

Overview of the Advanced Simulation and Computing Program (ASCI)

Contents

1. INTRODUCTION	1
2. ASCI OVERVIEW	2
2.1 ONE PROGRAM-THREE LABORATORIES.....	2
2.2 SCIENCE-BASED STEWARDSHIP.....	2
2.3 THE RIGHT ENVIRONMENT.....	2
3. DEFENCE APPLICATIONS AND MODELLING	3
3.1 ADVANCED APPLICATIONS DEVELOPMENT OVERVIEW.....	3
3.2 VERIFICATION AND VALIDATION OVERVIEW.....	3
3.3 MATERIALS AND PHYSICS MODELLING OVERVIEW.....	3
4. SIMULATION & COMPUTER SCIENCE	4
4.1 PROBLEM-SOLVING ENVIRONMENT OVERVIEW.....	4
4.2 DISTANCE AND DISTRIBUTED COMPUTING AND COMMUNICATIONS OVERVIEW.....	4
4.3 PATHFORWARD OVERVIEW.....	4
4.4 VISUAL INTERACTIVE ENVIRONMENT FOR WEAPONS SIMULATION.....	5
5. INTEGRATED COMPUTING SYSTEMS	6
5.1 PHYSICAL INFRASTRUCTURE AND PLATFORMS OVERVIEW.....	6
5.2 ON-GOING COMPUTING OVERVIEW.....	7
5.3 THE ASCI RED TOPS SUPERCOMPUTER.....	7
5.4 HARDWARE ENVIRONMENT ON ASCI BLUE-PACIFIC.....	9
5.5 THE ASCI BLUE MOUNTAIN 3-TOPS SYSTEM.....	9
5.6 HARDWARE ENVIRONMENT ON ASCI WHITE.....	11
5.7 TRI-LAB ASCI WHITECAP.....	15
5.8 ASCI Q.....	16
5.9 ASCI PURPLE.....	16
6. UNIVERSITY PARTNERSHIPS	17
6.1 ACADEMIC STRATEGIC ALLIANCES PROGRAM OVERVIEW.....	17
6.2 ASCI INSTITUTES.....	17

1. INTRODUCTION

On October 2, 1992, President Bush signed into law the FY1993 Energy and Water Authorisation Bill that established a moratorium on U.S. nuclear testing. President Clinton extended the moratorium on July 3, 1993. These decisions ushered in a new era by which the U.S. ensures confidence in the safety, performance, and reliability of its nuclear stockpile. The U.S. also decided to halt new nuclear weapon production. This decision meant that the nation's stockpile of nuclear weapons would need to be maintained far beyond its original design lifetime.

The United States Department of Energy/National Nuclear Security Administration (NNSA) oversees the nation's Stockpile Stewardship effort. Without underground testing, computer simulations are needed to make sure that the nuclear weapons stockpile is safe, reliable, and operational. NNSA's

supercomputers will compute the factors involved in a nuclear detonation - including a weapon's age and design - and eventually allow the NNSA to manage its entire stockpile of nuclear weapons without any real nuclear tests.

To implement these pivotal policy decisions, the Stockpile Stewardship Program was established. The goal of this program is to provide scientists and engineers with the technical capabilities to maintain a credible nuclear deterrent without the use of the two key tools used to do that job over the past 50 years: (1) underground nuclear testing and (2) modernization through development of new weapon systems.

The Advanced Simulation and Computing Program (ASC), historically referred to as ASCI, is the Department of Energy, NNSA's collaboration with Lawrence Livermore, Los Alamos, and Sandia national laboratories to ensure the safety and reliability of the nation's nuclear weapons stockpile. To accomplish their mission, the national laboratories in turn work in a strategic partnership with computer manufacturers and several of the nation's major universities. This powerful partnership has helped ASCI become the integrating force that preserves weapons design and testing experience, uses experimental data from aboveground test facilities along with the archive of nuclear test data, and improves the scientific understanding that provides high-confidence predictive simulation capabilities

2. ASCI OVERVIEW

The Accelerated Strategic Computing Initiative (ASCI) is a large, complex, multifaceted, and highly integrated research and development effort. The program's objective is to meet the science and simulation requirements of the Stockpile Stewardship Program for the National Nuclear Security Administration (NNSA).

2.1 One Program-Three Laboratories

The problems that ASCI will solve for the Stockpile Stewardship Program span the activities and responsibilities of the three DOE Defence Programs national security laboratories (Lawrence Livermore, Los Alamos, and Sandia). Co-operation among the laboratories is essential to solving these problems in an efficient and effective manner. ASCI is implemented by project leaders at each of the labs, guided by the NNSA Office of Advanced Simulation and Computing. The labs share ASCI code development, computing storage, visualization, and communication resources across laboratory boundaries in joint development efforts.

2.2 Science-Based Stewardship

ASCI provides computational and simulation capabilities that will help scientists understand ageing weapons, predict when components will have to be replaced, and evaluate the implications of changes in materials and fabrication processes. This science-based understanding is essential to ensure that changes brought about through ageing or re-manufacturing will not adversely affect the enduring stockpile.

Applications must achieve higher resolution, higher fidelity, three-dimensional physics, and full-system modelling capabilities to reduce reliance on empirical judgements. This level of simulation requires high-performance computing far beyond our current level of performance.

2.3 The Right Environment

A powerful problem-solving environment must be established to support application development and enable efficient and productive use of the new computing systems:

- ◆ High-performance, full-system, high-fidelity physics predictive codes are required that support weapon assessments, manufacturing process analyses, accident analyses, and certification.
- ◆ Creation of a computational infrastructure and operating environment is necessary to make these capabilities accessible and usable.
- ◆ Close co-operation with the computer industry to accelerate their business plans helps provide the computational platforms needed to support ASCI applications.

- ◆ Universities also play a critical role in developing computational tools and scientific understanding needed for this unprecedented level of simulations.

3. DEFENCE APPLICATIONS AND MODELLING

3.1 Advanced Applications Development Overview

ASCI is developing, on an accelerated schedule, the progressively higher performance software applications needed to provide the simulation tools and underpinnings for the Stockpile Stewardship Program (SSP). The key to reaching SSP objectives for initial implementation in 2004 and full implementation in 2010 is the ability to achieve ASCI's critical simulation and applications code milestones in the intervening years. ASCI will provide simulation tools and capabilities embodying the physical and chemical processes needed to predict the safety, reliability, performance, and manufacturability of weapon systems. It is a formidable challenge to replace the empirical factors and adjustable parameters used in current calculations with predictive physical models.

Meeting this challenge will require large, complex computer applications codes; this drives the scale of computing hardware and infrastructure. However, increased capability in hardware and infrastructure alone is insufficient. Much of the increased computational capability to be provided by ASCI must come from advances in the applications codes themselves.

These applications will integrate 3D capability, finer spatial resolution, and more accurate and robust physics. Adding these new capabilities, however, will strain the limits of the algorithms used in today's simulation codes. In addition, the necessity to do full-system or scenario simulations will require the development of new algorithms for coupled systems. As a consequence, the development and implementation of improved numerical algorithms to address these new capabilities will be a critical component of the applications strategies. Tightly integrated code teams-large interdisciplinary work groups consisting of scientists and engineers, along with computational mathematicians and computer scientists, devoted to producing coherent software for efficient predictive simulations-will develop these codes.

3.2 Verification and Validation Overview

Verification and validation provide high confidence in the computational accuracy of the simulations that support stockpile stewardship by systematically measuring, documenting, and demonstrating the predictive capability of the codes and their underlying models. Verification is the process of determining that a computational software implementation correctly represents a model of a physical process. Validation is the process of determining the degree to which a computer model is an accurate representation of the real world from the perspective of the intended model applications. Validation makes use of physical data and results from previously validated legacy codes.

3.3 Materials and Physics Modelling Overview

Experimental, theoretical, and computational capabilities are required to predict the physical properties of materials under conditions found in nuclear explosions. Of particular interest are the dynamic properties and response of materials under conditions of high strain and high-strain rates, impact, shock compression and quasi-isentropic loading.

Laboratory experiments and high-performance simulations will provide the basis for the development of predictive models and validated physical data of stockpile materials. Ultra-scale scientific computing platforms, multi-physics application codes, and unique experimental facilities have been deployed and integrated to establish these predictive capabilities.

High-performance simulations linking atomistic to continuum scales will lead to reliability predictions and lifetime assessment for corrosion, organic degradation, and thermal-mechanical fatigue of weapon electronics.

4. SIMULATION & COMPUTER SCIENCE

4.1 Problem-Solving Environment Overview

ASCI's unprecedented code development effort requires robust computing and development environments enabling codes to be developed rapidly. Through the Problem Solving Environment (PSE) program, ASCI is developing a computational infrastructure to allow applications to execute efficiently on the ASCI computer platforms and to provide accessibility from the desktops of scientists and engineers.

This computational infrastructure will include software development tools, run-time libraries, frameworks, solvers, archival storage, high speed interconnects, scalable I/O, local area networks, distributed computing environments, and software engineering.

The ASCI Defence Applications and Modelling program is the main driver for the PSE program requirements and the primary customer for PSE products and services. Thus, PSE is responsible for providing the weapon scientists and engineers, and the application developers the computational tools they need for the development and execution of applications on ASCI platforms.

Furthermore, in collaboration with third-party and platform partners, PSE is responsible for deploying system software on all ASCI platforms. PSE's program objective is to provide a complete software environment where products come from a variety of suppliers, including the platform partners, third-party Independent Software Vendors (ISVs), alliance partners, and laboratory R&D. PSE works closely with platform partners to define software requirements and priorities, and engages in collaborative development in critical areas where the platform providers may not have access to machine resources.

4.2 Distance and Distributed Computing and Communications Overview

DisCom² will assist in the development of an integrated secure information, simulation, and Modelling capability to support the design, analysis, manufacturing, and certification functions of the

Defence Programs complex through developments in two key strategic areas:

- Distance Computing will extend the environments required to support high-end computing to all sites. It will partner with the National Security Administration (NSA) to develop the high-speed encryptors required to interconnect the laboratories securely.
- Distributed Computing will develop an enterprise-wide integrated supercomputing architecture that will support nuclear-weapon science and engineering requirements for stockpile stewardship. It will take advantage of the ongoing revolution in commodity, cluster-based, distributed high-performance computing. It will adopt, support, and augment the open software approach to distributed cluster computing. A product will be an evolving set of commodity computing systems that compliment the ASCI terascale, ultra-computing platforms. These solutions will fit seamlessly into the secure ASCI computing environment.

4.3 PathForward Overview

A key ASCI strategy is to construct future high-end computing systems and environments by scaling commercially viable building blocks, both hardware and software. PathForward is the program designed to achieve such goals by entering into partnerships with U.S. industry to develop the key technologies necessary to accelerate the development of balanced 100-teraOPS-computer systems, and beyond.

These partnerships develop and accelerate technologies in the vendors' current business plans, but not available in either the time frame or the scale required by ASCI. Starting in 1998, PathForward focused on interconnect and data storage technologies, as well as systems software and tools for large-scale computing systems. These technologies were considered the most critical components in the construction of the 30-teraOPS-machine environment in 2001.

These are the essential scaling and integrating technologies that will enable ultra-scale computing systems to be engineered and developed out of both hardware and software commodity computing building blocks.

In FY2001, new projects in the areas of visualization and runtime system software were added to the PathForward portfolio.

4.4 Visual Interactive Environment for Weapons Simulation

The Visual Interactive Environment for Weapons Simulation (VIEWS) focuses on the problem of seeing and understanding the results of multiple teraOPS simulations and comparing results across and between simulations and experiments. One of the goals of VIEWS is to create an infrastructure called Data and Visualization Corridors that enable scientists to view and use the results of high-fidelity three-dimensional simulation. Strategies include

- Partnering with academia, industry, federal agency research and development.
- Visually exploring and interactively manipulating massive, complex data.
- Developing a means to orchestrate and effectively manage, extract, and deliver data.
- Developing efficient solutions for remote and collaborative scientific data exploration.
- Deploying the highest-performance data visualization facilities.

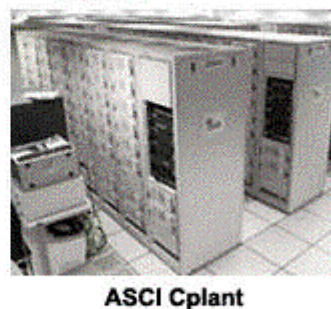
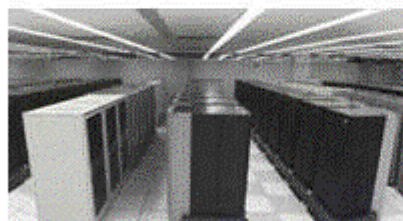
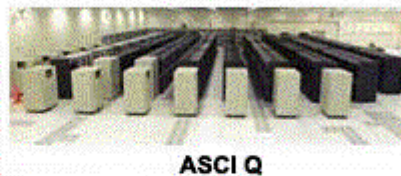
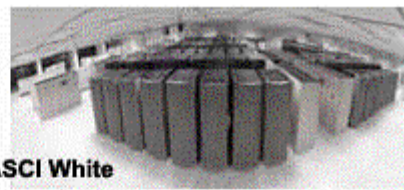
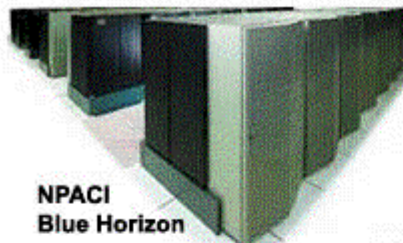
5. INTEGRATED COMPUTING SYSTEMS

5.1 Physical Infrastructure and Platforms Overview

The first supercomputers were developed for weapon applications in the 1960s as a partnership between the computer industry and the nuclear weapons laboratories. Throughout the 1970s and 1980s, this relationship continued. The Defence Programs laboratories were early user sites and a primary customer for new state-of-the-art high-performance computers and computing simulation capability.

In the late 1980s and early 1990s this relationship changed. The NNSA national security laboratories drastically reduced their partnerships with industry in developing computer systems. Due to significant budget reductions, they greatly reduced their purchases of supercomputers. This was also a period of momentous change within the computing industry, as minicomputers eroded the marketplace of mainframes, followed in turn by the cannibalization of the minicomputer market by microcomputers. As a result, the computing industry as a whole no longer viewed the NNSA laboratories as a primary customer for their most advanced computer designs.

Today, more powerful computing platforms are needed to achieve the performance simulation and virtual prototyping applications that the Stockpile Stewardship Program requires. ASCI has stimulated the U.S. computing industry to develop high-performance computers with speeds and memory capacities hundreds of times greater than currently available models and ten to several hundred times



greater than the largest computers that were likely to result from recent development trends. ASCI will

continue to partner with various U.S. computer manufacturers to accelerate the development of larger, faster computer systems and software that are required to run Defence Programs' demanding applications.

ASCI partnerships have brought about development and installation of the world's first teraOPS computer (the 1.8-teraOPS Intel machine at Sandia, Albuquerque), 3 three-plus teraOPS machines (at Los Alamos, in partnership with SGI; at Livermore, in partnership with IBM; and at Sandia, upgraded in 1999 to 3 teraOPS by Intel). An extension of the Lawrence Livermore/IBM contract allowed the development of a 12-teraOPS machine that was installed in Livermore in 2000.

5.2 On-going Computing Overview

The Ongoing Computing element is focused on making the computing resources needed to support stewardship available to the laboratories. This element is structured somewhat differently at each of the laboratories, but program-wide it is focused on the operation of the computer centres at the three laboratories. In general, that effort has two mission elements:

- (1) to provide ongoing stable production computing services to laboratory programs, and
- (2) to foster the evolution of simulation capabilities towards a production terascale environment as ASCI computer platforms evolve towards the 100-teraOPS level. This effort consists of software infrastructure, the networks, data storage, and output systems.

In the following sections we describe the hardware infrastructure associated with each of the following systems:

- 1) The ASCI Red TOPS Supercomputer
- 2) The ASCI Blue Pacific Supercomputer
- 3) The ASCI Blue Mountain Supercomputer
- 4) The ASCI White Supercomputer
- 5) ASCI Q , and the recently announced
- 6) ASCI Purple.

5.3 The ASCI Red TOPS Supercomputer

5.3.1 Introduction

The ASCI Red TOPS (Tera-Operations per Second) Supercomputer is the first step in the ASCI Platforms Strategy, which is aimed at giving researchers the five-order-of-magnitude increase in computing performance over current technology that is required to support "full-physics," "full-system" simulation by early next century. This supercomputer, installed at Sandia National Laboratories, is a massively parallel, MIMD (Multiple Instruction, Multiple Data) computer. It is noteworthy for several reasons. It was the world's first TOPS supercomputer. I/O, memory, compute nodes, and communication are scalable to an extreme degree. Standard parallel interfaces make it relatively simple to port parallel applications to this system. The system uses two operating systems to make the computer both familiar to the user (UNIX) and non-intrusive for the scalable application (Cougar). And it makes use of Commercial Commodity Off The Shelf (CCOTS) technology to maintain affordability.

5.3.2 Hardware

The ASCI TOPS system is a distributed memory, MIMD, message-passing supercomputer. All aspects of this system architecture are scalable, including communication bandwidth, main memory, internal disk storage capacity, and I/O.

The TOPS Supercomputer is organized into four partitions: Compute, Service, System, and I/O. The Service Partition provides an integrated, scalable host that supports interactive users, application development, and system administration. The I/O Partition supports a scalable file system and network services. The System Partition supports system Reliability, Availability, and Serviceability (RAS) capabilities. Finally, the Compute Partition contains nodes optimized for floating point performance and is where parallel applications execute. The system hardware parameters are summarized in Table 1.

Table 1. System hardware parameters

Compute Nodes (Red - Red / Black - Black)	4,510 (1,166 - 2,176 - 1,168)
Service Nodes (Red / Black)	52 (26 / 26)
Disk I/O Nodes (Red / Black)	73 (37 / 36)
System Nodes (Red / Black)	2 (1 / 1)
Network Nodes - Ethernet/ATM (Red / Black)	12 (6 / 6)
System Footprint	~2500 Square Feet
Number of Cabinets (Computer / Switch / Disk)	104 (76 / 8 / 20)
System RAM (Compute Nodes / I/O Nodes)	1212 GB Total (256 MB / 512 MB)
Topology	Mesh (38 X 32 X 2)
Node Link Bandwidth - Bi-directional	800 MB/s
Cross Section Bandwidth - Bi-directional	51.2 GB/s
Total Number of Pentium II Xeon Core Processors	9298
Processor to Memory Bandwidth	533 MB/s
Compute Node Peak Performance	666 MOPs
System Peak Performance	3.1 TOPs
Linpack Performance - Full System (Center + Red or Black / Red or Black)	2.38 TOPs (1.6333 TOPs / .581 TOPs)
RAID Disk Storage - Total / per Color	12.5 TB / 6.25 TB
RAID I/O Bandwidth - Total per Subsystem	4.0 GB/s 1.0 GB/s

5.3.3 Software

Software on the TOPS Supercomputer is a combination of operating systems tailored for specific tasks and standard programming tools to make the computer both familiar to the user and non-intrusive for the scalable application. To the application programmer, the system looks like a UNIX-based supercomputer. All the standard facilities associated with a UNIX workstation will be available to the user.

The operating system used for the Service, I/O, and System Partitions is Intel's distributed version of UNIX (POSIX 1003.1 and XPG3, AT&T System V.3 and 4.3 BSD Reno VFS) developed for the Paragon XP/S Supercomputer. The Paragon OS presents a single system image to the user. This means that users see the system as a single UNIX machine despite the fact that the operating system is running on a distributed collection of nodes.

The operating system in the Compute Partition is Cougar. Cougar is Intel's port of Puma, a light-weight operating system for the TOPS, based on the very successful SUNMOS system for the Paragon. (SUNMOS, and subsequently Puma, were developed by Sandia National Laboratories and the University of New Mexico.) System services and support for the interactive user are provided by a host OS (in this case, the Paragon OS running in the Service Partition). All access to hardware resources comes from the Q-Kernel, the lowest-level component of Cougar. Above the Q-Kernel sits the process control thread (PCT), which runs in user space and manages processes. At the highest level is the user's applications. As with most MPP systems, the basic programming model in Cougar is based on message passing.

FORTRAN77, FORTRAN90, C and C++ are supported. The interactive debugger and performance analysis tools understand these languages and map onto original source code.

5.3.4 Conclusion

The ASCI platform effort bridges the gap between giga-scale and tera-scale computing to accommodate the five-order-of-magnitude increase in performance required by "full-physics", "full-system" simulation.

5.4 Hardware Environment on ASCI Blue-Pacific

SST: Sustained Stewardship TeraOP
CTR: Combined Technology Refresh

Blue-Pacific System Attributes	SST System (Current)	CTR System (Current)
Total Nodes	1464 4-CPU SMP nodes consisting of 3x488-node sectors, S, K, & Y	280 4-CPU nodes
Compute Nodes	1296	256
Login Nodes	6 total (2 per sector)	2
Total CPUs	5856	1120
Total Compute CPUs	5184	1024
Memory per Node	1.5-2.5 GB (432 nodes with 2.5 GB)	1.5 GB
System RAM	2.6 Tbytes	420 Gbytes
Processor	332 MHz PowerPC 604e	332 MHz PowerPC 604e
Processors per Compute Node	4	4
Node to Node Bandwidth; Bi-directional	150 Mbyte/s	150 Mbyte/s
Processor to Memory Bandwidth	2.1 Tbyte/sec	2.1 Tbyte/s
Compute Node Peak Performance	2.656 GigaOPS	2.656 GigaOPS
System Peak Performance	3.9 TeraOPS	892 GigaOPS
RAID I/O Bandwidth	6.4 Gbytes/sec	320 Mbytes/sec
RAID Storage	62.5 TBytes	10 Tbytes
I/O Bandwidth to Local Disk	10.5 Gbytes/sec	4.7 Gbytes/sec
Disk I/O Nodes	168	20
System Control Machines	3	1
Network Connections (FDDI, HiPPi-800)	FDDI=6, HIPPI-800=12	0
Number of Node Cabinets	162	86

5.5 The ASCI Blue Mountain 3-Tops System

LANL's Blue Mountain System, mostly delivered between June and November of 1998, consists of 48 Silicon Graphics Origin 2000 shared memory multi-processor computers with 128 250-MHz processors on each machine (total of 6144 processors). These 48 machines have a composite of 1.5 Terabytes of

RAM and 76 Terabytes of fiber channel disk. Jointly, the machines represent a peak capacity of 3.072 TeraOps (3 trillion floating point mathematical operations) per second, with an expected sustained performance of 1 TeraOp per second on the demonstration code, *simplified Piecewise Parabolic Method*. In its full configuration, the Blue Mountain system is one of the most powerful computers installed on-site in the world.

A considerable challenge in the deployment of the ASCI Blue Mountain system is connecting the 48 individual machines into an integrated parallel compute engine. This challenge is currently being met with HIPPI-800 interconnects, which provide high communication bandwidth with great flexibility, without imposing a restricting topology. Each of the 48 machines has 12 HIPPI ports, connected via a 3-dimensional toroidal interconnect using 36 HiPPI-800 16 port switches.

In 1999, the interconnect was reconfigured with HiPPI-6400 32 port switches. HIPPI-6400 is a new ANSI standard for 6.4 gigabit/second data rates, with transport layer error control built into the hardware. Having this error control in the hardware permits the use of more lightweight protocols operating on each SMP node. The goal is to actually bypass the operating system that currently interacts with transfers between user space and the network. Another ANSI specification, Scheduled Transfer, provides the mechanism to remove this interaction and will be the technique used to increase interconnect performance between the 48 individual machines that make up the ASCI Mountain Blue system. For the high performance needed by ASCI applications, the multiple machines must be used together as a single machine. This is primarily accomplished via the Message Passing Interface (MPI) software. MPI uses the OS bypass, a low-level protocol which achieves low latency. The objective is to write portable applications by using MPI but to optimize performance through the use of OS bypass. To further optimize performance, LANL is also writing a library which will use OS bypass without MPI. In performance comparisons, the MPI library has provided a bandwidth of 90 Mbytes/second sustained with 144 microseconds one-way latency, and the OS bypass library has given 140 Mbytes/second bandwidth with 104 microseconds one-way latency.

The Load Sharing Facility (LSF) software from Platform Computing Corporation is used for job scheduling and control on the system. LSF distributes jobs across the 48 machines using features such as queue or machine limits, queue priorities, processor reservation, and job backfilling to provide efficient utilization of the system. In addition to queuing batch jobs, the software allows interactive work spanning multiple machines of the system, a capability which facilitates the development and testing of applications. Archival storage for the ASCI Blue Mountain system is provided by the High Performance Storage System, which is a new-generation storage system for extremely large amounts of data (petabytes) with the ability to access data at very high data rates (tens to hundreds of Mbytes/second). ASCI applications running on the system are expected to generate multigigabyte-sized files. A team of approximately 45 people, involving both LANL employees and SGI personnel, has been assembled on-site for the installation and support of the Blue Mountain 3Tops system. Areas of support include networking, user consultation, documentation, problem tracking, platform integration and system management, distributed resource management, security, applications support, development of parallel tools, data storage, operations, and facilities management.

The ASCI Blue System requires extensive facilities support. It uses:

- 10,000 square feet of floor space,
- 1.6 MWatts of power,
- 530 tons of cooling capability,
- 384 cabinets to house 6144 CPUs,
- 48 cabinets for the meta routers,
- 96 cabinets for the disks,
- 8 cabinets for the 36 HiPPI, switches, and
- ~476 miles of fiber cable.

The successful integration of the Blue Mountain 3Tops system represents one milestone on the road to scaling applications and supporting a fully operational simulation capability for stockpile stewardship.

The ASCI Blue Web site is <http://www.lanl.gov/asci/bluemtn/>. Unfortunately details of the machine, its current status are now treated as "secure" information and are not available to the interested reader.

5.6 Hardware Environment on ASCI White

<http://www.llnl.gov/asci/platforms/white/hardware/index.html>

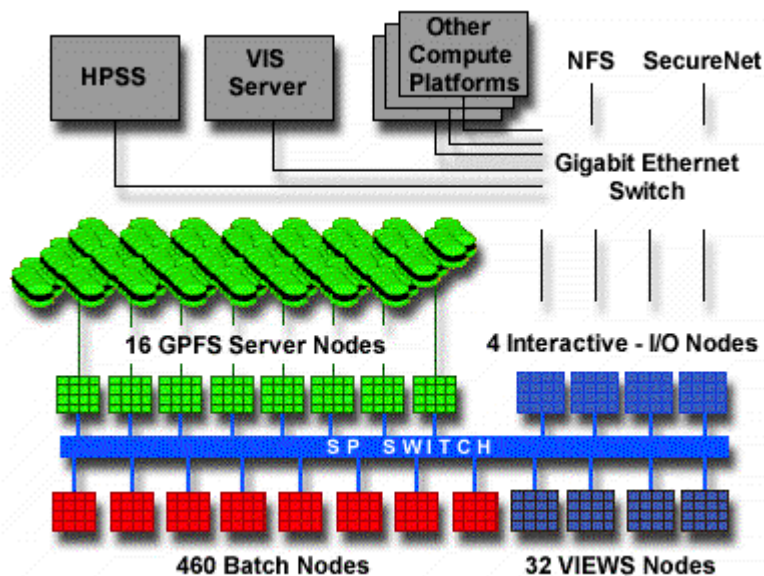
ASCI White is the third step in the DOE's five stage ASCI plan to achieve a 100 TeraOP/s supercomputer system by 2004. The ASCI White hardware environment comprises three IBM RS/6000 SP systems, White, Frost and Ice. White, the largest of these systems, is a 512-node, 16-way symmetric multiprocessor (SMP), classified system with a peak performance of 12+ TeraOP/s. Frost is a 68-node, 16-way SMP unclassified system and Ice is a 28-node, 16-way SMP classified system.

Complementing the IBM SP systems are arrays of external disk storage, GPFS parallel file systems, an HPSS archival storage system and visualization facilities. Specialized high-speed networking forms the backbone and interconnects all components of the ASCI White hardware environment.

The peak performance of the computer is 12.3 teraflops. The processors used are IBM RS6000 SP Power3's which run at 375 MHz. There are 8,192 of these processors in the core compute system. The total amount of RAM is 6Tb. The system is housed in over two hundred cabinets and fills a large room with an area the size of two basket ball courts. Located in a classified area at Lawrence Livermore National Laboratory, ASCI White covers a space the size of two basketball courts and weighs 106 tons. It has more than 160 TB of IBM TotalStorage 7133 Serial Disk System capacity, or enough to hold six times the entire book collection of the Library of Congress.

5.6.1 ASCI White Configuration

The diagram below depicts an overview of the classified ASCI White system configuration. The configuration table that follows provides more detailed information for all ASCI White systems.



ASCI White Configuration Schematic (classified White system)

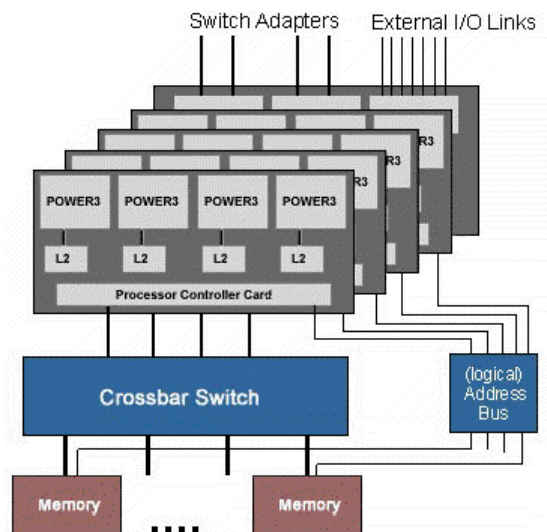
The IBM SP ASCI White systems are composed of many frames, with most frames containing four nodes. All nodes are of IBM's RS/6000 POWER3 symmetric multiprocessor 64-bit architecture. Each node is a stand-alone machine possessing its own memory, operating system, local disk and 16 CPUs. IBM produces several varieties of POWER3 nodes.

ASCI White Configuration Table

Classification	White	Ice	Frost
	Secure/Classified	Secure/Classified	Open/Unclassified
Processor Type	POWER3 375 MHz NH-2 16-way SMP	POWER3 375 MHz NH-2 16-way SMP	POWER3 375 MHz NH-2 16-way SMP
Total Number of Nodes	512	28	68
Total Number of Processors	8192	448	1088
System Peak Speed	12.3 TeraOP/s	0.7 TeraOP/s	1.6 TeraOP/s
Interactive Login Nodes	4	1	1
Debug Nodes	0 / 8 ¹	0	1
Batch/Compute Nodes	492 / 484 ¹	26	63
GPFS Server Nodes	16	2	2
Memory (per node)	16 GB	16 GB	16 GB
Total System Memory	8192 GB	448 GB	1088 GB
Total Local (on-node) Disk	36.3 TB	2.0 TB	4.9 TB
GPFS Global Disk	109 TB	5.7 TB	20.6 TB
SP Switch Type	SP Switch2	SP Switch2	SP Switch2
SP Switch Adapter Type	Colony double-single	Colony double-single	Colony double-single
Node-to-node bandwidth (bi-directional)	2 GB/sec	2 GB/sec	2 GB/sec
System Footprint	9,920 sq/ft; 106 tons		

¹ Current configuration / Eventual configuration

The ASCI White nodes are known as "Nighthawk-2" (NH-2) nodes.



Nighthawk-2 Node Schematic

POWER3 processors are super-scalar, (simultaneous execution of multiple instructions) pipelined, 64-bit RISC chips with two floating-point units and three integer units. They are capable of executing up to

eight instructions per clock cycle and up to four floating-point operations per cycle. All nodes are interconnected by the internal SP switch network. The "Omega" type design of the switch insures that application node-to-node bandwidth is independent of a node's location in the topology. Applications can use either the User Space (US) protocol or Internet Protocol (IP) for MPI task communications. US protocol is the faster and recommended protocol for most applications.

Nighthawk-2 Node Specs

Number of CPUs/Node	16	CPU Clock Speed	375 MHz
Node Peak Performance	19,840 MF/s	Memory	4-32 GB ¹
L1 Data Cache	64 KB	L1 Instruction Cache	32 KB
L1 Cache Line Size	128 bytes	L2 Cache	8 MB per processor
Maximum Disk	946 GB	I/O to Local Disk	40 MB/s

¹ All White nodes have 16 GB memory/node

Additional Information

- [IBM SP Hardware and Software Library](#)
- [Power3: Next Generation 64-bit PowerPC Processor Design](#)
This white paper describes technical details about the POWER3 architecture. Predates ASCI White technology.
- [375 MHz POWER3 SMP Wide Node](#)
This section of the IBM "RS/6000 SP Planning Volume 1, Hardware and Physical Environment" document discusses internal hardware specifications and environmental requirements. Predates ASCI White technology.

5.6.2 General Parallel File System (GPFS)

Each ASCI White system (White, Ice, Frost) is configured with its own parallel file system. The parallel file system hardware is comprised of dedicated server nodes directly attached to massive amounts of disk storage. These server nodes are also directly attached to the system's SP switch for high speed connectivity with the compute/client nodes where parallel jobs actually run. See the [ASCI White Configuration Schematic](#) as an example. The parallel file system software is IBM's General Parallel File System (GPFS), described below.

GPFS Overview : GPFS provides file system services to parallel and serial applications running in the SP environment. To the user, GPFS is designed to "look and feel" like a UNIX file system - all of the usual UNIX file commands (cp, mv, rm, etc) are supported and there are no new commands that a user must become familiar with. Applications run under GPFS as they would under other UNIX file systems.

Under GPFS, individual files are striped as a series of "blocks" distributed across multiple disk drives attached to multiple server nodes. This enables:

- Simultaneous reads/writes by multiple processes to non-overlapping regions of the same file
- Concurrent reads/writes to different files by multiple processes

File replication, data consistency and recoverability are integrated into the GPFS design.

Additional Information

- [A GPFS Primer](#)
Basic information. A good place to start.
- [General Parallel File System \(GPFS\) 1.4 for AIX Architecture and Performance](#)
An IBM white paper that describes the GPFS file system architecture and includes performance figures.

- [IBM's SP Library](#)
Technical documentation for GPFS. Mostly oriented to system administrators.

5.6.3 HPSS Archival Storage System

Both the Secure Computing Facility (SCF) and Open Computing Facility (OCF) provide researchers with an HPSS archival storage system interconnected by state-of-the-art, standards-based network technology.

For current HPSS hardware configuration information see <http://www.llnl.gov/icc/lc/dsg/systems.html>. These pages also provide links to other HPSS related LC projects and information.

5.6.4 Visualization

ASCI White users have access to Livermore Computing's visualization resources which include:

- Classified and unclassified "big data" visualization servers
- Assessment theaters
- Video production
- Graphics software
- Graphics consulting
- Graphics applications development

Visualization Servers

Server	Classification	Architecture	Processors # / Type	Memory (GB)	Disk (TB)	Graphics	Interconnect
Riptide	Unclassified	SGI Onyx-2	48 / R10000 250 MHz	36.6	10.5	8 IR2 pipes	1.6 GB/sec Numalink
Tidalwave	Classified	SGI Onyx-2	64 / R12000 300 MHz	24	9.2	16 IR2 pipes	1.6 GB/sec Numalink
Edgewater	Classified	SGI Onyx-2	40 / R12000 300 MHz	17.6	9.3	10 IR2 pipes	1.6 GB/sec Numalink
Whitecap	Classified	SGI Onyx 3800	96 / R12000 400 MHz	96	5	4 IR3 pipes	3.2 GB/sec Numalink

Assessment Theaters

Location	Classification	Server Connection	Displays
Bldg. 111	Classified	Edgewater	4x2 tiled powerwall 2x2 flat panel immersive (planned)
Bldg. 132	Classified	Tidalwave / Edgewater	5x3 tiled powerwall
Bldg. 451	Unclassified	Riptide	3x2 tiled powerwall

ASCI White VIEWS Pool

ASCI VIEWS program users have a dedicated, 32 node "views" pool available for interactive work on ASCI White in the classified environment.

Software, Services and Additional Information

- <http://www.llnl.gov/icc/sdd/img/>
Starting point for information about visualization and graphics resources and services.
- http://www.llnl.gov/icc/sdd/img/graphics_sw.shtml
Complete list of LC graphics software including versions and descriptions of each software package.

- <http://www.llnl.gov/computing/tutorials/graphics/>
Detailed information about the services, hardware and software provided by LC's Graphics group.
- http://www.llnl.gov/ascii/views_trilab/
Visual Interactive Environment for Weapons Simulation (VIEWS) program information including a program overview, contacts and project descriptions.
- <http://www.llnl.gov/str/Quinn.html>
Science & Technology Review article by Terri Quinn describing the various facets of the ASCII VIEWS program.

5.6.5 Programming Model

The ASCII White system is designed to support the mixed-mode parallel programming paradigm of clustered distributed memory with SMP shared memory. MPI is typically used for node-to-node distributed memory communications over the SP high-speed internal switch network. OpenMP or POSIX threads are used for on-node shared memory task communication. Uniprocessor, distributed memory only, and shared memory only applications are also supported in this environment.

A likely programming model for ASCII White is four MPI tasks per node, with four threads per MPI task. This model exploits both the number of CPUs per node and each node's switch adapter bandwidth. Job limits are 4,096 MPI tasks for US (high speed) protocol and 8,192 MPI tasks for IP (lower speed). The current MPI library is limited to 32-bit address space for message passing. A 64-bit MPI implementation is expected later.

5.7 Tri-Lab ASCII Whitecap

Whitecap is a shared tri-lab resource that is reserved for visualization work. Most common utilities and applications are available.

SGI Onyx 3800 Hardware and OS	
Capability	Status
Architecture	Symmetric Multiprocessor with distributed shared memory having nonblocking cache coherency between all nodes.
CPUs	Ninety-six 400-MHz R12000 processors on 48 nodes are available. Theoretical peak performance is 76.8 GFlops.
Memory	96 GB total RAM, installed as 1 GB with each CPU, are available.
IR graphic pipes	Four Infinite Reality 3 graphic pipes are currently available.
Infrastructure (modules, interconnects, etc.)	Some initial problems with "G bricks" and related hardware have been resolved.
Disk cache	6 TB of fibre channel disk arrays attached and fully functional. Mounted as /fc/temp1 and /fc/tmp2. Delivers about 241 MB/s for writes and 622 MB/s for reads. Greater write performance (~x2) is possible, but some reliability issues need to be investigated. Disk is structured as RAID-0 over RAID-3.
Operating system	IRIX v6.5.13f is currently installed.

Whitecap Software Applications

Capability	Status
EnSight	EnSight versions 7.4.1, 7.4, and 7.3.2 are available and tested within LLNL. Utility table for validating format of files. Needs final user testing from SNL/CA, SNL/NM, and
MeshTV	Version 4.3.1 is installed in /usr/gapps and tested successfully for serial and parallel modes.
SpeedSho	Available.
TeraScale	Version 1.30 is installed in /usr/gapps and tested.
VisIt	Version 1.0.6 is installed in /usr/gapps and tested.
Xmovie	Xmovie version 1.41 is installed in /usr/gapps and tested.

5.8 ASCI Q

ASCI Q, named to follow LANL's tradition of alphabetical names for computers, will have 11,968 processors, 12 terabytes of memory and 600 terabytes of disk storage.

ASCI Q is the [NNSA's](#) fifth machine in the sequence of high-performance computers, operating systems, and software applications as a part of its Stockpile Stewardship Program, with a goal of reaching 100 teraOPS in 2004.

The Los Alamos ASCI supercomputer, Q, is now being installed in its new facility, the Nicholas C. Metropolis Center for Modelling and Simulation, dedicated in May 2002.

5.9 ASCI Purple

At Supercomputing 2002 (Baltimore, November 19,2002) Energy Secretary Spencer Abraham announced that The Department of Energy (DOE) had awarded IBM a contract valued at \$216 to \$267 million to build the two fastest supercomputers in the world with a combined peak speed of up to 467 Teraflops. These two systems will have more combined processing power than the combined power of all 500 machines on the recently announced TOP500 List of Supercomputers.

The first system - called ASCI Purple - will offer the DOE the world's first supercomputer capable of up to 100 teraflops, more than twice as fast as the most powerful computer in existence today. ASCI Purple will consist of a massive cluster of POWER-based IBM™ systems and IBM storage systems, and represents a fifth-generation system under the ASCI Program.

The second supercomputer, a research machine called Blue Gene/L, will employ advanced IBM semiconductor and system technologies based on new architectures being developed in the ongoing partnership between IBM and the DOE for the government's ASCI Program. When completed, Blue Gene/L will have a theoretical peak performance of up to 367 teraflops with 130,000 processors running Linux. It will have the capability to process data at a rate of one terabit per second, equivalent to the data transmitted by ten thousand weather satellites. The supercomputer will be used by the three NNSA laboratories (Los Alamos, Sandia and Lawrence Livermore) and the ASCI University Alliance collaborators as well as other DOE laboratories in the future.

Blue Gene/L will be used to develop and run a broad suite of scientific applications including the simulation of very complex physical phenomena of national interest, such as turbulence, prediction of material properties, and the behavior of high explosives.

In addition to ASCI Purple, IBM also delivered Lawrence Livermore National Lab's (LLNL) previous most powerful supercomputers - ASCI White, unveiled in August 2001, and ASCI Blue Pacific, unveiled in October 1998. ASCI Purple will be delivered in stages with the first IBM systems arriving next year. The new machine will be installed in a dedicated building known as the Terascale Simulation Facility, currently under construction at LLNL in California.

The 100 teraflop ASCI Purple system will be powered by 12,544 POWER5 microprocessors, IBM's next generation microprocessor. These processors will be contained in 196 individual computers with a total memory bandwidth of 156,000 GBs. All of the computers are interconnected via a super-fast data highway with a total interconnect bandwidth of 12,500 GB. ASCI Purple will run IBM's [AIX 5L](#) operating system. The system will also contain 50 terabytes of memory (50 trillion units), which is 400,000 times more capacity than the average desktop PC and two petabytes of disk storage (two quadrillion units).

6. UNIVERSITY PARTNERSHIPS

6.1 Academic Strategic Alliances Program Overview

The purpose of this program is to engage the best minds in the U.S. academic community to help accelerate the emergence of new unclassified simulation science and methodology and associated supporting technology for high-performance computer modelling simulation. These alliances will support the development and credible validation of this simulation capability. ASCI will also work with the larger computing community to develop and apply commercially acceptable standards. ASCI plans to initiate exchange programs to bring top researchers directly into the project while allowing laboratory personnel to expand their experience base in external projects. An important step toward developing the next generation of scientists need for the national security programs at the Defence Programs labs. Research projects are implemented on three levels:

- Level One Strategic Alliances establishes five major centres engaged in long-term, large-scale, unclassified, integrated multidisciplinary simulation and supporting science and computational mathematics representing ASCI-class problems. These are
 - the Center for Simulating Dynamic Response of Materials at California Institute of Technology
 - the Center for Integrated Turbulence Simulation at Stanford University
 - the Center for Astrophysics Flash Phenomena at the University of Chicago
 - the Center for Simulation of Advanced Rockets at the University of Illinois, Urbana-Champaign
 - the Center for Simulation of Accident and Fire Environments at the University of Utah.
- Level Two Strategic Investigations establishes smaller discipline-oriented projects working in computer science and computational mathematics areas identified as critical to ASCI success.

Level Three Individual Collaborations establish focused projects initiated by individual ASCI researchers working on near-term ASCI-related problems.

[Academic Strategic Alliances Program Homepage\(ASAP\)](#)

6.2 ASCI Institutes

In addition to the three-level Alliances Program, ASCI's academic collaborations added a new component in FY2000 - The ASCI Institutes. The charter of the ASCI Institutes at the three NNSA national security laboratories is to create an environment for collaboration with academia on research topics in computer science, computational mathematics, and scientific computing that are relevant to the Stockpile Stewardship Program. These collaborations are conducted through a variety of mechanisms, ranging from one-day seminars to multi-month sabbaticals at the laboratories.

Each of the three Institutes has different topics of emphasis, depending on laboratory needs; however, they all co-ordinate and leverage their activities to ensure maximum benefit to the ASCI program.

Hiring qualified and experienced computer and computational scientists is extremely challenging in today's job market. One of the objectives of this effort is to enhance the laboratories' ability to attract academics to the laboratories.