Early Evaluation of the Cray XT3

S. R. Alam, T. H. Dunigan, Jr, M. R. Fahey,

P. C. Roth, J. S. Vetter, P. H. Worley

Oak Ridge National Laboratory

Oak Ridge, TN, USA 37831

http://www.csm.ornl.gov/ft
http://www.csm.ornl.gov/evaluations
http://www.nccs.gov

Acknowledgments

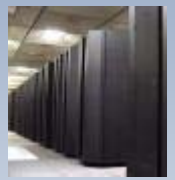
- Cray
  - Supercomputing Center of Excellence
  - Jeff Beckleheimer, John Levesque, Luiz DeRose, Nathan Wichmann, and Jim Schwarzmeier
- Sandia National Lab
  - Ongoing collaboration
- Pittsburgh Supercomputing Center
  - Exchanged early access accounts to promote cross testing and experiences
- ORNL staff
  - Don Maxwell
- This research was sponsored by the Office of Mathematical, Information, and Computational Sciences, Office of Science, U.S. Department of Energy under Contract No. DE-AC05-00OR22725 with UT-Batelle, LLC. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

NCCS Resources

September 2005 Summary

Network Routers

Control Network
1 GigE
10 GigE
UltraScience

Cray XT3 Jaguar	Cray X1E Phoenix	SGI Altix Ram	IBM SP4 Cheetah	IBM Linux NSTG	Visualization Cluster	IBM HPSS
						
5294 2.4GHz 11TB Memory	1024 .5GHz 2 TB Memory	(256) 1.5GHz 2TB Memory	(864) 1.3GHz 1.1TB Memory	(56) 3GHz 76GB Memory	(128) 2.2GHz 128GB Memory	Many Storage Devices Supported

Shared Disk	120TB	32TB	36TB	32TB	4.5TB	9TB	5TB

Backup Storage 5PB

Scientific Visualization Lab

- 27 projector Power Wall

Test Systems

- 96 processor Cray XT3
- 32 processor Cray X1E*
- 16 Processor SGI Altix

Evaluation Platforms

- 144 processor Cray XD1 with FPGAs
- SRC Mapstation
- Clearspeed
- BlueGene (at ANL)

7 Systems

Supercomputers
7,622 CPUs
16TB Memory
45 TFlops

238.5 TB

5 PB

Cray XT3 System Configuration – Jaguar

- 56 cabinets
- 5,212 compute processors (25 TF)
- 82 service and I/O processors
- 2 GB memory per processor
- 10.7 TB aggregate memory
- 120 TB disk space in Lustre file system
- Topology (X,Y,Z): 14 torus, 16 mesh/torus, 24 torus

Cray XT3 System Overview

- Cray's third generation MPP
  - Cray T3D, T3E
- Key features build on previous design philosophy
  - Single processor per node
    - Commodity processor: AMD Opteron
  - Customized interconnect
    - SeaStar ASIC
    - 3-D mesh/torus
  - Lightweight operating system – catamount – on compute PEs
  - Linux on service and IO PEs

Cray XT3 PE Design

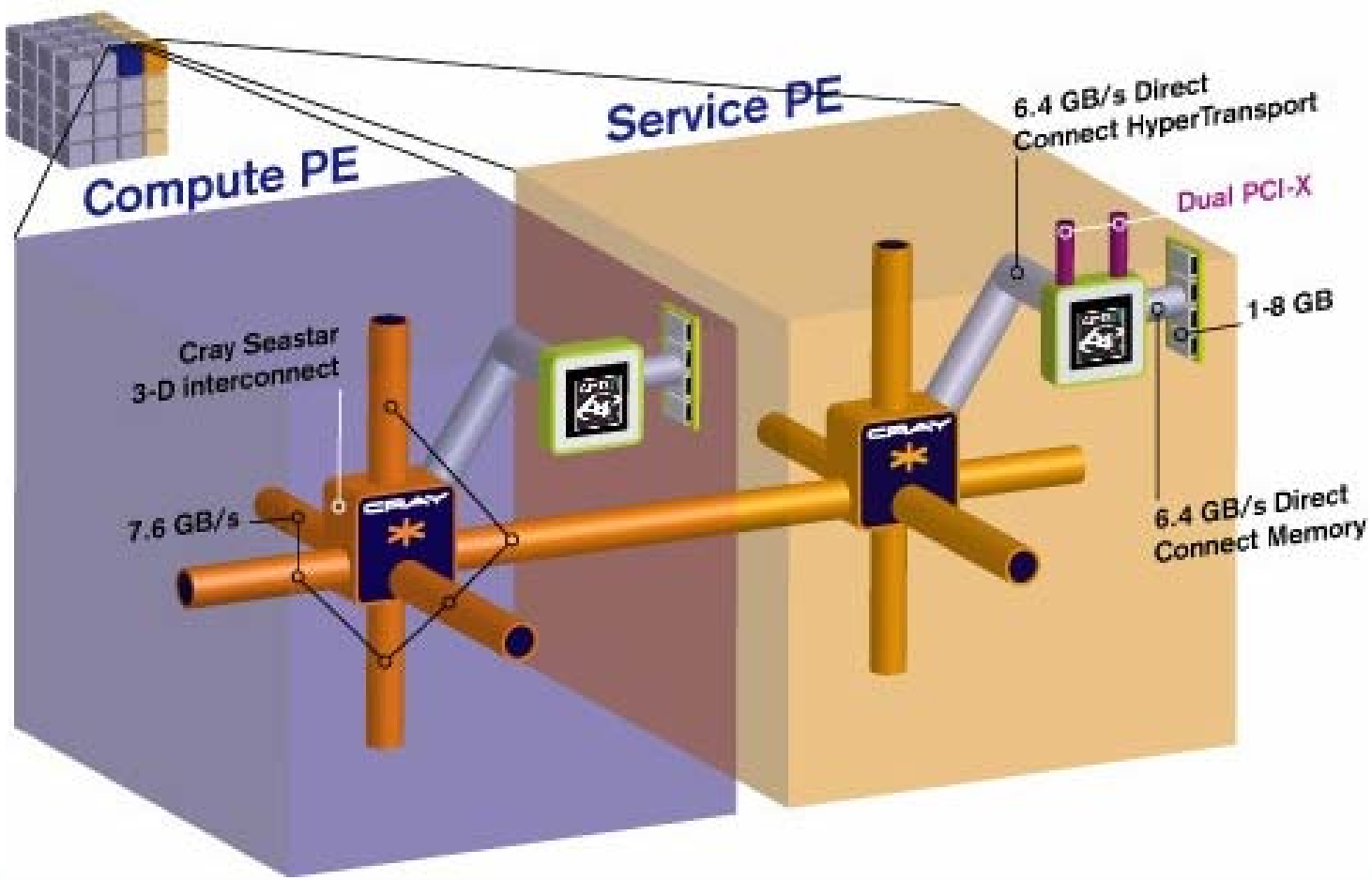


Image courtesy of Cray.

AMD Opteron

- AMD Opteron Model 150
  - Processor core / 2.4 Ghz
    - three integer units
    - one floating-point unit which is capable of two floating-point operations per cycle
    - 4.8 GFLOPS
  - Integrated memory controller
  - Three 16b 800 Mhz HyperTransport (HT) links
  - L1 cache: 64KB I and D caches
  - L2 cache: 1MB Unified
- Model 150 has three HT links but none support coherent HT (for SMPs)
  - Lower latency to main memory than SMP capable processors

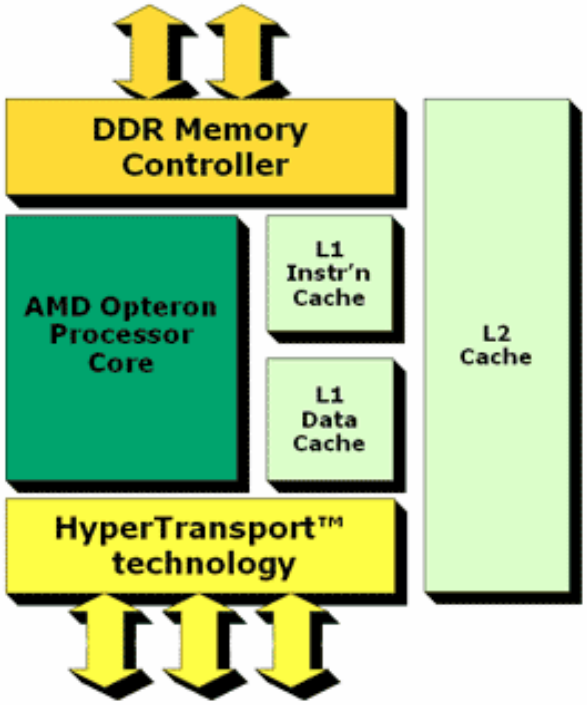


Image courtesy of AMD.

Cray SeaStar Interconnect ASIC

- Routing and communications ASIC
- Connects to the Opteron via 6.4 GBps HT link
- Connects to six neighbors via 7.6 GBps links
- Topologies include torus, mesh
- Contains
  - PowerPC 440 chip, DMA engine, service port, router
- Notice
  - No PCI bus in transfer path
  - Interconnect Link BW is greater than Opteron Link BW
  - Carries all message traffic in addition to IO traffic

Software

- Operating systems
  - Catamount
    - Lightweight kernel w/ limited functionality to improve reliability, performance, etc.
  - Linux
- Portals Communication Library
- Scalable application launch using Yod
- Programming environments
  - Apprentice, PAT, PAPI, mpiP
  - Totalview
- Filesystems
  - Scratch space through Yod
  - Lustre
- Math libraries
  - ACML 2.7

Evaluations

- Goals
  - Determine the most effective approaches for using the each system
  - Evaluate benchmark and application performance, both in absolute terms and in comparison with other systems
  - Predict scalability, both in terms of problem size and in number of processors

- We employ a hierarchical, staged, and open approach
  - Hierarchical
    - Microbenchmarks
    - Kernels
    - Applications
  - Interact with many others to get the best results
  - Share those results

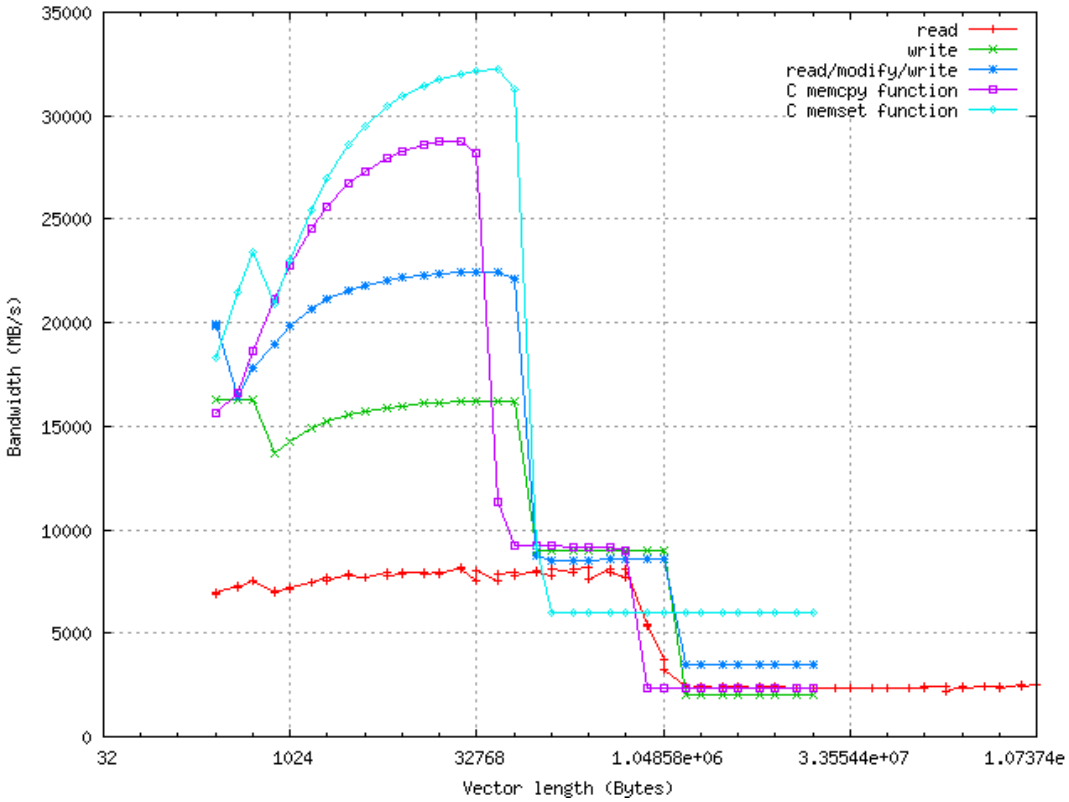
Microbenchmarks

- Microbenchmarks characterize specific components of the architecture

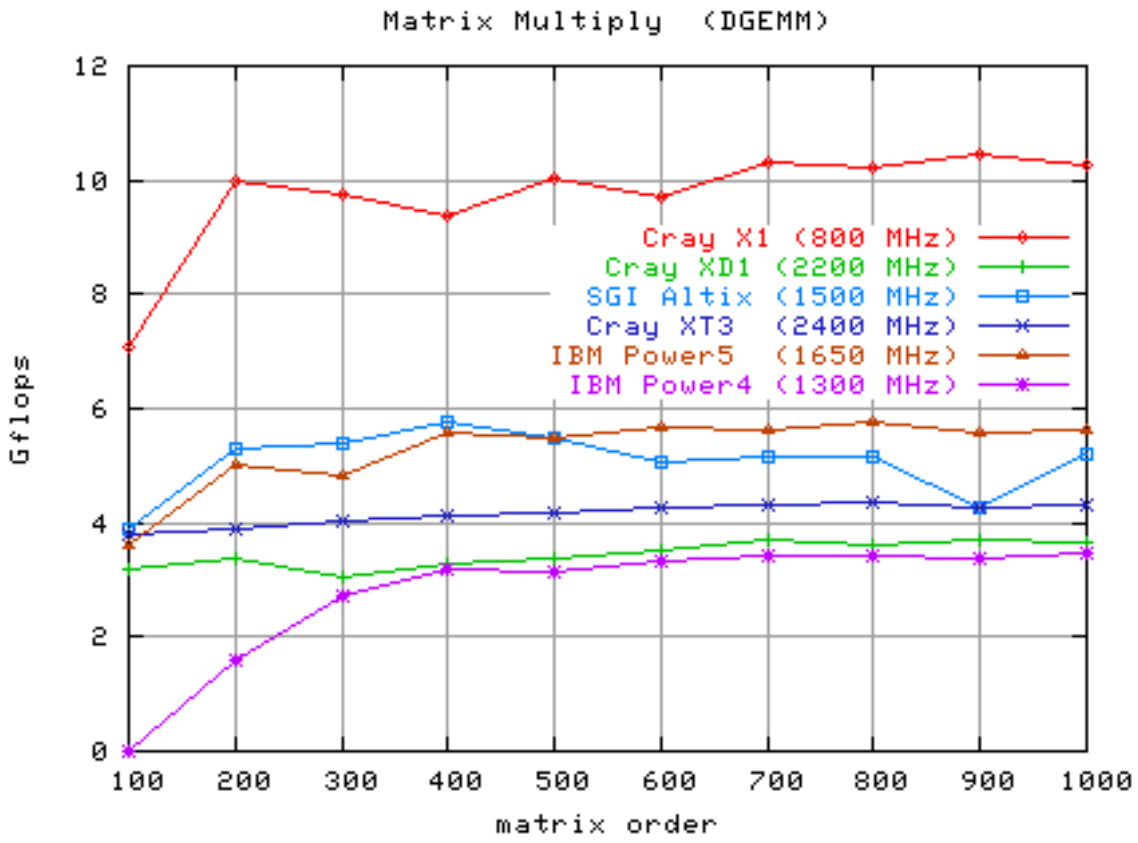
- Microbenchmark suite tests
  - arithmetic performance,
  - memory-hierarchy performance,
  - task and thread performance,
  - message-passing performance,
  - system and I/O performance, and
  - parallel I/O

Memory Performance

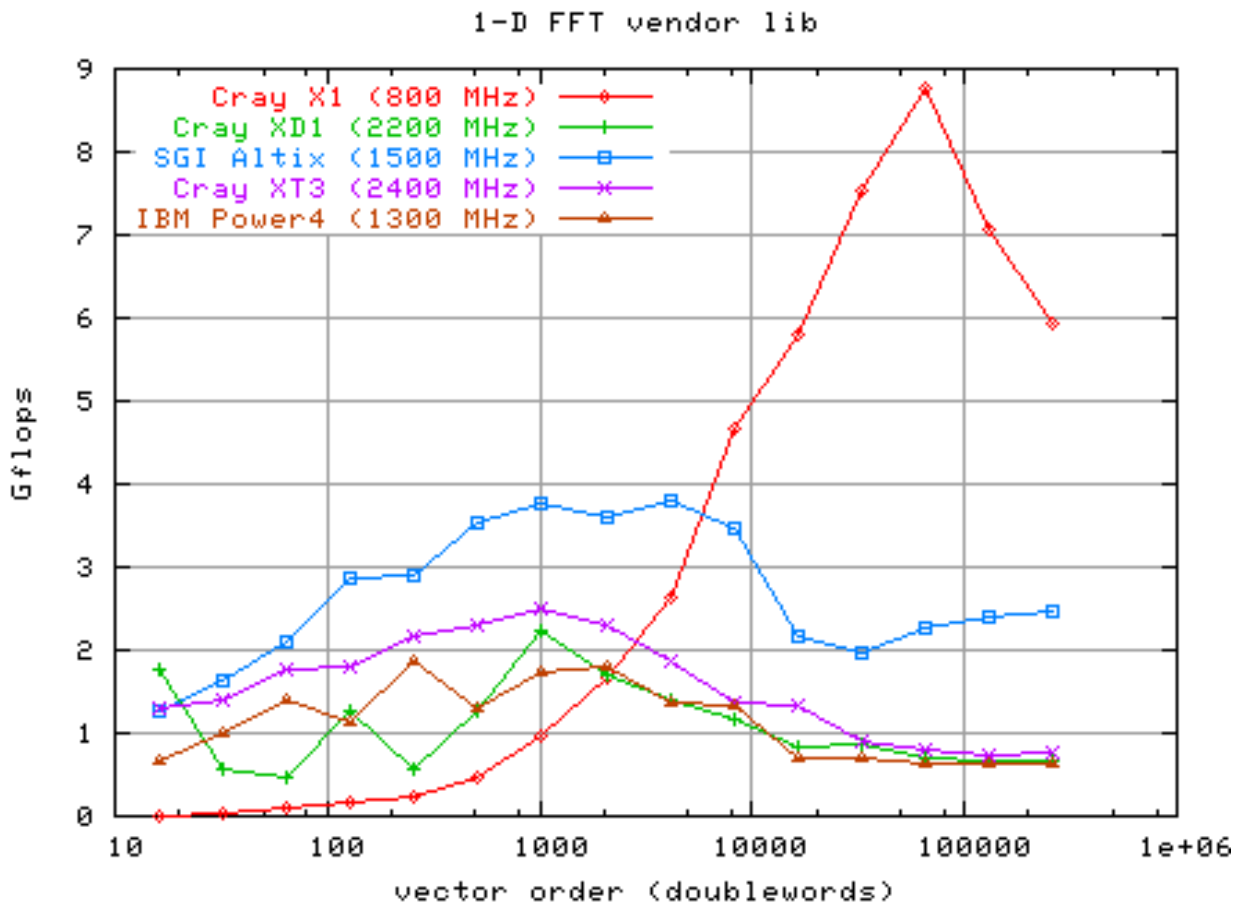
Platform	Measured Latency to Main Memory (ns)
Cray XT3 / Opteron 150 / 2.4	51.41
Cray XD1 / Opteron 248 / 2.2	86.51
IBM p690 / POWER4 / 1.3	90.57
Intel Xeon / 3.0	140.57



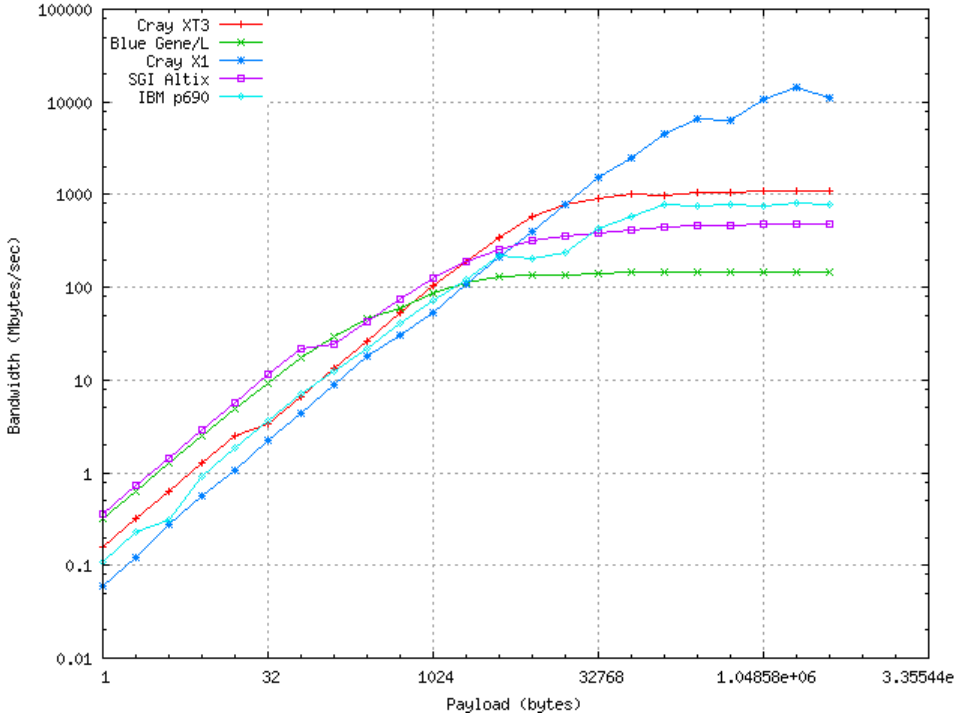
DGEMM Performance

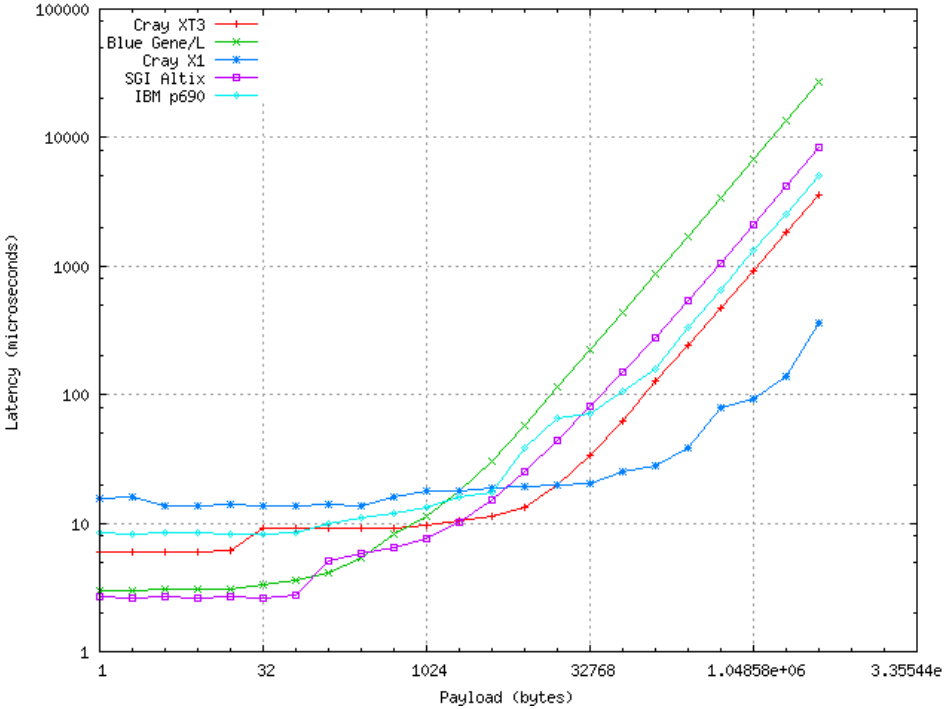


FFT Performance



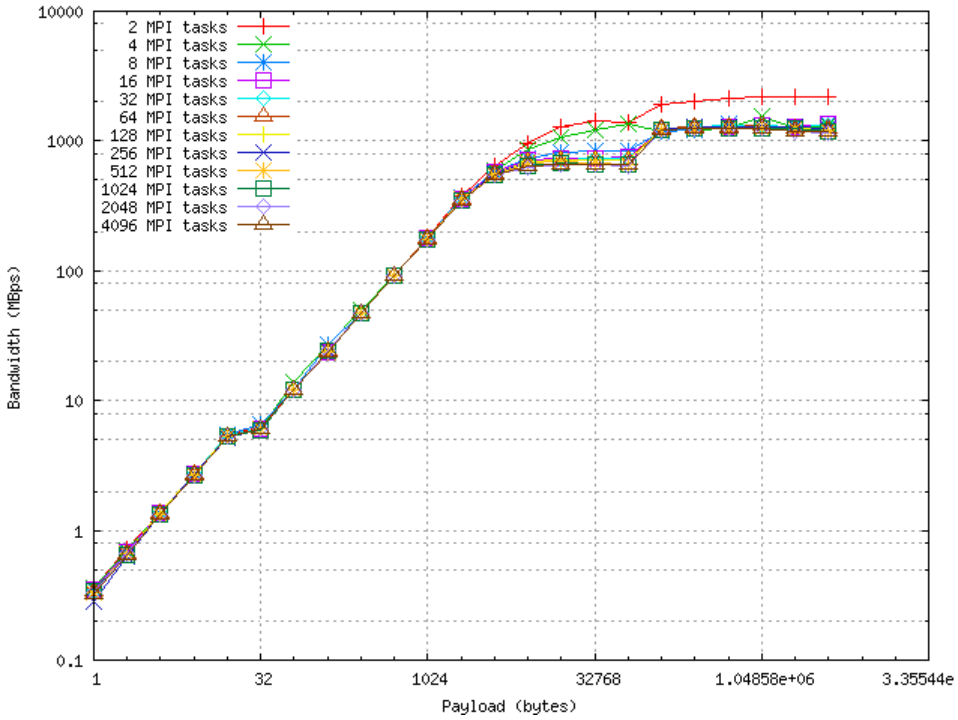
Ping Pong Performance

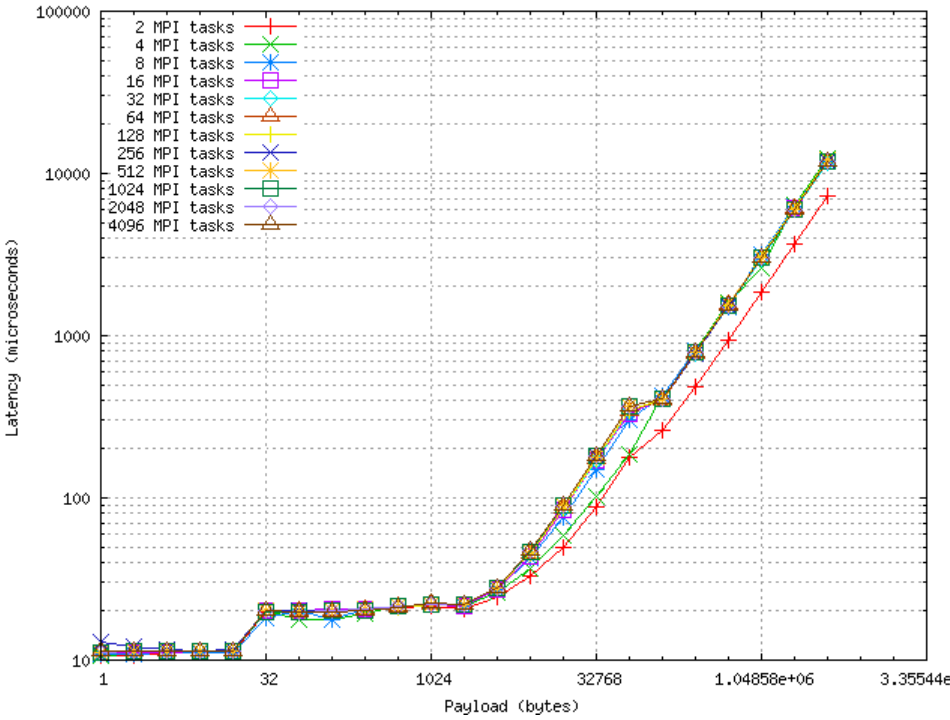




Exchange Performance

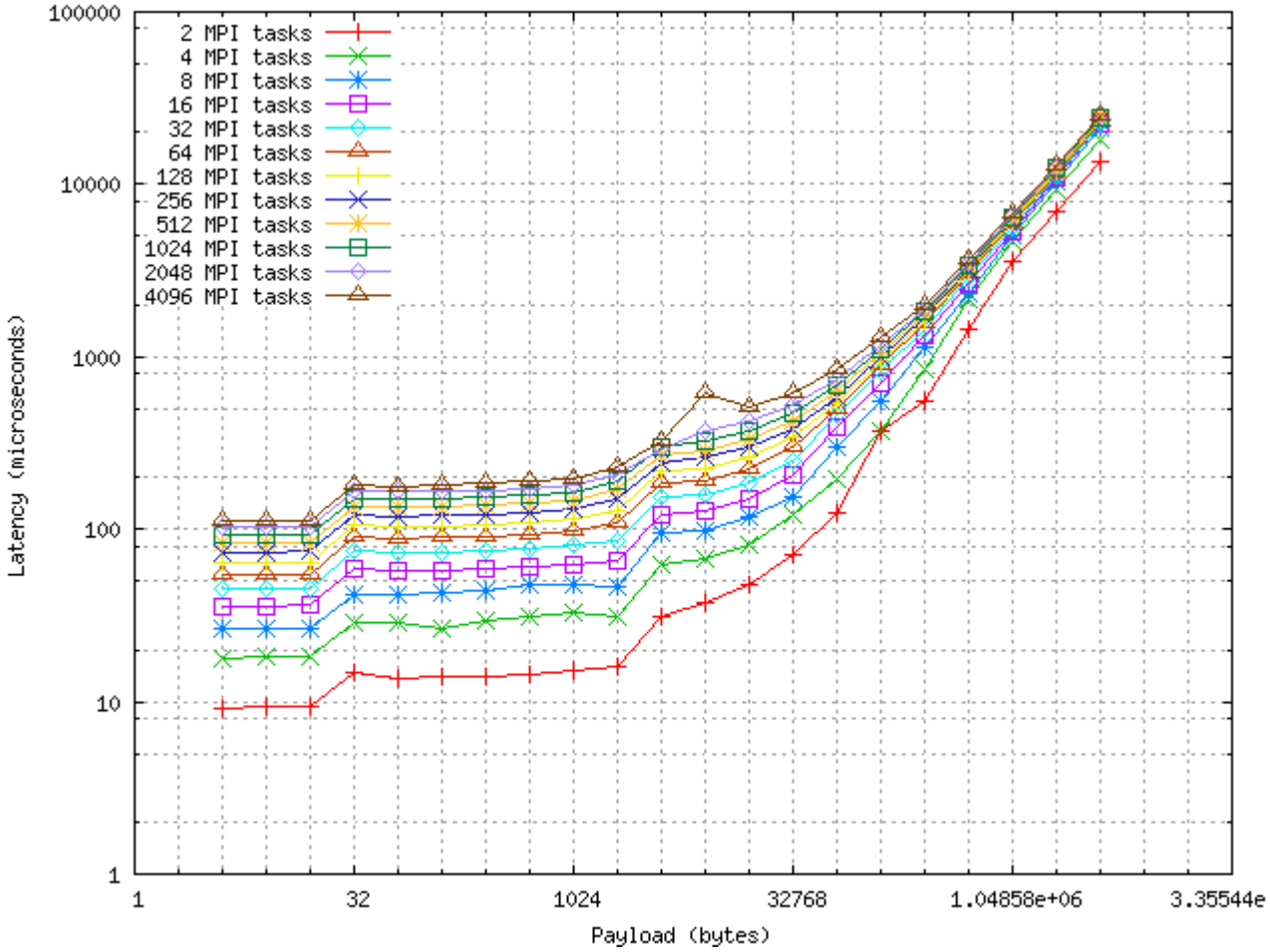
2 - 4096 processors





Allreduce Performance

2-4096 processors



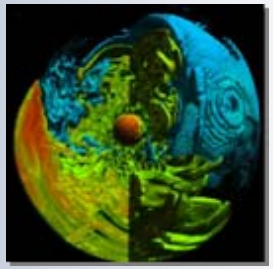
HPC Challenge Benchmark

http://icl.cs.utk.edu/hpcc/

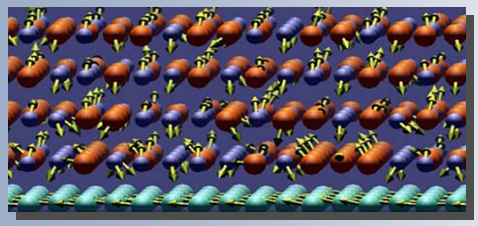
To be announced this week...

ORNL has Major Efforts Focusing on Grand Challenge Scientific Applications

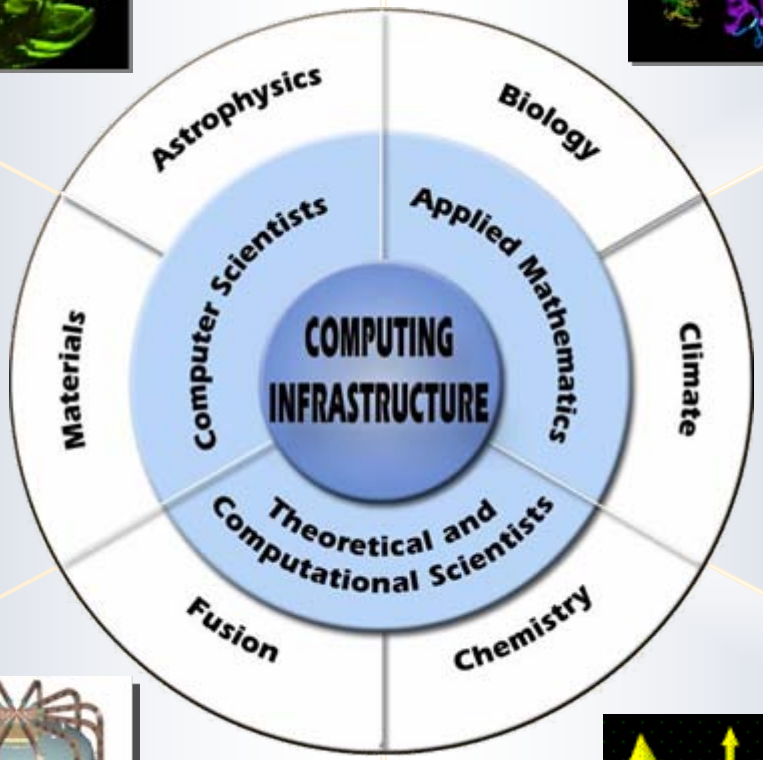
SciDAC Astrophysics

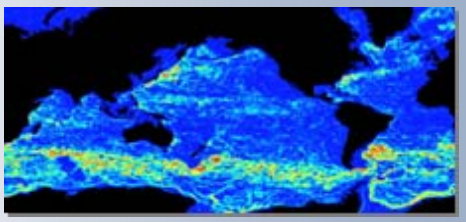


Genomes to Life



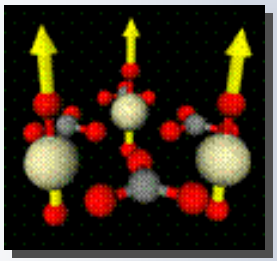
Nanophase Materials





SciDAC Climate

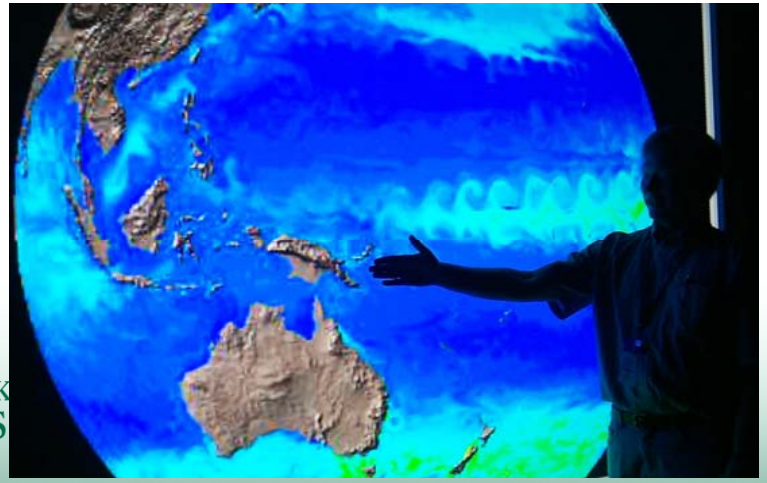
SciDAC Fusion



SciDAC Chemistry

Climate Modeling

- Community Climate System Model (CCSM) is the primary model for global climate simulation in the USA
  - Community Atmosphere Model (CAM)
  - Community Land Model (CLM)
  - Parallel Ocean Program (POP)
  - Los Alamos Sea Ice Model (CICE)
  - Coupler (CPL)
- Running Intergovernmental Panel on Climate Change (IPCC) experiments



Science and technology

The Economist November 13th 2004 87

Also in this section

89 Sorting sperm with optical tweezers

89 3D television

Climate change

A canary in the coal mine

The Arctic seems to be getting warmer. So what?

"CLIMATE change in the Arctic is a reality now!" So insists Robert Corell, an oceanographer with the American Meteorological Society. Wild-eyed proclamations are all too common when it comes to global warming, but in this case his assertion seems well founded.

Dr Corell heads a team of some 300 scientists who have spent the past four years investigating the matter in a process known as the Arctic Climate Impact Assessment (ACIA). The group, drawn from the eight countries with territories inside the Arctic Circle, has just issued a report called "Impacts of a Warming Arctic", a lengthy summary of the principal scientific findings. A second report, which will sketch out recommended policies, is due out in a few weeks. A third, far heftier tome detailing all the scientific findings will not come out for some months yet.

Already, though, the ACIA has made a splash. One reason is the inevitable wrangling over policy recommendations. News reports have suggested that the Bush administration has tried to suppress signs of support in the second, as yet unreleased, report, for the UN's Kyoto protocol or other mandatory policies for the control of greenhouse-gas emissions. But even setting politics aside, this week's scientific report has still created a stir with its bold assessment of polar warming.

At first sight, its conclusions are not so surprising. After all, scientists have long suspected that several factors lead to greater temperature swings at the poles than elsewhere on the planet. One is albedo—the posh scientific name for how much sunlight is absorbed by a planet's surface, and how much is reflected. Most of the polar regions are covered in snow and ice, which are much more reflective than soil or ocean. If that snow melts, the exposure of dark earth (which absorbs heat) acts as a feedback loop that accelerates warming. A second factor that makes the poles special is that the atmosphere is thinner there than at the equator, and so less energy is required to warm it up. A third factor is that less solar energy is lost in evaporation at the frigid poles than in the steamy tropics.

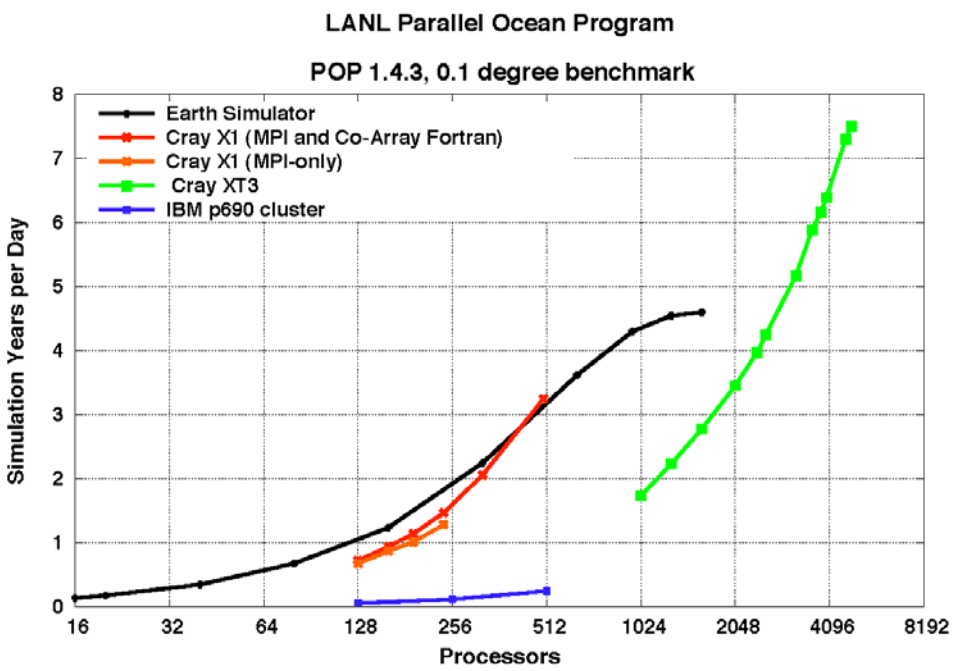
And yet the language of this week's report is still eye-catching: "the Arctic is now experiencing some of the most rapid and severe climate change on Earth." The last authoritative assessment of the topic was done by the UN's Intergovernmental Panel on Climate Change (IPCC) in 2001. That report made headlines by predicting a rise in sea level of between 10cm (four inches) and 90cm, and a temperature rise of between 1.4°C and 5.8°C over this century. However, its authors did not feel confident in predicting either rapid polar warming or the speedy demise of the Greenland ice sheet. Pointing to evidence gathered since the IPCC report, this week's report suggests trouble lies ahead.

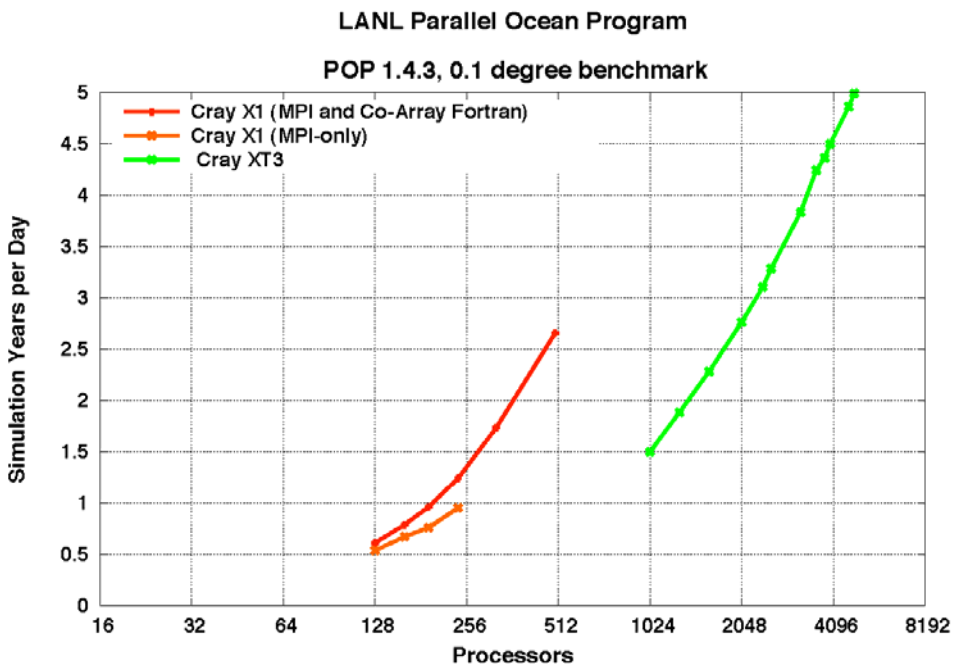
Hot on top

The ACIA reckons that in recent decades average temperatures have increased almost twice as fast in the Arctic as they have in the rest of the world. Sceptics argue that there are places, such as the high latitudes of the Greenland ice sheet and some buoys at sea, where temperatures seem to have fallen. On the other hand, there are also places, such as parts of Alaska, where they have risen far faster than average. Robin Bell, a geophysicist at Columbia University who was not involved in the report's compilation, believes that such conflicting local trends point to the value of the international, interdisciplinary approach of this week's report. As he observes, "climate change, like the weather, can be patchy and you can get fooled unless you look at the whole picture."

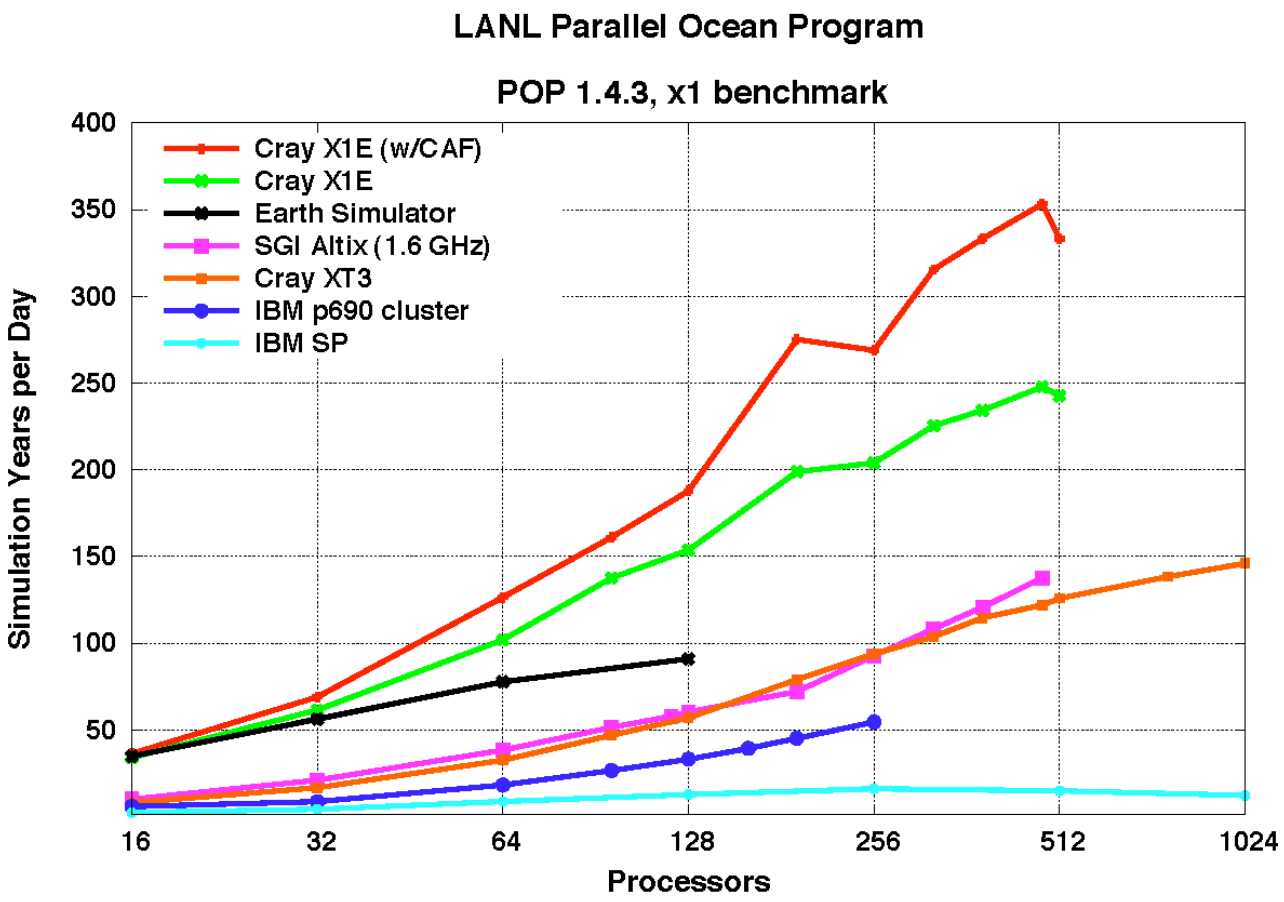
And there is other evidence of warming to bolster the ACIA's case. For example, the report documents the widespread melting of glaciers and of sea ice, a trend already making life miserable for the polar bears and seals that depend on that ice. It also notes a shortening of the snow season. The most worrying finding, however, is evidence—still preliminary—that the

Climate / Parallel Ocean Program / POP

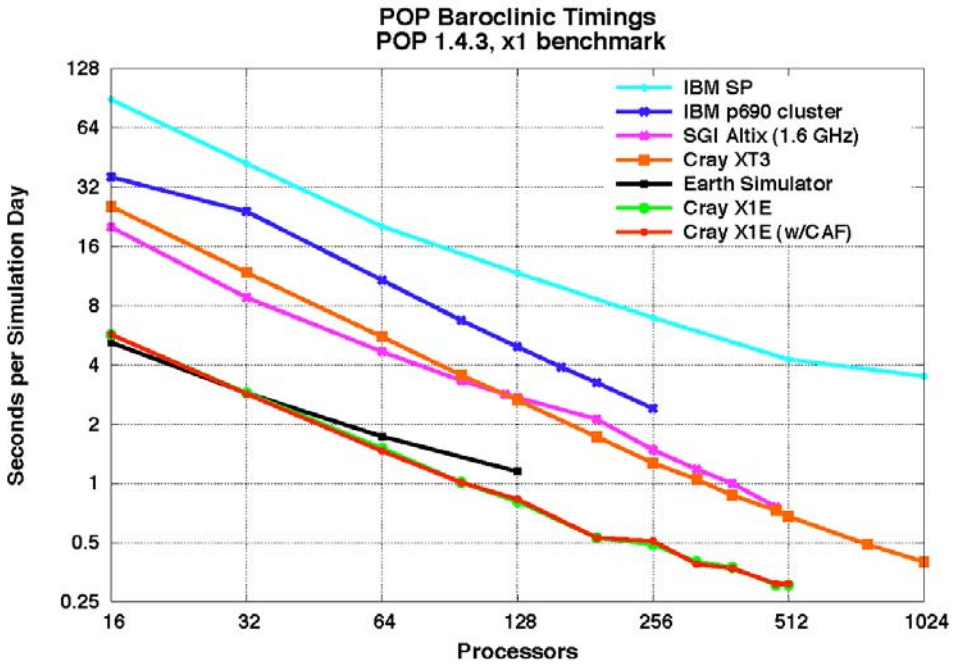


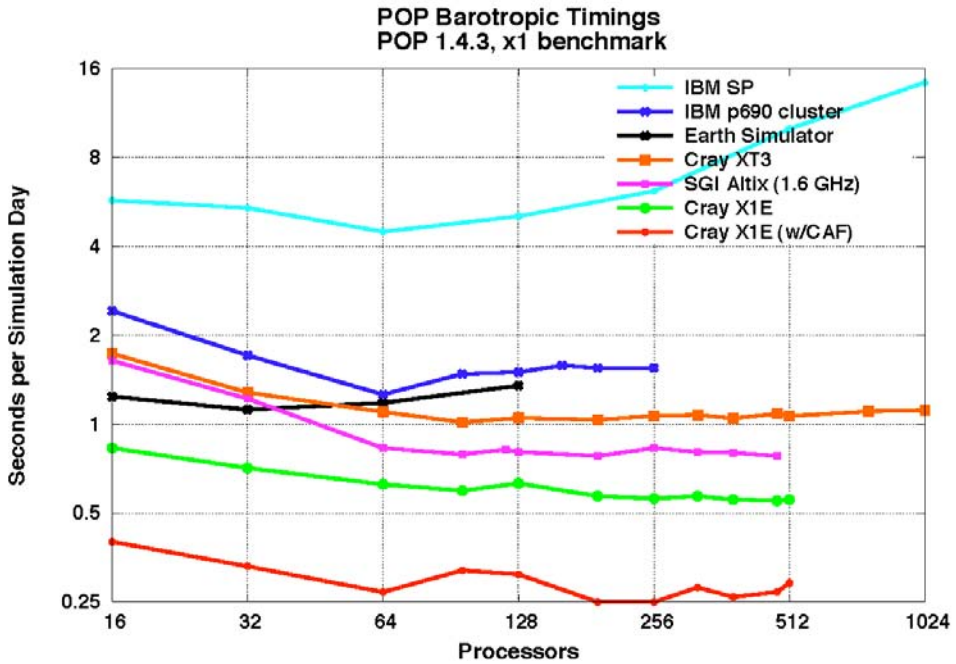


Climate / Parallel Ocean Program / POP

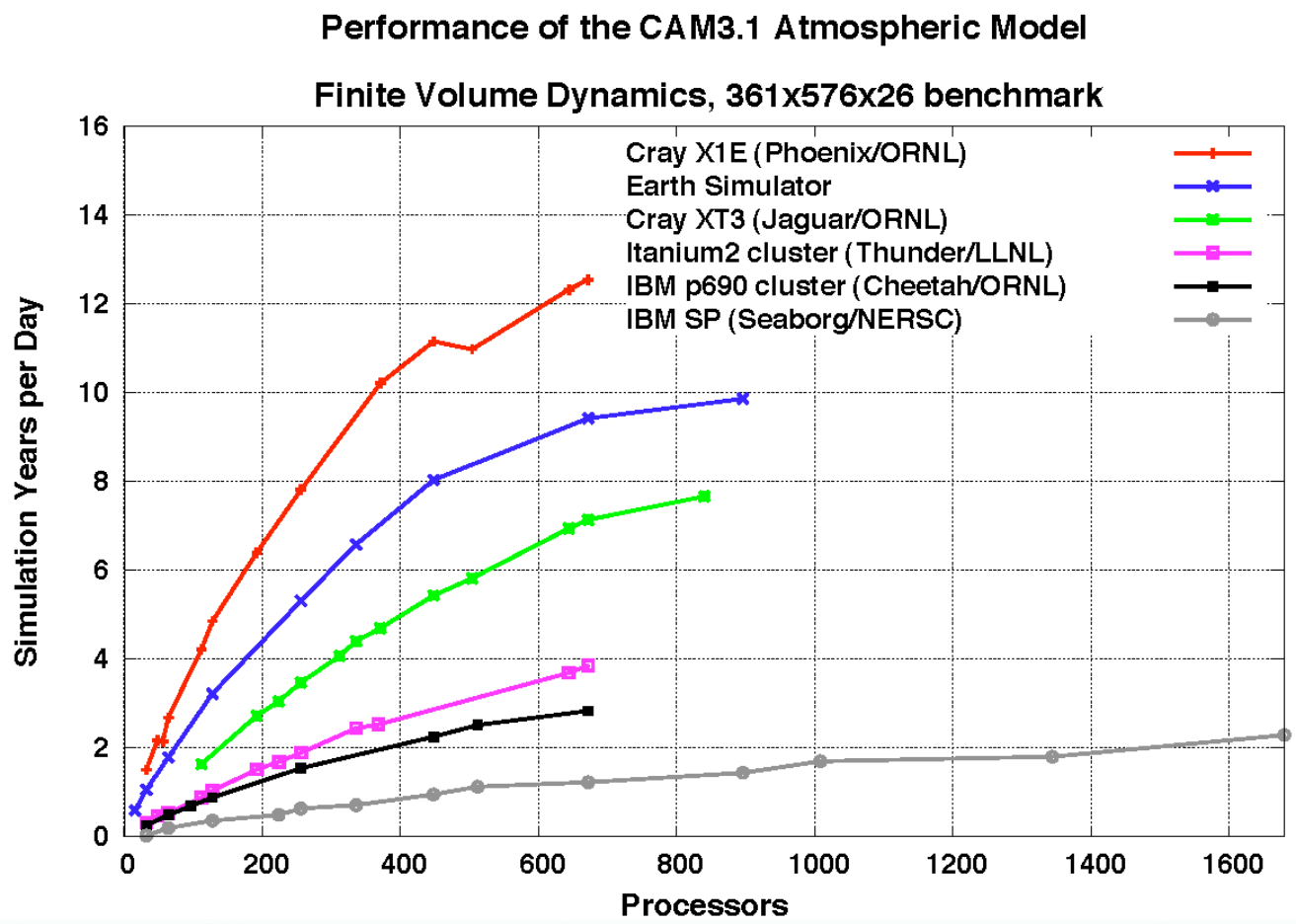


Climate / POP / Phase Timings



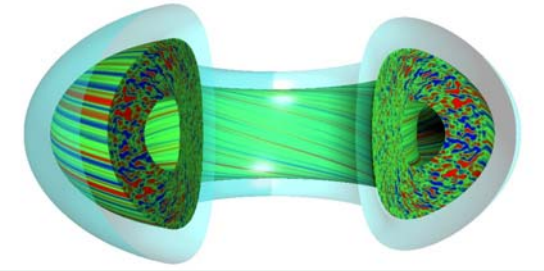


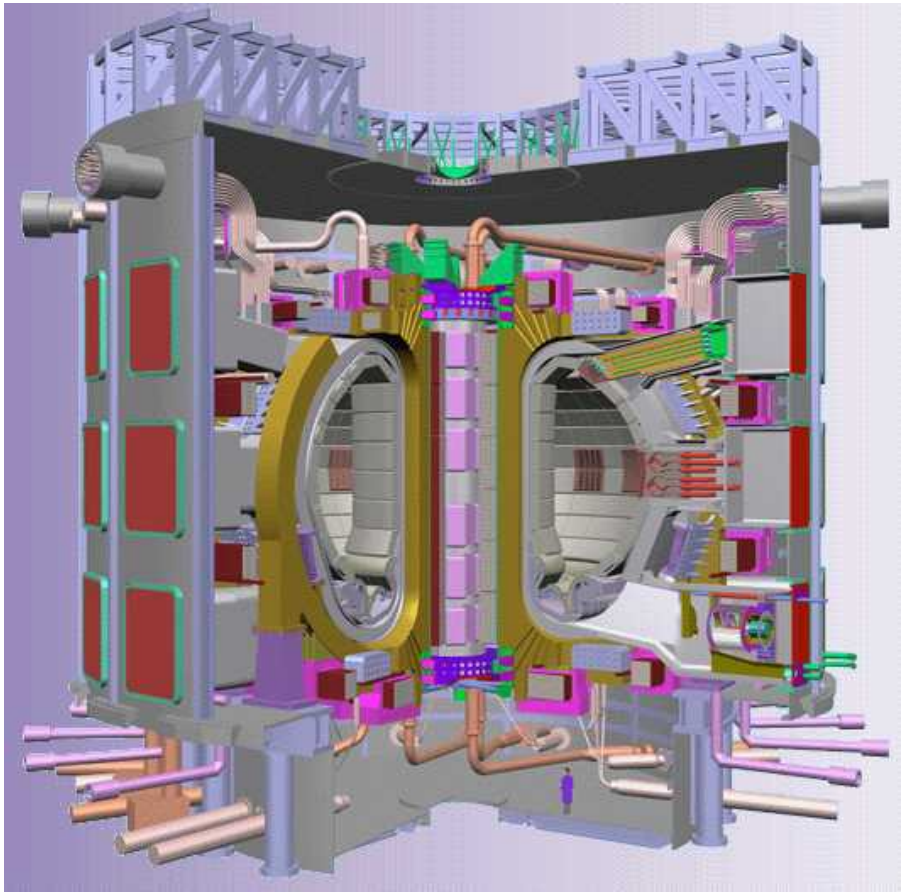
Climate / CAM


Performance of the CAM3.1 Atmospheric Model
Finite Volume Dynamics, 361x576x26 benchmark
Cray X1E (Phoenix/ORNL)
Earth Simulator
Cray XT3 (Jaguar/ORNL)
Itanium2 cluster (Thunder/LLNL)
IBM p690 cluster (Cheetah/ORNL)
IBM SP (Seaborg/NERSC)
Simulation Years per Day
Processors
0
2
4
6
8
10
12
14
16
0
200
400
600
800
1000
1200
1400
1600

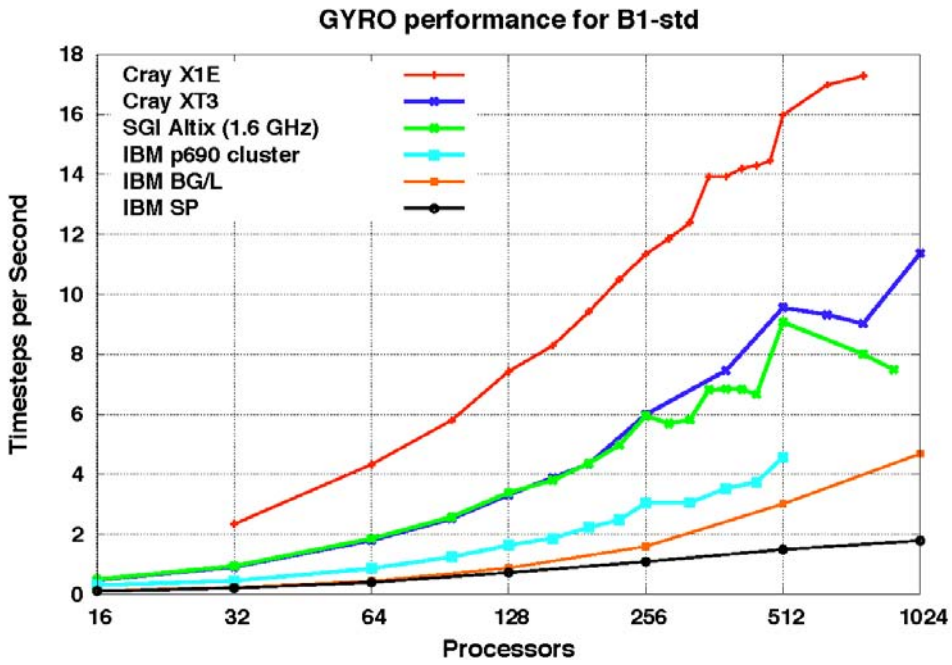
Fusion

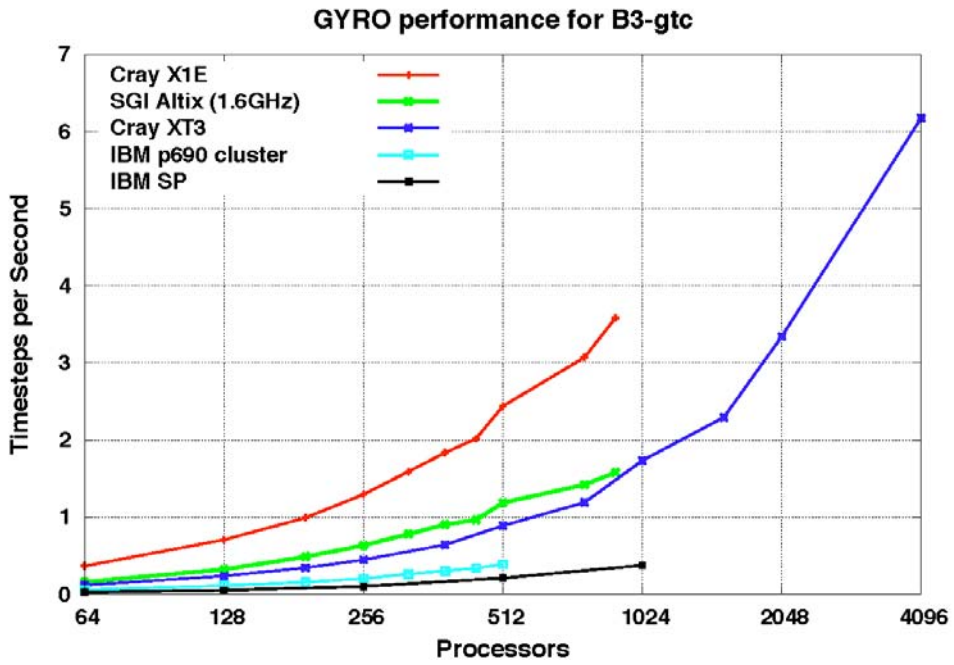
- Advances in understanding tokamak plasma behavior are necessary for the design of large scale reactor devices (like ITER)
- Multiple applications used to simulate various phenomena w/ different algorithms
  - GYRO
  - NIMROD
  - AORSA3D
  - GTC





Fusion / GYRO

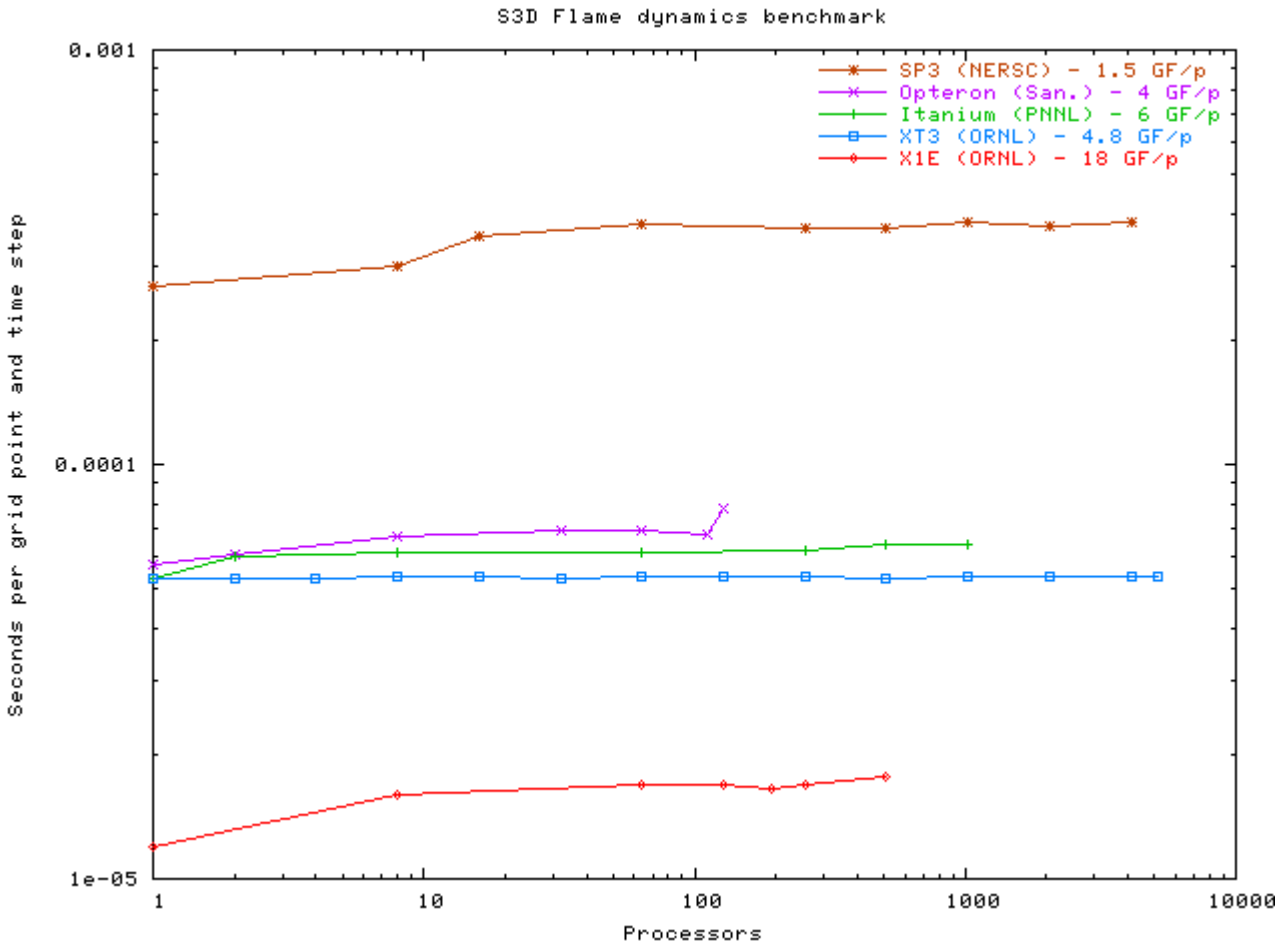

GYRO performance for B1-std
Cray X1E
Cray XT3
SGI Altix (1.6 GHz)
IBM p690 cluster
IBM BG/L
IBM SP
Timesteps per Second
Processors
0
2
4
6
8
10
12
14
16
18
16
32
64
128
256
512
1024


GYRO performance for B3-gtc
Cray X1E
SGI Altix (1.6GHz)
Cray XT3
IBM p690 cluster
IBM SP
Timesteps per Second
Processors
0
1
2
3
4
5
6
7
64
128
256
512
1024
2048
4096

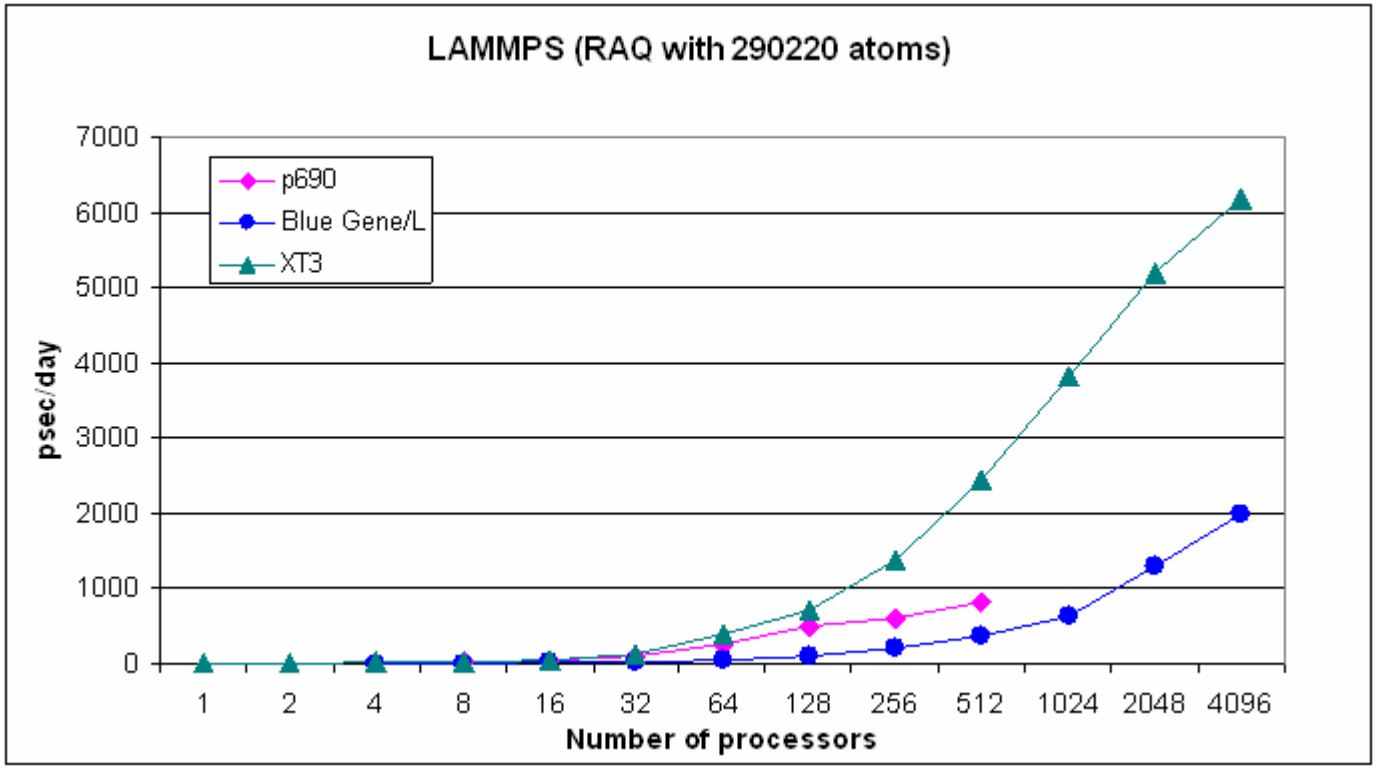
Fusion / GTC

NEED NEW DATA

Combustion / S3D



Biology / LAMMPS



Recent and Ongoing Evaluations

- Cray XT3
  - J.S. Vetter, S.R. Alam et al., "Early Evaluation of the Cray XT3 at ORNL," Proc. Cray User Group Meeting (CUG 2005), 2005.
- SGI Altix
  - T.H. Dunigan, Jr., J.S. Vetter, and P.H. Worley, "Performance Evaluation of the SGI Altix 3700," Proc. International Conf. Parallel Processing (ICPP), 2005.
- Cray XD1
  - M.R. Fahey, S.R. Alam et al., "Early Evaluation of the Cray XD1," Proc. Cray User Group Meeting, 2005, pp. 12.
- SRC
  - M.C. Smith, J.S. Vetter, and X. Liang, "Accelerating Scientific Applications with the SRC-6 Reconfigurable Computer: Methodologies and Analysis," Proc. Reconfigurable Architectures Workshop (RAW), 2005.
- Cray X1
  - P.A. Agarwal, R.A. Alexander et al., "Cray X1 Evaluation Status Report," ORNL, Oak Ridge, TN, Technical Report ORNL/TM-2004/13, 2004.
  - T.H. Dunigan, Jr., M.R. Fahey et al., "Early Evaluation of the Cray X1," Proc. ACM/IEEE Conference High Performance Networking and Computing (SC03), 2003.
  - T.H. Dunigan, Jr., J.S. Vetter et al., "Performance Evaluation of the Cray X1 Distributed Shared Memory Architecture," IEEE Micro, 25(1):30-40, 2005.
- Underway
  - XD1 FPGAs
  - ClearSpeed
  - EnLight
  - Multicore processors
  - IBM BlueGene/L
  - IBM Cell