DOE ASSESSMENT
SEAB Recommendations Related to High Performance Computing

1. Introduction

The Department of Energy (DOE) is planning to develop and deliver capable exascale computing systems by 2023-24. These systems are expected to have a one-hundred to one-thousand-fold increase in sustained performance over today's computing capabilities, capabilities critical to enabling the next-generation computing for national security, science, engineering, and large-scale data analytics needed to advance the Department's Office of Science and Energy and the National Nuclear Security Administration (NNSA) missions. Leadership in high-performance computing (HPC) and large-scale data analytics will advance national competitiveness in a wide array of strategic sectors. The U.S. semiconductor and high-performance computing (HPC) industries are poised to develop the necessary technologies for an exascale computing capability early in the next decade. However, an integrated government-industry-academia approach to the development of hardware, system software, and applications software, will be required to overcome the barriers of power efficiency, massive parallelism, and programmability in attaining maximum benefit from exascale computers.

The Department's exascale initiative plan, which has been in development since 2011, partners the Office of Science with the Office of Defense Programs in the National Nuclear Security Administration (NNSA), to coordinate and jointly conduct development of the required exascale technologies research and development (R&D), including applications software, and to acquire and deploy two or more exascale computers in the next decade. Seven DOE national laboratories (Argonne, Lawrence Berkeley, Oak Ridge, Pacific Northwest, Livermore, Los Alamos, and Sandia) are significantly involved in this exascale effort.

2. SEAB Review

In December 2013, Energy Secretary Ernest Moniz charged the Secretary of Energy Advisory Board (SEAB) to establish a Task Force to "*review the mission and national capabilities related to next generation high performance computing.*" Further, the Secretary charged that the Task Force "*examine the challenge problems and opportunities that drive the need for next generation high performance computing, as well as the advances and necessary steps to create and execute a successful path that will deliver next generation computational performance*" and that their "*report should include recommendations on whether and to what degree the U.S. Government should lead and accelerate the development of next generation high performance computing applications and systems.*"

In August 2014, the Task Force submitted its report to the SEAB. It contained six summary recommendations:

1. DOE, through a program jointly established and managed by the NNSA and the Office of Science, should lead the program and investment to deliver the next class of leading edge machines by the middle of the next decade. These machines should be developed through a co-design process that balances classical computational speed and data centric memory and communications architectures to deliver performance at the 1-10 exaflop level, with addressable memory in the exabyte range.
2. This program should be executed using the partnering mechanisms with industry and academia that have proven effective for the last several generations of leadership computing programs. The approximate incremental investment required is $3B over 10 years. This would include a roadmap of DOE acquisitions, starting with the CORAL program. Such a roadmap would focus industry on key system level deliverables.
3. DOE should lead, within the framework of the National Strategic Computing Initiative (NSCI), a co-design process that jointly matures the technology base for complex modeling and simulation and data centric computing. This should be part of a jointly tasked effort among the agencies with the biggest stake in a balanced ecosystem.
4. DOE should lead a cross-agency U.S. Government (USG) investment in "over-the-horizon" future high performance computing technology, including hardware, software, applications algorithms, operating systems, data analytics and discovery tools, agent based modeling, cognitive computing, neurosynaptic systems, and other forward looking technologies, including superconducting computing.
5. DOE should lead the USG efforts to invest in maintaining the health of the underlying balanced ecosystem in mathematics, computer science, new algorithm development, physics, chemistry, etc. but also including Independent Software Vendors (ISVs), the open source community, and other government entities.
6. The Path Forward requires operating in, and investing for, three timeframes and technology plateaus: (1) The greater Petascale timeframe (the next five years), (2) The Exascale timeframe (the next five to 10 years), and (3) Beyond Exascale. (SEAB commented: We note that the combined DOE investment in maintaining a healthy ecosystem and pursuing "over-the-horizon" technology identification and maturation is in the range of $100-150M per year.)

In October 2014, Secretary Moniz directed three follow-on questions to the SEAB Task Force:

1. The SEAB Task Force on High Performance Computing identified costs for a DOE exascale and beyond program and certain technology and engineering developments that would be met. Can the Task Force provide an additional level of granularity on the allocations against major technology areas and their timing that regulate success in getting to the 1-10 exascale range in a decadal timeframe?
2. The SEAB Task Force on High Performance Computing recognized that barriers exist to broadening U.S. industrial adoption of high-end HPC. The Deep Computing Solutions Center (DCSC) represents an initial DOE laboratory effort to address this problem. Could

the Task Force provide suggestions for what the Department could undertake to expand industrial high-end HPC use? This could include scaling up DCSC-like approaches or entirely different ways of using DOE laboratory expertise towards the stated goal.

3. Finally, any further thoughts on how a beyond-exascale research program (superconducting, quantum, neuromorphic…) might be structured would be appreciated.

In November 2014, the SEAB Task Force responded to the above questions.

This report summarizes the Department's analysis and responses to the recommendations provided by the SEAB resulting from the original charge and follow-on questions.

3. DOE Assessment and Response to Recommendations

The SEAB's 2014 report called on the Department to aggressively pursue next generation computing both for exascale and for the generation of technologies that will be required beyond Moore's law. Recommendations from the report and actions being undertaken by DOE are highlighted below.

SEAB Recommendation	DOE Assessment and Actions
1. **Management**. DOE, through a program jointly established and managed by the NNSA and the Office of Science, should lead the program and investment to deliver the next class of leading edge machines by the middle of the next decade. These machines should be developed through a co-design process that balances classical computational speed and data centric memory and communications architectures to deliver performance at the 1-10 exaflop level, with addressable memory in the exabyte range.	DOE concurs. Over the last six years, DOE Office of Science and the NNSA Office of Defense Programs have jointly pursued a program to develop and deliver exascale-class computing for DOE science and national security missions. This program has pioneered the use of "co-design," which exploits collaboration between DOE/NNSA labs and computer-vendor industries to develop hardware and software that optimally executes DOE/NNSA applications. A central goal of the DOE/NNSA program is to achieve first use of capable exascale computers in the international context, with performance in the 1-10 exaflop range with memory capacity approaching an exabyte.
2. **Partnering and Cost.** This program should be executed using the partnering mechanisms with industry and academia that have proven effective for the last several generations of leadership computing programs. The approximate incremental investment required is $3B over 10 years. This would include a roadmap of DOE acquisitions, starting with the CORAL program. Such a roadmap would focus industry on key system level deliverables.	DOE concurs. The DOE/NNSA exascale program emphasizes partnerships with industry and academia for the development of the needed hardware and software technologies. The recent CORAL joint NNSA-SC partnership will acquire pre-exascale systems in the 2018-19 timeframe, which will be followed by joint acquisitions of exascale computers in the 2023-24 timeframe. DOE/NNSA are refining the budget profile for the exascale project in preparation of the FY 2017 budget. The approximate incremental investment of $3B, cited by SEAB is consistent with current DOE/NNSA understanding of the funding level required to deliver exascale systems in the 2023-24 timeframe.
3. **Co-Design.** DOE should lead, within the framework of the National Strategic Computing Initiative (NSCI), a co-design process that jointly matures the	DOE concurs. DOE is coordinating with the interagency to align investment strategies to meet broader federal goals.

technology base for complex modeling and simulation and data centric computing. This should be part of a jointly tasked effort among the agencies with the biggest stake in a balanced ecosystem.	
4. **Beyond Moore's Law.** DOE should lead a cross-agency U.S. Government (USG) investment in “over-the-horizon” future high performance computing technology, including hardware, software, applications algorithms, operating systems, data analytics and discovery tools, agent based modeling, cognitive computing, neurosynaptic systems, and other forward looking technologies, including superconducting computing.	DOE generally concurs. Since January 2014, DOE has been conducting a process that will result in plans for R&D on next-generation computer hardware and software technologies that look past the anticipated end of progress in CMOS technology improvements. R&D under consideration comprises quantum, neuromorphic, and probabilistic computing and includes underpinning applied mathematics, computer science, materials development and fabrication, information science, and quantum theory. In February-June 2015, DOE/NNSA is conducting a series of research needs workshops to inform the FY 2017 budget process. In conjunction, DOE co-leads (with NSF and NIST) the newly constituted NSTC Quantum Information Science Interagency Working Group, which will publish its findings in late 2015. With regards to superconducting computing technologies, DOE observes that other federal agencies, notably DoD and the IC, emphasize this area of research. DOE is coordinating closely with these OFAs through the OSTP White House process.
5. **Ecosystem.** DOE should lead the USG efforts to invest in maintaining the health of the underlying balanced ecosystem in mathematics, computer science, new algorithm development, physics, chemistry, etc. but also including ISV’s, the open source community, and other government entities.	DOE concurs. Since June 2014, DOE has been coordinating with NSF on investments to promote the health of the U.S. underlying balanced ecosystem, across the areas of applied mathematics, computer science, new algorithm development, and the physical sciences. DOE has initiated in 2014-15 discussions with other federal agencies to promote collaborations in the application of peta and exascale computing, including NIH, NASA, DHS, DOC/NOAA, and DoD.

	With regards to ISVs, DOE observes that, although considerable effort has been expended in recent years in outreach to ISVs, we have found them reluctant to adopt modern, state-of-the-art HPC capabilities, largely due to business constraints. In response to their reluctance, DOE has, in the last three years, adopted a strategy of working with small businesses via the DOE SBIR program to achieve this objective.
6**. Path Forward**. The Path Forward requires operating in, and investing for, three timeframes and technology plateaus: (1) The greater Petascale timeframe (the next five years), (2) The Exascale timeframe (the next five to 10 years), and (3) Beyond Exascale. (SEAB commented: We note that the combined DOE investment in maintaining a healthy ecosystem and pursuing “over-the-horizon” technology identification and maturation is in the range of $100-150M per year.)	DOE concurs. The three timeframes SEAB cites coincide with the overarching strategy DOE has for HPC. In the immediate term, DOE/SC is continuing investments in its own petascale applications and is working closely with the DOE applied energy and environmental management program to apply petascale computing to their respective mission challenges. NNSA is developing the next class of nuclear weapons codes as well as national security modeling and analytics applications. In the 2015-2020 timeframe, DOE/NNSA are supporting development of exascale-class technologies, leading ultimately to deployment of capable exascale computers in the 2023-24 timeframe. To address the future of HPC in 2025 and beyond, DOE/NNSA plan to initiate investments in FY 2016 in “over-the-horizon” technologies, including initially quantum and ultimately neuromorphic and probabilistic computing. Although DOE does not disagree with the funding estimate cited by SEAB, we presently are developing plans and aim to have a validated estimate by late 2015.

4. Secretary's Response to SEAB Regarding the Exascale Task Force Supplemental Recommendations (November 2014)

The November 2014 responses of the SEAB Exascale Task Force to the supplemental questions posed by Energy Secretary Moniz are narrative in form, with some embedded additional recommendations. This Section of the report provides, for each supplemental question, a synopsis of the Task Force response, followed by the DOE analysis and actions.

Question 1: Can the Task Force provide an additional level of granularity on the allocations against major technology areas and their timing that regulate success in getting to the 1-10 exascale range in a decadal timeframe?

Summary of Task Force Response to Question 1: The Task Force response provides a discourse on the need to support both R&D and acquisitions and an estimate of the cost to achieve operational exascale computing. They

- envision a decade-long program will be required to achieve exascale computing, with intermediate steps, including prototypes, along the way;
- estimate that the funding profile for exascale will extend through 2024, with peaks in the range above $350 million per year but less than $400 million per year; and
- note that the cost of individual exascale computers will be in the $200-250M range, which is approximately $90M more than seen in the recent CORAL acquisitions.

DOE Response: Generally DOE agrees with the budget picture summarized in the SEAB Task Force response. Note that we are refining exascale budget estimates to the CD-1/2 level and will have "baseline" quality cost, scope, and schedule later this year.

FY 2016 funding, if appropriated, will enable timely uplift in exploratory hardware investments. This will need to be followed by a similar uplift in FY 2017 for the corresponding software stack R&D, with continuation of hardware and applications investments that are already underway. The current R&D approach encompasses technologies that target the relevant (foreseeable) options (GPU, non-GPU, system on a chip, memory, etc.). To reduce overall project risk, these multiple pathways must be supported until down select to 2 occurs in the 2019-20 timeframe. We are still working to balance the defense and non-defense investments in this program to help meet Recommendations 1-6. DOE notes that the current plan of record would start delivery of the first exascale-class computer in FY 2023, for full operation by FY 2024.

Question 2: Could the Task Force provide suggestions for what the Department could undertake to expand industrial high-end HPC use?

Summary of Task Force Response to Question 2: The Task Force response notes the history of successful HPC partnerships in DOE, with specific mention of the DOE INCITE program and LLNL's HPC Innovation Center. The response describes two broad classes of potential industry users: 1) large industries, for which the use of large scale computational modeling, simulations, and data analysis is standard and does not require “proof of concept” to justify its use, and 2) smaller companies that would benefit from expanded use of DOE HPC but, for reasons of size (people resources), capability (financial resources), or inexperience, do not currently make significant use of high performance computing in their operations. The Task Force itemizes four specific recommendations for DOE to undertake:

1. Create and support an easy to navigate DOE portal describing all publicly available computing resources and programs to access them, tools to determine the best fit to the problems and opportunities presented by the private sector, and clear instructions and guidelines on how to access the resources and programs. Specifically, each DOE facility offering such access should identify and support a single individual to serve as an initial point of contact for a new company wishing to explore access.
2. Continue to support competitive programs that provide access to leading edge HPC computing at the DOE. Ensure that such awards include not only a designated amount of computing time but also ensured access to the computational and domain specific expertise in the labs that support those capabilities. Access to such programs should be through three categories:
    a. Initial awards at no charge to the outside party, with priority put on new users or new applications.
    b. Subsidized access to small and medium enterprises aimed at expanding the national user base of high performance computing.
    c. Follow on awards to large enterprises on a full or partially subsidized pay-as-you-go basis. As an added incentive, such programs might give “credit” to large enterprises that bring new partners to a computational program.
3. As part of the broader enhanced technology efforts underway at DOE, the Department should support programs leading to the commercialization of new or matured codes so that such codes are available through the ISV model to the public user community.
4. The DOE should be a key partner with the university community, the national academic accreditation bodies, and the private sector in enhancing engineering and science degree programs, to ensure that graduates have the necessary background and skills needed for a future in which effective use of high performance computing will be a standard expected capability for a STEM career.

DOE Response: Expanding impact of HPC is a DOE goal aimed at broader utilization of petascale. Some of these efforts are focused on DOE mission areas, such as energy and environmental missions, while others seek to partner with small and large companies to develop next generation tools.

In the near term, we are adapting our requirements identification process to these areas and will have data in time to support FY 2017 discussions. At the same time, we are identifying potential exascale apps and the computer science and applied mathematics R&D required to achieve needed scalability.

We also have given the laboratories concrete performance goals to increase industrial impact of the resources they deploy.

With respect to the specific recommendations contained in the Task Force's response to Question 2,

1) DOE will explore a portal that describes all publicly available computing resources and programs to access them, tools to determine the best fit to the problems and opportunities presented by the private sector, and clear instructions and guidelines on how to access the resources and programs, and will identify and support a single individual to serve as an initial point of contact for companies wishing to explore access. Already, we have web sites for the INCITE program and the LLNL HPCIC. Extension to the full DOE set of HPC activities is straightforward.

2) In large measure, our existing allocation/utilization modes meet the 3 recommendations the SEAB Task Force listed. Two new features are contained in the recommendation: Subsidized access to small and medium enterprises aimed at expanding the national user base of high performance computing and giving "credit" to large enterprises that bring new partners to a computational program. We are exploring the former at the LCFs via the enhanced industry efforts associated with the FY 2016 performance goal, mentioned above. We will explore the latter recommendation starting in FY 2016 (based on planning that will be conducted in FY 2015).

3) We have now stood up a new Office of Technology Transitions (OTT), which can be used to develop software maturation plans. Separately, we are also pursuing software maturation via SBIR projects. Regarding ISVs, we observe they often are unreceptive to incorporation state-of-the-art DOE advances in their products. However, given the disruptive architecture changes we foresee in the shift to exascale, we will reengage the key ISVs to see if new progress can be made in this area.

Question 3: [Your] further thoughts on how a beyond-exascale research program (superconducting, quantum, neuromorphic…) might be structured would be appreciated.

Summary of Task Force Response: The Task Force states its view that there are three promising areas of advanced high performance computing currently being researched: quantum computing, superconducting circuits, and neuromorphic computing. The Task Force recommends

- rather than commit to specific technologies in "over-the-horizon" computational systems, DOE should invest to maintain and strengthen the computational ecosystem, including working with universities, which would allow DOE to understand what already is underway, while focusing on more advanced elements of "over-the-horizon" computing, including software development; and
- combining with the path to exascale, investments to sustain the advanced computing ecosystem and to look "over the horizon" should be funded at $100-$150 million per year. Included in this amount should be $20-$25 million per year to enable DOE to stay abreast of developments being sponsored by others

DOE Response: Both NNSA and SC are actively examining investment options for computing in the post CMOS era, in recognition of the fundamental physical limits imposed by lithographic feature size and Dennard scaling. Options being explored include quantum computing, neuromorphic computing, and probabilistic computing. The FY 2016 DOE budget request for the ECI initiates funding of exploratory next-generation R&D and efforts are presently underway to plan the detailed content of that portion of the ECI. NNSA/SC are keeping abreast of computing based on superconducting devices, both classical and quantum, but are not planning investments at this time because we believe adequate federal investments are being made through other federal agencies, primarily DoD and the Intelligence Community.

In the area of quantum computing, although NNSA and SC do not currently fund research, five of the National Laboratories (Lawrence Livermore, Los Alamos, Oak Ridge, Pacific Northwest, and Sandia) are funding quantum computing R&D through their respective LDRD (Laboratory Directed Research and Development) programs that are in alignment with other federal investments at their sites in this area. In FY 2014, funding for this LDRD work, aggregating across the five laboratories, was approximately $17M. This is comparable to the FY 2014 funding by the National Science Foundation ($15M) and the Air Force Office of Scientific Research ($18M), for which data are available. In preparation for initiating funding of quantum-computing activities FY 2016, NNSA and SC have conducted in FY 2014 and FY 2015 research opportunities workshops that cover quantum computing applications, quantum information science, and underpinning materials science, applied mathematics, computer science, and supporting technologies, such as cryogenics, and photonics (for optical interconnects). A basic research needs workshop for the materials sciences aspects of QIS technologies is planned to occur during the summer months of 2015.

Two other areas of post-CMOS technology under exploration at DOE are neuromorphic computing and probabilistic computing. The former, which is inspired by how the brain works, recognizes that biological circuits are intrinsically more energy efficient microelectronic circuits (albeit considerably slower, too) and process complexity in ways fundamentally different from digital electronic devices. The latter, which was not included in the SEAB recommendations, exploits the fact that future devices with very small feature sizes will encounter quantum effects, extreme variability, low signal to noise, and in general will exhibit a probabilistic nature of operation, rather than the deterministic nature of operation in present-day computers. This probabilistic nature of operation will produce results that are themselves probabilistic, exhibiting a "mean" approaching what would have been the deterministic result, and with a "standard deviation" representing the uncertainty in the result. In order to use computers based on probabilistic technologies, which offer the promise of orders-of-magnitude decreases in energy consumption, new basic research in applied and numerical mathematics is required.

For both of these areas, workshops are planned to occur in 2015, in time to inform the DOE FY 2017 budget submission.