Lawrence Livermore National Laboratory

Computation Directorate

Enabling Science and Technology

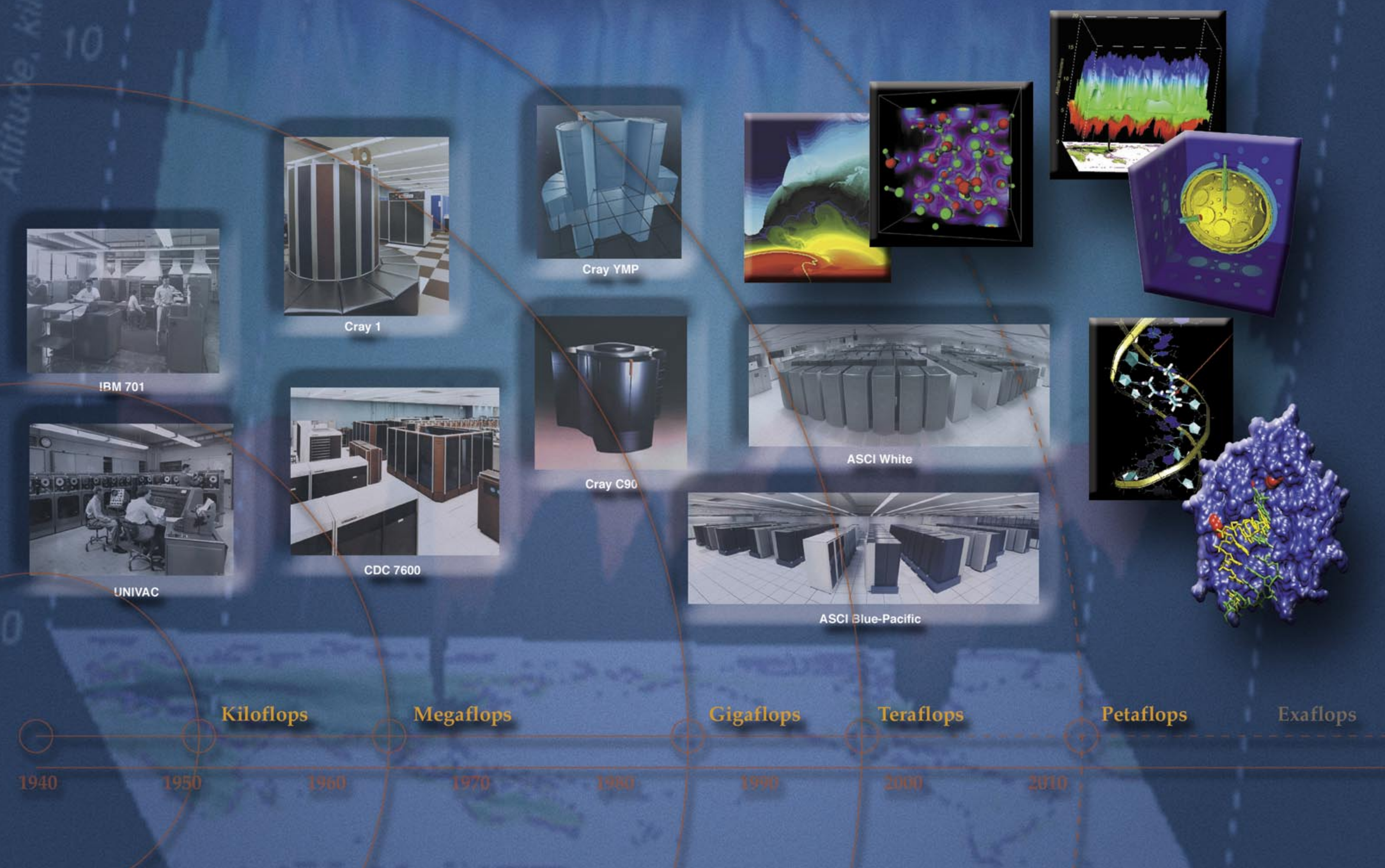
2003 Annual Report

UCRL-TR-202965

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

Contents

Advances in High-Performance Computing at LLNL



Dona L. Crawford
Associate Director for Computation

Big computers are icons: symbols of the culture, and of the larger computing infrastructure that exists at Lawrence Livermore. Through the collective effort of Laboratory personnel, they enable scientific discovery and engineering development on an unprecedented scale. For more than three decades, the Computation Directorate has supplied the big computers that enable the science necessary for Laboratory missions and programs.

Livermore supercomputing is uniquely mission driven. The high-fidelity weapon simulation capabilities essential to the Stockpile Stewardship Program compel major advances in weapons codes and science, compute power, and computational infrastructure. Computation's activities align with this vital mission of the Department of Energy.

Increasingly, non-weapons Laboratory programs also rely on computer simulation. World-class achievements have been accomplished by LLNL specialists working in multi-disciplinary research and development teams. In these teams, Computation personnel employ a wide array of skills, from desktop support expertise, to complex applications development, to advanced research.

Computation's skilled professionals make the Directorate the success that it has become. These individuals know the importance of the work they do and the many ways it contributes to Laboratory missions. They make appropriate and timely decisions that move the entire organization forward. They make Computation a leader in helping LLNL achieve its programmatic milestones. I dedicate this inaugural Annual Report to the people of Computation in recognition of their continuing contributions.

I am proud that we perform our work securely and safely. Despite increased cyber attacks on our computing infrastructure from the Internet, advanced cyber security practices ensure that our computing environment remains secure. Through Integrated Safety Management (ISM) and diligent oversight, we address safety issues promptly and aggressively. The safety of our employees, whether at work or at home, is a paramount concern.

Even as the Directorate meets today's supercomputing requirements, we are preparing for the future. We are investigating open-source cluster technology, the basis of our highly successful Mulitprogrammatic Capability Resource (MCR). Several breakthrough discoveries have resulted from MCR calculations coupled with theory and experiment, prompting Laboratory scientists to demand ever-greater capacity and capability. This demand is being met by a new 23-TF system, Thunder, with architecture modeled on MCR.

In preparation for the "after-next" computer, we are researching technology even farther out on the horizon—cell-based computers. Assuming that the funding and the technology hold, we will acquire the cell-based machine BlueGene/L within the next 12 months.

Achievements in 2003

Looking back over 2003, and looking forward to the possibilities of the next few years, I am excited about the ways Computation has enabled broad scientific opportunities and facilitated discovery, and the many ways we will continue to contribute to future LLNL initiatives. A number of achievements in 2003 enabled Laboratory initiatives, programs, and projects. They are briefly described below.

- **Major new computer capabilities augment LLNL's classified and unclassified HPC environment**
  The classified environment added three machines with more than 21-TF peak capability;

the unclassified 23-TF IA64 Thunder cluster began science runs.

- **HPC clusters using open-source systems software become available for production**
  Together with several industrial partners, Computation deployed a new version of the LLNL Linux Software Stack (operating system, parallel file system, and resource management system). Production-quality, large-scale HPC Linux clusters are now a reality. The multi-cluster simulation environment based on a single Lustre File System is impacting every program at LLNL.

- **New algorithms facilitate largest-ever ALE3D simulations**
  Within the nation's Stockpile Stewardship Program, large three-dimensional structural dynamics simulations are now performed on meshes with 610 million degrees of freedom, using 4032 processors of ASCI White. This is 100 times larger than the simulations of only three years ago, with 10 times the number of processors.

- **ICCS software operates in NIF Early Light**
  The NIF Integrated Computer Control System software is more than 75% complete. ICCS is used to commission and operate the first four beams as part of the successful NIF Early Light campaign, demonstrating NIF's end-to-end capability.

- **Bioinformatics efforts expand national biodefense**
  Computation partnered with NAI, BBRP, and Engineering to develop the Biological Aerosol Sentry and Information System (BASIS). This system enables early detection of biological pathogens. Computation researchers also developed a new, parallel algorithm that is ten times faster than the previous version and will enable the processing of genomes of larger organisms.

- **ARGUS integrated security system deploys to LANL**
  ARGUS, DOE's standard high-security system, protects assets at LLNL, Pantex, INEEL, DOE HQ, and now LANL. ARGUS includes personnel access control booths, alarm stations, map-based alarm reporting systems, and a closed circuit TV video assessment system.

- **Security improves through centralized management of usernames and passwords**
  Almost 100 LLNL business applications now authenticate through a single system, providing enhanced security of institutional services. Remote access is available through the One-Time Password (OTP) authentication system.

- **Major cyber vulnerabilities thwarted**
  Through internal collaboration, Windows vulnerabilities were identified and a response process was defined to both combat the vulnerabilities and to alert LLNL programs of the potential threat from malicious computer code. Several government and commercial sites were affected by the malicious-exploit computer code requiring them to disconnect from the Internet, but there were no major infections or operational impacts to the Laboratory.

- **Computation leads Lab-wide SQA efforts**
  Computation spearheaded the development of the institutional Software Quality Assurance (SQA) policy that was recently approved. The policy calls for a multi-tiered, risk-based, tailorable approach to software quality assurance and engineering practices. The Directorate is coordinating the development of an implementation plan.

- **Computation pilots the DHS ASC program**
  Computation is leading the Advanced Scientific Computing (CASC) research program for the Department of Homeland Security (DHS). LLNL organized a national workshop to define program needs and is now building a national program

Recognition and Awards in 2003

In addition to these major accomplishments, it is always gratifying to see Computation personnel and programs recognized by others for their outstanding contributions to the Directorate, the Laboratory, or the profession. A summary of such awards is noted below.

Teller Fellowship Award
Michel G. McCoy and Mark K. Seager received the fourth annual Teller Fellowship Award. Each award allows the recipient to do a year of self-directed work that will benefit the Laboratory. Mike and Mark plan to recruit a computer architect to focus on analyzing technologies that might scale to petaflop-class systems (10^{15} operations per second, peak speed) and beyond. Their goal is to explore cluster and alternative technologies for petaflop-class systems later this decade.

R&D 100 Award
Tom Slezak, Linda Ott, and Mark Wagner received a prestigious R&D 100 Award for their contributions to the BASIS team. BASIS, as described above, permits early detection of biological pathogens, and detectors have been successfully deployed at multiple locations across the country. Livermore team members serve four directorates: Nonproliferation, Arms Control and International Security; Biology and Biotechnology Research Program; Computation; and Engineering. An additional participant came from Los Alamos National Laboratory.

Service to the SC Conference Series
James McGraw chaired SC 2003, the premier conference in high-performance computing and networking. The annual conference, held in Phoenix, AZ, attracted more than 7,600 attendees, making it the largest and most successful in the 15-year history of the series.

Engineering Profession Distinguished Service Award
Linda Dibble accepted the Engineering Profession Distinguished Service Award from the San Joaquin Engineering Council. This award recognized her community outreach activities for the Lab, including the Tri-Valley Science & Engineering Fair and the Pleasanton Partnership in Education, among many others. Additionally, Linda has been a member of the Expanding Your Horizons consortium board, and since 1998 has co-chaired its annual San Joaquin conference, designed to nurture girls' interest in science and math courses, and to encourage them to consider science and math based career options such as engineering, computer science and biometrics.

Data Mining Awards and Recognition
Chandrika Kamath and Erick Cantú-Paz, both of the Sapphire Project, were issued a U.S. patent for a Parallel Object-Oriented Data Mining System.

Figure 1.00-1 The Terascale Simulation Facility will showcase ASCI Purple in 2005.

Erick Cantú-Paz is one of five initial fellows named to the International Society for Genetic and Evolutionary Computation.

Annual Report Overview

This first-ever Computation Directorate Annual Report compiles snapshots of programs and projects across the Directorate. It is divided into sections that explain and describe facets of our work. The following synopsis briefly outlines the report.

Section 2
Providing Desktops to Teraflops Computing
In 1995, at the beginning of the Advanced Simulation and Computing Program (originally the Accelerated Strategic Computing Initiative, or ASCI), we examined the kinds of physical phenomena we would need to simulate, when we would need to generate these simulations, and how quickly we would need calculation results returned. This analysis determined the computers we would acquire through partnerships with industry leaders. Our goal was to obtain, by 2004, a computer system capable of 100 trillion floating-point operations per second (100 TF).

Livermore, Los Alamos, and Sandia, the three national laboratories involved in ASCI, have fielded increasingly powerful massively parallel supercomputers. ASCI Purple, arriving at Livermore mid-2005, will fulfill the original 100-TF goal. But the story does not end there. Successful simulation environments require more than huge computers with maximum peak speeds. They require computing and communications environments integrated from desktops to teraflops, with associated support and services at all levels. They require infrastructure—storage systems, visualization capabilities, networks, compilers, and debuggers—all working together.

Section 3
Developing Applications Software
World-class science on advanced architectures such as ASCI Purple, Thunder, or BlueGene/L also requires the expertise of individuals who can direct advanced applications development. It requires code development, physics modeling, and algorithms improvements; it requires computer applications runs and analysis; it requires computer security compliance and technology integration; and, it requires information technology expertise.

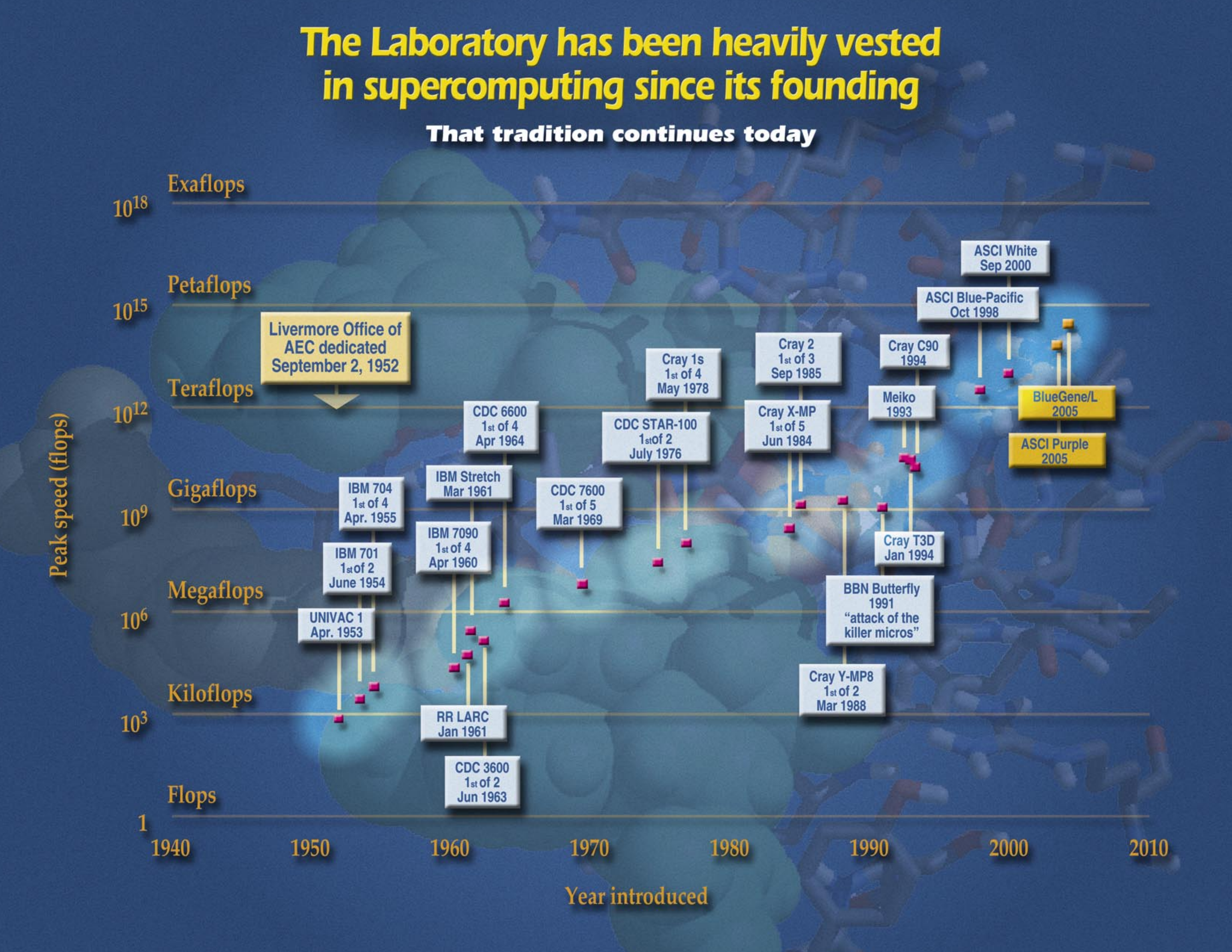
In addition, many projects require personnel who have real-time systems expertise, database management capability, specialized systems management capability, specialized systems knowledge, or specialized backgrounds in a particular area of computer science or mathematics. Regardless of the individual's background or project assignment, the work is undertaken in a balanced and integrated manner using a systems approach.

Section 4
Computing Research and Development
Directorate researchers actively advance the computational technologies that facilitate Laboratory terascale scientific simulation. Our research is broad in scope and consistent in vision. It enables Laboratory scientists to harness massively parallel machines with thousands of processors for predictive simulation of complex physical phenomena.

The Directorate fosters numerous research projects, including scalable numerical algorithms, discretization methodologies, object-oriented and component-based software, multiresolution data management and visualization, and system software. Computation personnel collaborate with programmatic partners to build and then to use these technologies in breakthrough scientific investigations in the defense, environmental, energy, and biological sciences.

Section 5
Additional Information
None of our work is performed in a vacuum. We collaborate often and extensively with almost 80 national laboratories, academic institutions, and industrial partners. We actively seek the innovation, sound judgment, and disciplined execution that lead to the collective success of the Laboratory's mission. Our partners and collaborators are listed in Section 5. Additionally, a cross-reference of acronyms is provided to assist the reader unfamiliar with LLNL abbreviations.

Section 2



The Laboratory has been heavily vested in supercomputing since its founding
That tradition continues today
Exaflops
Petaflops
Teraflops
Gigaflops
Megaflops
Kiloflops
Flops
10^{18}
10^{15}
10^{12}
10^{9}
10^{6}
10^{3}
1
Peak speed (flops)
1940
1950
1960
1970
1980
1990
2000
2010
Year introduced
Livermore Office of AEC dedicated September 2, 1952
UNIVAC 1 Apr. 1953
IBM 701 1st of 2 June 1954
IBM 704 1st of 4 Apr. 1955
IBM 7090 1st of 4 Apr 1960
IBM Stretch Mar 1961
RR LARC Jan 1961
CDC 3600 1st of 2 Jun 1963
CDC 6600 1st of 4 Apr 1964
CDC 7600 1st of 5 Mar 1969
CDC STAR-100 1st of 2 July 1976
Cray 1s 1st of 4 May 1978
Cray X-MP 1st of 5 Jun 1984
Cray 2 1st of 3 Sep 1985
Cray Y-MP8 1st of 2 Mar 1988
BBN Butterfly 1991 “attack of the killer micros”
Meiko 1993
Cray C90 1994
Cray T3D Jan 1994
ASCI Blue-Pacific Oct 1998
ASCI White Sep 2000
BlueGene/L 2005
ASCI Purple 2005

2.00 Providing the Desktops to Teraflops Computing Environment

The Computation Directorate provides LLNL with a world-class computing and networking environment capable of meeting laboratory mission and program needs. We support the Laboratory's computing and communications infrastructure, spanning users' needs from the desktop to the high-performance computing platforms. We assume responsibility for planning and operating the scientific computing facilities, developing tools that enable effective use of these facilities, providing expertise in desktop support, and running as site-wide network backbone for both classified and unclassified systems. We also undertake essential computational, communication, and computer security research required to sustain this computing environment.

At the highest level, two broad objectives are ubiquitous. First, we seek coordination of services so that the connections between the office environment and the HPC infrastructure become increasingly transparent. Second, from a strategic perspective, we require a framework for a petaflop strategy. Such a framework includes both strategic investments in simulation environments as well as in innovative architectures. The body of this Section reflects both of these objectives.

Sustaining a world-class scientific computing environment demands careful balancing of system components and planning of changes to exploit constant improvements in the technology. This environment consists of more than just the latest supercomputers. It includes production computing resources, fast I/O subsystems, high-capacity archival storage facilities, and high-speed network interconnects to link all of these components. Future planning includes detailed predictions of our expected needs and close collaborations with industry to tailor appropriate and cost-effective solutions. The first two reports in this Section describe our accomplishments and strategic planning for building, operating and evolving our high-performance computing systems and networks.

As these computing systems grow in size and complexity, the challenge of using them efficiently and effectively grows as well. To help users address this challenge, we develop tools, system software, and an application infrastructure, again in partnership with industry. We also provide user services, training, documentation, and consulting. The next two reports in this Section describe some of our most critical activities to support our users' ability to use the large systems well. These systems can generate terabytes and petabytes of data far faster than we can view and assess. One of the reports describes our latest development and deployment activities for visualization and data assessment. The next report describes additional support for HPC users, including access to information on how to use these systems, optimization tools and techniques, and strategies for submitting and tracking large production jobs that could take weeks or months to complete.

The last three reports in this Section complete the landscape needed by users. They must be able to work from their desktops, use backbone networks to access the HPC resources and external networks, and do everything in a safe, secure manner. The desktop environment efforts need to anticipate, integrate, communicate, and implement the information technology requirements of LLNL's programs and the institution. This includes technical support for Macintosh, Window, and Unix systems and servers, local help desks, and Web page development services. The network backbone must provide secure, reliable, effective access to information and computing resources from the desktop by delivering networks, centralized system administration services, and centralized enterprise services. Cyber security continues to be extremely challenging because we must maintain our ability to communicate and learn via Internet access and at the same time fend off increasingly sophisticated and persistent efforts to gain access to our internal networks.

Figure 2.00-1. The Multiprogrammatic Capability Resource combines open source software with cluster architecture to provide Advanced Simulation and Computing-level supercomputing power for unclassified research.

2.01 Platform Strategy and Systems, in Production and Planned

Problem Description

Our strategic and industrial collaborations in HPC center on delivering computing platforms to production environments in support of programs of national interest, with an eye to enabling realistic ramps to petaflop-scale systems. To that end, we work with multiple sources of computing technology to judge the boundary between promises and real computing capability, as well as to distinguish industrial trends. Based on these interactions with the computing industry we have developed a straddle strategy to deliver technology to the ASCI program and to the institution.

Technical Approach/Status

The strategy is depicted in Figure 2.01-1. The ASCI program has enjoyed success by riding Curve #1, the **Proprietary, Vendor Integrated SMP** cluster technology. As part of the ASCI tri-laboratory complex, LLNL and LANL played a significant role in establishing this technology path with the ASCI Blue procurement. This approach has taken the tri-lab HPC community from about 50 GF in CY1995, to 12.3 TF on ASCI White in late CY2000, and all the way to 20 TF on ASCI Q in late 2003. Curve #1 price–performance, however, is being eclipsed by Curve #2, **Open Source Commodity Clusters** with the Linux operating system (Beowulf technology).

In addition, on Curve #3, **Innovative Concepts**, such examples as IBM's system-on-a-chip (SOC) technology for embedded applications also have the potential for extreme price performance. Each of these curves has a different price–performance and risk trajectory. The SOC design shows high risk for ASCI-scale platforms today, but holds the promise of an affordable petaflop compute engine in the 2006–2008 timeframe. Thus, to simultaneously maximize the benefit and optimize the potential of emerging technologies, while also extracting benefits from mature technologies, we are engaging development of platforms and software on all three curves. When this strategy was launched, the intent was to move the most important programmatic work onto Curves #2 and #3 only when these technologies had matured, featured reduced risk, and were well understood as production environments.

During 2003, we saw a positive outcome of this approach. Experience with Thunder (IA64) and MCR (IA32) showed LLNL that Curve #2 could potentially deliver at ASCI scale with medium risk. We indicate this change in the trajectory by moving the curve to the left from the dashed blue (old) to the solid blue (new). With this new price–performance reality in mind, the Laboratory has renegotiated the ASCI Purple contract with IBM. This approach represents a win–win, because it accelerates IBM's trajectory in fielding very large, low-cost systems with the Power series, expanding the space of such solutions beyond INTEL and AMD. This will help maintain healthy competition and low costs.

Progress in 2003

A number of powerful systems were sited at LLNL in 2003—Thunder, Lilac, Violet, and Magenta. Other systems have been brought into production, MCR and ALC; and others have seen major upgrades — the Penguins, Adelie and Emperor. Much of this has been accomplished with the CHAOS (Clustered High-Availability Operating System) software stack.

In short, CHAOS augments the standard Linux Red Hat distribution with support for HPC clusters, including scalable system management and monitoring tools (primarily developed at LLNL), a high-performance interconnect (Quadrics Èlan3 and Èlan4), the Lustre parallel file system from CFS (heavily funded by ASCI) and an advanced resource manage-

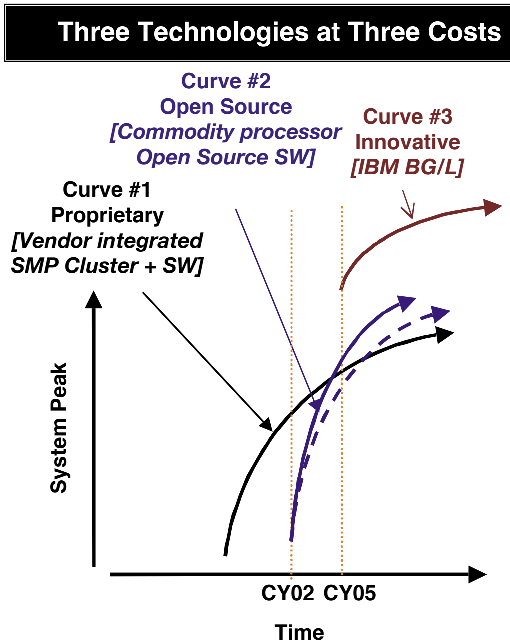


Figure 2.01-1. Price–performance straddle strategy to deliver ASCI technology.

ment and control apparatus (Simple Linux Utility Resource Management or SLURM), also developed at LLNL. These represent the tools necessary to field production HPC clusters, allow LC to integrate new systems rapidly, leverage in-house expertise to provide fast turnaround on bug reports and feature enhancements, and provide a framework for release management. The LC carefully keeps kernel modifications to a minimum to reduce friction with new releases, and leverages its relationship with Red Hat to get new releases into their distribution. It was the experience with this tool on computers of scale that accelerated LLNL's transition onto Curve #2.

Figure 2.01-2 summarizes recent progress. Systems noted in red are either new or have changed significantly in status during the past year, bringing the total peak across all systems on site or under integration close to 90 TF. By this relatively crude metric, LC is currently one of the larger HPC sites in the world.

Looking to 2005, the two IBM contract systems, the 100-TF Purple system and the 180–360-TF BlueGene/L (BG/L) system, seem remarkably well-aligned with the two foci of the new strategy for Advanced Simulation and Computing (still known as ASCI). These foci are: integration into the broader program through providing essential support for stockpile stewardship deliverables, in particular for Directed Stockpile Work (DSW) to support re-certification of weapons systems; and continual reduction in the phenomenology in the weapon simulation codes, including a deeper understanding, in quantitative terms, of their limitations. The latter is particularly important as weapons wander from their test base through aging. From this perspective, the Purple system will provide the must-have, time-critical cycles to the classified program in a highly reliable production environment.

BG/L will provide a research platform to increase prediction by understanding materials properties well enough to reduce phenomenology in the ASCI applications codes. It will contribute at all length and time scales for multiscale materials models (Figure 2.01-3). One can therefore look at the two systems as computational components, each vectored at a different aspect of the program strategy. This approach thus provides a vendor-integrated solution at Beowulf cluster cost and at scales heretofore unattained.

System	Program	Manufacturer & Model	Operating System	Interconnect	Nodes	CPUs	Memory (GB)	Peak GFLOP/s
Unclassified Network								47,847
Thunder	M&IC	California Digital	CHAOS 2.0	Elan4	1024	4,096	8,192	22,938
ALC	ASCI	IBM xSeries	CHAOS 1.2	Elan3	960	1,920	3,840	9,216
MCR	M&IC	Linux NetworX	CHAOS 1.2	Elan3	1152	2,304	4,608	11,059
Frost	ASCI	IBM SP	AIX 5.1	Colony DS	68	1,088	1,088	1,632
Blue	ASCI	IBM SP	AIX 5.1	TB3	264	1,056	396	701
TC2K	M&IC	Compaq SC ES40	Tru64 5.1b	Elan3	128	512	280	683
iLX	M&IC	RAND Federal	CHAOS 1.2	N/A	67	134	268	678
GPS	M&IC	Compaq GS320/ES45	Tru64 5.1b	N/A	49	160	344	277
PVC	VIEWS	Acme Micro	CHAOS 1.2	Elan3	64	128	128	614
Riptide	VIEWS	SGI Onyx2	Irix 6.5.13f	8 IR2 Pipes	1	48	37	24
Qbert	M&IC	Digital 8400	Tru64 5.1b	MC 1.5	2	20	24	25
Classified Network								41,171
Violet (pEDTV)	ASCI	IBM P655	AIX	Federation	128	1,024	2,048	6,144
Magenta (pEDTV)	ASCI	IBM p655	AIX	Federation	128	1,024	2,048	6,144
Lilac (xEDTV)	ASCI	IBM xSeries	CHAOS 1.2	Elan3	768	1,536	3,072	9,186
White	ASCI	IBM SP	AIX 5.1	Colony DS	512	8,192	8,192	12,288
Ice	ASCI	IBM SP	AIX 5.2	Colony DS	28	448	448	672
Blue-Pacific (S)	ASCI	IBM SP	AIX 5.3	TB3	488	1,952	1,164	1,296
Adelie	ASCI	Linux NetworX	CHAOS 1.2	Elan3	128	256	512	1,434
Emperor	ASCI	Linux NetworX	CHAOS 1.2	Elan3	128	256	512	1,434
Ace	ASCI	Rackable Systems	CHAOS 1.2	N/A	128	256	512	1,434
SC Cluster	ASCI	Compaq ES40/ES45	Tru64 5.1b	N/A	40	160	384	235
ICF Cluster	ICF	Compaq ES40/DS10L	Tru64 5.1b	N/A	12	36	12	48
GVIZ	VIEWS	Rackable Systems	CHAOS 1.2	Elan3	64	128	256	717
Whitecap	VIEWS	SGI Onyx3800	IRIX 6.5.13F	4 IR3 Pipes	1	96	96	77
Tidalwave	VIEWS	SGI Onyx2	Irix 6.5.13f	16 IR2 Pipes	1	64	24	38
Edgewater	VIEWS	SGI Onyx2	Irix 6.5.13f	10 IR2 Pipes	1	40	18	24

Figure 2.01-2. Progress summary: LC Systems either in production in March 2004, or currently being integrated, show rapid growth due to the maturation of cluster technologies.

The external community, including scientists from Office of Science and NSF laboratories, reviewed BG/L partnership progress twice in 2003. Both reviews endorsed the rate and quality of progress and encouraged increased engagement with the broad community. The July review panel, chaired by Mike Levine of Pittsburgh Supercomputer Center, reported that this work represented, "an important opportunity to substantially advance both the ASCI mission agenda and the development of very large-scale machines." This computer represents the most complex SOC design that IBM has ever built, yet no critical problems were found with first hardware. The first compute ASICs (application-specific integrated circuits) were delivered June 6, 2003, were booted, and were running applications two weeks thereafter. A 512-node prototype (128th scale of full system, Figure 2.01-4) ran Linpack at 1.412 TF using a slow clock (500 Mhz) rather than the final clock (666–700 Mhz) processors. The MPI latency is expected to come in at 4.3 µs (besting our 7 µs target). Further, IBM is achieving exceptionally high network bandwidth (~80% of theoretical max), accessible even at very small packet size (half the max speed at 500 bytes). Recently, 700-Mhz processors have been run using 512 MB of memory, increasing probability that the machine can be used for a wide variety of applications.

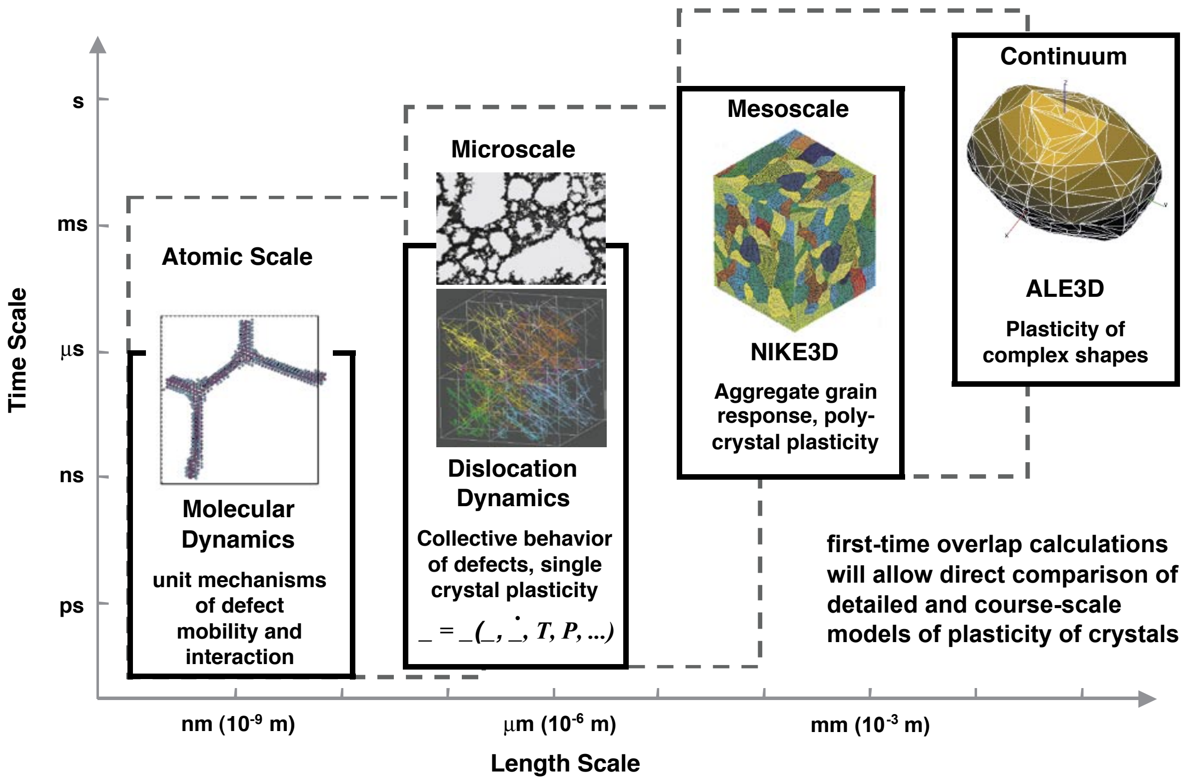


Figure 2.01-3. Multiscale comparisons: Length scales plotted as a function of time.

Early in 2004, Sandia, LANL and LLNL, with the concurrence of DOE HQ, agreed to move forward with the parts order in 2004. All three laboratories are targeting science applications at the system, to be delivered in phases in the first half of 2005. LLNL is funding the development of five applications, from ab initio molecular dynamics, to crack propagation, to turbulence. ASCI Alliances have been invited to submit proposals for access. LLNL is working with HQ to permit access by Office of Science laboratories collaborating with ASCI.

Figure 2.01-4. 512-node prototype system, operating at IBM Watson Research Center.

Problem Description

Translating LLNL's substantial investments in platforms and applications into successful science requires a balanced computer infrastructure. To ensure this, LC develops an annual **I/O Blueprint**, a planning document that collects platform I/O capabilities plus user requirements, and then presents architecture options, issues, action plans, deliverables, and budget scenarios along with a scope of work for the I/O infrastructure. I/O Blueprints have been used for many years at the LC to assure that investments across all ASCI infrastructure budgets are coordinated. ASCI White success depended heavily on the FY99 and the FY00 Blueprints.

The FY03 I/O Blueprint began by detailing a vision for the next three to five years, centered on an architecture in which computational and visualization resources share a high-performance parallel global file system to provide users with fast, cost-effective uniform access to a very large pool of online storage. This file system will be known as the Site-Wide Global File System (SWGFS), and is shown in Figure 2.02-1. Each of the I/O infrastructure teams began working toward this vision during the year. As outlined below, much progress has been made.

Progress on High-Level I/O Blueprint Deliveries in 2003

In the **networking** arena, LC continues to provide three independent networks to satisfy Labwide I/O needs: a small packet network tailored for interactive access from user desktops, a high-performance four-stripe parallel "jumbo" packet network to facilitate movement of large data sets, and a network for Center-wide NFS access. Anticipating the urgent need for more bandwidth, LC introduced a few 10-Gb Ethernet trunks into production and procured 50 more 10-Gb Ethernet ports. Requirements for greater connectivity were met by deploying newly available high port-density line cards, resulting in over 1300 Gigabit Ethernet ports on the unclassified network. The Visual Interactive Environment for Weapons Simulations (VIEWS) digital delivery effort collaborated with industry to demonstrate a technology for accessing LC's visualization resources from the desktop over the existing networks.

In the **archival** storage arena, LC deployed a new generation of high-performance archive-mover platforms and a Storage Area Network (SAN) disk cache. Upgrading to StorageTek's™ latest generation tape drives tripled tape performance and capacity. The High-Performance Storage System (HPSS) archival software underwent a significant upgrade and an ASCI PSE white paper addressed SWGFS/archive integration. The results of Blueprint-driven changes can be seen in Figure 2.02-2.

Network File System (NFS) upgrades were significant both in the amount of storage capacity provided and in the bandwidth offered. The upgrade plan included the use of RainStorage™ devices, which

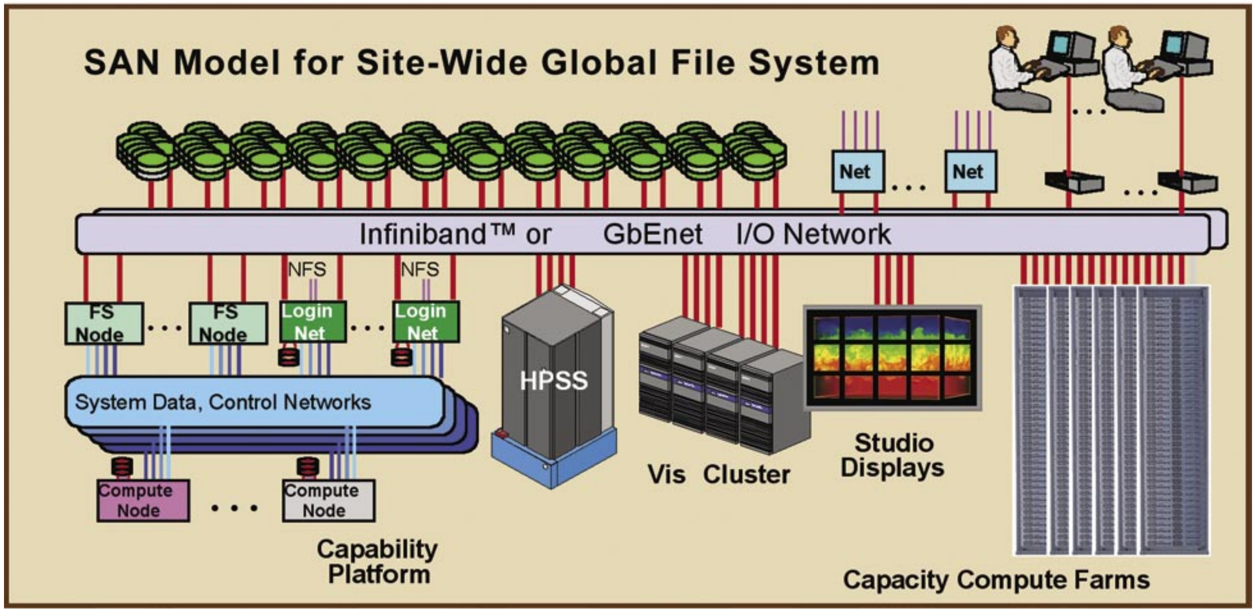


Figure 2.02-1. SWGFS architecture shares one file system, many computers and services.

made it possible to migrate to new NFS servers without unduly impacting normal operations. The upgrade allowed LC to respond to requirements for improvement in multiple areas: the addition of customer scratch space; a significant expansion in home directory space resulting in a large increase in customer quotas; the development of monitoring and planning tools; the deployment of a test platform for new server evaluation; and allowing selected user access to these NFS servers directly from their desktops.

The Center took its first steps toward **SWGFS in production** with the M&IC-funded compute resource MCR, closely followed by the ASCI-funded ALC cluster and the PVC visualization cluster. The Lustre file system, employed by "friendly users" since Fall 2002, is now providing acceptable functionality and performance on these systems. Today, Parallel Visualization Cluster (PVC) and MCR share a single global parallel file system, and the plan is to merge the IP-based Lustre storage infrastructure with the current high-performance parallel network infrastructure and make *all* the storage available to *all* hosts supporting the Lustre file system. This will allow the LC to move toward the visions of "one file system, many computers," and will cut costs significantly.

All of these accomplishments depended heavily on the existence of a flexible test bed infrastructure. The I/O Test Bed proof-of-concept environment continued as a vital tool used by all infrastructure components in identifying and developing reliable high-performance hardware and software solutions. Working together, and focused by the I/O Blueprint, networking, archive, NFS and SWGFS teams were able to provide LC customers with a balanced and world-class production environment for simulation science in 2003. We are also moving forward together toward state-of-the-art services for ASCI Purple and BG/L machines in 2005.

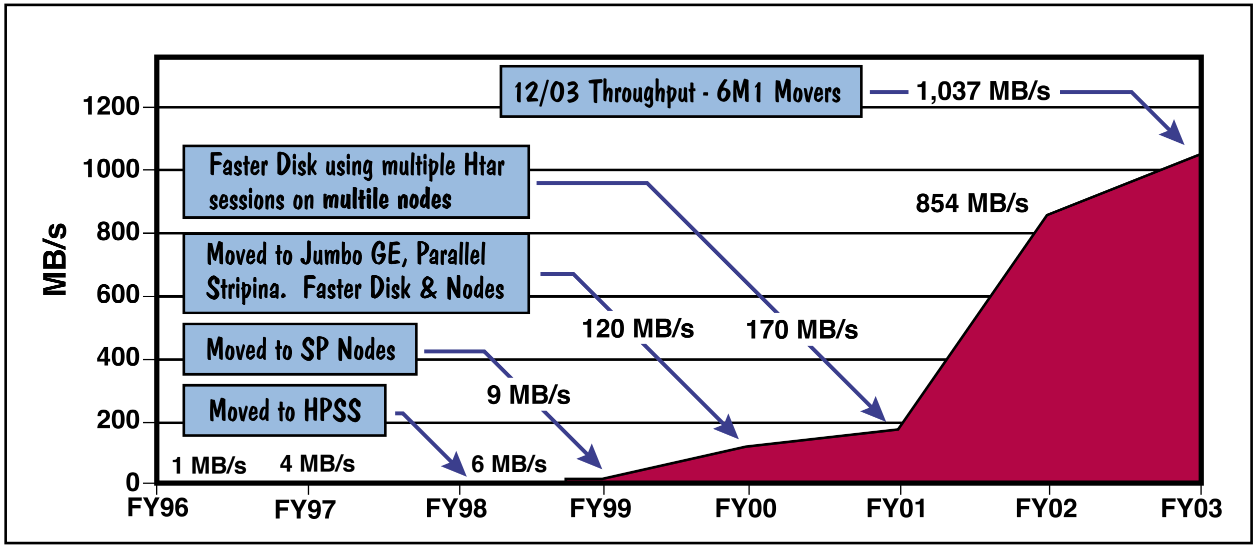


Figure 2.02-2. Maintaining SCF capability platform-to-HPSS performance ratios is an I/O Blueprint requirement.

Problem Description

The very large ASCI computers are prodigious data-generation engines. The I/O infrastructure is the communications and storage component necessary to preserve the data. The final component of the HPC architecture is the data assessment environment.

In support of Stockpile Stewardship, ASCI simulations require hardware and software tools to find, access, manipulate, and visualize the multi-terabyte scientific datasets resulting from large simulations, to compare results across simulations, and to compare between simulations and experiments. Traditional tools cannot cope with the size, scale, and complexity of terascale datasets. The challenge is to research, develop, and deploy tools that provide users the capability to "see and understand" their data. Although management and visualization of massive datasets is a problem addressed in other scientific and experimental contexts (e.g., satellite images, high-energy physics), accurate analysis of mesh-based ASCI simulation datasets of large magnitude is an ongoing challenge.

The LLNL strategy coordinates ASCI-supported research, development, engineering, deployment and applications support in visualization, data management and data exploration. A major direction targets research and development to create innovative technologies for scientific collaboration, data exploration, visualization, and understanding. Once the requisite technologies exist, they are integrated, tested, and evaluated by a representative set of users. Finally, the technologies are deployed in a generally available, operational and reliable environment for day-to-day use by ASCI users and applications. (Research achievements can be found in Section 4.) Here we concentrate on some of the more notable development and deployment activities for visualization and data assessment, with an emphasis on accomplishments in 2003.

Development and Deployment Activities in 2003

As noted earlier, the transition to a new architecture targeted toward clusters was completed through release of a new software stack. This layered approach provides application toolkits, as well as interfaces to standardized scaling, rendering, compositing, and image delivery libraries, in addition to job and session infrastructure tools. Included were new releases of (DMX) Distributed Multi-headed X11, Chromium, MIDAS and Telepath. DMX is an aggregate X11 server system. Chromium is a distributed, parallel OpenGL application-programming interface based on a dynamically filtered, streaming graphics model. MIDAS is an Open Source tool providing transparent, asynchronous transmission of application-generated imagery from remote visualization servers to desktops. Telepath supports the orchestration of a visualization session, including resource allocation, video switching and delivery and configuration of services.

Together, these packages represent a dynamic visualization applications environment capable of scaling with dataset size, display size and desired levels of performance, using commodity graphics enabled clusters. This software stack was first deployed on the new PVC visualization cluster, LLNL's first production, commodity, PC-based visualization engine. PVC provides direct visualization services for MCR and has already demonstrated its ability to handle multi-terabyte sets from several important codes. The PVC/MCR systems model will serve as a blueprint for future ASCI Purple-related deployments.

During 2003, the TeraScale Browser released its first production, out-of-core, surface-rendering engine, making it possible to render even the largest datasets at the end-user desktop. This release also included the first explicit support of DirectX 9-class graphics hardware in an end-user application, and has been measured to be up to five times faster than the previous release. VisIt, a powerful visualization tool used by DNT, had several phased releases this year, adding improved performance and new capabilities for scalable rendering, stereo, Adaptive Mesh Refinement (AMR) support and movie tools. VisIt was also ported to the ASCI Q machine at LANL, and to Mac OS X this year.

In addition, there was significant progress in the deployment of research tools for data discovery and data query (also see Section 4). Significant work continued to develop and deploy production-quality metadata and directory tool capabilities and fund development of interoperable data models and formats used by large LLNL ASCI code efforts.

Major enhancements were made to SimTracker (a data management tool) for high-level task coordination, remote data management and cross-site data sharing. SimTracker has been deployed with 18 simulation codes. Work has begun to add a comparison framework, change–audit features, and enhanced regression testing to SimTracker. Hopper, a graphical interface supporting HTAR, HPSS, SSH, and FTP, was released in beta form in 2003. It allows compact or detailed views of files and directories, a history mechanism, password management, and a search by name feature. Planned for Hopper in 2004 are content-based metadata searches, and closer integration with LLNL persistent-file transfer tools.

The Center also continued to upgrade ASCI data display capabilities (Figure 2.03-1) with new high-quality projectors and screen in the B132 Data Assessment Theater, a final layout design for the new Terascale Simulation Facility (TSF) Advanced Simulation Laboratory, procurement of high-bandwidth modems for high-resolution user office desktop displays, and evaluation of stereographic projectors.

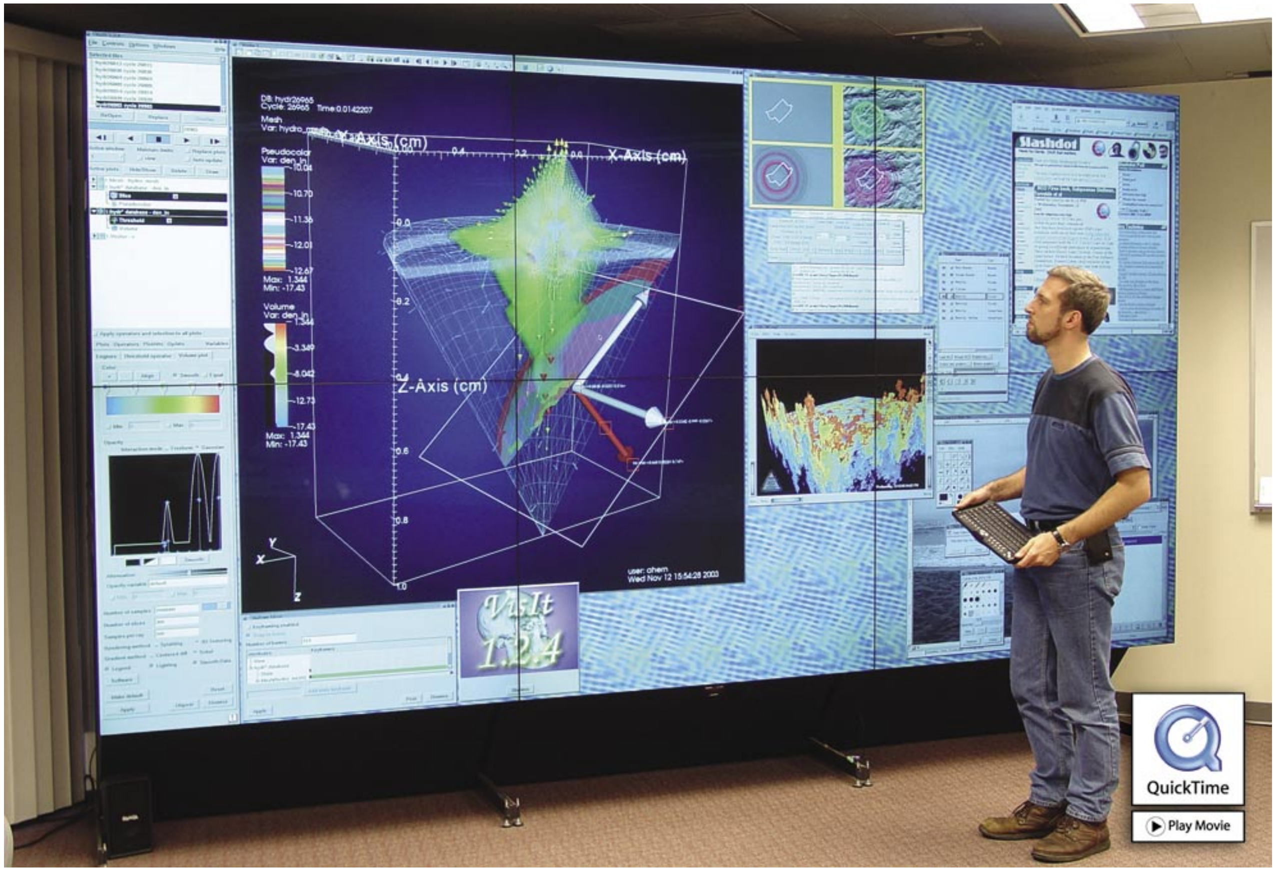


Figure 2.03-1. Interactive data analysis using multiple tools on a cluster-driven display wall.

Problem Description

Success in scientific computing is unlikely in the absence of first-rate consulting and services. Peak speed provides a potential for excellence, but quality service must be commensurate with the hardware on the floor. Our services encompass an "end-to-end cycle of user simulation science." Here, we allude to the cycle of user access, code development, job submission, and data assessment required by computational scientists. We develop tools, system software, and the application infrastructure that enable efficient and effective use of LLNL computing resources. We also provide the user services, training, documentation, and consulting that support the usability, accessibility, and reliable operation of LLNL's computing resources. The stages of this end-to-end cycle are described below.

Technical Approach and Progress in 2003

Stage 1: Enable unencumbered customer access to information to facilitate use of the HPC resources. We provide account services, online documentation, education and training, software quality assurance, and a tracking system for all reported questions and problems. In addition, a high-performance computing (HPC) helpdesk supports the HPC systems, and an institutional helpdesk provides desktop support services and consultation. The Remedy™ tracking system is being upgraded and deployed for several directorates. Both helpdesks ultimately will share this trouble-ticket database. Two such on-line documentation accomplishments for 2003 are noted below.

- Developed software to replace the DynaWeb™. legacy, proprietary document delivery system. The new web-based documentation system was written in XML, is searchable, and currently contains 36 user documents (4,000-plus pages of documentation).
- Replaced the legacy, proprietary document delivery system on the unclassified network with an XML web-based searchable system containing more than 4,000 pages of user documentation, including EZ manuals.

Stage 2: Provide support for scientists porting and optimizing their computer codes. We provide a stable leading-edge parallel application development environment that significantly increases application developers' productivity. Through collaborations with vendors and third-party developers, we ensure a robust environment with the most advanced development tools. We also assist scientists in the code development and maintenance process—compilers, libraries, debuggers, performance measurement and analysis tools, and I/O. Helping customers use these tools effectively speeds up the scientific applications, and therefore optimizes use of machine resources. Specific 2003 accomplishments focused on coding and performance tools, and

- Broadened our code development environment, including compilers, debuggers, performance analysis tools, I/O libraries, and expert consulting services to Linux environments. This contributed greatly to the success of the MCR, ALC, and Lilac Linux/Pentium/ Quadrics systems.
- Provided performance tools support, usage expertise, and code analysis that contributed to the successful completion of ASCI code milestones.

Stage 3: Simplify the submission and tracking of production jobs for users. The Livermore Computing Resource Manager (LCRM) tool provides users with a highly sophisticated fair-share scheduler for job and resource management of production computing assets. In addition, LLNL and Linux NetworX are jointly designing and developing SLURM, an open-source resource manager for Unix clusters. SLURM's primary functions are

- To manage a priority-ordered queue of pending work (typically parallel jobs).
- To allocate jobs exclusive or nonexclusive access to compute nodes.
- To provide a framework for initiating, monitoring, and managing jobs.

SLURM and LCRM form a critical part of the LLNL CHAOS cluster software stack, and together assure delivery of resources to the appropriate customer, with very high resource utilization (frequently exceeding 90%) in a contended environment with heterogeneous workloads. Specific 2003 accomplishments included these resource-management advances.

- LCRM v6.9 released April 2003: support for SLURM, better support for prioritizing jobs based on job-size, and replacement of the underlying NQS system with a new, streamlined TBS system.
- LCRM v6.10 released October 2003: support for visualization scheduling, variable job-sizes, and support for pre-emption on IBM systems. Visualization scheduling allows the user to specify a mix of viz and compute nodes.

Stage 4: Provide effective tools for data assessment (including visualization) of production computing runs. We provide extensive visualization and data-management support. Our experts develop and support tools on a wide range of platforms for users representing many disciplines, and for offices, theaters, work centers, and conference rooms. Services include consulting on scientific visualization packages, data-management tools and graphics utilities; authoring and producing movies and DVDs; demo support for PowerWalls; and the stewardship of three theaters. Specific 2003 accomplishments for data assessment included these.

- Produced visualizations in support of M&IC science runs and ASCI milepost reviews.
- Documented computational science results with Science of Scale video (with new material for SC2003).
- Supported numerous high-level demonstrations at all three PowerWalls for audiences from congressional staffers to military dignitaries.

Customer Profile, 2000 through 2003

Customer Profile	2000	2001	2002	2003
Number of Active Users	1959	2219	2304	2450
Classified	964	1080	1231	1360
Unclassified	1609	1826	1847	1904
Number of Remote Users	576	676	648	709
Sandia	85	128	133	163
LANL	70	88	108	119
ASCI Alliances	116	122	110	120
Other	305	338	297	307
Average Number of Hotline Contacts per day	100	110	116	109
4HELP Number of calls per day	63	65	62	77
Average Number of accesses to web pages per day	1080	7145	7308	8675
WWW Documentation				
Documents Available	33	37	35	37
Number of pages of documentation	5131	3798	4315	4331
Compiled Customer Assessments				
Number of classes offered	13	3	11	6
Class Evaluation	Avg. score 4.5 out of 5. Comments Favorable	4.9 out of 5.	4.7 out of 5.0	4.4 out of 5.0
Number of users registered	374	377 registered, 235 attended	514 registered, 426 attended	300
Number of sessions offered	33	20	34	41

Figure 2.04-1. Customer profile metrics help LC to monitor the quality of its services.

LC's success is evidenced by the customer profile metrics gathered over time (Figure 2.04-1), and the high effective-node utilization of the machines. Figure 2.04-2 shows how quickly our services and software helped users effectively utilize MCR when it opened to them; this high level of utilization continues today.

We also improved institutional user services significantly in 2003. Our institutional helpdesk acquired Open LabNet, One Time Password, and all remote-access support for the Laboratory. We increased phone coverage from two to three staff during the day, to four at all times during the second half of the year. Although call volume increased more than 14%, the abandonment rate dropped from 15.64% to 8.88%. In addition, during the second half of 2003, call wait dropped 38%. Remedy™ was successfully rolled out to NAI, NIF, Computation, and parts of Engineering, allowing all of these directorates to pass trouble tickets and gather important support metrics.

Our goal is to work with all LC customers so that we meet their time-critical requests to their full and complete satisfaction.

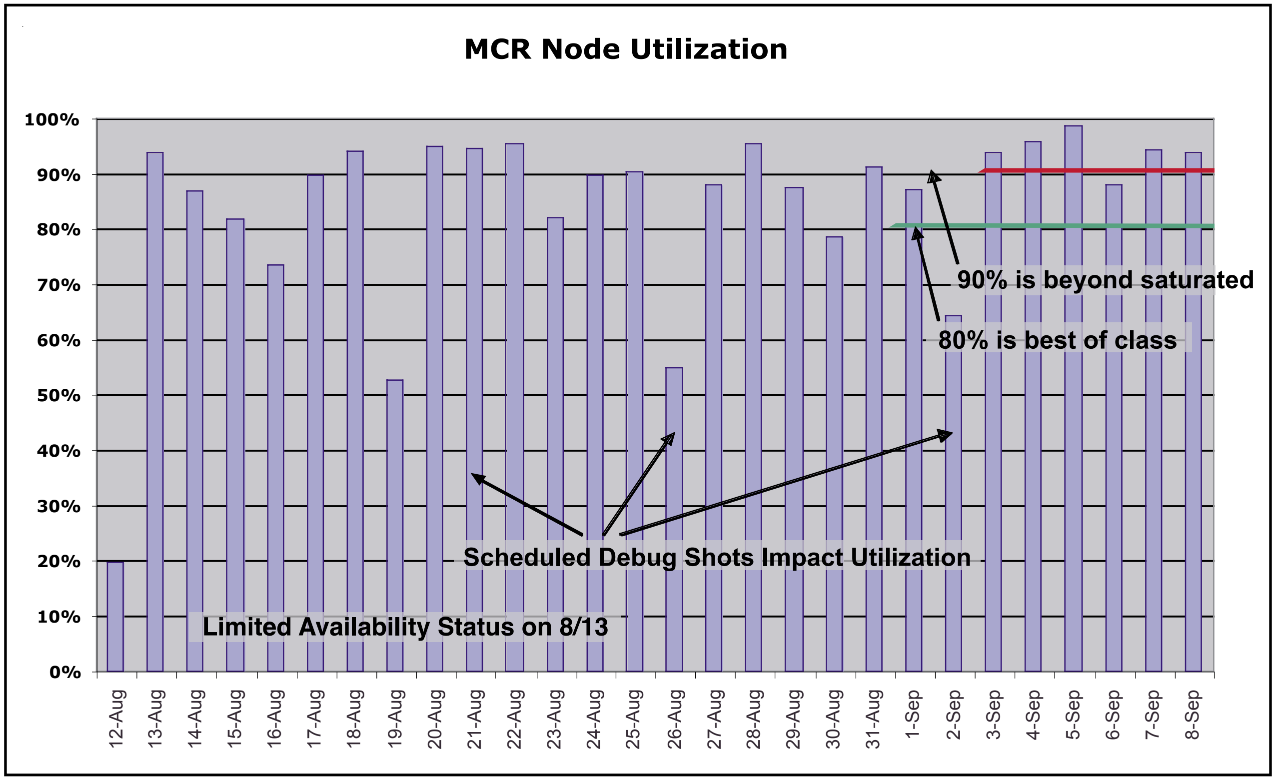


Figure 2.04-2. Complex scheduling algorithms deliver cycles appropriately and guarantee high utilization.

2.05 Desktop Support

Problem Description

All computer users, whether for HPC systems or administrative applications, start at the same place: desktop systems. We provide a competent, trained staff capable of supporting the Lab's wide variety of desktop platforms and operating systems. We recommend strategies for increased cost efficiency or improved performance to our users on an ongoing basis. We also give our users the information they need to make timely, informed decisions about their hardware and software budgets. Significant progress was made on several fronts this year all aimed at increasing desktop computing productivity for LLNL customers across all 13 Laboratory directorates.

Technical Approach/Status

An outside independent benchmarking company surveyed LLNL customers in early 2003. The benchmarking company made several high-level recommendations on ways to improve support. The recommendations were to focus on increasing support responsiveness, support availability and support expertise. An action plan was developed in response to the recommendations and improvements were targeted to areas that had particular issues. In one area, hours of the local support were expanded in direct response to customer feedback. To help solicit satisfaction feedback on an ongoing basis each support area has instituted a 30-second satisfaction survey on closed jobs. The survey is not sent to users on every job completion, but is sent periodically to do spot checks and solicit feedback to uncover issues on a more regular basis. Additionally, expected response-time matrices were created for each support area that indicate an expected problem resolution time by job priority and also define job priority levels. These matrices are intended to create a more consistent customer understanding of priority levels and response times. The final mechanism employed to improve support responsiveness is that technicians in most support areas are calling a handful of customers per week and personally communicating to them when they can expect their problem to receive attention. Our belief is that, in general, customers are more satisfied waiting a week to have their problem resolved when a reasonable cause for delay is communicated to them. Feedback from customers on the personal follow-up has been uniformly positive.

LLNL embarked on a "Total Cost of Ownership of Distributed Computing" study done by an independent benchmarking company. The preliminary report has been received and includes the following recommendations: provide stronger centralized IT Governance; strengthen the centralized help desk and services, implement more standards in the area of desktop operating systems and browsers; and use lifecycle management best practices to drive costs down. Another recommendation is to take a hard look at the number of desktops per customer, to ensure that our 2:1 ratio is required. We have not yet developed an action plan in response to these recommendations, however none of these recommendations comes as a complete surprise. We believe our overall strategy addresses most of them.

Progress in 2003

A major milestone was achieved in our effort to provide an institutional Active Directory (AD) production forest. Active Directory is Microsoft's directory service designed for distributed computing environments. AD allows organizations to centrally manage and share information on network resources and users, while acting as the central authority for network security. In addition to providing comprehensive directory services to a Windows environment, AD is designed to be a consolidation point for isolating, migrating, centrally managing, and reducing the number of directories that companies require.

A forest is the AD entity that is the root for a group of domains and organizational units. It is key to implementing computer security efficiently across LLNL desktops. The forest is up and functioning for two large Directorates (NIF and DNT), and several smaller support areas have migrated large portions of their environments to the institutional forest.

The past investments in our centralized tools have been paying unexpected dividends in the wake of the onslaught of Windows OS security issues. Radia, LLNL's Automated Software Delivery (ASD) tool now resides on more than 7000 LLNL PCs. It is routinely used to provide timely software updates. The NIF Directorate used Radia to deploy the blaster patch to 75% of 1200 targeted PCs in one recent overnight distribution. Even with this success, it is true that ASD is not designed to be an OS patching tool and requires substantial resources to use as a patching solution. Because of this, one of the efforts we have initiated is a search for a cost-effective Windows patching solution.

We have also increased our focus on technology strategy and vendor relationships. The concentrated effort in this area resulted in generation of the “Role of Technology Watch” document (Draft), visits to Apple, Microsoft and Dell to discuss roadmaps and strategies and an ongoing dialog with these vendors to fully understand and optimize the benefits available through existing maintenance contracts. We expect the technology watch role to expand to include focused programmatic requirements gathering, targeting of new technologies to particular advances in program ROI and Pilot studies of promising technologies.

This effort dovetails nicely with our efforts to formalize and publish LLNL recommended software, hardware and browser support strategies. Significant headway was made this year. A proposed strategy to formalize and publish support strategies has been created and vetted with the Desktop Advisory Group (DAG), and various support organizations at LLNL. The strategy includes four phases in the product lifecycle (target, current, containment, and no support) and includes definitions of what “support” means to various service providers at each phase in the lifecycle. The strategy is a multi-year plan for migrating to new software versions and moving off of old versions as vendors drop support for them. Figure 2.05-1 is the draft LLNL PC Operating System strategy. By providing a suggested road map to the institution, we hope to help programs proactively plan their desktop and software purchase and retirement decisions more effectively as well as give application providers a target to shoot for when developing or purchasing institutional applications.

Significance

In summary, much of the year has been spent conducting benchmarks and establishing a context from which to make ROI decisions while simultaneously pursuing improvements in infrastructure and tools (AD, ASD, patching), standards (software, hardware, and browser support strategies) and keeping an eye on new and relevant technologies. We continue to lay a strong foundation for continuing improvement in the out years. Strong central IT governance is fundamental to our ability to substantially increase ROI for our customers. This governance is vital to improving and consolidating the effort involved in effective communication, solicitation, and receipt of buy-in from more than 44 desktop support funding sources and many more stakeholders at LLNL.

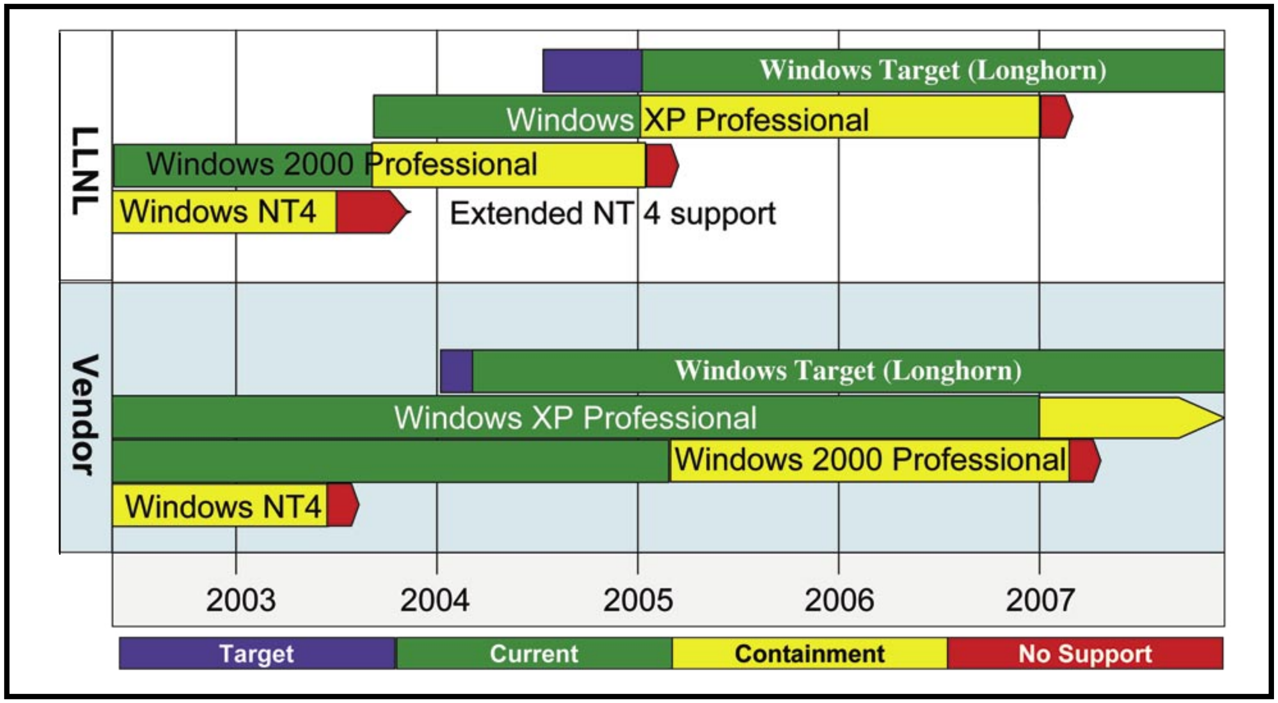


Figure 2.05-1. Windows operating system support lifecycle.

2.06 Laboratory Network Services

Problem Description

Laboratory network services enable effective, legitimate data communications among staff, collaborators, information repositories, and computing resources thus providing a basic set of user services to enhance the efficiency and productivity of all users and staff. A significant challenge is to provide these services ubiquitously and cost-effectively, with an appropriately high level of reliability, performance, and security.

Networks are the foundation of effective, secure communications with and within the Laboratory. Backbone and local network reliability and performance are improved by regular enhancements to the technology, architecture, and integrated network management and monitoring services. Firewalls and an intrusion detection/prevention system co-managed with the Computer Security Program form the first line of defense for information security. Remote-access services enable reliable, secure access for staff and collaborators to the Laboratory's computing resources. Laboratory staff benefit from a computing ecosystem that allows them to perform their technical and business activities successfully from the desktop.

Centralized Services and Accomplishments in 2003

Calendaring, email, ph (e.g., white pages), and Entrust encryption are among the centralized services provided. To protect computing resources, internal and external (i.e., to/from Internet), email is scanned for viruses, and external email is scanned for spam. Users require fewer passwords with centralized identification management for authentication now used by many business services at the Laboratory. To ease legitimate file sharing and protect information, development efforts are underway to demonstrate role management by enforcing a common set of business rules for granting access to sensitive information.

Calendar 2003 was a productive year for the staff developing and supporting the numerous centralized network services. Specific and significant accomplishments included the following.

- Migration to the new backbone was completed, Internet access was upgraded to 622Mbps, and the path between both was upgraded with redundant dual Gigabit Ethernet firewalls.
- Email was enhanced with improved metrics gathering and an anti-spam service.
- The server and clients for the centralized calendaring service were upgraded.
- The email list management service was upgraded to an enhanced new product.
- An exceptional multi-directorate effort resulted in just under 100 business applications using the centralized username and password database.
- Remote access services ("Best in Class" in the Gartner Survey) were enhanced to use One-Time Passwords (OTP).

Growth in LLNL Network Services, 1999–2003

	1999	2000	2001	2002	2003
Percent uptime for backbone	99.87%	99.77%	99.90%	99.84%	99.94%
Number of backbone connections	161	170	181	209	219
Number of network attached devices	25,720	27,551	33,653	36,854	46,420
Approx. weekly data forwarded by backbone	NA	NA	NA	NA	19.1 TB
Approx. weekly data to/from Internet	NA	NA	NA	NA	2.6TB in / 1.8TB out
Percent recevied email marked as spam	NA	NA	NA	NA	23%
Email virus scanner					
External messages scanned / viruses detected	NA	NA	10,878,555 / 31,040	16,401,745 / 119,170	18,825,825 / 829,498
Internal messages scanned / viruses detected	NA	NA	23,937,072 / 175	56,287,228 / 687	59,487,334 / 5,915
Number of central email users	9,694	9,749	10,972	12,531	12,936
Number of central POP users	8,033	8,156	8,809	8,913	10,080
Number of central calendaring users	NA	5,609	7,893	8,728	9,149
Number of Entrust (encryption) users	NA	1,684	2,750	2,940	3,378
Remote access for staff					
Number of VPN staff accounts	NA	NA	NA	NA	3,176
Number of VPN logins by staff	NA	NA	61,266	123,666	135,360
Number of IPA logins by staff	541	75,030	137,709	101,767	54,709
Number of other remote access (e.g., modem) logins by staff	NA	NA	NA	47,856	115,762
Remote access for collaborators					
Number of collaborator remote access accounts	NA	NA	NA	NA	726
Number of VPN&VPN-C remote access logins by collaborators	NA	NA	NA	182	5,737
Number of IPA remote access logins by collaborators	NA	NA	NA	6,008	20,323

Figure 2.06-1. Growth of network services has been dramatic for several years and continued in FY03.

Problem Description

The LLNL Computer Security Program (CSP) was established in February 2000. During 2003, CSP made significant progress toward enhancing the security of LLNL's computing resources. Every day, the Laboratory is bombarded with attempts to attack its internal networks and computers from the Internet. At the same time, there is an explicit DOE requirement to maintain the security of the Laboratory's infrastructure from possible insider threats. With the advances in computing technology and the growth in the number of Internet users, successfully solving these problems is difficult and ongoing, especially since the methods used to attempt unauthorized access are rapidly changing.

Technical Approach/Status

The CSP employs a variety of approaches to coordinate and manage cyber security functions, and to protect LLNL networks and computer infrastructure from intrusion and attack. Network protection is enhanced through the use of an Intrusion Detection and Response (IDR) fabric that overlays all three networks. Through risk analysis, assessments, and management, risks are identified and an optimum risk mitigation strategy is determined. At the same time, CSP employs an active vulnerability scanning function across all three networks as well as doing wireless "war driving" and modem "war dialing." Users with detected vulnerabilities are instructed to fix or mitigate the vulnerability, or have their network access disconnected. In addition, the CSP provides technical expertise to Laboratory programs to help solve and resolve, as appropriate, computer security problems or issues that are impacting their operations.

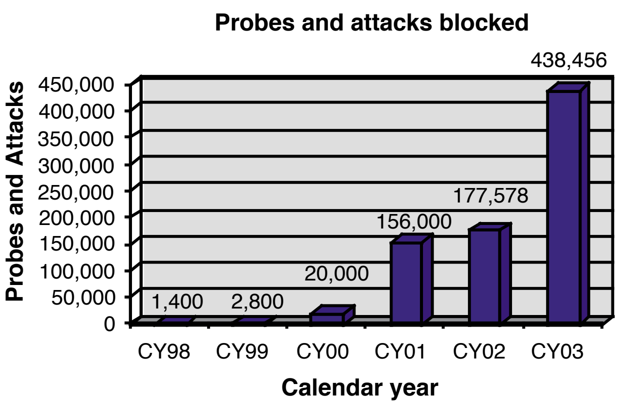
Progress in 2003

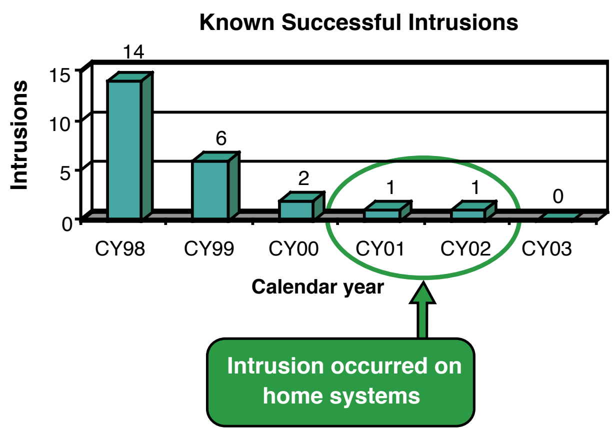
Major upgrades of the IDR systems on the Yellow network now accommodate OC-12 and Gigabit Ethernet speeds. Significant intrusions into our computer networks were prevented thanks to vulnerability scanning and identifying problems before they could be exploited.

We completed a risk assessment of voice-over-IP technologies to enable deployment on LLNL's classified network. After conducting the risk assessment, we received approval from DOE to remove the requirement for Personnel Security Assurance Program (PSAP) access authorization for File Interchange Systems (FIS) access.

Significance

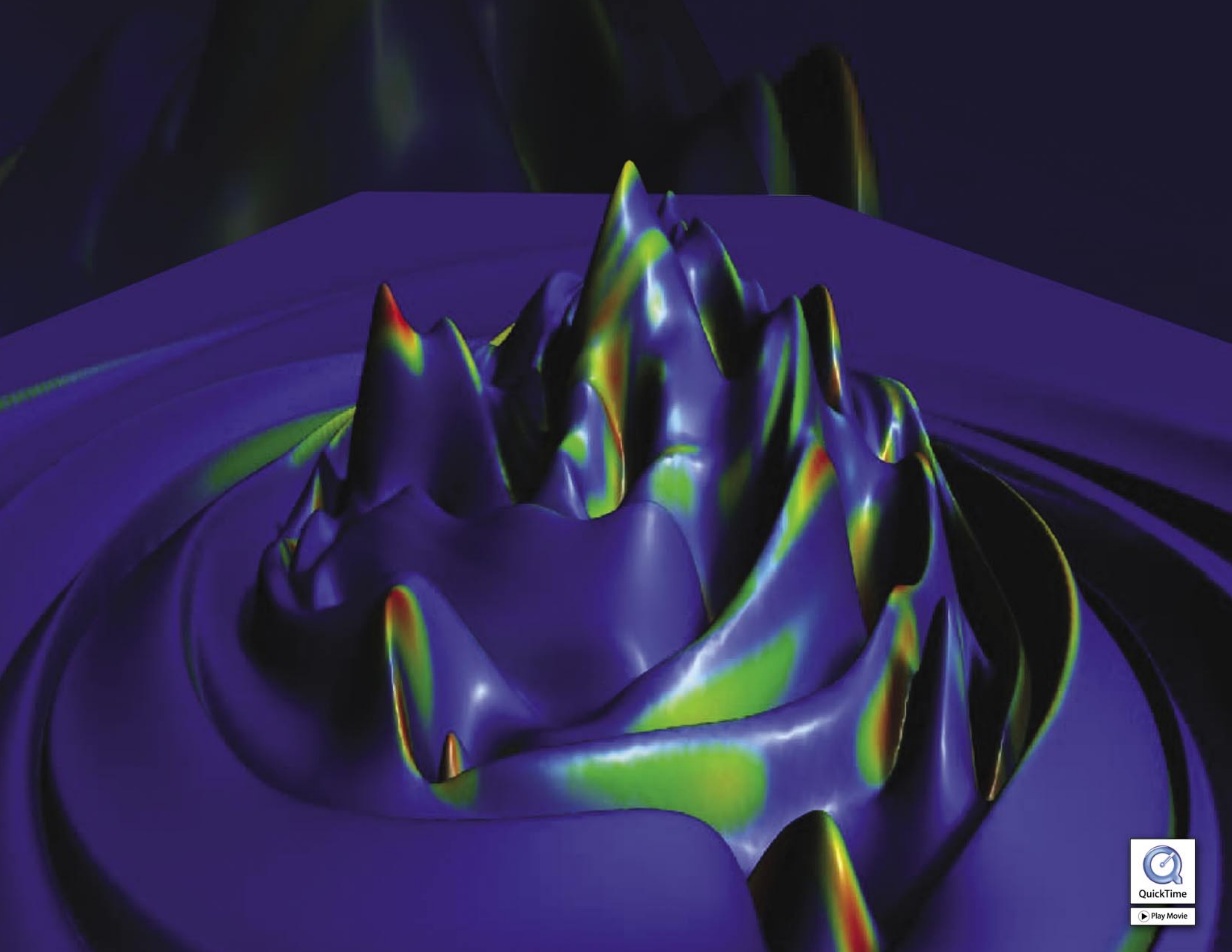
The continued progress made by the CSP is best characterized by the fact that there has not been a known intrusion into a computer on our network in more than 12 months. Through the IDR fabric and the virus and malicious code detection capabilities that are present on the networks, the Lab has not experienced a major disruption from malicious code and LLNL's users rarely have to worry about any intrusions. This "defense-in-depth" security infrastructure allows LLNL employees to concentrate on their work assignments instead of computer security.





Figures 2.07-1 (Probes) and 2.07-2.(Intrusions). Large-scale coordinated attacks were detected but were unsuccessful. No known successful intrusions occurred in 2003. Despite the high number of attempted intrusions, LLNL's network continues to be secure.

Section 3


QuickTime
Play Movie

3.00 Developing Applications Software

The Computation Directorate partners with every major Laboratory program in the development of software applications needed to meet programmatic milestones. One of the ways we do this is through the matrixing of highly-skilled computer science professionals who work side-by-side with domain experts to design, implement, and use these codes. The applications span a range of scientific and technical domains: simulation of fluid dynamics under extreme conditions; efficient, scalable, parallel multi-physics frameworks; computational modeling of materials; real-time control systems for large science experiments; and analysis of large sets of information from disparate sources.

Our Directorate employs more than 400 computer scientists and applications programmers, which makes us one of the largest applied computer science organizations in the world. We distinguish ourselves by combining outstanding discipline expertise with acquired domain knowledge, that is, an understanding of the application areas in which we work. This allows us to partner with the programs we support and helps to ensure that we provide innovative and relevant computing solutions to our customers. We strengthen these ties through the programmatic alignment of our application development efforts. This also has improved our responsiveness as a Directorate to Laboratory needs.

Figure 3.00-1 (facing page). The VIEWS visualization project simulated the dynamics of vortical flow, resulting in this height map of the vorticity (local fluid rotation) field.

The nature of the partnership between Computation and a given LLNL Program depends on the needs of that program, but the overall approach has many common features. Many programs rely on a small number of computer codes that grow and evolve over time (often decades) to meet specific needs. Computation may assign 2–20 computer scientists to each code to work with domain experts to define, implement, optimize and use these codes. The computer science expertise that must be brought to bear on each problem can differ dramatically from one problem to the next. For example, a stockpile stewardship application generally demands parallel computing and scalable algorithms expertise. These complex codes generate terabytes of data, the analysis of which requires sophisticated scientific visualization capabilities. In contrast, the real-time control system for NIF oversees every aspect of this unique experimental resource. This system has more than 60,000 control points and must manage timescales of events ranging from shot sequences (over hours) to laser pulse timing (trillionths of a second). Finally, in many applications, including bio-defense, the emphasis is on novel and domain-specific data analysis capabilities, for instance, the ability to identify unique DNA signatures via scalable pattern recognition techniques.

Underpinning all of our code development is the expectation that we will deliver high-quality, innovative, and maintainable software applications for our customers. We must develop a cost-effective approach that can be adapted to diverse programming situations, ranging from prototype research systems to mission-critical systems that cannot fail under any circumstances. Toward this end, we have spearheaded the Laboratory's new Software Quality Assurance initiative, to develop and promulgate risk-based, customizable software engineering practices. Finally, we strive to provide innovative software solutions by leveraging the Directorate's research activities (see Section 4). For example, our scalable algorithms research results are being integrated into ASCI and Office of Science applications.

The remainder of this Section highlights some of the work being done in Computation in partnership with LLNL Programs. Some of these projects are in the early stages of development, while others are reasonably mature and producing results. In each case, we try to highlight how we have applied our skills and expertise to contribute to solving the problems. We take pride in both our discipline capabilities and in the programmatic successes we enable through the timely and cost-effective use of these capabilities.

3.01 ASCI Code Development

Problem Description

The Defense and Nuclear Technologies Program relies heavily on large-scale scientific simulations in support of Scientific Stockpile Stewardship. These codes use a wide variety of physics and chemistry modeling techniques to predict and explain critical stockpile questions. We are an integral part of a team of physicists and engineers that solves problems of national importance.

Technical Approach/Status

Computation personnel assigned to each team contribute by developing, testing, optimizing, and enhancing these codes for a range of high-performance computing platforms. Overall, ASCI Code Development has had many advances this year in parallelization, algorithm development, and software quality assurance/software engineering (SQA/SE). This section will highlight two specific advances in parallelization and SQA/SE.

Progress in 2003

We have completed the initial parallelization of a Monte Carlo transport code that is showing excellent initial scaling results. The computer science advances include two options in parallelism that can work in combination. Spatial domain decomposition of the mesh and distribution of the particle load across the spatial domain allows for tracking a large number of particles in the code. Figure 3.01-1 depicts the two parallel modes. The verification on this code has shown it to compare favorably to other Monte Carlo particle transport codes on a subset of international criticality safety benchmark problems. The code is more flexible than similar Monte Carlo codes due to support for a variety of parallel run models, as well as capability to track several types of particles on problem geometries in 1, 2, and 3 dimensions. Finally, we have run hundreds of millions of particles on thousands of processors in 3D.

Several projects have developed and deployed enhanced automated testing and quality assurance software tools to handle multi-physics simulation codes and GUI testing. One testing tool, STAT quality assurance and performance monitoring system, continuously collects over 500 individual statistics from nightly regression tests and has found errors ranging from inefficient coding, to subtle errors in software and hardware platform changes. Tapestry and the Application Testing System (ATS) are regression-testing tools for parallel physics applications. Tapestry provides for MPI and OpenMP, batch and interactive testing, and several result comparison methods and reporting mechanisms. ATS provides a decentralized testing framework which allows someone who is an expert in a particular area to run a specific test in one or more ways while allowing someone who has no idea what the test does to run it without being told how. QTestViewer allows a developer working on multiple platforms to record GUI events, replay these events, save an event log, and then test the QT application systematically.

Significance

Using these enhancements and others developed by our team, we are now solving problems with higher quality and fidelity, more efficiently using valuable parallel resources, and increasing functionality and decreasing turn-around time for pre- and post-processing needs.

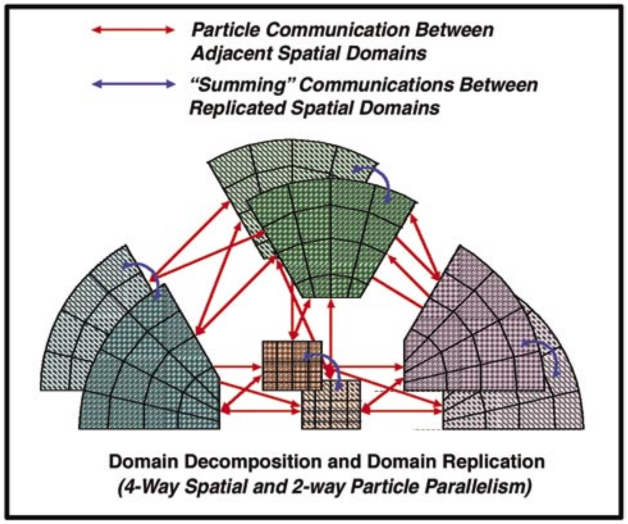


Figure 3.01-1. Domain decomposition and domain replication (4-way spatial and 2-way particle parallelism).

3.02 Verification and Validation Program in A/X Division

Problem Description

The mission of A/X Division's Verification and Validation (V&V) program is to provide scientific evaluation of the division's ASCI codes. A key aspect of the work consists of the design and implementation of analytic and semi-analytic tests that cover a broad range of physics. These tests are used to verify the physics algorithms implemented in the codes.

The results from at least one ASCI code are compared with results from legacy code, and with data from the National Ignition Facility, or other experiments that are selected on the basis of excellent data and are representative of key physics modules. One significant area of computer science support for this program is the development of a dynamic web site to report findings to the physics community.

Technical Approach/Status

The V&V web site is an interactive digital library that provides a means to publish internally the results from V&V studies. The web site is a modern web application utilizing Enterprise Java, commercial software, frameworks from the Apache Software Foundation, and emerging standards. At its core are an Apache web server running secure HTTP, an Oracle application server, and an Oracle database. The database serves as a repository for key parameters associated with a code run. Three categories of information are distinguished including biographical information about the run, key input parameters and figures of merit.

Progress in 2003

The physics reports, typically written in LaTeX, are converted to HTML and image files for browsing by using the tool Latex2Html. These are the only HTML pages in the system; the remainder of the content is generated by Java Server Pages (JSP) and servelets, which run as threads of execution under the control of the application server. The web site also incorporates a data browser integrated with the database. This is accomplished with an IDL session for on-line analysis of code results, and provides a graphical interface so that the user can visualize multiple data sets from database selections. In 2003, the web site became operational, with 10 reports and 350 runs registered in the database.

Significance

One of the key aspects of the web site is that a user can access results from a desktop by using any Web browser, as with Internet Explorer, for example. This eliminates a paper system and provides a means for rapidly distributing new information. The dynamic nature of the web site gives the user capabilities never before realized: the ability to interact with the report's data for both visual and analytic comparisons.

Figure 3.02-1. The W76 is one of the stockpile weapons for which capability has been demonstrated through verification and validation evaluations.

3.03 NIF Integrated Computer Control System

Problem Description

The National Ignition Facility, currently under construction at LLNL, is a stadium-sized facility containing a 192-beam, 1.8-megajoule, 500-terawatt, ultraviolet laser system together with a 10-meter diameter target chamber with room for nearly 100 experimental diagnostics. NIF is operated by the Integrated Computer Control System (ICCS), which will control more than 60,000 control points from a main control room (Figure 3.03-1). The control system is required to keep optical components precisely aligned over 1000 feet, and to finely orchestrate an automated shot sequence that takes place over hours, culminating in a billionths-of-a-second laser pulse that is kept in lock-step to a few trillionths of a second. The ICCS is constructed using a distributed object-oriented software framework that uses CORBA to communicate between languages and processors. This framework provides central services and patterns for building a layered architecture of supervisors and front-end processors.

Technical Approach/Status

The strategy used to develop ICCS calls for incremental cycles of construction and formal testing to deliver an estimated total of one million lines of code. Each incremental release allocates two to six months to implement targeted functionality consistent with overall project priorities. Releases culminate with successful formal off-line tests conducted by an independent Controls Verification and Validation (V&V) team in the ICCS Integration and Test Facility (ITF) and hardware integration labs. Tests are repeated on-line to confirm integrated operation and provide operator training in NIF. Offline tests in the ITF and in hardware integration labs, and these online tests in the NIF together identify 90% of software defects before the software is delivered to Operations. Test incidents are recorded and tracked from development to successful deployment by the verification team, with hardware and software changes approved by the appropriate change control board. Test metrics are generated by the verification team and monitored by the software quality assurance manager.

Progress in 20003

In 2003, NIF began its laser-commissioning program and has successfully operated the first four beams using the ICCS and by mid-2003 NIF had produced the highest energy 1ω, 2ω, and 3ω single laser energies in the world. All subsystems on NIF have been successfully fired for over 200 full system shots, achieving all scheduled project milestones. Approximately three-fourths of the NIF control systems software has been completed (including 250,000 lines of code in 2003) and used to commission and operate the first four beams of NIF.

Significance

Over the next several years, control system hardware commissioned on the first four beams will be replicated and installed to activate additional laser beamlines. Completing the remaining software is a large effort that involves completing shot automation software first for a bundle of eight beams and then for the remaining laser beams. During 2004, a separate testing effort is determining the performance limits of the control system and assuring the reliability needed to scale the control system to operate multiple bundles, and eventually 192 beams. The ICCS team is also structuring the higher-level server and shot automation software to readily meet the performance requirements as the laser is built out. This straightforward scaling flexibility is extremely important for the successful and reliable operation of NIF and was a key design goal when CORBA was chosen as the distribution mechanism for ICCS.

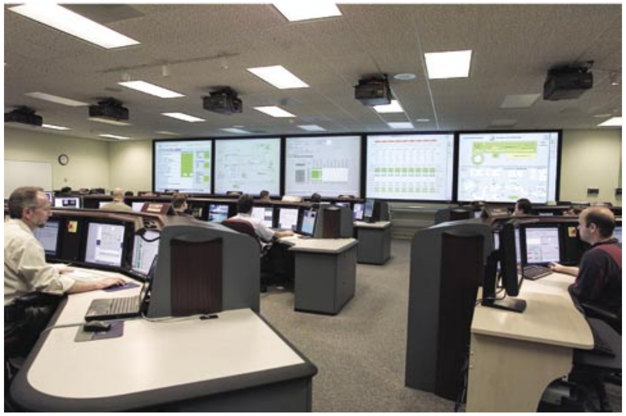


Figure 3.03-1. NIF is controlled from a main control room using the Integrated Computer Control System.

Problem description

Rapid identification of natural and genetically engineered pathogens is vital to our national security. Previously, efforts had concentrated on developing assays to detect organisms by targeting small portions of their genomes (e.g., single, well studied genes). The need arose for new tools that could rapidly identify multiple unique and specific regions of the genome, increase specificity, and reduce cost of development and time to deployment. This work is a continuation of a collaboration with BBRP biologists that counts the Human Genome Project among its successes.

Technical Approach/Status

Since August 2000, LLNL has pioneered the computational design of DNA signatures for detection of natural disease outbreaks or bioterrorism. Our DNA signature pipeline is a fully automated software system for identifying unique regions on pathogen genomes and selecting optimal sequences for development of real-time detection assays (Figure 3.04-1). These signatures are stringently tested by LLNL biologists and by external collaborators. They are validated by the Centers for Disease Control (CDC) and placed into regular use for environmental monitoring. Air samples are collected on filters, and then analyzed for the presence of pathogens. Results can be obtained in as little as 90 minutes. As of early 2004, our assays have more than 700,000 live bio-defense uses with zero false-positives.

Progress in 2003

We scaled our original DNA signature pipeline to run in parallel on a 24-CPU server. We can now process bacterial pathogens (e.g., Anthrax) in less than two hours. An invitation from the CDC to provide signature design help led to signatures for smallpox, monkeypox, and SARS—all in response to unplanned world events. Our system now automatically downloads new and updated microbial sequences from multiple public Web sites and checks all fielded signatures for potential false positive/negative results. We linked the DNA signature pipeline to a fully automated annotation system, which provides us with the additional capability of down-selecting from among many good signature candidates based on their association with genes of interest (e.g., virulence factors).

Leveraging our success with DNA signatures, in 2003 we prototyped a protein signature pipeline that can locate unique protein-based diagnostic targets on a whole-proteome scale. Our system identifies unique peptides, locates them on 3D protein models, and annotates for best-choice selection.

Significance

In 2003, our signatures entered BioWatch, which daily monitors the air in several dozen US cities. We have been asked to apply our computational system to a variety of human, animal, and plant pathogens. Success of our DNA signature design effort has established LLNL at the forefront of pathogen diagnostics. Analyses we performed in 2003 for the CDC have directly influenced the nation's policy on pathogen defense.

Figure 3.04-1. Hand-held detectors are important as the first line in Bio-Defense and pathogen identification.

3.05 Radiation Transport for Cargo Inspection

Problem Description

The Radiation Transport project develops a deterministic neutral particle transport simulation tool to model the time-dependent and steady state transport of neutrons and gamma rays through materials. This tool is being provided to DHS researchers to help in the design of active and passive radiation detectors used for detecting special nuclear material in cargo at inspection sites. It will provide them with a deterministic modeling capability that complements current Monte Carlo (MC) tools, is potentially much faster, can calculate solution sensitivities, and is in-house. Such a tool could be used in portable radiation detector configurations for DHS.

Technical Approach/Status

Neutron and gamma ray transport will be simulated in 1D, 2D, and 3D geometries, and will include delayed neutrons and gamma rays produced by fission, when appropriate. In many problems of interest to DHS, the current MC methodology is computationally intensive, requiring long run times. In such instances, an equivalent 1D or 2D deterministic capability would require much less work, and be extremely useful to DHS researchers. This is particularly true for simulations in highly diffusive media. Simulations in 3D may also be important. An added benefit is the ease with which solution and/or detector sensitivities with respect to design parameters can be calculated when using deterministic methods.

Progress in 2003

During 2003, we added the capability of solving the neutron kinetics equations to the transport code Ardra. Figure 3.05-1 shows a cross-section of a simple model of a cargo container containing a highly enriched uranium target (purple) shielded by simulated cargo (plywood, in green), a localized source (on the right), and a detector (black). Figure 3.05-2 is a snapshot of a time-dependent simulation showing a neutron pulse in the scalar flux of the 14MeV neutron energy group at 46μs as the pulse is traveling through the target.

Significance

Modeling is a significant part of the design process for active and passive radiation detectors. Providing a tool such as the one described above will allow for faster and more effective detector designs. The ability to track the delayed neutrons that result from fission is crucial to accurately model the delayed detector response seen in an actual experiment.

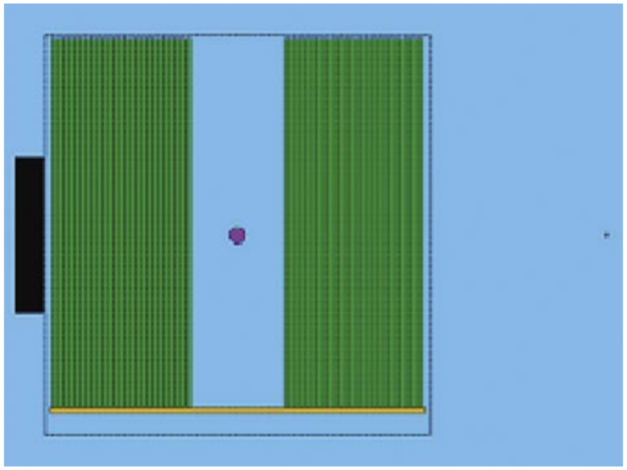


Figure 3.05- 1. A typical cargo container with a hidden target containing high-Z material.

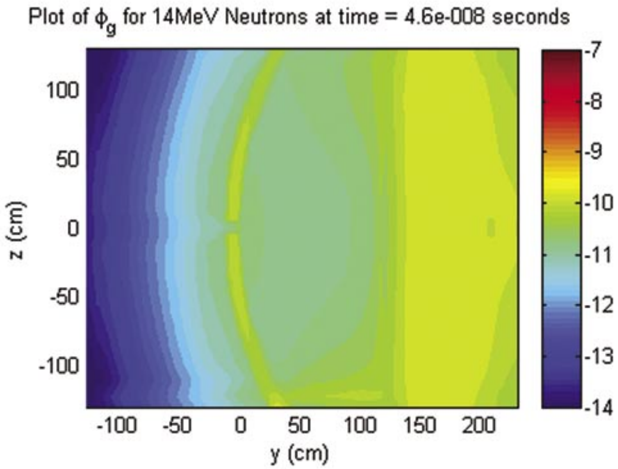


Figure 3.05-2. Scalar flux can be mapped as a pulse travels through the target.

Problem Description

The National Atmospheric Release Advisory Center (NARAC) provides to emergency managers the tools and services that map the probable spread of hazardous material (nuclear, radiological, chemical, biological, or natural emissions) accidentally or intentionally released into the atmosphere. Customer growth over the past few years has been significant—NARAC now supports many federal agencies, including DOE and DHS, as well as a growing number of state and local government agencies. To facilitate this increased external customer access, new systems and tools were developed, including the NARAC Enterprise System (NES), the NARAC iClient, and the NARAC Web. The NARAC software development team is composed primarily of Computation personnel.

Technical Approach/Status

NARAC is a distributed system, providing modeling and geographical information tools that run on an end-user's computer system, as well as real-time access to global meteorological and geographical databases and advanced three-dimensional model predictions from the NARAC Central System (NCS) at LLNL. The NCS is an object-oriented system, written in C++ and Java, using CORBA for communications and an OODBMS for object persistence. The iClient is written in Java and communicates with NES using SOAP. The NARAC Web is dynamic HTML that uses HTTPS to communicate with NES. The NES is written in Java, is built on J2EE, and uses a JDBC-compliant database; NES communicates with NCS using CORBA.

Progress in 2003

The NES and Web were designed, developed, and deployed within one year. Security, flexibility, ease of use, and future expansibility were the main design goals. User authentication, encrypted data communications, and fine-grained security were essential. For example, fine-grained security ensures that each user has access only to the capabilities and information for which he or she is authorized. By the end of 2003, there were more than 500 NARAC Web users at all levels of U.S. government.

Significance

The full NARAC System was used to support major exercises, alerts, and potential emergencies. For example, the Web and iClient were used extensively during TOPOFF2, the largest national emergency preparedness exercise since the terrorist attacks of September 11th. Set in Seattle, this exercise simulated a "dirty bomb" and involved emergency personnel from the city, county, and state governments, as well as 19 federal agencies, including DOE and DHS. Virtually all participants used NARAC predictions during this exercise. The Seattle Hazmat team and Incident Commander on scene used wireless communication and laptop-based NARAC iClient to submit and access NARAC predictions. Plume predictions were distributed using the NARAC Web to Seattle Fire and EOC and other county, state and federal agencies in real-time. Officials from the Mayor of Seattle, to the DHS Secretary, to White House personnel were briefed using NARAC predictions distributed over the Web.

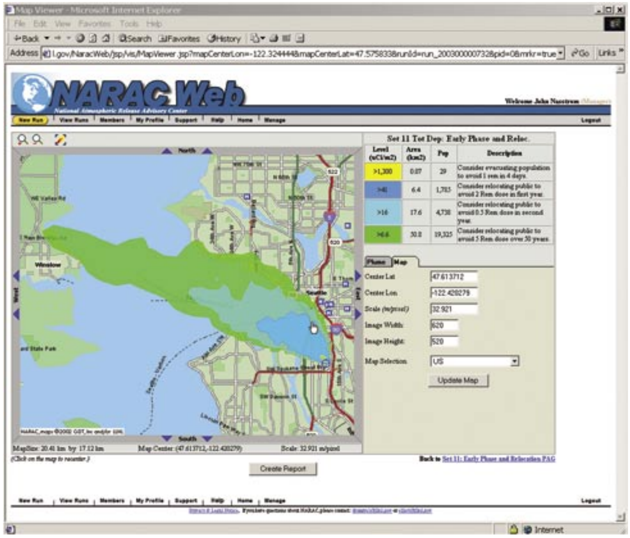


Figure 3.06-1. Sample NARAC Web plot from the TOPOFF2 emergency preparedness exercise.

3.07 Coupled Climate and Carbon Modeling

Problem Description

Contemporary climate models use prescribed carbon dioxide (CO_2) concentrations to predict the resulting climate. To properly assess the impacts of fossil fuel burning, however, one must instead base the climate computation on anthropogenic (human-induced) emissions of CO_2. Computation and Atmospheric Science Division (ASD) personnel have collaborated on the development of an integrated climate and carbon-cycle model that predicts the fate and climatic effects of fossil fuel-derived CO_2 and are applying it to analyze global warming and other related effects through the 21st century and beyond. This is the most comprehensive, and first American, fully coupled climate–carbon simulation system.

Technical Approach/Status

In this collaboration, Computation personnel have lead responsibility for the enabling technology, and ASD personnel for the scientific study. Computation members enhance and couple together the relevant component codes, parallelizing where necessary to create a scalable model that can execute on a multitude of high-performance architectures. In particular, they re-partition the land points in the terrestrial biosphere model to balance the computational load. This requires the institution of high-speed transposes to connect the terrestrial biosphere and atmospheric model domain decompositions. Researchers take advantage of both distributed and shared memory parallelism where possible.

Progress in 2003

With the model integration largely completed previously, 2003 was devoted to scientific simulation. Two main studies were completed, each involving several multi-century simulations, and each simulation taking roughly 50 days of active wall clock time. In the first study, project members analyzed the effects of CO_2 fertilization on the atmospheric concentration of CO_2. Depending on the extent to which CO_2 fertilization saturated with increasing CO_2 levels, this ranged from a doubling to a tripling of atmospheric CO_2 levels over the course of the 21st century (Figure 3.07-1).

In the second study, researchers varied the sensitivity of the climate to the radiative forcing of CO_2 and saw surface temperature increases of 3° to 8°K (Figure 3.07-2). While the project team believes these runs bracketed the degree of anthropogenic global warming, the variations underscored the importance of accurate, comprehensive modeling.

Significance

The experience and expertise of Computation personnel in high-performance computation, parallel code design, and modification and integration of large scientific programs has been instrumental in the success of this highly relevant endeavor. Having the first integrated climate and carbon modeling capability is an important step toward fulfilling DOE's mission to reduce uncertainties arising from climate–carbon feedbacks so that we can better address scientific and policy-related questions involving the climatic effects of burning fossil fuels.

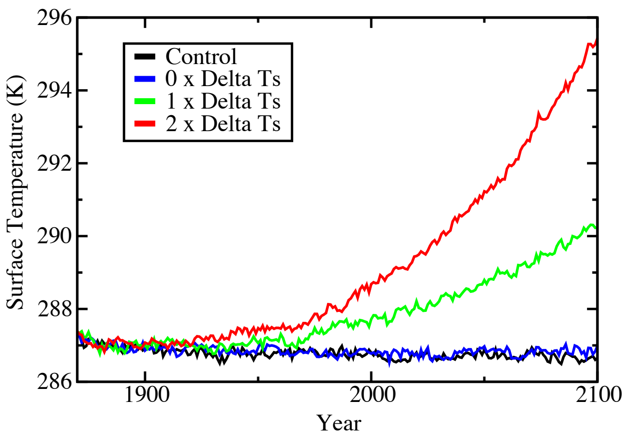


Figure 3.07-1. Simulated atmospheric CO_2 from 1870 through 2100, as a function of CO_2 fertilization (control run in black, standard case in green, saturated case in red). Saturated fertilization gives rise to a larger increase in CO_2.

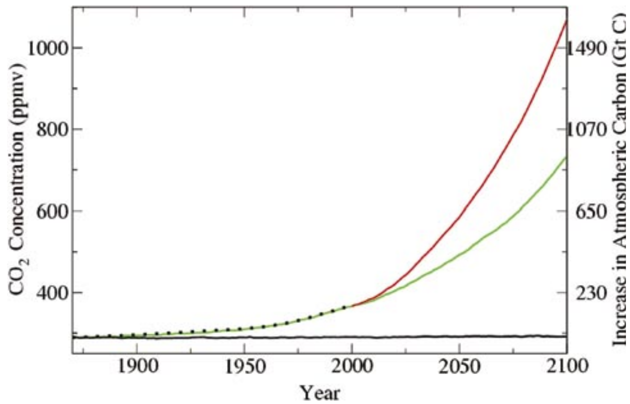


Figure 3.07-2. Evolution of globally averaged surface temperature, as a function of climate sensitivity to radiative forcing (control run in black, standard case in green, highly sensitive case in red). The expected increase in average surface temperature is anywhere between 3° and 8° K.

Problem Description

The properties of condensed matter at high pressure are both difficult to measure, and difficult to derive from what is known at atmospheric pressure. In particular, microscopic structural properties, (e.g., the type of chemical bonds present) and physical properties (e.g., the melting temperature) can be drastically modified when a substance is subjected to pressures reaching millions of atmospheres (Megabars). An experimental determination of these properties is often complex and expensive.

Technical Approach/Status

Recent progress in the technology of First-Principles simulations provides a new avenue for the exploration of properties of condensed matter in extreme conditions. First-Principles simulations are based on fundamental properties of matter derived from quantum mechanics, and do not rely on any empirical or adjusted parameters, thus providing a genuine theoretical prediction tool. The GP First-Principles simulation code developed in Computation has been used over the past years at LLNL to study the properties of fluids at high pressure.

Progress in 2003

In 2003, an important new capability was added to this simulation method by combining it with the so-called "two-phase" simulation approach. For the first time, it was possible to simulate accurately the solid–liquid interface of a molecular substance at high pressure and high temperature with a First-Principles approach. Using this new method, LLNL scientists in the Computation Directorate and in the Physics and Advanced Technologies (PAT) Directorate were able to predict the melting properties of lithium hydride up to a pressure of 200 GPa (2 Mbar). Results were published in *Physical Review Letters*.

Significance

This new type of simulation reaffirms and extends the role that First-Principles simulations will play in exploring the properties of condensed matter in extreme conditions. Being based on First Principles, this method is applicable to any other substance as well. The current parallel implementation, coupled with the powerful new high-end computing platforms available at the Laboratory, will further strengthen the position of LLNL at the forefront of high-pressure simulation research and Equation of State (EOS) calculations.

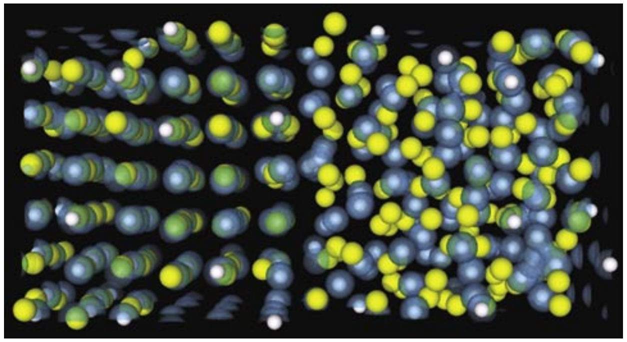


Figure 3.08-1. Two-phase First-Principles simulation of lithium hydride showing the coexistence of a solid phase (left) with a liquid phase (right). The simulation cell contains 432 atoms.

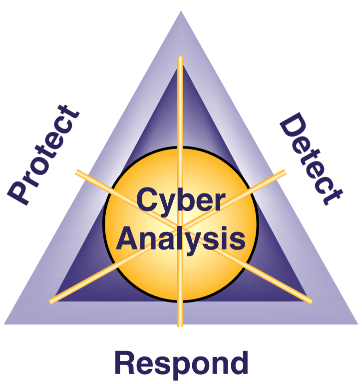
3.09 Computer Incident Advisory Capability

Problem Description

The Laboratory's Computer Incident Advisory Capability (CIAC) is a national DOE/NNSA cyber analysis center. CIAC notifies the Complex of vulnerabilities being exploited, recommends countermeasures, provides an overview of the current attack profile, and assists sites. CIAC focuses on specific threats and malicious activity targeting DOE/NNSA, and is developing a predictive analysis capability. This Advanced Warning and Response System (AWARE) assists DOE in preventing incidents rather than simply reacting to them after the fact.

Technical Approach/Status

AWARE strives to provide "Information-to-Insight-to-Action." It integrates technologies such as data mining, pattern recognition, statistical trending and traffic analysis, attacker profiling, and advanced visualization techniques. The system searches for trends and indications of possible attacks by analyzing potential threats, sites' sensor data, and data supplied by outside sources. AWARE will be able to pinpoint the right data at the right time, integrate sensor networks and databases into 24/7 operations, and allow actionable correlations to be made in near real-time.



Progress in 2003

Network traffic is analyzed within 15 minutes of receipt, by streamlined processing of 2 GB/day of data collected by sensors deployed at several DOE/NNSA sites. Statistics are calculated in parallel threads, synchronized, and loaded into an Oracle table for reporting and visualization.

The AWARE portal (alpha release) disseminates results of the hourly analysis. The portal provides authenticated access to site-specific information and non-site–attributed Complex-wide information to DOE/NNSA security personnel.

Clustering techniques profile IP addresses for normal behavior, so that current activity can be compared and aberrant behavior detected. This reveals key variants that distinguish types of behaviors that leave cyber fingerprints.

Significance

As cyber attacks continue to rise in sophistication and virulence, cyber indications and warning systems are more critical than ever. Vulnerability exploitation time is decreasing dramatically, while the cost of repairing the damage has doubled each year from 2001 through 2003. The sooner new exploits or vulnerabilities are detected, the earlier DOE/NNSA can take action against them.

AWARE provides DOE/NNSA with an effective cyber indications and warning capability. It proactively protects Departmental assets from compromise, thus averting potential incidents and their ensuing impacts on productivity throughout the restoration and recovery periods. Through AWARE, CIAC advances the current "Protect-Detect-Respond" security defense strategy to one facilitated by anticipating adversary attacks, assessing intrusions, and assisting the sites in adapting their security architectures to proactively counter the attack. CIAC's work in this area allows DOE/NNSA to advance technology beyond intrusion detection to intrusion forecasting.

Figure 3.09-1. CIAC provides 24/7 incident response, intrusion analysis, vulnerability response, and counterintelligence support.

Problem Definition

The Information Operations and Assurance Center (IOAC) is a capability with a focus on information operations, knowledge management, and analysis tool development. IOAC's information management and analysis tool development is a noteworthy strength that we are aggressively applying to several major homeland security mission areas including infrastructure protection and bio-defense analysis—the current program is handling and correlating diverse information feeds from multiple sources and creating a graph that interrelates the information from these sources.

Technical Approach/Status

The Information Fusion and Analysis area has been making progress in building large-scale information fusion systems. We have developed semantic graph-based technology to perform real-time threat analysis and warning. The semantic graph facilitates this by extracting important relationships and correlations from a plethora of diverse data sources. The Network Analysis Tools area is dedicated to development of science and technology solutions in support of the analysis of networks. This focus area develops tools to automatically build a network model, graphically visualize it, and analyze it for attributes and patterns of interest including identification of vulnerabilities.

Progress in 2003

Enhanced network analysis tool capabilities fused disparate data with network-related information and added a GIS front end to the mapping system. We initiated a collaborative program to develop the Information Fusion and Analysis technology for the DHS. Additionally, we completed the design and partial development of a prototype system for DHS to make inferences from diverse data sources and data types.

Significance

The 2003 deliverables provided new capabilities for the program sponsors. The information fusion engine will enable the DHS to automate the process of "connecting the dots" across numerous diverse data sources and data types. The system security infrastructure supports U. S. law and privacy requirements.

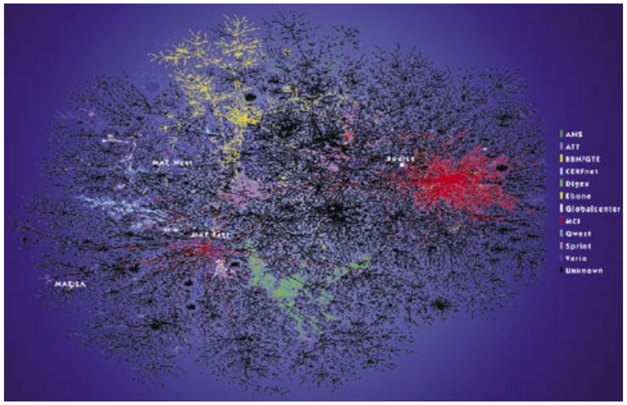


Figure 3.10-1. Network Analysis Tools automatically build, graphically visualize, and analyze a network model for attributes and patterns.

3.11 Software Quality Assurance

Problem Description

Outstanding software quality helps projects achieve their performance, cost, and schedule parameters. Each LLNL directorate has evolved its own Software Quality Assurance (SQA) approach, with varying degrees of documentation for those practices. In the absence of a consistently implemented Labwide SQA policy, we run the risk that external standards could be imposed on LLNL. Similarly, the Defense Nuclear Facility Safety Board has raised concerns that inconsistent practices might reduce confidence in the safety systems and safety analysis performed across the DOE Complex. These factors have prompted DOE and NNSA to raise the formality level of SQA for software used in critical applications.

Technical Approach/ Status

Led by Computation, software engineers from each directorate are collaborating to develop an institutional SQA Implementation Plan (SQA IP). LLNL will employ a multi-tiered, risk-based, tailorable approach to defining the SQA practices that apply to each project. Each project will grade its software components based upon the risk associated with their use. The graded approach will balance the hazards associated with the work the software is performing with the degree of rigor in the software quality practices. The SQA IP team will define various tiers of risk, and designate the software quality practices appropriate for each.

The IP will define these risk categories and associated practices based upon a set of over-riding SQA principles. The SQA IP is a blended model of external standards and current LLNL SQA best practices, to ensure added value to existing and future projects. In addition to clarifying appropriate SQA practices, this institutional approach satisfies the expectations of the external oversight organizations that increasingly scrutinize our software quality.

Progress in 2003

As the institutional approach is being developed, client programs have continued to enhance their own software engineering practices. ASCI sponsored a two-day, Tri-Lab workshop to share information about software engineering best practices, so that agencies might learn from one another. NIF continued to focus on its internal software engineering practices, and retained Raytheon to establish a unit-testing organization. NARAC developed an automated test suite to streamline its process. ARGUS increased its testing coverage with integration testing during the development phase, conducted by its own SQA team. This early testing protocol provided an earlier defect detection.

The Directorate also developed a series of seminars and training sessions to focus on software engineering practices. The seminar series for fiscal year 2004 includes usability, secure coding, testing, configuration management, risk management, and software quality updates. In an effort to better understand our customers' needs, Computation has begun to inventory current practices and tools used by Lab personnel in developing application software.

In 2003, six Computation employees became certified software quality engineers through the American Society for Quality. This was made possible by Computation's sponsorship of the software quality training and certification testing onsite. The Directorate is also hiring trained software quality experts to meet increased program demands. We are working closely with DOE in developing their SQA practices by stationing a LLNL employee at Headquarters.

Significance

Computation's SQA continues to enhance and enable improved software engineering practices Labwide. With Computation leading the effort, the current LLNL institutional SQA policy was approved on October 3, 2003. The primary policy objective was and is to deliver the best software products to our customers, and thereby to create confidence in the performance, cost, and delivery schedules. By developing a consistent software quality framework, we ensure that we will meet customer needs and comply with their software quality requirements.

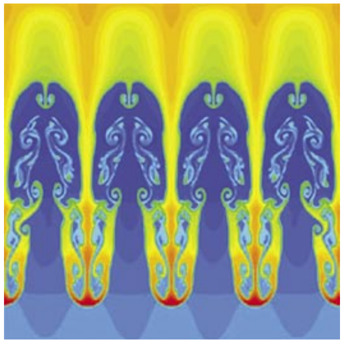
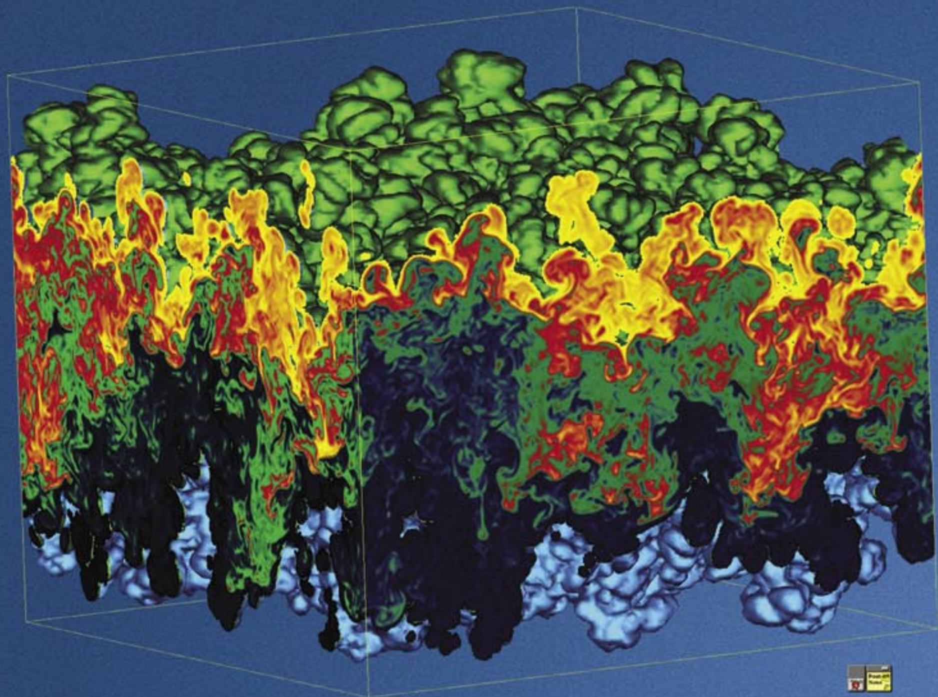
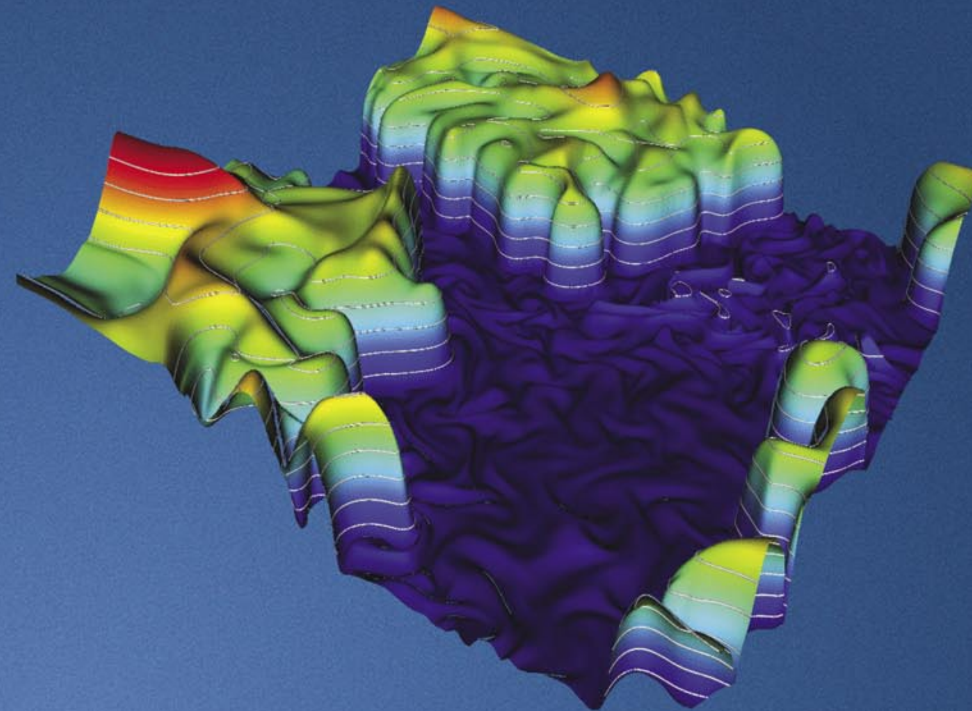
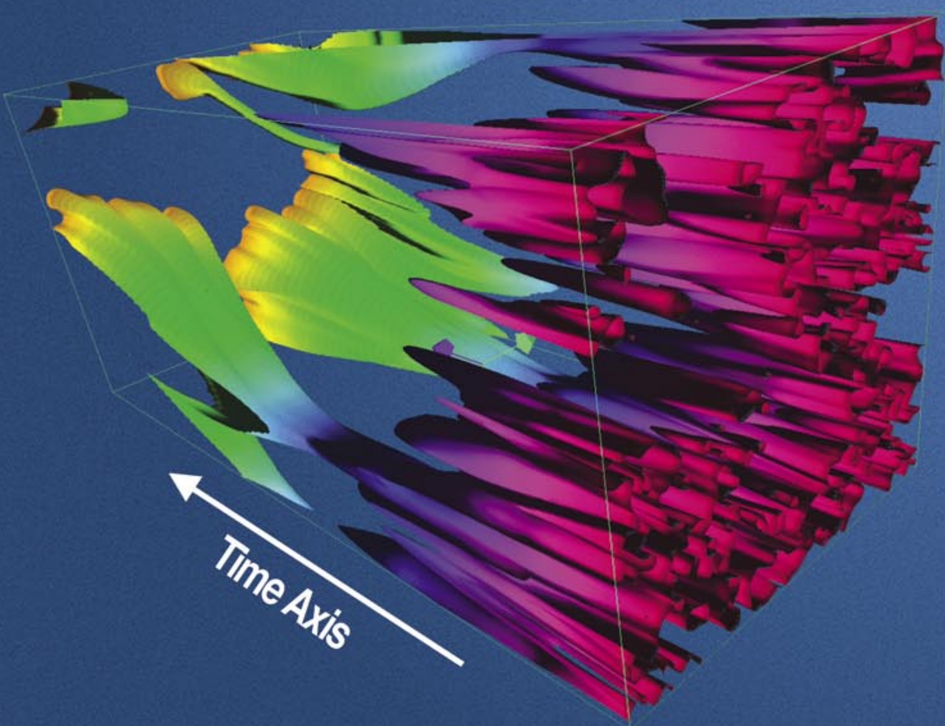


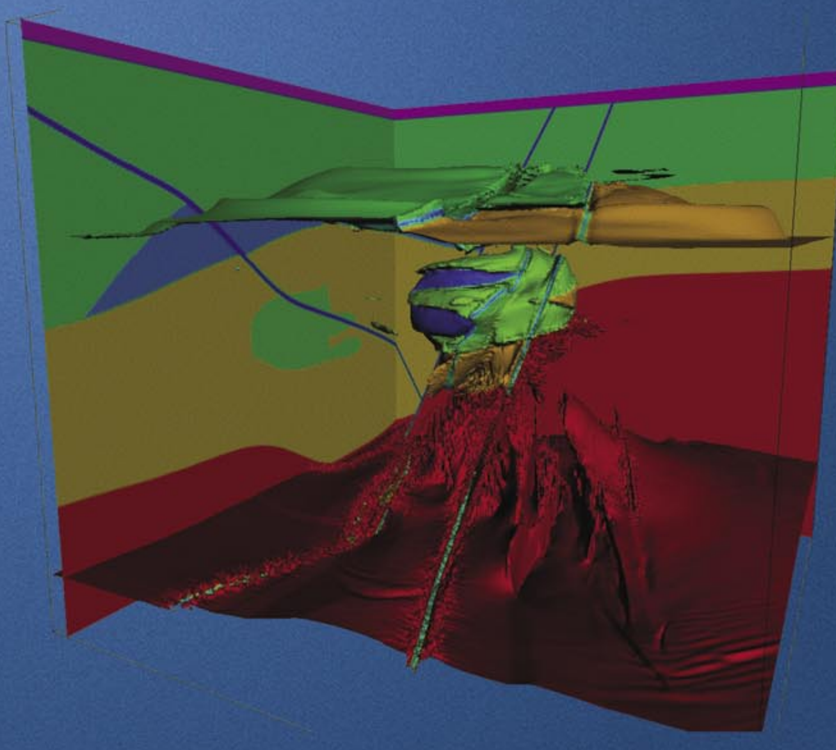
Figure 3.11-1. ASCI at LLNL relies on software quality assurance to produce accurate simulation results such as this hydrodynamic instability calculation.

Section 4






Time Axis



4.00 Engaging in Computing Research and Advanced Development

The Computation Directorate's responsibility for enabling science goes beyond just helping solve today's problems; we must be prepared to solve tomorrow's known and anticipated problems through research and advanced development. We conduct collaborative scientific investigations that require the power of high performance computers and the efficiency of modern computational methods. Our research and development activities are applications-driven, and focused on LLNL programmatic objectives that require advanced computational technologies. This section highlights progress in a selection of projects from our portfolio of research and advanced development.

Much Computation Directorate Research and Development is characterized by the aggressive use of massively parallel computing to solve problems of national interest. The problems typically involve large-scale simulations of complex systems. Classic problems of scientific interest are usually based on Partial Differential Equations (PDEs) or continuum descriptions, but there is now increasing interest in systems that are better treated using discrete simulation. Additionally, many complex problems involve both multi-physics and multi-scale issues that demand a rethinking of not only the original formulation, but also the computer science design for codes that will efficiently use available resources. Figure 4.00-1 illustrates one such example. The first four reports in this Section describe recent progress addressing these types of issues.

Figure 4.00-1 (facing page). These simulations were created by the streaming process described in Section 4.10.

Performance continues to be another critical area of research. We are concerned with both our ability to exploit available computing platforms effectively and our ability to optimize the time required to write new codes, evolve codes over time to meet new mission demands, and move codes to new platforms as they become available. The goal of our research is to simplify the construction of re-useable software libraries and to improve the performance of existing scientific software. Current approaches include object-oriented design, scripting approaches for scientific simulations, and component technologies. The next three reports exemplify our work in this direction.

"Data Science" is our umbrella term for describing research over a wide range of topics related to understanding and effectively using large-scale data. The Laboratory is challenged to extract insight from massive amounts of data arising from numerous sources, including: scientific simulations, experimental devices, sophisticated sensor systems, specialized data bases, and public Web pages. Our goal is to enable scientists to concentrate on science by minimizing the burden of physically managing data and computer resources. This goal drives research efforts in terascale visualization, large-scale pattern recognition, clustering and classification algorithms, genetic and evolutionary algorithms, video and image analysis, feature extraction, query infrastructures and data access and integration in dynamic environments. The final three contributions in this Section illustrate our progress in Data Science.

To accomplish our research and advanced development objectives, the Computation Directorate partners extensively with academia and industry. We benefit significantly by engaging these groups in working with us to address our needs and objectives. Mechanisms for these partnerships take a variety of different forms. In 2003, Computation hosted 63 summer students, eight sabbatical visits, and 162 other visits by a total of 125 different visitors. We collaborate with universities and other national laboratories through jointly funded federal research proposals. We also fund some research subcontracts in direct support of on-going research projects within the Directorate. These interactions have contributed to the R&D results reported here and to the entire set of research activities in the Directorate.

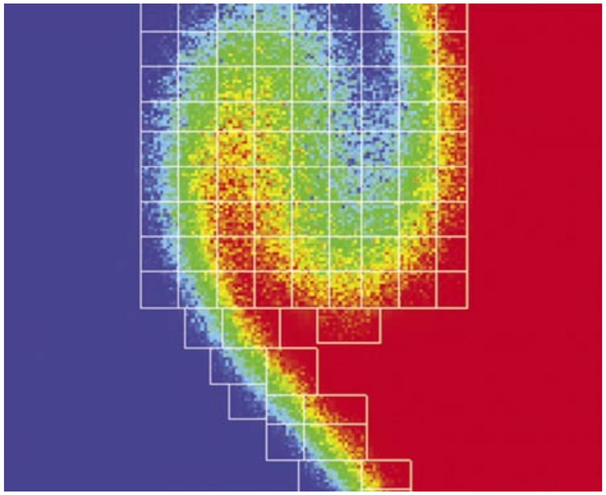


Figure 4.00-1. Hybrid multi-scale calculation using the SAMRAI (Adaptive Mesh Refinement) system. The fluid shear instability is resolved by employing continuum dynamics where applicable, and particle methods at the finest scale.

4.01 Scalable Linear Solvers

Problem Description

The B-Division code, ALE3D, uses implicit hydrodynamics techniques to generate structural dynamics simulations as part of the nation's Stockpile Stewardship Program. In 2003, the ALE3D team was interested in solving a very high-resolution spherical shell problem (Figure 4.01-1). The shell is composed of three layers and two different materials: steel for the inside and outside layers, and Lucite for the middle layer. One challenging aspect of this problem is the parallel solution of the linear systems that arise. In particular, the jumps in coefficients across material boundaries, the poor aspect ratios in the meshing of the steel layer, and the presence of so-called rigid body modes, create difficulties for the linear system solver.

Technical Approach/Status

The Scalable Linear Solvers (SLS) project is developing fast parallel multigrid algorithms and software for solving large, sparse linear systems of equations. The development of new linear solvers can often dramatically improve the capabilities of codes such as ALE3D, giving them the ability to simulate problems much faster and at much higher resolutions than ever before. Researchers are also investigating other numerical methods areas, including nonlinear solvers and sensitivity analysis. This research is contributing to programs at LLNL and elsewhere in the DOE, but for brevity, we have focused on one specific highlight in this report.

Progress in 2003

During 2003, two solver advances helped the ALE3D team accomplish record-breaking simulations. The first was the development of an automatic scheme for choosing smoothing parameters in algebraic multigrid. Using multigrid convergence theory, we derived formulas for pseudo-optimal smoothing parameters. These formulas require estimates for the largest eigenvalue of the smoother, which are computed by employing a known relationship between the conjugate gradient method and the Lanczos eigenvalue method. The second advance was the development of a new solver based on the smoothed-aggregation method. This latter solver exploits the availability of the rigid body modes and requires less memory and computations per iteration than the algebraic multigrid solver.

Significance

These solver advances have so far enabled solution of the spherical shell problem in Figure 4.01-1 on meshes with more than half a billion (610 million) degrees of freedom on 4032 processors of ASCI White in less than half an hour. This is 100 times larger, and run on 10 times the number of processors, than the simulations of only three years ago. It is also the largest implicit hydrodynamics calculation done to date in the ALE3D code.

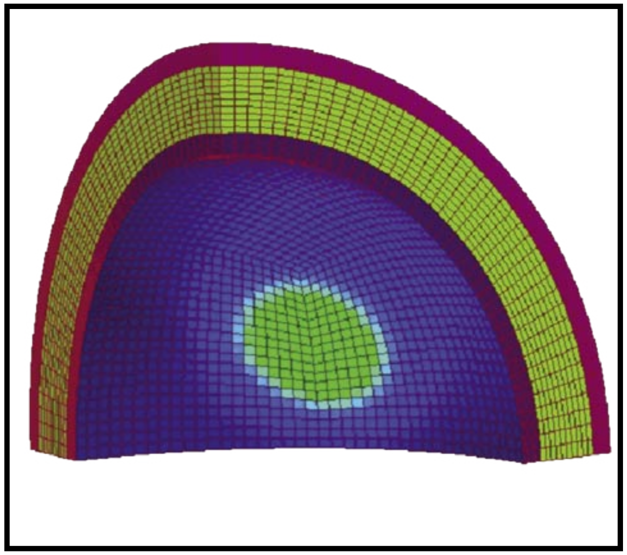


Figure 4.01-1. This simulation of a spherical shell implicit hydro problem used 4 million elements and 12.3 million unknowns. Advances in 2003 allow up to 610 million unknowns.

Problem Description

Many applications require the development and analysis of numerical methods for high-fidelity PDE-based simulations involving complex, possibly moving geometry. New techniques that address this application regime must be flexible, computationally efficient, and highly accurate.

Technical Approach/Status

We are developing advanced grid generation and discretization techniques that deliver highly flexible geometry representations while retaining the accuracy and efficiency advantages of simple single-block structured grids. In particular, we develop methods for 1) adaptive overlapping grids which consist of structured, logically rectangular curvilinear body-fitted component grids that overlap where they meet; 2) mixed-element or hybrid grids where the grid consists primarily of large regions of logically rectangular structured mesh; and 3) embedded boundary grids that represent complex geometry by cutting a structured component grid with a complex surface. In each case, we use highly efficient discretization techniques in structured grid regions and develop new techniques to handle grid overlap, mixed element meshes, or general polyhedral cells.

Progress in 2003

For overlapping grids, we developed new algorithms and grid generation capabilities for solving the incompressible Navier–Stokes equations for airflows around stadiums and cityscapes (Figure 4.02-1). We also performed basic research that led to a new understanding of stability properties of incompressible Navier–Stokes solutions. To develop overset grid techniques for free boundary dynamics, we analyzed non-Newtonian viscous fingering and developed novel time-stepping algorithms for coupling an elastic boundary to an incompressible fluid. With the Chemistry & Materials Science Directorate, we developed a 3D multi-block solver for rapid model prototyping of biochemical reactions using Overture software.

For embedded boundary methods, we developed a second-order accurate method for the second-order wave equation in general 2D domains. For the Neumann problem, we analyzed stability using a normal mode technique; for the Dirichlet problem, we developed a discrete boundary stencil that avoids the small-cell time-step restriction and devised the smooth startup procedure necessary to obtain second-order accurate gradients. We combined these techniques to solve Maxwell's equations written as a system of second-order wave equations for general 2D domains.

For hybrid grids, we designed and implemented 2D/3D unstructured , mixed element, second-order finite volume mesh operators, integrated them into Overture, and verified their accuracy. Using these operators, we completed preliminary work to solve Maxwell's equations on 2D mixed element meshes.

Significance

This work enables the solution of incompressible flow applications of dispersive modeling problems important to DHS, as well as shedding new light in basic research areas in biological computing and turbulence modeling. New hybrid and embedded boundary methods are expected to be of critical importance to maintain accuracy and efficiency as geometric domains become more complicated in applications such as accelerator modeling.

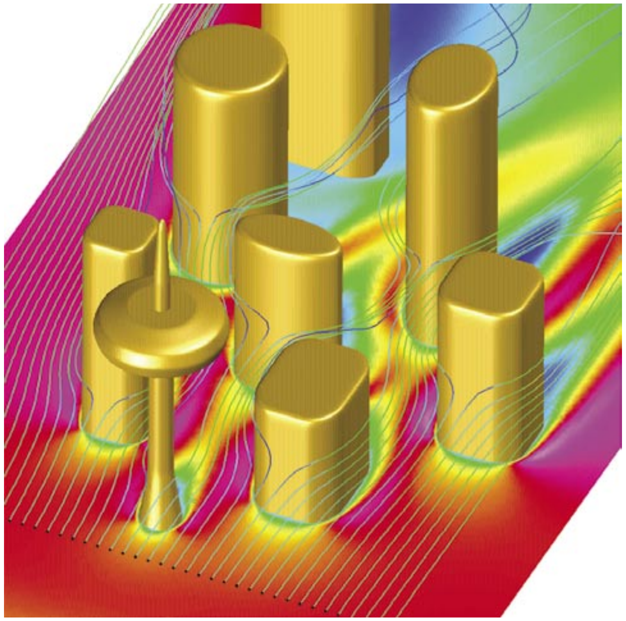


Figure 4.02-1 This flexible geometry simulation models dispersion in a cityscape domain.

4.03 SAMRAI

Problem Description

The SAMRAI project focuses on enabling structured adaptive mesh refinement (SAMR) technology for large-scale parallel applications. Simulations of science and engineering problems often require high resolution on multiple scales, both spatial and temporal. SAMR systematically focuses computational resources by dynamically adjusting resolution of the computational mesh.

Technical Approach/Status

The SAMRAI project is a research base for application, numerical software, and parallel computing issues associated with SAMR. The SAMRAI software library is employed in application development in various projects at LLNL and other institutions. SAMRAI software simplifies parallel SAMR development for computational scientists and allows software technology to be leveraged across multiple applications. Its object-oriented design allows SAMRAI capabilities to be enhanced and extended to meet unique problem requirements.

Project collaborations at LLNL involve problems in fluid dynamics, electronic structures, and adaptive ALE hydrodynamics. University collaborations involve hybrid continuum–particle methods, industrial fire simulations, and computational biology.

Progress in 2003

The ALE–AMR code couples SAMR and ALE methods using SAMRAI (Figure 4.03-1). The DNT Directorate is increasingly interested in this work. Initially, we are integrating SAMRAI capabilities into the DNT CALE code; we hope that this effort leads to adoption of SAMRAI and adaptive methods in other DNT codes. The DHS is using SAMRAI to develop high-resolution simulation models for releases of airborne contaminants in urban settings. This project involves the FEM3MP dispersion code from ASD, and complex geometry facilities of the Computation Directorate Rapsodi project.

Another project uses SAMRAI to develop adaptive methods for electronic structures calculations that will scale much better than existing state-of-the-art approaches. Our algorithm research crosscuts several projects. Recently developed combinatorial techniques significantly reduce adaptive meshing and communication costs in large-scale parallel applications. Nearly optimal scaling has been shown for canonical adaptive computations using more than 1000 processors (Figure 4.03-2).

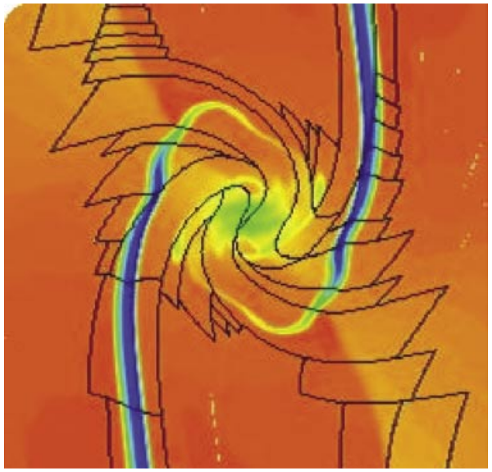


Figure 4.03-1. ALE–AMR combines deforming meshes and mesh refinement.

In 2003, we also organized an Institute for Scientific Computing Research (ISCR) workshop on multiscale methods, gathering researchers from DOE Labs, academia, and industry to discuss multiscale simulation techniques and their applicability to LLNL programs.

Significance

New and growing collaborations demonstrate that AMR is important for large-scale computational problems, and that SAMRAI is a viable software platform for programmatically-driven application development. By focusing computational resources, high-resolution simulations are achieved more efficiently than static mesh approaches that often yield inadequate resolution.

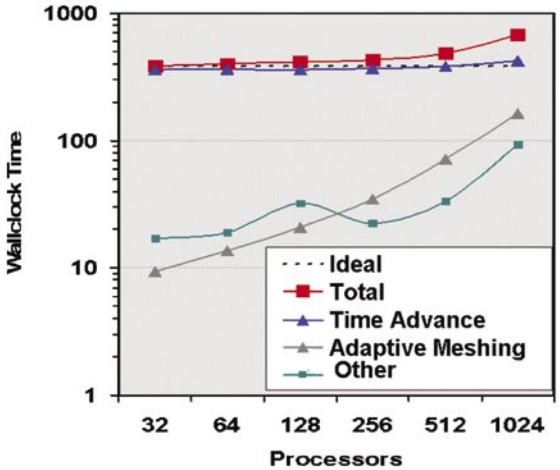


Figure 4.03-2. SAMRAI adaptive meshing operations show NlogN scaling on more than 1000 processors.

Problem Description

The AMTRAN code project works to scale deterministic (S_n) neutron transport problems up efficiently to thousands of parallel processors by means of spatial domain decomposition and load balancing for a finite-element code with block-structured adaptive mesh refinement on Cartesian grids. This project is of particular interest to DNT.

Technical Approach/Status

A directed binary spatial decomposition scheme was found that can be coupled with predetermined scripts for assigning subdomains to processors and for orchestrating the order of computational steps as these subdomains are repeatedly "swept" sequentially from multiple directions in parallel. These scripts can be represented diagrammatically as (logically) space-filling tiles composed of rectangles, each representing a computational unit of one sweep direction on a subdomain on a single sub-iteration (Figure 4.04-1). Identically shaped tiles are fitted together, like the pieces of a jigsaw puzzle, to minimize idle time as much as possible. As the degree of spatial parallelism and the number of processors increase, the tiles grow increasingly complex in a fractal-like manner.

Progress in 2003

The tiling method was discovered and implemented in 2003 by generalizing some simpler known cases of optimal scheduling and by assigning multiple subdomains to individual processors. The algorithm has been tested on cases involving from 4 subdomains in 2 dimensions to 512 subdomains in 3 dimensions, and scripts up to 4096 subdomains are ready for future parallel architectures. A simple example is shown in Figure 4.04-1 for a 4-by-4 set of domains in 2 dimensions, where efficiency is doubled from 40% to 80% by overlaying two half-tiles on one another. A plot of efficiencies of various configurations in three dimensions is reproduced as Figure 4.04-2. It shows that efficiencies approach 100% asymptotically as we successively double the number of subdomains for a given degree of parallelism. The three curves illustrate (from top to bottom) 16-way, 32-way, and 64-way spatial parallelism, respectively. The term "maximum theoretical processor usage" in the title means that we assume perfect load balance and ignore communication overhead costs.

Significance

The tiling method has led to significant improvements in efficiency and parallel scalability for the AMTRAN code. In addition, it provides for the first time a theoretical basis for projecting performance on more massively parallel computers planned for the future.

Master	Subdomain	Sub-iteration (without subdomain overloading)									
		1	2	3	4	5	6	7	8	9	10
		Sweep Directions									
1	(1,1)	0			1			2			3
2	(1,2)		0	1					2	3	
3	(1,3)		1	0					3	2	
4	(1,4)	1			0			3			2
5	(2,1)			0		2	1		3		
6	(2,2)				0	1	2	3			
7	(2,3)				1	0	3	2			
8	(2,4)			1		3	0		2		
9	(3,1)			2		0	3		1		
10	(3,2)				2	3	0	1			
11	(3,3)				3	2	1	0			
12	(3,4)			3		1	2		0		
13	(4,1)	2			3			0			1
14	(4,2)		2	3					0	1	
15	(4,3)		3	2					1	0	
16	(4,4)	3			2			1			0

Master	Subdomains	Sub-iteration (with subdomain overloading)									
		1	2	3	4	5	6	7	8	9	10
		Sweep Directions									
1	(1,1), (3,1)	0		2	1	0	3	2	1		3
2	(1,2), (3,2)		0	1	2	3	0	1	2	3	
3	(1,3), (3,3)		1	0	3	2	1	0	3	2	
4	(1,4), (3,4)	1		3	0	1	2	3	0		2
5	(2,1), (4,1)	2		0	3	2	1	0	3		1
6	(2,2), (4,2)		2	3	0	1	2	3	0	1	
7	(2,3), (4,3)		3	2	1	0	3	2	1	0	
8	(2,4), (4,4)	3		1	2	3	0	1	2		0

Figure 4.04-1. A 4-by-4 set of domains in two dimensions, where efficiency is doubled from 40% to 80% by overlaying two half-tiles on one another

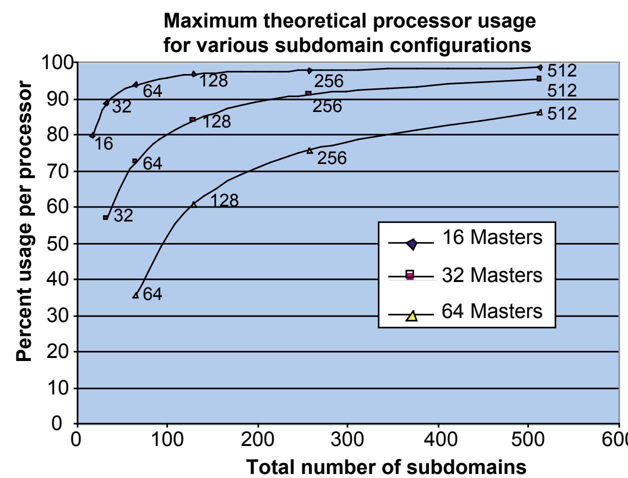


Figure 4.04-2. Theoretical efficiencies of various configurations in three dimensions.

4.05 Performance of Multi-Physics Applications

Problem Description

Supercomputer performance is often quoted in terms of the Linpack benchmark, which carries out dense linear algebra computations. By this measure, some systems can achieve 60–90% of their peak theoretical performance. ASCI applications, by contrast, typically use multiple coupled physics algorithms, each of which places unique demands on the computer. These applications are considered excellent performers when they reach 15% of peak on a single CPU for a given algorithm, or 5–8% aggregate over a time step. Understanding and explaining this dichotomy between benchmark and "real" application performance was a high priority in 2003, as several external review panels asked LLNL to study this topic.

Technical Approach/Status

Our approach to this problem was two-fold. First, we took detailed measurements of the algorithmic characteristics of their codes. This was done through a collaboration of code developers from DNT and computer scientists from the PSE/Tools Group. This group, known as APOMP (ASCI Performance Optimization and Modeling Project), was chartered to study and improve application performance. Characteristics such as CPU instruction mix, computational intensity (flops per memory op), cache hit ratios, parallel efficiency, and more were collected using a host of tools.

Second, a very simple model was developed to help determine why attaining performance near peak levels on a microprocessor-based architecture was impossible without significantly changing the algorithmic characteristics of the physics.

Progress in 2003

This model was applied first to the Power3 microprocessor architecture, the foundation of ASCI White. Peak performance on that chip requires two FMA (fused floating-point multiply-add) instructions to be issued every cycle, for a peak speed of 4 flops/cycle. This in turn requires the program to use each operand it fetches from memory three times for every result it stores.

By carefully measuring the instruction mix and computational intensity of the ASCI applications, we were able to show an upper bound on single-CPU performance of 23% of peak for data culled from several actual ASCI milepost calculations. This model did not take into account cache misses, integer instructions, parallel efficiency, or a host of other potential factors that would further close the gap between predicted and actual peak performance.

Significance

This simple model was presented to several external review committees (the JASONs, the ASCI Burn Code Review Committee, and the National Research Council), all of whom were keenly interested in application performance. By showing that we are reaching performance close to the "empirically derived upper bound" of the chip (versus theoretical peak performance), we allayed the concerns of the external reviewers and laid the groundwork to define new metrics for understanding the performance of ASCI applications.

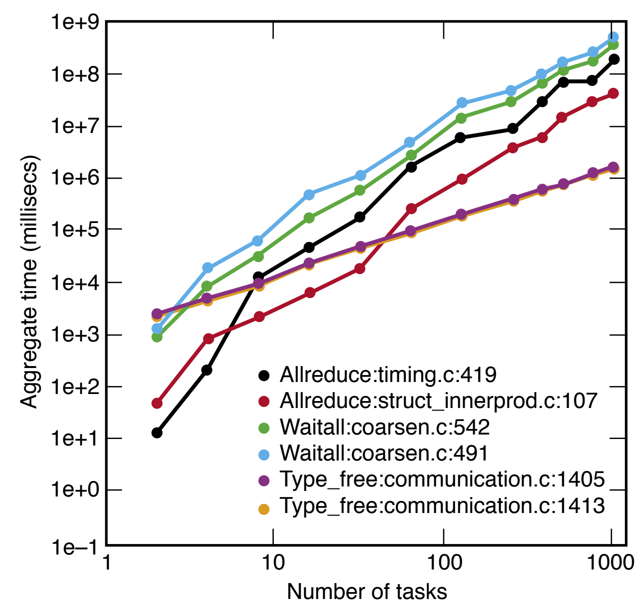


Figure 4.05-1. The mpiP tool shows how much time various MPI calls take as the number of tasks increases. Steep increases indicate possible scaling problems.

Problem Description

A major LLNL strategy in satisfying the need for production-quality capacity high-performance computing (HPC) cycles is the deployment of large-scale Linux clusters. These clusters take advantage of low-cost commodity processor components, which provides significantly more computing capacity than traditional proprietary HPC solutions. In conjunction with and critical to the success of this strategy is the development and deployment of an open source, production-quality Linux software stack.

Technical Approach/Status

The LLNL Linux software stack consists of three major components: an operating system, a parallel file system, and a resource management system. CHAOS is an in-house Red Hat-based Linux distribution that includes modifications for high-performance networks, cluster management and monitoring, and access control. Lustre is an open source high-performance parallel file system developed in part through a collaboration between LLNL and Cluster File Systems, Inc. (CFS) directed at achieving high performance within a single computing cluster, as well as in a shared environment. SLURM is a tool developed by LLNL and Linux NetworX to manage a queue of pending work, allocate access to nodes, and launch and manage parallel jobs.

Progress in 2003

CHAOS 1.2 was released in 2003 and is installed on all LLNL production Linux clusters. Several new or enhanced CHAOS components were released including a scalable, intra-cluster authentication service (*munge*) that enables secure and confidential application-level communication; a secure remote shell based on munge (*mrsh*); a multicast-based cluster heartbeat service (*whatsup*); a cluster configuration database (*genders*); and enhancements to the power management software (powerman). In addition, a pre-release of the 64-bit CHAOS 2.0 distribution needed for Thunder includes improved Itanium processor machine check support.

Lustre was first deployed in production on MCR including, late in the year, a multi-net capability that allows the PVC cluster to directly share MCR's file system, the first shared parallel file system at LLNL and an important step in enabling inter-system high-performance file sharing. Lustre was also deployed on a portion of the classified network. The reliability and performance of Lustre were improved significantly through the efforts of CFS and intensive testing by LLNL on the ALC cluster.

SLURM was installed on every Linux cluster at LLNL including MCR, ALC, PVC, Lilac, iLX, and Ace. It has proven to be reliable and highly scalable, with excellent performance characteristics. As a result of its success on Linux-based systems, SLURM is now being ported to AIX to deploy on the ASCI platforms in lieu of the existing resource management software.

Significance

LLNL's Linux software stack is critical to deploying production-quality HPC resources, and 2003 proved to be a major turning point in providing a production-capable software stack. Major advances made in CHAOS, Lustre, and SLURM have made production-quality, large-scale HPC Linux clusters a reality.

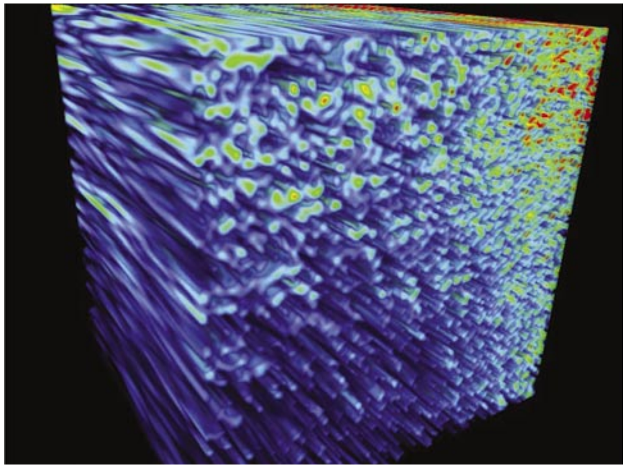


Figure 4.06-1. PF3D Simulation: 1920 processors on MCR cluster performed February 2003 (1/16th cross-section shown here).

4.07 Advancing the Code Development Environment

Problem Description

The goal for the parallel code development environment is to deliver the software tools and runtime libraries that allow applications to execute efficiently on the high-end computing platforms at LLNL.

Technical Approach/Status

Our parallel code development environment exploits a combination of local R&D and partnership development with academic and commercial sources. We test external products and provide requirements to external developers. The new large Linux clusters and planning for BG/L and Purple systems dominate our efforts. The LLNL user community insists that we stretch the limits of size and scalability, at the same time improving functionality and efficiency.

Progress in 2003

BG/L, a massively parallel cellular architecture system being developed jointly by IBM and LLNL, represents a significant architectural change from current ASCI systems. Adaptation of LLNL applications to this hardware prior to its delivery to LLNL in FY05 is critical to the overall project success. We are facilitating that adaptation through the use of BGLsim, a system simulator for parallel machines developed by IBM for hardware validation and software development.

Figure 4.07-1 shows that BGLsim models the complete BG/L hardware and system software environment. As a result, porting and tuning of applications on BGLsim directly correspond to results that will be seen on the as-yet-unbuilt machine. LLNL contributed to BGLsim by modifying the software so that it can use a variety of MPI implementations. In particular, LLNL modifications support the use of the native Quadrics MPI on Élan-based Linux systems, including the ASCI Linux Cluster. The simulator is installed on LLNL systems, and is instrumental in porting LLNL and ASCI Alliance applications to prototype BG/L hardware installed at IBM's Yorktown Heights facility.

The ROSE compiler project focused on sophisticated or domain-specific source-to-source translators that optimize existing scientific applications in C and C++. ROSE can trigger the generation of low-level platform-specific code to provide high performance, while preserving the simplicity and productivity of high-level of abstractions for the developer. Accomplishments during 2003 included the six-fold speedup of a C++ application, numerous loop optimizations, program analysis, and documentation.

The Etnus Totalview partnership added memory debugging functionality and a major speedup in launch of large jobs among the new product features. The Tool Gear infrastructure project created an interactive graphical interface for the mpiP communication-profiling tool to users' view source code annotated with communication performance information on both Linux and IBM AIX systems. Joint work with IBM on MPI and OS scalability work continued with development and test of a new kernel for the AIX OS that co-schedules interfering OS activities that are limiting scalability of collective operations

Significance

The thrust of effort in this area pushes the limits of scale, preparing for use of larger systems and also support of the large ASCI codes.

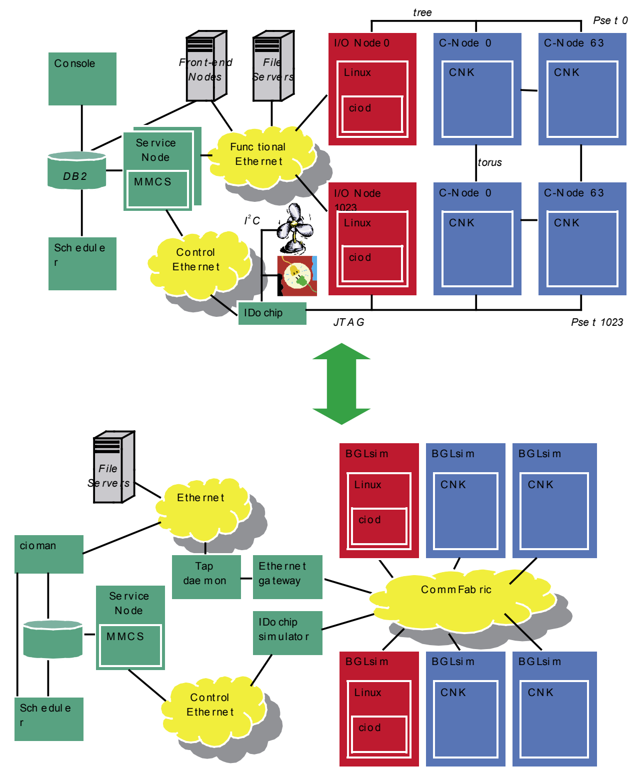


Figure 4.07-1. BGLsim, a system simulator for parallel machines developed by IBM for hardware validation and software development, models the complete BG/L hardware and system software environment.

Problem Description

A significant challenge in scientific data management is to improve scientists' interactions with huge data sets. Scientist's desire for ever-increasing resolution of their simulations is supported by ongoing increases in computational power. As a result, current simulations commonly produce data sets well over100 GB, and the average data set size is increasing. As data sets grow, it becomes increasingly difficult for scientists to gain insight from the data, because there is too much information to understand. To address this problem, scientists need a way to focus on relevant data and to eliminate unimportant information.

Technical Approach

To intelligently filter simulation data, we are developing an approximate, *ad hoc* query infrastructure. This infrastructure provides range-based queries against simulation data to identify regions of interest and produces a new data set containing only those regions. If an optional preprocessing step is performed, query accuracy can be traded for time, with the best answer possible within the given time constraints being returned.

Progress in 2003

A complex topology-based agglomeration algorithm was designed and implemented. A multi-resolution view of a data set is generated by repeatedly applying this algorithm to the corresponding mesh topology.

An incremental model creation algorithm was also designed and implemented, so that statistical models of the data could be created at the same time the hierarchy was built.

A single-pass clustering algorithm was designed and implemented. This algorithm identifies regions (i.e. nodes in the topology-based hierarchy) that are similar in multi-dimensional space and groups them together.

A beta version of the *ad hoc* query infrastructure that queries mesh files directly is currently in limited deployment. While it does not support trading time for accuracy, it is a parallel implementation that easily queries data sets of several hundred gigabytes. It has performed queries on 200-GB data sets such as shock wave tracking and selection of regions performing complex chemistry calculations.

Significance

The hierarchy creation and model-generation algorithms allow us to create multi-resolution hierarchies and models for complex mesh types, including unstructured and adaptive meshes. This dramatically increases the generality of our prototype.

The clustering algorithm will be used to improve the overall query performance by collocating similar regions on disk, so fewer reads are required to return relevant regions.

By performing a limited deployment of our prototype, we are able to make a direct impact on scientists' ability to understand their data, while also obtaining valuable insight into additional desired capabilities. This insight will allow us to focus our future research on the most valuable activities.

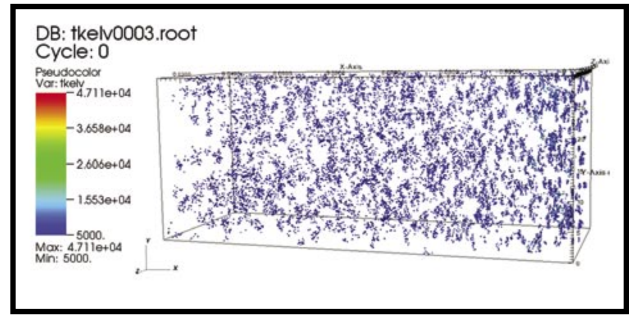


Figure 4.08-1. Regions within a time step in which complex chemistry calculations were performed.

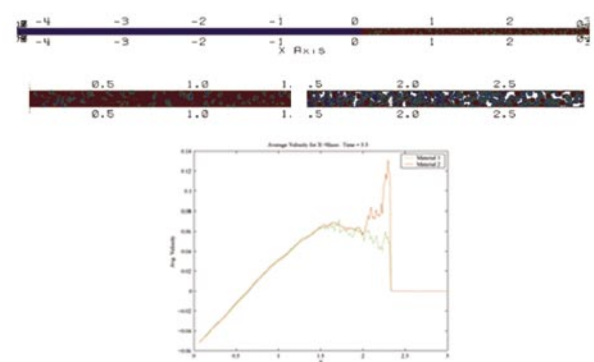


Figure 4.08-2. The top mesh is queried to generate the two meshes in the middle, and these meshes are analyzed to yield the resulting plot.

4.09 Scalable Scientific Data Mining

Problem Description

Advances in technology permit scientists to gather data from experiments, simulations, and observations at an ever-increasing pace. These massive, complex data sets are available as time-series or as images. Since it is impractical to manually analyze, explore, and understand this data, useful information is often overlooked, and the potential benefits of increased computational and data gathering capabilities are only partially realized.

Technical Approach/Status

The Sapphire project addresses the challenge of data overload by applying and extending ideas from the multi-disciplinary field of data mining. We conduct research in algorithms, incorporate the research into software, and apply the software to real-world problems at LLNL. The needs of these applications, in turn, drive our research.

We define data mining as the end-to-end process of extracting useful information from raw data. We focus on the compute-intensive activities—processing the data to extract objects and relevant features, dimension reduction techniques to identify key features, and pattern recognition techniques—to identify and characterize patterns in the data, which are then shown to the scientist for validation.

Our focus has evolved from the development of algorithms and software, to include the application of the software to problems of interest. Our object-oriented software currently supports all steps in the data mining process, with several algorithms provided for each step. The software is serial; many parts are embarrassingly parallel, and additional support for parallelism will be provided as required by the applications.

Progress in 2003

Our primary effort was in applications. We developed a Similarity-Based Object Retrieval (SBOR) system for retrieving objects in image and mesh data that are similar to a given query object (Figure 4.09-1). We are collaborating with AX Division on a code validation problem, and we are working with physicists at the DIII-D Tokamak to identify key features associated with the quiescent H-mode in the plasma. We began work on the detection and tracking of moving objects in video, a technique of interest in surveillance and computer simulations. We are also collaborating on computer security problems with CIAC.

We also investigated several algorithms for feature selection, the use of texture features for more efficient retrieval of high-resolution, remote-sensing imagery, and improved techniques for background subtraction to detect moving objects (Figure 4.09- 2).

Significance

Our progress in 2003 has helped LLNL scientists in several directorates, enabling us to apply cutting-edge data analysis techniques to their problems. Several of these problems are rather difficult and require further development of innovative approaches. Our research and involvement in professional activities continue to place Sapphire at the forefront of the scalable scientific data-mining field.

Figure 4.09-1. The Sapphire Similarity-Based Object Retrieval (SBOR) system allows a user to identify an object of interest in an image (left) and returns similar objects in the image database, ranked by similarity (right). This example shows that the system is invariant to rotation and translation.

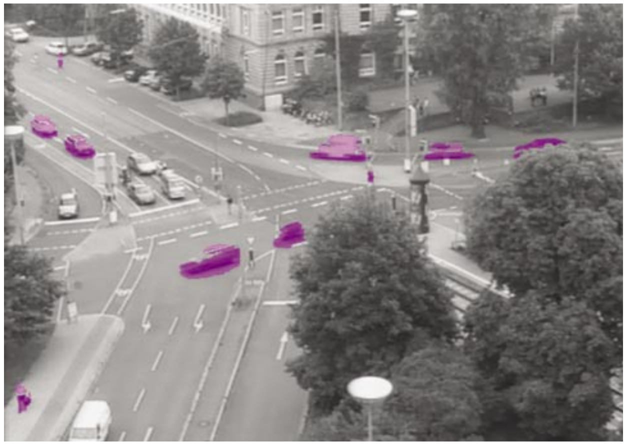


Figure 4.09-2. A frame from a video, with the moving objects highlighted. These objects were detected using background subtraction algorithms.

Problem Description

Laboratory scientists create very high-resolution models at an ever-increasing rate, which translates into their need to explore terabytes of data in real time. The challenge is to produce interactively any view of a high-resolution model by accessing the minimum amount of data. Moreover, one must help the user understand reliably the fundamental structures present in the model by exploring the minimum number of views.

Technical Approach

In response to this trend, the visualization group is implementing a long-term research plan based on the combination of two main strategies: redefinition of the visualization pipeline as a streaming process based on progressive and cache oblivious algorithms, and development of new data analysis tools that are tightly coupled with the visualization process and guide the user in navigating the data. This new visualization pipeline allows developing software tools that are intrinsically scalable with the size of the input problem and the performance of the computing resources. In our complementary research strategy, we introduce data analysis tools computing intrinsic topological and metric properties that help the user understand the structure of the data.

Progress in 2003

We completed our first prototype of the ViSUS Progressive Viewer, demonstrating the unprecedented capability of effectively exploring large data sets (e.g., 8-billion-node mesh) with resources as modest as a laptop computer or as large as a parallel visualization server driving a PowerWall display. We generated the Figure 4.10-1 images on a Dell laptop. Furthermore, we connected our streaming infrastructure to simulation codes developed independently with minimal code intrusion (one function call at the end of the time steps visualized).

Based on a Morse–Theoretical framework, we introduced algorithms for efficient and stable computation of critical points in a scalar field, their organization in the Morse–Smale Complex, and their persistence at different levels of resolution. The work is moving toward the definition of data comparison metrics and the analysis of dynamic structures.

Significance

Our streaming infrastructure optimizes the use of human and computing resources by providing three major new capabilities: large-scale visualization on low-end computers, remote visualization, and real-time monitoring of parallel simulations. A scientist can explore data sets that are tens to hundreds of GB on an office desktop workstation. Using a 10-MB network, we can access in real time remote data sets that are terabytes in size. We reduced our preprocessing time to a negligible amount so that even for high-resolution simulations we can have one time-step ready for visualization as the next one is still being computed. This enables new capabilities such as early detection and the elimination of wasteful executions.

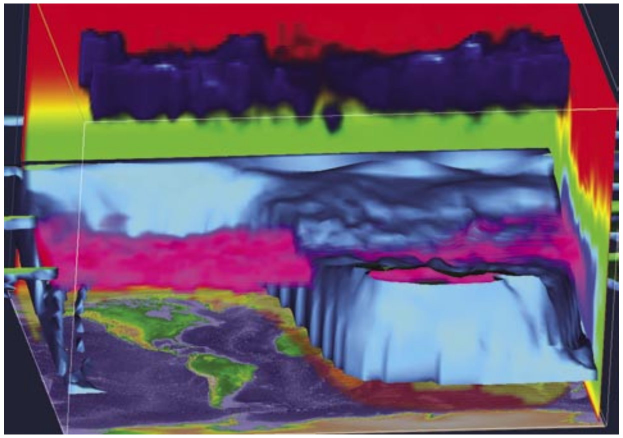


Figure 4.10-1. Researchers can explore terabytes of data in real time, via a streaming-infrastructure visualization pipeline.

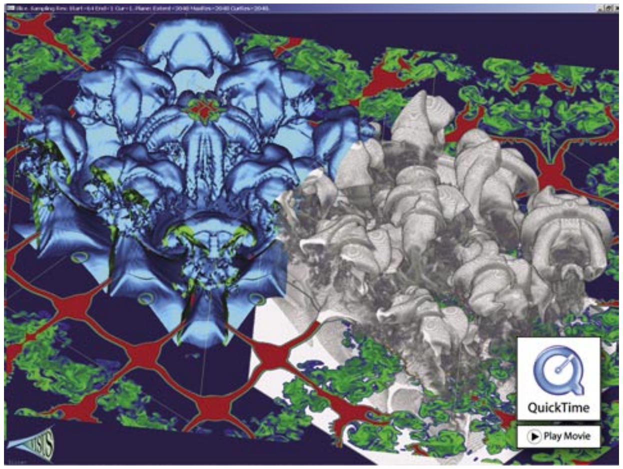


Figure 4-10-2. PPM Simulation, winner of a Gordon Bell Prize.

Academia

Austrian National Science University
Beckman Institute
Bristol University
California Institute of Technology
Carnegie–Mellon University
Columbia University
Cornell University
Dresden University of Technology
Duke University
Georgia Institute of Technology
Heidelberg University
Indiana University
New Mexico Institute of Mining and Technology
New York University
North Carolina State University
Northwestern University
Old Dominion University
Oregon Graduate Institute
Oregon Health and Science University
Penn State University
Purdue University
Rensselaer Polytechnic Institute
San Diego Supercomputer Center
San Jose State University
Southwest Texas State University
Stanford University
Technical University of Catalonia
Technical University of Vienna
Texas A&M University
U.S. Air Force Academy
Universitaet Bonn
Universitaet Erlangen–Nuernberg
University of California, Berkeley
University of California, Davis
University of California, Irvine
University of California, Los Angeles
University of California, San Diego
University of California, Santa Barbara
University of California, Santa Cruz
University of Chicago
University of Colorado, Boulder
University of Colorado, Denver
University of Illinois
University of Loeben
University of Maryland
University of Michigan
University of North Carolina
University of Oregon
University of San Francisco
University of Tennessee
University of Utah
University of Washington
University of Wisconsin
University of North Carolina
Warsaw University
Worcester Polytechnic Institute

Industry

BlueArc
Cluster File Systems
Etnus
HP
IBM
Intel
Krell Institute
Limit Point Systems
Linux NetworX
MPI Software Technology, Inc.
Network Appliance
Pallas
Quadrics
Red Hat
ZeroFault

National Labs/Government

Argonne National Laboratory
Brookhaven National Laboratory
Lawrence Berkeley National Laboratory
Los Alamos National Laboratory
Oak Ridge National Laboratory
Pacific Northwest National Laboratory
Sandia National Laboratories

5.02 Computation Contact Information

Topic	Authors	Web Sites
Computing at Lawrence Livermore	Dona Crawford	http://www.llnl.gov/comp/
Providing Desktops to Teraflops	Michel McCoy	http://www.llnl.gov/asci/
Platform Strategy and Systems	Jim Garlick, Terry Heidelberg, Moe Jette, Doug East	http://www.llnl.gov/asci/platforms/platforms.html
HPC Systems Area Network	Keith Fitzgerald	
Visualization and Data Assessment	Steve Louis	http://www.llnl.gov/asci/views/views.html
User Services	Brian Carnes	http://www.llnl.gov/icc/sdd/
Desktop Support	Kim Cupps	
Laboratory Network Services	Dave Wiltzius	
Cyber Security	Ed Matsche, Ted Michels	
Developing Applications Software	Steven Ashby	http://www.llnl.gov/car/
ASCI Code Development	Evi Dube	http://www.llnl.gov/asci/applications/applications.html
Verification & Validation Work	James McEnerney, George Christianson	
NIF Control System	Larry Lagin, Kim Minuzzo	http://www.llnl.gov/nif/
Bio-Defense	Carol Zhou, Tom Slezak	http://greengenes.llnl.gov/bbrp/html/mccreadyabst.html
Radiation Transport for Cargo Inspection	Peter Brown	http://www.llnl.gov/casc/Ardra/
NARAC	Joe Tull	http://narac.llnl.gov
Coupled Climate – Carbon Modeling	Art Mirin	
First Principles	Francois Gygi	http://www.llnl.gov/str/Gygi.html
Computer Incident Advisory Capability	Denise Sumikawa	http://www.ciac.org/ciac/CIACHome.html
Information Operations And Assurance	Diana Sackett	
Software Quality Assurance	Trish Damkroger, Joyce Moulden	
Engaging in Computing Research	Pete Eltgroth	http://www.llnl.gov/casc/
Scalable Linear Solvers	Robert Falgout, Jim Jones, Carol Woodward, Charles Tong	http://www.llnl.gov/CASC/linear_solvers/
Overture	Lori Feitag-Diachin	http://www.llnl.gov/casc/Overture/
SAMRAI	Richard Hornung	http://www.llnl.gov/casc/SAMRAI/samrai_home.html
Tiling Models for Spatial Decomp.	John Compton	
Performance of Multi-Physics	Rob Neely, John May	
Linux Cluster Software Stack	Doug East, Robin Goldstone, Moe Jette	
Advancing the Code Development Environ.	Mary Zosel, Dan Quinlan	
Scalable Data Management	Terence Critchlow, Randy Frank, Jeff Long	http://www.llnl.gov/CASC/datafoundry/
Scalable Data Mining	Chandrika Kamath	http://www.llnl.gov/casc/sapphire/
Scalable Interactive Data Exploration Tools	Valerio Pascucci	
Sections 5.01-5.02	Jim McGraw	http://www.llnl.gov/iscr/

5.03 Acronyms and Abbreviations

Selected Acronyms and Abbreviations, with Brief Descriptions

AD
Associate Director –LLNL Directorate senior manager
Active Directory – Microsoft proprietary product for administering large/complex systems

ALC—ASCI Linux cluster at LLNL

ALE3D—Arbitrary Lagrangian–Eulerian three-dimensional code

AMR—adaptive mesh refinement

ARGUS—DOE's standard high-security system, protects assets at LLNL, Pantex, INEEL, DOE HQ, and LANL; includes personnel access control booths, alarm stations, map-based alarm reporting systems and a closed-circuit TV video assessment system.

ASC—DHS's Advanced Scientific Computing research program

ASD—automated software delivery, via the tool Radia

ASCI—currently the Advanced Simulation & Computing program for NNSA/DOE, historically the national Accelerated Strategic Computing Initiative

ASCI Alliances—academic institutions hosting ASCI research & development

ASCI Blue–Pacific—existing LLNL/Tri-Lab IBM Silver 344-node system machine for unclassified access, allowing collaborative research access; peak of 1.3 TF

ASCI Purple—proposed 100-TF Tri-Lab machine to be installed at LLNL mid-2005

ASCI White—existing LLNL/Tri-Lab machine composed of three separate systems based upon IBM's POWER3 SP technology. The largest system is a 512-node SMP (16 CPUs/node) system with a peak speed slightly greater than 12 TF

ASIC—application-specific integrated circuit, first delivered to LLNL 6/6/03

BASIS—Biological Aerosol Sentry and Information System; enables early detection of biological pathogens, used at 2000 Olympics

BG/L—BlueGene/L, a 180–360-TF cell-based machine developed in partnership with IBM

BBRP—Biology and Biotechnology Research Program at LLNL

CHAOS—Clustered High-Availability Operating System software stack at LLNL, augments Linux Red Hat with support for HPC clusters

CORBA—Common Object Request Broker Architecture

CSP—Computer Security Program at LLNL

DAG—Desktop Advisory Group at LLNL

DHS—U.S. Department of Homeland Security

Directorate —LLNL organizational unit dedicated to a specific discipline or science, in particular, the Computation Directorate

DM—data mining, the extraction of relevant information from massive data sets

DNT—Defense and Nuclear Technologies Directorate at LLNL

DOE—U.S. Department of Energy
DOE/HQ DOE headquarters in Washington, DC

DSW—Directed Stockpile Work, supports re-certification of weapons systems

FIS—File Interchange Systems

GF—gigaflop, 10^9 floating-point operations/second

GigE—Gigabit Ethernet

HPC—high-performance computing

HPSS—high-performance storage system

HSI—high-speed interconnect

ICCS—NIF's Integrated Computer Control System software

IDR—intrusion detection and response security "fabric" over LLNL networks

INEEL—Idaho National Engineering & Environmental Laboratory (an NNSA Lab)

IPSO—Information Protection Support Organization

ISM—Integrated Safety Management program at LLNL to protect worker/occupational health and safety

LANL—Los Alamos National Laboratory

LC—LLNL Computing Center, the computing infrastructure at LLNL

Linux Software Stack—operating system, parallel file system, and resource management system

LLNL—Lawrence Livermore National Laboratory

M&IC—Mulitprogrammatic and Institutional Computing at LLNL

MCR—11.2-TF, 32-bit microprocessor-based cluster Mulitprogrammatic Capability Resource, combines open-source software with cluster architecture

MPI—message-passing interface

NAI—Nonproliferation, Arms Control and International Security Directorate at LLNL

NFS—network file system

NIF—National Ignition Facility, national research and test center for laser fusion, a Directorate at LLNL

NSF—National Science Foundation

OC–12—622 Mb/s interface

ORNL—Oak Ridge National Laboratory

OTP—one-time password authentication system.

PDE—partial differential equation

PF—petaflop, 10^{15} floating-point operations/second

PSAP—Personnel Security Assurance Program access authorization

PSE—ASCI Tri-Lab Problem Solving Environment

PVC—parallel visualization cluster

R&D 100 Award —R&D Magazine award winner for a product or technology first available for order or license to the private sector during the previous year

SAN—storage area network

SLURM —(Simple Linux Utility for Resource Management), a tool developed by LLNL and Linux NetworX to manage a queue of pending work, allocate access to nodes, and launch and manage parallel jobs

Slow clock —500 Mhz or slower

SNL—Sandia National Laboratories (/CA in California; /NM in New Mexico)

SOC—system-on-a-chip, IBM's proprietary technology for embedding applications

SQA—software quality assurance

SSP—Stockpile Stewardship Program that oversees the safety, security, and reliability using, amoung other methods, the nation's nuclear stockpile high-fidelity weapon simulation capabilities

SWGFS—Site-Wide Global File System

Teller Fellowship Award—LLNL Director's award modeled on the MacArthur Fellowship program, funds the recipient to do a year of self-directed work that will benefit LLNL

TF—teraflop, 10^{12} floating-point operations/second

Thunder—23-TF system, with architecture modeled on MCR

Tri-Valley—Livermore–Dublin–Pleasanton geographic and socio-economic area

TSF—Terascale Simulation Facility under construction at LLNL

VIEWS—ASCI Tri-Lab Visual Interactive Environment for Weapons Simulation