Master Informatics Eng.

2019/20

A.J.Proença

HPC systems: TOP500 lists

(some slides are borrowed)

Outline

1. TOP500
   a) TOP10 lists from Nov'17 to Nov'19
   b) Country distribution over the past 25 years
   c) PU chip technology evolution in the past 25 years and since last year
   d) Evolution of the accelerators since they were available
   e) Analysis of some relevant systems and architectures
2. GREEN500
   a) TOP10 lists from Nov'17 to Nov'19
   b) Analysis of some relevant systems
3. HPCG500
   a) HPCG *vs.* HPL: an overview
   b) TOP10 lists from Nov'17 to Nov'19
   c) Analysis of some relevant systems
4. GRAPH500
   a) Performance Metric (TEPS)
   b) Breadth-First Search (BFS) & Single Source Shortest Paths (SSSP)
5. And next?...

LINPACK benchmarks

From Wikipedia, the free encyclopedia
(Redirected from LINPACK benchmark)

For the software library, see LINPACK.

The **LINPACK Benchmarks** are a measure of a system's floating point computing power. Introduced by Jack Dongarra, they measure how fast a computer solves a dense *n* by *n* system of linear equations $Ax = b$, which is a common task in engineering.

The latest version of these benchmarks is used to build the TOP500 list, ranking the world's most powerful supercomputers.[1]

The aim is to approximate how fast a computer will perform when solving real problems. It is a simplification, since no single computational task can reflect the overall performance of a computer system. Nevertheless, the LINPACK benchmark performance can

LINPACK benchmarks

Original author(s)	Jack Dongarra, Jim Bunch, Cleve Moler, and Gilbert Stewart
Initial release	1979
Website	www.netlib.org /benchmark/hpl/

Top 10 HPC systems
Nov'17 TOP500

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
2	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P , NUDT National Super Computer Center in Guangzhou China	3,120,000	33,862.7	54,902.4	17,808
3	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland				
4	**Gyoukou** - ZettaScaler-2.2 HPC system, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 700Mhz , ExaScaler Japan Agency for Marine-Earth Science and Technology Japan				
5	**Titan** - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x , Cray Inc. DOE/SC/Oak Ridge National Laboratory United States				
6	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom , IBM DOE/NNSA/LLNL United States	1,572,864	17,173.2	20,132.7	7,890
7	**Trinity** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	979,968	14,137.3	43,902.6	3,844
8	**Cori** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/SC/LBNL/NERSC United States	622,336	14,014.7	27,880.7	3,939
9	**Oakforest-PACS** - PRIMERGY CX1640 M1, Intel Xeon Phi 7250 68C 1.4GHz, Intel Omni-Path , Fujitsu Joint Center for Advanced High Performance Computing Japan	556,104	13,554.6	24,913.5	2,719
10	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect , Fujitsu RIKEN Advanced Institute for Computational Science (AICS) Japan	705,024	10,510.0	11,280.4	12,660

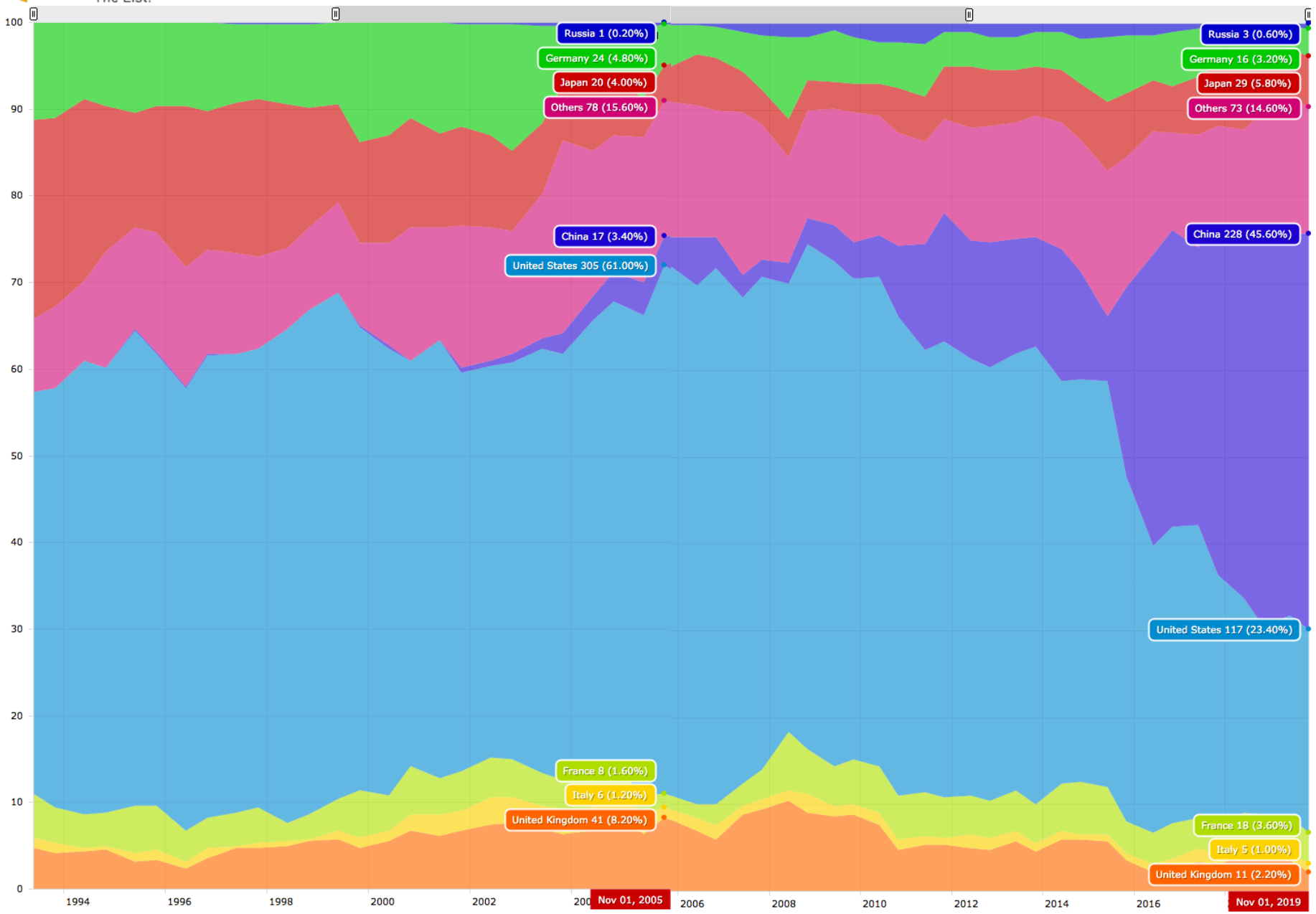
Top 10 HPC systems
Nov'18 TOP500

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory United States	2,397,824	143,500.0	200,794.9	9,783
2	**Sierra** - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438
3	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
4	**Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000 , NUDT National Super Computer Center in Guangzhou China	4,981,760	61,444.5	100,678.7	18,482
5	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	387,872	21,230.0	27,154.3	2,384
6	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	979,			
7	**AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2570 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR , Fujitsu National Institute of Advanced Industrial Science and Technology (AIST) Japan	391,			
8	**SuperMUC-NG** - ThinkSystem SD530, Xeon Platinum 8174 24C 3.1GHz, Intel Omni-Path , Lenovo Leibniz Rechenzentrum Germany	305,856	19,476.6	26,873.9	
9	**Titan** - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x , Cray Inc. DOE/SC/Oak Ridge National Laboratory United States	560,640	17,590.0	27,112.5	8,209
10	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom , IBM DOE/NNSA/LLNL United States	1,572,864	17,173.2	20,132.7	7,890

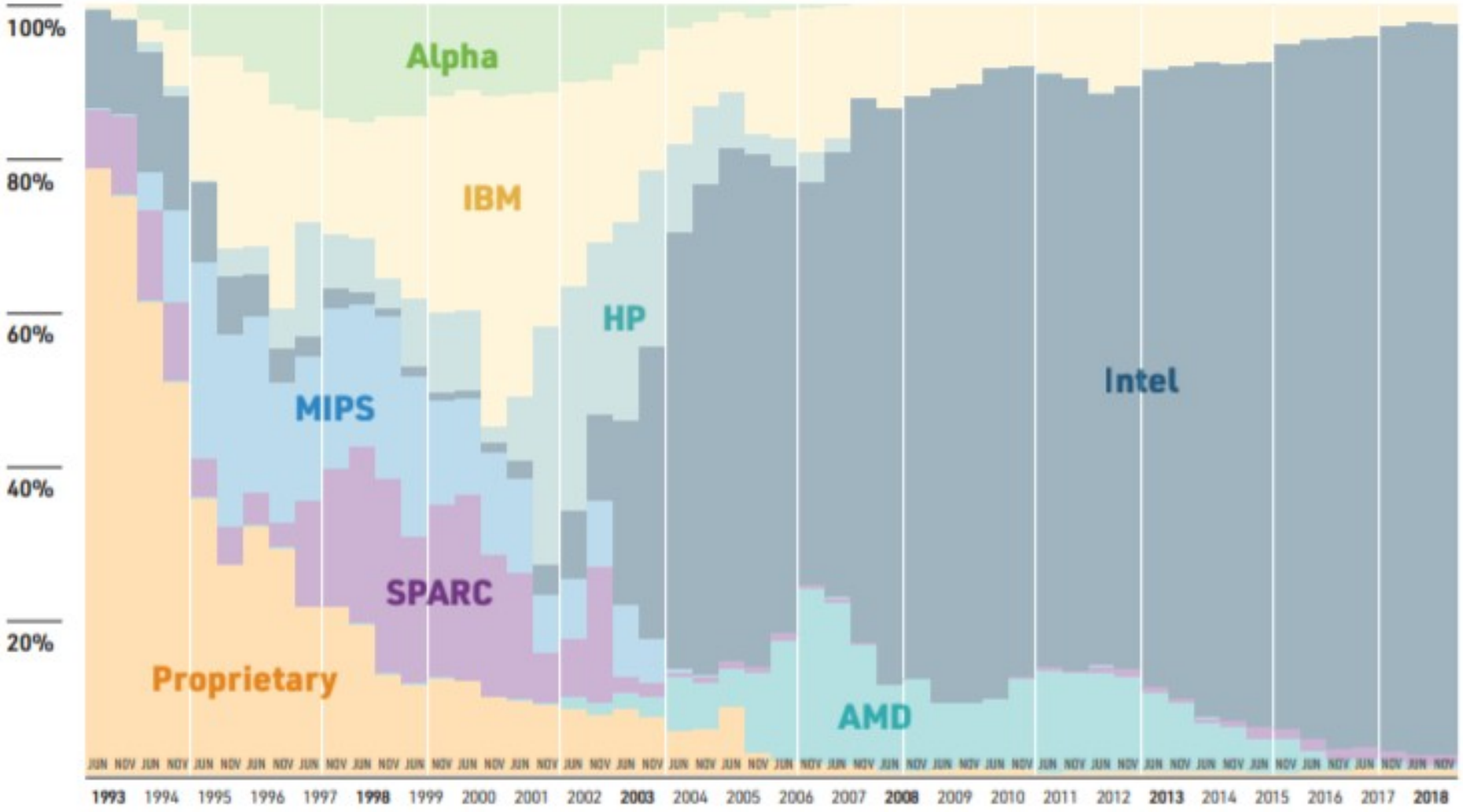
Top 10 HPC systems
Nov'19 TOP500

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	200,794.9	10,096
2	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438
3	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
4	**Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000 , NUDT National Super Computer Center in Guangzhou China	4,981,760	61,444.5	100,678.7	18,482
5	**Frontera** - Dell C6420, Xeon Platinum 8280 28C 2.7GHz, Mellanox InfiniBand HDR , Dell EMC Texas Advanced Computing Center/Univ. of Texas United States	448,448	23,516.4	38,745.9	
6	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray/HPE Swiss National Supercomputing Centre (CSCS) Switzerland	387,872			
7	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray/HPE DOE/NNSA/LANL/SNL United States	979,072			
8	**AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2570 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR , Fujitsu National Institute of Advanced Industrial Science and Technology (AIST) Japan	391,680	19,880.0	32,576.6	1,649
9	**SuperMUC-NG** - ThinkSystem SD650, Xeon Platinum 8174 24C 3.1GHz, Intel Omni-Path , Lenovo Leibniz Rechenzentrum Germany	305,856	19,476.6	26,873.9	
10	**Lassen** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Tesla V100 , IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	288,288	18,200.0	23,047.2	

Country distribution over the past 25 years



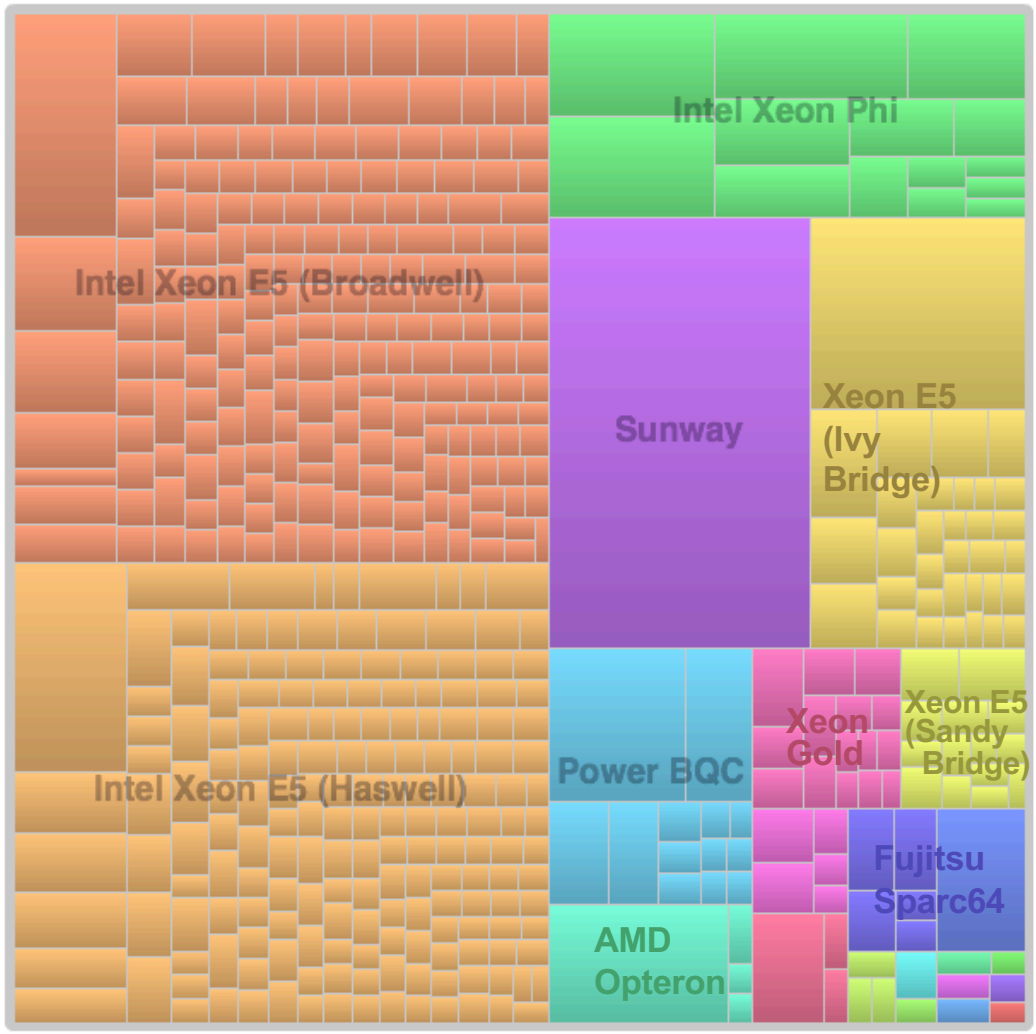
Chip technology in the past 25 years



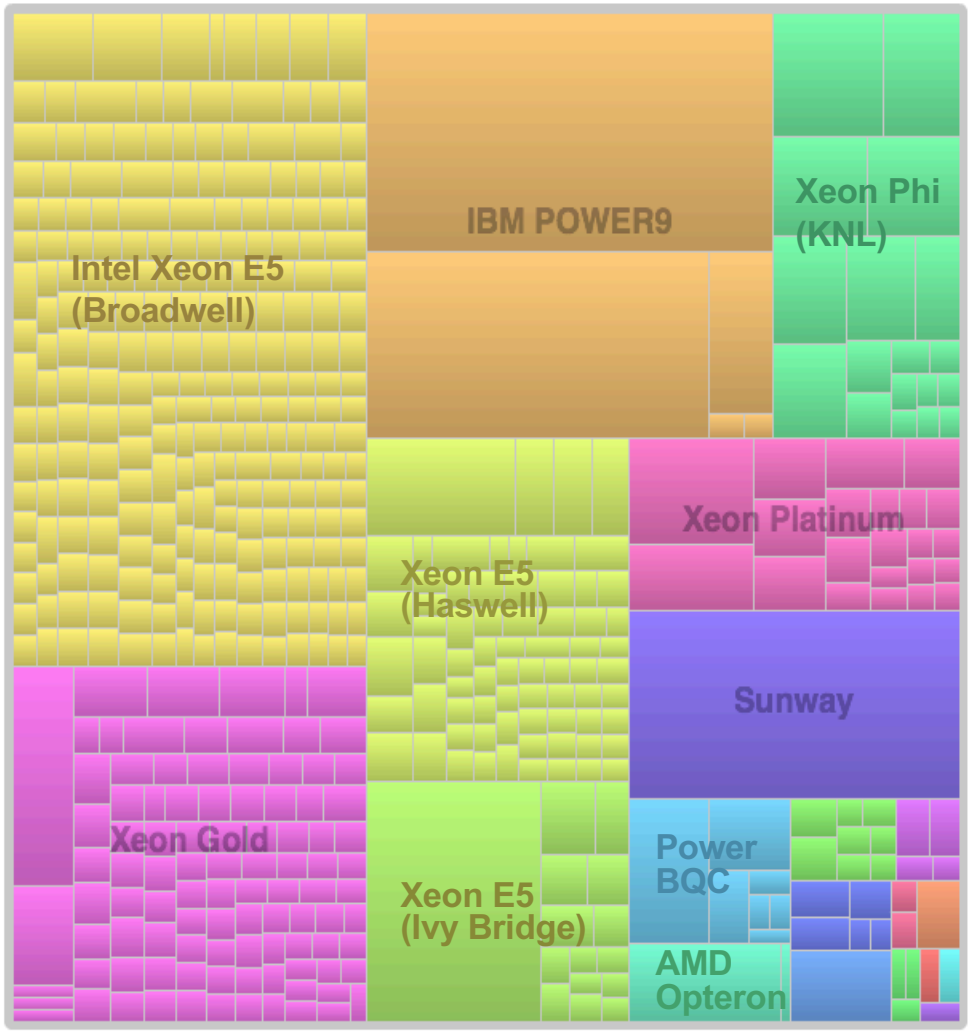
https://www.nextplatform.com/2018/11/12/the-widening-gyre-of-supercomputing/

Processor generations
Nov'17 & Nov'18

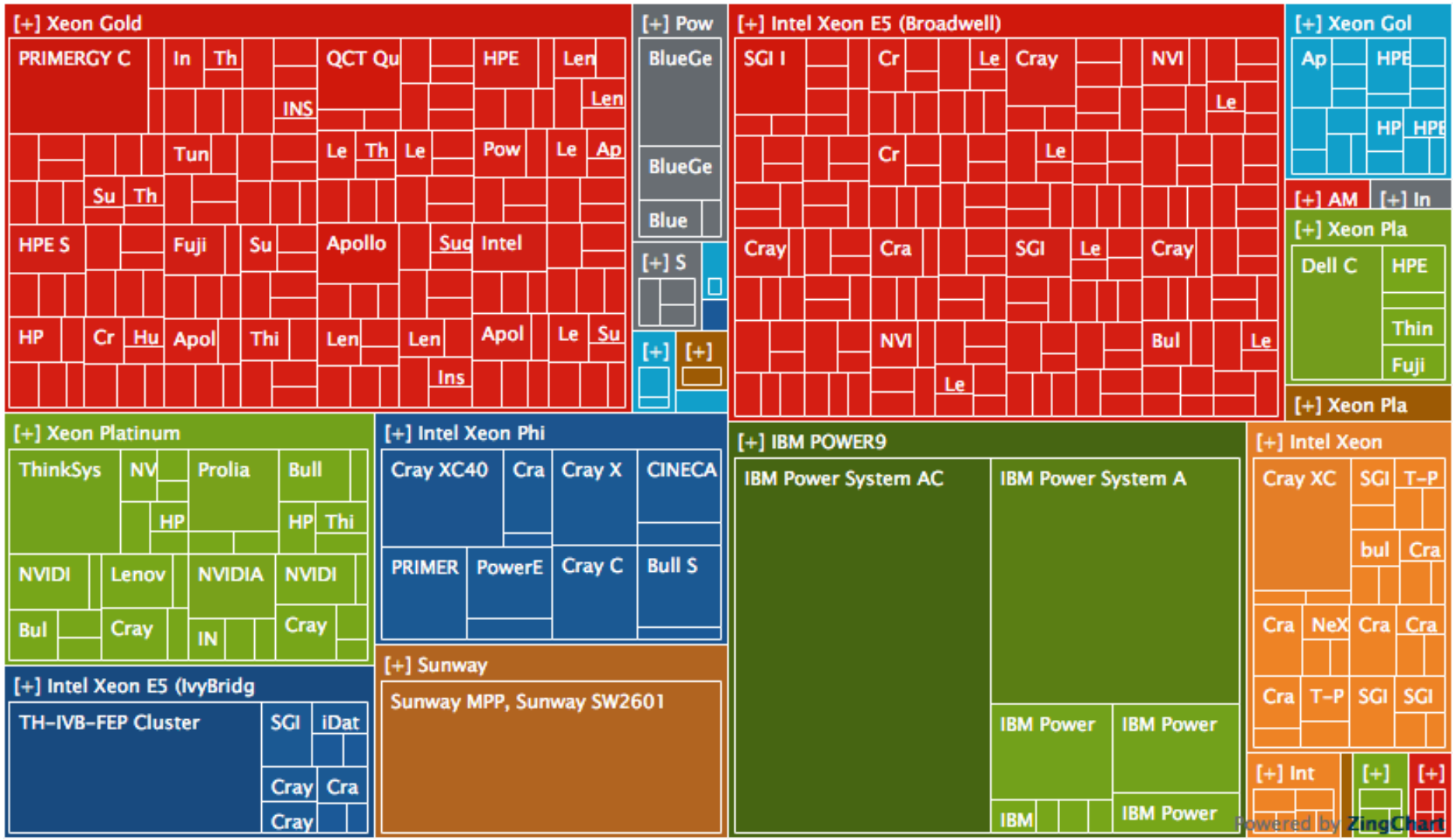
Nov'17



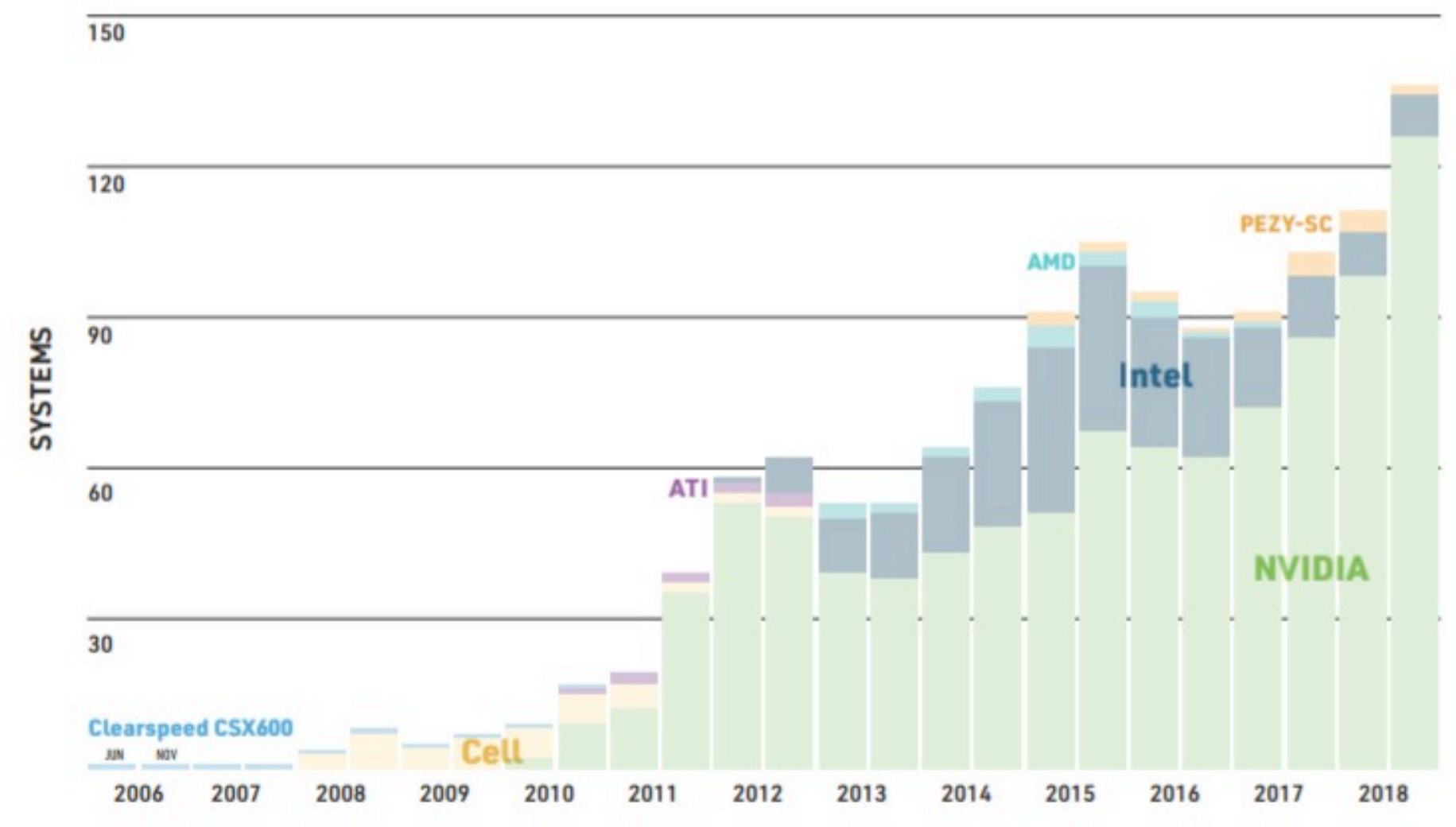
Nov'18



Processor generations
Nov'19

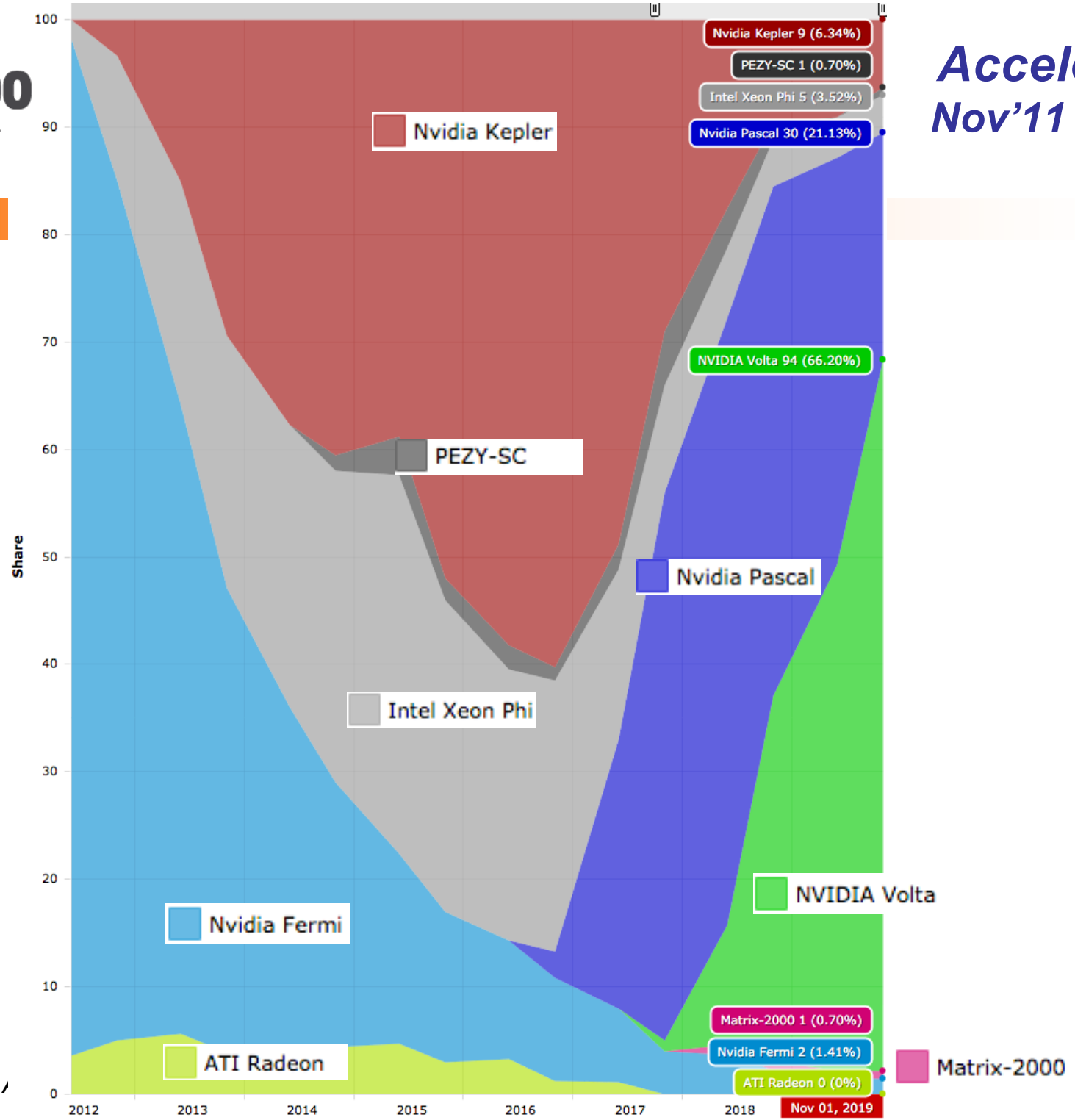


Accelerator families in the past 25 years



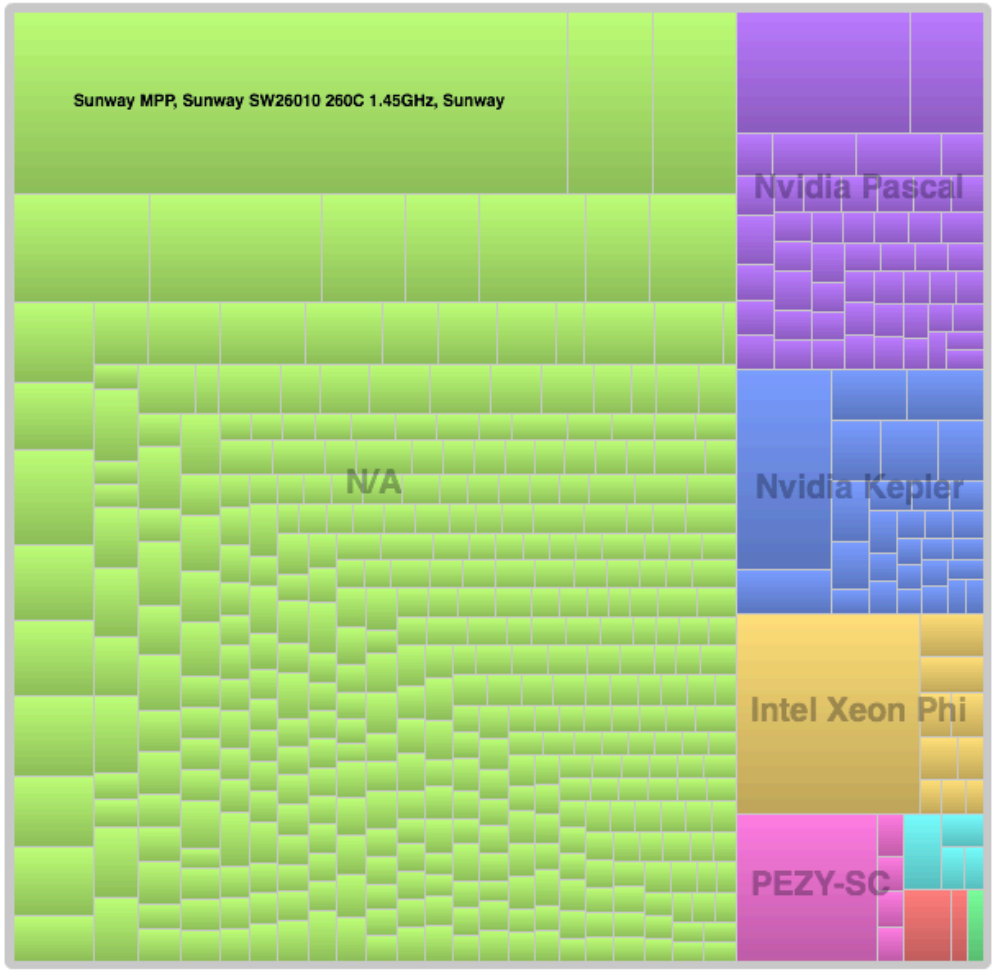
https://www.nextplatform.com/2018/11/12/the-widening-gyre-of-supercomputing/

Accelerators
Nov'11 - Nov'19

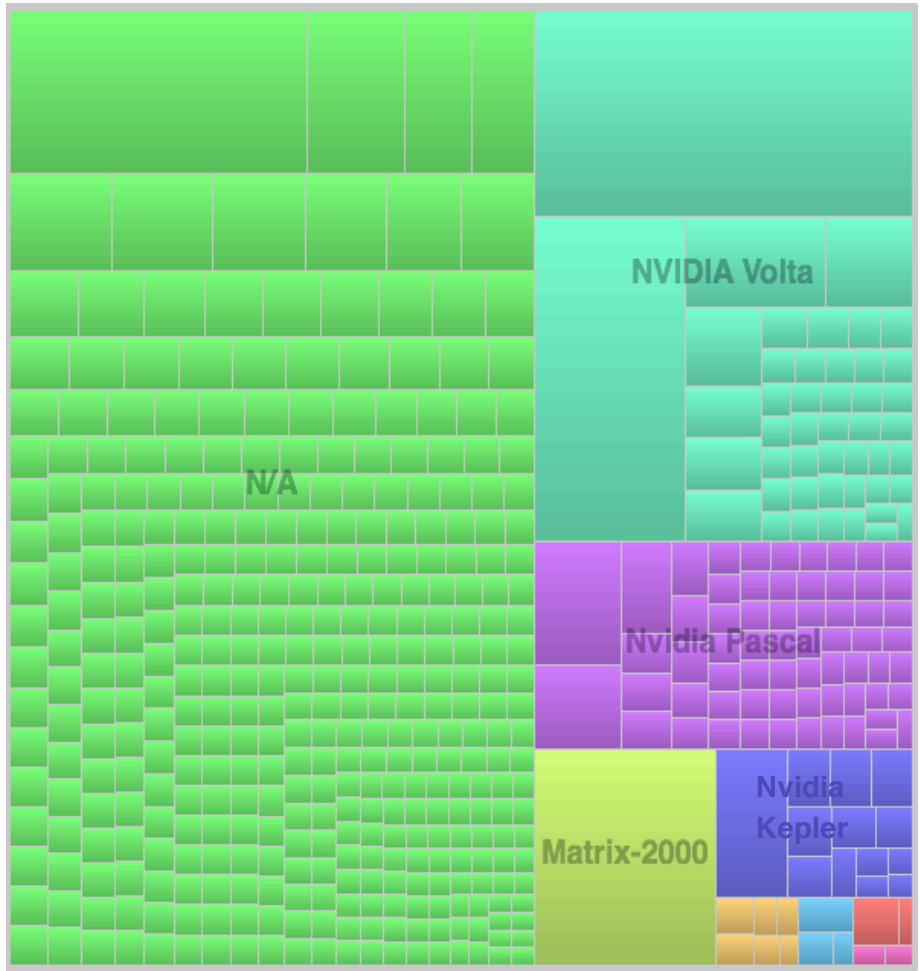


Accelerator family distribution
Nov'17 & Nov'18

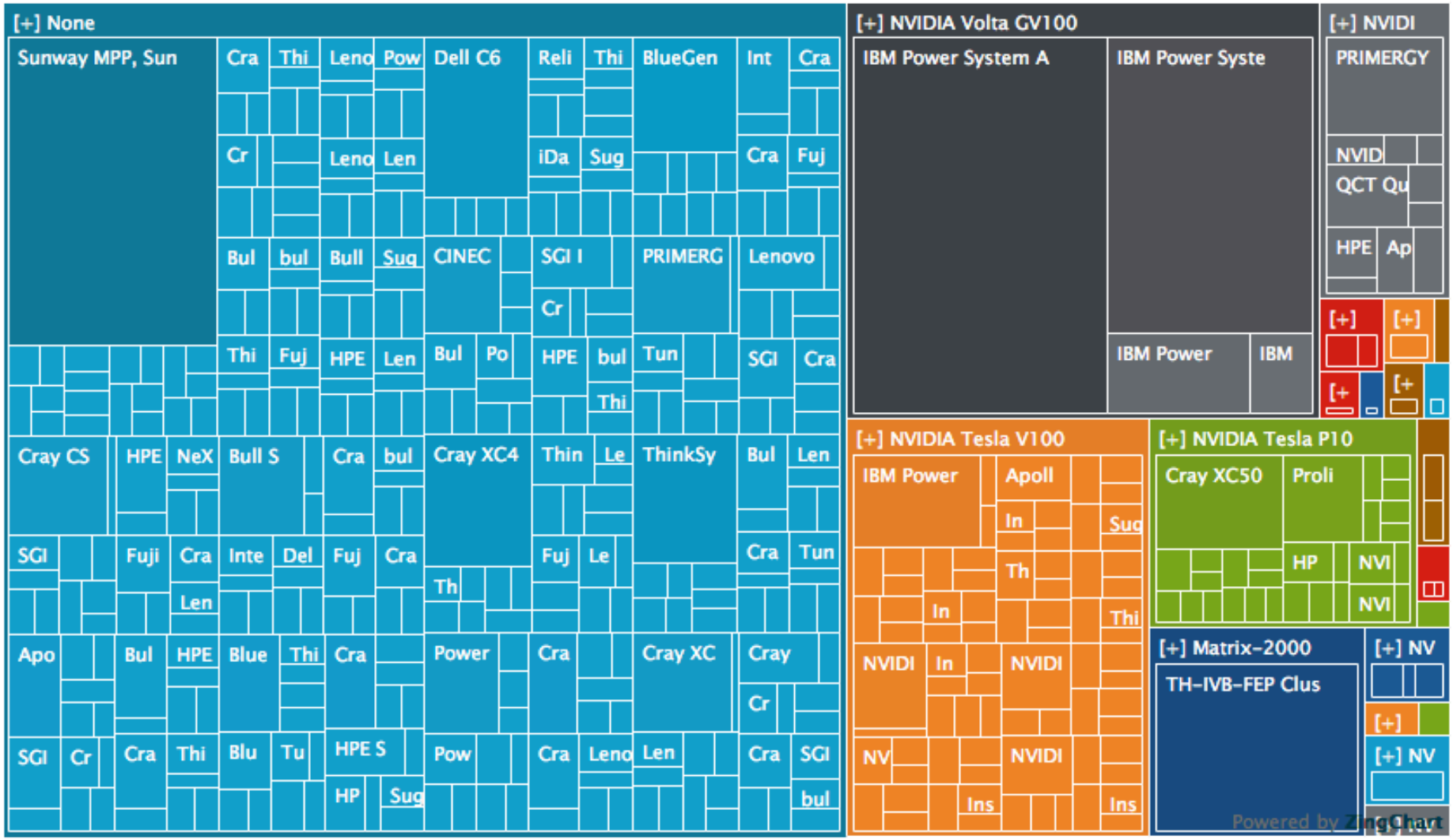
Nov'17



Nov'18



Accelerator family distribution
Nov'19



IBM POWER9 Summit

1 **Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM
DOE/SC/Oak Ridge National Laboratory
United States **Nov'19**

Summit Overview

OpenPOWER

Compute System

10.2 PB Total Memory
256 compute racks
4,608 compute nodes
Mellanox EDR IB fabric
200 PFLOPS
~13 MW

Compute Rack

18 Compute Servers
Warm water (70°F direct-cooled components)
RDHX for air-cooled components

Compute Node

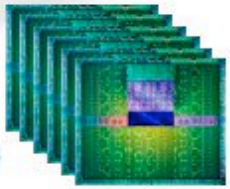
2 x POWER9
6 x NVIDIA GV100
NVMe-compatible PCIe 1600 GB SSD

25 GB/s EDR IB- (2 ports)
512 GB DRAM- (DDR4)
96 GB HBM- (3D Stacked)
Coherent Shared Memory

Components

IBM POWER9
- 22 Cores
- 4 Threads/core
- NVLink

NVIDIA GV100
- 7 TF
- 16 GB @ 0.9 TB/s
- NVLink



POWER9 Processor – Common Features

Power Systems

IBM

New Core Microarchitecture

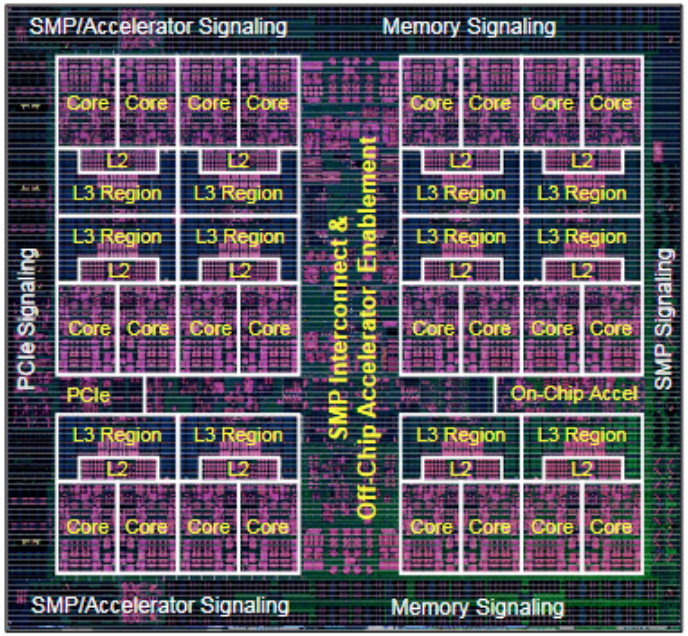
- Stronger thread performance
- Efficient agile pipeline
- POWER ISA v3.0

Enhanced Cache Hierarchy

- 120MB NUCA L3 architecture
- 12 x 20-way associative regions
- Advanced replacement policies
- Fed by 7 TB/s on-chip bandwidth

Cloud + Virtualization Innovation

- Quality of service assists
- New interrupt architecture
- Workload optimized frequency
- Hardware enforced trusted execution



14nm finFET Semiconductor Process

- Improved device performance and reduced energy
- 17 layer metal stack and eDRAM
- 8.0 billion transistors

Leadership Hardware Acceleration Platform

- Enhanced on-chip acceleration
- Nvidia NVLink 2.0: High bandwidth and advanced new features (25G)
- CAPI 2.0: Coherent accelerator and storage attach (PCIe G4)
- New CAPI: Improved latency and bandwidth, open interface (25G)

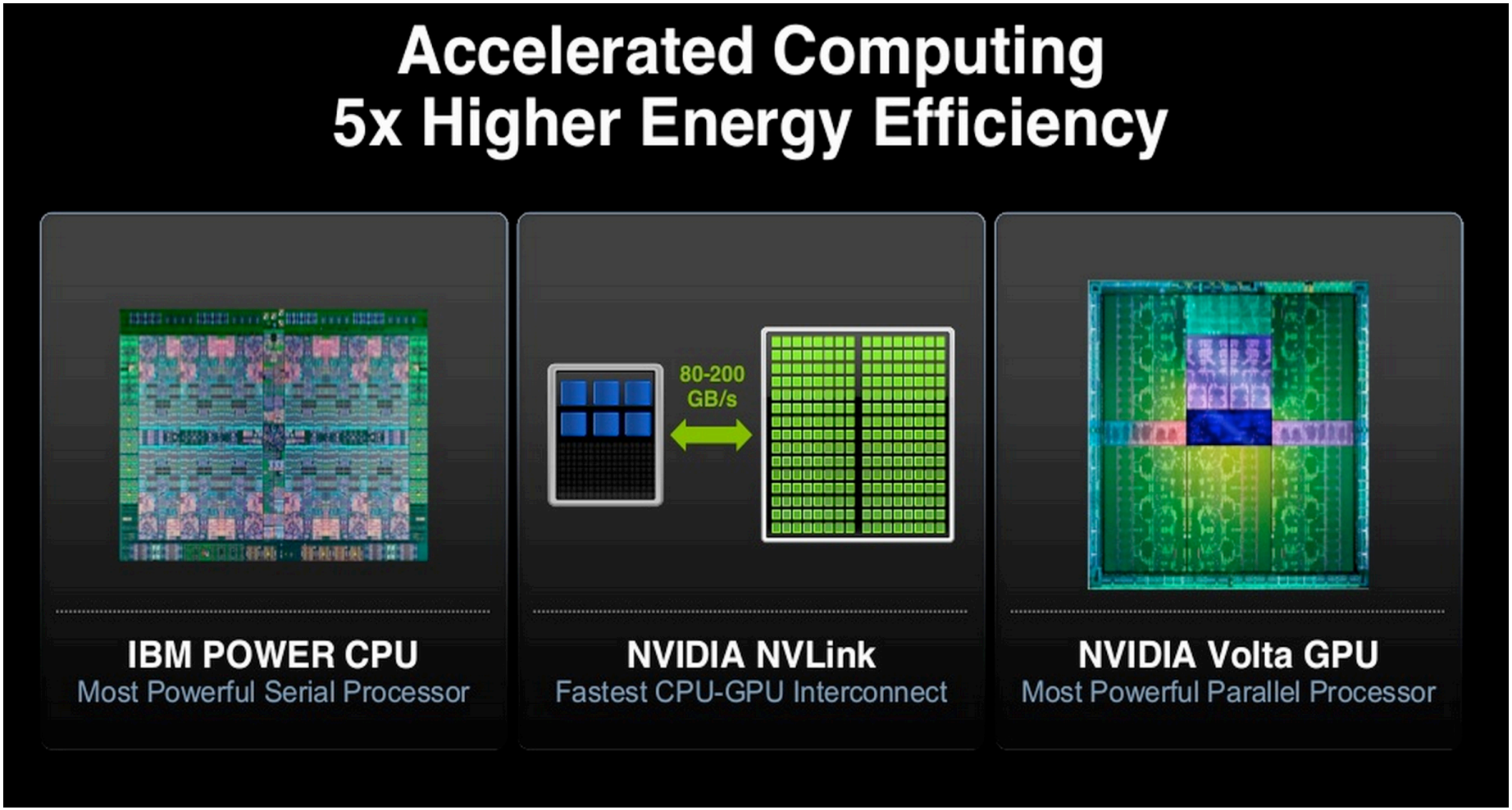
State of the Art I/O Subsystem

- PCIe Gen4 – 48 lanes

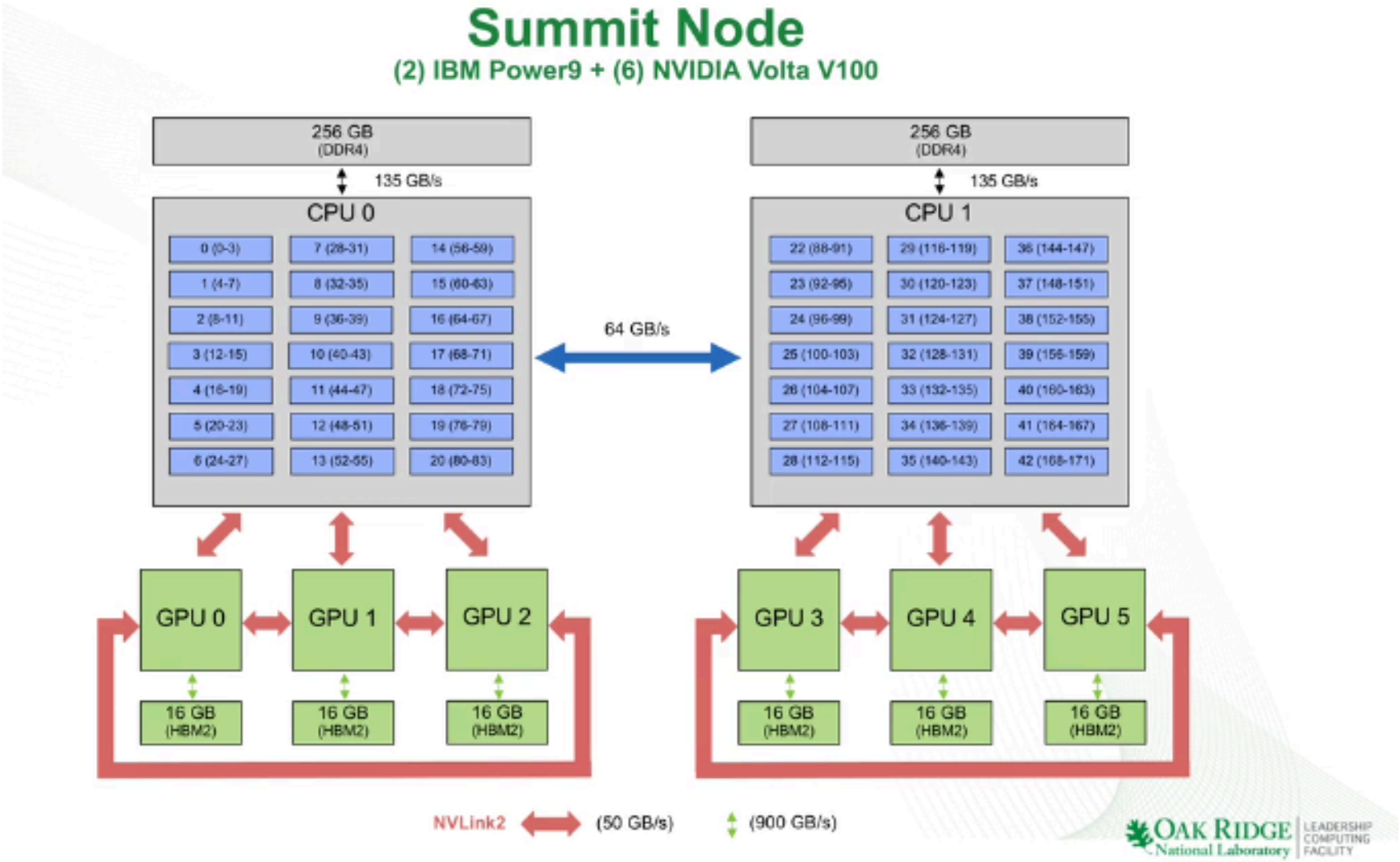
High Bandwidth Signaling Technology

- 16 Gb/s interface
  - Local SMP
- 25 Gb/s Common Link interface
  - Accelerator, remote SMP

IBM POWER9 + NVidia V100



Summit node architecture



https://en.wikichip.org/wiki/supercomputers/summit

IBM Power BlueGene/Q Compute (Sequoia)

Blue Gene/Q packaging hierarchy

1. Chip
16 cores

2. Module
Single Chip

3. Compute Card
One single chip module,
16 GB DDR3 Memory

4. Node Card
32 Compute Cards,
Optical Modules, Link Chips,
Torus

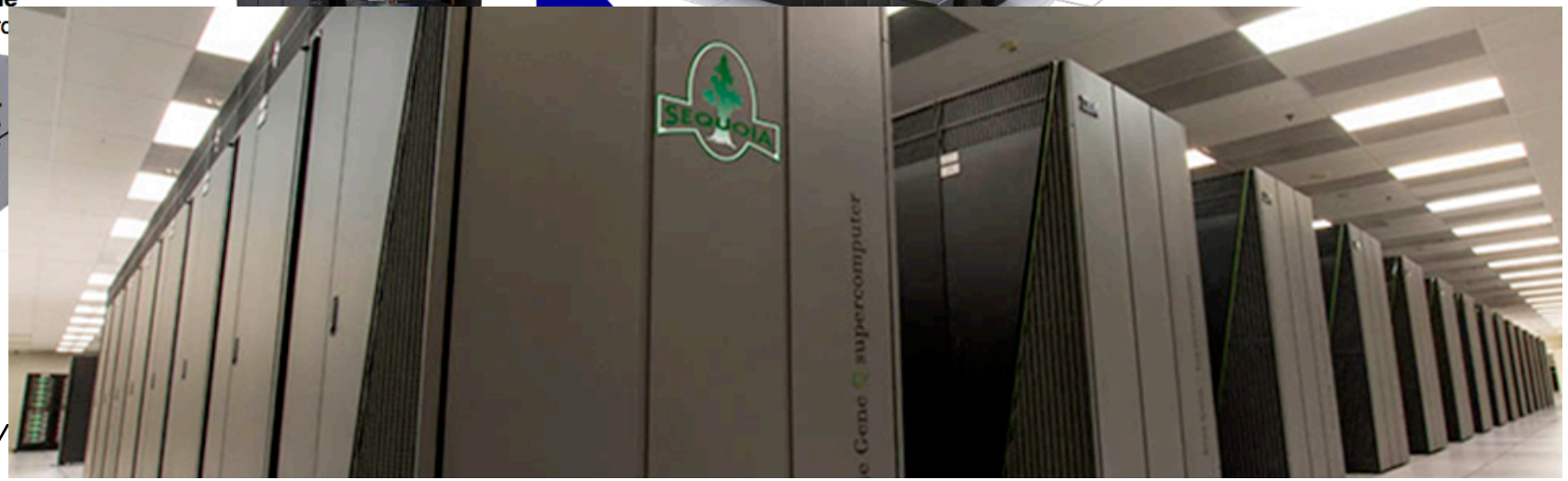
5a. Midplane
16 Node Card

5b. I/O Drawer
8 I/O Cards
8 PCIe Gen2 slots

6. Rack
2 Midplanes
1, 2 or 4 I/O Drawers

7. System
20PF/s

Ref: SC2010



TOP500 The List.

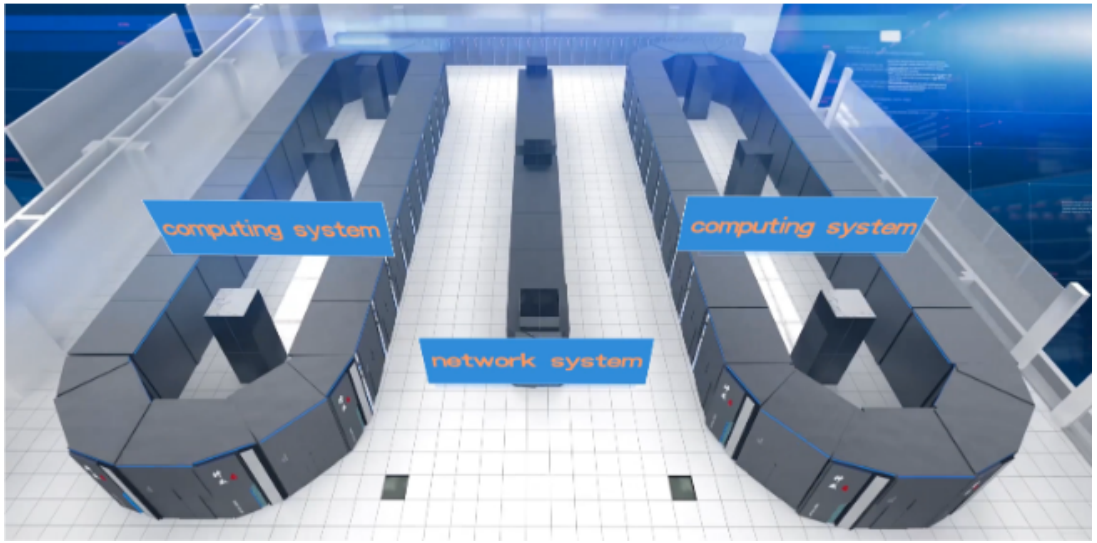
Jun'12: #1
Nov'12: #2
Jun'13: #3
Nov'13: #3
Jun'14: #3
Nov'14: #3
Jun'15: #3
Nov'15: #3
Jun'16: #4
Nov'16: #4
Jun'17: #5
Nov'17: #6
Jun'18: #8
Nov'18: #10
Nov'19: #12

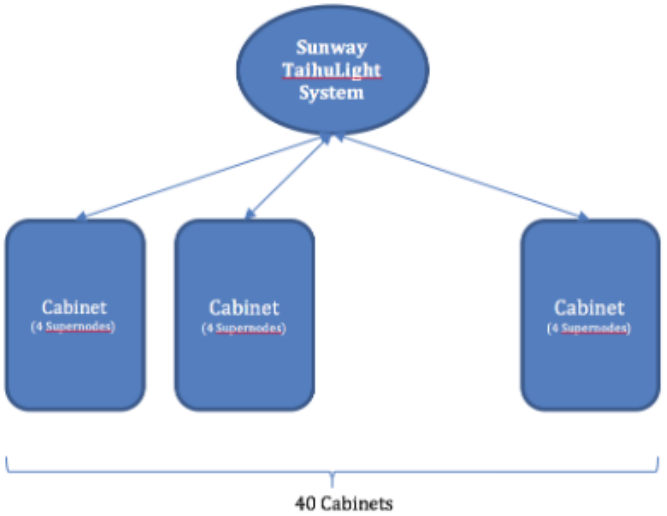
Sunway TaihuLight
(#1 in June'16 TOP500)

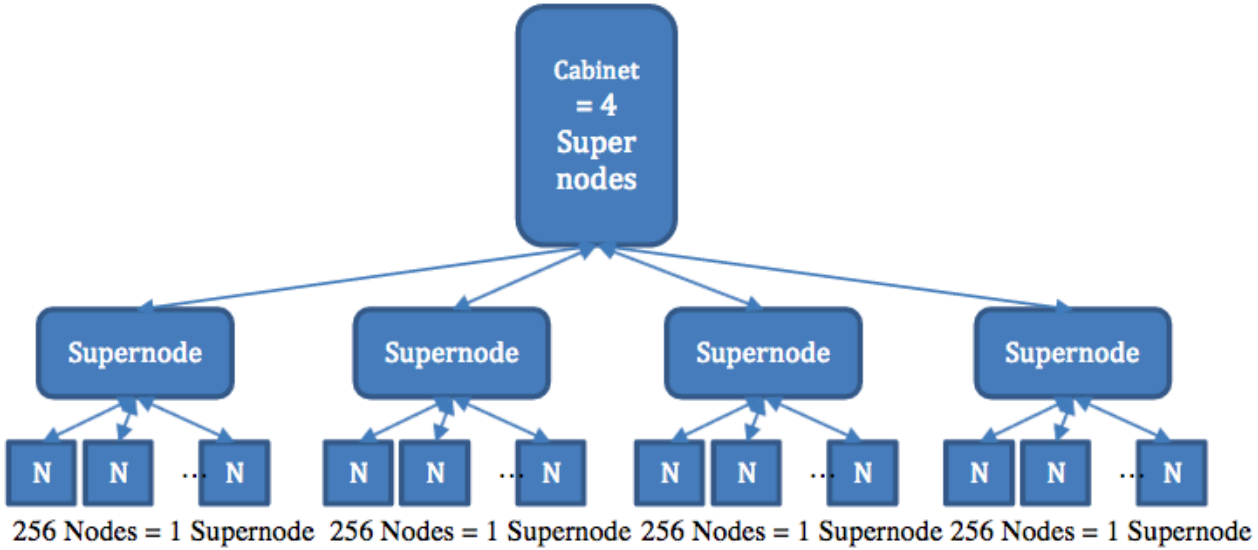
TOP 500 The List.

Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC
National Supercomputing Center in Wuxi
China **Nov'19**

Overview of the Sunway TaihuLight System



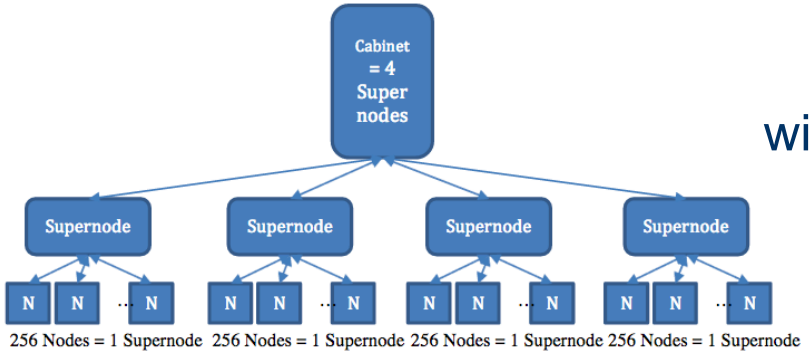




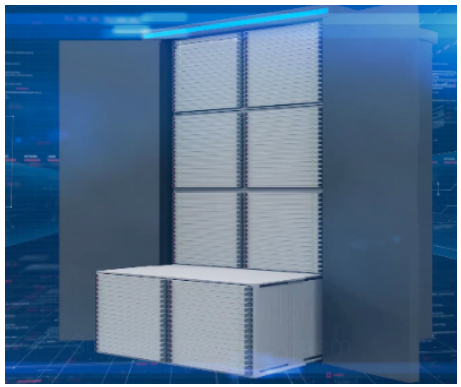
AJProença, Advanced Architectures,

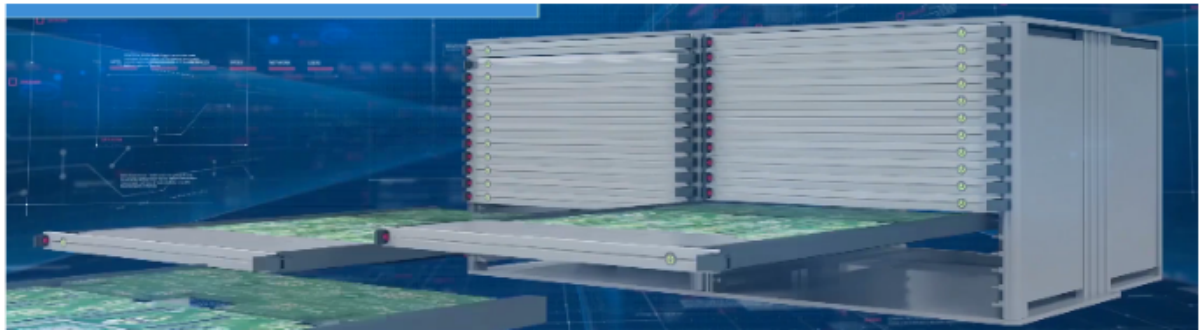
http://www.netlib.org/utk/people/JackDongarra/PAPERS/sunway-report-2016.pdf

Sunway TaihuLight
(#1 in June'16 TOP500)

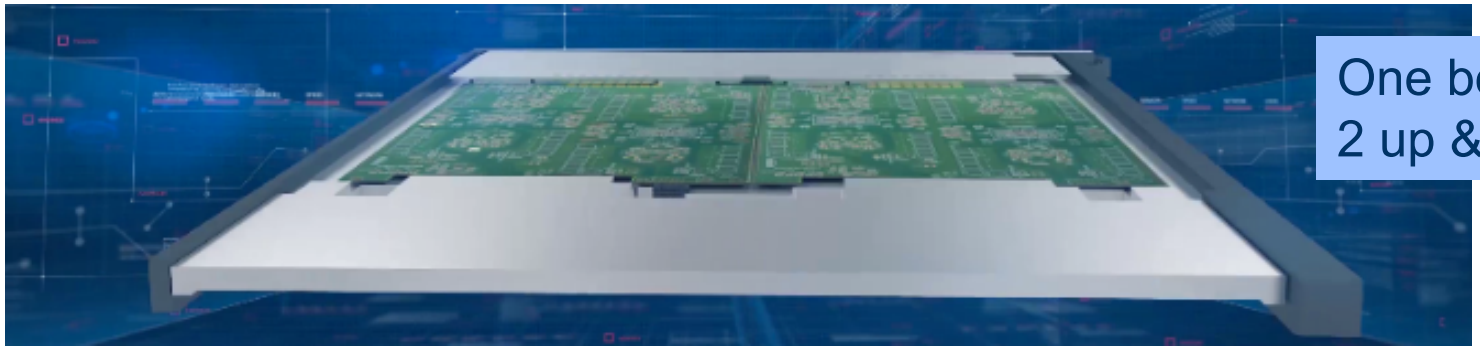


One cabinet
with 4 Supernodes





One Supernode
with 32 boards

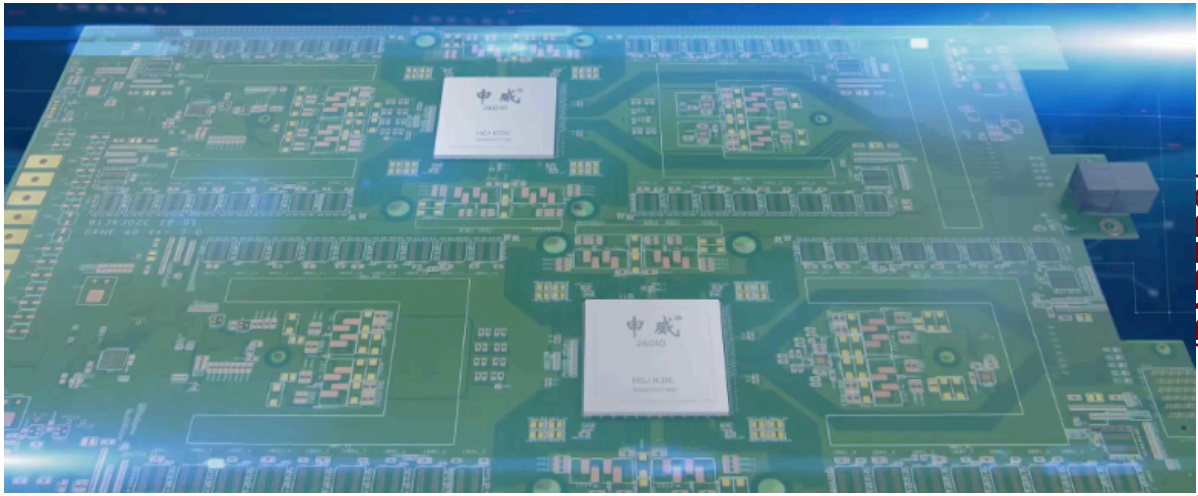


One board with 4 cards,
2 up & 2 down

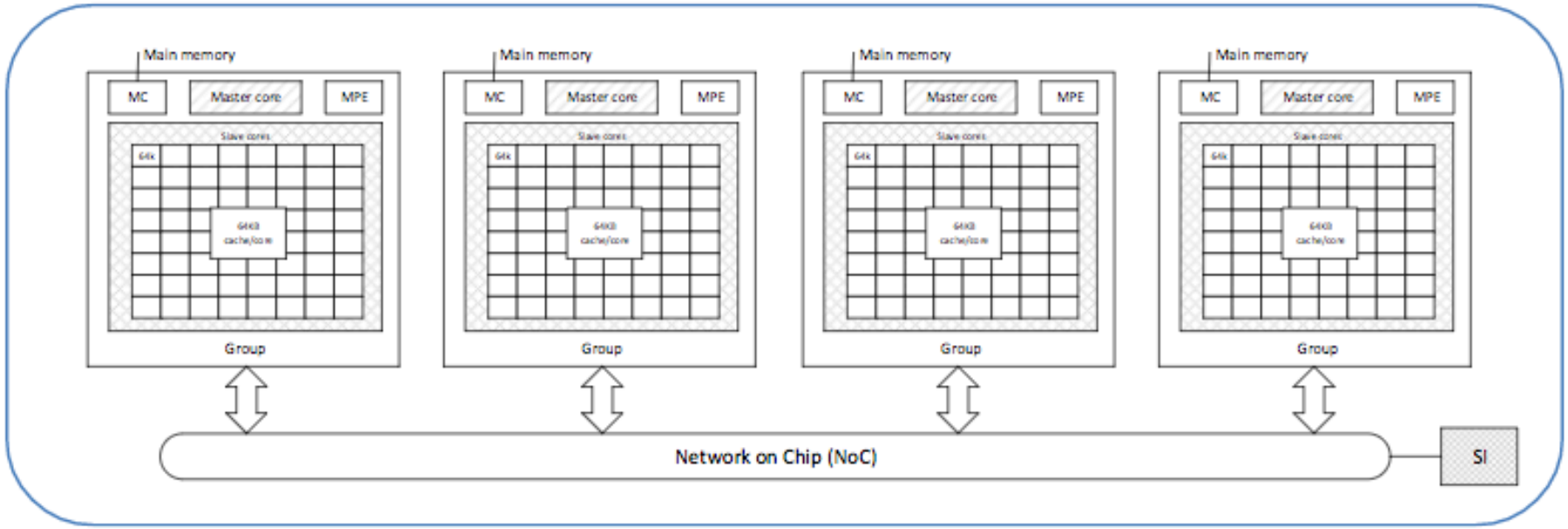
http://www.netlib.org/utk/people/JackDongarra/PAPERS/sunway-report-2016.pdf

Sunway TaihuLight
(#1 in June'16 TOP500)

One card with two nodes
(two SW26010 chips)



SW26010: the 4x64-core 64-bit RISC processor (w/ **256-bit vector** instructions & only cache L1)



Overview of Tianhe-2A

4 TOP500 The List.

Tianhe-2A - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000 , NUDT

Nov'19

Overview of Tianhe-2A

Comparison

Items	Milkyway-2	Milkyway-2A
Nodes & Performance	16000 nodes with Intel CPU + KNC	**17792 nodes with Intel CPU + Matrix-2000**
	54.9Pflops	**94.97Pflops**
Interconnection	10Gbps, 1.57us	**14Gbps, 1us**
Memory	1.4PB	**3.4PB**
Storage	12.4PB, 512GB/s	**20PB, 1TB/s**
Energy Efficiency	17.8MW, 1.9Gflops/W	**About 18MW, >5Gflops/W**
Heterogeneous software	MPSS for Intel KNC	**OpenMP/OpenCL for Matrix-2000**

https://www.nextplatform.com/2018/02/14/look-whats-store-chinas-tianhe-2a-supercomputer/

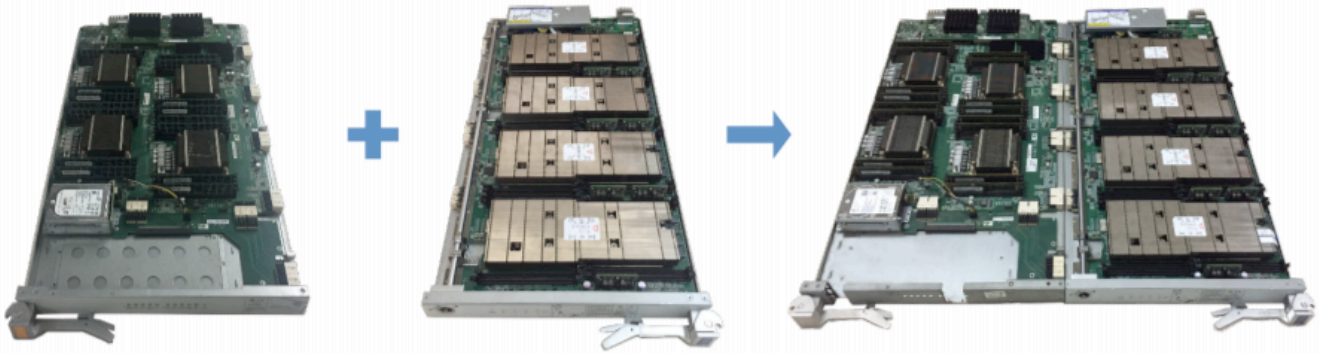
Overview of Tianhe-2A

Compute nodes

Heterogeneous Compute Blades

– Compute blade = Xeon part + Matrix-2000 part

4 Intel Xeon CPUs 4 FT Matrix-2000 2 Compute Nodes



– Use the Matrix-2000 part to replace the KNC part

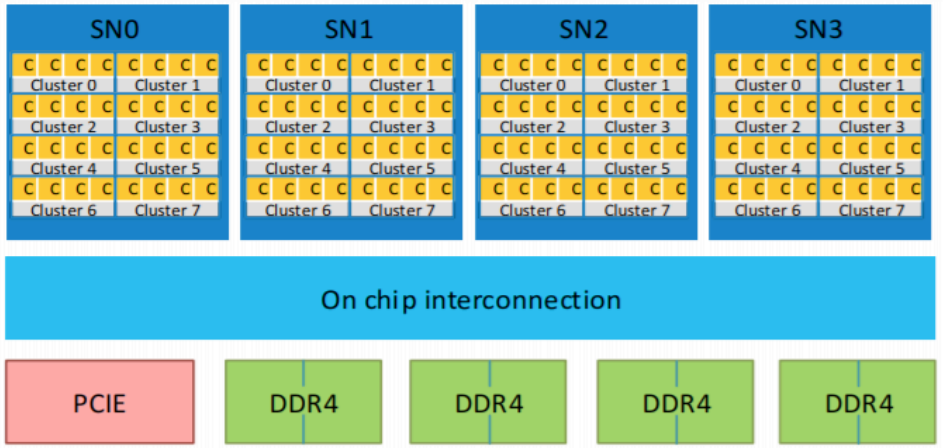
https://www.nextplatform.com/2018/02/14/look-whats-store-chinas-tianhe-2a-supercomputer/

Replacing the KNC in Tianhe-2A: the Matrix-2000 accelerator

Matrix-2000 accelerator

- **Chip specification**
  - 128cores
    - 4 super-nodes (SN)
    - 8 clusters per SN
    - 4 cores per cluster
    - Core
      - Self-defined 256-bit vector ISA
      - 16 DP flops/cycle per core
  - Peak performance: 2.4576Tflops@1.2GHz

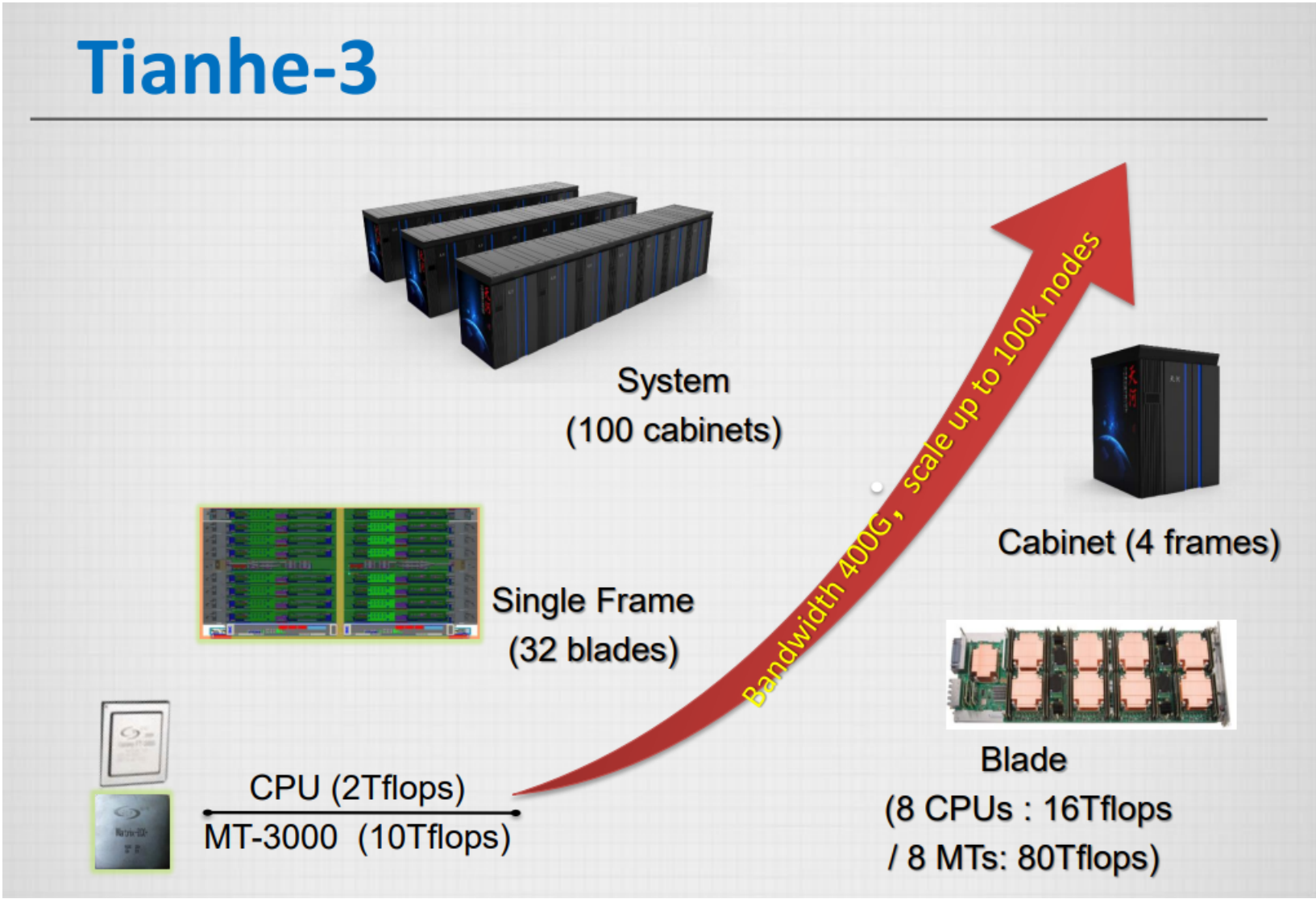
    4 SNs x 8 clusters x 4cores x 16 flops x 1.2 GHz = 2.4576 Tflops
  - Peak power dissipation: ~240w
  - Interface
    - 8 DDR4-2400 channels
    - X16 PCIE 3.0 EP Port



15

https://www.nextplatform.com/2018/02/14/look-whats-store-chinas-tianhe-2a-supercomputer/

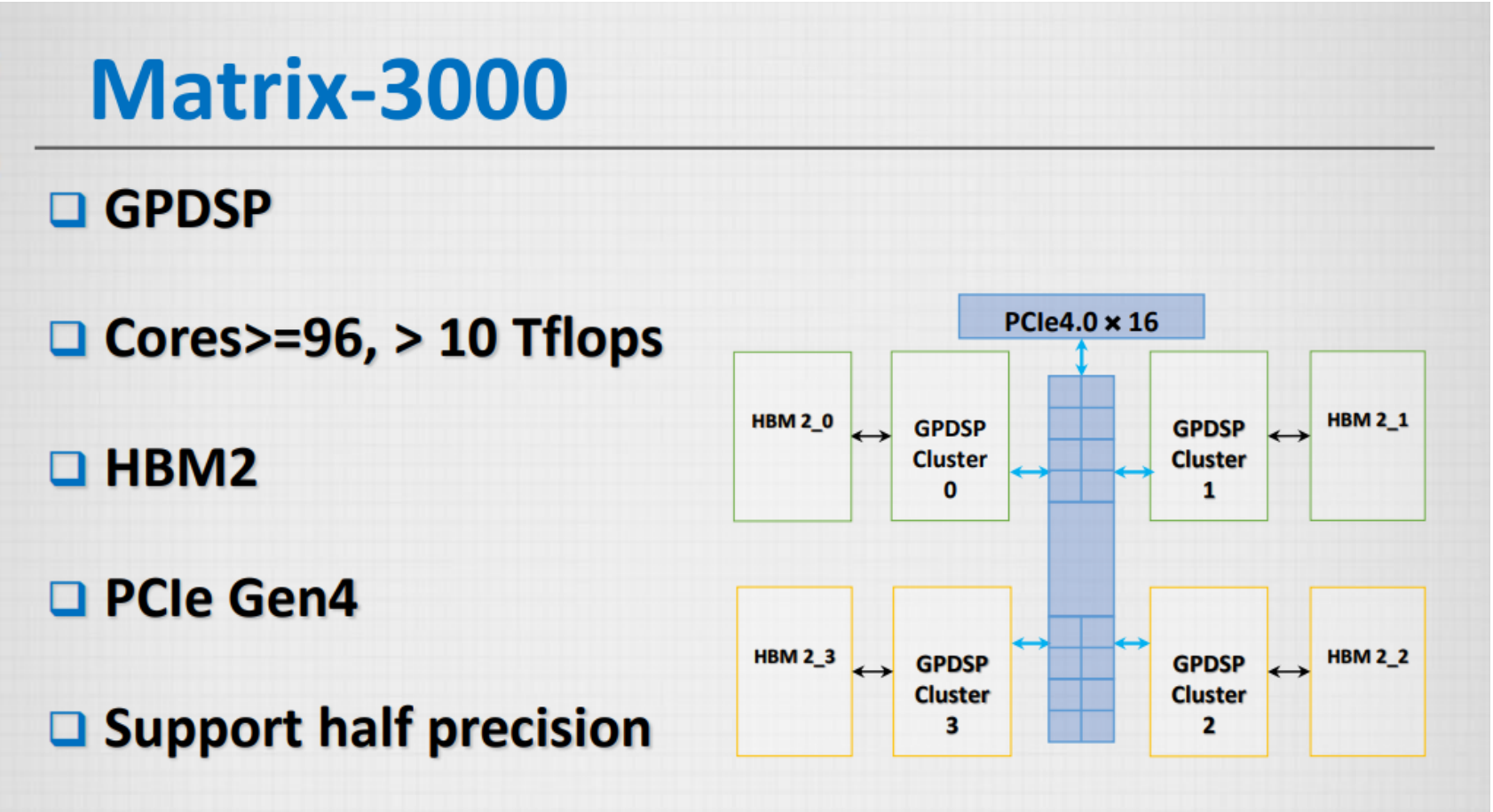
Next: Tianhe-3: Fujitsu A64FX ARM-SVE + Matrix-3000 accelerators



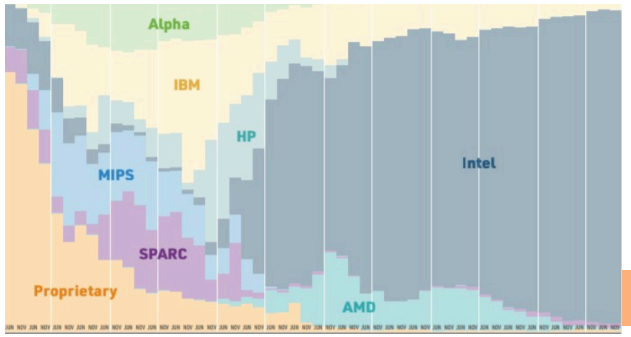
https://www.nextplatform.com/2019/05/02/china-fleshes-out-exascale-design-for-tianhe-3/

NUDT Matrix-3000 accelerator

National University of Defense Technology (NUDT)



https://www.nextplatform.com/2019/05/02/china-fleshes-out-exascale-design-for-tianhe-3/



Intel presence at TOP500:
the Xeon Processor Scalable Family
(formerly code-named Skylake-SP)

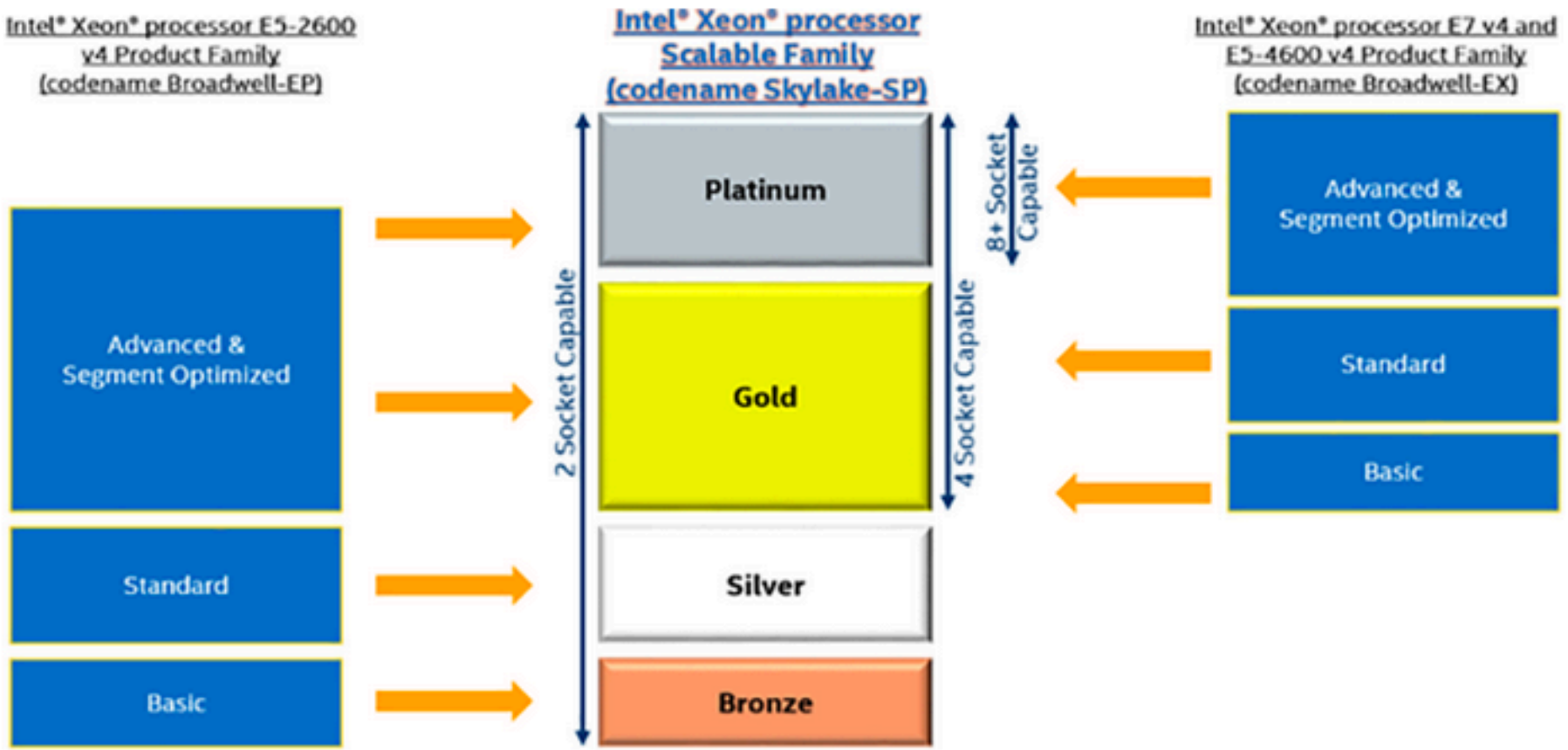
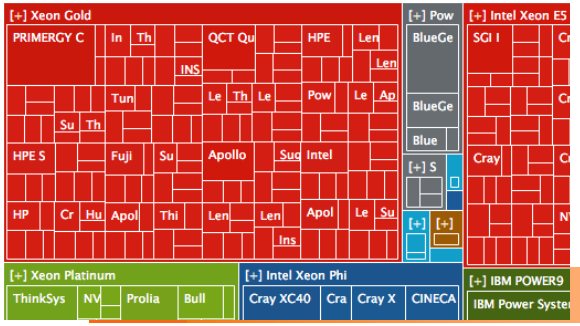


Figure 2. New branding for processor models.

https://software.intel.com/en-us/articles/intel-xeon-processor-scalable-family-technical-overview



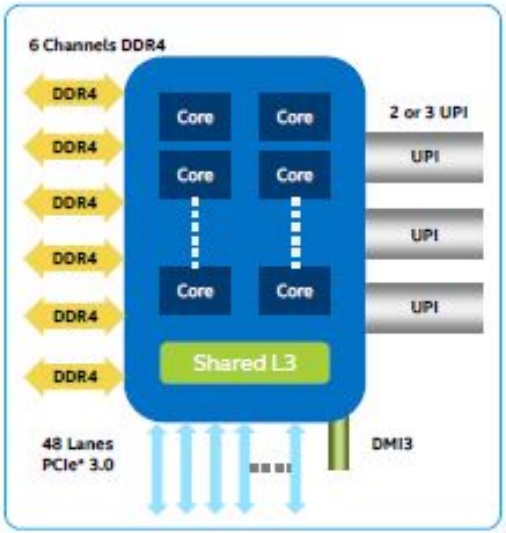
Intel Xeon Scalable Processor

First Generation Intel® Xeon® Scalable Processor

Introduced in July 2017

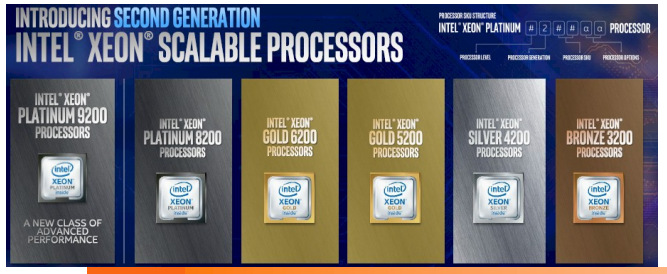
- Skylake-SP core microarchitecture with data center specific enhancements
- Intel® AVX-512 with 32 DP flops per cycle per core
- Data center optimized cache hierarchy – 1MB L2 per core, non-inclusive L3
- New Intel® Mesh architecture
- Enhanced 6 channel memory subsystem
- 48 lanes of PCIe Gen3 with integrated DMA, NTB, and VMD devices
- New Intel® Ultra Path Interconnect (Intel® UPI)

Features	Intel® Xeon® Scalable Processor
Cores and Threads Per CPU	Up to 28 cores and 56 threads
Last-level Cache (LLC)	Up to 38.5 MB (non-inclusive)
QPI/UPI Speed (GT/s)	Up to 3x UPI @ 10.4 GT/s
PCIe* Lanes/ Controllers	Up to 48 / 12 / PCIe 3.0 (2.5, 5, 8 GT/s)
Memory Population	Up to 6 channels of up to 2 RDIMMs, LRDIMMs, or 3DS LRDIMMs
Max Memory Speed	Up to 2666 MHz



Foundation for Accelerating Data Center Innovations

https://www.servethehome.com/resetting-scalable-expectations-intel-cascade-lake-sp-at-hot-chips-30/



Next PU generations for HPC: Intel Xeon Processor Scalable Family

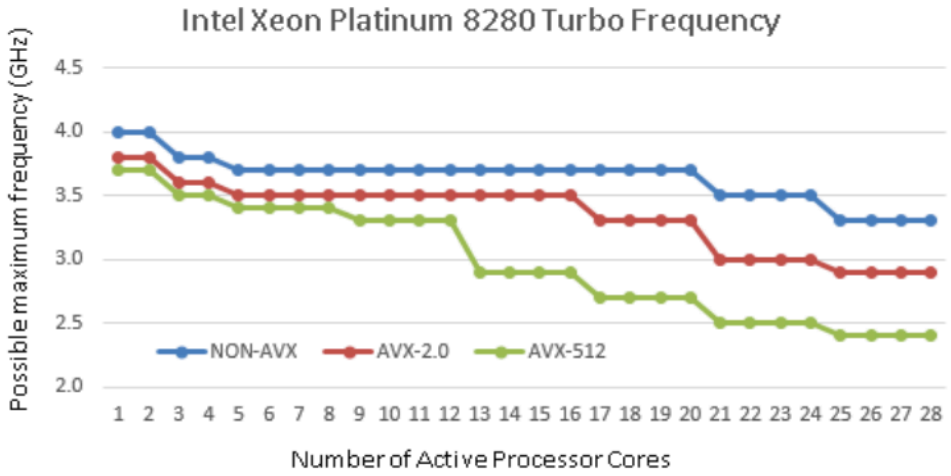
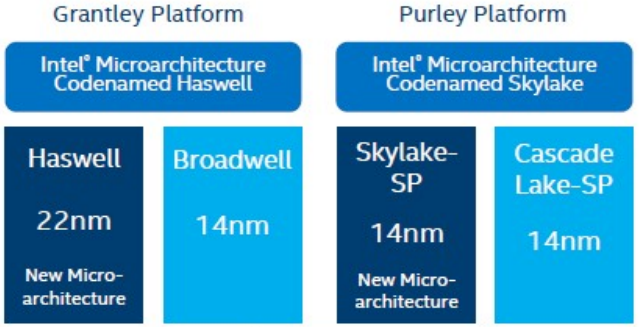


Figure 1 Intel Xeon Platinum 8280 Turbo Frequency



Features	Cascade Lake CPU
Cores and Threads	Up to 28 Cores and 56 Threads
Last-level Cache	Up to 38.5 MB (non-inclusive)
UPI Speed (GT/s)	Up to 3x UPI @ 10.4 GT/s
PCIe* 3.0 Lanes	Up to 48 lanes with 12 controllers
Memory Speed	Up to 6 channels @ up to 2666 MHz

Cache Hierarchy Changes

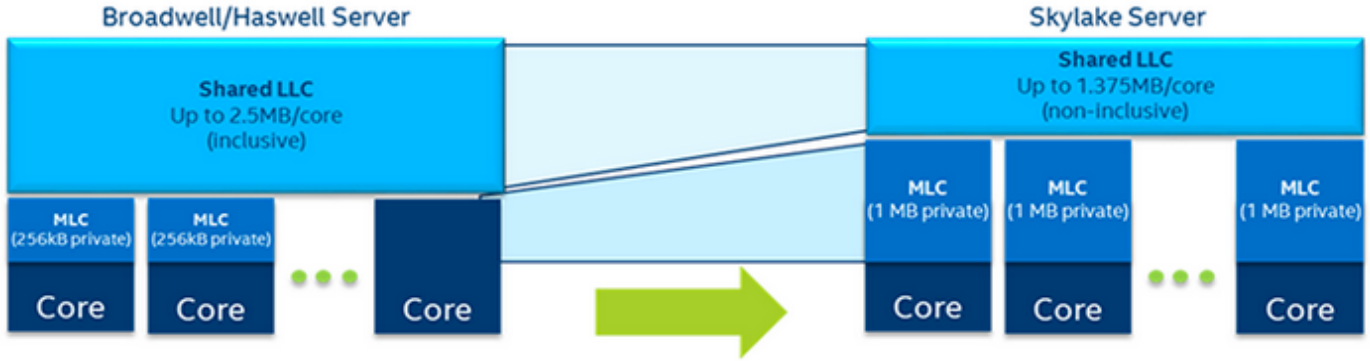
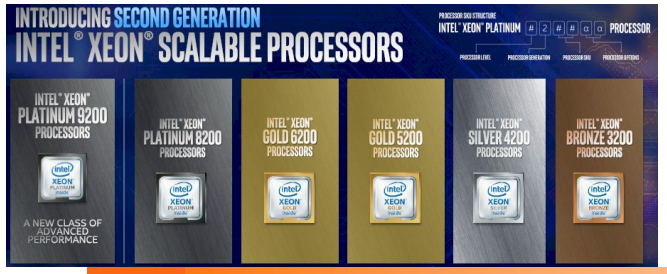


Figure 11. Generational cache comparison.

https://software.intel.com/en-us/articles/intel-xeon-processor-scalable-family-technical-overview



Intel Xeon Processor Scalable Family

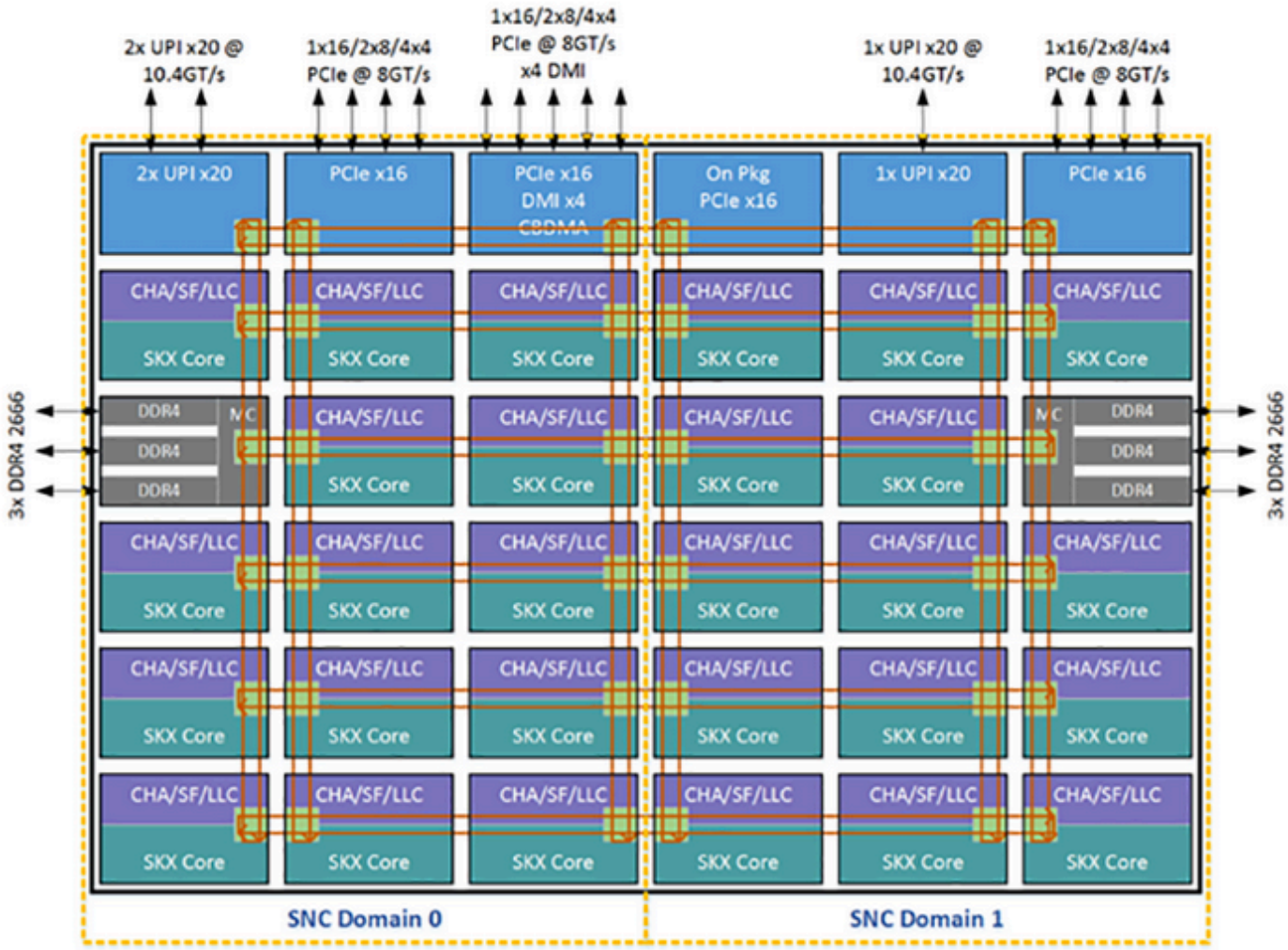
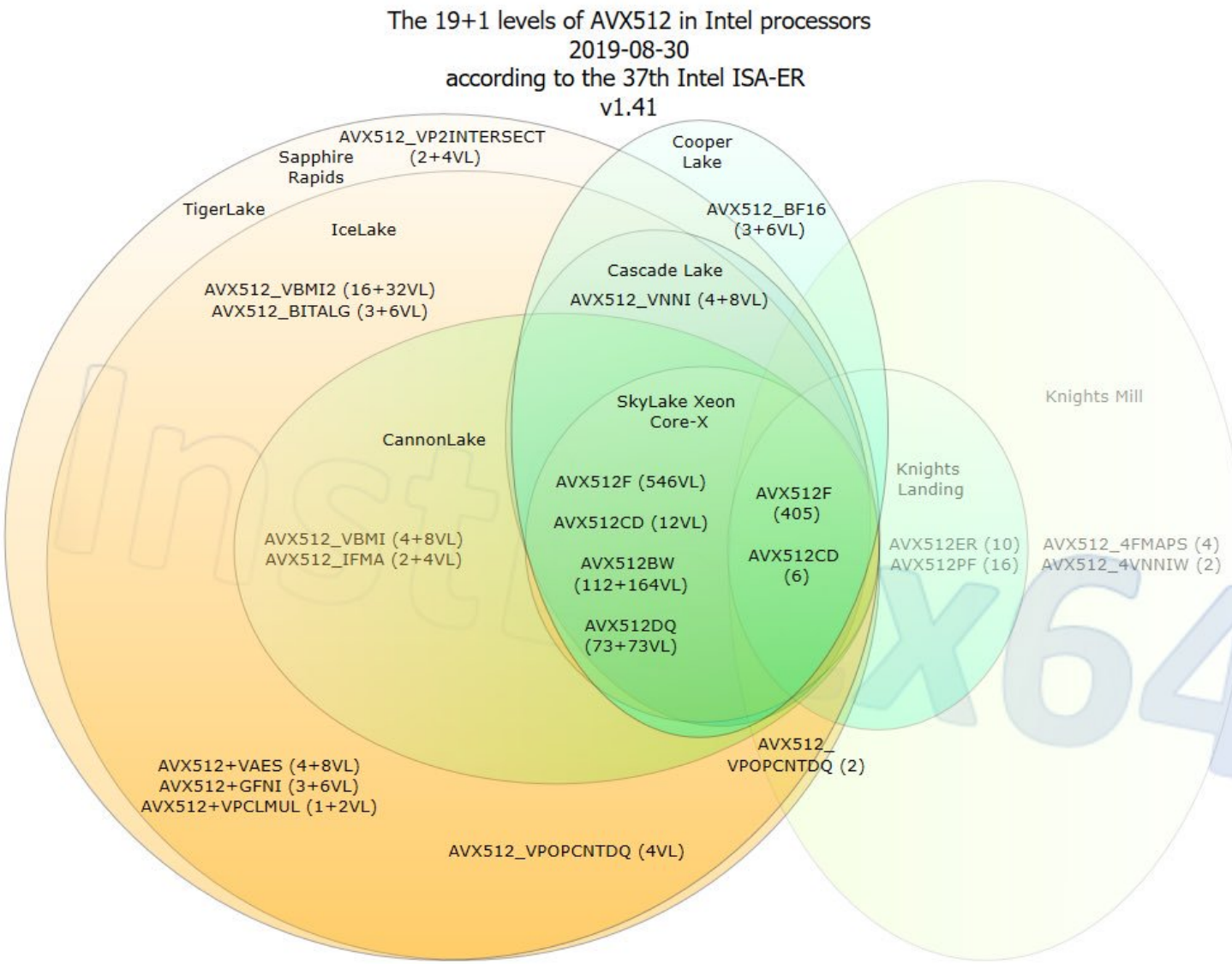


Figure 10. Sub-NUMA cluster domains.

https://software.intel.com/en-us/articles/intel-xeon-processor-scalable-family-technical-overview

Intel levels of AVX512



Intel Xeon Scalable Processor : the AVX512 VNNI extension

Cascade Lake Vector Neural Network Instructions

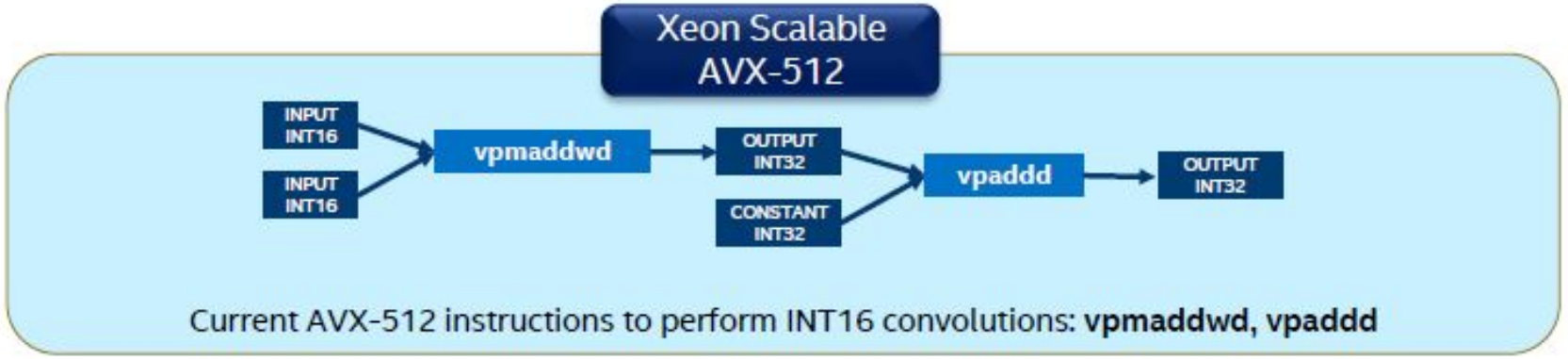
Vector Neural Network Instruction (VNNI) on Cascade Lake accelerates Deep Learning and AI inference workloads

- VNNI : A new set of Intel® Advanced Vector Extension (Intel® AVX-512) instructions
  - 8-bit (int8) new instruction (VPDPBUSD)
    - Fuses 3 instructions in inner convolution loop using int8 data type
  - 16-bit (int16) new instruction (VPDPWSSD)
    - Fuses 2 instructions in inner convolution loop using int16 data type

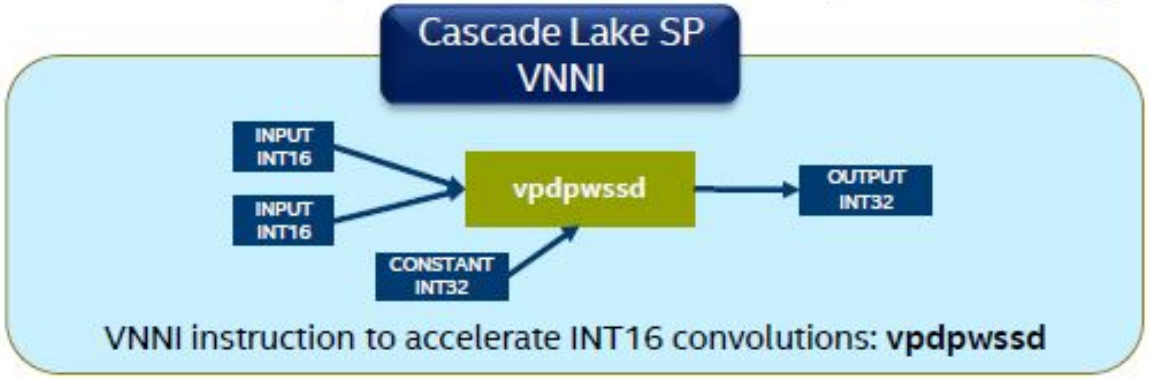
https://www.servethehome.com/resetting-scalable-expectations-intel-cascade-lake-sp-at-hot-chips-30/

Intel Xeon Scalable Processor:
the AVX512 VNNI extension

AI/DL Inference Enhancements on INT16 with VNNI

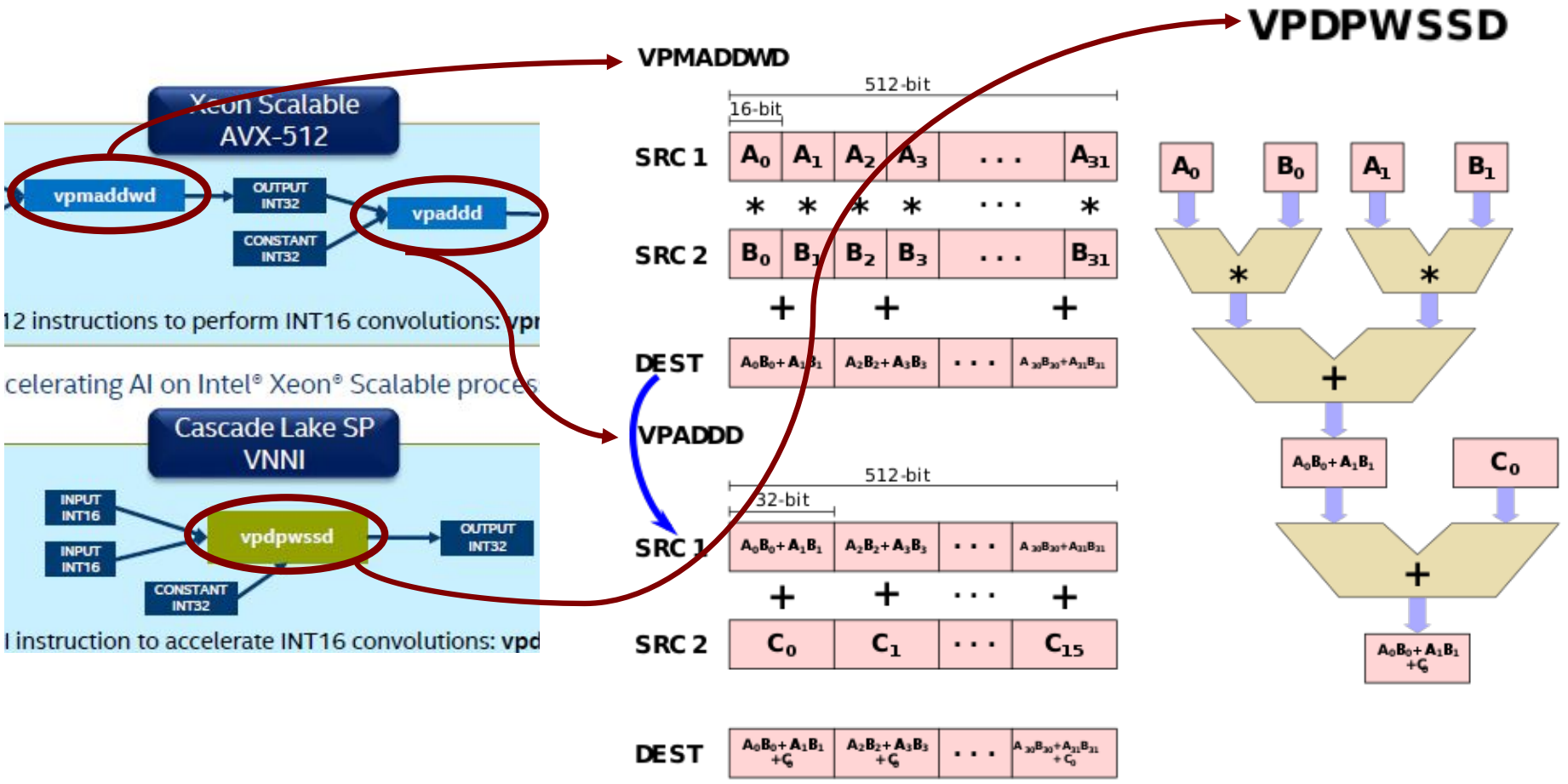


New instructions for accelerating AI on Intel® Xeon® Scalable processors using int16 data



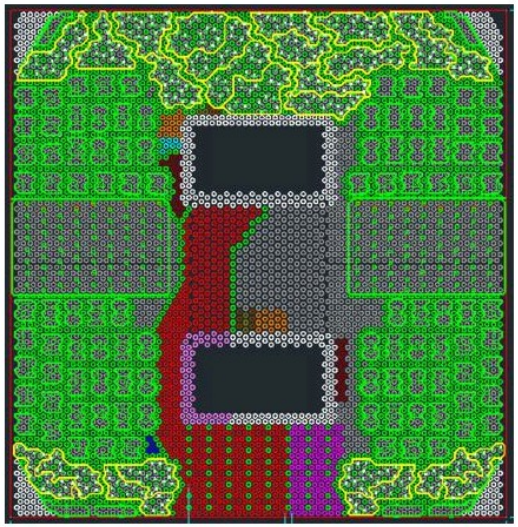
https://www.servethehome.com/resetting-scalable-expectations-intel-cascade-lake-sp-at-hot-chips-30/

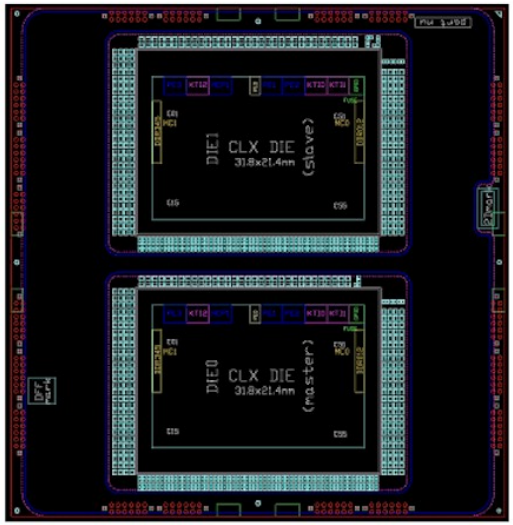
Intel Xeon Scalable Processor:
the AVX512 VNNI extension

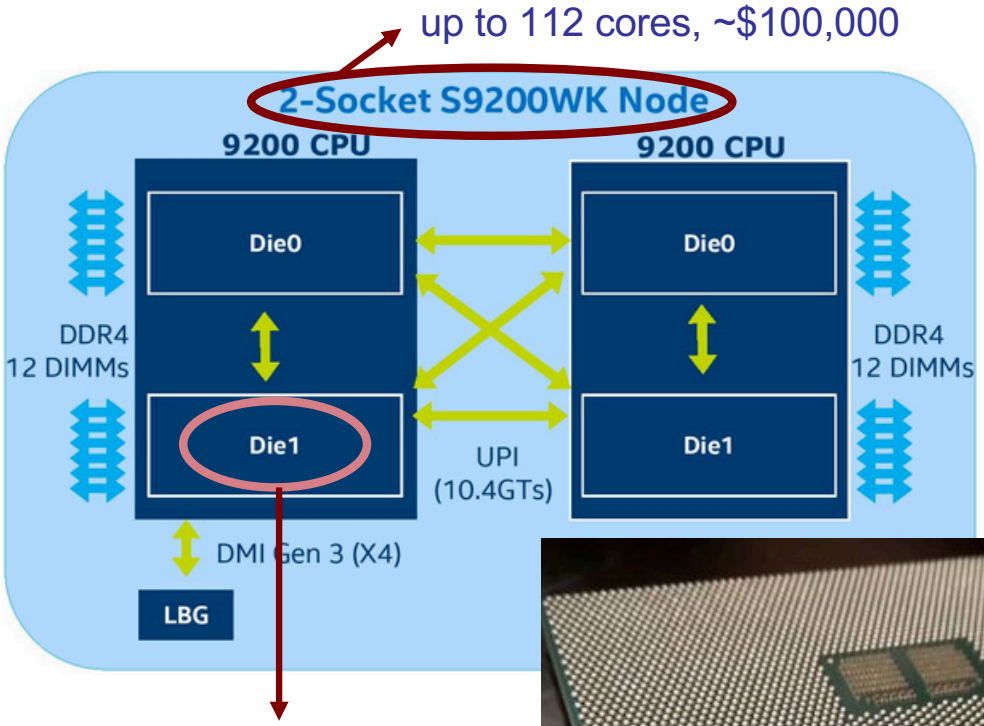


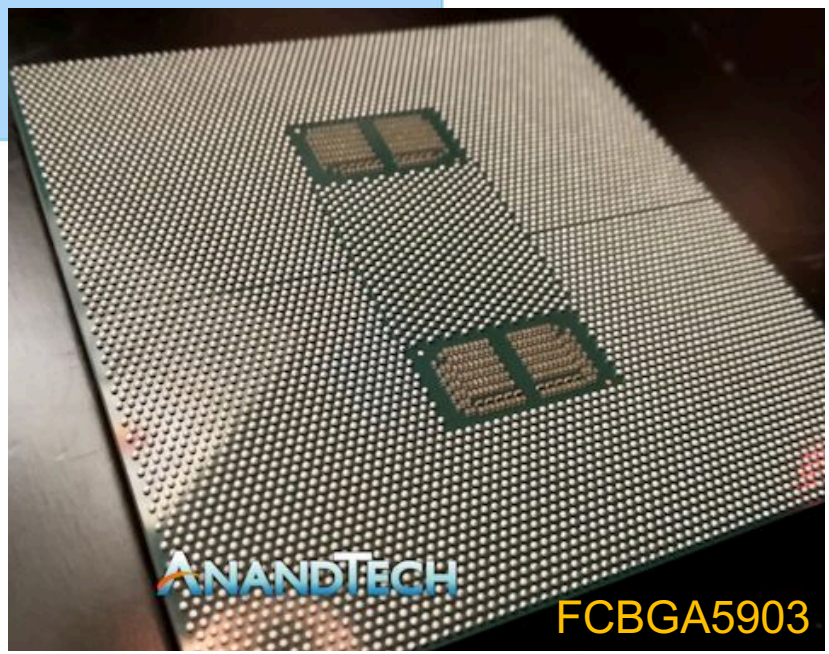
https://www.servethehome.com/resetting-scalable-expectations-intel-cascade-lake-sp-at-hot-chips-30/

Intel Xeon Scalable Processor:
the Advanced Performance (AP) serie









Next at Intel:
the Sunny Cove micro-architecture
(for the Ice Lake chip and ...)

SUNNYCOVE MICROARCHITECTURE

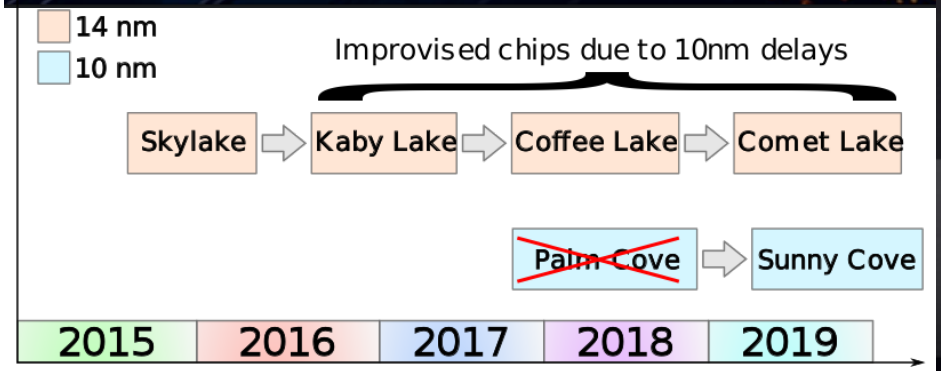
	HASWELL	SKY LAKE	ICE LAKE
L1 Data Cache	32KB	32KB	48KB
L2 Cache	256KB	256KB	512KB
L2 TLB	1024	1536 16 (1G)	2048 (4k) Shared 1024 for 2M/4M 1024 for 1G
µop Cache	1.5K µops	1.5K µops	2.25K µops
OoO Window	182	224	352
In-Flight Loads	72	72	128
In-Flight Stores	42	56	72

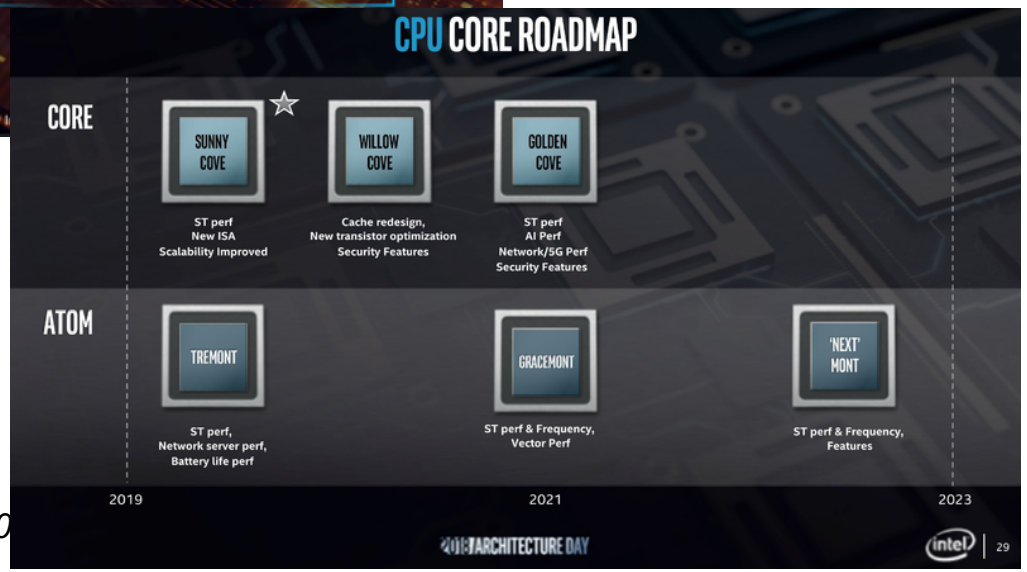
NEW CAPABILITIES

New Instructions for Crypto Performance
- Big Number Arithmetic (IFMA)
- Vector AES
- Vector Carryless Multiply
- Galois Field
- SHA

Additional Vector Capabilities
- DLBoost – Inference Acceleration
- VBMI (Permutes/Shifts)
- VBMI2 (Expand/Compress/Shifts)
- BITALG (POPCNT, Bit Shuffle)

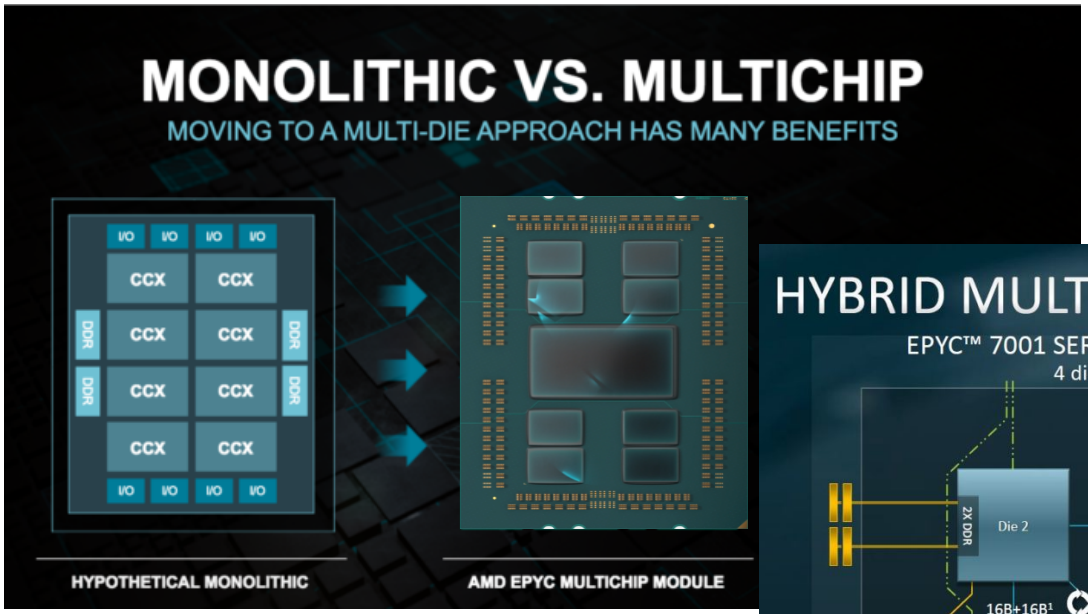
Security Features
- User Mode Instruction Prevention (UMIP)

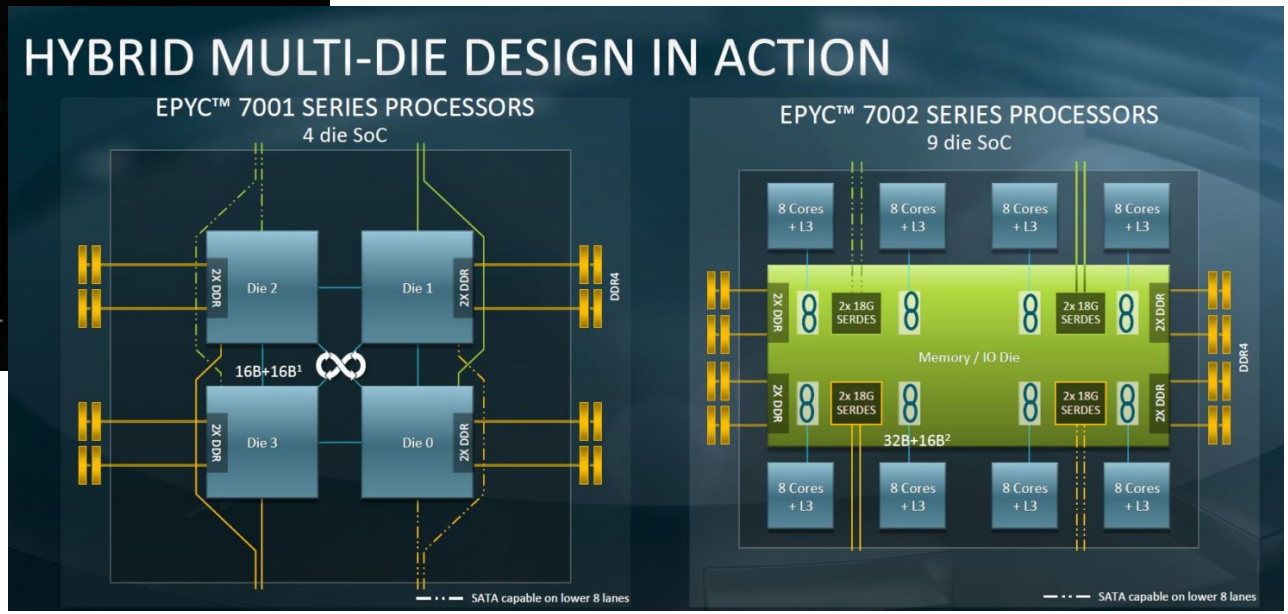




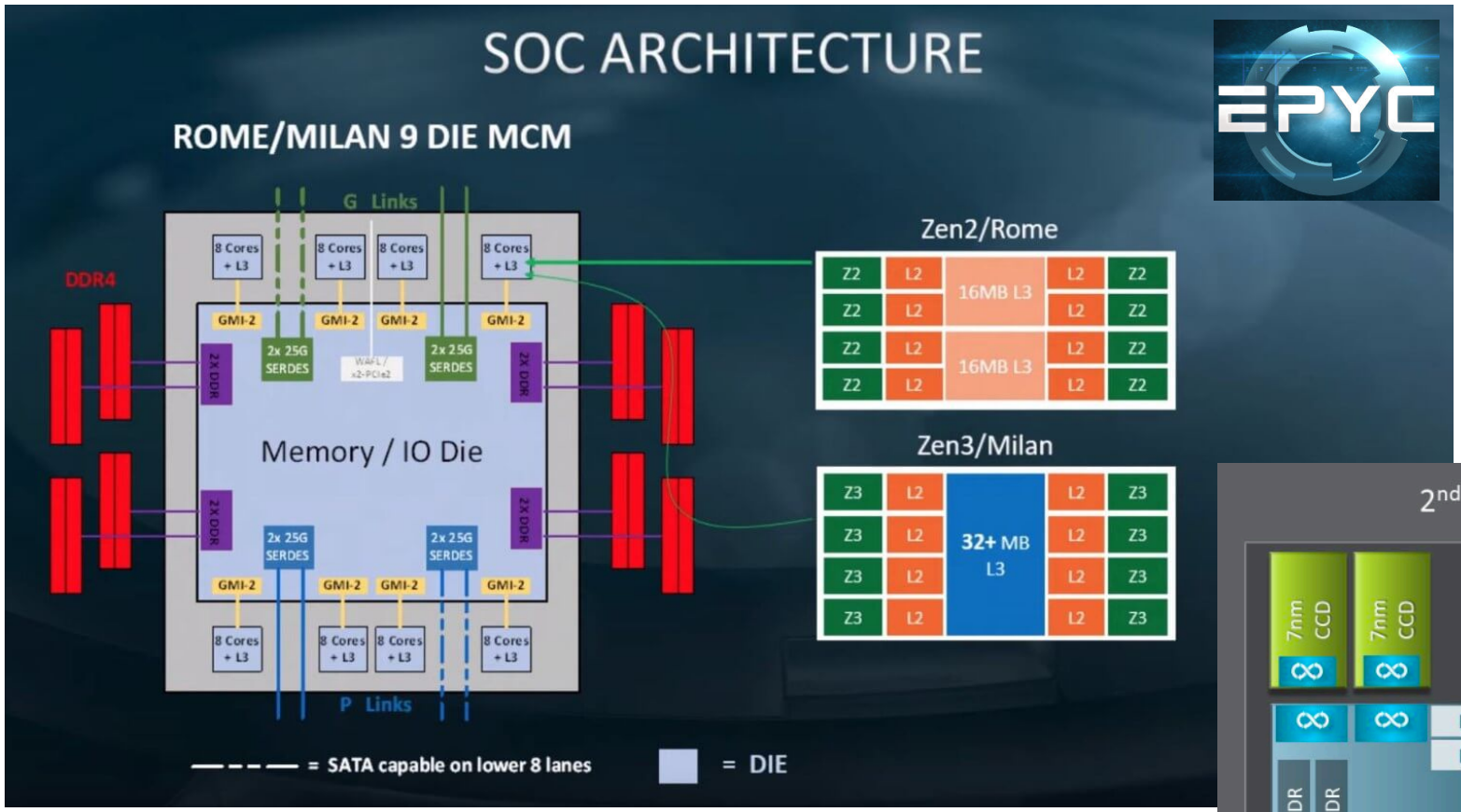
Key Intel Xeon competitor: AMD Epyc (Zen, Zen 2, 3, 4)

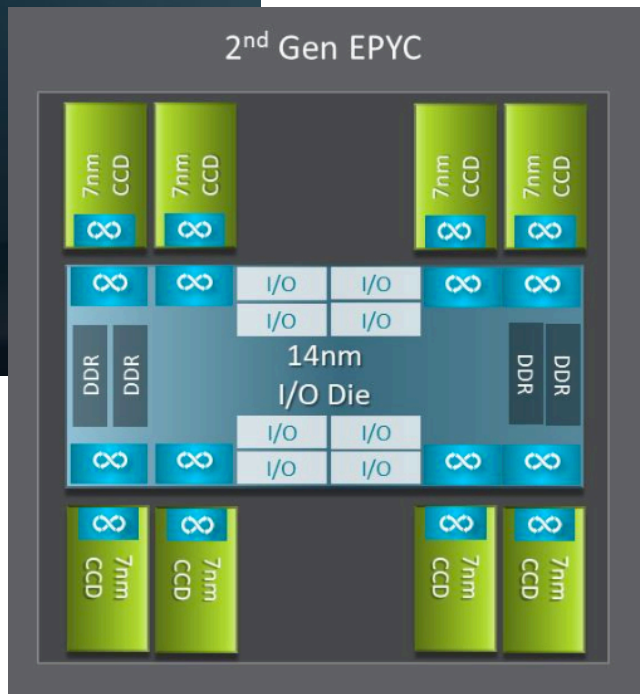
| AMD EPYC | SEPTEMBER 10TH 2019 | HPE CAST IN JAPAN





AMD Epyc: from Zen 2 (Rome) to Zen 3 (Milan)





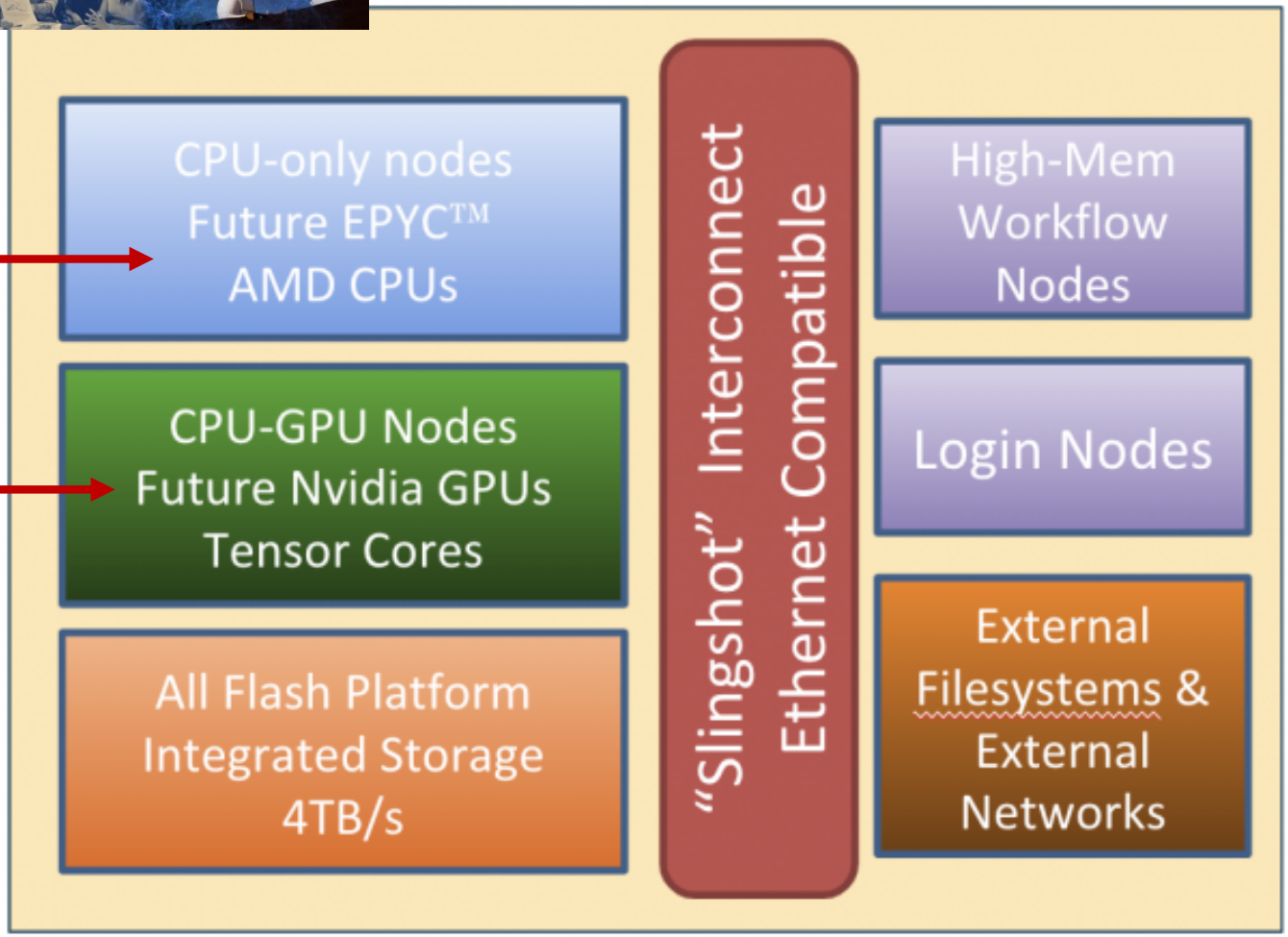
https://wccftech.com/amd-epyc-rome-7nm-server-cpu-official-launch-64-core-128-thread-128-pcie-gen4/

The new Cray cluster for NERSC: Perlmutter

← Design for the set of racks

AMD Epyc Milan
(Zen 3)

Nvidia Volta-next
(next-gen Volta)



Intel Xeon vs. AMD Epyc

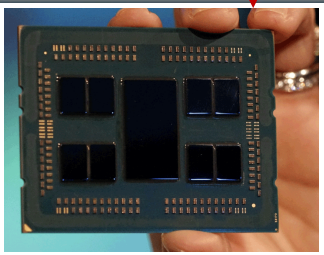
Microarchitecture Comparison						
	Skylake	Cannon Lake	Sunny Cove*		Zen	Zen 2
L1-D Cache	32 KiB/core 8-way	32 KiB/core 8-way	48 KiB/core 12-way		32 KiB/core 8-way	32 KiB/core 8-way
L1-I Cache	32 KiB/core 8-way	32 KiB/core 8-way	32 KiB/core 8-way		64 KiB/core 4-way	32 KiB/core 8-way
L2 Cache	256 KiB/core 4-way	256 KiB/core 4-way	512 KiB/core 8-way		512 KiB/core 8-way	512 KiB/core 8-way
L3 Cache	2 MiB/core 16-way	2 MiB/core 16-way	2 MiB/core 16-way		2 MiB/core	4 MiB/core
L3 Cache Type	Inclusive	Inclusive	Inclusive		Non-Inclusive	Non-Inclusive
Decode	4 + 1	4 + 1	4 + 1		4	4
uOP Cache	1.5k	1.5k	2.25k		2k	4k
Reorder Buffer	224	224	352		192	224
Execution Ports	8	8	10		10	11
AGUs	2 + 1	2 + 1	2 + 2		1 + 1	2 + 1
AVX-512	-	1 x FMA	1 x FMA		-	

** Sunny Cove numbers for Client. Server will have different L2/L3 cache and FMA, like Skylake*

Intel Xeon vs. AMD Epyc

(both launched 1st half 2019)

	AMD EPYC 7742 (Rome)	Intel Xeon Platinum 8280
	7 nm, I/O 14 nm 9-die package 2.25 GHz – 3.4 GHz	14 nm single-die, 2 SNC 2.7 GHz – 4.0 GHz AVX512: 1.8 – 3.7 GHz
Cores / Threads	64c / 128t	28c / 56t
L2/ L3 Cache	512 KiB/core/ 256 MB	1MiB/core/ 38.5 MB
Max Memory /Bandwidth	4 TB /190.7 GiB/s	1 TB/131.13 GiB/s
Memory Channels	8	6
PCI-E Lanes	128x PCI-E 4.0	48x PCI-E 3.0



14 nm
2-die 82xx, 56 cores
12 memory channels

Outline

1. TOP500
   a) TOP10 lists from Nov'17 to Nov'19
   b) Country distribution over the past 25 years
   c) PU chip technology evolution in the past 25 years and since last year
   d) Evolution of the accelerators since they were available
   e) Analysis of some relevant systems and architectures
2. GREEN500
   a) TOP10 lists from Nov'17 to Nov'19
   b) Analysis of some relevant systems
3. HPCG500
   a) HPCG *vs.* HPL: an overview
   b) TOP10 lists from Nov'17 to Nov'19
   c) Analysis of some relevant systems
4. GRAPH500
   a) Performance Metric (TEPS)
   b) Breadth-First Search (BFS) & Single Source Shortest Paths (SSSP)
5. And next?...

The Green500 list

About the Green500 List

The Green500 list ranks the top 500 supercomputers in the world by energy efficiency. The focus of performance-at-any-cost computer operations has led to the emergence of supercomputers that consume vast amounts of electrical power and produce so much heat that large cooling facilities must be constructed to ensure proper performance. To address this trend, the Green500 list puts a premium on energy-efficient performance for sustainable supercomputing.

The inaugural Green500 list was announced on November 15, 2007 at SC|07. As a complement to the TOP500, the unveiling of the Green500 ushered in a new era where supercomputers can be compared by performance-per-watt.

While the selection of any power-performance metric will be controversial, we currently opt for "FLOPS-per-Watt" given that it has already become a widely used metric in the community and for

Top 10 greener-HPC systems Nov'17 Green500

Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	Power (kW)	Power Efficiency (GFlops/watts)
1	259	**Shoubu system B** - ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 , PEZY Computing / Exascaler Inc. Advanced Center for Computing and Communication, RIKEN Japan	794,400	842.0	50	17.009
2	307	**Suiren2** - ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 , PEZY Computing / Exascaler Inc. High Energy Accelerator Research Organization /KEK Japan	762,624	788.2	47	16.759
3	276	**Sakura** - ZettaScaler-2.2, Xeon E5-2618Lv3 8C 2.3GHz, Infiniband EDR, PEZY-SC2 , PEZY Computing / Exascaler Inc. PEZY Computing K.K. Japan	794,400	824.7	50	16.657
4	149	**DGX SaturnV Volta** - NVIDIA DGX-1 Volta36, Xeon E5-2698v4 20C 2.2GHz, Infiniband EDR, NVIDIA Tesla V100 , Nvidia NVIDIA Corporation United States	22,440	1,070.0	97	15.113
5	4	**Gyoukou** - ZettaScaler-2.2 HPC system, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 700Mhz , ExaScaler Japan Agency for Marine-Earth Science and Technology Japan	19,860,000	19,135.8	1,350	14.173
6	13	**TSUBAME3.0** - SGI ICE XA, IP139-SXM2, Xeon E5-2680v4 14C 2.4GHz, Intel Omni-Path, NVIDIA Tesla P100 SXM2 , HPE GSIC Center, Tokyo Institute of Technology Japan	135,828	8,125.0	792	13.704
7	195	**AIST AI Cloud** - NEC 4U-8GPU Server, Xeon E5-2630Lv4 10C 1.8GHz, Infiniband EDR, NVIDIA Tesla P100 SXM2 , NEC National Institute of Advanced Industrial Science and Technology Japan	23,400	961.0	76	12.681
8	419	**RAIDEN GPU subsystem** - NVIDIA DGX-1, Xeon E5-2698v4 20C 2.2GHz, Infiniband EDR, NVIDIA Tesla P100 , Fujitsu Center for Advanced Intelligence Project, RIKEN Japan	11,712	635.1	60	10.603
9	115	**Wilkes-2** - Dell C4130, Xeon E5-2650v4 12C 2.2GHz, Infiniband EDR, NVIDIA Tesla P100 , Dell EMC University of Cambridge United Kingdom	21,240	1,193.0	114	10.428
10	3	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	361,760	19,590.0	2,272	10.398

The GREEN 500

Top systems Nov'18 Green500

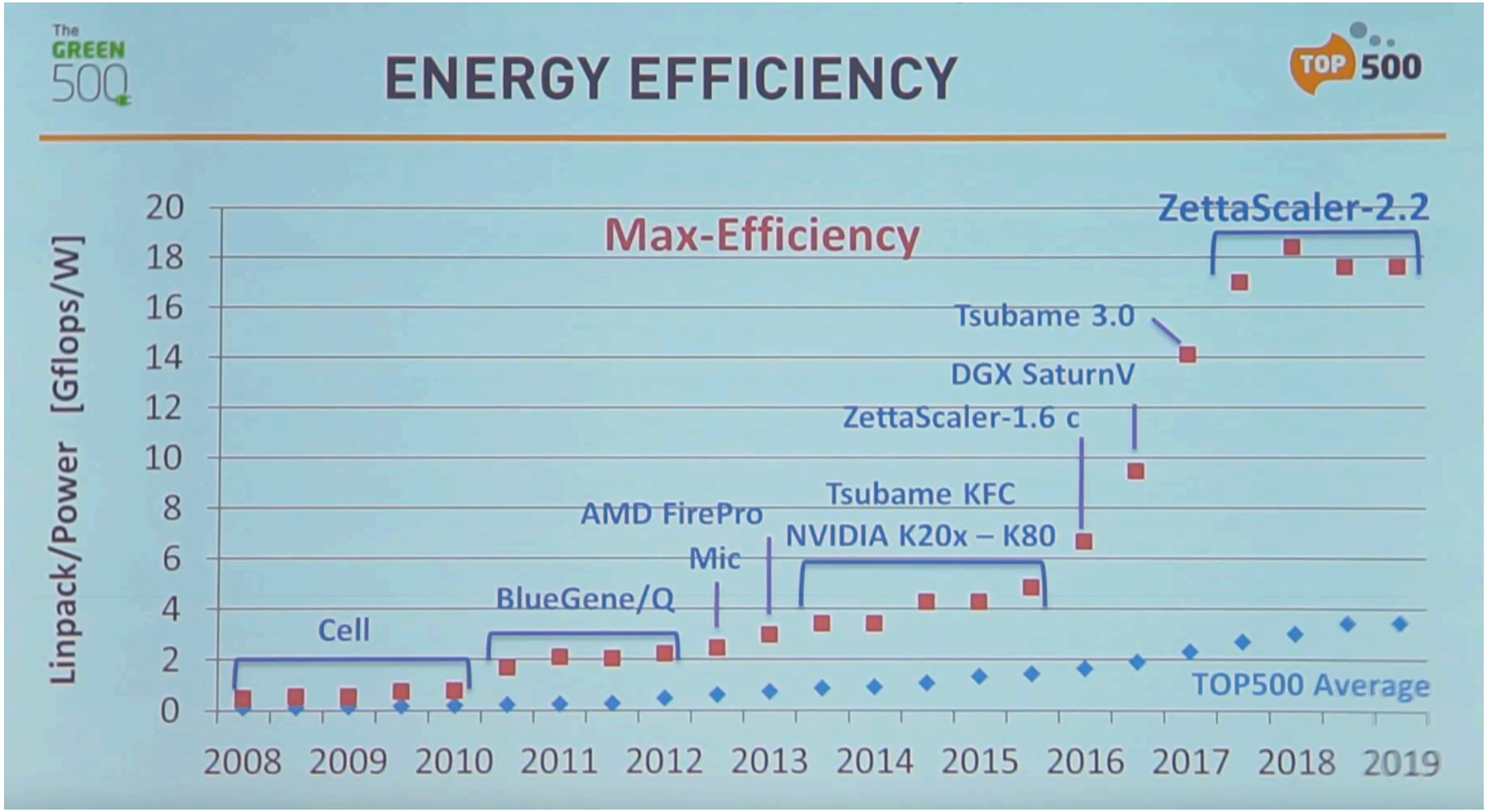
Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	Power (kW)	Power Efficiency (GFlops/watts)
1	375	**Shoubu system B** - ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2, PEZY Computing / Exascaler Inc. Advanced Center for Computing and Communication, RIKEN Japan	953,280	1,063.3	60	17.604
2	374	**DGX SaturnV Volta** - NVIDIA DGX-1 Volta36, Xeon E5-2698v4 20C 2.2GHz, Infiniband EDR, NVIDIA Tesla V100, Nvidia NVIDIA Corporation United States	22,440	1,070.0	97	15.113
3	1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,397,824	143,500.0	9,783	14.668
4	7	**AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2570 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR, Fujitsu National Institute of Advanced Industrial Science and Technology (AIST) Japan	391,680	19,880.0	1,649	14.423
5	22	**TSUBAME3.0** - SGI ICE XA, IP139-SXM2, Xeon E5-2680v4 14C 2.4GHz, Intel Omni-Path, NVIDIA Tesla P100 SXM2, HPE GSIC Center, Tokyo Institute of Technology Japan	135,828	8,125.0	792	13.704
6	2	**Sierra** - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	7,438	12.723
7	446	**AIST AI Cloud** - NEC 4U-8GPU Server, Xeon E5-2630Lv4 10C 1.8GHz, Infiniband EDR, NVIDIA Tesla P100 SXM2, NEC National Institute of Advanced Industrial Science and Technology Japan	23,400	961.0	76	12.681
8	411	**MareNostrum P9 CTE** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Tesla V100, IBM Barcelona Supercomputing Center Spain	19,440	1,018.0	86	11.865
9	38	**Advanced Computing System(PreE)** - Sugon TC8600, Hygon Dhyana 32C 2GHz, Deep Computing Processor, 200Gb 6D-Torus, Sugon Sugon China	163,840	4,325.0	380	11.382
10	20	**Taiwania 2** - QCT QuantaGrid D52G-4U/LC, Xeon Gold 6154 18C 3GHz, Mellanox InfiniBand EDR, NVIDIA Tesla V100 SXM2, Quanta Computer / Taiwan Fixed Network / ASUS Cloud National Center for High Performance Computing Taiwan	170,352	9,000.0	798	11.285

The GREEN 500

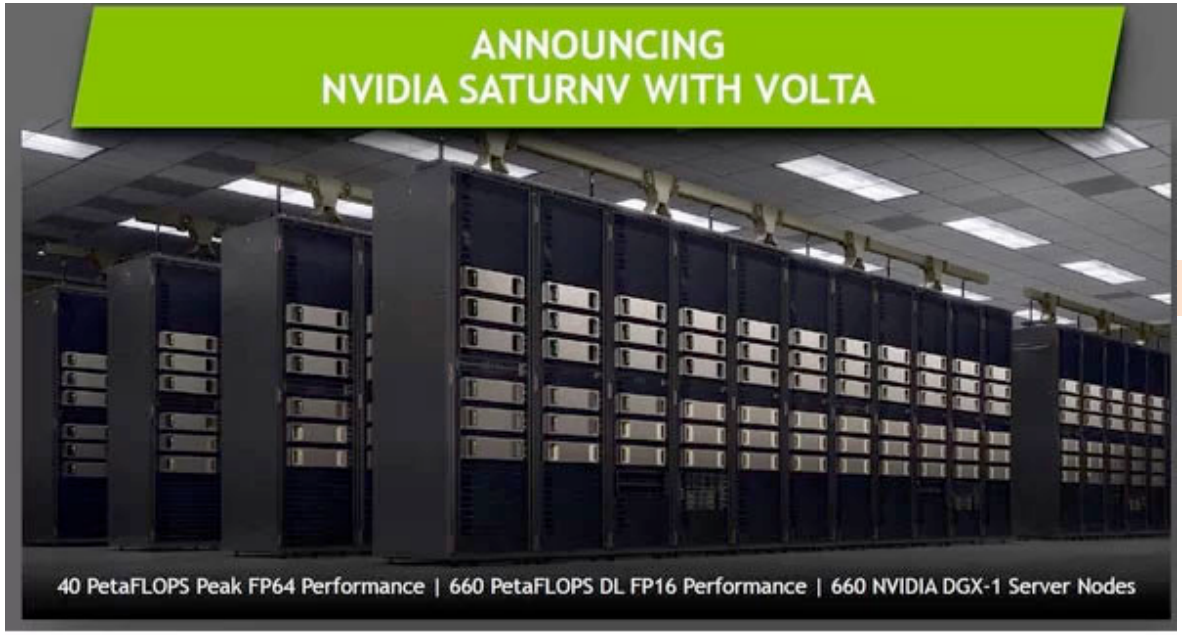
Top systems Nov'19 Green500

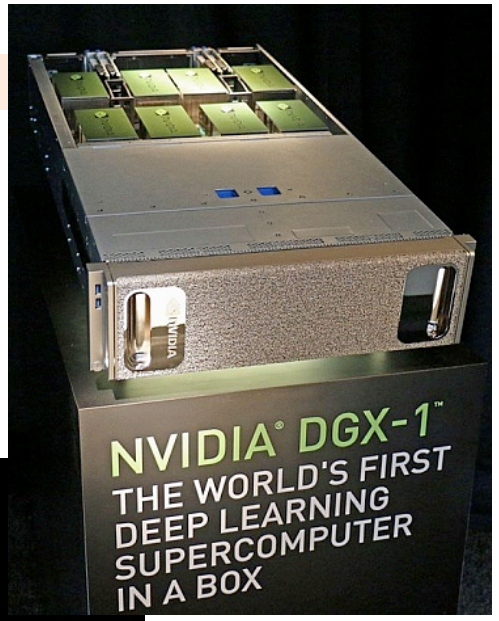
Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	Power (kW)	Power Efficiency (GFlops/watts)
1	159	**A64FX prototype** - Fujitsu A64FX, Fujitsu A64FX 48C 2GHz, Tofu interconnect D , Fujitsu Fujitsu Numazu Plant Japan	36,864	1,999.5	118	16.876
2	420	**NA-1** - ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 700Mhz , PEZY Computing / Exascaler Inc. PEZY Computing K.K. Japan	1,271,040	1,303.2	80	16.256
3	24	**AiMOS** - IBM Power System AC922, IBM POWER9 20C 3.45GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Volta GV100 , IBM Rensselaer Polytechnic Institute Center for Computational Innovations (CCI) United States	130,000	8,045.0	510	15.771
4	373	**Satori** - IBM Power System AC922, IBM POWER9 20C 2.4GHz, Infiniband EDR, NVIDIA Tesla V100 SXM2 , IBM MIT/MGHPCC Holyoke, MA United States	23,040	1,464.0	94	15.574
5	1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	10,096	14.719
6	8	**AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2570 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR , Fujitsu National Institute of Advanced Industrial Science and Technology (AIST) Japan	391,680	19,880.0	1,649	14.423
7	494	**MareNostrum P9 CTE** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Tesla V100 , IBM Barcelona Supercomputing Center Spain	18,360			
8	23	**TSUBAME3.0** - SGI ICE XA, IP139-SXM2, Xeon E5-2680v4 14C 2.4GHz, Intel Omni-Path, NVIDIA Tesla P100 SXM2 , HPE GSIC Center, Tokyo Institute of Technology Japan	135,828	8,125.0	792	13.704
9	11	**PANGEA III** - IBM Power System AC922, IBM POWER9 18C 3.45GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Volta GV100 , IBM Total Exploration Production France	291,024	17,860.0	1,367	13.065
10	2	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	7,438	12.723
11	48	**Advanced Computing System(PreE)** - Sugon TC8600, Hygon Dhyana 32C 2GHz, Deep Computing Processor, 200Gb 6D-Torus , Sugon Sugon China	163,840	4,325.0	380	11.382

Energy efficiency: Green500 in Jun'19

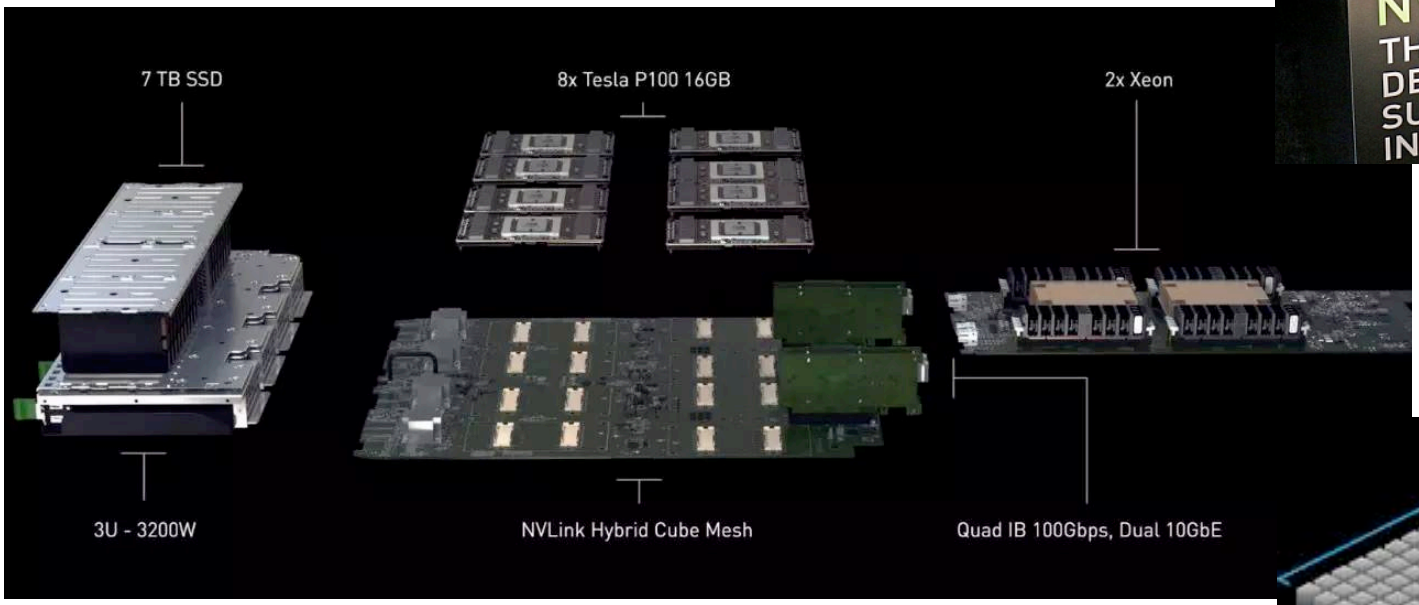


NVidia DGX-1 SaturnV



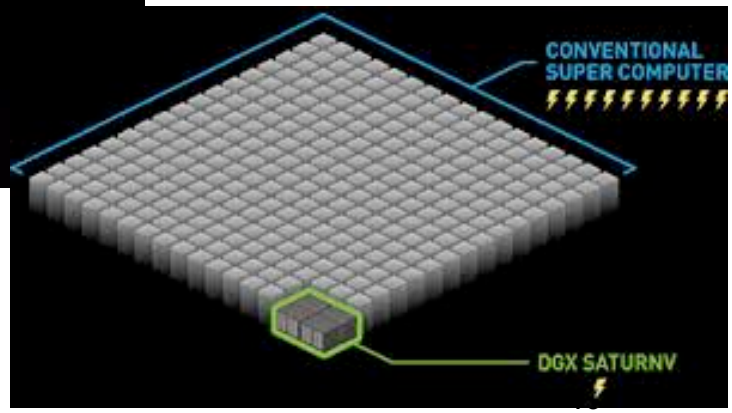


$149,000



2	374	**DGX SaturnV Volta** - NVIDIA DGX-1 Volta36, Xeon E5-2698v4 20C 2.2GHz, Infiniband EDR, NVIDIA Tesla V100 , Nvidia NVIDIA Corporation United States	22,440

Nov'18
Green500

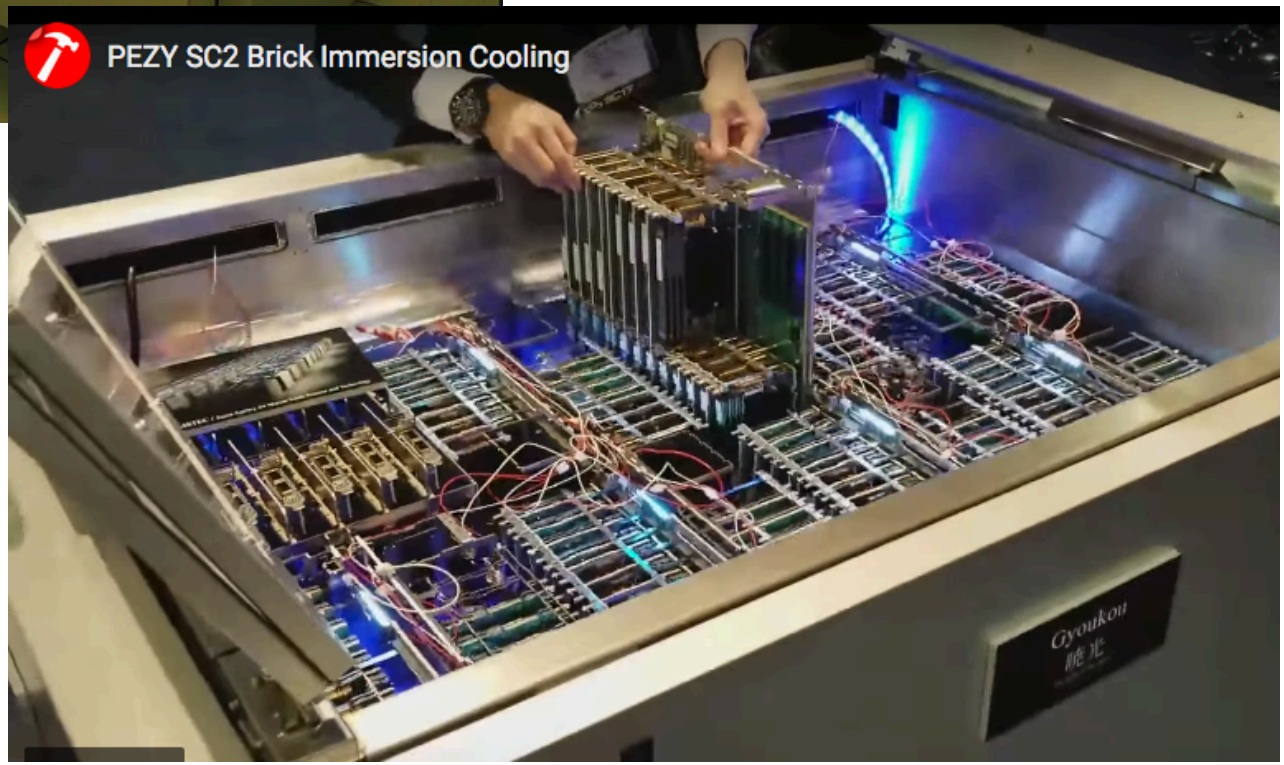


Gyoukou
ZettaScaler-2.2

10,000 PEZY-SC2 + 1,250 16-cores Xeon =
19.84 M PEZY cores + 20 K Xeon cores

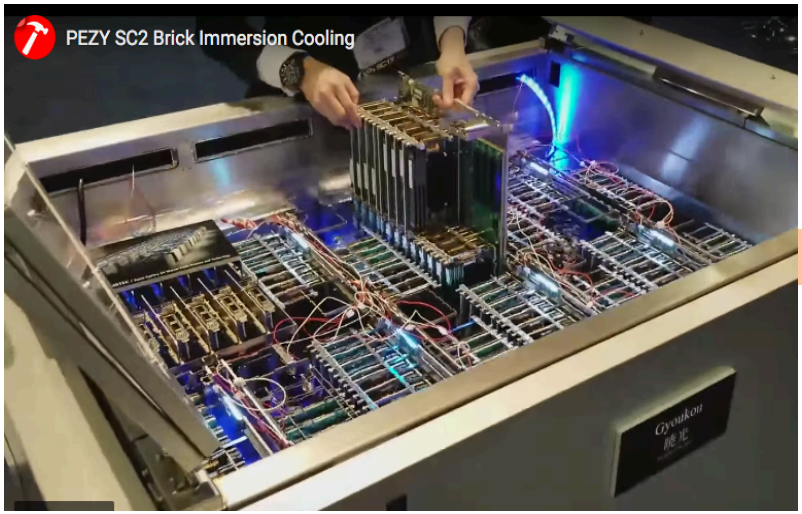
5	4	**Gyoukou** - ZettaScaler-2.2 HPC system, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 700Mhz , ExaScaler	19,860,000	19,1

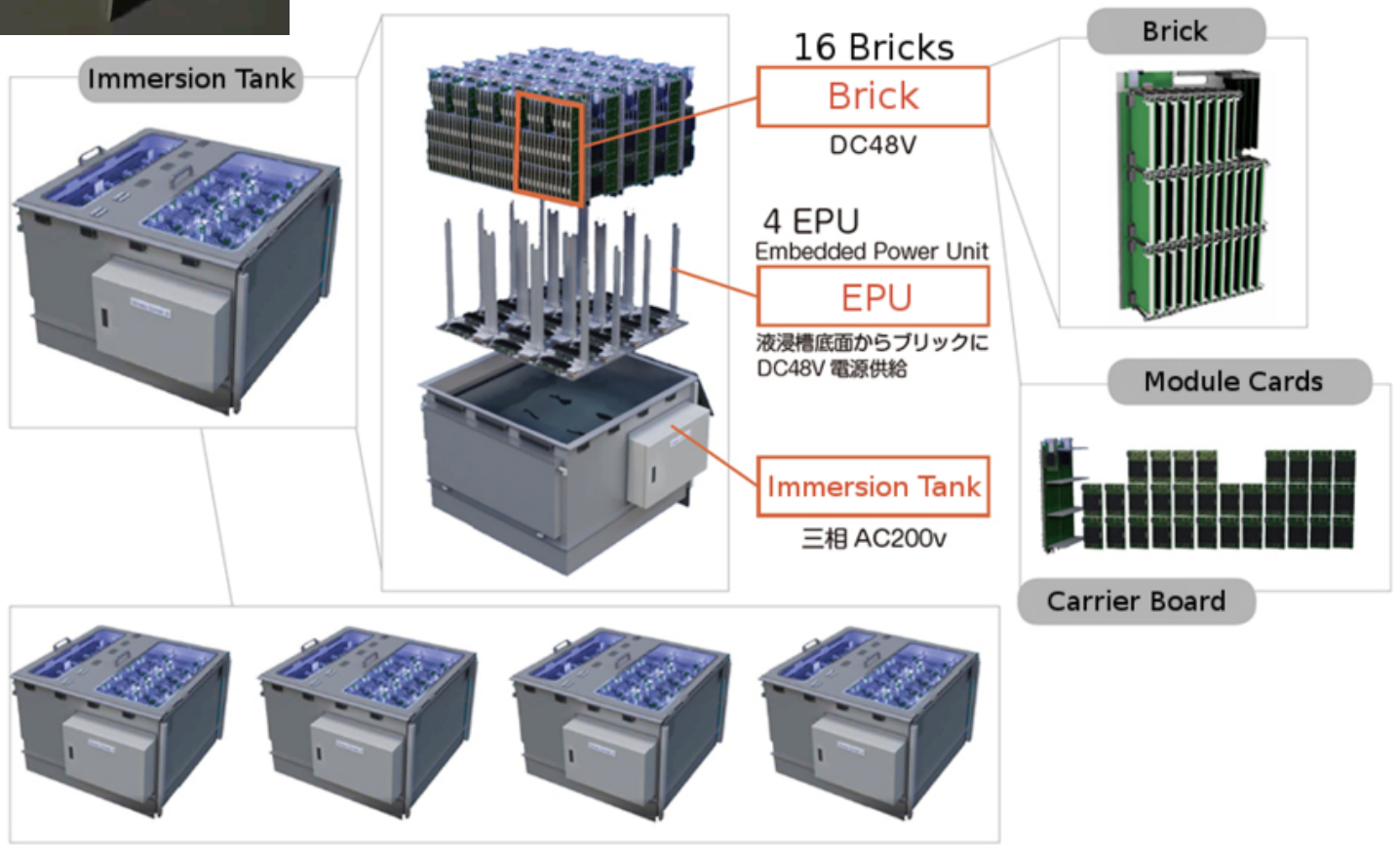
Nov'17 Green500



20 immersion tanks
each tank 16 bricks
each brick 32 PEZY
each PEZY ~2K
8-way SMT cores
=>
each tank ~1M cores

Gyoukou
ZettaScaler-2.2





Scalable (Internetwork of multiple immersion tanks)

Advanced Computing System (PreE)

(the Hygon Dhyana x86 device in a Sugon cluster)

tom's HARDWARE

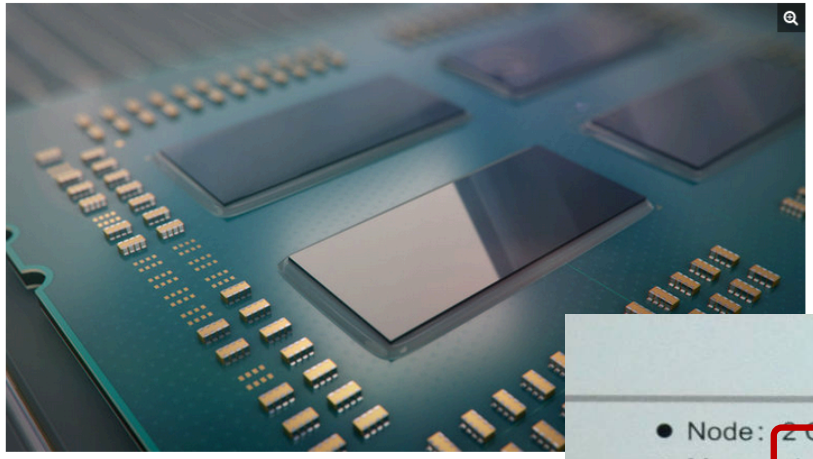
China Finds Zen: Begins Production Of x86 Processors Based On AMD's IP

by Paul Alcorn July 6, 2018 at 1:18 PM

海光 HYGON

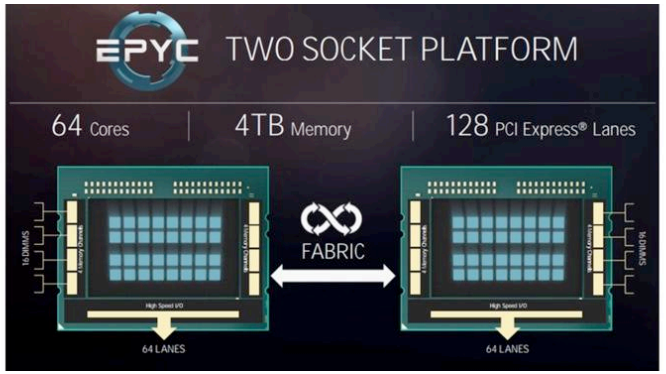
11	48	Advanced Computing System(PreE) - Sugon TC8600, Hygon Dhyana 32C 2GHz, Deep Computing Processor, 200Gb 6D-Torus , Sugon Sugon China	163,840

Nov'19 Green500



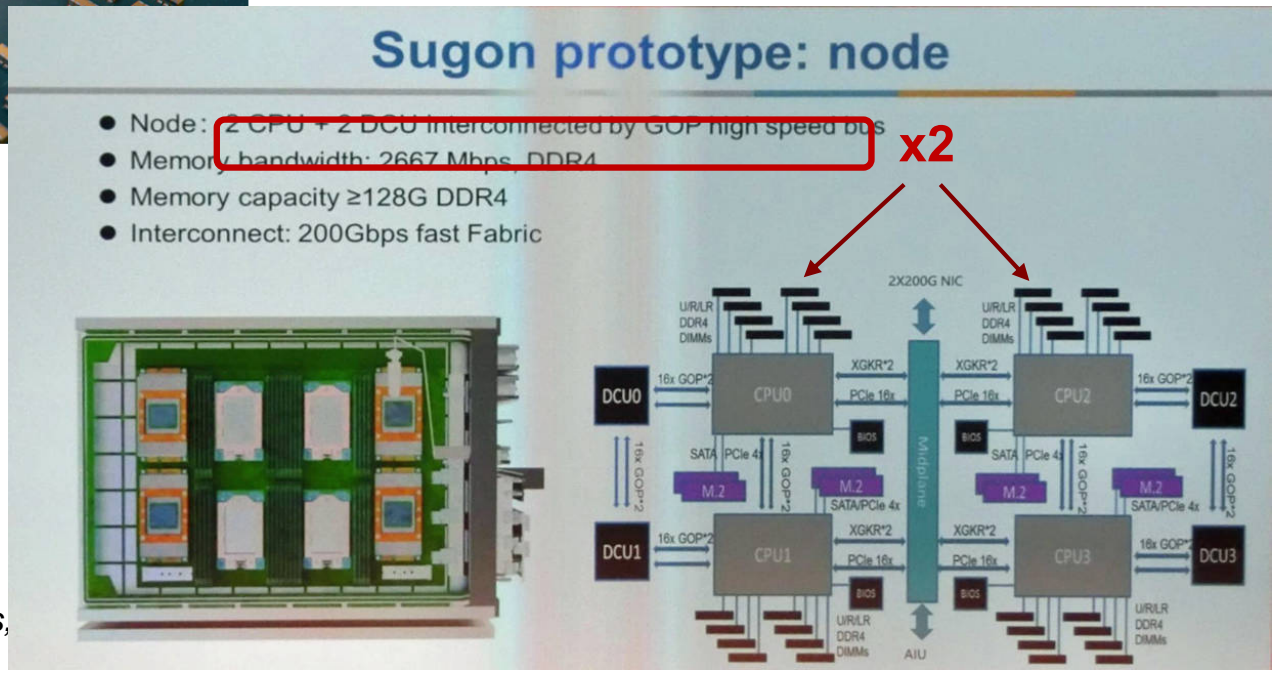
Hygon Dhyana:
32-core 2.0GHz

Similar to:
AMD EPYC 7501



"CPU":
32-core Hygon Dhyana

DCU:
GPU accelerator



Outline

1. TOP500
   a) TOP10 lists from Nov'17 to Nov'19
   b) Country distribution over the past 25 years
   c) PU chip technology evolution in the past 25 years and since last year
   d) Evolution of the accelerators since they were available
   e) Analysis of some relevant systems and architectures
2. GREEN500
   a) TOP10 lists from Nov'17 to Nov'19
   b) Analysis of some relevant systems
3. HPCG500
   a) HPCG *vs.* HPL: an overview
   b) TOP10 lists from Nov'17 to Nov'19
   c) Analysis of some relevant systems
4. GRAPH500
   a) Performance Metric (TEPS)
   b) Breadth-First Search (BFS) & Single Source Shortest Paths (SSSP)
5. And next?...

TOP500: HPCG vs. HPL

HPCG UPDATE: ISC'17

Jack Dongarra
Michael Heroux
Piotr Luszczek

HPCG Snapshot

- High Performance Conjugate Gradients (HPCG).
- Solves $Ax=b$, A large, sparse, b known, x computed.
- An optimized implementation of PCG contains essential computational and communication patterns that are prevalent in a variety of methods for discretization and numerical solution of PDEs

- Patterns:
  - Dense and sparse computations.
  - Dense and sparse collectives.
  - Multi-scale execution of kernels via MG (truncated) V cycle.
  - Data-driven parallelism (unstructured sparse triangular solves).
- Strong verification (via spectral properties of PCG).

TOP500: HPCG vs. HPL

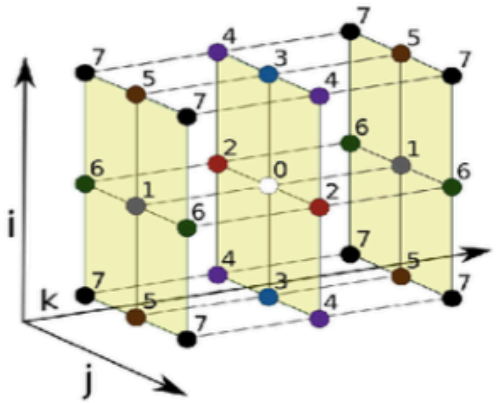
HPCG UPDATE: ISC'17

Jack Dongarra
Michael Heroux
Piotr Luszczek

hpcg-benchmark.org

Model Problem Description

- Synthetic discretized 3D PDE (FEM, FVM, FDM).
- Zero Dirichlet BCs, Synthetic RHS s.t. solution = 1.
- Local domain: $(n_x \times n_y \times n_z)$
- Process layout: $(np_x \times np_y \times np_z)$
- Global domain: $(n_x * np_x) \times (n_y * np_y) \times (n_z * np_z)$
- Sparse matrix:
  - 27 nonzeros/row interior.
  - 8 – 18 on boundary.
  - Symmetric positive definite.



27-point stencil operator

TOP500: HPCG vs. HPL

HPCG UPDATE: ISC'17

Jack Dongarra
Michael Heroux
Piotr Luszczek

Merits of HPCG

- Includes major communication/computational patterns.
  - Represents a minimal collection of the major patterns.
- Rewards investment in:
  - High-performance collective ops.
  - Local memory system performance.
  - Low latency cooperative threading.
- Detects/measures variances from bitwise reproducibility.
- Executes kernels at several (tunable) granularities:
  - nx = ny = nz = 104 gives
  - nlocal = 1,124,864; 140,608; 17,576; 2,197
  - ComputeSymGS with multicoloring adds one more level:
    - 8 colors.
    - Average size of color = 275.
    - Size ratio (largest:smallest): 4096
  - Provide a "natural" incentive to run a big problem.
- Full performance discussion:
  - http://www.hpcg-benchmark.org -> "Performance Overview" tab.

TOP500: HPCG vs. HPL

HPCG Benchmark

The High Performance Conjugate Gradients (HPCG) Benchmark project is an effort to create a new metric for ranking HPC systems. HPCG is intended as a complement to the High Performance LINPACK (HPL) benchmark, currently used to rank the TOP500 computing systems. The computational and data access patterns of HPL are still representative of some important scalable applications, but not all. HPCG is designed to exercise computational and data access patterns that more closely match a different and broad set of important applications, and to give incentive to computer system designers to invest in capabilities that will have impact on the collective performance of these applications.

HPCG is a complete, stand-alone code that measures the performance of basic operations in a unified code:

- Sparse matrix-vector multiplication.
- Vector updates.
- Global dot products.
- Local symmetric Gauss-Seidel smoother.
- Sparse triangular solve (as part of the Gauss-Seidel smoother).
- Driven by multigrid preconditioned conjugate gradient algorithm that exercises the key kernels on a nested set of coarse grids.
- Reference implementation is written in C++ with MPI and OpenMP support.

http://www.hpcg-benchmark.org

Top 10 HPC systems Nov'17 HPCG-500

Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	HPCG (TFlop/s)
1	10	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect , Fujitsu RIKEN Advanced Institute for Computational Science (AICS) Japan	705,024	10,510.0	11,280.4	602.736
2	2	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P , NUDT National Super Computer Center in Guangzhou China	3,120,000	33,862.7	54,902.4	580.109
3	7	**Trinity** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	979,968	14,137.3	43,902.6	546.124
4	3	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	361,760	19,590.0	25,326.3	486.398
5	1	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	480.8
6	9	**Oakforest-PACS** - PRIMERGY CX1640 M1, Intel Xeon Phi 7250 68C 1.4GHz, Intel Omni-Path , Fujitsu Joint Center for Advanced High Performance Computing Japan	556,104	13,554.6	24,913.5	385.479
7	8	**Cori** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/SC/LBNL/NERSC United States	622,336	14,014.7	27,880.7	355.442
8	6	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom , IBM DOE/NNSA/LLNL United States	1,572,864	17,173.2	20,132.7	330.373
9	5	**Titan** - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x , Cray Inc. DOE/SC/Oak Ridge National Laboratory United States	560,640	17,590.0	27,112.5	322.322
10	13	**TSUBAME3.0** - SGI ICE XA, IP139-SXM2, Xeon E5-2680v4 14C 2.4GHz, Intel Omni-Path, NVIDIA Tesla P100 SXM2 , HPE GSIC Center, Tokyo Institute of Technology Japan	135,828	8,125.0	12,127.1	188.6

HPCG List for November 2018

Top 10 systems Nov'18

Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	HPCG (TFlop/s)
1	1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory United States	2,397,824	143,500.0	2925.75
2	2	**Sierra** - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	1795.67
3	18	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect , Fujitsu RIKEN Advanced Institute for Computational Science (AICS) Japan	705,024	10,510.0	602.74
4	6	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	979,072	20,158.7	546.12
5	7	**AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2570 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR , Fujitsu National Institute of Advanced Industrial Science and Technology (AIST) Japan	391,680	19,880.0	508.85
6	5	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	387,872	21,230.0	496.98
7	3	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	480.85
8	13	**Nurion** - Cray CS500, Intel Xeon Phi 7250 68C 1.4GHz, Intel Omni-Path , Cray Inc. Korea Institute of Science and Technology Information Korea, South	570,020	13,929.3	391.45
9	14	**Oakforest-PACS** - PRIMERGY CX1640 M1, Intel Xeon Phi 7250 68C 1.4GHz, Intel Omni-Path , Fujitsu Joint Center for Advanced High Performance Computing Japan	556,104	13,554.6	385.48
10	12	**Cori** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/SC/LBNL/NERSC United States	622,336	14,014.7	355.44

HPCG List for November 2019

10 systems Nov'19

Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	HPCG (TFlop/s)	
1	1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	2925.75	
2	2	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	1795.67	
3	7	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray/HPE DOE/NNSA/LANL/SNL United States	979,072	20,158.7	546.12	
4	8	**AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2570 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR , Fujitsu National Institute of Advanced Industrial Science and Technology (AIST) Japan	391,680	19,880.0	508.85	
5	6	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray/HPE Swiss National Supercomputing Centre (CSCS) Switzerland	387,872	21,230.0	496.98	
6	3	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	480.85	
7	14	**Nurion** - Cray CS500, Intel Xeon Phi 7250 68C 1.4GHz, Intel Omni-Path , Cray/HPE Korea Institute of Science and Technology Information Korea, South	570,020	13,929.3	391.45	
8	15	**Oakforest-PACS** - PRIMERGY CX1640 M1, Intel Xeon Phi 7250 68C 1.4GHz, Intel Omni-Path , Fujitsu Joint Center for Advanced High Performance Computing Japan	556,104	13,554.6	385.48	
26	198	**Astra** - Apollo 70, Marvell ThunderX2 ARM CN9975-2000 28C 2GHz, 4xEDR Infiniband , HPE Sandia National Laboratories United States	143,640	1,833.0	90.90	

Fujitsu K computer

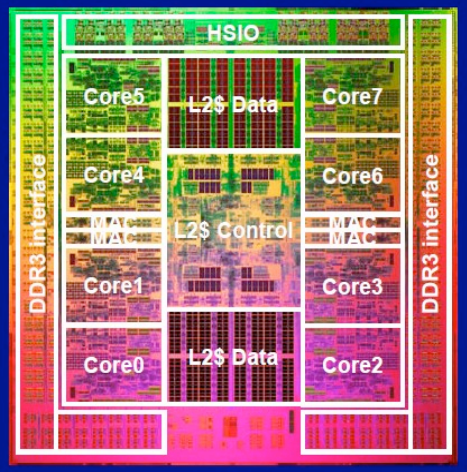
(the Japanese word "kei" (京) means 10 quadrillion, 10^{16})

		K computer
CPU	Name	SPARC64™ VIIIfx
	Performance	128GFlops@2GHz
	Architecture	SPARC V9 + HPC-ACE extension
	Cache configuration	L1(I) Cache:32KB/core, L1(D) Cache:32KB/core
		L2 Cache: 6MB(shared)
	No. of cores/socket	8
	Memory band width	64 GB/s.
Node	Configuration	1 CPU / Node
	Memory capacity	16 GB
System board	Node/system board	4 Nodes
Rack	System board/rack	24 System boards
	Performance/rack	12.3 TFlops

TOP500 The List.

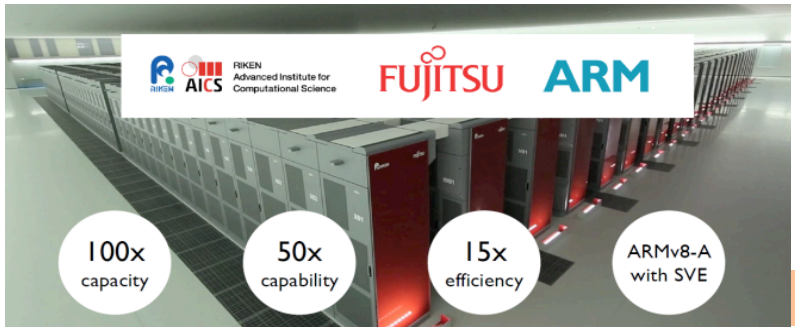
Jun'11: #1
Nov'11: #1
Jun'12: #2
Nov'12: #3
Jun'13: #4
Nov'13: #4
Jun'14: #4
Nov'14: #4
Jun'15: #4
Nov'15: #4
Jun'16: #5
Nov'16: #7
Jun'17: #8
Nov'17: #10
Jun'18: #16
Nov'18: #18
Jun'19: #20

SPARC64™ VIIIfx Chip Overvie



- **Architecture Featu**
  - 8 cores
  - Shared 5 MB L2$
  - Embedded Memory
  - 2 GHz
- **Fujitsu 45nm CMOS**
  - 22.7mm x 22.6mm
  - 760M transistors
  - 1271 signal pins
- **Performance (peak)**
  - 128GFlops
  - 64GB/s memory throughput
- **Power**
  - 58W (TYP, 30℃)
  - Water Cooling – Low leakage power and High reliability

SPARC64™ VIIIfx 12



Next PU generations for HPC: Fujitsu, from SPARC64 to ARMv8

Post-K: Powered by Fujitsu-designed CPU & Tofu

FUJITSU

- Fujitsu CPU cores support the ARM SVE instruction set architecture
- Post-K Fujitsu CPU cores & Tofu maintain the programming models and provide high application performance
- ARM's standard frameworks (SBSA, etc.) assure compatibility among platforms
- FP16 ("giant vector throughput") for supercomputers

	Functions & archtecture	Post-K	FX100	FX10	K
CPU Core	Instruction set architecture	ARMv8-A	SPARC V9		
	SIMD width	512bit	256bit	128bit	128bit
	Double precision (64bit)	✔	✔	✔	✔
	Single precision (32bit)	✔	✔	✔	✔
	Half precision (16bit)	✔	-	-	-
Interconnect	Tofu interconnect	Tofu3	Tofu2	Tofu	Tofu

ARM brand: a bit of history...

ARM architecture

From Wikipedia, the free encyclopedia

ARM, previously **Advanced RISC Machine**, originally **Acorn RISC Machine**, is a family of reduced instruction set computing (RISC) architectures for computer processors, configured for various environments. Arm Holdings develops the architecture and licenses it to other companies, who design their own products that implement one of those architectures—including systems-on-chips (SoC) and systems-on-modules (SoM) that incorporate memory, interfaces, radios, etc. It also designs cores that implement this instruction set and licenses these designs to a number of companies that incorporate those core designs into their own products.

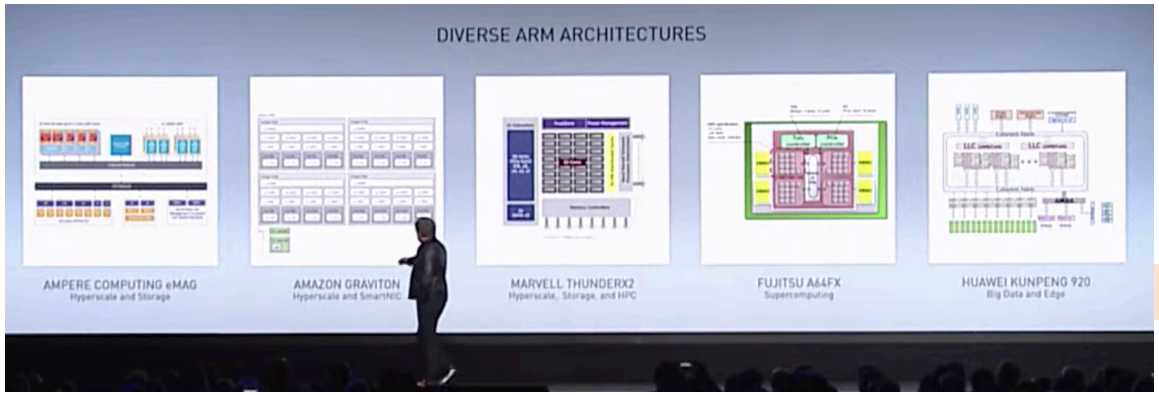
Processors that have a RISC architecture typically require fewer transistors than those with a complex instruction set computing (CISC) architecture (such as the x86 processors found in most personal computers), which

ARM architectures

The ARM logo

Designer	Arm Holdings
Bits	32-bit, 64-bit
Introduced	1985; 34 years ago
Design	RISC
Type	Register-Register
Branching	Condition code, compare and branch
Open	Proprietary

HPCs with ARMv8: server-level competitors

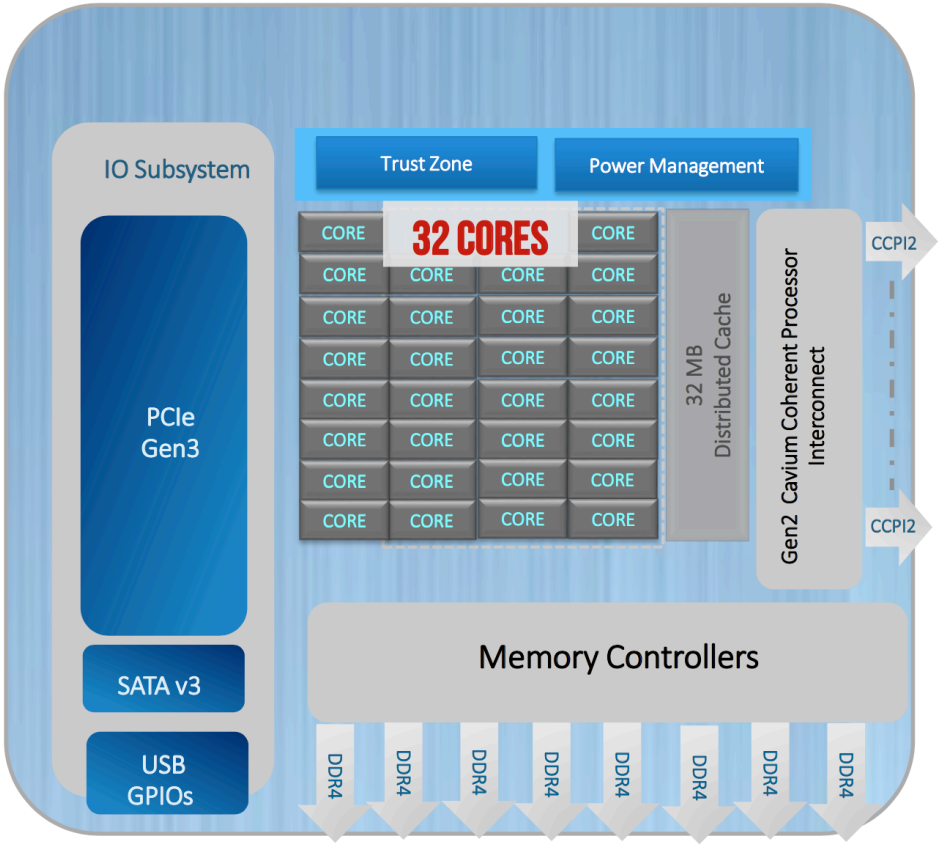


1. Marvell ThunderX product family
2. Fujitsu A64FX Arm chip
3. Ampere eMAG 8180 Arm Processor
4. Neoverse N1 hyperscale reference design
5. Huawei HiSilicon Kunpeng 920

https://www.eweek.com/pc-hardware/eweek-s-top-vendors-server-processor-makers

The Marvell/Cavium ThunderX2

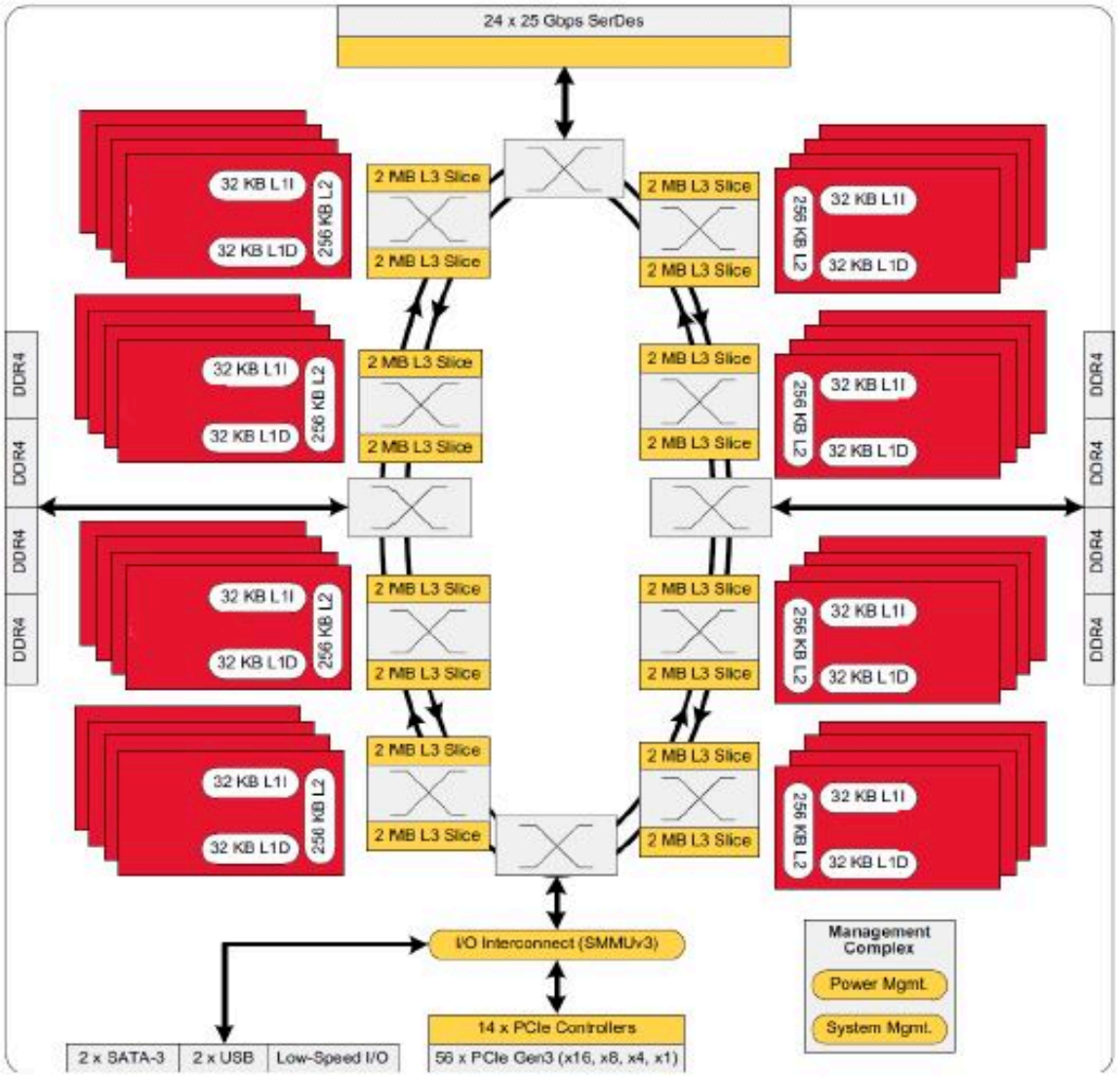
THUNDERX2® Family Key Features



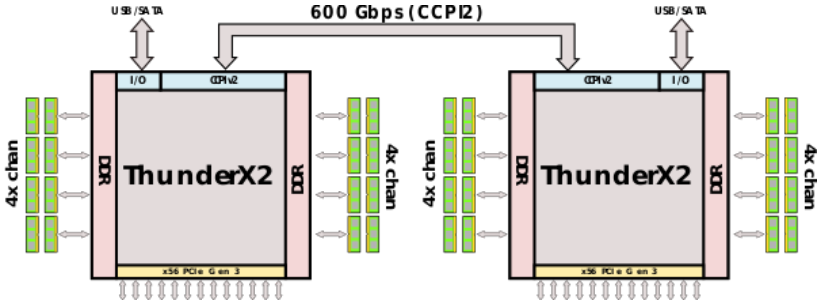
- Up to 32 custom Armv8.1 cores, up to 2.5GHz
- Full OoO, 1, 2, 4 threads per core
- 1S and 2S Configuration
- Up to 8 DDR4-2667 Memory Controllers, 1 & 2 DPC
- Up to 56 lanes of PCIe, 14 PCIe controllers
- Full SoC: Integrated SATAv3 USB3 and GPIOs
- Server class RAS & Virtualization
- Extensive Power Management
- LGA and BGA for most flexibility
- 40+ SKUs
- Volume SKU List Price: $1795 (180W) to $800 (75W)

https://www.hpcwire.com/2018/05/07/cavium-announces-ga-for-thunderx2-touts-oem-support/

The Marvell/Cavium ThunderX2 architecture block diagram

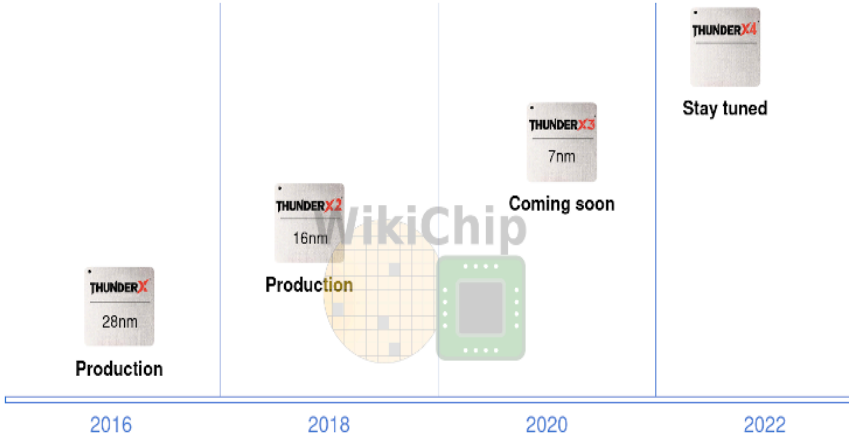


Scalability

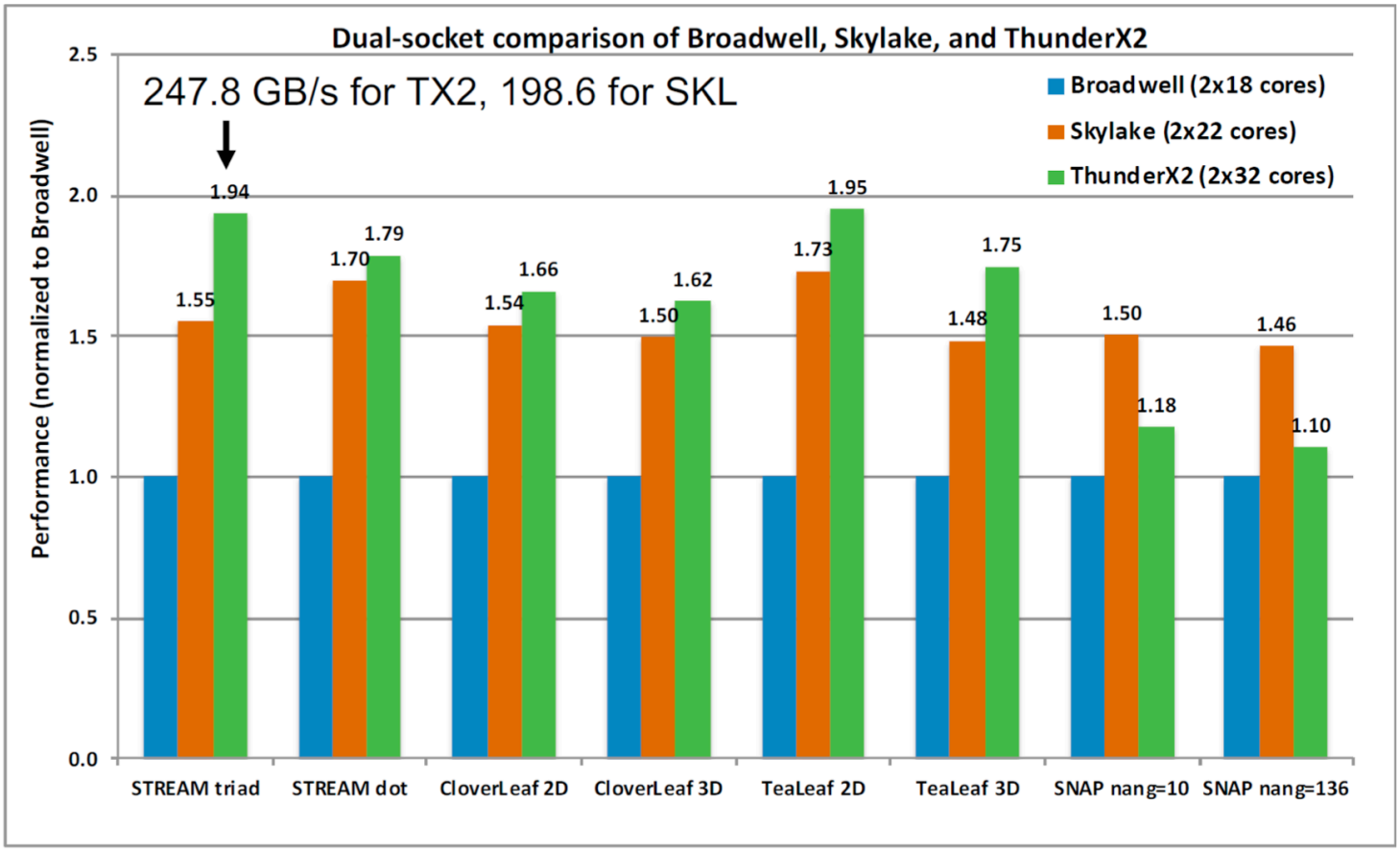


ThunderX® roadmap

Driving >2X generational performance improvement



Broadwell, Skylake & ThunderX2



https://www.hpcwire.com/2018/05/07/cavium-announces-ga-for-thunderx2-touts-oem-support/

Sandia National Laboratories

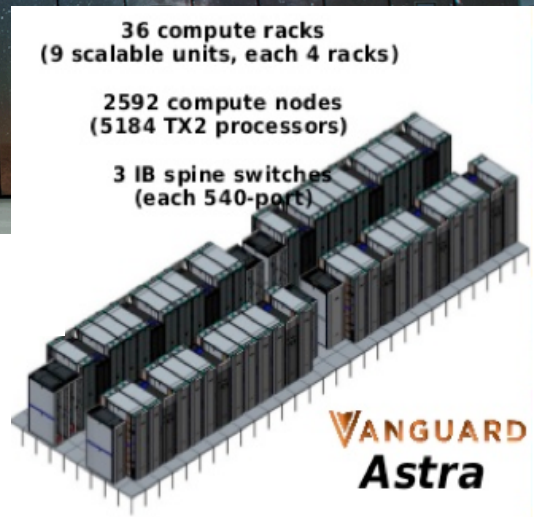
HPC systems with ARMv8: Astra Apollo 70

26 198

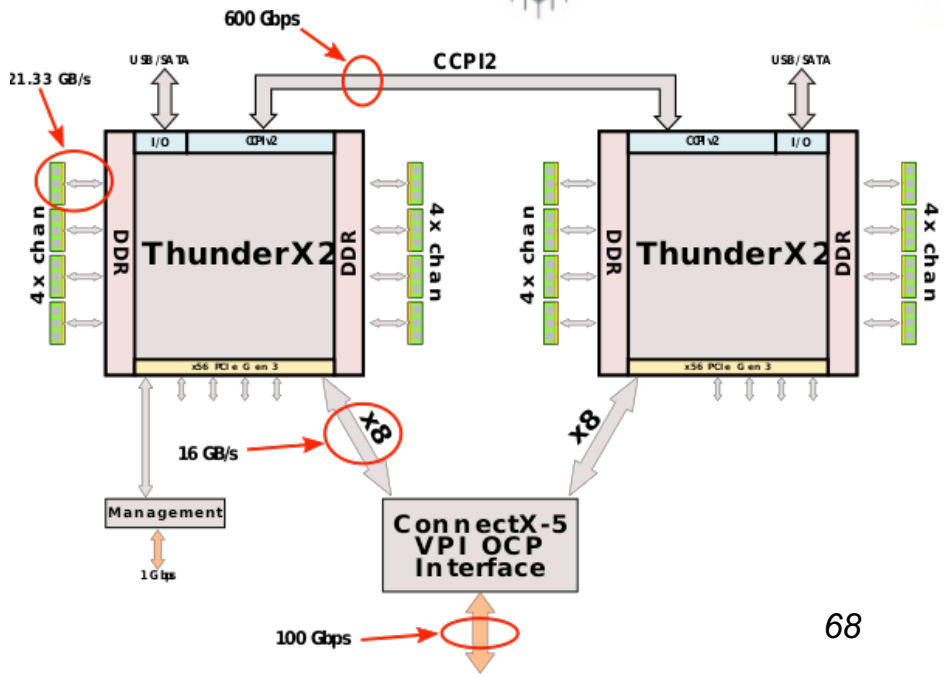
TOP 500 The List.

Astra - Apollo 70, Marvell ThunderX2 ARM CN9975-2000 28C 2GHz
4xEDR Infiniband , HPE
Sandia National Laboratories
United States

HPCG



Dual-socket ThunderX2 CN9975
28-core @ 2.0GHz
8 DDR4 controllers/socket
1x 8GiB DDR4-2666 DIMM/socket



HPC systems with ARMv8: Cray & BSC are also betting on ThunderX2

CRAY

NEWS RELEASE

CRAY CATAPULTS ARM-BASED PROCESSORS INTO SUPERCOMPUTING

Cray Adds Arm Processors with Complete Software Stack to the Cray XC50 Supercomputer

SEATTLE, Nov. 13, 2017 (GLOBE NEWSWIRE) -- Global supercomputer leader Cray Inc. (Nasdaq:CRAY) today announced the Company is creating the world's first production-ready, Arm®-based supercomputer with the addition of Cavium (Nasdaq:CAVM) ThunderX2™ processors, based on 64-bit Armv8-A architecture, to the Cray® XC50™ supercomputer. Cray customers will have a complete Arm-based supercomputer that features a full software environment, including the Cray Linux Environment, the Cray Programming Environment, and Arm-optimized compilers, libraries, and tools for running today's supercomputing workloads.

GW4, THE MET OFFICE, AND CRAY POWER UP THE LARGEST ARM-BASED SUPERCOMPUTER IN EUROPE

"Isambard" Supercomputer to Be Used for Scientific Research and Building the Next Generation of Computing Technology in the Exascale Era

SEATTLE, Nov. 11, 2018 (GLOBE NEWSWIRE) -- GW4, the Met Office and global supercomputer leader Cray Inc. (Nasdaq:CRAY) today announced that the Arm®-based supercomputer in Europe, named "Isambard," is now live in the United Kingdom. It is the largest Arm-based system in the world outside of the

Isambard System Details

The liquid-cooled Cray XC50 includes the Cray-developed Aries™ interconnect with Dragonfly topology. Each node includes two Marvell Arm ThunderX2 32-core,

About the GW4 Alliance

The GW4 Alliance brings together four of the most research-intensive and innovative universities in the UK: Bath, Bristol, Cardiff and Exeter. From the

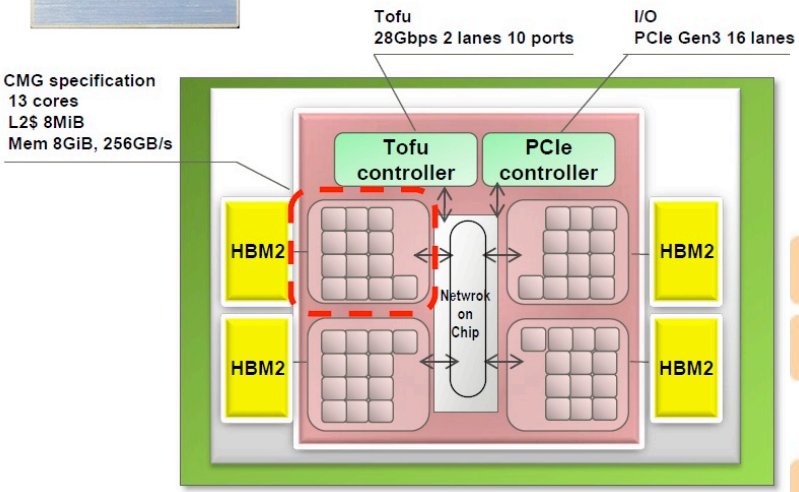
Bull sequana compute blade: X1310

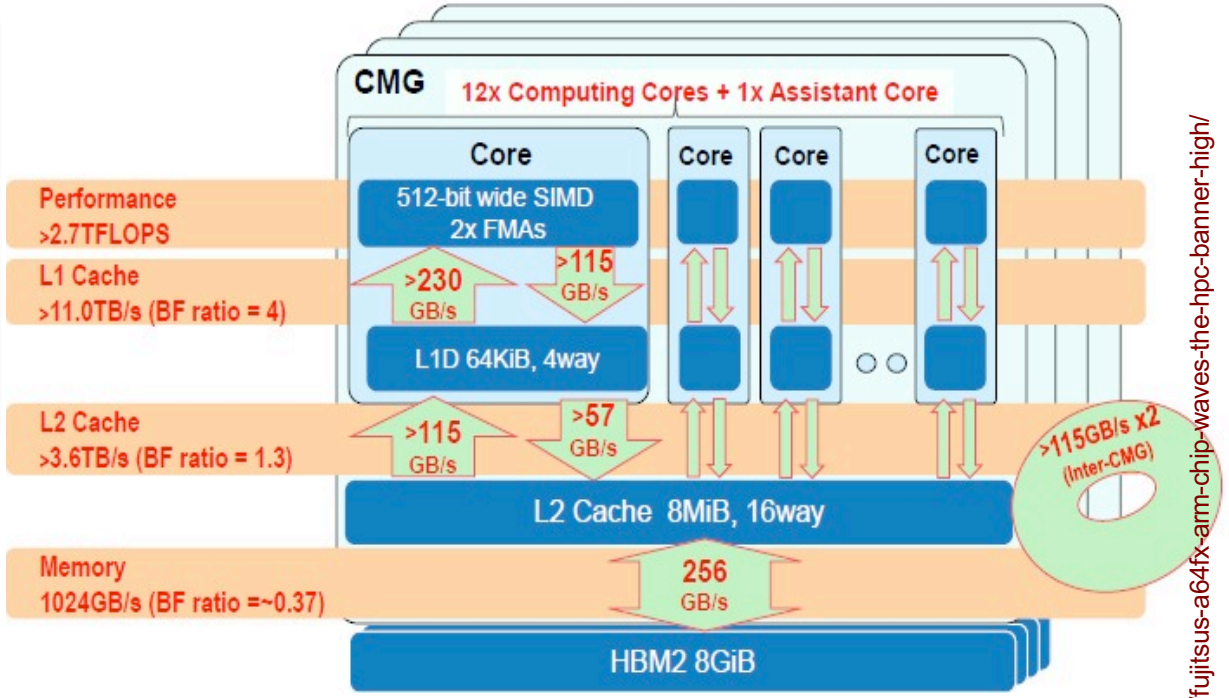
Cavium ThunderX2™ – ARMv8 processor

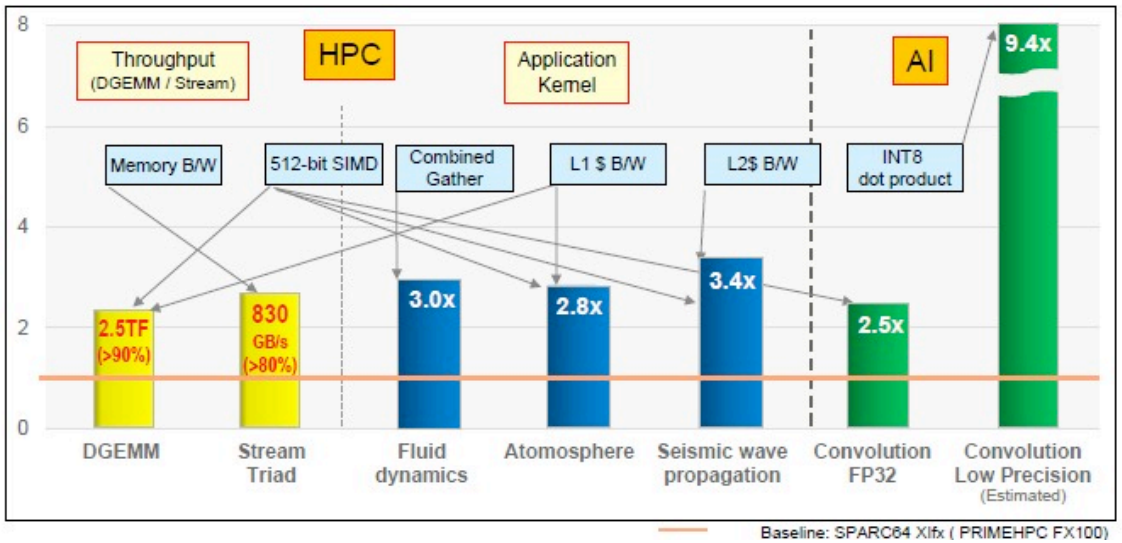
MONT BLANC

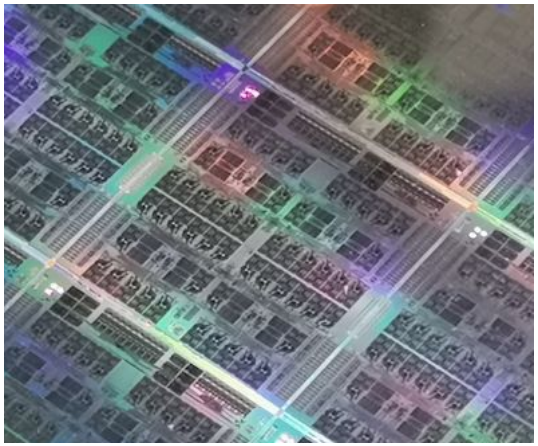
BSC **Barcelona Supercomputing Center**
Centro Nacional de Supercomputación

Fujitsu's A64FX ARM Chip

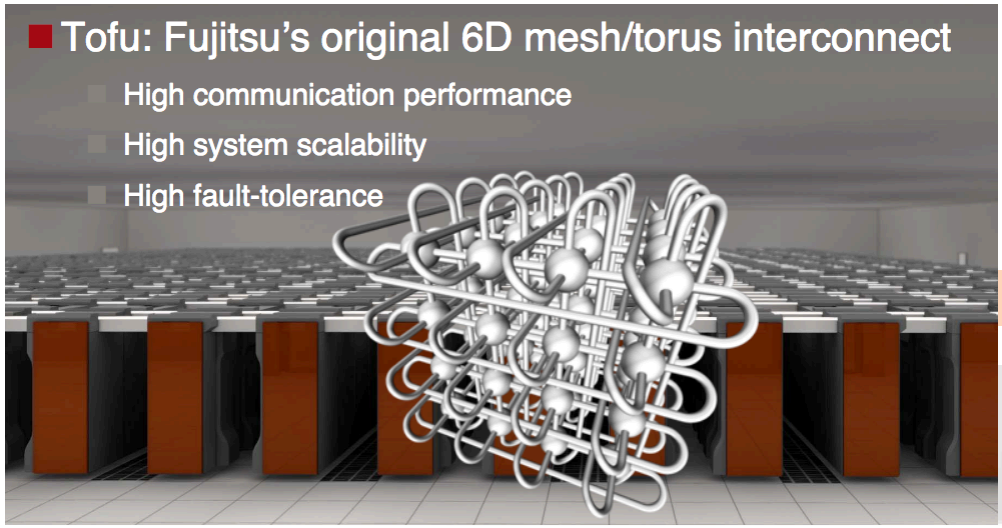








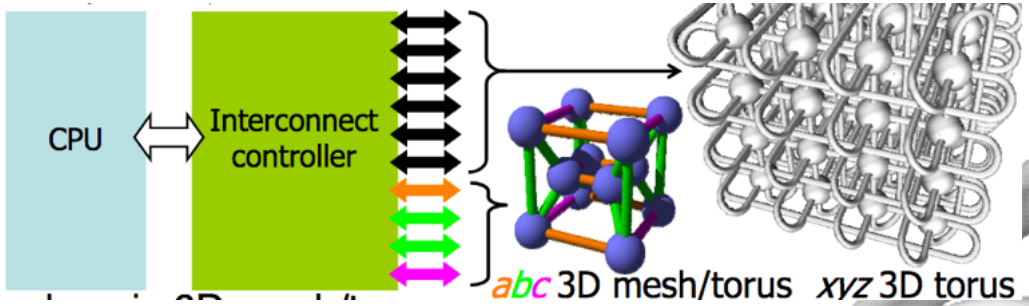
https://www.nextplatform.com/2018/08/24/fujitsus-a64fx-arm-chip-waves-the-hpc-banner-high/

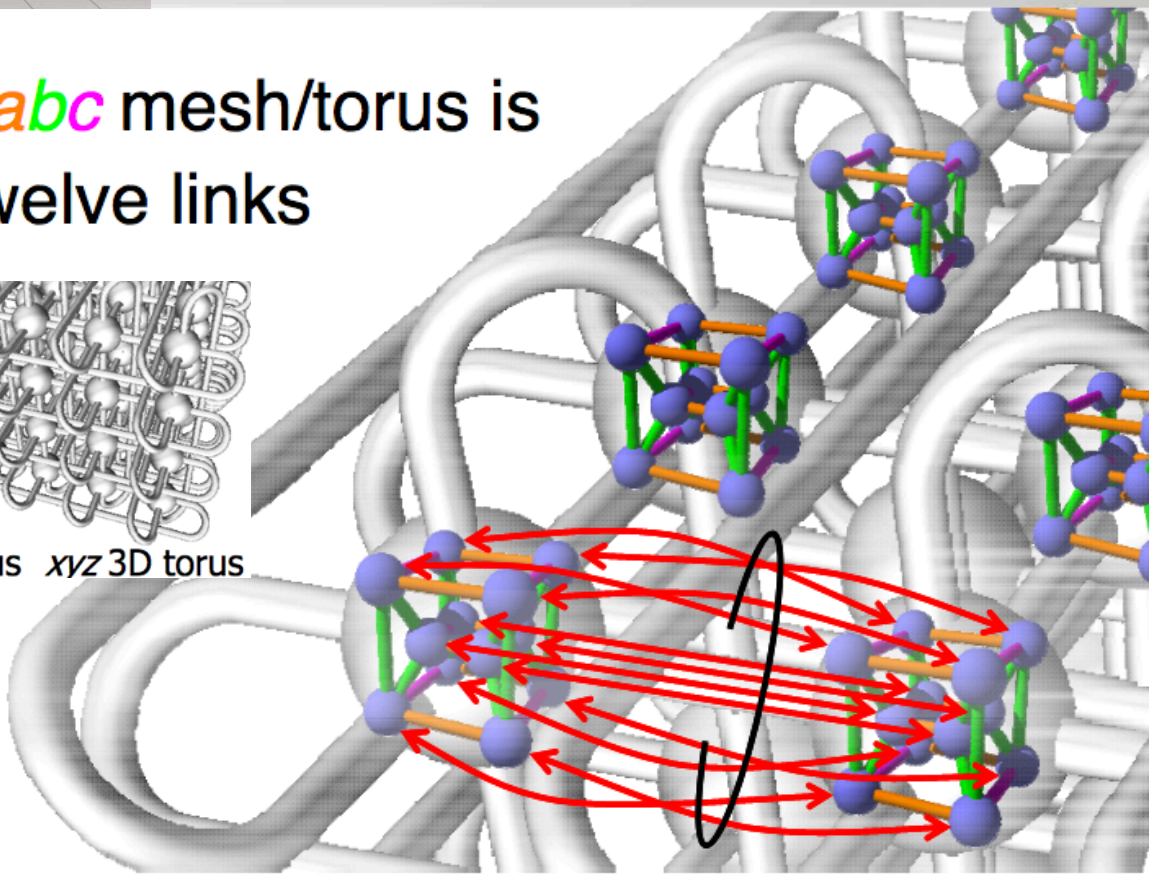


Tofu3: 6D mesh/torus interconnect

FUJITSU

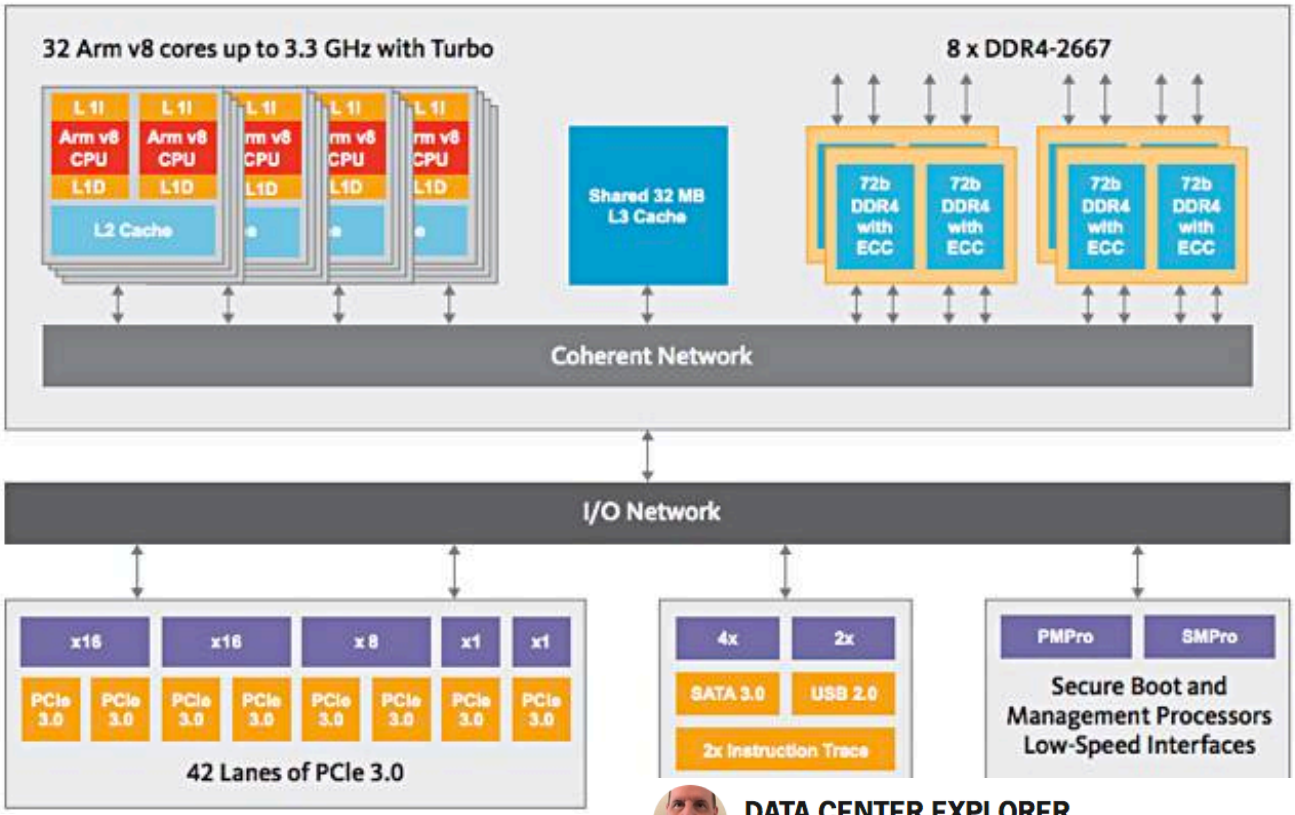
- Each pair of adjacent *abc* mesh/torus is interconnected with twelve links





Ampere eMAG chip

Ampere Computing ... led by former Intel president Renee James



eMAG 8180 Block Diagram

No support yet for 2-sockets

DATA CENTER EXPLORER

By Andy Patrizio, Network World | DEC 3, 2019 4:45 AM PST

About

Andy Patrizio is a freelance technology writer based in Orange County, California. He's written for a variety of publications, ranging from Tom's Guide to Wired to Dr. Dobbs Journal.

Ampere preps an 80-core Arm processor for the cloud

A new Ampere chip due out next year is single-threaded to avoid the 'noisy neighbor' problem that can impede customer workloads in multi-tenant cloud-provider networks.

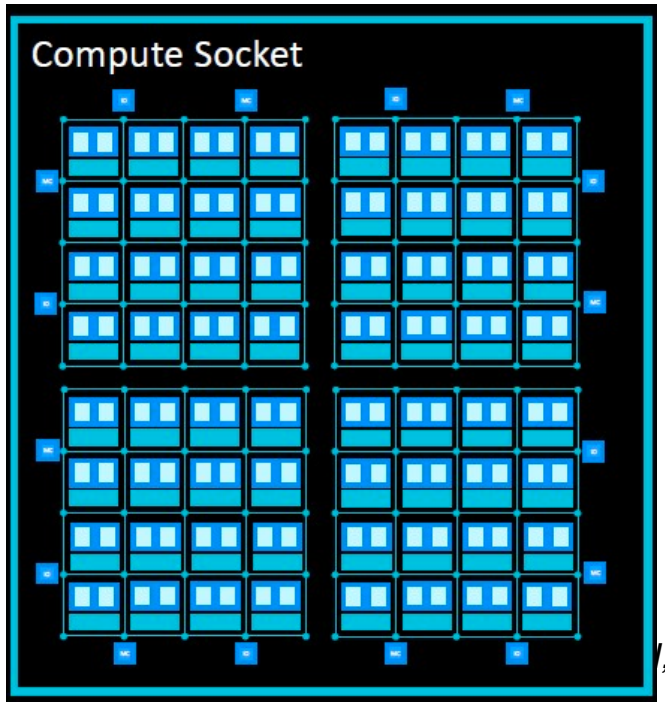
https://amperecomputing.com/wp-content/uploads/2019/01/eMAG8180_PB_v0.5_20180914.pdf

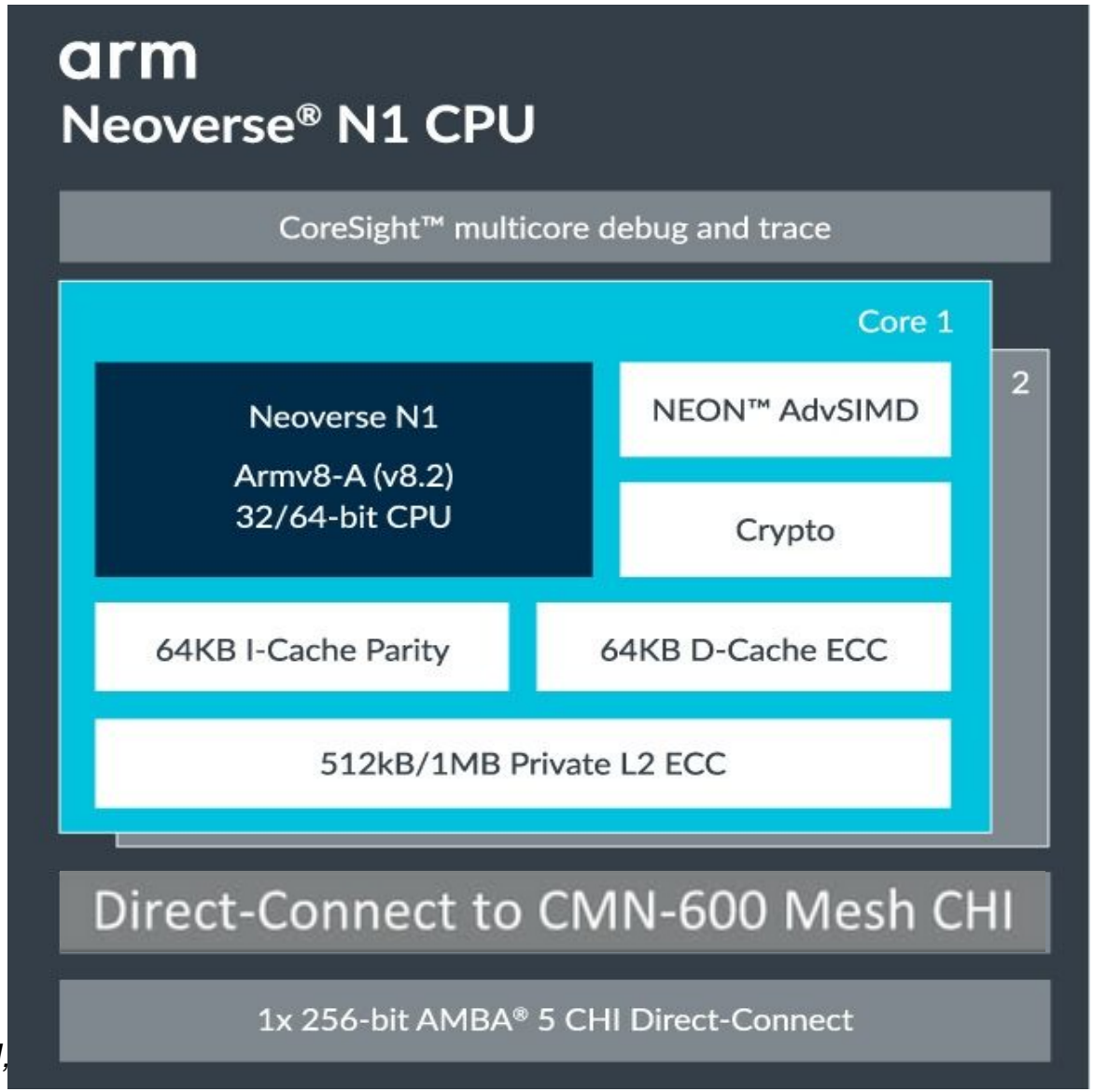
Arm Neoverse N1

Hyperscale Datacenter

150W and beyond

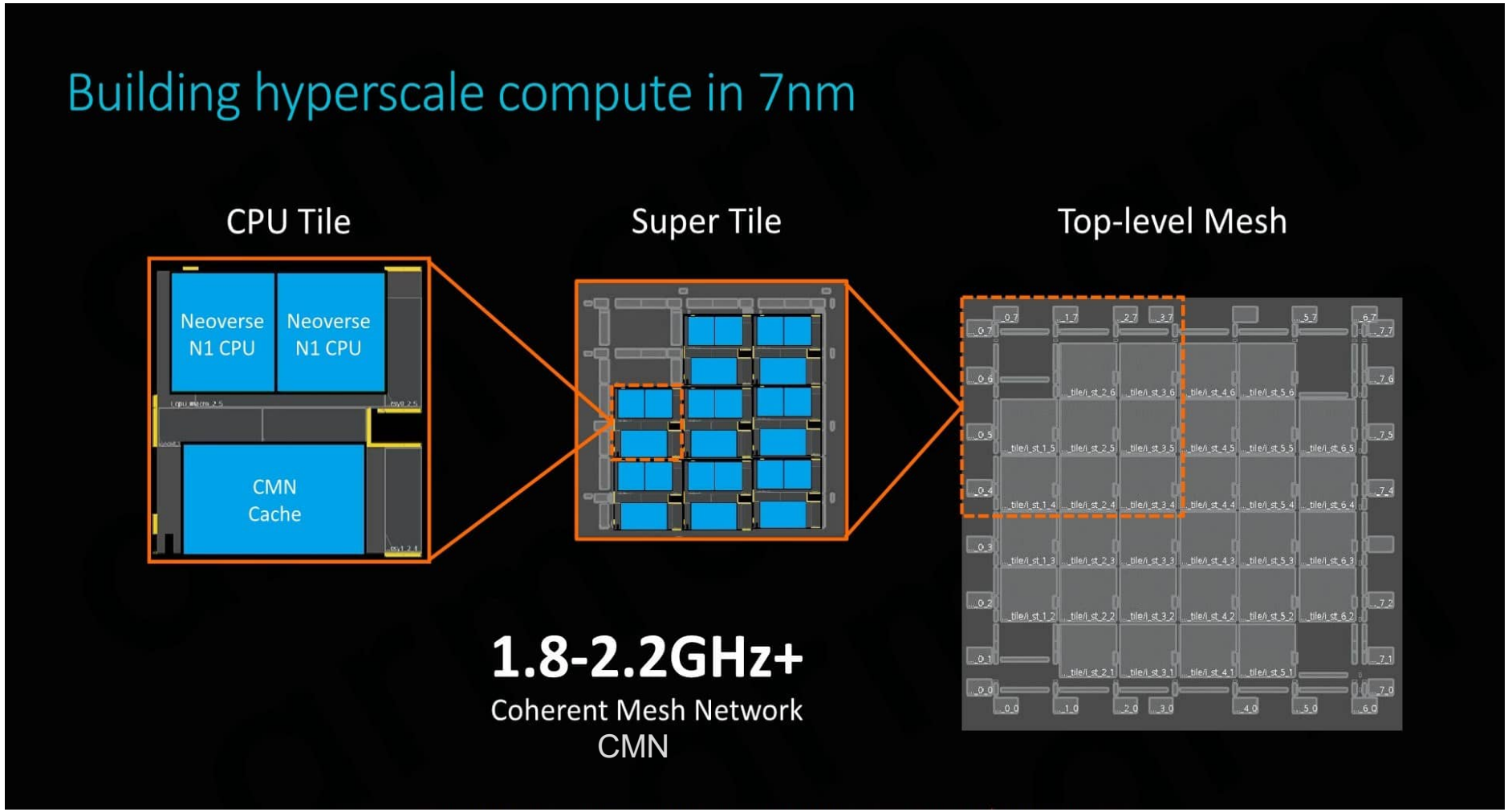
64-128 core





https://community.arm.com/developer/ip-products/processors/b/processors-ip-blog/posts/arm-neoverse-n1-platform-accelerating-the-transformation-to-a-scalable-cloud-to-edge-infrastructure

Arm Neoverse N1



Amazon Web Services (AWS):
the new Graviton2 with custom Arm Neoverse N1 cores

Neoverse N1 makes debut in new AWS cloud instances

December 03, 2019

By Chris Bergey, SVP and GM, Infrastructure Line of Business, Arm

The Huawei Kunpeng 920
(previously known as HiSilicon Hi1620)

TaiShan High-Performance ARM Processors

Hi1620 – The World's First 7nm Datacenter ARM Processor

- **Industry-leading performance with excellent memory and I/O capacity**
- **Fine-tuned for memory-bound workloads such as CAE/CFD, weather, and life-science**
- 128-bit Advanced SIMD unit

Processor	Hi1620
Core	ARM v8.2 architecture 2.4 -3.0 GHz, 24 -64 cores per socket
Cache	L1: 64 KB instruction cache and 64 KB data cache L2: 512 KB private per core L3: 24 -64 MB shared for all (1 MB/core)
Memory	8 DDR4 channels per socket, up to 3200 MHz
Coherent Interconnect	Coherent SMP interface for 2S & 4S 3 ports, up to 240 Gbit/s per port
I/O	40 PCIe Gen 4.0 lanes 2 x 100GE, RoCEv2/RoCEv1, CCIX x4 USB 3.0, x16 SAS 3.0, x2 SATA 3.0
Package	60 mm x 75 mm, BGA
Process	7 nm
Power	TDP: 100 -200 W

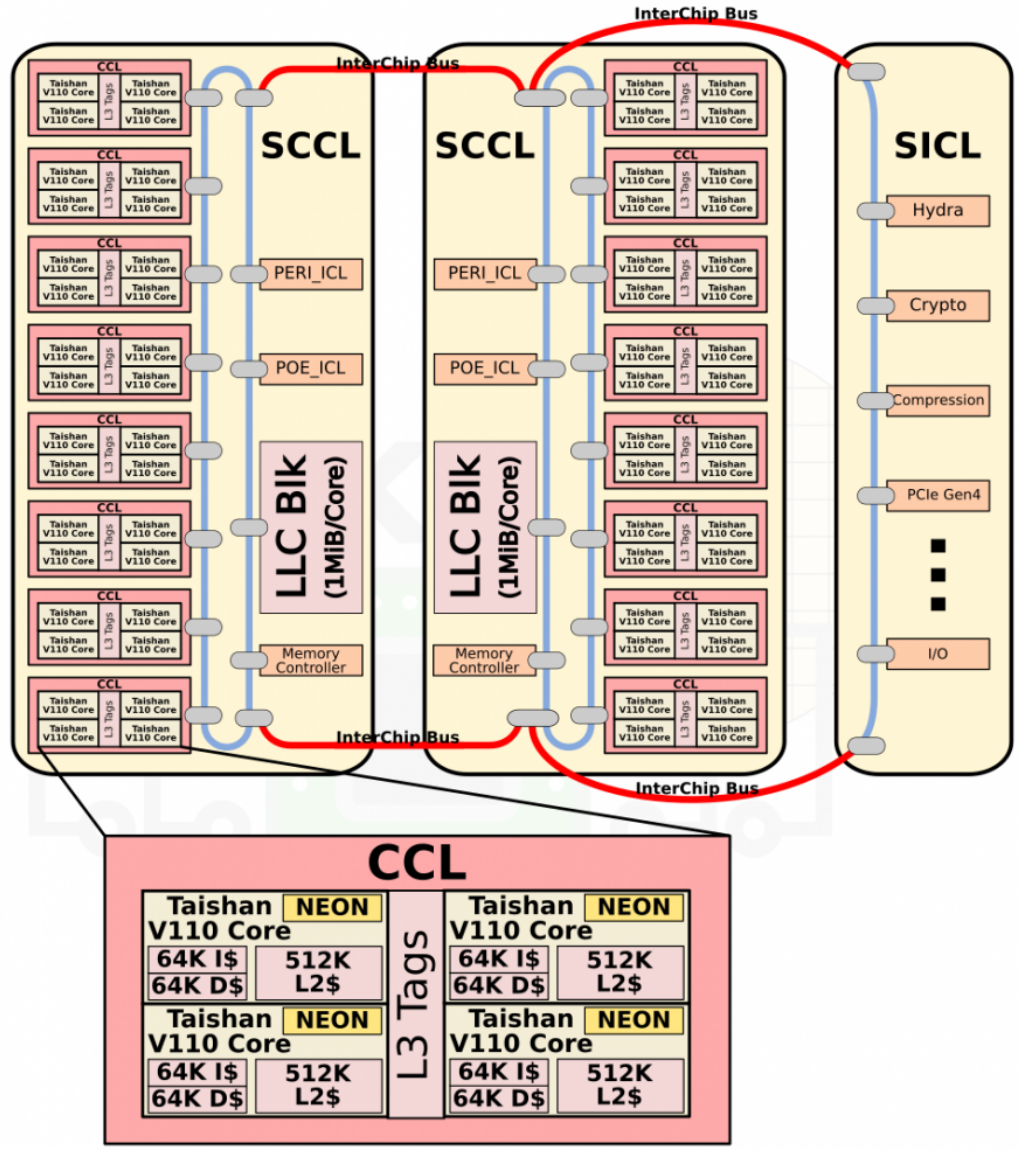
The Huawei Kunpeng 920 is based on the TaiShan V110 core, a semi-custom ARM Cortex-A72

The Huawei Kunpeng 920:
a multi-chip 48-64 cores

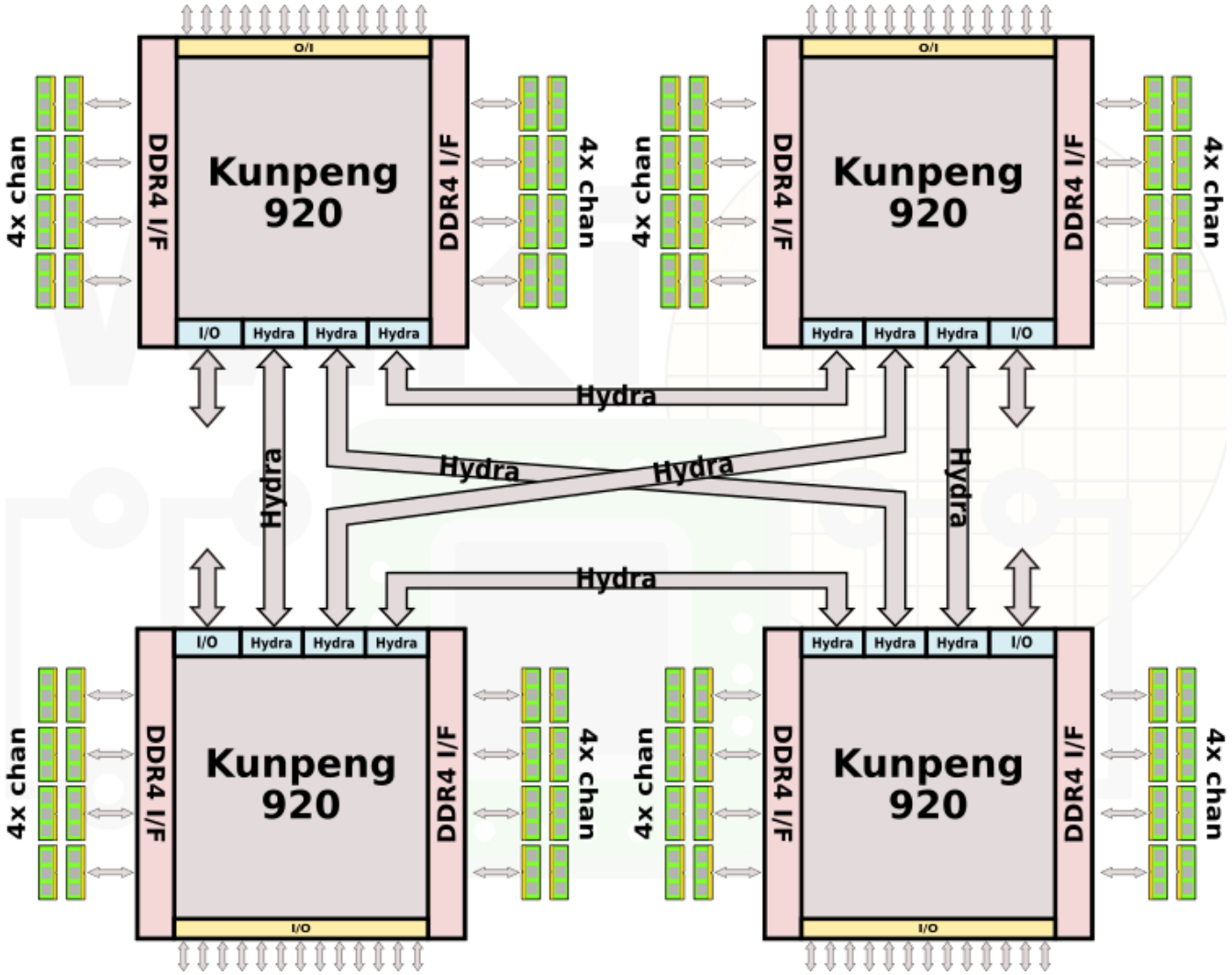
SCCL:
Super CPU cluster

SICL:
Super IO Cluster

CCL:
CPU Clusters



The Huawei Kunpeng 920:
multi-socket support



Outline

1. TOP500
   a) TOP10 lists from Nov'17 to Nov'19
   b) Country distribution over the past 25 years
   c) PU chip technology evolution in the past 25 years and since last year
   d) Evolution of the accelerators since they were available
   e) Analysis of some relevant systems and architectures
2. GREEN500
   a) TOP10 lists from Nov'17 to Nov'19
   b) Analysis of some relevant systems
3. HPCG500
   a) HPCG vs. HPL: an overview
   b) TOP10 lists from Nov'17 to Nov'19
   c) Analysis of some relevant systems
4. GRAPH500
   a) Performance Metric (TEPS)
   b) Breadth-First Search (BFS) & Single Source Shortest Paths (SSSP)
5. And next?...

- Description

  Data-intensive supercomputer applications are an increasingly important workload, but are ill-suited for platforms designed for 3D physics simulations. Application performance cannot be improved without a meaningful benchmark. Graphs are a core part of most analytics workloads. Backed by a steering committee of over 30 international HPC experts from academia, industry, and national laboratories, this specification establishes a large-scale benchmark for these applications. It will offer a forum for the community and provide a rallying point for data-intensive supercomputing problems. This is the first serious approach to augment the Top 500 with data-intensive applications.

- Performance metric: traversed edges per second (TEPS)

- List of benchmark problems
  - Breadth-First Search (BFS)
  - Single-Source Shortest Paths (SSSP)

The GRAPH500 BFS list
(November'19)

Top Ten from November 2019 BFS

	RANK	MACHINE	VENDOR	INSTALLATION SITE	LOCATION	COUNTRY	YEAR	NUMBER OF NODES	NUMBER OF CORES	SCALE	GTEPS
3	1	Sunway TaihuLight	NRCPC	National Supercomputing Center in Wuxi	Wuxi	China	2015	40768	10599680	40	23755.7
12	2	DOE/NNSA/LLNL Sequoia	IBM	Lawrence Livermore National Laboratory	Livermore CA	USA	2012	98304	1572864	41	23751
22	3	DOE/SC/Argonne National Laboratory Mira	IBM	Argonne National Laboratory	Chicago IL	USA	2012	49152	[illegible]	40	14982
1	4	OLCF Summit (CPU-Only)	IBM	Oak Ridge National Laboratory	Oak Ridge TN	United States	2018	2048	86016	40	7665.7
9	5	SuperMUC-NG	Lenovo	Leibniz Rechenzentrum	Garching	Germany	2018	4096	196608	39	6279.47
	6	Fermi	IBM	CINECA	Casalecchio Di Reno	Italy	2012	8192	131072	37	2567
13	7	NERSC Cori - 1024 Haswell partition	Cray	NERSC/LBNL	DOE/SC /LBNL/NERSC	United States	2017	1024	32768	37	2562.16
	8	Tianhe-2 (MilkyWay-2)	National University of Defense Technology	Changsha China	Changsha China	China	2013	8192	196608	36	2061.48
14	9	Nurion	Cray	Korea Institute of Science and Technology Information	Daejeon	Korea Republic Of	2018	1024	65536	37	1456.46
	10	Turing	IBM	CNRS/IDRIS-GENCI	Orsay	France	2012	4096	65536	36	1427

AJProença,

The GRAPH500 SSSP list
(November'19)

Top Ten from November 2019 SSSP

Show 10 entries

Search:

RANK	MACHINE	VENDOR	INSTALLATION SITE	LOCATION	COUNTRY	YEAR	NUMBER OF NODES	NUMBER OF CORES	SCALE	GTEPS
1	SuperMUC-NG	Lenovo	Leibniz Rechenzentrum	Garching	Germany	2018	4096	196608	37	1053.93
2	NERSC Cori - 1024 haswell partition	Cray	NERSC/LBNL	DOE/SC /LBNL/NERSC	United States	2017	1024	32768	36	558.833
3	Nurion	Cray	Korea Institute of Science and Technology Information	Daejeon	Korea Republic Of	2018	1024	65536	36	337.239
4	NERSC Cori - 512 KNL partition	Cray	NERSC/LBNL	DOE/SC /LBNL/NERSC	United States	2017	512	32768	35	229.188
5	Undisclosed Cray XE6	Cray	National Computing Facility	University	United States	2013	512	16384	34	134.173
6	Undisclosed Cray XE6	Cray	National Computing Facility	University	United States	2013	512	8192	31	12.88
7	Xeon Server	Dell	industry	BoiseID	United States	2017	1	40	23	3.09
8	RAIDEN CPU subsystem	Fujitsu	RIKEN AIP	Tokyo	Japan	2016	32	1024	33	2.66
9	Alkindi-CPU	Dell	The University of British Columbia	Vancouver	Canada	2015	1	28	22	1.23
10	Xeon Server	Dell	Industry	BoiseID	United States	2017	1	32	21	0.985