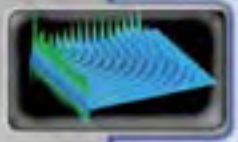







Information Technology Research and Development

INFORMATION TECHNOLOGY

The 21st Century Revolution

2001

Supplement to the President's FY 2001 Budget

National Science and Technology Council
Interagency Working Group on IT R&D

About the National Science and Technology Council:

President Clinton established the National Science and Technology Council (NSTC) by Executive Order on November 23,1993.This cabinet-level council is the principal means for the President to coordinate science, space, and technology policies across the Federal Government.NSTC acts as a "virtual agency" for science and technology to coordinate the diverse parts of the Federal research and development enterprise. The NSTC is chaired by the President. Membership consists of the Vice President, Assistant to the President for Science and Technology, Cabinet Secretaries,Agency Heads with significant science and technology responsibilities, and other White House officials.

An important objective of the NSTC is the establishment of clear national goals for Federal science and technology investments in areas ranging from information technologies and health research, to improving transportation systems and strengthening fundamental research. The Council prepares research and development strategies that are coordinated across Federal agencies to form an investment package that is aimed at accomplishing multiple national goals.

To obtain additional information regarding the NSTC, please contact the NSTC Executive Secretariat at (202) 456-6100.

About the Office of Science and Technology Policy:

The Office of Science and Technology Policy (OSTP) was established by the National Science and Technology Policy, Organization and Priorities Act of 1976.OSTP's responsibilities include advising the President in policy formulation and budget development on all questions in which science and technology are important elements;articulating the President's science and technology policies and programs; and fostering strong partnerships among Federal, State, and local governments, and the scientific communities in industry and academe.The Assistant to the President for Science and Technology serves as the Director of the OSTP and directs the NSTC on behalf of the President. OSTP Associate Directors co-chair the various committees of the NSTC.

To obtain additional information about OSTP, please call (202) 456-7116.

Dedication

This report is dedicated to Henry C. Kelly, who served as Assistant Director for Technology in the White House Office of Science and Technology Policy from 1993 to May 2000. As White House liaison to the President's Information Technology Advisory Committee (PITAC), Dr. Kelly encouraged the Committee to articulate boldly and fully its view of the Nation's information technology research needs, as the PITAC did in its highly regarded February 1999 report, "Information Technology Research: Investing in Our Future." He was instrumental in designing the Administration's Information Technology Research and Development (IT R&D) Programs,which are well aligned with the PITAC's recommendations and are the subject of this document. In June 2000, Dr. Kelly became President of the Federation of American Scientists. But his sage advice and tireless efforts on behalf of the Federal IT R&D enterprise are remembered with admiration and will long be appreciated.

Cover Design

This year's cover depicting the beginning of the new Information Technology millennium was designed by National Science Foundation artist James J. Caras.The cover images were selected from agency IT R&D activities described in this FY 2001 report.The contents of this report are set in Adobe Caslon, a modern revival of a typeface released by William Caslon in 1722. Due to its high degree of legibility, it was popular in the American Colonies, and was a particular favorite of Benjamin Franklin. The first printings of the U.S. Declaration of Independence and the Constitution were set in a version of Caslon.

Information Technology
Research and Development

INFORMATION TECHNOLOGY:

THE 21ST CENTURY REVOLUTION

A Report by the Interagency Working Group on IT R&D

National Science and Technology Council

OSTP LETTER

[FPO-Letter to be supplied]

Senior Principals Group

Chair
Neal Lane, Ph.D.
Assistant to the President for Science and Technology and
Director, White House Office of Science and Technology Policy

Members

D. James Baker, Ph.D.
Under Secretary for Oceans and Atmosphere and
Administrator, National Oceanic and Atmospheric Administration
Department of Commerce

Carol M. Browner
Administrator
Environmental Protection Agency

Rita Colwell, Ph.D.
Director
National Science Foundation

Delores M. Etter, Ph.D.
Deputy Under Secretary for Science and Technology
Department of Defense

Daniel S. Goldin
Administrator
National Aeronautics and Space Administration

Thomas A. Kalil
Special Assistant to the President for Economic Policy
White House National Economic Council

Ruth Kirschstein, M.D.
Acting Director
National Institutes of Health

Ernest J. Moniz, Ph.D.
Under Secretary
Department of Energy

Lori A. Perine
Deputy to the Associate Director for Technology,
Office of Science and Technology Policy
Executive Office of the President

Wesley Warren
Associate Director, Natural Resources, Energy, and Science
Office of Management and Budget

Interagency Working Group on IT R&D

Chair
Ruzena Bajcsy

NSF/CISE
Representative
Ruzena Bajcsy
Alternates
George O. Strawn
Robert R. Borchers

NSF/MPS
Representative
Robert A. Eisenstein
Alternate
Bradley D. Keister

DARPA
Representative
Shankar Sastry
Alternate
Mark Swinson

NASA
Representative
Lee B. Holcomb
Alternates
David B. Nelson
Betsy Edwards

NIH
Representative
Robert L. Martino
Alternates
Michael J. Ackerman
Judith L. Vaitukaitis

DOE Office of Science
Representative
C. Edward Oliver
Alternate
Norman H. Kreisman

NSA
Representative
George R. Cotter
Alternate
Norman S. Glick

NIST
Representative
William Mehuron
Alternate
Larry H. Reeker

NOAA
Representative
Thomas N. Pyke, Jr.
Alternate
William T. Turnbull

AHRQ
Representative
J. Michael Fitzmaurice

OSD
Representative
Charles J. Holland
Alternate
Cliff G. Lau

EPA
Representative
Joan H. Novak
Alternate
Robin L. Dennis

DOE/ASCI
Representative
Paul C. Messina
Alternate
José L. Muñoz

OMB
David S. Trinkle

OSTP
Lori A. Perine

NCO
Representative
Kay Howell
Alternate
Sally E. Howe

PCA Coordinating Groups and Team Chairs

High End Computing and Computation (HECC) Coordinating Group
Co-Chairs
Robert R. Borchers, NSF
José L. Muñoz, DOE

Human Computer Interface and Information Management (HCI & IM) Coordinating Group
Chair
Gary W. Strong, DARPA

Large Scale Networking (LSN) Coordinating Group
Co-Chairs
George O. Strawn, NSF
David B. Nelson, NASA

High Performance Networking Applications Team (HPNAT)
Chair
William T. Turnbull, NOAA

Internet Security Team (IST)
Chair
Christina M. McBride, NSA

Joint Engineering Team (JET)
Co-Chairs
Javad Boroumand, NSF
Paul E. Love, Internet2

Networking Research Team (NRT)
Co-Chairs
Mari W. Maeda, DARPA
Vacant

Software Design and Productivity (SDP) Coordinating Group
Co-Chairs
Janos Sztipanovits, DARPA
Michael Evangelist, NSF

High Confidence Software and Systems (HCSS) Coordinating Group
Co-Chairs
Helen Gill, DARPA
Vacant

Social, Economic, and Workforce Implications of IT and IT Workforce Development (SEW) Coordinating Group
Co-Chairs
C. Suzanne Iacono, NSF
William S. Bainbridge, NSF

Federal Information Services and Applications Council (FISAC)
Co-Chairs
Lawrence E. Brandt, NSF
G. Martin Wagner, GSA

Digital Government Liaison
Lawrence E. Brandt, NSF

Federal Statistics Team
Chair
Cathryn S. Dippo, BLS

Information Technology for Crises Management Team
Chair
Anngienetta R. Johnson, NASA

Next Generation Internet Team
Chair
William T. Turnbull, NOAA

Universal Access Team
Chair
Susan B. Turnbull, GSA

Digital Government Consortium Liaison
Valerie Gregg, NSF

Table of Contents

Executive Summary

Overview

At the dawn of the 21st century, the U.S. is enjoying an era of unprecedented possibilities and prosperity built on dramatic advances in science and technology. Federal investment in information technology research and development (IT R&D), leveraged by industry and academia, has led to technological breakthroughs that are transforming our society, driving economic growth, and creating new wealth. Revolutionary computing, networking, and communications tools now allow businesses of all sizes to participate in the global economy.

During the past decade, more than 40 percent of U.S. investments in new equipment have been in computing devices and information appliances, and since 1995 more than a third of all U.S. economic growth has resulted from IT enterprises. Today, more than 13 million Americans hold IT-related jobs, which are being added at six times the rate of overall job growth. More than 800,000 jobs were created by IT companies in the past year alone. This astonishing progress has been built on a foundation of Federal agency investments in research conducted in universities, Federal research facilities, and private industry.

Federal support for IT R&D has been essential to the flow of innovative ideas that will ensure America's continued leadership in the New Economy. IT has strengthened our national security, improved our quality of life, and unleashed an extraordinary era of industrial innovation and transformation. As computing and communications systems and software become more powerful and more useful, IT penetrates more deeply into our work, education, and home environments. More than half of U.S. classrooms are connected to the Internet today, compared with less than 3 percent in 1993. Nearly half of all American households now use the Internet, with more than 700 new households being connected every hour.

But obstacles to continued progress remain. Federal support for the kinds of far-reaching technologies that brought us the Internet is still not keeping pace with rapid developments in IT. Serious issues have arisen in education and in the workforce, both of which must adapt rapidly to a bewildering array of evolving technologies. And substantial numbers of U.S. citizens have little if any access to these new technologies—a pernicious "digital divide" that must be bridged.

To assure the health, prosperity, and economic competitiveness of future generations in our rapidly developing New Economy—as well as equal access for all to the knowledge base and tools that are making the New Economy an exciting reality—the Federal government must significantly increase its investments in those revolutionary technologies that will propel U.S. leadership in the 21st century.

Federal IT R&D Programs

The Federal government coordinates multiagency research in advanced computing, communications, and information technologies through the IT R&D Programs (successor to the High Performance Computing and Communications [HPCC] R&D Program and the Computing, Information and Communications [CIC] programs). IT R&D-coordinated activities are organized into distinct but interrelated Program Component Areas (PCAs). In FY 2000, a number of changes were made in the PCA framework, reflecting recommendations of the President's Information Technology Advisory Committee (PITAC) and key research themes emerging from the activities of the original five PCAs that date from FY 1997:

Program Component Areas (PCAs)

- **High End Computing and Computation (HECC).** To better characterize the breadth of the HECC investment, its budget is now reported as two new PCAs:
  - **High End Computing Infrastructure and Applications (HEC I&A)**
  - **High End Computing Research and Development (HEC R&D)**
- **Human Computer Interface and Information Management (HCI & IM)** succeeds the Human Centered Systems (HuCS) PCA, reflecting the increasing challenges of making large amounts of information easily available and useful to the widest variety of users.
- **Large Scale Networking (LSN)** activities include the Next Generation Internet (NGI) Initiative and Scalable Information Infrastructure (SII) R&D.
- **Software Design and Productivity (SDP)** is a new PCA in FY 2001, established in response to the PITAC's finding that not only is the demand for software exceeding our ability to develop it, but the software produced today is difficult to design, test, maintain, and upgrade. Research topics will include security, survivability, availability, reliability, and safety of IT systems and assurance in software- and information-centric systems through research in theoretical foundations, development of techniques and tools (with linkages to domain-specific languages), engineering and experimentation, and demonstrations and pilots.
- **High Confidence Software and Systems (HCSS)** was formerly the High Confidence Systems (HCS) PCA. Its new name and scope reflect the increasing need for adaptability, reliability, safety, and security in both the software and the systems that U.S. citizens count on each and every day.
- **Social, Economic, and Workforce Implications of IT and IT Workforce Development (SEW)** is the successor to the Education, Training, and Human Resources (ETHR) PCA. SEW's expanded R&D portfolio now includes assessment of the social and economic consequences of IT's transforming influence on the workplace as well as expanded research in education and worker training issues resulting from the rapid U.S. move to an information-based economy.

The PCAs are guided by Coordinating Groups (CGs) consisting of representatives from participating agencies. The CGs meet on a regular basis

throughout the year to review the progress of their PCA's research agenda and plan coordinated and collaborative activities.

The Federal Information Services and Applications Council (FISAC) works closely with the IT R&D Programs. The FISAC assists in the transfer of advanced computing and communication technologies developed by the IT R&D community to agency missions and systems across the Federal government.

HECC: High End Computing and Computation

HECC R&D extends the state of the art in computing systems, applications, and high end infrastructure to achieve the scientific, technical, and information management breakthroughs necessary to keep the U.S. in the forefront of the 21st century IT revolution. HECC researchers develop and apply the world's most advanced computing capabilities to modeling and simulation processes in physics, materials, biology, and environmental sciences; data fusion and knowledge engineering across all elements of the national, scientific, and industrial sectors; and complex and computationally intensive national security applications. Research activities are primarily supported at high end computing facilities that include the National Science Foundation's (NSF's) Partnerships for Advanced Computational Infrastructure (PACI) centers, National Aeronautics and Space Administration (NASA) testbeds, and Department of Energy (DOE) laboratories.

HECCCG, HEC I&A, and HEC R&D

Formerly the HECC Working Group (HECCWG), the HECC Coordinating Group (HECCCG) coordinates Federal R&D investments focused on maintaining and expanding U.S. leadership in high performance computing and computation, and promotes cooperation among Government laboratories, academia, and industry. In FY 2000, the HECCCG divided HECC R&D into two PCAs:

- HEC I&A research includes Federal agency mission applications development as well as the computing infrastructure needed to support this work. Focus areas include biomedical sciences, computational aerosciences, Earth and space sciences, weather forecasting and climate modeling, and tools to facilitate high end computation and analysis and display of the large data sets such computation generates.
- HEC R&D focuses on teraflops-scale (a trillion or more floating point operations per second) systems and computation. Research activities in this area support fundamental, long-term research to maintain and extend the U.S. lead in computing for generations to come. Current research focuses on advanced architectures such as hybrid technology multithreaded (HTMT), Beowulf for networked workstations, mass storage technologies, national-scale computational grids, and molecular, nano-, optical, quantum, and superconducting technologies.

HCI & IM: Human Computer Interface and Information Management

HCI & IM R&D develops advanced technologies that expand modes of human-computer interface and improve our ability to manage and make use of computing devices and information resources. The wide-ranging HCI & IM research agenda is generating new IT capabilities in many fields, including biomedicine, commerce, crisis management, education, law enforcement, library sciences, manufacturing, national defense, scholarship, and weather analysis. FY 2001 R&D areas include:

- Battlefield robotics, such as the Defense Advanced Research Projects Agency's (DARPA's) mobile autonomous robot software program, and remote/autonomous agents such as NASA's Remote Exploration and Experimentation (REE) project to develop autonomous supercomputing capabilities on spacecraft
- Collaboratories, visualization, and virtual reality, including DOE's DeepView scalable system for distributed collaborative research using online advanced microscopes; the National Institute of Standards and Technology's (NIST's) collaboratory, simulation, and 3-D visualization capabilities for manufacturing; and NSF's National Computational Science Alliance (Alliance) Access Grid for multisite teleconferencing, training, and research collaboration in 3-D immersive environments
- Web-based knowledge repositories and information agents for collecting and analyzing data, including the multiagency Digital Libraries Initiative Phase Two; NIST's Digital Library of Mathematical Functions (DLMF); and the Agency for Healthcare Research and Quality's (AHRQ's) COmputerized Needs-oriented QUality measurement Evaluation SysTem (CONQUEST), a tool enabling users to evaluate clinical practice performance
- Multimodal interactions between humans and computer systems, including DARPA's Communicator program to incorporate speech recognition in mobile computing devices; haptic devices, such as NASA's human factors program to add touch-feedback capabilities to simulation models; and intelligent systems, such as NSF's knowledge and cognitive systems (KCS) program to apply cognitive science to development of artificial intelligence in computing systems
- Multilingual translation, such as DARPA's Translingual Information Detection, Extraction, and Summarization (TIDES) program to generate English translations of materials in other languages

LSN: Large Scale Networking

LSN R&D has become a pivotal force in IT networking—including terrestrial optical, wireless, and satellite networking—generating critical advances in technologies that have been quickly adopted by the governmental, academic, and commercial sectors. LSN research supports key Federal agency missions and provides leadership in networking technologies, services, and performance required to create the scalable, reliable, secure very high-speed networks of the future. LSN's NGI and SII programs are designing and developing the prototypes for these future networks.

The LSN Coordinating Group (LSNCG)—formerly the LSN Working Group (LSNWG)—coordinates Federal networking R&D programs. Four teams, each of which includes non-Federal participants, report to the LSNCG to help coordinate Federal networking research and implement advanced networking technologies:

Joint Engineering Team (JET)

- The ***Joint Engineering Team (JET)*** coordinates the network architecture, connectivity, exchange points, and cooperation among Federal agency networks (FedNets)—DOE's Energy Sciences network (ESnet), NASA's Research and Education Network (NREN), NSF's very high performance Backbone Network Services (vBNS), and the Department of Defense's (DoD's) Defense Research and Engineering Network

(DREN)—and with other high performance research networks such as Abilene (a university/industry partnership), as well as with NSF's Chicago-based Science, Technology, And Research Transit Access Point (STAR TAP) for international connectivity, and connectivity to the geographically remote states of Alaska and Hawaii.

Networking Research Team (NRT)

- The ***Networking Research Team (NRT)*** coordinates agency networking research programs, shares networking research information among Federal agencies, and supports NGI R&D activities. The NRT provides outreach to end users by disseminating networking research information and coordinating activities among users and applications developers.

High Performance Networking Applications Team (HPNAT)

- The ***High Performance Networking Applications Team (HPNAT)*** coordinates Federal R&D to maintain and extend U.S. technological leadership in high performance networking applications in such fields as science and engineering, weather and the environment, biomedicine, and health care.

Internet Security Team (IST)

- The ***Internet Security Team (IST)*** facilitates testing of and experimentation with advanced network security technologies and serves as a forum for the exchange of security requirements and current and emerging security approaches.

NGI: The Next Generation Internet

The Federal NGI Initiative, tightly coupled with base LSN R&D and coordinated by the LSNCG, is creating the technical and infrastructure foundation for a more powerful, flexible, secure, and intelligent Internet in the 21st century. Authorized by Congress in the Next Generation Internet Research Act of 1998, the NGI Initiative:

- Develops, deploys, and demonstrates next-generation technologies that add functionality and improve quality of service (QoS) and performance in network reliability, robustness, and security; differentiated services such as multicast and audio/video; and network management, including allocation of bandwidth. These activities are supported by DARPA's SuperNet, an NGI testbed providing a 1,000-fold increase in end-to-end performance over 1997 Internet speeds, or approximately one gigabyte per second (GBps) for research end users.
- Develops and demonstrates revolutionary applications in enabling technologies such as collaboration technologies, digital libraries, distributed computing, privacy and security, and remote operation and simulation; and in disciplinary applications such as basic science, crisis management, education, the environment, Federal information services, health care, and manufacturing. NGI applications research is supported by an NGI testbed providing a 100-fold increase in end-to-end performance over 1997's Internet, or approximately 100 megabytes per second (MBps) for research end users. This book highlights many of these NGI applications.

SII: Scalable Information Infrastructure

The PITAC recommended a significant new effort and increased funding for networking R&D that includes interoperability and usability. Federal agencies responded to this challenge with a proposed new program in IT R&D. A major component of this program is SII, whose research goal is to develop sophisticated tools and techniques enabling the Internet to grow (scale) while

transparently supporting user demands that include the expanding areas of heterogeneous platforms and mobile and wireless computing. SII research will focus on deeply networked systems, anytime-anywhere connectivity, and network modeling and simulation. Anticipated FY 2001 research "new starts" include activities in agile networking infrastructures, network group collaboration, network security, and prototype access testbeds.

SDP: Software Design and Productivity

Recognizing that software is a key component of IT research, many Government agencies have worked on software design and productivity issues. But the sense of urgency underlined by the PITAC report has led to the creation of the new Software Design and Productivity (SDP) PCA for FY 2001 and the SDP Coordinating Group (SDPCG), which is developing its R&D agenda. SDP R&D is expected to focus on significantly improving the concepts, techniques, and tools that underpin our software infrastructure, including:

- Software engineering of complex systems
- Active software
- Software for autonomous systems
- Large-scale networks of sensors
- Component-based software design and development
- End-user programming
- Empirical software engineering research
- Software for embedded systems
- Model-based integration of embedded software
- Networked embedded systems

SDP activities will help educate Government, academic, and industrial software developers in well-founded and more cost-effective engineering to create useful, efficient, and reliable software. SDP's broad research agenda may overlap with other PCA areas, including HCSS.

HCSS: High Confidence Software and Systems

As mission-critical IT applications expand in national defense, medicine, crises management, aviation, and many other areas where lives and/or sensitive information are at risk, the need for stronger, more secure, and more stress-proof computing systems than exist today grows imperative. HCSS R&D concentrates on technologies that must perform without fail if computing systems are to achieve absolute reliability, safety, security, and survivability. HCSS activities include network and data security, cryptography, information survival, and system stress tolerance.

HCSS research includes the National Security Agency's (NSA's) high assurance computing platform (HACP), security management infrastructure, cryptography, active network defense, secure communications, secure network management, network security engineering, and wireless and optical technologies; NSF programs to develop "no surprise" software that behaves predictably under both normal and stressed conditions and component-based software; DARPA's formal methods (FM) program to develop concepts and tools for a formal science of software development; NIST's Internet security

architectures and Internet Protocol security (IPsec), intrusion detection, authorization management, and software fault analysis and specifications-based testing; and the Office of the Secretary of Defense's (OSD's) University Research Initiative (URI) five-year program to develop fault-tolerant protocols enabling continued network operation despite faults or attacks.

The HCSS Coordinating Group (HCSSCG) works closely with the other PCAs, including LSN and SDP, to coordinate the multiagency research focus on security in networks, software, and systems.

SEW: Social, Economic, and Workforce Implications of IT and IT Workforce Development

SEW was established as a PCA in FY 2000 to support R&D examining how IT is transforming our culture and inspiring innovative models for education and training in IT environments. Succeeding and expanding on the ETHR PCA, SEW reflects a broader research portfolio that focuses on the nature and dynamics of IT impacts on technical, educational, and social systems; the workforce development needs arising from the spiraling demand for workers who are highly skilled in technology; and the growing "digital divide" between Americans with access to information technology and those without. The SEW Coordinating Group (SEWCG) is shaping the new PCA's research agenda and coordinating plans for FY 2001.

FISAC: Federal Information Services and Applications Council

The FISAC assists in the transfer of advanced computing and communication technologies developed by the IT R&D community to agency missions and systems across the Federal government. FISAC disseminates information about Federal IT research priorities, activities, and results to the broad Government community, and advises the Interagency Working Group on IT R&D (IWG/IT R&D) about research needed to provide next and future generation capabilities required by the Federal government. FISAC carries out these activities through its IT for Crises Management, Federal Statistics (FedStats), NGI Applications, and Universal Access Teams, and by participating in NSF's Digital Government program to develop program announcements and solicit proposals for projects that bring computing and information technology researchers together with Federal agencies with significant information services missions. FISAC's work is funded through a combination of the IT R&D budget crosscut, IT R&D agencies, and non-IT R&D organizations.

ASCI: Accelerated Strategic Computing Initiative

DOE's Accelerated Strategic Computing Initiative (ASCI) applies advanced scientific and engineering computing to assuring the performance, safety, and reliability of the Nation's nuclear weapons without physical testing. Activities include building high end systems that scale to at least 30 trillion operations per second; developing high performance storage technologies; creating a visual interactive environment for weapons simulation (VIEWS); managing simulation data; developing a problem solving environment that includes distributed computing and scalable input/output; and the Academic Strategic Alliances Program (ASAP) to accelerate simulation science.

PITAC: President's Information Technology Advisory Committee

Comprising corporate leaders and research scientists from business and academic institutions, the 25-member PITAC was established by President Clinton in February 1997, as authorized by the High-Performance Computing Act of 1991, to provide expert independent guidance to the Federal government on maintaining America's preeminence in high performance computing and communications, information technology, and the Next Generation Internet.

The recommendations of the PITAC's influential February 1999 report, "Information Technology Research: Investing in Our Future," prompted an expanded focus in the Federal IT portfolio on key research areas, including high end computing, scalable information infrastructure, software development, and the socioeconomic impact of IT. The PITAC conducts annual reviews of the NGI Initiative, as mandated by the Next Generation Internet Research Act of 1998, and was asked by the White House to review its FY 2001 IT R&D budget proposals. The PITAC's February 2000 study, "Resolving the Digital Divide: Information, Access, and Opportunity," is helping shape Federal programs to address this issue. In FY 2000, PITAC panels are examining IT research needs in the digital divide, digital libraries, government, health care, international issues, learning, and open source software.

IT R&D budget and coordination

The proposed FY 2001 multiagency IT R&D budget is $2,137 million, representing a 38 percent increase over the estimated $1,546 million in FY 2000. Leadership for the IT R&D Programs is provided by the Senior Principals Group for IT R&D. This senior management group is chaired by the Assistant to the President for Science and Technology, who is Director of the White House Office of Science and Technology Policy (OSTP). The Interagency Working Group on IT R&D serves as the internal deliberative organization for the Senior Principals Group, providing policy, program, and budget guidance for the Executive Branch. The IWG works through the PCA Coordinating Groups.

NCO/CIC: National Coordination Office for Computing, Information, and Communications

The NCO is charged by OSTP with ensuring coordination of multiagency Federal IT research by providing technical and administrative support to the IWG, assisting in the preparation of multiagency planning, budget, and evaluation materials, and supporting other activities related to the IT R&D Programs. The NCO serves as the central source of information and documentation about Federal IT research activities. Working collaboratively, the NCO and the agencies participating in the IT R&D Programs craft the blueprints and implementation strategies for the comprehensive Federal effort to pioneer the next generation of advanced computing, network communications, and information technologies.

As the central points of contact for Federal IT initiatives, the IWG and the NCO meet often with representatives from Congress, Federal, state, and local organizations, academia, industry, professional societies, foreign organizations, and others to discuss the Government's IT programs and exchange technical and programmatic information. The NCO also supports the activities of the PITAC.

Each year, the NCO responds to thousands of inquiries with Web, print, and video information including IT R&D and PITAC publications, Congressional testimony, and meeting materials.

This report, prepared by the NCO in cooperation with participating Federal agencies, documents current and planned Federal IT R&D activities, highlighting representative FY 2000 accomplishments, major FY 2001 proposals, and the budget crosscut.

Information on the Web

Copies of HPCC, IT R&D, NGI, and PITAC publications, links to participating agency and related Web sites, and this report can be found at:

http://www.itrd.gov/ and
http://www.ngi.gov/

High End Computing and Computation

—Introduction and Overview

Federally supported High End Computing and Computation (HECC) programs conduct leading-edge research and development (R&D) in large, high performance computational systems, including hardware, software, architecture, and applications. HECC R&D extends the state of the art in computing systems, applications, and high end infrastructure to achieve the scientific, technical, and information management breakthroughs necessary to keep the U.S. in the forefront of the 21st century information technology (IT) revolution.

Federal HECC research continues to pave the way for revolutionary advances in science, technology, and national security and has become an important tool in the design and development of military and commercial products ranging from submarines and aircraft to automobiles. HECC researchers develop computation-intensive algorithms and software to model and simulate complex physical, chemical, and biological systems; information-intensive science and engineering applications; management and use of huge, complex information bases; and advanced concepts in quantum, biological, and optical computing.

HECC research continues to have substantial economic benefits. The President's Information Technology Advisory Committee (PITAC) noted that many of the underlying component technologies in today's mid-level and desktop computers and Internet communications devices were derived from yesterday's high end computers. Without the Federally funded, high-risk HECC research of past decades, today's personal computers, networks, and cell-phone infrastructures would not perform as well and U.S. economic growth would not be as robust.

Thus, the Federal HECC agenda is motivated by government, industrial, and scientific user applications demanding increasing levels of performance. Continuing U.S. leadership in critical areas of Federal agency mission responsibilities, such as national security, industrial production, and fundamental science, is at stake since they depend on U.S. leadership in high end computing. Areas of concern include:

- Modeling and simulation in biology, environmental sciences, materials sciences, and physics
- Data fusion and knowledge engineering demands of the new millennium across all elements of the national, scientific, and industrial sectors

- Complex and computationally intensive national security applications in weapons systems design and maintenance, cryptology, and battlespace dominance
- Industrial leadership in aerospace, automobile and other product design and in the manufacturing, energy, pharmaceutical, and chemical sectors

The current aims of HECC are to:

- Improve the usability and effectiveness of teraflops-scale systems (that is, those that can perform on the order of 10^{12} to 10^{14} operations per second)
- Pursue leading-edge research for future generations of computing, such as petascale computers (10^{15} to 10^{17} operations per second) and exabyte storage systems (10^{18} to 10^{20} bytes of information storage), based on current and advanced device technologies and subsystem components and innovative architectures
- Demonstrate prototype systems

Accomplishing these objectives requires investments in systems-software and applications-software research as well as efforts at the various layers of hardware and architecture from the smallest components through entire systems.

The High End Computing and Computation Coordinating Group (HECCCG) coordinates the HECC program with the support of the National Coordination Office (NCO) for Computing, Information, and Communications (CIC). The HECCCG facilitates interagency collaborations, identifies HECC R&D needs, and provides mechanisms for Federal cooperation with Government laboratories, universities, industry, and other private-sector organizations or entities. The Defense Advanced Research Projects Agency (DARPA), the Department of Energy (DOE), the Environmental Protection Agency (EPA), the National Aeronautics and Space Administration (NASA), the National Institutes of Health (NIH), the National Institute of Standards and Technology (NIST), the National Oceanic and Atmospheric Administration (NOAA), the National Security Agency (NSA), and the National Science Foundation (NSF) participate in the HECCCG.

High End Computing
—Infrastructure and Applications

Introduction

High end computing infrastructure and applications are needed for the advancement of science and technology, for large-scale information management approaches to product and process design, and for the support of national security—all of which are within or directly support the missions of various Federal agencies. HEC I&A encompasses projects to develop the software that underlies high end applications. Such software tends to support multiple applications or provide an application-level or middleware infrastructure where applications can run. This section describes that software as well as select high end applications. The IT R&D facilities section (page 34) describes HECC infrastructure facilities in detail.

Computing environments and toolkits

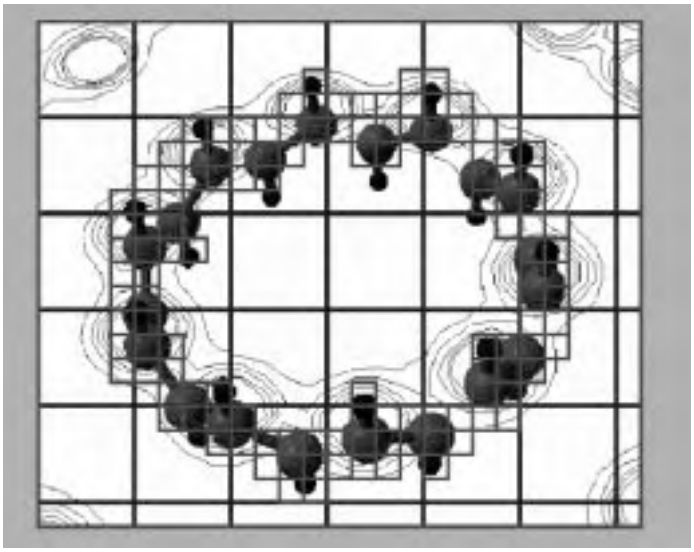
Architecture adaptive computing environment (aCe)

NASA's aCe is a data-parallel computing environment being designed to improve the adaptability of algorithms to diverse architectures. aCe will encourage programmers to implement applications on parallel architectures by assuring them that future architectures will run their applications with minimal modification, and will encourage computer architects to develop new architectures by providing an easily implemented software development environment and a library of test applications.

aCe will facilitate the ability of programmers to:

- Allow easy, architecture-independent expression of algorithms
- Port algorithms among diverse computer architectures
- Adapt algorithms to different architectures
- Easily and efficiently implement algorithms on diverse architectures
- Optimize algorithms on diverse computing architectures
- Develop applications on heterogeneous computing environments
- Develop programming environments for new computer architectures

Structured parallel execution is the concept behind aCe. Beginning with a virtual architecture that reflects the spatial organization of an algorithm, a programmer develops software reflecting the algorithm's temporal organization. Today, aCe has been implemented for a superset of C for Linux and Linux with parallel virtual machine (PVM), and has a Linux and Tera/Cray debugger.

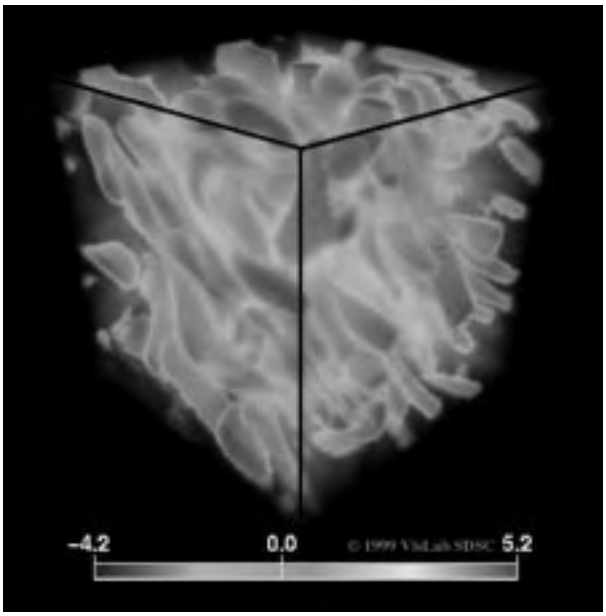


This KeLP application, pictured left, illustrates the results of a hierarchical adaptive mesh. The larger dark areas represent regions of high error, such as atomic nuclei in materials design applications.

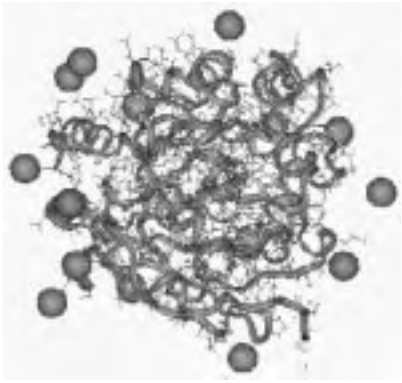
Kernel lattice parallelism (KeLP)

KeLP, developed under NSF funding at UCSD and NPACI, is a framework for implementing portable scientific applications on distributed memory parallel computing systems. KeLP supports coarse-grained data parallelism in distributed collections of structured data blocks. It is intended for applications with special needs, such as to adapt to data-dependent or hardware-dependent runtime conditions. KeLP is currently used in full-scale applications including subsurface modeling, turbulence studies, and first principles simulation of real materials.

The KeLP infrastructure, implemented as a C++ class library, provides high-level tools to computational scientists, allowing them to concentrate on applications and mathematics instead of low-level data distribution and interprocessor communication concerns, enabling them to develop complicated applications in a fraction of the time formerly required.



The visualization in this KeLP application—done at San Diego Supercomputer Center's (SDSC's) Advanced Scientific Visualization Lab—depicts small-scale structure in terms of the relative significance of rotational vs. straining motion. The red areas correspond to rotation-dominated (high vorticity) regions that concentrate into tube-like structures. The green regions are those with comparable rotation and strain, which tend to form sheets and surround the tube-like structures. The blue areas correspond to strain-dominated regions that indicate locally high-energy dissipation. This direct numerical simulation is the numerical solution to the exact 3-D time-dependent Navier-Stokes equations governing fluid motion.



This image is from a simulation of a halo-alkane dehalogenase enzyme after 100 picoseconds. The simulation was performed by molecular dynamics software that uses global arrays, part of DOE's ACTS toolkit.

Parallel algorithms and software for irregular scientific applications

This NSF-funded project at New Mexico State University is applying techniques from parallel computing and computational geometry to develop theoretically sound and practically efficient parallel algorithms for a class of irregular scientific applications that depend upon interactions among entities such as atoms located in 2-D or 3-D space. Researchers will explore techniques for efficient parallel execution of such applications and will develop software to aid applications programming in this environment. Applications include the N-body problem useful in astrophysics, plasma physics, molecular and fluid dynamics, computer graphics, numerical complex analysis, and protein-accessible surface area calculations for computational molecular biology.

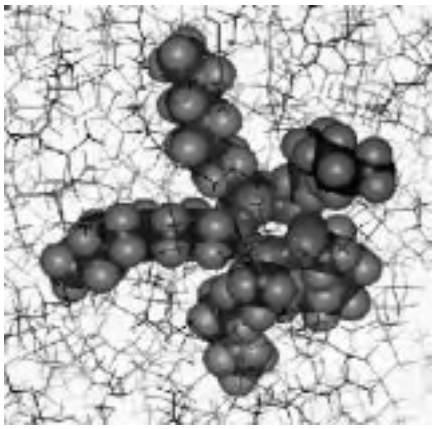
Advanced computational testing and simulation (ACTS) toolkit

DOE's ACTS toolkit is a set of software tools to help programmers write high performance scientific codes for parallel computers. It focuses primarily on software used inside an application, rather than software used to develop an application. Consisting primarily of software libraries, ACTS tools are designed to run on distributed memory parallel computing systems using the message passing interface (MPI) for communication, with portability and performance important considerations in their design. The tools fall into four broad categories:

- Numerical tools implement numerical methods and include sparse linear system solvers, ordinary differential equation solvers, and others.
- Frameworks provide infrastructure to manage some of the complexity of parallel computing, such as distributing arrays and communicating boundary information.
- Execution support provides application-level tools including performance analysis and visualization support.
- Developer support is provided transparently for tool developers.

Scalable visualization toolkits

Scientists routinely use desktop computers to visualize 100-megabyte data sets. But biomedical researchers, astronomers, oceanographers, and other scientists often need to analyze and visualize hundreds of gigabytes of data at a time. These files are so large that only a supercomputer can process them. Yet even a supercomputer's memory has difficulty accommodating the data. NPACI researchers are creating versatile supercomputer-based tools for rendering, visualizing, and interacting with very large data sets from a variety of scientific disciplines. These scalable visualization toolkits will support the next generation



This image displays the results of a simulation of liquid octanol. The 1,000-step simulation, which uses the ACTS molecular dynamics software NWArgos, included 216,000 atoms and was performed on a 1,300-node Cray T3E-900.

of large-scale simulations on teraflops computers, spurring new collaborations within and between scientific disciplines by providing a graphical user environment for sharing data and insights.

Tools to explore geometric complexity

Computational geometry began with the promise of unifying efforts to solve geometric problems in statistics, biology, robot motion planning, graphics, image analysis, virtual reality, and data mining. In two decades, the field has produced a number of tools for solving geometric algorithmic problems. A recurring feature in the design and analysis of geometric problems is the strong link between the computational and combinatorial aspects of the questions under investigation. Understanding the combinatorial geometry behind the problem is fundamental to being able to find an efficient solution. This NSF-sponsored project at the Polytechnic University of New York will explore combinatorial problems arising in geometric contexts to develop new tools and refine tools already available to design and analyze geometric algorithms, with a goal of constructing simpler and more efficient algorithms. The project is also investigating techniques to more realistically estimate the behavior of geometric algorithms on typical inputs.

NIST staff members discussing the result of a micromagnetic simulation.

Modeling tools

NASA Earth system modeling framework

In software engineering, a software architecture and a set of software entities (objects, programs, routines, interface definitions, type systems, and so forth) that allow the construction, storage, management, and aggregation of software components are called a "framework." Frameworks are used to:

- Foster reusability among software components and portability among computing architectures
- Reduce the time needed to modify research applications software
- Structure systems to better manage evolving software
- Enable software exchange among major research centers

This multiyear project, begun in FY 2000, will improve the interoperability, performance, and manageability of NASA's Earth and Space Science (ESS) applications through the development of a common Earth system modeling framework (ESMF). The overall goal of the ESS project is to demonstrate the potential afforded by balanced teraflops systems' performance to further our understanding of and ability to predict the dynamic interaction of physical, chemical, and biological processes affecting the Earth, the solar-terrestrial environment, and the universe. (For more about ESS, please see page 20.)

Micromagnetic modeling

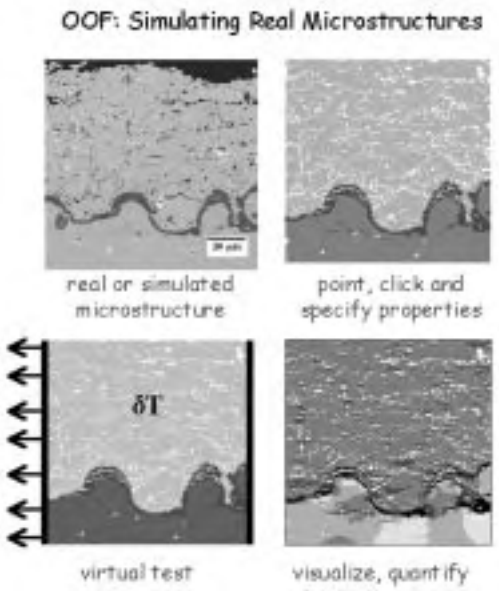
NIST's micromagnetic modeling project is developing computational tools for accurate and efficient micromagnetic calculations—essential in the magnetic disk drive industry to achieve higher densities and faster read-write times. NIST has released OOMMF, a modular object-oriented micromagnetic modeling framework and reference software that allows software developers to swap their code in and out as desired. OOMMF will help establish a baseline level of competence in 2-D modeling and compare competing algorithmic components. A 3-D version of the code is under development.

Modeling realistic material microstructures

The behavior of a material on the macroscopic scale depends to a large extent on its microstructure—the complex ensemble of polycrystalline grains, second phases, cracks, pores, and other features that are large compared to atomic sizes. Modeling such structures is challenging due, in part, to their complicated geometries. NIST has developed the OOF object-oriented finite-element software to analyze material microstructures and simulate physical property measurements. OOF allows materials scientists to study the influence of microstructure on a material's macroscopic properties through an easy-to-use graphical interface. By applying stresses, strains, or temperature changes, the user can measure the effective macroscopic material behavior or examine internal stress, strain, and energy density distributions. By modifying a microscopic material property, the user can find the effect of that property on the macroscopic behavior; by modifying the microstructure, the effect of geometry on a particular material can be determined. OOF is being extended to handle other properties in addition to thermoelasticity. The software won a 1999 Technology of the Year Award from *Industry Week* magazine.

Numerical and data manipulation techniques for environmental modeling

The primary objective of EPA's numerical and data manipulation techniques program is to improve the performance of key numerical algorithms that form the computational foundation of environmental models. This research develops



The graphic at right illustrates the steps in the use of NIST's object-oriented finite-element software OOF, which allows materials scientists to study the influence of microstructure on a material's macroscopic behavior by means of a graphical interface.

and evaluates parallel computing techniques encompassing interconnected workstations, vector and parallel supercomputers, parallel software and algorithms, and communication to determine the most effective approach to complex, multipollutant, and cross-media environmental modeling. Fundamental research is also conducted on computational techniques to quantify uncertainty as an integral part of the numerical computation.

HEC applications

HEC applications research harnesses the raw speed and data storage capacity of advanced computing platforms to science's most data-intensive, complex, and challenging problems, such as the design and properties of materials in weapons, aerospace, and industrial systems; the shapes and processes of biomolecular structures; and synthesis and analysis of terascale data sets. Computationally intensive high end applications include modeling, 3-D visualization, and tools for data mining and data fusion.

Biomedical applications

Neuroscience imaging

During the next century there is a real possibility that we will discover in detail how the brain works and how to treat or prevent common neurological diseases and traumas. Developments in modern computer-aided microscopes and advances in high performance computing promise to uncover new information about the structural and functional dynamics of the nervous system. Neuroscientists are involved in research covering a wide range of scales, from modeling molecular events and subcellular organelles to mapping brain systems. They are also investigating the ways in which single neurons and small networks of neurons process and store information. Newly possible detailed models of single neurons are being used to model the complex properties of neurons and neuronal networks. Breakthroughs in optical imaging and image processing provide opportunities for deriving information about the 3-D relationships among biological structures, and structure-function work is moving into 4-D (3-D plus time) imaging.

MCell

A growing interest in neuron modeling parallels the increasing experimental evidence that the nervous system is extremely complex. In fact, modeling is as

essential as laboratory experimentation in understanding structure-function relationships in the brain. The models may become as complex as the nervous system itself, thereby requiring use of advanced computing. In National Partnership for Advanced Computational Infrastructure (NPACI)-supported neuroscience research, both widely used and newly developed neuron modeling systems are being extended and linked to large-scale, high performance capabilities.

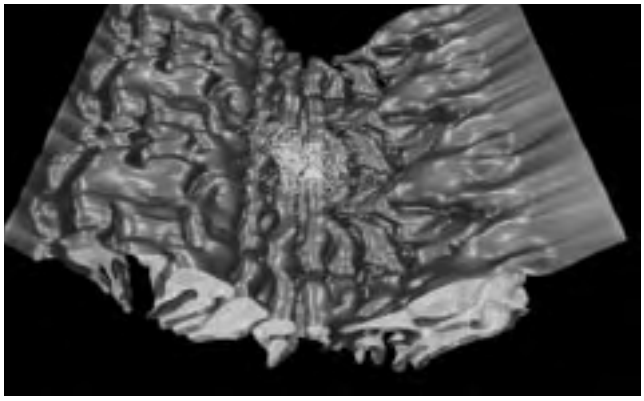
The ongoing NPACI MCell project has developed a general Monte Carlo (pseudo-random number-based) simulator of cellular microphysiology. Biological structures such as neurons show tremendous complexity and diversity at the subcellular level. For example, a single cubic millimeter of cerebral cortex may contain on the order of five billion interdigitated synapses of different shapes and sizes, and subcellular communication is based on a wide variety of chemical signaling pathways. A process like synaptic transmission encompasses neurotransmitter and neuromodulator molecules, proteins involved with exo- and endocytosis, receptor proteins, transport proteins, and oxidative and hydrolytic enzymes.

MCell incorporates high-resolution ultrastructure into models of ligand diffusion and signaling. Ligands and effectors—reaction mechanisms—and surfaces on which reactions take place are specified by the modeler, who uses the MCell model description language to build the simulation objects. MCell then carries out the simulation for a specified number of iterations, after which numerical results and images can be produced. Optimized software for widely used and newly developed models is being ported to the University of California-San Diego's (UCSD's) Cray T3E and IBM teraflops systems.

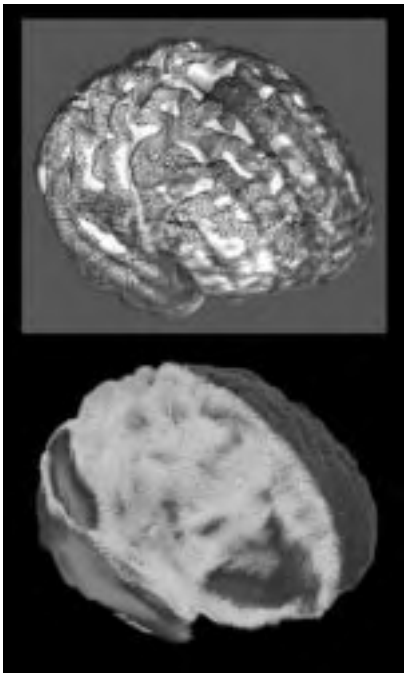
MCell has also been tested on a 40-machine NetSolve cluster. A collaborative effort among scientists at UCSD, the University of Tennessee, and Oak Ridge National Laboratories (ORNL), NetSolve turns a loosely associated collection of machines into a fault-tolerant, client-server compute cluster. The initial test of MCell on the NetSolve cluster demonstrated the need for a distributed file-checking mechanism that would allow NetSolve to support larger MCell runs.

Protein folding

Understanding how proteins form may yield exciting medical and scientific possibilities. In nature's ultimate origami, cells use information encoded in genes to construct a long chain of amino acids that compacts into a tangle of loops, helices, and sheets. A protein's unique geometry enables it to interact with other molecules and do the body's biochemical heavy lifting, regulating digestion, for example, or turning genes on and off during fetal development. Because of its



Simulation of a synapse between a nerve cell (not shown) and a mouse sternomastoid muscle cell. The neurotransmitter acetylcholine diffuses from a synaptic vesicle to activate receptors (pictured as a cloud of dots) on the muscle membrane. Snapshot is at 300 microseconds at peak activation.



Researchers at the University of California-Los Angeles's (UCLA's) Laboratory of Neuro-Imaging (LONI) are building population-based digital brain atlases to discover how brain structures are altered by disease. In 3-D maps of variability from an average cerebral cortex surface derived from 26 Alzheimer's disease patients, individual variations in brain structure are calculated based on the amount of deformation needed to drive each subject's convolution pattern into correspondence with the group average. Surface matching figures are computed using fluid flow equations with more than 100 million parameters. This requires parallel processing and very high memory capacity.

complexity, however, simulating protein folding and proteins' interactions with other molecules is one of the toughest problems in computational biology. Solvation models—so-called because water is the natural environment for proteins—calculate the forces acting between every possible pairing of the atoms in the protein as well as the surrounding solution, but such accuracy comes at a high cost. Simulating, with full atomic detail, just one-millionth of a second of the folding process in a small protein can take months of computation, even on today's high performance computers. "Cutoff models" that include only pairs of nearby atoms miss significant effects from greater distances. Methods that group atoms—originally developed in the 1980s to study interactions among stars—are yielding more accurate results. Research in this area is funded by NIH and is being carried out by National Computational Science Alliance (Alliance) scientists.

Emerge: Portable biomedical information retrieval and fusion

National Center for Supercomputing Applications (NCSA) researchers supported by NIH's National Cancer Institute (NCI) and NSF are addressing a recurring science problem—finding and relating information scattered across many data sources. To pinpoint the defective genes that cause cancer cells to run amok, for example, biological researchers comb the Internet weekly, scanning vast online databases for clues. The next essential clue to a tumor suppressor could lie hidden in the billions of bases of human DNA being archived in GenBank, or in any of dozens of other online databases. A user needs a skilled translator such as Emerge, a portable collection of information-retrieval programs developed at NCSA. Emerge translates a single query into the idioms of separate databases, collects the results, translates them back to a common computer language, and displays them on a user's screen.

A cancer researcher, for example, could enter the phrase "small-cell lung cancer" into a form displayed by a Web browser. The query is converted by Emerge into a data format called Extensible Markup Language (XML), a versatile offshoot of HyperText Markup Language (HTML) that may soon

become the foundational data format on the Web. The query is then sent to Gazebo, the heart of the Emerge software system. Gazebo translates the query into Z39.50, a standard language recognized by many library catalogs and databases of scientific literature.

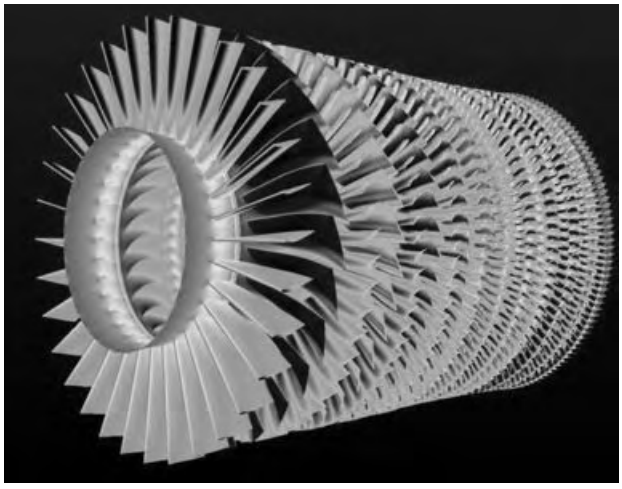
Because the cancer literature often uses synonymous terms to describe a single concept, NCI has funded an interface to the Unified Medical Language System (UMLS) metathesaurus developed over the past 10 years by NIH's National Library of Medicine (NLM) to integrate collections of medical terminologies. By next year, NCI plans to link cancer-related terms with Emerge, allowing patients to click on a highlighted term and search for related information from a universe of cancer databases. Physicians and cancer patients and their families may also discover, with a few clicks of the mouse, how many people suffer from a particular type of cancer, the status of cancer-related legislation, and information on potential drug treatments. Emerge is also part of a comprehensive science information system that has up-to-date news about research grants.

Aerospace applications

Computational Aerosciences (CAS)

The NASA Computational Aerosciences (CAS) project is working with industry toward the goal of trimming the time and cost of designing airplanes. Researchers propose to develop the high end computing hardware and systems and applications software to enable 1,000x speed-ups in systems performance. NASA-supported researchers have demonstrated a full compressor simulation in 15 hours—400 times faster than was possible in 1992. In overnight supercomputing calculations, numerical propulsion system simulation (NPSS) software will simulate a full range of engine functions. These simulations let designers try out potential changes without building and testing real hardware.

CAS is developing a framework to enable multidisciplinary design optimization of complete aircraft, which requires enormous computing resources. By integrating two Silicon Graphics, Inc. (SGI) Origin 2000 systems, CAS created a 100-gigaflop testbed that presents a single system image with global shared memory. CAS supports the development of cost-effective, high end computing solutions. For example, CAS found a 92 percent cost savings for certain design applications using 10 workstations as opposed to a single-processor supercomputer.



A simulation of the GE90's high-pressure compressor. NASA-supported CAS researchers have demonstrated a full compressor simulation in 15 hours—400 times faster than was possible in 1992.

Earth and Space Science (ESS)

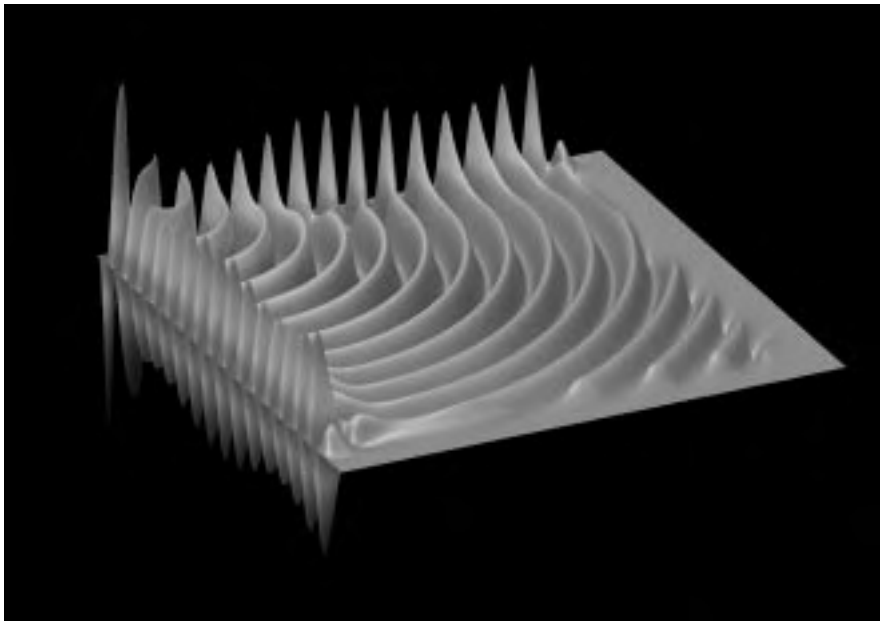
NASA's ESS R&D applies high end computing technologies to topics as cosmic as the collisions of ultra-dense neutron stars and as compelling as the fate of the Earth's climate. ESS, which is managed by NASA's Goddard Space Flight Center (GSFC) with the Jet Propulsion Laboratory (JPL), encompasses the following research areas:

- *Earth imaging science* employs multiple supercomputing systems to process and visualize satellite-collected radar data to monitor regional changes in the Earth's environment. A detailed imaging study that followed seasonal changes in the Amazon basin was recently completed.
- *Relativistic astrophysics* combines fluid dynamics and general relativity to probe the violent merger of ultra-dense neutron stars. These mergers may be the cause of still-mysterious hyper-energetic gamma-ray bursts.
- *Simulations of the Earth's interior* will gain insight into the chaotic processes that drive changes in the planet's interior.
- *Four-dimensional data assimilation* will build comprehensive views of the atmosphere by merging observations with climate models.
- *Earth system modeling* seeks to understand the Earth's climate via complex high-resolution models of coupled atmospheric/oceanic circulation and chemistry.
- *Fluids in microgravity* R&D studies fluids in low-gravity environments to simulate space-based manufacturing, life support, and safety systems.
- *Convection and dynamos in stars* focuses on the most fundamental and least understood turbulent processes in the interior of stars such as the sun.
- *Multiscale heliosphere modeling* uses computational studies to probe interactions of the solar wind with the local space environment.
- *Solar activity/heliospheric dynamics* investigates the tangled 3-D magnetic structures in the sun's corona or outer atmosphere. These structures play a key role in the physics of solar activity.

Advanced chemistry application

Understanding combustion

Most of the energy the world uses comes from the combustion of fossil fuels. Increases in computing power over the next few decades will make possible predictive computer models that will enable us to understand the complex interactions of fluid flow, chemistry, surface physics, and materials properties that determine the efficiency of combustion devices as well as the output of undesirable combustion byproducts such as soot and NO_x. In the past year, researchers at DOE's Lawrence Berkeley National Laboratory (LBNL) have brought to light aspects of methane combustion that have confounded scientists for a number of years. This research combined advanced adaptive mesh refinement technologies and a new understanding of chemical reaction rates to yield simulations that agree closely with experiments. Future extensions of the research must incorporate the chemistry of more complex hydrocarbons, such as diesel fuel, which have thousands of reaction pathways, as well as more realistic surface physics and more complex geometries.



Collaborators at DOE's LBNL and LLNL and the University of California-Davis have used supercomputers to obtain a complete solution of the ionization of a hydrogen atom by collision with an electron, the simplest nontrivial example of the problem's last unsolved component. Pictured at left is a representative radial wave function of two electrons scattered in the collision of an electron with a hydrogen atom.

Quantum physics application

Ionization by electron impact

For over half a century, theorists have tried and failed to provide a complete solution to scattering in a quantum system of three charged particles, one of the most fundamental phenomena in atomic physics. Such interactions abound. Ionization by electron impact, for example, is responsible for the glow of fluorescent lights and the ion beams that engrave silicon chips. Collaborators at DOE's LBNL and Lawrence Livermore National Laboratory (LLNL), and the University of California-Davis recently used supercomputers to solve the ionization of a hydrogen atom by collision with an electron—the simplest nontrivial example of the problem's last unsolved component. The breakthrough employs a mathematical transformation of the Schrödinger wave equation in which the wave functions of outgoing particles vanish at large distances from the nucleus rather than extending to infinity.

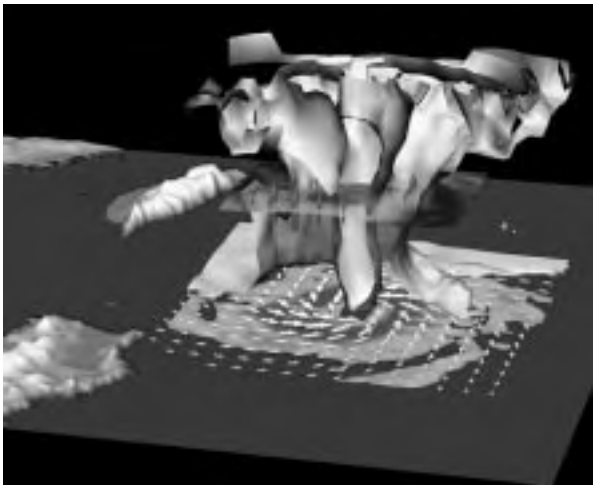
Weather applications

Hurricane intensity prediction

As part of NOAA's efforts to understand and forecast climate and weather, researchers at the agency's Geophysical Fluid Dynamics Laboratory (GFDL) seek to:

- Understand the genesis, development, and decay of tropical disturbances by investigating thermo-hydrodynamical processes using numerical simulation models
- Study small-scale features of hurricane systems such as the collective role of deep convection, the exchange of physical quantities at the lower boundary, and the formation of organized spiral bands
- Investigate the ability of numerical models to predict hurricane movement and intensity and transition those models to operational use

While the GFDL models are excellent in predicting intensities of weak to moderate hurricanes, better prediction of surface wind intensities in stronger



NOAA-supported researchers have simulated samples of hurricanes from today's climate and a projected greenhouse gas-warmed climate by linking information from GFDL's global climate model into the high-resolution GFDL hurricane prediction model (left). This now operational model has been used successfully by NOAA's National Centers for Environmental Prediction to predict tropical storm paths over the last several hurricane seasons.

hurricanes is anticipated in FY 2000-FY 2001 due to increased computing power, a result of HEC R&D, that allows hurricane models to operate at higher grid resolutions, account for asymmetries in storms, and improve physical parameterization. During FY 2000, developmental work to improve the hurricane model initialization, ocean interaction, model physics, and resolution continues; case studies will be used to evaluate the models' impact on forecasting skills. Work on assimilating more data into the forecast and analysis system will continue. The effects of evaporation of rain and sea spray, together with dissipative heating, will be evaluated.

Hurricanes and global warming

The strongest hurricanes in the present climate may be upstaged by even more intense hurricanes over the next century if the Earth's climate continues to be warmed by increasing levels of greenhouse gases in the atmosphere. Most hurricanes do not reach their maximum potential intensity before weakening over land or cooler ocean regions. However, those storms that do approach their upper-limit intensity are expected to be slightly stronger in the warmer climate due to the higher sea surface temperatures.

NOAA researchers have simulated samples of hurricanes from the present-day climate and from a projected greenhouse gas-warmed climate by linking information from GFDL's global climate model into the high-resolution GFDL hurricane prediction model. This is the operational model that has been used by NOAA's National Centers for Environmental Prediction to predict tropical storm tracks for the last several hurricane seasons. The simulation projects that wind speeds in the northwest tropical Pacific will increase by 5-12 percent if tropical sea surfaces warm by a little more than 2 degrees Centigrade. This study represents the first use of an operational model to study a phenomenon that was theorized a decade ago. It illustrates the use of high performance computing to investigate the potential impact of global climate change on weather systems.

High End Computing

—Research and Development

Overview

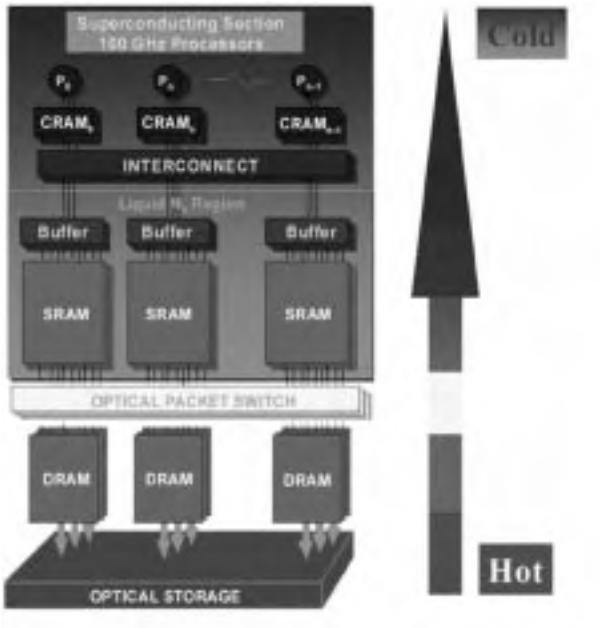
Assembling efficiently usable systems from hardware components—such as processors, memories, input/output devices, and switches—and from software components—such as languages, compilers, optimizers, schedulers, and memory managers—is a complex endeavor that must encompass a vast number of interactions to produce an architecture that, in turn, enables the development of a successful leading-edge system. These interactions include those among the components and their arrangement, power and cooling requirements, the hardware and the applications, and the systems software with both the underlying hardware and the applications. Projects described in this section illustrate the enormous range of approaches to high performance computing including workstation clusters, distributed mass storage, and computational grids.

Hybrid technology multithreaded (HTMT) architecture

Petaflops-scale computing faces severe challenges of cost, size, power, complexity, reliability, efficiency, generality, and programmability. Even with the extraordinary progress of complementary metal oxide semiconductor (CMOS) technology and massively parallel processor (MPP) architectures—systems based on conventional technologies that can sustain throughputs beyond a petaflop—petaflops-scale computing may not be feasible until after 2010.

The hybrid technology multithreaded (HTMT) interdisciplinary research project—supported by DARPA, NASA, NSA, and NSF—is attempting to exploit the superior properties of advanced device technologies to overcome current barriers to success. Such technologies include rapid single flux quantum (RSFQ) superconductor logic, optical communications and holographic storage, and processor-in-memory SRAM and DRAM chips. Superconductor logic is capable of 100 times the speed and power efficiency of conventional processors, while fiber optical communications using time division and wave division multiplexing can exceed wire-based channel bandwidth by a factor of 100 or more. Optical storage density and power efficiency using holographic photorefractive techniques may provide an order of magnitude advantage over today's semiconductor memory at comparable bandwidths. But efficient computation requires effective resource management in the presence of extremes in latency and parallelism. The HTMT architecture has been devised to incorporate adaptive latency-tolerant mechanisms based on multithreaded processors and an innovative memory-based proactive task management scheme called "percolation."

A design study, simulation, and analysis indicate that practical petaflops-scale computing based on the HTMT model is feasible by 2005, followed by rapid advances leading beyond 10 petaflops. In FY 2000, researchers are compiling a complete report on the sub-elements of HTMT, including preliminary tests,



An overview of the HTMT memory architecture.

data, simulations, sizing, produceability, performance estimates for a full system, cost estimates, and recommendations, if warranted, for the follow-on work. Beginning in FY 2000, a 15-month program will verify critical technologies, further simulate architecture to generate more accurate specification of the proposed hardware, and study the programming and execution of the proposed system. This effort may lead to the construction of a prototype machine.

Beowulf : High performance computing with workstation clusters and Linux

In the NASA-funded Beowulf project—which employs clusters of workstations to achieve high performance computing—every node is responsible for running its own copy of the operating system kernel, and nodes are generally sovereign and autonomous at the kernel level. The Beowulf software environment is implemented as an add-on to commercially available, royalty-free Linux distributions including all the software needed for a networked workstation. Beowulf has been migrated to the Red Hat Linux distribution that includes programming environments developed at various sites and development library packages. To present a more uniform system image to users and applications, researchers have extended the Linux kernel to allow a loose ensemble of nodes to participate in a number of global namespaces. These extensions will cause little increase in kernel size or complexity and, most importantly, will have negligible impact on individual processor performance.

Weather forecasting in a mixed distributed shared memory environment

NOAA has developed a portable message passing interface (MPI)-based version of its operational weather forecasting codes. To adapt the vector codes to NOAA's new IBM SP supercomputer, scientists converted the codes to column structure to optimize cache usage. A single column version of the global spectral model (GSM) produced by this conversion now uses observational input to diagnose differences in model physics. NOAA-supported researchers are exploring the concurrent use of MPI and OpenMP as they seek the most efficient programming model to use on future generation mixed distributed shared memory computing systems.

MVICH-MPI for virtual interface architecture

MVICH is a DOE/National Energy Research Supercomputer Center (NERSC)-funded project to provide portable high performance communications for cluster computing. It is an implementation of MPI for the virtual interface architecture (VIA). VIA is an industry-standard interface for system area networks (networks for clusters) that provides protected zero-copy user-space inter-process communication. MVICH provides a high performance MPI available on a wide range of commodity networks. MVICH is being developed at LBNL and is distributed with an open source license. It is part of the same project that is developing M-VIA, an implementation of VIA for Linux. MVICH implements four protocols to maximize performance over a range of message sizes. Over the next year, the project plans to add:

- Support for unreliable VIA. MVICH will handle dropped and out-of-order packets with negligible additional latency.
- Improved support for asynchronous communication using multiple threads.
- Thread safety. This requires a thread-safe VIA implementation as well as additional support for thread safety in MVICH.

Distributed-parallel storage system (DPSS)

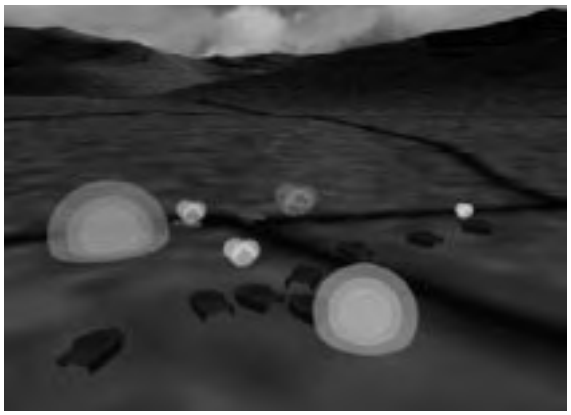
Modern scientific computing involves organizing, moving, visualizing, and analyzing massive amounts of data at multiple sites around the world. The distributed-parallel storage system (DPSS) provides scalable high performance distributed parallel architecture and data handling for building high end storage systems from low-cost commodity hardware components. It was developed in the DARPA-funded Multidimensional Applications and Gigabit Internetwork Consortium (MAGIC) testbed, with additional support from DOE. The DPSS architecture is a network striped disk array that lets applications determine optimal data layout, replication and/or code redundancy, security policy, and dynamic reconfiguration. This technology provides an economical, widely distributed architecture for caching large amounts of data that can potentially be used by many different users. Current performance results are 980 Mbps across a local area network (LAN) and 570 Mbps across a wide area network (WAN). DPSS is used in Terravision, the terrain navigation application for which it was designed; in an image viewing application from the U.S. Geological Survey's (USGS's) Earth Resources Observation Systems (EROS) data center; and in high-energy physics and medical imaging applications.

Globus

We do not think twice about accessing Web pages that are spread across the globe. The goal of Globus R&D is to bring about a similar revolution in computation. Globus—a joint project of DOE's Argonne National Laboratory (ANL) and the University of Southern California's (USC's) Information Sciences Institute funded by DARPA, DOE, and NSF—is developing the fundamental technologies to build computational grids, execution environments that enable an application to integrate geographically distributed instruments, displays, and computational and information resources. Such computations may link tens or hundreds of these resources.

Globus focuses on:

- Research into basic problems in resource management, security, fault tolerance, and algorithms



This image was generated by Synthetic Forces Express, a large-scale simulation developed at the California Institute of Technology (CalTech) that uses Globus to provide access to multiple supercomputer resources. SF Express conducted a record-breaking 100,298-vehicle simulation—the largest distributed, interactive battlefield simulation to date—executing it on 1,386 processors distributed over 13 computers at nine sites spanning seven time zones.

- Tools, such as prototype software that can run on a range of platforms
- Large-scale testbeds
- Large-scale grid-enabled applications, developed in collaboration with application scientists

Participants in the Globus Ubiquitous Supercomputing Testbed Organization (GUSTO) are testing Globus concepts on a global scale. GUSTO currently spans more than 40 institutions worldwide and includes some of the largest computers in the world. ANL and USC researchers have won the Global Information Infrastructure Next Generation Award for their development of GUSTO's Globus-based prototype for future computational grids. The GII awards recognize and promote best practices and new models in Internet and network technologies. The Next Generation category targets exemplary uses of the information infrastructure that demonstrate its direction and future potential for cutting-edge information and communications technology applications.

Globus: Smart instruments application

The Synthetic Forces Express (SF Express) project is investigating the use of high performance computing systems to support very large-scale distributed interactive simulations. SF Express conducted a record-breaking 100,298-vehicle simulation—the largest distributed, interactive battlefield simulation to date—executing it on 1,386 processors distributed over 13 computers at nine sites spanning seven time zones. Issues addressed to make this simulation possible included scalable communications, scenario distribution, resource configuration, resource management, information logging, monitoring, and fault tolerance. Global and local services were decoupled, allowing the application to run in a flexible, resource-aware environment. SF Express is incorporating emerging computational grid tools and techniques into the distributed interactive simulation environment to bring pervasive and dependable access to high end computation.

Legion: A worldwide virtual computer

Legion—an object-based metasystem software project at the University of Virginia funded by the Department of Defense/Naval Oceanographic Office (DoD/NAVO), NSF, and DOE—will build a system of millions of hosts and trillions of objects tied together with high-speed links. The interface will allow users working at home to access data and physical resources—such as cameras, digital libraries, linear accelerators, physical simulations, and video streams—as if these resources resided on their own disk drives. Groups of users will construct

shared virtual workspaces to exchange information and collaborate on research. Legion supports this with transparent scheduling, data management, fault tolerance, site autonomy, and security options, achieving high performance by allowing resource selection among available machines and parallel use of many resources. Legion can be used for parallel processing in a variety of applications and can execute a single application across geographically separate hosts or support meta-applications. Legion supports popular parallel libraries such as MPI and PVM; parallel languages such as MPL; wrapped parallel components so that legacy and new software can be aggregated into Legion objects; and exporting the runtime library interface to library, toolkit, and compiler writers. Legion is an open system whose runtime library is available and for which third-party development is encouraged.

Current and planned experimental applications of Legion include Chemistry at Harvard Molecular Mechanics (CHARMM) for molecular dynamics and mechanics such as protein-folding problems; direct simulation-Monte Carlo (DSMC), used at the University of Virginia to study vapor deposition onto surfaces; and coupled applications such as a federated climate model that ties together many lower-level models.

Java numerics

The rapid, widespread adoption of the Java language and environment for network-based computing has created a demand for reliable and reusable numerical software components to support scientific applications now under development. NIST is working with the Java Grande Forum (JGF) to assess the use of Java in high performance computing. The JGF is developing proposals for changing the language and its environment to overcome identified deficiencies and coordinate the development of standard applications program interfaces (APIs) for numerical computing. Several modifications to Java floating point semantics requested by the JGF Numerics Working Group were made this year, leading to performance improvements on common microprocessors, and progress is being tracked with SciMark, a new Web-based benchmark for numerical computations. A growing collection of proposed APIs for numerical computing in Java is available at: *http://math.nist.gov/javanumerics/*.

Research in advanced hardware components: Going where ordinary electronics cannot

Hardware components, such as microscopic transistors and wires, are the base of today's computing systems. They make up the processors, memories, switches, and linkages from which systems are built. The recent astonishing cost-performance progress in commodity computing systems is based primarily on progress at the hardware component level. At the high end, architectural advances are critical, but real breakthroughs are needed at the component technology level. The HTMT project (page 23) is investigating the use of RSFQ superconducting devices and holographic memories. The following projects illustrate how Federally funded basic research can go beyond current semiconductor-based approaches by using photonics, nuclear magnetic resonance, quantum mechanics, superconducting electronics, and biomolecules.

DARPA's very large-scale integration (VLSI) photonics program

It is not unusual today for commodity processor chips to perform simple arithmetic operations at speeds ranging from 300 to 600 MHz. These large computational bandwidths, however, are not yet matched by commensurate communication bandwidths between the chips. In fact, the wire interconnections between chips typically operate at only a fraction of the internal computational

bandwidth. Optics has the potential to solve the bandwidth interconnect problem between chips, between multichip modules, and at the backplane of boards.

Next generation advanced information processing systems, such as those expected to be used in realtime synthetic aperture radar (SAR) imaging, automatic target recognition, and intensive medical image processing, will require large aggregate computational and communication bandwidths. Since most of these systems are assembled from multiprocessor, memory, and special-purpose digital signal processing chips, the ideal communication bandwidths between them should be on the order of a terabit per second (Tbps). But Tbps bandwidths cannot, at this time, be achieved using conventional wire interconnections. The primary goals of DARPA's VLSI photonics program are to develop and demonstrate the basic technologies that will bring the benefits of optics to chip-scale interconnections.

Quantum computing

Theoretical computer science in the 1980s and 1990s has conjectured that a "quantum computer" can solve certain problems asymptotically more rapidly than conventional computers. Such a quantum computer can be thought of as an array of complex two-state systems—for example, the atomic nuclei in molecules with a spin quantum number of 1/2 as is found in hydrogen protons, each of which can store one quantum bit, or qubit, of information. NSA is conducting research into quantum computing using quanta of light as computing elements to demonstrate 1-qubit operations, perform experiments to achieve 2-qubit operations using the optical-lattice method of trapping atoms, and simulate the dynamics of a set of qubits in finer detail than previously achieved. Other NSA-supported projects include experimental research on quantum dots and Josephson junctions as possible qubits and an investigation of individual nuclear spins implanted in a silicon crystal as qubits.

In FY 2000, a consortium of university and Government laboratories is working on a scalable silicon-based nuclear spin quantum computer concept. Other research is under way on the characterization of spin coherence and Rabi oscillations in quantum nanostructures and on the measurement of decoherence times in superconducting qubits.

Quantum phase data storage and retrieval

NSF-funded researchers at the University of Michigan are exploring the possibility that atomic quantum phases can be used as a database. Data are assigned to a quantum state in a cesium atom, and a laser burst stores data in the assigned state by inverting the quantum phase. A second laser burst locates the stored data by amplifying the inverted state while suppressing the other states. Although speculative and long-term, this project illustrates the potential application of experimental physics to information technology.

Quantum information and computation

The goal of this DARPA-funded project at the California Institute of Technology (CalTech), the Massachusetts Institute of Technology (MIT), and USC is to build and operate devices that can store and process information in the quantum states of matter. The objectives are to:

- Develop and explore new ideas for implementing quantum gates in the laboratory
- Broaden the range of problems known to be efficiently solvable using quantum computing

- Design efficient networks of quantum gates to solve computationally difficult problems
- Improve the reliability of a quantum computer that operates under realistic conditions
- Develop a sequential simulator to validate and optimize models and circuits for quantum computers
- Fabricate devices and conduct experimental studies of their performance
- Develop quantum networks for distributed quantum computation and communication
- Explore the applicability of well-tested many-body methods to quantum computation

Ensemble quantum computer (EQC) nuclear magnetic resonance (NMR)

NMR spectroscopy is enabling the ensemble quantum computer, a new computational model that can, in principle, trade exponential growth in computation time for exponential increases in the system size needed for quantum computing, with the price being ensemble averages of microscopic observations. This DARPA-funded project focuses on several questions:

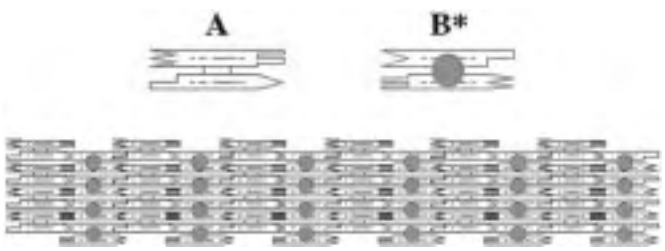
- How can the algorithms proposed for quantum computers be implemented efficiently using NMR?
- How can the errors due to imperfect instrumentation and decoherence be prevented, detected, and/or corrected?
- What computational models other than EQC can NMR implement?
- What macroscopic spin orders can be used to practically and efficiently encode information?
- What couplings and feedback among these orders can be most conveniently implemented?

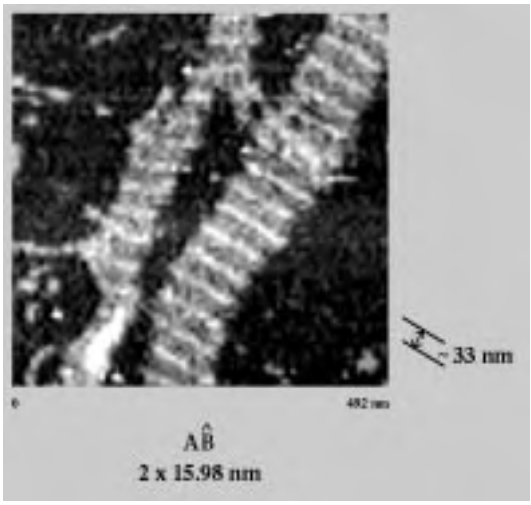
DNA data storage

One gram of DNA contains 10^{21} DNA bases, which is equal to 10^{8} terabytes of information storage. A DARPA- and NSF-funded project at a consortium of nine organizations led by Duke University is leveraging recombinant DNA techniques—appropriately modified to ensure error resiliency—to solve NP search problems. (NP problems grow exponentially harder as the number of items searched increases and becomes intractable using normal computing.) The key tasks are experimental demonstrations, nanoconstruction of new 3-D structures, applications, and mathematical models and software tools for simulation. The basic approach to NP search problems is to perform recombinant DNA operations to construct a vast number of possible solutions to the input problem and then perform further recombinant DNA operations to determine which of these correctly solve the problem.

Advanced microscopy tools for integrated circuit development

In order to develop and analyze the performance of the next and future generations of integrated circuits, NSA's microscopy program will develop knowledge, techniques, and instruments to fabricate, analyze, and manipulate the morphology of semiconductor surfaces and structures. The newly developed capabilities of in-situ direct imaging of surfaces (low energy electron microscopy [LEEM] and photo emission electron microscopy [PEEM]) with both spatial





A key goal of DNA nanotechnology is construction of periodic arrays in 2 and 3 dimensions. The project has produced 2-D arrays from antiparallel double crossover molecules. The first arrays were produced by using two different double crossover molecules (top). At the top of this drawing are two double crossover molecules, A and B, which are shown schematically. The complementarity between their sticky ends is represented as geometric complementarity. The * indicates that the B molecules contain DNA hairpins that project out of the plane of the helices; these hairpins act as topographic markers in atomic force microscopy (AFM) (left). The two molecules are approximately 4 nanometers wide, 16 nanometers long, and 2 nanometers thick. When these two tiles are mixed in solution, they form hydrogen-bonded 2-D arrays that are several microns long and hundreds of nanometers wide. The rows of projecting hairpins appear as stripes when visualized by AFM. The stripes are separated by about 32 nanometers, as expected. Thus, scientists can create specific structural features on the nanometer scale by self-assembly techniques.*

and temporal resolution will allow the intrinsic properties of surfaces to be used to design and characterize atomic-scale structures. Atomic force microscopy and nearfield scanning probe techniques will provide capabilities for manipulating, measuring, and confirming desired surface structures. Such capabilities will provide tools for creating solid state quantum computer test structures.

3-D diamond multichip module (MCM) cube computer

A 3-D computer architecture with a nanosecond system clock was completed in FY 2000 using test equipment developed by NSA and DARPA. After demonstrating a three-dimensionally interconnected stack of diamond aerosol spray-cooled multichip modules, the subnanosecond clock performance of five-stack 3-D interconnected diamond substrates verified the modeled performance for delay times through a "worst case nodal path." The emulated performance of a 40-layer stack was achieved through the use of "turn-around" boards at the top and bottom of the five-layer stack. The five-layer stack dissipated 2.5 kilowatts of heat while the maximum measured temperature on the 16 sensors per layer was 80 degrees centigrade in the center of the third MCM layer. The 500 watts per MCM were extracted by aerosol spray cooling of the exposed edges on two sides of the four-sided diamond substrate MCMs. Although this design used an older "fuzz button" interconnect technology for connecting the stacked thin-film-layered diamond substrates together (6,200 connections for each side of the MCM), no observed failures were recorded during the six-week qualifying tests. These results verified the 3-D concept for a diamond-substrate-based, aerosol spray-cooled supercomputer design.

Optical tape

Existing Government mass data storage capabilities are inadequate to meet near-term needs, and data rates fall short of what is needed to support advanced

computing. NSA is funding research to develop a higher-capacity medium compatible with existing systems to save millions of dollars that would otherwise be spent to expand floor space to house additional mass storage. Optical tape promises to combine high capacity with high data transfer rate in removable media. A 25 MBps prototype optical tape drive has been designed, fabricated, and tested, demonstrating that a user data capacity of one terabyte could be stored on a 3480-style cartridge. Funding partners for this project include NSA, NASA, and DOE. A planned follow-on effort is to modify the LOTS commercial tape drive to achieve data transfer rates of between 100 and 160 MBps in a single drive and to multiplex to provide a 320 MBps rate.

Superconducting electronics

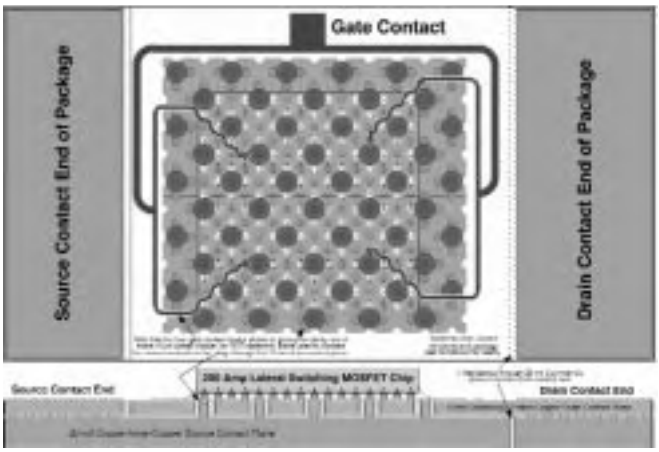
The NSA program in superconducting electronics focuses on high performance computing alternatives to current silicon and gallium arsenide technologies, which have speed and power limitations. Prior research suggests that superconducting supercomputers can deliver very high performance with very low power requirements. For example, the HTMT project reported above incorporates the use of superconducting electronics.

NSA is funding research to develop a superconductive crossbar switch, a 128 x 128 crossbar switch operating at 2.5 Gbps per port for use in supercomputing and networking applications. While the crossbar electronics operate at a temperature of 4° Kelvin and the cryogenic elements are cooled by a refrigerator, the input and output ports of the switch are at room temperature, providing the user with normal room-temperature support. Extended to higher speed and size, this switch is a candidate for use in HTMT. Layout of a 24-chip multichip module is 95 percent complete. Construction and assembly of the 128 x 128 crossbar will begin in FY 2001. The chips will be built and the electronics and housing of the MCM and cables will be completed.

NSA is pursuing sub-nanosecond memory. In order to take advantage of multigigahertz clock rate processors, multigigahertz memories must be locally available to the processors. This R&D will demonstrate sub-nanosecond access time superconductor and semiconductor memory technology and characterize its performance. The technology is applicable to both HTMT and to the superconductive crossbar switch when used as a room temperature switching system. So far, circuits have been designed and verified for the silicon-on-insulator (SOI) implementation, and calculations and simulations show its feasibility. Chips using sub-micron feature size are being built for testing. In FY 2001, the SOI version will be tested and characterized for speed and power, and the superconductive memory will be built and tested.

Smart memories

In FY 2000, NSA-supported smart memory R&D focused on the basic architecture of a computation tile—a building block that must efficiently execute applications exhibiting different types of parallelism. To guide the development, researchers examined how well the machine could emulate existing machine architectures developed for different application spaces. The machines chosen for emulation are the Stanford FLASH, Imagine, Hydra, and M-machine. The researchers also investigated architectures for on-chip interconnection networks, envisioning that future VLSI chips will be constructed with a common interconnection network independent of the function of the chip. The research aims at discovering new interconnection network architectures that best exploit the properties of these emerging on-chip networks.



Top and side views of a compact planar surface mount or clampdown package for a 200-ampere NMOS ultra-low Ron switching FET for low-voltage power converters. Using aerosol spray cooling to thermally manage the operation of a low-voltage, high-current power converter enables a radically new design for a silicon NMOS switching chip capable of operating at relatively high switching rates (>2MHz) while switching up to 200 amperes of current. The ultra-low resistance of the device allows power conversion efficiencies to reach as high as 90 percent—a remarkable value for such low-voltage converters. This inexpensive chip will play a critical role in high performance computation as bias voltages are reduced to meet the shrinking photolithographic demands of future machines that operate with concomitantly higher operating amperage.

FY 2001 plans include the design and fabrication of a reconfigurable wiring test chip, the development of programming models for smart memories, and completion of a proposed smart memory architecture.

Vendor partnerships

Vendor partnerships, which support the continuing development of supercomputing systems in the U.S., will benefit NSA and are also anticipated to be commercially successful. In October 1999, SGI/Cray demonstrated a boot of the operating system on the simulator of the SV2, a scalable hybrid scalar/vector system. The first node of the SV2 will be fabricated in early FY 2001. The first 64-processor system is expected in late FY 2001 or early FY 2002. NSA also has partnership agreements with Compaq/Digital Equipment Corporation (DEC) and Sun Microsystems, Inc., to improve scalar/integer calculations, reduce memory latency, and increase ease of use. Both vendors are incorporating NSA's UPC compiler in their commercial systems. DOE and the National Reconnaissance Office (NRO) participate in requirements specifications and reviews.

Molecular electronics

The goal of the DARPA program in "moletronics" is to demonstrate the integration of multiple molecules and/or nanoparticles into scalable, functional electronic devices that are interconnected as well as connected to the outside world in a realistic and practical manner. The aim is to provide moderate computational power and high-density memory in an extremely small, low-power format, which will not require multibillion-dollar fabrication facilities. Molecular electronics will use molecules and/or nanoparticles to achieve further miniaturization, greater functionality, and faster clock rates for advanced electronic systems that operate under a wide range of temperatures and preferably take advantage of 3-D architectures.

Nanocrystal devices

The aim of this NSF project is to develop and demonstrate nanocrystal (NC) and nanocrystal array (NCA)-based devices by first developing techniques for synthesizing NC-based thin films for device applications and understanding and optimizing their interfaces to other materials. Potential applications range from phosphors to photovoltaics to electronic switching and storage devices. Chemical techniques for fabricating NC-integrated systems and examining their structures will be developed. Their optical properties will be studied. Temperature-dependent electron transport, photovoltaics, and capacitance through NC and

NCA systems will be measured using electron-beam lithography. Interfaces will be characterized and controlled. This multidisciplinary scientific and engineering research will contribute basic understanding of key aspects of potentially highly useful materials and devices and will integrate research and education through student training.

NEW INITIATIVE

Measurement and calibration for the virtual sciences

The ongoing revolution in computing, communications, and information technologies can transform the foundation for industrial development from physical testing, prototypes, and pilot plants to a new paradigm based on modeling and simulation, where every material or process is conceived, designed, characterized, and optimized using advanced computation and information technology. Yet the methods and tools that allow users to validate, test, and calibrate models of designed materials, complex physical processes, and product performance are missing. Without them, industrial use of these advanced technologies will lag behind technology development.

Starting in FY 2001, NIST will begin a program that will include:

- Development and optimization of broadly applicable methods, algorithms, and associated data to accelerate industrial use of 21st century computing and information technologies for modeling and simulating complex materials and processes
- Development of methods and tools to support the validation, benchmarking, and comparison of models, algorithms, and software used for modeling and simulating materials and processes, including carefully constructed and analyzed open reference implementations, well-characterized test calculations and data, metrics and related tools for evaluating computed results, and testbeds for particular application domains
- Development of consensus standards, definitions, and tools for modeling and simulating information management, software integration, and software interoperability to enhance modeling and simulation in an application environment
- Development of user interfaces for advanced modeling and simulation software and integration with associated data and information resources from diverse sources including Web-based databases and archives

new starts

IT R&D Facilities

Overview

To provide an infrastructure to support science and technology research, NSF, DOE, NASA, NIH, NOAA, and EPA fund high end computing and communications facilities enabling scientists across the country to run large-scale applications at these sites via remote connections. This permits scientists to:

- Evaluate early prototype systems and provide feedback to developers
- Integrate visualization and virtual reality into high performance systems
- Run full-scale applications, including Grand Challenge and other breakthrough applications, on systems not otherwise available
- Develop parallel software using scaled down systems

Researchers at these facilities rely on enabling technologies, including high-speed networks, supercomputing systems—often with parallel processor architectures, massive data storage, and virtual reality display devices. Multidisciplinary research teams, including university faculty, facility managers and staff, hardware and software vendors, and industrial affiliates, contribute to the facilities' overall success. IT R&D funding is leveraged through equipment and personnel from vendors, discipline-specific agency funds, state and local funds, and industrial affiliate contributions. Industrial affiliation and outreach activities offer a low-risk environment for exploring and exploiting IT R&D technologies.

Applications software developers access these facilities over the Internet and experimental networks such as the NGI. All facilities provide extensive undergraduate educational opportunities and training for researchers, graduate students, and faculty members, and publish professional journal articles, annual reports, and newsletters. Most also offer K-12 educational programs.

The following systems—excluding workstation clusters—are funded by the named agencies and may receive additional funding from other IT R&D agencies. For example, funding for systems at NSF centers also comes from DARPA, NASA, and NIH.

NSF advanced computational partnerships and centers

The Partnerships for Advanced Computational Infrastructure (PACI) program supports the National Computational Science Alliance (Alliance), headquartered at the University of Illinois at Urbana-Champaign (UIUC), and the National Partnership for Advanced Computational Infrastructure (NPACI) at the San Diego Supercomputer Center (SDSC). Each of these leading-edge sites maintains a variety of high end computing systems and supports,

individually or in tandem, more than 60 geographically distributed partner institutions from 29 states that maintain smaller systems.

Taken as a whole, the PACI partnerships constitute a large, distributed computing environment connected via high-speed networks, over which the partners contribute to the infrastructure by developing, applying, and testing the necessary software and tools to drive further growth of this "national grid" of interconnected high performance computing systems. PACI provides the foundation for meeting the expanding need for high end computation and information technologies required by the U.S. academic community, supporting and developing information-intensive uses of the computing systems available on the grid and the technologies needed to support their use.

The partnerships provide:

- Access to a diverse set of advanced and mid-range computing engines, data storage systems, and experimental architectures
- Enabling technologies, by developing both software tools for parallel computation and software to access the partnership's widely distributed, architecturally diverse machines and data sources and effectively exploiting the partnership's very large distributed systems
- Application technologies, by allowing researchers in high end applications to develop and optimize discipline-specific codes and software infrastructures to make these available to the program as a whole and to researchers in other areas
- Education outreach and training, building an understanding of how to use high performance computing and communications resources and broadening the base of participation to ensure the Nation's continued world leadership in computational science and engineering

National Computational Science Alliance (Alliance)

The Alliance focuses on the emerging distributed shared memory (DSM) architecture and PC clusters running Windows NT and Linux. DSM combines the easy programmability of shared memory symmetric multiprocessors (SMPs) with the scalability of distributed memory MPPs. PC clusters provide unsurpassed price-performance for some applications. The Alliance is also engaged in grid-building activities (page 42) and the development of science portals.

Alliance resources

National Center for Supercomputing Applications (NCSA) at UIUC

- Twelve SGI Origin 2000s with an aggregate of 1,528 processors (768 at 500 Mflops peak each and 760 at 390 Mflops peak each), 680 Gflops peak, 618 GB memory, 4.3 TB disk storage
- NT Supercluster with 128 550 MHz Pentium III processors (550 Mflops peak each) and 32 330 MHz Pentium II processors (330 Mflops peak each), 81 Gflops peak, 144 GB memory
- HP/Convex Exemplar SPP-2000 with 64 180 MHz PA 8000 processors, 46 Gflops peak, 16 GB memory

Boston University

- Four SGI Origin 2000s with an aggregate of 192 195 MHz R10000 processors, 75 Gflops peak, 24 GB memory
- Four SGI Power Challenges with an aggregate of 42 processors and 6.2 GB memory

University of Kentucky

- HP/Convex Exemplar SPP-2200 with 64 200 MHz PA 8200 processors, 51 Gflops peak, 16 GB memory

University of New Mexico-Albuquerque HPCC

- Los Lobos Linux Supercluster with 512 733 MHz Pentium II processors, 375 Gflops peak, 256 GB memory (new in 2000)
- Roadrunner Linux Supercluster with 128 450 MHz Pentium II processors, 57.6 Gflops peak, 32 GB memory
- IBM SP2 with 96 66 MHz Power2 processors, 25 Gflops peak, 6 GB memory

University of New Mexico-Maui HPCC

- IBM SP with 200 222 MHz Power3 processors, 178 Gflops peak, 100 GB memory (new in 2000)
- IBM SP with 192 160 MHz Power2 processors, 123 Gflops peak, 100 GB memory

University of Wisconsin-Madison

- Condor flock—pool of approximately 400 machines of various types

National Partnership for Advanced Computational Infrastructure (NPACI)

NPACI is creating a continuous, ubiquitous, and pervasive national computational infrastructure. NPACI's focus includes providing computation and information resources to enable discovery at scales not previously achievable; developing and deploying integrated, easy-to-use computational environments to foster discovery in traditional and emerging disciplines; and promoting computational literacy to extend the excitement, benefits, and opportunities of computational science to all U.S. citizens.

NPACI resources

San Diego Supercomputer Center

- IBM RS/6000 SP with 1,152 222 MHz Power3 processors, 1.02 Tflops peak, 576 GB memory—the most powerful computing platform available to the U.S. academic community (new in 2000)
- Cray T3E with 272 300 MHz Alpha 21164 processors, 154 Gflops peak, 34 GB memory
- IBM SP with 128 160 MHz Power2 processors, 82 Gflops peak, 32 GB memory
- Sun HPC10000 with 64 400 MHz UltraSparc II processors, 51 Gflops peak, 64 GB memory

- Cray T90 with 14 processors, 24 Gflops peak, 4 GB memory
- Tera MTA with 16 processors, 8 Gflops peak, 16 GB memory (new in 2000)

University of Texas

- Cray T3E with 88 300 MHz Alpha 21164 processors, 34 Gflops peak, 11 GB memory
- Cray SV1 with 16 processors, 19.2 Gflops peak, 16 GB memory

University of Michigan

- IBM SP with 64 160 MHz Power2 processors, 31 Gflops peak, 64 GB memory

California Institute of Technology (operated jointly with NASA Jet Propulsion Laboratory)

- HP Exemplar X-Class with 256 processors, 184 Gflops peak, 64 GB memory
- HP V2500 with 128 processors, 128 GB memory (new in 2000)

National Center for Atmospheric Research (NCAR)

NCAR, located in Boulder, Colorado, provides computing facilities for the community of university and NCAR researchers in atmospheric, oceanic, and related sciences. The Climate Simulation Laboratory (CSL), a dedicated climate model computing facility, has also been established at NCAR to support the multiagency U.S. Global Change Research Program (USGCRP). NCAR is operated by the University Corporation for Atmospheric Research (UCAR) with NSF as primary sponsor.

NCAR resources

- IBM SP with 144 dual-processor (200 MHz Power3) nodes (128 compute nodes), 204 Gflops peak, 128 GB memory, 2.5 TB disk
- Compaq ES40 with 8 four-processor nodes, 32 Gflops peak, 32 GB memory
- Two Cray J924se/1024s with 24 processors, 8 GB memory, 1.5 Gflops sustained total
- Cray J920/512 with 20 processors, 4 GB memory, 60 Mflops/processor sustained
- Cray J916/256 with 16 processors, 2 GB memory, 60 Mflops/processor sustained
- SGI Origin 2000/128 with 128 250 MHz R10000 processors, 16 GB memory
- SGI Origin 2000/16 with 16 250 MHz R10000 processors, 16 GB memory

NASA testbeds

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

Numerical Aerospace Simulation (NAS) Facility, NASA Ames Research Center, Moffett Field, California

- SGI Origin 2000 with 512 CPUs, 192 GB memory
- SGI Origin 2000 with 256 CPUs, 64 GB memory
- SGI Origin 2000 with 64 CPUs, 16 GB memory
- SGI Origin 2000 with 24 CPUs, 7 GB memory
- Cray C90 with 16 CPUs, 8 GB memory
- SGI Origin 2000 with 12 CPUs, 3 GB memory (for long-term storage)
- Cray C90 with 7 CPUs, 2 GB memory

NASA Glenn Research Center, Cleveland, Ohio

- SGI Origin 2000 with 24 CPUs, 6 GB memory

NASA Goddard Space Flight Center, Greenbelt, Maryland

- Cray T3E-600, 1,296 processors, 162 GB memory
- Cray J932se, 32 processors, 8 GB memory
- Cray SV1, 24 processors, 8 GB memory

Jet Propulsion Laboratory, Pasadena, California

- SGI Origin 2000, 128 processors (R12000, 300 MHz), 32 GB memory, 6 TB disk
- Cray SV1-1A, 16 processors, 8 GB memory, 480 GB disk

NASA Langley Research Center, Langley, Virginia

- SGI Origin 2000 with 16 CPUs, 4 GB memory

DOE laboratories

DOE's Office of Science maintains a variety of high end supercomputing facilities throughout the Nation, including:

National Energy Research Scientific Computing Center (NERSC) at Lawrence Berkeley National Laboratory (LBNL)

The NERSC facility at LBNL in Berkeley, California, provides production computing for investigators supported by the Office of Science as well as researchers at universities and Federal laboratories. Some of these resources are available through a peer review allocation process. Research areas include high-energy and nuclear physics, fusion energy, materials sciences, chemistry, life sciences, environmental sciences, Earth and engineering sciences, and applied mathematics and computational science. NERSC's resources, which will rise from a total theoretical speed of 1.2 Tflops to 3.8 Tflops by December 2000, include:

- Cray T3E-900 with 692 processors, 623 Gflops peak, 177 GB memory, and 2.8 TB disk storage
- IBM RS/6000 SP with 608 Power3 200 MHz processors, 486 Gflops peak, 256 GB memory, and 10 TB disk storage

- Cray PVP cluster consisting of three Cray SV1s and a Cray J90se, with a total of 96 vector processors, 4 gigawords (GW) memory, and a peak performance of 83 Gflops
- Two HPSS tertiary storage systems with 1,000 TB total capacity
- SGI Onyx 2 server for scientific visualization from remote locations
- Parallel Distributed Systems Facility (PDSF), a networked, distributed system of 53 workstations and eight disk vaults with file servers, supporting large-scale computation for high-energy and nuclear physics investigations

Argonne National Laboratory (ANL)

ANL's facilities support computer science and computational science, providing testbeds for advanced visualization, cluster management, distributed computing, middleware, portable numerical libraries, and system software. Computational science applications include biology, chemistry, climate, combustion, and materials science. ANL's high end testbeds have a total theoretical speed of 0.5 Tflops and 100 TB storage. The testbeds include an SGI Origin 2000, an IBM SP-2, and the newest system, "Chiba City," which is a Linux SMP cluster consisting of 256 nodes, each with two Pentium III 500 MHz processors, 128 GB memory, 2.3 TB local storage, and a total theoretical speed of 256 Gflops.

Los Alamos National Laboratory (LANL)

Two kinds of high performance systems in the Advanced Computing Laboratory (ACL) at LANL support applications and power the ACL's software infrastructure:

- "Nirvana," an SGI/Cray system consisting of 16 shared memory multiprocessors, each with 128 processors and 32 GB of memory. The system interconnect is HiPPI 800 technology. Nirvana has 6.91 TB of disk storage with aggregate peak bandwidth of 4.8 GBps. This 2,048-processor ensemble machine provides 1 Tflops (theoretical), making it one of the highest-capability unclassified computer systems in the world. With 10 SGI Infinite Reality graphics engines that enable interactive visualization and analysis previously unavailable, Nirvana is the largest graphics supercomputer in the world. Nirvana supports applications for the Office of Science, universities in the DOE Accelerated Strategic Computing Initiative's (ASCI's) Academic Strategic Alliance Program (ASAP), and LANL institutional programs.
- A system of experimental, low-cost clusters of machines running Linux is supported for applications and research into making such clusters more effective. Currently, the largest of these is "Rockhopper," with 128 dual-CPU nodes (500 MHz Pentium IIIs), each with 1 GB of memory and 9 GB of disk storage. The system interconnect is Myrinet, with 14 usec latency and 1.2 GBps bandwidth. The system provides 128 Gflops peak performance and has shown per-processor performance parity with Nirvana on some applications.

Oak Ridge National Laboratory (ORNL)

The high end systems at ORNL include an IBM SP-3 and a Compaq Alphaserver with a total theoretical speed of 1.5 Tflops and 360 TB HPSS storage, dedicated to unclassified scientific computing:

- "Eagle" is an SMP cluster with 184 nodes, each with four IBM Power3 II (1.5 Gflops) processors, 1.08 Tflops/s theoretical speed, 372 GB memory, and 9.2 TB local storage
- "Falcon" is an SMP cluster with 80 nodes, consisting of four Compaq EV67 (1.3 Gflops) processors, 427 Gflop/s theoretical speed, 160 GB memory, and 5.5 TB local storage

Together, they form a high end production system and applications testbed for atmospheric radiation, biology, chemistry, climate, combustion, materials, nanotechnology, many-body physics, and spallation neutron source design.

Pacific Northwest National Laboratory (PNNL)

PNNL's high end systems have a total theoretical speed of 414 Gflops with 48 TB of EMASS storage. Uses include applied mathematics, atmospheric sciences, biology, chemistry, climate modeling, engineering, environmental molecular sciences, natural resources management, and subsurface reactive transport.

- "NWMPP1" is an MPP IBM SP with 512 0.5 Gflops processors, 247 Gflops theoretical speed, 262 GB memory, and 5.0 TB local storage

NIH computing systems

The Center for Information technology (CIT) has a 46-processor SGI Origin 2000 parallel computer and a 224-processor Beowulf cluster. Both systems and other high-performance computing resources are used by the NIH scientific staff in biomedical applications. NCI's Frederick Biomedical Supercomputing Center has a 96-processor Cray SV1 supercomputer, a 64-processor SGI Origin 2000 parallel computer, and a collection of biomedical software that are available to scientists who use the facility. The National Center for Research Resources (NCRR) supports various systems for biomedical research applications at its six High Performance Computing Resources Centers, which include:

NCRR Computing Resources Centers

- Resource for Concurrent Biological Computing, Beckman Institute, University of Illinois
- Supercomputing for Biomedical Research, Pittsburgh Supercomputing Center
- Theoretical Simulation of Biological Systems, Columbia University
- Parallel Computing Resource for Structural Biology, University of North Carolina, Chapel Hill
- Biomedical Computation Resource, University of California, San Diego
- Parallel Processing Resource for Biomedical Scientists, Cornell Theory Center, Cornell University

NCRR Scientific Visualization Resource Centers

NCRR also supports two Scientific Visualization Resource Centers:

- Interactive Graphics for Molecular Studies, University of North Carolina, Chapel Hill
- Special Research Resource for Biomolecular Graphics, University of California, San Francisco

NOAA laboratories

NOAA operates two high end computing centers that work closely with the computer science community to advance the art of programming highly parallel scalable systems for geophysical fluid dynamics problems:

- The Forecast Systems Laboratory in Boulder, Colorado, has a system from HPTi with 256 Compaq Alpha processors and 128 GB memory. This system is used to explore highly parallel regional and mesoscale forecast models under severe wall-clock constraints.
- The Geophysical Fluid Dynamics Laboratory (GFDL) in Princeton, New Jersey, has a Cray T-90 PVP and a 128-processor Cray T3E 900 MPP. These systems are used for the global climate modeling and weather forecasting Grand Challenges.

EPA systems

EPA's National Environmental Supercomputing Center (NESC) in Research Triangle Park, North Carolina, is dedicated to environmental research and problem solving for characterizing and quantifying risks to human health and to the integrity, resilience, and sustainability of ecosystems now and in the future. The NESC high performance computing resources include:

- Cray T3E-1200 with 72 processors and 18 GB memory, 76 Gflops peak
- Cray T3D with 128 processors and 8 GB memory, 19.2 Gflops peak
- Cray C94/3128 with 4 processors and 128 MW memory, 3 Gflops peak

ACCESS to the Access Grid

Overview

In April 1999, the NSF-funded National Computational Science Alliance (Alliance) opened a state-of-the-art communication facility in the Washington, D.C., metropolitan area to house the Alliance Center for Collaboration, Education, Science and Software (ACCESS DC). The first in a series of such sites being developed by the Alliance around the country, the 7,000-square-foot ACCESS DC complex—located adjacent to NSF headquarters in Arlington, Virginia—offers the business, education, and government communities hands-on introductory experience with the high end technologies and applications generated by Alliance IT R&D. ACCESS DC is designed to serve as a comprehensive information technology transfer point, where representatives of these major sectors, guided by skilled researchers, can examine, discuss, and experiment with the Nation's most advanced networking capabilities. The goal is to expand public understanding of how technology can strengthen business competitiveness, education, and government services, and improve the quality of life in the new century.

The Alliance, one of the two grant winners in NSF's PACI program, is a collaboration of more than 50 research institutions, industry, and Government formed in 1997 to prototype a next-generation computational and information infrastructure linking supercomputers, advanced visualization environments, remote scientific and medical instrumentation, and very large databases through high-speed networks such as vBNS and Abilene. The Alliance structure includes:

- Six application technologies teams—chemical engineering, cosmology, environmental hydrology, molecular biology, nanomaterials, and scientific instrumentation
- Three enabling technologies teams—parallel computing, distributed computing, and data and collaboration
- Four education, outreach, and training teams—enhancing education, universal access, government, and training
- Fifteen partners for advanced computational services—these are the research institutions that provide gateways to the National Alliance Grid, the Alliance's infrastructure for advanced computing
- Fifteen industrial partnerships, 10 strategic vendors, and Federal partnerships with DoD's High Performance Computing Modernization Program (HPCMP) and DOE's ASCI

The Computational Grid

The National Alliance Grid encompasses two components. The Computational Grid consists of a core group of high end parallel computing systems called supernodes, located at the Alliance headquarters at NCSA and at Boston University, the Ohio Supercomputing Center, the University of Kentucky, the University of New Mexico/Maui High Performance Computing Center, and the University of Wisconsin-Madison. The structure of the Computational Grid, which encompasses a variety of parallel processor architectures (massively parallel, parallel vector, distributed shared memory, shared memory symmetric multiprocessors, and clustered workstations), enables Alliance investigators to evaluate the relative performance of applications running on these architectures, and provides testbeds for researchers developing portable programming languages and environments.

The Access Grid

The Access Grid is the ensemble of resources, including hardware, software, and communications capabilities, enabling people across the Grid to collaborate in large-scale distributed meetings, collaborative teamwork sessions, seminars, lectures, tutorials, and training. Access Grid nodes are "designed spaces" that support the high end audio and visual technology—large-format media displays, presentation and interactive software environments, and interfaces to Grid middleware and remote visualization environments—needed to provide high-quality and productive user experiences. As the first fully equipped prototype location for such group-to-group, as opposed to desktop-to-desktop, interactivity, ACCESS DC is a primary host site for the Access Grid and demonstrates the Grid's collaboration and large-scale research capabilities.

Research activities

Access Grid research will explore:

- Scheduling mechanisms and models for group collaboration, such as tools to script and synchronize activities across the network
- Network audio, focusing on ways to improve the quality-of-service engineering needed for high-resolution audio streams to and from Grid nodes
- Scalable, high-resolution video techniques, including compression, panning and zooming capabilities, and video mosaics
- Designs for a network flow engine (NFE) and extensible routing technologies and tools to enable researchers to prioritize and route video and audio streams dynamically across the Grid
- Integration of remote visualization capabilities, including tools enabling interoperability and synchronization between differing visualization systems, and elaboration of the prototype Flatland 3-D collaborative visualization system
- Human factors and related workspace design and function issues, including physical and psychological responses to visualization environments and quantitative comparisons of total immersion and projection display technologies
- Remote 2-D and 3-D tools and applications

- Integration and coordination of the Access Grid, including discussions with industrial partners about commercial packaging of Grid node hardware to facilitate broader adoption of a standard hardware configuration

Access Grid nodes

Other Access Grid nodes are currently operational or under construction at the Air Force Research Laboratory, ANL, Boston University, the University of Illinois at Chicago's EVL, the University of Hawaii, the University of Kansas, the University of Kentucky, DOE laboratories (LBNL and LANL), the Maui High Performance Computing Center, NCAR, NCSA/UIUC, the University of New Mexico, North Dakota State, Ohio State, Princeton University, and the University of Utah. The Alliance will leverage the experience of the ACCESS DC site to guide expansion of the Access Grid utility to PACI partners and institutions in NSF's Experimental Program to Stimulate Competitive Research (EPSCoR), the Southeastern Universities Research Association (SURA), and IT R&D partnerships. The infrastructure at each site will vary with local priorities and programmatic needs, and some of the interactive sites will be independent of the Access Grid.

ACCESS DC

To bring visitors directly in contact with high end IT capabilities, such as visual supercomputing to create 3-D virtual environments, ACCESS DC is equipped with the latest immersive virtual reality, teleconferencing, and workstation technologies, all powered by cutting-edge software for interacting with distributed data and collaborating with colleagues in different locations. Supported by teams of experts, both on site and distributed across the Alliance, ACCESS DC visitors use the Grid and its applications to work in collaborative immersive realities, conduct multilocation video and multimedia conferences, and experiment with distributed distance learning capabilities.

Visitors and workshops

Since its inaugural open house, ACCESS DC has averaged nearly 200 visitors a month including representatives of K-12 and higher education, Federal, state, and local government agencies, and Alliance member institutions, and international visitors and dignitaries. Education and training activities have included:

- A workshop on parallel computing, hosted by NCSA
- A course and a workshop on Java for scientific programming with Tango, Syracuse University's software for Web-based distributed collaboration
- A distributed computing workshop hosted by NCSA and the National Laboratory for Applied Networking Research (NLANR)
- An ARL/NCSA/Maryland Virtual High School summer workshop on computational science for high school teachers and students, a week-long program providing hands-on experience with state-of-the-art computational tools for problem solving in science and engineering
- An SC99 supercomputing conference demonstration linking ACCESS DC with EVL and the NCSA exhibition booth over NSF's very high performance Backbone Network Services (vBNS) network

ACCESS DC, located near NSF in Arlington, Virginia, is equipped with the latest immersive virtual reality, teleconferencing, and workstation technologies, powered by cutting-edge software. At the upper left is ACCESS DC's ImmersaDesk. The ACCESS DC multiscreen interactive conference facility is pictured above. And ACCESS DC's state-of-the-art meeting room is pictured left.

Chautauquas 99

ACCESS DC participated in the Alliance's first major Access Grid event, called Chautauquas 99, which was designed to showcase for audiences of nonspecialists the Grid's varied uses in distance learning, distributed collaboration, and remote visualization. Three Access Grid nodes—Boston University, the University of Kentucky, and the University of New Mexico—each hosted a two-day Chautauqua in late summer 1999 open to interested members of the university and regional communities. About 500 people attended the three sessions, both on site and at multicast remote locations.

ACCESS DC hosted meetings of NSF's Digital Government Consortium, a nationwide network of researchers from academia, industry, and government working to improve access to government information and services through IT R&D, and of the Collaborative Community of Nations/Smithsonian National American Indian Museum. ACCESS DC is collaborating with DARPA and the ISI East network engineering firm to host the NGI Distinguished Lecture Series at the Northern Virginia facility.

Chautauquas 2000

In summer 2000, ACCESS DC hosted two Chautauquas 2000. The first, sponsored by the Ohio Supercomputing Center and the Committee on

Institutional Cooperation, a partnership of 12 midwestern research universities, was held June 13-15, 2000, with the supercomputing center serving as co-host site. The second was held August 1-3, 2000, under the sponsorship of the University of Kansas and NSF's EPSCoR K*Star research program. The university served as co-host site for this event.

Human Computer Interface and Information Management

Overview

HCI & IM R&D develops and enhances technologies to facilitate and improve communication between humans and computing devices. Federal investments in HCI & IM lead to increased computer accessibility and usability and to a greater understanding of computing systems, computer interfaces, and communication networks.

FY 2001 R&D areas include:

- Battlefield robotics
- Collaboratories for knowledge and data sharing, group decision-making, and operation of remote instruments
- Digital libraries
- Information agents for collecting and analyzing data
- Management, visualization, and exploitation of information and knowledge, including large knowledge repositories
- Multilingual document and speech-to-speech translation and understanding
- Multimodal interactions between humans and computer systems, including speech recognition tools, audio interfaces, and haptic devices
- Remote, autonomous agents
- Universal access
- Virtual reality environments

Scientists, engineers, physicians, educators, students, librarians, the workforce, the Federal government, and the public are all potential beneficiaries of HCI & IM technologies. This section describes the wide range of Federally sponsored IT R&D activities dedicated to expanding modes of human-computer interface and improving our ability to manage and make use of information resources.

Managing and "seeing" digital information

Digital Government program

NSF's Digital Government program encourages and supports research and interagency collaborative efforts to make Federal information and services more useful and widely available to the public without compromising the privacy and security of either citizens or the Government. Digital Government supports projects that bring computer and IT researchers together with Federal agencies

with significant information services missions. The program funds research; workshops and planning grants in areas such as statistical graphics; distributed geographic information system (GIS) image storage and retrieval in field data collection; regulatory compliance reporting; geospatial ontology and multimedia data mining; and coastal management and decision-making.

As part of an ongoing in-depth study of how IT R&D can more effectively support advances in Government uses of IT, Digital Government and NASA funded two workshops and reports in conjunction with the Computer Science and Telecommunications Board (CSTB) of the National Research Council. Findings of a workshop on IT research for crises management are available online at *http://www.cstb.org.* A second workshop, on February 9-10, 1999, provided an opportunity for IT researchers, IT research managers, and academic statisticians to discuss with Federal managers how they might more effectively collaborate to improve Federal statistics. The CSTB's formal report on these proceedings is also available online.

Other representative Digital Government activities include:

- An autumn 1999 meeting at the University of California-Santa Barbara on protocols for geographic naming in digital gazetteers
- A University of Maryland study, in partnership with the U.S. Bureau of the Census, the Federal Geographic Data Committee (FGDC), USGS, and Brazilian University, to develop new algorithms and systems for indexing, querying, and joining of GIS spatial data with outputs of tabular data (spreadsheets) rather than rendered maps
- A National Institute of Statistical Sciences effort to develop a query system for disclosure-limited statistical analysis of confidential data using metadata, risk reduction and visualization, and query history databases. Research partners include the Bureau of Labor Statistics (BLS), Carnegie Mellon University, the Census Bureau, Kansas State University, LANL, the MCNC computing center, the National Agricultural Statistics Service (NASS), the National Center for Health Statistics (NCHS), the Ohio State University, and the University of Maryland.
- A University of North Carolina-Chapel Hill study—in partnership with BLS, DOE's Energy Information Administration (EIA), NCHS, Syracuse University, Textwise, the University of California-Berkeley, and the University of Maryland—of ways to improve citizen access to Government statistical data with new techniques for metadata integration, tabular visualization and browsing, and natural language processing
- A Stanford University exploration of technologies focused on universal access to IT capabilities. With partners at the Census Bureau, the General Services Administration (GSA), Marconi, and Synapse, researchers work on technologies for the human/computer interface, including speech recognition, eye tracking, and haptic interfaces
- Development of new Digital Government testbeds and education programs such as:
    - A partnership between the Council on Excellence in Government and the Federal Web Consortium to develop a prototype Digital

Government resource center and a curriculum for a Digital Government Fellows program

- A USC/ISI partnership with the Census Bureau, BLS, and NCHS to create a survey authoring and administrative testbed for online information gathering and analysis

Some 20 Federal agencies are funded in the grant program, whose FY 2000 announcement drew more than 55 research proposals.

Systems Integration for Manufacturing Applications (SIMA)

NIST's Systems Integration for Manufacturing Applications (SIMA) program focuses on developing interfaces among advanced engineering and manufacturing software systems that enable seamless communication of information throughout extended manufacturing enterprises. Researchers are developing the technological infrastructure to allow engineers around the world to collaborate effectively, manufacturers to reduce time-to-market while producing and delivering products more efficiently, and customers to choose uniquely configured products from manufacturers without paying a premium. Recognizing that scientists and engineers developing new products and processes require access to authoritative information resources for fundamental technical data, SIMA R&D also focuses on applications for collating and delivering critically evaluated data in new and more intuitive ways. The NIST collaboratory activities described in the HCI & IM section are all part of SIMA.

Intelligent systems

Adaptive learning technology

DoD researchers are developing geographically distributed, cost-efficient, versatile, reusable, and adaptable systems to meet education and training goals for DoD's military and civilian workforces—technologies that can be deployed wherever and whenever the need arises. This multidisciplinary effort is aimed at going well beyond current state-of-the-art prototypes that can adapt to diverse learner communities.

The Office of the Secretary of Defense (OSD) University Research Initiative (URI)-supported adaptive learning technology program includes research on human factors leading to the design of effective, efficient, and user-friendly training environments; research on the cognitive and perceptual requirements for effective network-based learning, including the development of cognitive and psychomotor skills; tests of the effectiveness of new distributed training regimes, compared with traditional methods, especially in training in complex decision-making operations, command, control, and communications tasks, and other skills required of modern military personnel; research that develops a scientific basis for understanding how distributed training systems can best build competencies and skills that remain robust and effective under conditions of fatigue or stress; and research on how distributed training systems can assess a wide range of individual differences in cognitive, sensory, perceptual, or psychomotor performance.

Knowledge and cognitive systems (KCS)

NSF's knowledge and cognitive systems (KCS) program supports research fundamental to the development of intelligent machines—machines that can interact in an intelligent manner with humans (computer-aided machine intelligence) or on their own (autonomous intelligent agents). The program's three research components focus on artificial intelligence:

- Knowledge representation focuses on characterizing, specifying, storing, and transferring of all types of knowledge, including functional, categorical, structural, and relational knowledge, and forms of knowledge such as auditory, visual, tactile, and kinesthetic.
- Cognitive processing systems R&D investigates the way intelligent systems manage, transform, and use stored knowledge and incoming information to accomplish tasks.
- Machine learning and knowledge acquisition—the study of knowledge discovery and the way it is encoded—recognizes that machines, like humans, need to develop if they are to produce behavior that resembles cognitive functioning.

Some KCS research will explore knowledge representation in machines and studies of cognitive processes, which may be modeled on what is known of human or animal cognition in decision-making, linguistic cognition, machine learning, planning, reasoning, and sensory cognition. Application areas may include design and manufacturing, network management, medical diagnosis, data mining, and intelligent tutoring.

Smart spaces

The emergence of commodity high performance hardware and software is transforming the delivery, versatility, and cost of computing. This trend toward numerous, casually accessible computing devices connected by a ubiquitous network infrastructure is enabling more pervasive computing and fostering the development of new "smart spaces."

"Smart spaces"—worker-aware, perceptive, and connected computing environments—emerge when pervasive devices, networking technologies, and information retrieval capabilities are integrated into a functional work environment. A future interactive smart space might include voice interfaces, integrated multimedia, workload computers, wireless networking with personal communicators, language translators, large screen displays, and touch pads to support command, analysis, and design groups.

NIST is developing a smart space environment emphasizing sensor-based, perceptive interface experiments that can identify individual speakers and what they say. Researchers are addressing sensor fusion issues that must be resolved in order to combine visual and acoustic data streams—using signatures from facial images, vocal acoustics, and directional data from microphone arrays—to create a unique "person recognizer." A combined person recognizer, encompassing all individuals within the smart space environment, will provide higher accuracy, allow individual speakers to be associated with spoken utterances in a group, and allow continuous speech recognition to decode words. This technology can provide the basis for a new generation of perceptive interfaces that support collaborative work groups.

Speech technology

NIST has collaborated with both DARPA and NSA to develop and implement a series of benchmark tests to advance computer speech recognition technology, including near-real-time recognition of unconstrained vocabularies in speech typically encountered in radio and TV news broadcasts. For NSA, NIST has developed and implemented test protocols for transcribing conversational telephone-line speech, as well as tests of speaker recognition technologies. New information extraction measures and tests are being

implemented in FY 2000. In FY 2001, NIST plans to continue developing metrics and tests for information extraction from speech and for speech-based user interaction.

Multimodal capabilities

Communicator

Today, U.S. military operators are restricted in their access to information. Most often, military operators use voice over radio or formatted text messages to communicate during crises. To help the warfighter improve readiness and response, more flexible, timely, and dependable access to information is needed. In FY 2001, DARPA will continue its work on a new IT capability that allows people to literally converse with computers to create, access, and manage information and solve problems. The challenge of DARPA's Communicator program is to make relevant information accessible to warfighters in an organized, integrated manner that is readily usable by military personnel equipped with untethered information devices. By blending speech recognition technologies—the results of DARPA's past R&D investments—with natural language capabilities, Communicator will enable users to access and create information from networks without a laptop computer, batteries, keyboard, full-size display, and the time spent connecting to a browser. Key research challenges include dialogue management, context tracking, language generation, input language understanding, and hands-free and eyes-free interaction.

Human factors in aerospace systems

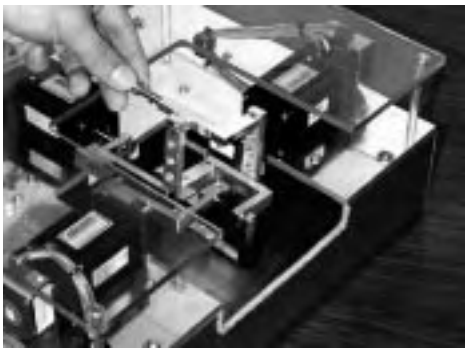
NASA's R&D in human factors will develop advanced human-centered information technology, models of human/system performance, principled methods for human factors design of human-centered computing systems, and human factors expertise to address aerospace challenges. In pursuit of this goal, scientists and engineers at NASA's Ames Research Center are conducting research in adapting the technologies of the future to humans in aerospace systems environments.

Virtual environments that offer only visual displays present interactive simulations that can be seen but not felt. The absence of palpable physics, such as the mechanical constraints imposed by real object surfaces or their weight as induced by altered gravity, compromises the accuracy and effectiveness of human manual interaction with these simulations. The mechanical feedback to the sense of touch necessary for realistic manipulation can be introduced into the simulation via haptic interfaces—force-reflecting hardware with appropriate software—to the virtual environment.

NASA and the University of California-Berkeley have built haptic interfaces for arm-scale and finger-scale manipulation. Both machines are based on a NASA-patented kinematic architecture that enables three degrees-of-freedom translations of a grasped endpoint to be coupled to the rotations of three motors fixed to a common base. Because it is composed solely of rigid links and ball bearing joints, this coupling mechanism permits a proportionally large workspace and is stiffer and more responsive than commercially available cable-driven systems. Researchers are implementing dynamic behaviors and textures in these haptic virtual environments for psychophysical studies.

Translingual Information Detection, Extraction, and Summarization (TIDES)

The burgeoning international information infrastructure offers unprecedented opportunities for rapid access to and fusion of information from sources around the world. However, the human language barrier remains a critical impediment



Miniaturized linkage using new kinematics for haptic displays. This machine is a portable desktop unit weighing 5 kg (including power amplifiers) with 4 N sustained force capacity. Its workspace is one-third the scale of the arm device, enough to allow a full range of fingertip motion when engaged in a precision pinch grip with the wrist supported. The machine is also partially balanced, reducing the portion of available actuator or human effort required to support the weight of the mechanism.

to effective access to and use of multilingual information, putting the U.S. at a distinct tactical disadvantage when operating in foreign lands and a strategic disadvantage in the international information arena. To help technical and military personnel overcome this barrier, DARPA has funded research on machine translation and algorithms for computer-based Translingual Information Detection, Extraction, and Summarization (TIDES) in FY 2000 and FY 2001.

The goal of the TIDES program is to enable English-speaking U.S. military users to access, correlate, interpret, and share multilingual information relevant to real-time tactical requirements, without requiring the user to have knowledge of a target language. This will require advances in cross-lingual information retrieval, machine translation, document understanding, information extraction, and summarization, as well as integration technologies yielding an end-to-end capability more valuable than these individual components. Achieving this goal will enable rapid correlation of multilingual sources of information to achieve comprehensive understanding of evolving situations for situation analysis, crises management, and battlespace applications.

Remote/autonomous systems

Remote Exploration and Experimentation (REE)

The goal of NASA's Remote Exploration and Experimentation (REE) project is to provide a new class of science mission that features fault-tolerant, high-reliability, low-power, and high performance supercomputing in space. These supercomputing systems will increase onboard processing capability by three to four orders of magnitude above what is currently available, permitting more data collection and processing, mitigating downlink limitations, and reducing ground station operations. With onboard supercomputing, a Mars rover, for example, would be capable of autonomous navigation and autonomous geological exploration and experimentation.

Additionally, REE is eliminating the need for radiation hardening, a task that puts space computers as much as five years behind Earth-bound technology. REE will allow NASA to use COTS computing technologies to ready computing devices for space flight 18 months after the technologies are available.

The robotic battlefield of the future

The battlefield of the future will demand weapons, unmanned combat vehicles, and communication systems that can navigate, reconfigure, and cooperate autonomously to accomplish time-critical commands. Many of these agents will be built from COTS products, while others, such as bio-inspired microrobots, are currently under development.

DARPA's mobile autonomous robot software project will develop and transition currently unavailable software technologies to program the operation of autonomous, mobile robots in partially known, changing, and unpredictable environments. The program aims to provide new software that will remove humans from the combat, conveyance, reconnaissance, and surveillance processes, thereby reducing danger to troops, lowering manpower costs, removing engineering constraints related to human physiology, and extending the range of military hardware. Researchers hope to demonstrate autonomous navigation by an unmanned military vehicle as well as interactions between humanoid robots, autonomous vehicles, and humans.

For robots to be fully integrated into the lives of humans in military, commercial, educational, or domestic contexts, they must be able to interact with humans in more meaningful, natural ways. DARPA is funding research to provide new ways for humans and robots to interact.

One potential way to bridge the gap between humans and robots is give them similar bodies. Given the premise that human-like intelligence requires humanoid interactions with the world, DARPA-funded MIT researchers are creating a robot, named Cog, as a set of sensors and actuators that approximate the sensory and motor dynamics of a human body. Cog's computational control is a heterogeneous network of different processors operating at different levels in the control hierarchy, ranging from small microcontrollers for joint-level control to digital signal processor networks for audio and visual preprocessing.

The Cog project's goal is to produce robots that can observe and respond to natural gestures and verbal instructions from a military commander. The commander should be able to demonstrate actions and supply auditory and visual cues to help the robot correctly perceive the instructions. For these imitative learning techniques to succeed, the robot must learn which aspects of the environment it should attend to and which actions it should reproduce.

The DARPA-funded Kismet project at MIT is training a robot head with eyebrows, eyelids, ears, and a mouth to discern social cues such as nodding or eye contact that are crucial in correctly guiding interaction.

Although successes with high-level, human-like modes of interaction are encouraging, some researchers are turning to other animals, such as ants and bees, for new approaches to robot design. DARPA's software for distributed robotics program is funding research to develop techniques for controlling large numbers of simple, computationally limited robots for missions such as deactivating minefields. For example, one DARPA-funded company is developing coin-sized robots that will respond to human direction and swarm targets such as chemical traces or human intruders.

Collaboration and virtual reality

Access Grid

NSF and DOE are participating in the Alliance's Access Grid, an ensemble of networked resources supporting group-to-group human interaction across the national technology grid. For further information, see page 42.

BioFutures

DARPA created the "Fundamental Research at the [Bio:Info:Micro] Interface" program, aimed at building interdisciplinary teams of researchers from the biology, IT, and microsystems technology fields to address fundamental research issues at the intersections of these three broadly defined areas.

The Bio component of the program includes disciplines from the molecular and cellular levels through the organism and population levels. The Info component includes the development of theories, algorithms, models and simulations, and scalable parallel and distributed systems. The Micro component includes development of sensors, materials, microfluidics, micromechanics, microphotonics, microelectronics, and large-scale systems created from such components. DARPA anticipates that long-term outcomes of this program will include the emergence of new interdisciplinary research communities and new science and technologies that will provide a foundation for revolutionary systems to satisfy future national and defense needs.

DeepView: A collaborative framework for distributed microscopy

DOE's DeepView is a scalable system integrating commercial software with unique computational components to enhance scientific problem solving within a collaborative framework. This "channel for distributed microscopy" offers a listing of online microscopes where users can participate in an experiment, acquire expert opinions, collect and process data, and store this information in electronic notebooks. Users can access capabilities including in situ microscopy, recovery of 3-D shapes through holographic microscopy, and image simulation for high-resolution transmission of electron microscopy.

Researchers are using DeepView to study the response of mammary tissue to low-dose ionizing radiation exposure. In the past, researchers have extrapolated the effects of low-dose exposure from the known effects of high exposure rates. Tissue response to radiation, and hence risk, is believed to be a composite of genetic damage, cell loss, and induced gene products. With DeepView, geographically distributed researchers can automate image acquisition and analysis to better understand the effects of radiation exposure and can capture and integrate the data into remote databases to compare data and generate simulations.

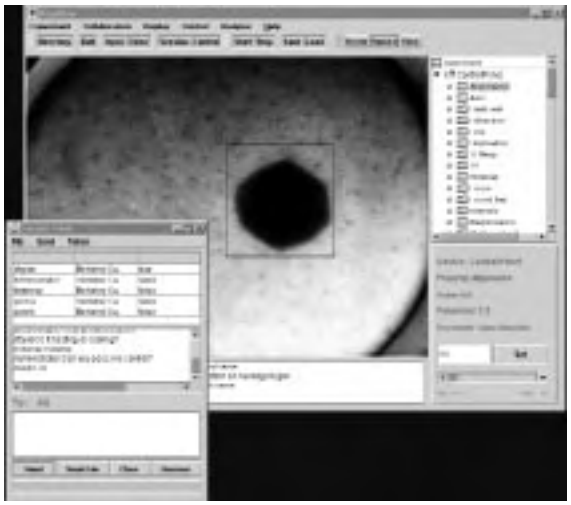
Distance visualization

A computational framework has been developed by ANL and the University of California to enable online 3-D tomographic image reconstruction using data, obtained from remote scientific instruments such as X-ray sources and electron microscopes, that are coupled to subsequent collaborative analysis. Using a combination of high performance networking and computing resources, parallel reconstruction algorithms, and advanced resource management, communication, and collaboration software, the research team demonstrated quasi-real-time 3-D imaging of samples on an advanced photon source tomographic beamline. Ten minutes after data collection commences, a 3-D image appears on the screens of project scientists at ANL and other institutions. Over the next 20 minutes, the image is progressively refined as more data are obtained. With this technology, scientists have, for the first time, the potential to change experimental parameters in the middle of an experiment.

The system has been demonstrated for data sources in Illinois and Japan with analysis devices located in California and Illinois. End users were in California, Florida, Illinois, and Japan. This research is supported by DARPA, DOE, NASA, and NSF.

Distributed collaboratories project

A scientific collaboratory environment is composed of many software components that work together to provide information, control, and a sense of presence to the remote researcher. These components include the infrastructures and tools needed to provide remote experiment monitoring and control,



DeepView is a scalable system integrating commercial middleware with computational components to enhance scientific problem solving within a collaborative framework. Researchers at several national facilities have used DeepView to conduct dynamic real-time collaborative experiments, such as this in-situ study of a lead inclusion at room temperature.

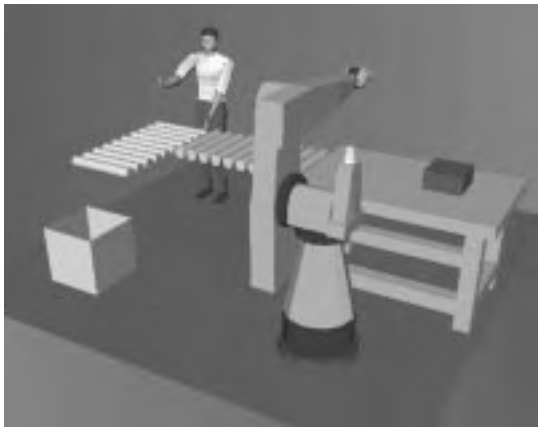
multimedia conferencing and telepresence, cross-platform compatibility, interface coherence and usability, and control coordination and conference management. The goal of DOE's distributed collaboratories project is to develop and deploy the technologies needed to enable access to and allow monitoring and control of remote experimental facilities from the home sites of researchers.

Through experience in building collaboratories, DOE researchers uncovered a need for better videoconferencing tools. Existing multicast-based tools are effective for broadcasting content and for one-to-one interaction, but more functionality is needed in the collaboratory environment. A collaborator must be able to activate a videoconference session, invite other participants to join, and carry on useful discussions with everyone participating equally. Additional tools needed include floor control, remote camera control, and remote control of the audio and video tools themselves.

Computer scientists at DOE laboratory LBNL have developed and deployed a system for remote control of video devices used in a videoconference. The system consists of a device server and a client that together let users control the videoconference cameras and video switcher over the Internet. The camera control tools are being adopted by the European Laboratory for Particle Physics (CERN) for use in its Internet conferencing software and have been used by the Internet Engineering Task Force (IETF) to broadcast meetings on the Internet.

Real-time collaborative access to chemical disaster information

NOAA's Pacific Marine Environmental Laboratory is exploring network-centric, Java-based methods for time-critical problem solving during chemical emergencies and other disasters involving hazardous materials. The initial phase of the project is focused on delivering situation-dependent data over the Internet from 12 source databases and integrating the data through a series of rules. With successful integration, a report previously run once every two years and in batch mode only will be updated and delivered with the latest data tailored to a specific situation as it evolves. NOAA researchers plan to surround this functionality with synchronous collaborative tools, enabling experts from around the U.S. to consult while maintaining a consistent shared view of the data. The research team is also working with EPA to ensure applicability and ease technology transfer.



Recent advances in shared distributed virtual environments (VEs) and embedded "smart" ubiquitous devices promise to revolutionize the ways we collaborate with people and interface with computing systems. NIST researchers are integrating these technologies and developing and applying tools for evaluating such environments. Applications include an interface to learn about and control devices in NIST's "smart room" and National Advanced Manufacturing Testbed (NAMT).

Manufacturing collaboratory

NIST is developing a manufacturing collaboratory to provide manufacturers, distributors, and researchers with structured methods and practices for implementing collaboratory technologies in manufacturing environments. Initial implementation of this collaboratory enhanced research in robotic arc welding, which requires asynchronous and synchronous collaboration support and the use of diverse data formats (such as video, still images, audio, text, database records, and image annotations) across a geographically dispersed community. This research resulted in a collaborative tool prototype with annotation capabilities that synchronize various data streams with indexing relative to time. The next step for the effort is to evaluate the collaboratory technologies in an industrial setting.

Manufacturing simulation and visualization

Simulation systems enable manufacturers to virtually prototype plant layouts, optimize material and component routings, and assess ergonomic factors prior to investing in plant redesigns or new factories. NIST researchers are investigating ways to augment commercial modeling and simulation software systems with programmable human ergonomic models, new human modeling programming, standard libraries for simulated manufacturing resources, integration of simulation systems in distributed computing environments, and validation of simulation algorithms. Researchers are developing natural language interfaces that can help simulate human tasks in a manufacturing operations environment.

NOAA's live access server (LAS)

Key requirements for distributed collaborations over the Web are the availability of data and common tools to browse, access, fuse, and analyze multiple distributed data sets. To meet this challenge, NOAA developed the live access server (LAS), a robust, extensible, collaborative data server for use by providers of gridded data from a variety of disciplines. Current users include:

- NOAA's Satellite Active Archive
- NOAA's Pacific Fisheries Environmental Laboratory
- NOAA's Pacific Marine Environmental Laboratory
- Office of Global Programs (OGP) Carbon Modeling Consortium
- Centre for Mathematical Modeling and Computer Simulation (India)
- Laboratoire de Météorologie Dynamique du Centre National de la Recherche Scientifique (France)

As the number of users grows, cross-disciplinary data sharing is expected to expand, leading to novel collaborations.

The tele-nanoManipulator

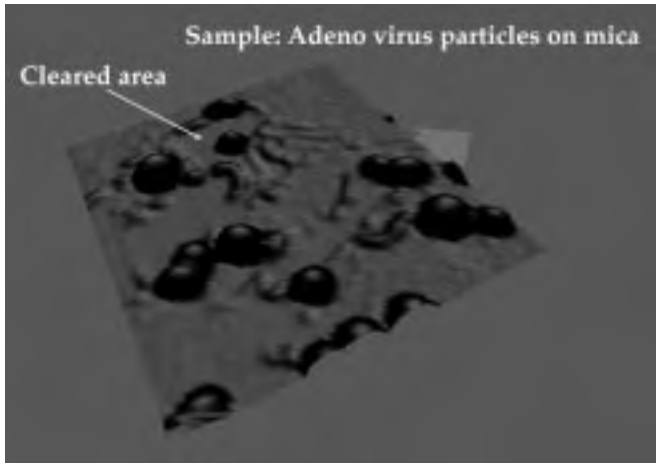
At the University of North Carolina-Chapel Hill (UNC), researchers have developed the tele-nanoManipulator system, allowing scientists to see, touch, and directly modify viruses, nanotubes, and other nanometer-scale objects over advanced networks using remote microscopes and graphics supercomputers in a unified system. The tele-nanoManipulator has been used to measure the rupture strength of DNA, measure the strength of the adenovirus capsid used in gene therapy, and build nanometer-scale circuits with carbon nanotubes. NIH, NSF, DoD's Army Research Office, DoD's Office of Naval Research, and industry sponsors support the UNC tele-nanoManipulator project. The tele-nanoManipulator team is also exploring how researchers, educators, and students anywhere on the Internet can participate in cutting-edge multidisciplinary science. During a visit to a local high school, UNC scientists used the system to enable the students to reach out and "touch" a virus over the Internet.

3-D visualization

NOAA inaugurated its collaborative immersive virtual environment testbed during FY 1999. In FY 2000, researchers are investigating new collaborative uses of the ImmersaDesk—a projection platform that "immerses" the user in a virtual reality environment—with the university community. These activities will move from the current desktop Virtual Reality Modeling Language (VRML, software for creating 3-D visual environments) world to larger immersive environments. Such 3-D techniques have already proven useful in a number of areas including data validation, where 3-D representations can highlight outliers and nonphysical attributes of geophysical data sets more quickly and intuitively than traditional 2-D plots.

Visualization and virtual reality for collaboration and manufacturing

NIST is partnering with industry to determine how 3-D visualization techniques and advanced collaboration tools can help improve commercial manufacturing processes and factory operations. Researchers are using VRML to interact with and visualize physical dynamics in virtual worlds and create virtual collaborative spaces that enable remote users to participate in design and engineering analyses. In FY 2000, NIST developed VRML "mirror worlds" that allow users to control and monitor the status of real-world devices. An existing camera interface is being extended within a JINI environment to enable



Researchers at the UNC Gene Therapy Center have used the tele-nanoManipulator to examine adeno virus particles. These particles are used as vectors in gene therapy, and scientists are seeking to understand how they stick to and move around on cell surfaces. A preliminary experiment is shown here, where the particles have been moved on a mica substrate. An area of the mica has been cleared of virus particles; the particles have been pushed together and then pulled apart.

additional services that allow users to control and monitor processes easily and intuitively.

Web-based tools for neutron research

NIST researchers are leveraging Internet capabilities to provide greater access to the agency's Center for Neutron Research, a unique national user facility that develops advanced instrumentation for materials research using neutron beams and makes it available to the private sector by developing methods that will allow external researchers to engage in experiments and analysis of data obtained using the facility via the Internet.

Web-based information resources

Digital Library of Mathematical Functions (DLMF)

In FY 2000, NIST began developing a digital library of certified reference data and associated information for the higher functions of applied mathematics—functions that aid engineers, scientists, statisticians, and others in scientific computation and analysis in areas as diverse as astronomy, atmospheric modeling, and underwater acoustics.

The NIST Digital Library of Mathematical Functions (DLMF) will be published on the Web within a structure of semantic-based representation, metadata, interactive features, and internal/external links. It will support user requirements such as simple lookup, search and retrieval in mathematical databases, formula validation and discovery, automatic rule generation, interactive visualization, custom data on demand, and pointers to software and evaluated numerical methodology. The online resource is expected to increase interaction between mathematicians who develop and analyze special functions and scientists and engineers who use them.

Some of the world's leading mathematicians—from the U.S., England, France, the Netherlands, and Austria—are participating in the project and developing much of the core material. NIST is exercising editorial control as well as developing and maintaining the Web site as a free public resource. The project, funded in part by NSF, is expected to be fully functional in FY 2003. The DLMF, successor to the classic 1964 National Bureau of Standards (NBS [now NIST]) Handbook of Mathematical Functions, is expected to contain more than twice as much technical information.

Clinical practice guidelines repository

The Agency for Healthcare Research and Quality (AHRQ), in partnership with the American Association of Health Plans and the American Medical Association, continued its development of a Web-based National Guideline Clearinghouse (NGC) in FY 2000. NGC is a publicly available electronic repository for evidence-based clinical practice guidelines aimed at helping health professionals improve the quality of care provided to patients. More than 500 clinical practice guidelines have been submitted to the NGC by physician specialty groups, medical societies, managed care plans, Federal and state agencies, and others.

Key components of the NGC include structured abstracts about the guideline and its development; a utility for comparing attributes of two or more guidelines in a side-by-side comparison; syntheses of guidelines covering similar topics, highlighting similarities and differences; links to full-text guidelines where available and/or ordering information for print copies; an electronic forum, NGC-L, for exchanging information on clinical practice guidelines; and

annotated bibliographies on guideline development methodology, implementation, and use.

Thousands of guidelines will ultimately be indexed, allowing rapid access to key recommendations and assessments on hundreds of topics. Individual physicians and other providers will be able to review and evaluate comprehensive sources of information to assist them with clinical decision-making and patient counseling. Health care systems and integrated delivery systems will use the information to adopt or adapt guidelines in their provider networks, and educational institutions will be able to incorporate NGC information into their curricula and continuing education efforts.

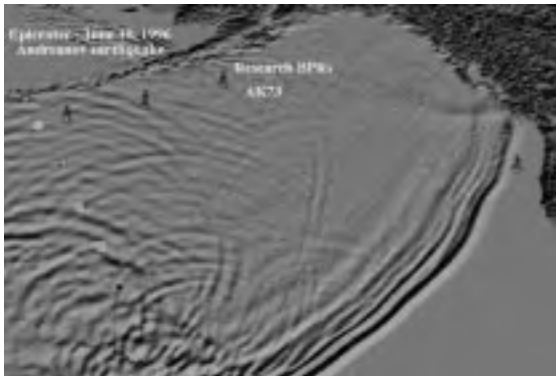
Surface wind analysis system

The surface wind analysis system at NOAA's Atlantic Oceanographic and Meteorological Laboratory (AOML) is a research tool for assimilating and synthesizing disparate weather observations—obtained from across the world's oceans by ships, buoys, satellites, and aircraft, and from coastal locations—into a consistent wind field. This tool takes advantage of computing and communications advances to obtain these data in near real time and provides specialists with an alternative to interpreting available data presented as text. NOAA is implementing a prototype database schema, developing object-oriented operational research and emergency management products with Web interfaces, and investigating efficient platform-independent programming technologies. Researchers have for several years used the system to provide experimental wind field guidance to forecasters, and have worked with hurricane specialists at NOAA's National Hurricane Center in analyzing wind observations during hurricane landfall episodes. System output will be formatted to work with FEMA's hazard loss estimation (HAZUS) wind module in GIS systems used by the emergency management community.

Web-based bioinformatics databases

Recent growth in the bioinformatics field has generated large sets of biological data on hundreds of thousands of living species, millions of DNA sequences, and tens of thousands of genes. Through its collaborative institute, the Center for Advanced Research in Biotechnology, NIST is developing two Web-based bioinformatics databases—one for molecular recognition and another for macromolecular structures—to help manage these data and advance experimental and computational research in macromolecules.

The molecular recognition database will be useful in drug design and chemical separations. It will contain the structure of pairs of target and reactant molecules, the properties of each, and their interactions and will be accessible via the Internet for submission of experimental results and content queries.



Researchers at NOAA's Pacific Marine Environmental Laboratory are conducting studies to understand how escarpments, ridges, and seamounts affect deep-water tsunami propagation in the Pacific Ocean, and to determine the accuracy and resolution of bottom topography needed for accurate simulation of tsunami propagation. The initial focus of this work is on tsunamis that are generated in the Alaska/Aleutian Subduction Zone and propagate southward toward Hawaii. The December 5, 1997, Kamchatka tsunami is simulated here.

The macromolecular database expands on the Protein Data Bank that NIST has owned since 1999. As a participant in the Research Collaboratory for Structural Bioinformatics, NIST is working to develop, refine, and apply evaluation algorithms to the contents to improve reliability and consistency and to transition to a new database management system.

Identification and security

Akenti: Collaborating in a secure environment

Collaborative technologies aimed at increasing communication between public and private research organizations require adequate and appropriate security that is flexible enough to accommodate the changing roles and requirements of the various participants.

Akenti is a security model and architecture developed at LBNL to provide scalable security services in highly distributed network environments. The goal is to express and enforce an access control policy without requiring a central enforcer and administrative authority. The resources that Akenti controls may be information, processing, or communication capabilities, or a physical system such as a scientific instrument. The Akenti model has been adopted at DOE and NASA.

Controlled sharing: The secure collaboratory

Researchers at DOE's LBNL and SNL have developed security tools to create a diesel combustion collaboratory that enables controlled sharing of data and resources among various national labs, diesel companies, and universities. With these security features, scientists can safely share their data and laboratories can grant outside access to powerful computing resources and instruments. In the collaboratory, members are identified by public key infrastructure (PKI) certificates, and resources are assigned access policies to control access to the shared resources.

Fingerprint and mug shot standards

Beginning with R&D in the 1960s in electronically comparing and matching fingerprint and mug shot images across different computerized systems, NIST and the Federal Bureau of Investigation (FBI) have expanded their collaboration to include developing standards and specifications for capturing and storing fingerprint and mug shot images and creating databases for evaluating techniques for processing those images.

As electronic means for exchanging fingerprint and other personal identification data replace the traditional hard copy form of this information, safeguards must be established to ensure the authenticity and integrity of the information. Beginning in FY 2000, NIST will develop a system in which a smartcard containing an individual's fingerprint and digital signature will be used in combination with the person's real-time scanned fingerprint image as a biometric to verify the authenticity of the electronic record's creator. The system will use a registry of legitimate digital signatures and the electronic signing of the transaction to ensure that the record is authentic and the data have not been corrupted since their creation.

Human ID

NIST researchers are conducting research to advance digital video human identification (ID) technologies. As an initial step in developing a baseline face recognition system, NIST developed the first support vector machine (SVM) face recognition system, which reduces recognition errors by a factor of two when compared to more conventional recognition methods. In partnership with

DARPA, NIST extended its human ID evaluation effort to include more complex video recognition tasks. In FY 2000, NIST began developing ground truthing procedures for digital video databases for the DARPA human ID program, implementing baseline digital video human ID systems, developing baseline evaluation procedures, collecting a digital video database, and performing computational psychophysics studies in collaboration with the University of Texas at Dallas.

Tools for analysis and productivity

Health care quality and clinical performance

AHRQ continued its development of COmputerized Needs-oriented QUality measurement Evaluation SysTem (CONQUEST), a tool permitting users to collect and evaluate health care quality measures and clinical performance measures. Clinical performance measures are tools that help assess the delivery of clinical services, such as appropriateness, safety, efficiency, timeliness, and competence. These measures enable comparisons over time within a provider group, among provider groups and organizations, and of performance results with goals. Because the methods must be uniformly specified and applied to all clinicians being compared, detailed specifications for data collection and analysis must first be developed and tested for comparability. CONQUEST uses two databases, one with almost 1,200 measures from 53 measure sets from a wide variety of organizations and the second with 57 clinical conditions, linked by a common language.

CONQUEST is used as a source of information about performance measures in the database, a source of information on specific conditions and the populations that experience them, and a tool for comparing and evaluating the worth of measures. It enables AHRQ to better disseminate knowledge about clinical performance measures and promote efficiency in clinical performance measurement.

Evaluating search engines

Since 1992, NIST and DARPA have sponsored the Text REtrieval Conference (TREC), a workshop series that evaluates search engines. A basic TREC task involves searching large amounts of text (around 2 GB) and producing ranked lists of documents that answer a specific set of test questions.

The TREC project has resulted in:

- More accurate text searching for all users
- Demonstrations of the scalability of statistical retrieval methods that provide an alternative to Boolean search methods
- New test collections
- Improved evaluation methods
- A forum for exchanging text retrieval research results
- An incubator for new research in cross-language retrieval (for example, English language questions to retrieve French documents), searching recorded speeches and news broadcasts, and question answering as opposed to document retrieval

The most recent TREC project involved 66 academic, industrial, and government groups from 16 countries.

Software usability testing

Testing business software generates information about qualities such as effectiveness, efficiency, and customer satisfaction. Computer scientists at NIST have launched a project to encourage software usability by standardizing how companies report their software testing and measuring how useful "common format" data are for potential purchasers.

Such standardization could save firms millions of dollars by reducing the lost productivity and huge training costs associated with buying software that is either poorly designed or inappropriate for a specific task. Participants in the project include many of the country's largest software producers and buyers: Boeing Co., Compaq Computer Corp., Eastman Kodak Co., Fidelity Investments, Hewlett-Packard Co., Microsoft Corp., State Farm Insurance Co., and Sun Microsystems Inc.

Web site usability tools

Usable Web sites are critical to industry for increasing e-commerce sales, improving worker productivity, and lowering costs and user frustration. Traditional methods to measure usability, however, are expensive, time consuming, and labor intensive. NIST is developing tools to help Web site designers and developers analyze sites for potential usability problems and to help usability professionals speed up their site evaluations. NIST researchers are working with industry to refine these tools and incorporate them into next-generation Web development and usability testing tools.

Digital Libraries

Overview

The invention of the printing press confronted 15th-century scholars and publishers with challenging problems as they attempted to transform hand-drawn and -lettered documents into multiple printed representations while developing libraries to house and organize them. Even so, by 1501 printers had produced some 20 million copies of 35,000 manuscripts, fueling the expansion of literacy and the spread of human knowledge beyond the social elites. By comparison, today's fast-evolving capabilities of computing systems and networks make it possible to recreate, archive, display, and manipulate exponentially greater quantities of electronically generated documents, data, images, sounds, and video streams—and to offer potential instant access to this knowledge to a significant portion of the world's population. But organizing "collections" of these many forms of electronic information and developing systems and software tools to make them available to end users require complex technical innovations, including collaboration among experts from widely disparate fields of knowledge.

Launched in 1994, the Digital Libraries Initiative (DLI) addresses the conceptual, structural, and computational challenges that must be met before we can realize the vision of universally accessible electronic repositories of human knowledge. Despite the modest scale of the initial DLI program, early interdisciplinary successes of participating researchers—demonstrated in such commercial spinoffs as the Lycos and Google search engines and Go2Net—attracted the attention of growing numbers of scholars and highlighted the enormous potential of digital information resources. DLI Phase Two, begun in FY 1999, spans a larger, more diverse set of research efforts that apply today's increasing computational and bandwidth capacities to the goal of making large-scale, distributed electronic collections accessible, interoperable, and usable through global knowledge networks. DLI Phase Two activities are jointly supported by NSF, DARPA, NIH/NLM, the Library of Congress, NASA, the National Endowment for the Humanities, and the FBI, in partnership with the National Archives and Records Administration, the Smithsonian Institution, and the Institute of Museum and Library Services.

DLI Phase Two directions

DLI Phase Two activities are drawing computer scientists and engineers from academia, industry, and government together with researchers and archivists in the humanities, the arts, and biomedical and physical sciences to develop new digital resource collections and testbed linkages among distributed archives; create frameworks, software, and network architectures that enable fusion of

multimedia materials into unified records; resolve semantic problems that currently prevent integration of digital resources from distributed collections; experiment with system designs to ensure the preservation, integrity, and privacy of data; and explore and codify educational applications of digital materials. Phase Two research focuses on three essential dimensions of digital libraries:

- *Human-centered research*—the ways in which digital libraries can improve and offer altogether new ways for people to create and use information
- *Content and collections*—the kinds of human knowledge digital libraries can house and make available to users
- *Systems-centered research*—the engineering, software, and taxonomic issues in creating and linking large-scale and disparate electronic collections via the Internet

Human-centered research

These activities explore next-generation methods, algorithms, and software that can empower expanded educational, professional, and personal uses of high-quality digital information resources. Research focuses on the development of intelligent search agents, improved abstracting and summarization techniques, advanced interfaces, and collaboration technologies and tools to enable individuals and groups to search for, retrieve, manipulate, and present electronic information archived in a variety of forms in a distributed network of source collections. Among the issues addressed by DLI Phase Two efforts:

Personalized Retrieval and Summarization of Image, Video, and Language Resources (PERSIVAL)

The explosion in Internet sites devoted to medical and health-related information makes it increasingly difficult for health care providers and consumers to find the most valuable and useful current resources. In the Personalized Retrieval and Summarization of Image, Video, and Language Resources (PERSIVAL) project, Columbia University researchers are experimenting with system designs to provide practitioners with quick and easy access to online medical resources tailored to individual patient needs. The goal is to develop personalized search and presentation tools to sort through distributed medical information, weed out repetitious and non-germane content, and summarize and present current findings that best match the real-time requirements of the practitioner or consumer. Using secure online patient records available at Columbia Presbyterian Medical Center as test models, the research team is linking a multimodal query interface with information extracted from a patient's medical record and user background to create a query graph for an online search of distributed medical resources. Search results are then filtered using natural language processing to provide the best matches with the patient's background. The results are presented in a customized multimedia format.

http://www.cs.columbia.edu/diglib/PERSIVAL/

Digital resources designed for children

The ways in which children ages 5 to 10 access, explore, and organize digital learning materials and the issues involved in creating learning environments suited to children's age-specific needs are the focus of a University of Maryland project. University researchers are fashioning developmentally appropriate tools for visualizing, browsing, querying, and organizing information in digital libraries designed for children. Audio, image, video, and text materials for the interdisciplinary research effort—which will include construction of a testbed digital collection about animals—are being made available by the Discovery

The University of Maryland has begun a partnership with children ages 5 to 10 and teachers from Yorktown Elementary School to create a multimedia children's digital library. This NSF-supported project will develop visual interfaces that support young children in querying, browsing, and organizing multimedia information, working with the children as "design partners" to develop new technologies that support the learning challenges in their age group.

Channel and the U.S. Department of the Interior's Patuxent (Maryland) Wildlife Research Center.

http://www.cs.umd.edu/hcil/kiddiglib/

Technologies and tools for students

Technologies and tools to make online educational resources more accessible and useful to communities of older learners, including college students and adults, are under development in several collaborative research efforts. For example, researchers at the Hypermedia and Visualization Laboratory (HVL) at Georgia State University and the Association for Computing Machinery (ACM) SIGGRAPH Education Committee are developing a model for a reusable national collection of peer-reviewed undergraduate educational applications in XML and improved navigation capabilities using information visualization techniques based on XML and 3-D Web graphics. Related work by researchers at the University of South Carolina in association with collaboratories at the University of Iowa and Georgia State focuses on creating a "Web-lab Library" of simulation software, experiments, and databases designed for students and researchers in the social and economic sciences.

http://econ.badm.sc.edu/beam/

Video information collage

Researchers at Carnegie Mellon University are creating an electronic workspace for video materials called a "video information collage," which will enable users to search for, view, and manipulate multiple video, text, image, and sound files from heterogeneous distributed sources. This will allow them to organize their discoveries into "chrono-collages" based on time relationships, "geo-collages" based on spatial relationships, or "auto-documentaries" preserving video's temporal nature. The research also involves creating a public video archive of recordings of historical, political, and scientific importance.

http://www.informedia.cs.cmu.edu/

Alexandria Digital Earth Prototype (ADEPT)

The Alexandria Digital Earth Prototype (ADEPT) program is a component of a large-scale digital library collaboration of the University of California-Berkeley, the University of California-Santa Barbara (UCSB), Stanford University, SDSC, and the California Digital Library (CDL). The ADEPT project builds on a DLI Phase One project that used UCSB's map and imagery collections to create a large-scale geospatial digital archive, called the Alexandria Digital Library (ADL), featuring maps, aerial photos, gazetteer items, and bibliographic records. In the ADEPT effort, researchers are constructing—and will evaluate the educational effectiveness of—customizable learning environments based on the ADL's geographically referenced contents, enabling students to bookmark and organize information from heterogeneous resources

and online services for multidisciplinary academic work. The ADEPT model employs a personalized interface called an Iscape, or Information landscape, with several layers of service and resource materials including meta-information tools indicating which resources in the personalized collection can be used collaboratively.

http://www.alexandria.ucsb.edu/adept/

Power browsers

Stanford researchers are experimenting with "power browsers"—handheld information appliances that access information sources, such as the Web, through wireless connections and software that maximizes the visual and navigation performance of very small displays. The software includes special information crawlers that save time by automatically performing certain search-related tasks. Researchers are also working on a large-scale "WebBase" database technology to store and index, for subsequent searching or analysis, millions of Web pages distributed across computers worldwide.

http://www-diglib.stanford.edu/

Content and collections

Researchers are creating novel digital archives of sound, image, and video as well as textual records from broad knowledge domains and specific disciplines in the sciences, arts, and humanities. They are evaluating methods of digital representation, preservation, and storage; exploring effective metadata systems (standard structures for presenting the intellectual context and pertinent related information about records in a collection); expanding access to educational materials and courseware; and developing technologies and protocols for addressing related legal and societal issues, such as copyright protection, privacy, and intellectual property management. Current research activities include:

Digital library for the humanities

Tufts University researchers, in partnership with the Max Planck Institute in Berlin, the Modern Language Association (MLA), the Boston Museum of Fine Arts, and the Stoa electronic publishing consortium, are developing the foundations for a scalable, interdisciplinary digital library accessible and useful to scholars as well as everyday Internet users. Materials included will date from ancient Egypt through 19th-century London. This site was processing 5 million requests per month in fall 1999.

http://www.perseus.tufts.edu

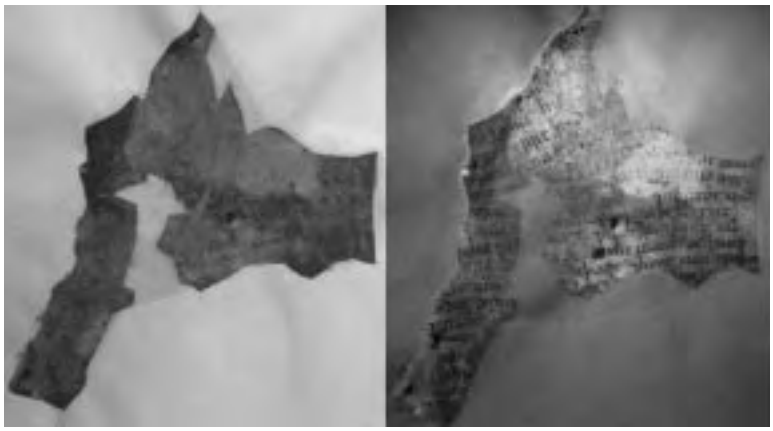
National Gallery of the Spoken Word (NGSW)

An interdisciplinary team at Michigan State University is building the Nation's first large-scale, fully searchable database and repository of historically significant audio materials spanning the 20th century. The "gallery" will also provide high-quality digital versions of such spoken words as Thomas Edison's first cylinder recordings and the voices of Babe Ruth and Florence Nightingale, with standard bibliographic and metadata access. A key research product will be a set of best practices for future Web sound development, including methods for conversion, preservation, access, and copyright compliance.

http://www.ngsw.org/app.html

National digital library for science, mathematics, engineering, and technology education (SMETE)

University of California-Berkeley researchers who developed the National Engineering Education Delivery System (NEEDS) digital library are exploring ways to expand the collection to encompass science, mathematics, and technology. The group is using its Web-based information portal, which



The NSF-supported Digital Atheneum project, based at the University of Kentucky, is developing state-of-the-art technologies to restore severely damaged manuscripts, ultimately presenting an electronic digital library of restored and edited images of previously inaccessible manuscripts. Illustrated here is a damaged manuscript as seen by the human eye (left), and with hidden markings revealed by ultraviolet digitization (right).

supports cataloguing, searching, displaying, and reviewing of digital learning materials and courseware, to begin developing a SMETE digital library, demonstrate the online resource's capabilities, and evaluate the initial SMETE testbed collection. The NSF-supported effort aims to create a broad-based digital learning resource for K-12 and postsecondary education.

http://www.needs.org

Digital Atheneum

NSF-funded researchers at the University of Kentucky, in partnership with the British Library and with support from IBM's Shared University Research (SUR) program, are developing state-of-the-art techniques to digitally restore and enhance aging and damaged original documents and create searchable archives of such materials. Working with documents from the British Library's Cottonian Collection (which contains Greek, Hebrew, and Anglo-Saxon manuscripts collected by 17th century antiquarian Sir Robert Bruce Cotton), they are testing new methods to illuminate otherwise invisible text and markings on documents and create digital annotation systems and semantic frameworks for domain- and data-specific searches of these materials.

http://www.digitalatheneum.org

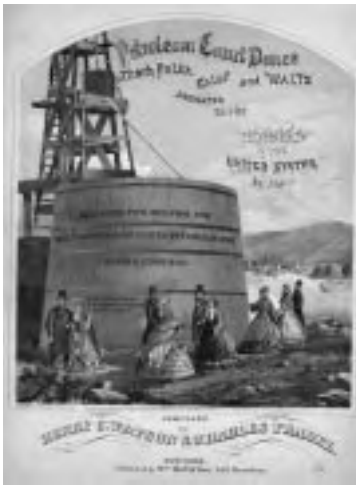
Digital workflow management

The more than 29,000 pieces of American popular sheet music in the Johns Hopkins University's Lester S. Levy Collection, already converted into digital records, will be made more accessible and usable through this project to create sound renditions and enhanced search capabilities. From items in the collection, which covers the period from 1790 to 1960, researchers will generate audio files and full-text lyrics using optical music recognition software written by staff of the Peabody Conservatory of Music at Johns Hopkins, and will develop workflow management tools to focus and reduce the amount of human labor involved. The activities will result in a framework, a tested process, and a set of tools transferable to other large-scale digitization projects.

http://levysheetmusic.mse.jhu.edu

Data provenance

Research at the University of Pennsylvania addresses one of the most difficult aspects of online resource collections: the questions surrounding the origin, or provenance, of an electronic record—such as how old it is, how it was originally generated, who produced it, and who has modified it. These questions are even more challenging in electronic than in traditional archives because the material involved ranges from a single pixel in a digital image to an entire database.



A treasure of the Special Collections Department of the Johns Hopkins University's Milton S. Eisenhower Library, the Lester S. Levy Collection of sheet music contains more than 29,000 pieces of American popular music spanning the period from 1790 to 1960, with a particular strength in the 19th century. All pieces are indexed at the library's online site, and visitors to the site can retrieve a catalog description of each piece. Additionally, for music published prior to 1923 and now in the public domain, a cover image and a page of music can be downloaded from the site.

Drawing on concepts from emerging software for presenting structured documents on the Web, researchers will develop prototype document "attachments" where annotations regarding provenance can be stored and queried, providing new data models, query languages, and storage techniques.

http://db.cis.upenn.edu/Research/provenance.html

Systems and testbeds

Systems research focuses on developing component technologies and the integration needed to create information environments that are dynamic and flexible; responsive at the individual, group, and institutional levels; and capable of continually adapting growing and changing bodies of data to new user-defined structures. These capabilities are prototyped and evaluated in testbed demonstrations that focus on media integration, software functionality, and breakthrough applications that offer transforming paradigms for social and work practices on a large scale.

New model for scholarly publishing

The current print model of academic publishing, based on centralized control and restricted distribution, originated long before the start of the information age. In another component of the large-scale DL collaboration at California institutions, University of California-Berkeley researchers are developing technologies and tools to create a distributed, continuous, self-publishing paradigm to use and disseminate scholarly information in this era of instantaneous global communication. The publishing system prototypes will be tested and demonstrated in the emerging CDL and on a testbed developed by SDSC.

http://elib.cs.berkeley.edu

Classification systems

Among the most complex technical challenges of digital archives is how to adapt or reinvent standardized identification and classification schemes for their contents, as well as interoperable search architectures that users need to locate these resources. On top of traditional print catalog taxonomies, archivists of electronic artifacts are juggling a number of new content categories (for example, video, image, sound, and software programs), formats (such as the jpeg and gif formats for graphic images), and related operational annotations. Researchers at the University of Arizona, in partnership with SGI, NCSA, NIH's NLM and NCI, GeoRef Information Services, and Petroleum Abstracts, are working on an architecture and associated techniques to automatically generate classification systems from large domain-specific textual collections and unify them with manually created classification structures. To generate and test prototypes, they are parallelizing and benchmarking computationally intensive scalable automatic clustering methods for keyword searching on large-scale collections with existing classification systems such as CancerLit (700,000 abstracts) and the NLM's UMLS (500,000 medical concepts) in medicine; GeoRef and Petroleum Abstracts (800,000 abstracts) and GeoRef Thesaurus (26,000 terms) in geoscience; and on Web applications including a collection with 1.5 million Web pages and the Yahoo! classification system (20,000 categories). Using simulations on parallel, high performance platforms, scientists will optimize and evaluate the output of the various algorithms and develop hierarchical display methods to visualize the results.

http://ai.bpa.arizona.edu/go/dl/

Virtual workspaces

Even when digital collections are structured and catalogued internally, many users also need something akin to a large, well-lit library table on which they can spread out items from various sources to work with and compare. Harvard-MIT Data Center researchers are jointly designing a Virtual Data Center (VDC) to manage and share quantitative social-science materials for research and teaching among multiple institutions and the public. The VDC will link with other research centers and databases, enabling participants to deposit data in many formats and set terms of access to their materials. Users will be able to download data containing only the variables they specify. The VDC's suite of open software tools ultimately will be offered as a free, portable product.

http://www.thedata.org

Security, quality, access, and reliability

In addition to effective classification systems and tools for users, the infrastructure of digital libraries, like that of their physical counterparts, requires systems ensuring the physical security of the collection, quality control, and remote access to the contents. Stanford University researchers are exploring ways to guarantee the long-term survival of digital information despite media obsolescence, natural disasters, and institutional change. They are prototyping techniques for automatically monitoring changes in collections and continuously "mirroring" the information into a large-scale archive that is automatically replicated at other sites. The prototype also uses mathematical models of projected failures in storage media to alert human operators to possible malfunctions.

At Cornell University, researchers are focusing on the integrity of digital library information, devising prototype administrative architectures to ensure that archived information is reliable and readily available, and that the intellectual-property rights of authors and the privacy rights of users are protected.

http://www.prism.cornell.edu

International efforts

In only a few years, DL activities have expanded to encompass not only digital construction work on important human records but also international collaborations to facilitate universal access to these new information resources.

U. S.-U.K. activities

An initiative of NSF and Britain's Joint Information Systems Committee, for example, supports international research to solve fundamental technical problems in linking and accessing geographically distributed materials in differing formats. These projects include the following:

- A University of California-Berkeley/University of Liverpool Library study will focus on enabling cross-domain searching in a multidatabase environment. The aim is to produce Cheshire, a next-generation online information retrieval system based on international standards, for Internet searches across collections of original materials, printed books, records, archives, manuscripts, museum objects, statistical databases, and full-text, geo-spatial, and multimedia data resources.
- HARMONY, a three-way partnership of Cornell University, the Australian Distributed Systems Technology Centre, and the University of Bristol (UK) will devise a metadata framework for describing networked collections of complex mixed-media digital objects. The research will draw together work on the Resource Description Framework (RDF),

XML, Dublin Core, and Moving Pictures Expert Group's MPEG-7—all of which are standards for representing and exchanging structured metadata. The goal is to enable multiple communities of library, education, and rights management experts to define overlapping descriptive vocabularies for annotating multimedia content.
http://www.ilrt.bris.ac.uk/discovery/harmony/

- A demonstration by Cornell University, the Los Alamos National Laboratory, and the University of Southampton (UK) will hyperlink each of the more than 100,000 papers in Los Alamos's online Physics Archive to every other paper in the archive that it cites. The project aims to highlight this powerful tool for navigating the scientific journal literature to encourage authors in other fields to join in creating similar hyperlinked online archives across disciplines and around the world.
http://journals.ecs.soton.ac.uk/x3cites/

- A tool for locating music itself online—a type of search that is not currently possible—is the goal of a collaborative effort by researchers at the University of Massachusetts and Kings College, London. This online music recognition and searching (OMRAS) tool will enable users to find musical information stored in online databases in formats ranging from encoded score files to digital audio.
http://journals.ecs.soton.ac.uk/x3cites/

- The IMesh Toolkit project, a partnership of the University of Wisconsin-Madison, the University of Bath (UK), and the University of Bristol (UK), uses an emerging approach to accessing Internet resources through highly selective, subject-specific Web sites called "subject gateways," and builds on the IMesh international collaboration among leading subject gateway developers. The research aims to advance the framework within which subject gateways and related services operate by defining an architecture specifying individual components and how they communicate. The architecture will allow interoperability and cross-searching between subject gateways developed in different countries.
http://www.imesh.org/toolkit/

- University of Michigan researchers, working with representatives of Britain's Consortium of University Research Libraries, are investigating the potential role of emulation in long-term preservation of information in digital form. The project will develop and test a suite of emulation tools; evaluate the costs and benefits of emulation as a preservation strategy for complex multimedia documents and objects; devise models for collection management decisions about investments in exact replication in preservation activity; assess options for preserving an object's functionality, "look," and "feel"; and generate preliminary guidelines for use of different preservation strategies such as conversion, migration, and emulation.

U.S.-Germany activities

In January 2000, NSF/DLI Phase Two and Germany's Deutsche Forschungsgemeinschaft issued a joint call for collaborative proposals from U.S. and German university researchers on developing and organizing internationally accessible digital collections.

NSF-EU working groups

The Joint NSF-European Union (EU) working groups on future directions for digital libraries research have completed their initial studies of national, technical, social, and economic issues and plans for common research agendas. Five working groups—each of which includes U.S. researchers from academia, industry, and government—addressed economic issues and intellectual property rights, interoperability among digital library systems, metadata, multilingual information access, and resource indexing and searching issues in globally distributed digital libraries. The final report, entitled "An International Research Agenda for Digital Libraries," and working papers can be found at: *http://www.si.umich.edu/UMDL/EU_Grant/home.htm* and *http://www.iei.pi.cnr.it/DELOS//NSF/nsf.htm.*

Large Scale Networking

Overview

In education, in health care, in science, and in business, astonishing advances in networking technologies are driving fundamental changes in society, dramatically influencing the way Americans pursue their everyday activities. From instantaneous stock market quotations to online art auctions, from Web-based U.S. postal zip code lookups to nationwide law enforcement databases, it is easy to see the significant impact of Federal agency support for Large Scale Networking (LSN) on today's evolving environment. LSN has become a pivotal force, generating critical advances in technologies that have been quickly adopted by the governmental, academic, and commercial sectors.

Federal LSN R&D includes:

- Traditional networking research to support agency mission requirements
- The Next Generation Internet (NGI) initiative
- Research in Scalable Information Infrastructure (SII)

LSN programs explore long-range fundamental networking research issues and transition developing LSN products into tools to support agency missions. Continuing the Federally supported R&D responsible for the core technologies that made the Internet and Internet applications possible, LSN focuses on technologies needed by the Federal agencies, infrastructure to support agency networking, and networking applications development.

Since its inception in 1998, the NGI initiative has been a primary focus of LSN, building on the LSN base programs to provide the networking research, testbeds, and applications needed to assure the scalability, reliability, and services required by the Internet over the next decade. In FY 2000, with the construction of its testbeds largely completed, NGI's original three goals were refocused as two thrusts that emphasize improved network performance and functionality and revolutionary networking applications.

In its 1999 report, "Information Technology Research: Investing in Our Future," the PITAC warned that Federal support for IT R&D was seriously inadequate. Research programs intended to maintain the flow of new IT ideas and train the next generation of researchers were funding only a small fraction of the necessary research. The PITAC recommended a significant new program and increased funding for IT R&D and an expanded Federal role in networking R&D that includes interoperability and usability.

Federal agencies responded to this challenge with a proposed new program in IT R&D. A major component is SII, whose research goal is to develop tools and

LSN teams

The LSN Coordinating Group (LSNCG)—formerly the LSN Working Group (LSNWG)—coordinates multiagency Federal networking R&D programs. Four teams report to the LSNCG to assist in this task and to help implement advanced networking technologies:

The Joint Engineering Team (JET) coordinates the network architecture, connectivity, exchange points, and cooperation among Federal agency networks (FedNets) and other high performance research networks, and provides close coordination of connectivity, interoperability, and services among Government, academia, and industry to improve end-to-end user performance and avoid duplication of resources and efforts. The JET also coordinates international connectivity and interoperability.

The JET cooperates with the academic community's gigabits per second points of presence (Gigapops), the Abilene network (a consortium among Qwest, Cisco, Nortel, and Indiana University), and the University Corporation for Advanced Internet Development's (UCAID's) Internet2 (I2). During FY 2000, the JET is developing and coordinating implementation of the Next Generation Internet Exchange Points (NGIXs), where connectivity and peering of the FedNets take place (page 89).

To date, NSF has made High Performance Connection (HPC) awards for NGI connectivity to 167 universities and four high performance computing centers, with other NGI sites supported by DARPA's SuperNet (15), NASA (5), DOE (5), and NLM (1). The JET continues to coordinate the connection of additional sites to the NGI testbeds, helping to implement multicast on the NGI testbed backbone and provide metrics to improve performance.

The JET supports cooperation among the agencies and with Abilene to provide improved, lower-cost services to geographically challenging areas such as Hawaii and Alaska. In FY 2000, the JET is coordinating an increase in research connectivity to Hawaii and Alaska.

The JET coordinates connectivity requirements for applications demonstrations, using NGI testbeds at supercomputing conferences and other demonstrations. The JET and other LSNCG Teams contributed to LSN's "Bridging the Gap Workshop" held August 11-12, 1999, at NASA's Ames Research Center at Moffett Field, California. This workshop brought together networking researchers and NGI applications developers to promote increased awareness and cooperation, providing roadmaps and schedules for developing and implementing emerging technologies.

The Networking Research Team (NRT) coordinates agency networking research programs, shares networking research information among Federal agencies, and supports NGI Thrust 1 activities. It provides outreach to end users by disseminating networking research information and coordinating activities among applications developers and end users.

The NRT is active in developing agency workshops on middleware. In cooperation with the other LSN Teams, the NRT organized the "Bridging the Gap Workshop," which identified three key areas of networking research—quality of service, multicast, and Internet security—as the highest priorities for achievable technology advances within the next one to three years. In December 1999, NRT coordinated an NGI research/Principal Investigator (PI) meeting with researchers from DARPA, NSF, and NIST to disseminate research results and increase collaboration.

The High Performance Networking Applications Team (HPNAT) coordinates Federal R&D to maintain and extend U.S. technological leadership in high performance networking applications, encouraging research that employs advanced networking technologies, services, and performance to support leading-edge applications. Advances will lead to new and more capable network applications that support Federal agency missions and build the foundation for the evolving national information infrastructure.

The HPNAT promotes cooperation in large-scale networking applications development among Federal agencies, Government laboratories, academia, and industry, and organizes information dissemination activities including technology demonstrations, workshops, and seminars. The HPNAT supports NGI Thrust 2 by helping organize NGI demonstrations at conferences such as SC99 in Portland, Oregon (page 99). The HPNAT coordinated the identification and presentation of 15 application case studies (see *http://www.nren.nasa.gov/case.html*) at the "Bridging the Gap Workshop." They were selected to represent nationally significant needs and broad application areas for the technologies examined in the workshop.

The Internet Security Team (IST) facilitates testing of and experimentation with emerging advanced network security technologies. It provides the LSNCG with feedback and direction for NGI research in network security by serving as a forum for the exchange of security requirements and needs and current and emerging security technologies. The IST encourages development and use of Internet security testbeds, helping LSN agencies and the JET to implement these testbeds and publicize testbed activities to national and international security research communities. In January 2000, the IST began planning for an ongoing series of "Workshops on Public Key Infrastructure (PKI) for Advanced Network Technologies," with the first held in April 2000. Workshop objectives are to provide a roadmap for developing and implementing PKI at the point of human-computer interface and to implement a multiagency/commercial testbed for developing Internet security standards.

techniques enabling the Internet to grow (scale) while transparently supporting user demands. An integral part of LSN, SII R&D complements the base LSN and the NGI efforts. SII research will focus on deeply networked systems, anytime, anywhere connectivity, and network modeling and simulation.

The following section of this report describes the FY 2000 accomplishments and FY 2001 plans for base LSN, NGI, and SII R&D.

LSN networking infrastructure support

Research networks

Agencies' base LSN R&D activities include providing the networking infrastructure for scientific research that addresses the mission requirements of participating agencies and developing enabling technologies and applications to expand the global-scale capabilities of the Internet. LSN supports research networks such as NASA's Research and Education Network (NREN), DOE's Energy Sciences Network (ESnet), and others, and network management such as at NSF's National Laboratory for Applied Network Research (NLANR). LSN infrastructure support includes implementing evolutionary leading-edge technologies for Federal agency use, including coordination of:

- Multimodal network testbeds
- Satellite interconnects
- Wireless technology development including standards and testing
- Network-attached devices
- Network-attached services
- Software objects
- Data sets

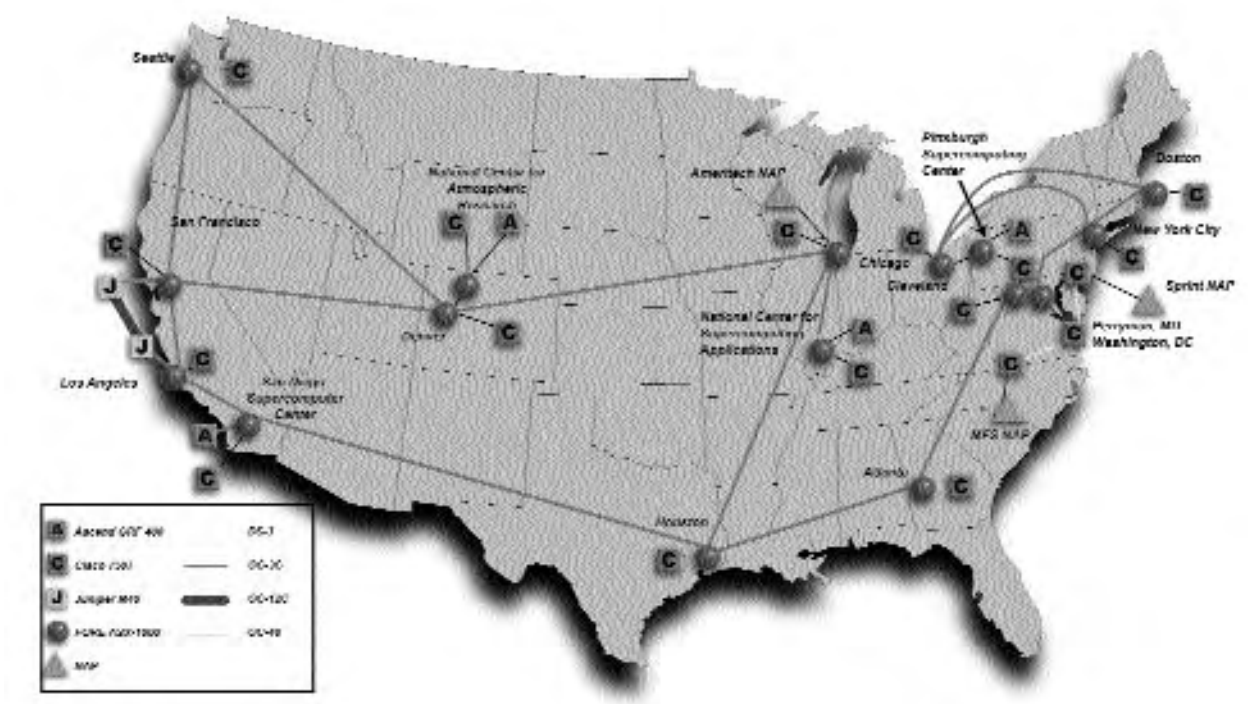
LSN coordinates R&D initiatives with other communities such as the Interagency Engineering Task Force+ (IETF+).

Measurement and network analysis

To meet the measurement and network analysis challenges posed by the rapid growth and expansion of the Internet, NLANR is building and operating a network analysis infrastructure with a primary focus on NSF High Performance Connections sites. R&D areas include active performance measurement, data collection from network entities via Simple Network Management Protocol (SNMP), data related to stabilities and status of Internet routing, and passive header trace data.

Fundamental networking research

NSF's networking research focuses on the fundamental science and technology needed to facilitate efficient, high-speed information transfer through networks and distributed systems. Funded projects range from network design and performance evaluation to middleware and software frameworks supporting applications on networks and distributed systems. They also address how networks and distributed systems interact with underlying communications technologies. Research areas include agent-based networks, high-speed networks, multicast, multimedia applications, multiple access protocols, network architectures, network design, network management, network security, network systems, object-oriented frameworks for networks, optical networks, performance



NSF's cooperative agreement with MCI WorldCom to provide very high performance backbone networking services (vBNS) to universities ended March 31, 2000. NSF has negotiated and completed a no-cost extension of these services.

evaluation, protocols, quality of service (QoS), resource management, traffic control, and wireless and mobile networks.

Special networking research projects

NSF special projects support larger and in some cases multidisciplinary efforts, specialized hardware and software research, and networks for networking systems research, and help develop research agendas and enhance community development. Accomplishments include production and distribution of a gigabit asynchronous transfer mode (ATM) switch kit and Internet Protocol version 6 (IPv6) development and source distribution. Current and anticipated projects may include all-optical networking, gigabit networking, communications, control theory, databases, devices, distributed systems, operating systems, research priorities in wireless and mobile communications and networking, software, and signal processing. Proof-of-concept demonstrations of novel networking ideas may range from laboratory experiments to national collaborations.

Advanced Networking Infrastructure (ANI)

The ANI program supports research, development, implementation, and testing of high performance network testbeds and related technologies to support the distributed IT goals of the U.S. research and education community. ANI makes the High Performance Connections awards to colleges and universities, stimulating improvements in campus research networking infrastructure and encouraging development of high performance applications requiring advanced networking services and speeds. The ANI program supports NLANR, which provides technical, engineering, and traffic analysis support for NSF High Performance Connections sites and high performance network service providers (HPNSPs) such as the NSF/MCI WorldCom vBNS.

In FY 1999, the high performance network increased its speed by activating an OC-48 (2.4 Gbps) research link between Los Angeles and San Francisco to serve the CalREN-2 university research and education network. The OC-48 upgrade allows for experimentation with multicast, IPv6, and QoS classes.

NSF's cooperative agreement with MCI WorldCom to provide very high performance Backbone Network Services (vBNS) to universities ended March 31, 2000. NSF has worked with MCI WorldCom, its vBNS+ service, and Abilene to effect a smooth transition for the universities being served.

Internet technologies program

The Internet technologies program focus areas include complex network monitoring and problem detection and resolution mechanisms; automated and advanced network tools, networked applications tools, and network-based middleware; and usable and widely deployable networking applications that promote collaborative research and information sharing. Current research includes:

- *Prototype testing and evaluation of wireless instrumentation for ecological research at remote field locations.* In conjunction with NSF's long-term ecological research (LTER) project, the goal is to develop remote sensing devices for retrieving field data, from the frozen lakes of Wisconsin to rugged mountain terrain in Puerto Rico.
- *Surveying the digital future.* This project is studying the impact of personal computing and Internet technologies on society and families. It will address the PITAC's call for research on socioeconomic issues.

STAR TAP

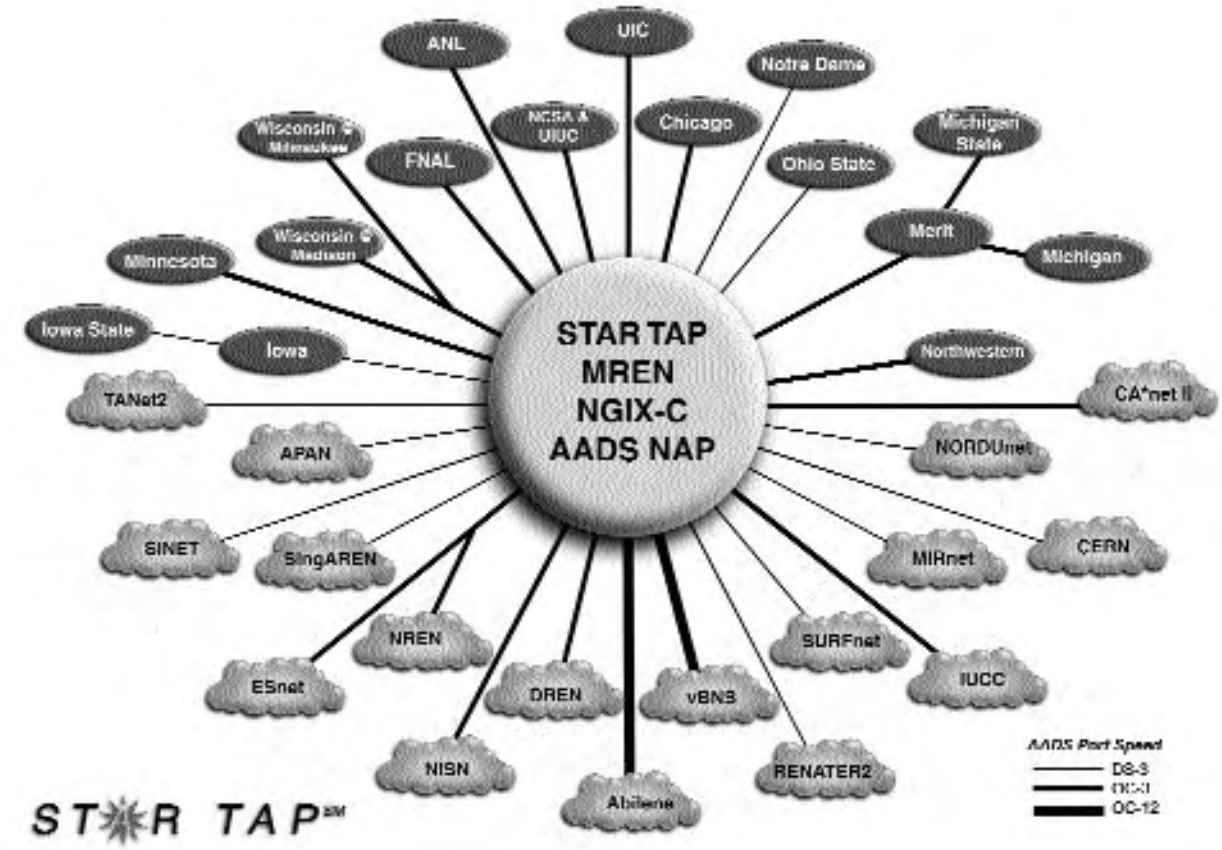
NSF's ANI program coordinates international networking activities that connect the U.S. research and education community with academic counterparts and resources around the world. ANI established the Science, Technology, And Research Transit Access Point (STAR TAP℠) in Chicago to interconnect the vBNS with international advanced networks to support high performance applications and develop new networking technologies. STAR TAP is managed by the University of Illinois at Chicago's EVL and operated by Ameritech Advanced Data Systems.

STAR TAP facilitates the long-term interconnection and interoperability of advanced international networking to support applications, performance measuring, and technology evaluations, anchoring the international vBNS connections program and allowing collaboration with the NGI initiative and the I2 community. STAR TAP has connections with:

Federal agency NGI backbone networks:

- DoD's Defense Research and Education Network (DREN) (155 Mbps)
- DOE's ESnet (155 Mbps)
- NASA's Integrated Services Network (NISN) and NREN (155 Mbps)
- NSF's vBNS (155 Mbps)

Other U.S. networks:



Schematic map of the Science, Technology, And Research Transit Access Point (STAR TAP) located in Chicago, Illinois. STAR TAP facilitates the long-term interconnection and interoperability of advanced international networking to support applications, performance measuring, and technology evaluations, anchoring the international vBNS connections program and allowing collaboration with the NGI initiative and the I2 community.

- Abilene/University Corporation for Advanced Internet Development (UCAID)
- vBNS (after March 31, 2000)

International networks:

- APAN (Asia-Pacific: Japan, Korea, Australia, Singapore [second connection] and Thailand, 70 Mbps)
- CA*Net (Canada, 45 Mbps)
- CERN, IUCC (Israel, 45 Mbps)
- MirNET (Russia, 6 Mbps)
- NORDUnet (Denmark, Finland, Iceland, Norway, Sweden, 45 Mbps)
- RENATER2 (France, 45 Mbps)
- SingAREN (Singapore, 14 Mbps)
- SURFnet (The Netherlands, 155 Mbps)

- TANet2 (Taiwan, 15 Mbps)
- TransPAC (Japan, Korea, Singapore, Australia, 35 Mbps)

Several other countries in Asia and South America have expressed their intention to connect to STAR TAP. Activities now focus on improving performance with higher-level services.

STAR TAP provides an integrated approach to the management, scheduling, consumption, and performance measuring of geographically distributed network, computing, storage, and display resources—a collection of resources called the international Grid (iGrid). The iGrid, with the support of STAR TAP, demonstrated international collaborations at SC99. In FY 2000, STAR TAP is moving to a distributed architecture with points of connectivity on both coasts and in Chicago. International STAR TAP participants provide U.S. transit.

Wireless standards

Today's new markets make extensive use of low-cost, distributed, embedded devices whose architecture supports mobile and self-configuring environments with a combination of wired and wireless technologies. Further development of these markets is hindered, however, by a lack of widely accepted standards. NIST is working with standards organizations on issues such as connecting portable wireless devices to traditional networks, service discovery, self-configuration and dynamic resource sharing, mobile code and data, and the software infrastructure needed to create and manage pervasive services and applications.

Active networks

DARPA's active networks program is creating a flexible and extensible networking platform to accommodate the rapid evolution and deployment of networking technologies and to support the increasingly sophisticated services demanded by defense applications. The "active nets" architecture is based on a highly dynamic runtime environment that supports finely tuned control of network services. The packet itself is the basis for describing, provisioning, or tailoring resources to meet delivery and management requirements.

Active networks goals include quantifiable improvement in network services such as fault tolerance, multitiered mobile security, dynamic access control, and audio/video synchronization and full-rate video over multicast. Some of the challenges in meeting these goals are the defining composite protocols; efficient, secure, and survivable "smart packet" processing; new strategies for routing and service provisioning in large networks with overlapping topologies and mobility requirements; and upgrading services to keep pace with network complexity.

Tolerant networking

DARPA's tolerant networking R&D will develop technologies to support continued operation during attacks on networks, addressing vulnerabilities and issues expected to arise in DoD's emerging network-centric warfare vision. Tolerant networking technologies will strengthen networks by introducing fault-tolerant defenses that emphasize integrity and availability and developing technologies to mitigate potential vulnerabilities in the dynamic creation and management of mission-driven coalitions. Examples are thwarting denial-of-service attacks by constraining an attacker's resources and enabling secure collaboration within dynamically established, mission-specific coalitions while minimizing potential threats from increased exposure or compromised partners.

Wireless information technology and networks

The coming generation of terrestrial wireless and satellite communications technologies promises a revolutionary leap forward in information accessibility and an increase in economic yield. In just 20 years, wireless personal communications services have gone from concept to connecting more than 300 million subscribers globally, or 40 percent of the world's installed wired telephone base. Current trends point to a revolution in wireless mobile IT within the next couple of decades that will provide complete freedom of location to the individual. Future-generation wireless systems R&D will enable the U.S. to compete successfully in a worldwide wireless market expected to grow to 700 million subscriber units by 2002.

Research topics include not just bandwidth but management to accommodate a wide mix of traffic types, power limitations, and scalability. Potential applications include crisis management, navigation (for example, in cars or through large buildings), telemedicine (not only for physicians, but for on-site paramedics), and learning, especially in remote locations. Current activities include:

- Research at Rice University on a seamless communication platform that will function in environments as different as high-speed indoor wireless and conventional cellular systems
- Columbia University research on a comprehensive QoS architecture for adaptive real-time services in mobile ad hoc networks, wireless flow management, and admission control algorithms that allocate bandwidth to adaptive real-time flows
- University of Southern California research on new concepts of directed diffusion and behavior networks to address the boundary problem between applications and networks
- University of Arizona research to develop a wireless architecture that uses "fine-grain configurability" and "fast coordinated adaptation" as cornerstones to provide QoS for applications in diverse and dynamically changing environments

Fault-tolerant networks

Denial-of-service attacks—which take advantage of high-cost protocol checks such as authentication in order to consume resources—can easily disrupt and cripple network operation by rapidly consuming all available network resources, rendering the network inaccessible to legitimate users. Fault-tolerant networks ensure continued availability and graceful degradation of the network infrastructure under partially successful attacks, maximizing network capacity for legitimate users.

Dynamic coalition management

Traditional system designs have "central nervous systems" that, if attacked, can react by completely disabling the system. Corrupted or malicious components can also lead to a malfunction of the entire system. Dynamic coalition management enables secure collaboration within dynamically established, mission-specific coalitions while minimizing potential threats from increased exposure or compromised partners. R&D in this area will develop the capability to dynamically manage and validate operational policy configurations across multiple theaters, securely manage information dissemination within large groups, and augment existing PKI technologies to accommodate rapid revocation and cross-certification.

Quorum: End-to-end mission success

Defense applications require seamless interoperability, distribution over multiple nodes, and information sharing to support rapidly organized joint and coalition missions. Such systems will be composed of rapidly evolving COTS components deployed in highly dynamic hostile environments. But today's commercial technology emphasizes functional interoperability with virtually no assurance of or control over mission-critical properties such as timeliness, precision, reliability, and security and acceptable tradeoffs among them.

Advances in networking and computing have spurred research and development in distributed computing, ranging from high performance workstation clusters to wide area information retrieval and collaboration

environments. These approaches typically emphasize either integrated solutions, which are tightly bound to particular operating systems and platforms, or overlays that support portability at the expense of performance. No single approach exhibits scalability from local to wide area environments, and none can provide the assured service and adaptivity necessary to support mission-critical Defense applications in the dynamic, shared, heterogeneous, wide area environments now emerging.

DARPA's Quorum program is developing technologies for an operating system to support global distributed computing for mission-critical applications. The Quorum program premise is that end-to-end QoS management across middleware, operating systems, and networks is the key to providing applications with the end-to-end assurance needed to guarantee mission success in highly dynamic, unpredictable networked environments; scalability over heterogeneous resources whose performance may span a range of several orders of magnitude; and evolvability to keep pace with technological advances. The Quorum program is structured as three interrelated technology development tasks and a fourth integration and demonstration task.

QoS architecture

Technologies include methods to specify application-level QoS constraints and allowable tradeoffs; protocols and assurance bounds for QoS negotiation; algorithms for mapping application-level expectations to individual resource and service constraints; instrumentation and analysis methods for dynamically monitoring delivered QoS; and protocols for providing feedback to applications and for triggering adaptation or renegotiation if necessary.

Translucent system layers

In distributed computing, functional transparency describes a situation in which implementation decisions that have an impact on mission-critical properties are visible to the user, yet frustratingly uncontrollable. R&D in this area will develop translucent services that preserve the benefits of functional transparency while remaining dynamically responsive to QoS constraints imposed by higher layers or feedback from lower layers or the environment, adapting their behaviors through selection or specialization of alternative implementations, policies, or mechanisms. Specific layers addressed include virtual machine APIs, distributed shared memory, distributed objects, communication services, and operating systems.

Adaptive resource manager

Quorum will develop the resource management technologies to dynamically discover, allocate, and schedule resources from a global heterogeneous pool to an application in accordance with its negotiated QoS constraints. Technologies being developed include methods for characterizing resource capabilities and mapping them to application requirements and profiles; technologies for collecting and maintaining a consistent global view of resource status; and dynamic resource allocation algorithms that yield near-optimal performance and support adaptation in response to workload demands, failures, information warfare attacks, or crisis modes.

Integration, demonstration, and validation

Realizing the Quorum vision requires coordinated development of constituent technologies, their integration into complete reference implementations of successively greater capability, and their evaluation and demonstration for realistic defense applications. The principal DoD testbed is the Navy's 21st

Century Surface Combatant program, which is exploring architectures and technologies for the next generation Aegis combat control system.

Sensor information technology (SensIT)

Microfabricated sensors are the interface between the physical world and the information world of the future. Delivering massive amounts of data cheaply and in real time, these sensors will be a crucial part of decision-making in the battlefield, surveillance with minimal manpower, and maintenance of equipment. DARPA's sensor information technology (SensIT) program is dedicated to maximizing the useful information that a network of thousands of sensors can deliver to key decision-making points in a timely manner. SensIT's mission is to develop all necessary software for a networked system of inexpensive, pervasive platforms that combine multiple microsensors, embedded processors, positioning ability, and wireless communication. The resulting technology will perform as if a supercomputer were miniaturized and distributed into the environment, with each node computing and collaborating to "see" into its sensor region.

Research challenges include developing large-scale sensor networks employing algorithms for self-assembly of highly dynamic ad hoc networks with minimal latency and maximal survivability, nanocryptography for security, easy-to-use querying and tasking languages, operating environments, distributed asynchronous algorithms for collaborative signal processing, and internetworked fixed and mobile sensors on humans, microrobots, and vehicles. (DARPA also supports software research in this area in the SDP PCA [page 113].)

Very high-speed networking

NSA's very high-speed networking program provides a high performance research network infrastructure characterized by multigigabit per second trunking speeds and will ultimately support sustained multigigabits per second data flows. In FY 2000, NSA will build on earlier experimentation with optical networking and networking protocols to:

- Delineate alternate network management solutions to reduce protocol layers
- Demonstrate optical multicasting by completing some of the agency's optical crossbar network and connecting it to ATDnet MONET equipment to create an all-optical internetwork
- Extend its transparent optical network to North Carolina in collaboration with MCNC and universities in the Raleigh-Durham area and connect to North Carolina's advanced education network and Internet2 to further investigate long distance issues
- Complete functionality/viability experiments begun in FY 1999

Extending earlier work in signaling, routing, addressing, and multicasting to meet NSA's addressing and bandwidth needs, in FY 2001 the agency will conduct research in congestion control and multidomain network management. NSA-supported researchers will examine a peer relationship among network management centers that exchange information in a controlled manner to enable end-to-end monitoring and fault isolation of network connections, helping address an environment in which there is little sharing of network management information among vendors, since such information can be used for competitive advantage.

DOE R&D

DOE's R&D addresses the need for a grid architecture that integrates networking, middleware, and applications to support wide area, data-intensive, and collaborative computing to link users with DOE's experimental and computation facilities. DOE has funded 12 projects to develop advanced network architectures and components, advanced middleware services, and advanced network monitoring tools and services with a goal of providing the performance guarantees required by applications.

DOE's implementation of Internet Protocol version 6 (IPv6)

One of the biggest long-term growth and scalability problems for the Internet is the lack of sufficient address space to globally address all systems with the current Internet Protocol version 4 (IPv4). A new version, IPv6, has been adopted as the next-generation network layer protocol standard. In August 1999, DOE's ESnet requested and was assigned the first production IPv6 addressing prefix by the American Registry for Internet Numbers (ARIN) and is using it to provide IPv6 services to ESnet users.

ESnet, which provides high-speed connectivity to thousands of scientific researchers at more than 30 DOE sites, has established a production IPv6 network initiative called the 6REN to encourage research and education networks worldwide to provide early production native IPv6 service. Because IPv6 uses 128-bit addresses, it offers a theoretical maximum of about 256 trillion, trillion, trillion addresses. This should allow sufficient addressing scalability to keep up with the current and future growth of the Internet, allowing universal accessibility. Other features designed into IPv6 include built-in security, dynamic automatic configuration, multicast, mobility, QoS, and an ability to allow routing systems to operate more efficiently. Full implementation of IPv6 is needed to effectively deploy large numbers of future wireless devices.

EMERGE

In FY 2000, DOE funded EMERGE, the ESnet/Midwest Research and Engineering Network (MREN) Regional Grid Experimental Testbed. EMERGE's goals are to place routers, Grid middleware, and applications test suites at DOE labs to make them differentiated service (DiffServ)- or QoS-enabled, explore interoperability of ESnet (an IP-over-ATM network) and Abilene (IP-over-SONET), and extend these QoS experiments to Europe and Asia via the NSF-funded STAR TAP.

EMERGE will facilitate advanced data flows not adequately addressed by today's "best efforts only" network. The testbed is driven by DOE computational science applications—particularly climate, combustion, and high-energy physics—that may require guaranteed bandwidth and low-latency networking. Tool development focuses on remote instrument control, data mining and extremely large data sets, collaborative environments, tele-immersion, ultra-high resolution rendered imagery, and unicast/multicast digital video.

Anticipated FY 2000 accomplishments include:

- 100 Mbps DiffServ networking available to science labs through EMERGE routers
- A Grid services package (GSP), a common suite of middleware services, deployed for use with DOE laboratory and university applications nationwide
- Deployment of a common test suite of tele-immersion applications

- Demonstration of DOE applications over the EMERGE testbed and ESnet
- Published quantitative measurements that document improvements when DiffServ is enabled
- Plans for testing alternative QoS strategies

Led by the University of Illinois at Chicago's EVL, in cooperation with ANL and ESnet, EMERGE is establishing a testbed with the following other sites:

Application sites:

- University of Wisconsin-Madison (Engine Research Center)
- University of Illinois at Urbana-Champaign (Center for Simulation of Advanced Rockets)
- University of Chicago (Center for Astrophysical Thermonuclear Flashes)

Tool development sites:

- University of Wisconsin-Madison (computer science: high-energy physics data grid)
- University of Illinois at Urbana-Champaign (NCSA: GSP)
- University of Illinois at Chicago (EVL: application-level tele-immersion network performance test suites)
- Northwestern University (International Center for Advanced Internet Research: video serving and Internet2 DiffServ interoperability over MREN)

Follow-on efforts with NIST will involve testing multiprotocol label switching (MPLS) to manage DiffServ and other flows, extending the GSP and incorporating visualization into monitoring tools, adding haptic and rendering flows to the tele-immersion network performance tests, creating a test suite for multiresolution compressed digital video, continuing interoperability testing and tuning with ESnet and Abilene, and increasing international cooperation. In the latter area, CERN, Holland, Russia, and Singapore have expressed interest in participating in DiffServ and MPLS tests.

Digital Collaboration Services (DCS)

ESnet is providing the DOE research community with the next generation of videoconferencing capabilities. Digital Collaboration Services (DCS) makes the audio, video, and data interactivity of a dedicated videoconference facility available on desktop computers, taking another step toward the virtual laboratory where scientists at distant DOE sites can collaborate in real time. DCS 2.0 is based on communications and applications standards adopted by the International Telecommunications Union (ITU) that allow interoperability with other networks and users.

DOE networking tools

In coordination with DARPA, DOE has developed Netlogger, a tool that provides real-time analysis for end-to-end performance monitoring. An easy-to-read graphical user interface (GUI) interface indicates bottlenecks in the network. DOE has also developed PingER, a network monitoring tool that

provides ping end-to-end reporting on Internet link performance. PingER measures response, jitter, loss, and reachability; provides national and international coverage of 536 remote nodes at 381 sites in 55 countries on six continents; and correlates with network performance monitoring tools. PingER primarily supports the high-energy physics community.

Global Observation Information Network (GOIN)

The GOIN demonstration project in Hawaii in March 1999 linked U.S. global observation researchers with partners in Japan. Using links supplied by NASA, DoD, and NSF, NOAA's PMEL demonstrated the first NOAA applications over the NGI, including OceanShare, a collaborative environment for oceanographic research, and 3-D tools using VRML to demonstrate the evolution of El Niño, fisheries larval drift, and fur seal feeding trips. Working with these innovative applications over the NGI provides greater opportunities for collaborating and accelerating scientific inquiry. Additional GOIN demonstrations used distributed real-time data to produce visualizations of natural Earth phenomena—such as sea ice movement and ocean thermal gradients—and used satellite thermal data to identify agricultural productivity.

Load balancing

NOAA has experimented with geographically distributed mirroring and load balancing to alleviate peak loads without wasteful over-provisioning. This is especially useful at the agency's National Hurricane Center, which experiences peak NOAA traffic during a hurricane landfall along the populous eastern seaboard. These experiments demonstrated their usefulness during the past hurricane season, when the load on NOAA's Web services exceeded that of any other governmental body as measured by unique accesses per month tracked by Media Metrix. The hurricane information was made available at three geographically distributed access points while critical operational weather data flowed uninterrupted.

LSN applications R&D

LSN applications R&D focuses on applying beta software from commercial providers and the research community to mission problems such as NOAA's weather modeling and NIH's medical collaboratories. Near-term development and deployment include Internet tools, products, and services such as high performance parallel interface (HiPPi) standards, application-specific multicast and QoS, and Internet videoconferencing development.

The Federal NGI initiative, in coordination with other Federal agency networking research programs, is creating the technical and infrastructure foundation for a more powerful, flexible, secure, and intelligent Internet in the 21st century. Tightly coupled with LSN network R&D and infrastructure support, NGI is helping build partnerships among academia, industry, and government to keep the U.S. at the cutting edge of information and communications technologies and stimulate the introduction of new multimedia applications in the Nation's schools, businesses, and homes.

The NGI initiative has two thrusts:

Thrust 1. NGI will advance R&D and experimentation in next-generation networking technologies to add functionality and improve performance in:

- Reliability
- Security
- Robustness
- Differentiated services, including multicast and audio/video—also known as quality of service (QoS) and class of service (CoS)
- Network management, including allocation and sharing of bandwidth

Other areas of concern for Thrust 1 include:

- Middleware
- Developing and deploying a software platform using high-speed Linux at a wide range of end-user sites
- Protocol stack tuning to improve performance

Thrust 1 is supported by DARPA's SuperNet, an NGI testbed providing a 1,000-fold increase in end-to-end performance over 1997 capabilities, or approximately 1 Gbps for research end users.

Thrust 2. NGI will develop and demonstrate revolutionary applications in enabling technologies such as:

- Collaboration technologies
- Digital libraries
- Distributed computing
- Privacy and security
- Remote operation and simulation

And in disciplinary applications such as:

- Basic science
- Crisis management
- Education
- The environment
- Federal information services
- Health care
- Manufacturing

Thrust 2 is supported by an NGI testbed providing a 100-fold increase in end-to-end performance over 1997's Internet, or approximately 100 Mbps for research end users.

THRUST 1 ACCOMPLISHMENTS AND PLANS

NGI Thrust 1 programs focus on R&D and testbed deployment and demonstration of technologies to provide network growth engineering and to enable the reliable and secure provisioning, management, and end-to-end delivery of differentiated classes of service.

NSF

NSF has implemented a wide range of networking research programs addressing network growth engineering, QoS, and security. Network engineering research projects focus on scaling up the Internet in size and speed, developing performance measurement and middleware, and addressing flow-based or aggregate-based QoS and congestion control. Specific projects address:

- Congestion management
- Terabit routing with QoS
- Advance reservations, routing, and scheduling
- Scheduling algorithms for high-speed switches with QoS guarantees
- QoS-enabled middleware for global distributed computing

NSF programs address network security and assurance, end-to-end security, and policy functions such as pricing and cost recovery. Specific projects include secure and robust agent technology and secure multicast.

High performance network service providers (HPNSPs)

NSF has designated a category of commercial HPNSPs that offer advanced network services over broadband networks to university and Federal agency sites to provide the high performance services needed by the NGI. The Abilene network—the first network to receive this designation—coordinates closely with the JET to provide connectivity and meet other requirements of the NGI community. Additional network providers have expressed interest in providing HPNSP services.

DARPA's SuperNet

The goal of DARPA's SuperNet program is guaranteed ultra-high bandwidth on demand over the shared national networks infrastructure. SuperNet will demonstrate multi-Gbps transmission end-to-end by:

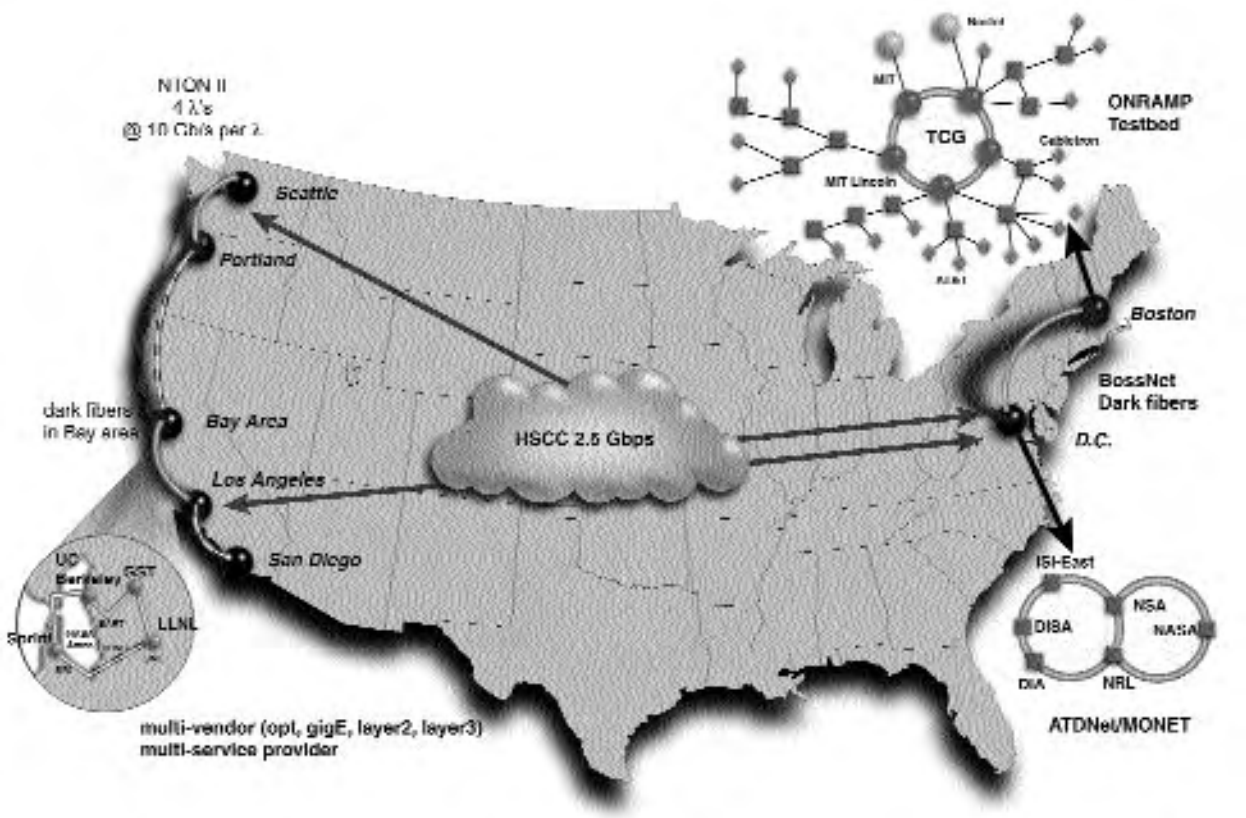
- Developing streamlined networking protocol stacks
- Stressing end-to-end architecture and performance
- Developing technologies for regional, metropolitan, and local area networks
- Implementing a dynamically controlled optical layer

At the SC99 conference, NGI researchers and industry partners set 2.4 Gbps world network speed and performance records, transmitting studio-quality high-definition television (HDTV) streams long distance over NGI testbeds to SC99.

NIST

Advanced encryption standard (AES)

NIST is developing AES, a royalty-free encryption standard to be used by government and industry to protect information over the next 30 to 50 years. The agency is testing candidate algorithms using multiple platforms and computing systems, comparing efficiency testing results and determining the validity of cryptanalysis, and validating conformance of COTS products.



The goal of DARPA's SuperNet program is guaranteed ultra-high bandwidth on demand over the shared national networks infrastructure.

Public key infrastructure (PKI)

NIST's PKI program will ensure the development of commercially available PKI products and services that are interoperable and sufficiently secure to meet the needs of Government agencies and the general public. NIST is helping users and suppliers establish PKI standards and specifications, interoperability, correctness, and quality, and is publishing its analysis of PKI component security requirements and guidance and developing pilots for automated key recovery systems and Web-based electronic certification.

Internet quality of service (QoS)

NIST is expediting commercial deployment of standardized Internet QoS technologies, devising tools to aid developers of adaptive Internet applications, developing techniques and tools to test distributed multiparty QoS routing and signaling protocols, and evaluating proposed algorithms and protocols for scalable QoS routing and signaling. For example, NIST has deployed NISTnet, a tool that emulates controllable Internet performance dynamics. NISTnet was demonstrated at SC99 (page 102).

Hybrid-fiber coax (HFC) access

Hybrid-fiber coax (HFC) is a means for connecting home computing systems to commercial communications networks. NIST is evaluating protocols for HFC media access control (MAC) and is studying end-to-end performance to improve ATM and TCP/IP traffic over HFC networks. NIST has incorporated HFC network protocols into its ATM network simulator and has published

results on contention resolution algorithms, bandwidth allocation, and priority schemes for end-to-end ATM and TCP/IP traffic over HFC networks.

Dense wave division multiplexing (DWDM)

NIST is accelerating the development of dynamically reconfigurable DWDM networks and evaluating proposed algorithms for wavelength assignment and routing in WDM networks. WDM dramatically increases bandwidth on existing fiber.

Agile networking infrastructures

NIST is developing agile networking technologies to enable programmable and reconfigurable communication infrastructures, middleware for adaptive, reconfigurable distributed systems, and measurement techniques to enable resource control in active networks. NIST is also evaluating measurement and standardization requirements for networked pervasive computing, which allows mobile wireless network access to the Internet.

THRUST 2
ACCOMPLISHMENTS AND PLANS

NGI Thrust 2 applications span a wide range of societal and technical areas, and the 100x testbed that supports applications development provides high performance, reliable service for a rigorous applications testing environment. The 100x testbed network:

- Develops and implements advanced technologies including QoS, multicast, security, network management, and protocol performance management
- Emphasizes end-to-end performance rather than point-to-point service
- Emphasizes transparency
- Includes campus environments (LANs) and user end systems
- Provides coordinated management of the network resources
- Provides reliable service through a robust architecture (NGI network backbone, regional and GigaPop structure, and Next Generation Internet Exchanges [NGIXs])

The network providers implement these capabilities, which are in turn coordinated by the JET. In FY 1999, the JET focused on interconnection points among the JETnets (vBNS, ESnet, NREN, DREN, NISN, and Abilene) at the NGIXs. NGIXs were implemented at NASA Ames Research Center (NGIX-West), Chicago (NGIX-Midwest), and the University of Maryland (NGIX-East), and peering arrangements are currently being developed at these exchanges. Further cooperation with the university community was provided by joint meetings among the Federal agencies, GigaPop operators, university campus representatives, and Internet2. This coordination has focused the university and Federal communities on NGI network performance measurements and upgrading campus networking capabilities and services.

In FY 2000 the JETnets implemented multicast in the NGI backbone. University campus networks need to implement and tune multicast tools in their networks to enable end users to effectively use multicast services, which are particularly effective in broadcasting lectures and meetings because they provide interaction among remote environments in real time.

IP QoS is being deployed in a collaborative testbed. ESnet, NREN, and vBNS participate in the Qbone project that is implementing interdomain

DiffServ. vBNS has offered reserved bandwidth service using RSVP/ATM. All JETnets are part of the 6Bone project that is developing IPv6, which provides protocols for a significantly increased address space and expedites network services such as QoS. vBNS is testing a native IPv6 service over separate routers.

Performance measurement and improvement

Measuring the performance of the Internet is critical in identifying problems such as bottlenecks and resolving them to provide improved network performance. Active measurement capabilities of the JETnets include the advanced measurement platform (AMP) and survey to provide throughput, round trip time (RTT), topology, one-way delay, delay variation, and packet loss data. Passive network performance measurements use Coral/OCXmon, which is non-invasive and provides detailed characteristics of individual flows. Additional measurement and visualization tools are under development. The JET is identifying large data flows and high-bandwidth users on the NGI networks to offer them assistance in improving the performance of their applications. The Team will provide fine-tuned laptops to measure end-to-end performance from specific subnets to assist local network operators in identifying bottlenecks and other network problems. The JET is fostering close cooperation among commercial vendors and the NGI agencies to develop Linux software on a platform to tune the protocol stack to provide improved, flexible Internet performance for a range of high performance applications.

NSF applications

Hundreds of applications are being developed by NSF-funded university researchers. While many require the high-bandwidth capabilities of the NGI, others require the advanced services being developed by the NGI initiative, such as QoS, security, multicast, collaboratory tools, and visualization software. Examples of the revolutionary applications funded by NSF include:

Ecology

Species analyst

The species analyst is a set of tools for simultaneous access to multiple biological collection databases that can be used to predict species distribution as a function of global climate change, species invasions, and species disturbances by humans. Led by the University of Kansas, this project is also being conducted at the University of California-Berkeley, Harvard University, the University of Michigan, the University of Nebraska, SDSC, and the University of Mexico.

Education and teaching

Collaborative development of 3-D life science educational resource

Yale University's School of Medicine and NLM are creating 3-D educational materials using NLM's life science image databases. To assess levels of acceptable video compression, Yale and NLM experimented with digital video encoding of the video records of patients suffering from neurological motion disorders such as Parkinson's disease.

Two data mining and visualization projects

The University of Illinois at Chicago's CAVERN project is conducting research in tele-immersion and intelligent data mining for use in computational science, collaborative distance-independent education, and the Every Citizen Interface to the National Information Infrastructure.

The Web TerraFly project at Florida International University allows users to fly over and manipulate data retrieved from a high performance semantic multimedia spatial database that includes satellite and aerial photography data.

Megaconference

With Ohio State University as the lead institution, this pilot project has established an ongoing videoconference collaboration linking 65 network engineers and researchers on three continents to discuss advanced networking technology.

Wide area interactive teaching

With Oregon State University in the lead, this application has provided the first regular scheduled Internet2 video class in graduate-level plant pathology among Oregon State, Kansas State University, and the University of Nebraska. The application uses enough bandwidth to enable high-quality video at these institutions.

Humanities, arts, and archaeology

Center for Electronic Reconstruction of Historical and Archaeological Sites (CERHAS)

CERHAS—a collaborative research project of the University of Cincinnati, Cleveland State University, and Wright State University—is building software for students and professors in the arts and humanities. CERHAS researchers are exploring real-time, networked motion capture, avatars, artificial personalities, and secure personal operating systems to enable collaborative virtual reality. Target applications will be distributed to users at participating institutions via Web-DVD.

Variations digital music library

Variations provides access to more than 5,000 titles of near CD-quality digital audio to computer users in the Cook Music Library of the Simon Music Library and Recital Center at the University of Indiana-Bloomington.

Manufacturing

Scaling Internet connections to support research applications

This NSF Science and Technology Center project focuses on automated machining direct from CAD sketches. Brown University, CalTech, Cornell University, the University of North Carolina, and the University of Utah are the collaborating institutions. Using the vBNS, Brown has been able to maintain its software on Utah systems, speeding research experiments and facilitating the display of 3-D graphics at remote locations to the point where applications appear to be running locally.

Multimedia

Large-scale video network prototype

Supported by the Southeastern Universities Research Association (SURA) and BBN/GTE Internetworking, this advanced networking project, which involves researchers at the University of North Carolina-Chapel Hill, the University of Tennessee-Knoxville, the Georgia Institute of Technology, and NYSERNet, is investigating and deploying model architectures for video services to support distributed video in higher education.

Remote science and networking

Remote observing

A new remote observing application will provide authorized scientists anywhere on an NGI network with access to the more than $1 billion in research instruments at the Mauna Kea Observatories in Hawaii, eliminating the need to travel and acclimate to the remote 15,000-foot-high facility. Institutions involved in this research include the University of Hawaii, the University of California, CalTech, the University of Washington, and others.

Telemedicine

Distance-independent telemedical diagnosis

Eastern Carolina University along with the Ohio Supercomputer Center, the Ohio State University, the Northeast Parallel Architecture Center, and Syracuse University are funded by NSF and NLM to study biomedical applications that require NGI capabilities such as:

The Mauna Kea Observatories sit at an altitude of 4,200 meters, near the summit of Mauna Kea on the Island of Hawaii. Currently, nine telescopes plus the Hawaii Antenna of the Very Long Baseline Array are in full operation. The cloud-free, dry, and stable atmosphere on Mauna Kea permits more detailed astronomical studies than are possible elsewhere. NGI technologies will enable scientists at remote locations to access the instruments at this unique site.

- Novel rendering techniques for extremely large volumetric data sets and methods to guarantee network performance
- Evaluating IP video over the network to determine the technologies needed to meet diagnostic quality needs and standards
- An investigation into the effect of real-time encryption of IP video to ensure patient confidentiality
- Qualitative analysis of how the NGI can be used to transmit cine-angiograms rapidly without degrading the images

Psychological services for the hearing-impaired

This University of Rochester research project uses teleconferencing technology to bring deaf consumers with mental health care needs into contact with sign language-fluent practitioners. Directly communicating with sign-fluent psychologists and social workers avoids problems caused by differences between American Sign Language and English as well as the awkwardness of using a sign language interpreter in a sensitive mental health setting.

Veterinary science
Virtual rounds

Virtual rounds is an application being developed by the colleges of veterinary medicine at the University of Tennessee-Knoxville, Auburn University, the University of Georgia-Athens, and North Carolina State University to share animal clinical cases via live videoconferencing. Sessions will be captured and stored on a video server to be available on demand, creating a potential for digital video libraries to support veterinary medical instruction.

Weather, atmospheric research
Advanced Regional Prediction System (ARPS)

ARPS is a fully automated, functionally complete numerical forecast environment that can predict weather on scales ranging from continents to cities. It was the first model of its kind in the world designed for massively parallel computers and is uniquely suited to accommodate observations from Doppler radars. NSF, NOAA, the Federal Aviation Administration (FAA), Allied Signal (now part of Honeywell), and American Airlines support this University of Oklahoma research.

Space Physics and Aeronomy Research Collaboratory

Space physics researchers at some 380 sites around the world study phenomena such as magnetic storms on the sun that can interfere with radio and television reception, disrupt electrical power transmission, and threaten orbiting spacecraft. They do so by controlling and gathering data from more than a dozen instruments located around—and above—the globe, directly accessing advanced supercomputer models of upper atmospheric phenomena and state-of-the-art communication tools that include "chat rooms" and a shared whiteboard. All research is recorded for replay, annotation, or asynchronous collaboration with colleagues.

DARPA applications

DARPA's NGI applications include:

- CSU-CHILL radar for remote sensing and meteorological analysis. Doppler radar data are collected at distributed sites, processed, and made available on the network at speeds of 240 MBps to 2.88 GBps. The data are used to determine meteorological characteristics such as rain, hail, ice crystals, and turbulence.
- Matisse computer microvision workstations optically monitor micro-electromechanical systems (MEMS) device response over varying focal planes. A scientific microscope provides ultra-high resolution motor control and stroboscopic light-emitting diode illumination, and a closed coupling device camera system provides a megapixel camera and frame grabber. A typical data set is 10 GB.
- Digital Earth. Part of an interagency initiative, this application is an open, distributed, scalable, multiresolution 3-D representation of the Earth into which massive quantities of geo-referenced information can be embedded. Digital Earth uses VRML and a standard browser and plug-in to navigate the 3-D model.

NIH applications

NLM awards

NLM initiated a three-phase NGI research program to develop innovative medical projects to demonstrate application and use of NGI capabilities, including:

- Quality of service
- Medical data privacy and security
- Nomadic computing
- Network management
- Infrastructure technology for scientific collaboration

NLM has awarded 15 Phase 2 awards to develop local telemedicine testbed applications. Those that prove effective will be implemented for end users under Phase 3 awards. Phase 2 awards under way or to be made in FY 2000 include:

Applications layer security solution for stationary/ nomadic environments

This project will evaluate security techniques within an open security architecture. The proposed solution is based on security shared among collaborating parties, nomadic computing, and the privacy of medical information.

Biomedical tele-immersion

By combining teleconferencing, telepresence, and virtual reality, tele-immersion enables teachers and students to interact with 3-D models, especially in surgical education. NGI data privacy and security guarantees allow tele-immersive environments to be derived from models of patient data.

Networked 3-D virtual human anatomy

The goal is to build an online virtual human cadaver, based on the Visible Human data set, that students can explore with a variety of tools.

Patient-centric tools for regional collaborative cancer care

This project will investigate the application of collaborative tools in the Seattle area Cancer Care Alliance (CCA) to:

- Enhance the CCA partners' existing clinical care programs to be highly collaborative patient-centered interdisciplinary efforts
- Allow a fully integrated team approach to cancer, including state-of-the-art diagnosis, treatment, and management through collaboration by distributed clinicians and researchers
- Accelerate the dissemination and application of new knowledge for cancer diagnosis and treatment

Radiation oncology treatment planning/care delivery application

This project will implement and evaluate NGI capabilities for radiation oncology treatment planning and care delivery, providing diagnostic support, treatment planning, and remote verification of proper equipment operation from the Comprehensive Cancer Center to a Johns Hopkins University treatment facility.

Rural health science education

This project is developing a plan to evaluate computer and interactive compressed video technologies to support rural health science education. Such an application will enable delivery of interactive educational programming, such as continuing medical education, clinical information systems, library services, and consultation. Beneficiaries include students and health care professionals.

DOE applications

Collaboratories

Funded in FY 1999, DOE's work in NGI applications was largely completed in FY 2000. DOE funded five collaboratories with multiple sites including

universities and national laboratories. DOE supports two testbeds to demonstrate advanced services to university sites and improve access and capabilities for university applications development researchers.

Distributed X-ray crystallography

Using tools being developed by the ANL Globus group and the DOE 2000 Common Component Architecture Forum, this project is building NGI network-based instrumentation including high-speed data collection, reduction, storage, and visualization, and real-time instrument control for the acquisition of macromolecular x-ray crystallographic data from the LBNL Advanced Light Source.

Integrated grid architecture and the Earth system grid

DOE is creating an integrated grid architecture built on a fabric of networks, archives, computers, display devices, and other technologies and associated local services. Grid services include protocols, authentication policy, resource management, instrumentation, and discovery. A remote data toolkit, remote computational toolkit, visualization toolkit, asynchronous collaboration, and remote sensors will help address the needs of end-user applications. One such application is the Earth system grid for climate modeling, projections, and impact assessments, in which more than 100 universities, laboratories, and centers participate.

Combustion Corridor

Combustion of fossil fuels accounts for approximately 85 percent of the energy expended in the U.S., and modeling this process is critical to increasing efficiency and reducing pollution. Combustion modeling codes can easily generate terabytes of data that must be analyzed by researchers dispersed across the U.S. The DOE Combustion Corridor application uses grid storage API, global naming services, the Globus resource reservation system, networks, disk caches, and PC clusters to interactively and collaboratively visualize these data.

Such visualizations typically represent complex, 3-D scientific problems varying over time, such as how two gases mix in a turbulent environment. To visualize these models, researchers previously required access to very powerful computing systems, and moving their large files onto local workstations was either impossible or impractical. In image-based rendering-assisted volume rendering (IBRAVR), developed at LBNL, a large data set is now partially pre-rendered on a large computational engine close to the data, and the final image rendering is performed on a workstation.

Corridor One

DOE's Corridor One—visualization at a distance—uses the grid to integrate data servers, analysis and manipulation engines, visualization backend servers, visualization clients, display device interfaces, and advanced networking services.

NASA applications

Biomedical image collaboratory

This experiment relies on a Visible Human Viewer developed on OpenStep, an Apple API. This application can show sections of a human body so a researcher can recognize anatomical objects. Anatomical terms can be attached using NLM's Unified Medical Language System (UMLS).

Collaborative electron microscopy of quarantined samples

This NGI experiment will remotely control and observe microscope imaging of quarantined samples returned from Mars at NASA's Ames Research Center (ARC) from remote sites at ARC, ORNL, and Oregon State University.

Digital Earth/Mars/Sky

Digital Earth (DE) is an interagency initiative to define and prototype a framework to make all U.S. geo-referenced data available via Web-style point-and-click user interfaces over high performance networks. NASA's DE research is a collaborative effort with the Federal Geographic Data Committee and the Open Geographic Information System Consortium, an organization of business, academic, and government officials concerned with interoperability of geographic data. This work supports NASA's goal of human telepresence throughout the Earth and solar systems, which requires high performance remote access and visualization of large Earth and space data sets.

Distributed video observation of shuttle processing and launch activities

This project will demonstrate the delivery of broadcast-quality video streams over a high performance IP network. Multiple video sources at the Kennedy Space Center will be transmitted to desktops at a variety of locations, and viewers will be able to choose among available sources. The primary goals are to use COTS technologies to provide greater video surveillance flexibility at reduced cost.

Virtual collaborative clinic (VCC)

NASA, in coordination with NLM, the Navajo Nation, Abilene, vBNS, and the CALREN-2 GigaPoP, is developing the virtual collaborative clinic (VCC) to demonstrate a high performance testbed that allows medical colleagues to simultaneously review medical images remotely in real time. This requires 30 to 50 Mbps and multicast technology.

NIST applications

NIST has deployed a pilot collaboratory for robotic arc welding and published preliminary results on usage. NIST researchers are deploying and assessing a manufacturing collaboratory testbed, evaluating collaboration processes and an industrial pilot collaboratory, and developing quantitative evaluation methods for collaboration technologies.

NIST has developed interfaces for virtual manufacturing applications, a testbed to demonstrate real-time, multiuser, interactive simulation of manufacturing equipment control, and extensions to VRML for device behaviors. NIST is building a remote interface to a weld cell with real-time display of a data-driven VRML weld controller model and synchronized real-time video with a VRML model.

In FY 2001 NIST plans to:

- Link the full immersion CAVE environment at the University of Illinois to a welding robot at NIST via the NGI
- Establish remote control of a welding robot in the virtual world

SII Program FY 2000-2001

The Scalable Information Infrastructure (SII) component of LSN R&D responds to the PITAC recommendation that the Federal government establish a high-priority focus on research seeking fundamental advances in the science and engineering of networking technologies that will enable the Internet to grow or "scale up" to meet requirements for ever-increasing numbers of users and types of devices. Begun in FY 2000, the SII R&D program is still being developed to address demands for 21st century technologies such as deeply networked systems, "anytime, anywhere" connectivity, and network modeling and simulation.

- *Deeply networked systems.* Researchers will develop software and network architectures enabling large numbers of diversified devices such as low-cost wireless sensors to provide real-time information from distributed sensors. Such devices could provide, for example, real-time information on air and water pollution and improve our ability to monitor the environment and respond to man-made disasters; "guardian angels" that monitor the health and safety of fire fighters, law enforcement officials, soldiers, and home health care patients; and crisis management centers that use sensors carried by response teams and airplanes to improve responses to forest fires, floods, and hurricanes. Research will focus on naming, addressing, and network configuration, and on developing less expensive network interfaces.
- *Anytime, anywhere connectivity.* Research will focus on wireless technology to provide all U.S. citizens with high-speed connections to the Internet, whenever and wherever they are needed. Wireless networks can extend services such as distance learning and telemedicine to remote rural areas in the U.S. and to markets in developing countries. This research will emphasize developing standards, improving bandwidths, and addressing wireless architectures to provide the wide geographic and mobile coverage needed for wireless Internet connectivity.
- *Network modeling and simulation.* The goals are to develop tools to model network behavior to test new network technologies before implementing them on testbeds and the full Internet, and to enable faster than real-time simulation to allow network operators to forecast network behavior, allowing them to intercede to prevent network congestion or collapse.

Federal agency SII programs

Federal agency SII activities will explore fundamental aspects of networking, including the ability to depict and troubleshoot very large systems through network modeling, simulation, and emulation; underlying network technologies such as optical and wireless; and middleware enabling large-scale systems, information management, and information and service survivability. DARPA- and NSF-supported researchers are developing the network architecture, middleware, and human interfaces for pervasive networking environments—systems capable of effectively and efficiently networking large numbers of very small distributed and embedded devices—including highly scalable and adaptable network protocols for evolving application requirements; QoS, such as for network control and

new starts

scheduling; and network security and privacy tools. Research addressing development of on-demand wideband access in scalable mobile networking will aim for multimegabit wireless access with seamless global roaming.

Agile networking infrastructures

Future networking communities will consist of highly fluid groups and locations accessing the Internet through a wide variety of information services, embedded devices, and facilities. In this environment, network infrastructures must be "agile," meaning able to respond quickly to variations in connectivity and bandwidth requirements and differences among access devices, and to dynamically reorganize end-to-end network services in response to changes in the composition of collaborating teams. The agile networking infrastructures effort will support the specification, standardization, measurement, and testing of protocols and software to support these capabilities. Program goals include the development of virtual overlay networks (VONs) that can be established and readjusted on demand in minutes by distributed mobile collaborative users.

Network research center

NSF and NIST plan to establish a high performance distributed network research center enabling researchers to study new protocols they are developing to improve Internet performance. The center will focus on discovering and managing the limiting features of networks, providing advanced performance features not now available on large-scale networks; distributed applications and their interaction with network protocols; middleware enabling efficient computer-aided software development; and tools and expertise to enhance the Internet knowledge base, particularly for minority-serving organizations.

Network security

NSF- and NIST-funded researchers will explore techniques, protocols, and procedures with the potential to enhance the overall security of packet-switched telecommunications networks, including the ability to support secure tele-collaboration activities.

Prototype access testbeds

Researchers will develop prototype testbeds to demonstrate very high performance campus networks that can serve as models for upgrading campus networks nationwide to take advantage of new backbone capabilities. The research will test potential delivery systems ranging from wireless to all optical and capabilities for interaction among wireless, optical, and wireline systems.

IT R&D Technology Demonstrations

Technology demonstrations provide an opportunity to make applications development visible to the general public while driving progress by establishing specific deadlines for operational capabilities and calling on resources—such as OC-48 network connectivity—that might not otherwise be available. They enable developers and scientists to see what kinds of tools, services, and innovations may be useful in their applications discipline, encourage new interactions and collaborations among researchers, and allow the public to participate in IT R&D research advances.

Federal agency IT R&D demonstrations are held several times a year to reach different audiences. In FY 1999-2000, an IT R&D Expo was held on Capitol Hill to foster greater understanding among legislators and their staffs of the progress on applications development. IT applications demonstrations, including NGI applications, were showcased at SC99, the national high performance networking and computing conference, to expand awareness in the science and commercial communities of Federally supported R&D efforts in high performance computing and networking. And IT R&D agencies were invited to conduct applications demonstrations in Asheville, North Carolina, in December 1999. Select applications demonstrations from these three events are highlighted in this section.

SC99 demonstrations

At the November 13-19, 1999, SC99 conference, NGI researchers and industry partners made news when they set 2.4 Gbps world network speed and performance records, transmitting studio-quality high-definition television (HDTV) streams long distance over NGI testbeds to the SC99 exhibition hall in Portland, Oregon. SC99, one of the largest gatherings devoted to advanced computing, drew more than 5,000 researchers and industry representatives from around the world.

For the second year, the NCO hosted a group of exhibits designed to broaden awareness in the academic and IT industry communities of the far-reaching results of Federally sponsored applications research, including the NGI. Six Federal agency members of the LSNCG—DARPA, DOE, NASA, NIH, NOAA, and NSF—demonstrated 10 cutting-edge applications and network technologies developed with Federal funding. Many other demonstrations of Federally funded R&D, including NGI, were held throughout the research area on the convention floor. Additional NGI applications are being developed by universities and laboratories.

Agency representatives, university researchers, and NCO staff prepare for the opening of SC99 in Portland, Oregon.

SC99 demos included:

World network land speed record

DARPA's SuperNet (part of NGI, page 87) was a critical element in setting the 2.4 Gbps packet-over-Sonet world land speed record. The network infrastructure for the long-distance, high-speed HDTV demonstration linked DARPA's National Transparent Optical Network (NTON-II), the University of Washington (UW)-led Pacific/Northwest GigaPop (P/NWGP), and Nortel Networks. Together, the networks delivered unprecedented levels of standard Internet capacity from the Microsoft Corporation and UW campuses, through the P/NWGP in Seattle, and on to the exhibition hall. There, Microsoft, the National Computational Science Alliance, Sony (in support of the ResearchTV consortium), and UW concurrently demonstrated two real-time gigabit applications, setting a record of more than 2 Gbps in aggregate throughput—by a wide margin the fastest real-time applications yet run over a WAN.

The demonstration was the equivalent of simultaneously transmitting the entire channel lineup of a 150-channel cable TV system, or 50 channels of broadcast quality HDTV, or five feature movies, or interactions among a large number of shared virtual realities.

Sponsors and participants

DARPA
Microsoft Corporation
National Computational Science Alliance
Nortel Networks
Pacific/Northwest GigaPop
ResearchTV
Sony
University of Washington

Virtual Laboratory (VLAB)

NASA's VLAB is a virtual reality environment enabling researchers working at desktop computers anywhere in the U.S. to monitor and actively participate in real-time experiments and training simulations at the Ames Research Center Flight Simulation Laboratories. The emphasis is on the perspective of the user, who can define the specific data and display configuration for an experiment, and multiple "players" at different sites can each customize their displays for the same simulation. For example, the VLAB interface lets engineers at the Johnson Space

Center, Boeing, Rockwell, and Lockheed Martin access tests at Ames' Vertical Motion Simulator, where researchers conduct astronaut training and study the physics and engineering requirements of rockets and space vehicles such as the Space Shuttle. New capabilities being developed for the VLAB system will allow researchers to create math models, displays, and control systems, validate models for higher-quality experiments, and do virtual prototyping for cockpit design and lab data system layouts. NASA researchers have developed a prototype Mars VLAB client and have deployed a full VLAB client at Johnson Space Center to support shuttle entry simulation.

VLAB concepts are broadly applicable to remote virtual control rooms such as those associated with wind tunnels, flight test facilities, and multiple interoperable laboratories. The NGI's high bandwidth and multicast technologies make it possible to provide the digital audio, MPEG-2, audio/video-conferencing, whiteboarding, and client/server support, as well as the real-time interactivity, needed by VLAB. The QoS afforded by the NGI ensures that visualization streams reach their destinations on time when networks are congested.

Participants

NASA Ames Flight Simulation Laboratory
NASA Johnson Space Center
NASA Research and Education Network (NREN)
Apple Computer, Inc.
Lockheed Martin Corporation
Rockwell International Corporation
The Boeing Company

Sponsor

NASA

SuperNet

DARPA's SuperNet is the set of interconnected NGI testbeds dedicated to prototyping the engineering, hardware, software, and connectivity of future networks capable of 1,000 times current Internet transmission speeds. Over the first two years of the NGI program, Federal, academic, and industry research partners developed and put in place the pioneering components necessary for a wholly new network infrastructure with capabilities far beyond those of today's Internet. At SC99, these partnerships reached a major developmental milestone, demonstrating for the first time full coast-to-coast interconnectivity across the network of SuperNet testbeds. The following components of SuperNet will be operational in FY 2000:

- NTON-II, DARPA's high-speed optical network on the West Coast, with four wavelengths each with 10 Gbps capacity
- High Speed Connectivity Consortium (HSCC), connecting Los Angeles and Washington, D.C., sites at 2.5 Gbps end to end
- ONRAMP, a Boston-area testbed operating at OC-48+ and fielding advanced metropolitan area and regional access technologies
- BOSSNET, a highly experimental WDM network developed by DARPA enabling physical layer networking and communications experiments over dark fiber, with connectivity from the Boston area to Washington, D.C., at OC-48+ speeds

Demonstrating DARPA's SuperNet at SC99. More than 40 research institutions use SuperNet in applications areas such as HDTV, reliable multicast, domain name system (DNS) security, wide area gigabit Ethernet (GbE), and active networks.

- Advanced Technology Demonstration Network/Multiwavelength Optical Networking (ATDNet/MONET), a 20 Gbps dynamically reconfigurable double ring network
- Collaborative Advanced Interagency Research Network (CAIRN), a coast-to-coast testbed for networking research with an infrastructure that is heterogeneous in hardware, interconnections, and speed

More than 40 research institutions use SuperNet components for network and applications research in areas such as HDTV, border gateway multicast protocol, reliable multicast, domain name system (DNS) security, DWDM switching, wide area gigabit ethernet (GbE), and active networks. Robust, scalable versions of these technologies will be needed to build out the NGI prototype networks into a full-scale, very high-speed infrastructure for research, commerce, and communication.

Participants

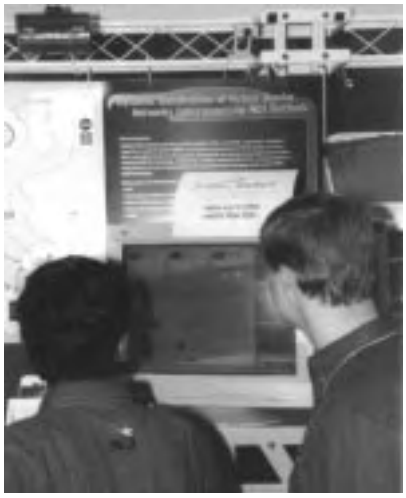
Carnegie Mellon University
Corporation for National Research Initiatives
Defense Information Systems Agency
Defense Intelligence Agency
GST
Laboratory for Telecommunication Sciences
Lawrence Livermore National Laboratory
MIT Lincoln Laboratory
NASA Goddard Space Flight Center
Naval Research Laboratory
Nortel
Qwest
USC Information Sciences Institute

Sponsor

DARPA

Dynamic construction of virtual overlay networks interconnecting NGI testbeds

The nationwide SuperNet interconnectivity demonstrated at SC99 was made possible by a new technology called multiprotocol label switching (MPLS) and a publicly available prototype MPLS research platform created by NIST called NISTSwitch. MPLS dynamically constructs virtual overlay networks that enable



SuperNet interconnectivity at SC99 was enabled by a new technology known as multiprotocol label switching (MPLS). MPLS is publicly available on a prototype platform known as NISTSwitch, a product of NIST-supported research. MPLS will enable NGI researchers to expand experiments with virtual overlay networks.

QoS-controlled data transmission from network to network. Virtual overlay networks are possible only with wide-bandwidth network infrastructures that permit WDM. In systems with WDM such as the SuperNet testbeds, the NISTSwitch platform operates as a kind of rapid-transit service hub, setting common communications and routing protocols for all carriers on all routes and working to maintain performance standards across the system as a whole.

MPLS will enable NGI researchers to expand their experimentation with virtual overlay networks for QoS, multicast, and security assurance on NGI testbeds. MPLS also holds promise in signaling and routing technology as a means of optimizing traffic engineering and constraint-based path selection in NGI networks.

Sponsors and participants

NIST
University of Southern California/Information Sciences Institute (USC/ISI)
NASA Goddard Space Flight Center
NASA Ames Research Center
DARPA

ImmersaDesk: An immersive virtual environment for oceanography and meteorology

The ImmersaDesk is a projection platform that "immerses" the user in a virtual reality environment using stereo glasses. Wearing the special glasses, a user can look into the ImmersaDesk's 4'x5' angled screen to view a computer-generated 3-D image simulated with near-perfect accuracy from complex statistical data. At the SC99 demo, visitors witnessed a tsunami approaching and crashing onto a coastal Japanese city, with realistic rendering of the wave, sea floor bathymetry, land topography, and the city itself. Using the device's interactive, animated, 3-D stereographic visualization capabilities, NOAA scientists are modeling other environmental influences such as the activity of hydrothermal vents, the effects of oceanographic forces on fisheries, and the characteristics of regional weather systems. The 3-D visualization of complex data sets reveals subtleties that scientists cannot see in 2-D graphics. For example, a 2-D graphic of an Alaskan fur seal's hunt for food shows only the seal's linear route through the water. A 3-D animation of the same data reveals that the seal makes numerous sharp upward and downward movements through the water along the way.

Demonstrating the ImmersaDesk at SC99. The ImmersaDesk is a large-scale projection platform that "immerses" the user in a virtual reality environment by means of stereo eyeglasses. NOAA scientists are using the ImmersaDesk in 3-D visualizations of complex data sets. The speed and bandwidth of NGI will enable climate and oceanographic researchers to work collaboratively in such 3-D virtual environments. NOAA, a non-NGI agency, has partnered with NGI agencies on several advanced networking applications from the beginning of the program.

NGI networks interconnecting heterogeneous computing resources and large data archives via a shared interface allow geographically distributed climate and oceanographic researchers to work collaboratively in such 3-D virtual environments.

Participants — *NOAA's Pacific Marine Environmental Laboratory (PMEL)*
Old Dominion University

Sponsor — *NOAA's HPCC program*

OceanShare

When a major oceanic oil spill occurred—before the advent of high performance computational and networking technologies—scientists routinely spent days and perhaps weeks gathering and analyzing the information needed to guide damage control and mitigation efforts. NOAA's OceanShare is a powerful new collaborative tool that enables researchers at NOAA laboratories to quickly access and work with large-scale oceanographic and meteorological data sets from multiple geographically distributed archives. This means that scientific analyses requiring heterogeneous information—weather conditions, ocean temperatures and currents, topographical data, and marine ecosystem data—from a variety of sources can be rapidly integrated and analyzed collaboratively by researchers, no matter where they are located.

OceanShare's data-intensive and bandwidth-intensive capabilities for collaborative research require the speed and QoS characteristics of the NGI, guaranteeing rapid response rates, a sophisticated shared interface, and uninterrupted flows of large-scale data. OceanShare combines interactive Java graphics, Java remote method invocation/common object request broker architecture (CORBA) network connections, and NCSA's object-oriented Habanero developer's framework to create collaborative networked access to distributed data sets. The new research tool enables NOAA/PMEL's Fisheries Oceanography program, in collaboration with NOAA's Alaska Fisheries Science Center and the University of Alaska, to study fish migrations, environmental conditions affecting fisheries, and related oceanographic influences on fish populations. NOAA's Hazardous Materials Response and Assessment Division

(HAZMAT) will use OceanShare to assess response needs when hazardous materials are released into the environment. The new tool will also provide networked access to climate data archived at NOAA's PMEL, Atlantic Ocean Meteorology Laboratory, and National Ocean Data Center, as well as at the University of Hawaii Sea Level Center and other research sites.

Participant — *NOAA's PMEL*

Sponsors — *NOAA's HPCC program*
NOAA's Environmental Services Data and Information Management (ESDIM) program

A virtual tour of 3-D oceanographic data sets using Virtual Reality Modeling Language (VRML)

From the TAO (Tropical Atmosphere-Ocean) Array—a group of some 70 moored buoys spanning the equatorial Pacific—NOAA scientists gather oceanographic and surface meteorological variables critical to improved detection, understanding, and prediction of the significant climate variations such as El Niño and La Niña that originate in the tropics. Today, using VRML software, scientists can turn the large-scale TAO data collected around the clock into 3-D objects that precisely mimic their real-world counterparts. These objects can be rotated and modified; the data display can be reorganized to show contour slices, color-coded poly-filled contours, surfaces, vectors, bathymetry, and topography. With the addition of Java scripts, these objects can be combined and animated in a 3-D VRML world that the user can move through and manipulate.

Researchers are able to "see," for example, what happens to ocean temperatures over time and at varying depths under El Niño and La Niña conditions. Scientists in NOAA's Carbon Exchange program use VRML to better understand another fundamental climatic influence—the fluxes in the vital transfer of carbon dioxide between the ocean and the atmosphere and their relationship to ocean temperatures. The tool enables the Fisheries Oceanography Combined Investigations (FOCI) program to examine the complex biological and physical oceanographic environments of fisheries.

Demonstrating OceanShare. NOAA's powerful new collaborative tool enables researchers at different NOAA labs to quickly access and work together with large-scale oceanographic and meteorological data sets from multiple geographically distributed archives.

NOAA researchers are using increasingly powerful scalable systems and software designs to improve the climate research and weather prediction needed by Government, industry, academia, and the general public.

VRML is a low-cost, Web-accessible application that can be widely used in education and research environments to model and interact with complex data in 3-D and stereographic 3-D. Broader deployment of these capabilities, however, will require the consistently high levels of reliability, QoS, and network speed now being prototyped in the NGI.

Participants

NOAA's PMEL
Old Dominion University

Sponsors

NOAA's HPCC and ESDIM programs

Computational challenges in climate and weather research

NOAA scientists increasingly use high end scalable computing systems and component-based software to improve the climate research and weather prediction needed by Government officials and policymakers, industries such as agriculture, transportation, and re-insurance, water managers and public health officials, academic researchers, and the general public. Advanced IT capabilities have improved scientists' ability to predict hurricanes and other severe weather events; to examine long-term influences on environment, such as the effect on hurricanes of increased carbon dioxide (CO^2) in the atmosphere and the ocean thermal response to climate warming trends; and to understand macro-features of the Earth's biosphere, such as eddies in the North Atlantic and atmospheric mixing in the Southern hemisphere.

Improved network speed and bandwidth will enable researchers to use increasingly powerful scalable systems and software designs to work collaboratively with more detailed and complex statistical models shared over high-speed networks.

Sponsor

NOAA

NIH biomedical collaboratory testbed

Researchers in structural biology and computational pathology working at geographically dispersed institutions are using collaboratory tools—including electronic notebooks, source code repositories, and data-sharing and teleconferencing technologies—to examine and evaluate the effectiveness of

these tools to improve their research. One group of collaborators is studying the biological functions of proteins using AMBER and CHARMM simulation software that depicts the structure and movement of proteins and DNA molecules. A computational pathology group is developing automated techniques to improve the speed and accuracy of tissue diagnosis. In collaboration with NCI researchers, they are working on 3-D visualization of prostate tissue samples and using DNA and RNA sequencing to correlate tissue characteristics with genetic tendencies.

Multicast technology and uniform routing protocols on the NGI will improve the high end forms of research collaboration including audio, data, and teleconferencing applications. Faster and more reliable connectivity among collaborating sites will make it possible to use and exchange large data files, 3-D volume-rendering graphics, visualization files, and source code.

Participants

NIH's National Cancer Institute
Pittsburgh Supercomputing Center
Scripps Research Institute
University of Pittsburgh Medical Center
University of Pittsburgh Center for BioMedical Informatics
University of California at San Francisco

Sponsor

NIH's National Center for Research Resources (NCRR)

BioCoRE and interactive molecular dynamics (IMD)

Using this portable collaboratory software for structural biology, researchers at different sites can start, visualize, and interactively steer molecular dynamics simulations on remote high performance computing platforms. BioCoRE, which stands for Biological Collaborative Research Environment, is a network-centered meta-application that has four components: a workbench that provides analysis tools, data sharing, resource allocation, simulation control, and interactive molecular dynamics; conferencing that includes audio and visual communication, visualization, and training; a notebook for recordkeeping; and a documentation capability. In the BioCoRE environment, researchers use its IMD molecular visualization program and simulation engine to display 3-D models of molecular systems such as proteins, nucleic acids, and lipid bilayer assemblies. In addition, a force feedback tool enables researchers to, in effect, poke at a system or change its shape to see how it responds. The application currently runs over NSF's vBNS network. The full package of BioCoRE tools is scheduled to be available to researchers in 2000.

BioCoRE's data-heavy and bandwidth-heavy applications require high-speed data transmission over networks and the high end storage capacity and operating speeds of supercomputers. The NGI will provide low latency to respond to user input and progressively higher bandwidths needed for increasingly complex simulations. The NGI will also enable transparent use of available distributed hardware, software, and databases.

Participants

NIH
Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana-Champaign

Sponsor

NIH's NCRR

Using BioCoRE portable collaboratory software, researchers at different sites can commence, visualize, and interactively steer molecular dynamics simulations on high performance platforms. Here, a researcher (left) is demonstrating the use of a force feedback tool to change a molecular shape and see how it responds (right). The full BioCoRE package of tools is scheduled to be generally available to researchers in 2000.

Terabyte Challenge 2000: Project Data Space

This collaboration among IT research institutions around the world is designed to establish standards, new network protocols and services, and performance monitoring tools for distributed data analysis and data mining across high performance networks. "Mining" is the work a computing system does to find and organize relevant data. The larger and more widespread the archives of potentially useful data, the more processing is required. The vast quantities of information stored on computers around the world make the capacity to mine it one of the most important technical computing challenges of this era.

Terabyte Challenge 2000 participants have developed a distributed testbed and knowledge network for experimentation and disciplinary studies in managing, mining, and modeling large (too large to fit in the memory of a single workstation), massive (large and distributed on both tape and disk), and geographically distributed data sets. The testbed, comprising clusters of workstations connected with a mix of traditional and high-speed networks, uses several software infrastructures—including PAPYRUS, a data-mining and predictive modeling system developed by the National Center for Data Mining at the University of Illinois at Chicago—to aim for the processing speed and data validity benchmarks required for high performance networked applications. The wide area data transfers called for in distributed data mining become practical only with the faster speed available on the NGI.

Terabyte Challenge members, whose demonstration at SC98 won the Most Innovative of Show award in the High Performance Computing Challenge category, are currently working with testbed applications that include:

- Anomaly prediction in brain scan data
- Digital Sky Survey: a classification of cosmological objects in widely distributed astronomical data
- Data mining and visualization of grid data in tele-immersion environments

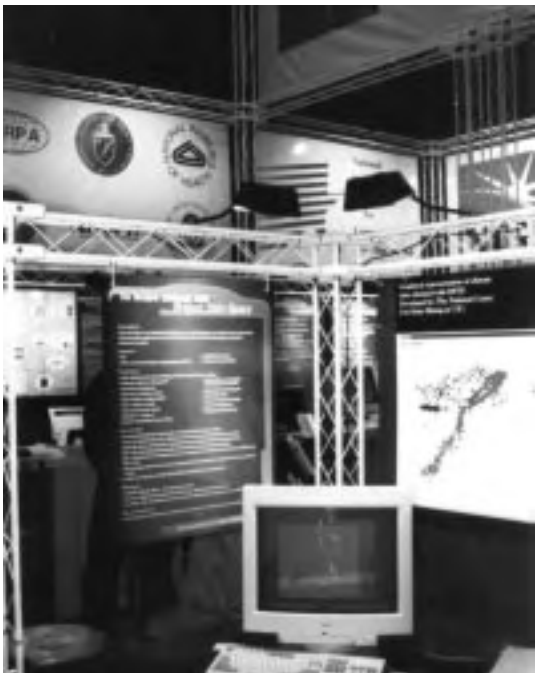
- Impact of data-intensive applications on next-generation networks
- Increasing the availability and usability of network traffic-monitoring data

Participants

National Center for Data Mining at the University of Illinois at Chicago (lead)
National Computational Science Alliance
National Laboratory for Applied Networking Research (NLANR)
New York University
Rensselaer Polytechnic Institute
University of California at Davis
University of Illinois at Chicago
University of Pennsylvania
Washington State University
Magnify, Inc., Chicago
Imperial College, London, England
National University of Singapore
University of Hong Kong
University of Toronto, Canada
ACSys, Canberra, Australia

Sponsors

NSF
DOE/ASCI
DOE's Office of Science
NASA Ames Research Center



DOE-supported Terabyte Challenge members, whose demonstration at SC98 won the Most Innovative of Show award in the High Performance Computing Challenge category, are currently working with testbed applications such as biomedical visualization and data mining. Wide area data transfers in distributed data mining become practical only with the faster speed available on the NGI.

Presentation at the NGI demonstrations conducted in Asheville, N.C., on December 6, 1999, at Blue Ridge Community College.

Asheville, North Carolina, demonstrations

The IT R&D agencies and NCO were invited by Congressman Charles Taylor (R-NC) to conduct demonstrations at Blue Ridge Community College in Asheville on December 6, 1999. Approximately 100 local educators, administrators, and technology leaders attended, including the Congressman and members of his local staff. The three-hour event included brief presentations by NCO, agency representatives, and the participants, followed by an open session of the demonstrations, many of which were tailored to local interests. Staffers at NOAA's National Climatic Data Center (NCDC) in Asheville, for example, requested a demo on the Digital Libraries Initiative (DLI). The presentations included:

Engineering and science tools of the future

The Next Generation Revolutionary Analysis and Design Environment (NextGRADE) and Immersive Visualization software suite provides a GUI that enables rapid assembly and analysis of both aircraft and spacecraft, allowing structural designers to quickly modify components and analyze design features. After the analysis is complete, the engineer can examine the results using an ImmersaDesk 3-D stereographic environment to view and interact with the structure from various perspectives in real time. These research tools will enable multiple geographically dispersed users to collaborate within multiscreen virtual reality environments.

Participant	*NASA Langley Research Center*
Sponsors	*NASA Cross-Enterprise Technology Development Program* *NASA Intelligent Synthesis Environment Initiative*

Digital library technologies for education

The Alexandria Digital Library (ADL) provides access to geospatial data such as maps, aerial photographs, and remote-sensing images taken from satellites, and to geographically referenced digital information. The Alexandria Digital Earth Prototype (ADEPT) project provides an interactive instructional environment that combines geo-referenced data with modeling analysis tools. (For more information, see page 65.)

Major Participants

University of California at Santa Barbara
University of California at Los Angeles
University of Georgia
Georgia Institute of Technology
San Diego Supercomputer Center

Major Sponsors

NSF
DARPA
NASA

Tele-nanoManipulator

The UNC-Chapel Hill tele-nanoManipulator system allows scientists to see, touch, and directly modify viruses, nanotubes, and other nanometer-scale (one-billionth, or 10^{-9} meter) objects. Using remote microscopes and graphics supercomputers over advanced networks, researchers can, for example, measure the rupture strength of DNA, build nanometer-scale circuits with carbon nanotubes (Buckytubes), and measure the strength of the adenovirus capsid that is used in gene therapy.

Participants

NIH's Resource for Molecular Graphics and Microscopy
UNC's Departments of Computer Science, Physics & Astronomy, Chemistry, and Psychology; Schools of Education and Information & Library Science; and Center for Gene Therapy

Sponsors

NIH's NCRR
NSF
Army Research Office
Intel Corporation
Office of Naval Research
SensAble Technologies, Inc.
Silicon Graphics, Inc.
ThermoMicroscopes

Telescience for advanced tomography applications

This project builds Web-based collaboration tools providing remote access to high resolution, 3-D microscopy in biomedical and neuroscience research. Using these tools, scientists are able to access and work with data from specialized imaging instruments such as high and ultra-high voltage electron microscopes at remote facilities. The telescience tools provide transparent access to supercomputing resources required to produce, refine, and analyze complex 3-D images of cellular and molecular structure and function. With geographically distributed image databases enabling comparisons among archived research results, the collaboration tools can help molecular biologists nationwide study the basic structures of cellular organelles like mitochondria and debilitating disorders such as Parkinson's and Alzheimer's disease.

Sponsors	*NIH's NCRR* *NSF's Partnerships for Advanced Computational Infrastructure Program*

Enhancing search engine effectiveness

The Text Retrieval Conference (TREC) project is described on page 61 of this book.

Participants	*Academic, industrial, and government organizations worldwide*
Sponsors	*NIST* *DARPA*

Computational challenges in climate and weather research

This NOAA demonstration is described on page 106 of this book.

Software Design and Productivity

Overview

In its eye-opening 1999 report to the President, the PITAC concluded that not only is the demand for software exceeding our ability to develop it, but the software produced today is difficult to design, test, maintain, and upgrade. Indeed, software complaints frequently make headlines in the daily news. The seemingly insignificant software failures that periodically shut down large parts of the Nation's telecommunications systems and the "Year 2000" (Y2K) problem that captured national attention as 1999 drew to a close are only two illustrations of what can go wrong with software design—and the immediate impact that such failures can have on a wide segment of the populace. Software today—unlike airplanes, office towers, or nuclear power plants—is constructed without the benefit of established standards by technologists and computer programmers who frequently lack thorough training in the engineering sciences of reliability, maintainability, and cost-effectiveness.

Recognizing that software is a key component of IT research, many Government agencies have worked on software design and productivity issues. But the sense of urgency underlined by the PITAC report has led to the creation of the new Software Design and Productivity (SDP) PCA for FY 2001, and the SDP Coordinating Group (SDPCG) is developing its R&D agenda. It is anticipated that IT R&D proposed in SDP will focus on significantly improving the concepts, techniques, and tools that underpin our software infrastructure. SDP activities will help educate Government, academic, and industrial software developers in well-founded and more cost-effective engineering to create efficient, reliable, and useful software. Participating SDP agencies include DARPA, NSF, NASA, NIST, NOAA, OSD/URI, NIH, and EPA. The SDPCG will encourage and facilitate interagency collaborations for Federal software design and productivity R&D. Some of DARPA's, NSF's, and NOAA's current or planned SDP R&D activities are highlighted in this section. Other agency activities will be highlighted in future reports.

Software engineering of complex systems

Today, we do not understand how to design and test complex software systems with millions of lines of code in the same way that we can verify that a bridge is safe. New fundamental software research will increase software productivity, make software more reliable and easier to maintain, and automate the discovery of errors. Beyond that, beginning in FY 2000 DARPA plans to study software development technologies to support continuous evolutionary development of long-lived military software systems. This research will include rationale management, evolutionary programming environments and languages, and system modeling analysis and composition.

Active software

Active software participates in its own development and deployment. The first examples of active software are already available on the Internet in the form of downloadable Java "applets." Research in active software will lead to software that can monitor its own progress toward a particular goal, discover a new capability needed for the task at hand, safely and securely download additional software needed to perform that task, and update itself. DARPA R&D will in particular support the study of software that can dynamically update and adapt due to physical interaction with its environment.

Software for autonomous systems

This DARPA-funded project is developing software to enable predictable, safe, and cooperative operation of free-ranging, autonomous systems. Research will lead to more intelligent software to control robots such as unmanned vehicles that keep U.S. Armed Forces from harm's way or devices that can explore planets or places on Earth that are unsafe for human travel; intelligent agents or "knowbots" that search the Internet on the user's behalf; robots and knowbots that plan, react appropriately to unpredicted changes, and cooperate with humans and other robots; and cars that drive themselves and automatically avoid collisions. These and other autonomous systems should be able to learn and adapt to change and uncertainty while improving with experience. For the warfighter, autonomous systems will enable revolutionary, asymmetric military capabilities, such as the ability to autonomously convey lethal and non-lethal military payloads anywhere on the battlefield without requiring human operators, and the ability to autonomously retrieve, process, and deliver information.

DARPA-supported scientists are conducting autonomous software research in the following areas:

Common software for autonomous robotics

R&D in common software for autonomous robotics will develop critical enabling software technologies reusable across a wide range of mobile autonomous robotic systems. In FY 2000, DARPA-led researchers will develop architectures to integrate deliberative, reactive, and learning behaviors, including knowledge representations, and will conduct laboratory demonstrations of alternative approaches to off-line learning, rapid sensor-motor mapping, engineered behaviors, and statistical control. In FY 2001, DARPA plans to support experimental evaluation of networking protocols for distributed robot controls that are more energy-efficient than conventional implementations, and to conduct a prototype demonstration and evaluation of software for distributed robotics that can coordinate 10 or more robotic devices in a collective task.

Software-enabled control

Software-enabled control R&D will leverage increased processor and memory capacity to increase our ability to maintain control over mobile devices through innovations such as predictive mode changes, dynamic control scheduling, composable coordinated control, and dynamic sensor and actuator allocation. In FY 2000, DARPA-supported researchers will specify an architecture for a hybrid control system that synthesizes the control law approach with computationally enabled mode logic scalable to very large state spaces of at least 100,000 states; develop active transition control and joint mode logic/control law designs; and design services for active model creation, augmentation, and query. In FY 2001, DARPA plans to prototype a multimode control architecture and framework; develop a predictive active model framework and parametric predictive and adaptive control frameworks; and complete multilevel, multimodal advanced design tools.

Agent-based negotiation

The goal of agent-based negotiation R&D is to allow the autonomous operation of large collections of agents negotiating real-time resource allocation issues such as those in logistics and countermeasures. In FY 2000, DARPA-led researchers will develop the framework for a bottom-up organization of autonomous software and define a strategy for consolidating and tasking responses to thousands of software agents with minimal human intervention. In FY 2001, DARPA plans to develop a strategy for predicting global behavior of large negotiating teams and to prototype autonomous software that negotiates in a logistics scenario.

Large-scale networks of sensors

Networking microsensors and embedded devices creates new stresses on hardware and software, including rapid self-assembly, timely acquisition, processing and exchange of sensor data, and energy-efficient operation. Accurate identification of events and information requires new ways of cooperation among these devices to process physical signals and integrate information in the network. Additionally, remote querying and accessing data collected by the sensor net should be simple, with easy-to-use interfaces.

DARPA activities in this area will extend software and networking R&D to geographically distributed microsensor networks. A major challenge is the development of software technologies that span a variety of sensor nets on ground, water, buildings, and humans. Another challenge is to design reliable networked embedded systems retaining only supervisory control, while automating traditional "in-the-loop" tasks. Sensor tasking, data collection, integration, and analysis must be fully automated to enable operation within time constraints far shorter than could be achieved by human operators.

In FY 2000, DARPA research in large-scale networks of sensors will:

- Specify diffusion-based approaches to networking and aggregation and distribution of information from large numbers of multitaskable sensor nodes
- Investigate declarative interfaces for querying and tasking networked embedded systems and develop a prototype based on relational database query technology and a lightweight operating environment
- Explore low-latency system designs and develop experimental platform and simulation capability
- Develop methods for collaborative signal processing and information integration

In FY 2001, DARPA will:

- Implement the FY 2000 automated information aggregation and diffusion prototype on a network with at least 50 nodes and 100 sensors
- Specify interfaces supporting common run-time services required by signal generation and processing applications
- Implement networked detection, estimation, tracking, and information integration
- Demonstrate multinode sensor network software and the benefits of collaborative signal processing in military operations such as fast-moving target detection and urban operations

- Prototype declarative interfaces for tasking and querying multitaskable sensor networks
- Investigate efficient interoperation of fixed and mobile sensors

DARPA is also supporting research in this area in the LSN PCA (page 82).

Component-based software design

NOAA's GFDL is working with Princeton University to redesign its major models using component-based software design. This program will produce a shared modeling infrastructure to enhance communication and software sharing while reducing redundancy. The program also supports the development of guidelines and standards for climate and weather model design that facilitate component reuse and modeling collaborations. At present, the new GFDL modeling system includes two global atmospheric models, a large assortment of atmospheric physics parameterizations, a comprehensive atmosphere-ocean-land-ice coupler, and an array of support tools. GFDL will incorporate this approach to redesign its Modular Ocean Model, used by oceanographers worldwide. This effort will assist GFDL's transition to scalable parallel computing architectures by isolating parallel memory management and I/O issues into a few modules that are shared by all components of the modeling system.

Promising research areas

Component-based software development

Today, most programs are written from scratch, one line of code at a time. Although interchangeable parts are fundamental in manufacturing and other engineering disciplines, the software industry lacks the equivalent of these components. Nonetheless, many in the software community—including the PITAC—think that the creation of truly usable component-based software development methods is one of the critical steps needed to achieve higher productivity and greater assurance in the end product. Component-based research will make it easier to find the right software component, to accurately predict the behavior of a software system assembled from components, and to support a growing electronic marketplace in software components.

Difficult issues in developing a component-based approach include the huge diversity of software products; the high rate of technological change; the lack of accepted standards for data representations, communication protocols, and component descriptions; the lack of models of software behavior that help predict the properties of component compositions; and the lack of a business model for the development and distribution of components. Real-world issues of code mobility, middleware infrastructures, distributed development teams, and the move away from the desktop model for computing must be understood and factored in. NSF anticipates funding R&D in software architecture to support product families, allowing recombinations of large components.

End-user programming

One way to address the critical shortage of programmers is to empower end users by creating more domain-specific development tools, such as user-friendly spreadsheets. Just as telephone companies solved the problem of scalability by putting dialing technology in the hands of end users, software developers can push portions of development out to users through various kinds of "smartware" —domain-specific and user-friendly templates or software "containers" that can be tailored to fit a variety of customer needs. Tools and techniques that make end-user programming more widespread will require significant advances in intelligent templates, domain-specific languages, and programming-by-example.

Empirical software engineering research

Software is an increasingly important $400-billion-a-year industry in the U.S., so it is essential to focus software engineering research on projects that will lead to improvements. Empirical validation can lend insight into the value of new ideas for improving the productivity of the software development process and creating higher-quality, more effective designs for software systems, thereby helping identify promising new ideas and pointing the way to others.

NSF plans to support projects that investigate such questions as: What causes software failures? What are useful sources of information for software maintenance? What are the development environment factors that affect the quality of software projects? NSF will also support longitudinal research that studies project artifacts to determine significant factors in the successful and unsuccessful evolution of software products. Other research will explore the efficacy of specific ideas for improving the productivity or quality of the development process and improving the design of software for fault tolerance, reliability, security, and similar characteristics.

NEW STARTS

Software for embedded systems

Beginning in FY 2000, DARPA researchers are focusing on the increasingly ubiquitous phenomenon of embedded software, or software for networked computing devices that are embedded in physical processes. Already critical in modern automotive engineering, where the power plant, braking system, and other elements are controlled by embedded software, such software may prove to be a major integrator in the years ahead for important military systems. DARPA's goals are to:

- Create conditions (technology, proof of concept) that support a viable market for reusable and customizable tools and frameworks
- Make embedded systems tools from third-party vendors accessible to large embedded system developers
- Encourage spill-over to other embedded software markets

This project will develop a new class of software to deal with the processing of physical world information by networked embedded devices. The rapid increase of processing power and the decreasing size and cost of today's microprocessors have created new devices and microsensors that will enable a new wave of DoD applications. For example, cheap, smart microsensors can be deployed quickly in large quantities in the battlefield to perform new monitoring and control functions, and a host of sensors can be attached to warfighters and assets to autonomously monitor safety and health information and equipment conditions.

Dynamic assembly for systems adaptability, dependability, and assurance (DASADA)

As software systems become more complex, they must be able to reconfigure and evolve themselves dynamically while the system is in operation. DARPA-led research will develop the dynamic gauges or measures of composability necessary for software components from multiple sources to support assured applications. The DARPA dynamic assembly for systems adaptability, dependability, and assurance (DASADA) program will ensure that the critical properties of

complex, heterogeneous software systems are maintained during and after composition, adaptation, and deployment. Researchers will conduct preliminary demonstrations of dynamic software component composability with multiple standard communications infrastructures including distributed component object model (DCOM), CORBA, distributed computing environment (DCE), or structuring (such as XML, RDF, and document object model [DOM]).

Model-based integration of embedded software

Tight integration of information processing with physical processes demands new technology for the integrated modeling of software and physical systems. These models will enable designers to capture complex physical constraints that the embedded software must satisfy. R&D in model-based integration of embedded software will use integrated models to analyze and verify the aggregate behavior of software and physical processes and to automatically customize and integrate system components. In FY 2000, DARPA researchers will:

- Investigate new modeling methods that identify physical constraints in embedded systems such as avionics and vetronics
- Develop customizable modeling tools that rapidly adjust to different modeling views and application domains
- Investigate new generation technology to configure, customize, and synthesize software directly from models

In FY 2001, DARPA plans to:

- Develop modeling tools to manage overlapping modeling views
- Investigate methods for the mathematical modeling and composition of model-based software generators
- Develop customizable frameworks for embedded software and demonstrate the rapid synthesis of embedded systems using customizable frameworks and model-based generators

Networked embedded systems

The large-scale networking of embedded and autonomous devices creates new requirements for embedded technologies that can achieve drastic reductions in costs while being compatible with a wide range of network and computation media; flexible mechanisms for naming, addressing, configuring, and administering communication and computation resources; and system design technologies that emphasize dynamic behavior guarantees. These challenges are addressed in DARPA's networked embedded systems R&D.

Programs under development at DARPA will:

- Establish methods to maintain and update critical information system-wide, without a centralized depository
- Investigate event/time-triggered system synthesis subject to time, functional, performance, safety, and security constraints
- Investigate design methods for embedded generators to guarantee select behaviors of the generated systems

High Confidence Software and Systems

Overview

From air traffic control systems and electrical power grids to emergency life-support systems and bank and insurance record-processing, advanced IT systems and networks power highly complex large-scale infrastructures critical to the safety, security, and well-being of society. Nearly everyone today relies on such networks for a vast assortment of everyday activities, from sending letters and paying bills to listening to music and conducting research. HCSS R&D focuses on the software and systems technologies necessary to ensure that the IT systems we count on achieve extremely high levels of reliability, availability, protection, restorability, and security. High confidence technologies make systems resistant to component failure and hostile intrusions—such as the massive denial-of-service (DoS) attacks on large commercial sites that took place in February 2000—and enable them to adapt or reconfigure themselves in response to damage or a perceived threat. Research in HCSS technologies encompasses all facets of IT systems, including the computing platform, the network, systems and applications software, the information in the system, and the interactions among all these components.

HCSS applications span national security, law enforcement, life- and safety-critical systems, transportation, telecommunications, personal privacy, and the protection of core elements of the Nation's information infrastructure. Federal agencies participating in HCSS research include NSA, NSF, DARPA, NIST, NASA, NIH, and OSD/URI. This section highlights some key current HCSS activities and FY 2001 goals.

NSA research

NSA's information assurance (IA) research focuses on emerging technologies that will have the most significant impact on Federal, private, and public mission-critical communications networks in the years ahead. Trends include continued evolution toward fiber optic networks, expansion of high-speed wired networks and mobile wireless technologies, growing needs for IA within and across networks, and proliferation of commercial security products, which will require improved identification and validation technologies. In FY 2000, NSA researchers are focusing on three major IA challenges: high assurance computing platform (HACP), security management infrastructure (SMI), and NSANet testbed applications.

High assurance computing platform (HACP)

HACP R&D aims to provide high assurance for the typical computing enclave of workstations and servers. End users in such enclaves must be able to share information, but the system must prevent information leakage and/or corruption. HACP requires a guarantee of controlled information sharing at different classification levels, enforcement of a customer-derived security policy, detection of and reaction to violations of that policy, and assurance features for the resulting configuration. Implementation steps include a proof-of-concept using limited trust and available cryptographic technology to demonstrate a COTS thin-client that can handle multiple classification levels of information with an acceptable level of assurance, and improvements to a commercial implementation of a Linux operating system. These efforts will help major COTS vendors realize a commercial HACP.

Security management infrastructure (SMI)

Effective control of virtually unlimited global connectivity will depend on SMI technologies. The fabric of this infrastructure will be woven from government and commercial key management infrastructures (KMIs)—systems to manage authentication and access protocols—augmented by security tools for software downloading, system audit, intrusion detection, and access management. Using HACP and NSANet, researchers will formulate public key infrastructure (PKI) and KMI roadmaps and develop or enhance security features for evolving SMI-related protocols and techniques for security-critical SMI functions.

Research topics include transport layer security; secure shells and a secure mail protocol; security features for ATM and multicast; SMI for optical and wireless networks; functional SMI elements such as cross-certification, certification revocation, and key recovery; and trust management in open systems.

NSANet testbed

To examine security issues in interconnecting secure and more-open networks, NSA research focuses on information systems and networks that support the national signals intelligence (SIGINT) system. An environment with many hierarchically structured enclaves, solid boundary protection, virtual communities of users with shared services, and distribution of information through Web services, NSANet will be used as the testbed secure network to integrate HACP and SMI prototypes, conduct research on detection, reporting, and response to hostile and sophisticated attackers, and demonstrate technologies for information sharing among independent and secure networks. In FY 2000, research is continuing in high assurance virtual private networking (HAVPN), a project started in FY 1999 to develop technologies enabling information sharing across high-speed, interconnected distributed networks built from predominantly COTS products that carry highly sensitive information.

Cryptography

Cryptography makes overall information assurance possible in IT systems. NSA provides the Federal government's cryptographic algorithms, backed by the highest level of crypto-mathematics expertise. Mathematical research provides a theoretical basis for designing algorithms for the unique requirements of the military, DoD, and the intelligence community. NSA's

multiyear research in public key cryptography will produce efficient public key algorithms and protocols, faster and more efficient arithmetic techniques, elliptic curve software, proactive authentication techniques, related technical support, and public key cryptography standards support.

NSA provides cryptographic technical consulting services to other Federal departments and agencies, working with NIST on hashing (generating a fixed-length value representing an original document of arbitrary length, designed so that a small change in the document will produce a large change in the fixed-length value), digital signatures, and key exchange algorithms; with NASA on command and control upgrades; with NSA programs in high-speed wireless and nuclear command and control cryptography; and with the military on, for example, next-generation Global Positioning System (GPS) upgrades.

Active network defense

Active network defense, a relatively new focus for NSA, provides research and advanced technology development for DoD's defensive information operations. This effort recognizes the technical reality that, while all possible attacks cannot be repelled, systems must nonetheless continue to provide information services, even at diminished capacity, until defenses can be mounted. In FY 1999, NSA initiated multiyear research investigations including:

- A project jointly sponsored with OSD and DARPA to explore a minimum essential information infrastructure (MEII) for DoD
- Development of novel, patternless anomaly detection technologies that identify unusual behavior in networks. Early test results of a prototype implementation correlate with anomalies identified by a proven intrusion detection system.
- An investigation of the vulnerabilities of SS7 commercial telephony signaling systems and strategies for strengthening them. The effort, undertaken in conjunction with GTE, extends the DARPA-funded Kingsman program, using an NSA testbed. Researchers are working with the University of Maryland on improving SS7 network security standards.
- An ATM mapping and monitoring tool that employs network-based sensors to monitor and respond in real time to certain ATM protocol events. The system allows multiple users to ask the same sensor to listen for different types of events and to respond to the user only when those events are encountered. The sensor can be used for monitoring tasks besides security, including network configuration and behavior and bandwidth management. Prototype implementations in multiple ATM domains found network problems within minutes. Researchers are enhancing the prototype, which offers capabilities not yet available in Federal or COTS implementations, to include additional monitoring tasks and an ATM network access control policy.

Secure communications

Secure communications research focuses on providing information security (INFOSEC) services for data moving over a public infrastructure or over public airways. High assurance capabilities are imperative in times of crisis and attack. This NSA research program encompasses the following technologies:

- *Speech coding,* to develop low bit rate algorithms often required for digitizing, encrypting, and transmitting tactical voice communication, and noise preprocessing and forward error correction research to enhance the quality of the voice signal.

  Researchers are concluding work, for example, on Mixed Excitation Linear Prediction (MELP), a system that will compete to be selected in March 2001 as NATO's standard system for seamless, end-to-end encryption of Allied voice communications.

- *Wireless communications,* to investigate and counteract the vulnerabilities of wireless services, use the results to influence the standards, provide select demonstrations for critical wireless technologies, and perform the testing, evaluation, and verification needed to ensure that the solutions work effectively.

  An NSA risk assessment of second-generation cellular technologies (TDMA, CDMA, and GSM) guided DoD's "Digital Battlefield" strategy for adding security features in the tactical environment. This included a military version of a CDMA phone developed by NSA to leverage commercial cellular technology and a next-generation cell phone for military operations. Using the DoD-developed Fortezza technology and a COTS operating system, NSA researchers have developed a wireless LAN security prototype that allows encrypted file transfer in a wireless Ethernet environment. An NSA analysis of vulnerabilities in commercial wireless standards supports agency efforts to promote security features and standards in commercial products with major industry and standards organizations. In FY 2000, NSA researchers will prototype a remote terminal testbed to demonstrate transmission of secure voice, video, and graphics to the desktop of a tactical warfighter.

- *High-speed secure communications,* including high-speed microelectronics and advanced packaging, techniques to incorporate cryptographic algorithms to get higher-speed performance, and consultations with customers to solve security problems.

  In FY 1999, NSA completed and published an architecture for a very flexible, high-speed ATM encryptor to be developed with DOE's Sandia National Laboratories (SNL). The design integrates two algorithms, enabling the product to interoperate with both public and classified-level networks. NSA fabricated a prototype 10 Gbps

encryption chip that researchers are using to consult on UltraFastLane, an ATM 2.5-10 Gbps encryptor used by the National Reconnaissance Office (NRO) and DOE. NSA collaborates with DARPA to study and test applications for high-speed semiconductor materials in high performance security systems. NSA research in high-speed cryptography developed hardware that can implement fast changes of key—a critical management capability in multicast key exchange, where large numbers of consumers need to communicate—and improved hardware and protocols in high-speed encryptors for satellite downlinks.

- *Encryption.* NSA supports NIST's development of an advanced encryption standard to replace the current data encryption standard (DES). NSA is implementing and evaluating the final candidate standard's algorithms and developing performance estimates.

- *Optical networks,* with a focus on developing proof-of-concept optical logic technology and switching devices tailored for cryptographic applications.

  To support Federal missions requiring transport of very large quantities of encrypted data—such as in DOE's ASCI—NSA is developing photonic circuits for encryption operating at speeds well beyond the upper limits of electronic circuits. Researchers are pursuing longer term research in multigigabit optical devices that can perform at the levels of complexity necessary for high-assurance cryptography—levels not attainable with today's devices. This requires photonic circuits with thousands of gates performing a cryptographic algorithm, occupying no more than one standard rack in a system. NSA is developing and demonstrating proof-of-concept optical logic and switching devices targeted toward implementing an all-optical key generator. Approaches under study include an optically controlled digital diode, a quantum effect device, and optical parallel processing.

Secure network management

Secure network management research supports SMI by developing secure protocols for information sharing, network control, and monitoring of events within networked information systems, and by participating in industry and standards organizations that shape network security standards and policies. NSA's development of the Internet Security Association Key Management Protocol (ISAKMP) standard through the IETF is enabling secure network connections and network security management. Research includes the following:

- In cooperation with NIST, NSA researchers developed Internet key exchange (IKE), a candidate standard and reference model for access and authentication on the Internet, and are simulating the integration

of several Internet key exchange and management capabilities. Researchers are working with the University of Maryland to demonstrate use of key formats and signatures such as the Digital Signature Standard as an authentication component for the Internet's DNS, in order to thwart attacks such as masquerading and interception. NSA proposed revisions to the IETF's SNMP, a basic element of network interconnectivity. Working with industry, NSA developed a reference implementation that strengthens the security features of the commercial standard.

- KMI allows the generation, production, distribution, control, and accounting of public key certificates and symmetric keys—user identification codes and their system-held matches, which must coincide to authenticate a user's identity. NSA researchers uncovered a certificate-masquerading vulnerability and produced a countermeasure, transmitting their results to commercial browser vendors. NSA research developed a way to update trusted certificates without user intervention. Research on security for the radio-based Tactical Internet focuses on authenticating command and control messages via digital signatures, understanding the limits of PKI, and maintaining confidentiality of command and control, employing group key concepts such as multicast.

- Research in multicast security and key management will develop and standardize efficient protocols for secure multicast. NSA has circulated a draft specification of a protocol for logical key hierarchy in Internet multicast that provides a flexible approach to re-keying a group in the presence of compromised users, and has tested the protocol. NSA-funded university research on multicast security concentrates on scalability and detecting malicious behavior.

Network security engineering

NSA's network security engineering research focuses on achieving high systemic levels of physical and operational performance and security, including network boundary definition and protection, security architectures, policy invocation and enforcement, assurance techniques such as trusted operating systems, and identification and authentication.

The DoS attacks by Internet hackers that briefly crippled leading Web sites such as Yahoo! and CNN.com in February 2000 made use of a technique called "spoofing" that is one subject of NSA's network security research. The hackers were able to spoof or fool Internet identification and authentication systems into helping them send myriad fake messages that then eluded the identification protocols of major sites, literally clogging their entranceways. NSA researchers are working to improve identification and authentication to provide a foundation for secure systems.

Research directions and activities include the following:

- Drawing on the scientific knowledge base in biology, physiology, and engineering, NSA develops and evaluates secure, cost-effective, accurate technologies and devices for detection and identification, such as fingerprint scanners and face-recognition systems. These provide a biometric component of authentication, eliminating spoofing vulnerabilities with techniques to assure that the person recognized by the system is alive and not an image. NSA's Tokeneer project combines smartcard tokens, biometrics, and cryptography to increase identification assurance to augment or replace password-based access systems. Researchers are evaluating a small fingerprint technology for remote log-in and an elliptic curve cryptographic algorithm for smartcards, and in partnership with DataKey/RSA have developed a downloadable cryptographic code for a Java-based smartcard.

- NSA is examining prototype distributed object computing security (DOCS) software products to raise security levels in a heterogeneous interoperable object-based networked environment.

- The goal of NSA research on a viable and secure trusted operating system (TOS) is to create an efficient architecture that provides the necessary support for security, executes programs user-transparently, and is attractive to vendors. NSA is working with DARPA, GSA, and NIST to integrate this architecture and security mechanisms to develop a secure version of Linux.

- NSA researchers developed and are evaluating Crackerbox—prototype system software to provide packet filtering and basic IP security. In demonstrations on an experimental platform, the software controlled access to system resources, provided a Web browser interface, operated as a firewall against network-based attacks, and secured hard drive files.

- NSA-developed SONET network interfaces have shown the ability to apply user-defined policies at OC-12 data rates to make sure that all incoming traffic is authorized. In collaboration with DARPA, researchers have developed prototype methods to streamline processing for application-level filtering in a high-speed firewall. In FY 2000, NSA is evaluating prototype security features on the NSANet testbed that provide access control for distributed enclaves without affecting quality of service, as well as monitoring the routing and switching layers of the network to detect anomalies. This product is being transitioned to a commercial network management software developer.

NSF research areas

NSF's HCSS research focuses on developing correct, safe, "no-surprise" software—software that behaves predictably in normal use and under

unanticipated conditions. Research activities span systems, software, networks, and communication. Information Technology Research (ITR), NSF's newest and largest IT program, begun in FY 2000, supports research needed to create and use the software-based systems of the future in as yet unimagined environments and applications.

Specific research areas include:

- Cryptography and other security mechanisms
- Hardware and software verification for guaranteeing system behaviors
- Formal software analysis techniques and fault-tolerant system concepts as bases for correctness and reliability

NSF-funded researchers are studying ways to integrate hardware and software analysis and testing to produce more reliable systems and exploring risk management and application of risk-based analyses to the design, construction, and certification of software systems. Project topics include certifying compilers, proof-carrying code, and data typing for automatic checking of desired software properties that traditionally have been difficult to assure.

Research in component-based software development includes design disciplines, testing methods, and methods that support systematic reuse of high-confidence components to reduce the sources of potential errors and development time. Projects explore different approaches, including theoretical investigations, empirical studies, construction of experimental systems, and creation of environments and tools used as exploratory testbeds.

Accomplishments of NSF research activities include advances in generating correct-by-construction software, behavior verification, security guarantees in distributed systems, and programming languages and environments that improve productivity and limit the possibilities for errors in constrained application domains.

DARPA's formal methods (FM) program

This FM program is developing the conceptual and technical frameworks for a body of engineering science principles and knowledge to govern construction of high-confidence computing systems, and the tools to test and validate system designs. The work grows out of the recognition among scientists, underscored in the PITAC's 1999 report, that the software infrastructure of today's computing systems and networks is a vast patchwork of often idiosyncratically designed and non-interoperable codes whose fragility manifests itself daily in unreliability, lack of security, performance lapses, errors, and difficulties in upgrading. At the same time, software complexity is increasing exponentially, including software used in safety- and mission-critical applications. This very complex software is extremely costly to develop but technically and economically nearly impossible to debug and validate.

These fundamental weaknesses in the contemporary software infrastructure led the PITAC to declare that scientifically grounded software was one of the Nation's top IT research priorities. The FM program starts

from the premise that public confidence in IT will remain limited unless research can generate radically higher levels of system security and reliability. FM activities are geared to enable systems and software designers to evaluate the quality of their work against rigorous scientific benchmarks, and customers to test and debug software products to be used in critical applications. Formal science-based methods will allow increased automation of software design, cost-effective approaches to the design process such as reusable software "modules," and a systemic engineering focus on interoperability and scalability—fundamental requirements for constructing the scalable information infrastructure of the future.

FM research covers a wide range of techniques and theories, including:

- Mathematics and logic
- Specification languages and methods (including both general purpose and domain-specific languages and logics)
- Formalization of desirable system properties
- Formal verification
- Application areas

Using these foundations from mathematics and computer science, FM research creates enabling technologies—tools for evaluating and verifying designs at various stages of development. FM tools will for the first time allow science-based analysis of requirements specifications, algorithm and protocol design, and executable programs.

Requirements specifications

Today, requirements specifications are underemployed. They are frequently not even written, and when written are often ignored and not kept up to date. They are regularly considered equivalent to documentation. The long-term objective of research in this area is to make writing requirements specifications seem more like programming, to accelerate system design by developing requirements specifications, and to keep specifications and implementation consistent. FM research will develop formal ways to specify aspects of systems, tools for consistency checking and debugging of specifications, and tools to check their correspondence with implementation.

Algorithm and protocol design

Software designs incorporate mathematical and logical formulas called algorithms and algorithm-based instructions called protocols that reflect agreement by the IT community on how software modules should interact with other parts of the system. FM advanced modeling tools can analyze the mathematical validity of algorithms and protocols, identify weaknesses and unexpected interactions, and define better solutions. This analytical capability is particularly vital in distributed computing networks, where algorithm and protocol designs are far more complex than those for stand-alone computing systems and play critical roles in security, telecommunications links, and fault

tolerance. The Verinet tool developed at the University of Pennsylvania, for example, can analyze network protocols from the algorithm level up through the routers that move messages through the Internet. A tool called Ensemble, developed at Cornell University to prove protocol properties, resulted in fast-path optimizations with no loss of system robustness.

Program analysis

Arguably the most difficult but important level of software to analyze, executable programs are substantially larger—often comprising millions of lines of code—than system specifications and thus nearly impossible to test thoroughly with current capabilities. FM research aims to exploit modularity in program design, in part to get at the problem that many system software problems arise from "misunderstandings" at interfaces where two programs must interact. By breaking large designs into modules and specifying interfaces for the points at which the modules interact, the research generates smaller, reusable program components whose interfaces are verifiable in testing. Programs designed in modular units would simplify design complexity and would reduce production time and cost. Mocha, a prototype modular verification tool developed by FM researchers, checks the validity of a design by comparing it with a simulated design derived from what it is supposed to accomplish (as formulated in the specifications). In a recent trial, Mocha found serious bugs in a VGI signal processing chip containing 96 processors.

National Information Assurance Partnership (NIAP)

In the National Information Assurance Partnership (NIAP) program (*http://niap.nist.gov*), NIST and NSA are developing security requirements for IT products and measures to evaluate their effectiveness. NIAP's long-term goal is to help increase consumer trust in information systems through cost-effective security testing, evaluation, and validation. Working closely with industry partners, NIAP also fosters R&D in security requirements definitions, test methods, tools, techniques, and assurance metrics, and facilitates developing a commercial security testing industry in the U.S.

Other NIAP accomplishments include:

- An automated tool to generate security requirements and security specifications for IT products and systems in accordance with the Common Criteria, an international standard
- Common Criteria-based protection profiles for technology areas such as commercial operating systems, database management systems, and firewalls
- An emerging evaluation and validation program for IT security
- A Health Care Security Forum that helps consumers define security requirements for health care information systems and applications

- A Smart Card Security Users Group that helps the financial payment industry define requirements for smartcards and appropriate accreditation criteria and methods for security testing laboratories
- A Telecommunications Security Forum that helps define security requirements for PBX and ATM switches, routers, and gateways
- A mutual recognition arrangement with Canada, the United Kingdom, France, Germany, Australia, and New Zealand to recognize results of IT product security evaluations against the Common Criteria

NIST high assurance Internet security architectures

In FY 2000, NIAP is continuing work toward a fully operational IT security evaluation and validation program employing accredited, private-sector security testing laboratories; Common Criteria certificates for validated IT products; Common Criteria-based protection profiles for technologies such as operating systems, database management systems, telecommunications and network devices, smartcards, and Internet browsers; and an automated tool for generating security tests. Another NIAP goal is to raise to 15 the number of nations participating in the mutual recognition protocol.

The rapid growth of e-commerce and pervasive computing has generated enormous demand for advanced security technologies. Buyers have many products to choose from, but frequently two or more products that individually are secure may actually weaken security through unexpected feature interactions. In addition, there are few widely accepted ways of evaluating the strength of security mechanisms. NIST is working with industry to develop new approaches for designing high assurance security architectures, efficiently using emerging technology products, testing the security of large networks, and evaluating the security of software products and systems. Goals are to develop:

- Definition of security services needed to meet the future needs of e-commerce
- Definition and specification of security components and applications
- A security testbed to evaluate interoperability and effectiveness requirements
- Metrics and benchmarks to develop, evaluate, and acquire security components, applications, and operating systems
- A methodology for testing and evaluating trusted operating systems, Internet security components, and applications

Traditional security components such as firewalls and encryption fail to provide a total security solution because the distinction between data and code is vanishing and damaging executable code is easily imported into large-scale networking environments. To ensure that attacks on networked servers can be defeated, the fundamental security layer must be moved down from the application to the operating system level, where decisions are made about access to file systems, devices, and processes. Trusted secure operating systems must be able, for example, to enforce security at each point of decision, denying or allowing access to a specific Web page or to specific fields in a database record. They also need accountability services that ensure swift and effective investigation of illicit transactions and activities.

NIST has developed and transferred to industry methods to integrate modern role-based access control (RBAC) mechanisms with trusted operating systems. To guarantee that all access control needs within Government and industry are met, NIST is prototyping a universal policy machine to serve as a model for future system development.

Internet Protocol security (IPsec)

NIST's IPsec research develops scalable technologies and tools to make the IP—the basic software framework enabling the routing and flow of Internet message traffic—more secure. IPsec enables a centrally controlled access policy and a multilevel security approach to provide security services including data origin authentication, connectionless integrity, replay protection, data confidentiality, limited traffic flow confidentiality, and key negotiation and management. The IETF has mandated the use of IPsec wherever feasible.

Cerberus, a NIST-designed reference implementation of the latest IPsec specifications, and PlutoPlus, a NIST reference implementation of the IPsec key negotiation and management specifications, are being used by the Internet industry in ongoing research on advanced issues in IPsec technologies. NIST's Web-based IPsec interoperability tester, known as IPsec-WIT, enables Internet researchers to conduct interoperability tests anytime and from any location without having to download test software or move the systems being tested. Plans include integration with PKI.

Mobile agent security and intrusion detection

In collaboration with Internet industry groups, NIST is studying mobile agent technologies to detect and defend against network security breaches. Objectives are to develop proof-of-concept prototypes to demonstrate mobile agents for network security testing and network management, develop standards for interoperable secure mobile agents, develop techniques to address security threats to mobile agents, and evaluate usefulness and scalability. NIST will use its database of network vulnerabilities, threats, and attacks to develop methods of detecting conditions that result in insecure systems.

Authorization management

New RBAC systems enable network managers to display and control the roles and privileges of network users and role and privilege inheritance. These tools simplify management of network authorizations but lack support

for multiple inheritance relationships, limiting their effectiveness in complex environments and increasing the chances of a security-compromising error.

NIST's prototype Role Control Center (RCC) ensures uniform treatment of privileges unique to a user and those assigned to a role, allows delegation of administrative responsibilities, and enables instantiation of RBAC users and roles on target systems. NIST has applied for an RCC patent and plans to integrate RCC concepts and tools in government and commercial implementations and to evaluate their costs and benefits.

Standards for critical infrastructure protection and e-commerce

Presidential Decision Directive 63 on critical infrastructure protection, issued in May 1998, requires Federal agencies to take concerted steps to eliminate vulnerabilities in the Nation's significant IT infrastructure systems. The directive established several entities to coordinate the effort, including the Critical Infrastructure Assurance Office (CIAO) and the Information Sharing and Analysis Centers (ISACs). NIST, as a participant in CIAO, works with the ISACs, which are public-private partnerships supporting research across all the critical infrastructure sectors, to expand the current understanding of vulnerabilities, threats, analysis, modeling, system interaction, and especially security and protection. NIST works with the e-commerce industry to reduce those vulnerabilities and threats. To help network managers distinguish between anomalies and real attacks, NIST has developed its I-CAT intrusion categorization tool to measure the level of interest in specific Internet penetration and denial-of-service attacks. Other NIST accomplishments related to e-commerce standards include:

- Cryptographic module validation program (CMVP), which validates cryptographic modules to Federal Information Processing Standard (FIPS) 140-1 and other cryptography-based standards
- Statistical profiles of the hacker community and published attack scripts
- IT Security Training and Awareness Resource Center

Expected results include:

- Identification of popular attack scenarios for critical infrastructure networks
- Security models and simulations for demonstrating essential security properties
- Security standards and test methods for critical infrastructure networks
- Security architectures to include security services for authentication, authorization, PKI, and intrusion detection

- Reference implementations to demonstrate security services and interoperability
- Security APIs to support the use and interoperability of cryptography, smartcards, access control, and intrusion detection

Software fault and failure data and analysis repository

NIST researchers are collecting empirical data on software failures and their causes, with the goal of improving software quality by establishing models of failures in real-world systems. A NIST study of failures in medical devices caused by software faults found, for example, that some medical device companies did not conduct generally accepted assurance activities to catch faults, and that a high percentage of failures resulted from conditions that would have been detected with only a small number of tests. Preliminary analysis of data from other software failure projects indicated that about one-third of faults were found in requirements activities and about one-fourth in system test or operation. Most software errors fell into the specification, logic, and computational fault categories. A NIST analysis of data from a large, complex distributed system found configuration management to be another major arena of software failures. This NIST effort has been endorsed by the software division of the American Society for Quality, and its concepts and tools are being used in graduate programs in computer science at Johns Hopkins and the University of Louisville.

Automatic test generation from formal specifications

In December 1999, *Business Week* reported that "bad software cost U.S. businesses $85 billion in lost productivity last year." To date, most research on improving software reliability through testing has focused on a limited and problematic automated approach called structural testing, which is based on execution paths and can only be performed on source code. NIST is developing formal methods to automatically produce software tests from specifications alone, using model checkers and mutation analysis. The goal is to substantially reduce the cost of testing software, which now consumes about 50 percent of software development budgets, cut time-to-market for companies producing software products, and provide a useful technique for organizations developing software standards. The NIST prototype, including a tool that allows developers to measure how completely any set of tests covers the behavior of a software product, is being evaluated by corporations for possible use.

OSD/URI fault-tolerant network protocols

In FY 2000, OSD's URI established a five-year research focus on real-time fault-tolerant network protocols. The aim is to develop protocols to allow the continuing operation and graceful degradation of large-scale computer and communication networks in spite of faults.

HCSS research agenda

Complementing the strong base in security established under the predecessor High Confidence Systems (HCS) Working Group, the HCSS Coordinating Group is completing a research agenda that outlines a possible

new initiative in assurance technologies and experimentation. The HCSSCG anticipates making this agenda available soon on the Web.

Social, Economic, and Workforce Implications

of IT and IT Workforce Development

Overview

New modes of learning, research, communication, commerce, and human services are proliferating so rapidly that we as a society have hardly paused to contemplate the changes wrought by IT or analyze its effects on people and institutions. SEW was established as a PCA in FY 2000 to support R&D dedicated to examining the ways IT is transforming our culture and to inspiring innovative models for education and training in IT environments. SEW is the successor to the Education, Training, and Human Resources (ETHR) PCA, reflecting the broader portfolio of issues this research now addresses. SEW R&D will focus on the nature and dynamics of IT impacts on technical and social systems, the workforce development needs arising from the spiraling demand for workers who are highly skilled in technology, and the growing "digital divide" between Americans with access to information technology and those without. Participating agencies include NSF, NASA, DOE, and NIH.

In its February 1999 report on IT research needs, the PITAC proposed such an expanded agenda for Federal R&D, arguing that the Nation "must invest in research to identify, understand, anticipate, and address" the problems posed by the increasing pace of technological transformation and "must ensure that all Americans are well-equipped to live and work in the changing world." The PITAC's February 2000 report, "Resolving the Digital Divide: Information, Access, and Opportunity," called for an intensive, coordinated national initiative supported by increased Federal funding to extend IT access and skills to groups bypassed by the opportunities of the Information Age.

SEW R&D activities encompass development and evaluation of advanced technologies for high-quality software learning tools, information-based models of educational systems and learning productivity, research on IT applications in cognitive processes, and demonstrations of innovative networking technologies. In addition, SEW R&D supports development of model curricula and course materials to promote IT literacy and graduate and postdoctoral programs to increase the number of IT professionals. The goal of SEW research is to maximize the societal benefits of IT by assuring that emerging information infrastructures productively serve the needs of all Americans and are sustainable over the long term.

Dynamics of social change

NSF's multidisciplinary SEW research builds on the agency's continuing work devoted to examining the scope, trajectory, and underlying processes of IT's transforming influence on public life, homes, and schools. NSF research is generating theories, models, and concepts to describe the dynamics of social changes being brought about by IT in many different arenas, such as in e-commerce, workplace organization and productivity, and scholarly research; new knowledge about the interactions among people, computers, and communication networks over distance and time; and scientifically grounded mapping of social, economic, and technological transformations across groups, organizations, institutions, and societies. Funded projects range from research studies to workshops and seminars.

FY 2000 activities and accomplishments include:

- Publication of "The Digital Dilemma: Intellectual Property in the Information Age." This NSF-funded in-depth study—conducted by the Computer Science and Telecommunications Board of the National Research Council—examines issues surrounding ownership of the electronic forms of creative "products" such as writings, music, and software.
- A pilot "Social and Economic Implications of Information Technologies Data Base," containing 4,000 listings and Web pointers for data sets, research papers, books, and Web sites. The citations are sorted into searchable "Road Maps" that cover IT implications for the community, commerce, education, employment and work, globalization, government, the home, institutional structure, productivity, science, and selected policy issues (*http://srsweb.nsf.gov/it_site/index.htm*).
- A Carnegie Mellon University (CMU) study examining how changing patterns in home use of telephones, television, and the Internet are affecting psychological and social well-being in families. The research will test the preliminary finding of an "Internet paradox"—that home Internet use decreases well-being.
- A Michigan State University study of computer and Internet use by African Americans. By examining the interactions of individuals newly introduced to home Internet connectivity, researchers aim to gain a deeper understanding of the factors contributing to the racial digital divide so that effective strategies to reduce it can be devised.
- A CMU workshop on social and organizational aspects of geographically and functionally distributed work will bring together researchers in anthropology, computer science, history, industrial engineering, information science, psychology, and sociology to discuss the new ways in which IT devices and capabilities are changing work arrangements and modes. One of the first such discussions to be held, the workshop will generate a research agenda, a Web site, and a book detailing research activities to date.
- A University of Maryland study of a trauma patient resuscitation team that examines how distributed teams of experts working in a highly dynamic, multitasking, and stressful environment coordinate their activities and use technology. The multidisciplinary medical research will

develop a model framework for coordination, identifying computation and communication technologies that could improve medical teamwork.

- An international workshop, co-sponsored by the European Science Foundation and the International Institute for Applied Systems Analysis, to identify critical challenges in deploying IT technologies to conduct global collaborative science research.
- Studies of informational and computational systems in formal organizations and electronic markets.
- Publication of "Science and Engineering Indicators 2000," the 14th edition of NSF's biennial publication documenting key trends in science and engineering research and education. Its chapter on information technologies provides a statistical and thematic overview of the emergence of IT as major factor in education, research, commerce, and cultural life.

Expanding the educational pipeline

Chief executives of leading corporations have identified the need to strengthen the technology workforce as the single greatest challenge to U.S. competitiveness over the next 10 years. There is a widening cultural gap between citizens who understand and use IT and those who do not. A July 1999 U.S. Department of Commerce report, using December 1998 Census Bureau data, showed that college graduates are at least eight times as likely to have a computer at home and nearly 16 times as likely to have home Internet access as those with an elementary school education. Households with incomes of $75,000 or higher are more than 20 times as likely to have access to the Internet as those at the lowest income levels and more than nine times as likely to have a computer at home. The persistent and growing digital divide also runs along racial and ethnic lines, with African-Americans and Hispanics substantially less likely to have access to the Internet in any location, and along geographic lines, with rural residents far less likely than urban dwellers to have access to network connectivity.

Several SEW activities focus on education and training needs, including public outreach programs designed to increase computer literacy, and scholarships, internships, and training initiatives at the undergraduate and graduate levels aimed at expanding the pool of Americans with advanced IT skills. The following sections highlight some of these activities.

Education, outreach, and training (EOT-PACI)

NSF's PACI, comprising the Alliance and NPACI, sponsor joint education, outreach, and training (EOT-PACI) activities to bring PACI-developed high performance hardware capabilities and software learning tools to bear in classrooms at all levels of education and in government, and provide training for teachers and other professionals in uses of new technologies. A number of EOT-PACI efforts focus on outreach to women, minorities, and individuals with disabilities, through such means as Web resources, mentoring programs, and IT internships. EOT-PACI projects include:

- EnVision for Kids is an NPACI project to develop a Web site and distribute an interactive hybrid (Mac/PC) CD-ROM aimed at 6th to 9th graders. Educational content will be based on current research in

NPACI's three thrust areas—technologies, applications, and education and outreach.

- The Computational Science Institute, an Ohio Supercomputer Center-sponsored program, is bringing teachers from around the state to the center for a week's training in using IT tools to stimulate students' interest in science and mathematics.
- Learning Technologies is a PACI project to develop, disseminate, and evaluate tools to enhance learning. The project will increase education researchers' access to PACI technologies and content for K-12 education and develop prototype educational materials and testbeds using PACI's collaborative resources.
- An Oregon State University research effort will develop a four-year curriculum leading to a Bachelor of Science degree in computational physical science. Once developed, the courses will be made available over the Web so that present degree holders can obtain a second bachelor's degree in that field.
- A project at the University of Wisconsin's NSF-supported National Institute for Science Education uses NPACI-developed video teaching and authoring tools to explore the concept of reversing the paradigm of lecture and homework. The project will measure the impact of the changed paradigm on student performance in a computational science course and evaluate the effect of advanced technology on the classroom experience.
- In workshops for faculty in the California State University System and on other NPACI partner campuses, San Diego State University's Education Center on Computational Science and Engineering (EC/CSE) is demonstrating how NPACI resources can be used in scientific investigation, discovery, and problem solving. Leveraging and customizing materials from NPACI partner sites, EC/CSE staff highlight technologies and tools that are applicable to undergraduate learning environments.
- A PACI project at the University of Houston-Downtown is developing a Web directory of minority institutions and their funded IT programs, to be housed at the university's Center for Computational Sciences and Advanced Distributed Simulation. The directory will be available on the Web site of the Association of Departments of Computer and Information Science and Engineering at Minority Institutions.
- The PACI Research Experience for Undergraduates (PACI REU) program is providing funding for 35 undergraduates this year to work with 18 PACI principal investigators (PIs) on their research.
- The Distributed Object Computation Testbed (DOCT), a prototype document management system created by an SDSC researcher in collaboration with the U.S. Patent and Trademark Office, enables officials to manage and update complex patent application documents from geographically distributed archives and computing platforms.




Elementary school students in Brownsville, Texas, share their Space Shuttle Simulation project via NASA's "Space Team Online" Web site and educational resource. NASA's Learning Technologies Project (LTP) applies the agency's vast resources to increasing public access to scientific databases, developing new applications and curriculum materials for K-14 students and teachers, and sponsoring educational programs that promote IT literacy and raise student interest in careers in the sciences.

- Collaborative Ecosystem Modeling, being developed at the University of Maryland with the NSF/EPA Partnership for Environmental Research, is creating techniques and tools to enable students, educators, policymakers, and stakeholders to build ecosystem models and run spatial simulations focusing on watershed management.

Training in biomedical informatics

As advanced networking applications spread to research, diagnostic, clinical, and training activities in medicine and biomedical science, the shortage of biomedical professionals equipped to use and maintain the complex computing and telecommunications systems supporting these activities has become increasingly acute. NLM, which pioneered the concepts of networked biomedical information archives and visualization tools, sponsors fellowship opportunities that prepare researchers and clinicians to integrate high performance computing technologies across the spectrum of health care environments.

NLM supports graduate, professional, and postdoctoral fellowships for one to three years of clinical or research training in biomedical informatics at Columbia University; the University of Missouri; the Oregon Health Sciences University; the University of Pittsburgh; the Regenstrief Institute; the Stanford University School of Medicine, the University of Utah; the Yale University Center for Medical Informatics; joint research training programs at Harvard University, the Massachusetts Institute of Technology, Tufts University, and the New England Medical Center; Duke University and the University of North Carolina; and the

W. M. Keck Center for Computational Biology at Rice University and the Baylor College of Medicine. In FY 2000, a total of 150 NLM fellowship positions were available at these universities, and NLM separately funded another 10 fellowships.

NASA's Learning Technologies Project

Nearly 42 years of Earth and space exploration at NASA have generated perhaps the world's richest archive of images and information about Earth, the universe, and the scientific and technological components of the discovery process. The agency's Learning Technologies Project (LTP) applies these vast resources to increasing public access to scientific databases, developing new applications and curriculum materials for K-14 students and teachers, and sponsoring educational programs that promote IT literacy and raise student interest in careers in the sciences. The LTP conducts these principal activities:

- K-12 Education Outreach Centers that integrate technological capabilities in curricula and disseminate these learning tools and materials throughout the education community
- K-14 Aeronautics Projects that use the Internet to engage students in science, mathematics, engineering, and aeronautics learning activities
- The Remote Sensing Public Access Center (RSPAC), which offers a user-friendly online hub of information, learning activities, and related science-learning sites, including RSPAC's popular Observatorium, a Web site that lets visitors make visual journeys around the Earth, across the solar system, and beyond
- Digital Libraries Technology (DLT) projects that support the development of new technologies—such as software tools, applications, and scalable software and hardware systems—that can accommodate more advanced user activities and substantial increases in public access to NASA via computer networks
- LTP's Special Projects, which foster broad public use, via the Internet, of Earth and space science databases generated by NASA and other agencies and encourage development of innovative software applications using the data

Social and economic implications of IT

In FY 2000, NSF began funding grants emphasizing broadly interdisciplinary and multi-institutional explorations of the social and economic implications of information technology. The research agenda calls for collaborations among specialists in IT design, integration, application, and socioeconomic implications (computer and information scientists, social informatics and information systems researchers, librarians, and industrial engineers); researchers in behavioral, cultural, economic, and social impacts of IT (historians, philosophers, and social and behavioral scientists); and researchers who study the ethical, legal, and social implications of IT in their disciplines (biologists, geologists, engineers, and sociologists). Key research areas include:

- Economic and technical systems that make up the digital economy, such as electronic markets and Internet commerce
- Causes and effects of unequal participation in IT by different social groups
- Interdependence of technologies, institutions, and communities
- Evolution and functioning of IT-based collaboratories and distributed work environments
- Impacts of IT on laws, ethics, and social norms
- Public access to government information
- IT involvement in public decision-making
- Development of new data and indicators for tracking IT
- Research methods for studying its socioeconomic impacts

NSF is especially encouraging:

- Studies that focus on the early life cycle of new information technologies, a period during which research can beneficially influence their future development and use
- Studies of extensively computerized or information-intense environments (such as wired communities, schools, universities, inter-organizational alliances) where research can "report back" on social and technical transformations
- Ecological studies that focus on the interdependence of IT, computerized groups, institutions, and communities, including social, political, and legal outcomes

The IT workforce (ITW)

Begun in FY 2000, NSF's ITW program addresses an aspect of the IT workforce shortage issue that is particularly worrisome to scientists and researchers—the underrepresentation of women and minorities in engineering and computer and information science occupations. ITW will support a broad set of scientific research studies focused on three basic themes:

- Environment and Culture—how the environment, culture, and other social contexts (such as households, neighborhoods, and communities) influence the attractiveness of IT, and how interest in and use of IT shapes the developmental environment, with particular emphasis on understanding the issues of different age groups
- IT Educational Continuum—how the educational environment influences students' progress from grade school to workforce entry, and why students with the potential to succeed in IT disciplines take educational paths that make it difficult for them to enter the IT workforce
- IT Workplace—why women and minorities with the potential to succeed in the IT workforce take alternative career paths, what barriers and obstacles they must overcome in an IT career path, and how the IT workplace can foster increased retention and advancement of these groups

ITW will emphasize multidisciplinary collaboration among researchers in IT, the social sciences, and education.

Federal Information Services and Applications Council

The Federal Information Services and Applications Council (FISAC) assists in the transfer of advanced computing and communication technologies developed by the IT R&D community to agency missions and systems across the Federal government. FISAC serves as a two-way information channel for officials and programs not closely involved in IT R&D. FISAC disseminates information about Federal IT research priorities, activities, and results to the broad Government community, and in turn advises the IWG/IT R&D about research needed to provide next-generation capabilities required by the Federal government. Because it is a forum for Federal end users of IT R&D applications, FISAC can identify technologies critical to agency missions, facilitate interagency collaboration, and provide the IWG with informed perspectives about technical requirements for Federal IT applications.

FISAC promotes the deployment of IT R&D products in Government by supporting multiagency demonstrations of advanced applications that offer promise for wider Federal implementation, encouraging pilot projects to assess the technologies required for specific Federal applications, and conducting briefings, workshops, and reviews on IT R&D developments. FISAC carries out these activities through its IT for Crises Management (ITCM), Federal Statistics (FedStats), NGI Applications, and Universal Access Teams, and by participating in NSF's Digital Government program. FISAC's work is funded through a combination of the IT R&D budget crosscut, IT R&D agencies, and non-IT R&D organizations.

Crises management

Managing crises to save lives, reduce economic loss, and preserve property requires an ability to assess changing situations, deploy life-saving resources quickly and effectively, and monitor results in the midst of an onslaught of information ranging from precise data to informal reports of uncertain reliability. Crises management is an activity in which government plays a key role and in which a broad range of players is involved. FISAC's ITCM Team identifies the IT R&D needed to field state-of-the-art technologies in emergency preparedness, crises response and mitigation efforts, and recovery coordination, and sponsors collaborative demonstrations and dissemination of these tools and applications. Team members include representatives from DoD, the Federal Emergency Management Agency (FEMA), NASA, NOAA, NSF, and other agencies.

Crises management research agenda

FISAC's ITCM Team has drafted and circulated for public comment a research agenda for developing capabilities to collect, evaluate, and synthesize large quantities and multiple types of information from many sources to generate real-time assessments of evolving situations, response scenarios, and distributed feedback systems. Such research would explore:

- Advanced networking, communication, deployment, and management in disasters, including pervasive sensing; rapid deployment of wireless networks at a disaster scene; nomadic technologies; adaptive and scalable networks for wireless, satellite, and wireline high-speed telecommunications under changing conditions; and use of high end data visualization corridor (DVC) and tele-immersion technologies at disaster centers.
- Distributed high end computing for disaster simulation. Researchers will develop models, similar to global climate and weather simulation systems, of data collection and simulation technologies for other natural disasters, such as earthquakes, fires, floods, tornadoes, tsunamis, and volcanic eruptions, and will explore computational grid or metacomputing strategies to access appropriate computing resources.
- Metacomputing techniques to assess disaster risks and guide mitigation strategies, such as building codes.
- Data-mining tools including data fusion of multimedia to discover the best sources of information on a particular hazard or during a crisis.
- Information triage systems to analyze, prioritize, and communicate information from varied sources to individuals involved in a crisis situation. For example, relief workers must be able to pinpoint the location of hazards at the disaster site and know whom to look for, while managers situated in a mobile control center must be able to direct the rescue efforts of many individual workers and target resources where they are most needed.
- Privacy and security protocols, especially in mobile technologies, and the ability to change them rapidly as crises progress. In order to treat an unconscious person, for example, emergency workers with appropriate clearance could access smartcards containing medical data normally available only to health care providers.
- Domain-specific judgment support and decision-making resources to assist individuals performing unfamiliar functions during or after a crisis.
- The viability and worthiness of tools and software for crises management, preparedness planning, training, and mitigation-strategy development. Researchers will prototype, test, and demonstrate these capabilities using testbeds and field tests with participating end users.
- Barriers to adopting IT tools within the crises management community. Such factors might include legal and regulatory barriers, lack of end-user training, absence of compelling cost-benefit analysis, or lack of technology transfer mechanisms.

The research agenda also identifies the need to develop educational programs and curricula for students and professionals and to transfer new technologies to appropriate organizations.

Digital Government

NSF's Digital Government program is aimed at using cutting-edge information technology to improve government services. Digital Government consults with FISAC in developing program announcements and soliciting proposals for projects that bring computing, networking, and IT researchers together with Federal agencies with significant information services missions. The program supports collaborative projects between these researchers and Federal agencies with information services challenges and issues. In FY 2000, NSF-funded Digital Government projects range from an application that lets users see fluctuations in gasoline data collected by DOE to a training program that will help Government managers use IT to improve delivery of Federal services. (For more detailed information about the Digital Government program, see page 47.)

Federal Statistics (FedStats)

More than 70 Federal agencies collect, analyze, and archive statistics of interest to the public in the course of carrying out their missions. The FedStats Team, jointly sponsored by FISAC and the 14 member agencies of the Interagency Council on Statistical Policy (ICSP), works cooperatively with agencies to identify IT R&D needed to make their data available online; construct an integrated national system of public statistical information; and promote partnerships with the IT research community through NSF's Digital Government program and related Federal initiatives. FedStats members participated in a February 1999 conference sponsored by Digital Government on IT issues in Federal statistics.

The ICSP, chaired by the Office of Management and Budget (OMB), maintains the FedStats Web site (*www.fedstats.gov*) to provide easy access to the full range of information produced by the statistical agencies for public use. The site has links to tools to access and view data from single agencies, and the multiagency Federal Electronic Research Review, Extraction and Tabulation Tool (FERRETT) that enables users to work with Census, BLS, and NCHS databases.

NGI applications

FISAC's NGI Applications Team (NGI/AT) works with the LSN HPNAT to make NGI applications opportunities available to the broad Federal community. NOAA, a non-NGI agency, has partnered with NGI agencies on several advanced networking applications from the beginning of the NGI program, and the USGS Biological Survey has begun participating in NGI discussions. Inquiries about extending NGI participation are welcome; currently under consideration are other Federal laboratories that have strong collaborations with academia such as the NOAA environmental and fisheries laboratories and data centers across the country, the USGS facilities in Menlo Park, California, and Golden, Colorado, and research facilities of the Department of Agriculture. FISAC NGI/AT agencies participated in the GOIN demonstrations and the "Bridging the Gap" workshop described in the LSN section of this report.

Universal access

The Universal Access Team focuses on research, development, and deployment of advanced technologies to eliminate technical barriers to broad

public participation in the digital environment. The Team helps agencies across the Government meet requirements that Federal services and employment opportunities be accessible to people with disabilities, and fosters partnerships among IT research institutions and agencies to design human-computer interface systems that improve accessibility. For example, the Team works on standards for validating IT performance in universally accessible systems in conjunction with the Underwriters Laboratory, in a project endorsed by the Enterprise Interoperability and Emerging IT Committee of the Government's Chief Information Officers Council, and on a new paradigm for interface protocols with the IT Accommodation Study Group of the National Committee for Information Technology Standards.

FY 2000 activities include:

- The Rapid Service Valuation and Preparation (RSVP) Access program, which brings agency experts and agency managers together to share resources, transfer best practices, and establish common standards for broadening the accessibility of IT services
- The Information Technology Testing for Accessibility Government-wide (IT-TAG) program, which is establishing an access performance validation capability with independent testing laboratories so that agencies can use the new Federal Acquisition Regulation procurement criteria
- Public Information Networks Need Accessible Collaborative Learning Environments (PINNACLE), a Web-based knowledge repository for learning and resource sharing that will enhance RSVP and IT-TAG, support virtual forums for Federal employees with disabilities, and allow IT service providers to discuss ways to improve the development pipeline for technologies that enable universal access

In FY 2001 the Universal Access Team will launch the Attuning to Disabled People in Society (ADP) program to rapidly scale up the Federal-state information infrastructure to comply with the accessibility requirements of Public Law 105-220, Section 508, and the Americans with Disabilities Act (ADA). ADP will ensure that Federal and state access standards and joint applications are interoperable; industry-led product development is focused and not fractured by disparate access standards; and users with disabilities can participate in developing Federal-state electronic services to prevent further costs to society attributable to exclusionary design.

DOE's ASCI Program

Overview

DOE's Accelerated Strategic Computing Initiative (ASCI) applies advanced capabilities in scientific and engineering computing to one of the most complex challenges in the nuclear era—maintaining the performance, safety, and reliability of the Nation's nuclear weapons without physical testing. A critical component of the agency's Stockpile Stewardship Program (SSP), ASCI research develops computational and simulation technologies to help scientists understand aging weapons, predict when components will have to be replaced, and evaluate the implications of changes in materials and fabrication processes for the design life of aging weapons systems. ASCI was established in FY 1996 in response to the Administration's commitment to pursue a comprehensive ban on nuclear weapons testing.

ASCI researchers are developing high end computing capabilities far above the current level of performance, and advanced simulation applications that can reduce the current reliance on empirical judgments by achieving higher resolution, higher fidelity, 3-D physics and full-system modeling capabilities for assessing the state of nuclear weapons. DOE has established an FY 2004 deadline for attaining working ASCI hardware and software and FY 2010 as the full implementation date for ASCI's R&D products.

ASCI activities center on development of the advanced applications software, more powerful computing platforms, and high end computational infrastructures needed to achieve four major objectives of the R&D effort:

- Performance—Create predictive simulations of nuclear weapons systems to analyze behavior and assess performance in an environment without nuclear testing
- Safety—Predict with high certainty the behavior of full weapons systems in complex accident scenarios
- Reliability—Achieve sufficient validated predictive simulations to extend the lifetime of the stockpile, predict failure mechanisms, and reduce routine maintenance
- Renewal—Use virtual prototyping and modeling to understand how new production processes and materials affect performance, safety, reliability, and aging. This understanding will help define the right configuration of production and testing facilities necessary for managing the stockpile throughout the next several decades.

DOE's three national Defense Program (DP) laboratories—LANL, LLNL, and SNL—collaborate on ASCI-related activities.

PathForward

The performance simulation and virtual prototyping applications required for the SSP call for far more powerful computing platforms than the industry now produces. ASCI's strategy, called PathForward, is to build the high end computing systems the program requires by scaling commercially viable building blocks—both hardware and software—to 30 teraops (30 trillion computing operations per second) and beyond. PathForward has established multiple partnerships with computer companies, government agencies, and academia to develop and accelerate technologies that are either not in the current business plans of manufacturers or not expected to be available in the timeframe or scale required by ASCI. For example, DOE and its three DP laboratories are partnering on a cost-sharing basis with Compaq/DEC, IBM, SGI, Cray, and Sun Microsystems to develop and engineer high-bandwidth, low-latency technologies to interconnect the 10,000 commodity processors needed to build a 30 teraops computer. In another PathForward partnership, DOE, DoD, and NASA are developing high performance storage technologies to reduce the physical size of ultra-scale data storage systems while significantly improving the speed at which data can be written into these systems. Such attributes are vital due to the massive storage requirements of the complex SSP simulations. The goal of this research is to develop optical tape drive technologies that can write 25 MBps (25 million bytes per second) to a 1 terabyte (1 trillion bytes)-capacity optical tape cartridge in a conventional sized unit.

ASCI computing platforms

A new IBM system dubbed "Baby Huey," the scalable prototype for a 10-teraops system that will be the world's fastest computing platform, has been installed at LLNL. Baby Huey consists of 16 IBM Nighthawk 1 nodes with a peak computing capacity of 114 Gflops and 32 GB of memory. It has been running ASCI applications to evaluate the performance of IBM's latest 64-bit computing technology in preparation for the scheduled full installation in summer 2000 of ASCI's 10-teraops "Option White" platform—a massively parallel system consisting of 512 multiprocessor nodes. It was just one 32-node RS6000 system (with specialized chess co-processors) named "Deep Blue" that defeated the world's leading chess champion in a highly publicized series of matches in 1997. Housing Option White requires 17,000 square feet of floor space and over 6.2 megawatts of electricity for power, cooling, and mechanical equipment—enough electricity to supply a small town with air conditioning. Option White is ASCI's fourth custom-built high-speed platform and the latest step toward the goal of having a 100-teraops system in place by 2004.

Visual Interactive Environment for Weapons Simulation (VIEWS)

ASCI's computing platforms enable DOE-supported scientists to store, retrieve, and manipulate complex data on a scale not possible on any other computing system. But researchers must develop equally advanced tools for organizing, managing, and visualizing the vast 3-D data sets representing the physical, chemical, and engineering properties of the nuclear weapons stockpile. ASCI applications will use extremely high fidelity 3-D models, on the order of one billion cells, to generate terabytes of raw data—a volume of information that would overwhelm scientists attempting to analyze it in the absence of tools that helped them manage it and "see" what it means. VIEWS integrates high

performance storage, high-speed networking, visualization hardware, and advanced data exploration and management software to provide the capabilities for high-level scientific data analyses.

The hardware infrastructure required to support this work includes a high performance scalable network of graphics workstations, visualization servers, and storage systems, all connected to the terascale computing platforms via high-speed interconnects. Through technologies such as video fiber modems, image compression and transmission, and hardware-parallel visualization systems, ASCI researchers can transmit real-time high-quality images from the visualization servers and supercomputers into offices and graphics labs. But the enormous sizes of ASCI data sets render existing visualization software ineffective for interactive data interrogation—that is, manipulating the data and visual representation for purposes of analysis in a distributed computing environment. Researchers are exploring parallel and demand-driven visualization strategies, including multiresolution and hierarchical techniques to moderate the data levels needed for visualization, and parallel and distributed algorithms to increase system capacity to meet scientists' visualization requirements.

Scientific data management (SDM)

VIEWS research is developing an organizational framework to speed and enhance a user's ability to browse and search the complex SSP data collections by integrating SDM-developed application software libraries and Web/Java-based components with commercial databases, mass storage systems, networking, and computing infrastructure. SDM efforts focus on capturing and sharing simulation data from application codes, organizing, searching, and managing a variety of data, and automating the computer-based data discovery process.

To have large-scale and diverse data sets flow smoothly among physics applications and between calculation and analysis, the data must be modeled in a machine- and application-neutral way. To meet this challenge, SDM researchers developed a common data model for ASCI simulation data based on principles from topology, along with a common API, and are working on metadata strategies to improve data organization and management. Such metadata range from documentation of the size, type, and creation date of a data set to a scientist's notes about the data. The researchers have developed a number of accompanying tools, including a calculation summary, a Web-accessible knowledge base of weapons data archives, a metadata editing and browsing capability, and Data Discovery—a suite of techniques and tools for automated querying, representation, and extraction of information from terascale simulation data sets.

Problem solving environment (PSE)

ASCI's problem solving environment (PSE) consists of the high end computational technologies and tools needed to conduct advanced scientific analyses on a secure, very high performance distributed computing system. PSE teams' work includes:

- ***ASCI distributed computing environment (DCE).*** The DCE team is establishing a common set of secure distributed computing services for use at each of its supercomputer sites, focusing on the middleware enabling desktop users to work smoothly across the heterogeneous ASCI network using heterogeneous computing and operating systems.

- ***High performance storage system (HPSS).*** More than 20 organizations representing academia, industry, and the Federal government (DOE, NASA, NOAA, and NSF) are collaborating to prototype a next-generation storage system based on commercially available products.
- ***Accelerated data transfer.*** Moving large amounts of data from one device to another is one of computing's most time-consuming tasks. ASCI researchers linked a new data-moving protocol to the HPSS parallel file transfer protocol, enabling dramatic improvements in data transfer rates.
- ***Scalable linear solvers.*** Because many currently used algorithms are not scalable, the computational workload for larger problems grows faster than the optimal linear rate. ASCI researchers are developing scalable algorithms that decrease computation time. For example, these algorithms can reduce a two-day run on a massively parallel processing machine to 30 minutes. Applications include studying complex physical phenomena, such as energy, the environment, and biological systems.
- ***Scalable input/output (I/O).*** This research aims to speed up the transfer of data through the various hardware and software components of a supercomputing system. The goal is an overall improvement in end-to-end performance, achieved by increasing data transfer speeds at every layer between an application and hardware.

Academic Strategic Alliances Program (ASAP)

By supporting ongoing technical interactions between the ASCI research and leading-edge academic R&D, ASAP accelerates new developments in simulation science and high performance technologies for computer modeling. In ASAP Level One, ASCI has established five major university centers to engage in long-term, large-scale unclassified research in simulation science and computational mathematics on advanced scientific problems. In Level Two, ASCI supports strategic investigations—smaller, discipline-oriented projects in computer science and computational mathematics critical to ASCI's success. ASAP Level Three supports individual collaborations focused on near-term ASCI research.

President's Information Technology Advisory Committee (PITAC)

Overview

On February 24, 1999, the PITAC delivered to the White House its report, "Information Technology Research: Investing in Our Future," a major examination of the Nation's IT research and development needs for the 21st century. Comprising corporate leaders and research scientists from business and academic institutions throughout the U.S., the 25-member Committee was established by President Clinton in February 1997 to provide expert guidance to the Federal government on maintaining America's preeminence in high performance computing and communications, information technology, and the Next Generation Internet. The PITAC was chartered by Congress under the High-Performance Computing Act of 1991 (P.L. 102-194) and the Next Generation Internet Act of 1998 (P.L. 105-305). In February 1999, the President issued an Executive Order extending the Committee's initial two-year mandate to February 2001.

The state of IT R&D: PITAC report

To prepare its report, the PITAC conducted an extensive review of Federal R&D in high performance computing, communications, and information technology, assessing current Federal research investments in light of growing demands for ever faster, more capable, and more robust technologies, and many more workers with the skills to develop and manage them. One indication of the need for significant R&D advances, the report observed, is the fact that the Federally-funded technologies underlying the Internet were designed to network a relatively small number of computers (2,000 as of 1985) but now connect some 70 million devices, with traffic doubling every 100 days.

Despite our manifest success in such advanced technologies—for example, production of semiconductors, computers, communications equipment, and software has accounted for a third of U.S. economic growth since 1992—the PITAC concluded that today Federal support for IT research is seriously inadequate. The report noted that investments in research programs are funding a fraction of what is needed and increasingly focus on short-term mission goals rather than long-term, high-risk activities. The PITAC report credited past Federal support of high-risk IT research for yielding "a spectacular return" in today's vibrant IT marketplace. But in a letter contained in the report, the PITAC's corporate leaders warned that this unprecedented engine of economic growth "could slow or disappear" without continued Federal support for innovation. "The government must increase its investment in the pipeline that generates ideas and the researchers to work on them," the industry officials wrote.

In view of these findings, the PITAC called for a visionary expansion of the Federal investment in IT research to restore and reinvigorate the flow of advanced innovations needed to fuel the information economy in the new century. Just as Federal research partnerships with academia and industry pioneered the concepts of digital computing machines and computer networking that have grown into today's worldwide communications infrastructure, the PITAC stated, national leadership again is necessary to drive "significant new research on computing and communication systems" ensuring that the cutting-edge hardware, software, and connectivity advances of the last quarter of the 20th century produce the next-generation breakthroughs of the 21st.

The PITAC report recommended that revitalized Federal support for information technology R&D begin with planned incremental increases in research funding in five areas of strategic importance to both Government and the private sector. The committee proposed raising the overall Federal commitment to IT research by $1.4 billion annually by FY 2004 in software, scalable information infrastructure, high end computing, human computer interface and information management, socioeconomic research and policy, and management of the Federal IT R&D enterprise.

The PITAC's key findings and recommendations are:

Findings and recommendations of the PITAC report to the President "Information Technology Research: Investing in our Future"

Key findings

- Total Federal information technology R&D investment is inadequate.
- Federal information technology R&D is excessively focused on near-term problems.
- The Federal information technology R&D funding profile is incomplete.

Key research recommendations

- Create a strategic initiative in long-term information technology R&D.
- Increase the investment for research in software, scalable information infrastructure, high end computing, and socioeconomic issues:
  - *Software.* The science and methods for efficiently creating and maintaining high quality software of all kinds, for ensuring the reliability of the complex software systems that now provide the infrastructure for much of our economy, for improving the interaction between people and computer-based systems and devices, and for managing and using information.
  - *Scalable information infrastructure.* Techniques for ensuring that the national information infrastructure—communications systems, the Internet, large data repositories, and other emerging systems—is reliable and secure, and can grow gracefully to accommodate the massive numbers of new users and applications expected over the coming two decades.

- *High end computing.* Continued invention and innovation in the development of fast, powerful computing systems and the accompanying communication systems needed to implement high end applications ranging from aircraft design to weather and climate modeling.
- *Socioeconomic issues.* Research on understanding the effects of information technology on our society, its economy, and the workforce should be funded. Furthermore, research should be focused on strategies for ameliorating information technology's potentially harmful effects and amplifying the benefits.

Key management recommendations

- Fund projects of larger scope and duration.
- Expand support for research carried out by teams of two to five researchers, possibly at different institutions, to address a single research project.
- Fund large centers for "Expeditions to the 21st Century," which would involve large teams of researchers in explorations of future information technologies and their societal effects.
- Establish "Enabling Technology Centers" to conduct research on the application of information technology to particular problems of national importance.

The full report is available at *http://www.itrd.gov/ac/report/*.

PITAC activities and initiatives

At the request of the President and/or as mandated by Congress, the PITAC periodically reviews aspects of the Government's IT R&D Programs and reports its findings. The committee also undertakes studies of information technology issues that are of significant national interest and develops reports on these subjects.

NGI reviews

In April 2000, the PITAC completed its second annual review of the Next Generation Internet (NGI) Initiative, as required by the NGI Act. The review covered advanced networking research, NGI testbeds, NGI applications, geographic reach, minority- and small-college reach, technology transfer, agency coordination, and IT leadership. In its report to Congress, the PITAC said the NGI agencies had logged significant achievements since the PITAC's April 1999 review, including:

- Implementation of network performance measurements. The new measurement systems found that between FY 1999 and FY 2000 NGI agencies had boosted maximum end-to-end performance on the NGI testbed networks from 80 Mbps to 900 Mbps.
- Increased high performance connectivity. The number of operational NGI sites grew from 154 to 177.

- Increased emphasis on developing end-user applications. The number of NGI applications rose from about 35 in the first PITAC review to more than 100.
- Expanded network reach. From the PITAC's first review, the number of NSF Experimental Program to Stimulate Competitive Research (EPSCoR) awards grew from 33 to 40, including two sites in Puerto Rico. Minority-serving institutions participating grew from one black and two Hispanic universities to two black and five Hispanic universities.

The PITAC noted that the agencies had implemented programmatic changes recommended in the first NGI review and had made "excellent progress," collaborating together and with academe and industry, toward NGI goals. The review found that about a dozen private-sector startup companies, capitalized at nearly $30 billion, had been launched from these NGI collaborations.

The review concluded with two recommendations:

- Funding for NGI should be extended through FY 2002 at the proposed funding levels, with planning for follow-on activities beginning immediately. More applications demonstrating the utility to end users of NGI's gigabit bandwidth, increased security, and enhanced quality of service should be funded.
- Since the NGI effort is not designed to address the reach issue, Congress should consider funding a separate infrastructure program in which NGI research institutions could serve as infrastructure mentors to nearby smaller or disadvantaged institutions.

IT R&D reviews

At the President's request, the PITAC last year reviewed the Administration's FY 2000 Information Technology for the Twenty-First Century, or IT^2, Initiative, and this year is reviewing the Administration's IT R&D budget proposal for FY 2002, which incorporates IT^2. The IT^2 Initiative, the FY 2000 review found, embodied the PITAC report's recommendations for more intensive research in software, scalable information infrastructure, and some aspects of high end computing. The review recommended increased emphasis on long-term research in advanced applications and middleware. It reiterated the PITAC report's call for increased funding for the IT R&D Programs as a whole and pointed to the special need for more high end facilities for the academic community.

In communications with the President and members of Congress, the PITAC urged continued bipartisan support for overall increases in IT R&D funding that "begin to refill the pipeline with ideas and human capital."

Digital Divide conference

In the 21st century, the PITAC report contended, it is essential that all Americans have access to the information infrastructure, along with the relevant tools and skills necessary to fully participate in the information age. IT tools and applications can provide opportunities that transcend barriers of race, gender, disability, age, income, and location. To pursue these ideas, the PITAC held an October 19, 1999, conference to explore the important issue of the "digital divide." "Resolving the Digital Divide: Information, Access, and Opportunity," hosted in association with the Joint Center for Political and Economic Studies and the Woodrow Wilson International Center for Scholars, focused on

information technology access for racial and ethnic groups in the U.S. The report on this conference, published in February 2000, makes the following recommendations:

- Resolving the digital divide demands a national initiative.
- Community relevance and community involvement are essential for solving the divide.
- Rethink educational approaches.
- Continue and expand Government programs and provide additional funding to resolve the digital divide.
- Rethink market approaches.
- More research, data collection, and evaluation are necessary to solve the digital divide.
- Better technology and more minority-owned businesses are necessary.

The PITAC is planning two additional conferences in the coming year on the issues of geographic disparities and small-university access to information tools.

New PITAC Co-Chairs

In August 1999, President Clinton appointed Raj Reddy of Carnegie Mellon University and Irving Wladawsky-Berger of IBM as the new Co-Chairs of PITAC. They succeeded Bill Joy of Sun Microsystems and Ken Kennedy of Rice University, who had been Co-Chairs for the two and a half years since the PITAC was established. In letters of appreciation, the President thanked the retiring Co-Chairs for their service to the Nation in developing the PITAC report and providing valuable counsel that helped shape the Government's IT R&D agenda.

IT challenges panels

In the fall of 1999, the PITAC established a group of panels to examine technological challenges to achieving the broad cultural transformations that the PITAC report called the critical opportunities of the information revolution. The report had identified 10 "National Challenge Transformations" in which information technology's promise can most benefit the Nation: the way we communicate, the way we deal with information, the way we learn, the practice of health care, the nature of commerce, the nature of work, the way we design and build things, the conduct of research, our understanding of the environment, and the activities of government. The panels, which include PITAC members and invited participants from academia and industry, will each develop a focused document detailing key IT research needs in a specific challenge area.

The PITAC panels are:

- Digital Divide Issues, chaired by Ching-chih Chen and John P. Miller
- Digital Libraries, chaired by David C. Nagel
- International Issues, chaired by Ching-chih Chen and David W. Dorman
- Open Source Software, chaired by Larry Smarr
- Transforming Government, chaired by David M. Cooper and Robert H. Ewald

- Transforming Health Care, chaired by Sherrilynne S. Fuller and Edward H. Shortliffe
- Transforming Learning, chaired by Susan Graham and Andrew J. Viterbi

In a December 1999 letter to the President describing these initiatives, the PITAC leaders said they planned to continue their emphasis on research to better understand the socioeconomic impact of the information technology revolution, paying particular attention to the growing "digital divide," the impact of globalization on the workforce, and key policy issues such as privacy, security, and intellectual property rights.

Committee membership

Committee Co-Chairs

Raj Reddy is Herbert A. Simon University Professor of Computer Science and Robotics at Carnegie Mellon University.

Irving Wladawsky-Berger is Vice President for Technology and Strategy of the Enterprise Systems Group at IBM Corporation.

Committee members

Eric A. Benhamou is CEO and Chairman of the 3Com Corporation.

Vinton Cerf is Senior Vice President for Internet Architecture and Engineering at MCI WorldCom.

Ching-chih Chen is Professor in the Graduate School of Library and Information Science at Simmons College.

David M. Cooper is Associate Director of Computation at the Lawrence Livermore National Laboratory.

Steven D. Dorfman is Retired Vice Chairman of the Hughes Electronics Corporation.

David W. Dorman is CEO of Concert.

Robert Ewald is President and CEO of the E-Stamp Corporation.

Sherrilynne S. Fuller is Head of the Division of Biomedical Informatics, Department of Medical Education at the University of Washington School of Medicine.

Hector Garcia-Molina is Leonard Bosack and Sandra Lerner Professor in the Departments of Computer Science and Electrical Engineering at Stanford University.

Susan L. Graham is Chancellor's Professor of Computer Science in the Department of Electrical Engineering and Computer Science at the University of California at Berkeley.

James N. Gray is a Senior Researcher in Microsoft's Scalable Servers Research Group and Manager of Microsoft's Bay Area Research Center.

W. Daniel Hillis is with Applied Minds, Inc.

Bill Joy is Founder and Vice President of Research at Sun Microsystems.

Robert E. Kahn is President of the Corporation for National Research Initiatives (CNRI).

Ken Kennedy is Director of the Center for Research on Parallel Computation and Ann and John Doerr Professor of Computer Science at Rice University.

John P. Miller is Director of the Center for Computational Biology and Professor of Biology at Montana State University.

David C. Nagel is President of AT&T Labs.

Edward H. Shortliffe is Professor and Chair of the Department of Medical Informatics at Columbia University's College of Physicians and Surgeons.

Larry Smarr is Professor of Computer Science and Engineering at the University of California-San Diego and Strategic Advisor to the National Computational Science Alliance.

Joe F. Thompson is William L. Giles Distinguished Professor of Aerospace Engineering in the Department of Aerospace Engineering at Mississippi State University.

Leslie Vadasz is Executive Vice President, Intel Corporation, and President, Intel Capital.

Andrew J. Viterbi is Vice Chairman and co-founder of QUALCOMM Incorporated.

Steven J. Wallach is Vice President of Chiaro Networks.

Change in membership

The PITAC thanks former member **David J. Farber,** Alfred Fitler Moore Professor of Telecommunication Systems (on leave) at the University of Pennsylvania, for his visionary contributions to the Committee. Dr. Farber resigned from the PITAC upon assuming the position of Chief Technologist at the Federal Communications Commission.

Coordination of IT R&D Programs

Office of Science and Technology Policy

The White House Office of Science and Technology Policy (OSTP) advises and supports the President in shaping broad national goals for Federal investments in science, space, and technology R&D and plays the senior role in coordinating the wide range of programs and activities in the Government's science and technology research portfolio. OSTP's responsibilities include policy formulation and budget development in science and technology areas, articulating the President's policies and programs, and fostering strong partnerships among Federal, state, and local governments and the scientific communities in industry and academia. OSTP oversees the activities of the National Science and Technology Council (NSTC). The Assistant to the President for Science and Technology serves as Director of OSTP.

Senior Principals Group for Information Technology (IT) R&D

This senior management team, chaired by and reporting to the Director of OSTP, provides policy leadership, guidance, and oversight for the Government's Information Technology (IT) R&D Programs. The Senior Principals Group and the NSTC help set programmatic priorities, ensure that the overall Federal IT R&D portfolio is balanced across agency missions and capabilities, and monitor progress. Members are senior officials from the IT funding agencies, including the Director of the National Science Foundation, the Administrator of the National Aeronautics and Space Administration, the Under Secretary of Energy, the Under Secretary of Commerce for Oceans and Atmosphere and Administrator, National Oceanic and Atmospheric Administration, the Director of the National Institutes of Health, the Deputy Under Secretary of Defense (Science and Technology), and senior officials from the Office of Management and Budget (OMB) and the National Economic Council (NEC).

Interagency Working Group on IT R&D

The Interagency Working Group on IT R&D (IWG/IT R&D, successor to the Subcommittee on Computing, Information, and Communications [CIC]) serves as the internal deliberative organization of the Senior Principals Group, providing policy, program, and budget guidance for the Executive Branch. The IWG coordinates the multiagency IT research efforts, including planning, budgeting, implementation, and review of research activities and interaction with the Congress, OMB, OSTP, academia, and industry. The IWG also provides technical assistance to the President's Information Technology Advisory Committee (PITAC) and works with Federal agencies that need advanced IT to identify their needs and accelerate development and deployment of appropriate technologies.

The NSF Assistant Director, Directorate for Computer and Information Science and Engineering (CISE), chairs the IWG. Agencies in the IWG are AHRQ, DARPA, DOE (Offices of Science and Defense Programs), EPA, NASA, NIH, NIST, NOAA, NSA, NSF (CISE and Mathematical and Physical Sciences Directorates), and OSD. Other participants include NEC, NCO, OMB, and OSTP. The NCO provides technical and administrative support to the IWG and select support to the IWG's Coordinating Groups.

Coordinating Groups

Six Coordinating Groups (previously called Working Groups) and one Council report to the IWG. These organizations confer regularly to coordinate the objectives and activities of the multiagency projects in their specialized domains. The Coordinating Groups are aligned with the major areas of IT research, called Program Component Areas (PCAs), that make up the Government's multiagency IT R&D portfolio, as follows:

- High End Computing and Computation Coordinating Group (HECCCG), which coordinates both HEC I&A and HEC R&D
- Human Computer Interface and Information Management Coordinating Group (HCI & IMCG)
- Large Scale Networking Coordinating Group (LSNCG), which includes the Next Generation Internet (NGI) Initiative and Scalable Information Infrastructure (SII) R&D and divides its activities among several teams:
  - High Performance Networking Applications Team (HPNAT)
  - Information Security Team (IST)
  - Joint Engineering Team (JET)
  - Networking Research Team (NRT)
- Software Design and Productivity Coordinating Group (SDPCG)
- High Confidence Software and Systems Coordinating Group (HCSSCG)
- Social, Economic, and Workforce Implications of IT and IT Workforce Development Coordinating Group (SEWCG)

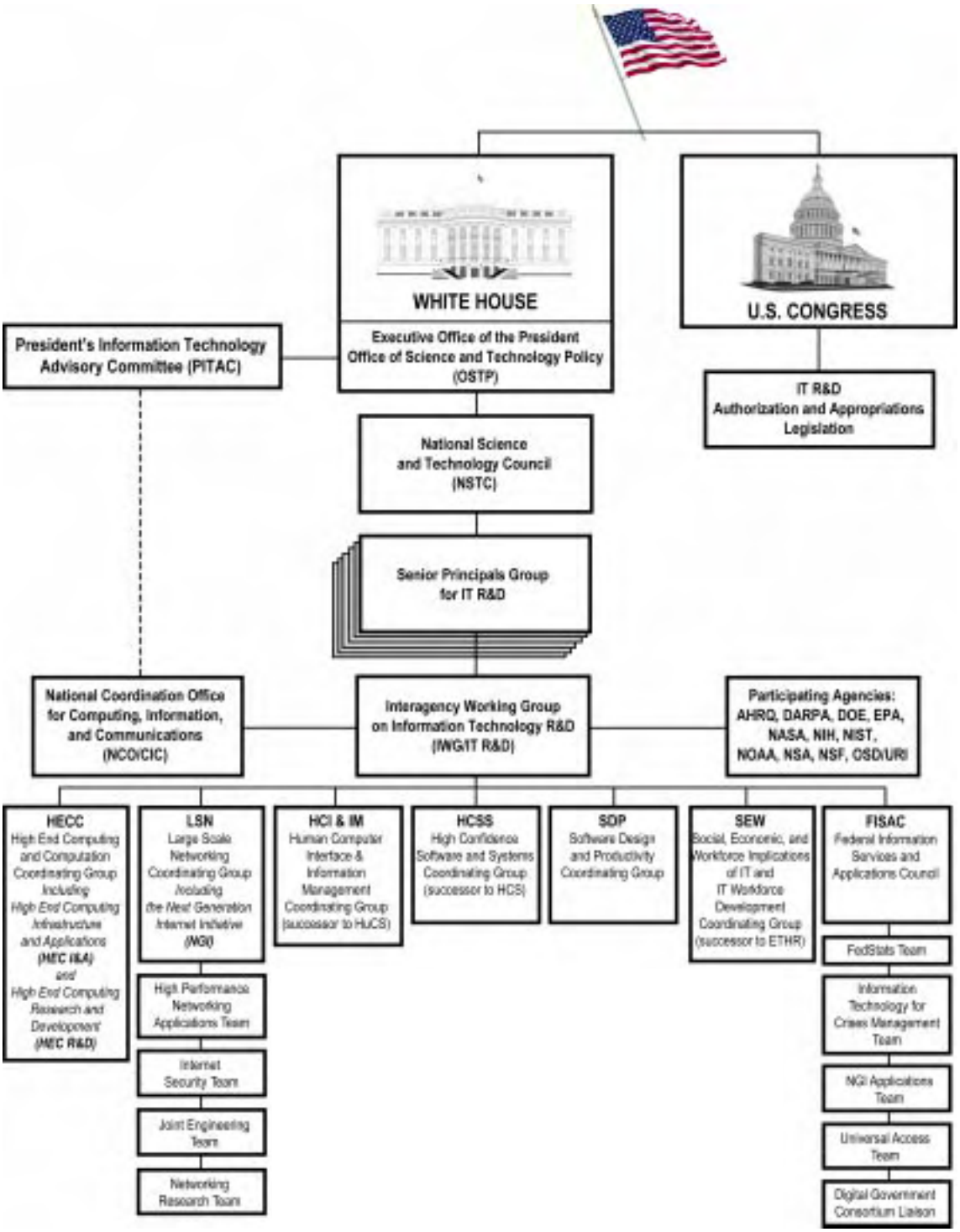
The Federal Information Services and Applications Council (FISAC), an outreach rather than a research organization, also reports to the IWG.

FY 2000 accomplishments and FY 2001 plans for the Coordinating Groups are highlighted in this book.

National Coordination Office for Computing, Information, and Communications (NCO/CIC)

The NCO supports the overall IT R&D Programs. Working collaboratively, the NCO and the agencies participating in the IT R&D Programs craft the blueprints and implementation strategies for the comprehensive Federal effort to pioneer the next generation of advanced computing, network communications, and information technologies. The NCO provides extensive technical and administrative support to the IWG, assisting in the preparation of multiagency planning, budget, and evaluation materials, and serves as a central source of information and documentation about Federal IT research activities.

In addition, the NCO supports the PITAC, the Presidentially appointed group of leaders in industry and academe that conducts independent evaluations


WHITE HOUSE
U.S. CONGRESS
President's Information Technology Advisory Committee (PITAC)
Executive Office of the President Office of Science and Technology Policy (OSTP)
IT R&D Authorization and Appropriations Legislation
National Science and Technology Council (NSTC)
Senior Principals Group for IT R&D
National Coordination Office for Computing, Information, and Communications (NCO/CIC)
Interagency Working Group on Information Technology R&D (IWG/IT R&D)
Participating Agencies: AHRQ, DARPA, DOE, EPA, NASA, NIH, NIST, NOAA, NSA, NSF, OSD/URI
HECC High End Computing and Computation Coordinating Group Including High End Computing Infrastructure and Applications (HEC I&A) and High End Computing Research and Development (HEC R&D)
LSN Large Scale Networking Coordinating Group Including the Next Generation Internet Initiative (NGI)
High Performance Networking Applications Team
Internet Security Team
Joint Engineering Team
Networking Research Team
HCI & IM Human Computer Interface & Information Management Coordinating Group (successor to HuCS)
HCSS High Confidence Software and Systems Coordinating Group (successor to HCS)
SDP Software Design and Productivity Coordinating Group
SEW Social, Economic, and Workforce Implications of IT and IT Workforce Development Coordinating Group (successor to ETHR)
FISAC Federal Information Services and Applications Council
FedStats Team
Information Technology for Crises Management Team
NGI Applications Team
Universal Access Team
Digital Government Consortium Liaison

of the Federal government's IT research activities. NCO staff members provide technical assistance for meetings of the PITAC and its subcommittees and panels, as well as for the PITAC's ongoing reviews of the Federal IT R&D Programs.

Outreach and presentations

As the central points of contact for Federal IT R&D initiatives, the NCO and the IWG meet often with representatives from Congress, Federal, state, and local organizations, academia, industry, professional societies, foreign organizations, and others to discuss the Government's IT R&D Programs and exchange technical and programmatic information.

At SC99, the national high performance networking and computing conference, for example, the NCO and the LSNCG demonstrated 10 advanced networking and computing applications developed by agencies participating in the IT R&D Programs and its NGI initiative (page 99).

The NCO each year also fulfills thousands of requests for information about Federal IT R&D activities from Congressional offices, academia, industry, and the public. Information resources provided by the NCO include print and video materials and Web sites constituting a full-text archive of all NCO publications since 1994, such as information on the IWG (and its predecessor Subcommittee on CIC R&D), details of the NGI initiative, Congressional testimony, reports and activities of the PITAC, links to other agencies in the IT R&D Programs, and important notices and official correspondence. The online resources are available at *http://www.itrd.gov/* and *http://www.ngi.gov/*.

High end computing and mass storage systems briefings

Continuing a program of targeted briefings to update Federal agency scientists, engineers, and contractors on trends in key advanced information technologies, the NCO and the HECCCG held a two-day meeting on mass storage systems June 28-29, 2000, at the National Library of Medicine's Lister Hill Auditorium in Bethesda, Maryland. The gathering featured presentations on a non-disclosure basis by vendors of digital storage technologies and discussions of major technical challenges posed by the fast-growing storage requirements of IT systems. Questions about next-generation storage technologies included, for example: Are tape backup systems becoming obsolete? What are the prospects for the linked storage units called storage area networks (SANs)? What implications for storage needs do emerging high performance computational and telecommunications applications present? And what impact does the great heterogeneity of contemporary computing environments have on mass storage design?

Buy American Report

Congress requires information concerning non-U.S. high performance computing and communications funding activities. In FY 2000, DARPA was the only IT R&D agency that entered into grants, contracts, cooperative agreements, or cooperative research and development agreements for IT R&D with either 1) a company other than a company that is either incorporated or located in the U.S. and that has majority ownership by individuals who are citizens of the U.S., or 2) an educational institution or nonprofit institution located outside the U.S. DARPA funded IT R&D-related awards of $164,000 to University College, London (UK) and $98,040 to Oxford University (UK). In FY 2000, no IT R&D procurement exceeds $1 million for unmanufactured articles, materials, or supplies mined or produced outside the U.S., or for manufactured articles,

materials, or supplies other than those manufactured in the U.S. substantially all from articles, materials, or supplies mined, produced, or manufactured in the U.S.

Agency IT R&D Budgets by Program Component Area

FY 2000 Budget (dollars in millions)

Agency	High End Computing Infrastructure and Applications (HEC I&A)	High End Computing Research and Development (HEC R&D)	Human Computer Interface and Information Management (HCI & IM)	Large Scale Networking (LSN)[a]	Software Design and Productivity (SDP)	High Confidence Software and Systems (HCSS)	Social, Economic, and Workforce (SEW)	Totals[b]
NSF	206.0	83.8	91.6	81.2	15.8	9.6	28.6	517
DARPA	20.3	16.2	40.3	69.9	30.0	17.9	0.0	195
NASA	99.4	25.4	5.5	20.1	10.0	6.5	6.7	174
NIH	30.9	3.2	71.8	63.0	0.6	6.2	6.9	183
DOE Office of Science	65.1	19.0	8.0	24.8	0.0	0.0	2.6	120
NSA	0.0	31.7	0.0	1.7	0.0	47.3	0.0	81
NIST	3.5	0.0	6.2	5.2	0.0	3.5	0.0	18
NOAA	13.2	0.0	0.5	2.7	1.4	0.0	0.0	18
AHRQ	0.0	0.0	4.1	3.7	0.0	0.0	0.0	8
OSD/URI	0.0	2.0	2.0	4.0	1.0	1.0	0.0	10
EPA	3.9	0.0	0.0	0.0	0.0	0.0	0.0	4
Subtotals	442.3	181.3	230.0	276.3	58.8	92.0	44.8	1,328
DOE ASCI[c]	86.6	27.3	0.0	23.5	39.9	0.0	40.2	218
Totals	528.9	208.6	230.0	299.8	98.7	92.0	85.0	1,546

Notes:

[a] The NGI budget is a part of the LSN budget.

[b] Subtotals do not add exactly to total due to roundoff.

[c] These totals vary from the President's FY 2001 Budget.

Agency IT R&D Budgets by Program Component Area

FY 2001 Budget Request (dollars in millions)

Agency	High End Computing Infrastructure and Applications (HEC I&A)	High End Computing Research and Development (HEC R&D)	Human Computer Interface and Information Management (HCI & IM)	Large Scale Networking (LSN)[a]	Software Design and Productivity (SDP)	High Confidence Software and Systems (HCSS)	Social, Economic, and Workforce (SEW)	Totals[b]
NSF	285.2	102.1	135.8	111.2	39.5	20.5	45.3	740
DARPA	54.6	56.5	48.0	85.3	55.0	8.0	0.0	307
NASA	129.1	25.8	17.9	19.5	20.0	9.1	8.3	230
NIH	34.5	3.4	99.6	65.6	0.7	6.5	7.0	217
DOE Office of Science	106.0	30.5	16.6	32.0	0.0	0.0	4.6	190
NSA	0.0	32.9	0.0	1.9	0.0	44.7	0.0	80
NIST	3.5	0.0	6.2	4.2	2.0	8.5	0.0	24
NOAA	13.3	1.8	0.5	2.7	1.5	0.0	0.0	20
AHRQ	0.0	0.0	8.1	7.4	0.0	0.0	0.0	16
OSD/URI	0.0	2.0	2.0	4.0	1.0	1.0	0.0	10
EPA	3.6	0.0	0.0	0.0	0.6	0.0	0.0	4
Subtotals	629.8	254.9	334.7	333.8	120.3	98.3	65.2	1,838
DOE ASCI[c]	131.8	36.5	0.0	35.0	40.2	0.0	55.7	299
Totals	761.6	291.4	334.7	368.8	160.5	98.3	120.9	2,137

Notes:

[a] The NGI budget is a part of the LSN budget.

[b] Subtotals do not add exactly to total due to roundoff.

[c] These totals vary from the President's FY 2001 Budget.

IT R&D Summary

IT R&D Goals

Assure continued U.S. leadership in computing, information, and communications technologies to meet Federal goals and to support U.S. 21st century academic, defense, and industrial interests

Accelerate deployment of advanced and experimental information technologies to maintain world leadership in science, engineering, and mathematics; improve the quality of life; promote long-term economic growth; increase lifelong learning; protect the environment; harness information technology; and enhance national security

Advance U.S. productivity and industrial competitiveness through long-term scientific and engineering research in computing, information, and communications technologies

IT R&D Agencies

AHRQ	*Agency for Healthcare Research and Quality, Department of Health and Human Services*
DARPA	*Defense Advanced Research Projects Agency, Department of Defense*
DOE Office of Science	*Department of Energy-Office of Science*
EPA	*Environmental Protection Agency*
NASA	*National Aeronautics and Space Administration*
NIH	*National Institutes of Health, Department of Health and Human Services*
NIST	*National Institute of Standards and Technology, Department of Commerce*
NOAA	*National Oceanic and Atmospheric Administration, Department of Commerce*
NSA	*National Security Agency, Department of Defense*
NSF	*National Science Foundation*
OSD/URI	*Office of the Secretary of Defense's University Research Initiative*

Evaluation Criteria for IT R&D Programs

Relevance/Contribution

The research must significantly contribute to the overall goals of the Federal Information Technology Research and Development (IT R&D) program, which include the goals of the seven Program Component Areas—High End Computing Infrastructure and Applications (HEC I&A), High End Computing Research and Development (HEC R&D), Human Computer Interface and Information Management (HCI & IM), Large Scale Networking (LSN), Software Design and Productivity (SDP), High Confidence Software and Systems (HCSS), and Social, Economic, and Workforce Implications of Information Technology and Information Technology Workforce Development (SEW)—to enable solution of Grand Challenge- and National Challenge-class applications problems.

Technical/Scientific Merit

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

Readiness

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

Timeliness

The proposed work must be technically/scientifically timely for one or more of the IT R&D Program Component Areas.

Linkages

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

Costs

The identified resources must be adequate, represent an appropriate share of the total available IT R&D resources (e.g., a balance among Program Component Areas), promote prospects for coordinated or joint funding, and address long-term resource implications.

Agency Approval

The proposed program or activity must have policy-level approval by the submitting agency.

Glossary

2-D
Two-dimensional.

3-D
Three-dimensional.

4-D
Four-dimensional.

Abilene
Advanced backbone network developed by Qwest, Cisco, Nortel, and Indiana University to support the University Corporation for Advanced Internet Development's Internet2 project.

ACCESS DC
The National Computational Science Alliance's Center for Collaboration, Education, Science and Software located in Arlington, Virginia.

Access Grid
The National Computational Science Alliance's hardware, software, and telecommunications capabilities for distributed research collaboration and teleconferencing.

aCe
NASA's architecture adaptive computing environment.

ACM
Association for Computing Machinery.

ACTS
DOE's advanced computational testing and simulation program.

ADEPT
Alexandria Digital Earth Prototype.

AES
Advanced encryption standard.

AFM
Atomic force microscopy.

AHCPR
Agency for Health Care Policy and Research. Renamed the Agency for Healthcare Research and Quality (AHRQ) in FY 2000.

AHRQ
Agency for Healthcare Research and Quality, part of the Public Health Service of the Department of Health and Human Services. Formerly the Agency for Health Care Policy and Research (AHCPR).

Akenti
A security model and architecture developed by LBNL.

Algorithm
A procedure designed to solve a problem. Scientific computing programs implement algorithms.

Alliance
The National Computational Science Alliance, one of NSF's two PACI partnerships.

AMP
Advanced measurement platform.

ANI
NSF Advanced Networking Infrastructure program.

ANIR
Advanced Networking Infrastructure and Research Division, part of NSF/CISE.

ANL
DOE's Argonne National Laboratory.

APAN
Asian-Pacific Advanced Network.

API
Applications program interface.

ARC
NASA's Ames Research Center.

ARIN
American Registry for Internet Numbers.

ARPS
Advanced Regional Prediction System.

ASAP
ASCI's Academic Strategic Alliances Program.

ASCI
DOE's Accelerated Strategic Computing Initiative.

ATDNet
Advanced Technology Demonstration Network.

ATM
Asynchronous transfer mode, a telecommunications technology, also known as cell switching, which is based on 53-byte cells.

Backbone network
A high capacity electronic trunk—for example, the NSF vBNS backbone—connecting lower capacity networks.

Bandwidth
A measure of the capacity of a communications channel to transmit information; for example, millions of bits per second or Mb/s.

Benchmark
A point of reference (artifact) to compare an aspect of systems performance (for example, a well-known set of programs). Also, to conduct and assess the computation or transmission capabilities of a system using a well-known artifact.

Beowulf
NASA-funded high performance operating system for networked workstations.

BGMP
Border gateway multicast protocol.

Bit
Binary digit.

BLS
Bureau of Labor Statistics.

BMC
Biomolecular computation.

BioCoRE
Biological Collaborative Research Environment.

BOSSNET
A Boston-area network providing SuperNet access. Supported by DARPA and a consortium of local institutions.

bps, or b/s
Bits per second.

Bps, or B/s
Bytes per second.

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

C
C programming language.

C++
C++ programming language, an object-oriented descendant of the C language.

CA*Net
Canada's high performance network.

CAD
Computer-aided design.

CAIRN
Collaborative Advanced Interagency Research Network.

CalREN-2
California Research and Education Network.

CalTech
California Institute of Technology.

CAS
NASA's Computational Aerosciences project.

CAVE
Cave Automatic Virtual Environment. A surround-screen, surround-sound, projection-based virtual reality system.

CD
Compact disc.

CDL
California Digital Library.

CDMA
Code-Division Multiple Access, a digital cellular technology that uses the available spectrum of frequencies by digitally encoding each individual conversation.

CERHAS
Center for Electronic Reconstruction of Historical and Archaeological Sites.

CERN
European Laboratory for Particle Physics.

CHARMM
Chemistry at Harvard Molecular Mechanics.

CIC
Computing, Information, and Communications.

CIO
Chief Information Officer.

CISE
NSF's Directorate for Computer and Information Science and Engineering.

CMU
Carnegie Mellon University.

Computational Grid
The National Computational Science Alliance's network of high end parallel computing systems.

CONQUEST
AHRQ's COmputerized Needs-oriented QUality measurement Evaluation SysTem.

Coral
A network performance measurement tool.

CORBA
Common object request broker architecture.

CoS
Class of service.

COTS
Commercial off-the-shelf. Describes hardware and software that are readily available commercially.

CSTB
Computer Science and Telecommunications Board, part of the National Research Council.

DARPA
Defense Advanced Research Projects Agency, part of DoD. Formerly ARPA.

DASADA
DARPA's dynamic assembly for systems adaptability, dependability, and assurance program.

DCS
DARPA's Digital Collaboration Services.

DE
Digital Earth, an interagency geo-referencing framework project.

DeepView
DOE's scalable software system for collaborative scientific microscopy.

DES
Data encryption standard.

DHHS
Department of Health and Human Services.

DiffServ
Differentiated services. Also a standard being developed by the IETF to implement differentiated services across provider networks.

DL
Digital Libraries.

DLI
Digital Libraries Initiative.

DLMF
NIST's Digital Library of Mathematical Functions.

DNA
Deoxyribonucleic acid, a biomolecule from which genes are composed.

DNS
Domain Name System, a core element of Internet architecture.

DoD
Department of Defense.

DOE
Department of Energy.

DoS
Denial of service, a form of network attack in which multiple messages overwhelm a system's capacity to process them.

DPSS
DARPA's distributed-parallel storage system.

DREN
DoD's Defense Research and Engineering Network.

DSM
Distributed shared memory.

DSMC
Direct simulation-Monte Carlo.

DWDM
Dense wave division multiplexing.

DVC
Data visualization corridor.

DVD
Digital video disc.

EIA
DOE's Energy Information Administration.

EMERGE
DOE's ESnet/MREN Regional Grid Experimental testbed.

Emerge
NCSA's prototype software for retrieving information from many databases with a single query.

EOT-PACI
The joint education, outreach, and training programs of NSF's Partnerships for Advanced Computational Infrastructure (PACI).

EPA
Environmental Protection Agency.

EPSCoR
NSF's Experimental Program to Stimulate Competitive Research.

EQC
Ensemble quantum computer.

EROS
U.S. Geological Survey's Earth Resources Observation Systems.

ESnet
DOE's Energy Sciences Network.

ESS
NASA's Earth and Space Sciences program.

ETHR
Education, Training, and Human Resources, formerly one of the IT R&D PCAs. Predecessor to the SEW PCA.

EVL
Electronic visualization laboratory.

Exa-
A prefix denoting 10^{18}, or a million trillion; for example, exabytes.

Extensible Markup Language
See XML.

FAA
Federal Aviation Administration.

FBI
Federal Bureau of Investigation.

FedNets
Federal agency networks.

FedStats
Federal Statistics.

FEMA
Federal Emergency Management Agency.

FGDC
Federal Geographic Data Committee.

FIPS
Federal Information Processing Standard.

FISAC
Federal Information Services and Applications Council, which reports to the IWG/IT R&D.

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

FM
Formal methods.

FY
Fiscal Year.

G, or giga-
A prefix denoting 10^{9}, or a billion; for example, Gflops or gigaflops, gigabytes, gigabits.

Gb
Gigabit.

GB
Gigabyte.

GbE
Gigabit Ethernet.

Gbps
Gigabits per second.

GBps, or GB/s
Gigabytes per second.

GEO
Geosynchronous Earth Orbiting, referring to a satellite.

GFDL
NOAA's Geophysical Fluid Dynamics Laboratory.

Gflops
Gigaflops, billions of floating point operations per second.

Gigapops
Gigabit per second points of presence.

GIS
Geographic information system.

Globus
A joint project of ANL and USC to build distributed computational grids.

GOIN
Global Observation Information Network.

GPS
Global Positioning System.

GSA
General Services Administration.

GSFC
NASA Goddard Space Flight Center.

GSM
NOAA's global spectral model.

GSM
Global System for Mobile Communication, a European and Asian cellular communications system based on the TDMA standard.

GUI
Graphical user interface.

GUSTO
Globus Ubiquitous Supercomputing Testbed Organization.

HACP
NSA's high assurance computing platform.

Haptic
Describes devices that simulate the sensations of touching and grasping.

HAVPN
High assurance virtual private networking.

HAZMAT
NOAA's Hazardous Materials Response and Assessment Division.

HAZUS
Hazard loss estimation methodology.

HCI & IM
Human Computer Interface and Information Management, one of the IT R&D PCAs. Successor to the Human Centered Systems (HuCS) PCA.

HCI & IMCG
Human Computer Interface and Information Management Coordinating Group. Successor to the HuCS Working Group (HuCSWG).

HCSS
High Confidence Software and Systems, one of the IT R&D PCAs. Successor to the High Confidence Systems (HCS) PCA.

HCSSCG
High Confidence Software and Systems Coordinating Group. Successor to the HCS Working Group (HCSWG).

HECC
High End Computing and Computation. Previously one of the IT R&D PCAs, but now subdivided into the High End Computing Infrastructure and Applications (HEC I&A) and the High End Computing Research and Development (HEC R&D) PCAs.

HECCCG
High End Computing and Computation Coordinating Group, which coordinates the HEC I&A and HEC R&D PCAs. Formerly High End Computing and Computation Working Group (HECCWG).

HEC I&A
High End Computing Infrastructure and Applications PCA.

HEC R&D
High End Computing Research and Development PCA.

Heterogeneous system
A system that contains computing systems, networks, or software from more than one vendor.

HFC
Hybrid-fiber coax, a type of cable connection between end users and commercial communications networks.

HiPPI
High performance parallel interface.

HP
Hewlett-Packard Company.

HPC
High performance computing.

HPCC
High Performance Computing and Communications. The Federal R&D effort launched by the High-Performance Computing Act of 1991. Immediate predecessor of the CIC R&D programs, which were succeeded by the current IT R&D Programs.

HPCMP
DoD's High Performance Computing Modernization Program.

HPNAT
High Performance Networking Applications Team, which reports to the LSNCG.

HPNSP
High performance network service provider.

HPSS
High performance storage system.

HSCC
High Speed Connectivity Consortium, a SuperNet testbed.

HTMT
Hybrid technology multithreaded, an experimental supercomputing architecture.

HuCS
Human Centered Systems, formerly one of the IT R&D PCAs. Predecessor of the HCI & IM PCA.

IA
Information assurance.

I&A
Identification and authentication.

IBM
International Business Machines Corporation.

ICSP
Interagency Council on Statistical Policy.

IEEE
The Institute of Electrical and Electronics Engineers, Inc.

IETF
Internet Engineering Task Force.

iGrid
International Grid.

IKE
Internet key exchange protocol.

INFOSEC
Information security.

Internet
The global collection of interconnected, multiprotocol computer communications networks including Federal, private, and international networks.

I2
Internet2.

I/O
Input/output.

IP
Internet Protocol.

IPv6
Internet Protocol, version 6. The first IP release to use 128-bit addressing—greatly expanding the possible total number of Internet addresses—and to include such NGI features as built-in security, multicast, dynamic configurability, and mobility.

IPsec
IP security protocol.

IPsec WIT
IPsec Web-based interoperability tester.

ISAKMP
Internet Security Association Key Management Protocol.

IST
Internet Security Team, which reports to the LSNCG.

IT
Information technology.

IT²
"Information Technology for the Twenty-First Century," an FY 2000 Presidential initiative to increase the Government's IT R&D investment.

ITR
NSF's Information Technology Research program.

ITCM
FISAC's Information Technology for Crises Management Team.

IWG/IT R&D, or IWG
Interagency Working Group on Information Technology R&D. Successor to the Subcommittee on CIC R&D.

Java
An operating system-independent programming language. Initially designed for use on the Internet as a simpler, object-oriented alternative to C++, Java can be employed in coding small interactive application modules (applets) used in Web pages.

JET
Joint Engineering Team, which reports to the LSNCG.

JPL
NASA's Jet Propulsion Laboratory.

K, or Kilo-
A prefix denoting 10^3, or a thousand; for example, kilobits.

Kbps
Kilobits per second or thousands of bits per second.

KCS
NSF's knowledge and cognitive systems program.

KeLP
Kernel lattice parallelism, a framework for implementing portable scientific applications on distributed memory parallel computers.

Kernel
The core software of an operating system.

KMI
Key management infrastructure.

LAN
Local area network.

LANL
DOE's Los Alamos National Laboratory.

LAS
NOAA's Live Access Server.

LBNL
DOE's Lawrence Berkeley National Laboratory.

Legion
An object-based metasystem software project.

Linux
An open source operating system.

LLNL
DOE's Lawrence Livermore National Laboratory.

LSN
Large Scale Networking, one of the IT R&D PCAs.

LSNCG
Large Scale Networking Coordinating Group. Formerly the Large Scale Networking Working Group (LSNWG).

M, or Mega-
A prefix denoting 10^6, or a million; for example, Mbps, or megabits per second.

Mb
Megabit.

MB
Megabyte.

Mbps, or Mb/s
Megabits per second or millions of bits per second.

MBps, or MB/s
Megabytes per second or millions of bytes per second.

MCell
NPACI's cellular microphysiology simulation project.

MCM
Multichip module.

MCNC
An advanced electronics, telecommunications, and high performance computing laboratory in North Carolina.

MEII
DoD's minimum essential information infrastructure.

Metadata
Information derived from and describing data, such as the author, creation date, format, and modifications to original.

Mflops
Megaflops, millions of floating point operations per second.

MirNET
U.S.-Russian network consortium.

MIT
Massachusetts Institute of Technology.

MONET
Multiwavelength optical networking.

Monte Carlo
A simulation technique using a pseudo-random number generator to develop multiple independent trials simulating a range and probability distribution of results.

MPI
Message passing interface.

MPICH
A multiplatform configurable implementation of MPI.

MPI-I/O
Message passing interface-input/output.

MPLS
Multiprotocol label switching.

MPP
Massively parallel processors.

MRI
Magnetic resonance imaging.

MTA
Multithreaded architecture.

M-VIA
An implementation of virtual interface architecture (VIA) for Linux.

MVICH
A DOE/NERSC project to provide portable high performance communications for cluster computing.

nano-
A billionth (10^{-9}). A nanosecond is one-billionth of a second.

NASA
National Aeronautics and Space Administration.

NC
Nanocrystal.

NCA
Nanocrystal array.

NCAR
National Center for Atmospheric Research.

NCDC
NOAA's National Climatic Data Center.

NCHS
National Center for Health Statistics, part of NIH.

NCI
National Cancer Institute, part of NIH.

NCO/CIC or NCO
National Coordination Office for Computing, Information, and Communications.

NCRR
National Center for Research Resources, part of NIH.

NCSA
National Center for Supercomputing Applications at the University of Illinois, Urbana-Champaign. Headquarters of the National Computational Science Alliance, part of NSF's PACI.

NERSC
DOE's National Energy Research Supercomputer Center.

NESC
EPA's National Environmental Supercomputing Center.

Netlogger
DOE tool providing real-time analysis of network performance.

NGI
Next Generation Internet, a Presidential initiative that is part of the IT R&D Programs.

NGIX
Next Generation Internet Exchange Point.

NGC
AHRQ's National Guideline Clearinghouse.

NIAP
NIST/NSA National Information Assurance Partnership.

NIH
National Institutes of Health, part of DHHS.

NIST
National Institute of Standards and Technology, part of the Department of Commerce.

NLANR
National Laboratory for Applied Networking Research, sponsored by NSF.

NLM
National Library of Medicine, part of NIH.

NMR computing
Nuclear magnetic resonance spectroscopy-based computing.

NOAA
National Oceanic and Atmospheric Administration, part of the Department of Commerce.

NORDUnet
A network spanning the Nordic countries.

NPACI
National Partnership for Advanced Computational Infrastructure. Part of NSF's PACI.

NPSS
Numerical propulsion system simulation software.

NREN
NASA Research and Education Network.

NRL
Naval Research Laboratory, part of DoD.

NRT
Networking Research Team, which reports to the LSNCG.

NSA
National Security Agency, part of DoD.

NSF
National Science Foundation.

NSFNET
An NSF computer network program, predecessor to vBNS.

NSTC
The Presidential National Science and Technology Council.

NTON-II
DARPA's National Transparent Optical Network on the West Coast.

NYSERNet
New York State Education and Research Network.

OC
Optical carrier. Typically used with a speed designation. OC-1 is 51.85 Mbps and all other OC-x are multiples of this speed, e.g., OC-3 = 155.52 Mbps.

OCXmon
OCx monitor. A generic name for non-invasive performance measurement for OC network links, e.g., OC-12mon, OC-48mon.

OMB
White House Office of Management and Budget.

ONR
Office of Naval Research, part of DoD.

ONRAMP
A Boston-area SuperNet testbed.

OOF
NIST's object-oriented finite element software.

OOMMF
NIST's object-oriented micromagnetic modeling framework.

OpenMP
Open message passing.

Option White
DOE's new 10-teraops computing platform at LLNL.

OS
Operating system.

OSD
Office of the Secretary of Defense.

OSTP
White House Office of Science and Technology Policy.

P/NWGP
Pacific/Northwest GigaPop.

PACI
NSF's Partnerships for Advanced Computational Infrastructure, comprising the National Computational Science Alliance (Alliance) and the National Partnership for Advanced Computational Infrastructure (NPACI).

PathForward
DOE's ASCI program of partnerships with industry, academia, and other Federal agencies to accelerate development of technologies needed for ultra-scale computing systems.

PCA
Program Component Area. The IT R&D Programs is organized into PCAs: High End Computing and Computation (HECC, subdivided into HEC I&A and HEC R&D); Human Computer Interface and Information Management (HCI & IM); Large Scale Networking (LSN, including NGI); Software Design and Productivity (SDP); High Confidence Software and Systems (HCSS); and Social, Economic, and Workforce Implications of IT and IT Workforce Development (SEW). Each PCA spans an area in which multiple agencies have activities.

Penn State
Pennsylvania State University.

Peta-
A prefix denoting 10^{15}, or a thousand trillion; for example, petabits.

Pflops
Petaflops, 10^{15} flops.

PingER
DOE network-monitoring tool providing ping end-to-end reporting.

PITAC
President's Information Technology Advisory Committee.

PKI
Public key infrastructure.

PMEL
NOAA's Pacific Marine Environmental Laboratory.

PSE
ASCI's problem solving environment.

PVM
Parallel virtual machine.

PVP
Parallel vector processor.

QBone
I2's end-to-end testbed to accelerate the development of interdomain quality of service.

QC
Quantum computing.

QoS
Quality of service.

qubit
One quantum bit.

Quorum
DARPA program to develop seamless, interoperable distributed defense networks from COTS products.

R&D
Research and development.

RBAC
Role-based access control.

Red Hat
A company that distributes and supports the Linux OS.

REE
NASA's Remote Exploration and Experimentation project.

RENATER2
French network.

RSFQ
Rapid single flux quantum, a high-speed superconductor logic.

RSPAC
NASA's Remote Sensing Public Access Center.

SAN
Storage area network.

SCxx
Abbreviation for the national high performance networking and computing conference, formerly called SuperComputing, e.g., SC99.

Scalable
A system is scalable if it can be made to have more (or less) computational power by configuring it with a larger (or smaller) number of processors, amount of memory, interconnection bandwidth, input/output bandwidth, and mass storage.

SDM
ASCI's scientific data management program.

SDP
Software Design and Productivity, one of the IT R&D PCAs. Established in FY 2000.

SDPCG
Software Design and Productivity Coordinating Group.

SDSC
San Diego Supercomputer Center.

SensIT
DARPA's sensor information technology program.

SEW
Social, Economic, and Workforce Implications of IT and IT Workforce Development, one of the IT R&D PCAs. Successor to the ETHR PCA.

SEWCG
Social, Economic, and Workforce Implications of IT and IT Workforce Development Coordinating Group. Successor to the ETHR Working Group.

SF Express
Synthetic Forces Express, a project using high performance distributed computing systems to support large-scale interactive simulations.

SGI
Silicon Graphics, Inc.

SIGINT
Signals intelligence.

SII
Scalable information infrastructure.

SIMA
NIST's Systems Integration for Manufacturing Applications program.

SingaREN
Singapore Advanced Research and Education Network.

SMI
NSA's security management infrastructure.

SMP
Shared memory symmetric multiprocessor.

SNL
DOE's Sandia National Laboratories.

SNMP
Simple Network Management Protocol.

SOI
Silicon-on-insulator.

SONET
Synchronous optical network.

SSP
DOE's Stockpile Stewardship Program.

STAR TAP
NSF's Science, Technology, And Research Transit Access Point, an international transit network meeting point in Chicago.

Subcommittee on CIC R&D
Subcommittee on Computing, Information, and Communications Research and Development. Predecessor of the Interagency Working Group on Information Technology R&D (IWG/IT R&D).

SuperNet
A DARPA NGI network to demonstrate multi-Gbps end-to-end transmission.

SURA
Southeastern Universities Research Association.

SURFnet
Netherlands network.

T, or Tera-
A prefix denoting 10^{12} or a trillion; for example, terabits, teraflops.

TANnet2
Taiwan network.

Tbps, or Tb/s
Terabits per second.

TBps, or TB/s
Terabytes per second.

TCP
Transmission Control Protocol.

Tele-nanoManipulator
A system enabling scientists to see, touch, and manipulate nanometer-scale objects, such as viruses, using remote microscopes over high-speed networks.

TDMA
Time Division Multiple Access, a digital wireless technology enabling multiple simultaneous data channels on a single radio frequency.

TIDES
DARPA's Translingual Information Detection, Extraction, and Summarization project.

TransPAC
U.S./Asia-Pacific Consortium.

TREC
Text REtrieval Conference, sponsored by DARPA and NIST.

UCAID
University Corporation for Advanced Internet Development, which operates the Abilene network supporting Internet2 project research universities.

UCLA
University of California-Los Angeles.

UCSB
University of California-Santa Barbara.

UCSD
University of California-San Diego.

UIC
University of Illinois at Chicago.

UMLS
NLM's Unified Medical Language System.

UPC
A programming language that combines features of AC with features of Split-C and PCP, two other parallel C languages.

URI
OSD's University Research Initiative.

USC
University of Southern California.

USGS
United States Geological Survey.

U.S.
United States.

UW
University of Washington.

vBNS
NSF's very high performance Backbone Network Services.

VCC
NASA's virtual collaborative clinic.

VE
Virtual environment.

VGI
Virtual graphics interface.

VIA
Virtual interface architecture, an industry-standard interface for system area networks.

VIEWS
DOE's Visual Interactive Environment for Weapons Simulation.

Virginia Tech
Virginia Polytechnic Institute and State University.

VLAB
NASA's Virtual Laboratory.

VLSI
Very large-scale integration.

VON
Virtual overlay network.

VPN
Virtual private network.

VRML
Virtual Reality Modeling Language.

WAN
Wide area network.

WDM
Wavelength division multiplexing.

Web
A reference to the World Wide Web.

Windows NT
A Microsoft network operating system.

Wireless technologies
Communications technologies that use radio, microwave, or satellite communications channels versus wire, coaxial cable, or optical fiber.

WWW
World Wide Web.

XML
Extensible Markup Language, an offshoot of the hypertext markup language HTML that enables formatting of multimedia Web pages.

Contacts

National Coordination Office for Computing, Information, and Communications (NCO/CIC)

Suite II-405
4201 Wilson Boulevard
Arlington, VA 22230
(703) 292-ITRD (292-4873)
FAX: (703) 292-9097
nco@itrd.gov
ngi@itrd.gov

Internet Web Servers:

http://www.itrd.gov/
http://www.ngi.gov/

NCO

Kay Howell
Director

Sally E. Howe, Ph.D.
Associate Director

Yolanda Comedy, Ph.D.
Technical Staff (Presidential Advisory Committee Liaison)

Michael Hammon
Special Projects

Vicki L. Harris
Administrative Assistant

Larry Janicki , Ph.D.
Senior Systems Administrator

Martha K. Matzke
Technical Writer (Blue Book Associate Editor)

Betty McDonough
Webmaster

Grant Miller, Ph.D.
Technical Staff (LSN Liaison)

Laurie Mitchell
Executive Secretary

Krish Namboodiri, Ph.D.
Technical Staff (Implementation Plan Liaison)

Terrence L.Ponick, Ph.D.
Technical Writer (Blue Book Editor)

Ann Rutherford
Administrative Officer

Alan Tellington
Systems Administrator

Carolyn Van Damme
Chief of Staff and Outreach Coordinator

Robert I. Winner, Ph.D.
Technical Staff (Presidential Advisory Committee Support)

AHRQ

J. Michael Fitzmaurice, Ph.D., FACMI
Senior Science Advisor for Information Technology,
Immediate Office of the Director
Agency for Healthcare Research and Quality
2101 East Jefferson Street, Suite 600
Rockville, MD 20852
(301) 594-3938
FAX: (301) 594-2168

BLS

Cathryn S. Dippo, Ph.D.
Associate Commissioner, Office of Survey Methods Research
Bureau of Labor Statistics
2 Massachusetts Avenue N.E., Suite 4915
Washington, DC 20212
(202) 691-7372
FAX: (202) 691-7426

DARPA

Shankar Sastry, Ph.D.
Director, Information Technology Office
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-2228
FAX: (703) 696-4534

Helen Gill, Ph.D.
Program Manager, Information Technology Office
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-7461
FAX: (703) 696-4534

Gary M. Koob, Ph.D.
Program Manager, Information Technology Office
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-7463
FAX: (703) 696-4534

Sri Kumar, Ph.D.
Program Manager, Information Technology Office
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-0174
FAX: (703) 696-4534

Mari W. Maeda, Ph.D.
Program Manager, Information Technology Office
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-2255
FAX: (703) 696-4534

W. Douglas Maughan, Ph.D.
Program Manager, Information Technology Office
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-2373
FAX: (703) 696-4534

Jean Scholtz, Ph.D.
Program Manager, Information Technology Office
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-4469
FAX: (703) 696-4534

Stephen L. Squires, Ph.D.
Special Assistant for Information Technology
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-2400
FAX: (703) 696-2209

Gary W. Strong, Ph.D.
Program Manager, Information Technology Office
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-2259
FAX: (703) 696-0564

Col. Mark Swinson
Program Manager, Information Technology Office
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-2247
FAX: (703) 696-4534

Janos Sztipanoits, Ph.D.
Program Manager, Information Technology Office
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714
(703) 696-2234
FAX: (703) 696-4534

DoD

Charles J Holland, Ph.D.
Director of Information Systems, DUSD (S&T)
Department of Defense
1777 North Kent Street, Suite 9030
Rosslyn, VA 22209
(703) 588-7443
FAX: (703) 588-7560

Rodger Johnson
Program Manager, Defense Research and Engineering Network
DoD High Performance Computing Modernization Office
1010 North Glebe Road, Suite 510
Arlington, VA 22201
(703) 812-8205
FAX: (703) 812-9701

DOE

C. Edward Oliver, Ph.D.
Associate Director, Office of Advanced Scientific Computing Research (ASCR), Acting Director, Mathematical, Information, and Computational Sciences (MICS) Division, ASCR
Department of Energy
OASCR/MICS, SC-30,
19901 Germantown Road
Germantown, MD 20874
(301) 903-7486
FAX: (301) 903-4846

Richard A. Carlson
Program Manager, Mathematical, Information, and Computational Sciences (MICS) Division, Office of Advanced Scientific Computing Research (ASCR)
Department of Energy
OASCR/MICS, SC-31
19901 Germantown Road
Germantown, MD 20874-1290
(301) 903-0073
FAX: (301) 903-7774

Thom H. Dunning, Jr., Ph.D.
Assistant Director, Office of Science
Department of Energy
19901 Germantown Road
Germantown, MD 20874-1290
(301) 903-9550
(301) 903-4846

Daniel A. Hitchcock, Ph.D.
Program Manager, Mathematical, Information, and Computational Sciences (MICS) Division, Office of Advanced Scientific Computing Research (ASCR)
Department of Energy
OASCR/MICS, SC-31
19901 Germantown Road
Germantown, MD 20874-1290
(301) 903-5800
FAX: (301) 903-7774

Frederick C. Johnson, Ph.D.
Program Manager, Mathematical, Information, and Computational Sciences (MICS) Division, Office of Advanced Scientific Computing Research (ASCR)
Department of Energy
OASCR/MICS, SC-31
19901 Germantown Road
Germantown, MD 20874-1290
(301) 903-3601
FAX: (301) 903-7774

Thomas A. Kitchens, Ph.D.
Program Director, Mathematical, Information, and Computational Sciences (MICS) Division, Office of Advanced Scientific Computing Research (ASCR)
Department of Energy
OASCR/MICS, SC-31
19901 Germantown Road
Germantown, MD 20874-1290
(301) 903-5152
FAX: (301) 903-7774

Norman H. Kreisman
Advisor, International Technology
Department of Energy, SC5
Mailstop 3H049-FORS
1000 Independence Avenue, S.W.
Washington, D.C. 20585
(202) 586-9746
FAX: (202) 586-7152

Paul C. Messina, Ph.D.
Senior Technical Advisor, Defense Programs,
Office of Strategic Computing and Simulation
Department of Energy
Room 1J077
1000 Independence Avenue, S.W.
Washington, D.C. 20585
(202) 586-1101
FAX: (202) 586-2168

José L. Muñoz, Ph.D.
Lead Computer Scientist, Office of Simulation and Computer Science
Department of Energy, DP-14
Forrestal Building
1000 Independence Avenue, S.W.
Washington, D.C. 20585
(202) 586-5132
FAX: (202) 586-8005

Mary Anne Scott, Ph.D.
Program Manager, Mathematical, Information, and Computational Sciences (MICS) Division, Office of Advanced Scientific Computing Research (ASCR)
Department of Energy
OASCR/MICS, SC-31
19901 Germantown Road
Germantown, MD 20874-1290
(301) 903-6368
FAX: (301) 903-7774

George R. Seweryniak
Program Manager, Mathematical, Information, and Computational Sciences (MICS) Division, Office of Advanced Scientific Computing Research (ASCR)
Department of Energy
OASCR/MICS, SC-31
19901 Germantown Road
Germantown, MD 20874-1290
(301) 903-0071
FAX: (301) 903-7774

EPA

Joan H. Novak
HPCC Program Manager, MD-80
Environmental Protection Agency
Research Triangle Park, NC 27711
(919) 541-4545
FAX: (919) 541-1379

Robin L. Dennis, Ph.D.
Senior Science Program Manager, MD-80
Environmental Protection Agency
Research Triangle Park, NC 27711
(919) 541-2870
FAX: (919) 541-1379

EPA FedEx address:
79 T.W. Alexander Drive
Building 4201
Research Triangle Park, NC 27709

GSA

G. Martin Wagner
Associate Administrator for Governmentwide Policy
General Services Administration
1800 F Street, N.W. (M) Room 5240
Washington, D.C. 20405
(202) 501-8880
FAX: (202) 501-8898

Keith Thurston
Assistant to Deputy Associate Administrator, Office of Governmentwide Policy
General Services Administration
1800 F Street, N.W. (MK) Room 2239
Washington, D.C. 20405
(202) 501-3175
FAX: (202) 501-2482

Susan B. Turnbull
Director, Center for IT Accommodation
General Services Administration
1800 F Street, N.W. (MKC) Room 1234
Washington, D.C. 20405-0001
(202) 501-6214
FAX: (202) 501-6269

NASA

Lee B Holcomb
Chief Information Officer
National Aeronautics and Space Administration
Code AO
300 E Street, S.W.
Washington, DC 20546
(202) 358-1824
FAX: (202) 358-3063

David B. Nelson, Ph.D.
Deputy Chief Information Officer
National Aeronautics and Space Administration
Code AO
300 E Street, S.W.
Washington, DC 20546
(202) 358-1817
FAX: (202) 358-3063

Betsy Edwards
Chief Information Officer, Office of Aero-Space Technology
National Aeronautics and Space Administration
Code RS
300 E Street, S.W.
Washington, DC 20546
(202) 358-4639
FAX: (202) 358-3550

Kul B. Bhasin, Ph.D.
Thrust Area Manager for High Rate Data Delivery
National Aeronautics and Space Administration
NASA Glenn Research Center at Lewis Field
21000 Brookpark Road
Mail Stop 54-2
Cleveland, OH 44135
(216) 433-3676
FAX: (216) 977-7444

Wayne H. Bryant
Assistant Chief, Flight Electronics Technology Division
NASA Langley Research Center
National Aeronautics and Space Administration
Mail Stop 150
Hampton, VA 23681
(757) 864-1690
FAX: (757) 864-8821

William J Campbell
Head, Applied Information Sciences Branch, NASA Goddard Space Flight Center
National Aeronautics and Space Administration
Code 935 - NASA/GSFC
Greenbelt, MD 20771
(301) 286-8785
FAX: (301) 286-1776

James R. Fischer
Project Manager, NASA HPCC/ESS Project
NASA Goddard Space Flight Center
National Aeronautics and Space Administration
Code 930 - NASA/GSFC
Greenbelt, MD 20771
(301) 286-3465
FAX: (301) 286-1634

Kenneth Freeman
HPCC/NREN Project Manager
National Aeronautics and Space Administration
Ames Research Center
Mail Stop 233-21
Moffett Field, CA 94035
(650) 604-1263

Anngienetta R. Johnson
Director, Program Planning and Development Division, Office of Earth Science
National Aeronautics and Space Administration
Code YF
300 E Street, S.W.
Washington, DC 20546
(202) 358-4717
FAX: (202) 358-2769

Nand Lal, Ph.D.
Computer Scientist, Digital Library Technologies
NASA Goddard Space Flight Center
National Aeronautics and Space Administration
Code 933 - NASA/GSFC
Greenbelt, MD 20771
(301) 286-7350
FAX: (301) 286-1775

Stephen Scott Santiago
Chief Information Officer, NASA Ames Research Center
National Aeronautics and Space Administration
Ames Research Center
Mail Stop 233-7
Moffett Field, CA 94035-1000
(650) 604-5015
FAX: (650) 604-6999

Eugene Tu, Ph.D.
Program Manager, High Performance Computing and Communications Program
National Aeronautics and Space Administration
Ames Research Center
Mail Stop 258-3
Moffett Field, CA 94035-1000
(650) 604-4486
FAX: (650) 604-4036

NIH

NIH IT R&D Coordinator and CIT contact:

Robert L. Martino, Ph.D.
Associate Director, Office of Computational Bioscience
Chief, Computational Bioscience and Engineering Laboratory, Center for Information Technology (CIT)
National Institutes of Health
12 South Drive, MSC 5624
Building 12A, Room 2033
Bethesda,MD 20892-5624
(301) 496-1112
FAX: (301) 402-2867

Don R. Preuss
Chief Technology Officer, Center for Information Technology (CIT)
National Institutes of Health
12 South Drive, MSC 5654
Building 12A, Room 3033
Bethesda,MD 20892-5654
(301) 496-5703
FAX: (301) 402-1754

Alan S. Graeff
Chief Information Officer, Center for Information Technology (CIT)
National Institutes of Health
12 South Drive, MSC 5654
Building 12A, Room 3033
Bethesda, MD 20892-5654
(301) 496-5703
FAX: (301) 402-1754

NCI program:

Jacob V. Maizel, Ph.D.
Biomedical Supercomputer Center, National Cancer Institute, Frederick Cancer Research and Development Center
National Institutes of Health
P.O. Box B, Building 469,Room 151
Frederick, MD 21702-1201
(301) 846-5532
FAX: (301) 846-5598

Carol A. Dahl, Ph.D.
Director, Office of Technology and Industrial Relations National Cancer Institute
National Institutes of Health
31 Center Drive, MSC 2590
Building 31,Room 11A03
Bethesda, MD 20892-2590
(301) 496-1550
FAX: (301) 496-7807

NCRR program:

Judith L. Vaitukaitis,M.D.
Director, National Center for Research Resources
National Institutes of Health
31B Center Drive
Building 31,Room 3B11
Bethesda,MD 20892-2128
(301) 496-5793
FAX: (301) 402-0006

Richard M. DuBois,Ph.D.
Head, Computer Technology Section, BT National Center for Research Resources
National Institutes of Health
One Rockledge Center
6705 Rockledge Drive, Room 6148
Bethesda,MD 20892-7965
(301) 435-0758
FAX: (301) 480-3659

Michael Marron, Ph.D.
Director, Biomedical Technology
National Center for Research Resources
National Institutes of Health
One Rockledge Center
6705 Rockledge Drive, Room 6160
Bethesda, MD 20892-7965
(301) 435-0755
FAX: (301) 480-3775

NLM program:

Donald A.B. Lindberg, M.D.
Director, National Library of Medicine
National Institutes of Health
Building 38, Room 2E17B
8600 Rockville Pike
Bethesda, MD 20894
(301) 496-6221
FAX: (301) 496-4450

Michael J Ackerman, Ph.D.
Assistant Director for High Performance Computing and Communications, National Library of Medicine
National Institutes of Health
Building 38A, Room B1N30
8600 Rockville Pike
Bethesda, MD 20894
(301) 402-4100
FAX: (301) 402-4080

NIGMS program:

James C Cassatt
Director, Division of Cell Biology and Biophysics, National Institute of General Medical Sciences
National Institutes of Health
4500 Center Drive, MSC 4500
Bethesda, MD 20892
(301) 594-0828
FAX: (301) 480-2004

NIST

William Mehuron, Ph.D.
Director, Information Technology Laboratory
National Institute of Standards and Technology
100 Bureau Drive, Stop 8900
Gaithersburg, MD 20899-8900
(301) 975-2900

Larry H. Reeker, Ph.D.
Senior Computer Scientist, Information Technology Laboratory
National Institute of Standards and Technology
100 Bureau Drive, Stop 8970
Gaithersburg MD 20899-8970
(301) 975-5147
FAX: (301) 948-1784

Ronald F. Boisvet, Ph.D.
Chief, Mathematical and Computational Sciences Division, Information Technology Laboratory
National Institute of Standards and Technology
100 Bureau Drive, Stop 8910
Gaithersburg, MD 20899-8910
(301) 975-3812
FAX: (301) 990-4127

James Fowler
Manager, Systems Integration for Manufacturing Applications Program (SIMA), Manufacturing Engineering Laboratory
National Institute of Standards and Technology
100 Bureau Drive, Stop 8260
Gaithersburg MD 20899-8260
(301) 975-3180
FAX: (301) 258-9749

Martin Herman, Ph.D.
Chief, Information Access and User Interfaces Division, Information Technology Laboratory
National Institute of Standards and Technology
100 Bureau Drive, Stop 8940
Gaithersburg, MD 20899-8940
(301) 975-4495
FAX: (301) 975-5287

Kevin L. Mills, Ph.D.
Chief, Advanced Network Technologies Division, Information Technology Laboratory
National Institute of Standards and Technology
100 Bureau Drive, Stop 8920
Gaithersburg, MD 20899-8920
(301) 975-3618
FAX: (301) 590-0932

Doug Montgomery
Manager, Internetworking Technologies Group, Advanced Network Technologies Division
National Institute of Standards and Technology
100 Bureau Drive, Stop 8920
Gaithersburg MD 20899-8920
(301) 975-3630
FAX: (301) 590-0932

Steven R Ray, Ph.D.
Chief, Manufacturing Systems Integration Division, Manufacturing Engineering Laboratory
National Institute of Standards and Technology
100 Bureau Drive, Stop 8260
Gaithersburg, MD 20899-8260
(301) 975-3524
FAX: (301) 258-9749

Robert Rosenthal
Computer Security Division, Information Technology Laboratory
National Institute of Standards and Technology
100 Bureau Drive, Stop 8930
Gaithersburg, MD 20899-8930
(301) 975-3603
FAX: (301) 948-1233

David C. Stieren
Program Manager, National Advanced Manufacturing Testbed, Manufacturing Engineering Laboratory
National Institute of Standards and Technology
100 Bureau Drive, Stop 8202
Gaithersburg, MD 20899-8202
(301) 975-3197
FAX: (301) 926-8730

NOAA

Thomas N. Pyke, Jr.
Director for HPCC
National Oceanic and Atmospheric Administration
Room 15300
1315 East-West Highway
Silver Spring, MD 20910
(301) 713-3573
FAX: (301) 713-4040

William T. Turnbull
Deputy Director for HPCC
National Oceanic and Atmospheric Administration
Room 15300
1315 East-West Highway
Silver Spring, MD 20910
(301) 713-3573
FAX: (301) 713-4040

Alexander E. MacDonald, Ph.D.
Director, Forecast Systems Laboratory
National Oceanic and Atmospheric Administration
325 Broadway
Boulder, CO 80303
(303) 497-6378
FAX: (303) 497-6821

Jerry Mahlman, Ph.D.
Director, Geophysical Fluid Dynamics Laboratory
National Oceanic and Atmospheric Administration
Forrestal Campus U.S. Route 1
P.O. Box 308
Princeton, NJ 08542-0308
(609) 452-6502
FAX: (609) 987-5070

Bruce B Ross, Ph.D.
Deputy Director, Geophysical Fluid Dynamics Laboratory
National Oceanic and Atmospheric Administration
Forrestal Campus, U.S. Route 1
P.O. Box 308
Princeton, NJ 08542-0308
(609) 452-6504
FAX: (609) 987-5070

Louis Uccellini
Director, National Centers for Environmental Prediction
National Oceanic and Atmospheric Administration
5200 Auth Road, Room 101
Camp Springs, MD 20746
(301) 763-8000
FAX: (301) 763-8434

NSA

George R. Cotter
Chief Scientist
National Security Agency
9800 Savage Road, Suite 6217
Fort George G. Meade, MD 20755-6217
(301) 688-6434
FAX: (301) 688-4980

Norman S. Glick
Senior Computer Scientist
National Security Agency
9800 Savage Road, Suite 6217
Fort George G. Meade, MD 20755-6217
(301) 688-8448
FAX: (301) 688-4980

Howard Gordon
Senior Computer Scientist
National Security Agency
9800 Savage Road, Suite 6648
Fort George G. Meade, MD 20755-6648
(301) 688-9513
FAX: (301) 688-9454

Christina M. McBride
Secure Networks and Computer Systems Division
National Security Agency
9800 Savage Road, Suite 6534
Fort George G. Meade, MD 20755-6534
(301) 688-0847
FAX: (301) 688-0255

Robert V. Meushaw
Technical Director, INFOSEC Research and Technology
National Security Agency
9800 Savage Road, Suite 6529
Fort George G. Meade, MD 20755-6529
(301) 688-0840
FAX: (301) 688-0255

William J. Semancik
Director, Laboratory for Telecommunications Sciences
National Security Agency
c/o U. S. Army Research Laboratory
Adelphi Laboratory Center
2800 Powder Mill Road, Building 601, Room 131
Adelphi, MD 20783-1197
(301) 688-1709
FAX: (301) 291-2591

NSF

Ruzena Bajcsy, Ph.D.
Assistant Director, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201 Wilson Boulevard, Suite 1105
Arlington, VA 22230
(703) 306-1900
FAX: (703) 306-0577

George O. Strawn, Ph.D.
Executive Officer, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201 Wilson Boulevard, Suite 1105
Arlington, VA 22230
(703) 306-1900
FAX: (703) 306-0577

S. Kamal Abdali, Ph.D.
Deputy Division Director, Computer-Communications Research Division, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201 Wilson Boulevard, Suite 1145
Arlington, VA 22230
(703) 306-1910
FAX: (703) 306-1947

Frank D. Anger, Ph.D.
Program Director, Software Engineering and Languages, Computer-Communications Research Division, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201 Wilson Boulevard, Suite 1145
Arlington, VA 22230
(703) 306-1911
FAX: (703) 306-1947

William S. Bainbridge, Ph.D.
Science Advisor, Office of the Assistant Director, Directorate for Social, Behavioral, and Economic Sciences
National Science Foundation
4201 Wilson Boulevard, Suite 905
Arlington, VA 22230
(703) 306-1741
FAX: (703) 306-0495

Robert R. Borchers,Ph.D.
Division Director, Advanced Computational Infrastructure and Research Division, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201Wilson Boulevard,Suite 1122
Arlington, VA 22230
(703) 306-1970
FAX: (703) 306-0632

Javad Boroumand
Associate Program Director, Advanced Networking Infrastructure and Research Division, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201Wilson Boulevard,Suite 1175
Arlington, VA 22230
(703) 306-1949
FAX: (703) 306-0621

Lawrence E. Brandt
Program Director for Digital Government, Division of Experimental and Integrative Activities, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201Wilson Boulevard,Suite 1160
Arlington, VA 22230
(703) 306-1981
FAX: (703) 306-0589

Aubrey M. Bush,Ph.D.
Division Director, Advanced Networking Infrastructure and Research Division, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201Wilson Boulevard,Suite 1175
Arlington, VA 22230
(703) 306-1950
FAX: (703) 306-0621

Y.T. Chien, Ph.D.
Senior Scientist, Multidisciplinary Research, Information and Intelligent Systems Division, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201Wilson Boulevard,Suite 1160
Arlington, VA 22230
(703) 306-1980
FAX: (703) 306-0589

Robert A. Eisenstein, Ph.D.
Assistant Director, Directorate for Mathematical and Physical Sciences
National Science Foundation
4201Wilson Boulevard,Suite 1005
Arlington, VA 22230
(703) 306-1801
FAX: (703) 306-0545

Michael Evangelist, Ph.D.
Division Director, Computer-Communications Research Division, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201Wilson Boulevard,Suite 1145
Arlington, VA 22230
(703) 306-1910
FAX: (703) 306-1947

Michael J Foster, Ph.D.
Program Director, Experimental Partnerships, Experimental and Integrative Activities, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201Wilson Boulevard,Suite 1160
Arlington, VA 22230
(703) 306-1980
FAX: (703) 306-0589

Valerie Gregg
Program Manager, Federal World Wide Web Consortium, Advanced Computational Infrastructure and Research Division, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201Wilson Boulevard,Suite 1122s
Arlington, VA 22230
(703) 306-1980
FAX: (703) 306-0610

Stephen M.Griffin
Program Director, Digital Libraries Initiative, Information and Intelligent Systems Division, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201Wilson Boulevard,Suite 1115
Arlington, VA 22230
(703) 306-1930
FAX: (703) 306-0599

C. Suzanne Iacono, Ph.D.
Program Director, Computation and Social Systems, Information and Intelligent Systems Division, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201 Wilson Boulevard, Suite 1115
Arlington, VA 22230
(703) 306-1927
FAX: (703) 306-0599

Bradley D. Keister, Ph.D.
Program Director, Nuclear Physics, Physics Division, Directorate for Mathematical and Physical Sciences
National Science Foundation
4201 Wilson Boulevard, Suite 1015
Arlington, VA 22230
(703) 306-1891
FAX: (703) 306-0566

Charles H. Koelbel, Ph.D.
Program Director, Advanced Computational Research, Advanced Computational Infrastructure and Research Division, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201 Wilson Boulevard, Suite 1122
Arlington, VA 22230
(703) 306-1962
FAX: (703) 306-0632

Michael Lesk, Ph.D.
Division Director, Information and Intelligent Systems Division, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201 Wilson Boulevard, Suite 1115
Arlington, VA 22230
(703) 306-1930
FAX: (703) 306-0599

Stephen R. Mahaney
Senior Advisor for Budget, Policy, and Planning, Directorate for Computer and Information Science and Engineering
National Science Foundation
4201 Wilson Boulevard, Suite 1160
Arlington, VA 22230
(703) 306-1900
FAX: (703) 306-0577

OSD/URI

Cliff G. Lau, Ph.D.
Associate Director for Corporate Programs
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5660
(703) 696-0431
FAX: (703) 588-1013

OMB

David S. Trinkle
Program Examiner, Office of Management and Budget
Executive Office of the President
New Executive Office Building, Room 8225
725 17th Street, N.W.
Washington, DC 20503
(202) 395-4706
FAX: (202) 395-4652

OSTP

Duncan T. Moore, Ph.D.
Associate Director for Technology, Office of Science and Technology Policy
Executive Office of the President
Old Executive Office Building Room 423
17th Street and Pennsylvania Avenue, N.W.
Washington, DC 20502
(202) 456-6046
FAX: (202) 456-6023

Lori A. Perine
Deputy to the Associate Director for Technology, Office of Science and Technology Policy
Executive Office of the President
Old Executive Office Building Room 423
17th Street and Pennsylvania Avenue, N.W.
Washington, DC 20502
(202) 456-6039
FAX: (202) 456-6023

FY 2001 Editorial Group

Executive Editor
Sally E. Howe
National Coordination Office

Editor
Terrence L. Ponick
National Coordination Office

Associate Editor
Martha K. Matzke
National Coordination Office

NCO Writing Group
Sally E. Howe
Kristin Janger
Martha K. Matzke
Grant Miller
Krish Namboodiri
Terrence L. Ponick
Robert J. Winner

Contributors
Frank D. Anger, NSF
Richard M. Dubois, NIH
James R. Fischer, NASA
Helen Gill, DARPA
Norman S. Glick, NSA
Stephen M. Griffin, NSF
Martin Herman, NIST
Craig W. Hunt, NIST
C. Suzanne Iacono, NSF
Thomas A. Kitchens, DOE
R. J. (Jerry) Linn (formerly NIST)
Anne Richeson (formerly NSF)
Mary Anne Scott, DOE
Gary W. Strong, DARPA
Janos Sztipanovits, DARPA
Mark Swinson, DARPA
William T. Turnbull, NOAA

Acknowledgments

This Supplement to the President's FY 2001 Budget, the largest Blue Book to date, reflects not only the increasing breadth and accomplishments of the Federal IT R&D Programs but the editorial contributions of many individuals, to whom we extend sincerest thanks.

The agency program managers who developed the research goals and program plans highlighted in this report shaped its scope and substance.

The Editorial Group's colleagues at the NCO provided ready and able support in all aspects of the production process, especially proofreading, preparation of the Web version, and distribution.

Norman Glick of NSA, who has for years materially assisted in preparing the final document for publication, again contributed invaluable editorial services.

Abstract

The Federal Information Technology Research and Development (IT R&D) Programs conduct long-term research and development activities to assure continuing U.S. leadership in computing, information, and communications and to enable Federal departments and agencies to fulfill their missions successfully in the 21st century. This report summarizes the FY 2000 goals, objectives, and accomplishments of the IT R&D programs by Program Component Area (PCA), and presents an overview of some of the FY 2001 plans in each PCA.

The PCAs are:

HECC: High End Computing and Computation, including

HEC I&A High End Computing Infrastructure and Applications

HEC R&D High End Computing Research and Development

HCI & IM: Human Computer Interface and Information Management

LSN: Large Scale Networking, including Next Generation Internet Initiative (NGI) and Scalable Information Infrastructure (SII) activities

SDP: Software Design and Productivity

HCSS: High Confidence Software and Systems

SEW: Social, Economic, and Workforce Development Implications of IT and IT Workforce Development

The report also contains special sections on the President's Information Technology Advisory Committee (PITAC), the Federal Information Services Applications Council (FISAC), DOE's Accelerated Strategic Computing Initiative (ASCI), the Digital Libraries Initiative (DLI) Phase Two, and the National Computational Science Alliance's (Alliance's) Access Grid.

For additional copies or further information please contact:

National Coordination Office for Computing, Information,
and Communications
4201 Wilson Boulevard, Suite II-405
Arlington, VA 22230
Voice: (703) 292-ITRD (292-4873)
Fax: (703) 292-9097

NCO:
Web: http://www.itrd.gov/
E-mail: nco@itrd.gov

NGI:
Web: *http://www.ngi.gov/*
E-mail: ngi@itrd.gov