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.

Design of innovative nanomaterials
Understanding of microbial molecular and cellular systems
100 yr Global climate to support policy decisions
Predictive simulations of fusion devices

1 Petaflop/s Cray Supercomputer
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

<table>
<thead>
<tr>
<th>System</th>
<th>CPU</th>
<th>Memory</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cray XT3 Jaguar</td>
<td>2.6GHz</td>
<td>21TB Memory</td>
<td>120 TB</td>
</tr>
<tr>
<td>Cray X1E Phoenix</td>
<td>0.5GHz</td>
<td>2 TB Memory</td>
<td>44 TB</td>
</tr>
<tr>
<td>SGI Altix Ram</td>
<td>1.5GHz</td>
<td>1.1TB Memory</td>
<td>36 TB</td>
</tr>
<tr>
<td>IBM SP4 Cheetah</td>
<td>1.3GHz</td>
<td>7GB Memory</td>
<td>32 TB</td>
</tr>
<tr>
<td>IBM Linux NSTG</td>
<td>3.0GHz</td>
<td>128GB Memory</td>
<td>4.5 TB</td>
</tr>
<tr>
<td>Visualization Cluster</td>
<td>2.2GHz</td>
<td>5TB Memory</td>
<td>9 TB</td>
</tr>
<tr>
<td>IBM HPSS</td>
<td></td>
<td>5TB Memory</td>
<td>5 TB</td>
</tr>
</tbody>
</table>

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

August 2006 Summary
- 7 Systems
- Supercomputers 13,834 CPUs
- 26TB Memory
- 74 TFlops
- Total Shared Disk 250.5 TB
- Data Storage 5 PB
Phoenix – 18.5 TF Cray X1E vector system

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

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

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!
FY 2006 LCF system utilization

Jaguar Utilization
Phoenix Utilization
Phoenix Uptime
Jaguar Uptime

Allocations made and former users removed

Jaguar Software Upgrade
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

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

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

- 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

Leadership-class Computing Facility

Computing Environment
Common look and feel across diverse hardware

Grand Challenge Teams

Research team
National priority science problem
Tuned codes

Breakthrough Science

Software & Libs
User support

Platform support
Leadership Hardware

Leadership Computing Facility
FY2006 allocations and user engagement

Jaguar allocations

Phoenix allocations

NCCS Hosts Users Meeting for FY2006 Allocations on Jaguar and Phoenix

On February 16-17, 2006, the National Center for Computational Sciences (NCCS) held a workshop for project teams with FY 2006 allocations on the Oak Ridge National Laboratory’s (ORNL’s) Jaguar and Trinity (Phoenix) supercomputers. The purpose of the gathering was to establish good two-way communication between the NCCS staff, project scientists, and other users. The NCCS workshop included an overview of the center, Jaguar and Phoenix architectures and software, and details about the support services that are provided to the research teams. The workshop also included sessions on advanced topics such as high performance computing, parallel programming, and scientific computing tools, and application domain expertise, allowing users to take maximum advantage of their computer allocations.

The three days of the meeting included an overview of the center, Jaguar and Phoenix architectures 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 2 INCITE projects presented overviews of their research goals. The group spent an afternoon organizing a User Council and Task Council to find a way to communicate 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 in various topics ranging from porting and optimization questions. User representatives gave workshops on Jaguar and Phoenix. Users also had an opportunity to hear a detailed presentation on visualization using VESTAS.

Jeffrey Hughes, Director of the NCCS, was presiding over the meeting. He welcomed 80 participants to this first Users Meeting, with almost every LCF and INCITE project represented. “The wide range of projects presented opened up new perspectives for our users. Jaguar and Trinity are important to DOE—climate, fusion, nanoscale, materials, chemistry, and biology—but also important to industry such as aerospace design and simulation. We look forward to contributing to the success that ends of these teams will experience on their way to accomplishing breakthrough science in each of their application disciplines.”

Amidst the usual excitement about the meeting and the opportunity to give them a chance to interact with technical and support staff at the NCCS, several users of the Center praised the help line and told the new users, “If you need help, just ask for it.”
## FY06 Allocated Projects by Science Area

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

**Total:** $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)

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

Galaxy NGC 4
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

Next Generation Simulations in Biology
PI: Pratul Agarwal, ORNL
Quantum conductance of amphoterically doped single-walled carbon nanotubes

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

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

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

**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

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

**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

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

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