Developing the HPC compute cores of tomorrow

February 10th, 2014

Ofri Wechser

Intel Fellow, Visual and Parralel Group, Intel

Decomposing Emerging Applications

Execution

Evaluation

Clustering / Classification
Bayesian Network
Markov Model
Decision Trees
Forests of Trees
Neural Networks
Probabilistic Networks
Optimization-based models:
Linear/Non-linear/Stochastic
Time-series models

Modeling (Recognition)

What is …?

Mining

Is it …?

Synthesis

What if …?

Data Mining
Web Mining
Semantic Search
Streaming Data Mining
Distributed Data Mining
Content-based Image Retrieval
Query By Humming
Video Mining
Intrusion Mining
Strategy Simulation
Audio Synthesis
Video/Image Synthesis
Summary Synthesis
Machine Translation

Intel Transforming HPC

Circa 2003

MIPS
SGI

PowerPC
IBM

X86
Dell, HP, IBM

PA-RISC
HP

Itanium
HP, SGI, IBM

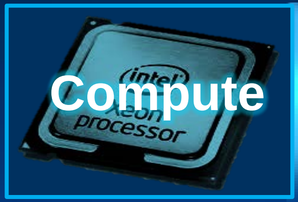
Vector
Cray, NEC

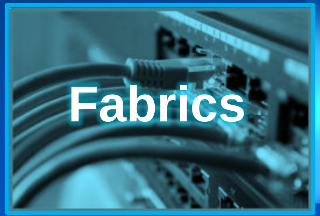
SPARC
Sun

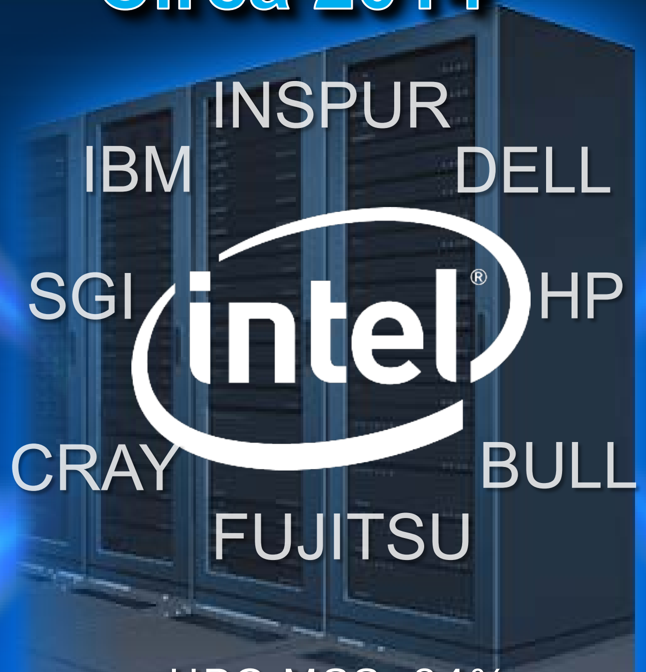
ALPHA
HP

intel ~20% of Top 500

Circa 2014







HPC MSS: 84%
80% of Top 500

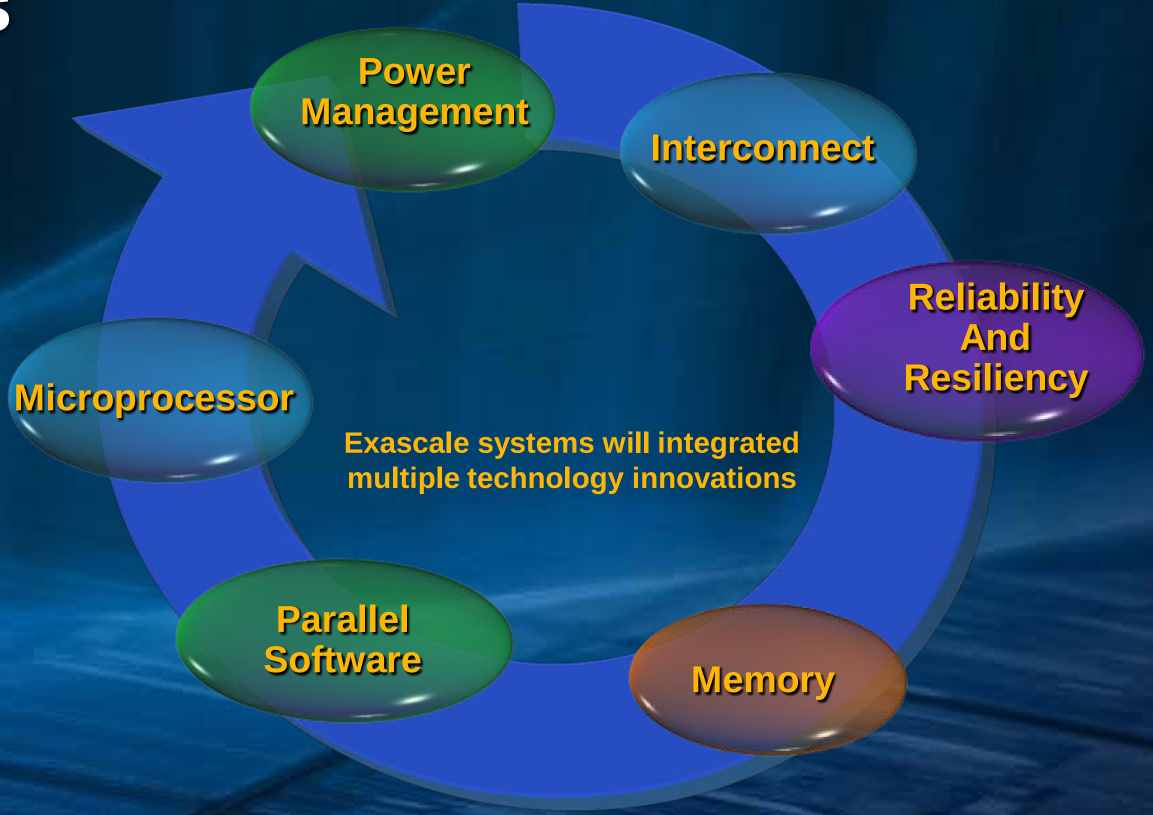
Source: Top 500.org

intel Look Inside.™

Key technical challenges

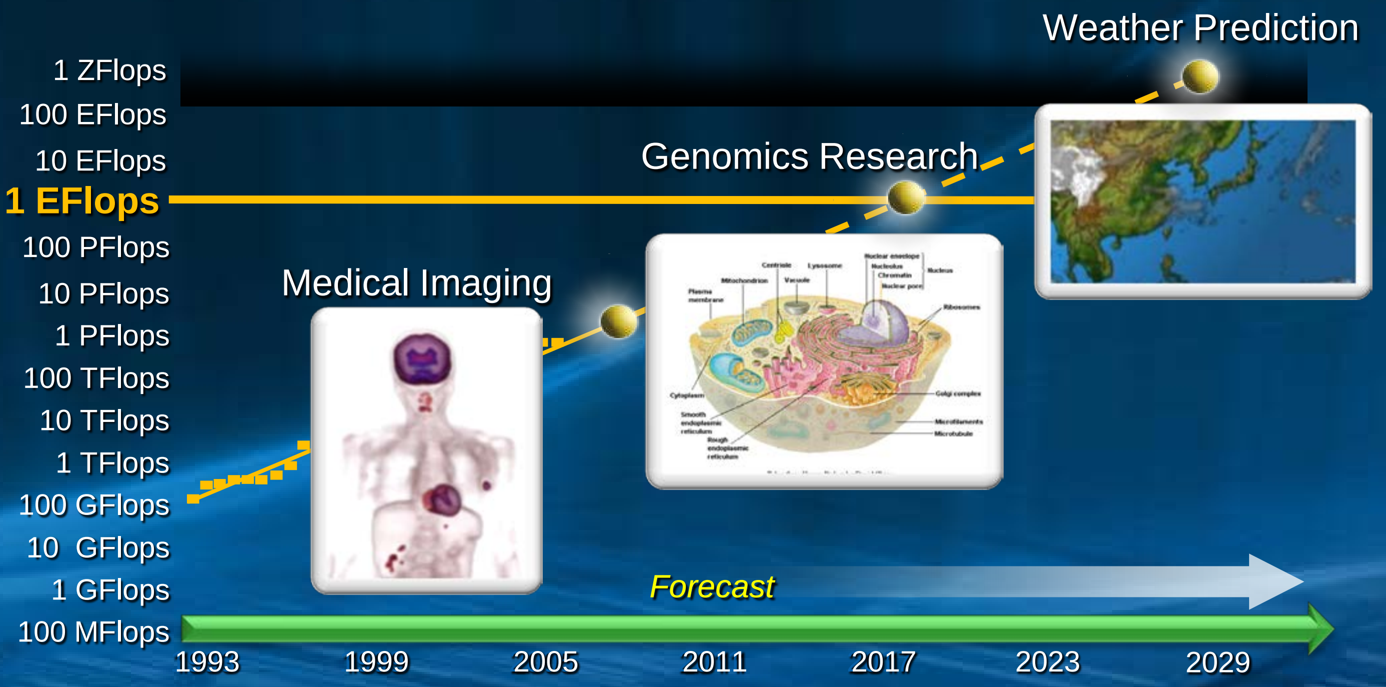
Key HPC system level challenges

- Compute density, efficiency
- Power efficiency
- Memory capacity, performance
- Network performance
- Reliability



Exascale

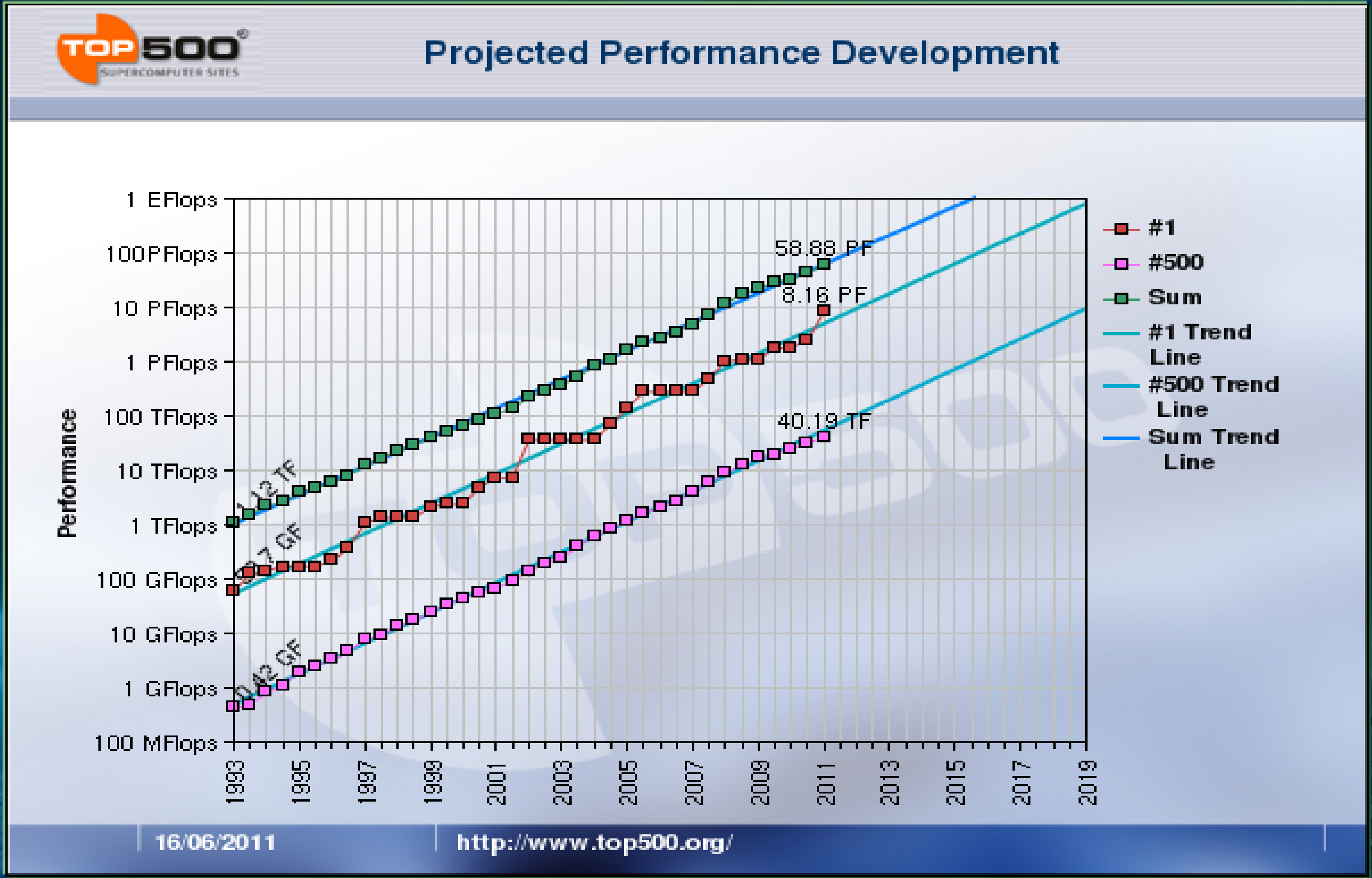
Goal: 1-ExaFlops (10^{18}) within 20 MW by 2018



Solve many yet impossible life changing problems

Make PFlop HPC computing affordable and ubiquitous

Trends to Exascale Performance



Roughly 10x performance every 4 years.

Predicts that we'll hit Exascale performance in 2018-19

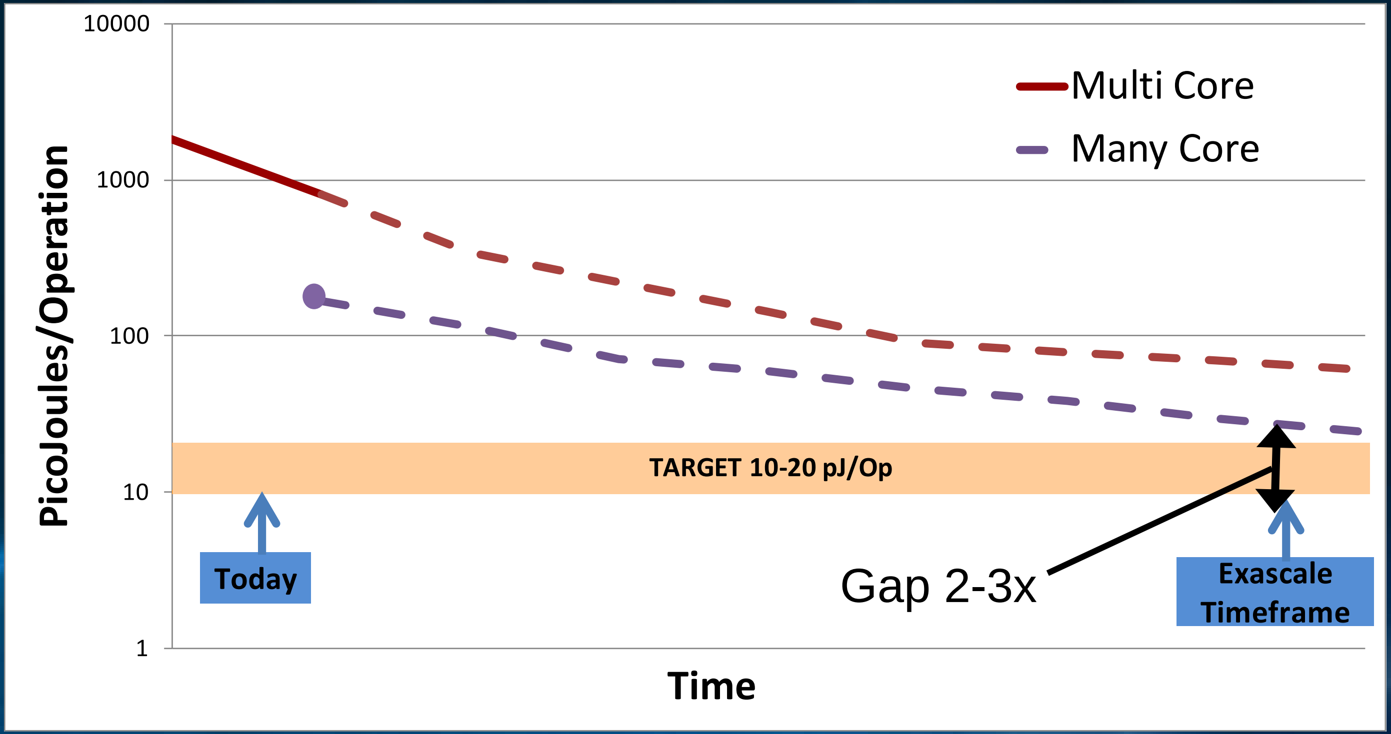
Quick status

At System Level:

- **Today:** 10 PF, 12 MW
  - **→1200 pJ/Op**
- **Exaflop:** 1000 PF, 20 MW
  - **→20 pJ/Op**

Needs improvements in all system components

- **Processor portion needs to reduce to 10 pJ/OP**
- **~60x improvement needed for Exascale**

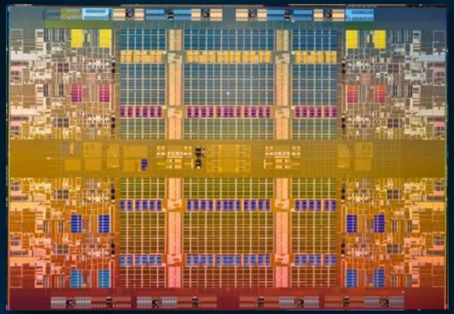


Can expect 1.5-3X improvement per generation

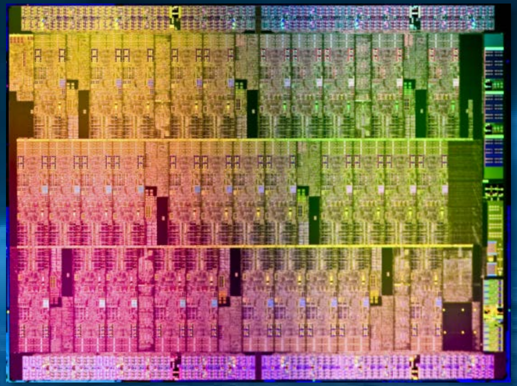
- **Process, uArch**

- **Still 2-3X gap to resolve ..**

Intel’s Many Core and Multi-core Engines



Multi-core Intel® Xeon® Processor at 2 - 3.5 GHz



Intel® Many Integrated Core (MIC) Architecture at 1-1.33 GHz

** Die size not to scale*

Intel® Xeon® processor:

- Intel’s Foundation of HPC Performance
- Suited for full scope of workloads
- Industry leading performance/watt for serial & highly parallel workloads.

Intel® Xeon Phi™ Architecture:

- Optimized for highly parallelized compute intensive workloads
- Common programming model & S/W tools with Xeon processors, enabling efficient app readiness and performance tuning
- Launched on 22nm with up to 61 cores and required B/W to provide outstanding performance for highly parallel HPC uses

The “Knights” Family

(* Project code names, not product names)

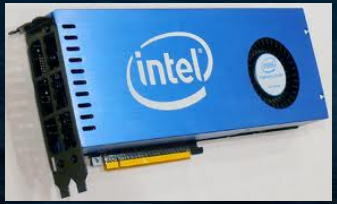
Knights Landing and Future Knights Products

Knights Corner (KNC)

- 1st Intel® Xeon Phi™ product
- 22nm process
- Up to 61 Intel Architecture Cores
- PCIe

Knights Ferry (KNF)

Software Development Platform



Knights Corner is…

Optimized, Highly Parallel
Intel Co-Processor
(connected via PCIe to Intel® Xeon® processor host)

Runs *Complete* Applications

IP Addressable

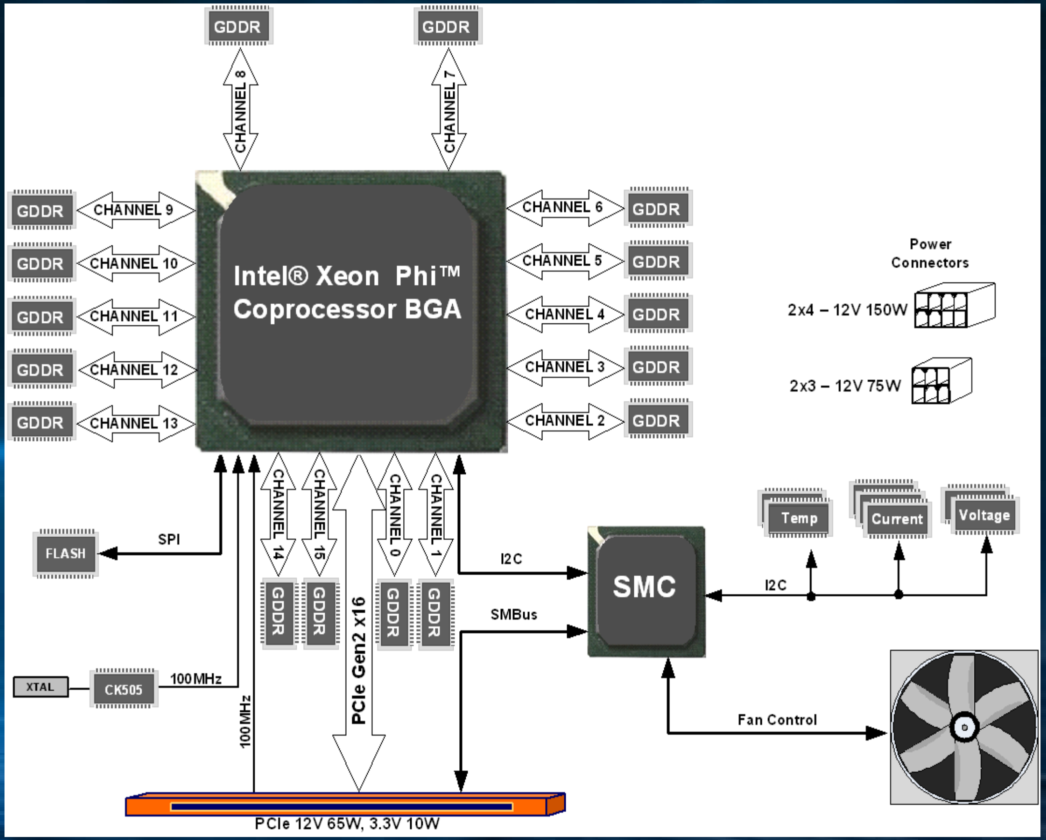
Open Source Linux OS

Common Source Code

Standard models of clustering

Builds on / Advances
State of the Art in Parallelism

Intel Developer tools



More Than 1 TFLOP/sec (DP) Sustained

ASCI Red: 1 TeraFlop/sec
December 1996

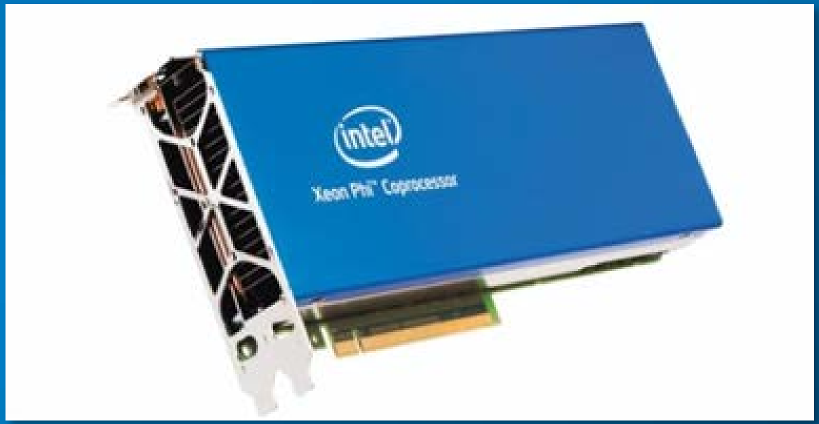
1996 First System 1 TF/s Sustained

7264 Intel® Pentium® Pro Processors**

OS: Cougar

72 Cabinets

** (with 2/3 system built…Full system later upgraded
to 3.1 TeraFlops/sec with 9298 Intel® Pentium II Xeon processors)



Knights Corner: 1 TeraFlop/sec
November 2011

2011 First Chip 1 TF/s Sustained

One 22nm Chip

OS: Linux*

One PCI Express* slot

Source" http://en.wikipedia.org/wiki/ASCI_Red Photo: Photo: http://www.top500.org/system/details/2758

Knights Corner Chip-level Architecture

- Up to 61 Cores

- High Performance On-Die Bidirectional Interconnect
  - Fully Coherent L2 Caches
  - Distributed Global Tag Directory

- Memory
  - 8 Memory Controllers
  - 16 GDDR5 Channels @ 5.5 GT/s
  - Up to 16GB Capacity (4Gb dev)
  - Clamshell Supported

- PCIe Gen2 x16 (EP)
  - Up to 14 GB/s w/ 256B Packets
  - Support for P2P transactions

- Reliability Features
  - Parity on L1 caches
  - ECC on L2 caches
  - ECC on Memory
  - CRC on Memory IO
  - Command-Address Parity on Memory IO

In the #1 System on the Top 500 (June '13)

Milky Way -2

3,120,000 Core Supercomputer

Source: www.top500.org, June 2013

#	Site	Manufacturer	Computer	Country	Cores	Rmax [Pflops]	Power [MW]
1	National University of Defense Technology	NUDT	Tianhe-2 NUDT TH-IVB-FEP, Xeon 12C 2.2GHz, IntelXeon Phi	China	3,120,000	33.9	17.8
2	Oak Ridge National Laboratory	Cray	Titan Cray XK7, Opteron 16C 2.2GHz, Gemini, NVIDIA K20x	USA	560,640	17.6	8.21
3	Lawrence Livermore National Laboratory	IBM	Sequoia BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	1,572,864	17.2	7.89
4	RIKEN Advanced Institute for Computational Science	Fujitsu	K Computer SPARC64 VIIIfx 2.0GHz, Tofu Interconnect	Japan	795,024	10.5	12.7
5	Argonne National Laboratory	IBM	Mira BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	786,432	8.59	3.95
6	Texas Advanced Computing Center/UT	Dell	Stampede PowerEdge C8220, Xeon E5 8C 2.7GHz, Intel Xeon Phi	USA	462,462	5.17	4.51
7	Forschungszentrum Juelich (FZJ)	IBM	JuQUEEN BlueGene/Q, Power BQC 16C 1.6GHz, Custom	Germany	458,752	5.01	2.30
8	Lawrence Livermore National Laboratory	IBM	Vulcan BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	393,216	4.29	1.97
9	Leibniz Rechenzentrum	IBM	SuperMUC iDataPlex DX360M4, Xeon E5 8C 2.7GHz, Infiniband FDR	Germany	147,456	2.90	3.52
10	National SuperComputer Center in Tianjin	NUDT	Tianhe-1A NUDT TH MPP, Xeon 6C, NVidia, FT-1000 8C	China	186,368	2.57	4.04

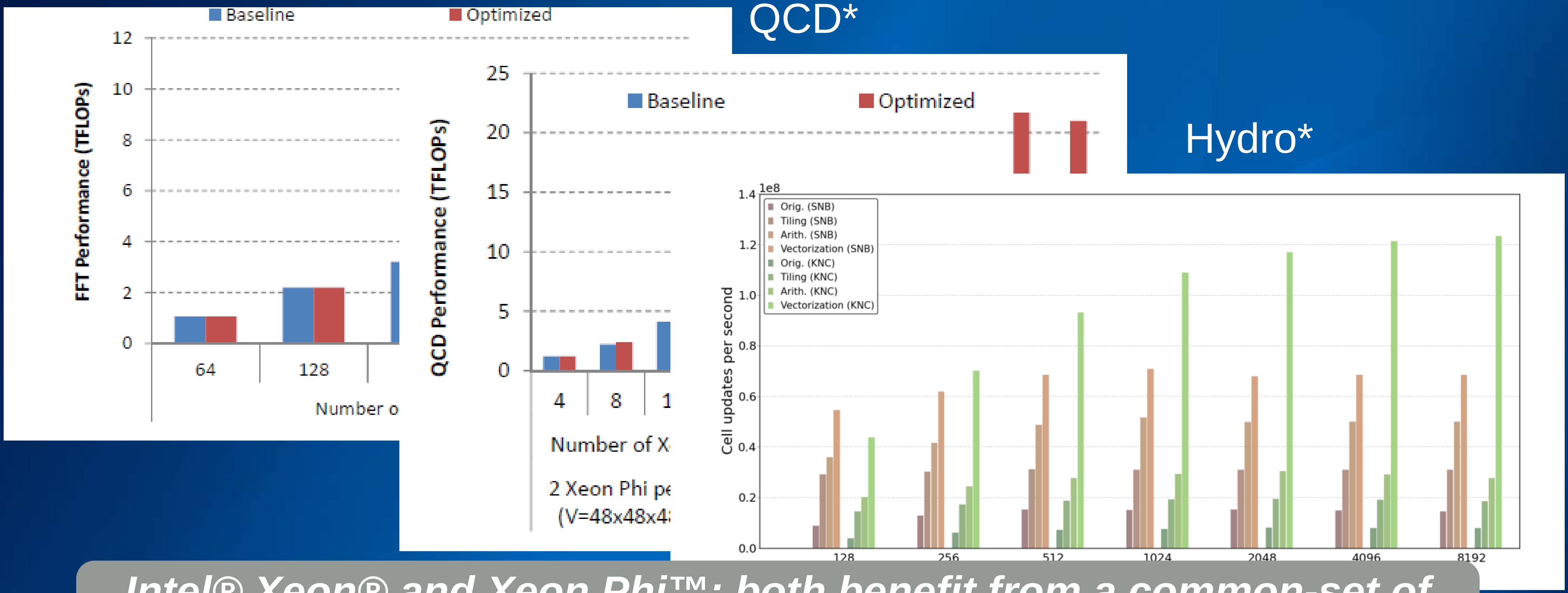
- Developed by Inspur & National University of Defense Technology
- 32000 IVT CPU's and 48000 KNC Cards
- System Peak Performance = 55 PF (33.9 PF delivered)
- Total Memory = 1.4PB, Total Disk Space = 12.4PB
- System Power = 17.8 MW

KNC Results

FFT*

QCD*

Hydro*

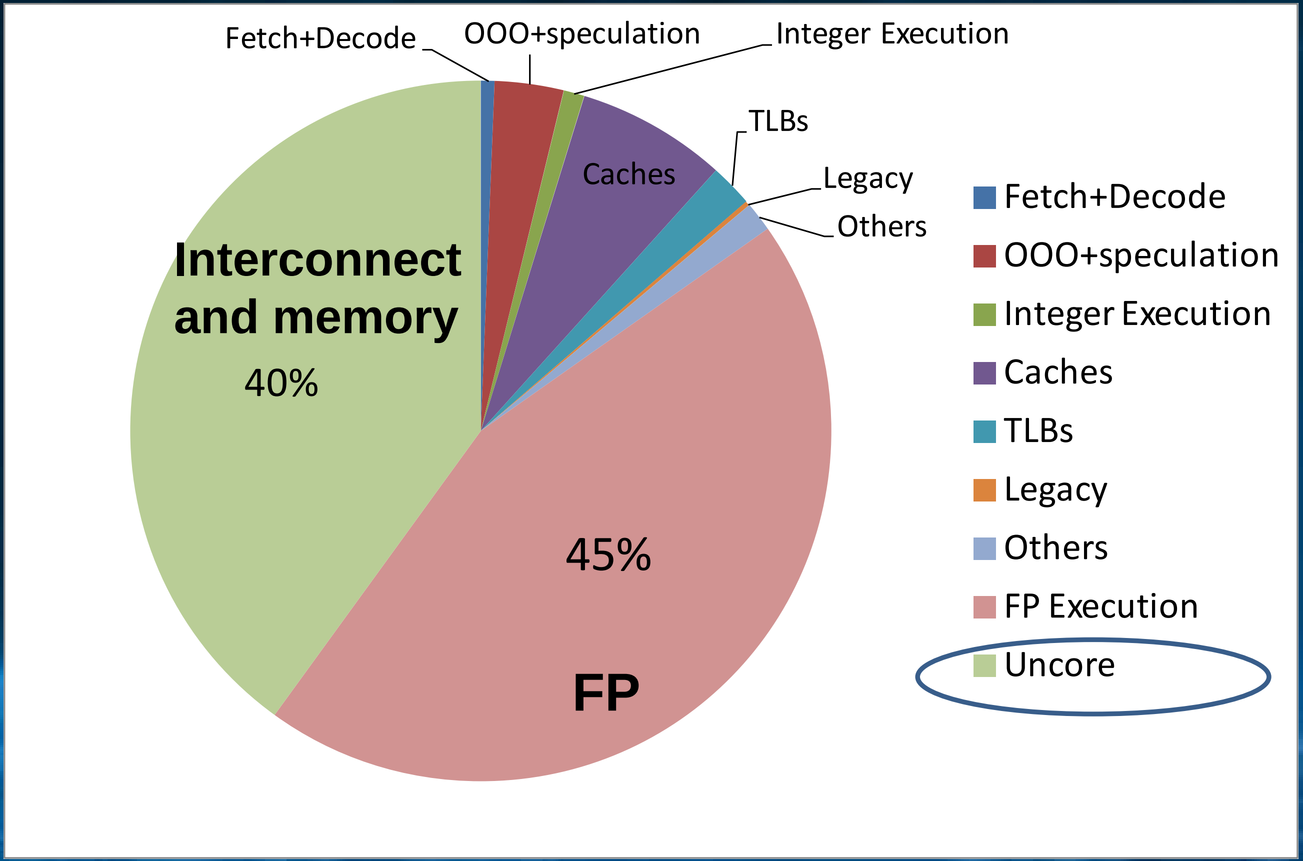


Intel® Xeon® and Xeon Phi™: both benefit from a common-set of manycore-friendly optimizations

* See Backup slides 24-26 for benchmark and system details

Optimization Notice

Chip-level Power Distribution



At chip level core power is even smaller portion (~15%).

X86 support, OOO, TLBs ~6% of the chip power

Benefits outweigh the gains from removing them

Next Intel® Xeon Phi™ Processor: Knights Landing

Designed using Intel's cutting-edge 14nm process

Not bound by "offloading" bottlenecks

Standalone CPU or PCIe coprocessor

Leadership compute & memory bandwidth

Integrated on-package memory

All products, computer systems, dates and figures specified are preliminary based on current expectations, and are subject to change without notice.

Intel® Advanced Vector Extensions 512 (Intel® AVX-512)

- 512-bit SIMD Instructions
- 8x 64-bit F.P./INT or 16x 32-bit F.P./INT per 512-bit register
- 32x 512-bit registers (ZMM0-ZMM31)
- 8x mask registers
- First implemented in the future Intel® Xeon Phi™ processor and coprocessor known by the code name Knights Landing
- Also supported by some future Intel® Xeon® processors

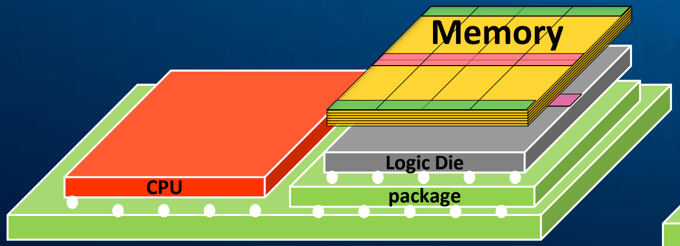
For testing the Intel® Software Development Emulator (Intel® SDE) has been extended for Intel AVX-512 and is available at http://www.intel.com/software/sde.

Memory going forward

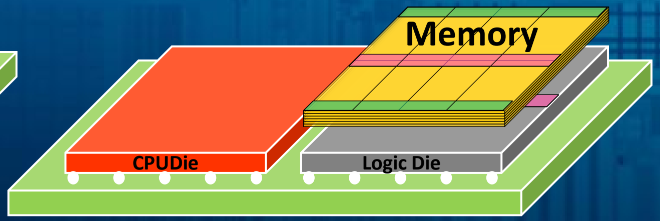
- Significant power consumed in Memory
  - Need to drive 20 pj/bit to 2-3 pJ/bit
- Balancing BW, capacity and power is hard problem

- More hierarchical memories
- Progressively more integration

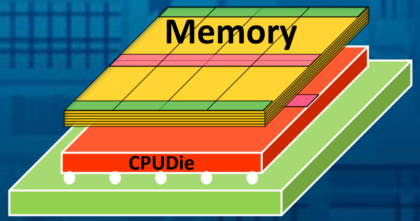
Multi-package Usage



Multi-chip Package Usage

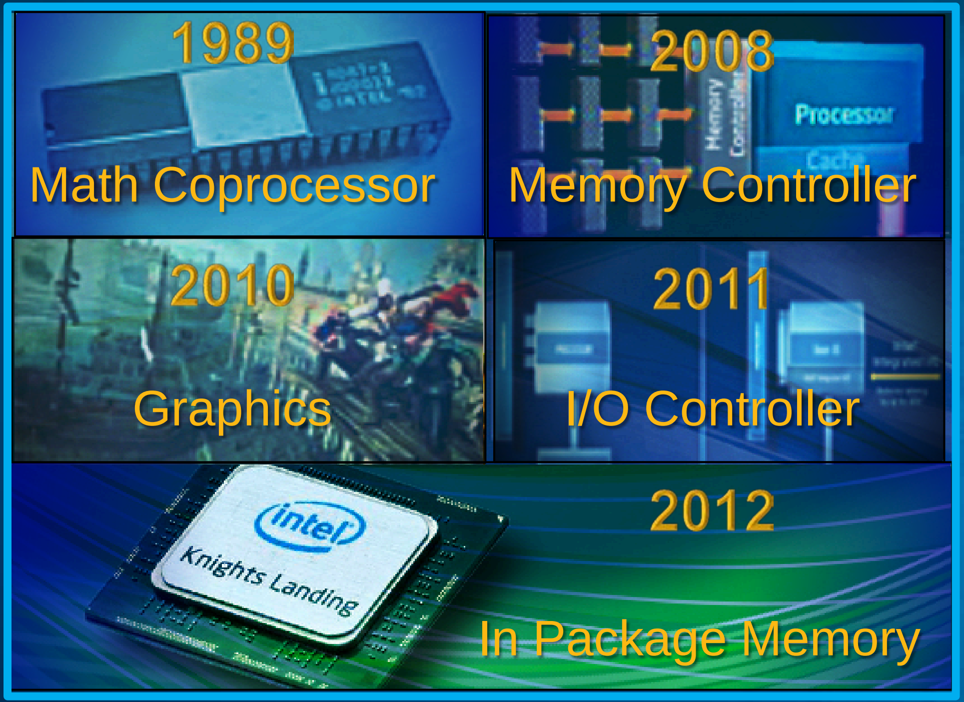


Direct Attach Usage

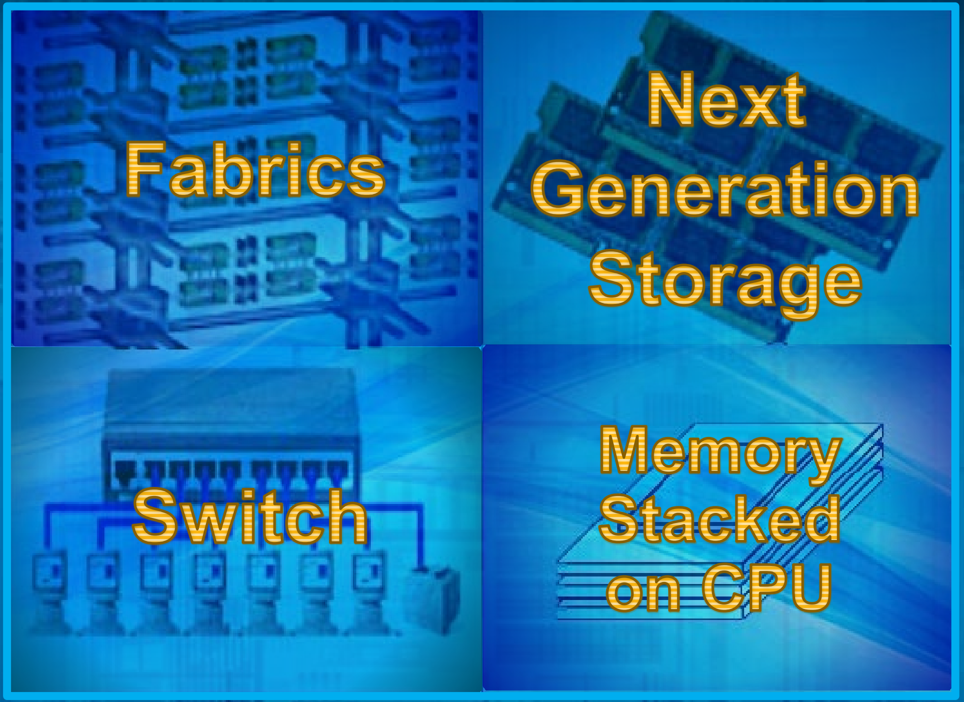


More integration

Unprecedented Innovations Only Enabled by the Leading Edge Process Technology



Integrated Today



Coming Tomorrow

SYSTEM LEVEL BENEFITS IN COST, POWER, DENSITY, SCALABILITY, & PERFORMANCE

Remaining questions going forward

- Math libraries and h/w accelerators
  - H/W acceleration 10-100X more efficient than general programmable cores/GPU’s
  - Usually hard to predict upfront which functions if hardened will be impactful across many applications
  - Basic math libraries are probably worth looking at. Need a solution for both compute as well as data delivery
- Data parallelism and vector processing
  - Vector processing highly efficient (PPA)
  - Still hard to program, create compilers
  - Some problems will never parallelize
  - Several potential directions going forward
- Memory architecture
  - Need high capacity, high bandwidth
  - Certain apps need low latency
  - Need low power
  - Caches vs. scratch pad
- Network architecture

Summary

- Intel leading with HPC solutions around Xeon/Xeon-Phi
- KNC meeting expectation and showing encouraging results
- Knights Landing and subsequent generations under development
- Key system level questions will require cross-industry efforts

intel®