

Advanced Scientific Computing Research: Delivering Computing for the Frontiers of Science

Executive Summary

This plan presents the strategic vision for the Advanced Scientific Computing Research (ASCR) program in the Department of Energy's Office of Science for the next 10 years. It responds to the challenge of providing the computing and networking knowledge, tools, and facilities to enable DOE and the Nation's world leadership in critical areas of science, such as the following:

- Enabling new materials through nanoscience;
- Enabling the design and engineering of fusion power plants to produce energy without CO₂;
- Understanding the regional effects of global climate change;
- Developing new bacteria that can produce hydrogen, sequester carbon, and clean up toxic wastes;
- Understanding the fundamental nature of matter; and
- Understanding the processes that underpin combustion of fossil fuels to reduce pollution and increase efficiency.

ASCR will address these challenges through an integrated program that brings together a world-class basic research effort; strong, focused partnerships with application scientists across the Office of Science; and networking and computing facilities that enable the science of the next decade. The strategies laid out in this plan will increase the coupling between ASCR research efforts, its network facilities, and its computing resources—including evaluation testbeds, high-performance production capabilities and a new category of leadership-class computers. These leadership-class computers will play a key role in ASCR's portfolio, similar to that of the largest light sources and accelerators in the other programs of the Office of Science, and will provide capabilities over 100 times greater than the computers available for open science in the United States today.

The ASCR program will contribute significantly to the success of government-wide initiatives to revitalize high-end computing in the United States and to develop the cyber infrastructure needed to support the national research community. Specifically, ASCR will establish close partnerships with other Federal agencies—DOD, NSA, DARPA, and NSF—that will build on ASCR's strengths to make unique contributions to the Nation.

Contents

EXECUTIVE SUMMARY	2
INTRODUCTION	5
PROGRAM-WIDE STRATEGIC ISSUES.....	7
MAINTAINING A WORLD-CLASS BASIC RESEARCH EFFORT	7
WORKING EFFECTIVELY WITH APPLICATION SCIENTISTS.....	8
<i>Defining Requirements.....</i>	8
<i>Partnerships and Pilots.....</i>	12
<i>Long-Term Support and Maintenance of Software.....</i>	13
MANAGING INTERAGENCY PARTNERSHIPS.....	14
PROGRAM INTEGRATION AND PRIORITY SETTING	17
COMPUTING AND NETWORK FACILITIES AND TESTBEDS	19
RESEARCH	22
ANALYSES OF PROGRAM ELEMENTS: STRENGTHS, WEAKNESSES, OPPORTUNITIES, AND THREATS.....	22
APPLIED MATHEMATICS RESEARCH.....	22
<i>Contribution to Overall ASCR Strategic Goals</i>	22
<i>Planning Horizon.....</i>	22
<i>Current Areas of Investment</i>	22
<i>Transfer of Knowledge to Application Scientists</i>	22
<i>Strengths</i>	22
<i>Weaknesses</i>	22
<i>Opportunities</i>	22
<i>Threats</i>	22
<i>Planning Process.....</i>	22
<i>Gap Analysis.....</i>	22
COMPUTER SCIENCE RESEARCH.....	22
<i>Contribution to Overall ASCR Strategic Goals</i>	22
<i>Planning Horizon.....</i>	22
<i>Current Areas of Investment</i>	22
<i>Transfer of Knowledge to Application Scientists</i>	22
<i>Strengths</i>	22
<i>Weaknesses</i>	22
<i>Opportunities</i>	22
<i>Threats</i>	22
<i>Planning Process.....</i>	22
<i>Gap Analysis.....</i>	22
INTEGRATED NETWORK ENVIRONMENT.....	22
<i>Contribution to Overall ASCR Strategic Goals</i>	22
<i>Planning Horizon.....</i>	22
<i>Current Areas of Investment</i>	22
<i>Transfer of Knowledge to Application Scientists</i>	22
<i>Strengths</i>	22
<i>Weaknesses</i>	22
<i>Opportunities</i>	22
<i>Threats</i>	22
<i>Planning Process.....</i>	22
<i>Gap Analysis.....</i>	22
FACILITIES AND TESTBEDS.....	22
<i>High-Performance Production Capability: NERSC</i>	22
Contribution to Overall ASCR Strategic Goals.....	22

Planning Horizon	22
Current Areas of Investment	22
Transfer of Knowledge to Application Scientists.....	22
Strengths	22
Weaknesses.....	22
Opportunities	22
Threats	22
Planning Process	22
<i>Advanced Computing Research Testbeds.....</i>	<i>22</i>
Contribution to Overall ASCR Strategic Goals.....	22
Planning Horizon	22
Current Areas of Investment	22
Transfer of Knowledge to Application Scientists.....	22
Strengths	22
Weaknesses.....	22
Opportunities	22
Threats	22
Planning Process	22
Program Element Gap Analysis:.....	22
<i>ESnet.....</i>	<i>22</i>
Contribution to Overall ASCR Strategic Goals.....	22
Planning Horizon	22
Current Areas of Investment	22
Transfer of Knowledge to Application Scientists.....	22
Strengths	22
Weaknesses.....	22
The weaknesses that ESnet has to face are as follows:	22
Opportunities	22
Threats	22
Planning Process	22
Gap Analysis.....	22
APPENDIXES.....	22
APPENDIX A: EXTERNAL ENVIRONMENT.....	22
APPENDIX B: PLANNING PROCESS	22
APPENDIX C: MEASURES OF SUCCESS	22
APPENDIX D: LIST OF WORKSHOPS	22

Introduction

The mission of the Office of Advanced Scientific Computing Research (ASCR) is to

deliver forefront computational and networking capabilities to scientists nationwide that enable them to extend the frontiers of science, answering critical questions that range from the function of living cells to the power of fusion energy.

Computational science is increasingly central to progress at the frontiers of almost every scientific discipline and to our most challenging feats of engineering. Computer-based simulation enables us to predict the behavior of complex systems that are beyond the reach of our most powerful experimental probes or our most sophisticated theories. Computational modeling has greatly advanced our understanding of fundamental processes of nature, structure, and reactivity. We can now design novel catalysts and high-efficiency engines on computers. Through modeling and simulation, we will be able to explore the interior of stars and learn how protein machines work inside living cells. The other research programs in the Office of Science depend on the success of ASCR to enable them to answer many of the important questions facing their disciplines, such as the following:

- What new, useful properties do materials display as we move from the classical or macroscopic world to objects composed of a few atoms to a few thousand atoms or molecules? What range of optical, mechanical, catalytic, electrical, tribological, and other properties can be achieved by designing devices and materials at the molecular scale?
- How do we design new and revolutionary technologies and processes, using and combining principles of biological and physical systems that offer new solutions for challenges from medicine to environmental cleanup?
- Can we successfully control a burning plasma that shares the characteristic intensity and power of the sun?
- How were the nuclei of the chemical elements we find on earth formed in stars and supernovae?
- What is the nature of dark energy? Of dark matter? Why do they account for so much of the universe? What are their origins?

Scientific questions such as these require ASCR to advance beyond current computational abilities. To accomplish its mission and to enable the critical science of the next decade, ASCR must address the following challenges:

- What new mathematics is required to effectively model systems such as the Earth's climate or the behavior of living cells that involve processes taking place on vastly different time scales or length scales?
- Which computational architectures and platforms will deliver the most benefit for the science of today and the science of the future?

- What advances in computer science and algorithms are needed to increase the effectiveness with which supercomputers solve problems for the Office of Science?
- What operating systems, data management, analysis, model development, and other tools are required to make effective use of future-generation supercomputers?
- Is it possible to defeat geography by making all scientific resources readily available to scientists, regardless of whether they are at a university, national laboratory, or industrial setting?

The solutions to all of these challenges build on the successes of the ASCR program and its predecessors over the past half-century, from the establishment of the applied mathematics research program in the 1950s through the establishment of the first national open supercomputer center in 1974, to a leadership role in the transition to parallel computing and electronic collaboration in the 1990s. Today's successful partnerships with the other programs in the Office of Science—for example, Scientific Discovery through Advanced Computing (SciDAC) and Genomes to Life (GTL)—are built on these successes.

ASCR's ability to meet its mission goals is based on three fundamental strengths:

- A world-leading basic research effort in the areas of applied mathematics, computer science focused on enabling the high-performance computing, and integrated high-performance network environments (computer networks and collaboratories) of the future. These research efforts center on the requirements of the scientific disciplines.
- Strong, focused partnerships with other programs in the Office of Science and application scientists to effectively test, transfer, and validate ASCR research and to identify important opportunities for future research.
- World-class computing and network facilities that enable scientists to advance the forefront of discovery.

Each of these elements is critical to the success of ASCR in meeting its mission. ASCR's integrated approach to managing these strengths, combined with the expertise of the Office of Science in managing multidisciplinary partnerships, represents a unique resource for the nation.

This plan begins with a discussion of three program-wide issues: maintaining a world-class basic research effort, ensuring effective interaction with application scientists, and managing interagency partnerships. Following this are sections on the ASCR methodology for establishing priorities and allocating resources to program elements. The body of this plan focuses on analyses of four individual program elements: applied mathematics, computer science, integrated network environments, and facilities and testbeds. The analysis is based on results from a July, 2003 workshop held in Washington, D.C.

The appendices discuss the constraints the external environment places on ASCR, the planning process used by ASCR, and the metrics used by ASCR to measure success. Also provided is a list of links to the workshop reports that support this document.

Program-wide Strategic Issues

Maintaining a World-Class Basic Research Effort

In order for ASCR to succeed in its mission, it must maintain a world-class basic research effort in applied mathematics, computer science, and integrated network environments that are the foundation for the partnerships and facilities that enable the application scientists to succeed. We note that the time lapse between early research in a new mathematical or computer science idea and its broad use by application scientists is 10 years or longer. Therefore, ASCR must build the foundation today for the tools the application scientists will need a decade from now.

ASCR supports research activities that can be transferred to applications scientists in the next 3–5 years as well as research that is not expected to be widely used by application scientists for 10 years or more. It is relatively straightforward to plan for, or at least establish the requirements for, short-term projects by examining the issues faced by today’s application scientists. However, it is much more difficult to determine the types of long-term research that are needed. Although it might seem desirable to focus only on near-term activities, this strategy would drain the pipeline of ideas and have serious consequences for ASCR, the Office of Science, and the nation. The current state of research in high-performance computer architecture is one example of this problem, because the failure of the Government to address basic research in this area has resulted in a significant gap in our understanding of the architectures that will be needed in the next decade.

Details of the planning processes used by each of the ASCR research activities are addressed in the appropriate program element analysis. Each area must balance support for evolution of ongoing approaches with support of revolutionary approaches. For example, the Applied Mathematics program element has been supporting research on the solution of large linear systems on computers for decades, and we expect this work to evolve to address the challenges of such systems on computers with tens to hundreds of thousands of processors. On the other hand, research in wavelets or chaos theory represented radical approaches to critical problems, which were initially supported because of the strength of the peer review of the proposals and the vision of the DOE program manager.

Because work on these long-term research issues must inform and transfer to the shorter-term research portfolio, leaders of the ASCR research communities must work on both types of activity. Equally critical is long-term support for high-quality research groups to tackle the hardest problems. Moreover, because many of ASCR’s research questions take a decade to answer, we must pay attention to developing the applied mathematicians, computer scientists, integrated network environment researchers, and computational scientists of the future. One such effort is the Computational Science Graduate Fellowship (CSGF) program; another key effort is ASCR support of graduate students and postdoctoral researchers. In addition, ASCR has outreach efforts to minority-serving institutions to expand opportunities develop the next generation of leaders for the nation.

Working Effectively with Application Scientists

To ensure effective networking with application scientists, ASCR must first understand the requirements for the future, both for facilities and for research. Then ASCR must develop plans that fit within available budget envelopes to have the greatest chance of meeting the needs of the application scientists. Finally, ASCR must develop effective ways for delivering facilities as well as research results to application scientists.

Defining Requirements

We first discuss the process of defining requirements for computing and networking facilities, since this is somewhat simpler than the process of defining requirements for research.

The Office of Science has a well-defined process for monitoring the growth in requests for resources, as well as a detailed planning process to evaluate future requirements on a three-year timeframe. The results of this process, and an analogous process for SciDAC projects, are displayed in Figure 1. As the figure shows, within the current target-funding envelope, the demand for resources will dramatically exceed the supply through 2009. In addition, we note that this process of defining requirements captures primarily the capacity needs for running existing simulation software either more frequently or at modestly increased scale. It does not capture the requirements of the scientific community for computations at dramatically larger scale to address important scientific questions.

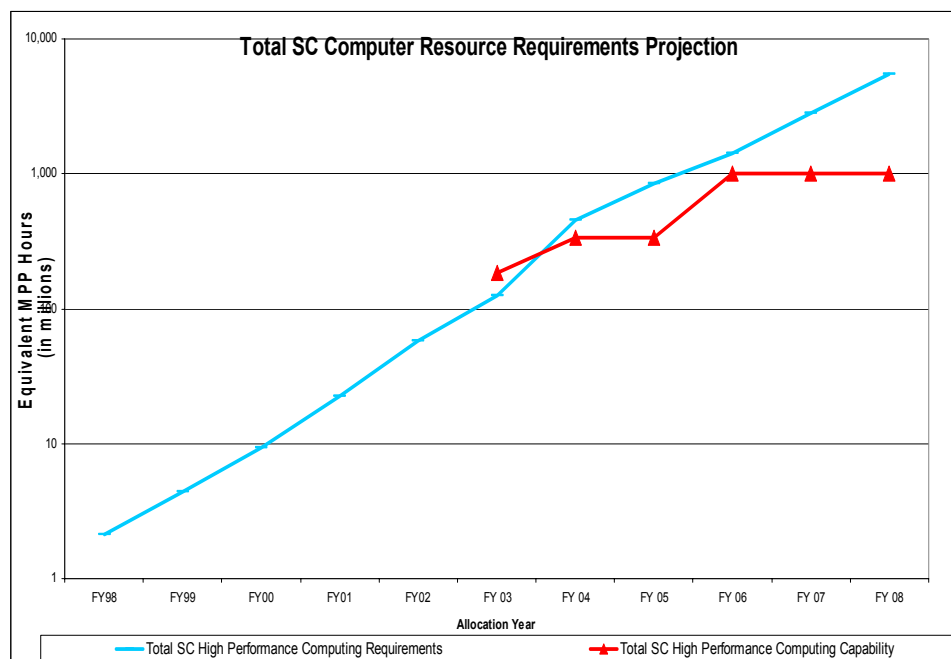


Figure 1: Projection of Computer Resource Requirements

To determine the scale of computing that would be needed to enable qualitatively new simulations, ASCR convened a workshop on June 23–24, 2003, with nearly 300 experts in the scientific disciplines as well as mathematics and computer science. The full results of the

workshop are at www.pnl.gov/scales. Table 1, extracted from the report, displays the new science that would be enabled by an increase in computational capability of a factor between 100 or 1000.

Table 1: New Science Enabled by Dramatically Increased Computational Capability

Research Programs	Major Scientific Advances
Biological and Environmental Sciences	<ul style="list-style-type: none"> • Provide global forecasts of Earth’s future climate at regional scales using high-resolution, fully coupled and interactive climate, chemistry, land cover, and carbon cycle models. • Develop predictive understanding of subsurface contaminant behavior that provides the basis for scientifically sound and defensible cleanup decisions and remediation strategies. • Establish a mathematical and computational foundation for the study of cellular and molecular systems and use computational modeling and simulation to predict and simulate the behavior of complex microbial systems for use in mission application areas.
Chemical and Materials Sciences	<ul style="list-style-type: none"> • Provide direct 3-dimensional simulations of a turbulent methane-air jet flame with detailed chemistry and direct 3-dimensional simulations of autoignition of <i>n</i>-heptane at high pressure, leading to more-efficient, lower-emission combustion devices. • Develop an understanding of the growth and structural, mechanical, and electronic properties of carbon nanotubes, for such applications as strengtheners of polymers and alloys, emitters for display screens, and conductors and nonlinear circuit elements for nanoelectronics. • Provide simulations of the magnetic properties of nanoparticles and arrays of nanoparticles for use in the design of ultra-high-density magnetic information storage. (Nanomagnetism is the simplest of many important advances in nanoscience that will affect future electronics, tailored magnetic response devices, and even catalysis.) • Resolve the current disparity between theory and experiment for conduction through molecules with attached electronic leads. (This is the very basis of molecular electronics and may well point to opportunities for chemical sensor technology.)
Fusion Energy Sciences	<ul style="list-style-type: none"> • Improve understanding of fundamental physical phenomena in high-temperature plasmas, including transport of energy and particles, turbulence, global equilibrium and stability, magnetic reconnection, electromagnetic wave/particle interactions, boundary layer effects in plasmas, and plasma/material interactions. • Simulate individual aspects of plasma behavior, such as energy and particle confinement times, high-pressure stability limits in magnetically confined plasmas, efficiency of electromagnetic wave heating and current drive, and heat and particle transport in the edge region of a plasma, for parameters relevant to magnetically confined fusion plasmas. • Develop a fully integrated capability for predicting the performance of magnetically confined fusion plasmas with high physics fidelity, initially for tokamak configurations and ultimately for a broad range of practical energy-producing magnetic confinement configurations. • Advance the fundamental understanding and predictability of high-energy density plasmas for inertial fusion energy. (Inertial fusion and magnetically confined fusion are complementary technological approaches to unlocking the power of the atomic nucleus.)
High Energy Physics	<ul style="list-style-type: none"> • Establish the limits of the Standard Model of elementary particle physics by achieving a detailed understanding of the effects of strong nuclear interactions in many different processes, so that the equality of Standard Model parameters measured in different experiments can be verified (or, if verification fails, signal the discovery of new physical phenomena at extreme sub-nuclear distances). • Develop realistic simulations of the performance of particle accelerators, the large and complex core scientific instruments of high-energy physics research, both to optimize the design, technology, and cost of future accelerators and to use existing accelerators more effectively and efficiently.
Nuclear Physics	<ul style="list-style-type: none"> • Understand the characteristics of the quark-gluon plasma, especially in the temperature-density region of the phase transition expected from quantum chromodynamics (QCD) and

- currently sought experimentally.
- Obtain a quantitative, predictive understanding of the quark-gluon structure of the nucleon and of interactions of nucleons.
- Understand the mechanism of core collapse supernovae and the nature of the nucleosynthesis in these spectacular stellar explosions.

Similarly, through the planning activities of the ESnet Steering Committee we have developed projections of the core network bandwidth to satisfy current applications (see Figure 2). Here, again, our ability to fulfill the needs of the scientists may be compromised sometime between FY 2005 and FY 2007.

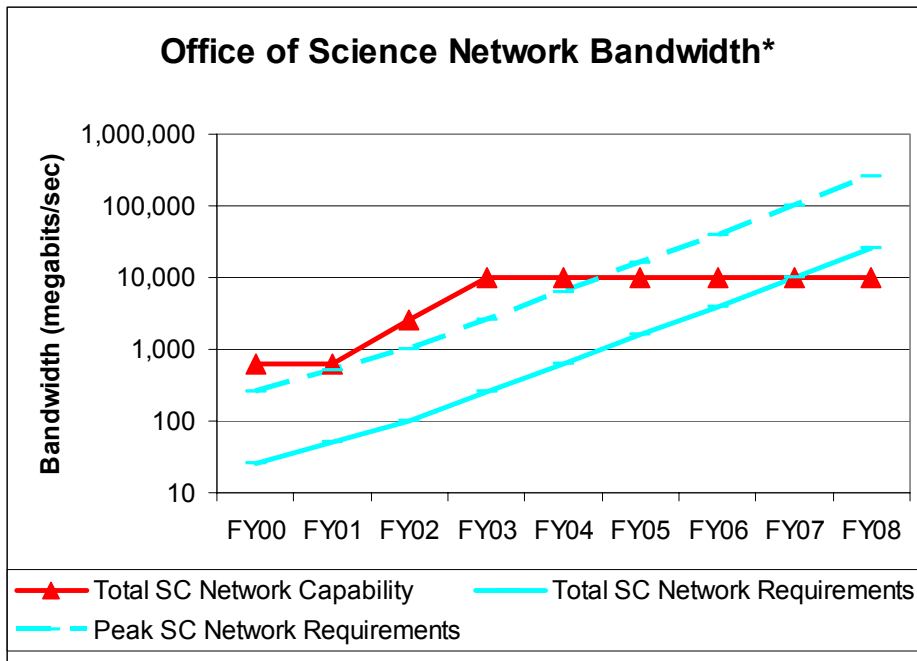


Figure 2: Projection of Network Needs

In addition, we recognize that the emergence of new applications and access to petascale facilities could alter our projections. To understand these issues, we conducted a workshop to develop a roadmap for ESnet. The results of this workshop are available on the Web at gate.hep.anl.gov/may/ScienceNetworkingWorkshop. Table 2 summarizes some of the network capability demands that we are likely to face.

Table 2: Network Demands in Selected Disciplines

1995–1999	2002–2004	2007–2009
Climate In 1998 there were about 5 TB/year of experimental and simulation climate data going to media. About	Climate experimental data and modeling data at the three largest U.S. facilities currently totals 100	By 2008 network accessible climate experimental and simulation data in the U.S. will be increasing at rate of

<p>this time, the DOE and other agencies launched a long-range program to acquire experimental data and support simulations.</p>	<p>TB (NERSC – 40 TB, ORNL– 40 TB, and NCAR (non -DOE) – 20 TB) and is being added at a rate of 20 TB/year.</p>	<p>3 PB/year. This is due to both greatly enhanced experimental measurements and simulations.</p>
<p>Fusion Energy Plasma physics/fusion research at DOE’s three main experimental facilities, General Atomics, MIT, and PPPL, and numerical simulations generated 2 TB of data in 1998 (mostly from experiments).</p>	<p>Present plasma physics/fusion experiments and simulations are generating 20 TB/year of data (each contributing roughly half).</p>	<p>Driven mainly by large scale advanced simulations and preparation for a burning plasma experiment, fusion researchers will be generating 1 PB/year of data by 2008. They also expect the necessary collaborative tools to be full partners in the international program.</p>
<p>Hadron Structure Investigation of the quark-gluon structure of the nucleon and nuclei resulted in 50 TB of data and analysis the first full year of operation of all of the experimental facilities CEBAF at JLab in 1998.</p>	<p>Currently, CEBAF experiments and analysis, including those associated with the discovery of the pentaquark, produce 300 TB/year of data.</p>	<p>CEBAF’s upgrade to 12 GeV to investigate quark confinement and detailed quark distributions will produce several PB/year.</p>
<p>Quark-Gluon Plasma The goal for the RHIC at BNL is discovering the quark-gluon plasma thought to exist at the edge of the Big Bang. RHIC began operations in 2000.</p>	<p>RHIC has early results that indicate that it may have discovered the quark-gluon plasma and is currently putting 600 TB/year to media.</p>	<p>By 2008, RHIC will increase the amount of data going to media to 5 PB/year as it details its information on the quark-gluon plasma.</p>
<p>Materials Science – Neutrons Neutron Science is critical for investigating the properties of materials by neutron scattering.</p>	<p>The SNS is currently under construction at ORNL. It will increase U.S. neutron science capabilities by more than an order of magnitude.</p>	<p>The SNS will turn on in late 2006 and achieve full operation in 2008, at which time it will produce 200 TB/year of data and analysis.</p>
<p>Materials Science – Photons The four DOE-funded light sources (APS, ALS, SSRL and NLS) are used to investigate the properties of materials and the structure of biological molecules such as proteins. In 1998 they accumulated 3 TB of data.</p>	<p>Currently the four light sources are acquiring and sending data at the rate of 30 TB/year over ESnet.</p>	<p>The drive to understand the dynamics as well as the structure of materials and biological molecules using greatly enhanced detectors will result in at least a fivefold increase in the acquisition of data at the light sources by 2008 to 150 TB/year.</p>
<p>Chemistry – Combustion Simulations for combustion are critical to improve our use of energy. The simulations were generating 100 GB/year in 1998.</p>	<p>Construction of a web-based archive for collaborative sharing and annotation of a broad range of chemical science data is now under way. Combustion is currently generating 3 TB/year and storing annotated feature and data subsets to this archive.</p>	<p>In 2007 combustion simulations will produce several PB/year of data to be collaboratively visualized, mined, and analyzed. In addition, there will be several hundreds of TB/year of experimental data generated, plus publication and annotation in Web-accessible archives of hundreds of TB/year for collaborative research.</p>
<p>Chemistry – Environmental EMSL at PNNL came on-line in 1997 with the mission of</p>	<p>EMSL’s unique combination of simulations, high field magnetic</p>	<p>As high-rate proteomic and nanoscale facilities and high-end</p>

understanding and controlling the molecular processes that underlie our environmental problems. In 1998 it put 250 GB to media.	resonance instruments, high-performance mass spectrometers, optical imaging instruments, and so forth generate 100 TB/year to media.	supercomputers come on-line, EMSL's rate of putting data to media will increase to 2 PB/year by 2008.
Genomes to Life In the area of proteomics and metabolites for Genomes to Life (GTL) there was less than 10 GB of data on-line in the world in 1998.	Proteomics and metabolomics currently are capable of generating 400 TB/year . GTL information for a single microbe generates 20 PB of proteomic data and 16 PB of metabolite data.	Proteomics and metabolomics data generation has the potential to increase to the level of tens of PB/year by 2008.

In addition to evaluating the needs of the application scientists for computing and network facilities, ASCR must evaluate the needs of the application scientists for research in applied mathematics, computer science, and advanced network environments to support science. Such an evaluation is particularly difficult because of the long time lag, often 10 years, between basic research in these areas and application by scientists. Therefore, the planning process must abstract the needs of application scientists, identify multiple research threads that might satisfy these abstract needs, and support the research until it matures. To accomplish this objective, ASCR uses multidisciplinary workshops, such as SCaLES, partnerships with application scientists, and the expert judgments of the program managers.

Partnerships and Pilots

Partnerships and pilots play a crucial role in the ASCR strategy to deliver advanced computing and an integrated network environment to scientists. The SciDAC effort builds on ASCR's decades of experience in building partnerships with application scientists, from the grand challenges of the HPCC initiative, though the scientific applications partnerships and collaboratory partnerships initiated in the mid-1990s, to the SciDAC partnerships initiated in 2001 and the Genomes to Life partnership with BER. These partnerships enable ASCR to transfer the best research results in information technology and applied mathematics to application scientists. In some cases this transfer has delivered a hundred fold capability increase to scientific applications. In addition, these partnerships have identified new areas for research in applied mathematics and computer science. Furthermore, these interdisciplinary partnerships enable systematic assaults on major scientific questions such as the relationship of global and regional climate, formation of soot in fossil fuel combustion, and design of new particle accelerators.

Nevertheless, we are still trying to arrive at optimal strategies for managing these partnerships. Two important issues are under discussion:

- **How to manage these teams.** We believe it could be useful to bring in organizational psychology professionals to help document the factors that make teams succeed and develop training for the future. Several members of ASCAC have proposed leaders from the commercial sector to help ASCR with these important issues.
- **How to divide the authority for funding decisions.** Effective decision-making requires effective division of authority between the Federal program managers and the project

leader in interorganization teams.

We expect that formal documentation of the SciDAC experience and processes will provide valuable ideas about formulating future partnership strategies.

In addition to these issues, a number of other issues are critical for partnerships to succeed:

- **Software engineering.** Since software is developed by teams, having appropriate software engineering practices becomes critical—particularly because many scientific software systems last for decades. Software engineering will become even more important as we move to more complex software developed by larger teams for entire communities. This is both an issue of choosing and enforcing an existing methodology for a team and a research issue of defining more effective ways of engineering scientific software. ASCR-supported research in software components is one approach to this problem.
- **Intellectual property.** Because the teams are creating copyrightable software, they must manage the intellectual property. ASCR believes that software developed under its auspices should be available to the research community under an open source license. We have not taken a position on whether a gnu-style license (where all derived products are also open source) or a Berkeley-style license (which permits development of proprietary derived products) is more appropriate.

Long-Term Support and Maintenance of Software

ASCR, as well as a number of other programs in the Office of Science, faces a significant issue in the maintenance of software that results from its research and partnerships. Much of ASCR's software is used by thousands of scientists. The research to develop improved versions of this software is clearly within ASCR's mission. The SciDAC mathematics and computer science Integrated Software Infrastructure Centers were required to have a software maintenance plan in their proposals. However, the post development support, maintenance, and testing of software such as PVM, MPICH, PETSC, and a number of linear algebra libraries is a significant challenge. In many respects this software is a new type of virtual facility that must be maintained if the full potential of the software is to be reached.

In the next three years ASCR will research and compare a number of models for this long-term support:

- Direct funding of support at DOE laboratories.
- Exploration of commercial models including small business. Perhaps an agreement that enabled a single company to commercialize all the software would generate enough revenue.
- Open source models, including establishment of free-standing, not-for-profit organizations to manage this software.

We emphasize that the software developed by ASCR benefits scientists funded by many other Federal agencies. Hence, the sale of support probably requires negotiation of MOUs with those agencies, similar to the MOUs DOE and NSF have negotiated to govern large, high-energy

physics projects.

We also emphasize that participation in and support for standards-setting activities are important for achieving broad interoperability—especially for open source models that, in their best instantiation, involve a wide community of users.

Managing Interagency Partnerships

As Figure 3 shows, DOE-SC is an integral part of the U.S. research infrastructure. Therefore, not only do ASCR research programs need to be viewed in the context of other IT research programs but ASCR facilities need to be viewed in the context of other national computing and network facilities. In the past this coordination has been most critical for network facilities, since ESnet relies on university networks to provide end-to-end service between the thousands of university scientists who use DOE facilities and those facilities. ASCR has accomplished this coordination through the Joint Engineering Team of the Large Scale Network coordinating group and its predecessors for over a decade. However, there has been only informal coordination between ASCR-funded computing facilities (see Figure 3) and those funded by other agencies.

One of the results of the HECRTF will be a closer coordination of high end computing research, facility and testbed activities. This coordination is required to support a large enough number of research and evaluation prototypes with different architectures (at least 5) to reduce the government-wide mission risk from uncertainty about computer architecture. Coordination will also be required to enable scientists to have access to the diverse advanced computers they need to make scientific progress, including systems that enable the largest calculations. These leadership-class systems must be managed by the host agency as a resource for the nation.

Because of the need for long-term planning to accomplish the installation and operation of high-performance computers, we believe that shared-funding models across agencies are inappropriate because they raise the risk.

In addition to the facilities activities cited, research activities funded by the ASCR program are coordinated with other federal efforts (see Figure 4) through the Interagency Principals Group, chaired by the President's Science Advisor, and the Information Technology Working Group (ITWG). The ITWG evolved through an interagency coordination process that began under the 1991 High Performance Computing Act as the High Performance Computing, Communications, and Information Technology (HPC/CIT) Committee. The Federal IT R&D agencies have established a 10-year record of highly successful collaborative accomplishments in multiagency projects and in partnerships with industry and academic researchers. The multiagency approach leverages the expertise and perspectives of scientists and technology users from many agencies who are working on a broad range of IT research questions across the spectrum of human uses of information technology. DOE has been an active participant in these coordination groups and committees since their inception, and the ASCR program will continue to coordinate its activities through these mechanisms, including an active role in implementing the Federal IT R&D FY 2002–2006 Strategic Plan under the auspices of the National Science and Technology Council and the President's Science Advisor.

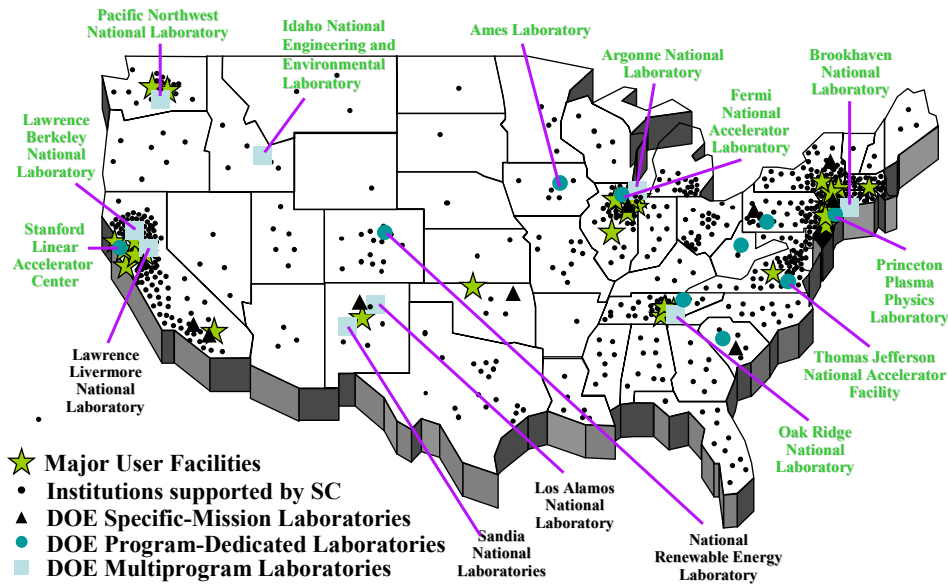


Figure 3: Major User Facilities

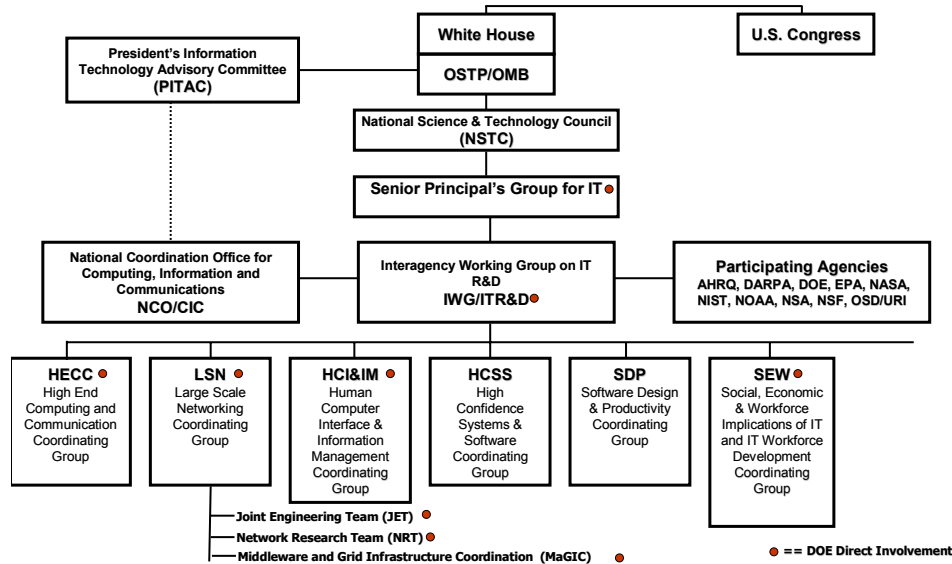


Figure 4: Organization Chart of Federal Agency Information Technology Research Coordination

In addition to the general coordination that occurs under the IWG framework, ASCR has a number of ongoing partnerships with DARPA and NSA in the area of high-performance computing. Furthermore, ASCR has a close partnership within the Department of Energy with the National Nuclear Security Administration. Areas of common interest include

- management of scientific data,
- analysis and visualization of petabyte (1 petabyte is 1 million gigabytes) datasets,
- computer operating systems, and
- mathematical algorithms and software for solving complex problems.

In order to ensure the effectiveness of ASCR and NNSA programs, an ongoing collaborative planning activity has been established. The objective of this joint planning is to identify areas of common interest and establish appropriate coordination of the efforts. This planning involves three general types of interaction:

1. Areas of basic research that are supported by ASCR and will be important to ASCI in the future but are not important enough to justify ASCI investment at this time. Examples of this type of effort include basic research supported by ASCR in the areas of programming models for parallel computers and next-generation message-passing protocols for parallel computers. In these areas, the appropriate level of interaction between the programs is information sharing so that when the appropriate time comes, ASCI can take advantage of the results from these efforts.
2. Areas of R&D where ASCI is making major investments and the results of these investments are more or less directly usable by the unclassified research community. Examples include ASCI work to test parallel programming tools on ASCI computers, PathForward partnerships between NNSA and industry, and ASCI-supported research on secure global parallel file systems. In these areas, ASCR's investments are focused on adapting and supporting the results of ASCI investments for use by the SC computational science research community.
3. Areas that are major research barriers to both ASCR and ASCI where research investments must be coordinated to ensure maximum effectiveness and leverage. Examples of this type of effort include software to manage scientific data, development of component architectures for high-performance software, and ongoing development of the High Performance Storage System (HPSS) in partnership with IBM.

Coordination is accomplished in several ways. First, program managers from ASCR and ASCI meet to inventory and discuss current work and categorize the efforts using documents such as ASCR project descriptions and the ASCI Implementation Plan. In cases where detailed coordination is required, program managers arrange joint site visits and participate in each other's review teams.

Program Integration and Priority Setting

To succeed in its mission, ASCR views scientific discovery as a process and—through its research, partnerships and facilities—provides insight and tools that support every aspect of this process. Figure 5 depicts the process of discovery and shows how different aspects of the ASCR portfolio contribute to this process.

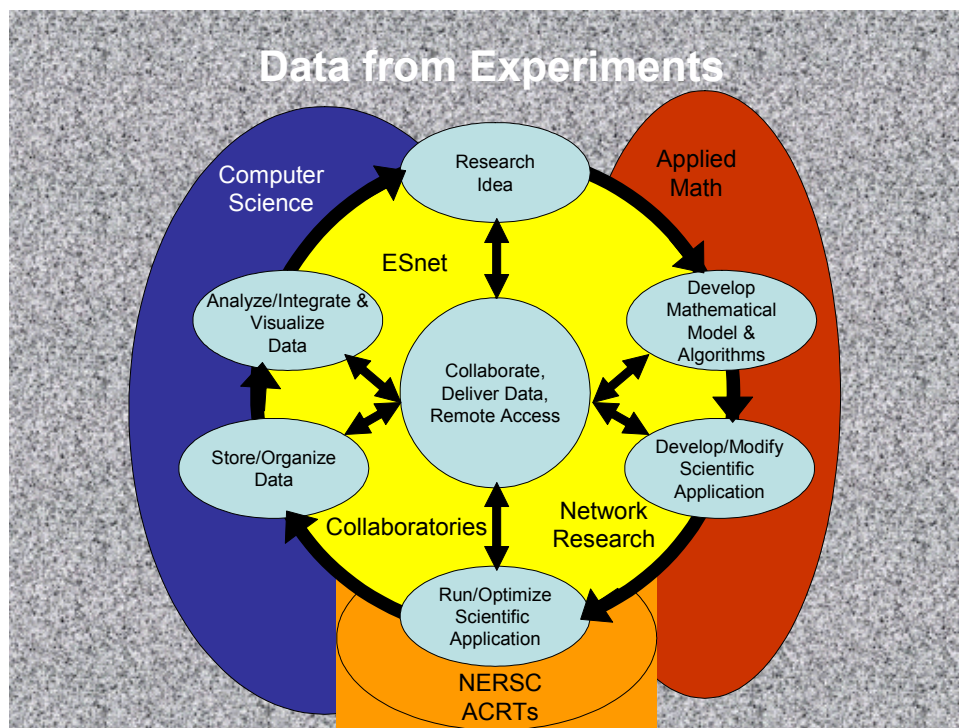


Figure 5: Discovery Process

The individual program element analyses in the following sections identify important opportunities and ways in which each of the ASCR research and facilities efforts contribute to the ability of the rest of the Office of Science to meet its missions. ASCR ranks and prioritizes these opportunities across the various research and facilities efforts to deliver the best program. In addition, because ASCR supports all aspects of the process of scientific discovery, a version of Amdahl's law pertains: "The pace of scientific discovery is determined by the slowest link in the process." Therefore, ASCR investments must be balanced between the areas.

A number of other strategic principles guide ASCR investments:

- Focus on investments, especially in facilities, that support multiple scientific disciplines.
- Abstract the needs of the application scientists to develop research programs that ensure that the pipeline of ideas for the future is full.
- Manage external risk through a diversity of approaches, especially in areas where ASCR is the lead Federal investor.

- Don't try to do everything. Rely on other agencies in parts of the research portfolio; however, constantly evaluate the risk.

To implement these principles, ASCR has organized its investments as shown in Figure 6.

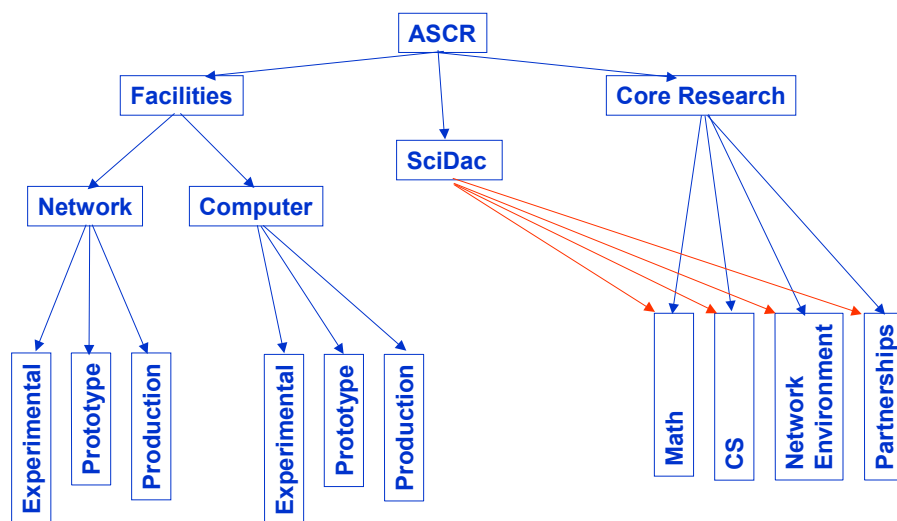


Figure 6: Organization of ASCR Investments

Not all of the elements are populated. In particular, ASCR has virtually no current investment in experimental or prototype networks or experimental computer facilities. A great strength of this methodology for allocating resources is that it enables effective management of the tension between supporting research with a 10-year horizon and supporting facilities that are in great demand.

However, this method of allocating resources has been less effective in coupling ASCR research to ASCR facilities. This situation is particularly important as the scientific community moves into a future with high performance computer facilities in the range of 50–1,000 teraflops and networks that must deliver terabit per second end-to-end performance. In order to address these needs, the current facilities strategy—where procurements of production computers are informed by, but independent of, investments in research and evaluation prototypes or software—is unlikely to succeed. Much closer coupling of these activities is required. In addition, the computer science research community will require testbeds that are independent of the research and evaluation prototypes. This future argues for a much more holistic view of the facilities, where a portion of the computer science–network environment research is tightly integrated with a portfolio of experimental, prototype, and production computing and networking facilities.

Computing and Network Facilities and Testbeds

To solve next-generation scientific problems, researchers will require a complex portfolio of facilities and testbeds. The elements of this portfolio are listed in Table 2.

Table 2: Elements of Facilities Strategy

Compute Facilities and Testbeds	Network Facilities and Testbeds
<p>Experimental</p> <ul style="list-style-type: none"> • Proof of concept • Small-scale research projects <p>Research and Evaluation (R&E) Prototype</p> <ul style="list-style-type: none"> • Sufficient scale to enable evaluation of scientific potential • Research projects • Enables new science for a few application scientists, who are willing to compute on a research system. <p>High-Performance Production Capability</p> <ul style="list-style-type: none"> • Stable, multiuser capability environment • Large user support, consulting and training investment • Direct support of agency mission <p>Leadership Class</p> <ul style="list-style-type: none"> • Most capable system available for a class of applications • Small number of projects • Resource for national science community, managed similar to light source or high energy physics facility <p>High-Performance Capacity</p> <ul style="list-style-type: none"> • Geometric mean of high-performance production and desktop 	<p>Experimental</p> <ul style="list-style-type: none"> • Proof of concept • Small-scale research projects <p>Research and Evaluation (R&E) Prototype</p> <ul style="list-style-type: none"> • Sufficient scale to enable evaluation of scientific potential • Research projects • Enables new science for a few application scientists <p>High-Performance Production Network</p> <ul style="list-style-type: none"> • Stable capability environment • Large user support, consulting and training investment • Direct support of agency mission

Even within the production computing facilities there must exist a range of capabilities, from the computers that can perform the largest and most demanding calculations for the first time anywhere, to the high-performance computers that enable scientists across the Office of Science to do the important science that does not require the most capable computers. ASCR focuses on large-scale facilities that are used by multiple disciplines. We do not consider the smaller mid-range systems that are purchased by individual research groups to be within our mission.

Figures 7 and 8 show how the individual elements of the compute and network facilities portfolios are interrelated.



Compute Facilities and Testbeds Timeline

Office of Science

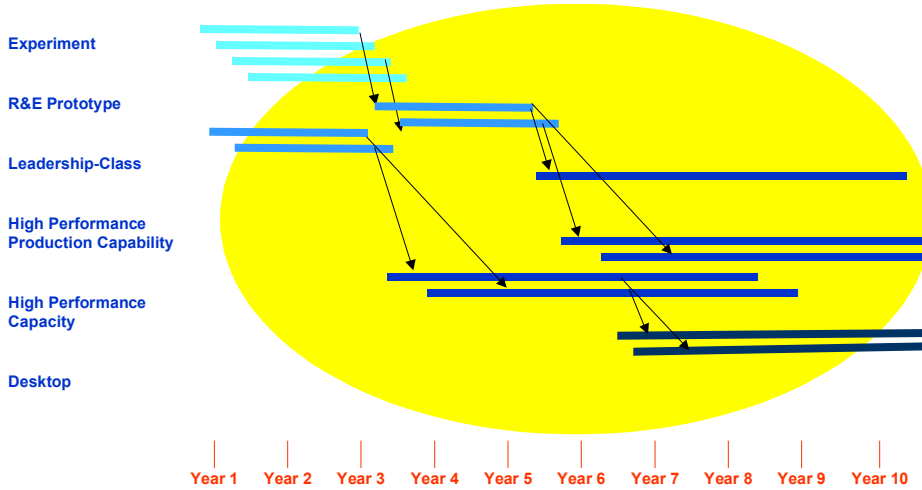


Figure 7: Timeline for Compute Facilities and Testbeds



Network Facilities and Testbeds Timeline

Office of Science

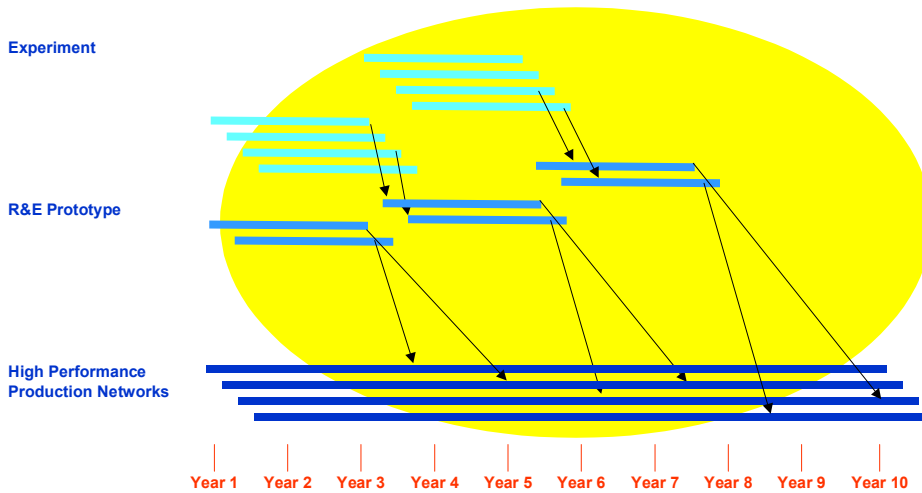


Figure 8: Timeline for Network Facilities and Testbeds

The future of high-performance computing facilities, in this vision, includes a portfolio of R&E prototypes, which must be managed as a portfolio across the Federal Government. Based on external review, near the end of the evaluation period, some of these systems become candidates either as high-performance production systems or as leadership-class systems. We envision that high-performance production systems will continue to be managed by individual agencies to support their mission requirements. Leadership-class systems, on the other hand, will be resources for the Nation and must be managed as such. The DOE Office of Science has a long history of providing this type of facility for the scientific community and believes that a computing capability at this scale is critical for scientific leadership. As with other major facilities, we believe that ASCR should be the sole developer and operate leadership-class facilities should be in ASCR, with access managed through peer-reviewed proposals. Funding models combining multiple appropriations to support major facilities were tried and abandoned in the 1970s.

In this vision of the future there must also be much tighter integration of software research—both for systems and for mathematical libraries—to enable the effective use of these computers.

Based on this vision, ASCR has the following strategy for high performance computer facilities and testbeds.

- Maintain ASCR’s investment in high-performance production capability computing. This capability is key to the ability of DOE to accomplish its science mission.
- Increase the number of R&E prototypes under evaluation, and strengthen ASCR’s coupling to R&E prototype evaluations funded by other Federal programs. One possible method for accomplishing this objective would be through a government-wide Performance Evaluation Research Center. R&E prototypes would be selected based on peer-reviewed research proposals, with strong interactions with the scientific disciplines in the review process.
- Institute a major review of each R&E prototype at approximately the three-year point to qualify the architecture for installation at either high-performance production capability or leadership scale. The results of these reviews would be broadly available across the government.
- Improve the coupling between software research and facilities, the SciDAC Integrated Software Infrastructure Centers.
- Implement a leadership-class computer as a resource for the Nation.

Detailed issues related to this strategy remain to be worked out.

- “Managing” interagency portfolio of R&E prototypes to ensure competition and breadth. Federal agencies will need to coordinate proposal review, evaluation, and access more closely than in the past.
- Determining what to do with successful R&E hardware. After the three-year review, this

hardware will still have a number of years of useful life, at least for a small class of applications. Because the high-performance production or leadership-class follow-on most likely would be updated hardware, perhaps under an option in the original contract, the existing hardware should be transferred to the high-performance production capability or capacity portfolio. This transfer probably would not involve physically moving the hardware; however, the level of software integration achievable remains an issue.

- Managing the transition from R&E prototype to high-performance production or leadership scale. We anticipate that the decisions about what architecture to install at either scale and where to site the hardware will need to be managed independently. High-performance production and leadership-class installations may have significantly different requirements in terms of issues such as user support, system management, software and hardware infrastructure, space, and network access from locations that conduct successful R&E prototype evaluations. Therefore, each alternative needs to be reviewed separately. This type of transition is different from the current procurement process in which the production centers conduct requests for proposals that are independent of any information from the R&E centers. This raises a number of issues related to ensuring appropriate competition in the marketplace to deliver hardware cost-effectively and engage a number of vendors in the process.
- Maintaining world-class support and consulting organization for high-performance production capability. One of the most critical resources at the high-performance production capability installations is the staff. Effective consulting, user support and training are critical to the scientists working on these facilities. Therefore, independent of where a leadership-class computer is located, incentives are needed to prevent the erosion of this staff.
- Maintaining the capability at at least two locations to conduct evaluations of R&E prototypes. This capability includes teams of experts, and physical and supporting computer hardware infrastructure. It is simply not sensible or cost effective to site R&E systems at institutions that do not have this capability in place.
- Managing access to leadership-class computers. Leadership-class computers are resources for the Nation; therefore, such resources must be used only to do important science that cannot be accomplished at other facilities. Peer review is the accepted method for judging scientific quality and must be a core of this process. However, peer review works best when proposals from a single discipline are under consideration. In addition, because the leadership-class systems are resources for the Nation, there will be situations in which the individual agencies have extremely high-priority tasks such as design of multibillion dollar experiments or analysis of critical incidents such as the space shuttle Columbia. Detailed access procedures would need to be established between the agencies and OSTP.
- Managing an inherently distributed infrastructure. Managing a high-performance computer is a complex task moving to a future where these resources are distributed poses a number of significant challenges in policy, configuration management, distributed file systems, and the integrated network environment.

The network facility and testbed strategy is similar to the strategy for high performance computer facilities with high-performance production networks, R&E prototypes, and experimental networks. However, the management issues are different.

- Managing an interagency portfolio from experiment to R&E prototype to production. Many of the issues involved in providing end-to-end performance (the only kind of performance the application scientists care about) require interagency and federal-commercial coordination to resolve.
- Inserting specialized requirements for science into a global, multi-Autonomous System environment. The DOE Office of Science in particular and the U.S. scientific community in general are a small part of the global network market. Therefore, inserting new technologies into the global infrastructure is a significant issue.
- Managing the transition from R&E prototype to high-performance production network. The high-performance network that scientists rely on needs hardened production software, and its performance measures are different from those of prototypes.
- Managing complex new provisioning models and expansion up the protocol stack. New provisioning models for DWDM include leasing wavelengths and owning fiber. These models transfer the management responsibility for the physical layer of the network back to ESnet (possibly with significant performance and cost benefits). In addition, many of the core network services above the transport layer need support in the evolving Grid environment. Finally, at levels above 10 mbps, TCP may not be an appropriate protocol because of inherent limitations in its error control and flow management capabilities.

Research

Within a broad range, portfolio investments in research are equivalent to investments in facilities. For example, in certain areas of solvers for elliptic problems, the historical data suggests that the increased capability resulting from mathematical methods is comparable to the increase resulting from computer hardware improvements. Similarly, advanced scheduling software that maintains time to solution for large jobs and delivers increased use of the hardware is equivalent to having more hardware. In addition, because the mission of ASCR is an end-to-end one, investments in mathematics, computer science, and the integrated network environment can all be evaluated on their potential to accelerate scientific discovery.

Nevertheless, two factors make a comparison of investment in research vs. facilities challenging. First, the time scale between research and deployment is long. Our experience is that introduction of new ideas, such as computer languages, takes at least 10 years of sustained investment. In addition, it is almost never possible to predict what performance improvement a basic research project will deliver in 10 years. This is an inherent issue with evaluating basic research. Therefore, many quantitative methods for balancing the research portfolio are simply not applicable.

Two general principles from real options theory do have value in a basic research portfolio related to the impact of risk in decision making. The first principle is that uncertainty virtually

always adds value to the option. The second is that risks that are external to or internal to the decision process affect the decision process in different ways.

For example, external uncertainties such as the price of computers, number of undergraduates entering technical disciplines, or availability of dark fiber are not affected by ASCR decisions vis-à-vis applied mathematics research or investments in ESnet. In this case, the most important factor to consider is the downside risk, because waiting always reduces the uncertainty. Thus, computer acquisitions must be timed to take advantage of external factors.

On the other hand, uncertainties about the applicability of wavelet approaches to quantum chemistry are strongly coupled to decisions by ASCR. In this case, the only way to reduce the uncertainty is to invest, and the most important factor to consider is the value of success.

Based on these considerations and on the need to balance evolutionary and revolutionary approaches, the research portfolio must build on its current successful programs to address:

- Computers with tens to hundreds of thousands of processors – operating systems, systems software, mathematical algorithms, and tools to enable scientists to use and program them
- The mathematics of multiscale and complex systems – systems that occur across the Office of Science, for which simply adding computer power to the mathematical models of today cannot succeed
- Petascale and exascale data management, analysis and visualization – file systems, intelligent data movement, high-dimensional data, feature extraction, visualizing complex datatypes
- The integrated network environment of the future – high-speed transport, navigation in cyberspace, security, and usability by applications scientists

Analyses of Program Elements: Strengths, Weaknesses, Opportunities, and Threats

Applied Mathematics Research

Contribution to Overall ASCR Strategic Goals

The Applied Mathematics Research (AMR) program element delivers mathematical analysis and algorithmic tools to enable rigorous understanding and high-fidelity simulation of physical, chemical, and biological processes of interest to the Office of Science and to the Department of Energy as a whole. This program element has two components: a base research activity and a number of SciDAC Integrated Software Infrastructure Centers (ISICs).

The investments made in the AMR base span the entire range of research and development of both fundamental applied mathematical techniques and robust and efficient numerical software, including the following:

- Development of well-posed mathematical models to describe linear and nonlinear physical systems with ever increasing complexity.
- Rigorous mathematical analysis of the behavior of these models under various conditions, including existence of solutions, definition of appropriate boundary conditions, and analysis of the presence of singularities.
- Development of solvable discrete versions of these (generally continuous) mathematical models that are appropriate for translation into computational simulations. This work includes the generation and iterative improvement of underlying computational meshes and the subsequent discretization of the continuous equations into discrete representations of computationally efficient basis elements.
- Development and software expression of efficient algorithms for solving the discretized models. These algorithms include basic elements such as numerical integration, solutions of linear systems of equations, eigenvalue and eigenvector computations, time-integrators, and nonlinear solvers.
- Rigorous analysis of the sources and magnitude of error in the computational solution. This work includes predictability analysis and uncertainty quantification, both for model reduction and for determination of levels of confidence in the results.
- Development of optimization techniques for engineering design optimization, for discrete optimization problems, and for optimization problems under a complex array of constraints.
- Exploration of new areas of mathematical analysis and computational algorithms, dictated by need or opportunity. The AMR program element has invested or is currently investing in such far-ranging areas as dynamical systems, multiresolution analysis, multiscale mathematics, and ultrascale algorithms.

These investments are *required* for ASCR to deliver high-performance computing to advance the frontiers of science. In the one area that has been most extensively evaluated, the solution of a discretized system corresponding to an elliptic operator, the contribution to scientific capability from mathematical advances over the past three decades is at least as great as the advances enabled by hardware speedup (see Figure 9).

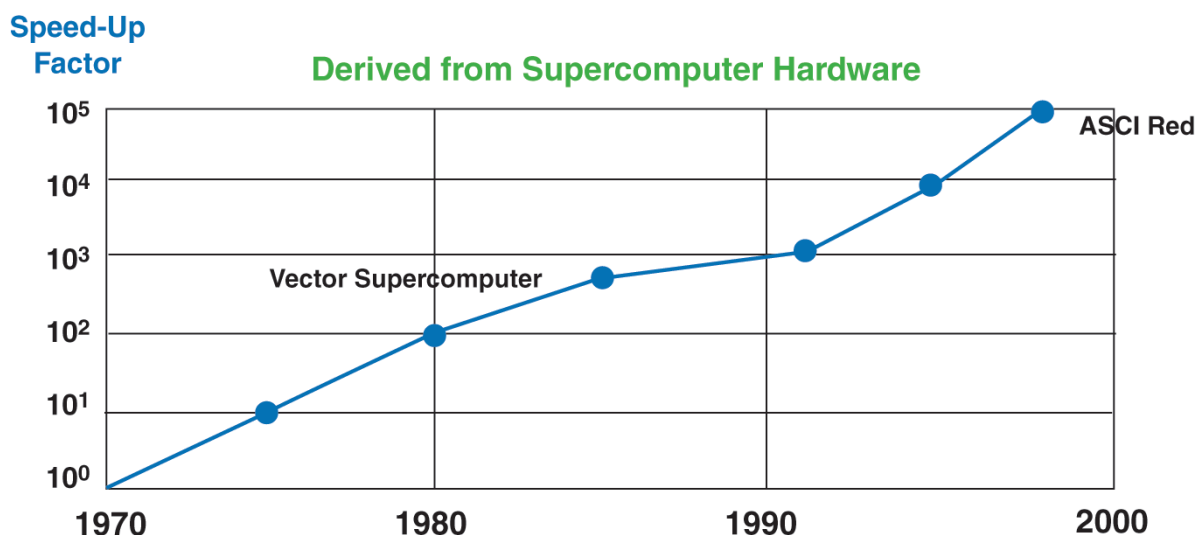
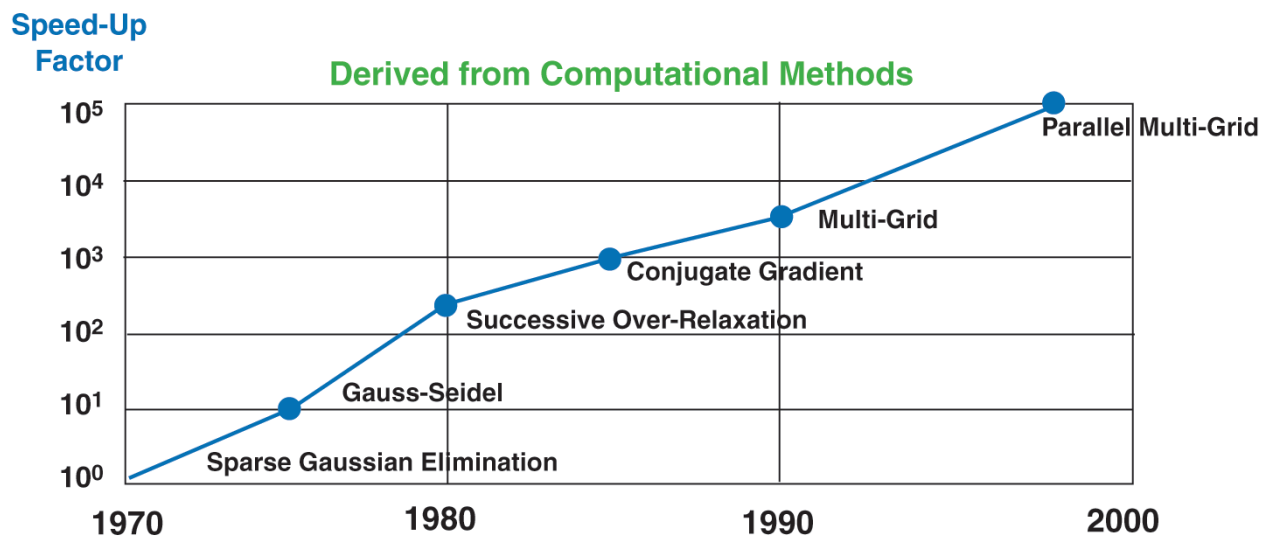


Figure 9: Speedup over the past 30 years in (a) computational methods and (b) supercomputer hardware

The ISIC program focuses on research, development, and deployment of software to (1) accelerate the development of SciDAC application codes, (2) achieve maximum efficiency on terascale computers, (3) enable a broad range of scientists to use computational modeling and simulation in their research, and (4) protect long-term investments in these codes. ISIC investments support applied mathematics research and development to bridge the gap between the advanced computing technologies developed by high-end research community and the computational needs of SciDAC applications. The activity supports three applied mathematics research efforts: the Terascale Simulation Tools and Technologies (TSTT) Center, the Algorithmic and Software Framework for Applied Partial Differential Equations Center (APDEC), and the Terascale Optimal PDE Simulations (TOPS) Center.

Planning Horizon

The base research activity in the AMR program element must maintain a balance between short-term (1–3 year) and long-term (3–10 year) horizons for planning purposes, whereas the ISIC activity is focused on the shorter term only.

Short-Term Planning. Some investments, such as improved solver techniques (including advanced preconditioners), mesh generation, and the development of numerical software tools for delivery of advanced algorithms to applications, should have short-term payoff. Identification of appropriate short-term investments is generally dictated by the immediate needs of the applications scientists (e.g., faster solvers, better meshes for discretizing mathematical models) or by the evolution of promising lines of research begun earlier (e.g., automatic differentiation, predictability analysis). In particular, fruitful lines of research that were once in the long-term planning process ultimately transition to short-term planning.

Short-term planning is relatively straightforward. Periodic workshops that bring together researchers in a specific area of applied mathematics can provide valuable input for planning the future course of research in that area. For example, the applied mathematics program recently conducted a workshop on predictability analysis and uncertainty quantification to bring the community together to assess the current state of research and to provide input on the future challenges in this area.

Long-Term Planning. Maintaining a vibrant and effective basic research program in applied mathematics requires identification of future barriers to scientific progress in modeling and simulation, before those barriers are reached by the current state of the art. For example, the AMR program started investing in scalable algorithms beginning in the early 1980s, nearly a decade before multiprocessing became the preferred means of attaining high performance in computational science. Similarly, investments in predictability analysis were motivated by the recognition that output from large-scale, highly complex computational models had to be accompanied by some measure of confidence in the results.

More recently, three barriers to progress have emerged in computational science that must be addressed through applied mathematics research: multiphysics applications, multiscale mathematics, and ultrascale algorithms. New mathematical analysis and new algorithmic techniques for handling *multiphysics applications* (which involve the coupling of disparate models) or *multiscale mathematics* (which involves models that have a wide range of interacting scales) are highly complex and will take many years to mature to a state where they directly improve our ability to solve scientific problems. Still other investments are higher risk, but have a substantial likelihood of payoff if progress is achieved. Examples include *ultrascale algorithms*—algorithms that effectively scale to tens of thousands or hundreds of thousands of processors.

In all cases, the long-term planning is fraught with challenges, since the proper avenues of approach for dealing with interacting collections of highly complex physical models on many scales often is not known. Workshops can provide valuable community input, both from the mathematics community and from the applications.

Ultimately, however, the program must be sufficiently flexible to pursue several avenues at once. Indeed, without a careful balance between short- and long-term investments, emerging barriers to scientific advancement in computational science will go unchallenged.

Current Areas of Investment

The AMR program element base activity currently invests in two major categories of applied mathematics research: analysis and numerical algorithms.

- **Analysis.** Investments in analysis are intended to provide a basic understanding of the behavior and characteristics of mathematical models of physical systems, independent of their expression in computable code. Examples include research into fluid dynamics, general conservation laws, equations of state, and boundary conditions, all falling under the heading of “analysis of PDEs.” Additionally, some research in “analysis of algorithms” is included, such as analysis of discretization techniques, mathematical treatment of singularities, and theoretical analysis of scalability of algorithms.
- **Numerical Algorithms.** Numerical algorithms research includes the development of advanced numerical software for the solution of PDEs and integral equations using techniques such as mesh generation, refinement and evaluation tools, the solution of systems of algebraic equations, multigrid techniques, boundary integral methods, and asymptotically fast methods.

In addition to these two categories of research, the applied mathematics research program element invests in the development of human resources, through several programs targeted at young researchers, including graduate students, postdoctoral fellows, and academic researchers early in their research careers.

The following list summarizes the current portfolio of investment in the basic research program of applied mathematics within ASCR:

PDEs	Fast Methods (e.g., FFTs, Fast Multipole, multigrid)
CFD	Scalable Methods
Meshing & Adaptive Mesh Refinement	Software Tools (e.g., PETSc, TAO, EBCombo, MPSalsa, LOCA, Trilinos, Hypre, SuperLU, FronTier)
Solvers (linear, nonlinear, eigenvalue)	Future Workforce Development
Optimization (continuous and discrete)	<ul style="list-style-type: none"> • Five named fellowships at national laboratories • Early Career Principal Investigator (ECPI) Program • Computational Sciences Graduate Fellowship (CSGF) Program
Dynamical Systems	
High-Performance Computation	
Automated Reasoning	
Boundary Integral Methods	
Interface tracking methods (e.g., FronTier, Level Set)	
Statistics	
Predictability Analysis / Uncertainty Quantification	

The ISIC activity currently invests in three areas:

- **Terascale Simulation Tools and Technologies (TSTT) Center** – Enable application scientists to easily use multiple mesh and discretization strategies within a single simulation. Examples focus on high-quality hybrid mesh generation for representing complex and evolving domains, high-order discretization techniques for improved numerical solutions, and adaptive strategies for automatically optimizing a mesh to follow moving fronts or to capture important solution features.
- **Algorithmic and Software Framework for Applied Partial Differential Equations Center (APDEC)** – Develop a high-performance algorithmic and software framework for solving partial differential equations arising from problems in fusion, accelerator design and combustion. This framework will provide new simulation capabilities based on locally structured grid methods, including adaptive meshes for problems with multiple length scales; embedded boundary and overset grid methods for complex geometries; efficient and accurate methods for particle and hybrid particle/mesh simulations; and high-performance implementations on distributed memory multiprocessors.

Terascale Optimal PDE Simulations (TOPS) Center – Develop and deploy an integrated toolkit of open source, near-optimal complexity solvers for nonlinear partial differential equations in SC applications. Of particular importance are multilevel methods to reduce computational bottlenecks by one to three orders of magnitude on terascale systems. The focus is on solution of nonlinear PDEs, PDE eigenanalysis, and optimization of PDE-constrained systems. The goal is to attain highest possible computational efficiency by side-stepping memory bandwidth limitations of hierarchical memory architectures. There are five areas of concentration: adaptive time integrators for stiff systems, nonlinear implicit solvers, optimization, linear solvers, and eigenanalysis.

Transfer of Knowledge to Application Scientists

In a mission-oriented agency such as DOE, transferring research results to applications is an important component of maintaining a healthy research program. The AMR program element strives to balance immediate relevance with long-term research goals.

For the basic research activity, knowledge transfer occurs naturally in several different ways:

- Active collaboration with applications scientists as part of an interdisciplinary team of researchers in computational science
- Participation in software development, usually both as a testbed for new ideas or algorithms and as a mechanism for deploying the resulting algorithms to applications scientists
- Complementary funding expressly designed for knowledge transfer or for strengthening the interdisciplinary nature of a team pursuing a common goal
- Strong motivation by a specific scientific or engineering challenge

- Strong connection between an academic researcher and a DOE national laboratory or other scientific research facility

We emphasize, however, that technology transfer cannot be a litmus test for the base program, since this would allow for the support of only short-term research.

For the ISIC activity, on the other hand, transfer of the technologies to the SciDAC applications teams is a critical measure of success and is managed through shared personnel and explicit funding of mathematicians on the applications teams to ensure close coupling.

Strengths

The AMR program element has been supporting world-class research in applied mathematics for over 50 years. Following the suggestion of John von Neumann, the AEC put in place the Applied Mathematical Sciences (AMS) program in the early 1950s. Ever since, the DOE AMS program has played a key role in developing the algorithms, software, and other mathematical tools that underlie high performance computation and simulation for science and engineering applications. In the 1990s the AMS subprogram of the Basic Energy Sciences program was integrated into the newly formed Office of Advanced Scientific Computing Research. The AMR program element is unique because it promotes close ties between academic research and research at the DOE national laboratories, building on their complementary institutional strengths, and because it encourages a close interaction between mathematicians and the applications community.

Other important strengths of the AMR program include the following:

- A long-standing and firm commitment from DOE's Office of Science, with explicit recognition that a basic research program in applied mathematics is of fundamental importance to the mission of the Office of Science and of DOE.
- The dedication of scores of talented researchers who are highly productive in achieving research results but who also feel a sense of ownership and are eager to help nurture the program.
- A very high visibility of SC-supported researchers at both national and international meetings. The applied mathematics research program element supports a very high percentage of plenary speakers, invited speakers, and conference organizers of leading conferences in applied mathematics.
- A continuing history of support for strong research teams.
- A DOE mission focus that provides direction.
- A history of supporting research motivated by the capability of interdisciplinary teams of researchers to multiply their achievements.

Weaknesses

The AMR program element is under increasing pressure to support research whose benefits are immediately apparent. Though striving to maintain a long-term component of the program, the pressure to provide measures of progress tends to lead to insufficient investment in speculative

or “risky” investments that have the potential to address the most challenging barriers to progress in computational science.

Other weaknesses include an underinvestment or lack of investment in several critical areas:

- No investment in mathematical techniques and algorithms for dealing with mathematical models of physical phenomena occurring at a wide range of overlapping and interacting scales (multiscale mathematics)
- No investment in the development of numerical algorithms that scale effectively to tens or even hundreds of thousands of processors (ultrascale algorithms)
- Underinvestment in all areas of discrete mathematics, particularly where techniques in discrete mathematics can address scientific problems with an asymptotically exponential computational complexity
- Underinvestment in statistics
- Underfunding of almost all research teams at national laboratories, as concluded by a recent review of one portion of the AMR program element laboratory programs.

Opportunities

The AMR program element has several opportunities for growth and exciting new research on the near horizon.

- Investments in multiscale mathematics now can remove roadblocks years down the road. Funding for this is included in the FY 2005 President’s Request. We have initiated a series of workshops to develop directions for this effort.
- Identification of some investments in discrete mathematics and combinatorics could conceivably have huge payoffs for some applications, including the understanding and modeling of metabolic networks, homeland defense applications, and the national infrastructure).
- Encouragement of more women and minorities in applied mathematics research.

Threats

Among the main threats faced by the AMR program element is an ongoing erosion of its ability to maintain identity from other agencies’ research programs. In addition, the pressure to shorten the “payoff” horizon for research results has serious implications for the ability of the program to study the questions that will result in the deployment of new software tools to applications in the next decade.

Planning Process

Planning for the AMR program element is the responsibility of the Federal program manager, who with the rest of the program managers in the Office of Advanced Scientific Computing

Research must identify the future requirements of the DOE and the Office of Science in advanced computing. Input from the computational science research community is essential to this process and is obtained in several ways. Leading conferences on various topics in computational science and mathematics provide community input to the program manager. In addition, the ASCR program frequently convenes workshops whose participants provide a written report summarizing the conference findings. Recent examples include a series of workshops on the role of mathematics in the Genomes to Life (computational biology) program, a workshop on the role of mathematics in the theory and modeling of nanoscience, and a workshop on the role of computational science in fusion energy.

The Advanced Scientific Computing Advisory Committee (ASCAC) is also responsible for providing input to the program management of ASCR regarding critical research elements that should be considered for the program.

Gap Analysis

The following gaps in the program have been identified:

- **Multiscale mathematics** – investments in this area will be required to address highly complex phenomena spanning many interacting length scales and time scales. Computational models are evolving, from a single module representing a subset of behavior of a physical system, to comprehensive and high-fidelity models capable of simulating the entire physical system.
- **Ultrascale algorithms** – investments in this area will be required to ensure that future generations of high-end architectures that must necessarily involve many thousands of processing elements are optimally used by applications simulations.
- **Discrete mathematics** (and its role in mainstream scientific simulation) – investments in this area have a significant potential to make orders-of-magnitude gains in the computational complexity of some applications areas, including computational biology and homeland defense.
- **Statistics** – investments in this area are required to deal with extracting knowledge from the oceans of data that large-scale simulations will produce.
- **Multiphysics** – investments in this area are required to properly formulate complex models of multiphysical systems, either by coupling existing modules of separate subsystems or by designing new models capable of representing the entire range of physical behavior of a highly complex system.

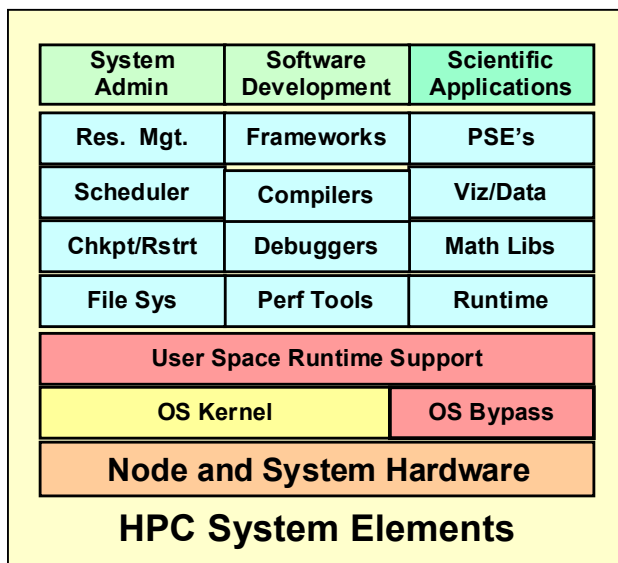
Computer Science Research

Like the AMR program element, the Computer Science (CS) program element consists of two major components: the base program and the Computer Science ISICs.

The base program supports academic and laboratory research efforts that enable DOE scientists to use very high performance computers effectively and efficiently and to produce timely computational results that support, complement, and extend experimental and theoretical scientific advances. Research is supported in the context of a vision of a comprehensive and

unified computational environment for massively parallel computers that enables computational scientists to effectively use the most advanced computers available without being overwhelmed by system inefficiency and complexity. Topics of interest include programming models, language and library interoperability, performance evaluation and modeling, data management and parallel I/O, scalable operating systems and systems management, scientific visualization, feature detection, and data mining.

Figure 9: System Software Elements



The ISIC program focuses on research, development, and deployment of software to (1) accelerate the development of SciDAC application codes, (2) achieve maximum efficiency on terascale computers, (3) enable a broad range of scientists to use computational modeling and simulation in their research, and (4) protect long-term investments in these codes. ISIC investments support computer science research and development to bridge the gap between the advanced computing technologies developed by high-end computer vendors and the computational needs of SciDAC applications. The activity supports four computer science research efforts: high-performance software components for interoperability, high-end performance evaluation, scientific data management, and scalable systems management for terascale clusters.

Contribution to Overall ASCR Strategic Goals

Computational modeling and simulation are particularly important for the solution of problems that are insoluble by traditional theoretical and experimental approaches, hazardous to study in the laboratory, or time-consuming or expensive to solve by traditional means. This program element provides research, development, testing, and evaluation of high-performance software infrastructure to support the effective use of high-performance computing systems for scientific applications. The results of this program element enable researchers in the scientific disciplines to analyze, model, simulate, and predict complex physical, chemical, and biological phenomena important to the DOE mission. CS research efforts provide the software infrastructure that

computational scientists build on to efficiently implement these models on the highest-performance computers available and to store, manage, analyze and visualize the massive amounts of computational data that are produced. Successful CS research leads to software that is made available to application scientists at NERSC, ASCI terascale systems, and other computational facilities throughout DOE.

The ASCR Computer Science program element is the largest and most active computer science program in Federal research, with the goal of with making computational science an effective third branch of science fully integrated with experimental and theoretical science. Research activities are motivated by software issues arising from next-generation architectures—architectures that will be increasingly complex and difficult to use but that offer the promise of much-needed performance improvements. The CS program element has a long record of making high-quality open-source software available to DOE application scientists, the academic research community, and the high-end computer vendor and third-party software industries.

Planning Horizon

Application scientists want to do research in their own fields of science, not computer science. They require their software to be stable and long lived and are rarely interested in investing significant amounts of time in mastering complex research-quality software tools or in making many modifications to their code to incorporate new software technology that promises only modest improvements in execution performance. Consequently, the process of research, development, testing, and evaluation of high-performance software infrastructure is lengthy and requires stable investment over a period of many years.

The typical planning timeframe of the CS base program is ten years or longer. In some areas, such as programming language research, ten years is a minimum timeframe from inspiration to full implementation that is broadly used by application scientists. Recognition of these time and support requirements is fundamental to the success of the CS base research program.

The planning horizon for the ISICs is different. The approach adopted by the SciDAC program is to take a snapshot of some of the best results of CS base research and focus them on the application of terascale computers to key applications in the Office of Science. Success is measured first and foremost by application progress, not by computer science advances, and the planning horizon is shorter, typically five years. Lessons learned from the successes and failures of applying advanced CS methods to SciDAC applications are extremely valuable and help shape the future directions of the base research program.

Current Areas of Investment

The CS program element currently invests in five major areas, each of which represents approximately 20% of the program. These areas are listed below together with examples of currently funded activities. At present, about 25% of the investments are in university research and 75% are in national laboratory research.

- Interoperability/Portability Tools

- Common Component Architecture ISIC – High-performance parallel component technology for enabling interoperability and reuse of simulation software and for enabling large-team scientific software development
 - Performance Evaluation and Optimization
 - Performance Evaluation Research Center (PERC) ISIC – Memory-based methods for performance characterization of system architectures and scientific application to better predict and improve application performance
 - Paradyn/Dyninst –Binary-based performance monitoring and optimization tools
- Systems Software Environment
 - Scalable Systems Software ISIC –Scalable resource management, job initiation, system monitoring, and other system management tools for high-performance clusters having thousands of nodes
 - Cougar lightweight kernel (LWK) evaluation – Characterization of the application performance impact of an LWK vs. a full operating system kernel
 - Science Appliance SSI cluster –Cluster software that delivers an effective single system image to the user
- Programming Models/Runtime
 - Programming Models Center – Improved performance and interoperability of major parallel programming models (MPI, OpenMP, Global Arrays), emerging models (UPC, Co-array Fortran, Titanium); development of common runtime support
- Visualization/Data Understanding
 - Scientific Data Management ISIC – Improved high-performance input/output and secondary storage systems, efficient data-subset retrieval, scalable data-management methods for petabyte-scale datasets, and feature detection and data-mining algorithms
 - Scalable rendering via clusters of Linux or Windows-based PCs and commodity graphics cards
 - New visualization techniques and software frameworks allowing the graphics community to share tools and techniques across application domains
 - Coupling of visualization tools with collaborative environments, such as the Access Grid

Transfer of Knowledge to Application Scientists

A fundamental method of transfer of knowledge to application scientists is through open source code made available through various Web-based mechanisms. This program element has a strong record of community involvement, through workshops and similar face-to-face mechanisms, of engaging the application community in helping define the problem and the promising approaches to solution. The success of MPI is based, in no small measure, to such an approach. Another fundamental mechanism is providing direct support and training to application scientists. This activity is based at NERSC and is an essential mechanism to help bridge the gap between CS tools and end users. Other important transfer mechanisms include

publications, PI meetings, laboratory staff interactions, conferences, invited talks, program reviews, workshops, joint application/computer science SciDAC PI meetings, poster sessions at technical meetings and workshops, and program management emphasis on application relevance.

Strengths

The CS program element has a long and distinguished record of accomplishments and has received frequent recognition, including several R&D 100 awards, Gordon Bell prizes for high-performance computing, and best paper awards at the SCxx series of conferences.

The CS program element has produced numerous fundamental advances in high-end systems: the first interactive operating system for high-performance computers (CTSS at NERSC); development of the HiPPI high-performance interface between supercomputers, storage systems and other devices; creation of the Parallel Virtual Machine (PVM) model for enabling distributed computers to work in concert on a single problem; and leadership in the creation of the Message Passing Interface (MPI) programming model, which has become the lingua franca of scientific parallel computing. The MPICH version of MPI developed at Argonne National Laboratory is an open source version of MPI that is the basis for all computer vendor versions of MPI. Similarly, Argonne's ROMIO implementation of high-performance parallel input/output has become a standard for the high-performance computing community. Recently the CS program element has supported a joint effort called Open Source Cluster Application Resources (OSCAR), involving universities, laboratories, and industry (including Intel, IBM, Dell, and SGI), that has released an open source, easy-to-install software package for high-performance cluster management and operations; OSCAR already has been widely adopted.

The program is recognized by the high-performance community as having particular strengths in a number of areas, including performance modeling, evaluation and optimization; parallel programming models, especially MPI and Global Arrays; parallel operating system environments; and high performance component architecture. The program is as comprehensive as resources allow and is tightly focused on high-performance computing research issues. It supports many large research activities that are required to effectively address many of the fundamental technical challenges, and there are frequent partnerships between the national laboratories and the academic community.

The CS program also is very active in the interagency community and has strong ties and relationships with the DARPA HPCS program, the NNSA ASC (ASCI) program, NSA research programs, and the DOD Defense Modernization program. Coordinated or cofunded research activities are under way with most of these agencies. The CS program was a major contributor to the national High-End Computing Revitalization Task Force.

Weaknesses

The CS program element is making a significant investment in scalable software, but the evaluation of this software is an ongoing challenge and is emerging as a significant weakness because of the lack of adequate test facilities. The largest available testbed has only 256 nodes and 512 processors. Even with the use of virtual-node mechanisms, such a system is totally

inadequate for testing system software and tools that need to scale to 10,000 to 100,000 processors. It is very difficult, if not impossible, to conduct software tests on production machines because these tests frequently cause a system crash—an unacceptable situation in a production environment. More than one testbed is required because architectural differences, such as support for a global address space, cannot be accommodated in a single system. In addition, special test capabilities are needed to support data management and visualization research. The lack of adequate resources to fund the breadth of essential test facilities is the most prominent weakness of the program.

Other ongoing challenges include the need to promote a CS research community that is focused on the high end and that is recognized in academic departments. No professional computer science society encompasses both high-end computer science research and application. In other words, there is no CS analogue of the Society for Industrial and Applied Mathematics. It remains an ongoing challenge to effectively engage far-flung application communities, both in the problem definition phase and in the technology transfer phase. The lengthy technology development pipeline for high-end software infrastructure exacerbates this difficulty. Programs such as SciDAC that are focused on engaging both communities are a promising step, but much remains to be done.

Opportunities

The CS activity will be continually challenged to anticipate the application performance characteristics of next-generation computer architectures and next-generation scientific applications. As a consequence of Federal Government investment, including the NSA support of the development of the Cray X1 and the DARPA High Productivity Computer Systems program, we are entering an era of much greater high-end architectural diversity. This diversity—plus the continuing growth of high-end clusters to encompass tens of thousands of nodes, hundreds of thousands of processors, and deep-memory hierarchies—will require renewed emphasis on improvements in sustained application performance, portability, functionality, and fault tolerance. Specific areas of research interest include the following:

- Global parallel operating system and runtime environments that are designed to satisfy the enormous demands of petascale systems for scalability, reliability, and fault management
- Common, standardized compile and runtime infrastructures and tool interfaces
- Fault tolerant software that also enables both computer systems and applications to automatically recover from node failure and to run to completion
- Problem solving environments that can be readily customized for specialized application needs and that are capable of incorporating and comparing computational and experimental results
- Intelligent program development environments that incorporate expert systems and other artificial intelligence methodology to automatically offer guidance and choices for all aspects of application development, including data structures and mathematical methods
- Data management, file system, input/output, and scientific visualization software

infrastructure that tracks vendor technology and translates mass market capability to scientific usage

- New programming models and languages that are accepted by the application community

Threats

From a scientific software infrastructure perspective, it is entirely possible to have too many system architectures. For the previous decade or so, almost all high-end systems in the nation were based on relatively minor variations of a single architecture—clustered symmetric multiprocessor (SMP) systems. We are now entering an era in which there are vast differences in scale (Blue Gene/L with 128,000 processors) or node/interconnect functionality (ASC Red Storm and the Cray X1), and next-generation petascale systems from the DARPA HPCS program are likely to incorporate substantial new levels of complexity. The software infrastructure must continually adapt and advance to shield the application scientist from complexity and scale while at the same time enabling the application scientist to maximally benefit from new architectures. Many experts predict the end of Moore's law in the 2015 timeframe, so it is likely that architectures will continue to evolve in new directions as fundamental barriers to performance are reached.

Coping with petascale scientific data requires new computer science approaches to scientific data management. One issue is that as dataset sizes increase, human ability to scan such datasets is quickly overwhelmed, even with the best visualization tools available. Thus, we must develop efficient and advanced data extraction and analysis capabilities so that only relevant summarized subsets are presented to the scientist for evaluation. Another issue is that massive scientific datasets are essentially immovable. Although raw network capacity is increasing rapidly, data quantities are growing even faster. In consequence, much of the process of data extraction, subsetting, and summarization must be done remotely, before the data is transmitted. This requirement changes our data analysis software methodology dramatically and raises a number of subtle issues. Reliable and robust data networking is essential, along with appropriate quality-of-service guarantees, since much of the remote computing will be interactive. Moreover, we need middleware to manage intermediate storage systems in the network and mechanisms for scheduling and running analysis jobs at locations remote from the customer.

The more successful this CS program element is, the larger and more challenging the issue of long-term software support and evolution becomes. Today the high-end software infrastructure consists of many millions of lines of code and represents the investment of thousands of Ph.D. years of research and development. The original development teams currently maintain almost all of this code. New mechanisms and approaches are needed to provide end users with necessary assurances of long-term support and to free research developers to undertake new efforts. A fundamental aspect of the problem is that the high-end segment of the marketplace is too small to be attractive to third-party software firms. Open source software is essential to enable long-term support in most instances, but it does not address the underlying economic issues. Clearly, some type of long-term investment by the government in high-end software maintenance will be required to address this problem.

Planning Process

Input to the CS research planning process comes from a variety of sources:

- Feedback from application scientists who help identify needs and gaps in existing research activities. This feedback occurs during focused workshops, laboratory visits, SciDAC , and a variety of other mechanisms.
- Close coordination and interaction with other research agencies to develop a broad understanding of research priorities and gaps from a national perspective.
- Technical workshops, PI meetings, and formal or informal interactions with the high-end research community to help determine opportunities and potential new directions.
- Discussions with members of the Advisory Committee, with program managers in ASCR and with program managers throughout the Office of Science.

Examples of recent planning efforts include three workshops in the past two years on scalable and reliable operating systems for high-performance systems (the FAST-OS workshop series) and a workshop on high-end scientific data management frameworks. In addition, a workshop on end-to-end petascale data management will be held early next year. Workshops such as these provide an effective means for community involvement in the identification of end-user needs and critical research issues that must be addressed.

Gap Analysis

The software infrastructure must be provided to enable new architectures to efficiently run legacy application code and promote the development of new application codes that take maximum advantage of the high-performance characteristics inherent in the architectural designs.

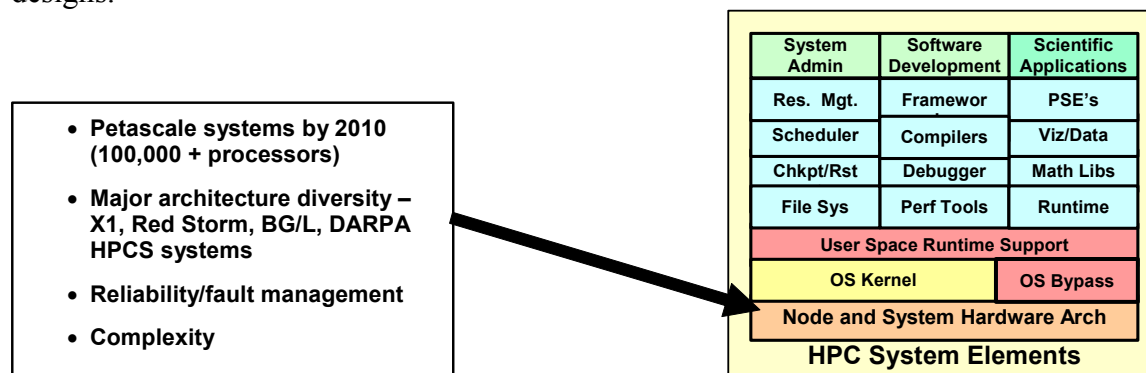


Figure 10: Systems Software Challenges

As sketched in Figure 10, the challenge is to “swap out” the legacy hardware layer at the bottom and to “swap in” a new hardware layer and have the entire software infrastructure between the low-level hardware and the high-level applications, system management, and software development environments continue to function. This is a daunting challenge and will require a maximum research effort for the foreseeable future.

Another emerging gap is the problem of coping with petascale datasets. Simulations in astrophysics, climate modeling, computational biology, fusion, high energy physics, combustion, materials, and other scientific applications are now routinely producing datasets containing tens to hundreds of terabytes and petabyte datasets will be here soon. While processing power typically doubles every eighteen months following Moore's law, the quantity of data stored online quadruples in the same period.

Integrated Network Environment

Contribution to Overall ASCR Strategic Goals

A scalable, secure, integrated network environment for ultrascale distributed science is being developed to make it possible to combine resources and expertise to address complex questions that no single institution could manage alone. Unique resources, large teams, and large-scale computing and data are characteristics of many DOE science programs. This program element is creating the means for research teams to integrate unique and expensive DOE research facilities and resources for remote collaboration, experimentation, simulation, and analysis. This environment is closely coupled with the production network provided by ESnet.

Planning Horizon

The rapid changes in computing and communication technologies mandate a flexible planning horizon for both the short term and the long term. Such planning is also called for to keep the research pipeline feeding the development of useful tools.

Short-term planning addresses the integration, prototyping, testing, and accelerated deployment of advanced computing, communications, and middleware technologies that are outcomes of previous research and development efforts. Driving the long-term planning are the projected requirements of future scientific applications—fundamental research issues for advanced collaborative and network capabilities that must be resolved to satisfy requirements of the scientific applications.

Current Areas of Investment

The Integrated Network Environment program element currently invests in three areas:

Middleware – research and development of technology to enable universal, ubiquitous, easy access to remote resources and to ensure that distributed teams work together easily. Enabling high performance for scientific applications is a strong focus, and it includes services that allow applications to adapt to changing network conditions.

Network Research – leading-edge networking technologies, such as transport protocols and services for ultra-high-speed data transfers; techniques and tools for ultra-high-speed network measurement and analysis; and scalable cybersecurity technologies that protect resources in the open science environment.

Collaboratory Pilot Projects – early implementations of virtual laboratories that give scientists the technology to collectively observe and attack problems using combinations of ideas,

methodologies, and instrumentation that do not exist at any single location. These virtual laboratories test and validate the enabling technologies that unite distributed expertise, instruments, and computers for discipline-specific applications.

Transfer of Knowledge to Application Scientists

The DOE national laboratories, the unique instruments and facilities at those laboratories, and the university community contributing to the DOE missions create a complex, distributed system that is conducting scientific research on a wide range of complex problems that depend, increasingly, on high-performance network infrastructure. By developing a scalable, secure, integrated environment to support network-intensive science applications, ASCR is in a unique position to create a model for complex interdisciplinary science research through the testing and validation of advanced concepts and technologies developed in this research program as well as in other Federal agencies, universities, and industry. The main vehicle is collaboratory pilots, which are partnerships with other program offices.

Fundamental and applied research efforts in collaborative and network technology across the government are coordinated by participation in the working groups and teams under the Interagency Working Group on Information Technology Research and Development (ITRD), an OSTP function, which has representatives from OMB, OSTP, NSF, DARPA, NASA, NSA, DOE/MICS and DOE/ASCI. Most of this coordination occurs as part of the Large Scale Networking Coordinating Group of the ITRD. The two teams under LSN, Middleware and Grid Infrastructure Coordination (MAGIC) and the Network Research Team (NRT), have conducted planning workshops as well as workshops focused on specific critical issues.

Extensive working relationships exist within the broader scientific community addressing Grid technologies, collaboration technologies, and network research. Researchers involved in these areas are active participants in a number of standards bodies such as IETF, the Global Grid Forum, and Oasis.

Strengths

The built-in goal of supporting collaborations and cooperative research increases the level of interaction, integration, and cooperation in these research projects and programs. This is a clear advantage in the transfer of knowledge to application scientists. Beginning with the DOE2000 program in 1997 and continuing with the SciDAC program in 2001, a culture of joint program planning and execution, both among DOE laboratories and among organizations within DOE, has been nurtured as an integral part of this element. This provides a core group of researchers who are leading the cultural change within DOE and its laboratories, a change that will make the integration of DOE R&D efforts more successful in the future.

The collaboratory pilot projects are pivotal to this process and to understanding the requirements for a range of scientific applications that frequently exceed present state of the art in computing and communications. In addition to providing rigorous tests of the emerging technologies, the collaboratory pilots help define promising areas for future research and are the initial transfer of the technology to scientists in those disciplines.

The networking research projects also make their software products and information available to other DOE efforts and the larger scientific community through project Web sites and open source software distributions. Moreover, infrastructure designs are often incorporated into community standards efforts.

Other outreach activities include newsletters, reports, involvement in standards development, and demonstrations at topical conferences aimed at educating potential users.

Weaknesses

Three major weaknesses have been identified:

Cultural Inertia

- Community building across disciplines and organizations is still hard and requires considerable attention to get right, especially when introducing in new elements and players.
- No mechanism or migration path exists for institutionalizing and incorporating new services, developed through research, into the production network infrastructure so that they are made available for the broad scientific community.
- Success metrics for production networking conflict with success metrics for development of network-based advanced services.

Technology Barriers

- Technology still requires considerable investment of time and effort, an investment that some adopters do not feel is justified for the benefit gained.
- Lack of investment in fault tolerance, troubleshooting, error detection, and so forth is a barrier for broad use of the technology by some applications.
- Advanced applications have limited ability to test at scale in the production network, in part because of the inherent conflict of objectives between research and production activities.
- No infrastructure is available to support experimental testbed activities for network research areas such as ultra high-throughput transport protocols.
- Mature, supported (long-term) code is necessary for broad community adoption.

Organizational Barriers

- The organization is ineffective in integrating program responsibilities and accountabilities when moving from research to production environments.
- Middleware technologies demonstrated and services generated do not have an institutional migration path into the broad community infrastructure.
- Research and evaluation networking testbeds are supporting application projects that stress the network capabilities. These can provide early production experience needed for early adoption of advanced network technologies, but there is no migration path or resources to move these capabilities into production.

Opportunities

Ten years ago the National Research Council report *National Collaboratories – Applying Information Technology for Scientific Research* forecast the following:

The fusion of computers and electronic communications has the potential to dramatically enhance the output and productivity of US researchers. A major step toward realizing that potential can come from combining the interests of the scientific community at large with those of the computer science and engineering community to create integrated, tool-oriented computing and communication systems to support scientific collaboration. Such systems can be called *collaboratories*. (National Academy Press, Washington, D.C., 1993)

It took about five years for the concept to be embraced in a measurable number of areas. Today, the technology has advanced to the point where scientists (with high-energy physicists leading the way) are beginning to rely on that technology to do their science.

With sufficient investments in integrated advanced infrastructure over the next five years, it will be possible to effect paradigm shifts, increasing the scale and productivity of science well beyond what we have today. The revolutionary shifts in the variety and effectiveness of how science is done can arise only from a well-integrated, widely deployed, and highly capable distributed computing and data infrastructure, and not just any one element of it.

In five years, the infrastructure and the services it provides must be capable of handling the 10 petabytes of data per year that will be coming from the LHC experiment in CERN so that U.S. scientists can participate fully in the search for the Higgs boson. The same infrastructure should be capable of making the climate experimental and simulation data in the United States, which will be increasing at a rate of 3 PB per year, as easily accessible as data on a desktop hard-drive, so that all climate researchers can assess the potential consequences of climatic changes.

Opportunities need to be exploited for aggressive application projects structured as true partnerships between application scientists, computer scientists, and network engineers. These are drivers for the underlying technology and provide a means of evangelizing new application communities such as bioinformatics and nanotechnology. The collaboratory pilots will continue to be challenged by rapid technology growth in computing and communications technology.

Research needs to be conducted in the context of applications for the following:

- Effective network caching and computing, to stage large datasets and rapidly access computing
- Ultra-high-throughput transport protocols to improve the end-to-end throughput of distributed science applications
- Scalable technologies for cybersecurity systems
- Monitoring and problem diagnosis for end-to-end and top-to-bottom information for science applications and services
- Management, in a community setting, of the authoring, publication, curation, and evolution

of scientific data, programs, computations, and other products

- Real-time collaborative control and data streams with stable end-to-end control channels
- Dynamic provisioning and demand-assignment bandwidth
- Facilitating and automating of the scientific workflow in a very dynamic environment of distributed scientific information
- Common middleware tools that enable domain developers to efficiently construct and extend complex scientific computation and analysis environments

Threats

A major threat to the routine realization of collaborative applications is the heterogeneity of devices, mechanisms, and policies for wide-area, multi-institutional computing and communications environments. The research challenge is to develop and broadly deploy a set of uniform services that hide the details of the local environments and provide enhanced capabilities to applications scientists. In order to be useful, these services must become part of the persistent, base-networking infrastructure—which includes the production network, ESnet. We note, however, that this does not necessarily solve the end-to-end performance problem when data transmissions move across parts of the commercial Internet as well as the DOE infrastructure.

Other threats include the following:

- Long-term planning for facilities and new programs seldom includes wide-area networking—it is not culturally engrained and is not as straightforward as planning for computer cycles. Network issues are generally postponed until facilities and new programs are near deployment.
- Onerous policies are dictated without a clear understanding of the open science environment. For example, security approaches have been promulgated using business practices as a model rather than adopting a trust model more appropriate for the open science environment where it is more important to protect resources than to protect information.
- Networking requirements call for highly specialized technologies. Few commercial components are available for designing very high-speed networks. Experimental networks have historically led in this area, driving industry to accommodate, for example, *diffserv* in router capabilities.
- As DOE pushes the limits of networks capabilities and middleware, the burden to develop, deploy, and maintain very highly specialized components will increase and will become a potential major strategic issue. DOE has traditionally leveraged commercially available components to develop networks with enhanced capabilities. The reliance on specialized technologies will require commitment and leadership in technology areas useful to DOE but with very little immediate commercial payoff.
- Graduates of traditional computer science curriculums are not trained to handle the unique networking, collaborative technology, and middleware challenges facing DOE.

Planning Process

As with other element of ASCR, the planning process must address two factors. First is the need for input from the scientific communities for which remote collaboration, supercomputing, data-intensive activities, and remote resource access is important. Second is the need to take into account efforts supported by other Federal agencies. The inputs from many sources must be integrated to develop strategic plans and roadmaps. These sources include workshop reports on critical areas, findings by advisory committees, and reports of program and project reviews.

For example, a planning workshop held last year focused on understanding the requirements of high-impact science applications over the next five to ten years on middleware, high-performance network research, and the existing production network. A subsequent workshop addressing the fundamental challenges of developing ultra high-speed transport protocols and related technologies was held to identify research issues more clearly and to develop a research agenda and understand the priorities for testing in an experimental network testbed. A related workshop sponsored by an interagency group covered middleware and Grid infrastructure research. Most recently, using those workshops as basis to build on, a workshop was held to develop a comprehensive technology roadmap that will take the Office of Science networking infrastructure (that includes experimental network testbeds, research and evaluation networks, and a production network) to a state that will satisfy the networking requirements of high-impact science applications over the next five years.

Gap Analysis

Rapid advances in optical network technologies in the past decade have resulted in abundant deployed optical network bandwidth that can be exploited by the Office of Science to create new cost-effective models of high-performance networks to satisfy current and future network requirements of science applications. There is a limited time window to take advantage of these changes in the telecommunications industry to pursue a significantly different approach for network provisioning. Options include procurement and lighting of dark fiber or obtaining multiple lambdas from a commercial vendor.

Networking is advancing much faster than computing and has the potential to transform research as well as our daily lives. Through collaboratory pilots and experimental network testbeds, discipline scientists can work in collaboration with computer scientists to think of radically new ways to enable their science. The development of advanced experimental networks is an opportunity that can be exploited to maintain leadership in networking with extreme capabilities. These experimental networking testbeds, along with the advanced deployment and evaluation of new networking technologies, and the exploration of advanced networking concepts will be aimed at accelerating the adoption of emerging networking technologies and transferring networking R&D results into production networks that support science applications (including ESnet). Teaming with other agency experimental or research testbeds provides leverage for all stakeholders and ensures that the networks that evolve from those testbeds will interoperate. The

issues of end-to-end performance can also be addressed in this way, especially those related to policies.

Extraordinary advances in supercomputing technologies are making possible petaflops computers in the next few years. Providing efficient access to these supercomputers will require terabits/sec networks with corresponding high-performance middleware that addresses issues such as network caching and computing, fault tolerance and error detection/correction, and real-time collaborative control and data streams.

Facilities and Testbeds

High-Performance Production Capability: NERSC

Contribution to Overall ASCR Strategic Goals

NERSC is the Office of Science's flagship high-performance production capability computing center. NERSC provides a balanced introduction of the best new technologies for complete computational and storage systems, coupled with the advanced development activities necessary to wisely incorporate these new technologies. NERSC is currently the source of one of the most powerful terascale unclassified computational platforms in the United States as well as a petascale data storage facility. Having been recently upgraded, NERSC-3E is now a 10 teraflop facility with over 6,000 compute processors. NERSC's staff provide the entire range of support activities, from high-quality operations and client services to direct collaborative scientific support, which enables a broad range of scientists to effectively use the NERSC systems in their research. NERSC provides support for SciDAC and other computational science teams, with the goal of bridging the software gap between currently achievable and peak performance on the new terascale platforms through new and innovative software and algorithm development.

Planning Horizon

The current planning horizon for High Performance Production Capability computing is five years. This is described in the NERSC Strategic Proposal FY2002–2006 and the accompanying document, the NERSC Implementation Plan FY2002–2006, which were reviewed by ASCR. This planning horizon is determined by a balance between two objectives:

- Addressing the computing capability and capacity requirements of the DOE Office of Science programs
- Increasing computational capability using emerging and constantly improving scientific computing technology

Current Areas of Investment

NERSC's computational resources support all of the program offices and areas in the Office of Science, especially

- Accelerator Physics
- Astrophysics
- Chemistry
- Climate and Environmental Science
- Computational Science and Mathematics
- Earth and Engineering Science
- Fusion Energy Science
- Lattice Gauge Theory
- Life Sciences
- Materials Science
- Nuclear Physics

In order to support these applications NERSC provides the highest capability production computing resource available at any time to the SC research community. This investment is balanced by investments in supporting infrastructure as well as expert staff that provides direct support to the researchers in all of the SC program offices. Results from the NERSC staff effort in utilizing new hardware and software are made available to all SC program offices. The allocation of computer resources across all SC program offices is coordinated by the Office of Science Supercomputing Allocation Committee (SAC), which comprises one representative from each of the SC program offices.

A new vision of NERSC as a *virtual* center is also evolving. In conjunction with the Research and Evaluation Program (R&E), NERSC will manage those systems that migrate from the R&E Program into the high-performance production capability class. These systems will be integrated into the Energy Research Computing Allocations Process (ERCAP) at NERSC. This will provide the program offices with a single vision of the resources across the entire computational portfolio. It also will provide the ability to optimally match the applications with the appropriate computational platform. NERSC will participate in the DOE Science Grid. SC's Innovative and Novel Computational Impact on Theory and Experiment (INCITE) Project will continue to make 10% of NERSC's computational resources available to the worldwide scientific community.

Transfer of Knowledge to Application Scientists

NERSC scientific staff members have extensive experience working in direct research collaboration with NERSC clients on computational science projects, with the goal of promoting and enhancing the use of the facility. With expertise in applied mathematics, computer science, and various scientific disciplines, the scientific staff can effectively collaborate with NERSC clients and the broader scientific community. They are involved in timely development of state-of-the-art methodologies and strategies for computational sciences that are suitable for massively

parallel computation, thus opening ways to do new science that is otherwise impossible. They also perform research in the design and implementation of highly efficient computational kernel algorithms for current and future NERSC architectures and applications. Consulting staff provide software support for a complex set of applications, libraries, tools, and environments that exist on some or all the NERSC systems, such as Gaussian, NAG, MPI, Totalview, and performance analysis tools. A variety of software packages are supported, including visualization and client interface software, some of them commercial and some from the open scientific community. Consulting support focuses on directly helping the DOE scientific community become more productive in their computational and data management work. It provides direct client assistance, managing and resolving client problem reports. It is important that the client community be able to interact and ask for assistance in the way most comfortable and effective for them—not just in the most efficient way for NERSC. Thus, NERSC supports telephone, email, and Web interactions with timely acknowledgment and response resolution.

Under the SciDAC program, NERSC staff is actively involved in selected Scientific Challenge Teams and their corresponding Integrated Software Infrastructure Centers. NERSC interacts closely with those ISICs that deal specifically with numerical tools and performance issues. Scientific staff provide high-level support in porting, evaluating, and deploying tools developed by ISICs on the terascale computing platforms at NERSC. The staff also help in understanding and resolving issues, with the goal of enhancing the performance of such tools.

NERSC provides advanced training and client instruction in the use of the latest technology. NERSC staff develop skills in new areas and share them with clients by creating, updating, and presenting all the relevant external and internal training information related to using NERSC systems. These activities include video and teleconferences, multiple days of intensive classes, lectures, seminars, and symposia presented in collaboration with other groups. All of these training materials are also available on the NERSC Web server. Likewise, NERSC provides value-added on-line documentation that complements, simplifies, and clarifies vendor documentation and other information.

NERSC maintains and supports a diverse collection of visualization software on NERSC production systems for remote researchers. The software ranges from simple plotting packages through high-performance 3D applications. Some software is commercial, some is freeware, and some is custom developed. The maintenance activities are focused on keeping this collection of software up to date and operational in the production environment. The support activities include consulting and providing accurate and up-to-date online documentation. Some consulting activities are short, consisting of only simple questions from application scientists and succinct answers. Others are more substantial, involving a visualization consultant generating complex visualizations. In some cases, these more involved consulting engagements evolve into long-term collaborations that seek to find solutions to problems that are not easily solved with existing software or approaches. NERSC performs outreach activities that are intended to increase awareness in the user community of the NERSC visualization services and capabilities as well as the potential for visualization to improve scientific insight and productivity. NERSC engages in investigations that are intended to produce solutions to today's hard visualization problems, as well as to evaluate approaches and solutions to tomorrow's problems. These activities often involve collaborations with staff in other NERSC groups or other institutions.

NERSC staff also balance high utilization, large-capability computing and responsive systems. For most clients, frequent turnaround (every night) of jobs is important. Some disciplines, such as climate, have to run jobs in a certain order because the simulation result of one period is the input for the next year. NERSC works with the user community to develop guidelines that balance the need to utilize systems well with the requirement to provide good turnaround. Priority schemes and specialized services are used to assure a fair access to the system resources. Using the guidelines, NERSC balances the needs and desires of multiple groups, and it implements methods to make the systems most effective for the scientific community. NERSC is able to respond to special requests from the NERSC community for processing and services. The NERSC computational systems are known for regularly handling full-configuration jobs with frequent turnaround. NERSC also is able to provide one-of-a-kind processing, be it high priority, very long runs, very large memory applications, or massive data or other requirements, because the system managers can configure the systems to respond to diverse needs for resources.

Strengths

NERSC provides extremely high-quality support of large-scale scientific computing by Ph.D.-level scientists, mathematicians, and computer scientists. It has one of the largest unclassified computing capabilities in the United States. It also has the largest data storage facility of any computer system in the Office of Science. NERSC is a world leader in its ability to evaluate the performance of and to procure and deploy the highest-end production computing systems. NERSC has also established partnerships with IBM in the Blue Planet project looking at new computational architectures and in the Software Storage Consortium, which is developing modern High Performance Storage Systems.

NERSC has active collaborations with the other programs of the Mathematical, Informational and Computational Sciences (MICS) Division. In the area of the Unified Science Environment, NERSC leverages the work in Grids and collaborative technologies performed elsewhere in the MICS program and will begin to place large-scale production computing facilities in the DOE Science Grid. In other areas of technology development, NERSC maintains close connections and strategic collaborations with programs funded by MICS in the other DOE laboratories, as well as universities.

Weaknesses

NERSC's five-year planning horizon does not appear to be meeting the growing requirements for the computer resources needed by the base research program, the SciDAC research projects and the INCITE project. Unparalleled growth in the base program and the introduction of SciDAC and INCITE after the initial projections were made in the five-year plan have created a tremendous computational resource shortfall. A similar situation has arisen in data storage. Data-intensive science, especially in the areas of high energy and nuclear physics and climate simulation, is increasing the need for archival storage at a much higher rate than predicted in the five-year plan. Finally, there has been an insufficient investment in developing and deploying architectures specifically designed for scientific computing in the production environment. As a

result, today's high-performance computing architectures are composed of commercial off-the-shelf architectures more suited for Web serving and transaction processing.

Opportunities

Three significant opportunities exist for ASCR's high-performance production capability.

The first is to implement the Grid technology at NERSC that is needed for an integrated science environment, combining experiment, simulation, and theory by facilitating access to computing and data resources, as well as to large DOE experimental instruments.

The second is to allow NERSC to become a *virtual* center. By adding the hardware that migrates from the R&E Projects, NERSC can expand its computational portfolio. In order to migrate from R&E to high performance capability computing, the hardware must have demonstrated efficient performance on some set of DOE applications. Since NERSC provides computational resources to the full spectrum of DOE applications, those that run the most efficiently on the new hardware could be moved there from the flagship center. This would provide more computational resources at the flagship center for other applications. Clearly, a number of cost, service, and integration issues must be faced in this transition; however, it promises significant improvements in the overall capability.

The third opportunity is to improve the coupling between the evaluation of Research and Evaluation computers at the ACRTs and decisions on future hardware installations at NERSC. Using the results of evaluations to decide on hardware installations has the potential to improve the decisions as well as improve integration between R&E and production computing.

Threats

Since 35% of NERSC users are not in DOE laboratories (45% if the INCITE awards are to application scientists outside DOE laboratories), this user base could be lost if NERSC cannot keep up with the demand for computational resources. These application scientists, who are critical to DOE programs, could develop alternative computational resources in alternative research areas not in DOE's mission. Keeping up with the demand for computational resources also requires leading-edge technology. The inability to keep leading-edge computational technology will lead to staff attrition as they will seek this technology at other facilities. NERSC users are spread across the globe. The increased restrictions being imposed because of cybersecurity, especially access of foreign nationals to computational resources, threatens the existence of NERSC and other DOE computational facilities, because these restrictions could drive the large number of scientists at both DOE laboratories and universities are not U.S. citizens to compute at other computational facilities which do not have the same limitations..

Planning Process

Planning at NERSC is guided by both internal and external influences. The most critical goal of NERSC is to meet the needs of its users. To achieve this goal, the NERSC Users Group produces the *Greenbook* every three years. This report includes discussions of the ways in which

NERSC should evolve in order to meet the needs of DOE's scientists to continue to do forefront science. Internally, NERSC's Advanced Systems Group works closely with the Future Technologies Group in the Computational Research Division to keep up to date on the latest computational technologies. These groups have close contact and collaborations with computer hardware vendors, other DOE laboratories, and other computational facilities in order to be aware of the latest technological advances and the performance characteristics of existing technologies. Finally, in conjunction with the programs planning horizon there is a periodic *clean sheet* planning process. This was most recently done in May of 2001 and resulted in the production of two documents, the *NERSC Strategic Proposal FY2002–2006* and the *NERSC Implementation Plan FY2002–2006*.

Advanced Computing Research Testbeds

Contribution to Overall ASCR Strategic Goals

The Advanced Computing Research Testbeds (ACRT) program element supports the acquisition, testing, and evaluation of advanced computer hardware testbeds to assess the prospects for meeting future computational needs of the Office of Science, such as SciDAC and special-purpose applications. The ACRTs provide two types of computer testbeds for evaluation: early systems and experimental systems. Each testbed involves significant research and architecture design activities. These research and evaluation (R&E) prototypes have been identified as a critical element in the HECRTF plan because they enable early partnership with vendors to tailor architectures to scientific requirements. The results from these partnerships also play a key role in the choice of both high-performance production systems and potential leadership-class systems government-wide. Currently, the only ACRT is at the Oak Ridge National Laboratory - Center for Computational Sciences (CCS).

In order to succeed, the ACRTs must provide an integrated computing center with the most up-to-date storage, networking, visualization, and high-end computing resources, along with a staff of application experts to assist the application developers, who play a key role in the evaluation. For example, the CCS has application research institutes in climate and carbon research, nanophase materials science, and genomes to life. The CCS is located at Oak Ridge National Laboratory in a new 40,000 ft² facility. This modern infrastructure is an important asset of ASCR and the Office of Science and is critical to enable the advanced evaluations under this activity.

Planning Horizon

The planning cycle for this program element is determined by three factors:

- Availability of novel computer architectures in prototype, about 3 years
- Timescale for impact on future generations of computer architectures, about 6 years
- Timescale for evolution of scientific applications, 5–10 years.

Current Areas of Investment

Currently the largest ACRT investment is in the evaluation of the Cray X1 at ORNL. ACRT technology evaluation activities include the following:

- Early evaluation of new computer architectures
 - Distributed-memory clusters using the IBM Federation interconnect
 - Large, shared-memory symmetric multiprocessors using COTS processors with the SGI Altix system
 - Distributed storage, networking, and data analysis testbed for SciDAC ISICs and computational biology
- Strong interactions with both the computer industry and industrial users of high-end computing
 - Computer and applications code performance evaluation;
 - Climate modeling with DOE/BER and NSF
 - Fusion simulations with DOE/FES
 - Materials science with DOE/BES
 - Computational biology and genomes to life with DOE/BER and DOE/ASCR
 - Computational chemistry with DOE/BES
 - Astrophysics with DOE/HEP
- Assessment of tools and software environments
- Assessment of new database and networking technologies

Transfer of Knowledge to Application Scientists

As an ACRT, the CCS provides critical information to high-performance production capability facilities as well as applications scientists through the dissemination of test performance data. In addition, the CCS provides critical information to the vendors on the suitability of their prototypes for scientific computing and works with the vendors to correct deficiencies and, more important, plan for the performance needs of scientific applications in future generation computer systems. This information is transferred through scientific papers, conferences, workshops, and training sessions with both vendors and the user community. The CCS maintains an active Web site (<http://www.csm.ornl.gov/evaluation>) with the results of early evaluation benchmarks.

Application scientists play a critical role in the evaluation and are not interested in “demonstrations” where they spend time getting their codes to work for one or two “biggest-ever” runs. Because of the significant investment in these computers, the successful R&E

computers must be run as early production resources for selected application scientist communities, in order to make the best use of the dollars invested.

The CCS serves a few major projects in the DOE and maintains close contact with the user community through the Scientific Application Support Group. Each of the major projects has a CCS staff member who serves as a member of the project team and as liaison to the CCS. These people have extensive experience with parallel and vector programming and are experts in optimizing the application codes with the project's application scientist. The CCS further draws on the resources of the mathematics, statistics, data management, and computer science groups in the division to assist in specialized areas that are of general use to all applications. ORNL has the principal investigator or co-investigators of three of the SciDAC Integrated Software Infrastructure Centers as well as PIs in every scientific discipline in the SciDAC program. This depth of scientific and computational science expertise ensures that CCS users have the very best support available anywhere in the world.

The CCS regularly conducts workshops and small-group training sessions to acquaint users with the best ways to use new computer systems. For example, over the past year the CCS has held several workshops to provide training and tuning information about the new Cray X1 computer system. Since this machine has a radically different architecture and programming paradigm from the prevalent computer systems in use today, it was necessary to provide more extensive training and guidance to users wishing to use the system. General workshops were held prior to the arrival of the system, and a number of discipline-specific workshops have been held since delivery to help scientists understand the best way to use the computer. Moreover, the CCS Web site provides a wealth of information and documentation for users, from the novice to the experienced computational scientist who needs a reference manual on a specific topic.

Strengths

The ACRT program builds on a strong history of successful efforts in this area and has a number of sites with the specialized expertise to field novel architectures and attract the scientific applications needed to test them.

The CCS also has an extremely strong operations staff of system administrators and programmers who are able to deploy and administer the many "serial number one" systems procured by the center. In addition, these systems require the proper balance of secondary and tertiary storage as well as network connectivity, and the technical staff are equally adept at supporting these peripheral systems. Ultimately these high-end computer systems must be delivered to application scientists. The center provides excellent scientific application support in order to deliver maximum science to the DOE community.

Weaknesses

The current portfolio of systems under evaluation is too narrow. In addition, information dissemination to production centers as well as government-wide portfolio management in this area is not sufficient.

Opportunities

After a period of relative stagnation, it appears that a much more varied set of architectures will be available in the next 3–5 years. Early evaluation of these prototype systems will play a critical role in any effort to deliver petaflop-level computing for science. The HECRTF offers the opportunity to improve the portfolio management and information dissemination weaknesses identified above.

Threats

A major threat to the ACRT program is the fact that the computer systems must be procured quickly if the evaluations are to be useful both to the government and to the vendors, because the specifications for the next generation of computer are often set 24–36 months before it is delivered. These requirements are inconsistent with the timings built into OMB-mandated IT capital planning processes. In addition, the current focus on net present value as a financial measure distorts the process and is widely understood to systematically undervalue R&D.

Three threats other threats must also be addressed by the ACRT program:

- The research systems represent significant financial commitment.
- Many of the issues that confront the applications appear only on systems with thousands of processors. It is simply not possible to predict the performance of a scientific application on 5,000 processors based on its performance on 128.
- Application scientists, who play a critical role in enabling realistic evaluations of new architectures, are unwilling to invest the effort needed to enable their applications to run on the new architectures *unless there is a clear possibility for significant scientific advances*.

Planning Process

ACRT evaluations are based on reviewed proposals from organizations that have a peer-reviewed history of successful evaluations of systems at scale. It is important to install novel systems in locations that can leverage personnel and infrastructure resources; however, it is not necessarily the case that ACRT infrastructure needs to replicate production facilities.

Program Element Gap Analysis:

Three significant gaps exist in the ACRT program element. First, the number of architectures under study needs to be expanded to ensure better interactions with vendors and the architecture research community on the scientific computers of the future. Second, the coupling and

information transfer between the R&E prototype evaluations funded under this program element and other parts of the ASCR program and other parts of the Federal Government needs strengthening. Third, the infrastructure and teams at a number of locations are a critical capital resource that must be maintained so that these activities can be effectively carried out.

ESnet

Contribution to Overall ASCR Strategic Goals

Since 1985, the Energy Sciences Network (ESnet) has provided the wide area network that has been a critical infrastructure service for all of the scientific programs within the Office of Science. It enables researchers at national laboratories, universities, and other institutions to communicate with each other using the collaborative capabilities needed to address some of the world's most important scientific challenges. It provides worldwide access to SC facilities, including light sources, neutron sources, particle accelerators, fusion reactors, spectrometers, high-end computing facilities, and massive data resources. The ESnet and its associated services thus are essential to the continued success of DOE science.

Planning Horizon

The short-term planning for tactical deployments of infrastructure is supported by direct community input for requirements through the Energy Sciences Steering Committee (ESSC).

Long-term planning is driven by the projected requirements of scientific applications. For example, from 2003 to 2008 there will be a 500-fold to 1,000-fold increase in the amount of data going to media at many Office of Science facilities. This roughly matches the historical doubling seen every year in the amount of traffic moving across ESnet since its inception.

Current Areas of Investment

Three major areas require research and development before additional production-quality services can be added to the present infrastructure and made available to the user community. ESnet provides support in each of these three areas.

Connectivity. Connectivity research and development seeks to understand the dynamics of a very high-speed nationwide network.

- Distributed network performance measurement and analysis – techniques to measure the dynamics of the network from various sites and to perform analysis to improve end-to-end performance
- High-performance transport protocols – protocols that scale to very high-speed and ultrascale networks or can make effective use of OC192 (10 Gbps) networks and beyond
- Multicast and secure group communication – methods for secure data distribution to a large number of scientists simultaneously
- High-speed, ubiquitous, and reliable backbone infrastructure – technology that enables scientists to communicate effectively irrespective of time and location in the world.
- Quality-of-service (QoS) services – services to improve end-to-end performance on existing high-speed networks

- New protocol implementation (IPv6) – for better management of the limited IP address space and for better security

Collaborative technologies: A number of DOE large-scale science projects critically depend on collaborations of multidisciplinary researchers who collectively represent capabilities that are unavailable at any single national laboratory or university. Seamless access to distributed resources that include high-end computers, experimental facilities, and data repositories by the researchers is essential to carrying out DOE missions, and the “network” and the associated collaboratory or Grid tools have become critical components of the modern scientific infrastructure. Research in distributed technologies is required to allow a widely dispersed research community to effectively collaborate on research.

- Access Grid technology – allows multisite simultaneous video and data conferencing using multicast technology
- Ad hoc H.323 (IP-based) conferencing – allows WAN-based video and interactive data collaboration.

Security: ESnet supports the development of the DOE Grids Certificate Services to support scientists and engineers working on DOE-related scientific efforts, including the recent introduction of the Public Key Infrastructure (PKI) services into the production services. This PKI service is designed to support the Grids being deployed around the world, as well as a number of DOE virtual organizations that require the use of certificates that are trusted in the global research community. ESnet actively coordinates with the Global Grid Forum, the European DataGrid, and Cross Grid CA managers to ensure that DOE Grids certificates have the widest possible acceptance.

Fundamental and applied research efforts in collaborative and network technology across the government are coordinated by participation in the working groups and teams under the Interagency Working Group on Information Technology Research and Development (ITRD), an OSTP function, which has representatives from OMB, OSTP, NSF, DARPA, NASA, NSA, DOE/MICS and DOE/ASCI. Most of this coordination occurs as part of the Large Scale Networking Coordinating Group of the ITRD. Under the LSN, the Joint Engineering Task Force (JET) coordinates operational activities for Federal networks and involves multiple agencies as well as vendors.

ESnet participates extensively in community forums, task forces, working groups, committees, and subcommittees in which standards, protocols, and acceptable practices are developed. Some of the more important ones areas follows:

- Energy Sciences Steering Committee (ESSC)
- Energy Sciences Site Coordinating Committee (ESSC), which holds two joint meetings a year with Internet 2 and the LSN/JET
- Internet Engineering Task Force (IETF) partnerships
- Multi agency/Vendor Joint Engineering Task Force (JET) membership
- Workshops, conferences, and publications
- North American Network Operators (NANOG) Group membership

Transfer of Knowledge to Application Scientists

ESnet is a high-performance network infrastructure that supplies the DOE science community with capabilities and services not available through commercial networks. As such, it provides services to all application scientists who require that infrastructure. It participates in the development of the global Internet by being an early adopter of new technologies such as dual-protocol routers, Ipv6, Slow Start, and ATM. The use of such emerging technologies in a production-quality network helps advance the state of technology within the wider Internet community.

Strengths

ESnet provides end-to-end connectivity for the DOE research community that is not available over the commodity Internet. It has strong community involvement, as well as support through the ESSC and ESCC in the operation and management of the infrastructure. ESnet collaborates with the network research and middleware programs, both in DOE and other federal agencies, to expand the advanced services available to the user community. ESnet has developed many long term working relationships with national and international peers and has a long track record and extensive experience in meeting the ever-expanding programmatic requirements. It is able to gain significant leveraging of overall effort and cost savings with central support staff on 24/7 basis and has proved to be very responsive to the special demands of the research community. The strong cooperation between ESnet and its users allows rapid and effective response to evolving mission needs. As such, ESnet provides a single, positively branded, national and international identity for DOE networking.

Weaknesses

The weaknesses that ESnet has to face are as follows:

- Success metrics for production networks conflict with success metrics for R&D networks. Users expect a 99.999% available network, while research requires that one break the network occasionally.
- Users demand production network services (i.e., “need it yesterday”).
- The inherent conflict of research and production objectives makes collaboration between those elements of the overall program difficult, but collaboration is necessary to move new services into production.
- Users have difficulty dealing with any outages, even less than 1%, as evidenced by ESnet’s current performance.
- Less than mature router/switch code from vendors for high-end networking limits how fast new services are introduced into the production network. The fact that vendors are under pressure to deliver services may result in deployment of partially tested code.
- Self-imposed barriers sometimes limit effective working relationships (e.g., funding transfer constraints, “not invented here” mentality).

Opportunities

A number of opportunities exist:

- Core networks have abundant optical capacity to develop cost-effective agile network infrastructure to support high-impact science applications. ESnet is working with Qwest and National Lambda Rail (NLR) to get low-cost pricing on dark fiber.
- ESnet is exploring deployment of ultra high-speed metropolitan networks to interconnect data analysis and management centers associated with petascale computers.
- Growth in Grids will create requirements for new central services. ESnet is working closely with the Grid community and actively participates in Grid meetings and conferences.
- ESnet continues to apply research network technology improvements to accelerate the development of advanced networking services
- ESnet is developing additional distributed security services to protect the network and provide user security without cumbersome user services

Threats

A number of threats will affect the ability of ESnet to meet the DOE research needs in the next few years:

- Science programs do not have a culture accustomed to defining long-range programmatic networking needs in a timely manner. The prevailing mentality is “Adequate networking will always be there.” Yet it takes months and even years to plan and deploy the ultrascale links and develop the protocols and put them into production. The science programs have to become much better at projecting their ultrascale network needs.
- Many user communities feel that the network is free and do not fully understand the resources required to operate a high-speed production network.
- Current cyber threats require tighter constraints and use valuable network resources. As these threats evolve, more resources are being spent on prevention and mitigation. In a zero-sum budget, this is becoming a serious weakness.

Programmatic needs are projected to outstrip network projected growth rate and budget. Increased emphasis and need for ultrascale bandwidth and new services cannot be accommodated indefinitely.

- Security policies may neutralize needed performance advances if mandated from above as “one size fits all” without regard to the specialized needs of the DOE research community.
- There is potential reduced longevity of at-risk telecomm vendors because of current market conditions. With the rash of telecom companies that recently have left the marketplace and those that are in trouble, telecom partners that can deliver the long-range goals required for ultrascale networking are becoming more difficult to find. This will continue to be an

issue until the marketplace sorts itself out.

- Highly specialized network technologies are beyond the current scope of industry. Most of the industry requires many low- to medium-speed connections. For example, very few vendors can provide quality of service and latency guarantees.

Planning Process

To develop strategic plans, ESnet relies on two major activities: input from the scientific community for which the production-quality network is important; and a good working relationship with other worldwide networks and service providers for peering and connectivity. Input from the scientific community comes from active participation in numerous committees to gain an understanding of the programmatic requirements for DOE science. Other sources include workshops on critical areas for science, findings by advisory committees, and reports of program and project reviews. In addition, ESnet has a steering committee (ESSC) and a site coordinating committee (ESCC) and participates in numerous forums (IETF, NANOG, I2) and working groups (LSN, JET). ESnet also held a road-mapping workshop earlier this year, which defined the strategy for how ESnet would fulfill DOE requirements.

Gap Analysis

Historically, science communities have expected network capabilities to be readily available when needed, and generally they have been. In recent years, however, the gap between ready out-of-the-box end-to-end performance and theoretical performance has been widening, approaching three orders of magnitude (see Figure 11). Coupled with unprecedented requirements of large-science applications, this gap will be a severe bottleneck. For instance, the leading operational link speeds are currently at 10 gigabit/sec level (OC192) but are available only at the backbone. At the applications end-hosts, the transport throughputs typically reach only a few tens of megabits, and routinely reach only a small number of hundreds of megabits with considerable effort and constant attention. Furthermore, bandwidth is not the only measure of performance needed. There are no technologies available in current operational wide-area networks that will provide either the guaranteed stability needed for real-time steering and control over long-haul Internet connections or the agility needed for instantly redirecting massive visualization flows.

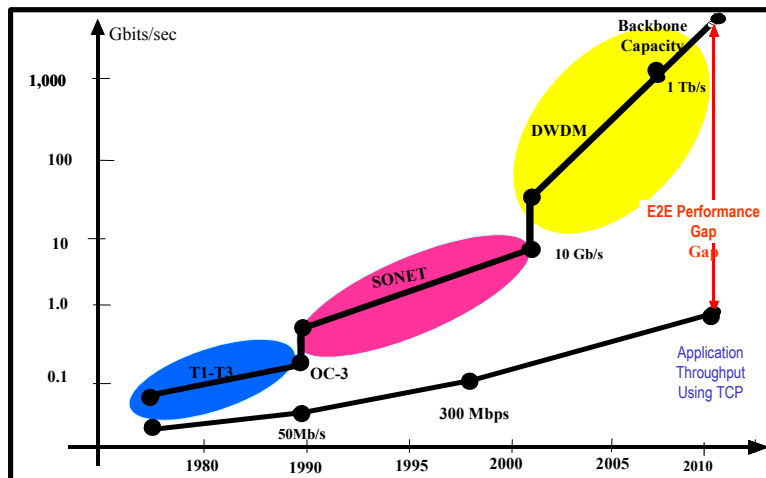


Figure 11: Gap between actual and theoretical performance

Appendixes

Appendix A: External Environment

A number of external factors influence the ability of ASCR to achieve its missions.

- The evolution of the commercial market for high-performance computing and networking hardware and software for science.
- Strategic and programmatic decisions made by other (non-DOE) Federal agencies and by international entities.
- The availability of a world-class research community to work on ASCR problems.
- Evolution of public and government attitudes about risk that lead to restrictive security and program evaluation methodologies.
- The fundamentally multidisciplinary, collaborative, distributed, and often international nature of the “big science” questions that DOE is charged with attacking.
- The evolution of government and DOE priorities that affect the level of resources available to accomplish ASCR missions.

The commercial market for high-performance computing and networking hardware and software has a profound effect on the ability of ASCR to meet its missions. We do not fabricate or design any of the underlying technology.

Technology trends and business forces in the U.S. computer system industry over the past decade caused most domestic vendors to curtail or abandon the development of high-end systems designed to meet the most demanding requirements of scientific research. Instead, large numbers of smaller commercial systems have been combined and integrated into large systems that try to achieve the peak performance levels required for agency missions in computational science. The hardware is complicated, unwieldy, and not balanced for scientific applications. Enabling software has been developed for scientists to take advantage of these new computers. However, this software is extraordinarily complex because of the requirements for implementing parallelism. Even if hardware were better optimized for the requirements of science, the inherent complexity of managing large-scale parallelism would make the software complex. Consequently, DOE, primarily through the MICS subprogram, and other Federal agencies whose missions depend on high-performance computing, must make basic-research investments to adapt high-performance computing and networking hardware into tools for scientific discovery. Current Federal efforts, including DARPA’s High Productivity Computing Systems project and the interagency High End Computing Revitalization Task Force, are designed to ensure the existence of future high-performance systems optimized for scientific applications.

The explosion in data also challenges our ability to make effective use of hardware and software designed for business applications. High-performance computers and large experimental facilities deliver petabytes of data. The required computing systems, capable of efficiently dealing with petabytes, are not expected to emerge as commercial off-the-shelf products in the near term. Storage systems must be able to manage hundreds of petabytes and also allow interactive analysis of 100-terabyte files. In this area the current commercial focus is on capacity, not speed. As computing performance evolved from 1999 to the present, one debilitating constraint remained—the large time required to access data on storage. The dominance of storage access latency exists even with aggressive assumptions about the evolution of storage device data transfer performance. Mount times and positioning for removable volumes will see only small, if any, improvement. With storage and retrieval times of 1–2 hours to satisfy a large data query, interactive access is infeasible through expected technology extrapolation alone. In addition, the fact that the doubling times for the performance of various component technologies is different implies that high-performance computing and data challenges will require significant I/O parallelism to satisfy the storage throughput requirements. More precisely, the data manipulation environment will require highly parallel I/O systems for dynamic distribution of data across parallel disks and tapes, persistent storage for massive data collections, support for creation of derived data products and the associated simulation code, manipulation of remote data products, and catalogs to maintain metadata describing the data and algorithm collections.

ASCR investments in network facilities and the computer science research that underpins the integrated network environment are significantly affected by the current transition in the telecommunications industry, combined with the opportunities for direct fiber ownership and DWDM wavelength leases. In addition, even though ASCR requirements to move a small number of very large data streams are different from the requirements of a commercial network for aggregation of a large number of small flows, ASCR research in these areas is constrained because the results must be deployable in an international telecommunications infrastructure over which we have no direct control.

Also affecting ASCR investments in network facilities and the computer science research that underpins the integrated network environment are the current government-wide approaches to security. Since the September 11 attacks on the United States, the government movement to mitigate risk whatever the cost has accelerated. New laws, such as the Federal Information Security Management Act, and OMB policy make research in advanced network technologies much more difficult and thus impact the distributed collaborative science that these advanced technologies enable. Strategies to provide appropriate protection for cyber assets, without compromising our ability to do the needed research, must be explained and justified to authorities whose goal is to reduce risk, especially risk of embarrassment.

ASCR relies on a world-class research community. It is well known that the United States is not producing enough undergraduates trained in science as a whole, much less in high performance computing and networking. The current security environment also has had a dramatic impact on the ability of U.S. universities to attract the best foreign graduate students.

Appendix B: Planning Process

The planning process for ASCR must take into account two important factors:

- The need for input from all of the scientific communities that ASCR research and facilities support.
- The relationship of ASCR research and facilities to efforts supported by other Federal agencies and other parts of DOE, especially NNSA.

For this reason the ASCR staff must integrate inputs from many sources to develop strategic plans and roadmaps. These sources include Federal staff in other SC program offices; the Advanced Scientific Computing Advisory Committee (ASCAC) and the other Federally chartered advisory committees for SC programs; interagency coordination of information technology R&D under processes established by OSTP; Federal staff at the National Nuclear Security Administration, to ensure close coupling with ASCI; the worldwide scientific research community in scientific disciplines important to DOE; and the nationwide research community in the scientific disciplines in the ASCR portfolio.

Because of the diverse set of inputs needed to support strategic planning, ASCR uses a variety of techniques to communicate with the various external communities. One of these techniques is a formal charge to ASCAC to evaluate high-performance computing facilities. In addition, ASCR participates in interagency planning processes such as the High End Computing Revitalization Task Force (HECRTF). Further, ASCR relies on the results of technical workshops sponsored by ASCAC (sometimes in partnership with other Advisory Committees), by interagency committees (such as HECRTF), by ASCR in partnership with other SC program offices (e.g., Genomes to Life), or by ASCR alone. A list of the major activities used to inform this strategic plan is in Appendix A. While the planning horizon for this document is 10 years, it is anticipated that major revisions will be needed more frequently as a result of changes in the external environment, DOE priorities, and the results of evaluations of the ongoing efforts.

One recent meeting, the Science Case for Large Scale Simulation (www.pnl.gov/scales), made a number of specific recommendations:

- Major new investments in computational science are needed in all of the mission areas of the Office of Science in DOE, as well as those of many other agencies, so that the United States may be the first, or among the first, to capture the new opportunities presented by the continuing advances in computing power. Such investments will extend the important scientific opportunities that have been attained by a fusion of sustained advances in scientific models, mathematical algorithms, computer architecture, and scientific software engineering.
- Multidisciplinary teams, with carefully selected leadership, should be assembled to provide the broad range of expertise needed to address the intellectual challenges associated with translating advances in science, mathematics, and computer science into simulations that can take full advantage of advanced computers. Extensive investment in new computational facilities is strongly recommended, since simulation now cost-effectively complements experimentation in the pursuit of the answers to numerous scientific

questions. New facilities should strike a balance between capability computing for those “heroic simulations” that cannot be performed any other way and capacity computing for “production” simulations that contribute to the steady stream of progress.

- Investment in hardware facilities should be accompanied by sustained collateral investment in software infrastructure. The efficient use of expensive computational facilities and the data they produce depends directly on multiple layers of system software and scientific software, which, together with the hardware, are the “engines of scientific discovery” across a broad portfolio of scientific applications. Additional investments in hardware facilities and software infrastructure should be accompanied by sustained collateral investments in algorithm research and theoretical development. Improvements in basic theory and algorithms have contributed as much to increases in computational simulation capability as improvements in hardware and software over the first six decades of scientific computing.
- Computational scientists of all types should be proactively recruited with improved reward structures and opportunities as early as possible in the educational process so that the number of trained computational science professionals is sufficient to meet present and future demands. Sustained investments must be made in network infrastructure for access and resource sharing as well as in the software needed to support collaborations among distributed teams of scientists, in recognition of the fact that the best possible computational science teams will be widely separated geographically and that researchers will generally not be colocated with facilities and data.
- Federal investment in innovative, high-risk computer architectures that are well suited to scientific and engineering simulations is both appropriate and needed to complement commercial research and development. The commercial computing marketplace is no longer effectively driven by the needs of computational science.

Appendix C: Measures of Success

The strategic plan for ASCR builds on a long history of important research and facility outcomes, many of which have changed the way science is done in the world. The following is a brief partial list of important accomplishments enabled by ASCR.

- **Established First National Supercomputer Center.** In 1974, DOE established the National Magnetic Fusion Energy Computing Center (the predecessor to the National Energy Research Scientific Computing Center, NERSC), and pioneered the concept of remote, interactive access to supercomputers. Before that time, scientists using supercomputers had to travel to the location of the computer to use it. In addition, application scientists were able to use these computers only by submitting jobs and waiting for hours or days to see the output. The Mathematical, Information, and Computational Sciences (MICS) subprogram developed the first interactive operating system for supercomputers, Cray Time Sharing System (CTSS), as well as a nationwide network to allow remote users to have effective access to the computers. This operating system revolutionized access to supercomputers by enabling users to monitor their jobs as they executed. When the National Science Foundation (NSF) initiated its Supercomputer Centers program in the 1970s, the CTSS operating system was adopted by the San Diego Supercomputer Center and the National Center for Supercomputing Applications to enable users to access NSF's first CRAY machines.
- **Enabled Proof That the Universe Is Flat.** Since ancient times, the geometry of the universe has been a topic of speculation and inquiry. An international team of scientists reported in *Nature* that the universe is, as Euclid thought, flat. They obtained their results by combining cosmic microwave radiation data collected during an Antarctic balloon flight and extensive analysis, amounting to 50,000 hours of computer time, using NERSC's Cray T3E supercomputer and software written at NERSC. Called BOOMERANG for "Balloon Observations of Millimetric Extragalactic Radiation and Geophysics," the collaboration includes over two dozen researchers from seven countries. Supercomputers at NERSC, along with software developed there, were crucial to extracting fundamental cosmological parameters from the data, the largest and most precise set of cosmic microwave background (CMB) data yet collected. From the dataset, the BOOMERANG team was able to make the most detailed map of the CMB's temperature fluctuations ever seen. From a CMB map, cosmologists derive a "power spectrum," a curve that registers the strength of these fluctuations on different angular scales and that contains information on such characteristics of the universe as its geometry and how much matter and energy it contains. The power spectrum derived from the BOOMERANG Antarctic flight data is detailed enough to allow the determination of fundamental cosmic parameters to within a few percent, indicating that the geometry of the universe is flat and that the universe will expand forever. The calculation required would have taken almost six years to complete if run on a desktop personal computer. On the NERSC Cray T3E, processing time over the life of the project totaled less than 3 weeks.
- **Contributed to the Development of the Internet: Slow-Start Algorithm for the Transmission Control Protocol.** Transmission Control Protocol (TCP), part of TCP/IP (Internet Protocol), is responsible for ensuring that packets arrive at their destination. In 1987, as DOE and the other Federal agencies were interconnecting their networks to form the core of the Internet, critical parts of the infrastructure began to fail. There was concern

that this represented a fundamental flaw in the TCP/IP architecture; however, a researcher at LBNL applied ideas from fluid flow research to understand the problem and develop a solution. This new TCP algorithm was incorporated in virtually every commercial version of Internet software within 6 months and enabled the Internet to scale from a small research network to today's worldwide infrastructure.

- **Developed High Speed Interconnects for Supercomputers: High Performance Parallel Interface.** In order to develop a standard interface between supercomputers and other devices, such as disk arrays and archival tape systems, and visualization computers, DOE laboratories developed the High Performance Parallel Interface (HiPPI) and led a consortium of vendors to make it the industry standard for the highest bandwidth interconnects between computers and peripheral devices. Many research issues in high-speed signaling, data parallelism, and high-speed protocol design had to be understood to enable this advance.
- **Led the Transition to Massively Parallel Supercomputing: Parallel Virtual Machine (PVM) and Message Passing Interface (MPI).** DOE researchers developed PVM and MPI to enable scientists to make effective use of networks of workstations and massively parallel computers. Both of these software packages have become standards in the industry and are implemented by virtually all of the high performance computer manufacturers in the world. Both of these developments were enabled by over a decade of basic research in message passing and distributed computing supported by DOE along with many experiments to apply these techniques to real scientific problems.
- **Achieved Unprecedented Levels of Performance on Large-Scale Computers.** MICS-supported researchers from U.S. Department of Energy laboratories have been honored by the high-performance scientific computing community as recipients of the Gordon Bell prize at several prestigious SupercomputingXX conferences. The 1998 award was for "best performance" of a supercomputing application; two other awards, in 1999 and 2001, were in the "special category," which emphasizes high-quality algorithms and software libraries.
- **Laid Mathematical Foundations for High-Performance Computing: Numerical Linear Algebra Libraries.** Today's high-performance scientific computations rely on high-performance, efficient libraries of numerical linear algebra software: LINPACK, EISPACK, LAPACK, SCALAPACK. These libraries, which are the core of numerical efforts in the solution of differential and integral equations, are the direct result of decades of DOE funding of basic research in this area. The libraries are used by thousands of researchers worldwide and are a critical part of the world's scientific computing infrastructure.
- **Extended the Frontiers of High-Performance Networks for Science:** Scientific simulations to meet Office of Science missions frequently involve accessing large data files on the order of millions to billions of megabytes. Measurements, experiments, and simulations at many locations around the world are generating these data files. Reliable access to these data requires investments in high-speed, high-bandwidth networks, and in robust, efficient network software. To highlight the special features of these requirements, the supercomputing conference series initiated a Network Bandwidth Challenge in 2000, in which researchers were invited to demonstrate their ability to maximize network

performance for their application. In both 2000 and 2001, the first prize for optimal use of the network went to a DOE laboratory-led application. In 2001, the prize-winning application was based on an interactive scientific simulation running at two separate supercomputers. The results of the simulation were sent to the conference floor over the network and visualized at a sustained network performance level of 3.3 gigabits per second, or approximately 1,000 times faster than commercially available DSL.

- **Promoted Education of Outstanding Graduate Students.** The Computational Science Graduate Fellowship Program has been in existence for 10 years supporting over 120 students at approximately 50 universities. These graduates have gone on to leadership positions in government laboratories, universities, and industry.
<http://www.krellinst.org/csgf/mag/alumni.cgi>
- **Enabled Management of Scientific Data.** Terascale computing and large scientific experiments produce enormous quantities of data that require effective and efficient management. The task of managing scientific data is overwhelming. Researchers at Lawrence Berkeley National Laboratory have developed a specialized index for accessing very large datasets that contain a large number of attributes that may be queried. This new index performs 12 times faster than the previous best-known method and 100 times faster than conventional indexing methods in commercial database systems. Researchers in high-energy physics and combustion modeling are using the prototype index.

Building on this history of success, ASCR continues to pursue its mission. ASCR uses a number of techniques to measure its progress.

- Expert review by committees of scientists organized as subpanels of ASCAC
- Peer review to ensure scientific quality
- Management reviews of high performance computing and network facilities
- Tracking of annual targets and long-term outcomes through the OMB Program Assessment Rating Tool

Appendix D: List of Workshops

- Blueprint for Future Science Middleware and Grid Research and Infrastructure, August 2002
 - <http://www.nsf-middleware.org/MAGIC/default.htm>
- DOE Science Network Meeting, June 2003
 - <http://gate.hep.anl.gov/may/ScienceNetworkingWorkshop/>
- DOE Science Computing Conference, June 2003
 - <http://www.doe-sci-comp.info>
- Science Case for Large Scale Simulation, June 2003
 - www.pnl.gov/scales/
- Workshop on the Road Map for the Revitalization of High End Computing
 - <http://www.cra.org/Activities/workshops/nitrd/>
- Cyberinfrastructure Report
 - <http://www.cise.nsf.gov/evnt/reports/toc.htm>
- ASCR Strategic Planning Workshop
 - <http://www.fp-mcs.anl.gov/ascr-july03spw>