SAND2011-2015C

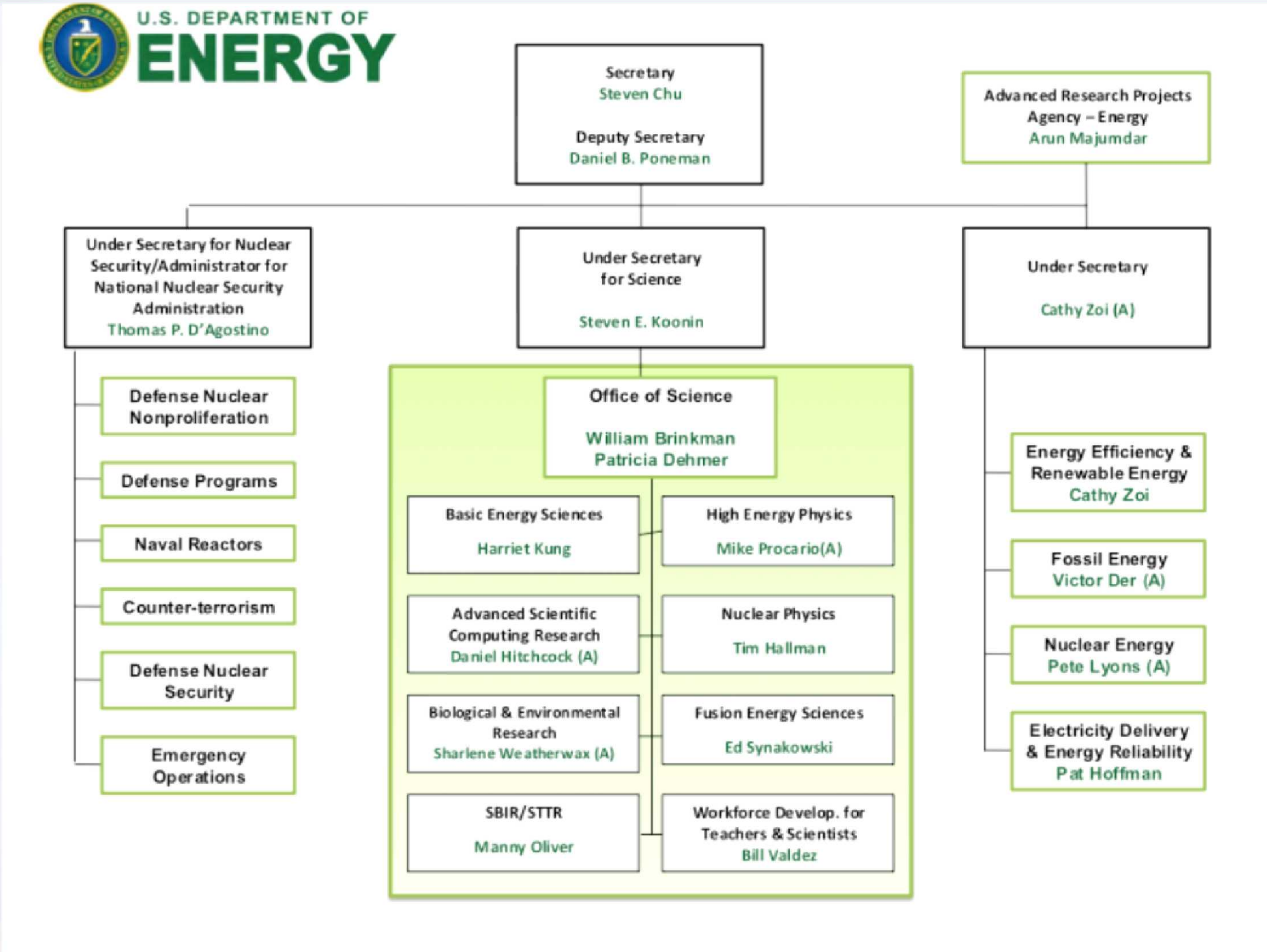
Bill Spotz, Sandia National Laboratories

(Formerly on assignment to US DOE Office of Science)

Monday, March 28, 2011

EXASCALE CO-DESIGN AND CLIMATE MODELING

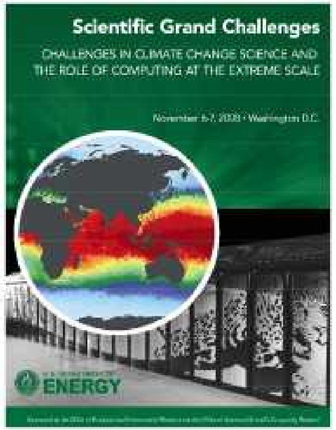
Sandia National Laboratories


U.S. DEPARTMENT OF
ENERGY
Secretary
Steven Chu
Deputy Secretary
Daniel B. Poneman
Advanced Research Projects Agency – Energy
Arun Majumdar
Under Secretary for Nuclear Security/Administrator for National Nuclear Security Administration
Thomas P. D'Agostino
Under Secretary for Science
Steven E. Koonin
Under Secretary
Cathy Zoi (A)
Defense Nuclear Nonproliferation
Defense Programs
Naval Reactors
Counter-terrorism
Defense Nuclear Security
Emergency Operations
Office of Science
William Brinkman
Patricia Dehmer
Basic Energy Sciences
Harriet Kung
High Energy Physics
Mike Procario(A)
Advanced Scientific Computing Research
Daniel Hitchcock (A)
Nuclear Physics
Tim Hallman
Biological & Environmental Research
Sharlene Weatherwax (A)
Fusion Energy Sciences
Ed Synakowski
SBIR/STTR
Manny Oliver
Workforce Develop. for Teachers & Scientists
Bill Valdez
Energy Efficiency & Renewable Energy
Cathy Zoi
Fossil Energy
Victor Der (A)
Nuclear Energy
Pete Lyons (A)
Electricity Delivery & Energy Reliability
Pat Hoffman

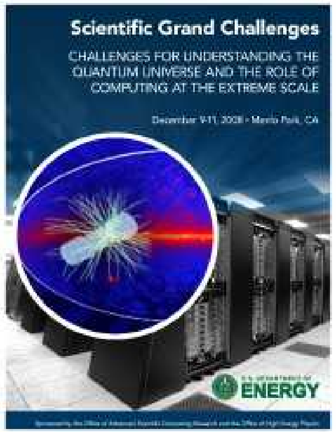
Exascale Issues

- **Clock speed ceiling requires increased concurrency for improved performance (billion-way)**
- **Floating point operations use less power than data movement on and off chip:**
  - Remap multiphysics to put as much work per location on same chip
  - Include embedded UQ to increase concurrency
  - Reformulate to trade flops for memory use
  - Weak scaling approaches (constant memory/flop) probably will not work
- **I/O to disk will be slower relative to FLOPS and in-core data movement:**
  - Part of the file system may be on the node
  - Include in-line data analysis if you can for more concurrency
  - Trigger output to only move important data off machine
- **There will be silent errors as well as lower mean time to interrupt**
  - Uncertainty Quantification including hardware variability
  - New approaches to check-pointing and resilience
- **Current commercial roadmap targets memory energy per bit about 4× too high (assuming 20MW machine)**

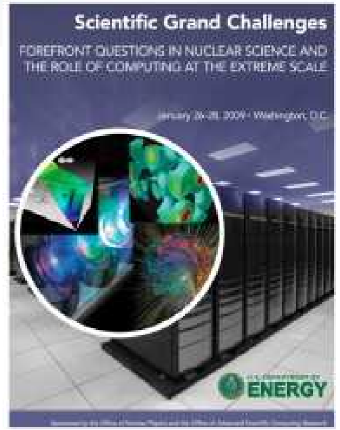
DOE ASCR Extreme Scale Computing Workshops



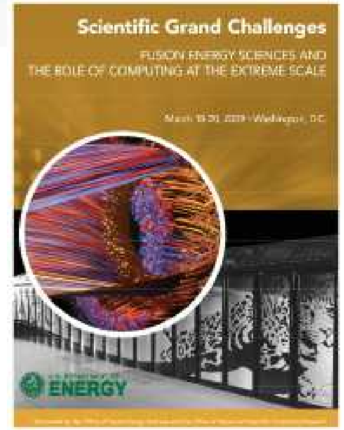
Climate
11/08



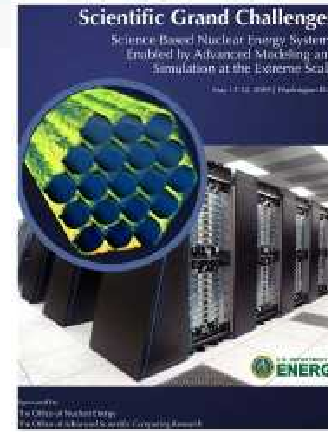
High Energy Physics
12/08



Nuclear Physics
1/09

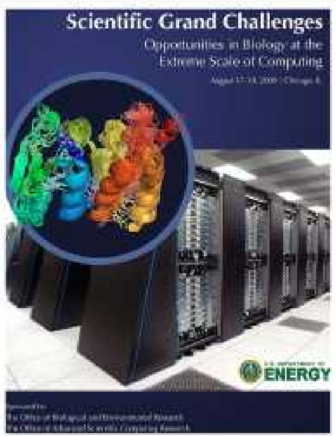


Fusion Energy
3/09

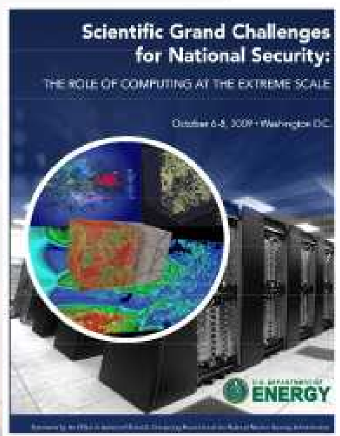


Nuclear Energy
5/09

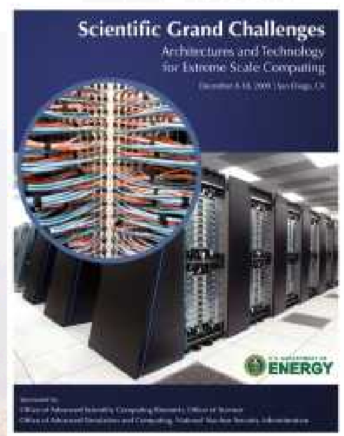
Basic Energy
8/09



Biology
8/09



National Security
10/09



Architecture
12/09

Crosscut
2/10

Sandia National Laboratories

Climate Workshop Summary

- **Model Development and Integrated Assessment**
  - How will the sea level, sea-ice coverage and ocean circulation change as the climate changes?
  - How will the distribution and cycling of water, ice, and clouds change with global warming?
  - How will extreme weather and climate change on the local and regional scales?
  - How do the carbon, methane, and nitrogen cycles interact with climate change?
- **Algorithms and Computational Environment**
  - Develop scalable algorithms for non-hydrostatic atmospheric dynamics with quasi-uniform grids, implicit formulations, and adaptive multiscale and multiphysics coupling
  - Foster international consortia for parallel input/output, metadata, analysis and modeling tools for regional and decadal multimodel ensembles
  - Develop multicore and deep memory languages to support parallel software development
- **Decadal Predictability and Prediction**
  - Identify sources and mechanisms for decadal predictability
  - Develop strategies for tapping into this predictability and ultimately benefiting society
- **Data, Visualization, and Computing**
  - Develop new, robust techniques for dealing with the input/output, storage, processing, and wide-area transport demands of extreme scale data
  - Integrate diverse and complex data
  - Dedicate resources to the development of standards, conventions, and policies

Potential System Architectures

Systems	2009	2015	2018
System peak	2 Peta	100-300 Peta	1 Exa
Power	6 MW	~15 MW	~20 MW
System memory	0.3 PB	5 PB	64 PB (+)
Node performance	125 GF	0.5 TF or 7 TF	1-2 or 10TF
Node memory BW	25 GB/s	1-2TB/s	2-4TB/s
Node concurrency	12	O(100)	O(1k) or 10k
Total Node Interconnect BW	3.5 GB/s	100-200 GB/s	200-400GB/s
System size (nodes)	18,700	50,000 or 500,000	O(100,000) or O(1M)
Total concurrency	225,000	O(100,000,000) *O(10-50) to hide latency	O(billion) * O(10-100) to hide latency
Storage	15 PB	150 PB	500-1000 PB (>10x system memory is min)
I/O	0.2 TB/s	10 TB/s	60 TB/s
MTTI	days	O(1 day)	O(1 day)

Exascale Co-design

- “Co-design” first mentioned in fusion energy workshop report (3/09); important focus of crosscut workshop report (2/10)
- Concept emerged from embedded systems
- Need to allocate complexity between hardware, systems software, libraries, and applications
- Potential for applications to modify designs at all levels
- Need to consider reformulating as well as re-implementing
- Need to include uncertainty quantification, in line data analysis, and resilience in applications
- Co-adapt applications to new programming models and perhaps languages
- Impact of massive multithreaded nodes and new ultra-lightweight operating systems.

ASCR Exascale Co-design FOA

- **In June, 2010, ASCR released a Funding Opportunity Announcement requesting proposals for multi-institution Exascale Co-design Centers, focused around a scientific application**
- **In July, 2010, proposals for 21 centers were submitted**
- **Five proposals received program development money:**
  - Materials in extreme environments
  - Chemistry for energy conversion and storage
  - Fusion energy
  - High-energy density physics
  - Advanced reactors
- **Two proposals encouraged to reformulate:**
  - Climate
  - Combustion

ASCR Exascale Research Kickoff Meeting

- **Held 3/7/11 in San Diego, CA**
- **Participants:**
  - Advanced Architecture, X-Stack, Scientific Data Management and Analysis at Extreme Scales awardees
  - X-Stack planning team: An Open Integrated Software Stack for Extreme Scale Computing
  - Seven co-design proposal teams
- **Large amounts of data collected:**
  - Coverage assessment
  - Nature and likelihood of collaborations
  - Crosscut topics, including concerns
  - Gap analysis
- **Five crosscutting topics were identified based on researcher input:**
  1. Programming models
  2. Resilience and fault tolerance
  3. Data management & I/O architecture
  4. Memory / data movement and energy efficiency
  5. Application architecture
- **Data still being analyzed…**

Concluding Remarks

- **ASCR anticipated exascale challenges several years ago**
- **ASCR has been engaging the community to develop a path forward to achieve exascale by ~2018**
- **ASCR and NNSA will partner to achieve DOE exascale goals**
- **X-Stack computer science projects have already been funded**
- **Co-design has been identified as an important and necessary approach to exascale computing**
- **Exascale co-design centers:**
  - Five proposals granted program development money
  - Two proposals encouraged to reformulate
  - More were initially envisioned…
  - Path forward is continually re-examined