Committee of Visitors Report

Advanced Scientific Computing Research

July 2012

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

0.0 Executive Summary

The Committee of Visitors (COV) for the Office of Advanced Scientific Computing Research (ASCR) program in Computer Science met July 10 - 11, 2012 at the DOE facility in Germantown, MD. The COV is grateful to the program officers and other ASCR staff who gave of their time and knowledge to help the committee in its deliberations.

The specific charge to the COV included the following five elements:

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.

Findings and Recommendations:

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

Findings:

Based on the presentations and interviews with program officers and management, and on examination of project folders in the Computer Science (CS) program, 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.
- Explore possible mechanisms to relax the present, very stringent approach to excluding reviewers with potential conflicts of interest in order to achieve a better balance between external reviewers and reviewers familiar with DOE's and ASCR's mission.
- Investigate the feasibility of providing a longer-term, more coherent schedule of solicitations, recognizing budget contingencies and ongoing research advances.
- Devise some new mechanism for funding the exploration of promising new ideas that might not conform to the planned research programs.

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

Findings:

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.

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:

Findings:

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 to ensure that the present DOE Lab CS research teams that provide users of the present state-of-the-art supercomputers with innovations in software libraries and tools are maintained at a critical mass. This is related to a second concern of 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 would be integrated to meet the strategic goal of delivering a coherent Exascale software system. Such integration could naturally be undertaken by the base CS research program at the DOE Labs but this requires maintaining their core base of research expertise.

Recommendations:

- ASCR's CS program should maintain an appropriate balance between funding for the Exascale research program and the base CS research program at the DOE labs.
- More prominence should be given to research into energy-efficient computing, machine learning and data analytics in future solicitations since these topics are important more generally than just in the context of the Exascale initiative.

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:

Findings:

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/Data Mining technologies.

Recommendations:

- Review panels should ideally contain a mix of external and DOE Laboratory researchers.
- In its scientific data management and analysis program the CS program should work with the BES and BER experimental data communities as well as ASCR's traditional simulation and modeling community.
- ASCR should consider setting up a research program to build expertise in Machine Learning/Data Mining technologies in support of the Office of Science's mission for Big Data and data-intensive science.

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:

Findings:

A result of the present uncertainty in funding for the ASCR's Exascale initiative, there is a danger that the schedule for delivery of an Exascale computing system could be significantly delayed. However, the COV notes that other countries, including Japan and China, have adopted an aggressive timescale and constitute a significant, 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 funding for the Exascale initiative 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 Observations:

Findings:

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 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 so far only been progress in allocating one new FTE to these vacancies.

Recommendations:

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

1.0 Introduction

The U.S. Department of Energy's Office of Science founded the Advanced Scientific Computing Research (ASCR) program to develop the algorithms, computer programs and hardware that advance scientific research. The Department of Energy has long recognized that development of high-performance computers, the networks to connect them and the software to run them is crucial to America's research lead. The Office of ASCR is organized as two divisions: the Computational Science Research and Partnerships Division and the Facilities Division. The research program is divided into Applied Mathematics, Computer Science, Next Generation Networking for Science and the Partnerships program (SciDAC) which includes the Co-Design Centers.

The ASCR website describes the CS research program as follows:

"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."

This COV has been asked to review the management processes for the Computer Science (CS) elements of the ASCR program and assess the operations of the CS programs during the fiscal years 2009, 2010, and 2011.

1.1 Method of Review

A charge letter from the Director of the Office of Science (SC-1) to the Chair of ASCAC, dated October 18, 2011, established the Computer Science Committee of Visitors (COV). The Director of ASCR in consultation with the ASCAC Chair selected the COV chair and announced details of the COV members at the March 2012 ASCAC meeting. The list of participants in the COV is provided in Attachment I and the charge letter is provided in Attachment II. The COV conducted telephone and email exchanges with CS program directors and had a site visit on July 10th and 11th 2012.

This report presents the findings and recommendations of the COV. The committee met with ASCR members for an on-site visit at the DOE Germantown location on Tuesday, July 10th and Wednesday, July 11th, 2012. The COV Chair, Tony Hey, introduced the members of the committee and summarized the charge to the committee. The Associate Director, Dan Hitchcock, gave an overview presentation of the Office of Science's activities and the role of ASCR.

- He explained the context of ASCR within the Department of Energy and the Office of Science and described the organizational structure of ASCR.
- The ASCR funded facilities were discussed – High-End and Leadership Computing; Research and Evaluation Prototypes and the Energy Sciences Network (ESnet).
- He then outlined the challenges and investments needed for Exascale Computing. Extrapolation of existing technologies to create an Exaflop system would require 200 MW of power: the 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.

CS Program Manager Lucy Nowell then gave a detailed presentation of the core Computer Science program. This covered the context for the ASCR Core CS Program; the CS mission and budget; CS personnel issues; proposal processing and reviewing practices; CS Program planning; Solicitations, Proposal statistics and Awards for FY09, FY10 and FY11; the CS portfolio profile; Exascale research conferences and workshops; and the Data Tsunami.

- Nowell explained the 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; and how do we extract scientific information from large data from experiments and simulations.
- Factors providing important context for the CS program are the ASCR facilities; Research and Evaluation Prototypes (REP); the Exascale Co-Design Centers; and the Applied Mathematics and Next Generation Networking Programs.
- The budget for the CS program had risen from $30,782K in FY 2009 to $47,301K in FY 2011.
- In terms of ASCR Program Managers supporting the CS program, Lucy Nowell and Sonia Sachs were now full-time federal employees but it was clear that the CS program was still under-resourced in terms of support.
- The CS research program falls into five general categories: operating and file systems; performance and productivity tools; programming models; data management and visualization; and extreme-scale architectures. In addition, the joint Applied Mathematics-Computer Science Institutes were being phased out and a new theme, simulation of advanced architectures had been 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 initially the scientific challenges posed by Extreme Scale computing and latterly on the technology issues for Exascale computing.
- Detailed statistics of 4 FOAs that made awards during the COV time-frame were then given. These calls included the 2008 FOA 08-19 on Petascale tools which received 30 unique project proposals and funded 12 projects in 2009. There were three FOA calls in 2010: FOA 10-255 on Advanced Architectures (28 unique project proposals, 6 projects funded); FOA 10-256 on Scientific Data Management and Analysis at Extreme Scale (37 unique project proposals, 10 projects funded); and FOA 10-257 X-Stack Software (61 unique project proposals, 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.

The COV spent the rest of the first day discussing the presentations and were able to talk with Director Bill Harrod on the morning of day 2. He gave us his view of the opportunity for the Exascale initiative to fundamentally change the way we do computing. In view of the

importance of the Exascale initiative to the whole of ASCR's activities, and to the CS program in particular, it was unfortunate he was unable to be present on the first day. It would have been very helpful for the committee to have heard a more detailed account of his views of how the activities of CS program related to the overall ASCR Exascale initiative, prior to their detailed review of the CS program. Although the COV was charged with evaluating the ASCR CS program, since Exascale extends across the whole ASCR portfolio – and in particular includes the Co-Design Centers and REP program – it was difficult for the committee to judge the significance of the Exascale component of the CS portfolio as a part of the ASCR-wide initiative. In particular, the absence of an approved Exascale plan from DOE is clearly creating an undesirable amount of uncertainty but the COV acknowledges that this is not in ASCR's control.

The COV then spent time examining samples of peer reviewed applications.

1.2 COV Charge

The specific charge to the COV included the following five elements:

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.

2.0 Efficacy and Quality of the Program's Processes

The COV considers the CS program to be generally effective and reasonably well managed. The program officers are clearly dedicated and competent public servants who have considerable

knowledge of their portfolios (even with such relatively short tenure) and of the relevant communities of practice.

Charge 1 (a): Assess the efficacy and quality of the processes used during the past three years to solicit, review, recommend, and document application and proposal actions.

The COV spoke with the program officers about the solicitation and review processes and examined some electronic files for the submission, review, and decision documentation of proposals. The solicitation and review processes appear to be effective and fair, and to be well managed. At the time of the last COV review, the CS program was seriously understaffed. It remains understaffed, but the program is benefiting significantly from the contributions of the two program managers now in place. Proposals continue to be evaluated primarily by panel review rather than by mail review. Panels evaluate each proposal but are not expected to reach consensus or to rank the proposals. This gives the program manager considerable flexibility to balance the investment portfolio according to mission needs.

Findings:

The documentation of the solicitation and review processes and the capture of associated summary statistics are much improved since the last COV review. All submissions, reviews, and documentation are captured in a uniform manner. However, it still appears necessary to create some reports "by hand" that should be available by simple queries. It should be possible to provide longitudinal reports, both individually and in aggregate, about PI's, about submitting institutions, about reviewers, and about both declined and funded proposals. DOE is not alone in needing this kind of information – it is important for any funding organization. DOE should continue to improve its online information management and to learn from other organizations how best to plan and stage its improvements. For example, NSF and NIH are moving to an integrated approach via research.gov that allows consideration of whole sets of reviews – a capability that would be helpful to DoE program managers as well.

The quality of the reviewers and the diversity of institutions at which they work have been strengthened considerably. We have some concern that the pool of reviewers for some solicitations is greatly limited because of the breadth of the conflict-of-interest strictures. There is value in being able to use reviewers who are not only expert in the technical merits of the proposed research, but also familiar with the DOE mission and its research context. We encourage ASCR to explore means to include more DOE-knowledgeable reviewers by refining the criteria by which conflicts of interest are determined.

The calls for proposals (FOA's) have been irregularly scheduled in the three years under review – there were none in 2009, three in 2010, and none in 2011. The seemingly ad hoc and irregular pattern of FOA's and the somewhat unpredictable choice of topics, together with the

phasing out of block grants, are a detriment to planning on the part of the research community. Although it is essential that researchers continue to do high-quality research that is responsive to the agency's mission in order to receive funding, it is also important that the best researchers enjoy some security and predictability in their careers that encourages them to continue to contribute their talents to the DOE mission. In times of financial pressure, FOA's with fewer or smaller awards are preferable to an absence of solicitations, and further forward planning would be beneficial.

It has also been the practice to fund unsolicited proposals (those not submitted in response to a FOA) only if funds become available unexpectedly. This appears to be the only mechanism to infuse the program with new out-of-the-box ideas. This seems to the COV to be a lost opportunity. We encourage ASCR to provide some funds on a regular basis for exploration of novel promising ideas that would contribute to the DOE mission.

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.
- Explore possible mechanisms to relax the present, very stringent approach to excluding reviewers with potential conflicts of interest in order to achieve a better balance between external reviewers and reviewers familiar with DOE's and ASCR's mission.
- Investigate the feasibility of providing a longer-term, more coherent schedule of solicitations, recognizing budget contingencies and ongoing research advances.
- Devise some new mechanism for funding the exploration of promising new ideas that might not conform to the planned research programs.

Charge I (b): Assess the efficacy and quality of the processes used during the past three years to monitor active awards, projects and programs

An important role of the ASCR office is to monitor progress of the efforts funded by the Program. Such monitoring is essential to ensure that the research conducted is achieving the

desired results, that those results can be effectively applied to the mission of the Office of Science, and that those results are disseminated to the broader research community. Since the program has a history of sequential awards with a number of the institutions and investigators it funds, effective monitoring is especially important to insure that such relationships continue to be of value to the mission.

Findings:

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.

The proposed mechanisms for oversight seem to place an unreasonably heavy burden on both staff and awardees by requiring frequent face to face meetings. This is especially onerous for the program managers given that staff numbers are so low. The present plan requires that each awardee participate in face to face meetings with staff and ASCR management at six month intervals. This seems hard to achieve given the calendars of all of the people involved, including the Division Director. It may also not necessarily be the most efficient way to develop an overall view of each project's activities. For example, the program managers could provide awardees with detailed material covering the project monitoring expectations. Perhaps such material could be made more generally available as an explicit part of the award conditions.

The program managers use a number of mechanisms to assess the impact of the numerous workshops held by Computer Research. The major mechanism appears to be the gathering of attendee feedback but such surveys are inevitably incomplete. The program may be better served if staff developed some alternative mechanisms for follow-up and for measuring the effectiveness of the workshops. These mechanisms could include items such as the development of new solicitations, the establishment of new collaborations, or the use of social media throughout the meeting to provide a persistent record of the interactions.

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.

3.0 Effect of the Award Process on Portfolios

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

The last COV report in 2009 found that:

'The ASCR CS program has clearly played a major role in supporting the leadership role of the DOE Office of Science in capability computing. The researchers funded by the program and their research are prominent in the international community and recognized as world-class. Software libraries and tools funded by the CS program are widely used not only in the US but also in Europe and Asia.'

These software libraries and tools are vital to users of high-end systems and are constantly being improved and updated. This is an important non-Exascale component of the CS program.

Since the last COV, a large part of the ASCR Computer Science program has been focused on the "Exascale challenge": the goal of deploying an energy-efficient Exascale computer, capable of addressing DOE's mission needs, within the next decade. In this context, ASCR issued three Computer Science solicitations in 2010:

1. Advanced Architectures and Critical Technologies for Exascale Computing (FOA 10-255), which funded 6 projects addressing topics such as energy efficiency, the co-design of hardware and software, and memory technologies. Total $5M/year.
2. Scientific Data Management and Analysis at Extreme Scale (FOA 10-256), which funded 10 projects addressing topics such as file systems and I/O, data analysis and scientific discovery. Total $5M/year.
3. X-Stack Software Research (FOA 10-257) which funded 11 projects. X-Stack addresses the need to substantially, or totally, re-write the system software stack for Exascale computers, and to assist the migration of MPI-based applications to handle billion-way parallelism. The funded projects address topics such as auto-tuning, languages and runtime systems, productivity tools, and fault tolerance. Total $8.5M/year.

In addition to the above, in FY09-FY11, funding for 8 existing projects was renewed, and 15 unsolicited proposals were funded. In the same period ten projects were funded under the Early Career Research Program (ERCP).

Findings:

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. The emphasis on Exascale is clearly to be welcomed and is introducing challenging research issues for energy efficiency, chip design, memory systems and computer architectures. However, one concern for the COV was how to ensure that the present DOE Lab CS research teams that provide users of the present state-of-the-art supercomputers with innovations in software libraries and tools are maintained at a critical mass. This is related to a second concern of 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 will be integrated to meet the strategic goal of delivering a coherent Exascale software system. Such integration could naturally be undertaken by the base CS research program at the DOE Labs but this requires maintaining their core base of research expertise.

From our examination of the present research portfolio, the COV thought that the areas of energy-efficient computing, machine learning, and data analytics are under-represented, although there are a few individual projects that address these issues. These are all important topics in Exascale research. For example, the use of machine learning algorithms will likely be essential in dynamically scheduling tasks on heterogeneous Exascale platforms, and also in auto-tuning applications for such platforms. Powerful data analytics tools will be needed to extract information from the enormous amount of data generated by Exascale simulations.

Recommendations:

-
- ASCR's CS program should maintain an appropriate balance between funding for the Exascale research program and the base CS research program at the DOE labs.
- More prominence should be given to research into energy-efficient computing, machine learning and data analytics in future solicitations since these topics are important more generally than just in the context of the Exascale initiative.

Charge 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 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?

With the end of our ability to increase the clock speed of silicon microprocessors in around 2005, multi-core processors are now the norm and exploitation of parallelism is essential for increased performance. Modern supercomputers incorporate multi-core chips, complex memory hierarchies and high bandwidth, low latency interconnection networks. Increasingly, more complex options incorporating GPGPUs are becoming attractive to application programmers and there is a need to improve the software tools and architectures to better exploit the potential of these hybrid systems. This is the subject of the first question.

The second question is concerned with the vast volumes of data generated both by supercomputer simulations and by experimental facilities. This involves a variety of approaches involving hardware systems and memory storage architectures. In addition, the extraction of meaningful scientific information from such large datasets requires new tools for both visualization and analysis. This importance of this component of the ASCR CS mission is underlined by the recent charge to ASCR 'to assemble a sub-committee to examine the potential synergies between the challenges of data-intensive science and exascale.'

Findings:

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. FOA 10-255 on Advanced Architecture also has some relevance: of the 6 funded projects, 5 are concerned with hardware-software Co-Design systems and memory architectures, and one is explicitly concerned with the energy implications of Exascale architectures.

FOA 10-256 was concerned with Scientific Data Management and Analysis at Extreme Scale. The call resulted in 37 project proposals and 10 projects were recommended for funding. The awards covered the following areas: File Systems, IO and Storage ($2.0M); Triage and Analysis ($1.3M); Visual Analytics ($0.8M); Integration ($0.5M); Knowledge Representation and Machine Reasoning ($0.4M). While the COV welcomes the introduction of a wider pool of expert reviewers it found surprising that none of the reviewers for this call were from any of the DOE Laboratories. It would seem that a better balance of reviewers including some with experience and knowledge of DOE's mission and practices would be desirable.

A workshop on "Exascale Data Management, Analysis, and Visualization" was held in February 2011 and a joint workshop with BES on "Data and Communications in Basic Energy Sciences: Creating a Pathway for Scientific Discovery" took place in October 2011. In addition to the FOA 10-256 portfolio of 'data projects', another dozen data projects were presented at the Exascale Research Conference in Portland in April 2012.

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. The majority of the present research projects around data are more focused on data management than data analytics. The Machine Learning/Data Mining agenda does not seem strong at the DOE Labs and is also not represented in the Applied Mathematics program. The CS program in ASCR should seriously consider initiating such a research program on behalf of the whole of the Office of Science.

Recommendations:

- Review panels should ideally contain a mix of external and DOE Laboratory researchers.
- In its scientific data management and analysis program the CS program should work with the BES and BER experimental data communities as well as ASCR's traditional simulation and modeling community.
- 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.

Charge 2 (c): Within the boundaries defined by DOE missions and available funding, comment on how the award process has affected 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.

The previous COV report on the CS program included the following recommendations:

"The program should continue its leadership in high end computing and expand its collaborations broadly with the international community."

"The program should launch strategic initiatives in all mission relevant aspects of data-intensive computing, data management and analysis."

The 2009 COV report also noted that:

"The CS program should continue its support and leadership of the International Exascale Software Project and its initiatives in support of new tools, architectures and technologies to support Exascale systems. In addition, the ASCR program should consider taking on a similar international leadership role in database management, visualization, mining and curations of the multi-Petabyte heterogeneous data sets generated by experiments and simulations."

Findings:

In contrast to the previous COV's recommendations, this committee was informed that the IESP has been judged to have come to the end of its usefulness. In the future, the intention is now to have an annual workshop in the US to which the international Exascale community will be invited. The justification for this decision was not apparent to this committee and is a deliberate reversal of the previous, more collaborative approach.

A result of the present uncertainty in funding for the ASCR's Exascale initiative, there is a danger that the schedule for delivery of an Exascale computing system could be significantly delayed. However, the COV notes that other countries, including Japan and China, have adopted an aggressive timescale and constitute a significant, 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.

There was no evidence of consideration of a similar international collaborative approach to the problems of 'Big Data' – in terms of management, visualization, mining and curation – as that evidenced by the International Exascale Software Project.

Recommendations:

- ASCR should do all that it can to ensure that it receives sufficient funding for the Exascale initiative 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.

4.0 Overarching Observations and Summary

Findings:

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 so far only been progress in allocating one new FTE to these vacancies.

Recommendations:

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