



Master Informatics Eng.

2018/19

A.J.Proença

HPC systems: on top of the wave

(most slides are borrowed)



1. TOP500

- a) TOP10 lists from Nov'18 and Nov'17
- 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: IBM Summit & BlueGene/Q, Sunway TaihuLight, Tianhe-2A
- f) Intel presence in TOP500: current shift into the Scalable Family

2. GREEN500

- a) TOP10 lists from Nov'17 and Nov'18
- b) Analysis of some relevant systems: NVidia SaturnV with Volta, the Chinese system with Hygon Dhyana (IP from AMD Epyc 7501)

3. HPCG500

- a) HPCG vs. HPL: an overview
- b) TOP10 lists from Nov'17 and Nov'18
- c) Analysis of some relevant systems: Fujitsu K and Post-K, HPE Astra
- d) Next HPC systems based on ARMv8: Cray XC50, the MontBlanc project

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

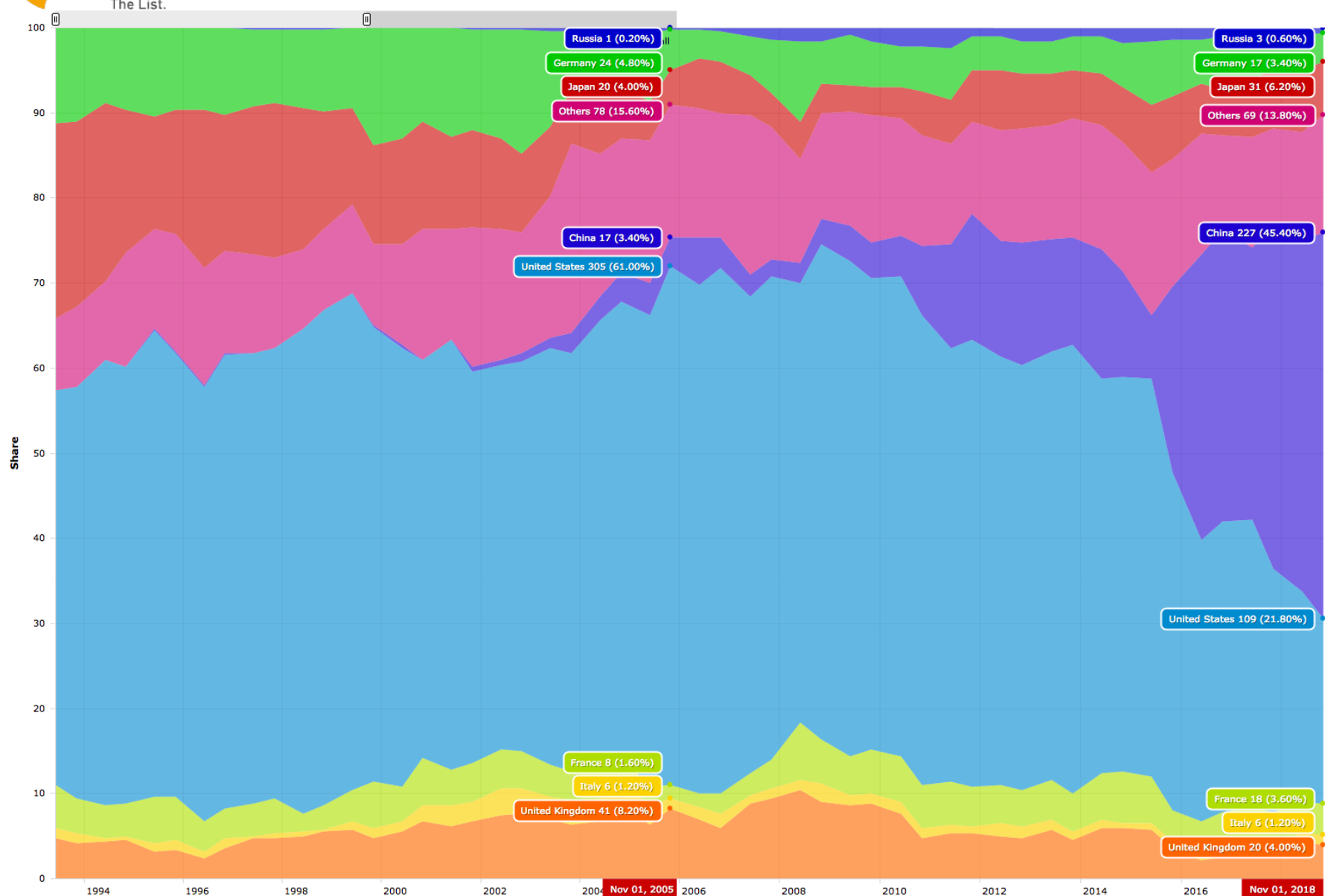
Top 10 HPC systems Nov'18 TOP500

Top 10 HPC systems

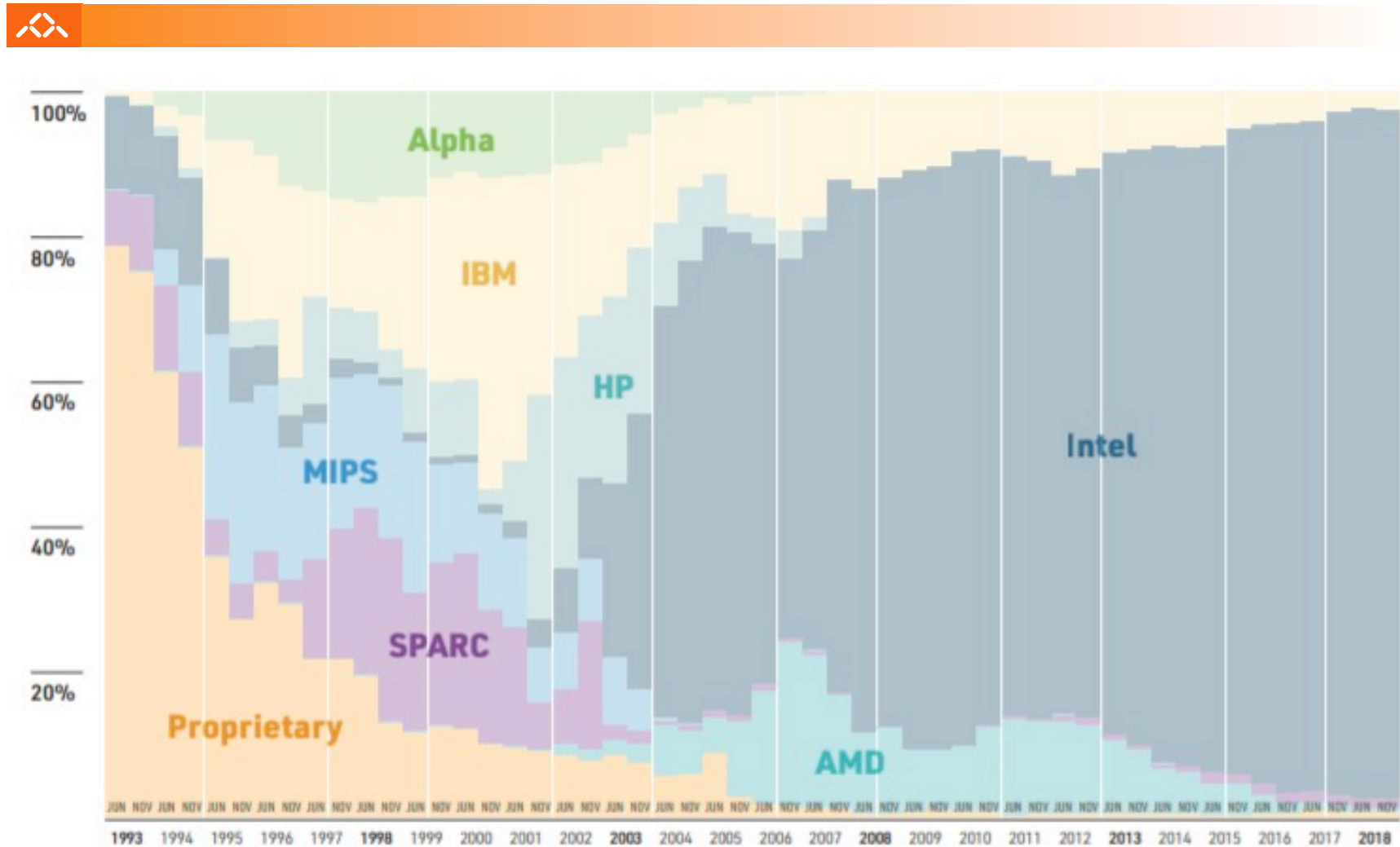
Nov'17 TOP500

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	
		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
3	Piz Daint - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	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
4	Gyokou - 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	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
5	Titan - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x , Cray Inc. DOE/SC/Oak Ridge National Laboratory United States	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

Country distribution over the past 25 years



Chip technology in the past 25 years

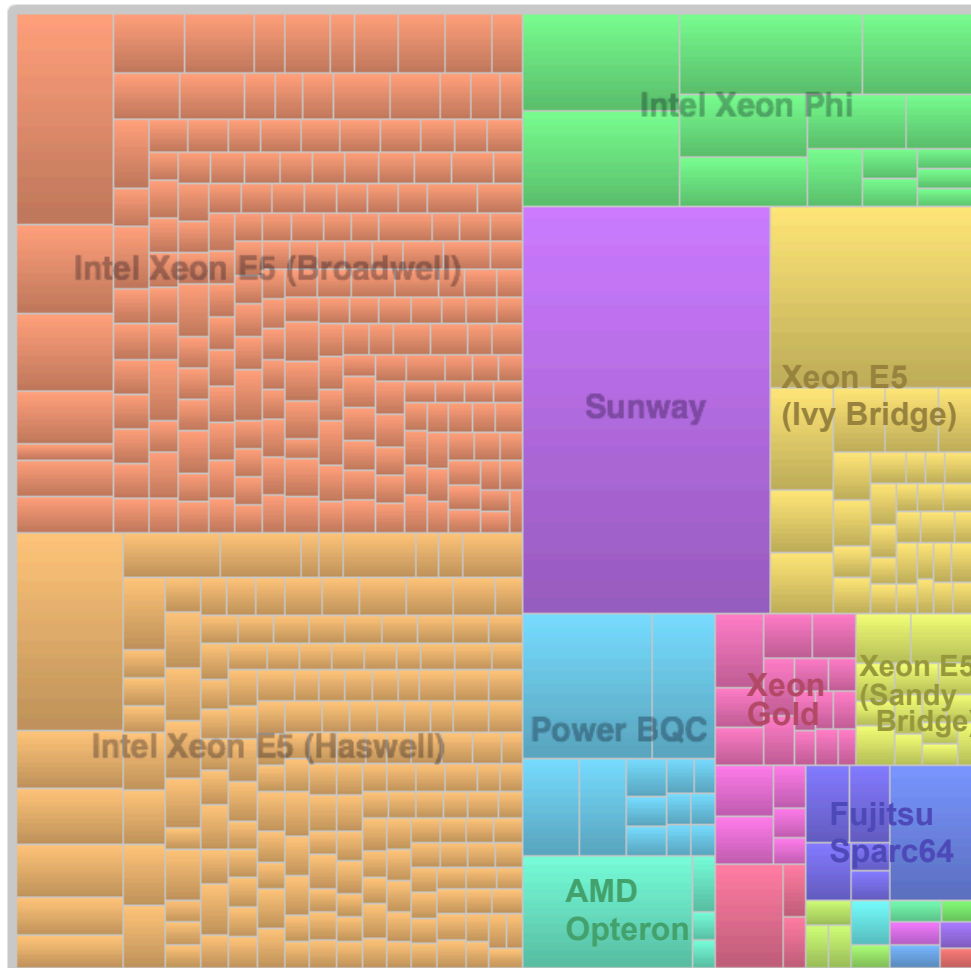


<https://www.nextplatform.com/2018/11/12/the-widening-gyre-of-supercomputing/>

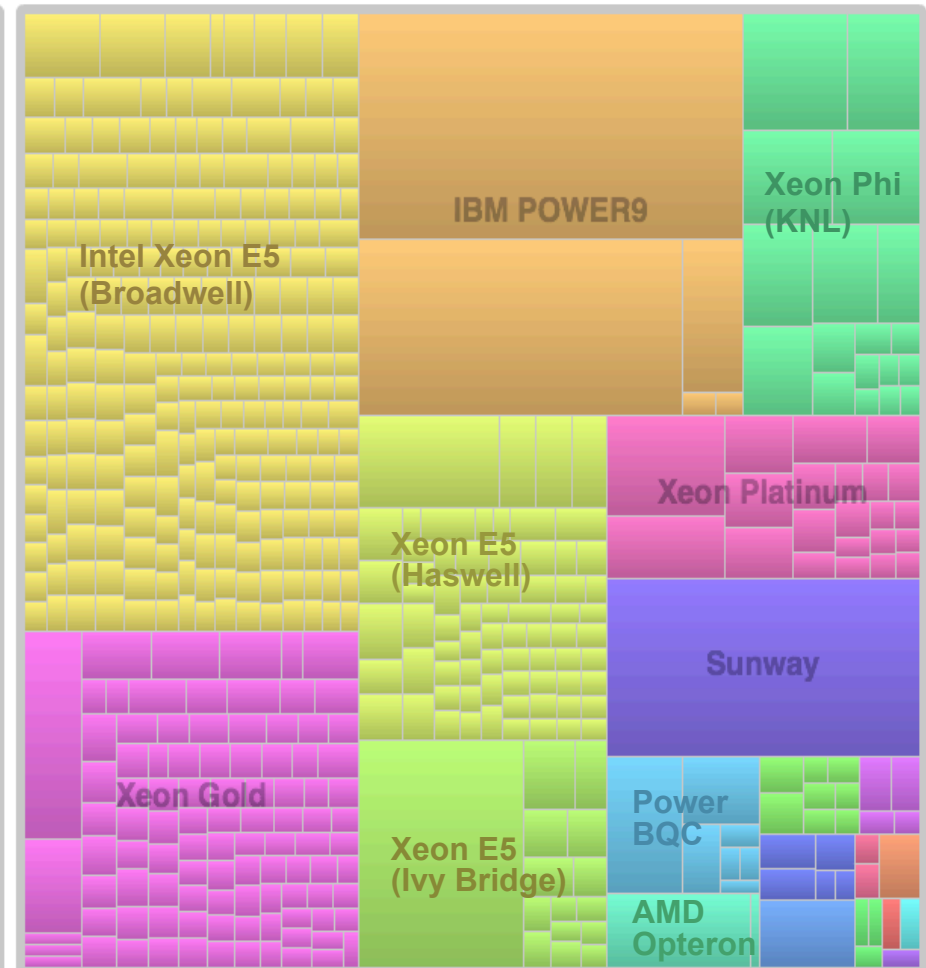
Processor generations Nov'17 & Nov'18



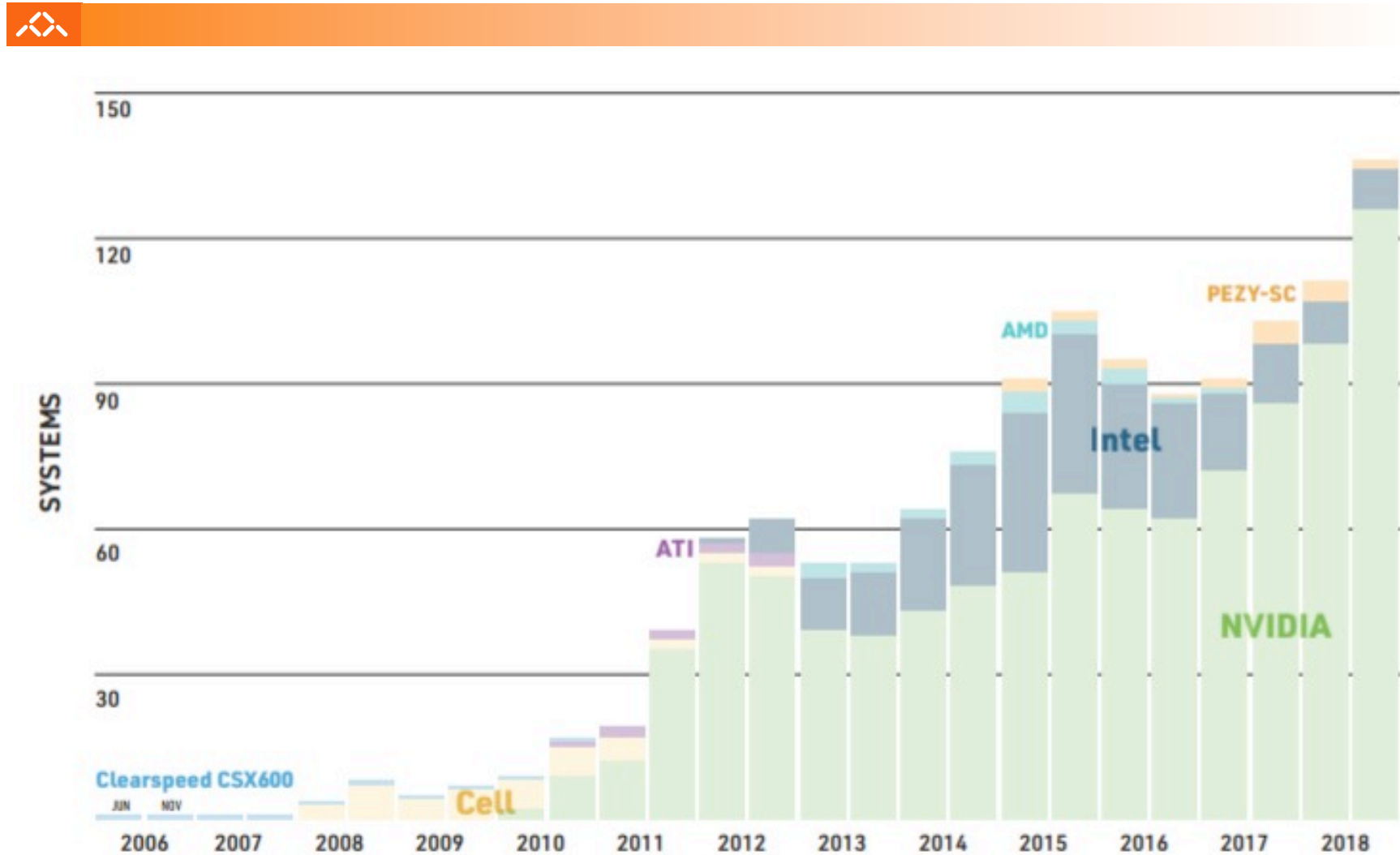
Nov'17



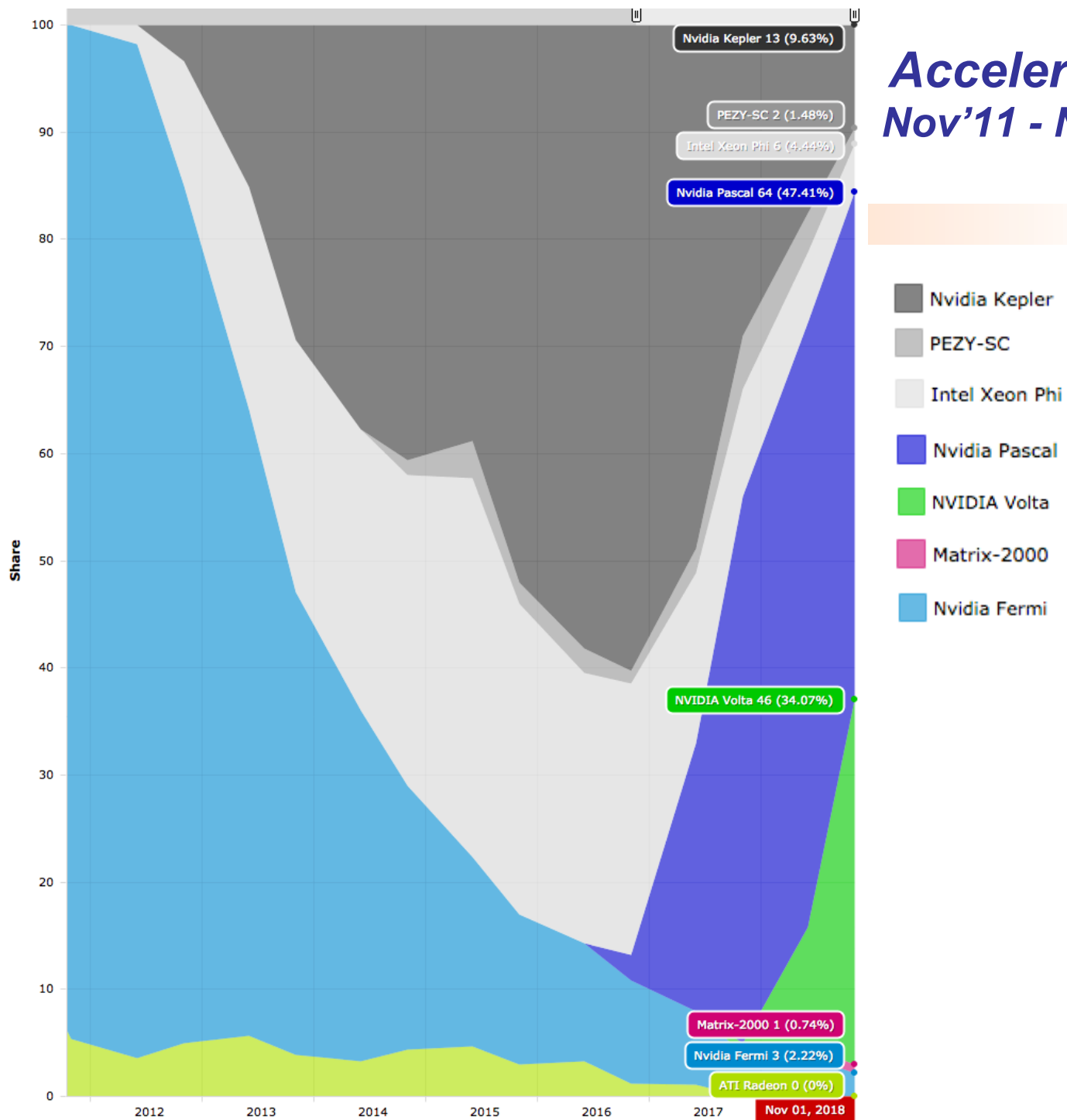
Nov'18



Accelerator families in the past 25 years



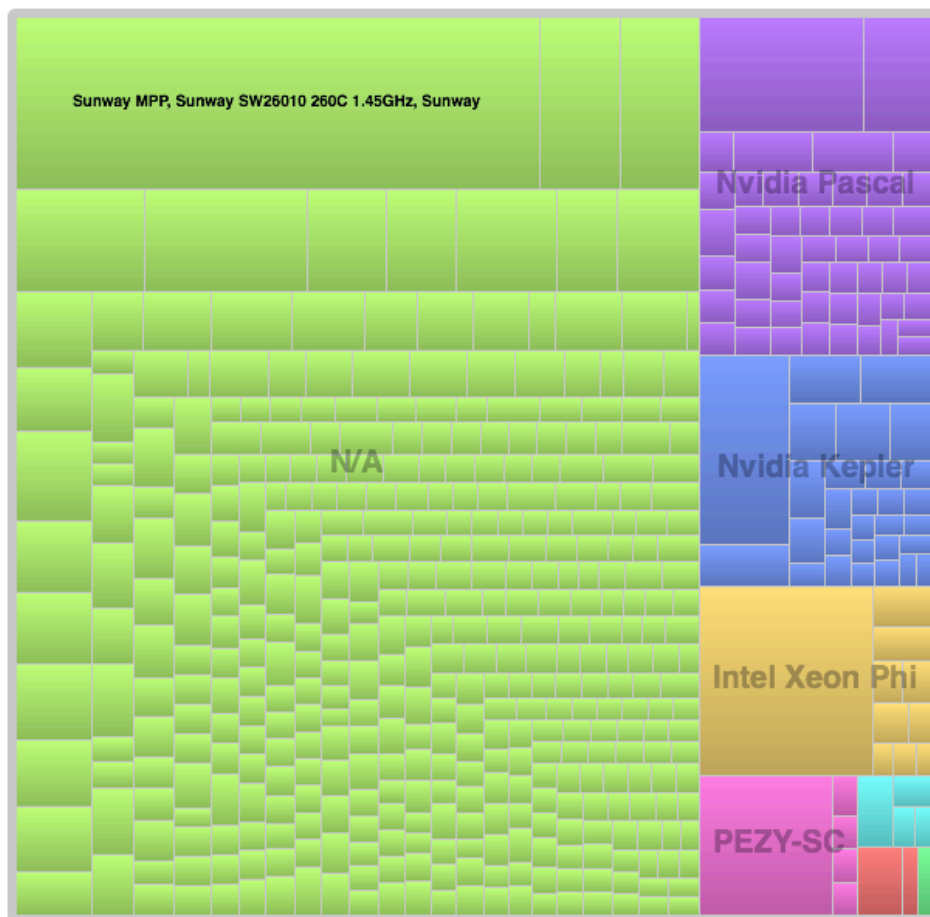
<https://www.nextplatform.com/2018/11/12/the-widening-gyre-of-supercomputing/>



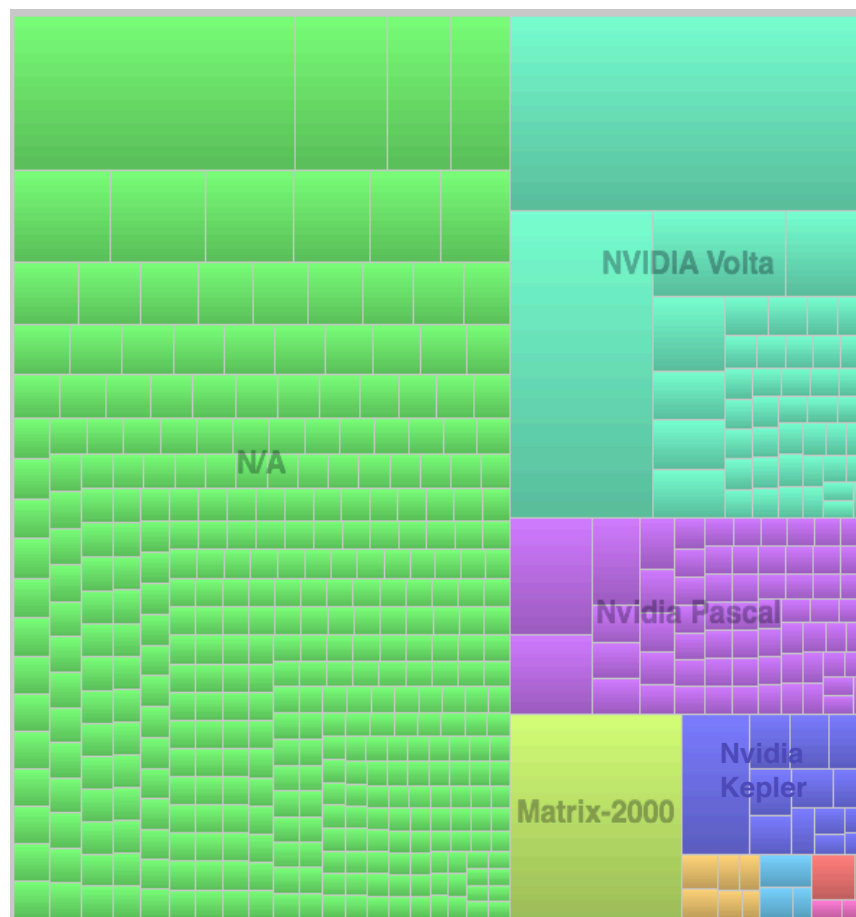
Accelerator family distribution Nov'17 & Nov'18



Nov'17



Nov'18





IBM POWER9 Summit (Nov'18 #1 TOP500)

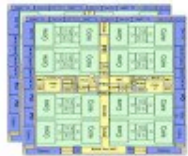
Summit Overview



Components

IBM POWER9

- 22 Cores
- 4 Threads/core
- NVLink



NVIDIA GV100

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



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

Compute Rack

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



Compute System

10.2 PB Total Memory

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





22-core IBM POWER9



POWER9 Processor – Common Features



New Core Microarchitecture

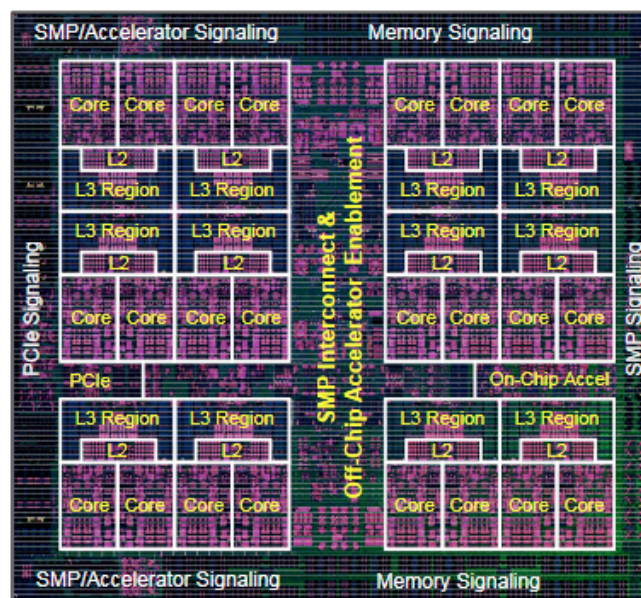
- Stronger thread performance
- Efficient agile pipeline
- POWER ISA v3.0

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

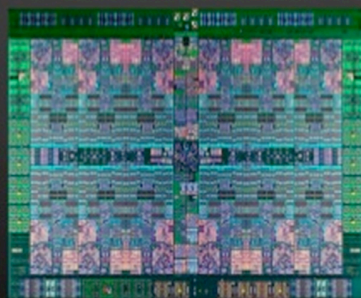
- 16 Gb/s interface
 - Local SMP
- 25 Gb/s Common Link interface
 - Accelerator, remote SMP



IBM POWER9 + NVidia V100



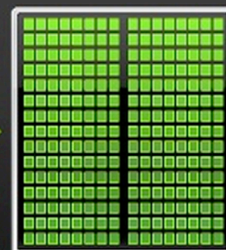
Accelerated Computing 5x Higher Energy Efficiency



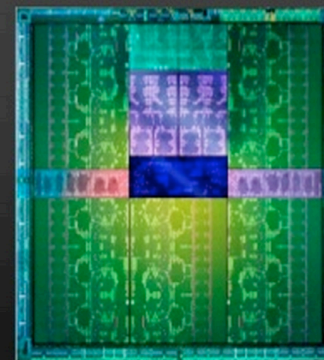
IBM POWER CPU
Most Powerful Serial Processor



80-200
GB/s



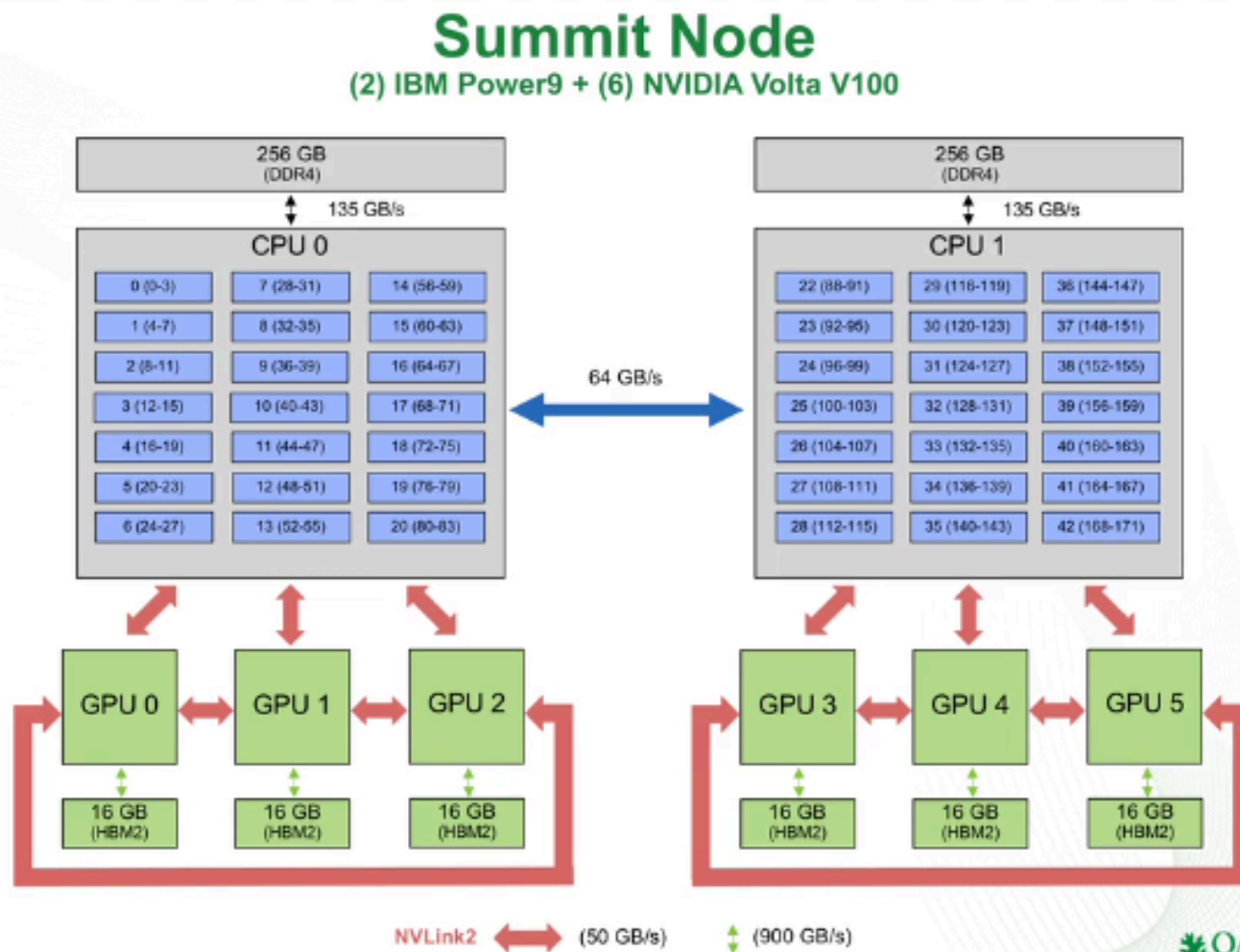
NVIDIA NVLink
Fastest CPU-GPU Interconnect



NVIDIA Volta GPU
Most Powerful Parallel Processor



Summit node architecture



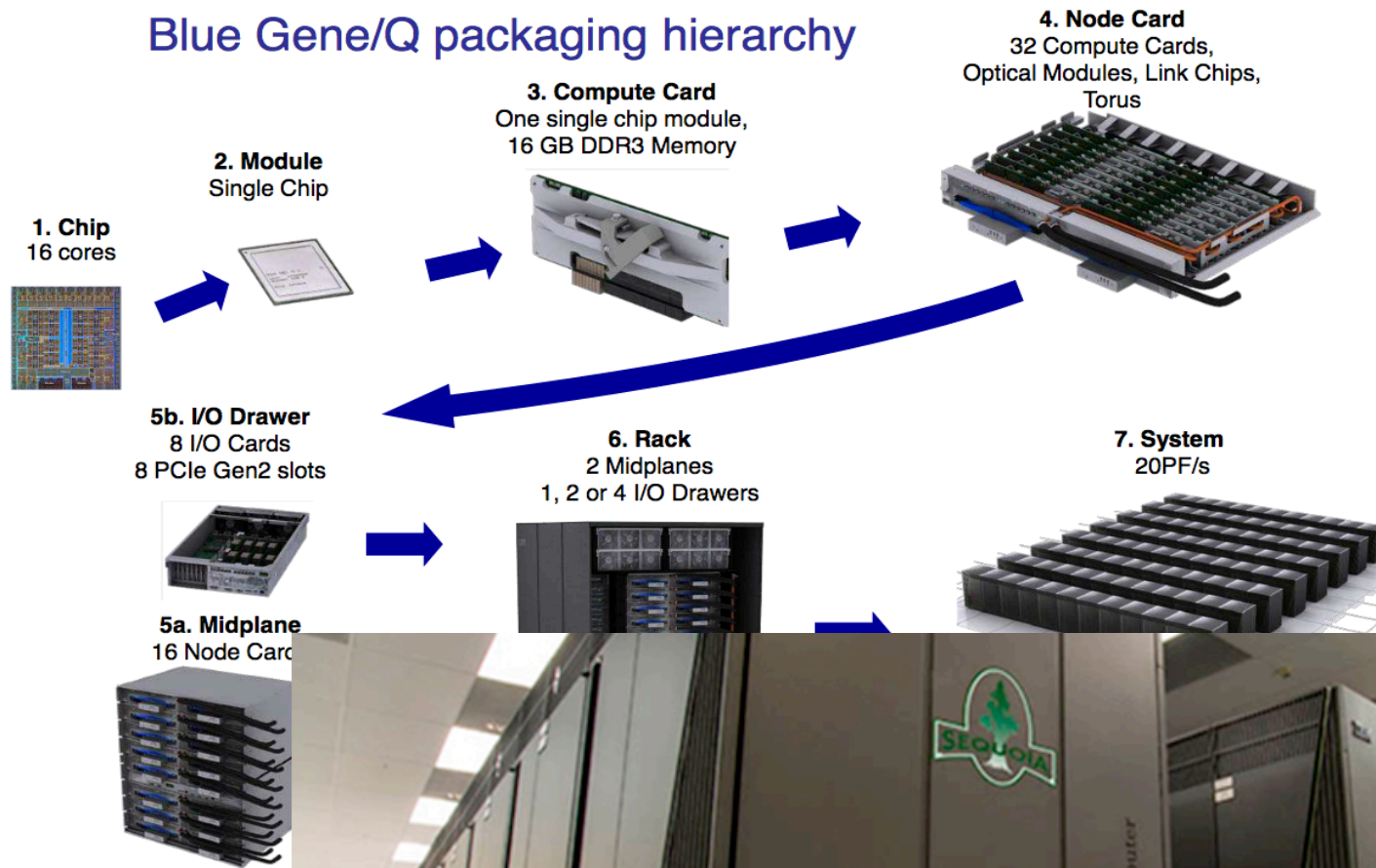
<https://en.wikipedia.org/wiki/supercomputers/summit>



IBM Power BlueGene/Q Compute (Sequoia)



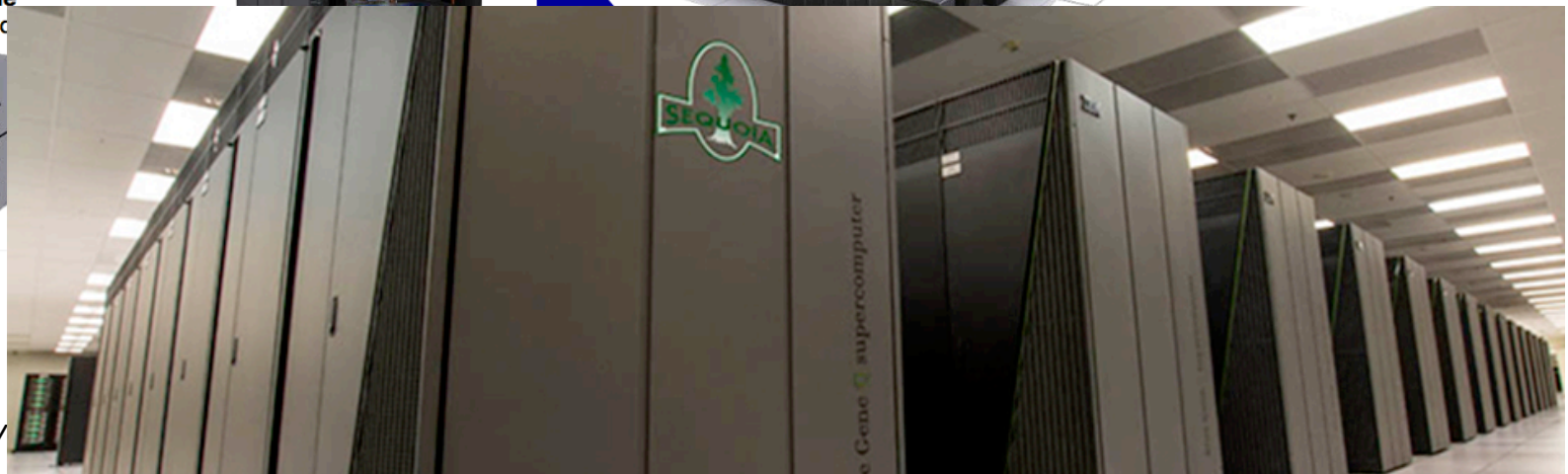
Blue Gene/Q packaging hierarchy



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

Ref: SC2010

AJProença, Adv

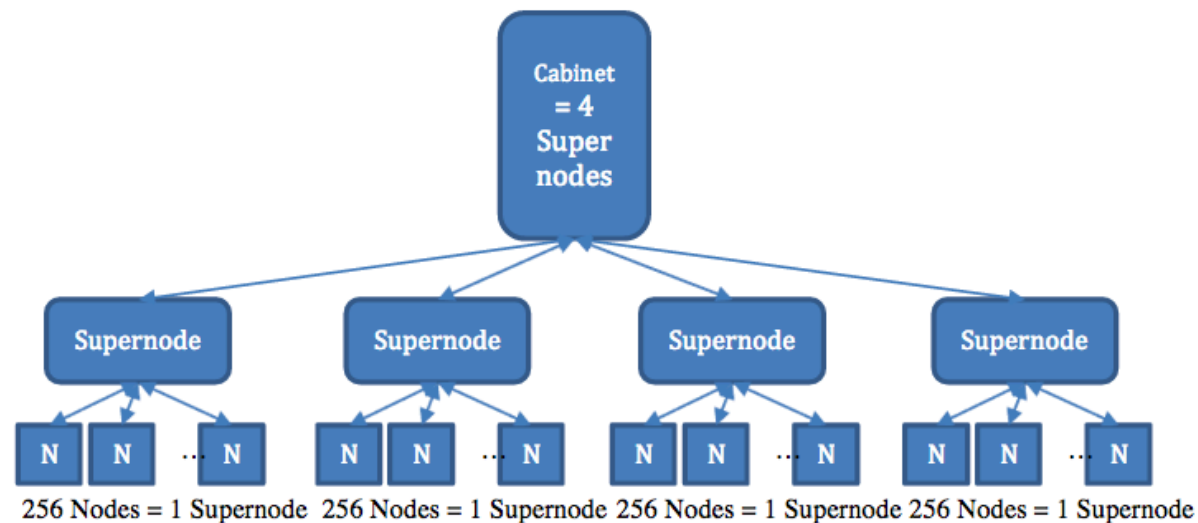
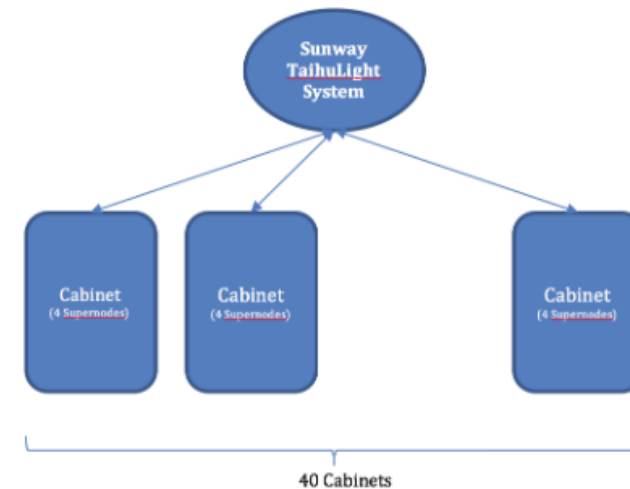
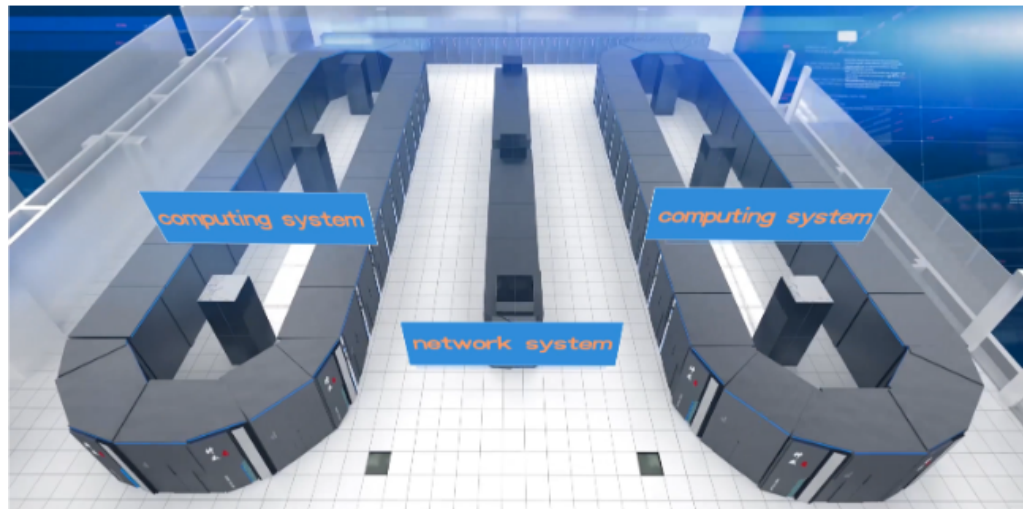




Sunway TaihuLight (#1 in June'16 TOP500)



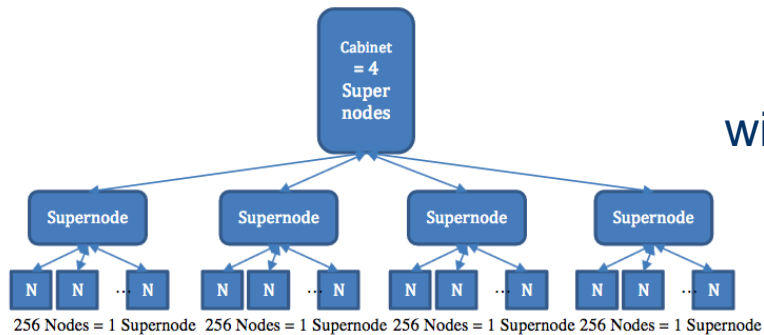
Overview of the Sunway TaihuLight System



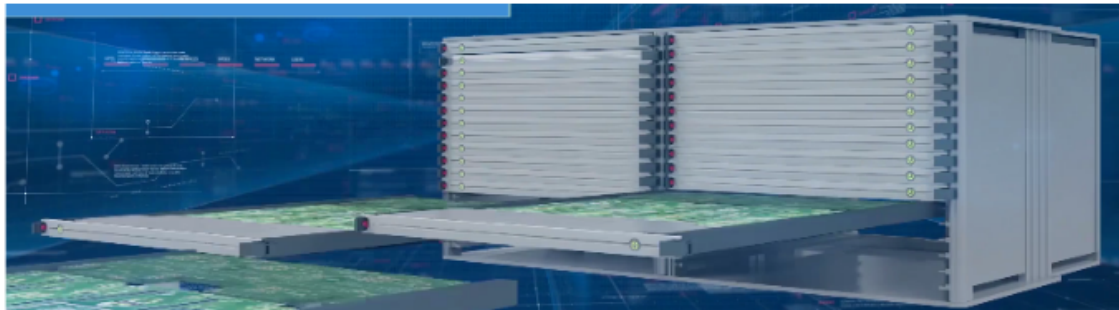
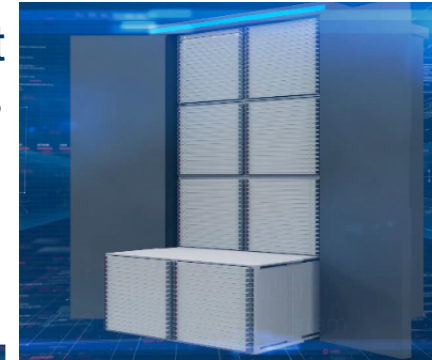
AJProença, Advanced Architectures,



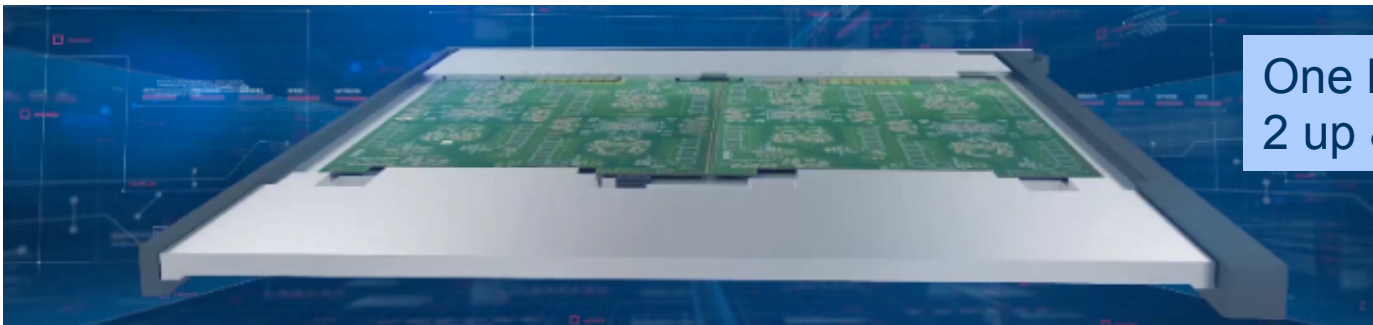
Sunway TaihuLight (#1 in June'16 TOP500)



One cabinet
with 4 Supernodes



One Supernode
with 32 boards

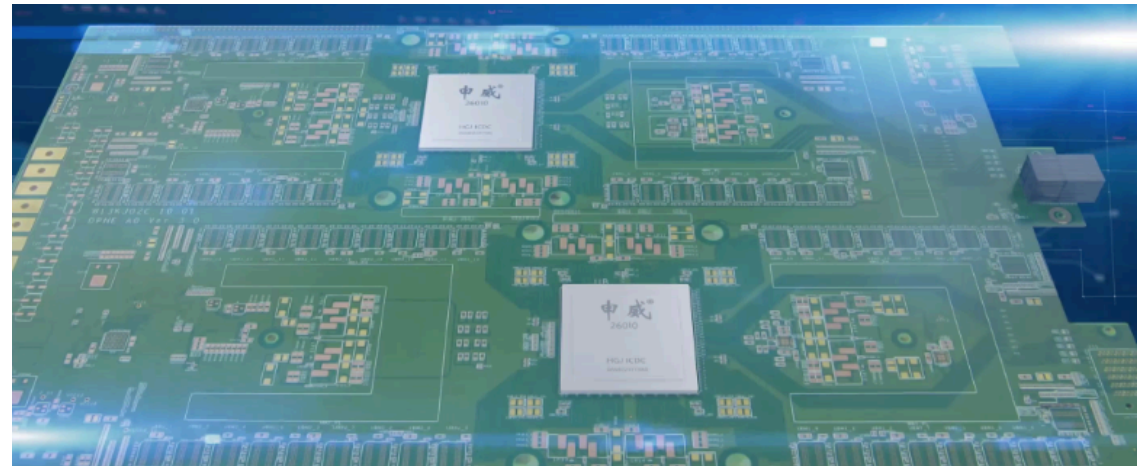


One board with 4 cards,
2 up & 2 down

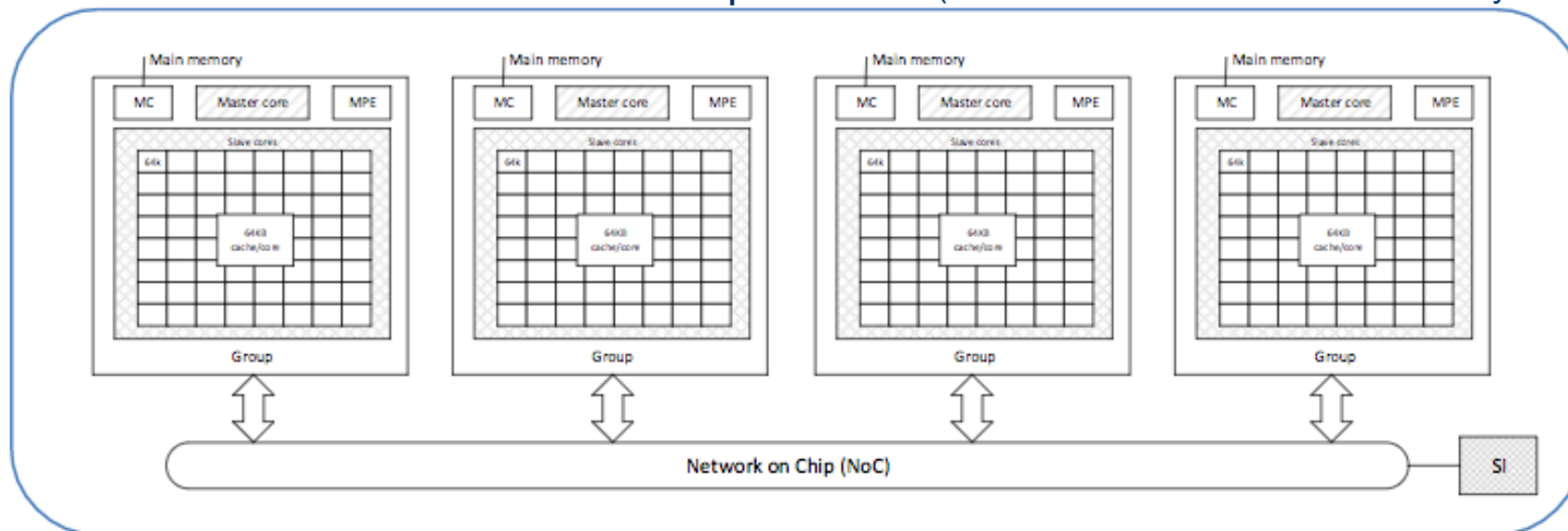


Sunway TaihuLight (#1 in June'16 TOP500)

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



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



Overview of Tianhe-2A



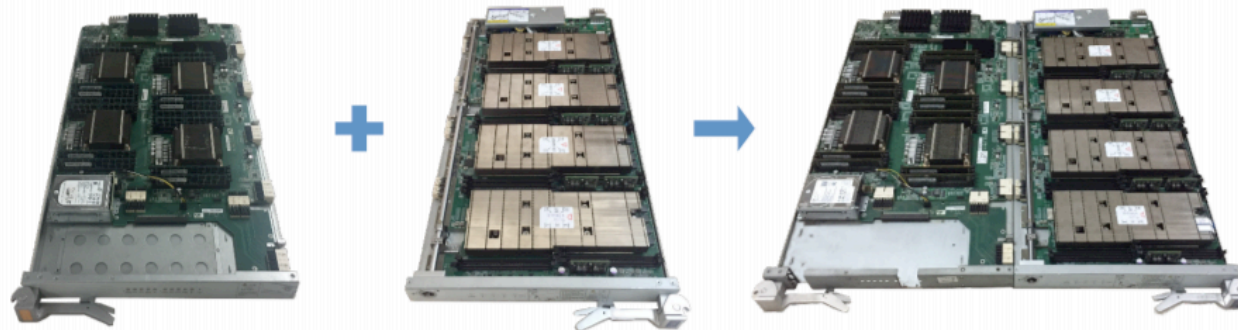
Compute nodes



● 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



Overview of Tianhe-2A



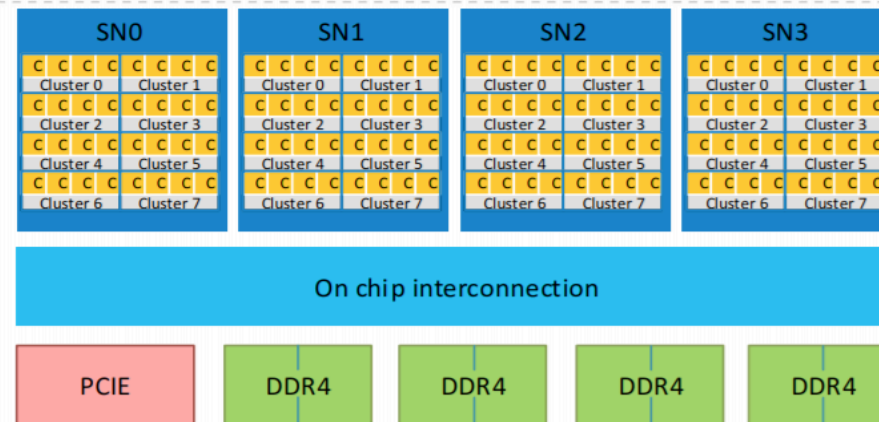
Matrix-2000 accelerator



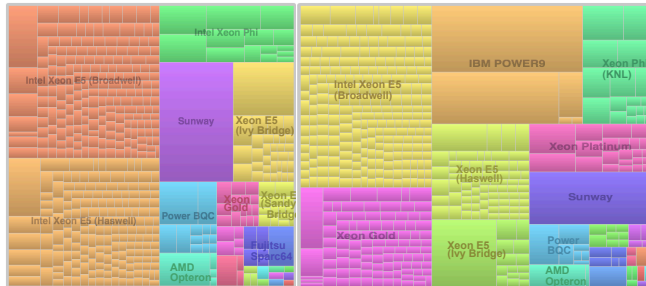
Chip specification

- 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**
- Peak power dissipation: ~240w
- Interface

- 8 DDR4-2400 channels
- X16 PCIE 3.0 EP Port



$$4 \text{ SNs} \times 8 \text{ clusters} \times 4 \text{ cores} \times 16 \text{ flops} \times 1.2 \text{ GHz} = 2.4576 \text{ Tflops}$$



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

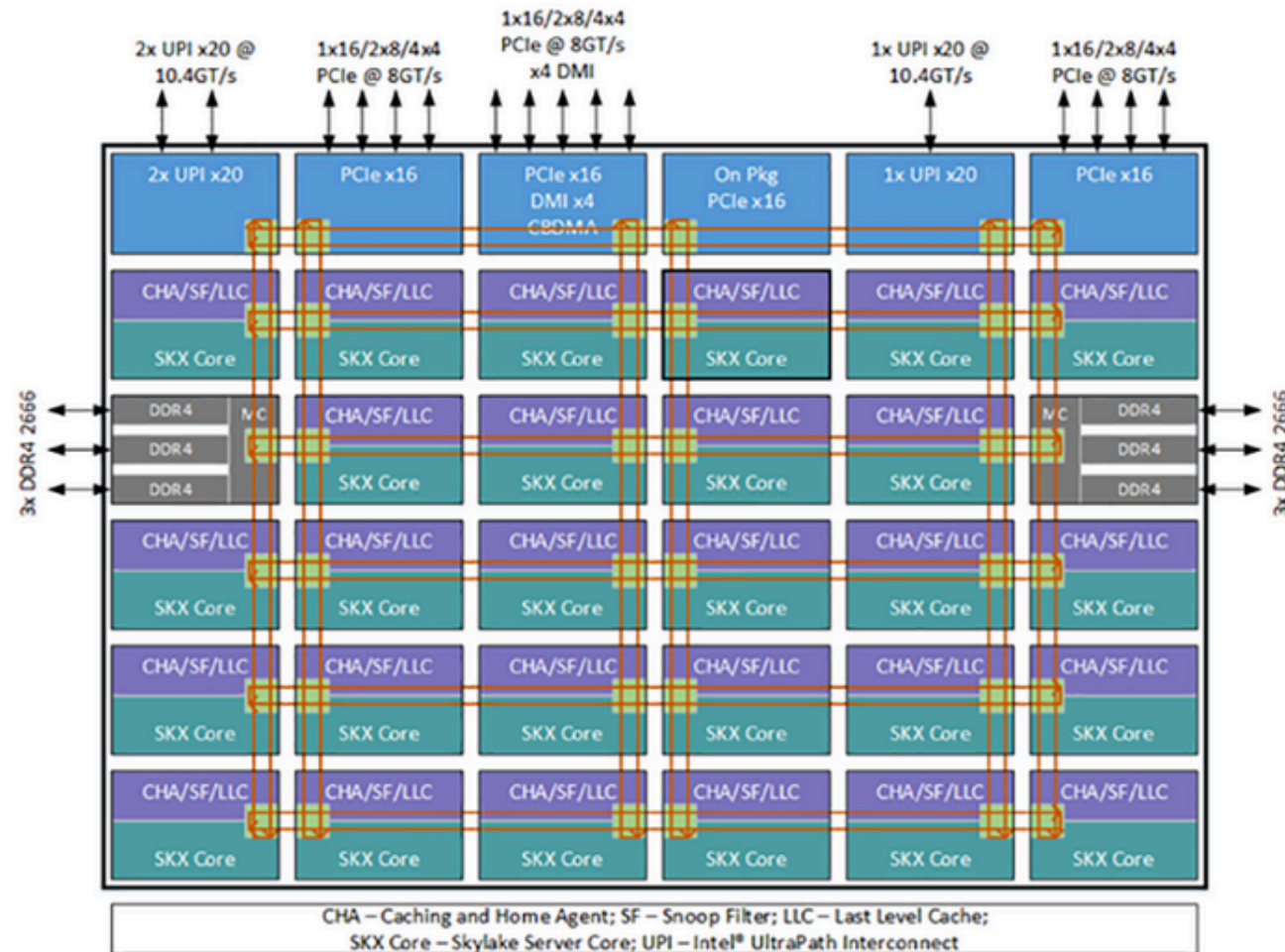
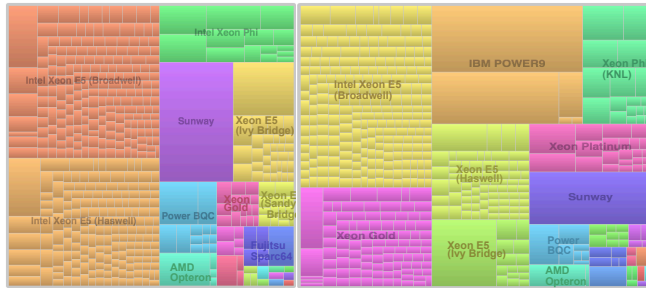


Figure 5. Intel Xeon processor Scalable family mesh architecture.



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

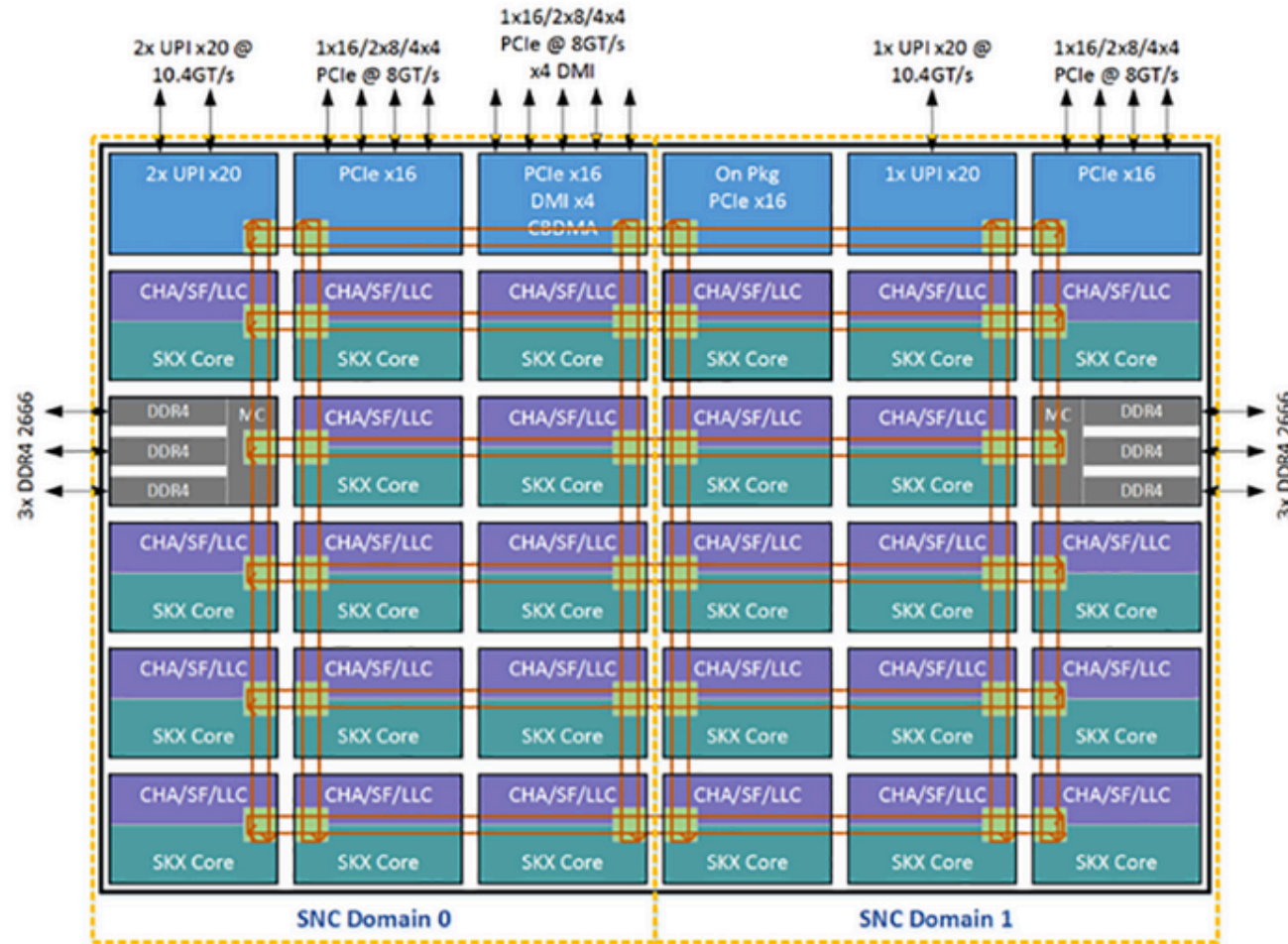
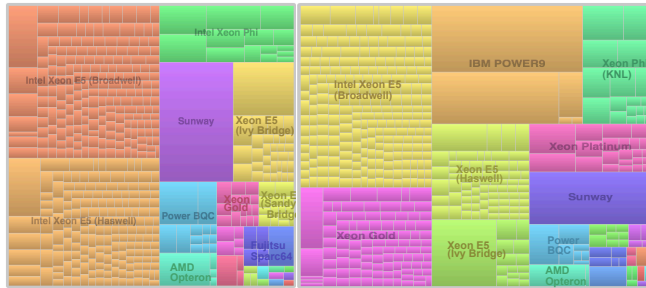


Figure 10. Sub-NUMA cluster domains.



**Next PU generations for HPC:
Intel® Xeon® Processor Scalable Family**

Cache Hierarchy Changes

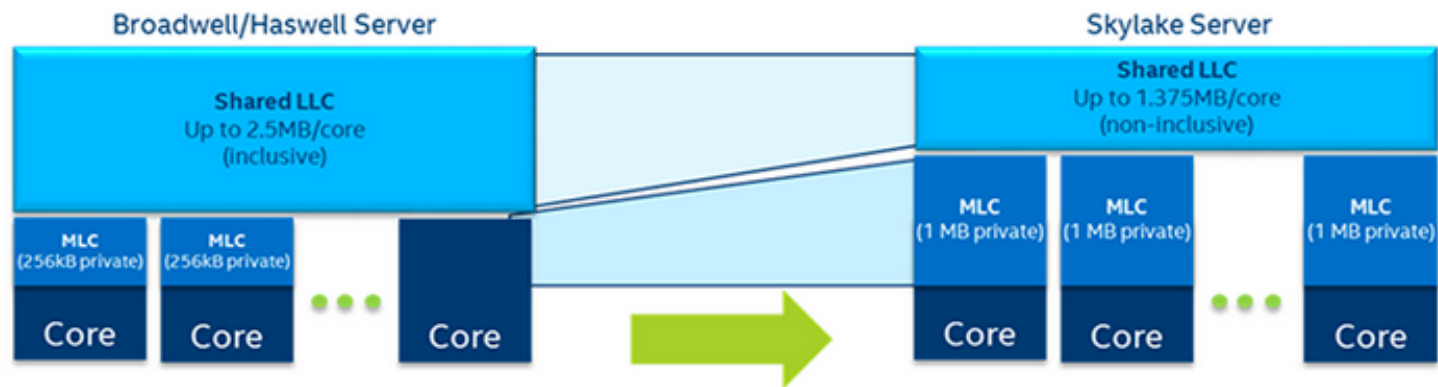


Figure 11. Generational cache comparison.



Top 10 greener-HPC systems Nov'17 Green500

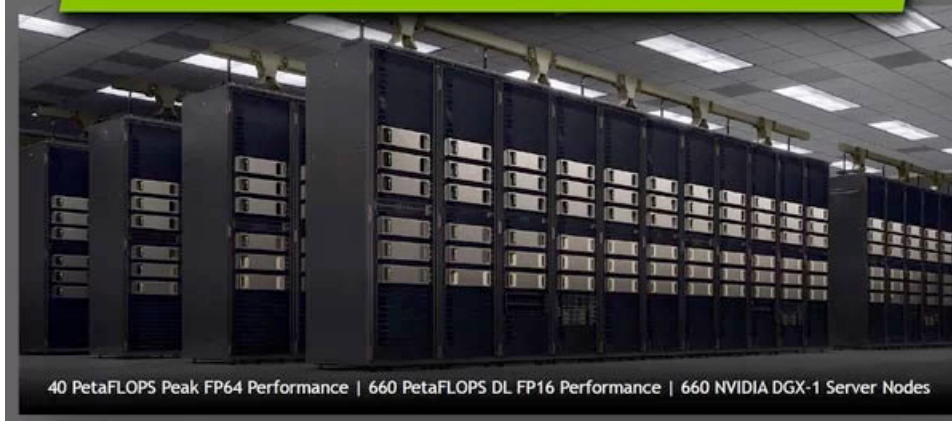
TOP500							
Rank	Rank	System	Cores	Rmax [TFlop/s]	Power [kW]	Power Efficiency [GFlops/watts]	
1	259	Shoubu system B - ZettaScaler-2.2, Xeon D-1571 16C 1.3GHz, Infiniband EDR, PEZY-SC2 , PEZY Computing / Exascaler Inc. Advanced Center for Computing and Communication, RIKEN Japan	794,400	842.0	50	17.009	
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	Gyokkou - 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	



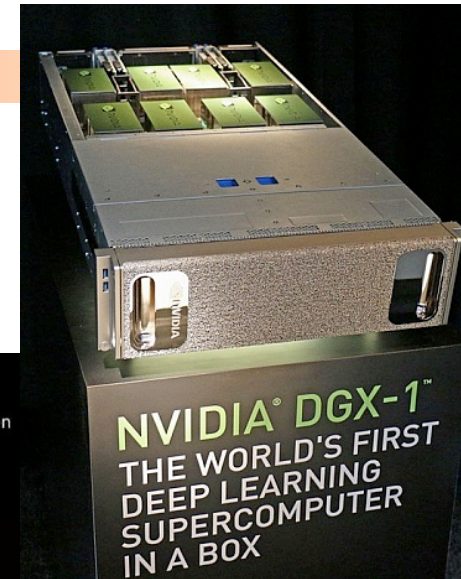
Top systems Nov'18 Green500

TOP500							
Rank	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	

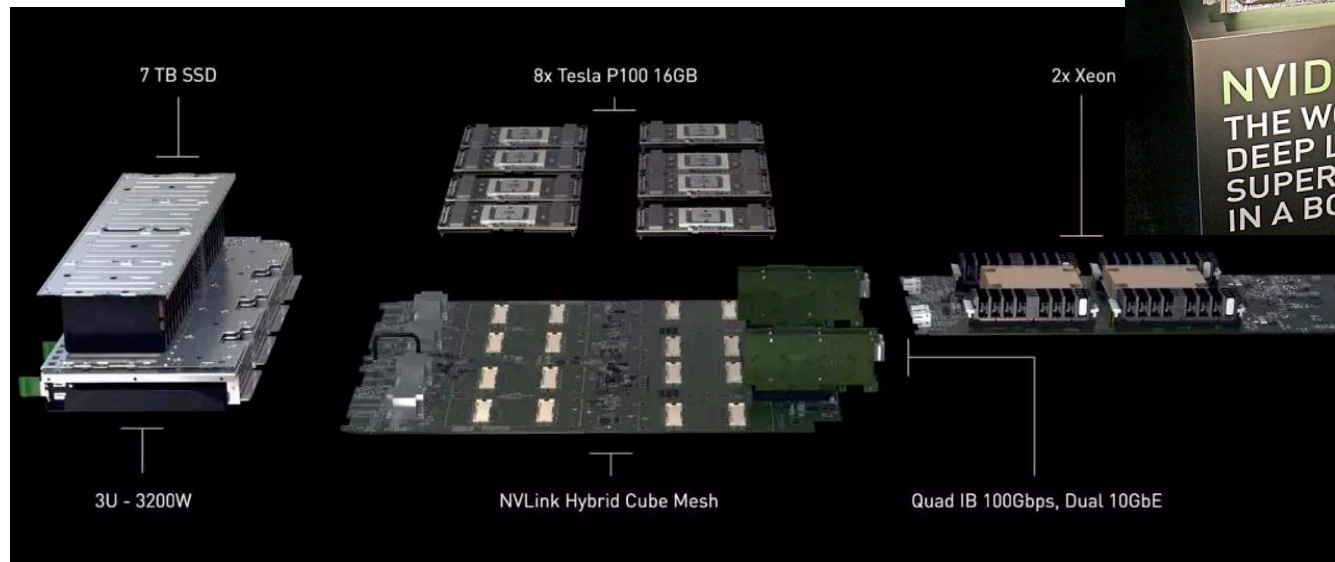
ANNOUNCING NVIDIA SATURNV WITH VOLTA



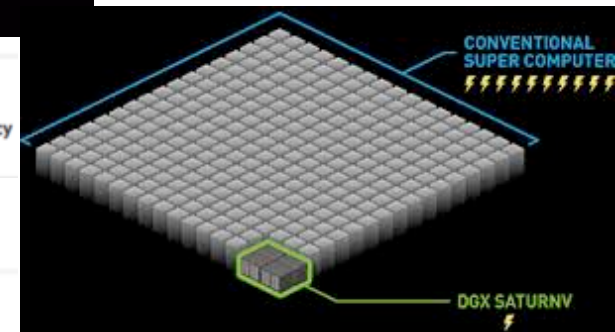
NVidia DGX-1 SaturnV



\$149,000



2	374	DGX SaturnV Volta	NVIDIA DGX-1 Volta36, Xeon	22,440	1,070.0	97	15.113
		V100, Nvidia	E5-2698 v4 2.2GHz, Infiniband EDR, NVIDIA Tesla				
			NVIDIA Corporation				
			United States				



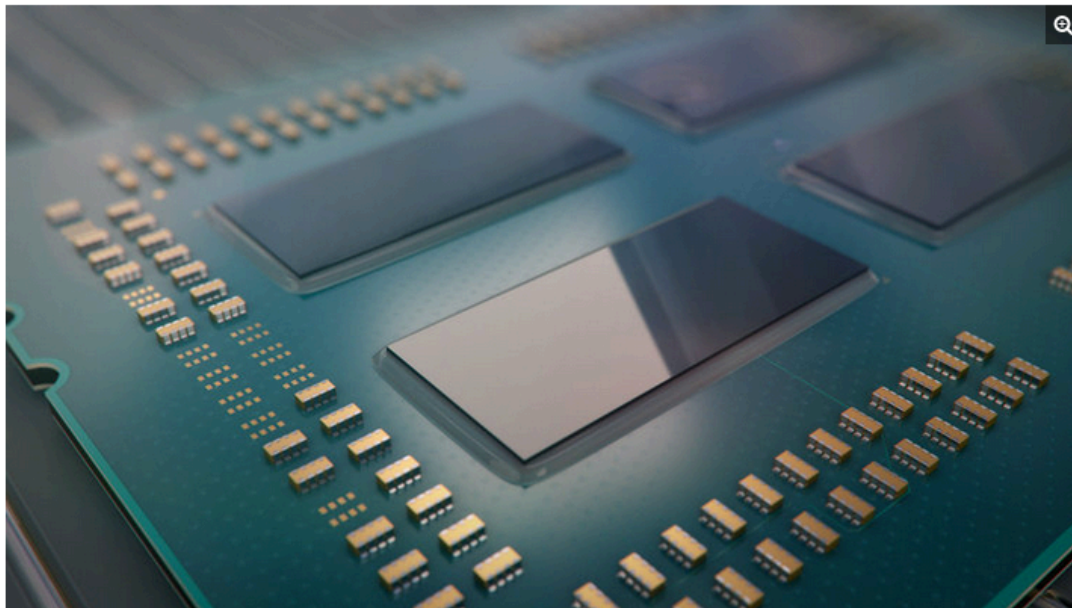
Advanced Computing System (PreE) (the Hygon Dhyana x86 device)



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

62
COMMENTS

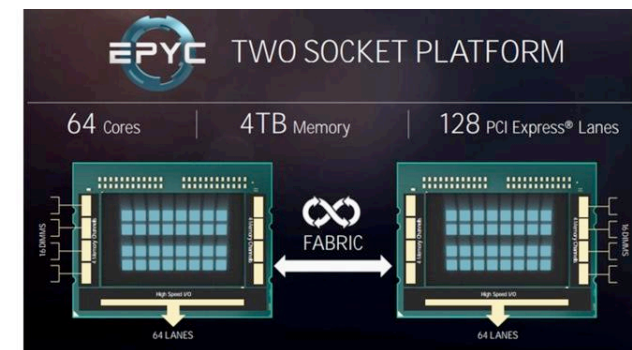
by Paul Alcorn July 6, 2018 at 1:18 PM



海光
HYGON

Hygon Dhyana:
32-core 2.0GHz CPU

Similar to:
AMD EPYC 7501



Chinese-designed "Dhyana" x86 processors based on AMD's Zen microarchitecture are beginning to surface from Chinese chip producer Hygon. The processors come as the fruit of AMD's x86 IP

TOP500: HPCG vs. HPL

HPCG UPDATE: ISC'17

Jack Dongarra
Michael Heroux
Piotr Luszczek



hpcg-benchmark.org

3

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

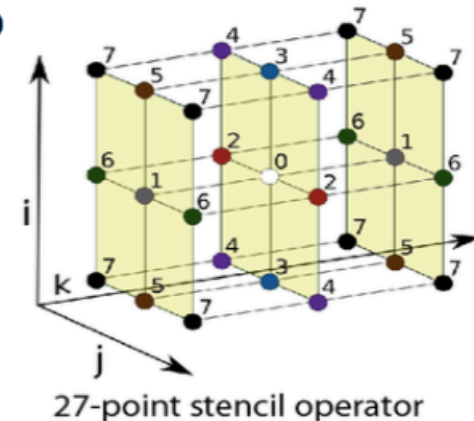
Jack Dongarra
Michael Heroux
Piotr Luszczek



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:
 $(np_x \times np_y \times np_z)$
- Process layout:
 $(n_x * np_x) \times (n_y * np_y) \times (n_z * np_z)$
- Global domain:
- Sparse matrix:
 - 27 nonzeros/row interior.
 - 8 – 18 on boundary.
 - Symmetric positive definite.



TOP500: HPCG vs. HPL

HPCG UPDATE: ISC'17

Jack Dongarra
Michael Heroux
Piotr Luszczek



hpcg-benchmark.org

5

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:
 - $n_x = n_y = n_z = 104$ gives
 - $n_{\text{local}} = 1,124,864; 140,608; 17,576; 2,197$
 - 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



TOP500				Rmax	Rpeak	HPCG							
Rank	Rank	System	Cores	(TFlop/s)	(TFlop/s)	(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	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
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	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
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	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
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	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
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	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

AJProença, Advanced Architectures, MiEI, UMinho, 2018/19



HPCG List for November 2018

TOP500						
Rank	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	

**Top 10 systems
Nov'17**



Fujitsu K computer

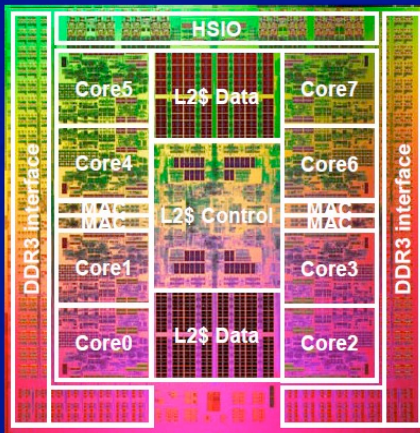
(the Japanese word "kei" (京) means 10 quadrillion, 10^{16})



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

		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

SPARC64™ VIIIfx Chip Overview



- **Architecture Features**
 - 8 cores
 - Shared 5 MB L2\$
 - Embedded Memory Controller
 - 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°C)
 - Water Cooling – Low leakage power and High reliability

SPARC64™ VIIIfx

12

All Rights Reserved, Copyright© FUJITSU LIMITED 2009





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

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



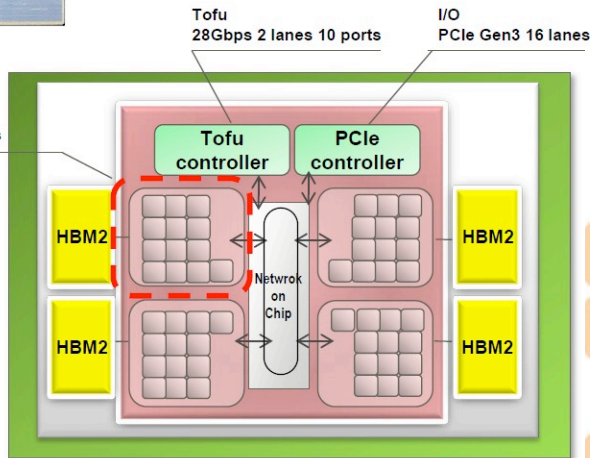
- Fujitsu CPU cores support the ARM SVE instruction set architecture
- Post-K Fujitsu CPU cores & Tofu maintain the programming models and provide high application performance
- ARM's standard frameworks (SBSA, etc.) assure compatibility among platforms
- **FP16 ("giant vector throughput") for supercomputers**

	Functions & architecture	 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	Enhanced	Tofu2	Tofu	Tofu



Fujitsu's A64FX ARM Chip

CMG specification
13 cores
L2\$ 8MiB
Mem 8GiB, 256GB/s

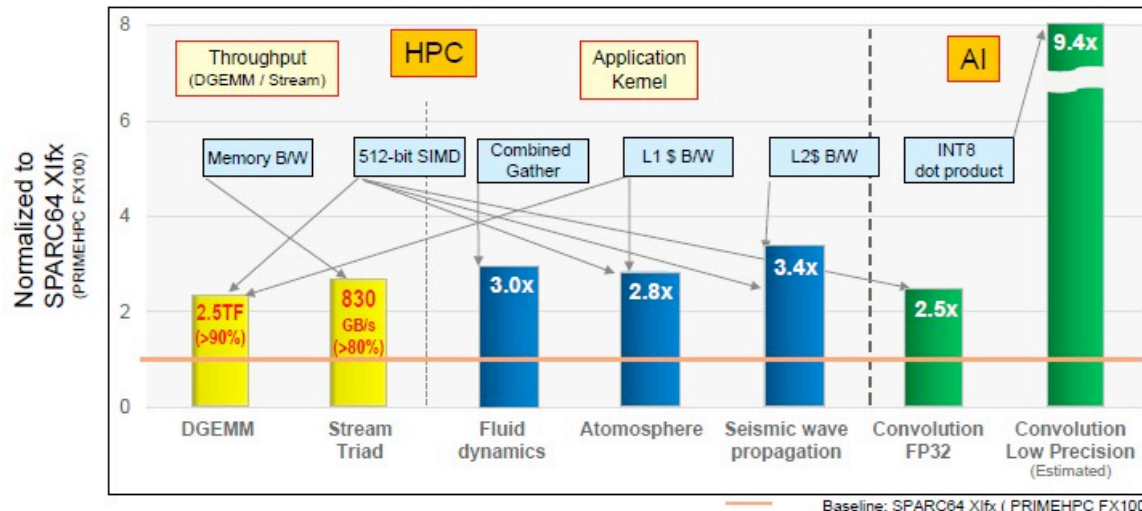
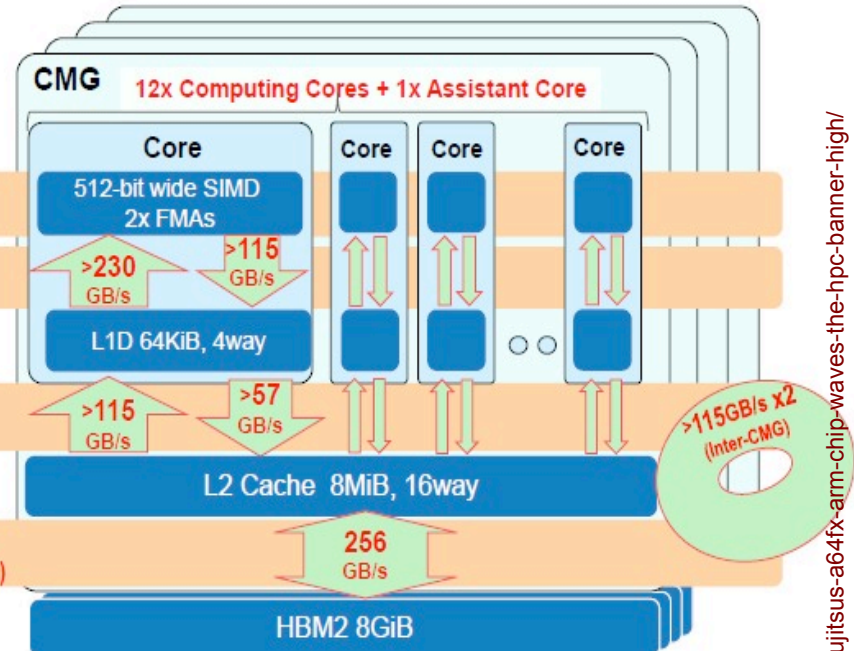


Performance
>2.7TFLOPS

L1 Cache
>11.0TB/s (BF ratio = 4)

L2 Cache
>3.6TB/s (BF ratio = 1.3)

Memory
1024GB/s (BF ratio ≈ 0.37)





HPCs with ARMv8: HPE Apollo 70, Astra



Sandia
National
Laboratories

11 | Vanguard-Astra System Packaging

HPE Apollo 70 Chassis: 4 nodes



HPE Apollo 70 Rack

18 chassis/rack

72 nodes/rack

3 IB switches/rack
(one 36-port switch
per 6 chassis)

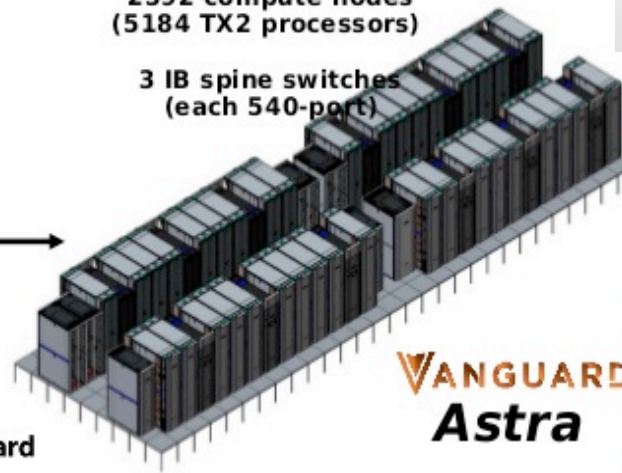


Hewlett Packard
Enterprise

36 compute racks
(9 scalable units, each 4 racks)

2592 compute nodes
(5184 TX2 processors)

3 IB spine switches
(each 540-port)

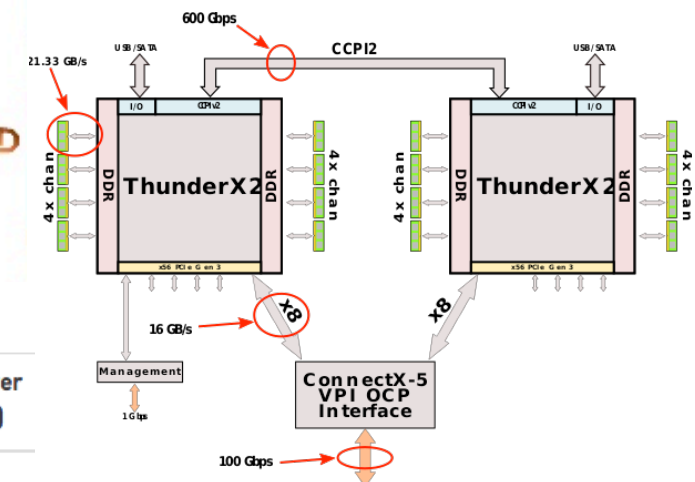


VANGUARD
Astra



HPE Apollo 70
Cavium TX2 Node

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



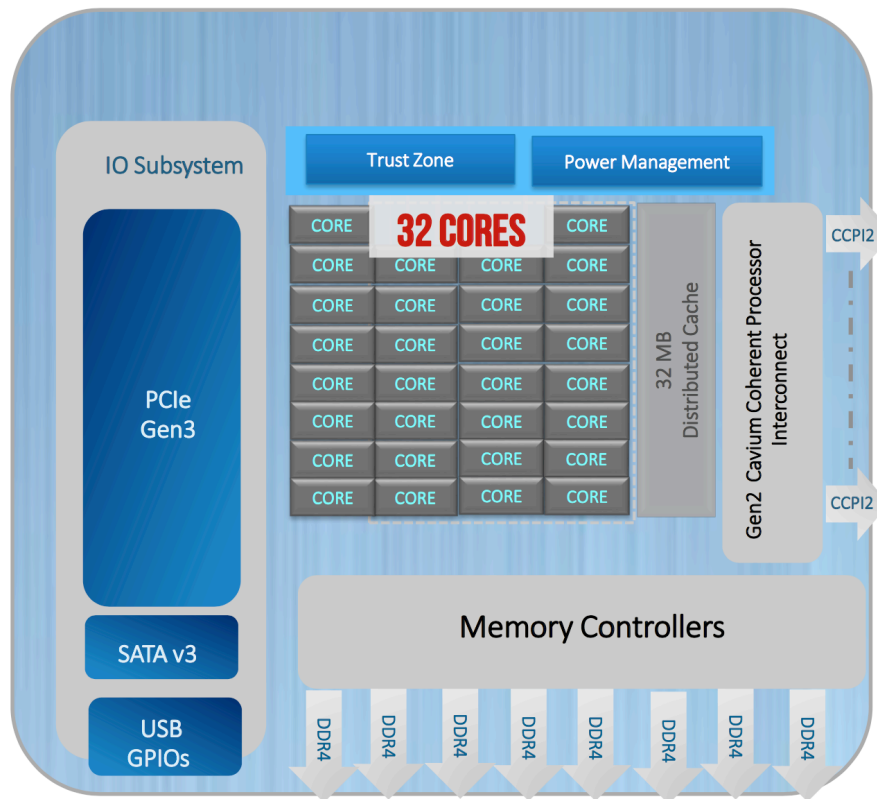
RANKING

List	Rank	System	Vendor	Total Cores	Rmax (TFlops)	Rpeak (TFlops)	Power (kW)
11/2018	204	Apollo 70, Cavium ThunderX2 CN9975-2000 28C 2GHz, 4xEDR Infiniband	HPE	125,328	1,529.0	2,005.2	



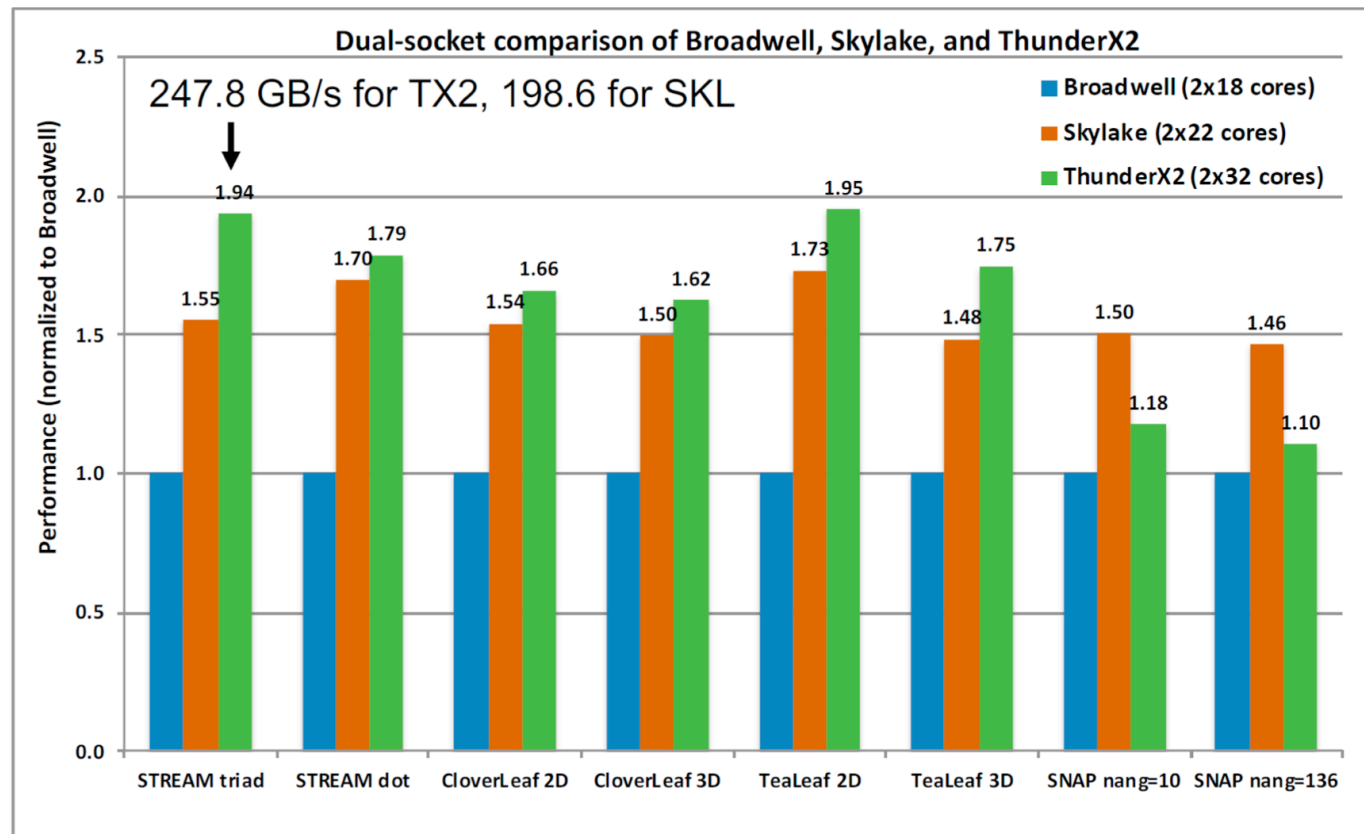
HPC systems with ARMv8: the Marvell/Cavium ThunderX2

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)

Broadwell, Skylake & ThunderX2



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.

Bull sequana compute blade: X1310 Cavium ThunderX2™ – ARMv8 processor

**MONT
BLANC**



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

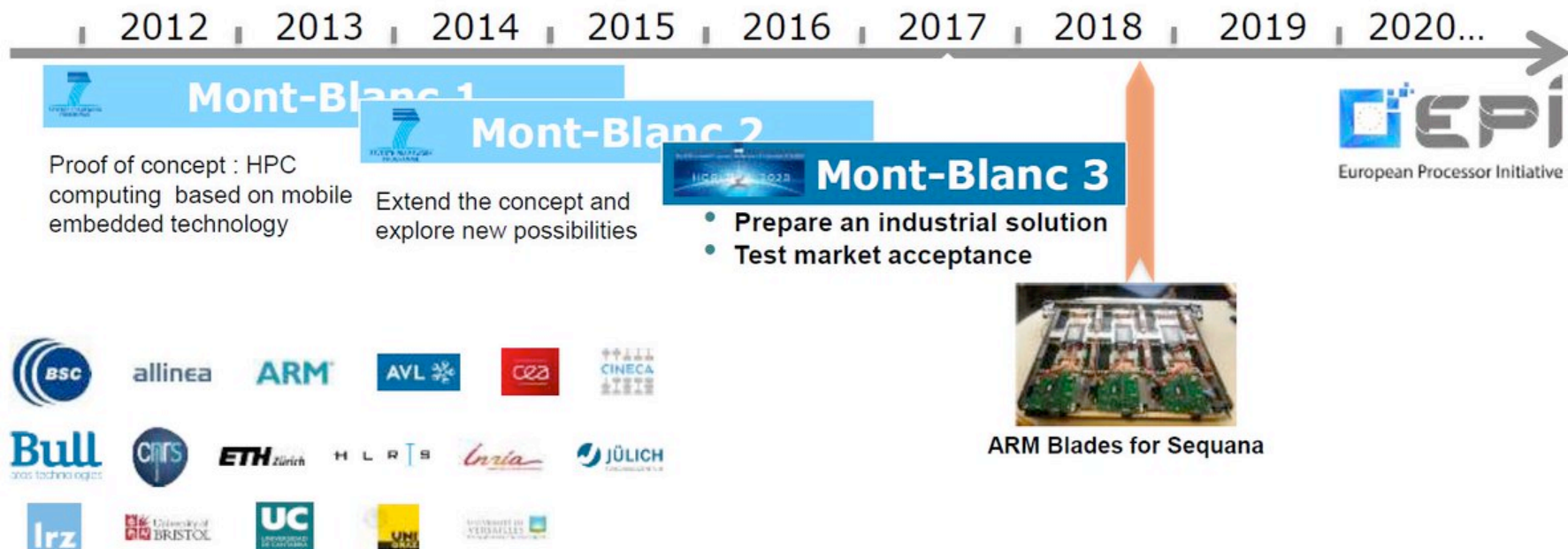


AJProença, Advanced Architectures, MiEI, UM

HPC systems with ARMv8: Mont-Blanc roadmap at BSC



Atos HPC and ARM *A long time engagement*



2 | 7-3-2018 | Eric Eppe | © Atos - Commercial in confidence
GBL HPC & Quantum Solution | HPC User Forum – TERATEC – Bruyères-le-Châtel

MONT-BLANC