Written Testimony
Professor Daniel A. Reed
Vice President for Research and Economic Development, The University of Iowa

Before the
Subcommittee on Energy
Committee on Science, Space and Technology
U.S. House of Representatives

Hearing on "America's Next Generation Supercomputer: The Exascale Challenge"

May 22, 2013

Chair Lummis, Ranking Member Swalwell, and Members of the Subcommittee, my name is Dan Reed, and I am the Vice President for Research and Economic Development at the University of Iowa. Thank you for the opportunity to share perspectives on the opportunities and challenges surrounding exascale computing and to respond to your questions regarding the American High-end Computing Leadership Act. I appreciate the time and attention that the Committee is spending on this topic, and I commend you for advancing the dialogue on computational science and high-performance computing.

My testimony begins by emphasizing the importance of high-performance computing as an enabler of scientific discovery and innovation across all disciplines, which distinguishes it from other scientific instruments. It summarizes key points in the shifting technology base of high-performance computing and the critical dependence of that base on continued investments in basic research. It then outlines some of the key recommendations from past reviews of the U.S. advanced computing investment strategy. It also emphasizes the interdependence of high-performance computing and the broader computing ecosystem, with implications for the future of U.S. competitiveness. Finally, it concludes by providing a set of recommendations and next steps for the Federal government and others to allow the U.S. to develop next-generation high-performance computing systems and to maintain its global leadership.

High-Performance Computing: A Universal Amplifier

The English scientist Sir Humphrey Davy could well have been speaking about high-performance computing when he said, two centuries ago:

> *Nothing tends so much to the advancement of knowledge as the application of a new instrument. The native intellectual powers of men in different times are not so much the causes of the different success of their labors, as the peculiar nature of the means and artificial resources in their possession.*

In a phrase – success accrues to the talented and trained who have access to the most effective and powerful tools, whether computers, telescopes, particle accelerators, or genetic sequencers. Computing, and particularly high-performance computing, is unique among these and other scientific instruments, distinguished by its universality as an intellectual amplifier.

New telescopes advance astronomy and deepen our understanding of the universe's origins and cosmological future, but do not illuminate biological processes and the origins of life. New particle accelerators test the limits of the Standard Model and our understanding of fundamental physics, but do not yield new insights into the Earth's geological processes or and the exogeology of other planets in our solar system.

In contrast, **new, more powerful supercomputers and improved computational models yield new insights into all scientific disciplines, for they breathe life into the underlying mathematics of scientific models, allowing us to understand nuanced predictions and to shape experiments more efficiently. They also help capture and analyze the torrent of experimental data being produced by a new generation of scientific instruments and sensors, themselves made possible by advances in computing and microelectronics. Consequently, high-performance computing has emerged as the third pillar of the research portfolio, complementing theory and experiment across all disciplines.**

High-Performance Computing: Past and Present

At any moment, high-performance computing (HPC) is most accurately defined by its impact – those computing systems with transformative power to enable breakthrough scientific discoveries, ensure defense preeminence and maintain international competitiveness. Thus, these HPC systems integrate the most advanced microprocessors and computational accelerators, the highest speed, lowest latency networks and the highest capacity storage systems. Their system software also embodies advanced techniques for resource management and systemic resilience, and the applications integrate complex numerical techniques that span a wide range of temporal and spatial scales. In short, they embody the most advanced computing technology currently available.

In the past thirty years, we have seen repeated shifts in HPC hardware and software technologies, themselves consequences of long-term, U.S. government-funded basic research, with concomitant changes in computing systems deployments across industry, academia and our national laboratories. In the 1980s, vector supercomputing dominated, as embodied in the eponymously named systems designed by the late Seymour Cray. The 1990s saw the rise of massively parallel processing (MPPs) and shared memory multiprocessors (SMPs), built by Thinking Machines, Silicon Graphics and others. In turn, clusters of commodity (Intel/AMD x86) and purpose-built processors (e.g., IBM's BlueGene), dominated the previous decade. Today, those clusters have been augmented with accelerators and GPUs. Each of these technology transitions brought dramatically higher performance – from gigaflops (10^9 arithmetic operations (flops) per second) through teraflops (10^{12} flops/second) to petaflops (10^{15} flops/second) – and new scientific and technical insights via higher fidelity computational models.

Today, leading edge HPC systems at the Department of Energy and the National Science Foundation allow researchers to explore the frontiers of phenomena in scientific and engineering domains as diverse as high-energy physics, materials science, combustion dynamics, biophysics and computational chemistry, structural mechanics, and molecular biology. From understanding the subtleties of airflow in turbomachinery and underhood cooling through chemical molecular dynamics for consumer products to biomass feedstock for fuel cells, these and other systems also support advanced design and manufacturing, in partnership with U.S. industry.

Across government agencies, these systems have also played an essential role in ensuring the safety and reliability of our nuclear stockpile and in protecting our national security in an uncertain and dangerous world. Large-scale data analytics also now enable extraction of insights from the unprecedented volumes of scientific and biomedical data being created by scientific instruments, as well as helping ensure information superiority for national security. High-speed networking and the global Internet also facilitate research collaboration and information sharing.

High-Performance Computing: Looking to the Future

With every new generation of high-performance computing technology, the Department of Energy and its national laboratories have been at the forefront, collaborating with universities, other agencies and industry in the design, deployment and operation of the world's most powerful supercomputers. DOE's Advanced Scientific Computing Research Program (ASCR) has been a crucial element of this activity, funding basic and applied computing research and system development, and developing new computational science applications. ASCR has also worked closely with its DOE partner, the National Nuclear Security Agency (NNSA), on advanced technologies and system deployments.

Today, HPC systems from DOE's Oak Ridge National Laboratory, Lawrence Livermore National Laboratory and Argonne National Laboratory occupy the first, second and fourth places on the list of the world's fastest computers, based on the Top500 list. From this, one might surmise that all is well. After all, in today's 21st century knowledge economy, the importance of U.S. leadership in high-performance computing and computational science would seem self-evident.

Yet today's U.S. leadership in both deployed HPC capability and in the underlying technologies that are needed to create the future generations of HPC systems is now under unprecedented challenge. Other nations are investing strategically in HPC and computational science to advance their national and regional priorities. The U.S. research community has repeatedly issued warnings and alarms about this erosion of U.S. leadership in information technology and high-performance computing.

In 2004, I testified to this committee on this same topic while serving as the Director of the NSF-funded National Center for Supercomputing Applications (NCSA), which for twenty-five years has provided HPC services to the national science and engineering community, most recently via the NSF Blue Waters petascale HPC system. At the time of my 2004 testimony, I had recently chaired the 2003 community workshop on the *Roadmap for the Revitalization of High-end Computing*,[1] which had been convened in response to a request from the interagency High-end Computing Revitalization Taskforce (HECRTF). The workshop report's executive summary noted,

> *The common theme throughout these recommendations is the need for sustained investment in research, development, and system acquisition. This sustained approach also requires deep collaboration among academic researchers, government laboratories, industrial laboratories, and computer vendors. … Rather,* ***multiple cycles of advanced research and development, followed by large-scale prototyping and product development, will be required to develop systems that can consistently***

[1] Community workshop on the *Roadmap for the Revitalization of High-end Computing*, 2003, organized by the Computing Research Association (CRA), available at http://archive.cra.org/Activities/workshops/nitrd/

achieve a high fraction of their peak performance on critical applications, while also being easier to program and operate reliably.

In 2005, as a member of the President's Information Technology Advisory Committee (PITAC), I chaired a review of U.S. computational science capabilities, which produced a report to the President entitled *Computational Science: Ensuring America's Competitiveness*.[2] The report's principal finding was

> *Computational science is now indispensable to the solution of complex problems in every sector, from traditional science and engineering domains to such key areas as national security, public health, and economic innovation. Advances in computing and connectivity make it possible to develop computational models and capture and analyze unprecedented amounts of experimental and observational data to address problems previously deemed intractable or beyond imagination. Yet, despite the great opportunities and needs, universities and the Federal government have not effectively recognized the strategic significance of computational science in either their organizational structures or their research and educational planning. These inadequacies compromise U.S. scientific leadership, economic competitiveness, and national security.*

Based on this finding, the PITAC's principal recommendation was the following:

> *To initiate the required transformation,* ***the Federal government, in partnership with academia and industry, must also create and execute a multi-decade roadmap directing coordinated advances in computational science and its applications in science and engineering disciplines.***

Today, we are poised on the threshold of a new era, one defined by exascale computing (10^{18} flops/second) and trans-petascale data analysis. It brings the promise of new scientific discoveries and insights, but also difficult technical and engineering challenges. Exascale system design and construction will require solutions to some deep and fundamental problems in semiconductor processes, energy-efficient computing and data movement, primary and secondary memory design, packaging and cooling, resilience and reliability, resource management and programming. It will also require development of new numerical algorithms, data analysis techniques and scientific and engineering applications.

These solutions will not be simply incremental extensions of current technologies, nor will those solutions be derived from current industry research and development paths alone. Equally importantly, the fruits of such collaboration can have far deeper benefits than simply the construction of an exascale computing platform. They will be the innovative disruptions that will help position the U.S. information technology industry for the future. Our global competitors are well aware of this disruption opportunity -- there are now active and well-funded initiatives for hardware, software and applications in the European Union, Japan, China and India.

[2] *Computational Science: Ensuring America's Competitiveness*, President's Information Technology Advisory Committee (PITAC), June 2005

Basic Computing Research: Partnerships and Innovative Disruption

In the United States, it has long been axiomatic that we are the world's leader in information technology and the application of that technology to business, science, engineering and government. In the 1960s, the birth of System/360 mainframe computing and its support for business processes made IBM a global leader in computing. In the 1970s and 80s, minicomputers such as the PDP-11 and VAX brought computing to research laboratories and smaller businesses, making Digital Equipment (DEC) a global brand. In the 1980s and 90s, personal computing made Intel and Microsoft large and successful companies. Today, Apple, Google and Amazon are icons of the smartphone and Internet age.

Each of these companies has been the beneficiary of Federal investments in long-term basic research, including investments in high-performance computing. The microprocessors and software in our PCs and smartphones embody architectural research, resource management and programming abstractions developed over four decades of research. Indeed, many of these ideas were first tested and validated in systems designed for high-performance computing.

Today's Internet originated from DARPA network research investments in the 1970s and 80s, and from NSFnet, which the National Science Foundation (NSF) created to connect NSF's supercomputing centers and provide open access to high-end computing facilities. This environment spawned the Mosaic graphical web browser at the University of Illinois's NSF-funded National Center for Supercomputing Applications (NCSA) sparking the 1990s dot.com boom and the explosive growth of the Internet. That environment and further investment in search and indexing research led to the search engines, social networks and cloud services that define our daily interactions.

Make no mistake; global computing leadership is not a U.S. birthright; it has been repeatedly earned and hard fought, based on a continued commitment to basic research investments by the Federal government, translation of those basic research insights into technological innovations, and strategic investment and business acumen to create and deliver new computing systems and products.

Andrew Grove, the former CEO of Intel, highlighted the importance of continual innovation in his famous computing aphorism, "only the paranoid survive." What he really said is far more subtle and important, "Success breeds complacency. Complacency breeds failure. Only the paranoid survive." Simply put, past success can lull one into complacency at precisely the time that changes and strategic investment are needed to ensure future success.

The computing industry is replete with telling examples of Grove's maxim, when technology breakthroughs spawned disruptive innovations. The rise of the personal computer made Microsoft and Intel large and successful, but it also required IBM to reinvent itself to continue to prosper. In that same period, DEC failed to make that transition successfully, despite its talented people and technology base. More recently, the birth of the Internet and the rise of smartphones and tablets have had similar disruptive effects on the computing ecosystem, with important consequences for our future.

The Internet and web services revolution is now global and U.S. influence, though still substantial, is being diluted. Notwithstanding Apple's phenomenal success, the majority of smartphones and tablets are now designed, built and purchased outside the U.S., and the annual sales volume of smartphones and tablets already exceeds that of PCs and servers. In short, this exploding "post-PC" market is international in scope, with U.S. consumers an increasingly small minority of users.

This ongoing shift in consumer preferences and markets is accompanied by another seismic technology shift. Smartphones and tablets are based on low power, energy-efficient microprocessors (a key component of proposed exascale computing designs) and systems-on-a-chip (SoCs) using the U.K.-created ARM architecture. Unlike Intel and AMD, which design and manufacture the x86 chips found in today's PCs and most leading edge servers and HPC systems, ARM does not manufacture its own chips. Instead, it licenses the design to others, who incorporate the architecture into custom SoCs that are manufactured by global semiconductor foundries such as TSMC.[3] Thus, the ARM hardware ecosystem is global in scope, and U.S. vendors, led by NVIDIA, Qualcomm and Texas Instruments, are but three of the international competitors in the ARM SoC market.

As a member of the National Academies Board on Global Science and Technology (BGST), in 2012, I chaired a study on this and other shifts and their implications for the United States and its future defense capabilities. The resulting report, entitled *The New Global Ecosystem in Advanced Computing: Implications for U.S. Competitiveness and National Security*,[4] made several salient points relevant to this discussion, of which the following is notable:

> *Over time, the increasing presence and establishment of foreign markets that are larger, are potentially more lucrative, and have better long-term growth potential than in the United States and other developed countries could also have significant implications.* ***Any shift in the global commercial center of gravity could lead to a shift in the global R&D center of gravity as international firms are required to locate in these markets if they are to remain competitive and to meet the requirements of government regulations in the target markets.***

These observations are equally apt for the future of HPC and exascale programs. U.S. competitiveness and continued HPC leadership are predicated on a vibrant U.S. computing industry that can continue to invest in the development of new technologies – advanced chips and architectures, novel networks and hardware systems, and new system software and applications – while leveraging continued investment in basic computing research by Federal research agencies, universities and national laboratories.

Actionable Recommendations

The global computing ecosystem is in flux, and other nations are investing strategically in high-end computing. In the U.S., we also face difficult decisions about Federal investment priorities, given current economic realities. Thus, it has never been more important that we act strategically and thoughtfully as we consider the future of funding for basic computing research and for high-end computing. I believe the following are essential elements of a successful U.S. strategy.

1. **Advanced HPC system deployments are crucial, but the computing research and development journey is more important than any single system deployment by a pre-determined date.** The basic and applied research in algorithms, software, applications, semiconductor technologies, storage systems, energy management, integration and packing, resilience and scaling, among others, will produce unexpected discoveries and technology breakthroughs, as well as enable design

[3] Taiwan Semiconductor Manufacturing Company (TSMC), http://www.tsmc.com/english/default.htm

[4] *The New Global Ecosystem in Advanced Computing: Implications for U.S. Competitiveness and National Security*, National Academies Board on Global Science and Technology, 2012, available at http://www.nap.edu/catalog.php?record_id=13472

and deployment of effective exascale systems. Those discoveries and the people who made them are the lifeblood of computing innovation and future U.S. competitiveness. They are the true enablers of sustainable exascale computing.

2. **High-end data analytics (big data) and high-end computing (exascale) are both essential elements of an integrated computing research and development agenda; neither can be sacrificed or minimized to advance the other.** From web search, social networks and business processes, through government efficiency and service optimization to large-scale scientific instrumentation and sensors, big data has been and will be transformational. Cloud computing infrastructure and services and high-performance computing systems and services have much in common, and insights from each can benefit the other. Global leadership in both is essential to the our future.

3. **Research and development of next-generation algorithms, software and applications is as crucial as investments in semiconductor devices and hardware, and we have historically underinvested in these areas relative to hardware.** Despite this underinvestment, experience has shown that over the past thirty years performance increases in high-performance computing systems has been due in equal parts to hardware and software advances. The massive and unprecedented scale of current and future high-performance computing systems is bringing new challenges in programmability and systemic resilience, resource scheduling and numerical stability. We must invest in a balanced way.

4. **Much deeper and sustained interagency collaboration is needed. The Department of Energy, particularly the Advanced Scientific Computing Research Program (ASCR), has led the development of an exascale computing research and development agenda, but it cannot succeed alone.** In the past, the National Science Foundation, the Department of Defense, the National Institute of Standards and Technology, and the National Institutes of Health have been active and engaged partners in the high-performance computing research and development agenda. Today, that is much less true.

   The historical strength of the U.S. research strategy in high-performance computing has been the complementarity and diversity of its participating research agencies. We must renew and reenergize that partnership, given the unique role that each agency plays:
   - NSF – basic computer science research in the enabling technologies; data management and sustainable cyberinfrastructure for national science and engineering academic community
   - DoD – advanced technology research and prototyping; mission-oriented deployments
   - NIST – standards and cybersecurity
   - NIH – computational modeling, big data analytics and biomedical applications for higher quality, lower cost health care
   - DOE – computational science, systems research and prototyping; large-scale system deployments, building on the research and operations staff of the Office of Science and NNSA laboratories

5. **We must change the model for research, development, acquisition and deployment of high-end computing systems if the U.S. is to sustain the leadership needed for future scientific discovery and national security.** As the HECRTF report noted, we must support and sustain multiple cycles of advanced research and development, followed by large-scale prototyping and product development. In a budget-constrained world, we must work more efficiently and collaboratively, which will require new and deeper interagency prioritization and budget allocations, along with

long-term industry partnerships. To ensure that, **there should be verifiable metrics of interagency collaboration, community engagement and technical progress that are tied to agency funding**.

6. Finally, the global information technology ecosystem is in flux, with the transition to a new generation of low power, mobile devices and cloud services. **We must recognize and embrace the need for "dual use" technology research and development that enables high-performance computing systems and scientific discovery while also ensuring the competitiveness of U.S. industry, both in information technology and in the use of computing to advance U.S. businesses. Our long-term national security depends on this.**

I believe we face both great opportunities and great challenges in high-performance computing. Scientific discovery via computational science truly is the "endless frontier" of which Vannevar Bush spoke so eloquently in 1945. The challenges are for us to sustain the research, development and deployment of the high-performance computing infrastructure needed to enable those discoveries. To do so, we must adapt our model of collaboration, retaining the strength of its diversity while focusing our resources efficiently.

Finally, let me again commend this committee and its continued leadership and commitment to high-performance computing, including the American High-end Computing Revitalization Act.

Daniel A. Reed
Vice President for Research and Economic Development
The University of Iowa

Daniel A. Reed is Vice President for Research and Economic Development, as well as University Chair in Computational Science and Bioinformatics and Professor of Computer Science, Electrical and Computer Engineering and Medicine, at the University of Iowa. In this role, he is responsible for the university's research processes and compliance, campus-wide research and policy centers, technology evaluation and licensing, and economic development.

Prior to joining the University of Iowa, from 2007 to 2012, he was a senior leader at Microsoft, first serving as Microsoft's scalable and multicore computing strategist and later as Corporate Vice President for Extreme Computing and Technology Policy. In these roles, he helped shape Microsoft's long-term vision for technology innovations in parallel and cloud computing and the company's associated policy engagement with governments and institutions around the world.

Previously, he was the Chancellor's Eminent Professor at UNC Chapel Hill, as well as the Director of the Renaissance Computing Institute (RENCI) and the Chancellor's Senior Advisor for Strategy and Innovation for UNC Chapel Hill. Prior to that, he was Gutgsell Professor and Head of the Department of Computer Science at the University of Illinois at Urbana-Champaign (UIUC) and Director of the NSF National Center for Supercomputing Applications (NCSA). He was also one of the principal investigators and chief architect for the NSF TeraGrid. He received his PhD in computer science in 1983 from Purdue University.

In addition to his technical activities, Reed has been deeply involved in policy initiatives related to science, technology and innovation, both in the United States and internationally. He has served as a member of the U.S. Federal Communications Commission's Technical Advisory Committee, as a member of the U.S. President's Council of Advisors on Science and Technology (PCAST) and chair of the computational science subcommittee of the President's Information Technology Advisory Committee (PITAC). He currently serves as chair of the capability computing review committee for Los Alamos National Laboratory, as a consultant to Department of Energy, and as a member of the National Academies Board on Global Science and Technology.