HPC
ADVISORY COUNCIL

SPECFEM3D Performance Benchmark and Profiling

December 2009

Note

- **The following research was performed under the HPC Advisory Council activities**
  - Participating vendors: Jülich, ParTec, and Mellanox
  - Compute resource - Jülich Supercomputer JUROPA
- **For more info please refer to**
  - www.mellanox.com, http://www.fz-juelich.de/jsc/, http://www.parastation.com/

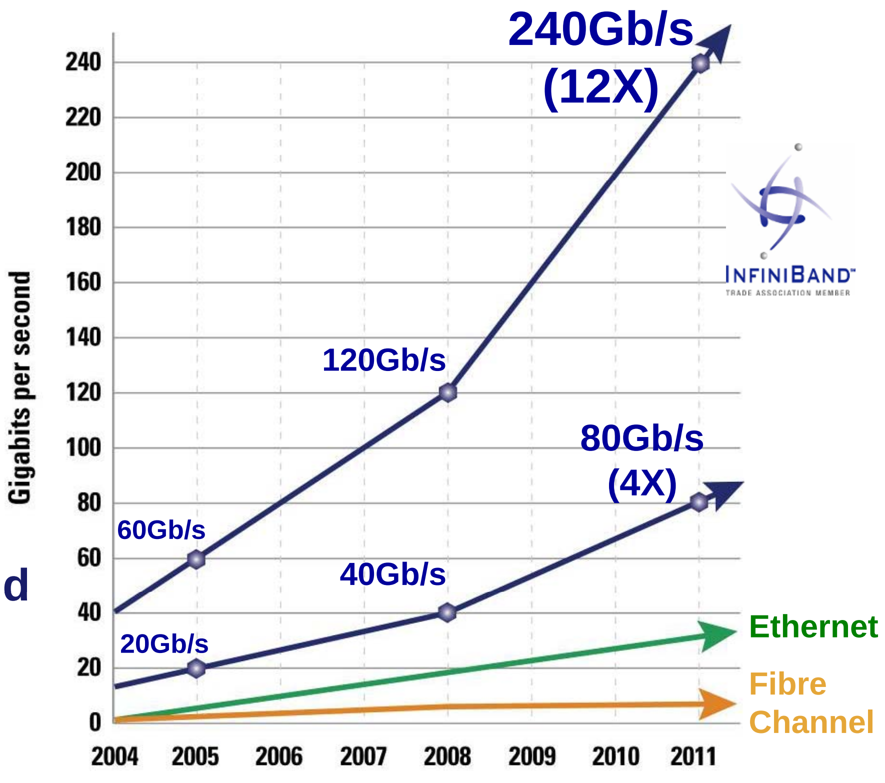
SPECFEM3D

- **SPECFEM3D**
  - Simulates seismic wave propagation in sedimentary basin
  - Can be used to simulate seismic wave propagation in complex three-dimensional geological models such as
    - Anisotropy
    - Attenuation
    - Fluid-solid interfaces
    - Rotation, self-gravitation
    - Crustal and mantle models
- **The package is written in Fortran90 and based on MPI**
- **SPECFEM3D is open source developed by**
  - Dimitri Komatitsch at University of Pau, France
  - California Institute of Technology
  - Princeton University

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**

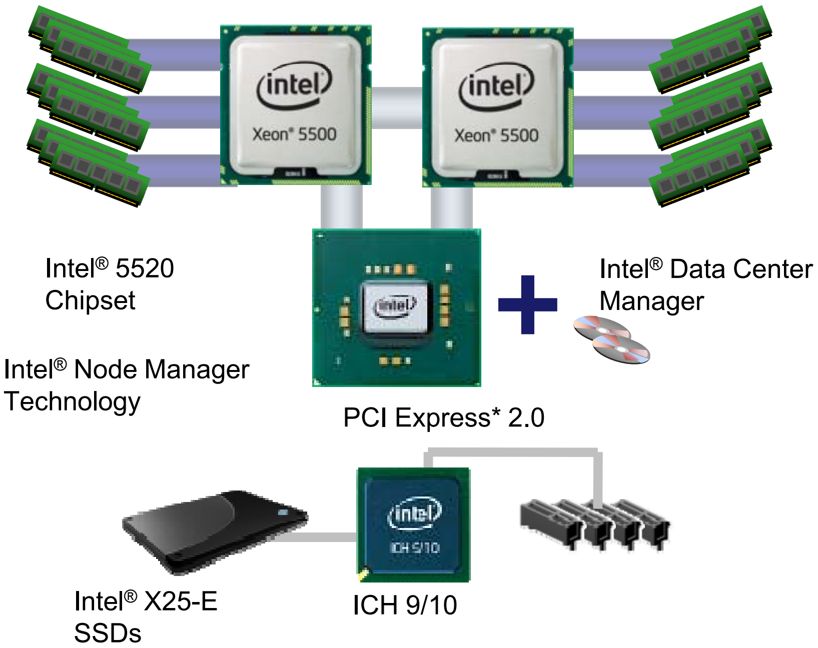
The InfiniBand Performance Gap is Increasing



InfiniBand Delivers the Lowest Latency

Delivering Intelligent Performance

Next Generation Intel® Microarchitecture



Intel® 5520 Chipset

Intel® Node Manager Technology

Intel® Data Center Manager

PCI Express* 2.0

Intel® X25-E SSDs

ICH 9/10

Bandwidth Intensive

- Intel® QuickPath Technology
- Integrated Memory Controller

Threaded Applications

- 45nm quad-core Intel® Xeon® Processors
- Intel® Hyper-threading Technology

Performance on Demand

- Intel® Turbo Boost Technology
- Intel® Intelligent Power Technology

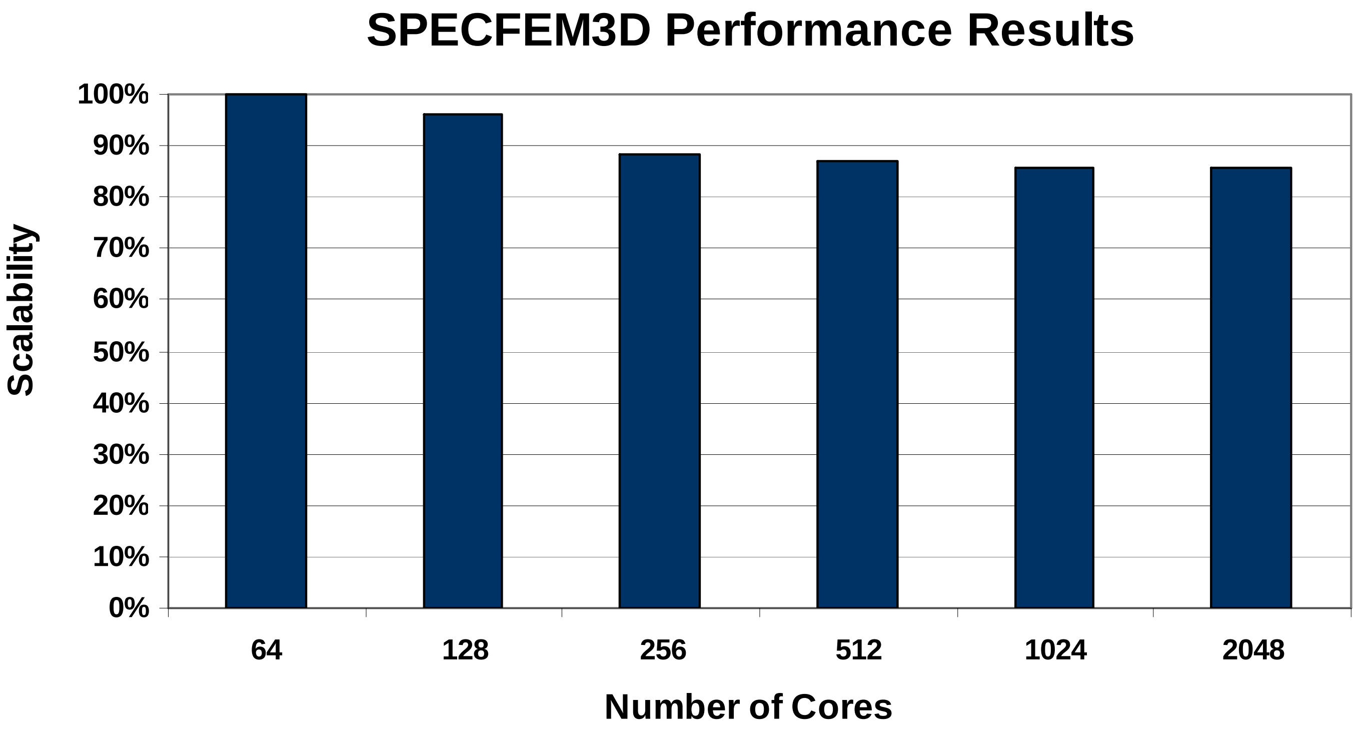
Performance That Adapts to The Software Environment

Test Cluster Configuration

- **Jülich - JuRoPa**
  - Quad core Intel Xeon X5570 2.93 GHz
  - Mellanox IB QDR HCAs and Mellanox based switches
  - Fat tree, non blocking fabric
  - Memory: 24GB memory per node (DDR3, 1066 MHz)
- **OS: SUSE SLES 11, OFED 1.4.1 InfiniBand SW stack**
- **MPI: ParTec MPI**
- **Application: SPECFEM3D-1.4.3**

SPECFEM3D Benchmark Results

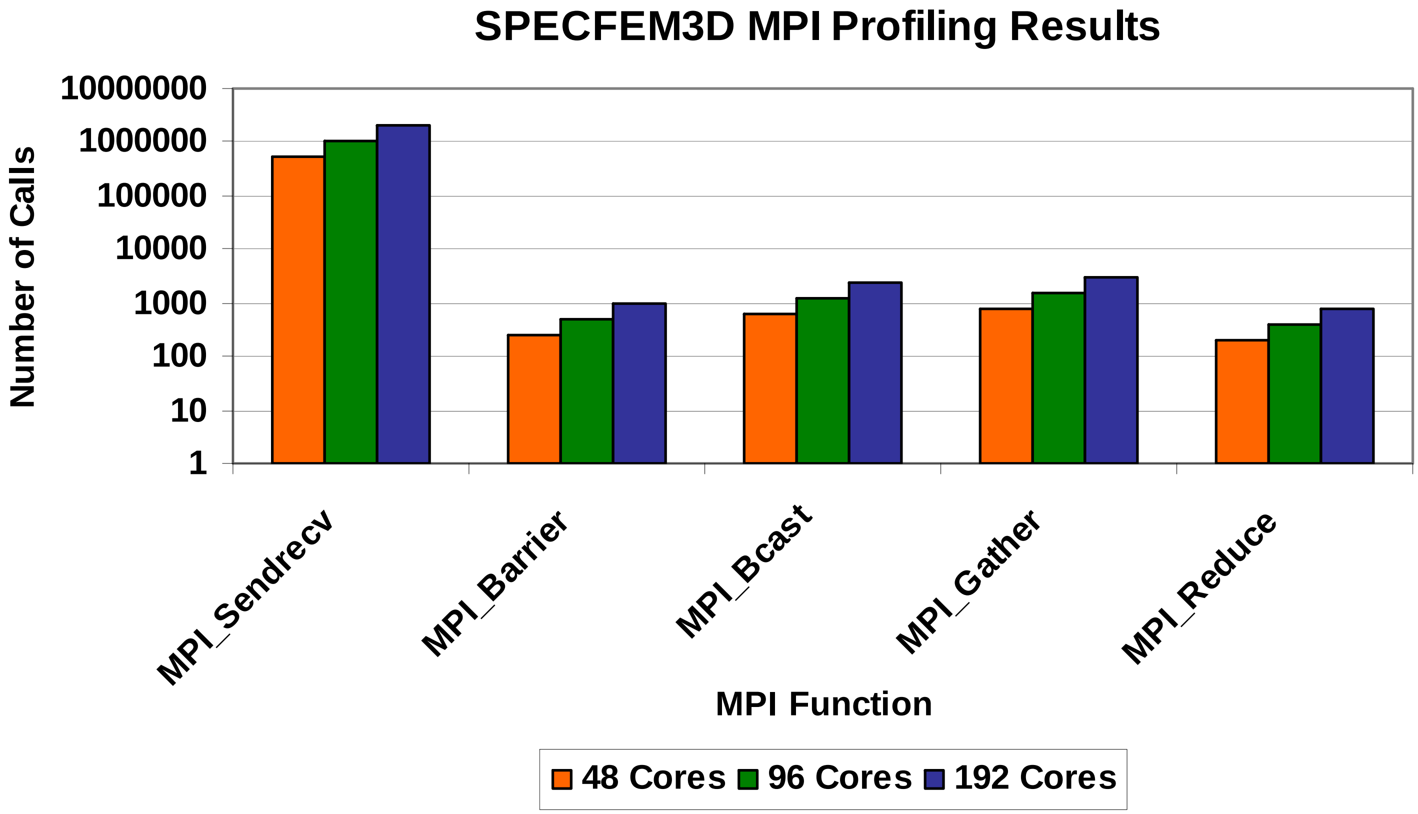
- **Input Dataset - Harvard_LA**
  - 3D model based upon the high-resolution Los Angeles basin
- **InfiniBand QDR enables high performance and scalability**



Higher is better

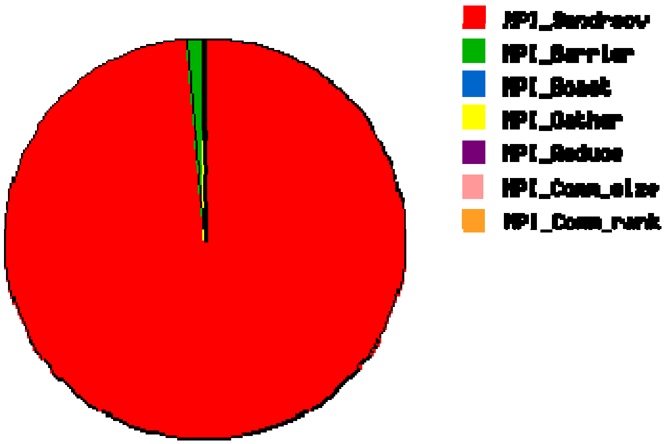
SPECFEM3D Profiling Results

- **Number of messages increases linearly with number of processes**

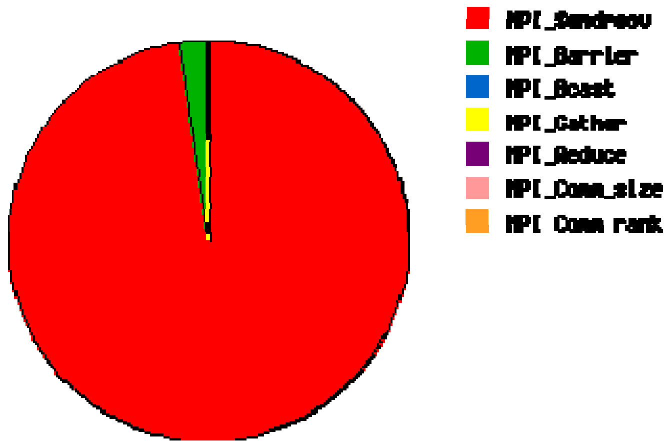


SPECFEM3D Profiling Results

- MPI_Sendrecv creates largest communication overhead
- MPI_Barrier overhead grows as cluster size increases

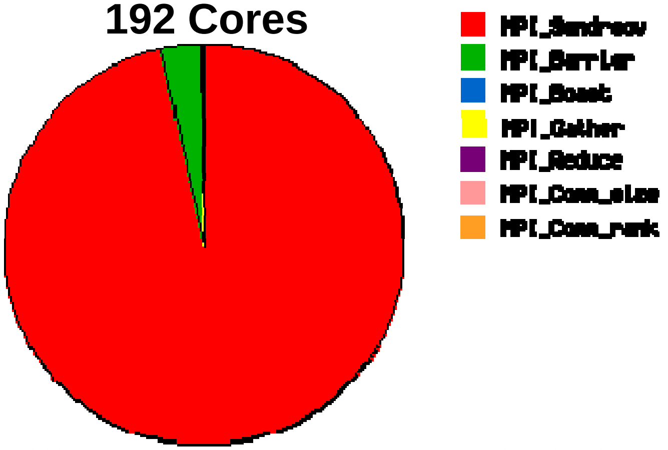


48 Cores



96 Cores

192 Cores

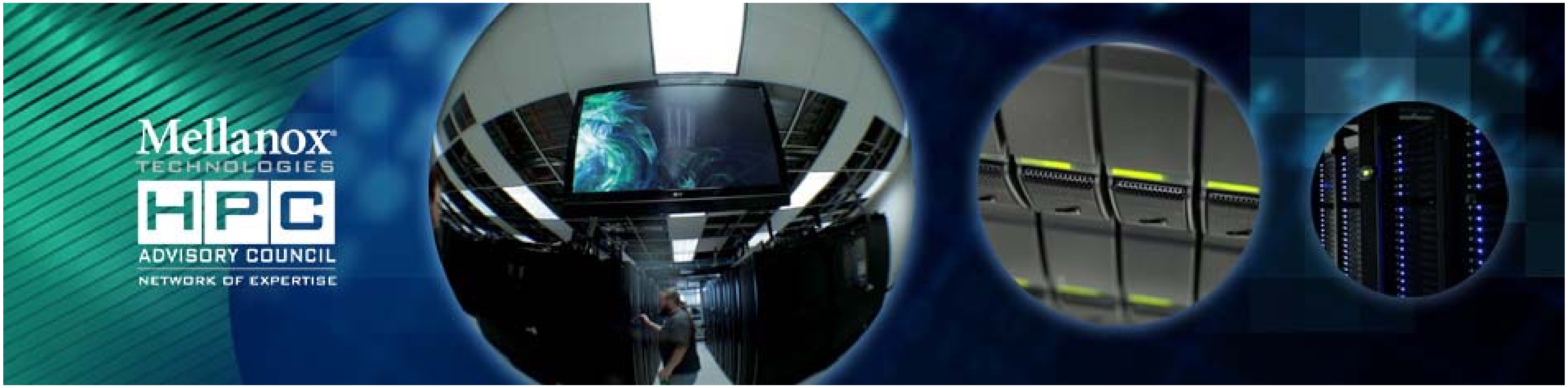


Summary

- **Linear increase of messages impose growing demand of high speed interconnect**
  - The faster interconnect can handle the messages, the better application performance will be achieved
- **Communication overhead of MPI_Barrier increases faster relative to other MPI functions in SPECFEM3D**
  - Mellanox CORE-Direct technology can offload MPI_Barrier to InfiniBand card to accelerate application performance
- **SPECFEM3D demonstrated great scalability over large cluster system**
  - InfiniBand QDR provides low latency and high bandwidth to enable SPECFEM3D scalability
  - ≥86% scalability over 2000 cores
  - Similar scalability is expected over even larger system

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