National Aeronautics and Space Administration

NASA Advanced Supercomputing Facility

Pleiades Overview

Bob Ciotti
Supercomputing Systems Lead/System Architect

OFA Sonoma 2009

Agenda

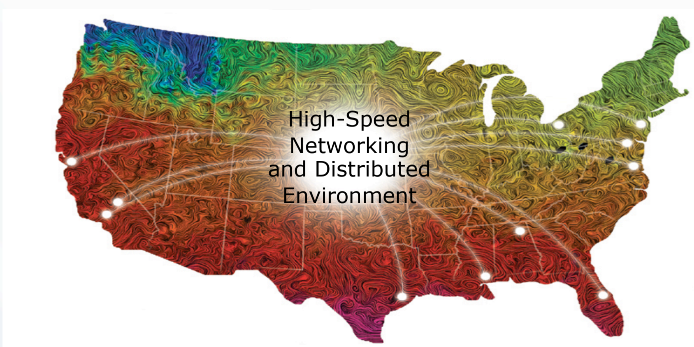
- The NAS Mission
- Columbia
- Pleiades
- Infiniband Directions

NAS Mission and Capabilities

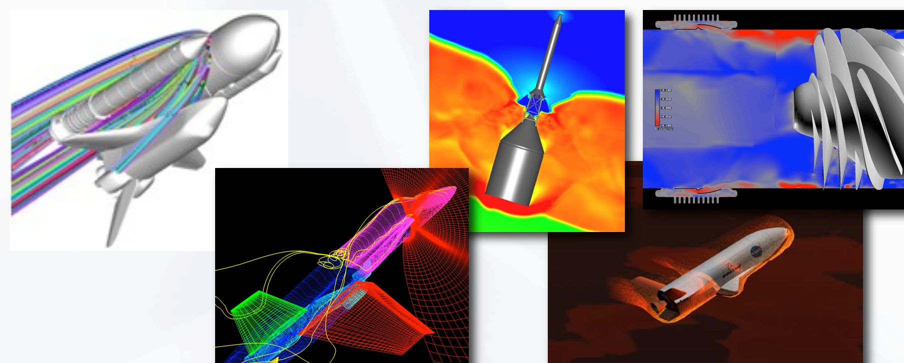
Mission: To develop and deliver the most productive high-end computing environment in the world, including supercomputing platforms, large-scale data storage, high-speed networking, application development and performance optimization, advanced data analysis and visualization, and high-fidelity modeling and simulation; enabling NASA to extend technology, expand knowledge, and explore the universe.

Supercomputing Operations

Supercomputing User Services



NASA-Wide Supercomputing Environment

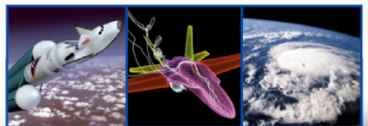


High-Fidelity Modeling and Simulation

NAS: Full Spiral Support for Advanced Computational Modeling and Simulation

NAS Mission: Develop and deliver the most productive high-end computing environment in the world, enabling NASA to extend technology, expand knowledge, and explore the universe.

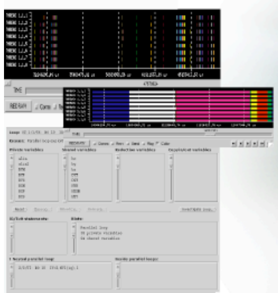
NASA Scientists and Engineers



Scientists and engineers set up computational problems, selecting the best-suited codes and resources to solve NASA's complex mission problems

Outcome: Dramatically enhanced understanding and insight, accelerated science and engineering, improved accuracy, and increased mission safety

Performance Optimization



Data Analysis and Visualization

Visualization experts apply advanced data analysis and rendering techniques to help scientists explore and understand large, complex data sets

Software experts utilize tools to parallelize and optimize codes, dramatically increasing simulation performance while decreasing turnaround time

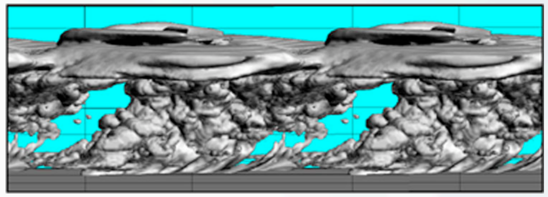
Supercomputers, Storage and Networks

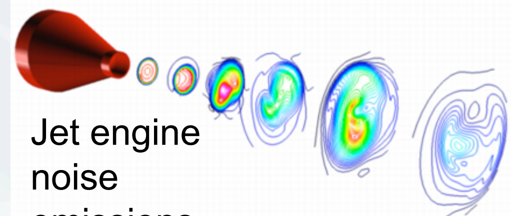
Supercomputing environment (hardware, software, network, and storage) is used to execute optimized codes to solve NASA's large computational problems

Supercomputing Support for NASA Missions

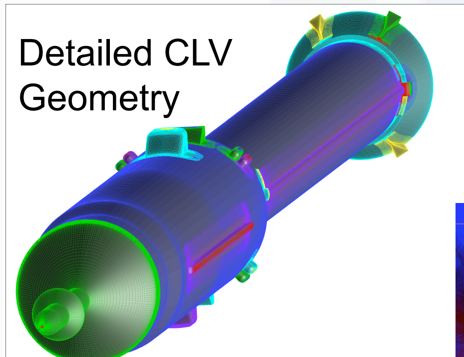
- ARMD
  - LaRC: Jet wake vortex simulations, to increase airport capacity and safety
  - GRC: Understanding jet noise simulations, to decrease airport noise
- ESMD
  - ARC: Launch pad flame trench simulations for Ares vehicle safety analysis
  - MSFC: Correlating wind tunnel tests and simulations of Ares I-X test vehicle
  - ARC/LaRC: High-fidelity CLV flight simulation with detailed protuberances
- SMD
  - Michigan State: Ultra-high-resolution solar surface convection simulation
  - GSFC: Gravity waves from the merger of orbiting, spinning black holes
- SOMD
  - JSC/ARC: Ultra-high-resolution Shuttle ascent analysis
- NESC
  - KSC/ARC: Initial analysis of SRB burn risk in Vehicle Assembly Building

Jet aircraft wake vortices



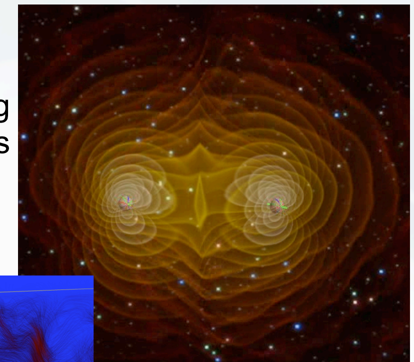


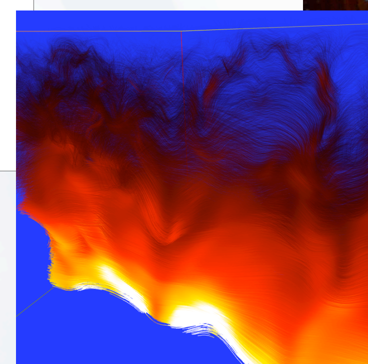
Jet engine noise emissions



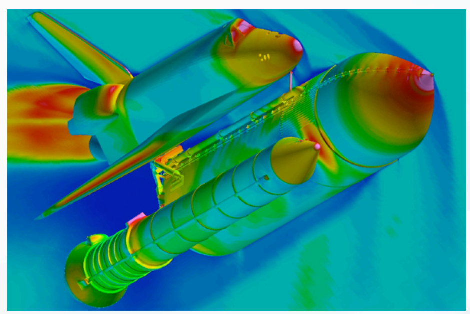
Detailed CLV Geometry

Orbiting, Spinning Black Holes



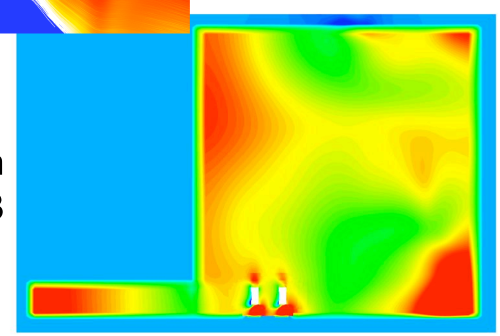


Solar surface convection

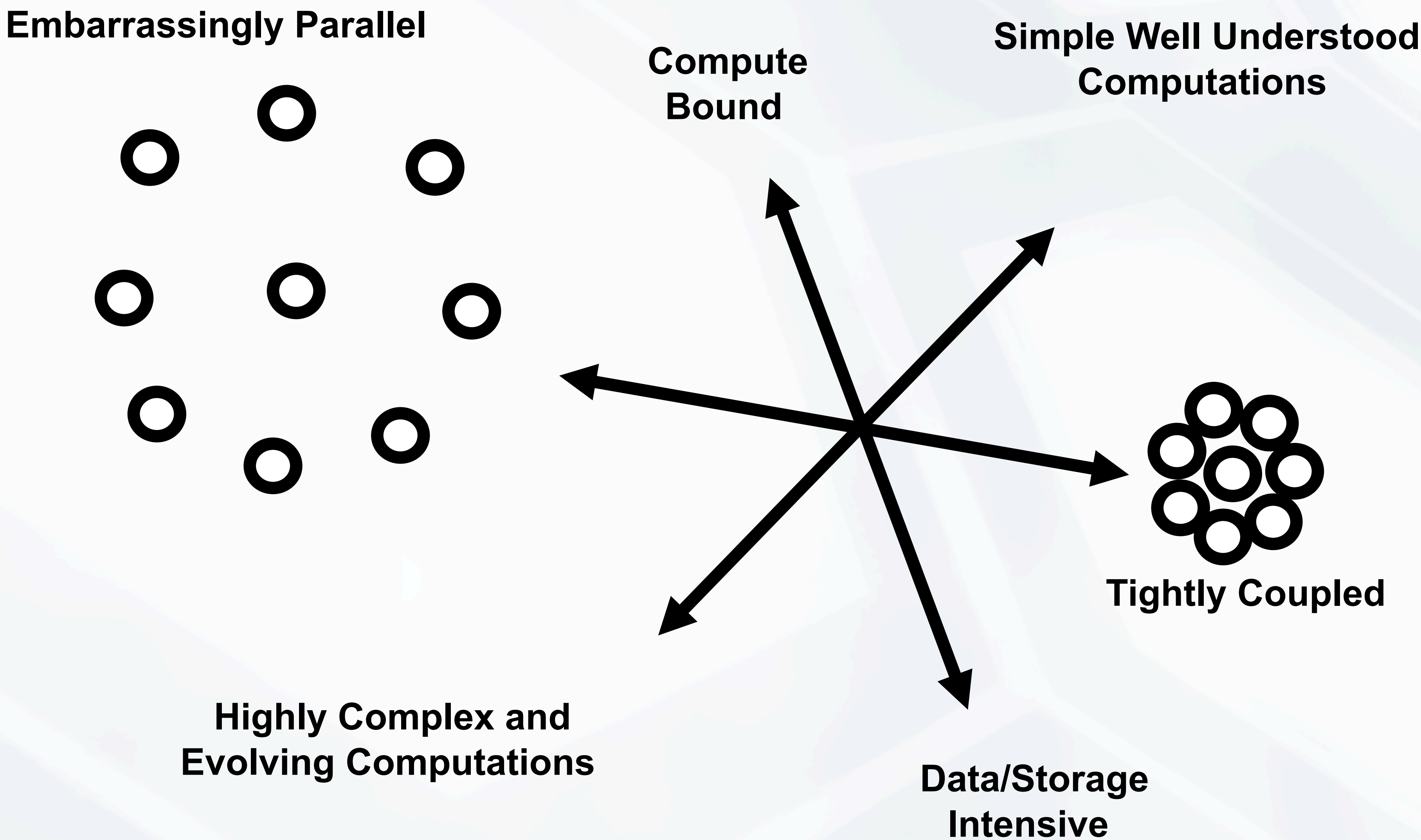


Shuttle Ascent Configuration

2-SRB Burn in VAB



NASA's Computational Landscape



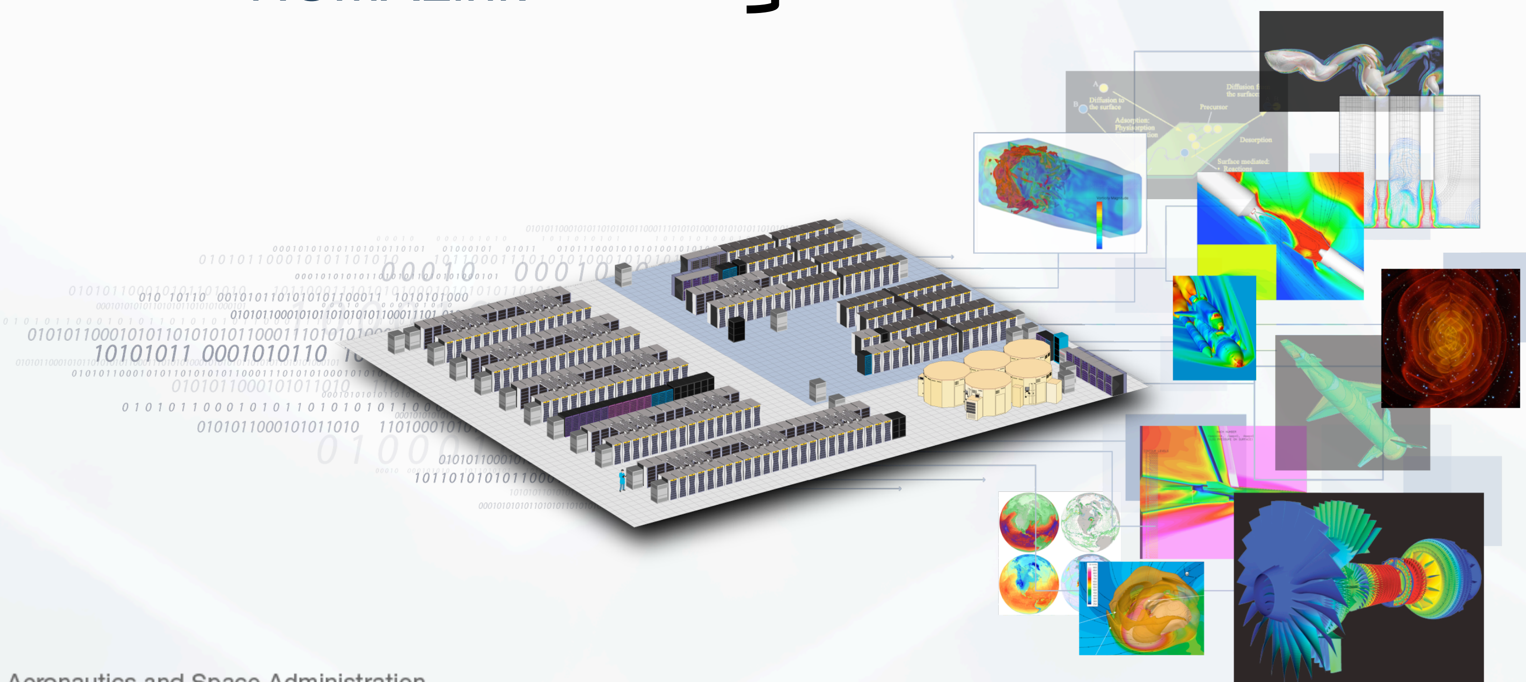
Columbia System - October 2004

Columbia in '04

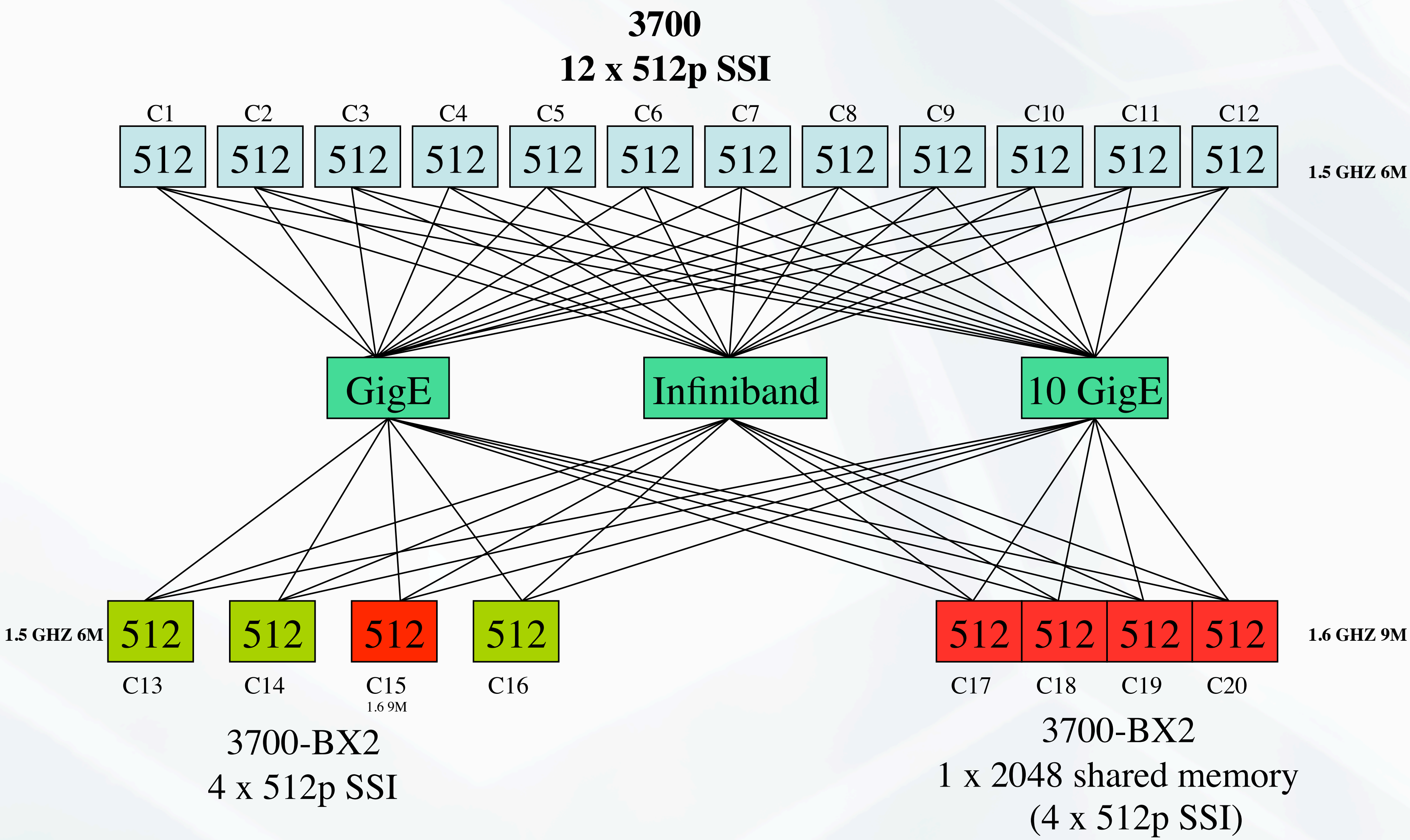
- Columbia '04
  - 12 SGI Altix 3700
  - 8 SGI Altix 3700 BX 2
    - 4 connected with NUMALink

62 TF

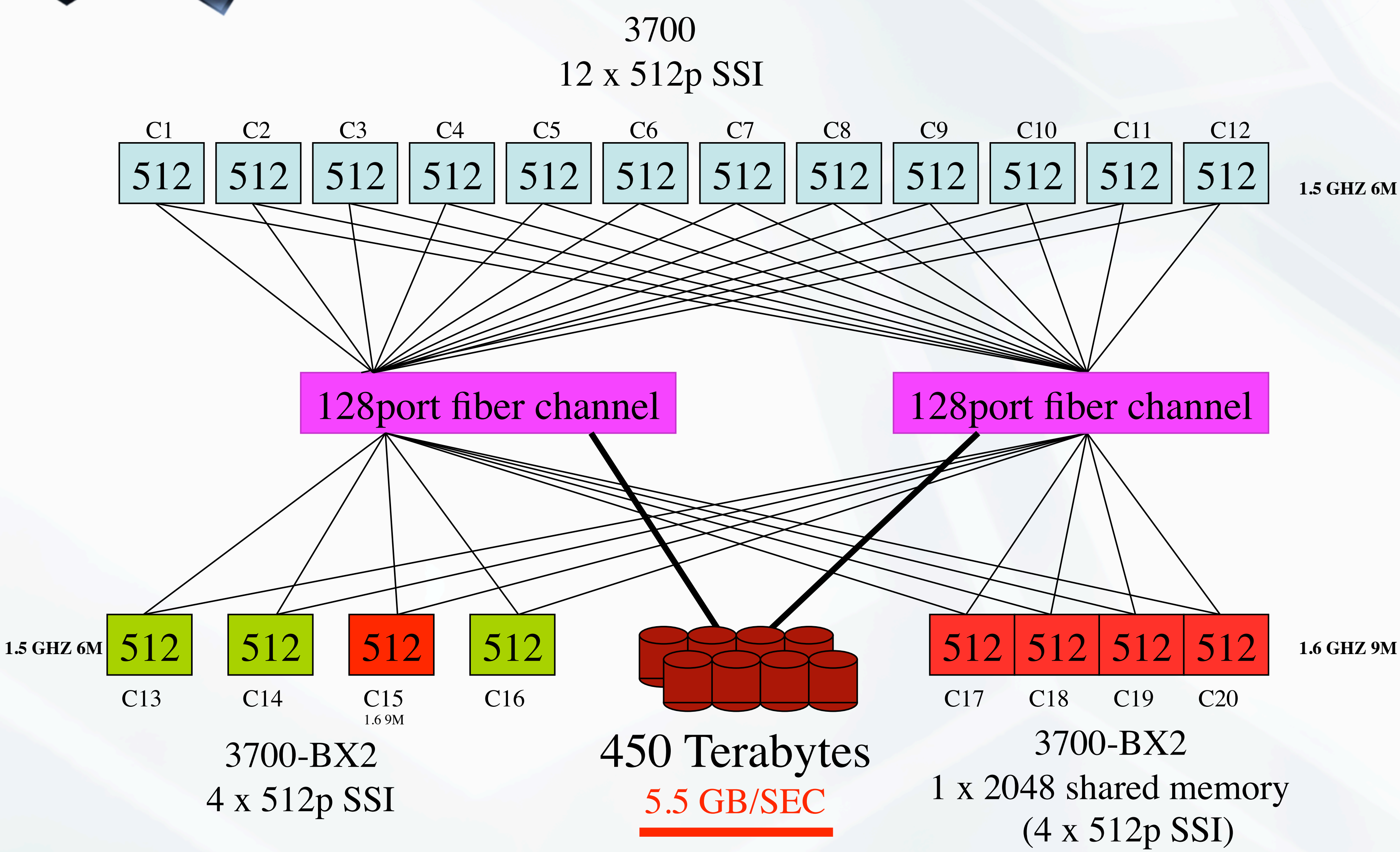
10160 SBUs



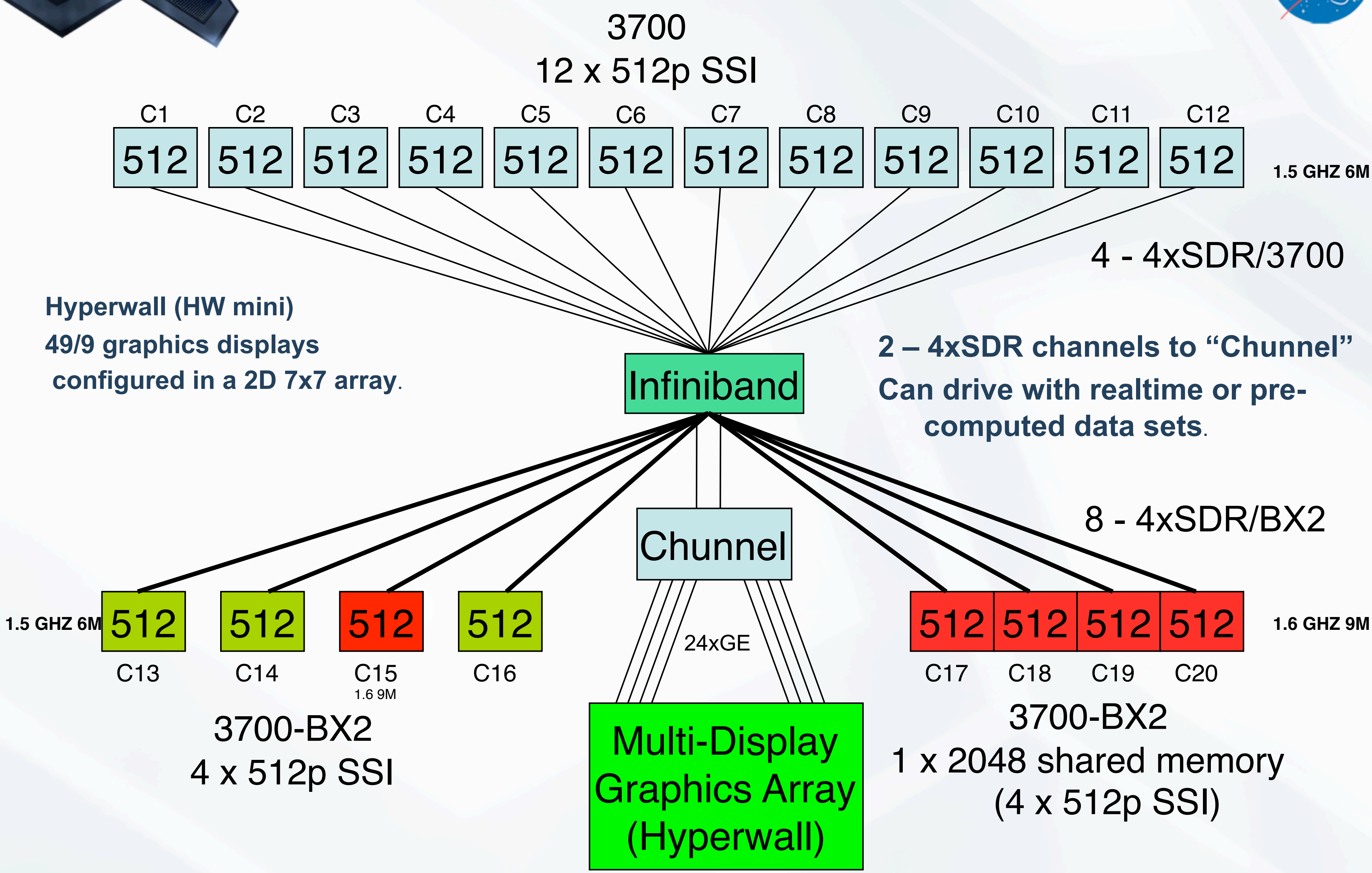
Columbia – Compute Configuration



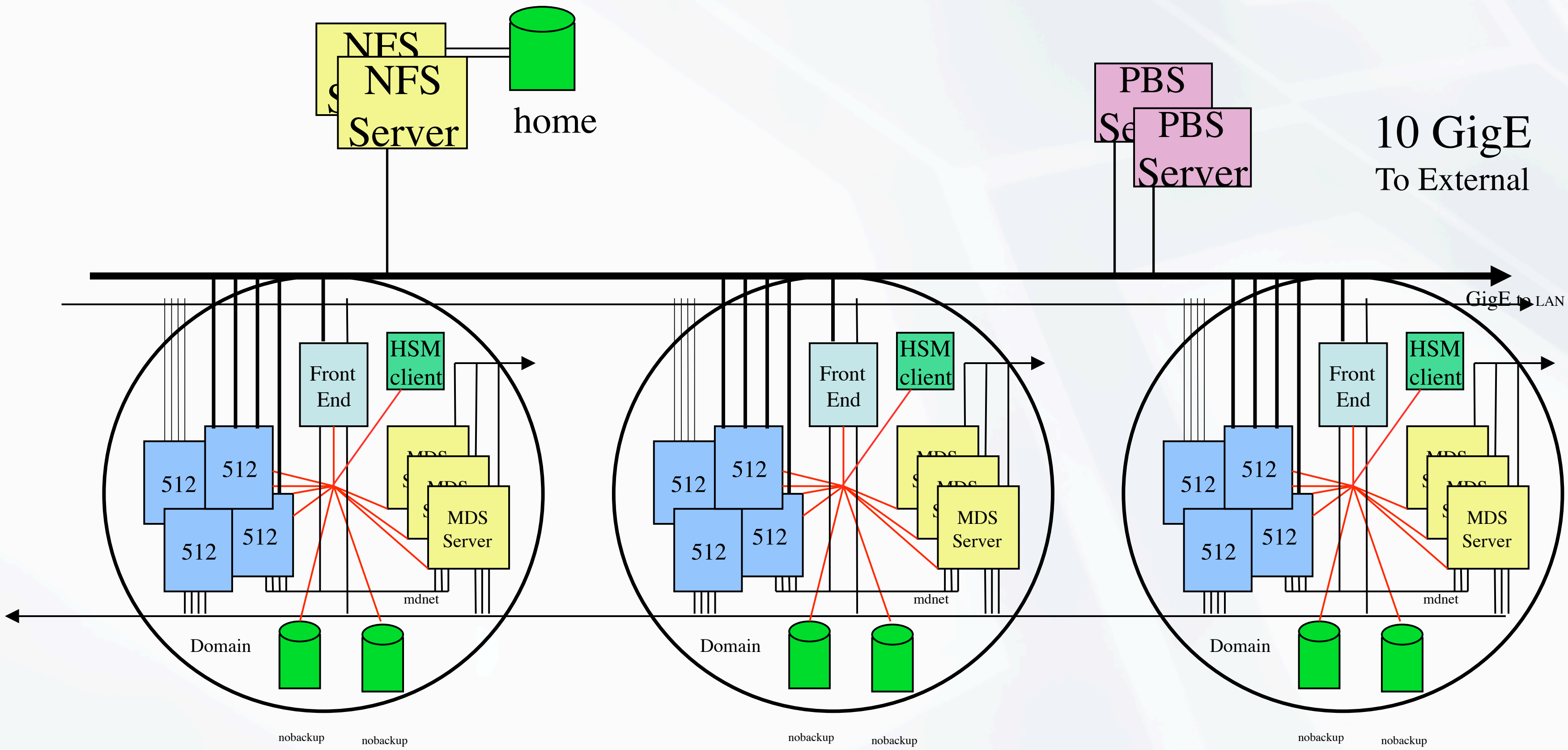
Columbia - SAN Fabric



Columbia – Graphics Configuration



Architecture Target - Phase III



- 3 "Domains"
- Domain relates to global filesystem span (CXFS)
- Systems can float between domains based on load/requirements
- Job scheduling is independent of filesystem domain – i.e. span domain
- MPI jobs can use infiniband between all systems
    - mostly to soak up the hard to schedule – last remaining cpus in a large SSI

NAS Early 2008

- Columbia (89TF) (14224 SBU)
  - 23-node SGI Altix Supercluster
    - 11 512c 3700s
    - 8 512c 3700BX2s
    - 1 512c 4700
    - 2 1024c 4700s
    - 1 2048c 4700
- RTJones (43TF) (2540 SBU)
  - 512 dual-socket blades
  - 1024 Quad-core Intel Xeon
- Schirra (4.8TF) (1016 SBU)
  - 40 IBM 9118-575 nodes
  - 320 dual-core IBM P5+

137 TF

17780 SBUs

RTJones Prototype Testbed

- 512 nodes
  - 4096 X5355 (Clovertown) Xeon cores @2.66ghz
  - Greengreek MCH
  - Infinihost III HCA
  - Served as a prototype test bed for Pleiades
  - 24-port Mellanox switch chip (MT47396)
  - Dual-Plane Hypercube
    - 2:1 oversubscription of node/switch:switch uplinks
      - 16 node:8 switch (8d with 1 port reserved I/O)
      - 2048 nodes/16k cores - max standard config
  - FibreChannel Back End disk subsystem
    - ~5gb peak bandwidth
    - 1 Lustre Filesystem
      - 1 - IS220 MDS
      - 2 - DDN 9550

Pleiades Base Configuration

- 5888 nodes
  - 47,104 X5472 (Harpertown) Xeon cores @3.0ghz
  - Seaburg MCH
  - Connect-X HCA
  - 24-port Mellanox switch chip (MT47396)
  - Dual-Plane Hypercube
    - 1:1.6 node/switch:switch uplinks
      - 8 node:13 switch (12d with 1 port reserved I/O)
      - 32k nodes/256k cores - max standard config
      - Less ports used for Express links
  - Infiniband Back End disk subsystem
    - ~30gb peak bandwidth
    - 3 Lustre filesystems
      - 1 - IS220 MDS
      - 2 - DDN 9900

Pleiades Final Configuration

- Merge of RTJones and Pleiades
  - 6400 nodes
  - 51,200 X5472/X5355 Xeon cores @3.0/2.66ghz
  - 24-port Mellanox switch chip (MT47396)
  - Dual-Plane Hypercube
  - Back End disk subsystem
    - ~35gb peak bandwidth
    - 4 Lustre filesystems

NAS September 2008

- Pleiades (608TF) (44934 SBU)
  - 6400 dual-socket blades
  - 12,800 quad-core Intel Xeon processors
  - 51,200 cores
- Columbia (86TF) (13716 SBU)
  - 22-node SGI Altix Supercluster
    - 10 512c 3700s
    - 8 512c 3700BX2s
    - 1 512c 4700
    - 2 1024c 4700s
    - 1 2048c 4700
- Schirra (4.8TF) (1016 SBU)
  - 40 IBM 9118-575 nodes
  - 320 dual-core IBM processors

699 TF

59666 SBUs

Pleiades – Infiniband Fabric

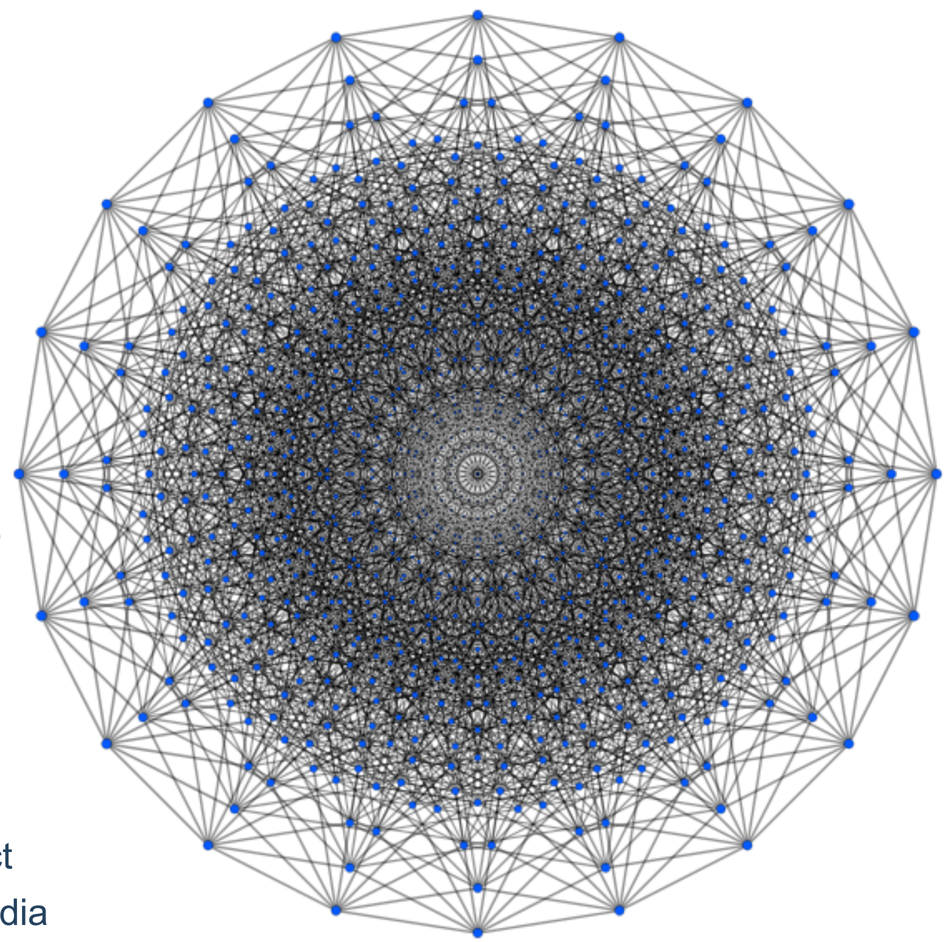
- SGI Design
  - IMHO -> “signpost system” commodity mesh
- DDR
  - 22 miles of cables
    - combined optical/passive copper
  - 6400 CAs (node endpoints) x 2 active ports/endpoint
  - 1600 24port DDR switches
- Dual Independent Planes
  - 2 complete subnets (12,800+ active end point ports)
  - 10d Partial Hypercube Topology
- Outboard Switches
- Failover
  - Currently Only MPI
  - Looking at Lustre/NFS via APM
- SAN
  - NFS and Lustre both serve >6,400 clients, NFS for /home, Lustre for /nobackup
  - NFS – both IPoIB and RDMA
  - Lustre – DDN/OST and LSI/MDT on separate IB subnet

Infiniband – Subnet Discovery

Partially Populated 10d Hypercube

- Subnet manager algorithm
  - Minimum Hop Count
  - Break Ties by Port Number

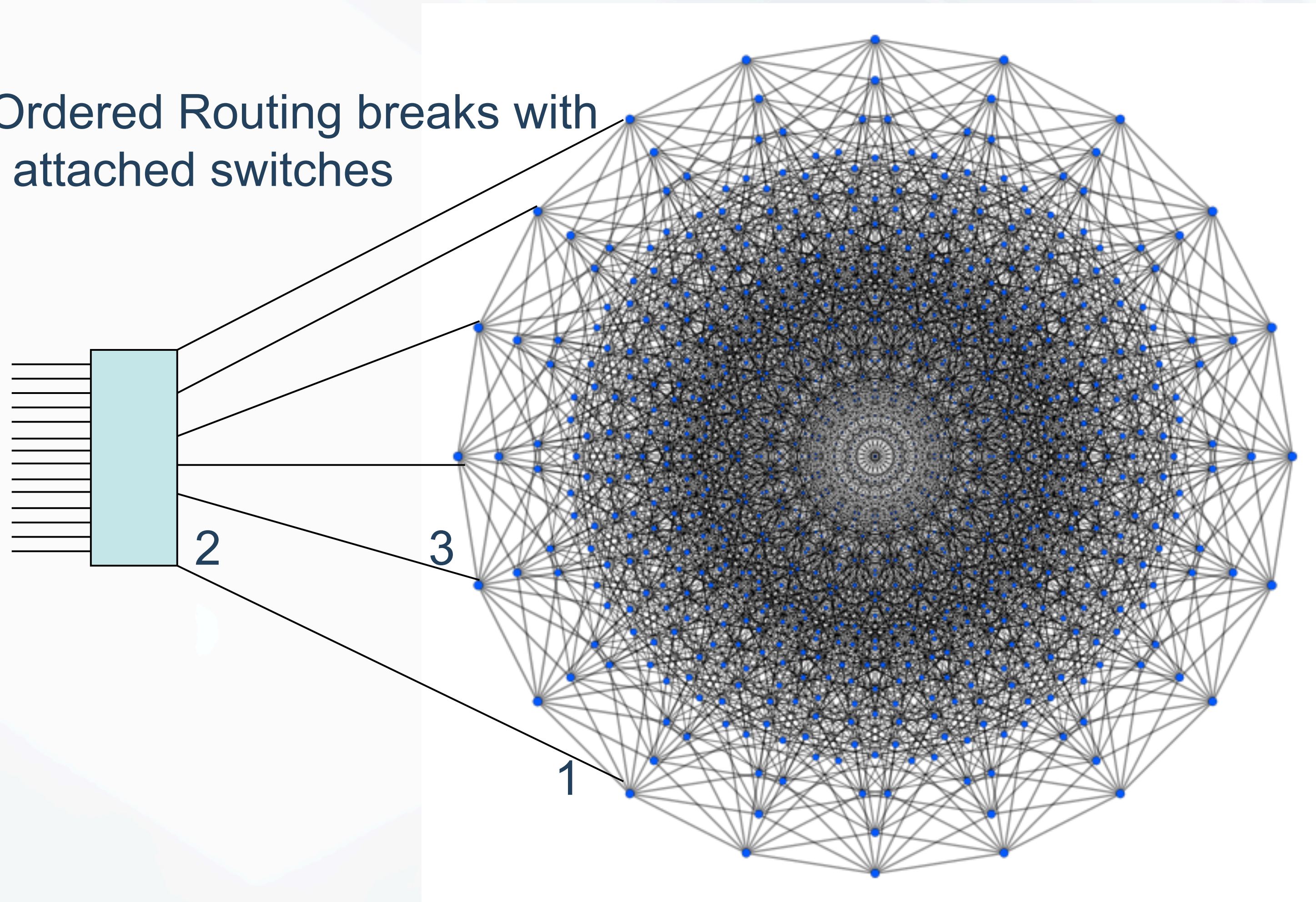
→Dimension Ordered Routing (DOR)



(Orthographic demidekeract
by Claudio Rocchini, wikipedia
Copyright GNU http://en.wikipedia.org/wiki/GNU_Free_Documentation_License
Creative Commons 3.0 http://creativecommons.org/licenses/by/3.0/)

Infiniband – Subnet Discovery

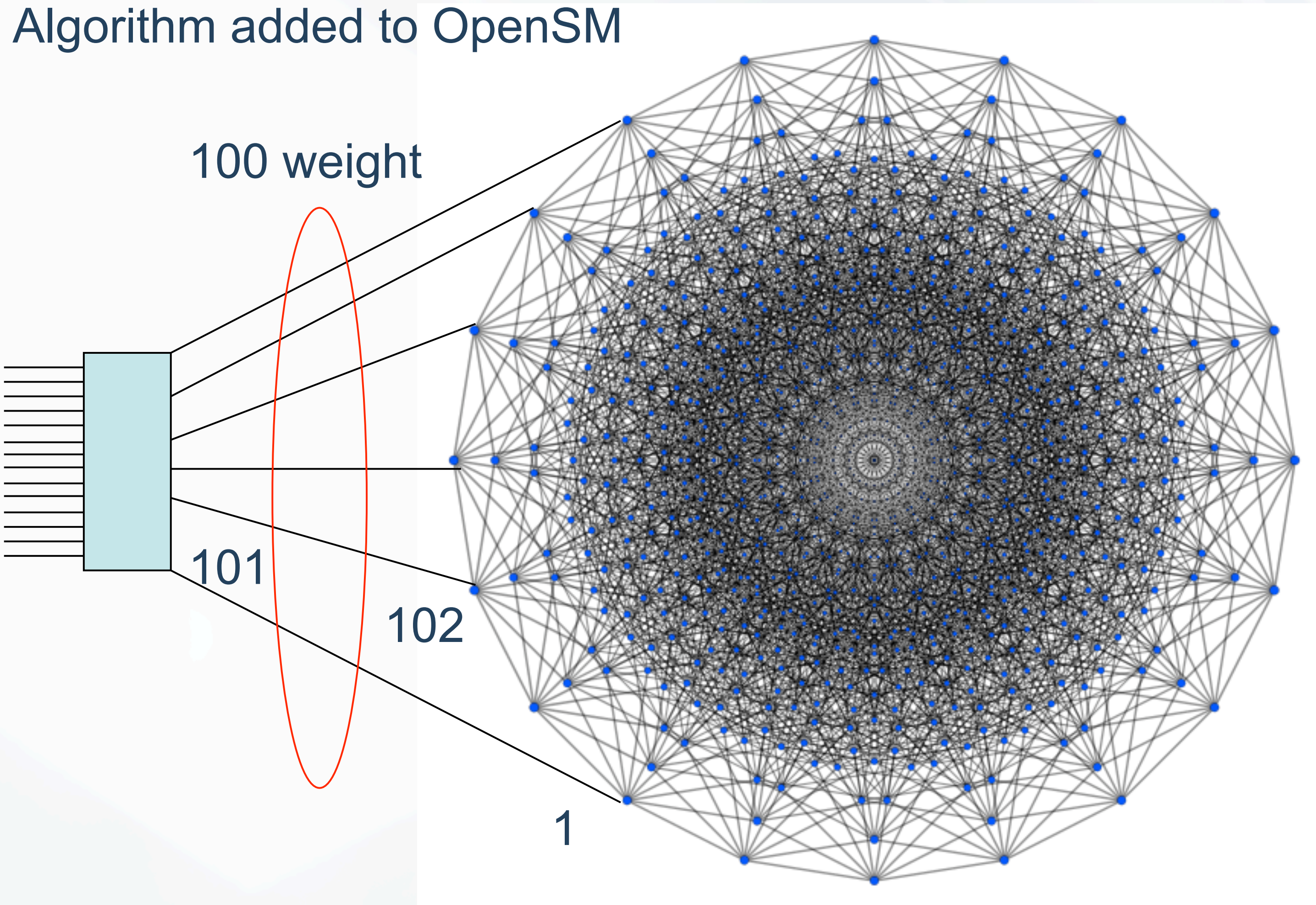
Dimension Ordered Routing breaks with externally attached switches



Orthographic demidekeract
by Claudio Rocchini, wikipedia

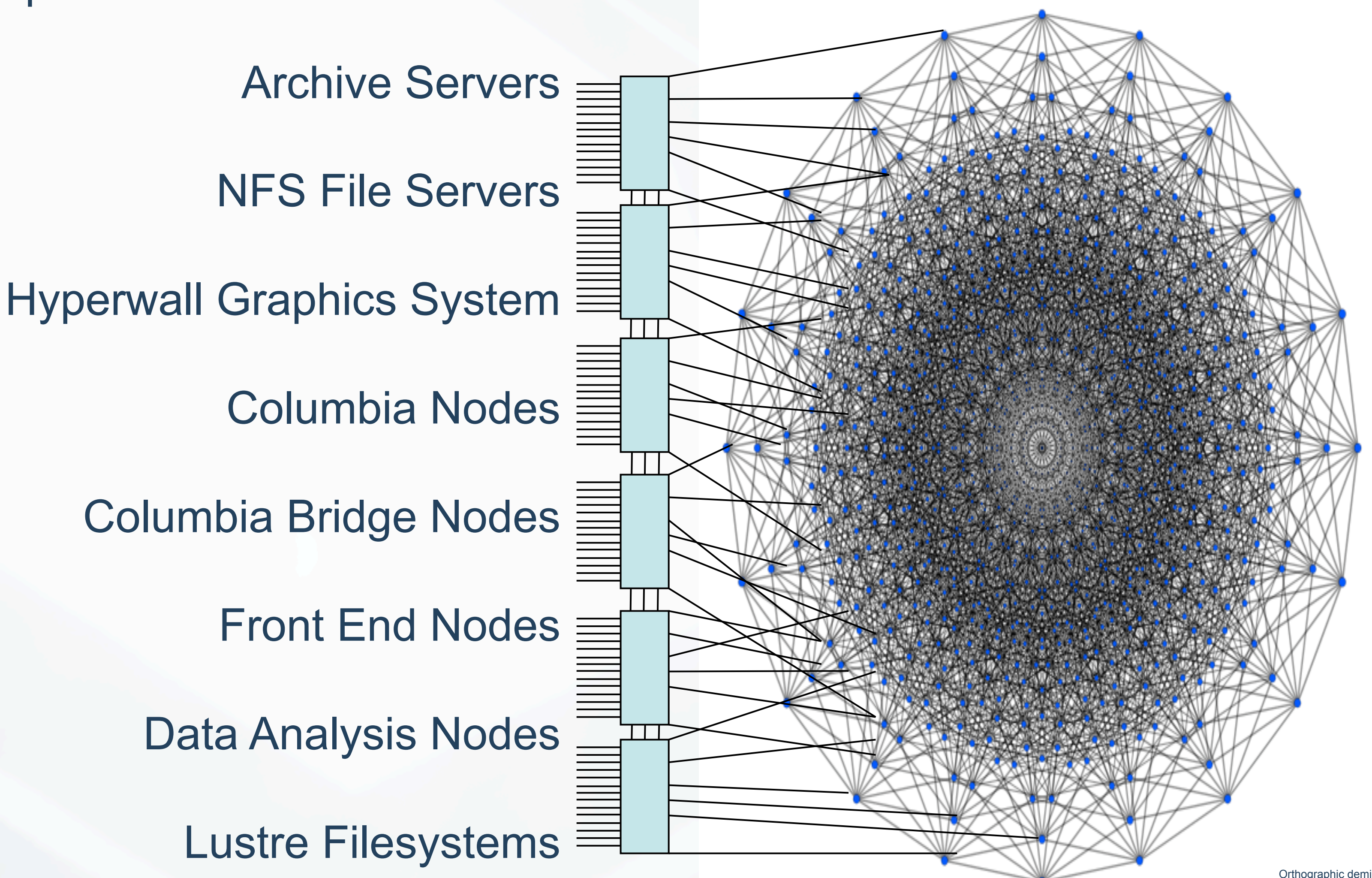
Infiniband – Subnet Discovery

- Weighting Algorithm added to OpenSM

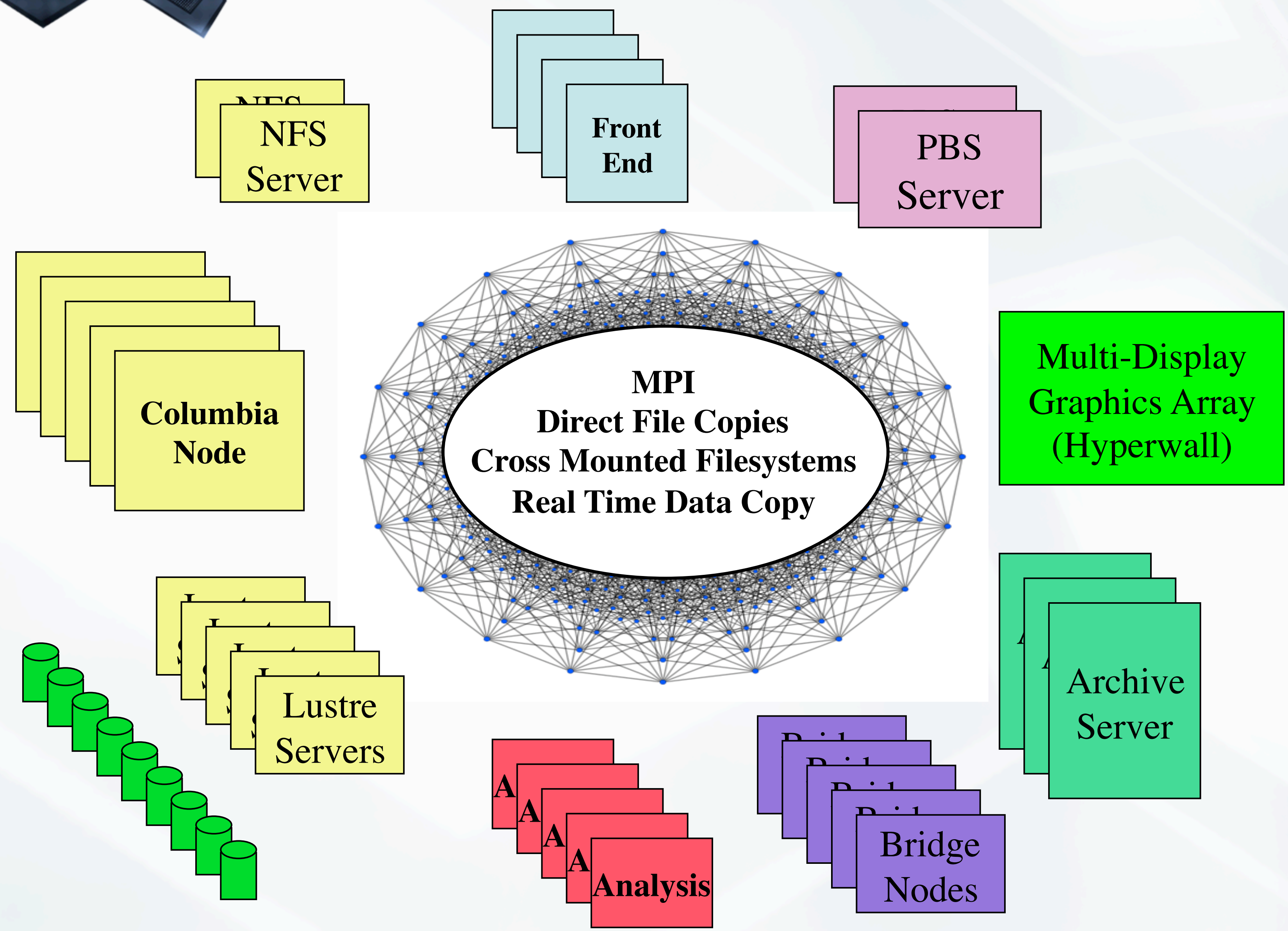


Infiniband Subnet LAN

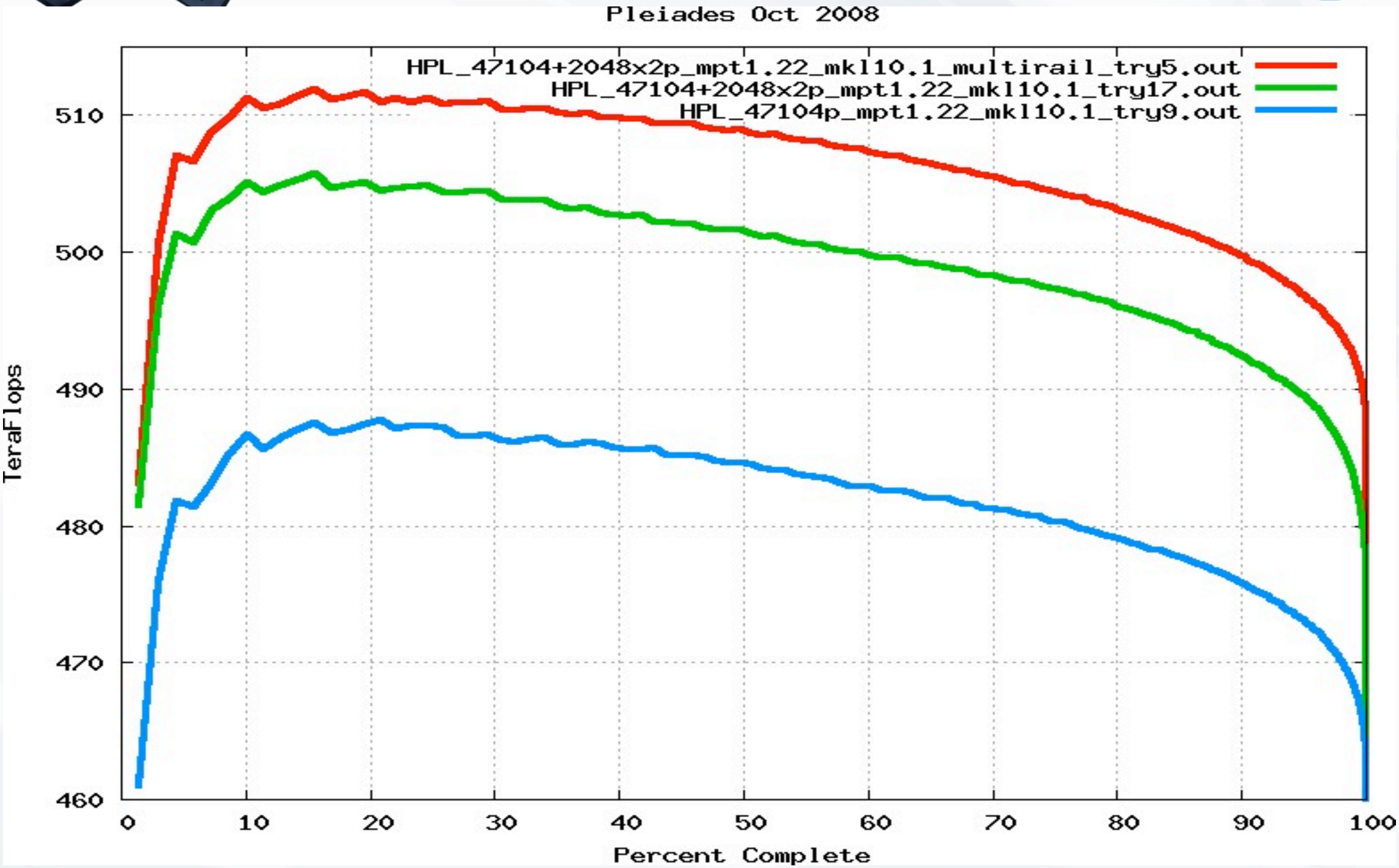
LAN Implemented with out board IB switches



Architecture Target



LINPACK

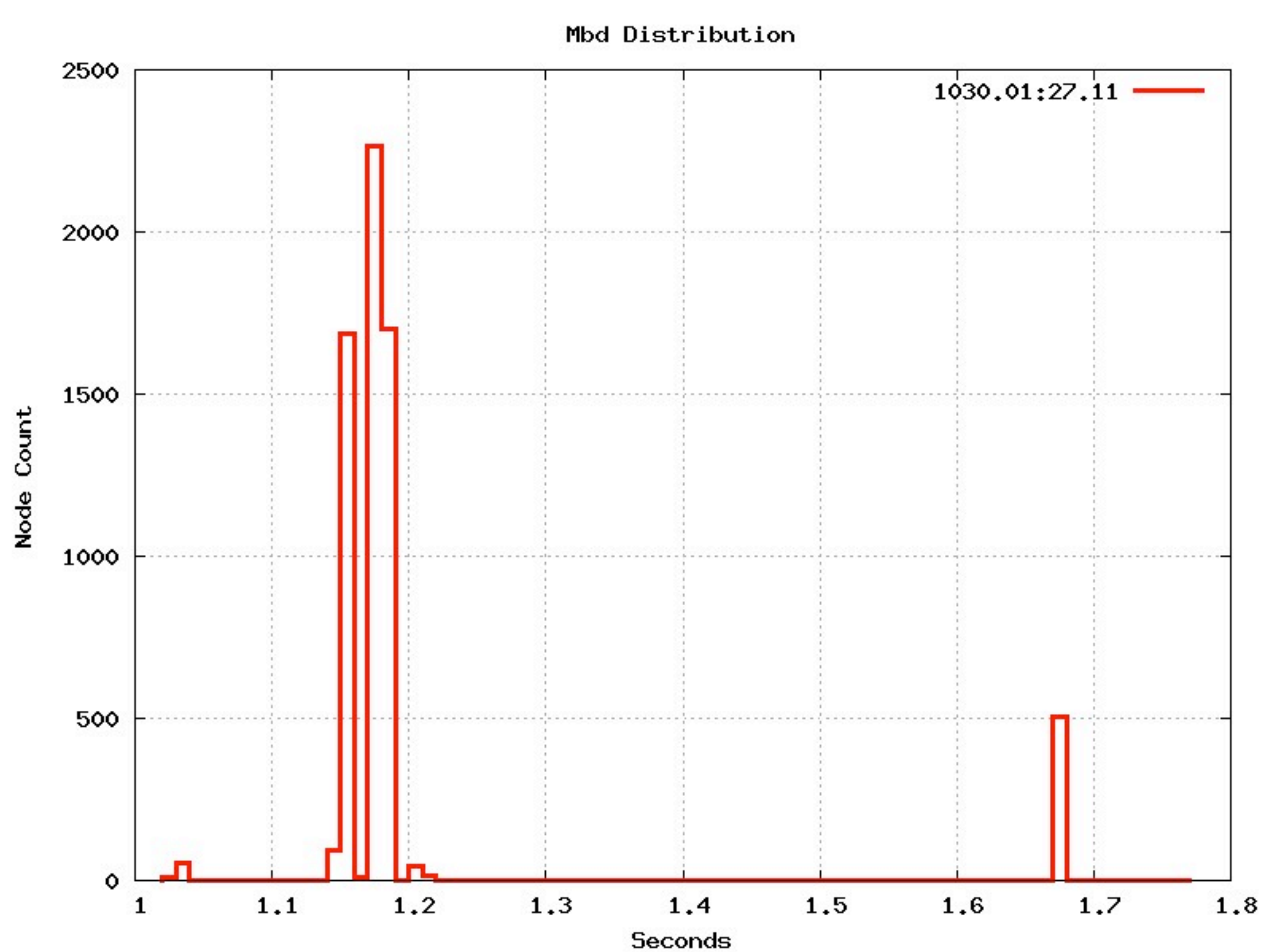


Pleiades Oct 2008
HPL_47104+2048x2p_mpt1.22_mkl10.1_multirail_try5.out
HPL_47104+2048x2p_mpt1.22_mkl10.1_try17.out
HPL_47104p_mpt1.22_mkl10.1_try9.out
TeraFlops
460
470
480
490
500
510
0
10
20
30
40
50
60
70
80
90
100
Percent Complete

System Confidence Tests

- SW Tools
  - Tree version of pdsh
    - Hits 6400 nodes ~3 seconds
- HW scrub
  - Mbd – CPU/Memory Diagnostic
  - Dasw.err – Dump IB Error Counters
  - Box – link load check (Lat/BW)
  - Ix – Intersect Bad Hardware List

System Confidence Tests
MBD – Node Diagnostic



System Confidence Tests
BOX– Link Diagnostic

- BOX
  - IB analog of MBD
  - Measures latency and bandwidth between CA’s
  - Slow runs can indicate problem areas
    - » Congestion
    - » Bad HW not logging errors

System Confidence Tests

Dump Switch Errors

```
Thu Oct 30 19:00:21 PDT 2008
/u/ciotti/bin/scripts/dasw.err -lp1
last lid/guid  map on Thu Oct 30 17:29:26 PDT 2008 r1i0n0 1
last counter reset at Thu Oct 30 17:37:12 PDT 2008
last counter  read at Thu Oct 30 17:44:29 PDT 2008

Checking 368 sw0 switches
Checking 368 sw3 switches
Checking 32 rtjones switches
         .     .     .       .      .                SwPort:   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20   21  22  23  24
         .     .     .       .      .                   sw0:  n2  n3   x  n0  ha  n1  n5  n4  d1  n7   x  n6  d2  d3  d4  d5  d6  d7  d8  d9  d10 d11 d12 rlc
         .     .     .       .      .                   sw3: n13 n12   x n15  ha n14 n10 n11  d1  n8   x  n9  d2  d3  d4  d5  d6  d7  d8  d9  d10 d11 d12 rlc
 r50i0n15: SwLid 12894 ib0-sw3 port-1            RcvErrors:   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .    1   .   .   .
 r50i0n15: SwLid 12894 ib0-sw3 port-1          XmtDiscards:   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   . 2045   .   .   .
 r50i0n15: SwLid 12894 ib0-sw3 port-1         LinkRecovers:   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   26   .   .   .
 r58i0n15: SwLid 13280 ib0-sw3 port-1 RcvRemotePhysErrors:   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   1   .   .    .   .   .   .
r122i0n15: SwLid 13064 ib0-sw3 port-1            RcvErrors:   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .    1   .   .   .
r122i0n15: SwLid 13064 ib0-sw3 port-1          XmtDiscards:   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .  274   .   .   .
r122i0n15: SwLid 13064 ib0-sw3 port-1         LinkRecovers:   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   26   .   .   .
         .     .     .       .      .                   rtj:   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20   21  22  23  24
         .     .     .       .      .                   rtj:  n2 n11  n3  n9  n1 n10  n5 n12  n4 n14  n6 n13  d1 n15  n7  d2  d3  d4  d5  n8   n0  d6  d7 rlc
p4fe1 ~/IB.LOGS/1030 >
```

- Code queries all active counters on the system
- Returns all non-zero values ~10 seconds (too slow)

System Confidence Tests
IX – intersect bad hardware

- Ix – Intersect Bad Hardware List
  - MPI kicks out bad CA pairs
  - Ibtracert between CA’s
  - Creates a list of switches/switch ports/CA’s
  - Combine with dasw.err output
  - intersect!

System Management
Continuous Operations

- Goal – never take the whole system down.
  - Software version conflicts
  - Rolling updates of various components
    - NFS, Lustre servers
    - NFS, Lustre clients (compute nodes)
    - Various admin, front ends, bridge nodes

- Dynamically Split fabric
  - Developed capability to split system for testing
  - Simply uses ibportstate to down a port
    - e.g. blocking port 21 splits the system into two pieces, one 64 racks and the other 36 racks

OFED Score Card

- What is OFED doing right
  - It works at scale – and I'm surprised
    - Power of Open Source – May the Source be with you
  - We have ridden through several switch and port failures
- What is OFED doing wrong
  - All-to-all ping
  - Lack of parallelism
  - Maybe too much reliance on the way IP works
    - E.g. ARP cache, DNS, multicast
  - Error/Performance monitoring
  - Security? (users with hostile intent)
- OpenSM/SA
  - Quite reliable
  - Path discovery (bug?, performance)
  - Might consider credit loop checking (DOR)/work around

Architecture Directions

- Storage network design change – this is the future
  - Cost effective/high performance
  - Fully consolidated onto infiniband (done)
    - (OK – we do limited bridging to GE/10GE)
    - Preferred path is IB

- No computer is an island
  - No computer is “A” supercomputer
  - Continuous operation 24/7/365
  - Old way of thinking about computers/networks
    - 1 port/node
    - → switch/node integration (2^1 -> 2^N)
  - Need for Fat Tree switches?

Future Directions

- Topologically aware scheduling
- Job aware routing
- Round trip path equivalence
- Express links/LMC
- Multi-Plane latency minimization
- Name Service
- ARP Configuration/Design
- Path Queries
- Performance/Error monitoring
- Cycle Prevention