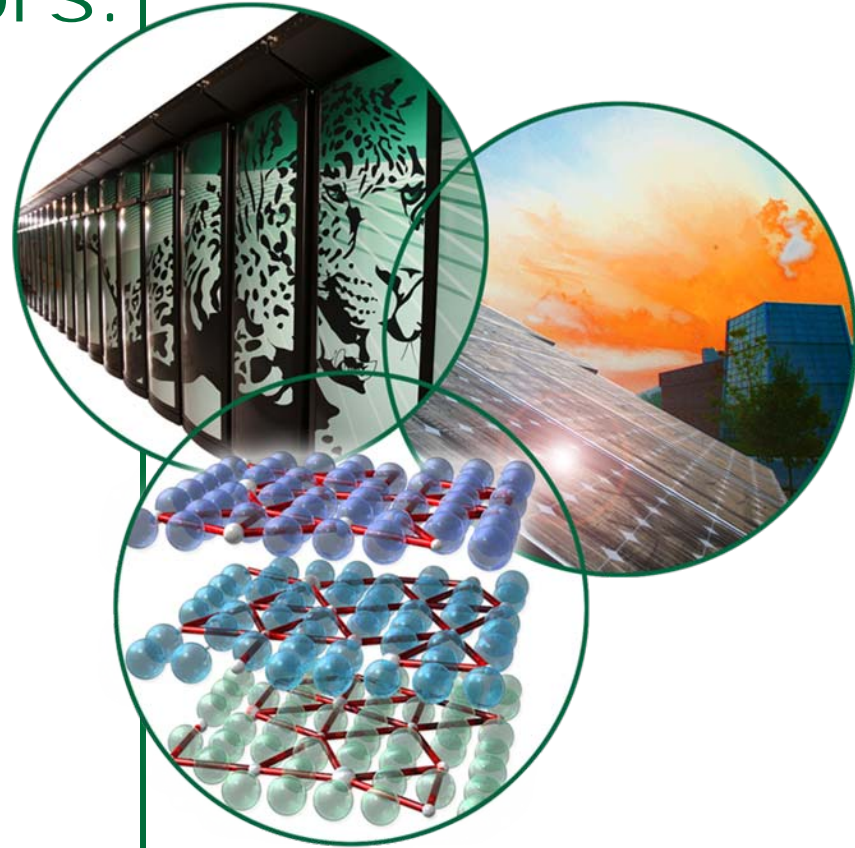


HPC Landscape - Application Accelerators: Deus ex machina?

Presented to the
**High Performance Embedded
Computing 2009 (HPEC)**

Jeffrey Vetter
Future Tech Group Leader, ORNL
Joint Professor, Georgia Institute of Technology

22 September 2009



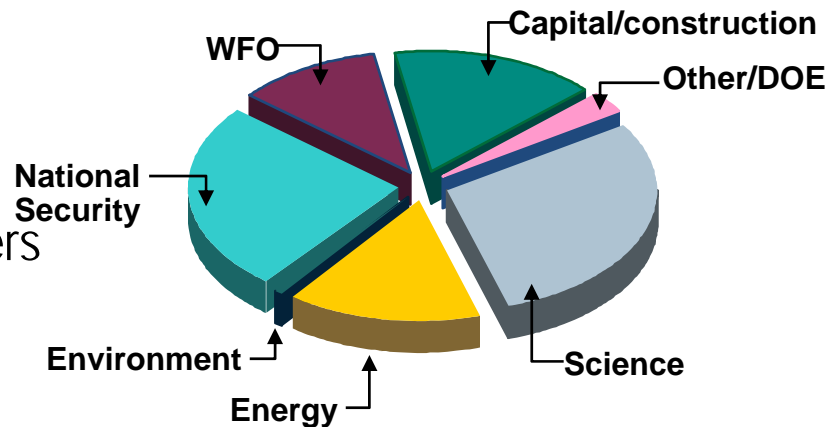
ORNL in the Early Days

- 1943 - Manhattan Project
 - Los Alamos
 - Oak Ridge
 - Hanford
- 1950-60s
 - Nuclear energy, medicine, materials
- 1970s - Divided into two sites
 - Y12 – NNSA
 - X10 – DOE Office of Science



ORNL Today

- Managed by UT-Battelle since April 2000
- \$350M modernization project
- Key capabilities
 - Neutron science
 - Ultrascale computing
 - Systems biology
 - Materials science at the nanoscale
 - Advanced energy technologies
 - National security
- 4,600 staff
- 3,000 guest researchers
- 30,000 visitors
- FY08 budget: \$1.4B



Spallation Neutron Source and Center for Nanophase Materials Sciences



Mission: Conduct basic and applied research and development to create scientific knowledge and technological innovations that enable the solution of compelling national problems

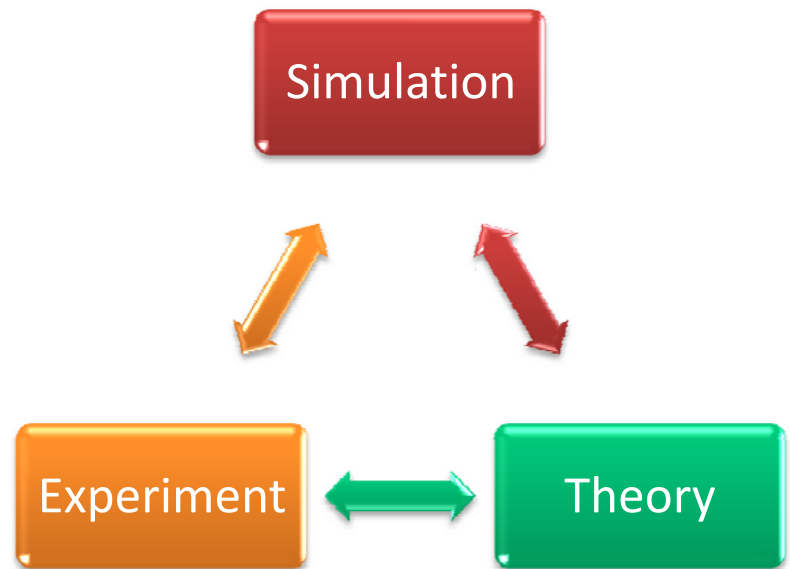
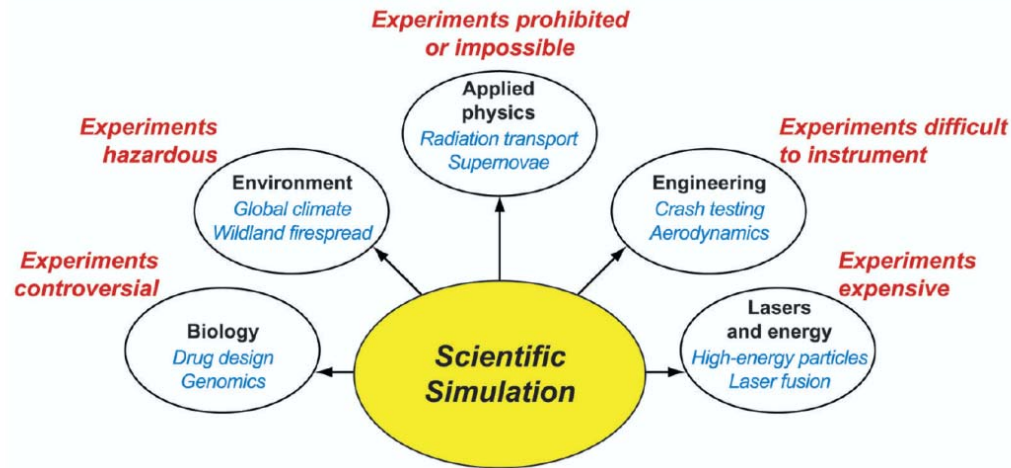
We have a three pronged strategy for sustained leadership and programmatic impact

- Provide the nation's most powerful open resource 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**



Scientific Simulation

- Simulation is becoming an accepted form of science
- Complex phenomena that
- Examples
 - Political limits
 - Nuclear test ban treaty
 - Experimentation limits
 - Climate modeling
 - Atomistic scale materials, etc



Motivating Example: Climate Modeling

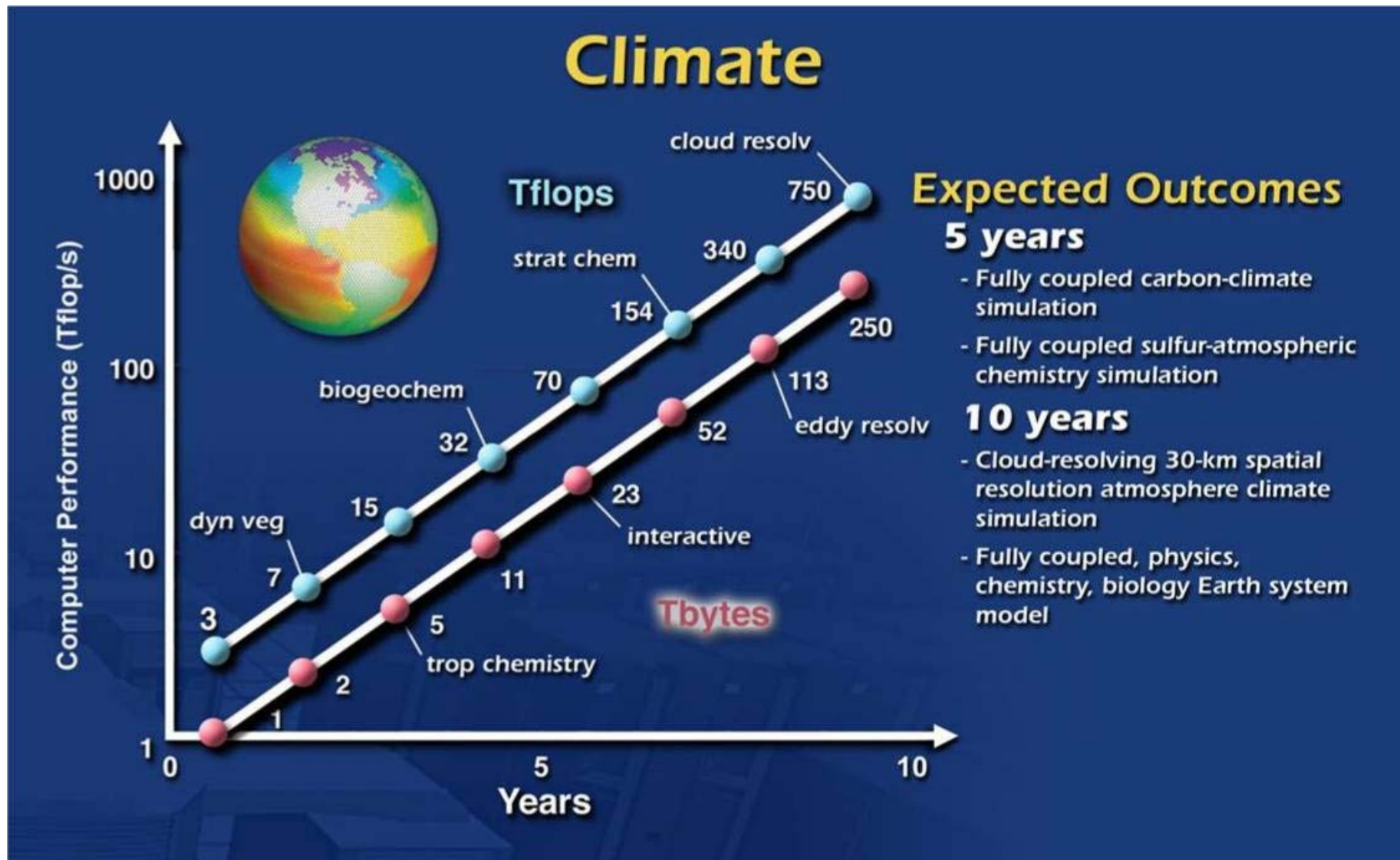
- Intergovernmental Panel On Climate Change
 - Sponsored by UNEP and WMO
 - Thousands of scientists from around the world
 - Won Nobel Prize w/ Gore in 2007
 - Provide policy-neutral scientific information
 - Assess basis, impact, adaptation, vulnerability, mitigation
- UN Conference TODAY



Climate Modeling - Future

- Observations and natural events provide effective tools
- Scientific simulation provides a critical tool to assess future scenarios
- Community Climate System Model is used in these assessments
- Simulation
 - ORNL has provided considerable computing resources over the past five years for IPCC simulations
 - NICS Cray XT5 at ORNL will be dedicated to climate simulations for 3 months
 - NOAA will provide \$215M over the next 5 years to sponsor climate research at ORNL

Climate Computational Projected Requirements



ORNL - Scientific and Technical Computing



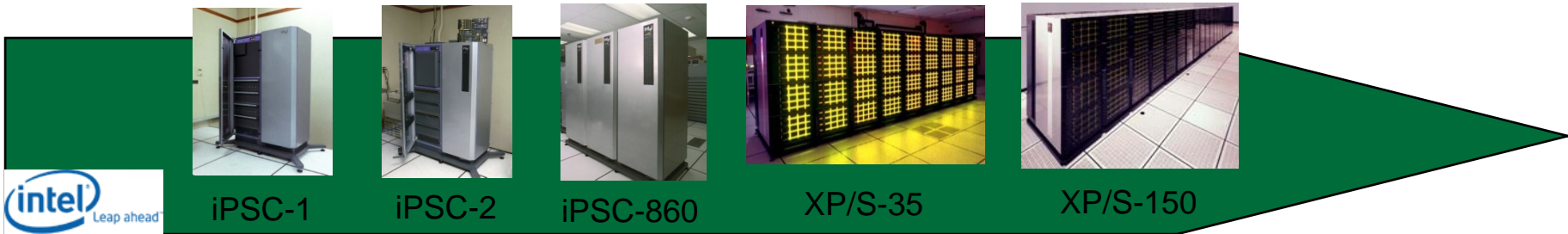
CRAY
THE SUPERCOMPUTER COMPANY

X-MP J90 X1 XD1 XT3 X1E XT4 "Baker"



IBM

360-195 SP2 4341 3033 Power3 and 3+ Power4 and 4+ BG/P



intel Leap ahead

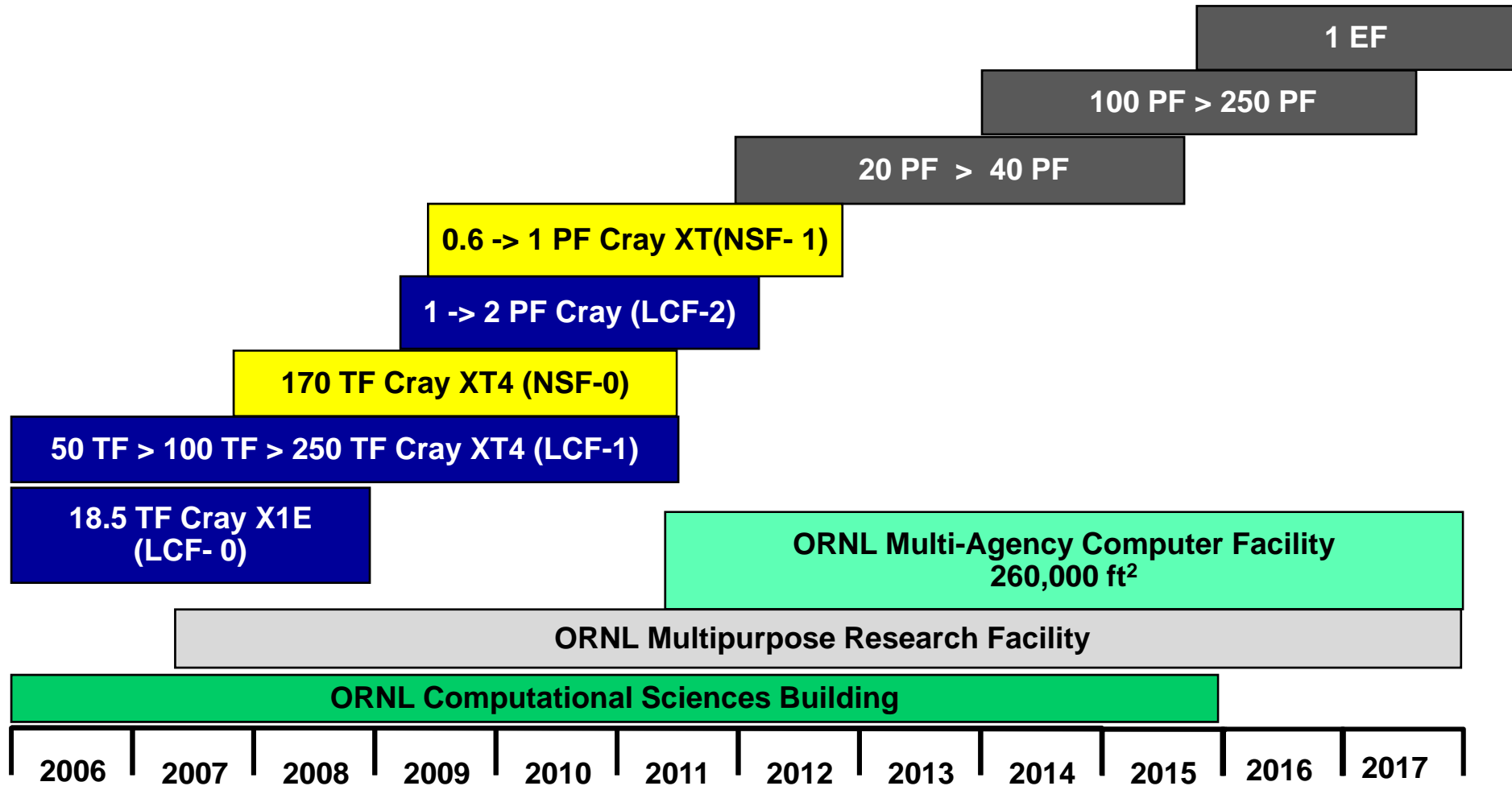
iPSC-1 iPSC-2 iPSC-860 XP/S-35 XP/S-150



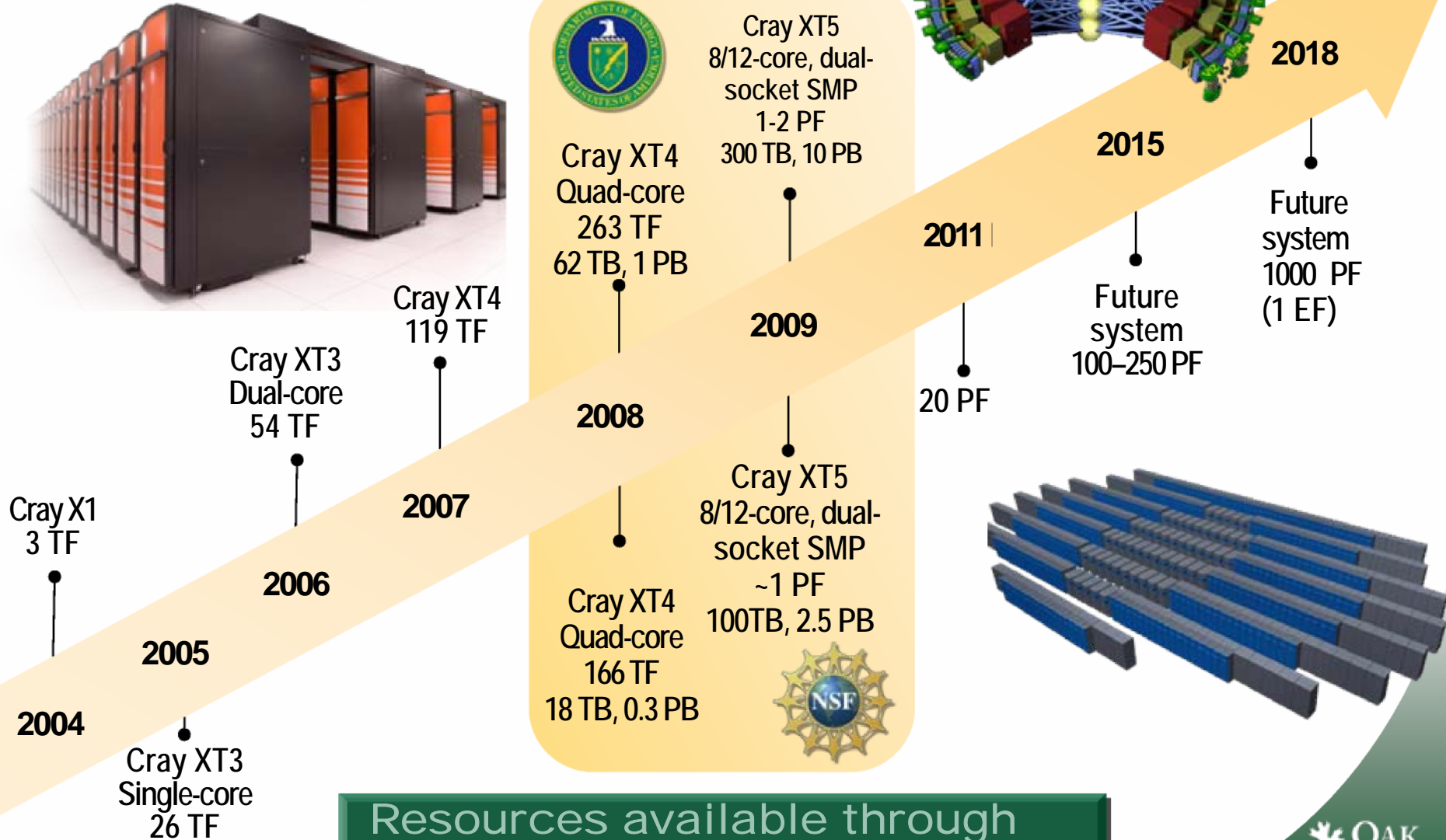
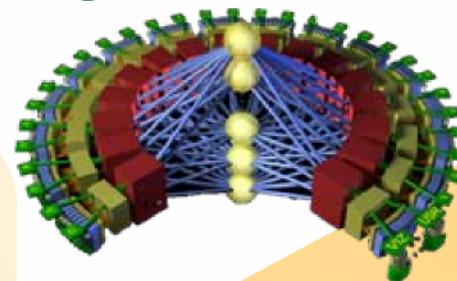
Others ORACLE KSR-1 SGI Origin Compaq AlphaServer SRC-6 SGI Altix

ORNL
BRIDGE
National Laboratory

ORNL Roadmap to Exascale

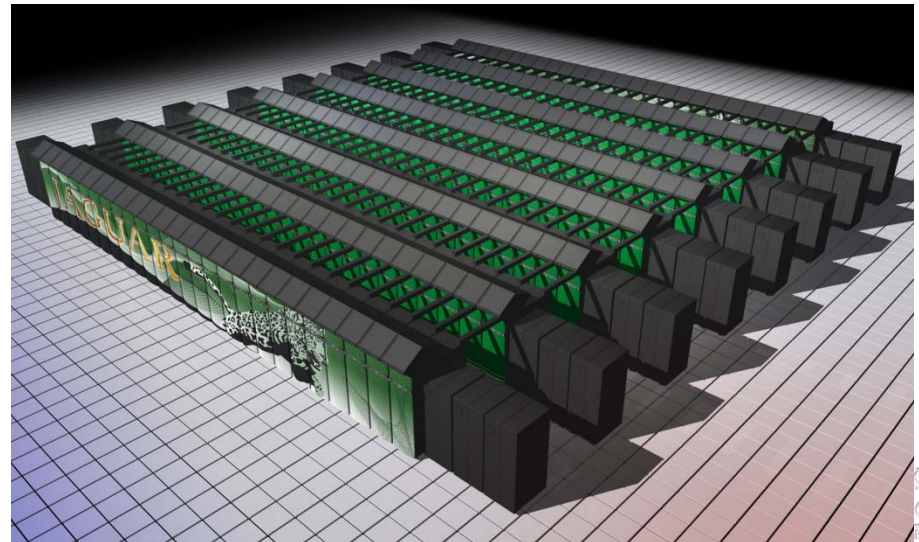
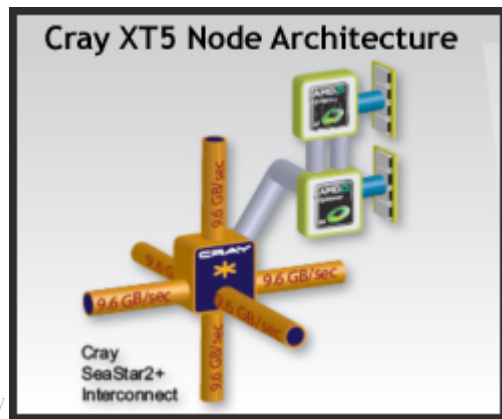


1000x increase in computing and data capabilities



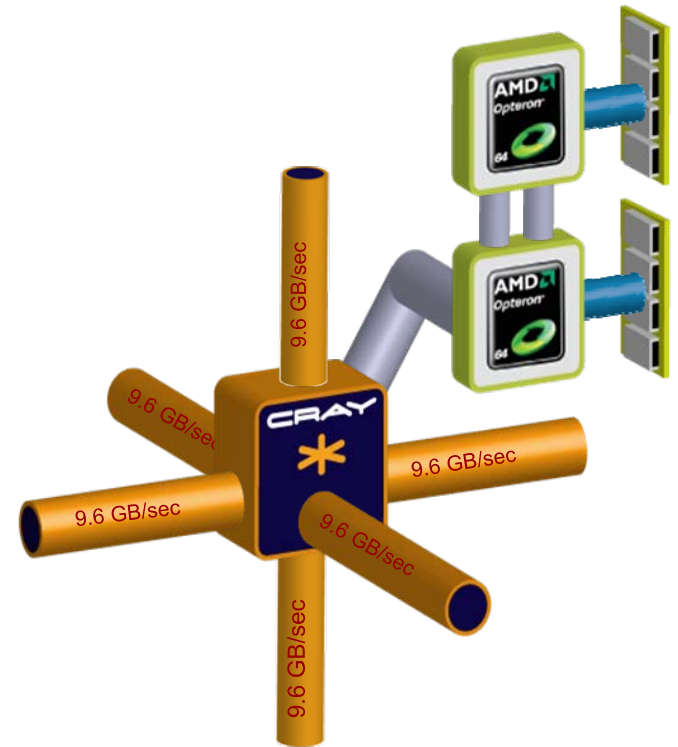
Jaguar Today

Jaguar Specifications	Total	XT5	XT4
Peak Teraflops	1,645	1,382	263
Quad-Core AMD Opterons	45,376	37,544	7,832
AMD Opteron Cores	181,504	150,176	31,328
Compute Nodes	26,604	18,772	7,832
Memory (TB)	362	300	62
Disk Bandwidth (GB/s)	284	240	44
Disk Space (TB)	10,750	10,000	750
Interconnect Bandwidth (TB/s)	532	374	157
Floor Space (feet ²)	5,800	4,400	1,400
Cooling Technology		Liquid	Air



Current Upgrade of XT5 to Istanbul Opteron is Underway – 3/5 of system already complete

- Upgrade of existing XT5 system to 6-core “Istanbul” processor (2H 2009 upgrade)
- Socket replacement of the 4-core 2.3 GHz “Barcelona” Opteron processors with 6-core 2.6 GHz “Istanbul” processors
- Upgrades XT5 from 1.375 PF to 2.3 PF
- Provides compute boost for applications and a platform for scaling applications to 225K cores
- Node peak performance goes from 73.6 GF to 124.8 GF, an increase of 70%
- 21-27% increase in memory bandwidth, maybe more if we can run in ungangled mode



Highly visible science output

PHYSICAL REVIEW LETTERS

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

T. A. Maier^{1,2}, D. Puffinger,³ and D. J. Scalapino^{1,2}

¹Center for Nonlinear Materials Science and Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6048, USA

²Yukawa Institute for Physics, Kyoto University, Uji, 611-0192, Japan, Japan

³Department of Physics, University of California, Santa Barbara, California 93106-8030, USA

(Received 20 January 2008; published 15 May 2008)

The question of whether one should speak of a "pairing glue" in the Hubbard and t - J models is basically a question about the dynamics of the pairing interaction. If the dynamics of the pairing interaction arise from virtual states, whose energies correspond to the Matsubara gap, and give rise to the exchange coupling J , the interaction is instantaneous on the relative time scale of interest. In this case, while we would like to see "instantaneous" in the context of the relative time scale of interest, it is not yet clear whether sufficient information is definitively and pairing, one says that the open fluctuations provide the pairing glue. The question of whether there is a pairing glue is answered by showing different pairing mechanisms are relevant to the pairing interaction in the Hubbard and t - J models. The frequency dependence suggests that the exponentiated by the Matsubara frequency dependence of the Nambu self-energy.

Physical Review Letters: High temperature superconductivity

ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Combustion and Flame

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^a, Philip J. Smith^a

^aDepartment of Chemical Engineering, University of Utah, Salt Lake City, UT 84095, USA


Combustion and Flame: 3D flame simulation

A four-step, three-moment, semiprecursor soot model is employed. Previous two-dimensional decaying turbulence simulations have shown the importance of multidimensional flame dynamical effects on soot. Liu, Chen, P.J. Smith, T. Lu, C.K. Law, Combust. Flame 151 (1–2) (2007) 2–20. It is strongly impacted by the diffusive motion of the flame relative to soot (which is slow), resulting in soot being differentially transported toward or away from the flame and to the flame directly influencing soot reactivity and radiative properties. Here, the dimensions in a temporal jet configuration with mean shear. Results show that similar of strain and curvature are important, but that enhanced turbulent mixing of fuel and effect on transport of soot toward flame zones. Soot modeling in turbulent flames complexity of soot formation and transport processes and the lack of detailed experimental data. The present direct numerical simulation provides the first step toward providing a better understanding of the soot formation process.

JULY 2008

VOLUME 16 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'Arco, R. E. Barrett, S. D. Allen, D. W. Swain, D. B. Hatchler, R. W. Harvey, J. R. Myra, D. A. D'Angelo, C. K. Phillips, E. Valeo, D. V. Suthier, P. T. Bonoli, J. C. Wright, and H. Qiu

MEMBER SUBSCRIPTION COPY
Library or Other Institutional Use Prohibited Until 2013

AMERICAN INSTITUTE OF PHYSICS

Physics of Plasmas: ICRF heating in ITER

LETTERS

Pulsar spins from an instability in the accretion shock of supernovae

John M. Blondin¹ & Anthony Mezzacappa²

Rotation-powered pulsars are born with individual initial spin periods of order 20 ms (some as short as 20 ms in some cases). 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. An alternative picture has been proposed to explain the distribution of pulsar spins, suggesting that the birth rotation is either too slow or too fast. Here we report a robust mechanism for the initial accretion shock in core-collapse supernovae that is able to generate a strong rotational flow in the vicinity of the accretion-protonation zone. 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. If, at any time, the accretion correlation between the spin period of a neutron star and the progenitor star is broken, the accretion correlation between the spin period of a neutron star and the progenitor star is broken. The collapse of a massive star's core that triggers a supernova is followed by a brief episode of low shear in which the accretion shock wave stalls at a radius that is not too far from the center, and the supernova is initiated. Hydrodynamical simulations show that this stalled shock is subject to the magnetorotational instability, or MRI. However, these two simulations show only axisymmetric results and, hence, dynamics cannot affect the rotation of the accretion zone. In three-dimensional non-axisymmetric simulations, the accretion shock is subject to a series of instabilities that lead to a final spin period that is consistent with observations. We find that the nonlinear evolution of the MRI is dominated by a low-order non-axisymmetric mode characterized by a quadrupole structure.

LETTERS

Clumps and streams in the local dark matter distribution

J. Diemand, M. Kuhlen, P. Madau, M. Zemp, B. Moore, D. Potter, & J. Stadel

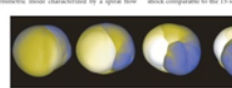
In cold dark matter cosmological models, the structure forms and evolves in a hierarchical fashion. Numerical simulations within the last decade have shown that the local dark matter distribution is not smooth, but is instead characterized by a complex structure of clumps and streams. The local dark matter distribution is not smooth, but is instead characterized by a complex structure of clumps and streams. The local dark matter distribution is not smooth, but is instead characterized by a complex structure of clumps and streams.

Nature: Astrophysics



The cold dark matter (CDM) model has been remarkably successful at describing the large-scale mass distribution of our Universe from the first Big Bang to the present. However, the nature of the dark matter particles is best tested on small scales, where its interaction properties manifest themselves by modifying the structure of galaxy halos and their substructures. CDM theory predicts that the growth of cosmic structures begins early, on Earth-like mass scales, and continues from the horizon up until galaxy clusters form that are 20 orders of magnitude more massive. Simulating multi-scale structures is extremely challenging, as the range of length scales, and structures that need to be simulated is immense. We have performed the highest-resolution CDM theory simulation to date via Lattice II of the assembly of the Galaxy (LII). The simulation includes the growth of a Milky Way system from redshift 100 to the present. It provides the most accurate predictions on the small-scale distribution of dark matter as far as density and phase structure in the local halo environment and properties. We used the parallel tree code PENCILBOX and the smoothed particle hydrodynamics (SPH) code GADGET to simulate the evolution of the system.

Figure 1: The evolution of the accretion shock illustrates the relation of the spin period of the SAS. The blue portion of the shock surface represents the leading edge of the SAS, and the red portion represents the trailing edge. The SAS is shown propagating from right to left across the front face of the shock. The accretion shock is shown propagating from right to left across the front face of the shock. The accretion shock is shown propagating from right to left across the front face of the shock.



Transformational science enabled by advanced scientific computing

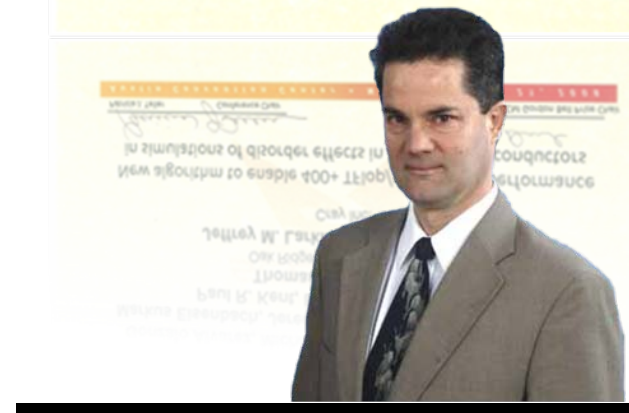
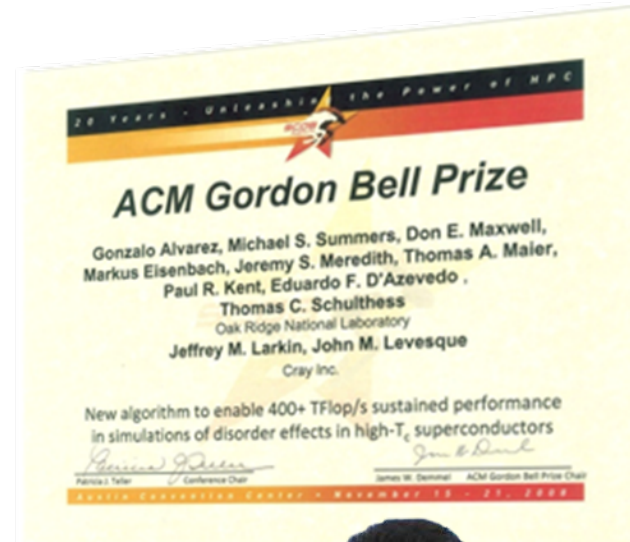


2008 Gordon Bell Prize: 3 of 6 finalists ran on Jaguar

- Prize awarded to ORNL team led by Thomas Schulthess for DCA++ simulation of disorder effects in high-temperature superconductors
 - Simulation achieved 1.352 petaflops MP on ORNL's Cray XT Jaguar
 - Modifications of algorithms and software design boosted performance tenfold

Finalists

DCA++	ORNL	Jaguar
LS3DF	LBL	Jaguar
SPECFEM3D	SDSC	Jaguar
RHEA	TACC	Ranger
SPaSM	LANL	Roadrunner
VPIC	LANL	Roadrunner



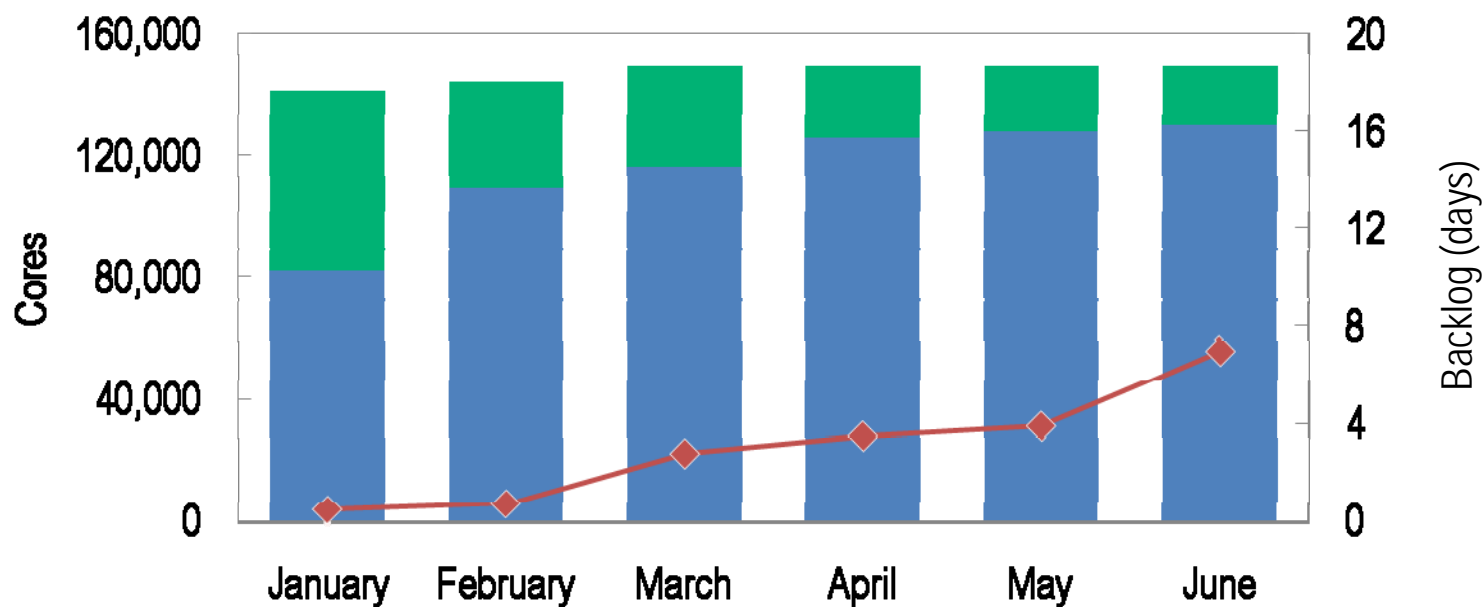
Science applications are scaling on Jaguar

Science area	Code	Contact	Cores	Performance
Materials	DCA++	Schulthess	150,144	1.3 PF MP
Materials	LSMS	Eisenbach	149,580	1.05 PF
Seismology	SPECFEM3D	Carrington	149,784	165 TF
Weather	WRF	Michalakes	150,000	50 TF
Climate	POP	Jones	18,000	20 simulation years/day
Combustion	S3D	Chen	144,000	83 TF
Fusion	GTC	PPPL	102,000	20 billion particles/second
Materials	LS3DF	Lin-Wang Wang	147,456	442 TF
Chemistry	NWChem	Apra	96,000	480 TF
Chemistry	MADNESS	Harrison	140,000	550+ TF



Jaguar XT5 status, April 2009

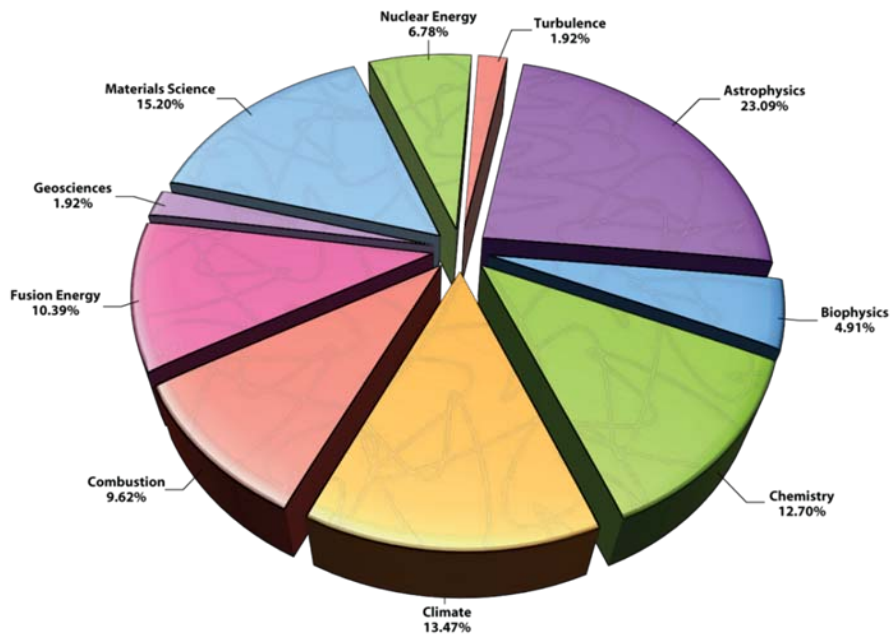
Stable	<ul style="list-style-type: none"> • MTTI: 32 hours • MTTF: 52 hours 	<ul style="list-style-type: none"> • Driver for downtimes: Spider testing • System has been up as long as 10 days
Available	<ul style="list-style-type: none"> • Scheduled at 94% • Overall availability at 86% 	<ul style="list-style-type: none"> • Driver for downtimes: Spider testing • Target completion date: Mid-July
Usable	<ul style="list-style-type: none"> • Fully loaded utilization: 84% 	<ul style="list-style-type: none"> • Steadily climbing since start of transition to operations on January 12, 2009
In demand	<ul style="list-style-type: none"> • Current queue backlog: 3.9 days 	<ul style="list-style-type: none"> • Delivered 73M hours in April 2009



Jaguar XT5 science workload

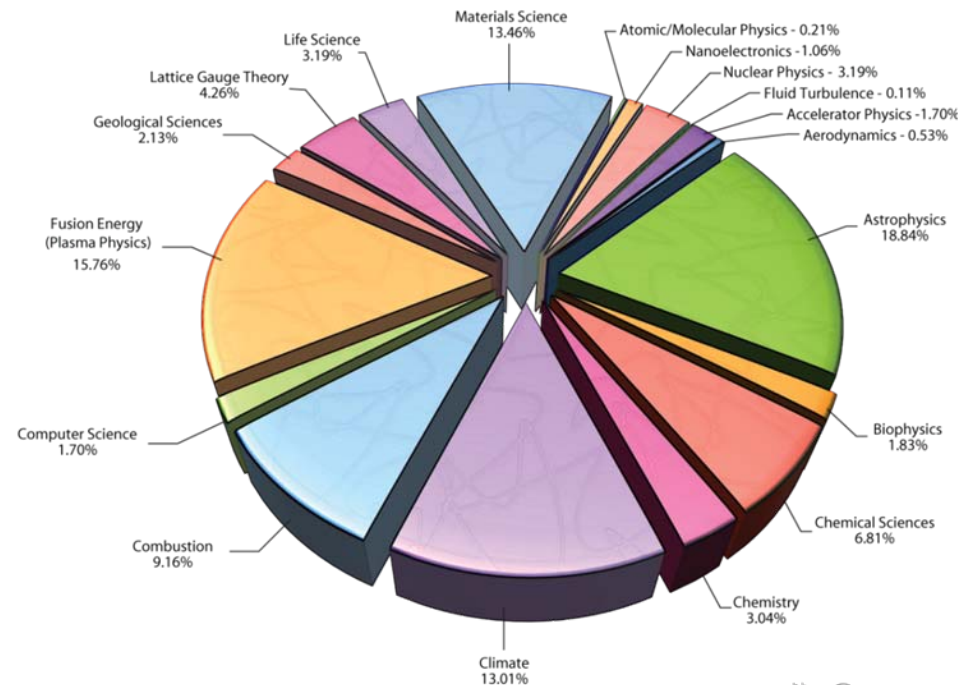
Current: Transition to operations

- 27 early-access projects, January–July
 - Total allocation: 540.8M hours
 - Usage since January: 247M hours (159 users)
- ALCC project: FY09 Joule metric applications
 - Total allocation: 25M hours
 - Usage since January: 17M hours (17 users)



Fall 2009: General availability

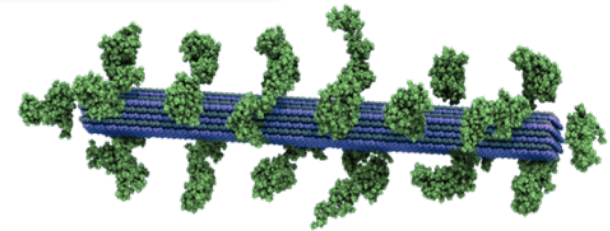
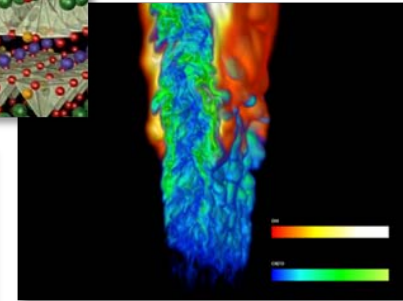
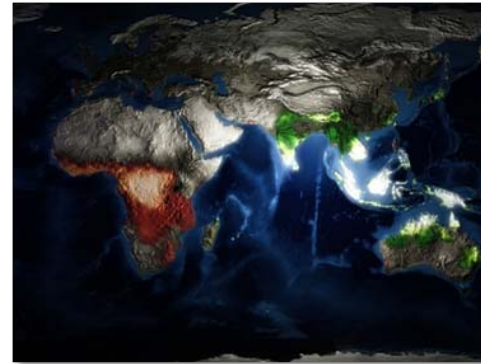
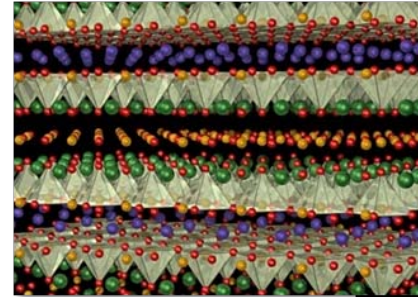
- 38 INCITE projects transitioned to system
 - Total allocation: 469M hours
 - Usage (XT4) since January: 91M hours (448 users)
- Discretionary projects: Climate AR5 production
 - Total allocation: 80M hours
 - 2 projects: NCAR/DOE and NOAA (~50 users)



Jaguar XT5 early science projects: Tackling compelling problems

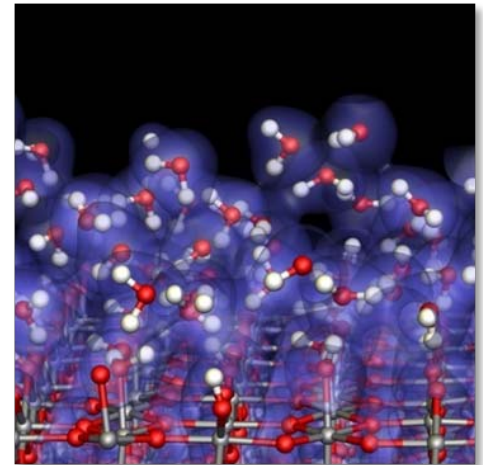
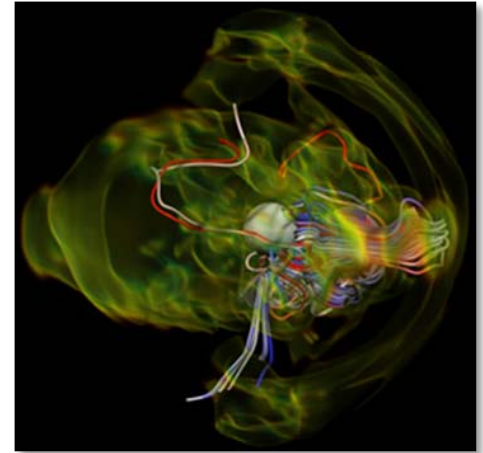
- Energy for environmental sustainability

- Climate change: Carbon sequestration, weather event impacts on global climate, decadal climate predictive skill in aerosol forcing, global climate at unprecedented resolution, mantle convection
- Energy storage: Charge storage and transfer in nano-structured supercapacitors
- Combustion: Stabilizing diesel jet flames for increased efficiency and decreased emissions
- Bioenergy: Recalcitrance in cellulosic ethanol
- Solar: Nonequilibrium semiconductor alloys
- Energy transmission: Role of inhomogeneities in high-temperature superconducting cuprates
- Fusion: ITER design, optimization, and operation
- Nuclear energy: Fully resolved reactor core neutron state



Jaguar XT5 early science projects: Tackling compelling problems

- Materials and nanoscience
 - Structure of nanowires, nanoelectronics, nanorods, and strongly correlated materials (magnets)
- Fundamental science
 - Astrophysics: Decipher core-collapse supernovae and black hole mergers
 - Chemistry: Water structure in biological and aqueous-phase systems; water bonding on graphene and carbon nanotubes
 - Nuclear physics: Probe the anomalously long lifetime of carbon-14
 - High energy physics: Generate lattices at the physical values of the light quark (u, d, s) masses
 - Turbulence: Dispersion relative to air quality modeling and bioterrorism; physics of shock/turbulence interaction
 - Biology: Effect of an embolization on hemodynamics in the entire neurovasculature



Reality: Facilities Designed for Exascale Computing

Open Science Center (40k ft²)

- ▶ Upgrading building power to 15 MW
- ▶ 210 MW substation, upgradeable to 280 MW
- ▶ Deploying a 6,600 ton chiller plant
- ▶ Tripling UPS and generator capability



National Security Center (32k ft²)

- ▶ Capability computing for national defence
- ▶ 25 MW of power and 8,000+ ton chillers

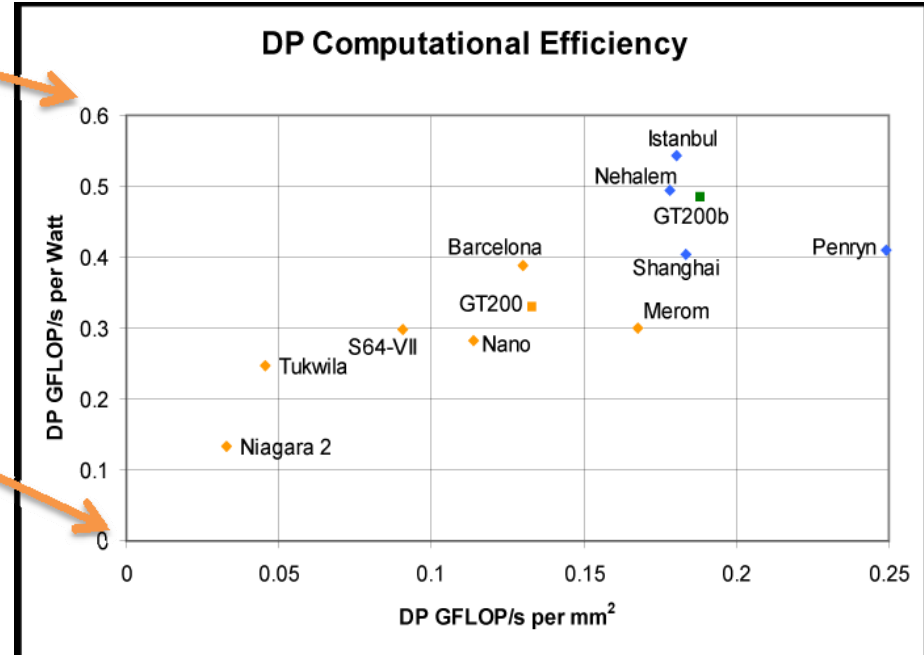
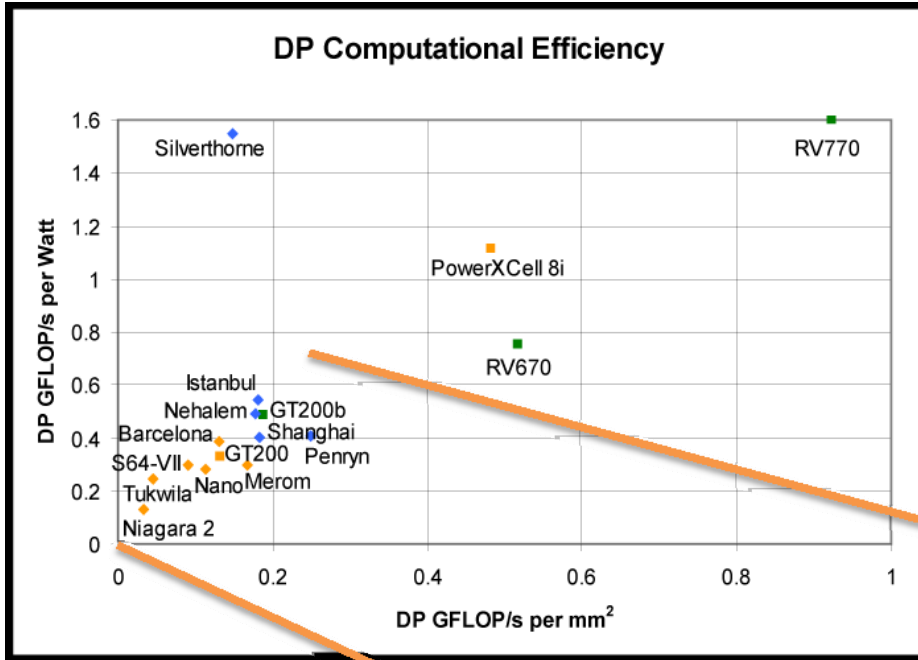


New Computer Facility (260k ft²)

- ▶ 110 ft² raised floor classified; same unclassified
- ▶ Shared mechanical and electrical
- ▶ Lights out facility



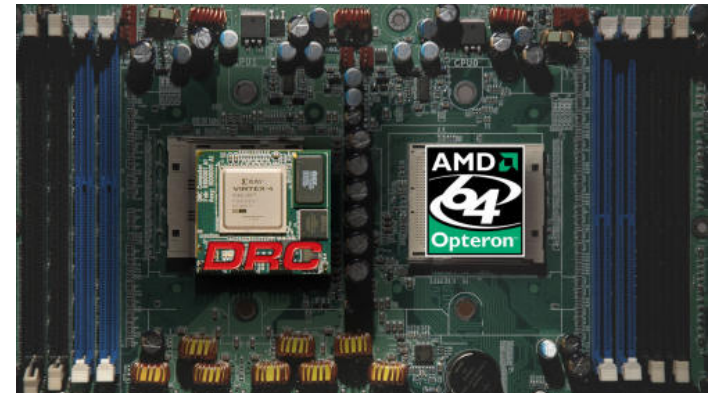
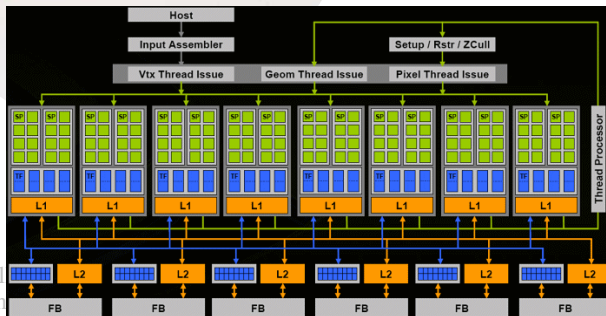
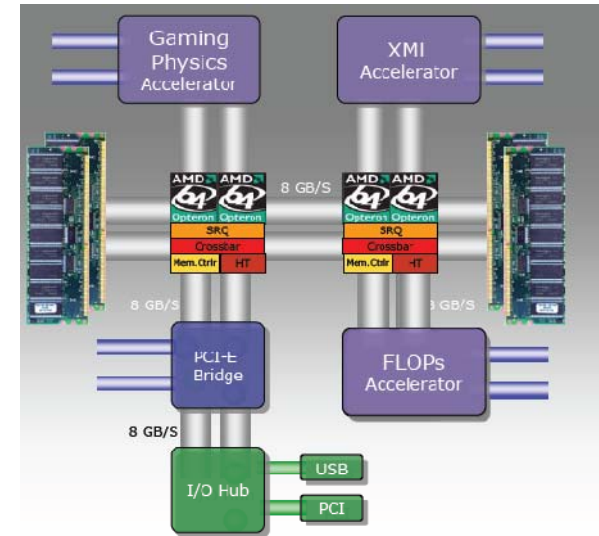
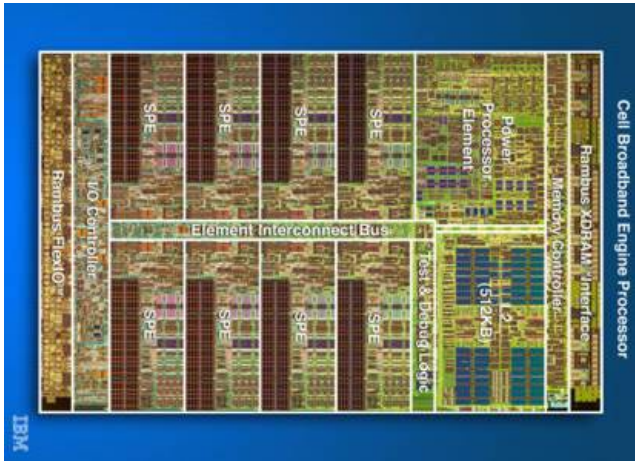
Power Efficiency! See Dr. Kogge's Talk



Source: <http://www.realworldtech.com>

Architectural Complexity - Heterogeneity, Specialization, Multi-paradigm

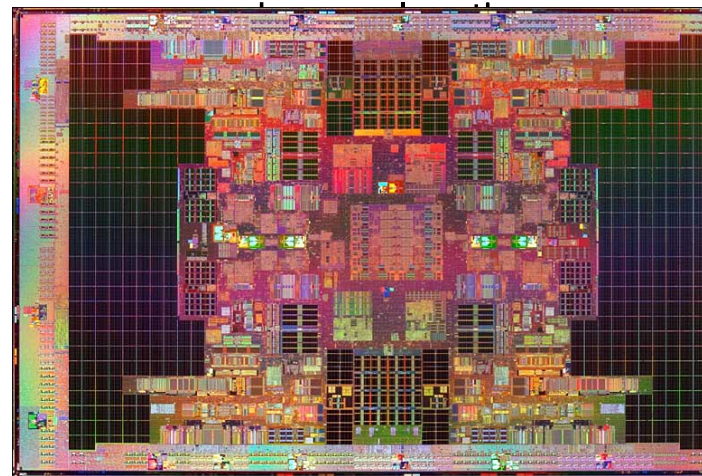
Architectures target specific workloads: games, graphics, business, encryption, media



Trends will easily scale performance achieved to 1 TFlop on a node/socket.

Several Factors Motivate Specialization

- Sacrifice generality and compatibility to address specific algorithms
- Computational power
 - Asymmetric, dynamic cores can provide performance advantages [1]
 - Use transistors differently
- Tukwila – First 2 billion transistor chip
 - 80486 had ~1.2M transistors, ~50MHz, 1989
 - Specialization can be free
- Power
 - Combination of these features provide more ops per watt for your targeted application
- Memory
 - Include different cache or memory organization tuned for data access patterns
- Performance of general purpose

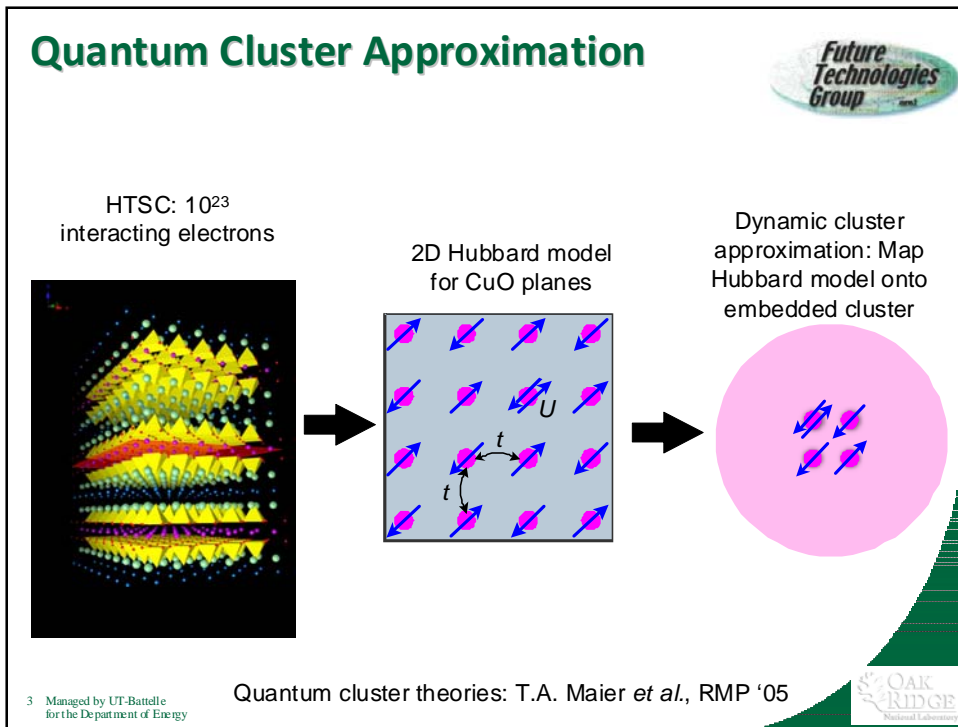


Applications

- In depth application analysis of existing applications

Application	Domain	Usability and Availability			Motifs										Models							
		Scalability	Restrictions	Readiness	Structured Grid	Unstructured Grid	Multiresolution Grid	Dense Matrix	Sparse Matrix	Spectral/FFT	MD/N-Body	MapReduce / Monte Carlo	Graphs	PIC	Algo Load Imbalance	Adaptive	C/C++	FORTRAN	1-Sided Model Pref	MPI	OpenMP	
MILC	QCD	3	3	3	3				3	3				3			3			3		
NEK5000	Nuclear Physics / CFD	3	3	3					3		3							3			3	
AMG	Multigrid solver	3	3	3						3											3	
IRS	implicit radiation solver	3	3	3	3				3												3	
LAMMPS	classical MD code	3	3	3							3	3						3			3	
Sphot	monte carlo photon transport	3	3	3									3								3	
UMT	radiation transport	3	3	3		3															3	
GTC	Fusion	3	3	3	2	2			3					3				1	3		3	3
MILC	QCD	3	3	3	3				3	3				3				3			3	
DCA++	Materials	3	2	3					3				3								3	
GTC	Fusion	3	2	3	2	2			3					3				1	3		3	3
POP	Climate	2	3	3	3				3					3					3		3	3
S3D	Combustion	3	2	3	3														3		3	
MILC	QCD	3	3	3	3				3	3				3				3			3	
GAMESS	Chemistry	1	3	3					3						1				3	3		2
HYCOM	Ocean	1	3	3	3																3	
WRF	Weather	3	3	3		3	3									1		3	3		3	1
FLASH	Astrophysics	3	1	3											3						3	
Flash IO	IO Benchmark	3	1	3																	3	
NAMD	Bio	1	3	3																	3	
QBOX	Materials	3	1	3					3		3			3				3			3	
AMR Solver Kernel	Solver	2	3	3	3		3							3	3						3	
IOR	IO Benchmark	2	3	3														3			3	
PARATEC	Materials	2	3	3	3				3											3		3
Flash	Astrophysics	3	2	3					3							3					3	
Flash IO	IO Benchmark	3	2	3					3												3	
LSMS	Materials	3	2	3	3				3									2	2		3	
WRF	Weather	3	3	3		3	3								1		3	3	3		3	

Computational Materials - Case Study



CPU: single 2.0GHz Opteron

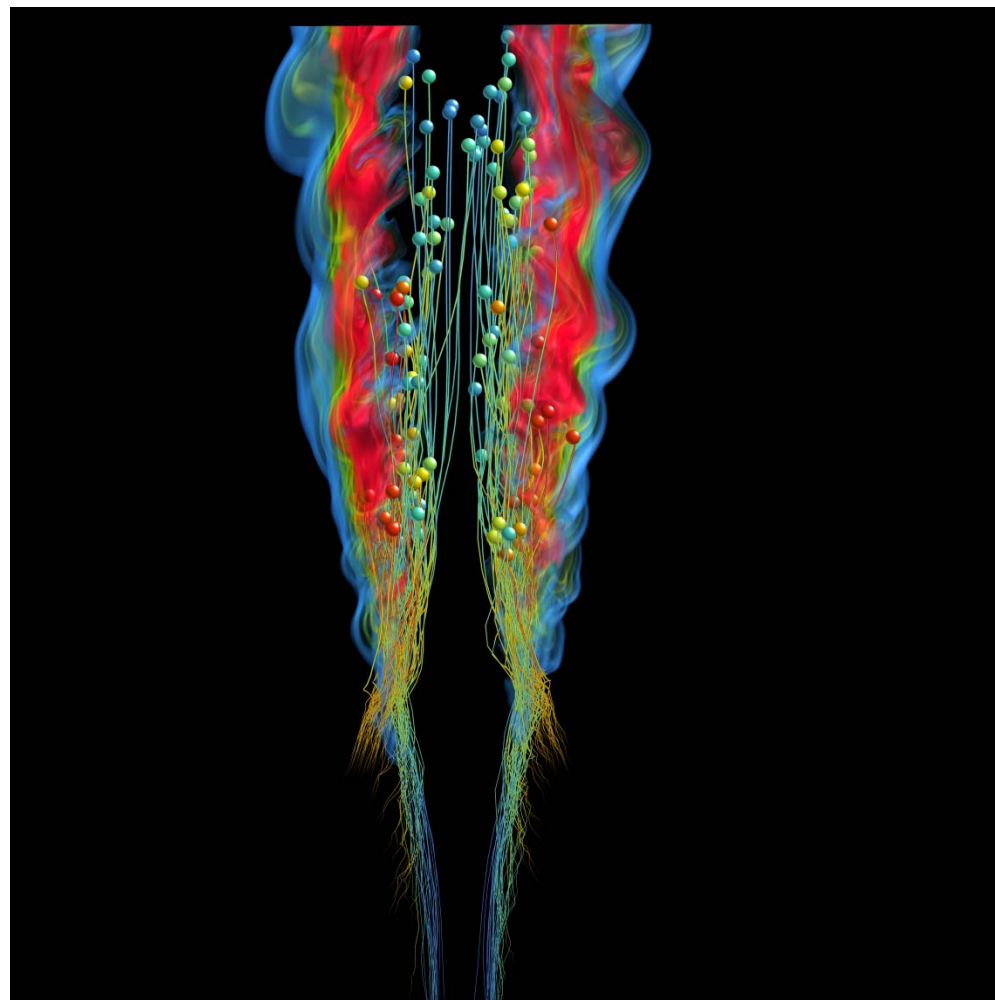
GPU: single 500MHz/96SP G80 (8800GTS 320MB)

- 10x speedup
 - Offload BLAS to GPU, transfer all data
- 13x speedup
 - Offload BLAS, smart data orchestration
- 19x speedup
 - Offload BLAS + additional functions, smart data orchestration

DCA++: J.S. Meredith, G. Alvarez *et al.*, "Accuracy and performance of graphics processors: A Quantum Monte Carlo application case study," *Parallel Comput.*, 35(3):151-63, 2009.

S3D – A case study

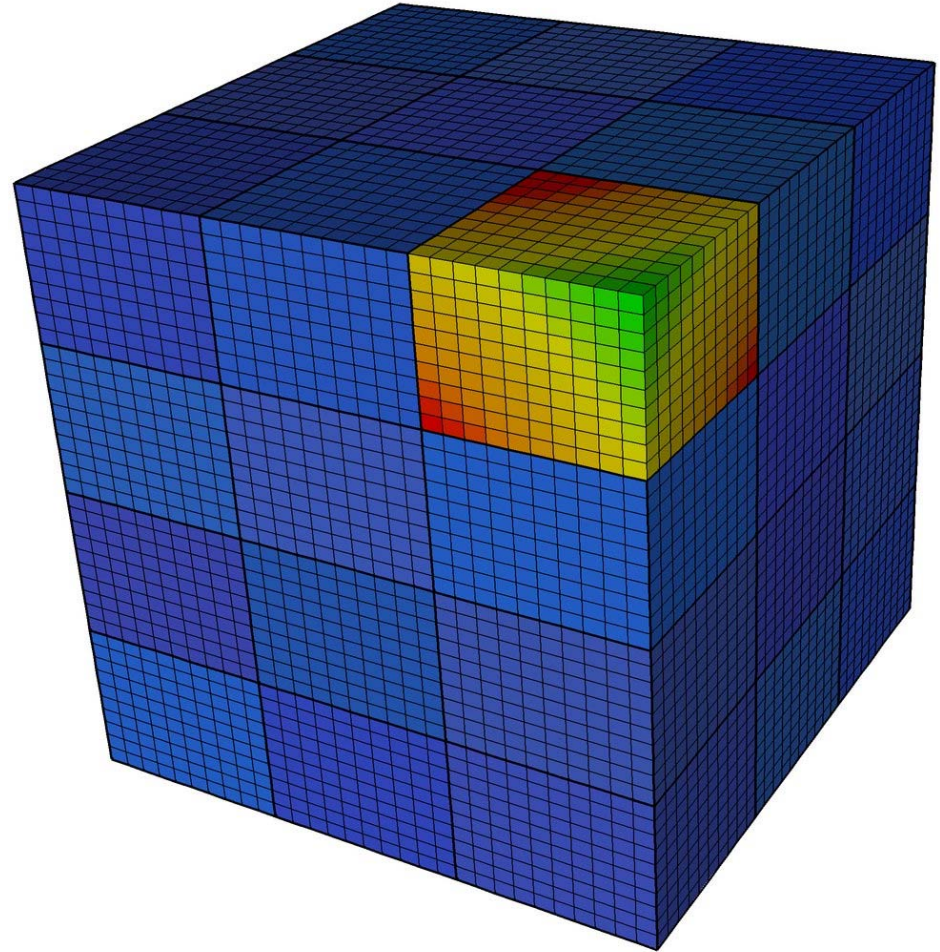
- S3D is a massively parallel direct numerical solver (DNS) for the full compressible Navier-Stokes, total energy, species and mass continuity equations coupled with detailed chemistry.
- Dept. of Energy marquee application for combustion
- Highly scalable code, written entirely in Fortran with MPI



K. Spafford, J. Meredith *et al.*, “Accelerating S3D: A GPGPU Case Study,” in *Seventh International Workshop on Algorithms, Models, and Tools for Parallel Computing on Heterogeneous Platforms (HeteroPar 2009)*. Delft, The Netherlands, 2009

Challenges: Tailoring S3D to the GPU

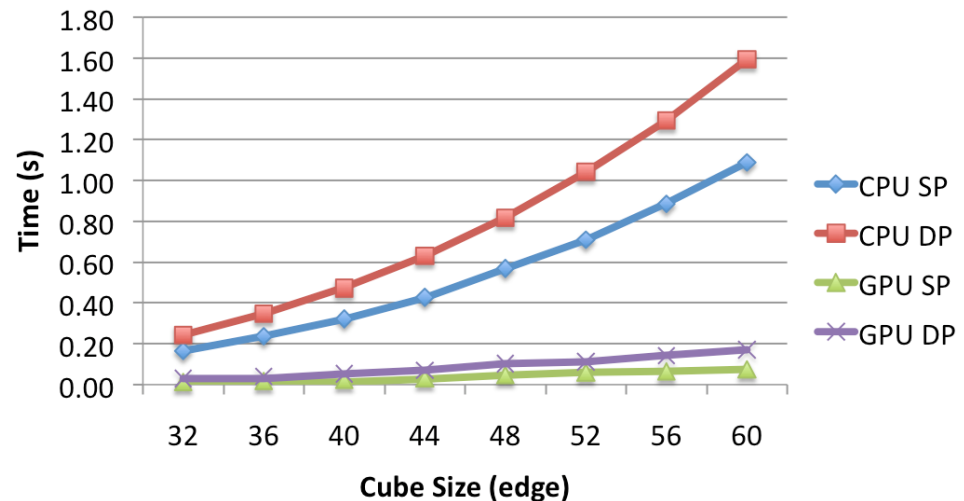
- Main Tasks
 - S3D has a huge code base written in Fortran, porting the entire application without proof-of-concept is infeasible—must choose a kernel for acceleration
 - Expose fine-grained data parallelism
 - Modify algorithms to exploit the GPUs memory hierarchy
- Chemistry calculations – GetRates kernel
 - Based on temperature, pressure, and mass fraction, GetRates computes the speed at which chemical reactions are occurring
 - Comprises roughly 45-60% of runtime, varying with grid size
 - Expected to increase with more complex chemistry
 - Operates on a regular, structured 3-D grid



Exploiting the GPU Memory Hierarchy

- Coalesce reads/writes to global memory
 - Single most important factor for performance
- Interleave accesses to global memory with intense calculation
- Use shared memory when possible
- Constant memory: almost as fast as register access on cache hit
- Transfer data to/from GPU only when absolutely necessary

- SP: 13-14.5x DP: 8.1x-9.3x



Keeneland – In a nutshell

- NSF Track 2D/XHPC
- NSF award
 - \$12M
 - 5 years from 2010
- Team
 - Georgia Tech, Lead
 - ORNL, UTK, NICS
- Field two GPGPU accelerated HPC systems
 - Initial Delivery, 2010
 - Final Delivery, 2012
 - FD decision point
 - Operations and support
 - Apps and software development
- Strong emphasis on outreach and education
 - Tutorials, Classes, Workshops



Georgia Institute
of Technology®



DOE/ORNL is considering this option

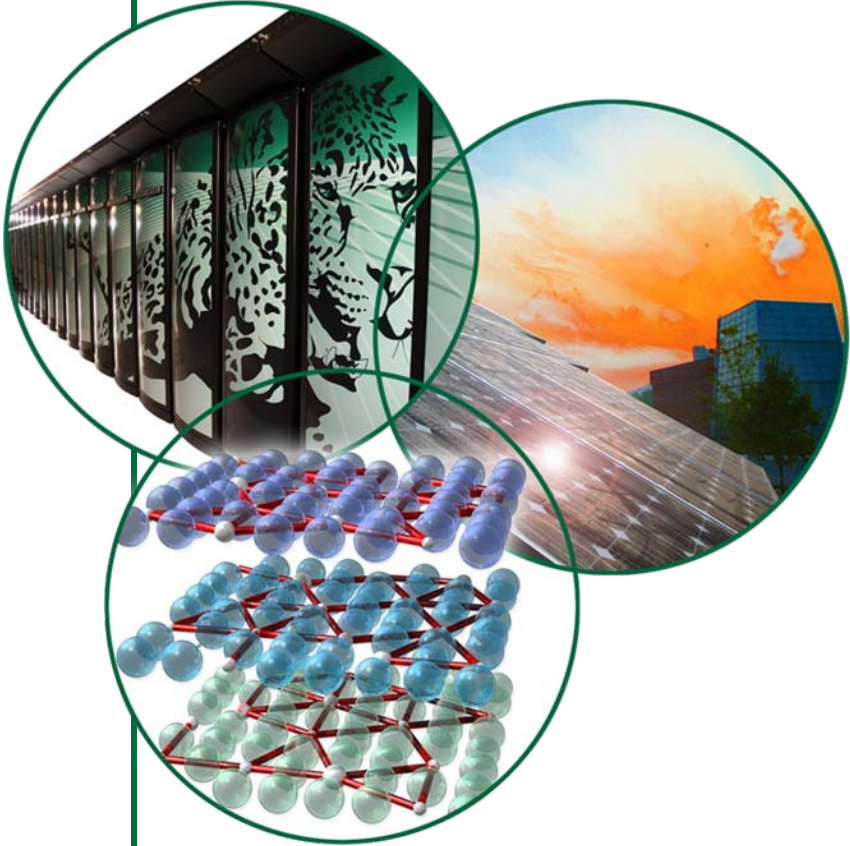
- 2011
- 10-20 PF
- Multicore Commodity processors paired with accelerators
- Applications
 - Tier 1
 - S3D
 - DCA++
 - CAM-HOMME
 - LAMMPS
 - MADNESS
 - Tier 2
 - Chimera
 - Denovo
 - AORSA
 - gWL-LSMS
 - GTC

Deus ex machina?

- Drama
 - Increasing appetite for scientific simulation
 - Increasing size of systems leading to huge facilities
- Enter accelerators/GPUs
- Will GPUs untangle this plot?

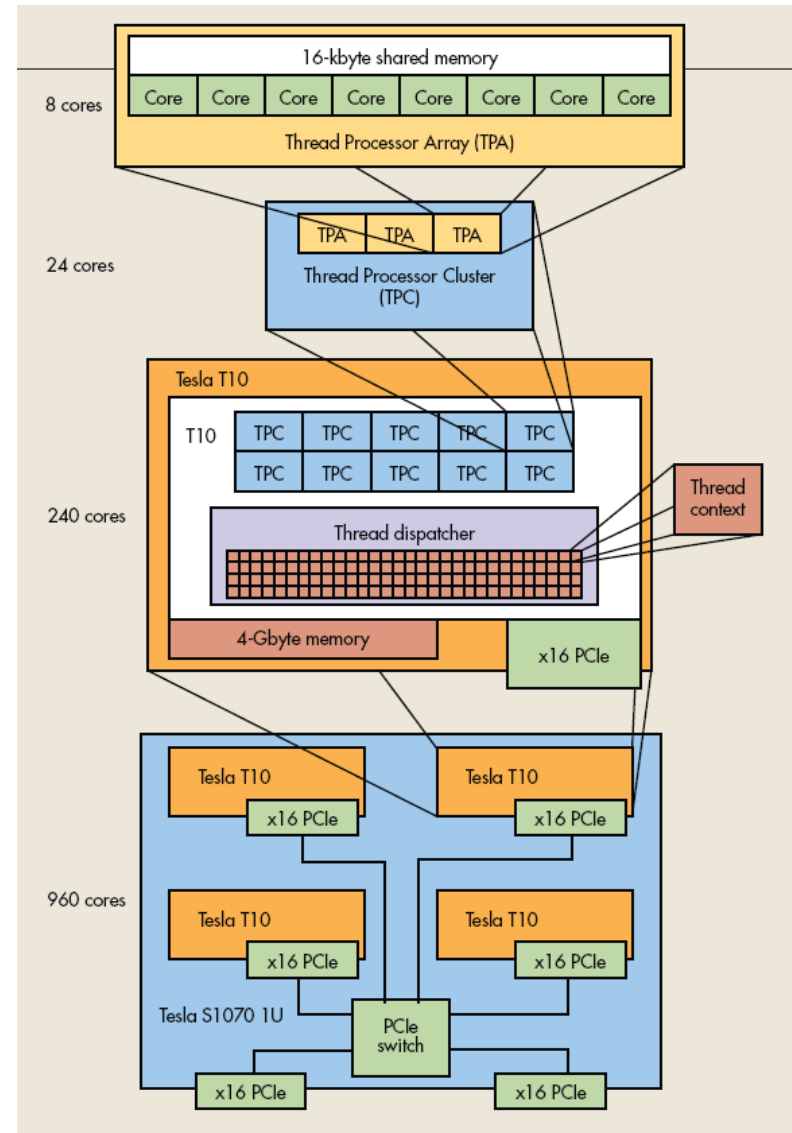


Bonus Slides



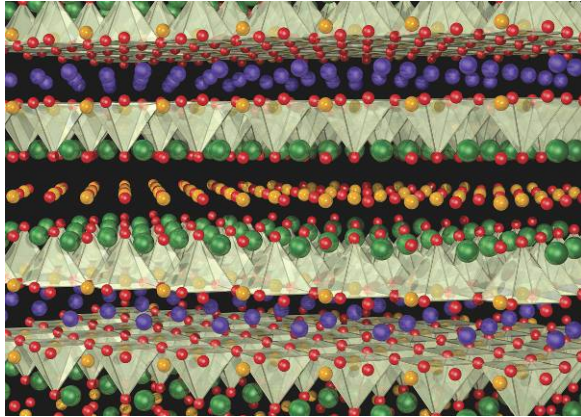
Concurrency - Observations

- S3D has many levels of concurrency
 - MPI
 - Host – GPU including data orchestration
 - SIMT
 - One GPU can have 1000s of thread contexts
 - Cores have internal concurrency

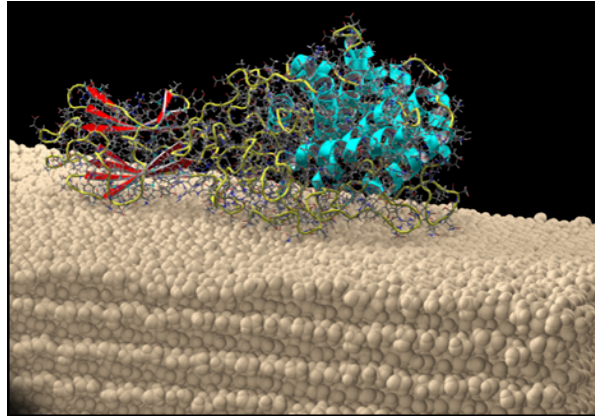


<http://electronicdesign.com>

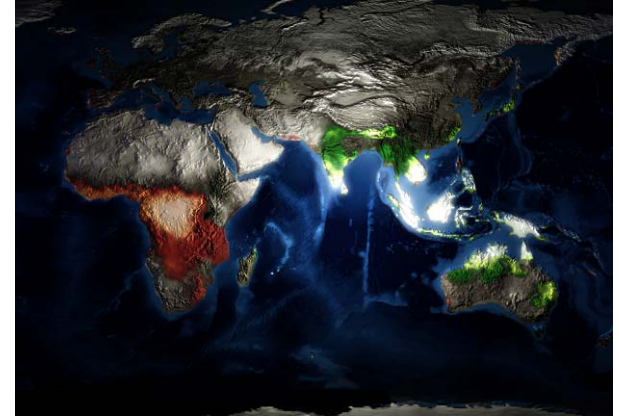
Leadership computing is advancing scientific discovery



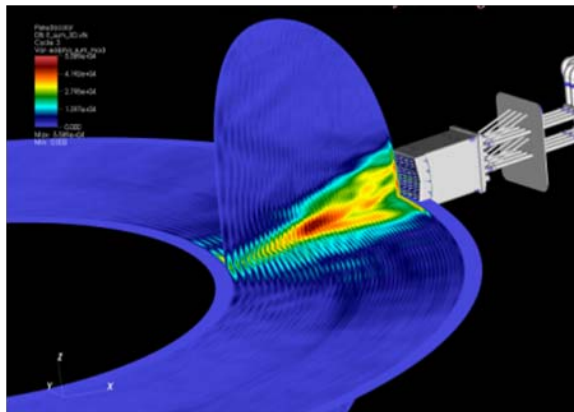
Resolved decades-long controversy about modeling physics of high temperature superconducting cuprates



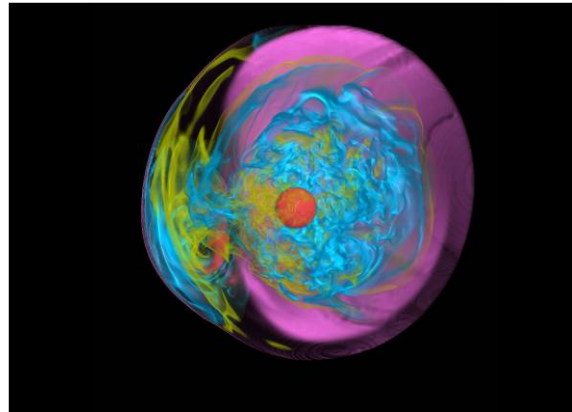
New insights into protein structure and function leading to better understanding of cellulose-to-ethanol conversion



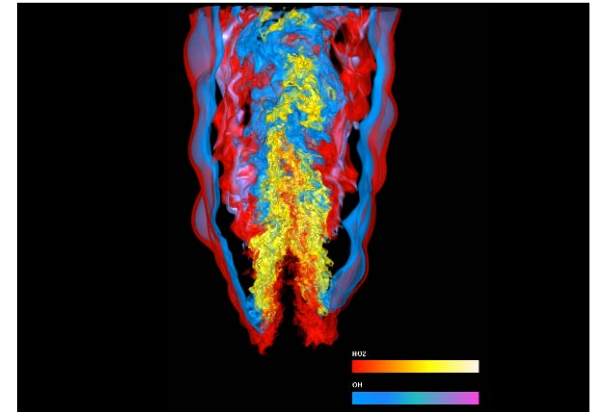
Addition of vegetation models in climate code for global, dynamic CO₂ exploration



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

ORNL FTG – Capabilities and Project Sampling

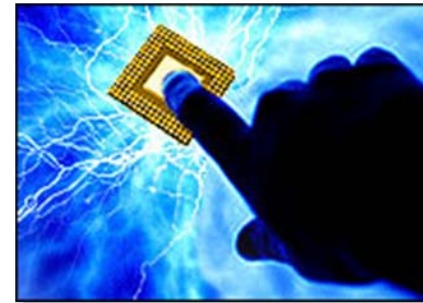
- Emerging architectures
 - Graphics Processors (e.g., NVIDIA)
 - FPGAs (e.g., Xilinx, Cray, SRC)
 - IBM CELL
 - Massive Multithreaded architectures (e.g., Cray XMT, Niagara 1 & 2)
 - Large Memory – Global addressing
- Performance
 - Scalable tools and infrastructures
 - mpiP MPI performance tool (over 10k downloads)
 - MR-Net scalable infrastructure
 - Prediction
 - Performance models up to 1M cores
 - Simulation of memory behavior
 - Ongoing Benchmarking
 - Nehalem, Shanghai, BlueGene/P
- Programming Systems
 - MPI3 One-sided and RMW operations
 - Global Distributed Transaction Memory support (e.g., Chapel global **atomic** ops)
- System Software
 - Parallel IO
 - Improved scalability for MPI-IO Lustre on Cray
 - PNFS over Lustre
 - TLB support for superpages
 - Virtualization for improved system management and scalability testing

ORNL Future Technologies Group

<http://ft.ornl.gov/>

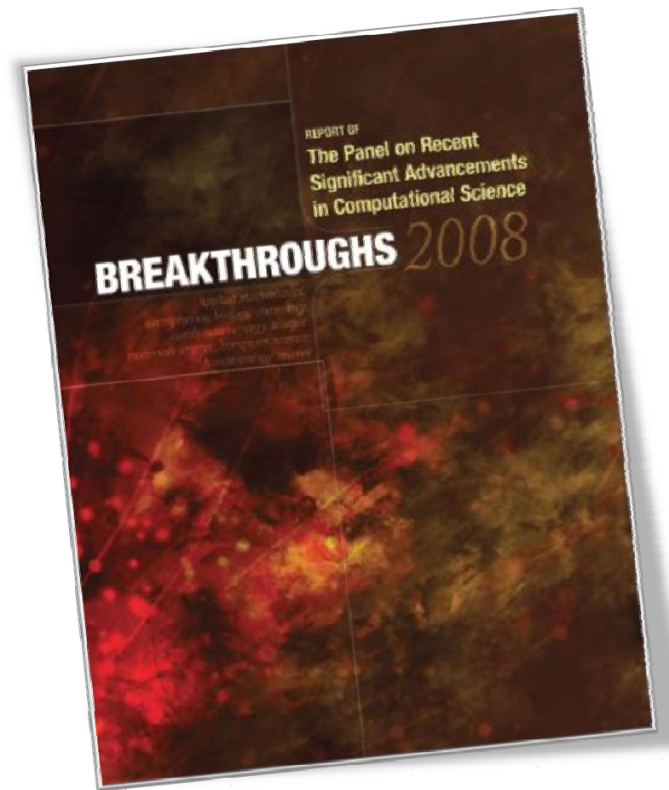
Mission - performs basic research in enabling technologies for future generations of high-end computing, including experimental computing architectures and software, with the goal of improving the performance, reliability, and productivity of these systems for our sponsors.

- Staff
 - Jeffrey Vetter (Group Leader)
 - Sadaf Alam
 - David Bader (joint w/ Georgia Institute of Technology)
 - Micah Beck (joint w/ University of Tennessee-Knoxville)
 - Anthony Danalis (joint w/ University of Tennessee-Knoxville)
 - Deb Holder (Group Secretary)
 - Gabriel Marin (in April)
 - Collin McCurdy
 - Jeremy Meredith
 - Kenneth Roche
 - Philip Roth
 - Thomas Sterling (joint w/ Louisiana State University)
 - Olaf Storaasli
 - Vinod Tipparaju
 - Weikuan Yu (joint w/ Auburn University)
- Sponsors
 - DOE Office of Science
 - DARPA
 - Intelligence Community
 - LDRD
 - NCCS
 - NIH
 - SBIR
 - DOE Institute for Advanced Architectures and Algorithms



ORNL Leadership Computing Facility: Delivering important breakthroughs

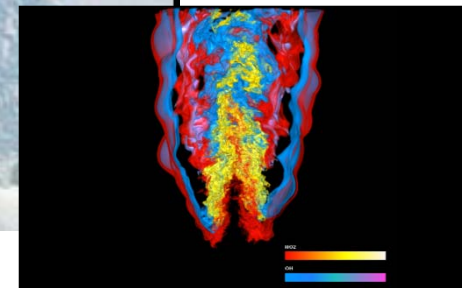
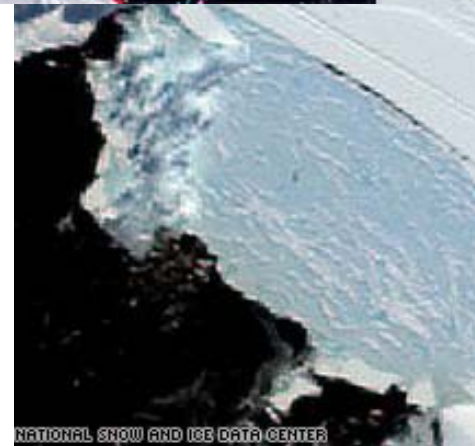
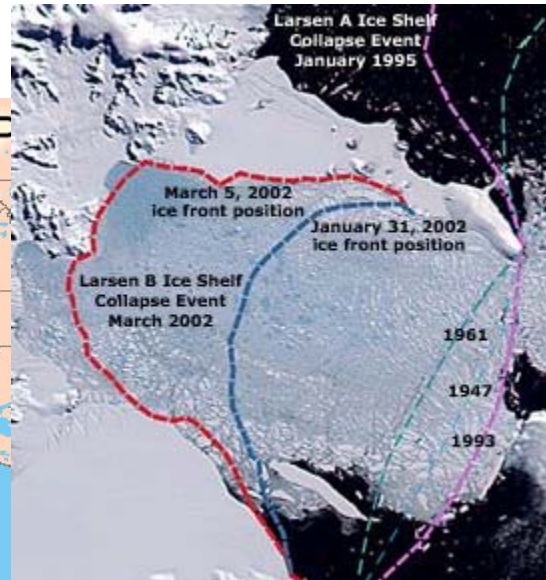
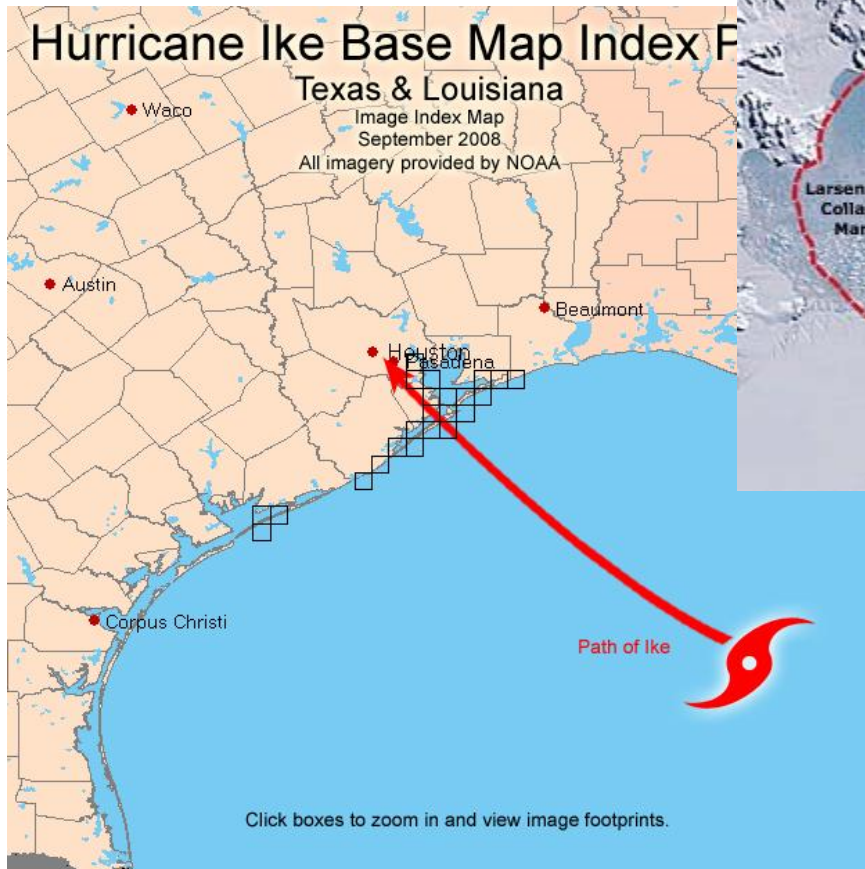
- February 2008: DOE's Office of Advanced Scientific Computing Research chartered a panel of computational scientists, applied mathematicians, and computer scientists to identify recent breakthroughs in computation science and enabling technologies
 - 10 major recent advances in simulation were identified
 - 5 of 10 were achieved using the Cray XT4 Jaguar at ORNL



Jaguar XT5: Delivering leadership computing

- Job size distribution indicates leadership usage
 - At least 30K cores: 50% of utilization
 - At least 45K cores: 32% of utilization
 - At least 90K cores: 18% of utilization
- 2009 Gordon Bell submissions: 5 from applications running on Jaguar XT5
- Application readiness and scaling demonstrate effectiveness of ORNL Leadership Computing Facility model

Changing Priorities of Workloads: Systems Must Perform Well for All



Application Complexity

- Today's applications are quickly becoming more complex than earlier applications
 - Exascale Town Hall report [1]
- Increased fidelity
- Answering more complex questions
 - Scientists [2,3] working on model that combines
 - Climate model
 - Energy economics
 - Population movements, demographics
- New application areas
 - Search
 - Massive Data
- Multiple phases of a multi-physics application
 - E.g., Climate
- Multiple languages, libraries, runtime systems

Applications design and implementation are already complex; Writing and optimizing code for each new architecture and programming model is impractical (and only going to happen w/ heroic efforts/funding.)

[1] H. Simon, T. Zacharia, and R. Stevens, Eds., Modeling and Simulation at the Exascale for Energy and the Environment, 2007.

[2] S.W. Hadley, D.J. Erickson et al., "Responses of energy use to climate change: A climate modeling study," *Geophysical Research Letters*, 33(17), 2006. the U.S. Department of Energy

[3] D. Vergano, "Air conditioning fuels warming," in USA Today, 2006.