Paving the Road to Exascale
Interconnecting Your Future

December 18, 2015 | PC Cluster Symposium
メラノックステクノロジーズジャパン株式会社

japan_sales@mellanox.com
Leading Supplier of End-to-End Interconnect Solutions

Enabling the Use of Data

Comprehensive End-to-End InfiniBand and Ethernet Portfolio

ICs | Adapter Cards | Switches/Gateways | Software and Services | Metro / WAN | Cables/Modules
--- | --- | --- | --- | --- | ---
ConnectX-4 | ConnectX-3 | SwitchX-2 | SwitchIB-2 | Store

At the Speeds of 10, 25, 40, 50, 56 and 100 Gigabit per Second
TOP500 Interconnect Trends

- "LENOX" EDR InfiniBand system
  Lenovo HPC innovation center
- Shanghai Supercomputer Center
  Magic Cube II supercomputer
InfiniBand Switch Portfolio

**Modular Switches**
- 648 port
- 324 port
- 216 port
- 108 port

**Edge Switches**
- 36 ports externally managed
- 18 port externally managed
- 8-12 ports externally managed
- 36 ports managed
- 18 port managed
- 12 port managed

**Long Distance**
- metroX

**Bridge – VPI**
- Bridging
- SwitchX²

**Management**
- Interface
- Performance
- Topology
- Security
- Traffic Management
- Facilitator
- LMS Manager
High-Performance Designed 100Gb/s Interconnect Solutions

**Adapters**
- ConnectX-4: 100Gb/s Adapter, 0.7us latency
- 150 million messages per second
- (10 / 25 / 40 / 50 / 56 / 100Gb/s)

**Switch**
- SwitchIB2: 36 EDR (100Gb/s) Ports, <90ns Latency
- Throughput of 7.2Tb/s
- 7.02 Billion msg/sec (195M msg/sec/port)

**Switch**
- Spectrum: 32 100GbE Ports, 64 25/50GbE Ports
- (10 / 25 / 40 / 50 / 100GbE)
- Throughput of 6.4Tb/s

**Interconnect**
- LinkX: Transceivers
- Active Optical and Copper Cables
- (10 / 25 / 40 / 50 / 56 / 100Gb/s)
- VCSELs, Silicon Photonics and Copper
Enter the Word of Scalable Performance – 100Gb/s Switch

Best Performance
- 90ns switch latency
- Throughput of 7.2 Tb/s in 1U
- 195M messages per second
- 136W ATIS reported power

Enhanced Capabilities
- InfiniBand Router
- Adaptive Routing (AR)
- Fault Routing (FR)

High Resiliency
- High efficiency power supplies
- AC / DC / BBU power supplies option
- Class 4 (3.5W) supported on all ports

x86 Powerful CPU
- Improved software upgrade time
- Up to 2048 nodes in cluster
- Run Virtual Machine (VM)
Benefits of InfiniBand Router

- Enable scaling an HPC cluster over 48K LIDs
- Enable sharing a common storage network by multiple disconnected subnets
  - Limit congestion spread to source subnet
- Allow running HPC/MPI jobs efficiently on the joint network
  - Maintaining large bisectional bandwidth between the subnets
  - Low latency penalty for crossing subnets
- Isolation of SM responsibilities
- Simple administration and out-of-the-box experience
EDR InfiniBand Performance Leadership

**OptiStruct Performance**  
(Engine_Assy.fem)

- **Altair**  
- 40% increase

---

**RADIOSS 13.0 Performance**  
(NEON1M11, MPP)

- **Altair**  
- 25% increase

---

**LS-DYNA Performance**  
(neon_refined_revised)

- 15% increase

---

For all graphs: higher is better.
**ConnectX-4: Highest Performance Adapter in the Market**

- **InfiniBand:** SDR / DDR / QDR / FDR / EDR
- **Ethernet:** 10 / 25 / 40 / 50 / 56 / 100GbE

- 100Gb/s, <0.7us latency
- 150 million messages per second

- OpenPOWER CAPI Technology
- CORE-Direct Technology
- GPUDirect RDMA
- Dynamically Connected Transport (DCT)
- Ethernet Offloads (HDS, RSS, TSS, LRO, LSOv2)
### ConnectX-4 EDR 100G InfiniBand

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Uni-Directional Throughput</td>
<td>100 Gb/s</td>
</tr>
<tr>
<td>Bi-Directional Throughput</td>
<td>195 Gb/s</td>
</tr>
<tr>
<td>Latency</td>
<td>0.61 us</td>
</tr>
<tr>
<td>Message Rate</td>
<td>149.5 Million/sec</td>
</tr>
</tbody>
</table>
### InfiniBand Adapters Performance Comparison

<table>
<thead>
<tr>
<th>Mellanox Adapters</th>
<th>ConnectX-4</th>
<th>Connect-IB</th>
<th>ConnectX-3 Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Port Performance</td>
<td>EDR 100G</td>
<td>FDR 56G</td>
<td>FDR 56G</td>
</tr>
<tr>
<td>Uni-Directional Throughput</td>
<td>100 Gb/s</td>
<td>54.24 Gb/s</td>
<td>51.1 Gb/s</td>
</tr>
<tr>
<td>Bi-Directional Throughput</td>
<td>195 Gb/s</td>
<td>107.64 Gb/s</td>
<td>98.4 Gb/s</td>
</tr>
<tr>
<td>Latency</td>
<td>0.61 us</td>
<td>0.63 us</td>
<td>0.64 us</td>
</tr>
<tr>
<td>Message Rate</td>
<td>149.5 Million/sec</td>
<td>105 Million/sec</td>
<td>35.9 Million/sec</td>
</tr>
</tbody>
</table>
Mellanox QSFP 100Gb/s Cables

Complete Solution of 100Gb/s Copper and Fiber Cables

Copper Cables
VCSEL AOCs
Silicon Photonics AOCs

Making 100Gb/s Deployments as Easy as 10Gb/s
The Ever Growing Demand for Higher Performance

Performance Development

Terascale  Petascale  Exascale


The Interconnect is the Enabling Technology

SMP to Clusters  Single-Core to Many-Core  Co-Design

Application
Software
Hardware

© 2015 Mellanox Technologies
- Mellanox Confidential -
Co-Design Architecture to Enable Exascale Performance

CPU-Centric

Limited to Main CPU Usage
Results in Performance Limitation

Co-Design

Creating Synergies
Enables Higher Performance and Scale
Introducing Switch-IB 2 World’s First Smart Switch

- The world fastest switch with <90 nanosecond latency
- 36-ports, 100Gb/s per port, 7.2Tb/s throughput, 7.02 Billion messages/sec
- Adaptive Routing, Congestion Control, support for multiple topologies

The Only Approach to Deliver 10X Performance Improvements
SHArP (Scalable Hierarchical Aggregation Protocol) Technology

SHArP Enables Switch-IB 2 to Manage and Execute MPI Operations in the Network

Switch-IB 2 Enables the Switch Network to Operate as a Co-Processor

Delivering 10X Performance Improvement for MPI and SHMEM/PGAS Applications
Mellanox delivers the HIGHEST return on investment for ANY scale deployment!
Mellanox InfiniBand Solutions Deliver Highest ROI for Any Scale

Smart Network For Smarter Systems
RDMA, Acceleration Engines, Programmability

- Higher Performance
- Unlimited Scalability
- Higher Resiliency
- Proven!

Power Consumption
Per Switch Port

- 25% Lower

Message Rate

- 44% Higher

Switch Latency

- 20% Lower

Scalability
CPU efficiency

- 2X Higher

100 Gb/s
Link Speed

200 Gb/s
Link Speed

2014

Gain Competitive Advantage Today

2017

Protect Your Future

Gain Competitive Advantage Today

Proven!
Mellanox Solutions

Offload Technology

Proven Solutions

Eco-System

Standard

Speed-Up Your Present, Protect Your Future
Mellanox Delivers Best Interconnect

**Higher Performance**
- 100Gb/s throughput at 0% CPU utilization
- Adapter: 150 Million messages/sec on today’s systems, 44% higher
- Switch: 7.02 Billion messages/sec (195 Million per port)
- 20% lower switch latency, with deterministic latency!

**Lower TCO**
- 25% lower power consumption per switch port
- Standard-based solutions, large eco-system support
- Backward and future compatibility – protect investments
- Offloading Architecture (RDMA, GPUDirect etc.) delivers highest system efficiency

**Higher Reliability**
- 1,000X higher reliability - Mellanox delivers Bit Error Rate of $10^{-15}$ versus $10^{-12}$
- Superior signal integrity
- Support for Multiple data integrity mechanisms (FEC\(^1\), LLR\(^2\), COD\(^3\) and more)

---

1 - Forward Error Correction; 2 - Link Level Retransmission; 3 - Correction on Demand
Thank You
Highest Performance and Scalability for x86, Power, GPU, ARM and FPGA-based Compute and Storage Platforms

10, 20, 25, 40, 50, 56 and 100Gb/s Speeds

Smart Interconnect to Unleash The Power of All Compute Architectures