ASC™

Advanced Simulation and Computing

Computing Strategy

National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

LAWRENCE LIVERMORE NATIONAL LABORATORY

Los Alamos NATIONAL LABORATORY

Sandia National Laboratories

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

SAND 2013-3951P
Issued by Sandia National Laboratories for NNSA's
Office of Advanced Simulation & Computing, NA-114

For more information, contact: Thuc Hoang at
Thuc.Hoang@nnsa.doe.gov

Advanced Simulation and Computing
Computing Strategy

Advanced Simulation and
Computing Program, NA-114

Authors:

James A. Ang
Paul J. Henning
Thuc T. Hoang
Rob Neely

May 2013

A Publication of the Office of Advanced Simulation & Computing,
NNSA Defense Programs

ACKNOWLEDGMENTS

This Computing Strategy is an ASC community product with numerous authors, reviewers and editors without whose contributions this document would not be so comprehensive in details. Invaluable contributions to the overall document were made by Bert Still (LLNL), Terri Quinn (LLNL), Becky Springmeyer (LLNL), Kim Cupps (LLNL), Bill Archer (LANL), Mark Anderson (LANL), Sue Kelly (SNL), Fred Johnson (SAIC), Charlie Slocomb (SAIC), Tina Macaluso (SAIC), and Emily Simpson (SAIC). Superb graphics and publication expertise were provided by Karla Weaver (SNL), David Rogers (SNL), and Douglas Prout (SNL). Most importantly, I would like to especially acknowledge and thank Paul Henning (LANL), Jim Ang (SNL), and Rob Neely (LLNL) for being my co-authors and untiring editors of the document as we went through numerous drafts.

Thuc T. Hoang
NA-114

CONTENTS

FOREWORD

The essential task of the Stockpile Stewardship program, and by extension ASC, is to ensure the U.S. will never have to resort to nuclear testing to guarantee a safe, secure and effective nuclear deterrent. Our ultimate success depends on our ability to provide the next generation of stockpile stewards with simulation tools which serve to enable broad and deep knowledge of the individual processes involved in a nuclear weapons explosion, as well as a comprehensive understanding of the complex interactions among these processes. The simulation goal is to inform critical stockpile stewardship decisions through detailed behavior prediction, uncertainty quantification, and validation through comparison with comprehensive experimental results and past tests.

Over the past fifteen years, progress in transistor technology, characterized by Moore's law and Dennard scaling, has resulted in phenomenal growth in the size and computational capability of ASC high performance computing systems - from the first terascale system, ASCI Red, in 1997 to today's petascale systems such as Sequoia, Cielo, and Roadrunner. This 10,000-fold increase in system performance has been harnessed by NNSA laboratory mathematicians, computer scientists and domain science experts to incorporate new scientific understanding and perform increasingly accurate predictive simulations of ever more complex and challenging stockpile problems.

Expected challenges of stockpile stewardship over the next decade drive a continued need to advance our science-based simulation capabilities. The process of moving from petascale to exascale will be fundamentally different due to technology limitations that are forcing change in overall system design and usage. The success of recent past will not extrapolate so easily into the future. Most prominently, power and data movement have emerged as critical factors affecting all aspects of future systems and codes. Planning has begun for developing new generations of computational resources and simulation codes in order to proactively address these constraints and develop the necessary agility to adapt to looming changes in the next generation of high performance computing platforms.

This new ASC Computing Strategy discusses these issues in depth and describes the initial steps by ASC in moving to a new era of challenge and opportunity for predictive simulation. Partnerships with industry, DOE sister organizations, and other federal agencies are a key part of our approach. Agility and adaptability are essential to ensure that future generations of ASC simulation capabilities and resources provide cost-effective solutions that will continue to underwrite our nation's resolve to forgo nuclear testing.

Robert E. Meisner, Director

Office of Advanced Simulation and Computing and Institutional Research and Development

Defense Programs

National Nuclear Security Administration

Washington, DC

EXECUTIVE SUMMARY

This ASC Computing Strategy will guide future ASC acquisitions of high-end computational platforms and the development of supporting software infrastructure to ensure the timely, cost-effective availability of simulation and computing resources needed to maintain a safe, secure and effective stockpile without underground testing. This document presents key principles that will guide the transition to future generations of computer architectures, applications and infrastructure consistent with critical NNSA mission requirements and available resources. Six important points below provide a robust framework for our future decisions.

1. Support the NNSA Nuclear Security Enterprise (NSE) Strategic Plan:
   - Provide an agile, efficient, and integrated enterprise responsive to mission needs and strategic goals.
   - Synchronize planning and execution with evolving stakeholder priorities as documented in the Predictive Capability Framework (PCF) and Directed Stockpile Work (DSW).
2. Align a balanced acquisition strategy with mission needs:
   - Acquire Commodity and Advanced Technology systems for cost-effective services in support of stockpile requirements.
   - Collaborate with industrial partners and other experts to foster the timely availability of future computing systems that are readily useable by application developers.
3. Ensure user productivity through best-value system acquisition and a persistent computing environment:
   - Deliver production-level Commodity and Advanced Technology systems to users as rapidly and cost effectively as possible.
   - Promote the availability and use of robust programming models, development environments and performance tools responsive to user requirements.
4. Execute a comprehensive and enduring approach to ensure a smooth transition for the ASC code base to the future computing paradigm:
   - Maintain effective working relationships with other DOE and federal partners, industry, and academia to overcome critical technology challenges.
   - Promote and coordinate teams to achieve a comprehensive understanding of the impact of future architectures on ASC production codes and subsequent modernization thereof.
5. Foster complex-wide integration and collaboration:
   - Create and evolve standards for hardware and software through coordinated tri-lab procurements of Commodity Technology systems and deployment of a robust, common software environment.
   - Operate production-level Advanced Technology platforms as national user facilities that employ premium resources for NSE-wide priorities.
6. Emphasize importance of a vital workforce and supporting infrastructure:
   - Recruit technical talent to keep NNSA at the forefront of high performance computing (HPC) and simulation technologies.
   - Develop supporting infrastructure to ensure that ASC platforms are optimally usable and productive.

INTRODUCTION

Established in 1996, the Advanced Simulation and Computing (ASC) Program continues to be a cornerstone of the National Nuclear Security Administration (NNSA) Stockpile Stewardship Program (SSP). It provides simulation capabilities and computational resources to support annual stockpile assessment and certification; studies advanced nuclear weapons manufacturing processes; analyzes accident scenarios and weapons aging; and provides tools to enable stockpile Life Extension Programs (LEPs) and resolve Significant Finding Investigations (SFIs). This requires a balanced program, including technical staff, hardware, simulation software, and computer science solutions.

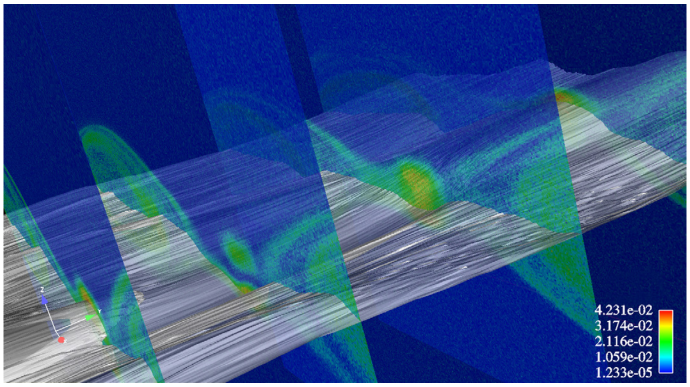
ASC must continue to meet three objectives in order to provide necessary simulation and computing services to the NNSA weapons program:

- Objective 1. *Robust Tools*. Develop robust models, codes, and computational techniques to support stockpile needs such as SFIs, LEPs, annual assessments, as well as evolving future requirements.
- Objective 2. *Prediction through Simulation*. Deliver verified and validated physics and engineering codes to 1) enable simulations of nuclear weapons performance in a variety of operational environments and physical regimes, and 2) enable risk-informed decisions about the performance, safety, and reliability of the stockpile.
- Objective 3. *Balanced Operational Infrastructure*. Implement a balanced computing strategy of platform acquisition and operational infrastructure to meet DSW and SSP needs for production and advanced simulation capabilities.

The *2013 ASC Computing Strateg*y concentrates on providing the computational infrastructure required by present and future ASC platform users within the budgetary constraints of the base ASC program. The computing infrastructure described in this strategy is a complex environment that integrates many types of hardware and software products. Whenever possible, ASC utilizes products from commercial vendors and the open source software community. However, when the required technology is not available from these sources, ASC invests in internal research and development (R&D) and vendors' non-recurring engineering (NRE) activities to close the gaps.

The NNSA SSP needs and fiscal constraints dictate the tempo and means by which ASC acquires computational systems, whereas Moore's Law and industry's response to the end of Dennard scaling affect the cadence of commercial progress. Moore's Law is the statement that the number of transistors on integrated circuits, and, by corollary, system performance, doubles about every two years. Dennard scaling is the ability to drive smaller transistors at a reduced voltage and higher clock speed; it magnified the performance impact of Moore's Law. Dennard scaling ended in 2003 when the commodity processor industry introduced the first dual core processor. Consequently, commercial computing technology offerings must be adapted to prevent frequent disruptions to existing Engineering and Physics Integrated Code (EPIC) applications. These applications, however, cannot be shielded indefinitely from upcoming radical changes in computing technology. ASC must continue to identify and invest in promising new computing technology that supports HPC simulation while providing a more step-wise path to exploiting its full potential.

In addition to strategic investments, ASC will continue to identify areas to consolidate lab-specific activities into a common tri-lab environment to better maximize the value of its investments and provide enhanced fail-over capabilities should a disaster or natural emergency occur to any NNSA lab.



Magnetic islands and drift-kink instability observed in 3D large-scale fully kinetic VPIC simulations run on the Roadrunner base system of magnetic reconnection in electron-positron plasma. (LANL)

PLATFORMS

A computing platform is an integrated system of hardware and software that ultimately provides an environment in which a weapon analyst or designer can run simulations and analyze results. It is not just a computer; it is a host of hardware and software components, often developed independently from one another (e.g., compute nodes, networks, file systems, long-term storage, operating systems, compilers, and numerical libraries). The HPC system vendors do not provide all the necessary system software and user tools on the acquired systems that the ASC users need for their diverse application code use. Therefore, the ASC program is required to expend additional resources to develop in-house (at NNSA National Laboratories) or acquire such supplementary software packages either through third party vendors or academia.

Understanding the Computing Workload

The primary users of ASC platforms are designers, analysts, and computational scientists in the NNSA weapons program. Each of these users is attempting to address a stockpile issue using a set of computer simulations as a part of their investigation. A single simulation, on its own, provides little insight because it approximates physical reality that is sensitive to the modeling parameters specified for the simulation. To understand how the results change as the simulation model parameters are changed, users have to examine or analyze a series or ensemble of simulations. Two major categories of changes are fidelity and uncertainty.

Fidelity refers to how accurately a simulation calculates physics, length, time, or other quantities. For example, a coarse fidelity simulation of a car travelling from Chicago to Manhattan might tell you how long it took, the number of tanks of fuel consumed, and perhaps arrival time at each refueling stop. A high fidelity simulation would provide a more continuous profile of the vehicle's velocity and fuel usage patterns by tracking arrival time at each mile marker along the path and correlating those times to local events such as traffic patterns, road conditions, and weather. The difference between these two levels of fidelity governs the user's ability to answer questions like "Is it possible for a car to make the trip from Chicago to Manhattan?" versus "Will this particular car, starting on this day, make the trip safely?" At some point, increasing the fidelity of the simulation no longer has appreciable effects on the outcome because the simulation is "converged." Before running a simulation, it is especially difficult to know how much fidelity is required to achieve convergence when investigating complex multi-physics simulations. Each additional level of fidelity requires greater computational resources.

Even after a simulation has reached a sufficiently converged solution, there is still the problem of demonstrating confidence in the results. Physical systems and their simulators are full of uncertainties, making it impossible to make predictions with confidence unless these uncertainties are understood. Returning to the driving example above, questions might include "How well do we know the tire pressures at all times of the trip? How good are the traffic and weather models that we used? Which of these things really matter in the end?" From a computational standpoint, this uncertainty quantification (UQ) increases resource demands by either running ensembles of simulations with varying parameters or running much more complicated simulators that can propagate uncertainties within a single simulation.

Often users generally begin by performing a number of coarse fidelity simulations, and then progressively refine the simulations for higher accuracy and confidence as needed. This creates a wide mix of the quantity and size of the simulation jobs performed, consisting of many smaller jobs and fewer larger jobs, which directly drives the ASC platform acquisition strategy.

Platform Classes

In the pioneering days of the Accelerated Strategic Computing Initiative (ASCI) (1996-2003), all NNSA platforms were leading edge systems. There was no commodity marketplace for high performance computing (HPC) systems. Each new system required 12-24 months to stabilize and the EPIC codes had to be ported to each new architecture. ASCI bought these systems to run challenging, cutting-edge three-dimensional (3D) simulations that were driving game-changing advances in weapons science. However, as pointed out in the 2003 JASON ASCI Requirements review, the workload was increasingly comprised of smaller simulations that employed coarsely resolved or 2D calculations designed to provide sufficient insight before investing in a full-system run. In addition, early ASCI investments in commodity networking technologies were enabling HPC clusters to come to market at lower price points. Accordingly, the *ASC 2007 Platform Strategy* identified three classes of computing platforms:

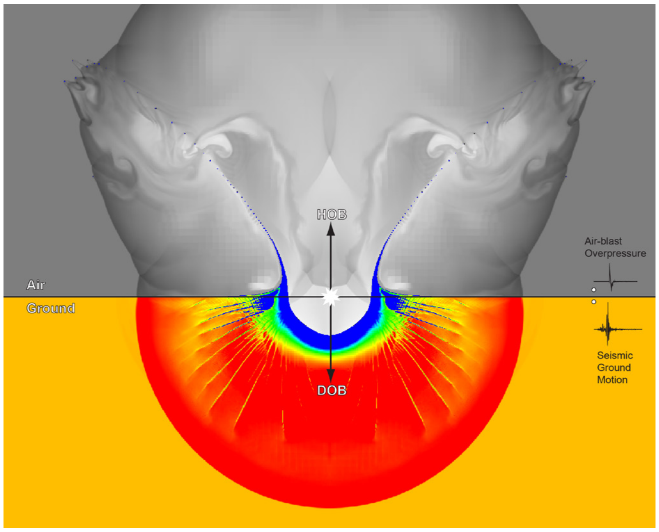
- *Capacity:* low-risk, cost-effective systems intended for more modest parallel computing challenges;
- *Capability:* general-purpose, first-of-a-kind systems involving modest technological risk, intended for the most challenging problems; and
- *Advanced Architectures*: special-purpose, higher-risk, first-of-a-kind systems for exploring new technology and code development strategies needed to tackle currently intractable problems.

To meet the program's Capability and Advanced Architecture platform requirements, ASC was not able to procure commodity HPC systems because they were deficient in key performance characteristics such as interconnection network bandwidth, processing speed, or operating system scalability.

First-of-a-kind platform acquisition is not a simple, one-time procurement. Before the platform can be delivered, there are often NRE costs associated with ensuring that the new technology can be brought to market in the required timeframe. After the acquisition, there are two types of additional costs: operational and opportunity. Operational costs include expenditures for the facilities (buildings, cooling, and other infrastructure), power, and support staff. Opportunity costs are due to the downtime required to integrate a supercomputer into an existing computing environment (e.g. networks and storage systems) and include the period of time between initial system start up and its general availability to the user community for routine, production-level operations.

In recent years, platform vendors and hardware integrators have been embracing the concepts of multi-core processors, large-scale networking, and new computer architectures that were novel five years ago. At the same time, market forces have shifted vendor attention away from large HPC systems to cloud computing, data analytics, and mobile computing, especially in the consumer space. The HPC market constitutes only a tiny fraction of the primary computing market, greatly limiting the influence of the HPC community on the direction of computing technology. ASC now faces a set of formidable challenges in platform acquisition:

- Ensuring that we can acquire systems that meet NNSA mission needs, now and in the future;
- Addressing these mission needs within the cost, power, and reliability constraints of the ASC program;
- Adapting to disruptive architecture technology supported by broader markets, while preserving the investment in our validated application code base.



Monitoring of nuclear tests requires predictions of complex seismic wave propagation to differentiate earthquakes and explosions. (LLNL)

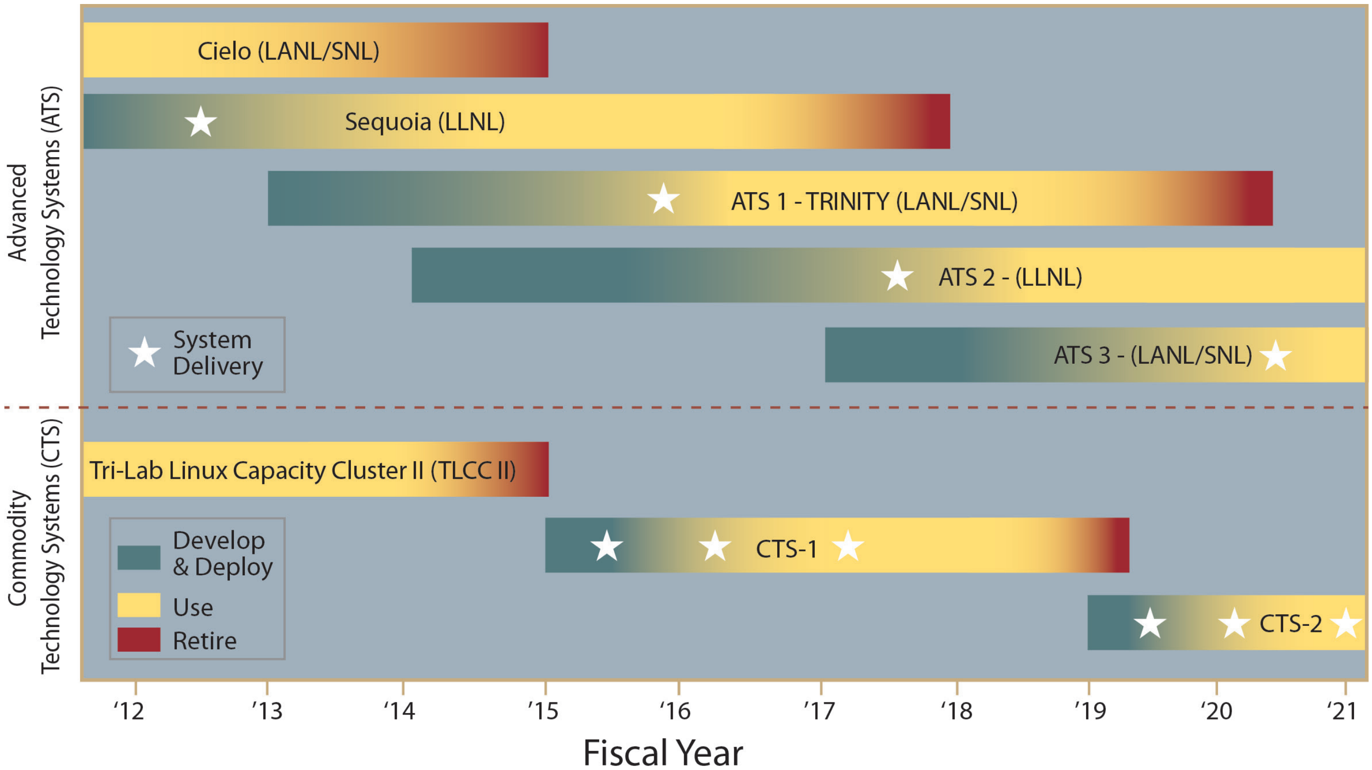


Figure 1. ASC Platform Acquisition Plan for FY12-21. The first of the AT systems will be called Trinity, and it is expected to be in operation by 2016. AT systems will alternate between the New Mexico and California sites and be deployed every five years at either site. The cadence for new generations of CT systems will be every four years.

The new ASC platform acquisition plan (see **Figure 1** above) includes two computing platform classes: *Commodity Technology* (CT) systems and *Advanced Technology* (AT) systems. The CT systems provide computing power to a large percentage of the design and analysis community by leveraging predominantly commodity hardware and software. The goal of these systems is to minimize software changes and maximize availability to end-users. In contrast, the AT systems are the vanguards of the HPC platform market and incorporate features that, if successful, will become future commodity technologies. These large, first-of-a-kind systems will require application software modifications in order to take full advantage of exceptional capabilities offered by new technology.

Two significant developments have made it logical for the program to move to two platform classes. First, operating system and supporting system software improvements now allow CT systems to support many application jobs that previously required an AT system. The second development has been rapid and broad industry support for advanced architectures with complex, hierarchical compute nodes that utilize commodity central processing units (CPUs) to coordinate and drive accelerators (e.g., general purpose-graphical processing units or GP-GPUs), co-processors (e.g., Xeon Phi), and other, often heterogeneous, subsystems, pioneered by Roadrunner. ASC recognizes that while these advanced architecture systems currently address the energy challenges for some of the ASC single physics and materials science applications, they are not likely a good match for the program's EPIC applications. On the other hand, the fact that the larger global community that is developing HPC applications for these advanced architectures makes it a viable option for CT systems. To address the performance, energy, and resilience challenges for EPIC applications, ASC will need to direct its NRE technology investments on the development of new AT systems to support these types of applications. While there will certainly be more ideas for NRE technology development than ASC can afford to fund, a key consideration will be how to invest in advanced technologies that help preserve the value of the ASC application code base.

FastForward Investments

The ASC program has a long history of making strategic investments that foster the development of advanced computing technologies necessary to improve the scalability and performance of HPC platforms in general, and for ASC's computing needs specifically. From the beginning of ASC, these investments in advanced technology development and NRE projects supported both hardware and software efforts to accelerate the availability and increase the scalability of technologies that the U.S. computer industry would not otherwise provide. In mid-FY12, ASC resurrected a vendor partnership effort that was previously known as PathForward but is now called FastForward. ASC will apply investments in areas unique to the HPC community, provide early access to promising new technologies to allow the ASC codes to adapt, and provide ASC computer and computational scientists a direct pathway to have co-design dialogues with the HPC industry collaborators.

Joint investments by the ASC and Department of Energy (DOE) Office of Science Advanced Scientific Computing Research (ASCR) programs in FastForward will support R&D efforts by the national labs and vendors to produce new software and hardware technologies. Examples may include technology to reduce data motion costs, to provide increased memory capacity within a constrained power budget, or to improve overall system resilience. Investments may also include small prototype demonstration systems. **Figure 2** below illustrates the relationship between FastForward R&D investments and the system acquisitions within the new ASC Computing Strategy. AT systems will take advantage of technology developed in successful FastForward projects, and present a natural opportunity for integration of FastForward-developed technologies into production computing platforms. Conversely, CT systems are less likely to be early adopters of FastForward technology.

Development of co-design capabilities within the CSSE (Computational Systems and Software Environment) and FOUS (Facility Operations and User Support) sub-programs targeted at supporting the ASC co-design project will likely identify multiple options for new capabilities that can and should be implemented in future computing technologies. We intend to quantify the beneficial impact of adopting these changes, while reducing the barriers and risks to industry adoption of innovative advances. FastForward investments are a key element of this strategy and the active interaction between the FastForward partners and the ASC co-design project will enable laboratory scientists and engineers to have early performance data and gain insight into impacts of the advanced architectures on the ASC codes. The ASC co-design approach is detailed further on page 10.

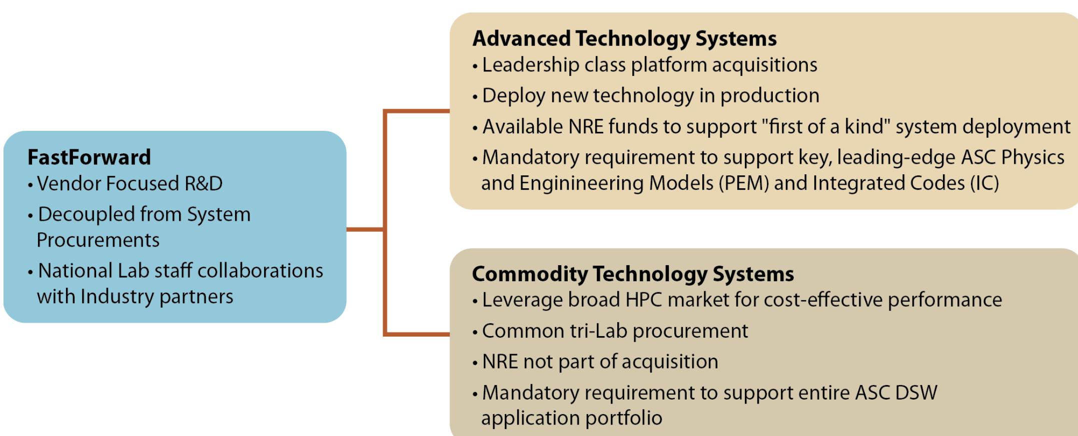


Figure 2. Interfaces of the ASC platform components. Early HPC hardware and software technology investments are made through FastForward projects. Promising FastForward technologies are further developed with non-recurring engineering (NRE) funding that is integrated into acquisition of first-of-a-kind AT systems. Successful AT system technologies may appear in future CT systems. However, some FastForward technologies could directly transition into CT systems (e.g., new memory or storage technology advances).

FastForward investments will help accelerate and implement the innovations ASC expects to identify and influence through co-design, vendor feedback, and early experiences with the research community.

Transitioning to the New Platform Classes

With the exception of the most demanding calculations, CT systems will accommodate the needs of much of the designer and analyst community. This approach minimizes both acquisition cost and the effect of technological change on mission critical applications. Furthermore, CT systems may consist of a mix of system sizes that matches the workload, e.g. many small systems and a few large systems.

The primary objective of AT systems is to meet the mission needs of ASC that are beyond the current performance capabilities of CT systems. A secondary objective is to bring beneficial, new technologies to market so they can be incorporated in future CT systems. The economics of introducing new technology typically results in AT systems being at large scales. However, specific system size is based on mission needs rather than simple performance targets (such as peak FLOPs) that do not directly indicate performance of the ASC weapons production codes. From the perspective of EPIC applications, the sequence of AT and CT system acquisitions over time should align with a managed progression of technology, which minimizes the amount of weapons application software that must be rewritten to achieve performance goals.

Additional Considerations

Each platform class comes with a unique set of expectations that guide the acquisition process. CT systems are expected to:

1. Meet mission requirements for simulations that do not require the scale of AT systems.
2. Emphasize full-system stability and availability, including minimal delays between system delivery and general availability.
3. Provide a common tri-lab computing environment.
4. Maximize value in acquisition, operation, and code development costs.

Given these expectations, certain platform characteristics arise naturally. CT systems will predominantly consist of commodity hardware and software, employing technology that EPIC codes can utilize without necessitating disruptive code adaptations. However, specialization in situations where a small investment brings a large benefit to the user community will be considered for further evaluation and potential procurement. CT systems are designed and provisioned to maximize availability to the end-user and are managed and operated as a production-level resource. Disruptions due to system maintenance activities will be kept to a minimum. If problems should arise during system maintenance activities, the entire platform will not be rendered unavailable and will be returned to its production state as quickly as possible while the problems are resolved.

The AT systems are expected to:

1. Meet mission requirements for the most challenging engineering and physics simulations required for predictive capability – supporting both increased fidelity and quantification of uncertainties.
2. Incorporate new computing technology that provides benefits (e.g., scale, speed, stability, energy efficiency, programmability, manageability, etc.) which are not presently available in existing commodity offerings.
3. Cultivate and develop path-finding technologies that could appear in future CT systems.

The benefit of a new technology will be measured from the perspective of the ASC end users. This means that a faster processor or esoteric architecture may not be the correct approach. Other platform features may be more beneficial to ASC users, such as:

1. Better I/O through traditional interfaces
2. Increased memory capacity and bandwidth
3. Improved network characteristics
4. Higher system stability and availability
5. Increased programmer productivity

AT systems are also intended to be generally available to end users as an ASC production-level resource. However, due to the more advanced technologies involved, a larger percentage of the available time may be allocated to system software debugging at scale. It is anticipated that once an AT system becomes generally available to the end user community, time on the machine will be reserved through a lightweight proposal process that allows programmatic tri-lab priority to dictate usage.

COMPUTING ENVIRONMENT ECOSYSTEM

It is important to consider the environment in which future AT and CT systems will operate. To be successful, the platforms must integrate within an ecosystem of workflow processes, user tools suite, storage systems, and communication devices. Additionally, the facilities infrastructure and operational support services are critical to successful deployment.

User Environment

While variations exist, a typical end user workflow requires the components shown in **Figure 3** below: meshing tools and set-up software, ensembles of runs and a smaller set of large-scale runs, data analysis, visualization, and data storage. Ultimately, the ASC user environment needs to be persistent and consistent, to the extent possible, among all ASC currently deployed and future platforms as different-sized problem sets are required to run interchangeably on either CT or AT systems.

Software Environment

To date, ASC applications predominately rely on bulk-synchronous coarse grain parallelism using message passing interface (MPI). However, a few application development teams have introduced finer grain, on-node parallelism into code kernels using threading. This trend of moving away from an MPI-everywhere programming model will likely continue and will increasingly introduce additional complexities such as accelerators, co-processors, processor-in-memory, deep memory hierarchies, user-controlled resilience and power management. Since the precise vector away from the MPI paradigm remains unknown, the software stack must continue to enable the MPI bulk-synchronous model while creating an environment with with rich support for software development that exploits the multi or many-accelerator core processors that hardware vendors are releasing.

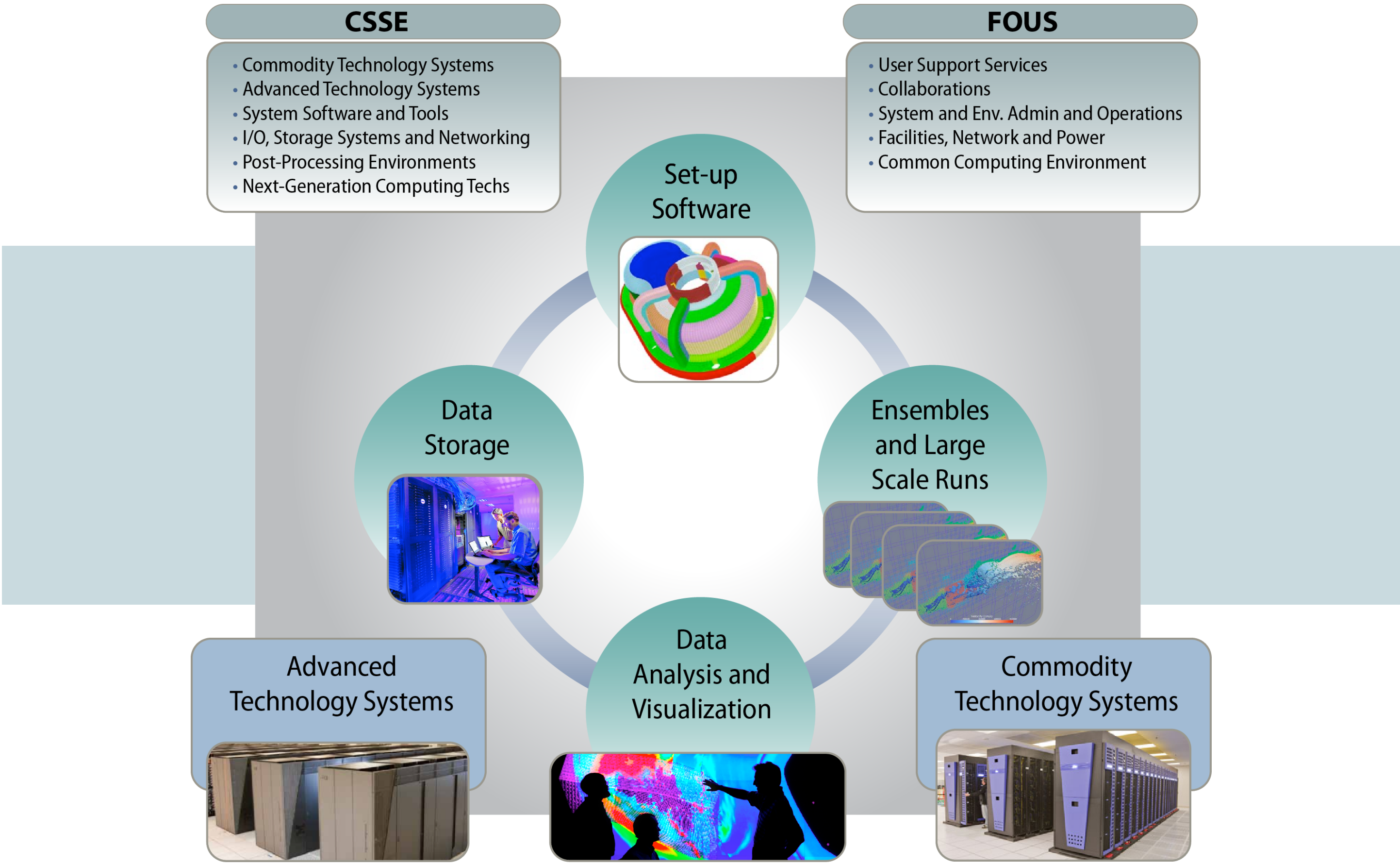
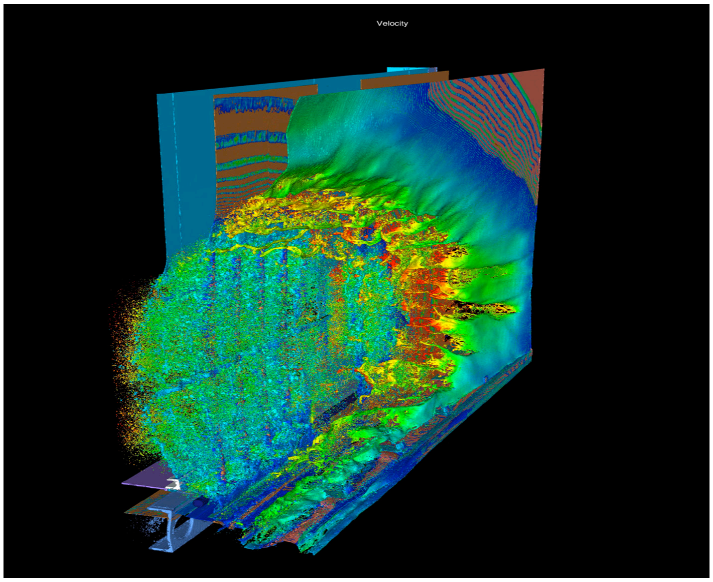


Figure 3. ASC User Environment: The ASC Computing Environment is delivered by the CSSE and FOUS sub-programs which are responsible for the deployment of the platforms and the necessary provision of user support services, such as tools and networking, data analysis and storage, facility operations and user hotline, etc.



A 32,768 CTH simulation helps designers understand the response structures under severe blast loading conditions so that the robustness of these structures may be improved. A structure composed of sheet metal is loaded by an explosive blast. The plot gives the velocities of various components of the structure as they are torn apart. Accurately resolving these thin parts required a very fine mesh resolution which can presently be achieved using only very large scale computing platforms. (SNL)

Programming Environment – The languages and libraries that implement programming models must adapt to the growing levels of on-chip parallelism and increasing depth of memory hierarchies. Abstraction layers can hide the underlying details of the HPC architectures, making the "program once, run everywhere" requirement easier to fulfill. To achieve maximum utilization of the new architectures, performance and correctness tools are needed as developers enter these new programming regimes.

Data Analysis and Visualization – Higher mesh resolution, data dimensionality, and finer time increments stress current visualization tools and supporting infrastructure, such as I/O. New techniques such as in-situ and in-transit data analytics offer promise for relaxing the infrastructure requirements imposed by post-processing visualization. Another dimension of data analysis is the ability to compare sets of related data. This is crucial for code validation, as well as a thorough understanding of the physics.

Input/Output, File Systems and Storage – The capacity and performance of the disk/tape technologies that underlie ASC storage systems are on disparate trajectories, with performance gains on the worst (flattest) slope. This is further exacerbated by the higher levels of node performance. While there may be some further gains with respect to increased I/O parallelism, ultimately hybridized solutions will need to be explored and implemented. New approaches will likely be required for "scratch space" (while calculations are running) and for the backup and recovery systems. These problems cannot be solved with just new hardware solutions. Sophisticated software packages will be needed to make the hardware accessible and high performing.

Operating Systems, Runtime, and RAS (Reliability, Availability, and Serviceability) – In many cases, ASC has been able to utilize consumer-grade software, such as Linux with modification, for its HPC platforms. With the introduction of many-core chips for the mass market, industry should be in a position to provide ASC with operating systems suitable for massively parallel processing. However, the commercial usage model is very task oriented with tasks that are either naturally parallel (e.g., search queries, transaction updates) or totally distinct from each other (e.g., check for mail, scan for viruses, and play music). ASC must motivate HPC vendors to explore other solutions. The runtime systems must be more adaptable and be closely integrated with the RAS functions.

Application Programming Interface (API) Development – High energy consumption and insufficient reliability continue to be a concern. While solutions have been proposed in the hardware domain, all levels of the software stack must help facilitate application progress in the face of ever-increasing hardware failure rates and ever-decreasing floating point operations per joule. A strategy that includes application interfaces at all levels of the software stack is needed.

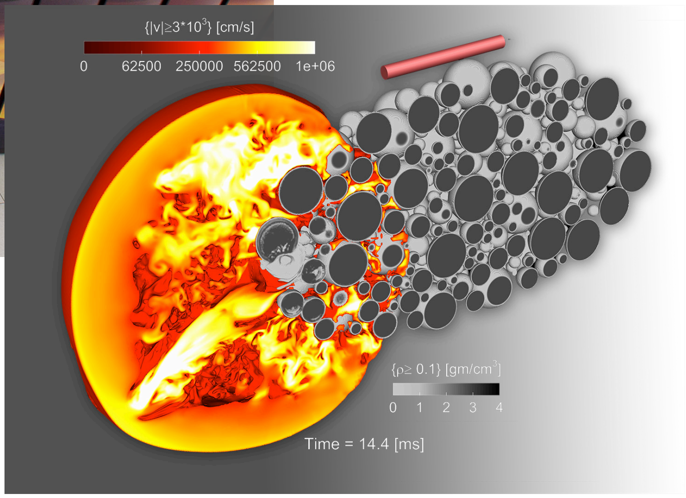
Facilities and Operations

Facilities are the support skeletons of the ASC computing infrastructure, providing space, power, cooling, and systems monitoring to increasingly complex and densely packaged computing and storage systems. As ASC looks to the challenges of the next generation, the demands on facilities will increase. Preparations must begin for changes in the facility hardware and in management practices. The ultimate goal of successful operations is maximizing the availability of computing platforms to the end users with minimum expense and effort required of the computing center.

Power and Infrastructure – Energy conservation in ASC facilities is critical for reducing operational costs, especially as next generation computers drive the requirements to new levels. Power considerations are a primary concern of facilities, with power budget for systems approaching and exceeding 20 megawatts (MW) for multi-petascale and exascale systems. Given the long lead-time to coordinate with state and local power providers for significant upgrades of electrical capacity in either new or existing facilities, ASC must plan far in advance and build in sufficient flexibility to facility designs to account for demands well into the future. Likewise, as future computing systems focus on dense packaging and liquid cooling to reach performance requirements, the weight per area of these machines may require significant upgrades to existing structural floor supports. It is likely that this trend will continue as innovation in packaging helps set system integrators apart from each other. The variety of options being planned will stress the facilities' ability to adapt.

Networking – Both local area networks (LANs) and wide area networks (WANs) must support the two platform classes. Latency requirements will be driven by response time for visualization. File transfers dictate the bandwidth needs. Both the CT and AT systems will place demands on each site's LANs. As their capabilities increase, so will the load on the networks. Each AT system is unique and is a shared ASC tri-lab resource, which will in turn stress the WAN. Of particular concern is maintaining and upgrading the encryption devices used for the ASC classified computing services.

Metrics and Best Practices – In order to measure progress and areas for improvement in operations, the metrics collected need to quantify efficiencies and areas for improvement more accurately. Whereas DOE computing facilities were once the largest in the world, data centers run by industry now eclipse ASC centers in terms of floor space and power availability, and in some cases, efficiency metrics. While the usage of ASC systems compared to industry data centers are quite different, the facility issues are much the same. Aggressive collaboration with other HPC and data centers to identify best practices and attainable standards is necessary.



Astroid simulation: Full-physics, full-geometry, 3D simulation using 32,000 processors on Cielo to model a 500-kiloton nuclear blast to explode a 5-kilometer asteroid (Cielo/LANL)

CO-DESIGN APPROACH

As we move through today's petascale era, ASC is embracing a co-design process to evolve and transition codes into the next decade's exascale era. Broadly speaking, co-design is a process of end-to-end optimization—from a code's fundamental physics and numerical methods, through its algorithms and data structures, to the hardware and system software that the code will ultimately run on. This process of co-design is often implemented through formation and management of collaborative, multi-disciplinary teams that include DOE computational scientists and representatives from hardware and software vendors. The team conducts its investigation through a tightly coupled cycle of application, algorithm and system software development, performance modeling, hardware simulation, and hardware design. Through the co-design process, optimal design tradeoffs are identified for hardware, the run-time environment, parallel file systems, physics algorithms and packages, and programming models. The inter-relationships amongst the team are fluid and agile, reorganizing and realigning themselves as necessary throughout the lifetime of the effort.

The NNSA co-design project will investigate a diverse set of technical areas: architecture-aware algorithms, programming models, system software, hardware architectures, resiliency, power management, etc. The key tools applied for these investigations are:

1. Proxy Applications,
2. HPC Architectural Simulators, and
3. Advanced Architecture Test Beds.

Characteristics that impact performance should be understood as early as possible in the analysis and design of new computers. Furthermore, it is often the case that there are multiple ways to design and implement the algorithms used in an application, and the choice can have a dramatic impact on application performance. Hardware architectural choices, if properly exposed through the system software to the application developers, can likewise have a tremendous impact on performance.

While the IC program element has the lead responsibility for the ASC co-design project, the multi-disciplinary nature of co-design indicates that the CSSE and FOUS program elements will contribute to and develop technical capabilities for the project.

A LANL researcher discusses details of a multi-million molecule simulation of a ribosome with a LANL intern in the Los Alamos RAVE (Reconfigurable Advanced Visualization Environment) facility. (LANL)

Two visualization scientists navigate through a simulation of a supernova in the Los Alamos RAVE. (LANL)

Proxy Application Development

Application performance is determined by a combination of many choices: hardware platform, runtime environment, languages and compilers, algorithm choice and implementation, and more. In this complex environment, the use of mini-applications, skeleton applications, and kernels (small self-contained software products collectively referred to as proxy applications) is an important approach for rapidly exploring the parameter space of all these choices. Furthermore, the use of proxy applications enriches the interaction among application, library, system software and hardware architecture developers by providing explicit functioning software and concrete performance results that lead to detailed, focused discussions of design trade-offs, algorithm choices, and runtime performance issues.

Recent work in application performance analysis takes advantage of two important properties of many applications. Firstly, although an application may have one million or more lines of source code, performance is sometimes dominated by a relatively small subset. Secondly, for the remaining code these applications often contain many physics models that are mathematically distinct but have very similar performance characteristics. Mini-applications take advantage of these two application properties by encapsulating only the most important computational operations and consolidating physics capabilities that have the same performance profiles. The large-scale application developer, who is tasked with designing and developing the proxy application, guides the decisions, resulting in a code that is a small fraction of the original application size, yet still captures the primary performance behavior.

HPC Architectural Simulation

Architectural simulator capabilities are important tools to enable co-design to close the loop back to computer architects and hardware component designers. Without this capability, co-design threatens to retreat to “business as usual,” in which new HPC systems are procured and ASC code teams, algorithm developers, and system software developers are then tasked with extracting the best performance possible from the HPC systems. The intent behind HPC architectural simulation is to obtain quantitative data to guide the technology development and design of all elements of the integrated HPC system.

ASC industry partners have a tradition of using simulators to analyze and model processors, interconnection networks, and other features of their proprietary designs. Some simulation capabilities are cycle-accurate and highly proprietary. To the extent that the ASC program can access and use these simulators, or provide proxy applications to drive these proprietary simulators, important quantitative data can be obtained to inform the co-design process. Processor models can be integrated with memory subsystem and network interface models to provide a node-level model. At a lower level of fidelity, it may be acceptable to give up on cycle-accurate processor models in order to enable simulations of entire HPC systems that integrate node-level models with models of the interconnection networks and to include analysis into holistic properties, such as energy consumption, thermal stresses, and system reliability. Again, proxy applications can be used to drive HPC architectural simulators to provide quantitative data to the co-design process.

ASC will support a co-design project for deep interactions with the vendor community through the use of proxy applications, architecture simulations, and experimental test beds.

Advanced Architecture Test Beds

Looking forward over the five years of this plan (and beyond), it is likely that some revolutionary technologies will be required, particularly if one considers the low fraction of peak performance in HPC utilization today and the need to avoid prohibitive power costs. It is also likely that many technologies for HPC will continue on an evolutionary track. ASC will adopt revolutionary technology changes, up to and including rewriting our applications, if that is what is required to meet mission needs. Recall that the original ASCI program made sustained and substantial investments to transform the application code base from vectors to MPI to create the current ASC application portfolio. In much the same way, the ASC program must undertake the challenge of transforming the EPIC codes to incorporate additional or different programming models.

It is critical to have a diverse set of experimental architecture test beds between now and 2015 to guide evolutionary versus revolutionary technology investment decisions. As a community, access to and experience with these experimental architecture test beds will allow ASC to become more informed collaborators in co-design processes, more adaptable to changes in hardware, and have stronger basis for making programming model changes. Perhaps more importantly, this experience will provide a foundation for decision makers to determine the path to exascale while the program continues to meet mission obligations.

A Singular Focus on FLOPS Leads to Unbalanced System Performance

The computer industry has long used FLOPS as a measure of the size of a computer. One way to calculate the FLOPS of a system is a standardized linear algebra benchmark called LINPACK. This benchmark measures the “sustained” FLOPS of a computer, which is a fraction of the peak theoretical FLOPS that can actually be used once other costs are accounted for. Interested parties can submit the LINPACK results for their computers to be included in a ranking list called the TOP500 (http://top500.org). In 2008, ASC fielded the Roadrunner system at LANL, which was the first computer on the TOP500 to break the sustained petaFLOPS (a million billion floating point operations per second) barrier. The ASC Sequoia system at LLNL was #1 on the June 2012 list and #2 on the November 2012 list with 16.32 petaFLOPS.

Unfortunately, there are issues with using this metric to specify or compare computers. The LINPACK test is not representative of the multi-physics simulation codes upon which the NNSA missions depend. Therefore, a high ranking on the TOP500 is not an indication that it is a good computer for meeting NNSA mission needs. The fact that Sequoia was also the number one system in the June 2012 Green500 and Graph500 lists is a testament to the attention LLNL and IBM paid to these other performance dimensions.

It makes little sense to focus solely on FLOPS targets for future ASC computers. Instead, the ASC program will identify a set of performance goals for each platform that will meet the computational needs of the program. Through co-design efforts, ASC is developing a portfolio of proxy applications that represent how real applications use (and stress) computer architectures. These proxy applications, and associated benchmarks derived from their development and evolution, will be the performance targets for future AT and CT systems.

EXTERNAL PARTNERSHIPS

The execution of the ASC Computing Strategy requires communication, alignment, and partnerships with many organizations and institutions. These include U.S. computer companies, universities, and the DOE/SC ASCR program.

Active Engagement with the U.S. Computer Industry

The ASC Program has been extremely successful in leveraging commodity computer technology to create systems without incurring the tremendous costs of fully custom designs. However, commodity computing, especially in the consumer markets, is rapidly moving away from the types of designs successfully leveraged in the past. The scientific HPC community, and NNSA in particular, must maintain close collaborations with computer vendors to ensure that features essential to its needs are continuously identified, integrated, and supported in future products. Appendix A summarizes the lessons learned since the publication of the *2007 ASC Platform Strategy*.

While close collaboration with computer vendors will mitigate the risk of a catastrophic loss of HPC capability, ASC simulation codes are unable to exploit enhanced hardware features without employing new programming approaches. This represents a significant challenge for ASC application developers, as mission drivers for new capabilities in simulation codes already stretch resources. The ASC program must make hard decisions on programming approaches that meet its current needs and for the future, while managing the timing of any technological disruptions.

In addition, the ASCR and ASC's FY12 joint collaboration on FastForward is intended to jump start new collaborative R&D partnerships with computer vendors. The AT systems will include funding to support NRE product development costs associated with the delivery and integration of first-of-a-kind systems.

Training and Education of a New Generation of Scientists and Engineers

The soon-to-be established new Predictive Science Academic Alliance Centers (PSAAP II) are expected to explore predictive science for multi-physics applications that use future generations of advanced computer architectures. This program, spanning more than eight major universities over the last 15 years, exposes young graduate students and post-doctoral candidates to the power of interdisciplinary, large-scale scientific simulation and computation as they work on the Centers' overarching, unclassified application problems and also during the requisite lab visits. ASC not only funds research in disciplines of high interest to the weapons program, but also makes enormous unclassified computing resources available to the Centers' research staff, post-doctoral candidates, students, and their mentors involved in these activities.

Strategic Partnership with DOE/SC ASCR

As described above, the ASC program is partnering with ASCR to leverage the respective programs' core talents and establish a set of collaborative R&D projects with industry and the national laboratories. Plans are now underway to execute collaborative procurements of the ASC AT systems with ASCR leadership class systems for future deployments. This is an invaluable opportunity to leverage the NRE investments of critical technologies to benefit both programs.

Located in Livermore's TSF computing facility, Sequoia is a 96-rack IBM BlueGene Q supercomputer used by researchers at the Los Alamos, Sandia, and Lawrence Livermore national labs.

SUMMARY

Successful stockpile stewardship is fundamentally dependent upon simulations and analyses of extraordinarily complex devices and physical processes in order to advance our scientific understanding and inform critical weapon system decisions. The ASC Program has the responsibility for ensuring that all necessary scientific expertise, computational platforms, and software infrastructure are available to meet current and future stockpile stewardship needs. The program must accomplish this with maximum efficiency within programmatic budget realities, while balancing conflicting priorities and preparing application codes and infrastructure for anticipated major changes in HPC architecture. The overriding objective is to maximize user and developer productivity while simultaneously enabling scientific code improvements and enhanced confidence in simulations of device and system performance outside of the data range previously provided by the nuclear test base.

The orderly evolution of HPC architectures that characterized the past two decades has been replaced by significant and potentially disruptive new architectural paradigms. External market and technology factors, including the shift to embedded and mobile systems, the slowing of Moore's law, and system power constraints, are driving CPU architecture towards designs not ideal for current ASC application codes. Effective industrial partnerships which emphasize co-design with ASC domain applications, computer science, and mathematical experts are a key element in working with industry to foster systems which will not require extensive code rewriting.

The ASC program continues to reap substantial benefits from lessons learned from the past. Many have led to substantial reductions in the cost of system acquisition and operations. The rapidly increasing computational capability of CT systems means that a substantial fraction of the stockpile stewardship workload can be performed on these systems at significant savings. And AT systems will be focused on satisfying mission needs beyond the capability of CT systems and on early adoption of promising new platform technologies. At the same time, ASC has emphasized the development and use of a common software infrastructure to promote ready access and use by scientists throughout the weapons complex.

In summary, the technical challenges associated with stockpile stewardship are placing increasing demands on computational analysis and simulation in order to meet critical mission needs and strategic goals. ASC is responding with a substantially revised *ASC Computing Strategy* to ensure the availability of critical computational resources in a timely and cost-effective fashion. Key elements of the new strategy include shifting to two platform acquisition classes: Commodity and Advanced Technology; reengaging and partnering with the nation's HPC industry via FastForward investments; and emphasizing co-design activities with industry, academia, and other federal HPC programs. The strategy is motivated and driven by the requirements for end user productivity, application code evolution, and managed response to technological changes.

ACRONYMS

API	Application Programming Interface
ASC	Advanced Simulation and Computing
ASCI	Accelerated Strategic Computing Initiative
ASCR	Advanced Scientific Computing Research
AT	Advanced Technology
COTS	Commodity off the shelf
CPU	Central Processing Unit
CSSE	Computational Systems and Software Environment
DAM	Defense Applications and Modeling
DOE	Department of Energy
DOE SC	Department of Energy Office of Science
CT	Commodity Technology
EPIC	Engineering and Physics Integrated Codes
FLOPS	Floating-point operations per second
FOUS	Facilities, Operations and User Support
GPGPU	General Purpose Graphics Processing Unit
HPC	High Performance Computing
IC	Integrated Codes
I/O	Input/Output
LAN	Local Area Network
LANL	Los Alamos National Laboratory
LLNL	Lawrence Livermore National Laboratory
MPI	Message Passing Interface
NNSA	National Nuclear Security Administration
NPR	Nuclear Posture Review
NRE	Non-Recurring Engineering
NSE	Nuclear Security Enterprise
PEM	Physics and Engineering Models
PSAAPII	Predictive Science Academic Alliance Program II
RAS	Reliability, Availability, and Serviceability
R&D	Research & Development
SNL	Sandia National Laboratories
SSP	Stockpile Stewardship Program
TLCC	Tri-lab Linux Capacity System
UQ	Uncertainty Quantification
V&V	Verification & Validation
WAN	Wide Area Network

APPENDIX A. UPDATED LESSONS LEARNED

Lesson 1: A stable, modestly changing computing environment significantly increases the productivity of code developers, designers, and analysts.

Discussion: Our highest-end simulation needs have often required new technical features only provided by the most advanced vendor-provided solutions—features often developed by partnering between the computer scientists and users at the NNSA laboratories and the computer vendors. To implement these solutions, ASC recognized the balance between short-term user productivity disruptions and the long-term productivity increases that result from more capable codes running on more powerful systems. To implement new technologies with minimum disruption to users, ASC developed a bulk-synchronous processing with explicit message-passing programming model that allows our applications to evolve independently from the characteristics of a particular generation of high-end computers. This model values optimizing long-term code portability over solutions that increase processor efficiency.

Lesson 2: The weapons workload benefits from a mix of computer systems available to match cost-performance to problem needs.

Discussion: Over the past several years, the marketplace evolved to a state in which high-end, commodity-based systems met the needs of many ASC capacity problems at substantially less cost. In the first 10 years of the ASCI/ASC Program, the focus was on commercial-off-the-shelf (COTS) processors, but the inter-processor communications fabric and software were customized because no alternatives existed. Now COTS providers have expanded to include the communications fabric and systems software including the Linux operating system, as well as a variety of open-source software for debugging and performance tools, system monitoring and control, job scheduling, and file systems. ASC investments contributed to this evolution of COTS technology. This has made possible the acquisition of TLCC systems, which can handle a substantial fraction of the workload at a significantly reduced cost.

Lesson 3: Investing in market-based supercomputers has proven to be a successful strategy for balancing system costs and progress in scientific computations.

Discussion: From its inception, the ASC Program decided to work with the computing industry to leverage its business models to build supercomputers for scientific applications. One benefit from such partnership was that commodity-based solutions provided an evolutionary path for applications, ensuring that code investments could cost-efficiently carry over to future generations. Another was that while market-based supercomputing platforms were expensive, they were still more affordable than custom-built architectures, and vendors were able to build, test, and deliver them in a relatively short period of time. Furthermore, given the low sales volume available from the scientific community, these business-based solutions leveraged a much larger market and provided a stable basis for producing ongoing generations of supercomputers. While the ASC Program does not target one-of-a-kind system investments, it encouraged innovations in computer hardware and software that increased the capability and efficiency of high-end systems.

Lesson 4: Bringing leading-edge systems to a production level is both a time- and resource-consuming process that requires a strong partnership between the laboratories and vendors.

Discussion: To meet requirements and to ensure that needed petascale computers would exist in the future, the ASC Program procured systems that accelerated the business plans of its vendor partners. This resulted in both an invigorated HPC industry and a series of "serial-number-1" systems which were acquired two to four years in advance of market offerings from a cross section of the industry. Over the past two decades, the ASC approach fostered competition and brought systems to market that would not have existed otherwise. Such systems provided a means to explore problem spaces previously not possible, but the application of such systems to production work introduced unforeseen problems in hardware and software reliability and system features. The tri-lab system integration teams worked closely with the ASC applications groups and system vendors and to ensure that when the applications uncovered bugs in hardware or software, the issues were dealt with quickly and the solutions were implemented in a practical manner. Often understated, this was an essential ingredient for success in an advanced development environment.

While one-of-a-kind systems are not the goal of ASC, the effort to push the state of the art sometimes results in systems that may not be commercially successful. While this is a disappointing outcome, from the perspective of the larger program it is an indication that ASC is pushing the state of the art. Finally, the experience and lessons learned for lab personnel creates a foundation for the next state-of-the-art system.

Lesson 5: Innovative architectural approaches provide significant future capabilities even though use of the advanced technology/architecture may be confined, in the early stages, to a subset of the important physics simulations for which ASC is responsible.

Discussion: AT systems, though higher-risk endeavors, have provided significant returns to ASC in terms of our understanding of innovative architectural features. For example, the IBM BlueGene/L (BG/L) computer with 131,000 non-COTS processors demonstrated efficient use of floor space and low power consumption, and the Cray Red Storm machine demonstrated high scalability with its advanced interconnect technology. Our use and investigation of the architectural innovations in AT systems with industrial partners ensures that we understand how to use the next-generation computers to solve stockpile stewardship problems and demonstrates to us that future HPC systems are suitable for our problems. We foresee that strategic investments in AT systems, which may begin as only applicable to a subset of stockpile issues but later expand in scope to become powerful general-purpose production engines, is a viable model for ASC success. By fielding an AT system while executing our day-to-day production computing responsibilities, we see a number of benefits:

- We learn how to write code for such future machines;
- We begin to understand how our applications can be made to work efficiently on new and possibly revolutionary architectures;
- We begin to port the large EPIC codes in advance of the production phase; and
- We train users and system personnel in its use.

Lesson 6: Computing-at-a-distance continues to be feasible for resolving even the most complex weapons system problems.

Discussion: Advanced, powerful ASC computers, whether deployed at Los Alamos, Livermore, or Sandia, are ASC tri-lab resources with major cycle allocations determined for each laboratory. This is a successful model and enables scientists at each of the laboratories to compute effectively from their home laboratory on the most powerful systems available within the NNSA complex. This success is enabled by robust classified networking resources and by data assessment tools that can be run in a variety of ways to fit particular programmatic requirements and platforms. Future AT systems will continue to follow this proven, successful usage model.

Lesson 7: U.S. competitiveness and leadership in high-end computing, enabled by government investment and industry commitment, are necessary for progress in science-based stockpile stewardship.

Discussion: Science-based stockpile stewardship could not have succeeded without a sufficiently healthy high-end computing industry. Our particular mission for national nuclear security has required computing performance beyond the capabilities normally available in the commercial marketplace. Designing and delivering state-of-the-art supercomputers that meet our stockpile stewardship needs within practical cost and schedule constraints have been possible through the efficiency and innovation of a healthy and competitive industry. Government support, through the ASC Program and a limited number of other agencies, helped to engender this competitiveness and leadership through both competitive procurements and a commitment to share some NRE costs for development of capabilities needed to scale the systems upward to meet our mission computing needs. The common denominator for continued U.S. competitiveness remains to be successful long-term partnerships between government agencies and industry.

APPENDIX B. ASC NATIONAL WORK BREAKDOWN STRUCTURE