НОВОЕ ПОКОЛЕНИЕ ПРОЦЕССОРОВ INTEL® XEON PHI™
Legal Notices and Disclaimers

- INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY RELATING TO SALE AND/OR USE OF INTEL PRODUCTS, INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT, OR OTHER INTELLECTUAL PROPERTY RIGHT. Intel products are not intended for use in medical, life-saving, life-sustaining, critical control or safety systems, or in nuclear facility applications.

- Intel products 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 may make changes to dates, specifications, product descriptions, and plans referenced in this document at any time, without notice.

- This document may contain information on products in the design phase of development. The information herein is subject to change without notice. Do not finalize a design with this information.

- Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them.

- Intel Corporation or its subsidiaries in the United States and other countries may have patents or pending patent applications, trademarks, copyrights, or other intellectual property rights that relate to the presented subject matter. The furnishing of documents and other materials and information does not provide any license, express or implied, by estoppel or otherwise, to any such patents, trademarks, copyrights, or other intellectual property rights.

- Wireless connectivity and some features may require you to purchase additional software, services or external hardware.

- Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, visit Intel Performance Benchmark Limitations.

- 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 © 2016 Intel Corporation. All rights reserved.
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© 2016, Intel Corporation. All rights reserved. Intel, the Intel logo, Atom, Xeon, Xeon Phi, Core, VTune, and Cilk are trademarks of Intel Corporation in the U.S. and other countries.

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
Agenda

Xeon Phi code named Knights Landing

- Introduction
- KNL Architecture
- Cluster Modes
- Memory Modes

Getting Performance – AVX512
WHAT IS KNIGHTS LANDING
A Paradigm Shift for Highly-Parallel Server Processor and Integration are Keys to Future

Knights Landing (KNL)

- **Memory Bandwidth**: >400 GB/s STREAM
- **Memory Capacity**: Over 25x* KNC
- **Power Efficiency**: Over 25% better than card¹
- **Cost**: Less costly than discrete parts¹
- **Flexibility**: Limitless configurations
- **Density**: 3+ KNL with fabric in 1U³

*Comparison to 1st Generation Intel® Xeon Phi™ 7120P Coprocessor (formerly codenamed Knights Corner)

¹Results based on internal Intel analysis using estimated power consumption and projected component pricing in the 2015 timeframe. This analysis is provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance.

²Comparison to a discrete Knights Landing processor and discrete fabric component.

³Theoretical density for air-cooled system; other cooling solutions and configurations will enable lower or higher density.
Intel® Xeon Phi™ Product Family
Highly-parallel processing to power your breakthrough innovations

Available Today
Knights Corner
Intel® Xeon Phi™
x100 Product Family
- 22 nm process
- Coprocessor only
- >1 TF DP Peak
- Up to 61 Cores
- Up to 16GB GDDR5

Available Today
Knights Landing
Intel® Xeon Phi™
x200 Product Family
- 14 nm process
- Host Processor & Coprocessor
- >3 TF DP Peak\(^1\)
- Up to 72 Cores
- Up to 16GB HBM
- Up to 384GB DDR4\(^2\)
- >400 GB/s STREAM
- Integrated Fabric\(^2\)

Future
Knights Hill
3\(^{rd}\) generation
- 10 nm process
- Integrated Fabric (2\(^{nd}\) Generation)
- In Planning…

\(^*\)Results will vary. This simplified test is the result of the distillation of the more in-depth programming guide found here: https://software.intel.com/sites/default/files/article/383967/sxeon-phi-right-for-me.pdf
All products, computer systems, dates and figures specified are preliminary based on current expectations, and are subject to change without notice.

\(^1\) Over 3 Teraflops of peak theoretical double-precision performance is preliminary and based on current expectations of cores, clock frequency and floating point operations per cycle. FLOPS = cores x clock frequency x floating point operations per second per cycle.

\(^2\) Host processor only
Knights Landing (Host or PCIe)

**Host Processor**
- Knights Landing

**Host Processor w/ integrated Fabric**
- Knights Landing

**Groveport Platform**

**Knights Landing Processors**
- Host Processor for Groveport Platform
- Solution for future clusters with both Xeon and Xeon Phi

**Knights Landing PCIe Coprocessors**
- Ingredient of Grantley & Purley Platforms
- Solution for general purpose servers and workstations
KNL w/Intel Omni-Path

- Omni-Path Fabric integrated on package
- First product with integrated fabric
- Connected to KNL die via 2x16 PCIe* ports
  - Output: 2 Omni-Path ports
    - 100 Gb/s/port

- Benefits:
  - Low cost, latency and power
  - High density and bandwidth
  - High scalability

* On package connect with PCIe semantics, with MCP optimisations for physical layer
KNIGHTS LANDING ARCHITECTURE
Knights Landing

- > 8 billion transistors
- 14nm process
- Up to 72 cores
Knights Landing: Architecture

Diagram is for conceptual purposes only and only illustrates a CPU and memory – it is not to scale and does not include all functional areas of the CPU, nor does it represent actual component layout.

- Over 3 TF DP peak
  - Full Xeon ISA compatibility through AVX-512
  - ~3x single-thread vs. compared to Knights Corner

- Up to 72 cores (36 tiles)
  - 2D mesh architecture

- 6 channels DDR4
  - Up to 384GB
  - ~90 GB/s

- Common with Grantley PCH

- On-package 2 ports OPA
  - Integrated Fabric

- Up to 16GB high-bandwidth on-package memory (MCDRAM)
  - Exposed as NUMA node
  - >400 GB/s sustained BW

- 2x 512b VPU per core
  - (Vector Processing Units)

- Based on Intel® Atom Silvermont processor with many HPC enhancements
  - Deep out-of-order buffers
  - Gather/scatter in hardware
  - Improved branch prediction
  - 4 threads/core
  - High cache bandwidth
  - & more

- Wellsburg PCH

- PCIe Gen3 x36

- DMI

- HFI

- 2 VPU HUB 2 VPU

- Core 1MB L2 Core
Mesh of Rings
- Every row and column is a (half) ring
- YX routing: Transmit in Y -> Turn -> Transmit in X
- Messages arbitrate at injection and on turn

Cache Coherent Interconnect
- Distributed directory to maintain cache coherency
  - CHA: caching/home agent keeps L2s coherent
  - Address hashes used to service L2 misses
  - MESIF protocol (F = Forward)
Microarchitecture Details

- Core based on Intel® Atom™ Silvermont (with many HPC enhancements)
  - 4 threads / core
  - Threading: back-to-back fetch and issue per thread
  - Core resources dynamically repartitioned (shared) between threads at thread selection points
  - 2x out-of-order buffer depth\(^3\)
  - Gather/scatter in hardware
  - Advanced branch prediction
  - VPU:
    - 32SP and 16DP
    - x87, SSE and AVX support
  - Address bits: 46/48 Physical/Virtual

\(^3\)Compared to the Intel® Atom™ core (based on Silvermont microarchitecture)
KNL ISA

- First processor that supports AVX-512
  - Binary compatible with Intel® Xeon® Processor:\(^1\):
    - Prior Intel® Xeon® processor binaries will run on KNL without recompilation
    - KNC Code will need recompilation to run on KNL
  - Yes: x87, MMX, SSE, AVX1 and AVX2.
  - No: TSX instructions. In HSW, under separate CPUID bit
- KNL Adds:
  - AVX512F: 512b vector extensions with mask support.
  - AVX512PFI: New Prefetch Instructions
  - AVX512CDI: Conflict Detection Instructions: To enable more vectorizing
  - AVX512ERI: New Exponential and Reciprocal Instructions

\(^1\)Binary compatible with Intel® Xeon® Processors v3 (Haswell) with the exception of Intel® TSX (Transactional Synchronization Extensions).
Microarchitecture Details

- Cache:
  - KNL provides high cache bandwidth
  - Icache: 32KB 8-way
  - Dcache: 32KB 8-way, 2x64B load ports, 1 store port
  - 2x B/W between Dcache and L2\(^3\)
  - Faster unaligned cache-line access support

\(^3\)Compared to the Intel® Atom™ core (based on Silvermont microarchitecture)
Microarchitecture Details

- Processor Buffers:
  - 2-wide decode/rename/retire
  - 72 inflight uops/core out-of-order buffers
  - Up to 6-wide at execution
  - INT (2x12) and FP (2x20) RS OoO
  - MEM RS (1x12) in-order with OoO completion
  - Recycle buffer holds memory ops waiting for completion
  - INT and MEM RS hold source data while FP RS does not
Microarchitecture Details

- TLB (Translation Lookaside Buffer):
  - 1st level uTLB (64 entries for 4K pages)
  - 2nd level dTLB (256 entries for 4K pages, 128 for 2MB, 16 for 1G)
- DMI (Direct Memory Interface): 4 lanes for chipset
- NTB: non-transparent bridge to create PCIe coprocessor (processor commonality)
- Performance monitoring reference manual for device driver developers (link)
CLUSTER MODES
Cluster Modes

- KNL has three on-die cluster modes:
  - All-to-All
  - Quadrant
  - Sub-NUMA Clustering (SNC)
All-2-All

- Address uniformly hashed across all distributed directories
  - No affinity between Tile, Directory and Memory
  - Most general mode.
    - Lower performance than other modes
    - “Mode of last resort”
- Typical Read L2 miss:
  1. L2 miss encountered
  2. Send request to distributed directory
  3. Miss in the directory. Forward to memory
  4. Memory sends the data to the requestor
- Chip divided into four virtual Quadrants
  - Address hashed to a Directory in the same quadrant as the Memory
    - Equally likely to go to any CHA in quadrant
  - Then uses round-robin across the memory channels
  - Affinity between the Directory and Memory
  - Lower latency and higher BW than All-to-All
  - Less traffic crossing quadrant boundaries
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 nodes
  - SW needs to be NUMA optimised to get the benefits
    - Running one MPI rank per NUMA region will ensure locality-of-access, and may improve bandwidth.
MEMORY MODES
3 Memory Modes

- Mode selected at boot
- Cache mode: MCDRAM covers all DDR
- Flat mode: MCDRAM is explicitly allocatable
Flat MCDRAM: SW Architecture

- Memory allocated in DDR by default
  - Keeps low bandwidth data out of MCDRAM.
- Apps explicitly allocate important data in MCDRAM
  - “Fast Malloc” functions: Built using NUMA allocation functions
  - “Fast Memory” Compiler Annotation: For use in Fortran.

MCDRAM exposed as a separate NUMA node

Intel® Xeon® with 2 NUMA nodes

≈

KNL with 2 NUMA nodes

Flat MCDRAM using existing NUMA support in Legacy OS
MEMORY MODES

- MCDRAM as Cache
  - Upside:
    - No software modifications required.
    - Bandwidth benefit.
  - Downside:
    - Latency hit to DDR.
    - Limited sustained bandwidth.
    - All memory is transferred DDR -> MCDRAM -> L2.
    - Less addressable memory.

- Flat Mode
  - Upside:
    - Maximum bandwidth and latency performance.
    - Maximum addressable memory.
    - Isolate MCDRAM for HPC application use only.
  - Downside:
    - Software modifications required to use DDR and MCDRAM in the same application.
    - Which data structures should go where?
    - MCDRAM is a limited resource and tracking it adds complexity.
SUMMARY
Summary

- Knights Landing is a high-throughput successor to Knights Corner:
  - Socketed, bootable processor with access to large amounts of RAM
  - General purpose, support for standard SW stack
  - Greatly improved single-thread performance vs KNC
  - Very high bandwidth, flexible MCDRAM
  - Optional on-chip interconnect (Omni Path)
  - High performance, versatile AVX512 instructions