Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020

Jon Eisenberg, CSTB director

Committee on Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science in 2017-2020
Computer Science and Telecommunications Board
Division on Engineering and Physical Sciences

The National Academies of
SCIENCES • ENGINEERING • MEDICINE
Committee

WILLIAM D. GROPP, University of Illinois, Urbana-Champaign, Co-Chair
ROBERT J. HARRISON, Stony Brook University, Co-Chair
MARK R. ABBOTT, Woods Hole Oceanographic Institution
ROBERT L. GROSSMAN, University of Chicago
PETER M. KOGGE, University of Notre Dame
PADMA RAGHAVAN, Pennsylvania State University
DANIEL A. REED, University of Iowa
VALERIE TAYLOR, Texas A&M University
KATHERINE A. YELICK, University of California, Berkeley

JON EISENBERG, study director
Charge

Develop a framework for future decision making about NSF’s advanced computing strategy and programs.

(No recommendations concerning the level of federal funding for computing infrastructure.)
Some Background

• NSF success supporting science and engineering through advanced computing has increased demand exponentially

• At the same time, the nature of computing is changing in multiple ways
  – Increasing use of data from many sources
  – End of Dennard (frequency) scaling forcing major changes in architecture, in turn requiring investment in new algorithms and software

• Internationally, advanced computing recognized as vital to scientific progress
Recommended Goals For Advanced Computing

1. Position the United States for continued leadership in science and engineering
2. Ensure that resources meet community needs
3. Aid the scientific community in keeping up with the revolution in computing, and
4. Sustain the infrastructure for advanced computing
1. POSITION THE UNITED STATES FOR CONTINUED LEADERSHIP IN SCIENCE AND ENGINEERING
Recommendations

Recommendation 1 NSF should sustain and seek to grow its investments in advanced computing—to include hardware and services, software and algorithms, and expertise—to ensure that the nation’s researchers can continue to work at frontiers of science and engineering.

Recommendation 2 As it supports the full range of science requirements for advanced computing in the 2017-2020 timeframe, NSF should pay particular attention to providing support for the revolution in data-driven science along with simulation. It should ensure that it can provide unique capabilities to support large-scale simulations and/or data analytics that would otherwise be unavailable to researchers and continue to monitor the cost-effectiveness of commercial cloud services.
Observations

• Increased capability has historically enabled new science. Without continued growth, some research will have difficulty making advances
• Many fields now rely on a greater aggregate amount of computing than a typical university can provide
• “Converged” systems that support both compute-and data-intensive applications can play a role; more specialized systems may also be needed to meet some requirements
Cloud computing

- Demonstrates value of service model that “democratizes” access
- Leverage and ease access to rich software stacks and large data sets

- Does not currently support very large, tightly coupled parallel applications
- Positioned today to play a growing role for data-centric applications
  - Largest systems at scale larger than any computational science platform

- Cost savings? NSF computing centers already exploit economies of scale and load sharing
- Rapidly changing; price (cost) and types of services likely to change
- Cost of commercial cloud could be greatly reduced by reducing or eliminating overhead, bulk purchase, or partnerships with commercial cloud providers.
  - Generally considered a service on which overhead is charged, which makes purchase more attractive to each research group but more expensive in the aggregate.
Resources

- Maintaining science leadership will be challenging
- Resources for advanced computing inherently limited even as demand is growing
- If NSF cannot increase or better leverage advanced computing resources:
  - Unable to meet future aggregate demand
  - Will have to reduce the size of the largest research projects supported by advanced computing
2. ENSURE THAT RESOURCES MEET COMMUNITY NEEDS
Recommendation 3. To inform decisions about capabilities planned for 2020 and beyond, NSF should collect community requirements and construct and publish roadmaps to allow the Foundation to set priorities better and make more strategic decisions about advanced computing.

Recommendation 4.2 NSF should inform users and program managers of the cost of advanced computing allocation requests in dollars to illuminate the total cost and value of proposed research activities.
Roadmaps

• Reflect visions of communities supported by NSF
  – Both largest users and long tail
• Brief documents with overall strategy/approach, not details
• Look ~5 years ahead with vision ~10 years out

• Purposes:
  – Guide investment
  – Inform users about future facilities
  – Enable more effective partnerships within NSF and with other federal agencies

• Models:
  – Academies astronomy and astrophysics decadal surveys
  – DOE’s Particle Physics Project Prioritization Panel
  – But...
    • Input must be collected from a much wider set of users
    • Requirements must be aggregated at a much higher level
# Japan’s Second Tier Plans

Japanese “Leading Machine” Candidates

Roadmap of the 9 HPCI University Centers

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hokkaido</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tohoku</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEC SX-9 + Exp6800 (31TF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEC SX-ACE 800TFlop’s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tsukuba</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HA-PACS (800 TF, GPU)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tokyo</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2K Todai (140 TF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fujitsu FX10 (1PFlops, 150TB, 408 TB/s), Hitachi SR16000M1 (514 TF, 10.9 TB, 5.376 TB/s)</td>
<td>100+ PF</td>
<td>10+ PF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PostT2K JHPCA (~30 PF, 100+ TF, 600TB, 4.0+ PB/s, 0.68+ PB/s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tokyo Tech.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsubame 2.5 (50+ TF, 110+ PF, 5-6PB/s, 1.84MW)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsubame 3.0 (20-25 PF, 5-6PB/s) 1.84MW (Max 2.3MW)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsubame 4.0 (100-200 PF, 20-40PB/s, 2.3-1.68MW)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nagoya</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fujitsu M16000(3.6TF, 17TBs), H86000(26.6TF, 6.6TB/s), FX(10.7TF, 30 TB/s)</td>
<td>50+ PFlops</td>
<td>10+ PF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fujitsu FX10 (30.9TF, 31.8 TB/s), CX400(470.6TF, 55 Post FX10 Upgrade (3PF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Kyoto</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cray XE6 (300TF, 92.7TB/s), GreenBlade 8000</td>
<td>10+ PF</td>
<td>10+ PF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cray XC30 (420TF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cray XC40 Phi(384TF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Osaka</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(243TF, 61.5 TBs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SX-4 + SX-9 (21.7 TF, 3.3 TB, 50.4 TB/s)</td>
<td>10+ PF</td>
<td>50+ PF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Kyushu</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hitachi SR16000M1 (172 TF, 22TB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hitachi HA6000chi(240)(500TF, 215 TB, 90.8TB/s), Xenon Phi(672 TF, 6.6TB, 4.4 TB/s)</td>
<td>(5+ PB/s)</td>
<td>(5+ PB/s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fujitsu FX10(270TF, 65.28 TB/s), CX400(510TF, 152.5 TB, 151.14 TB/s), GPGPU(256TF, 30 TB, 53.52 TB/s)</td>
<td>10+ PF</td>
<td>50+ PF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

~17PF April 2015, Japan-wide ~40PF (incl. K), ”SINET5” (100GBps nationwide)
## DOE’s Roadmap

<table>
<thead>
<tr>
<th>System attributes</th>
<th>NERSC Now</th>
<th>OLCF Now</th>
<th>ALCF Now</th>
<th>NERSC Upgrade</th>
<th>OLCF Upgrade</th>
<th>ALCF Upgrades</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name Planned Installation</td>
<td>Edison</td>
<td>TITAN</td>
<td>MIRA</td>
<td>Cori 2016</td>
<td>Summit 2017-2018</td>
<td>Theta 2016</td>
</tr>
<tr>
<td>System peak (PF)</td>
<td>2.6</td>
<td>27</td>
<td>10</td>
<td>&gt; 30</td>
<td>150</td>
<td>&gt; 8.5</td>
</tr>
<tr>
<td>Peak Power (MW)</td>
<td>2</td>
<td>9</td>
<td>4.8</td>
<td>&lt; 3.7</td>
<td>10</td>
<td>1.7</td>
</tr>
<tr>
<td>Total system memory</td>
<td>357 TB</td>
<td>710TB</td>
<td>768TB</td>
<td>~1 PB DDR4 + High Bandwidth Memory (HBM)+1.5PB persistent memory</td>
<td>&gt; 1.7 PB DDR4 + HBM + 2.8 PB persistent memory</td>
<td>&gt; 480 TB DDR4 + High Bandwidth Memory Local Memory and Persistent Memory</td>
</tr>
<tr>
<td>Node performance (TF)</td>
<td>0.460</td>
<td>1.452</td>
<td>0.204</td>
<td>&gt; 3</td>
<td>&gt; 40</td>
<td>&gt; 3</td>
</tr>
<tr>
<td>Node processors</td>
<td>Intel Ivy Bridge</td>
<td>AMD Opteron</td>
<td>Nvidia Kepler</td>
<td>64-bit PowerPC A2</td>
<td>Intel Knights Landing many core CPUs Intel Haswell CPU in data partition</td>
<td>Multiple IBM Power9 CPUs &amp; multiple Nvidia Voltas GPUs</td>
</tr>
<tr>
<td>System size (nodes)</td>
<td>5,600 nodes</td>
<td>18,688 nodes</td>
<td>49,152</td>
<td>9,300 nodes 1,900 nodes in data partition</td>
<td>&gt;3,500 nodes</td>
<td>&gt;2,500 nodes</td>
</tr>
<tr>
<td>System Interconnect</td>
<td>Aries</td>
<td>Gemini</td>
<td>5D Torus</td>
<td>Aries</td>
<td>Dual Rail EDR-IB</td>
<td>Aries</td>
</tr>
<tr>
<td>File System</td>
<td>7.6 PB 188 GB/s, Lustre®</td>
<td>32 PB 1 TB/s, Lustre®</td>
<td>26 PB 300 GB/s GPFS™</td>
<td>28 PB 744 GB/s Lustre®</td>
<td>120 PB 1 TB/s GPFS™</td>
<td>10PB, 210 GB/s Lustre Initial</td>
</tr>
</tbody>
</table>
NSF roadmap
Why Make Value of Allocations Known?

• Goal is to inform decisions
  – A resource that is free is often wasted
  – Can inform choices for code tuning, algorithm development
  – Could pilot project to allow groups to trade resources between software help (tuning, algorithm implementation) and compute cycles
    • Pilot needed to identify unintended consequences

• Not intended for chargeback
3. AID THE COMMUNITY IN KEEPING UP WITH THE REVOLUTION IN COMPUTING
Observations

• Computer architectures, hardware, program models, are changing rapidly
• Better software tools, technical expertise, and more flexible service models can boost productivity

• Leadership role for NSF in defining future advanced capabilities and helping researchers use them effectively will help ensure that:
  – Software and systems remain relevant to science portfolio
  – Researchers are prepared to use current and future capabilities
  – Investments are aligned with future directions
Recommendations

Recommendation 5. NSF should support the development and maintenance of expertise, scientific software, and software tools that are needed to make efficient use of its advanced computing resources.

Recommendation 6. NSF should invest modestly to explore next-generation hardware and software technologies to explore new ideas for delivering capabilities that can be used effectively for scientific research, tested, and transitioned into production where successful. Not all communities will be ready to adopt radically new technologies quickly, and NSF should provision advanced computing resources accordingly.
Software is a Large Part of Advanced Computing Infrastructure

• There is a very large investment in software for computational science
• Much of this will need to be rewritten for the new architectures
• Few groups are ready for this
• Required:
  – New ideas to automate as much as possible
  – Investment in engineering software
4. SUSTAIN THE INFRASTRUCTURE FOR ADVANCED COMPUTING
Observations

• Expertise and physical infrastructure are essential, long-lived assets

• Recent NSF strategy of acquiring facilities and creating centers relies on:
  – Irregularly scheduled competition among institutions roughly every 2 to 5 years
  – Equipment, facility, and operating cost sharing by states, institutions, and vendors

• Challenges with this approach:
  – Relies on cost sharing that may no longer be viable due to mounting costs and budget pressures
  – Repeated competitions can lead to proposals designed to win a competition rather than maximize scientific returns
  – Most importantly, doesn’t provide long-term support needed to develop and retain talent needed to manage systems, support users, and evolve software
Recommendations

• **Recommendation 7** NSF should manage advanced computing investments in a more predictable and sustainable way.
  – Funding models that emphasize continuity of support
  – Longer-term commitments to center-like entities that can support both resources and expertise
  – Regular, rigorous review of centers
Managing advanced computing investments in a more predictable and sustainable way

• Creates opportunities to apply expertise more broadly within NSF, such as for large scale science projects with large, long-term needs
• Creates new opportunities to address long-term data storage, preservation, and curation challenges
Questions?

cstb.org
jeisenbe@nas.edu
202.664.1235