INTEL® XEON® SCALABLE PROCESSOR FOR HPC
Growing Challenges in HPC

**System Bottlenecks**
“The Walls”
- Memory
- I/O
- Storage
- Energy Efficient Performance
- Space
- Resiliency
- Unoptimized Software

**Divergent Workloads**
- Machine Learning
- Visualization
- HPC
- Big Data

**Barriers to Extending Usage**
- Democratization at Every Scale
- Cloud Access
- Exploration of New Parallel Programming Models

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

Optimizing for Cloud
Intel® Scalable System Framework

MODELING & SIMULATION  |  HPC DATA ANALYTICS  |  MACHINE LEARNING  |  VISUALIZATION

MANY WORKLOADS - ONE FRAMEWORK

A Flexible Framework for Today & Tomorrow

Enabling Breakthrough System Performance
Unified Intel® Xeon® Scalable Platform

**Intel® Xeon® Processor E7**
Targeted at mission-critical applications that value a scale-up system with leadership memory capacity and advanced RAS

**Intel® Xeon® Processor E5**
Targeted at a wide variety of applications that value a balanced system with leadership performance/watt/$
### Overview

- Skylake core microarchitecture, with data center specific enhancements
- Intel® AVX-512 with 32 DP Flops per core
- Data center optimized cache hierarchy – 1MB L2 per core, non-inclusive L3
- New mesh interconnect architecture
- Enhanced memory subsystem
- Modular IO subsystem with integrated devices
- New Intel® Ultra Path Interconnect (Intel® UPI)
- Intel® Speed Shift Technology
- Security & Virtualization enhancements
- Optional Integrated Intel® Omni-Path Fabric (Intel® OPA)

#### Features

<table>
<thead>
<tr>
<th>Features</th>
<th>Intel® Xeon® Processor E5-2600 v4</th>
<th>Intel® Xeon® Scalable Processor (Skylake-SP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cores Per Socket</td>
<td>Up to 22</td>
<td>Up to 28</td>
</tr>
<tr>
<td>Threads Per Socket</td>
<td>Up to 44 threads</td>
<td>Up to 56 threads</td>
</tr>
<tr>
<td>Last-level Cache (LLC)</td>
<td>Up to 55 MB</td>
<td>Up to 38.5 MB (non-inclusive)</td>
</tr>
<tr>
<td>QPI/UPI Speed (GT/s)</td>
<td>2x QPI channels @ 9.6 GT/s</td>
<td>Up to 3x UPI @ 10.4 GT/s</td>
</tr>
<tr>
<td>PCIe® Lanes/Controllers/Speed(GT/s)</td>
<td>40 / 10 / PCIe 3.0 (2.5, 5, 8 GT/s)</td>
<td>48 / 12 / PCIe 3.0 (2.5, 5, 8 GT/s)</td>
</tr>
<tr>
<td>Memory Population</td>
<td>4 channels of up to 3 RDIMMs, LRDIMMs, or 3DS LRDIMMs</td>
<td>6 channels of up to 2 RDIMMs, LRDIMMs, or 3DS LRDIMMs</td>
</tr>
<tr>
<td>Max Memory Speed</td>
<td>Up to 2400</td>
<td>Up to 2666</td>
</tr>
<tr>
<td>TDP (W)</td>
<td>55W-145W</td>
<td>70W-205W</td>
</tr>
</tbody>
</table>
Core Microarchitecture Enhancements

Data Center Core

- Larger and improved branch predictor, higher throughput decoder, larger window
- Improved scheduler and execution engine, improved throughput and latency of divide/sqrt
- More load/store bandwidth, deeper load/store buffers, improved prefetcher
- Data center specific enhancements ➔ Intel® AVX-512 with 2 FMAs per core, larger 1MB L2 per core

### Data Center-Specific Enhancements to the Core

<table>
<thead>
<tr>
<th></th>
<th>Broadwell uArch</th>
<th>Skylake uArch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Out-of-order Window</td>
<td>192</td>
<td>224</td>
</tr>
<tr>
<td>In-flight Loads + Stores</td>
<td>72 + 42</td>
<td>72 + 56</td>
</tr>
<tr>
<td>Scheduler Entries</td>
<td>60</td>
<td>97</td>
</tr>
<tr>
<td>Registers – Integer + FP</td>
<td>168 + 168</td>
<td>180 + 168</td>
</tr>
<tr>
<td>Allocation Queue</td>
<td>56</td>
<td>64/thread</td>
</tr>
<tr>
<td>L1D BW (B/Cyc) – Load + Store</td>
<td>64 + 32</td>
<td>128 + 64</td>
</tr>
<tr>
<td>L2 Unified TLB</td>
<td>4K+2M: 1024</td>
<td>4K+2M: 1536: 1G: 16</td>
</tr>
</tbody>
</table>

![Diagram of Data Center Core]
Intel® Advanced Vector Extensions 512 (Intel® AVX-512)

- 512-bit wide vectors
- 32 operand registers
- 8 64b mask registers
- Embedded broadcast
- Embedded rounding

<table>
<thead>
<tr>
<th>Microarchitecture</th>
<th>Instruction Set</th>
<th>SP FLOPs / cycle</th>
<th>DP FLOPs / cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skylake</td>
<td>Intel® AVX-512 &amp; FMA</td>
<td>64</td>
<td>32</td>
</tr>
<tr>
<td>Haswell / Broadwell</td>
<td>Intel® AVX2 &amp; FMA</td>
<td>32</td>
<td>16</td>
</tr>
<tr>
<td>Sandybridge</td>
<td>Intel® AVX (256b)</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>Nehalem</td>
<td>SSE (128b)</td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intel AVX-512 Instruction Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intel AVX-512-F</td>
</tr>
<tr>
<td>AVX-512-VL</td>
</tr>
<tr>
<td>AVX-512-BW</td>
</tr>
<tr>
<td>AVX-512-DQ</td>
</tr>
<tr>
<td>AVX-512-CD</td>
</tr>
</tbody>
</table>

**INTEL® AVX-512 DOUBLES THE NUMBER OF FLOPS PER CYCLE**
New Mesh Interconnect Architecture

Xeon® E7 v4 24-core die

Skylake-SP 28-core die

DUAL-RING IN BROADWELL SERVER (INTEL® XEON® PROCESSOR E5/E7) V. MESH IN SKYLAKE-SP
Re-architected L2 & L3 Cache Hierarchy

**Previous Architectures**

- **Shared L3** 2.5MB/core (inclusive)
  - L2 (256KB private)
  - L2 (256KB private)
  - L2 (256KB private)

**Core**

**Skylake-SP Architecture**

- **Shared L3** 1.375MB/core (non-inclusive)
  - L2 (1MB private)
  - L2 (1MB private)
  - L2 (1MB private)

**Core**

- **On-chip cache balance shifted from shared-distributed (prior architectures) to private-local (Skylake architecture):**
  - Shared-distributed ➔ shared-distributed L3 is primary cache
  - Private-local ➔ private L2 becomes primary cache with shared L3 used as overflow cache

- **Shared L3 changed from inclusive to non-inclusive:**
  - Inclusive (prior architectures) ➔ L3 has copies of all lines in L2
  - Non-inclusive (Skylake architecture) ➔ lines in L2 *may not* exist in L3
IO Performance

>50% aggregate IO bandwidth improvement in line with memory bandwidth increase for a balanced system performance

Source as of June 2017: Intel internal measurements on platform with Xeon Platinum 8180, Turbo enabled, UPI=10.4, SNC1, 6x32G DDR4-2400/2666 per CPU, 1 DPC, and platform with E5-2699 v4, Turbo enabled, 4x32GB DDR4-2400, RHEL 7.0. 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
Skylake-SP with Integrated Fabric

Single on-package Storm Lake Host Fabric Interface (HFI) port
Fabric component interfaces to SKL using 16 PCIe* lanes
Fabric PCIe* lanes are incremental to the existing 48 PCIe* lanes available on SKL-F
Single cable from SKL-F package connector to QSFP module at rear IO panel
Same socket for SKL-SP and SKL-F processors

- Intel® Xeon® Scalable Processor-based platforms can be designed to support both processors
- Platform designs will require an expanded keep-out zone and additional board components to accommodate both processors
June 2017 Top500 Analysis

**Intel® OPA 22% more total entries than EDR IB***

- Intel® OPA: 38 entries
- EDR IB*: 31 entries

**Intel® OPA 7.6% higher average efficiency at scale than EDR IB***

- Average efficiency on Intel® Xeon® processors:
  - Intel® OPA: 81.9%
  - EDR IB*: 74.3%

**Intel® OPA 30% more top 100 entries than EDR IB***

- Top 10: Intel® OPA 1, EDR IB*: 4
- Top 15: Intel® OPA 6, EDR IB*: 3
- Top 50: Intel® OPA 13, EDR IB*: 10

**Share of 100Gb Flops – 67.1PF OPA vs 38.7PF EDR**

- Intel® OPA: 67.1PF
- EDR IB*: 38.7PF

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

www.top500.org
June 2017 list
Why OPA has been winning in HPC

HPC Performance – Major Buying Criteria

- In most cases, Intel OPA has equal to or better HPC performance (latency, bandwidth, and applications)

Price / Performance – Next Evaluation/Buying Point

- Leadership price/performance at any scale – from entry to large supercomputers
- Build a better cluster – more FLOPS and storage

OPA Features – Key Differentiators

- Builds on key technologies from QLogic and Cray needed to support ever-increasing HPC cluster sizes & workloads
- Features: CPU/Fabric Integration, HPC Optimization and Enhanced Fabric Functionality
Intel® OPA’s Impact across Many Segments

100’s of Accounts Deployed

Supercomputers

Artificial Intelligence

Traditional HPC

Enterprise R&D

HPC Cloud

15
Application Performance - Intel® Xeon® Platinum 8170 processors

Intel® Omni-Path Architecture (Intel® OPA) vs EDR InfiniBand*

Best performance using either Intel® MPI or Open MPI for both fabrics

1. Not every application was tested with both Intel MPI and Open MPI. See configuration slide for necessary details.

2. VASP benchmark comparison at 8 nodes only. Some workloads use different rank counts. See configuration slide for detail.

Relative Performance Intel® OPA vs EDR IB*
1.63x Average Gains on High Performance Compute Apps

Results have been estimated based on internal Intel analysis and are provided for informational purposes only. Geomean of Weather Research Forecasting - Conus 12Km, HOMME, LSTCLS-DYNA Explicit, INTES PERMAS V16, MILC, GROMACS water 1.5M_pme, VASP/256, NAMD/stmv, LAMMPS, Amber GB Nucleosome, Binomial option pricing, Black-Scholes, Monte Carlo European options. Any difference in system hardware or software design or configuration may affect actual performance. 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 information go to http://www.intel.com/performance/datacenter. Configurations: see page 54,55

Benefits of Intel® Xeon® processor Scalable family
- Intel® AVX-512 and Higher IPC/Per Core performance
- Balanced IO and memory
- Predictable latency with mesh/memory architecture improvements

Additional options to scale performance
- Intel® Optane™ SSD DC P4800X Series for fast storage
- Intel® Omni-Path Fabric to scale cluster performance
### Latency, Bandwidth, and Message Rate - Intel MPI Benchmarks

**Intel® Xeon® Processor E5-2697A v4 & Intel® Xeon® Platinum 8170 processors**  
**Intel® Omni-Path Architecture (Intel® OPA)**

| Metric                                      | Intel® Xeon® E5-2697A v4 CPU¹  
|                                             | 2.6 GHz, 16c | Intel® Xeon® Platinum 8170 CPU²  
|                                             | 2.1 GHz, 26c |
| Latency (one-way, 1 switch, 8B) [ns] ; PingPong | 910          | 940          |
| Bandwidth (1 rank per node, 1 port, uni-dir, 1MB) [GB/s] ; Uniband | 12.3         | 12.3         |
| Bandwidth (1 rank per node, 1 port, bi-dir, 1MB) [GB/s] ; Biband | 24.4         | 24.5         |
| Message Rate (max ranks per node, uni-dir, 8B) [Mmps] ; Uniband | 112.0        | 152.0        |
| Message Rate (max ranks per node, bi-dir, 8B) [Mmps] ; Biband | 131.7        | 211.2        |

**Last updated: April 2017**

Dual socket servers with one Intel® OPA Edge switch hop. Intel® Turbo Boost Technology enabled, Intel® Hyper-Threading Technology enabled. Intel MPI Benchmarks 4.1. Intel® OPA: Open MPI 1.10.4-hfi as packaged with IFS 10.3.1.0.22. Benchmark processes pinned to the cores on the socket that is local to the Intel® OP Host Fabric Interface (HFI) before using the remote socket.  
1. Intel® Xeon® E5-2697A v4 processor, 2.60 GHz 16 cores, 2133 MHz DDR4 memory. 32 ranks per node for message rate tests. RHEL* 7.2., 3.10.0-327.36.3.el7.x86_64 kernel. BIOS settings: IOU non-posted prefetch disabled. Snoop timer for posted prefetch=9. Early snoop disabled. Cluster on Die disabled.  
2. Intel® Xeon® Platinum 8170 processor, 2.10 GHz 26 cores, 64 GB 2666 MHz DDR4 memory per node. 52 ranks per node for message rate tests RHEL* 7.3, 3.10.0-514.el7.x86_64 kernel.  

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](http://www.intel.com/performance). Copyright © 2017, Intel Corporation. Other names and brands may be claimed as the property of others.
Intel® Xeon® Platinum Processor 8180 vs. Intel® Xeon® Processor E5-2699v4

SGEMM Performance

![Graph showing SGEMM Performance comparison between SKX and BDX]

DGEMM Performance

![Graph showing DGEMM Performance comparison between SKX and BDX]

<table>
<thead>
<tr>
<th></th>
<th>DGEMM</th>
<th>SGEMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>Up to 2.26×</td>
<td>Up to 2.31×</td>
</tr>
<tr>
<td>20Kx20Kx20K</td>
<td>Up to 2.25×</td>
<td>Up to 2.32×</td>
</tr>
</tbody>
</table>

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](http://www.intel.com/performance)

Source: Intel measured as of June 2017. Configuration details on slide 27.
Increased SPEC MPI 2007 performance with the 2S Intel® Xeon® Gold 6148. Performance metric is a geomean over 13 apps.

**SPEC MPI 2007**

**Application / Workload Description:**
- SPEC MPI* 2007 benchmark suite is for evaluating MPI-parallel, floating point, compute intensive performance across a wide range of cluster and SMP hardware
- Workload can be run on either single node and cluster (this is a single node comparison)

**Key Takeaway:**
- SPEC gives users the most objective and representative benchmark suite for measuring and comparing high-performance computer systems.

**Performance Factors:**
- Intel® AVX-512 contributes up to 18% performance boost per component.
- Memory bandwidth improvements speedup application performance up to 71% in geomean.
- Memory bandwidth contributes to performance increase.

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

1. Testing conducted on ISV* software comparing Intel® Xeon® Gold 6148 processor to 2S Intel® Xeon® Processor E5-2697 v3 and to 2S Intel® Xeon® Processor E5-2697 v4. Testing done by Intel. For complete testing configuration details, see slide 28.
Continued scientific progress ... increasingly depends on advances in simulation and data analysis, which will only be possible through code modernization...

—Sudip Dosanjh, Director at NERSC

Notices and Disclaimers

Intel technologies’ features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.

No computer system can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit http://www.intel.com/performance.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Statements in this document that refer to Intel's plans and expectations for the quarter, the year, and the future, are forward-looking statements that involve a number of risks and uncertainties. A detailed discussion of the factors that could affect Intel's results and plans is included in Intel's SEC filings, including the annual report on Form 10-K.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Intel, the Intel logo and others are trademarks of Intel Corporation in the U.S. and/or other countries. *Other names and brands may be claimed as the property of others.

© 2017 Intel Corporation.
Optimization Notice

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel.

Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Notice revision #20110804