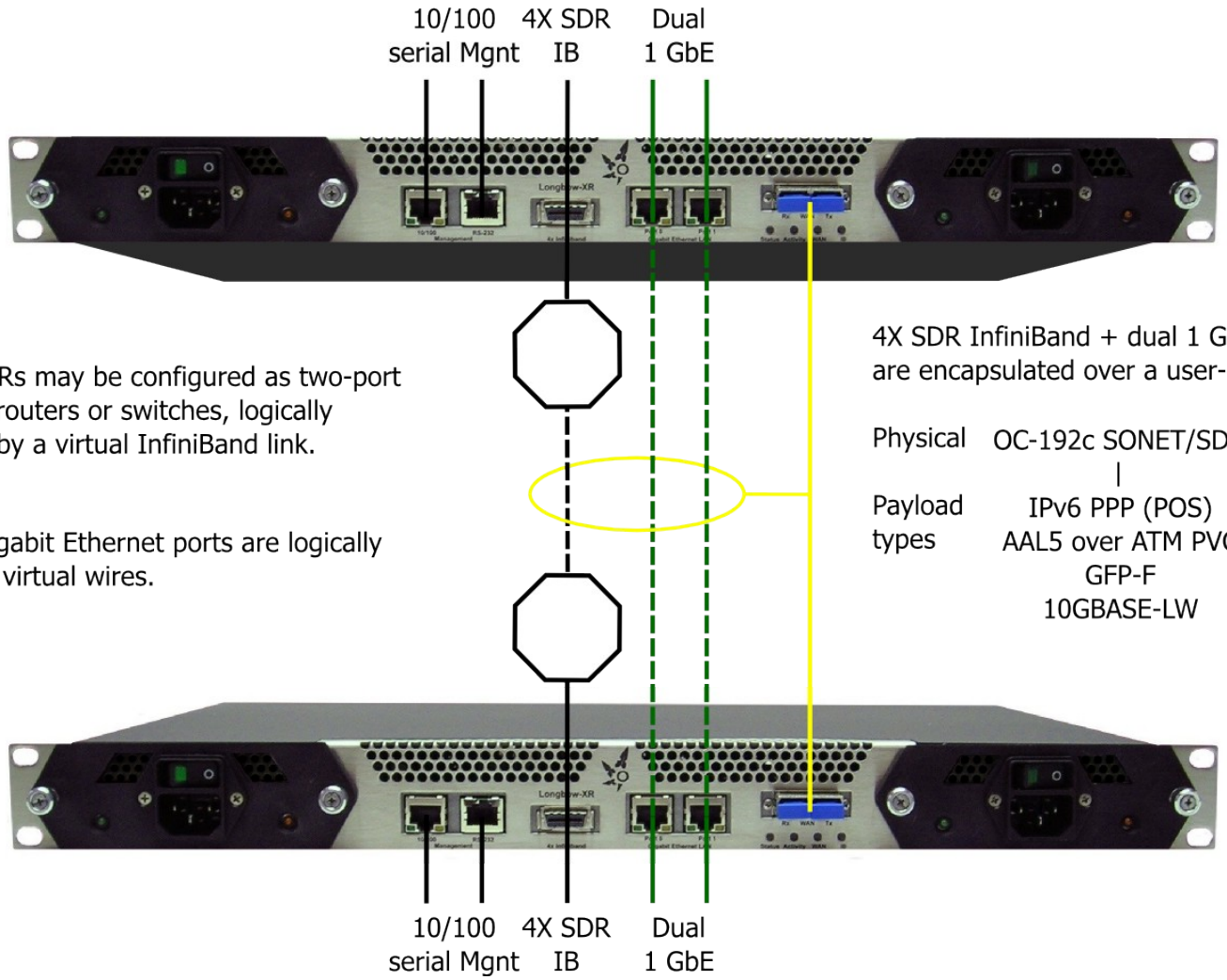
Adventures in Wide Area InfiniBand

OPENFABRICS ALLIANCE

OBSIDIAN

David Southwell

Longbow XR in production



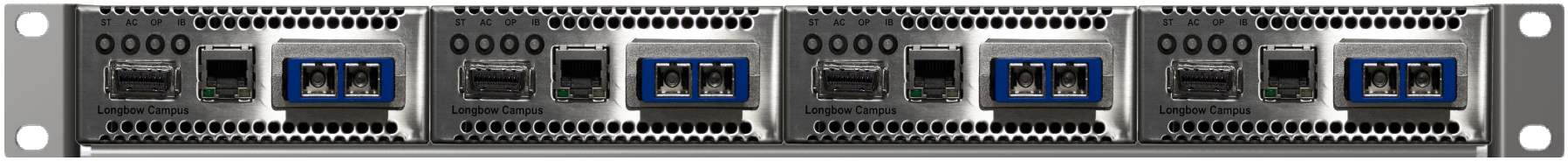
Longbow XRs may be configured as two-port InfiniBand routers or switches, logically connected by a virtual InfiniBand link.

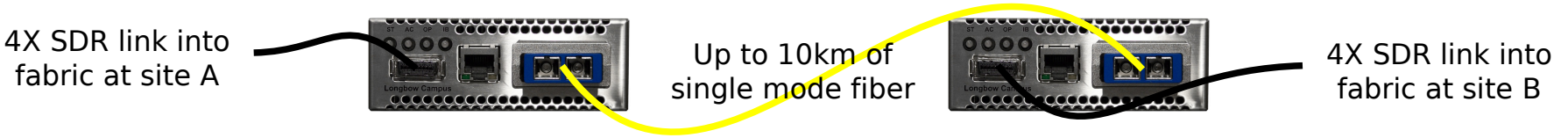
The dual 1 Gigabit Ethernet ports are logically connected by virtual wires.

4X SDR InfiniBand + dual 1 Gigabit Ethernet links are encapsulated over a user-configurable WAN:

Physical	OC-192c SONET/SDH	10GBASE-LW/LR
	|	|
Payload types	IPv6 PPP (POS) AAL5 over ATM PVC GFP-F 10GBASE-LW	IPv6

Longbow Campus in production



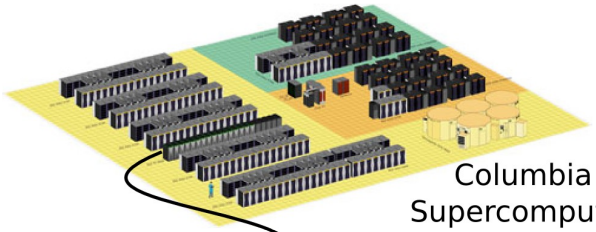


SC|06 NASA/NLR

NASA Ames Research Center - Mountain View, CA

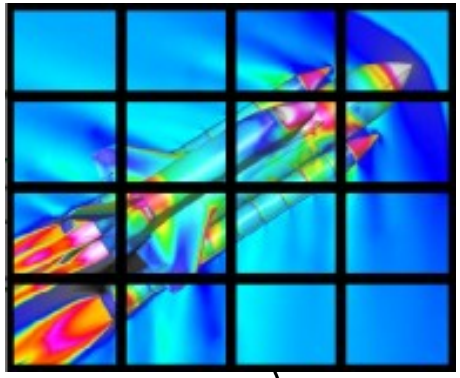
NASA booth (917) Supercomputing 2006 - Tampa, FL

InfiniBand based 3x3 Hyperwall visualization system



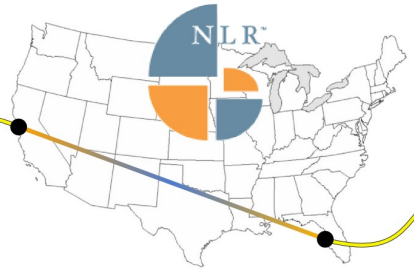
Columbia Supercomputer

InfiniBand fabric



Obsidian Longbow XR

Obsidian Longbow XR

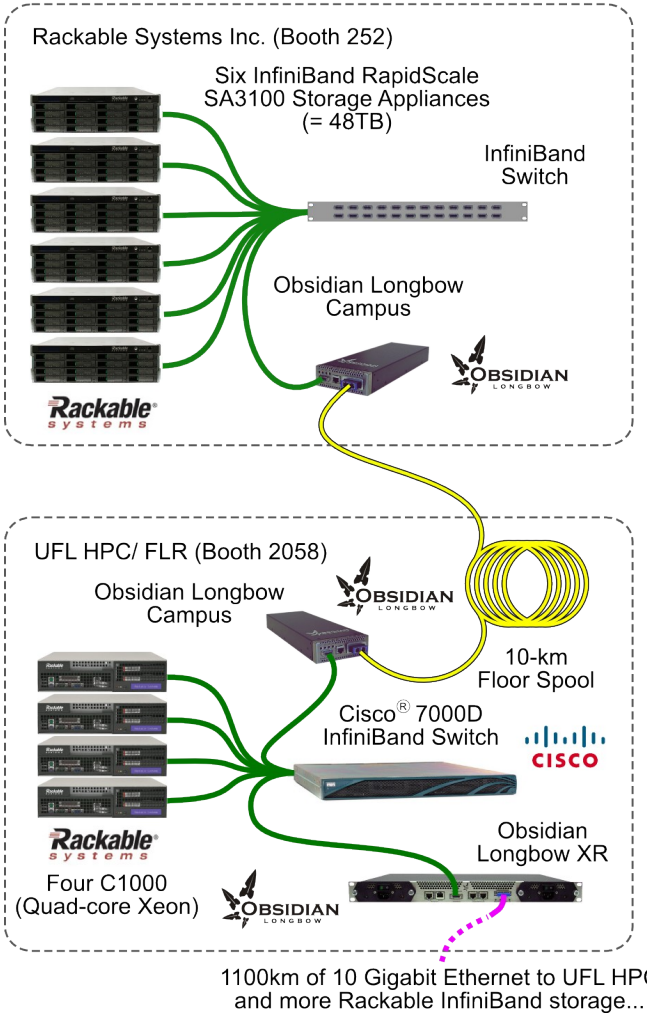


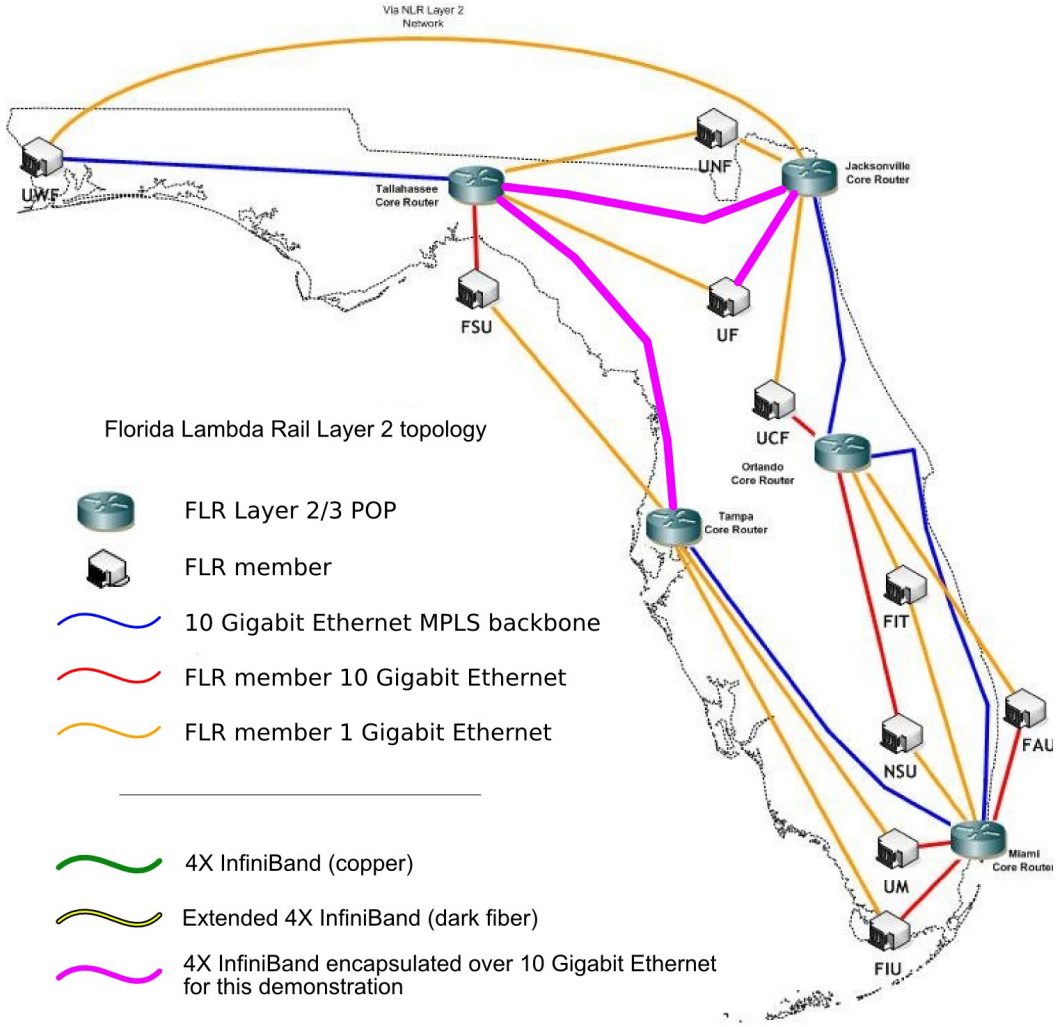
10 Gigabit Ethernet
Cisco 6509 with
IPSec option

10 Gigabit Ethernet
Cisco 6506 with
IPSec option

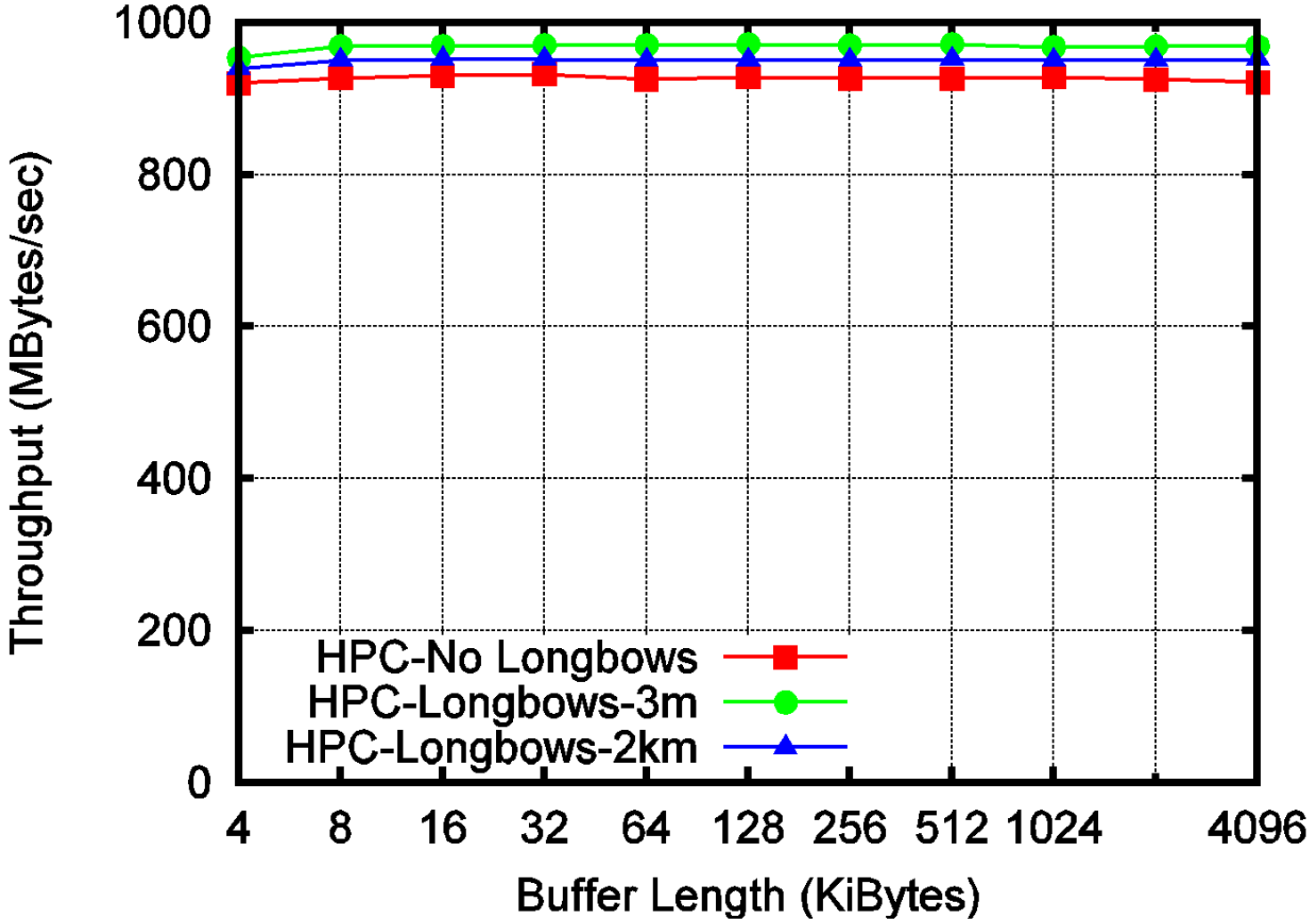
6,600km of 10Gigabit Ethernet layer 3

SC|06 UFLHPC/ FLR/ Rackable

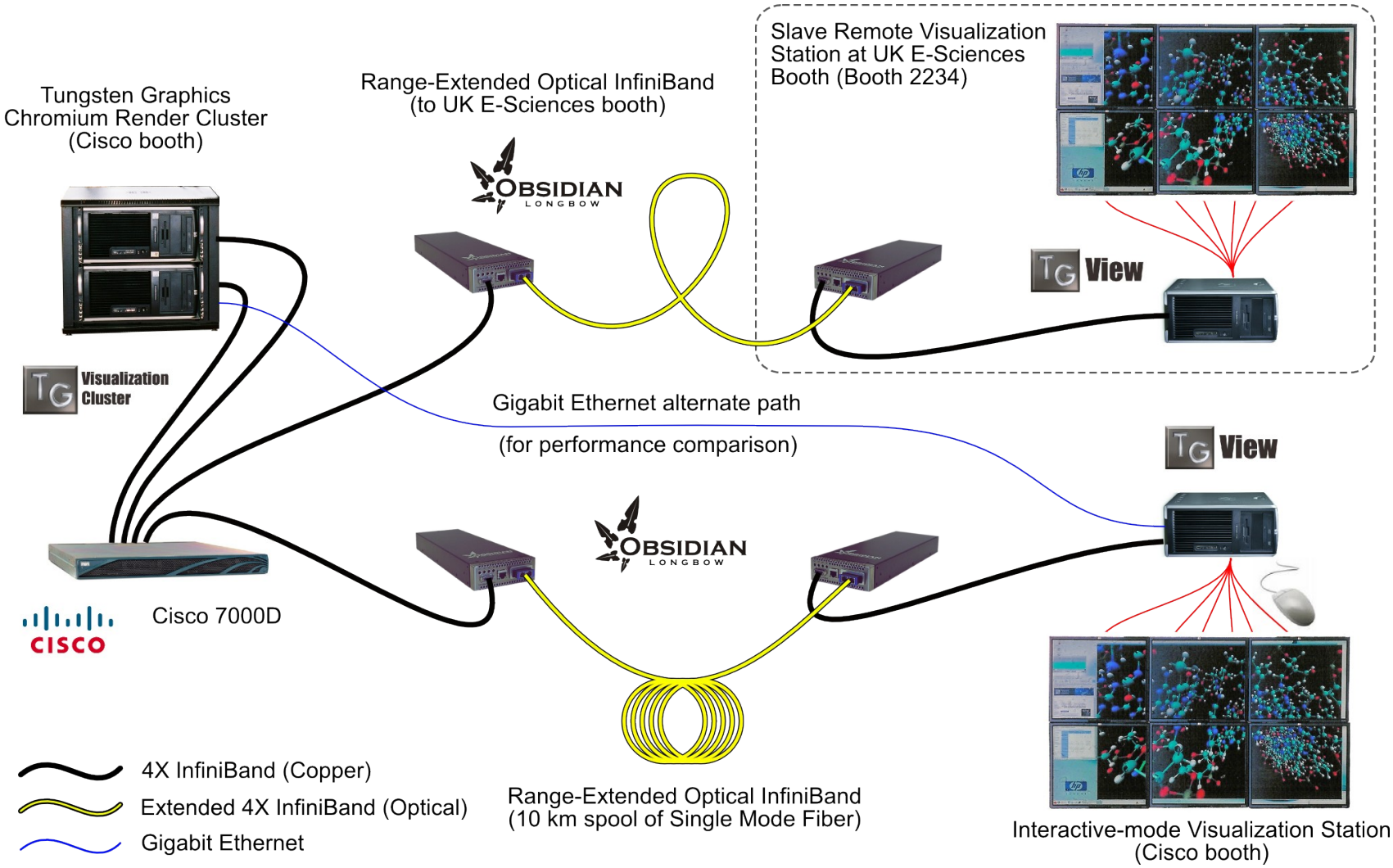




(UFL HPC Campus)

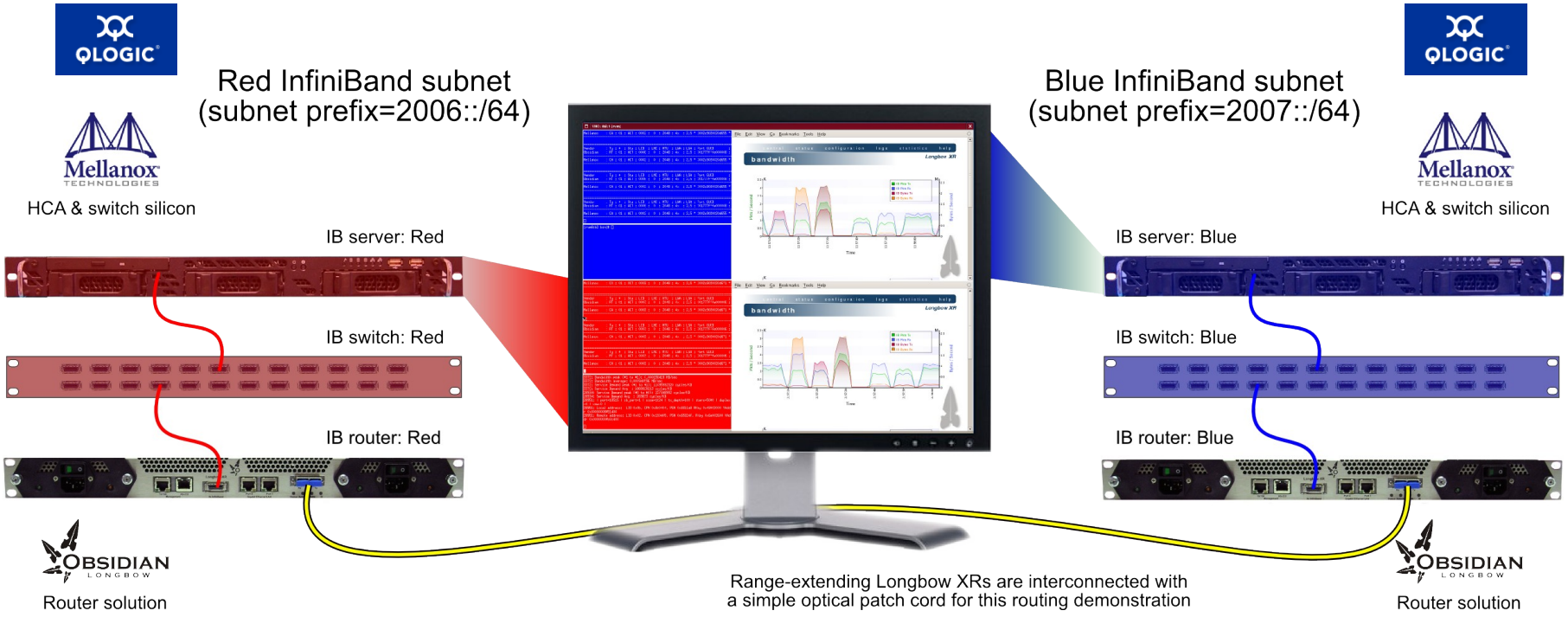


SC|06 Cisco/ TungstenGraphics

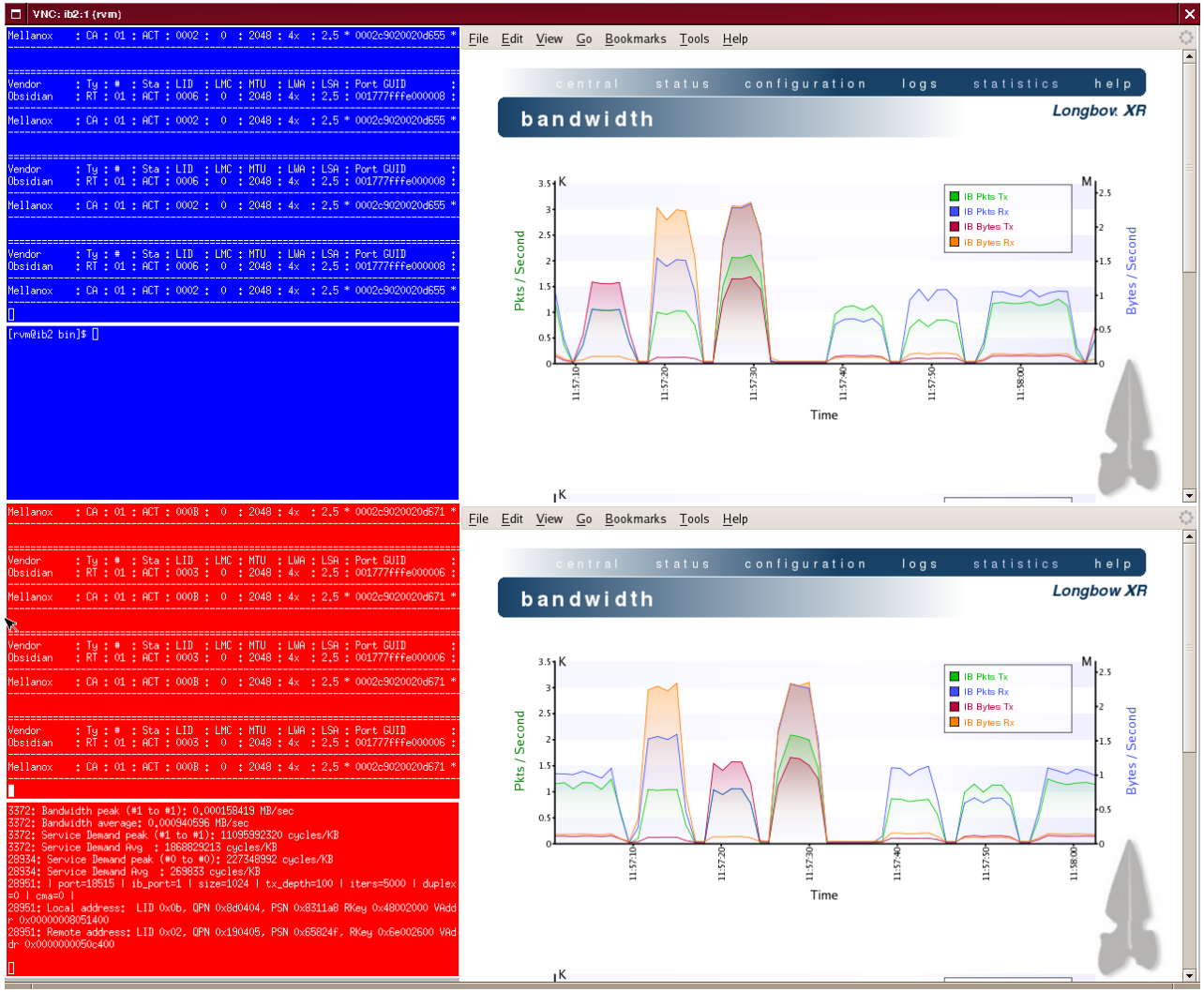


SC|06 Cisco/ TungstenGraphics

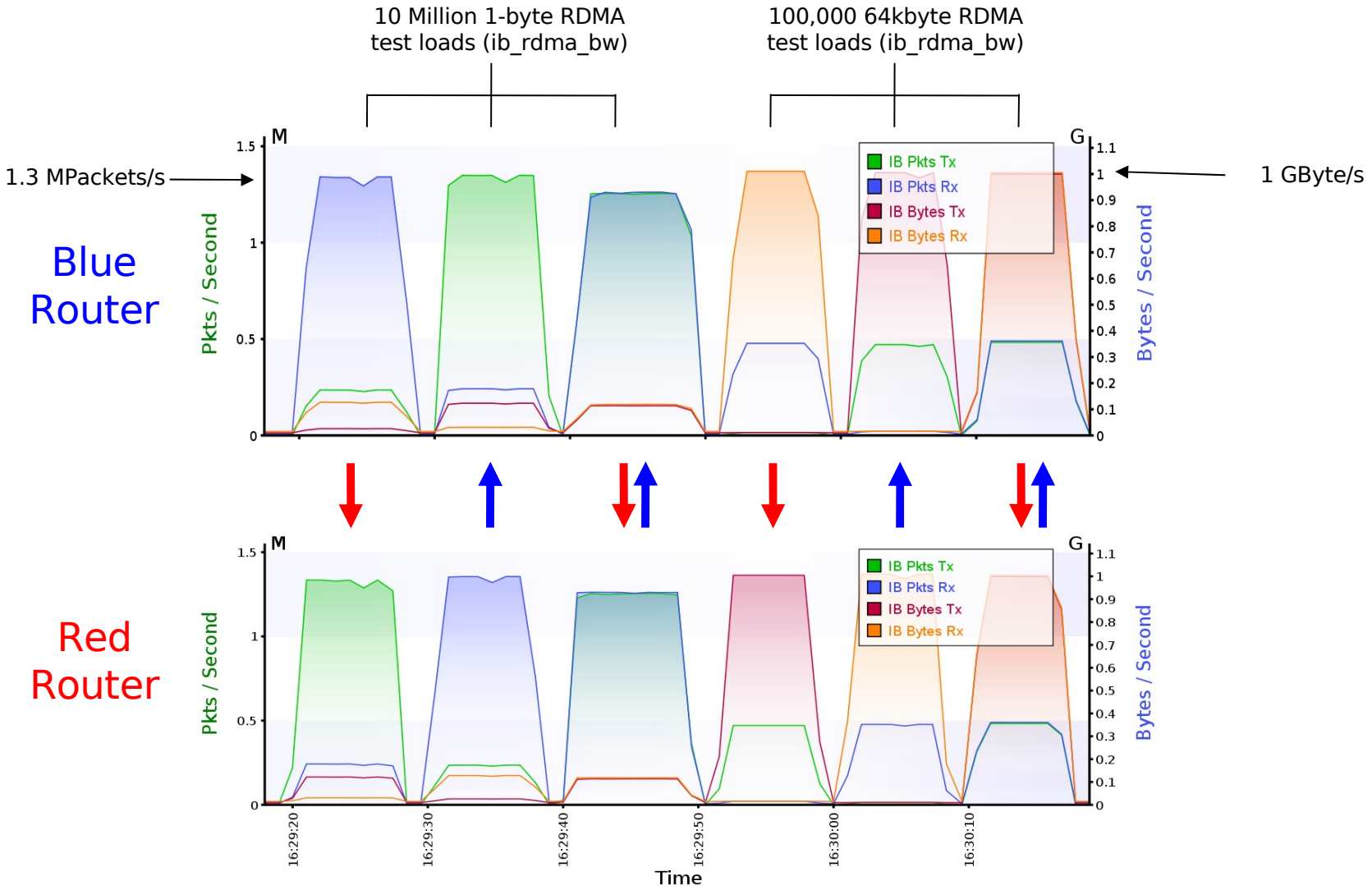
SC|06 Mellanox/ QLogic



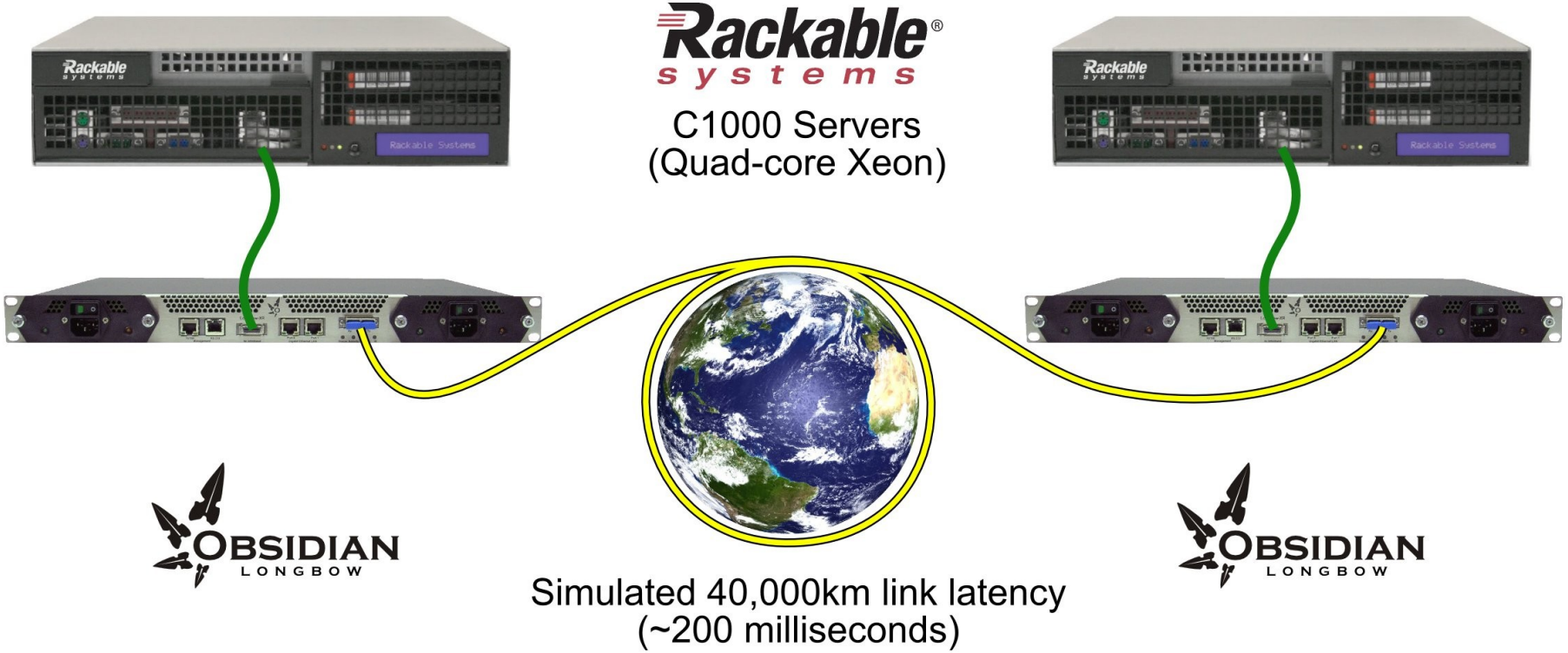
SC|06 Mellanox/ QLogic



SC|06 Mellanox/ QLogic



AFCEA West 2007



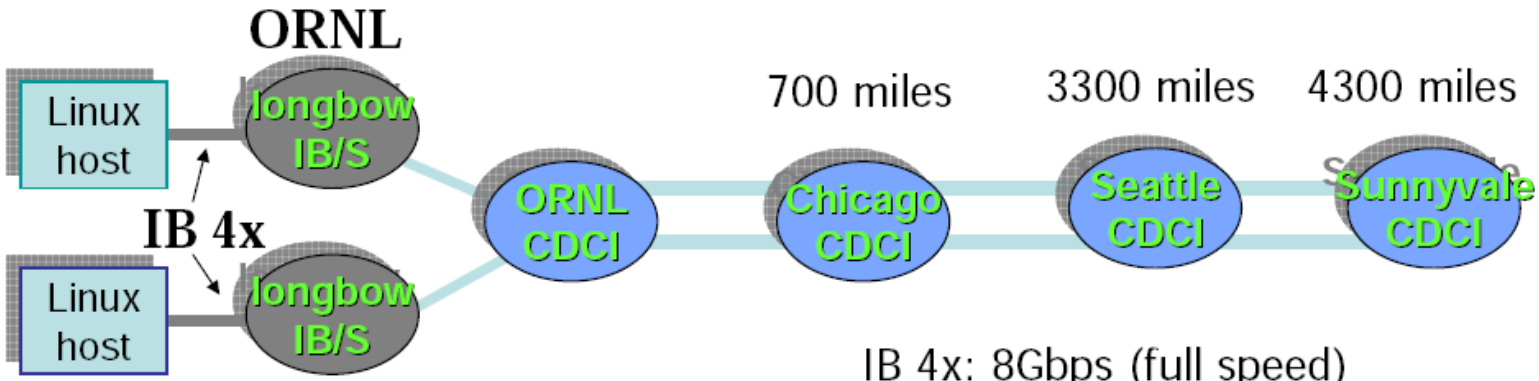
JCTD-Large Data

Oak Ridge National Labs

Infiniband Over SONET

- Infiniband is effective data transport protocol for storage networks (few miles):
- TCP is not easily extended and not optimal for such data transfers

Question: Is IB effective over wide-area? - Yes



IB 4x: 8Gbps (full speed)
Host-to-host local switch:7.5Gbps

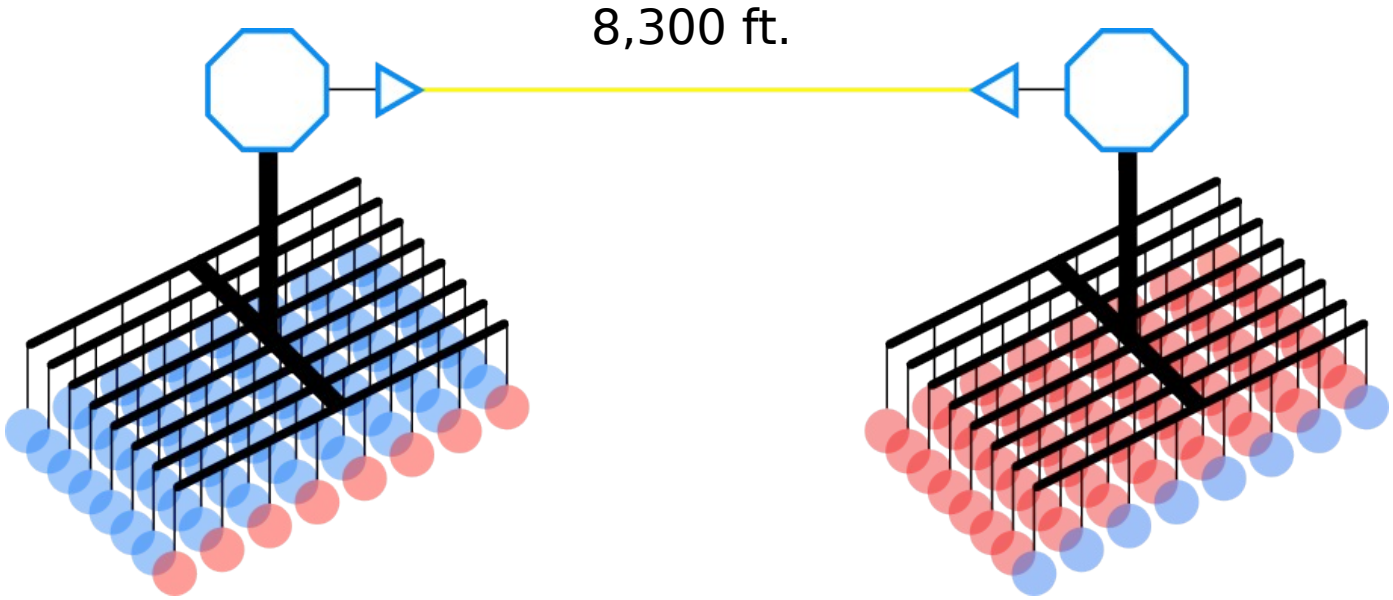
ORNL loop -0.2 mile: **7.5**Gbps

ORNL-Chicago loop – 1400 miles: **7.46**Gbps

ORNL- Chicago - Seattle loop – 6600 miles: **7.23**Gbps

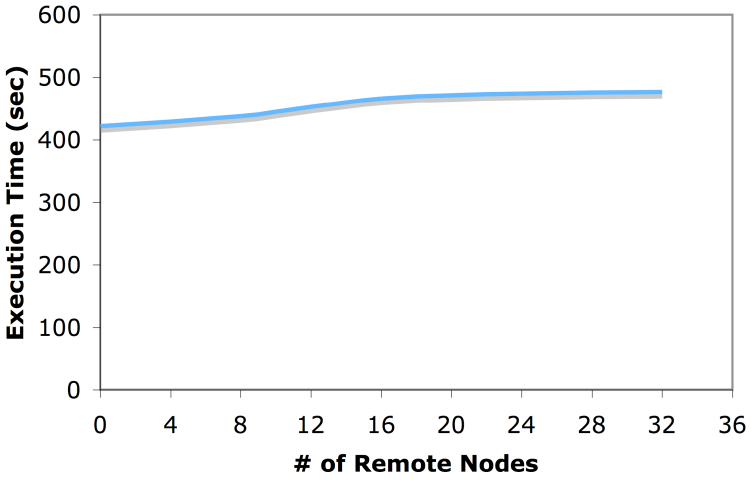
ORNL – Chicago – Seattle - Sunnyvale loop – 8600 miles: **7.20**Gbps

Arizona State University

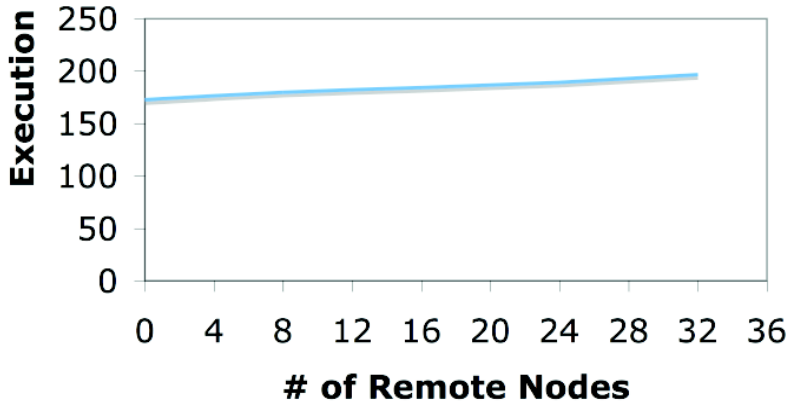


Σ # active cores = 64 (constant)

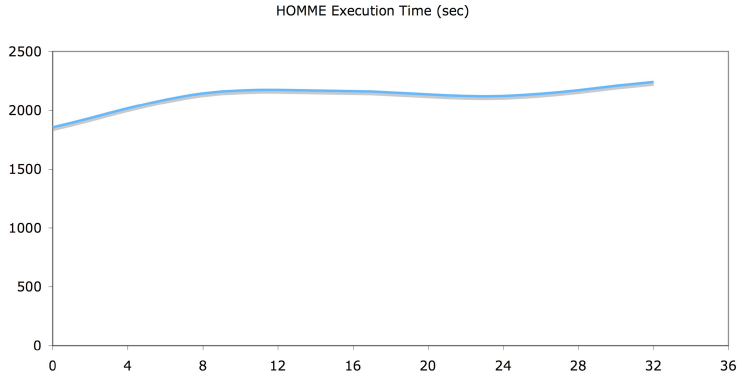
Arizona State University



MIMD Lattice Computation (MILC)

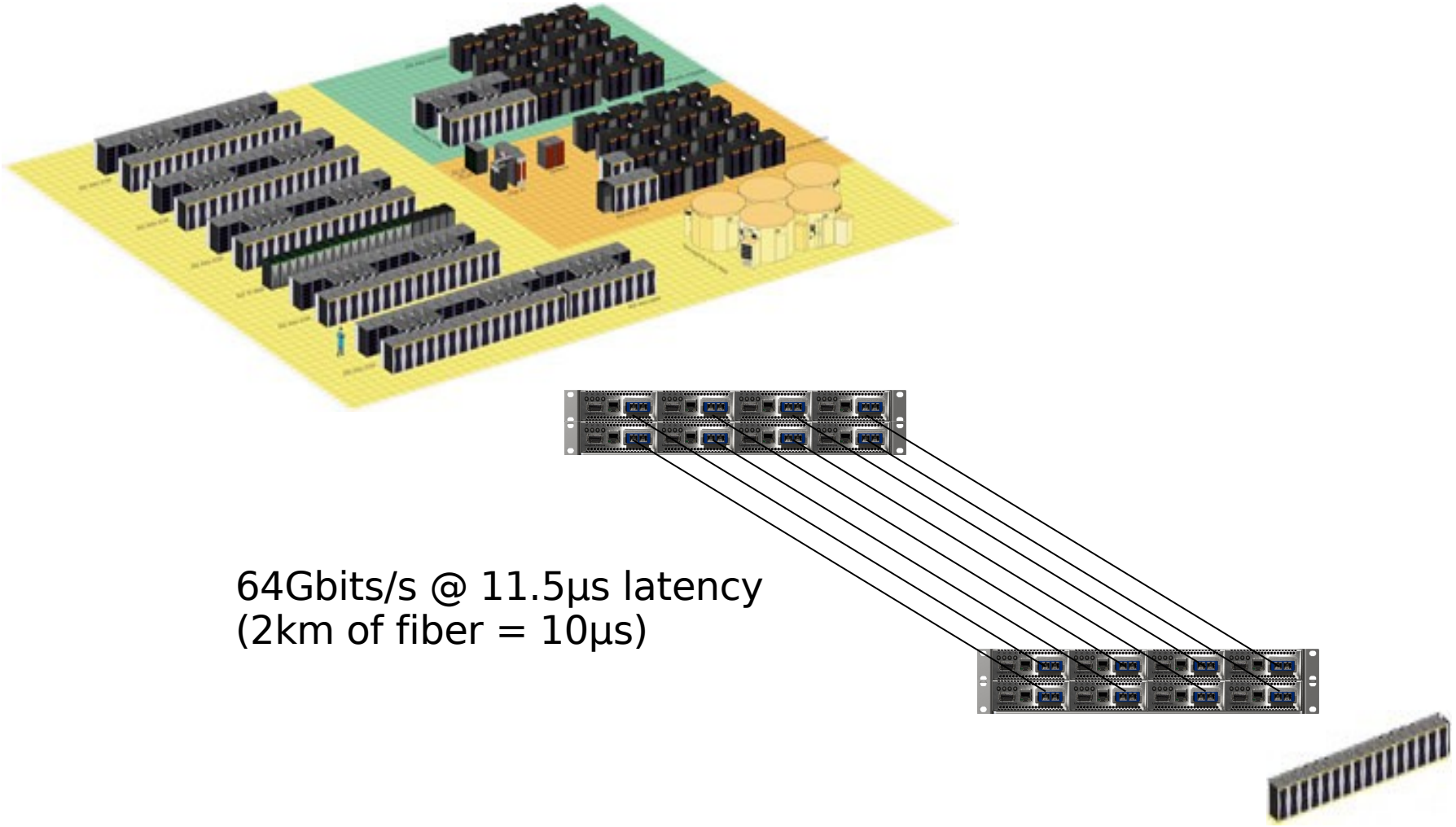


Weather Research and Forecasting (WRF) Code



High Order Methods Modelling Environment (HOMME)

NASA Columbia Expansion



64Gbits/s @ 11.5µs latency
(2km of fiber = 10µs)

Closing...

New InfiniBand application domains are being validated:

- Efficient WAN transport – reliable RDMA without TCP/IP tuning issues
- Data center replication/ disaster recovery
- High fidelity remote visualization
- Low latency optical networking
- InfiniBand storage area networks
- Cluster clustering – Campus Area Grid
- Supercomputer expansion – building annexation
- Streaming media transport
- Streaming big science transport

...but hurdles remain...

Closing...

We see a strong need for :

- OpenFabrics support for InfiniBand routing for subnet isolation
- Efficient IPoIB (to facilitate selling fat WAN links dedicated to InfiniBand)
- Parallel file system support to tolerate much higher link latencies

Thank You

OPENFABRICS
ALLIANCE