LA-UR- 11-03818

Approved for public release;
distribution is unlimited.

Title:	The Cielo capability Supercomputer
Author(s):	Manuel Vigil Douglas Doerfler
Intended for:	JOWOG 34

Los Alamos
NATIONAL LABORATORY
EST.1943

Form 836 (7/06)

The Cielo Capability Supercomputer

A B S T R A C T

Presentation gives an overview of the Cielo supercomputer design, architecture, and is also a progress report on its installation and future at Los Alamos National Laboratory. One early application simulation created on Cielo (Slide 13)is included as a still and as a brief video. this material has been reviewed and approved by a DC as unclassified.

The Cielo Capability Supercomputer

Manuel Vigil (LANL), Douglas Doerfler (SNL)

Los Alamos NATIONAL LABORATORY EST. 1943

Sandia National Laboratories

Operated by the Los Alamos National Security, LLC for the DOE/NNSA

NNSA

cielo

NNSA New Mexico Alliance for Computing at Extreme Scale

The Cielo Platform

- Cray, Inc. selected to deliver Cielo Platform by ACES
  - Cielo Platform (1.03 Peak PetaFlops in FY10)
    - 5% Cielo Nodes with Additional Memory (for Viz partition)
    - Application Regression Testbed (Cielito)
    - System Development Testbed (x2)
  - Additional Delivery Option (.33 additional PFs in FY11)
  - Total of 1.37 PF system (with room for expansion)
- Cielo - provide petascale production capability computing for the ASC tri-lab community (LLNL, SNL, LANL)
  - Capable of running a single application across the entire machine
  - Usage Model will follow the ASC's Capability Computing Campaign (CCC) process

cielo

NNSA New Mexico Alliance for Computing at Extreme Scale

Design Philosophy & Goals

- Petascale production capability to be deployed in Q1FY11
  - Take over the role Purple currently plays
  - Usage Model will follow the Capability Computing Campaign (CCC) process
  - Capability: Capable of running a single application across the entire machine
- Easy migration of existing integrated weapons codes
  - MPI Everywhere is the nominal programming model
  - 2GB memory per core (minimum) to support current application requirements
- Productivity goal is to achieve a 6x to 10x improvement over Purple on representative CCC applications
  - Memory subsystem performance will be the major contributor to node performance
  - Interconnect performance will be major contributor to scaling performance
  - Reliability will be major contributor to CCC total time to solution
- Upgrade path to allow increased capability in out years
- Key challenges: Reliability, Power, HW and SW Scalability, Algorithmic Scaling to 80K to 100K MPI ranks

cielo
NNSA New Mexico Alliance for Computing at Extreme Scale

Cielo at a Glance

- Platforms
  - Cielo – Petascale Capability Computing Platform
  - Cielito – Small application development testbed
- Cray XE6 Architecture
  - 3D Torus Topology using Cray Gemini high-speed interconnect
  - AMD Magny-Cours based nodes
- Cray Linux Environment (CLE) System Software
  - ALPS Runtime with Moab batch scheduling
  - CrayPat & Apprentice2 performance analysis tools
  - TotalView debugger
  - PGI, Cray, Intel and GNU compiler suites
  - Etc.

- Integrated Visualization & Analysis Partition
  - 64 GB memory per node partition
- Integrated into LANL's Parallel Scalable Backbone Network (PaScalBB)
  - 10 PB of user available storage
  - > 200 GB/s of network bandwidth
  - 160 GB/sec of parallel file system bandwidth

Los Alamos NATIONAL LABORATORY EST. 1943

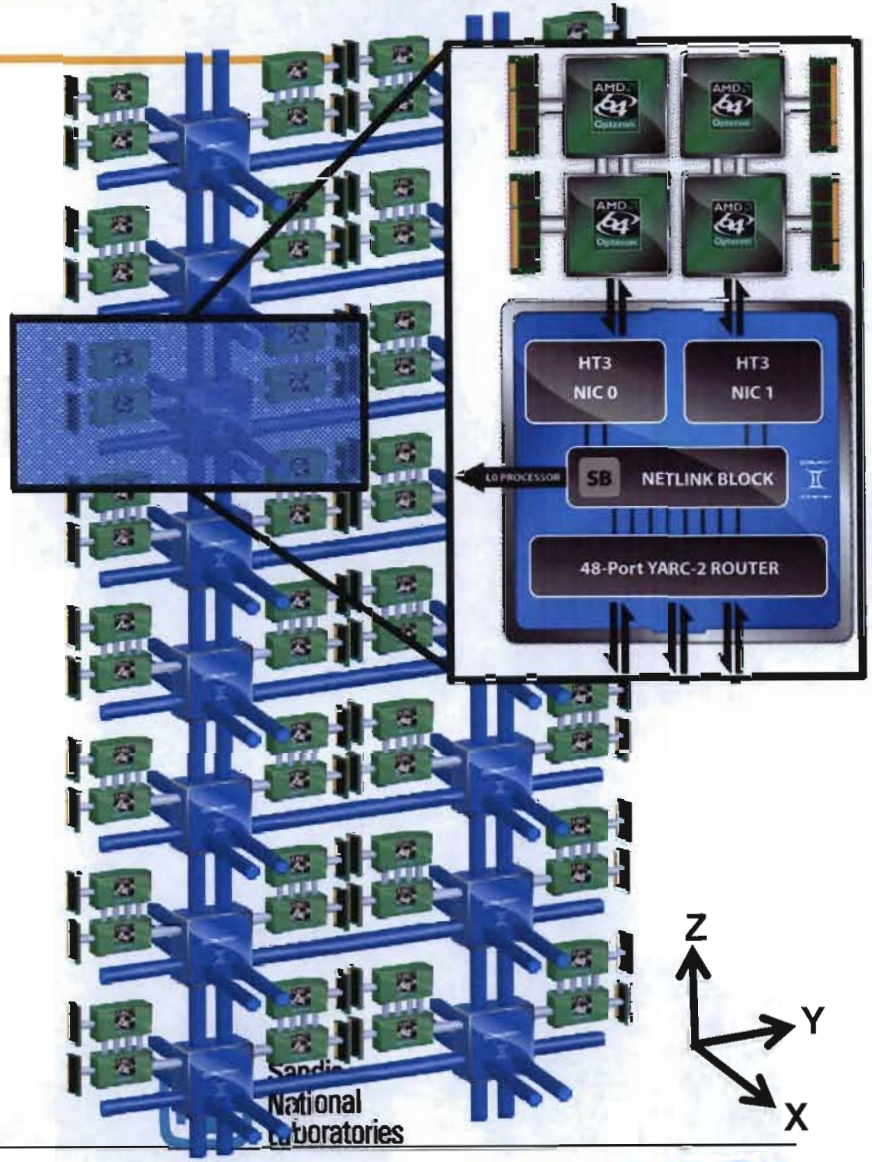
Sandia National Laboratories

cielo

NNSA New Mexico Alliance for Computing at Extreme Scale

Cielo Hardware Architecture

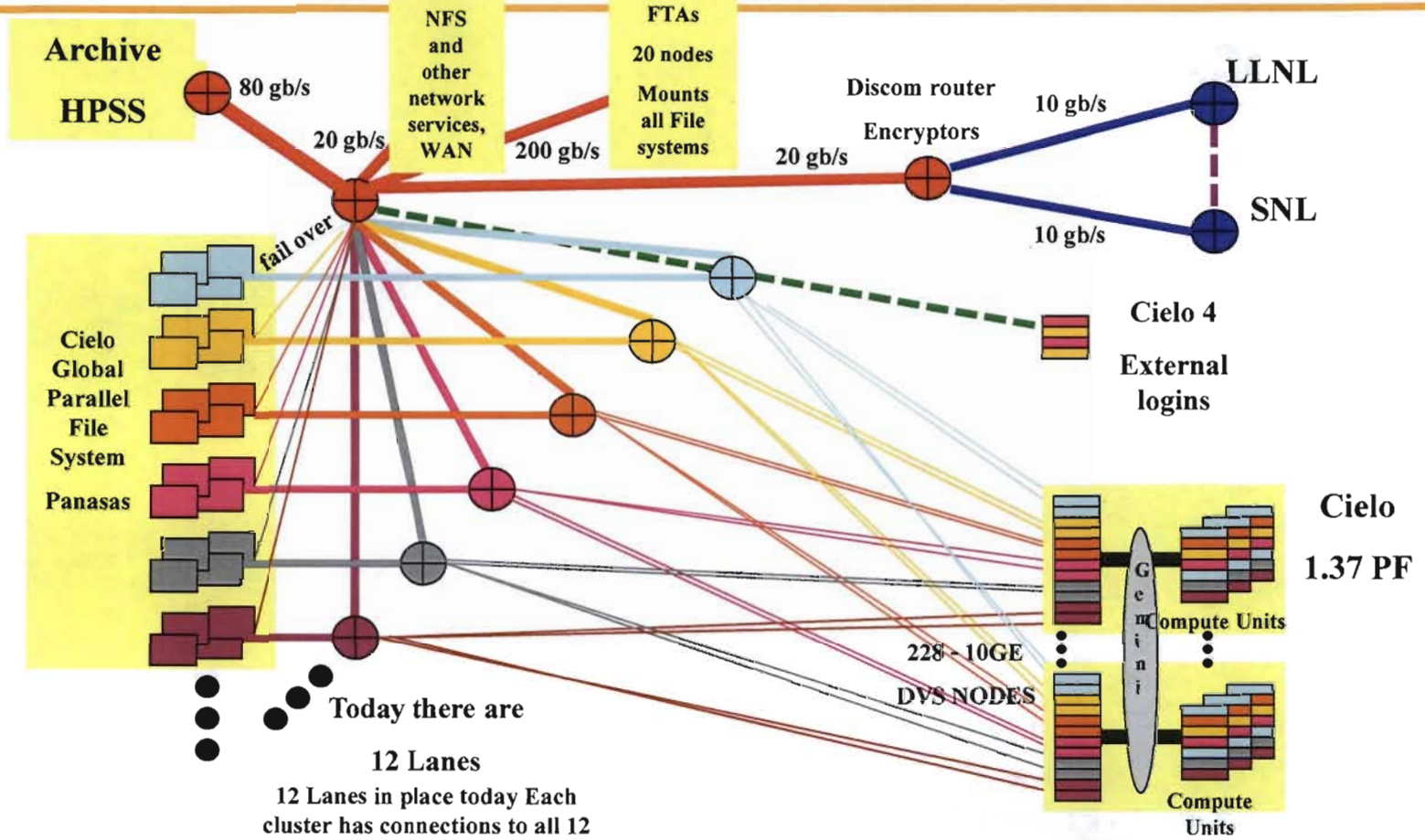
- AMD Magny-Cours Node
  - Dual-socket AMD 6136 Processors
  - 2 x 8 = 16 total cores
  - 2.4 GHz core frequency
  - 32 GB of 1333 DDR3 memory
    - 64 GB for Visualization Nodes
  - 153.6 peak DP GFLOPs
  - 85.3 peak GB/s memory BW
- Gemini High-Speed Interconnect
  - 3D Torus topology
  - Phase 1: 18x8x24
    - X bisection: > 4.38 TB/s
    - Y bisection: > 4.92 TB/s
    - Z bisection: > 3.92 TB/s
  - Phase 2: 16x12x24
    - X bisection: > 6.57 TB/s
    - Y bisection: > 4.38 TB/s
    - Z bisection: > 4.38 TB/s
  - Node Injection
    - > 6 GB/s/dir sustained BW
    - > 8 MMsgs/sec sustained



cielo

NNSA New Mexico Alliance for Computing at Extreme Scale

Integration into LANL's PaScalBB



Los Alamos NATIONAL LABORATORY EST. 1943

Sandia National Laboratories

cielo

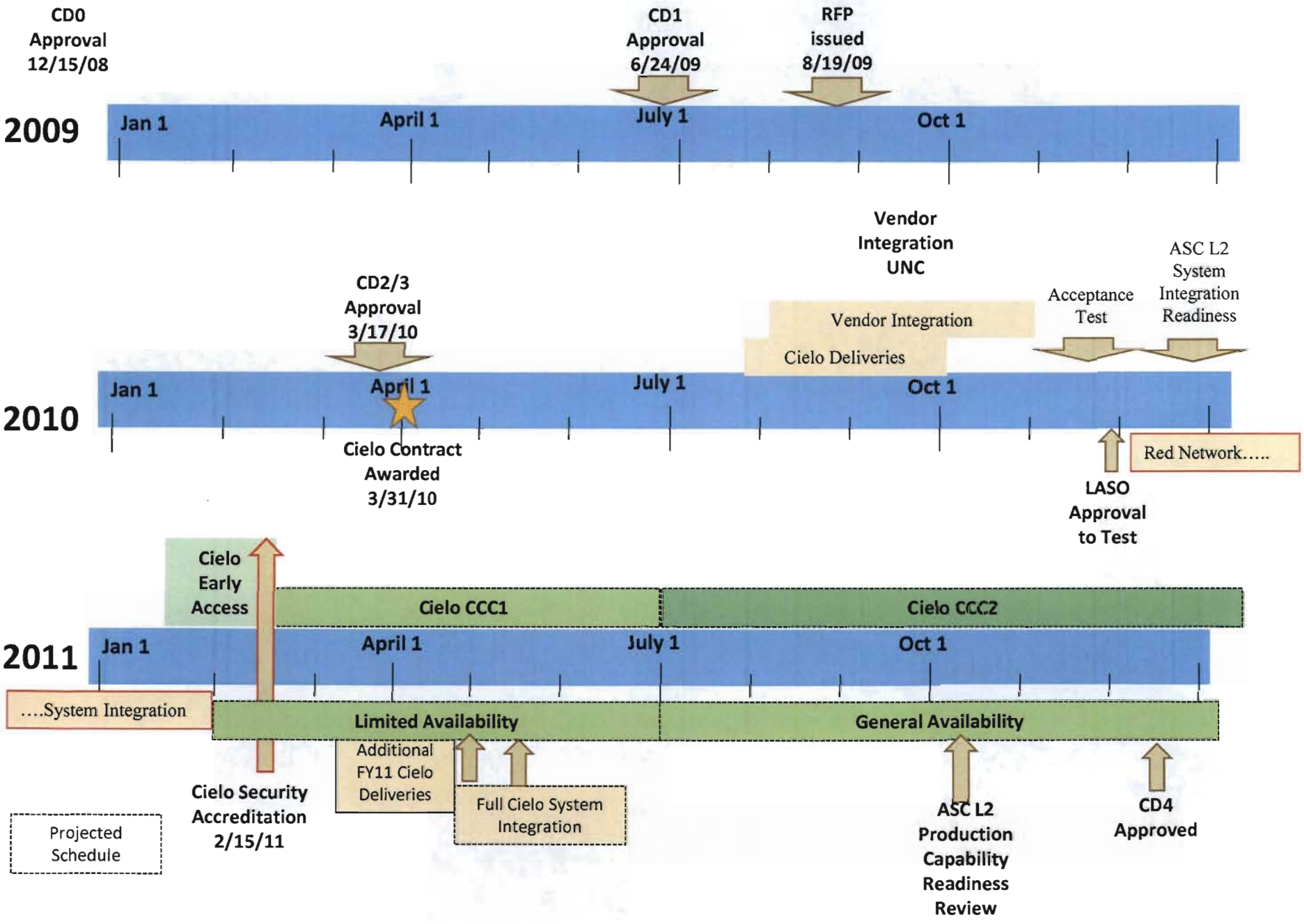
NNSA New Mexico Alliance for Computing at Extreme Scale

Cielo By Numbers

	Phase 1	Phase 2	Cielito
# of Cabinets	72	96	1
# of Service Nodes	208	272	14
# of Compute Nodes	6,704*	8,944*	68
# of Visualization Nodes	(376)	(376)	(4)
# of Compute Cores	107,264	143,104	1,088
Peak Memory BW	572 TB/s	763 TB/s	5.8 TB/s
Memory Capacity per Core	2 GB (4 GB)	2 GB (4 GB)	2 GB (4 GB)
Compute Memory Capacity	226.6 TB	298.2 TB	2.3 TB
Peak Compute FLOPS	1.03 PF	1.37 PF	10.4 TF
Sustained PFS BW	> 160 GB/s		TBD
System Power	~3 MW w/HPL	~4 MW w/HPL	
Full System Job MTBI	> 25 hours		
System MTBI	> 200 hours		

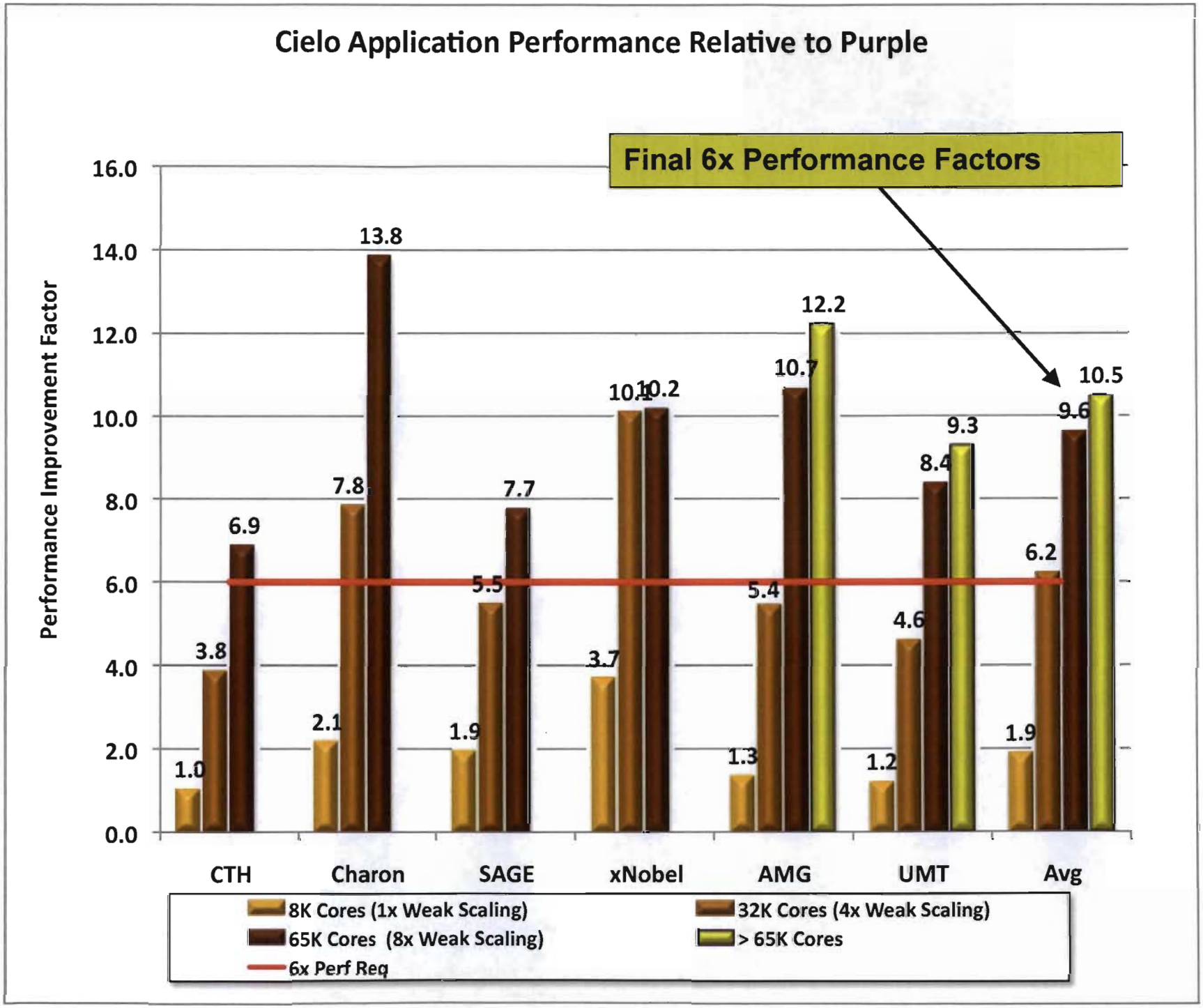
* Total compute nodes including Viz nodes and nodes allocated for other services

Cielo Platform Schedule Highlights 2009-2011



Cielo Highlights

- Integration and Acceptance – Dec 2010 (ASC Level 2)
  - Demonstrated 9.6x improvement in capability relative to Purple
- Capability Computing Campaign 1 (CCC1) – Feb 2011
  - Over 70 users: 3 early application campaigns plus porting and scaling of integrated weapons codes
- Upgrade on schedule: April 30th through May 20th
  - Grows Cielo from 1.03 to 1.37 PF (107,264 to 143,104 compute cores)
- CCC2 will begin in July
- Cielo Project being nominated for DOE Defense Program Award of Excellence
- Positive feedback from users
  - "… one of the biggest real data visualizations that ParaView has done anywhere in the world"
  - "… solve times are comparable to a similar problem having a factor of 100 fewer elements that I ran on the LLNL Purple machine"
  - "Going to Muzia and then to Cielo, there have been *NO* significant porting issues to work through in getting our ITS-CAD capability running"


Cielo Application Performance Relative to Purple
Final 6x Performance Factors
Performance Improvement Factor
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
1.0
3.8
6.9
2.1
7.8
13.8
1.9
5.5
7.7
3.7
10.1
10.2
1.3
5.4
10.7
12.2
1.2
4.6
8.4
9.3
1.9
6.2
9.6
10.5
CTH
Charon
SAGE
xNobel
AMG
UMT
Avg
8K Cores (1x Weak Scaling)
32K Cores (4x Weak Scaling)
65K Cores (8x Weak Scaling)
> 65K Cores
6x Perf Req

cielo

NNSA New Mexico Alliance for Computing at Extreme Scale

Cielo Usage and Operational Model

- The ASC Capability Computing Campaign (CCC) process will be used to allocate resources on Cielo.
- Cielo is operated in an environment very similar to the last capability system (Purple).
- The Cielo Usage Model is our "contract" with users for how to best use the machine.
  - The Usage Model will form baseline requirements for the 2011 Production Capability Readiness L2 Milestone.

cielo
NNSA New Mexico Alliance for Computing at Extreme Scale

Cielo Early Applications

- pF3D – Steve Langer, LLNL
  Successfully ran a Laser-Plasma simulation of a NIF experiment on 32K Cielo Processors

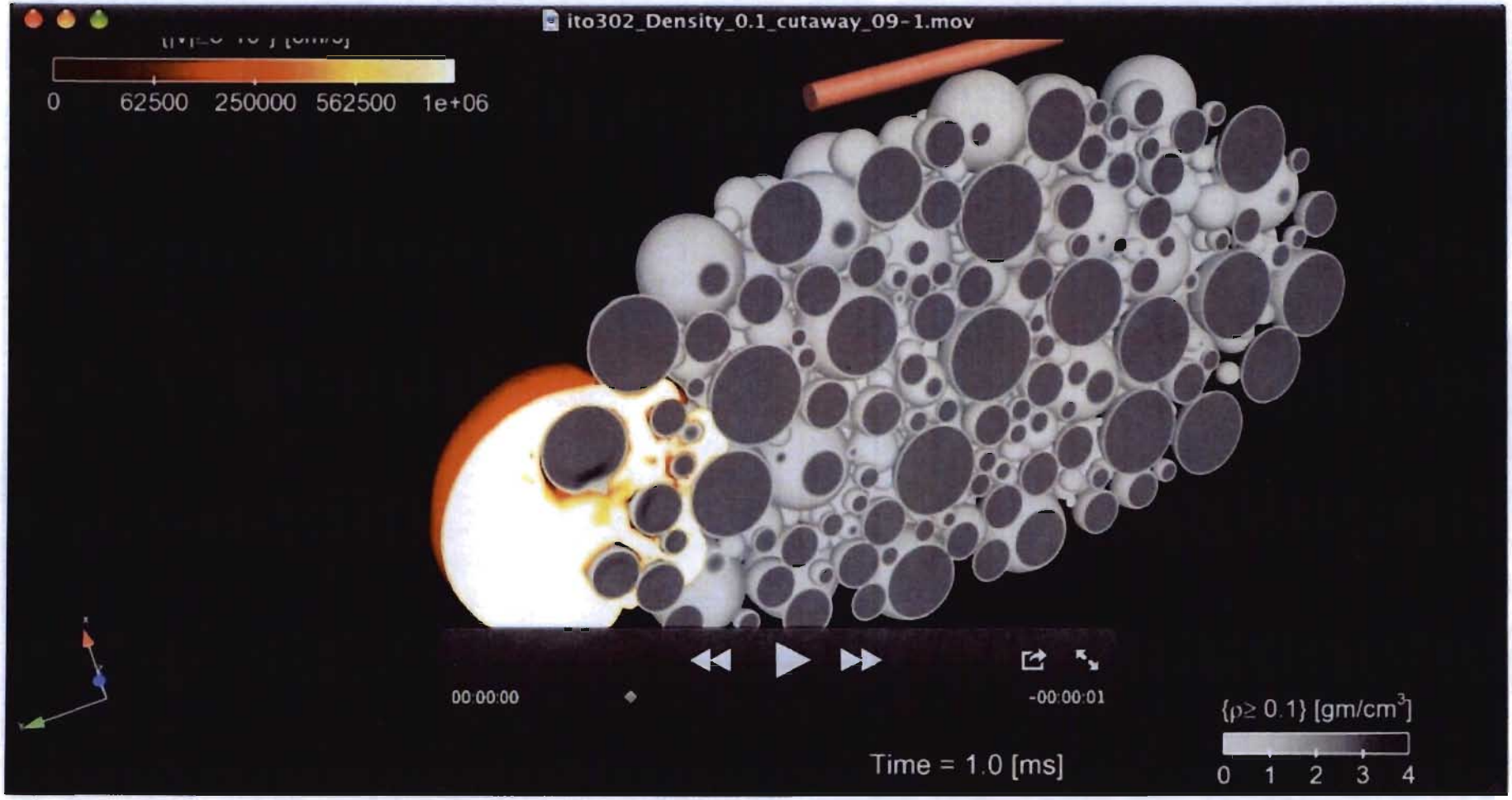
- CTH Simulations - Steve Attaway, Shivonne Haniff, Joel Stevenson, Jason Wilke, SNL
  Successfully ran a series of calculations using 32K PEs and has simulated ~900 microseconds of a 2,000 microsecond calculation

- Stockpile Calculations Results on Cielo – Robert Weaver, LANL
  Successfully completed five different full-system events running for stockpile calculations of interest. We have demonstrated using Cielo for high-fidelity runs using an advanced physics package that was accelerated for Roadrunner, but will run on Cielo using a higher processor count instead of accelerators.

cielo
NNSA New Mexico Alliance for Computing at Extreme Scale

3D RAGE simulations on the Cielo supercomputer to simulate a 1Mton surface explosion on Asteroid 25143 Itokawa

- **Significant public interest in this topic**
  - **Several Hollywood movies; interest from government; popular articles on this topic**
  - **We use the shape of the Asteroid Itokawa, which is not a near-Earth hazard, simply to have a nonspherical geometry**
- **Many methods of Potentially Hazardous Objects (PHOs) mitigation have been proposed:**
  - **Nuclear options: Explosive disruption; stand-off momentum/velocity transfer**
  - **Non-nuclear methods: gravity attractors; solar energy absorption (paint) etc.**
- **For this simulation we use the RAGE hydrocode in 3D with a 1 Mton energy source on the surface of the object**
- **Here we use realistic (nonspherical) shapes and explore a "rubble piles" composition, i.e., where the asteroid has experienced many disruptive interactions, recombined, and is composed of many smaller "rocks."**
- **This simulation currently has run only to 25 ms; interesting mitigation occurs after ~5 s for this explosion energy. This is still running on Cielo.**
- **Never tried before: Cielo is the first computer big enough to try this problem in 3D at Los Alamos.**


ito302_Density_0.1_cutaway_09-1.mov
0
62500
250000
562500
1e+06
00:00:00
-00:00:01
Time = 1.0 [ms]
{ρ≥ 0.1} [gm/cm3]
0
1
2
3
4

cielo
NNSA New Mexico Alliance for Computing at Extreme Scale

Cielo - Next Steps

- System upgrade from 1.03 PF/s to 1.37 PF/s (72 to 96 Cabinets)
  - Scheduled completion is May 20th, 2011
  - Acceptance will be in conjunction with continuing CCC1
- Continue CCC1
  - Additional applications being added
- Call for CCC2 (May 2011)
  - Similar to Purple CCC model
  - Job size categories for Cielo (based on governance model)
  - Dedicated application time for larger jobs
- Start CCC2 in July 2011
- Planning for ASC L2 Production Capability Milestone Review (Oct., 2011)
- *Original project schedules have been met*
- *ACES, Cray, and Panasas team interactions are working well*

QUESTIONS?