Are we there yet?:
10Gb Ethernet for HPC

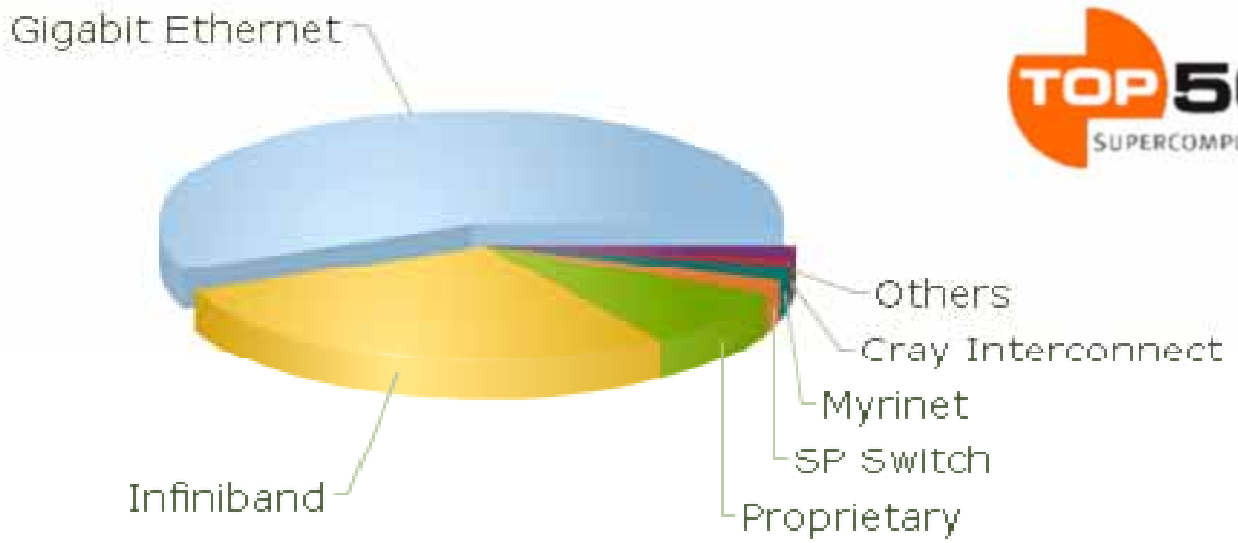
Name: Dan Tuchler
Title: VP Strategy and Product Management – BLADE Network Technologies
Date: January 28, 2009

BLADE
NETWORK TECHNOLOGIES

This is the year of 10 Gigabit Ethernet in HPC !!

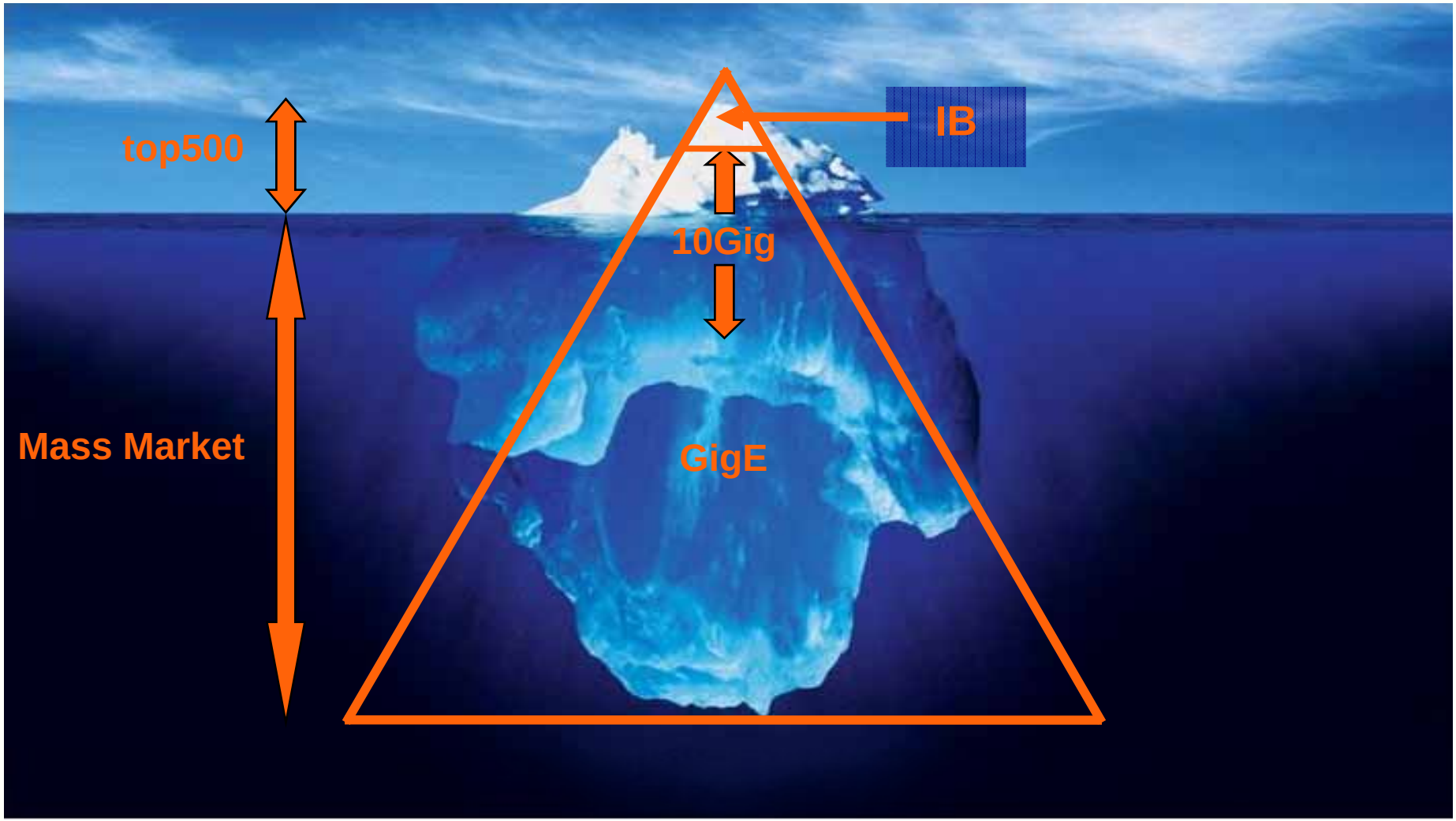
2009?
2010?
2006?

HPC Interconnect Landscape



Interconnect	Count	Share
Gigabit Ethernet	282	56.4%
10 Gig Ethernet	0	0.0%
Infiniband	141	28.2%
Myrinet	10	2.0%
Others	67	13.4%

The bigger picture for HPC



HPC Forecast: Strong Growth Over Next Five Years ($ Millions)

	2007	2012	CAGR
Supercomputer	$2,682	$3,512	5.5%
Technical Divisional	$1,610	$3,092	13.9%
Technical Departmental	$3,384	$5,763	11.2%
Technical Workgroup	$2,400	$3,193	5.9%
Total	$10,076	$15,617	9.2%

Source: IDC, 2008

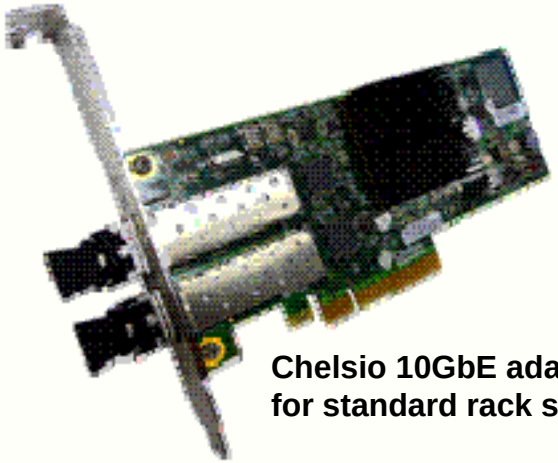
10 Gig E issues

What's holding up adoption?

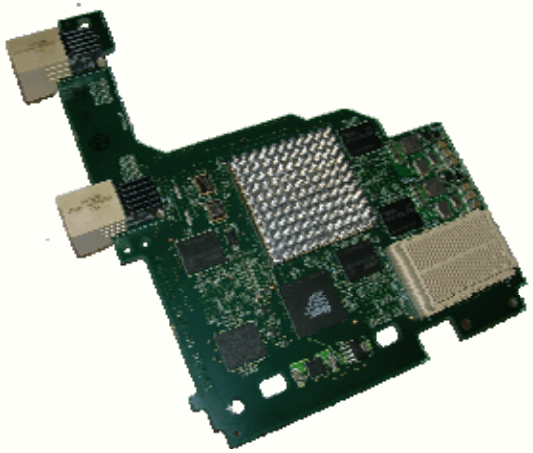
- 10 Gig NICs
- Price of Switches
- Switch Scaling
- PHY Confusion
- Proof of Performance

10 Gig NICs

- Prices are dropping fast
- Major server vendors are including 10 Gig Ethernet as standard server feature (LOM)
- Several NIC vendors proving mature and stable for HPC



Chelsio 10GbE adapter for standard rack servers

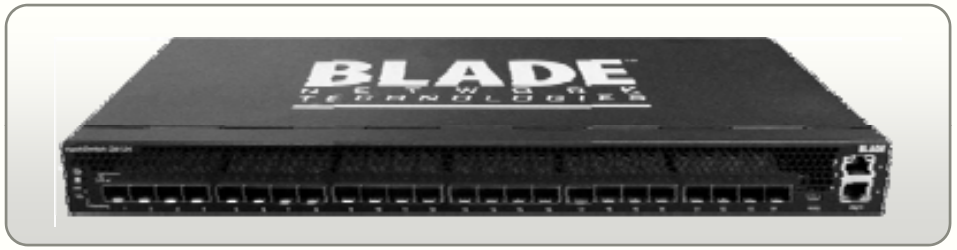


Chelsio 10GbE adapter for IBM BladeCenter-H servers

Price of 10 Gig Ethernet Switches

- Switch ports used to cost more than servers!
- 10 Gig E switches now list for <$500 / port

BLADE's Nortel 10G
Blade Switch



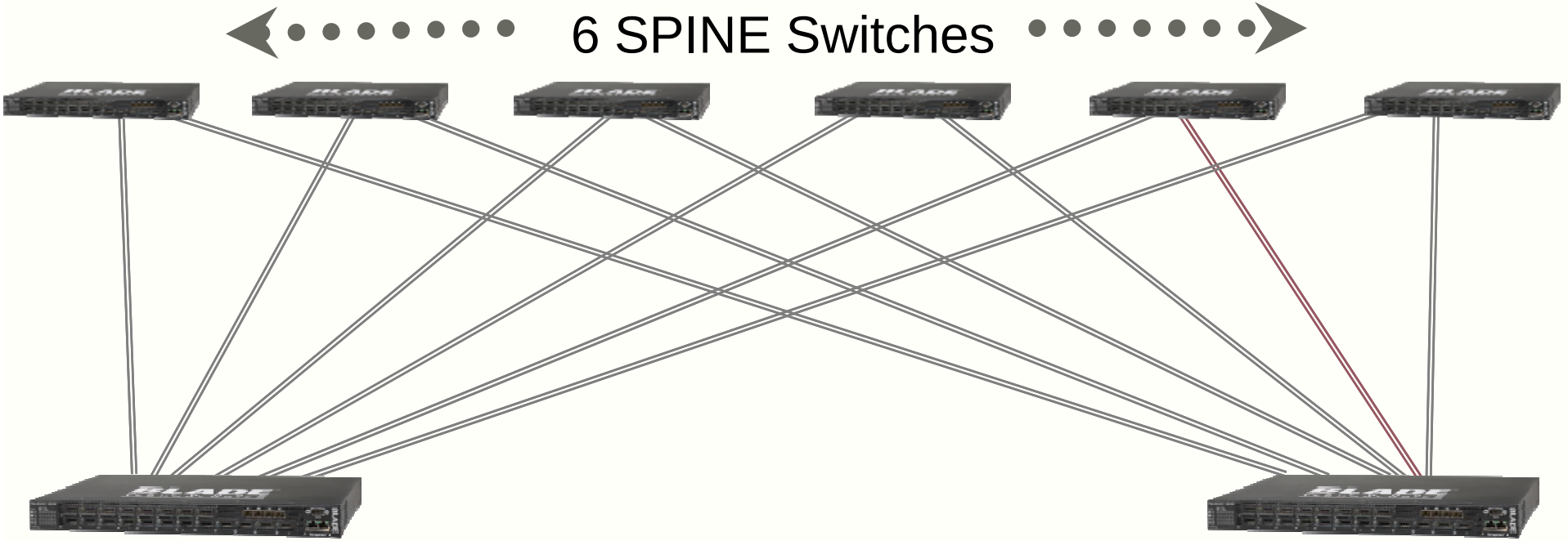
RackSwitch G8124
10 Gig SFP+ Switch

RackSwitch G8100
10 Gig CX4 Switch

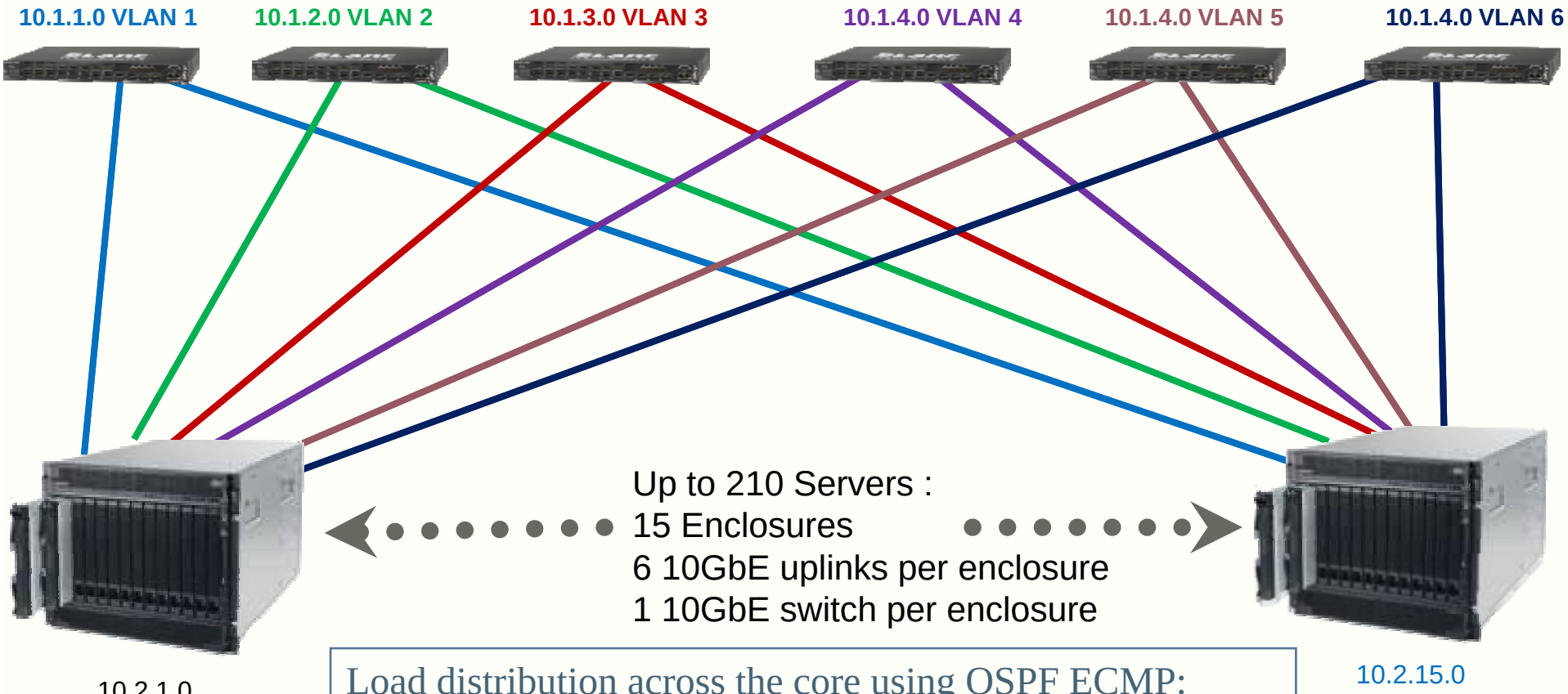
Switch Scaling

Typical CLOS Topology - 144 10GbE Ports

- 2-tier design scales to 288 ports



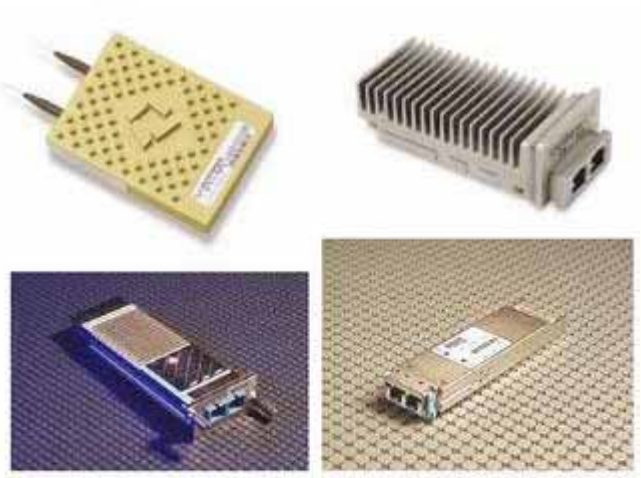
HPC Topology – up to 208 10GbE Servers
IBM BladeCenter Design



Load distribution across the core using OSPF ECMP:
Separate IP subnet for each enclosure
Separate VLAN and IP subnet for each RackSwitch
Full network path redundancy – not just link level
No Spanning Tree
2.3 to 1 oversubscription

PHY Confusion

- Optical standard interfaces for 10 Gig E:
  - Fixed optics
  - XENPAK
  - X2
  - XFP
  - SFP+
- 10GBase-T (i.e. Cat5, RJ45)
- CX4



Users have been unwilling to bet on a survivor!

New developments:

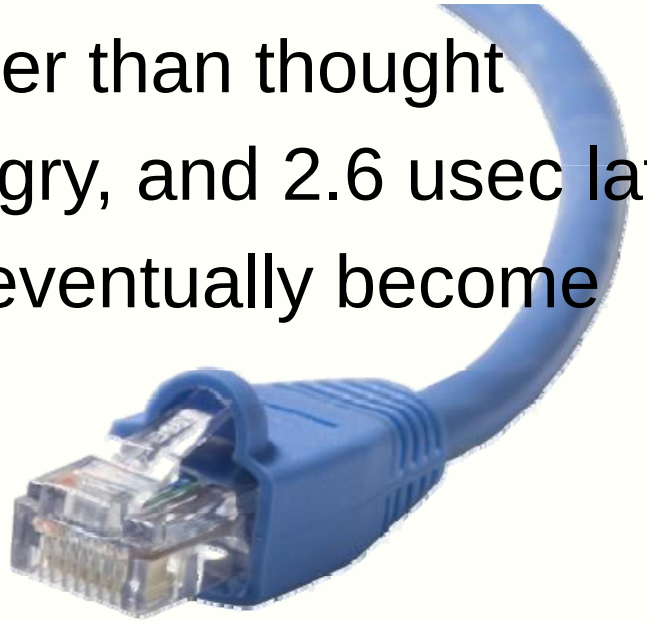
- SFP+ Direct Attach Cables
  - Passive cables with SFP+ ends
  - Low cost - $40 – $50
  - High density – same as RJ45

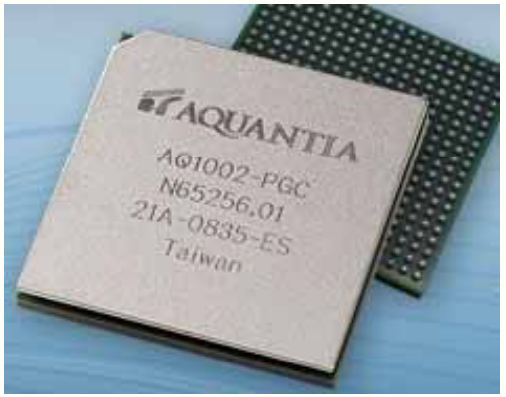
SFP+ Copper (Twinax) cable

CX4 – in use today

10GBase-T

- The problem was harder than thought
- Expensive, power-hungry, and 2.6 usec latency
- But – 10GBase-T will eventually become widespread



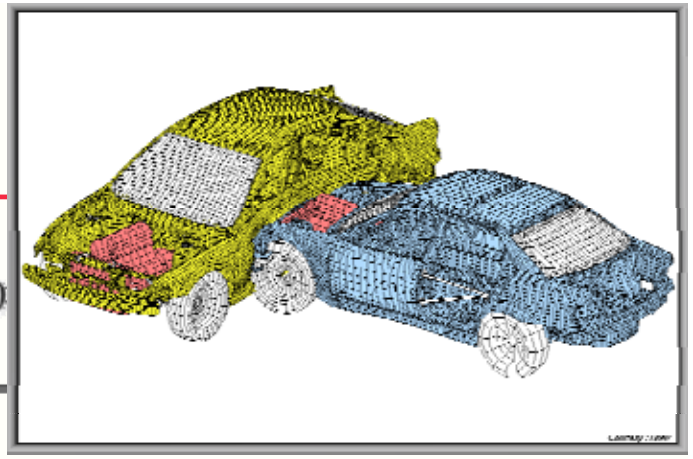


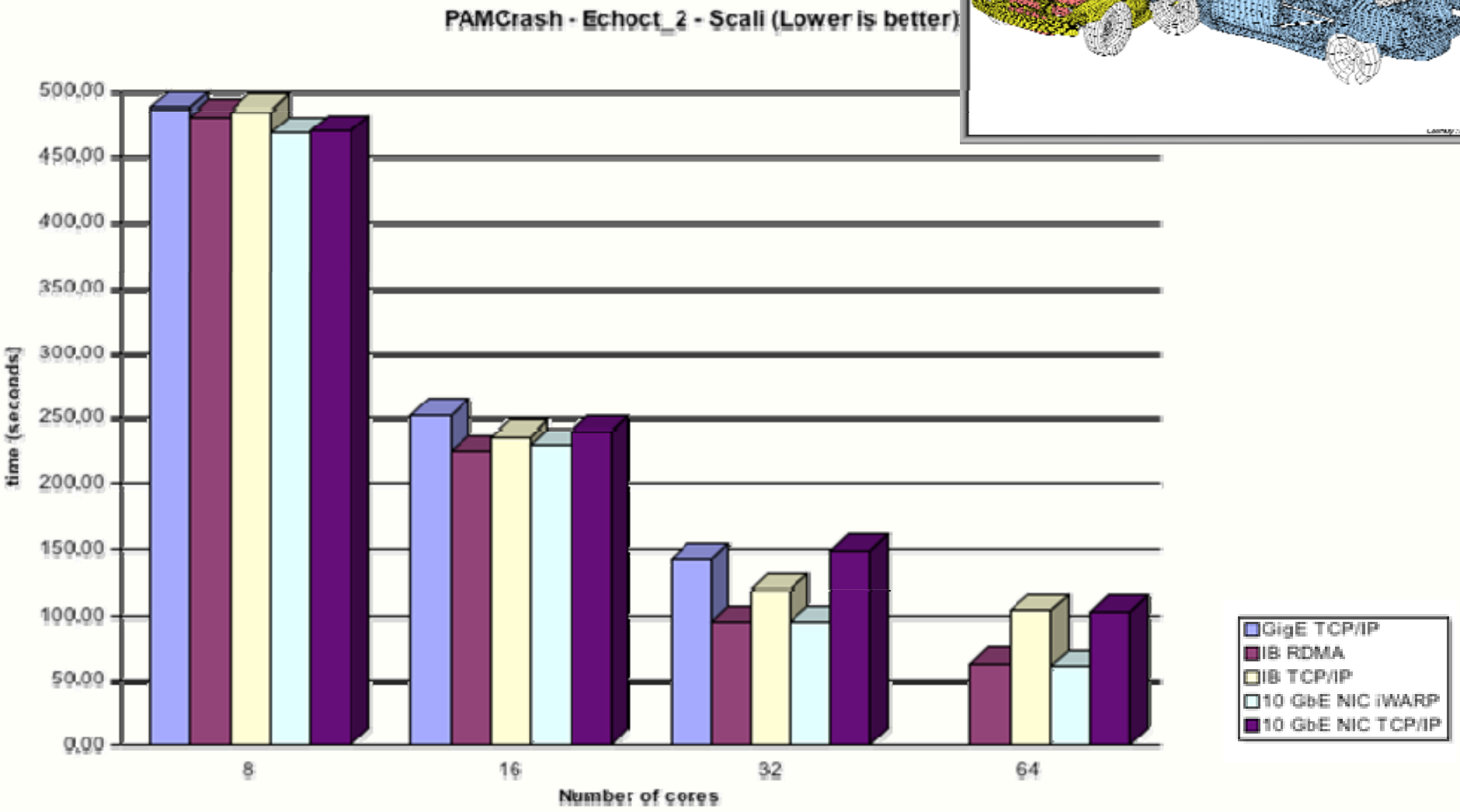
Performance

10 Gig Ethernet offers:

- Same familiar operating environment
- Ease of use, debug, and management
- Path to 40 and 100 Gig Ethernet
- 10x bandwidth and 8x better latency vs. Gig Ethernet
- *But – do applications run faster !??!?*
  - Vendors talk about micro-benchmarks
  - Most users care about execution time

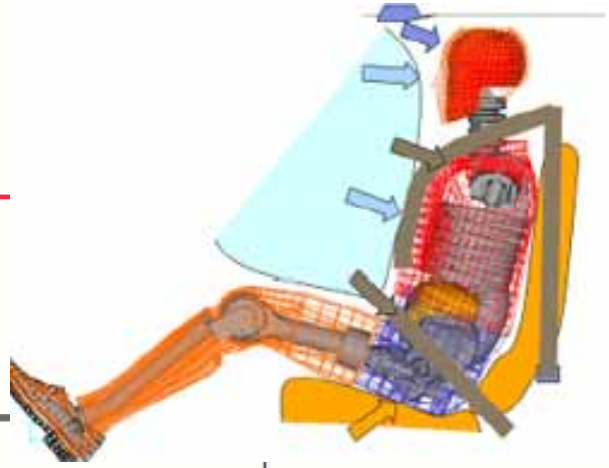
PAM CRASH:
Elapsed Time (sec)

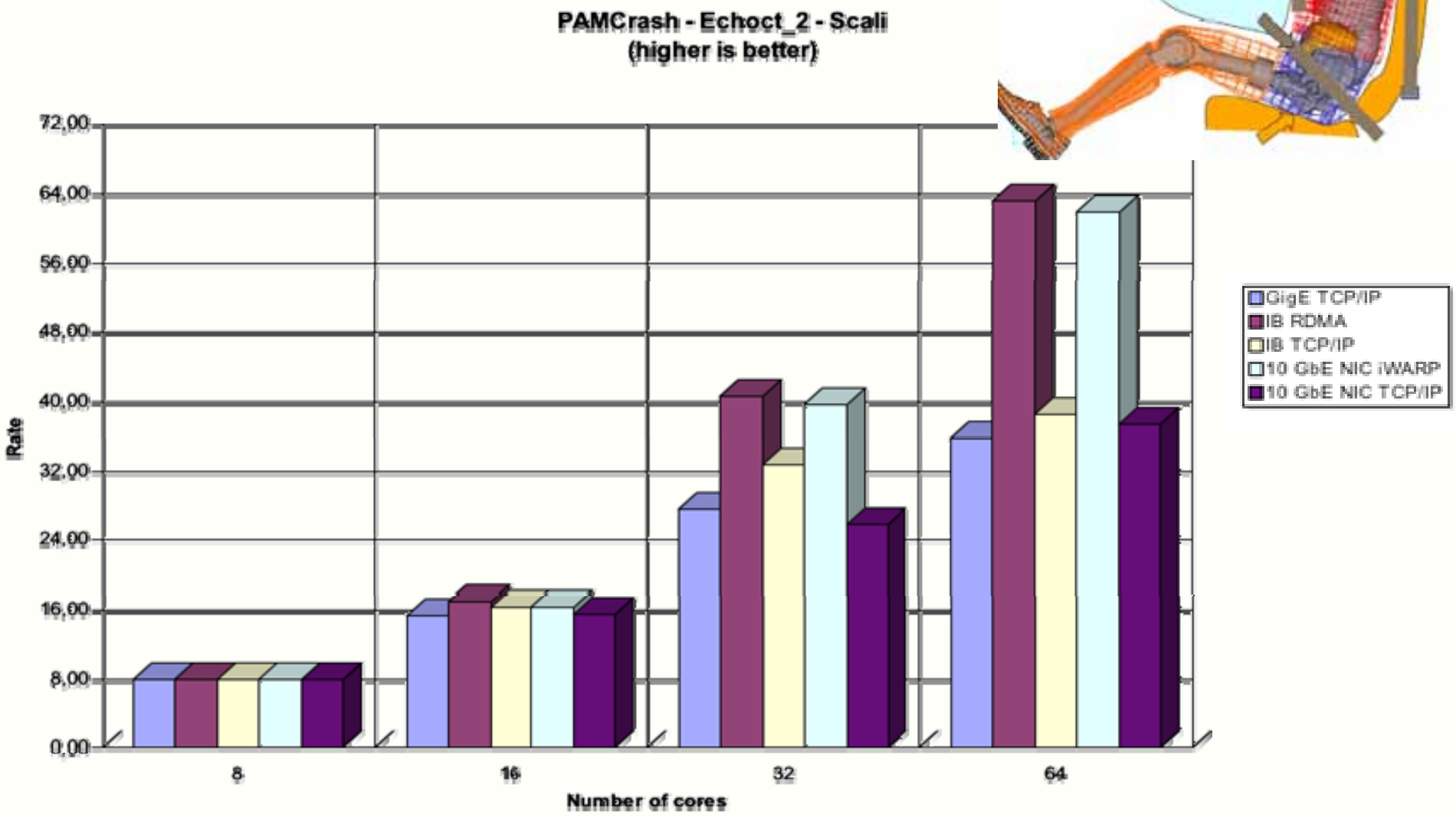




10GE 32% faster than 1G, equal to IB DDR, for 32 cores

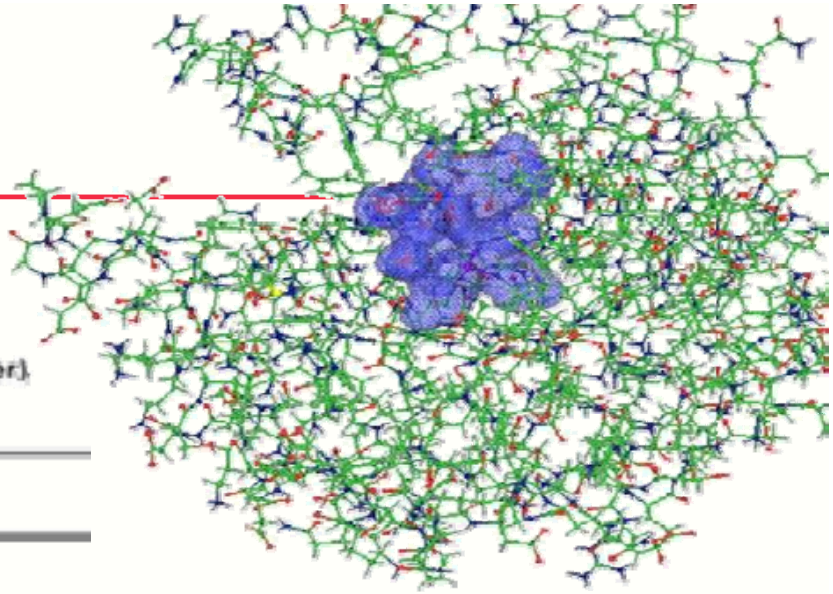
PAM CRASH: Speed Up



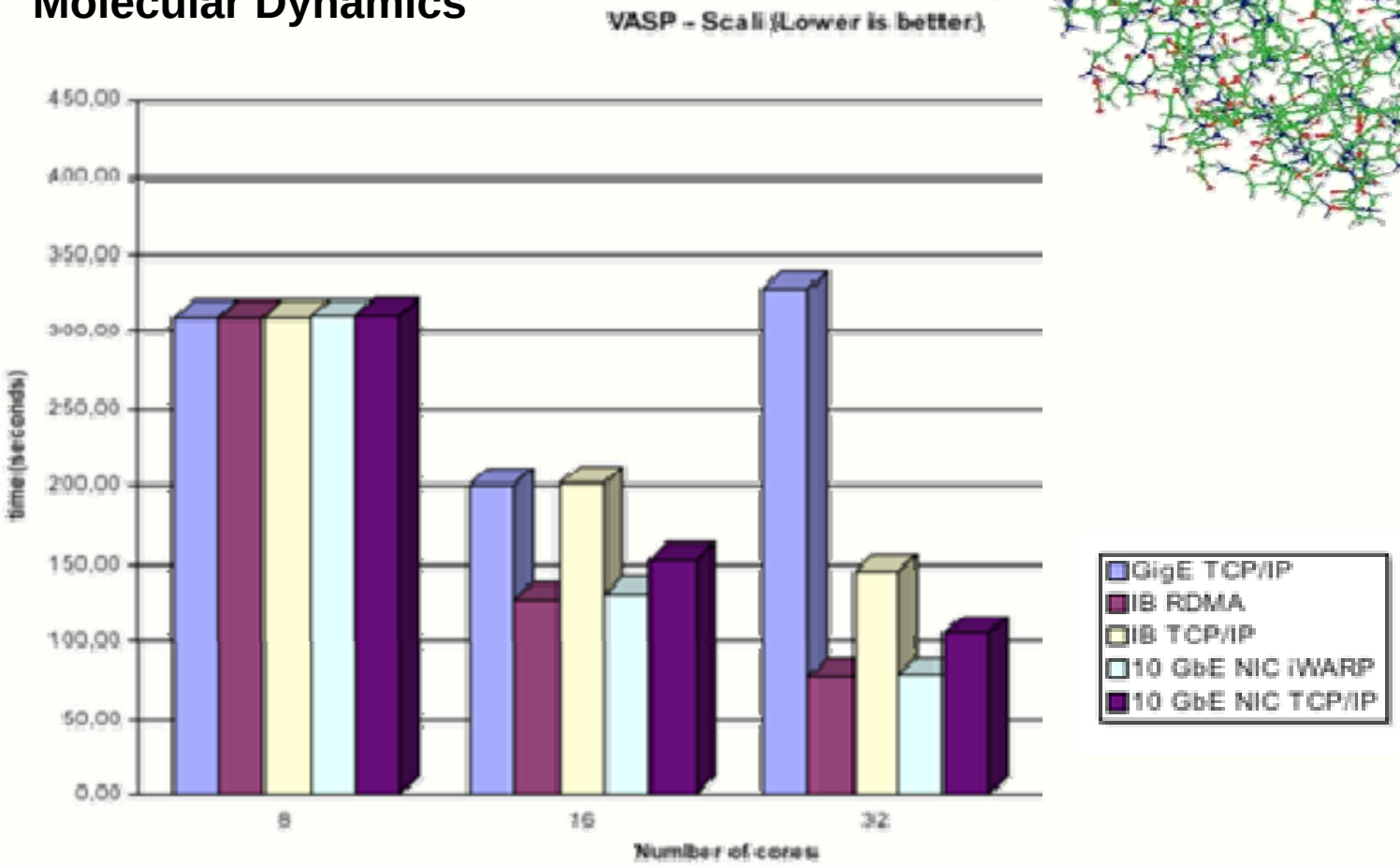


10GE 70% faster than 1G, equal to IB DDR, for 64 cores

VASP 4.6.28:
Elapsed Time (sec)

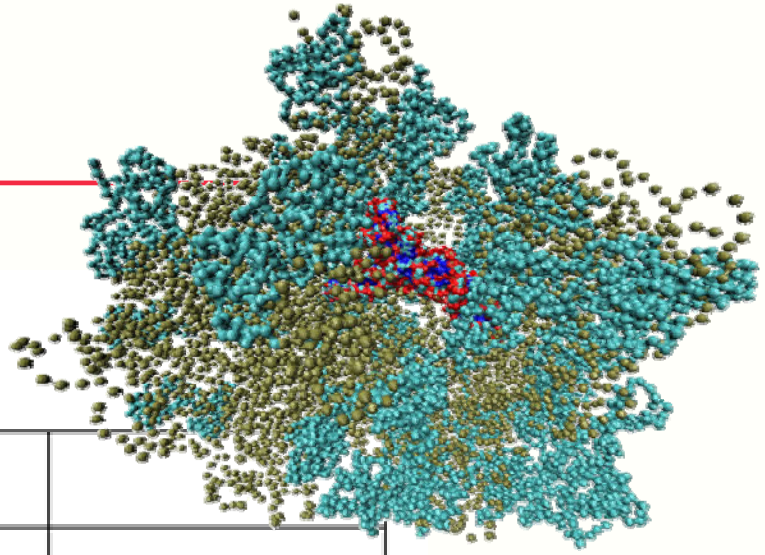


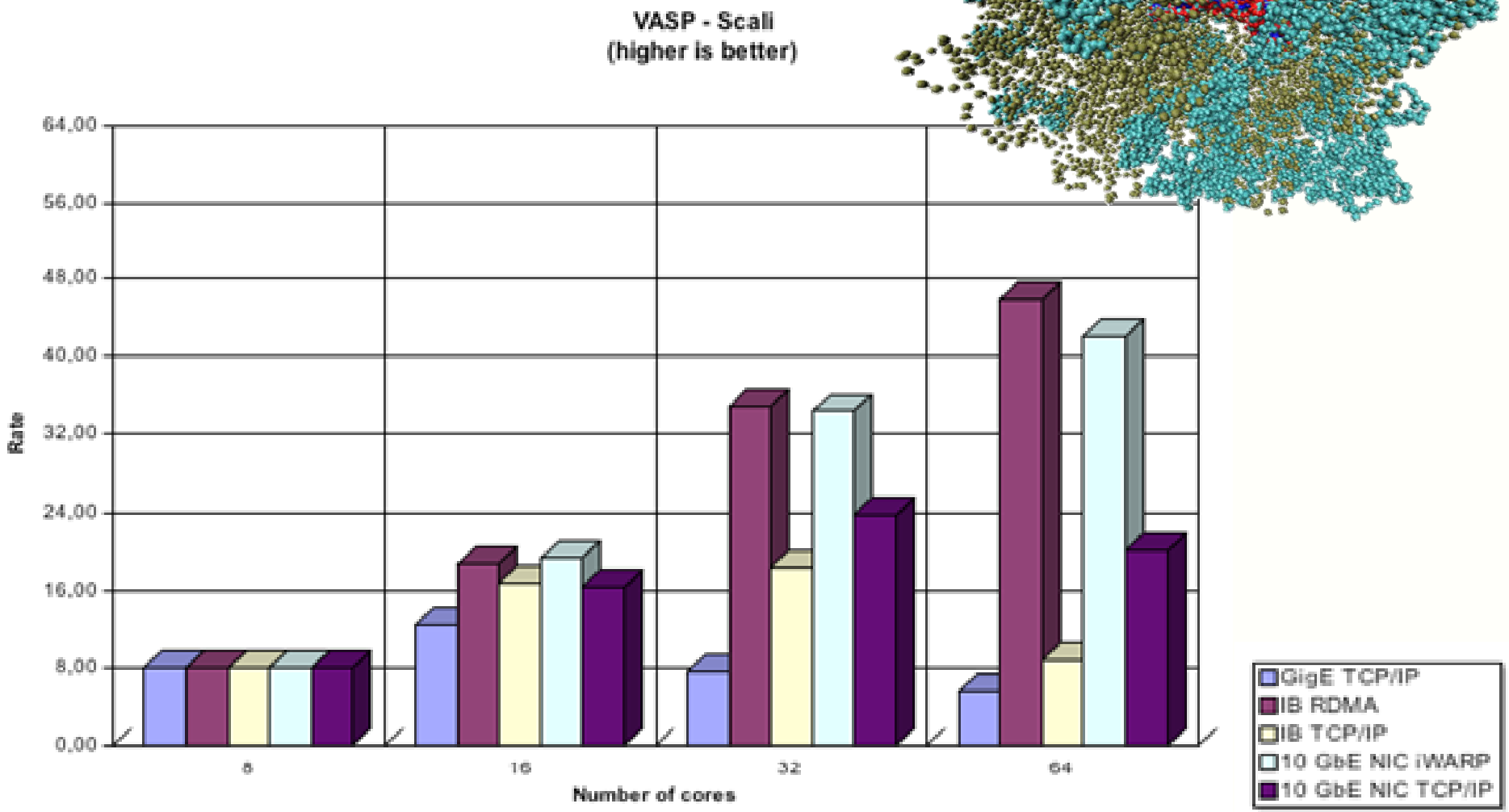
Vienna Ab-initio Simulation Package
Molecular Dynamics



10GE 4.25x faster than 1G, equal to IB DDR, for 32 cores

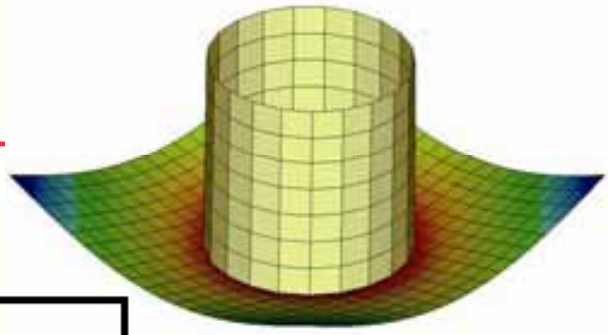
VASP 4.6.28: Speed Up



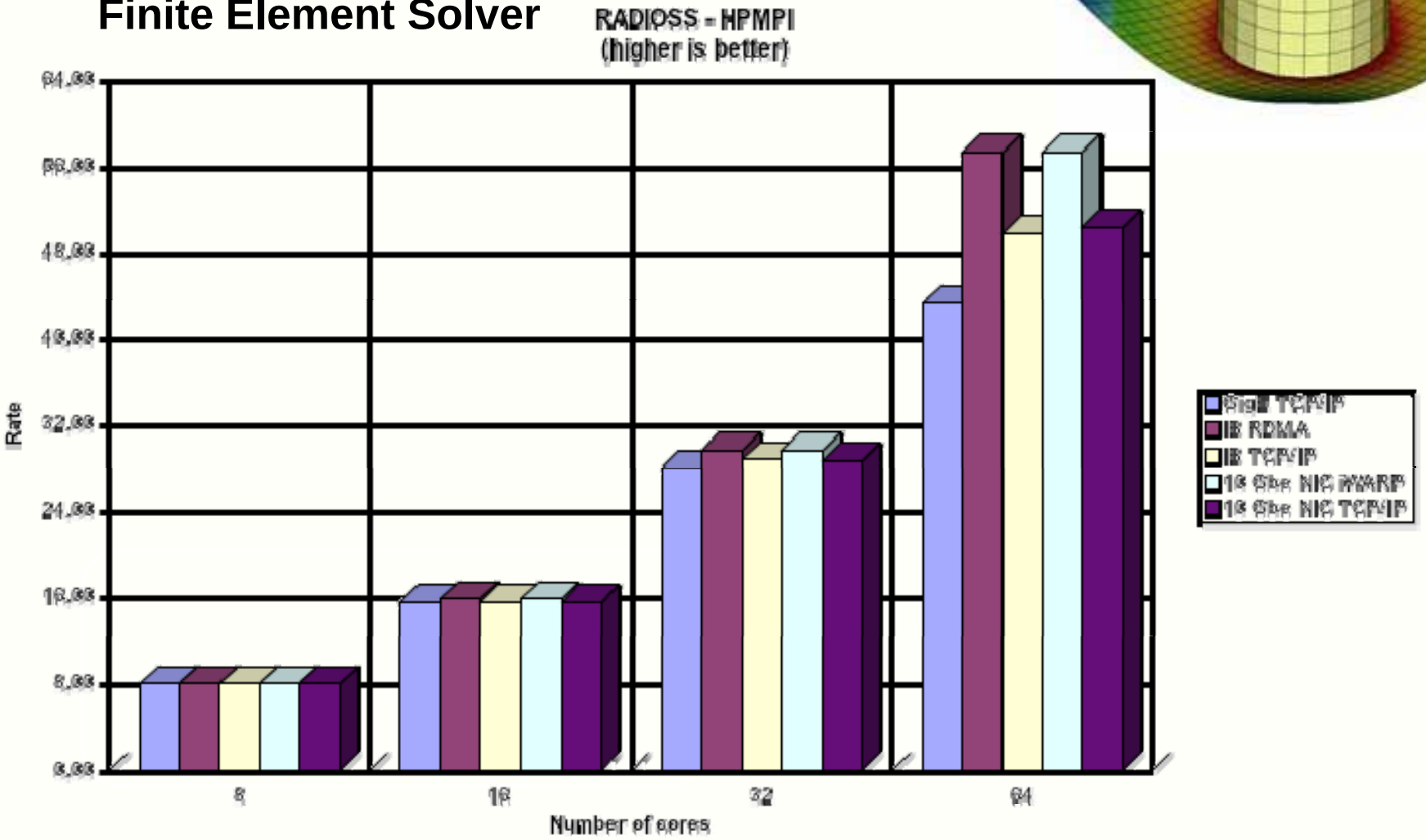


10GE 6.3x faster than 1G, almost equal to IB DDR, for 64 cores

RADIOSS 9.0: Speed up



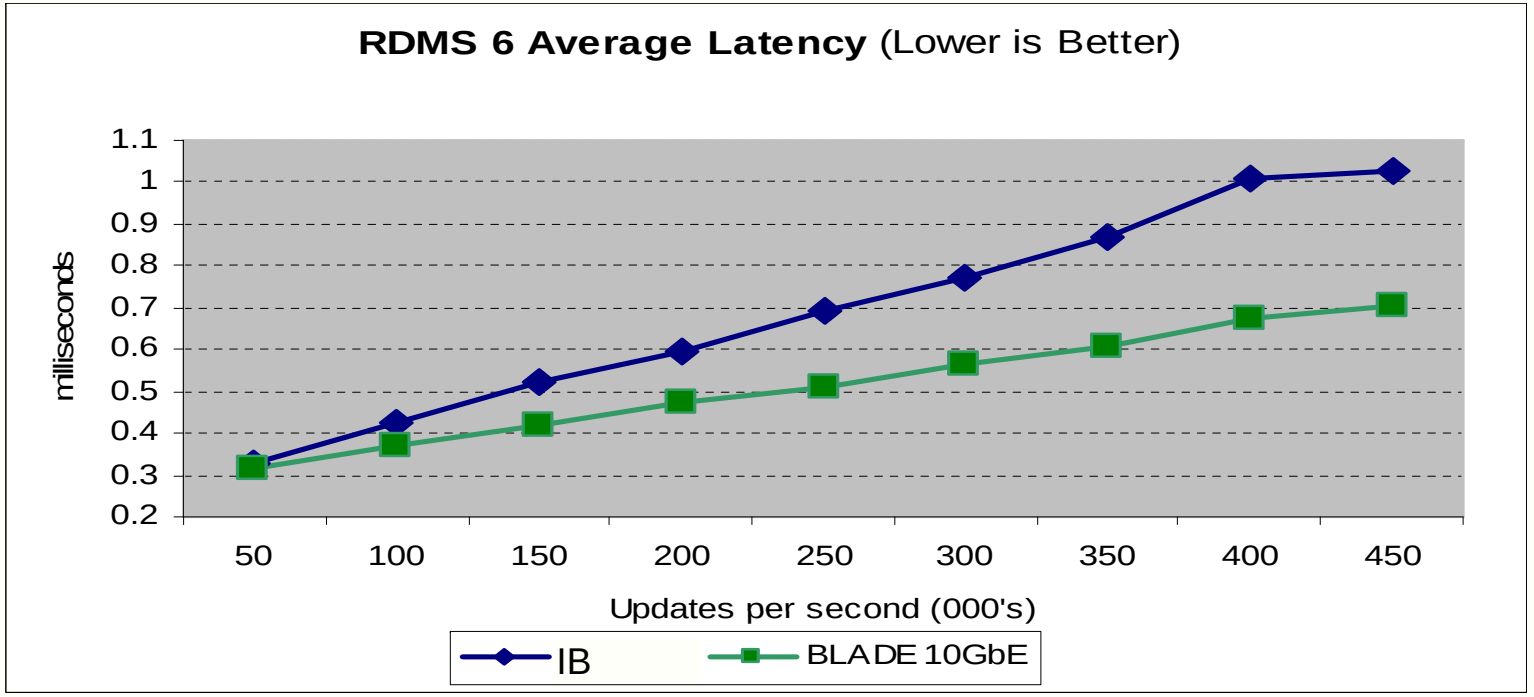
Finite Element Solver



10GE 30% faster than 1G, equal to IB DDR, for 64 cores

RMDS Performance
BLADE's 10GbE vs. InfiniBand

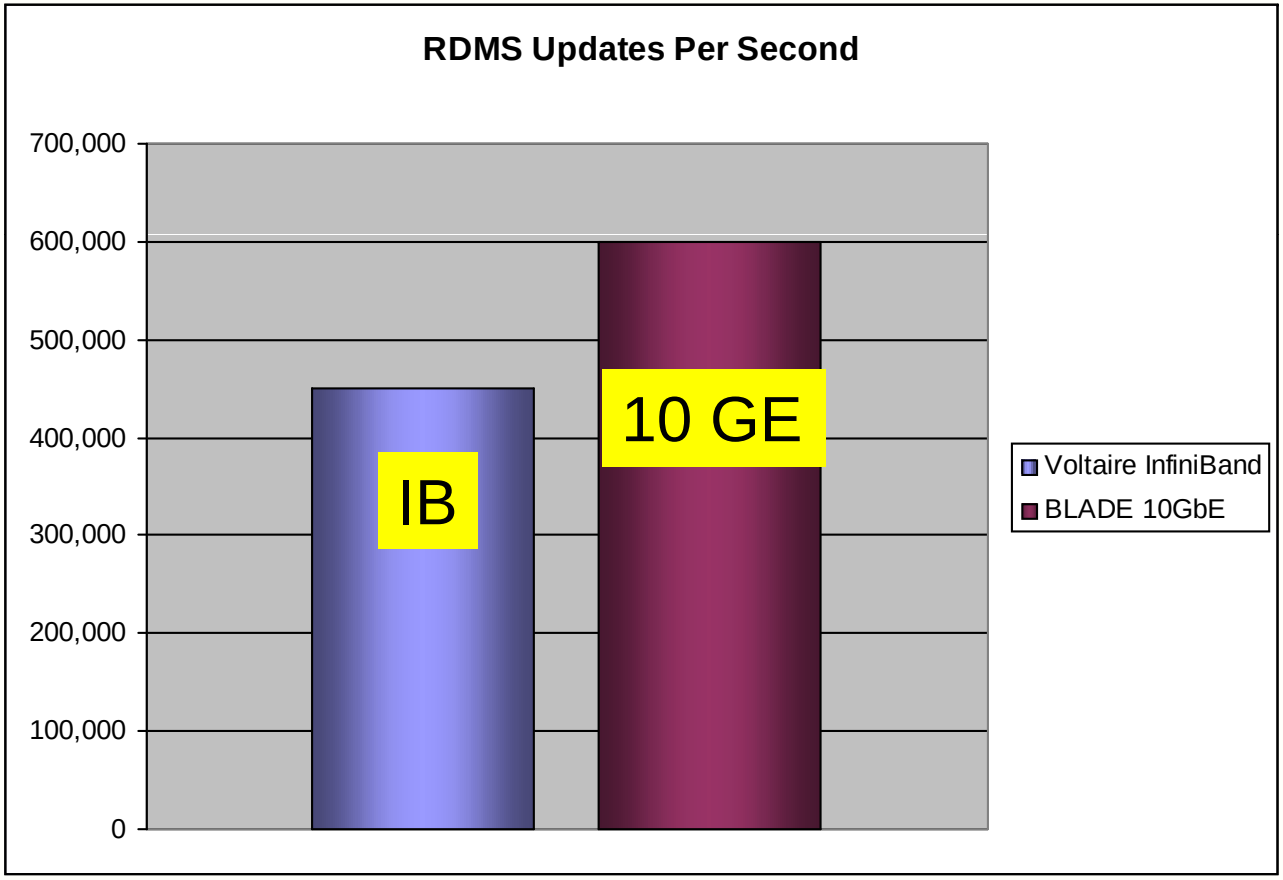
- BLADE's 10GbE outperformed InfiniBand
  - Significantly higher updates per second
  - 31% Lower latency than InfiniBand



**Voltaire and BLADE tests used similar 3 GHz Xeon 5160 based servers with 4MB L2 cache*

RMDS Performance
BLADE's 10GbE vs. InfiniBand

- BLADE's 10GbE outperformed InfiniBand



**Voltaire and BLADE tests used similar 3 GHz Xeon 5160 based servers with 4MB L2 cache*

Why 10G Ethernet Now?

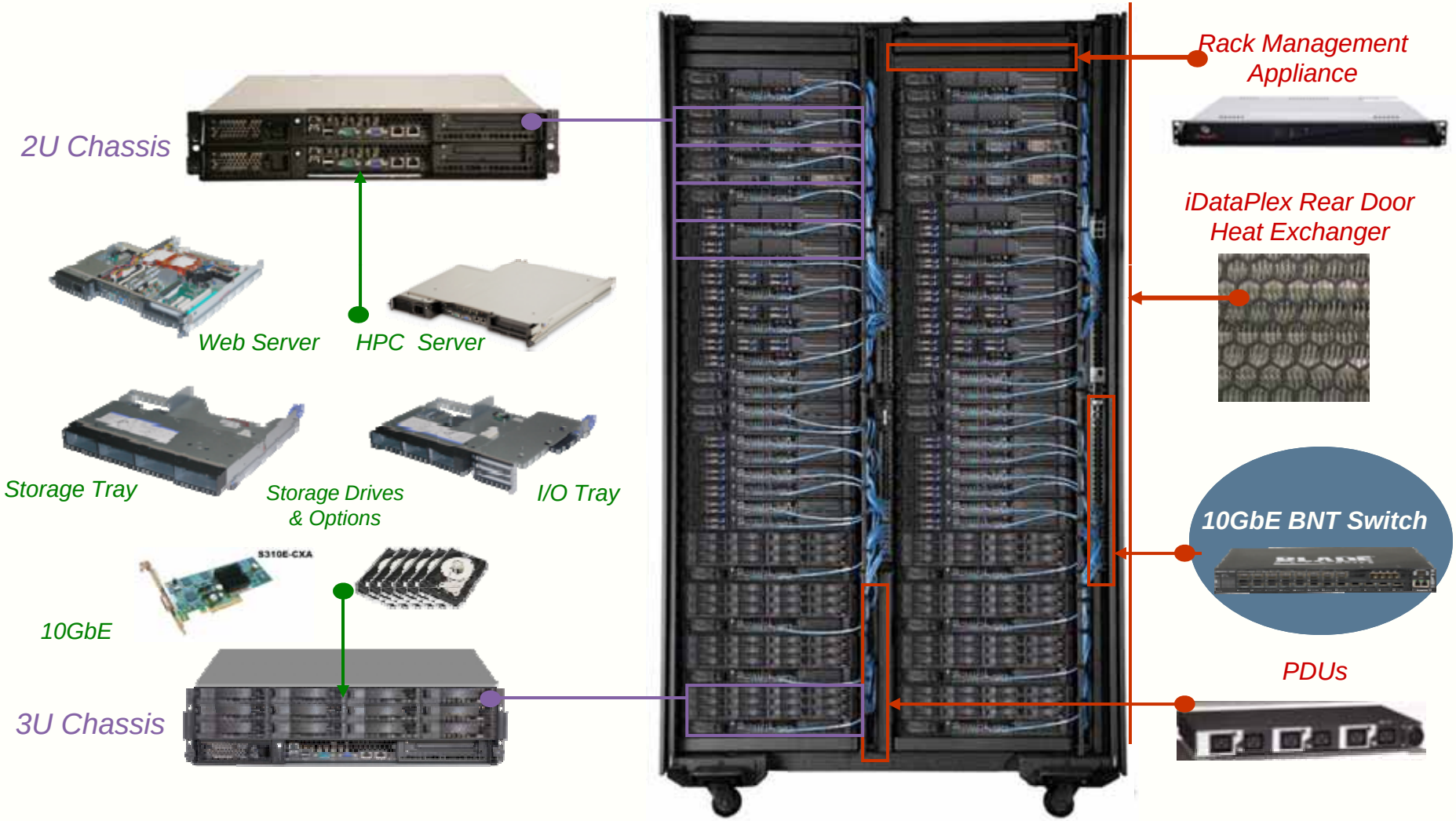
- Increasing demands of processors like quad core
- Prices are dropping
  - 10 Gig NICs
  - Switches under $500/port
  - Very attractive price performance
- IT Skill sets – easier to move to 10G Ethernet
- Technologies are more proven
  - CX4 and SFP+ are becoming the preferred PHY connections
  - Benchmarks are emerging
  - Early adopters and testing environments are delivering proof points of 10G Switch Scaling

BLADE Network Technologies

BLADE is the market-leading supplier of Gigabit and 10G Ethernet networking infrastructure solutions for blade server based environments

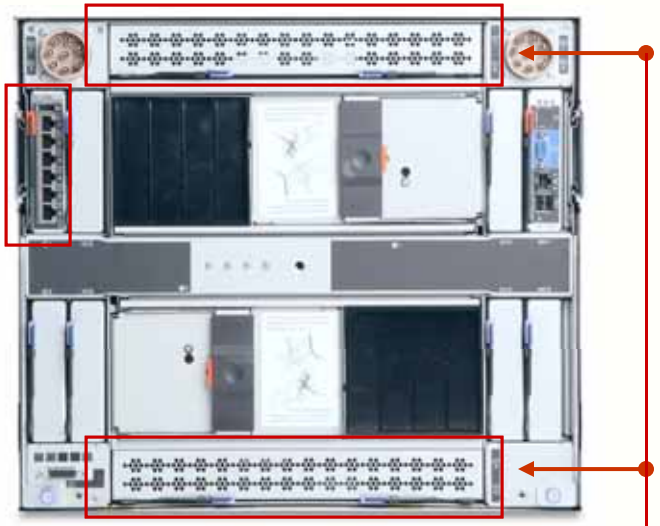
- First blade switch delivered in 2003
  - BLADE was a former division of Nortel and has been fully independent of Nortel since 2006
- Eight embedded Nortel Switch Modules for IBM BladeCenter
  - And growing!
- Over 45% blade networking market share
  - For every Cisco blade switch out there are 2 Nortel switches
  - Over 5 million ports connected to over 1 million blades
  - In over half the Fortune 500
- 6 Million hours of actual MTBF
- Management & Network Virtualization Tools
  - SmartConnect™ (with VMReady™) & BLADEHarmony™

IBM System x iDataPlex



IBM BladeCenter

Front View



Rear View

Blade server

10GbE Adapter

Switches

Question & Answer

Thank you !!

iDataPlex Cluster 1350

RackSwitch G8124 & G8100

BLADE's Nortel 10G Blade Switch for BladeCenter