Mellanox TECHNOLOGIES
HPC ADVISORY COUNCIL

NETWORK OF EXPERTISE

CPMD Performance Benchmarks and Profiling

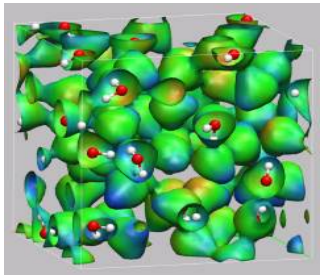
January 2009

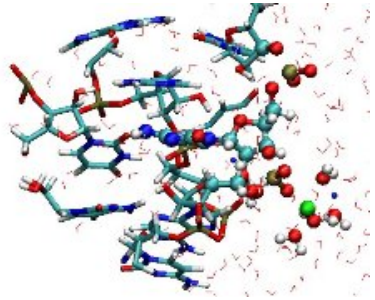
Note

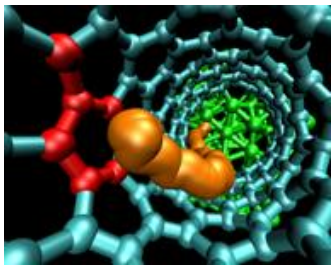
- **The following research was performed under the HPC Advisory Council activities**
  - AMD, Dell, Mellanox
  - HPC Advisory Council Cluster Center
- **The participating members would like to thank the CPMD consortium for their support**
- **For more info please refer to**
  - www.mellanox.com, www.dell.com/hpc, www.amd.com

Car-Parrinello Molecular Dynamics (CPMD)

- **CPMD**
  - A parallelized implementation of density functional theory (DFT)
  - Particularly designed for ab-initio molecular dynamics
  - Brings together methods
    - Classical molecular dynamics
    - Solid state physics
    - Quantum chemistry
  - CPMD supports MPI and Mixed MPI/SMP
  - CPMD is distributed and developed by the CPMD consortium







Objectives

- **The presented research was done to provide**
  - CPMD performance benchmarking
  - Cluster Interconnect effect on CPMD performance
  - CPMD performance comparison with different MPI
  - Understanding CPMD communication pattern

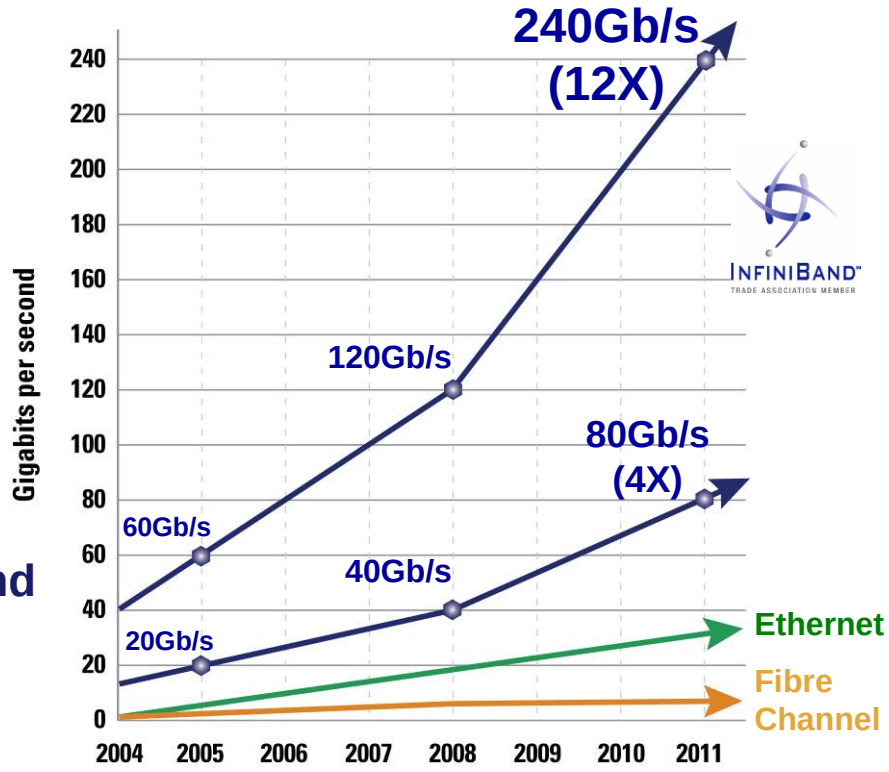
Test Cluster Configuration

- **Dell™ PowerEdge™ SC 1435 24-node cluster**
- **Quad-Core AMD Opteron™ 2382 processors (“Shanghai”)**
- **Mellanox® InfiniBand ConnectX® DDR HCAs**
- **Mellanox® InfiniBand DDR Switch**
- **Memory: 16GB memory, DDR2 800MHz per node**
- **OS: RHEL5U2, OFED 1.3 InfiniBand SW stack**
- **MPI: Open MPI 1.3, Platform MPI 5.6.5,**
- **Application: CPMD 3.13 with BLAS/LAPACK libraries**
- **Benchmark Workload**
  - C120
  - Wat32

Mellanox InfiniBand Solutions

- **Industry Standard**
  - Hardware, software, cabling, management
  - Design for clustering and storage interconnect
- **Performance**
  - 40Gb/s node-to-node
  - 120Gb/s switch-to-switch
  - 1us application latency
  - Most aggressive roadmap in the industry*
- **Reliable with congestion management**
- **Efficient**
  - RDMA and Transport Offload
  - Kernel bypass
  - CPU focuses on application processing
- **Scalable for Petascale computing & beyond**
- **End-to-end quality of service**
- **Virtualization acceleration**
- **I/O consolidation Including storage**

The InfiniBand Performance Gap is Increasing

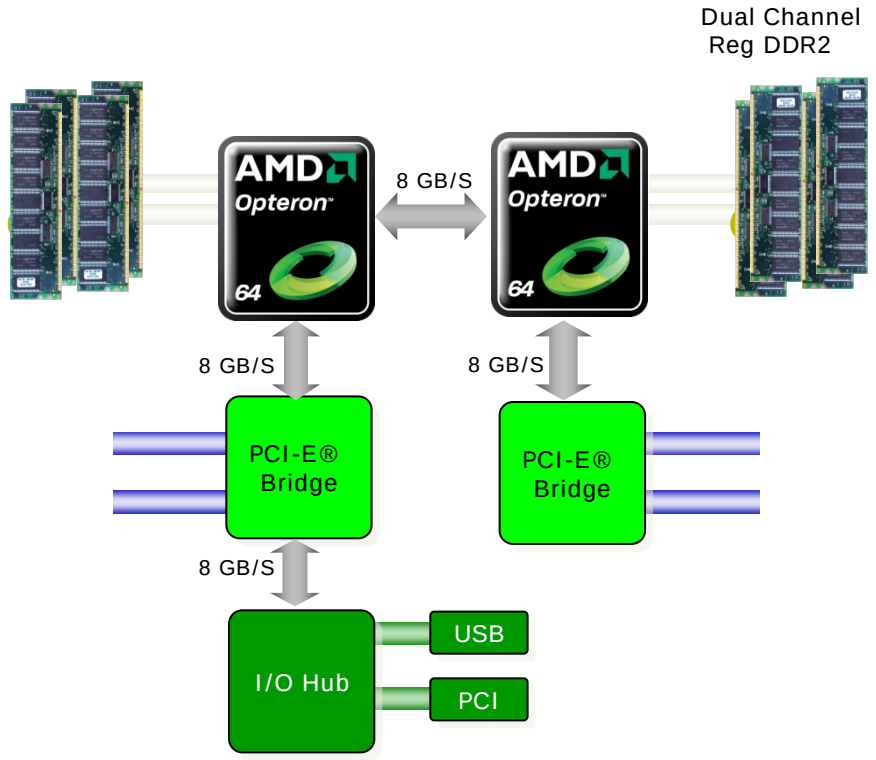


InfiniBand Delivers the Lowest Latency

*** http://www.infinibandta.org/itinfo**

Quad-Core AMD Opteron™ Processor

- **Performance**
  - Quad-Core
    - Enhanced CPU IPC
    - 4x 512K L2 cache
    - 6MB L3 Cache
  - Direct Connect Architecture
    - HyperTransport™ Technology
    - Up to 24 GB/s peak per processor
  - Floating Point
    - 128-bit FPU per core
    - 4 FLOPS/clk peak per core
  - Integrated Memory Controller
    - Up to 12.8 GB/s
    - DDR2-800 MHz or DDR2-667 MHz
- **Scalability**
  - 48-bit Physical Addressing
- **Compatibility**
  - Same power/thermal envelopes as 2nd / 3rd generation AMD Opteron™ processor

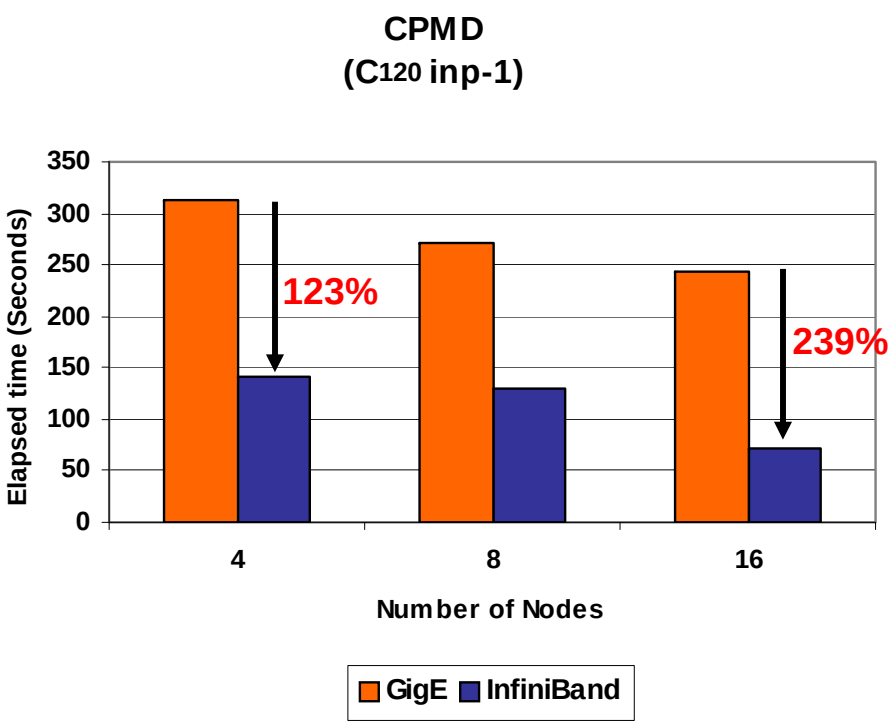


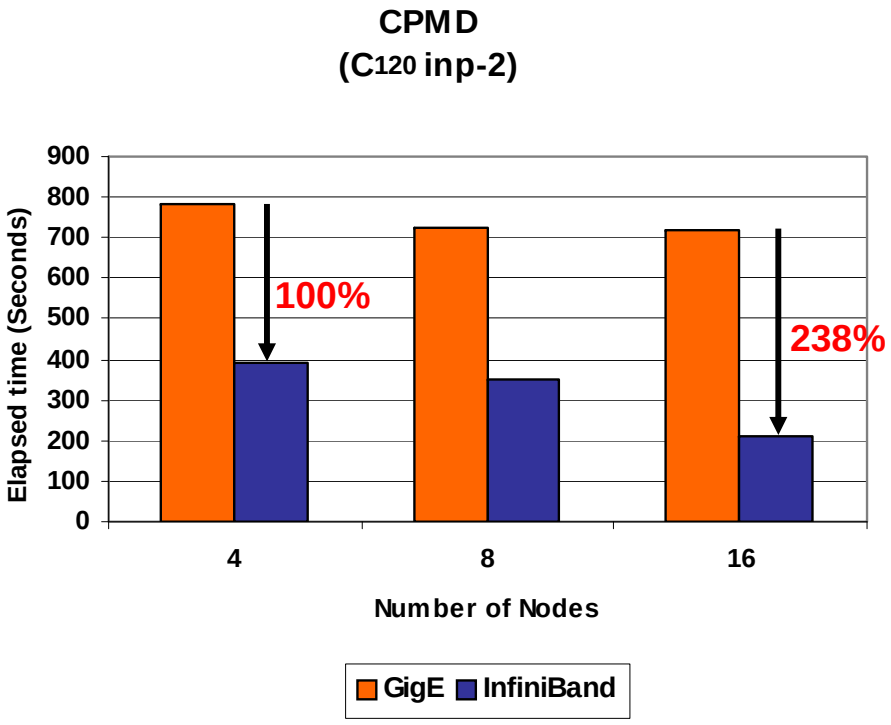
Dell PowerEdge Servers helping Simplify IT

- **System Structure and Sizing Guidelines**
  - 24-node cluster build with Dell PowerEdge™ SC 1435 Servers
  - Servers optimized for High Performance Computing environments
  - Building Block Foundations for best price/performance and performance/watt
- **Dell HPC Solutions**
  - Scalable Architectures for High Performance and Productivity
  - Dell's comprehensive HPC services help manage the lifecycle requirements.
  - Integrated, Tested and Validated Architectures
- **Workload Modeling**
  - Optimized System Size, Configuration and Workloads
  - Test-bed Benchmarks
  - ISV Applications Characterization
  - Best Practices & Usage Analysis

CPMD Results – C_{120}

- **Case C_{120} - 120 carbon atoms**
  - inp-1: Wavefunction optimization
  - inp-2: Molecular dynamics simulation
- **InfiniBand outperforms GigE in every cluster size**



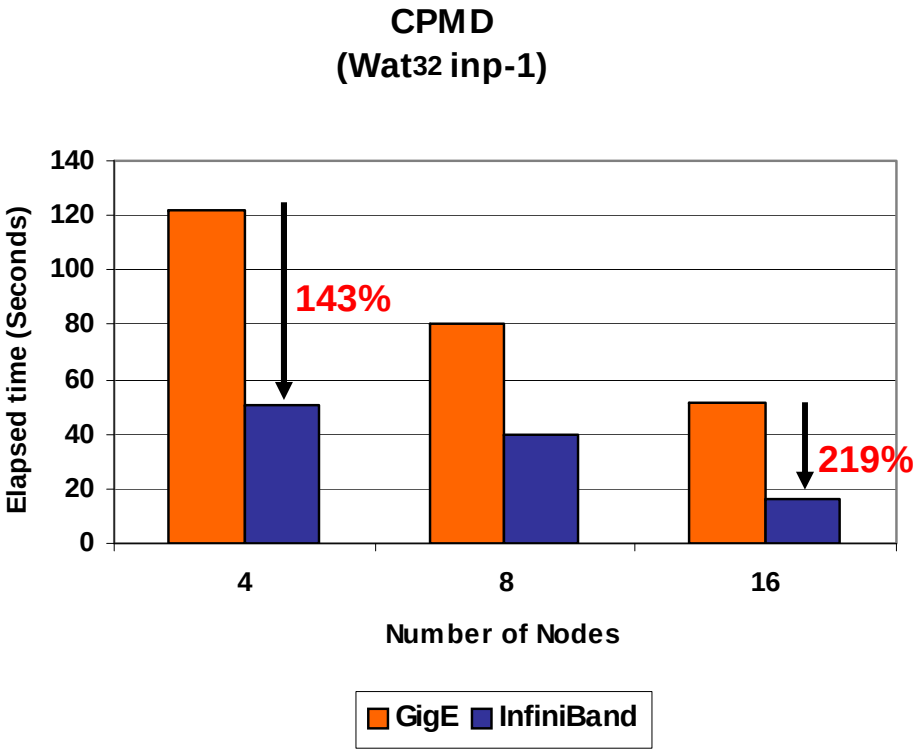


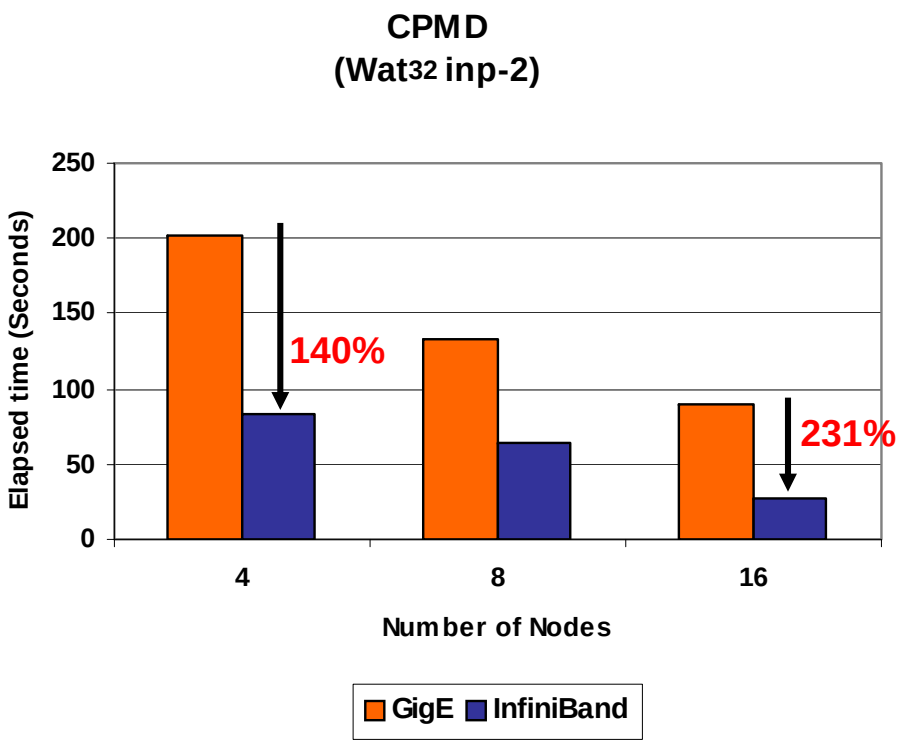
Lower is better

Platform MPI

CPMD Results – Wat32

- **Case Wat32 - a cluster of 32 water molecules**
  - inp-1: Wavefunction optimization
  - inp-2: Molecular dynamics simulation
- **Similar to case C120, InfiniBand enables superior scaling compared to GigE**



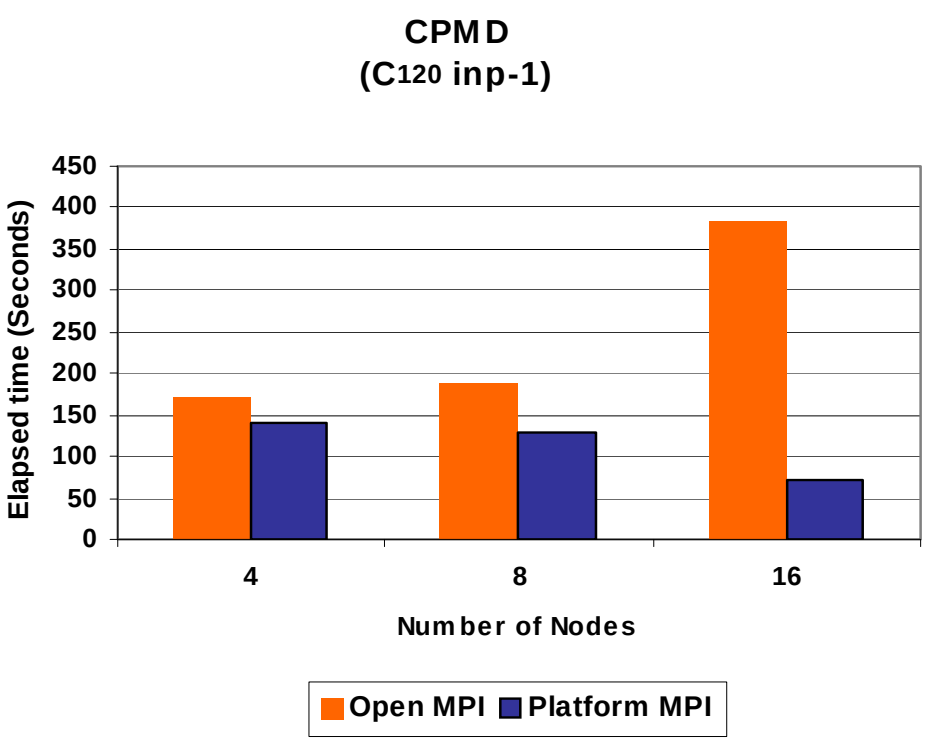


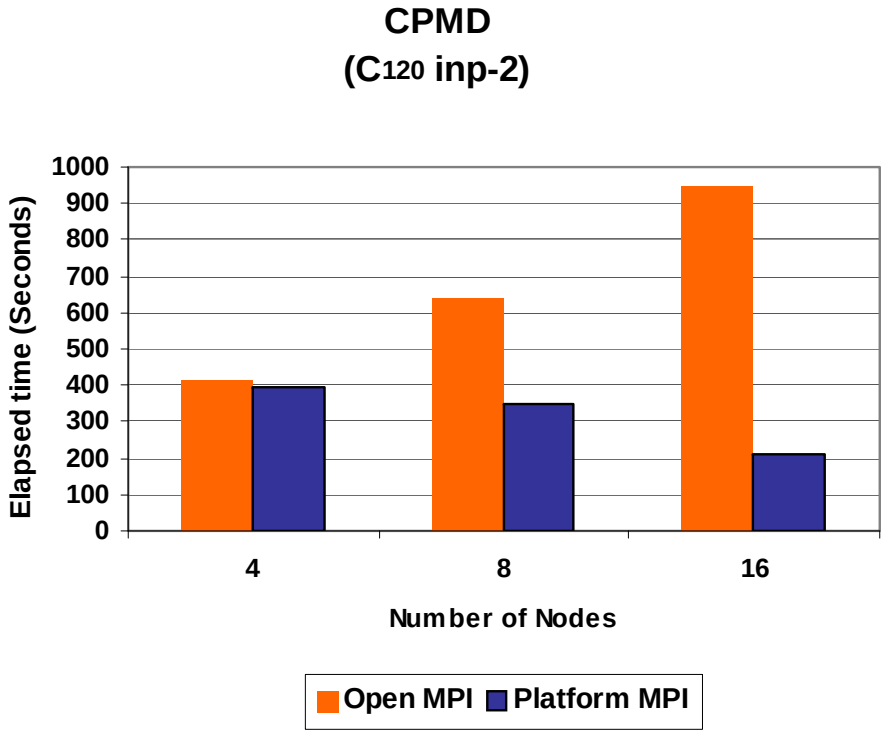
Lower is better

Platform MPI

CPMD Performance Comparison - MPI

- **Platform MPI shows performance advantage over Open MPI**



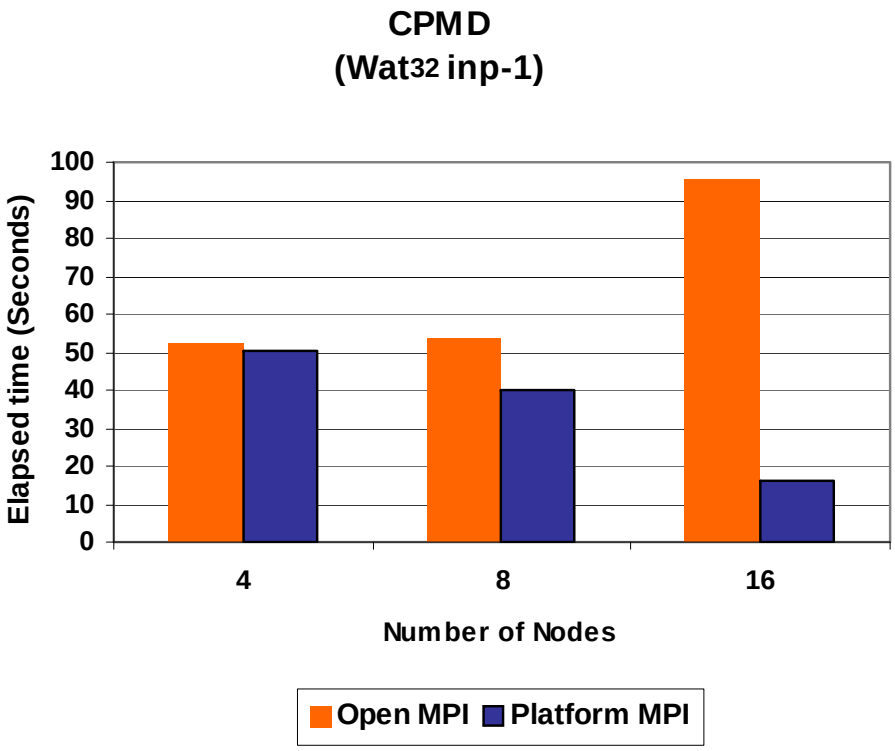


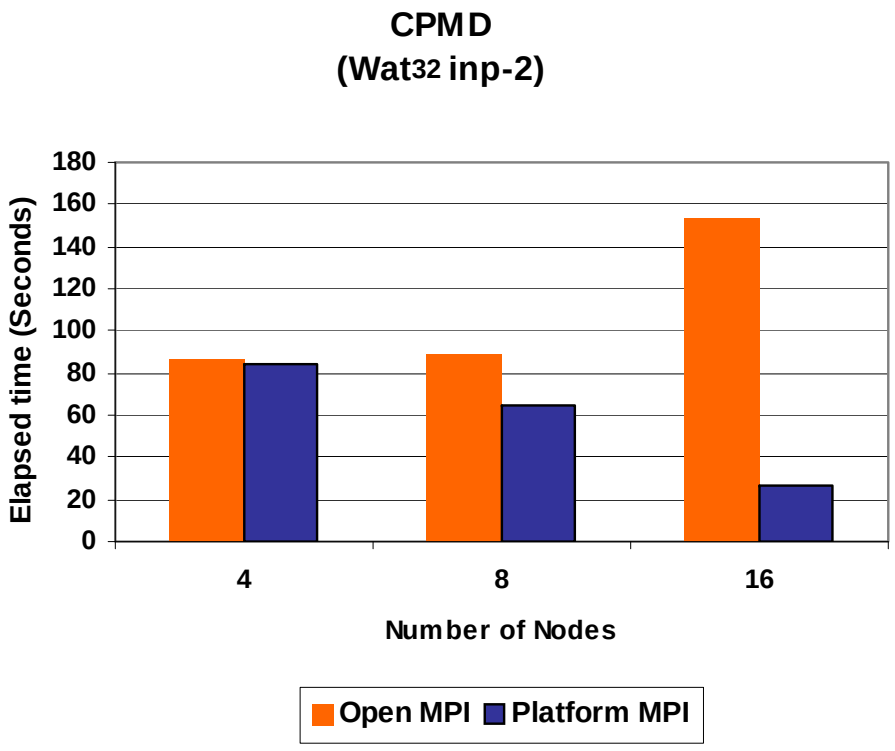
Lower is better

These results are based on InfiniBand

CPMD Performance Comparison - MPI

- **Platform MPI shows better scalability over Open MPI**



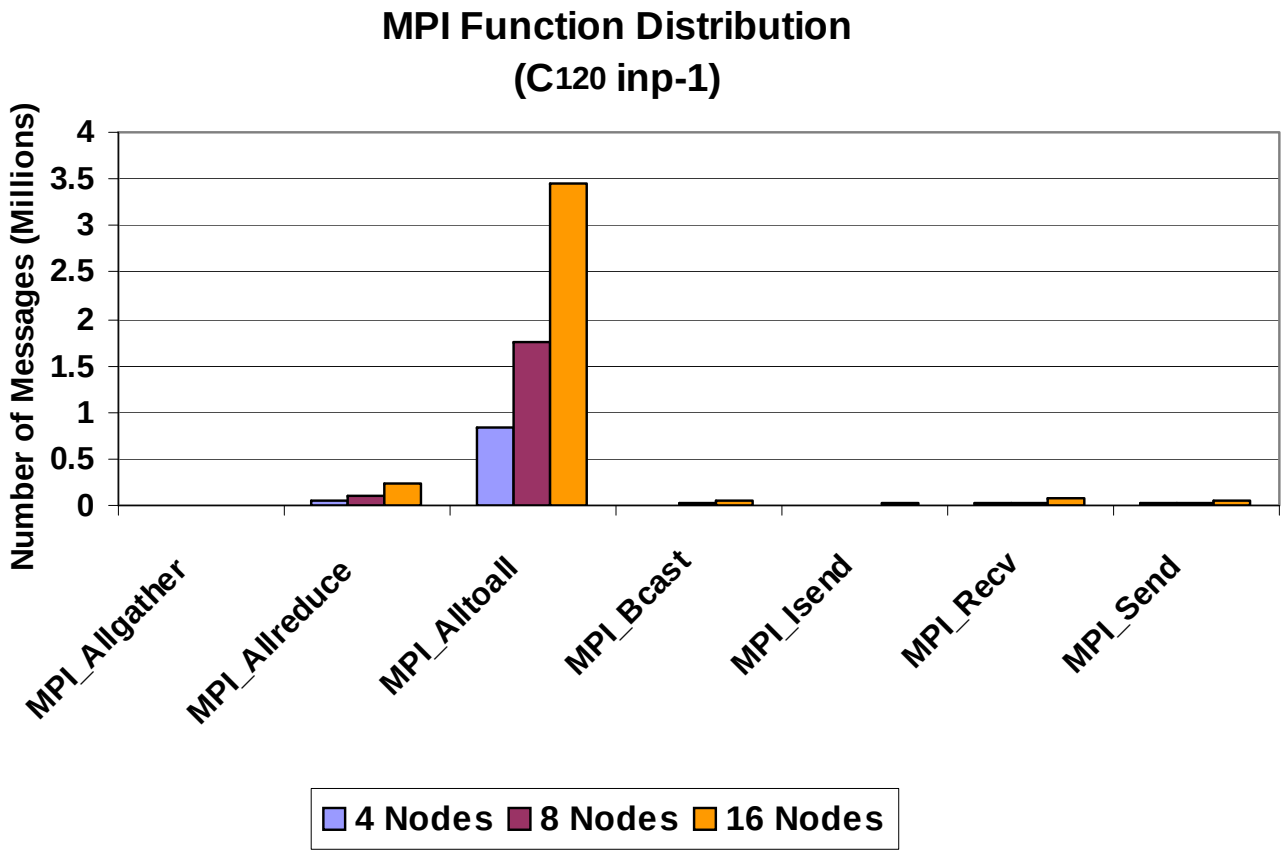


Lower is better

These results are based on InfiniBand

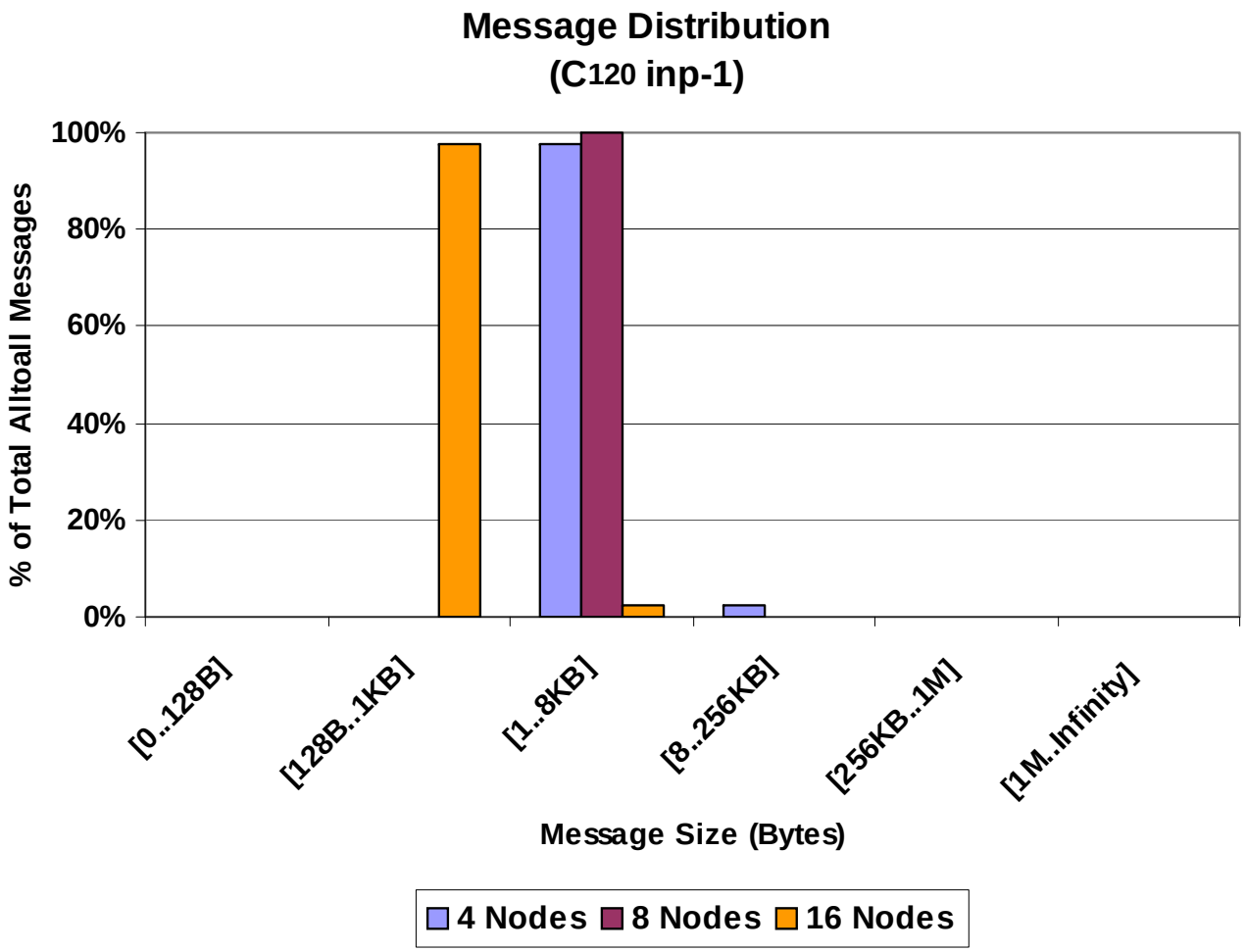
MPI Collectives in CPMD

- **MPI_AlltoAll is the key collective function in CPMD**
  - Number of AlltoAll messages increases dramatically with cluster size



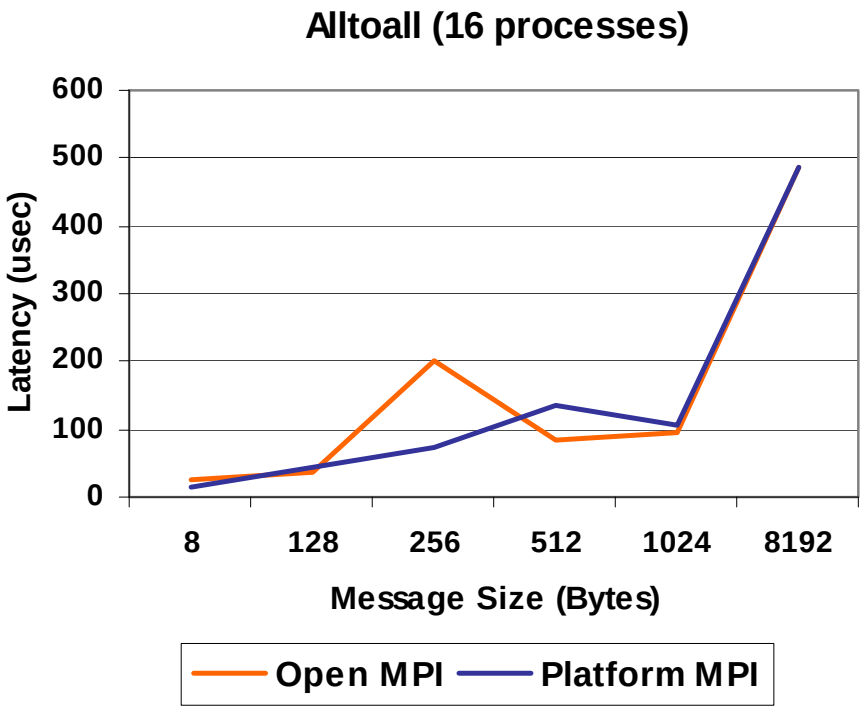
MPI Message Size Distribution

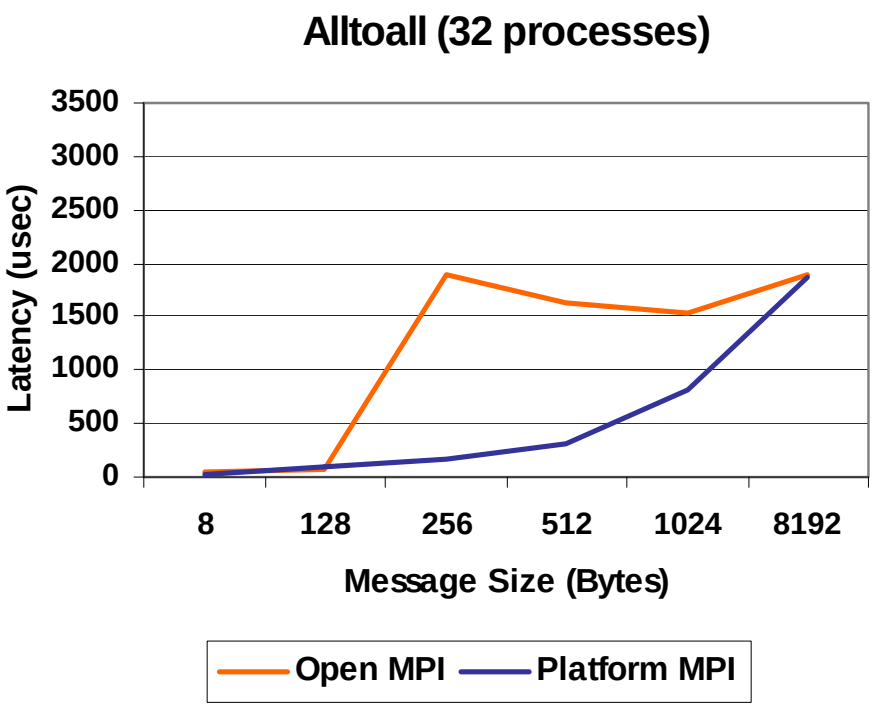
- **Majority messages are medium size**



MPI Alltoall Performance Comparison

- **Platform MPI has better MPI Alltoall performance versus Open MPI**



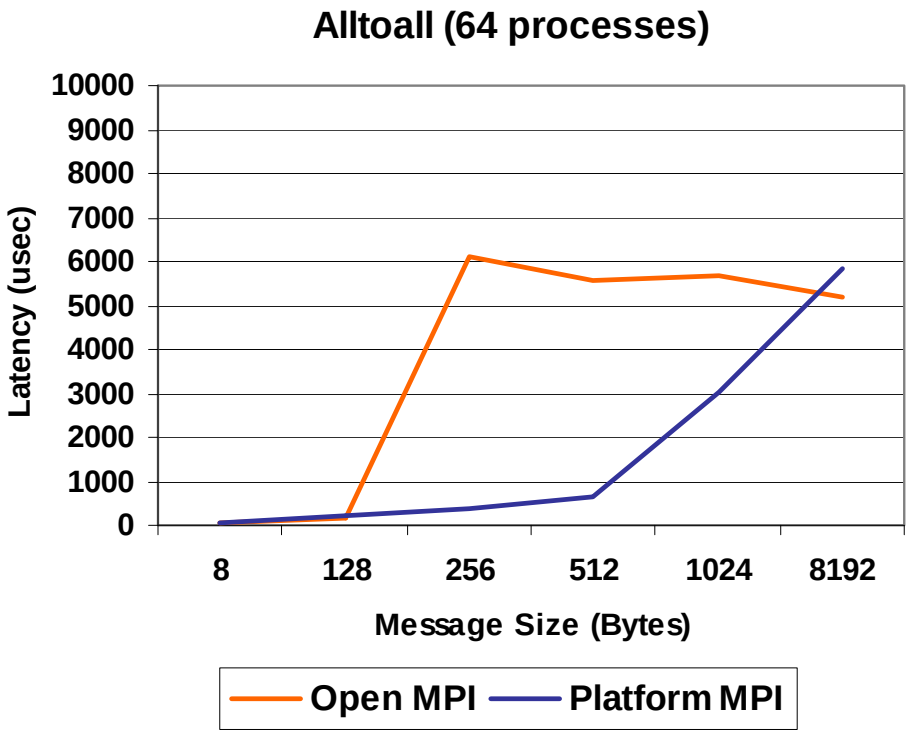


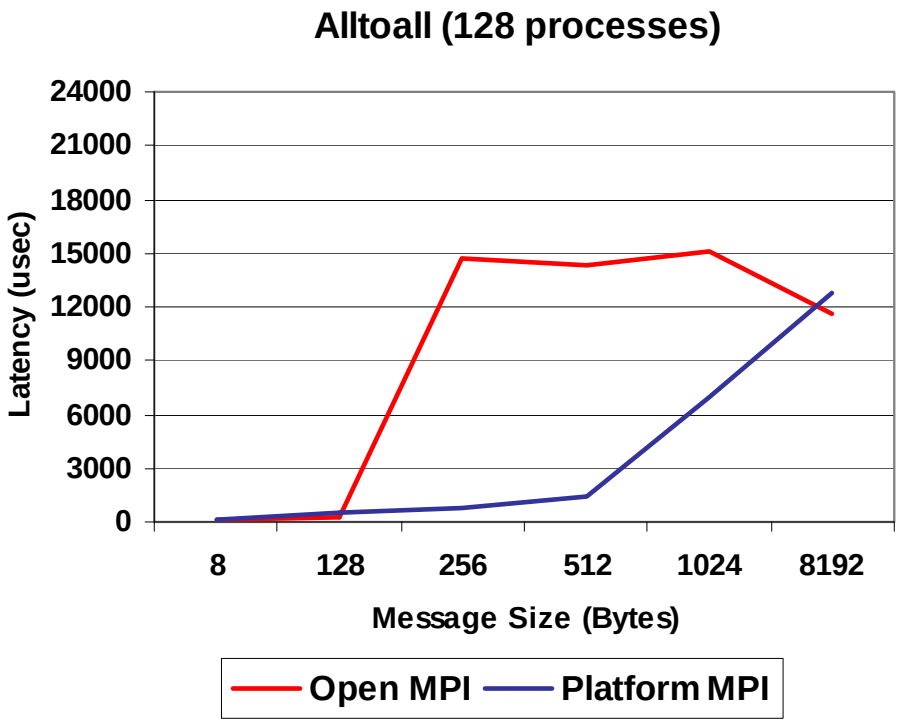
Lower is better

These results are based on InfiniBand

MPI Alltoall Performance Comparison

- **Performance advantage of Platform MPI increase with node count**





Lower is better

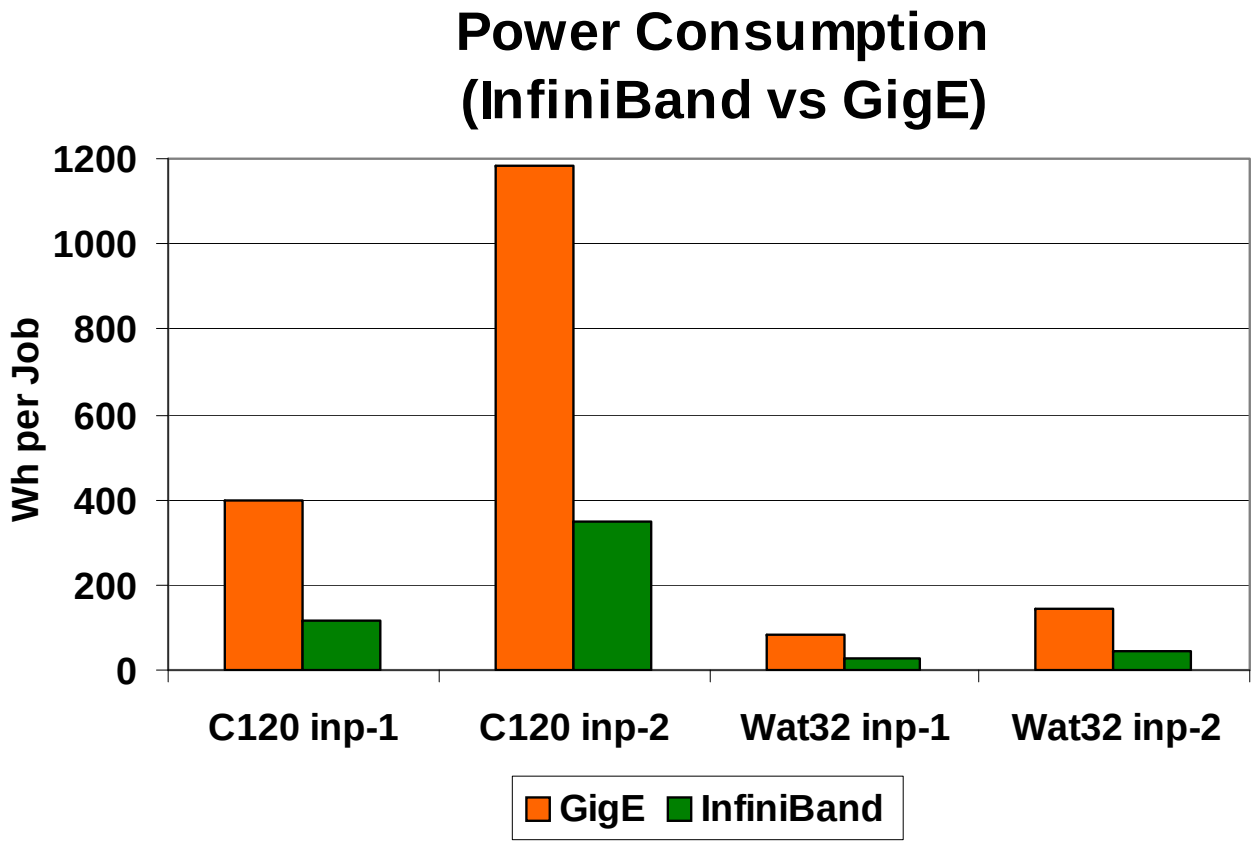
These results are based on InfiniBand

CPMD Profiling Summary

- **CPMD was profiled to understand its communication pattern**
- **Message size distribution**
  - Most used message are ranging from 128B to 8KB
  - Percentage of smaller messages (128B to 1K) increases with cluster size
- **MPI function in CPMD**
  - MPI Alltoall is the dominant MPI function
  - Mostly effects overall CPMD performance

Power Consumption Comparison - Interconnect

- **InfiniBand enables power efficient simulations**
- **Reducing system power/job consumption by up to 70%**



16 nodes cluster

Summary of CPMD

- **CPMD relies on interconnect with low latency and high throughput**
  - Most messages transferred between processes are 128Bytes - 8KB messages
  - Number of messages scales up quickly as number of processes increases
- **InfiniBand enables CPMD performance scalability**
  - InfiniBand performance is up to 239% higher than GigE
- **CPMD attains better performance with Platform MPI versus Open MPI**
  - MPI AlltoAll performance capability leads to better CPMD performance
- **Power Efficiency**
  - InfiniBand enables green computing by dramatically reduce the power per job
  - More than 70% reduction in power/job

Thank You

HPC Advisory Council

HPC@mellanox.com

All trademarks are property of their respective owners. All information is provided "As-Is" without any kind of warranty. The HPC Advisory Council makes no representation to the accuracy and completeness of the information contained herein. HPC Advisory Council Mellanox undertakes no duty and assumes no obligation to update or correct any information presented herein