Department of Energy
Advanced Scientific Computing Advisory Committee
Exascale Computing Initiative Review

August 2015

1. Introduction and Overview

On November 19, 2014, the Advanced Scientific Computing Advisory Committee (ASCAC) was charged with reviewing the Department of Energy's conceptual design for the Exascale Computing Initiative (ECI). In particular, this included assessing whether there are significant gaps in the ECI plan or areas that need to be given priority or extra management attention. ASCAC chair, Professor Roscoe Giles, in turn, requested ASCAC member, Professor Daniel Reed, to lead a subcommittee to conduct the review. The subcommittee included the following members:

- Professor Daniel Reed, University of Iowa (subcommittee chair and ASCAC member)
- Professor Martin Berzins, University of Utah (ASCAC member)
- Dr. Robert Lucas, Livermore Software Technology Corporation
- Professor Satoshi Matsuoka, Tokyo Institute of Technology
- Dr. Robert Pennington (retired)
- Professor Vivek Sarkar, Rice University (ASCAC member)
- Professor Valerie Taylor, Texas A&M University

Any advanced scientific computing plan builds on a broad array of other high-performance computing research, development, and deployment programs. It also rests on a commercial ecosystem of hardware, software and application vendors, a talented and trained workforce, and a global community of computational science researchers and developers. The DOE ECI is no exception. It has antecedents in the long history of advanced computing at DOE, cognate activities at other U.S. agencies, the private sector, and both international collaborators and competitors. The European Union, Germany, Japan, and China are all pursuing advanced computing initiatives.

With this backdrop, the review drew on conversations with key staff from the Office of Science (SC) and the National Nuclear Security Administration (NNSA) at the Department of Energy and its laboratories, as well as members of the international high-performance computing community. In addition, the subcommittee reviewed the November 2014 *Preliminary Conceptual Design for an Exascale Computing Initiative*[1] and its appendix (*ExaRD Detailed Technical Descriptions*[2]), as well as previous reports from DOE and other U.S. agencies on the proposed exascale initiative. The subcommittee also reviewed European and Asian high-performance computing program plans, selected research papers, and public presentations.

These materials provide a wealth of technical material on scientific applications, national needs and opportunities, economics and competitiveness, as well as the technical challenges of

1 *Preliminary Conceptual Design for an Exascale Computing Initiative, http://science.energy.gov/~/media/ascr/ascac/pdf/meetings/20141121/Exascale_Preliminary_Plan_V11_sb03c.pdf, November 2014*

2 *ExaRD Detailed Technical Descriptions, http://science.energy.gov/~/media/ascr/ascac/pdf/meetings/20141121/ExaRD_Appendix_v5-141113.pdf, November 2014*

scaling, resilience, and efficiency in algorithms, applications, system software and runtime systems, computer architecture and hardware. Given the breadth and depth of previous reviews of the technical challenges inherent in exascale system design and deployment, the subcommittee focused its assessment on organizational and management issues, considering technical issues only as they informed organizational or management priorities and structures.

This report presents the observations and recommendations of the subcommittee. Ranked in order of importance, from highest to lowest, they are:

1. Develop a detailed management and execution plan that defines clear responsibilities and decision-making authority to manage resources, risks, and dependencies appropriately across vendors, DOE laboratories, and other participants.
2. Unlike other elements of the hardware/software ecosystem, application performance and stability are mission critical, necessitating continued focus on hardware/software co-design to meet application needs.
3. As part of the execution plan, clearly distinguish essential system attributes (e.g., sustained performance levels) from aspirational ones (e.g., specific energy consumption goals) and focus effort accordingly.
4. Mitigate software risks by developing evolutionary alternatives to complement more innovative, but untested alternatives.
5. Remain cognizant of the need for the ECI to support both data intensive and computation intensive workloads.
6. Given the scope, complexity, and potential impact of the ECI, conduct periodic external reviews by a carefully constituted advisory board.
7. Where appropriate, work with other federal research agencies and international partners on workforce development and long-term research needs, while not creating dependences that could delay or imperil the execution plan.

Finding: *Like any ambitious undertaking, DOE's proposed ECI involves some risks, as it seeks to develop and deploy advanced computing technologies and systems of unprecedented scale. Despite the risks, the benefits of the initiative to scientific discovery, national security and U.S. economic competitiveness are clear and compelling. The subcommittee strongly endorses the DOE plan for exascale computing development and deployment.*

The subcommittee believes the ECI is a well-crafted plan designed to meet DOE mission needs while also advancing broader national security and competitiveness goals, and the DOE has a successful record of managing complex projects of this type. Thus, the subcommittee's recommendations should be viewed as suggestions for enhancements to the plan, rather than criticisms.

2. Recommendations

2.1. Management and execution plan

This ambitious program brings together the Department of Energy's Office of Science (SC), its National Nuclear Security Administration (NNSA), nine national laboratories, and vendor partners over an eight-year period. Furthermore, the ECI will unfold concurrently with, and interact with, many other DOE computational science programs. Therefore, resource management, execution planning, and risk management are essential for success.

Recommendation: *Develop a detailed management and execution plan that defines clear responsibilities and decision-making authority to manage resources, risks, and dependencies* *appropriately across vendors, DOE laboratories, and other participants. The present joint leadership structure spanning SC and NNSA has been effective in articulating a coherent vision for the future. For this high-level vision to be translated into a realizable and detailed execution plan, the subcommittee recommends that DOE establish a leadership structure that operates below and in concert with the present, high-level leadership at DOE headquarters. This leadership structure's sole focus should be the exascale program.*

Structurally, this leadership structure could take several forms, ranging from a small management group to an overall leader – one person for whom this is their primary, full-time responsibility. Regardless, there must be a clear chain of command and well-articulated mechanisms for identifying and resolving issues.

The initial focus of this leadership activity should be the creation of a detailed execution plan for project activities and deliverables, with adaptation based on evolving technologies and risks. It should include resource scoping and responsibilities, including availability of additional personnel to meet program needs. *Because the program includes both research as well as deliverable hardware and software, DOE should develop a formal risk and assessment plan in concert with detailed execution planning,*

The broad sweep and long timescale of the ECI will undoubtedly require additional project management staff to ensure activities remain on track. In addition, the project must also ensure that the pool of researchers and developers is adequate and that the distributed team has the skills and experience to address the entire scope of exascale hardware, software, and application challenges.

During the lifetime of the project, new computing technologies will undoubtedly emerge. Indeed, developing some of those technologies is one of the ECI's objectives. *However, the primary rationale for the ECI is the new scientific discoveries and technical capabilities it will enable.* DOE must quantify what that means, ensuring there are credible application and discovery measures for the success of the ECI. Simply stating that the throughput of a benchmark, or even an application ensemble, is one hundred times that of a current system (e.g., ORNL's Titan) is not adequate.

The ECI will necessarily require management adaptation and risk assessment as technologies mature and project milestones are reached or adjusted based on new insights. These adaptations will be particularly important in addressing issues such suggested limits on power (20 MW) and total memory capacity (128 PB). Aspirational visions are motivating, but they should not be non-negotiable requirements, particularly when technology is shifting rapidly. Rather, there must be a well-defined process to monitor technology developments, potential risks and benefits; careful co-ordination across stakeholders; and rigorous assessment of project priorities and directions --- all of which must be viewed through the lens of the ECI's primary goal of enabling new scientific discoveries and technical capabilities.

Recommendation: *Given the scope, complexity, and potential impact of the ECI, conduct periodic external reviews by a carefully constituted advisory board. Such a board may also be useful in resolving complex resource issues and in assessing research and development risks when technology is changing quickly. It can also help define success metrics.*

2.2. PARTNERSHIPS, COLLABORATION AND WORKFORCE DEVELOPMENT

There are broad and extensive opportunities for cooperation among the federal agencies that depend on advanced computing. For example, DOE's *Preliminary Conceptual Design (PCD)* document highlights three major computational application areas that are essential to the DOE mission needs: (a) computational materials science, (b) next generation climate models and (c) nuclear stockpile stewardship. The PCD document also identified four primary technical challenges that must be addressed for exascale systems to meet the needs of DOE's mission: parallelism, resilience, energy, and memory/storage.

The February 2015 DOE ASCAC report investigated these and other technical issues in more depth, identifying a "top 10" list of research challenges[3]. The ASCAC report also identified the need for long-term partnerships among the stakeholders, including government agencies, academia and vendors, to address these fundamental requirements, as derived from science and mission needs.

History has shown that the required advances in hardware, architecture, algorithms, software and tools will not arise holistically and solely from a single agency, a vendor or a program, regardless of focus and mission. Instead, they accrue from extensive collaboration and community buy-in across government laboratories, industry, and academia. The DOE ECI can significantly influence the broader advanced computing research and development directions that will contribute to the effective co-design of the exascale system. It can also benefit from collaborations that advance those developments.

Interagency Collaboration and Workforce Development

Purely open-ended research programs are not compatible with DOE's mission need for a targeted, outcome-driven execution plan that contains clear deliverables and integration. The current, *Preliminary Conceptual Design* document notes only that the exascale computing initiative "may be expanded to include R&D efforts in other federal agencies." The subcommittee encourages such collaborations.

The Department of Defense (DOD) has a mission-driven interest in exascale computing similar to DOE's. In areas such as chemistry, materials science, and engineering, there is a great deal of overlap. DOE should endeavor to build a strong collaboration with relevant DOD organizations.

Another example of possible research collaboration is the development of new mathematical algorithms that minimize data movement and maximize resilience, as recommended by the previous ASCAC report. Many of the exascale programming challenges relate to the complexity of distributed data structures and code tuning for specific algorithm and architecture combinations. The management and analysis of very large, multi-dimensional data sets is another such example. These computing research problems overlap with fundamental research supported by the National Science Foundation in the areas of advanced computing and data systems.

There are previous and extant examples of interagency collaborations that can serve as models for a joint understanding of scientific requirements and approaches. These include projects in

[3] Top Ten Exascale Research Challenges, DOE ASCAC Subcommittee Report, http://science.energy.gov/~/media/ascr/ascac/pdf/meetings/20140210/Top10reportFEB14.pdf, February 2014

astronomy (e.g., the Dark Energy Survey and the Large Synoptic Survey Telescope) as well longstanding joint efforts in high-energy physics. Successful, cooperative projects are driven by mission needs and outcomes, and the ultimate success of the DOE exascale effort will depend on creation of an ecosystem that enables those outcomes.

Finally, the training, workforce development and educational issues for advanced computing extend beyond the scope of existing DOE computing activities and DOE's need address the hardware and software requirements specific to the ECI. Ideally, such a training and development program would focus on developing the scientific and technical expertise at the very large parallel scales needed for successful exascale platform development and deployment. This will require specific additional recruitments and the ability to compete successfully in the commercial workforce marketplace, an increasing challenge given the importance of information technology.

Given the wide range of technical challenges, the subcommittee encourages DOE to develop plans for interagency research collaborations and mechanisms to incorporate salient research results, while not creating dependencies that could delay or imperil its execution plan. More broadly, interagency research collaborations would expand and accelerate development of a highly trained and flexible workforce that is aware of, contributing to and utilizing exascale systems. This will be essential to the on-going and long-term success of the DOE computing based research programs. Early career programs, including highly promising graduate students and faculty, could be one area for a strong initiative in workforce development with multi-agency cooperation.

International Collaborations and Community Building

Although Europe's plans for post-petascale computing are yet to crystalize, Germany's yearly procurement program is likely to move in the exascale direction through its largest computers. Consistent funding levels at about 400M euros will allow the development of an internationally competitive computing infrastructure. Currently, the German program is strongly academic and industrially focused, targeting better performance from traditional architectures (i.e., Intel or x86), with less emphasis on novel and more energy efficient architectures. However, the German program could yield interesting collaborations on algorithmic and software fronts and provide an important reference point for the DOE program.

In addition, there are exascale projects of potential relevance in the European Union's Horizon 2020 initiative, notably the DEEP-ER and Mont-Blanc 2 activities. Also, the Horizon 2020 "Flagship" $1 billion euro/10 year Human Brain Project has announced the 2017 acquisition of a pre-exascale platform at the Juelich supercomputing center.

Japan is currently planning a domestically designed and built pre-exascale "post-K" supercomputer to succeed the highly successful "K-computer." Designed and built by Fujitsu, the K-computer has been operated by AICS (Advanced institute for Computational Science) as Japan's flagship supercomputer. The post-K system will also be built by Fujitsu and is expected to include advanced technologies such as 3D-stacked memory, an optical interconnect, and aggressive power saving techniques to meet its 30 MW target. In June 2014, DOE and Japan's Ministry of Education, Sports, Culture, Science and Technology (MEXT) signed an agreement to collaborate on system software for future exascale systems.

Finally, as an indicator of international competition and its global implications, China is embarked on an ambitious program to build a successor to its Tianhe-2 system. Tianhe-2 is currently ranked first on the Top500 list of the world's fastest computing systems, and its successor is expected to be based on chips developed within China. This has substantial implications for the global marketplace and national security.

The subcommittee believes there are additional opportunities for bilateral and multilateral international collaborations to ensure development of consistent and interoperable software ecosystems and applications; such collaborations could accelerate the transition of scientific applications to new hardware and software. It encourages DOE to foster those collaborations, while not creating dependences or risks from international collaboration.

Recommendation: *Where appropriate, work with other federal research agencies and international partners on workforce development and long-term research needs, while not creating dependencies that could delay or imperil the execution plan.*

2.3. SOFTWARE AND TOOLS

Exascale software development has two distinct goals. The first is creation of a functioning software ecosystem that allows applications to execute at scale as quickly as possible, with minimal change. The second is more complex and challenging -- shifting the software base to post-petascale architectures and ensuring broader uptake and use of exascale systems. In both cases, the scale of the proposed systems brings new challenges in resilience, parallelism, energy efficient algorithms and software, and memory and data management.

As previous ASCAC studies have noted, extreme scale brings resilience and reliability challenges due to the large number of system components and continuing shrinkage of semiconductor feature sizes. In turn, the need for resilience has deep and profound implications for every software layer, from operating systems through runtime systems and communications protocols to applications. It is likely that several, if not all software layers, will need to identify hardware failures and take appropriate error correction steps, or fail gracefully when recovery is not possible.

In addition, the need to rethink programming models to reflect the scale, relatively slow speeds and large energy costs of communication, and the potential fragility of the future systems requires investigation of new programming models and middleware. In the limiting case, it may require a complete reconceptualization of parallel programming. Although such new models appear to be appropriate for development of future codes, the need to migrate existing codes to new architectures requires evolutionary approaches also be investigated. These may include introduction of portability layers that allow computationally intensive code sections to be mapped onto different architectures or possibly domain-specific languages.

Addressing the technical challenges while supporting both evolutionary and revolutionary software approaches will require collaboration among academic researchers, national laboratory staff and industrial software teams, with appropriately funded and structured research and development programs. DOE laboratories and their vendors have extensive experience in delivering software for new machines, and current and projected system procurements provide a basis for planning and execution. However, the challenges of exascale software development are deep and substantive.

Recommendation: *Mitigate software risks by developing evolutionary alternatives to complement more innovative, but untested alternatives. This would ensure that applications have both an evolutionary and a revolutionary path to exascale execution.*

2.4. HARDWARE AND ARCHITECTURE

As noted earlier, the primary goal of the ECI is meeting next generation DOE mission objectives of scientific discovery, engineering design, and national security. Because device technologies and architectures are rapidly evolving, it is very important to analyze architectural performance targets such as total power consumption with respect to application requirements, while carefully assessing the limits of technology evolutions, and adapting accordingly. This is the essence of effective co-design.

Several disruptive opportunities such as UV lithography, multiple variants of non-volatile memory, aggressive 3-D memory stacking, and advanced photonics are on the horizon. This may allow exascale designs to exceed some current design targets by substantial margins. Conversely, the uncertainty in other areas may make certain targets infeasible within the project timeframe. Due to such unpredictability, it is important to distinguish the irreducible requirements derived from applications from aspirational project goals, and iteratively refine outcomes and approaches based on research and development outcomes, lest undesirable consequences arise.

For example, certain target numbers such as 20 MW power consumption and 128 PB memory capacity may be difficult to attain simultaneously. The DRAM power reduction ratio over time is predicted to be rather small, as device characteristics challenge further lithography and voltage reductions. If target numbers are publicized and shape activities prematurely, there is a danger that the project will be perceived as a failure for not reaching merely aspirational objectives.

As a precursor to the ECI, DOE has already launched research and development projects with multiple vendors as part of its “Forward” projects. In addition, there are staged deployments of multiple systems (i.e., Trinity/Cori, CORAL, and APEX) with diverse architectures that form milestones along the path to exascale operations. The architectural properties and target performance of the development projects and future intermediate systems should be continuously reviewed and possibly altered based on application and technology assessments.

Finally, it is important to recognize that a second, important purpose of the ECI is advancing the economic and technical competitiveness of the U.S. information technology industry. Unless there are commoditizable technologies capable of meeting DOE mission objectives, the technologies developed and the machine architectures deployed risk becoming so exotic and/or expensive such that they are only applicable to a small number of large machines.

Rather, a broad ecosystem should result from the project, with competitive downscaling and commoditization that can be adopted widely across the computing community. Such an ecosystem should ideally be capable of delivering systems with an appropriate level of performance on real applications, including those with smaller numbers of floating point operations per byte, rather than embracing impressive peak performance that is unattainable by real applications. Failure to create such a broad ecosystem will very likely disincentivize both the users and the vendors, and as a result, will fail to leverage their mainstream research and development efforts, ultimately resulting in fewer technological advances and lower overall performance.

At a point when many new technologies components are still maturing, the ECI must not commit prematurely. ECI systems should build on the successes of existing programs, both to reduce risk and to minimize disruptions in the execution model presented to scientific applications. Wherever possible, the ECI should leverage commercial technology and collaborate with other

U.S. and International partners to minimize the cost and risks associated with exascale development and deployment.

Recommendation: *As part of the execution plan, clearly distinguish essential system attributes (e.g., sustained performance levels) from aspirational ones (e.g., specific energy consumption goals) and focus effort accordingly.*

2.5. ALGORITHMS AND APPLICATIONS

The goal of the ECI is not to just to produce exascale systems, but rather to also enable exascale scientific discoveries, engineering design and national security. Therefore, the associated applications must be drivers of ECI planning and execution. The subcommittee is pleased that the ECI has adopted a co-design process, as recently embraced by DOE.

Application-driven co-design will be critical to ensuring that exascale systems are tailored to the needs of the problems most critical to DOE's mission. From its outset, the ECI should identify a set of these mission-critical applications from its target domains (computational materials science, next generation climate models, stockpile stewardship) and make them yardsticks against which exascale systems are evaluated.

Recommendation: *Unlike other elements of the hardware/software ecosystem, application performance and stability are mission critical, necessitating continued focus on hardware/software co-design to meet application needs.*

With the end of Dennard scaling,[4] the number of arithmetic logic units and processor cores in high-performance computing systems is growing exponentially, while the power density is no longer constant. Hardware concurrency levels are expected to reach $O(10^9)$ (i.e., billion way) at exascale. Achieving this extraordinary level of concurrency will be extremely challenging for most scientific applications, and DOE research programs must focus on algorithms, tools, and techniques to support application concurrency scaling. Not only must tools and techniques identify this concurrency, perhaps three orders-of-magnitude more than that realized today, they may also need to contribute to the solution of energy efficiency and resilience challenges expected at exascale, programming concepts for which today there may only be research prototypes.

Examples of the changes required in algorithms abound. New algorithms that trade increases in arithmetic and logical operations for reduced data movement, such as higher-order finite elements, need to be considered. It likely will be necessary to have communication-avoiding algorithms. Stores may have to be minimized to avoid damaging persistent memories, such as NAND FLASH. Likewise, a renewed emphasis on strong scaling in algorithms will be required to meet performance goals.

As noted previously, the architecture of high-end systems is changing at a pace not seen in the last twenty years, and the future will likely see different families of exascale systems with very different properties. As noted in the software and tools discussion, this will require new programming abstractions, as languages based on half-century old machine models are increasingly incapable of being mapped efficiently to newer systems.

[4] Dennard scaling is a MOSET scaling rule proposed in a 1974 paper by Dennard *et al.* In its simplest form, it states that as transistors shrink, their power density remains constant and power use is proportional to chip area. It broke down in 2004-2007, leading to today's multicore chips.

The ECI must develop new programming abstractions that enable successful use of the attributes of the new systems. New concepts could include processor affinity, or affinity of tasks to specific data objects, as well as non-uniform memory access (NUMA) policies. They could also include hints for the runtime system regarding energy, resilience, and performance optimization. Domain specific languages and architecture-aware libraries might be ways to separate the specification of a scientific problem from its realization on a specific exascale platform, allowing for code and performance portability in an increasingly diverse high-end environment.

The ECI must also provide an evolutionary path forward for existing, previously validated scientific codes. Not only would it be inefficient to recreate working code, the applications must be validated to regain the trust of the user community. New abstractions, such as task-based concurrency scheduled at runtime, need to be interoperable with the today's predominant execution model, realized as MPI+X, where X includes Pthreads, OpenMP, OpenACC and CUDA. While part of this gap may be bridged by advances in standards to include, say, task-based approaches, scientists should be able to evolve their algorithmic and coding investments.

The ECI must also develop a new generation of tools to help scientific software developers and users understand application behavior, and learn how to debug, optimize, and use them effectively. This challenge will be exacerbated by systems that dynamically adapt to energy and faults, changing clock speeds and migrating work. It will likely be impossible to perform repeatable timing experiments, and already there can be more threads in an application than there are pixels on a display, making it hard to imagine how to convey information on application and system behavior to developers, users, and system operators.

Further, a new generation of data analytic tools and libraries are needed to aid in the interpretation and validation of the data generated from exascale applications. Much more data will be generated, making it important to explore possible *in situ* analytics or other methods for rapid analysis and visualization. More broadly, this includes end-to-end workflows that encompass verification and validation (V&V), as well as uncertainty quantification (UQ).

Although performance is an important metric for evaluation of methods and systems, scientific productivity is also crucial. For example, with climate modeling, one important metric of productivity is the number of years of climate change that can be generated in a day. For each of the application drivers, the application scientists will need to identify appropriate productivity metrics.

Finally, the developers of mathematical libraries and scientific applications will need early access to prototypes of exascale systems. This is necessary to test new algorithmic concepts and verify that they perform as anticipated on exascale systems when they emerge. Analysis of the successes, and failures, achieved on these early systems will also provide feedback to the designers of the exascale systems, who will be working in their own co-design process.

The subcommittee recommends that the ECI focus on the scientific discoveries and technical capabilities that will be enabled by exascale systems, and use those to establish its success metrics. In turn, this will require advances in scientific applications, their mathematical algorithms, and co-design of systems that can support them. Suitable enhancements to programming abstractions and tools will also be needed to make development of these applications tractable. The developers of the exascale applications will need to be supplied with early examples of systems to demonstrate that both the applications and the systems are making adequate progress toward exascale capability.

3. Summary

The benefits of investments in advanced computing technologies and applications extend far beyond the core missions of DOE's SC and NNSA. Dramatic decreases in computing costs, coupled with equally dramatic increases in performance and capability, have transformed our world. Computing technologies and applications have reshaped business, communications and entertainment, scientific discovery, medicine and healthcare. By any measure, computing has been responsible for a substantial fraction of U.S. economic growth over the past forty years.

In addition to computational models of ever-greater resolution, fidelity, and predictive power, advanced computing increasingly incorporates advanced data analytics. The latter arises from both computational modeling and from scientific and environmental sensors. Modeling and data analysis are inextricably intertwined enablers of innovation and discovery; both draw on the same ecosystem of hardware and software technologies; and both are crucial elements of DOE's ECI.

Recommendation: *Remain cognizant of the need for the ECI to support both data intensive and computation intensive workloads.*

Exascale computing is not a destination, but rather the next milestone in a journey toward greater computing capability and the scientific discoveries and national capabilities it will enable. Therefore, DOE must also support and conduct research to anticipate and shape the high-performance computing field beyond exascale levels. The DOE research portfolio must be balanced and include both the near-term and the longer-term goals.

Finding: *Given the centrality of advanced computing to scientific discovery, economic competitiveness, and national security, it is critical that the U.S. continue to maintain leadership in both enabling computing technologies and in their applications to problems of national importance. Thus, the subcommittee believes the proposed DOE ECI is an important element of ensuring the continued national security and economic competitiveness of the United States.*

In summary, the subcommittee strongly endorses the DOE plan for exascale computing development and deployment. It is a well-crafted plan designed to meet DOE mission needs while also advancing broader national security and competitiveness goals.