

Perspectives on Petascale Computing for Earth System Modeling

Presented to the
11th International Specialist Meeting on the Next
Generation Models on Climate Change and
Sustainability for Advanced High Performance
Computing Facilities

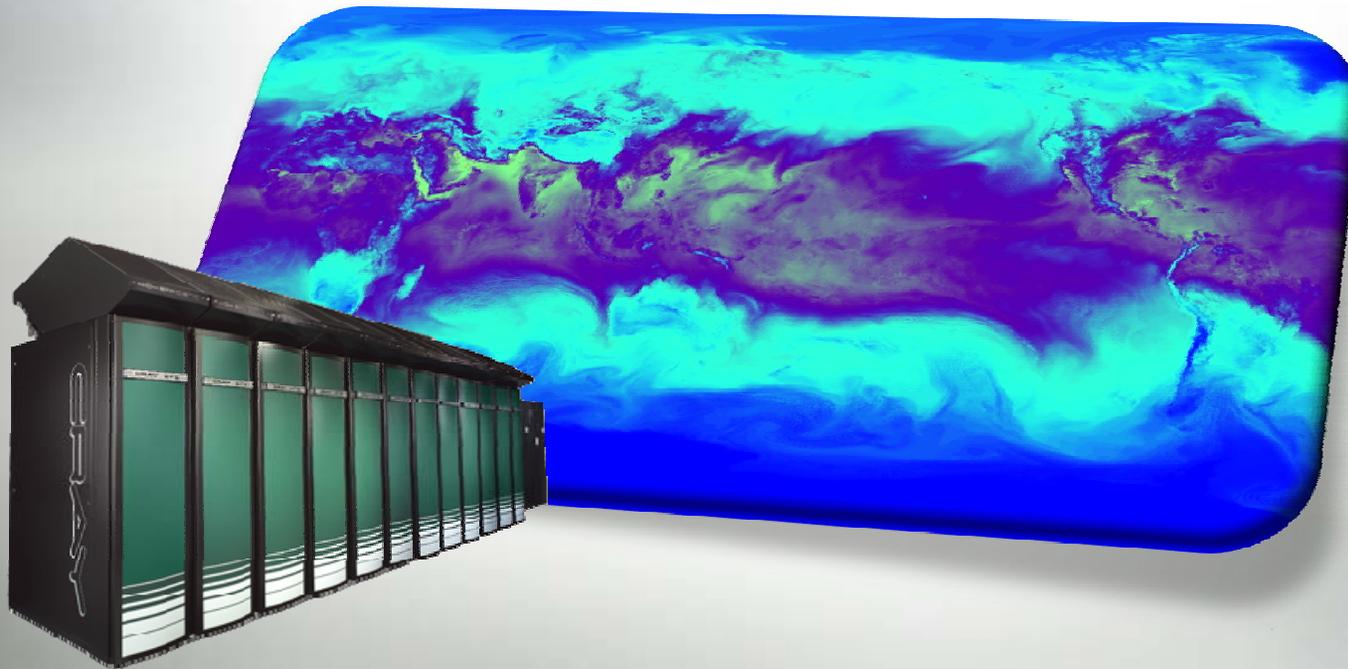
16-18 March 2009

John Levesque
levesque@cray.com

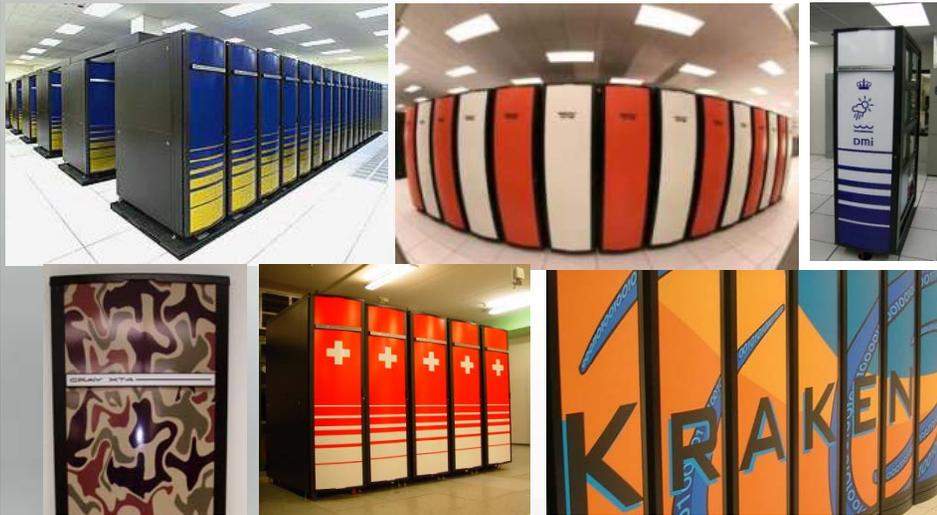
Per Nyberg
nyberg@cray.com

- General Cray Update
- Cray Technology Directions
- Perspectives on Petascale Computing

General Cray Update



HPC is Cray's Sole Mission



- Used by leading HPC centers worldwide.
- Significant R&D centered around all aspects of high-end HPC:
 - From scalable operating systems,
 - To advanced cooling technologies.
- Cray XT MPP architecture has gained mindshare as prototype for highly scalable computing:
 - From 10's TF to Petaflop.



Market & Business Momentum

- >\$700M in contracts in 2006~2008.
- Continued strong international momentum.
- We shipped our 1000th Cray XT cabinet on 4 September 2008 (*and 100% are still in service*).
- First Petaflop system was delivered to ORNL and accepted by end of 2008.
- 607 TF NSF University of Tennessee system accepted in February 2009.
- \$250M DARPA HPCS Phase III contract award.
- R&D agreement with Intel to create custom compute processors with microprocessor technology as part of HPCS Cascade program.

2008~2009 XT Accomplishments

- Over 50 upgrades or systems built:
 - >11K Compute blades
 - >70K sockets (280K cores)
 - >390 XT cabinets
- Introduced ECOphlex Liquid-Cooling
- Lustre file system at ORNL breaks 100 GB/s (some tests break 150 GB/s).
 - Plan to hit 240 GB/s.
- ORNL JaguarPF designed, built and delivered on schedule!
- 607 TF NSF University of Tennessee system accepted in February 2009.



University of Tennessee 607 TF Cray XT5 - Kraken

- Officially entered full production on 2 February 2009.
- Next scheduled upgrade is in late 2009.
- NSF's largest supercomputer and the world's fastest university managed supercomputer.
- NSF Track 2 award to the University of Tennessee.
- Housed at the University of Tennessee – Oak Ridge National Laboratory Joint Institute for Computational Sciences.



THE UNIVERSITY of TENNESSEE

KNOXVILLE, CHATTANOOGA, MARTIN, TULLAHOMA, MEMPHIS

Mar-2009

ORNL Climate Meeting



Slide 7

ORNL Petascale “JaguarPF” System

- Installed at the National Center for Computational Sciences (NCCS) at ORNL.
 - XT5 with ECOphlex liquid cooling.
- Will enable petascale simulations of:
 - Climate science
 - High-temperature superconductors
 - Fusion reaction for the 100-million-degree ITER reactor
- Not just more of the same, but unprecedented simulations.
- The only open science petascale system in the world.



What does the ORNL Petascale Jaguar System Represent the Earth System Modeling Community ?



- Milestone capability that has been in demand by climate community for years.
- Capabilities highlighted by Raymond Orbach at the November 2008 DoE sponsored “Challenges in climate change science and role of computing at the extreme scale”.
- For modelers – either you have run at 150,000 cores, or you have not.
- For scientists – unprecedented results in the near-term and a capability to model unknown phenomena.



Mar-2009



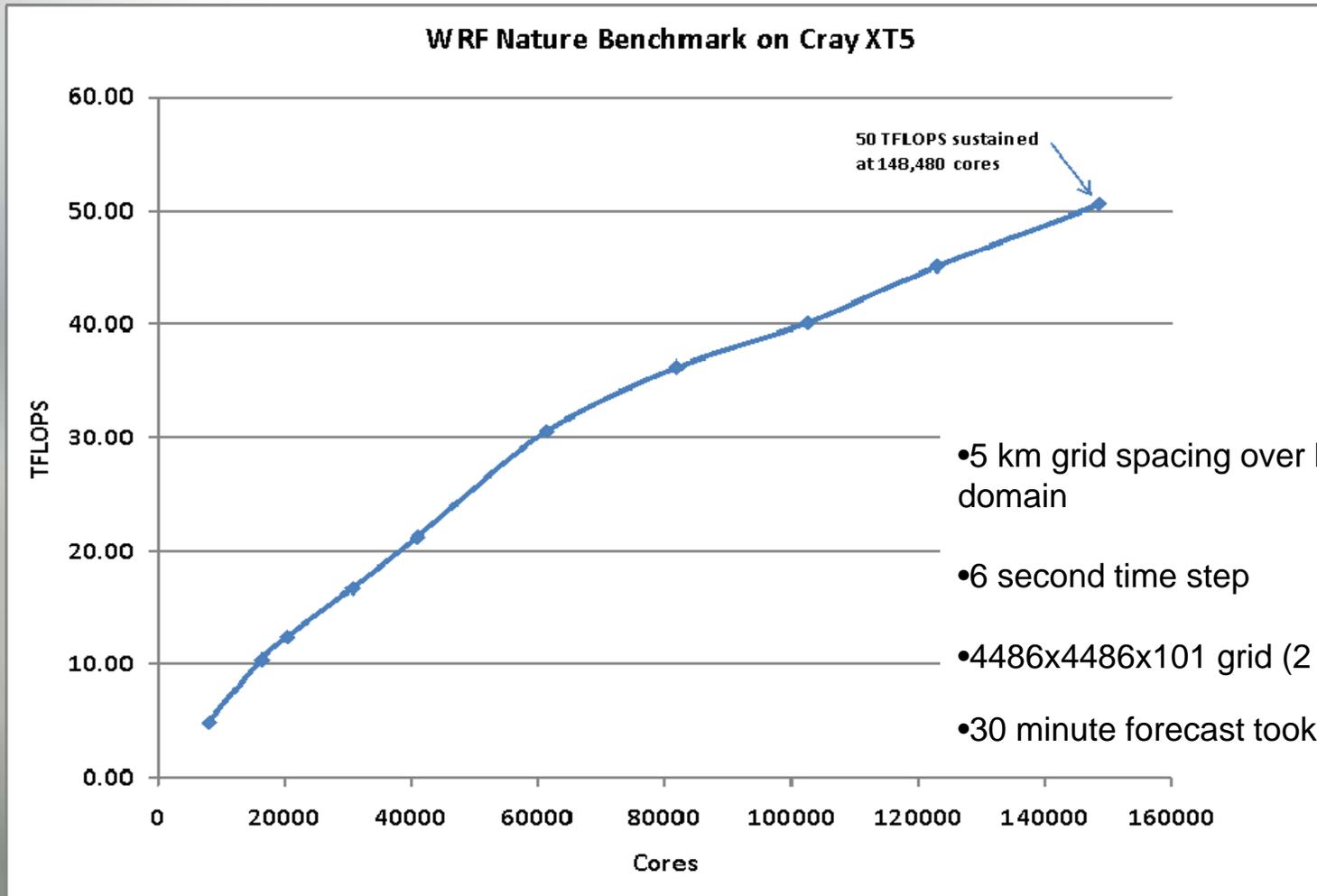
ORNL Climate Meeting



Slide 9

WRF 'nature' Benchmark on ORNL Cray JaguarPF

Breaks the standing Earth Simulator record for sustained performance on a meteorological application



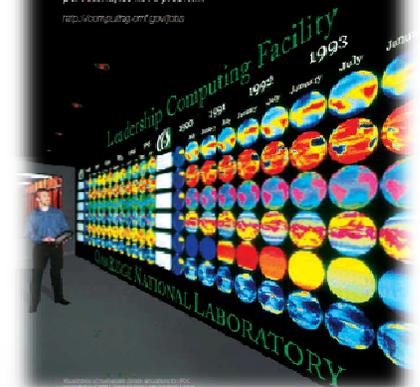
- 5 km grid spacing over hemispheric domain
- 6 second time step
- 4486x4486x101 grid (2 billion cells)
- 30 minute forecast took 69 seconds

Climate Usage of DoE Cray Systems

- DoE / NSF Climate End Station (CES)
 - An interagency collaboration of NSF and DOE in developing the Community Climate System Model (CCSM) for IPCC AR5.
 - A collaboration with NASA in carbon data assimilation
 - A collaboration with university partners with expertise in computational climate research.
- DoE / NOAA MoU
 - DoE to provide NOAA/GFDL with millions of CPU hours.
 - Climate change and near real-time high-impact NWP research .
 - Prototyping of advanced high-resolution climate models.

Understanding climate change sometimes requires the world's most powerful open science computer.

With more than 100 trillion calculations per second, it's not a problem.
<http://www.computing.doe.gov/data>



NOAA NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
 UNITED STATES DEPARTMENT OF COMMERCE

Department of Energy to Provide Supercomputing Time to Run NOAA's Climate Change Models

September 8, 2008

The U.S. Department of Energy's (DOE) Office of Science will make available more than 10 million hours of computing time for the U.S. Commerce Department's National Oceanic and Atmospheric Administration (NOAA) to explore advanced climate change models at three of DOE's national laboratories as part of a three-year memorandum of understanding on collaborative climate research signed today by the two agencies.

NOAA will work with climate change models as well as perform near real-time high-impact (non-production) weather prediction research using computing time on DOE Office of Science resources including two of the world's top five most powerful computers – the Argonne National Laboratory's 557 TF IBM Blue Gene/P and Oak Ridge National Laboratory's 263 TF Cray XT4. NOAA researchers will also receive time on DOE's National Energy Research Scientific Computing Center at Lawrence Berkeley National Laboratory.

NOAA Administrator Conrad C. Lautenbacher, Jr. (left) and DOE Under Secretary for Science Dr. Raymond L. Orbach (right).
 High resolution (Credit: NOAA)

Advanced, high-resolution climate models from NOAA's Geophysical Fluid Dynamics Laboratory (GFDL) will be prototyped and compared to other models like the NSF-DOE sponsored Community Climate System Model. This partnership is also consistent with the goals of the U.S. Climate Change Science Program, which is responsible for facilitating the creation and application of knowledge of Earth's global environment through research, observations, decision support, and communication. NOAA and DOE scientists play key roles in national and international assessments, for example, the Nobel Prize winning Intergovernmental Panel on Climate Change.

The Climate End Station: Moving to CESM v1.0

CCSM Development Peter Gent NCAR

Collaborators: ~100 Scientists from DOE Labs, NCAR, and University Community

Development of the fourth version of the CCSM. One of the biggest improvements is the simulation of ENSO.



Use of the CCSM to produce short-term climate forecasts to 2030. Higher atmosphere resolution (3.5%) produces better rainfall in SE



Note Improved SE USA rainfall.

Climate Change Warren Washington NCAR

CCSM Climate Change Working Group: Critical science development and applications in support of climate-informed decision making

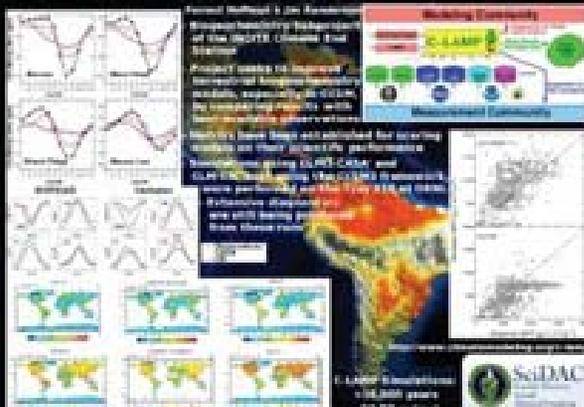
- DOE CSPP (AR2-1A) "Low Emissions Scenario"
- Coupled for short runs
- Near-Term High-Resolution Climate Prediction
- Climate Change 2100 and beyond
- Preoptic Carbon Aerosol experiments



Carbon Assimilation Don Anderson NASA

Collaborators: Donald Anderson, Steven Fawson, and Mike Späthlin

Simulation and Assimilation of Atmospheric Carbon Species via Multi-year simulations of atmospheric carbon species at 0.5-degree spatial resolution using the GEOS-5 AGCM using specified emission boundary fluxes for CO and CO2. The AGCM will be perturbed in order to produce estimates of sensitivity of the tropospheric trace gases to different parameters in the model's physical parameterizations. Analysis of the results will quantify how different choices of physical parameters impact the quality of simulations, including agreement with ground-based CO and CO2 measurements and space-based radiances.



SciDAC ESM w/ Sulfur Philip Cameron-Smith LLNL

Status: Components validated. Ocean spin-up underway.

- Scientists involved:
- Philip Cameron-Smith (PI, LLNL)
 - Rub Jacob (LANL)
 - Scott Eby (LANL)
 - Phil Jones (LANL)
 - Mat Moulton (LANL)
 - Art Minobe (LLNL)
 - Jean-Francois Lamarque (NCAR)
 - John Drake (ORNL)
 - Dave Erickson (ORNL)
 - Steve Glass (PNNL)
 - * other SciDAC collaboration
 - Focus: Implementing ESM on Super



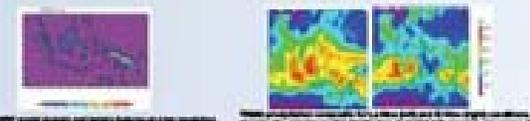
Eddy Resolving POP Phil Jones LANL

Mathew Malmud (LANL), Frank Bryan (NCAR), Phil Jones (LANL)

Configure and spin up a 1/10 degree, 42 level stand-alone POP simulation. Then use the result as the initial condition in a fully coupled CCSM run. Compare the results with a similarly configured 1 degree ocean in order to quantify the impact of using an eddying ocean model.



WRF Resolution Studies Ruby Lueng PNNL



Objective: To systematically assess impacts of spatial resolution on simulating cloud processes and their interactions with the circulation.

Approach: Simulations will be performed with the Weather Research and Forecasting (WRF) model. One year cloud resolving simulations at 1, 2, and 4 km resolution in the Pacific warm pool. One year mesoscale simulations at 10, 20, and 40 km resolution in the Pacific warm pool. Simulations using the global WRF on an equi planet with embedded domains at 10 and 12 km resolution in the Pacific warm pool.

CAM Development Jim Hack ORNL

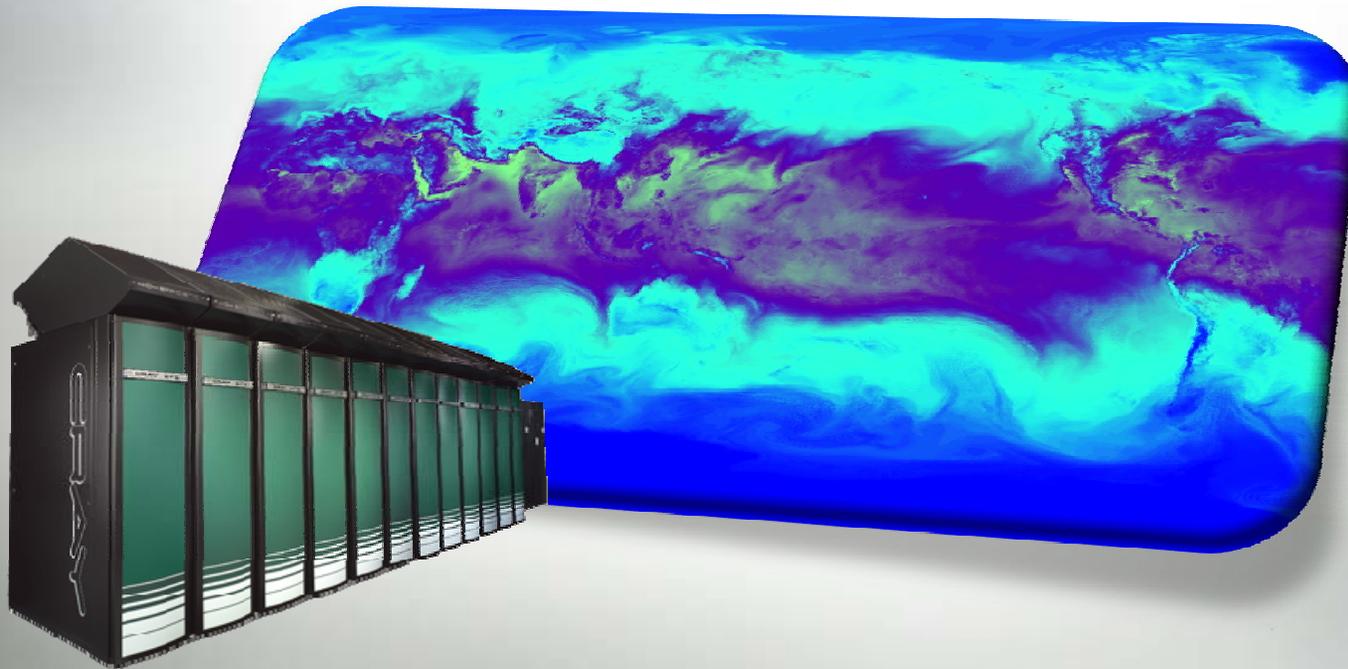
Improvements in mean climate properties were seen when the horizontal resolution in CAM was doubled. However, only modest improvements in transient behavior were observed, suggesting that large-scale systematic biases in mean climate are far more related to the parameterized treatment of non-resolvable physics, such as clouds, convection, radiation and boundary layer. But they also demonstrate that some transient features of the simulation are far more realistic, including the clear presence of tropical storms, at the T426 truncation. In addition to the spectral Eulerian dynamical core, we are experimenting with the prototype GCM using the PV dynamical core at resolutions as fine as 0.25 dp, which should have solution accuracy equivalent to the T426 spectral model.

University Component David Dickinson ORNL

The magnitude and ubiquity of deforestation in the Amazon basin has become increasingly clear in recent years with overwhelming evidence that vegetation is not passively responding to changes in climate and weather, but is dynamically evolving with climate, human and natural disturbances to vegetation. The MIT team (Knox Bras Chouinard) is simulating this interaction via an ensemble of Amazonian land-atmosphere states via coupled land-atmosphere limited-area model simulations. These simulations will use high and low carbon emission GCM (CCSM-CM3.0) output as lateral boundaries, and forced anthropogenic deforestation over boundary conditions from both business-as-usual (aggressive deforestation) and conservation type.



Cray Technology Directions



Opposing Forces in Today's HPC Marketplace

HPC wire

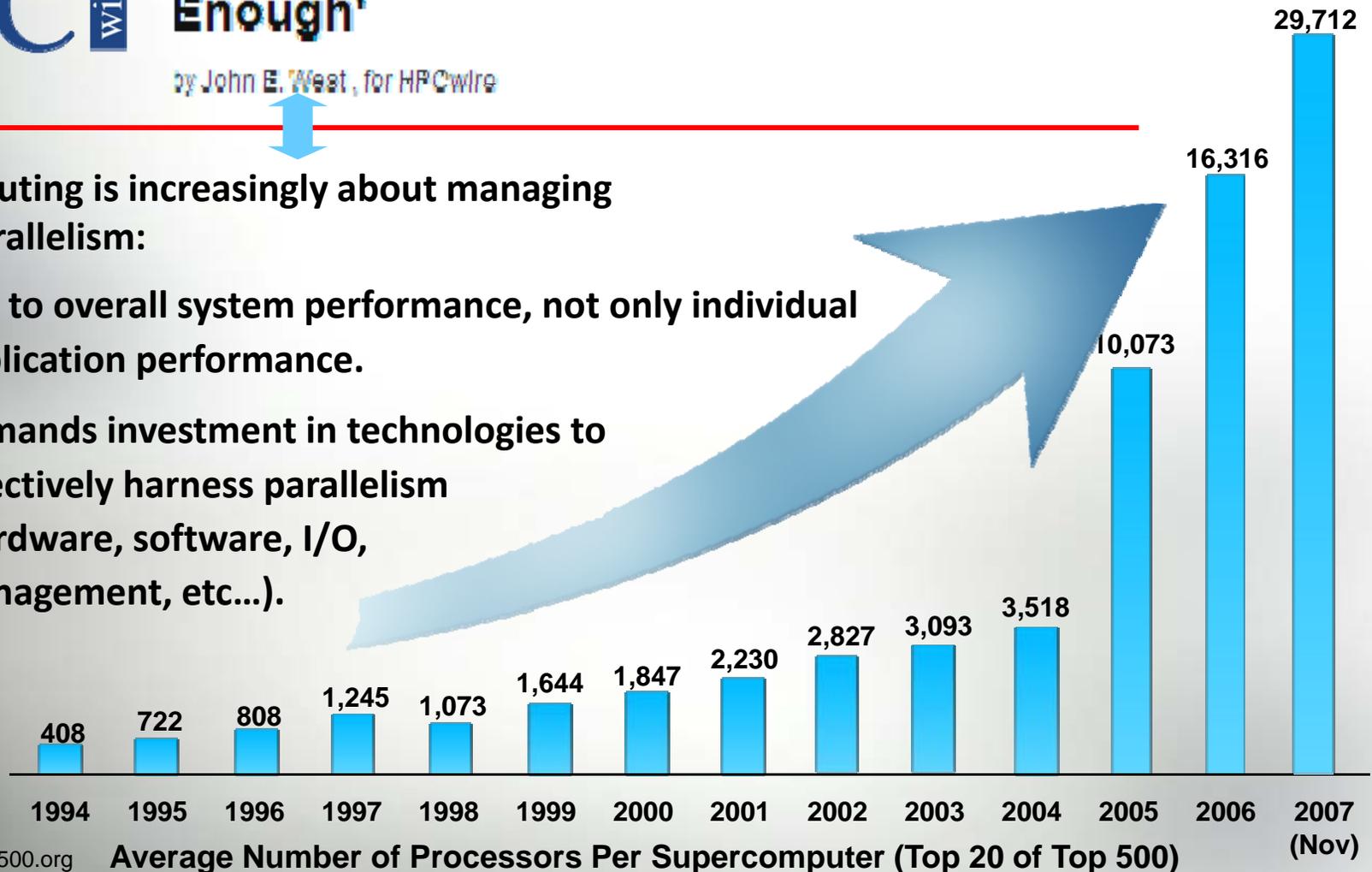
April 28, 2008

HPC Innovation In the Era of 'Good Enough'

by John E. West, for HPCwire

Supercomputing is increasingly about managing massive parallelism:

- Key to overall system performance, not only individual application performance.
- Demands investment in technologies to effectively harness parallelism (hardware, software, I/O, management, etc...).



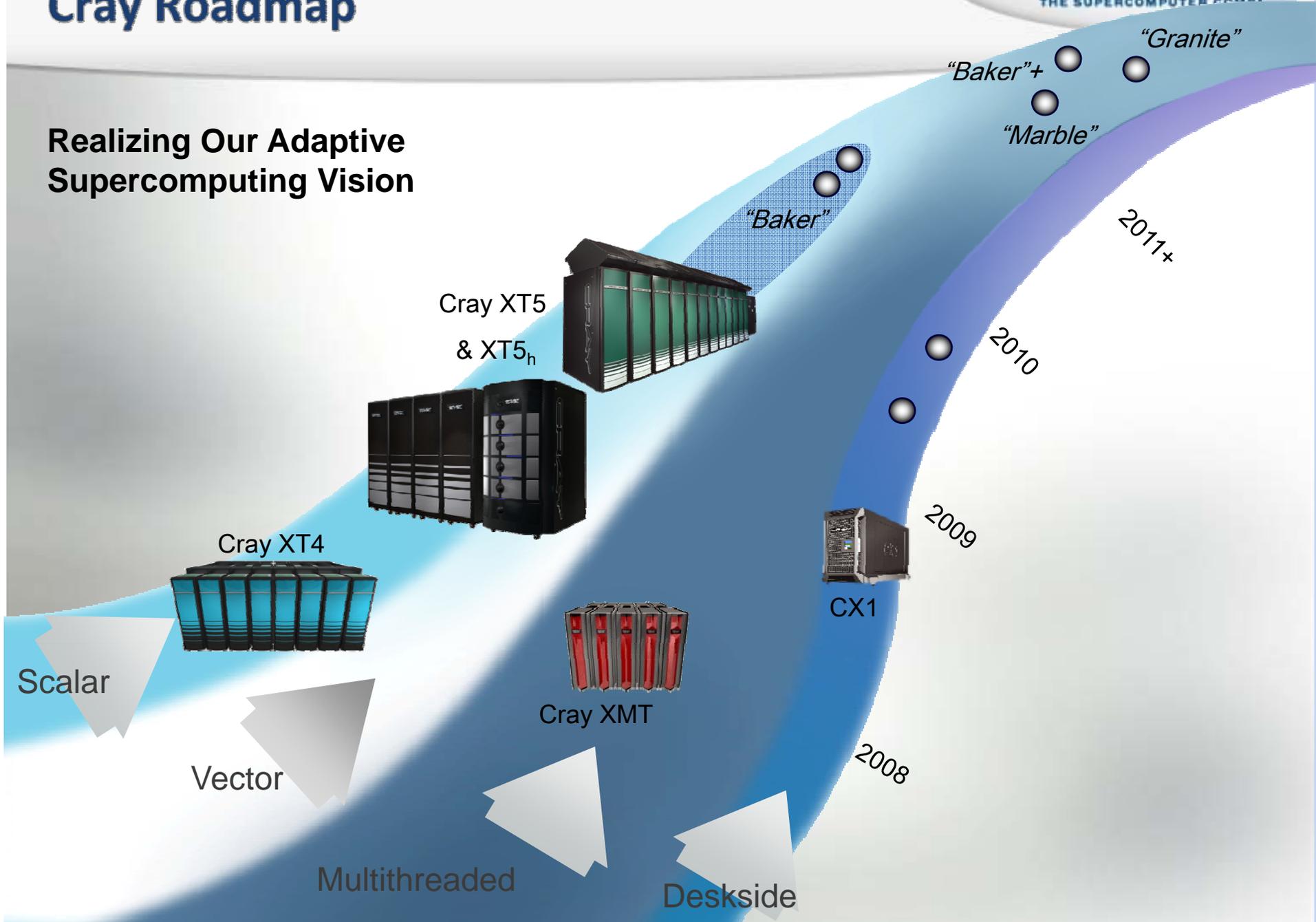
Source: www.top500.org

Average Number of Processors Per Supercomputer (Top 20 of Top 500)

Cray Roadmap



Realizing Our Adaptive Supercomputing Vision

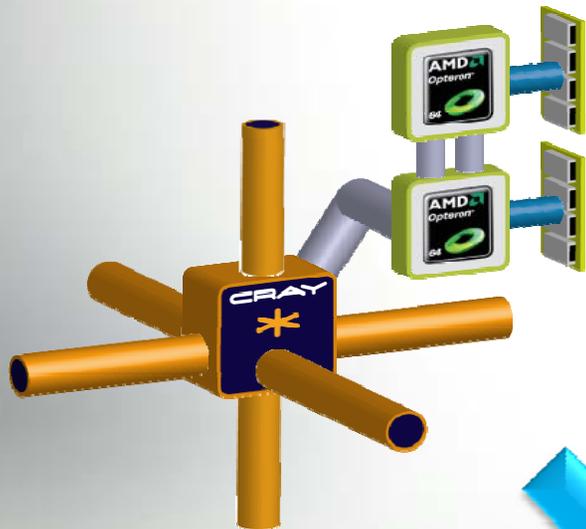


Cray Technology Directions

- Scalable System Performance:
 - Enabling scalability through technologies to manage and harness parallelism.
 - In particular interconnect and O/S technologies.
 - Cray is a pioneer and leader in light weight kernel design.
 - System performance as a whole to support a petascale infrastructure.
- Green Computing: packaging, power consumption and ECOphlex cooling:
 - Simple expansion and in-cabinet upgrade to latest technologies, preserving investment in existing infrastructure.
 - Lowest possible total electrical footprint.
- Provide investment protection and growth path through R&D into future technologies: HPCS Cascade Program.

Cray XT Architecture Building Blocks

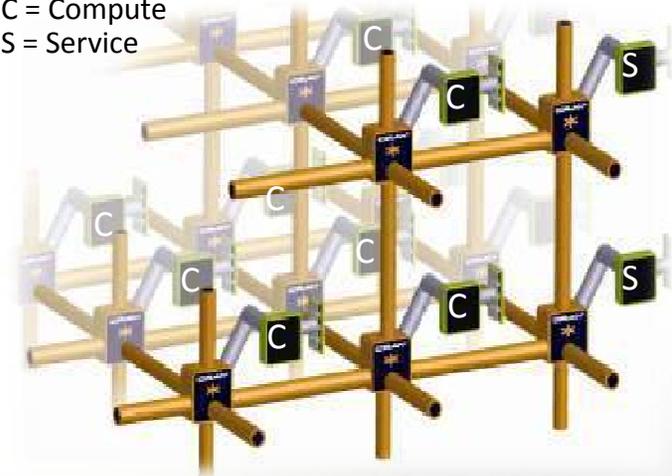
Microprocessor Based Node



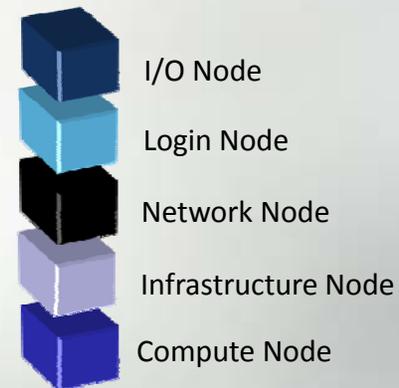
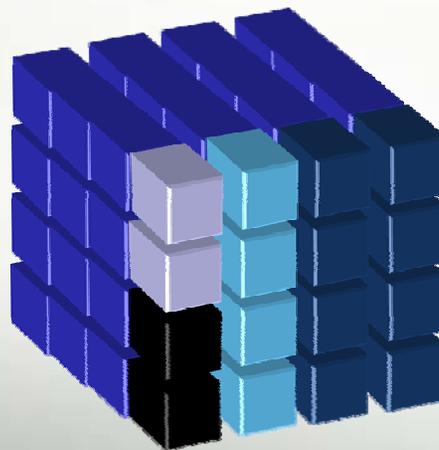
Cray Custom Interconnect

3-D Torus

C = Compute
S = Service

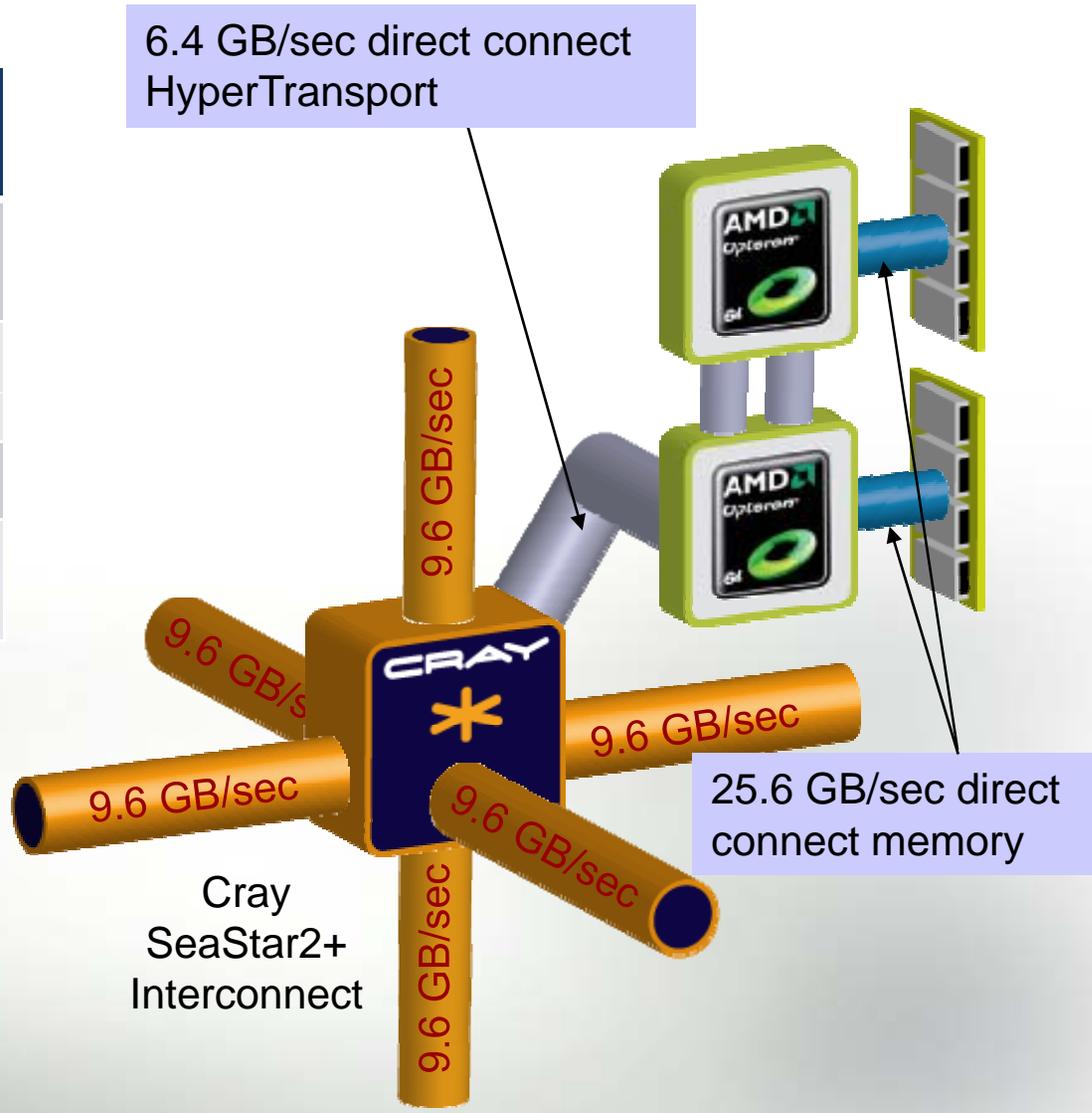


Unified System Architecture
Designed for Performance and
Management at Scale



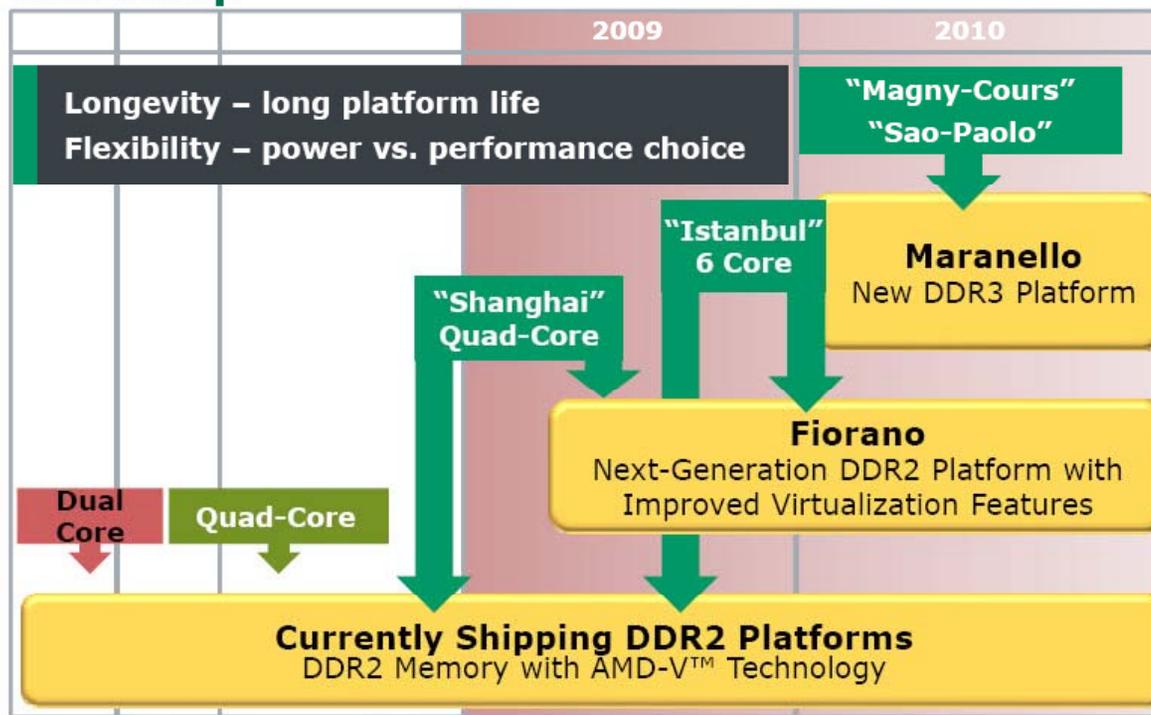
Cray XT5 Node - Today

Cray XT5 Node Characteristics	
Number of Cores	8
Peak Performance	73-86 Gflops/s
Memory Size	8-32 GB per node
Memory Bandwidth	25.6 GB/sec



Leveraging the AMD Roadmap

AMD Cross-Generation x86 Server Platforms Roadmap

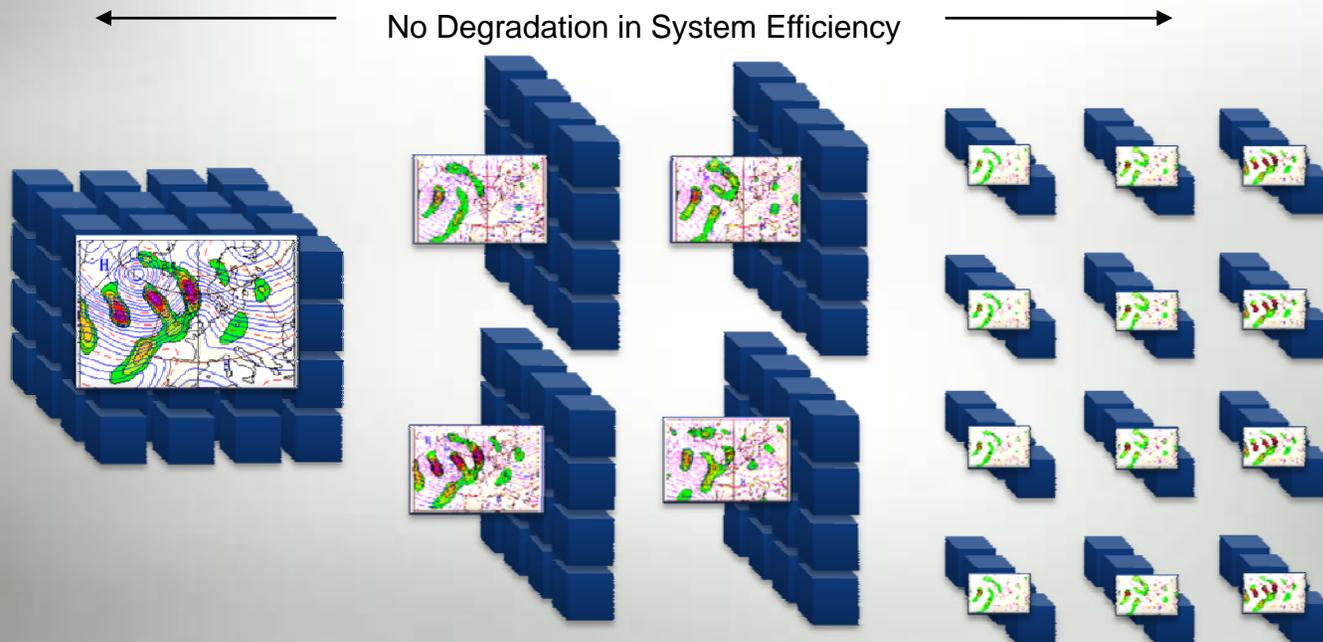


5 | AMD Opteron™ for Science | February 25, 2009



XT Architecture Support for NWP and Climate Workload Efficiency

- Cray XT architecture is uniquely positioned to provide both:
 - Single model performance at scale – deterministic forecasts.
 - Throughput performance – ensemble modeling.
- Maximizes:
 - Model development and execution.
 - System administration.
 - Facilities and total cost of ownership.
- In a single, unified system architecture.



Why the Interconnect Matters

Table 1. Key characteristics of computer systems

System name	Location	Processor	Clock speed (GHz)	Max flops/clock	Peak Gflop/s/core	Cores / node	Node	Interconnect
DataStar	SDSC	IBM Power4+	1.5/1.7	4	6.0/6.8	8	IBM p655	Custom fat tree
Franklin	LBNL/NERSC	AMD Opteron	2.6	2	5.2	2	Cray XT4	Custom 3D torus
Lonestar	TACC	Intel Xeon (Woodcrest)	2.66	4	10.6	4	Dell Power-Edge 1955	InfiniBand SDR fat tree
Ranger	TACC	AMD Opteron (Barcelona)	2.0	4	4.0	16	SunBlade x6420	InfiniBand SDR fat tree

From paper:

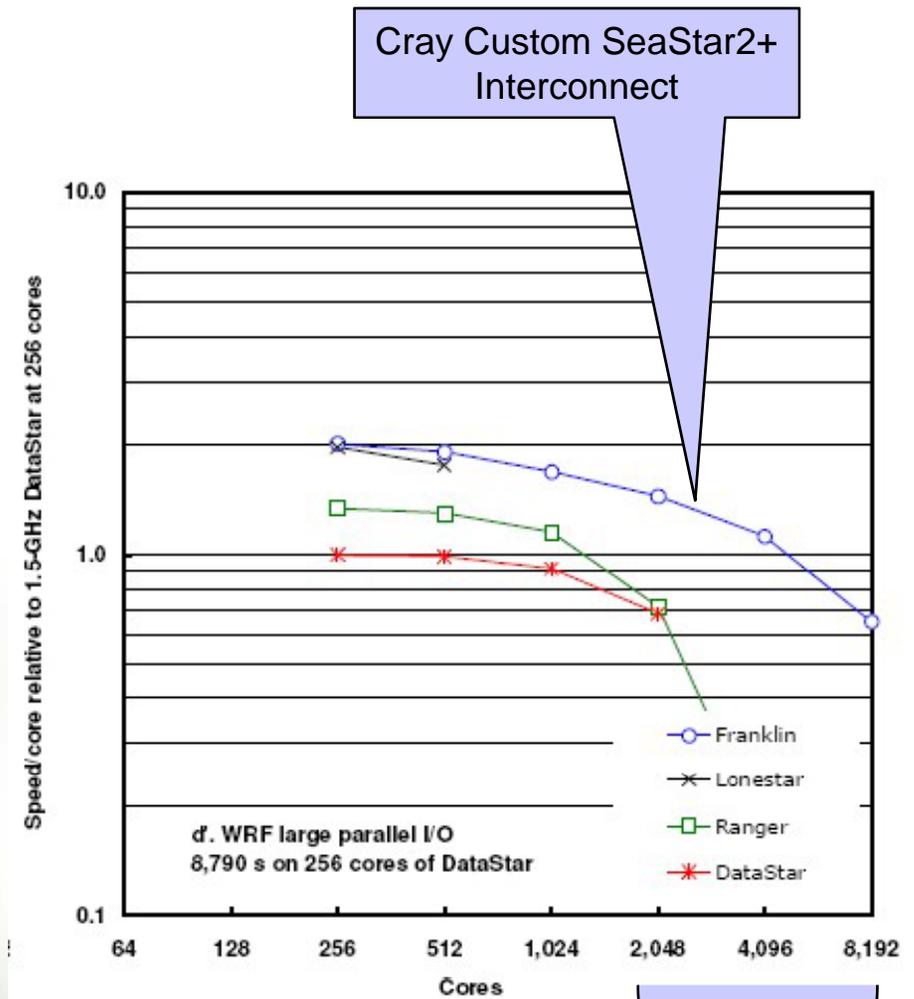
“Characterizing Parallel Scaling of Scientific Applications using IPM”

Nicholas J. Wright, Wayne Pfeiffer, and Allan Snaveley

Performance Modeling and

Characterization Lab

San Diego Supercomputer Center



4x the number of cores

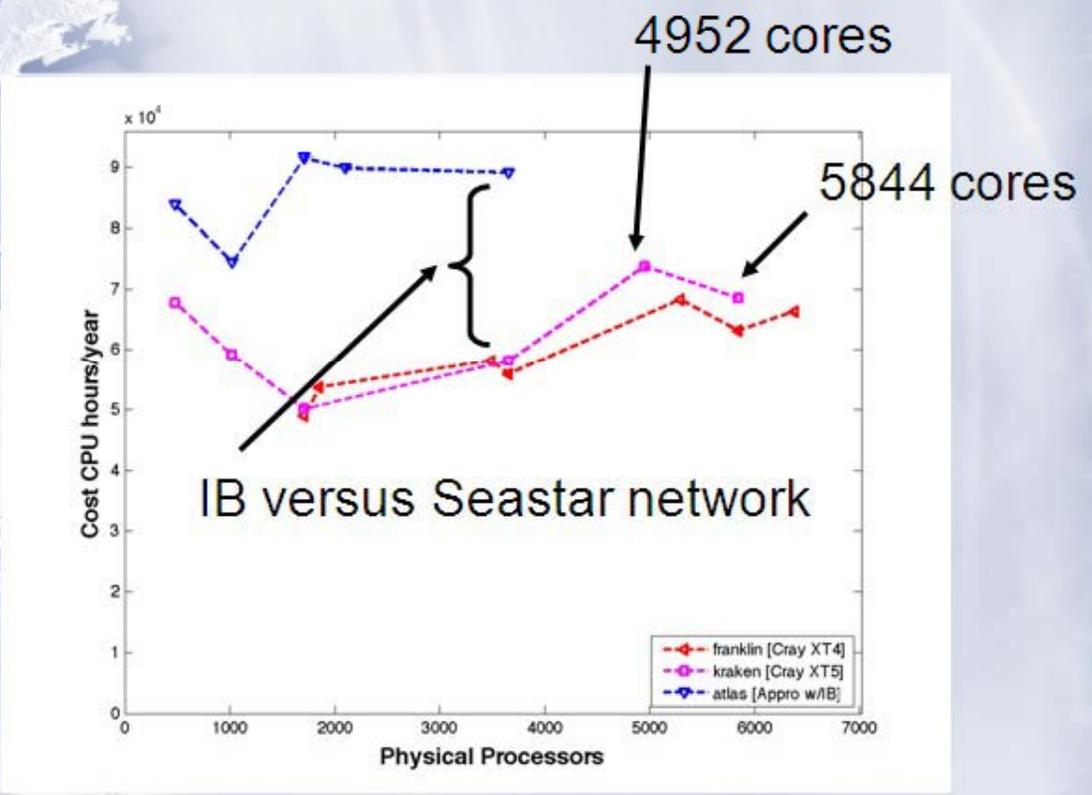
Why the Interconnect Matters



Performance Evaluation of Ultra-High-Resolution Climate Simulations

John M. Dennis: dennis@ucar.edu
 Mariana Vertenstein: mvertens@ucar.edu
 Tony Craig: tcraig@ucar.edu
 March 10, 2009

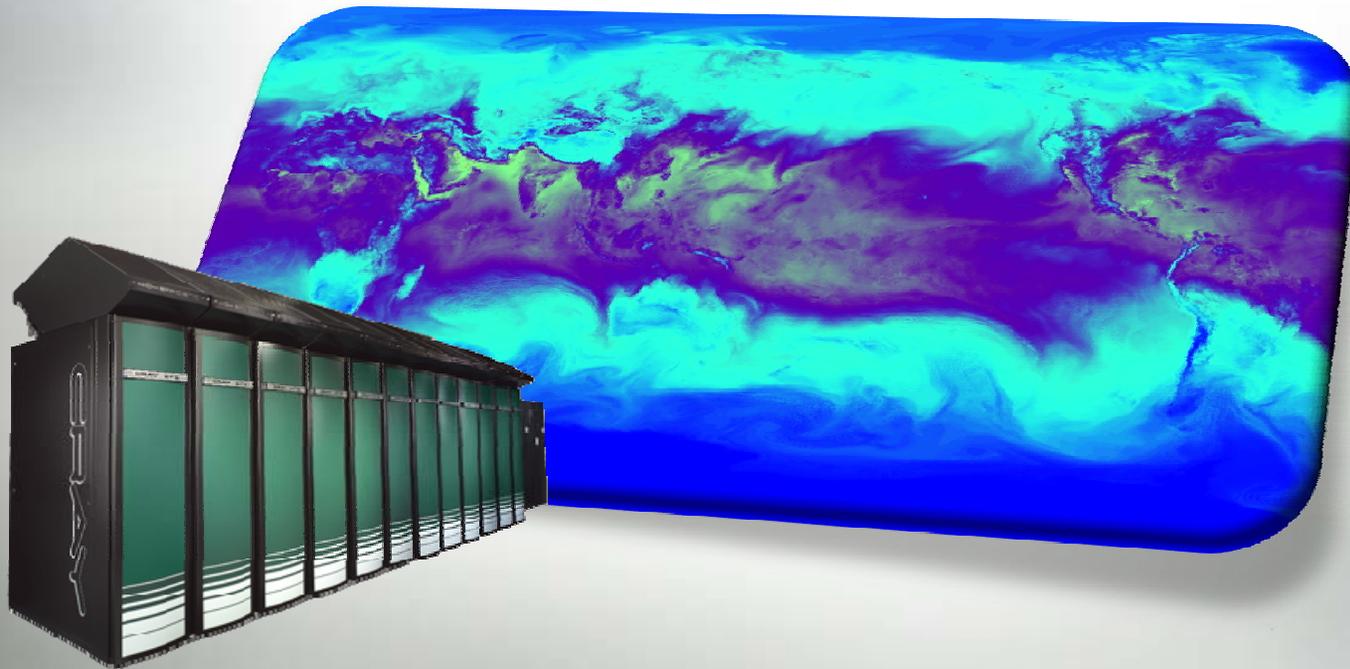
Simulation cost for CCSMhr05A



Mar 10, 2009

24

Perspectives on Petascale Computing



Scientific and HPC Advances in Earth System Modeling

Science

Improved scientific representation
Increased scientific complexity
Increased observational data and data assimilation
Increased resolution, integration lengths
Use of ensembles and inter-disciplinary coupled modeling

Today

In the past speed was based on exponential growth in single CPU performance.
Today it is through exponential growth in parallelism.
Efficient scalability is the challenge.
Every model and system aspect must be addressed.
Developer access to scalable systems is essential.

Today

Cost per grid point is increasing due to physics.
Code complexity increasing.
Not all components run at the same efficiency.
Greater degree of dependence on the overall computing, data and support environment.



High Performance Computing and Modeling

Algorithmic improvements (applied math & domain specific)
Computer Science improvements
Raw processing speed

Petascale (or even Terascale) Perspectives

- There remains a tremendous number of models and application areas that have not yet reached even the terascale level.
- Those that can scale have benefited from a focused, iterative multi-year algorithmic optimization effort:
- Optimization strategies do not remain stagnant and must take advantage of evolving hardware and software technologies.
- Ongoing access to scalable, leadership class systems and support is essential:
 - Either you have actually run on 5K, 10K, 20K, 50K... processors, or you have not ! Theory vs. Reality

Evolving Path to Petascale

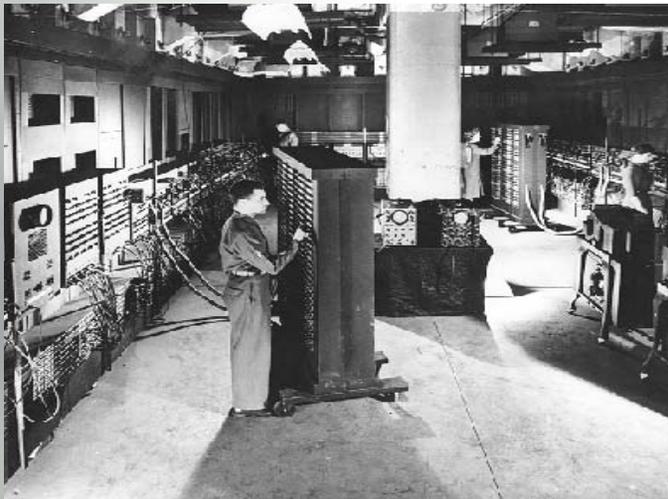
- Three basic paths are available to users today:
 - Standards based MPP
 - Low power
 - Accelerators
- Cray Cascade program is intended to bring these together to provide greater flexibility within a single architecture.
- Will be accomplished through:
 - Continued investment in core MPP and emerging technologies.
 - Continued investment in customer relationships.
- Cray's wealth of experience in pushing the boundaries of scalability will continue to positively impact entire HPC community.

Summary

- HPC is Cray's only business.
- Cray's MPP technologies are playing a key role in supporting the HPC community in preparing for and using Petascale capabilities.
- Cray's wealth of experience in pushing the boundaries of scalability will continue to positively impact entire HPC community.

58 years after the first NWP forecast...

1950 NWP Forecasts on ENIAC



- 736 km grid spacing over U.S.
- 15 x 18 x 1
- 1,000,000 computations
- 24 hour forecast took 24 hours

2008 WRF Nature Run on Cray XT



- 5 km grid spacing over hemispheric domain
- 4486 x 4486 x 101 grid (2 billion cells)
- Nearly 3.5 quadrillion floating point operations
- 30 minute forecast took 69 seconds

Thank you for your attention