Archived Material

Historical Purposes Only

Federal HPCC Program testimony - HPCC Program Successes

Testimony before the Subcommittee on Science of the House Committee on Science, Space, and Technology (5/10/94)

HPCC Program Successes
Supplement to testimony

Supplement to May 10, 1994 Testimony before the Science Subcommittee of the House Science, Space, and Technology Committee by representatives of the HPCC Program

At the May 10 hearing, Rep. Boucher, Chairman of the Science Subcommittee, invited HPCC Program representatives to supplement their testimony before the Subcommittee with a report describing the major successes of the HPCC Program to date. The representatives thank Mr. Boucher for his invitation and submit the following report, which describes overall Program accomplishments and individual agency successes.

The document entitled, "High Performance Computing and Communications: Technology for the National Information Infrastructure" (commonly referred to as the Blue Book) is a Supplement to the President's Fiscal Year 1995 Budget that itemizes six major Program successes. These accomplishments, brought up to date, are as follows:

1. More than a dozen high performance computing research centers are in operation nationwide. More than 100 scalable high performance systems are in operation at these centers. These include large scale parallel systems, vector parallel systems, hybrid systems, workstations and workstation clusters, and networked heterogeneous systems. The largest of these systems now provides almost 150 gigaflops (billions of floating point operations per second) performance on benchmarks. The HPCC Program's 1996 goal of developing high performance systems capable of sustained teraflops (trillions of floating point operations per second) performance is well on the way to being met.

2. The Internet, seeded by HPCC funding, continues to experience phenomenal growth in size, speed, and number of users. This network of networks (with the HPCC-funded networks providing the initial core) enables researchers to access high performance computers and advanced scientific instruments easily, and has begun to allow independence of geographical location.

3. One illustration of the global reach of HPCC technology is that the Internet now extends across the country and throughout much of the world. The Internet links nearly 3 million computers, more than 18,000 networks in the U.S. and more than 13,000 outside the U.S., and 1,200 4-year colleges and universities, 100 community colleges, 1,000 high schools, and 300 academic libraries in the U.S.

4. More than half a dozen gigabit testbeds conduct research in high capacity networks. These network testbeds connect 24 sites, including many of the high performance computing research centers. Seven Federal agencies, 18 telecommunications carriers, 12 universities, and two state high performance computer centers participate. The testbeds develop technologies to handle the NII's (National Information Infrastructure) increased demand for computer communications, along with greater accessibility, interoperability, and security. The Program goal of demonstrating gigabit (billions of bits) per second transmission speeds by 1996 is well on the way to being met.

5. Teams of researchers are using scalable systems to discover new knowledge and demonstrate new capabilities that were not possible with earlier technologies. These researchers are addressing the Grand Challenges, fundamental problems in science and engineering with broad economic and scientific impact whose solutions can be advanced by applying high performance computing techniques and resources. Many of these challenges are associated with HPCC agency missions. Agencies are increasingly working with U.S. industry to use their Grand Challenge applications software and develop new software that will improve commercial and industrial competitiveness.

   Users of the new scalable computing systems with their high performance on large problems and larger memories are able to address more complex, more realistic problems. We are beginning to understand our world better and to improve our lives:

   - Improved modeling of the Earth and its atmosphere has sharpened our ability to predict the movement and characteristics of storms and other forms of severe weather. With improved forecasts, there will be more lives saved and an overall positive economic impact due to reduced property loss and evacuation of smaller endangered areas along the coast and elsewhere.

   - New air and water quality models are enabling improved environmental decision making.

   - Improvements in the design and manufacture of goods are yielding better products. Both the production processes and products such as cars and airplanes are becoming more energy efficient.

   - We are learning more about how the human body functions and are improving our ability to diagnose and treat diseases.

6. The thousands of researchers who develop fundamental HPCC and NII technologies and applications form the vanguard that hardware and software vendors rely upon to promote the use of high performance computers and communications throughout the U.S. economy. Hundreds of teachers and thousands of students access HPCC resources, and the Program conducts hundreds of training events for thousands of trainees. Dozens of small systems have been installed at colleges and universities. The goal of these efforts is to train a nation of knowledgeable users and thereby fully incorporate HPCC and NII technologies and applications into the U.S. economy.

In addition to these multi-agency successes, participating agencies have had notable successes both individually and jointly in developing and applying HPCC technologies. We asked each agency to provide some examples:

1. Advanced Research Projects Agency (ARPA)

Overview

During the past decade, ARPA investments in information technologies enabled fundamental advances that have been a major factor in the continued leadership of the U.S. in this key area. Previous ARPA investments in information technology tended to focus on relatively narrow technology sectors such as timesharing, computer graphics, computer networks, personal computing, and associated systems software and early applications. Each of these has had a fundamental effect on defense, commerce, and education. ARPA investments of the past decade have built upon and jointly leveraged all of these, and have transformed them into a new information technology that is also transforming society.

These results have recently become visible in the form of the Federal HPCC Program and its extension to include the technologies and pilot applications for the NII. Dramatic advances in these technologies have produced a new scalable information technology base that can be configured over a wide range of computing and networking performance, and distributed in the form of national- and global- scale systems.

New scalable technologies provide a combination of higher performance for the larger systems, with greater cost effectiveness for the smaller systems. In addition to the new technologies, a new generation of people who understand the technologies, as well as the new applications, has begun to emerge. The scalable nature of the technologies, coupled with the emergence of a pilot information infrastructure using Internet technologies, enables the accelerated diffusion of ideas, methods, technologies, and human resources.

Beyond these accomplishments, the Internet (used as a pilot information infrastructure) is transforming the way in which the information technology research community operates, the way it collaborates with industry, the way the scientific and engineering communities work with each other, and the way in which even broader sectors operate. To see this, one need only look at the explosive growth of the Internet, the number of users, computers, networks, and information servers, and the distribution of these systems by geography, by discipline, or by sector. These technologies and their application have considerably influenced Vice President Gore's National Performance Review, and are now forming the foundation for dramatically transforming the way the Federal Government operates into a new form of "electronic government."

Below we highlight successes in five major technology areas: (a) scalable computing systems, (b) microsystems, (c) scalable software, (d) networking, and (e) national- scale information enterprises.

1.a. Scalable Computing Systems

Over the past decade, ARPA invested in the development of a wide range of scalable computer architectures to explore the use of parallelism to increase performance. The technology innovations include processor architectures, cache and memory system architectures, interconnects, disk storage systems, and high performance communications network interfaces.

More specifically, ARPA pioneered the development of RISC-based computer systems, spanning workstations, shared memory multiprocessors, and large scale parallel architectures, including the associated operating systems and compiler technologies. The Sun SPARC and MIPS instruction set architectures can be viewed as direct commercializations of ARPA-funded research projects, and other architectures such as the IBM PowerPC are derived from these. The National Science Foundation (NSF) also played a key role in supporting the fundamental research that led to RISC concepts and systems.

ARPA provided support for demonstrating RAID (Redundant Arrays of Inexpensive Disks) systems with hundreds of small formfactor disk drives. This is now the architecture of choice for high performance, high capacity, high reliability I/O (input/output) systems, and has spawned commercial developments at over 20 companies, including Compaq, DEC, IBM, NCR, StorageTek, and Sun. NSF also provided research support for the early proof of the RAID concept.

A critical HPCC Program milestone was recently achieved when ARPA demonstrated 100 gigaops class systems as part of cooperative projects including Cray Research (T3D), IBM (SP1 and SP2), Intel (Paragon), Kendall Square Research (KSR 1), and Thinking Machines (CM-5). Many of these systems were first evaluated at NSF Supercomputer centers. ARPA worked closely with vendors to accelerate developments in scalable architectures. For example, through a cooperative agreement with ARPA, the Cray T3D was delivered in 26 months on schedule and within budget. An early beta version of the machine was made available at NSF's Pittsburgh Supercomputer center through joint NSF and ARPA sponsorship.

1.b. Microsystems

The design technology developed and validated within the microsystems program has created the advanced architectural building blocks for scalable parallel architectures. ARPA fueled the development of the computer-aided design software industry by supporting early and fundamental university-based research in integrated circuit layout, simulation, and synthesis tools. Many of the existing CAD software vendors can trace their roots to design software efforts that came from this program. Companies such as Cadence and Synopsis have built directly upon CAD software originally developed under ARPA sponsorship.

ARPA funded the technology developments that made it possible to demonstrate the first "virtual factory" integrated circuit process simulation shadowed by a real semiconductor fabrication line. State-of-the-art integrated circuit fabrication lines cost as much as $1 billion, making it imperative to validate the process prior to final implementation. The virtual prototyping techniques developed under ARPA sponsorship directly addressed this need by demonstrating approaches to optimize and tune the process before committing it to fabrication lines.

Scalable computer backplane communications technology provides the foundation for a majority of university and commercial scalable architectures. Work has been funded at Caltech, MIT, the University of Washington in Seattle, and others. Caltech's achievement in mesh routing has led to its adoption in various commercial systems, and the same components are now being exploited in high bandwidth, low latency, low cost local area networks.

ARPA funded the development of MOSIS (metal oxide semiconductor implementation service), the first network-based rapid implementation service for VLSI (very large scale integration) systems and an early prototype for NII-based design and commercial services. NSF worked with ARPA to make this service available to educators and students throughout the U.S., as well as to NSF's research community. This has led to a vibrant and well trained engineering community, able to capitalize on new technologies and to retain U.S. leadership. In addition, ARPA funded the development of FAST, the first-of-its-kind network-based broker for electronic and other catalog components. Time to order and acquire parts has been reduced by a factor of five to 10. FAST is being evaluated by the Defense Logistics Agency for possible operational use within DOD.

1.c. Scalable Software

ARPA's scalable software investments have led to many advances in the underlying technologies that support large, complex, and distributed applications. ARPA funded the development of microkernel operating system technology, which allows applications-level customization of operating system services for greater performance and flexibility. The microkernel software architecture forms the intellectual basis of the WindowsNT Operating System from Microsoft. The ARPA- developed Mach OS is the operating system of two major scalable computing vendors (DEC and Hewlett-Packard), and forms the foundation for continued innovations in OS.

ARPA developed the Computer Emergency Response Team (CERT) in response to the first large scale attack on Internet hosts. CERT has become the world leader in the fight against computer intruders. In addition to being the focus for intrusion detection and response, CERT has become a locus for creation and dissemination of protective software, tools for system security administration, and establishment of "best practice" for site computer security. ARPA and the National Institute of Standards and Technology (NIST) later collaborated to form the Forum of Incident Response and Security Teams (CERT-FIRST) to provide a government interface for CERT activities and extend the capabilities among Federal agencies.

The widespread deployment of the Internet has required significant advances in protection and trust mechanisms. ARPA has led in the development and deployment of security and privacy technology. ARPA developed privacy-enhanced mail for secure, trusted electronic mail interchange on open networks. ARPA, in cooperation with NIST and the National Security Agency (NSA), also developed the technology for Trusted Operating Systems. Through close coordination with developers of commercial versions of Mach, "Trusted Mach" or TMach, will inherit both performance enhancements and functional enhancements from its commercial variant. Another ARPA developed system, the ISIS distributed interprocess communication system, decreases the complexity of developing distributed applications by effectively eliminating most common distributed systems software failure modes. The strategies it employs are so effective that it has been commercialized by its developers and is being used for critical real time applications such as the American Stock Exchange trading system and the French Air Traffic Control System.

Acceptance and integration of Internet technologies in the commercial and military environments requires the ability to protect organizational systems. ARPA has supported the development of a suite of protective technologies for computer systems including public domain network firewall technology, controlled execution Unix, and trusted auditing tools. These tools enable systems administrators to increase the resistance of their systems to intrusion.

Libraries of high performance software executable over a range of heterogeneous machines are major enablers of scalable technologies. ARPA funded the development of SCALAPACK (the scalable version of the linear algebra package LAPACK), one of the most widely used libraries of mathematical software for high performance systems. This was a proof of concept demonstration that performance tuned mathematical software libraries enable applications to be written in machine independent fashion while achieving high efficiency on various high performance computer architectures. NSF supported the development of select portions of SCALAPACK.

ARPA also supported the development of new and extended programming languages for parallel machines, including FORTRAN-D, High Performance Fortran (HPF), Split-C, and data parallel and object parallel C++, as well as the associated advanced compilation technology for modern RISC and signal processing microprocessors. Compilers and compiler technology developed at Rice University under ARPA sponsorship have been transferred to IBM, Intel, Motorola, Silicon Graphics, Tera, and Texas Instruments. NSF also supports work in scientific compilers and languages through its Science and Technology center at Caltech, Rice University, and Syracuse University.

1.d. Networking

Recognizing early the importance of computer communications, ARPA has invested in a variety of network developments over the past two decades. Initial funding was for ARPANET, the first packet-switched network and the predecessor of a broad industry of computer networks. ARPA funded the development of the Internet and its associated TCP/IP network protocol and, with the collective leadership of NSF and DOE, saw its evolution into the globe-spanning Internet. The Internet today includes more than 30,000 networks and 2,500,000 computers and enables communications amongst more than 20 million people around the world. More than half of today's Internet-connected organizations are commercial, and many of the Internet communication services are provided by commercial vendors. The development of TCP/IP was responsible for the creation of the network router industry that in 1994 is expected to have over $3 billion in sales. The TCP/IP protocol was recently adopted by Microsoft as the basis for all of its future networking products, which should result in complete penetration of the computer market. In addition, the ARPANET and its successor Internet have done more for cross discipline collaboration than any other program, activity, or incentive before or since.

ARPA developed the SURAN (Survivable Radio Network) family of networking protocols that permits reliable delivery of data in a hostile environment. This seminal work is a direct predecessor to the Cellular Digital Packet Data (CDPD) standards being created and deployed today. ARPA pioneered the development of packet-switched voice protocols in the 1970s and the packet video protocols in the 1980s, which provided the early proof of concept that packet technology would work for these media. This is the basis for the Asynchronous Transfer Mode (ATM) protocols used within the telephone system to integrate voice, video, and data. ARPA-sponsored contractors, such as Fore Systems, are among the world's leaders in supplying ATM switching technology for local area networks. In collaboration with the Department of Energy's (DOE) Lawrence Berkeley Laboratory, ARPA developed advanced video, audio, and shared whiteboard collaboration capabilities in conjunction with developing a multicast capability layered over the Internet. This multimedia conferencing capability is used regularly to disseminate information from conferences and other meetings as well as support informal meetings over the Internet, thereby further enhancing cross discipline cooperation. In a strong collaboration with NSF, ARPA deployed five gigabit-class networking testbeds (described in section 2.f. below) and a sixth ARPA-only testbed. All six focus on network research as well as high performance applications. These testbeds have fostered interoperability among telecommunications service providers and have accelerated the availability of high bandwidth networking services to commercial customers. In addition, they demonstrated the need for gigabit-class or better network infrastructure for certain high performance applications such as three- dimensional radiation treatment planning and visualization.

1.e. National-Scale Information Enterprises

ARPA's leadership in network-based technologies and services yielded numerous early examples of NII capabilities within the ARPANET and later the Internet. ARPA pioneered the development, deployment, and operation of electronic commerce capabilities on networks for more than 10 years. These highly successful efforts include MOSIS, the silicon broker, and FAST, the semi-automated sourcing and buying system. Recently, FAST initiated a commercial pilot project with Citibank to interface the brokering system to a payment system, thus automating billing, credit authorization, and payment processing. FAST currently supports 135 customers and over 3,000 vendors.

In addition, ARPA has led the way in developing network-based tools for collaboration. These include the MBONE (multicast backbone) infrastructure supporting shared video over the Internet, as well as associated tools for shared design spaces.

ARPA has taken a pioneering position in digital libraries technologies by establishing early testbeds and working closely with the Library of Congress to prototype new publishing and recordation mechanisms suitable for the electronic world envisioned for the NII. In addition, ARPA's programs in scalable I/O, mass storage, and image processing systems all provide crucial technology that will underlie digital libraries.

As a specific recent accomplishment, ARPA supported MIT in developing the White House electronic document distribution system over the Internet. Over 2,000 documents have been distributed since January 20, 1993, to 4,500 direct subscribers and an estimated 120,000 end users. It is estimated that over 4 million page equivalents have been distributed in this fashion. This technology is a prototype of advanced subscription and dissemination capabilities needed for the NII.

---

2. National Science Foundation (NSF)

2.a. Supercomputer centers

The four NSF Supercomputer centers have become recognized leaders in the national HPCC effort. By providing reliable high performance computing services, they influence a broad range of scientific disciplines both within the academic community and across a wide range of industries, from computer vendors to producers of products for the general public. In this latter context, the centers pioneered partnerships with the private sector to apply HPCC technologies and practices to specific design and manufacturing problems. Two illustrative examples are now described:

2.a.1. Cornell Theory center - Merck Research Laboratory Collaboration:

The Cornell Theory center (CTC), one of the four Supercomputer centers, and Merck Research Laboratories (MRL) recently announced that the company has joined CTC's Corporate Research Institute (CRI) to explore scalable parallel computing technologies. A division of Merck & Co., Inc., the world's largest prescription drug company, MRL will be applying CTC's parallel computing resources to computer-aided drug discovery and structure determination. An interdisciplinary research center at Cornell University in Ithaca, NY, CRI was established to manage the exchange of knowledge and transfer of technology from CTC and its affiliated community to industry.

Dr. James B. Doherty, Executive Director of Biophysical Chemistry and Chemical and Biological Systems at Merck stated, "We are excited by the opportunity to interact with the scientists of the Cornell Theory center as we explore the potential of scalable parallel computer systems. We believe such systems will play an important role in meeting the ever-increasing computational demands of computer-aided drug discovery. More importantly, we believe these technologies will be useful in our efforts to develop important new medicines."

Under the partnership agreement, Merck scientists have access to CTC's computational resources, most notably the IBM Scalable POWERparallel system (SP1), as well as to CTC consulting and training programs. CTC acquired the first SP1 in April 1993 with support from New York State. The SP2, a 512 processor system that will be capable of performing at 125 gigaflops (1/8 teraflops) speed, is scheduled for installation later this year. CTC has a strategic relationship with IBM to apply SP technologies to problems in disciplines such as fluid dynamics and rational drug design.

In addition to accessing CTC computing resources via network connections, MRL researchers and staff participate in CTC education and training programs. These programs range from an introduction to parallel computing to highly technical training in specific applications with hands-on experience.

Other CRI partners include Abbott Laboratories, BIOSym Technologies, Corning, and Xerox. CRI offers a rich variety of opportunities for its partners: a robust high performance computing environment, experimentation with a variety of highly parallel scalable systems, advanced algorithm research, and a strong record in interdisciplinary approaches to scientific and engineering problems critical to industry.

2.a.2. Mosaic

The second example of NSF Supercomputer center activities is the freely distributed Mosaic software suite developed by the National center for Supercomputer Applications (NCSA). Mosaic enables dynamic browsing, retrieval, and display of multimedia data (text, images, sounds, movies) from the World-Wide Web (WWW), an information retrieval system on the Internet. Mosaic uses hypertext to allow documents to contain words or images as direct links to additional information. Clicking on linked words or images retrieves and displays new files and information without having to know the location or network address of that information. Retrieved hypertext files in turn contain links to more information, creating a global web of information involving over 2,300 databases. In less than a year, Mosaic has become the fastest growing Internet-based application, with over 30,000 copies being delivered each month.

Recently, Spyglass Inc. and NCSA have entered into an agreement that will put Mosaic on the desktops of millions of people. The agreement gives Spyglass, which was formed in 1990 and has commercialized other NCSA technologies, full rights to enhance, commercialize, and broadly relicense Mosaic. Spyglass is making a multimillion dollar commitment to NCSA and will focus initially on developing a commercially enhanced version of Mosaic that other companies will incorporate in the products they distribute to their customers.

"We're committed to evolving Mosaic so it becomes a robust, commercial tool with complete documentation, technical support, and advanced features," explained Tim Krauskopf, cofounder of Spyglass. "We'll be collaborating with NCSA and other key partners to create new tools and establish standards that will help organizations build robust World-Wide Web information servers quickly and inexpensively."

Spyglass has redesigned Mosaic so it will be a more robust and full featured tool. Mosaic enhancements available from Spyglass include improved installation, better memory management, increased performance, new forms capabilities, online hypertext-based help, support for a proxy gateway, and user interface improvements such as support for multiple windows. Future versions will include enhanced security and authentication, which will enable credit card and other business transactions to take place on the Internet; filters that will enable documents from popular document readers to be read seamlessly by Mosaic; and integration with emerging editing and document management tools.

A number of businesses are already using Mosaic and WWW to publish magazines, deliver goods and services, provide technical support to customers, and conduct other forms of business electronically. For example, Mosaic and WWW are part of the recently announced $12 million CommerceNet project, a public- and private- sector-backed initiative exploring ways to conduct commerce over the Internet and other networks.

2.b. Grand Challenge Applications Groups

NSF, with ARPA support, funds 16 Grand Challenge Applications Groups pursuing solutions to scientific and engineering problems whose computational demands go beyond those provided by today's most powerful computing systems. Some highlights are summarized below.

2.b.1. Geophysical and Astrophysical Fluid Dynamics

One part of the NSF Grand Challenge Application Group effort in Geophysical and Astrophysical Fluid Dynamics at the University of Colorado and at the University of

Minnesota is simulating turbulence in a computational cube of the finest resolution possible in order to understand the fundamental mechanisms driving turbulent flows. One goal is completing a series of calculations on a computational grid of one billion processors, that is, in a cube with a thousand computational zones in each of the three dimensions. In order to be practical, such a calculation requires on the order of 28 gigabytes of addressable memory and many hours of dedicated system time executing in the range of several gigaflops.

Professor Paul Woodward, a member of the team from the University of Minnesota, presented the requirements to Silicon Graphics (SGI) as a possible demonstration of that company's high performance computing capability. The result was a collaboration in which SGI agreed to put together a network of high end workstations configured specifically to support the execution of Woodward's piecewise parabolic method hydrodynamics software on a grid of one billion computational zones. For the Grand Challenge team, questions of compelling scientific importance could be explored; for its part, SGI could demonstrate that a cluster of their workstations could attack one of today's largest computational problems.

Over a three week period, SGI configured 16 of its Challenge XL systems with 20 processors each, and in collaboration with the Minnesota team, addressed specific software issues that had to be resolved in order to conduct a successful simulation. The periodic boundary conditions used in the simulation necessitated interconnecting the systems in a 2x2x4 three-dimensional toroidal array using FDDI (fiber distributed data interface) network technology. Thus, the experiment represents one of the few times that the scientific application requirements drove computing system design; normally, a manufacturer provides a system and users must adapt their algorithms to that system's design characteristics.

The numerical simulation was carried out over the space of a week using the full system at 30 percent of peak efficiency including communication time overhead. This resulted in a sustained computation rate of 4.9 gigaflops. It will take much longer to analyze the output and describe the findings. To facilitate this process the results were written on 350 4-gigabyte tapes organized to support the analysis as well as restarting the calculations from specific points if necessary.

This billion-zone turbulence computation demonstrated that workstation computing involving hundreds of processors must be taken seriously by the high performance computing community. It was made possible by the arrival of a new generation of workstations, represented by the SGI Challenge XL Servers used here, which have very fast interconnections of tens of powerful new microprocessors with multiprocessing software tools to match. Simultaneously, "off the shelf" high speed interconnection networks now can produce highly effective arrangements of tens of these workstations with relatively little effort or cost. Additionally, high speed disk and tape peripherals can now be attached to this new generation of workstations. The effort also demonstrated the power of some of the extensions to the Fortran programming language that the Minnesota team has developed and used in their simulations. SGI is investigating the use of these extensions in their product line as well as their possible impact on standards efforts such as the High Performance Fortran Forum, which focuses on developing a portable Fortran that supports parallelism.

2.b.2. Understanding Human Joint Mechanics through Advanced Computational Models

At Rensselaer Polytechnic Institute, Professor Rober Spilker leads a group of engineers, mathematicians, and computer scientists in an effort to develop realistic computational models of human skeletal joints using actual anatomic geometries and multiphasic properties of joint tissues. Early results focused on the role of cartilage in joint movement. In particular, accurate models for cartilage deformation and for the observed time-dependent and velocity-dependent friction coefficient of cartilage have been developed. These models will be used to better understand the interaction between cartilage and the interstitial fluid in joint motion.

2.b.3. High Performance Computing for Learning

Researchers at MIT under the direction of Professor Robert Berwick seek to understand how a person or a machine can learn visual, motor, language, and other skills, thereby laying the foundation for a future generation of high performance computers that include powerful brain-like functions. The approach is to combine computational models with biologically driven learning models in order to design VLSI chips that enhance machine learning. For example, a novel memory- intensive VLSI chip has been demonstrated as a PC "coprocessor" that delivers high performance computer performance on the face recognition problem.

2.b.4. Earthquake Ground Motion Modeling in Large Basins

Professor Jacob Bielak leads a team of researchers centered at Carnegie Mellon University who are developing the ability to predict the ground motion in large basins during strong earthquakes, with particular attention to better understanding the Greater Los Angeles Basin. To this end the team is building tools to help the application specialists solve mechanics problems on parallel computers. One such tool is Archimedes, a special purpose compiler that supports the mapping onto parallel systems of unstructured grid computations required in basin modeling. Such discipline-specific tools demonstrate that non-specialists can make effective use of parallel systems without becoming intimately involved with the idiosyncrasies of a particular architecture.

2.c. Science and Technology centers

NSF supports four Science and Technology centers whose missions are directly related to the HPCC Program: The center for Cognitive Science at the University of Pennsylvania, the center for Computer Graphics and Scientific Visualization at the University of Utah, the center for Discrete Mathematics and Theoretical Computer Science at Rutgers University, and the center for Research in Parallel Computation (CRPC) at Rice University. These centers support a broad program from basic research to technology transfer to educational programs aimed at exposing K-12 students to the excitement of science.

CRPC conducts basic research in parallel languages, compilers, and programming. It is a leader in the High Performance Fortran language specification effort that is supported by most parallel systems vendors. It has formed strong ties with industrial and scientific applications in order to provide feedback on the relevance of the research developments. For example, the Geosciences Parallel Computation Project involves participants from a number of universities and the petroleum industry focused on using parallel computation to simulate enhanced oil recovery methods. Study areas include flow in porous media, seismic analysis, and optimal well placement in the context of developing advanced tools for parallel scientific programming.

2.d. Education and Training

NSF supports a variety of efforts aimed at K-12 through graduate education and postgraduate and industrial training. The Supercomputer centers and the Science and Technology centers play a major role in this activity. The SuperQuest Program provides a specific example.

SuperQuest is a nationwide summer program for high school students and their teachers. Each SuperQuest entry team has teacher-coaches and three or four students who work on computational science projects that are judged in a competition among all submitted projects. The student finalists and advisers attend a three week educational and training seminar that involves "hands-on" experience with a variety of high performance vector and parallel computers, including executing applications software related to their particular science projects. These seminars are held at designated SuperQuest sites, each of which hosts four winning SuperQuest teams.

Students and teachers attend classes in subjects ranging from advanced mathematics to English, including training in effectively presenting results to an audience. They also benefit from associating with academicians, industrial researchers, mentors, and consultants at the sites. The teams return home to continue their computational research, where they use advanced workstations awarded to their schools by equipment vendors through the SuperQuest program. Each team also receives one year of Internet connectivity so it may continue to access the computers and communicate with the researchers at the SuperQuest sites.

In 1993 the SuperQuest sites were CTC, NCSA, Reed College, and the University of Alabama at Huntsville; in 1994 the University of North Carolina, Chapel Hill, will be added. Similar programs have been initiated in Arizona, California, Minnesota, and New Mexico.

2.e. Institutional Infrastructure -- Minority Institutions Program

NSF provides grants to computer and information science and engineering departments in minority institutions (usually African American or Hispanic) to aid in research, education, and outreach activities. The program supports equipment acquisition and Internet access in order to lower barriers to full research participation by minority institutions. The education component requires that the university increase minority participation in computer science and engineering by upgrading faculty qualifications and undergraduate research participation. The outreach component has two parts: (1) active recruitment of entering freshmen through visits to high schools and K-12 participation in science projects, and (2) cooperative agreements between minority and majority institutions in order to provide opportunities for graduate work for students from the minority institutions.

For example, the University of North Carolina A&T, an historically black university, supports both research and education and is addressing the shortage of African Americans in computer science and engineering through its Communications Methods Research Laboratory, Signal Processing Laboratory, and VLSI Design Laboratory. The school is linked to Duke University, Michigan State University, Stanford University, and the University of Michigan as preferential graduate schools for NC A&T graduates. Under this program, NC A&T has provided 30 minority students to the Ph.D. pipeline and strengthened its research programs.

2.f. Gigabit Testbeds

NSF and ARPA jointly fund five gigabit network testbeds led by the Corporation for National Research Initiatives (CNRI) for research on very high speed communications networks and Grand Challenge applications. Testbed collaborators come from universities, national laboratories, high performance computer centers, and major industrial organizations that greatly leverage the Federal investment. The testbeds include three wide area networks, AURORA, BLANCA, and CASA, and two metropolitan area networks, NECTAR and VISTANet. Researchers are exploring issues important for developing and using more widely deployed gigabit networks, including network architectures; interfaces between the network and high performance computers, graphics engines, and workstations; high speed and real time protocols; distributed computing; shared memory; stream-lined, network-friendly operating systems; traffic and network control; and application software environments. The testbeds are used to explore Grand Challenge applications that cannot be investigated without the combined use of resources linked by gigabit networks, including high performance computers, graphics engines, high performance workstations, and sensors as varied as MRI scanners and astronomy observatories. These applications include climate modeling that combines ocean and atmospheric models, medical imaging, multimedia digital libraries, multiple remote visualization and control of simulations, radiation treatment planning, and radio astronomy imaging.

---

3. Department of Energy (DOE)

DOE and its predecessor agencies and commissions have a long history in high performance computing, both in defense programs and in unclassified energy research programs. The

National Energy Research Supercomputer center (NERSC) was established in 1975 to provide for fusion energy's high performance computing needs, and was made available to all energy researchers in 1984. DOE also established the predecessor of the Energy Sciences Network (ESnet) to provide access by NERSC users, the majority of whom were located off site. ESnet is now an integral part of the Internet.

DOE was one of the founders of the HPCC Program. It participates in all components of the Program but emphasizes the computational applications. The HPCC Program enabled DOE to establish two new HPCC Research centers, dramatically improve ESnet, undertake eight Grand Challenge Projects, move its important applications to parallel processing through advanced parallel algorithm development, and initiate educational programs in computational science. Many of the most significant DOE accomplishments are described in HPCC reports to Congress. Several more recent accomplishments are described below.

3.a. Sandia/Intel Shatter Computing Speed Record

A team from Sandia National Laboratories and Intel achieved a world record 143.4 double-precision gigaflops on the Massively Parallel LINPACK benchmark. The previous record of 124.5 gigaflops was set in August 1993 by Fujitsu's Numerical Wind Tunnel, a one-of-a-kind system owned by the Japanese National Aerospace Laboratory. A key technology enabling this performance was SUNMOS, the Sandia UNM (University of New Mexico) Operating System. SUNMOS uses only two percent of each processor's memory and operates at peak communication bandwidth.

"Sandia's achievement certainly validates the soundness of Intel's scalable architecture and the leading role that Sandia National Laboratories plays in pushing the envelope of high performance computing," said Ed Masi, president of Intel's Supercomputer Systems Division and vice president of Intel Corporation. "The LINPACK record represents a win not just for Sandia and Intel, but for the U.S. Advanced Research Projects Agency, the Department of Energy and the entire U.S. HPCC community."

3.b. Technical Conferencing Capabilities

With the increasing size and the geographical diversity of scientific collaborations, the role of high speed networking in supporting collaborative work must be expanded. Under the HPCC Program, computer scientists at the Lawrence Berkeley Laboratory work with scientists at Xerox Palo Alto Research center to develop tools for network-based collaboration and conferencing.

Such collaboration has three basic needs: an audio tool, a video tool, and a shared workspace or "whiteboard." In addition, network protocols needed a number of modifications in order to support video conferencing. In order to provide good voice transmission, a time stamp was added to each packet to compensate for the variable time delay in network communication. A multicast protocol was developed in order to minimize the bandwidth requirements of video transmission. This protocol ensures that only one copy of any video conference is transmitted on a single network link. When the video signal reaches a point where it must follow two or more separate links, multiple copies of the conference are generated by the multicast routers.

Used together, these tools have become an invaluable part of ESnet and the Internet. Beginning with the networking and computer science communities, use of conferencing tools has now spread to many DOE science programs and beyond. On April 15, 1994, the first Multicast Symposium in the medical community was held at the Long Beach Regional Medical Education center at the Long Beach VA Medical center in California. This Pancreatic Islet Micro-circulation Symposium brought many experts in the field together to establish and summarize the current state of knowledge of islet microcirculation. Most of the conference was broadcast directly from the center but one presenter, Susan Bonner Weir of Harvard, was unable to attend in person and gave her presentation remotely from the MIT Laboratory of Computer Science. The symposium's success demonstrated the important role that video conferencing tools can play in the medical community.

3.c. Simulation of Grinding Process in Virtual Reality

Researchers in Argonne's Mathematics and Computer Science Division are using the CAVE (Cave Automatic Virtual Environment) to gain new insights into the grinding process, an essential process in manufacturing. This is collaborative work with Argonne's Reactor Engineering Division and with industrial scientists at Surface Finishes Co.

Grinding a component involves several controllable parameters such as materials properties, rotational speed, surface roughness, and force/interface between the block and the grinding wheel. Virtual reality enables the researcher to better understand the effects of these parameters on the grinding process. Using CAVE tools in conjunction with Argonne's large scale parallel IBM SP1 computer, researchers observe temperatures and stresses within the block and wheel. Sounds of the surface motions are played through the CAVE audio system, ablated material is ejected as small particles and displayed as sparks, and the researchers have tactile feel of the dials and knobs that control the process.

Argonne and its collaborators envision several benefits from the CAVE simulations: (1) information gleaned from the virtual reality environment can aid in the development of new manufacturing processes, (2) the CAVE/SP1 integration will enable researchers to explore alternative designs in real time, and (3) virtual reality will facilitate training in performing manufacturing operations.

3.d. HiPPI/SONET Gateway: Paving the Information Superhighway

As a result of the HPCC Program, several high performance local area computer networks that use gigabit network technology based on the High Performance Parallel Interface (HiPPI) have been built across the country. This is the highest performance local area networking infrastructure commercially available today. The Los Alamos HiPPI interconnects the CM-5, Cray T3D, High Performance Data System (HPDS) (described below), SGI visualization high performance computers, plus various workstations and workstation clusters. While these local area networks can reach up to a few kilometers, interconnecting testbeds that are farther apart was a critical need.

With HPCC support, Los Alamos developed the HiPPI/SONET Gateway to solve this distance limitation. This Gateway uses SONET (Synchronous Optical NETwork), which is the commercial infrastructure now being implemented by local and long distance telephone carriers. The Gateway uses SONET 155 megabit channels (OC-3C) to obtain the full HiPPI bandwidth by "striping" seven of these channels to achieve the full 0.8 gigabit/second bandwidth along with forward error correction.

The Gateway is accomplishing two important milestones: (1) the link installed in August 1993 between Los Angeles and San Diego holds the world speed record of 660 megabits per second between two widely separated high performance computers, and (2) the installation in the second quarter of 1994 of the link between Los Alamos and San Diego. These links, especially the latter, will be the fastest in operation among the gigabit testbeds. The Los Alamos link will also have the highest bandwidth - distance product of any testbed now in operation.

3.e. Global Climate Modeling and the CHAMMP Program

Global climate simulations offer the best scientifically grounded framework for studying our changing planet. The HPCC Program together with the DOE CHAMMP (Computer Hardware, Advanced Mathematics, and Model Physics) program has been instrumental in defining a climate simulations development path that will use the power of large scale parallel processors. Oak Ridge National Laboratory (ORNL), in collaboration with Argonne and the National center for Atmospheric Research (NCAR), has provided the first full implementation of a state-of-the-art atmospheric climate model on a large scale parallel computer. To explore the feasibility of using such computers for climate research, the Parallel Community Climate Model PCCM2 has been optimized and configured for parallel I/O. The high I/O requirements of climate models is currently driving the large scale parallel computer vendors to advance computing technology in this key area. One-year climate simulations have been performed on the ORNL Intel Paragon XP/S 35 to validate climate statistics in preparation for an even more accurate 10 year simulation using high spatial resolution. This simulation will explore the importance of fine scale processes in climate prediction. This simulation has not been performed on conventional high performance vector computers because of the high memory requirements and the dedicated computational resources required to complete the study.

3.f. The Los Alamos High-Performance Data System (HPDS)

Under the auspices of the HPCC Program, the implementation of the Los Alamos High-Performance Data System (HPDS) has resulted in significant technological breakthroughs in data access and data storage system science and has laid the groundwork for future data storage systems.

The HPDS was developed to meet the enormous data storage and data access requirements of DOE Grand Challenge problems, and is the first fourth-generation data storage system to be put into production. It can handle data generated by problems running on the most advanced, large scale parallel, large memory computers and high speed cooperative processing networks in the world. HPDS is currently being used extensively to handle massive amounts of data from global climate model simulations. In addition, knowledge and experience gained by implementing and using HPDS is being used in designing and implementing the High-Performance Storage System (HPSS), one of the largest collaborative efforts between national laboratories and industrial partners. Requests for HPDS have opened the door for future collaborations and technology transfer. Without HPCC support, HPDS development and implementation would not have happened, and data storage technology advancement would be substantially delayed.

3.g. Manufacturing for the Semiconductor Industry

Sandia's large scale parallel Direct Simulation Monte Carlo (DSMC) software is being used to simulate plasma etch and deposition reactor designs for the microelectronics industry. The software models fluid flow in the high vacuum regime where continuum equations break down. The software is being used extensively to model process physics under a Sandia/SEMATECH CRADA. Lam Research has already saved money by using the DSMC software to evaluate the parameter space for a new design; previous empirical tests had failed to meet the design goals. Companies including Applied Materials, IBM, Intel, Lam Research, and Motorola are working with Sandia to set up their own problems and have them executed on the Intel Paragon and nCUBE high performance computers at Sandia.

3.h. Domain Mapping Software Advances Parallel Computing

Sandia researchers have developed Chaco, a software package that provides reliable, automated methods for mapping problems onto large scale parallel computers. Efficient parallel execution requires that a problem be decomposed into equally sized subproblems that are assigned to processors in a way that minimizes data exchange between processors. Sandia researchers have recently made important progress in solving this problem. Combining graph theory, numerical analysis, and advanced computer science techniques, Chaco is a sophisticated software package implementing a variety of new and advanced mapping algorithms. It is designed to cover a broad cost/quality range, allowing users to choose methods that give reasonably good mappings very quickly or that provide nearly optimal mappings in moderate time. In addition, the new techniques are quite general and can be applied without expertise to many scientific computations.

Since its release in October 1993, Chaco has been licensed for use at more than 35 external sites, gaining rapid acceptance at universities, corporate research facilities, and other national laboratories.

---

4. NASA

Overview

NASA's HPCC program is an integral part of the Federal HPCC Program. It is structured to contribute to broad Federal efforts while addressing agency-specific computational problems that are beyond projected near-term computing capabilities. NASA's HPCC research and development work fully supports the Federal Program's efforts to dramatically improve U.S. capabilities in high performance computing and communications. The NASA program will enable the agency to demonstrate the testbed technology for a thousand-fold increase in computing performance and to apply that technology to NASA Grand Challenges.

The goals of NASA's HPCC program are to accelerate the development, application, and transfer of high performance computing technologies to meet the engineering and science needs of the U.S. aeronautics and Earth and space science communities, and to accelerate the implementation of the NII. These goals are supported by four specific objectives: (1) to develop algorithm and architecture testbeds that are fully able to use high performance computing concepts and increase overall performance, (2) to develop high performance computing systems scalable to sustained teraflops performance, (3) to demonstrate HPCC technologies on U.S. aeronautics and Earth and space science research problems, and (4) to develop services, tools, and interfaces essential to the NII.

NASA has Grand Challenges in Computational Aerosciences (CAS) and Earth and Space Sciences (ESS). These Grand Challenge applications were chosen for their potential and direct impact on NASA, their national importance, and their technical challenge.

4.1 Computational Aerosciences (CAS) Project

The Grand Challenge in Computational Aerosciences is to accelerate the development and availability of high performance computing technology of use to the U.S. aerospace community, to facilitate the adoption and use of this technology by that industry, and to hasten the emergence of a viable commercial market in which hardware and software vendors can exploit these advantages.

To support these goals, NASA has pioneered the implementation of design and simulation software on parallel systems. This marks the first implementation of aerospace simulation software on parallel systems, showing the way toward making them useful to industry. Additionally, NASA developed the most widely accepted performance evaluation and tuning software for applications on parallel systems; developed the only reliable set of benchmarks with which to judge the many parallel architectures; and provides the most complete set of testbed systems to the aeroscience community. Among its most important accomplishments, NASA has:

- Provided an entire suite of first generation parallel systems for evaluation by the aerospace community
- Led the implementation of aerospace design software on the new parallel systems at the NAS (Numerical Aerodynamic Simulation) facility (Intel Paragon with 208 processors and Thinking Machines CM-5 with 128 processors installed)
- Awarded a Cooperative Research Agreement (CRA) for the HPCC second generation parallel testbed and related research to a consortium headed by IBM

NASA recently enabled remote high speed network access to its testbed systems and successfully demonstrated aerodynamic and multidisciplinary simulation of HSCT (High Speed Civil Transport) configurations using large scale parallel systems.

4.2 Earth and Space Sciences (ESS) Project

The Grand Challenge in Earth and Space Sciences is to demonstrate the potential afforded by teraflops systems to further our understanding of and ability to predict the dynamic interaction of physical, chemical, and biological processes affecting the solar-terrestrial environment and the universe. The computing techniques developed as part of NASA's HPCC program will further the development of a suite of multidisciplinary models leading ultimately to scalable global simulation coupling many models, or for treating highly energetic multiple scale problems in astrophysics.

The Earth and space science community is interested in the ESS research achievements that relate key enabling computational science areas. Top space and Earth scientists are Principal Investigators (PIs) and collaborators on eight ESS Grand Challenge Teams and 21 Phase-1/Phase-2 Guest Computational Investigations selected by NASA Headquarters through the ESS NASA Research Announcement (NRA) process; this has established computational metrics for these teams. Using large scale parallel computing systems to conduct research has become routine for these investigators, but how best to transfer their experience to the broader scientific community has yet to be determined.

A new core ESS performance evaluation activity will produce ESS parallel benchmarks to expose underlying factors that determine performance. NASA has joined with NSF in this effort by forming the Joint NSF-NASA Initiative in Evaluation (JNNIE).

Working with Goddard Space Flight center (GSFC), the ESS project has begun deploying ATM-based LANs within GSFC and connecting to DOD's (Department of Defense) local 2.4 gigabit/second ring around Washington, DC.

The ESS project has developed and demonstrated advanced parallel computational techniques in collaboration with ESS PI Teams. These are interdisciplinary critical mass teams that are developing complete Grand Challenge applications scalable to teraflops. Some of the critical scientific areas are:

- Scalable Hierarchical Particle Algorithms for Cosmology and Accretion Astrophysics. Dr. Wojciech Zurek at the Department of Energy's (DOE) Los Alamos National Laboratory (LANL) is the PI. The scientific Grand Challenge is to provide physical insight into astrophysical phenomena of cosmological significance through numerical modeling. The computational challenge is to improve the numerical statistics of the simulations and better represent the physical processes involved in controlling the evolution of the universe.

- Earth System Model: Atmosphere/Ocean Dynamics and Tracers Chemistry. Dr. Roberto Mechoso of the University of California at Los Angeles is the PI. The scientific Grand Challenge is to develop a global coupled model of the atmosphere and the oceans, including chemical tracers and biological processes, to be used to model seasonal cycle and interannual variability. The computational challenges are to make it possible to test the impact of changing parameter values in these models rapidly and to visualize time-accurate model output in real time.

- Large Scale Structure and Galaxy Formation. The PI is Dr. George Lake at the University of Washington. The scientific Grand Challenge is to allow a physically meaningful comparison between empirical data and hydro-dynamic models of large scale structure and galaxy formation.

The ESS project has also developed a teaching curriculum for parallel computational techniques for the annual NASA Summer School in High Performance Computational Physics at GSFC.

Further, the ESS project established the center of Excellence in Space Data and Information Sciences (CESDIS) University Research Program in Parallel Computing and linked each research project with a major NASA activity. Under this project, ESS worked closely with Convex, Cray Research, Intel Parallel Systems Division, and MasPar, in developing their scalable parallel systems, and installed major testbeds from these manufacturers at Jet Propulsion Laboratory (JPL) and GSFC to make their systems (Convex SPP-1, Cray T3D, Intel Paragon, and MasPar MP-2 Cluster) available for evaluation by ESS investigators.

4.3 K-12 Project

The NASA K-12 Project supports the IITA's (Information Infrastructure Technology and Applications, a component of the HPCC Program) efforts to establish standards and provide working models for commercial communications infrastructure. NASA has developed a plan by integrating a combination of 33 schools, NASA HPCC K-12 projects, and Regional Training centers, located in 14 states and the District of Columbia. The project's goal is to significantly contribute to NASA's role in meeting national goals for science and mathematics K-12 education in areas appropriate for high performance computing and communications technology. The objectives are to involve all NASA centers participating in the HPCC program, involve other centers as appropriate, obtain a broad national presence, impact early stages of education (middle school and earlier), inspire students with NASA mission applications, enable students to become electronic information explorers, and facilitate communication among separated communities to empower synergistic collaborations.

One example of project implementation is the current effort in the State of Maryland that provides connectivity to every school district within the State. This was accomplished through the Maryland/Goddard Science Teacher Ambassador Program. That Program's overall objective is to introduce secondary school Earth or environmental science teachers to NASA Earth and Space Science data, information, Internet tools and applications, for subsequent dissemination to students and other educators.

---

5. National Institutes of Health (NIH)

5.a. Visible Human Project -- National Library of Medicine (NLM)

Images are an important part of biomedical knowledge, and high performance computing has enabled new methods for visualizing three-dimensional anatomic structure. With HPCC resources, NLM has undertaken the first steps toward building a collection of three-dimensional digital images of human structure at sub- millimeter level resolution. Named the Visible Human Project (VHP), this image collection is derived from Computed Tomography, Magnetic Resonance Imaging, and digitized cryosections of cadavers. As a reference standard for three- dimensional human anatomy, the VHP data will enable creation of "living" textbooks of anatomic structure that allow students and health care practitioners to view internal body structures from any viewpoint, using interactively generated three-dimensional imaging. Extending these methods to individualized treatment planning will provide an unprecedented ability to plan accurate and minimally invasive surgery.

5.b. Telemicroscopy -- National center for Research Resources (NCRR)

The increasing availability of high speed computer networks makes feasible a system in which a scientist working at a graphics workstation in a laboratory can perform a research investigation by deploying tasks to specialized scientific instruments and high performance computers located at remote sites and interconnected to the workstation over the network. The Microscopist's Workstation project explores such a "distributed laboratory" in which a researcher will be able to remotely acquire and process images of a biological specimen using a special electron microscope in combination with high performance computers. The project is designed to extend the availability of an intermediate high voltage electron microscope, a unique scientific laboratory instrument located at the San Diego Microscopy and Imaging Resource (SDMIR), through the use of advanced networking technology.

The microscope can image considerably thicker biological specimens with greater contrast than more commonly available lower voltage instruments. Images of thick specimens provide important information about the specimen's three-dimensional structure. The microscope is being used to study several cellular phenomena, including the disruption of nerve cell components in Alzheimer's disease.

The microscope is controlled by a local workstation connected via high speed networking to the remote researcher's workstation. Network voice and data links allow the investigator to communicate with the microscope operator and to collect images in a variety of forms in near real time. These images can be routed via the network to a high performance computer for automatic processing into a digital volume representing the interior of the specimen.

The Microscopist's Workstation project is a collaborative effort among the SDMIR at the University of California San Diego (UCSD) School of Medicine, the San Diego Supercomputer center, the Scripps Research Institute, and the UCSD Department of Computer Science and Engineering. Costs are being shared by NCRR, NIH, NSF, the State of and University of California, and industrial partners (Datacube, Network Systems, and Sun).

5.c. Predicting Molecular Structure From Genetic Code -- National Cancer Institute (NCI)

Large scale parallel computation and high speed network communications have become important parts of the research effort at NCI's Laboratory of Mathematical Biology. An optimized and parallelized version of the Zuker RNA folding software, which predicts molecular structure by analyzing the linear sequence of RNA (the messenger molecule produced by genes), has been adapted to a MasPar MP-2 large scale parallel computer. The program is able to analyze RNA sequences that are over 9,000 nucleotides in length; this is substantially larger than the maximum length that can be analyzed by the Cray YMP computer. The HIV virus is an example of such a sequence. NCI researchers have demonstrated that the MP-2 can fold such sequences 4.5 times faster than a single processor of the YMP, thus giving a considerable price/performance advantage. Efficiently folding such large sequences will enable a detailed comparison of structural motifs that may be common in viral families such as HIV, and support both vaccine development and targeted drug design.

---

6. National Institute of Standards and Technology (NIST)

6.a. Guide to Available Mathematical Software (GAMS)

A wealth of reusable software is freely available for use by scientists and engineers for solving recurring types of mathematical and statistical problems. As high performance computers become more widely available, it is increasingly important that this high quality software be easily accessed and used. Unfortunately, the average user has difficulty locating the most appropriate software within the volume of specialized software distributed over computer networks.

The NIST Guide to Available Mathematical Software (GAMS) project develops techniques and tools to help scientists and engineers locate and use computer software to solve mathematical and statistical problems. The products of this work include a problem classification scheme, a database of information about available software, and an information retrieval system for user access. The information system functions as a virtual software repository; rather than maintaining a repository itself, it provides a convenient cross-index of repositories managed by others. Moreover, GAMS uses Internet resources to connect to these repositories and obtain source code, documentation, and tests for its users. For example, some 1,600 software modules from the well-known netlib repository have been cataloged. GAMS retrieves these items from the netlib server at ORNL. The complete GAMS database contains information on more than 9,200 software modules from some 60 software packages at four physical repositories.

Anonymous telnet access to the GAMS system is available to Internet users. Both a generic and a sophisticated X11 networked graphical user interface are supported. Some 1,700 connections from about 150 Internet hosts were logged during its first eight months of (largely unadvertised) operation. Recently, a new network access mechanism was developed based on NCSA Mosaic. GAMS availability via Mosaic has had a strong impact on user access -- more than 17,000 GAMS Mosaic access requests were recorded during April 1994.

The GAMS project supports the goals of the ASTA component of the HPCC Program in facilitating the widespread reuse of algorithms and software developed for high performance computers. It depends upon Internet advances in order to effectively deliver services to its users and on Mosaic software developed by NCSA under NSF HPCC funding. GAMS is one of the software exchange mechanisms being studied and coordinated by NASA.

6.b. The MultiKron Performance Measurement Chip

The basic measure of computer performance is the length of time required to complete a task. But then comes the question, "Why did it take that long?" The answer is particularly difficult to obtain in parallel computers, where problems are broken into many parts and solved simultaneously by many processors. This contrasts with traditional computers where problem parts are solved sequentially by one processor.

Scalable parallel computing systems constitute a growing market encompassing both vast machines at the high end of computing and desk-side personal workstations. Yet the technology is still immature. Large scale parallel programming is not easy and system features are not firmly established. Different programming techniques are required by each parallel architecture.

The NIST parallel processing group investigates performance measurement methods that promote scalable systems. Accurate measurement permits understanding the source of performance bottlenecks and therefore leads to techniques to scale system designs upwards. This approach addresses concerns in achieving the economic promise of affordable large scale parallel systems.

In cooperation with ARPA, NIST has designed the MultiKron chip technology for measuring the performance of advanced processors and very high speed networks. MultiKron technology achieves low-perturbation data capture for assessing the performance of scalable parallel computing systems. This technology was transferred to Intel, as well as to other companies and researchers. Intel applied the MultiKron measurement principles in its scalable parallel Paragon system, whose designers and users can obtain accurate performance information and use that information to achieve maximum throughput.

NIST also designed, developed, and distributed a designer and experimenter's kit for using the industry standard VME bus to integrate MultiKron instrumentation into systems, and used MultiKron instrumented systems to assess the performance of TCP/IP communications protocols over a HiPPI-based network.

---

7. National Security Agency (NSA)

7.a. SPLASH 2 System Development

Researchers at the Supercomputing Research center (SRC) of the Institute for Defense Analyses, a federally funded research center, and NSA have developed SPLASH 2, a general purpose hardware accelerator board that can deliver high performance computer processing performance for a variety of applications at greatly reduced cost. In the most basic terms, SPLASH 2 provides a collection of raw logic gates that can be programmed through software. The primary processing elements of SPLASH 2 are Xilinx 4010 field programmable gate arrays (FPGAs). Each FPGA is connected to a 256K x 16-bit RAM. Communication between the processing elements is performed via a linear data path and a crossbar interconnection network. The linear data path enables SPLASH 2 to perform systolic algorithms, while the crossbar connection makes it possible to implement algorithms that require irregular communication. The combination of these two communication patterns facilitates the data transfers required for a diverse set of applications. A SPLASH 2 system consists of a Sun SPARCstation host, an interface board, and between one and 16 SPLASH 2 array boards. SPLASH 2 is programmed in the VHSIC (Very High Speed Integrated Circuit) Hardware Description Language (VHDL), which is an IEEE and DOD standard programming language. This language allows users to configure the system hardware to perform specific processing functions. Applications implemented in this manner are able to run at the speed afforded by hardware while maintaining the flexibility provided by software. Thus SPLASH 2 is an excellent computing platform for researchers who do not have the resources to design and build hardware for every new high speed processing task they wish to perform.

Several applications have been implemented on the 10 existing SPLASH 2 prototype systems. Researchers at Virginia Tech are using SPLASH 2 to process image data at real time video rates. Image processing algorithms that have been successfully implemented on SPLASH 2 include edge detection, Fourier transforms, image convolution, and histogram generation. Work is scheduled to begin at Michigan State University that will use SPLASH 2 for fingerprint identification at rates that are orders of magnitude faster than those achieved using current methods. SPLASH 2 algorithm research is not confined to universities. The National Cancer Institute is using SPLASH 2 to test newly identified DNA strands against an existing database so that these new strands can be evaluated and categorized. Although these applications begin to demonstrate the variety of tasks that can be efficiently executed on SPLASH 2, other applications in signal processing and prototype development environments may also benefit from its processing capabilities. Future work will concentrate on refining the SPLASH 2 hardware and making the technology available to researchers for examining these unexplored application areas.

7.b. ATM/SONET Project

Over the past two years, the NSA Information Security Research organization has successfully addressed the problem of incorporating encryption in ATM and SONET networks. This has consisted of defining the methods of operation and building and demonstrating proof-of-concept encryptors to operate in test versions of these networks. To define the systems, NSA has been working with the vendor community, standards groups, and users at the national and international levels to create command structures, protocols, and formats that allow efficient use and management of encryption in these environments. NSA is also developing the necessary vendor independent interfaces between host communications equipment and the encryptors.

This program addresses the secure use of future common carrier communications systems. ATM is the technology with the most promise for these applications. Unlike present carrier services where capacity is allocated and billed regardless of actual information transfer, in ATM the capacity of a physical link is allocated only as needed based upon demand for transferring information. This is the concept of a "virtual circuit" -- the user appears to have an allocated circuit, but the network only transfers information for the user when information is present. This efficient sharing of network capacity potentially represents dramatic cost savings to telecommunications users. Because of the seemingly random mixing of information flows along a communications path, each packet of information potentially belongs to a different virtual circuit. Any encryption concept, therefore, has to switch between user states at the packet rate of the network. This "Key-Agile" encryption allows multiple ATM virtual circuits/channels to operate on a per demand basis, in conformance with the principles of ATM virtual circuit communications, while maintaining confidentiality between the various sets of users without degrading network capacity. The proof-of-concept models have focused on the ATM networks, but the same basic concepts are baselined into KG- 189, for securing SONET links, and into FASTLANE, a production encryptor for securing ATM virtual circuits. Verifying the validity of the operational concepts in a realistic environment is an important part of this program. NSA is a node on ATDnet (Advanced Technology Demonstration network), DOD's Global Grid Washington-area ATM/SONET testbed, which is presently completing installation and checkout. Once stable, ATDnet will be used to test and demonstrate NSA's prototype encryption equipment, thereby paving the way for fully operational and approved encryption equipment.

7.c. TERASYS Development and Cray-PIM Supercomputer

The Terabit Operation Per Second (TERASYS) project goal is to demonstrate large scale parallel computing in the memory of a general purpose computer. The system is based on a specially designed Processor-In-Memory (PIM) chip containing 64 processors. Five TERASYS workstations were produced during the past year. Each contains 512 PIM chips and computes at approximately the same speed as an eight- processor Cray YMP. A software environment has been written to make the hardware ready for user-written programs.

The next step is incorporating PIM chips into the Cray-3 computer. If successful, this will result in a prototype of a commercial PIM high performance computer. Cray Computer would then have the expertise to build production machines. Using existing memory density technology, the prototype is expected to achieve a speed equivalent to 200 to 800 Cray YMP processors at a fraction of the cost of current high performance computers. This machine would be suited for parallel applications that require a small amount of memory per process. Examples of such applications are laminar flow over a wing, weather modeling, image processing, and other large scale simulations or bit-processing applications. Because it is based on memory technology, PIM technology is readily adaptable to future memory density advances.

---

8. National Oceanic and Atmospheric Administration (NOAA)

NOAA is a mission agency responsible for describing, predicting, and understanding changes in the Earth's environment. To perform and to advance this mission, NOAA develops and applies leading-edge computation and communications technologies to perform Grand Challenge research in climate and weather modeling, predict the weather, understand and predict ocean processes, and manage the Nation's fisheries.

To advance its capabilities in high performance computing, NOAA has acquired a Cray C90 high performance system for its National Meteorological center (NMC, Suitland, MD) for operational weather forecasting. It has acquired a 208-processor Intel Paragon for its Forecast Systems Laboratory (FSL, Boulder, CO) to be used as a state-of-the-art testbed for developing advanced algorithms in climate, ocean, and weather modeling. This acquisition was accomplished in conjunction with NOAA's partners in the ARPA-sponsored National Consortium for High Performance Computing and as part of the Boulder Front Range Consortium. NOAA is conducting a major technological upgrade to its high performance computer facility at the Geophysical Fluid Dynamics Laboratory (GFDL, Princeton, NJ) to support research in numerical algorithms for environmental modeling.

8.a. High Performance Computing

NOAA has long been a major developer and user of advanced computational technologies in Grand Challenge applications. Over the past several years, NOAA has had major successes in advancing our understanding of the environment through the use of advanced high performance computing technologies. Among NOAA's recent accomplishments are:

- The Parallel Ocean Program (POP), a variant of the GFDL Modular Ocean Program (MOM), has been successfully transported by scientists at DOE/Los Alamos National Laboratory (LANL) and GFDL, to the LANL 1,024-processor Thinking Machines CM-5. Scientists are now using this model to perform very high resolution experiments to produce multi-year eddy-resolving simulations of the world oceans.

- As part of the NOAA/DOE collaboration, GFDL's SKYHI global atmospheric grid-point model (used to study climate, middle atmosphere dynamics, and atmospheric ozone) has been thoroughly redesigned and transported to the LANL CM-5. GFDL and LANL scientists are now preparing to conduct global atmospheric experi- ments with half degree horizontal resolution (producing 50 kilometer grid spacing over the United States) to investigate the effect of small scale features on stratosphere and mesosphere dynamics. These very high resolution experiments require computing resources that are beyond the capabilities of the current GFDL computer.

- NOAA scientists, collaborating with DOD/Naval Research Laboratories (NRL) scientists, have successfully implemented a version of the NMC atmospheric spectral model to the NRL Thinking Machines CM-200.

- A regional atmospheric model has been restructured by NMC and FSL scientists for execution on large scale parallel systems. NOAA's FSL scientists have developed a parallel version of the Mesoscale Analysis and Prediction System (MAPS) as a functional prototype for both the Federal Aviation Administration and the National Weather Service. MAPS is the first of several strategic weather models to be parallelized for the Aviation Weather Program. These models will provide high resolution forecasts on both national and regional scales to support operational and aviation meteorology.

- NOAA's Pacific Marine Environmental Laboratory (PMEL) has entered into a cooperative agreement with the Arctic Region Super- computing center (ARSC) of the University of Alaska, which will provide PMEL and other NOAA scientists with access to ARSC's high performance computing resources, including a 128-processor Cray Research T3D. This collaboration will allow PMEL scientists to incorporate new parallel computing capabilities into their research models for ocean circulations of the Gulf of Alaska and the Bering Sea.

8.b. Environmental Information Dissemination through Networks

In performing its mission, NOAA collects, archives, and disseminates environmental data and information. It operates three National Data centers and more than 30 additional major environmental information archive and distribution facilities. To fulfill this mission, NOAA has taken a pioneering role in advancing the use of the Internet and the envisioned NII to open its vast archives of environmental data and information for easy public access. NOAA is applying new information discovery tools that have emerged from the HPCC Program, such as Mosaic, to its environmental information systems to provide ready user access across the diversity of NOAA information. HPCC technologies play a strategic role in allowing NOAA to provide an integrated view across geographically distributed information systems so that a virtual "NOAA environmental information system" can be developed.

NOAA's recent accomplishments in providing timely, universally accessible information services include the following:

- NOAA now has 29 information servers on the Internet providing a wide variety of real time and historical environmental data and information to the world.

- As of the latest count, NOAA has implemented 15 Mosaic servers that access major NOAA databases, thereby providing simple, highly graphical, multimedia access to a broad range of environmental information.

- Using the new information technologies, NOAA has developed the ability to provide scientists, policy makers, educators, and students with immediate access to information that, until recently, would have taken weeks or months to find and retrieve from archives and distribute to the requestor. NOAA's information servers connected to the Internet provide convenient, instantaneous access to centuries of global weather observations, the current status of solar X-rays and particles as they are detected by NOAA's Geostationary Operational Environmental Satellites (GOES), current and recent past winds and sea temperatures across the tropical Pacific used to monitor El Nino events, and a wide range of other atmospheric, geophysical, and oceanographic information.

- A prototype Satellite Active Archive system connected to the Internet provides access to vast quantities of near real time NOAA satellite imagery as well as to historical satellite imagery.

- The NOAA Data Directory Services allows Internet users to browse information about the agency's environmental data holdings including information about applicability for a variety of uses and how to obtain the data.

9. Environmental Protection Agency (EPA)

9.a. Environmental Modeling

High performance computing has enabled the linkage of air and water models to address the problem of nutrient overloading in the Chesapeake Bay that can cause dramatic decreases in the productivity of the shellfish industry. Atmospheric deposition is an important pathway for nitrogen and toxic chemicals, possibly accounting for 25 to 35 percent of the nitrogen going into the Bay. Through the use of new technology and optimized computer software, EPA scientists have been able to use nitrogen oxide deposition estimates from an air quality model as input to water quality models to determine how reductions in emissions of nitrogen oxides translate into changes in nutrient load in the Bay. High performance computing has created possibilities for greater cost efficiency in solving the problem because, beyond a certain point, air controls may be relatively less expensive than water controls. HPCC has enabled a more complete cross-media examination of options, allowing decision makers to take advantage of added benefits to regional ozone control by reducing air emissions of nitrogen.

High performance computing has facilitated the use of advanced air quality models in a retrospective study of the costs and benefits of the 1970 Clean Air Act. The use of advanced models for this type of assessment was typically infeasible due to computational limitations. High performance computing allowed more pollutants to be modeled, bringing in regional acid deposition, sulfate, visibility on an annual basis and regional ozone on a seasonal basis, while at the same time enabling uncertainty analysis and model evaluation to ensure scientific credibility of the results. This had never been done before using a single, self-consistent modeling system.

As part of efforts to develop an environmental decision support system, researchers at MCNC have demonstrated the use of dependency graphs for automated execution of multiple dependent programs including meteorology and emissions processing for the Urban Airshed Model. The processing graphs are created using the Heterogeneous Network Computing Environment (HeNCE), a public domain integrated graphical environment for creating and running parallel programs over heterogeneous computers. MCNC also completed a portable suite of modules, data, and networks for an Applications Visualization System (AVS)-based analysis component for the environmental decision support system. Linkages between AVS and ARC, from Environmental Systems Research Institute, were constructed to facilitate visualization and three-dimensional animation of data in geographical information systems.

A multidisciplinary team at Carnegie Mellon University has developed a Geographic Environmental Modeling System (GEMS) designed to facilitate in-depth study of the environmental effects of air pollution. GEMS provides a way to explore the relative effectiveness of alternative emissions control strategies. It is transportable to a number of computer systems, accesses data remotely via servers, and uses a graphical user interface for ease of use.

Researchers at the University of Wisconsin have demonstrated the applicability of three-dimensional animation techniques for air quality models. New volume rendering and interactive data selection techniques were developed for VIS-5D (VISualization of 5-Dimensional data sets) to accommodate the extensive data requirements of air quality models.

9.b. Computational Techniques

Researchers at Caltech have implemented six advection algorithms on both sequential (IBM and Sun workstations) and parallel (a network of IBM RISC workstations and a 256-processor Intel Touchstone Delta) platforms and performed comparative testing. Advection is a major component of air quality models. Results indicate that the accurate space derivative method coupled with the Forester filter produces the most accurate results, although at the expense of more computation time. However, a speed-up of about 88 is achieved with the parallel implementation on the Intel.

A gas-phase chemistry quasi-steady state approximation solver has been transported to a 32-processor Kendall Square Research KSR 1 by researchers at North Carolina State University. Numerous comparative studies indicate that this algorithm is more suited to vector processors. Experiments with other algorithms indicate the importance of optimizing local cache performance. The energy portion of a quantum mechanical model of large molecules has also been transported to the KSR 1 by researchers at Duke University.

As part of a joint research effort between EPA and DOE, researchers at Argonne National Laboratory have implemented a parallel version of the Mesoscale Meteorology model (MM5) using a portable message passing library on an IBM SP1 and an Intel Delta. Almost 2 gigaflops was achieved using 120 processors on the SP1 and about 3 gigaflops was achieved using 480 processors on the Delta. These are encouraging results given the importance of meteorological models for air quality simulation.

Four new cooperative agreements for research on parallel algorithms for ground water flow and transport, water quality, aerosol chemistry, and prediction of reactivity and structure of organic and biochemical compounds in solution are being awarded in FY 1994.

EPA's Scientific Visualization center (SVC) completed tests of distributed visualization techniques involving all available computing resources at the Research Triangle Park and Bay City Visualization centers. Capabilities include access to large remote data sets, distributed parallelization of computationally intensive tasks, time sequenced visualization, animation, and recording. These techniques are now being incorporated into the overall visualization infrastructure to support agency-wide distributed visualization.

EPA has established a desktop video conferencing capability between scientists at Research Triangle Park and the EPA's National Environmental Supercomputing center (NESC) in Bay City, Michigan using MBONE (multicast backbone) digitized video techniques over the Internet. This link is used routinely for planning and status meetings, training sessions, and user support for software optimization and visualization. These capabilities are being extended to additional EPA remote Sites.

9.c. Network Connectivity

EPA established a FDDI research network in Research Triangle Park, NC with a high speed (T3 or 45 megabits per second) Internet connection to NESC. T1 (1.5 megabits per second) connectivity has been provided to the Environmental Research center in Georgia and to the Chesapeake Bay Office in Maryland and agreements are being established to extend network connectivity to the Ozone Transport Commission in New York and the Tennessee Valley Authority. This connectivity provides the bandwidth needed to satisfy heterogeneous computing, data management, and collaborative visualization requirements for EPA's HPCC research and technology transfer pilot programs.

9.d. Education/Training

A workshop was held for Federal, state, and industrial personnel participating in technology transfer pilot programs to begin an assessment of training needs and user requirements for use of air quality decision support systems. A prototype regulatory modeling and analysis support system, UAMGUIDES, was demonstrated for use in urban ozone non-attainment studies and is being provided to all participants for their continued use and evaluation. Multimedia electronic "tutor and help" approaches, televisualization, as well as collaborative approaches to provide expert advice to remote locations are being evaluated.

EPA's HPCC program helps support EarthVision, a computational science educational program for high school students and teachers. Saturday tutorials at NESC provide basic knowledge and skills required to prepare project proposals. Competitively selected teams participate in a three week summer education program. During the academic year, the teams conduct environmental research using scientific workstations and EPA's high performance computer.

EPA researchers have been given hands-on visualization training, and a student summer research program provides undergraduate opportunities for summer employment involving HPCC activities. Numerous undergraduate and graduate students and a postdoctoral student are supported directly through cooperative research agreements with universities.