STATEMENT BY

J. STEPHEN BINKLEY

ACTING DIRECTOR, OFFICE OF SCIENCE,

U.S. DEPARTMENT OF ENERGY

BEFORE THE

HOUSE SCIENCE, SPACE AND TECHNOLOGY COMMITTEE,

SUBCOMMITTEE ON ENERGY,

ON

ACCELERATING DISCOVERY: THE

FUTURE OF SCIENTIFIC COMPUTING AT THE DEPARTMENT OF ENERGY

MAY 19, 2021

Thank you, Chairman Bowman and Ranking Member Weber. I am pleased to discuss the scientific computing capabilities of the Department of Energy (DOE), including the forthcoming exascale systems, the implications of these capabilities for other scientific disciplines, and their relevance to pressing societal challenges.

Introduction

DOE computing has deep historical roots, going back to the Manhattan Project, where computers were extensively used, including one of the earliest IBM punch card systems. During the 1950s, John von Neumann, a commissioner of the Atomic Energy Commission (AEC) and a pioneer in computing, advocated for a Mathematics program that would advance computer development. This started a succession of investments over the years into ever more powerful computing capabilities at the National Laboratories, beginning with what are now two National Nuclear Security Administration labs, Lawrence Livermore and Los Alamos National Laboratories.

With the eventual dissolution of the AEC and the establishment of the Department of Energy in 1977, DOE became the heir and steward of these capabilities. DOE and its predecessor agencies support for applied mathematics and computer science, along with major investments in computer hardware and computational science, have driven progress in high-performance computing (HPC) in the United States, a major force spurring U.S. computing industry.

The scope of DOE computing applications has expanded from their original national defense focus to encompass an increasingly broad portfolio of scientific research and significant use by industry, especially with the establishment of the Leadership Computing Facilities at Argonne and Oak Ridge National Laboratories in 2004.

Today, DOE computing is a partnership between the National Nuclear Security Administration and the Office of Advanced Scientific Computing Research within the Department's Office of Science. Our two organizations work hand-in-hand to advance high-performance computing, including our partnership in the Exascale Computing Project (ECP).

Strategic Importance

Over the decades, the strategic importance of high-performance computing has grown enormously, and it is fair to say that a nation's capabilities in high-performance computing are one of the most important measures of its overall competitive standing in the global economy.

Thus, high-performance computing has become an essential pillar not just of America's national security but also of our leadership in science and—increasingly—of our national economic competitiveness.

A key measure of this competitiveness is the power of hardware, the raw capabilities of our most capable systems, and this informs the deployment of exascale machines. Additionally, an entire research and development ecosystem both contributes to, and benefits from, these capabilities. The defining strength of DOE computing is the vitality of the total effort involving a whole of government interagency approach and the multiple partnerships among the computing and microelectronics industries, the national laboratories, universities, and industry at large.

A key term is "co-design," meaning a highly integrated and cooperative effort where end users and application requirements contribute to the design of systems, and the systems in turn enable scientific discovery through advances in hardware, software, algorithms, and applications. A recent report by the Advanced Scientific Computing Advisory Committee speaks of "the enviable culture of co-design teams consisting of scientific users, instrument providers, theoretical scientists, mathematicians and computer scientists," which is an apt characterization of the approach.

Leading with the Science

When DOE's Office of Science established the Office of Advanced Scientific Computing, or ASCR, in 1999, their first major initiative was to establish Scientific Discovery through Advanced Computing, or SciDAC, a program that thrives to this day. The aim was to make scientific research and discovery the driver of the computational effort, to lead with the science and thereby let the quest for scientific discovery drive progress in computing.

SciDAC supports multi-disciplinary partnerships of computer scientists and applied mathematicians with domain scientists supported by the Office of Science, which has brought major computational-driven advances in a range of fields, from climate science, fusion energy, and high energy and nuclear physics, to materials science, chemistry, particle accelerator design, and biology, to name just some areas. SciDAC has also spurred major progress in software development and algorithm design.

The program encompasses all major Office of Science program offices along with DOE's Office of Nuclear Energy.

To mention just a few notable achievements:

- Computational scientists at Lawrence Berkeley National Laboratory and climate scientists at the Los Alamos National Laboratory and the University of Bristol have teamed to create the most accurate model for understanding the retreating Greenland and Antarctic ice sheets. By providing sub-kilometer resolution of key areas and less computationally demanding resolution of less important regions, it provides a much more accurate picture of the dissipating mass, which threatens to contribute to significant sea-level rise over the coming century.
- Fusion holds the promise of abundant, carbon dioxide-free energy. Progress toward fusion will depend on experiments run on ITER, currently under construction in France. ITER success requires understanding the interactions between the burning plasma contained in the reactor and the reactor walls. These involve extraordinarily high temperatures—millions of degrees in the case of the plasma and up to a thousand degrees for reactor walls—that are impossible to reproduce in a laboratory and difficult to simulate accurately at less than exascale capabilities. Scientists as Oak Ridge National Laboratory and University of Tennessee have developed a new approach using machine learning that accurately simulates the interaction between the plasma and the tungsten coating of the reactor walls—using a fraction of computing power normally required.
- The nature of supernova explosions is one of the enduring mysteries of astrophysics. With support by SciDAC, scientists led by Princeton University used Argonne National Laboratory's Cray XC40 system, called Theta, created some of the first three-dimension simulations of supernova explosions. The development of fully three-dimensional simulations is opening the way to more detailed comparison between simulation and observed events, facilitating a new understanding of the mechanism by which supernovae explode and collapse into neutron stars.

SciDAC continues to be an important force for progress in both domain sciences and computational science and engineering.

<u>The Advent of Exascale</u>

Exascale computing will provide the capability to tackle even more complex problems. Exascale has the capacity to deepen our understanding of climate change and hasten the development of clean energy. It will aid in the development of advanced manufacturing and holds major promise in the area of cancer research. The potential applications are too numerous to name.

The first exascale system, a Cray system called Frontier, is scheduled for delivery at the Oak Ridge Leadership Computing Facility (OLCF) beginning in July, with deployment completion expected in October. The Argonne Leadership Computing Facility's (ALCF) Aurora system is expected to be deployed in 2022. Lawrence Livermore National Laboratory is expecting delivery of the Department's third exascale machine, El Capitan, in 2023.

The drive toward exascale has been a major co-design effort, on two levels.

First, there has been a co-design partnership of Office of Science and NNSA with major computing and microelectronic vendors, beginning in Fiscal Year 2012. Through a series of five "forward" programs—FastForward, DesignForward 1 and 2, Fastforward 2, and the current ECP PathForward program—DOE-supported researchers have worked with vendors including AMD, Cray, IBM, Intel, NVDIA, and HPE to overcome a series of key technical hurdles, including surmounting power and speed limitations through the deployment of Graphic Processing Units (GPUs), improving interconnect, and developing new approaches to memory. In total DOE, through ASCR in the Office of Science and NNSA combined, has invested $460 million in this effort, while industry contributed at least an additional $307 million toward the research. DOE invited multiple agencies to participate in the project reviews as observers along the way. This has significantly disseminated the knowledge gained through the ECP project development process to other agencies.

Many of the challenges to be overcome stem from the impending end of Moore's Law. Moore's Law is shorthand for the progressive process of microchip miniaturization over the past several decades that has led processor speeds to double approximately every two years. As this process of miniaturization reaches its limits, and feature sizes on chips narrow toward the width of atoms, significant creativity is required to address these limitations. Two successful recent examples from the ECP include the partnership with AMD to develop architectures based on AMD's chiplets, which are separate pieces of silicon within a single processor package; and a partnership with Intel to develop interconnect based on Silicon Photonics, which communicate via photons rather than electrons—with potential increases in speed and major power savings.

These research partnerships with DOE have been critical in enabling the U.S. computer industry to enter the exascale era.

A second, parallel and overlapping, co-design effort encompasses scientific end-users. A key goal of this effort is to ensure that the community is ready with scientific applications to run as soon as exascale systems are deployed and available. These partnerships echo those of SciDAC, where applied mathematicians, computer scientists, and software developers—and often hardware designers—team with domain scientists to harness computation for research.

Challenges and Opportunities

At the threshold of exascale deployments, the scientific environment continues to evolve and to pose new challenges and opportunities. Current and planned upgrades to Office of Science user facilities—including light sources, neutron scattering sources, nanoscale and genomic facilities, and the supercomputers themselves—will bring more sophisticated and precise observations but also vastly larger data outputs. Means must be found to tame these mountains of data, process them, and extract what is meaningful. Artificial intelligence (AI) and machine learning (ML) are likely to play a key role here.

Data do not just pose a challenge; they also represent an enormous opportunity. In the coming years, we expect a great deal of scientific discovery to be heavily data-driven, meaning discovery will come from new methods of extracting insights from massive data sets, many of which exist today as still largely untapped resources. That is one reason why the Office of Science recently designated several important existing data sets as Publicly Reusable Research (PuRE) Data Resources, which are to be carefully curated and made available to the public at large on a more systematic basis. Included are data sets from the Materials Project, the Atmospheric Radiation Measurement program, the DOE Joint Genome Institute, the DOE Systems Biology Knowledgebase, the Particle Data Group, and the National Nuclear Data Center, with more resources likely to be added. The Office of Science is committed to careful management of these resources, stewarding them for the benefit of the community. The expectation is that the community will take the opportunity to mine these resources for new insights. AI and ML are likely to be at the center of these efforts.

AI also holds the promise of more sophisticated and autonomous facility operations. Currently operations at facilities such as x-ray light sources are heavily dependent on human monitoring and control, which involves a measure of trial and error. AI has the potential to monitor observations and adjust instrument operations in real time to ensure that observations capture what is needed and beam time use is optimized.

Already important progress is being made in this direction. The Office of Science's Energy Sciences Network, or ESnet, now connects the Linac Coherent Light Source at SLAC National Accelerator Laboratory with the Cori supercomputer at Lawrence Berkeley National Laboratory's National Energy Research Scientific Computing Center, providing real-time analysis ported to users' laptops, enabling on-the-fly adjustments to experiments based on computational results.

Increasing integration of AI into this process will require not only increased computational power but also enhanced networking. The Office of Science ESnet user facility provides ultra-high broadband connectivity to the DOE laboratories and research institutions across the U.S. and internationally. This connectivity will be increasingly vital as facility operations are controlled computationally, potentially by physically distant computational resources.

Beyond Exascale

Even as we stand on the threshold of the deployment of exascale systems, we are looking beyond exascale to new frontiers. Quantum Information Science (QIS) is one of the most important such frontiers. We have seen early progress in quantum networking. Last year, Brookhaven National Laboratory and Stony Brook University demonstrated a three-node network prototype over approximately 87 miles. Scientists from Argonne National Laboratory and the University of Chicago achieved photon entanglement over a network stretching 52 miles through the Chicago suburbs. These early results bode well.

Through the National Quantum Initiative Act, we established five National QIS Research Centers, led respectively by Argonne National Laboratory, Brookhaven National Laboratory, Fermi National Accelerator Laboratory, Lawrence Berkeley National Laboratory, and Oak Ridge National Laboratory. They are pursuing research in quantum networking, quantum sensing, and quantum computing and will position us well as quantum applications, and quantum computing in particular, become more capable.

In addition, we are supporting early investigations into the potential of neuromorphic computing, or computing architectures modeled on the neural structure of the human brain.

However, quantum computing at any scale is still well over the horizon, and neuromorphic computing even further out.

<u>The Imperative of Co-Design</u>

That is why it is essential that we continue our co-design efforts with vendors and with the microelectronics industry in the context of current technologies. We must continue to promote and support the development of new approaches, new materials, and new architectures to ensure that our nation is invested in the technologies that our researchers will need to remain on the cutting edge. Leadership in science remains indispensable to our security and prosperity, and today there is no leadership in science without leadership in high-performance computing. International competition is fierce. We need to be investing today in the immediate next-generation systems to stay in the game and to lead.

Federal investment is essential, since our activities in high-performance computing serve the national interest but do not in any sense drive the market. Absent Federal investment, vendors have little incentive to pursue these systems.

A skilled HPC workforce is also imperative. The Computational Science Graduate Fellowship, which is available, with varying eligibility requirements, to undergraduate seniors through the first year of Ph.D. study. Since its establishment in 1991, the program has sponsored over 450 fellows from more than 60 universities. Currently there are 89 fellows using high-performance computing in their research over a wide range of fields. The Computational Science Graduate Fellowship plays a major role in building and sustaining the nation's HPC workforce and is a Federal investment that thereby pays dividends.

DOE Scientific Computing in the Pandemic

As a final point, crises have a way of revealing the true value of things. The COVID-19 pandemic has made abundantly clear the enormous value of DOE's high-performance computational resources—for molecular modeling, drug screening, epidemiology, and a host of other applications in the war against the disease. Early in the pandemic, OLCF's Summit supercomputer screened thousands of compounds for potential effectiveness against the SARS-CoV-2 virus. As requests for computer time multiplied and other public and private resources were made available, DOE joined with 15 partner institutions in March 2020 to establish the COVID-19 High Performance Computing Consortium. IBM established the consortium website linking to a central gateway, based on the NSF-funded XSEDE portal, for submission of proposals, which were reviewed by a panel of experts and assigned to appropriate resources first in the United States, then across the globe. Eventually, the consortium counted 43 members with over 600 petaflops of computing capacity. At last count, as many as 175 proposals were received from researchers in 15 countries.

An indication of the power of HPC to accelerate discovery in the fight against the pandemic is provided by a study titled, "AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics," which was awarded a 2020 Special Gordon Bell Prize for COVID-related applications. As part of this research, using AI, and building on multiple experiments supported by NIH, NSF, as well as DOE, a team involving Argonne scientists found a way to significantly speed the screening of molecules for effectiveness against the virus—increasing the speed of screening by a factor of 50,000. Extending the approach, the team managed to create an automated pipeline capable of screening literally a billion drug candidates per day. The work was supported through the National Virtual Biotechnology Laboratory, a consortium of DOE national laboratories focused on response to COVID-19, with funding provided by the Coronavirus CARES Act. (Like the HPC Consortium, NVBL employs a single gateway for efficient expert review of proposals and assignment of facilities.)

This last example of accelerated discovery hints at the enormous promise of the current era of high-performance computing, an era marked by the advent of exascale systems, the rapid development of AI and ML, and the establishment of computation as a powerful tool of discovery on equal footing with theory and experiment. The United States stands poised to lead in this new era of computing and discovery, and our future security and prosperity will depend on it.