REQUIREMENTS FOR ASCI

Executive Summary

STUDY LEADER
Roy Schwitters

CONTRIBUTORS INCLUDE:

Henry Abarbanel
Michael Brenner
John M. Cornwall
Alvin Despain
Paul E. Dimotakis
Sidney D. Drell
Douglas M. Eardley
Robert Grober
Raymond Jeanloz
Jonathan Katz
Steven Koonin
Darrell Long
Dan Meiron (Consultant)
Frances Perkins
Christopher Stubbs
Peter Weinberger

JSR-03-330E

July 29, 2003

The MITRE Corporation
JASON Program Office
1820 Dolley Madison Blvd
McLean, Virginia 22102
(703) 883-6997

1 Executive Summary

This section summarizes the conclusions and recommendations of the 2003 JASON summer study commissioned by the National Nuclear Security Administration (NNSA) to identify the distinct requirements of its stockpile stewardship program (SSP) in relation to the hardware procurement strategy of the Advanced Simulation and Computing (ASCI) program. In particular, we were tasked to evaluate the increased risk to the nuclear weapons (NW) stockpile and the scientific program of SSP as a result of delaying computer acquisitions intended to advance computing capability. We were also charged to consider the confidence in our NW simulation capability and the appropriate balance between near-, intermediate- and long-term SSP needs for acquiring new hardware with increased computing capability.

Today, ASCI comprises high-performance computing hardware, suites of large codes built on a validated scientific/engineering base, experienced people and connections to the greater scientific, computing and national security communities. The tools and methods developed under ASCI have evolved to the point where they are today *essential* to stockpile stewardship. So are the people supported through ASCI who, working closely with NW designers, engineers and managers, have acquired invaluable expertise in developing and optimizing ASCI tools and in establishing and improving the scientific credibility of NW simulations. Some notable ASCI accomplishments are described in the body of this report.

Two commonly used measures of the overall productivity of ASCI platforms are *Capability* and *Capacity*. The common unit of measure for both is peak floating-point operations, noted as TeraFlops or TF (10^{12} operations per second). Used in this context, *Capability* refers to the maximum processing power possible that can be applied to a single job and *Capacity* represents the total processing power available from all machines capable of operating ASCI codes. A given amount of *Capability* implies *Capacity* in two ways: 1) by its direct contribution to total capacity and 2) because a high-*Capability* machine can be re-configured into multiple lower-*Capability* machines to run multiple shorter jobs, often with somewhat improved overall performance.

Today, the ASCI platforms of highest *Capability* are LLNL's "White" at 12.3 TF and LANL's "Q" at 20 TF. The next planned acquisitions are SNL's "Red Storm" projected to be 40 TF and LLNL's "Purple C" at 100 TF. SSP requirements over the next few years call for a few large jobs which need the largest available *Capability* but the most important trend is a factor-of-two oversubscription in ASCI *Capacity* (well supported in our view by distinct technical requirements) which is projected to go on and potentially worsen in the foreseeable future. This level of demand means that jobs are selected to run by some combination of administrative fiat and management priority-setting; it creates strong incentives for all involved in ASCI computing to improve performance of algorithms and platforms for greater delivered performance. A strong and valuable effort has been made by the ASCI program to increase the efficiency of performance (ratio of computing operations delivered to peak performance) to current levels; in practice, about 0.5–15% of the peak processing speed is realized, depending on algorithms and details of implementation. Similar efficiencies are found in many commercial, engineering and scientific applications. While applauding efforts within ASCI to improve efficiency, a continuing investment in improving efficiency is called for.

Within 10 years, estimates of the demand for *Capability* and general physics arguments indicate a machine of 1000 TF = 1 PetaFlop (PF) will be needed to execute the most demanding jobs. Such demand is inevitable; it should not be viewed, however, as some plateau in required *Capability*—there are sound technical reasons to expect even greater *Capability* demand in the future.

We were charged to evaluate the increased risk to the stockpile and to the scientific program it supports that would result from delaying acquisitions to advance *Capability*. To assess this additional risk, we constructed an acquisition scenario where FY 04 funding was reduced by $33M (24%), requiring that the procurements of Red Storm and Purple C be stretched out. Given the approximately $50M shortfall in requested FY 03 funds, the FY 04 reduction was assumed to lead to further reductions in FY 05, 06 and 07, which we modeled by an approximately flat acquisition budget through FY 08 at a level of approximately $120M–$130M. The net effect of such a stretch-out is to reduce overall *Capacity* to below 1/3 of demand during the critical program years FY 05–08. In this period, several program milestones are to be completed including the refurbishment and first production units of one major weapons system, and the qualification of critical components of another. We judge the risk to SSP of such a delay to be *high*, not so much from the delayed *Capability*, but from the very serious reduction in overall *Capacity*. The large jobs projected to require 100 TF-level *Capability* might be deferred at moderate risk to the program, but the resulting very large oversubscription in *Capacity* risks becoming unmanageable. Purchasing the proposed large platforms—Purple C and Red Storm—on a stretched-out, suboptimal schedule where their CPUs and other components are bought after their performance/cost prime strikes us as being unwise both in terms of delivered capability and capacity.

In evaluating NNSA's current ASCI platform acquisition strategy, we see two areas of substantial risk. The first is related to the current and projected oversubscription in *Capacity*. A factor-of-two oversubscription is probably manageable; however, larger demand on *Capacity* could become unmanageable. The likely result would be the delay in meeting SSP technical milestones and/or overly-cautious decision-making leading to expensive—at the $50M–$100M level—mitigation programs from elsewhere in the SSP (for addressing SSP life-extension-program or significant-finding-investigation issues, for example) that might not be necessary if sufficient confidence were provided by timely simulations. The second area of risk is the lack of a credible "road map" to acquiring the next generation of machines with PF *Capability*, which will be needed within a decade. Scaling to 1 PF using present machine architectures implies very large numbers of processors—of order 100,000, perhaps—might be needed. Such large numbers raise serious questions of scalability of code performance and of machine reliability.

JASON recommends SSP management consider four general areas for mitigating the risks associated with its present ASCI acquisition strategy:

1. Platform Acquisition:

    (a) Modify the allocation of resources in the current acquisition plan to provide additional *Capacity* platforms starting as soon as possible.

    (b) Lay the groundwork for future *Capability* machines. This may involve acquisition of "*Capability*-exploration" machines focused on optimizing efficiency in computation for ASCI problems in order to gain experience with architectures that might plausibly be extended to the PF level. These acquisitions can probably begin in the FY 06–07 time frame and may be able to replace currently planned greater-than 100 TF-scale machines that are scheduled to follow the Purple C and Red Storm acquisitions.

2. SSP Requirements:

   Set priorities in the SSP requirements and assign ASCI resources accordingly. This is essential to reduce excess computing demand and to assure that the high priority problems are addressed to meet goals as scheduled. In particular, some of the stockpile-to-target-sequence requirements identified for Cold War scenarios place significant demands on ASCI resources (and, in fact, on the entire SSP). They should be reviewed carefully in the context of current and anticipated US security needs. Those judged to no longer meet a compelling cost/benefit standard should be either relaxed or assigned an appropriately lower ranking in the queue of high-priority tasks for ASCI's available and planned future resources.

3. ASCI Operations:

   (a) Expand access to ASCI "most-capable" systems to best align ASCI *Capability* with overall SSP priorities and needs. We applaud efforts to make ASCI platforms available to workers across the complex, regardless of where a given machine or work-group is located, but present quotas should be re-examined. We expect additional benefits to accrue from enhanced scientific communications and competition. Depending on the architecture of future PF-level machines, it is possible that future acquisitions will entail fewer high-*Capability* platforms, in which case convenient and flexible access across the complex becomes essential.

   (b) Continue and expand, as appropriate, investment in computational science investigations directed toward improving the delivered performance of algorithms relevant to ASCI; consider dedicating regularly scheduled machine time on capability platforms for efficiency studies.

4. Nuclear Weapons Science:

   Encourage the advance of NW science at every opportunity in the SSP and ASCI programs. Better science is the most cost-effective way to reduce risk in the stewardship program and the only possible way to achieve sufficiency in the modeling and understanding of these complex systems. Some excellent new science is beginning to emerge in association with ASCI. In the body of this report, we comment on some of this new science and suggest possible extensions.

To summarize, our major concern is in providing as soon as practical much needed capacity to the ASCI program lest the resulting very large oversubscription in *Capacity* becomes unmanageable. In addition, a road map must be developed to deliver to the program machines of the requisite *Capability*.