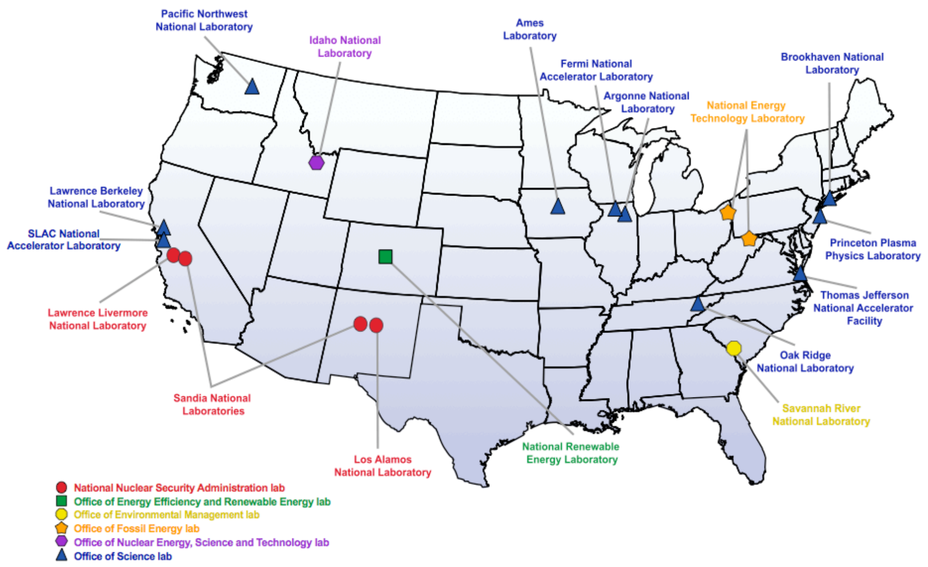
HPC Computing and Trends toward Exascale Computing in the USA

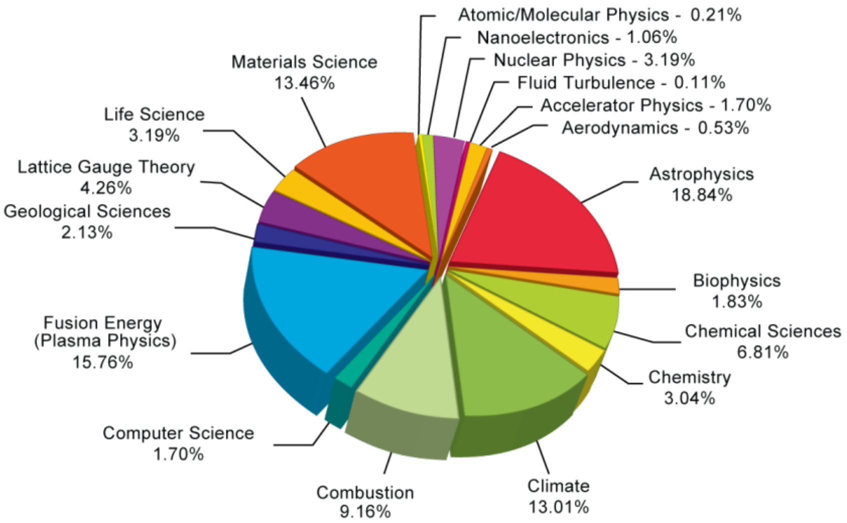
Pete Beckman
Director of the Argonne Leadership Computing Facility

DEPARTMENT OF ENERGY NATIONAL LABORATORIES



Previous INCITE Awards

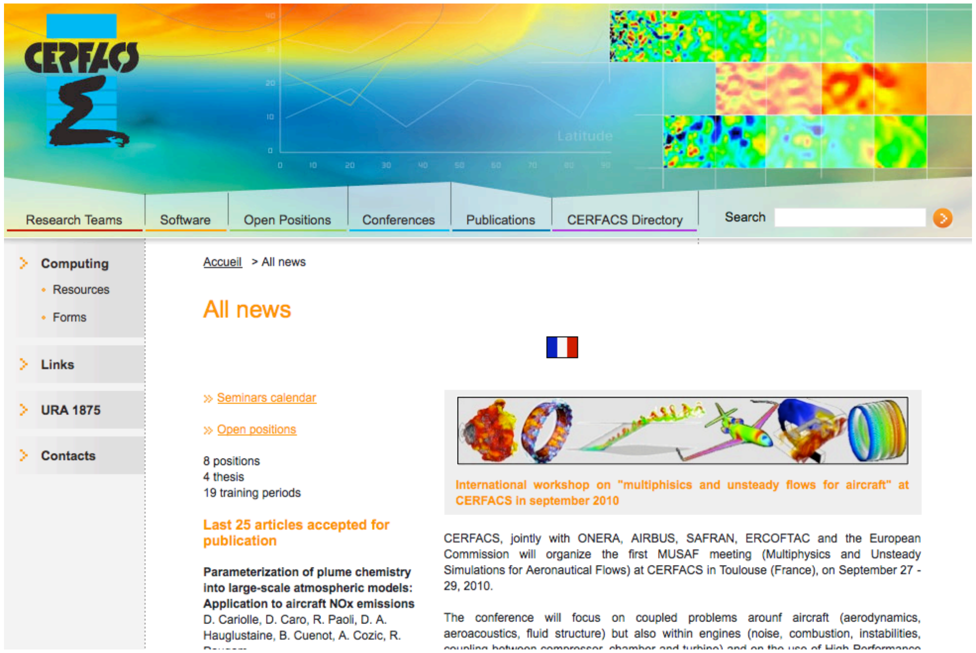
In **2009** nearly 900 million processor-hours were awarded to 25 new projects and 41 renewal projects



Argonne's IBM Blue Gene/P – 556 TFs

Partnerships:

IBM
LLNL
ANL



All news

» Seminars calendar

» Open positions

8 positions
4 thesis
19 training periods

Last 25 articles accepted for publication

Parameterization of plume chemistry into large-scale atmospheric models: Application to aircraft NOx emissions
D. Cariolle, D. Caro, R. Paoli, D. A. Hauglustaine, B. Cuenot, A. Cozic, R.

International workshop on "multiphysics and unsteady flows for aircraft" at CERFACS in september 2010

CERFACS, jointly with ONERA, AIRBUS, SAFRAN, ERCOFTAC and the European Commission will organize the first MUSAF meeting (Multiphysics and Unsteady Simulations for Aeronautical Flows) at CERFACS in Toulouse (France), on September 27 - 29, 2010.

The conference will focus on coupled problems arounf aircraft (aerodynamics, aeroacoustics, fluid structure) but also within engines (noise, combustion, instabilities,

---

Insight into Parkinson's Disease

INCITE

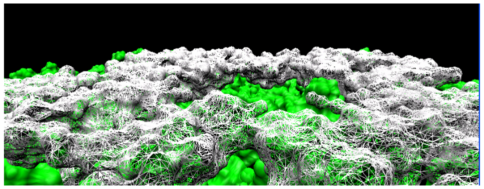
Igor Tsingelny
University of California, San Diego

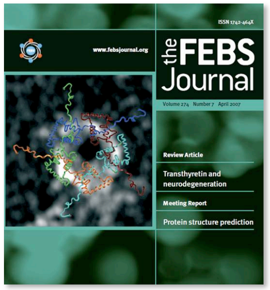
Science

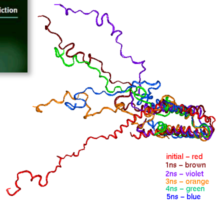
- Parkinson's Disease affects 5 million in US and Europe
- Increased aggregation of α-*synuclein* protein is thought to lead to harmful pore-like structures in human membranes
- Simulations show α-synuclein complexes, and β-synuclein prevents creation of propagating α-synuclein complexes

Methods and Challenges

- Using NAMD and MAPAS on Blue Gene at ALCF and SDSC
- α-syn proteins are unstructured







α-synuclein coformer at membrane (left), completed pentamer (above), evolution of α–synuclein over 5ns (right)

Better Reactors Faster at Scale

Paul Fischer ANL

Pressure distribution of coolant flow

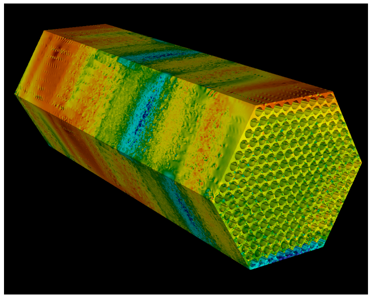
Breakthrough computation of thermal-hydraulics with Nek5000

2.95M spectral elements

1 billion grid points

- Full physical configuration of 217 wire-wrapped fuel pins
  - *Production runs on 32K cores*
  - *80% parallel efficiency on 124K cores (strong scaling)*
  - *Culmination of 3 year INCITE with results from lower pin counts*
- Historical Build
  - 7-pins: Strong correlation with LES and RANS providing path for saving compute resources in the future
  - 19-pins: Proof that time-saving boundary conditions can be used

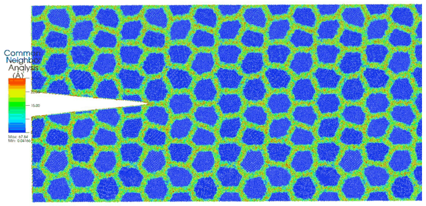
- Innovations over 3 Years
  - Scalable spectral element multigrid solver for the pressure
  - 4th generation coarse-grid solver (algebraic multigrid)
  - Elimination of all arrays scaling with global element count
  - Communication algorithms to discover processor topology
  - Scalable grid partitioner
  - Parallel I/O rewrite (subcommunicator)
  - Parallel visualization



Designing Better Materials for Nuclear Reactors

Priya Vashista
University of Southern California

- **47 million-atom ReaxFF MD simulations exploring fracture modes in materials**
  - 50M CPU-Hours
- **Key to design of next-generation nuclear reactor**
  - Revealed a missing link between sulfur-induced intergranular amorphization and embrittlement

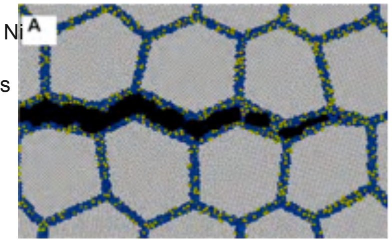


With Sulfur

Color: common-neighbor parameter that characterizes atomistic defects

> *".. the proposed upgrade of the ALCF resources to a 10-20 Petaflops Blue Gene Q system would enable our research group to make major breakthrough simulations on a number of applications of high relevance to DOE. We look forward to working with the ALCF on their new platform. "*

(A) close-up fracture simulation in nanocrystalline Ni with amorphous sulfide GB layers

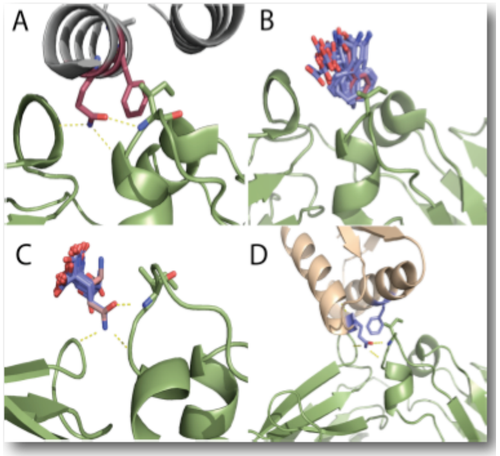


INCITE

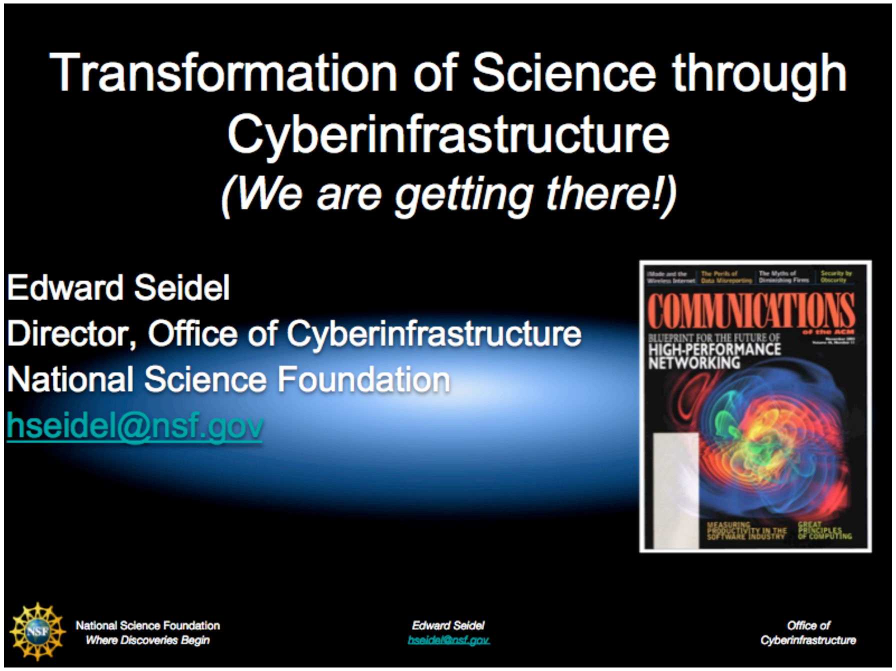
Computational Protein Structure Prediction and Protein Design

David Baker
University of Washington

- Computationally design protein-based inhibitors towards pathogens like H1N1
- Rapid turn around of huge campaigns on ALCF reinvented how the science is done and enables new research
- Rapidly determine an accurate, high-resolution structure of any protein sequence up to 150-200 residues
- Incorporating sparse experimental NMR data into Rosetta to allow larger proteins



The interfaces of protein-protein complexes often exhibit a handful of key interactions, termed hot-spots. At right, the original protein (A) is replaced by an easy-to-manufacture custom scaffold (D)



Transformation of Science through Cyberinfrastructure
(We are getting there!)

Edward Seidel
Director, Office of Cyberinfrastructure
National Science Foundation
hseidel@nsf.gov

NSF TeraGrid Shared Environments

Computers

Data services

Visualization services

People

Modeling and simulation

Data analysis & visualization

User support

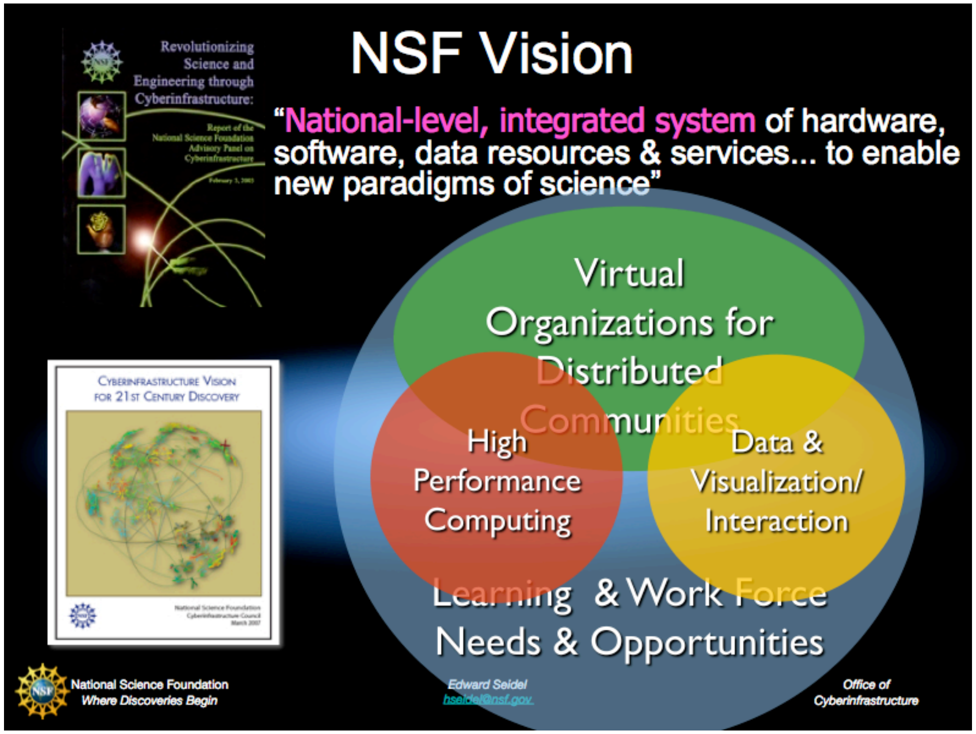
Training

Common user environments

Tools for educators

Science Gateways

Courtesy Ed Seidel, NSF



NSF Vision

Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Advisory Panel on Cyberinfrastructure February 3, 2003

“National-level, integrated system of hardware, software, data resources & services... to enable new paradigms of science”

Cyberinfrastructure Vision for 21st Century Discovery
National Science Foundation Cyberinfrastructure Council March 2007

Virtual Organizations for Distributed Communities
High Performance Computing
Data & Visualization/ Interaction
Learning & Work Force Needs & Opportunities

National Science Foundation *Where Discoveries Begin*
Edward Seidel hseidel@nsf.gov
Office of Cyberinfrastructure

Data-Driven Multiscale Collaborations* for Complexity

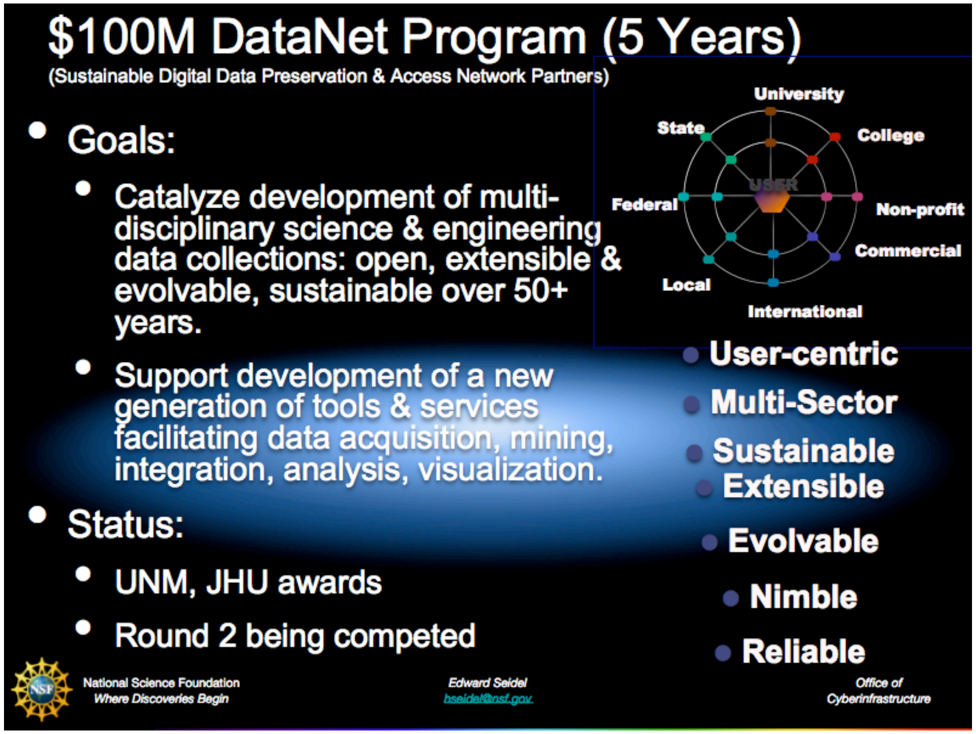
Great Challenges of 21st Century

- Multiscale Collaborations
  - General Relativity, Particles, Geosciences, Bio, Social...
  - And all combinations...
- Science and Society being transformed by CI and Data
  - Completely new methodologies
  - “The End of Science” (as we know it)
- CI plays central role
  - No community can attack challenges
  - Technical, CS, social issues to solve
- *Places requirements on computing, software, networks, tools, etc*

The End of Science
The quest for knowledge used to begin with grand theories. Now it begins with massive amounts of data. Welcome to the Petabyte Age.

*Small groups still important!

National Science Foundation *Where Discoveries Begin*
Edward Seidel hseidel@nsf.gov
Office of Cyberinfrastructure



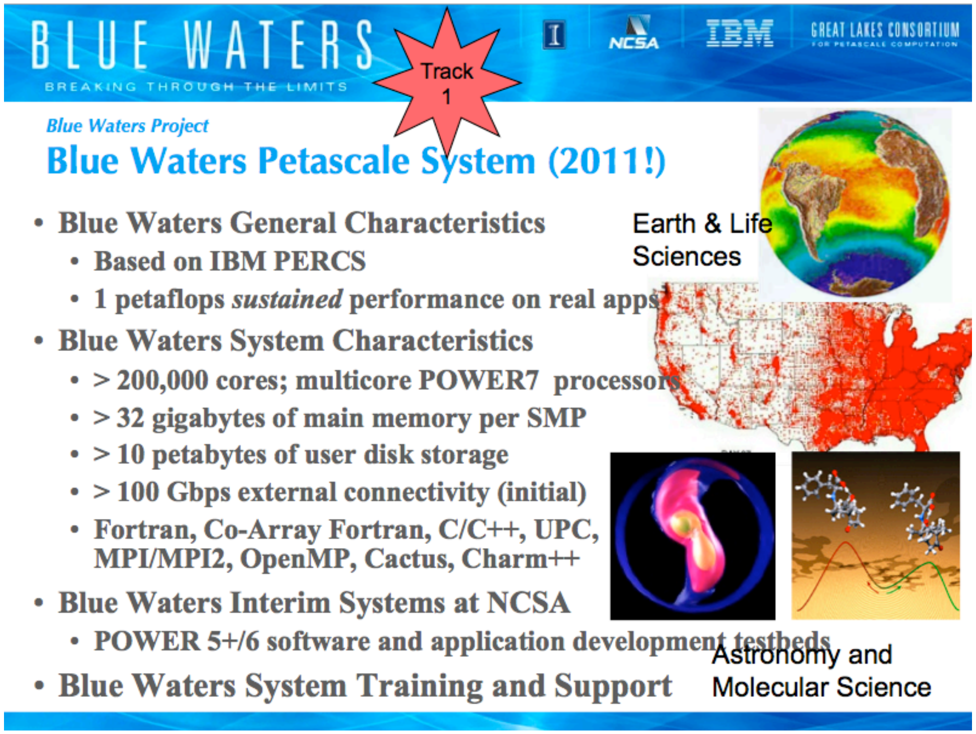
$100M DataNet Program (5 Years)

(Sustainable Digital Data Preservation & Access Network Partners)

- Goals:
  - Catalyze development of multi-disciplinary science & engineering data collections: open, extensible & evolvable, sustainable over 50+ years.
  - Support development of a new generation of tools & services facilitating data acquisition, mining, integration, analysis, visualization.
- Status:
  - UNM, JHU awards
  - Round 2 being competed

- User-centric
- Multi-Sector
- Sustainable
- Extensible
- Evolvable
- Nimble
- Reliable

National Science Foundation *Where Discoveries Begin*
Edward Seidel hseidel@nsf.gov
Office of Cyberinfrastructure



BLUE WATERS
BREAKING THROUGH THE LIMITS

Track 1

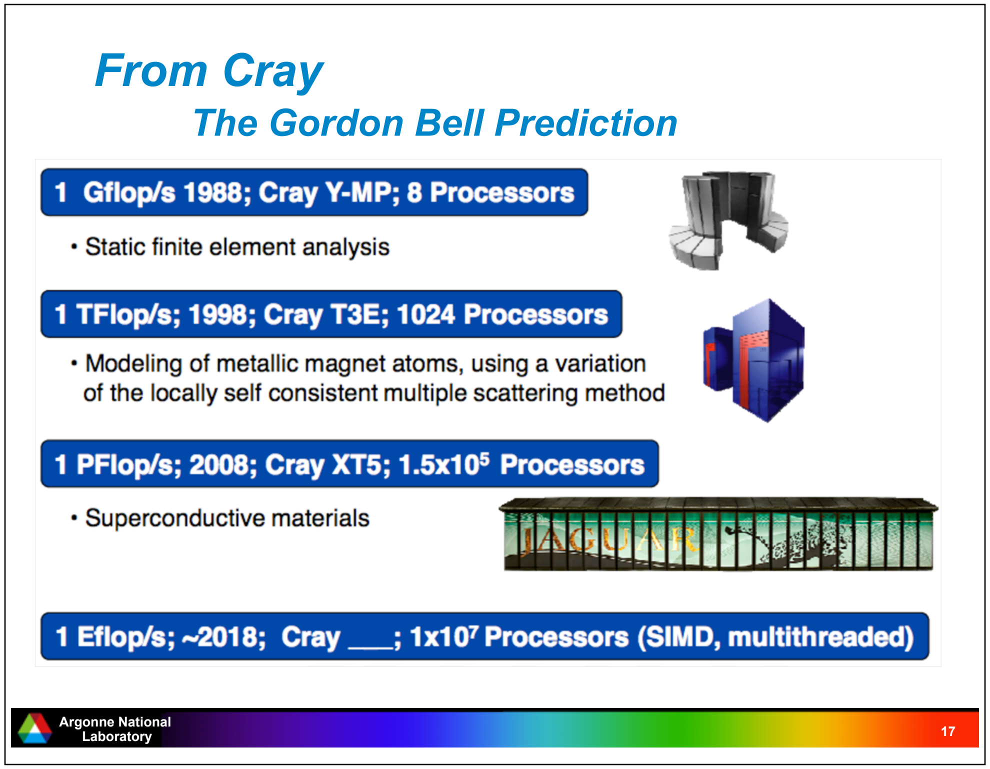
Blue Waters Project

Blue Waters Petascale System (2011!)

- Blue Waters General Characteristics
  - Based on IBM PERCS
  - 1 petaflops *sustained* performance on real apps
- Blue Waters System Characteristics
  - > 200,000 cores; multicore POWER7 processors
  - > 32 gigabytes of main memory per SMP
  - > 10 petabytes of user disk storage
  - > 100 Gbps external connectivity (initial)
  - Fortran, Co-Array Fortran, C/C++, UPC, MPI/MPI2, OpenMP, Cactus, Charm++
- Blue Waters Interim Systems at NCSA
  - POWER 5+/6 software and application development testbeds
- Blue Waters System Training and Support

Earth & Life Sciences

Astronomy and Molecular Science



From Cray

The Gordon Bell Prediction

1 Gflop/s 1988; Cray Y-MP; 8 Processors

- Static finite element analysis

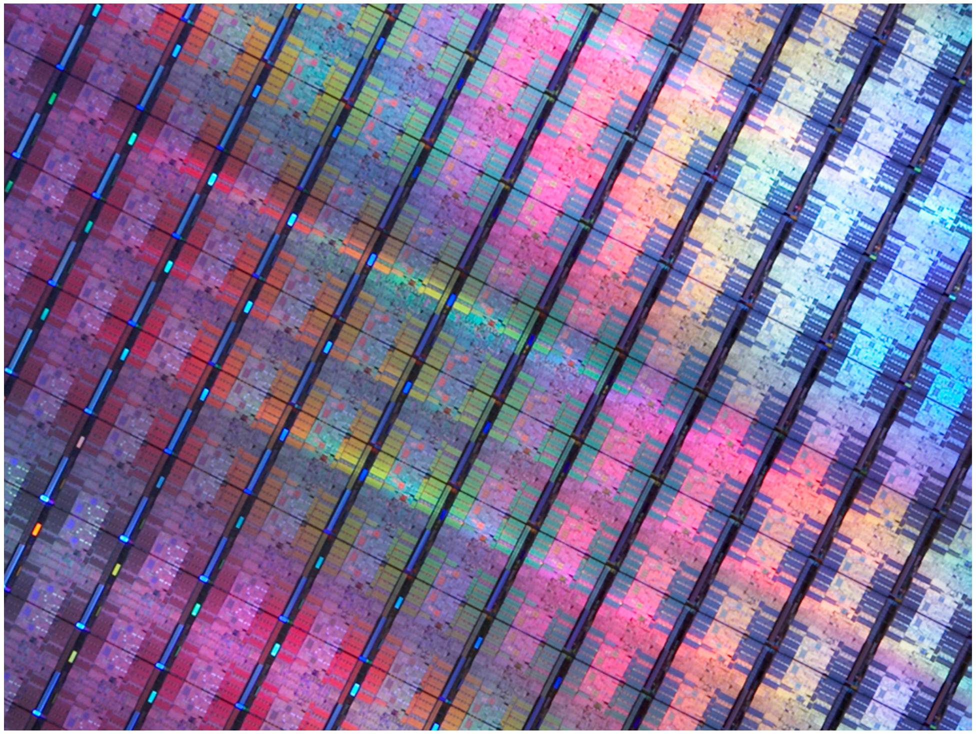
1 TFlop/s; 1998; Cray T3E; 1024 Processors

- Modeling of metallic magnet atoms, using a variation of the locally self consistent multiple scattering method

1 PFlop/s; 2008; Cray XT5; 1.5x10^5 Processors

- Superconductive materials

1 Eflop/s; ~2018; Cray ___; 1x10^7 Processors (SIMD, multithreaded)



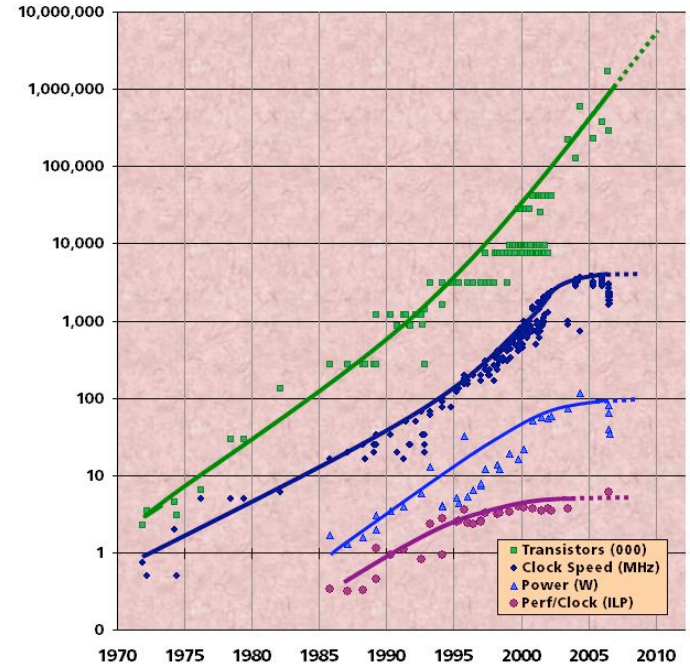
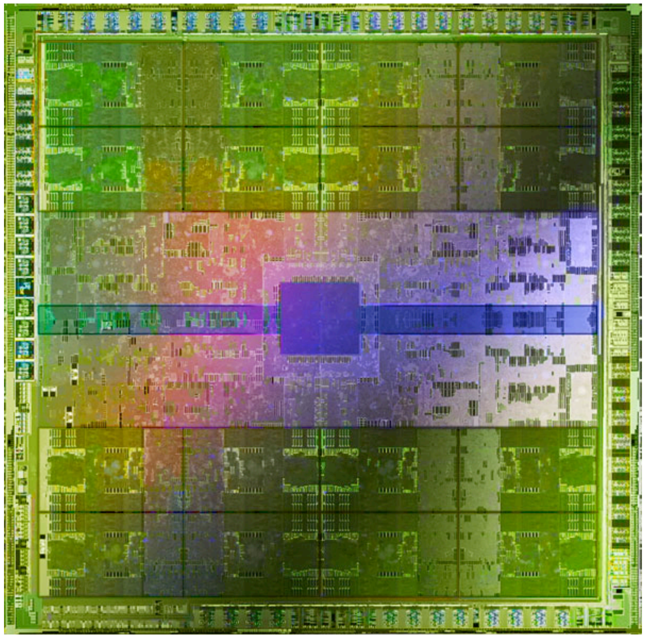


Figure courtesy of Kunle Olukotun, Lance Hammond, Herb Sutter, and Burton Smith

Nvidia Fermi

- 512 Cores
- IEEE FP
- ECC on Memory
- 64 bit address space
- Caches



"It took a decade to be able to efficiently utilize a 10X increase in processor parallelism, to expect that 1000X can be handled in less than that is a long stretch"

A Thought Experiment

lim

cores → ∞

power → ∞

mem/core → 0

C, C++, FORTRAN

OPENCL

OPENMP

?

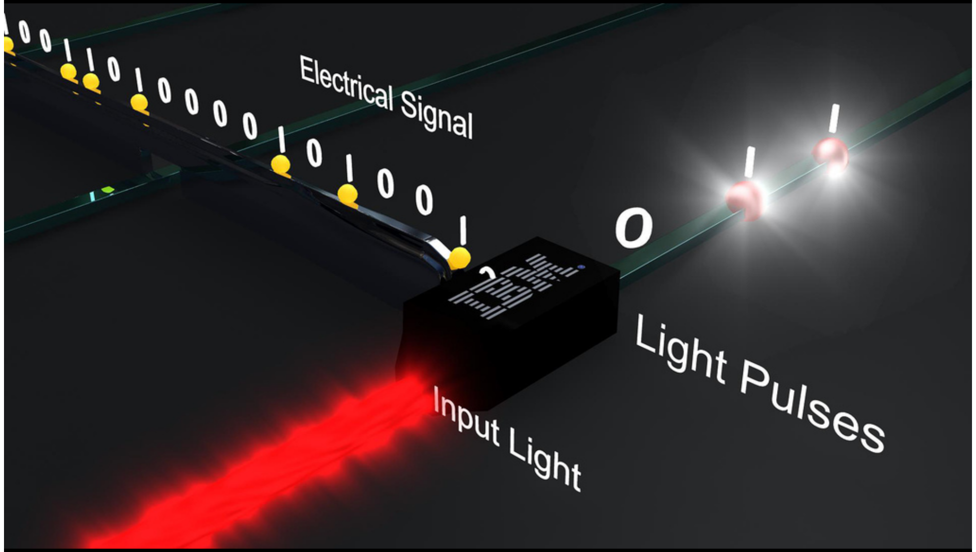
CUDA

MPI, UPC

Big Research Problem

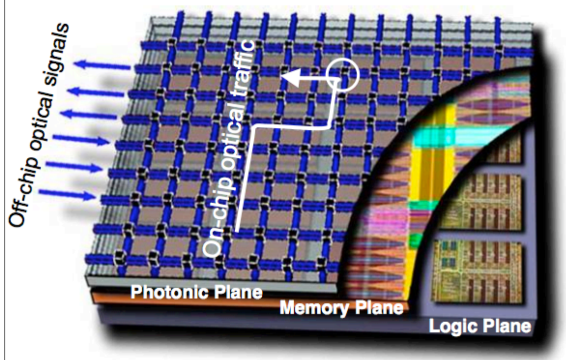
PACKAGING

Silicon Nanophotonics



100X Faster, 10X Less Power

Vision for 22nm CMOS (circa 2018) - 10 TFLOPs on a 3D chip



36 "Cell" chip (~300 cores)

System level study:
IBM, Columbia, Cornell, UCSB

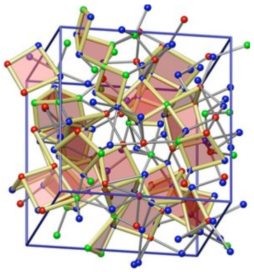
Co-PIs:
Jeff Kash (IBM)
Keren Bergman (Columbia)
Yurii Vlasov (IBM)

Logic plane	~300 cores
Memory plane	~30GB eDRAM
Photonic plane	On-Chip Optical Network >70Tbps optical on-chip >70Tbps optical off-chip

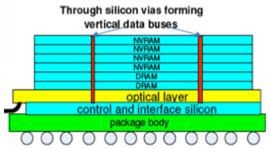
Photonic layer is not only connecting various cores, but also routes the traffic

All future dates and specifications are estimations only. Subject to change without notice.

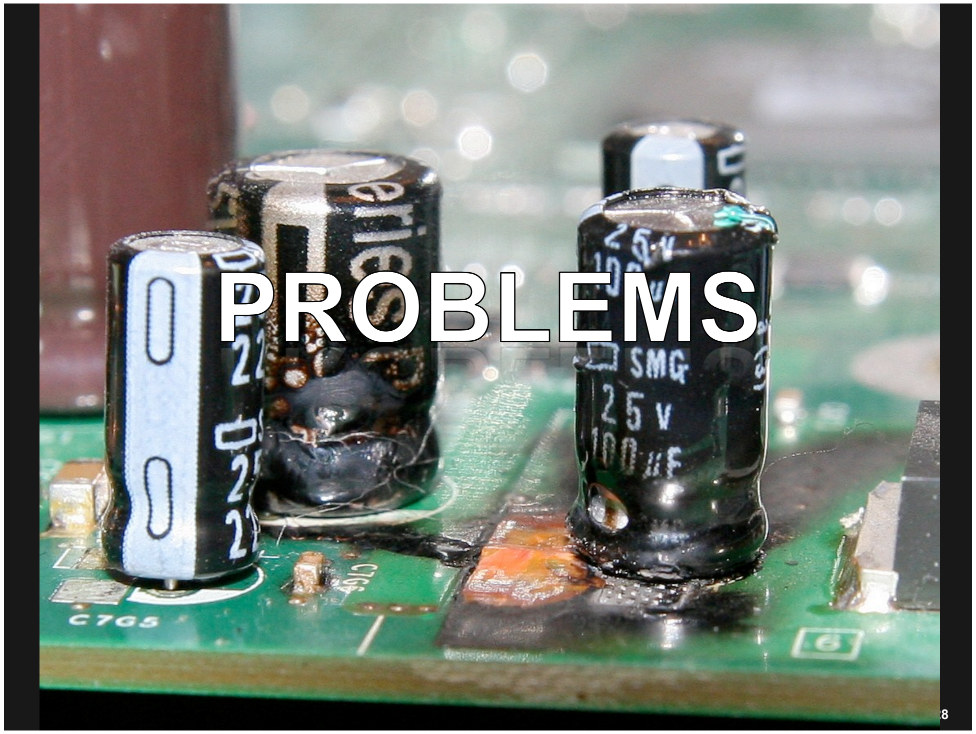
Change the Packaging
Change the Paradigm



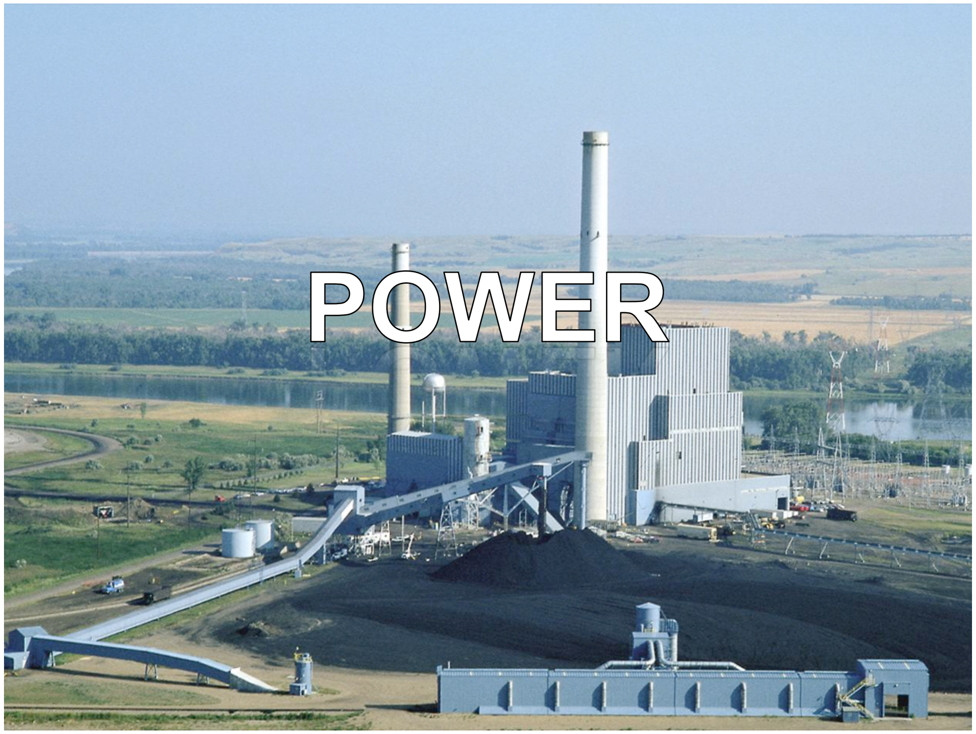
Handy, Close, Near, Far



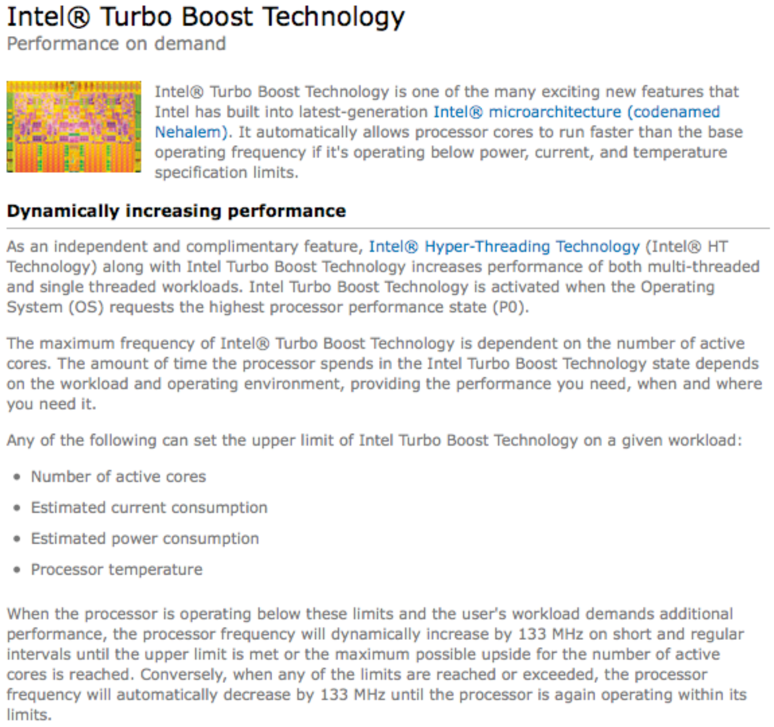
Big Research Problem



PROBLEMS



POWER



Intel® Turbo Boost Technology

Performance on demand

Intel® Turbo Boost Technology is one of the many exciting new features that Intel has built into latest-generation Intel® microarchitecture (codenamed Nehalem). It automatically allows processor cores to run faster than the base operating frequency if it's operating below power, current, and temperature specification limits.

Dynamically increasing performance

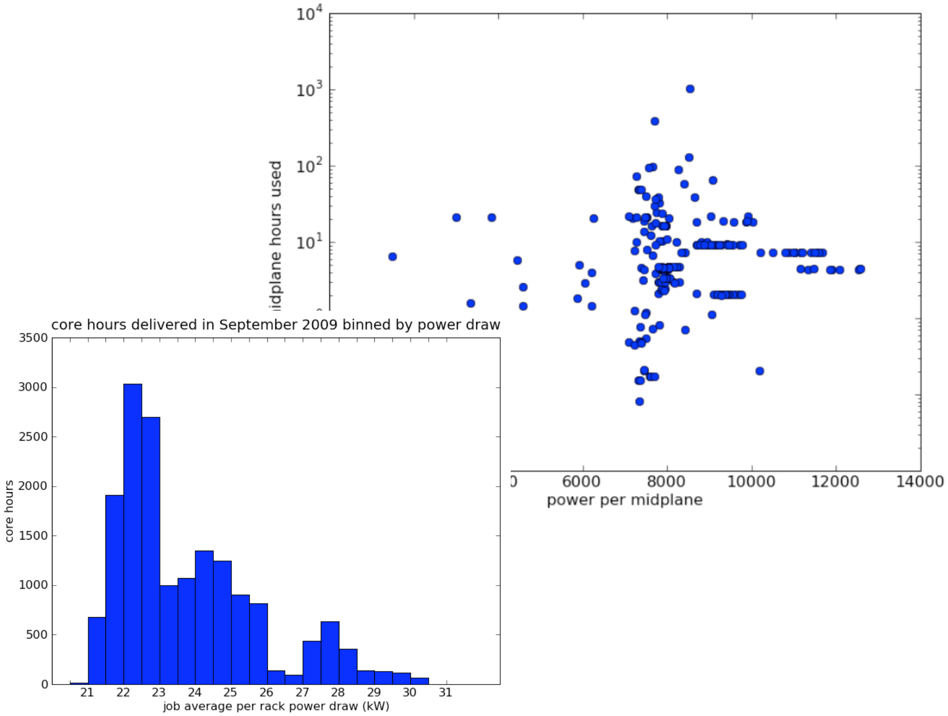
As an independent and complimentary feature, Intel® Hyper-Threading Technology (Intel® HT Technology) along with Intel Turbo Boost Technology increases performance of both multi-threaded and single threaded workloads. Intel Turbo Boost Technology is activated when the Operating System (OS) requests the highest processor performance state (P0).

The maximum frequency of Intel® Turbo Boost Technology is dependent on the number of active cores. The amount of time the processor spends in the Intel Turbo Boost Technology state depends on the workload and operating environment, providing the performance you need, when and where you need it.

Any of the following can set the upper limit of Intel Turbo Boost Technology on a given workload:

- Number of active cores
- Estimated current consumption
- Estimated power consumption
- Processor temperature

When the processor is operating below these limits and the user's workload demands additional performance, the processor frequency will dynamically increase by 133 MHz on short and regular intervals until the upper limit is met or the maximum possible upside for the number of active cores is reached. Conversely, when any of the limits are reached or exceeded, the processor frequency will automatically decrease by 133 MHz until the processor is again operating within its limits.



“At Exascale, where fetching a word from DRAM will cost more power than performing a double precision floating point multiply-accumulate operation with it, there will be a new imperative to minimize state size and eliminate unnecessary references to deep in the memory hierarchy”

GF/W must improve 1000x for Exascale
How will software evolve?

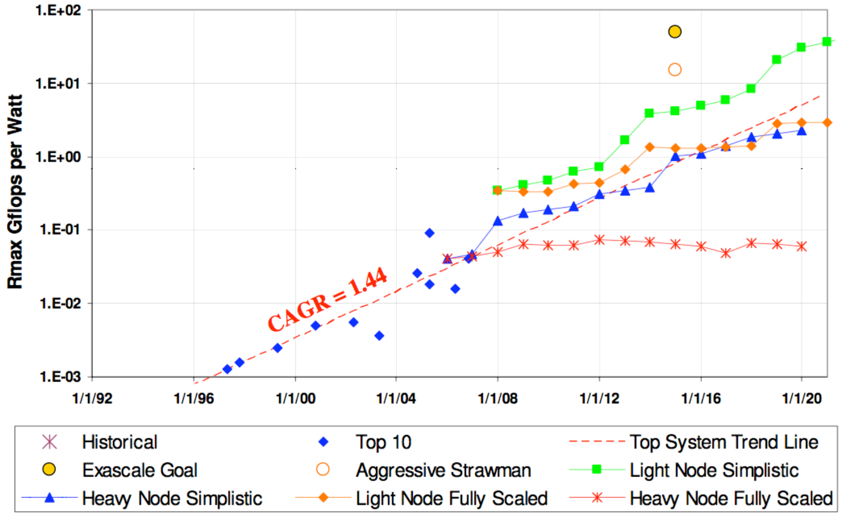


Figure 8.3: The power challenge for an Exaflops Linpack.

Source: DARPA Exascale Report

PARALLELISM

PACKAGING

PROBLEMS

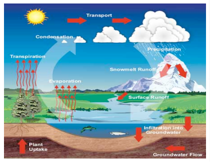
POWER

Simulations are critical for understanding climate change and predicting impacts

Ocean and ice
Improved modeling of land ice to reduce uncertainties in predictions for sea level rise and changes in ocean circulation

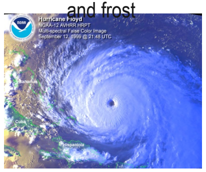
Greenland ice
NASA

Hydrological cycle Reduce uncertainties in prediction of cloud radiative effects and precipitation due to climate change



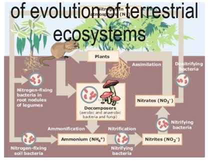
Hydrological cycle
NWS

Extreme weather
Use multi-scale ESMs to understand statistics of cyclones, blizzards, meso-scale storms, heat waves, droughts, and frost



Hurricane Floyd
NOAA

Carbon, methane, and nitrogen cycles
Use ESMs with improved chemical and biological process models to reduce uncertainties in predictions of evolution of terrestrial ecosystems

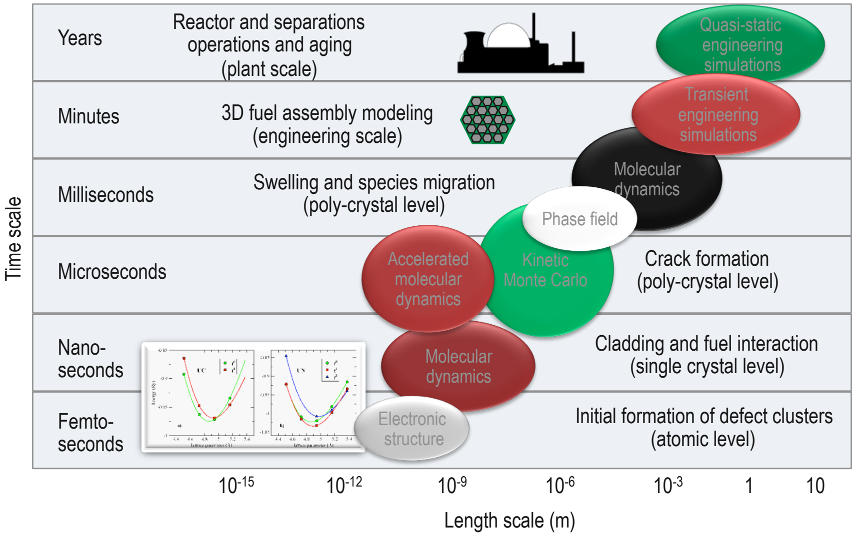


Nitrogen cycle
EPA

"Given these drivers . . . it is clear that exascale computers and ultra fast networks,
data systems and computational infrastructure will be required by 2020."

Challenges in Climate Change Science and the Role of Computing at Extreme Scale, November 2008

Predictive simulations are a critical capability for nuclear energy



Years — Reactor and separations operations and aging (plant scale)

Minutes — 3D fuel assembly modeling (engineering scale)

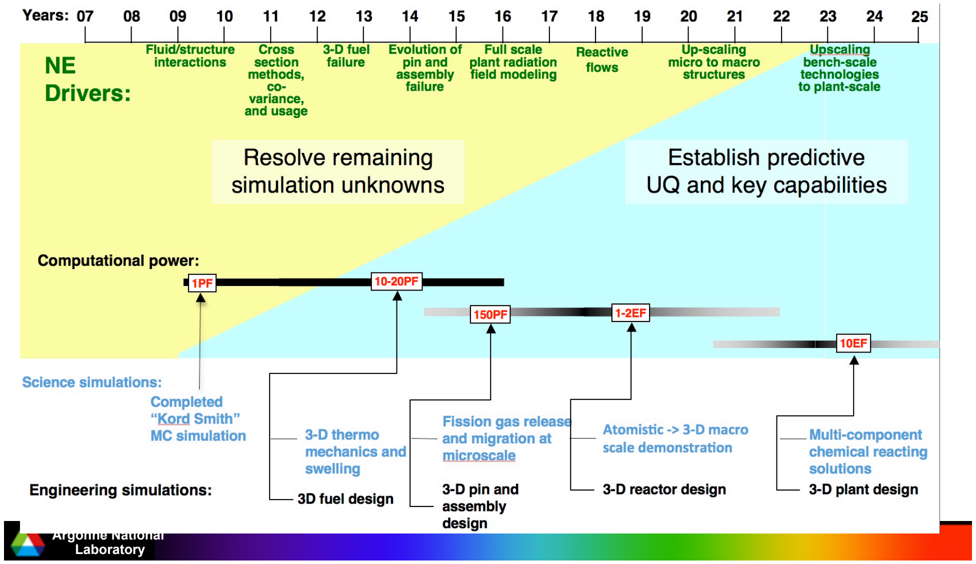
Milliseconds — Swelling and species migration (poly-crystal level)

Microseconds — Crack formation (poly-crystal level)

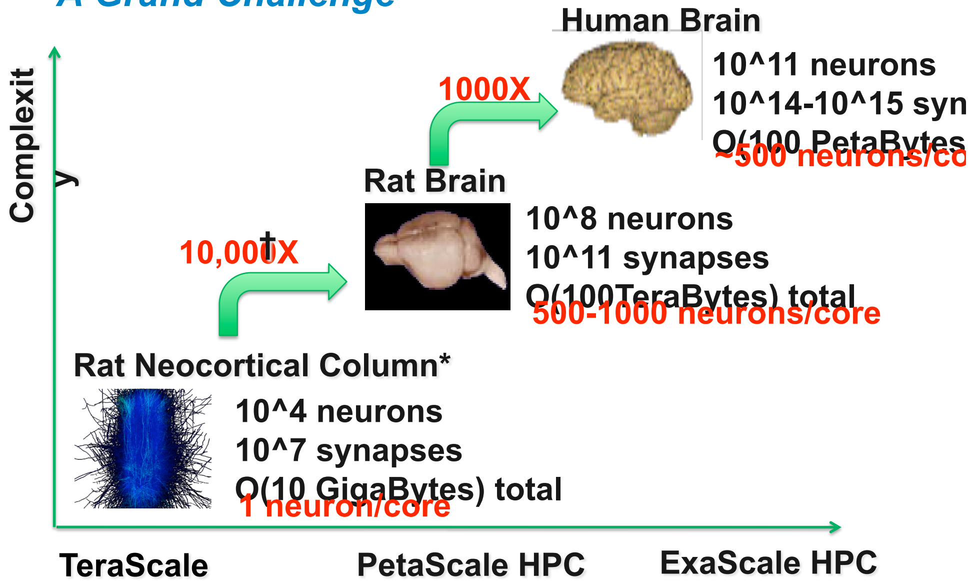
Nano-seconds — Cladding and fuel interaction (single crystal level)

Femto-seconds — Initial formation of defect clusters (atomic level)

Computational Requirements for Nuclear Energy Modeling



Case Study: Cellular Level Brain Model – A Grand Challenge



* 2007 [illegible] ject/EPFL, Switzerland using a 4-rack IBM BlueGene/L

†scaling factor includes additional scaling on Teras

INTERNATIONAL EXASCALE SOFTWARE PROJECT

JUNE 28-29, 2009 PARIS, FRANCE

Improving HPC Software

- Pete Beckman & Jack Dongarra

IESP Goal

Improve the world's simulation and modeling capability by improving the coordination and development of the HPC software environment

Workshops:

Build an international plan for developing the next generation open source software for scientific high-performance computing

Where We Are Today:

- ☐ SC08 (Austin TX) meeting to generate interest
- ☐ DOE's Office of Science funding
- ☐ Santa Fe meeting April 6-8, 2009
    - ☐ 60 people
- ☐ NSF's Office of Cyberinfrastructure funding
- ☐ European meeting June 28-29, 2009
    - ☐ 70 people
    - ☐ Draft Roadmap
    - ☐ Outline Report
- ☐ Asian meeting (Tsukuba Japan) October 18-20, 2009
    - ☐ Refine roadmap
    - ☐ Refine Report
- ☐ SC09 (Portland OR) BOF to inform others
    - ☐ Public Comment
    - ☐ Draft Report presented

Nov 2008

Apr 2009

Jun 2009

Oct 2009

Nov 2009

Building a Software Roadmap

- *Musings on the Path Toward Exascale*, Robert Lucas – ISI/USC
- *BSC Vision Towards Exascale*, Mateo Valero, BSC
- *Software Challenges of Extreme Scale Computing*, Michael Heroux – Sandia National Laboratory
- *Software and Exascale Computing*, Bill Camp – Intel Corporation
- *Application Analysis and Porting in the PRACE Proj* Michielse – Netherlands National Computing Facili (NCF)
- *The Application Perspective – Seeking Productivity Performance*, David Barkai – Intel Corporation
- *EDF white paper*, J.Y. Berthou and J.F. Hamelin – ED
- *The Biggest Need: A New Model of Computation*, T – Louisiana State University
- *NSF IESP Whitepaper*, Abani Patra, Rob Pennington Office of Cyberinfrastructure, National Science Fou
- *A Proposal for a Capability Centers Consortium*, Bi Snir – NCSA and the University of Illinois at Urbana
- *Slouching Towards Exascale*, Rusty Lusk, Argonne Laboratory
- *A Collaboration and Commercialization Model for Software Research*, Mark Seager and Brent Gorda, Livermore National Laboratory
- *The Case for A Hierarchal System Model for Linux* Seager and Brent Gorda, Lawrence Livermore Natio
- *IESP Whitepaper: PDE-based applications and solv scale*, DavidKeyes, Columbia University & SciDAC T
- *Developing a high performance computing/numer roadmap*, Ann Trefethen, Nick Higham, Ian Duff, a Coveney

Whitpapers and notes on crosscutting issues for Paris Meeting

- *Performance at Exascale*, Bernd Mohr (Jülich Supercomputing Centre) and Matthias S. Mueller (Wolfgang E. Nagel Center for Information Services and HPC)
- *Resource Management*, Barney McCabe (ORNL) and Hugo Falter (ParTec)
- *Programmability Issues*, Vivek Sarkar (Rice U.), Jesus Labarta (UPC), Mitsuhisa Sato (U. of Tsukuba), Barbara Chapman (U. of Houston)
- *Models of Computation – Enabling Exascale*, Thomas Sterling, Louisiana State University.
- *Major Computer Science Challenges at Exascale*, Al Geist (ORNL) and Robert Lucas (ISI)
- *Towards Exascale File I/0*, Yutaka Ishikawa, University of Tokyo
- *Co-design of Architectures and Algorithms*, Al Geist (ORNL) and Sudip Dosanjh (SNL)
- *IESP Exascale Challenge: Resilience and Fault Tolerance*, Al Geist (ORNL) and Franck Cappello (INRIA)

QUESTIONS?