Community Input on the Future of High Performance Computing

Report of a Workshop held on July 29, 2010

High-Performance Computing Task Force

Co-Chairs:
Thomas Zacharia
University of Tennessee and Oak Ridge National Laboratory,
Oak Ridge, Tennessee

James L. Kinter III
Center for Ocean-Land-Atmosphere Studies
Calverton, Maryland

Prepared for the
National Science Foundation
Advisory Committee for Cyberinfrastructure
Table of Contents

Executive Summary ....................................................................................... v
Introduction .....................................................................................................1
Findings..............................................................................................................3
Recommendations .......................................................................................... 10
Acknowledgements..................................................................................... 11
Appendix A: AC CI Task Forces ................................................................. 13
Appendix B: HPC Workshop Series ............................................................. 15
Appendix C: Workshop Attendees................................................................. 17
Appendix D: Charge to the Workshop ......................................................... 19
Appendix E: Participant White Papers......................................................... 21
Executive Summary

On July 29, 2010, the High Performance Computing (HPC) Task Force established by the Advisory Committee for Cyberinfrastructure (ACCI) of the National Science Foundation (NSF) held the second of three planned workshops to gather input and advice from the U.S. computing community on how NSF should proceed in building and sustaining a broad portfolio of HPC investments that will stimulate transformative scientific discovery over the next 5–10 years.

This workshop focused on application drivers for exascale computing and data infrastructure. The 42 scientists, vendors, and resource providers invited to the workshop were asked to consider the following questions:

1. What is exascale computing and data cyberinfrastructure (CI) essential for?
2. What new science does exascale computing and data CI enable?
3. What are the challenges to reach exascale?
4. What scientific areas have not yet considered exascale, and how do we help them get there?

The workshop produced four major recommendations, summarized below. Findings and additional recommendations are discussed in the body of the report.

**Recommendation 1:** Given the major challenges involved in the transition to HPC at the exascale, NSF should consider new models for partnerships, such as expanded collaborations with industry, academia, and other agencies, to accelerate the development of exascale systems and applications.

**Recommendation 2:** Given the promise of exascale computing and data CI to accelerate innovation and enhance competitiveness, NSF should expand its efforts to engage new user communities in HPC, with particular emphasis on broadening industry use of HPC and on applying new and expanded capabilities to data-intensive fields of research.

**Recommendation 3:** Given the opportunities and challenges presented by the generation of exabytes of digital data, NSF should provide funding for a digital data framework designed to address the issues of knowledge discovery in the exascale ecosystem, including the co-location of archives and community data resources with compute and visualization resources as appropriate.
Recommendation 4: Given the rapidly changing environment for HPC, NSF should establish a continuing process for soliciting community input on its plans for HPC investments to facilitate the growth of an ecosystem for exascale science.

The HPC Task Force plans to hold a third workshop on December 3, 2010, with a focus on broader engagement and workforce development for HPC. Findings and recommendations from all three workshops will be integrated into a final report.
Introduction
The National Science Foundation (NSF) has commissioned an Advisory Committee for Cyberinfrastructure (ACCI), consisting of distinguished scientists and engineers from academia, government and industry with expertise in different disciplines and a computational focus, to advise it on all cyberinfrastructure (CI) activities. Additional information on the ACCI is presented in Appendix A.

In April 2009, the NSF commissioned six task forces within the ACCI to assist with strategic planning in the following areas:

- Campus bridging
- Data and visualization
- Grand challenge communities and virtual organizations
- High performance computing (HPC)
- Learning and workforce development
- Software

These task forces have 12–18 months to conduct studies, hold workshops, and make recommendations. Then NSF will develop programs based on their findings.

The HPC Task Force has organized a workshop series, as described in Appendix B, to gather input and advice from the U.S. computing community on how NSF should proceed in building and sustaining a broad portfolio of HPC investments that will stimulate transformative scientific discovery over the next 5–10 years.

The first of three planned workshops was held on December 4, 2009. Position papers, presentation materials, and the final report on this workshop, which focused primarily on HPC requirements for the next generation of scientific discovery, are available at http://www.nics.tennessee.edu/dec09workshop.

The second workshop, held on July 29, 2010, addressed application drivers for exascale computing and data CI. Participants (listed in Appendix C) were asked to consider four questions:

1. What is exascale computing and data CI essential for?
2. What new science does exascale computing and data CI enable?
3. What are the challenges to reach exascale?
4. What scientific areas have not yet considered exascale, and how do we help them get there?
The full charge to the workshop is provided in Appendix D. Participants were encouraged to prepare brief position papers; those submitted are included in Appendix E. These position papers are also available online at http://www.nics.tennessee.edu/july10workshop, as are presentation materials from the workshop and other supporting information. The findings and recommendations of this workshop are described in this report.

A third workshop, scheduled for December 3, 2010, will focus on broader engagement and workforce development for HPC.

The draft reports on the findings and recommendations of the three workshops will be distributed to the participants before final submittal to NSF. Feedback from the attendees will be integrated into a final report. These reports will be delivered to NSF and will be made available to the entire community on the web at http://www.nics.tennessee.edu/workshop.
Findings

Increases in computational power over the past decade have enabled the development of simulations and models of unprecedented fidelity and speed, with profound implications for discovery and innovation. As a result, HPC is now widely viewed as essential to scientific research, and the U.S. scientific community has come to rely heavily on the CI provided by NSF.

The petascale systems that are now moving into use are accelerating progress in scientific understanding by enabling the exploration of complex systems, from astrophysical phenomena to global climate to biomolecules and nanoparticles. In addition, they are providing industry with opportunities to shorten design cycles, reduce development costs, and improve the performance of manufactured items.

These petascale systems also provide tantalizing glimpses of the transformational impact of the exascale systems that are expected to become available within the next decade. Realizing this impact will require a substantial investment (on the order of several billion dollars)—not only in the development of hardware and software for these systems, but also in the translation of today’s forefront scientific applications to the new architectures that will enable exascale computing and in the education and training of a workforce that can effectively use these tools and develop new ones that can deliver accurate solutions to complex problems, not only in basic science but also in downstream industrial applications.

Simultaneously, the “data deluge” associated with the generation and management of exabytes of data, not only from increasingly detailed simulations but also from experiments and observations, creates new opportunities and challenges for research and development (R&D). New developments in data analysis, visualization, and management are needed to enable the use of massive quantities of data from disparate sources, both to guide research and to directly advance science.

Workshop participants were asked to consider four questions focused on the evolution of computing and data CI from the petascale to the exascale and its consequences. Findings from the workshop are presented in the context of these questions.

1. What is exascale computing and data CI essential for?

Extensive reviews by the scientific community have identified a number of scientific “grand challenges” that can exploit computing at
extreme scales to bring about dramatic progress toward their resolution; for example, a 2007 report from the U.S. Department of Energy (DOE) on modeling and simulation at the exascale for energy and the environment\(^1\) highlights a number of significant opportunities for exploiting computing at the exascale to solve compelling problems for energy and the environment. Presentations at the workshop addressed the potential of exascale science to provide new insights into biomolecular structures; advance the development of high-resolution models for understanding the impacts of climate change; enable dramatic advances in radio astronomy; support knowledge-based design of new energy technologies; understand the interaction of the solar wind with the magnetosphere; explore functional interactions between cells, organs, and systems and quantify how these interactions change in disease states; improve long-term seismic hazard analysis; and perform core collapse supernovae simulations with the spatial resolution required to properly model critical aspects of explosion dynamics.

Past experience with the application of HPC to these and other fields provides confidence that vastly expanded compute and data power will enable continuing advances as simulation fidelity improves and/or more time steps can be computed. Progress in uncertainty quantification will enhance acceptance of the results of increasingly sophisticated models of complex physical systems.

As computational resources, sensor networks, and other large-scale instruments and experiments expand to the exascale, the amount of data generated by these sources is also growing commensurately. Opportunities exist for new discoveries in data-rich fields—not only in areas such as biology and medicine, chemistry, climate science, earth science, and astrophysics, but also in public health, economics, and social science. However, new exascale CI will be essential, not only to manage and extract useful information from these massive data sets, but also to ensure the preservation of persistent and definitive data repositories and to facilitate and encourage sharing of data from high-cost experiments and simulations across communities.

For example, raw data is produced in enormous quantities at facilities such as the radio telescopes of the National Radio Astronomy Observatory. The Large Synoptic Survey Telescope (LSST), an 8.4-m ground-based telescope, will require the transfer of \(>15\) terabytes (TB) of raw data nightly; processing of these data to provide near-

\(^1\) *Modeling and Simulation at the Exascale for Energy and Environment*,
http://www.sc.doe.gov/ascr/ProgramDocuments/ProgDocs.html
real-time feedback to the telescope to optimize imaging and to alert the astronomy community to interesting observations will produce >75 TB nightly.\(^2\) The next-generation Square Kilometer Array is projected to generate a data flow of ~100 TB/s with 100 TB/day archived, a data volume that mandates pipelining from observing to archive with no human intervention. These extreme data requirements call for data management systems that can reliably process unprecedented data volumes, ensure consistent data quality without manual intervention, accommodate both scientific and computing technology evolution over at least a decade, and serve data products to a diverse community of users located across multiple continents.

Similarly, the Center for Analysis and Prediction of Storms (CAPS) at Oklahoma University is developing and demonstrating techniques for the numerical analysis and prediction of high-impact local weather and environmental conditions, with emphasis on the assimilation of observations from Doppler radars and other advanced in-situ and remote sensing systems. CAPS produces real-time, high-resolution weather analyses and forecasts throughout the year, with enhanced forecasts (real-time 400-m-resolution low-level wind analyses, updated every 5 minutes) during the spring storm season. Data from the Center for Collaborative Adaptive Sensing of the Atmosphere (CASA) network and other sources are incorporated into these analyses. Very short range numerical weather prediction (NWP) forecasts for a 2-hour window at 1-km resolution are produced every 10 minutes when active weather exists within the CASA Oklahoma network. This environment is producing large amounts of data that need to be shared across the country.\(^3\)

A third example can be drawn from experience at the Southern California Earthquake Center (SCEC),\(^4\) which is building earthquake modeling capabilities with the goal of transforming seismology into a predictive science. Although the location, time, and magnitude of large earthquakes cannot be reliably and skillfully predicted, numerical simulations of seismic radiation from complex fault ruptures and wave propagation through three-dimensional crustal structures have now advanced to the point where they can usefully predict strong ground motions from anticipated earthquake sources. These simulations require robust, on-demand supercomputing and rapid access to very large data sets.

\(^2\)http://www.lsst.org/lsst
\(^3\)http://forecast.caps.ou.edu/
\(^4\)http://www.scec.org/
Other areas in which exascale computing and data CI can be expected to have a transformational impact include social networks, which can be used in predicting and modeling the spread of infectious diseases; social and economic systems; and linguistics and text analysis.

2. What new science does exascale computing and data CI enable?
Exascale computing and data CI will make it possible to deliver more accurate solutions to a variety of complex problems. DOE’s Extreme Computing Workshop Series assessed the impact of computational modeling and simulation on several areas of “domain science,” including climate science, high-energy and nuclear physics, fission and fusion energy, basic energy sciences, and biology. In all of these areas, exascale computing was viewed as an essential tool for advancing science and accelerating innovation.

In addition, high-end simulation is the most credible vehicle for accelerating the application of knowledge from basic science to the design of new energy technologies, including advanced photovoltaic materials and energy storage systems. Experience at DOE’s Leadership Computing Facilities demonstrates that engineering applications (e.g., modeling the complex geometry of engineered systems) provide both drivers and challenges for HPC; for example, Boeing has used these capabilities for the development of large-scale computational tools that have accelerated the design of transport airplanes, while fully incorporating the complex chemistry of biofuels into combustion models for advanced transportation and power generation systems will require exascale resources.

As computational resources, sensor networks and other large-scale instruments and experiments grow, the datasets generated by these sources are also growing. Other communities—such as libraries, industry, and the social sciences—suddenly find themselves in possession of vast quantities of digital data as well. These data are being captured by emerging ecosystems and cyberenvironments that cross a number of technological and social boundaries in both scientific and nonscientific fields. This new resource, if effectively managed, presents a remarkable opportunity for data-driven and data-enabled science.

3. What are the challenges to reach exascale?
The research community has extensively studied the challenges associated with reaching the exascale. The Defense Advanced

---

Research Projects Agency (DARPA) Exascale Computing Study\textsuperscript{6} provides an excellent overview of the technology challenges for exascale systems. The International Exascale Software Project, cosponsored by NSF, DOE, and other organizations, is addressing the technical challenges for software infrastructure that will accompany the novel architecture and extreme scale of emerging systems.\textsuperscript{7} DOE’s scientific grand challenge workshop series\textsuperscript{5} included consideration of the R&D required in mathematical models, applied mathematics, and computer science to meet the needs of a spectrum of applications at the exascale. These efforts make it clear that a broad effort will be needed to deliver exascale science.

In particular, application development is critical both in broadening the use of petascale computing and in advancing to the exascale. NSF’s Accelerating Discovery in Science and Engineering through Petascale Simulations and Analysis (PetaApps) program provides some support, but more will be needed to develop applications that will enable wider use of HPC resources as computing power increases and to facilitate the porting of established codes to new systems.

Approaches discussed at the workshop included the development of application teams charged with matching hardware to software, working with different platforms to find the most effective match. DOE’s Scientific Discovery through Advanced Computing (SciDAC) program provides a useful model in its development of a common set of base libraries that are ported to new systems by a core group. The development of modular codes was also discussed, with consideration given to the need for advances in design and for extensive data transfer. It was noted that code development often benefits from competition; independent sites can both compete and collaborate. Closer collaboration with the international computing community (especially Europe and Japan) would be beneficial.

Of particular importance, “co-design” of architectures and algorithms, in which scientific problem requirements influence architecture design and technology constraints inform the formulation and design of algorithms and software, offers the opportunity to improve the effectiveness of both petascale and exascale systems. This approach is being aggressively pursued by DOE and DARPA.


Workshop participants also discussed the need for training programs, both to encourage use of resources at the lower levels of the “Branscomb pyramid” and to address the new operating models and different memory hierarchies expected for exascale systems.

The cost of reaching the exascale presents an additional challenge. A single exascale system could cost between $300M and $400M—an amount that is likely to limit the number of these systems that can reasonably be constructed. Thus, in advocating for these systems the scientific community must be able to state clearly what problems they will solve. The ability to attack previously insoluble problems (as noted in the findings for question 2, above) will be a key “selling point.”

The utility of exascale computers for more applied work will also be important in justifying investment in these resources. Engineering applications (e.g., modeling the complex geometry of engineered systems) provide both drivers and challenges for exascale computing, but private-sector support for their development is limited at best. A number of applications developed by the research community have been adopted by industry; for example, pharmaceutical companies are using molecular dynamics (MD) codes developed by the academic community. The recent establishment of DOE’s energy innovation hub for nuclear energy modeling and simulation, the Consortium for Advanced Simulation of Light Water Reactors (CASL), was cited as a welcome step in the right direction for the nuclear industry and one that will help to demonstrate the power of HPC in accelerating innovation. Further efforts to strengthen the connection between HPC and industry is imperative.

Digital data has already reached the exascale. Efficient and effective reduction and analysis, management, storage, service, and sharing of the exabytes of data produced by simulations, experiments, and observations present significant technical challenges for data CI. In particular, researchers will need high-performance visualization and data analysis tools and techniques to extract insights from large amounts of data. However, policy issues associated with organization, curation, and lifecycle of the data may present even greater challenges.

In terms of hardware, the simple storage disk drives with tape access of 20 years ago have been replaced by multiple types of disk drives, hierarchical storage systems with complicated software, and robotic tape subsystems. In the past, some applications regenerated data files instead of retrieving them because it was faster; this is no longer an effective way to manage compute resources. Today, a typical large
compute machine procurement dedicates 10–15% of the overall budget to extreme storage environments specifically for that compute environment. Significant investments in additional center-wide high-speed data management storage environments have also become essential. Large, dedicated data-intensive machines with large, extreme storage systems for all types of data are also beginning to populate the compute machine landscape.

In parallel with these changes, data management has expanded to include information lifecycle management and collection management, and data curation has become an important factor in the design of the underlying CI. The need to selectively retain data for future research presents challenges to funding agencies, data owners, data users, and service organizations. In general, funding for computational research has focused on near-term problems and their solution; no provision has been made for preservation of and continuing access to the data generated by the research. As a result, the intellectual capital being produced is not always managed with the same diligence as traditional scientific output.

The management of data has also become a forefront issue in scientific publishing: as the number of journal articles accompanied by datasets increases, publishers are faced with decisions about their role in managing and preserving this data.

Action is needed to make the immense quantities of digital data now being acquired and stored, both within and beyond the scientific community, available for advancing science and engineering. Action is also needed to address issues such as data provenance, validation, access authorization, expiration, and cost to regenerate.

4. What scientific areas have not yet considered exascale, and how do we help them get there?

In most scientific areas, a few high-level users are taking advantage of petascale CI and considering the exascale. However, the difficulty of using even today's HPC resources deters many users, and the challenges of exascale systems will exacerbate this problem. Programs are needed:

- to assist users at all levels of the “Branscomb pyramid”;
- to educate students, faculty, and practicing scientists in the use of today’s computational resources and techniques; and
- to attract the next generation of computational and domain scientists who will use exascale computing and data CI to make discoveries.
**Recommendations**

The presentations, breakout sessions, and discussion led to three major recommendations aligned with the workshop theme of application drivers for exascale computing and data CI, and to one general recommendation.

**Recommendation 1:** Given the major challenges involved in the transition to HPC at the exascale, NSF should consider new models for partnerships, such as expanded collaborations with industry, academia, and other agencies, to accelerate the development of exascale systems and applications.

In the area of applications development, NSF should consider providing additional support for the work of porting established codes to new systems. As an example, DOE’s Office of Science has commissioned an effort aimed at having several key science applications ready for production runs when its next-generation system is delivered. DOE’s SciDAC program provides another useful model in its development of a common set of base libraries that are ported to new systems by a core group. The development of modular codes should be encouraged, taking into account the need for advances in design and for extensive data transfer.

**Recommendation 2:** Given the promise of exascale computing and data CI to accelerate innovation and enhance competitiveness, NSF should expand its efforts to engage new user communities in HPC, with particular emphasis on broadening industry use of HPC and on applying new and expanded capabilities to data‐intensive fields of research.

NSF should consider establishing an outreach program to industry to communicate the value of HPC as a means of increasing innovation and productivity. NSF should also explore opportunities to assist industry in accessing modeling and simulation technology, perhaps through the development of one or more technology centers. Such centers could provide a “workshop” environment for development of exascale applications, graduate student education, and showcasing of successes, fostering interactions across the community. The development of the National Center for Atmospheric Research provides a starting point, as do the “renaissance teams” formed at the National Center for Supercomputing Applications in the 1980s.

Such a center could also provide a focal point for building and expanding connections to data-intensive fields that can benefit from broader use of HPC, especially given the challenges of effectively using
extreme-scale systems. Virtual organizations are also expected to be useful in expanding access to HPC resources and tools.

**Recommendation 3:** Given the opportunities and challenges presented by the generation of exabytes of digital data, NSF should provide funding for a digital data framework designed to address the issues of knowledge discovery in the exascale ecosystem, including the co-location of archives and community data resources with compute and visualization resources as appropriate.

Specifically, new methods, management structures, and technologies are needed to manage the size, diversity, and complexity of current and future datasets and data streams and to facilitate use of digital data in deriving new scientific insights and driving innovation.

**Recommendation 4:** Given the rapidly changing environment for HPC, NSF should establish a continuing process for soliciting community input on its plans for HPC investments to facilitate the growth of an ecosystem for exascale science.

This recommendation, although not directly tied to the questions posed to workshop participants, is based on discussions of the growing need for HPC resources at all levels of the “Branscomb pyramid” as critical tools for research and education. As NSF considers future acquisition, development, and provision of these resources, community input can provide valuable insights into the opportunities and challenges of applying HPC to the advancement of science and technology.

**Acknowledgements**

This work was supported in part by the National Science Foundation under OCI-0960905, Community Input on the Future of High-Performance Computing.
Appendix A: ACCI Task Forces

The National Science Foundation (NSF) vision for cyberinfrastructure includes a national-level, integrated system of hardware, software, data resources, and services to enable revolutionary advances in science and engineering. This vision, as described in a 2007 NSF report,\(^8\) includes virtual organizations for organizing the efforts of distributed communities; high-performance computing (HPC) hardware and software; data management, analysis and visualization hardware and software; and learning and work force development efforts.

To realize this vision, the NSF Office of Cyberinfrastructure (OCI) is funding the research, development and deployment of a comprehensive cyberinfrastructure to enable the solution of the most challenging problems in science and engineering. OCI, in strong collaboration with other offices, directorates, and agencies, provides stewardship for computational science and engineering at NSF. It also supports the preparation and training of current and future generations of researchers and educators to use cyberinfrastructure to further the nation’s research and education goals.

NSF has commissioned an Advisory Committee for Cyberinfrastructure (ACCI) that consists of distinguished scientists and engineers from academia, government and industry with expertise in different disciplines and a computational focus. This committee advises NSF on all cyberinfrastructure activities. NSF oversees the vision and strategic direction with operational oversight left to the committee. ACCI membership and other details can be found at: http://www.nsf.gov/od/oci/advisory.jsp.

In April 2009, NSF commissioned six task forces within the ACCI to assist with the strategic planning. The six task forces are the following:

- Campus bridging,
- Data and visualization,
- Software,
- High performance computing,
- Learning and workforce development, and
- Grand Challenges Communities and Virtual Organizations.

These task forces have 12–18 months to conduct studies, hold workshops and make recommendations. Then NSF will develop programs based on their findings.

The drivers that the task forces must consider include: requirements for accessing the advanced computing resources needed in the 2011–2015 time frame, the development and support of applications, the new opportunities being presented by innovations in computer science and engineering, integration of research and education to support the development of the next generation of scientists and engineers, and policies and programmatic activities needed to allow each of the drivers to be realized.
Appendix B: HPC Workshop Series

A core group of leading scientists, vendors and resource providers representing the high-performance computing (HPC) community has been selected in consultation with National Science Foundation (NSF) program managers to convene for a series of three one-day meetings. This group includes participants from NSF, the U.S. Department of Energy (DOE), the U.S. Department of Defense (DOD), and computing centers in other nations. The HPC Task Force is acting as an organizing committee to design the agenda and assignments for the workshops. Attendance at the HPC workshops is by invitation only; input from the larger research and education community has been solicited via a website.

At the first workshop, on December 4, 2009, participants discussed requirements for the next generation of scientific discovery. In addition, directors from NSF supercomputer centers described lessons learned from the inception of the program from 1985 through the current Track 1 and 2 programs, including the benefits and challenges of the current architectures and deployments. Participants were encouraged to prepare brief (1–2 page) position papers addressing these issues and their proposals for future programs. These position papers are available at http://www.nics.tennessee.edu/dec09workshop, along with presentation materials from the workshop.

At the second workshop, on July 29, 2010, 42 participants discussed application drivers for exascale computing and data cyberinfrastructure. Participants were again encouraged to prepare position papers. These position papers are available at http://www.nics.tennessee.edu/july10workshop, along with presentation materials from the workshop.

A third workshop, scheduled for December 3, 2010, will focus on broader engagement and workforce development for HPC.

The draft reports on the findings and recommendations of the three workshops will be distributed to the participants before final submittal to NSF. Feedback from the attendees will be integrated into a final report. This report will be delivered to NSF and will be made available to the entire community on the web.
Appendix C: Workshop Attendees
Amy Apon, University of Arkansas
Bill Barth, University of Texas
John Connelly, University of Kentucky
Rhonda Dias, Silicon Graphics, Incorporated
Mark Fahey, University of Tennessee
Rob Fowler, Renaissance Computing Center
Tom Furlani, State University of New York at Buffalo
Omar Ghattas, University of Texas
Galen Gisler, University of Oslo
Matthias Gobbert, University of Maryland
Dennis Goo, Intel
Teri Hagan, Oak Ridge National Laboratory
David Halstead, National Radio Astronomy Observatory
Robert Harrison, Oak Ridge National Laboratory
Thuc Hoang, National Nuclear Security Administration
Eric Jakobsson, University of Illinois
Brad Jones, Silicon Graphics, Incorporated
Homa Karimabadi, University of California – San Diego
George Karniadakis, Brown University
Dan Katz, University of Chicago
Jim Kinter, Center for Ocean-Land-Atmosphere Studies, HPC Task Force Co-chair
Patricia Kovatch, University of Tennessee
William Kramer, University of Illinois
John Levesque, Cray, Incorporated
Rick Linger, Carnegie-Mellon University
Bruce Loftis, University of Tennessee
Dick McCombie, Cold Spring Harbor Laboratory
Phil Maechling, University of Southern California
Bronson Messer, Oak Ridge National Laboratory
George Michaels, Intel
Jose Munoz, National Science Foundation
Bonnie Nestor, Oak Ridge National Laboratory
Esmond Ng, Lawrence Berkley National Laboratory
Rob Pennington, National Science Foundation
Irene Qualters, National Science Foundation
Dave Randall, Colorado State University
Klaus Schulten, University of Illinois at Urbana-Champaign
Mark Shephard, Rensselaer Polytechnic Institute
Jeffrey Vetter, Georgia Institute of Technology
Community Input on the Future of High Performance Computing

July 29, 2010

Nancy Wilkins-Diehr, San Diego Supercomputer Center
Thomas Zacharia, University of Tennessee
and Oak Ridge National Laboratory, HPC Task Force Co-chair
Appendix D: Charge to the Workshop
The goal of the workshop is to develop recommendations to assist NSF in developing future programs for advanced scientific computing. The charge is to provide specific advice to NSF on how to build and sustain a broad portfolio of HPC investments that stimulates transformative scientific discovery over the next 5 to 10 years. The HPC investments include the development of science and engineering applications capable of fully exploiting trans-petaflops computers, advanced computing systems capable of sustained multi-petaflops performance, systems and application development software to facilitate the use of these systems, and the networks and cyberinfrastructure needed to connect researchers with these resources. The advice should address the following objectives:

- Building and sustaining a state-of-the-art cyberinfrastructure that stimulates transformative science,
- Creating and sustaining a savvy, capable community that fully exploits HPC resources and drives further development of both scientific codes and the computing infrastructure,
- Incorporating research advances on cyberinfrastructure into production cyberinfrastructure, and
- Developing training and educational programs to grow and foster a vital, diverse computational science community.

The 42 scientists, vendors, and resource providers invited to the workshop were asked to consider the following questions:

1. What is exascale computing and data cyberinfrastructure essential for?
2. What new science does exascale computing and data CI enable?
3. What are the challenges to reach exascale?,
4. What scientific areas have not yet considered exascale, and how do we help them get there?
Appendix E: Participant White Papers

E.1 “The Consumer Digital Infrastructure and Scientific Computing”
David P. Anderson, Space Sciences Laboratory,
University of California, Berkeley

E.2 “An Exascale Brain Initiative”
Eric Jakobsson, National Center for Supercomputing Applications,
University of Illinois

E.3 “Preparing Applications for Exascale Computing”
John Levesque, Cray Inc.

E.4 “Thoughts on Exascale Computing”
David A. Randall, Colorado State University
Summary: the consumer digital infrastructure has the potential to increase the amount of computing power available to U.S. scientists by orders of magnitude, with no hardware investment by funding agencies. The technology to realize this potential is already in place; what is needed is an appropriate organizational structure.

Today’s consumer digital infrastructure (CDI) consists of mass-market devices (e.g., desktop, laptop and tablet computers, game consoles, PDAs, media players, smart phones, VTRs) and the communication networks that connect them (the commodity Internet, cell phone and other radio networks). Driven by the requirements of Internet streaming video and 3-D graphical games, consumer products have become very powerful. Today’s PC has 4 GB RAM, 1 TB disk, and a 1 TeraFLOPS GPU; it is capable of running almost any scientific application. Home network connections are on the order of 10 Mbps, increasing soon to 100 Mbps and then to the Gbps range as optical fiber to homes is deployed.

The CDI currently includes over 1 billion privately-owned PCs and 100 million GPUs capable of general-purpose computing. These have a total computing capability of about 20 ExaFLOPS, growing to 100 ExaFLOPS by 2012. They have on the order of 10 Exabytes of free disk space, accessible via 1 Peta-bps of network bandwidth. In order to handle peak loads, the CDI is over-provisioned and hence underutilized. The average CPU and home Internet connection are used only a few percent of the time they’re available – which, even with power-saving modes, is a significant fraction of the time.

The CDI is an ideal platform for almost all types of high-performance scientific computing, and has several advantages relative to traditional HPC resources:

- The CDI has much larger processing and storage capacity, and hence enables otherwise infeasible science. For example, the Square Kilometer Array radio telescope will generate 0.1 to 1 TB/sec of data. Using the CDI, data could be stored longer (months versus hours) and analyzed in more ways.
- The CDI’s HVAC, electrical power and hardware are paid for by consumers, and the hardware is continuously upgraded to state-of-the-art components. The value of the CDI’s PCs alone is roughly $1 trillion.
- The CDI is self-maintaining: consumers fix their own software and hardware problems.
- Wide geographic, political and infrastructural distribution increases system robustness by eliminating single points of failure.
- Consumer products are the main focus of computing research and development, and have a substantial price/performance advantage over data-center hardware.

To use the excess capacity of the CDI for scientific computing, we first need the consent of the resource owners. This requires incentives that reward consumers for the use of their computing resources, and publicity that makes the public aware of this possibility. Experience has shown that consumers can be motivated by their support for the goals of the research, by the chance to participate in online communities, and by competition based on computational contribution. This is called **volunteer computing**.

Second, we need a software system to dispatch jobs from scientists’ servers to consumer computers, and to execute jobs on those computers. This system must handle a variety of factors that are unique to volunteer computing:

- Scale: millions of nodes and tens of millions of jobs per day.
• Heterogeneity: substantial diversity of software, hardware, and network connectivity.
• Trust and reliability: incorrect computational results may be intentionally returned by malicious participants.
• Communication access: many systems are behind firewalls that allow only outgoing HTTP traffic.
• Availability: sporadic presence and a high churn rate.
• Ease of use: the client software must be extremely simple to install and must work with no configuration or user intervention.

BOINC, an NSF-funded project at UC Berkeley, has developed the only software platform that addresses these issues. BOINC provides server software that lets scientists create volunteer computing projects, and client software, available for all major platforms, that lets volunteers participate in any combination of these projects. BOINC was released in 2004, and there are now about 50 volunteer computing projects. The volunteer population consists of 500,000 people and 1 million computers, providing an average throughput of 12 PetaFLOPS.

Volunteer computing is useful for most “bag of tasks” applications: parameter sweeps, simulations with perturbed initial conditions, compute-intensive data analysis, genetic algorithms. BOINC is used by projects from many institutions, doing research in many areas, including astrophysics, cosmology, climate study, biochemistry, epidemiology, environmental science, cognitive science, genetics, mathematics, nanotechnology, particle physics, quantum computing, and seismology. BOINC supports many types of scientific applications:

• Data-intensive applications; for example, the Einstein@home project analyzes Petabyte-scale data from the LIGO gravitational-wave observatory and the Arecibo radio observatory. BOINC includes a data management system that optimizes network traffic for data-intensive applications.
• Applications in FORTRAN, C, Java, Python, CUDA, OpenCL, or other language systems.
• Multi-thread applications and applications that run in virtual machines.
• Legacy applications for which only an executable is available.
• Jobs that take seconds, days, or months of processing time.
• Jobs with extreme memory, disk, and/or latency requirements; such jobs are sent only to hosts that are able to handle them.

The viability of volunteer computing has been established, but its potential has not been approached. To do so, we must 1) make volunteer computing available to all scientists, and 2) grow the volunteer population from a half million people to tens of millions.

Achieving these goals is, I believe, primarily a matter of organizational structure. Currently, most volunteer computing projects are operated by a single research group. Relatively few groups have the necessary combination of skills and resources, and the proliferation of “brands” makes marketing unproductive. A preferable alternative is as follows:

• Volunteer computing projects would be operated by HPC providers such as supercomputing centers, Hubs, and national Grids, which would port applications, operate BOINC servers, and interface workflow tools to the BOINC back end. The resources of the CDI would then be transparently available to the thousands of scientists served by these organizations, and the specialized computing resources owned by the providers would be freed up for those applications that require them.
• The volunteer computing marketing effort would be consolidated in a single national brand (say, ScienceUSA.org) representing a consortium of funding agencies and HPC providers. Consumers would volunteer resources to research areas (biomedicine, environment, physics, etc.) rather than to specific projects or applications. The allocation of resources to applications would be controlled by the consortium.
E.2 An Exascale Brain Initiative

Eric Jakobsson, National Center for Supercomputing Applications, University of Illinois

**Overall Rationale:** Exascale computing should be a unique tool for unpacking complexity, if the data exist to provide the bases for analysis and modeling of the complex system. To the best of our knowledge, the brain is the most complex bit of matter in the universe. There is an unprecedented and accelerating flood of data about all biological systems at all levels of detail, which unprecedented computing power for its analysis and modeling. To understand human biology, including the human brain, is a special challenge because experiments that can be done on other organisms are too invasive to be done on humans. Hence the tools of comparative computational biology must be used to extrapolate from experimental biology on other organisms to understanding humans.

**Complexity of Biological Systems in General and the Brain in Particular:** At the nanoscale, living cells are extraordinarily complex. In the first place, they are soft condensed matter, which is an unusually complex type of matter. In the second place, by virtue of the compartmentalization of cells by membranes, this particular type of condensed matter displays an extraordinary amount of dielectric heterogeneity at the nanoscale. This heterogeneity is so pronounced that it is highly inaccurate to use the common shorthand of dielectric constant to describe the dielectric properties for this type of matter. Rather one must consider the polarizability of the individual molecules—or rather their polarizabilities (both electronic and rotational, both dynamic and static). The heterogeneity is on the nanoscale, because of the fundamental mode by which biological systems are self-assembled; i.e., the hydrophilic/hydrophobic effect acting on amphipathic nanometer-size molecules in an aqueous environment. It is now possible to capture the same mechanism for self-assembly of synthetic systems, but the synthetic systems fall far short of the complexity of the living systems, because the protocols for synthesis are far less complex and specific than the biological protocols. The “reference book” for the biological protocols is coded into the cell’s DNA. The specific implementation of these protocols is continually interacting with, and continually and cumulatively modified by, the cell’s environment.

All of the above could be said about any living cells. What sets the brain apart as uniquely complex is the degree of cellular heterogeneity, interconnectedness, and functional connection to the organism’s social and physical environment, of the brain. No other collection of cells has this combination of attributes.

The dynamics of the time scale for understanding the brain range from tiny fractions of a second (for responding to stimuli and beginning the restructuring associated with remembering and learning) to hundreds of millions of years (the evolutionary time to the last common ancestor of genes, and the motifs and domains within genes, that the human brain shares with bacteria). The time scales are inextricably linked; what happens on one time scale affects what happens on all the others.
Specific Computational Challenges

Atomic to Molecular Level: The atomically detailed force fields for molecular dynamics simulations need to be improved to cope with the nature of biological soft matter, for example the dynamic continual reordering of biological membranes. To improve force fields substantially from the present level will require extensive calculations at the quantum level, including both high level quantum Monte Carlo and lower level (density functional theory) ab initio md. Additional atom types will be needed to provide accurate interaction potentials in the context of the variety of macromolecular structures, and the creation of each of these interaction potentials will be compute intensive. Even with exascale computing, extension of molecular dynamics simulations will require coarse-grained molecular dynamics and mean field Langevin dynamics to extend the time and distance scales for modeling and simulating cell dynamics. The creation of the interaction potentials for coarse-grained and mean field simulations will be extensive, even by petascale and exascale standards. All of the above tools depend on known macromolecular structures, which up to now have been almost exclusively experimentally determined. Petascale and exascale computing will make possible the accurate prediction of many more structures than can be experimentally determined, and thus extend the scope of dynamical simulations.

Pathway, Network, and Cellular Level: Exascale computing, combined with ever-increasing amounts of gene and protein sequence data from many organisms, will permit the accurate prediction of macromolecular interaction networks by the tools of evolutionary correlation analysis. These techniques, compute-intensive but also potentially very powerful, exploit the fact that for genes that are components of evolutionarily conserved pathways, evolution must be correlated or the interactions will not be conserved. Regulatory networks exist within cells at every level in the process of information transfer from the genotype to the phenotype: transcription of DNA to messenger RNA, alternative splicing of messenger RNA to multiple protein products, and post translational modification of raw protein to fully functional protein. They are comprised of all possible classes of macromolecular interactions: protein-protein, protein-nucleic acid, and nucleic acid-nucleic acid interactions. Reconstruction of these networks from the raw data, and modeling the dynamics of the reconstructed networks, will take as much computer power as one can imagine, let alone build. Interaction of the cell with the extracellular environment, mediated through receptors, channels, and transporters on the cell surface, adds other dimensions of complexity. The most accurate models of the signaling pathways must take into account stochastic behavior of small numbers of molecules; computationally more modest differential equation models will not suffice to accurately describe the nonlinear dynamics.

Brain Organization: Increasing amounts of digitized data on brain structures and interconnectivity have the potential, combined with computational modeling at the cellular, molecular, and pathway levels, to facilitate our understanding of how higher brain functions emerge from the lower levels of biological organization. Clearly this has to do with the above-mentioned cellular heterogeneity and
proliferation of interactions in the brain. (A single cell may receive synaptic input from hundreds of other cells and in turn send output to hundreds of others.)

*Population Modeling:* Humans and many other organisms are highly social. Social interactions are essentially interactions among brains. Population models focusing on relatively simple interactions among humans (spread of infectious disease) are already slated for petascale application. Ideas are as infectious as microbes, but much more complex, so social interaction models based on realistic models of brain function, are bound to be exascale applications.

*Spinoffs:* All of the modeling at the molecular level will have application to computational nanoscience, and vice versa; i.e., computational nanoscience will be highly relevant to biomolecular computation. All of the analysis and modeling technology at the pathway, network, and cellular level will be relevant to the corresponding modeling of other types of cells and tissues. Population models will be put in the context of ecological and economic models to enhance their accuracy and scope.

**Understanding the Brain as the ultimate Problem**

Understanding the brain is the ultimate problem in the sense that the brain is our instrument for understanding everything else. Just as understanding the workings of a measuring instrument is essential for understanding the data from that instrument, so also should we expect that understanding the brain will color our understanding of everything.
E.3 Preparing applications for Exascale computing
John Levesque, Cray Inc.

The path-forward to exascale computing will be difficult given the current state of applications. The potential Exascale architectures will rely on very powerful nodes, either multi-core(16-24) with accelerators or many-core (100-200) nodes. Both of these architectures necessitate that the applications are able to utilize the NUMA memory architecture on the nodes. If the current applications are all MPI, scaling to high processor/core counts will generally not be possible. Utilization of the on node parallelism and memory hierarchy will be difficult; however, it will be the only way that an application can hope to sustain an exaflop in performance.

The last time the community faced such a programming challenge was the move to MPI from shared memory Vector systems. As in that case, the application restructuring will be significant and as in that case there will not be a silver bullet that automatically converts an existing code to effectively utilize the target architecture.

In order to move existing all-MPI applications to the exascale architectures a concerted effort must be put in place to facilitate the refactoring of the application. The work will consist of identifying high level shared memory parallelism in the applications. Only a few existing DoE applications have taken such a step. Most notable is Community Atmospheric Model (CAM) which is written with very high level OpenMP structures. Consequently, CAM performs well on the multi-core nodes, with MPI between nodes (or sockets) and OpenMP within the node (or socket). Another application that already has the OpenMP at a high level is the fusion code GTC. It is important to understand that low level OpenMP, which is easily achieved by automatic parallelization by the compiler is not sufficient to provide the granularity required to utilize the shared memory node.

In the case of multi-core sockets with accelerators, the vectorization of the application is also a requirement to achieve a performance gain on the accelerator. While CAM may have good threading at a high level, its computational routines are not very vectorizable. Significant restructuring of the low level routines is required to achieve the desired performance on the accelerator.

The refactoring of the application may take algorithm changes, significant source code changes and it could take several years to accomplish. Additionally, the re-write of the application should be performed in a manner that assure performance portability across both exascale designs as well as existing multi-core MPPs. While initial accelerator ports have relied heavily on writing CUDA routines, the directive based approach is showing great promise. The directive based approach uses a proposed set of extension to OpenMP that allows the user to direct the compiler on how to most effectively utilize the target accelerator. An application written with such OpenMP extensions at a high level would be compilable for either the of the two aforementioned architectures as well as existing multi-core MPP systems.
This application restructuring effort should be done with teams that include code developers, users of the application and optimization specialists from the vendor. Additionally, programming tools need to be developed to facilitate the porting and optimization of the applications.

The DoE Office of Science has commissioned such an effort to address the application readiness of several important applications for the next generation Oak Ridge National Laboratory (ORNL) system.

**ORNL Application Readiness effort**

For the past year, teams have been established to take six of the major DoE applications and ready them for the OLCF3 system which is proposed to be a multi-core MPP system with accelerators on each node. The teams consist of representatives from ORNL, system and processor vendors, and application developers. The six applications areas are Climate, Combustion, Material Science, Bio-molecular simulation, Nuclear Safety and Carbon Sequestration.

The process is to first identify the important Science problem, whose successful analysis would advance the state-of-the-art in the field. The problem would be one that could not be solved without the proposed system. Analysis is then performed on the application to identify the profile of performance for the target problem on the target system. Once this work is completed, work is identified to move the current application to a form that would be portable across existing and future exascale architecture. This analysis can only be performed with the appropriate members of the teams.

The goal for each of the teams is to first identify the most important kernels in the application and use hand-coded CUDA as well as directive based implementation to compare their performance to the comparable state-of-the-art multi-core system. Once the individual kernels are developed and analyzed, the communication and memory management of the resultant hybrid application is examined to identify kernels that can be run asynchronously with other kernels and communication. Given the power of the accelerator, overlapping communication on the accelerator with computation on the multi-core socket and communication between nodes is paramount.

At this point an accurate projection can be obtained for moving the application to the proposed system. As mentioned above, one of the applications is requiring an algorithmic change and another will take several years to complete.

The intent is to have some subset of the applications ready for production science runs when the machine is delivered.

**Development of an exascale programming environment**

The work of refactoring the initial six applications is feeding into the design of an exascale programming environment to facilitate the porting and optimization of the multitude of other DoE applications. No one expects all application developers to go through the effort put forth by the initial code porting
projects discussed above. For example, with the comparisons between the CUDA hand coding and the directive based approach, the compilers can be fine tuned to give comparable performance. Supplying a Fortran, C or C++ language interface for the target system will in itself significantly reduce the amount of work to port/optimize applications.

An enormous part of the programming environment is the maturity of compilers to generate efficient code for the target architectures; however, no one expects the compiler to perform the parallelization and vectorization automatically. There must be a mechanism to allow the user to assist the compiler in the analysis of the application. This is where a programming environment built around the compiler can significantly aid the analysis and refactoring process. The programming environment must include:

1) Statistics gathering to facilitate the identification of high level parallel structures
2) Interactive parallelization facility that allows the user to participate in the parallel analysis
3) Facility to allow the user to examine variable usage throughout the application
4) Interactive scoping facility to aid in classifying the variables and identifying what variables need to be transferred to the accelerator, back from the accelerator or resident on the accelerator
5) Statistics gathering to allow for fine tuning of the accelerator code
6) Facility to generate portable application with OpenMP w/extensions

The importance of integrating the programming environment with the compiler cannot be overstressed. To perform most of the tasks mentioned, the target compiler must be invoked to supply the parallel analysis. Today all compilers perform some form of inter-procedural analysis which is required in parallelizing high level looping structures.

In addition to the programming tools the runtime on the accelerator must allow for parallel nested regions. In many cases a single looping structure will not have enough parallel iteration. Allowing multiple parallel loop nests will allow for the required amount of parallelism. Additionally the runtime must allow the application to call other routines on the accelerator. For example, the physics loop in CAM calls 200-300 routines, if the compiler and runtime allows such a loop to be executed on the accelerator without inlining, the task of porting that application would be reduced significantly.

**Recommendations**

1) When investigating the porting/optimization of an application to the exascale architecture, code teams must be established that include users of the code, developers of the code and the vendor of the exascale system.
2) Teams should strive to develop a performance portable application that runs on existing multi-core MPP as well as the target exascale system
3) Teams should work closely with the vendor compilers to improve the capability of the compilers to generate efficient code from an application decorated with appropriate directives
4) Programming tools tightly integrated with the vendor’s compiler must be developed and used to reduce the effort for future application conversions
E.4 Thoughts on Exascale Computing
David A. Randall, Colorado State University

• What is Exascale computing and data cyberinfrastructure essential for?
The emergent behavior of maximally multiscale systems in which 1) the smallest important scales are many orders of magnitude smaller than the largest scales, and 2) all or many intermediate scales manifest their own emergent behaviors which then affect the larger scales.

The climate system is an example.

A living thing is another example.

• What new science does Exascale computing and data cyberinfrastructure enable?
It is interesting to make a list of scientific problems that Exascale computing does NOT enable. In climate research, a good example is simulation of ice age cycles. A convincing simulation would have to run about one million simulated years. It appears that emerging Exascale technologies will NOT enable such simulations.