BAYNCORE

Intel technology platform for HPC

Intel® Xeon® and Intel® Xeon Phi™ processor update

Manel Fernández

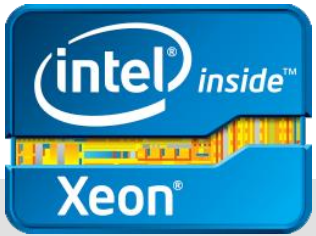
Intel HPC Software Workshop Series 2016

HPC Code Modernization for Intel® Xeon and Xeon Phi™

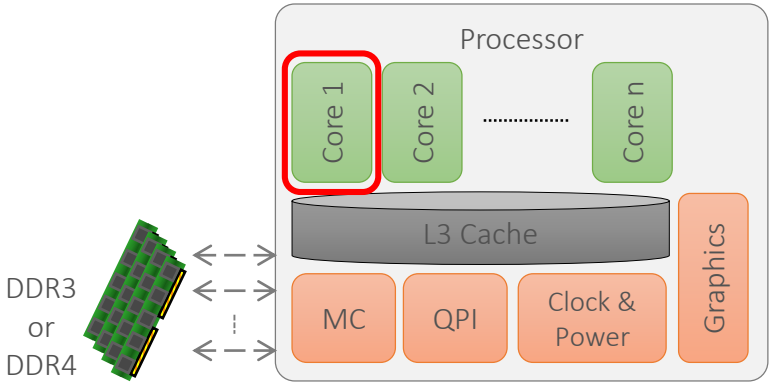
February 17th 2016, Barcelona

Today's Intel solutions for HPC

The multi- and many-core era



Multi-core	Many integrated core (MIC)
C/C++/Fortran, OMP/MPI/Cilk+/TBB	C/C++/Fortran, OMP/MPI/Cilk+/TBB
Bootable, native execution model	PCIe coprocessor, native and offload execution models
Up to 18 cores, 3 GHz, 36 threads	Up to 61 cores, 1.2 GHz, 244 threads
Up to 768 GB, 68 GB/s, 432 GFLOP/s DP	Up to 16 GB, 352 GB/s, 1.2 TFLOP/s DP
256-bit SIMD, FMA, gather (AVX2)	512-bit SIMD, FMA, gather/scatter, EMU (IMCI)
Targeted at general purpose applications Single thread performance (ILP) Memory capacity	**Targeted at highly parallel applications** High parallelism (DLP, TLP) High memory bandwidth



Intel® Core™ architecture

Intel® Core™ architecture roadmap

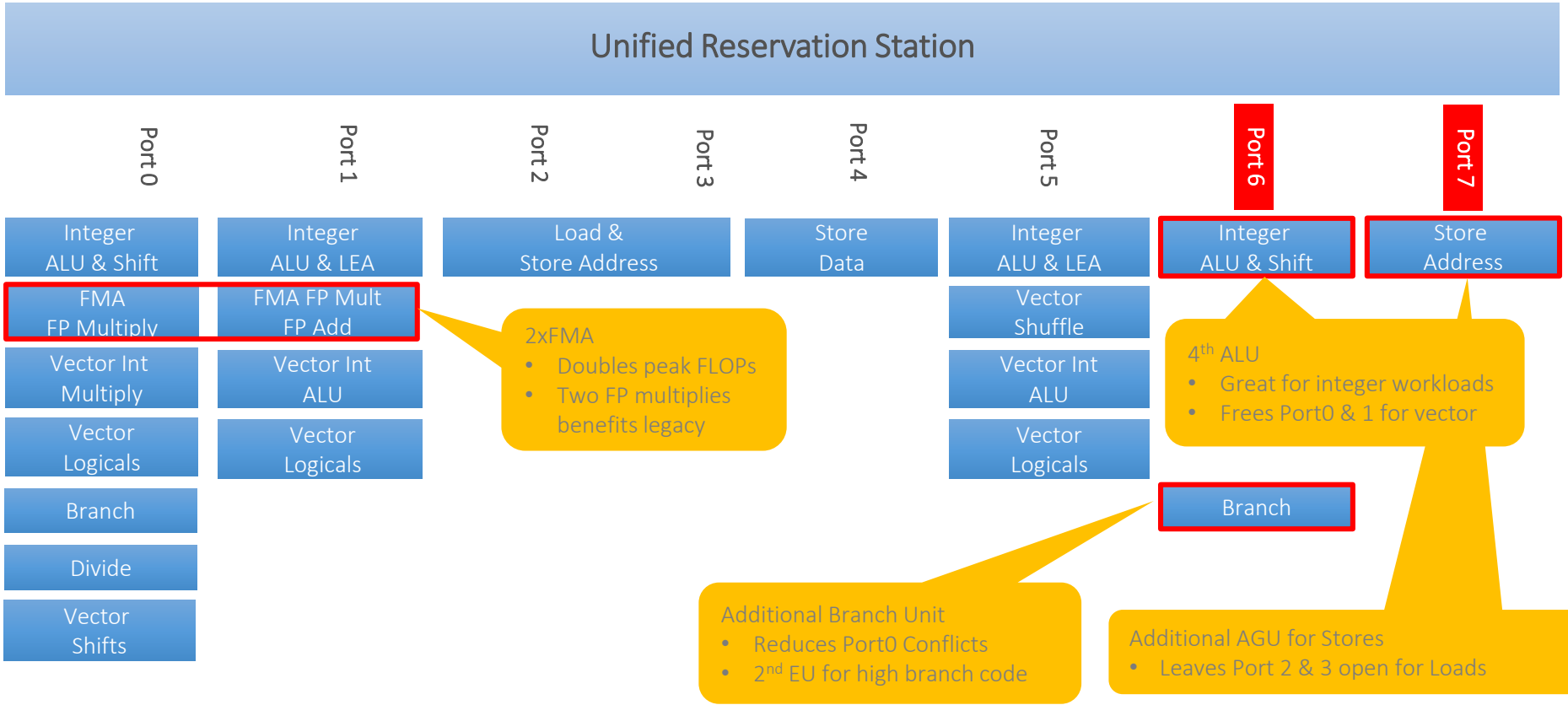
The "Tick-Tock" roadmap model

Intel® Core™ "Nehalem" microarchitecture		2nd generation	3rd generation	4th generation	5th generation	6th generation
Nehalem	Westmere	Sandy Bridge	Ivy Bridge	Haswell	Broadwell	Skylake
45nm	32nm		22nm		14nm	
2008	2009	2011	2012	2013	Sep 2014	2015
SSE4.2	AES	AVX	RNRAND, etc.	AVX2	ADX and 3 other new instructions	MPX, SGX AVX-512 (Xeon only)

Tick (shrink, new process technology)

Tock (innovate, new microarchitecture)

Haswell execution unit overview



Microarchitecture buffer sizes

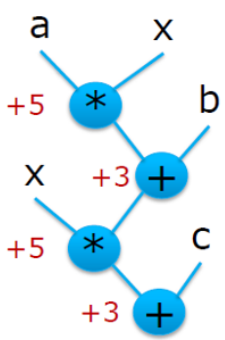
Extract more parallelism on every generation

	Nehalem	Sandy Bridge	Haswell	Skylake
Out-of-order Window	128	**168**	**192**	**224**
In-flight Loads	48	**64**	**72**	72
In-flight Stores	32	**36**	**42**	**56**
Scheduler Entries	36	**54**	**60**	**97**
Integer Register File	N/A	**160**	**168**	**180**
FP Register File	N/A	**144**	**168**	168
Allocation Queue	28/thread	28/thread	**56**	**64/thread**

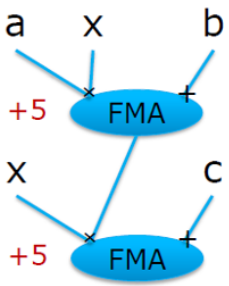
Fused Multiply and Add (FMA) instruction

Example: polynomial evaluation

$$ax^2 + bx + c$$
$$=$$
$$x(ax + b) + c$$



16 cycle latency
2 cycle throughput



10 cycle latency
1 cycle throughput

Micro-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 (lower is better)
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)

Improves accuracy and performance for commonly used class of algorithms

Broadwell: 5th generation Intel® Core™ architecture

Microarchitecture changes

FP instructions performance improvements

- Decreased latency and increased throughput for most divider (radix-1024) uops
- Pseudo-double bandwidth for scalar divider uops
- Vector multiply latency decrease (from 5 to 3 cycles)

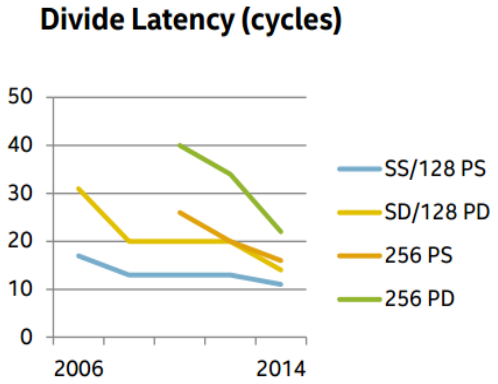
STLB improvements

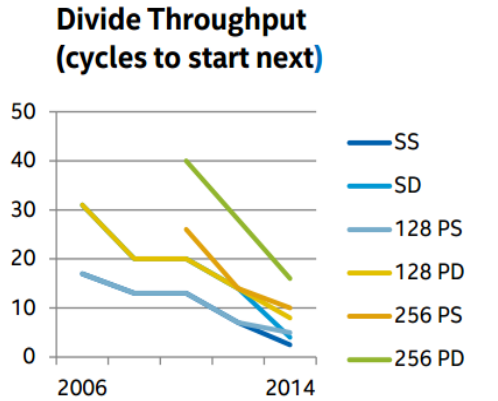
- Native, 16-entry 1G STLB array
- Increased size of STLB (from 1kB to 1.5kB)

Enabled two simultaneous page miss walks

Other ISA performance improvements

- ADC, CMOV – 1 uop flow
- PCLMULQDQ – 2 uop/7 cycles to 1 uop/5 cycles
- VCVTPS2PH (mem form) – 4 uops to 3 uops





Skylake: 6th generation Intel® Core™ architecture

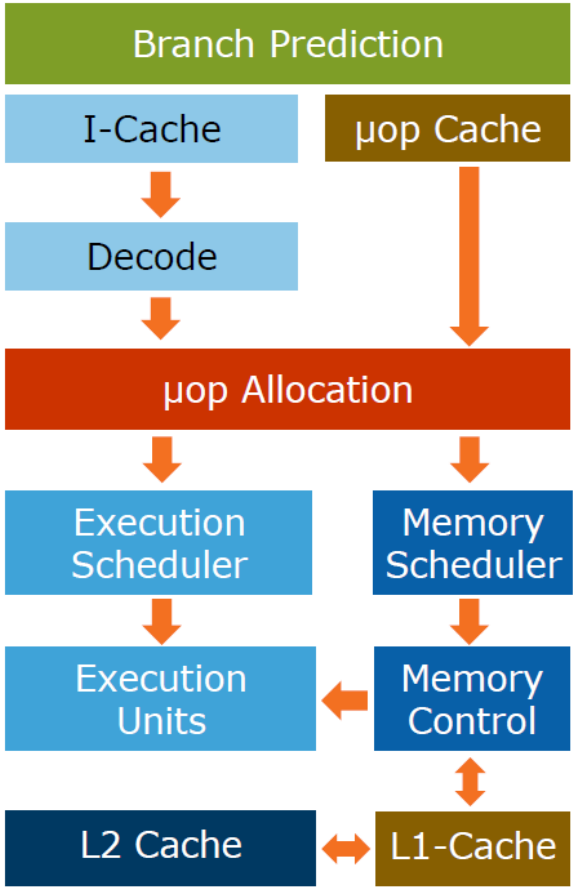
Dedicated server and client IP configurations

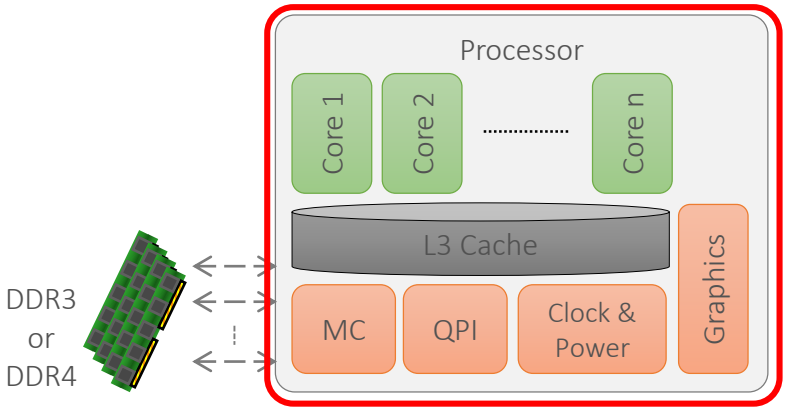
Improved microarchitecture

- Higher capacity front-end (up to 6 instr/cycle)
- Improved branch predictor
- Deeper Out-of-Order buffers
- More execution units, shorter latencies
- Deeper store, fill, and write-back buffers
- Smarter prefetchers
- Improved page miss handling
- Better L2 cache miss bandwidth
- Improved Hyper-Threading
- Performance/watt enhancements

New instructions supported

- Software Guard Extensions (SGX)
- Memory Protection Extensions (MPX)
- AVX-512 (Xeon versions only)

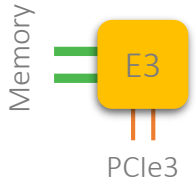




Intel® Xeon® processor architecture

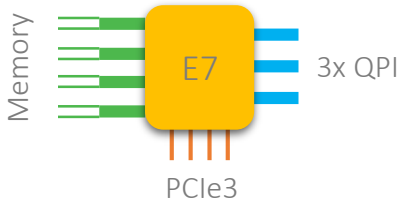
Intel® Xeon® processors and platforms

Intel® Xeon® E3



Intel® Xeon® E5

Intel® Xeon® E7



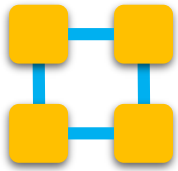
Intel® Xeon®
E5-1xxx
E3-1xxx

CPU/Socket

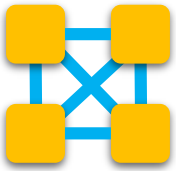
Intel® Xeon®
E5-2xxx

QPI

Intel® Xeon®
E5-4xxx



Intel® Xeon®
E7-4xxx

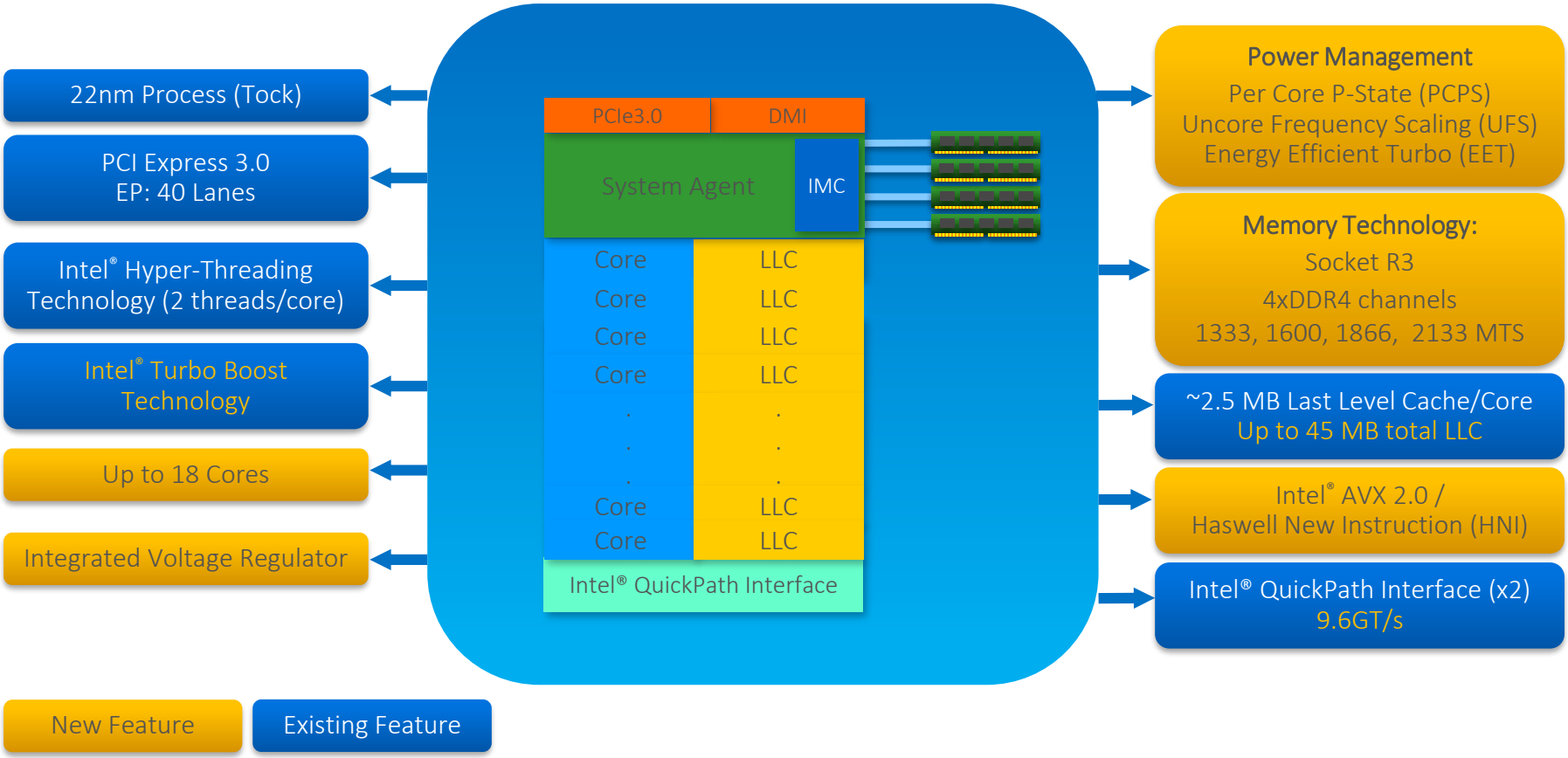


Intel® Xeon®
E7-xxxx

>4S

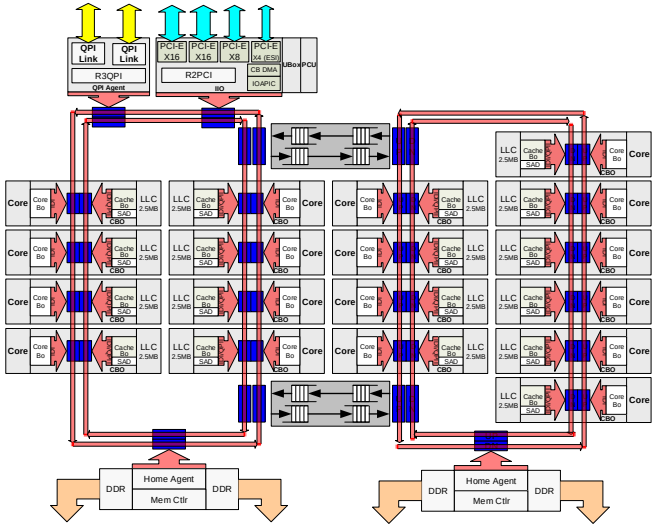
E5-2600 v3 (Haswell EP)
E5-2600 v4 (Broadwell EP)

Intel® Xeon® E5-2600 v3 processor overview

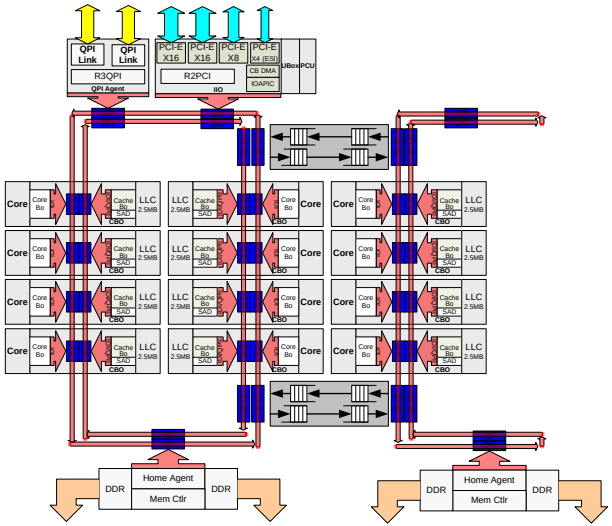


Haswell EP die configurations

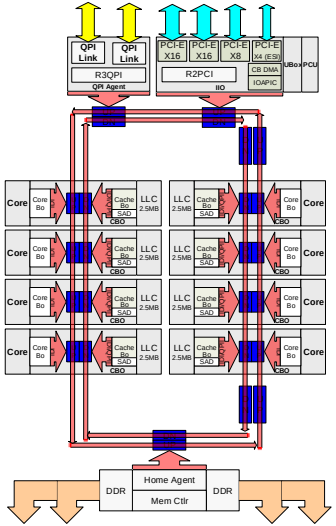
14-18 Core (HCC)



10-12 Core (MCC)



4-8 Core (LCC)



Chop	Columns	Home Agents	Cores	Power (W)	Transitors (B)	Die Area (mm²)
HCC	4	2	14-18	110-145	5.69	662
MCC	3	2	6-12	65-160	3.84	492
LCC	2	1	4-8	55-140	2.60	354

Cluster on die (COD) mode

Supported on 1S/2S HSW-EP SKUs with 2 Home Agents (10+ cores)

In memory directory bits and directory cache used on 2S to reduce coherence traffic and cache-to-cache transfer latencies

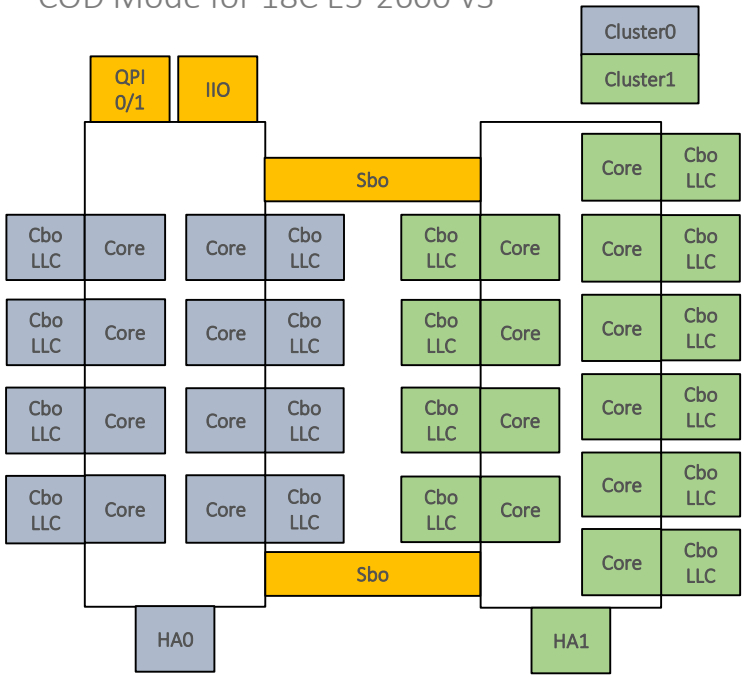
Targeted at NUMA optimized workloads where latency is more important than sharing across Caching Agents (Cbo)

- Reduces average LLC hit and local memory latencies
- HA sees most requests from reduced set of threads which can lead to higher memory bandwidth

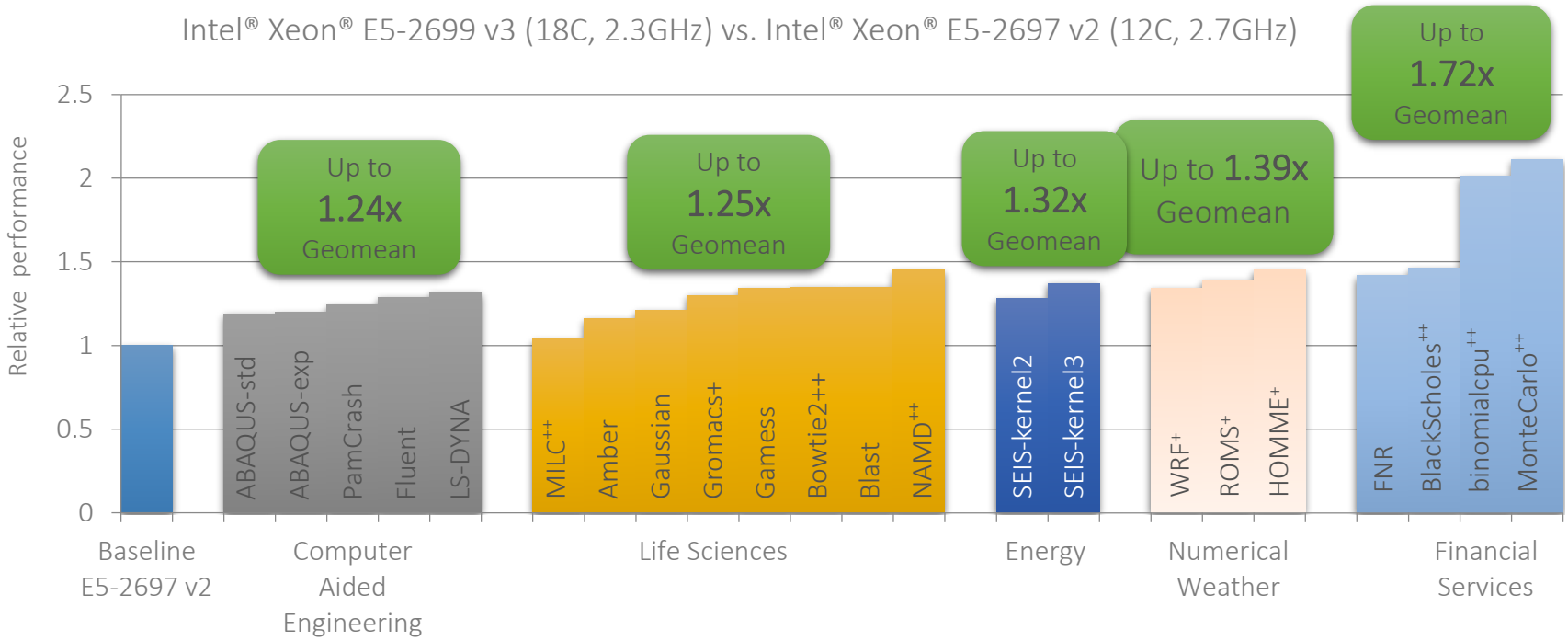
OS/VMM own NUMA and process affinity decisions

- Exported as 2 NUMA nodes

COD Mode for 18C E5-2600 v3



Intel® Xeon® E5-2600 v3 product family



More information at http://www.intel.com/performance

Intel® Xeon® E5-2600 v4 processor overview

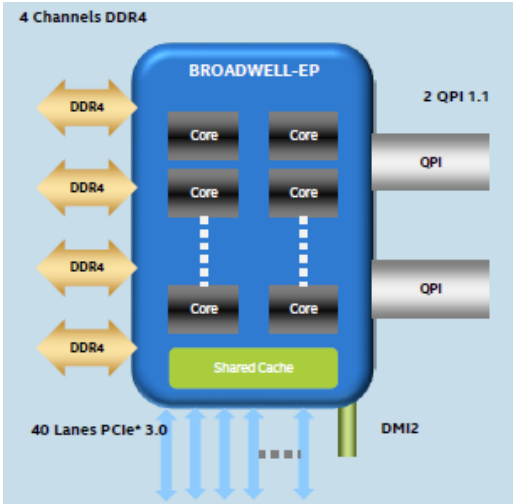
New features

Broadwell uarch, 14nm
Optimize cloud/datacentre performance
Socket compatible on Grantley platform

New processor technologies

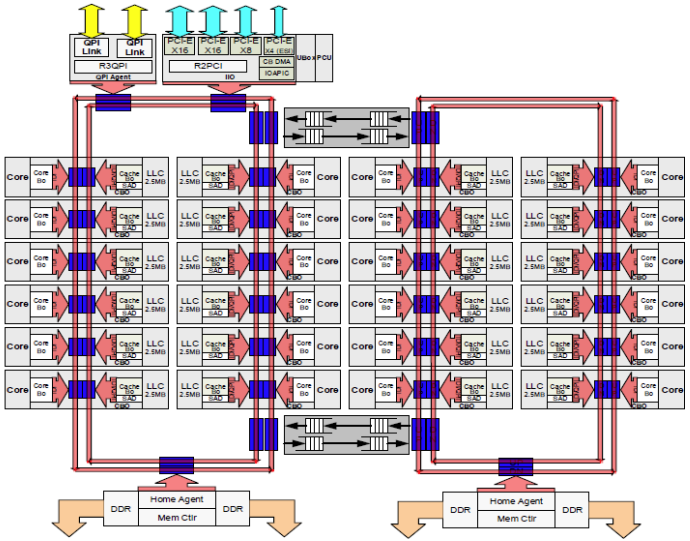
Memory BW monitoring
Improved virtualization and security
TSX (HLE/RTM) enabled

Features	Xeon E5-2600 v3 (Haswell-EP)	Xeon E5-2600 v4 (Broadwell EP)
Cores per socket	Up to 18	Up to 22
Threads per socket	Up to 36 threads	Up to 44 threads
Last-level cache (LLC)	Up to 45 MB	Up to 55 MB
QPI speed (GT/s)	2x QPI 1.1 channels 6.4, 8.0, 9.6 GT/s	
PCIe lanes/Controllers/Speed (GT/s)	40/10/PCIe 3.0 (2.5, 5, 8 GT/s)	
Memory population	4 channels of up to 3 RDIMMs of 3 LRDIMMs	+3DS LRDIMM
Max memory speed	Up to 2133	Up to 2400
TDP (W)	160 (Workstation only), 145, 135, 120, 105, 90, 85, 65, 55	

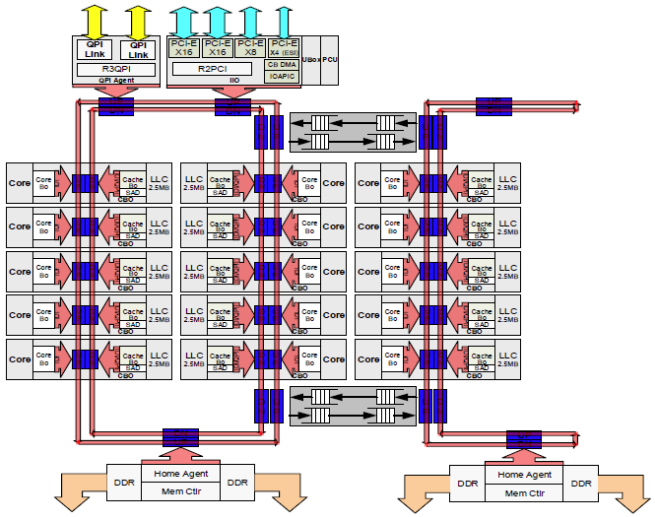


Broadwell EP die configurations

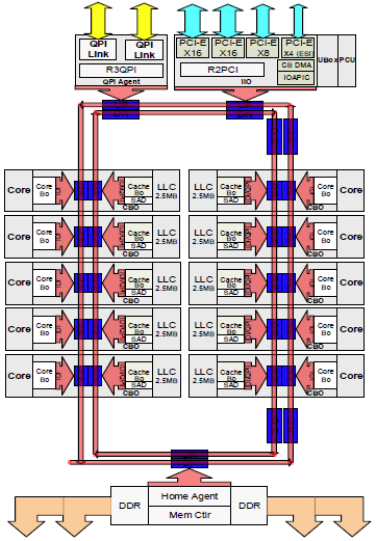
Up to 22 Core (HCC)



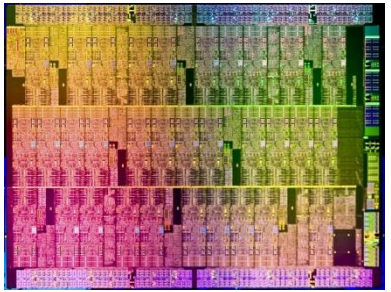
Up to 14 Core (MCC)



Up to 8 Core (LCC)



Chop	Columns	Home Agents	Cores	Power (W)
HCC	4	2	12-22	105-145
MCC	3	2	8-14	85-120
LCC	2	1	6-8	85



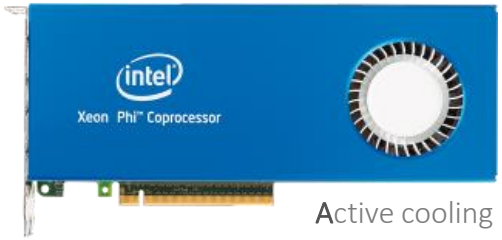
Intel® Xeon Phi™ (co)processor architecture

Intel® Many Integrated Core architecture (Intel® MIC)

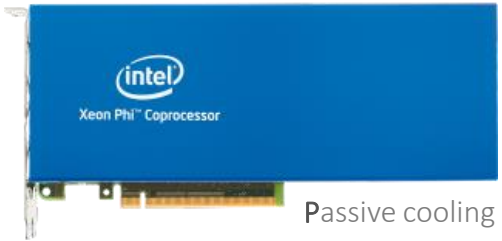
Intel® Xeon Phi™ architecture family

Intel® Xeon Phi™ coprocessor x100 product family **"Knights Corner"**	Intel® Xeon Phi™ coprocessor x200 product family **"Knights Landing"**	Upcoming generation of the Intel® MIC architecture **"Knights Hill"**
2013	2H'2015	2017?
22 nm process	14 nm process	10 nm process
1 TeraFLOP DP peak	3+ TeraFLOP DP peak	?
57-61 cores In-order core architecture 1 Vector Unit per core	72 cores (36 tiles) Out-of-order architecture based on Intel® Atom™ core 2 Vector Units per core Up to 3x single thread performance w.r.t. Knights Corner	?
6-16 GB GDDR5 memory	On package, 8-16 GB high bandwidth memory (HBM) with flexible models: cache, flat, hybrid Up to 768 GB DDR4 main memory	?
Intel® Initial Many Core Instructions (IMIC)	Intel® Advanced Vector Extensions (AVX-512) Binary compatible with AVX2	?
PCIe coprocessor	Stand alone processor and PCIe coprocessor versions	?
Intel® True Scale fabric	Intel® Omni-Path™ fabric (integrated in some models)	2nd generation Intel® Omni-Path™ fabric

Intel® Xeon Phi™ coprocessor product lineup



Active cooling



Passive cooling

Dense form factor

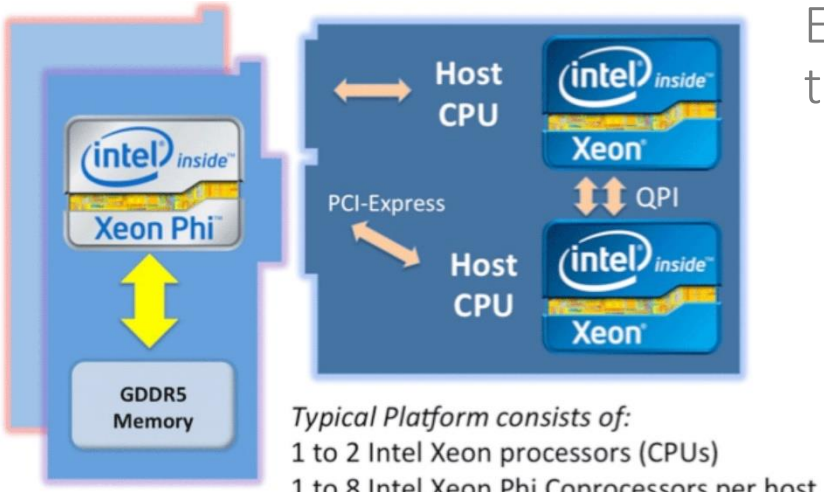
No thermal solution (X)

Family	Specifications	Product name
7 Family Highest performance, more memory Performance leadership	61 cores 16GB GDDR5 352 GB/s > 1.2TF DP 270-300W TDP	7120P (Q2'13) 7120X (Q2'13) 7120D (Q1'14) 7120A (Q2'14)
5 Family Optimized for high density environments Performance/watt leadership	60 cores 8GB GDDR5 320-352 GB/s > 1TF DP 225-245W TDP	5110P (Q4'12)+ 5120D (Q2'13)+
3 Family Outstanding parallel computing solution Performance/$ leadership	57 cores 6-8GB GDDR5 240-320 GB/s > 1TF DP 270-300W TDP	3120A (Q2'13) 3120P (Q2'13) 31S1P (Q2'13)*

(*) Special promotion under $200

(+) Special offer with a free 12-month trial of Intel® Parallel Studio XE Cluster Edition

Intel® Xeon Phi™ platform architecture



Typical Platform consists of:
1 to 2 Intel Xeon processors (CPUs)
1 to 8 Intel Xeon Phi Coprocessors per host

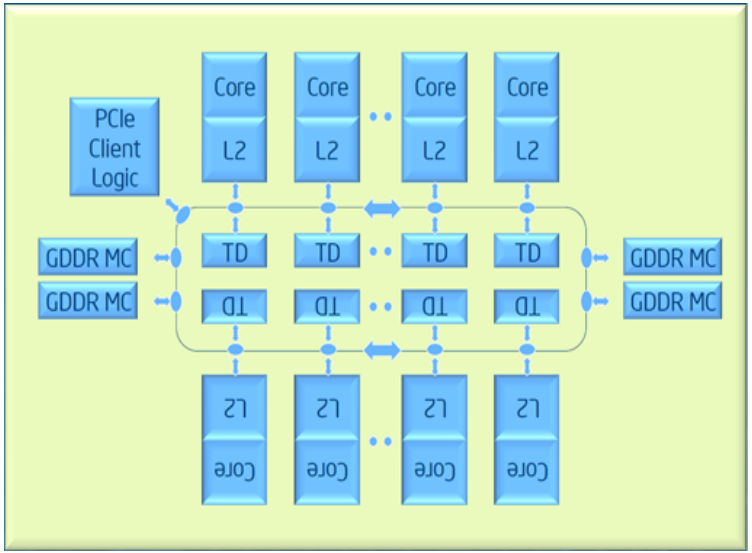
Each coprocessor connected to one host through PCIe bus

- PCIe Gen 2 (client) x16
  - Between 6-14 GB/s (relatively slow)
- Up to 8 coprocessors per host
- Inter-node coprocessors communication through Ethernet or InfiniBand
  - InfiniBand allows PCIe peer-to-peer interconnect without host intervention

Each coprocessor can be accessed as a network node

- It has its own IP address
- Runs a special uLinux OS (BusyBox)
  - Intel® Many Core Software Stack (MPSS)

Intel® Xeon Phi™ uncore architecture



(c) 2013 Jim Jeffers and James Reinders, used with permission.

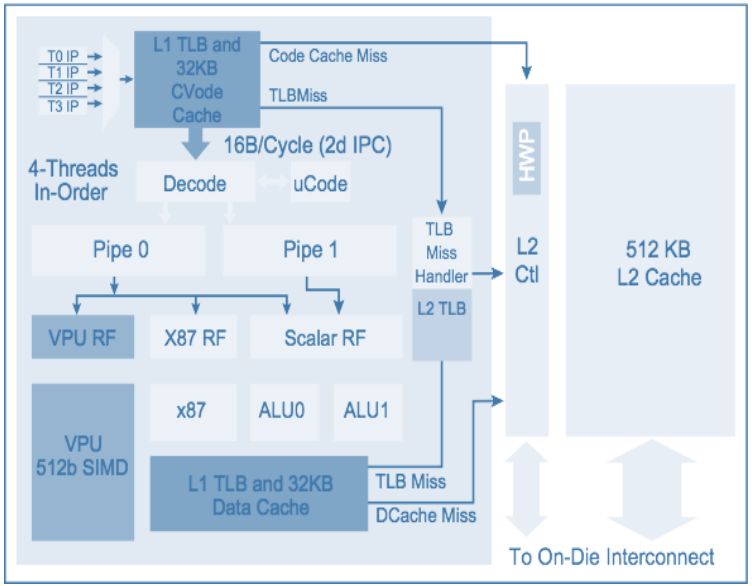
High bandwidth interconnect
- Bidirectional ring topology

Fully cache-coherent SMP on-a-chip
- Distributed global tag directory (TD)
- About 31 MB of “L2 cloud”
  - >100-cycle latency for remote L2 access

8-16 GB GDDR5 main memory (ECC)
- 8 memory controllers (MC)
  - >300-cycle latency access
- 2 GDDR5 32-bit channels per MC
- Up to 5.5 GT/s per channel
- 352 GB/s max. theoretic bandwidth
  - Practical peak about 150-180 GB/s

ECC on GDDR5/L2 for reliability

Intel® Xeon Phi™ core architecture



(c) 2013 Jim Jeffers and James Reinders, used with permission.

Based on Intel® Pentium® family

- 2-wide decode/issue pipeline
- In-order execution

4 hardware threads per core

512-bit SIMD vector unit

- Intel® Initial Many Core Instructions (IMCI)

64-bit datapath / 64-bit addressing

Per core L1/L2 caches

- 3-cycle/15-cycle latency access
- IL1 32KB, DL1 32KB, UL2 512 KB
- L2 HW streaming prefetcher

BAYNCORE

Significant performance improvement with Intel® Xeon Phi™ performance improvement

- Xeon = Intel® Xeon® processor
- Xeon Phi = Intel® Xeon Phi™ coprocessor

Segment	Application/Code	Performance vs. 2S Xeon*
DCC	NEC/Video Transcoding case study	Up to 3.0x[2]
Energy	Seismic Imaging ISO3DFD Proxy 16th order Isotropic kernel RTM Seismic Imaging 3DFD TTI 3- Proxy 8th order RTM (complex structures) Petrobras Seismic ISO-3D RTM (with 1,2,3 or 4 Intel® Xeon Phi™ coprocessors)	Up to 1.45x[3] Up to 1.23x[3] Up to 2.2x, 3.4x, 4.6x or 5.6x[4]
Financial Services	BlackScholes SP/DP Monte Carlo European Option SP/DP Monte Carlo RNG European SP/DP Binomial Options SP/DP	SP: Up to 2.12x[3] ; DP Up to 1.72x[3] SP: Up to 7x[3] ; DP Up to 3.13x[3] SP: Up to 1.58x[3] ; DP Up to 1.17x[3] SP: Up to 1.85x[3] ; DP Up to 1.85x[3]
Life Science	BWA/Bio-Informatics Wayne State University/MPI-Hmmer GROMACS/Molecular Dynamics	Up to 1.5x[4] Up to 1.56x[1] Up to 1.36x[1]
Manufacturing	ANSYS/Mechanical SMP Sandia Mantevo/miniFE case study	Up to 1.88x[5] Up to 2.3x[4]
Physics	ZIB (Zuse-Institut Berlin)/Ising 3D (Solid State Physics) ASKAP tHogbomClean (astronomy) Princeton/GTC-P (Gyrokinetic Torodial) Turbulence Simulation IVB	Up to 3.46x[1] Up to 1.73x[3] Up to 1.18x[6]
Weather	WRF /Code WRF V3.5	1.56x[6]

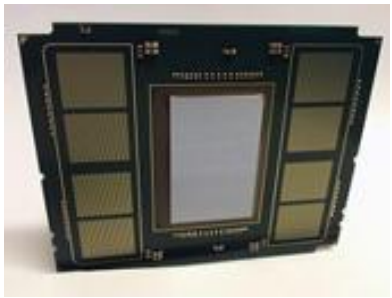
More information at Intel® Xeon Phi™ Applications and Solutions Catalog

1. 2S Xeon E5 2670 vs. 2S Xeon* E5 2670 + 1 Xeon Phi* coprocessor (Symmetric)
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)

Knights Landing: 2nd generation Intel® Xeon Phi™

Performance
3+ TeraFLOPS of double-precision peak theoretical performance per single socket node
3x Single-Thread Performance compared to Knights Corner
Most of today's parallel optimizations carry forward to KNL simply by recompile

Integration	
Intel® Omni Path™ fabric integration	
High-performance on-package memory (MCDRAM)	Over 5x STREAM vs. DDR4 (Over 400 GB/s vs 90 GB/s)
	Up to 16GB at launch
	NUMA support
	Over 5x Energy Efficiency vs. GDDR5
	Over 3x Density vs. GDDR5
	In partnership with Micron Technology
	Flexible memory modes including cache and flat

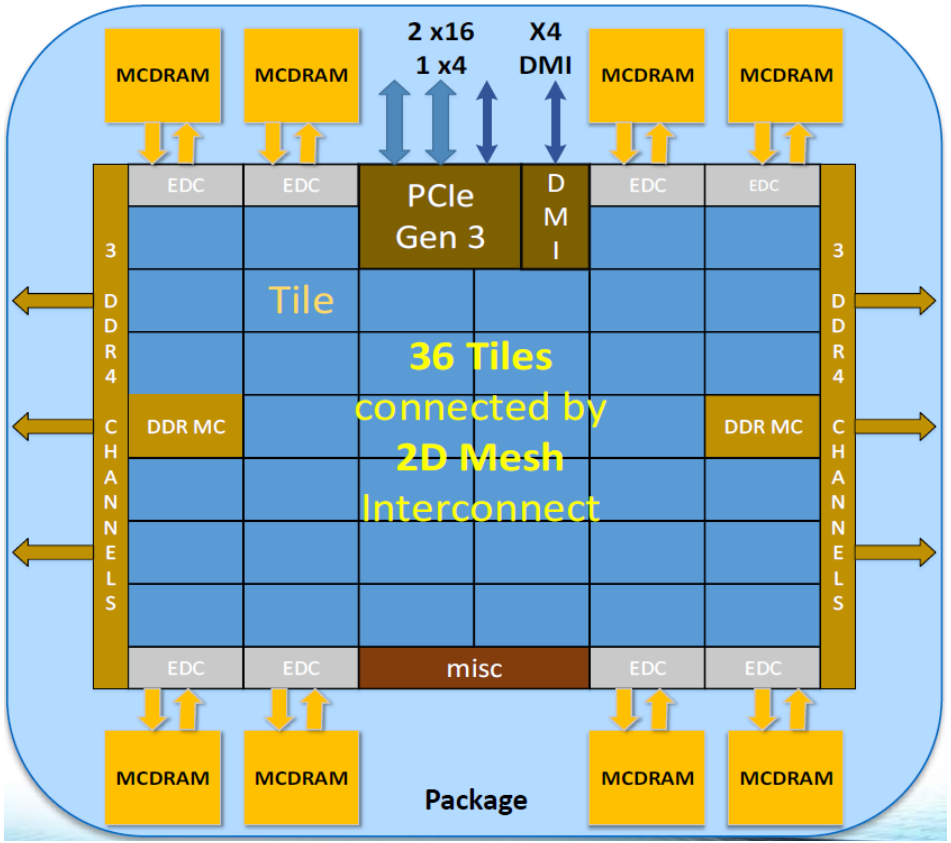


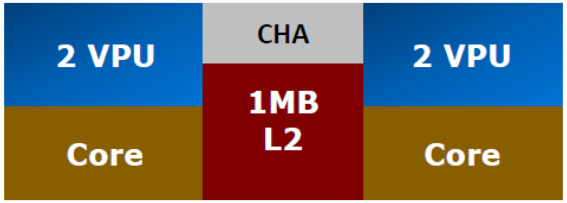
Microarchitecture	
Over 8 billion transistors per die based on Intel's 14 nanometer manufacturing technology	
Binary compatible with Intel® Xeon® Processors with support for Intel® Advanced Vector Extensions 512 (Intel® AVX-512)	
72 cores in a 2D Mesh architecture	
2 cores per tile with 2 VPUs per core	
1MB L2 cache shared between 2 cores in a tile (cache-coherent)	
Cores based on Intel® Atom™ (Silvermont) microarchitecture with many HPC enhancements	4 Threads / Core
	2X Out-of-Order Buffer Depth
	Gather/scatter in hardware
	Advanced Branch Prediction
	High cache bandwidth
	32KB Icache, Dcache
	2 x 64B Load ports in Dcache
	46/48 Physical/virtual address bits
Multiple NUMA domain support per socket	

Server processor
Standalone bootable processor (running host OS) and a PCIe coprocessor (PCIe end-point device)
Platform memory: up to 384GB DDR4 using 6 channels
Reliability ("Intel server-class reliability")
Power Efficiency (Over 25% better than discrete coprocessor) → Over 10 GF/W
Density (3+ KNL with fabric in 1U)
Up to 36 lanes PCIe* Gen 3.0

Availability
First commercial HPC systems in 2H'15
Knights Corner to Knights Landing upgrade program available today
Intel Adams Pass board (1U half-width) is custom designed for Knights Landing (KNL) and will be available to system integrators for KNL launch; the board is OCP Open Rack 1.0 compliant, features 6 ch native DDR4 (1866/2133/2400MHz) and 36 lanes of integrated PCIe* Gen 3 I/O

Knights Landing platform overview





Single socket node

- 36 tiles connected by coherent 2D-Mesh
- Every tile is 2 OoO cores +
  2 512-bit VPU/core + 1 MB L2

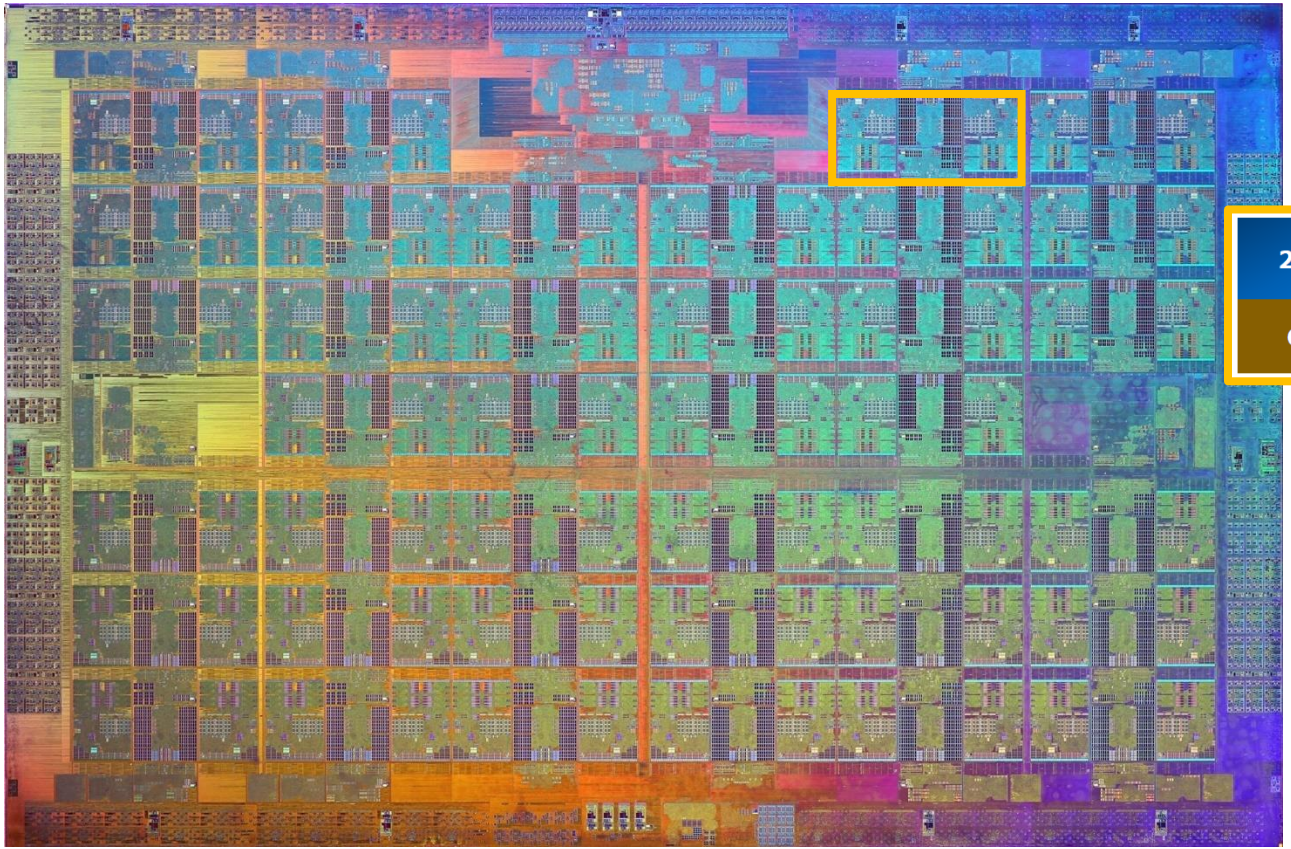
Memory

- MCDRAM, 16 GB on-package; High BW
- DDR4, 6 channels @ 2400 up to 384GB

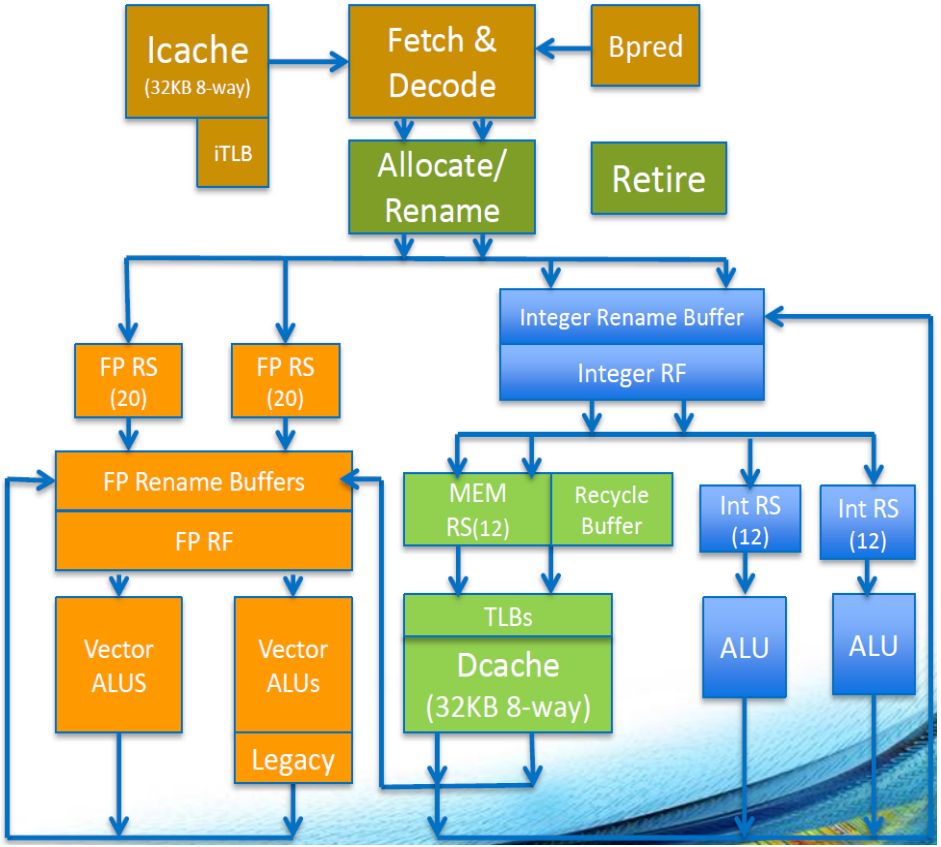
IO & Fabric

- 36 lanes PCIe Gen3
- 4 lanes of DMI for chipset
- On-package Omni-Path fabric

Intel® Knights Landing die



Knights Landing core architecture



OoO core w/ 4 SMT threads

- 2-wide decode/rename/retire
- Up to 6-wide at execution
- Int and FP RS OoO
- 2 AVX-512 VPUs

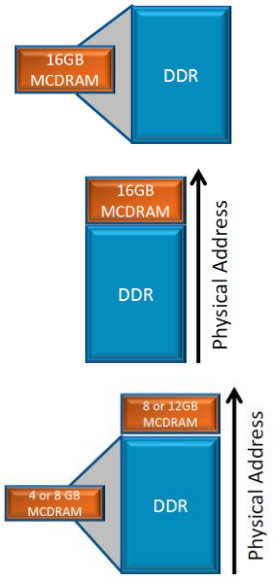
$s/TLBs

- 64-bit Dcache ports (2-load & 1-store)
- 1st level uTLB w/ 64 entries
- 2nd level dTLB w/ 256-4K, 128-2M, 16-1G pages

Others

- L1 (IPP) and L2 prefetcher.
- Fast unaligned support
- Fast gather/scatter support

Knights Landing's on package HBM memory



Cache Model

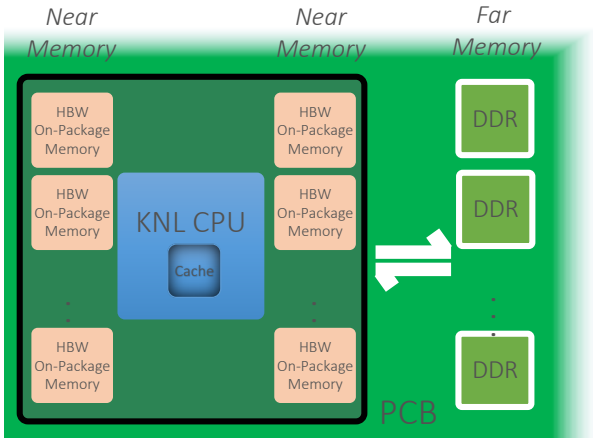
Let the ***hardware automatically manage*** the integrated on-package memory as an "L3" direct-mapped cache between KNL CPU and external DDR

Flat Model

Manually manage how your application uses the integrated on-package memory and external DDR for peak performance

Hybrid Model

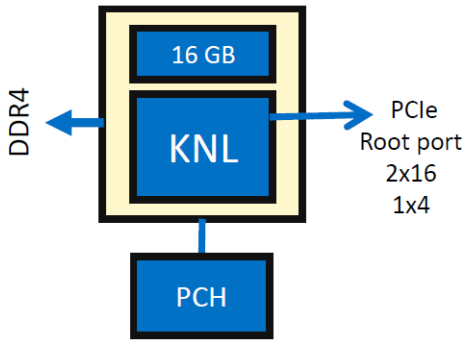
Harness the ***benefits of both*** cache and flat models by segmenting the integrated on-package memory



Maximizes performance through higher memory bandwidth and flexibility

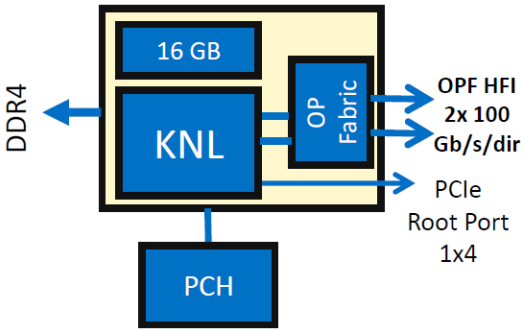
- Explicit allocation allowed with open-sourced API, Fortran attributes, and C++ allocator

Knights Landing products



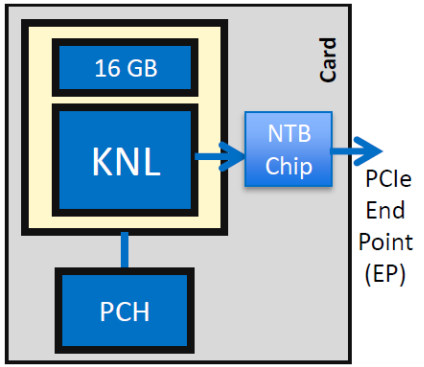
KNL

6 DDR channels
Up to 16 GB MCDRAM
36-lanes Gen3 PCIe (root port)



KNL w/ Omni-Path

6 DDR channels
Up to 16 GB MCDRAM
4-lanes Gen3 PCIe (root port)
Omni-Path fabric (200 Gb/s/dir)



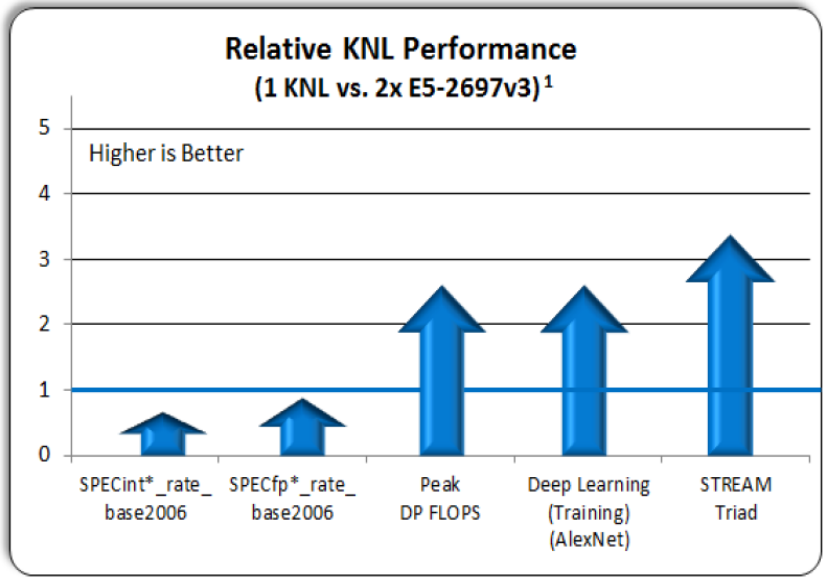
KNL Card

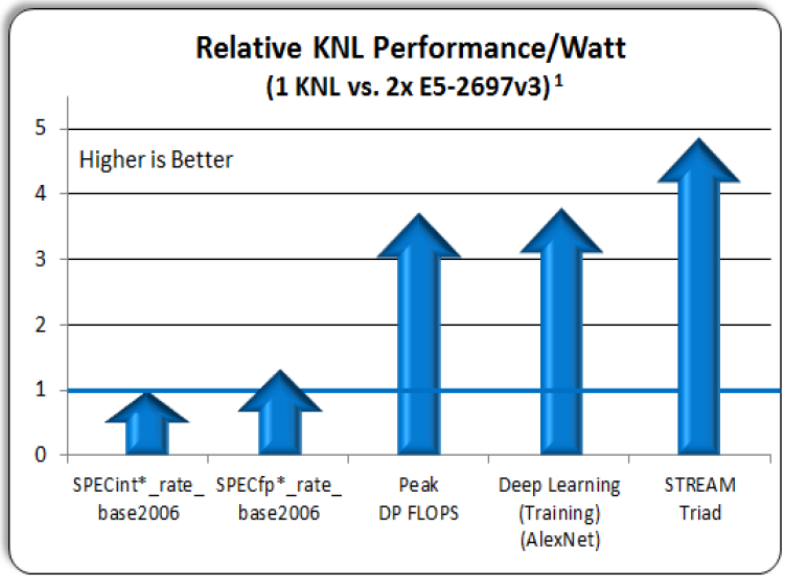
No DDR Channels
Up to 16 GB MCDRAM
16-lanes Gen3 PCIe (end point)
NTB Chip to create PCIe EP

Self boot socket

PCIe Card

Knights Landing performance





[1]Projected KNL Performance (1 socket, 200W CPU TDP) vs. 2 Socket Intel® Xeon® processor E5-2697v3 (2x145W CPU TDP)

Significant performance improvement for compute and bandwidth sensitive workloads, while still providing good general purpose throughput performance

Key new features for software adaptation to KNL

Large impact: **Intel® AVX-512 instruction set**

- 32 512-bit FP/Int vector registers, 8 mask registers, HW gather/scatter
- Slightly different from future Intel® Xeon™ architecture AVX-512 extensions
- Backward compatible with SSE, AVX, AVX-2
- Apps built for HSW and earlier can run on KNL (few exceptions like TSX)
- Incompatible with 1st Generation Intel® Xeon Phi™ (KNC)

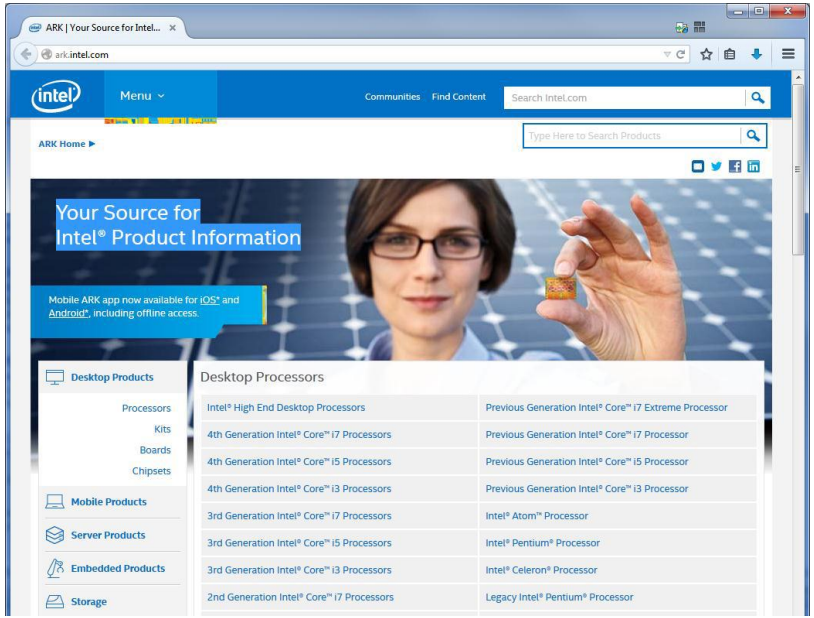
Medium impact: **new, on-chip high bandwidth memory (HBM)**

- Creates heterogeneous (NUMA) memory access
- Can be used transparently too however

Minor impact: **differences in floating point execution/rounding**

- New HW-accelerated transcendental functions like exp()

Reference sites online



Detailed information about available Intel products

Encoding scheme of Intel processor numbers