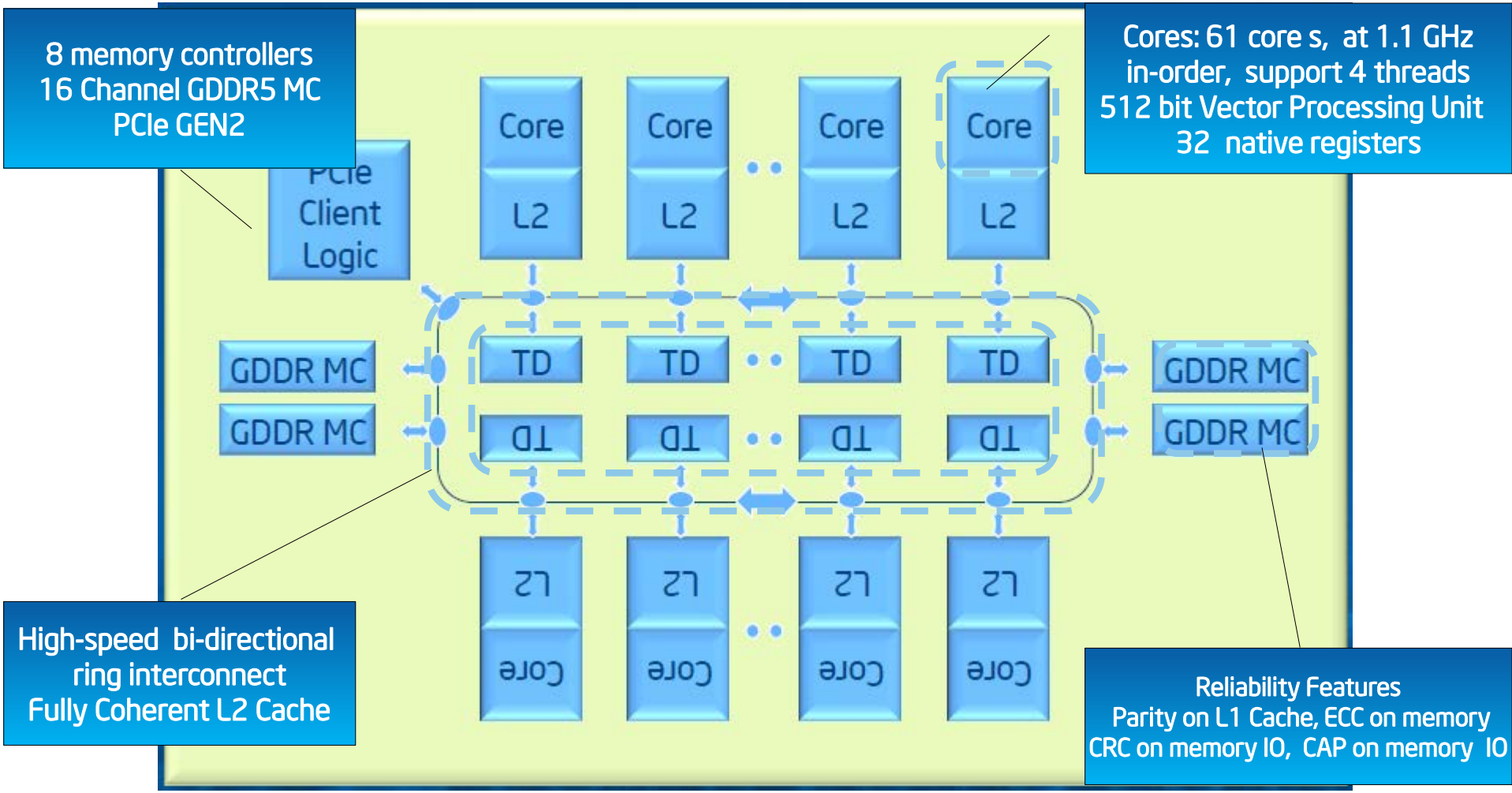
A quick introduction to the Intel® Xeon Phi™

Stephen Blair-Chappell, Intel

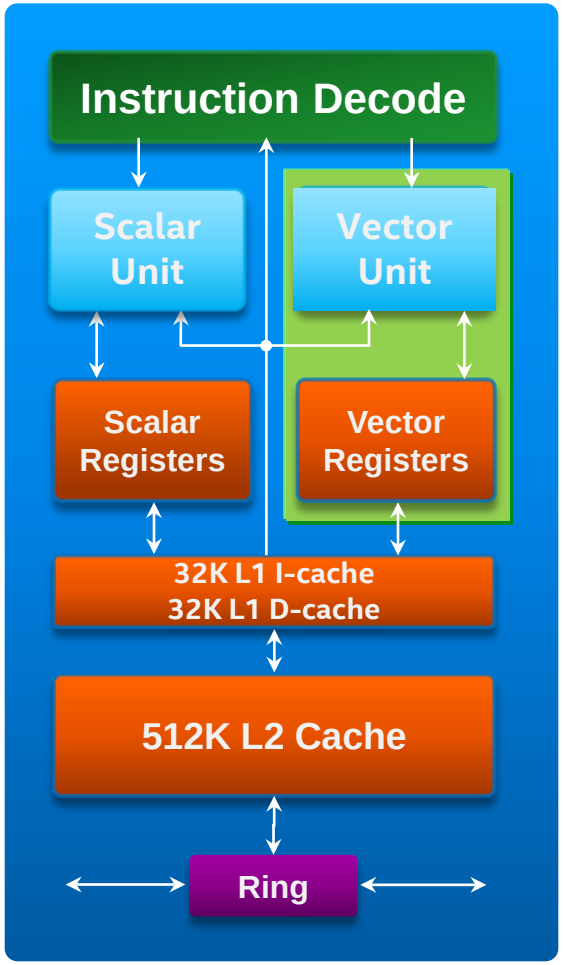
What is it?

- Co-processor
  - PCI Express card
  - Stripped down Linux operating system (busybox/dash)
- Dense, simplified processor
  - Simplifications for power savings In-order
  - Wider vector unit
  - Wider hardware thread count
- Lots of names
  - Many Integrated Core architecture, aka MIC
  - Knights Corner (code name)
  - Intel Xeon Phi Co-processor SE10P (product name)

Intel® Xeon Phi™ Architecture Overview



Core Architecture Overview



60+ in-order, low power IA cores in a ring interconnect

Two pipelines

- Scalar Unit based on Pentium® processors
- Dual issue with scalar instructions
- Pipelined one-per-clock scalar throughput

SIMD Vector Processing Engine

4 hardware threads per core

- 4 clock latency, hidden by round-robin scheduling of threads
- Cannot issue back to back inst in same thread

Coherent 512KB L2 Cache per core

Key Differentiators
Xeon Phi vs Workstation

More Cores

Slower Clock Speed

Wider SIMD registers

Faster Bandwidth

In-order pipeline

A Tale of Two Architectures

	Intel® Xeon® processor	Intel® Xeon Phi™ Coprocessor
Sockets	2	1
Clock Speed	2.6 GHz	1.1 GHz
Execution Style	Out-of-order	In-order
Cores/socket	8	Up to 61
HW Threads/Core	2	4
Thread switching	HyperThreading	Round Robin
SIMD widths	8SP, 4DP	16SP, 8DP
Peak Gflops	692SP, 346DP	2020SP, 1010DP
Memory Bandwidth	102GB/s	320GB/s
L1 DCache/Core	32kB	32kB
L2 Cache/Core	256kB	512kB
L3 Cache/Socket	30MB	none

Theoretical Peak Flops Performance Example

Frequency * Num Sockets * Num Cores * Vector Width * FP Ops

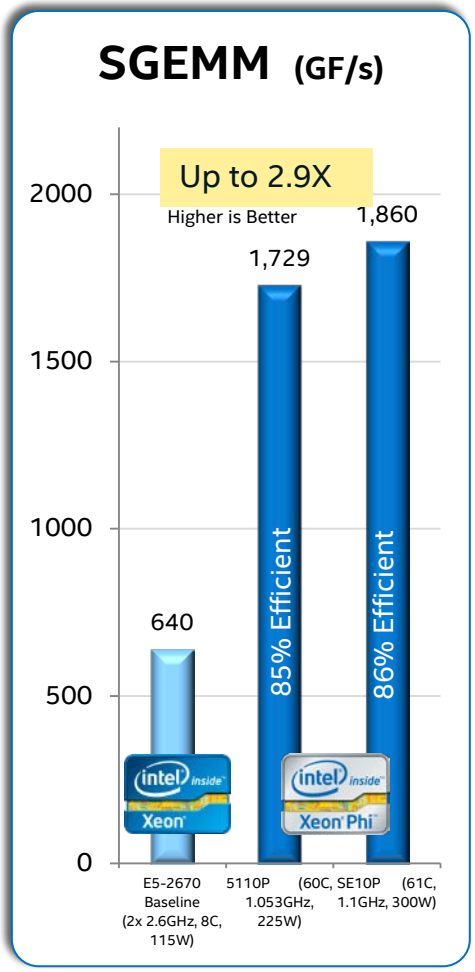
Two socket Intel® Xeon® E5-2670 Processor

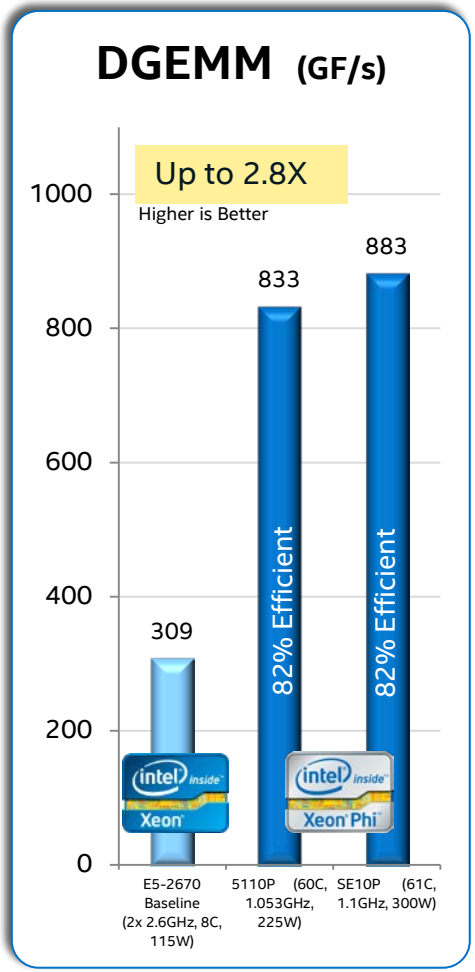
Freq	Sockets	Num Cores	Vector Width	FP Ops	GFlops
2.6	2	8	4	2	666

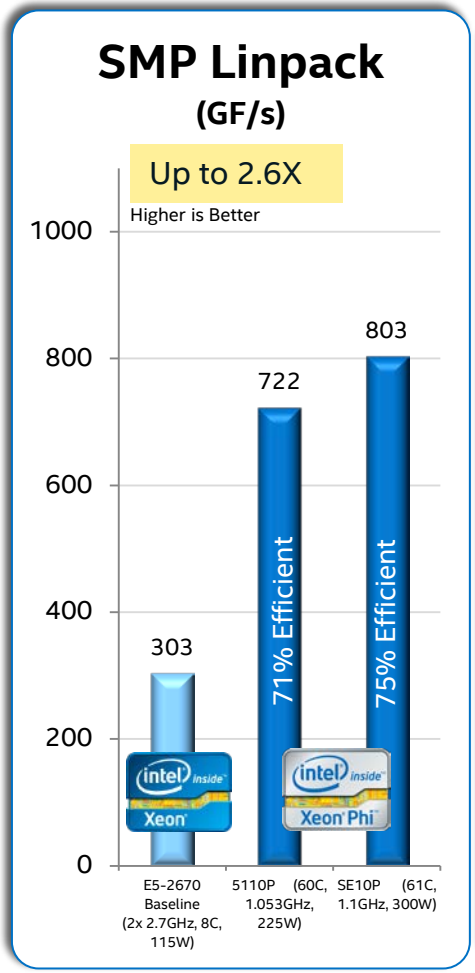
Single card Xeon Phi Coprocessor (B0)

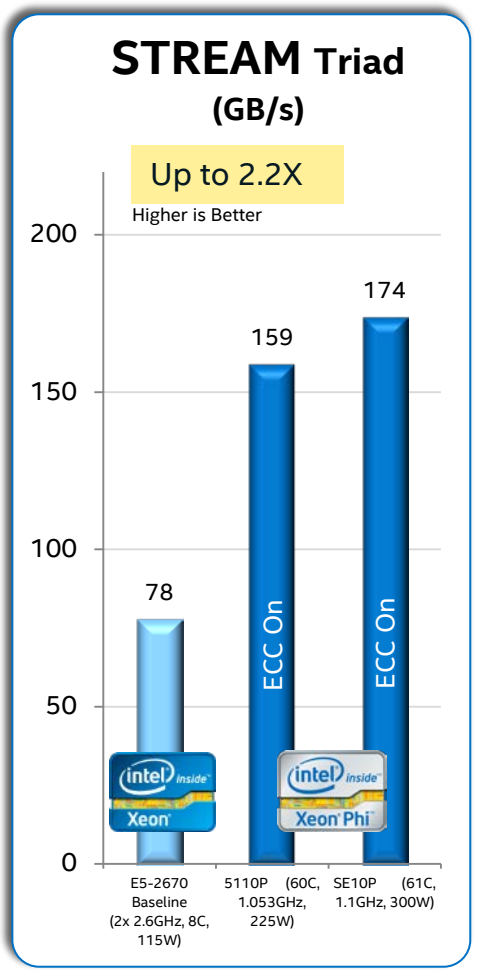
Freq	Sockets	Num Cores	Vector Width	FP Ops	GFlops
1.091	1	61	16	2 (using FMA)	2,128

Synthetic Benchmark Summary (Intel® MKL)









Coprocessor results: Benchmark run 100% on coprocessor, no help from Intel® Xeon® processor host (aka native)

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 Measured results as of October 26, 2012 Configuration Details: Please reference slide speaker notes.
For more information go to http://www.intel.com/performance

Intel® Xeon Phi™ Coprocessor:

Increases Application Performance up to 10x

Updated

Segment	Customer	Application	Performance Increase[1] vs. 2S Xeon*
Energy	Acceleware	8th order isotropic variable velocity	Up to 2.23x
	Sinopec	Seismic Imaging	Up to 2.53x[2]
	CNPC (China Oil & Gas)	GeoEast Pre-Stack Time Migration (Seismic)	Up to 3.54x[2]
Financial Services	Financial Services	BlackScholes SP Monte Carlo SP	Up to 7.5x Up to 10.75x
Physics	Jefferson Labs	Lattice QCD	Up to 2.79x
Finite Element	Sandia Labs	miniFE (Finite Element Solver)	Up to 2x[3] Up to 1.3x[5]
Solid State Physics	ZIB (Zuse-Institut Berlin)	Ising 3D (Solid State Physics)	Up to 3.46x
Digital Content Creation/Video	Intel Labs	Ray Tracing (incoherent rays)	Up to 1.88x[4]
	NEC	Video Transcoding	Up to 3.0x[2]
Astronomy	CSIRO/ASKAP (Australia Astronomy)	tHogbom Clean (Astronomy image smear removal)	Up to 2.27x
	TUM (Technische Universität München)	SG++ (Astronomy Adaptive Sparse Grids/Data Mining)	Up to 1.7x
Fluid Dynamics	AWE (Atomic Weapons Establishment - UK)	Cloverleaf (2D Structured Hydrodynamics)	1.77x

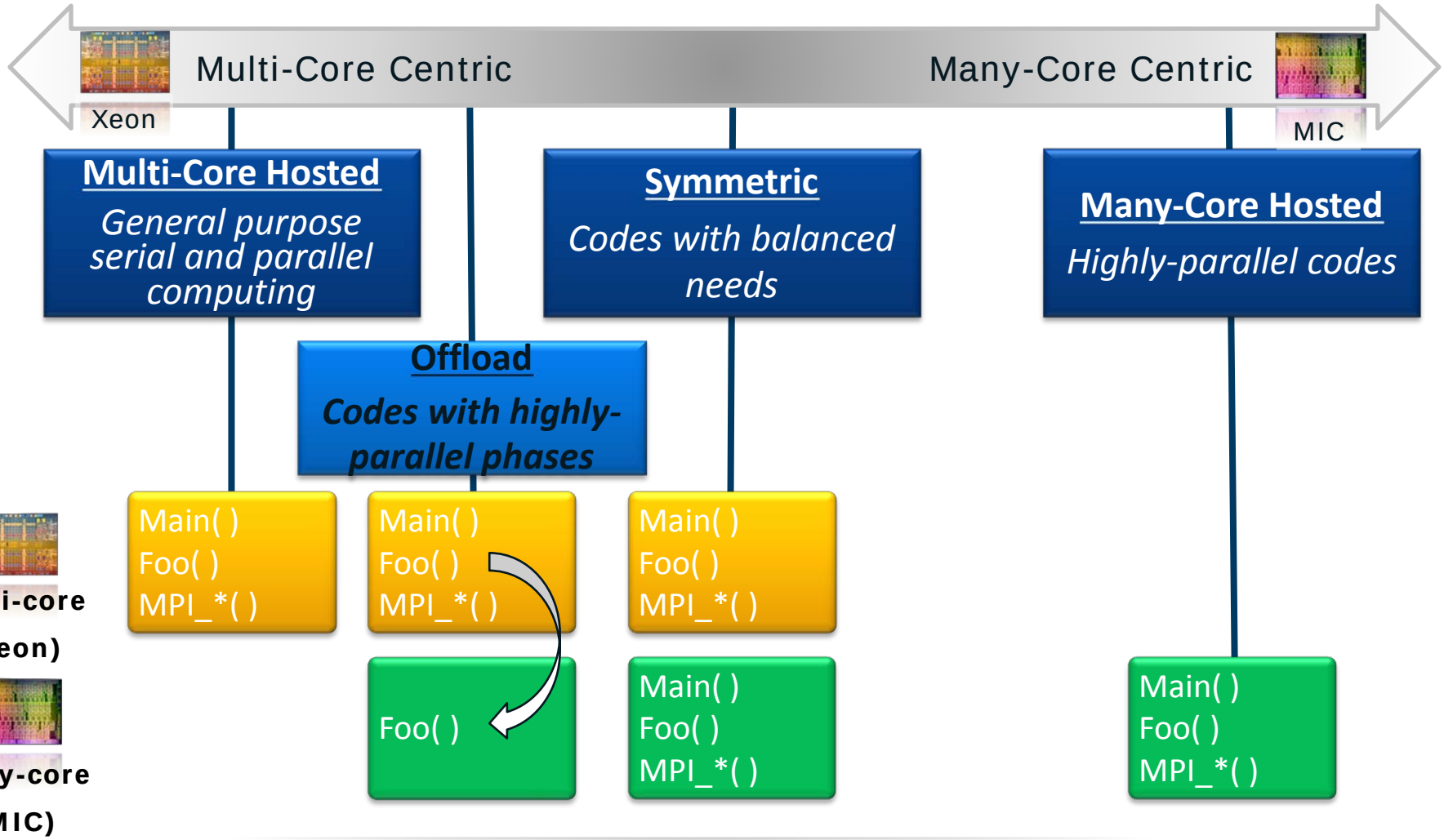
Notes:

1. 2S Xeon* vs. 1 Xeon Phi* (preproduction HW/SW & Application running 100% on coprocessor unless otherwise noted)
2. 2S Xeon* vs. 2S Xeon* + 2 Xeon Phi* (offload)
3. 8 node cluster, each node with 2S Xeon* (comparison is cluster performance with and without 1 Xeon Phi* per node) (Hetero)
4. Intel Measured Oct. 2012
5. 8 node cluster, each node with 2S Xeon* (comparison is cluster performance with Xeon only vs. Xeon Phi *only (1 Xeon Phi* per node) (Native)

Source: Customer Measured results as of October 22, 2012 Configuration Details: Please reference slide speaker notes.

For more information go to http://www.intel.com/performance

Programming Models and Mindsets



Range of models to meet application needs

Examples of Offloading

C/C++ Offload Pragma

```
#pragma offload target (mic)
#pragma omp parallel for reduction(+:pi)
for (i=0; i<count; i++)  {
        float t = (float)((i+0.5)/count);
        pi += 4.0/(1.0+t*t);
}
pi /= count;
```

MKL Implicit Offload

```
//MKL implicit offload requires no source code changes,
     simply link with the offload MKL Library.
```

MKL Explicit Offload

```
#pragma offload target (mic) \
     in(transa, transb, N, alpha, beta) \
     in(A:length(matrix_elements)) \
     in(B:length(matrix_elements)) \
     in(C:length(matrix_elements)) \

out(C:length(matrix_elements)alloc_if(0))
     sgemm(&transa, &transb, &N, &N, &N, &alpha,
          A, &N, B, &N, &beta, C, &N);
```

Fortran Offload Directive

```
!dir$ omp offload target(mic)
!$omp parallel do
      do i=1,10
      A(i) = B(i) * C(i)
      enddo
!$omp end parallel
```

C/C++ Language Extensions

```
class _Shared common  {
    int data1;
    char *data2;
    class common *next;
    void process();
};
_Shared class common  obj1, obj2;

…
_Cilk_spawn _Offload obj1.process();
_Cilk_spawn          obj2.process();

…
```

KNL Public Knowledge

- Knights Landing is the code name for the **2nd generation** product in the Intel® Many Integrated Core Architecture
- Knights Landing targets Intel's **14 nanometer** manufacturing process
- Knights Landing will be productized as a **processor** (running the host OS) and a **coprocessor** (a PCIe end-point device)
- Knights Landing will feature on-package, **high-bandwidth memory**
- **Flexible memory modes** for the on package memory include: flat, cache, and hybrid modes
- Intel® Advanced Vector Extensions **AVX-512**

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.

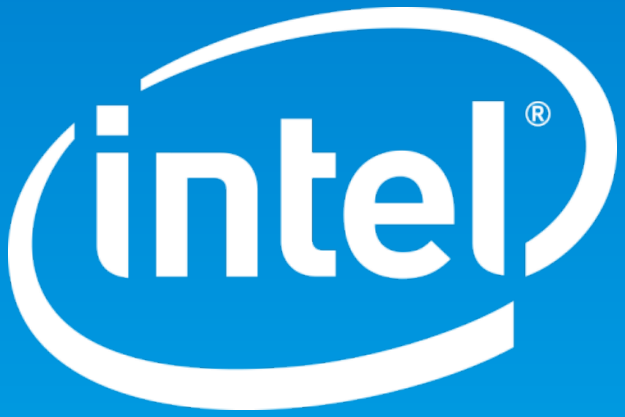
Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may 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.

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

Backup


intel