



# STAR-CCM+ and STAR-CD Performance Benchmark and Profiling

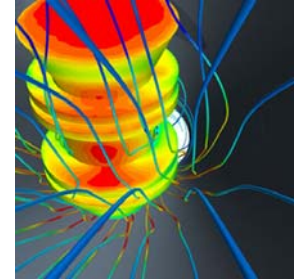
March 2009



- **The following research was performed under the HPC Advisory Council activities**
  - Participating vendors: AMD, Dell, Mellanox
  - Compute resource - HPC Advisory Council Cluster Center
- **The participating members would like to thank CD-adapco for their support and guidelines**
- **For more info please refer to**
  - [www.mellanox.com](http://www.mellanox.com), [www.dell.com/hpc](http://www.dell.com/hpc), [www.amd.com](http://www.amd.com)

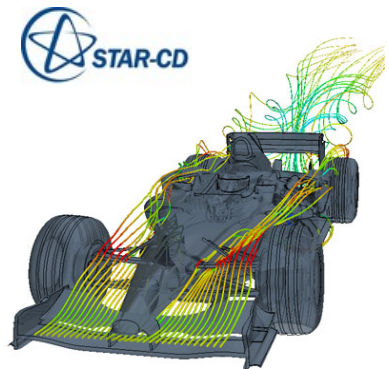
- **STAR-CCM+**

- An engineering process-oriented CFD tool
- Client-server architecture, object-oriented programming
- Delivers the entire CFD process in a single integrated software environment



- **STAR-CD**

- An integrated platform for multi-physics simulations
- A long established platform for industrial CFD simulation
- Bridging the gap between CFD and structural-mechanics



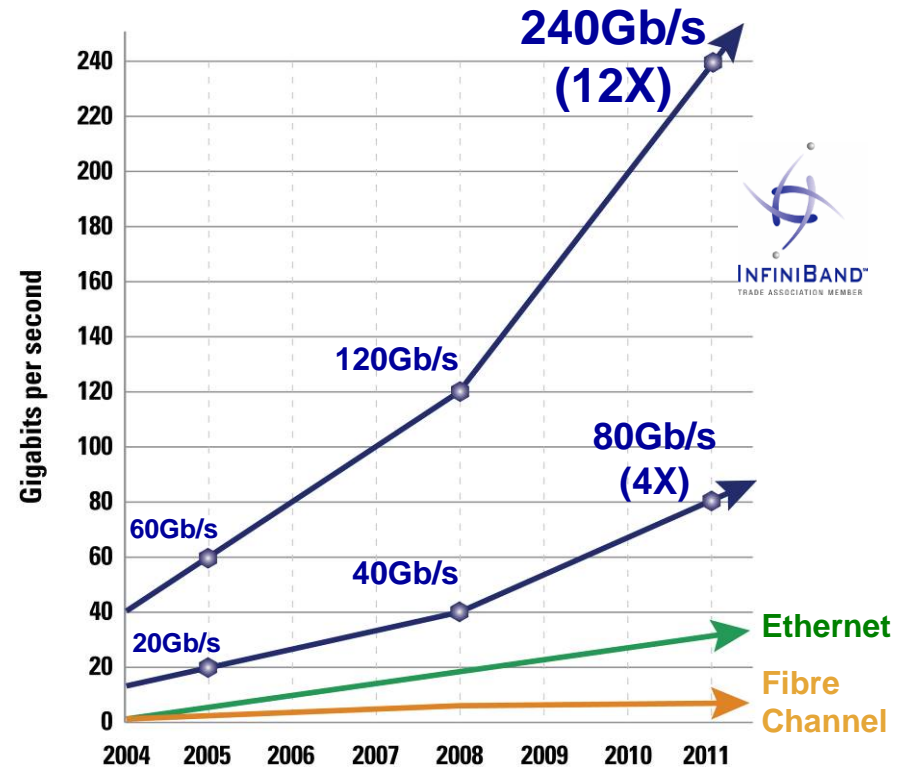
- **Developed by CD-adapco**

- **The presented research was done to provide best practices**
  - STAR-CCM+ and STAR-CD performance benchmarking
  - Interconnect performance comparisons
  - Ways to increase STAR-CCM+ and STAR-CD productivity
  - Understanding STAR-CD communication patterns
  - MPI libraries comparisons

- **Dell™ PowerEdge™ SC 1435 24-node cluster**
- **Quad-Core AMD Opteron™ 2382 (“Shanghai”) CPUs**
- **Mellanox® InfiniBand ConnectX® 20Gb/s (DDR) HCAs**
- **Mellanox® InfiniBand DDR Switch**
- **Memory: 16GB memory, DDR2 800MHz per node**
- **OS: RHEL5U2, OFED 1.4 InfiniBand SW stack**
- **MPI: Platform MPI 5.6.5, HP-MPI 2.3**
- **Application: STAR-CCM+ Version 3.06, STAR-CD Version 4.08**
- **Benchmark Workload**
  - **STAR-CD: A-Class (Turbulent Flow around A-Class Car)**
  - **STAR-CCM+: Auto Aerodynamics test**

- **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

# 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/clock peak per core

- Integrated Memory Controller

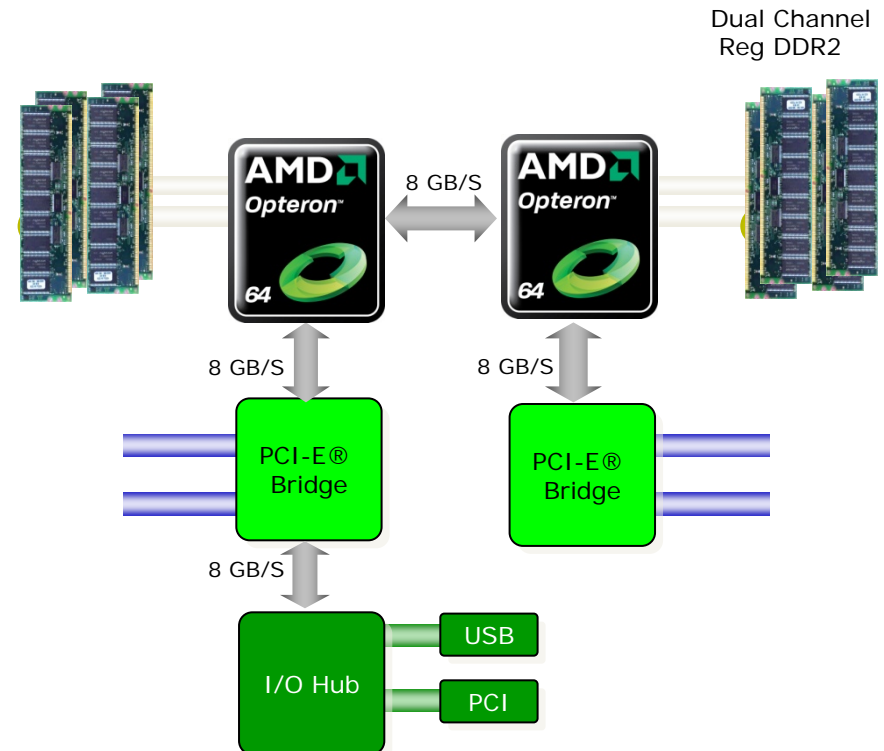
- Up to 12.8 GB/s
- DDR2-800 MHz or DDR2-667 MHz

- **Scalability**

- 48-bit Physical Addressing

- **Compatibility**

- Same power/thermal envelopes as 2<sup>nd</sup> / 3<sup>rd</sup> generation AMD Opteron™ processor



- **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



- **Input Dataset**

- Auto Aerodynamics test

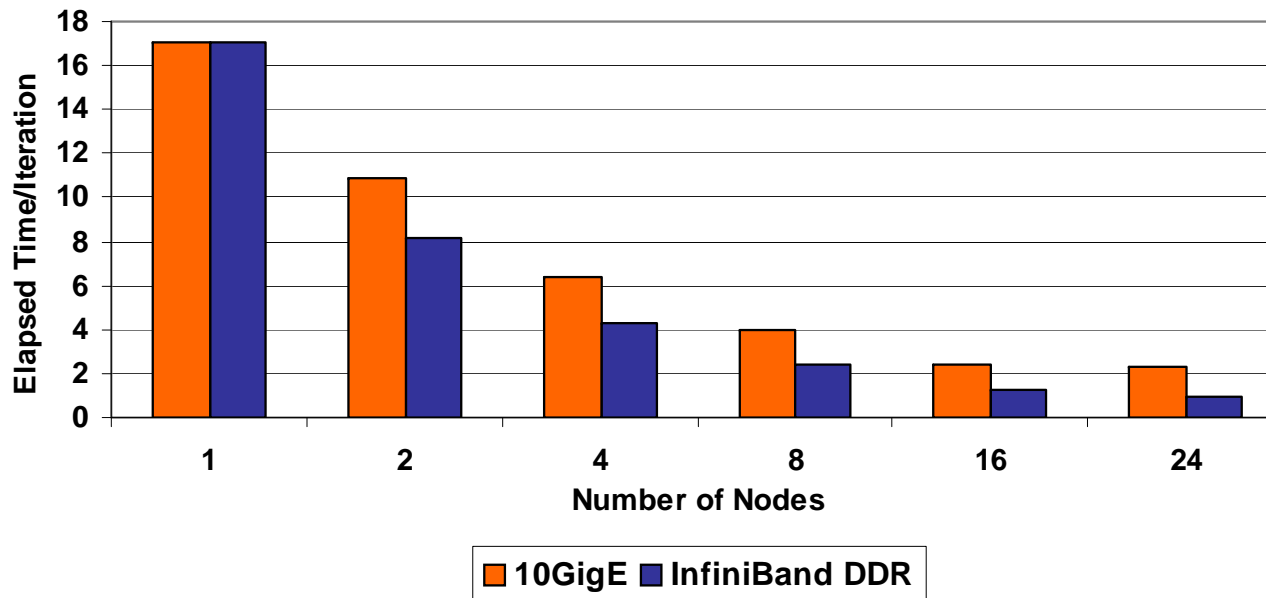
- **InfiniBand DDR delivers higher performance and scalability**

- For any cluster size

- Up to 136% faster run time



**STAR-CCM+ Benchmark Results**  
(Auto Aerodynamics Test)



*Lower is better*

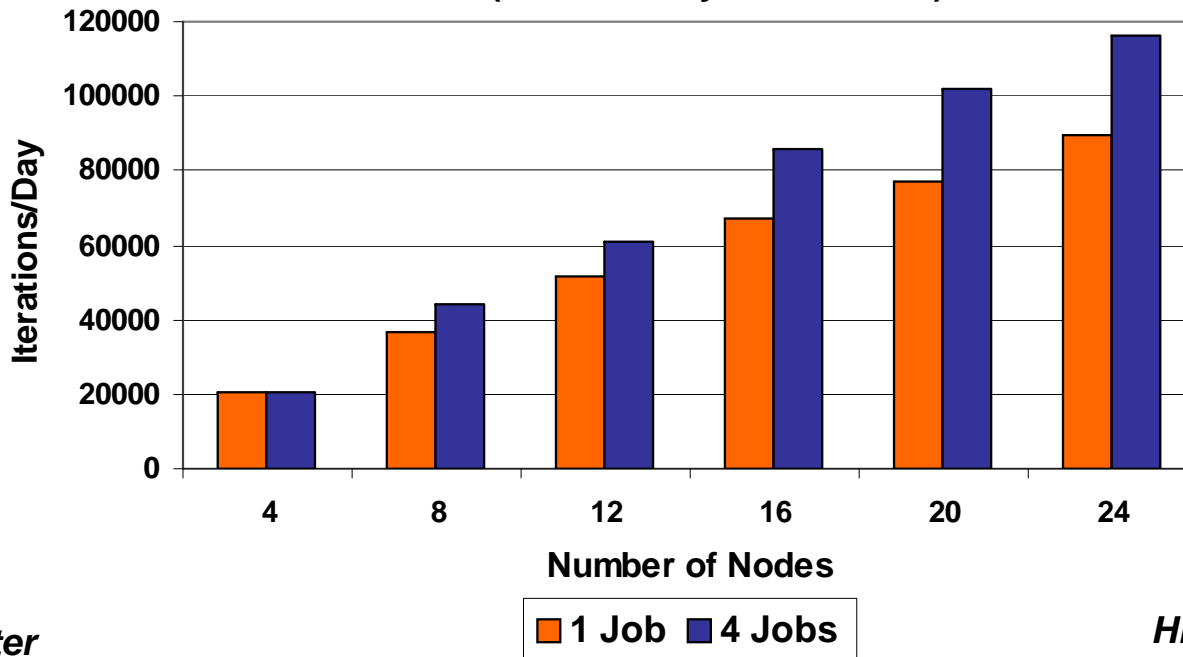
**HP MPI**

# STAR-CCM+ Productivity Results

- **Two cases are presented**
  - Single job over the entire systems
  - Four jobs, each on two cores per server
- **Productivity increases by allowing multiple jobs to run simultaneously**
  - Up to 30% increase in the system productivity



**STAR-CCM+ Productivity**  
(Auto Aerodynamics Test)



*Higher is better*

*HP MPI over InfiniBand*

# STAR-CCM+ Productivity Results - Interconnect

- **Test case**

- Four jobs, each on two cores per server

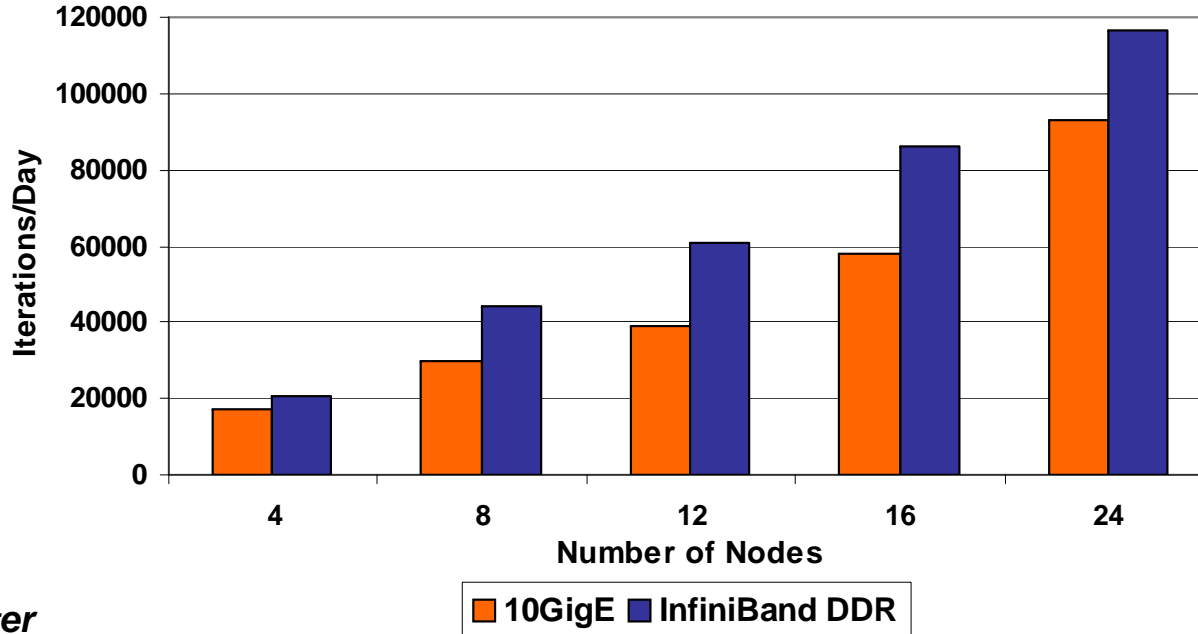
- **InfiniBand DDR provides higher productivity compared to 10GigE**

- Up to 25% more iterations per day

- **InfiniBand maintains consistent scalability as cluster size increases**



**STAR-CCM+ Productivity**  
(Auto Aerodynamics Test)



*Higher is better*

**HP MPI**



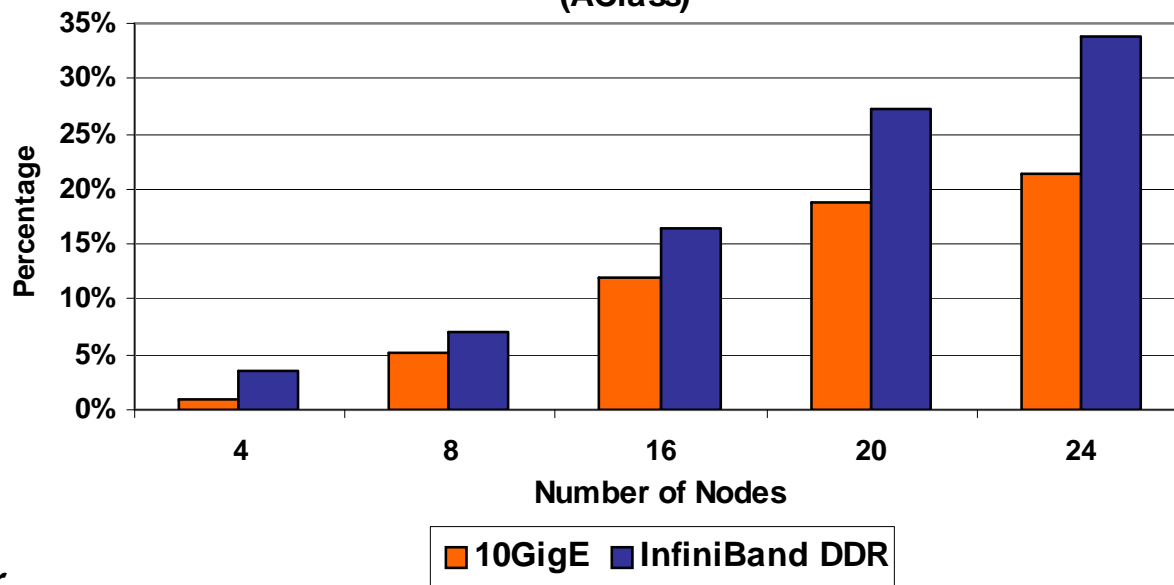
- **Test case**

- Single job over the entire systems
- Input Dataset (A-Class)

- **InfiniBand DDR enhances performance and scalability**

- Up to 34% and 10% more jobs/day compared to GigE and 10GigE respectively
- Performance advantage of InfiniBand increases as cluster size scales

**STAR-CD Performance Advantage of  
InfiniBand and 10GigE over GigE  
(A-Class)**



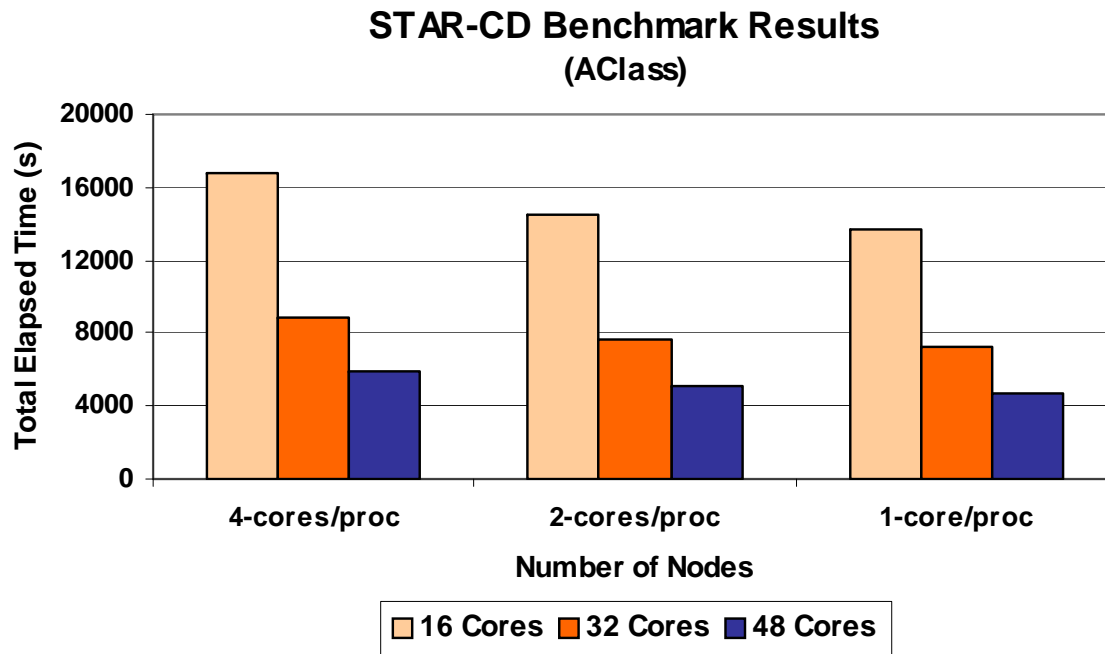
*Higher is better*

*Platform MPI*

# Maximize STAR-CD Performance per Core



- **Test case**
  - Single job over the entire systems
  - Using one, two or four cores in each quad-core AMD processor
    - Remaining cores kept idle
- **Using partial cores per simulation improves single job run time**
- **It is recommended to have multiple simulations simultaneously**
  - For maximum productivity (slide 10)

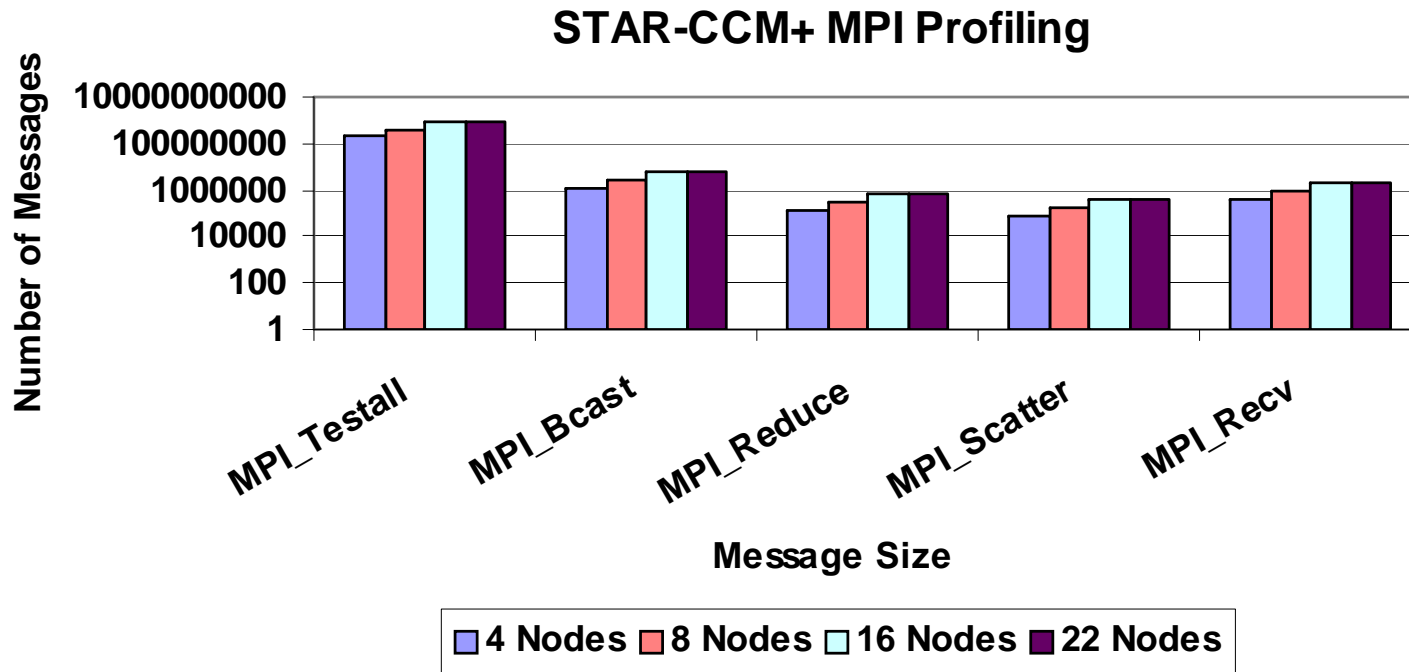


*Lower is better*

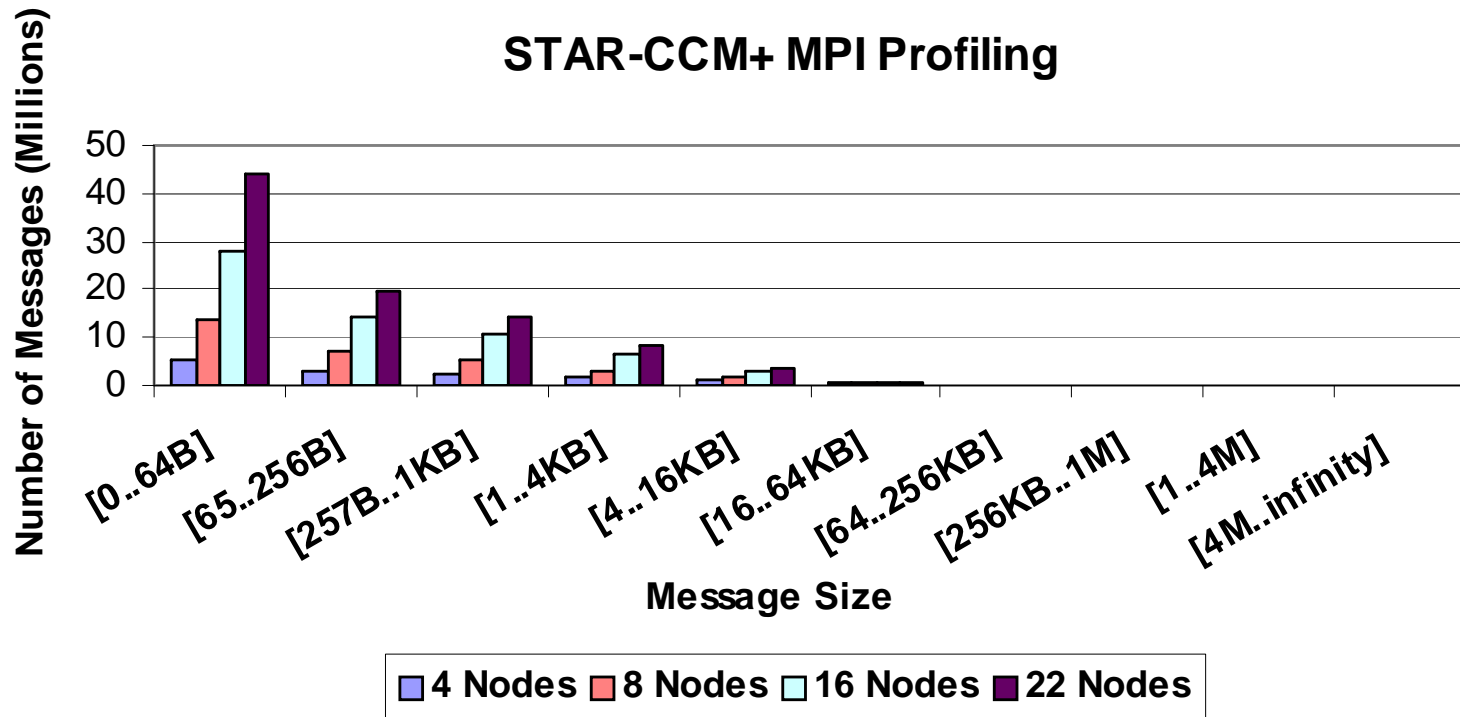
*Platform MPI*

# STAR-CCM+ Profiling – MPI Functions

- MPI\_Testall, MPI\_Bcast, and MPI\_Recv are the mostly used MPI functions



- Most MPI messages are within 4KB in size
- Number of messages increases with cluster size

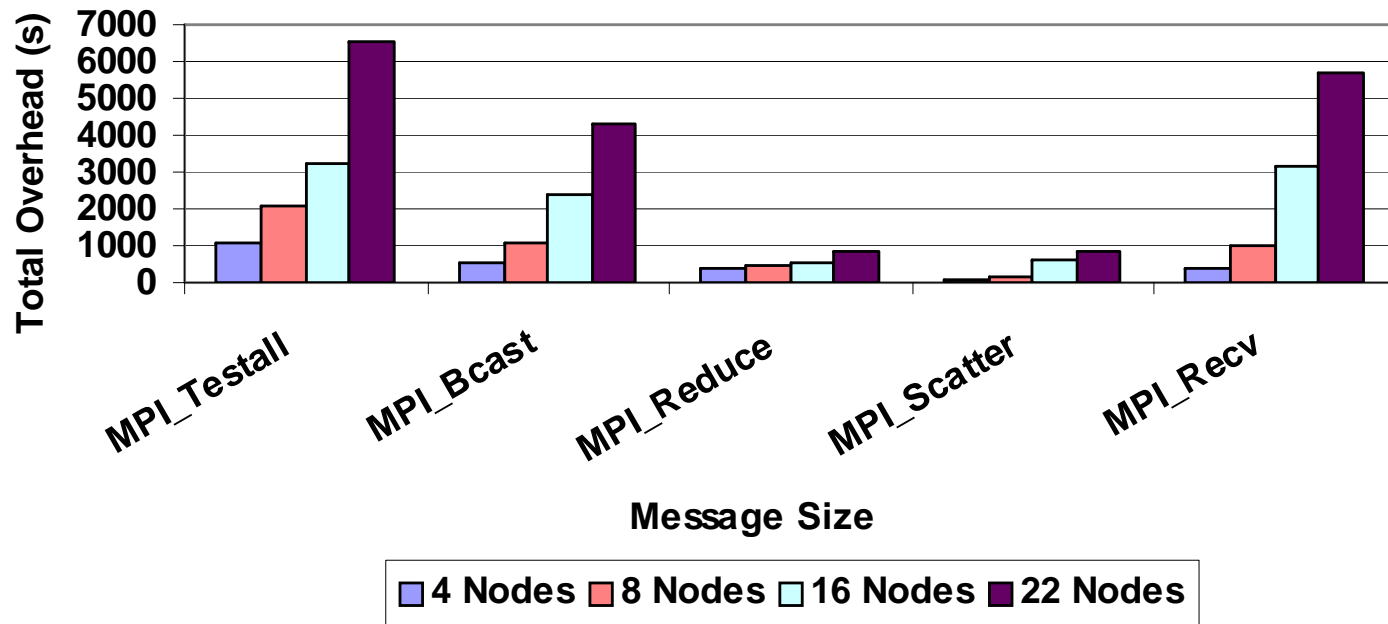


# STAR-CCM+ Profiling – Timing

- MPI\_Testall, MPI\_Bcast, and MPI\_Recv have relatively large overhead
- Overhead increases with cluster size

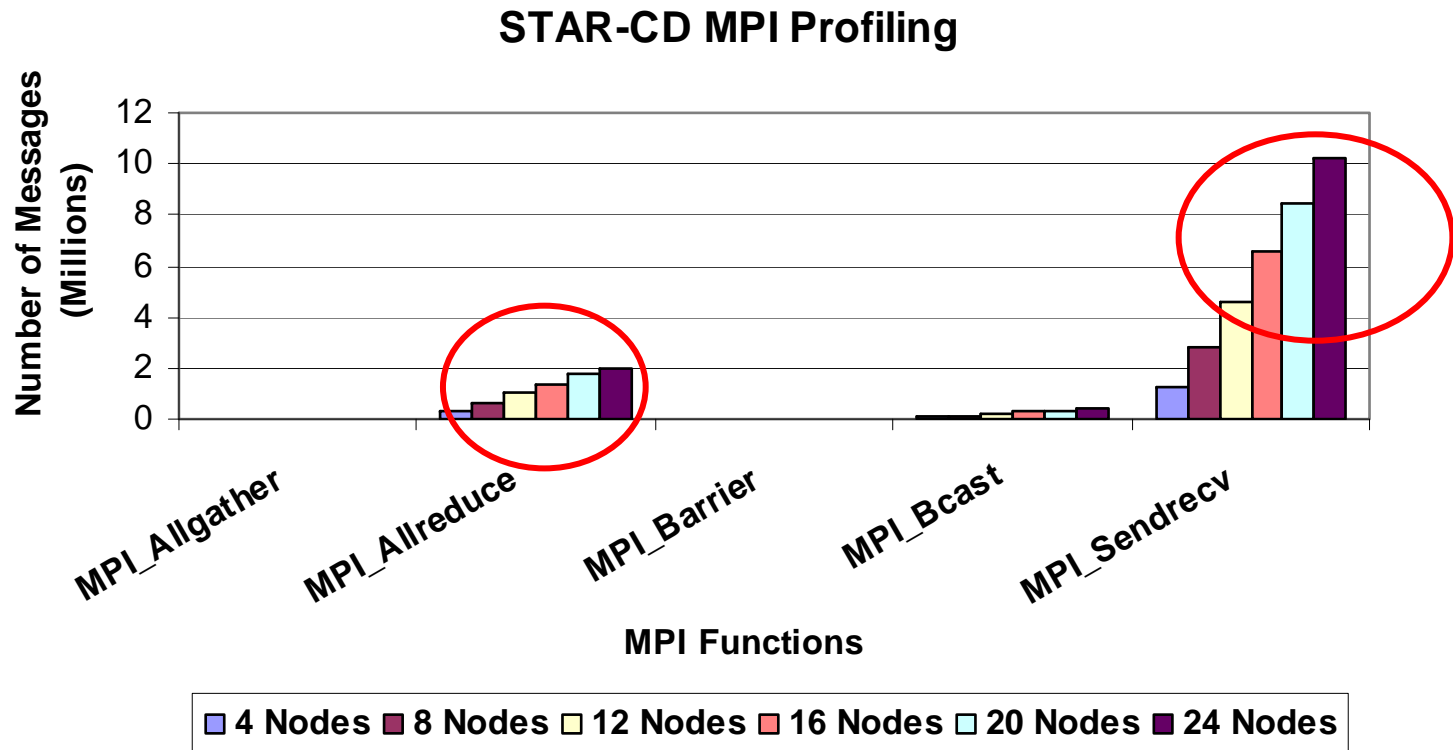


STAR-CCM+ MPI Profiling



- **STAR-CCM+ was profiled to determine networking dependency**
- **Most used message sizes**
  - <4KB messages
  - Number of messages increases with cluster size
- **Interconnects effect to STAR-CD performance**
  - Both interconnect latency and throughput influences STAR-CCM+ performance due to its messaging pattern

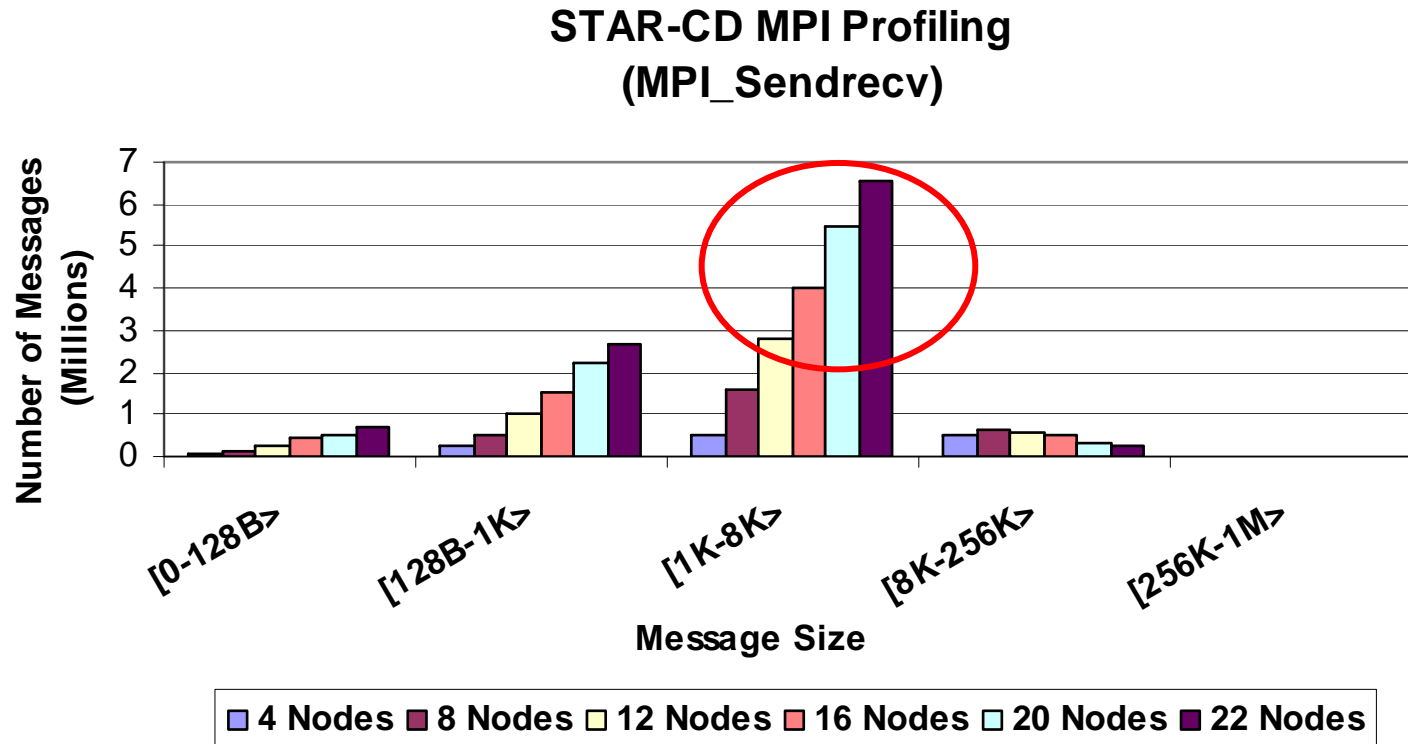
- **MPI\_SendRecv** and **MPI\_Allreduce** are the mostly used MPI functions



# STAR-CD Profiling – Data Transferred



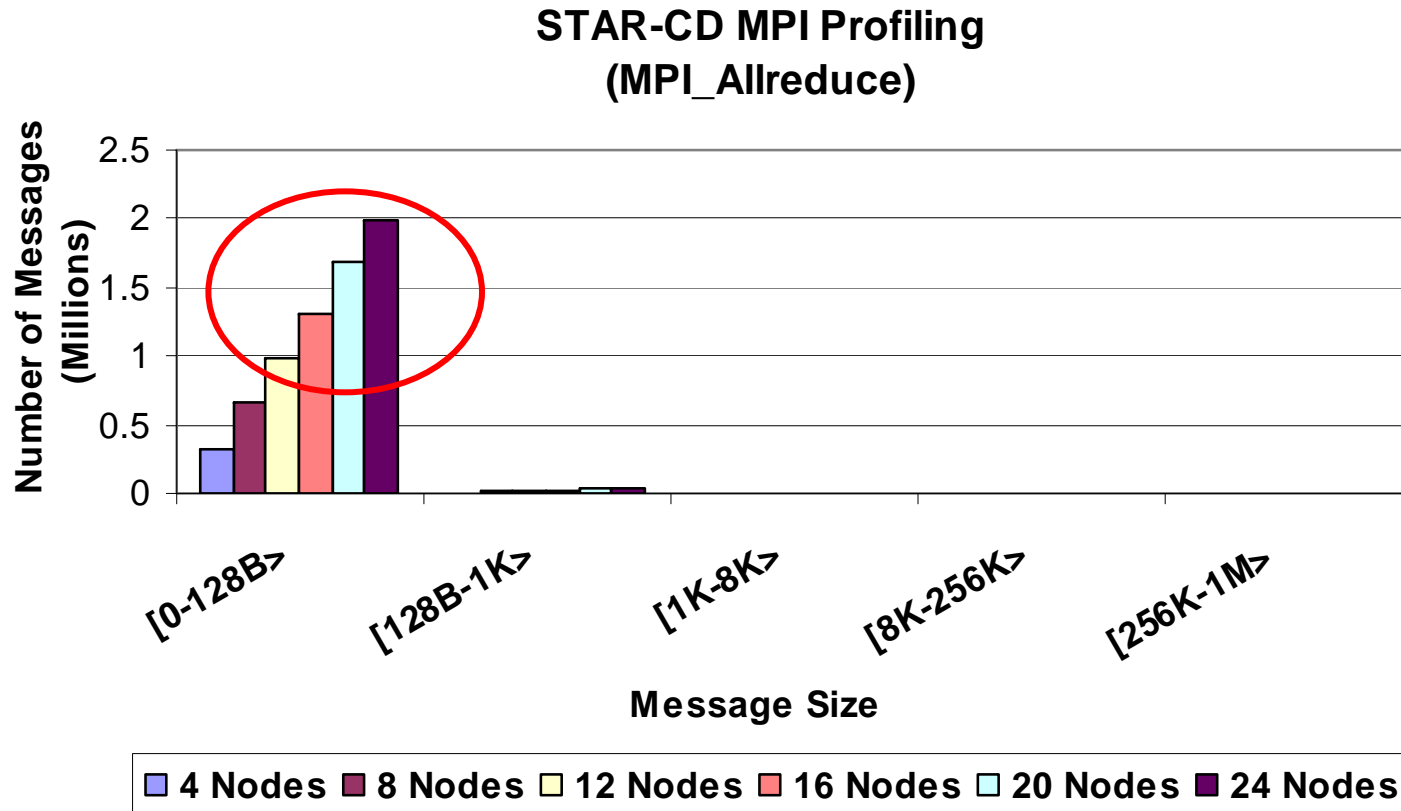
- Most point-to-point MPI messages are within 1KB to 8KB in size
- Number of messages increases with cluster size



# STAR-CD Profiling – Data Transferred

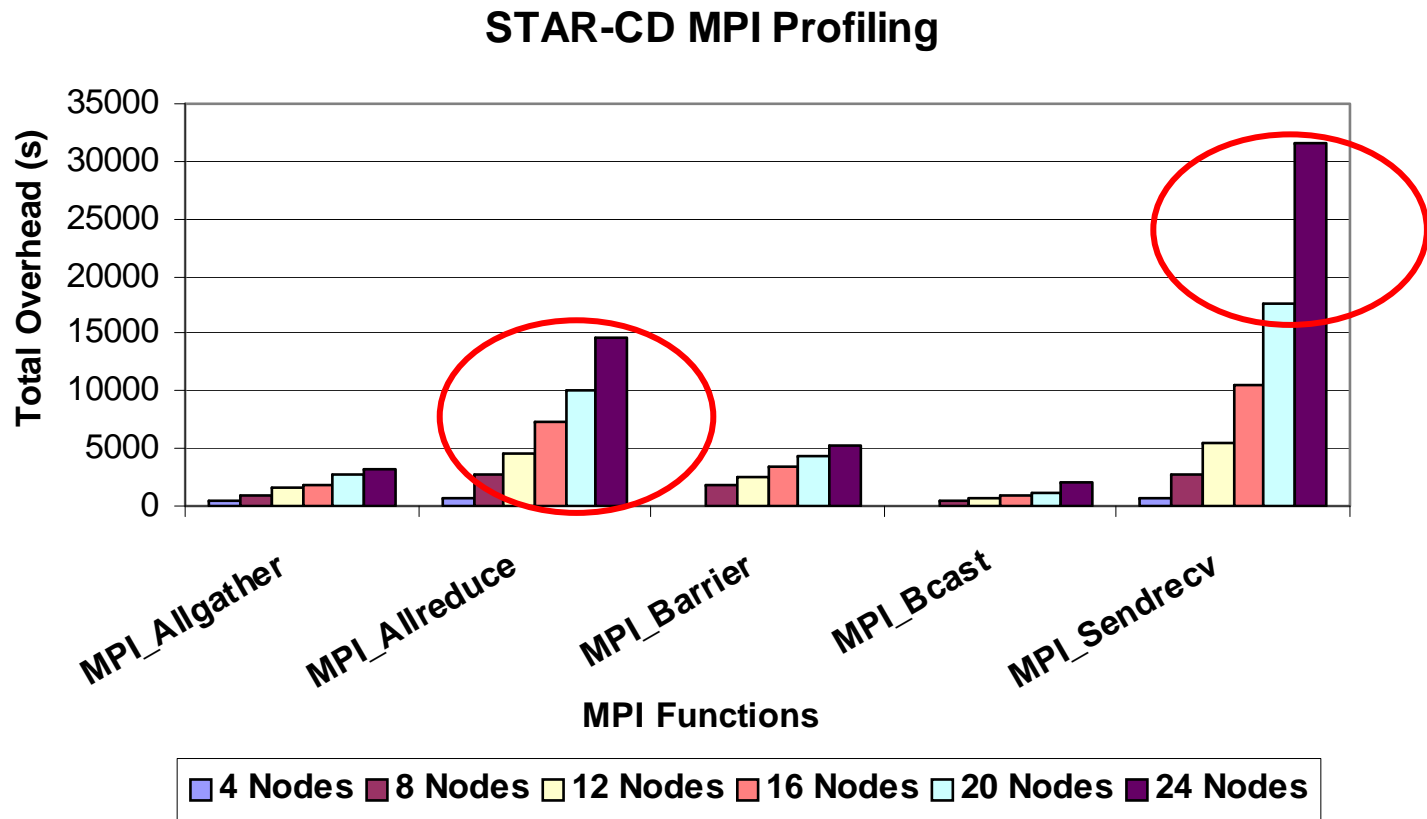


- Most MPI collective messages are smaller than 128Bytes
- Number of messages increases with cluster size



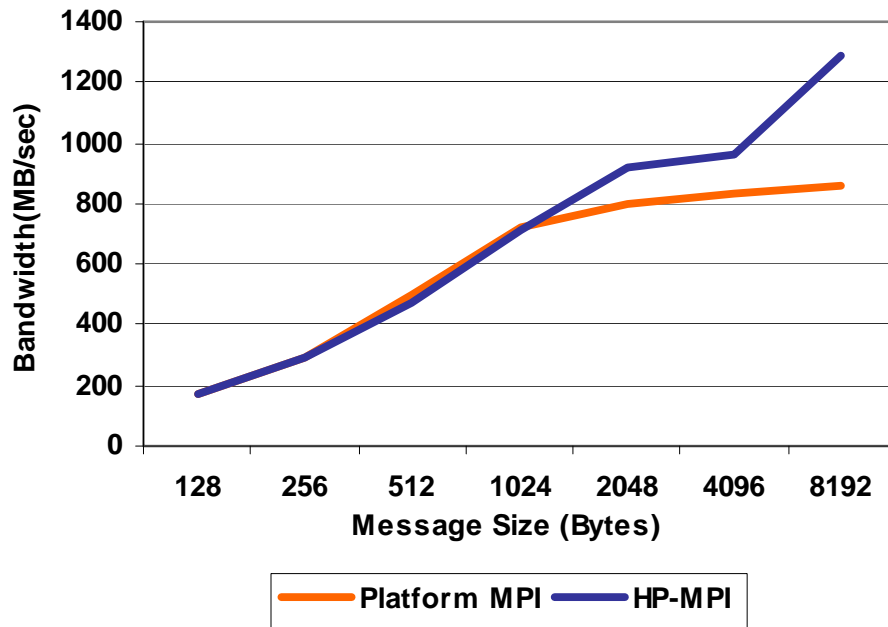
# STAR-CD Profiling – Timing

- **MPI\_Allreduce and MPI\_Sendrecv have large overhead**
- **Overhead increases with cluster size**

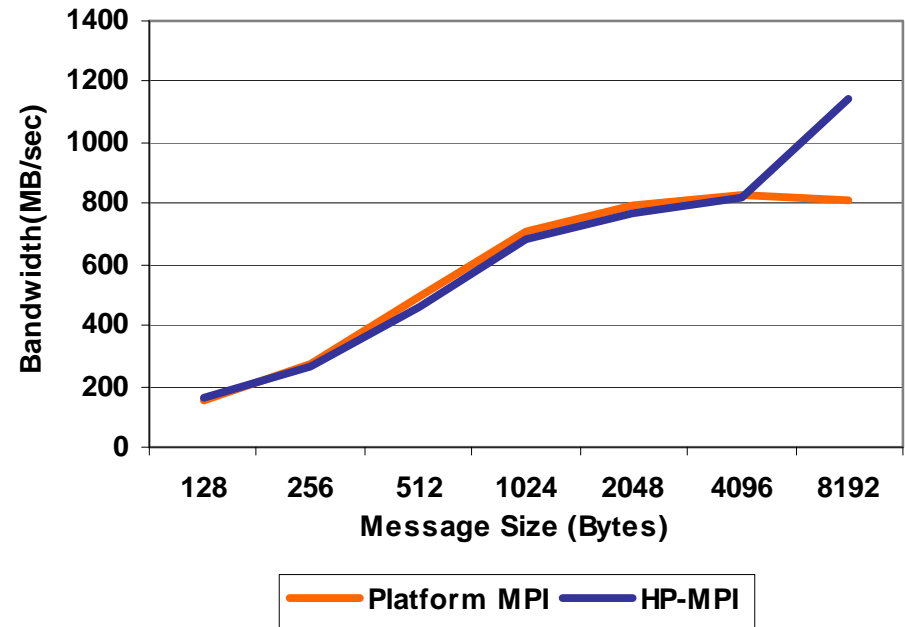


- HP MPI demonstrates better performance for large message

MPI\_Sendrecv  
(32 Processes)



MPI\_Sendrecv  
(192 Processes)



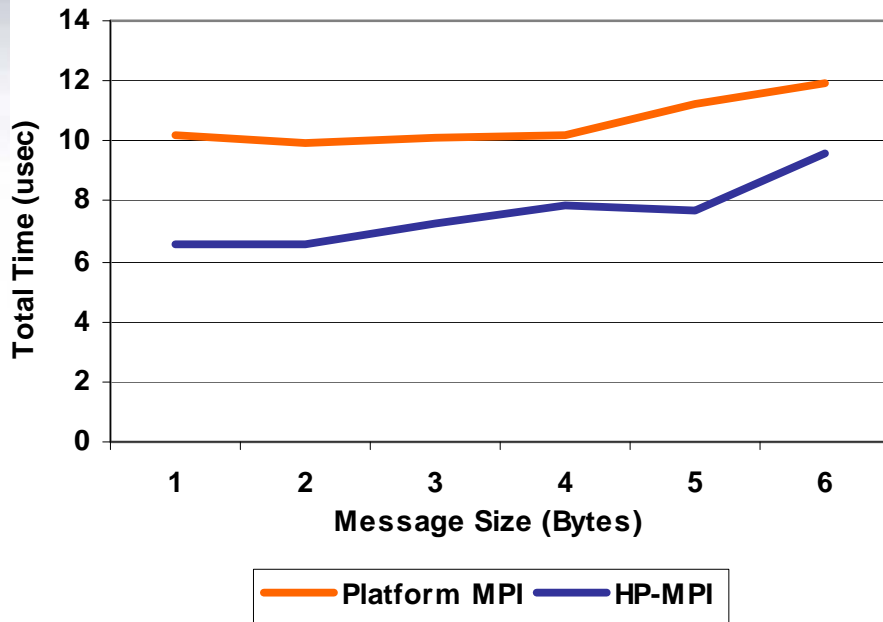
*Higher is better*

*InfiniBand DDR*

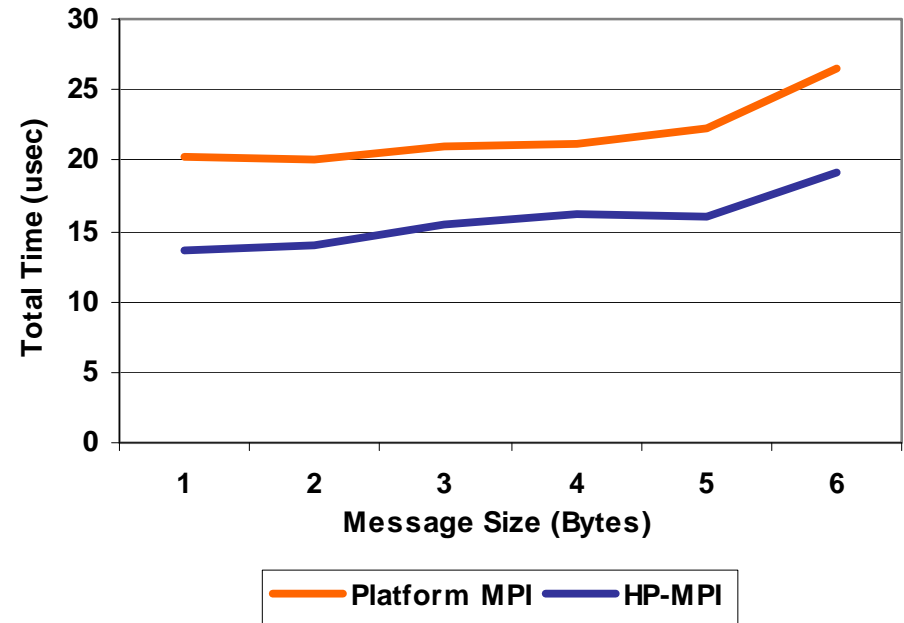
# MPI Performance Comparison – MPI\_Allreduce

- HP MPI demonstrates lower MPI\_Allreduce runtime for small messages

MPI\_Allreduce  
(32 Processes)



MPI\_Allreduce  
(192 Processes)

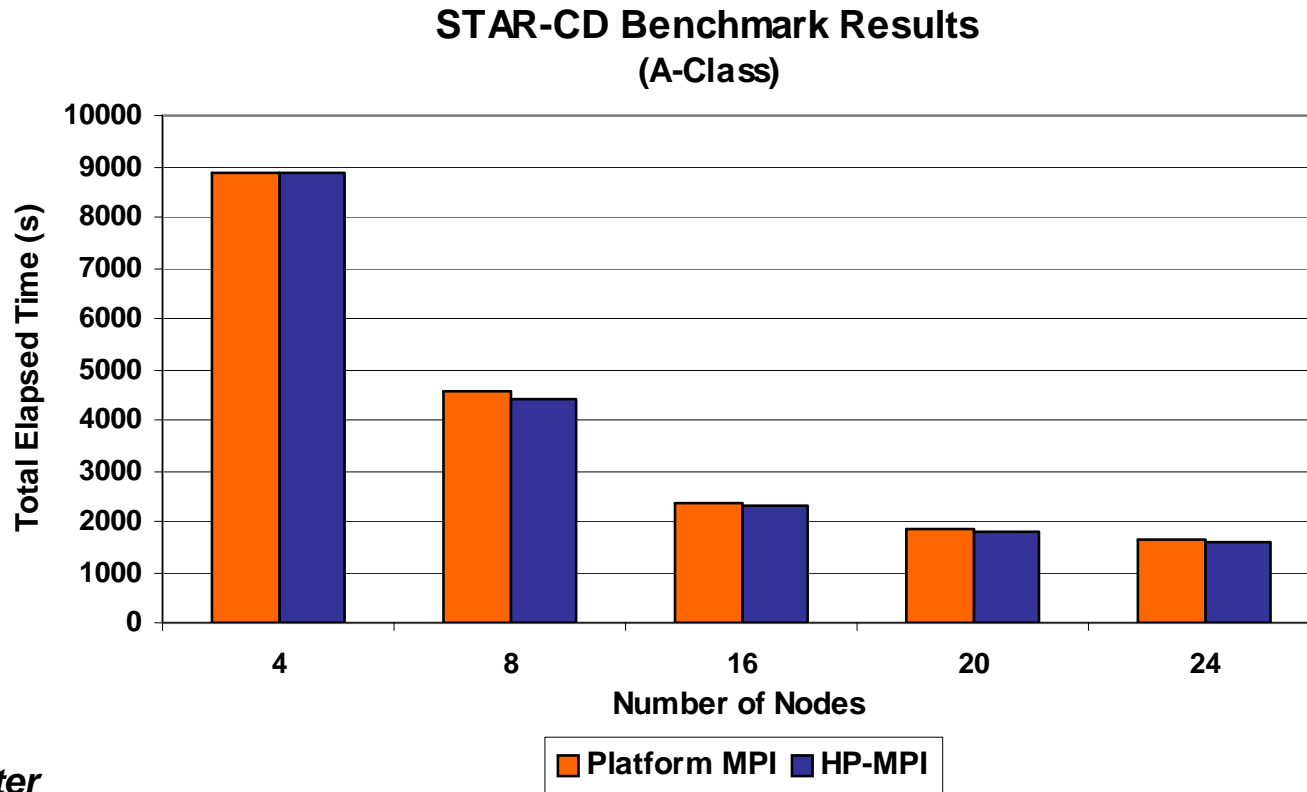


*Lower is better*

*InfiniBand DDR*



- **Test case**
  - Single job over the entire systems
  - Input Dataset (A-Class)
- **HP-MPI has slightly better performance with CPU affinity enabled**



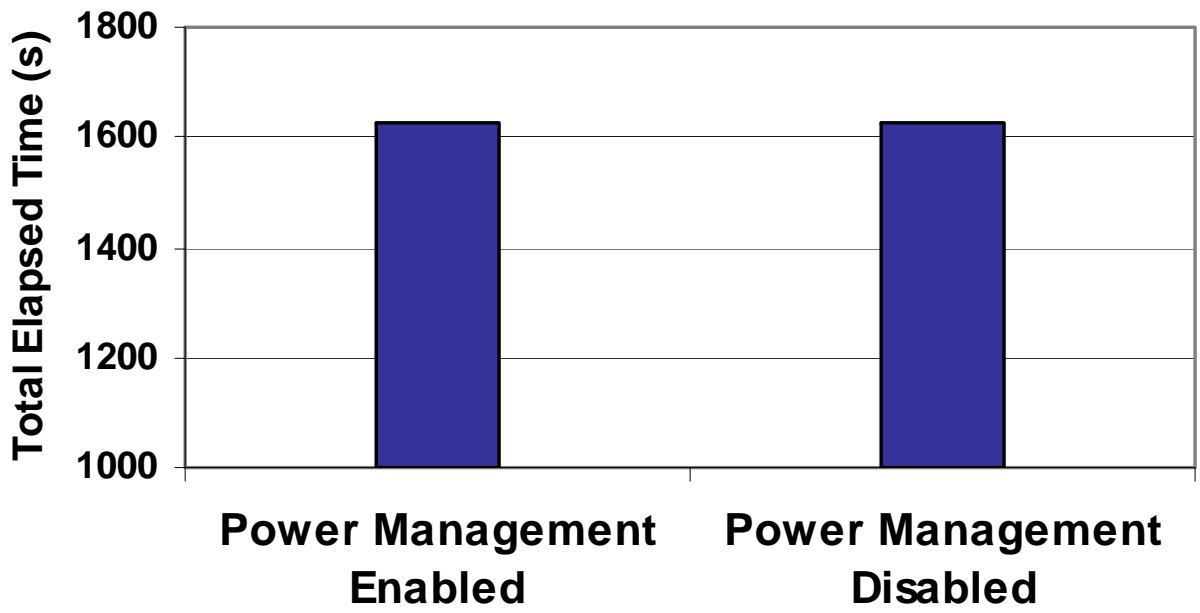


- **STAR-CD was profiled to determine networking dependency**
- **Majority of data transferred between compute nodes**
  - Medium size messages
  - Data transferred increases with cluster size
- **Most used message sizes**
  - <128B messages – MPI\_Allreduce
  - 1K-8K – MPI\_Sendrecv
- **Total number of messages increases with cluster size**
- **Interconnects effect to STAR-CD performance**
  - Both interconnect latency (MPI\_Allreduce) and throughput (MPI\_Sendrecv) influences application performance



- **Test Scenario**
  - 24 servers, 4-Cores/Proc
- **Nearly identical performance with power management enabled or disabled**

**STAR-CD Benchmark Results  
(A Class)**



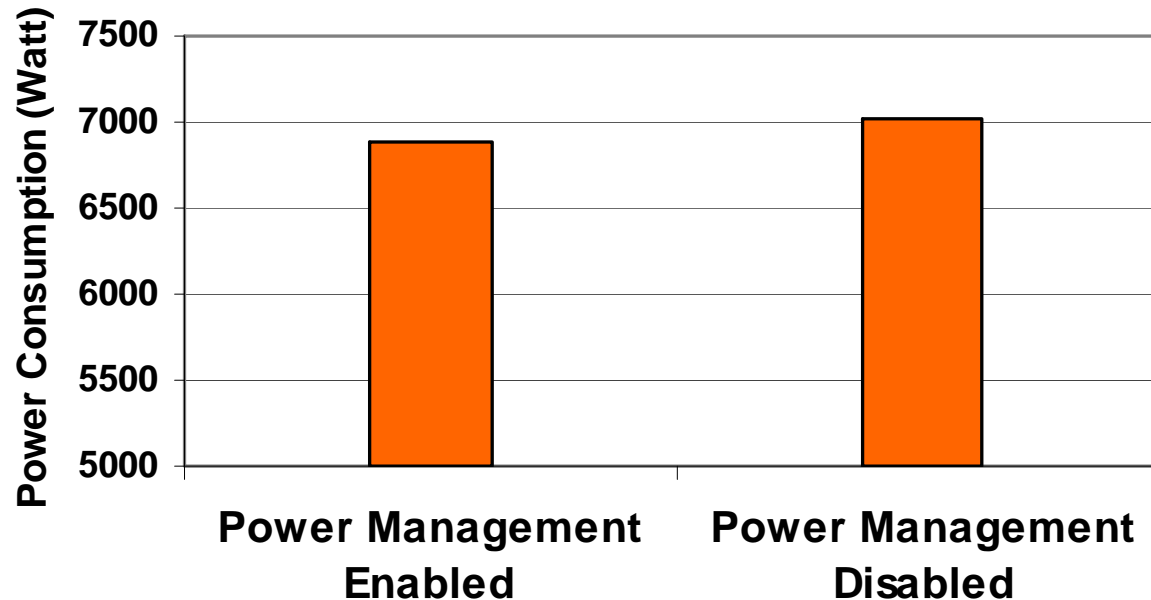
*Lower is better*

*InfiniBand DDR*

- Power management reduces 2% of total system power consumption



## STAR-CD Benchmark Results (A-Class)



*Lower is better*

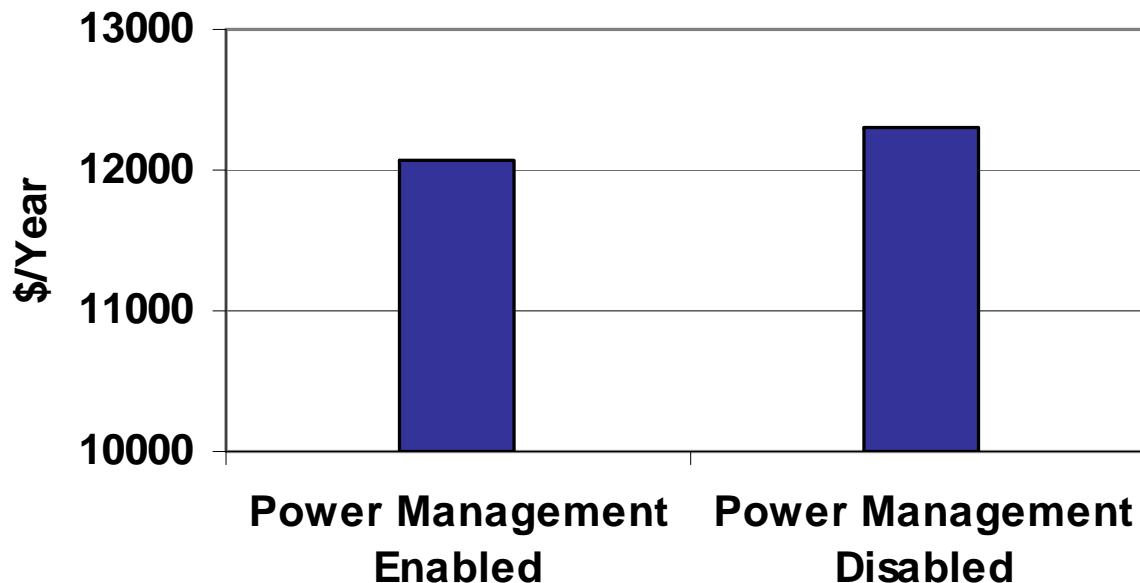
*InfiniBand DDR*

# Power Cost Savings with Power Management



- Power management saves 248\$/year for the 24-node cluster
- As cluster size increases, bigger saving are expected

**STAR-CD Benchmark Results**  
**Power Cost Comparison**



**24 Node Cluster**

$\$/year = Total\ power\ consumption/year\ (KWh) * \$0.20$

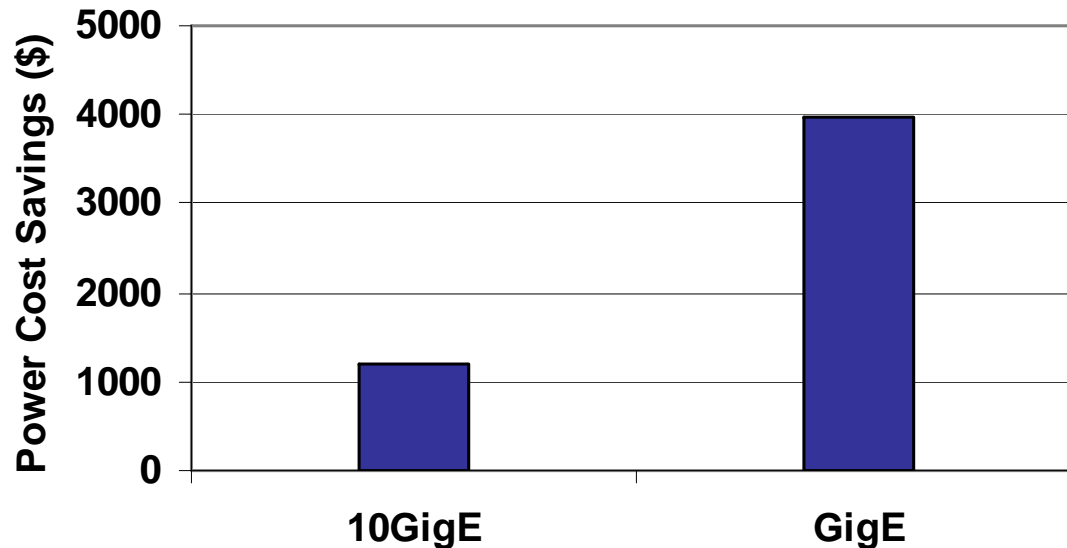
For more information - <http://enterprise.amd.com/Downloads/svrpwrusecompletefinal.pdf>

**InfiniBand DDR**



- InfiniBand saves ~\$1200 and ~\$4000 power to finish the same number of STAR-CD jobs compared to 10GigE and GigE
  - Yearly based for 24-node cluster
- As cluster size increases, more power can be saved

**Power Cost Savings**  
(InfiniBand vs 10GigE and GigE)



24 Node Cluster

$\$/KWh = KWh * \$0.20$

For more information - <http://enterprise.amd.com/Downloads/svrpwrusecompletefinal.pdf>

Power Management enabled

- **STAR-CD and STAR-CCM+ are widely used CFD simulation software**
- **Performance and productivity relies on**
  - Scalable HPC systems and interconnect solutions
  - Low latency and high throughput interconnect technology
  - Reasonably process distribution can dramatically improves performance per core
- **Interconnect comparison shows**
  - InfiniBand delivers superior performance in every cluster size
  - Low latency InfiniBand enables unparalleled scalability
- **Power management provide 2% saving in power consumption**
  - Per 24-node system with InfiniBand
  - \$248 power savings per year for 24-node cluster
  - Power saving increases with cluster size
- **InfiniBand saves power cost**
  - Based on the number of jobs can be finished by InfiniBand per year
  - InfiniBand enables \$1200 and \$4000 power savings compared to 10GigE and GigE

# Thank You

## HPC Advisory Council



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