

UT-BATTELLE



AMES LABORATORY



ARGONNE NATIONAL LABORATORY



Los Alamos NATIONAL LABORATORY



Lawrence Livermore National Laboratory



NCAR

Pacific Northwest National Laboratory



PPPL



Sandia National Laboratories

NATIONAL LEADERSHIP



COMPUTING FACILITY

# National Leadership Computing Facility

Presented to  
DOD Common Operating Environment Workshop  
March 17, 2005

Buddy Bland  
**Director of Operations**  
**Center for Computational Sciences**  
**Oak Ridge National Laboratory**

THE CENTER FOR  
COMPUTATIONAL SCIENCES

OAK RIDGE NATIONAL LABORATORY  
U. S. DEPARTMENT OF ENERGY

# Outline

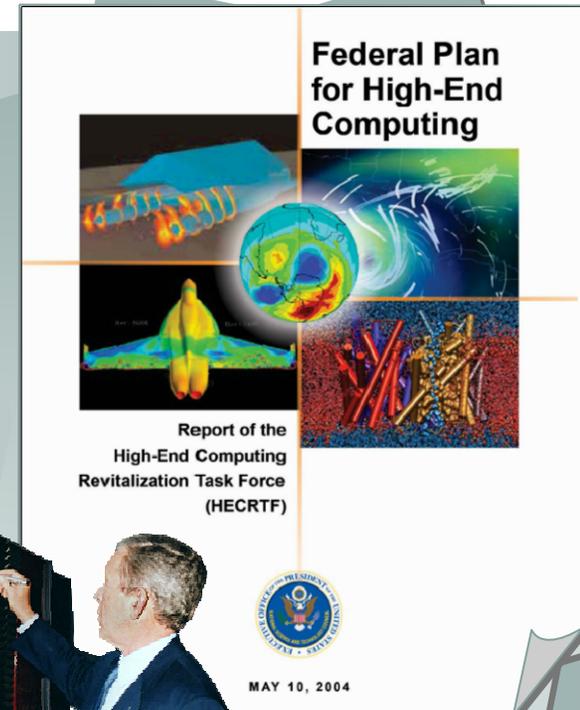
- Background of the Leadership Computing Facility
  - **Federal initiative**
  - **Scientific drivers**
  - **Why ORNL?**
  - **Computer Systems**
- Others from ORNL will speak about our applications
  - **Thomas Schulthess – Grand Challenge Teams**
  - **Richard Barrett – Applications and Tools**
  - **Trey White – Climate Modeling**

# Leadership computing is a White House priority

**“The goal of such systems [leadership systems] is to provide computational capability that is at least 100 times greater than what is currently available.”**

**“...Leadership Systems are expensive, typically costing in excess of \$100 million per year....”**

**– Page 29, Federal Plan for High-end Computing**



# Leadership computing is a congressional priority

- Appropriated \$30M in FY04 for leadership computing
- Additional \$30M appropriated in FY05
- Public Law 108-423 Department of Energy High-End Computing Revitalization Act of 2004



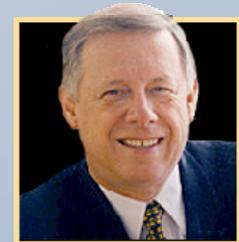
# Leadership computing is a State of Tennessee priority

- **\$9M State of Tennessee Investment in the Joint Institute for Computational Sciences**
- **\$10M for National Academy Level Joint Faculty**
- **\$12M for high speed networks for research and education**
- **\$1M/year for Computational Science Initiative for graduate student training and outreach**



**“I have recommended funds ...to attract more nationally-recognized faculty members (jointly with ORNL).... There is an opportunity today.... to rapidly become world class in some areas like supercomputers, materials science, and nanotechnology.**

**....Our pioneer ancestors wouldn't have known what supercomputers were, but I believe they would have understood our aspirations perfectly.”**



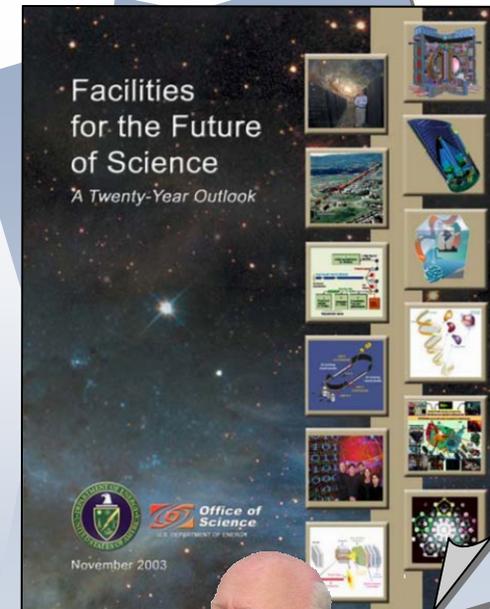
**- Gov. Bredesen, State of the State Speech, January 31, 2005**

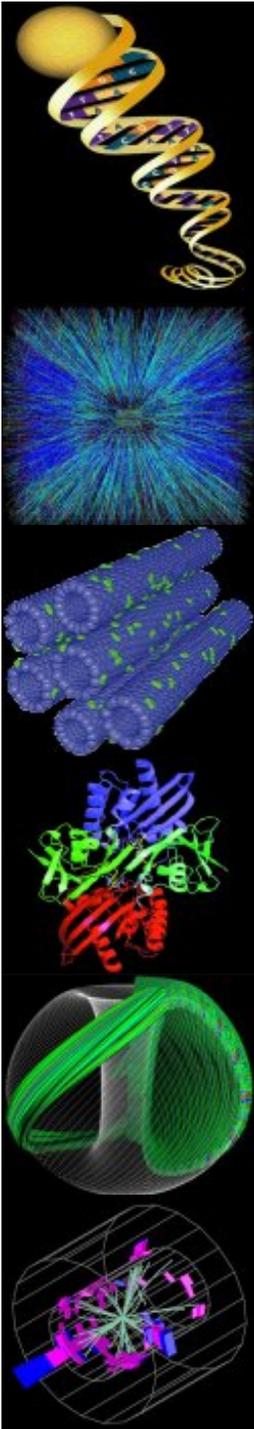
# Leadership computing is the highest domestic priority for Office of Science

- **Ray Orbach has articulated his philosophy for the SC laboratories**
  - Each lab will have world-class capabilities in one or more areas of importance to Office of Science
  - ORNL: SNS and NLCF will underpin world-class programs in materials, energy, and life sciences
- **20-year facilities plan being used to set priorities among projects**

*“I am committed to the concept of a Leadership Class Computing facility at Oak Ridge National Laboratory. The facility will be used to meet the missions of the Department and those of other agencies. I can assure you that I understand the important role supercomputing plays in scientific discovery.”*

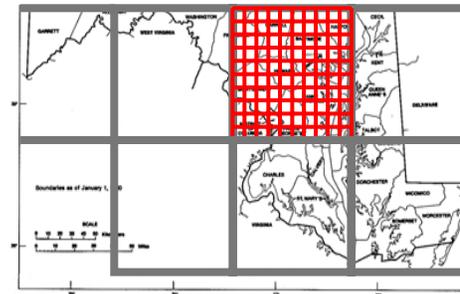
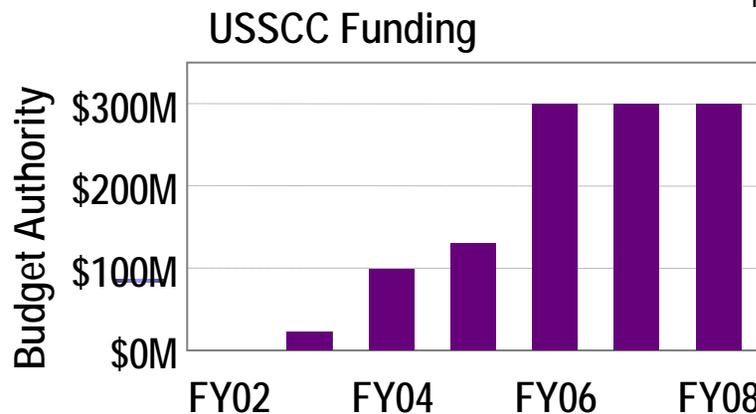
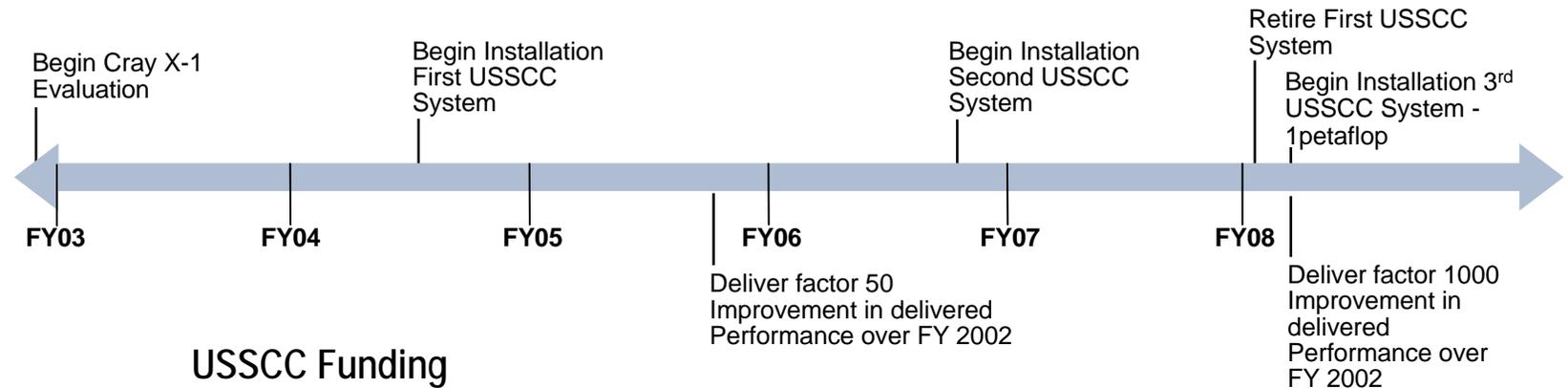
*Secretary Bodman*





# Reassert U.S. leadership in high performance computing for science in strategic areas

GOALS:	By 2005:	By 2008:
Deliver UltraScale Scientific Computing Capability (USSCC) to:	Deliver computational performance 50 times greater than now achieved for selected scientific modeling problems	Deliver computational performance 1,000 times greater than now achieved for selected scientific modeling problems



**Graphical depiction of increased spatial resolution in climate models - a map of DC area showing today's grid in black and the USSCC grid in red.**

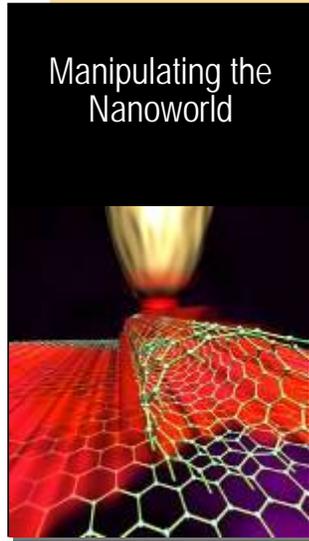
Ray Orbach, 2004

OAK RIDGE NATIONAL LABORATORY  
U. S. DEPARTMENT OF ENERGY

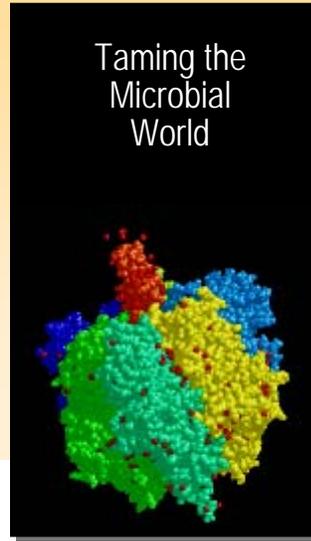
# Leadership Computing for Science

Critical for success in key national priorities

Office of Science research priorities



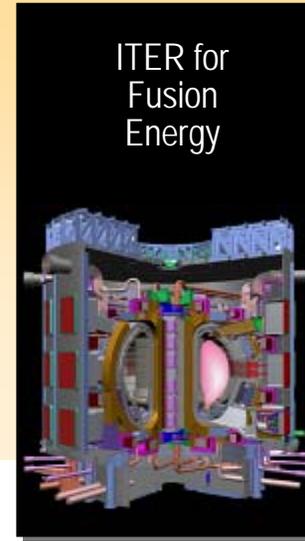
Computational design of innovative nanomaterials



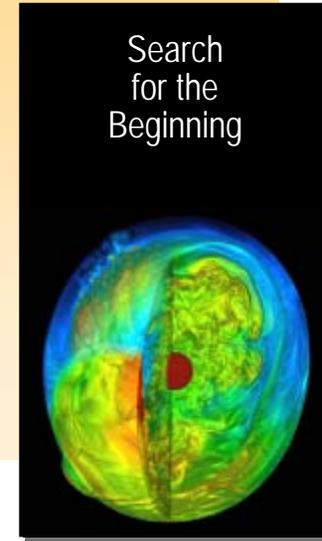
Predictive understanding of microbial molecular and cellular systems



Full carbon cycle in climate prediction, IPCC



Simulation of burning plasma, Fusion Simulation Project



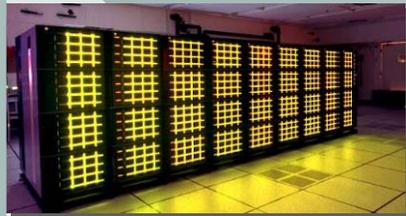
Terascale Supernovae Simulation

Theory, Mathematics, Computer Science

National Leadership-Class Computing Facility for Science

# Center for Computational Sciences performs three inter-related activities for DOE

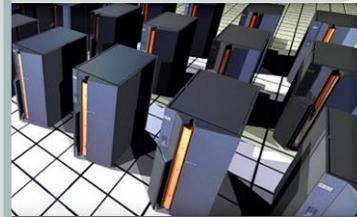
- Deliver National Leadership Computing Facility for science
  - **Focused on grand challenge science and engineering applications**
- Principal resource for SciDAC and (more recently) other SC programs
  - **Specialized services to the scientific community: biology, climate, nanoscale science, fusion**
- Evaluate new hardware for science
  - **Develop/evaluate emerging and unproven systems and experimental computers**



**Intel Paragon:  
World's fastest  
computer**



**IBM Power3:  
DOE-SC's first  
terascale system**



**IBM Power4:  
8th in the world  
(2001)**



**Cray X1:  
Capability  
computer  
for science**

# CCS Terascale Systems



Cray X1 – Phoenix (6.4 TF)  
Largest X1 in the world and first in DOE  
Scalable Vector Architecture  
512 Multi-Streaming Processors 400 MHz

IBM Power4 – Cheetah (4.5 TF)  
First Power4 system in DOE - #8 on Top500  
Cluster using IBM Federation Interconnect  
864 IBM Power4 processors 1.3 GHz



SGI Altix – Ram (1.5 TF)  
Large Globally Addressable Memory System  
256 Intel Itanium2 Processors 1.5 GHz  
Linux with single operating system image

IBM Power3 – Eagle (1 TF)  
First Terascale machine in DOE-SC  
Cluster of SMP Nodes  
736 IBM Power3 Processors 375 MHz



# New world-class facility capable of housing leadership class computers

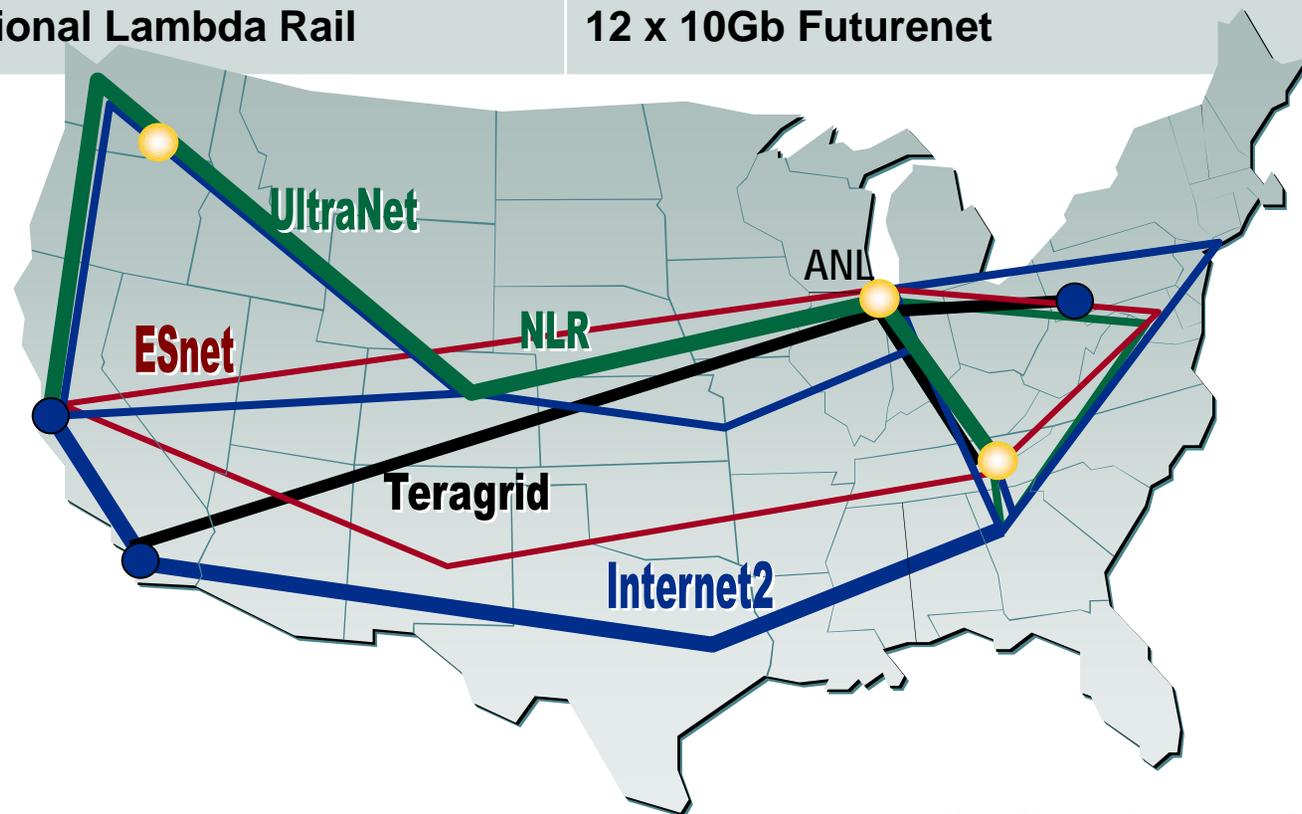
- \$72M private sector investment in support of leadership computing
- Space and power:
  - **40,000 ft<sup>2</sup> computer center with 36-in. raised floor, 18 ft. deck-to-deck**
  - **8 MW of power (expandable)**
- High-ceiling area for visualization lab (Cave, Powerwall, Access Grid, etc.)
- Separate lab areas for computer science and network research



# High bandwidth connectivity to NLCF enable efficient remote user access

## Connected to major science networks

OC48 to ESNET (provisioned by ESNET)	1 - 4 x 10Gb to NSF Teragrid
10Gb to Internet2	2 x 10Gb Ultranet
2 x 10Gb to National Lambda Rail	12 x 10Gb Futurenet



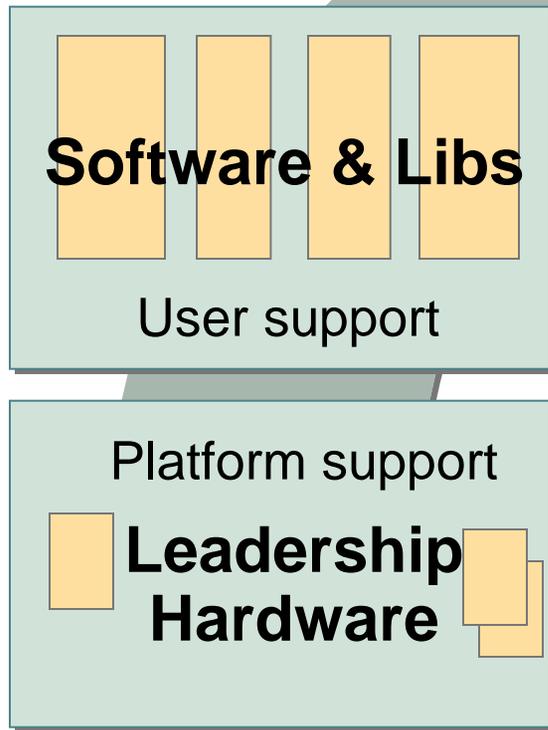


# Our plan of action to deliver leadership computing for DOE

- **Rapidly field most powerful open capability computing resource for scientific community**
  - Providing clear upgrade path to at least 100 teraflop/s (TF) by 2006 and 250 TF by 2007/2008
- **Deliver outstanding access and service to the research communities**
  - Utilizing most powerful networking capability extant coupled with secure and highly cost-effective operation by proven team
- **Deliver much higher sustained performance for major scientific applications than currently achievable**
  - Developing next generation models and tools
  - Engaging computer vendors on hardware needs for scientific applications
- **Engage research communities in climate, fusion, biology, materials, chemistry, and other areas critical to DOE-SC and other federal agencies**
  - Enabling high likelihood of breakthroughs on key problems
- **Conduct in-depth exploration of most promising technologies for next-generation leadership-class computers**
  - Providing pathways to petaflop/s (PF) computing within decade

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

## Leadership-class Computing Facility



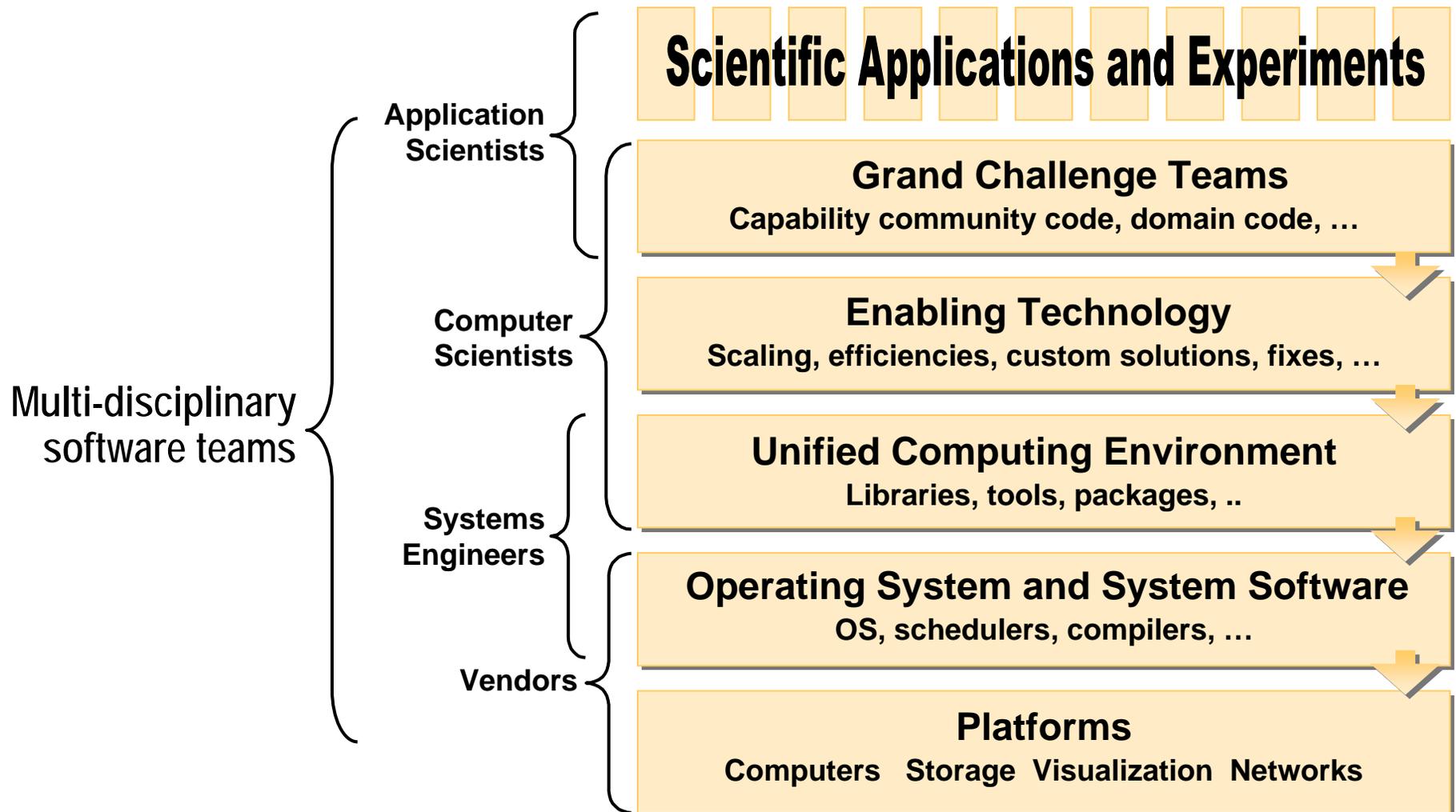
**Computing Environment**  
Common look and feel across diverse hardware

## Grand Challenge Teams



## Breakthrough Science

# A unified enabling software strategy



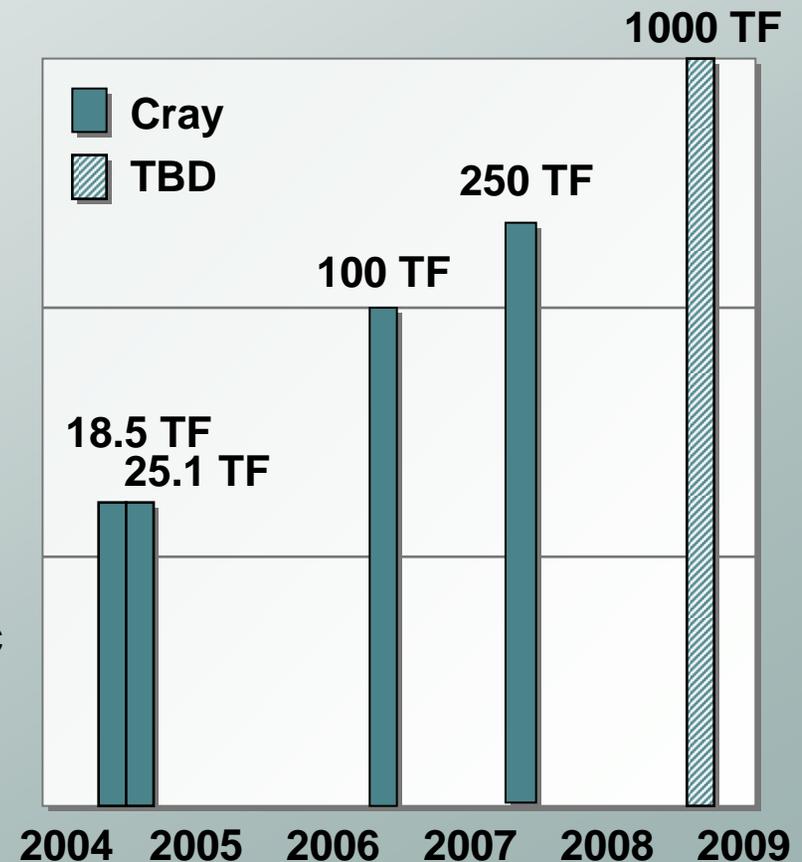
# NLCF hardware roadmap

**In 2005,**

- Deploy 18.5TF Cray X1E and 25.1TF Cray XT3 systems
- Cray forms and supports “Supercomputing *Center of Excellence*”
- Develop/deploy complementary software environment
- Full operational support of NLCF as a capability computing center
- Deliver computationally intensive projects of large scale and high scientific impact through competitive peer review process

**In 2006, deploy 100TF Cray XT3**

**In 2007-8, deploy 250TF Cray Rainier**



# Phoenix – The CCS Cray X1

- Largest Cray X1 in the world
- 2 TB globally addressable memory
- 512 processors
  - 400 MHz, 800 MHz vector units
- 32 TB of disk
- Most powerful processing node
  - 12.8 GF CPU, 2-5x commodity processors
- Highest bandwidth communication with main memory
  - 34.1 GB/sec
- Highly scalable hardware and software
- High sustained performance on real applications



# Cray X1E Supercomputer

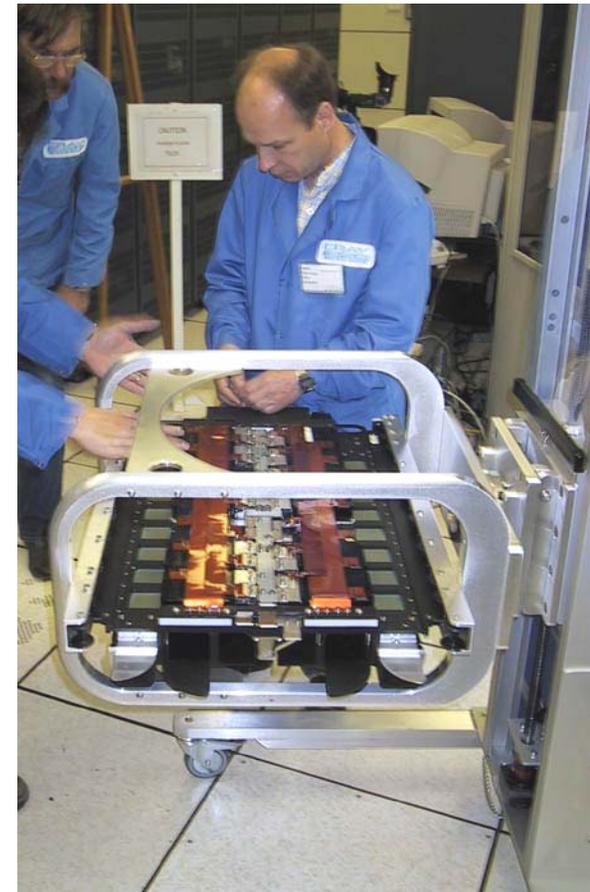
- **Re-implement X1 in 0.13 $\mu$ m technology**
  - Double processor density
  - 41% performance increase per processor
  - Cache scales in bandwidth with processor
  - Significantly reduces cost per processor
- **Approximately triples X1 performance per cabinet**
- **Upgradeable from X1 by processor swap**
- **Low-risk upgrade**
  - Software that runs on Cray X1E is faster, but *not* different
    - Users do not need to recompile
    - Supported features identical between X1 and X1E
    - Uses one code base for both products

# Phoenix – Cray X1 to X1E upgrade

- One-fourth of machine will be upgraded at a time
- X1 and X1E nodes will be partitioned into separate systems until upgrade completed
- System goes from 6.4 TF to 18.5 TF



Upgrades begin in June and continue through summer



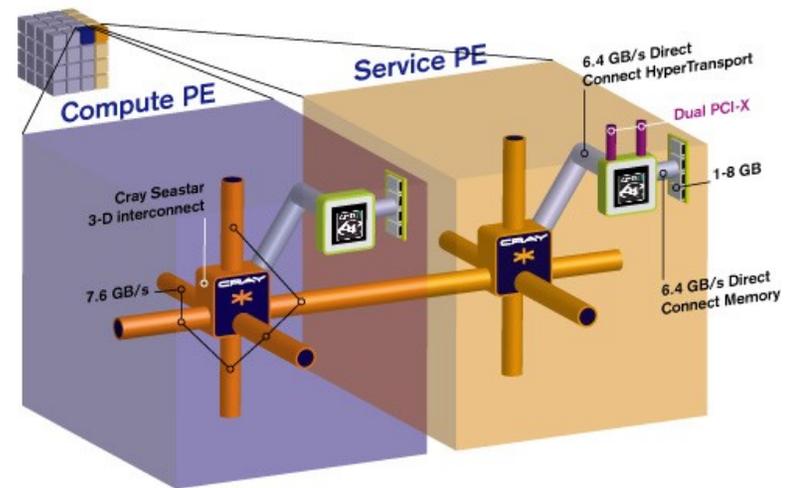
# Cray XT3 Architecture

## Design Goals

- High-performance commodity microprocessor
- Surround with balanced or “bandwidth rich” environment
- Eliminate “barriers” to scalability
  - SMPs do not help here
  - Eliminate Operating System Interference (OS Jitter)
  - Reliability must be designed in
  - Resiliency is key
  - System Management
  - I/O



Cray XT3 Scalable Architecture



- XT3 is 3<sup>rd</sup> generation Cray MPP
- System implements many T3E architectural concepts in current best-of-class technologies
- Scales to many 1,000s of processors
  - Scalable single-PE architecture
  - Scalable reliability and system management
  - Scalable SW environment
  - Scalable I/O subsystem

# Jaguar – The CCS XT3 system

Cabinets	Performance	Processors	Memory	Disk Space	I/O Bandwidth
56	25 TF	5,304	10.5 TB	120 TB	15 GB/s
120	54.6 TF	11,374	23 TB	240 TB	30 GB/s
120	109 TF	22,748	46 TB	480 TB	60 GB/s



“Jaguar” first 11 cabinets

# Jaguar – CCS XT3 system delivery schedule

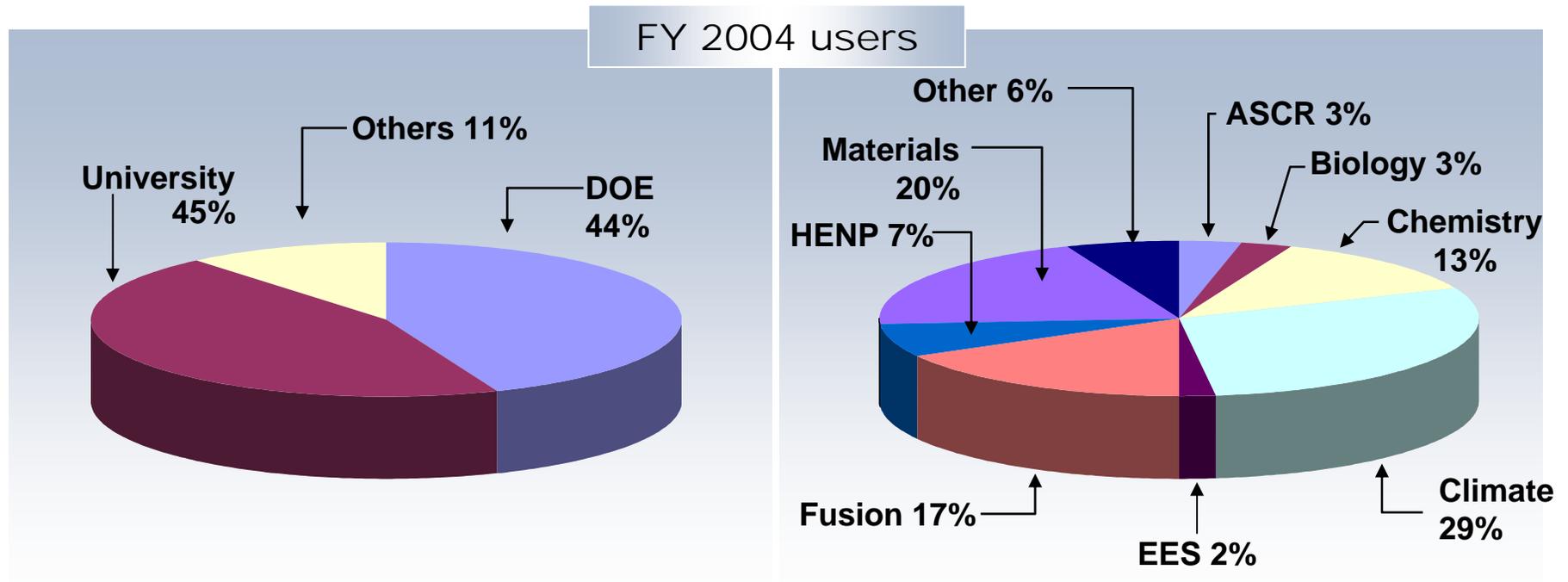
- Initial cabinet delivered December 2004 to begin application porting
- 11 additional cabinets delivered February 2005
- 32 additional cabinets to be delivered in April
- 12 additional cabinets to be delivered in June

## **And subject to additional funding:**

- 64 additional cabinets (50 TF)
- Upgrade all cabinets to dual-core processors and double the memory and I/O bandwidth (100 TF)



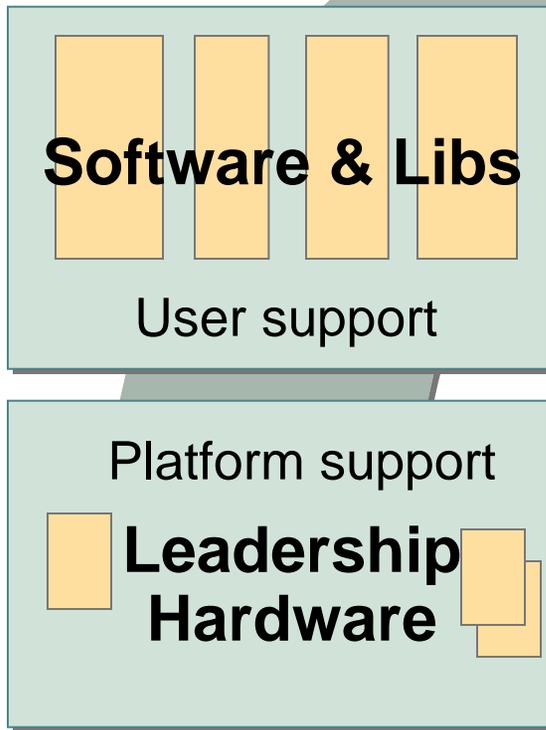
# CCS: A diverse user community using many architectures



With four architectures and users who run at many centers, we need to provide consistency for our users.

Our users need common tools to enable their science

**Leadership-class Computing Facility**



**Computing Environment**  
Common look and feel across diverse hardware

**Grand Challenge Teams**



**Breakthrough Science**

OAK RIDGE NATIONAL LABORATORY  
U. S. DEPARTMENT OF ENERGY

# Questions?

