CPU ARCHITECTURES

Dr Christopher Dahnken
SSG DRD EMEA Datacenter
INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS”. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO THIS INFORMATION INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

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.

Copyright © 2017, Intel Corporation. All rights reserved. Intel, Pentium, Xeon, Xeon Phi, Core, VTune, Cilk, and the Intel logo are trademarks of Intel Corporation in the U.S. and other countries.

<table>
<thead>
<tr>
<th>Optimization Notice</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
</tr>
</tbody>
</table>

Notice revision #20110804
Agenda

• Xeon Phi: Knights Landing
• Xeon: Sky Lake
• Short outlook – Non-Volatile Memory/Optane
KNL Architecture Overview

**ISA**
Intel® Xeon® Processor Binary- Compatible (w/ Broadwell)

**On-package memory**
Up to 16GB, ~500 GB/s STREAM at launch

**Platform Memory**
Up to 384GB (6ch DDR4-2400 MHz)

**Fixed Bottlenecks**
- 2D Mesh Architecture
- Out-of-Order Cores
- 3x single-thread vs. KNC

TILE: (up to 36)
- 2VPU
- HUB
- 1MB L2
- 2VPU

Enhanced Intel® Atom™ cores based on Silvermont™ Microarchitecture

---

x4 DMI2 to PCH
36 Lanes PCIe* Gen3 (x16, x16, x4)

MCDRAM

DDDR4

2VPU Core

2VPU Core

1MB L2

HUB

Tile

EDC (embedded DRAM controller)

IMC (integrated memory controller)

IIO (integrated I/O controller)
KNL Mesh Interconnect

Mesh of Rings
- Every row and column is a (half) ring
- YX routing: Go in Y → Turn → Go in X
- Messages arbitrate at injection and on turn

Cache Coherent Interconnect
- MESIF protocol (F = Forward)
- Distributed directory to filter snoops

Three Cluster Modes
1. All-to-All
2. Quadrant
3. Sub-NUMA Clustering
Cluster Mode: All-to-All

Address uniformly hashed across all distributed directories

No affinity between Tile, Directory and Memory

Lower performance mode, compared to other modes. Mainly for fall-back

Typical Read L2 miss
1. L2 miss encountered
2. Send request to the distributed directory
3. Miss in the directory. Forward to memory
4. Memory sends the data to the requestor

1) L2 miss, 2) Directory access, 3) Memory access, 4) Data return
Cluster Mode: Quadrant

Chip divided into four virtual Quadrants

Address hashed to a Directory in the same quadrant as the Memory

Affinity between the Directory and Memory

Lower latency and higher BW than all-to-all. SW Transparent.

1) L2 miss, 2) Directory access, 3) Memory access, 4) Data return
Cluster Mode: Sub-NUMA Clustering (SNC)

Each Quadrant (Cluster) exposed as a separate NUMA domain to OS.

Looks analogous to 4-Socket Xeon

Affinity between Tile, Directory and Memory

Local communication. Lowest latency of all modes.

SW needs to NUMA optimize to get benefit.

1) L2 miss, 2) Directory access, 3) Memory access, 4) Data return
KNL Core and VPU

Out-of-order core w/ 4 SMT threads
VPU tightly integrated with core pipeline
2-wide decode/rename/retire
2x 64B load & 1 64B store port for D$
L1 prefetcher and L2 prefetcher
Fast unaligned and cache-line split support
Fast gather/scatter support
KNL Memory Modes

- Mode selected at boot
- MCDRAM-Cache covers all DDR

Flat Models

Physical Address

Cache Model

Hybrid Model
## MCDRAM: Cache vs Flat Mode

<table>
<thead>
<tr>
<th>Mode</th>
<th>Software Effort</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDR Only</td>
<td>No software changes required</td>
<td>Not peak performance.</td>
</tr>
<tr>
<td>MCDRAM as Cache</td>
<td>Change allocations for bandwidth-critical data.</td>
<td>Best performance.</td>
</tr>
<tr>
<td>MCDRAM Only</td>
<td>Limited memory capacity</td>
<td>Optimal HW utilization + opportunity for new algorithms</td>
</tr>
<tr>
<td>Flat DDR + MCDRAM</td>
<td>Recommended</td>
<td></td>
</tr>
<tr>
<td>Hybrid</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Summary
- **DDR Only**: Not recommended due to limited memory capacity.
- **MCDRAM as Cache**: Recommended for optimal HW utilization and opportunity for new algorithms.
- **MCDRAM Only**: Recommended for peak performance.
- **Flat DDR + MCDRAM**: Recommended for best performance.
KNL Instruction Set

- **Xeon 5600** "Nehalem"
- **Xeon E5-2600** "Sandy Bridge"
- **Xeon E5-2600v3** "Haswell"
- **Xeon Phi** "Knights Landing"
- **Xeon** Sky Lake
# 2-socket+ Intel® Xeon® Roadmap

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Thurley Platform</td>
<td>Intel® Microarchitecture Codenamed Nehalem</td>
<td>45nm</td>
<td>New Microarchitecture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>32nm</td>
<td>Westmere</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intel® Microarchitecture Codenamed Sandy Bridge</td>
<td>32nm</td>
<td>Sandy Bridge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Romley Platform</td>
<td>Intel® Microarchitecture Codenamed Haswell</td>
<td>22nm</td>
<td>Ivy Bridge</td>
<td>Haswell</td>
<td>Broadwell</td>
<td></td>
</tr>
<tr>
<td>Grantley Platform</td>
<td>Intel® Microarchitecture Codenamed Skylake</td>
<td>14nm</td>
<td>Skylake</td>
<td></td>
<td>Future</td>
<td></td>
</tr>
<tr>
<td>Purley Platform</td>
<td></td>
<td>14nm</td>
<td>Future</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Brickland Platform is Ivy Bridge-EX, Haswell-EX, and Broadwell-EX

**Skylake microarchitecture delivers ~10% (geomean) IPC improvement v. Broadwell**
New Skylake Uncore Interconnect Architecture

**Broadwell Server 24-core die – dual-ring interconnect**

**Skylake (or Cascade Lake) Server 28-core die – mesh interconnect**

Mesh interconnect (Skylake Server) replaces dual-ring interconnect (BDW E5/E7)
Core Microarchitecture Enhancements

- Larger and improved branch predictor, higher throughput decoder, larger window to extract ILP
- 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 MLC

ABOUT 10% PERFORMANCE IMPROVEMENT PER CORE ON INTEGER APPLICATIONS AT SAME FREQUENCY
## Intel® Xeon® Scalable Processor Feature Overview

### Feature Details

<table>
<thead>
<tr>
<th>Feature</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Socket</strong></td>
<td>Socket P</td>
</tr>
<tr>
<td><strong>Scalability</strong></td>
<td>2S, 4S, 8S, and &gt;8S (with node controller support)</td>
</tr>
<tr>
<td><strong>CPU TDP</strong></td>
<td>70W – 205W</td>
</tr>
<tr>
<td><strong>Chipset</strong></td>
<td>Intel® C620 Series (code name Lewisburg)</td>
</tr>
<tr>
<td><strong>Networking</strong></td>
<td>Intel® Omni-Path Fabric (integrated or discrete)</td>
</tr>
<tr>
<td></td>
<td>4x10GbE (integrated w/ chipset)</td>
</tr>
<tr>
<td></td>
<td>100G/40G/25G discrete options</td>
</tr>
<tr>
<td><strong>Compression and Crypto Acceleration</strong></td>
<td>Intel® QuickAssist Technology to support 100Gb/s comp/decomp/crypto 100K RSA2K public key</td>
</tr>
<tr>
<td><strong>Storage</strong></td>
<td>Integrated QuickData Technology, VMD, and NTB</td>
</tr>
<tr>
<td></td>
<td>Intel® Optane™ SSD, Intel® 3D-NAND NVMe &amp; SATA SSD</td>
</tr>
<tr>
<td><strong>Security</strong></td>
<td>CPU enhancements (MBE, PPK, MPX)</td>
</tr>
<tr>
<td></td>
<td>Manageability Engine</td>
</tr>
<tr>
<td></td>
<td>Intel® Platform Trust Technology</td>
</tr>
<tr>
<td></td>
<td>Intel® Key Protection Technology</td>
</tr>
<tr>
<td><strong>Manageability</strong></td>
<td>Innovation Engine (IE)</td>
</tr>
<tr>
<td></td>
<td>Intel® Node Manager</td>
</tr>
<tr>
<td></td>
<td>Intel® Datacenter Manager</td>
</tr>
</tbody>
</table>

### Diagram

- **Skylake-SP CPU**: 2 or 3 Intel® UPI
- **3x16 PCIe* Gen3**: 1x 100Gb OPA Fabric
- **DDR4 2666**: 3x16 PCIe Gen3
- **Supplementary Features**:
  - BMC: Baseboard Management Controller
  - PCH: Intel® Platform Controller Hub
  - IE: Innovation Engine
  - Intel OPA: Intel® Omni-Path Architecture
  - Intel QAT: Intel® QuickAssist Technology
  - ME: Manageability Engine
  - NIC: Network Interface Controller
  - VMD: Volume Management Device
  - NTB: Non-Transparent Bridge

---

*PCle*: PCI Express
Platform Topologies

2S Configurations

4S Configurations

8S Configuration

**INTEL® XEON® SCALABLE PROCESSOR SUPPORTS CONFIGURATIONS RANGING FROM 2S-2UPI TO 8S**
intel®