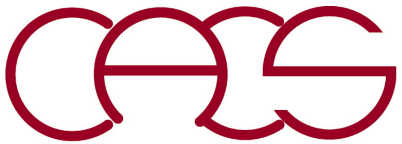
Intel Xeon Phi Programming

Aiichiro Nakano

Collaboratory for Advanced Computing & Simulations
Department of Computer Science
Department of Physics & Astronomy
Department of Chemical Engineering & Materials Science
Department of Biological Sciences
University of Southern California

Email: anakano@usc.edu

Goal: Multithreading on Intel Xeon Phi

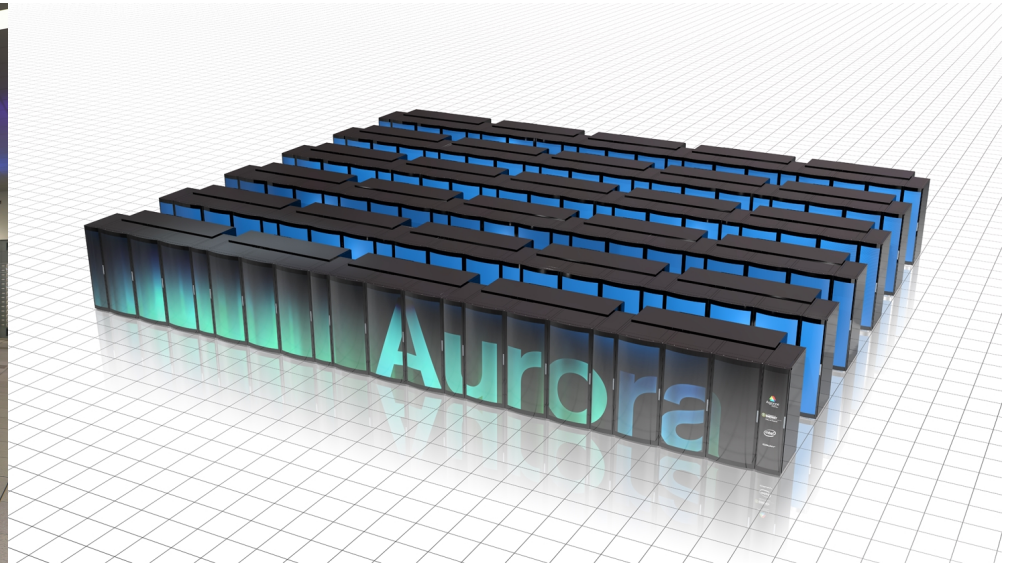


Two Supercomputing Parties in the US

GPU

Titan: Oak Ridge Nat'l Lab
17.6 Petaflop/s
AMD Opteron + NVIDIA K20x

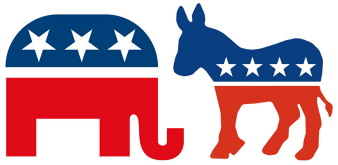
Phi



Aurora: Argonne Nat'l Lab (2019)
180-450 Petaflop/s
Intel Xeon Phi

CRAY THE SUPERCOMPUTER COMPANY
AMD
NVIDIA

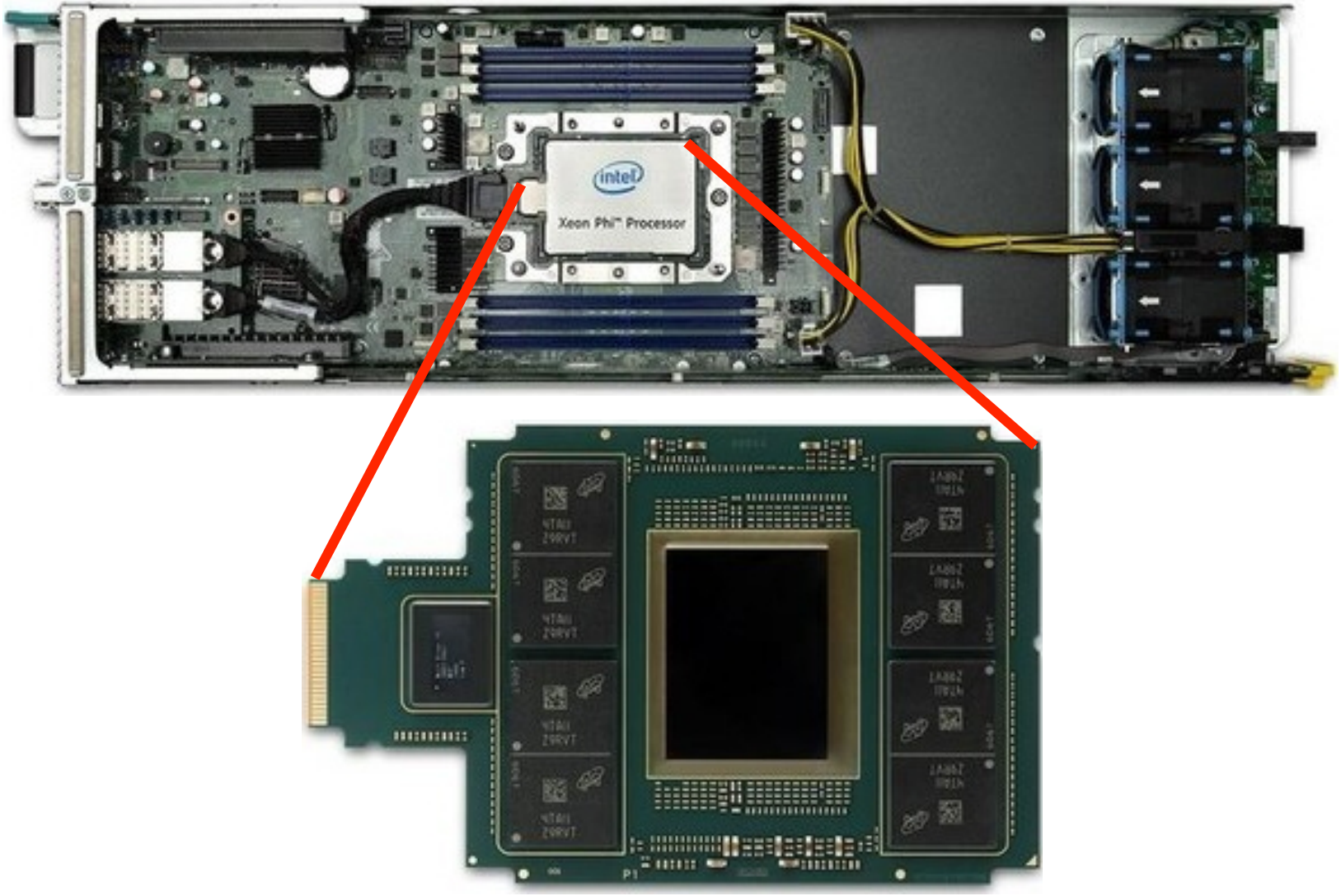
GPU *vs.* Phi



CRAY THE SUPERCOMPUTER COMPANY
intel

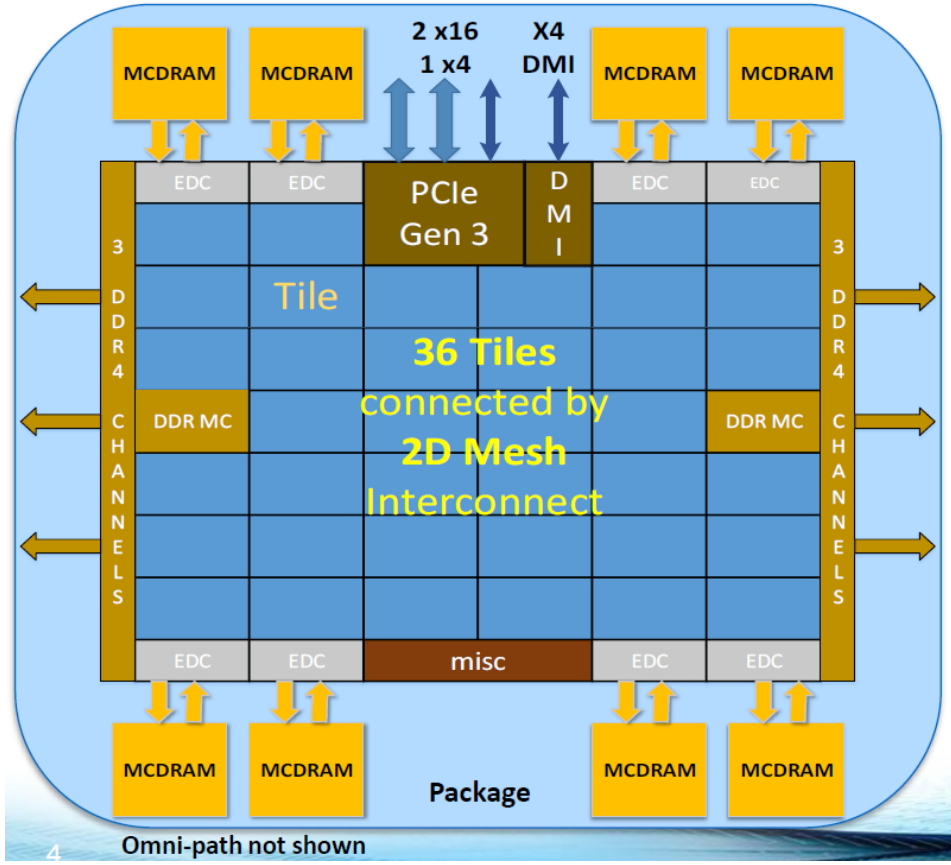
Intel Xeon Phi Processors

Current Knights Landing (KNL) is a predecessor of the Knights Hill (KNH) processor in Aurora

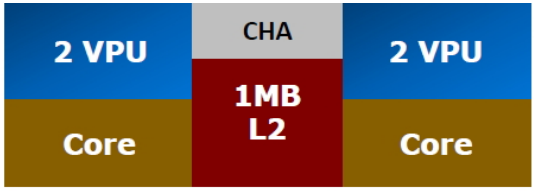


Knights Landing (KNL)

Knights Landing Overview



TILE



Chip: 36 Tiles interconnected by **2D Mesh**
Tile: 2 Cores + 2 VPU/core + 1 MB L2

Memory: MCDRAM: 16 GB on-package; High BW
DDR4: 6 channels @ 2400 up to 384GB
IO: 36 lanes PCIe Gen3. 4 lanes of DMI for chipset
Node: 1-Socket only
Fabric: Omni-Path on-package (not shown)

Vector Peak Perf: 3+TF DP and 6+TF SP Flops
Scalar Perf: ~3x over Knights Corner
Streams Triad (GB/s): MCDRAM : 400+; DDR: 90+

Source Intel: All products, computer systems, dates and figures specified are preliminary based on current expectations, and are subject to change without notice. KNL data are preliminary based on current expectations and are subject to change without notice. 1Binary Compatible with Intel Xeon processors using Haswell Instruction Set (except TSX). 2Bandwidth numbers are based on STREAM-like memory access pattern when MCDRAM used as flat memory. Results have been estimated based on internal Intel analysis and are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance.

VPU: Vector processing unit
MCDRAM: Multi-channel dynamic random access memory (4× bandwidth of DRAM)

KNL Parallel Programming

- **Standard MPI+OpenMP programming is supported**
- **Should utilize fast on-chip MCDRAM (multi-channel dynamic random access memory) shared by 72 cores**
- **Should take advantage of AVX-512 (512-bit or 8 double-precision) SIMD operations on vector processing units (VPUs)**