High Performance Computing for Breakthrough Science at ORNL

Presented at the

ORAU Annual Council Meeting

James J. Hack, Director
National Center for
Computational Sciences

25 February 2009

Managed by UT-Battelle
for the Department of Energy

ORNL is the U.S. Department of Energy’s largest science and energy laboratory

- $1.3B budget
- 4,250 employees
- 3,900 research guests annually
- $350 million invested in modernization

- World’s most powerful computing facility
- Nation’s largest concentration of open source materials research

- Nation’s most diverse energy portfolio
- The $1.4B Spallation Neutron Source in operation
- Managing the billion-dollar U.S. ITER project

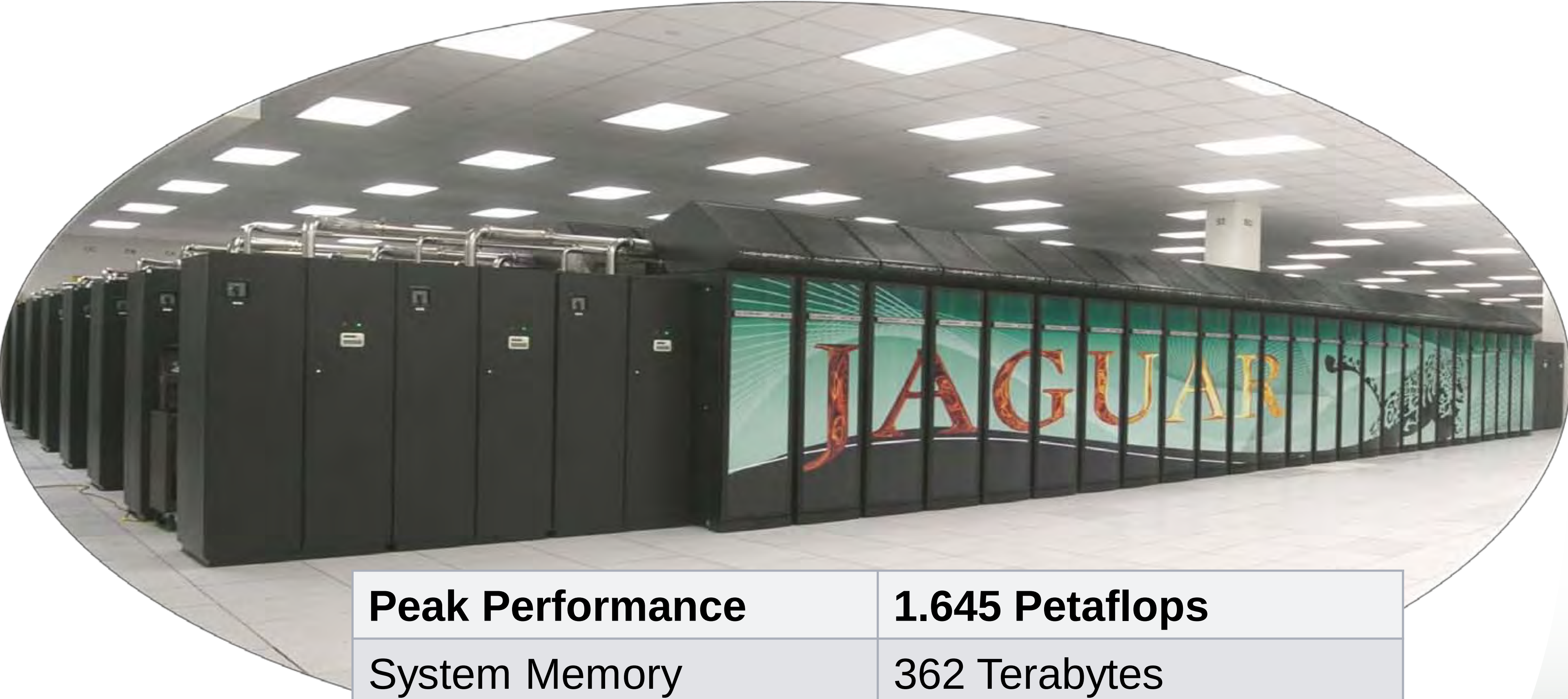
National Center for Computational Sciences
Oak Ridge National Laboratory

Mission: Deploy and operate the computational resources required to tackle global challenges

- Providing world-leading computational resources and specialized services for the most computationally intensive problems
- Providing stable hardware/software path of increasing scale to maximize productive applications development
- Deliver transforming discoveries in materials, biology, climate, energy technologies, etc.
- Ability to investigate otherwise inaccessible systems, from supernovae to energy grid dynamics

Jaguar – 1.64 PF Cray XT: 45,376 Quad-Core Processors, 362 TB memory

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	240+ Gigabytes/second
Interconnect Bandwidth	532 Terabytes/second

High Performance Linpack Benchmark

- **#2 on November 2008 list**
- **1059 TeraFLOPS (76.7% of peak)**
- **Ran on 150,152 cores**
- **Largest HPL run ever, by a huge margin**
- **Ran on the XT5 portion of the machine *41 days after delivery of a 200 cabinet system!!!***
- ***Ran for 18.3 hours without a failure!!!!***

```
T/V                N    NB     P     Q               Time                 Gflops
--------------------------------------------------------------------------------
WR03R3C1     4712799   200   274   548           65884.80              1.059e+06
--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV--VVV-
Max aggregated wall time rfact . . . :              13.67
+ Max aggregated wall time pfact . . :              10.99
+ Max aggregated wall time mxswp . . :              10.84
Max aggregated wall time pbcast  . . :            6131.91
Max aggregated wall time update  . . :           63744.72
+ Max aggregated wall time laswp . . :            7431.52
Max aggregated wall time up tr sv  . :              16.98
------------------------------------------------------------------------------------
||Ax-b||_oo/(eps*(||A||_oo*||x||_oo+||b||_oo)*N)=        0.0006162 ...... PASSED
================================================================================
```

HPC Challenge Benchmarks

Four “Class 1” benchmarks:

HPL	**902 TFLOPS**	**#1**
G-Streams	**330**	**#1**
G-Random Access	**16.6 GUPS**	**#1 Baseline**
G-FFTE	**2773**	**#3**

Working on further optimizations, but just ran out of time.

A balanced, high performance supercomputer.

What does the system look like?

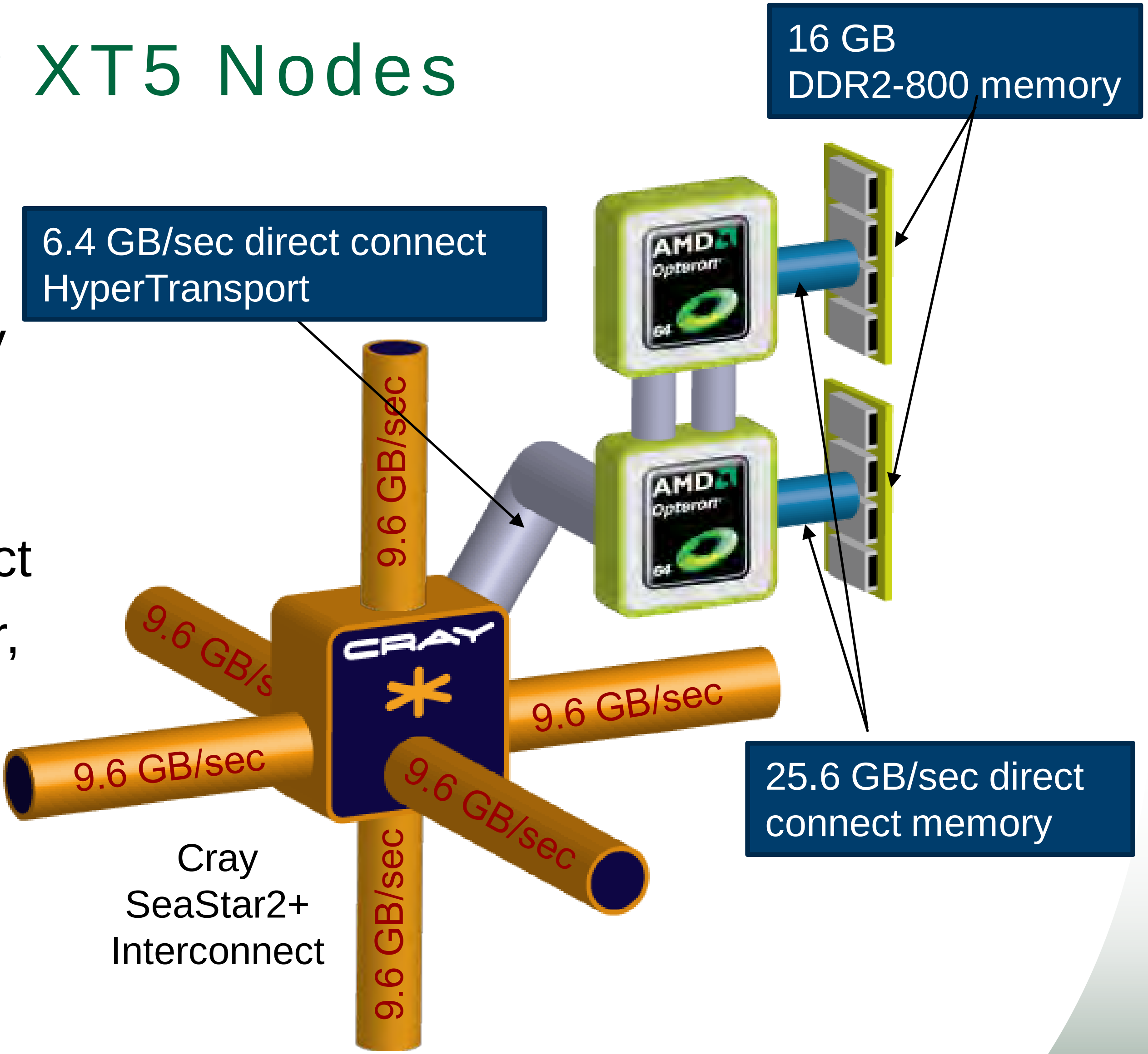
Jaguar combines the existing 263 TF Cray XT4 system at ORNL's NCCS with a new 1,382 TF Cray XT5 to create a 1.64 PF system

System attribute	XT5	XT4
Quad-core AMD Opteron™ Processors	37,544	7,832
Node Architecture	Dual socket SMP	Single Socket
Memory per core / Node (GB)	2 / 16	2 / 8
Total System Memory (TB)	300	62
Disk Capacity (TB)	10,000	750
Disk Bandwidth (GB/s)	240	44
Interconnect	SeaStar2+ 3-D Torus	SeaStar2+ 3-D Torus

Jaguar's Cray XT5 Nodes

- 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



16 GB
DDR2-800 memory

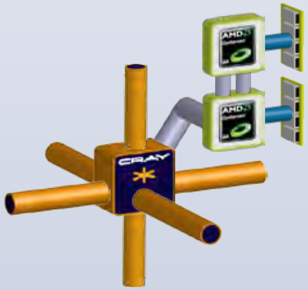
6.4 GB/sec direct connect HyperTransport

25.6 GB/sec direct connect memory

Cray
SeaStar2+
Interconnect

Building the Cray XT5 System

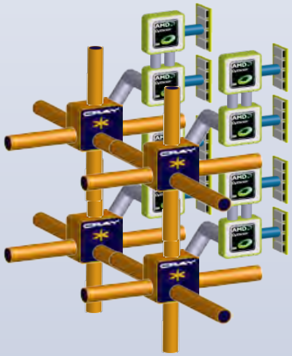
Node
73.6 GF
16 GB



1x1x1

4x

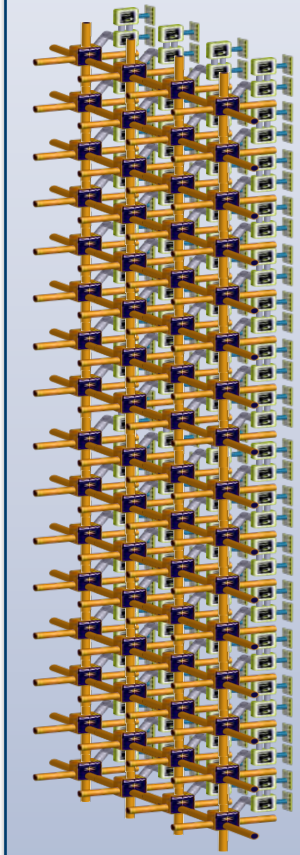
Blade
294 GF
64 GB



1x2x2

24x

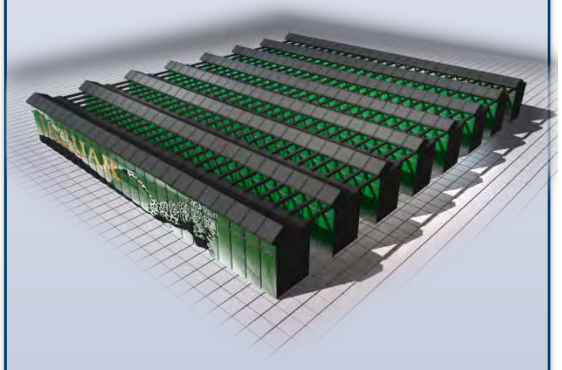
Rack
7.06 TF
1.54 TB



1x4x16

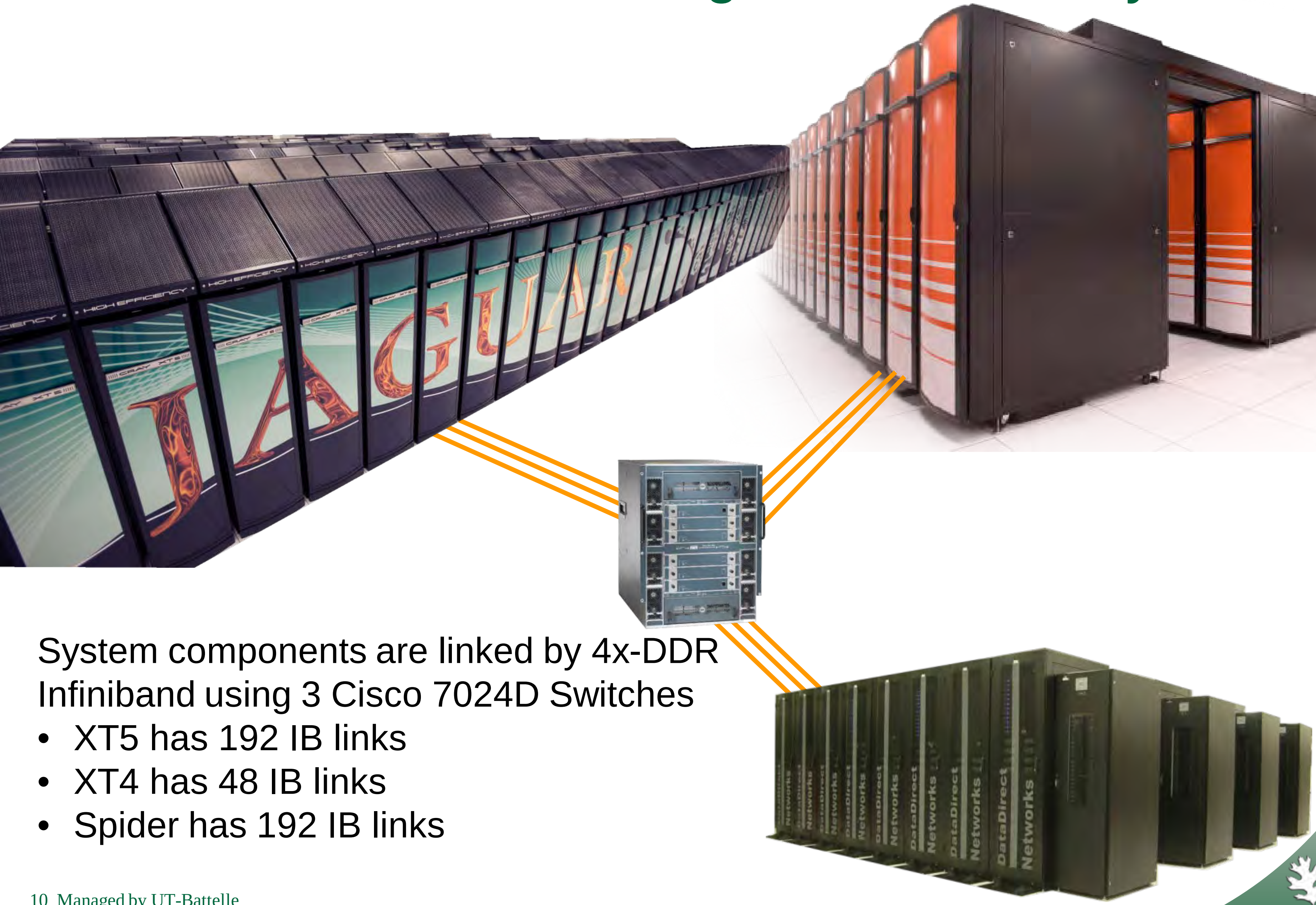
200x

System
1382 TF
300 TB



25x32x16

Jaguar combines a new 1.38 PF Cray XT5 with the existing 263 TF Cray XT4



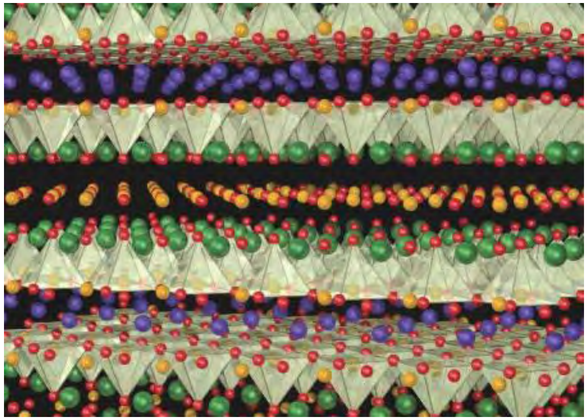
System components are linked by 4x-DDR Infiniband using 3 Cisco 7024D Switches

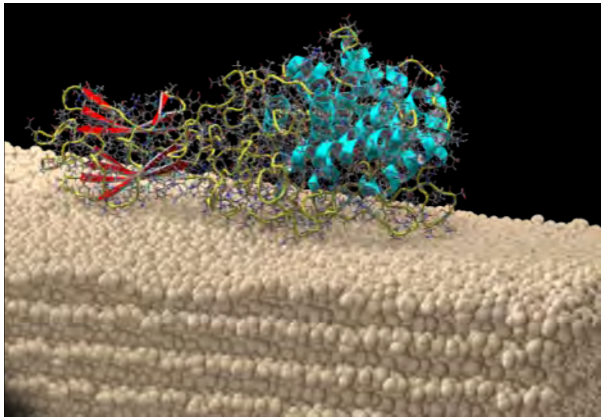
- XT5 has 192 IB links
- XT4 has 48 IB links
- Spider has 192 IB links

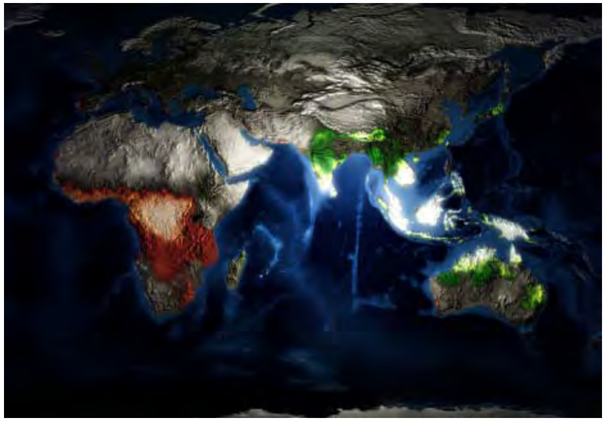
Storage for an avalanche of data

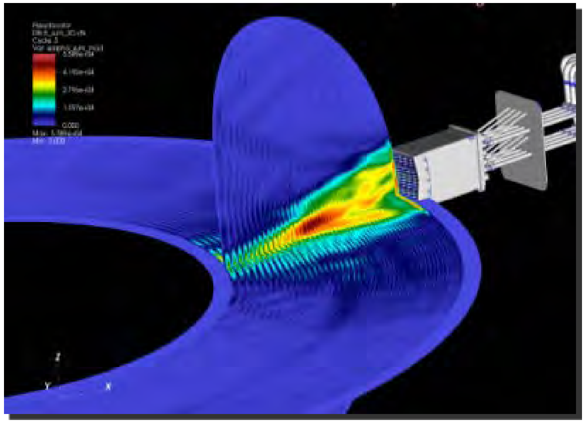
- “Spider” is being installed to provide a shared, parallel file system for all systems
  - Based on Lustre file system
- Bandwidth of over 240 GB/s
- Over 10 PB of RAID6 Capacity
  - 13,440 1 TB SATA Drives
- 192 Storage servers
  - 3 TeraBytes of memory
  - 14 TeraFlops
- Available from all systems via our high-performance scalable I/O network
  - Over 3,000 InfiniBand ports
  - Over 3 miles of cables
  - Scales as storage grows
- Engineered for high availability

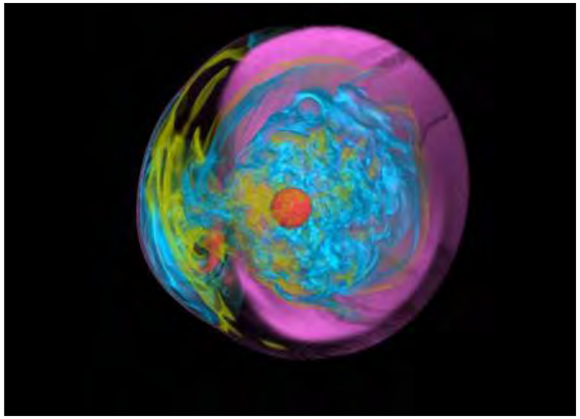
We are advancing scientific discovery

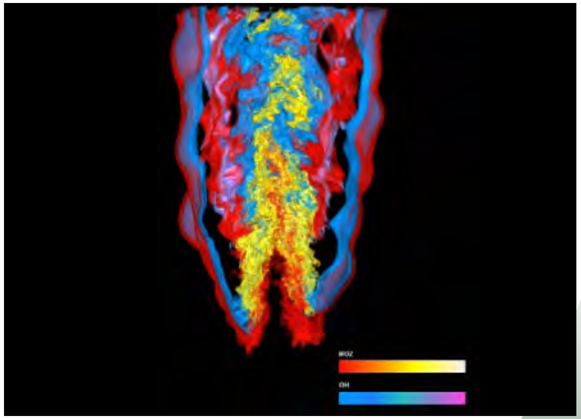

Resolved decades-long controversy about modeling physics of high temperature superconducting cuprates


New insights into protein structure and function leading to better understanding of cellulose-to-ethanol conversion

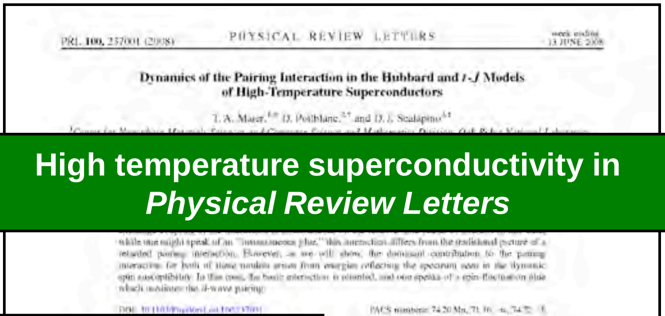

Addition of vegetation models in climate code for global, dynamic CO_2 exploration


First fully 3D plasma simulations shed new light on engineering superheated ionic gas in ITER


Fundamental instability of supernova shocks discovered directly through simulation


First 3-D simulation of flame that resolves chemical composition, temperature, and flow

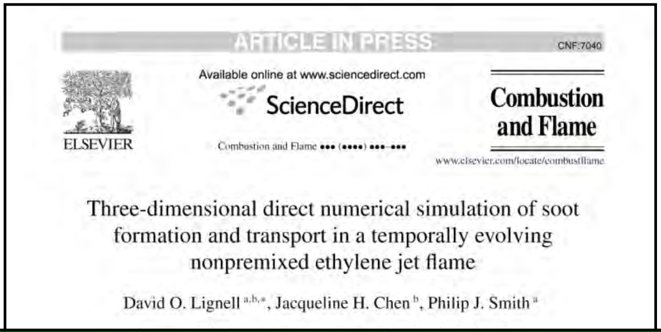
Recent and Highly Visible Science Output


PRL 100, 237001 (2008) PHYSICAL REVIEW LETTERS week ending 13 JUNE 2008

Dynamics of the Pairing Interaction in the Hubbard and t-J Models of High-Temperature Superconductors

T. A. Maier,[1,*] D. Poilblanc,[2,†] and D. J. Scalapino[3,‡]

High temperature superconductivity in *Physical Review Letters*


ARTICLE IN PRESS CNF:7040

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Combustion and Flame ••• (••••) •••–•••

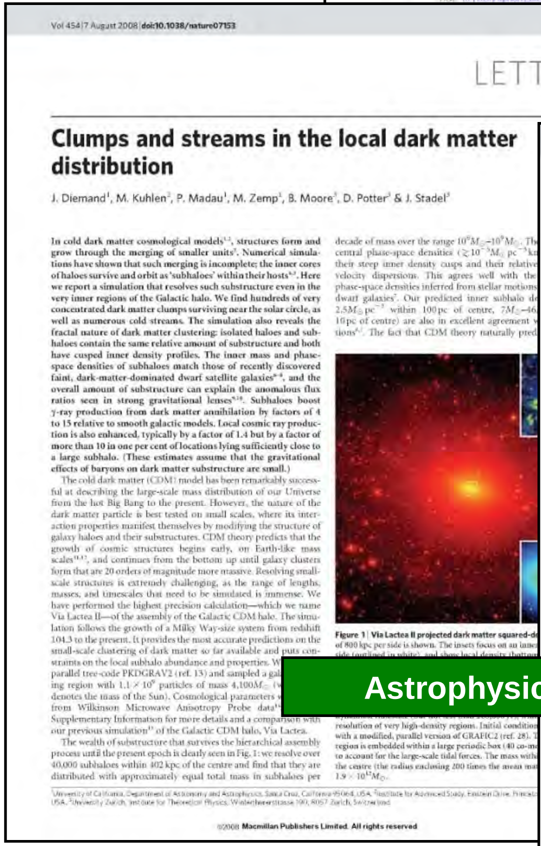
Combustion and Flame

www.elsevier.com/locate/combustflame

Three-dimensional direct numerical simulation of soot formation and transport in a temporally evolving nonpremixed ethylene jet flame

David O. Lignell[a,b,*], Jacqueline H. Chen[b], Philip J. Smith[a]

Combustion in *Combustion and Flame*


Vol 454 | 7 August 2008 | doi:10.1038/nature07153 nature

LETTERS

Clumps and streams in the local dark matter distribution

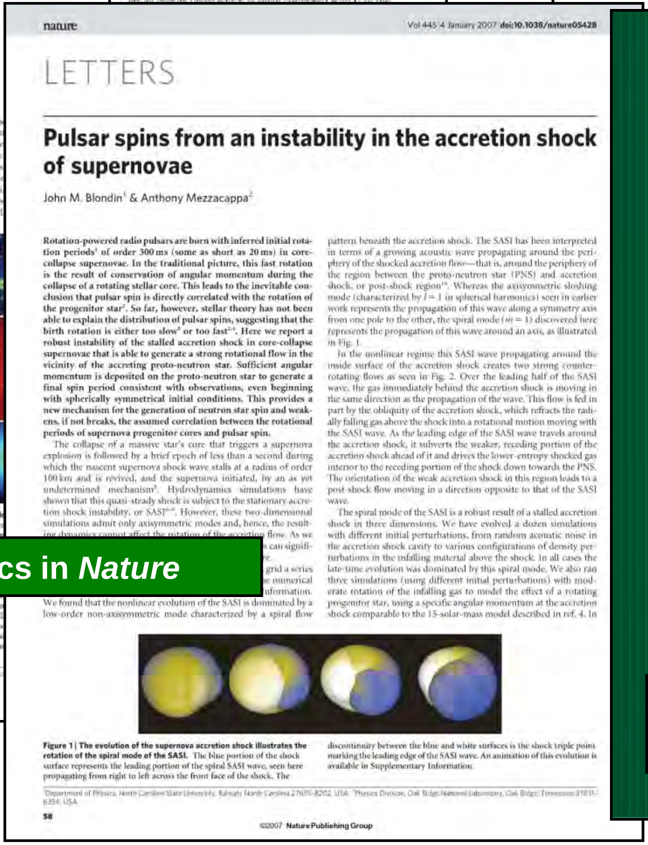
J. Diemand[1], M. Kuhlen[2], P. Madau[1], M. Zemp[1], B. Moore[3], D. Potter[3] & J. Stadel[3]

In cold dark matter cosmological models[1,2], structures form and grow through the merging of smaller units[3]. Numerical simulations have shown that such merging is incomplete; the inner cores of haloes survive and orbit as 'subhaloes' within their hosts[4,5]. Here we report a simulation that resolves such substructure even in the very inner regions of the Galactic halo. We find hundreds of very concentrated dark matter clumps surviving near the solar circle, as well as numerous cold streams. The simulation also reveals the fractal nature of dark matter clustering: isolated haloes and subhaloes contain the same relative amount of substructure and both have cusped inner density profiles. The inner mass and phase-space densities of subhaloes match those of recently discovered faint, dark-matter-dominated dwarf satellite galaxies[6-8], and the overall amount of substructure can explain the anomalous flux ratios seen in strong gravitational lenses[9,10]. Subhaloes boost γ-ray production from dark matter annihilation by factors of 4 to 15 relative to smooth galactic models. Local cosmic ray production is also enhanced, typically by a factor of 1.4 but by a factor of more than 10 in one per cent of locations lying sufficiently close to a large subhalo. (These estimates assume that the gravitational effects of baryons on dark matter substructure are small.)

The cold dark matter (CDM) model has been remarkably successful at describing the large-scale mass distribution of our Universe from the hot Big Bang to the present. However, the nature of the dark matter particle is best tested on small scales, where its interaction properties manifest themselves by modifying the structure of galaxy haloes and their substructures. CDM theory predicts that the growth of cosmic structures begins early, on Earth-like mass scales

Figure 1 | Via Lactea II projected dark matter squared-density

©2008 Macmillan Publishers Limited. All rights reserved


nature Vol 445 | 4 January 2007 | doi:10.1038/nature05428

LETTERS

Pulsar spins from an instability in the accretion shock of supernovae

John M. Blondin[1] & Anthony Mezzacappa[2]

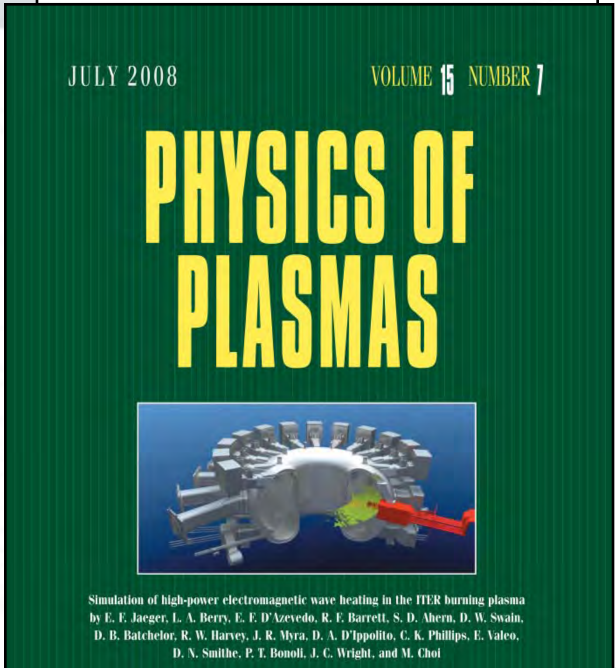
Rotation-powered radio pulsars are born with inferred initial rotation periods[1] of order 300 ms (some as short as 20 ms) in core-collapse supernovae. In the traditional picture, this fast rotation is the result of conservation of angular momentum during the collapse of a rotating stellar core. This leads to the inevitable conclusion that pulsar spin is directly correlated with the rotation of the progenitor star[2]. So far, however, stellar theory has not been able to explain the distribution of pulsar spins, suggesting that the birth rotation is either too slow[3] or too fast[2,4]. Here we report a robust instability of the stalled accretion shock in core-collapse supernovae that is able to generate a strong rotational flow in the vicinity of the accreting proto-neutron star. Sufficient angular momentum is deposited on the proto-neutron star to generate a final spin period consistent with observations, even beginning with spherically symmetrical initial conditions. This provides a new mechanism for the generation of neutron star spin and weakens, if not breaks, the assumed correlation between the rotational periods of supernova progenitor cores and pulsar spin.

The collapse of a massive star's core that triggers a supernova explosion is followed by a brief epoch of less than a second during which the nascent supernova shock wave stalls at a radius of order 100 km and is revived, and the supernova initiated, by an as yet undetermined mechanism[5]. Hydrodynamics simulations have shown that this quasi-steady shock is subject to the stationary accretion shock instability, or SASI[6-9].

Figure 1 | The evolution of the supernova accretion shock illustrates the rotation of the spiral mode of the SASI.

58 ©2007 Nature Publishing Group

Astrophysics in *Nature*


JULY 2008 VOLUME 15 NUMBER 7

PHYSICS OF PLASMAS

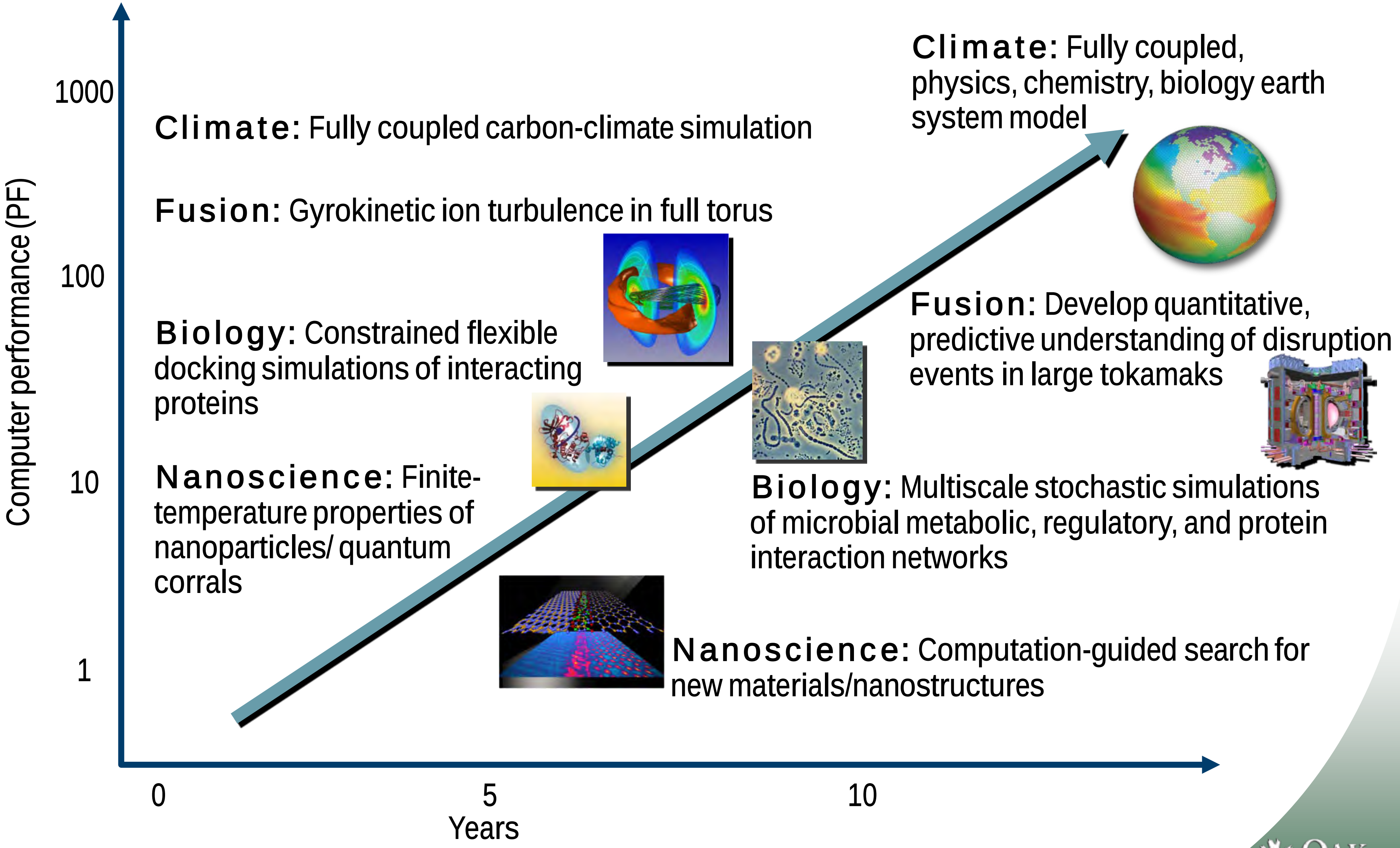
Simulation of high-power electromagnetic wave heating in the ITER burning plasma by E. F. Jaeger, L. A. Berry, E. F. D'Azevedo, R. F. Barrett, S. D. Ahern, D. W. Swain, D. B. Batchelor, R. W. Harvey, J. R. Myra, D. A. D'Ippolito, C. K. Phillips, E. Valeo, D. N. Smithe, P. T. Bonoli, J. C. Wright, and M. Choi

Fusion on the cover of *Physics of Plasmas*

Science Prospects and Benefits with High End Computing in the Next Decade

Opportunity	Key Application Areas	Goal and Benefit
Materials science	Nanoscale science, manufacturing, and material lifecycles, response and failure	Design, characterize, and manufacture materials, down to the nanoscale, tailored and optimized for specific applications
Earth science	Weather, carbon management, climate change mitigation and adaptation, environment	Understand the complex biogeochemical cycles that underpin global ecosystems and control the sustainability of life on Earth
Energy assurance	Fossil, fusion, combustion, nuclear fuel cycle, chemical catalysis, renewables (wind, solar, hydro), bioenergy, energy efficiency, power grid, transportation, buildings	Attain, without costly disruption, the energy required by the United States in guaranteed and economically viable ways to satisfy residential, commercial, and transportation requirements
Fundamental science	High energy physics, nuclear physics, astrophysics, accelerator physics	Decipher and comprehend the core laws governing the Universe and unravel its origins
Biology and medicine	Proteomics, drug design, systems biology	Understand connections from individual proteins through whole cells into ecosystems and environments
National security	Disaster management, homeland security, defense systems, public policy	Analyze, design, stress-test, and optimize critical systems such as communications, homeland security, and defense systems; understand and uncover human behavioral systems underlying asymmetric operation environments
Engineering design	Industrial and manufacturing processes	Design, deploy, and operate safe and economical structures, machines, processes, and systems with reduced concept-to-deployment time

Science advances for the next decade require leadership computing



Climate: Fully coupled carbon-climate simulation

Fusion: Gyrokinetic ion turbulence in full torus

Biology: Constrained flexible docking simulations of interacting proteins

Nanoscience: Finite-temperature properties of nanoparticles/ quantum corrals

Climate: Fully coupled, physics, chemistry, biology earth system model

Fusion: Develop quantitative, predictive understanding of disruption events in large tokamaks

Biology: Multiscale stochastic simulations of microbial metabolic, regulatory, and protein interaction networks

Nanoscience: Computation-guided search for new materials/nanostructures

We have a multi-agency strategy for sustained leadership in computational sciences

- **Provide the nation’s most powerful open resources for capability computing**
- **Follow a well-defined path for maintaining national leadership in this critical area**
- **Deliver cutting-edge science relevant to the missions of key federal agencies**
- **Synergy of requirements and technology**
- **Unique opportunity for multi-agency collaboration for science**

NSF

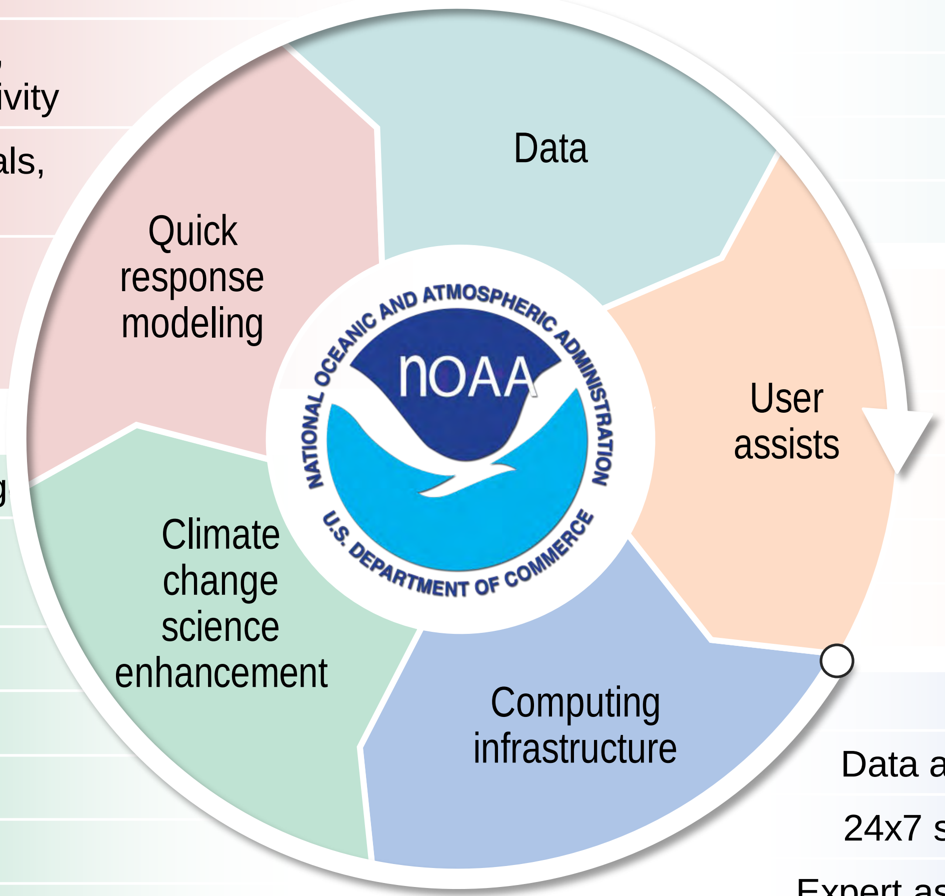
Department of Defense United States of America

NOAA National Oceanic and Atmospheric Administration U.S. Department of Commerce

We are positioning ourselves to be a strategic climate science partner with NOAA

Climate change	Scenario, topic, time period, location
Impacts	Topic, system, climate sensitivity
Adaptation	Topic, potentials, limits
Mitigation	Emissions, technologies, policies

Modeling and simulation	Climate chang
	Impacts, adaptation
	Mitigation
Improving decision-related science	Uncertainties
	Thresholds
	Feedbacks
	Model coupling
	Integrating observations/ experiments with models



Regional climate projections
Climate impacts
GHG emissions (CDIAC)
Adaptation possibilities
ARM archive

Visualization tools
Expert system shells
Semantic web searches
Simplified models
Climate End Station
Scientific liaison to project

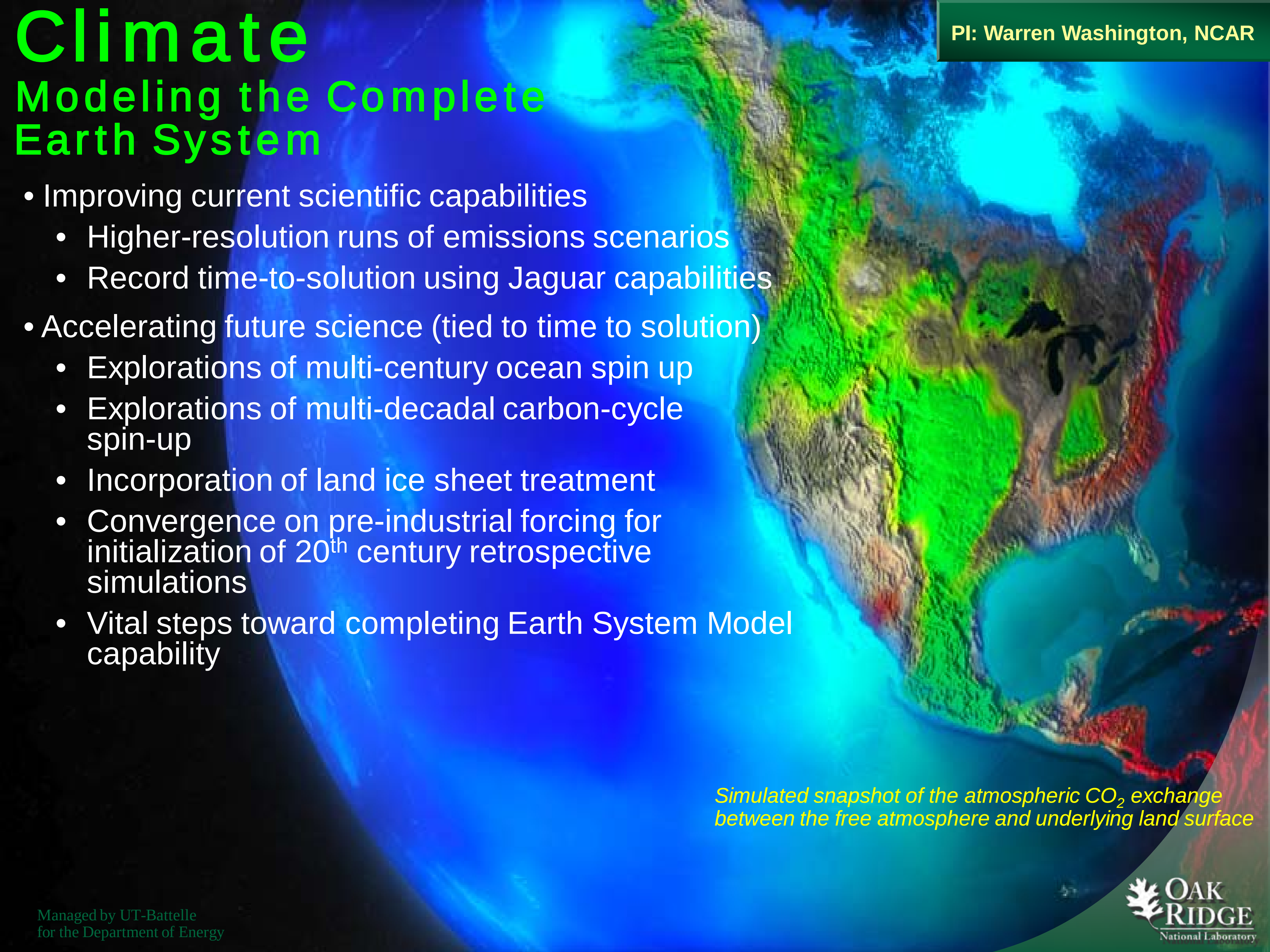
Petascale computers
Data analysis systems and tools
24x7 system and facility support
Expert assistance with applications
Gigabit to terabit per second networks
Petabyte file systems and data archives

Climate

Modeling the Complete Earth System

PI: Warren Washington, NCAR

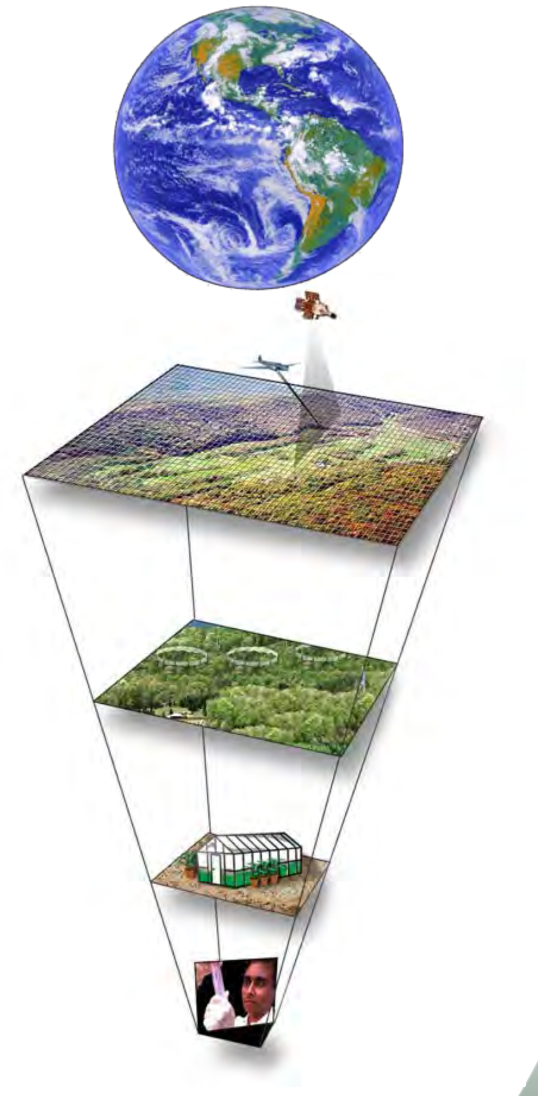
- Improving current scientific capabilities
  - Higher-resolution runs of emissions scenarios
  - Record time-to-solution using Jaguar capabilities
- Accelerating future science (tied to time to solution)
  - Explorations of multi-century ocean spin up
  - Explorations of multi-decadal carbon-cycle spin-up
  - Incorporation of land ice sheet treatment
  - Convergence on pre-industrial forcing for initialization of 20th century retrospective simulations
  - Vital steps toward completing Earth System Model capability



Simulated snapshot of the atmospheric CO_2 exchange between the free atmosphere and underlying land surface

Energy-Carbon-Water Challenges for (DOE) Climate Change Science

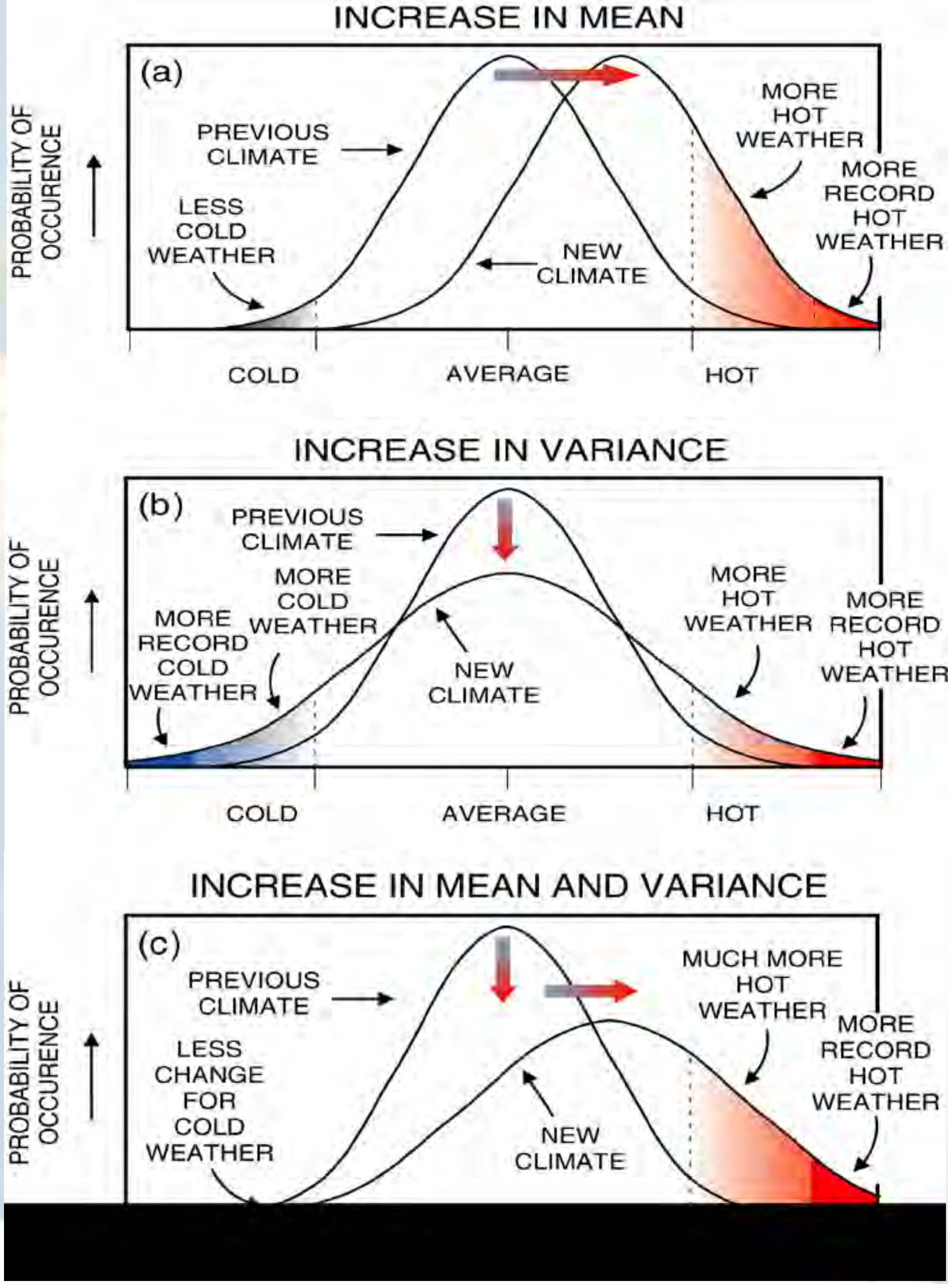
- **Provide the scientific basis to enable evaluation of climate change consequences in the context of sustainable production and use of interrelated resources**
- **Develop and apply an integrated earth system model combined with systematic experiments and observations to understand and project impacts of climate change**
- **Emphasize long-term (decadal) consequences at regional to global scales**



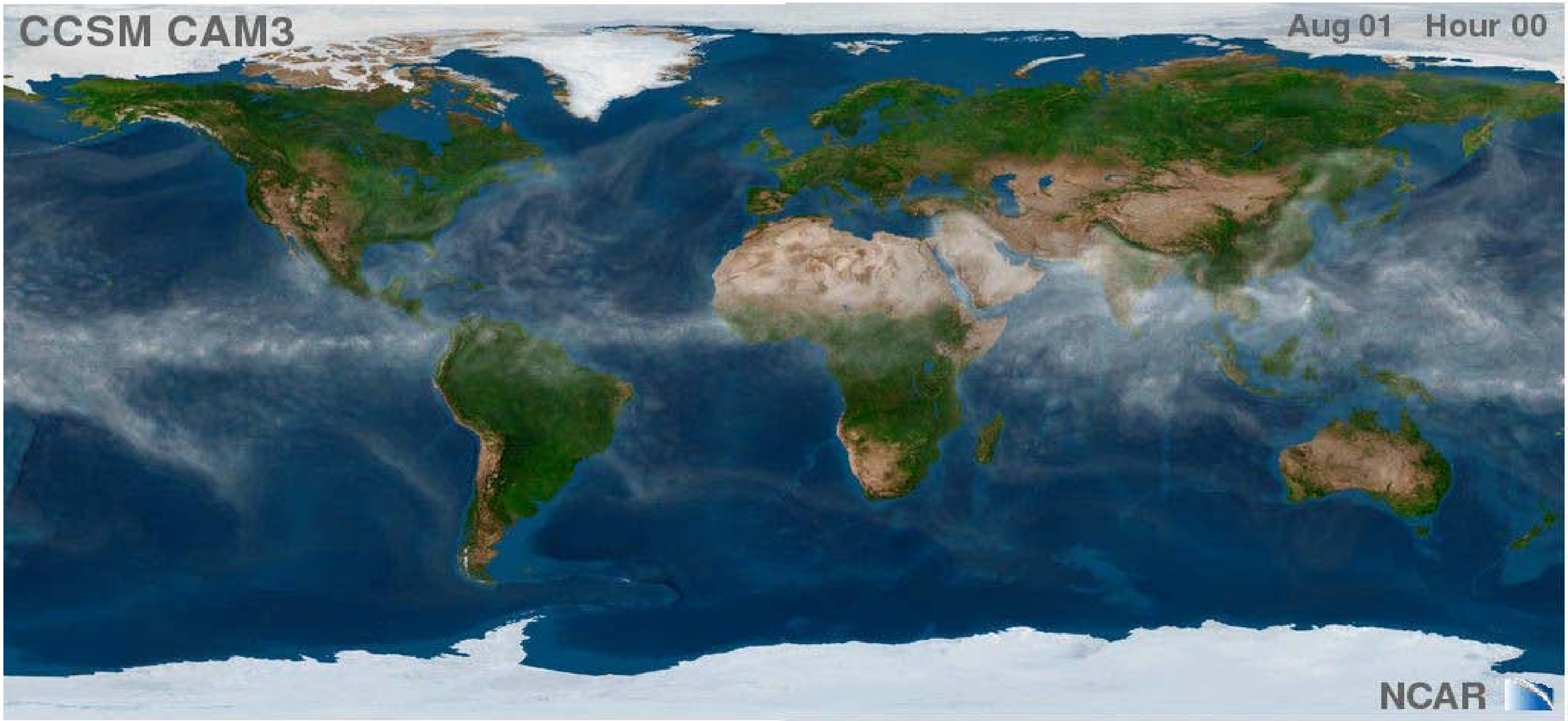
Approach

- **Leverage unique ORNL competencies in climate science**
  - *Long history of contributions to global climate modeling*
    - *fundamental contributions to mathematical & computational algorithms*
    - *fundamental contributions to computational implementation*
    - *fundamental knowledge of model component integration*
    - *development of global carbon cycle component models*
  - *Long history of contributions to carbon cycle science*
    - *measurements of terrestrial carbon cycle and ecosystem response*
    - *novel experimental capabilities to quantify terrestrial ecosystem sensitivities*
    - *extensive data holdings on carbon cycle observations*
    - *terrestrial ecosystem process model development*
  - *Unmatched leadership computational and data facilities*
    - *enabling and pacing technologies for making scientific progress*
  - *Widely recognized expertise in Integrated Assessment*
    - *unique capabilities in adaptation and vulnerabilities*

What do we mean by climate change?
What do we mean by climate extremes?

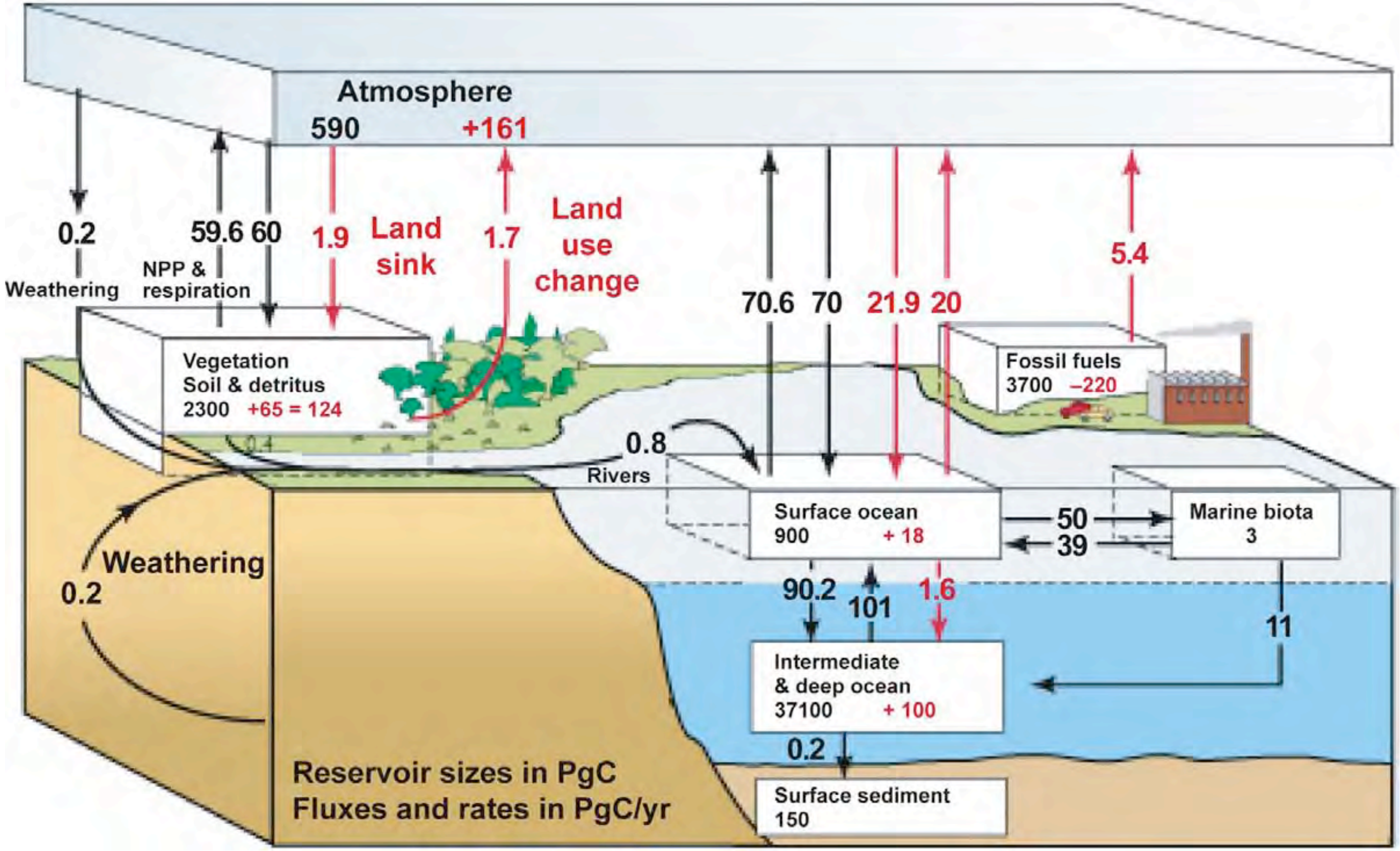


Climate Change Science: simulating the statistics of weather

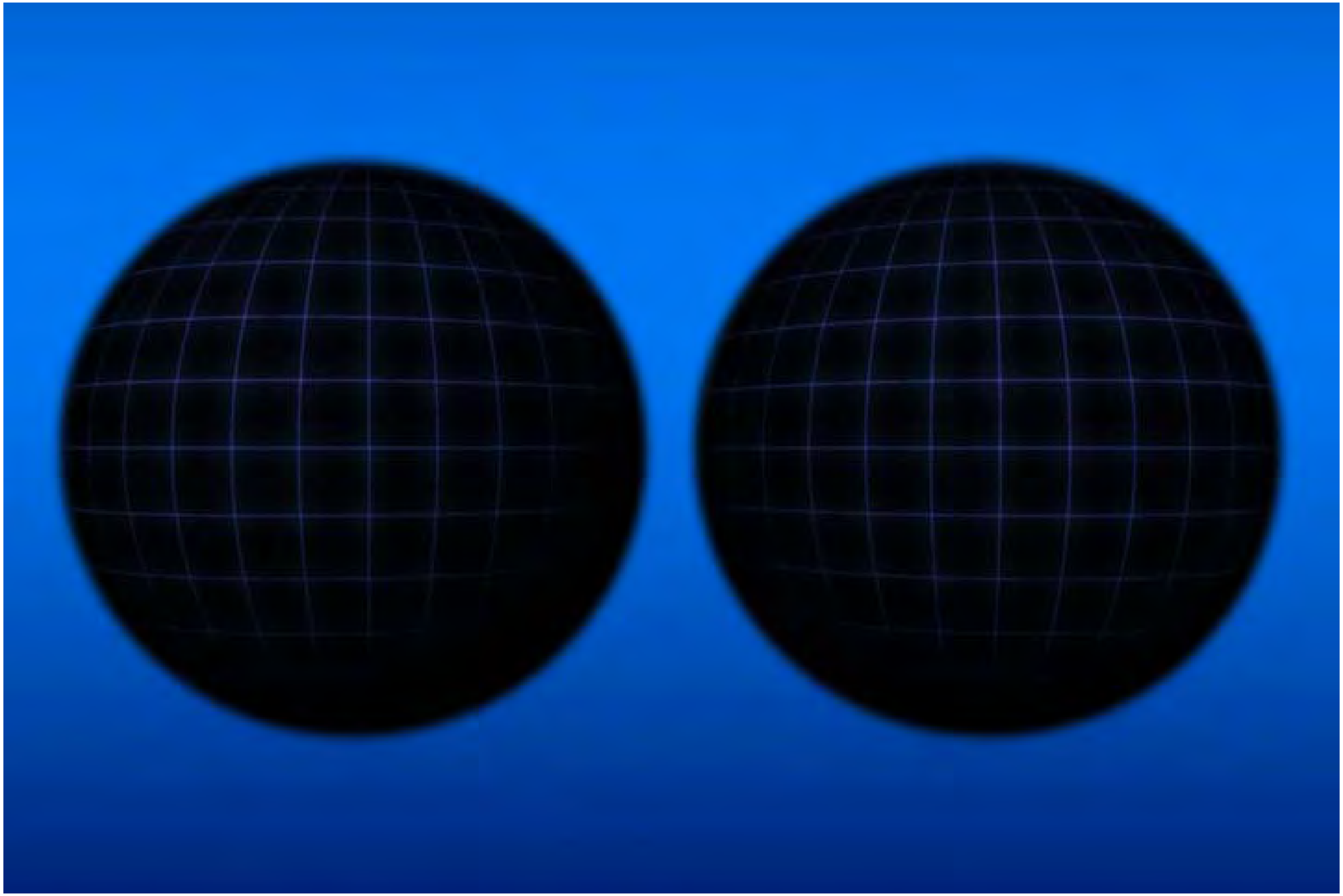


Column Integrated Water Vapor

Earth System Models add the carbon cycle



Carbon Cycle Modeling



Exchange of Carbon between surface and free atmosphere

Visualization and Data Analytics

Visualization

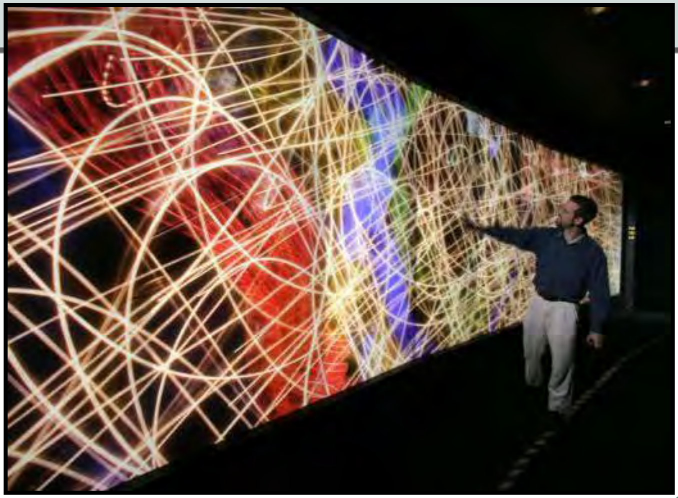
Once users have completed their runs, the Visualization task group helps them make sense of the sometimes overwhelming amount of information they generate.

- Viewing at a 30'x8' PowerWall
- Upgraded cluster with GPUs for remote visualization

End-to-End Solutions

Researchers must analyze, organize, and transfer an enormous quantity of data. The End-to-End task group streamlines the work flow for system users so that their time is not eaten up by slow and repetitive chores.

- Automate routine activities, ex. job monitoring at multiple sites
- Data Analysis



Million-fold increase in computing and data capabilities

2004

Cray X1
3 TF

2005

Cray XT3
Single-core
26 TF

2006

Cray XT3
Dual-core
54 TF

2007

Cray XT4
119 TF

2008

Cray XT4
Quad-core
166 TF
18 TB, 0.3 PB

Cray XT4
Quad-core
263 TF
62 TB, 1 PB

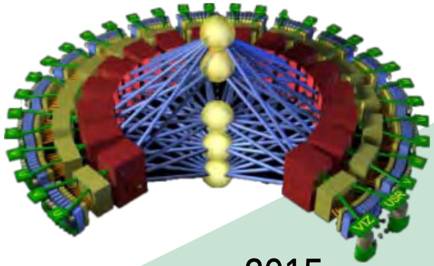
2009

Cray "Baker"
8/12-core, dual-socket SMP
~1 PF
100 TB, 2.5 PB

Cray "Baker"
8-core, dual-socket SMP
1.4 PF
300 TB, 10 PB

2011

DARPA
HPCS
20 PF

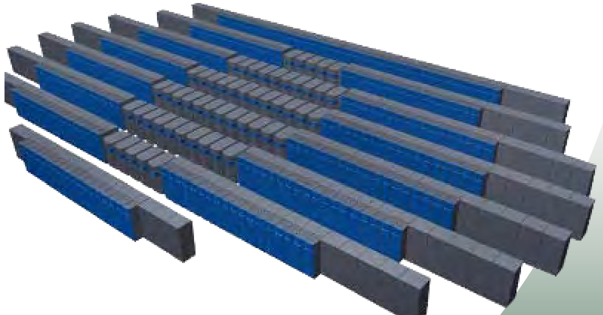


2015

Future system
100–250 PF

2018

Future system
1000 PF
(1 EF)



ORNL's Current and Planned Data Centers

Computational Sciences Building (40K ft^2)

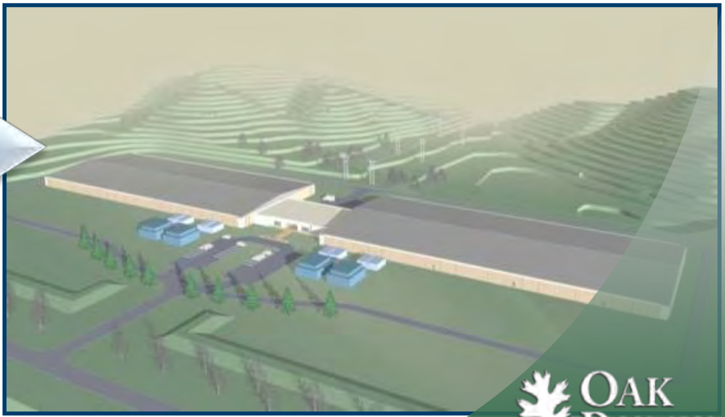
- Upgraded building power to 25 MW
- Deployed a 6,600 ton chiller plant
- Tripled UPS and generator capability

Multiprogram Research Facility (32K ft^2)

- Capability computing for national defense
- Expanded to 25 MW of power and 8,000 ton chiller

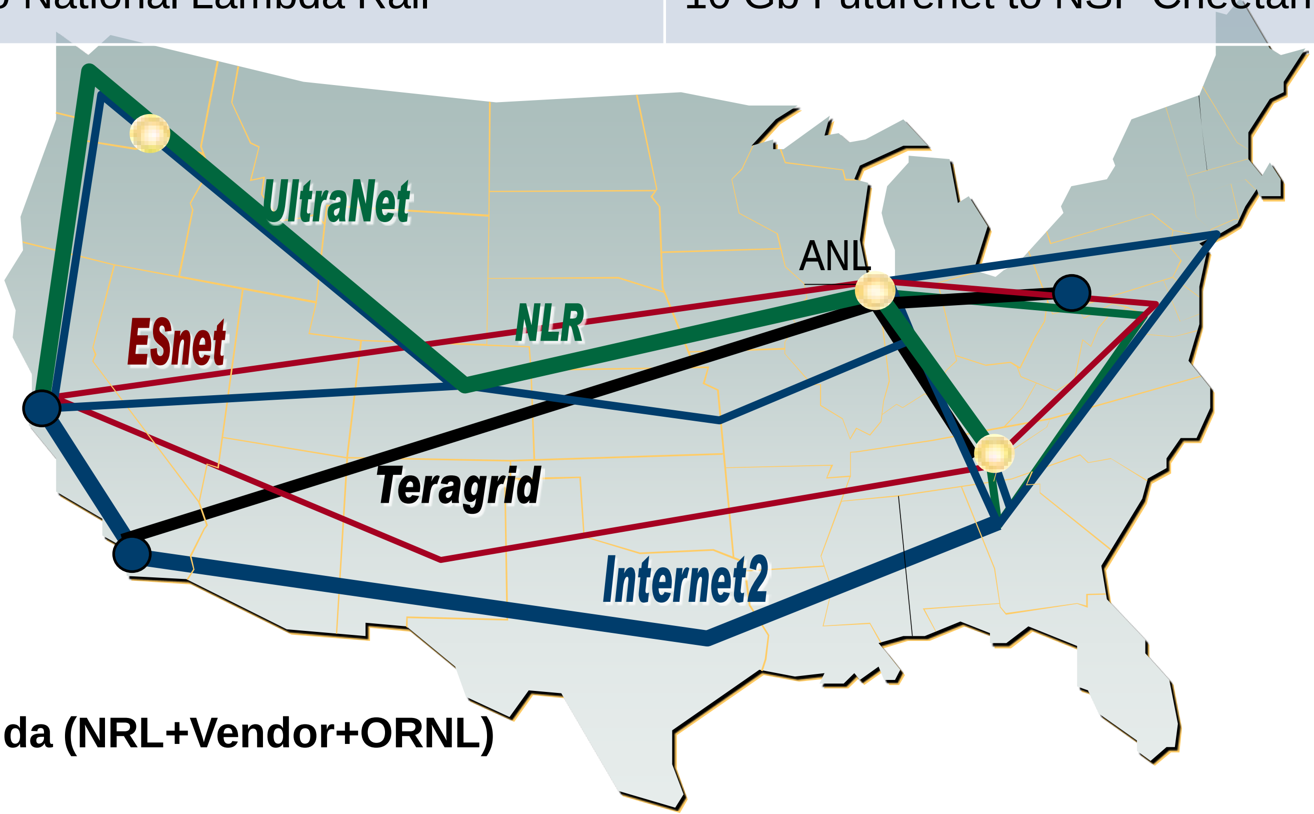
Multiprogram Data Center (260K ft^2)

- 110K ft^2 classified; 110K ft^2 unclassified
- Shared mechanical & electrical infrastructure
- Build out 25K ft^2 on each side as needed
- Lights out facility



High Bandwidth Connectivity to NCCS Enables Efficient Remote User Access

Connected to Major Science Networks	
OC192 to ESNET with backup OC48	1 - 4 x 10 Gb to NSF Teragrid
10 Gb to Internet2	2 x 10 Gb UltraScienceNet
4 x 10 Gb to National Lambda Rail	10 Gb Futurenet to NSF Cheetah net



Target

- **100Gb/lamda (NRL+Vendor+ORNL)**

Take away message

- ORNL and UT lead the world in High Performance Computing (HPC)
- Many of the nation’s most challenging problems cannot be solved without simulation using HPC
- HPC is a $200M per year enterprise in our area
- 500+ people directly employed by ORNL and UT in HPC

Questions?

Visit us at: http://computing.ornl.gov

Jim Hack	Buddy Bland	Doug Kothe
jhack@ornl.gov	BlandAS@ornl.gov	Kothe@ornl.gov