HIGH PERFORMANCE COMPUTING AND COMMUNICATIONS

FOUNDATION FOR AMERICA'S INFORMATION FUTURE

A Report by the Committee on Information and Communications

National Science and Technology Council

Second Printing

THE WHITE HOUSE

WASHINGTON

MEMBERS OF CONGRESS:

I am pleased to forward with this letter "High Performance Computing and Communications: Foundation for America's Information Future" prepared by the High Performance Computing, Communications and Information Technology (HPCCIT) Subcommittee of the National Science and Technology Council's Committee on Information and Communications (CIC). This report, which supplements the President's Fiscal Year 1996 Budget, describes the interagency High Performance Computing and Communications Program.

America stands preeminent today among the nations of the world in its mastery and deployment of information technologies. Under the HPCC Program, government, industry, and academia have helped to lay the technological foundations for the National and Global Information Infrastructures and to maintain our global leadership in science and engineering. The "virtual agency" concept, which the HPCC Subcommittee has successfully implemented, is a key element of our collective effort to re-invent the Federal government to achieve greater efficiencies and productivity.

As noted in a recent National Research Council report requested by Congress to study the HPCC program: "Publicly funded research in information technology will continue to create important new technologies and industries, some of them unimagined today, and the process will take 10 to 15 years. The High Performance Computing and Communications Initiative is the main vehicle for public research in information technology." Sustained, long term commitment to the fundamental science and advanced technologies for our information future is essential, and this Program represents an opportunity to accelerate opportunities for innovation. The HPCC program has demonstrated its ability to solve the next set of long term problems in fast moving technology areas which is the key to laying the foundation for America's information future.

This past year marked the first full year of operation for the National Science and Technology Council (NSTC) and the successful development of a comprehensive Strategic Implementation Plan by the CIC. It is the beginning of an important dialogue to guide future public R&D investments in computing and communications. HPCC technologies are at the core of the overall strategy.

The HPCCIT Subcommittee and its member agencies have maintained a close working relationship with the private sector and academia, which is illustrated throughout the document. John C. Toole, Chair of the HPCCIT Subcommittee, Donald A. B. Lindberg, former chair, the many people in the HPCC agencies, their associates, and staff are to be commended for making HPCC a role model for streamlined, efficient, and highly effective government investment in advanced technology.

John H. Gibbons
Assistant to the President
for
Science and Technology

Table of Contents

Executive Summary

America is entering the Age of Information. The future of U.S. leadership in advanced computing, communications, and information technology will be determined by long-term research and development conducted by government, industry, and academia. The Federal High Performance Computing and Communications (HPCC) Program continues to be a driving force for advancing these technologies and their application to challenges in science and engineering. The Program is cited as a model for multi-agency planning and coordination and helps provide the foundation for future investments in information and communications R&D that is the cornerstone of the emerging Global Information Infrastructure (GII). This report complements "America in the Age of Information," the Strategic Implementation Plan prepared by the Committee on Information and Communications (CIC) of the National Science and Technology Council (NSTC), to which the HPCC Program reports.

HPCC advances enable U.S. leadership in science, mathematics, and engineering; economic growth and job creation; enhanced national security; improved environmental quality; a healthier, better educated citizenry; and effective harnessing of information technology. Information technologies are critical to the worldwide scientific community, our country's economic competitiveness, our national defense, and the quality of life of every American citizen. As the pace of advancement accelerates and competition increases, it is vital that we sustain U.S. leadership and accelerate the application of these technologies throughout our society. The Federal government plays a crucial role in this effort — the HPCC Program, working with industry and academia, is helping to meet the mission needs of participating Federal departments and agencies and the broader needs of the citizenry.

Congress authorized the HPCC Program with bipartisan support when it passed the High Performance Computing Act of 1991 (Public Law 102-194). Created as a dynamic R&D program, it has provided the sustained focus needed for developing these technologies and has adapted to the needs and opportunities of a changing world. The Program's major accomplishments, which are described in this document, provide the foundation for R&D the country needs as we prepare to enter the 21st century.

Since its inception, HPCC Program goals (listed on page 105) have focused on R&D in a wide range of high performance computing and communications technologies, including interdisciplinary and long-range problems for enabling solutions to an evolving set of computationally intensive Grand Challenges — fundamental problems in science and engineering that have broad economic and scientific impact. In addition, the Program focuses on developing technologies to enable the information-intensive National Challenges — fundamental applications that have broad and direct impact on the Nation's competitiveness and the well-being of its citizens, and advanced prototyping of a future integrated and interoperable GII capable of supporting societal needs.

Numerous reports, studies, and interactions influence the Program, including those of the participating departments and agencies, the National Research Council (NRC), the Government Accounting Office, industry, and academia. Responding to a request from Congress, in February 1995 the NRC released the report, "Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure." That report affirms the value of the Program, its contributions to the advancement of high performance computing and communications, and the need for increased emphasis on R&D of technologies for a large-scale, integrated information infrastructure.

As Dr. John H. Gibbons, Director of the Office of Science and Technology Policy, has said, "if information is power, HPCC will be the key to power for us as citizens and for the Nation as a whole." The lessons learned in the HPCC Program from attacking the Grand Challenges and from understanding the requirements of the National Challenges will enable the development of the GII.

In FY 1996, 12 Federal departments and agencies will participate in the HPCC Program by coordinating their R&D activities and accelerating technology transfer into key computationally intensive and information-intensive applications areas. The estimated FY 1995 HPCC Program budget for the nine participating Federal organizations was $1,038 million. For FY 1996, the President requested $1,143 million for the 12 organizations participating in the HPCC Program. The 12 organizations are:

ARPA	Advanced Research Projects Agency
NSF	National Science Foundation
NASA	National Aeronautics and Space Administration
DOE	Department of Energy
NIH	National Institutes of Health
NSA	National Security Agency
NIST	National Institute of Standards and Technology
VA	Department of Veterans Affairs
ED	Department of Education
NOAA	National Oceanic and Atmospheric Administration
EPA	Environmental Protection Agency
AHCPR	Agency for Health Care Policy and Research

The National Coordination Office (NCO) for High Performance Computing and Communications coordinates the activities of the participants through the High Performance Computing, Communications, and Information Technology (HPCCIT) Subcommittee of the CIC. On March 1, 1995, the founding Director of the NCO, Donald A.B. Lindberg, who served concurrently as Director of the National Library of Medicine at NIH, completed the successful term that he began on September 1, 1992. A new full-time NCO Director, John C. Toole, was announced jointly by Dr. Gibbons and Dr. Anita K. Jones, Chair of the CIC. The Director will coordinate the HPCC Program and the emerging information and communication initiatives as directed by the CIC.

The HPCC Program is the cornerstone of the CIC Strategic Implementation Plan, part of a dynamic planning process involving government, industry, and academia. That plan identifies six Strategic Focus Areas for R&D that will benefit the Nation's diverse users of information:

- Global-Scale Information Infrastructure Technologies that build advanced application building blocks and widely-accessible information services

- High Performance/Scalable Systems to support "high performance" and "low end" applications in a seamless fashion

- High Confidence Systems that will provide the availability, reliability, integrity, confidentiality, and privacy needed by the Nation's emerging ubiquitous information infrastructure

- Virtual Environments and simulations that will continue to transform scientific experimentation and industrial practice, and will play an increasingly important role in education and training

- User-Centered Interfaces and Tools to provide easier development, navigation, "mining," and general use of information resources

- Human Resources and Education both to educate the next generation of industrial and academic leaders in information science and technology, and to establish a foundation for new learning technologies

While the full impact of HPCC Program accomplishments may not be known today and may not be felt until the next decade, the Program's careful and balanced investments have already led to significant achievements. Some highlights from the many accomplishments reported in this document are:

- Based on research in software and networking technology, HPCC-funded communications technologies form the basis of the Internet, which continues its phenomenal growth in size, speed, and number of connected users.

- A combination of gigabit-speed (billions of bits per second) networking technology and computational science have demonstrated that massive scientific calculations can be executed across parallel processing systems located more than 1,000 miles apart. These speeds represents a thousand-fold improvement over the fastest networks in existence in 1991.

- Scalable parallel high performance computing systems are being developed that today represent performance that is a thousand-fold improvement over the most advanced systems available in 1991; have regained world benchmark records with speeds of more than 250 billion floating point operations per second; and promise improvements leading to meeting the goal of a trillion operations per second (teraops) performance on large scientific and engineering problems

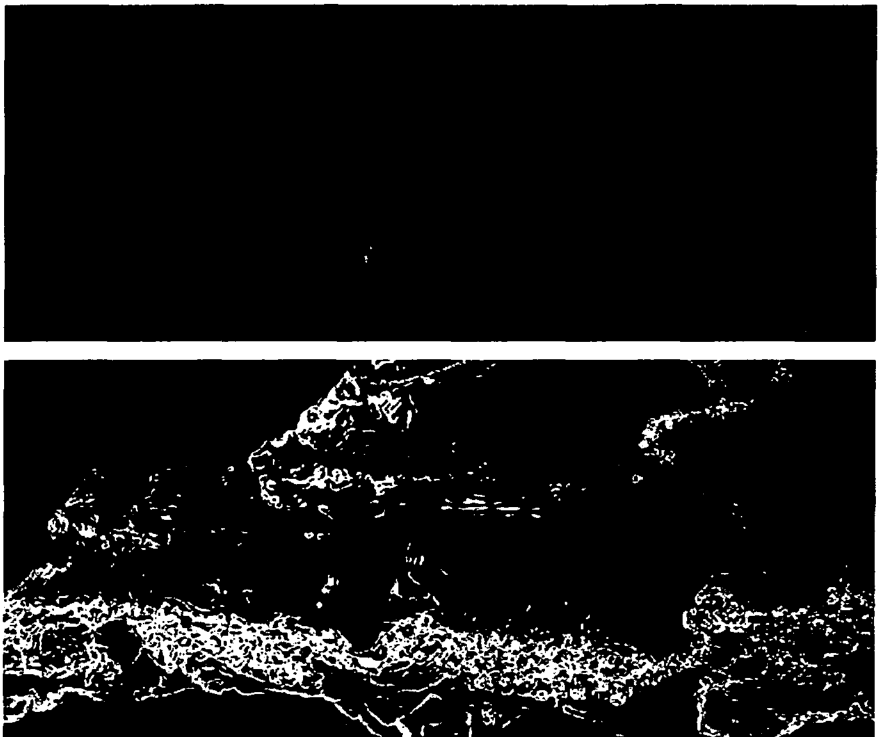
- HPCC software investments have been substantial. The National HPCC Software Exchange was established in 1994 to serve as a single software distribution point for peer-evaluated software, including new languages and tools, and benchmarking information.

- HPCC organizations manage more than a dozen high performance computing facilities that house parallel, vector, and scalar systems ranging from small to large scale; hybrid systems; workstations and workstation clusters; mass storage systems; visualization systems; and heterogeneous systems connected by high speed networks. These facilities provide environments in which systems performance can be analyzed and improved and new science can emerge. One example is the virtual reality CAVE, in which users are immersed in a manipulable three-dimensional virtual environment.

- Grand Challenges are fundamental problems in science and engineering that have broad economic and/or scientific impact and whose solution can be advanced by applying high performance computing techniques and resources. Use of HPCC technologies to address these problems has resulted in qualitative leaps in scientific understanding and represents a new paradigm of multidisciplinary research involving collaborative teams of computer and computational scientists and subject matter experts to solve more complex, more realistic problems in timely ways.

- In 1995, the Program began demonstrating select National Challenge applications and leveraging the HPCC investment through these demonstrations. National Challenges are fundamental applications that have broad and direct impact on the Nation's competitiveness and the well-being of its citizens, and that can benefit from the application of HPCC technologies and resources. Results from R&D in the Grand Challenges and in information infrastructure technologies are being integrated into these National Challenge applications to improve our productivity, the quality of our environment, and the quality of our lives.

- Technologies that directly support the missions of participating organizations are being demonstrated. For example, design change time has been significantly reduced using mediator technology, demonstrating improvements essential for streamlining Department of Defense weapons system acquisition and affordable design of large systems in industry.



Advances in Ocean Modeling due to HPCC

These figures illustrate the progress made by the HPCC Program. The top figure shows the highest-resolution global ocean simulations carried out just before the Program began. The approximation to the Earth's topography contained only three "islands," and the calculated ocean currents had only gross features. The calculations were insufficient to predict regional climates that require an accurate representation of the global ocean structure. Time sequences of these simulations were also inadequate for global climate prediction because few computing systems were powerful enough to carry out these simulations before 1992.

The bottom figure is the highest resolution simulation carried out in 1995 with the Parallel Ocean Program (POP) developed at DOE's Los Alamos National Laboratory. In this case the topography contains 80 islands. It shows ocean currents that transport heat between ocean regions (brighter colors represent higher velocities). The simulations were carried out on the 1,024-node Thinking Machines CM-5 supercomputer using approximately 10 X 10 miles resolution. Improved numerical methods and increased power of the CM-5 made possible longer and more realistic simulations that more accurately represent the ocean structure, as is evidenced from the existence of the known strong boundary currents such as the Gulf Stream. This model is the foundation for a new coupled atmosphere-ocean climate model that should reduce many uncertainties associated with model-based climate change prediction. Further information can be found at:

http://www.acl.lanl.gov/HPCC/climate_1.html

This project is jointly sponsored by the HPCC Program and the Federal Global Change Research (GCR) Program. It illustrates the beneficial symbiotic relationship between the two interagency programs.

- The HPCC Program has instituted educational programs and new curricula that have brought computational science into classrooms at all levels and increased the diversity of students pursuing scientific disciplines.

As cited by the NRC study, it often takes 10 to 15 years for publicly-funded research to reach commercial success. This document highlights a number of high performance computing and communications research efforts, begun a decade or more ago, that are now having profound scientific and economic impact on the Nation — examples include the Internet and R&D in global climate change and in weather forecasting. These examples point out not only the fundamental long-term contributions that federally supported research in information technology plays, but also the need to maintain fundamental research with long-term objectives in these critical areas.

The NRC report cited "the importance of retaining the HPCCI's [HPCC Initiative] momentum at just the time when its potential to support improvement in the nation's information infrastructure is most needed." The HPCC Program has provided a single national focus for high performance computing and communications technologies. This position is strongly endorsed by CIC's Strategic Implementation Plan and by industry groups.

As it goes forward, the Program will help create the very fabric of tomorrow's information-oriented society. Much remains to be done before a vigorous information economy can emerge. Systems and networks must have significantly higher throughput, bandwidth, security, and assurance to support the needs of the diverse user communities. The same infrastructure that spurs business will also carry education to both traditional learners and a new class of next-generation students. Leading-edge computational science and engineering must continue to discover new knowledge and develop new capabilities using the most advanced computing systems and communications networks. Digital libraries and collaboration techniques developed under the HPCC Program will be important enabling applications. Software must be easier to develop, execute, and adapt to all types of users. Computing systems must operate in more distributed environments. Both industry and government must be able to employ the most advanced information technology in the most affordable way possible. And, as always, well educated students will remain a critical component of technology transfer.

Our country must maintain leadership in computing and communications technologies, while applying them to our society through the Grand and the National Challenges. This in itself is not an easy task, and innovation in basic research, development, and the application of information technology will continue to yield significant new tools, techniques, and affordable applications in the decades to come. Solutions will rely on new system architectures, on advances in our understanding of the fundamentals of computational techniques and networking technologies, and on the ability of these endeavors to capture the insight and imagination of strong, well-trained minds. Only a carefully coordinated Federal effort, coupled to industry, academia, and the citizenry, can lay the long-term foundation for the Nation's information future.

NOTE: This book takes the form of a printed version and an on-line version available over the Internet. The on-line version was prepared using tools developed under the HPCC Program. It contains "hypertext links" to additional information that are identified by URLs (Uniform Resource Locators) in the printed version. These URLs change over time, and the on-line version will attempt to provide the most current links. That version can be found at the HPCC Program's "home page" whose URL is:

http://www.hpcc.gov/

How to use the Internet and World Wide Web browsers such as Mosaic to find out more about the HPCC Program

This document takes two forms -- a printed report and its companion on-line version. Both were created using the networking and information-finding tools developed by the HPCC Program itself. While the printed report summarizes many of the Program's major accomplishments, the on-line version presents a more comprehensive record. Internet users can find that version and other information about the Program by using an Internet/World Wide Web (WWW) browser such as Mosaic (developed at the National Center for Supercomputing Applications (NCSA)) to go to the following Uniform Resource Locator (URL):

http://www.hpcc.gov

That on-line version presents not only the text and images found in the printed report but also accesses the printed URLs to link to more detailed project descriptions from hundreds of HPCC research groups. Some of these include multimedia presentations. Because these groups update their information, the on-line version will be more current than the printed version. These URLs change over time, and the on-line version will attempt to provide the most current links.

The Internet is a global collection of interconnected, multi-protocol computer networks upon which these new capabilities run. Its evolution is described on page 10. The World Wide Web (WWW or Web) is a set of protocols and services for identifying and accessing files located on computer "servers" on the Internet. WWW was initially developed at CERN, the European laboratory for particle physics in Geneva, Switzerland, to meet data management needs of the high energy physics community. More information about the intent of the original development is at:

http://info.cern.ch/

These files are usually written in html (hypertext markup language), a subset of the federally- recognized Standard Generalized Markup Language. They may contain multimedia information (text, image, sound, and video) and hypertext, which is special text containing URLs that identify the locations of other files of (usually related) information located on any Internet/WWW server. Commonly an organization's server has a "home page" file that describes the organization and has hypertext "links" to other files on that server and elsewhere.

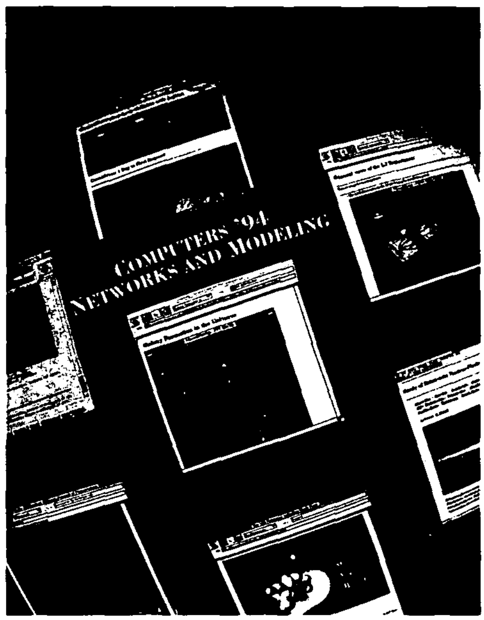
To access such information a user installs and runs an "Internet/WWW browser" (such as Mosaic, described opposite) on his/her "client," generally a workstation or personal computer that is connected to the Internet. The user requests information from a specific file on a specific server by providing the browser with the appropriate URL. The user can then follow the hypertext links in that file to other information. As these servers become numerous, the labyrinth of URLs becomes rich with conveniently accessible information. Material for this document was obtained by accessing such URLs, and the document in turn identifies more than 150 URLs. This method offers a new paradigm of how information can be easily organized, presented, and explored.

Mosaic — A browser that made a difference

The development of Mosaic, a user-friendly graphical Internet/WWW browser, is a particularly notable but unpredicted HPCC Program success. Mosaic allows a user to move through Internet/WWW information in an intuitive manner (either by simply "pointing and clicking" on the hypertext link, denoted by colored text, or on an icon or an image, or by explicitly naming the URL), and the connection is transparently made using http (hypertext transport protocol). The NCSA Mosaic home page has links to "What's New" files containing date-stamped announcements of new servers or new information on existing servers:

http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/NCSAMosaicHome.html

The announcement in November 1993 of the availability of Mosaic on the Macintosh, PCs, and Unix workstations has been followed by unprecedented activity by organizations setting up servers and users accessing them. Today nearly 10,000 organizations have Internet/WWW/Mosaic servers — these include organizations within the Federal government and other governments, corporations, and schools (from universities to elementary schools). The NCSA Mosaic client software, which is freely available worldwide, is being distributed over the Internet at the rate of about 100,000 copies per month. WWW traffic on the Internet is growing faster than any other kind of traffic. In FY 1994 NCSA entered into agreements enabling commercialization of Mosaic, and several products have been marketed. NCSA maintains a public-domain reference implementation that has stimulated both research in the field and innovation in the commercial market.



The cover of Science *August 12, 1994, which has special coverage on Computing in Science. An article on pages 895 to 901 by B.R. Schatz and J.B. Hardin contains an account of the development of WWW and NCSA Mosaic. Reprinted with permission from* Science *volume 265, August 12, 1994, cover. Copyright American Association for the Advancement of Science.*

The White House home page is:

http://www.whitehouse.gov/

Some U.S. Congress and Library of Congress URLs are:

gopher ://gopher.senate.gov/

http://www.house.gov/

http://thomas.loc.gov/

Introduction

America is entering the Age of Information. The Federal High Performance Computing and Communications (HPCC) Program continues to be a driving force for progress in computing, communications, and information technologies; their application to fundamental challenges in science and engineering; and the R&D needed to realize the emerging Global Information Infrastructure (GII). As part of the Committee on Information and Communications (CIC) of the National Science and Technology Council (NSTC), the HPCC Program provides an important component of America's investment in its information future.

Congress authorized the HPCC Program with bipartisan support when it passed the High Performance Computing Act of 1991. Created as a dynamic R&D program, it has provided the sustained focus needed for developing these technologies and has adapted to the needs and opportunities in a changing world.

The spectacular images transmitted around the world of the recent collision of the Shoemaker-Levy 9 comet with the planet Jupiter illustrate the impact of HPCC technologies across multiple Federal agencies. It is an example of enabling new science and computational models, run on a new generation of scalable computing systems, communicated almost instantaneously on a web of networks, using the Hubble space telescope in unforeseen ways. This document describes numerous such major accomplishments and plans as the Program enters its fifth year. Some other highlights include:

- High performance communications R&D has exploded from the evolution of the Internet to a new leading edge focus on high speed technologies for interoperable very high performance systems around the globe.

- High performance computing systems have brought scalable parallel computing into a new era; today's focus is on new operating systems capabilities such as real-time systems, new component technologies, embedded systems, mass storage, and specialized high performance architectures.

- Advanced software technologies R&D continues to make marked progress in microkernel operating systems, programming languages, and tools for developing software for parallel computing systems. Progress in scalable I/O, computational techniques, and performance measurement are key to on-going advances in addressing the Grand Challenges.

- Enabling technologies for information infrastructure have focused on distributed computing; reliability; mobility; security and privacy; and tools for building distributed applications.

- Interdisciplinary Grand Challenges R&D has led to new science and engineering in a wide range of disciplines. By modeling aircraft and their engines, combustion, the oceans, the atmosphere, the weather, pollution, climate, groundwater, earthquakes, vegetation, the human body, proteins, enyzmes, the human mind, materials, chemicals, car crashes, the galaxies, or a comet colliding with Jupiter, high performance computing and communications have brought the Nation new knowledge and new abilities.

- The Program strongly supports enabling technologies for National Challenge applications, with emphasis on digital libraries and long-term electronic commerce technologies. Specific advanced applications include computer-based patient records, manufacturing testbeds, and public access to scientific information.

- Basic research and education remain a foundation of the Program. Fundamental new ideas are being explored across the participating organizations. This report documents the strong commitment to education and training.

This supplement describes the accomplishments and plans of dedicated and creative scientists, engineers, individuals, and organizations throughout the U.S. It is organized to showcase both technologies and major application areas across the five major components of the Program (listed on page 107). More information about individual research efforts is available on line as described on the next two pages. Such direct and immediate connectivity between producers and users empowers individuals and groups across the U.S. and around the world.

HPCC Program Accomplishments and Plans

1. High Performance Communications

High performance communications activities are a highly productive HPCC Program area. During the past year we have seen continued explosive growth in the use of high performance communications for information access, research at the frontiers of science, and commercial endeavors. High performance communication activities include continued research and development in internetworking technologies to support these new communities and applications, research into network systems operating at a billion bits per second and faster, wireless technologies, integrating computing and communication across wide areas, and enhanced Internet connectivity.

Internetworking R&D

The Internet was created by ARPA, and that agency, other HPCC agencies, and industry and academia have contributed to its evolution and phenomenal growth (details about its evolution are given on page 10). As new communities of education, commercial, public sector, and individual users join the Internet, fundamental changes must be made to support increased use, new applications such as video and imagery, service quality, security, and new modes of communication. Internetworking R&D supports these activities across the HPCC Program.

Work at ARPA, NSF, and other HPCC agencies in protocols is directed at ensuring that different types of networks continue working together as a cohesive system. New services, such as multicast and guaranteed service quality, are being created, evaluated, and introduced. Research into new routing and scaling techniques for extremely large networks anticipates an Internet that interconnects millions of sub-networks and billions of end users across an increasingly diverse set of media. Network security from the perspectives of protecting the network and protecting users is being incorporated into the fundamental operating principles.

- In FY 1995 ARPA will demonstrate network bandwidth, delay, and service reservation guarantees to support new services, such as video, imagery, collaborative work, and interactive data services. In conjunction with other DOD agencies, ARPA will test and evaluate SONET and ATM (described on page 10) encryption technologies at 155 Mb/s (megabits per second or millions of bits per second). Also in FY 1995, next generation multicast-based services with improved ease of use will be demonstrated.

- In FY 1995 NSF plans to support integrative research, including experimental projects, that spans computer architecture, operating systems, compilers, I/O systems, network interfaces, and networks. The objective is to reduce latency and increase performance of the network integrated computing system.

- In FY 1995 DOE will support R&D and deployment to integrate scalable I/O with network R&D.

Gigabit-Speed Networking R&D

A major HPCC activity is the support of five gigabit testbeds jointly funded by ARPA and NSF with support from DOE through their national laboratories, and a sixth testbed funded by ARPA. Nine Federal agencies, thirteen telecommunications carriers, twelve universities, eight corporations, and two state supercomputer centers participate in these six testbeds that connect 24 sites. Transmission equipment and fiber optic cabling is provided by the carriers and industry at no cost to the Federal government. An important success of the gigabit networking project is the high level of cooperation among the academic, industrial, and governmental participants leading to new capabilities in the sci-

The Evolution of the Internet

The Internet began in the late 1960s with the development of ARPANET. ARPANET provided a limited number of researchers with shared, interactive communications between computing systems at different locations and developed key innovative technologies on which the Internet still depends. Some of these include:

- "Packet switching" in which a message is divided into packets that are transmitted to their destination and reassembled — rather than using a dedicated circuit as in the telephone system or "store and forward" that is used in the telegraph system

- A distributed rather than a centralized network — even when some nodes fail, the remaining network continues to function, although perhaps at reduced capability

- Adaptivity — packets can be transmitted to the same destination over different parts of the network based on current network load and connectivity

- The Transmission Control Protocol (TCP) and the Internet Protocol (IP) were developed in the 1970s and early 1980s under ARPA sponsorship to enable different networks to interoperate. TCP/IP made it possible for ARPANET, the ARPA packet radio network, and the packet satellite network, each with different packet sizes and network transmission speeds, to interoperate.

While initially an experimental system, the need for a stable network emerged as the user community grew. MILNET was split from the ARPANET for use by Department of Defense users. Federal agencies established networks to support their R&D communities: NSF created NSFNET for the university research community, NASA built the NASA Science Internet (NSI), and DOE created the Energy Sciences Network (ESnet). These networks and an ever growing profusion of local area networks and regional networks form the HPCC component of the Internet, a "network of networks," and demonstrate the fundamental strength of new networking technologies created in the early ARPANET.

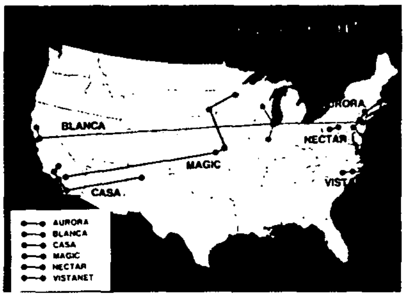
The recent phenomenal growth of the Internet beyond the HPCC community is the result of educational, public service, private sector, and personal investment. Today there are more than 27,000 networks in the U.S., an increase of 350 percent over that reported on these pages a year ago. Non-U.S. networks number more than 21,000, an increase of 225 percent. More than 1,500 U.S. colleges and universities, 1,500 high schools and 1,700 elementary schools in the U.S. have full Internet connectivity.

The HPCC Program has directly stimulated the emergence of a vigorous and highly competitive private sector industry in Internet hardware, software, and connectivity in which the U.S. is a world leader.

ence community, new markets for the industrial community, and new visions for high performance communications.

http://WWW.CNRI.Reston.VA.US:4000/public/gigabit.html.

One of the technologies further developed and deployed in the testbeds is Asynchronous Transfer Mode or ATM, a "fast packet switched" cell relay technology in which small packets of fixed 53-byte size are rapidly routed over a network. ATM is a technology being investigated by the HPCC Program and the U.S. telecommunications industry to integrate data, voice, and video services using a single protocol. "ATM/SONET" refers to transmitting ATM packets over a Synchronous Optical Network



HPCC-supported gigabit testbeds funded jointly by NSF and ARPA test high speed networking technologies and their application in the real world. Courtesy Corporation for National Research Initiatives.

(SONET) link, a standard for transmitting data across fiber optic cable.

Some of the accomplishments of the gigabit testbed project are:

- Operation of high speed data transmission systems — The testbeds used SONET systems to transmit data bits over optical fibers at speeds not previously achieved in long distance networks. Most testbeds operated at 2.4 billion bits per second (gigabits per second or Gb/s) across thousands of kilometers. A typical comprehensive encyclopedia contains about one billion bits.

- Transport protocols for wide-area networks — The testbeds experimented with new very high-speed transport protocols. A number of testbeds used ATM as their technology for transport and switching. Others adapted the High Performance Parallel Interface (HiPPI), a standard interface for accessing supercomputers, to transport data between sites. New transmission records of 500-800 Mb/s between hosts over long distances were set with these transport protocols.

- ATM switches — New ATM switches were developed to route data at high speed across the testbeds. Maximum network and switching speeds at the start of the project were T3 (45 Mb/s). The testbeds required higher speed switches and pushed switching to 155 Mb/s, 622 Mb/s, and faster.

- Host interfaces — Computing systems vendors have delivered fast processors, but the development of fast communication interfaces to attached devices such as mass storage systems and to high-speed networks has lagged. Testbed researchers worked independently and with vendors to increase the speeds of host I/O and interfaces to the network. For example, one group demonstrated 400 Mb/s between two Digital Equipment Corporation (DEC) Alpha workstations using ATM/SONET interfaces.

- Distributed shared memory — Network speeds increased, and improved host interface throughput enabled new approaches to sharing information across the network. Two testbeds explored implementations of a single virtual address space across multiple computers, significantly increasing the capability of a distributed computer system.

- Several groups of researchers developed prototype protocols to address network demands such as high speeds and additional performance requirements such as efficiently using network resources while providing different qualities of service guarantees needed by different types of applications. One group designed and demonstrated Tenet, a real-time protocol suite that delivers required quality of service bounds specified by the application, including delay jitter, throughput, and packet loss probability bounds. A second group developed (1) the Dynamic Time Windows protocol to avoid, and where necessary control, congestion for end-to-end flow control, and (2) Network Traffic Reporting that includes algorithms for measuring and controlling traffic and for scheduling and load control. A third group worked on cell-based integrated services at gigabit speeds to address similar problems using ATM.

- Several application environments were developed to facilitate using distributed computing systems. These include the Data Transfer Mechanism (DTM), Express, and Dome. DTM is a message passing facility to simplify interprocess communication and facilitate creating sophisticated distributed applications in a heterogeneous computing environment. Express is an integrated software package that addresses all phases of the application development cycle in an architecture-independent manner, including algorithm design, implementation, debugging, and performance analysis. The Dome (Distributed Object Migration Environment) project is developing libraries of distributed objects (that is, scalable software) for producing applications software for parallel execution on networks of computing systems.

- Applications — All the testbeds included demanding applications that drive gigabit speed network development and provide criteria for evaluating effectiveness. These applications include chemical dynamics, global climate modeling, medical treatment planning, radio astronomy, and terrain visualization. Results from distributing select applications over multiple heterogeneous computing systems in the CASA testbed show superlinear increases in performance: that is, two computers working together solve the problem in less than half the time of either single computer.

Descriptions of four gigabit networks, along with an image from a VistaNET application, follow.

Irradiation of healthy tissue was reduced in an experimental radiation therapy medical treatment planning application that was distributed over a 622 Mb/s HiPPI/ATM/SONET network in the VistaNET testbed. Calculations were performed on a Cray Research Y-MP and displayed on a Pixel Planes workstation. This application provided data for real world network traffic analysis.

CASA

CASA is a 2.4 Gb/s SONET testbed connecting the NSF San Diego Supercomputer Center (SDSC), the DOE Los Alamos National Laboratory (LANL), NASA's Jet Propulsion Laboratory (JPL), and California Institute of Technology (Caltech). Each of these sites has significant computing resources, and CASA focused on interconnecting these supercomputers. While LANL and the other sites are about 1,600 km apart, the link is 2,000 km long. This link holds world speed records of 500 Mb/s for TCP/IP file transfers and 792 Mb/s for raw HiPPI (High Performance Parallel Interface), a world record for host-to-host bandwidth-distance product of any ground-based testbed. The distance and speed presented new problems for network researchers. For example, in CASA there are over 800,000 bytes of data in flight at any instant. Any error in transmission causes unacceptable delay in processing. The CASA HiPPI/SONET gateway incorporates a forward error correction scheme to avoid re-transmission of data when an error is detected. The CASA testbed allows two groups of climate researchers, one in California and the other in New Mexico, each working on different parts of a problem, to combine their software models into a single metacomputer-based execution using heterogeneous computing systems.

http://www.noc.lanl.gov/lanp/project.html#casa

MAGIC

The MAGIC (Multidimensional Applications and Gigabit Internetwork Consortium) project was established to develop a very high speed, wide-area networking testbed to address challenges in heterogeneous computing, distributed storage, coordination of multiple data streams, and techniques for managing the effects of network delays in a defense applications context. The MAGIC team demonstrated TerraVision, a terrain visualization application that allows a user to view and navigate through a photorealistic landscape. With the addition of GPS-derived (GPS stands for Global Positioning System) data representing positions of units and vehicles, multiple commanders can get bird's eye views of situation and flow during training exercises as shown below. TerraVision images are stored on a distributed set of Image Server Systems (ISSs) connected by a high speed network. ISS images are transmitted to displays at Fort Leavenworth, KS, or Overland Park, KS. Sustained rates of several hundred Mb/s have been demonstrated.

http://www.magic.net/

To create a three-dimensional terrain scene in real time, distributed storage servers transmit data streams containing portions of the image in parallel over a high speed network to a user's workstation. In this view of a military training facility, the real-time vehicle position is shown by the blue line.

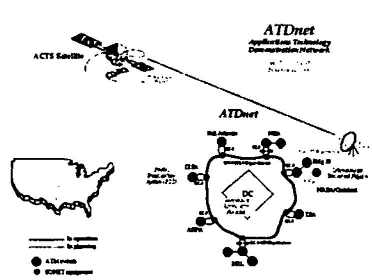
ATDnet

The Advanced Technology Demonstration Network (ATDnet) is a recently inaugurated high performance networking testbed in the Washington, DC area. It is intended to be representative of possible future Metropolitan Area Networks. Initially architected by NSA and the Naval Research Laboratory and funded by ARPA to enable collaboration among DOD and other Federal agencies, ATDnet has a primary goal to serve as an experimental platform for diverse network research and demonstration initiatives. ATDNet provides an opportunity for early access and familiarization for the DOD, especially those components dependent on state-of-the-art information technology. Emphasis is on early deployment, operation, and management of emerging ATM/SONET networks.

Six Federal agencies participate. ARPA and the Defense Information Systems Agency (DISA) serve as co-chairs of the ATDnet program. Other members are the Defense Intelligence Agency (DIA), NASA, the Naval Research Laboratory (NRL), and NSA.

In FY 1995 ATDNet is used for experiments in telemedicine, distributed simulation, security and encryption techniques, ubiquitous video teleconferencing, large ATM network signaling research, ship design and analysis visualization, and network experiments. New network management techniques for SONET and ATM systems are being developed that will give DOD fast-start access to equivalent commercial services once they are tariffed.

http://www.disa.atd.net/disacfe/disacfe.html



ACTS

In FY 1994 NASA launched its Advanced Communications Technology Satellite (ACTS). ACTS provides a "network in the sky" and is capable of multiple low speed access, four 155 Mb/s full duplex connections, or a single 622 Mb/s full duplex connection. In FY 1995 NASA expects to demonstrate gigabit speed applications using ACTS in two experiments: "A Performance Study of an ATM/SONET Satellite and Terrestrial Network for an Engine Inlet Simulation," and "High Data ACTS Experiment for Performing Global Science: Keck (Hawaii) Telescope and Global Climate Model." ARPA and NASA will also demonstrate interconnection of terrestrial gigabit networks in FY 1995.

http://cesdis.gsfc.nasa.gov/hpccm/accomp/94accomp/ess94.accomps/ess3.html

http://kronos.lerc.nasa.gov/acts/experiment-program-overview-6.html

Additional gigabit networking R&D includes:

- Beginning in FY 1995 ARPA plans to develop advanced opto-electronic network component technology and network architecture for scalable and modular networks. The aggregate network bandwidth will be in the range of terabits per second, and the network will handle multimedia service for both digital and analog signals.

- ARPA and NSF will continue to co-manage the 10 Gb/s testbed at Washington University initiated under the Technology Reinvestment Program (TRP).

- In FY 1995 and FY 1996 NASA and DOE plan to demonstrate interoperability among independently managed ATM-based networks supplied by multiple vendors and used by multiple agencies for collaborative efforts. DOE will demonstrate constant bit rate services and other non-traditional data usage over ATM.

- NSA has a perennial requirement for the fastest networking technology in order to satisfy the National Security Grand Challenge. The goals of its networking efforts are to realize multi-gigabit per second trunking speeds and to support sustained data flows of at least hundreds of megabits per second. In the longer term, the work aims to achieve sustained data flows of many gigabits per second. In FY 1995 NSA will install an in-house optical crossbar network testbed; beginning in FY 1996 NSA will conduct R&D in high speed optical network architecture and optical network management techniques.

- In FY 1995 and FY 1996 NIST will collaborate with the ATM Forum to develop standardized conformity tests for ATM-based technology, and will collaborate with industry and users to establish an ATM testbed to test evolving products. NIST R&D includes simulation and design to assess interactions among reliability of data transmission, flow control, congestion control, and the right to transmit information, in order to reduce the likelihood of losing information during periods of high demand. Design flaws and ways to minimize data loss will be reported to developers and to appropriate standards organizations.

Wireless Technologies

As part of its program in National-Scale Information Enterprise, ARPA is extending the information infrastructure into the mobile environment. This includes wireless communications and networking technologies described here, and supporting services for mobile information systems (described on page 29):

- An environment is being implemented that allows a high density of in-building mobile users to simultaneously access multimedia computing systems and information servers connected to high bandwidth network backbones. In this environment a lightweight portable multimedia notebook called the InfoPad is used to access a pico-cell wireless network.

http://infopad.eecs.berkeley.edu

- A design environment for the wireless networking environment is also being developed: a point-to-point wireless multimedia link using a PC notebook has been demonstrated, validating the overall design approach.

  http://millennium.cs.ucla.edu/wamis.html

- A high speed ATM-based wireless communication system is being developed that will be adaptive at both the network and the host interface levels. The system will allow for rapid deployment and response to a changing environment, thereby realizing the advantages of ATM for wireless networks.

  http://www.tisl.ukans.edu/RDRN/

In FY 1996 NSF will support basic research in wireless networks, including transmission, networking, and computing issues. The objective is to enable network access in a secure, seamless, and highly mobile environment, and to enable the handling of voice, video, and data communications.

In FY 1995 and FY 1996 NIST plans to assess the vulnerabilities and security requirements of emerging wireless technologies and to assess performance and establish benchmarks.

R&D for Network Integrated Computing

The computing systems and instruments used in research on the Grand Challenges and the National Challenges, as well as on Defense Challenges, are located throughout the country, and the researchers who use these resources travel around the world. One goal of the HPCC Program is to develop a networking and computing environment in which the number of researchers working on a problem and their locations are incidental to conducting the research. Another goal is to enable the extension of such an environment to an ever larger user community. Activities include the following:

- ARPA is developing technologies to move information efficiently among applications across widely distributed networks.

- ARPA is enabling dynamic configuration of networked computing resources; resource scheduling; toleration of failures peculiar to a wide area environment; toleration, avoidance, or hiding of unavoidable latency in a wide area network; and work with heterogeneous resources and across different administrative domains.

- In FY 1995 ARPA expects to integrate DOD and commercial networks and demonstrate crisis management technologies.

- DOE has on-going research in packetized video and voice, telecommuting, on-line facilities access and control, and collaborative workspaces. In FY 1995 DOE expects to release a version of its collaborative work technologies to other agencies for review and use and to complete packetization of commercial codex video/audio output that allows ESnet to carry packetized video encapsulated in IP (Internet Protocol). Such MBONE (multicast backbone) services are used in some 135 conferences or meetings each month and support NASA's multicast from Antarctica and the Space Shuttle as well as multicasting the Internet Engineering Task Force (IETF) and other Internet standards activities.

- In FY 1996 DOE plans to design and prototype protocols, techniques, and mechanisms to support enhanced network and workstation-based audio/video capabilities as well as virtual and electronic collaborative "lab spaces."

  http://www.mcs.anl.gov/home/stevens/labspace/root.html

- The DOE-funded project called ComPaSS (Communications Package for Scalable Software) at Michigan State University addresses the design of communications libraries and supporting tools for distributed-memory computing platforms. A better understanding and implementation of collective communication for wormhole-routed networks, tools for data redistribution and resource management on workstation clusters, and host interface software for distributed computing across ATM networks from this work are evidenced in the incorporation of Michigan State's wormhole routing techniques into commercial switches as well as the development of analytical network tools to monitor and measure various application traffic implementing collective communications (multicast) over both raw ATM and TCP/IP over ATM.

http://www.cps.msu.edu/~mckinley/ComPaSS/

Enhanced Internet Connectivity

At the beginning of the HPCC Program in 1991, NSF was upgrading the NSFNET backbone to T3 speed, and by 1992 full T3 backbone service was operational. This was the first network to provide T3 service. Today, as the NSFNET program prepares to switch over to a new architecture described below, commercial network service providers are offering T3 backbone service as are several regional networks. In fact all regional networks routinely provide T1 (1.5 Mb/s) service or better on their backbones.

As the Internet community grows, higher bandwidth and enhanced service networks are needed to handle more users, the increased size of transmitted objects (for example, images and video), and new applications such as collaborative work in which voice, video, and interactive data are transmitted. In the next year HPCC agencies will continue to enhance Internet capabilities. Examples are:

- By the end of FY 1995 NSF will complete the transition to a new architecture that consists of a set of Network Access Points (NAPs), funded by NSF, and accessed by network service providers (NSPs) typically at T3 speeds. NSPs will provide connectivity to the NAPs from the regional networks. A Routing Arbiter, also funded by NSF, will assist the cross network traffic at each NAP.

- In FY 1995 NSF will implement a very high speed Backbone Network Service (vBNS) to link the NSF Supercomputer Centers and NAPs at 155 Mb/s. NSF is issuing a solicitation in FY 1995 for provision of additional international Internet service at higher capacities than at present. Finally in FY 1995 NSF plans to announce a new Connections Program to support the use of innovative technologies for schools and libraries, to continue its regular program of funding connections for institutions of higher education, and to fund high bandwidth Internet connectivity for qualifying applications.

- DOE is expanding its ESNet with commercial ATM service in cooperation with NASA. DOE will upgrade ESNet to 45 Mb/s ATM services and will begin upgrading to 155 Mb/s at up to four sites in FY 1995. ESnet provides worldwide access to Energy Research facilities including Light Sources, Neutron Sources, Particle Accelerators, Spectrometers, Genome Centers, and data bases.

  http://www.es.net/

- In FY 1995 NASA completed its upgrade to T3 connectivity and began its upgrade to 155 Mb/s among five NASA centers using a commercial ATM service: Ames Research Center at Moffet Field in Mountain View, CA; Goddard Space Flight Center in Greenbelt, MD; Langley Research Center in Hampton, VA; Lewis Research Center in Cleveland, OH; and the Jet Propulsion Laboratory (JPL) in Pasadena, CA. In FY 1996 NASA will demonstrate 622 Mb/s interoperation between Goddard and JPL.

- In FY 1995 at least 24 NOAA National Data Centers will be connected to the Internet through high bandwidth links; and in FY 1996 bandwidth will be increased to accommodate rapidly increasing user demand.

HPCC agencies continue to connect more users in the research and education communities:

- NSF, the lead HPCC agency in connecting academic institutions, continues to fund connections for institutions of higher education at the rate of about 100 per year.

- At NIH, the National Library of Medicine's (NLM) Medical Connections program, in conjunction with NSF, provides "jump start" funding to academic medical centers, community hospitals, and other health care organizations to connect to the Internet. 170 U.S. medical schools and health care facilities have been connected during the past three years under the NSF and NLM Connections Program. Special emphasis is given to linking medical libraries with health care delivery organizations and database servers in support of timely and accurate clinical decision making. The program also supports the creation of regional consortia of health care institutions for sharing medical information and distribution of Internet capability within an institution. Approximately 15 grants are being awarded in FY 1995. The overall goal is to provide Internet connectivity to the top 3,000 health care institutions in the U.S.

  http://www.nlm.nih.gov/ep.dir/rfa_internet_conn.html

- The Department of Veterans Affairs (VA), an agency new to the HPCC Program in FY 1996, has an agency-wide network used to exchange medical and other data. This network is connected to the Internet via gateways, providing expanded functionality, wider access, and greater security. Additional capacity for transmitting multimedia data and using browsers such as Mosaic is being implemented.

- EPA has a dedicated T3 line between its National Environmental Supercomputing Center in Bay City, MI, and its National Computing Center and HPCC research network in North Carolina; this network facilitates technology transfer to state environmental groups.

HPCC agencies also continue to provide and upgrade Internet connectivity to other countries:

- In FY 1995 NSF's International Connections Management Program (ICM) upgraded service to Stockholm and to London to 4 Mb/s each, and service to Paris to 2 Mb/s. Several additional Latin American and Caribbean countries are connected to/or will presently connect to the ICM infrastructure. In the Pacific, Malaysia upgraded its connection from 64 Kb/s (kilobits per second or thousands of bits per second) to 1.5 Mb/s; at least two 64 Kb/s connections from the People's Republic of China were made; and Japan NACSIS upgraded its connection to 2 Mb/s. South Africa doubled its bandwidth from 128 Kb/s to 256 Kb/s. In all, about 40 countries are connected to the Internet by means of the ICM infrastructure and with the exception of the European connections, the connectivity is largely paid for by the other countries (NSF shares cost with the European countries on those connections).

- In FY 1995 DOE is upgrading links to Japan to 512 Kb/s, adding a 128 Kb/s link to Germany, installing data compressors on the link to Italy, and adding links to Russia, all in support of multi-national collaboration such as the International Thermonuclear Engineering Reactor.

2. High Performance Computing Systems

When the HPCC Program began in 1991, traditional vector computers were still the primary high performance computing systems. Those systems were known to be approaching their physical limits, and a number of computing systems vendors were developing parallel systems that promised to overcome those limits. Today all major U.S. vendors have adopted parallel technology. Products span a wide range, including scalable parallel, fine- and coarse-grained parallel, vector and vector/parallel, networked workstations with high speed connectivity, and heterogeneous systems connected by high speed networks.

Performance Accomplishments

One of the fastest of these systems, an Intel Paragon with 1,904 nodes (each consisting of two Intel i860 processors) and 38 GB memory at

DOE's Sandia National Laboratories, achieved a world record 143.4 gigaflops (Gflops) on the Massively Parallel (MP) Linpack benchmark.

By loosely coupling two large Paragons with a total of 2,256 of Intel's new multiprocessing nodes (each consisting of three i860 XP microprocessors, for a total of 6,768 processors), world records of 281 Gflops on the MP Linpack benchmark and 328 Gflops on a double-precision complex LU factorization code were realized. The systems were connected by 16 HiPPI channels with over 3 GB/s capacity. Both ran SUNMOS (Sandia/UNM Operating System), whose development was funded by DOE. SUNMOS is a lightweight custom message-passing OS kernel optimized for high performance and low latency. Applications demanding this level of performance are described beginning on page 41.

http://www.cs.sandia.gov/HPCCIT/main.html

Applications exhibiting half teraflops performance over a network are expected to be demonstrated on the exhibition hall floor at Supercomputing '95 in November 1995 in San Diego. The HPCC Program remains well on track toward demonstrating technologies capable of sustaining one trillion operations per second (teraops) performance on select large scientific and engineering problems.

Some of the members of the team that broke the world computing systems speed record.

ARPA is the lead HPCC agency developing the scalable computing technologies capable of sustained teraflops and faster performance. The ARPA program will enable industry to develop such systems for broad commercial use while enabling defense agencies to procure large-scale versions from the low-cost commercial technology base for special applications without redeveloping architectures, operating systems, or applications software. In FY 1995 ARPA will support the hardware, software, and system architecture design of teraflops-scale systems expected to emerge in FY 1996. Efforts include support for the convergence of architectural models (for example, message passing and distributed shared memory, or scalable vector and massively parallel). That will result in systems that run critical defense parallel applications more efficiently (for example, automatic target recognition, sensor fusion), and critical mobile targets. These systems will take advantage of ARPA's efforts in microsystems that are expected to produce computer backplane data rates approaching one Gb/s per wire. FY 1995 ARPA efforts will also create early prototypes of significantly higher performance and secure operating systems. In workstation cluster computing, ARPA will explore this more loosely coupled scalable computing architecture, including work on resource discovery, operating environments, and new models of computing.

Microsystems

ARPA's microsystems program provides the scalable microarchitecture building blocks specifically targeted at key limitations in power and computational efficiency for defense embedded computing applications for next generation computing systems. FY 1995 efforts include design of programmable protocol accelerator integrated circuits that support multiple communications protocols on the backplanes of next generation systems. Efforts in high speed signaling will result in improvements in inter-node communication bandwidth. New microarchitectures that use VLSI (very large scale integration) technology are being investigated as part of an effort to develop parallel-friendly architectures that support security, work cooperatively with compilers, have wider bandwidth to memory, and have new caching strategies. Microsystems also support CAD tools and environments for designing complex digital systems. Computational prototyping is used to accelerate the design process: in FY 1995 it will be used in virtual design environments for atomic-level to system-level modeling.

Embedded Systems

Work in embeddable high performance computing systems enables the transfer of high performance scalable technologies to military applications. Prototypes of embedded versions of emerging high performance computing systems lower future risk. In FY 1995 efficient programming environments will be developed for this class of systems. Real-time operating systems that meet performance, size, and security needs will emerge. Open interfaces for scalable computing systems will be developed, including new communications standards enabling low latency high performance interconnection of elements in a heterogeneous distributed system for embedded applications. In FY 1995 interoperability mechanisms that insulate application development from the hardware system will be explored. Scalable instrumentation will be developed to speed the debugging process and to validate performance. In FY 1996 ARPA plans to prototype scalable embedded modules containing memory and power on a single unit of replication.

Networks of Workstations (NOW)

In an ARPA-funded effort at the University of California at Berkeley, the NOW project is developing hardware and software support for using a network of workstations as a distributed computing system on a building-wide scale. This project addresses distributed scalable computing technologies typical of large embedded military programs. Advances in switched networks such as ATM have made it possible to closely integrate processors, memory, and disks. This approach connects a cluster of workstations into an integrated high performance system almost as powerful as traditional supercomputers. Parallel scientific software, computer-aided engineering software, databases, file servers, and large scale commercial information servers can benefit. R&D issues include network interface hardware, communications protocols, network-wide resource management, distributed scheduling, and parallel file systems. Features of a 100-workstation system that will be demonstrated include (1) high performance (delivering large portions of the capacity to demanding sequential and parallel applications while guaranteeing good performance to interactive users), (2) incremental scalability by adding workstations, (3) fault tolerance (the system remains usable even when a workstation fails), and (4) easy administration. Commercially available systems are used for fast prototyping.

http://now.cs.berkeley.edu/

Rapid Prototyping Facility

R&D in high performance computing systems involves government, industry, and academia. An example is the NSF-funded high performance rapid prototyping facility at the University of Michigan that can be accessed over the Internet. In FY 1995 NSF began funding university-based R&D in the design and analysis of memory architectures, and in the use of high performance computing in prototyping and manufacturing. In FY 1996 the agency plans to support additional research addressing the continuing imbalance between processor and memory speeds, which is becoming a major roadblock to advances in high performance computing systems. In FY 1996 NSF's Engineering Directorate plans to fund university-based R&D in optical and optoelectronic technologies that will enable future advances in ultra high capacity computing and communications. Together with NSF's Computer and Information Science and Engineering Directorate (CISE), they will begin to support the implementation of wireless network architectures and their interface to optical networks; research on integration at the interface of device and system can help accelerate the implementation of these technologies.

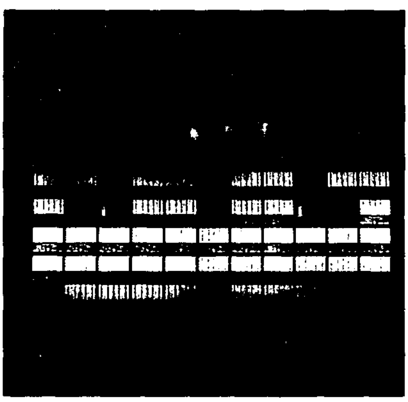
Specialized Very High Performance Architectures

NSA R&D is directed at order of magnitude improvements for deriving information using mathematical and signal processing approaches. Activities include:

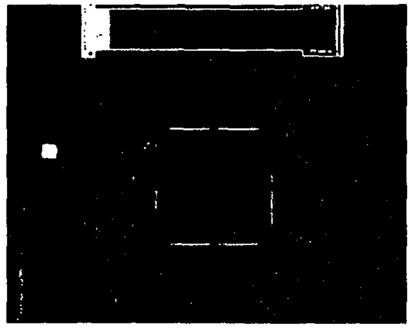
- Prototyping a PIM (Processor in Memory) system integrated into a Cray Computer Cray-3 high performance vector system and making it available to NSA and non-NSA users

- Continued R&D of SPLASH, a field programmable gate-array-based architecture that is attached to a workstation and uses software to reconfigure the hardware to match an application's computation, thereby providing high performance at low cost

- Continuing R&D in heterogeneous data repository software, workstation-class systems with very high speed interfaces and switches, distributed operating systems, a next generation super massively parallel processor computing system, and superconducting electronics



A wafer containing multiple PIM (Processor-in-Memory) chips, each with 128kb of memory and 64 processors. 0.25M of these processors with memory have passed initial testing in a single Cray-3 quadrant as part of the Cray-3/SSS (Super Scalable System), a joint venture between NSA and Cray Computer Corporation.



Components of a prototype superconductive crossbar switch being developed at NSA. Data are transferred (via ribbon cable) from room temperature to cryogenic temperatures and back to room temperature at 2.5 Gb/s. A full 128-by-128 configuration is intended for use as a switch for massively parallel computer memory data transfers.

Mass Storage

High performance computing systems handle substantially more data — both input and output — than traditional systems. Large scale simulations, experiments, and observational projects generate large multidimensional datasets on meshes of space and time. Accurate modeling requires that the mesh be as dense as possible. Fast simulations require that relevant subsets of information in large datasets be accessed quickly — in seconds or minutes rather than hours. For example, scientific investigations of environmental and earth science phenomena require ever increasing volumes of data in order to develop accurate models that can explain and predict these phenomena in a timely fashion.

The mass storage industry has developed technologies for handling petabytes (10^{15} bytes) of data today and is developing technologies for handling exabytes (10^{18} bytes) in the future. Technology advancements include increasing the density of storage media such as disks and tapes, RAID (Redundant Arrays of Inexpensive Disks), and robotic tape storage systems. Software that has been developed to manage mass storage systems includes file and volume managers and systems for updating, archiving, and backup.

On October 21-23, 1994, twelve mass storage vendors and two other organizations presented non-disclosure briefings to the HPCCIT Subcommittee on issues including trends, obstacles, standards, forums, international issues, the HPCC Program, and current and planned product lines.

The High Performance Storage System (HPSS) Consortium was one of the two non-vendor organizations at the October briefings. Consisting of DOE's Oak Ridge (ORNL), Lawrence Livermore (LLNL), Los Alamos, and Sandia National Laboratories, and IBM Federal, HPSS is developing a high performance parallel network-centered data storage and access system capable of GB/s transfers. Experimental implementations are in place at the development laboratories and at Cornell University, with general availability in late 1995.

Several HPCC-funded projects are addressing specific mass storage issues, including the following:

- The National Storage Laboratory (NSL) and HPSS at LLNL
- The DOE-funded OPTIMASS project has documented up to a 50-fold speedup in accessing large spatio-temporal datasets from mass storage systems such as robotic tape systems. The improved timings result from reorganizing the data to match common access patterns rather than the order in which the data were obtained or generated. For example, in preparing for a simulation, a region of interest is often identified using summary data, and raw data from the region are needed for more detailed analysis. In this case, data are best organized into "clusters" associated with summarized regions. Optimization algorithms have been developed for reorganizing the original dataset to match intended use, a workbench helps scientists select the most desirable reorganization, and storage server protocol enhancements have been designed to permit both control over physical placement of data on storage devices and efficient cluster management. This system has been implemented at NSL and demonstrated for typical queries in climate modeling visualization.

  http://gizmo.lbl.gov/optimass.html

3. Advanced Software Technologies

Today's high performance computing environment differs in three fundamental ways from the environment of 1991 when the HPCC Program began:

- The environment includes parallel systems that are significantly more complicated and powerful than scalar and vector systems.
- One application program can be executed on several computing systems. These systems can be scalar or vector or parallel or a combination thereof; they can range in size from workstations to the highest performance systems; they can include special purpose systems such as workstations for scientific visualization or expensive remote instruments such as telescopes or microscopes; they can come from different vendors; and they can be located at sites hundreds or thousands of miles apart that are connected by high speed networks. The most demanding application programs are beginning to be executed in exactly this mode (such applications are described beginning on page 41).
- Users access this environment from scientific workstations that often are critical parts of the environment itself — the workstations display progress of the execution and results from the computations, and users in turn interact with the execution in real time.

This contrasts with the 1991 environment in which users commonly ran their applications on single scalar or vector computers accessed via comparatively low speed local area networks and interacted little with the execution, or they ran their applications on their own workstations.

Several software developments have made the new environment possible. These include the following:

Systems Software

New scalable microkernel operating systems are the product of R&D during earlier years of the HPCC Program and have become more robust, more stable, and more complex. These operating systems manage dozens to thousands of parallel processors and associated memory (as well as vector processors in some cases); control I/O to peripherals such as disk drives, tape drives, printers, and advanced mass storage systems that incorporate robotics; and facilitate communications with more complicated and higher speed networks.

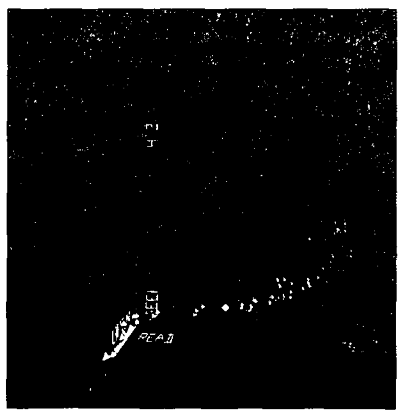
The microkernel basis for HPCC prototype operating systems has now been accepted by the broader computing community. In the past year, DEC, Hewlett-Packard, IBM, and Microsoft committed to use microkernel operating systems as the basis of their own offerings. The overhead involved in transitioning into and out of the microkernel has received considerable attention in the past two years. Technology developments include software fault isolation, co-location, and extensible microkernels.

Scalable I/O

Begun in FY 1994, the Scalable I/O Grand Challenge addresses the need to complement fast computing systems with fast input/output to memory. Teraflops systems that access terabytes of data will require Gb/s I/O. By instrumenting I/O intensive Grand Challenge applications, insights into I/O behavior are providing opportunities for I/O performance optimization. By adding features to operating systems, languages, compilers, memory management, interfaces to mass storage systems, and runtime support, I/O performance increases of 10-fold to 100-fold are possible. The initiative is funded by ARPA, DOE, NASA, and NSF, and involves vendors and researchers from more than 30 institutions. Testbeds include a 512-node Intel Paragon at Caltech's Concurrent Supercomputing Consortium, and a 128-node IBM Scalable POWERparallel (SP) at DOE's Argonne National Laboratory (ANL). Products will include an integrated set of tools and software systems providing scalable I/O capabilities available to the U.S. high performance computing community, and a scalable I/O benchmark suite. One example is the Pablo software used for performance instrumentation, analysis, and visualization. Pablo development was funded by ARPA and NSF; it has been commercialized by Intel.

http://www.ccsf.caltech.edu/bluebook_96/nsfgcpiom.html

http://www-pablo.cs.uiuc.edu/HPCC.html



Snapshot of the dynamic patterns of read behavior in a parallel version of software to calculate electron-molecule cross-sections using a 128-processor Intel Paragon at Caltech's Concurrent Supercomputing Consortium. The axes are file open duration, file seek duration, and file read duration. The locations of the octahedra are the current values of each processor's performance metric. History ribbons show the last N positions for three select octahedra (red is most recent). The Pablo software was used to produce this image.

Programming Languages and Compilers

High Performance Fortran (HPF)

The HPF language definition was completed in 1994. It allows straightforward expression of data parallel constructs. HPF has become a de facto standard language for parallel computing systems, providing Fortran programmers with a familiar portable language. Both computing systems vendors and independent software vendors have written compilers that translate HPF commands into machine instructions that distribute the computation across the processors, memory, and networks. The HPF Forum is a coalition of government, high performance computing systems vendors, and academic groups that is coordinated by the Center for Research on Parallel Computation (CRPC). The Forum is studying extensions that support task parallelism and scalable I/O. ARPA and NSF directly support this effort.

http://www.erc.msstate.edu/hpff/home.html

High Performance C++ (HPC++)

ARPA supports a multi-institution collaborative effort that is defining a minimal set of extensions to C++ that support both task and data parallel constructs, for use by C and C++ programmers. ARPA and NSF are also developing common runtime support environments for both HPF and HPC++ in order to increase efficiency and reduce barriers in compiler development.

High-Efficiency Languages and Compilers

NSA has long supported research in high performance languages and compilers for specific high performance computing systems with the goal of achieving near-peak performance in many applications and helping to achieve advertised price/performance characteristics of existing systems. These joint NSA/industry efforts include the AC extended C compiler for the Thinking Machines CM-5, which shows dramatic speedups over the vendor's data parallel compiler; the DBC bit serial data parallel C compiler for the joint Cray Computer/NSA Cray-3/PIM system; and AC for the Cray Research T3D that has mechanisms allowing efficient emulation of shared-memory programming on distributed memory multiprocessors. Consistency and clarity of the mapping between language and architecture underlie these successes. Continued research has the goal of extending these models to other architectures and using them as a basis for higher level programming models.

Software Tools

Parallel Virtual Machine (PVM)

PVM software permits a heterogeneous collection of networked computing systems, all of which use the Unix operating system, to be used as a single large computer. Developed early in the HPCC Program with multi-agency support, PVM is used at hundreds of sites worldwide both for problem solving and as a tool to teach parallel programming.

http://www.netlib.org/pvm3/index.html

Message Passing Interface (MPI)

MPI is a specification for a standard portable library of subprograms for message passing that can be called from programs written in Fortran or C. Begun at Supercomputing '92 in November 1992, the first phase of MPI was completed in May 1994 with ARPA, DOE, and NSF support. Message passing gives control of parallelism to the application developer rather than to the hardware or the compiler. MPI can be implemented on any of today's parallel computing systems or networked heterogeneous workstations. DOE funded the public-domain portable version that was implemented by Argonne National Laboratory and Mississippi State University. Future DOE products will include an instrumented version that produces communications statistics and conversion of linear algebra and partial differential equations libraries to use MPI.

http://www.mcs.anl.gov/Projects/mpi/index.html

Parallel Tools Consortium (Ptools)

Ptools is an open community of vendors (both systems and software), Federal organizations, and university researchers developing public-domain portable reference implementations of tools for parallel environments. The Consortium seeks to identify and build tools that users want and will use, and to involve users in tool development and refinement. Ptools was established in 1993 and is headquartered at Oregon State University. Current projects include distributed array visualization, a lightweight corefile browser, a message queue manager, parallel Unix commands, and portable timing routines.

http://www.llnl.gov/ptools/ptools.html

Computational Techniques

Algorithms for numerical computations and for finding and moving data are widely used in Grand Challenge applications. They are developed by experts and included in general purpose libraries of reusable software. The HPCC Program has funded such software development since its inception, and much of that software is becoming available through the National HPCC Software Exchange (NHSE). A catalog of available software can be found at:

http://www.netlib.org/nse/

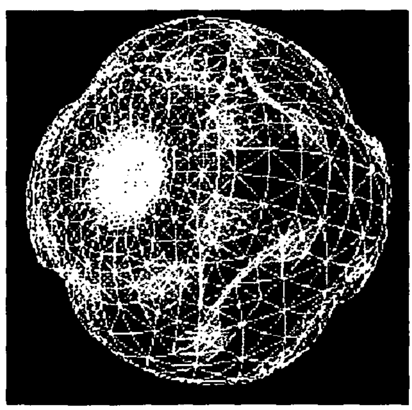
Multipole Method for Solving Differential Equations

Performance improvements have been the result of algorithm development as much as hardware speedup. An example of this trend is the development of the multipole method, which also illustrates the contributions of basic research to HPCC. The algorithm provides the solution of a differential equation approximated at N points in order N operations. Thus doubling the number of points in order to obtain a more accurate solution requires only twice as much work; other algorithms might require many times the work to obtain the same solution. The original research for the method was supported by NSF at Yale University. The algorithm is now a cornerstone for major simulations including Grand Challenges at Caltech and the University of Illinois.

Unstructured Mesh Computation with PUMAA3D

Both computations and parallel data management are issues in unstructured mesh computations, in which a complex surface or solid object is partitioned into a collection of covering triangles or trapezoids. An application package then solves a system of equations at the nodes of this mesh. A practical application is finite element analysis of the effects of pressure and temperature on disk brakes. The PUMAA3D software developed at DOE's Argonne National Laboratory implements efficient parallel algorithms for unstructured mesh generation, adaptive mesh refinement (for example, where the pressure or temperature are greatest), mesh partitioning, and the solution of the sparse linear systems that commonly occur in these applications. The software is in the public domain, flexible, and portable to a wide variety of distributed memory systems. PUMAA3D development has been done on the Intel Delta and the IBM SP system.

http://www.mcs.anl.gov/Projects/meshacc94.html



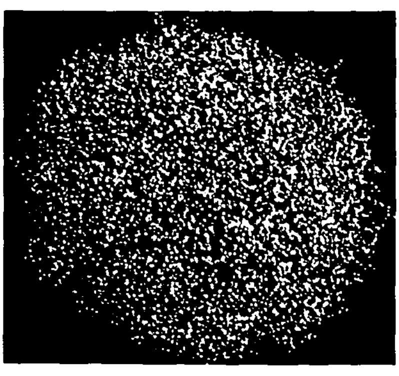
New techniques adaptively refine, de-refine, and partition meshes to accurately model rapidly changing solutions such as those that arise in simulating layered high temperature superconductors.

ScaLAPACK

Funded by ARPA and NSF, ScaLAPACK is a scalable software library for common linear algebra computations. These include linear and eigen solvers for both dense and sparse matrices. Many sparse problems have internal data representations that make it infeasible to use standard libraries. A book of templates for building custom sparse linear solvers has been developed. The use of reverse communication that allows users to customize solvers to their data structures is being investigated.

Handling Irregular Data using CHAOS

A number of scientific computing applications have unstructured, sparse, adaptive, or block-structured data. Two such applications are computational fluid dynamics and molecular dynamics. With ARPA funding the portable CHAOS runtime support library has been developed to help parallelize these applications. Library features include (1) coordinated interprocessor data movement, (2) off-processor data management, (3) support for a shared name space, and (4) coupled runtime data and workload partitioners to compilers. The library can be called from the Fortran D and HPF programming languages and, in the future, from C++. CHAOS has been used to parallelize the CHARMM molecular dynamics application to run on several systems. Current efforts address handling hierarchical data structures and optimizing disk access. This work is being conducted at the University of Maryland, where it is being used for the land cover dynamics Grand Challenge, and at LANL, Caltech, Rice University, and Syracuse University. It is being used by the ARPA-funded Parallel Compiler Runtime Consortium and the Scalable I/O initiative.



Input molecule for the CHARMM molecular dynamics software that has been parallelized on multiple systems using the CHAOS runtime library. Key portions of CHARMM have been automatically parallelized using an enhanced version of the Fortran D compiler.

http://www.cs.umd.edu/projects/hpsl/projects/blue_book.html

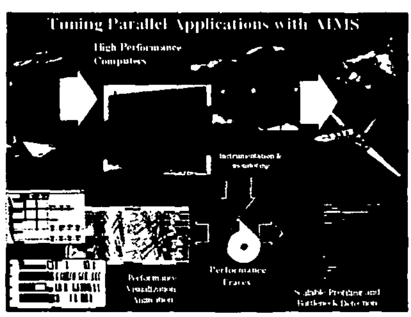
Performance Measurement

Automated Instrumentation and Monitoring System (AIMS)

NASA has developed AIMS tools to help detect performance bottlenecks and suggest ways to eliminate them. The following features have been included or are under development:

- Scalable visual representations of execution traces — A number of methods for limiting the amount of information gathered and displayed will be developed, and results will be used in developing fast and flexible trace-browsing capabilities.

- Users need not change their applications in order to use these tools, and easy-to-use interfaces involve minimal user interaction.



- The issue of data distribution and alignment will be addressed using compiler support, and the feasibility and usefulness of data-oriented views will be demonstrated for NASA computational fluid dynamics applications.

NASA has also developed a version of AIMS that supports performance debugging in distributed heterogeneous computing environments and that exploits the portable message passing interface and multitasking provided by PVM.

AIMS is used at NASA and DOE facilities and in universities for both teaching and evaluation. Convex Computer Corporation produced CXTRACE that is based on AIMS and runs on their SPP-1 and HP cluster. AIMS will be implemented on the IBM SP-2 and is under evaluation by other vendors.

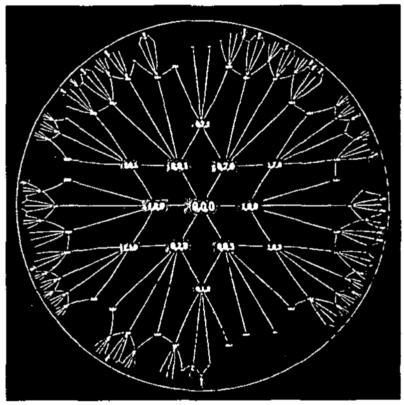
http://cesdis.gsfc.nasa.gov/hpccm/accomp/94accomp/cas94.accomps/cas3.html

Performance of a Toroidal Architecture

The NSA-funded Supercomputing Research Center, part of the Institute for Defense Analyses, has developed a performance monitoring system that displays the dynamic performance of system components such as memory references, memory bank stalls, network traffic, and processor stalls (shown at the right). Features include discrete group actions in hyperbolic geometry, scrolling the workstation screen with hyperbolic transformations, and displaying function graphs over network nodes and links.

Performance Instrumentation

NIST, with ARPA support, has developed the MultiKron chip that unobtrusively monitors memory busy traffic and can gather statistics or traces at user request. This chip has been installed on Intel Paragons. The NIST S-Check project, also supported by ARPA, and the Pardyn project at the University of Wisconsin, which receives ARPA and DOE support, automatically instruments a program and gathers runtime statistics useful in understanding performance.



This hyperbolic scrollable display that maps the three-dimensional toroidal Cray Research T3D network onto the two-dimensional workstation screen without false crossings was developed at the Supercomputing Research Center.

HINT (Hierarchical INTegration)

HINT evaluates computer performance by measuring the quality of an answer as a function of time, instead of counting operations as done in the usual benchmarking. The unit of measurement is QUIPS (QUality Improvement Per Second). HINT reveals the characteristics of a computing system, demonstrating, for example, the impact of cache, memory size, and operating system overhead. The software is easy to transport to any type of modern computing system, making the benchmarking of parallel systems as easy as conventional systems. Because HINT was designed to predict application performance, it can be used as a computer design tool. HINT was developed by DOE's Ames Laboratory.

http://www.scl.ameslab.gov/scl/Projects/hint1.html

Benchmarking

NASA Parallel Benchmarks

These benchmarks reflect the diverse computational demands of NASA's mission-oriented science and engineering programs. There are two sets of benchmarks: the Parallel Benchmarks for Numerical Aerodynamic Simulation (NAS) and the Parallel Benchmarks for Earth and Space Sciences (ESS). The NAS Parallel Benchmarks were developed in 1991 to evaluate the performance of parallel computing systems for workloads that typify those used by the aerospace engineering community. The ESS Parallel Benchmarks are a new set of test programs typifying those used by the Earth and space sciences community. Both are designed to enable portability across disparate classes of parallel architectures and system configurations while retaining the basic nature of the computations to be performed. They increasingly serve as procurement criteria in industry and academia, and aid internal development of current and future systems by parallel computing systems vendors.

http://cesdis.gsfc.nasa.gov/hpccm/accomp/94accomp/bench.html

Joint NSF-NASA Initiative on Evaluation (JNNIE)

The JNNIE study, a collaboration between NSF and NASA high performance computing centers, is evaluating the effectiveness of various parallel architectures for large-scale scientific computing, determining the level of effort required to transport realistic-sized codes to parallel environments, determining the extent to which portability is achievable across diverse architectures (and at what price in performance), and disseminating this information to academia and industry. To this end, participants have been conducting scalability studies and cross-architecture comparisons of a number of academic and third-party application codes in the multi-vendor scientific computing environment.

Software Sharing

National HPCC Software Exchange (NHSE)

In September 1994 the HPCC Program funded NHSE to collect software (or software descriptions) for high performance computing systems and make it available on the Web. The top level organizational structure for the software is: application level, library software, parallel programming environments, parallel programming languages, parallel programming tools, and performance visualizations. NHSE also contains a hardware and software vendor catalog and information about reports, journals, and professional associations.

http://www.cs.rice.edu/CRPC/bluebook/bluebook.html

Visualization

Critical to efficiently understanding results from high performance computations is rapid three-dimensional high-resolution color display of results from simulations, often overlaid with other data such as experimental observations, with which the user can interact. These displays can be single images or video. Several projects are developing improvements in this field:

Cave Automatic Virtual Environment (CAVE)

The CAVE was developed at the Electronic Visualization Laboratory at the University of Illinois at Chicago. Two additional CAVEs have been built at the National Center for Supercomputing Applications (NCSA) and Argonne National Laboratory. It is a multi-person, room-sized, surround-screen, surround-sound, projection-based virtual reality environment. Graphics are rear projected in three dimensions onto the walls and floor and viewed with stereo glasses. Several people can explore an environment together in the room. As the principal viewer wearing a location sensor moves within the display boundaries or controls the image using a "wand" (the CAVE equivalent of a "mouse"), the perspective and stereo projections of the environment are updated. Each institution pursues a wide range of interactive applications including molecular modeling (drug design), product design, medical imaging, manufacturing, cosmology, and education. In addition, research continues to develop and extend CAVE capabilities such as coupling these virtual environments to supercomputing resources for large-scale background computations. A key goal is demonstrating distributed collaboration in which the three CAVEs are linked by high-bandwidth network connections for visualizing the same application in the three locations.

http://www.ncsa.uiuc.edu/EVL/docs/html/CAVE.html

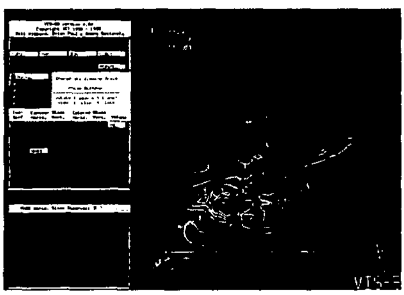
http://www.mcs.anl.gov/anlcave.html

http://www.ncsa.uiuc.edu/VR/VR/VRHomePage.html

Vis5D

The Space Science and Engineering Center (SSEC) at the University of Wisconsin at Madison in cooperation with EPA has developed an interactive three-dimensional visualization tool (shown below) for use at desk-top workstations that is based on SSEC's Vis5D for visualizing output from atmospheric and ocean models. Vis5D has also received non-HPCC NASA funding.

http://ssec.wisc.edu/~billh/epa.html



Application of Vis5D to EPA's Regional Acid Deposition Model shows transparent volume rendering of sulfur dioxide (the red fog) and a horizontal slice with iso-lines of nitric acid over a topographic map of the Eastern U.S. The icons on the left give the user interactive control over the three-dimensional images as they are animated. Vis5D makes this interactive exploration possible by compressing data sets to fit in workstation memories. Vis5D has been used for experiments over the Blanca Gigabit Testbed and has been adapted to run in the virtual reality CAVE (described above); it is freely available over the Internet.

VolVis

VolVis is a multi-featured visualization package developed at the State University of New York at Stony Brook through DOE's Partnership in Computational Sciences (PICS) consortium centered at ORNL. It accommodates both regular and irregular grids and includes (1) manipulation and rendering tools, (2) flexible I/O, (3) volumetric navigation, (4) a key-frame animation generator, (5) quantitative analysis tools, and (6) a protocol for communicating with three-dimensional input devices. It is available for systems with XWindows and Motif. Parallel versions of key algorithms including output visualization algorithms for several platforms have been developed.

Other Research Activities

ARPA is developing technology and manufacturing capability for high definition displays and scalable image processing systems. Components include projection, head-mounted and direct-view displays based on multiple technologies, display architectures and processors, compression algorithms, and high speed data transmission. In FY 1996 ARPA plans to establish a testbed for interoperability standards for display interfaces and to demonstrate a prototype high-resolution progressive scan digital camera.

Synthetic cutaway view of Crater Lake, OR. The upper figure is the original digital elevation model that contains 307,000 triangles and requires approximately 10 seconds to display. A new algorithm was used to produce the lower figure, a simplified model with 800 triangles that can be displayed in real time.

ARPA has also funded the development of multi-resolution modeling algorithms to simplify surface models so that they can be displayed faster. The algorithms display nearby objects using the detailed input data but display distant objects more coarsely. Related projects include (1) automating shape acquisition, (2) modeling of enclosing environments such as rooms, factories, and ports for architecture, simulation, and training applications, and (3) radiosity algorithms that simulate indirect lighting.

A virtual reality Grand Challenge is described on page 52.

4. Technologies for the Information Infrastructure

The HPCC Program is developing technologies underlying the National Information Infrastructure (NII) to facilitate addressing National Challenge problems of significant social and economic impact (HPCC National Challenges are described beginning on page 75). These technologies build upon and extend those needed to accomplish the goals of the original HPCC Program. They will also contribute to the Government Services Information Infrastructure (GSII), the Defense Information Infrastructure (DII), and the Global Information Infrastructure (GII). Other non-HPCC technologies to support the NII are under development by Federal departments and agencies. Such efforts are being followed and in many cases coordinated with HPCC Program activities.

Information Infrastructure Services Technologies

NII technologies must be integrated and interoperable — information must be transmitted over different bitways among different computing systems in a seamless fashion; services such as authorization and accounting must be integrated and interoperable; and they must serve people with different communications preferences (for example, voice rather than visual).

Integration and interoperability are most easily accomplished by using tools that implicitly provide them. Toward this end, ARPA has initiated a program to develop the software and services technologies needed to facilitate a marketplace of advanced distributed applications that will interoperate over the underlying networking infrastructure. The following are examples of some current activities in this program.

HORUS

The HORUS project is developing an environment for reliable distributed computing. Its "virtual synchrony" environment integrates distributed computing with fault tolerance (detection of and reaction to failures) and will make it easier to develop distributed software. An example is military mission planning and control software that integrates data from space, air, and ground resources.

http://www.cs.cornell.edu/Info/Projects/HORUS/ARPA/arpa.html

SILDS

Billing, payment, accounting, and associated privacy mechanisms are critical to the NII, especially the National Challenge in Electronic Commerce. ARPA is developing the underlying technologies to provide these capabilities in an interoperable manner, building upon underlying security services. An example is the SILDS project that is developing an infrastructure to support authorization, authentication, accounting, and related security services for the Internet that is layered over the Kerberos authentication system.

http://nii-server.isi.edu/gost-group

Adaptive Communication

Portable information and communications devices will be extensively used in the NII. ARPA is developing the required underlying mobile and wireless technologies (wireless networking technologies are described on page 14). Toolkits are needed for developing applications that adapt to changes in the computing and communications environment (such as platform mobility, sporadic connectivity, and dynamically varying bandwidths). An example is "middleware" services that can automatically or semiautomatically adapt the amount and form of communications between a mobile computer and the infrastructure to vastly different operating circumstances.

http://www.mcl.cs.columbia.edu/

TRAVLER

The TRAVLER project funded by ARPA is packaging systems software and services for mobile computers. These include support for adaptive agents, data consistency, database support, predictive caching of files, real-time modeling and simulation, resource discovery, security, user-level replication, and virtual networking.

http://www.cs.ucla.edu/ficus/

Wide-Area File System

ARPA funds Transarc Corporation in developing a system for organizing and locating information in a Wide-Area File System. Each file has a "synopsis" that summarizes its contents, while a "digest" contains multiple synopses. The system lets users control who can access their data. It builds on the AFS (Andrew File System) software developed at Carnegie Mellon University and Transarc and is used by more than 200,000 people at more than 120 government, industry, and academic Internet sites.

ftp://grand.central.org/afs/transarc.com/public/www/Department/ARPA/summary.html

FICUS

The FICUS project is developing a peer-to-peer file replication system that keeps multiple full-function copies of files so that (1) workers can improve their efficiency by keeping multiple copies of key files on systems both at work and at home; (2) individuals collaborating at a distance can each have copies of their working files, and updates are propagated; and (3) copies of files can be stored on portable systems. Replicating and updating files is done over the Internet. Ficus is built using its "stackable layers" technology in which software modules providing new services (such as these new file system functions) are added to existing operating systems; this technology can also be used to add other file system functions such as encryption and data compression.

http://www.cs.ucla.edu/ficus/

Nile Project

The NSF-funded Nile Project is building scalable systems for reliable access to distributed data using distributed computing resources. The immediate application is processing enormous amounts of data accumulated by CLEO, a large High Energy Physics experiment. While the goal of Nile is to solve the computing and network problems specific to this particular scientific endeavor, accessing 10 to 1,000 TB (terabytes or trillions of bytes) of distributed data and processing those data on 100 to 1,000 distributed computing systems is becoming increasingly important to such diverse areas as environmental science, medical imaging, seismology, and video-on-demand. Two particular features of the Nile project are the importance of low-latency access to the data and working with an enormous amount of data on tape. Nile collaborating institutions include Cornell University, the University of Florida, the University of Texas at Austin, and the University of California at San Diego. CLEO itself is a collaboration of 23 universities in the U.S. and Canada, and several members of CLEO are also involved with Nile.

http://www.nile.utexas.edu/NC/Report/Report-950118.html

Video Teleconferencing

Video teleconferencing explicitly requires integration and interoperability as well as resource management. DOE is addressing the need for standard file systems, graphics, spreadsheets, serial interfaces for access and configuration, and sliding bar control of video interaction. Another DOE effort addresses merging telephone, FAX, e-mail, and video into one resource.

World Wide Web (WWW) and NCSA Mosaic

WWW and Mosaic are described on pages 6 and 7. Assuring the continued coordination of a global infrastructure for information is critical to the success of the NII. The HPCC Program will continue R&D in WWW and browsers such as Mosaic. In addition HPCC agencies support and ARPA funds the World Wide Web Consortium, headquartered at MIT.

Issues being addressed at NCSA and elsewhere include collaborative document preparation (with appropriate security), coupling to mass storage and database management systems, forms for user queries and submission of information/data to the server, interactivity, needs of the disabled in the Web environment, next generation WWW architectures, replacing the transient Web link addresses with permanent identifiers (today if a file is moved, its address changes, and every hypertext link to it would have to be changed in order to refer to the same file), scaling issues, searching (and the information that is used in the search), security and privacy, tools for performance measurement and load management, and usage metering/accounting/payment. Furthermore, the development and extension of Web standards continues to be a priority.

Security and Privacy

The NII will place increased demands on security and privacy systems. Two examples are medical records and financial records — both users and providers need to know that there is no unauthorized access to or updating of the information. Already because of the growing dependence on information systems and their interconnection, these systems are more exposed and vulnerable to attack.

In response to this vulnerability, incident response teams have been established to monitor and react to unauthorized activities and potential network intrusions. Each team serves a specific constituency such as an organization or a network. Two of the first such teams, CERT and CIAC, were founded in 1989 as a result of an increase in wide-spread, malicious network-based attacks such as the Morris Worm, and provide round the clock incident handling:

- CERT is the Computer Emergency Response Team based at the Software Engineering Institute in Pittsburgh, PA. CERT was established by ARPA to support Internet users without response teams and continues to serve as the response team for much of the Internet. In 1994 CERT was contacted more than 33,000 times to report computer security incidents or to request security-related information that affected more than 40,000 different Internet sites. Community outreach included issuing 15 CERT advisories. As the incident rate has grown, so has the sophistication of the intruders. To counter these threats, CERT is expanding its program to increase emphasis on incident prevention through technology improvement, security awareness and training, and security assessment activities.

  http://www.sei.cmu.edu/SEI/programs/cert.html

- CIAC is the computer security incident response team for DOE and its contractors. CIAC provides training classes, white hat reviews, and awareness documents in the form of bulletins, advisories, electronic notes, tools, and references. CIAC's hotline number is 510-422-8193; its e-mail address is ciac@llnl.gov; and its ftp server is ciac@llnl.gov.

  http://ciac.llnl.gov/

CERT and CIAC are founding members of FIRST, the Forum for Incident Response and Security Teams. Other Federal agencies participating in FIRST include DOE, NASA, NIST, VA, the Navy, and the Small Business Administration. Members work together in handling complex incidents, testing vulnerability fixes, and sharing incident trend information. Members participate in the annual incident handling workshop sponsored by FIRST; the next workshop will be held September 18-22, 1995, in Karlsruhe, Germany; details about FIRST and the workshop can be found at:

http://www.first.org/first/

HPCC agencies are also responding by increasing emphasis on security research and prototype deployment in FY 1996 and beyond. ARPA, DOE, NASA, NIST, and NSA are addressing issues including preventing unauthorized entry into computing systems, protecting the network infrastructure (the Internet, higher speed networks as they are deployed, and other "bitways" such as satellite, cable TV, cellular, and broadcast) from external and internal attack, protecting information in repositories and in transit, providing data security controls within applications, privacy protection for medical and other sensitive applications, secure electronic commerce, secure internetworking for distributed simulations, secure collaborative work (for example, video teleconferences), security in emergencies or crises or hazards (for example robots in hazardous areas), and emergency response. The long-term challenge is to provide security solutions that can scale to emerging high performance computing technologies, such as multimedia, ultra-high data rates, mobile computing, and very large-scale distributed information storage and retrieval.

For example, one-time password technology and the Kerberos authentication system are helping to alleviate this problem. Kerberos is a cryptographic-based network system, originally developed at MIT, that provides a mechanism for client/server computing authentication. Several HPCC-funded sites are implementing such technologies:

- The NSF Supercomputer Centers are collaborating to become a "metacenter" by sharing their resources. As the metacenter evolves, a researcher will be able to direct that in the middle of the night a large program is to be run using computers and archival storage systems at several of the other centers. In order to do this, all involved systems and networks must be assured that the work is authorized. Extending such models for use throughout the Internet can minimize future attempts to exploit security vulnerabilities.

- The ESnet-based DOE laboratories added cross-realm functionality to Kerberos in order to complete their 1994 multi-site and multi-organization demonstration of authentication to support collaborative research.

http://nii-server.isi.edu/gost-group/

http://www.es.net/pub/esnet-doc/auth-and-security/

The secure-http protocol is an enhanced version of the WWW http protocol that supports privacy, authentication, and digital signatures. Secure http enhancements have been added to NCSA Mosaic to create Secure-Mosaic. It has been incorporated into CommerceNet, a TRP-funded government/industry consortium, which is the first full-scale market trial of electronic commerce on the Internet (the Electronic Commerce National Challenge is described on page 83).

http://www.commerce.net

Through a new joint Defense effort, ARPA, NSA, and the Defense Information Systems Agency will coordinate their research programs in digital signatures, e-mail security, secure operating systems, secure distributed applications over a single administrative domain, secure routing protocols, security checking, and survivability and recoverability.

Security tools and services developed by government, industry, and academia will be integrated, scalable, and suitable for use by commercial carriers, by third-party providers of security services, and by applications for embedded security functions. Resultant technologies will be flexible in meeting specific needs and in providing varying degrees of protection.

These security and privacy systems must be both policy flexible and cost effective. For example, medical and financial records require more security and privacy than information in libraries, although library users needs privacy guarantees. The Synergy system being developed at NSA, as well as other security technologies developed by

NSA and ARPA, is being integrated into basic research in computing systems that will result in the next generation of commercial products. This strategy will lead to products that are easily specialized for a wide variety of security needs.

NASA supports a coordinated effort to improve the protection of sensitive but unclassified data used in collaborative aeronautics engineering both between the Federal government and industry and among industrial partners. The aeronautics community will use encryption tools and the security architecture developed for electronic commerce as a basis for communicating data that require protection from premature or unauthorized release.

Information Infrastructure Applications Technologies

An easy-to-use NII requires complex interfacing of humans to intelligent information management, control, and sensor systems. Performance and usability also depend on the ability of those systems to interface among each other. In FY 1995 and 1996, HPCC agencies are addressing many issues in this area, including the following:

- Research on software environments focuses on application developers having a design environment that matches the application. Tools developed in this effort are being used to construct human-computer interfaces used in applications ranging from software testing to visualizing thermal signatures in the B-2 tailpipe (Northrop). Test and evaluation tools used by Hughes Aircraft improve testing productivity by approximately 25 percent. The Analytical Science Corporation uses the same tools to develop an Avionics Verification and Validation system for Wright Labs. Science Applications International Corporation uses tools developed by this research community to provide interoperability between C and Ada software for U.S. Army simulators. Under related software research programs, tools are being developed to reverse engineer software to extract design information and provide tailored explanations of that software. The Army MICOM Software Engineering Directorate is evaluating some of these tools for supporting the Patriot upgrade program.

- ARPA and NSF provide support to addressing problems in long-term management and evolution of complex software — system lifetimes are long, applications cluster in families with common characteristics, and change is rapid. The effort builds on the architecture-based approach of defining product families by adding capabilities for recording information about the rationale underlying a design so that modifications can be made more rapidly without introducing errors. Common components will be shared, reducing the cost of providing and maintaining productivity enhancers such as domain-specific architectures for families of systems and very high level languages (fourth generation languages). In FY 1996 consensus Architecture Description Language and Interactive Architecture Synthesis Tool for complex systems will be defined, cluster applications identified, and applied demonstrations initiated.

- ARPA is developing technology to enable the integration of DOD's existing legacy data systems into more effective functionally integrated information services — without developing large-scale monolithic information systems. This effort has demonstrated that by using an information-integrating mediator, F-22 engineers can execute design changes in 80 percent less time, which is essential for streamlining the weapon systems acquisition process. A mediation module developed from this effort enables air campaign planners to locate the highest priority targets for a particular military objective in 10 minutes instead of 48 hours, providing military advantage in time of crisis. It is enabling military doctors to locate patient information within minutes instead of days, providing a life saving difference in emergencies. These advances will be incorporated in other applications in FY 1996.

- ARPA is developing enabling technologies and design tools that potentially will aid affordable acquisition and upgrade of next generation DOD platforms. In FY 1996 cognitive support tools will improve the ability of engineers to generate, store, retrieve, track, explore, and analyze design alternatives. Today humans generate alternatives in an ad hoc manner and explore them by detailed physical simulations that are time consuming to create and execute. Consequently, most complex design efforts explore a small number of solutions rather than an expanded design space. Rockwell is developing "Design Sheet," a tool used in the Air Force's Airborne Laser program to integrate weight, performance, and cost data and models from teams representing several disciplines and companies; the time to build an integrated model was reduced from two years to six months.

- ARPA is developing tools to support military planners in creating, executing, monitoring, and adapting courses of action and plans for logistical and transportation support. Interactive tools have been a key focus area; several have been developed and are in varying phases of transition. Planning is a labor-intensive, time-consuming, detailed activity that underlies nearly all aspects of military operations. ARPA's efforts have advanced the state of the art in planning and have produced several systems that support specific planning functions and are being transitioned. Applications of this technology to the operating forces include (1) DART, a transportation analysis and planning tool that supported logistics planning for Desert Storm by reducing the time required to develop transportation plans from a few weeks to a few hours (DART has been fielded to every commander-in-chief, where it is in daily use); (2) TARGET, a distributed collaborative planning tool that supports the split-base concept and lets multiple geographically distributed planners work together in a single shared planning workspace (TARGET is being transitioned into the Global Command and Control System Leading Edge Environment); and (3) Air Campaign Planning Tool (ACPT), a system that guides a user through the analysis of national objectives into military objectives, military objectives into tasks, and tasks into specific actions, finally resulting in a Master Attack Plan with traceability of all activities back to national objectives (ACPT is being incorporated into the joint air planning system known as CTAPS).

- ARPA and NSF support development of core technology in human language processing and the transfer of this technology into productive use. Language-based applications help improve overall readiness of our forces by allowing the user to interact directly and naturally with complex systems. Intuitive interaction using natural human language will make for rapid acquisition of expert skills. For example, the TIPSTER system uses text understanding technology to support detection and extraction of key information for intelligence analysts from NSA and the CIA. In FY 1996 tools for speech understanding will be applied to command and control problems including language interaction with tactical warfare training simulators, language-based access for joint service logistics planning, and language-based access to in-transit visibility of logistical assets.

- ARPA is developing new technology to enable humans and computers to interact effectively, efficiently and naturally, including (1) adaptive presentation and input that is sensitive to the context of the user, task, and application domain, (2) multimodal interaction incorporating speech and gestures, and (3) problem and information visualization, focusing on clearly presenting symbolic and numeric data. Emerging technologies and methods are being inserted into the Joint Task Force Advanced Technology Demonstration (ATD), Planning and Decision Aids, and Healthcare Information Systems programs.

In FY 1995 and FY 1996, NSF will enhance its support for research in intelligent systems, especially human-machine information systems. The research will advance the underlying scientific knowledge and technologies needed to create an intelligent service layer that will enable a significantly broader base of information providers, developers, and consumers, while reducing existing barriers to accessing and using information and computing resources for real-world applications. Work in this area includes:

- Human language technology — Development of new technologies for speech recognition, text understanding, and multilingual language processing, including machine-aided language translation

- Multi-modal human-computer interfaces — Image processing and analysis, computer vision, and integration of images with sound, text, and other forms of information source and presentation

- Virtual environments/collaboration technology — Sharable computing and communication environments that many can access, interact with, and use effectively across time/geographical and physical/artificial boundaries

- Very large knowledge repositories — Models, tools, and technologies for storing, representing, accessing, and using large amounts and different kinds of objects, data, and knowledge sources

5. High Performance Computing Research Facilities

It is at these facilities that (1) full-scale production systems are used on Grand Challenge scale applications that cannot be scaled down without modifying the problem being addressed or requiring unacceptably long execution times, (2) early prototype systems are evaluated, feedback is provided to developers, and systems are made more mature and robust, (3) small versions of mature systems are used for developing parallel applications, and (4) advanced visualization systems are integrated into the high performance computing environment. The largest of the Grand Challenge scale applications are being run on multiple high performance systems located at these HPCC facilities across the country and networked at gigabit speeds. These facilities are demonstrating a new paradigm for conducting advanced R&D by the Federal government and American industry and academia.

The facilities are a virtual meeting place of an interdisciplinary group of experts. These groups include facility staff, hardware and software vendors, Grand Challenge applications researchers, industrial affiliates that want to develop industry-specific software, and academic researchers interested in advancing the frontier of high performance computing. HPCC funding is heavily leveraged by discipline-specific agency funds, equipment and personnel from hardware vendors, funds from state and local governments and universities, with industrial affiliates paying their fair share. Industrial affiliation offers a low risk environment for exploring and ultimately exploiting HPCC technology.

Production-quality operating systems software and software tools are developed at these facilities. Applications software developers access their resources over the Internet. Production-quality applications are often first run at these facilities, with user access increasingly at gigabit speeds. With their wide range of hardware and applications software, these are often sites for benchmarking systems and applications, and feedback is provided to hardware and software developers and vendors.

Other broadly common features of the facilities are extensive K-12 and undergraduate educational opportunities; training for researchers, graduate students, and faculty; and publication of articles in professional journals, annual reports, and newsletters.

Many of the systems listed below are funded by the named agency and receive additional funding from other HPCC agencies. For example, funding for systems at NSF centers also comes from ARPA, NASA, and NIH.

The resources at each facility and their key focus areas are as follows:

NSF Supercomputer Centers

NSF funds four Supercomputer Centers and augments the computing facilities at NCAR (the National Center for Atmospheric Research). The systems at these centers are listed below.

The term Metacenter refers to the joint cooperative activities of these centers and others in naturally overlapping research and technology areas. The Metacenter facilitates collaboration, communication, technical progress, and interoperability among participating institutions. To encourage further extension and collaborative integration of Metacenter activities with those of other providers of high performance computing, communications, and information infrastructure, in FY 1994 NSF initiated a program of Metacenter Regional Alliances (MRAs) resulting in six awards. In addition to augmenting national support activities, these MRAs are intended to complement, expand, and strengthen existing Metacenter activities at the regional, state, or local level. Participants in MRAs are expected to prototype new local and regional activities having the potential for being broadly replicated. NSF intends to add a similar number of MRAs in FY 1995.

http://pscinfo.psc.edu/MetaCenter/MetaScience/welcome.html

Cornell Theory Center (CTC), Ithaca, NY

The main CTC system is a 512-processor IBM SP-2. One CTC focus area is a globally scalable computing environment, including mass storage, I/O capability, networking, archival storage, data processing power, and graphics power.

http://www.tc.cornell.edu/ctc.html

National Center for Supercomputing Applications (NCSA), Urbana-Champaign, IL

Resources include:

- Convex C-3880 (with 8 processors, 4 GB memory, and 100 GB high speed disk)
- Convex Exemplar (one 8-processor hypernode, 512 MB memory, and 4 GB disk, with an upgrade to a system with 64-processor, 8 GB memory, and 180 GB high speed disk in FY 1995)
- Silicon Graphics Power Challenge (with 16 processors, 4 GB memory, and 100 GB high speed disk)
- Thinking Machines CM-5 (with 512 compute nodes, 16 GB memory, and 140 GB high speed parallel disk)

The three-tiered NCSA network consists of (1) Ethernet or FDDI to the desktop; (2) FDDI backbone between buildings, high-end systems, and the Internet; and (3) HiPPI between high performance computing systems, mass storage, and high-end peripherals.

NCSA is also involved in ATM research in (1) a local area network, (2) a trans-continental 155 Mb/s (SONET OC-3) national network, and (3) the BLANCA gigabit testbed at 622 Mb/s (SONET OC-12).

NCSA's virtual reality CAVE is described on page 27, and NCSA Mosaic is described on pages 6 and 7.

http://www.ncsa.uiuc.edu/General/NCSAHome.html

Pittsburgh Supercomputer Center (PSC), Pittsburgh, PA

Resources include:

- The first single-vendor heterogeneous system consisting of a Cray Research T3D (with 512 processors, each with 64 MB of memory) coupled to a C90 (with 16 processors and 4 GB of memory)

- 14-processor DEC Alpha workstation cluster
- HiPPI and FDDI network connecting these resources to each other and to storage devices

http://pscinfo.psc.edu/

San Diego Supercomputer Center (SDSC), San Diego, CA

Resources include:

- Cray Research C90 (with 8 vector processors)
- Intel Paragon (with 400 processors)
- Thinking Machines CM-2 (with 8,192 processors)
- Eight-workstation DEC Alpha cluster

An environmental modeling case study appears on page 87.

http://www.sdsc.edu/SDSCHome.html

National Center for Atmospheric Research (NCAR), Boulder, CO

NSF HPCC funds enabled NCAR to acquire a 64-processor Cray Research T3D and an 8-processor IBM SP-1 for use in the global climate modeling Grand Challenge.

http://www.ucar.edu/homepage.html

NSF Science and Technology Centers

Each of these four Centers addresses a particular research area; common to all four is cross-disciplinary focus, knowledge transfer and links to the private sector, and education and outreach. The Centers are:

The Center for Research in Parallel Computation (CRPC) at Rice University

CRPC aims to make parallel computing systems as easy to use as conventional computing systems — efforts include HPF, PVM, MPI, and NHSE, HPC++, algorithms for physical simulation, algorithms using parallel optimization, and ScaLAPACK. These are described beginning on page 23.

http://www.cs.rice.edu/CRPC/bluebook/bluebook.html

The Center for Computer Graphics and Scientific Visualization at the University of Utah

This Center is building and displaying models that are visually and measurably indistinguishable from real world entities.

http://www.graphics.cornell.edu/GVSTC.html

The Center for Discrete Mathematics and Theoretical Computer Science headquartered at Rutgers University

This Center is applying discrete mathematics and theoretical computer science to solving fundamental problems in science and engineering; in FY 1995 a special year on Mathematical Support for Molecular Biology, focusing on DNA sequencing and protein structure, was begun; in FY 1996 a special year on Logic and Algorithms, focusing on the relationship between mathematics and computational algorithms, will begin.

http://dimacs.rutgers.edu/

The Center for Cognitive Science at the University of Pennsylvania

This Center studies the human mind through the interaction of disciplines such as psychology, philosophy, linguistics, logic, and computer science. Work in human cognition, perception, natural language processing, and parallel computing has applications in robotic and manufacturing systems, human-machine interfaces, and language teaching and translational tools.

http://www.cis.upenn.edu/~ircs/homepage.html

NASA Testbeds

NASA maintains testbeds throughout the country to offer a diversity in configuration and capability. The testbeds are:

Ames Research Center, Moffett Field, CA

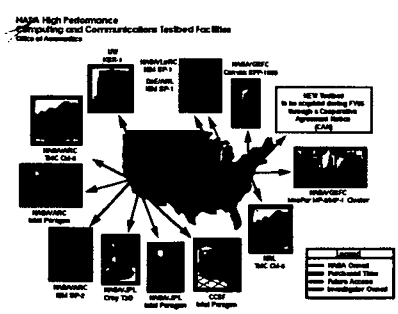
Resources include:

- IBM SP-2 (with 160 processors, 20 GB memory, 320 GB mass storage, 11.6 Gflops sustained performance)
- Intel Paragon (with 208 compute nodes, 16 service nodes, 7 GB memory, 16 Mflops peak speed)
- Thinking Machines CM-5 (with 128 compute nodes — each consisting of a SPARC processor and 4 vector processors, 4 GB memory, 48 GB scalable disk array, peak speed 16 Gflops), which is also used by the Naval Research Laboratory

Goddard Space Flight Center, Greenbelt, MD

Resources include:

- Convex SPP-1 (with 8 processors)
- MasPar MP-2 (16,384 processors, 1 GB memory, 22 GB parallel disk array, 6.2 Gflops peak speed, SIMD (Single Instruction Multiple Data)); MP-1 (8,192 processors, 512 MB memory, 0.6 Gflops); 2 MP-1 (4,096 processors, 256 MB memory, 0.3 Gflops each). These four systems have all been connected by a 4-by-4 HiPPI switch. Peak performance is 7.5 Gflops. MasPar applications are being modified to take advantage of the HiPPI network in order to distribute the workload across the four MasPar systems. NASA Parallel Benchmark codes are being implemented on the cluster in FY 1995. This cluster demonstrates combining SIMD and MIMD (Multiple Instruction Multiple Data) programming styles and enables MasPar to move beyond its current 16,384-processor ceiling.



http://cesdis.gsfc.nasa.gov/hpccm/accomp/94accomp/94accomps.html

Jet Propulsion Laboratory (JPL), Pasadena, CA

Resources include:

- Cray Research T3D (with 256 processors — each processor is a DEC Alpha with peak speed 150 Mflops, 16 GB memory, and 103 GB disk)
- Intel Delta (with 528 processors) at Caltech
- Intel Paragon (with 56 compute nodes each with 32 MB memory, 4 I/O nodes each with 32 MB memory and 4.8 GB RAID, one HiPPI node with 32 MB memory, one user service node with 32 MB memory, and one boot node with 32 MB memory and 4.8 GB RAID)

The Paragon came into service in August, 1993, for Earth and Space Science Grand Challenge applications development. Peak performance is 5.6 Gflops in single precision (using 56 nodes with 1.8 GB aggregate memory and more than 20 GB aggregate on-line disk). The system is available to NASA HPCC investigators and select collaborators. Interactive access to all compute nodes is available 24 hours a day, seven days a week.

Langley Research Center, Hampton, VA

Resources include an Intel Paragon (with 72 compute nodes, 3 service nodes, 8 I/O nodes, 2 Ethernet nodes, 2 HiPPI nodes, and 2 FDDI nodes, each with 32 MB memory and 38 GB disk).

Lewis Research Center, Cleveland, OH

Resources include an IBM SP-1 (with 16 processors, 2 GB disk, 800 Mflops peak speed).

Other Resources

These include:

- IBM RS6000 workstation cluster (32 nodes, 3 GB memory, 1.6 Gflops peak speed) that is part of a 128-node IBM SP-1 consortium at Argonne National Laboratory

- Kendall Square Research KSR-1 (56 nodes, 1.8 GB memory, 3 GB disk, MIMD) at the University of Washington

Further information about these NASA testbeds is at:

http://cesdis.gsfc.nasa.gov/hpccm/accomp/94accomp/ess94.accomps/ess4.html

DOE Laboratories

National Energy Research Supercomputer Center (NERSC), Lawrence Livermore National Laboratory (LLNL), Livermore, CA

The Supercomputer Access Program at NERSC provides production computing for investigators supported by the Office of Energy Research in the following areas: material sciences, chemistry, geosciences, biosciences, engineering, health and environmental research, high energy and nuclear physics, fusion energy, and applied mathematics and computational science. The Center serves more than 4,000 accounts involved in some 700 projects. NERSC resources include:

- Cray Computer Cray-2 (with 8 processors, 128 megawords (millions of 64-bit words (Mw) memory, and 84 GB of rotating disk)

- Cray Computer Cray-2 (with 4 processors, 128 Mw memory, and 64 GB of rotating disk)

- Cray Research C90 (with 16 processors, 256 Mw memory, 4 GB of solid state disk, and 150 GB of rotating disk)

- The National Education Supercomputer, a four processor Cray Research X-MP EL provided by Cray Research that is available to high schools over the Internet

Anticipated FY 1995 accomplishments include providing initial production use of a storage system based on National Storage Laboratory (NSL) technology to Energy Research programs, and bringing in a pilot early production massively parallel system. FY 1996 plans include replacing this system with a fully configured system, increasing disk capacity and robotic tape capabilities for the storage system, installing Kerberos-based security for NERSC services, and providing client access to the Distributed File System (DFS) from major NERSC computing platforms. DFS is a distributed file system based on the Andrew File System (AFS), which is now part of the Open Software Foundation's Distributed Computing Environment (DCE).

Los Alamos National Laboratory (LANL), NM, and
Oak Ridge National Laboratory (ORNL), TN

These DOE HPCC Research Centers provide full-scale high performance computing systems for work on Grand Challenge applications and use in scalability studies. These applications must be run on large prototype systems — they cannot be scaled down without removing essential aspects of their physics.

LANL operates a Thinking Machines CM-5 (1,024 compute nodes, 32 GB memory, 128 GB Scalable Disk Array using four HiPPI channels). This system has 128 Gflops theoretical performance with 50 Gflops observed on several codes.

ORNL resources include:

- Intel Paragon XP/S 150 (with 1,024 MP nodes (with 3,072 processors), 960 of the nodes have 64 MB of memory and 64 of the nodes have 128 MB of memory for a total of 70 GB of memory, plus five service nodes and 127 I/O nodes each connected to a 4.8 GB RAID disk for a total of 600 GB of system disk space). An MP node has two compute processors and one message passing processor, though under certain conditions all three can be used for computations.

- Intel Paragon XP/S 35 (with 512 GP nodes, each with 32 MB of memory for a total of 16 GB of memory, plus 5 service nodes and 27 I/O nodes each connected to a 4.8 GB RAID disk for a total of 130 GB of system disk space). A GP node has one compute processor and one message-passing processor, though under certain conditions both processors can be used for computations.

- Intel Paragon XP/S 14 (with 96 MP nodes — 32 nodes having 128 MB and 64 nodes having 64 MB for a total of 8 GB of memory, and 38.4 GB of system disk space); and Intel Paragon XP/S 5 (with 66 GP nodes, each with 16 MB of memory for a total of 1 GB of memory, and 24 GB of system disk space).

- Kendall Square KSR 1 (with 64 processors, each with 32 MB of memory for a total of 2 GB of memory, and two 10 GB RAID disks) used to study shared memory algorithms

- nCube2 (with 8 processors, each with 32 MB of memory for a total of 256 MB of memory, and a 2 GB disk on the Sun front end) used by high school students in the Adventures in Supercomputing program

  http://www.ornl.gov/ocnm/AiS/AiS.html

Several system upgrades are anticipated in FY 1995. FY 1996 plans include expanding DOE storage capabilities by implementing the HPSS network-centered parallel storage system management software.

NIH Systems

The Division of Computer Research and Technology (DCRT) has a 128-processor Intel iPSC/860 with 2 GB memory. In FY 1995 DCRT will install a 10- to 20-Gflops massively parallel system. Both systems are used by NIH staff in biomedical applications.

http://hpccwww.dcrt.nih.gov/

The National Cancer Institute (NCI) Frederick Biomedical Supercomputing Center has an 8-processor Cray Y-MP and a MasPar MP-2 with 4,096 processors along with a comprehensive collection of biomedical software available to all scientists who use the facility.

http://www-lmmb.ncifcrf.gov/

http://www-ips.ncifcrf.gov/

http://www-pdd.ncifcrf.gov

The National Center for Research Resources (NCRR) supports systems for biomedical research applications at its High Performance Computing Resources Centers at CTC, PSC, SDSC, NCSA, and Columbia University.

NOAA Laboratories

The Forecast Systems Laboratory in Boulder, CO, has a 221-processor Intel Paragon, with 6.5 GB memory and 28.8 GB disk, rated 15 Gflops peak. This system is used to parallelize regional and mesoscale forecast models.

The Geophysical Fluid Dynamics Laboratory (GFDL) in Princeton, NJ, is acquiring a high performance computing system including massively parallel capabilities in FY 1995, and the National Meteorological Center in Camp Springs, MD, will acquire a system in FY 1996. These systems are to be used for the global climate modeling and weather forecasting Grand Challenges. The GFDL procurement is funded in part by NOAA HPCC funds; it will provide an order of magnitude improvement in performance, have more advanced data archiving facilities, and use the new scalable models now being developed to address important climate problems.

EPA Systems

EPA's National Environmental Supercomputing Center in Bay City, MI, has a Cray C90 (with 3 processors, 64 Mw memory, and 90 GB disk). In FY 1995 EPA plans to acquire a scalable massively parallel system for installation at this center. These systems are dedicated to environmental research and problem solving.

6. Grand Challenge Applications

A key goal of the HPCC Program is to demonstrate the use of high performance computing and communications technologies to discover new knowledge and illustrate new capabilities that were not possible with earlier technologies. The Program has conducted R&D in Grand Challenge applications, which are fundamental problems in science and engineering with broad economic and scientific impact whose solution can be advanced by applying HPCC technologies. In addition to scientific importance, selection criteria included potential for cost sharing with sources directly concerned with the specific applications and the potential for leveraging across disciplines.

A common feature of many of these Grand Challenges is that they involve simulation. In part because of HPCC technologies, simulation has become recognized as the third paradigm of science, the first two being experimentation and theory. In some cases it is the only approach available for further advancing knowledge — experiments may not be possible due to size (very big or very small), speed (very fast or very slow), distance (very far away), dangers to health and safety (toxic or explosive), or the economics of conducting the experiments. In simulations, mathematical models of physical phenomena are translated into computer software that specifies how calculations are performed using input data that may include both experimental data and estimated values of unknown parameters in the mathematical models. By repeatedly running the software using different data and different parameter values, an understanding of the phenomenon of interest emerges. The realism of these simulations and the speed with which they are produced affect the accuracy of this understanding and its usefulness in predicting change.

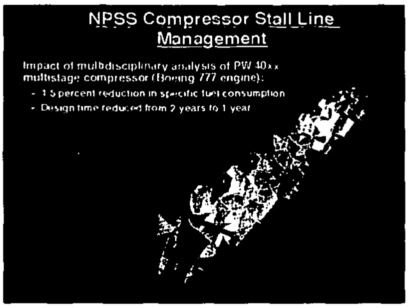
Due to limitations such as speed and memory in computing systems available at the beginning of the HPCC Program, many simulations could not be completed with sufficient accuracy and timeliness to be of interest. Through efforts requiring collaboration among computer scientists, mathematicians, computational scientists, and subject matter specialists, these limitations are being removed. In particular over the past year many Grand Challenge teams have reported advances that have been the result of faster run times, larger memory, higher resolution and more realistic modeling. Several dozen Grand Challenge applications projects and Grand-Challenge-scale applications are now described. Many of these involve multi-agency cooperation and support. In particular, all of the NSF Grand Challenges described are also supported by ARPA.

Applied Fluid Dynamics

Computational Aeroscience

NASA's Numerical Propulsion Simulation System (NPSS) focuses on reducing both the cost and the time to develop aeropropulsion systems. NPSS will build a simulation environment that allows for the construction of arbitrary engine configurations for design and analysis. The environment will permit a choice of analysis techniques, analysis complexity, languages with which to describe the algorithms, and the ability to access and manage data from various sources.

NPSS MOD1 was released in 1994. It employs an object-based model for engine simulations: engine components, such as the compressor, combustor, turbine, and shaft, can be modeled as independent entities that can be replaced with models of greater fidelity that execute on differ-



Analysis to define the flow physics involved in compressor stall. It suggested a variety of approaches to improve the performance of compression systems, while providing increased stall margins. A Cray Research C-90, IBM SP-1, and IBM workstation cluster were used to formulate and develop this model.

ent computing platforms in a dynamic environment. An engineer can configure engines with ease using this environment and a graphical user interface.

This work is being conducted at NASA Lewis Research Center in Cleveland, OH. Researchers at Lewis are working with the U.S. aeropropulsion industry to define a U.S. standard for select design codes.

One component of NPSS is the MSTAGE multistage compressor analysis software that has been implemented on the Cray Research C90, IBM SP-1, and IBM workstation cluster at NASA Lewis. This software has been used to analyze the flow physics involved in compressor stall that has suggested several approaches to improving compression system performance and increasing stall margins. This work is part of a NASA/MIT/Pratt and Whitney effort. A 1.5 percent reduction in Specific Fuel Consumption (SFC) for a large commercial aircraft engine was demonstrated at Pratt and Whitney; this result was achieved in half the historical design time.

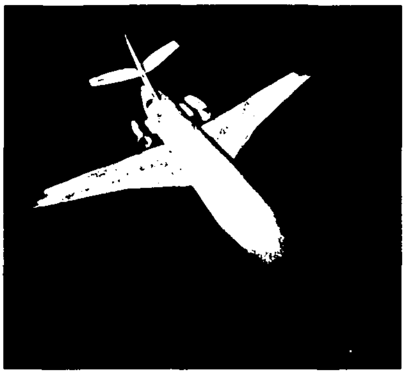
Other NASA accomplishments include the installation of a 160-node IBM SP-2, located at NASA Ames Research Center, which performs at a sustained rate of 11.64 Gflops in NASA Parallel Benchmark codes; this was accomplished through a Cooperative Research Agreement, a new method of doing business for NASA. Additionally, the Framework for Interdisciplinary Design and Optimization (FIDO) project, located at NASA Langley Research Center, is developing a general computational environment for performing multidisciplinary design using networked heterogeneous computing systems. Such a graphical interface provides easier understanding and access to data than previous text-based methods, and requires less training for users.

http://cesdis.gsfc.nasa.gov/hpccm/accomp/94accomp/cas94.accomps/cas4.html

Coupled Field Problems and GAFD (Geophysical and Astrophysical Fluid Dynamics) Turbulence

This project represents engineering and scientific Grand Challenges linked through common enabling computer science advancements: High Performance Computational Methods for Coupled Field Problems and Coherent Structures and Dynamics in Geophysical and Astrophysical Turbulent Flows. The former addresses three engineering problems that exhibit strong interaction between components modeled by different disciplines: (1) aeroelasticity of a complete aircraft, (2) distributed control of flexible structures, and (3) coupling of electromagnetic, thermal, and superconducting phase-change behavior. The latter addresses four problems in oceans, atmospheres, and stars in which turbulence coexists with large-scale coherent structures and mean flows: (1) geostrophic turbulence, (2) ocean convection, (3) deep convection in planetary atmospheres, and (4) compressible convection in stars like the sun, all constrained by the effects of rotation and stratification. Both efforts are enabled by advances in networking, algorithm development, programming environments, and performance prediction and analysis.

Aeroelasticity studies the mutual interaction between aerodynamic and elastic forces for an aerospace vehicle. A flexible aircraft structure immersed in a flow is subjected to surface pres-



An image from a video illustrating the flutter analysis of a FALCON jet under a sequence of transonic speed maneuvers. Areas of high stress are red; areas of low stress are blue.

sures induced by that flow. Moreover, structural dynamic motions induced by these pressures in turn change the boundary conditions of the flow. The accurate prediction of aeroelastic phenomena requires extensive computation to solve simultaneously the coupled fluid and structural equations of motion. For example, the aeroelastic response of a detailed wing-body configuration using potential flow theory requires about five CPU hours on a Cray Computer Cray-2. To establish the transonic flutter boundary for a given set of aeroelastic parameters, about 30 aeroelastic response analyses are typically required, which brings the total CPU time to six days. The aeroelastic simulation of a complete aircraft (shown above) is made possible by the new generation of massively parallel computing systems as well as by methods developed under this Grand Challenge. This simulation required the simultaneous solution of 463,674 nonlinear fluid equations and 45,108 linear structural equations, thousands of times. These computations were carried out in a heterogeneous mode on a 128-processor Cray Research T3D, a 128-processor IBM SP-2, and a 128-processor Intel Paragon XP/S.

In another effort, a billion-zone (1,024-by-1,024 by 1,024) turbulence problem was solved on an array of 16 Silicon Graphics Challenge XL workstations with 28 GB total system memory. The computation required a week and achieved a sustained computational rate of 4.9 Gflops. The workstations were configured as a 2-by-2-by-4 toroidal array using FDDI networking. This computation demonstrated the high performance of a networked cluster of shared memory multiple-processor machines.

This team is also installing their software on systems from different vendors, including the Cray Research T3D. In these efforts they are early users of implementations of High Performance Fortran (HPF) compilers, participate in the Message Passing Interface Forum, and develop performance analysis tools. Participants in this NSF-funded project include the University of Colorado, NCAR, and the University of Minnesota.

http://lcd-www.colorado.edu/Blue_book/bb_joint.html

Combustion Modeling: Adaptive Grid Methods

Combustion is a major source of energy, plays a dominant role in transportation, and is an important factor in many industrial processes. Requirements for energy efficiency and for emission reduction have led industry to increased use of computer simulations to design combustion devices. One such device is the pulse combustor, which is characterized by a periodic combustion process. Available simulation software has been limited in its ability to represent the detailed physical processes and the complex geometries of such practical engineering devices. This Grand Challenge team is developing parallel software that models fully three-dimensional fluid dynamics in the combustion chamber and incorporates lower-dimensional approximations for the inlet valve that controls the injection of fuel and the tail pipe whose acoustic properties control the cyclic behavior of the device.

Participants in this project include researchers at DOE's Lawrence Livermore and Los Alamos National Laboratories, the Courant Institute for the Mathematical Sciences, and the University of California at Berkeley. They are collaborating with Coen Co. of Burlingame, CA, and Babcock

Fuel flow around the stagnation plate in a pulse combustor. A burning cycle drives a resonant pressure wave, which in turn enhances the rate of combustion, resulting in a self-sustaining, large-scale oscillation. The figure shows the injection phase when the pressure in the combustion chamber is low. Fuel enters the chamber, hits the stagnation plate and becomes entrained by a vortex ring formed by flow separation at the edge of the splash plate. Researchers are developing computational models to study the interplay of vortex dynamics and chemical kinetics and will use their results to improve pulse combustor design.

and Wilcox of Alliance, OH, in extending and validating their numerical methods to simulate low-NOx natural gas burners. They also participate in benchmarking software for modeling natural gas burners sponsored by the Gas Research Institute that is funded by the gas production industry.

http://www.nersc.gov/doc/Comp_Research/CCSE/bb96.html

Oil Reservoir Modeling: Parallel Algorithms for Modeling Flow in Permeable Media

This DOE-funded Grand Challenge has the goal of using high performance parallel processing to simulate (1) the behavior of petroleum reservoirs to enhance oil recovery, and (2) groundwater aquifers to aid removal of contaminants from fresh water aquifers. Improved understanding of petroleum reservoirs leads to better reservoir management and more efficient U.S. oil and gas production. Numerical modeling of fluid flow in permeable media is critical to the management and protection of groundwater supplies. These two problems have common scientific and engineering fundamentals involving multiphase transport in permeable media.

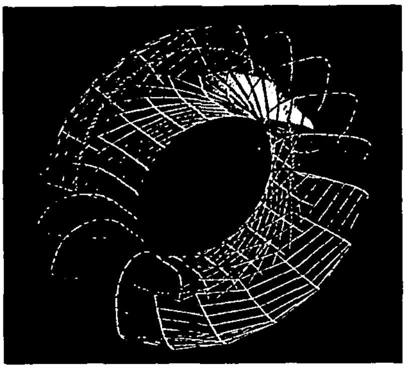
Using current scalar and vector computing systems, most industrial reservoir engineering codes use fewer than 100,000 gridblocks and model fewer than 1,000 wells. Key geological, physical, and chemical features are only crudely approximated. Preliminary studies suggest that a million gridblocks are feasible with parallel computing technologies, enabling more accurate and more useful results.

This work is being conducted by interdisciplinary teams in applied mathematics, computational science, and chemical, petroleum, and environmental engineering at the University of Texas at Austin, Rice University, and industrial affiliates. Participants have developed accurate and efficient serial and parallel numerical algorithms for solving linear and nonlinear, coupled partial differential equations, including advective dominated transport equations and elliptic/parabolic flow equations; have developed efficient parallel domain decomposition solvers for the large sparse linear systems that result from temporal and spatial discretizations; and work with computational science applications in modeling subsurface multiphase flow and multi-component reactive transport, surface water, and root-soil systems, and the interpretation of microscopic data on macroscopic scales. The reservoir simulation study was used to evaluate the use of horizontal wells with vertical drainholes for tertiary oil recovery using carbon dioxide.

http://www.pe.utexas.edu/HPCC/hpcc.html

Numerical Tokamak Project (NTP)

The goal of DOE's NTP Grand Challenge is to develop and integrate particle and fluid plasma models on large-scale parallel computing systems as part of a multidisciplinary study of Tokamak fusion reactors. The kinetic particle simulations and the fluid simulations have different physics and computational attributes and advantages, and comparisons of their results enable calibration and lead to improvements in



Particle trajectories and electrostatic potentials from a three-dimensional implicit tokamak plasma simulation employing adaptive mesh techniques. The boundary is aligned with the magnetic field that shears around the torus. The strip in the torus is aligned with the local magnetic field and is color mapped with the local electrostatic potential. The yellow trajectory is the gyrating orbit of a single ion.

each and to the development of hybrid models that embody aspects of both.

Accomplishments include (1) the development and optimization of a suite of fluid and kinetic three-dimensional simulation codes for massively parallel computing systems, (2) testing and comparison of various MPP (massively parallel processing) systems and programming paradigms and studies of MPP performance scaling with the number of processors, (3) improvement of code performance by a factor of 10 to 100 with careful MPP optimization and by an additional factor of 10 to 100 with optimized spatial grids, (4) development of advanced perturbative, implicit, and hybrid algorithms to improve simulation efficiency further, (5) development of new tools to handle storage and retrieval of large data sets and to post-process and visualize the data interactively on distributed communications networks, (6) building portable code and code modules for MPPs and clusters, class libraries, parallel fast Fourier transforms and elliptic matrix solvers, convolution routines, and parallel I/O using NetCDF software.

These accomplishments have enabled or accelerated (1) larger, more efficient three-dimensional toroidal gyrofluid and gyrokinetic simulations relevant to Tokamak experiments using a Thinking Machines CM-5 and a Cray Research C90, (2) code comparisons leading to closer agreement between gyrokinetic and gyrofluid simulations of a Princeton Tokamak Fusion Test Reactor (TFTR) test case and studies of the scaling of the turbulent transport with respect to the physical parameters of interest and the isotopic composition of the plasma, (3) advances in simulated thermal diffusivities, fluctuation spectra, and parameter scaling agreeing more closely with experiments (for example TFTR discharges), (4) quantitative determination of the importance of nonlinear self-generated sheared flows in influencing turbulence, (5) progress in simulations of ion-temperature-gradient, trapped electron, and alpha-particle-driven toroidal Alfven eigenmode instabilities, and (6) inclusion of more physics in the simulation (impurities, collisions, kinetic electrons, velocity shear, toroidicity, curvature drives and resonances, and magnetically trapped particles).

Algorithmic improvements and use of large-scale parallel computing systems result in 100- to million-fold performance improvements. Information processing and visualization tools accelerate comparison of computational models to each other, to experimental data, and to analytic theory, enabling better understanding of the target physics.

This work is being conducted at NERSC, LLNL, LANL, and PSC.

http://www.acl.lanl.gov/GrandChal/Tok/BlueBook.html

Meso- to Macro-Scale Environmental Modeling

NOAA, EPA, NASA, DOE, and NSF are involved in these Grand Challenges and those in the next section on Ecosystem Simulations. These agencies collaborate on developing shared models of the Earth's atmosphere and oceans that the individual agencies use for global climate modeling and agency-specific applications. NOAA uses them for weather modeling (for use in weather forecasting), EPA for pollution modeling (using meteorological models that are modified weather models), DOE for use in groundwater management and environmental remediation, and NSF for basic research in subjects such as earthquake motion and land cover dynamics. The global climate modeling being conducted by these agencies builds on these basic models, mission-specific applications, and data. This work is now described in detail.

Massively Parallel Atmospheric Modeling Projects

Atmospheric models are used in weather modeling (for example, in understanding the mechanisms that control the development of severe storms) and in environmental modeling (for example, in understanding the impact of emissions controls on air quality). They have been optimized for vector supercomputers, where they have been run for years. These computers have several limitations, notably limited memory, high cost, and uncertain future scalability. Memory limits restrict the geographical extent and resolution of the simulations, and their cost prohibits the regional deployment of the models that will be required for planning future air quality measures.

Parallel computing offers significant improvements in raw speed and cost performance. Speed improvements allow scientists to address larger problems, more complex model processes, and finer resolutions. Cost performance improvements allow wider deployment — a local airport can run its own weather model, a state government its own air quality model. Such a cost-effective portable scalable parallel atmospheric circulation model has been developed by researchers at Argonne and NCAR. The model is a parallel version of NCAR's Mesoscale Model 5 (MM5) that can resolve distance scales ranging from continental to one kilometer or less through the use of specialized numerical solvers and support for nested grids. The model incorporates new algorithms that scale to the number of processors and address issues such as load balancing that are unique to massively parallel systems. New programming techniques were developed to simplify the production of portable efficient parallel models. The model has been run on systems with hundreds to thousands of processors that provide the large memory and computational power needed for high resolution long duration simulations; these include the Cray Research T3D, IBM SP, and Intel Delta and Paragon. It can be run on small clusters of workstations connected by high speed networks, providing cost effective simulations when high resolution is not required.

MM5 is a building block for EPA's next-generation air quality modeling and decision support system, Models-3, that is being developed to improve both the scientific accuracy and accessibility of modeling tools and the data used in air quality management. By using distributed computing techniques, a model can be run on high-end massively parallel computing systems in research labs as well as on small workstation clusters in local government offices. Deployment is expected to result in improved aircraft safety, air quality control measures, and understanding of meteorological processes.

HPCC funding comes from EPA and DOE; the U.S. Air Force provides additional support. This project complements the DOE CHAMMP (Computer Hardware, Advanced Mathematics, and Model Physics) Program that supports development of the Parallel Community Climate Model, a scalable global atmospheric model, at Argonne, Los Alamos, and Oak Ridge.

http://www.mcs.anl.gov/home/itf/epa.html

The SKYHI general circulation model developed by NOAA/GFDL has been used for a decade to investigate the dynamical behavior of the stratosphere and mesosphere. The new version is scalable and is designed to run efficiently on both shared and distributed memory systems under either data parallel or message passing programming models. It allows scientists to incorporate new physics software modules of atmospheric features such as radiation and convection.

A new scientific experiment is investigating how waves generated by tropospheric disturbances, such as thunderstorms and flow over mountains, affect winds in the stratosphere and mesosphere. This experiment will be run on the 1,024-node Thinking Machines CM-5 at DOE/LANL and will use a grid resolution greater than can be achieved on the supercomputer at NOAA/GFDL. This collaboration is supported by HPCC funds through NOAA and DOE/CHAMMP.

Parallel Ocean Modeling

The GFDL Modular Ocean Model (MOM) is public-domain software that is being used by more than 200 scientists in more than 20 countries. Research topics include climate studies (both present day and paleoclimate), data assimilation, predictability, and basic research into understanding world ocean processes on a number of space and time scales. A new version, MOM-2, is being released in FY 1995. One feature of the new version is a flexible memory window that allows solution for a group of latitude rows at one time; as a result MOM-2 can be run on a range of computer architectures with the number of latitude rows chosen to fit the particular system. The modular design of the model also permits alternate ways to treat the atmosphere-ocean boundary at the top of the model and provides an improved way to represent the effects of rough bottom terrain.

http://www.hpcc.noaa.gov/Ocean94

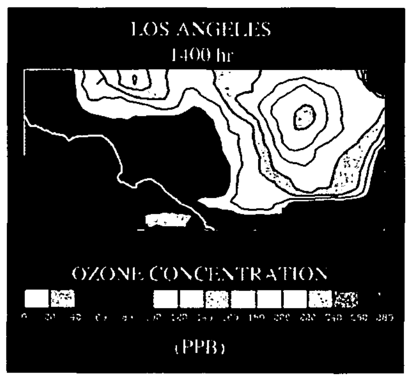
The Parallel Ocean Program (POP) at DOE/LANL was developed to perform high resolution ocean simulations. Recent POP results are shown in the lower image on page 4, where the significance of the calculations is also described.

http://www.acl.lanl.gov/

Mathematical Modeling of Air Pollution Dynamics

Massively parallel computing systems provide an avenue for overcoming the computational requirements in the study of atmospheric chemical dynamics. Implementation issues include domain decomposition strategies, algorithm design and evaluation, portability, modularity, and buffering techniques used in I/O operations. The Caltech urban air pollution model has been implemented on distributed memory MIMD systems including a 512-node Intel Paragon and a workstation cluster.

The central challenge in developing a parallel air pollution model is implementing the chemistry and transport operators used to solve the atmospheric reaction-diffusion equation. The chemistry operator is generally the most computation-



Ozone concentrations for the California South Coast Air Basin predicted by the Caltech research model show a large region in which the national ozone standard of 120 parts per billion (ppb) are exceeded. Measurement data corroborate these predictions. Scientific studies have shown that human exposure to ozone concentrations at or above the standard can impair lung functions in people with respiratory problems and can cause chest pain and shortness of breath even in the healthy population. This problem raises concern since more than 30 urban areas across the country still do not meet the national standard.

Combining Grand Challenge and National Challenge Technologies

When the HPCC Program began, it focused on "Grand Challenges," computationally intensive problems in science and engineering. With the addition of the Information Infrastructure Technology and Applications component, the HPCC Program also began addressing the "National Challenges," information-intensive applications impacting U.S. competitiveness and societal well being. There is a class of applications that span both of these types of challenges. One example, Global Climate Modeling, was described on page 4. A second, Weather Forecasting, appears opposite. Three others appear on pages 85 through 87.

ally intensive step in atmospheric air quality models, and a new method based on Richardson extrapolation to solve the chemical kinetics has been developed. The transport operator (advection equation) is the most challenging to solve numerically. Because of its hyperbolic nature, non-physical oscillations and/or negative concentrations appear near steep gradient regions of the solution. Six algorithms for solving the advection equation have been evaluated for their suitability for use in parallel photochemical air quality models. A speedup factor of 94.9 has been measured when chemistry, transport, and I/O are done in parallel. This work provides the computational infrastructure needed to incorporate new physico-chemical phenomena in the next generation of urban- or regional-scale air quality models. HPCC provides the tools essential to develop our understanding of air pollution further.

Funding for this work came from EPA and the IBM Environmental Research Program. The research was performed at Caltech using the Intel Delta and Paragon systems at Caltech's Concurrent Supercomputing Consortium.

http://nicarao1.che.caltech.edu/~dabdub

Additional EPA-funded research on optimizing the performance of atmospheric chemistry solvers conducted at North Carolina State University demonstrated superlinear speedup, that is, the ratio of execution time on a single processor to that of multiple processors is greater than the number of processors. The explanation for this anomalous result is that the data size of the problem exceeds the memory of a single processor, and thus incurs a data-movement penalty in the single-processor case. Using multiple processors allows the data to be spread over the system, accommodating much larger problem sizes. The research indicates excellent scalability for the Cray Research T3D from 8 to 128 processors.

http://www.csc.ncsu.edu/departmental/proposals/

A Distributed Computational System for Large Scale Environmental Modeling

In the past, the only way to determine the efficacy of solutions to various environmental problems was to implement a set of control strategies and measure the results. When the problem was air quality, this meant using the only experimental laboratory available, namely the atmosphere. More recently, computational models of the physical and chemical processes that take place in the atmosphere allow trials with different control strategies without the expense and difficulty of real-world experiments. These computationally intensive mathematical models are benefiting from the use of parallel and distributed algorithms and mass storage resources.

An NSF-funded Grand Challenge team is studying the effects of six different alternative fuels in each of 12 scenarios. While the study will require over 1,000 CPU hours (45 CPU days), it will be done in parallel on a collection of high performance workstations and will be completed in a week.

This work is conducted at Carnegie Mellon University and MIT in coordination with EPA's

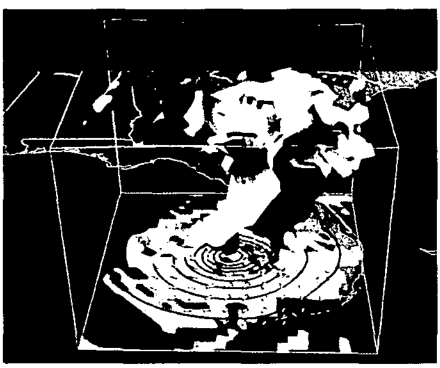
A Case Study Combining Grand Challenge and National Challenge Technologies:

Weather Forecasting

NOAA's Geophysical Fluid Dynamics Laboratory has developed a Hurricane Prediction System that is the result of more than a decade of R&D by a small group of GFDL scientists. During the last two hurricane seasons scientists at GFDL and at NOAA's National Meteorological Center (NMC) in Camp Springs, MD, have been comparing this new system with their operational hurricane forecast models.

During the 1993 hurricane season, GFDL ran the model in semi-operational test mode on its own computing system located in Princeton, using observational data and global forecasts provided by NMC. Early in the season, the model's prediction for the storm track of Hurricane Emily demonstrated its potential as a new forecasting tool. When Emily was 72 hours from possible landfall, the GFDL system predicted that the storm would move to the northeast, while all other forecasts predicted that Emily would move due west. In addition, GFDL's forecast predicted that the storm would approach the Outer Banks of North Carolina and then would make an abrupt turn to the northeast, moving back out to sea. This initial prediction, which was repeated in subsequent predictions in which the model used more recent observations, was followed closely by the actual storm track.

Because of its success in forecasting Emily and other tropical storms during the 1993 hurricane season, NMC decided to run the GFDL system as part of its operational hurricane forecast suite in a parallel test mode during the 1994 season. In order to do this, the model was rewritten to improve the wall-clock turn-around of its forecasts during storm emergencies. The redesigned, parallelized model exhibited a ten-fold improvement in wall-clock performance on NMC's Cray Research C-90 system compared to the model's performance on a single processor.

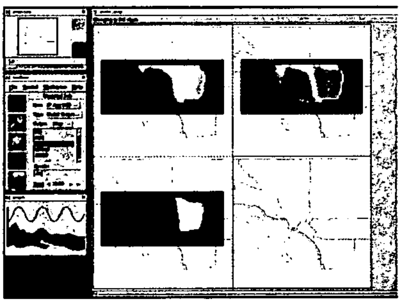


This cut-away view of a 72-hour forecast for Tropical Storm Gordon was made by the GFDL hurricane model. It was made on the morning of August 13th at which time Gordon was located in the Caribbean just southeast of Cuba. At this time, the other models forecast Gordon either would head toward the central Gulf or into the Atlantic east of Florida. The GFDL 72-hour forecast position was within approximately 40km (25 mi) of the observed position. The isosurface shown is the 85 percent relative humidity surface, and the color shading near the Earth's surface represents the instantaneous precipitation rates, indicating that the most intense rains were expected on the east side of the storm. This is typical of a weak tropical storm undergoing shearing by upper-level westerlies that in this case turned the system onto the Florida peninsula. The white arrows indicate the winds at the surface, and are drawn only where wind speeds exceed gale force. Contour lines of the sea level pressure show the characteristic circular pattern of a tropical storm.

During the 1994 season, the GFDL system was run for tropical systems at all development stages, from weaker tropical depressions to full-blown hurricanes. It forecasted 60 cases for the Atlantic, 148 cases for the eastern Pacific, and a few experimental cases for typhoons in the western Pacific. Comparisons of the GFDL model's storm track forecasts with those from current NMC hurricane models for 1994 tropical storms in the Atlantic indicate that the GFDL system is in the top performance group for forecasts out to 36 hours and is superior to all other forecast models at 48 and 72 hours.

Hurricane Gordon, one of the more important storms in the Atlantic during the 1994 season, provided a difficult test for all forecasting systems because of its complicated storm track. The GFDL model was the first forecast system running at NMC to predict, three days in advance, that Gordon would move west from Cuba through the Florida Strait and then turn northeast to cross the Florida peninsula. Later, as the storm moved north off the east coast, predictions from the GFDL system hinted at Gordon's abrupt U-turn back toward the south.

http://www.hpcc.noaa.gov/Weather94



Ozone concentration over Los Angeles for two scenarios. The upper left image shows predicted ozone concentrations if automobiles use compressed natural gas (CNG) fuel instead of standard gasoline. The upper right image shows estimated ozone concentrations without any change in the vehicle fleet. The bottom image shows the difference between the two panels; such visualizations allow rapid comparison.

research activities. Computing platforms available include a DEC Alpha Supercluster and the Cray Research T3D and C-90; in addition, parts of the simulation are available on the Intel Paragon and iWarp systems.

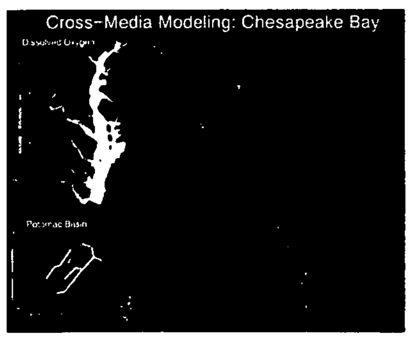
http://www.cs.cmu.edu:8001/afs/cs.cmu.edu/project/gems/www/current.html

Cross-Media (Air and Water) Linkage

Nitrogen is a major cause of eutrophication in coastal estuaries. It is the controlling nutrient in the Chesapeake Bay. Atmospheric nitrogen accounts for 25 to 35 percent of the nitrogen input (or "loading") to the Bay. Control of this atmospheric nitrogen loading may be crucial to efforts to restore coastal estuaries and appears to be critical to Bay restoration. A Bay Agreement that was signed in 1987 and renewed in 1992 calls for a 40 percent reduction in the amount of controllable nutrients reaching the Bay by the year 2000.

EPA is linking together air and water simulation models in order to (1) assess the impact of air pollution controls on nitrogen loading to the Bay, (2) assess the benefit of these controls on Bay restoration, and (3) link the models more effectively by reducing the temporal and spatial mismatches between the individual air and water models.

The three models that are being linked are (1) the Regional Acid Deposition Model (RADM) for atmospheric deposition, (2) the Hydrologic Simulation Program Fortran model for nutrient flow in the watershed to the Chesapeake Bay, and (3) the three-dimensional Chesapeake Bay Water Quality Model (CBWQM) of response to nutrient loading. In addition a weather model is used to drive the atmospheric model, and a separate three-dimensional hydrology model is used to simulate flows in the Bay. RADM and CBWQM were developed on Cray Research systems. The atmospheric model is the most computationally intensive and has the greatest temporal/spatial mismatch. It is being moved to a scalable parallel system.



(1) Dissolved oxygen in Chesapeake Bay, (2) nitrate loading in the Potomac Basin, and (3) atmospheric nitric acid and wet deposition across the Eastern U.S. Three air and water models are linked together for cross-media modeling of the Chesapeake Bay. Atmospheric nitrogen deposition predicted by the atmospheric model (right) is the input load to the watershed model and the three-dimensional Bay model. The watershed model (lower left) delivers nitrate loads from each of the water basins to the three-dimensional Bay model (upper left).

The grid resolution of the atmospheric model was reduced from 80 kilometers to 20 kilometers over the northeastern U.S. airshed. This will result in more accurate spatial linkage with the water quality models and in improved resolution of urban and point-source emission influences. This increased resolution quadruples computing time.

The airsheds for the Chesapeake Bay watershed and the Bay itself were determined by two sets of simulations. The first estimated average annual deposition of nitrogen oxide emissions for representative meteorology from ten subregions. The second looked at sulfur dioxide emissions. The simulations took 1,400 hours on a Cray Research Y-MP. The airshed was found to be four times the size of the watershed.

This work has been nominated for the annual Smithsonian Computer World Award in the environmental category.

http://www.epa.gov/docs/HPCC/homep.html

Adaptive Coordination of Predictive Models with Experimental Observation

This NSF-supported Grand Challenge, centered at Stanford University, addresses several issues related to geophysical simulation and prediction, the use of observing systems for data acquisition, and the assimilation of the measured data into numerical simulations. The basic idea is that a predictive computer model, carrying out a simulation faster than real time, can be used to estimate what data need be gathered, as well as the location and resolution of these data, to enable accurate prediction of the future behavior of a complex nonlinear fluid system, such as the atmosphere or the ocean. The data can then be acquired at different resolutions in each region in accord with the predictions of the computer model. Because the density of data needed to accurately describe flows varies greatly over the flow domain, the interactive use of adaptive models with error estimation and control together with the automated observing systems will allow significant reduction in the amount of measured data required to achieve this goal.

A laboratory-testbed interactive system similar to the Earth Observing System that is being developed for use with atmospheric and oceanic models is under construction. It is impossible to study the intended applications directly at present because the observing systems and necessary computing power do not exist. Accordingly, the laboratory experiment will involve a rotating annular tank, with a sloping outer wall and filled with a stratified saline solution, which generates fluid motions very similar to those of geophysical fluid dynamics.

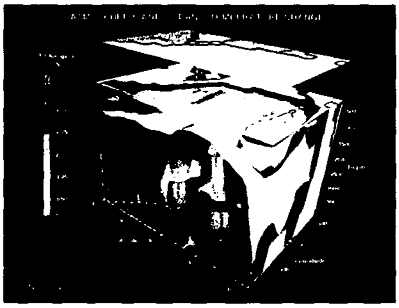
The application of the powerful adaptive composite grid method to this problem and its implementation in a parallel algorithm is essential to the successful completion of the project. The simulations will be carried out on parallel systems at NASA Ames Research Center, and the experimental data acquisition and computer systems will communicate via the Bay Area Regional Research Network.

http://www-cs.stanford.edu/hpcc.html

Global Climate Modeling

A coupled atmosphere-ocean model is used to simulate the entire atmosphere-ocean system for integration time scales ranging from months to tens or hundreds of years. The model being used at GFDL today consists of a spectral atmosphere general circulation model and the Modular Ocean Model (MOM) (described on page 47). Both models are being redesigned, and the spectral model is being modularized for consistency with the SKYHI grid-point model (described on page 46). These two atmospheric models will more easily share physics modules. The new model's dynamic core is designed to run efficiently on both shared memory and distributed memory systems. The model contains a more flexible vertical coordinate (for handling mountain terrain) and more efficient algorithms for solving the model's equations of motion. This work is illustrated on page 52.

http://www.hpcc.noaa.gov/Climate94



The colored plane floating above the block represents the simulated atmospheric temperature change at the earth's surface, assuming a steady one percent per year increase in atmospheric carbon dioxide to the time of doubled carbon dioxide. The surfaces in the ocean show the depths of the 1.0 and 0.2 degree (Celsius) temperature changes. The Southern Hemisphere shows much less surface warming than the Northern Hemisphere. This is caused primarily by the cooling effects of deep vertical mixing in the oceans south of 45 degrees South latitude. Coupled ocean-atmosphere climate models such as this one from NOAA/GFDL help improve scientific understanding of potential climate change.

Four-Dimensional Data Assimilation for Massive Earth System Data Analysis

Scientists are often faced with the task of sifting through voluminous data that come from observations or computational experiments. Virtual reality promises a means to search and manipulate such data rapidly by taking advantage of human vision and motor capabilities.

NASA is developing methods for analyzing this type of data generated in its Earth and Space Sciences (ESS) Grand Challenges. To do so it adapted the Flow Analysis Software Toolkit (FAST) developed at NASA Ames Research Center. Its virtual reality adaptation VR-FAST is being used by ESS researchers in atmospherics and in hydrospheric processes. Plans include incorporating additional data exploration capabilities and providing feedback from ESS researchers.

NASA is analyzing the applicability of the Vis5D software to virtual reality.

http://cesdis.gsfc.nasa.gov/hpccm/accomp/94accomp/ess94.accomps/ess2.html

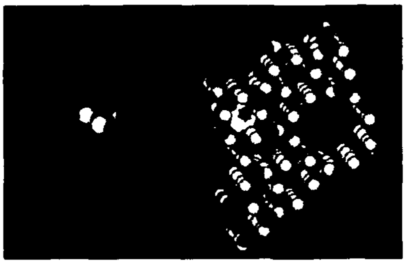
A scientist uses NASA's virtual reality modeling resources to explore the Earth's atmosphere as part of the Earth and Space Science Grand Challenge.

Ecosystem Simulations

Environmental Chemistry

DOE's Computational Chemistry Grand Challenge addresses the problem of modeling large molecular systems. One application area is complex problems in environmental chemistry such as the disposal of chlorofluorocarbons and the percolation of toxic chemicals through clay-based landfills and waste sites, which are of interest to both DOE and industry. Other application areas are product design and process optimization (for example the design of polymers and composite materials) and molecular biology (for example, drug design).

Quantum chemical self-consistent field (SCF) methods for studying molecular structure on vector computing systems were limited in scalability and by inefficiencies from extra computation. The new parallel SCF implementation overcomes these limitations. For example, to speed data communications among processors, a new programming strategy makes it easier to



The 38-atom carbonate system on the left illustrates the most advanced modeling capability at the beginning of the HPCC Program; the 389-atom zeolite system on the right was produced by a recent simulation. Computational complexity effectively grows as the cube of the number of atoms, implying a thousand fold increase in computational power between the two images.

write efficient software for computing systems with nonuniform memory access costs. This strategy is incorporated in the Fortran-callable Global Array communications library. The research has also enabled scientists to determine a scaling factor used to assign the optimal number of processors to the molecule under study.

This work is being conducted by researchers from DOE's Argonne and Pacific Northwest laboratories and pharmaceutical and chemical companies such as Allied Signal. They have run the new parallel software on the IBM SP-1, the Intel Delta and Paragon, the Kendall Square Research KSR-2, the Cray Research T3D, and workstation clusters.

http://www.mcs.anl.gov/Projects/chemacc94.html

Groundwater Transport and Remediation

More than half of the U.S. population depends on groundwater for its water supply. Groundwater is also an important source of irrigation and industrial process water. In many regions, available sources of groundwater are a fundamental constraint on development and economic activity. Groundwater supplies are increasingly threatened by organic, inorganic, and radioactive contaminants introduced into the environment by improper disposal or accidental release. Estimates of remediation costs of U.S. government sites alone range into the hundreds of billions of dollars. Protecting the quality of groundwater supplies is a problem of broad societal importance.

Remediation methods remain extremely (and potentially prohibitively) expensive and unpredictable in their success. The software developed under the DOE-sponsored Grand Challenges has been critical in developing effective remediation strategies. Grand Challenge numerical modeling of groundwater transport on massively parallel computing systems improves U.S. competitiveness by (1) directly applying groundwater technologies to groundwater problems, (2) applying these technologies to related industrial processes, and (3) applying generic massively parallel computational methods to industrial processes.

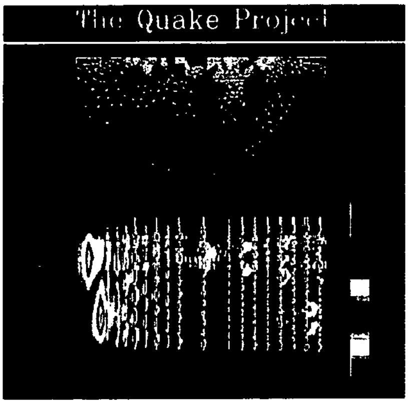
http://www.ccs.ornl.gov/GC/gw/gw.html

Earthquake Ground Motion Modeling in Large Basins: The Quake Project

A tool is being developed to simulate earthquake ground motion on parallel computing systems in order to determine how the intensity and duration of earthquake ground motion varies over a region. The knowledge gained will be used in designing earthquake-resistant structures for local conditions, leading to greater economy and safety. This work is illustrated on page 54.

This NSF-funded project is being conducted by engineers, computer scientists, and seismologists at Carnegie Mellon University, the University of Southern California, the Southern California Earthquake Center, and the National University of Mexico.

http://www.cs.cmu.edu:8001/afs/cs/project/quake/public/www/quake.html

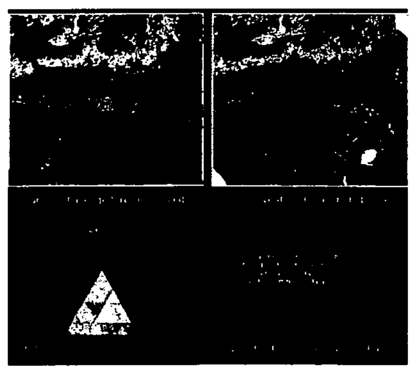


The upper image shows a computational model of a valley that has been automatically partitioned for solution on a parallel computing system, one processor to a color. The lower image shows the response of the valley as a function of frequency and position within the valley. It is well known that the response of a building to an earthquake is greatest when the frequency of the ground motion is close to the natural frequency of the building itself. These results show that damage can vary considerably depending on building location and frequency characteristics. Obtaining this kind of information for large basins such as the Greater Los Angeles Basin requires high performance computing.

High Performance Computing for Land Cover Dynamics

Understanding land cover dynamics is critical to the study of global climate change. Databases of land cover dynamics are needed for global carbon models, and biogeochemical, hydrological, and ecosystem response modeling. Over the span of several decades, changes in vegetation take place at a scale of less than 1 kilometer and require analysis of high resolution satellite images. Portable scalable software for a variety of image and map data processing applications is being developed and in the future will be integrated with new models for parallel I/O of large-scale images and maps.

An initial focus area is generating maps of the world's tropical rain forests over the last three decades. The image below illustrates the application of a new algorithm for solving the mixture modeling problem to a remotely-sensed image of part of Africa. Mixture modeling allows environmental scientists to estimate the proportions of different vegetation types present in a single pixel, thereby characterizing the vegetation more realistically than a classification that labels each pixel as a single vegetation type. Accurate descriptions of the land surface are important boundary conditions for climate and other global environmental models. Using the new algorithm, vegetation proportions are estimated by comparing observed reflectance measurements within a pixel to measurements expected if the pixel were purely of one type and solving for the proportions using mathematical optimization procedures. The algorithm, which is based on solving similar image restoration problems, is more accurate and faster than the labor-intensive classical methods.



This figure encodes the proportions of desert, grass, and forest within each pixel of a satellite image using color mixing. The Grand Challenge result, on the left, was produced using a new parallel algorithm and is a much more accurate estimate of mixture proportions than the least squares algorithm traditionally employed by environmental scientists.

This NSF-funded work is being conducted at the University of Maryland.

http://www.umiacs.umd.edu/labs/GC/blue_book.html

Massively Parallel Simulation of Large-Scale, High-Resolution Ecosystem Models

The objective of this research is to demonstrate a high performance computing system for large-scale high-resolution ecosystem simulation. To date a DEVS-GIS (Discrete Event System Specification interfaced to Geographic Information System) has been implemented in C++ to run on a single processor or in a multiprocessor environment using PVM. An object-based interface enabling models to transparently access commonly-used GIS databases has also been implemented. Watershed models have been developed and tested on a single processor DEVS-GIS system and are being ported to the Thinking Machines CM-5. The next step is extensive performance analysis. This will include the development of interconnect "middleware" and visualization software specifically for the DEVS-GIS environment. The long-term goal of this NSF-supported research at the University of Arizona is to enable experimentation with complex ecosystem models that is impossible without high performance computing.

http://www-ais.ece.arizona.edu

Biomedical Imaging and Biomechanics

Images are an important part of biomedical knowledge. New computational technologies provide the opportunity to supplement traditional two-dimensional biological and medical images with dynamic three-dimensional images that can be viewed, rotated, and reversibly dissected in a manner analogous to the physical objects they represent.

Visible Human Project

NIH/NLM is building and evaluating digital image libraries of human anatomy. Full use and understanding of the anatomical structures depicted in such libraries require the integration of HPCC technologies with technologies used in medical imaging systems including computed tomography (CT) and magnetic resonance imaging (MRI). Combining this library with efficient rendering algorithms will provide new educational tools for researchers, healthcare providers, students, and the general public. NIH/NLM is working with industry and academia to encourage the development of interoperable methods for representing and communicating such electronic images.

The Visible Human Data Set consists of images from a male and a female cadaver. The data set derived from the male cadaver was completed in November 1994 and made available both by Internet ftp (file transfer protocol) and DAT tape; NCAR assisted in distributing the data set. The size of the male data set is about 15 GB. More than 150 non-financial license agreements to access the data set have been signed by government, commercial, and academic organizations. Proposed applications include multimedia educational materials for educational levels from K through post-graduate medical education and home education, virtual reality programs includ-

Cryosectional image from the Visible Human Male.

ing surgical simulators, and modeling applications.

The data set from the female cadaver, which is expected to become available early in FY 1996, will be more detailed than the male.

In order to facilitate retrieval, methods need to be developed to link image data to symbolic text-based data that includes names, hierarchies, principles, and theories. Basic research is needed in the description and representation of structures and the connection of structural-anatomical to functional-physiological knowledge. The long-term project goal is to link the print library of functional-physiological knowledge with the image library of structural-anatomical knowledge transparently into one unified resource of health sciences information.

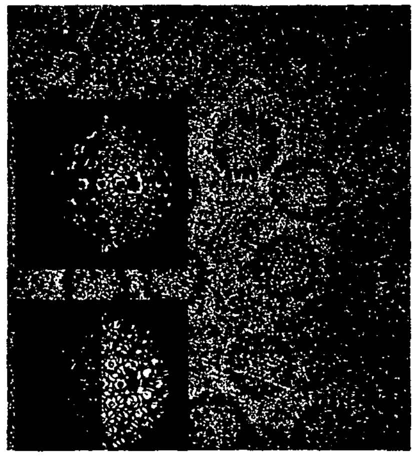
http://www.nlm.nih.gov/extramural_research.dir/visible_human.html

Reconstruction of Positron Emission Tomography (PET) Images

A PET image is formed through a computational reconstruction process. The computational time for the reconstruction and the quality of the resultant image depend primarily on the reconstruction algorithm that is used. Fourier methods have traditionally been used — they are fast but can lead to artifacts. The iterative expectation maximization method (EM) based on a maximum likelihood criterion is known to yield reconstructions that are as good as or better than Fourier methods and have lower patient dose, but their computational demands have limited their use. New generation PET scanners allow for the retraction of their lead shields, and current research is studying whether the additional information that is gathered improves the quality of the image, though at the price of increased computation. The EM algorithm has been implemented on the Intel iPSC/860 at NIH/DCRT for use with the GE Advance scanner, but can be used for other three-dimensional scanners. One iteration of the EM algorithm for reconstructing a brain-size image using data from retracted scanners on 32 processors in the Intel system took 55 minutes. Reconstructions of body-size images will require larger and faster computing systems, more memory, faster I/O to disk, and improved algorithms.

Image Processing of Electron Micrographs

Large icosahedral viruses can be reconstructed using high-resolution electron microscopy and three-dimensional image reconstruction. The input to the reconstruction is a set of two-dimen-



Three-dimensional reconstruction of large icosahedral viruses. Shown are images of herpes simplex virus type 1 capsids, which illustrate the potential of new parallel computing methods. They show the location of a minor capsid protein called VP26 as mapped in experiments in which VP26 was first extracted from purified capsids by treatment with guanidine hydrochloride and then rebound to the capsids. The right half of the top image shows the depleted capsid and the rebound VP26 capsid, and the left half shows the three-dimensional reconstruction, as it would be obtained with a conventional sequential computer. Parallel computing extended the analysis to obtain the lower images, which improved the signal-to-noise ratio and the resolution from approximately 3.5 to under 3.0 nanometers. The clusters of six VP26 subunits, shown together in the top image, are clearly resolved in the bottom image. This work was conducted at NIH in collaboration with the University of Virginia.

sional projections of virus particles obtained from electron micrographs. Estimating the two-dimensional orientation of these particles and using that information to perform the three-dimensional reconstruction are computationally intensive and memory-intensive calculations. Parallel algorithms have been developed to (1) estimate each particle's orientation by distributing the particles and associated computations across the computing systems' processors, and (2) reconstruct the complete three-dimensional image by distributing the micrograph data to balance the computations across the processors. The discovery of the location of individual proteins in the human herpes virus was due in part to the use of and improvements to these algorithms. This work was conducted at DCRT.

Automated Interactive Microscope (AIM)

A key feature of light microscopy is the ability to observe movement and activity in living cells. In this NSF-funded Grand Challenge at the Center for Light Microscope Imaging and Biotechnology at Carnegie Mellon University (an NSF Science and Technology Center), these microscopes will be coupled with high performance computing systems that are up to 1,000 times more powerful than they were several years ago. This environment will make possible:

- Image processing in near real time that was previously done off-line in hours or days — The research biologist can view images while an experiment is in process and while experimental intervention is still possible. Such real-time monitoring is important in studying phenomena such as the absorption process of a drug.

- Previously computationally infeasible image analysis and visualization techniques — An example is use of virtual reality for three-dimensional time-dependent display calculated from two-dimensional images — a specific illustration is embryonic development.

- Automated detection of "events of interest" followed by alerting the investigator, data gathering, or automatic intervention.

Light microscopy images of a keratocyte (an epidermal cell of a fish scale) obtained using Differential Interference Contrast to enhance visibility of nearly transparent structures both in raw form (left) and with boundary outlined by the "snake" method (right). Beginning with a curved line in the general vicinity of the boundary, the method then improves the fit by compromising between visible edges and preference for smooth boundaries without breaks or kinks not clearly dictated by the image. Edge finding is needed in automated determination of cell types, analysis and recording of cell motion, and searching for unusual events. One application area is automated techniques in medical diagnosis.

Accomplishments to date include acquiring a custom microscope, high speed camera, and local graphics workstation; installing high speed networking between Carnegie Mellon University and Pittsburgh Supercomputer Center, site of a Cray Research C90 and T3D and mass storage resources; obtaining/developing and integrating key software; and educating biologists and computer scientists about each other's techniques and problems. Next will come integration and use, including development of parallel and pipelined image processing software and user interface software to improve ease of use. Industrial partners collaborate in both developing and applying this technology.

http://www.stc.cmu.edu/aim/96bb/

A related "Collaboratory for Microscopic Digital Anatomy" National Challenge is described on page 92.

Understanding Human Joint Mechanisms

Joints in the human musculoskeletal system (for example, the knee) carry large forces when functioning normally, and potentially damaging loads in extreme and traumatic events. These forces pass through relatively thin layers of soft hydrated tissue, articular cartilage, which must function as frictionless, load-bearing surfaces. The thin layers slide over each other in what an engineer would term a sliding contact problem.

The goal of this Grand Challenge is to understand this sliding contact problem, enabling better understanding of how a joint carries load in normal and pathological situations and contributing to improved clinical treatment. High performance computing resources are needed to conduct the precise three-dimensional simulations of contact of two-phase (consisting of solid and fluid parts) tissues over complicated three-dimensional layer shapes with realistic loading conditions and material properties.

Underlying the overall goals of this project are research studies of importance and interest to the scientific computing community. They include (1) computational methods to solve the sliding contact problems of three-dimensional bodies, (2) methods to automatically control the error in a large scale numerical simulation, (3) constructing three-dimensional solid computer models of anatomical entities that have been digitally measured by methods such as CT scan, MRI, stereophotogrametry, or three-dimensional digitization, (4) parallel algorithms that can be used to solve nonlinear time-dependent problems (such as the deformation of soft tissues, in this project), and (5) using computer predictions with experimental data to help understand complicated nonlinear materials such as soft tissue.

This is a five year interdisciplinary effort involving bioengineering, computational mechanics, mechanical engineering, mathematics, and computer science supported by NSF at Rensselaer Polytechnic Institute and Columbia University.

http://www.scorec.rpi.edu/status2.html

Molecular Biology

Protein and Nucleic Acid Sequence Analysis

Advances in biotechnology have produced massive volumes of biological sequences and their associated biological and bibliographical information. For example, the current release (87.0) of GenBank from NLM's National Center for Biotechnology Information (NCBI) contains 269,478 sequences from over 11,000 different species and consisting of more than 248 million nucleic acid bases. Researchers who discover new sequences search the database for other sequences that are similar or otherwise relevant to their studies. The biological information associated with similar sequences may provide clues to the structure and function of the newly discovered sequences. Integrated structural modeling using all known sequences and structures in conjunction with powerful computational and visualization tools available this year at NCBI will further enhance this discovery process. NCBI's information systems containing data related to molecular biology, biochemistry, and genetics now handle 20,000 searches daily over the Internet.

http://www.ncbi.nlm.nih.gov/

To find similar sequences, the query sequence is compared to every other sequence in the database. The computationally intensive dynamic programming algorithm for comparing two sequences developed by Smith and Waterman has been commonly used. The execution time for this algorithm grows exponentially with the size of the database, making it impractical on conventional sequential computing systems. DCRT has developed a new parallel algorithm that greatly reduces the search time. Taking advantage of the fact that the comparison of the query sequence with each database sequence can be done independently and in parallel, this "static" algorithm sorts the sequences in the database by decreasing length, assigns each sequence in this sorted list to that "bucket" whose contents have the smallest total sequence length, and parcels out the buckets to the processors. Given a query sequence with 50 bases, execution time was 2.3 minutes on a 128-processor Intel iPSC/860 versus 3.3 hours with just one processor. This algorithm was more than twice as fast as a previously developed dynamic load balancing algorithm.

Protein Folding Prediction

The protein folding prediction problem is to determine the three-dimensional structure of a protein molecule given only its amino acid sequence. Understanding this three-dimensional structure is needed in studying the function of proteins in living systems and designing new ones for biological and medical purposes. Amino acid sequences of proteins are being discovered at an explosive rate, but experimental procedures for determining their three-dimensional structure (for example, x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy) are slow, costly, and complex. Theoretical and computational methods can help by predicting structure from sequence.

One reason why this problem is unsolved is that the biochemical rules governing the folding and the stability of proteins are not known. If they were known, an algorithm simulating the folding could be implemented. One alternative available today is an algorithm that finds the "best" conformation by searching all possible conformations; however because this space of all possible conformations is much too large to search, the search is restricted to a lattice. The restrictions in this lattice-space Monte Carlo method introduce distortions and consequently imperfect results.

DCRT is developing parallel algorithms so that more conformations can be searched and a more realistic energy function can be computed. These algorithms implement computationally intensive strategies for searching a large number of structures and calculating each one's free energy, particularly the hydrophobic potential that is proportional to the solvent-accessible surface area (ASA) of the protein molecule. The dihedral angle space Monte Carlo (DASMOC) algorithm combines knowledge-based rules and global minimization of free energy to search a large and complex energy space in order to find a protein structure in the most stable energy state. Current joint DCRT and NIH/NCI activities include implementing DASMOC on DCRT's Intel and developing parallel search algorithms that can work with DASMOC; DCRT is also developing parallel methods for ASA calculations.

Protein folding can generally be considered to occur in three phases or structures: a primary structure in which amino acids are linked together end-to-end according to its sequence; a secondary structure in which helices and beta sheets are formed; and a tertiary structure that is the final functional form of the protein. The secondary structure of proteins up to a certain size can be determined using NMR. NIH/NCRR has supported the development of a new parallel algorithm that uses the secondary structure obtained from NMR experiments and amino acid sequence information to calculate a lowest energy structure, which is the predicted tertiary (stable final form) structure. This algorithm has been tested, and the predicted results are accurate for a number of proteins whose structure is known.

Ribonucleic Acid (RNA) Structure Prediction

The ribonucleic acid (RNA) molecule serves several roles in nature. It is chemically very similar to DNA, but the differences are enough to distinguish its activities from the DNA molecule. The molecule is a key component for protein synthesis, serving as a template for protein generation (messenger RNA or mRNA) and being part of associated molecules for the synthesis of the proteins (ribosomal RNA or rRNA, and transfer RNA or tRNA). Many viral infections are also caused by RNA molecules, including the common cold, human immunodeficiency virus (HIV), and polio. The mechanisms for infection may vary. Some may be used like mRNA while others, like HIV, actually go through a reverse transcription process into the DNA. The viral packaging usually carries the virus' viability.

NCI has developed a new experimental "genetic algorithm" for RNA folding. This algorithm is highly parallelizable and rapidly convergent to solutions in a large conformational search space of RNA structures. It borrows from the processes of biological evolution using operations such as mutation, recombination and reproduction, and a selection criterion based on the idea of the survival of the fittest. The algorithm has been designed to run on the MasPar MP-2 massively parallel computer. It computes 16,384 RNA conformations at each generation. Using a random but structured information exchange at each generation allows it to iterate towards an optimal solution.

NCI also developed a parallel dynamic programming algorithm that generates suboptimal solutions for RNA folding. The program on the MasPar MP-2 is capable of rapidly folding RNA sequences that are over 9,000 nucleotides in length. HIV and the common cold virus are some examples of large sequences that have been folded with this algorithm.

http://www-lmmb.ncifcrf.gov

Molecular Design and Process Optimization

Biological Applications of Quantum Chemistry

Ab initio quantum mechanical methods are being developed for massively parallel computing systems. Unlike empirical force field (or molecular mechanics) or semi-empirical methods, *ab initio* methods are not parameterized and therefore can be used to describe previously unknown chemical substances with a good degree of accuracy. Because they are computationally intensive, to date these methods have been applied to small chemical systems that generally have fewer than 30 atoms. Systems of biological interest usually have more than 100 atoms and will require the gigaflops speed of massively parallel systems.

The Hartree-Fock Self-Consistent Field (SCF) approximation to the time-dependent Schroedinger equation has been implemented on a MIMD distributed memory parallel system. The computational bottleneck is in calculating a large number of integrals, saving the large data set of results on disk, and reading a subset of the data at every iteration in the approximation. The alternative direct SCF method that recalculates a smaller number of integrals at each iteration and saves the results in memory lends itself to parallelization. A new "mpqc" software and its underlying library written in C perform direct method calculations. One feature of the library is that matrices are distributed across the processors, eliminating the restriction on matrix size to the size of a processor's memory.

The mpqc software and extensions and enhancements to it are the building blocks for optimizing molecular geometries. Such geometries can be used in determining other molecular properties and the energy of chemical reactions. Optimization methods written in C++ have been implemented in mpqc. This software has been used to optimize geometries of molecules documented in the scientific literature in 10 iterations compared with other methods that took 90 iterations or in some cases were not converging after similar numbers of iterations.

This is joint work involving DCRT and Sandia National Laboratories. Future efforts involve converting mpqc libraries to C++ and applications in other areas of computational chemistry.

Biomolecular Design

Physical models and mathematical algorithms for numerical simulation of biomolecular systems are being developed at the University of Houston with NSF support. The research team combines expertise in chemistry, biophysics, biochemistry, chemical engineering, computer science, and mathematics.

The acetylcholinesterase dimer (AChE), partly shown in the figure at the right, is an enzyme responsible for degrading the neurotransmitter acetylcholine in species from man on down to insects. AChE is a target for many commonly used drugs and toxins; among the drugs that bind to AChE are therapeutic agents for Alzheimer's disease, myasthenia gravis, and glaucoma. Using numerical simulation, a second "back door" to the active site was recently discovered, creating the likelihood that substrates can come in one door and exit through the other. The Intel Touchstone Delta at Caltech and the Intel Paragon at SDSC will now be used for more

Electrostatic field, shown in yellow, of the acetylcholinesterase enzyme. The known active site is shown in blue; the second 'back door' to the active site is thought be at the spot where the field lines extend toward the top of the picture.

extensive simulations that should explain how this back door works: what makes it open and close, and what can pass through it and what can't.

The computations are being done using EulerGromos, the scalable parallel molecular dynamics software developed at the University of Houston. Using 256 processors on the Delta, simulations of the full solvated dimer involving 131,653 dynamical atoms take approximately 20 seconds per time step and require on the order of 50,000 time steps.

Recent work by the team and others on data parallel programming languages and techniques has resolved many of the core programming issues for multicomputers. The programming languages that have been developed require further refinement based on experience gained by implementing the sophisticated numerical algorithms the simulations require. Software engineering techniques have been developed to facilitate efficient porting of several codes to both distributed-memory and shared-memory processors. Furthermore, at least three different parallel programming paradigms are being used as appropriate for different applications, and in most cases automatic translation between them can be accomplished easily.

http://sina.tcamc.uh.edu/tcamc

Biomolecular Modeling and Structure Determination

In this NSF-funded Grand Challenge, the Theoretical Biology Group at the University of Illinois at Urbana-Champaign in collaboration with Duke University, Yale University, and New York University are developing tools for using high performance computing systems for research in structural biology. Their MDScope product includes (1) molecular visualization software for interactive display of molecular systems, (2) molecular dynamics software designed for performance, scalability, modularity, and portability, and (3) a protocol and library that functions as the unifying communication agent between the other two components.

MDScope allows scientists to explore the attributes of macromolecules in an immediate and visual way, and facilitates research into more complex systems than could not be readily understood using traditional methods. It will have uses in computer-aided drug design and protein structure refinement and prediction. Ongoing projects include specific membrane proteins, protein-DNA complexes, muscle proteins, and virus coats.

A portion of the Glucocorticoid Receptor bound to DNA; the receptor helps to regulate expression of the genetic code.

This work was conducted on Silicon Graphics and HP workstations.

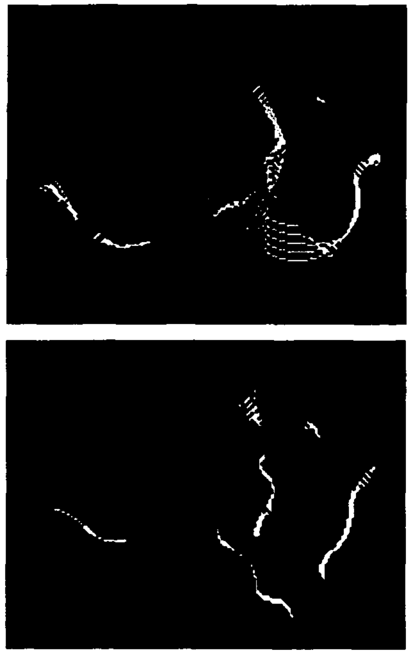
http://www.ks.uiuc.edu:1250/NSF_HPCC/

Computational Structural Biology

The so-called *de novo* protein structure prediction problem is important to both pharmaceutical drug designers and molecular biologists because three-dimensional protein structure essentially determines the protein's biological function. This DOE Grand Challenge has two goals: (1) the development and validation of algorithms and procedures for three-dimensional protein structure modeling and prediction, and (2) the implementation of a Protein Folding Workbench that allows biologists and biochemists to use large-scale parallel computing resources to explore various approaches to protein structure modeling and prediction. Biologically relevant models that can incorporate general experimental data and that provide interfaces oriented to the needs of the biology and biochemistry communities are emphasized. A hierarchical approach to the structure prediction problem moves from a "coarse grained" discrete lattice representation to a "fine grained" full atom representation.

This work was conducted by researchers from Caltech, the University of Washington, Argonne, and CRPC. Resources included an Intel Paragon, an IBM SP-1, 12 networked IBM workstations, and an ATM fiber optic network funded by Pacific Bell.

http://www.compbio.caltech.edu



The upper figure shows the known structure of the protein crambin from the Brookhaven Protein Data Base (PDB), and the lower figure is the best selection from a large ensemble of candidate chains, generated on a fcc (face-centered cubic) lattice using a guided replication Monte Carlo chain generation algorithm. Development of the algorithm and its serial and parallel implementations was funded by the HPCC Program. The three-dimensional structure prediction procedure was benchmarked at about 6 minutes on a 500-node Intel Paragon versus 24 hours on a single-processor IBM RS6000 workstation, a 225-fold speedup.

Computational Methods for Enzyme Catalysis

Modern genetic engineering provides methods for modifying biological molecules (biomolecules) such as proteins to have properties for applications outside of living systems — for example in chemical manufacturing, environmental remediation, and materials development. This project focuses on developing computational methods for relating the atomic structure of such proteins to their properties and function.

In biological systems, chemical reactions are usually catalyzed by protein molecules called enzymes. Accurate computational methods that use quantum-mechanical models have been developed for studying these enzymatic reactions. These models describe the making and breaking of chemical bonds in a reaction that is

Graphical representation of the bovine pancreatic ribonuclease enzyme. Many high-resolution X-ray structures are available for this enzyme, which makes it an ideal candidate for verifying new modeling methods.

controlled by the enzymatic catalyst. Because these enzymes typically contain thousands of atoms, fully quantum methods are intractable even on the largest computing systems, and new computational methods have been developed. The new methods are hybrids in which the active portion of the enzyme, which is involved directly in the chemical reaction, is modeled with quantum mechanics while the bulk of the molecule and the solvent are treated by computationally less demanding classical methods.

The core of the computation is the GAMESS (Generalized Atomic and Molecular Electronic Structure System) quantum chemistry software that has been adapted for parallel computing systems and allows complex enzyme active sites to be modeled for the first time. Special modifications of GAMESS allow the interfacing of quantum and classical models. This project is joint between NIST and Iowa State University, which maintains and freely distributes the software to hundreds of government, industrial, and academic research laboratories worldwide.

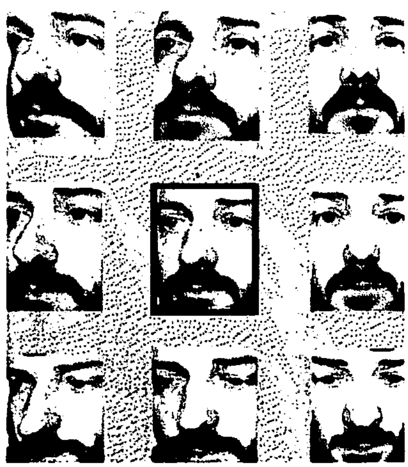
http://gams.cam.nist.gov/hpcc/

Cognition

High Performance Computing for Learning

The most effective approaches to problems in learning are computationally intensive, using optimization techniques in high-dimensional spaces with hundreds or thousands of parameters and simulation of neurobiological and developmental systems, and in addition require a grasp of realistic and cognitive constraints. Integrating machine perception and learning into the HPCC environment provides an opportunity to develop more robust and flexible practical systems for tasks such as visual inspection and can aid understanding of how the brain works.

Under some conditions, image examples can be used directly (without physically-based three-dimensional models) for image analysis and graphics synthesis using learning techniques. With this approach, many example images of an object under different imaging conditions (such as poses and expressions) are gathered, and an "analysis network" learns to associate to every image a vector of parameters such as pose and expression. A "synthesis network" learns the inverse task of producing the corresponding image given certain pose and expression parameters. When only one example image is available, new "virtual" images can be generated by learning from prototypical examples, as illustrated in the image below. This learning exploits a class of multidimensional interpolation networks



The central image is the original camera shot and the surrounding images were generated from the original using image synthesis/analysis.

that approximate the nonlinear mapping between vector input and vector output. In addition to face recognition, application areas include computer graphics, special effects, very low bandwidth teleconferencing, interactive multimedia, and object recognition systems.

Conducted at the Center for Biological and Computational Learning at MIT, this work is funded by NSF, other Federal organizations, and corporations.

http://www.ai.mit.edu/projects/cbcl/HPCCBlueBook/BlueBook.html

A New View of Cognition

Understanding how the brain directs natural behavior requires changing the environment on millisecond timescales at key points in the ongoing task. To study this problem, simulated worlds have been generated by special purpose high performance computing systems interfaced with head-mounted visual displays and visual and kinetic monitoring equipment.

This virtual reality approach funded by NIH has already led to unexpected results. Recent findings concern the brain's ability to rapidly switch its focus and absorb just the right amount of information for its current task. This process is called temporal fractionation of physical properties.

Observations of the common properties of physical objects show that they cohere; for example, the color and location of an object are unified in normal human perception when timescales are equal to or longer than a second. Behavioral experiments using innovative virtual reality equipment have found that the brain processes color and relative location at two different times during the course of solving the task. This result was entirely unanticipated, since the color and location of an object are unified in normal human perception. This new view of cognition revises theories of how the brain prepares information for problem solving.

Other virtual reality research includes hand-eye virtual environments and driving simulators. These projects are yielding interesting research results and have promising applications in addressing problems such as perception and distance deficiencies in elderly drivers. These studies will also serve as diagnostic aids for a wide variety of illnesses such as Huntington's and Parkinson's diseases that are characterized by short-term memory deficits.

Fundamental Computational Sciences

Quantum Chromodynamics

DOE's Oak Ridge and Los Alamos National Laboratories are using massively parallel computing systems for numerical simulations of quantum chromodynamics (QCD), the fundamental theory of the strong interactions of high energy physics. Such work can help unlock fundamental secrets of nature. For example, lattice QCD is a branch of elementary particle physics that seeks to further understanding of the properties of quarks and gluons (the most basic constituents of matter) by computational modeling of their interactions.

QCD software developed for several parallel systems (including the Cray Research T3D, IBM SP-2, Intel Paragon, Thinking Machines CM-5, and workstation clusters) provides stress tests for all key system features — integer and floating point operations, inter-node communications, and I/O. In one instance previously undetected problems were revealed, and large sections of code are now included in a vendor's diagnostic software.

http://physics.indiana.edu/~sg/qcd_doe.html

High Capacity Atomic-Level Simulations for the Design of Materials

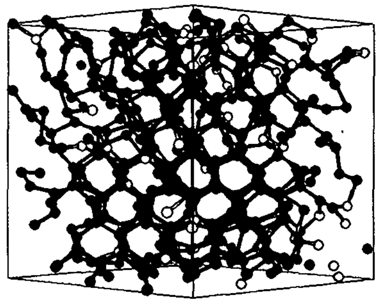
The objective of this NSF-supported Grand Challenge at Caltech is to develop theoretical methods for practical computations of the structures and properties of real materials for use in industrial process design for manufacturing new materials. There is an enormous gap between current methods for atomistic simulations and the level needed to accurately describe the relevant properties of industrially important materials. The strategy is to transcend from the most fundamental theory, quantum mechanics, where simulations are limited to small systems (10 to 20 atoms), through molecular dynamics where systems of a few thousand atoms are possible, to new techniques suitable for practical chemical engineering software directed at complex molecules.

A potentially significant industrial application for atomistic simulations is predicting the glass temperature T of a polymer: above T the polymer is soft and can be formed, while below T it is stiff. Industry would like to adjust T to desired values by manipulating the polymer. This is done empirically today, leading to costly and wasteful experiments. Excellent predictions of T were recently obtained for a variety of cases (including Teflon) in a collaboration involving Chevron, BF Goodrich, and Asahi Glass. The mechanism controlling T was also explained via the simulations, suggesting that in the future it may be possible to control T by making specific changes to the polymer.

http://www.wag.caltech.edu

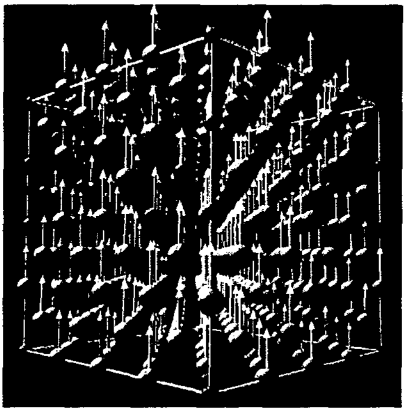
First Principles Simulation of Materials Properties

The work on the DOE-funded Grand Challenge in Materials Sciences is distributed among Oak Ridge National Laboratory, Brookhaven National Laboratory, and Ames Laboratory at Iowa State University. The effort is to use high performance computing resources to design materials. A key to designing material for structural, magnetic, optic, electrical, and high temperature applications is our understanding and ability to control synthesis and processing at the atomic level. As many of the crucial macroscopic properties of materials actually depend on defects, clusters, and microscopic structures involving hundreds to thousands of atoms, it is only with the availability of modern high performance computing systems that first principles modeling of these structures and related materials properties can be undertaken. To this end a hierarchy of increasingly accurate and computationally intensive techniques have been developed, tested, and are now being applied — classical molecular dynamics, tight binding molecular dynamics, and *ab initio* methods. Parallel computing combined with new algorithms that scale linearly with the number of atoms are being used to calculate efficiently the electronic structure and quantum-mechanical forces for systems of up to 500 atoms for *ab initio* methods and up to 10,000 atoms for tight binding molecular dynamics using an Intel Paragon XP/S 35.



The melting of diamond.

The melting and pressure-temperature phase diagram of carbon is an example in which conditions are too extreme for laboratory experiments, but in which accurate molecular dynamics simulations are leading to new insights for understanding natural and artificial diamond synthesis. The image above shows a snapshot of 512 carbon atoms in a diamond lattice in the process of melting (T > 4000K). The red atoms indicate four-fold bonded (diamond-like) atoms, the blue atoms indicate three-fold (graphitic) bonded atoms, and there are also a number of two-fold and five-fold coordinated atoms. The large number of three-fold atoms indicates that the liquid phase is less dense than the four-fold diamond phase. This is in contrast to silicon, in which the liquid phase has a higher average coordination than in the diamond structure. By running such simulations for the coexistence conditions of the solid and liquid phases, the melting temperatures of diamond as a function of pressure are determined.



The magnetic structure of a Ni-rich NiCu alloy.

Magnetic alloys are at the heart of a wide range of technological applications from the oldest of structural materials to the next generation of data storage and retrieval devices. However, a detailed microscopic understanding of alloy magnetism is lacking, hindering further development of these technologies. Using a new *ab initio* method, researchers have been able to study, for the first time, the nature of the magnetic state in disordered alloys. In the image above, Ni- (large blue spheres) and Cu- (small red spheres) atoms occupy the lattice sites of a 256-atom/unit cell model of a Ni-rich disordered NiCu alloy. The local Ni-site magnetic moment is distributed inhomogeneously, varying from a minimum of approximately 0.29 Bohr magnetons (short blue arrows) to a maximum of approximately 0.6 Bohr magnetons (long red arrows). Interestingly, the magnetic moment on a Ni-site correlates with the total magnetic moment on the nearest neighbor shell of atoms surrounding it: large red arrows tend to be surrounded by other reddish arrows, while small blue arrows are surrounded by either Cu sites having no moment or other blue arrows. The results of these simulations are being used to re-interpret results of neutron-scattering measurements of magnetic correlations in these alloys and to provide new insights into the properties of magnetic alloys.

http://www.ccs.ornl.gov/GC/materials/MShome.html

Black Hole Binaries: Coalescence and Gravitational Radiation

Black holes are formed by pressureless dust. The three-dimensional spiraling coalescence of two black holes is a problem of fundamental importance in astrophysics and general relativity. Such an event would produce a strong source of gravitational radiation that will be detectable by LIGO (Laser Interferometer Gravitational-wave Observatory) by the turn of the century. Solving this problem requires using advanced computational techniques to solve the Einstein field equations. The solution is computationally intensive and requires new methods for data management and visualization. Adaptive gridding and multigrid techniques applied to hyperbolic and parabolic systems of equations can be used in other computationally intensive problems in science and engineering. Tackling such problems in turn will stimulate new developments in architectures and algorithms for massively parallel and vector systems. This NSF-funded Grand Challenge is centered at the University of Texas and involves a large number of collaborators from other institutions.

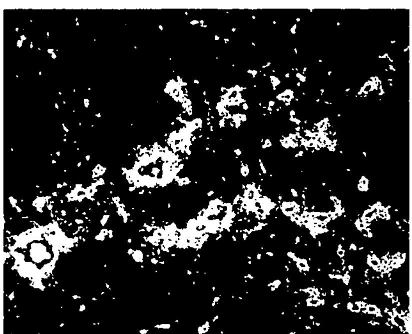
http://godel.ph.utexas.edu/Center/GC/page1.html

NASA has developed galaxy formation models that display different views of the same simulation.

Scalable Hierarchical Particle Algorithms for Galaxy Formation and Accretion Astrophysics

This NASA Grand Challenge has the goals of (1) understanding structure formation on distance scales from sub-galactic to cosmological, and (2) studying accretion problems such as stellar collisions and disruptions and accretion onto a black hole. For these studies scalable parallel particle software — N-body, smoothed-particle hydrodynamics (SPH), and hybrid — based on hierarchical tree data structures is being implemented on the IBM SP-1, Intel Paragon, and Thinking Machines CM-5. Load balancing is achieved by (1) using the self-similar (Morton ordered) curve that traverses the volume of the simulated domain, (2) assigning particles corresponding to different pieces of the curve to different processors, and (3) assigning adjacent particles to the same processor, thereby minimizing communication between processors. This approach can be implemented to be independent of the nature of interparticle forces, making it applicable to similarly formulated problems in physics, chemistry, molecular biology, and engineering.

http://cesdis.gsfc.nasa.gov/hpccm/accomp/94accomp/94accomps.html



Simulation of gravitational clustering of dark matter. This detail shows one sixth of the volume computed in a cosmological simulation involving 16 million highly clustered particles that required load balancing on a massively parallel computing system. Many particles are required to resolve the formation of individual galaxy halos seen here as red/white spots.

Radio Synthesis Imaging

This NSF-funded effort at the University of Illinois is implementing prototype next-generation astronomical telescope systems — remotely located telescopes connected by high speed networks to high performance computing systems and on-line databases accessed by astronomers over gigabit speed networks. The current prototype links the Berkeley-Illinois-Maryland Millimeter Array (BIMA) to NCSA for real-time data transfer into a database and for archiving on the NCSA mass storage system. Computationally intensive software for processing radio synthesis array data is being implemented on parallel systems. This work is applicable to other remotely controlled data-intensive facilities.

http://atlas.ncsa.uiuc.edu/hpcc-radioastro

Large Scale Structure and Galaxy Formation

This NSF-funded Grand Challenge is developing new parallel algorithms and software development strategies in order to use teraflops computing systems to answer two of the most fundamental questions in the physical sciences: What is the origin of large-scale structure in the universe, and how do galaxies form?

Cosmological simulations require computing the motion of millions of particles subject to their mutual gravitational attraction. To do this calculation efficiently on massively parallel computing systems is a challenge when the particles become highly clustered. It requires load balancing, that is, giving each processor an equal amount of work. A perfectly load-balanced algorithm that permits simulations with tens of millions of particles has been developed for the Thinking Machines CM-5.

The same algorithm has been applied to more earthly problems involving the motion of many mass particles — two examples are the spread-ing of oil droplets after a marine oil spill and modeling fuel injection sprays in internal combustion engines. A modified algorithm that incorporates contact forces is being applied to the flow of food or drug pellets in a processing machine.

This work is being conducted by seven institutions that comprise GC3, the Grand Challenge Cosmology Consortium that is anchored by NCSA and PSC.

http://zeus.ncsa.uiuc.edu:8080/BlueBook96/BB96.html

Grand-Challenge-Scale Applications

The HPCC Program funds the development of applications other than the Grand Challenges. This work goes on in particular at NSF and DOE facilities and at NASA. Examples of these activities are described below. Additional NSF highlights are described at:

http://pscinfo.psc.edu/MetaCenter/MetaScience/welcome.html

Simulation of Chorismate mutase

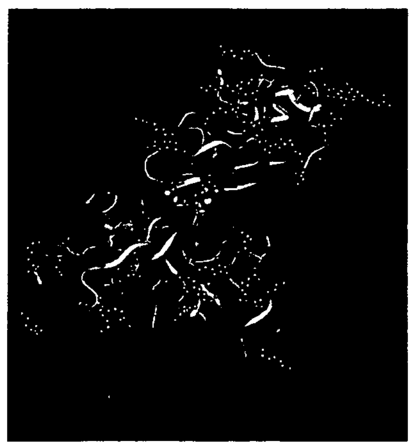
Chorismate mutase is an enzyme involved in the synthesis of amino acids. Bacteria use it to speed up chemical reactions by a factor of more than a million. Because it does not occur in the human body, inhibitors of this enzyme may be safe antibacterial agents. Using high performance computing at the Cornell Theory Center, researchers have discovered strong concentrated lines of electrostatic force leading to the "active site," that is, the spot where molecules are processed. This may provide the secret to the enzyme's potency and lead to improved drug design.

Simulation of chorismate mutase showing lines of electrostatic force.

A five-minute high-resolution animation containing images of the complexity shown in the accompanying figure required 40 hours of computing on each of 10 processors of an IBM SP-2 and produced approximately 16 GB of data.

Simulation of Antibody-Antigen Association

The first stage in molecular recognition in antibody-antigen association is the encounter between the antigen and antibody under the influence of mutual intermolecular forces. Researchers at NCSA have simulated the encounter between the antibody fragment of the monoclonal antibody HyHEL-5 and the antigen, hen-egg lysozyme, accounting accurately for the electrostatic steering and orientational forces.



The complex between the fragment of a monoclonal antibody, HyHEL-5, and hen-egg lysozyme. The key amino acid residues involved in complexation are displayed by large spheres. The negatively charged amino acids are in red and the positively charged ones in blue. The small spheres highlight other charged residues in the antibody fragment and hen-egg lysozyme.

The computational process involves solving two fundamental equations, the Poisson-Boltzmann equation for obtaining realistic electrostatic fields due to the antibody and the diffusional equation for obtaining the probability of encounter between the protein and antibody. The electrostatics calculation requires a large-memory computer, while a massively parallel computer is ideal for computing the large number of trajectories necessary in the diffusional simulations. The overall simulation involved using the Convex C3, the SGI Power Challenge, and the Thinking Machines CM-5 at NCSA. The antibody fragment has over 200 amino acids, while the hen-egg lysozyme is made up of 129 amino acids. Each electrostatics computation of the antibody molecule involved several hours on the Convex C3, while the brownian dynamics trajectories on each of the 512-processor partitions on the CM-5 involved an hour.

This simulation yielded for the first time experimentally realistic rates of encounters between antibodies and proteins. Extensions to mutant proteins and antibodies are of value in protein design.

A Realistic Ocean Model

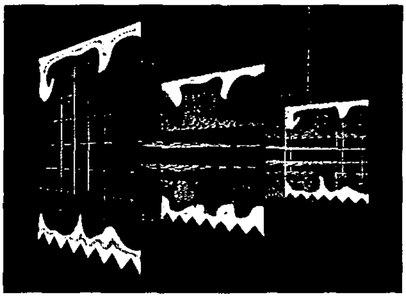
Satellite measurements show ocean levels rising one to three millimeters a year, which raises the question: What are the potential effects for coastal population centers? The uncertainties inherent in a system as complex as the Earth's climate means that the only tool available for assimilating the multitude of variables and trying to make rational predictions is computer modeling.

This year researchers at the Pittsburgh Supercomputing Center made a notable step toward a realistic model of ocean circulation. Exploiting the parallel-processing ability of the Cray Research T3D, a model of circulation in the North Atlantic correctly predicted the course of the Gulf Stream. No other circulation model of the entire North Atlantic has been able to do this. This very high resolution simulation ran for 10 days on 256 T3D processors. The results have proved the feasibility of a revised approach to ocean modeling. Prior approaches have been vexed by distortions in the interaction between the warm ocean surface and deeper, colder regions. The new model implements a set of equations that eliminate this spurious heat diffusion. This work is significant for ocean modeling and climate modeling in general, confirming that with sufficient computing capability it is possible to overcome obstacles that have bedeviled this area of research.

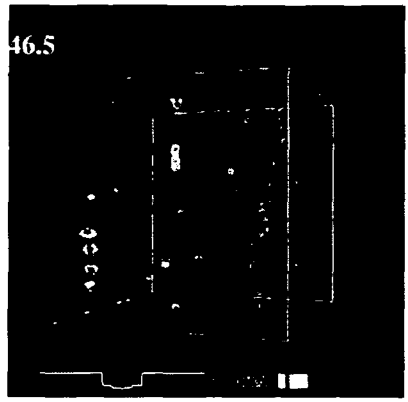
Simulation of circulation in the North Atlantic. Color shows temperature, red corresponding to high temperature. In most prior modeling, the Gulf Stream turns left past Cape Hatteras, clinging to the continental shoreline. In this simulation, however, the Gulf Stream veers off from Cape Hatteras on a northeast course into the open Atlantic, following essentially the correct course.

Drag Control

Paradoxically, riblets, or corrugated grooves, have been found to reduce drag when used on aircraft surfaces. Using computational fluid dynamics techniques and the Intel Paragon at the San Diego Supercomputer Center, a research group is looking for correlations between the vorticity field of the flow and riblet configurations, in order to develop a generic understanding of controlling wall turbulence. The appropriate use of riblets would allow for better control of drag, and hence more fuel-efficient aircraft. This work is illustrated on page 70.



Simulations on SDSC's Intel Paragon of turbulence over surfaces mounted with streamwise riblets. Computed turbulence intensities indicate that the reduction of fluctuations near the wall with riblets (bottom) results in a six percent drag reduction in this geometry.



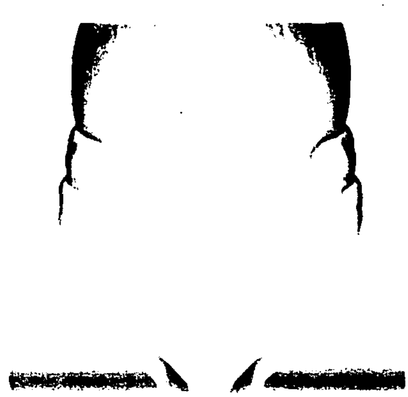
This image is a single frame from a volume visualization rendered from a computer model of turbulent fluid flow. The color masses indicate areas of vorticity that have stabilized within the volume after a specified period of time. The colors correspond to potential vorticity, with large positive values being blue, large negative values being red, and values near zero being transparent.

The Impact of Turbulence on Weather/Climate Prediction

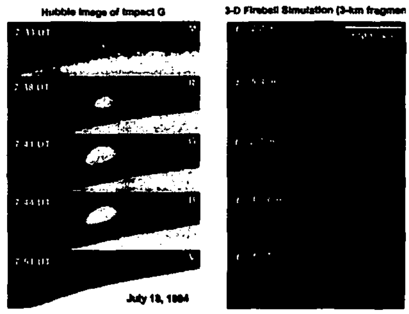
The primary goal of this National Center for Atmospheric Research project is to calculate fluid turbulence under the influences of environmental rotation and variable density at unprecedentedly high resolution by using the most powerful computers. The turbulence models being run under this project can be used to assist in the study of similar fluid motion phenomena in the atmosphere and oceans. This is because most of the volume of the Earth's atmosphere and oceans have a gravitationally stable density stratification, with lighter fluid above heavier. In combination with the planetary rotation, this causes the most energetic motions to occur mainly in horizontal planes. However, when buoyancy fluxes create a gravitationally unstable density stratification, vigorous vertical motions ensue. This convection is common in atmospheric clouds and also in polar regions of the ocean with strong surface cooling. The convection primarily occurs in plumes that carry parcels with anomalous density over long vertical distances within the unstable zone.

Shoemaker-Levy 9 Collision with Jupiter

Initial computer simulations had shown that the comet — known as Shoemaker-Levy 9 — would end quietly, being softly caught by the largest planet in the solar system. But NSF and DOE studies indicated that a large fragment would explode after penetrating about 60 kilometers into the planet's atmosphere creating a plume of superheated debris, shooting hundreds of kilometers above Jupiter's layered clouds. The predictions, which accurately foretold the event, were performed independently on the Pittsburgh Supercomputer Center's Cray Research C90 and Sandia's Intel Paragon. The simulations would not have been possible without the large memories now available on these powerful supercomputers. Observers used the Pittsburgh simulations to plan their monitoring of the crash and these simulations now appear to have provided an accurate description of the resulting explosions. In a recent keynote address, comet co-discoverer Eugene Shoemaker said Sandia's simu-



Impact of the comet fragment. Image height corresponds to 1,000 kilometers. Color represents temperature, ranging from tens of thousands of degrees Kelvin (red), several times the temperature of the sun, to hundreds of degrees Kelvin (blue).



Comparison of Hubble Space Telescope (HST) images of the fireball resulting from the Comet Shoemaker-Levy 9 fragment G impact on Jupiter (left) with a Sandia computational simulation (right). The impact-generated fireball and debris plume were imaged over the horizon of Jupiter by HST over the eighteen-minute sequence shown here. Because of the viewing geometry, the lower 400 km of the fireball are beyond and below the horizon of Jupiter, and the lower 1,500 km or so are in Jupiter's shadow from the sunlight. The fireball is incandescent at early times. At late times, the illumination is due only to scattered sunlight. Information about the size of fragment G can be determined only by comparing the simulations with observations. The Sandia simulations for a 3 km-diameter ice fragment are remarkably similar to the HST data but suggest fragment G was somewhat smaller.

lations had been responsible for alerting astronomers to look for fireball plumes. An animation of the Pittsburgh simulations can be viewed at:

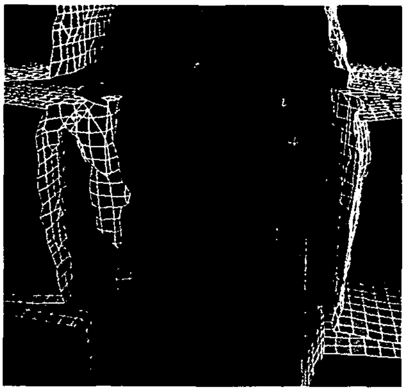
http://www.psc.edu/MetaCenter/MetaScience/Articles/MacLow/MacLow.html

http://www.cs.sandia.gov/HPCCIT/jupiter.html

Numerical Simulations Answer Questions about Vortex Structure and Dynamics in Superconductors

One of the main obstacles in the development of practical high-temperature superconducting materials is dissipation. Dissipation is caused by the motion of magnetic flux quanta called vortices. The entry and subsequent diffusion of vortices into a superconductor have been the focus of numerous experiments. However, the limited resolution of physical experiments can provide only a rough view of aggregates of vortices.

Numerical simulations on the IBM SP system at DOE's Argonne National Laboratory have provided a promising new approach for studying vortices. Numerical simulations exploit the extraordinary memory and speed of massively parallel computers, thereby attaining the extremely fine temporal and spatial resolution needed to model complex vortex behavior. For the first time, applied mathematicians and computer scientists have identified a superstructure in materials subject to strong current. The superstructure is characterized by regions of slowly varying vortex density separated by stationary fault lines. The simulations on the SP computer also revealed that vortex lattices with misoriented grains gradually ``heal" when subject to a weak current. These new discoveries are providing new insights about vortex behavior. More



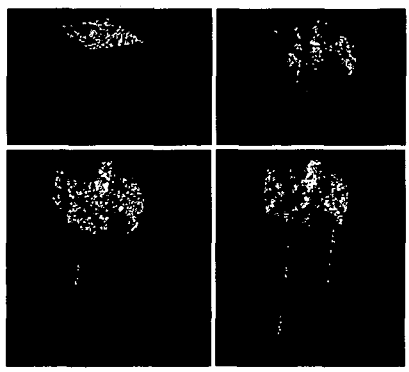
Early stages in the formation of a magnetic flux vortex. The figure shows the penetration of a magnetic field into a thin strip of high-T_c superconducting material, which is imbedded in a normal metal, and the formation of a magnetic flux vortex. The red surface is an isosurface for the magnetic induction. The isosurface follows the top and bottom of the superconducting strip (not shown). The field penetrates from the left and right sides. Thermal fluctuations cause ``droplets" of magnetic flux to be formed in the interior of the strip. As time progresses, these droplets may coalesce into vortices. One vortex is being spawned from the left sheet of the isosurface. These computations were done on Argonne's IBM SP system.

important, the numerical simulations may help researchers find ways to enhance the lattice defects, bringing superconductivity one step closer to practical application. Further details about on-going superconductivity simulation studies (including animations and other advanced visualizations) are at:

http://www.mcs.anl.gov/superconductivity.html

Molecular Dynamics Modeling

In collaboration with researchers in the Semiconductor Research Consortium (SRC) CRADA researchers are using the SPaSM molecular dynamics (MD) software to model ion-implantation processes. The simulations will be used to obtain more information on the major types of defects generated by ion implantation in silicon and their ability to migrate or diffuse in temperatures ranging from room temperature to approximately 400-500 Kelvin. Today industry is largely using table look-up methods in their process simulators. These methods attempt to include species, energy, dose, angle, rotation, and surface film (such as screen oxides) effects. These effects have been locally calibrated over the range of typical processing conditions. These tables are not easily extensible, limiting the overall range of this method and making it difficult to interpolate or extrapolate. MD simulations promise extensibility to new processing conditions such as multilayer implantation of dopants. They have the potential to generate initial damage profiles that are critical for transient enhanced diffusion (TED) calculations using point defect diffusion models, another factor for the industry. However, this MD approach has been calibrated over a much smaller processing window and is not generally used in industry. Second, such physically-based MD calculations will permit



MD simulation of a crystal block of 5 million silicon atoms as 11 silicon atoms implanted, each with an energy of 15keV. The simulation exhibits realistic phenomena such as amorphization near the surface and the channeling of some impacting atoms. These snapshots show the atoms displaced from their crystal positions (damaged areas) and the top layer (displayed in gray) at times 92 and 277 femtoseconds (10^{-15} seconds) after the first impact.

simulations of structures with arbitrary geometries and surface films. Third, these simulation methods promise simulation of new dopants (such as indium) and new targets (such as silicide) combinations without extensive calibration since the physics of the materials will be modeled directly.

The industry and its university collaborators (including the University of California at Berkeley, Stanford University, and the University of Texas at Austin) need this kind of information in order to develop more accurate, physically-based, and computationally efficient models for predicting implant-induced damage of the crystal and the following recrystalization with defects. Atomistic analysis and simulations are needed to support the development of more advanced phenomenological models in the simulator codes used extensively in the semiconductor industry.

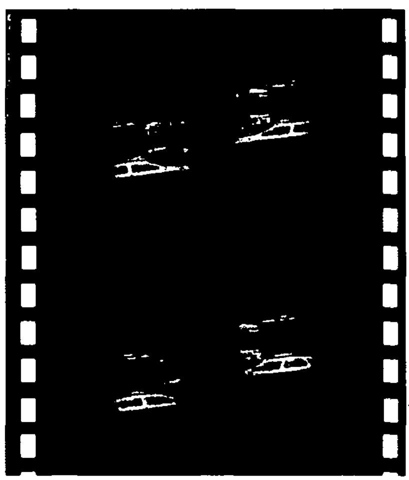
http://www.acl.lanl.gov/HPCC/data11-semicdtr.html

Advanced Computational Research in Crash Simulation

Automobile manufacturers must crash test their new cars to determine if they meet safety and crashworthiness standards for protecting passengers during accidents. The cost of a single car crash test can run from $50,000 to $750,000. The crash is usually performed at the end of the design process when the possibilities for design changes are limited. An alternative to actual testing is simulation that can provide structural information in early design stages. Design for crashworthiness is becoming more important as lighter-weight materials (aluminum alloys, magnesium, and polymer composites) are used in new cars to make them lighter and thus more energy efficient. In studies employing conventional material models, it is not unusual for an analysis to require a week of computer time on current supercomputers.

The Intel Paragon massively parallel systems at DOE's Oak Ridge National Laboratory, which combine hundreds of processors that can operate concurrently on a problem, can deliver the computational power required to perform analyses of complex crash situations in a relatively short time. Using data supplied by the U.S. Department of Transportation, researchers at Oak Ridge have performed computational analyses of a 4-door sedan crashing into a lamppost at 35 miles per hour, and of an offset head-on collision between two cars. In the course of computational analysis, the processors calculated the local deformation and the energy absorbed during a crash for each of 56,000 points, or finite elements. Simulation included detailed physical models such as 3,000 spot welds as well as 248 different structural materials. The analysis time has been reduced from 48 to 8 hours compared to current industry standards, and work is underway to further increase efficiency and incorporate detailed material models and design optimization.

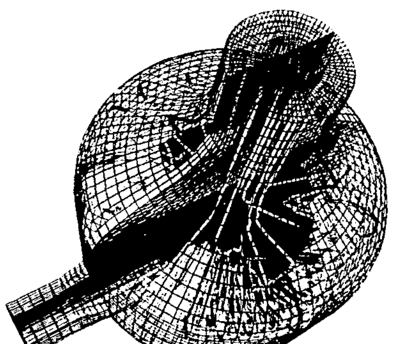
http://www.ccs.ornl.gov/ccs_info/CCS94.html



Illustrative of the computing power at the Center for Computational Science is the 50 percent offset crash of two Ford Taurus cars moving at 35 mph shown here. The Taurus model is detailed; the results are useful in understanding crash dynamics and their consequences. These results were obtained using parallel DYNA-3D software developed at Oak Ridge. Run times of less than one hour on the most powerful machine are expected.

Massively Parallel Numerical Methods for Advanced Simulation of Chemically Reacting Flows

SALSA is a new three-dimensional massively parallel (MP), chemically-reacting-flow software has been developed at DOE's Sandia National Laboratories. It allows simulations of complex three-dimensional fluid flows with complex reaction kinetics. SALSA has been used to simulate the deposition of a silicon carbide (SiC) mechanism with 19 chemical species undergoing over 40 chemical reactions. This chemical vapor deposition (CVD) process is of interest to a number of U.S. semiconductor companies. Using advanced MP algorithms, over 65 billion operations per second performance has been obtained in the solution phase of the simulation. This is 46 percent of the peak performance of 1,904 processors of Sandia's Intel Paragon and represents a significant increase in computational performance over state-of-the-art chemically-reacting-flow simulations. For this reason, SALSA was chosen as a finalist in the prestigious 1994 Gordon Bell Prize Competition for MP scientific applications. The development of enabling technologies, such as load balancing techniques and MP iterative solution methods, was funded by DOE's Office of Scientific Computing.



View of fluid streamlines and the center plane temperature distribution in a vertical disk, chemical vapor deposition reactor. Simulations such as these allow designers to produce higher uniformity semiconductor materials by eliminating unwanted detrimental effects such as fluid recirculation.

This complex chemically-reacting-flow simulation software is important to technology areas of interest to Federal agencies such as DOE and DOD (including ARPA), as well as U.S. industry. These areas include combustion research for transportation, atmospheric chemistry modeling for pollution studies, chemically reacting flow models for analysis and control of manufacturing processes, and CVD process modeling for production of advanced semiconductor materials.

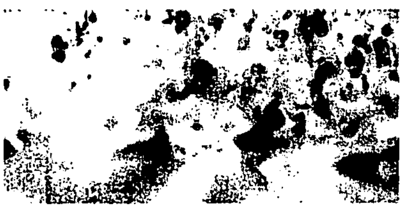
http://www.cs.sandia.gov/HPCCIT/crf.html

Convective turbulence and mixing in astrophysics

NASA has developed a new generation of portable production software for hydrodynamics and magnetohydrodynamics (MHD) for use in astrophysics that takes advantage of the large memory and fast speeds of modern parallel systems.

Managed by Goddard Space Flight Center (GSFC), the project is using tools provided by Argonne National Laboratory to develop portable codes for distributed memory environments; is working with Argonne computer scientists to refine these tools; is using University of Colorado performance tools to evaluate and improve parallel code efficiency, and to study and improve network communications (including ATM); and is comparing parallel computing strategies by investigating alternative parallelization tools based on shared memory constructs. The project has:

- Developed application software for MIMD architectures, including (1) software for three-dimensional compressible hydrodynamics, with nonlinear thermal conduction, based on a higher-order Godunov method, and a nonlinear multigrid elliptic solver, (2) a three-dimensional fully parallel fast Fourier transform for spectral or pseudospectral hydrodynamics and MHD software, (3) com-



NASA simulation of temperature fluctuations (dark: cool; light: hot) in a layer of convectively unstable gas (upper half) overlying a convectively stable layer (lower half) within the deep interior of a Sun-like star. This simulation was performed on the Argonne IBM SP-1.

pressible hybrid software, which performs spectral transforms on individual processors, and finite-difference software across processors (for convection in rotating systems), and (4) a collaboration with GSFC and JPL scientists to develop new hydrodynamic codes on massively parallel processors.

- Used the software for new calculations including (1) compressible penetrative convection in Sun-like stars, (2) core convection within A stars, (3) compressible turbulent convection constrained by rotation, (4) evolution of magnetic fields in turbulent conducting fluids, and (5) compressible turbulent magnetoconvection, and subgrid modeling for convection

- Studied the performance of parallel systems and communications, including (1) testing of ATM networks, and (2) testing and refining a library for scientific computing.

NASA has developed the first generation of fully portable application codes for turbulent mixing problems on massively parallel architectures, representing a broad range of techniques for solving hydrodynamic problems. This software achieves the high performance required to study new frontier problems in astrophysical fluid dynamics.

http://astro.uchicago.edu/rranch/Computing/

7. National Challenge Applications

Since FY 1994 one of the goals of the HPCC Program has been to address the National Challenges, fundamental applications with broad and direct impact on the Nation's competitiveness and the well-being of its citizens. These Challenges rely on the computing, communications, and information infrastructure technologies developed by the Program and in many cases also on development of Grand Challenge applications. Two of the National Challenge applications described below, Digital Libraries and Electronic Commerce, are fundamental enabling applications that will be used by many of the other National Challenges.

Digital Libraries

Wide use of library information requires both the information itself and convenient access methods. The past year has seen unprecedented growth in the amount of information available electronically and user access to it. This growth has resulted from several developments, some key ones of which were either funded by the HPCC Program or were begun by HPCC agencies prior to the creation of the Program and continued since then. They include:

- The Internet and its successful exponential growth (described on page 10)

- Client/server technology in which a user at a workstation (a client) could connect to a remote computer (a server)

- The World Wide Web (WWW), providing uniform mechanisms (using URLs, html, and http) for identifying and accessing information located on servers throughout the Internet (described on page 6)

- NCSA Mosaic, a browser that provides a single easy-to-use feature-rich flexible and extensible environment for accessing WWW files (described on page 7)

Joint Digital Libraries Research Initiative

In FY 1994 NSF, ARPA, and NASA jointly funded a four-year digital libraries research and technologies development activity. Awards were made to six consortia, each led by a university and including more than 10 other organizations such as libraries, museums, publishers, schools, and computing and communications industry companies.

The initiative's objective is to advance the means to collect, store, and organize information in digital forms, and make that information generally available for searching, retrieval, and processing locally and via communication networks — all in user-friendly ways. The projects address related research issues such as human-computer interaction, pricing and charging, privacy and security, collections management and archiving, librarianship, and user evaluation. The projects also contain extensive outreach and education components. The six projects are:

- "The Environment Library: A Prototype of a Scalable, Intelligent, Distributed Library" led by the University of California at Berkeley. This project will produce a prototype digital library focusing on the California environment. The information will include not only environmental impact statements and reports, but also photos, videos, computer models, county general plans, maps, and databases of environmental information. The multimedia information will be made available through a database-oriented client-server system to users engaged in environmental planning and studies at the state and local levels. Research thrusts include automated indexing, intelligent retrieval and search processes, data base technology, and data compression and communication tools.

  http://http.cs.berkeley.edu/~wilensky/proj-html/proj-html/

- The University of Michigan Digital Libraries Research Project is creating a large evolving multimedia testbed of Earth and space science data. Its contents will range from page images to interactive, compound documents, and eventually to real-time interaction with scientific data, replays of collaborative sessions, and human expertise. The research will focus on building a cooperating set of software agents — user interface, mediator, and collection interface agents. Potentially connecting thousands of users and information repositories, the system will be designed to systematize the vast amount of information on the Internet on an array of topics.

  http://sils.umich.edu/UMDL/HomePage.html

- The Alexandria Project is based at the University of California at Santa Barbara. It focuses on providing easy access to large and diverse collections of maps, images, and other pictorial material as well as a full range of new electronic library services. The initial objective is to turn a large cartographic collection — now totaling more than 4 million items — into electronically accessible forms, and to build a system that allows users to locate maps and other spatially-indexed materials in geographically dispersed digital libraries and databases. The testbed architecture is divided into user interface, catalog, storage, and ingest components. Content-based search will be developed for the catalog component.

  http://alexandria.sdc.ucsb.edu

- The objective of the Stanford Integrated Digital Library Project (SIDLP) is to develop the enabling technologies for a single integrated "virtual" library capable of providing uniform access to a large number of emerging networked information sources and collections. Targeted information resources range from personal information collections to collections that one finds today in conventional libraries and to the large data collections shared by scientists. The main objective is to build the "glue" that will make worldwide collections usable as a unified entity, by people with a wide range of expertise — all in a scalable and economically viable fashion.

  http://www-diglib.stanford.edu

- The University of Illinois project will include a customized version of NCSA Mosaic. Prototypes will be built to make library information readily accessible to Internet browsers by forming the "Interspace" in which information is linked coherently and displayed in appropriate formats. Initially focusing on engineering and science journals and magazines, the comprehensive collection will contain full articles, including texts, pictures, tables, equations, and references.

  http://www.grainger.uiuc.edu/dli

- The Informedia Project led by Carnegie Mellon University is developing an interactive digital video library system. Speech, image, and natural language understanding technologies are used to build the library corpus and to perform content-based search and retrieval of audio and video materials. The library testbed will initially contain 1,000 hours of video from public television and educational archives. Early users of the library will include primary and secondary school students.

  http://fuzine.mt.cs.cmu.edu/im/im-proposal.html

Digital Library Technology Projects

NASA funds R&D in digital library technology to support access to and distribution of remote sensing images and data, particularly scalable applications in both the public and private sectors. In addition to participating in the joint initiative described immediately above, NASA funded seven digital library technology projects in FY 1994 and early FY 1995. Awardees include public schools, museums, colleges and universities, and small and large businesses. The awards follow. The URL is:

http://sdcd.gsfc.nasa.gov/ISTO/DLT/dlt_projects.html

- Universal Spatial Data Access Consortium (Bell Communications Research) — Will develop tools for locating, accessing, browsing, transporting, and reusing NASA imagery and other geospatial data. Tools will integrate existing public domain software with new software for graphical user interface design, data screening, data compression, and distributed database technology.

  http://superbook.bellcore.com/USDAC/home_page.html

- Retrieval of Digital Images by Means of Content Search (IBM) — Will improve the speed of image-retrieval algorithms for content-based search. Current systems are limited by either the kind of search that they can perform, the size of database they can effectively support, or both. The testbed will test the technology on up to 100 GB of information.

  http://www.ibm.com/News/950222/nasa-00.html

- Reaching NASA from Home: Internet Access via Cable TV (Computer Science Corporation) — Will conduct a field trial showing how various pieces of the existing infrastructure can be integrated to provide Internet access to every household served by cable TV.

  http://sdcd.gsfc.nasa.gov/ISTO/DLT/abstracts.html

- Creating the Public Connection: Interactive Experiences with Real-Time Earth and Space Science Data (Rice University) — Will link the Houston Museum of Natural Science, with over 2 million visitors annually, to Rice University, a major hub on the Internet, so that NASA near-real-time data and imagery can be transmitted from Rice to the Museum and the public in the form of planetarium programs, computer-based interactive kiosks, and space and Earth science problem-solving simulations.

  http://space.rice.edu/hmns/connect.html

- Project Horizon (University of Illinois) — Will provide easy-to-use scalable digital library technologies for the public to locate, integrate, move, and analyze both Earth and space science data over the Internet. Activities include developing a "scientific data server" that integrates data in many common formats, enhancing client software such as Mosaic, and R&D on scalable servers and information systems.

http://sdcd.gsfc.nasa.gov/ISTO/DLT/abstracts.html

- Compression and Progressive Transmission of Digital Images (University of Wisconsin) — To provide a new compression scheme and transmission protocol along with servers and visualization software. A prototype has reduced bandwidth requirements by factors as large as 100.

  http://jerry.ssl.wisc.edu/~jwp/can.html

- SAIRE — A Scalable Agent-based Information Retrieval Engine (Loral AeroSys) — Will develop software to provide both expert and novice users access to metadata about Earth and space science datasets available from the U.S. Global Change Master Directory over the Internet. With the aid of intelligent software agents, this software will accept simple descriptions of a request in a domain-restricted natural language format, structured query, or menus, then correct errors or add missing information, learn the user's preferences, and shield the user from complex querying mechanisms.

  http://sdcd.gsfc.nasa.gov/ISTO/DLT/abstracts.html

NASA awards to support public use of Earth and space science data over the Internet are described beginning on page 79. In late FY 1995, NASA will solicit for a second round of digital library technology cooperative agreements and grants with industrial and academic partners. The awards are to be made in FY 1996.

http://rsd.gsfc.nasa.gov/rsd/CAN2.html

Satellite Weather Data Dissemination

NOAA's Satellite Active Archive is a digital library of real-time and historical satellite images from NOAA's Polar-orbiting Operational Environmental Satellites (POES). The two-TB (terabyte) library resides on an IBM robotic tape storage system that can be upgraded to handle 50 TB.

Users can search inventories of satellite images, preview representative Earth images, and download images for further processing and analysis. Applications include weather analysis and forecasting, climate research and prediction, global sea surface temperature measurements, atmospheric soundings of temperature and humidity, ocean dynamics research, volcanic eruption monitoring, forest fire detection, and global vegetation analysis.

http://www.saa.noaa.gov/

Environmental Decision Support

In FY 1996 EPA plans to offer research grants to stimulate development of technologies supporting intelligent interactive browsing, retrieval, and integrated analysis of large distributed collections of heterogeneous environmental data. Research targeted to address data storage, indexing, query specification and processing, filtering and summarization, and interoperability of information management systems will enable more integrated environmental decision support and risk assessment.

Computer Science Technical Reports Testbeds

Under ARPA funding, CNRI is developing methods for linking distributed heterogeneous electronic libraries; creating an experimental testbed of Computer Science Technical Reports and related information; exploring alternate network-based search, display, and retrieval techniques; and investigating techniques for electronic copyright management in a networked environment. This effort is being carried out in collaboration with five universities and the Library of Congress. In FY 1995 the project is demonstrating:

- Prototypes of distributed digital library technologies including techniques for scalable storage management and data repositories, persistent object bases, and multimedia objects

- A copyright management system providing proof of concept of fully electronic copyright registration, recordation, rights transfer, and management

http://www.cnri.reston.va.us/home/cstr.html

Unified Medical Language System (UMLS)

The disparity in biomedical terminology used to describe related concepts in different digital libraries prevents researchers and practitioners from retrieving and integrating relevant biomedical information from different sources such as biomedical literature, medical databases, expert knowledge systems, and clinical records. To address this situation, NIH/NLM has a collaborative program to create a Unified Medical Language System. This program, which now has 500 experimental users worldwide, will help health professionals retrieve electronic biomedical information from a variety of sources, regardless of how the concepts are expressed. The UMLS Metathesaurus is the core of the VA Clinical Lexicon currently being tested by more than 40 VA hospitals in the context of the Electronic Patient Record. The goal is to make it easy for users to link information from patient record systems, bibliographic databases, factual databases, and expert systems. NLM is using UMLS to provide a single point of information access. In FY 1996 it plans to have fully operational intelligent-agent-mediated multi-database searching available to all NLM users with Internet access.

http://www.nlm.nih.gov/factsheets.dir/UMLS.html

CALS Library

In FY 1995 and FY 1996 NIST plans to establish a demonstration library that incorporates robust search methods and provides access to information on CALS (Continuous Acquisition Lifecycle Support) to serve manufacturing interests.

Public Access to Government Information

The following are HPCC-funded digital libraries that reside on Internet/WWW servers accessible using browsers such as Mosaic.

Earth Data

In FY 1994 NASA established a Remote Sensing Public Access Center to provide access to information about numerous digital library technology and remote sensing database application projects. This Center is a combined effort of government, large and small companies, and universities.

http://www.rspac.ivv.nasa.gov/

At the end of FY 1994 and in early FY 1995 NASA funded 18 projects for remote sensing database application projects through grants and cooperative agreements. Awardees include numerous museums, schools and school districts, universities, and large and small companies. The awards are:

- Athena (Science Applications International Corporation) — Athena will develop curricular and resource material using geophysical and other data sets accessed over the Internet and prepare this material to form part of the science, math, and technology curricula for grades K-12 now under development.

  http://rsd.gsfc.nasa.gov/rsd/awards/Athena.html

- BADGER (Lockheed Missile and Space Corporation) — In collaboration with Smart Valley, Inc., a locally-based team led by Lockheed Palo Alto Research Laboratories will develop the Bay Area Digital GeoResource (BADGER). This non-profit community resource will be an on-line visualization system for geographic data of the San Francisco Bay Area developed from satellite and aerial images, maps, and databases. The system will be accessible over the Internet to municipalities, utilities, and other commercial, environmental, and non-profit interests. Environmental features, underground services, property lines, municipal zones, census data, and demographic information will all be depicted in BADGER's multi-layer visualization system.

http://www.svi.org/badger.html

- Dissemination of Atmospheric Sciences and Space Science Data and Information for K-12 and the Public (University of Washington) — A Pacific Northwest approach with special emphasis on products for use in science and mathematics instruction.

  http://rsd.gsfc.nasa.gov/rsd/awards/Puget.html

- Earth System Science Community Curriculum Testbed (ECOlogic Corporation) — The Curriculum facilitates interactions and communications among members of the Earth System Science Community by providing curricula designed to guide student investigations of the Earth system, and access to data, tips, techniques and guides, tools, and expert support.

  http://www.circles.org/ESSCC.html

- Emergency and Crisis Management (University of North Texas) — Develop Internet application demonstrating the usefulness of NASA's remote sensing data in mitigation, preparation, response, and recovery from natural and technological disasters

  http://pearl.ias.unt.edu/ias/projects.html

- Enhanced Access for Forest Management Planning (University of Minnesota) — Use LANDSAT images, digitized aerial photography, and ground-based forest databases to manage forest resources

  http://wwwrsl.forestry.umn.edu:10000/fornet/

- Enhancing the Teaching of Science in Elementary Education through the Application of NASA Remote Sensing Databases and Internet Technology (The Analytic Science Corporation) — Develop weather-based curriculum for grades K-6

  http://rsd.gsfc.nasa.gov/rsd/awards/FIFE.html

- Exploring the Environment (Wheeling Jesuit College, WV) — The NASA Classroom of the Future (COTF) program at the College is engaged in a three-year effort to develop Environment Earth Science modules. The modules will demonstrate educational application of NASA's remote sensing technologies and will be available over the Internet/WWW. While NASA and associated agencies hold terabytes of remote sensing data that can be used to support educational activities over the Internet, few teachers or students know of the resources or are able to access them. This project will enable access to these databases and other near-real-time information through learning modules.

  http://rsd.gsfc.nasa.gov/rsd/awards/COTF.html

- Flood Management Enhancement Using Remotely Sensed Data (SENTAR, Inc.) — The SENTAR Cooperative Agreement will use remotely-sensed data from the Distributed Active Archive Centers (DAACs) to enhance the capabilities of flood disaster management at the Alabama Emergency Management Agency (AEMA).

  http://iquest.com/~sentar/Flood_Management.html

- NASA Digital Image Data Distribution for Education, Public Access and Tourism in Hawaii (University of Hawaii and TerraSystems, Inc.) — Current data and images of the Hawaiian Islands over the Internet for use by the tourism industry and education, television, and researchers

  http://satftp.soest.hawaii.edu/space/hawaii/

- Passport to Knowledge: Electronic Field Trips to Scientific Frontiers Via Interactive TV and the Internet (The Childhood Project, Inc.) — Destinations include the Antarctic, the Kuiper Airborne Observatory (at NASA Ames Research Center), and the Hubble Space Telescope.

  http://quest.arc.nasa.gov/livefrom/passport.html

- Public Access to Earth and Space Science Data via Television (WRC-TV) — Develop visualizations of current Earth and space science data for inclusion in daily weather and news reports and for use in science classes with Internet access

http://rsd.gsfc.nasa.gov/rsd/awards/WRC-TV.html

- Satellite Data Driven Real-time Agricultural Management Decision Aids (University of Wisconsin) — Develop four end-user applications: (1) irrigation scheduling for on-farm use, (2) irrigation electrical-demand prediction system for power generation decisions by utility companies, (3) estimation of the duration of leaf wetness leading to foliar disease prediction in potatoes, and (4) prediction of frost damage for protection of cranberry crops.

  http://rsd.gsfc.nasa.gov/rsd/awards/Wisconsin.html

- A Science Infrastructure for Access to Earth and Space Science Data Through the Nation's Science Museums (University of California at Berkeley) — This project is creating a national Science Information Infrastructure (SII) linking the Nation's science museums to each other. The infrastructure is founded on an established framework of robust Internet connectivity spanning the country between public resource centers in science museums, which are linked closely with research centers. The information is made available to the community through the resource centers created in collaboration with local teachers. In turn teachers will be assisted in extracting information from the centers and applying it to developing science curricula.

  http://cea-ftp.cea.berkeley.edu/Education/SII/HomePage.html

- Surfing the Net: Aquatic Applications of Archival Satellite Imagery (Gulf of Maine Aquarium) — Investigate the land/sea interface, oceanographic applications, and the effects of human activities on the environment for grades K-12

  http://rsd.gsfc.nasa.gov/rsd/awards/Surfing.html

- Using Science and the Internet as Everyday Classroom Tools (Smithsonian Astrophysical Observatory) — One aim of this project is to make software available to elementary school students so they can use computers and the Internet to experience the essential quality of scientific research. To do so, research-quality astronomical software is being modified for the K-12 classroom. This includes the widely-used SAO image display program and astronomical analysis programs that are appropriate for elementary students.

  http://rsd.gsfc.nasa.gov/rsd/awards/SAO.html

- VolcanoWorld (University of North Dakota) — This project will establish interactive information centers with Internet-updated current information on volcanoes in the U.S., around the world, and on other planets. An education testbed in Grand Forks, ND, schools will be established for grades 4-8. Visitor information kiosks will be established at Mt. St. Helens and Hawaii Volcanoes National Parks.

  http://volcano.und.nodak.edu/

- Windows to the Universe (University of Michigan) — An Earth and Space Science Internet-based active-learning system for the general public with simulation-guided animation and voice overlays for museums and libraries.

  http://rsd.gsfc.nasa.gov/rsd/awards/Windows.html

NASA awards in support of digital library technology are described beginning on page 77. In late FY 1995 NASA will solicit for a second round of remote-sensing database application cooperative agreements and grants with industry and academic partners. The awards are to be made in FY 1996.

http://rsd.gsfc.nasa.gov/rsd/CAN2.html

NOAA is the steward of environmental data and information for the Nation. These data represent the environmental history of the country since its inception. Over this past year significant amounts of the data have become available on line, either through direct access or on-line orders. The URL of the NOAA Data Catalog containing more than 5,000 data sets is:

http://www.esdim.noaa.gov/NOAA-Catalog/NOAA-Catalog.html

The three NOAA National Data Centers manage and disseminate these data. Over the past year each has explored new ways of providing data using HPCC technologies. The National Climatic Data Center (NCDC) has developed and put on line the Global Climate Perspectives System that lets users manipulate climate data and see the results on line. The National Geophysical Data Center (NGDC) provides a broad range of data from the surface of the sun to the geophysics of the Earth and from recent DMSP satellite images back to paleoclimate data. The National Oceanographic Data Center (NODC) provides on-line access to the most recently acquired oceanographic data from sources around the world. Their URLs are:

http://www.ncdc.noaa.gov/

http://www.ngdc.noaa.gov/ngdc.html

http://www.nodc.noaa.gov/

In addition, NOAA has more than 40 locations disseminating on-line information to the Nation.

In FY 1996 NOAA plans to develop a data dissemination pilot that uses advanced data access tools on the Internet and makes select heterogeneous NOAA environmental data, distributed geographically at NOAA data centers, accessible to users in a more timely and complete way.

Education

The Department of Education (ED) has developed the Internet-based AskERIC (Educational Resources Information Center) that provides (1) a question and answer service; (2) a Virtual Library of teacher resources such as lesson plans, print and video materials (from CNN, the Discovery Channel, and PBS, for example), and research information; and (3) the National Parent Information Network, developed in conjunction with the National Urban League and others, that contains material to help parents support child development. FY 1995 and FY 1996 efforts address expanding the Virtual Library to include multimedia and hypermedia, completing the digital ERIC bibliographic database, and availability of full text of documents indexed in the ERIC bibliographic database in FY 1995. AskERIC's e-mail address is:

askeric@ericir.syr.edu.

In FY 1995 and FY 1996 ED's Office of Educational Research and Improvement (OERI) is building a core set of searchable databases containing current R&D activity, promising and exemplary programs, educational statistics, and departmental information accessible via INet, its Institutional Communications Network. OERI is exploring and testing enhancements to Mosaic useful to K-12 education and users with disabilities.

Health Care Data

NLM's computer-based Medical Literature Analysis and Retrieval System (MEDLARS) was established to achieve rapid access to the Library's vast store of biomedical and health-related information. Through world-wide communications networks, MEDLARS search services are available on line to individuals and institutions. It is searched more than eighteen thousand times a day.

In order to make searching easier and provide a user-friendly way to use MEDLARS, NLM developed Grateful Med, a PC- and Macintosh-compatible software package. The availability of Grateful Med, which can take advantage of the rapid communications afforded by the Internet, has resulted in an upsurge of usage by health professionals.

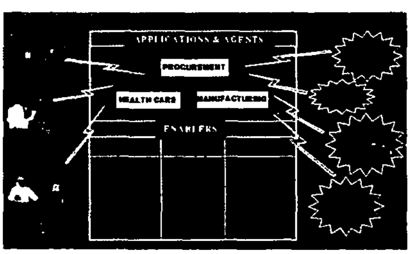
The National Center for Biotechnology Information (NCBI) is creating automated systems for storing and analyzing knowledge about molecular biology, biochemistry, and genetics. NCBI conducts research in advanced methods of computer-based information processing for analyzing the structure and function of biologically important molecules. It facilitates the use of and coordinates databases and software by biotechnology researchers and medical personnel worldwide through Internet access.

The National Cancer Institute (NCI) provides free CancerFax and CancerNet (Internet e-mail) information including PDQ (Physician Data Query) statements on cancer screening, prevention, treatment, and supportive care (in English and in Spanish); cancer fact sheets; and cancer literature. NCI is developing Internet-accessible multimedia digital libraries.

Electronic Commerce

NIST and ARPA collaborate in this National Challenge to make business transactions more efficient and cost effective. These transactions include electronic bidding, ordering and payments, and exchange of digital product specifications and design data. The National Performance Review recommended harnessing these technologies for the benefit of the Federal government and U.S. taxpayers.

In 1994 NIST established an Electronic Commerce Integration Facility that deploys and integrates technology developed by several agencies, including ARPA, NSA, and NSF. More than 24 vendors and suppliers participate via cooperative R&D agreements (CRADAs), equipment and software loans for demonstration purposes, and on-line access to electronic catalogs and value-added networks. The latter are commercial networks that add additional functions such as routing of messages and mailboxes where requests for quotations, requests for information, purchase orders, and other transactions can be exchanged among trading partners.



Electronic Commerce integrates applications software, data communications, data management, data interchange, and security functions to exchange business transactions and related data.

CommerceNet, a prototype electronic bid solicitation system that uses the WWW, has been developed and demonstrated. Funding came from a NIST CRADA with Enterprise Integration Technologies/CommerceNet and ARPA's TRP (Technology Reinvestment Program). On behalf of a purchaser's agent, an intelligent agent automatically solicits bids from supplier lists of desired products and, when the bid closes, presents the bids sorted by price to the purchaser. Complementary technologies (1) use smart card security technology and electronic mail to create digital signatures; (2) use either public key or secret key cryptography to encrypt transactions carried by e-mail; and (3) use security technologies to authenticate the identity of the party initiating the transaction (for example, the bidders) and the contents of the transaction, and to ensure confidentiality of a transaction (for example, a bid). Candidate future activities include integrating these technologies with Mosaic or e-mail into a single package.

http://www.commerce.net

To enable information retrieval and update from heterogeneous distributed databases, NIST integrated RDA (standardized remote database access, a communications protocol for establishing connections between database clients and servers) and SQL (Structured Query Language, a language to query and retrieve information from relational databases). This enables information retrieval and update from heterogeneous distributed databases. The Internet-based RDA/SQL prototype can access products from three vendors today, demonstrating how distributed information can be managed and accessed as a single virtual relational database, and how the information can be fused into more complex objects supporting Electronic Commerce and other applications.

http://waltz.ncsl.nist.gov/HPCC/commerce.html

In FY 1995 and FY 1996 NIST will collaborate with industry in developing a strategy for using the Standard for the Exchange of Product Model Data (STEP) and other data interchange standards, and will implement the strategy in NIST demonstration facilities. In FY 1996 NIST will provide an Internet-based interface for information access, query, selection, ordering, and disseminating its Standard Reference Data and Standard Reference Materials.

Through its funding for FAST, ARPA is addressing electronic commerce issues associated with automated parts acquisition. The need for flexible, low-overhead product development and support has been demonstrated in on-going FAST business operations in both DOD and industry. R&D issues include (1) handling high volume communications in electronic commerce, (2) improving the productivity of government buyers (and lowering government cost), and (3) interfaces with "foreign" code (facilitating interoperability with existing software and consequently more rapid dissemination).

http://info.broker.isi.edu/1/fast

In recognition of the synergy between digital libraries and electronic commerce, ARPA plans in FY 1996 to demonstrate prototype information services through a testbed incorporating the two subjects, including experimental charging mechanisms.

Civil Infrastructure

The Sustainable Management of Civil Infrastructure

Civil infrastructure systems such as roadways and bridges, water and gas distribution systems, and communications systems are large, complex, heterogeneous, distributed, dynamic systems that affect, and are affected by, technological, environmental, social, political, and economic factors. The life cycle of any such system includes design, construction, operation, maintenance, repair, rehabilitation, and disposal; each phase requires coordination among experts in diverse fields addressing the technical and social aspects of the system. In order to address the coordination issues, NSF has funded a National Challenge group at the University of Illinois. Research will focus on modeling the activity and communications within human teams in the context of the available media, technology, and organizational structures. Consideration will be given to how communication and work practices affect and are affected by technological possibilities. The model will be tested against the gas pipeline and road systems of Fort Gordon, GA.

Education and Lifelong Learning

The publicly available digital libraries developed by HPCC agencies are resources for education and lifelong learning. Those libraries and special efforts to use them for education and lifelong learning are described under the "Public Access to Government Information" National Challenge (page 79) and under Education and Training (page 97).

Energy Management

In order to enhance the existing capabilities of utilities for real-time energy supply and demand management, new technology for distributed systems needs to be developed in areas such as interoperability, authentication, privacy control, and multicast data aggregation. DOE plans to fund the development and implementation of both wide-area-based and distributed network tools, services and protocols that enable energy utilities to improve efficiency, conservation, billing and customer service, and that promote end-user interaction and control over their energy use.

Environmental Monitoring

NOAA's Pacific Marine Environmental Laboratory (PMEL) in Seattle, WA, focuses on environmental monitoring in both the coastal and open ocean to support predicting ocean processes on time scales of days to decades. The tropical Pacific Ocean is being monitored in order to understand and predict El Niño events. El Niño events are known to influence strongly the weather patterns over the U.S. as well as fisheries populations along *(continued on page 88)*

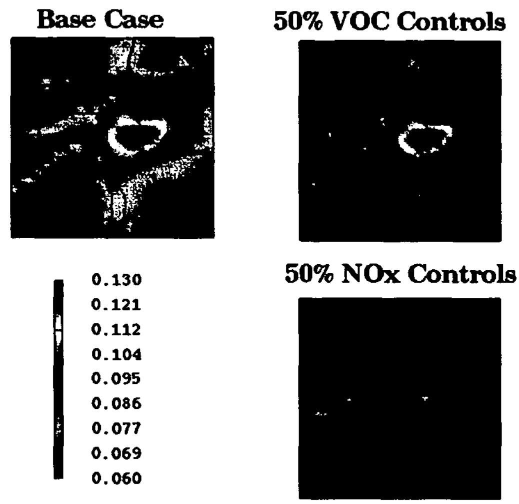
A Case Study Combining Grand Challenge and National Challenge Technologies

Air Quality Management: The North Carolina Compliance Modeling Project

Ground level ozone, or "smog," threatens animal and plant life. Occurring naturally in low concentration, ozone is harmless. However, air pollution can increase ozone concentrations to the point where it can severely irritate lungs, nasal passages, and eyes; reduce disease resistance; reduce crop yields; and injure trees. Crop losses alone due to ozone are estimated to be $3 to $5 billion annually in this country. The problem is so pervasive that newspapers and television stations in many areas daily report pollution level indicators in the summer. Periodically health officials issue warnings for children and the elderly to remain indoors if ozone concentrations are dangerously high.

Ozone is produced through complex chemical reactions primarily involving volatile organic compounds (VOCs) and oxides of nitrogen (NOx). These ozone precursors are produced by automobiles, factories, and even by natural processes such as lightning and vegetative growth. To ensure that ozone levels are within acceptable ranges for a given region, the VOCs and/or NOx emissions may need to be controlled. These controls are usually costly, and ill-suited controls may not even make much difference in the ozone levels that a region experiences. Since 1980, over $800 million has been spent in Atlanta on ozone reduction measures with only limited success. Ozone levels have not been reduced, but they have not increased either, despite significant population growth.

The Charlotte, NC, area is another region that has experienced unhealthy ozone levels. Previous ozone modeling efforts have indicated that VOC reductions lower ozone levels more than NOx reductions do. However, these results were obtained for regions that have more significant ozone problems than Charlotte has, regions that are generally more urbanized. Given that the character of the ozone problem in Charlotte may not match what is found in more urban areas, and considering the enormity of potential control costs, the North Carolina Division of Environmental Management (NCDEM) has employed HPCC technologies to determine an effective control strategy for the Charlotte region. The NCDEM used the Urban Airshed Model (UAM) to model the effects of numerous emissions controls. The figure shows that reducing VOCs in the Charlotte region has a minimal effect on modeled ozone, whereas NOx reductions significantly reduce the modeled ozone concentrations. This indicates that the ozone problem in the Charlotte vicinity differs from that in the more urban areas and should be addressed accordingly.



Ozone concentrations (ppm) simulated by the Urban Airshed Model for the Charlotte, NC, area on June 17, 1988, at 5 pm EST, using 1999 projected emissions data.

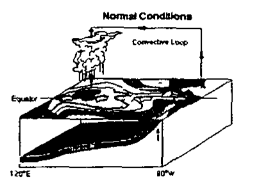
The importance of HPCC in modeling control strategies for ozone is not so much in determining whether global controls should be placed on VOCs or NOx, but rather in quantifying the effects that particular plants, factories, or highways have on the modeled ozone field. To extract this amount of detail, the UAM must be run with numerous control strategies for a variety of meteorological conditions. This process requires hundreds of UAM executions, each execution requiring significant computer resources. Outside of an HPCC environment, these model runs simply could not be made in a timely manner. Important decisions affecting the economics and health of an area would be made with a skeletal amount of modeling information, possibly resulting in an ill-conceived control strategy. Such a plan might place an inordinate economic burden on some industrial segments, while not significantly improving the air quality of the region. More serious polluters, on the other hand, might not be targeted for meaningful controls.

The principal collaborators in this project include EPA, the North Carolina Supercomputer Center, the North Carolina Division of Environmental Management, the University of North Carolina at Chapel Hill, Duke University, Carolina Power and Light, and Systems Applications International. This collaboration ensured that the economic concerns of potential control targets would be addressed, resulting in a control package that is not only environmentally friendly but also economically feasible.

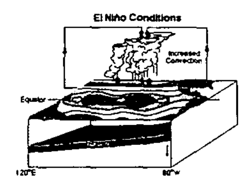
http://www.iceis.mcnc.org/publications/bluebook/

A Case Study Combining Grand Challenge and National Challenge Technologies: *What is an El Niño?*

El Niño is a disruption of the ocean-atmosphere system in the tropical Pacific having important consequences for weather around the globe. Among these consequences are increased rainfall across the southern tier of the U.S. and in Peru, which has caused destructive flooding, and drought in the west Pacific, sometimes associated with devastating brush fires in Australia. Observations of conditions in the tropical Pacific are considered essential for predicting short-term (a few months to one year) climate variations. To provide necessary data, NOAA operates a network of buoys that measure temperature, currents, and winds in the equatorial band. These buoys daily transmit data that are available in real time to researchers and forecasters around the world.

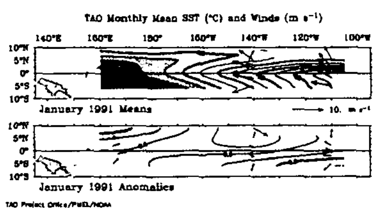


In normal, non-El Niño conditions (top panel of schematic diagram), the trade winds blow towards the west across the tropical Pacific. These winds pile up warm surface water in the west Pacific, making sea surface temperature about 8 degrees Celsius higher than the east off South America. This cold water is nutrient rich, supporting high levels of primary productivity, diverse marine ecosystems, and major fisheries.

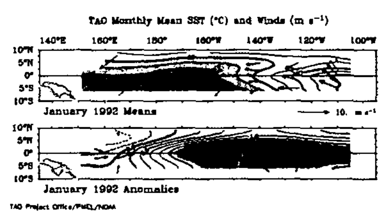


Schematic diagram of normal and El Niño conditions in the Pacific Ocean

During El Niño (bottom panel of the schematic diagram), the trade winds relax in the central and western Pacific leading to a deepening of the layer of warm surface water in the eastern Pacific, and an elevation of this layer in the west. This leads to a reduction in the upwelling and cuts off the supply of nutrient rich deep water to the upper ocean layers. The result is a rise in sea surface temperature and a drastic decline in primary productivity, the latter of which adversely affects higher trophic levels of the food chain, including commercial fisheries in this region. Rainfall follows the warm water eastward, with associated flooding in Peru and drought in Indonesia and Australia. The eastward displacement of the atmospheric heat source overlaying the warmest water results in large changes in the global atmospheric circulation, which in turn force weather changes in regions far removed from the tropical Pacific.



A normal year -- January 1991



An El Niño year -- January 1992

El Niño can be seen in Sea Surface Temperature in the Equatorial Pacific Ocean

El Niño can be seen in measurements of the sea surface temperature, such as those shown above, which were made from the Tropical Atmosphere-Ocean Array of moored buoys. In January 1991, the sea surface temperatures and the winds were near normal, with warm water in the Western Pacific Ocean (in red on the top panel of January 1991 plot), and cool water, called the "cold tongue" in the Eastern Pacific Ocean (in green on the top panel of the January 1991 plot). The winds in the Western Pacific are very weak (see the arrows pointing in the direction the wind is blowing towards), and the winds in the Eastern Pacific are blowing towards the west (towards Indonesia). The bottom panel of the January 1991 plot shows anomalies, the way the sea surface temperature and wind differs from a normal January. In this plot, the anomalies are very small (yellow/green), indicating a normal January.

January 1992 was the peak of an El Niño year. In January 1992, the warm water (red in the top panel of the January 1992 plot) has spread from the western Pacific Ocean towards the east (in the direction of South America), the "cold tongue" (green color in the top panel of the January 1992 plot) has weakened, and the winds in the western Pacific, usually weak, are blowing strongly towards the east, pushing the warm water eastward. The anomalies show clearly that the water in the center of the Pacific Ocean is much warmer (red) than in a normal January.

An animation of the 1991 El Niño is on the PMEL server at *http://www.pmel.noaa.gov/toga-tao/el-nino-story.html*

A Case Study Combining Grand Challenge and National Challenge Technologies:

Modeling San Diego Bay to Aid Resource Management

Analytical methods and software models are being developed in order to increase understanding of San Diego Bay. One application is to study environmental risks such as contaminant spills, and the dynamics of animal communities. The tools that are being developed will provide access to the best and most comprehensible information available to aid public policy makers in managing the use of the bay. The methods and procedures developed will be able to be generalized to other geographic areas in the U.S. and around the world.

San Diego Bay is one of the great natural harbors of the world. First sighted by Juan Rodriguez Cabrillo in 1542, it is 22 square miles in size, with its deep-water harbor sheltered by the peninsulas of Point Loma and Coronado. Five municipalities (San Diego, Chula Vista, National City, Imperial Beach, and Coronado) share the coastline of the bay, and many more depend on it for domestic and international commerce and diverse recreational activities. In addition, the Nation depends on San Diego Bay to support the Pacific fleet for national defense.

Regulatory oversight of the bay is assigned to some 26 organizations and municipalities involved in using and managing it. They include Federal, state, and local organizations: EPA, U.S. Fish and Wildlife Service, California Department of Fish and Game, and the County Health Department. Legislation has established the San Diego Bay Interagency Water Quality Panel (Bay Panel) to coordinate the environmental monitoring activities of these groups; the Panel funds this project. Bay Panel participants also include NOAA, the U.S. Navy, the City of San Diego, the Port of San Diego, the Audubon Society, and the Environmental Health Coalition.

The central focus of this effort is to develop a visual three-dimensional model of the bay based on physical, biological, and chemical data contributed by 35 data collection programs. Such visualization will provide the Panel with a complete picture of what is and is not known about the complex conditions of the bay and will help monitor the success or failure of policy decisions over time.

The first year's activities focus on acquiring and reviewing the data, establishing quality control to ensure that units of measurement are consistent and that anomalies are revealed and evaluated, and centralizing the data on a server at SDSC that is accessible over the Internet. The resulting database will be used to develop thematic data layers for presentation in the visual model, find correlations among interdisciplinary data, and support the examination of gaps in the spatio-temporal record and redundancies in monitoring programs, which waste labor and financial resources. Then work will turn to developing the visual model and doing preliminary hydrodynamic modeling. The visualization will include animated sequences over time and space to illustrate variations in population densities, water quality, and sediment characteristics. The second and third years call for development of analytical and predictive models. Model results and animations accessible via three-dimensional browsers will be available on the server.

The underlying bathymetry of San Diego Bay.

http://www.sdsc.edu/SDSC/Research/Comp_Bio/sdbay/sdbay.html

(from page 84) the Pacific coast of North and South America. On-going international research efforts are targeted at understanding the mechanisms by which El Niño events impact weather and fish catch. It is anticipated that this understanding will lead to a monitoring program that will allow forecasters to predict unusually wet or dry weather for large regions in the country months in advance and to anticipate economic impacts of reduced yields of Pacific fisheries. PMEL provides Internet access to near-real-time graphical depictions of tropical Pacific Ocean temperature measurements. These observations are telemetered from buoys via satellite link. An El Niño case study is described on page 86.

http://www.pmel.noaa.gov/toga-tao/el-nino-story.html

Health Care

Building upon the Grand Challenges in biomedical imaging and biomechanics and molecular biology, this National Challenge can improve the quality, effectiveness, and efficiency of the U.S. health care system. The HPCC agencies working in this area include the Agency for Health Care Policy and Research (AHCPR); the Department of Veterans Affairs (VA); NIH's NLM, NCRR, and DCRT; ARPA; and NSF. AHCPR and the VA are new to the HPCC Program effective with the FY 1996 budget cycle, and their HPCC efforts are in this National Challenge and in supporting network connectivity:

- AHCPR seeks to use HPCC technologies to improve the quality, appropriateness, effectiveness, and accessibility of health care

- With its 171 VA Medical Centers, 362 outpatient clinics, 129 nursing homes, and 35 domicilliaries (ambulatory nursing home centers), the VA operates the largest centrally directed health care system in the U.S. Focus areas include the use of HPCC technologies for more effective administrative operation and improved patient care.

Activities at several individual agencies are first described, followed by multi-agency activities in computer-based patient records and several projects.

NLM and AHCPR fund three-year grants, initiated in FY 1994, designed to help physicians practice better medicine by using advanced computing and networking capabilities. Testbed networks are being developed to share information sources including computerized patient records and their associated medical images, to allow access to clinical research protocols and on-line reference sources, to provide for remote patient monitoring, teleconferencing, and health professions education. These networks will operate between major medical centers and underserved rural areas, cross-country between medical centers, and within an urban environment between the medical center and underserved urban communities. An evaluation component is also planned to assess the cost and patient outcome changes that result from using such technologies.

NLM grantees will explore the application of virtual reality to medicine by creating and evaluating advanced computer simulations of human anatomic structure that support surgical planning and health professions education. Supported telemedicine projects include those to share neuroradiology images for consultation and patient monitoring (CA), to diagnose skin lesions via teleconsultation (OR), to create a statewide regional telemedicine system for a rural section of the country (WV), and to establish a National Laboratory for the study of Rural Telemedicine (IA). Methods for automatically linking the textual component of a computerized patient record with diagnostic images including pathology, radiology, and ultrasound will be developed. NLM and AHCPR will fund additional cooperative agreements in FY 1995 and FY 1996.

Beginning in FY 1993 NLM also funded a consortium of nine West Virginia institutions to use advanced networking to deliver health care to both rural and urban areas. In FY 1995 the VA Medical Center in Clarksburg, WV, joined this Mountaineer Doctor TV Network.

NCRR is using virtual reality for applications that use scientific instruments for molecular biology research and support for select surgical and other therapeutic interventions such as radiation planning. The researcher needs to visualize large amounts of data in a "natural environment." In clinical care the ability to "see" deep inside the body, especially the brain, can determine whether surgery or other intervention can be performed successfully with minimal patient trauma. HPCC technologies are needed to provide realistic real-time virtual reality visualization. NCRR efforts include its atomic force microscope, room-filling-molecule technology to permit better wayfinding, force-arm technology for molecular docking and folding studies, and brain behavior research.

ARPA is developing tools for use in decision making by health care providers and patients/consumers in ambulatory care and in combat casualty and trauma care. Key components include (1) a human/computer interaction manager, (2) a task and context manager, and (3) an information broker. The architecture stresses multi-modal user interfaces; integration of information from heterogeneous sources; multimodal information acquisition, communication, and presentation; intelligent on-line support for medical personnel; logistical planning and resource allocation; and simulation-based systems for decision makers in health care management.

Computer-Based Patient Records (CBPR)

The goals of this activity are to improve the accuracy, uniformity, and retrievability of patient care data by the health care community and to use these data to improve clinical decision making. This requires the development of clinical data standards and the deployment of distributed integrated computing systems throughout health care institutions and to health care providers. Thorough testing is required as these systems, decision support algorithms, and knowledge servers are brought into hospitals, the offices of doctors and other health care providers, and patients' homes. These efforts will bring the benefits of HPCC to health care providers and consumers throughout the U.S.

In FY 1995 AHCPR and NLM plan to fund R&D to incorporate guidelines, decision aids, expert systems, and reminder systems in CBPR systems.

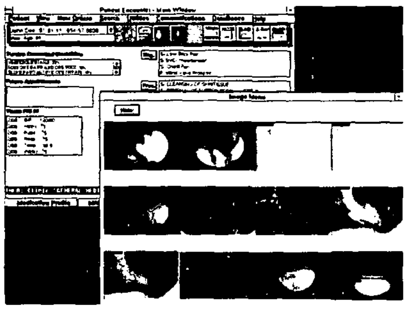
In FY 1995 AHCPR supports studies in the use of CBPR for patient care, outcome analysis, payment, quality assessment, research, and data transmission among institutions and providers. In FY 1996 AHCPR plans to support:

- Private sector development and testing of standards for (1) definition and coding of medical terms, (2) content of specific data sets for decision making, and (3) electronic data exchange

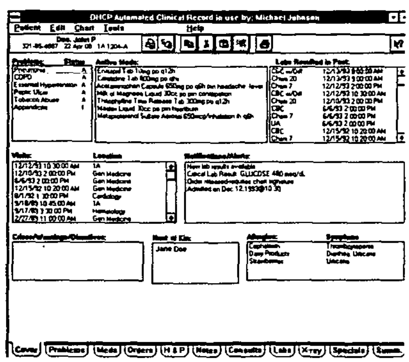
- Evaluation of (1) how CBPR applications are received by health care providers and patients, (2) how physician and patient behavior changes, and (3) how patient outcomes, productivity, and cost of care are affected

The VA Decentralized Hospital Computer System is an integrated automated system that includes more than 60 clinical and management modules. Its information sources include facility-based integrated patient-based information and databases on compact disks. Patient medical images are integrated with CBPR on clinical workstations at some VA medical centers. Plans include increased access to other networked resources such as NLM's Medline and other Internet-based libraries. Efforts focus on data capture, order entry of procedures and services (for example, laboratory tests and diet), standards, security, and a blend of in-house development, federal sharing initiatives, and private industry systems. The agency is developing clinical software for workstations and networking distributed clinical workstations to extend and expand access to CBPR. Another VA project addresses sharing medical information with the DOD; issues include exchanging data and interoperability. VA work is illustrated on page 90.

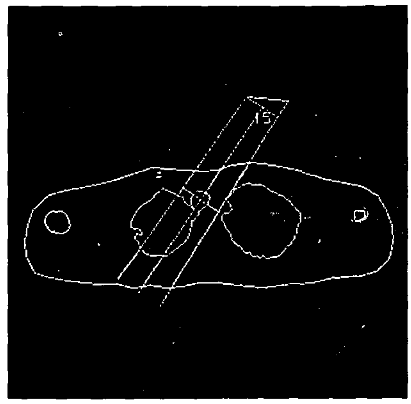
http://www.va.gov/



A visual medical chart containing multidisciplinary images is integrated into a medical record containing a patient's demographic and clinical data. Developed by the VA, this prototype will upgrade the visual chart system used at two VA medical centers.



VA is prototyping OE/RR (Order Entry/Results Reporting), a component of a patient's electronic medical record. Using it and the patient's health summary data contained in a hospital's information system, a clinician can enter laboratory reports and view results, including special alerts to critical conditions.



A source image slice with a beam placed and some contours drawn. The contours denote regions of different density and are subsequently used in the radiation dose calculation in place of the source image. The beam specifies the path of the central ray, width, placement, and the presence of a blocking wedge.

Radiation Treatment Planning

In treating a cancer patient, a radiation oncologist seeks to determine the optimum placement, blocking, and intensity of radiation beams so that the body volume to be irradiated receives the maximum dosage while surrounding tissue has minimum damage. Computationally intensive techniques have proved useful in planning such treatment. The input is a series of images such as from a CAT scan; bones, organs, and the tumor are identified; a beam placement plan is developed; and the radiation is simulated. The output is a series of two-dimensional images showing isodose contour maps and relative radiation absorption. The output from using different beam placement plans is compared in order to select the most effective plan. While today all beams lie in the same plane, researchers have begun assembling the computing resources and implementing simulations for oblique beam placement plans that promise to be more effective. At DCRT, these resources include an Intel iPSC/860, high speed networks, and graphical workstations.

Functional Neurological Image Analysis

Researchers at several NIH institutes, including DCRT, the National Institute on Aging, the National Institute of Mental Health, and the

Single slices of MRI scans of two normal children of different ages. The leftmost scan is warped to have the form of the middle scan using the tie-points identified by the squares. The warped image is shown at right. This work was conducted at NIH's National Institute of Mental Health.

National Institute of Neurological Disorders and Stroke, and at the UCLA School of Medicine, are developing software systems to analyze brain image data. In order to compare images from different subjects or from the same subject at different ages or disease stages, linear and nonlinear transformations are used to "warp" the images into a standard "stereotactic space." The warping algorithm developed at UCLA was implemented on DCRT's Intel iPSC, and is being tested with patient data.

DCRT continues its own R&D in warping algorithms and in applications that use these algorithms. This includes (1) enhancing and using the Hammersmith Hospital's Statistical Parametric Mapping (SPM) warping software for PET scans, (2) implementing SPM in the C language and realizing significant speed increases, (3) implementing principal components analysis for functional brain image data, and (4) investigating methods for segmenting magnetic resonance images into regions of gray matter, white matter, and cerebrospinal fluid, which can be used as structural templates for functional neuroimaging studies.

Project Hippocrates: HIgh PerfOrmance Computing for Robot-AssisTEd Surgery

Advanced planning, simulation, and execution technologies for the next generation of computer-assisted surgical robots are being developed in a collaborative effort involving researchers in robotics, computational mechanics, computer science, surgery, and bioengineering. Hip replacement surgery, one of the most common procedures in orthopedic surgery, is targeted because of anticipated benefits from the high precision and accuracy.

The biomechanics-based surgical simulator that is being developed will allow surgeons to evaluate mechanical consequences of a proposed surgical plan. By coupling the simulator with precise surgical robots, the surgeon can plan an "ideal" surgery and ensure that it is carried out. To execute the plan with a robot, the system must be able to register (determine the position and orientation of) a bone in a clinical environment. Surface-based registration is preferred because it does not require surgically-implanted markers to properly align the pre-operative plan with the patient's anatomy, thus reducing the potential for patient trauma associated with the added risk of infection, anesthesia risk, and post-operative pain stemming from a second surgical procedure. Success of both surgical registration and pre-operative simulation depends strongly on the realism of the geometric and physical models. High performance computing is needed to overcome the computational complexities in modeling, simulation, and registration. Advanced human-computer interaction techniques will be used to ensure clinical viability.

This NSF-funded National Challenge is being conducted at Carnegie Mellon University in collaboration with orthopedic surgeons at Pittsburgh's Shadyside Hospital.

http://www.cs.cmu.edu:8001/afs/cs/project/mrcas/www/hippocrates.html

Prototypes for Clinic-Based Collaboration

A team of physicians, computer scientists, and sociologists based at the University of Michigan will design, implement, test, and evaluate a prototype system for viewing medical images over a distance, allowing primary care physicians to collaborate with radiologists out of the region. Research issues in this NSF-funded National Challenge include adapting conventional medical diagnostic practice to using remote collaboration technology and assessing the impact of this technology.

http://www.sils.umich.edu/~weymouth/Medical-Collab/

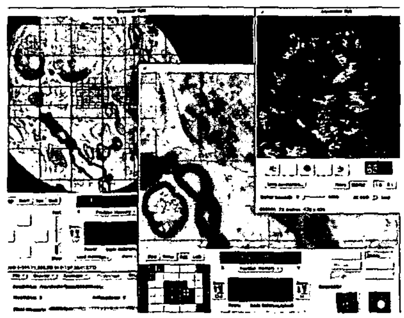
Trusted Interoperation of Health Care Information Systems

This National Challenge effort will develop and demonstrate techniques for the interoperation of health care databases containing sensitive data that mismatch in semantics, representations, and security/privacy policies. Areas being addressed include the relationship and difference between security and privacy; how to access disparate databases without compromising the security and privacy policies of each; and capturing information from multiple databases with different structures. With NSF support, the research will be conducted by the Departments of Computer Science and Medicine at Stanford University in collaboration with Inova Health Systems.

http://www.csl.sri.com/sri-csl-db/bluebook.html

Collaboratory for Microscopic Digital Anatomy (CMDA)

This NSF-funded National Challenge is a collaboratory providing remote acquisition, analysis, and visualization of data from the Intermediate High Voltage Electron Microscope at the University of California at San Diego. CMDA will provide transparent distribution of computationally intensive tasks such as those involved in the acquisition and derivation of three-dimensional representations of biological structure using computed electron tomography. A scientist can use the remote Gridbrowser interface (illustrated above right) to control the microscope; focusing and registration of images will be automatic. When a tomographic representation is desired, the images will be acquired, automatically transferred to appropriate high performance computing systems for tomographic reconstruction and volume rendering, and automatically transferred back to the scientist's workstation for viewing and analysis. Scientists



This Gridbrowser interface shows (1) a low magnification survey with gridlines identifying the source of the higher magnification view, (2) cross-hairs identifying the current position of the microscope stage (which can be changed remotely), and (3) a red-green stereo view of the three-dimensional volume derived from acquired data.

at different sites will be able to use Gridbrowser for collaborative analysis.

http://szechuan.ucsd.edu/CMDA.html

Distributed Imaging Over Gigabit Networks

The focus of this NSF-supported National Challenge at Washington University in St. Louis is on distributed imaging applications over gigabit networks and the underlying protocols and networking technology needed to support those applications. There are two principal applications. One uses a computational optical sectioning microscope to visualize the three-dimensional structure and function of developing organisms at the cellular level. In this application, gigabit networks are used to transfer raw data from the microscope to mass storage devices and to massively parallel image processors that synthesize the images and then transfer them (again, across the network) to display stations at sites around the campus. The second application is in neurosciences: three-dimensional data sets of the brains of individual specimens are mapped onto a labeled reference brain in order to help identify significant features and conduct comparative studies. The enabling technologies include a local gigabit network connecting biologists, neurologists, computer scientists, and electrical engineers; and operating system, visualization, and database software for near-real-time interaction by the participants with one another and with the experiments.

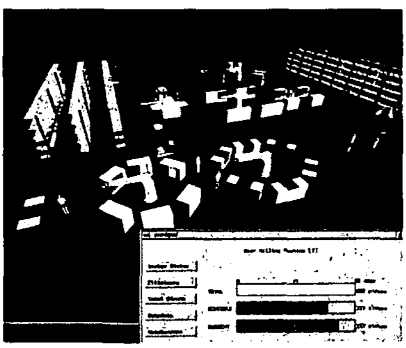
http://arl.wustl.edu/arl/projects/nca/intro.html

Manufacturing Processes and Products

HPCC technologies are being applied to the design of processes and products as well as to the processing and manufacturing of products. Key aspects include the development of new standards for product data that are used throughout the design and production process and the integration of design with processing and manufacturing. Of particular interest are the mechanical, electronics, construction, and chemical industries.

NIST's Manufacturing Testbed

As part of its Systems Integration for Manufacturing Applications (SIMA) program, NIST established the Advanced Manufacturing Systems and Networking Testbed (AMSANT).



An example of the types of user interfaces required to visualize data on manufacturing activities in a production facility. A prototype facility was simulated to provide for real-time views into the factory control system database and to simulate manufacturing data access by multiple users.

AMSANT serves as a demonstration site for industrial suppliers and users and assists industry in developing and implementing voluntary standards. Prototype systems and interface specifications are communicated to appropriate standards organizations. On-going work addresses interface protocols, information models, and integration architectures. In FY 1995 the AMSANT for mechanical parts production applications and process plants industries was implemented. Products are developed in collaboration with other HPCC agencies and U.S. industry and are made available through workshops, training materials, Internet-based data repositories, and pre-commercial prototype systems that can be installed by potential vendors for testing and evaluation. A prototype STEP Conformance Testing System and Abstract Test Suite Developer's Toolkit, and a software development environment for STEP Applications Protocols were demonstrated.

http://elib.cme.nist.gov/fasd/projs/sima-pm/sima.htm

Information Integration for Simulation-Based Design and Manufacturing

This NSF-funded National Challenge, involving researchers at the University of Iowa and Rensselaer Polytechnic Institute (RPI), is investigating how the NII can be used by engineering organizations wishing to collaborate on design and manufacturing problems. RPI has developed core technologies used to demonstrate the exchange of a solid model of a real steering pump between different CAD (Computer Aided Design) systems at two automobile companies. This was the first such exchange using STEP, which was recently released as an ISO (International Organization for Standards) standard. STEP serves as a neutral format for exchanging product data between any pair of different CAD systems.

This approach is common to computer science — if for example there are 10 different software packages that do fundamentally the same computations but use different data formats, then establish a neutral format and develop 20 translators, one for each package in each direction. Use one translator to move from one package to the neutral format and another translator to move to the other package. This approach is preferable to developing 90 translators, one for each pair of different packages and both directions. The comparison is much more pronounced when the number of packages is larger, as is the case with thousands of CAD systems.

In FY 1996 research topics include simulating mechanical systems for multidisciplinary design over the Internet and other wide-area networks, design of neutral databases for integrated engineering systems, and distributed control systems for processing engineering changes. The simulation environment will use the PDES/STEP standard for engineering data exchange, the OMG CORBA standard for object service modeling, and Mosaic for data presentation. It will be tested by rebuilding models of components of the Hummer Vehicle's suspension assembly using different CAD software packages and Iowa's simulation-based design environment.

Micromechanical Characteristics of Composite Materials

This NSF-funded National Challenge located at the University of Texas will develop a computationally-based toolbox for the optimal design of composite materials. Mechanical engineers and computer scientists will create a collection of computational components with which engineers can design and analyze composite materials directly from the microstructural level. The microstructural properties will be provided by industrial collaborators such as ALCOA and COMCO of Austin, TX. Research issues include micromechanical modeling of composites, finite element analysis, parallel algorithms, and data structures.

In FY 1995 ARPA expects to:

- Demonstrate a mobile computing system CAD environment through the design of early prototype, high bandwidth, pico-cellular, and wireless access points to the wireline infrastructure

- Demonstrate network-based access to the Multichip Module fabrication services

- Develop information infrastructure services for manufacturing (for example access to engineering analysis and rapid prototyping)

In FY 1996 ARPA plans to:

- Demonstrate design environments supporting simulation and synthesis of wireless systems spanning integrated circuits to network applications

- Expand network design and manufacturing services to include factory simulation and reusable product and process design libraries

8. Basic Research and Human Resources

Basic Research

HPCC basic research has two goals — to develop new methods for addressing fundamental technological limitations encountered as the field advances, and to develop the foundations for next and future generations of HPCC technologies.

The underlying basic research that develops the foundations for future generations of HPCC technologies can be characterized in seven broad areas: components, communications, computing systems, software and tools, intelligent systems, information management, and applications. Examples of accomplishments resulting from NSF basic research illustrate its impact. These include the following:

- The Ray Casting Engine project used VLSI (very large scale integration) design techniques and MOSIS (metal oxide semiconductor implementation service), an ARPA- and NSF-funded service for academic research and education, to design and build a new, application-specific, massively parallel computer to compute solid geometry quickly for mechanical design. This project resulted in new approaches to solid modeling, metrology, and the treatment of sculptured objects and led to examination of completeness and utility of ray representations.

- The Supercomputer ToolKit project used inexpensive, fast commodity hardware modules with static reconfigurable interconnect and dedicated software on each module to build a cost-effective application-specific computing machine that achieved supercomputer performance in simulating a model of the complete solar system. This dedicated system was used to carry out a 100 million year integration and proved planetary orbits to be chaotic. It was also used in engineering design to carry out real-time computations with sensors and actuators to design dynamically stable bridges.

- A project on the Correctness of Hardware and Software Systems used theoretical models to study complex systems with a large number of states, to study how to scale techniques for state-space blowup, and to verify that models of such systems satisfy specifications. The resulting techniques were used to identify coherency and potential deadlock errors found in the IEEE Futurebus+ protocol for application systems developed under the Navy's "Next-Generation Computer Resource" program.

- A Reduced Complexity Robotics project used modular hardware, simple sensors and actuators, efficient software design, rigorous scientific models and principles, theories on completeness, correctness, and complexity of plans for flexible parts handling and assembly systems that can be simulated, built, and programmed in a matter of weeks versus months and years. Such systems are directly and seamlessly controllable to produce any quantity of a selected assembly within a prescribed family of products, programmed directly from piece-part CAD files and manufacturing process sheets.

- A Multi-Media, Knowledge-Based, Evolutionary Biomedical Database System studies the retrieval of images by content rather than image ID; the integration of images, text, and domain knowledge; approximate or fuzzy query answering; and application and demonstration of these techniques on biomedical images and data, in particular hand growth and brain tumor studies. The work has led to the development of new spatio-temporal models for image representation, and the prototype system has potential for use in radiology and other areas using image databases such as manufacturing and digital libraries.

- A Robust Spoken Language Systems project that ARPA also supports focuses on the development of next-generation systems more capable of handling continuous speech, speakers' variabilities, disturbances from environmental noise, and larger vocabularies. The resulting system architecture has been licensed to U.S. West for enhancing telephone voice transactions. The U.S. 1995 Census will field test the system for interactive telephone interviews.

- The Mentat project is developing an object-oriented programming system that transparently distributes computations across geographically distributed, heterogeneous, computing resources. Applications in biochemistry for comparing sequence libraries and in electrical engineering for testing combinatorial circuits are executing in an environment of Sun, IBM, and SGI systems. Mentat software has been distributed to more than 150 sites.

- A Computational Molecular Biology project is studying the problem of how to locate genes and important functional signals in "raw sequences." Efficient combinatorial matching algorithms and algorithms for sequence reconstruction along with an innovative special biological chip as a DNA sample holder have been developed. This will allow the reconstruction of unknown DNA fragments from a set of common DNA sequences. The resulting binary sequencing by hybridization and customized DNA chips are expected to impact medical genetic diagnosis, DNA sequencing, DNA fingerprints for legal systems, and recombinant DNA technology.

Other applications that benefit directly from high performance computing and communications research include physics (for example, astrophysics, high energy physics, low temperature physics, and material sciences), chemistry (from basic chemical properties to chemical processing), biology (including genome understanding, molecular dynamics, neural and cellular models, and pharmacology), and a full spectrum of engineering disciplines (such as fluid dynamics, petroleum engineering, and structural dynamics).

NSF supports long-term investigator-initiated basic research in science and engineering disciplines represented by its Directorates for Biological Sciences, Engineering, Geosciences, Mathematical and Physical Sciences, and Social, Behavioral, and Economic Sciences. This research is generally focused on improving the ability of researchers to use high performance computing and communications to advance fundamental understanding in those disciplines. It frequently takes the form of advanced software development that ultimately has impact outside of the particular discipline. These Directorates also participate in Grand Challenge and National Challenge activities. Some specific activities are as follows:

- Through group awards and individual investigator awards, graduate and postdoctoral students are supported in order to increase the pool of talent trained in scientific computing and information processing. In FY 1995 NSF received 317 preproposals in response to a solicitation for multidisciplinary group-oriented research in Grand Challenges, National Challenges, enabling technologies, and challenges in computer science, mathematical sciences, and problem-solving environments. Between 15 and 20 multi-year awards are anticipated in FY 1995.

- Through its institutional infrastructure and research instrumentation activities, in FY 1995 NSF provided resources for conducting research in scalable I/O, network applications in education and manufacturing, gigabit-speed networking research, virtual manufacturing, multimedia systems, basic experimental research, and for the NIE program described immediately below. In FY 1996 NSF will evaluate existing programs, develop infrastructure support for basic research needed for educational use of the NII, continue infrastructure support for virtual manufacturing and other NII activities, and increase NIE participation.

- The Computer and Information Science and Engineering (CISE) and the Education and Human Resources (EHR) Directorates at NSF have established the Networking Infrastructure for Education (NIE) program. NIE's goals are to build synergy among technology and education researchers, developers, and implementers; encourage innovation and experimentation by educational groups; and integrate networking technology with education reform. Co-funded with ARPA, NSF awarded 19 planning grants and policy studies, 11 multi-year projects, and 10 one-year supplements to existing awards. In FY 1995 between 100 and 125 proposals are expected. One focus of the new awards will be on bringing networking information to previously underrepresented communities such as tribal colleges.

- NSF funds basic research in gigabit speed networking including work at the gigabit testbeds described beginning on page 9; gigabit switching systems; protocols and software structures for network management; resource discovery among collaborative information spaces in large decentralized environments; network information theory; multi-sender and multi-receiver network security; modulation, detection, and coding of reliable information storage and retrieval; all-optical networks; and optical technologies for computing and communications. In FY 1996 NSF plans to fund research aimed at more effective use of large information resources, solicit and make awards for wireless access testbeds, continue collaboration in a 10 Gb/s testbed with ARPA, and examine the convergence of computing, entertainment, and telecommunications from the standpoint of Internet experience.

NASA funds basic research in computer architectures, fundamental algorithms, computational complexity, networked and distributed computation, numerical analysis, and application-specific algorithms. NASA supports the Illinois Computer Laboratory for Aerospace Systems and Software (ICLASS) at the University of Illinois and supports research centers at its Ames, Langley, and Goddard facilities.

DOE supports basic research in applied mathematics, computer science, and computational science toward solving large scientific problems. Work is conducted at ten DOE laboratories and more than 30 universities; more than 60 undergraduate students and more than 40 graduate students are supported. FY 1995 accomplishments include:

- Software based on adaptive mesh refinement techniques for resolving three-dimensional fluid flow phenomena in complex geometries

- Security procedures for voice and data transmission based on sending the data encapsulated in a chaotic signal and decoding the signal using techniques for controlling chaos

- Software for resolving large-scale stochastic programming problems such as those found in large commercial electric power distribution systems, networks, and airline crew equipment scheduling operations

- Development and implementation of sparse matrix techniques for finite element software that performs crash simulation and analysis

FY 1996 DOE plans include:

- Research in inverse scattering, nondestructive evaluation, and discrete mathematics applied to parallel programming and genome sequencing and mapping

- Multidisciplinary, multi-sector research in DOE's Grand Challenge applications

- Initiating a graduate student program based at industrial and laboratory sites

NLM grants provide training in HPCC technologies for biomedical professionals and provide training for computing professionals in medical and health applications.

Education and Training

During the three years that the HPCC Program has formally existed, the community of expert users has grown from a small enclave of researchers in computing and communications technologies to large groups of researchers and students in all fields. The community of able users spans all sectors of the economy and all sectors of society throughout the U.S. and around the world. This growth in knowledge and ability was initially the result of hands-on and formal training at HPCC-funded facilities (described beginning on page 35) and in industry; today this type of training and education is being conducted on line; at universities, colleges, elementary and secondary schools; at other places of learning such as museums and libraries; and in industrial training programs throughout the country.

In FY 1994 NSF established a Native American Telecommunications Technical Assistance think tank, joined with ED and the National Telecommunications and Information Administration (NTIA, part of the Department of Commerce) in sponsoring a workshop on the role of state networks in developing statewide networking infrastructure, and developed museum-based networking support activities in urban centers.

On-going NSF activities include the following:

- Joint NSF/ARPA funding of more than 4,000 student projects in VLSI fabrication by MOSIS

- The Superquest program that provides high school students and teachers with research experiences at its Supercomputer Centers, undergraduate course and curriculum development, and research experiences for undergraduate students

- Pilot Educational Networks project to develop, implement, test, and evaluate applications of computing and communications to education. The NIE (described on page 96) addresses issues of large scale networking for education. Issues include scalability, cost/benefits, policy, effective applications, and educational impact for user-driven models of computer networking in education.

- Pilots and large-scale models for educational applications of digital libraries

- The Maryland "Virtual High School for Computational Science" on the Internet, where students throughout Maryland learn science and mathematics by participating in collaborative application projects using computational science, guided by experienced teachers and other mentors. Three groups participate:

  — Montgomery Blair High School, which has developed a collaborative teaching approach based upon computational science applications

  — Maryland high school teachers and students, including those at isolated rural, inner-city, and suburban schools

  — Government agencies, business groups, and network groups concerned about the schools' abilities to respond to the promise of advanced technologies

The NSF postdoctoral training program in computational science and engineering and experimental computer science continues. To date some 80 people received two years of support to pursue computational science across a spectrum of disciplines and in experimental computer science.

NASA has a broad-based program designed to produce rapid improvement in both K-12 and lifelong learning in science, mathematics, and engineering. The two components of this program are (1) open competitive solicitations called "cooperative agreement notices," and (2) an educational outreach program at seven NASA field centers nationwide.

In FY 1995 NASA solicited for projects to use the Internet in lifelong learning through the cooperative agreement notice "Education, Training, and Lifelong Learning in Aeronautics." Details can be found at:

ftp://quest.arc.nasa.gov/00/Pub/CAN95

Outreach program accomplishments at the field centers include:

- Ames Research Center (Moffett Field, CA) — Produced several popular videos on approaches to and benefits from using the Internet in the classrooms

  http://quest.arc.nasa.gov/

  — Established a popular educational resource Web page supporting teachers nationwide

  http://quest.arc.nasa.gov/trc/toc.html

- Dryden Flight Research Center (Edwards, CA) — Investigating how to use the Internet to assist challenged and emotionally disturbed students in the classroom

- Goddard Space Flight Center (Greenbelt, MD) — Established an Internet-based program for Goddard scientists to work with teachers in all 24 Maryland school districts to improve the teaching of Earth and environmental science

  http://www.gsfc.nasa.gov/GSFC_homepage.html

- Jet Propulsion Laboratory (Pasadena, CA) — Produced multimedia software to guide school children through a space exploration mission from defining objectives through designing and selecting instruments and platforms to analyzing scientific data

  http://www.jpl.nasa.gov/educ/education.html

- Johnson Space Center (Houston, TX) — Developed a knowledge robot (Knobot (TM)) to discover and provide materials on

Gonzaga High School students using NASA's Earth System Science Community Curriculum Testbed. These students are using the Internet to access and review student-published "chapters" submitted by students from many schools, including the Model Secondary School for the Deaf at Gallaudet University.

the Internet tailored to the needs of K-12 teachers

http://www.jsc.nasa.gov/stb/ILIAD/Mosaic/iliad.html

- Langley Research Center (Hampton, VA) — Deployed Internet networks to public schools economically by using the local carrier telephone infrastructure

  http://k12mac.larc.nasa.gov/Networks/Models.html

- Lewis Research Center (Cleveland, OH) — Demonstrated prototype high bandwidth, low-cost Internet access to schools using wireless radio frequency technology

  http://www.lerc.nasa.gov:80/Other_Groups/K-12/K-12_homepage.html

On-going DOE activities include Adventures in Supercomputing, which trains in-service high school teachers and middle school teachers at select sites, and the National High School Honors Program, which provides summer supercomputing enrichment for gifted high school students.

In FY 1996 DOE plans to complete its electronic computational science textbook. In FY 1995 and FY 1996 DOE plans to develop a similar science textbook. The agency also plans to conduct a networking technology assessment project to identify educational technology tools and curricula for a variety of schools and to produce a catalog of effective technologies for use by teachers, administrators, and state and local education officials.

DOE supports more than 50 students in its Computational Science Graduate Fellowship program.

NLM funds both planning and implementation grants to academic medical centers with the goal of developing, testing, and implementing generalizable systems linking administrative, clinical, educational, and research databases so that they appear as one to the user. Four grants continue; approximately seven planning awards and two implementation awards will be made in FY 1995.

NCRR funds undergraduate and graduate students in the biomedical sciences and biomedical scientists in the use of HPCC technologies. NCRR trains K-12 teachers to use interactive learning tools in classroom activities and in scientific-computer literacy.

NSA is developing a skill-based assessment and training model in order to maintain and enhance employee abilities in today's constantly changing workplace.

EPA is developing a "virtual support center" through which advanced environmental modeling tools and assistance are provided to a target user community. That community consists of Federal government personnel, environmental decision makers from state governments, and industrial personnel representing the Consortium for Advanced Modeling of Regional Air Quality. The user community works with developers and trainers to ensure that system capabilities and interfaces meet their needs. The center will be Internet-accessible to environmental decision makers across the country. EPA also funds its EarthVision computational science educational program for high school students, graduate student support, training fellowships, conferences, and workshops. In FY 1996 EPA will develop environmental simulation software to expose high school students to computational environmental analysis.

The ten ED regional educational laboratories have created a Technology Task Force to develop Internet-based information systems. Connectivity of all ten labs will be completed in FY 1995 along with information servers at nine of the labs.

ED's Teacher Networking Project funds networking and applications for teacher professional development, with priority given for teachers serving poor families.

HPCC Program Organization

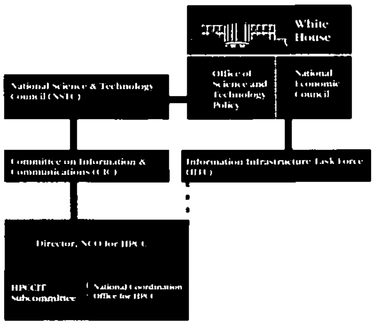
The HPCC Program in FY 1996 includes 12 Federal departments and agencies. The High Performance Computing, Communications, and Information Technology (HPCCIT) Subcommittee consists of representatives from each of these organizations. HPCCIT reports to the Committee on Information and Communications (CIC) R&D Committee of the National Science and Technology Council (NSTC). The Director of the National Coordination Office for High Performance Computing and Communications (NCO) reports jointly to the Director of the Office of Science and Technology Policy (OSTP), Executive Office of the President, and the Chair of the CIC. The NCO serves as a liaison to the U.S. Congress, state and local governments, foreign governments, industry, universities, and the public.

The Program is organized in five components:

High Performance Computing Systems (HPCS)

- Accelerated development of scalable computing systems, with associated software, including networks of heterogeneous systems ranging from affordable workstations to large scale high performance systems

- Technologies to enable the use of advanced component, packaging, mass storage, and communications technologies for the design of large scale parallel computing systems



National Research and Education Network (NREN)

- Broadened network connectivity of the research and education communities to high performance computing and research resources

- Accelerated development and deployment of networking technologies

Advanced Software Technology and Algorithms (ASTA)

- Prototype solutions to Grand Challenge problems

- Improved algorithms, software technologies, and software tools for more efficient use of scalable computing systems

- Deployment of advanced high performance computing systems for testing, technology integration, and evaluation

Information Infrastructure Technology and Applications (IITA)

- Prototype solutions to National Challenge problems using HPCC enabling technologies

- Accelerated development and deployment of NII enabling technologies

Basic Research and Human Resources (BRHR)

- Support for research, training, and education in computer science, computer engineering, and computational science; and infrastructure enhancement through the addition of HPCC resources

Interaction with Government, Industry, and Academia

The NCO and HPCC agencies meet frequently with representatives from the Federal and other governments, industry, and academia. The HPCC Program actively seeks input and advice from all interested parties.

During FY 1994 the NCO, HPCCIT Subcommittee, and participating departments and agencies held numerous meetings with representatives from the U.S. Congress, Federal and Federally-charted organizations, state and local organizations, industry, academia, professional societies, foreign governments, and others, to exchange information about technical and programmatic issues, trends, and needs.

Between November 1993 and June 1994 the NCO and HPCCIT held five open meetings — with representatives from the computing systems, telecommunications, and software industries, and with directors of both Federal and non-Federal computing centers. In October 1994 the Subcommittee held non-disclosure briefings by 12 mass storage vendors and two other organizations conducting R&D in this area. HPCC departments and agencies cosponsored the Second Pasadena Workshop on Systems Software for High Performance Computing Systems held in January 1995.

Program Coordination

The HPCCIT Subcommittee and its Executive Committee coordinate Program planning, budgeting, implementation, and review. This is done through monthly meetings that include information exchanges, common development of multi-organization programs, and the review of the plans and budgets of participating organizations. HPCCIT charters two groups that coordinate activities in specific areas:

- The Scientific and Engineering Computing Working Group (S&ECWG), chaired by NASA, has as its mission to identify, promote, transfer, and coordinate software developed under HPCC. The Working Group also facilitates coordination of the HPCC Grand Challenge applications research and fosters testbed and software tools evaluation. The S&ECWG planned the mass storage briefings, HPCC participation in the Pasadena Workshop, and is planning computing systems briefings to be held later in 1995.

- The IITA Task Group, co-chaired by ARPA and NSF, is chartered to acquire information about on-going and planned IITA-related activities in HPCC departments and agencies, to develop a plan for coordinating IITA activities among those organizations, to assist the HPCCIT Subcommittee in reporting requirements related to IITA, and to provide advice on technology and policy related to IITA to the HPCCIT Subcommittee and its parent committees. The Task Group is planning a Digital Libraries Workshop scheduled for mid 1995.

The HPCC Program is responsible for Federal networking R&D. The HPCC Program and the Federal Networking Council (FNC), which is chartered by NSF, work together to establish an effective interagency forum and long-term strategy to oversee the operation and evolution of the federally-funded portion of the Internet in support of science, research, and education. FNC members represent Federal agencies that need to operate and use increasingly advanced networking facilities, mainly for research and education.

The NCO and HPCC agencies work with members of the Information Infrastructure Task Force in addressing NII policy issues, and with other Federal agencies involved in helping schools, libraries, and medical facilities become part of the NII.

Program Evaluation

Federally-chartered and Federal organizations, industrial and academic organizations, and professional societies provide critical analyses of the Program through conferences, workshops, and reports. These efforts help make Program goals and accomplishments better understood, and strengthen Program planning and management.

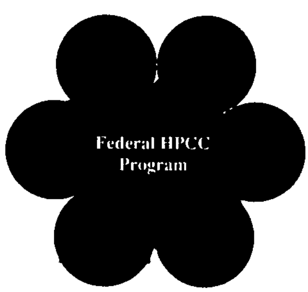
In FY 1994 the NCO and members of the HPCCIT Subcommittee met with representatives from the National Research Council's Computer Science and Telecommunications Board, which was tasked by Congress with evaluating aspects of the HPCC Program. The report, "Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure," was published in March 1995 by the National Academy Press, Washington, DC.

HPCC and CIC

Building on the successes of the HPCC Program, in March 1995, the CIC released a Strategic Implementation Plan entitled "America in the Age of Information." As stated in the plan's abstract, "The 21st century will be the Age of Information. Implementation and support of this plan will help guarantee America's preeminence in information technology as our country enters the 21st century."

The plan sets a high level strategy for Federal R&D in information and communications technologies. It identifies a broad base of opportunities for investments that leverage the technical strengths of individual Federal departments and agencies while meeting their mission-specific information and communications technology goals. These investments support the science and technology base essential to continued economic competitiveness and national defense. The plan will be revised and extended through interchanges among government, industry, and academia. The Plan is built around six Strategic Focus Areas designed to focus fundamental information and communications research and accelerate development in ways that are responsive to the NSTC's overarching goals, agency mission goals, and our Nation's long-term economic and defense needs. The six Strategic Focus Areas are:

Information & Communications R&D Investment



Conceptual Relationships of Total I&C R&D Investment
Strategic Focus Areas & HPCC Investment (1994)

Global-scale information Infrastructure technologies

These are advanced applications building blocks and widely-accessible information services (such as usage metering and payment) upon which large scale integrated and distributed applications such as National Challenge applications will be constructed.

High Performance / Scalable Systems

These systems will enable broad deployment of information and communications technologies so that "high performance" and "low end"applications can operate in a seamlessly integrated fashion.

High confidence systems

These systems will provide the availability, reliability, integrity, confidentiality, and privacy needed by the increasingly diverse users of the Nation's emerging ubiquitous information infrastructure

Virtual environments

These environments will enable interaction in real time and in increasingly realistic ways over large distances by diverse users; they will transform scientific experimentation and practice (telemedicine is an example) and will play increasingly important roles in education and training.

User-centered interfaces and tools

"Human centered" information systems will encourage broader use of NII (and GII and DII) technologies, easier navigation and "mining" of information resources by these users, and will allow end users to develop and tailor services and applications easily in order to suit their needs.

Human resources and education

The CIC program includes educating and training people so that they have the knowledge, skills, and insights to lead research in science and technology and to apply their discoveries to national needs. CIC and CET (NSTC's Committee on Education and Training) R&D will enable new educational technologies to provide the diverse 21st century workforce with the tools it needs in order to participate in the information revolution.

The CIC R&D activity areas include components, communications, computing systems, support software and tools, intelligent systems, information management, and applications. These applications fall into three classes:

- High performance applications for science and engineering
- High confidence applications for dynamic enterprises
- High capability applications for the individual

Buy American Report

The Congress has requested that the HPCC Program inform it about certain HPCC funding of non-U.S. activities.

ARPA is the only HPCC agency that in FY 1995 and FY 1996 is entering into a grant, contract, cooperative agreement, or cooperative research and development agreement, for HPCC research and development with either (1) a company other than a company that is either incorporated or located in the U.S., and that has majority ownership by individuals who are citizens of the U.S., or (2) an educational institution or non-profit institution located outside the U.S. That activity is with University College, London, for approximately $120,000 per year. It is part of the Program's NREN effort and focuses on developing the underlying technology to be used in pilot demonstrations of multi-country, multi-way, wide-area multimedia services.

In FY 1995 and FY 1996, no HPCC procurement exceeds $1 million for unmanufactured articles, materials, or supplies mined or produced outside the U.S.; or for manufactured articles, materials, or supplies other than those manufactured in the U.S. substantially all from articles, materials, or supplies mined, produced, or manufactured in the U.S.

Reports about the HPCC Program

The NCO provides electronic, print, and video materials to hundreds of media representatives in the U.S. and abroad, and responds to thousands of requests for information from Congressional offices, industry, academia, and the public. The NCO distributed more than 11,000 copies of its FY 1995 report entitled "High Performance Computing and Communications: Technology for the National Information Infrastructure." A video with the same name as this report is available from the NCO.

The NCO has both Web and gopher servers that contain the FY 1994, FY 1995, and soon this FY 1996 annual report, information about sources of funding for HPCC R&D, links to the servers of HPCC agencies, and related information. In January 1995, users of more than 5,000 computing systems accessed the Web server and more than 2,000 accessed the gopher server; these users were from 50 countries.

http://www.hpcc.gov

HPCC Program Summary

HPCC Program Goals

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

Provide wide dissemination and application of the technologies to speed the pace of innovation and to improve the national economic competitiveness, national security, education, health care, and the global environment

Provide key enabling technologies for the National Information Infrastructure (NII) and demonstrate select NII applications

HPCC Agencies

AHCPR – Agency for Health Care Policy and Research, Department of Health and Human Services

ARPA – Advanced Research Projects Agency, Department of Defense

DOE – Department of Energy

ED – Department of Education

EPA – Environmental Protection Agency

NASA – National Aeronautics and Space Administration

NIH – National Institutes of Health, Department of Health and Human Services

NIST – National Institute of Standards and Technology, Department of Commerce

NOAA – National Oceanic and Atmospheric Administration, Department of Commerce

NSA – National Security Agency, Department of Defense

NSF – National Science Foundation

VA – Department of Veterans Affairs

HPCC Program Strategies

Develop, through industrial collaboration, high performance computing systems using scalable parallel designs and technologies capable of sustaining at least one trillion operations per second (teraops) performance on large scientific and engineering problems such as Grand Challenges

Support all HPCC components by helping to expand and upgrade the Internet

Develop the networking technology required for deployment of nationwide gigabit-speed networks through collaboration with industry

Demonstrate the productiveness of wide-area gigabit networking to support and enhance Grand Challenge and National Challenge applications collaborations

Demonstrate prototype solutions of Grand Challenge problems that achieve and exploit teraops performance

Provide and encourage innovative use of high performance computing systems, emerging software and algorithms, and network access technologies for solving Grand Challenge, National Challenge, and other advanced applications to enable new capabilities in missions such as national security, public health, public safety, environment, and education

Create an infrastructure, including high performance computing research centers, networks, and collaborations that encourage the diffusion and use of high performance computing and communications technologies in U.S. research and industrial applications

Work with industry to develop information infrastructure technology to support the Global Information Infrastructure

Leverage the HPCC investment by working with industry to implement National Challenge applications

Enhance computational science as a widely recognized discipline for basic research by establishing nationally recognized and accepted educational programs in computational science at the pre-college, undergraduate, graduate, and postgraduate levels

Increase the number of graduate and postdoctoral fellowships in computer science, computer engineering, computational science and engineering, and informatics, and initiate undergraduate computational sciences scholarships and fellowships

Overview of the Five HPCC Program Components

Five integrated components represent the key areas of high performance computing and communications:

HPCS – High Performance Computing Systems

Extend U.S. technological leadership in high performance computing through the development of scalable computing systems, with associated software, capable of sustaining at least one trillion operations per second (teraops) performance. Scalable parallel and distributed computing systems will be able to support the full range of usage from workstations through the largest-scale highest-performance systems. Workstations will extend into portable wireless interfaces as technology advances.

NREN – National Research and Education Network

Extend U.S. technological leadership in computer communications by a program of research and development that advances the leading edge of networking technology and services. NREN will widen the research and education community's network connectivity to high performance computing and research centers and to electronic information resources and libraries. This will accelerate the development and deployment of networking technologies by the telecommunications industry. It includes nationwide prototypes for terrestrial, satellite, wireless, and wireline communications systems, including fiber optics, with common protocol support and applications interfaces.

ASTA – Advanced Software Technology and Algorithms

Extend U.S. technological leadership in the development of advanced software and algorithms, enabling the use of HPCC-developed tools and methods, particularly to address large-scale problems such as the Grand Challenges. Software and early prototype applications are tested and evaluated on high performance computing and communications systems. Grand Challenge problems are computationally intensive problems such as forecasting weather, predicting climate, improving environmental modeling, building more energy-efficient cars and airplanes, designing better drugs, and conducting basic scientific research.

IITA – Information Infrastructure Technology and Applications

Extend U.S. leadership in the development of advanced information technologies supporting distributed applications, intelligent interfaces to information systems, virtual environments, and National Challenge applications. National Challenge problems are information-intensive applications such as education and lifelong learning, digital libraries, health care, advanced manufacturing, electronic commerce, and environmental monitoring. IITA efforts will strengthen the HPCC technology base, broaden the market for these technologies, and accelerate industry development of the Global Information Infrastructure.

BRHR – Basic Research and Human Resources

Support research, training, and education in computer science, computer engineering, and computational science, and enhance the infrastructure through the addition of HPCC resources. Initiation of pilot projects for K-12 and lifelong learning will support expansion of the NII.

Evaluation Criteria for the HPCC 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, information infrastructure, and basic research, to enable solution of the Grand Challenges and the National 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.

Agency Approval

The proposed program or activity must have policy-level approval by the submitting agency.

Agency Budgets by HPCC Program Components

FY 1995 Budget (Dollars in Millions)

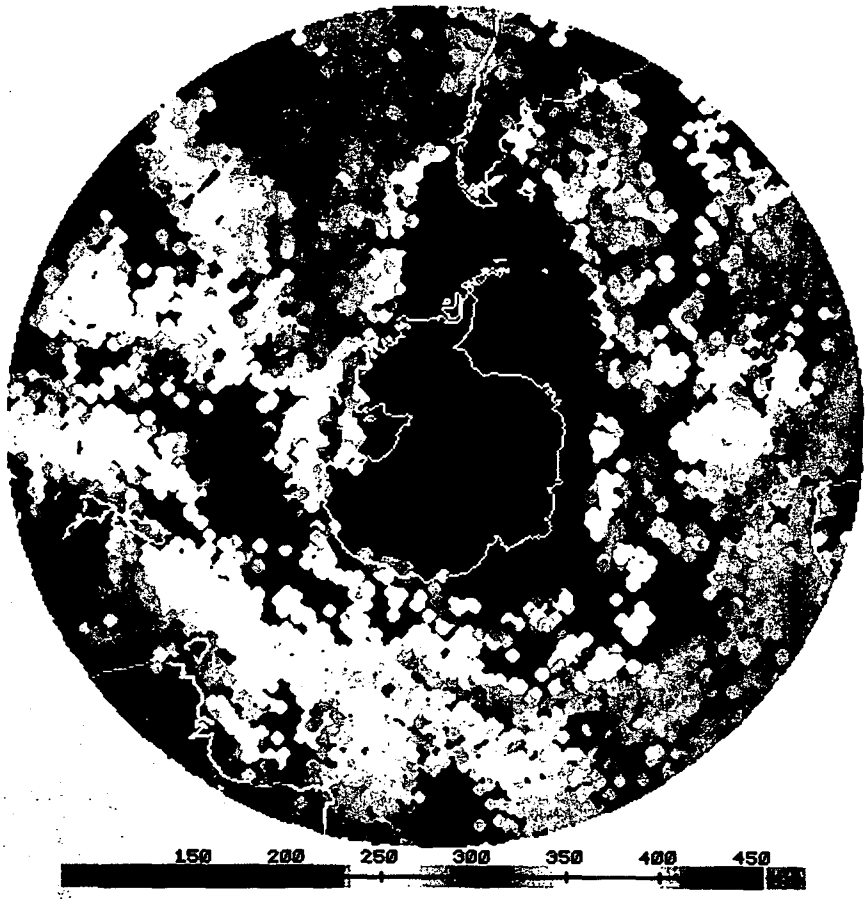
Agency	HPCS	NREN	ASTA	IITA	BRHR	TOTAL
ARPA	111.4	55.8	31.3	133.0	12.7	**344.2**
NSF	22.1	52.1	127.9	38.8	56.1	**297.0**
NASA	9.6	19.1	77.8	21.3	3.6	**131.4**
DOE	8.6	16.0	64.5	3.0	20.9	**113.0**
NIH	4.9	2.7	25.0	24.4	11.5	**68.5**
NSA	21.3	11.3	6.9	0.2	0.2	**39.9**
NIST		2.2	3.6	19.3		**25.1**
EPA		0.7	10.5		1.3	**12.5 ***
NOAA		3.6	2.8			**6.4**
TOTAL	**177.9**	**163.5**	**351.3**	**240.0**	**106.3**	**1,039.0 ***

FY 1996 Budget Request (Dollars in Millions)

Agency	HPCS	NREN	ASTA	IITA	BRHR	TOTAL
ARPA	91.0	51.7	33.9	168.4	18.1	**363.1**
NSF	23.5	54.7	132.6	42.7	60.1	**313.6**
NASA	7.6	20.9	77.0	21.8	4.0	**131.3**
DOE	8.7	17.0	64.3	3.0	20.5	**113.5**
NIH	4.1	2.3	28.1	32.6	11.7	**78.8 ***
NSA	16.8	10.4	12.4	0.2	0.2	**40.0**
NIST		2.2	3.6	28.3		**34.1**
VA		0.2		23.4		**23.6**
ED			11.4	2.3	3.5	**17.2**
NOAA		7.5	7.4	0.5		**15.4**
EPA		0.7	9.3	1.0	1.0	**12.0**
AHCPR						**0.0****
TOTAL	**151.7**	**167.6**	**380.0**	**324.2**	**119.1**	**1,142.7**

* *Updated from the President's FY 1995 and 1996 Budget Requests*

** *As a new agency to the HPCC Program, $8.4 million AHCPR funds are not included in the total.*



A TOVS Ozone Picture of the 1994 Antarctic Ozone Hole

This image was generated from approximately 1,000 individual TIROS Operational Vertical Sounder (TOVS) ozone measurements by the NOAA 12 satellite on October 17, 1994. Each dot represents one observation. The purple and violet regions represent the Antarctic Ozone Hole. TOVS ozone measurements are not generally feasible over the cold Antarctic Plateau, or over cold clouds; these areas are in gray. TOVS permits both daytime and nighttime satellite measurements of the Earth's ozone shield on a real-time basis. The High resolution InfraRed Spectrometer (HIRS) 9.7 micrometer channel is particularly sensitive to artificial kinds of chlorofluorocarbon (CFC) induced ozone depletions that occur in the lower stratosphere. With this tool, scientists can monitor the formation and evolution of this important atmospheric feature.

HPCCIT Subcommittee

Office of Science and Technology Policy (OSTP)

Director: John H. Gibbons

National Science and Technology Council (NSTC)

NSTC Secretariat
Angela Phillips Diaz, Executive Secretary

Committee on Information and Communications (CIC)

Chairs: Anita K. Jones
Lionel S. Johns

Vice Chair: Paul Young

High Performance Computing, Communications, and Information Technology Subcommittee (HPCCIT)

Chair: John C. Toole

Vice Chair: Paul H. Smith

Executive Secretary: Charles R. Kalina

Scientific and Engineering Computing Working Group

Chair: Lee B. Holcomb, NASA

Information Infrastructure Technology and Applications (IITA) Task Group

Co-Chairs: Barry Leiner, ARPA
Y.T. Chien, NSF

ARPA

Representative
Howard Frank

Alternates
Stephen L. Squires
Robert Parker

NSF

Representative
Paul Young

Alternates
Merrell Patrick
Robert G. Voigt

NASA

Representative
Lee B. Holcomb

Alternates
Paul E. Hunter
Phillip L. Milstead

DOE

Representative
John S. Cavallini

Alternates
Norman H. Kreisman
Thomas A. Kitchens

NIH

Representative
Robert L. Martino

Alternates
Michael J. Ackerman
Judith L. Vaitukaitis

NSA

Representative
George R. Cotter

Alternate
Norman S. Glick

NIST

Representative
Frederick C. Johnson

Alternate
R. J. (Jerry) Linn

VA

Representative
Daniel L. Maloney

Alternate
Rebecca L. Kelley

ED

Representative
Linda G. Roberts

Alternate
Alexis T. Poliakoff

NOAA

Representative
Thomas N. Pyke, Jr.

Alternates
William T. Turnbull
Ernest J. Daddio

EPA

Representative
Joan H. Novak

Alternate
Robin L. Dennis

AHCPR

Representative
J. Michael Fitzmaurice

Alternate
Stanley J. Phillips

OMB

Dennis E. Smith

OSTP

Lionel S. Johns
Michael R. Nelson

Notes: ARPA, NSF, NASA, and DOE hold permanent positions on the HPCCIT Executive Committee. Additional positions rotate among the other agencies.

Representatives of the Centers for Disease Control and Prevention, the Indian Health Service, and the Food and Drug Administration of the Public Health Service of the Department of Human Health and Services; the Department of Agriculture; and the National Communications System attend the HPCCIT Subcommittee meetings as observers.

References

Reports about the HPCC Program

The following are some of the reports produced by other Federal agencies, federally-chartered organizations, industry, and academia, about the Program.

- *Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure*, Computer Science and Telecommunications Board, National Research Council, Washington, DC, 1995

- *From Desktop to Teraflop: Exploiting the U.S. Lead in High Performance Computing*, NSF Blue Ribbon Panel on High Performance Computing, National Science Foundation, August 1993

- *High Performance Computing and Communications - New Program Direction Would Benefit From a More Focused Effort*, U.S. GAO Report to the Chairman, Committee on Armed Services, House of Representatives, GAO/AIMD-95-6, November 1994

- *Information Superhighway: Issues Affecting Development*, United States General Accounting Office, GAO/RCED-94-285, September 1994

- *The National Information Infrastructure: Agenda for Action*, Information Infrastructure Task Force, September 15, 1993

- *National Information Infrastructure: Industry and Government Roles*, Information Technology Association of America, July 1993

- *National Information Infrastructure Progress Report September 1993-1994*, Information Infrastructure Task Force, Department of Commerce, September 1994

- *Perspectives on the National Information Infrastructure: CSPP's Vision and Recommendations for Action*, Computer Systems Policy Project, Washington, DC, January 12, 1993

- *Promoting High-Performance Computing and Communications*, Congress of the United States Congressional Budget Office, June 1993

- *Putting the Information Infrastructure to Work: A Report of the Information Infrastructure Task Force Committee on Applications and Technology*, Department of Commerce, May 1994

- *R&D for the NII: Technical Challenges*, American Electronics Association, Computer and Business Equipment Manufacturers Association, Computing Research Association, Computer Systems Policy Project, Council on Competitiveness, Cross Industry Working Team, Educom, Gaithersburg, MD, February 28 - March 1, 1994

- *Realizing the Information Future: The Internet and Beyond*, Computer Science and Telecommunications Board, National Research Council, Washington, DC, 1994

- *Science in the National Interest*, President William J. Clinton, Vice President Albert Gore, Jr., The White House, August 1994

- *Technology for America's Economic Growth: A New Direction to Build Economic Strength*, President William J. Clinton, Vice President Albert Gore, Jr., The White House, February 22, 1993

- *Vision for a 21st Century Information Infrastructure*, Council on Competitiveness, Washington, D.C., May 1993

Information from the HPCC Program

Publications produced by the Program, by organizations to which it reports or has reported, participating agencies, and workshop convenors, include the following:

- *Enabling Technologies for Peta(FL)OPS Computing: February 22-24, 1994*, California Institute of Technology, CCSF-45, Pasadena, CA
- *The Federal High Performance Computing Program*, Executive Office of the President, Office of Science and Technology Policy (OSTP), and the FCCSET Committee on Physical, Mathematical and Engineering Sciences (CPMES), 1989
- *FCCSET Initiatives in the FY 1994 Budget*, OSTP, April 8, 1993
- *Grand Challenges 1993: High Performance Computing and Communications, The FY 1993 U.S. Research and Development Program*, Executive Office of the President, OSTP, and FCCSET/CPMES, 1992
- *High Performance Computing Act of 1991* (P.L. 194-102)
- *High Performance Computing and Communications: Toward a National Information Infrastructure*, Executive Office of the President, OSTP, and FCCSET/CPMES, 1993
- *High Performance Computing and Communications: Technology for the National Information Infrastructure*, Executive Office of the President, OSTP, and FCCSET/CPMES, 1994
- *High Performance Computing and Communications: Program Successes*, Executive Office of the President, OSTP, and the Committee on Information and Communications of the National Science and Technology Council, 1994. This is a supplement to May 10, 1994, testimony before the Science Subcommittee of the House Science, Space, and Technology Committee.
- *High Performance Computing and Communications: Supercomputing '94: The Federal HPCC Program*, November, 1994. This documents the HPCC Program panel.
- *A Research and Development Strategy for High Performance Computing*, the Federal Coordinating Council for Science, Engineering, and Technology (FCCSET), 1987
- *System Software and Tools for High Performance Computing Environments, A Report on the Findings of the Pasadena Workshop: April 14-16, 1992*, Society for Industrial and Applied Mathematics, Philadelphia, 1993
- *Workshop and Conference on Grand Challenge Applications and Software Technology: Pittsburgh, May 4-7, 1993*, Argonne National Laboratory

Many of the more recent of these reports can be accessed from the NCO Web server:
http://www.hpcc.gov

Glossary

ACTS
Advanced Communications Technology Satellite, a NASA-sponsored program. A joint NASA/ARPA collaboration will demonstrate high speed ATM/SONET transmission over the ACTS satellite, and will provide interface and operations experience in mating high speed terrestrial communications systems with high speed satellite communications systems.

AERONet
Aeronautics Network in NASA

AFS
Andrew File System

AHCPR
Agency for Health Care Policy and Research, part of the Public Health Service of the HHS

AIMS
Automated Instrumentation and Monitoring System

Algorithm
A procedure designed to solve a problem. Scientific computing programs contain algorithms.

ARPA
Advanced Research Projects Agency, part of DOD

ARPANet
Advanced Research Projects Agency Network

ASTA
Advanced Software Technology and Algorithms, a component of the HPCC Program

ATDNet
Advanced Technology Demonstration Network

ATM
Asynchronous Transfer Mode, a telecommunications technology, also known as cell switching, which is based on 53-byte cells

Backbone Network
A high capacity electronic trunk connecting lower capacity networks, for example the NSFNET backbone

Bandwidth
A measure of the capacity of a communications channel to transmit information; for example, million of bits per second or Mb/s

Benchmark
A point of reference (artifact) to compare an aspect of systems performance (for example, a well-known set of programs); also, to conduct and assess the computation (or transmission) capabilities of a system using a well-known artifact

Bit
Acronym for binary digit

Bps, or B/s
An acronym for bytes per second

bps, or b/s
An acronym for bits per second

BRHR
Basic Research and Human Resources, a component of the HPCC Program

Browser
A system that provides access to and rendering of distributed information objects located at network-based repositories; that is, services allowing a user to locate, access, and display information composed of text and still images; to animate moving images; and to play associated sound tracks. Mosaic is an example of a public-domain browser.

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

C
C programming language

C++
C++ programming language, an object-oriented descendant of the C language

CAD
Computer Aided Design

CERT
Computer Emergency Response Team

CIAC
Computer Incident Advisory Center

CIC
Committee on Information and Communications of the NSTC

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

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.

CRPC
Center for Research in Parallel Computation

CTC
Cornell Theory Center

Database
A collection of information

Database server
A system that provides database access services to users via client software systems

DCE
Distributed Computing Environment

DCRT
Division of Computer Research and Technology, part of NIH

DFS
Distributed File System

DII
Defense Information Infrastructure

DOC
Department of Commerce

DOD
Department of Defense

DOE
Department of Energy

ED
Department of Education

e-mail
electronic mail

Encryption
A reversible transformation of the encoding of information into a form that is unintelligible to unauthorized recipients in order to ensure the privacy of the information being communicated – an algorithm, the encoded information, and a key are used to encrypt (transform) and decrypt (reverse-transform) information.

EPA
Environmental Protection Agency

ESnet
Energy Sciences Network

ESS
Earth and Space Sciences

Exa-
A prefix denoting 10^{18}, or a million trillions of ... (for example, exabytes)

FDDI
Fiber Distributed Data Interface

Flops
Acronym for floating point operations per second. The term "floating point" refers to that format of numbers that 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.

FNC
Federal Networking Council

ftp
file transfer protocol

G, or Giga-
A prefix denoting 10^9, or billions of ... (for example, Gflops or gigaflops; gigabytes, gigabits)

GAO
Government Accounting Office

Gateway
A system that interconnects networks (or applications) that communicate using different protocols, and bridges their differences by transforming one protocol (message) into another

GB
An acronym for Gigabyte

Gb
An acronym for Gigabit

Gb/s
Gigabits per second

GFDL
Geophysical Fluid Dynamics Laboratory

Gflops
Gigaflops, billions of floating point operations per second

GII
Global Information Infrastructure

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

GSII
Government Services Information Infrastructure

Heterogeneous system
A distributed system that contains more than one kind of computer

HHS
Department of Health and Human Services

High performance computing
Covers the full range of advanced computing systems including workstations, networks of workstations with servers, scalable parallel systems, vector parallel systems, and more specialized systems. Scalable input/output interfaces, mass storage systems, and archival storage are components of these systems. Included also are system software and software development environments that enable users to view their workstations and the rest of their computing environments as a unified system.

HiPPI
High Performance Parallel Interface, 800 Mb/s

HPCC
High Performance Computing and Communications

HPCCI
High Performance Computing and Communications Initiative

HPCCIT
High Performance Computing, Communications, and Information Technology Subcommittee, part of the CIC

HPCS
High Performance Computing Systems, a component of the HPCC Program

HPF
High Performance Fortran

HPSS
High Performance Storage System

html
Hypertext markup language, a text composition language that allows an author to compose the contents of a document (file) that may contain text, still and moving images, and audio (multimedia objects) and have embedded links to files (multimedia objects) located on other computing systems via uniform resource locators (URLs). The ability to embed links (URLs) to multimedia objects located at distributed servers on the Internet and the ability of browsers to link to and render multimedia objects is what realizes hypertext.

http
Hypertext transport protocol, the communications protocol used by WWW servers and client systems to identify sources of information and transfer the files containing html-based multimedia objects

I/O
Acronym for Input/Output

IETF
Internet Engineering Task Force

IITA
Information Infrastructure Technology and Applications, a component of the HPCC Program

Interconnectivity
The ability of two or more computers to easily locate and communicate with each other over an infrastructure that provides the speed and clarity to accomplish a proposed task

Internet
The global collection of interconnected, multiprotocol computer networks including Federal, mid-level, private, and international networks.

Interoperability
The ability of any two computers that are interconnected to understand each other and perform mutually supportive tasks such as client/server computing

IP
Internet Protocol

ISS
Image Server Systems

JNNIE
Joint NSF-NASA Initiative on Evaluation

JPL
Jet Propulsion Laboratory

K, or Kilo-
A prefix denoting 10^3, or thousands of ... (for example, kilobits/second)

Kb/s
Kilobits per second or thousands of bits per second

Kerberos
A cryptographic-based network system that provides a mechanism for client/server computing authentication

LANL
Los Alamos National Laboratory

LLNL
Lawrence Livermore National Laboratory

M, or Mega-
A prefix denoting 10^6, or millions of ... (for example, Mbps, or megabits per second; Mflops)

MB
Megabyte

Mb
Megabit

Mb/s
Megabits per second or millions of bits per second

MBONE
Multicast backbone

Metacenter
A virtual supercomputing center composed of the networked high performance computing resources at several supercomputing centers

Mflops
Megaflops, millions of floating point operations per second.

MILNET
Military Network

MIMD
Multiple Instruction Multiple Data

MOSIS
Metal Oxide Semiconductor Implementation Service

MP
Massively Parallel

MPI
Message passing Interface

MPP
Massively parallel processor

MRAs
Metacenter Regional Alliance

Multicast
A mode of communications that allows for simultaneous distribution of information to multiple designated recipients in a single transmission

NAPs
Network Access Points, networks or switches where multiple networks interconnect to exchange traffic

NASA
National Aeronautics and Space Administration

National Challenge
A fundamental application that has broad and direct impact on the Nation's competitiveness and the well-being of its citizens, and that can benefit from the application of HPCC technology and resources

National Information Infrastructure (NII)
The integration of hardware, software, and skills that will make it easy and affordable to connect people with each other, with computers, and with a vast array of services and information resources

NCAR
National Center for Atmospheric Research

NCBI
National Center for Biotechnology Information

NCI
National Cancer Institute, part of NIH

NCO
National Coordination Office for High Performance Computing and Communications

NCRR
National Center for Research Resources, part of NIH

NCSA
National Center for Supercomputing Applications

NERSC
National Energy Research Supercomputer Center

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

NIH
National Institutes of Health, part of HHS

NIST
National Institute of Standards and Technology, part of DOC

NLM
National Library of Medicine, part of NIH

NOAA
National Oceanic and Atmospheric Administration, part of DOC

NREN
National Research and Education Network, a component of the HPCC Program

NSA
National Security Agency, part of DOD

NSF
National Science Foundation

NSFNET
NSF computer network program

NSI
NASA Science Internet

NSL
National Storage Laboratory

NSTC
National Science and Technology Council

OMB
Office of Management and Budget

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.

Optoelectronic
A descriptor of technology which combines optical and electronic components

ORNL
Oak Ridge National Laboratory

OS
Operating system

OSTP
White House Office of Science and Technology Policy

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

PC
Personal computer

Peta-
A prefix denoting 10^{15}, or a thousand trillion of ... (for example, petabits)

PICS
Partnership in Computational Sciences

PMEL
Pacific Marine Environmental Laboratory

Port
Transport a computer program from one computer system to another

Portable
Property of some computing software that allows the software to be run with little or no change on many kinds of computing systems

Prototype
The original demonstration model of what is expected to be a series of systems. Prototypes are used to prove feasibility, but often are not as efficient or well-designed as later production models.

PSC
Pittsburgh Supercomputer Center

PVM
Parallel Virtual Machine

RAID
Redundant Arrays of Inexpensive Disks

R&D
Research and development

Scalable
A system is scalable if it can be made to have more (or less) computational power by configuring it with a larger (or smaller) number of processors, amount of memory, interconnection bandwidth, input/output bandwidth, and amount of mass storage.

SDSC
San Diego Supercomputer Center

SIMD
Single Instruction Multiple Data

SONET
Synchronous Optical Network

STEP
Standard for the Exchange of Product Model Data

Superlinear
A relationship that is greater than a simple linear relationship (for example, multiple or fraction), but one that is not an exponential or geometric relationship

T, or Tera-
A prefix denoting 10^{12} or trillions of ... (for example, terabits, teraflops)

T1
Network transmission of a digital signal at 1.5 Mb/s (DS-1)

T3
Network transmission of a digital signal at 45 Mb/s (DS-3)

TCP/IP
Transmission Control Protocol/Internet Protocol

TRP
Technology Reinvestment Program

URL
Uniform Resource Locator

VA
Department of Veterans Affairs

vBNS
Very high speed Backbone Network Services

VLSI
Very large scale integration

VR
Virtual reality

WAIS
Wide Area Information Service

Web
A reference to the World Wide Web; a subset of the Internet supported by a related set of protocols, services, and software tools including browsers.

Wireless technologies
Communications technologies that utilize radio, microwave, or satellite communications channels versus wire, coaxial or optical fiber

WWW
World Wide Web

Contacts

National Coordination Office for High Performance Computing and Communications (NCO)

Suite 665
4201 Wilson Boulevard
Arlington, VA 22230
(703) 306-HPCC (4722)
FAX: (703) 306-4727
nco@hpcc.gov

Internet/World Wide Web Server:

http://www.hpcc.gov/

NCO

John C. Toole
Director
toole@hpcc.gov

Paul H. Smith, Ph.D.
Deputy Director
phsmith@hpcc.gov

Charles R. Kalina
Executive Secretary, HPCCIT Subcommittee
kalina@hpcc.gov

Sally E. Howe, Ph.D.
Assistant Director for Technical Programs
howe@hpcc.gov

Donald M. Austin, Ph.D.
Assistant Director for Planning
austin@hpcc.gov

Margaret L. Simmons
Assistant Director for Program Integration
simmons@hpcc.gov

John J. Dumbleton
ComSci Fellow
jdumble@hpcc.gov

Vicki L. Harris
Administrative Assistant
harris@hpcc.gov

AHCPR

J. Michael Fitzmaurice, Ph.D.
Director, Office of Science and Data Development
Agency for Health Care Policy and Research
2101 East Jefferson Street, Suite 604
Rockville, MD 20852
(301) 594-1483
FAX: (301) 594-2333
mfitzmau@po5.ahcpr.gov

Stanley J. Phillips
Deputy Director, Office of Science and Data Development
Agency for Health Care Policy and Research
2101 East Jefferson Street, Suite 602
Rockville, MD 20852
(301) 594-1483
FAX: (301) 594-2333
Sphillip@po5.ahcpr.gov

ARPA

Howard Frank, Ph.D.
Director, Information Technology Office
Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-2228
FAX: (703) 696-2202
HFrank@arpa.mil

Stephen L. Squires
Special Assistant to the Director
Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-2226
FAX: (703) 696-2202
Squires@arpa.mil

Robert Parker
Assistant Director, Scalable Computing Technologies, Information Technology Office
Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-2255
FAX: (703) 696-2202
RParker@arpa.mil

Barry Leiner, Ph.D.
Assistant Director, Networking and Information Management, Information Technology Office
Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-2227
FAX: (703) 696-2202
BLeiner@arpa.mil

DOE

David Nelson, Ph.D.
Associate Director, Office of Computational and Technology Research (OCTR)
Department of Energy
OCTR, ER-30
19901 Germantown Road
Germantown, MD 20874-1290
(301) 903-5800
FAX: (301) 903-7774
nelson@er.doe.gov

John S. Cavallini
Director, Mathematical, Information, and Computational Sciences (MICS) Division, Office of Computational and Technology Research (OCTR)
Department of Energy
OCTR/MICS, ER-31
19901 Germantown Road
Germantown, MD 20874-1290
(301) 903-5800
FAX: (301) 903-7774
cavallini@nersc.gov

Norman H. Kreisman
Advisor, International Technology
Department of Energy, ER-5
Mailstop 3H049-FORS
1000 Independence Avenue, S.W.
Washington, DC 20585
(202) 586-9746
FAX: (202) 586-7152
kreisman@er.doe.gov

Thomas A. Kitchens, Ph.D.
Program Manager, MICS Division, Office of Computational and Technology Research (OCTR)
Department of Energy
OCTR/MICS, ER-31
19901 Germantown Road
Germantown, MD 20874-1290
(301) 903-5152
FAX: (301) 903-7774
kitchens@er.doe.gov

Daniel A. Hitchcock, Ph.D.
Program Manager, MICS Division, Office of Computational and Technology Research (OCTR)
Department of Energy
OCTR/MICS, ER-31
19901 Germantown Road
Germantown, MD 20874-1290
(301) 903-6767
FAX: (301) 903-7774
hitchcock@er.doe.gov

ED

Linda G. Roberts, Ed. D.
Special Advisor on Educational Technology, Office of the Deputy Secretary
Department of Education
FB10
600 Independence Avenue, S.W.
Washington, DC 20202
(202) 401-1000
FAX: (202) 401-3093
linda_roberts@ed.gov

Alexis T. Poliakoff
Network Planning Staff
Department of Education
OHRA/IRG
600 Independence Avenue, S.W.
Washington, DC 20202
(202) 708-5210
FAX: (202) 260-5669
alex_poliakoff@ed.gov

EPA

Joan H. Novak
HPCC Program Manager, MD-80
Environmental Protection Agency
Research Triangle Park, NC 27711
(919) 541-4545
FAX: (919) 541-1379
novak.joan@epamail.epa.gov

Robin L. Dennis, Ph.D.
Senior Science Program Manager, MD-80
Environmental Protection Agency
Research Triangle Park, NC 27711
(919) 541-2870
FAX: (919) 541-1379
rdennis@hpcc.epa.gov

Internet use:

Bruce P. Almich
EPA Federal Networking Council Representative, MD-34
Environmental Protection Agency
Research Triangle Park, NC 27711
(919) 541-3306
FAX: (919) 541-0950
almich.bruce@epamail.epa.gov

EPA FedEx address:

79 Alexander Drive
Building 4201
Research Triangle Park, NC 27709

NASA

Lee B. Holcomb
Director, HPCC Office
Office of Aeronautics
National Aeronautics and Space Administration
Code RC
300 E Street, S.W.
Washington, DC 20546
(202) 358-2747
FAX: (202) 358-3557
l_holcomb@aeromail.hq.nasa.gov

Paul E. Hunter
HPCC Program Manager
National Aeronautics and Space Administration
Code RC
300 E Street, S.W.
Washington, DC 20546
(202) 358-4618
FAX: (202) 358-3557
p_hunter@aeromail.hq.nasa.gov

NASA (continued)

Phillip L. Milstead
Technical Manager, Scientific and Engineering Computing, Office of Aeronautics
National Aeronautics and Space Administration
Code RC
300 E Street, S.W.
Washington, DC 20546
(202) 358-4619
FAX: (202) 358-3557
p_milstead@aeromail.hq.nasa.gov

NASA express/commercial/courier address:

NASA High Performance Computing and Communications Office
Mail Code RC
Attn: Receiving & Inspection (Rear of Building)
300 E Street, S.W.
Washington, DC 20024-3210

NIH

NIH HPCC Coordinator and DCRT contact:

Robert L. Martino, Ph.D.
Chief, Computational Bioscience and Engineering Laboratory, Division of Computer Research and Technology
National Institutes of Health
9000 Rockville Pike
Building 12A, Room 2033
Bethesda, MD 20892
(301) 496-1111
FAX: (301) 402-2867
martino@alw.nih.gov

DCRT program:

David Rodbard, M.D.
Director, Division of Computer Research and Technology
National Institutes of Health
9000 Rockville Pike
Building 12A, Room 3033
Bethesda, MD 20892
(301) 496-5703
FAX: (301) 402-1754
rodbard@nih.gov

William L. Risso
Deputy Director, Division of Computer Research and Technology
National Institutes of Health
9000 Rockville Pike
Building 12A, Room 3033
Bethesda, MD 20892
(301) 496-8277
FAX: (301) 402-1754
risso@nih.gov

NCI program:

Jacob V. Maizel, Ph.D.
Biomedical Supercomputer Center, National Cancer Institute, Frederick Cancer Research and Development Center
National Institutes of Health
P.O. Box B, Building 469
Frederick, MD 21702-1201
(301) 846-5532
FAX: (301) 846-5598
jmaizel@ncifcrf.gov

Cherie Nichols
Chief, Planning Evaluation and Analysis Branch, Office of Program Operations and Planning, National Cancer Institute
National Institutes of Health
Building 31, Room 11A21
31 Center Drive, MSC 2590
Bethesda, MD 20892
(301) 496-5515
FAX: (301) 402-1225
nicholsc@od.nci.nih.gov

NCRR program:

Judith L. Vaitukaitis, M.D.
Director, National Center for Research Resources
National Institutes of Health
9000 Rockville Pike
Building 12A, Room 4011
Bethesda, MD 20892
(301) 496-5793
FAX: (301) 402-0006
vaitukaitis@nih.gov

Caroline Holloway, Ph.D.
Director, Office of Science Policy
National Center for Research Resources
National Institutes of Health
One Rockledge Center
6705 Rockledge Drive, Suite 5046
Bethesda, MD 20892-7965
(301) 435-0866
FAX: (301) 480-3654
carolineh@ep.ncrr.nih.gov

NLM program:

Donald A.B. Lindberg, M.D.
Director, National Library of Medicine
National Institutes of Health
Building 38, Room 2E17
8600 Rockville Pike
Bethesda, MD 20894
(301) 496-6221
FAX: (301) 496-4450
lindberg@nlm.nih.gov

Michael J. Ackerman, Ph.D.
Assistant Director for High Performance Computing and Communications, National Library of Medicine
National Institutes of Health
Building 38A, Room B1N30
8600 Rockville Pike
Bethesda, MD 20894
(301) 402-4100
FAX: (301) 402-4080
ackerman@nlm.nih.gov

NIST

Frederick C. Johnson, Ph.D.
Associate Director for Computing, Computing and Applied Mathematics Laboratory
National Institute of Standards and Technology
Technology Building, Room B122
Gaithersburg, MD 20899
(301) 975-2700
FAX: (301) 963-9137
fjohnson@nist.gov

R. J. (Jerry) Linn
Associate Director for Program Implementation, Computer Systems Laboratory
National Institute of Standards and Technology
Technology Building, Room B164
Gaithersburg, MD 20899
(301) 975-3624
FAX: (301) 948-1784
linnrj@nist.gov

NOAA

Thomas N. Pyke, Jr.
Director for HPCC
National Oceanic and Atmospheric Administration
Room 15300
1315 East-West Highway
Silver Spring, MD 20910
(301) 713-3573
FAX: (301) 713-4040
tpyke@hpcc.noaa.gov

William T. Turnbull
Deputy Director, HPCC
National Oceanic and Atmospheric Administration
Room 15300
1315 East-West Highway
Silver Spring, MD 20910
(301) 713-3573
FAX: (301) 713-4040
wturnbull@hpcc.noaa.gov

NOAA (continued)

Ernest Daddio
Program Officer
National Oceanic and Atmospheric Administration
Room 15400
1315 East-West Highway
Silver Spring, MD 20910
(301) 713-1262
FAX: (301) 713-1249
edaddio@hpcc.noaa.gov

Directors of high performance computing centers:

Jerry Mahlman, Ph.D.
Director, Geophysical Fluid Dynamics Laboratory
National Oceanic and Atmospheric Administration
Forrestal Campus, U.S. Route 1
P.O. Box 308
Princeton, NJ 08542
(609) 452-6502
FAX: (609) 987-5070
jm@gfdl.gov

Ron MacPherson, Ph.D.
Director, National Meteorological Center
National Oceanic and Atmospheric Administration
5200 Auth Road, Room 101
Camp Springs, MD 20746
(301) 763-8016
FAX: (301) 763-8434
ronmcp@nic.fb4.noaa.gov

Sandy MacDonald, Ph.D.
Director, Forecasting Systems Laboratory
National Oceanic and Atmospheric Administration
325 Broadway
Boulder, CO 80303
(303) 497-6378
FAX: (303) 497-6821
macdonald@fsl.noaa.gov

Internet Use:

Terrance Seldon
Manager, NOAA Network Information Center
National Oceanic and Atmospheric Administration
Federal Building 4
Suitland and Silver Hill Roads
Suitland, MD 20746
(301) 457-5165
FAX: (301) 457-5199
tseldon@nesdis.noaa.gov

NSA

George R. Cotter
Chief Scientist
National Security Agency
9800 Savage Road
Fort Meade, MD 20755-6000
(301) 688-6434
FAX: (301) 688-4980
grcotte@afterlife.ncsc.mil

Norman S. Glick
Senior Computer Scientist
National Security Agency
9800 Savage Road
Fort Meade, MD 20755-6000
(301) 688-8448
FAX: (301) 688-4980
nsglick@afterlife.ncsc.mil

NSF

Paul Young, Ph.D.
Assistant Director, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201 Wilson Boulevard, Suite 1105
Arlington, VA 22230
(703) 306-1900
FAX: (703) 306-0577
pryoung@nsf.gov

Melvyn Ciment, Ph.D.
Deputy Assistant Director, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201 Wilson Boulevard, Suite 1105
Arlington, VA 22230
(703) 306-1900
FAX: (703) 306-0577
mciment@nsf.gov

Merrell Patrick, Ph.D.
Acting Executive Officer, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201 Wilson Boulevard, Suite 1105
Arlington, VA 22230
(703) 306-1900
FAX: (703) 306-0577
mpatrick@nsf.gov

Robert G. Voigt, Ph.D.
Acting HPCC Coordinator, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201 Wilson Boulevard, Suite 1105
Arlington, VA 22230
(703) 306-1900
FAX: (703) 306-0577
rvoigt@nsf.gov

Y.T. Chien, Ph.D.
IITA Activities, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201 Wilson Boulevard, Room 1115
Arlington, VA 22230
(703) 306-1930
FAX: (703) 306-0599
ytchien@nsf.gov

VA

Daniel L. Maloney
Director, Washington Information Systems Center
Department of Veterans Affairs
8403 Colesville Road, Suite 200
Silver Spring, MD 20910
(301) 427-3700
FAX: (301) 427-3711
maloney.dan@forum.va.gov

Rebecca L. Kelley
Management Analyst, Veterans Health Administration Technology Service
Department of Veterans Affairs
8403 Colesville Road, Suite 200
Silver Spring, MD 20910
(301) 427-3750
FAX: (301) 427-3711
kelley.rebecca@forum.va.gov

OSTP

Lionel S. Johns
Associate Director for Technology, Office of Science and Technology Policy
Executive Office of the President
Old Executive Office Building, Room 423
17th Street and Pennsylvania Avenue, N.W.
Washington, DC 20500
(202) 456-6046
FAX: (202) 456-6023
ljohns@ostp.eop.gov

Michael R. Nelson, Ph.D.
Special Assistant for Information Technology Office of Science and Technology Policy
Executive Office of the President
Old Executive Office Building, Room 423
17th Street and Pennsylvania Avenue, N.W
Washington, DC 20500
(202) 456-6046
FAX: (202) 456-6023
mnelson@ostp.eop.gov

OSTP (continued)

Paul A. Regeon
CIC Executive Secretary, Office of Science and Technology Policy
Executive Office of the President
Old Executive Office Building, Room 423
17th Street and Pennsylvania Avenue, N.W.
Washington, D.C. 20500
(202) 456-6042
FAX: (202) 456-6023
pregeon@ostp.eop.gov

Editorial Group for High Performance Computing and Communications FY 1996

Editor

Sally E. Howe
National Coordination Office

Writing Group

Thomas A. Kitchens, DOE, Chair

Donald M. Austin, NCO
J. Michael Fitzmaurice, AHCPR
Norman S. Glick, NSA
Frederick C. Johnson, NIST
Rebecca L. Kelley, VA
Robert L. Martino, NIH
Phillip L. Milstead, NASA
Joan H. Novak, EPA
Alexis T. Poliakoff, ED
William T. Turnbull, NOAA
Glenn Ricart, ARPA
Robert G. Voigt, NSF

Acknowledgments

Many people contributed to this book, and we thank them for their efforts. This especially includes the researchers who provided descriptions of their work. From the Cognitive Science Branch at the National Library of Medicine (NLM), we thank Branch Chief Alexa McCray, Jeff Bridgers, and Zoe Stavri for preparing the index and for managing the NCO servers, and we thank Jeff Bridgers for preparing the WWW and gopher versions of this book. From the NLM Audiovisual Program Development Branch, we thank Joe Fitzgerald for the cover design, and both Troy Hill and Joe Fitzgerald for their artistic contributions. We thank the staff of the National Coordination Office for their assistance in distributing this report.

Index