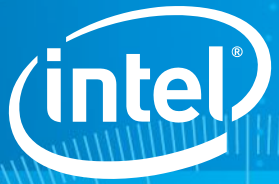
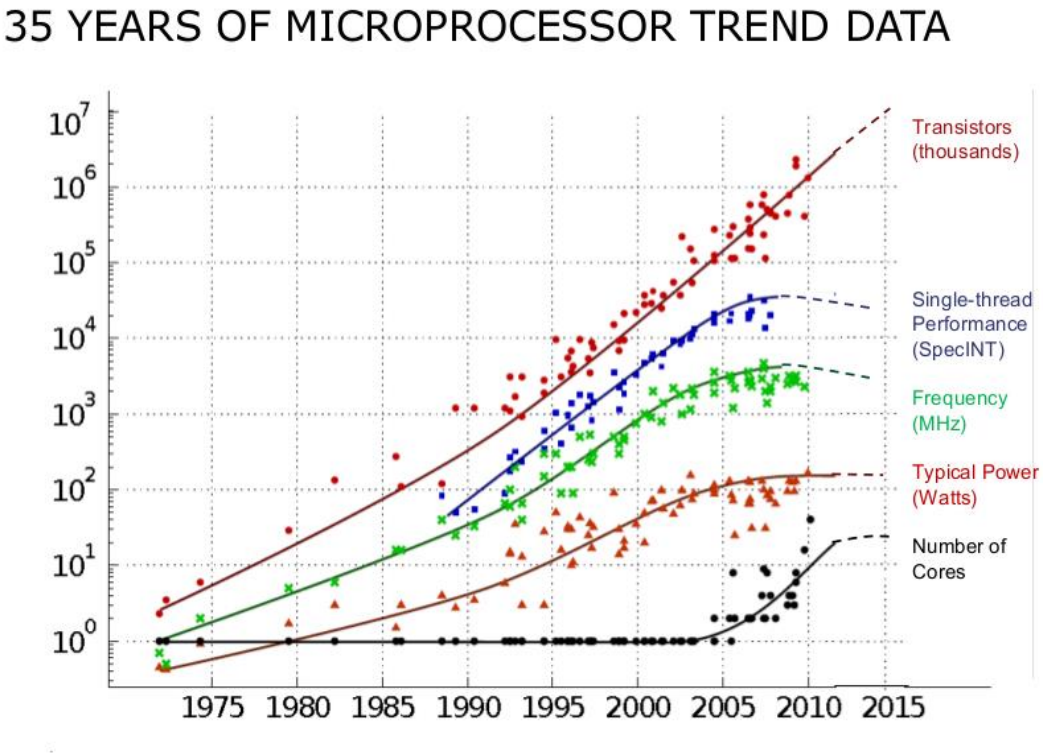


High Performance Computing Ecosystem and Trends

4 July 2016

Nikolay Mester

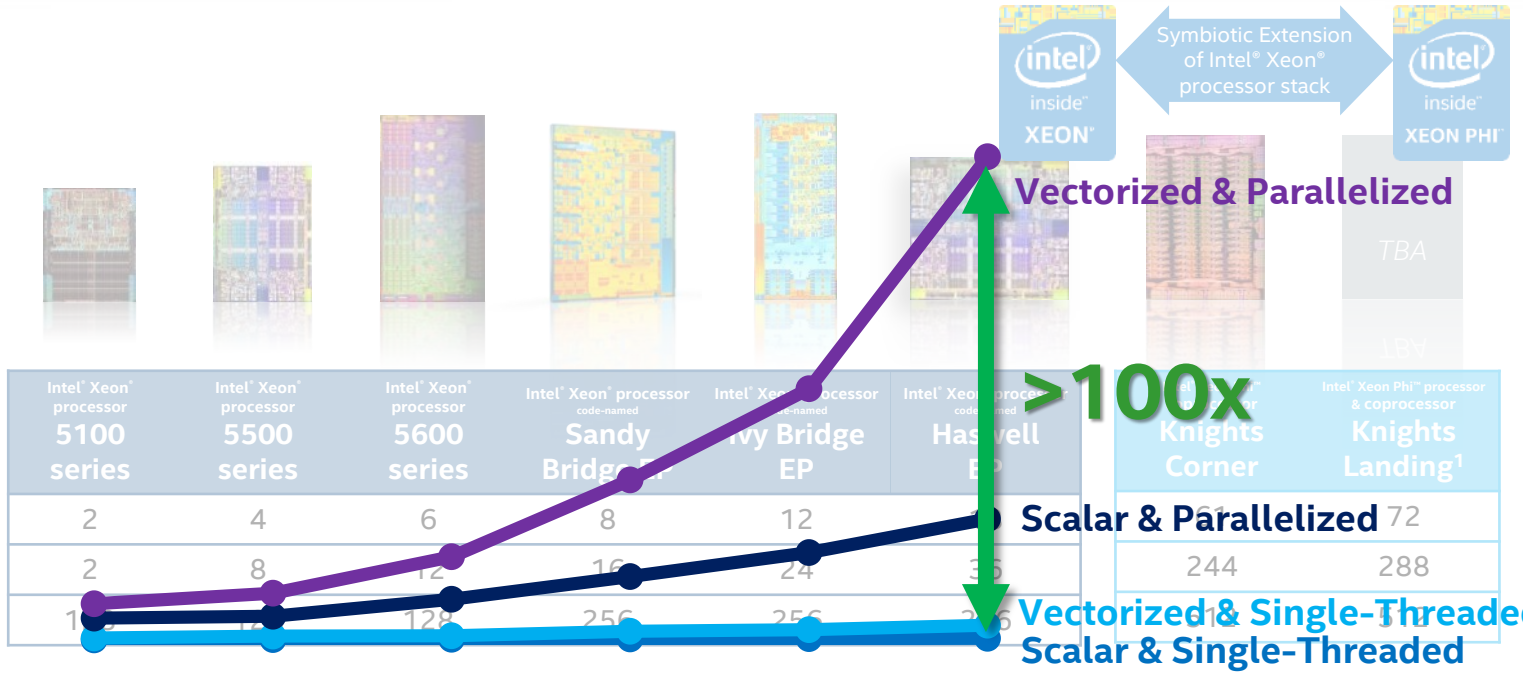
Moore's Law and Parallelism



Original data collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond and C. Batten
Dotted line extrapolations by C. Moore

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

CPU Parallelism is Already a MUST



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. Configurations: Intel Performance Projections as of Q1 2015. For more information go to http://www.intel.com/performance . Results have been estimated based on internal Intel analysis and are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance. Copyright © 2015, Intel Corporation

Chart illustrates relative performance of the Binomial Options DP workload running on an Intel® Xeon® processor from the adjacent generation.

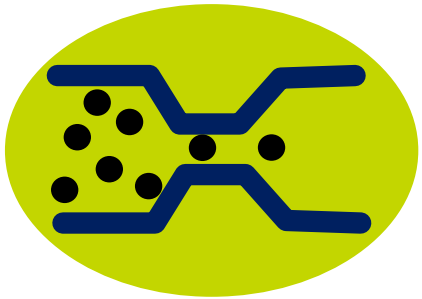
*Product specification for launched and shipped products available on ark.intel.com.

[1]Not launched

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

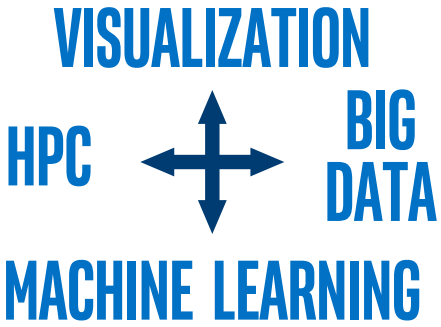
Growing Challenges in HPC

"The Walls" System Bottlenecks



Memory | I/O | Storage
Energy Efficient Performance
Space | Resiliency |
Unoptimized Software

Divergent Infrastructure



Resources Split Among Modeling and Simulation | Big Data Analytics | Machine Learning | Visualization

Barriers to Extending Usage

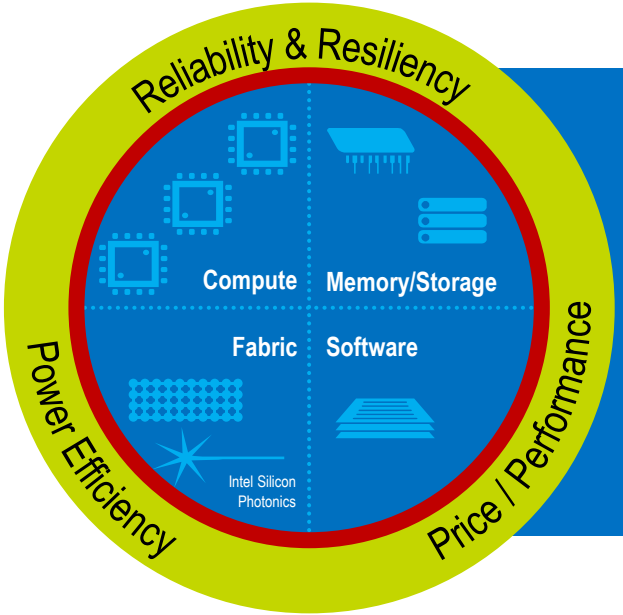
Democratization at Every Scale | Cloud Access | Exploration of New Parallel Programming Models

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

intel

Intel® Scalable System Framework

A Holistic Design Solution for All HPC Needs



Small Clusters Through Supercomputers

Compute and Data-Centric Computing

Standards-Based Programmability

On-Premise and Cloud-Based

Intel® Xeon® Processors
Intel® Xeon Phi™ Processors
Intel® Xeon Phi™ Coprocessors
Intel® Server Boards and Platforms

Intel® Solutions for Lustre*
Intel® Optane™ Technology
3D XPoint™ Technology
Intel® SSDs

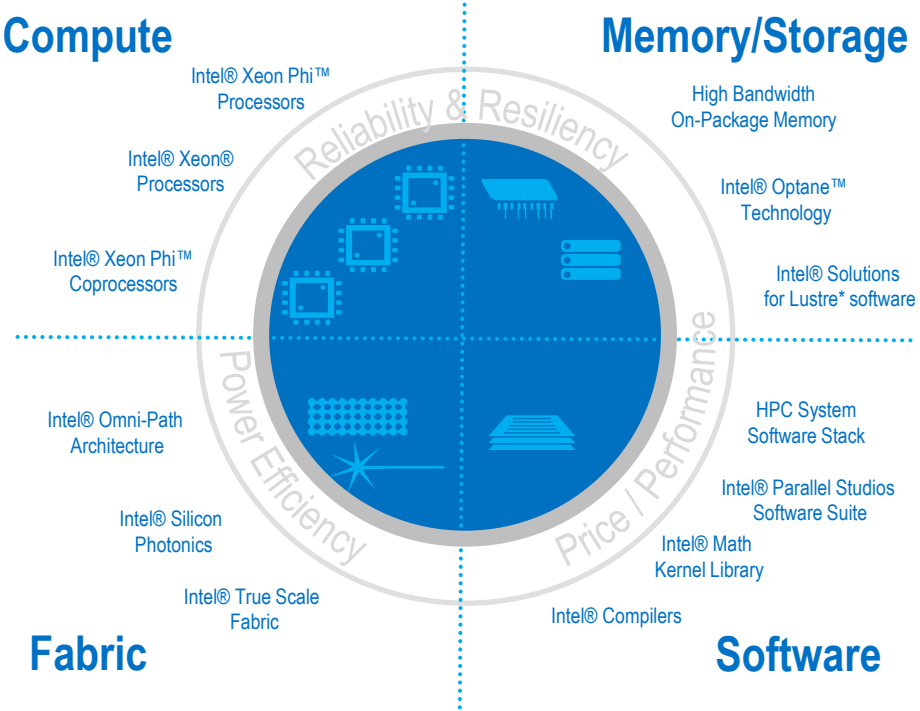
Intel® Omni-Path Architecture
Intel® True Scale Fabric
Intel® Ethernet
Intel® Silicon Photonics

HPC System Software Stack
Intel® Software Tools
Intel® Cluster Ready Program
Intel Supported SDVis

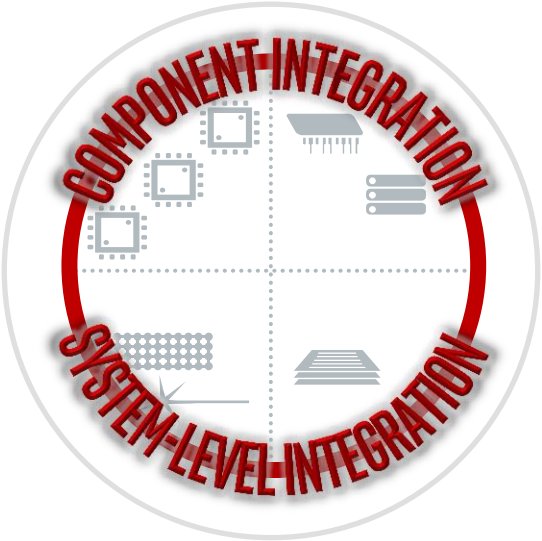
Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

How It Works

Innovative Technologies



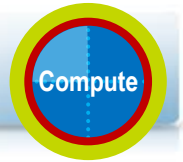
Tighter Integration
and Co-Design



Increased System Density
Reduced System Power Consumption

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

High Performance Compute



Intel® Scalable System Framework

Intel® Xeon Phi™ Processors

Optimized for **Highly Parallel** and **Highly Vectorized** Apps

Intel® Xeon® Processors

Optimized for **Serial and Parallel** Applications

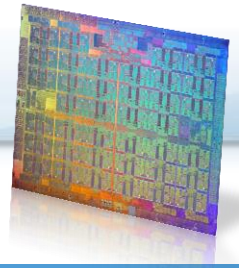
Common Programming Model

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

intel

Intel® Xeon Phi™ x200 Product Family

(codename Knights Landing)



Intel® Scalable System Framework

Platform Memory

Up to 384 GB DDR4

Compute

- Intel® Xeon® Processor Binary-Compatible
- **3+ TFLOPS, 3X ST** (single-thread) perf. vs KNC
- 2D Mesh Architecture
- Out-of-Order Cores

On-Package Memory

- Up to **16 GB** *at launch*
- **5X** Bandwidth vs DDR (over 400GB/s)[2]

Fabric (optional)

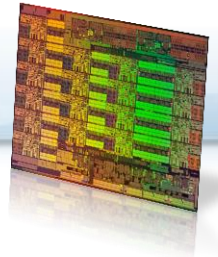
- 1st Intel processor to integrate

1. Over 3 Teraflops of peak theoretical double-precision performance is preliminary and based on current expectations of cores, clock frequency and floating point operations per cycle.
2. Projected result based on internal Intel analysis of STREAM benchmark using a Knights Landing processor with 16GB of ultra high-bandwidth versus DDR4 memory only with all channels populated.
3. Source: Intel internal information

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Intel® Xeon® Processors

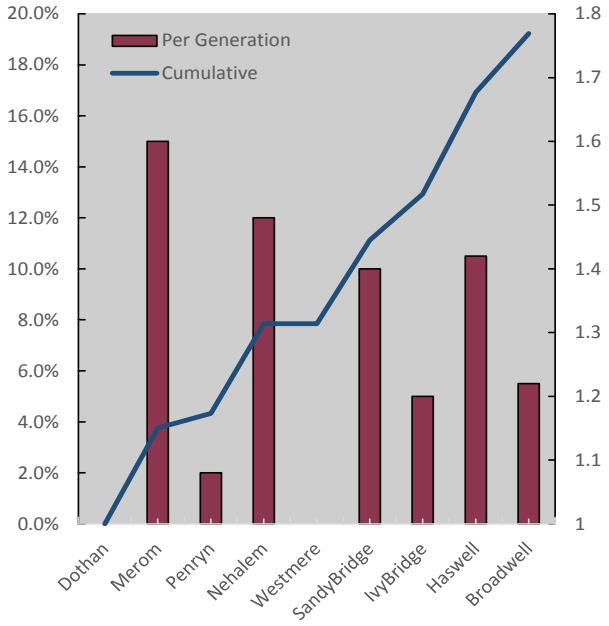
At the Heart of Intel® Scalable System Framework



Intel® Scalable System Framework

Feature	Xeon E5-2600 v3 (Haswell-EP, 22nm)	Xeon E5-2600 v4 (Broadwell-EP, 14nm)
Cores Per Socket	Up to 18	**Up to 22**
Threads Per Socket	Up to 36 threads	**Up to 44 threads**
Last-level Cache (LLC)	Up to 45 MB	**Up to 55 MB**
QPI Speed (GT/s)	2x QPI 1.1 channels 6.4, 8.0, 9.6 GT/s	
PCIe* Lanes/ Controllers/Speed(GT/s)	40 / 10 / PCIe* 3.0 (2.5, 5, 8 GT/s)	
Memory Population	4 channels of up to 3 RDIMMs or 3 LRDIMMs	**+ 3DS LRDIMM&**
Max Memory Speed	Up to 2133	**Up to 2400**
TDP (W)	160 (Workstation only), 145, 135, 120, 105, 90, 85, 65, 55	

Core Single Thread IPC Performance



THE HEART OF THE DATA CENTER

Requires BIOS and firmware update
& 3D Stacked DIMMS depend on market availability

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Intel® Xeon® Processor E5 v4 Family: Core Improvements

Intel® Scalable System Framework

Extract more parallelism in scheduling uops

- Reduced instruction latencies (ADC, CMOV, PCLMULQDQ)
- Larger out-of-order scheduler (60->64 entries)
- New instructions (ADCX/ADOX)

Improved performance on large data sets

- Larger L2 TLB (1K->1.5K entries)
- New L2 TLB for 1GB pages (16 entries)
- 2nd TLB page miss handler for parallel page walks

Improved address prediction for branches and returns

- Increased Branch Prediction Unit Target Array from 8 ways to 10

Floating Point Instruction performance improvements

- Faster vector floating point multiplier (5 to 3 cycles)
- 1024 Radix divider for reduced latency, increased throughput
- Split Scalar divides for increased parallelism/bandwidth
- Faster vector Gather

All products, computer systems, dates and figures specified are preliminary based on current expectations, and are subject to change without notice.
Intel may make changes to specifications and product descriptions at any time, without notice

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

intel

High Performance Memory and Storage

Intel® Scalable System Framework

High-Bandwidth Memory

Configurable Modes

Integrated into the Processor

Intel® Optane™ Technology

(built with 3D XPoint™ Technology)

SSDs

DIMMs

Intel Solutions for Lustre* Software

The **Most Widely Used** File System for **HPC**

New Technologies Are Bringing Memory Closer to Compute

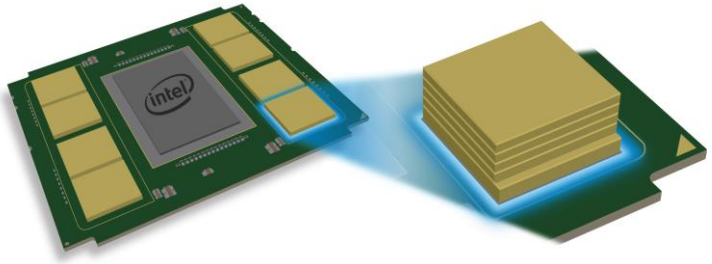
Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Bringing Memory Back Into Balance

Memory

Intel® Scalable System Framework

up to 16 GB of High Bandwidth on-package memory in Knights Landing



3 Modes of Operation:

Flat Mode: Acts as Memory

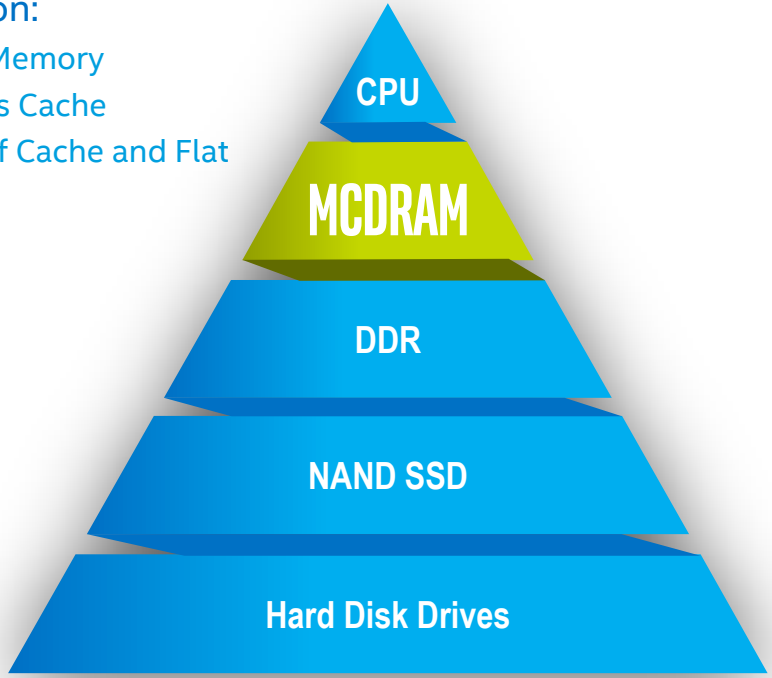
Cache Mode: Acts as Cache

Hybrid Mode: Mix of Cache and Flat

5X
BANDWIDTH
VS. DDR4[1], >400 GB/s[1]

>5X
ENERGY EFFICIENT
VS. GDDR5

>3X
DENSITY
VS. GDDR5[2]



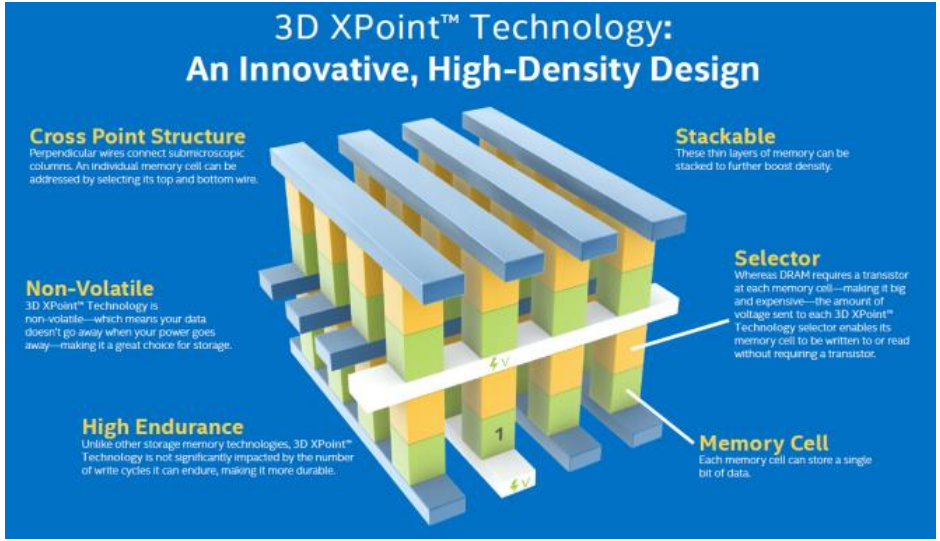
[1] Projected result based on internal Intel analysis of STREAM benchmark using a Knights Landing processor with 16GB of ultra high-bandwidth versus DDR4 memory with all channels populated.
[2] Projected result based on internal Intel analysis comparison of 16GB of ultra high-bandwidth memory to 16GB of GDDR5 memory used in the Intel® Xeon Phi™ coprocessor 7120P.

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

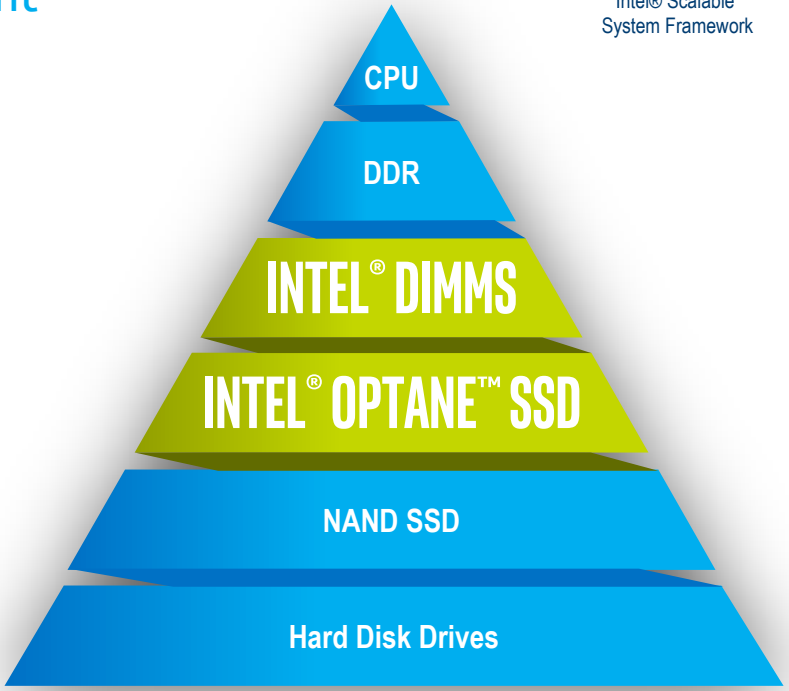
intel

Bridging the Memory-Storage Gap

Intel® Optane™ Technology Based on 3D XPoint™



Intel® Scalable System Framework



SSD

- 10x More Dense than Conventional Memory[3]
- Intel® Optane™ SSDs 5-7x Current Flagship NAND-Based SSDs (IOPS)[1]

DRAM-like performance

- Intel® DIMMs Based on 3D-XPoint™
- 1,000x Faster than NAND[1]
- 1,000x the Endurance of NAND[2]

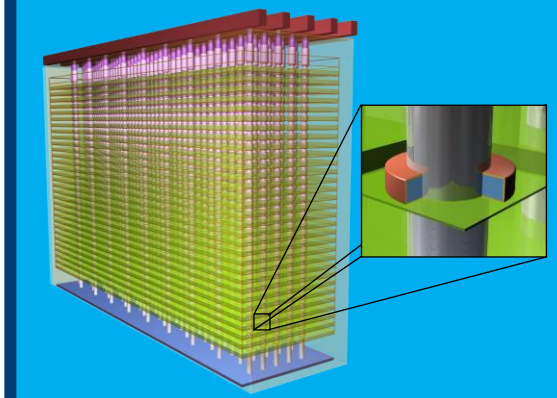
[1] Performance difference based on comparison between 3D XPoint™ Technology and other industry NAND
[2] Density difference based on comparison between 3D XPoint™ Technology and other industry DRAM
[2] Endurance difference based on comparison between 3D XPoint™ Technology and other industry NAND

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

NAND Flash and 3D XPoint™ Technology

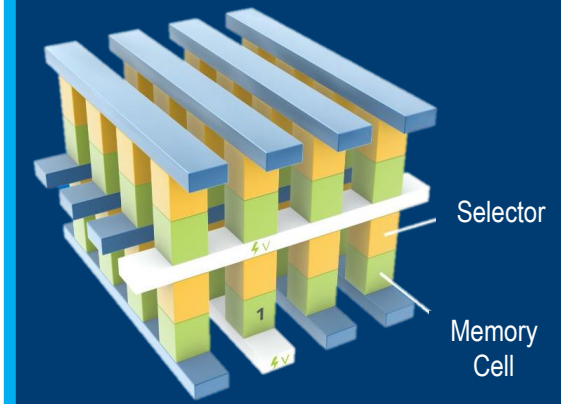
Intel® Scalable System Framework

3D MLC and TLC NAND



Enabling highest capacity SSDs at the lowest price

3D XPoint™ Technology



Enabling highest performance SSDs and expanding use cases

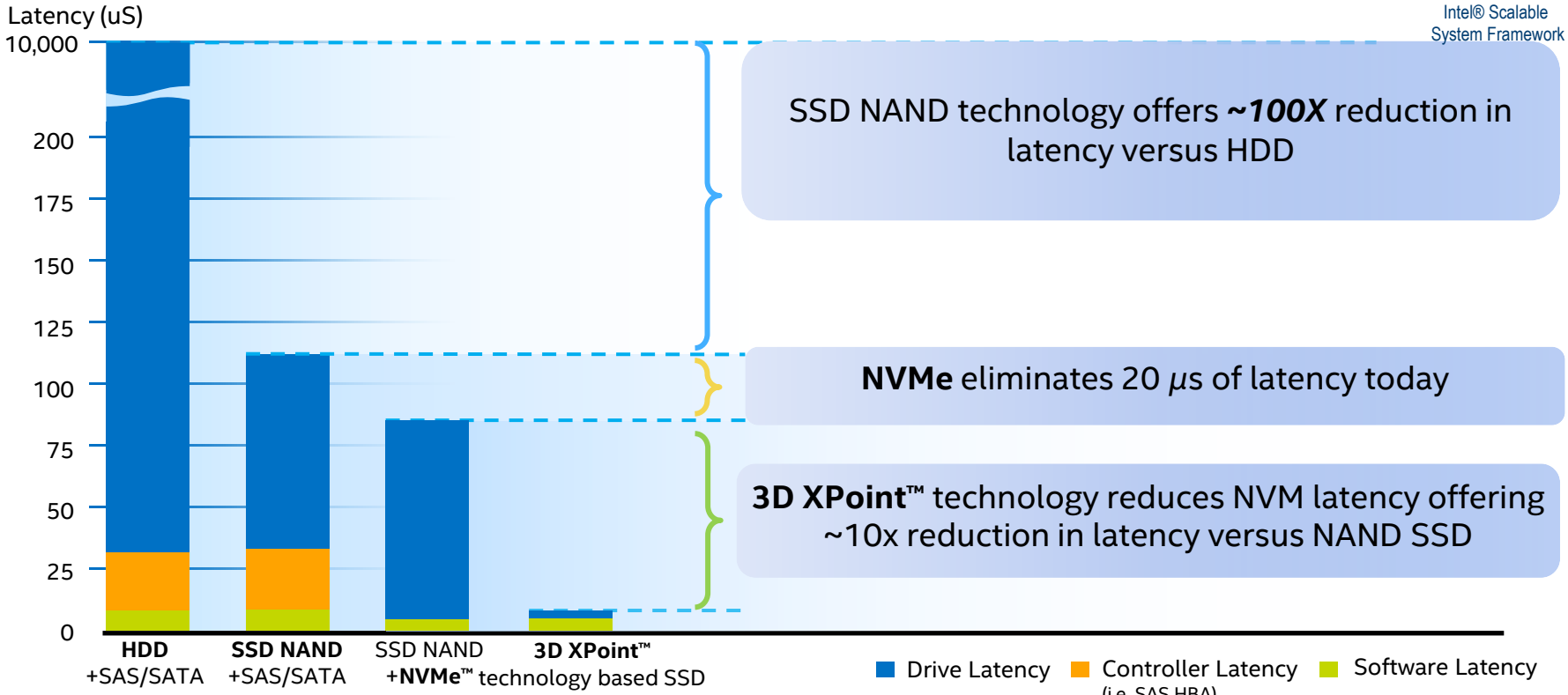
Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

intel

NVMe™ and 3D Xpoint™ are the next Quantum Leaps!

Memory

Intel® Scalable System Framework



SSD NAND technology offers ***~100X*** reduction in latency versus HDD

NVMe eliminates 20 μs of latency today

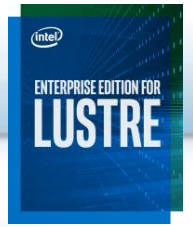
3D XPoint™ technology reduces NVM latency offering ~10x reduction in latency versus NAND SSD

Source: Storage Technologies Group, Intel. Comparisons between memory technologies based on in-market product specifications and internal Intel specifications.

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

intel

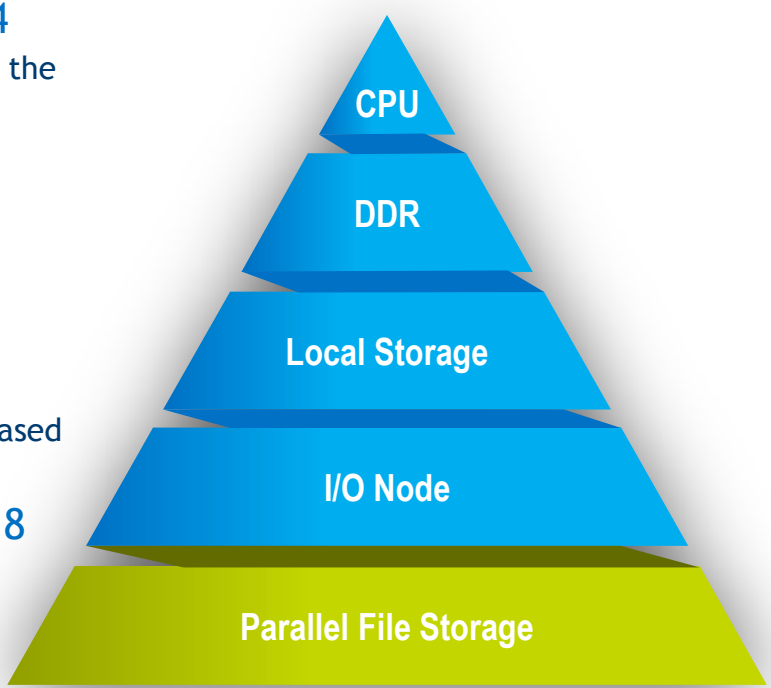
Intel® Solutions for Lustre* Software



Intel® Scalable System Framework

The Speed of Lustre* with the Support of Intel

- Intel® Enterprise Edition for Lustre* Software v2.4
  - Support for "Distributed Namespace" (DNE) Feature to Scale Out the Metadata Performance of Lustre*
  - Support for the Latest OS: Red Hat* 6.7-7 and SUSE* 11sp4-12
  - Parallel Read IO Performance & HSM Scalability Improvements
- Intel® Cloud Edition for Lustre* Software v1.2
  - Support for Over-the-Wire and Storage Encryption
  - Disaster Recovery from File System Snapshots
  - Simplified File System Mounting on Clients
  - Support for Intel® Xeon® Processor E5-2600 v3 Product Family-Based Instances
- Intel® Foundation Edition for Lustre* Software v2.8
  - Delivers the Latest Functions and Features
  - Fully Supported by Intel



EXTREME SCALE STORAGE FOR HPC

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

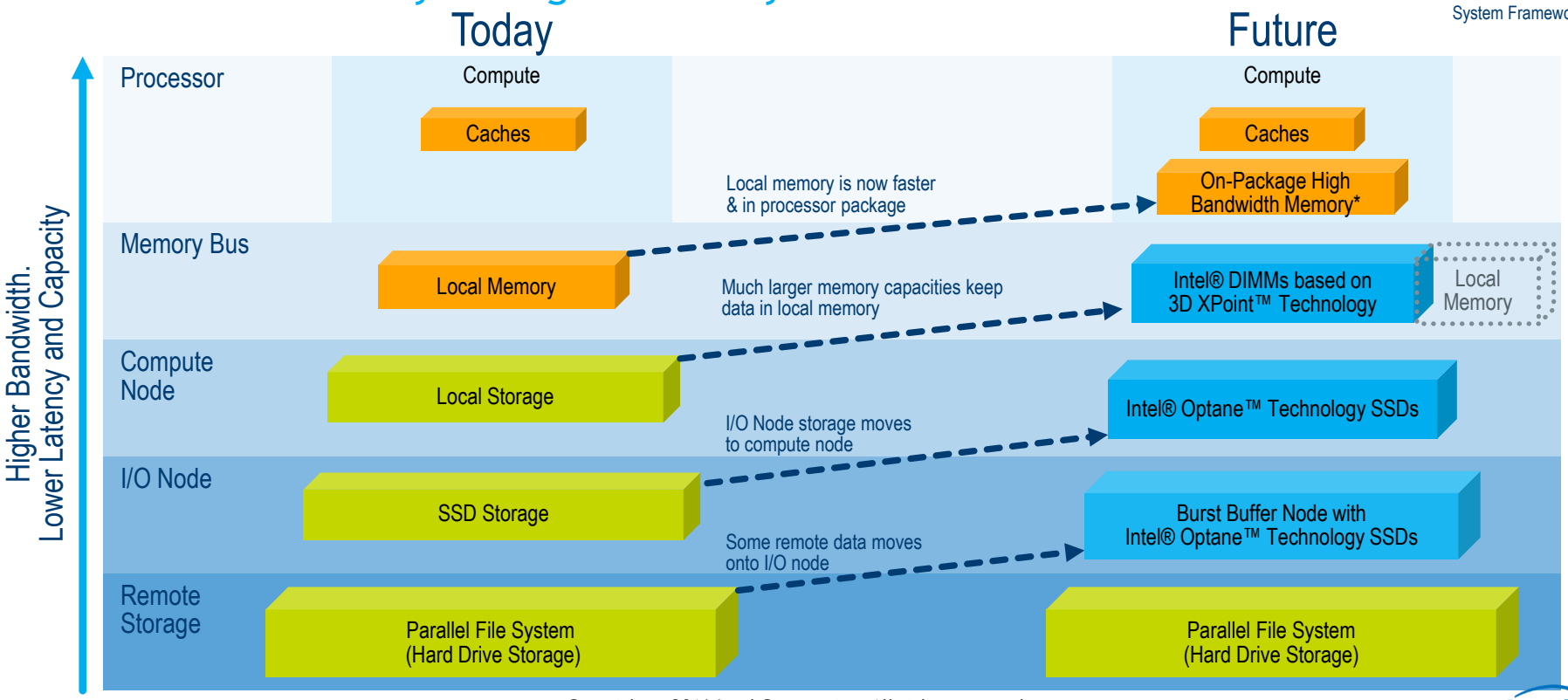
intel

Tighter System-Level Integration

Memory

Intel® Scalable System Framework

Innovative Memory-Storage Hierarchy



Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

intel

Intel® Omni-Path Architecture

Intel® Scalable System Framework

Evolutionary Approach, Revolutionary Features, End-to-End Solution

HFI Adapters
Single port
x8 and x16

x16 Adapter (100 Gb/s)

x8 Adapter (58 Gb/s)

Edge Switches
1U Form Factor
24 and 48 port

48-port Edge Switch

24-port Edge Switch

Director Switches
QSFP-based
192 and 768 port

768-port Director Switch (20U chassis)

192-port Director Switch (7U chassis)

Silicon
OEM custom designs
HFI and Switch ASICs

HFI silicon Up to 2 ports (50 GB/s total b/w)

Switch silicon up to 48 ports (1200 GB/s total b/w

Software
Open Source
Host Software and Fabric Manager

Cables
Third Party Vendors
Passive Copper Active Optical

Better Scaling vs. EDR

- 48 Radix Chip Ports
- Up to 26% More Servers than InfiniBand* EDR within the Same Budget[1]
- Up to 60% Lower Power and Cooling Costs[2]

Configurable / Resilient

- Job Prioritization (Traffic Flow Optimization)
- No-Compromise Resiliency (Packet Integrity Protection and Dynamic Lane Scaling)

Robust product offerings and ecosystem

- End-to-end Intel product line
- >100 OEM designs[3]
- Strong ecosystem with 70+ Fabric Builders members

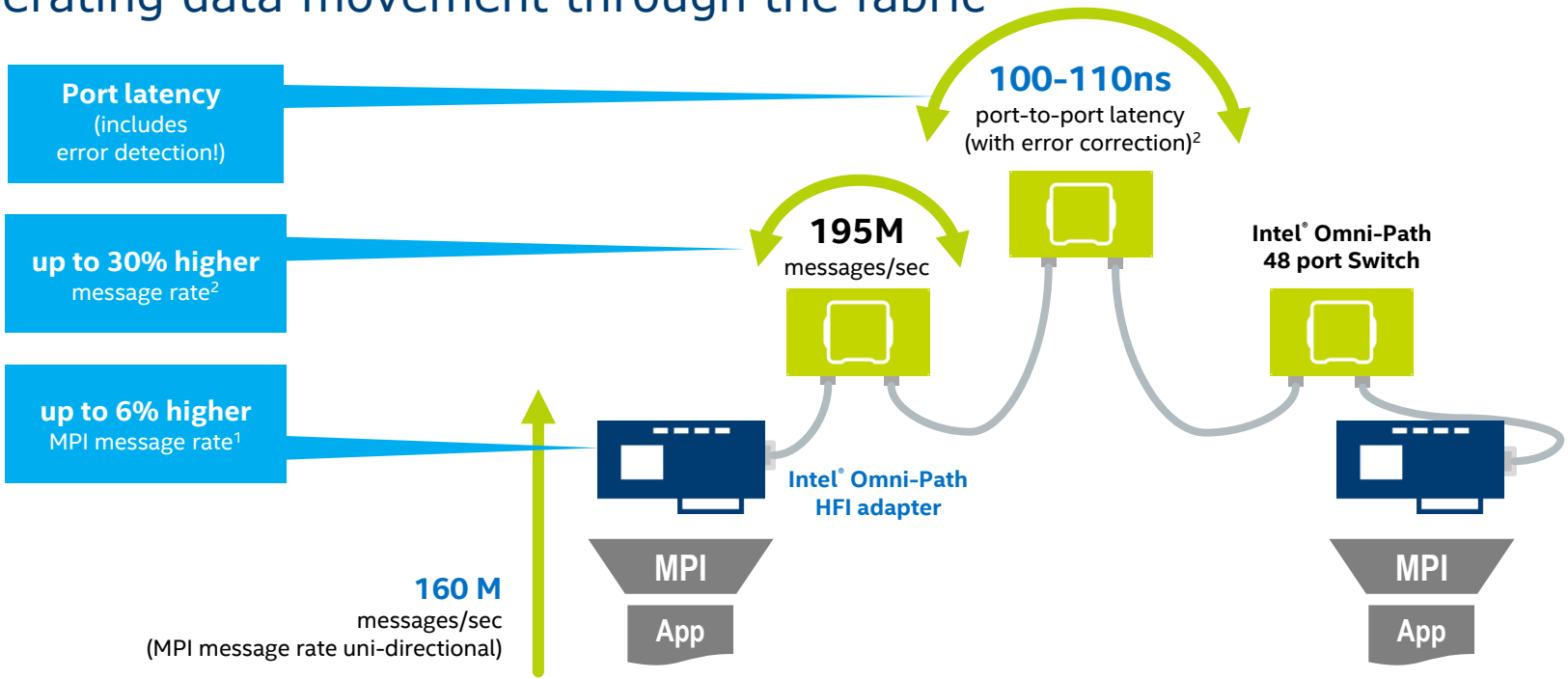
Maximizes price-performance, freeing up cluster budgets for increased compute and storage capability

1. 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 Intel MSRP pricing on ark.intel.com.
2. 2. 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 director switches and edge switches. Mellanox power data based on Mellanox CS7500 Director Switch, Mellanox SB7700/SB7790 Edge switch, and Mellanox ConnectX-4 VPI adapter card installation documentation posted on www.mellanox.com as of November 1, 2015. Intel OPA power data based on product briefs posted on www.intel.com as of November 16, 2015. Intel® OPA pricing based on estimated reseller pricing based on Intel MSRP pricing on ark.intel.com. 3. Intel internal information. Design win count based on OEM and HPC storage vendors who are planning to offer either Intel-branded or custom switch products, along with the total number of OEM platforms that are currently planned to support custom and/or standard Intel® OPA adapters. Design win count as of November 1, 2015 and subject to change without notice based on vendor product plans. *Other names and brands may be claimed as property of others.
3. Source: Intel internal information. Design win count based on OEM and HPC storage vendors who are planning to offer either Intel-branded or custom switch products, along with the total number of OEM platforms that are currently planned to support custom and/or standard Intel® OPA adapters. Design win count as of November 1, 2015 and subject to change without notice based on vendor product plans. *Other names and brands may be claimed as property of others.

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Intel® Omni-Path Architecture

Accelerating data movement through the fabric



[1] Based on Intel projections for Wolf River and Prairie River maximum messaging rates, compared to Mellanox CS7500 Director Switch and Mellanox ConnectX-4 adapter and Mellanox SB7700/SB7790 Edge switch product briefs posted on www.mellanox.com as of November 3, 2015.

[2] Latency reductions based on Mellanox CS7500 Director Switch and Mellanox SB7700/SB7790 Edge switch product briefs posted on www.mellanox.com as of July 1, 2015, compared to Intel measured data that was calculated from difference between back to back osu_latency test and osu_latency test through one switch hop. 10ns variation due to "near" and "far" ports on an Intel® OPA edge switch. All tests performed using Intel® Xeon® E5-2697v3 with Turbo Mode enabled.

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

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.

Copyright © 2016 Intel Corporation. All rights reserved.

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

Intel® Software Solutions

Intel® Scalable System Framework

Intel® Software Defined Visualization

Low Cost
No Dedicated Viz Cluster

Excellent Performance
Less Data Movement, I/O
Invest Power, Space, Budget in Greater Compute Capability

High Fidelity
Work with Larger Data Sets – Not Constrained by GPU Memory

Intel® Parallel Studio

Faster Code
Boost Application Performance on Current and Next-Gen CPUs

Create Code Faster
Utilizing a Toolset that Simplifies Creating Fast and Reliable Parallel Code

HPC System Software Stack

An Open Community Effort
Broad Range of Ecosystem Partners
Open Source Availability

Benefits the Entire HPC Ecosystem
Accelerate Application Development
Turnkey to Customizable

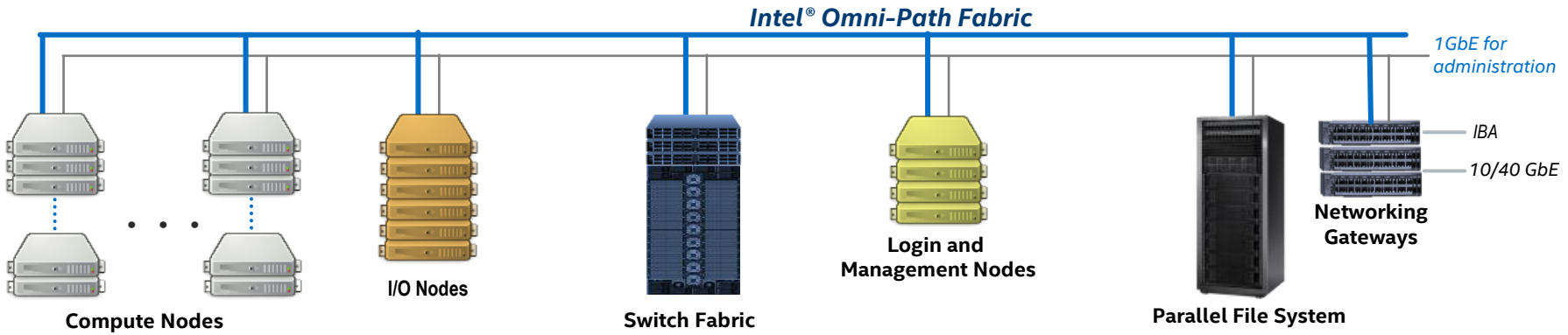
Open Software Available Today!

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

intel

What Makes a Great HPC Solution?

Intel® Scalable System Framework



Intel® Compute
Intel® Xeon Phi™ Processors
Intel® Xeon® Processors
Intel® Optane™ Technology
Intel® Omni-Path Fabric

Burst Buffer
Intel® Xeon® Processors
Intel® Omni-Path Fabric
Intel® Optane™ Technology

Intel® Networking
Intel® Omni-Path Fabric
Intel® Silicon Photonics

Intel® Software Tools
Intel® Parallel Studio
Intel® Node Manager
Intel® Trace Analyzer

Intel® Solutions for Lustre*
Intel® Enterprise Edition for Lustre*
Intel® Foundation Edition for Lustre*
Intel® Cloud Edition for Lustre*

Reference Architecture
Intel® Cluster Ready

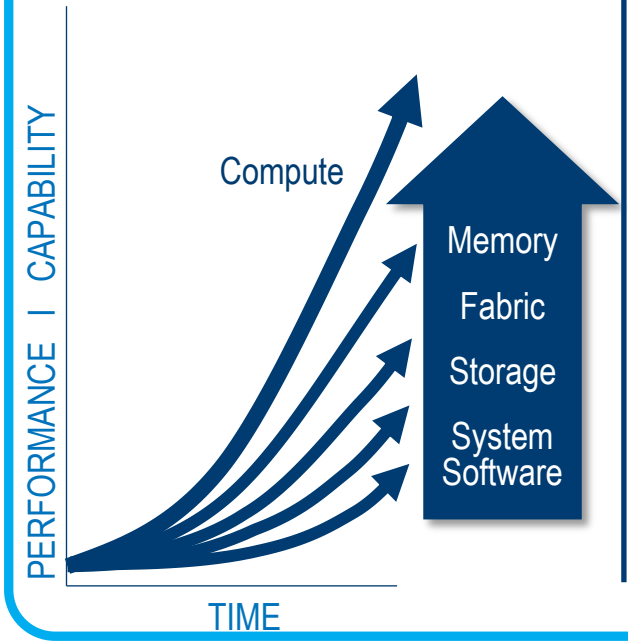
Actual configurations depend on specific OEM offerings and implementation.
Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Summary: a Holistic Architectural Approach

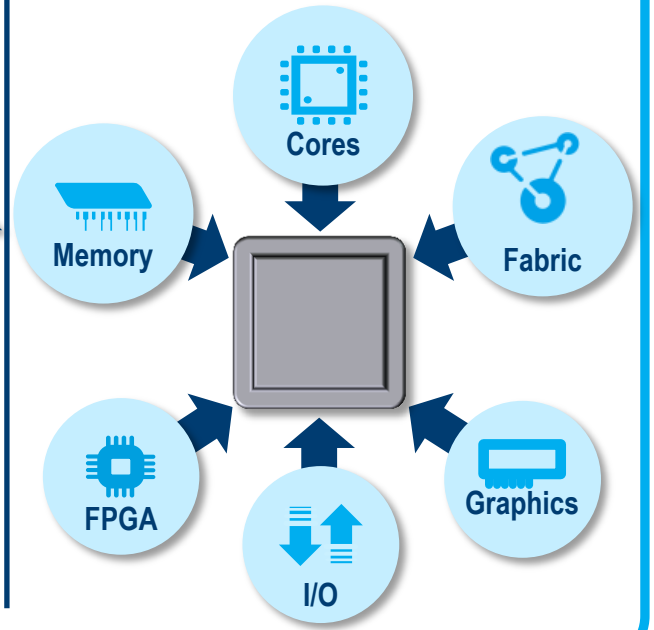
Intel® Scalable System Framework

System

Innovative Technologies



Tighter Integration



Application

Modernized Code

Copyright © 2016 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

intel

Спасибо!

nikolay.mester@intel.com