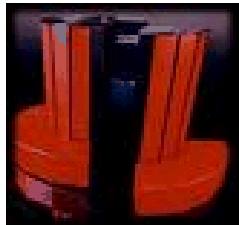


High Performance Computing (HPC) Overview

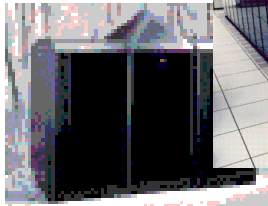
Constantine “Dino” Pavlakos
John Zepper

November 4, 2010

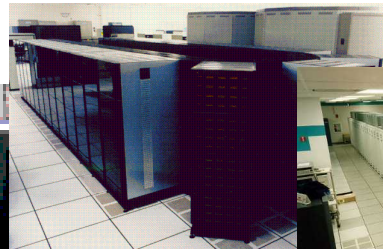
Sandia has a legacy of success in High Performance Computing



Cray XMP



nCUBE



Paragon



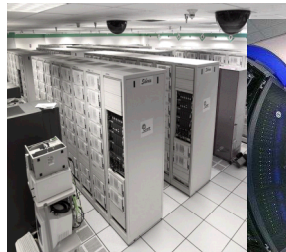
ASCI Red

• 1.3 Tflops



Red Storm (Cray XT series)

• #2 in 2006, now 284 Tflops, ~31.6K cores



Cplant™



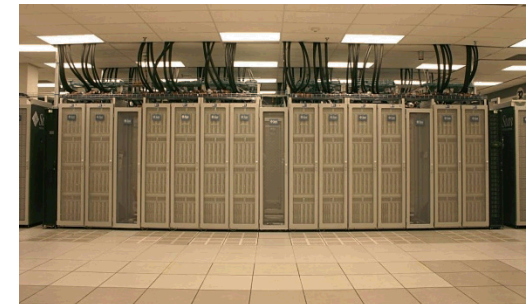
Thunderbird

• #6 in 2006, 53 Tflops,
~8.6K cores



TLCC

• 3 machines, 38 Tflops
ea, ~4.3K cores ea



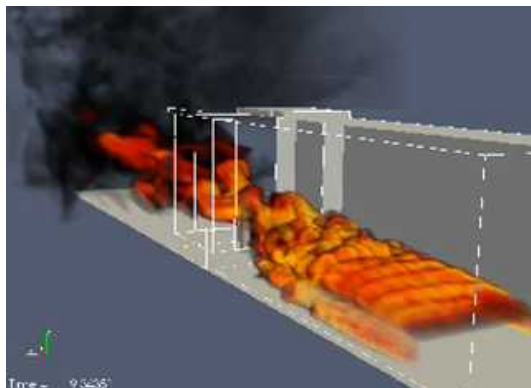
Red Sky/Mesa

• #10 in 2009 (aggregate), 433
Tflops, ~43K cores

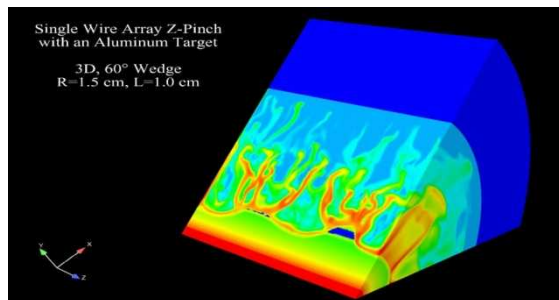
- Seminal work & Gordon Bell prize for demonstrating application speed-up on parallel machines (Benner, Gustafson, Montry)
- Key contributions to massively parallel system architectures and system software
- Key design contributions to the success of the Cray XT series

HPC capabilities rooted in Sandia's weapons programs are being used as part of a systems-based approach to solve complex problems in science & engineering

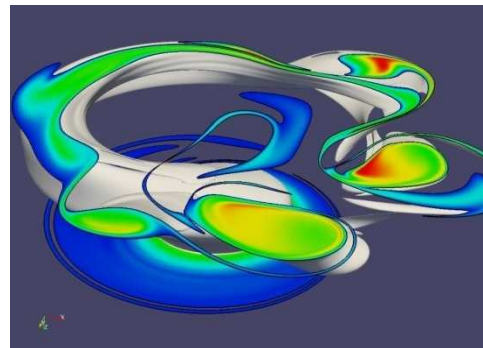
The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.



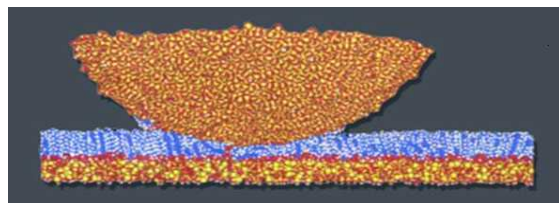
Fire Simulations



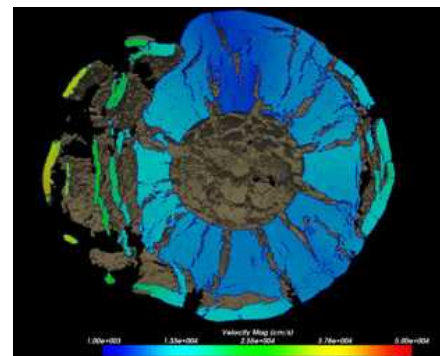
Pulsed Power –
Z-pinch Implosion Studies



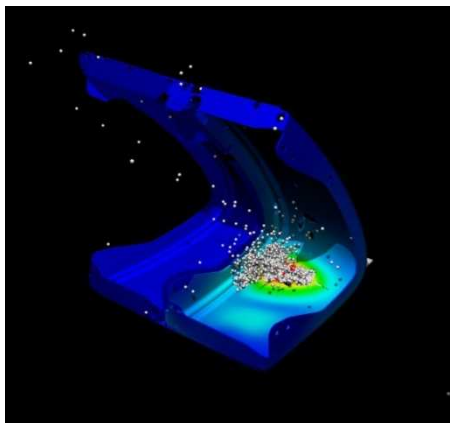
Climate Modeling



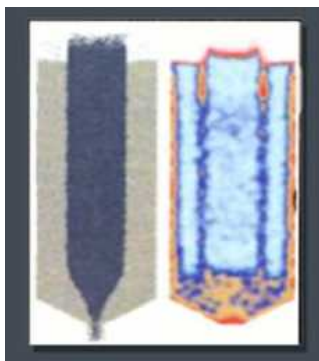
Microsystems Design/Analysis



Asteroid simulations

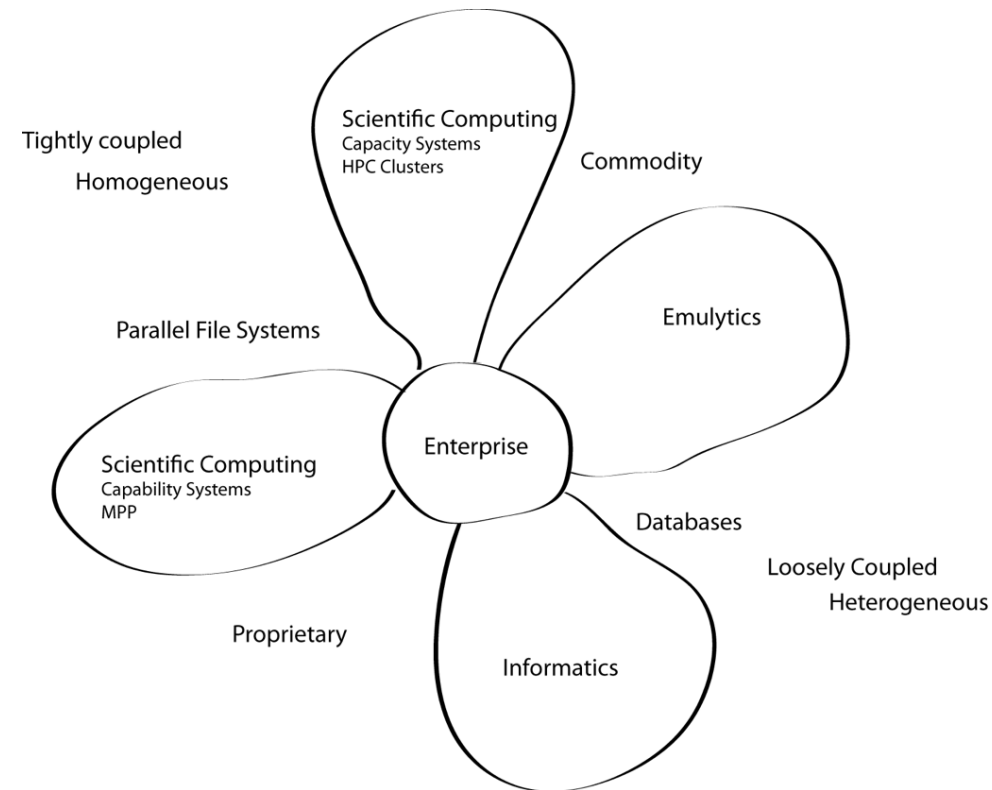


Space Shuttle Safety Analysis



Pebble Bed Nuclear
Reactor Design

Five Flavors of FI(ops)



Enterprise Computing

Desktop, Servers, Networking, Security...

Examples: Sandia Restricted Network (SRN)...

Scientific Computing

Capacity — aka Clusters

Examples: Red Sky, Unity, Glory, ...

Capability — aka Big Iron, MPP, etc.

Examples: Cielo, Red Storm, ...

Informatics — HPC-driven information analysis

Threaded vs. data intensive architectures

Examples: XMT + Netezza, ...

Emulytics, or Emulative Computing

Emulation of large-scale systems using smaller-scale systems at varying levels of fidelity

Live-Virtual-Constructive environments

Examples: Black ICE, NCR, I/O Range, ...



Red Sky Main themes

■ Cheaper

- 5X capacity of Tbird at 2/3 the cost
- Substantially cheaper per flop than our last large capacity machine purchase

■ Leaner

- Lower operational costs
- Three security environments via modular fabric
- Expandable, upgradeable, extensible
- Designed for 6yr. life cycle

■ Greener

- 15% less power ... 1/6th power per flop
- 40% less water ... 5M gallons saved annually
- 10X better cooling efficiency
- 4x denser footprint



CCF Room X50 Power Usage Effectiveness (PUE) Calculation Comparison of Current State to Alternate Scenarios

Total Power Used	2279 kW	Current State for all of X50 with: 33.4 kW per Rack 0.510 kW per ton Chiller Efficiency
IT Equipment Used	1794 kW	
Power Usage Effectiveness	1.27	

Total Power Used	2231 kW	Current State including incremental efficiencies: 0.434 per ton Chiller Efficiency 60% per day Lights Out
IT Equipment Used	1794 kW	
Power Usage Effectiveness	1.24	

Total Power Used	2138 kW	Current State including incremental efficiencies: 0.434 kW per ton Chiller Efficiency 60% per day Lights Out 0.2 kW per Plate Frame Heat Exchanger utilized 4.5 months per year
IT Equipment Used	1794 kW	
Power Usage Effectiveness	1.19	

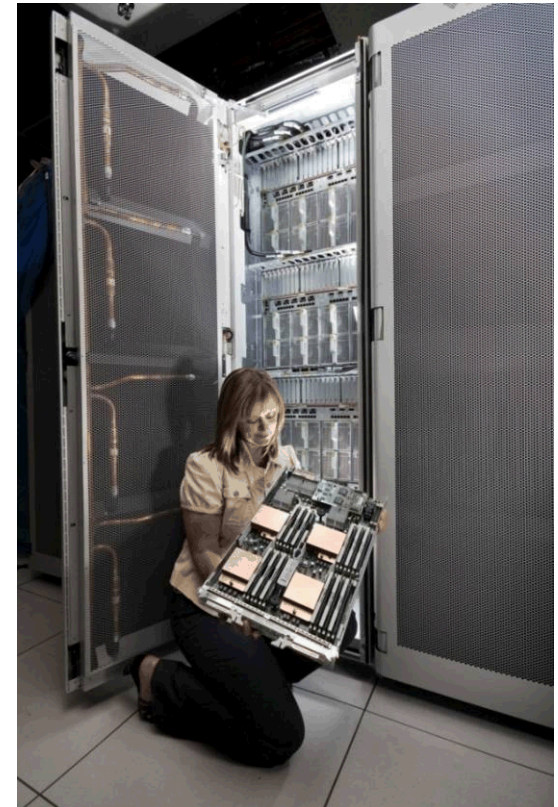
Total Power Used	1803 kW	Red Sky only: Removed non-Red Sky IT Equipment Removed all outside utilities except chilled water pump
IT Equipment Used	1694 kW	
Power Usage Effectiveness	1.06	



Red Sky Major innovations

In Partnership with Sun (now Oracle) and Intel

- **Bridging from capacity to capability**
 - Many “Red Storm” characteristics (scaling) at commodity price
 - 2-3X *faster* than Red Storm in mid range
 - 1/3 operational costs
- **Red Sky innovations**
 - Petascale midrange architecture
 - Intel Nehalem processor (2.93 vs 3.2GHz)
 - QDR Infiniband
 - 3D mesh/torus
 - Optical cabling
 - Optical Red/Black switching
 - Refrigerant cooling/glacier doors
 - Power distribution
 - Routing & Interconnect resiliency
 - Minimal Ethernet (RAS & mgmt. only)
 - Boot over IB
 - Software RAID (Intel processor)





Hardware facts:

505 TF Peak

Red Sky: 325 TF

Red Mesa: 180 TF

5,386 nodes (2,693 Sun X6275 blades)

Total for Rows A-F (Red Sky + Red Mesa)

2.93 GHz quad core, Nehalem X5570 processor

43,088 cores total

12 GB RAM per node (1.5 per core)

64 TB RAM total

3D torus InfiniBand

QDR via Mellanox ConnectX on motherboard and InfiniScale IV in QNEM

1,440 IB cables = 9.1 miles (220 miles of optical strands)

96- J4400 JBOD storage enclosures providing a total of 2 PB (raw) for scratch file systems

2304 1TB Seagate disks

R134a based cooling doors

1.7 MW power (PUE of 1.27)

1,848 square feet of space in 6 rows

68 Sun C48 cabinets

up to 96 nodes per rack

up to 768 cores per rack

Software facts:

CentOS 5.3

OFED 1.4.1

SNL modified OpenSM (Subnet manager) with custom routing engine for 3D Torus (Torus-2QoS)

Diskless boot over IB using a custom isolinux boot strap or GPXE

SNL developed system management toolset

SNL developed RAS system

Lustre 1.8.x with patchless clients

X6275 Blade (2 nodes)





Red Sky – User Feedback Summary

User response suggests that Red Sky is a unique capability for Sandia's high performance community that will be in high demand. Users are reporting performance speedups as high as 3 to 4 times that of Sandia's other capacity computing clusters. A sampling of user feedback includes:

"Red Sky is great, very fast machine. The transition to Red Sky was easy."

"Red Sky is an extremely fast machine that is robust for jobs in the range of four thousand cores and smaller."

"I got my first job running in less than 10 minutes after porting input/restart files."

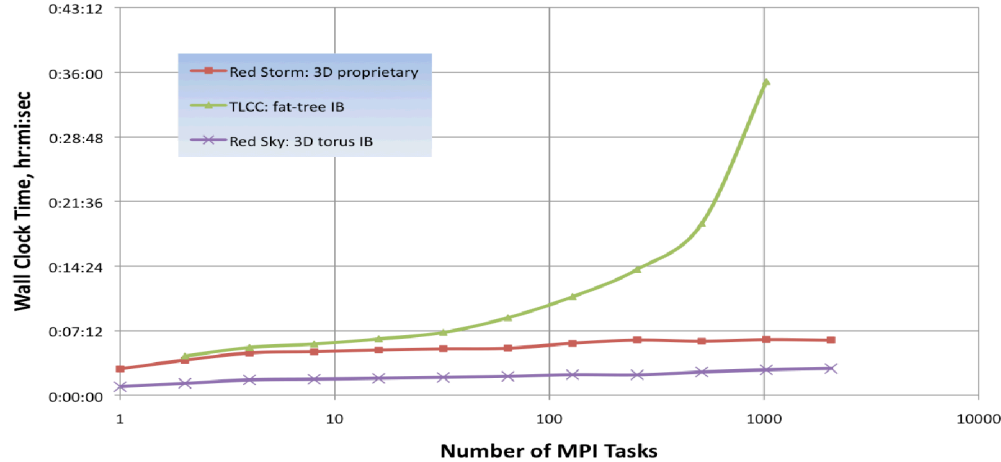
"My experience on Red Sky set the record for me as far as minimum time required from originally getting on a machine to being able to run."

"Red Sky is also backed by a superb and very helpful support team."

"The transition was remarkably smooth."

"Quite honestly I have never seen such a speed-up since the N-Cube to Janus transition about 10 years ago!"

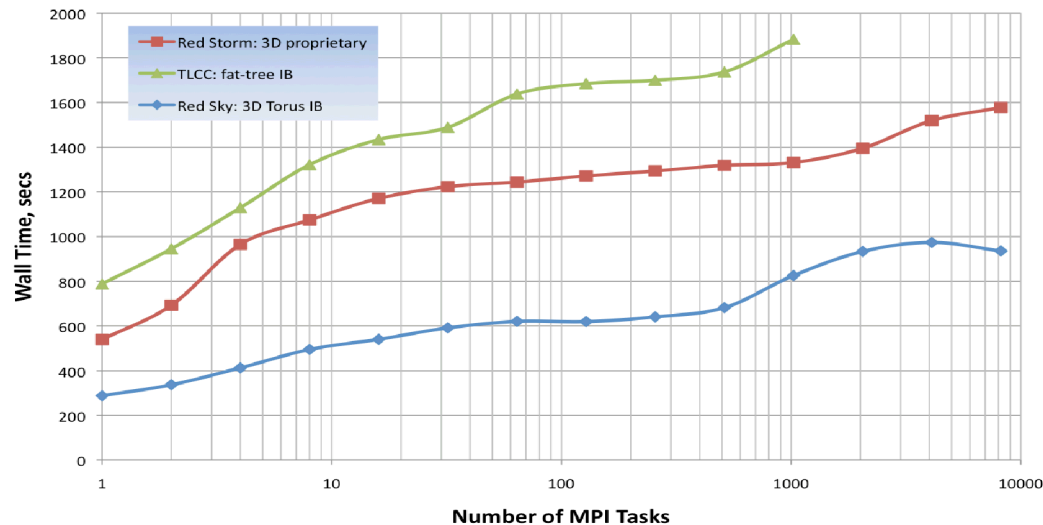
PRESTO 4.14.1: Walls Collision (ACME) Weak Scaling
10,240 Elements/task; 596 Time Steps



Red Sky performance

Application performance analysis and scaling results have been impressive

CTH Shape Charge: Wall Time for 100 time Steps:
Weak Scaling with 80x192x80 Cells/core





Let's go have a look ... Red Sky / Red Mesa Layout

