The Intel® Platform for High Performance Computing

Dr-Ing. Michael Klemm
Software and Services Group
Intel
(michael.klemm@intel.com)
Legal Disclaimer & Optimization Notice

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

Intel Confidential

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
Transforming the Economics of HPC

Executing to Moore’s Law

Predictable Silicon Track Record – well and alive at Intel. Enabling new devices with higher performance and functionality while controlling power, cost, and size.

Future options subject to change without notice.
Driving Innovation and Integration
Enabled by Leading Edge Process Technologies

Integrated Today

Coming in the Future
The Magic of Integration
Moore's Law at Work & Architecture Innovations

1970s
150 MFLOPS
CRAY-1

2013
1000000 MFLOPS
Intel® Xeon Phi™

6666x
#1 TOP500 June 2013

33 PFLOPS HPL

54 PFLOPS Peak

32000 Intel® Xeon® E5v2 Processors

48000 Intel® Xeon Phi™ Coprocessors
TOP500 Highlights

427 of 500 (85%) of all systems
111 of 114 (97%) of new systems

PRACE ISC Award—2014
1st Sustained 1PFlop Real Science Performance on an IA-based System
Intel® Xeon Processor Architecture
## Intel “Tick-Tock” Roadmap – Part I

<table>
<thead>
<tr>
<th>Micro Architecture</th>
<th>Codename “Nehalem”</th>
<th>2nd Generation Intel® Core™ Micro Architecture</th>
<th>3rd Generation Intel® Core™ Micro Architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merom</td>
<td>NEW Micro architecture</td>
<td>65nm</td>
<td>NEW Process Technology</td>
</tr>
<tr>
<td>Penryn</td>
<td>NEW Process Technology</td>
<td>45nm</td>
<td>Nehalem</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Westmere</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sandy Bridge</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ivy Bridge</td>
</tr>
</tbody>
</table>

**Intel® Core™ MicroArchitecture**

- **Merom**
  - NEW Micro architecture
  - 65nm
  - **TOCK**
  - 2006 SSSE-3

- **Penryn**
  - NEW Process Technology
  - 45nm
  - **TICK**
  - 2007 SSE4.1

- **Nehalem**
  - NEW Micro architecture
  - 45nm
  - **TOCK**
  - 2008 SSE4.2

- **Westmere**
  - NEW Process Technology
  - 32nm
  - **TICK**
  - 2009 AES

- **Sandy Bridge**
  - NEW Micro architecture
  - 32nm
  - **TOCK**
  - 2011 AVX

- **Ivy Bridge**
  - NEW Process Technology
  - 22nm
  - **TICK**
  - 2012 RDRAND etc

*Other brands and names are the property of their respective owners.*

Copyright © 2014, Intel Corporation. All rights reserved.
# Intel “Tick-Tock” Roadmap – Part II

*Future Release Dates & Features subject to Change without Notice!*

<table>
<thead>
<tr>
<th>4th Generation Intel® Core™ Micro Architecture</th>
<th>TBD</th>
<th>TBD</th>
<th>TBD</th>
<th>TBD</th>
<th>TBD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Haswell</strong></td>
<td><strong>Broadwell</strong></td>
<td><strong>Skylake</strong></td>
<td><strong>TBD</strong></td>
<td><strong>TBD</strong></td>
<td><strong>TBD</strong></td>
</tr>
<tr>
<td>22nm</td>
<td>14nm</td>
<td>14nm</td>
<td>10nm</td>
<td>10nm</td>
<td>7nm</td>
</tr>
<tr>
<td><strong>TICK</strong></td>
<td><strong>TOCK</strong></td>
<td><strong>TICK</strong></td>
<td><strong>TOCK</strong></td>
<td><strong>TICK</strong></td>
<td><strong>TOCK</strong></td>
</tr>
<tr>
<td>2013</td>
<td>September 2014!</td>
<td>2015/6 (?)</td>
<td>???</td>
<td>???</td>
<td>???</td>
</tr>
<tr>
<td>AVX-2</td>
<td>5 new Inst.</td>
<td>AVX-512 (?)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Copyright © 2014, Intel Corporation. All rights reserved. *Other brands and names are the property of their respective owners.*
Recap: Sandy Bridge and Ivy Bridge Execution Units

256K L2 Cache (Unified)

Scheduler (54)
- Port 0
  - ALU
  - V-Mul
  - V-Shuf
  - FDiv
  - 256 FP Mul
  - 256 FP Blend
- Port 1
  - ALU
  - V-Add
  - V-Shuf
- Port 5
  - ALU
  - JMP
  - 256 FP Shuf
  - 256 FP Bool
  - 256 FP Blend
- Port 2
  - Load
- Port 3
  - Load
- Port 4
  - STD

Memory Control

Allocating/Renaming/Retiring (4)

Inst. Queue
Decoders (4)

Branch Predictor

Pre-Decode

32K L1 ICache (8way)
32K L1 DCache (8way)

48 byte/cycles

16 byte/cycle

Out-of-order (168)

165x164
16byte/cycle

Copyright © 2014, Intel Corporation. All rights reserved. *Other brands and names are the property of their respective owners.
Haswell Core at a Glance

Next generation branch prediction
- Improves performance and saves wasted work

Improved front-end
- Initiate TLB and cache misses speculatively
- Handle cache misses in parallel to hide latency
- Leverages improved branch prediction

Deeper buffers
- Extract more instruction parallelism
- More resources when running a single thread

More execution units, shorter latencies
- Power down when not in use

More load/store bandwidth
- Better prefetching, better cache line split latency & throughput, double L2 bandwidth

No pipeline growth
- Same branch misprediction latency
- Same L1/L2 cache latency
Haswell Execution Unit Overview

Unified Reservation Station

Port 0
- Integer ALU & Shift
- FMA
- FP Multiply
- Vector Int Multiply
- Vector Logicals
- Branch
- Divide
- Vector Shifts

Port 1
- Integer ALU & LEA
- FMA FP Mult
- FP Add
- Vector Int ALU
- Vector Logicals

Port 2
- Load & Store Address
- 2xFMA
- Doubles peak FLOPs
- Two FP multiplies benefits legacy

Port 3
- Store Data

Port 4
- Integer ALU & LEA

Port 5
- Integer ALU & Shift
- Vector Shuffle
- Vector Int ALU
- Vector Logicals

Port 6
- Store Address
- Branch
- New Branch Unit
- Reduces Port0 Conflicts
- 2nd EU for high branch code

New AGU for Stores
- Leaves Port 2 & 3 open for Loads

4th ALU
- Great for integer workloads
- Frees Port0 & 1 for vector

Intel® Microarchitecture (Haswell)
## Haswell Buffer Sizes

Extract more parallelism in every generation

<table>
<thead>
<tr>
<th></th>
<th>Nehalem</th>
<th>Sandy Bridge</th>
<th>Haswell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Out-of-order Window</td>
<td>128</td>
<td>168</td>
<td>192</td>
</tr>
<tr>
<td>In-flight Loads</td>
<td>48</td>
<td>64</td>
<td>72</td>
</tr>
<tr>
<td>In-flight Stores</td>
<td>32</td>
<td>36</td>
<td>42</td>
</tr>
<tr>
<td>Scheduler Entries</td>
<td>36</td>
<td>54</td>
<td>60</td>
</tr>
<tr>
<td>Integer Register File</td>
<td>N/A</td>
<td>160</td>
<td>168</td>
</tr>
<tr>
<td>FP Register File</td>
<td>N/A</td>
<td>144</td>
<td>168</td>
</tr>
<tr>
<td>Allocation Queue</td>
<td>28/thread</td>
<td>28/thread</td>
<td>56</td>
</tr>
</tbody>
</table>
## Core Cache Size/Latency/Bandwidth

<table>
<thead>
<tr>
<th>Metric</th>
<th>Nehalem</th>
<th>Sandy Bridge</th>
<th>Haswell</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L1 Instruction Cache</strong></td>
<td>32K, 4-way</td>
<td>32K, 8-way</td>
<td>32K, 8-way</td>
</tr>
<tr>
<td><strong>L1 Data Cache</strong></td>
<td>32K, 8-way</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>
<td>4 cycles</td>
</tr>
<tr>
<td>Load bandwidth</td>
<td>16 Bytes/cycle</td>
<td>32 Bytes/cycle (banked)</td>
<td>64 Bytes/cycle</td>
</tr>
<tr>
<td>Store bandwidth</td>
<td>16 Bytes/cycle</td>
<td>16 Bytes/cycle</td>
<td>32 Bytes/cycle</td>
</tr>
<tr>
<td><strong>L2 Unified Cache</strong></td>
<td>256K, 8-way</td>
<td>256K, 8-way</td>
<td>256K, 8-way</td>
</tr>
<tr>
<td>Fastest load-to-use</td>
<td>10 cycles</td>
<td>11 cycles</td>
<td>11 cycles</td>
</tr>
<tr>
<td>Bandwidth to L1</td>
<td>32 Bytes/cycle</td>
<td>32 Bytes/cycle</td>
<td>64 Bytes/cycle</td>
</tr>
<tr>
<td><strong>L1 Instruction TLB</strong></td>
<td>4K: 128, 4-way, 2M/4M: 7/thread</td>
<td>4K: 128, 4-way, 2M/4M: 8/thread</td>
<td>4K: 128, 4-way, 2M/4M: 8/thread</td>
</tr>
<tr>
<td><strong>L1 Data TLB</strong></td>
<td>4K: 64, 4-way, 2M/4M: 32, 4-way, 1G: fractured</td>
<td>4K: 64, 4-way, 2M/4M: 32, 4-way, 1G: 4, 4-way</td>
<td>4K: 64, 4-way, 2M/4M: 32, 4-way, 1G: 4, 4-way</td>
</tr>
<tr>
<td><strong>L2 Unified TLB</strong></td>
<td>4K: 512, 4-way</td>
<td>4K: 512, 4-way</td>
<td>4K+2M shared: 1024, 8-way</td>
</tr>
</tbody>
</table>
# New Instructions in Haswell

<table>
<thead>
<tr>
<th>Group</th>
<th>Description</th>
<th>Count *</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVX-2</td>
<td><strong>SIMD Integer Instructions promoted to 256 bits</strong></td>
<td>170 / 124</td>
</tr>
<tr>
<td></td>
<td>Adding vector integer operations to 256-bit</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Gather</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Load elements using a vector of indices, vectorization enabler</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Shuffling / Data Rearrangement</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blend, element shift and permute instructions</td>
<td></td>
</tr>
<tr>
<td>FMA</td>
<td>Fused Multiply-Add operation forms (FMA-3)</td>
<td>96 / 60</td>
</tr>
<tr>
<td>Bit Manipulation and Cryptography</td>
<td>Improving performance of bit stream manipulation and decode, large integer arithmetic and hashes</td>
<td>15 / 15</td>
</tr>
<tr>
<td>TSX=RTM+HLE</td>
<td>Transactional Memory</td>
<td>4 / 4</td>
</tr>
<tr>
<td>Others</td>
<td>MOVBE: Load and Store of Big Endian forms</td>
<td>2 / 2</td>
</tr>
<tr>
<td></td>
<td>INVPCID: Invalidate processor context ID</td>
<td></td>
</tr>
</tbody>
</table>

* Total instructions / different mnemonics
FMA: Fused Multiply Add Instruction
Improves accuracy and performance for commonly used class of algorithms

Mirco-Architecture | Instruction Set | SP FLOPs per cycle | DP FLOPs per cycle
--- | --- | --- | ---
Nehalem | SSE (128-bits) | 8 | 4
Sandy Bridge | AVX (256-bits) | 16 | 8
Haswell | AVX2 (FMA) (256-bits) | 32 | 16

2x peak FLOPs/cycle (throughput)

Latency (clocks) | Xeon E5 v2 | Xeon E5 v3 | Ratio *
--- | --- | --- | ---
MulPS, PD | 5 | 5 | *
AddPS, PD | 3 | 3 | *
Mul+Add /FMA | 8 | 5 | 0.625

>37% reduced latency
(5-cycle FMA latency same as an FP multiply)

Increased performance potential for Technical Computing workloads like Structural Analysis, CFD, EMF computation, Cosmology, .... *

Copyright © 2014, Intel Corporation. All rights reserved. *Other brands and names are the property of their respective owners.
Intel® Xeon Platforms
Intel® Xeon® Processors

Intel® Xeon® E3

Intel® Xeon® E5

Intel® Xeon® E7

Memory

E3

PCIe3

E5

PCIe3

Memory

2x QPI

PCIe3

Memory

3x QPI

PCIe3
Intel® Xeon® Processors & Platforms

Intel® Xeon®
- E5-1xxx
- E5-2xxx
- E5-4xxx
- E7-xxxx

Intel® Xeon®
- E3-1xxx

CPU/Socket

QPI

>4S
Intel® Xeon® E5-2600v3 Processor Overview

- 22nm Process (Tock)
- PCI Express 3.0 EP: 40 Lanes
- Intel® Hyper-Threading Technology (2 threads/core)
- Intel® Turbo Boost Technology
- Up to 18 Cores
- Integrated Voltage Regulator

Power Management
- Per Core P-State (PCPS)
- Uncore Frequency Scaling (UFS)
- Energy Efficient Turbo (EET)

Memory Technology:
- Socket R3
- 4xDDR4 channels
- 1333, 1600, 1866, 2133 MTS
- ~2.5 MB Last Level Cache/Core
- Up to 45 MB total LLC

New Feature
- Intel® AVX 2.0 / Haswell New Instruction (HNI)
- Intel® QuickPath Interface (x2)
- 9.6GT/s

Existing Feature
- Intel® QuickPath Interface
- System Agent
- DMI
- PCIe3.0
- IMC
Key Differences Between E5-2600 v2 & E5-2600 v3

<table>
<thead>
<tr>
<th></th>
<th>Xeon E5-2600 v2</th>
<th>Xeon E5-2600 v3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core Count</td>
<td>Up to 12 Cores</td>
<td>Up to 18 Cores</td>
</tr>
<tr>
<td>Frequency</td>
<td>TDP &amp; Turbo Frequencies</td>
<td>TDP &amp; Turbo Freq AVX &amp; AVX Turbo Freq</td>
</tr>
<tr>
<td>AVX Support</td>
<td>AVX 1 8 DP Flops/Clock/Core</td>
<td>AVX 2 16 DP Flops/Clock/Core</td>
</tr>
<tr>
<td>Memory Type</td>
<td>4xDDR3 channels RDIMM, UDIMM, LRDIMM</td>
<td>4xDDR4 channels RDIMM, LRDIMM</td>
</tr>
<tr>
<td>Memory Frequency</td>
<td>1866 (1DPC), 1600, 1333, 1066</td>
<td>RDIMM: 2133 (1DPC), 1866 (2DPC), 1600 LRDIMM: 2133 (1&amp;2DPC), 1600</td>
</tr>
<tr>
<td>QPI Speed</td>
<td>Up to 8.0 GT/s</td>
<td>Up to 9.6 GT/s</td>
</tr>
<tr>
<td>TDP</td>
<td>Up to 130W Server, 150W Workstation</td>
<td>Up to 145W Server, 160W Workstation (Increase due to Integrated VR)</td>
</tr>
<tr>
<td>Power Management</td>
<td>Same P-states for all cores Same core &amp; uncore frequency</td>
<td>Per-core P-states Independent uncore frequency scaling Energy Efficient Turbo</td>
</tr>
</tbody>
</table>
# Intel® Xeon® Processor E5-2600 v3 Product Family SKU Stack

## Advanced
- 25-30 MB LLC cache
- 9.6 GT/s QPI
- DDR4-2133
- Intel® Hyper-Threading
- Intel® Turbo boost

<table>
<thead>
<tr>
<th>CPU Count</th>
<th>Frequency</th>
<th>Power</th>
<th>SKU</th>
</tr>
</thead>
<tbody>
<tr>
<td>12C</td>
<td>2.6 GHz</td>
<td>135W</td>
<td>E5-2690v3</td>
</tr>
<tr>
<td>12C</td>
<td>2.5 GHz</td>
<td>120W</td>
<td>E5-2680v3</td>
</tr>
<tr>
<td>12C</td>
<td>2.3 GHz</td>
<td>120W</td>
<td>E5-2670v3</td>
</tr>
<tr>
<td>10C</td>
<td>2.6 GHz</td>
<td>105W</td>
<td>E5-2660v3</td>
</tr>
<tr>
<td>10C</td>
<td>2.3 GHz</td>
<td>105W</td>
<td>E5-2650v3</td>
</tr>
</tbody>
</table>

## Standard
- 15-20 MB LLC cache
- 8.0 GT/s QPI
- DDR4-1866
- Intel® Hyper-Threading
- Intel® Turbo boost

<table>
<thead>
<tr>
<th>CPU Count</th>
<th>Frequency</th>
<th>Power</th>
<th>SKU</th>
</tr>
</thead>
<tbody>
<tr>
<td>8C</td>
<td>2.6 GHz</td>
<td>90W</td>
<td>E5-2640v3</td>
</tr>
<tr>
<td>8C</td>
<td>2.4 GHz</td>
<td>85W</td>
<td>E5-2630v3</td>
</tr>
<tr>
<td>6C</td>
<td>2.4 GHz</td>
<td>85W</td>
<td>E5-2620v3</td>
</tr>
</tbody>
</table>

## Basic
- 15 MB LLC cache
- 6.4 GT/s QPI
- DDR4-1600

<table>
<thead>
<tr>
<th>CPU Count</th>
<th>Frequency</th>
<th>Power</th>
<th>SKU</th>
</tr>
</thead>
<tbody>
<tr>
<td>6C</td>
<td>1.9 GHz</td>
<td>85W</td>
<td>E5-2609v3</td>
</tr>
<tr>
<td>6C</td>
<td>1.6 GHz</td>
<td>85W</td>
<td>E5-2603v3</td>
</tr>
</tbody>
</table>

## Segmented Optimized
- 45-35 MB LLC cache
- 9.6 GT/s QPI
- DDR4-2133
- Intel® Hyper-Threading
- Intel® Turbo boost

<table>
<thead>
<tr>
<th>CPU Count</th>
<th>Frequency</th>
<th>Power</th>
<th>SKU</th>
</tr>
</thead>
<tbody>
<tr>
<td>18C (2U)</td>
<td>2.3 GHz</td>
<td>145W</td>
<td>E5-2699v3</td>
</tr>
<tr>
<td>16C (1U)</td>
<td>2.3 GHz</td>
<td>135W</td>
<td>E5-2698v3</td>
</tr>
<tr>
<td>14C (2U)</td>
<td>2.6 GHz</td>
<td>145W</td>
<td>E5-2697v3</td>
</tr>
<tr>
<td>14C (1U)</td>
<td>2.3 GHz</td>
<td>120W</td>
<td>E5-2695v3</td>
</tr>
<tr>
<td>14C (1U)</td>
<td>2.0 GHz</td>
<td>120W</td>
<td>E5-2693v3</td>
</tr>
</tbody>
</table>

## Frequency Optimized
- ≥2.5MB/core LLC cache
- 9.6 GT/s QPI
- DDR4-2133
- Intel® Hyper-Threading
- Intel® Turbo boost

<table>
<thead>
<tr>
<th>CPU Count</th>
<th>Frequency</th>
<th>Power</th>
<th>SKU</th>
</tr>
</thead>
<tbody>
<tr>
<td>8C (20M)</td>
<td>3.2 GHz</td>
<td>135W</td>
<td>E5-2667v3</td>
</tr>
<tr>
<td>6C (20M)</td>
<td>3.4 GHz</td>
<td>135W</td>
<td>E5-2643v3</td>
</tr>
<tr>
<td>4C (15M)</td>
<td>3.5GHz</td>
<td>135W</td>
<td>E5-2637v3</td>
</tr>
</tbody>
</table>

## Workstation Only
- 10MB LLC cache
- 8.0GT/s QPI
- DDR4-1866
- Intel® Hyper-Threading
- Intel® Turbo boost

<table>
<thead>
<tr>
<th>CPU Count</th>
<th>Frequency</th>
<th>Power</th>
<th>SKU</th>
</tr>
</thead>
<tbody>
<tr>
<td>10C (2S)</td>
<td>3.1 GHz</td>
<td>160W</td>
<td>E5-2687Wv3</td>
</tr>
</tbody>
</table>

## Low Power
- 20-30 MB LLC Cache
- 9.6 GT/s QPI
- DDR4-2133
- Intel® Hyper-Threading
- Intel® Turbo boost

<table>
<thead>
<tr>
<th>CPU Count</th>
<th>Frequency</th>
<th>Power</th>
<th>SKU</th>
</tr>
</thead>
<tbody>
<tr>
<td>12C</td>
<td>1.8 GHz</td>
<td>65W</td>
<td>E5-2650Lv3</td>
</tr>
<tr>
<td>8C</td>
<td>1.8 GHz</td>
<td>55W</td>
<td>E5-2630Lv3</td>
</tr>
</tbody>
</table>
On-Die Interconnect Enhancements

**E5-2600 v2**

- PCIe
- QPI
- Shared L3 Cache (30MB)
- Memory Controller

**E5-2600 v3**

- PCIe
- QPI
- Buffered switch
- Shared L3 Cache (45MB)
- Memory Controller
Haswell EP Die Configurations

14-18 Core (HCC)

10-12 Core (MCC)

4-8 Core (LCC)

Not representative of actual die-sizes, orientation and layouts – for informational use only.

<table>
<thead>
<tr>
<th>Chop</th>
<th>Columns</th>
<th>Home Agents</th>
<th>Cores</th>
<th>Power (W)</th>
<th>Transitors (B)</th>
<th>Die Area (mm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCC</td>
<td>4</td>
<td>2</td>
<td>14-18</td>
<td>110-145</td>
<td>5.69</td>
<td>662</td>
</tr>
<tr>
<td>MCC</td>
<td>3</td>
<td>2</td>
<td>6-12</td>
<td>65-160</td>
<td>3.84</td>
<td>492</td>
</tr>
<tr>
<td>LCC</td>
<td>1</td>
<td>1</td>
<td>4-8</td>
<td>55-140</td>
<td>2.60</td>
<td>354</td>
</tr>
</tbody>
</table>
## Haswell Processor Improvements

<table>
<thead>
<tr>
<th>Area</th>
<th>Change</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>On-die interconnect</strong></td>
<td>• Two Fully Buffered Rings</td>
<td>• Enables higher core counts and provides higher bandwidth per core.</td>
</tr>
<tr>
<td><strong>Home Agent / Memory Controller</strong></td>
<td>• DDR4</td>
<td>• Increased memory bandwidth and power efficiency</td>
</tr>
<tr>
<td></td>
<td>• Two Home Agents in more SKUs</td>
<td>• Greater socket BW with more outstanding requests</td>
</tr>
<tr>
<td></td>
<td>• Directory Cache</td>
<td>• Lower average memory latency</td>
</tr>
<tr>
<td><strong>LLC</strong></td>
<td>• Cluster On Die (COD) mode</td>
<td>• Increased performance, reduced latency</td>
</tr>
<tr>
<td></td>
<td>• Improved LLC allocation policy</td>
<td>• Enables improved performance by better application placement in a virtualized environment</td>
</tr>
<tr>
<td></td>
<td>• Cache Allocation Monitoring</td>
<td></td>
</tr>
<tr>
<td><strong>Power Management</strong></td>
<td>• Separate clock and voltage domains for each core and uncore (enables PCPS, UFS)</td>
<td>• Better performance per watt</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lower socket idle (package C6) power.</td>
</tr>
<tr>
<td><strong>QPI 1.1</strong></td>
<td>• Increase to 9.6GT/s</td>
<td>• Multi-socket coherence performance</td>
</tr>
<tr>
<td><strong>Integrated IO-Hub (IIO)</strong></td>
<td>• LLC cache tracks IIO cache line ownership</td>
<td>• Improves PCIe bandwidth under conflicts (concurrent accesses to the same cache line).</td>
</tr>
<tr>
<td></td>
<td>• Increased PCIe buffers and credits</td>
<td>• Increase PCIe bandwidth and latency tolerance</td>
</tr>
<tr>
<td><strong>PCI Express 3.0</strong></td>
<td>• DualCast - Allows a single write transaction to multiple targets.</td>
<td>• Utilized to minimize memory channel bandwidth – data can be sent to memory and on the NTB port. Storage applications are typically memory bandwidth limited.</td>
</tr>
<tr>
<td></td>
<td>• Relaxed ordering</td>
<td></td>
</tr>
</tbody>
</table>
DDR4 Benefits

Lower Power
- Lower voltage (1.5v -> 1.2v) DIMMs
- Smaller page size (1024 -> 512) for x4 devices
- Initial results show savings of ~2W per DIMM at the wall.

Improved RAS
- Command/Address Parity error recovery

Higher bandwidth
- 14% higher STREAM results (DDR4-2133 vs. DDR3-1866)
- Increased DIMM frequency when multiple DIMMs per channel are installed
Cluster on Die (COD) Mode

- Supported on 1S & 2S SKUs with 2 Home Agents (10+ cores)
- In memory directory bits & directory cache used on 2S to reduce coherence traffic and cache-to-cache transfer latencies
- Targeted at NUMA optimized workloads where latency is more important than sharing across Caching Agents
  - Reduces average LLC hit and local memory latencies
  - HA sees most requests from reduced set of threads potentially offering higher effective memory bandwidth
- OS/VMM own NUMA and process affinity decisions
Intel® Turbo Boost Technology 2.0 and Intel® AVX*

- Intel® Turbo Boost Technology 2.0 automatically allows processor cores to run faster than the Rated and AVX base frequencies if they’re operating below power, current, and temperature specification limits.

- Amount of turbo frequency achieved depends on the type of workload, number of active cores, estimated current & power consumption, and processor temperature.

- Due to workload dependency, separate AVX base & turbo frequencies will be defined for Xeon® processors starting with E5 v3 product family.

[Diagram showing Frequency comparison between Previous Generations and E5 v3 & Future Generations]
How does frequency change with AVX workloads?

Core detects presence of AVX instructions
  - AVX instructions draw more current & higher voltage is needed to sustain operating conditions

Core signals to Power Control Unit (PCU) to provide additional voltage & core slows the execution of AVX instructions
  - Need to maintain TDP limits, so increasing voltage may cause frequency drop
  - Amount of frequency drop will depend on workload power & AVX frequency limits

PCU signals that the voltage has been adjusted & core returns to full execution throughput

PCU returns to regular (non-AVX) operating mode 1ms after AVX instructions are completed
Intel® Xeon® Processor E5-2600 v3 Product Family
High Performance Computing Performance

Intel® Xeon® E5-2699 v3 (18C, 2.3GHz) vs. Intel® Xeon® E5-2697 v2 (12C, 2.7GHz)

- Up to 1.24x Geomean
- Up to 1.25x Geomean
- Up to 1.32x Geomean
- Up to 1.39x Geomean
- Up to 1.72x Geomean

*Optimized for AVX. **Optimized for AVX2. Source as of June 2014: Intel internal measurements on platform with two E5-2697 v2, 8x8GB DDR3-1866, RHEL6.3. Platform with two E5-2699 v3, 8x8GB DDR4-2133, RHEL 6.3, NUMA-COD mode for all except Energy-ES mode. 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.
Intel® Many Integrated Core Architecture (Intel® MIC)
Intel® Xeon Phi™ Coprocessor
## Intel Architecture Multicore and Manycore

<table>
<thead>
<tr>
<th>Processor Family</th>
<th>Core(s)</th>
<th>Threads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intel Xeon 64-bit</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Intel Xeon 5100 series</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Intel Xeon 5500 series</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Intel Xeon 5600 series</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Intel Xeon E5 Product</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Family Ivy Bridge</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>Intel Xeon Haswell</td>
<td>18</td>
<td>36</td>
</tr>
<tr>
<td>Intel Xeon Phi code name</td>
<td>61</td>
<td>244</td>
</tr>
</tbody>
</table>

Intel® Xeon Phi™ Coprocessor extends established CPU architecture and programming concepts to highly parallel applications.

*Images do not reflect actual die sizes. Actual production die may differ from images.*
Each Intel® Xeon Phi™ Coprocessor core is a fully functional multi-thread execution unit

- Core unit based on Intel® Pentium® processor family
  - Two pipelines (U and V)
    - Dual-issue on scalar instructions
  - Scalar pipeline 1 clock latency
  - 64-bit data path

- 4 hardware threads per core
  - Thread context: GPRs, ST0-7, etc.
  - “Smart” round-robin scheduling
    - Prefetch buffers 2 inst-bundles / context
    - Next ready context selected in order
Each Intel® Xeon Phi™ Coprocessor core is a fully functional multi-thread execution unit

Instruction decoder is fully pipelined but is designed as a 2-cycle unit

- Enables significant increase to maximum core frequency, but...
  - Core cannot issue instructions from same context in adjacent cycles
  - Means minimum two threads per core to use all available compute cycles
Each Intel® Xeon Phi™ Coprocessor core is a fully functional multi-thread vector unit

Vector unit width 512 bits!

- 32 512-bit vector registers per context
  - Each holds 16 floats or 8 doubles
  - ALUs support int32/float32 operations, float64 arithmetic, int64 logic ops
  - Ternary ops including Fused-Multiply-Add
  - Broadcast/swizzle support, float16 up-convert
  - 8 vector mask registers for per lane conditional operations
  - Most ops: 4-cycle latency 1-cycle throughput
    - Matches 4-cycle round robin of integer unit
  - Mostly IEEE 754 2008 compliant
    - Not supported: MMX™ technology, Streaming SIMD Extensions (SSE), Intel® Advanced Vector Extensions (Intel® AVX)
Individual cores are tied together via fully coherent caches into a bidirectional ring on the Intel® Xeon Phi™ coprocessor.

**Bidirectional ring**
- 180 GB/sec
- Distributed Tag Directory (DTD) reduces ring snoop traffic
- Gen2x16 PCI Express* 64-256 byte packets peer-to-peer R/W

**GDDR5 Memory**
- 16 32-bit channels
- Up to 5.5 GT/sec
- 8 GB - 275 ns access latency

**L1** 32K I/D-cache per core
- 1 cycle access latency
- 3 cycle addr-gen interlock 1
- 8-way associativity
- 64-byte cache line
- ~38 concurrent access/core

**L2** 512K cache per core
- 11 cycle raw latency
- 8-way associativity
- 64-byte cache line
- Streaming HW prefetcher
- ~38 concurrent access/core
## Cache Hierarchy

<table>
<thead>
<tr>
<th>Parameter</th>
<th>L1</th>
<th>L2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coherence</td>
<td>MESI</td>
<td>MESI</td>
</tr>
<tr>
<td>Size</td>
<td>32KB + 32 KB</td>
<td>512 KB</td>
</tr>
<tr>
<td>Associativity</td>
<td>8-way</td>
<td>8-way</td>
</tr>
<tr>
<td>Line Size</td>
<td>64 Bytes</td>
<td>64 Bytes</td>
</tr>
<tr>
<td>Banks</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Access Time</td>
<td>2 cycle</td>
<td>11 cycle</td>
</tr>
<tr>
<td>Policy</td>
<td>Pseudo LRU</td>
<td>Pseudo LRU</td>
</tr>
<tr>
<td>Duty Cycle</td>
<td>1 per clock</td>
<td>1 per clock</td>
</tr>
<tr>
<td>Ports</td>
<td>Read or Write</td>
<td>Read or Write</td>
</tr>
</tbody>
</table>

There is no L3 cache!
## Intel® Xeon Phi™ Coprocessor Product Lineup

### 7 Family
**Highest Performance, Most Memory**
Performance leadership

<table>
<thead>
<tr>
<th>Model</th>
<th>GDDR5</th>
<th>Memory Rate</th>
<th>Performance</th>
<th>TDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>7120P</td>
<td>16GB</td>
<td>352GB/s</td>
<td>&gt;1.2TF DP</td>
<td>300W</td>
</tr>
<tr>
<td>7120X</td>
<td>8GB</td>
<td>&gt;300GB/s</td>
<td>&gt;1TF DP</td>
<td>225-245W TDP</td>
</tr>
<tr>
<td>7120A</td>
<td>6GB</td>
<td>240GB/s</td>
<td>&gt;1TF DP</td>
<td>300W</td>
</tr>
</tbody>
</table>

### 5 Family
**Optimized for High Density Environments**
Performance/watt leadership

<table>
<thead>
<tr>
<th>Model</th>
<th>GDDR5</th>
<th>Memory Rate</th>
<th>Performance</th>
<th>TDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>5110P</td>
<td>16GB</td>
<td>352GB/s</td>
<td>&gt;1.2TF DP</td>
<td>300W</td>
</tr>
<tr>
<td>5120D</td>
<td>8GB</td>
<td>&gt;300GB/s</td>
<td>&gt;1TF DP</td>
<td>225-245W TDP</td>
</tr>
</tbody>
</table>

### 3 Family
**Outstanding Parallel Computing Solution**
Performance/$ leadership

<table>
<thead>
<tr>
<th>Model</th>
<th>GDDR5</th>
<th>Memory Rate</th>
<th>Performance</th>
<th>TDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>3120P</td>
<td>6GB</td>
<td>240GB/s</td>
<td>&gt;1TF DP</td>
<td>300W</td>
</tr>
<tr>
<td>3120A</td>
<td></td>
<td></td>
<td></td>
<td></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 information go to [Intel Performance](http://www.intel.com/performance).
**Intel® Xeon Phi™ Coprocessor:**
*Increases Application Performance up to 7x*

<table>
<thead>
<tr>
<th>Segment</th>
<th>Application/Code</th>
<th>Performance vs. 2S Xeon*</th>
<th>Code Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCC</td>
<td>NEC / Video Transcoding Embree 2.0</td>
<td>Up to 3.0x² Up to 1.89x¹ (native)</td>
<td>• NEC Case Study</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Intel Developer Zone⁷</td>
</tr>
<tr>
<td>Energy</td>
<td>Seismic Imaging ISO3DFD Proxy 16th order Isotropic kernel <strong>RTM</strong> Seismic Imaging 3DFD TTI 3- Proxy 8th order <strong>RTM</strong> Petrobras Seismic ISO-3D <strong>RTM</strong> (2.4 Intel® Xeon Phi™ coprocessors)</td>
<td>Up to 1.45x³ Up to 1.23x² Up to 3.4x, 5.6x⁴</td>
<td>• Contact Intel representative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Proprietary code</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Proprietary code</td>
</tr>
<tr>
<td>Financial Services</td>
<td>BlackScholes SP / DP Monte Carlo European Option SP / DP Monte Carlo RNG European SP / DP Binomial Options SP / DP</td>
<td>SP: Up to 2.12x¹; DP Up to 1.72x³ SP: Up to 7x¹; DP Up to 3.13x³ SP: Up to 1.58x²; DP Up to 1.17x³ SP: Up to 1.85x¹; DP Up to 1.85x³</td>
<td>• Contact Intel representative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Contact Intel representative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Contact Intel representative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Contact Intel representative</td>
</tr>
<tr>
<td>Life Science</td>
<td>BWA/Bio-Informatics Wayne State University/mpi-Hmmer GROMACS / Molecular Dynamics</td>
<td>Up to 1.5x⁴ Up to 1.56x¹ (symmetric) Up to 1.36x¹ (symmetric)</td>
<td>• Contact Intel representative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Intel Developer Zone⁷</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Intel Developer Zone⁷</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>ANSYS / Mechanical SMP Sandia Mantevo / miniFE</td>
<td>Up to 1.88x⁵ Up to 2.3x⁴</td>
<td>• Intel Developer Zone⁷</td>
</tr>
<tr>
<td>Physics / Astronomy</td>
<td>ZIB (Zuse-Institut Berlin) / Ising 3D (Solid State Physics) ASKAP HogbomClean (astronomy) Princeton / GTC-P (Gyrokinetic Toroidal Turbulence Simulation IVB</td>
<td>Up to 3.46x¹ (symmetric) Up to 1.73x¹ Up to 1.18x⁶</td>
<td>• Contact Intel representative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Intel Developer Zone⁷</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Intel Developer Zone⁷</td>
</tr>
<tr>
<td>Weather</td>
<td>WRF /Code WRF V3.5</td>
<td>1.56x⁶</td>
<td>• Intel Developer Zone⁷</td>
</tr>
</tbody>
</table>

1. 2S Xeon E5 2670 vs. 2S Xeon* E5 2670 + 1 Xeon Phi® coprocessor
2. 2S Xeon E5 2670 vs. 2S Xeon E5 2670 + 2 Xeon Phi® coprocessor
3. 2S Xeon E5-2697v2 vs. 1 Xeon Phi coprocessor (Native Mode)
4. 2S Xeon E5-2697v2 vs. 2S Xeon E5 2697v2 + 1 Xeon Phi® coprocessor (Symmetric Mode) (for Petrobras, 1, 2, 3 or 4 Xeon Phi’s in the system)
5. 2S Xeon E5 2670 vs. 2S Xeon* E5 2670 + 1 Xeon Phi® coprocessor (Symmetric) (only 2 Xeon cores used to optimize licensing costs)
6. 4 nodes of 2S E5-2697v2 vs. 4 nodes of E5-2697v2 + 1 Xeon Phi® coprocessor (Symmetric)

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.

Source: Intel & Customer Measurement

---

* Enabling >60 codes for Intel® Xeon Phi™ coprocessors

---

* Xeon = Intel® Xeon® processor
* Xeon Phi = Intel® Xeon Phi™ coprocessor
## Intel® Xeon Phi™ Coprocessor: Public Code

<table>
<thead>
<tr>
<th>Segment</th>
<th>Application/Code</th>
<th>Performance¹ vs. 2S Xeon*</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCC</td>
<td>DreamWorks/Embree Open Source NEC / Video Transcoding</td>
<td>Up to 1.53x</td>
</tr>
<tr>
<td></td>
<td>Seismic Imaging ISO3DFD Proxy 16th order isotropic kernel <strong>RTM</strong></td>
<td>Up to 1.45X¹</td>
</tr>
<tr>
<td></td>
<td>Seismic Imaging 3DFD TTI 3-Proxy 8th order <strong>RTM</strong> (complex structures)</td>
<td>Up to 1.24 (24 threads)²</td>
</tr>
<tr>
<td></td>
<td>Petrobras Seismic ISO-3D <strong>RTM</strong> 2s XEON and 2 cards</td>
<td>Up to 1.97 (16 threads)²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Up to 3.4x²</td>
</tr>
<tr>
<td>Energy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finance</td>
<td>Black Scholes SP</td>
<td>Up to 2.21x¹</td>
</tr>
<tr>
<td></td>
<td>Monte Carlo EMEA Option SP</td>
<td>Up to 7.05x¹</td>
</tr>
<tr>
<td></td>
<td>Monte Carlo RNG SP</td>
<td>Up to 1.58x¹</td>
</tr>
<tr>
<td>Life Science</td>
<td>BWA/Bio-Informatics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wayne State University/MPHimer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GROMACS / Molecular Dynamics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TUM (Technische Universität München) / SG++ Astronomy</td>
<td></td>
</tr>
<tr>
<td>Manufacturing</td>
<td>ANSYS / Mechanical SMP</td>
<td>Up to 1.76x²</td>
</tr>
<tr>
<td></td>
<td>Sandia Mantevo / miniFE (Finite Element Solver)</td>
<td>Up to 2.3x²</td>
</tr>
<tr>
<td>Physics</td>
<td>ZIB (Zuse-Institut Berlin) / Ising 3D (Solid State Physics)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AWE (Atomic Weapons Estab/ReverFE) (2D Structured Hydrodynamics)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Princeton / GTC-P (Gyrokinetic Toroidal) Turbulence Simulation</td>
<td></td>
</tr>
<tr>
<td>Weather</td>
<td>WRF /Code wRF V3.5</td>
<td>1.56x²</td>
</tr>
</tbody>
</table>

### Notes:
1. 2S Xeon E5 2670 vs. 2S Xeon E5 2670 + 1 Xeon Phi* (preproduction HW/SW & Application running 100% on coprocessor unless otherwise noted)
2. 2S Xeon E5 2697v2 vs. 2S Xeon® E5 2697v2 + 1 Xeon Phi* (offload)
3. 2S Xeon E5 2697v2 vs 2S Xeon E5v2 = 2 Xeon Phi
4. 8 node cluster, each node with 2S Xeon* (comparison is cluster performance with and without 1 Xeon Phi* per node) (Hetero)
Intel® Xeon Phi™ Product Family
based on Intel® Many Integrated Core (MIC) Architecture

2013:
Intel® Xeon Phi™ Coprocessor x100 Product Family
“Knights Corner”
22 nm process
Up to 61 Cores
Up to 16GB Memory

2015:
Intel® Xeon Phi™ Coprocessor x200 Product Family
“Knights Landing”
14 nm
Processor & Coprocessor
Up to 72 cores
On Package, High-Bandwidth Memory

Future Knights:
Upcoming Gen of the Intel® MIC Architecture
In planning
Continued roadmap commitment

*Per Intel’s announced products or planning process for future products.
Next Intel® Xeon Phi™ Processor
Codename: Knights Landing

Designed using Intel’s cutting-edge 14nm process

Not bound by “offloading” bottlenecks
Standalone CPU or PCIe Coprocessor

Leadership compute & memory bandwidth
Integrated On-Package Memory
Next Intel® Xeon Phi™ Processor

Knights Landing

Compute: Energy-efficient IA cores
- Microarchitecture enhanced for HPC
- 3x Single Thread Performance vs Knights Corner
- Intel Xeon Processor Binary Compatible

On-Package Memory:
- up to 16GB at launch
- 5X Bandwidth vs DDR4
- 1/3x the Space
- 5X Power Efficiency

Jointly Developed with Micron Technology
Modernizing Community Codes... *Together*

<table>
<thead>
<tr>
<th>AMBER</th>
<th>WRF</th>
<th>VISIT</th>
<th>VASP</th>
<th>UTBENCH</th>
<th>SU2</th>
<th>SG++</th>
<th>SeisSol, GADGET, SG++</th>
<th>ROTOR SIM</th>
<th>R</th>
<th>Quantum Espresso</th>
<th>Optimized integral</th>
<th>OPENMP/mpi</th>
<th>Openflow</th>
<th>NWChem</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVBP (Large Eddy)</td>
<td>Blast</td>
<td>BUDE</td>
<td>CAM-5</td>
<td>CASTEP</td>
<td>Castep</td>
<td>CESM</td>
<td>CFSv2</td>
<td>CIRCAC</td>
<td>CliPhi (COSMOS)</td>
<td>COSA</td>
<td>Cosmos codes</td>
<td>DL-MESO</td>
<td>DL-Poly</td>
<td>ECHAM6</td>
</tr>
</tbody>
</table>

Intel® Parallel Computing Centers

Other brands and names are the property of their respective owners.
Programming for Intel Architecture
Highly Parallel Applications

Efficient vectorization, threading, and parallel execution drives higher performance for suitable scalable applications.

Theoretical acceleration of a highly parallel processor over a Intel® Xeon® parallel processor (<1: Intel® Xeon® faster) – For illustration only.
Parallel Programming for Intel® Architecture

- **NODES**: Use Intel® MPI, Co-Array Fortran
- **CORES**: Use threads directly (pthreads) or via OpenMP®, C++11
  Use tasking, Intel® TBB / Cilk™ Plus
- **VECTORS**: Intrinsics, auto-vectorization, vector-libraries
  Language extensions for vector programming (SIMD)
- **BLOCKING**: Use caches to hide memory latency
  Organize memory access for data reuse
- **DATA LAYOUT**: Structure of arrays facilitates vector loads / stores, unit stride
  Align data for vector accesses

Parallel programming to utilize the hardware resources, in an abstracted and portable way
Heterogeneous Programming

- Intel® MKL
- OpenMP®
- OpenCL
- Intel® TBB
- Intel® Cilk™ Plus
- C/C++
- Fortran
- Directive-based offloading & Virtual Shared Memory model
- Messaging libraries and internal infrastructure

Copyright © 2014, Intel Corporation. All rights reserved. *Other brands and names are the property of their respective owners.
Native Programming for Intel Xeon Phi

Parallelization
- Intel® Math Kernel Library
- OpenMP®
- Intel® Threading Building Blocks
- Intel® Cilk™ Plus
- POSIX threads*

Vectorization
- Intel® Math Kernel Library
- Auto vectorization
- Semi-auto vectorization: #pragma (vector, ivdep, simd)
- Intel® Cilk™ Plus Array Notations
- C/C++ Vector Classes (F32vec16, F64vec8)
- Intrinsics

Ease of use
Fine control
Flexible Execution Models for Heterogeneous Platforms

Compilers, Libraries, Runtime Systems

Source Code

Serial and Moderately Parallel Code

Highly Parallel Code

Multicore Only

Multicore Hosted with Manycore Offload

Symmetric

Manycore Only (Native)