Leading Supplier of End-to-End Interconnect Solutions

Server / Compute

Virtual Protocol Interconnect
56G InfiniBand

10/40/56GbE & FCoE

Switch / Gateway

Virtual Protocol Interconnect
56G InfiniBand

10/40/56GbE

Storage Front / Back-End

Comprehensive End-to-End InfiniBand and Ethernet Portfolio

ICs

Adapter Cards

Switches/Gateways

Host/Fabric Software

Metro / WAN

Cables/Modules
Mellanox Enables Most Efficient HPC Platforms

**TOP500 Interconnect Trends**

- InfiniBand
- Ethernet
- BlueGene
- Cray
- Other

**PetaFlop Capable Systems on the TOP500 list**

- InfiniBand (Mellanox): 24
- Proprietary: 5
- Cray: 11

**TOP500 - TOP25 Most Efficient System**

© 2014 Mellanox Technologies
InfiniBand’s Unsurpassed System Efficiency

- TOP500 systems listed according to their efficiency
- InfiniBand is the key element responsible for the highest system efficiency; in average 30% higher than 10GbE
- Mellanox delivers efficiencies of more than 99% with InfiniBand

**World Leading Compute Systems Efficiency Comparison**

Average Efficiency
- InfiniBand: 87%
- Cray: 79%
- 10GbE: 67%
- GigE: 40%
Architectural Foundation for Exascale Computing
The 7th generation of Mellanox interconnect adapters

World’s first 100Gb/s interconnect adapter (dual-port FDR 56Gb/s InfiniBand)

Delivers 137 million messages per second – 4X higher than competition

Support the new innovative InfiniBand scalable transport – Dynamically Connected
Connect-IB Provides Highest Interconnect Throughput

Higher is Better

Unidirectional Bandwidth

<table>
<thead>
<tr>
<th>Message Size (bytes)</th>
<th>ConnectX2-Pcie2-QDR</th>
<th>ConnectX3-Pcie3-FDR</th>
<th>Sandy-ConnectIB-DualFDR</th>
<th>Ivy-ConnectIB-DualFDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>12810</td>
<td>12485</td>
<td>6343</td>
<td>3385</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>256</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1024</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4K</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16K</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64K</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>256K</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bidirectional Bandwidth

<table>
<thead>
<tr>
<th>Message Size (bytes)</th>
<th>ConnectX2-Pcie2-QDR</th>
<th>ConnectX3-Pcie3-FDR</th>
<th>Sandy-ConnectIB-DualFDR</th>
<th>Ivy-ConnectIB-DualFDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>256</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1024</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4K</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16K</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64K</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>256K</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

 Gain Your Performance Leadership With Connect-IB Adapters

Source: Prof. DK Panda

© 2014 Mellanox Technologies - Mellanox Confidential - 8
200% Higher Performance Versus Competition, with Only 32-nodes

Performance Gap Increases with Cluster Size

WIEN2k Performance

WRF Performance (conus12km)
Switch-IB EDR
100G InfiniBand

- 7th generation of Mellanox interconnect switch
- 36 EDR (100Gb/s) Ports
- <130ns latency
- Throughput - 7.2 Tb/s
- Software compatible with FDR InfiniBand
- x86 CPU management
- InfiniBand router
- Multiple topologies (Fat-Tree, Torus, Dragonfly +)
Take Advantage of EDR Aggregation for FDR Clusters

648-node FDR cluster – 1:1

648-node EDR cluster – 2:1 (FDR2EDR)

EDR Network Aggregation Improves Cost-Performance

- Future proof
- Less real estate (less switches, less cables)
- Lower latency
- Wider pipes reduces congestion
HPC-X™

- Complete MPI, PGAS/OpenSHMEM/UPC package for HPC environments
  - Fully optimized for Mellanox InfiniBand and VPI interconnect solutions
  - Supports 3rd party solutions

- Components
  - Communication libraries: ScalableMPI, ScalableSHMEM, ScalableUPC
  - Acceleration libraries: MXM – Messaging Accelerator, FCA – Fabric Collectives Accelerator
  - Tools: Integrated Performance Monitoring Tool (IPM), benchmarks etc.

20%-70% Performance Improvement
HPC-X™ Performance Advantage – CD-Adapco

21% Performance Advantage!

HPC-X™ Performance for Commercial CFD
HPC-X™ Performance Advantage – VPS (PAM-CRASH)

8% Performance Advantage!

VPS 2013.01 Performance (NEON_FINE_CAR2CAR)

HPC-X™ Performance for Commercial FEA
GPUDirect RDMA

Native support for peer-to-peer communications between Mellanox HCA adapters and NVIDIA GPU devices
Evolution of GPUDirect RDMA

Before GPUDirect
- Network and third-party device drivers, did not share buffers, and needed to make a redundant copy in host memory.

With GPUDirect Shared Host Memory Pages
- The network and GPU can share “pinned” (page-locked) buffers, eliminating the need to make a redundant copy in host memory.
- Eliminates CPU bandwidth and latency bottlenecks
- Uses remote direct memory access (RDMA) transfers between GPUs
- Resulting in significantly improved MPISendRecv efficiency between GPUs in remote nodes
Performance of MVAPICH2 with GPUDirect RDMA

**GPU-GPU Internode MPI Latency**

- Lower is Better
- Source: Prof. DK Panda

**GPU-GPU Internode MPI Bandwidth**

- Higher is Better

67% Lower Latency

5X Increase in Throughput
Mellanox PeerDirect™ with NVIDIA GPUDirect RDMA

- HOOMD-blue is a general-purpose Molecular Dynamics simulation code accelerated on GPUs
- GPUDirect RDMA allows direct peer to peer GPU communications over InfiniBand
  - Unlocks performance between GPU and InfiniBand
  - This provides a significant decrease in GPU-GPU communication latency
  - Provides complete CPU offload from all GPU communications across the network
- Demonstrated up to 102% performance improvement with large number of particles

### HOOMD-blue Performance (LJ Liquid Benchmark, 16K Particles)

<table>
<thead>
<tr>
<th>Number of Nodes</th>
<th>Without GPUDirect RDMA</th>
<th>With GPUDirect RDMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2900</td>
<td>3300</td>
</tr>
<tr>
<td>2</td>
<td>3000</td>
<td>3500</td>
</tr>
<tr>
<td>3</td>
<td>3300</td>
<td>3900</td>
</tr>
<tr>
<td>4</td>
<td>3600</td>
<td>4200</td>
</tr>
</tbody>
</table>

**Performance Improvement:** 21%

### HOOMD-blue Performance (LJ Liquid Benchmark, 512K Particles)

<table>
<thead>
<tr>
<th>Number of Nodes</th>
<th>Without GPUDirect RDMA</th>
<th>With GPUDirect RDMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2500</td>
<td>5000</td>
</tr>
<tr>
<td>2</td>
<td>5000</td>
<td>10000</td>
</tr>
<tr>
<td>4</td>
<td>10000</td>
<td>20000</td>
</tr>
<tr>
<td>8</td>
<td>20000</td>
<td>40000</td>
</tr>
<tr>
<td>16</td>
<td>40000</td>
<td>80000</td>
</tr>
<tr>
<td>32</td>
<td>80000</td>
<td>160000</td>
</tr>
<tr>
<td>64</td>
<td>160000</td>
<td>320000</td>
</tr>
<tr>
<td>96</td>
<td>320000</td>
<td>640000</td>
</tr>
</tbody>
</table>

**Performance Improvement:** 102%
Questions?
Thank You