INTEL XEON SCALABLE (SKYLAKE) PROCESSOR AND PURLEY PLATFORM
Legal Notices and Disclaimers

All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance estimations and benchmark results were obtained prior to implementation of recent software patches and firmware updates intended to address exploits referred to as "Spectre" and "Meltdown". Implementation of these updates may make these results inapplicable to your device or system. 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 www.intel.com/benchmarks.

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. Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance.

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 SSSE3 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
Legal Notices and Disclaimers Continued

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. **No computer system can be absolutely secure.** Check with your system manufacturer or retailer or learn more at [intel.com].

Intel technologies may require enabled hardware, specific software, or services activation. Check with your system manufacturer or retailer.

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

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

Intel, the Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

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

Copyright © 2019 Intel Corporation. All rights reserved.
Agenda

SuperMUC-NG system overview

Purley Platform Features

Skylake processor

- Uncore Architecture
- Core Architecture

Some abbreviations used in the presentation:
AVX – Advanced Vector Extensions
ISA – Instruction Set Architecture
SKX – Skylake Server, also Skylake-SP: codename for Intel Xeon Scalable Processor
LBG – Lewisburg: codename for Intel C620 Series Chipsets
MLC – Mid-Level Cache, usually 2nd level cache
LLC – Last Level Cache, usually Level-3 cache of a processor
OPA – Intel Omni-Path Architecture fabric
UPI – Intel Ultra Path Interconnect
QPI – Intel QuickPath Interconnect

* Other names and brands may be claimed as the property of others.
## Characteristics of SuperMUC-NG

<table>
<thead>
<tr>
<th>Compute Nodes</th>
<th>Thin Nodes</th>
<th>Fat Nodes</th>
<th>Total (Thin + Fat)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Processor</strong></td>
<td>Intel Xeon Platinum 8174 (Skylake-SP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cores per Node</strong></td>
<td>48</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td><strong>Memory per node (GByte)</strong></td>
<td>96</td>
<td>768</td>
<td></td>
</tr>
<tr>
<td><strong>Number of Nodes</strong></td>
<td>6,336</td>
<td>144</td>
<td>6,480</td>
</tr>
<tr>
<td><strong>Number of Cores</strong></td>
<td>304,128</td>
<td>8,912</td>
<td>311,040</td>
</tr>
<tr>
<td><strong>Rpeak @ nominal (PFlop/s)</strong></td>
<td>26.3</td>
<td>0.6</td>
<td>26.9</td>
</tr>
<tr>
<td><strong>Memory (TByte)</strong></td>
<td>608</td>
<td>111</td>
<td>719</td>
</tr>
</tbody>
</table>

### High-Performance Fabric

Intel Omni-Path Architecture fabric

“Inverted” fat tree: 8 thin-node islands non-blocking within 792 nodes

### Filesystems

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High Performance Parallel Filesystem (HPPFS)</strong></td>
<td>50 PB @ 500 GB/s for $SCRATCH and $WORK</td>
</tr>
<tr>
<td><strong>Data Science Storage (DSS)</strong></td>
<td>20 PB @ 70 GB/s for $PROJECT</td>
</tr>
<tr>
<td><strong>Home Filesystem</strong></td>
<td>256 TB for $HOME</td>
</tr>
</tbody>
</table>

### Infrastructure

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cooling</strong></td>
<td>Direct warm water cooling</td>
</tr>
<tr>
<td><strong>Waste Heat Reuse</strong></td>
<td>Reuse for producing cold water with adsorption coolers</td>
</tr>
</tbody>
</table>

### Software

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Operating system and provisioning</strong></td>
<td>SuSE Linux (SLES 12 SP3) and xCat, OpenHPC-compliant, network boot</td>
</tr>
<tr>
<td><strong>Batch system, containers</strong></td>
<td>SLURM 18.08, containers (e.g., CharlieCloud, etc.)</td>
</tr>
<tr>
<td><strong>Development Environment</strong></td>
<td>Intel Parallel Studio XE 2019, Intel MPI 2019</td>
</tr>
</tbody>
</table>

Source: [https://www.lrz.de/services/compute/supermuc/supermuc-ng/](https://www.lrz.de/services/compute/supermuc/supermuc-ng/)

* Other names and brands may be claimed as the property of others.
System design: islands and fabrics

**Storage Island**
- 1152p OPA1 DCS
  - 4x 48 ports leaf
  - 12x 32 ports leaf

**Service Island**
- 1152p OPA1 DCS
  - 6x 48 ports
  - 2x 32 ports

**Thin-node island 1**
- 788 (+4 spare)
- 11 racks

**Thin-node island 8**
- 788 (+4 spare)
- 11 racks

**Bandwidth** 1.6 TB/sec

**Ratio** 3.75:1

- 73x Y cables 100m

**DR HOME:**
- 10x cables 100m

**Cloud** [100GbE]

**LRZ backbone [40GbE]**

**HRR**

- 2 FireWall servers

* Other names and brands may be claimed as the property of others.
System design: compute node and chassis


* Other names and brands may be claimed as the property of others.
Intel Server Platforms Transitions

### Romley-EP Platform
- **Processor Brand**: Xeon E5-2600, Xeon E5-2600 v2, Xeon E5-2600 v3, Xeon E5-2600 v4, Xeon Scalable
- **Core Codename**: Sandy Bridge, Ivy Bridge, Haswell, Broadwell
- **Process**: 32nm, 22nm
- **Max Cores**: 8, 12
- **New Lead ISA**: AVX, AVX2
- **Memory**: 4ch DDR3-1600, 4ch DDR3-1866
- **Bus speed**: QPI 8GT/s

### SuperMUC Phase 2
- **Processor Brand**: Xeon E5-2600 v3, Xeon E5-2600 v4
- **Core Codename**: Haswell, Broadwell
- **Process**: 22nm, 14nm
- **Max Cores**: 18, 22
- **New Lead ISA**: AVX2
- **Memory**: 4ch DDR4-2133, 4ch DDR4-2400
- **Bus speed**: QPI 9.6 GT/s

### Purley-SP Platform
- **Processor Brand**: Xeon Scalable
- **Core Codename**: Skylake, Cascade Lake
- **Process**: 14nm, 28
- **Max Cores**: 28
- **New Lead ISA**: AVX-512
- **Memory**: 6ch DDR4-2666
- **Bus speed**: UPI 10.4 GT/s

Source: https://ark.intel.com
* All processors support prior available x86 ISA extensions up to Intel® SSE4.2 and several application-specific accelerator instructions, such as AES-NI
* Other names and brands may be claimed as the property of others.
Significant Performance Gains on HPC Applications
Intel® Xeon® Scalable Processor

HPC Application Performance

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>MILC</td>
<td>1.4</td>
</tr>
<tr>
<td>OpenFoam</td>
<td>1.5</td>
</tr>
<tr>
<td>miniFE</td>
<td>1.5</td>
</tr>
<tr>
<td>VASP</td>
<td>1.8</td>
</tr>
<tr>
<td>LAMMPS</td>
<td>2.2</td>
</tr>
</tbody>
</table>

GeoMean 1.65

Higher is better

1.65x average gains on HPC workloads comparing Intel® Xeon® Scalable processor to prior generation (geomean of MILC, VASP, OpenFoam, miniFE, LAMMPS).

Intel internal measurements as of June 2018. Configuration details: see backup. * Other names and brands may be claimed as the property of others.

Performance results are based on testing as of June 2018 and may not reflect all publicly available security updates. See configuration disclosure for details. No product can be absolutely secure.

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 www.intel.com/benchmarks.
Grantley Platform Performance with Xeon E5-2697 v3

~1 TFlops/sec*

~50 GB/s**

~50 GB/s**

up to 40 GB/s+

up to 76.8 GB/s+

TBG

PCH

DDR4 2133

2x1GbE

SATA

USB

Intel® QPI

PCIe Gen3

PCIe Gen3

DGEMM benchmark, ** STREAM Triad benchmark, + peak bandwidth

Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your 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.
Purley Platform Performance with Xeon Platinum 8174 (205W)

Purley platform delivers up to 2x bandwidth across multiple performance metrics

~2.8 TFlops/sec*

~100 GB/s **

up to 83 - 124 GB/s

4x10GbE

SATA

USB

PCle Gen3

Intel® UPI

up to

48 GB/s+

DDR4 2666

~100 GB/s **

PCle Gen3

DDR4 2666

* DGEMM benchmark, ** STREAM Triad benchmark, * peak bandwidth

Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your 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.
Skylake Server Processor Overview

14nm Process Technology with Skylake Core Microarchitecture

Intel® AVX-512 with 2 FMAs/Core Virtualization Enhancements
New Memory Protection Features (PPK, MPX, XU/XS)

Rebalanced Cache Hierarchy: Increased MLC to 1MB/Core
1.375 MB Last Level Cache/Core

Up to 28 Cores on a 2D-Mesh

Intel® Hyper-Threading Technology (2 threads/core)

Intel® Turbo Boost Technology

Existing Feature

New Skylake Server Feature

Power Management:
Per Core P-State (PCPS)
Uncore Frequency Scaling (UFS)
Energy Efficient Turbo (EET)
On die PMAX detection (NEW)
Intel® Speed Shift Technology (NEW)

Memory Technology:
6x DDR4 channels
2133, 2400, 2666 MT/s
RDIMM, LRDIMM

48 Lanes of PCI Express* 3.0
Intel® QuickData Technology Enhancements – 2x bandwidth

Intel® UPI: 10.4GT/s, 30% better eff.

Integrated Voltage Regulator
New On-Chip Interconnect Architecture

Haswell Server: up to 18 cores per die
Skylake Server: up to 28 cores per die

Mesh improves scalability with higher bandwidth and reduced latencies
Skylake Server Socket Diagram
Intel® Ultra Path Interconnect (Intel® UPI)

- Keizer Technology Interconnect (KTI) is now Intel® Ultra Path Interconnect (Intel® UPI), replacing Intel® QPI in the Purley and future platforms
- New, faster coherent link with greater message efficiency
  - Increased bandwidth and performance over QPI
  - Improved messaging efficiency, multiple requests per packet

UPI is not compatible with QPI.
Based on Intel internal measurements

**Idle Power**

<table>
<thead>
<tr>
<th>LO</th>
<th>L0p QPI</th>
<th>L0p UPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>25%</td>
<td></td>
</tr>
</tbody>
</table>

**Data Efficiency**

- 4% to 21% (per wire)

**Data Rate**

- 9.6 GT/s QPI
- 10.4 GT/s UPI

UPI enables high throughput and power efficiency

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.
Coherency Optimizations in Purley Platform

Sub-NUMA Cluster Mode
- Associates LLC slice with nearest memory controller
- Applications use NUMA primitives to achieve lower LLC/memory latency

XPT Prefetch
- Core miss initiates local memory access in parallel with LLC access
- Uses history-based prediction to avoid unnecessary prefetches

UPI Prefetch
- Similar to XPT prefetch, but for UPI requests from remote socket

Local and Remote Direct-to-Core
- Data sent directly from memory controller or UPI to requesting core

Direct-to-UPI
- For UPI requests from remote socket, memory controller directly sends data to UPI instead of going through CHA

UPI Optimizations
- Opportunistic snoop broadcast – avoid directory read to save memory BW
- HitME cache – directory cache for frequently used lines
- IO Directory Cache – directory cache for remote IO writes
Sub-NUMA Clusters (SNC)

Prior generation supported Clusters-On-Die (COD)

SNC provides similar localization benefits as COD, without some of its downsides
- Only one UPI caching agent required even in 2-SNC mode
- Latency for memory accesses in remote cluster is smaller, no UPI flow
- LLC capacity is utilized more efficiently in 2-cluster mode, no duplication of lines in LLC

Downside with SNC
- Addresses from remote cluster never get cached in local cluster LLC, resulting in larger latency compared to COD in some cases

SNC provides extra NUMA node pathways just as COD does. Therefore when setting affinity on Skylake use the same methodology as you would for Haswell/Broadwell

<table>
<thead>
<tr>
<th></th>
<th>Clusters per socket</th>
<th>XPT/UPI Prefetch</th>
<th>SNC Domain 0</th>
<th>SNC Domain 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNC off</td>
<td>1</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-SNC</td>
<td>1</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-SNC</td>
<td>2</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- SNC off: No prefetch in UMA
- 1-SNC: Like SNC off, but can do prefetch
- 2-SNC: Yes

- Addresses from remote cluster never get cached in local cluster LLC, resulting in larger latency compared to COD in some cases
Sub-NUMA Clusters – 2 SNC Example

**Local SNC Access**

<table>
<thead>
<tr>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>Core</td>
<td>Core</td>
<td>Core</td>
<td>Core</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>Core</td>
<td>Core</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mem Ctrl</th>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
<th>Mem Ctrl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>Core</td>
<td>Core</td>
<td>Core</td>
<td>Mem Ctrl</td>
</tr>
</tbody>
</table>

**Remote SNC Access**

<table>
<thead>
<tr>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>Core</td>
<td>Core</td>
<td>Core</td>
<td>Core</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>Core</td>
<td>Core</td>
<td>Core</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mem Ctrl</th>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
<th>Mem Ctrl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>Core</td>
<td>Core</td>
<td>Core</td>
<td>Mem Ctrl</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
<th>LLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>Core</td>
<td>Core</td>
<td>Core</td>
<td>Core</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Core</th>
<th>Core</th>
<th>Core</th>
<th>Core</th>
<th>Core</th>
</tr>
</thead>
</table>
Intel® Xeon® Processor Scalable Family
Dual-socket memory latency & bandwidth*

100% local memory read
Sequential address

* Source as of May 2017: Intel internal measurements of BW/latency on platform with Skylake-SP H0 28C internal sample, Core=turbo, CLM=turbo, KTI=10.4, SNC1, 6x32GB DDR4-2400/2667 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 information go to http://www.intel.com/performance

*Other names and brands may be claimed as the property of others.
On-Chip Cache Coherency

Distributed LLC or L3 cache – 1.375MB slice physically located with each core

- Cache line address hashed across all L3 slices
- L3 is not inclusive of L2 and L1, acts as victim cache for L2 evictions
- Same L1/L2 tracking capability as SF

Snoop Filter (SF) tracks presence of lines in L2 and L1

- Uses same address hash as L3
- SF+L3 is inclusive of L2 & L1 cache
Rebalanced Cache Hierarchy

- Shift cache balance from shared-distributed to private-local by enlarging Mid Level Cache (MLC)
- Shared Last Level cache (LLC) retained to benefit shared data and to enable capacity balancing
- Snoop Filter to track all core owned lines
- Total per Core/LLC tile cache size lower than Broadwell/Haswell
- Higher hit rate on lower latency MLC than Broadwell/Haswell

Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your 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 complete information visit http://www.intel.com/performance.
Inclusive L3 v. Non-Inclusive L3

1. Memory reads fill directly to the MLC, no longer to both the MLC and LLC
2. When a MLC line needs to be removed, both modified and unmodified lines are written back
3. Data shared across cores are copied into the LLC for servicing future MLC misses

Skylake Server cache hierarchy architected and optimized for server use cases:
- Multithreaded workloads can operate on larger data per thread (due to increased L2 size) and reduce uncore activity
- Virtualized use-cases get larger private L2 cache free from interference
Effect of Cache Rebalancing

Multi-Threaded, Batch, or Virtualized Workloads

• Effectively utilizes MLC capacity private to each core
• Reduces cache pollution effects from “noisy neighbors”

Single-Thread per socket Workloads

• Does not utilize MLC capacity in the cores not executing the thread
• Larger MLC reduces effective memory latency and performance variability

Concurrent and virtualized workloads effectively utilize available cache capacity
Single thread per socket performance may see some cache capacity deficit
### Core Cache Size/Latency/Bandwidth

<table>
<thead>
<tr>
<th>Metric</th>
<th>Xeon E5 v4 (Broadwell)</th>
<th>Xeon SP (Skylake)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 Instruction Cache</td>
<td>32K, 8-way</td>
<td>32K, 8-way</td>
</tr>
<tr>
<td>L1 Data Cache</td>
<td>32K, 8-way</td>
<td>32K, 8-way</td>
</tr>
<tr>
<td>Fastest Load-to-use</td>
<td>4 cycles</td>
<td>4 cycles</td>
</tr>
<tr>
<td>Load bandwidth</td>
<td>64 Bytes/cycle</td>
<td>128 Bytes/cycle</td>
</tr>
<tr>
<td>Store bandwidth</td>
<td>32 Bytes/cycle</td>
<td>64 Bytes/cycle</td>
</tr>
<tr>
<td>L2 Unified Cache</td>
<td>256K, 8-way</td>
<td>1024K, 16-way</td>
</tr>
<tr>
<td>Fastest load-to-use</td>
<td>11 cycles</td>
<td>14 cycles</td>
</tr>
<tr>
<td>Bandwidth to L1</td>
<td>64 Bytes/cycle</td>
<td>64 Bytes/cycle</td>
</tr>
<tr>
<td>L1 Instruction TLB</td>
<td>4K: 128, 4-way</td>
<td>4K: 128, 4-way</td>
</tr>
<tr>
<td></td>
<td>2M/4M: 8/8 thread</td>
<td>2M/4M: 8/8 thread</td>
</tr>
<tr>
<td>L1 Data TLB</td>
<td>4K: 64, 4-way</td>
<td>4K: 64, 4-way</td>
</tr>
<tr>
<td></td>
<td>2M/4M: 32, 4-way</td>
<td>2M/4M: 32, 4-way</td>
</tr>
<tr>
<td></td>
<td>1G: 4, 4-way</td>
<td>1G: 4, 4-way</td>
</tr>
<tr>
<td>L2 Unified TLB</td>
<td>4K+2M shared: 1536, 12-way</td>
<td>4K+2M shared: 1536, 12-way</td>
</tr>
<tr>
<td></td>
<td>1G: 16 entries</td>
<td>1G: 64 4-way entries</td>
</tr>
</tbody>
</table>

Note: Red = Differences
Memory hierarchy evolution summary

Based on Intel engineering estimates. Diagrams scale not representing actual metric values.

Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your 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.
High-Level Skylake Core Microarchitecture Improvements

**Improved front-end**
- Improved and larger Branch Predictor
- Wider and Deeper Instruction Supply

**Deeper Out-of-Order Buffers**
- Extract more instruction parallelism

**More execution units, shorter latencies**
- Improved Divider (Radix1024), 2x 128 bit

**More Load/Store Bandwidth than BDW**
- Prefetcher improvements
- Deeper store buffer, fill buffer and WB buffer
- Higher bandwidth

**Expandable core for datacenter specific enhancements**
- Second AVX-512 FMA
- Larger 1MB MLC

Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance.
Skylake Server Peak FLOPS

Skylake Server Peak Compute Throughput
- AVX-512 with 2 FMAs per core provide 2x peak FLOPs/cycle of Haswell/Broadwell
- 2x cache bandwidth to feed wider vector units
  - 64-byte load/store from L1
  - 2x L2 bandwidth

<table>
<thead>
<tr>
<th>uArch</th>
<th>Instruction Set</th>
<th>SP FLOPs per cycle</th>
<th>DP FLOPs per cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nehalem</td>
<td>SSE (128-bits)</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Sandy Bridge</td>
<td>AVX (256-bits)</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>Haswell / Broadwell</td>
<td>AVX2 &amp; FMA</td>
<td>32</td>
<td>16</td>
</tr>
<tr>
<td>Skylake Server</td>
<td>AVX512 &amp; FMA</td>
<td>64</td>
<td>32</td>
</tr>
</tbody>
</table>

Intel® AVX-512 Instruction Types

- **AVX512-F**: AVX-512 Foundation Instructions
- **AVX512-VL**: Vector Length Orthogonality: ability to operate on sub-512 vector sizes
- **AVX512-BW**: 512-bit Byte/Word support
- **AVX512-DQ**: Additional D/Q/SP/DP instructions (converts, transcendental support, etc.)
- **AVX512-CD**: Conflict Detect: used in vectorizing loops with potential address conflicts

Based on Intel engineering estimates.
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, 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.
## Intel® AVX (Sandy Bridge & Ivy Bridge)
- 16 SP/8 DP Flops/Cycle
- 256-bit basic FP
- 16 registers
- NDS (and AVX128)
- Improved blend
- MASKMOV
- Implicit unaligned

## Intel® AVX2 (Haswell & Broadwell)
- 32 SP/16 DP Flops/Cycle
- Float16
- 2 256-bit FP FMAs (Fused Multiply-Add)
- 256-bit integer
- PERMD
- Gather

## Intel® AVX-512 (Skylake-SP)
- Up to two 512-bit FMAs
- 512-bit FP and Integer
- 32 registers
- 8 mask registers
- 64 SP/32 DP Flops/Cycle (SKUs with 2 512-bit FMAs)
- 32 SP/16 DP Flops/Cycle (SKUs with 1 512-bit FMA)
- Embedded rounding
- Embedded broadcast
- Scalar/SSE/AVX “promotions”
- Native media additions
- HPC additions
- Transcendental support
- Gather/Scatter
**AVX512 features**

- **Higher throughput**
- **Greatly improved unrolling and inlining opportunities**

**32 vector registers, 512b wide: zmm0 through zmm31**
- Overlaid on top of existing YMM arch state
- Writing to xmm zeroes bits [511:128]
- Writing to ymm zeroes bits [511:256]

**8 mask registers, 64b wide: k0 through k7**
- KNL only uses bits [15:0] though (PS,PD,D,Q)
- EVEX.aaa=000 is an indicator of “no mask”
  - \{k0\} is illegal

<table>
<thead>
<tr>
<th>Feature</th>
<th>Skylake Server</th>
<th>Xeon Phi Knights Landing</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVX-512 foundation (F)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AVX-512 conflict detection (CDI)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AVX-512 exponential and reciprocal (ERI)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>AVX-512 prefetch (PFI)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>AVX-512 Byte and Word Instructions (BW)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>AVX-512 Doubleword and Quadword Instructions (DQ)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>AVX-512 Vector Length Orthogonality (VL)</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

The diagram illustrates the different types of registers and their bit ranges.

- **ZMM0-31**: 64 bytes
- **YMM0-15**: 32 bytes
- **XMM0-15**: 16 bytes

The red squares indicate features that are not supported in certain architectures.
Skylake Core Microarchitecture Enhancements

- 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

Server specific enhancements ➔ 2nd 512-bit FMA, larger 1MB L2 per core (vs. 256KB per core for client core)

**Skylake microarchitecture delivers ~10% (geomean) IPC improvement vs. Broadwell**

### Comparison Table

<table>
<thead>
<tr>
<th>Feature</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>64/4-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</td>
</tr>
<tr>
<td></td>
<td>1G: 64</td>
<td>1G: 64</td>
</tr>
</tbody>
</table>
Prior generation server CPUs typically decreased frequencies by 1 bin for each additional active core.

Skylake server processor reduces frequencies more gradually as the number of active cores increases.

Opportunity for large frequency increase at intermediate active core counts.
## Processor Frequency for SSE & AVX workloads

Intel Xeon E5-2697 v3 and Intel Xeon Platinum 8174 Processor at 205W TDP used for illustration purposes

<table>
<thead>
<tr>
<th>Non-AVX</th>
<th>Non-AVX Base Frequency</th>
<th>Non-AVX max All-Core Turbo frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7</td>
<td>1.8</td>
<td>2.6</td>
</tr>
<tr>
<td>2.1</td>
<td>2.2</td>
<td>2.7</td>
</tr>
<tr>
<td>2.3</td>
<td>2.4</td>
<td>2.8</td>
</tr>
<tr>
<td>2.5</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>3.0</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>3.2</td>
<td></td>
<td>3.3</td>
</tr>
<tr>
<td>3.4</td>
<td></td>
<td>3.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AVX2</th>
<th>AVX2 base frequency</th>
<th>AVX2 max All Core Turbo frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7</td>
<td>1.8</td>
<td>2.2</td>
</tr>
<tr>
<td>2.1</td>
<td>2.2</td>
<td>2.3</td>
</tr>
<tr>
<td>2.4</td>
<td>2.5</td>
<td>2.6</td>
</tr>
<tr>
<td>2.7</td>
<td>2.8</td>
<td>2.9</td>
</tr>
<tr>
<td>3.0</td>
<td>3.1</td>
<td>3.2</td>
</tr>
<tr>
<td>3.3</td>
<td>3.4</td>
<td>3.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AVX-512</th>
<th>AVX-512 base frequency</th>
<th>AVX-512 max All Core Turbo frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7</td>
<td>1.8</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1</td>
<td>2.2</td>
<td>2.3</td>
</tr>
<tr>
<td>2.4</td>
<td>2.5</td>
<td>2.6</td>
</tr>
<tr>
<td>2.7</td>
<td>2.8</td>
<td>2.9</td>
</tr>
<tr>
<td>3.0</td>
<td>3.1</td>
<td>3.2</td>
</tr>
<tr>
<td>3.3</td>
<td>3.4</td>
<td>3.4</td>
</tr>
</tbody>
</table>

AVX-512 “Heavy” Usage Examples

Significant gains for vector workloads while balancing power and frequency

Intel® AVX is designed to balance power consumed by lowering frequency when needed, while delivering significant performance gains and reduced runtimes.
Basic software tools enabling Skylake support

Compiler

- Targeting AVX512 ISA:
  -xCORE-AVX512
  -xSKYLAKE or -xSKYLAKE-AVX512
  -mtune=skylake or -mtune=skylake-avx512

- Controlling 512-bit register usage with -qopt-zmm-usage=<keyword>
  - high - generate zmm code without restrictions
  - low – tell compiler that is should avoid using 512-bit wide registers

Intel MPI Library

I_MPI_PLATFORM=skx
I_MPI_SHM=[skx_avx512 | skx_avx2 | skx_sse]
  with I_MPI_FABRICS=shm:ofi
Purley/Skylake Call To Action

The Skylake family of server processors brings amazing performance on demanding workloads

- Take advantage of Intel® AVX-512 on Intel® Compiler 19.0+ and Parallel Studio XE 2019 Update 3
- Take advantage of new larger mid-level cache size for reduced latency and higher core count for improved parallelism
- Explore localization benefits such as reduced latency for memory accesses and an increase in efficiency of the LLC due to elimination of duplicate cache lines
## Skylake Cache Hierarchy

<table>
<thead>
<tr>
<th></th>
<th>L1-I</th>
<th>L1-D</th>
<th>L2 (MLC)</th>
<th>Snoop Filter Slice</th>
<th>L3 (LLC) Slice</th>
<th>Directory in memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line Size (B)</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>N/A</td>
<td>64</td>
<td>2-bits/line</td>
</tr>
<tr>
<td>Sets</td>
<td>64</td>
<td>64</td>
<td>1024</td>
<td>2048</td>
<td>2048</td>
<td>N/A</td>
</tr>
<tr>
<td>Associativity</td>
<td>8-way</td>
<td>8-way</td>
<td>16-way</td>
<td>12-way</td>
<td>11-way</td>
<td>N/A</td>
</tr>
<tr>
<td>Cache Size</td>
<td>32KB</td>
<td>32KB</td>
<td>1MB</td>
<td>Eqv. 1.5MB</td>
<td>1.375MB</td>
<td>N/A</td>
</tr>
<tr>
<td>Latency (Cyc)</td>
<td>N/A</td>
<td>4</td>
<td>14</td>
<td>Variable</td>
<td>Variable</td>
<td>Variable</td>
</tr>
<tr>
<td>Peak BW (B/Cyc)</td>
<td>N/A</td>
<td>128+64</td>
<td>64</td>
<td>N/A</td>
<td>32</td>
<td>N/A</td>
</tr>
<tr>
<td>Inclusivity</td>
<td>N/A</td>
<td>N/A</td>
<td>L1-I, L1-D</td>
<td>All L2s &amp; IO</td>
<td>Non-incl</td>
<td>All L2 &amp; L3</td>
</tr>
<tr>
<td>Fill Policy</td>
<td>N/A</td>
<td>On a miss</td>
<td>On a miss</td>
<td>L2 Evict</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Update Policy</td>
<td>N/A</td>
<td>Writeback</td>
<td>N/A</td>
<td>Writeback</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>
Performance 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 www.intel.com/benchmarks.

**Configuration details:**

7. 1.6X average gains on HPC workloads comparing Intel® Xeon® Scalable processor to prior generation (geomean of MILC, VASP, OpenFoam, YASK is3dfd, miniFE, LAMMPS, DGEMM). Intel internal measurements as of June 2018: 1-node, 2-sockets of Intel® Xeon® Gold 6148, Platform: Wolf Pass / S2600WF/H48104-850, Memory configuration: 12 slots / 16 GB/ 2666 MT/s DDR4 RDIMM, Total Memory per Node: 192, Hyper-Threading: Yes, Turbo: Off, ucode: x043, OS: Red Hat Enterprise Linux* 7.4, Kernel: 2.10.0-693.11.0.el7.x86_64, Score: [MILC= 57.4 GFLOPs/sec, VASP=116.1 sec, OpenFoam=553, miniFE=30.4 Gflops, LAMMPS=74, DGEMM=223.9GFlops] vs. 1-node, 2-sockets of Intel® Xeon® E5-2699 v4, Platform: Grantley / S2600WT/H48296-300, Memory configuration: 8 slots / 16 GB/2400 MT/s DDR4 RDIMM, Total Memory per Node: 128, Hyper-Threading: Yes, Turbo: Off, ucode: 0x02A, OS: Red Hat Enterprise Linux* 7.4, Kernel: 3.10-693.21.1.el7.x86_64, Score: [MILC=40 GFLOPs/sec, VASP=209.4 sec, OpenFoam=879.4, miniFE=19.3GFlops, LAMMPS=33.3, DGEMM=1413 GFlops]

1c. Up to 1.55x on integer throughput performance - estimates based on the Intel® Xeon® Platinum 8180M processor on Wolf Pass SKX with 384 GB Total Memory on Red Hat Enterprise Linux* 7.4 using Benchmark software: SPEC CPU® 2017, Compiler: Intel® Compiler IC18 OEM, Optimized libraries: AVX512. ucode: 0x043, Data Source: Request Number: 40, Benchmark: SPECrate*2017_int_base, Score: 281 Higher is better vs. 1-Node, 2 x Intel® Xeon® Processor E5-2699 v4 on Wildcat Pass with 256 GB Total Memory on Red Hat Enterprise Linux* 7.4 using Benchmark software: SPEC CPU® 2017 v1.2, Optimized libraries: IC18.0_20170901, Other Software: MicroQuill SMART HEAP, Script / config files : xCORE-AVX2. ucode: 0x02A, Data Source: Request Number: 40, Benchmark: SPECrate*2017_int_base, Score: 181 Higher is better

1d. Up to 1.55x on technical compute app throughput - estimates based on Intel internal testing as of June 2018 on SPECfp* _rate_base2006: 1-Node, 2 x Intel® Xeon® Platinum 8180M Processor on Wolf Pass SKX with 384 GB Total Memory on Red Hat Enterprise Linux* 7.4 using Benchmark software: SPEC CPU® 2017, Compiler: Intel® Compiler IC18 OEM, Optimized libraries: AVX512. ucode: 0x043, Data Source: Request Number: 39, Benchmark: SPECrate*2017_fp_base, Score: 236 Higher is better vs. 1-Node, 2 x Intel® Xeon® Processor E5-2699 v4 on Wildcat Pass with 256 GB Total Memory on Red Hat Enterprise Linux* 7.4 using Benchmark software: SPEC CPU® 2017 v1.2, Optimized libraries: IC18.0_20170901, Other Software: MicroQuill SMART HEAP, Script / config files : xCORE-AVX2. ucode: 0x02A, Data Source: Request Number: 39, Benchmark: SPECrate*2017_fp_base, Score: 148 Higher is better

1e. Up to 1.6x on est STREAM - triad - estimates based on Intel internal testing as of June 2018 on STREAM - triad: 1-Node, 2 x Intel® Xeon® Platinum 8180M Processor on Wolf Pass SKX with 384 GB Total Memory on Red Hat Enterprise Linux* 7.4 using Benchmark software: STREAM, Compiler: Intel® Compiler IC17, Optimized libraries: AVX512. ucode: 0x043, Data Source: Request Number: 37, Benchmark: STREAM - Triad, Score: 201.24 Higher is better vs. 1-Node, 2 x Intel® Xeon® Processor E5-2699 v4 on Wildcat Pass with 256 GB Total Memory on Red Hat Enterprise Linux* 7.4 using Benchmark software: STREAM, Optimized libraries: IC16, Other Software: AVX2. ucode: 0x02A, Data Source: Request Number: 37, Benchmark: STREAM - Triad, Score: 124.78 Higher is better

1i. Up to 2.2x Linpack throughput - estimates based on Intel internal testing as of June 2018 on Intel® Distribution of LINPACK: 1-Node, 2 x Intel® Xeon® Platinum 8180M Processor on Wolf Pass SKX with 384 GB Total Memory on Red Hat Enterprise Linux* 7.4 OS Kernel: 3.10.0-693.11.6.el7.x86_64, Update uCode: 0x043 using Benchmark software: MP Linpack 2018.0.006, Compiler: l_mpi_2018.1.163, Optimized libraries: AVX512, Array 80000, Data Source: Request Number: 38, Benchmark: Intel® Distribution of LINPACK, Score: 3367.5 Higher is better vs. 1-Node, 2 x Intel® Xeon® Processor E5-2699 v4 on Wildcat Pass with 256 GB Total Memory on Red Hat Enterprise Linux* 7.4 OS Kernel: 3.10.0-693.21.1.el7.x86_64, uCode: 0x02A using Benchmark software: MP Linpack 2018.0.006, Optimized libraries: l_mpi_2018.1.163, AVX2, Array 80000, Other Software: MicroQuill SMART HEAP, Script / config files : xCORE-AVX2. Benchmark: Intel® Distribution of LINPACK, Score: 1427.23 Higher is better