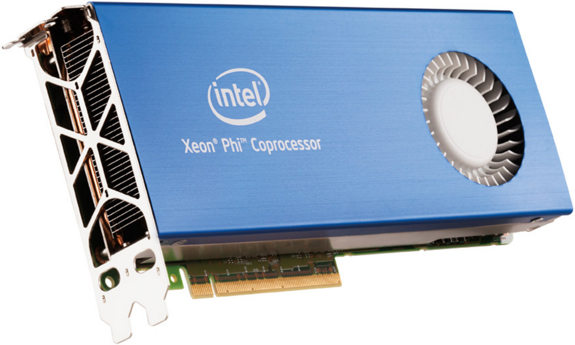
# Intel Xeon Phi Knights Landing (KNL)

SNUG TechTalk

SciNet
www.scinet.utoronto.ca
University of Toronto
Toronto, Canada

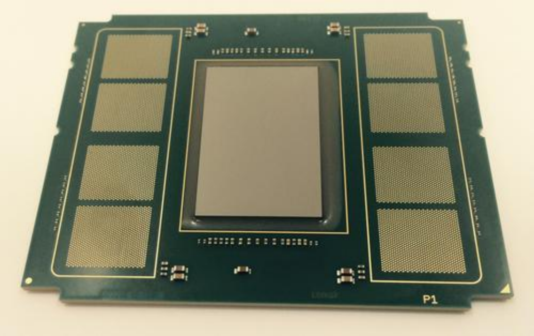
October 12, 2016

# Xeon Phi



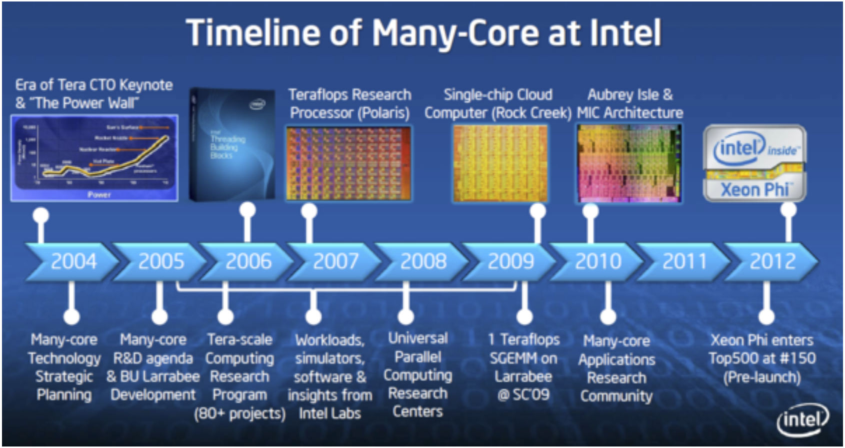
http://www.intel.com

# Xeon Phi



http://www.intel.com

# History



## Timeline of Many-Core at Intel

Era of Tera CTO Keynote & "The Power Wall"

Teraflops Research Processor (Polaris)

Single-chip Cloud Computer (Rock Creek)

Aubrey Isle & MIC Architecture

2004 2005 2006 2007 2008 2009 2010 2011 2012

Many-core Technology Strategic Planning

Many-core R&D agenda & BU Larrabee Development

Tera-scale Computing Research Program (80+ projects)

Workloads, simulators, software & insights from Intel Labs

Universal Parallel Computing Research Centers

1 Teraflops SGEMM on Larrabee @ SC'09

Many-core Applications Research Community

Xeon Phi enters Top500 at #150 (Pre-launch)

http://www.intel.com

# History

## Xeon "Phi" History

- Larrabee (concept video card design)
- Knights ferry (MIC development platform)
  - 32 core, 2GB
  - $\sim$ 750 GFlops SP
- Knights Corner (1st Gen)
  - 60 core, 8/16GB
  - $\sim$ 1 TFlops DP
- Knights Landing (2nd Gen)
  - 72 core, 16 + upto 384GB
  - $\sim$ 3 TFlops DP
  - native processor
  - omni-path interconnect
- Knights Hill (3rd Gen 2018)
  - $>$ 5 TFlops DP
  - omni-path 2 interconnect

# Compute Canada Xeon Phi Resources

## SciNet - KNL

- 4 x Intel Xeon Phi 7210 (64 @1.3 GHz cores and 96+16GB)
- ssh knl0[1-4]
- module load intel/17.0.0

## Calcul Quebec - Guillimin

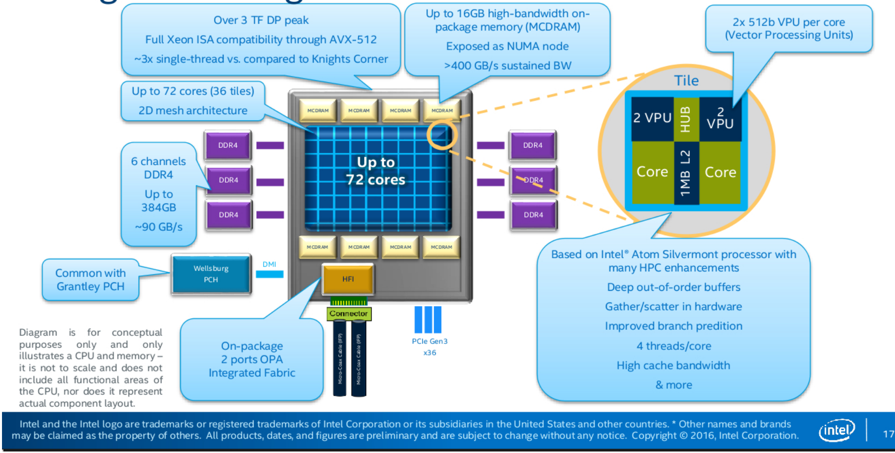
- 50 nodes (2 x 8-core Intel Sandy Bridge Xeon, 64GB)
- 2 x Intel Xeon Phi 5110P (60 1.053GHz cores and 8GB)

# 2nd Generation Xeon Phi - KNL

## Knights Landing (KNL)

- 64-72 core @ 1.3-1.4 GHz
- 4 threads/core
- 16GB MCDRAM ( 430 GB/s STREAM)
- 384GB DDR4 ( 90 GB/s STREAM)
- AVX-512
- $\sim$ 3 TFlops DP
- Self Host or PCIE coprocessor
- Integrated Intel Omni-Path Fabric *

# Knights Landing: Architecture



Over 3 TF DP peak

Full Xeon ISA compatibility through AVX-512

~3x single-thread vs. compared to Knights Corner

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)

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

Based on Intel® Atom Silvermont processor with many HPC enhancements

Deep out-of-order buffers

Gather/scatter in hardware

Improved branch predition

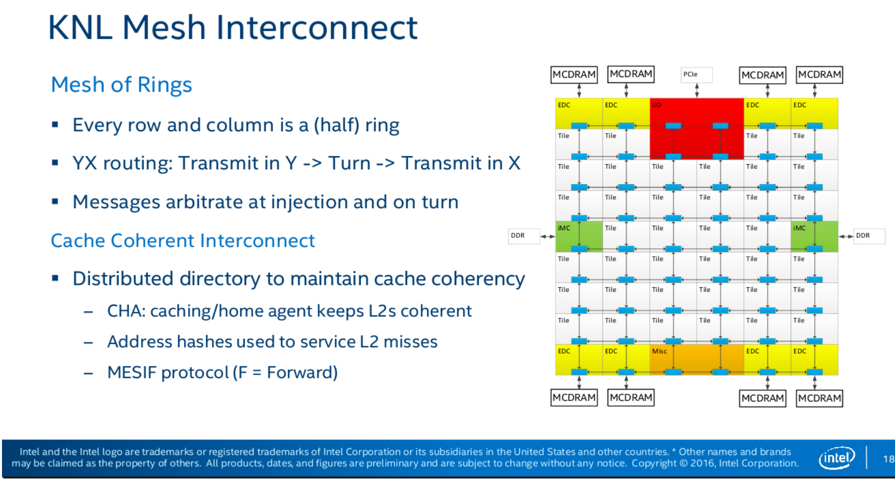
4 threads/core

High cache bandwidth

& more

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.

http://www.intel.com



# KNL Mesh Interconnect

## 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)

MCDRAM MCDRAM PCIe MCDRAM MCDRAM
EDC EDC IIO EDC EDC
Tile Tile Tile Tile
Tile Tile Tile Tile Tile Tile
Tile Tile Tile Tile Tile Tile
DDR iMC Tile Tile Tile Tile iMC DDR
Tile Tile Tile Tile Tile Tile
Tile Tile Tile Tile Tile Tile
Tile Tile Tile Tile Tile Tile
EDC EDC Misc EDC EDC
MCDRAM MCDRAM MCDRAM MCDRAM

http://www.intel.com

# Cluster Modes

## Cluster Modes

- All-to-All
  - Memory Addressed uniformly
  - Lowest performance
- Quadrant (default on knl0[1-4])
  - Four Virtual Quadrants
  - Lower Latency and higher BW than All-to-All
  - SW transparent
- Sub-NUMA Clustering (SNC)
  - 4 Separate NUMA domains
  - Lowest latency
  - SW needs to be NUMA optimized

# Xeon Phi Memory Modes

What to do with 16GB of MCDRAM?

## Memory Modes

- Cache (knl0[1-2]])
  - OS Treats it like an L2 Cache
  - User only sees/controls DDR memory
- Flat (knl0[3-4]])
  - Explicitly allocatable
  - NUMA regions (0 DDR / 1 MCDRAM)
- Hybrid
  - 50% / 50% cache/memory
  - 25% / 75% cache/memory

# Xeon Phi Memory Modes

## Memory Modes

- Cache
    - No code modifications required
    - Latency hit to DDR ( DDR → MCDRAM → L2)
    - Less total memory addressable
- Flat
    - Maximum bandwidth and latency performance
    - Max Memory addressable
    - Code modifications to use both in the same application

# Xeon Phi Memory Modes - Numactl

Numactl - Control NUMA policy for processes or shared memory

```
user@knl03 $numactl -H
```

```
user@knl03 $numactl --membind 1 ./run-app
```

# Xeon Phi Memory Modes - Memkind Library

The memkind library is an extensible heap manager allowing partitioning of heap between “kinds” of memory. A hbwmalloc interface simplifies usage, built on top of memkind.

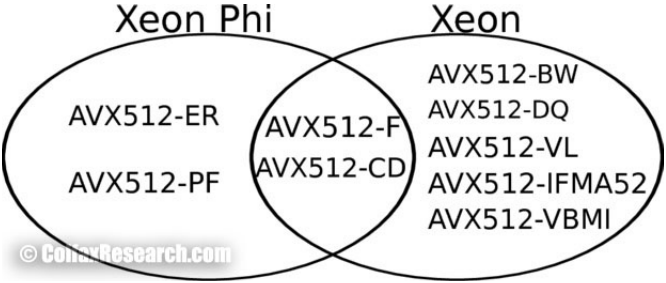
```
#include <hbwmalloc.h> // hbwmalloc interface
const int n = 1<<10;
// Allocation to HBM
double* A = (double*) hbw_malloc(sizeof(double)*n);
// Deallocate
hbw_free(A);
```

```
user@knl03 $icpc memkindtest.cc -lmemkind -o test-
memkind user@knl03 $
```

# KNL Vectorization

## Vector Support

- supports x87, MMX, SS3, AVX and AVX2
- AVX-512



http://www.colfaxresearch.com

# Programming

## Languages

- C, C++, Fortran
- MPI, OpenMP 4.0
- TBB, Cilk+

## Tools

- Intel Compilers (icc, icpc, ifort)
- Intel MPI
- Intel Tools (VTune, Advisor, Inspector, etc.)

# Which to Choose?

## GPU vs. Phi vs. CPU Considerations

- CUDA/OpenCL vs. native C/C++/Fortran & OpenMP/MPI
- 1500* vs 72 vs 24 cores
- GPU requires heterogeneous
- GPU has theoretically higher Flops (SP & DP)
- KNL has MCDRAM
- AVX-512 in KNL & Skylake
- Power8/9 have CAPI/Nvlink

# KNL Summary

## Xeon Phi Knights Landing

- High computational intensity (Flops/Watt)
- $\sim$ 3 TFlops DP
- native processor
- 16GB MCDRAM
- Intel's Blue Gene Q (Alan Gara)
- Top 500 list (#117 Stampede-KNL)
- May compete more with Skylake than Pascal/Volta

# Xeon Phi (KNL) Resourecs

## Useful Resources

- http://colfaxresearch.com/get-ready-for-intel-knights-landing-3-papers/
- http://dap.xeonphi.com/#implinks
- https://wiki.scinet.utoronto.ca/wiki/index.php/Knights_Landing
- https://software.intel.com/en-us/articles/getting-ready-for-KNL