# Petascale Hardware Update

| | | |
|---|---|---|
| **1:00** | **Session goals** | **Buddy Bland (ORNL)** |
| **1:05** | **Roadrunner** | **Andy White (LANL)** |
| **1:30** | **Intrepid** | **Ray Bair (ANL)** |
| **1:55** | **Zia** | **Sudip Dosanjh (SNL)** |
| **2:15** | **Kraken** | **Jim Ferguson (NICS)** |
| **2:30** | **Jaguar** | **Buddy Bland (ORNL)** |
| **2:45** | **Future Directions** | **Panel (All speakers)** |

# Petascale Computing at ORNL

**Buddy Bland**

**Project Director**

**Leadership Computing Facility**

**Oak Ridge National Laboratory**

***presented to:***

**Fall Creek Falls Conference**

**September 9, 2008**

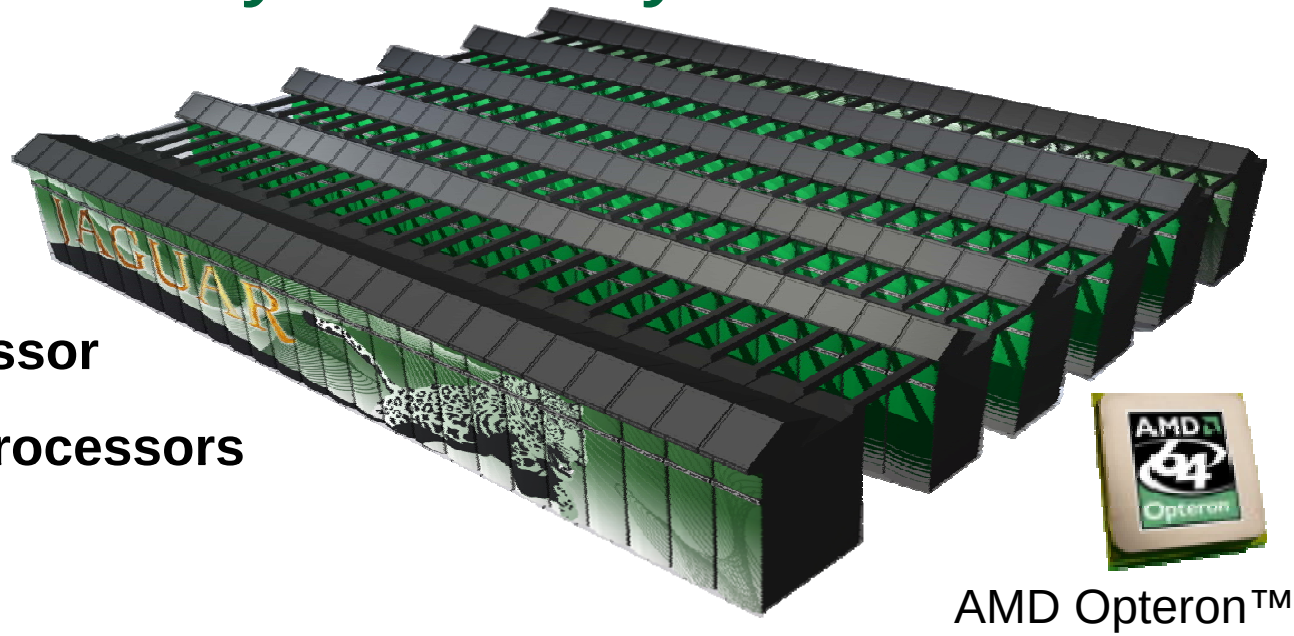
Managed by UT-Battelle
for the Department of Energy

# Jaguar – Cray XT4 Upgraded to 250 TF

- **We have upgraded Jaguar from single-core to, dual-core to quad-core Opteron™ processors**
  - **Replaced 7,832 processors and added 15,664 2GB DIMMS**

| | |
|---|---|
| Quad-core Processors | 7,832 |
| Memory / Core | 2 GB |
| System Memory | 62 TB |
| Disk Bandwidth | 44 GB/s |
| Disk Space | 900 TB |
| Node Size | 4-core, 35 GF |

# 1 Petaflops Cray XT5 System




## 4Q 2008: Cray XT5

- **1 Petaflops system**
- **37 Gigaflops processor**
- **27,888 quad-core processors Barcelona 2.3 GHz (111,552 cores)**
- **2 GB/core; 223 TB total**
- **200+ GB/s disk bandwidth**
- **13,944 dual-socket, 8-core SMP "nodes" with 16 GB**
- **6.0 MW system power**
- **150 cabinets, 3,400 ft$^2$**

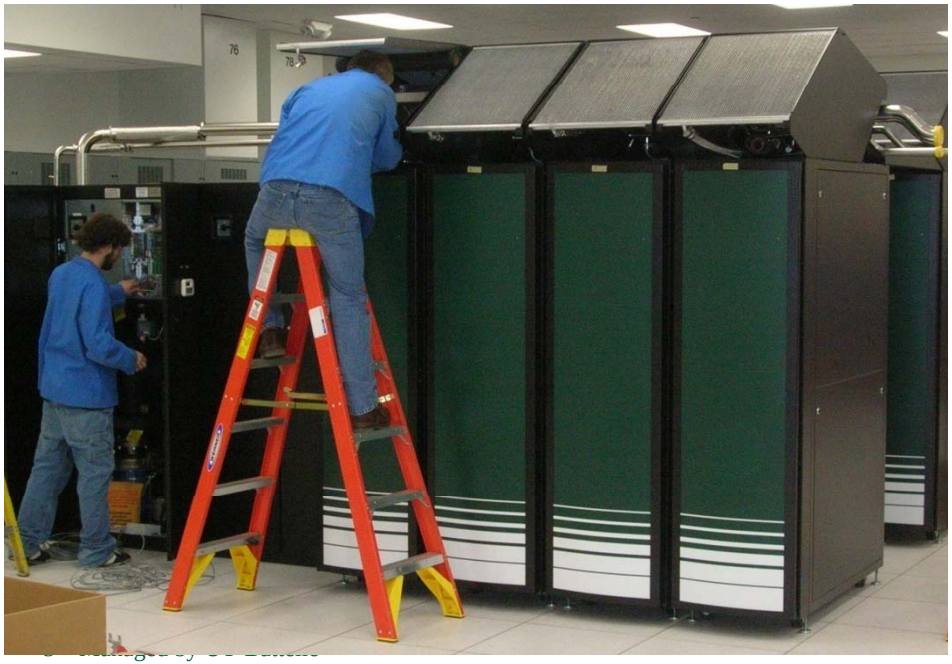



AMD Opteron™
Quad-Core
2.3 GHz

- **Compute Node Linux operating system**
- **Torus interconnect**

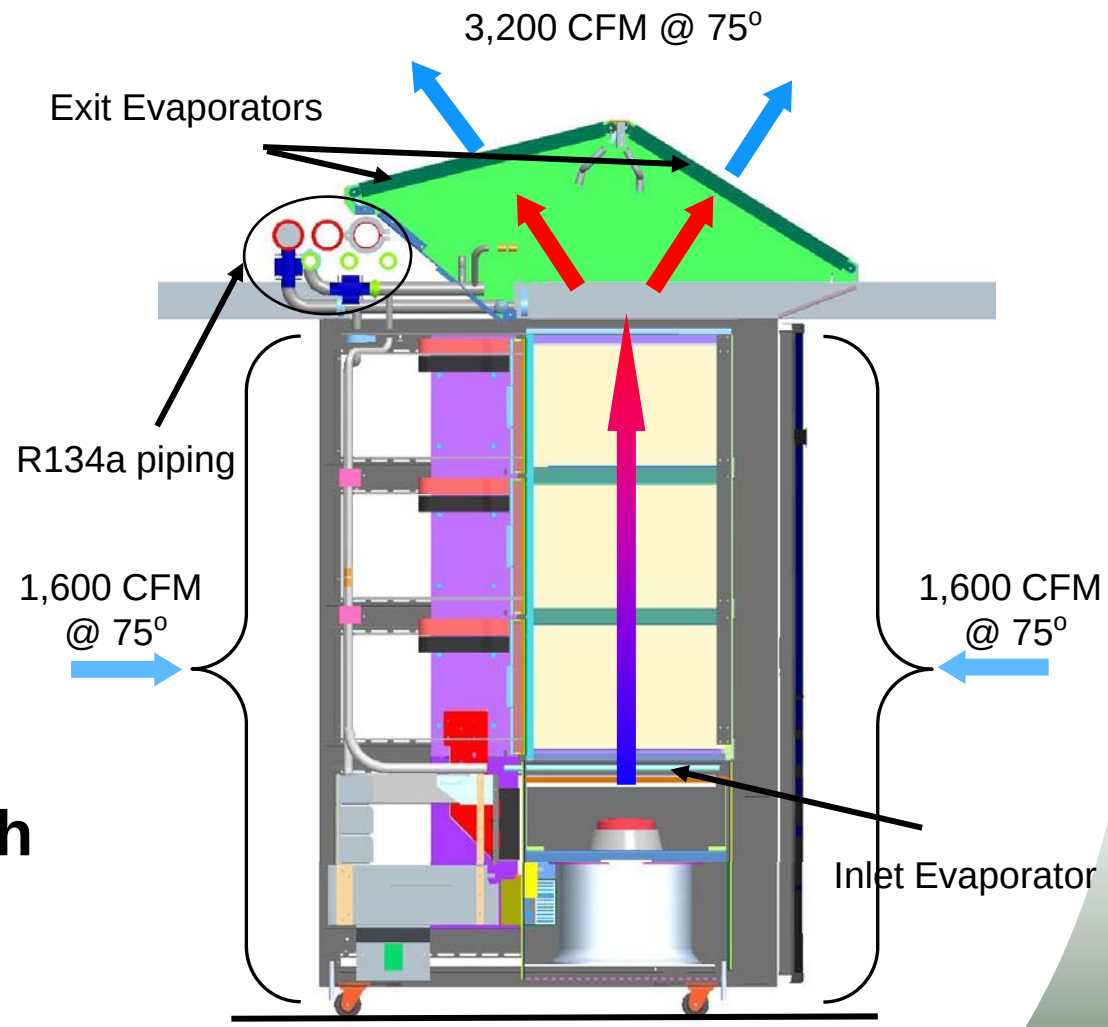
# First cabinets have started arriving

- **Liquid cooling is working (in fact it is cooling the room!)**
- **Cabinets continue to be delivered through December**



# Cray XT5-High Efficiency Liquid Cooling

- **Liquid Cooled design removes heat to liquid before it leaves the cabinet**
- **Saves about 900KW of power just in air movement**
- **Phase change liquid to gas removes heat much more efficiently**




**Cray ECOphlex™**
**Liquid Cooling Technology**

# Storage Infrastructure

## *Center -wide shared file systems*

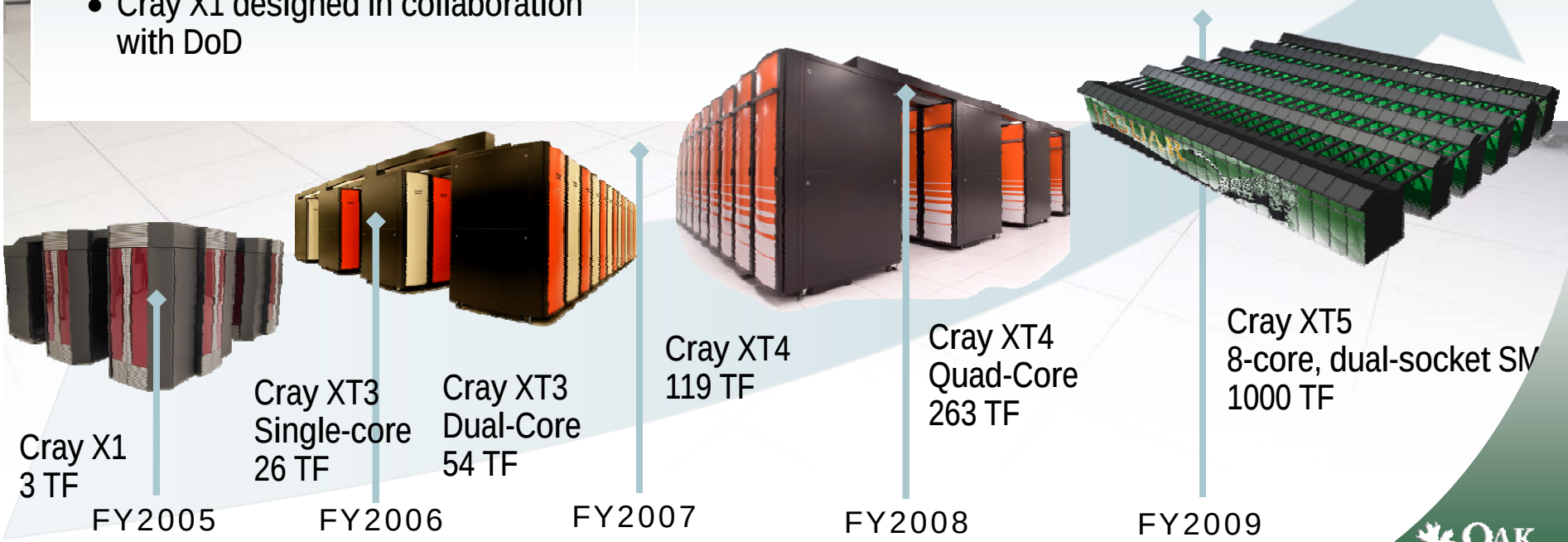
- **"Spider" is being installed to provide a shared, parallel file system for all systems**
  - **Based on Lustre file system**
- **Bandwidth of over 200 GB/s with 10 petabytes capacity**
- **HPSS provides archival storage for all system**
- **HPSS has been upgraded with two additional tape libraries to add additional capacity and bandwidth**

Challenge: Management of over 14,000 disk drives

HPSS

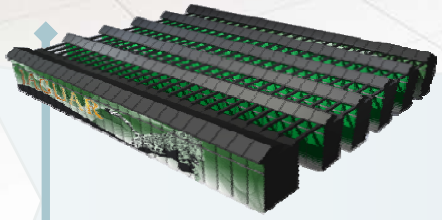
# We have increased system performance by more than 300 times since 2004

| Hardware scaled from single-core through dual-core to quad-core and dual-socket SMP nodes | Scaling applications and system software is the biggest challenge |
|---|---|
| • NNSA and DoD have funded much of the basic system architecture research<br>  • Cray XT based on Sandia Red Storm<br>  • IBM BG designed with Livermore<br>  • Cray X1 designed in collaboration with DoD | • SciDAC program is funding scalable application work that has advanced many science apps<br>• DOE-SC and NSF have funded much of the library and applied math as well as tools<br>• Computational Liaisons key to using deployed systems |

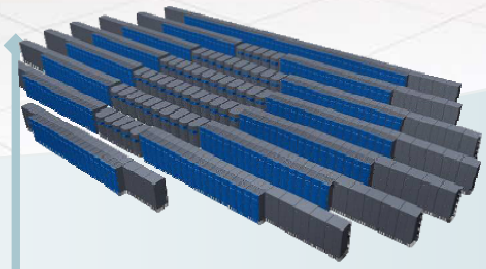


# We will advance computational capability by 1000x over the next decade

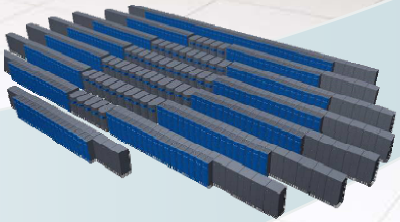
| Mission: Deploy and operate the computational resources required to tackle global challenges | Vision: Maximize scientific productivity and progress on the largest scale computational problems |
|---|---|
| • Deliver transforming discoveries in materials, biology, climate, energy technologies, etc.<br>• Ability to investigate otherwise inaccessible systems, from supernovae to energy grid dynamics | • Providing world-class computational resources and specialized services for the most computationally intensive problems<br>• Providing stable hardware/software path of increasing scale to maximize productive applications development |



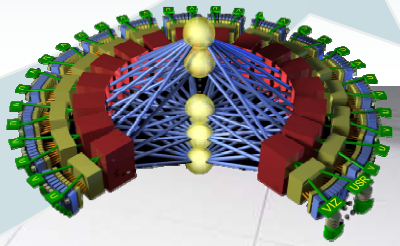
Cray "Baker": 1 PF
Leadership-class
system for science



DARPA HPCS: 20 PF
Leadership-class
sustained PF system



100-250 PF



Future system: 1 EF

FY2009 FY2011 FY2015 FY2018

# 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$^2$)**

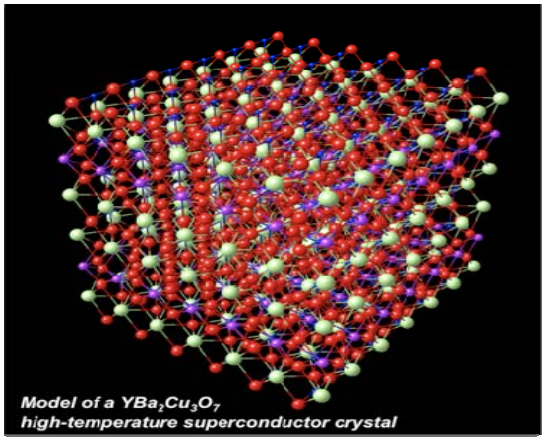
- Capability computing for national defense
- Expanded to 25 MW of power and 8,000 ton chiller

**Multiprogram Data Center (260K ft$^2$)**

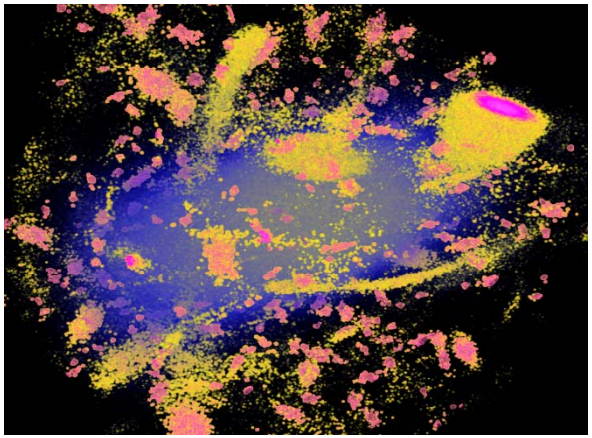
- 110K ft$^2$ classified; 110K ft$^2$ unclassified
- Shared mechanical & electrical infrastructure
- Build out 25K ft$^2$ on each side as needed
- Lights out facility

# Advancing Scientific Discovery

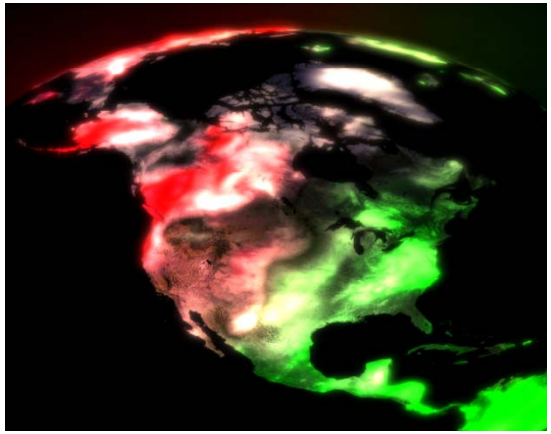
5 of the top 10 ASCR science accomplishments over the past 18 months used LCF resources & staff



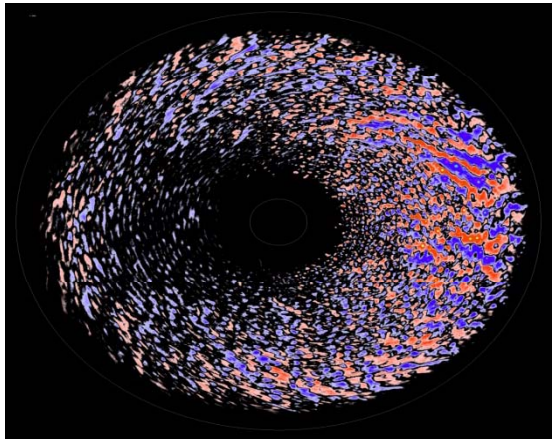

**Electron pairing in HTC cuprates***
The 2D Hubbard model emits a superconducting state for cuprates & exhibits an electron pairing mechanism most likely caused by spin-fluctuation exchange. *PRL* (2007, 2008)



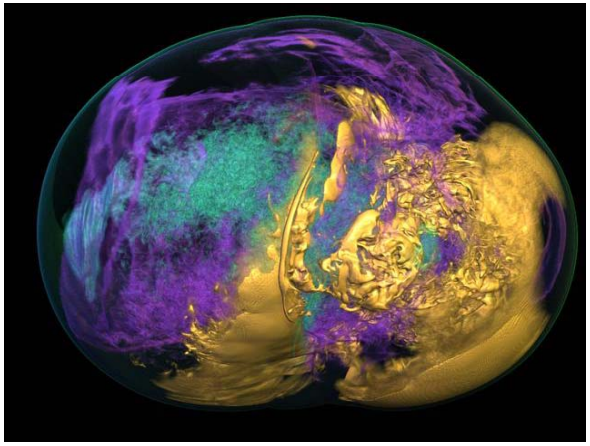
**Shining the light on dark matter***
A glimpse into the invisible world of dark matter, finding that dark matter evolving in a galaxy such as our Milky Way remains identifiable and clumpy. Aug 7, 2008 issue of *Nature*



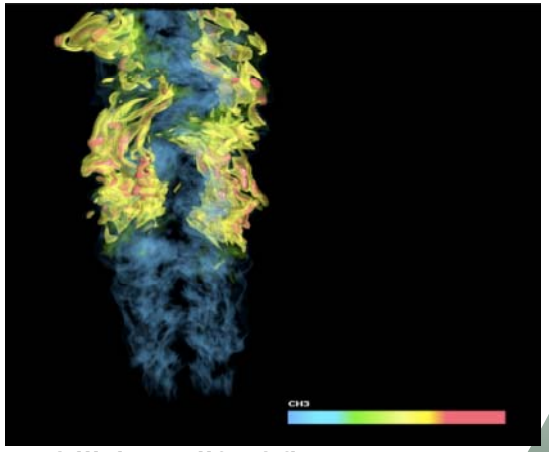
**Modeling the full earth system**
Simulated time evolution of the atmospheric $CO_2$ concentration originating from the land's surface



**Taming turbulent heat loss in fusion reactors*** Advanced understanding of energy loss in tokamak fusion reactor plasmas,. *PRL*(vol 99) and *Physics of Plasmas* (vol 14)



**How does a pulsar get its spin?***
Discovered the first plausible mechanism for a pulsar's spin that fit observations, namely the shock wave created when the star's massive iron core collapses. Jan 4, 2007 issue of *Nature*



**Stabilizing a lifted flame***
Elucidated the mechanisms that allow a flame to burn stably above burners, namely increasing the fuel or surrounding air co-flow velocity *Combustion and Flame(2008)*

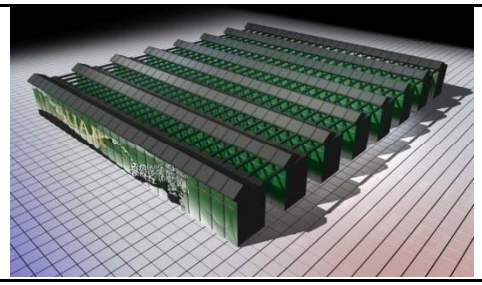
# Questions?

Buddy Bland
National Center for Computational Sciences
Oak Ridge National Laboratory
http://nccs.gov

# Panel Discussion

## Future Directions for Leadership Computers – How Big Can These Systems Get?

# The Size and Power Requirements for Computer Centers Continues to Grow

| | | | |
|---|---|---|---|
| ORNL Computational Science Building, 2003, 28 MW, 320,000 $ft^2$ | |  | Jaguar, Cray XT5 1PF, 6.0 MW |
| LANL Metropolis Center, 2002, 7 MW, 303,000 $ft^2$ | | | Roadrunner IBM, 1.4 PF, 3.6 MW |
| LLNL Terascale Simulation Facility, 2004, 25 MW, 253,000 $ft^2$ | | | BlueGene/L, IBM, 596 TF, 3.0 MW |
| Google "Project 2", 2006, 200,000(?) $ft^2$ | | | Undisclosed |

# Panel Discussion

**Future Directions for Leadership Computers – How Big Can These Systems Get?**

- **Power requirements?**
- **Reliability requirements?**
- **Programmability?**
- **Technology?**
  - **Hybrid systems**
  - **Memory bandwidth wall**
  - **Disk Bandwidth**
  - **Prices of systems**
  - **Vendor Viability**
  - **Application Scalability**