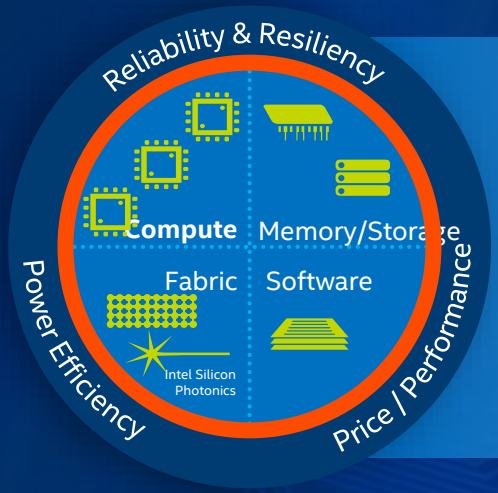
INTEL® SCALABLE SYSTEM FRAMEWORK

A CONFIGURABLE DESIGN PHILOSOPHY EXTENSIBLE TO A WIDE RANGE OF WORKLOADS



Small Clusters Through Supercomputers

Compute and Data-Centric Computing

Standards-Based Programmability

On-Premise and Cloud-Based

Intel® Xeon® Processors	Intel® Solutions for Lustre*	Intel® Omni-Path Architecture	HPC System Software Stack
Intel® Xeon Phi™ Processors	Intel® SSDs	Intel® True Scale Fabric	Intel® Software Tools
Intel® Xeon Phi™ Coprocessors	Intel® Optane™ Technology	Intel® Ethernet	Intel® Cluster Ready Program
Intel® Server Boards and Platforms	3D XPoint™ Technology	Intel® Silicon Photonics	Intel® Visualization Toolkit

INTEL 100G OMNI-PATH FABRIC - ITOC2016

YANG YANGUO

May 2016

Agenda

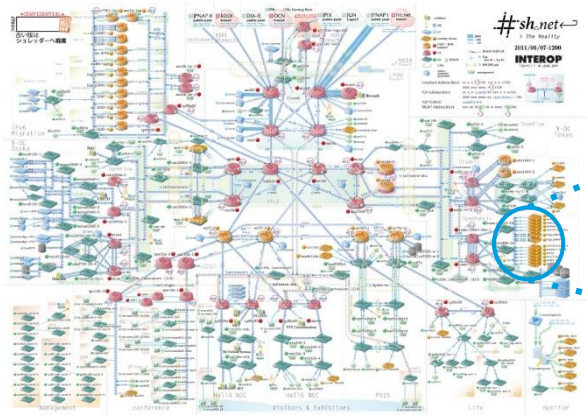
- Quick Overview: HPC Fabrics
- What is Intel® 100Gb Omni-Path Architecture(OPA)?
- Why is Intel 100Gb OPA
- Summary

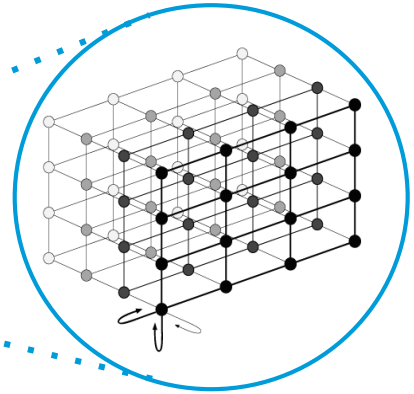
QUICK OVERVIEW: HPC FABRICS

What is Different Between Networks and Fabrics?

Network: Universal interconnect designed to allow any-and-all systems to communicate

HPC Fabric: Optimized interconnect allows many nodes to perform as a single system





Intel® Omni-Path Architecture or Infiniband

Key NETWORK (Ethernet) Attributes:

- Flexibility for any application
- Designed for universal communication
- Extensible configuration
- Multi-vendor components

Key FABRIC Attributes:

- Targeted for specific applications
- Optimized for performance and efficiency
- Engineered topologies
- Single-vendor solutions

Fabric: InfiniBand* and OPA

InfiniBand/OPA is a multi-lane, high-speed serial interconnect (Copper or Fiber)

- Typically presented as a 4x solution
- Speeds: 40Gb/s (M & Intel QDR), 56Gb/s (M FDR), 100Gb/s (EDR & Intel OPA)

High bandwidth, low latency HPC interconnect for commodity servers

- Ethernet switch latency is typically measured in µs, but InfiniBand/OPA is in *nanoseconds*
- Lower CPU load
- Lower cost than Ethernet
  - 100GbE measured in multiple $1,000's per switch port
  - 100Gb OPA is ~$1k per switch port (target for Intel® OPA list pricing)

Major HPC Fabric Components

Host Channel Adapter (HCA) / Intel® OPA Card (Host Fabric Interface, HFI)

- Terminates a Fabric link and executes transport-level functions

Switch

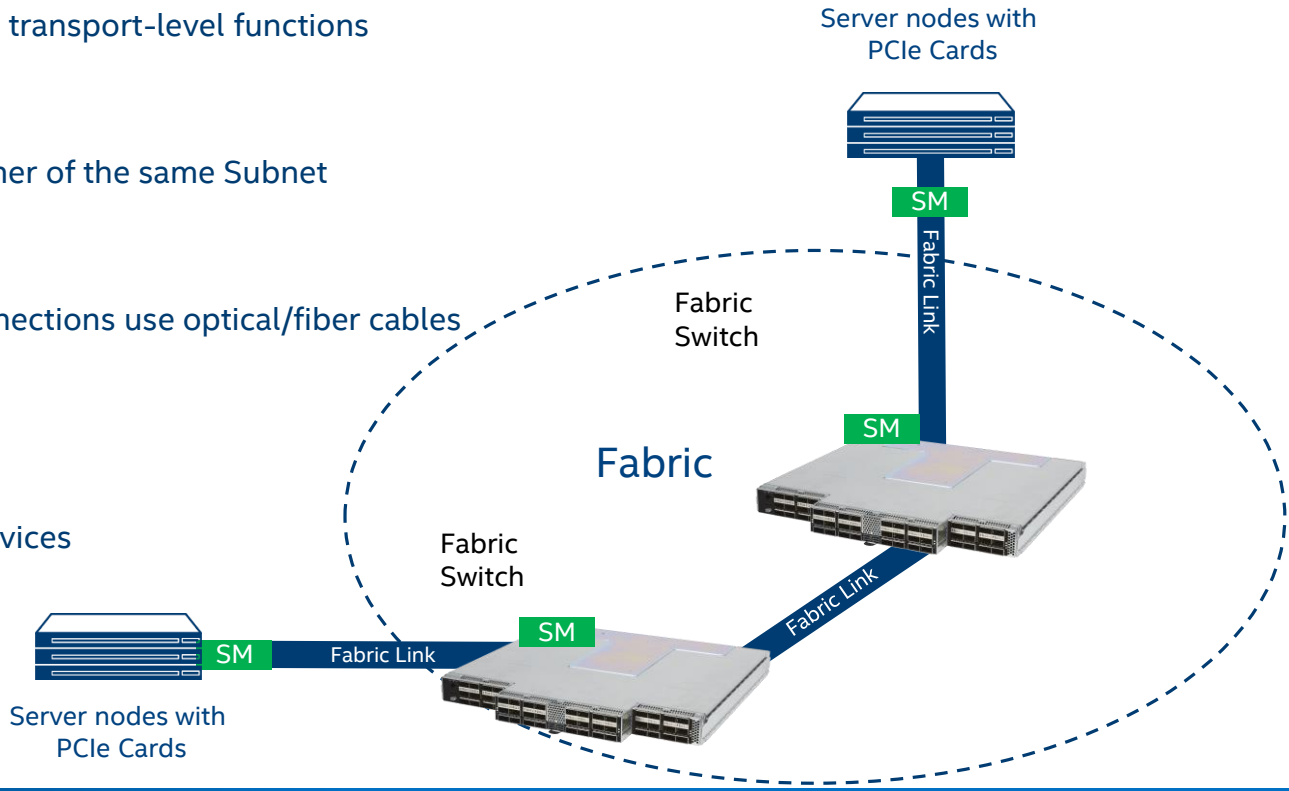
- Routes packets from one link to another of the same Subnet

Cables

- Copper cables are typical, longer connections use optical/fiber cables
- Connectors are QSFP/QSFP28

Subnet Manager

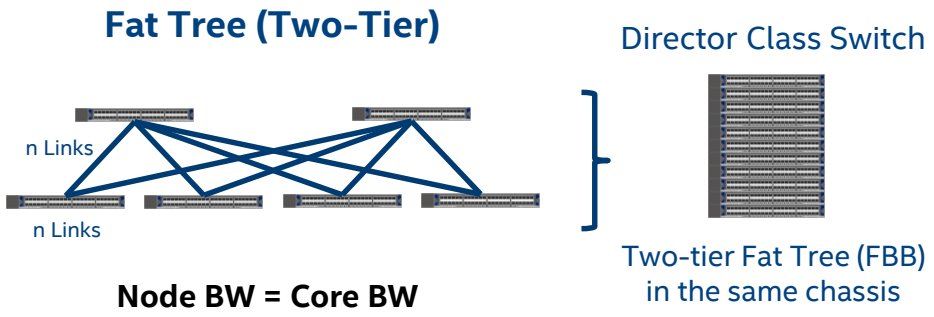
- Discovers and configures attached devices and manages the fabric



HPC Fabric Configurations

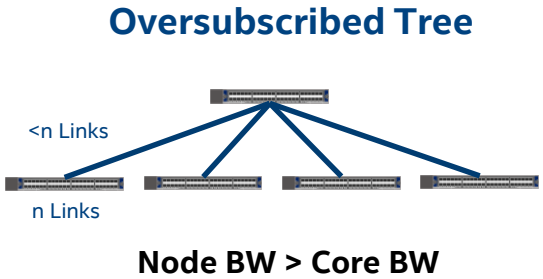
Fat Tree [most popular]:

Network supports Full Bisectional Bandwidth (FBB) between a pair of nodes

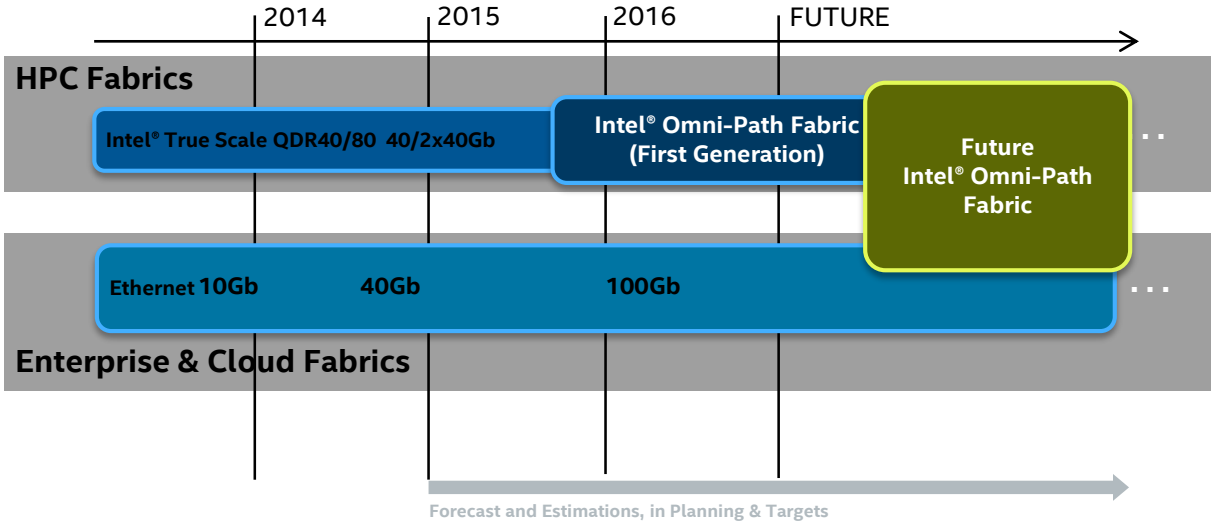


Oversubscribed Fat Tree [next most popular]:

Constant Bisectional Bandwidth (CBB) can be less than FBB between a pair of nodes.



The Intel® Fabric Product Roadmap Vision



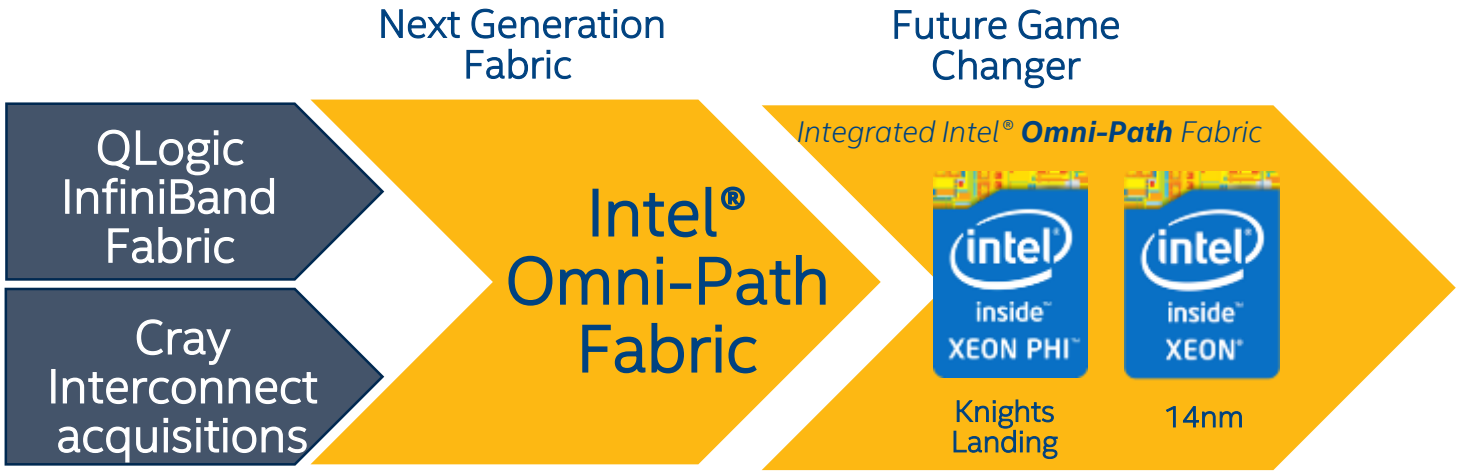
Establish in HPC with the first generation Intel® Omni-Path Architecture
Expand to broader market segments in successive generations

Potential future options, subject to change without notice. All timeframes, features, products and dates are preliminary forecasts and subject to change without further notification.

INTEL® OMNI-PATH FABRIC: 100G OPA

INTEL® 100G OMNI-PATH

EVOLUTIONAY APPROACH, REVOLUATIONARY FEATHERS, END-TO-END PRODUCTS



最大的结合Intel® True Scale fabric and Cray Aries

添加创新的新特性，提高性能，可靠性和QoS

基于现存的 OpenFabrics Alliance* 软件，二进制代码兼容 Infiniband

完整的端到端的产品线

未来芯片集成
→ 降低网络采购成本，简化部署安装成本

Fabric Solutions Powered by Intel® Omni-Path Architecture

PCIe Adapters


Intel Part #	100HFA018LS 100HFA018FS	100HFA016LS 100HFA016FS
Description	Single-port PCIe x8 Adapter, Low Profile and Std Height	Single-port PCIe x16 Adapter, Low Profile and Std Height
Availability[1]	Q2'16	Q2'16
Speed	58 Gbps	100 Gbps
Ports, Media	Single port, QSFP28	Single port, QSFP28
Form Factor	Low profile PCIe Std Height PCIe	Low profile PCIe Std Height PCIe
Features	Passive thermal – QSFP heatsink, supports up to Class 4 max optical transceivers	Passive thermal – QSFP heatsink, supports up to Class 4 max optical transceivers
Sandy Bridge	X	X
Ivy Bridge	X	X
Intel® Xeon® processor E5-2600 v3 (Haswell-EP)	✓	✓
Intel® Xeon® processor E5-2600 v4 (Broadwell-EP)	✓	✓

Edge Switches / Director Switches







Intel Part #	100SWE48UF2 / R2 100SWE48QF2 / R2	100SWE24UF2 / R2 100SWE24QF2 / R2	100SWD24B1N 100SWD24B1D 100SWD24B1A	100SWD06B1N 100SWD06B1D 100SWD06B1A	100SWDLF32Q	100SWDSPINE	100SWDMGTSH
Description	48 Port Edge Switch ("Q" = mgmt card)	24 Port Edge Switch ("Q" = mgmt card)	24-slot Director Class Switch, Base Config	6-slot Director Class Switch, Base Config	Director Class Switch Leaf Module	Director Class Switch Spine Module	Director Class Switch Management Module
Availability[1]	Q2'16	Q2'16	Q2'16	Q2'16	Q2'16	Q2'16	Q2'16
Speed	100 Gbps	100 Gbps	100 Gbps	100 Gbps	100 Gbps	100 Gbps	100 Gbps
Max External Ports	48	24	768	192	32	N/A	N/A
Media	QSFP28	QSFP28	10/100/1000 Base-T USB Gen2	10/100/1000 Base-T USB Gen2	QSFP28	Internal high speed connections	10/100/1000 Base-T USB Gen2
Form Factor	1U	1U	20U	7U	Half-width module 2 modules per leaf	Full width module, 2 boards/module	Half-width module
Features	Forward / reverse airflow and mgmt card options, up to 2 PSU	Forward / reverse airflow and mgmt card options, up to 2 PSU	Up to 2 mgmt modules, up to 12 PSUs, AC and DC options	Up to 2 mgmt modules, up to 6 PSUs, AC and DC options	Hot swappable	96 internal mid-plane connections, hot swappable	N+1 redundancy, hot swappable

Passive Copper Cables

0.5M	1.0M	1.5M	2.0M	3.0M
100CQQF3005 100CQQH3005 (30 AWG)	100CQQF3010 100CQQH3010 (30 AWG)	100CQQH2615 (26 AWG)	100CQQH2620 (26AWG)	100CQQH2630 (26AWG)

Active Optical Cables

3.0M	5.0M	10M	15M	20M	30M	50M	100M
100FRRF0030	100FRRF0050	100FRRF0100	100FRRF0150	100FRRF0200	100FRRF0300	100FRRF0500	100FRRF1000

[1] Production Readiness / General Availability dates

Intel® Omni-Path Edge Switch

100 Series 24/48 Port: Features[1]

Compact Space (1U)
- 1.7”H x 17.3”W x 16.8“L

Switching Capacity
- 4.8/9.6 Tb/s switching capability

Line Speed
- 100Gb/s Link Rate

Standards-based Hardware Connections
- QSFP28

Redundancy
- N+N redundant Power Supplies (optional)
- N+1 Cooling –Fans (speed control, customer changeable forward/reverse airflow)

Management Module (optional)

No externally pluggable FRUs

Power	Copper		Optical (3W QSFP)	
Model	Typical	Maximum	Typical	Maximum
24-Ports	146W	179W	231W	264W
48-Ports	186W	238W	356W	408W

24-port Edge Switch

48-port Edge Switch

This presentation discusses devices that have not been authorized as required by the rules of the Federal Communications Commission, including all Intel® Omni-Path Architecture devices. These devices are not, and may not be, offered for sale or lease, or sold or leased, until authorization is obtained.

[1]Specifications contained in public Product Briefs.

Intel® OPA Director Class Systems 100 Series 6-Slot/24-Slot Systems[1]

Power	Copper		Optical (3W QSFP)	
Model	Typical	Maximum	Typical	Maximum
6-Slot	1.6kW	2.3kW	2.4kW	3.0kW
24-Slot	6.8kW	8.9kW	9.5kW	11.6kW

Highly Integrated

- 7U/20U plus 1U Shelf

Switching Capacity

- 38.4/153.6 Tb/s switching capability

Common Features

- Intel® Omni-Path Fabric Switch Silicon 100 Series (100Gb/s)
- Standards-based Hardware Connections – QSFP28
- Up to Full bisectional bandwidth Fat Tree internal topology
- Common Management Card w/Edge Switches
- 32-Port QSFP28-based Leaf Modules
- Air-cooled, front to back (cable side) air cooling
- Hot-Swappable Modules
  - Leaf, Spine, Management, Fan , Power Supply
- Module Redundancy
  - Management (N+1), Fan (N+1, Speed Controlled), PSU (DC, AC/DC)
- System Power : 180-240AC

6-Slot Director Switch

24-Slot Director Switch

[1]Specifications contained in public Product Briefs.

This presentation discusses devices that have not been authorized as required by the rules of the Federal Communications Commission, including all Intel® Omni-Path Architecture devices. These devices are not, and may not be, offered for sale or lease, or sold or leased, until authorization is obtained.

Intel® Omni-Path Host Fabric Interface

100 Series Single Port[1]

Low Profile PCIe Card

- 2.71”x 6.6“ max. Spec compliant.
- Standard and low profile brackets

Wolf River (WFR-B) HFI ASIC

PCIe Gen3

Single 100 Gb/s Intel® OPA port

- QSFP28 Form Factor
- Supports multiple optical transceivers
- Single Link status LED (Green)

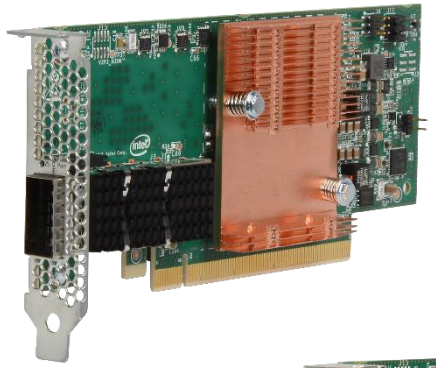
Power

	Copper		Optical (3W QSFP)	
Model	Typical	Maximum	Typical	Maximum
X16 HFI	7.4W	11.7W	10.6W	14.9W
X8 HFI	6.3W	8.3W	9.5W	11.5W

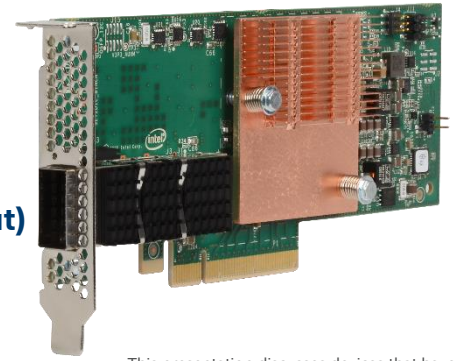
Thermal

- Passive thermal - QSFP Port Heatsink
- Standard 55C, 200lfm environment

[1]Specifications contained in public Product Briefs.



x16 HFI
(100Gb Throughput)

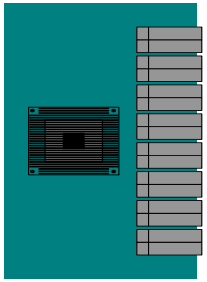


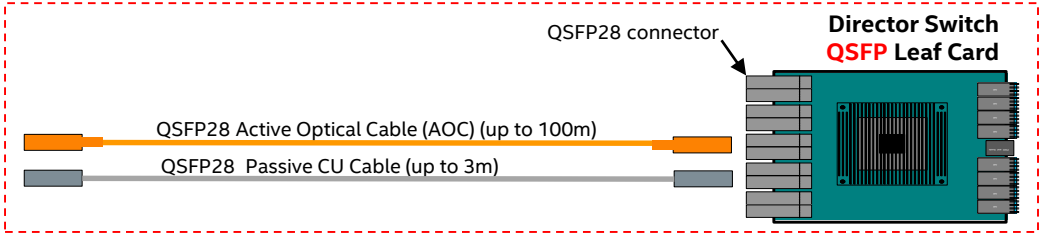
x8 HFI
(~58Gb Throughput)
PCIe Limited

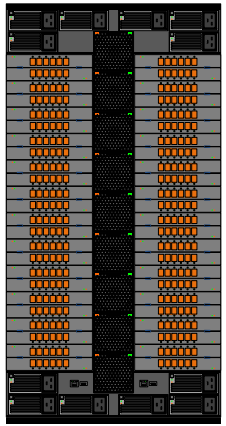
This presentation discusses devices that have not been authorized as required by the rules of the Federal Communications Commission, including all Intel® Omni-Path Architecture devices. These devices are not, and may not be, offered for sale or lease, or sold or leased, until authorization is obtained.

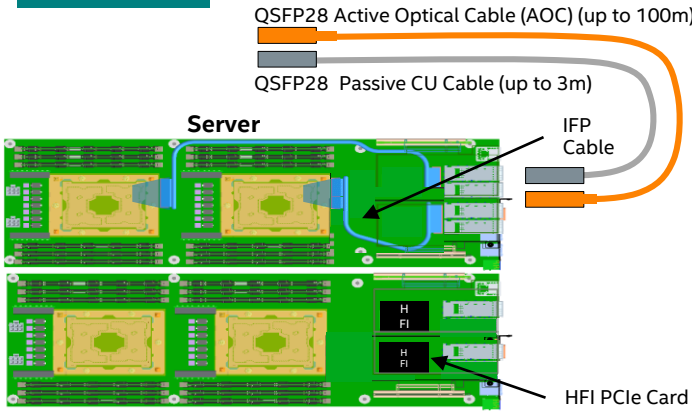
Intel® Omni-Path Architecture Fabric Cabling Topology

Edge Switch
Up to 48p





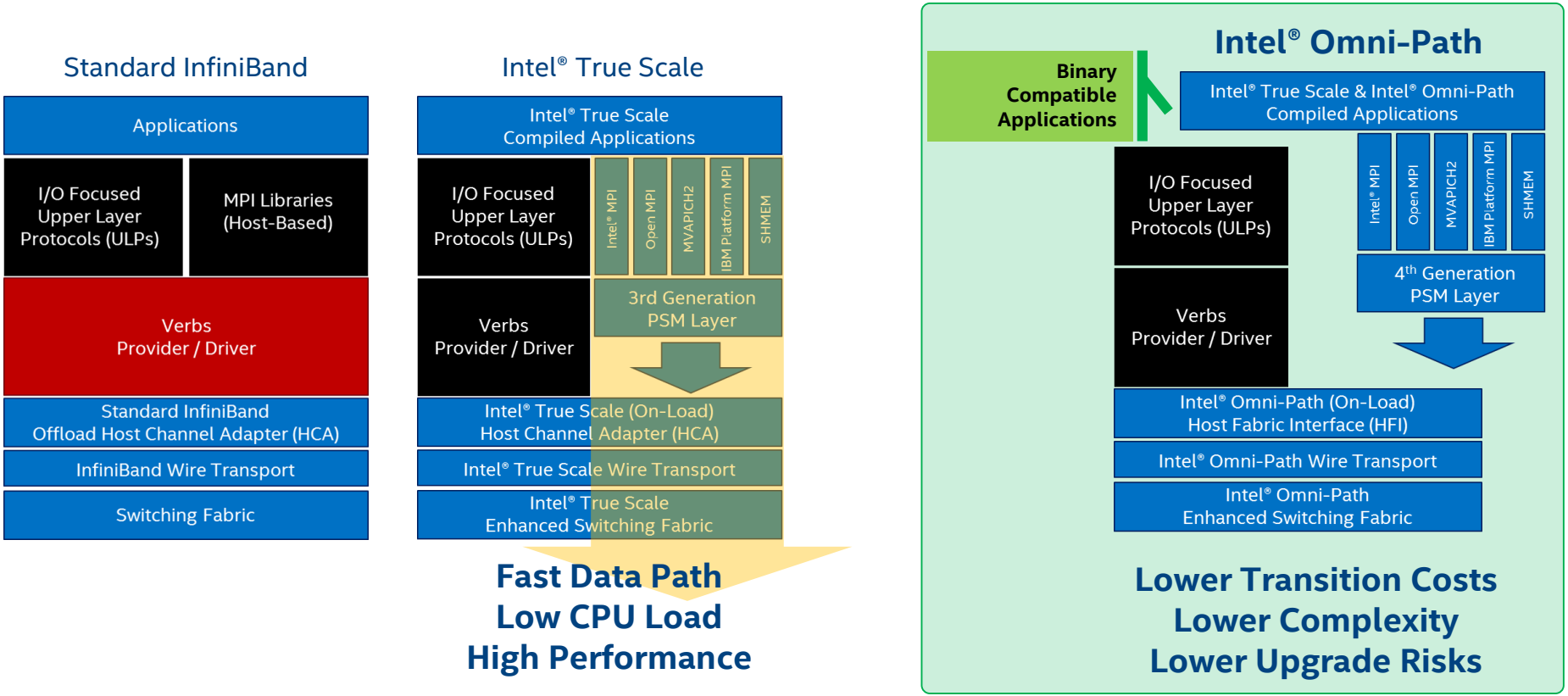




Legend:
Transceiver

Host Layer Optimization:

Optimize HPC Code Path and Generational Compatibility



PERFORMANCE

Intel® OPA MPI Performance Measurements

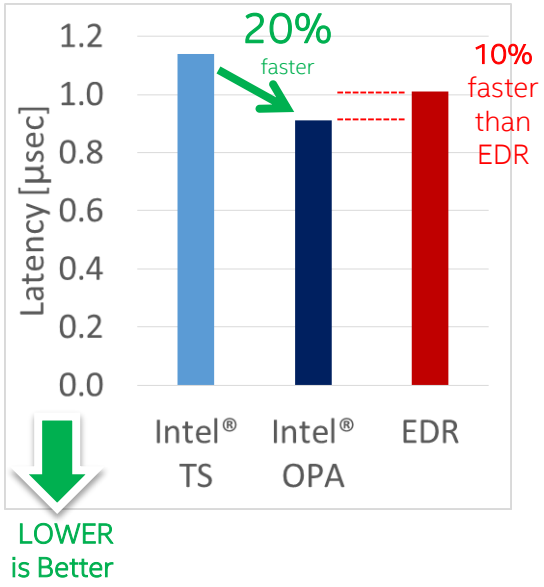
Metric	Intel® Xeon® CPU E5-2697 v3 with Intel® Omni-Path Fabric[1]
LATENCY **OSU Latency Test (8B)**	
Latency (one-way, b2b nodes) [2]	790 ns
Latency (one-way, 1 switch) [2]	900 ns
MESSAGING RATES (rank = rank pairs) **OSU Message Bandwidth Test (8B, streaming)**	
Message Rate (1 rank, uni-dir)[3]	5.3 M msg/s
Message Rate (1 rank, bi-dir)[3]	6.3 M msg/s
Message Rate (max ranks, uni-dir) [3]	108 M msg/s
Message Rate (max ranks, bi-dir) [3]	132 M msg/s
BANDWIDTH (rank = rank pairs) **OSU Message Bandwidth Test (512 KB, streaming)**	
BW (1 rank, 1 port, uni-dir) [3]	12.3 GB/s
BW (1 rank, 1 port, bi-dir) [3]	24.5 GB/s

Intel® Xeon® CPU E5-2699 v4 with Intel® Omni-Path Fabric[4]
143 M msg/s
172 M msg/s

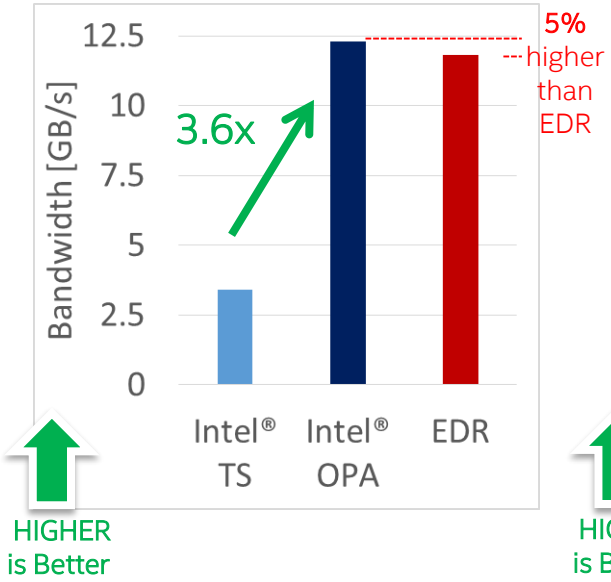
All tests performed by Intel with OSU OMB 4.4.1.
1 Intel® Xeon® processor E5-2697 v3 with Intel® Turbo-Mode enabled. 8x8GB DDR4 RAM, 2133 MHz. RHEL7.0.
2 osu_latency 1-8B msg. w/ and w/out switch. Open MPI 1.10.0-hfi packaged with IFS 10.0.0.0.625.
3 osu_mbw_mr modified for bi-directional bandwidth measurement. w/switch Open MPI 1.10.0-hfi packaged with IFS 10.0.0.0.625. . IOU Non-Posted Prefetch disabled in BIOS. snp_holdoff_cnt=9 in BIOS.
4 Intel® Xeon® processor E5-2699v4 with Intel® Turbo-Mode enabled. 8x8GB DDR4 RAM, 2133 MHz. RHEL7.0. IFS 10.0.0.991.35. Open MPI 1.8.5-hfi. B0 Intel® OPA hardware and beta level software.

Intel® OPA MPI Performance Improvements

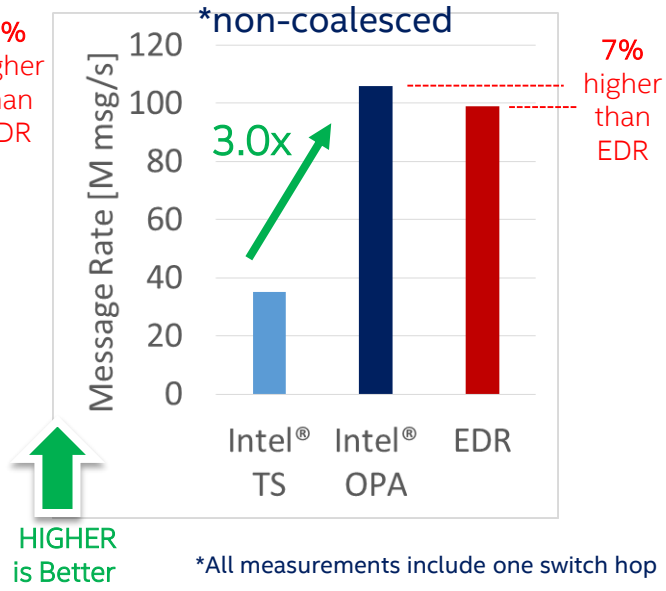
MPI Latency[1]



MPI Bandwidth[2]



MPI Message Rate[3]



*All measurements include one switch hop

Tests performed by Intel on Intel® Xeon® Processor E5-2697v3 dual-socket servers with 2133 MHz DDR4 memory. Turbo mode enabled and hyper-threading disabled. Ohio State Micro Benchmarks v. 4.4.1. Intel OPA: Open MPI 1.10.0 with PSM2. Intel Corporation Device 24f0 – Series 100 HFI ASIC. OPA Switch: Series 100 Edge Switch – 48 port. IOU Non-posted Prefetch disabled in BIOS. EDR: Open MPI 1.8-mellanox released with hpcx-v1.3.336-icc-MLNX_OFED_LINUX-3.0-1.0.1-redhat6.6-x86_64.tbz. MXM_TLS=self,rc tuning. Mellanox EDR ConnectX-4 Single Port Rev 3 MCX455A HCA. Mellanox SB7700 - 36 Port EDR Infiniband switch. Intel® True Scale: Open MPI. QLG-QLE-7342(A), 288 port True Scale switch. 1. osu_latency 8 B message. 2. osu_bw 1 MB message. 3. osu_mbw_mr, 8 B message (uni-directional), 28 MPI rank pairs

ANSYS*

Fluent* 17 Computational Fluid Dynamics

*"Thanks to Intel® **OPA** and the latest Intel® **Xeon® E5-2600 v4** product family, ANSYS Fluent* is able to achieve performance levels **beyond our expectations**. Its unrivaled performance enables our customers to simulate higher-fidelity models without having to expand their cluster nodes ."*[1]

***Dr. Wim Slagter* – Director of HPC and cloud marketing, ANSYS**

- Intel® Omni-Path Architecture (Intel® OPA) is a powerful low latency communications interface specifically designed for High Performance Computing.
- Cluster users will get better utilization of cluster nodes through better scaling.
- Cluster performance means better time-to-solution on CFD simulations.
- Coupled with Intel® MPI, and utilizing standard Fluent runtime options to access TMI, Fluent is ready and proven for out-of-the-box performance on Intel OPA-ready clusters.

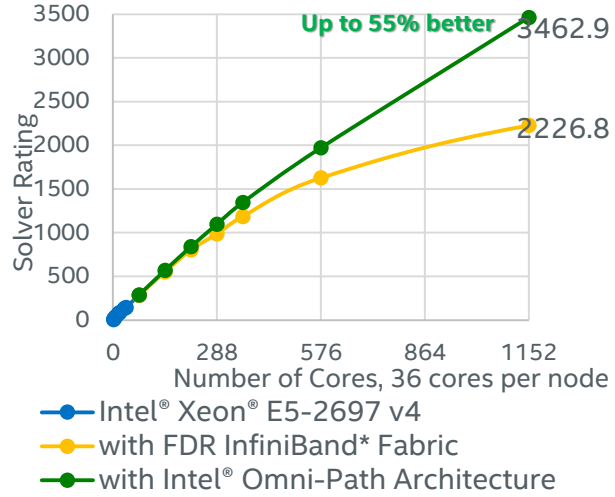
Up to 55% performance advantage with Intel® OPA compared to FDR fabric on a 32 node cluster

ANSYS

www.ansys.com

Technical Computing

ANSYS Fluent* 17 solver rating increased by up to 1.55X with Intel® Omni-Path Architecture scaling on a 32-node cluster

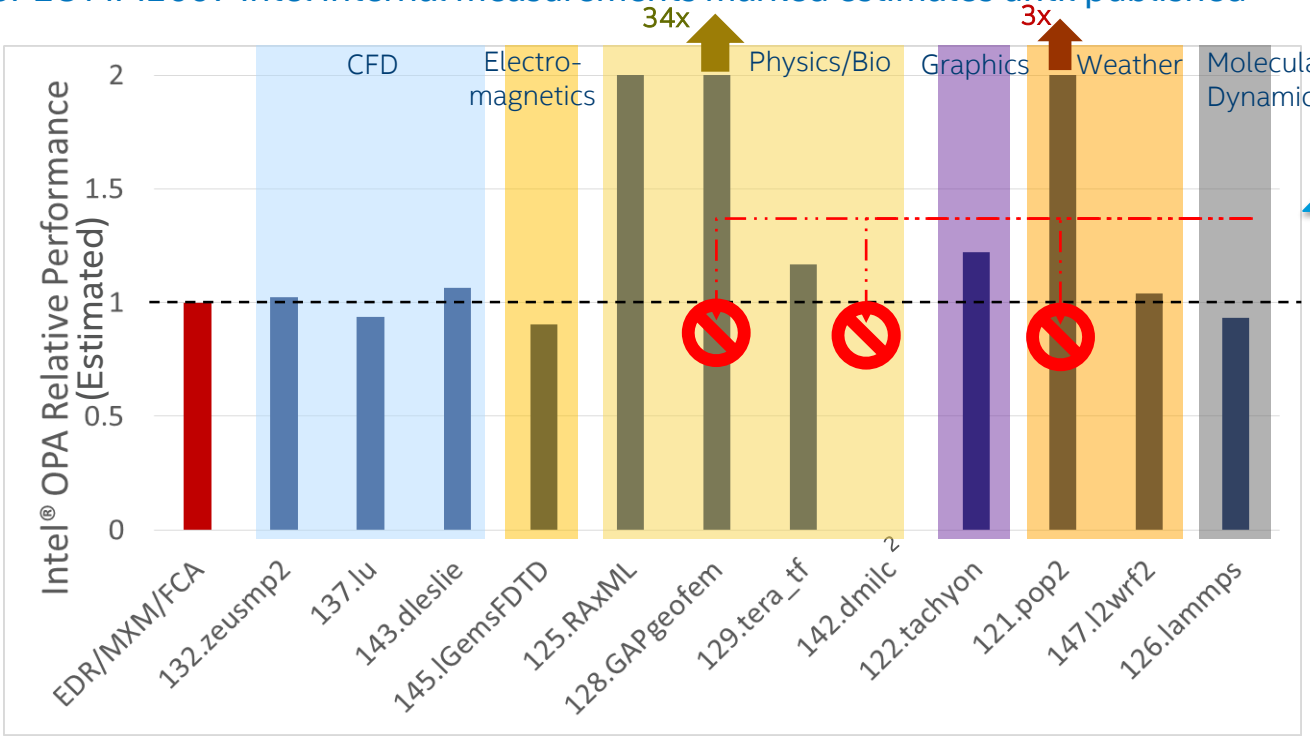


Workload: 12 million cell combustor model, part of the Fluent benchmarking suite. Fluent 17.0

1 - Testing conducted on ISV* software on 2S Intel® Xeon® Processor E5-2697 v4 comparing Intel® OPA to FDR InfiniBand* fabric. Testing done by Intel. For complete testing configuration details, go here.
Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance.

Real Application Performance* - Intel® OPA vs EDR/MXM-FCA

*SPEC MPI2007 Intel internal measurements marked estimates until published



HIGHER is Better

Up to 12% Higher Performance[1]

Does not include three indicated workloads

16 nodes
448 MPI ranks

[1] Overall advantage does not include 121.pop2, 128.GAPgeofem, or 142.dmilc, for which EDR has significant performance/execution problems

[2] 142.dmilc does not run with EDR/Open MPI 1.8-mellanox released with hpcx-v1.3.336-icc-MLNX_OFED_LINUX-3.0-1.0.1-redhat6.6-x86_64.tbz

Tests performed by Intel on Intel® Xeon® Processor E5-2697v3 dual-socket servers with 2133 MHz DDR4 memory. 16 nodes/448 MPI ranks. Turbo mode and hyper-threading disabled. Intel® OPA: Intel Corporation Device 24f0 – Series 100 HFI ASIC. OPA Switch: Series 100 Edge Switch – 48 port. OPA: Open MPI 1.10.0 with PSM2. Mellanox EDR based on internal measurements: Open MPI 1.8-mellanox released with hpcx-v1.3.336-icc-MLNX_OFED_LINUX-3.0-1.0.1-redhat6.6-x86_64.tbz. Mellanox EDR ConnectX-4 Single Port Rev 3 MCX455A HCA. Mellanox SB7700 - 36 Port EDR Infiniband switch. SPEC MPI2007, Large suite, https://www.spec.org/mpi/

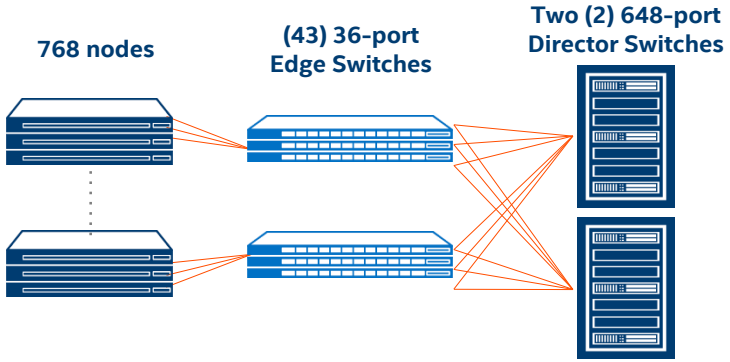
COST BENEFITS

Intel® Omni-Path Fabric's **48 Radix Chip**

It's more than just a 33% increase in port count over a 36 Radix chip

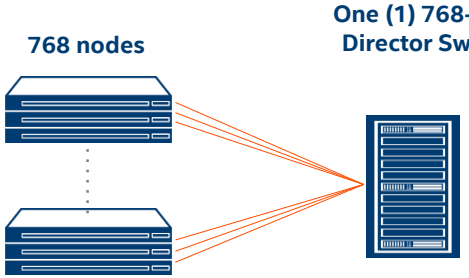
InfiniBand* EDR (36-port Switch Chip)

FIVE-hop Fat Tree



Intel® Omni-Path Architecture (48-port)

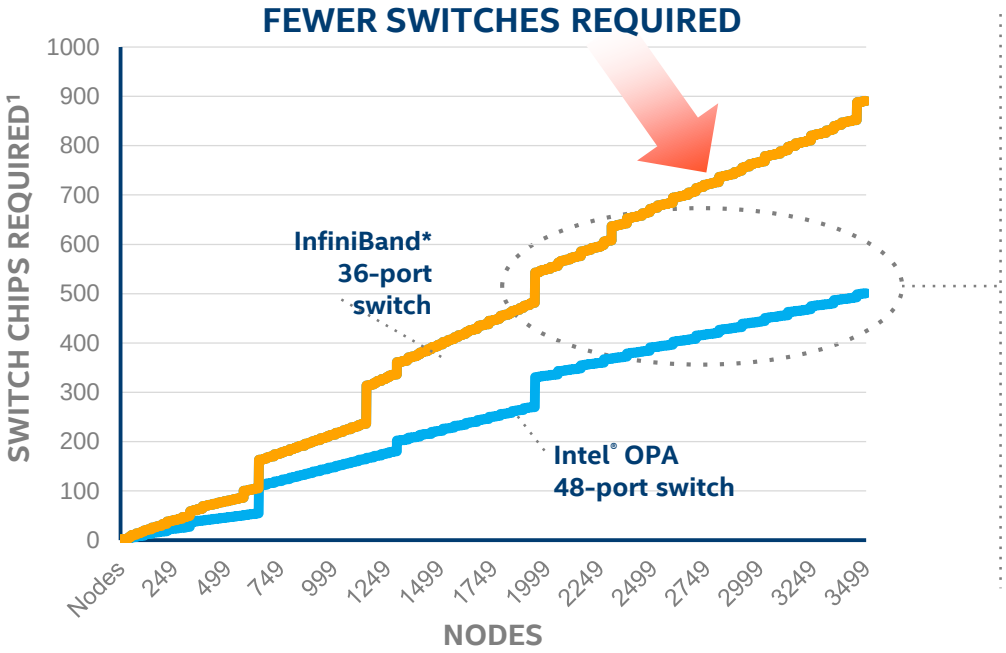
THREE-hop Fat Tree

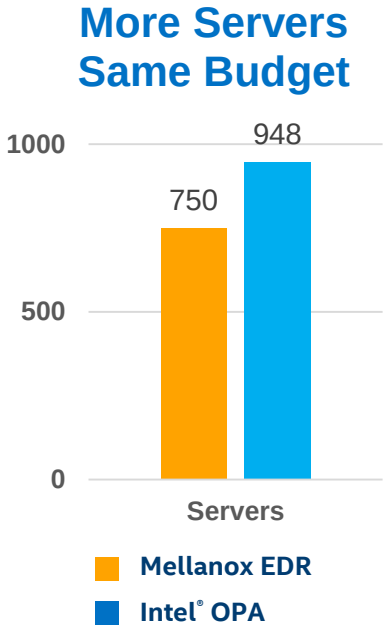


			% Reduction
(43) 36-port	**Edge Switches**	Not required	**100%**
1,542	**Cables**	768	**50%**
99u (2+ racks)	**Rack Space**	20u (<½ rack)	**79%**
~680ns (5 hops)	**Switch Latency[1]**	300-330ns[2] (3 hops)	**51-55%**

[1] Latency numbers based on Mellanox CS7500 Director Switch and Mellanox SB7700/SB7790 Edge switches. See www.Mellanox.com for more product information.

Are You Leaving **Performance** on the Table?



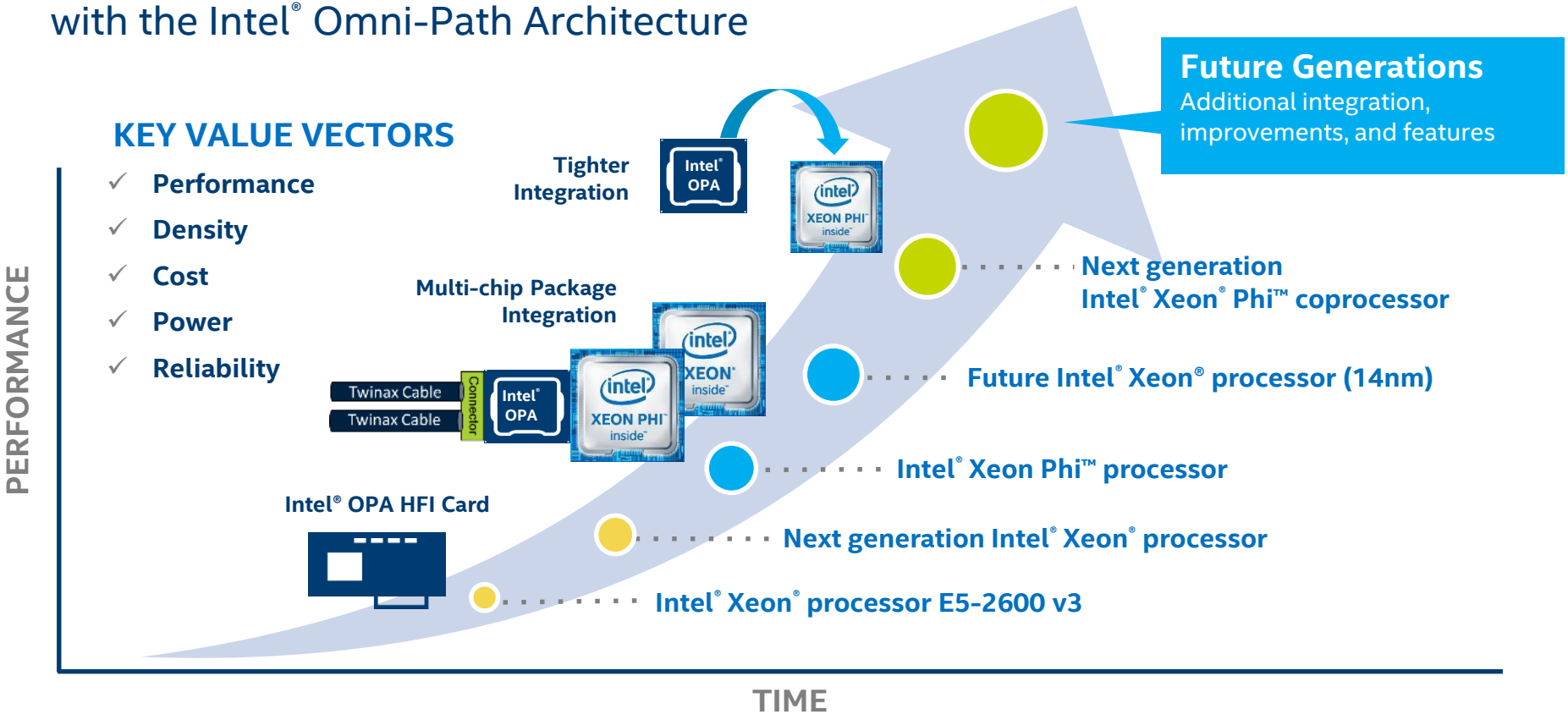


Up to **26%** more servers[1]

[1] Configuration assumes a 750-node cluster, and number of switch chips required is based on a full bisectional bandwidth (FBB) Fat-Tree configuration. Intel® OPA uses one fully-populated 768-port director switch, and Mellanox EDR solution uses a combination of 648-port director switches and 36-port edge switches. Mellanox component pricing from www.kernelsoftware.com, with prices as of November 3, 2015. Compute node pricing based on Dell PowerEdge R730 server from www.dell.com, with prices as of May 26, 2015. Intel® OPA pricing based on estimated reseller pricing based on projected Intel MSRP pricing at time of launch. * Other names and brands may be claimed as property of others.

CPU-Fabric Integration

with the Intel® Omni-Path Architecture



Intel® OPA HFI Option Comparison

	PCIe Card x8 (Chippewa Forest)	PCIe Card x16 (Chippewa Forest)	Knights Landing-F	Skylake-F (single –F CPU populated)	Skylake-F (two –F CPUs populated)	Notes
Ports per node	1	1	2	1	2	▪ Assumes single CHF card populated, although multiple cards in a single node is supported
Peak bandwidth	7.25 GB/s	12.5 GB/s	25 GB/s	12.5 GB/s	25 GB/s	▪ Total platform bandwidth
Latency	1 us	1 us	1 us	1 us	1 us	▪ No measurable difference in MPI latency expected since both use a PCIe interface
CPU TDP adder	n/a	n/a	15W	0W, 10W, 15W	0W, 10W, or 15W	▪ TDP adder per socket, dependent on SKL-F SKU
Power	▪ 6.3W typ ▪ 8.3W max	▪ 7.4W typ ▪ 11.7W max	n/a	n/a	n/a	▪ Estimated power numbers with passive Cu cables
PCIe slot required?	Yes	Yes	No	No	No	▪ Custom mezz card mechanically attached to board or chassis. Requires power and sideband cables
PCIe slot option	Low profile x8 PCIe slot, or custom mezz card	Low profile x16 PCIe slot, or custom mezz card	PCIe carrier card with x4 PCIe connector	PCIe carrier card with x4 PCIe connector	PCIe carrier card with x4 PCIe connector	▪ SKL-F (dual –F CPU) can use a single 2-port PCIe carrier card, similar to KNL PCIe carrier card ▪ Carrier card requires a PCIe connector routed for power, but not necessarily routed for PCIe signals
PCIe lanes used (on board)	8	16	32 [4 lanes available]	0	0	▪ SKL-F includes dedicated PCIe lanes for –F SKUs ▪ Assumes PCIe carrier card uses a x4 PCIe slot only routed for power and not PCIe signals

TECHNOLOGY COMPARISONS

Product Comparison Matrix

Feature	Intel® Omni-Path	EDR	Notes
Switch Specifications			
Link Speed (QSFP28)	100Gb/s	100Gb/s	Same Speed
Port Count: Director - Edge -	192, **768 (66% more per 1U)** 48, 24	216, 324, **648** 36	**+ 18.5% Ports** **+ 33% Ports**
Latency: Director - Edge -	**300-330ns** (Includes PIP) 100-110ns (Includes PIP)	<500ns[1] (Should be 3 x 90ns?) **90ns[1] (FEC Disabled)**	**Up to 32% Advantage** FEC increases power up to 50% per port
Redundant Power/Cooling	Yes (Director AC and/or AC-DC Power)	Yes	
Packet Rate Per Port: Switch Host	195M msg/sec **160M** msg/sec (CPU Dependent)	150/195M msg/sec - Switch-IB/Switch-IB 2 150M msg/sec	Mellanox claims are not for MPI Messages. Most HPC applications use MPI as transport
Power Per Port ***(Typical Copper)***[2] : – 24/18-Slot Director – 48/36-Port Edge (M) – 48/36-Port Edge (U)	 **~8.85 Watts** 3.87 W **3.48 W**	 14.1 Watts 3.78 W 3.78 W	 **37.2% Lower Power** EDR Power for FEC and Mgmt Card missing EDR Power for FEC missing
Director Leaf Module: Size/Qty	32 / **(24-Slot),** (6-Slot)	**36** / (18-Slot), (6-Slot)	**+33% modules in single large director**
Largest 2 Tier Fabric (Edge/Director)	18,432	11,664	~1.6x (QSFP28)
Host Adapter Specifications			
Host Adapter Model	Intel® OPA 100 Series (HFI)	HCA (ConnectX-4)	
Protocol	Intel® OPA	InfiniBand	
Speed Support (Host)	x16 = 100Gb/s – x8 = 58Gb/s	All Prior IB Speeds[1]	CX4 includes a rate locked FDR version[1]
Power Per Port ***(Typical Copper)***[2] : – 1-Port x16 HFI – 1-Port x8 HFI	 **7.4 W Copper** **6.3 W Copper**	13.9 W Copper	**46.7% Lower Power**

[1] Mellanox Datasheets: December, 19 2015 [2] Power ratings assume fully loaded systems

Intel® Omni-Path High Level Feature Comparison Matrix

Features	Intel® OPA	EDR	Notes
Link Speed	100Gb/s	100Gb/s	Same Link Speed
Switch Latency – Edge/DCS	100-110ns/300-330ns	90ns/~500ns	Intel® OPA includes "Load-Free" error detection • Application Latency Most important
MPI Latency (OSU pt2pt)	Less Than 1µs	~1µs	Similar 1 Hop Latency • Intel's OPA HFI improves with each CPU generation
Link Enhancements – Error Detection/Correction	Packet Integrity Protection (PIP)	FEC/Link Level Retry	Intel OPA is a HW detection solution that adds **no latency or BW penalty**
Link Enhancements – Data Prioritization across VLs	Traffic Flow O… (TFO)		Over and above VL prioritization. Allows High priority traffic to preempt in-flight low priority traffic (~15% performance improvement)
Link Enhancements – Graceful Degradation	Dynamic Lane Scaling (DLS)	No	Non-Disruptive Lane(s) failure. Supports asymmetrical traffic pattern. Avoids total shutdown,
RDMA Support	Yes	Yes	RDMA underpins verbs. Intel® OPA supports verbs. TID RDMA brings Send/Receive HW assists for RDMA for larger messages
Built for MPI Semantics	Yes – PSM (10% of code)	No - Verbs	Purpose designed for HPC
Switch Radix	48 Ports	36 Ports	Higher Radix means less switches, power, space etc.
Fabric Router	No	Future	Limited need to connect to older fabric technologies except for storage – Still not available

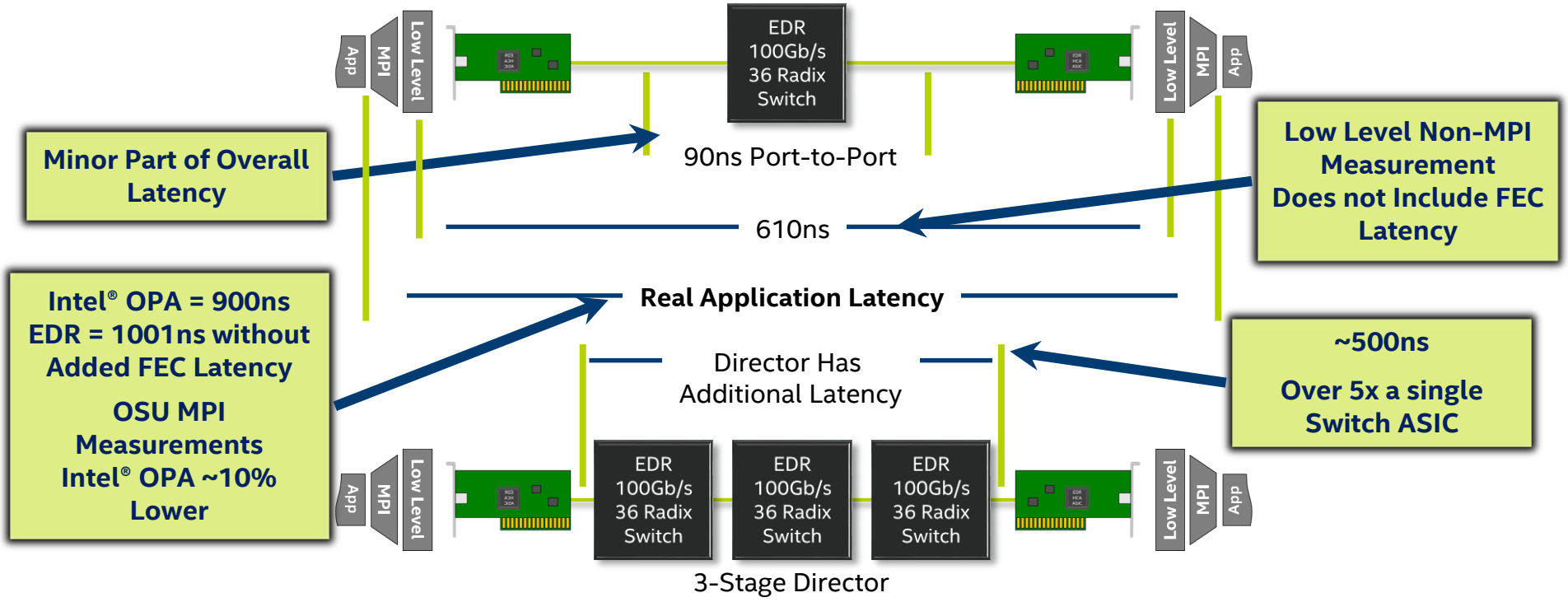
Multi-Stage Fabric Latency Protection

EDR Source: Publically Available Data
OPA Features: Based on Design

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration.

Switch Latency

Understanding Switch Latency Comparisons



Tests performed by Intel on Intel® Xeon® Processor E5-2697v3 dual-socket servers with 2133 MHz DDR4 memory. Turbo mode enabled and hyper-threading disabled. Ohio State Micro Benchmarks v. 4.4.1. Intel OPA: Open MPI 1.10.0 with PSM2. Intel Corporation Device 24f0 – Series 100 HFI ASIC. OPA Switch: Series 100 Edge Switch – 48 port. IOU Non-posted Prefetch disabled in BIOS. EDR: Open MPI 1.8-mellanox released with hpcx-v1.3.336-icc-MLNX_OFED_LINUX-3.0-1.0.1-redhat6.6-x86_64.tbz. MXM_TLS=self,rc tuning. Mellanox EDR ConnectX-4 Single Port Rev 3 MCX455A HCA. Mellanox SB7700 - 36 Port EDR InfiniBand switch 1. osu_latency 8 B message.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance.

RDMA Support

Intel® Omni-Path Architecture (Intel® OPA)
RDMA Support

Intel® OPA has always supported RDMA Functions for MPI-Based applications via PSM

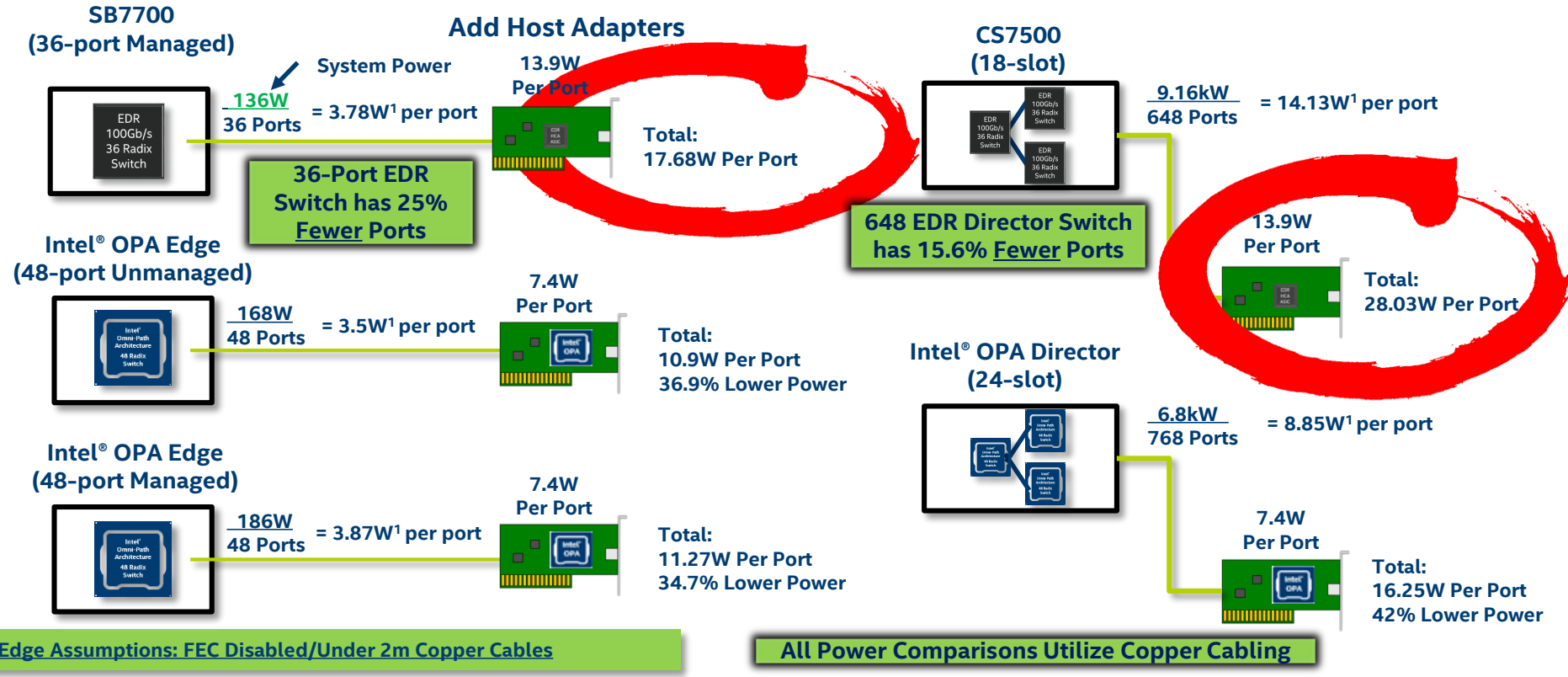
- 16 Send DMA (SDMA) engines and Automatic Header Generation provide HW-assists for offloading large message processing from the CPU

Intel® OPA supports RDMA for Verbs I/O

- RDMA is the underlying protocol for Verbs
- Storage runs over verbs
- Additional performance enhancements are coming
- 8K MTU supported to further reduce CPU interrupts for I/O

Power Usage

Intel® OPA vs. EDR: End-to-End Power Comparison:



[1]Assumes that all switch ports are utilized. All power measurements are typical. All Mellanox power from 12/23/15 documents located a www.mellanox.com. Mellanox Switch 7790 power from datasheet. Host Adapter power from ConnectX®-4 VPI Single and Dual Port QSFP28 Adapter Card User Manual page 45. CS7500 Director power from 648-Port EDR InfiniBand Switch-IB™ Switch Platform Hardware User Manual page 75

OPA视频

Proven Technology Required for Today's Bids:

Intel® OPA is **the Future** of High Performance Fabrics

QLOGIC intel Aries

Highly Leverages
existing Intel, Aries and Intel® True Scale technologies

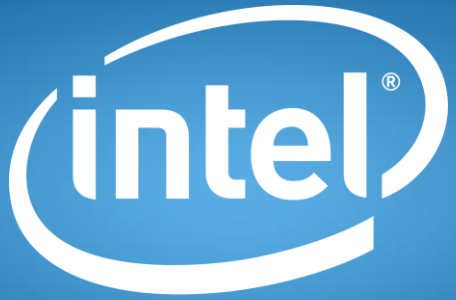
Leading Edge Integration
with Intel® Xeon® processor and Intel® Xeon Phi™ processor

Innovative Features
for high fabric performance, resiliency, and QoS

Open Source software and supports standards like the **OpenFabrics Alliance***

Robust Ecosystem
of trusted computing partners and providers

*Other names and brands may be claimed as property of others.



experience
what's inside™