Committee of Visitors Report

Advanced Scientific Computing Research

July 2012

Review of Computer Science Program
Fy 2009, 2010, 2011

Background

Date of COV: July 10 – July 11, 2012
Program: Computer Science
Fiscal Years Being Reviewed: 2009, 2010, 2011
Office: Advanced Scientific Computing Research (ASCR)

Committee Membership:

- Susan Graham, UC Berkeley

- Tony Hey, Microsoft Research

- Sylvia Spengler, NSF

- Anne Trefethen, Oxford University

- David Walker, Cardiff University

Charge to Committee

1. **Assess the efficacy and quality of the processes used during the past three years to:**
   (a) solicit, review, recommend, and document proposal actions and
   (b) monitor active awards, projects and programs.

2. **Within the boundaries defined by DOE missions and available funding, comment on how the award process has affected:**
   (a) the breadth and depth of portfolio elements;
   (b) the degree to which the program is addressing the challenges of multi-core hybrid computing and peta-to-exascale scientific data management;
   (c) the national and international standing of the program with regard to other computer science research programs that are also focused on the demands of high performance scientific computing and analysis of petascale datasets.

ASCR CS Program

- The Computer Science research agenda fills a critical gap in scientific computing. The computing resources required to fulfill the Office of Science mission exceed the state-of-the-art by a significant margin. Furthermore, the software tools, libraries and the distributed software environments needed to accelerate scientific discovery through modeling and simulation are beyond the realm of commercial interest. Yet, the computing resources and the applications that run on them are vital to maintaining the United States' competitiveness in the world economy.

- The Computer Science program supports research that enables computing at extreme scales and the understanding of extreme scale data from both simulations and experiments. It aims to make scientific computers as easy and effective to use as possible. Extreme scale refers to the use of Exascale computing platforms that will be operational in the 2018-2020 timeframe. Exascale computing platforms will be capable of up to 1 quintillion (10^{18}) floating point operations per second.

- In order to ensure the efficiency and productivity of the supercomputing systems managed and operated by the Office of Science, the Computer Science program addresses challenges in advanced computer architectures; programming models, languages, and compilers; execution models, operating, runtime, and file systems; performance analysis and productivity tools; and data management and data analytics, including visual analysis."

Method of Review (1)

Overview Presentation by Dan Hitchcock:

- Context of ASCR within the Department of Energy and the Office of Science and described the organizational structure of ASCR.

- The ASCR funded facilities – High-End and Leadership Computing; Research and Evaluation Prototypes (REP) and the Energy Sciences Network (ESnet).

- Extrapolation of existing technologies to create an Exaflop system would require 200 MW of power: target is to deliver an Exaflop computer by 2020 that requires only 20 MW.

- Since communication is expensive in both time and energy, software and algorithms are needed that minimize data movement.

- The ASCR Exascale Co-Design Centers will enable us to understand how to allocate complexity between hardware, systems software, libraries, and applications.

Method of Review (2)

Presentation by PM Lucy Nowell:

- Explained context of ASCR's Core CS program within a national context coordinated by the National Information Technology R&D (NITRD) group.
- ASCR's CS program is addressing two fundamental questions:
  - How can we make today's and tomorrow's leading edge computers tools for science;
  - How do we extract scientific information from large data from experiments and simulations.
- Context for the CS program from the ASCR facilities; Research and Evaluation Prototypes (REP); the Exascale Co-Design Centers; and the Applied Mathematics and Next Generation Networking Programs.
- Budget for the CS program had risen from $30,782K in FY 2009 to $47,301K in FY 2011.
- Lucy Nowell and Sonia Sachs were now full-time federal PMs but it was clear that the CS program was still under-resourced in terms of support.

Method of Review (3)

Presentation by PM Lucy Nowell (continued):

- The ASCR CS research program falls into five general categories:
    - Operating and file systems;
    - Performance and productivity tools;
    - Programming models;
    - Data management and visualization;
    - Extreme-scale architectures.
- In addition, the joint Applied Mathematics-Computer Science Institutes were being phased out and a new theme, 'simulation of advance architectures' was added in the FY12 budget request.
- Details were given of both the university and laboratory proposal submission processes and of the peer review criteria and process.
- In terms of CS Strategic Planning, during the period 2008 to 2011, the program had convened 14 workshops on the scientific challenges posed by Extreme Scale computing and latterly on the technology issues for Exascale computing.

Method of Review (4)

Presentation by Lucy Nowell (continued):

- Detailed statistics of 4 FOAs that made awards during the COV time-frame were then given:
  - 2008: FOA 08-19 on Petascale tools (12 projects funded in 2009).
  - 2010: FOA 10-255 on Advanced Architectures (6 projects funded);
  - 2010: FOA 10-256 on Scientific Data Management and Analysis at Extreme Scale (10 projects funded);
  - 2010: FOA 10-257 X-Stack Software (10 projects funded).
  - Details of the FOA reviewers were also given.

- In addition to these major FOAs, details were given of the 2009 FOA for Early Career Research Program and of unsolicited proposals and CS renewals in FYs 2009, 2010 and 2011.

- After a brief review of the CS portfolio funding trends, Nowell's presentation concluded with a discussion of the challenges of data-driven science.

Method of Review (5)

- Dialog with Director Bill Harrod (day 2)
  - Vision for Exascale
  - Discussion of Exascale initiative and its relation to the base DOE CS Program
  - Discussion of international collaboration
- Review of peer-reviewed applications from FOAs
  - COV examined a sample set of proposals and reviews
  - No anomalies were found

Primary Findings and Recommendations:

1(a) Efficacy and quality of the processes used to solicit, review, recommend and document application and proposal actions

The COV considers the CS program to be generally effective and well managed. The solicitation and review processes appear to be effective and fairly administered. The documentation of these processes and the capture of associated summary statistics are much improved since the last COV review.

Recommendations:
- Continue to improve the online information management capabilities of the program (and related ASCR programs that incorporate computer science research), informed by an overall plan, and by best practices from other funding organizations such as NSF and NIH.
- Expand the information management capabilities to incorporate a reviewer database that records areas of expertise, quality of past reviews, responsiveness, and conflicts of interest, and a PI database that identifies previous successful and unsuccessful DOE proposals, links to research and project websites, and all currently active DOE-funded projects.
- Introduce mechanisms to provide balanced and knowledgeable reviewers by using a less crude, more refined approach to conflicts of interest.
- Provide a longer-term, more coherent schedule of planned solicitations, adapted as necessary to budget contingencies and ongoing research advances.
- Incorporate some mechanism for funding the exploration of promising new ideas that might not conform to the planned research programs.

Primary Findings and Recommendations:

1(b) Efficacy and quality of the processes used to monitor active awards, projects and programs

The CS research program managers use generally effective mechanisms, including site visits, meetings and progress reports, to monitor ongoing awarded projects. The COV was impressed by the effort that program managers put into maintaining effective oversight of the current awards. The time and intellectual commitment are significant, as displayed by the calendars and activities of the individual managers. The effectiveness of the program managers could be enhanced by considering additional mechanisms that do not rely on such frequent face to face meetings.

Recommendations:
- Computer science program managers should be encouraged to consider how new technologies and new media, including social environments and hubs, could be used to provide more efficient oversight.
- Better metrics should be developed for evaluating the impact and future needs for workshops and other conferences used as oversight mechanisms.
- A team approach needs to be developed to utilize the staff of ASCR and the Computer Science program managers most efficiently while maintaining adequate oversight of current research activities.

Primary Findings and Recommendations:

2(a) Within the boundaries defined by DOE mission and available funding, comment on how the award process has affected the breadth and depth of portfolio elements.

Overall, the awards process (open solicitation, peer review, decision by ASCR) has resulted in the funding of a broad range of projects relevant to DOE's mission. The new strategic focus on Exascale computing has resulted in a good balance between mission-critical and horizon-scanning elements, and it is apparent to the COV that the ASCR Computer Science program has continued, in general, to support high-quality, leading-edge research. One concern for the COV was how ensure that the present CS research teams that provide users of the present state-of-the-art supercomputers with ever improving software libraries and tools are maintained at a critical mass. A second concern was how to coordinate the funded research projects selected by the peer review process for each FOA. In particular, it was not clear how these independent projects could be integrated to meet a strategic goal such as delivering a coherent Exascale software system.

Recommendations:
- It is important that ASCR's CS program maintains a balance between its focus on Exascale research and the traditional research strengths of the CS research groups at the DOE labs.
- The CS program should consider the importance of research into energy-efficiency, machine learning and data analytics for Exascale systems within the context of its overall planning for the Exascale computing, and more prominence should be given to these topics in future solicitations.

Primary Findings and Recommendations:

2(b) Within the boundaries defined by DOE missions and available funding, comment on how the award process has affected the degree to which the program is addressing the challenges of multi-core hybrid computing and peta-to-exascale scientific data management.

The challenges of multi-core hybrid computing are being addressed through both the awards made in response to the Petascale Tools FOA 08-19 and through the Exascale awards to FOA 10-257 X-Stack Software. Since the last COV report, a start has been made in addressing the data management and analysis agenda. The ASCR team is also participating in the Office of Science's Digital Data Working Group. In this group, the Basic Energy Sciences (BES) program has significant data management challenges from their SNS neutron and LCLS X-ray Laser facilities which will generate Terabytes of data each day. Similarly, the Biological and Environmental Research (BER) program has major data challenges in its genomic activities and in its research on biogeochemical systems. What is surprising is the lack of any significant ASCR support for research into Machine Learning and Data Mining technologies.

Recommendations:
- The review panel should ideally contain a mix of university and DOE Laboratory researchers.
- The CS program should work with the BES and BER experimental data communities as well as ASCR's traditional simulation and modeling community in its scientific data management and analysis program.
- ASCR should consider setting up a research program to build expertise in Machine Learning and Data Mining technologies in support of the Office of Science's data mission.

Primary Findings and Recommendations:

2(c) Within the boundaries defined by DOE mission and available funding, comment on how the award process has affected the national and international standing of the portfolio elements.

The schedule for delivery of Exascale computing also appears to have been significantly delayed, with 2023 now considered a more realistic timescale for the DOE to deliver an Exascale system. However, the COV notes that other countries, including Japan and China, have adopted a more aggressive timescale and are a potential risk to the US's leadership both in terms of the delivery date for Exascale systems and also in chip development and production.   It is also the view of the committee that an annual workshop in the US is not likely to provide sufficient engagement with international activities to allow significant collaboration. .

Recommendations:
- ASCR should do all that it can to ensure that it receives sufficient investment in Exascale for the US to remain internationally competitive.
- The program should maintain its leadership role in high end computing by continuing to engage with the international community.

General Findings and Recommendations:

A significant problem for the ASCR CS program and the DOE research community is the present level of uncertainty with respect to funding for the Exascale initiative. The COV believes that significant additional funding is required for ASCR to successfully execute on delivering an Exascale computing platform. To retain US leadership at Exascale it is imperative that the uncertainties about both the details of the plan and funding be resolved as soon as possible. The COV considers the CS program to be effective and well managed. The documentation of these processes and the capture of associated summary statistics are much improved since the last COV review. The CS program managers use generally effective mechanisms, including site visits, meetings and progress reports, to monitor ongoing awarded projects. The COV was impressed by the effort that program managers put into maintaining effective oversight of the current awards. However, the committee also believes that the number of permanent staff currently allocated to the CS program is insufficient for sustaining these processes for the long term. Although three additional CS staff positions have been approved in principle, there has been no progress in allocating FTEs to these vacancies.

Recommendations:
- ASCR should work with ASCAC and the Office of Science to do everything possible to secure adequate additional funding for the Exascale initiative and protect US leadership in supercomputing technology.
- The COV recommends that ASCR negotiate to be allowed to fill the approved CS vacancies as quickly as possible.