The National Strategic Computing Initiative

Office of Science and Technology Policy

June, 2016
Overview

• The “What” and “Why” of NSCI
• Retrospective on Year 1
• Feedback from Community at Anniversary event
• Getting to “What’s Next?”
• Dwarfs and Giants
EXECUTIVE ORDER

CREATING A NATIONAL STRATEGIC COMPUTING INITIATIVE

By the authority vested in me as President by the Constitution and the laws of the United States of America, and to maximize benefits of high-performance computing (HPC) research, development, and deployment, it is hereby ordered as follows:

The NSCI is a whole-of-government effort designed to create a cohesive, multi-agency strategic vision and Federal investment strategy, executed in collaboration with industry and academia, to maximize the benefits of HPC for the United States.
NSCI Policy Principles

It is the policy of the United States to sustain and enhance its scientific, technological, and economic leadership position in HPC research, development, and deployment through a coordinated Federal strategy guided by four principles:

1. The United States must deploy and apply new HPC technologies broadly for economic competitiveness and scientific discovery.

2. The United States must foster public-private collaboration, relying on the respective strengths of government, industry, and academia to maximize the benefits of HPC.

3. The United States must adopt a "whole-of-government" approach that draws upon the strengths of and seek cooperation among all Federal departments and agencies with significant expertise or equities in HPC in concert with industry.

4. The United States must develop a comprehensive technical and scientific approach to efficiently transition HPC research on hardware, system software, development tools, and applications into development and, ultimately, operations.

This order establishes the NSCI to implement this whole-of-government strategy, in collaboration with industry and academia, for HPC research, development, and deployment.
Strategic Objectives

(1) Accelerating delivery of a capable exascale computing system that integrates hardware and software capability to deliver approximately 100 times the performance of current 10 petaflop systems across a range of applications representing government needs.

(2) Increasing coherence between the technology base used for modeling and simulation and that used for data analytic computing.

(3) Establishing, over the next 15 years, a viable path forward for future HPC systems even after the limits of current semiconductor technology are reached (the "post-Moore's Law era").

(4) Increasing the capacity and capability of an enduring national HPC ecosystem by employing a holistic approach that addresses relevant factors such as networking technology, workflow, downward scaling, foundational algorithms and software, accessibility, and workforce development.

(5) Developing an enduring public-private collaboration to ensure that the benefits of the research and development advances are, to the greatest extent, shared between the United States Government and industrial and academic sectors.
Federal Agency Roles

The order identifies lead agencies, foundational R&D agencies, and deployment agencies:

• **Lead** agencies (DOE, DOD, NSF) are charged with developing and delivering the next generation of integrated HPC capability and will engage in mutually supportive research and development in hardware, software, and workforce to support the objectives of the NSCI.

• **Foundational** R&D agencies (IARPA, NIST) are charged with fundamental scientific discovery work and associated advances in engineering necessary to support the NSCI objectives.

• **Deployment** agencies (NASA, FBI, NIH, DHS, NOAA) will develop mission-based HPC requirements to influence the early stages of design of new HPC systems and will seek viewpoints from the private sector and academia on target HPC requirements.
Retrospective on Year 1

- **July 29, 2015** – Executive 13702 Order Issued
- **July 30, 2015** – NSCI Private Roundtable (Academia, Private Sector, Government) at the White House
- **August 26, 2015** – Inaugural meeting of the NSCI Executive Council
- **September 15, 2015** – RFI on Science Drivers for Capable Exascale issued
- **October 20-21, 2015** – White House NSCI Workshop (Academia, Private Sector, Government)
- **February, 2016** – President’s FY 2017 budget request submitted
- **June 20, 2016** – Sunway TaihuLight leads new Top500 ranking
- **July 26, 2016** – Public Implementation Plan issued
- **July 30, 2016** – NSCI Anniversary Workshop
NSCI Exascale Request for Information
(RFI NOT-GM-15-122)

• Specific scientific and research challenges that need 100-fold increase in performance
• Potential impact of the research to the scientific community, national economy, and society.
• Specific limitations/barriers HPC systems must overcome
• Related research areas that would benefit from this level of augmented computational capability.
• Important computational and technical parameters of computational problems in 10 years
• Alternative models of deployment and resource accessibility for exascale computing
• Capabilities needed by the end-to-end system, including data analytics and visualization tools, shared data capabilities, and data services
• Other issues, e.g., training, workforce development and collaboration environments.
Highlighted Applications:
- Assuring National CyberSecurity
- Self Assembly based Nano-manufacturing
- Rapid Design of organic-polymer based Thermoelectrics
- Optimizing the Power Grid
- Realistic Hypersonic flow for Flight Vehicles
- Regional Scale Seismic predictions
- High Resolution Atmospheric & Climate Models
- N-by-N Comparison of All Patients in US

DOE National Laboratories Responses: 135
Academic Responses: 94
Industry Responses: 8
Foreign Responses: 2
Others: 5
NSCI Workshop

A White House Workshop on the National Strategic Computing Initiative was held October 20-21, 2015, with around 250 participants representing industry, government, academia, and other organizations. During the workshop, many individual opinions were expressed. Several ideas emerged that could potentially inform the NSCI implementation:

• The evolutionary path for HPC is more uncertain than during the previous decades. There will be different ways of coupling simulation with data analytics.

• There is reason to be optimistic for convergence of analytics and HPC in the long term – but diversity of approach is key in the short term.

• A number of hardware technology and architectural innovations will be attempted to overcome physical limitations for charge-based CMOS. NSCI must accommodate this breadth of choice and avoid any premature down select of technology.

• Although existing clouds lack the performance required to satisfy the most demanding HPC applications, they have already proved suitable for a number of scientific workloads. Clouds could be a viable model for NSCI broad deployment.

• Deeper engagement with the industrial (non-computing) sector will be key to achieve broad deployment and advance the Nation’s economic competitiveness.
NSCI in the FY 2017 President’s Budget

The 2017 President’s Budget Request was sent to Congress on February 9, 2016. The total funding explicitly allocated to the NSCI was $318 million.

• DOE National Nuclear Security Administration, ($95 million)
  – Advanced Simulation and Computing: Activities and research leading to deployment of exascale capability for national security applications in the early 2020’s

• DOE Office of Science ($190 million)
  – Advanced Scientific Computing Research ($154 million) R&D and design to ultimately achieve capable exascale systems with 1000x performance of current HPC
  – Basic Energy Sciences ($26 million): basic research resulting in codes to predictively design functional materials and chemical processes
  – Biological and Environmental Research ($10 million): develop science base for increasingly complex climate modeling and data analytic applications

• National Science Foundation ($33 million):
  – Multi-directorate: Scientific discovery for HPC, HPC for scientific discovery, the broader HPC ecosystem, and workforce development

Above in addition to existing core programs aligned with NSCI goals
Themes from NSCI Anniversary (1/3)

• **Perspectives on the Top500**
  – Don’t respond as if TaihuLight is a stunt. Have a plan, play the long game, and innovate.
  – It’s not about FLOPS, it’s about performant software.
  – Leadership. It’s not about the Top500, it’s about “materials to application” leadership.

• **Funding and funding stability**
  – If we believe in the NSCI, we need real investments.
  – By providing longer-term grants, we can better tap into the strengths of academia.
  – To succeed, the NSCI needs long-term investments and partnerships, not just one year USG money.

• **Moonshots**
  – Major programs can drive capability development and accessibility.
  – A small number of outcome based “moonshots” should be developed to drive the efforts forward, which would also provide contextual basis for programs.
Themes from NSCI Anniversary (2/3)

• **Software**
  – Standard interfaces that enable contributions from the broader community are urgently needed to ensure impacts of new HPC systems have maximum value.
  – More investment in software is needed, with attention to pruning what we don’t need in the future. Often we are not good at down-selects but this may be a time to do so with more intent.
  – Need to fund end-to-end community codes, creating a pipeline from programming languages to problems.
  – Apply the workflow concept to HPC software – how can we get more efficiency from coders AND enable more scientists to leverage HPC?

• **Partnerships**
  – Enduring public-private collaboration, emphasizing effective sharing of information between USG, industry, and academia is necessary but not sufficient. Strong commitment in the Executive Branch will also be necessary for the NSCI to succeed.
  – Opportunities to think about hubs/institutes across national labs, universities that are open to private sector to create “one stop shopping” ecosystem that lowers barriers to entry by providing the “right” level of abstraction
  – Differentiated partnerships are needed – “not peanut butter across all performers”
Themes from NSCI Anniversary (3/3)

• **Access/broad deployment**
  – For economic competitiveness, the NSCI must drive HPC into small and medium enterprises.
  – Government can do more to increase access to HPC systems.
  – Need to leverage different sectors to accelerate pace of development and broaden deployment.
  – Make it accessible!

• **Workforce**
  – We must generate more people that can exploit HPC. Perhaps 100 times as many as we do today.
  – We need people teaching others to teach.
  – There are training and education opportunities by partnerships between universities and private sector, but we can do better.

• **HPC development and deployment is becoming more expensive for everyone. Private sector is looking for more unity from the Federal Government in terms of requirements and procurement strategies.**
Getting to “What’s Next?”

- Working on a new governance and collaboration model, where lead agencies would contribute some modest staffing for joint NSCI management activities
- Agencies are actively reviewing and revising their FY17 plans
- OSTP and agencies are continuing outreach
- OSTP and NITRD are working on a software development concept to increase productivity and portability of HPC applications
Dwarfs and Giants

• **Seven Dwarfs**
  – In 2004, Phil Colella identified seven numerical patterns for high-end simulation in the physical sciences

• **Seven Giants**
  – In 2013, the National Research Council identified seven computational tasks for massive data analysis

• **Recall that**
  – NSCI Anniversary Workshop participants repeatedly highlighted the importance of reducing the barrier to entry for HPC programming
  – The NSCI seeks convergence of data analytic systems with simulation and modeling and broad deployment for leadership in scientific computing and economic competitiveness
Opportunity: Petascale in a Rack

- **Low power revolution in hardware (cellphones) now reaching HPC.**
  - By 2024: Petascale for $200K, 20KW (1% of what it is now).
  - Cost of 2-3 years of a postdoctoral researcher, or one year of a mid-level engineer.
  - Power / cooling needs are modest.
  - This price / performance point makes HPC potentially accessible to a much broader range of users.

- **Opportunity: HPC for S&T as a ubiquitous technology (analogous to cellphone apps)**

- **Challenge: HPC for S&T is presently a niche activity. Teams of highly-trained specialists develop one-off applications codes. This doesn’t scale to ubiquitous technology!**
Reducing the Barrier to Entry

• **Low-power technologies lead to more complex architectures on a single chip or board.**
  – High degree of parallelism (1000x on a chip), hierarchically organized.
  – Complex memory hierarchies at the board and chip level.
  – Programming at high performance is a challenge.

• **Complexity of architectures leads to ever-widening semantic gap.**
  – 1996: two lines of mathematics required 20 lines of code to obtain high performance.
  – 2016: two lines of mathematics requires 100’s (up to 1000) lines of code to obtain high performance. The details depend on the algorithm and on the architecture.
  – Such a semantic gap is a barrier to a productive and agile software environment.

• **Is a productive, agile, performant software ecosystem technically feasible?**

• **Closing the semantic gap is a necessary (but not sufficient) condition for such an ecosystem.**
Increasing the Level of Abstraction

A higher level of abstraction in the programming system is essential to close the semantic gap.

• **Algorithmic patterns:** combinations of computation and data access that are derived from the underlying mathematical structure of particular ways of representing S&T problems on a computer.
  – Big data patterns: basic statistics, “n-body”, graphs, linear algebra, optimization, integration, alignment (e.g. genomics).

• **Approach:** programming system based on a high-level representation of these patterns, combined with tools to automate the generation of optimized code from HL pattern code.
  – Economies of scale: patterns are common across multiple S&T applications.
  – The mathematical structure of the patterns are available to the rest of the programming system as a guide to the generation of optimized code.
  – The required software technologies are available.
**Layered Software Stack**

<table>
<thead>
<tr>
<th>Applications: use libraries and pattern-based toolset, general-purpose software stack.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Libraries: use pattern-based toolset, general-purpose software stack.</td>
</tr>
<tr>
<td>Pattern-based toolset: uses general-purpose software stack</td>
</tr>
<tr>
<td>Shared tools: translation tools, autotuners, performance tools</td>
</tr>
<tr>
<td>Hardware and associated general-purpose software stack</td>
</tr>
</tbody>
</table>
Metrics for Success

• Expressiveness: no more than 10x expansion going from a mathematical description of an algorithm to HL pattern code.

• Performance: measured performance of at least 50% of theoretical peak performance *for the algorithm*.

• Portability: no more than 10% of the HL pattern code must change in moving to a new platform.

• Supported software: design, user documentation; regression testing, bug tracking; worked examples.

• S&T adoption.
  – Short term: scientific successes by early adopters.
  – Long-term: richly-populated and widely-used software ecosystem.
Building a Whole-of-Nation Effort

• We are proud of the level of coordination and collaboration we have achieved across Federal agencies, but...
  – We know we don’t have a monopoly on good ideas
  – We need all of them to succeed

• Must establish similar level of coordination and collaboration with academia, private industry, and international partners
  – Have held one industry roundtable, and two NSCI workshops
  – Joint Request for Information in the Fall
  – Continue outreach to industry and academia to help us shape and promote follow-on activities
Outlook: A Lasting Initiative

• We have achieved a high level of coordination and collaboration across Federal agencies

• We are still in the process of establishing similar level of coordination and collaboration with academia and private sector, but are encouraged by the level of engagement.

• Our long-term strategy for High Performance Computing
  – Was initially established in a government-only NSCI Implementation Plan at the end of November in 2015;
  – Was shared in part in the President’s 2017 budget request;
  – Was expanded upon with the release of the Strategic Plan; and
  – Continues to evolve to reflect external events and inputs from industry and academia