

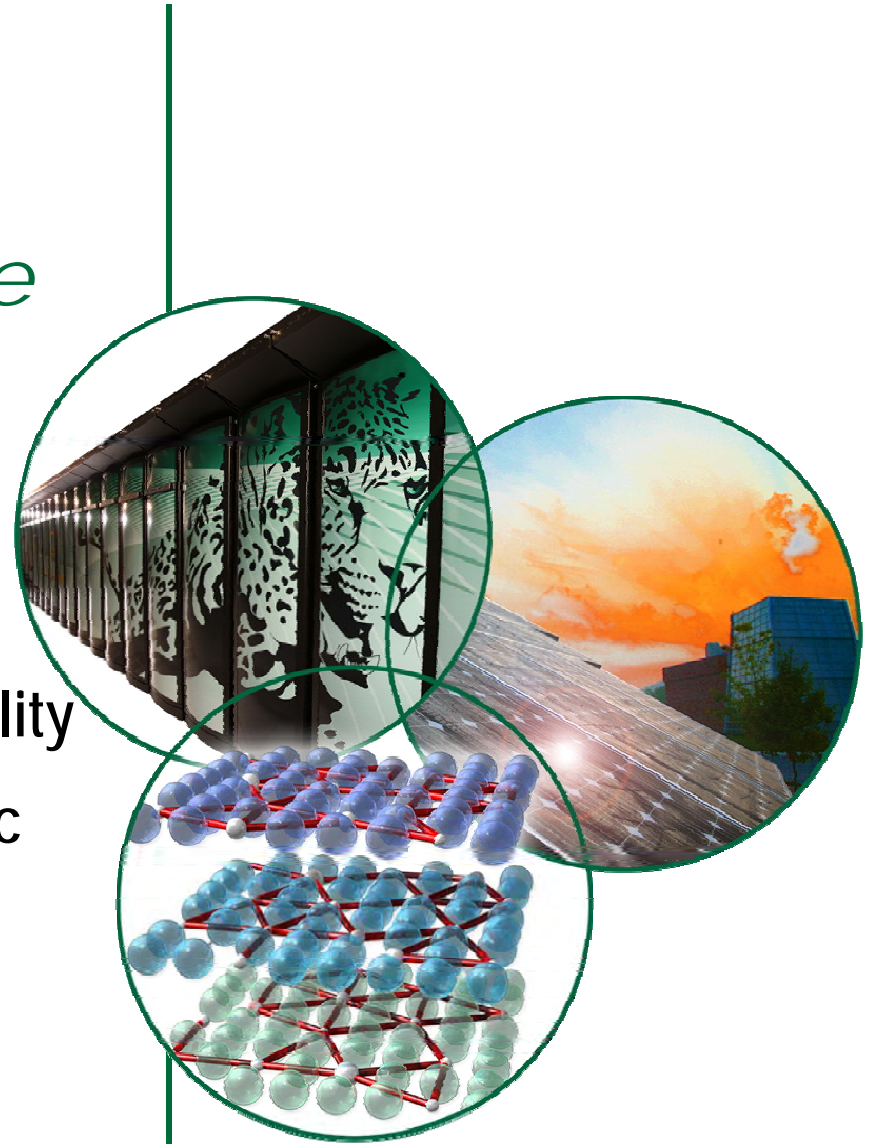
Establishing Petascale Scientific Computing

Buddy Bland

Oak Ridge Leadership Computing Facility

Presentation to the Advanced Scientific
Computing Advisory Committee

March 3, 2009



An outstanding launch for petascale computing in ASCR and ORNL at SC'08

Only 41 days after assembly of a totally new 150,000 core system

- Jaguar beat the previous #1 performance on Top500
- Jaguar had two *real* applications running over 1 PF
 - DCA++ 1.35 PF Superconductivity problem
 - LSMS 1.05 PF Thermodynamics of magnetic nanoparticles problem
- Jaguar won two of the four HPC Challenge awards

ASCR Sweep of HPC Challenge Awards



- HPC Challenge awards are given out annually at the Supercomputing conference
- Awards in four categories, result published for two others; tests many aspects of the computer's performance and balance
- Must submit results for all benchmarks to be considered
- Unfortunately, ORNL team only had two days on the machine to get the results. Got a better G-FFT number (5.804) the next day. ORNL submitted only baseline (unoptimized) results.

G-HPL (TF)		EP-Stream (GB/s)		G-FFT (TF)		G-Random Access (GUPS)		EP-DGEMM (TF)		PTRANS (GB/s)	
ORNL	902	ORNL	330	ANL	5.08	ANL	103	ORNL	1,257	SNL	4,994
LLNL	259	LLNL	160	SNL	2.87	LLNL	35.5	ANL	362	LLNL	4,666
ANL	191	ANL	130	ORNL	2.77↑	SNL	33.6	LLNL	162	LLNL	2,626

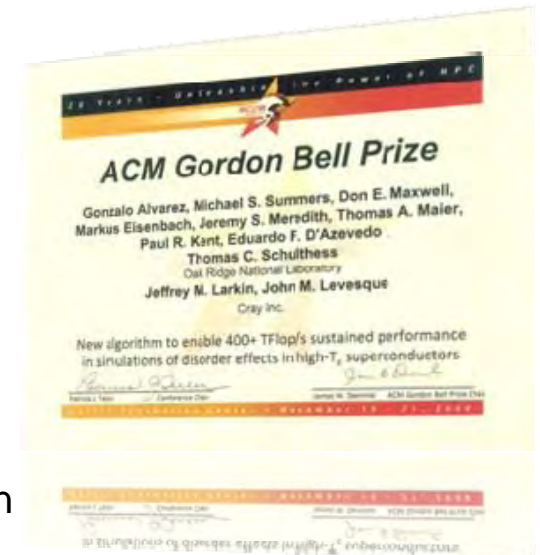
HPC CHALLENGE

Gordon Bell prize awarded to ORNL team



Three of six GB finalist ran on Jaguar

- A team led by ORNL's Thomas Schulthess received the prestigious 2008 Association for Computing Machinery (ACM) Gordon Bell Prize at SC08
- For attaining fastest performance ever in a scientific supercomputing application
- Simulation of superconductors achieved 1.352 petaflops on ORNL's Cray XT Jaguar supercomputer
- By modifying the algorithms and software design of the DCA++ code, the team was able to boost its performance tenfold





Gordon Bell Finalists

✓DCA++	ORNL
✓LS3DF	LBL
✓SPECFEM3D	SDSC
•RHEA	TACC
•SPaSM	LANL
•VPIC	LANL



Science Applications are Scaling on Jaguar

Science Area	Code	Contact	Cores	Total Performance	Notes
Materials	DCA++	Schulthess	150,144	1.3 PF*	Gordon Bell Winner 
Materials	LSMS	Eisenbach	149,580	1.05 PF	
Seismology	SPECFEM3D	Carrington	149,784	165 TF	Gordon Bell Finalist
Weather	WRF	Michalakes	150,000	50 TF	
Climate	POP	Jones	18,000	20 sim yrs/ CPU day	
Combustion	S3D	Chen	144,000	83 TF	
Fusion	GTC	PPPL	102,000	20 billion Particles / sec	
Materials	LS3DF	Lin-Wang Wang	147,456	442 TF	Gordon Bell Winner 
Chemistry	NWChem	Apra	96,000	480 TF	
Chemistry	MADNESS	Harrison	140,000	550+ TF	

Jaguar Completed Acceptance on December 23rd

- Four phase acceptance test

- Hardware Checkout
- Functionality
- Performance
- Stability

“We unanimously agreed that the procedures had been followed by the LCF team and that the installed 1PF Cray XT5 system passed all the tests.” - Acceptance Review Panel

- Peer Review Panel met on December 29th

Lawrence Buja –UCAR	Philip Jones –LANL
Jackie Chen - Sandia	John Drake –ORNL
Jonathan Carter –LBL	Thomas Schulthess –CSCS/ORNL
Susan Coghlan –ANL/ALCF	Barbara Helland – DOE/ASCR Observer

Pushing Back the Frontiers of Science

Petascale Early Science Projects Tackle National/Global Problems

- **Energy for environmental sustainability**
 - **Climate change:** carbon sequestration, weather event impacts on global climate, decadal climate predictive skill in aerosol forcing, global climate at unprecedented resolution
 - **Bioenergy:** recalcitrance in cellulosic ethanol
 - **Solar:** non-equilibrium semiconductor alloys
 - **Energy storage:** charge storage and transfer in nano-structured supercapacitors
 - **Energy transmission:** role of inhomogeneities in high-T superconducting cuprates
 - **Combustion:** stabilizing diesel jet flames for increased efficiency & decreased emissions
 - **Fusion:** ITER design, optimization, and operation
 - **Nuclear energy:** fully-resolved reactor core neutron state
- **Materials and nanoscience**
 - **Structure of nanowires, nanorods, & strongly correlated materials (magnets)**
- **Fundamental science**
 - **Astrophysics:** decipher core-collapse supernovae & black hole mergers
 - **Chemistry:** elucidate water structure in biological & aqueous-phase systems
 - **Nuclear physics:** probe the anomalously long lifetime of Carbon-14
 - **Turbulence:** dispersion relative to air quality modeling and bioterrorism

Pushing Back the Frontiers of Science

Petascale Early Science Projects Tackle National/Global Problems

- **20+ projects**
 - All need petascale resource to achieve their goals
 - 100+ of the best computational scientists from all over the world
 - Scientists hail from universities, national laboratories, government agencies, private industry
 - ANL, SNL, LANL, LLNL, ORNL, PNNL, LBNL, PPPL, NCAR, NOAA/GFDL, NASA, ...
- **500+ million hours allocated**
 - 6 months period
 - Twice the entire INCITE 2008 allocation
 - Jaguar/XT5 can deliver 3.6M hours *daily*



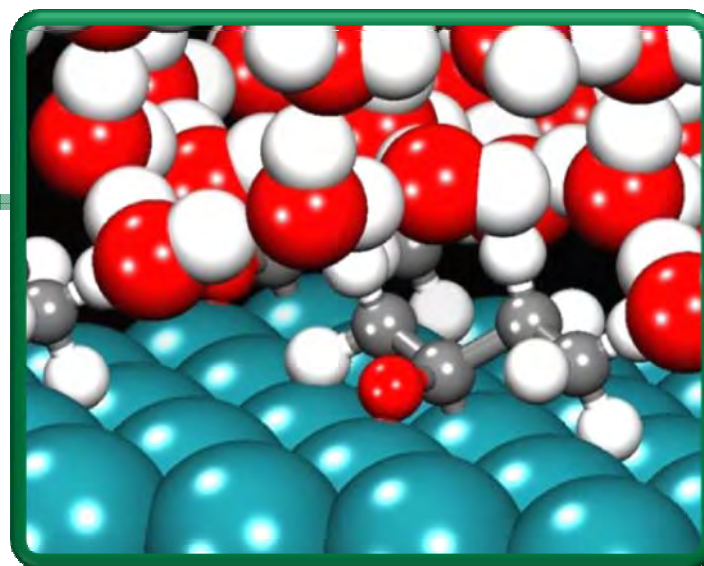
Jaguar/XT5

<http://www.nccs.gov/leadership-science/petascale-early-science/>

Chemical nanoscience at the petascale

chemistry

This project will apply density functional theory to the workings of carbon tube supercapacitors— nanostructures that store two to three orders of magnitude more energy than conventional capacitors—and provide a nanoscale look at the physical and chemical processes that limit storage capacity, useful lifetime, and peak power output.



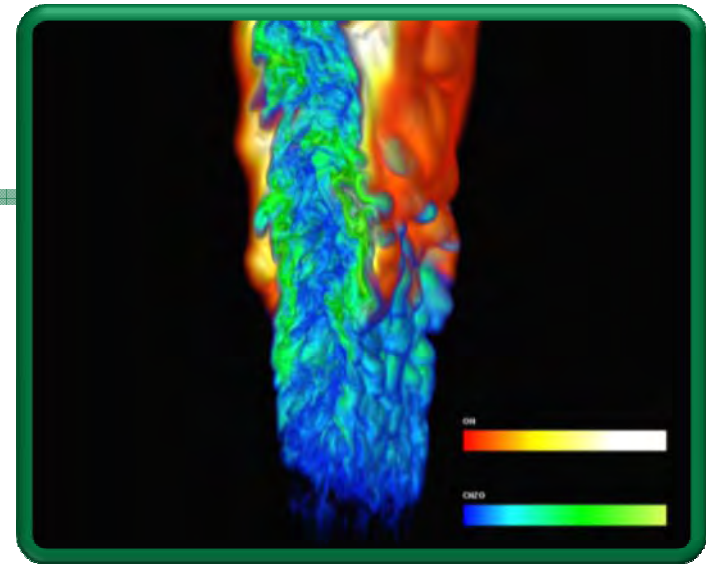
PI: Robert Harrison, ORNL/UT
Code: MADNESS and NWChem
Allocation: 30,000,000 hrs

MADNESS is running at over 140,000 cores
NWChem is running at 96,000 cores at 550 TF

Direct numerical simulation of diesel jet flame stabilization at high pressure

combustion

A massively parallel direct numerical simulation of combustion in a low-temperature, high pressure environment will boost the search for cleaner and more efficient diesel fuels by showing exactly how jet flames ignite and become stable in high-pressure diesel engines.



PI: Jackie Chen, SNL

Code: S3D

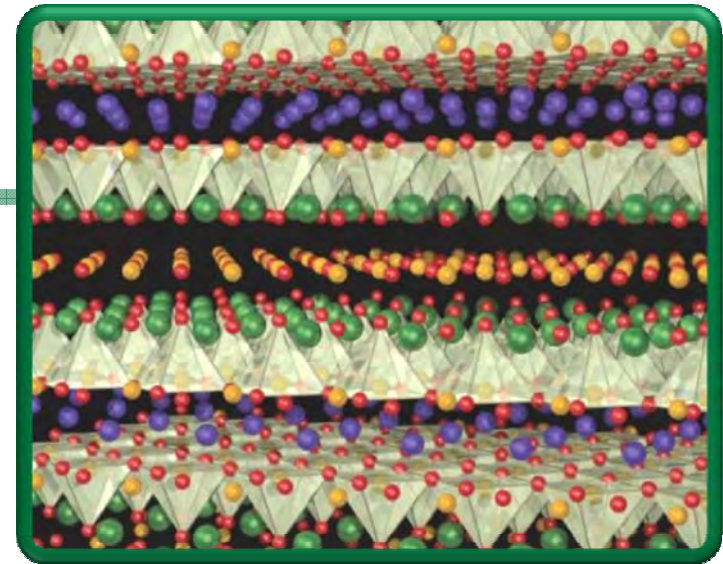
Allocation: 50,000,000 hrs

S3D is running at 144,000 cores

Investigations of the Hubbard Model with Disorder

materials science

This project will use the world's fastest scientific computing application to investigate Cooper pairs—the electron pairings that turn materials into superconductors—with the goal of understanding why they develop at different temperatures in different materials. The eventual goal is to develop superconductors that do not require cooling.



PI: Thomas Schulthess, ETH/ORNL
Code: DCA++
Allocation: 60,000,000 hrs

DCA++ is running at 1.35 PF on 150,000 cores
Winner of Gordon Bell Prize, 2008

High resolution climate explorations using a new scalable dynamical core

climate

This project will conduct a series of climate and deterministic forecast experiments designed to evaluate the role of individual weather events—such as hurricanes—on lower frequency variability in global climate. These simulations will be conducted at resolutions even finer than for current deterministic weather forecasting, allowing scientists to more completely explore simulation requirements for addressing questions about regional climate change



PI: Max Suarez, NASA/GMAO

Code: GEOS-5

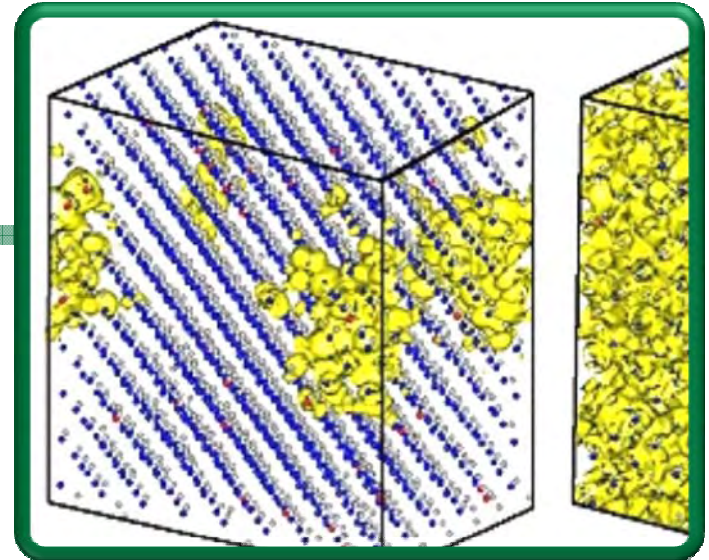
Allocation: 20,000,000 hrs

This project is a great example of interagency cooperation, between NASA, NOAA, and DOE, on pushing the frontiers in climate science research with a scientific plan that complements other pioneering initiatives – James Hack

Charge Patching Method for electronic structures and charge transports of organic and organic/inorganic mixed nanostructures

materials science

This project will apply a new, linear-scaling code—LS3DF—to the internal electronic state of a nanoscale system, a problem that has confounded theoreticians, experimentalists, and computational researchers alike. The researcher team will calculate the internal electric fields of large nanocrystals as well as the piezo electric effect of such fields.



PI: Lin-Wang Wang, LBNL

Code: LS3DF

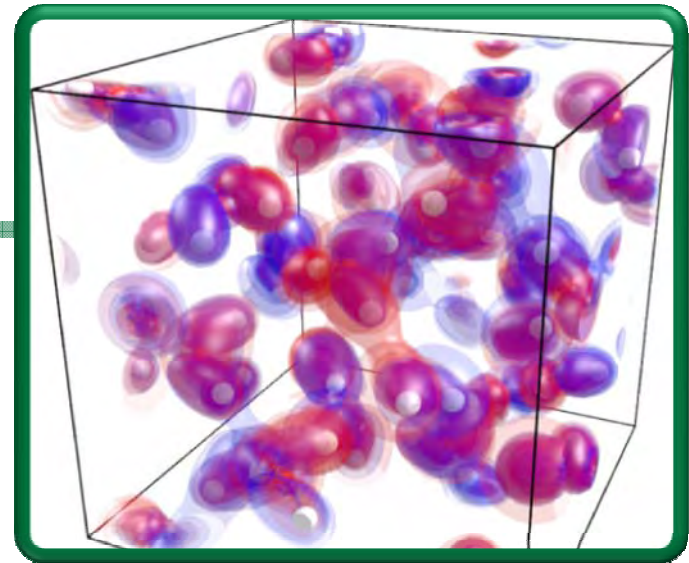
Allocation: 4,000,000 hrs

LS3DF on Jaguar at 147,000 cores and 442 TF
Winner of Gordon Bell Special Prize, 2008

Quantum Monte Carlo calculation of the energetics, thermodynamics and structure of water and ice

chemistry

This project will use a refined Quantum Monte Carlo approach to perform first-principles simulations of the physics of liquid water. An enhanced understanding of this ubiquitous, indispensable, but inadequately understood substance will be invaluable in applications ranging from biology to chemical manufacturing to energy production.



PI: David Ceperley, Univ. of Illinois

Code: QMCPACK

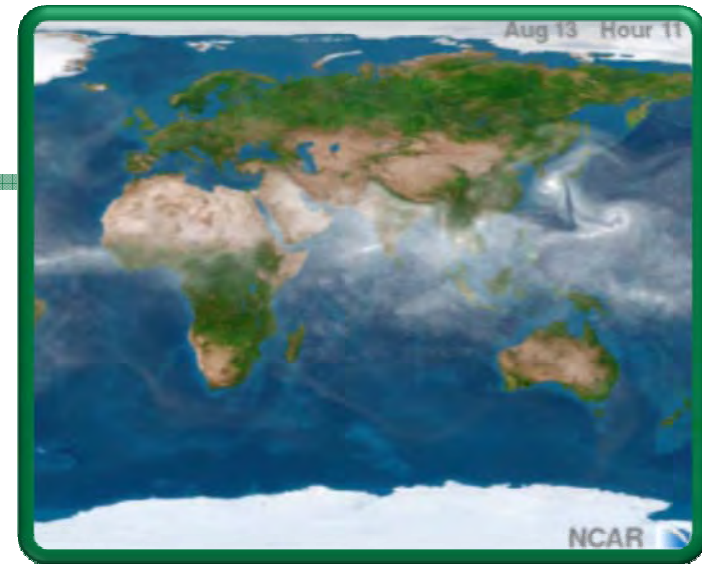
Allocation: 32,000,000 hrs

“[First principles simulation of liquid water] will have breakthrough impacts on atomistic computational biology research . . . ” – David Ceperley

Explorations of decadal predictive skill using the community climate system model

climate

This project will conduct an ensemble of ten high-resolution, retrospective, multi-decadal climate simulations from 1960 to present day. The experimental framework will exploit observations of the climate system's response to volcanic eruptions. These simulations will also enhance researchers' ability to evaluate anthropogenic greenhouse gas and aerosol impacts on climate on regional space scales.



PI: Kate Evans, ORNL

Code: CCSM

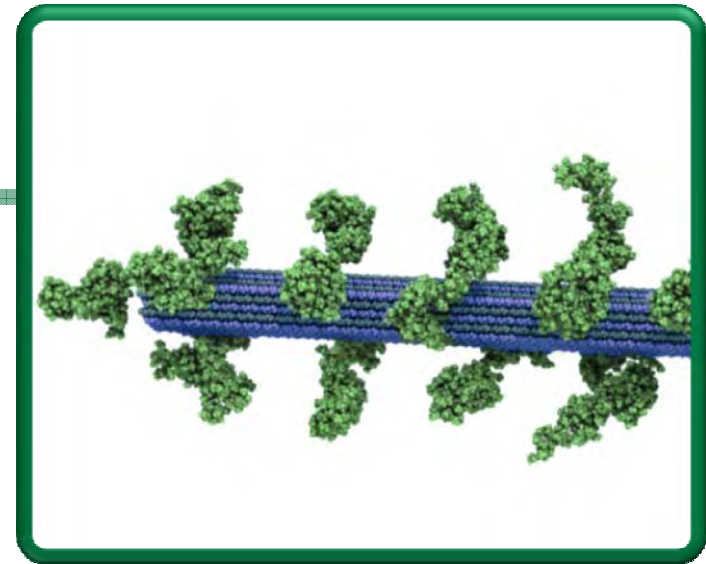
Allocation: 30,000,000 hrs

Jaguar's capabilities provide unprecedented resolution and long simulation periods, which allows for prediction of climate events on a regional and decadal scale. - Kate Evans

Cellulosic ethanol: A simulation model of lignocellulosic biomass deconstruction

biology

This project, a massive molecular-level simulation of the woody plant material used to produce ethanol, promises to reveal the factors that complicate the conversion of fibrous cellulose into fermentable glucose. In doing so, it will open the door to more efficient and less costly production of this relatively clean, renewable fuel.



PI: Jeremy Smith, UT

Code: GROMACS

Allocation: 25,500,000 hrs

On a 5.4 million atom system, Jaguar delivers more than 100 nanoseconds of simulated time per day. This is phenomenal productivity. - Jeremy Smith

Enabling breakthrough science

"But one thing is clear from an operational point of view, the Cray XT admin. team did a superb job, and the machine held up very well as we completed the [Q4 Joule] runs on the full system."

- Thomas Schulthess, ORNL

"These are large, demanding jobs. Our resource needs will only increase as we continue our transformation from code development to science."

- Mike Zingale, Stony Brook University

"The queuing policy in Jaguar allows our large simulations to run in a timely manner. The analysis capabilities have also been very valuable. [The NCCS liaison] has been extremely helpful."

-Jeremy Smith, University of TN/ORNL

These are one-of-a kind simulations. There is nothing like them in the literature to date in terms of breadth and methodology."

-Chris Mundy, PNNL

"It's amazing that a code this complex could be ported to such a large system with so little effort."

-Laura Carrington, San Diego Supercomputer Center

"The support of the User Assistance Center is highly beneficial to our project. As an example, the support staff has allowed us to use queues that make the completion of our downscaling simulation much faster to produce results needed for a collaborative effort to assess climate change impacts."

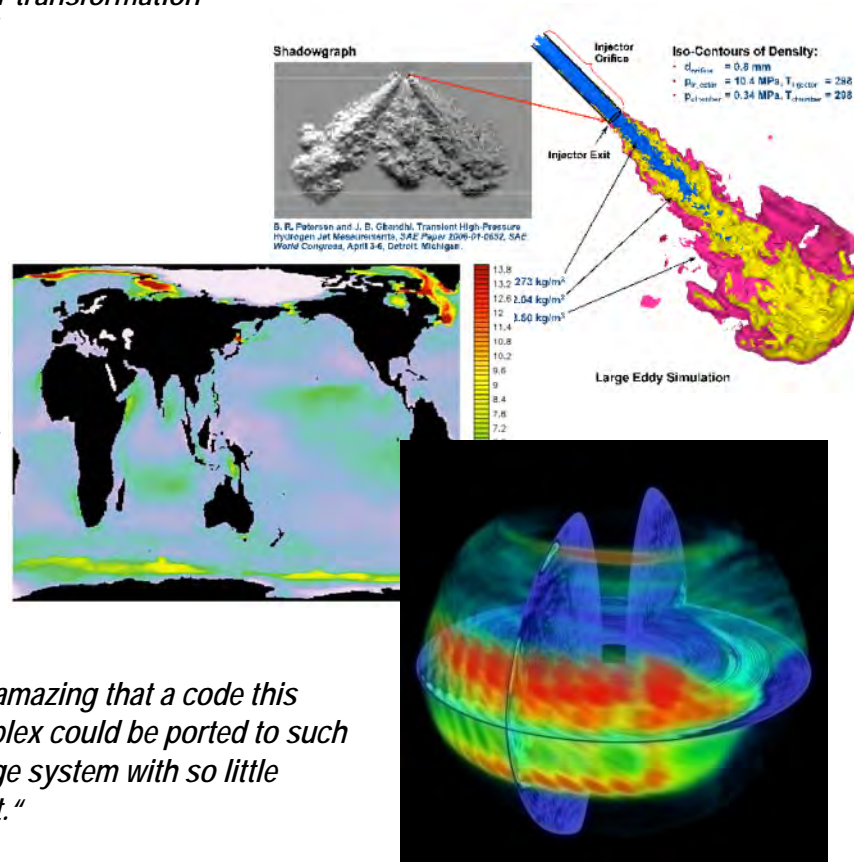
- The WRF team, Climate End Station

"In comparison with other systems, we consider the LCF as one of the best. Support was always both beneficial and fast."

- Hermann Fasel, University of Arizona

"Our NCCS staff liaison has been extremely helpful and often proactive in helping us. When asked technical questions, he has always been very responsive. In fact, in all our interactions with NCCS staff, we have found them to be professional and courteous. We greatly appreciate their dedication to us."

- Robert Sugar, University of California-Santa Barbara



DOE's broad range of science challenges demand a balanced, scalable, general purpose system



Projects	2006	2007	2008	2009
Accelerator physics	1	1	1	1
Astrophysics	3	4	5	5
Chemistry	1	1	2	4
Climate change	3	3	4	5
Combustion	1	1	2	2
Computer science	1	1	1	1
Fluid Dynamics			1	1
Fusion	4	5	3	5
Geosciences		1	1	1
High energy physics		1	1	
Life sciences	2	2	2	4
Materials science	2	3	3	4
Nuclear physics	2	2	1	2
Industry	2	3	3	3
Total Projects:	22	28	30	38
CPU Hours (Millions):	36	75.5	145	470

Jaguar: World's most powerful computer Designed for science from the ground up

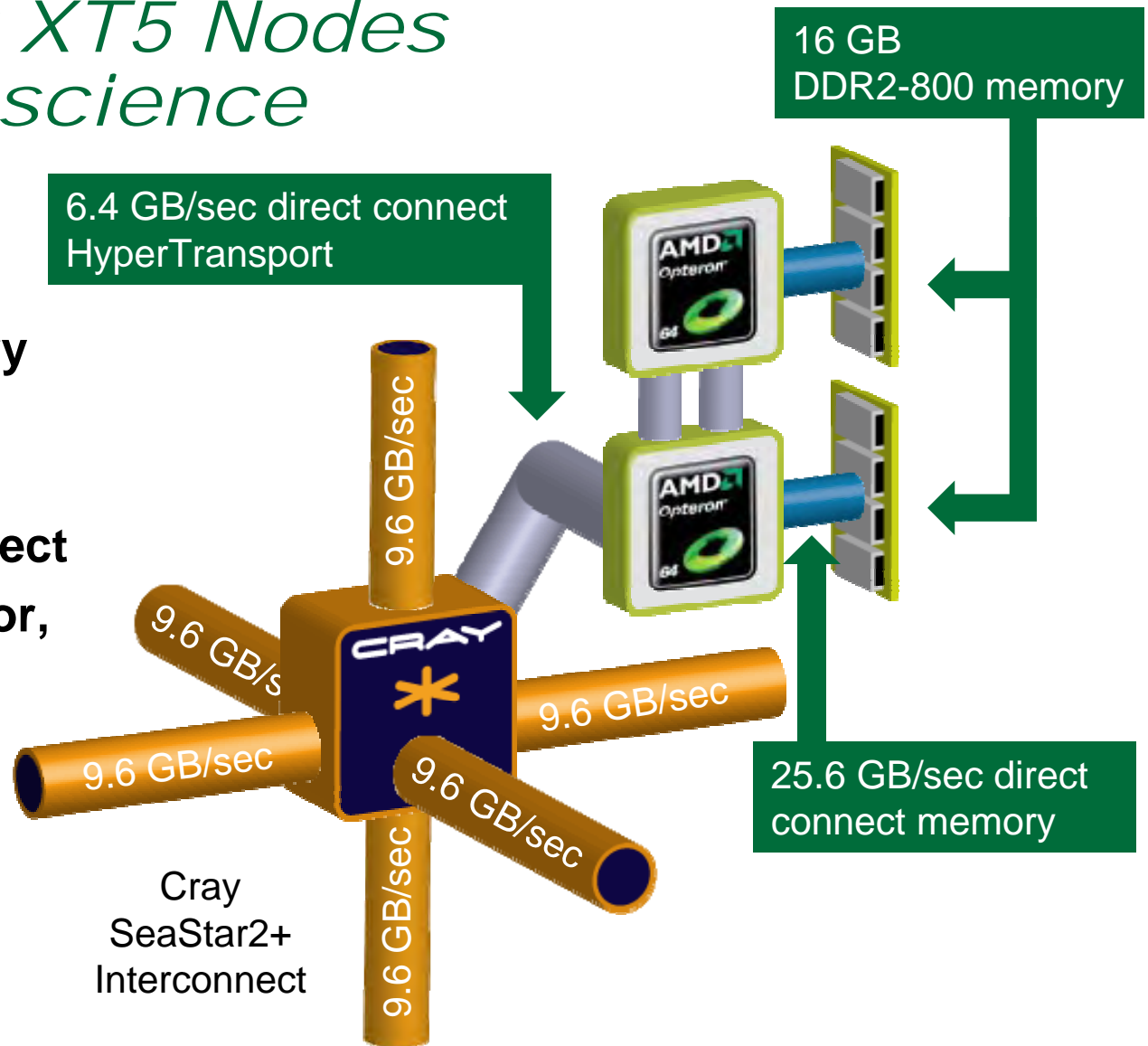


Peak performance	1.645 petaflops
System memory	362 terabytes
Disk space	10.7 petabytes
Disk bandwidth	200+ gigabytes/second

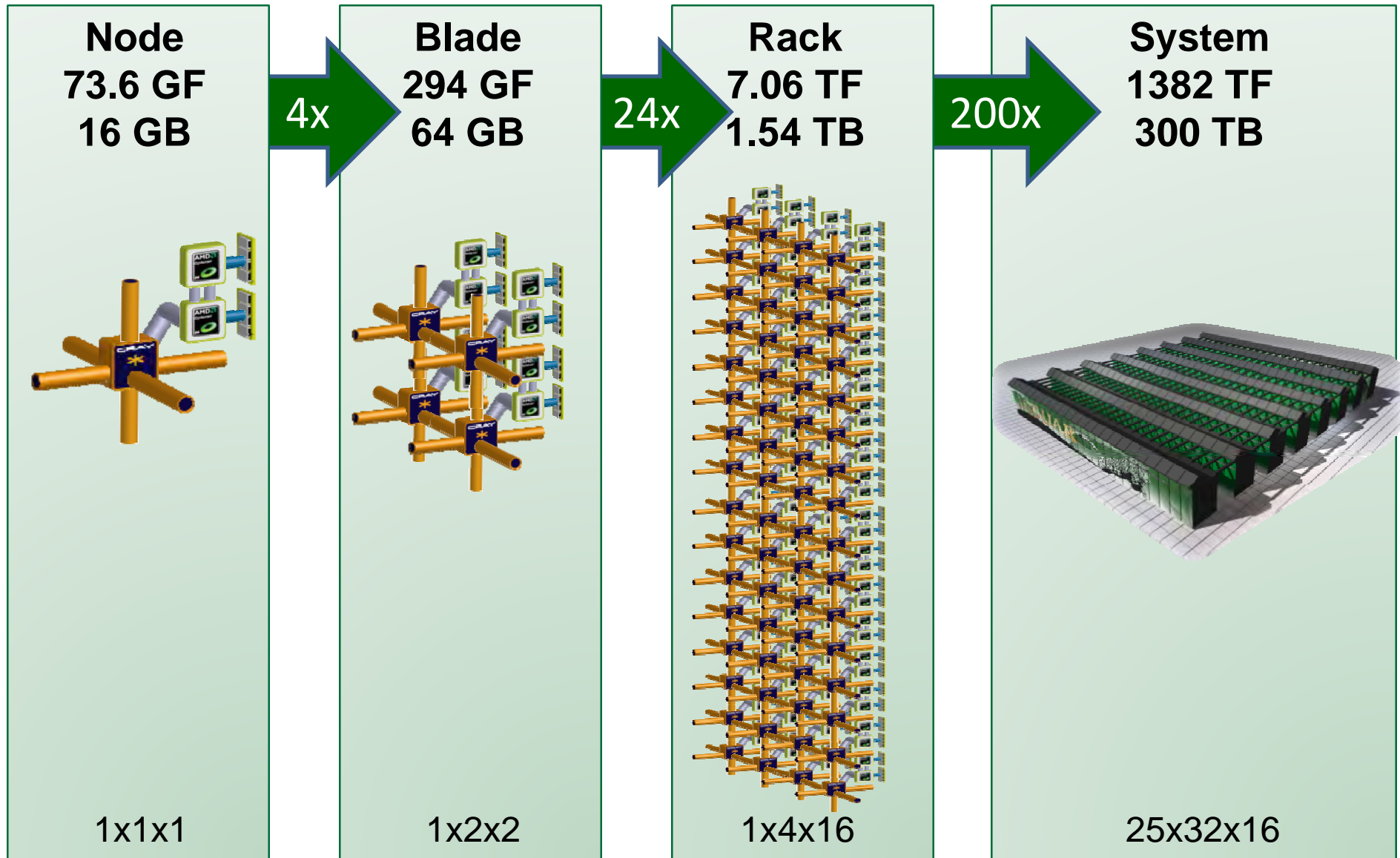
Jaguar's Cray XT5 Nodes Designed for science

- **Powerful node improves scalability**
- **Large shared memory**
- **OpenMP Support**
- **Low latency, High bandwidth interconnect**
- **Upgradable processor, memory, and interconnect**

GFLOPS	76.3
Memory (GB)	16
Cores	8
SeaStar2+	1



Building the Cray XT5 System



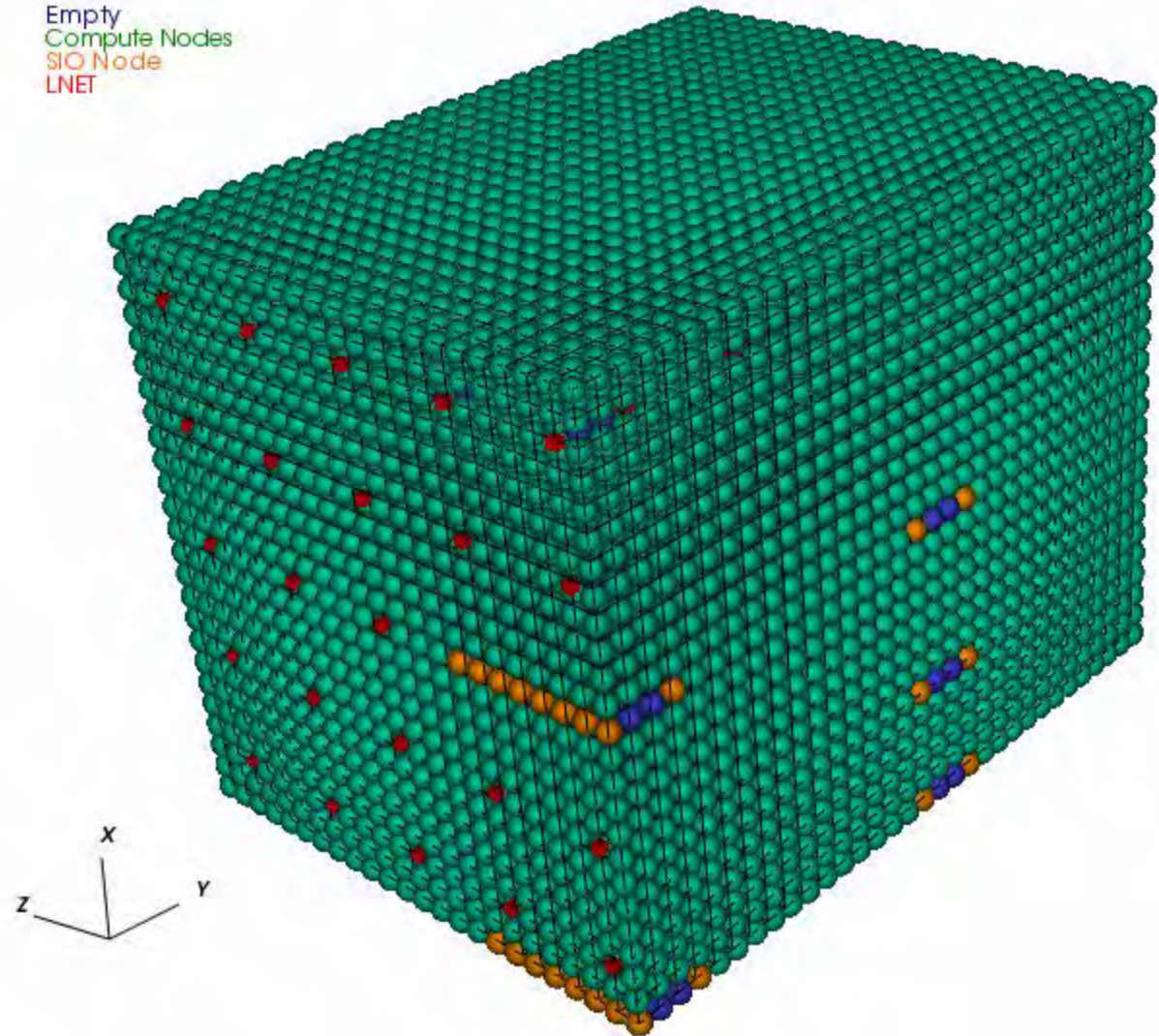
XT5 I/O Configuration Driven by application needs

XT5 Topology

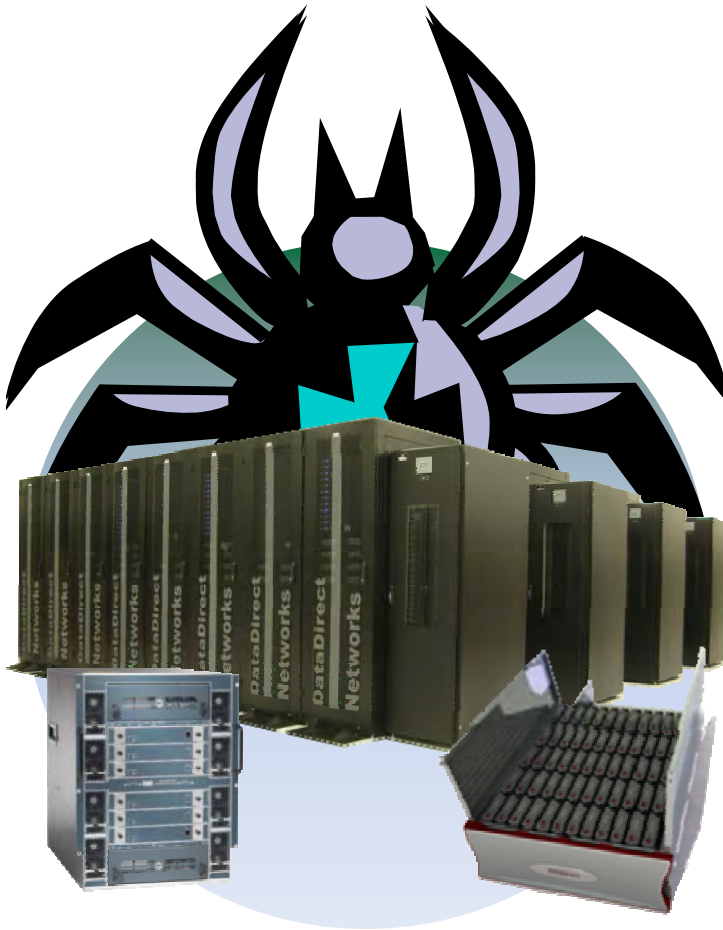
Features of I/O nodes

- 192 I/O nodes
- Each connected via non-blocking 4x DDR Infiniband to Lustre Object Storage Servers
- Fabric connections provides redundant paths
- Each OSS provide 1.25 GB/s
- I/O nodes spread throughout the 3-D torus to prevent hot-spots

Empty
Compute Nodes
I/O Node
LNET

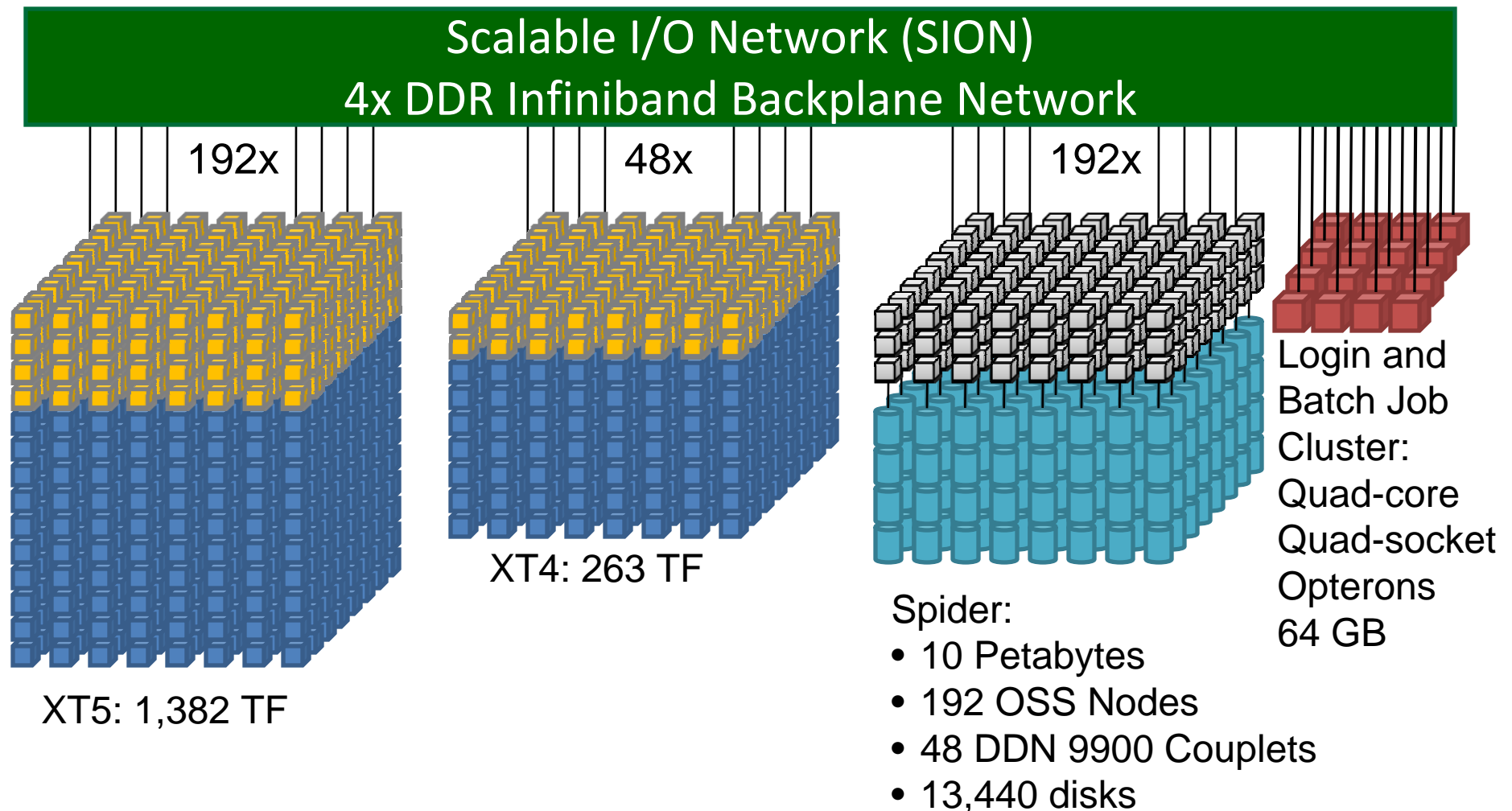


Center-wide File System

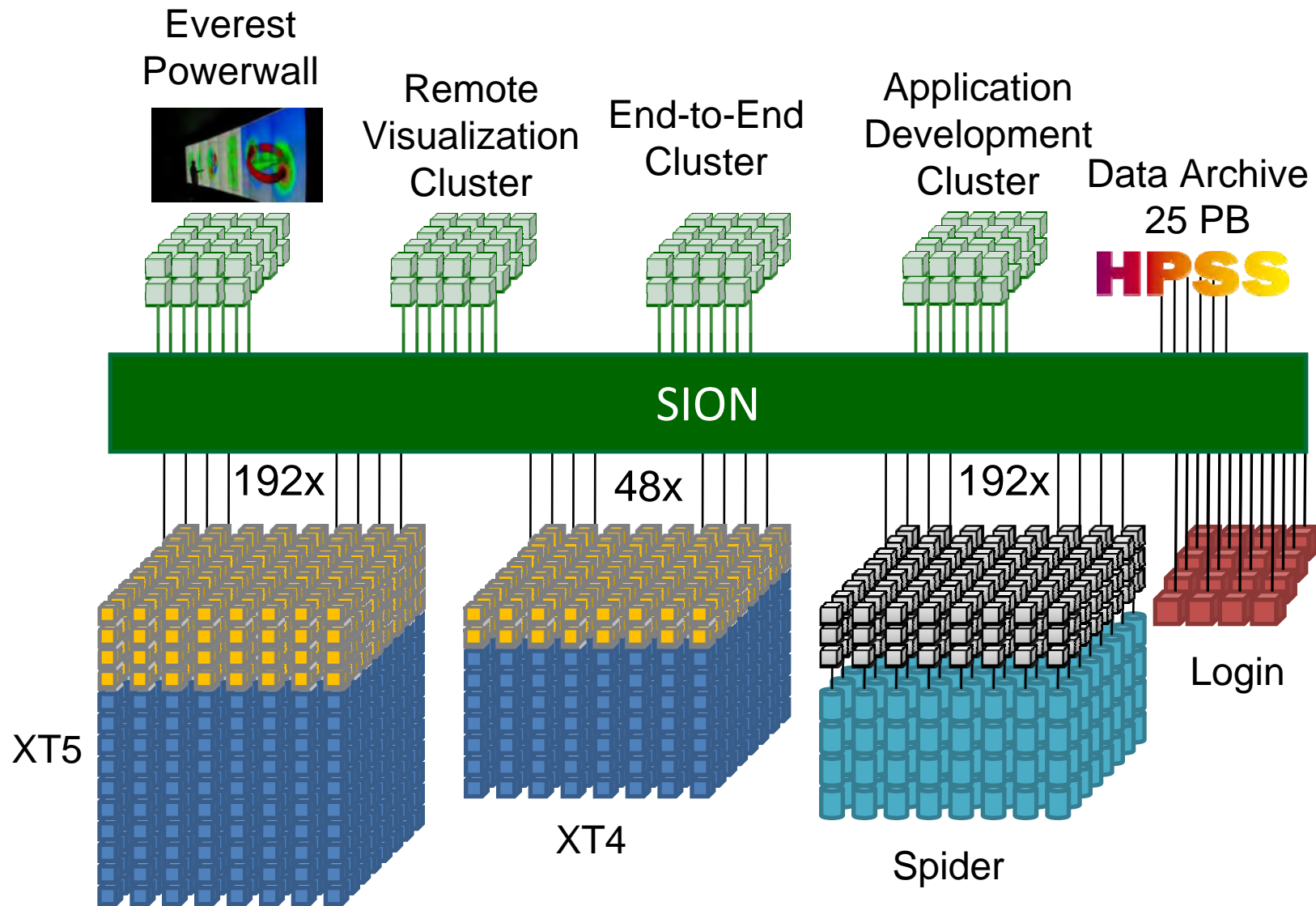


- “Spider” will provide a shared, parallel file system for all systems
 - Based on Lustre file system
- Demonstrated bandwidth of over 200 GB/s
- Over 10 PB of RAID-6 Capacity
 - 13,440 1-TB SATA Drives
- 192 Storage servers
 - 3 TB of memory
- Available from all systems via our high-performance scalable I/O network
 - Over 3,000 InfiniBand ports
 - Over 3 miles of cables
 - Scales as storage grows
- Undergoing system checkout with deployment expected in summer 2009

Combine the XT5, XT4, and Spider with a Login Cluster to complete Jaguar



Completing the Simulation Environment to meet the science requirements



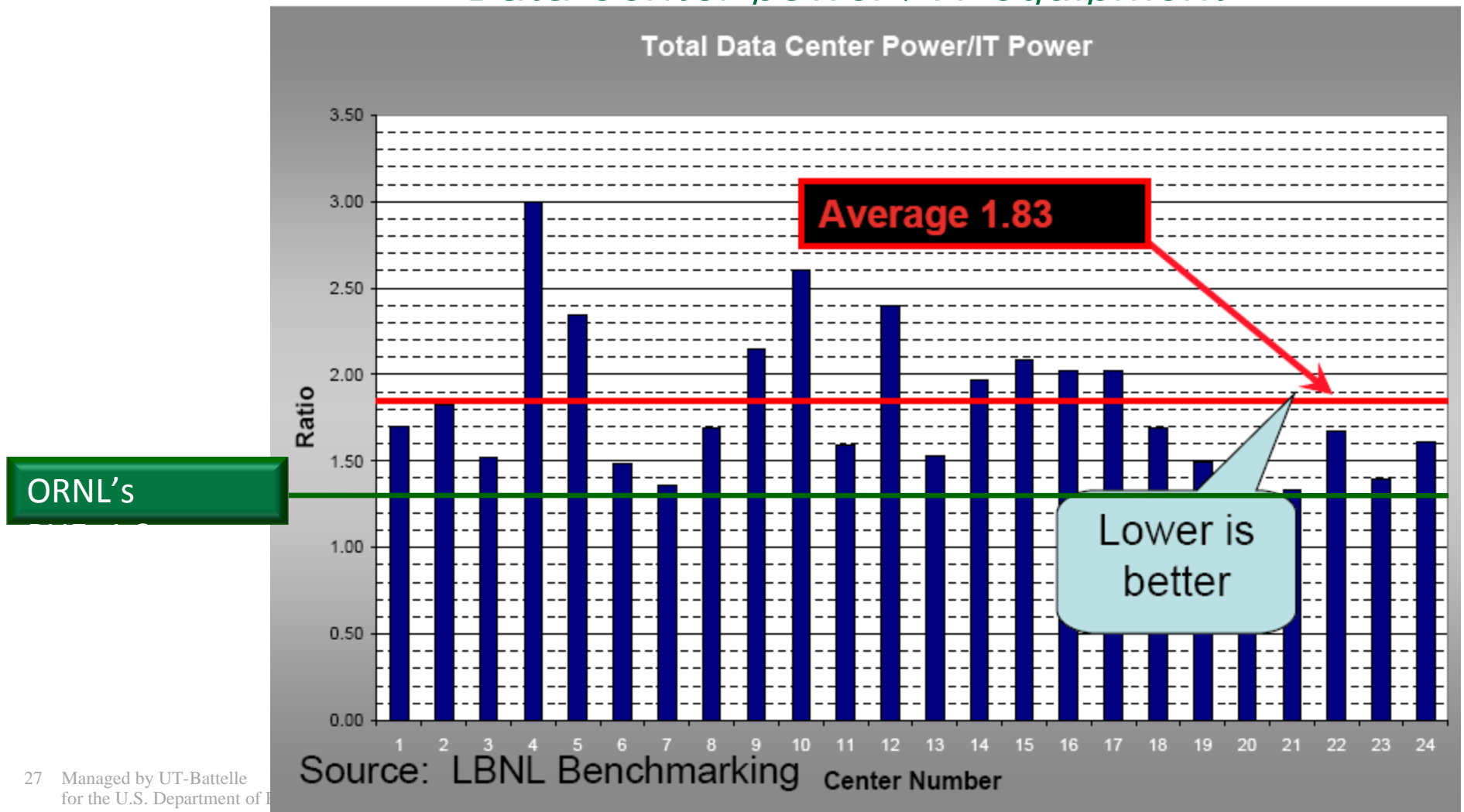
In 2003 we built a facility for Leadership Computing

- Space and power:
 - 40,000 ft² computer center with 36-in. raised floor, 18 ft. deck-to-deck
 - 8 MW of power (expandable)
- Office space for 400 staff
- Classroom and training areas for users
- New state-of-the-art visualization facility
- Separate lab areas for computer science and network research



Today, ORNL's facility is among the most efficient data centers

*Power Utilization Efficiency (PUE) =
Data Center power / IT equipment*



Electrical Systems Designed for efficiency

**13,800 volt power into the building saves
on transmission losses**



**480 volt power to cabinets saves \$1M in
installation costs**



**High efficiency power supplies in the
cabinets**

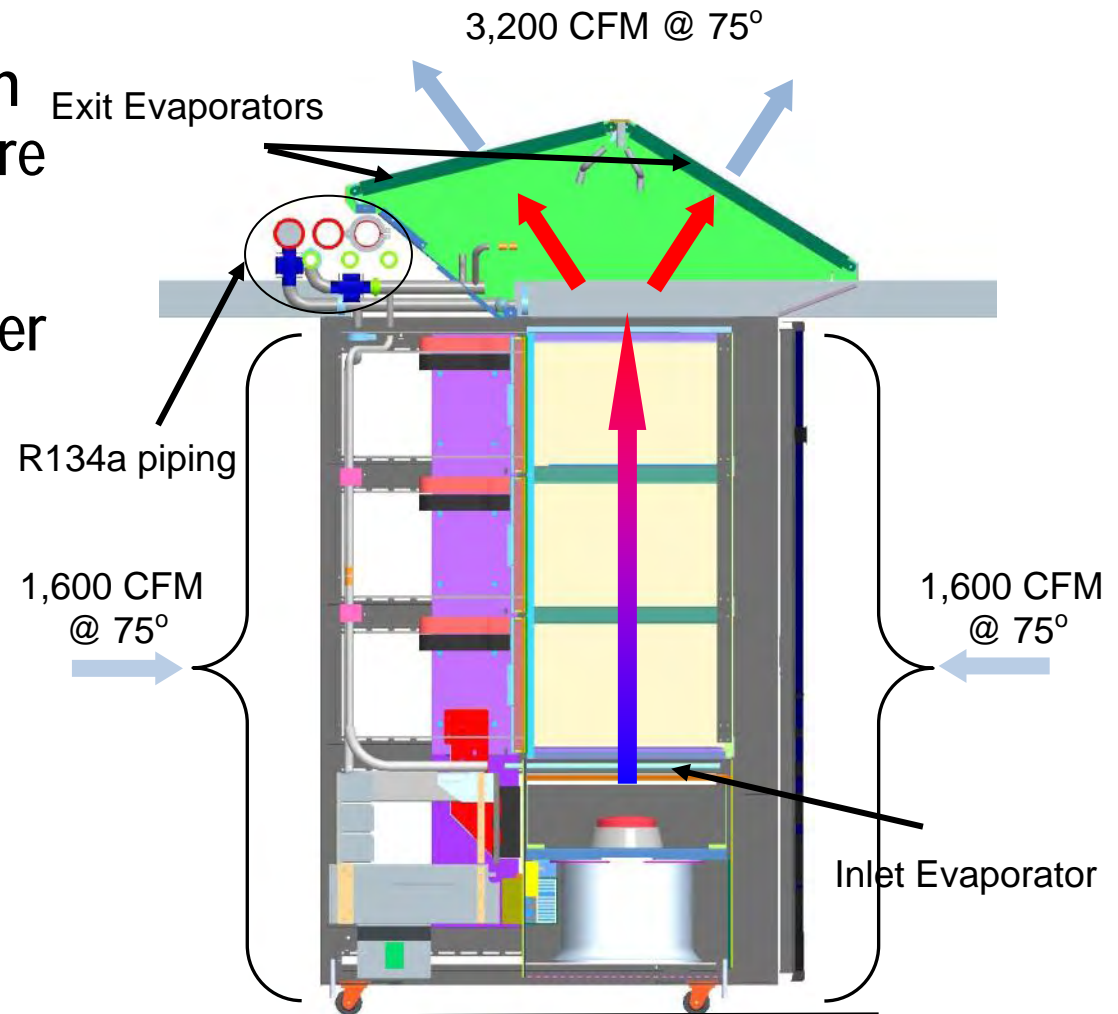


**Flywheel based UPS for highest
efficiency**



High Efficiency Liquid Cooling Required to build such a large system

- Newer Liquid Cooled design removes heat to liquid before it leaves the cabinet
- Saves about 900KW of power just in air movement and 2,500 ft² of floor space
- Phase change liquid to gas removes heat much more efficiently
- Each XDP heat exchanger replaces 2.5 CRAC units using less power and one-tenth the floor space



Timeline and Lessons Learned

- 7/30/2008: First cabinet arrived
- 9/17/2008: 200th cabinet arrived
- 9/29/2008: First full machine application
- 11/17/2008: Completed benchmarks and checkout
- 12/1/2008: Started acceptance testing
- 12/23/2008: Completed acceptance test

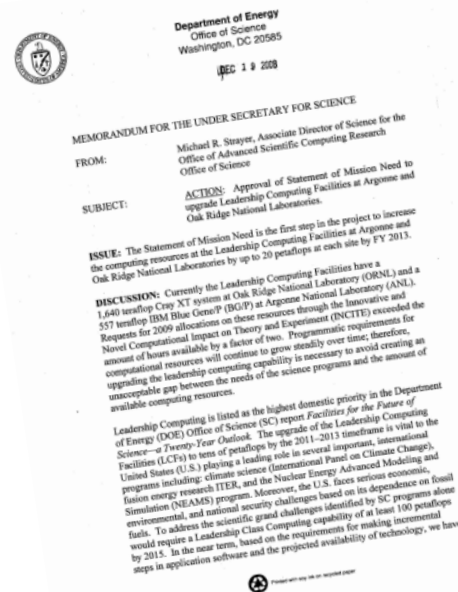
- Design was driven by science requirements. This was our roadmap.
- Had a detailed plan. These are big projects and project management is critical.
- When the inevitable problems occur, go back to the science requirements.
- We started preparing the applications on the XT4 in the spring & summer
- The liaison model of supporting the INCITE projects is the right model
 - Assign a computational scientist to each project
 - Begin early to get the apps scaled
 - Be ready to go as soon as the machine is delivered
- Applications are the hard part. Keep a consistent architecture for as long as possible.

DOE-SC and ASCR are leading in modeling and simulation for critical international problems

- ASCR sweep of major awards at SC'09
 - All four HPC Challenge awards
 - Both Gordon Bell awards
- Important, time-critical problems in Energy Assurance, Climate Change, Materials Science, and Basic Science are running today solving problems that were unapproachable just a few years ago
- Our strategy of early engagement with the vendors, working closely with the science teams, and continuous improvement of the systems and facilities is the right approach to computational science at the cutting edge

Where do we go from here?

- Mission Need Statement (CD-0) signed by Dr. Raymond Orbach
- Upgrades Leadership Computing Facilities to 20–40 petaflops in 2011–2013
- Required for mission of DOE-SC
 - Climate change
 - Energy assurance
 - Basic science
- Calls for 2 systems to provide capability and provide architectural diversity



How Will We Change The World?

“We finally have a true leadership computer that enables us to run calculations impossible anywhere else in the world. The huge memory and raw compute power of Jaguar combine to transform the scale of computational chemistry.

Now that we have NWChem and MADNESS running robustly at the petascale, we are unleashing a flood of chemistry calculations that will produce insights into energy storage, catalysis, and functionalized nano-scale systems.”



*Robert Harrison
ORNL and University of Tennessee*

Questions?



**We are only beginning to realize
the transforming power of computing
as an enabler of discovery and innovation**