BLUE WATERS

SUSTAINED PETASCALE COMPUTING

Brainstorming High Performance Computing and Data Workshop Summary

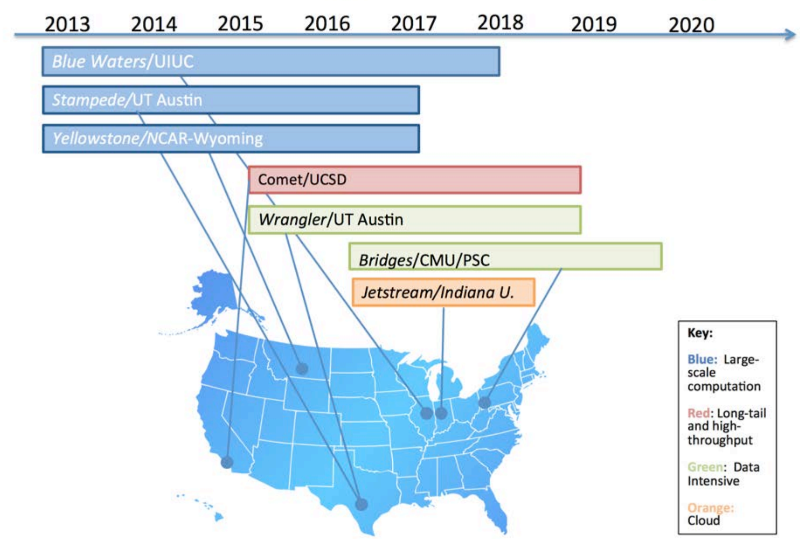
William Kramer
National Center for Supercomputing Applications, University of Illinois
http://bluewaters.ncsa.illinois.edu

NCSA NSF GREAT LAKES CONSORTIUM FOR PETASCALE COMPUTATION CRAY

Motivations

- Blue Waters, Stampede and other systems are reaching maximum impact
- NSF chartered a National Academy Study "Committee On Future Directions For Nsf Advanced Computing Infrastructure To Support U.S. Science In 2017-2020"
  - *The contribution of high-end computing to U.S. leadership and competiveness in basic science and engineering and the role that NSF should play in sustaining this leadership;*
  - *Expected future national-scale computing needs;*
  - *Complementarities and trade-offs that arise among investments in supporting advanced computing ecosystems;*
  - *The range of operational models for delivering computational infrastructure,*
  - *Expected technical challenges to affordably delivering the capabilities needed for world-leading scientific and engineering research*

NSF-Supported Computational Investments Reflect Increasing National Diversity



Workshop Charters

- First workshop –Requirements and Needs Workshop
  - concentrate on *gathering comprehensive information about the needs of the "NSF sponsored science and engineering research community" for advanced capacity through capability computing, including data-intensive and compute-intensive. The workshop will focus on the research community needs at the high end of spectrum and will attempt to articulate the value, contributions and impacts of at scale computing for the advancement of science, engineering and other research.*
- Second workshop – Implementation Alternatives Workshops
  - concentrate on *alternative operational models for providing computing capability, including the role of private industry, NSF centers, support of science workflows, and data management. This covers both how resources are provided to scientists and how they are managed.*

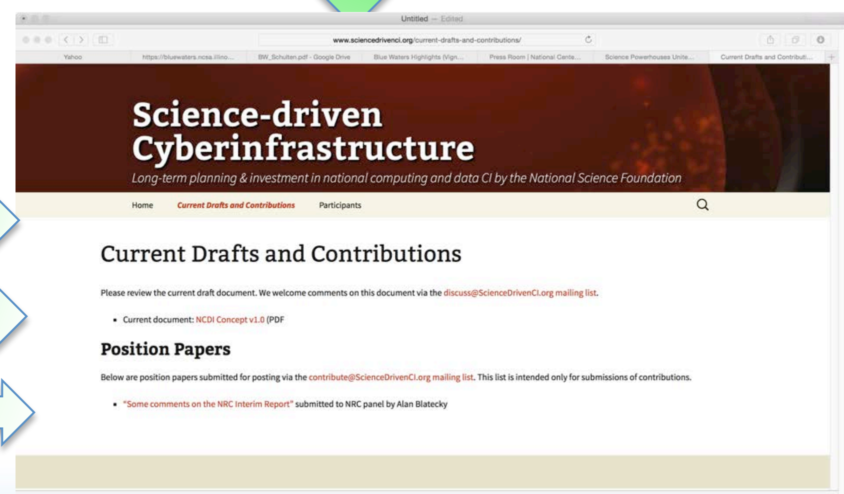
Workshop Focus

- High End of the Advanced Digital Ecosystem includes:
  - leadership class computing capabilities – levels 1 and 2
  - leadership class data capabilities – levels 1 and 2
  - leadership class networking
  - support services and service models for leadership class: multi-disciplinary computational science expertise
  - advanced applications software capable of fully exploiting the resources
  - workforce development for current and next generation researchers and technical staff to enable effective use of the ecosystem
  - outreach and training to enable new researchers and new domains

ww.brainstormhpcd.org and http://www.sciencedrivenci.org

High End of the Advanced Digital Ecosystem:

- leadership class computing capabilities – levels 1 and 2
- leadership class data capabilities – levels 1 and 2
- leadership class networking
- support services and service models for leadership class: multi-disciplinary computational science expertise
- advanced applications software capable of fully exploiting the resources
- workforce development for current and next generation researchers and technical staff to enable effective use of the ecosystem
- outreach and training to enable new researchers and new domains



Workshop Process

- Invited almost 200 people to share their input
  - From many areas
    - Range science and engineering and research disciplines
    - Experimentalists, data analysts, simulators
    - HPC traditional users, non-traditional users
    - Service Providers (limited)
    - Large experiments
- Methods of participating
  - Attending
  - Position papers
  - Informal input
  - Advocate direct input to NRC

Attendees

- **Workshop 1**
- Stuart Anderson, LIGO/Caltech
- Richard Arthur, GE Global Research
- Klaus Bartschat, Drake University
- Greg Bauer, NCSA/University of Illinois
- Martin Berzins, SCI Institute University of Utah
- Wes Bethel, LBNL
- Tom Cheatham, University of Utah
- Said Elghobashi, University of California, Irvine
- Jim Fonseca, Purdue University
- Steven Gottlieb, Indiana University
- Bruce Harmon, Iowa State University
- Anna Hasenfratz, University of Colorado Boulder
- Peter Kasson, University of Virginia
- Fatemeh Khalili-Araghi, University of Illinois at Chicago
- William Kramer, NCSA/University of Illinois
- John Levesque, Cray Inc
- David Lifka, Cornell University
- Paul Mackenzie, Fermilab
- Pieter Maris, Iowa State University
- Michael Norman, University of California San Diego
- Steve Oberlin, NVIDIA
- Nikolai Pogorelov, University of Alabama in Huntsville
- Thomas Quinn, University of Washington
- Ralph Roskies, Pittsburgh Supercomputing Center
- Mark Scheel, California Institute of Technology
- Ed Seidel, NCSA/University of Illinois
- Todd Simons, Rolls-Royce Corporation
- Dan Stanzione, University of Texas at Austin
- John Towns, NCSA/University of Illinois
- Shaowen Wang, NCSA/University of Illinois
- Paul Woodward, University of Minnesota
- Donald Wuebbles, University of Illinois
- Pui-kuen (P.K) Yeung, Georgia Institute of Technology
- Shiwei Zhang, College of William & Mary

- **Workshop 2**
- Stuart Anderson, LIGO
- Dinshaw Balsara, University of Notre Dame
- Bill Barth, TACC
- Greg Bauer, NCSA
- Cristina Beldica, NCSA
- Jerzy Bernholc, North Carolina State University at Raleigh
- David Dixon, University of Alabama
- Thom Dunning, Univ of Washington
- Rama Govindaraju, Google
- Robert Harrison, NRC Observer/Stonybrook
- Thomas Hauser, University of Colorado Boulder
- Victor Hazlewood, University of Tennessee – JICS
- Curtis Hillegas, Princeton
- Fatemeh Khalili-Araghi, University of Illinois at Chicago
- Kevin Kissel, Google
- Bill Kramer, NCSA
- Mike Levine, PSC
- Honggao Liu, Louisiana State University
- Jagnnathan Ramanujam, Center for Computation and Tech – LSU
- Barry Schneider, NIST
- Ed Seidel, NCSA
- John Towns, NCSA
- Frank Tsung, UCLA
- Jorge Vinals, Minnesota Supercomputing Institute
- Liqiang Wang, University of Wyoming
- Nancy Wilkins-Diehr, San Diego Supercomputing Center
- Steve Wolff, Internet2
- Paul Woodward, Minnesota Stellar Explosions

Institutions

Workshop 1

1. California Institute of Technology
2. College of William & Mary
3. Cornell University
4. Cray Inc
5. Drake University
6. Fermilab
7. GE Global Research
8. Georgia Institute of Technology
9. Indiana University
10. Iowa State University
11. LBNL
12. LIGO/Caltech
13. NCSA/University of Illinois Urbana-Champaign
14. NVIDIA
15. Pittsburgh Supercomputing Center
16. Purdue University
17. Rolls-Royce Corporation
18. University of Alabama in Huntsville
19. University of California San Diego
20. University of California, Irvine
21. University of Colorado Boulder
22. University of Illinois at Chicago
23. University of Illinois Urbana-Champaign
24. University of Minnesota
25. University of Texas at Austin
26. University of Utah
27. University of Virginia
28. University of Washington

Workshop 2

1. GE Global Research
2. Georgia Institute of Technology
3. Google
4. Internet2
5. LIGO/California Institute of Technology
6. Louisiana State University
7. Minnesota Supercomputing Institute
8. National Institutes of Standards
9. NCSA/University of Illinois Chicago
10. NCSA/University of Illinois Urbana-Champaign
11. North Carolina State University at Raleigh
12. Pittsburgh Supercomputing Center
13. Princeton University
14. Rolls-Royce Corporation
15. San Diego Supercomputing Center
16. University of Alabama in Huntsville
17. University of California San Diego
18. University of California, Los Angles
19. University of Colorado Boulder
20. University of Illinois at Chicago
21. University of Illinois Urbana-Champaign
22. University of Minnesota
23. University of Notre Dame
24. University of Tennessee
25. University of Texas at Austin
26. University of Washington
27. University of Wyoming

Position Papers

- **Searches for Gravitational Waves with the Laser Interferometer Gravitational-Wave Observatory (LIGO)** — Stuart Anderson, California Institute of Technolgoy
- **Position Paper to National Research Council** — Richard Arthur, GE Global Research
- **PGAS Languages: An Easy-Entry Paradigm for Peta/ExaScale Computing** — Dinshaw S. Balsara, University of Notre Dame
- **Comments on Topics 4 and 6 on the NRC List** — Klaus Bartschat, Drake University
- **Quantum Simulations in Materials Design, Nano Science and Technology** — Jerry Bernholc, Center for High Performance Simulation and Department of Physics, NC State University
- **The Blurry Line Between Exascale and Big Data** — E. Wes Bethel, Lawrence Berkeley National Laboratory
- **Supplying Cycles for High Performance Computing with a Focus on Chemistry: Community Needs to be Considered** — David A. Dixon, Department of Chemistry, The University of Alabama
- **Comments on Interim Report: "Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020"** — Thom Dunning, Pacific Northwest National Laboratory & University of Washington
- **No Results Left Behind: The Case for Testing in Computational Science and a Test Case in Nanoelectronics Modeling** — Jim Fonseca, Purdue University
- **Lattice Field Theory for High Energy Physics** — Steven Gottlieb, Indiana University; Anna Hasenfratz, University of Colorado; Paul Mackenzie, Fermilab; Robert Sugar, UCSB

Position Papers

- **Lattice Field Theory Computations** — Steven Gottlieb, Indiana University; Anna Hasenfratz, University of Colorado; Paul Mackenzie, Fermilab; Robert Sugar, UCSB
- **NSF 2nd HPCD Workshop 2015** — Rama Govindaraju, Google
- **Change or Perish** — Bruce Harmon, Iowa State University
- **It's going to take more than hardware to advance the state of the art in scientific computation** — John M. Levesque, Cray Inc.
- **Tighter development cycles between scientific applications and advanced computational infrastructure** — Peter Kasson, University of Virginia
- **Molecular Dynamics simulations of Biological Systems** — Fatemeh Khalili-Araghi, University of Illinois at Chicago
- **Accelerating Scientific Discovery and Engineering Practice through Advanced, High Spectrum Computing and Data Analysis (tables)** — William Kramer, NCSA/University of Illinois; et al.
- **Thoughts on the NSF Future Directions Interim Report** — Glenn K. Lockwood
- **Position Paper for HPCD Brainstorm Workshop** — Pieter Maris, Iowa State University
- **appendixB_answers_Maris** — Pieter Maris, Iowa State University
- **Needs and Opportunities In Cosmology In The Era of Ultra-Deep Surveys** — Michael L. Norman, UCSD
- **The Other 90 Percent** — Steve Oberlin, NVIDIA

Position Papers

- **Thoughts on The Future of NSF-supported Advanced Computing from a Numerical Relativity and Computational Astrophysics Perspective** — Christian Ott and Mark Scheel, Caltech
- **High-Performance Computing Challenges in Space Physics Simulations** — Nikolai V. Pogorelov, University of Alabama in Huntsville
- **Computational Challenges for Galaxy Formation** — Thomas Quinn, University of Washington
- **Position Paper for 2nd Brainstorming HPCD Workshop** — Barry I. Schneider, NIST
- **High Performance Computing Policy Paper** — Todd Simons, Rolls-Royce
- **Reflections on an Observation of the Interim Report of the Committee on Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science in 2017-20** — Dan Stanzione, Texas Advanced Computing Center
- **Towards HPC in the Cloud** — Liqiang Wang, Department of Computer Science, University of Wyoming
- **Scalable CyberGIS Analytics for Solving Complex Environmental, Geospatial, and Social Scientific Problems** — Shaowen Wang, University of Illinois
- **Observations and Recommendations Regarding NSF's Support for High-Performance Computing** — Nancy Wilkins-Diehr, Wayne Pfeiffer, and Richard L. Moore, SDSC
- **High-Resolution Earth System Global and Regional Modeling for Climate Assessment and Policymaking Require Advanced Computing Infrastructure** — Donald Wuebbles, University of Illinois at Urbana-Champaign; Warren Washington, Gerald Meehl, and Tom Bettge, National Center for Atmospheric Research

Position Papers

- **Fluid Dynamics and Turbulence: the Case for NSF-HPC in 2017-2020** — P.K. Yeung, Georgia Tech
- **An Integrative, Cross-Foundation Cyberinfrastructure for Science & Engineering Research** — multiple authors
- **Simulating the First Galaxies and Quasars: The Bluetides Cosmological Simulation** — Blue Waters PI: Tiziana Di Matteo, Carnegie Mellon University
- **Solving Prediction Problems in Earthquake System Science** — Blue Waters PI: Thomas H. Jordan, SCEC
- **Enabling Breakthrough Kinetic Simulations of the Magnetosphere** — Blue Waters PI: Homayoun Karimabadi, UCSD
- **Accelerating Nanoscale Transistor Innovation with NEMO5** — Blue Waters PI: Gerhard Klimeck, Purdue University
- **Design and Management of Satellite Assets to Advance Space-Based Earth Science** — Blue Waters PI: Patrick Reed, Cornell University
- **Simulation of Turbulent Stellar Hydrodynamics** — Blue Waters PI: Paul Woodward, University of Minnesota
- **High-Resolution Climate Simulations** — Blue Waters PI: Donald Wuebbles, University of Illinois at Urbana-Champaign

WS 1 Working Groups

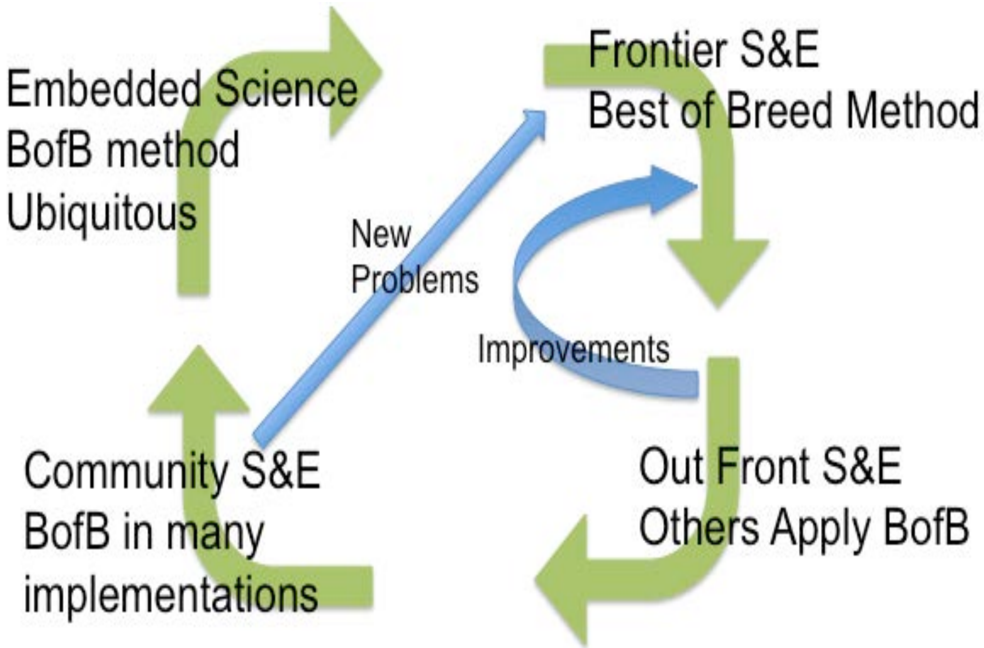
- Working Group 1 – *High Spectrum Computing and Analysis Needs and Priorities*
  - Identify the characteristics for advanced high spectrum computing and data infrastructure that enables integrated discovery involving experiments, observations, analysis, theory, and simulation. Evaluate the match between resources and demand for the high spectrum of systems, for both compute- and data-intensive applications, and the impacts on the research community if NSF can no longer provide state-of-the-art computing and data analysis for its research community. Provide prioritization guidelines for NSF to meet the needs of high spectrum computing and analysis.
- Working Group 2 – *Access to Sufficient High Spectrum Resources*
  - There are challenges facing researchers in obtaining access to advanced computational and data analysis resources. These range from allocation processes, over subscription of resources and the need to frequently migrate from one resource to another. Some have said the easiest part is actually getting the *at-scale* application to work. In other cases, the communities may be self limiting requests to known resources rather than requesting what is actually needed. In reviewing the white papers and other sources, estimate the true computational requirements that would enable all areas of science and engineering to make timely progress in both best of breed problems and common practice problems. Where possible, provide quantitative data on computing needs.
- Working Group 3 – *Risk, Opportunities and NSF's Role Fostering High Spectrum Science and Engineering*
  - There are multiple technical challenges to building future, more capable advanced computing and data systems for the next decade. Technology limitations will make some approaches more difficult for applications to use in a productive manner. How should NSF best respond to the challenges for sustained application performance and researcher productivity in the future? What is the optimum ways and frequency to collect requirements and status in the rapidly changing environments? What is the best way to motivate and enable science teams to be ready to use high spectrum resources in a highly efficient manner to do best of breed problems in the most timely manner? What are the risks and opportunities to scientific leadership in current US plans for extreme scale computing and what can NSF do to address and reduce the risks and maximize opportunities
- Working Group 4 – *Computation and Data Analysis*
  - In reviewing the submitted white papers and other information, analyze the requirements for computation and data analysis for open science and engineering that will need to be met to enable high-spectrum science and engineering in the next decade. Provide a break-down of the system and architectural requirements that are anticipated to enable high spectrum science and engineering.
- Working Group 5 – *Storage and Data Movement*
  - In reviewing the submitted white papers and other information, analyze the requirements for storage and data movement for open science and engineering that will need to be met to enable high-spectrum science and engineering in the next decade. Data movement is expensive both in terms of energy, investment and human effort. Identify strategies to minimize data movement to make science and engineering teams as productive as possible. Discuss the current number of storage hierarchies (memory, rotating disk, tape) and project the types and levels of hierarchies that will be likely and effective over the next decades.
- Working Group 6 – *Workflow and Methods*
  - In reviewing the submitted white papers and other information, analyze the requirements for advanced workflows and methods that will need to be met to enable high-spectrum science and engineering in the next decade. Time to insight in many science areas is not only related to the largest scale work steps, but may be dominated by other workflow steps. Today, many workflows have been developed on a project or team basis, with many different assumptions. Project, where possible, the commonalities of workflows and where there could be opportunities for synergies and optimizations.

WS 2 Working Groups

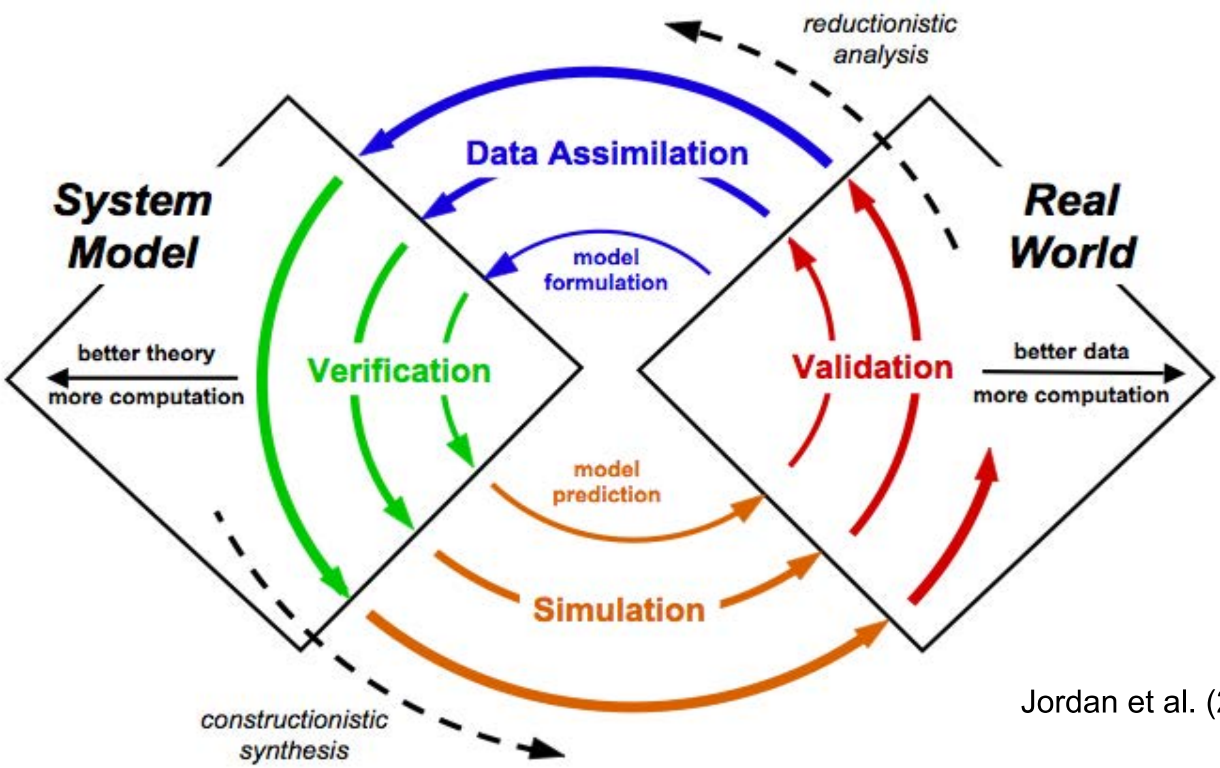
- Breakout # 1: Alternative Design Models
  - Identify and describe the alternative design models for providing resources and services. Example models could be implementation by discipline orientation (common resources shared by many/all disciplines to resources subdivided for discipline specific resources); system integration and scale (tightly integrated to very loosely integrated); resource distribution (localized into a few facilities to highly geographically distributed); etc. Identify the points of "diminishing returns" for each of the models and the potential strengths and weaknesses for each. Each alternative should be described in two three sentences.
- Breakout # 2: Alternative Provisioning Methods
  - Identify the primary alternatives that NSF could use to provision the necessary resources for high spectrum science, engineering and research. Triage these alternatives to identify the three to five that would likely be mostly feasible and implementable. For each alternative, identify the role(s) that NSF, universities, private industry and other federal agencies might have in making the alternative succeed. Each alternative should be described in two three sentences.
- Breakout # 3: Alternative Analysis Criteria
  - Develop the important evaluation criteria that should be used to assess and select alternative methods for NSF to deploy high spectrum resources. These criteria should include perspectives from all stakeholders and potential uses, but needs to be aggregated so that there are no more than 10 criteria for assessment. Additionally, identify the important cost model components and describe them so the alternatives can be assessed. Each criteria should be described in two three sentences.
- Breakout #4: NSF Administrative Implementation Alternatives
  - Identify the primary implementation alternatives that NSF might implement to provision the necessary resources for high spectrum science, engineering and research. What processes are available and feasible for NSF to implement (current methods of site/ equipment awards, contracts, intergovernmental agreements, MFREC processes, etc.) Triage these alternatives to identify the three to five that would likely be mostly feasible and implementable. For each alternative, identify the role(s) that NSF, universities, private industry and other federal agencies can have in making the alternative succeed.
- Breakout #5: Alternative Evaluation and Assessment – 1
  - Given the alternatives developed in Breakouts # 1 and 2, and the evaluation criteria developed in Breakout #3, evaluate the alternatives with strengths and weaknesses; risks and uncertainties and perform an initial qualitative cost assessment.
- Breakout #6: Alternative Evaluation and Assessment – 2
  - Given the alternatives developed in Breakouts # 1 and 2, and the evaluation criteria developed in Breakout #3, evaluate the alternatives with strengths and weaknesses; risks and uncertainties and perform an initial qualitative cost assessment.

Flow of Influence

- It often takes tremendous computing power to develop new ways to solve the most challenging problems
- Very specialized approaches are needed
- Improving the algorithms (methods of solving problems) decreases the time it takes to solve a problem at least as much as new hardware.
- What is done on a high end systems typically becomes common practices a decade later on other systems, and is used for many standard things within another decade



Inference Spiral of System Science



Jordan et al. (2010)

- **As models become more complex and new data bring in more information, we require ever increasing computational resources**

Slide courtesy of T Jordan - SCEC

Characteristics for Future @Scale

- **Dramatically increase fidelity** in models and simulations to improve insights and address new problems.
  - Increasing use of multi-scale and multi-physics. These are needed to accurately explore simulated phenomena.
  - Increasing resolution.
  - Increasing complexity.
  - Increased number of "ensemble" trials.
- **Longer simulated time periods**
  - often required to accurately simulate the system of interest
  - simulations of larger systems often require longer periods of time to stabilize
- **Increased number of problems to address**
  - The first 100 million all-atom simulations were completed in 2013. By 2020 there will be tens to hundreds of teams doing hundreds to thousands of 100 million atom simulations

Characteristics for Future @Scale

- **Changing workflow methods**
  - Deadline driven analysis for experimental and observational data
  - visualization to interpret and understand the simulation and analysis results
  - Malleable/elastic resource management for application load balancing and resiliency.
  - Automation through workflows to support repeatability of computational/ analytical solutions.
  - Use of data model programming methods,
- **Increased integration with data sources and increased use of simulation** data products.
  - data from multiple experiments and observations
  - Observation data assimilation
  - Track-1 systems enables them to produce community data sets that are then useful for others

Characteristics for Future @Scale

- **Changing algorithmic methods**
  - Substantially improve their algorithmic methods to reach new research goals over the next five to ten year
  - Not just to address new computer architectures
  - Also to improve the time to solution independent of hardware changes and to develop the algorithms needed for multi-physics and multi-scale simulations.
  - Use of adaptive gridding and malleable/elastic resource management
  - Applications load balancing and resiliency will expand. Improving load balancing is critical to overcoming both Amdahl's law limits and the increasing variation in system component performance
  - Need resources to re-engineer, test and validate

CONSENSUS OUTCOMES

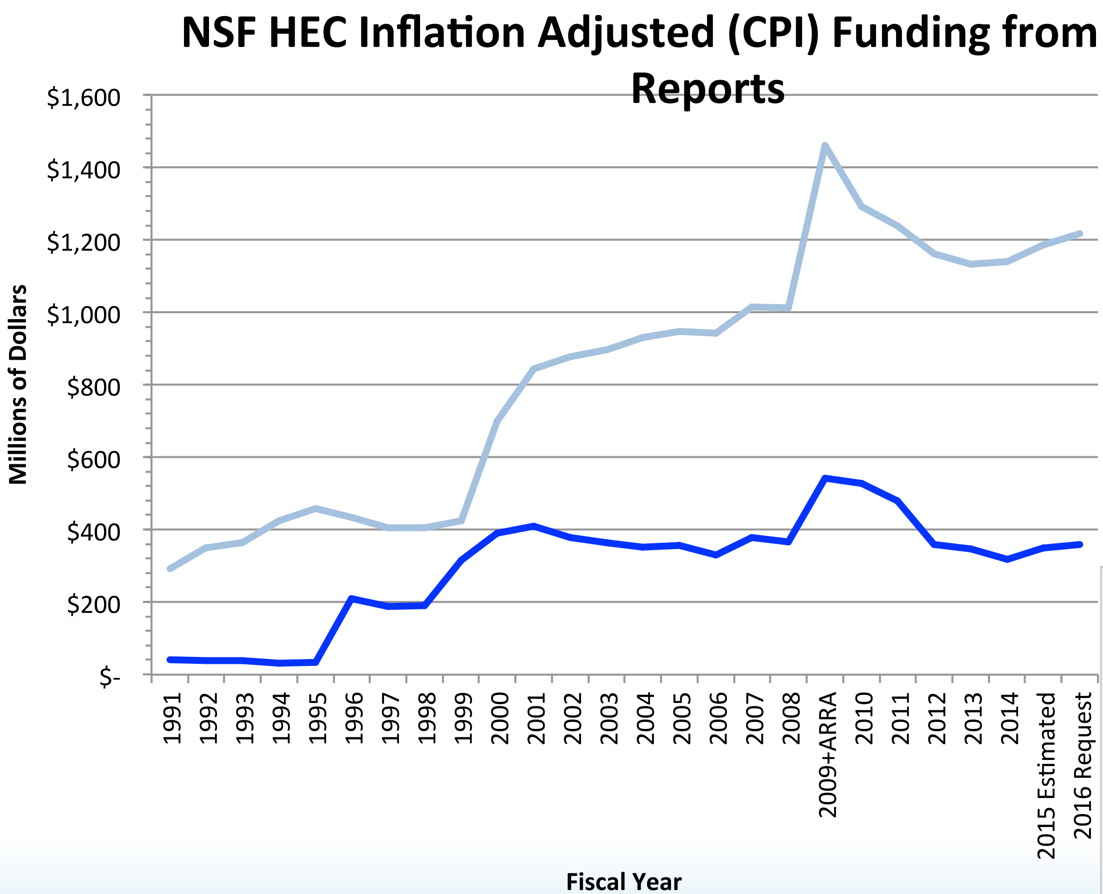
Workshop 1 and 2

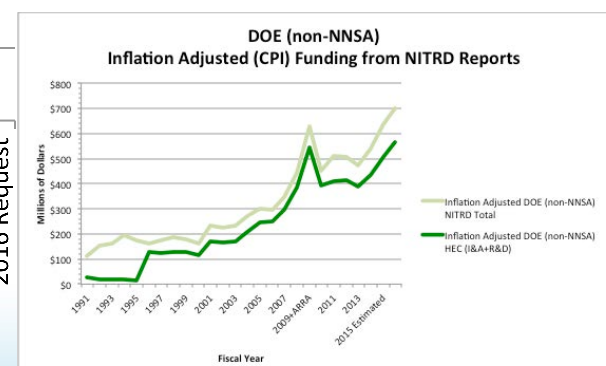
- There is compelling evidence that computational and data intensive computing are are now implicitly intertwined with all areas of discovery.
- High end computational and data intensive computing resources are critical enablers for discovery in nearly all fields of science and engineering as well as for technological advancement.
- NSF has made great progress in expanding the use of high end computing and data analysis throughout the research communities
- The NSF is the primary agency responsible for basic science in the US across all fields.
  - NSF provides advanced computational and data resources for NSF grantees and researchers supported by other agencies such as the DoE, NIH, NASA, etc.
  - At last count ~50% of all of the NSF supported XSEDE compute cycles go to non-NSF supported research.

Workshop 1 Conclusions

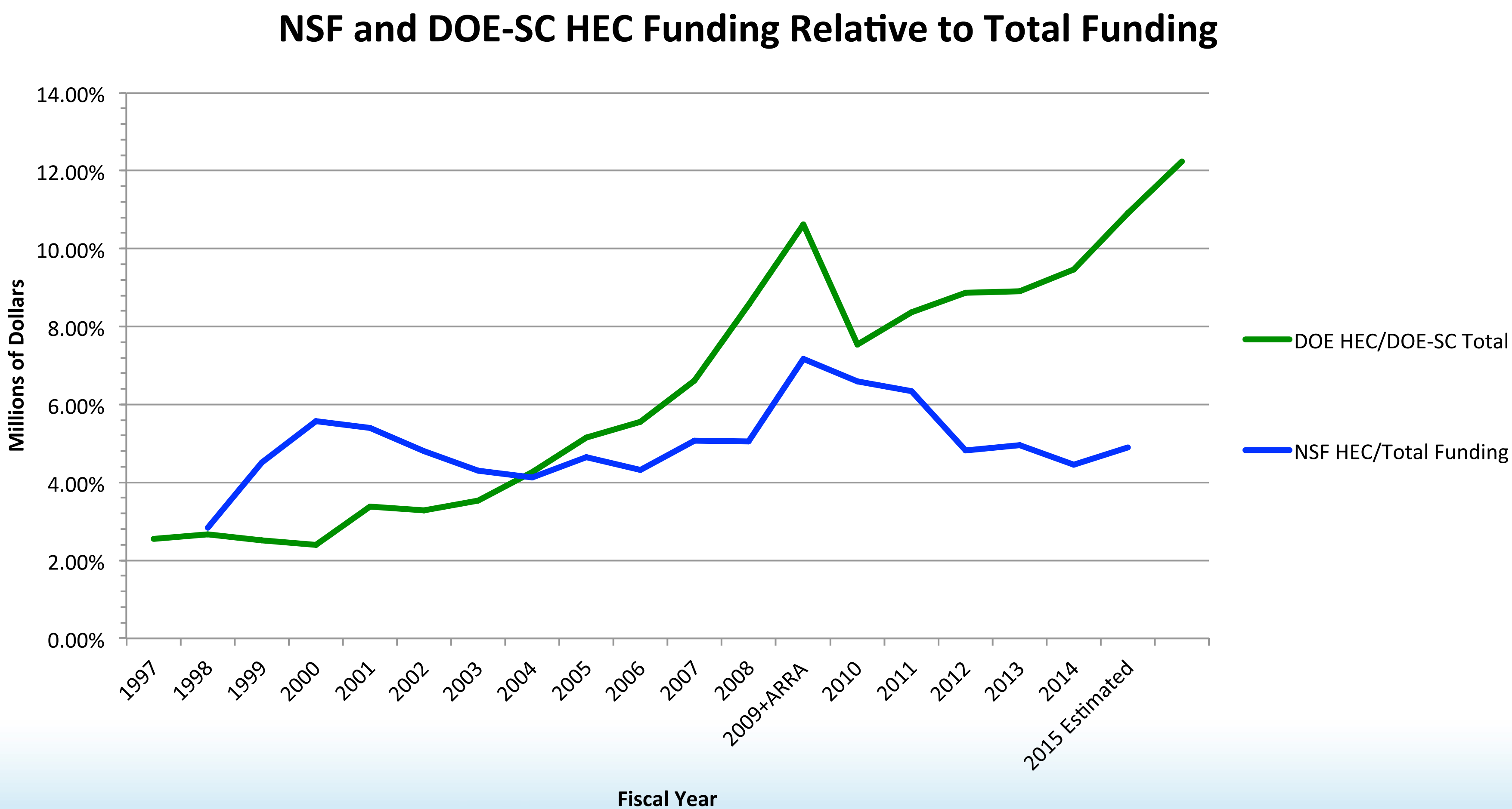
- Because NSF has the broad, national mission of advancing science and engineering and insuring the scientific and economic competitiveness of the nation, it is vitally important for NSF to provide a series over time of High End computational and data analysis tier 1 and tier 2 systems to meeting increasing demands of high spectrum digital science.
  - The NSF CI has to include systems of the scale and capability
  - Benefits for a variety of platforms, but more than 2 to 4 systems will create inefficiencies
    - Minimum size/capability arerequired
- Scientific needs increasingly require high performance capability.
- The sustained funding for HPCD has led to tremendous advances in many fields and NSF can not let that progress and momentum stop.
- “Frontier” science and engineering” will be seriously inhibited if NSF does not to invest in high end computing and data infrastructure the support

NSF Funding For HECD





Comparing Percent of HEC Funding to Total



Examples of Science Goals for Next Generations

- Full Cell and then Human Cell all atom simulations
- Full cycle (multiple days) of Space Weather
- Seismic Impacts for all buildings including 1 story structures
- … several more from vignettes

Workshop 1 Conclusions

- Petabyte scale data repositories become more necessary.
  - NSF should invest in resources to support community based, curated, large data repositories to be used by 100's to 1000's of scientists.
    - Organized, High End simulation data products typically have a useful life of 5-8 years
    - Some products may be valuable for a decade of more
- Important to enable complex workflows that are converging to be simultaneously computation-intensive and data –intensive.
- Invest in SW development and improvement
  - NSF should be investing in HPCD university programs related
  - Foster parallel programmed computational science curriculum.
- Industry and therefore economic support will migrate to where the expertise exists to do industrial based research
  - SW tool chain, libraries, algorithms are necessary enablers
  - Validation of codes and reference data sets are required
  - Allocations for 'non-discovery' (e.g meshing) use is critical for economic impacts

Workshop 2 Conclusions

- Computational and data intensive computing are a critical enablers for discovery in all fields of science and engineering and for technological advancement.
- The success of NSF's HPCD investments and the pervasiveness of it created an environment in which that such HPCD resources are now assumed.
- Need a NSF-wide assessment of needs and risks with respect to provisioning high spectrum resources and services to enable CDS&E.
- Investments in such facilities are equivalent to investments in experimental facilities (e.g. LIGO, OOI, NEON,...) with long term investment plan.
- Investment in high spectrum resources and services are necessary to maintain momentum;
- Needs to be a significant increase in coherence for NSF's investments in CI
- At least a real doubling of investment in high spectrum resources over the next 5 years.
- Geographical consolidation (2-4 major facilities with high performance network access) is being driven by the need co-locate compute and data resources.

Workshop 2 Conclusions

- NSF must have a frontier science-driven digital ecosystem to support the broad open science community
  - Should cede this responsibility to other agencies given that NSF has no control of prioritization due to missions of other agencies.
  - There are issues to access to resources at mission agencies.
- The existence of High End CI resources enabled nternational competitiveness of US researchers including expanding the expertise of the US workforce
  - Must provide the resources necessary for them to keep pace with their research collaborators.
- NSF must evaluate and develop funding methods/programs for advanced CI that recognizes the need for coherence and continuity for suppliers, providers and consumers of these capabilities in order to enhance the productivity of the scientific and engineering research endeavors.
- Technology paths forward for advanced computation are better defined than those for data-intensive science; therefore need more technology development efforts for resources supporting data-intensive science

Summary

Because NSF has the broader, national mission of advancing science and engineering and ensuring the scientific and economic competitiveness of the nation, it is vitally important NSF provide a series of HPCD Track 1 and Track 2 systems and services over time to meet the increasing demands of high spectrum digital science with increased funding and coherence

This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.