UT-BATTELLE

ORNL Leadership Computing Facility

Presented to

Advanced Scientific Computing Advisory Committee

Thomas Zacharia

Associate Laboratory Director

Oak Ridge National Laboratory

Washington D. C.

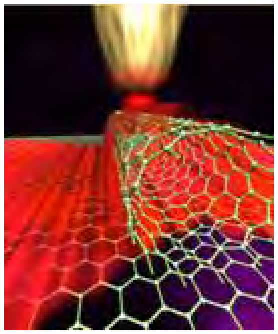
August 8-9, 2006

Leadership Computing Mission and vision

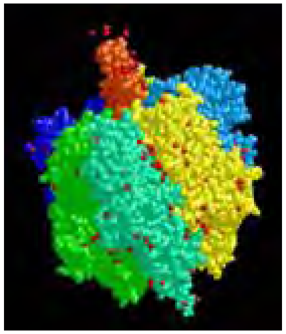
Focus on computationally intensive projects of large scale and high scientific impact through competitive peer review process

Provide the capability computing resources needed to solve problems of strategic importance to the nation.

1 Petaflop/s Cray Supercomputer

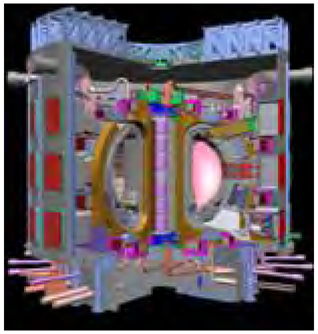


Design of innovative nanomaterials



Understanding of microbial molecular and cellular systems

100 yr Global climate to support policy decisions

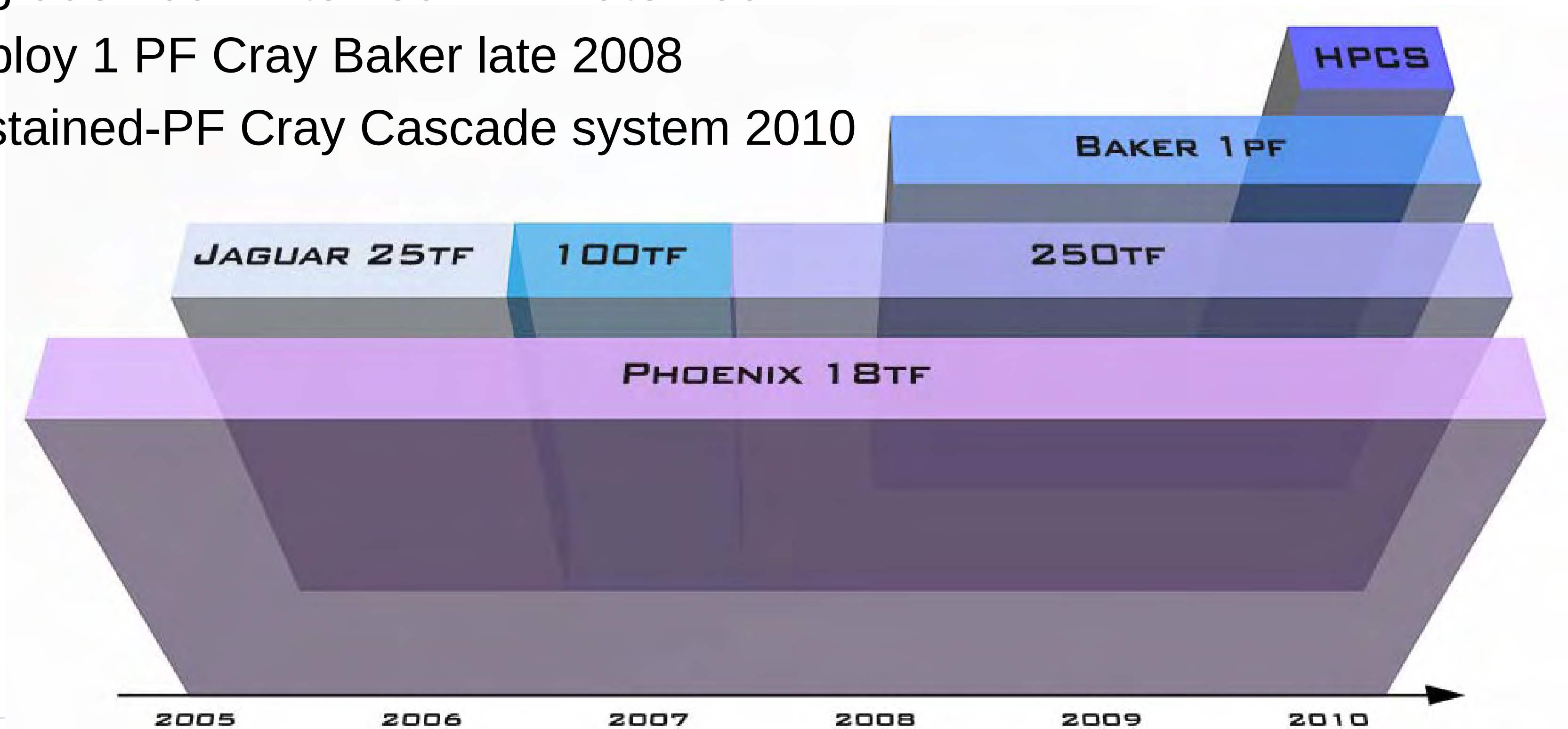


Predictive simulations of fusion devices

LCF project milestones: Deliver 1PF system in 2008 Deliver 250 TF by 2007

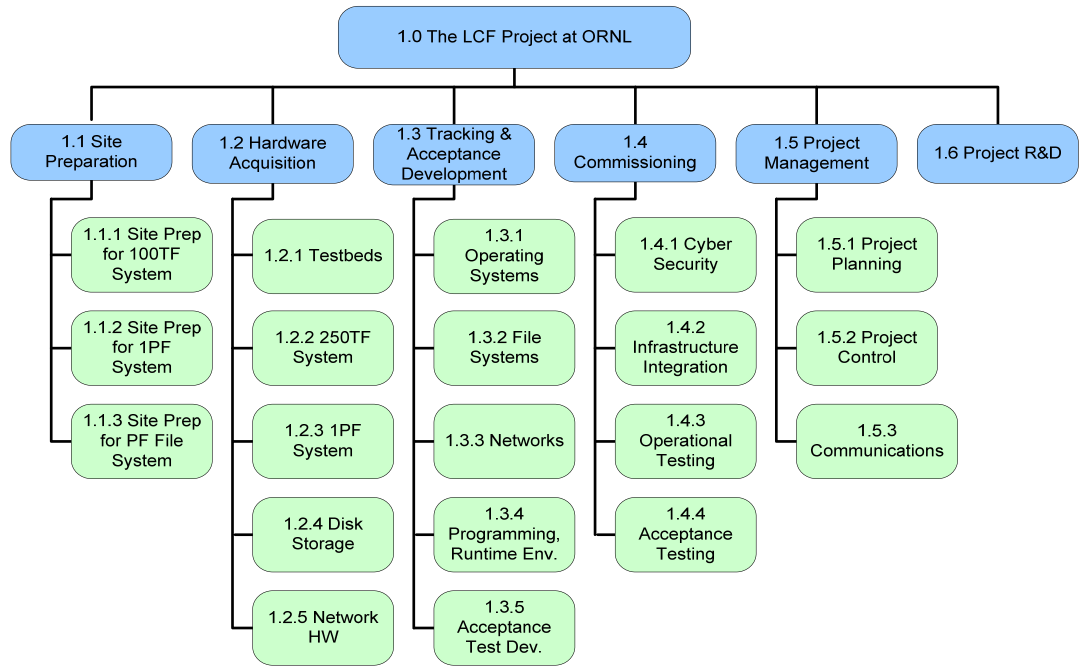
Roadmap

- Upgrade existing 50 TF XT3 to dual-core 100 TF system in 2006
- Upgrade 100 TF to 250 TF in late-2007
- Deploy 1 PF Cray Baker late 2008
- Sustained-PF Cray Cascade system 2010



LCF managed as a major DOE project

- **The LCF Project for the delivery of the 250 TF and 1 PF computer systems is being managed to a 200+ element WBS with detailed scope, cost, and schedule.**



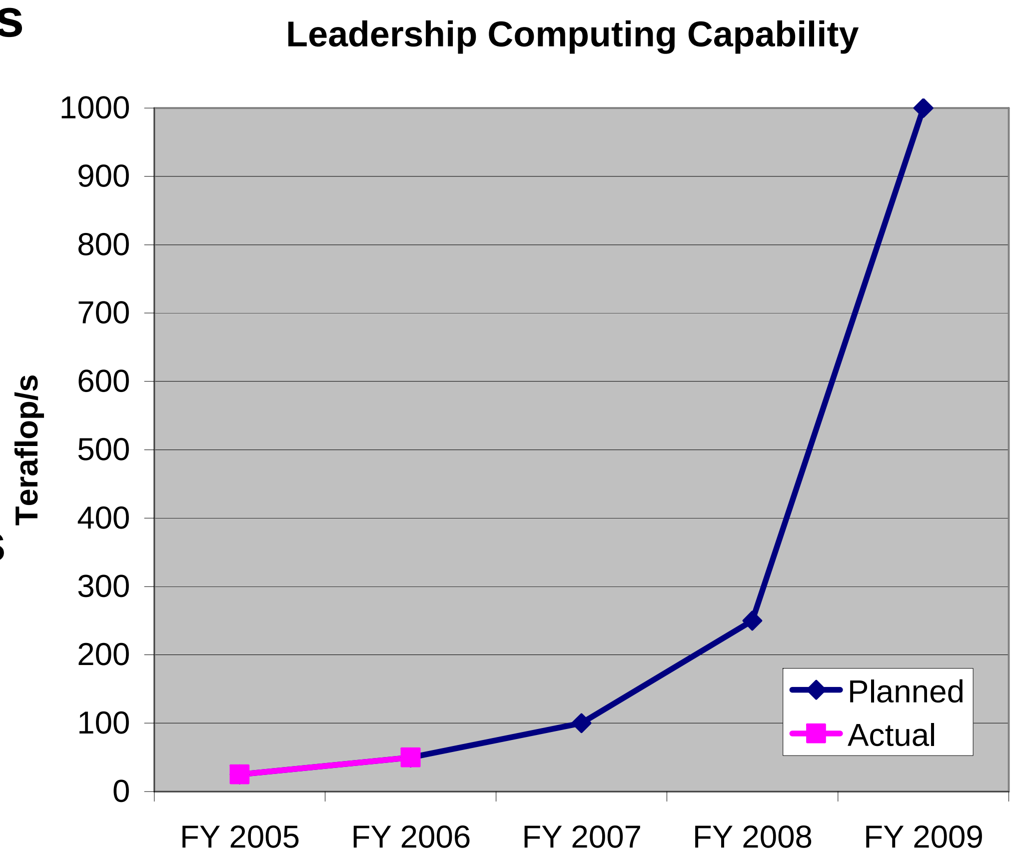
- **The LCF has developed a detailed Risk Management Plan and is actively tracking and mitigating major project risks.**

Major Project Risks

1. Operating System for multi-core processor systems
2. Scalable file system for 1 PF system
3. Applications readiness
4. Market forces could delay multi-core parts from AMD

LCF Project:
Delivering on Schedule, Scope, and Budget

- **Phoenix was upgraded from 256 to 512 to 1,024 processors on schedule.**
- **Jaguar was installed, accepted, and turned over to users on schedule.**
- **The upgrade of Jaguar to dual-core processors with 21 TB of memory and 54 TF is done and acceptance is on schedule.**



LCF resources

Network Routers

Control Network
1 GigE
10 GigE
UltraScience

August 2006 Summary

7 Systems

Cray XT3 Jaguar

(10,506) 2.6GHz 21TB Memory

120 TB

Cray X1E Phoenix

(1,024) 0.5GHz 2 TB Memory

44 TB

SGI Altix Ram

(256) 1.5GHz 2TB Memory

36 TB

IBM SP4 Cheetah

(864) 1.3GHz 1.1TB Memory

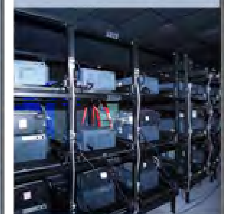
32 TB

IBM Linux NSTG

(56) 3GHz 76GB Memory

4.5 TB

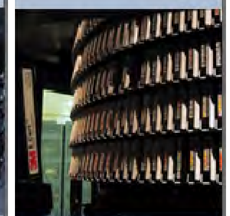
Visualization Cluster



(128) 2.2GHz 128GB Memory

9 TB

IBM HPSS



Many Storage Devices Supported

5 TB

Supercomputers
13,834 CPUs
26TB Memory

74 TFlops

Total Shared Disk 250.5 TB

Evaluation Platforms

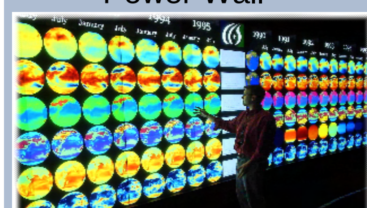
- 144-processor Cray XD1 with FPGAs
- SRC Mapstation
- Clearspeed
- BlueGene (at ANL)

Test Systems

- 96-processor Cray XT3
- 32-processor Cray X1E*
- 16-processor SGI Altix

Scientific Visualization Lab

35 megapixels Power Wall



Backup Storage

5PB

5 PB Data Storage

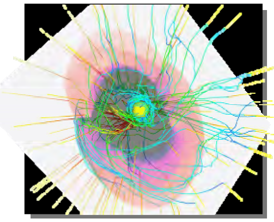
Phoenix – 18.5 TF Cray X1E vector system

- **Highly scalable hardware and software**
- **High sustained performance on key applications**

CRAY X1E
PHOENIX

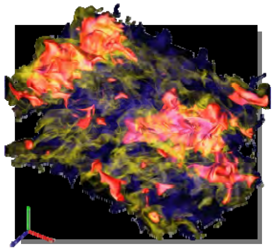
(1,024) 0.5GHz
2 TB Memory

44 TB



Astrophysics

Simulations have uncovered a new instability of the shock wave and a resultant spin-up of the stellar core beneath it, which may explain key observables such as neutron star "kicks" and the spin of newly-born pulsars



Combustion

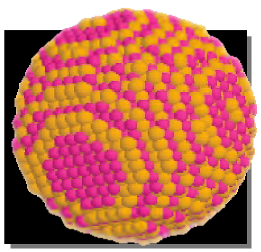
Calculations show the importance of the interplay of diffusion and reaction, particularly where strong finite-rate chemistry effects are involved

Jaguar – 54 TF Cray XT3

Upgraded to 10,424 processors and 21 TB of memory in July 2006

CRAY XT3
JAGUAR

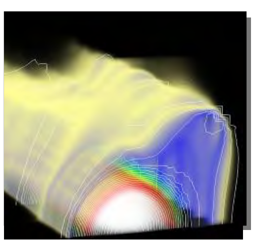
(5,294) 2.6 GHz
21TB Memory
120TB



Materials Science

Nanoparticles present capacity for information storage dramatically greater than bulk materials

Over 81% of theoretical peak performance was achieved for non-collinear magnetic structure calculation of FePt particles



Plasma Turbulence

Largest-ever simulation of plasma behavior in a tokamak crucial to harness power of fusion reactions; simulation used 60% of Jaguar resources

Jaguar Dual-Core Upgrade

54TF Hardware Upgrade

July 11-18, 2006

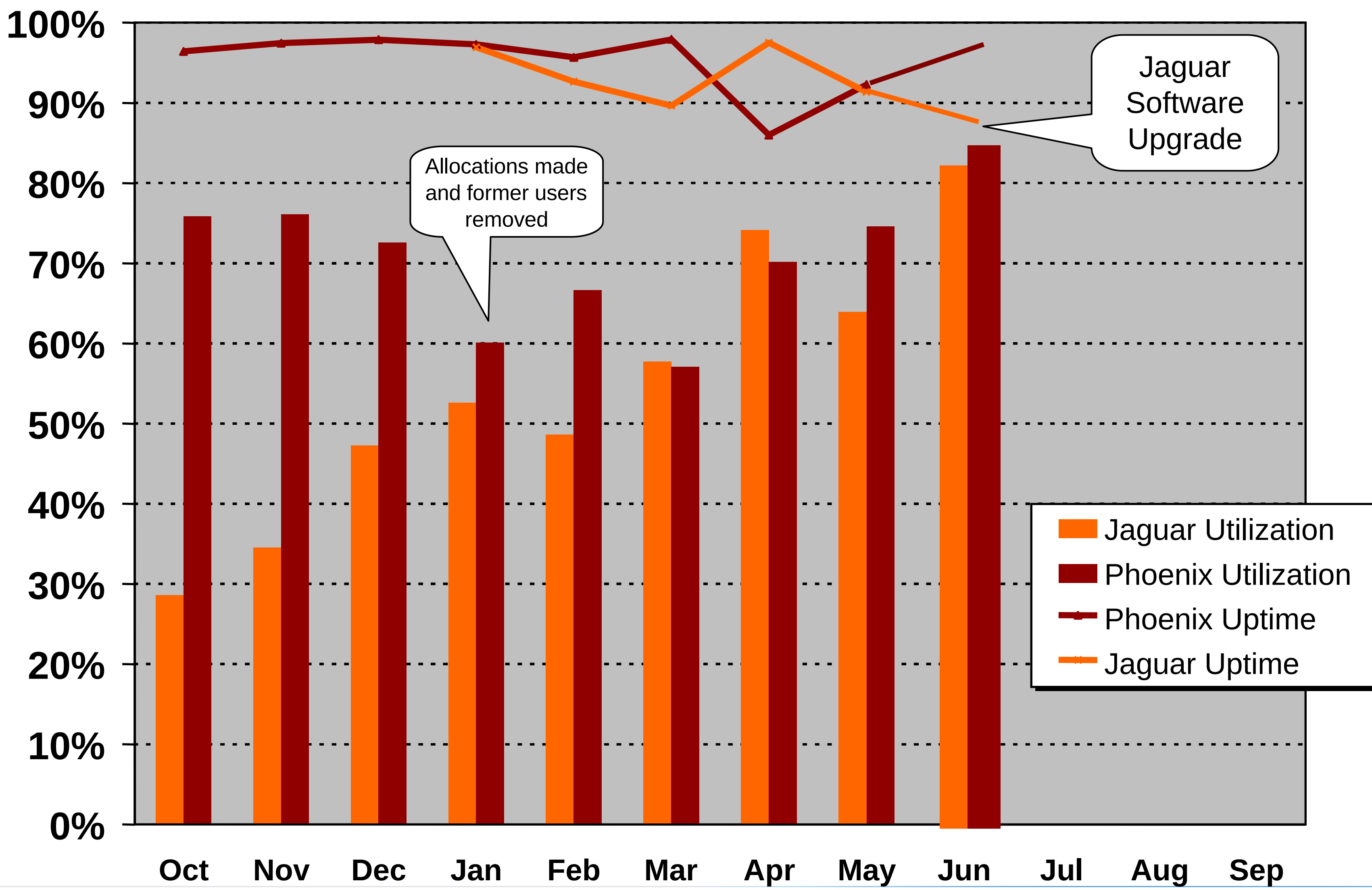
- ☑ **Field replaced 5,212 Opteron processors (8 failures)**
- ☑ **Added 5,212 memory DIMMS (13 failures)**
- ☑ **Rewired interconnect to double the bisection bandwidth**
- ☑ **Added power supplies and upgraded system firmware**
- ☑ **Passed HW acceptance**

First XT3 dual-core upgrade performed by Cray was completed on schedule with 0.2% component failure!

Replacing processors on boards

Replacing an upgraded board

FY 2006 LCF system utilization



LCF supports broad spectrum of scientific domains and associated application codes

- **Wide spectrum of science domains**
- **FY06 LCF projects**
  - Accelerator design
  - Chemistry
  - Climate
  - Combustion
  - Fusion
  - Materials
  - Molecular biology
  - Nuclear structure
  - Supernova ignition

- **Wide spectrum of capability requirements (within each science domain)**
- **Bigger problems**
  - Higher resolution
  - More grid points or particles
- **Harder problems**
  - More physics
  - More-expensive grid points
  - More time steps
  - Tightly coupled

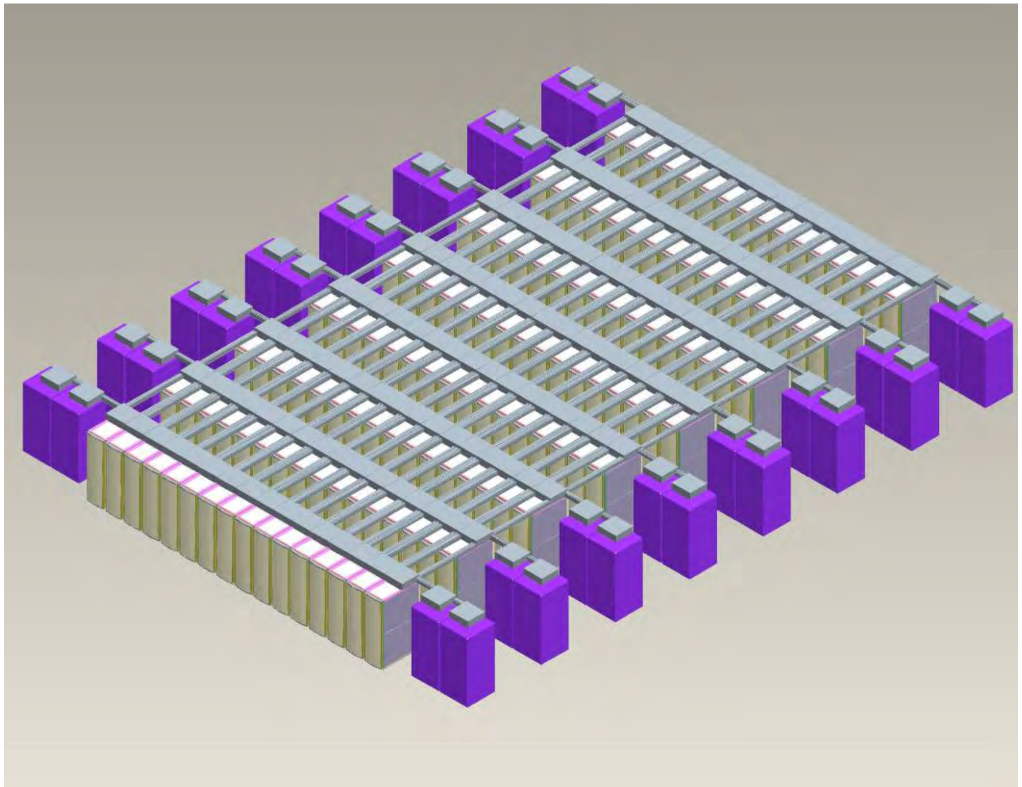
Designing petascale system to match applications

System Attribute	Application Behaviors and Properties Benefiting from Attribute
Node Peak Flops	All computationally intensive algorithms Critical to increasing performance of non-scalable algorithms
Mean Time to Interrupt (MTTI)	Applications with primitive restart capability or large restart files
WAN Bandwidth	Domain areas with community data/repositories; remote visualization and analysis
Node Memory Capacity	High degrees of freedom per node, multi-component/multi-physics, volume visualization, data replication parallelism, subgrid models (PIC)
Local Storage Capacity	Time series algorithms, out-of-core algorithms, debugging at scale
Archival Storage Capacity	Large data that must be preserved for future analysis or comparison; for community databases; expensive to recreate;
Memory Latency	Algorithms with random data access patterns for small data
Interconnect Latency	Global reduction; explicit algorithms using nearest-neighbor or systolic communication; interactive visualization; iterative solvers; pipelined algorithms
Disk Latency	Naïve out-of-core memory usage; many small I/O files
Interconnect Bandwidth	Big messages, global reductions of large data; implicit algorithm with large degrees of freedom per grid point;
Memory Bandwidth	Large multi-dimensional data structures and indirect addressing; lots of data copying or transposition; sparse matrix operations
Disk Bandwidth	Reads/writes large amounts of data;; well-structured out-of-core memory usage

1000 TF Baker system in 2008

System configuration

- **1 PF peak**
- **23,936 multi-core processors**
- **136 cabinets**
- **32x34x24 topology**
- **34 heat exchange units**
- **7 MW power**



1 PF Cray system in 2008

Partner with TVA to solve Leadership Computing power requirements
Significant challenge for the industry

System	Peak Performance	No. of Cores	Power
Jaguar	25TF	5,212	0.9 MW
Jaguar	50 TF	10,424	1.2 MW
Jaguar	100TF	23,480	2.8 MW
Jaguar	250TF	36,536	3.0 MW
Baker	1PF	95,744	7.0 MW
Top500 June 2006	2.79PF	873,595	~100 MW
Intel 1 yr Production	1600PF	300,000,000	~50 GW

- Partner with TVA to deliver reliable, cost-effective, power
- 70MW LCF substation under construction – upgradeable to 170 MW
- Upgraded transmission capability – Three redundant feed circuits
- TVA load shed 1.9GW last week without interruption to LCF

"Intel expects to sell 60 million dual-core chips this year, accounting for about a quarter of total processor sales"

-Justin Rattner, Intel CTO

Baker system can be upgraded to Cray’s HPCS “Cascade” system

- **Faster processors**
- **More and faster memory**
- **Vector and multi-treaded processors**
- **FPGAs**
- **Improved RAS system**

- **Reuse Baker Infrastructure**
  - Cabinets
  - Power
  - Cooling infrastructure
- **Builds on Baker software stack**
  - Compute node microkernel
  - Full featured service and I/O nodes
  - Scalable parallel file system

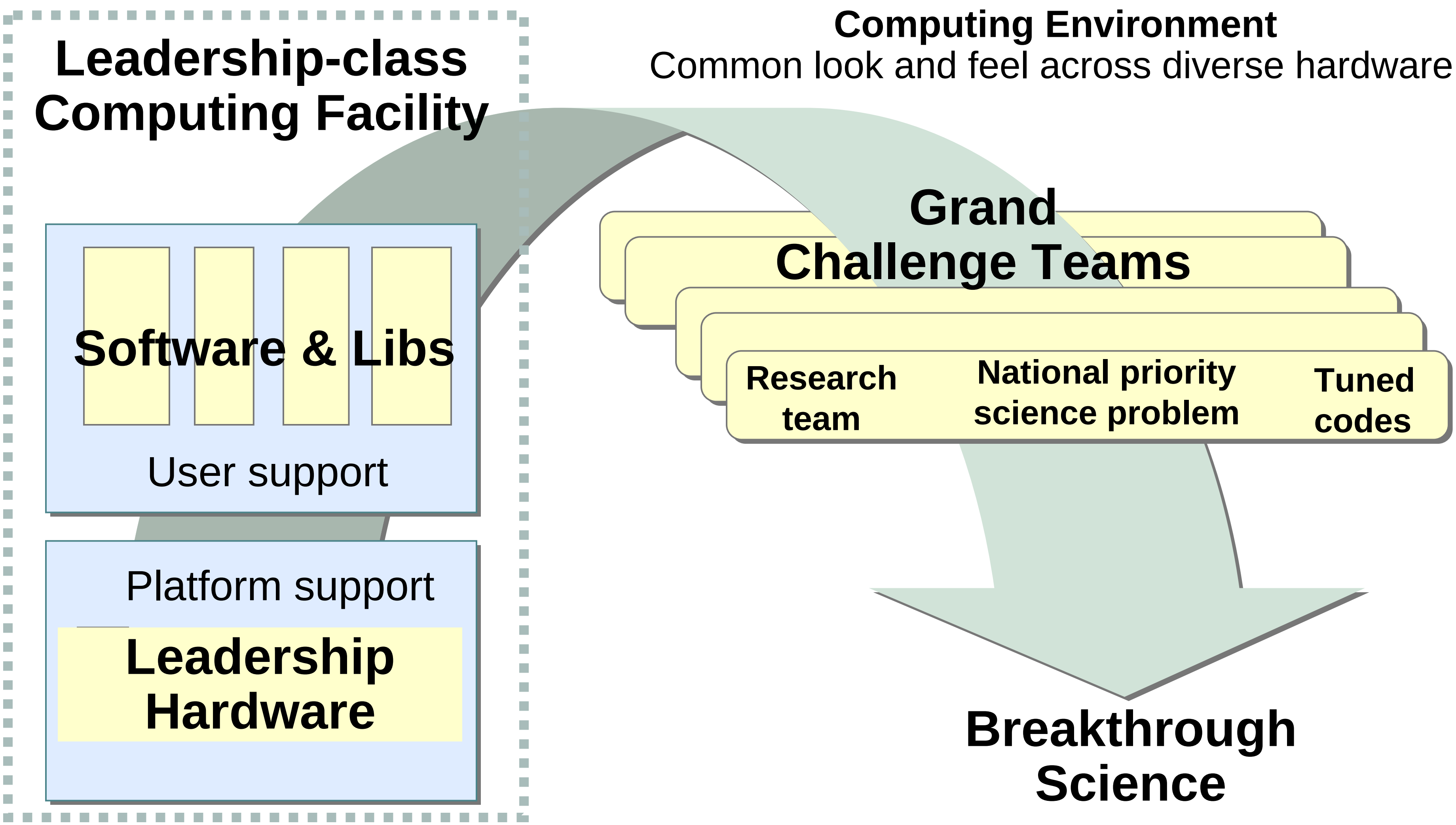
Deliver high productivity systems through improved system software and tools

Critical that 1 PF and 250TF systems deliver science on day one

- Identify and mitigate key gaps in system software and tools
  - **Critical: software required for hero programmers to make science breakthroughs on petascale system**
    - **Highly tuned MPI and math libraries**
    - **Light Weight Kernel – Linux version tuned and stable**
    - **Petascale I/O and file system**
    - **Networking the 1PF Baker system to rest of LCF and world**
    - **Reliability and fault tolerance for**
      - **science applications, middleware, and system software**
  - **Important: software required for the typical LCF user to make productive, efficient use of the petascale system (plug and play)**
    - **Common programming environment**
    - **Advanced debugging**
    - **Automated performance analysis**
- Engage laboratory community on advanced system software and tools

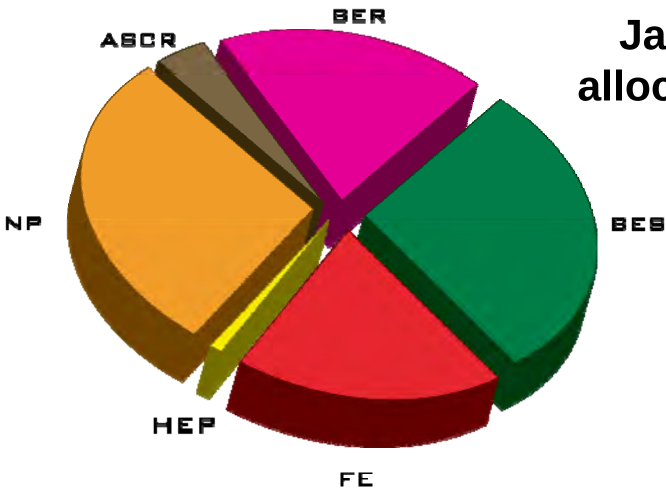
The Goal is Science

Facility plus hardware, software, and science teams all contribute to Science breakthroughs

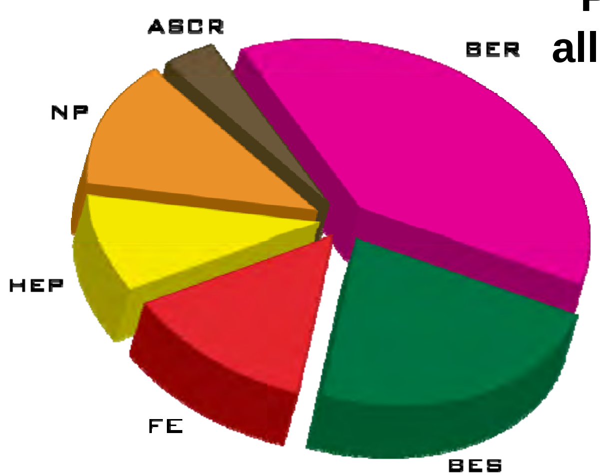


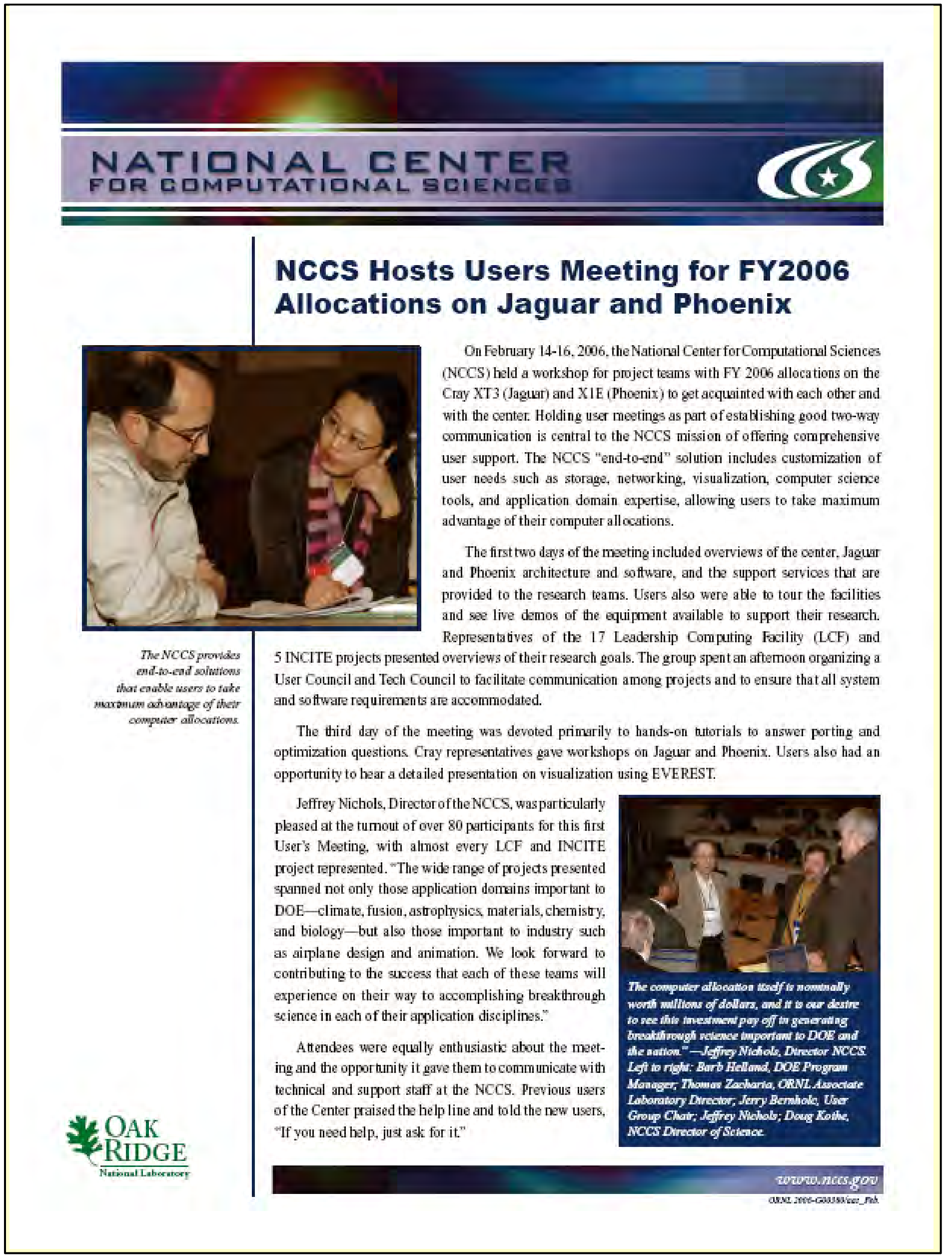
FY2006 allocations and user engagement

Jaguar allocations



Phoenix allocations





NATIONAL CENTER FOR COMPUTATIONAL SCIENCES

NCCS Hosts Users Meeting for FY2006 Allocations on Jaguar and Phoenix

On February 14-16, 2006, the National Center for Computational Sciences (NCCS) held a workshop for project teams with FY 2006 allocations on the Cray XT3 (Jaguar) and X1E (Phoenix) to get acquainted with each other and with the center. Holding user meetings as part of establishing good two-way communication is central to the NCCS mission of offering comprehensive user support. The NCCS "end-to-end" solution includes customization of user needs such as storage, networking, visualization, computer science tools, and application domain expertise, allowing users to take maximum advantage of their computer allocations.

The NCCS provides end-to-end solutions that enable users to take maximum advantage of their computer allocations.

The first two days of the meeting included overviews of the center, Jaguar and Phoenix architecture and software, and the support services that are provided to the research teams. Users also were able to tour the facilities and see live demos of the equipment available to support their research. Representatives of the 17 Leadership Computing Facility (LCF) and 5 INCITE projects presented overviews of their research goals. The group spent an afternoon organizing a User Council and Tech Council to facilitate communication among projects and to ensure that all system and software requirements are accommodated.

The third day of the meeting was devoted primarily to hands-on tutorials to answer porting and optimization questions. Cray representatives gave workshops on Jaguar and Phoenix. Users also had an opportunity to hear a detailed presentation on visualization using EVEREST.

Jeffrey Nichols, Director of the NCCS, was particularly pleased at the turnout of over 80 participants for this first User's Meeting, with almost every LCF and INCITE project represented. "The wide range of projects presented spanned not only those application domains important to DOE—climate, fusion, astrophysics, materials, chemistry, and biology—but also those important to industry such as airplane design and animation. We look forward to contributing to the success that each of these teams will experience on their way to accomplishing breakthrough science in each of their application disciplines."

Attendees were equally enthusiastic about the meeting and the opportunity it gave them to communicate with technical and support staff at the NCCS. Previous users of the Center praised the help line and told the new users, "If you need help, just ask for it."

The computer allocation itself is nominally worth millions of dollars, and it is our desire to see this investment pay off in generating breakthrough science important to DOE and the nation." —Jeffrey Nichols, Director NCCS. Left to right: Barb Helland, DOE Program Manager; Thomas Zacharia, ORNL Associate Laboratory Director; Jerry Bernholc, User Group Chair; Jeffrey Nichols; Doug Kothe, NCCS Director of Science.

OAK RIDGE National Laboratory

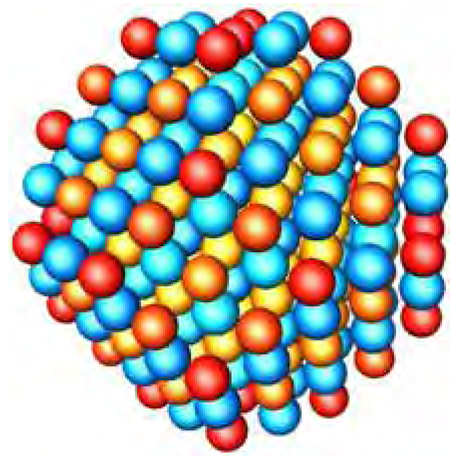
www.nccs.gov

FY06 Allocated Projects by Science Area

Project	Jaguar Allocation	Percent of Jaguar	Phoenix Allocation	Percent of Phoenix	Type	Description	Domain Science	PI
AST003	1,250,000	4.1%	0	0.0%	LCF	Multi-dimensional Simulations of Core-Collapse Supernovae	Astrophysics	Burrows
AST004	3,000,000	9.9%	0	0.0%	LCF	Ignition and Flame Propagation in Type Ia Supernovae	Astrophysics	Woosley
AST005	3,550,000	11.7%	700,000	11.9%	LCF	Multi-dimensional Simulations of Core-Collapse Supernovae	Astrophysics	Mezzacappa
BIO014	500,000	1.7%	0	0.0%	LCF	Next Generation Simulations in Biology	Biology	Agarwal
BIO015	1,484,800	4.9%	0	0.0%	INCITE	Molecular Dynamics Simulations of Molecular Motors	Biology	Karplus
CHM022	1,000,000	3.3%	300,000	5.1%	LCF	Rational Design of Chemical Catalysts	Chemistry	Harrison
CLI016	0	0.0%	29,000	0.5%	LCF	Role of Eddies in Thermohaline Circulation	Climate	Cessi
CLI017	3,000,000	9.9%	2,000,000	33.9%	LCF	Climate-Science Computational End Station	Climate	Washington
CLI018	1,496,856	4.9%	0	0.0%	LCF	Studies of Turbulent Transport in the Global Ocean	Climate	Peacock
CSC023	1,000,000	3.3%	200,000	3.4%	LCF	PEAC End Station	Computer Science	Worley
CSC026	950,000	3.1%	0	0.0%	INCITE	Real-Time Ray-Tracing	Computer Science	Smyth
EEF049	3,500,000	11.6%	300,000	5.1%	LCF	Simulations in Strongly Correlated Electron Systems	Materials Science	Schulthess
EEF050	0	0.0%	200,000	3.4%	INCITE	Large Scale Computational Tools for Flight Vehicles	Engineering	Hong
EEF051	500,000	1.7%	0	0.0%	INCITE	Numerical Simulation of Brittle nd Ductile Materials	Materials Science	Ortiz
FUS011	2,000,000	6.6%	225,000	3.8%	LCF	Gyrokinetic Plasma Simulation	Fusion	Lee
FUS012	0	0.0%	440,240	7.5%	LCF	Tokamak Operating Regimes Using Gyrokinetic Simulations	Fusion	Candy
FUS013	3,000,000	9.9%	0	0.0%	LCF	Wave-Plasma Interaction and Extended MHD in Fusion Systems	Fusion	Batchelor
FUS014	0	0.0%	400,000	6.8%	INCITE	Interaction of ETG and ITG/TEM Gyrokinetic Turbulence	Fusion	Waltz
HEP004	30,000	0.1%	0	0.0%	LCF	Reconstruction of CompHEP-produced Hadronic Backgrounds	High Energy Physics	Newman
HEP005	0	0.0%	500,000	8.5%	LCF	Design of Low-loss Accelerating Cavity for the ILC	Accelerator Physics	Ko
NPH004	1,000,000	3.3%	0	0.0%	LCF	Ab-inito Nuclear Structure Computations	Nuclear Physics	Dean
SDF022	3,000,000	9.9%	600,000	10.2%	LCF	High-Fidelity Numerical Simulations of Turbulent Combustion	Combustion	Chen
Total	30,261,656		5,894,240					

Insight into next generation recording media: magnetic properties of FePt nanoparticles

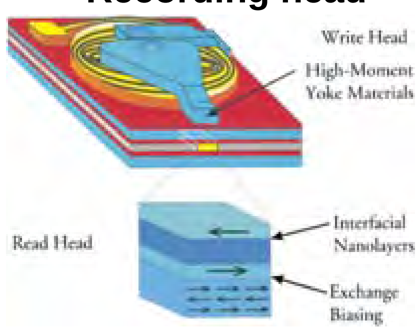
PI: Thomas Schulthess, Oak Ridge National Laboratory



FePt magnetic nanoparticles

Material identified by industry for next generation magnetic recording (>1Tbpsi)

Recording head



Highly optimized calculations on NLCF

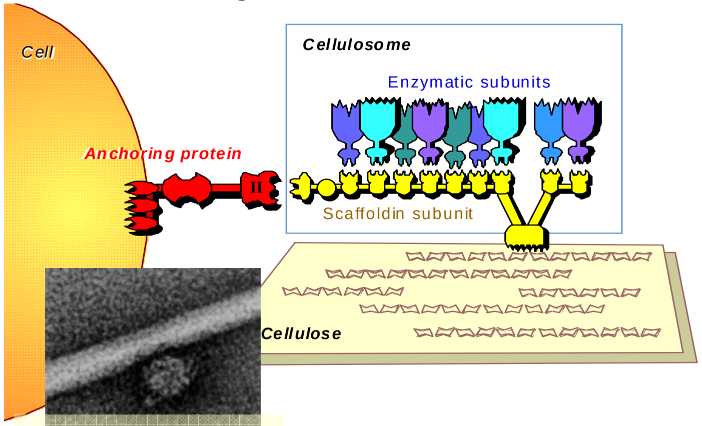
- **DFT calculations using codes optimized with Cray Center of Excellence**
- **Largest, most complex calculations of this type to date**
- **~50% of peak on 512 Jaguar processors (1TFLOP) for ~800 atoms**
- **2000 atoms planned**

Summary of results

- **Revealed, for first time, strong influence of nanoscale on magnetic structure of FePt nanoparticles**
- **Sensitivity of magnetic structure to small changes demonstrates importance of calculations for materials design and optimization**

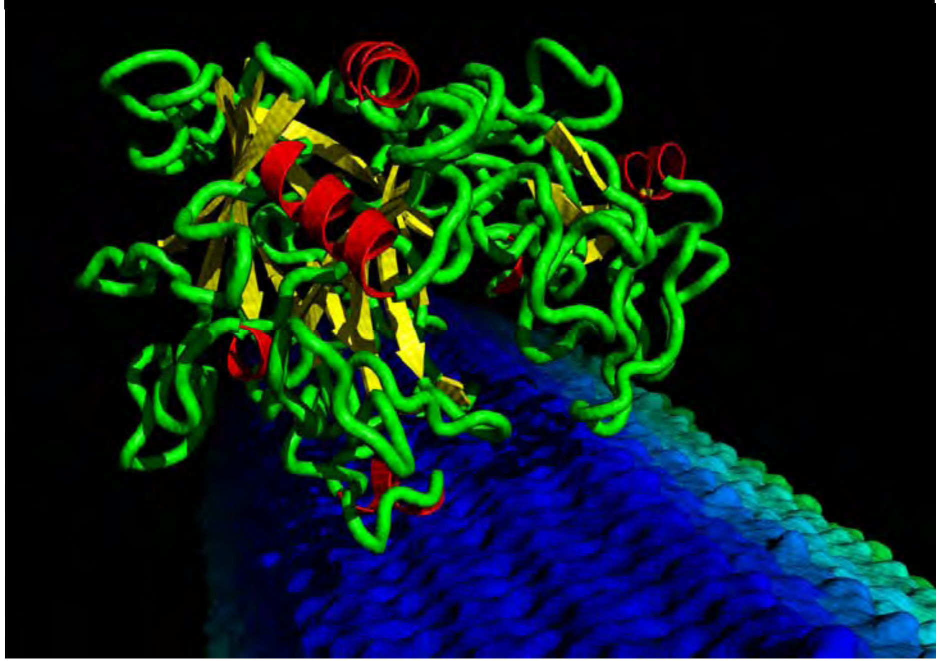
Creation of efficient enzymes for cellulose degradation through protein engineering

- **Renewable energy: ethanol production from cellulose**
- **Detailed understanding of cellulase enzyme mechanisms from multi-scale modeling**
  - 1-100 ns trajectories for systems with over 800,000 atoms
- **Simulations with different substrates & mutant enzymes**



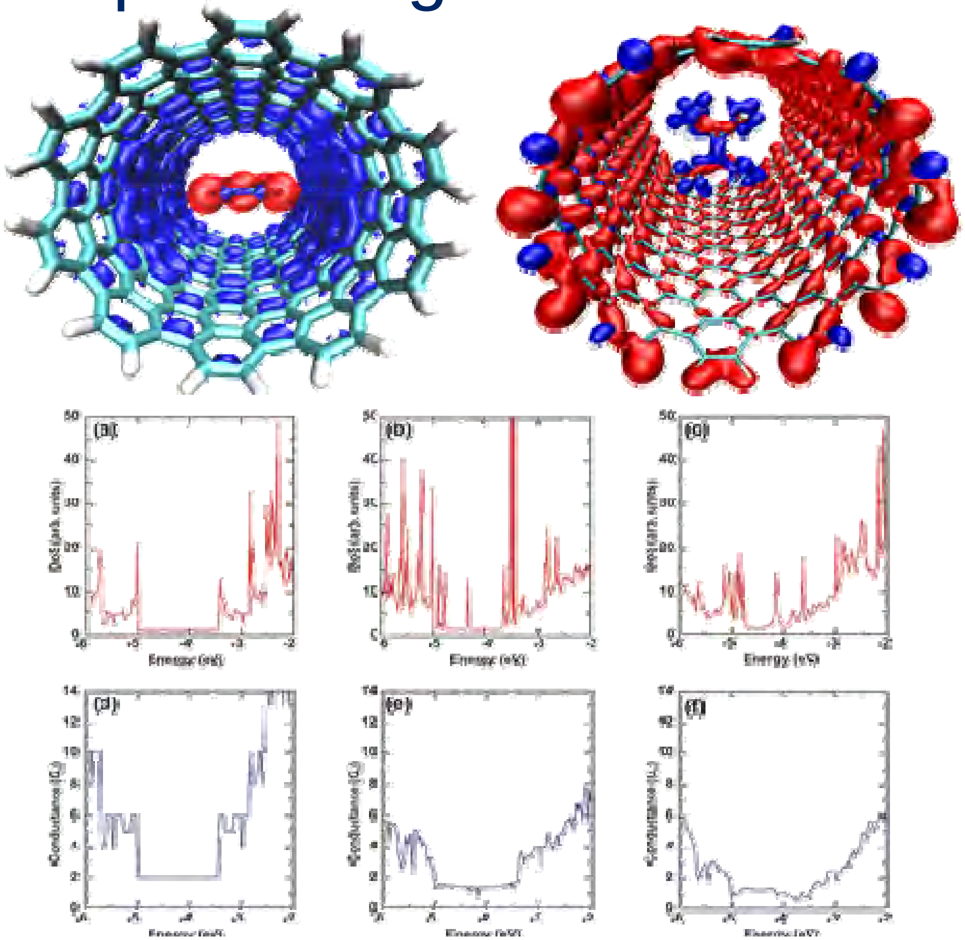
***C. thermocellum* growing on cellulose**

Metabolism of Cellulose by Exocellulase Enzyme

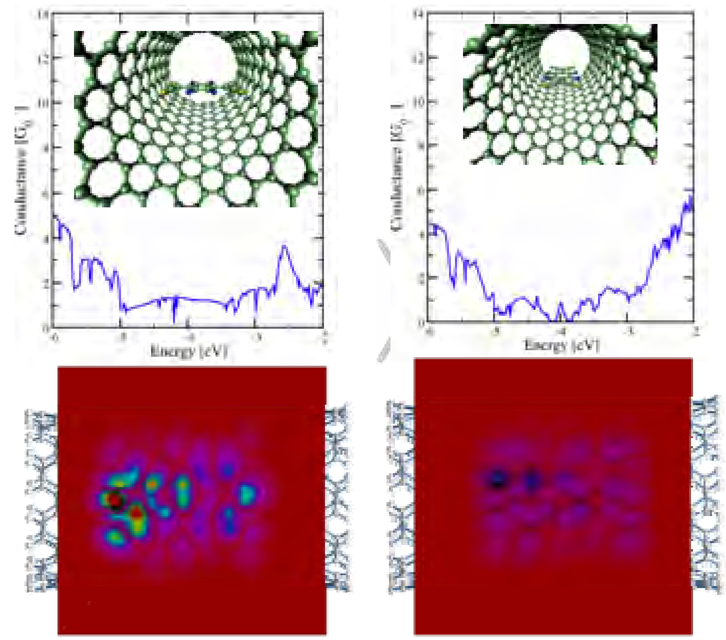


Next Generation Simulations in Biology
PI: Pratul Agarwal, ORNL

Quantum conductance of amphoterically doped single-walled carbon nanotubes



Molecular switch



Scattering states analysis of switching effects

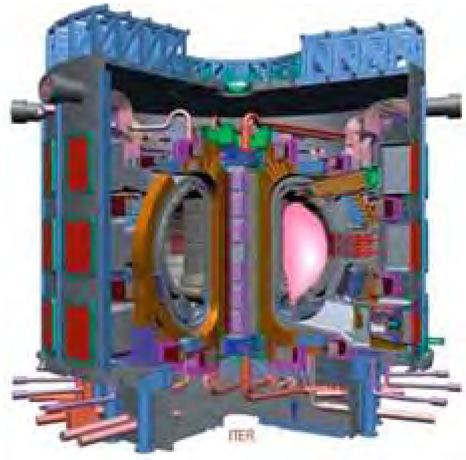
Implementing hybrid approach for examination of electronic properties of molecular-based structures an efficient and accurate procedure demonstrated for studying effects of amphoteric doping of SWCNTs

Munier and Sumpter, *J. Chem. Phys.* 128, 024705 (2005)

New ability to model coupled ETG/ ITG turbulence in shaped plasma in nonlinear phase

PI: Ron Waltz, General Atomics (waltz@fusion.gat.com)

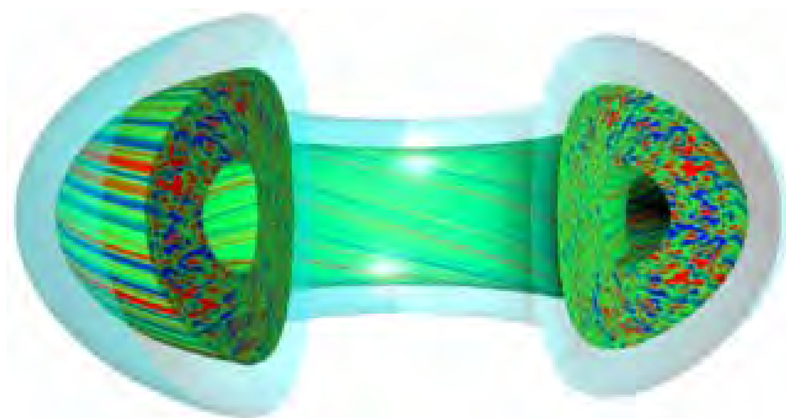
Impact: Modeling and understanding of plasma turbulence is crucial for development of stable and efficient fusion devices



Problem: Computational modeling of interaction of turbulence on ion and electron spatial and temporal scales

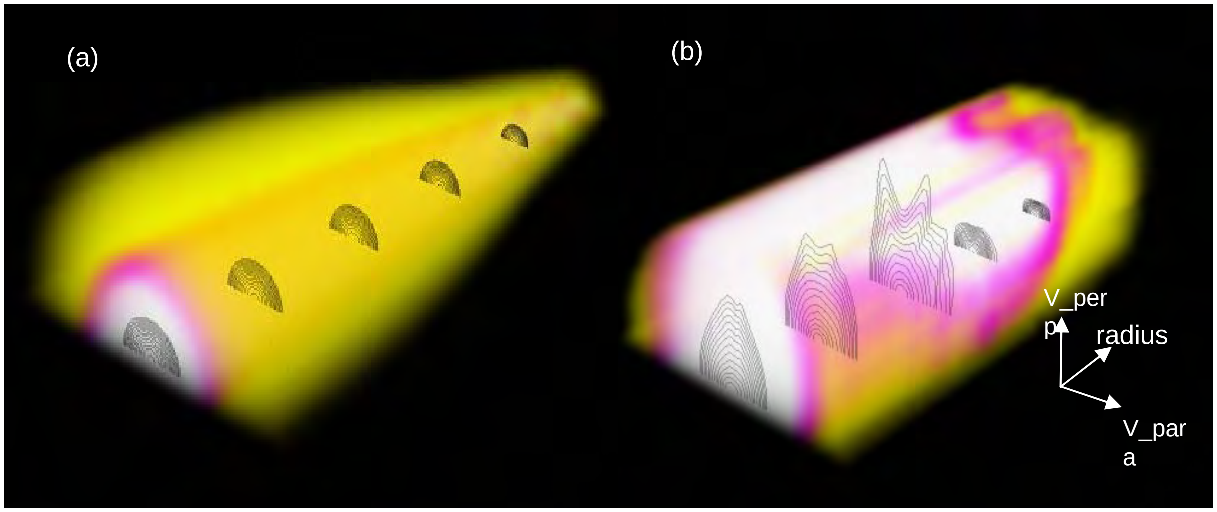
Scales differ by orders of magnitude and have traditionally been treated by separate simulations

Results: Simulations shed new light on how short-wavelength ETG turbulence comes into play as long-wavelength (ITG/TEM) turbulence is suppressed in pedestal (an edge transport barrier)



Evolution of non-thermal plasma distributions during ion cyclotron resonance heating of Tokamak plasmas

- **Interaction of radio-frequency (RF) waves with fusion alpha particles must be understood and optimized for successful fusion power production**
  - Increasing current knowledge will help to plan and analyze ITER experiments
- **Recent self-consistent simulations of evolution of energetic ions in plasma heated by RF waves in ion cyclotron range of frequencies (ICRF) was carried out on the LCF systems – *a significant step***
- **AORSA code coupled to CQL3D Fokker-Planck code can now simulate the evolution of ion distributions during ICRF heating**



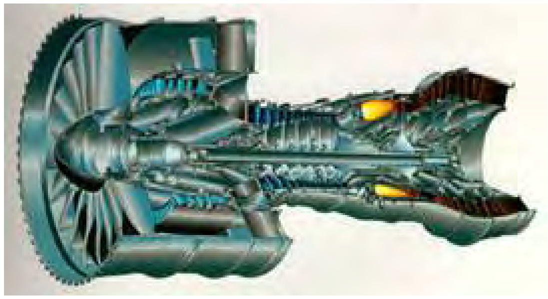
Bounce-averaged minority hydrogen distribution function in Alcator C-MOD shot 1051206002 at *f* = 80 MHz: (a) 0th iteration; (b) 4th iteration

Simulation of Wave-Plasma Interaction and Extended MHD in Fusion Systems
PI: Don Batchelor, ORNL

Improving Energy Security through new insights into lean premixed combustion

PI: Jackie Chen, Sandia National Laboratories (chen@snl.gov)

Impact: Increase Thermal Efficiency and Decrease Emissions in New Land-Based Natural Gas Turbine Designs



Problem: Understand the flamelet and thin-reaction zones where lean premixed combustion occurs

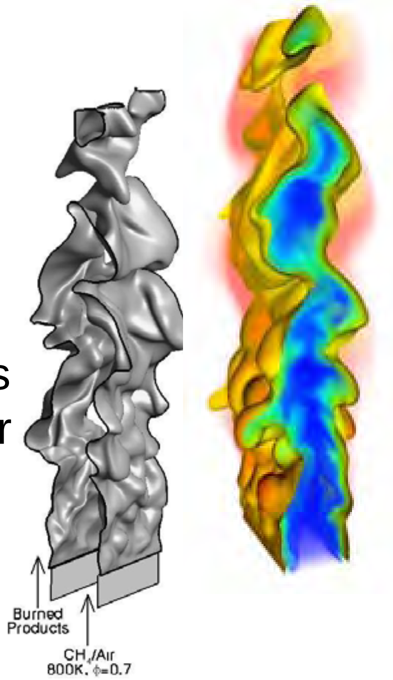
Characterized by strong turbulence and chemistry interactions

Challenges: Combustion at the lean flammability limit is hard

- Prone to extinction,
- unburned hydrocarbon emission,
- large amplitude pressure oscillations,
- emission of toxic carbon monoxide

Results: Flame structure is penetrated by small scale eddies, but mean reaction rates still resemble a strained laminar flame

New insights into source terms that influence flame thickness



Thank you