Technologies and application performance

Marc Mendez-Bermond

HPC Solutions Expert - Dell Technologies

September 2017

The landscape is changing

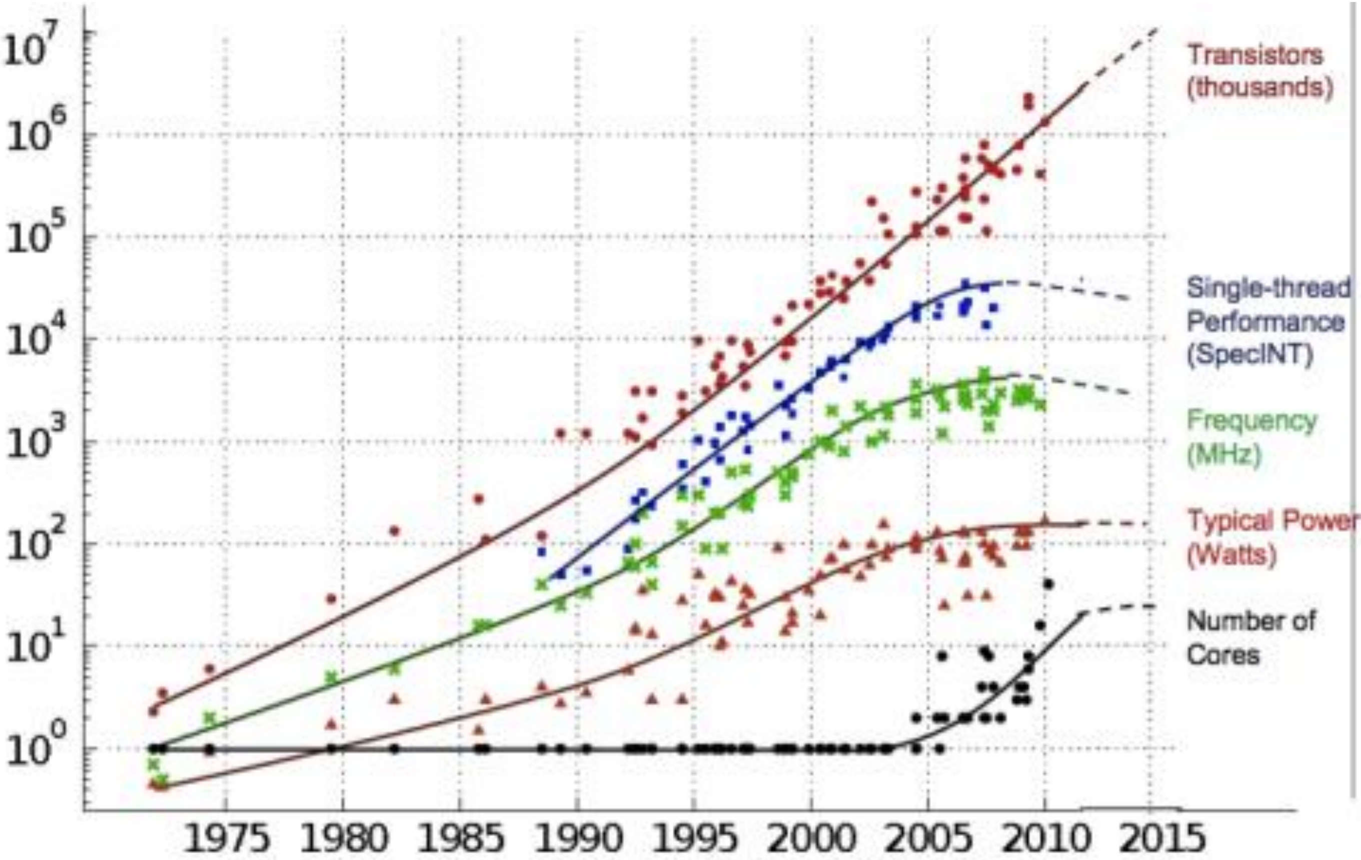
“We are no longer in the general purpose era… the argument of tuning software for hardware is moot. Now, to get the best bang for the buck, you have to tune both.”

- Kushagra Vaid, general manager of server engineering, Microsoft Cloud Solutions

https://www.nextplatform.com/2017/03/08/arm-amd-x86-server-chips-get-mainstream-lift-microsoft/amp/

Moore’s Law (Technology)

- The clock speed plateau
- The power ceiling
- IPC limit

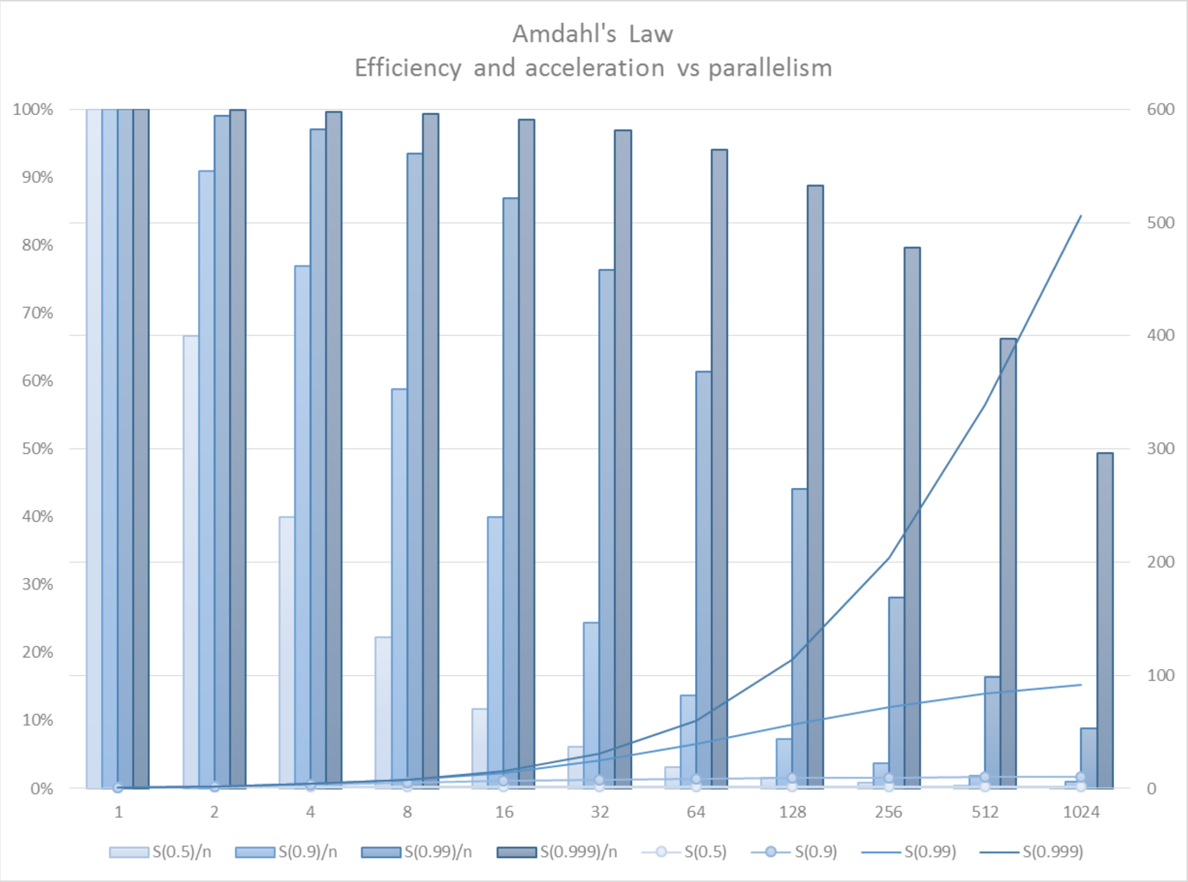


Chuck Moore, "DATA PROCESSING IN EXASCALE-CLASS COMPUTER SYSTEMS", The Salishan Conference on High Speed Computing, 2011

Amdahl’s Law (Application)

- Amdahl’s law predicts performance from your app parallelization
- 50% : x2 max
- 99% : x100 max
- 99.9% : x1000 max

- But you should also check the efficiency here :
  - 99.9% parallel, at 1024 processors, x509 and efficiency at 49% …



WARNING : Legacy Slide from back in 2014 (ENS/PSMN cluster inauguration)

Still valid conclusions !!!

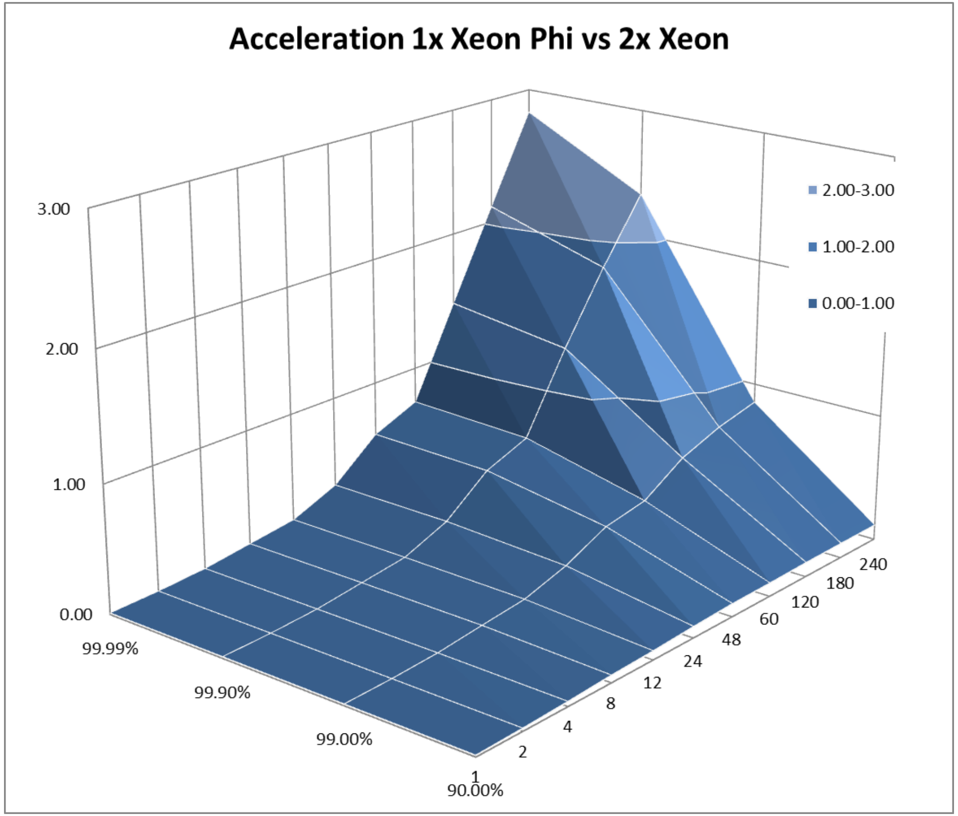
Computing technologies

Intel Xeon Phi : a few considerations

- x86_64 programming models
- Cache coherency
  - Dual-ring interconnect
  - 8 (soon 16) GB RAM
- Right to the point cores
  - No « out of order» execution
  - No branch prediction
  - 4 Hyper-threads per core
  - Wide vectors (16 op/c/core)
- PCIe connectivity to host

App should fit in onboard memory,
Parallelism > 99.9%,
Vectorization > 95%

Xeon Phi : 60 cores @ 1 GHz vs 2 Xeon : 8 cores @ 2.6 GHz



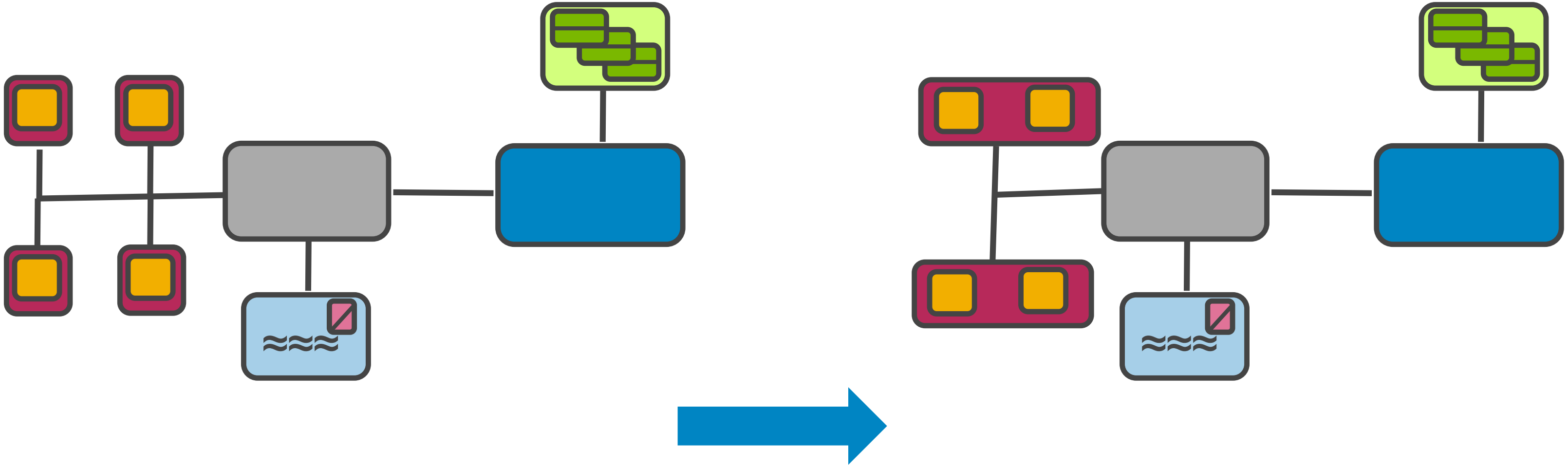
]

Moore’s Law vs Amdahl's Law - “too Many Cooks in the Kitchen”

Industry is applying Moore’s Law by adding more cores

Meanwhile Amdahl’s Law says that you cannot use them all efficiently

System trend over the years (1)

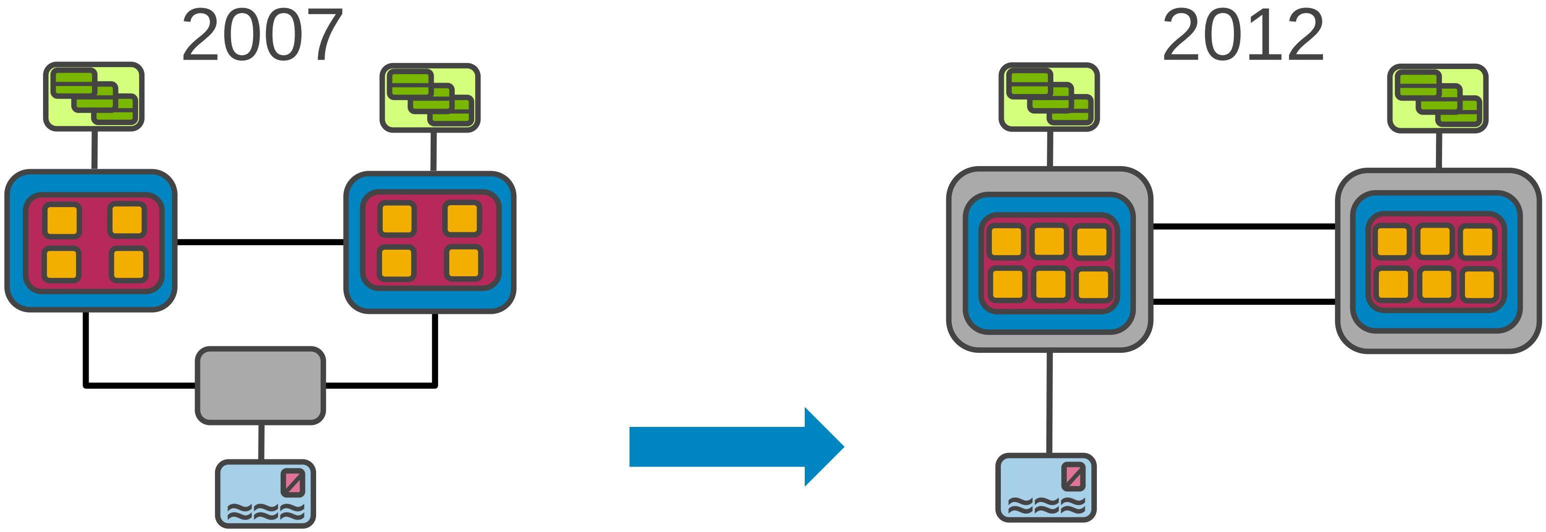


~1970 - 2000

2005

Multi-core: TOCK

System trend over the years (2)



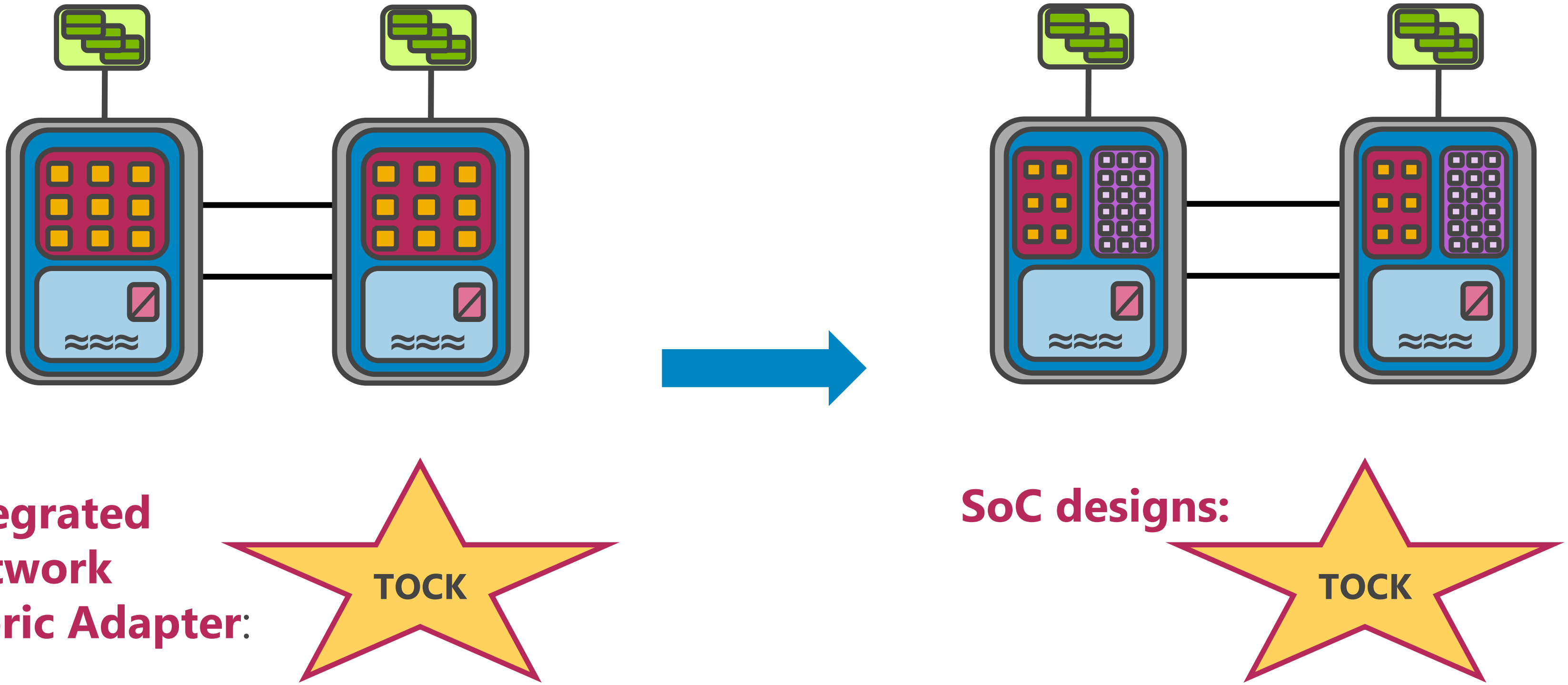
Integrated Memory controller:

TOCK

Integrated PCIe controller:

TOCK

Future



Integrated Network Fabric Adapter:

SoC designs:

Improving performance - what levels do we have?

- Challenge: Sustain performance trajectory without massive increases in cost, power, real estate, and unreliability
- Solutions: No single answer, must **intelligently turn** “Architectural Knobs”

$$\underbrace{\overset{1}{(Freq)} \times \overset{2}{\left(\frac{cores}{socket}\right)} \times \overset{3}{(\#sockets)} \times \overset{4}{\left(\frac{inst\ or\ ops}{core\ \times\ clock}\right)}}_{\text{Hardware performance}} \times \underbrace{\overset{5}{(Efficiency)}}_{\substack{\text{What you really get} \\ \text{Software performance}}}$$

Turning the knobs 1 - 4

1 Frequency is unlikely to change much - Thermal/Power/Leakage challenges

2 Moore’s Law still holds: 130 -> 14 nm - LOTS of transistors

3 Number of sockets per system is the easiest knob.

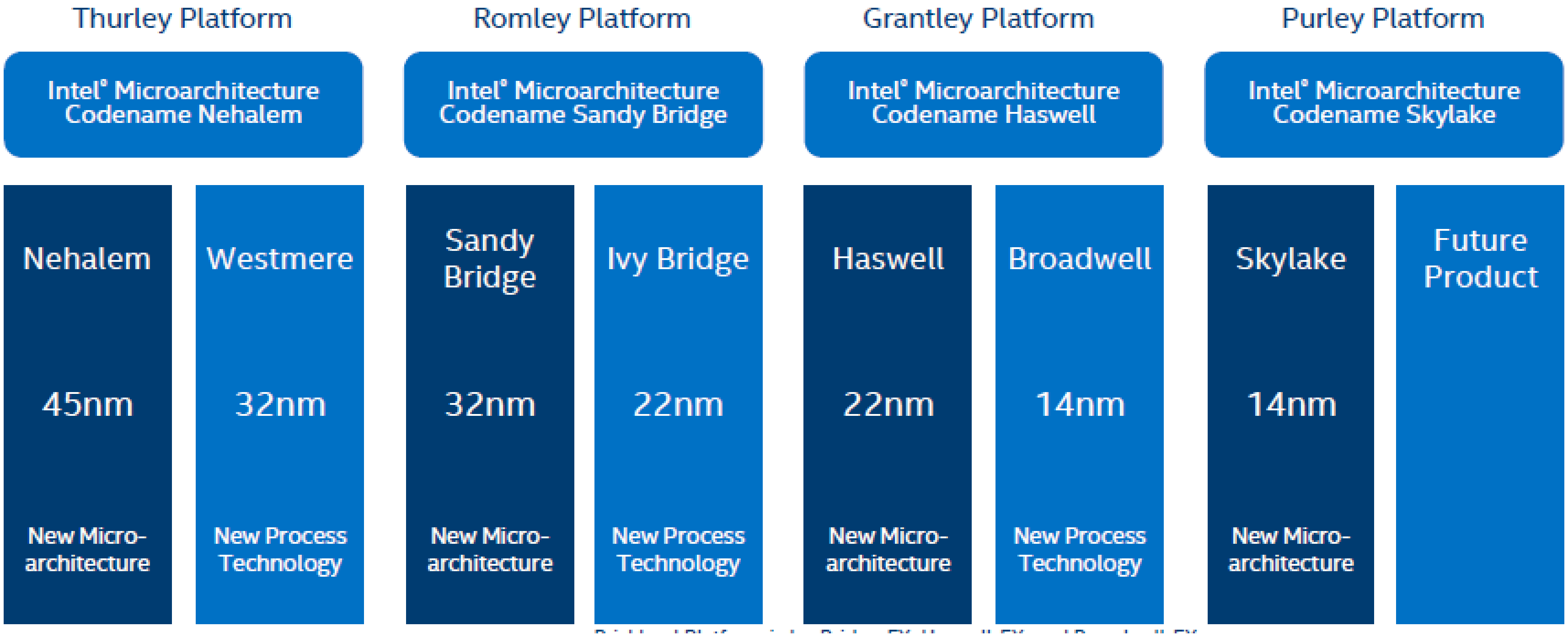
Challenging for power/density/cooling/networking

4 IPC still grows

FMA3/4, AVX, FPGA implementations for algorithms

Challenging for the user/developer

New capabilities according to Intel



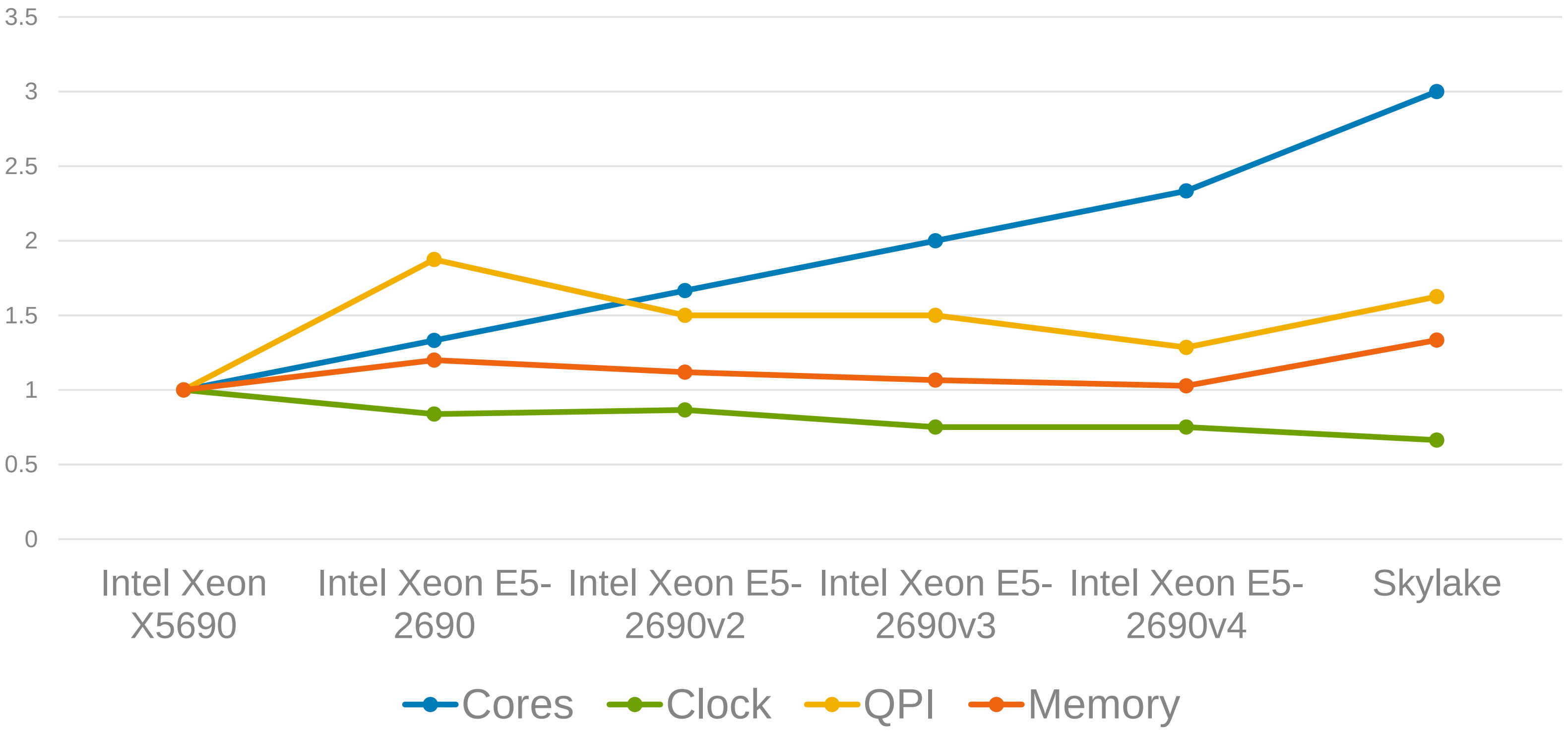
Thurley Platform		Romley Platform		Grantley Platform		Purley Platform	
Intel® Microarchitecture Codename Nehalem		Intel® Microarchitecture Codename Sandy Bridge		Intel® Microarchitecture Codename Haswell		Intel® Microarchitecture Codename Skylake	
Nehalem	Westmere	Sandy Bridge	Ivy Bridge	Haswell	Broadwell	Skylake	Future Product
45nm	32nm	32nm	22nm	22nm	14nm	14nm	
New Micro-architecture	New Process Technology	New Micro-architecture	New Process Technology	New Micro-architecture	New Process Technology	New Micro-architecture	
SSSE3	*SSE4*	*AVX*	*AVX*	*AVX2*	*AVX2*	*AVX-512*	
2007	*2009*	*2012*	*2013*	*2014*	*2015*	*2017*	

The state of ISV software

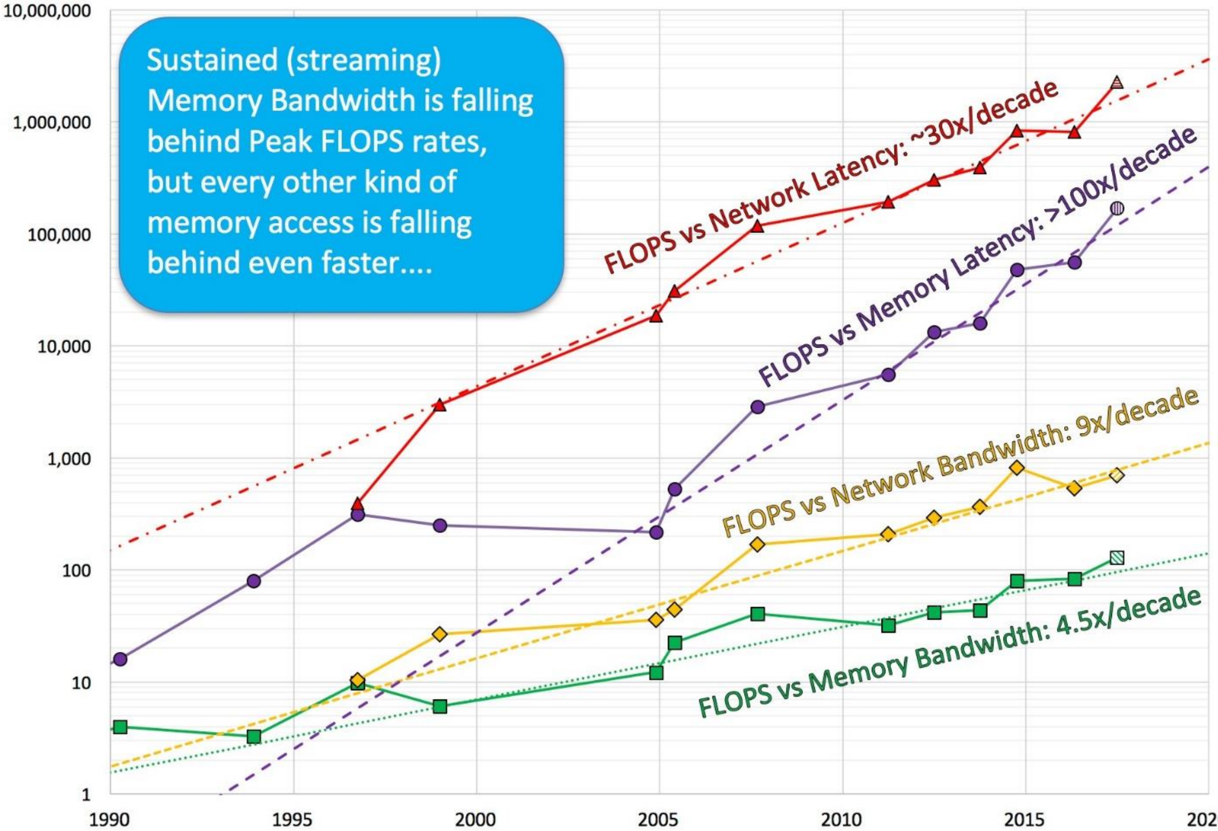
Segment	Applications	Vectorization support
CFD	Fluent, LS-DYNA, STAR CCM+	Limited SSE2 support
CSM	CFX, RADIOSS, Abaqus	Limited SSE2 support
Weather	WRF, UM, NEMO, CAM	Yes
Oil and Gas	Seismic processing	Not applicable
	Reservoir Simulation	Yes
Chemistry	Gaussian, GAMESS, Molpro	Not applicable
Molecular dynamics	NAMD, GROMACS, Amber,…	PME kernels support SSE2
Biology	BLAST, Smith-Waterman	Not applicable
Molecular mechanics	CPMD, VASP, CP2k, CASTEP	Yes

Bottom line: ISV support for new instructions is poor. Less of an issue for in-house developed codes, but programming is hard

Meanwhile the bandwidth is suffering



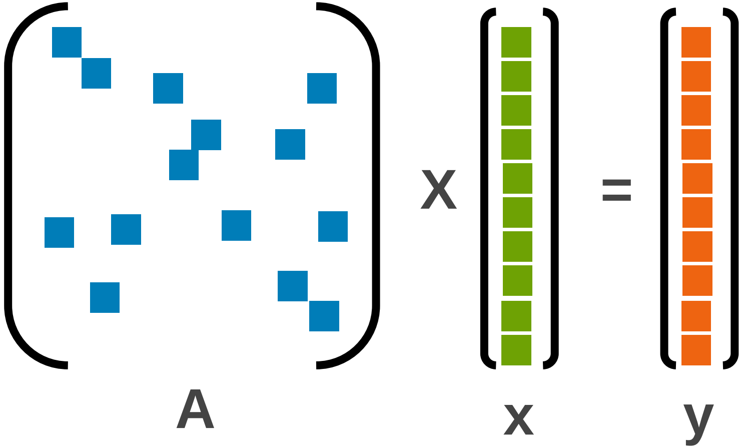
Add to this the Memory Bandwidth and System Balance



Sustained (streaming) Memory Bandwidth is falling behind Peak FLOPS rates, but every other kind of memory access is falling behind even faster....

Obtained from: http://sc16.supercomputing.org/2016/10/07/sc16-invited-talk-spotlight-dr-john-d-mccalpin-presents-memory-bandwidth-system-balance-hpc-systems/

And data is becoming sparser (think “Big Data”)

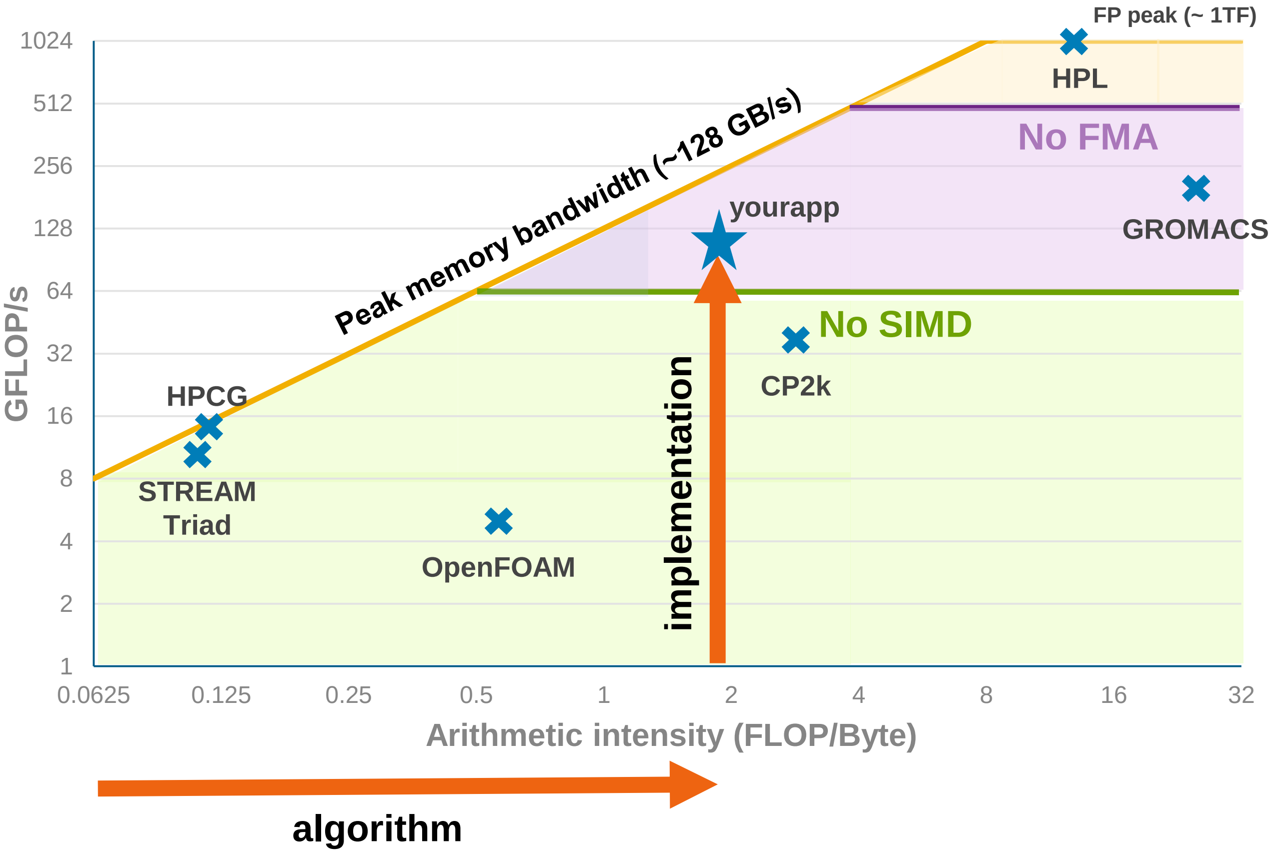


Sparse Matrix “A”

- Most entries are zero
- Hard to exploit SIMD
- Hard to use caches

- This has very low arithmetic density and hence memory bound
- Common in CFD, but also in genetic evaluation of species

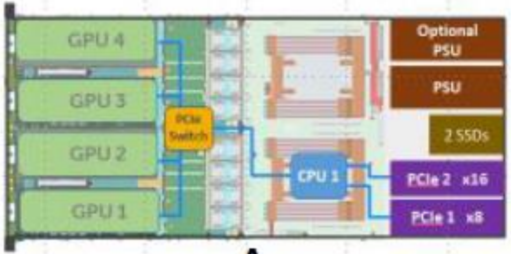
Xeon roofline model (v4)



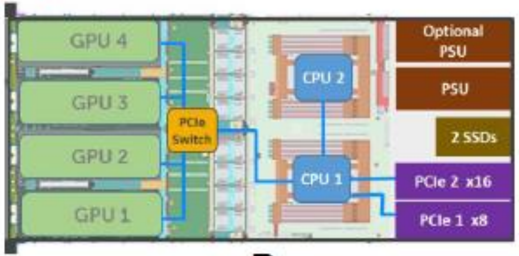
What does Intel do about these trends?

Problem	Westmere	Sandy Bridge	Ivy Bridge	Haswell	Broadwell	Skylake
QPI bandwidth	No problem	Even better	Two snoop modes	Three snoop modes	Four (!) snoop modes	• UPI • COD snoop modes
Memory bandwidth	No problem	Extra memory channel	Larger cache	Extra load/store units	Larger cache	• Extra load/store units • +50% memory channels
Core frequency	No problem	• More cores • AVX • Better Turbo	• Even more cores • Above TDP Turbo	• Still more cores • AVX2 • Per-core Turbo	• Again even more cores • optimized FMA • Per-core Turbo based on instruction type	• More cores • Larger OOO engine • AVX-512 • 3 different core frequency modes

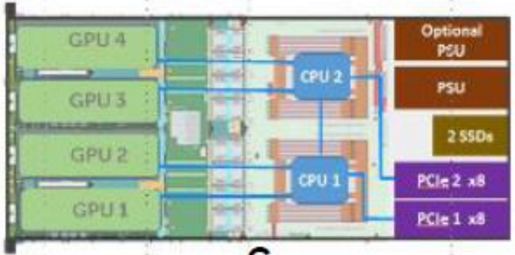
C4130 – Ten supported variations



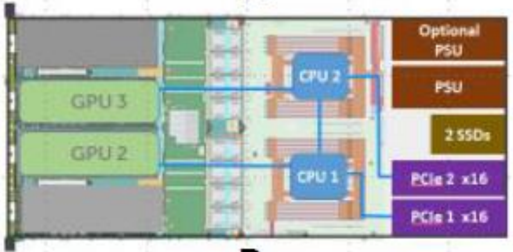
A



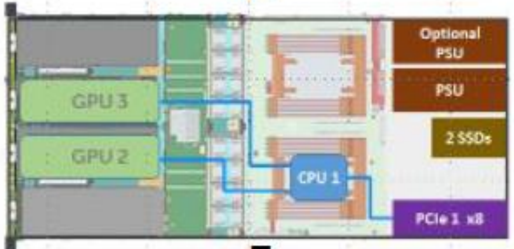
B



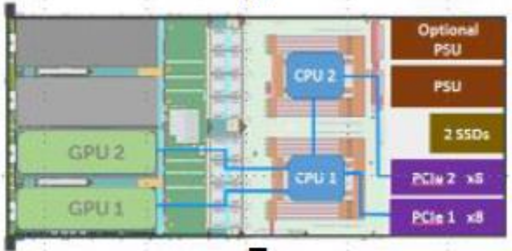
C



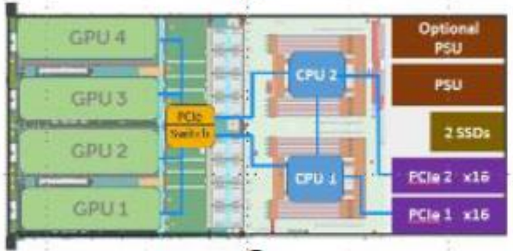
D



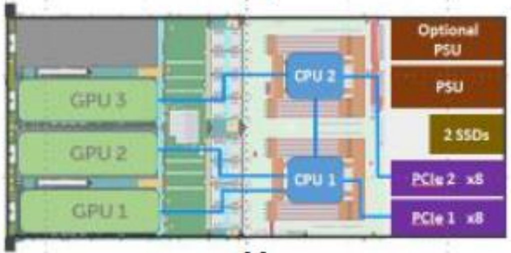
E



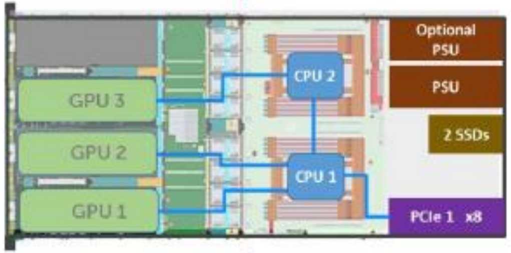
F



G



H



I

Pragmatic computing

Parallelize

Take advantage of multicore

Vectorize

Take advantage of large-vector units

Amdahl’s law : limiting factor

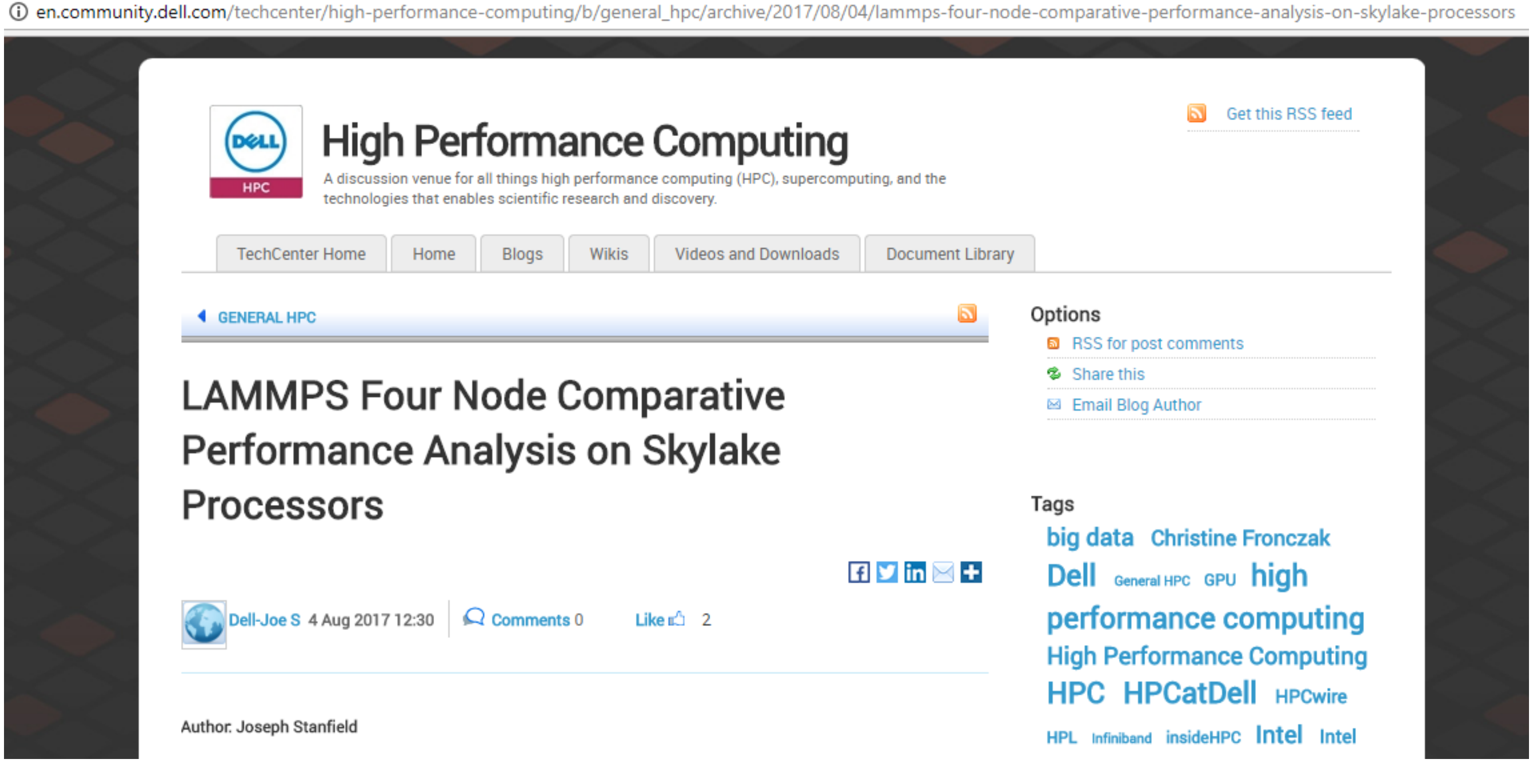
Moore’s law : benefiting factor

Optimize

- Intrinsic optimization
- Execution optimization

Efficiency of implementation

Public benchmark data



en.community.dell.com/techcenter/high-performance-computing/b/general_hpc/archive/2017/08/04/lammps-four-node-comparative-performance-analysis-on-skylake-processors

High Performance Computing

A discussion venue for all things high performance computing (HPC), supercomputing, and the technologies that enables scientific research and discovery.

LAMMPS Four Node Comparative Performance Analysis on Skylake Processors

Dell-Joe S 4 Aug 2017 12:30 | Comments 0 Like 2

Author: Joseph Stanfield

Portfolio: Ready Solutions for HPC

Benefits

BUILD

Maximum flexibility
Validated for use case
Heterogeneity with lower risk
Component lifecycle automation and control

Consumption models

BUY

Fastest time to value
Optimized and tuned for use case
Greatest risk reduction
Solution lifecycle automation

Solutions

Scale

STORAGE READY BUNDLES

Dell EMC Ready Bundle for HPC NFS Storage

Scales from a minimum of 48TB to 480TB of raw capacity in a single name space

Dell EMC Ready Bundle for HPC Lustre Storage

Lustre parallel file storage system scales from 120TB to petabytes of data

SYSTEMS FOR A RANGE OF USE CASES

Dell EMC HPC System for Life Sciences

Fully integrated for pharma/biotech applications

Dell EMC HPC System for Manufacturing

Fully integrated for compute-aided engineering (CAE) workloads

Dell EMC HPC System for Research

General purpose compute cluster for multiple research workloads

DELL EMC

HPC Innovation Lab World-Class Infrastructure

Dedication to Research and Development:

- 13K sq. ft (1200m²) with 1300+ Servers and ~10PB
- Leverage Expertise in HPC
- Test New Technologies
- Tune your applications for performance and efficiency

Zenith

- Top500 class system based on Intel Scalable Systems Framework (OPA, KNL, Xeon, OpenHPC)
- 256-nodes with dual 2697v4 processors, non-blocking OPA fabric and 270TFlops sustained performance

Rattler

- Research/development system in collaboration with Mellanox and NVIDIA
- 80 nodes configured with Infiniband EDR and 2660v3 processors

Merci !

marc_mendez_bermond@dell.com

DELL EMC