EDR InfiniBand

OFA UM 2015

January 2015

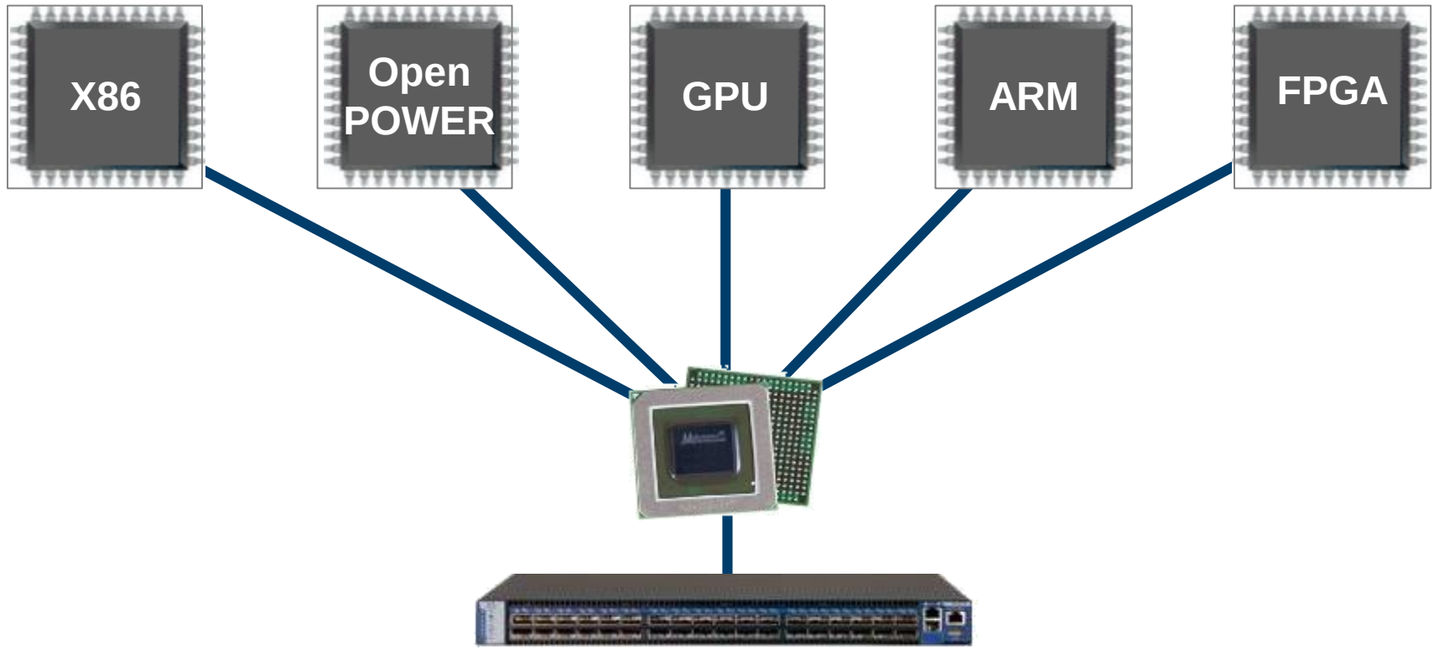
Mellanox TECHNOLOGIES

Connect. Accelerate. Outperform.™

End-to-End Interconnect Solutions for All Platforms

Highest Performance and Scalability for

X86, Power, GPU, ARM and FPGA-based Compute and Storage Platforms



Smart Interconnect to Unleash The Power of All Compute Architectures

Technology Roadmap

Mellanox → **20Gbs** → **40Gbs** → **56Gbs** → **100Gbs** → **200Gbs** →

Terascale

3rd

TOP500 2003
Virginia Tech (Apple)

Petascale

1st

“Roadrunner”
Mellanox Connected

Exascale

OAK RIDGE National Laboratory
“Summit” System

Lawrence Livermore National Laboratory
“Sierra” System

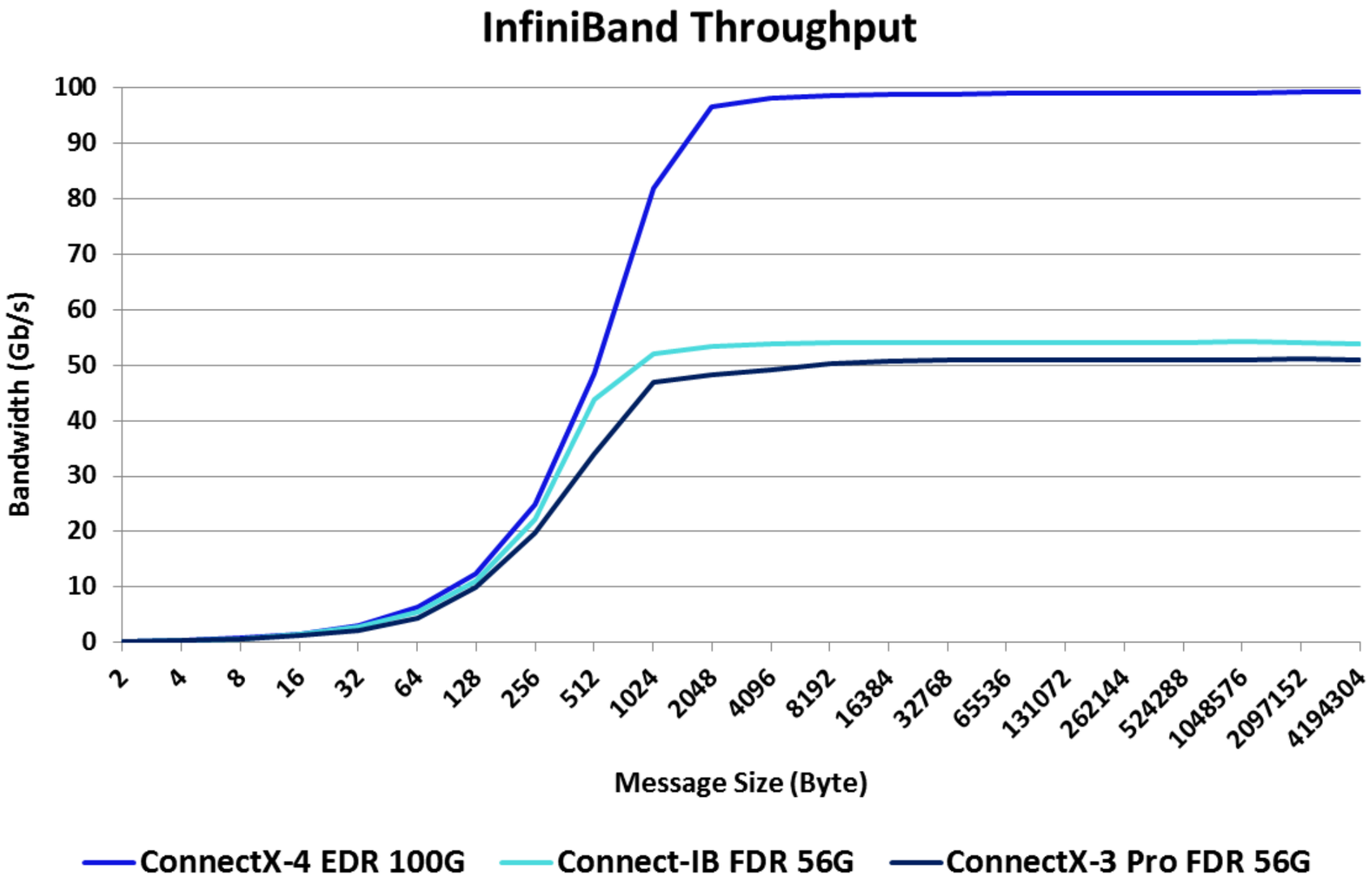
2000 **2005** **2010** **2015** **2020**

InfiniBand Adapters Performance Comparison

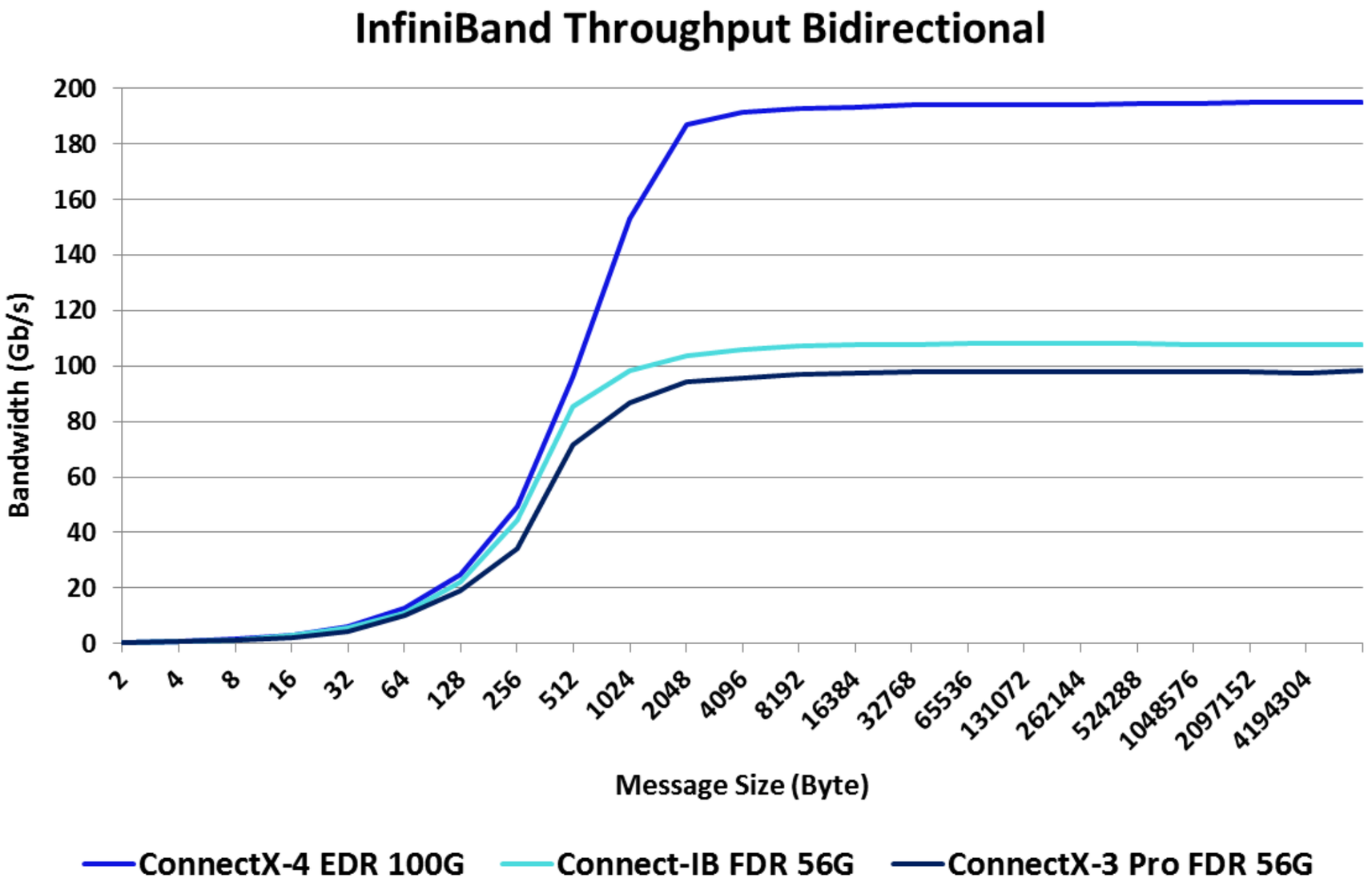
	ConnectX-4 EDR 100G*	Connect-IB FDR 56G	ConnectX-3 Pro FDR 56G
InfiniBand Throughput	100 Gb/s	54.24 Gb/s	51.1 Gb/s
InfiniBand Bi-Directional Throughput	195 Gb/s	107.64 Gb/s	98.4 Gb/s
InfiniBand Latency	0.61 us	0.63 us	0.64 us
InfiniBand Message Rate	149.5 Million/sec	105 Million/sec	35.9 Million/sec
MPI Bi-Directional Throughput	193.1 Gb/s	112.7 Gb/s	102.1 Gb/s

***First results, optimizations in progress**

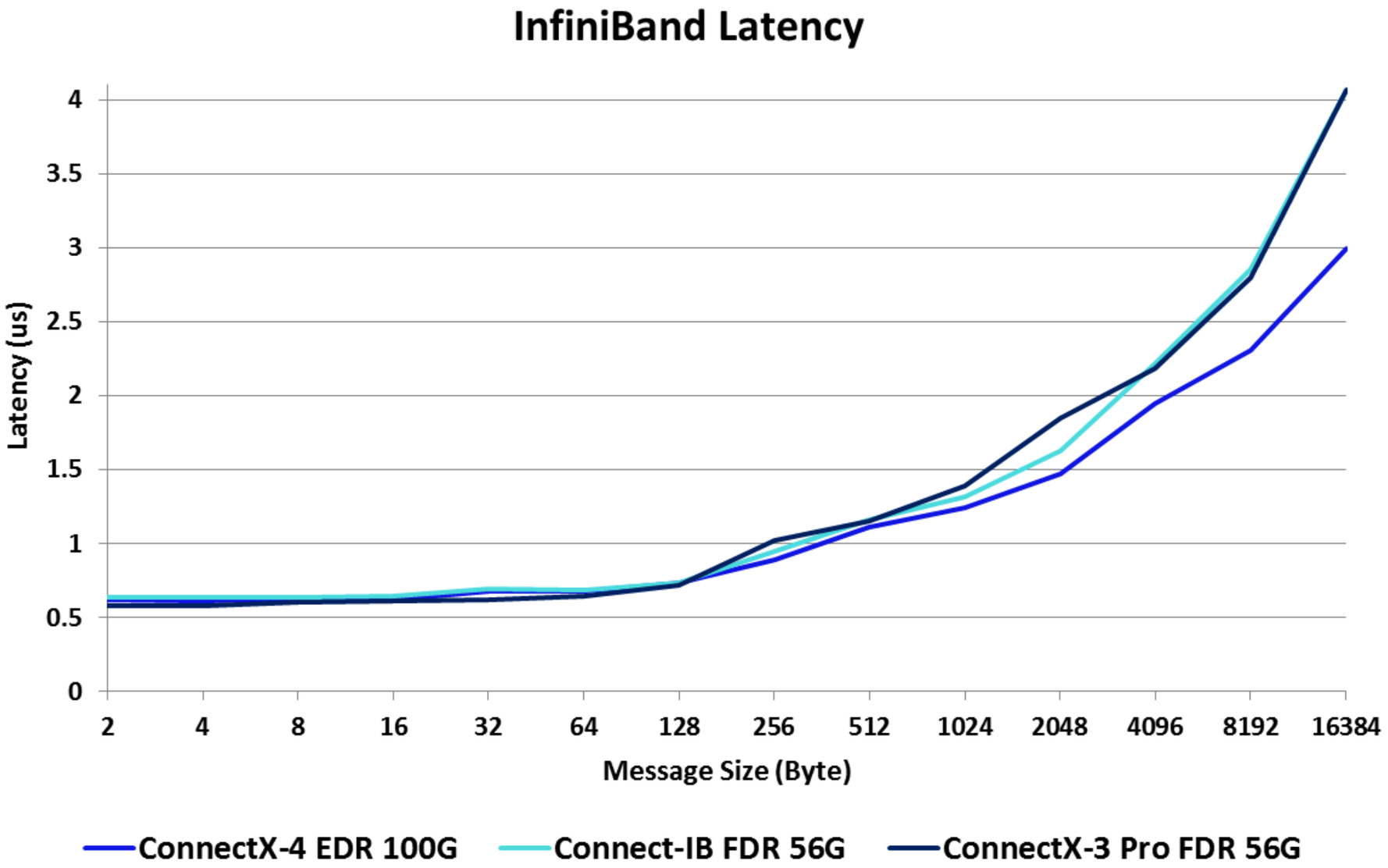
InfiniBand Throughput



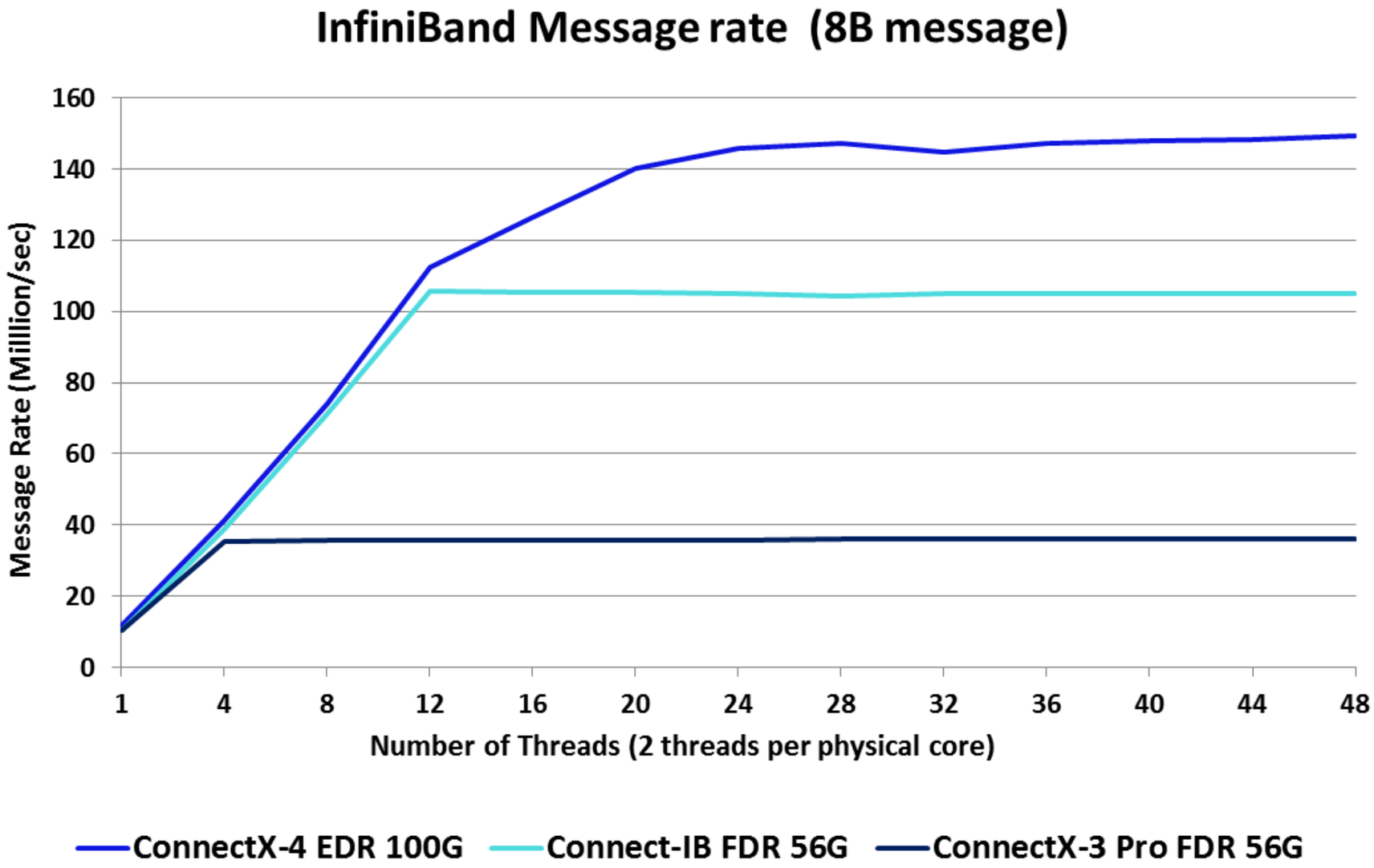
InfiniBand Bi-Directional Throughput



InfiniBand Latency



InfiniBand Message Rate



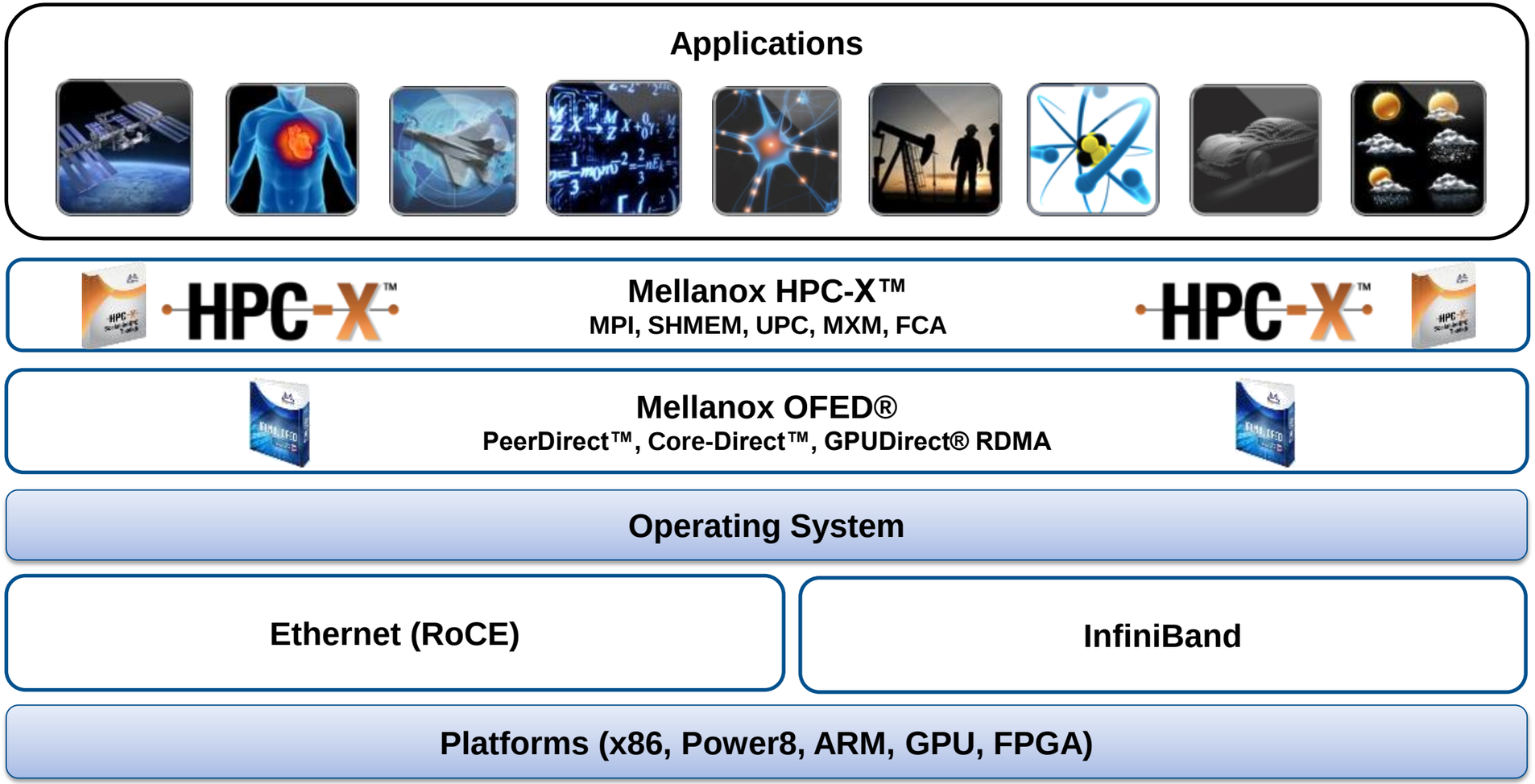
Comprehensive HPC Software

Mellanox HPC-X™ Scalable HPC Software Toolkit

HPC-X™

- MPI, PGAS OpenSHMEM and UPC package for HPC environments
- Fully optimized for standard InfiniBand and Ethernet interconnect solutions
- Maximize application performance
- For commercial and open source usage
- OpenMPI based, Berkley UPC based

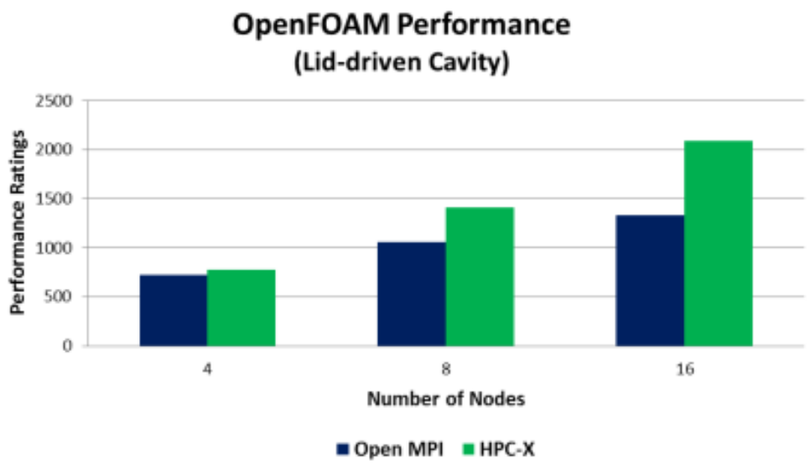
Enabling Applications Scalability and Performance



Comprehensive MPI, PGAS/OpenSHMEM/UPC Software Suite

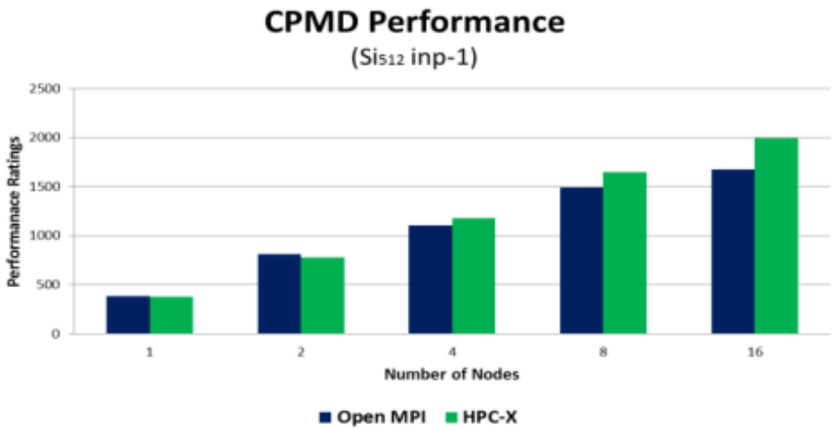
HPC-X Performance Advantage

58% Performance Advantage!



OpenFOAM

20% Performance Advantage!



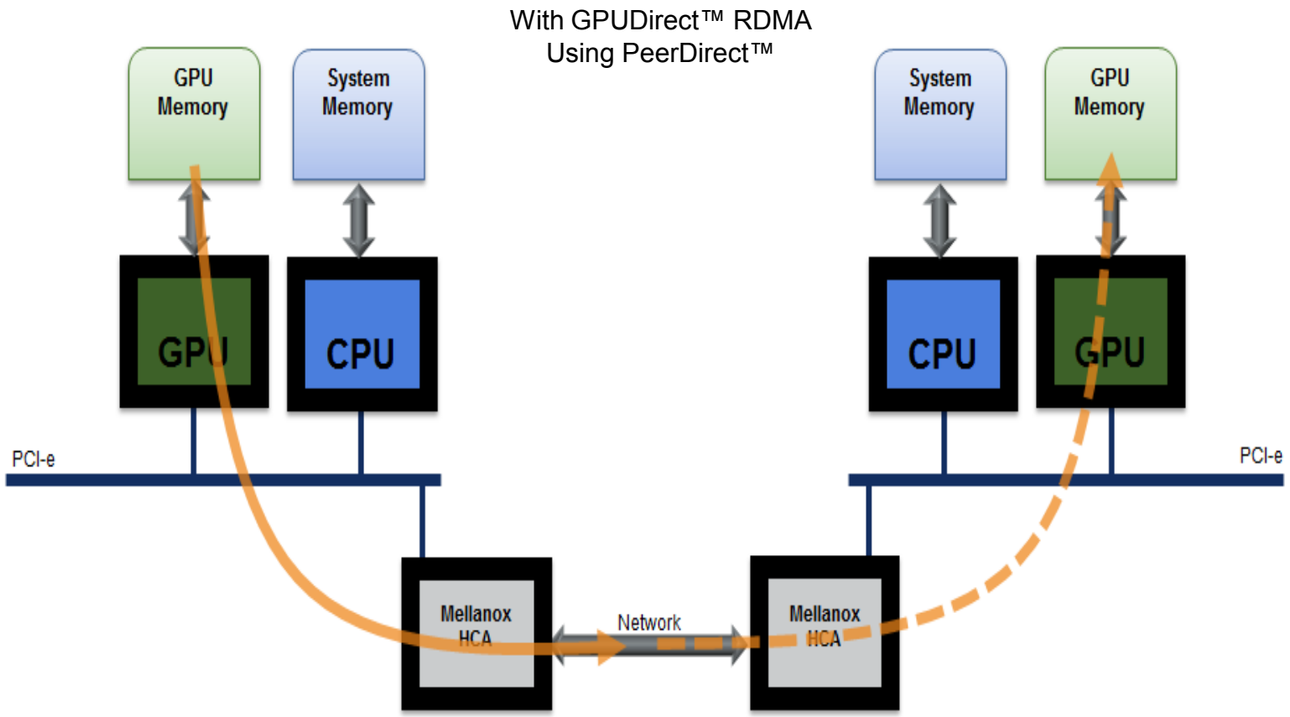
CPMD

Enabling Highest Applications Scalability and Performance

GPUDirect RDMA

Accelerator and GPU Offloads

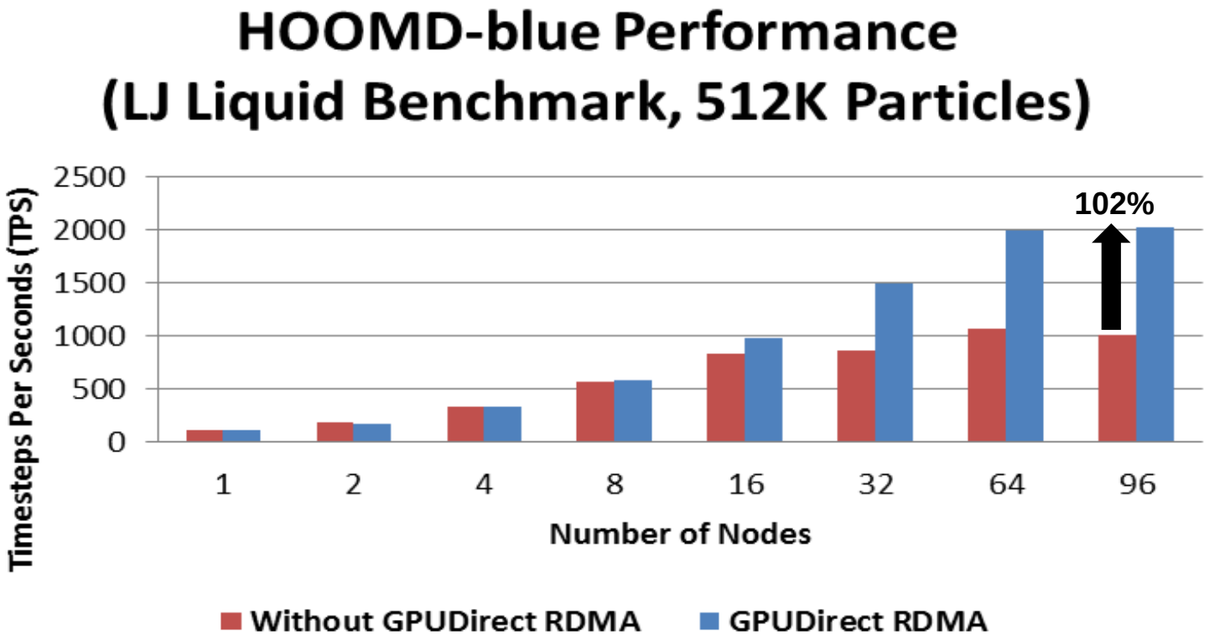
GPUDirect™ RDMA (GPUDirect 3.0)



- Eliminates CPU bandwidth and latency bottlenecks
- Uses remote direct memory access (RDMA) transfers between GPUs
- Resulting in significantly improved MPI SendRecv efficiency between GPUs in remote nodes
- Based on PeerDirect technology

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

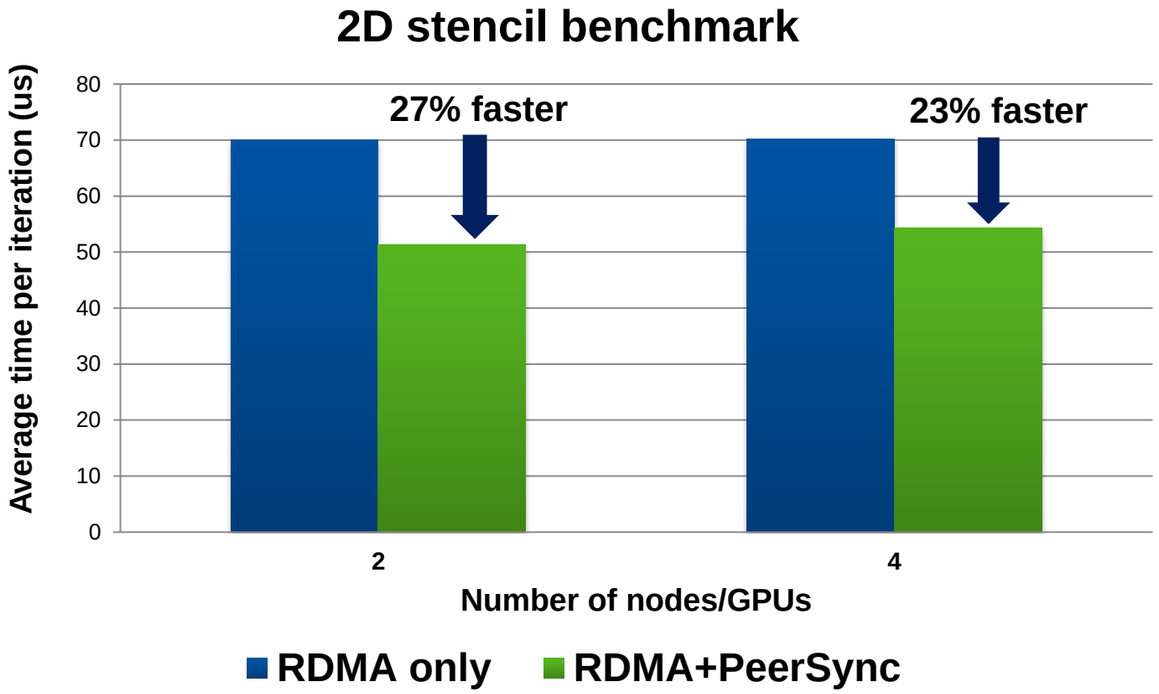


GPUDirect Sync (GPUDirect 4.0)

- GPUDirect RDMA (3.0) – direct data path between the GPU and Mellanox interconnect
  - Control path still uses the CPU
    - CPU prepares and queues communication tasks on GPU
    - GPU triggers communication on HCA
    - Mellanox HCA directly accesses GPU memory

- GPUDirect Sync (GPUDirect 4.0)
  - Both data path and control path go directly between the GPU and the Mellanox interconnect

Maximum Performance For GPU Clusters



Thank You

Mellanox TECHNOLOGIES

Connect. Accelerate. Outperform.™