Master Informatics Eng.

2022/23

A.J.Proença

Top HPC systems in TOP500 lists

(some images are borrowed)

What is TOP500?

TOP500

From Wikipedia, the free encyclopedia

The **TOP500** project ranks and details the 500 most powerful non-distributed computer systems in the world. The project was started in 1993 and publishes an updated list of the supercomputers twice a year. The first of these updates always coincides with the International Supercomputing Conference in June, and the second is presented at the ACM/IEEE Supercomputing Conference in November. The project aims to provide a reliable basis for tracking and detecting trends in high-performance computing and bases rankings on HPL,[1] a portable implementation of the high-performance LINPACK benchmark written in Fortran for distributed-memory computers. Currently the latest TOP500 list is the 60th, published in November 2022. Since June 2022, USA's Frontier is the world's most powerful supercomputer, reaching 1102 petaFlops (1.102 exaFlops) on the LINPACK benchmarks.[2] The United States has by far the highest share of total computing power on the list (nearly 50%),[3] while China currently leads the list in number of systems with 173 supercomputers, with the USA not far behind in second place.

TOP500

Key people	Erich Strohmaier Jack Dongarra Horst Simon Martin Meuer
Established	24 June 1993
Website	www.top500.org

149

The TOP500 list is compiled by Jack Dongarra of the University of Tennessee, Knoxville, Erich Strohmaier and Horst Simon of the National Energy Research Scientific Computing Center (NERSC) and Lawrence Berkeley National Laboratory (LBNL), and, until his death in 2014, Hans Meuer of the University of Mannheim, Germany.

The TOP500 project lists also Green500 and HPCG benchmark list.

LINPACK benchmarks (HPL)

HPL, written in C, measures the sustained floating-point rate (GFLOPs/s) to solve a dense system of linear equations using double-precision floating-point arithmetic

LINPACK benchmarks

From Wikipedia, the free encyclopedia

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.

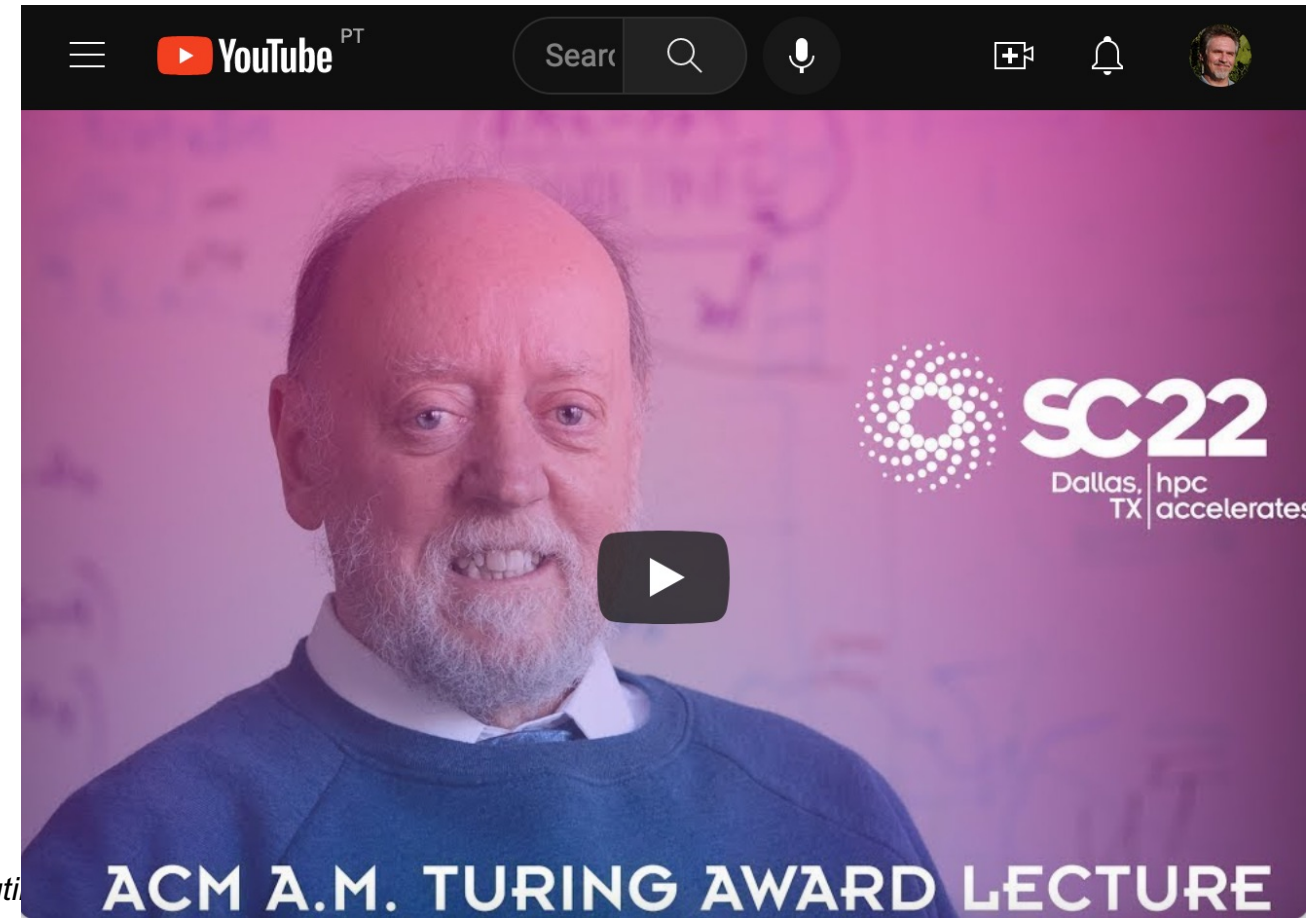
The latest version of these benchmarks is used to build the TOP500 list, ranking the world's most powerful supercomputers.[1]

LINPACK benchmarks

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

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 provide a good correction over the peak performance provided by the manufacturer. The peak performance is the maximal theoretical performance a computer can achieve, calculated as the machine's frequency, in cycles per second, times the number of operations per cycle it can perform. The actual performance will always be lower than the peak performance.[2] The performance of a computer is a complex

ACM Turing Award 2021: Jack Dongarra



https://www.youtube.com/watch?v=IsnRP9akCDk

Outline

1. **TOP500** *(LINPACK)*
   a) TOP10 lists from Nov'17 to Nov'22
   b) Analysis of some relevant systems and architectures
   c) Country distribution over the past 25 years
   d) PU chip technology evolution in the past 25 years & since last year
   e) Evolution of the accelerators since they were available
2. **GREEN500**
   a) TOP5 list from Nov'19 to Nov'22
   b) Analysis of some relevant systems and architectures
3. **HPCG**
   a) HPCG *vs.* HPL: an overview
   b) TOP10 in Nov'22
4. **HPL-AI**
   a) High-performance Linpack (HPL) and artificial intelligence (AI) workloads

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

3x systems in Top10
w/ Xeon Phi KNL

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

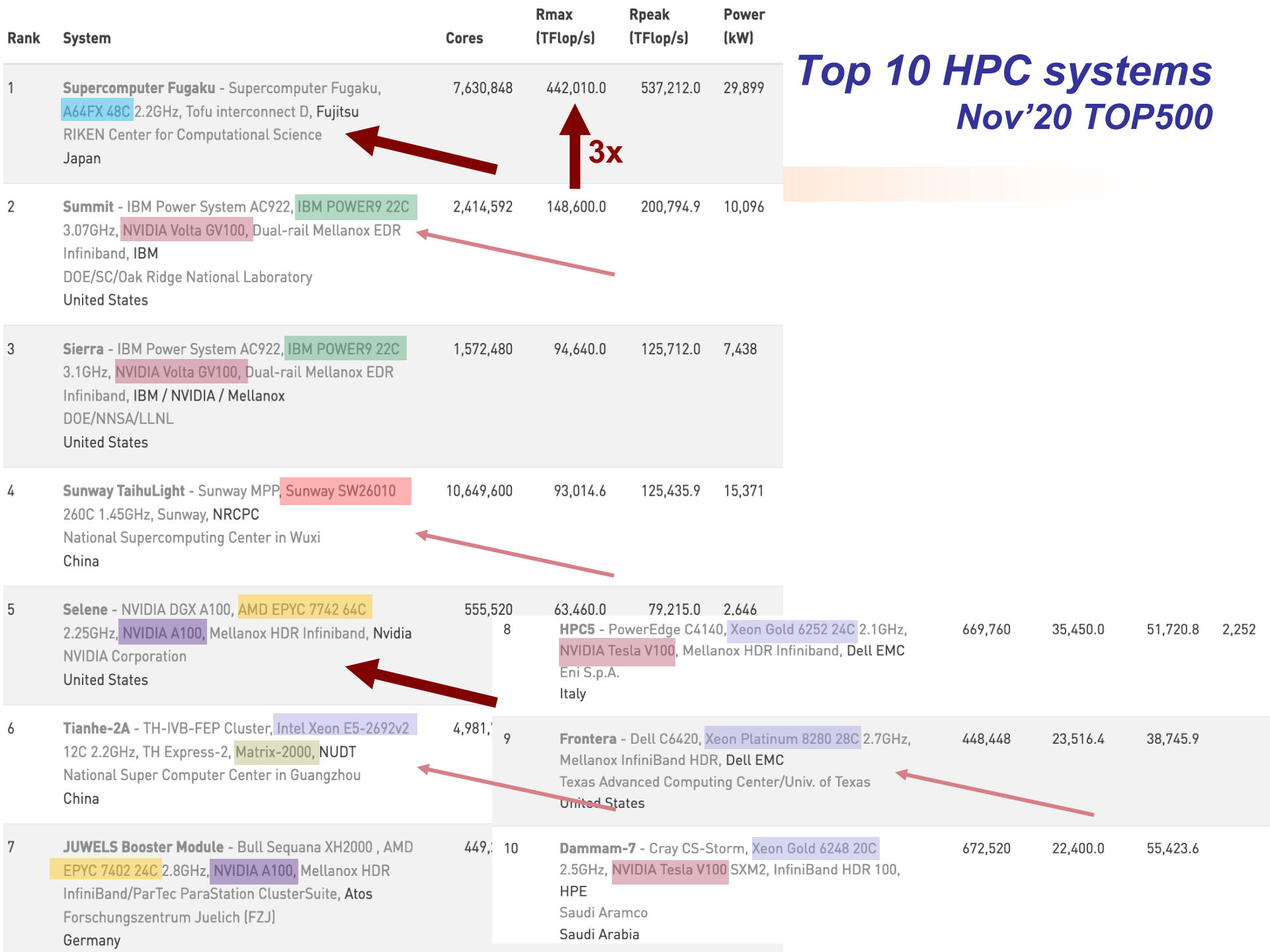
Top 10 HPC systems
Nov'19 TOP500

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	

Frontera (TACC):
successor of Stampede2

Top 10 HPC systems
Nov'20 TOP500

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442,010.0	537,212.0	29,899
2	**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
3	**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
4	**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
5	**Selene** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia NVIDIA Corporation United States	555,520	63,460.0	79,215.0	2,646
6	**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,[illegible]			
7	**JUWELS Booster Module** - Bull Sequana XH2000 , AMD EPYC 7402 24C 2.8GHz, NVIDIA A100, Mellanox HDR InfiniBand/ParTec ParaStation ClusterSuite, Atos Forschungszentrum Juelich (FZJ) Germany	449,[illegible]			
8	**HPC5** - PowerEdge C4140, Xeon Gold 6252 24C 2.1GHz, NVIDIA Tesla V100, Mellanox HDR Infiniband, Dell EMC Eni S.p.A. Italy	669,760	35,450.0	51,720.8	2,252
9	**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	
10	**Dammam-7** - Cray CS-Storm, Xeon Gold 6248 20C 2.5GHz, NVIDIA Tesla V100 SXM2, InfiniBand HDR 100, HPE Saudi Aramco Saudi Arabia	672,520	22,400.0	55,423.6	



Top 10 HPC systems
Nov'21 TOP500

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442,010.0	537,212.0	29,899
2	**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
3	**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
4	**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
5	**Perlmutter** - HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10, HPE DOE/SC/LBNL/NERSC United States	761,856	70,870.0	93,750.0	2,589
6	**Selene** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia NVIDIA Corporation United States	55			
7	**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	4,98			
8	**JUWELS Booster Module** - Bull Sequana XH20[illegible], AMD EPYC 7402 24C 2.8GHz, NVIDIA A100, Mellano[illegible] HDR InfiniBand/ParTec ParaStation ClusterSuite, [illegible]tos Forschungszentrum Juelich (FZJ) Germany	449,280	44,120.0	70,980.0	1,764
9	**HPC5** - PowerEdge C4140, Xeon Gold [illegible]52 24C 2.1GHz, NVIDIA Tesla V100, Mellanox HDR In[illegible]iband, DELL EMC Eni S.p.A. Italy	669,760	35,450.0	51,720.8	2,252
10	**Voyager-EUS2** - ND96amsr_A100_v4, AMD EPYC 7V12 48C 2.45GHz, NVIDIA A100 80GB, Mellanox HDR Infiniband, Microsoft Azure Azure East US 2 United States	253,440	30,050.0	39,531.2	

New in the TOP10 list since June'21, with AMD Milan (Zen3)

Microsoft Azure, pushed TACC Frontera to 13th…

Top 10 HPC systems
Nov'22 TOP500

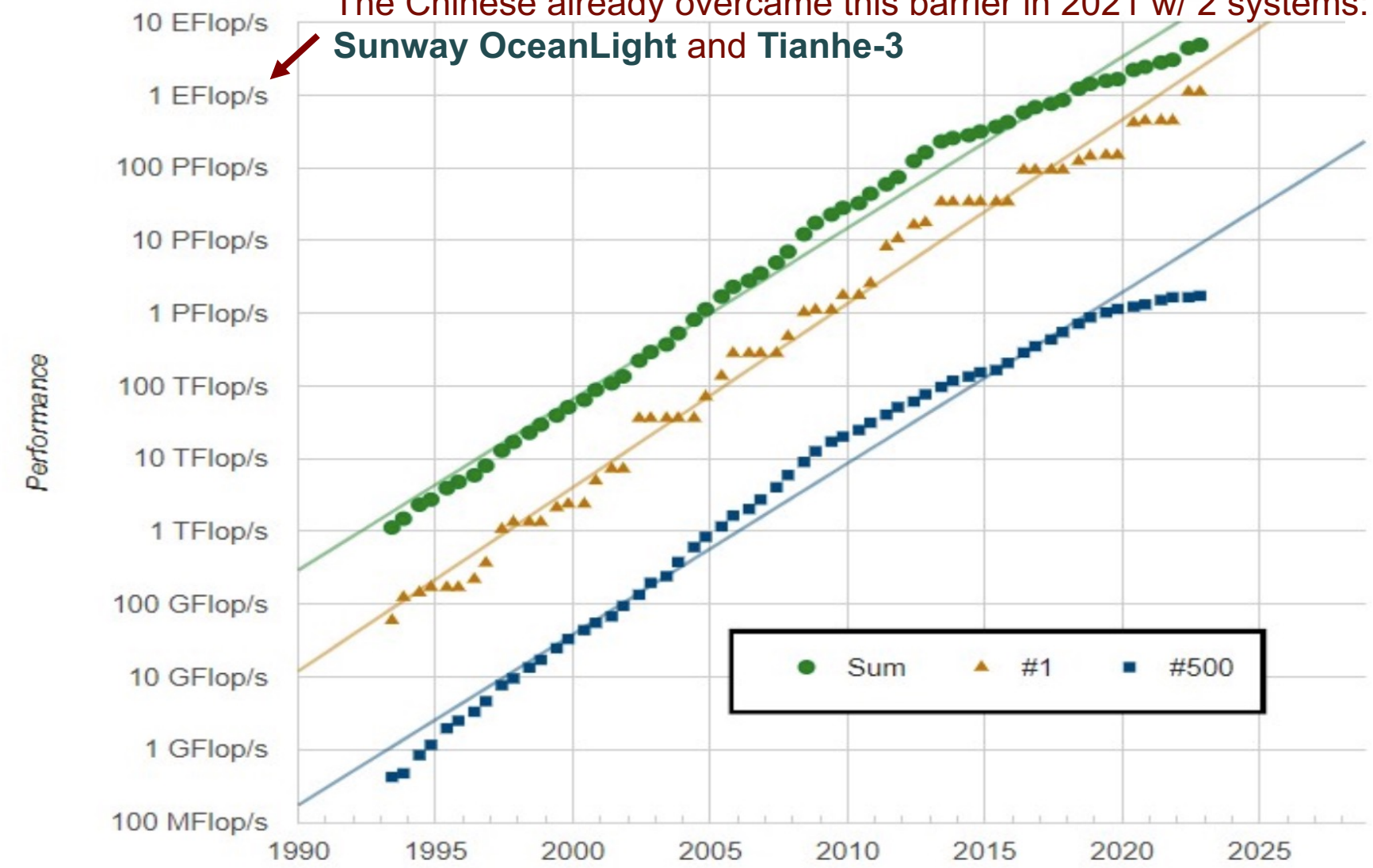
Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,730,112	1,102.00	1,685.65	21,100
2	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
3	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,220,288	309.10	428.70	6,016
4	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, Atos EuroHPC/CINECA Italy	1,463,616	174.70	255.75	5,610
5	**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	148.60	200.79	10,096
6	**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				
7	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC National Supercomputing Center in Wuxi China				
8	**Perlmutter** - HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10, HPE DOE/SC/LBNL/NERSC United States	761,856	70.87	93.75	2,589
9	**Selene** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia NVIDIA Corporation United States	555,520	63.46	79.22	2,646
10	**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	61.44	100.68	18,482

Exascale!

New in the TOP10 list since June'22, with AMD Milan "Trento" & AMD Radeon Instinct MI250X

New in the TOP10 list, with 3rd Gen Xeon & A100

Peak performance
1993 to 2022



https://www.nextplatform.com/2022/11/14/testing-the-mettle-of-the-top-supercomputing-iron-on-earth/

Analysis of the key systems in 2022

1. **#1** in Nov'22: **Frontier** *(AMD Epyc Trento 64c + AMD Instinct MI250x)*
2. **#2** in Nov'22 (#1 in Jun'20): **Fugaku** *(Fujitsu A64FX, 48 cores)*
3. **#4** in Nov'22: **Leonardo**, *(3rd Gen Xeon, 32c + NVidia Ampere A100)*
4. **#5** in Nov'22 (#1 in Nov'18): **Summit** *(IBM POWER9, 22 cores + NVidia Volta GV100)* **+Sierra,** follow-up of #1 in Jun'12, **Sequoia** *(IBM POWER BGQ, 16 cores)*
5. **#7** in Nov'22 (#1 in Nov'17): **TaihuLight** *(Sunway SW26010, 260 c)*
6. **#9** in Nov'22: **Selene** *(AMD Epyc Rome 64 c + NVidia A100)*
7. **#10** in Nov'22:**Tianhe-2A** (MilkyWay-2A) *(Xeon, 12c + Matrix-2000)*, follow-up of #1 in Jun'13, **Tianhe-2** (MilkyWay-2) *(Xeon, 6 c + Xeon Phi 31S1P)* #1 in Nov'10, **Tianhe-1A** (MilkyWay-1A) *(Xeon, 6 c + NVidia Fermi)*

Frontier at ORNL

TOP 500 The List.

1 **Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE
DOE/SC/Oak Ridge National Laboratory
United States

Nov'22

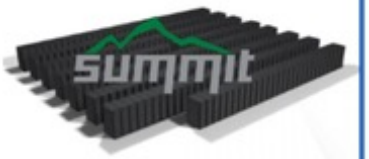
Competitive procurement asking for:

- 50–100× application performance of Titan
- Support for traditional modeling and simulation, high-performance data analysis, and artificial intelligence applications
- Peak performance >1000 PF
- Smooth transition for existing and future applications

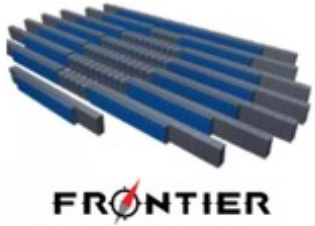
OAK RIDGE National Laboratory

2008 — Jaguar: 2.3 PF
World's Fastest

2012 — Titan: 27 PF
Accelerated Computing
World's Fastest

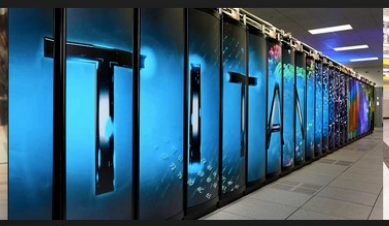


2017 — Summit: 200 PF
Accelerated Computing
5–10× Titan Performance



2021 — Frontier: >1000 PF
Competitive Procurement
5-10× Summit
Performance

JAGUAR
No. 1 in 2009, 2010



TITAN
No. 1 in 2012

SUMMIT
No. 1 in 2018

FRONTIER
No. 1 in 2022

Frontier architecture

Frontier Overview

Extraordinary Engineering

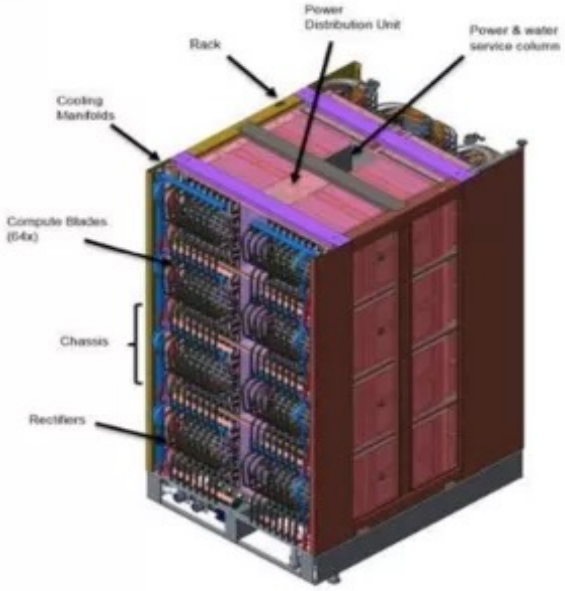
System

- 2 EF Peak DP FLOPS
- 74 compute racks
- 29 MW Power Consumption
- 9,408 nodes
- 9.2 PB memory (4.6 PB HBM, 4.6 PB DDR4)
- Cray Slingshot network with dragonfly topology
- 37 PB Node Local Storage
- 716 PB Center-wide storage
- 4000 ft^2 foot print

Built by HPE

Olympus rack

- 128 AMD nodes
- 8,000 lbs
- Supports 400 KW



Powered by AMD

AMD node

- 1 AMD "Trento" CPU
- 4 AMD MI250X GPUs
- 512 GiB DDR4 memory on CPU
- 512 GiB HBM2e total per node (128 GiB HBM per GPU)
- Coherent memory across the node
- 4 TB NVM
- GPUs & CPU fully connected with AMD Infinity Fabric
- 4 Cassini NICs, 100 GB/s network BW

Compute blade

- 2 AMD nodes

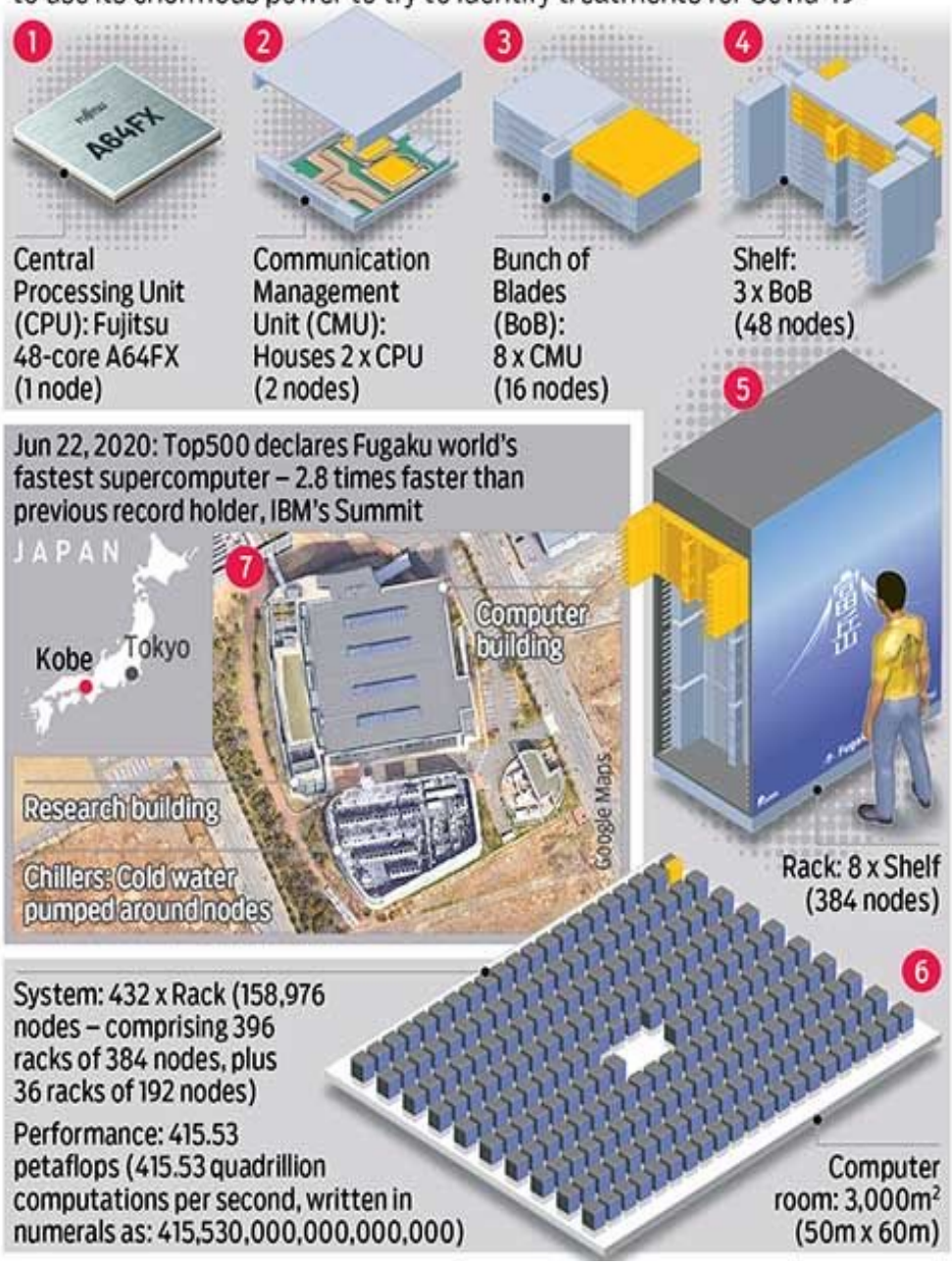
All water cooled, even DIMMS and NICs

2 富岳 Fugaku

2 **Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu
RIKEN Center for Computational Science
Japan
Nov'22

Supercomputer to seek Covid-19 cure

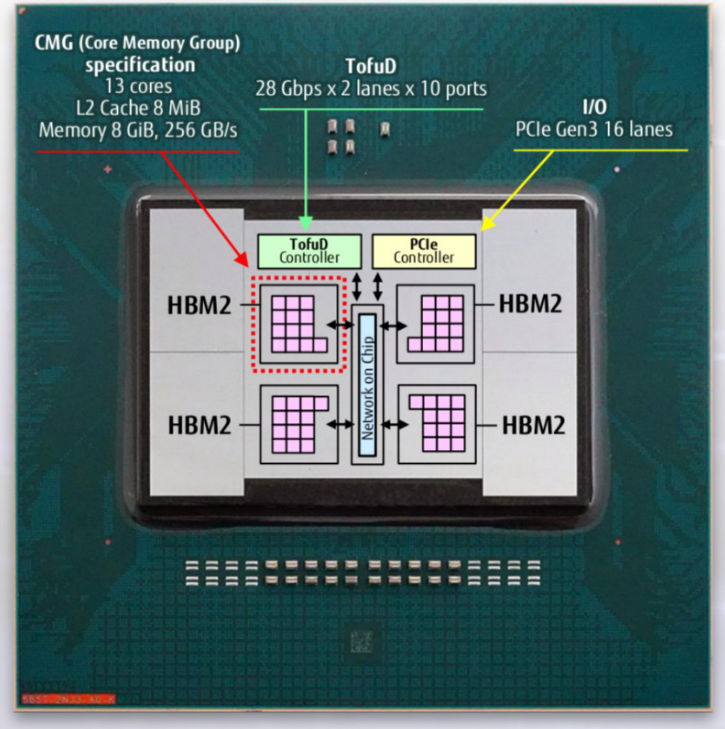
The world's fastest supercomputer, Japan's $1.2 billion Fugaku, is to use its enormous power to try to identify treatments for Covid-19



Fujitsu A64FX in Fugaku

1. High-Performance Arm CPU A64FX in HPC and AI Areas

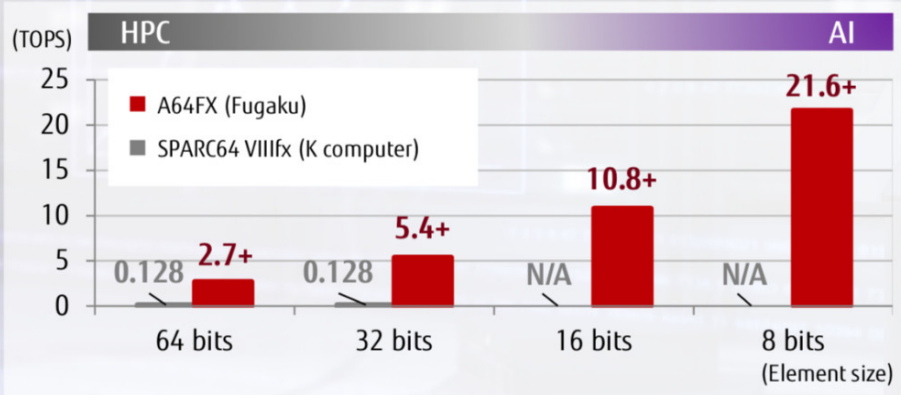
FUJITSU



Architecture features

ISA	Armv8.2-A (AArch64 only) SVE (Scalable Vector Extension) arm
SIMD width	512-bit
Precision	FP64/32/16, INT64/32/16/8
Cores	48 computing cores + 4 assistant cores (4 CMGs)
Memory	HBM2: Peak B/W 1,024 GB/s
Interconnect	TofuD: 28 Gbps x 2 lanes x 10 ports

Peak performance (Chip level)



4

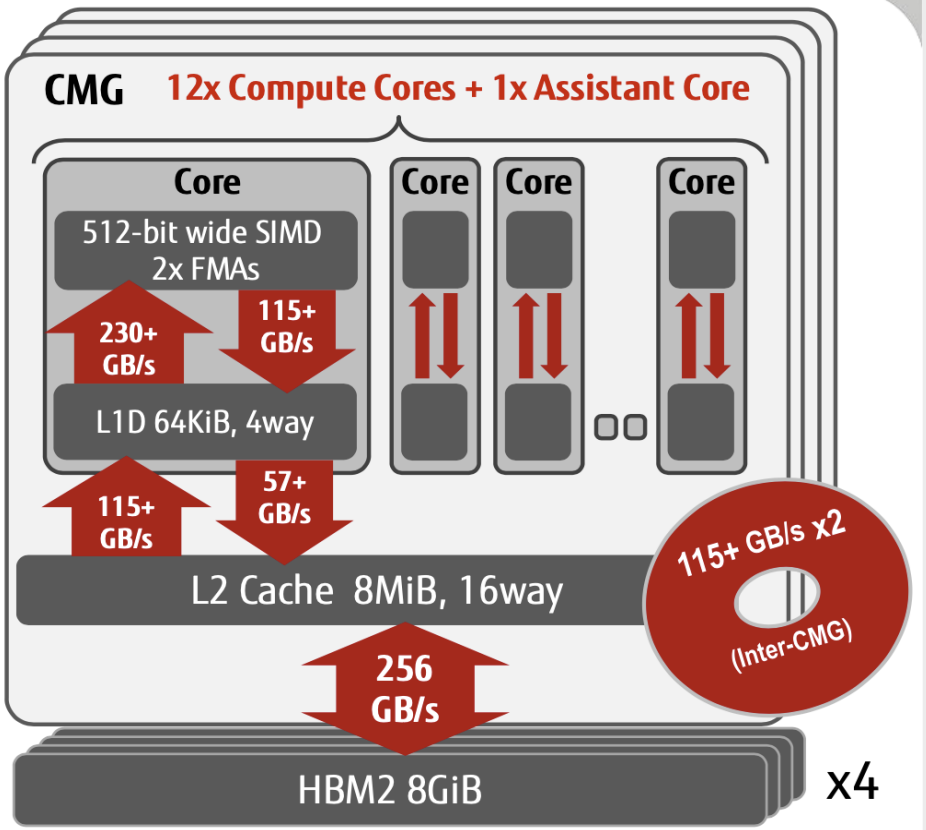
Fujitsu A64FX in Fugaku

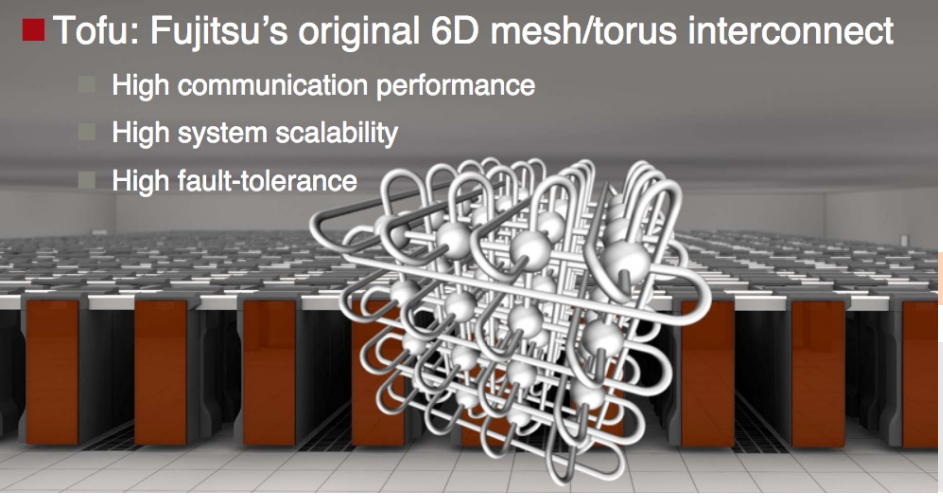
Fujitsu-designed CPU Core w/ High Memory Bandwidth

FUJITSU

- A64FX out-of-order controls in cores, caches, and memories achieve superior throughput

BW and calc. perf.	A64FX	B/F
DP floating perf. (TFlops)	2.7+	-
L1 data cache (TB/s)	11+	4
L2 cache (TB/s)	3.6+	1.3
Memory BW (GB/s)	1024	0.37

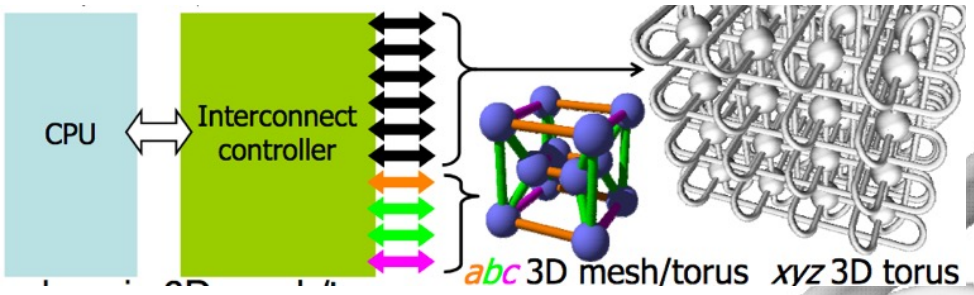


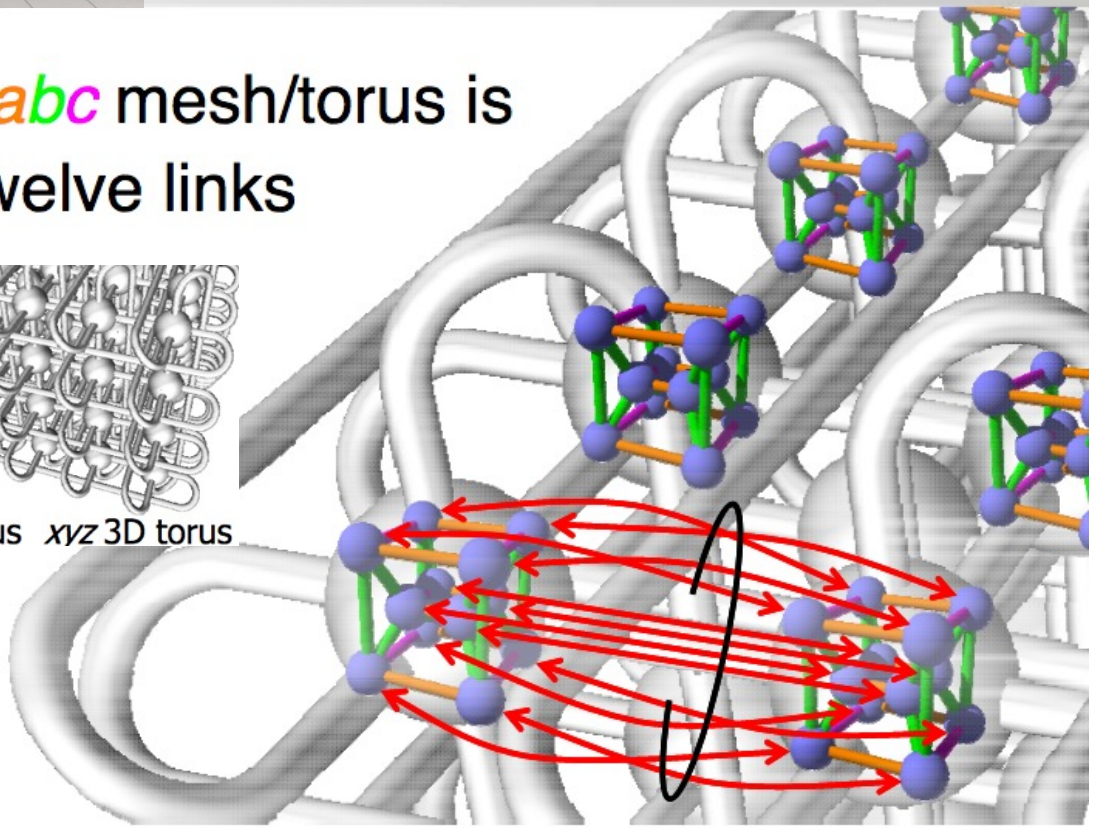


Tofu3: 6D mesh/torus interconnect

FUJITSU

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





3

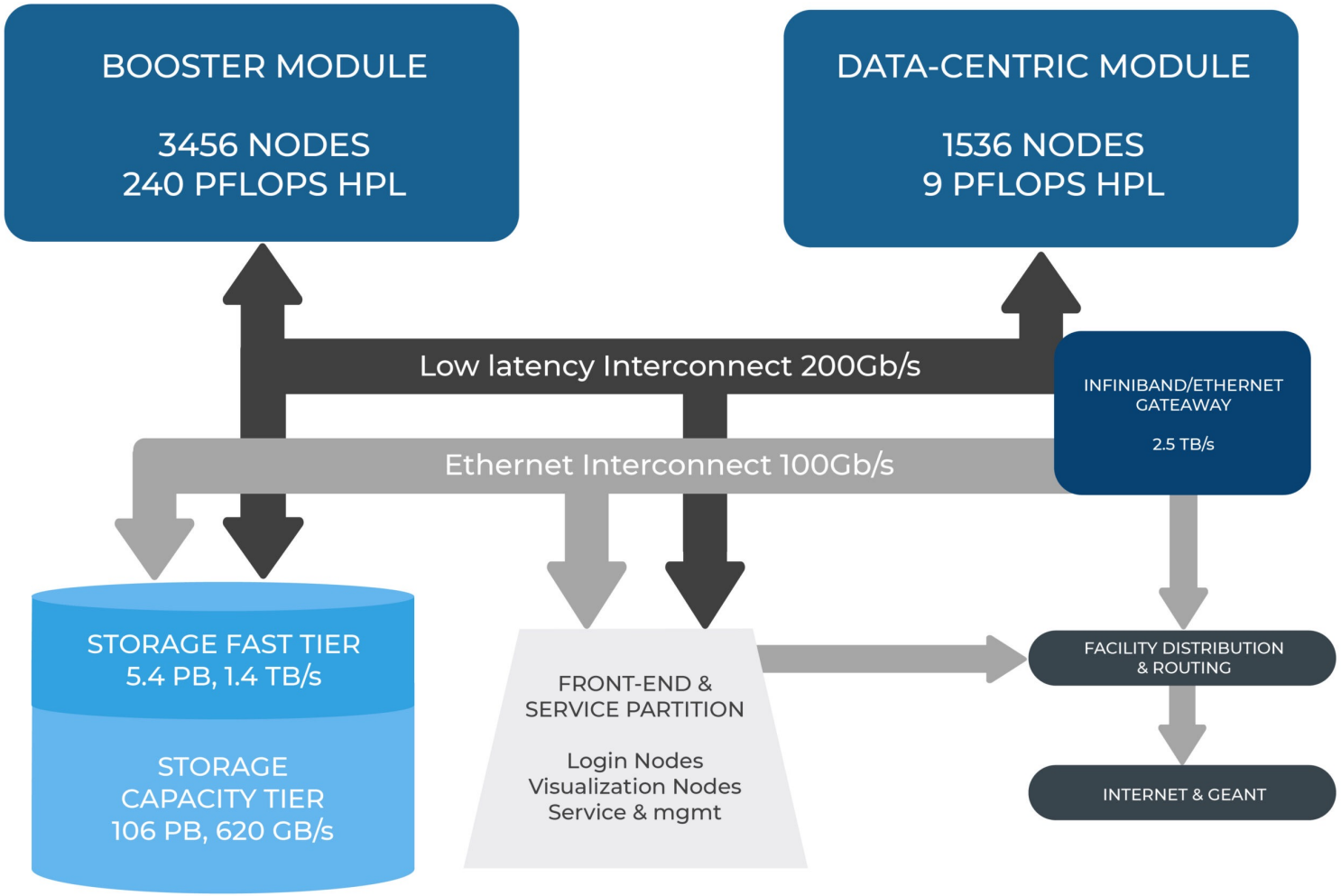
EuroHPC Supercomputers at the top5 in TOP500

TOP500 The List. Nov'22

LUMI

Similar to **Frontier**

Leonardo at CINECA



https://leonardo-supercomputer.cineca.eu/hpc-system/

Leonardo at CINECA

BOOSTER MODULE — 3456 nodes

Features a **custom BullSequana X2135 "Da Vinci" blade**, composed of:

Ice Lake

- 1 x CPU Intel Xeon 8358 32 cores, 2,6 GHz
- 512 (8 x 64) GB RAM DDR4 3200 MHz
- 4 x NVidia custom Ampere GPU 64GB HBM2
- 2 x NVidia HDR 2×100 Gb/s cards

Performance per node: 89,4 TFLOPs peak

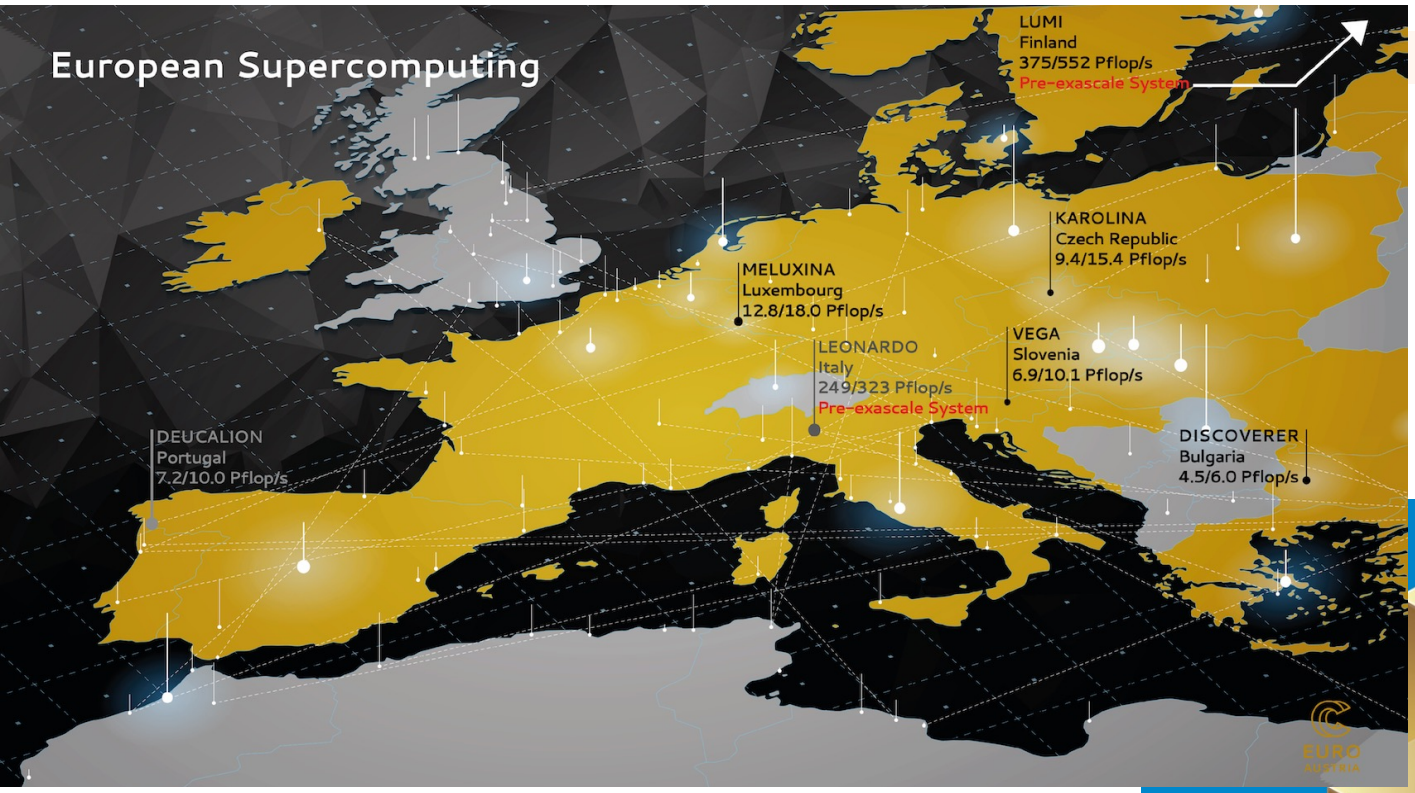
DATA CENTRIC MODULE — 1536 nodes

Features a **BullSequana X2610 compute blade.**

Each computing node is composed of:

- 2x Intel Sapphire Rapids, 56 cores, TDP 350 W
- 512 (16 x 32) GB RAM DDR5 4800 MHz
- 1 x NVidia HDR100 100 Gb/s card
- 8 TB NVM

European Supercomputing



EuroHPC
Joint Undertaking

IBM POWER9 Summit

(Nov'19 #1 TOP500)

5 TOP500 The List.

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'22

Summit Overview

OpenPOWER

Compute Node

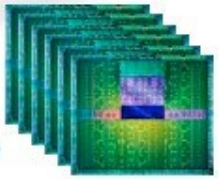
2 x POWER9
6 x NVIDIA GV100
NVMe-compatible PCIe 1600 GB SSD

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



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

Power Systems

IBM

New Core Microarchitecture

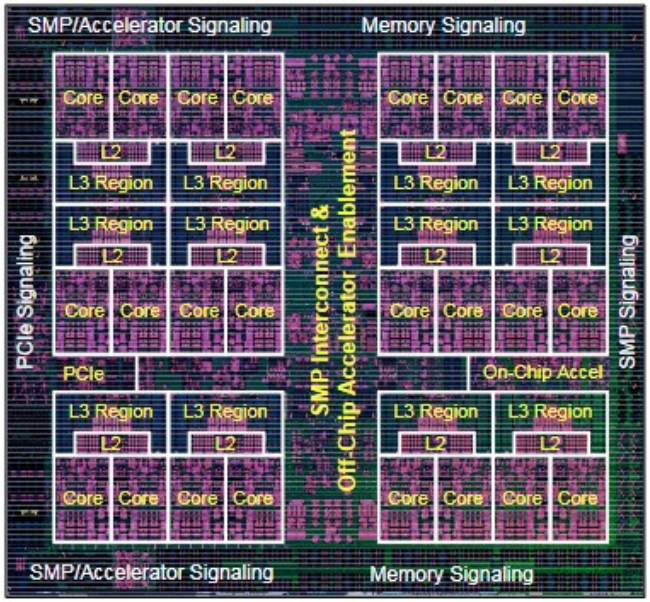
- Stronger thread performance
- Efficient agile pipeline
- POWER ISA v3.0

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)

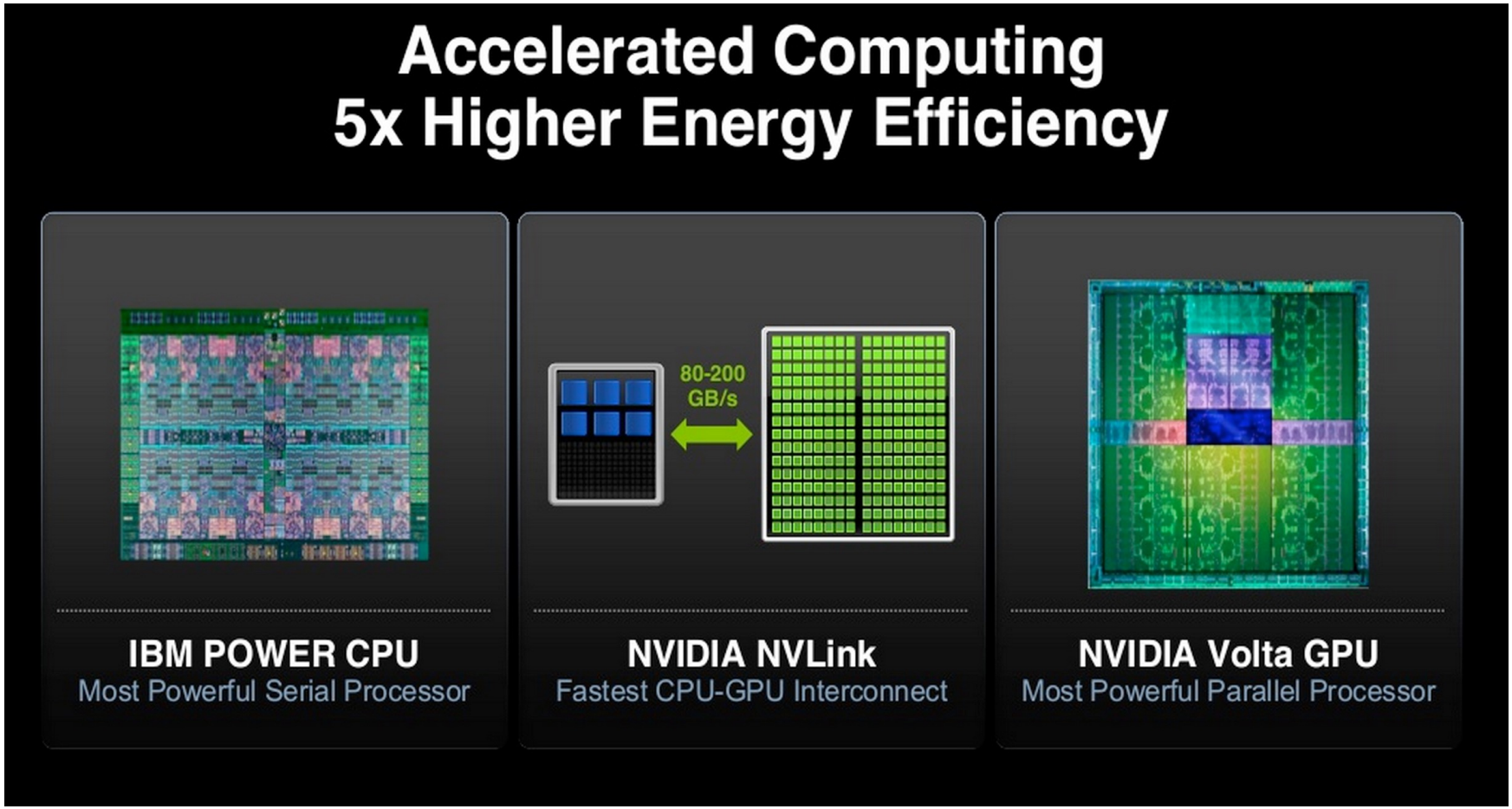
State of the Art I/O Subsystem

- PCIe Gen4 – 48 lanes

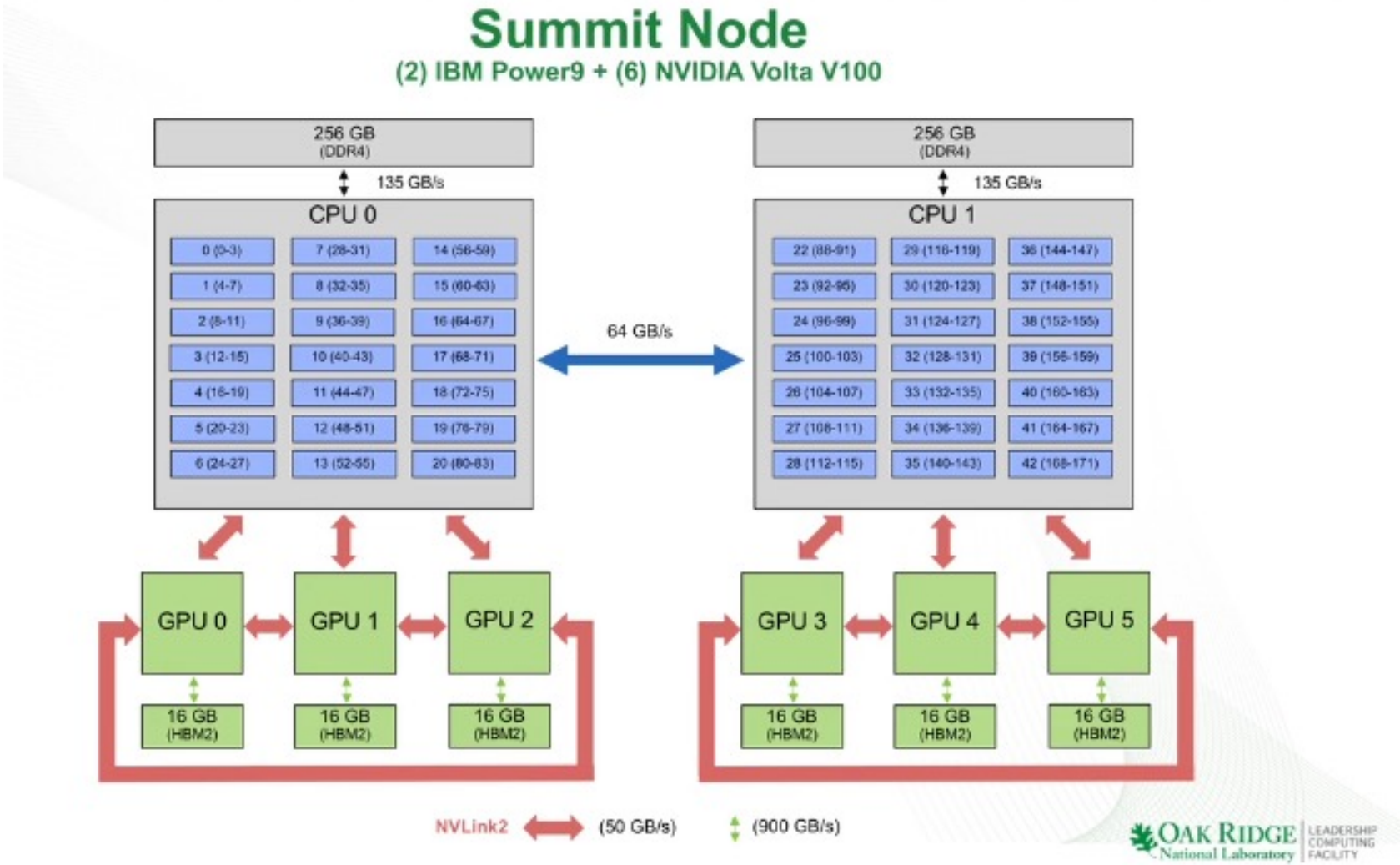
High Bandwidth Signaling Technology

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

IBM POWER9 + NVidia V100



Summit/Sierra node architecture



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

Sunway TaihuLight

(#1 in June'16 TOP500)

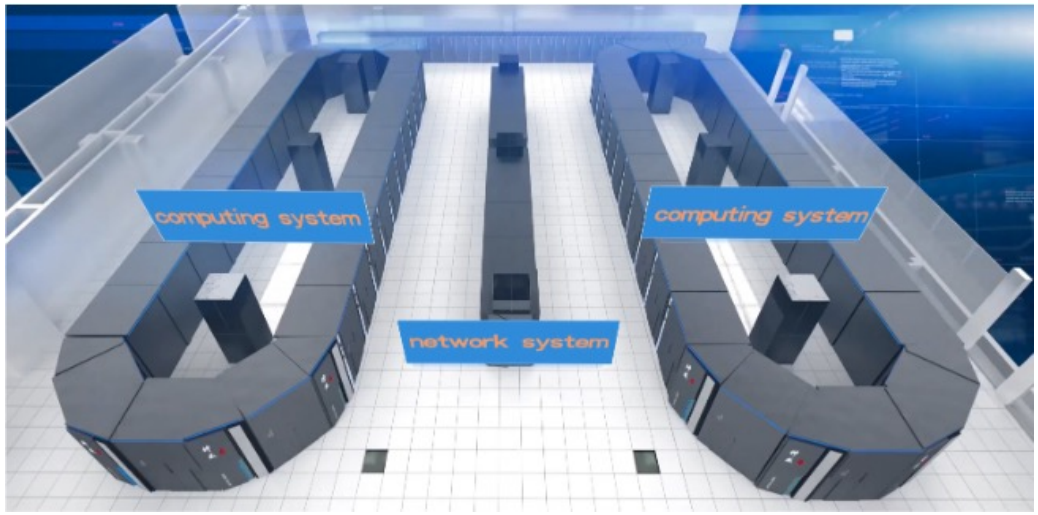
7

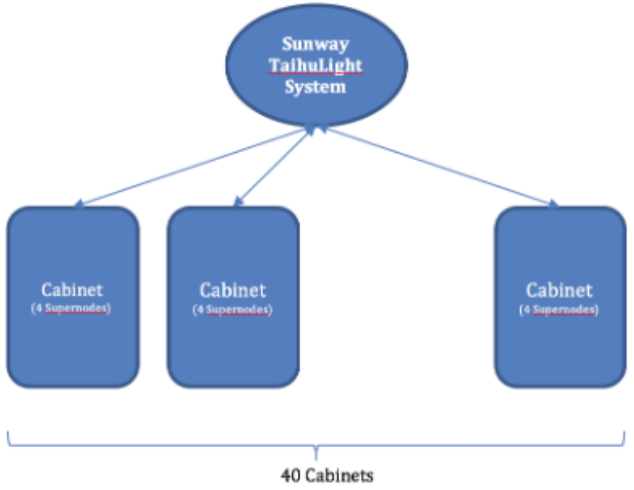
Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC
National Supercomputing Center in Wuxi
China

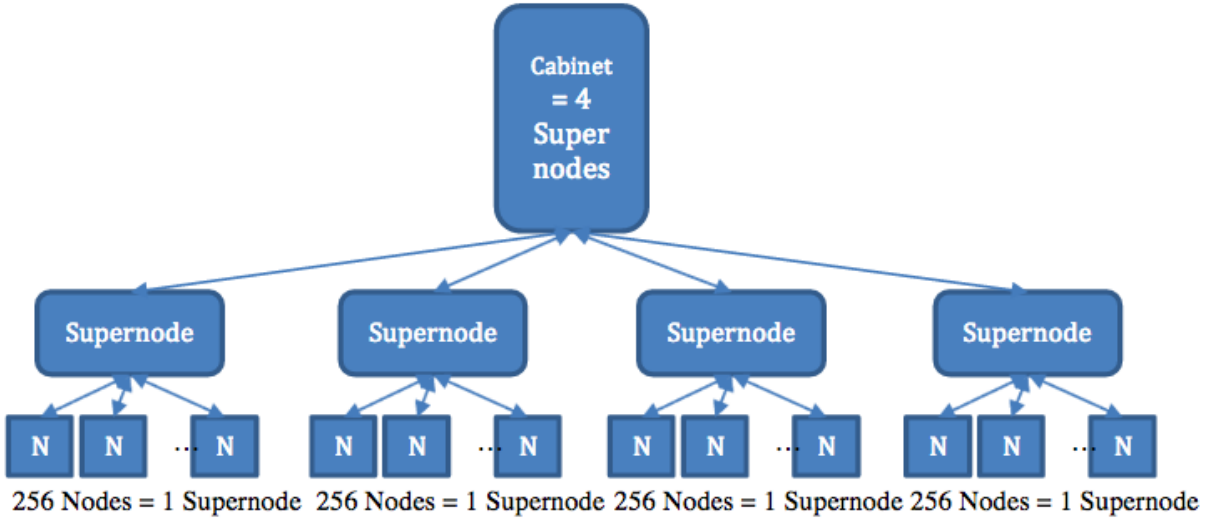
TOP500 The List.

Nov'22

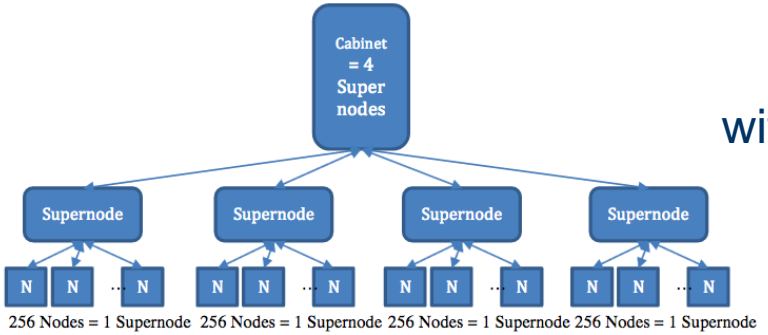
Overview of the Sunway TaihuLight System



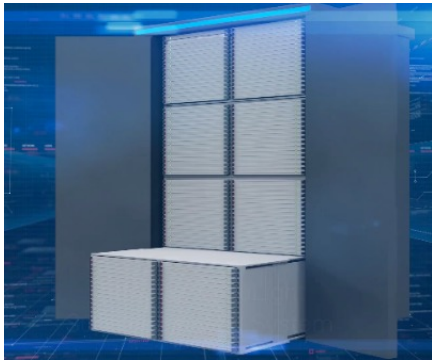




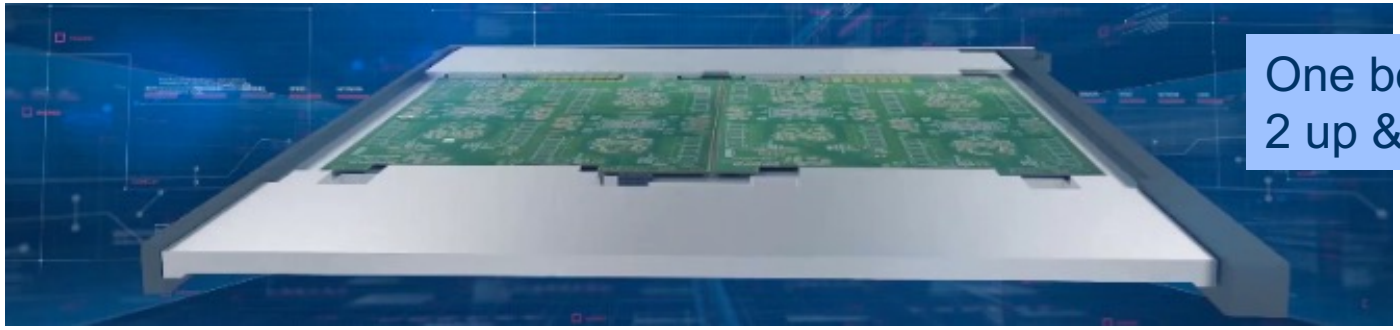
Sunway TaihuLight



One cabinet with 4 Supernodes



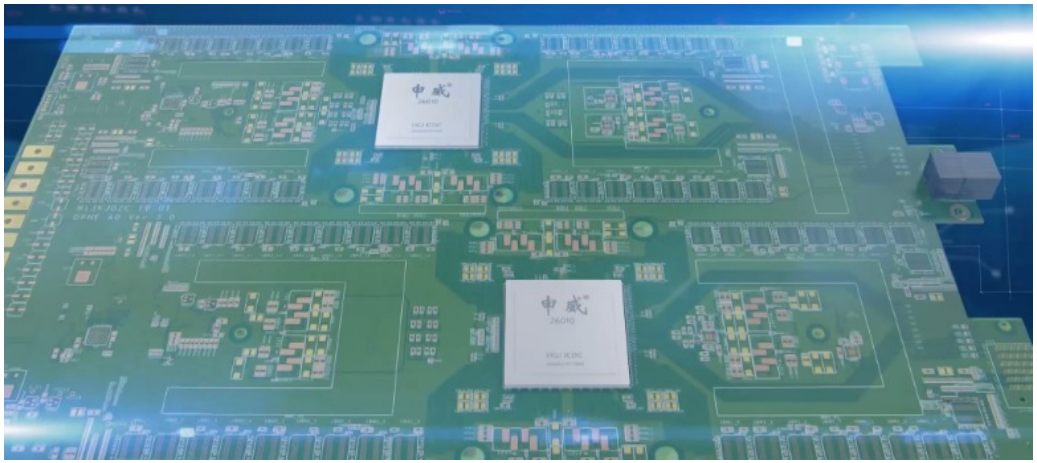
One Supernode with 32 boards



One board with 4 cards, 2 up & 2 down

Sunway TaihuLight

One card w/ two PU devices *(two SW26010 chips)*

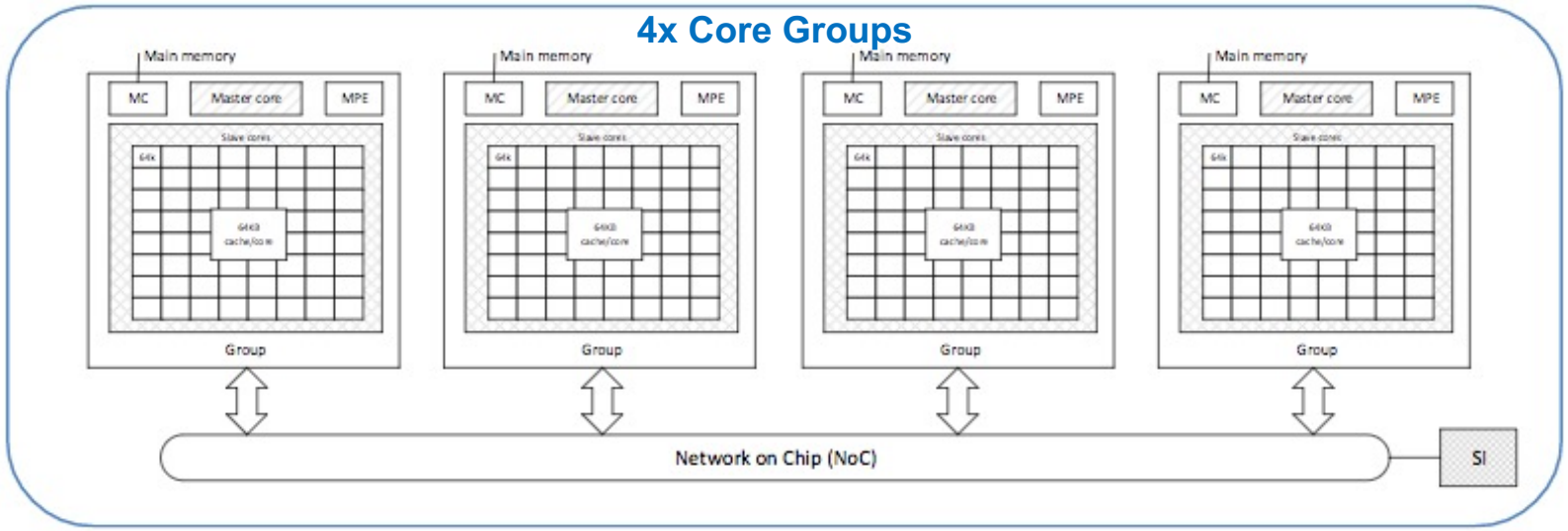


SW26010 chip with
4x NUMA Core Groups (CG).

Each CG follows a hybrid approach:

- 1 fat-core (MPE) for serial work, OoO execution, superscalar, L1 & L2
- 8x8 grid of skinny-cores (CPE), L1 private & L2 shared by the grid
- all cores are 64-bit RISC PU and all support 256-bit vector instructions

SW26010 chip

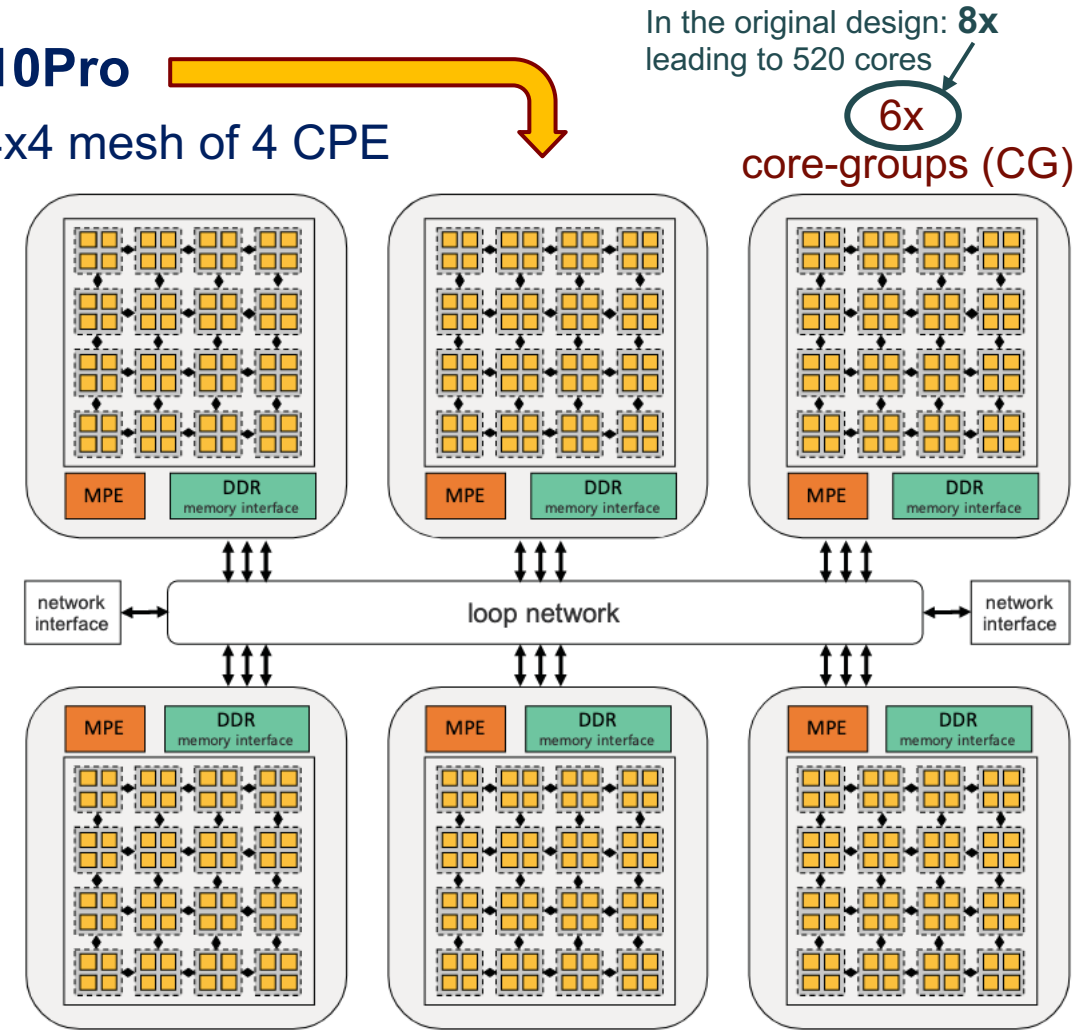


AJProença,

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

The new Sunway OceanLight supercomputer (April'21)

- Based on the chip **SW26010Pro**
  - 6x CG, each 1x fat-core & 4x4 mesh of 4 CPE
  - overall **390 cores** (= 6 x (1 + 64))
  - mem controller at each CG accesses 16 GiB DDR4 with 51.2 GiB/s bandwidth
  - each chip: accesses 96 GiB DDR4 with 307.2 GiB/s bandwidth
- Single-socket nodes
- More nodes, more cores: from 7.6 millions cores to over **41.9 millions cores**!



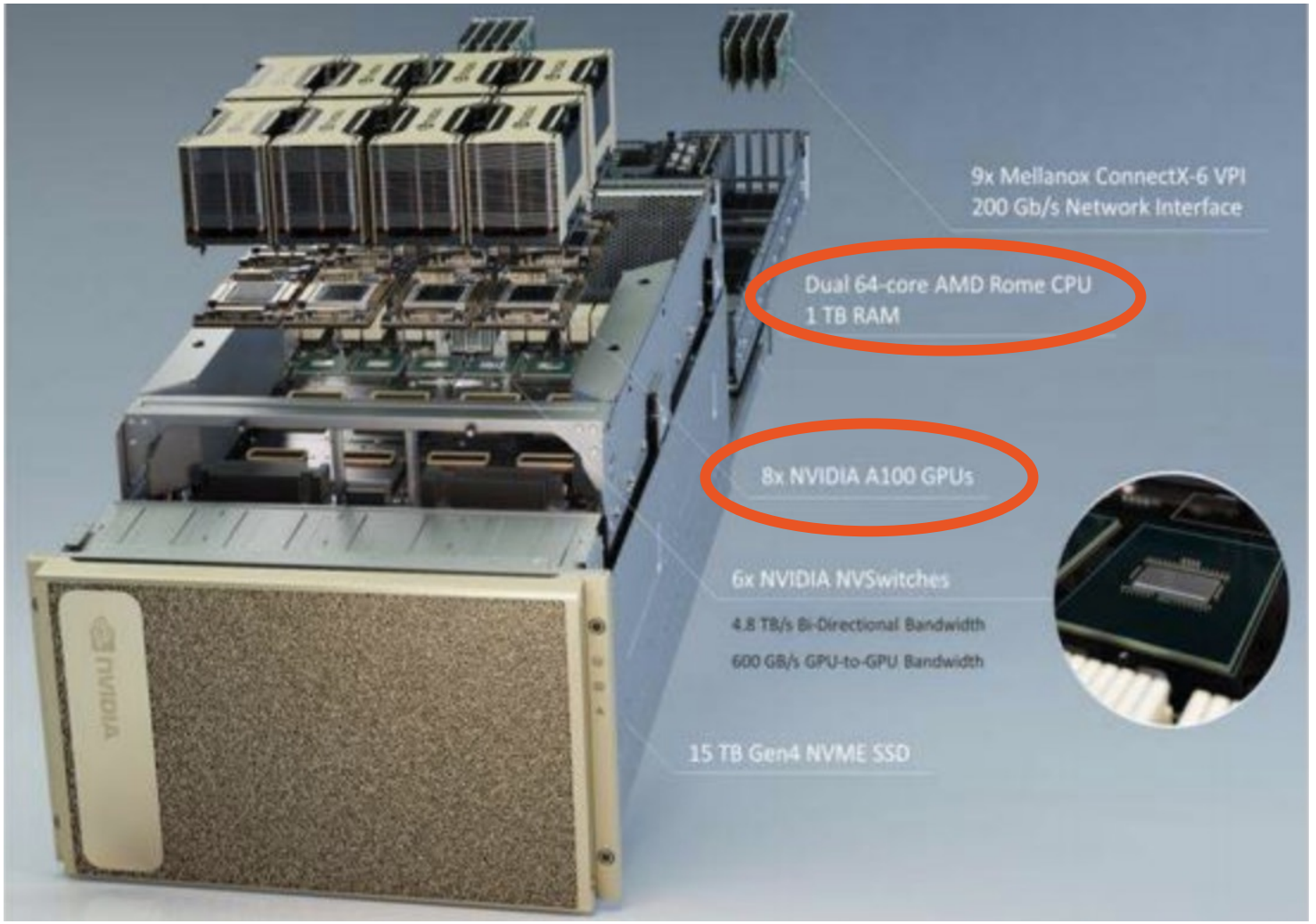
https://dl.acm.org/doi/pdf/10.1145/3458817.3487399

NVidia Selene: 280 DGX A100 nodes

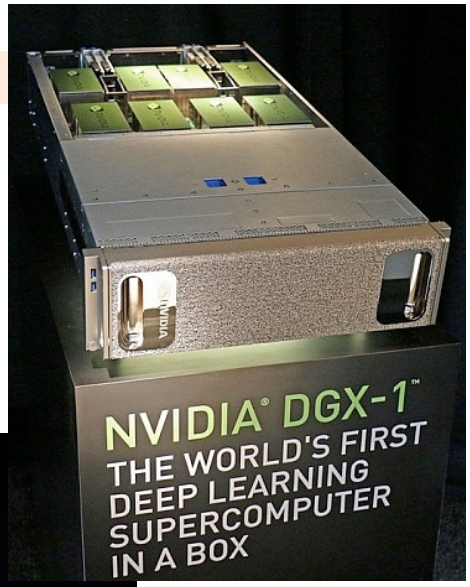
9 **Selene** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia
NVIDIA Corporation
United States
Nov'22

DGX A100 node

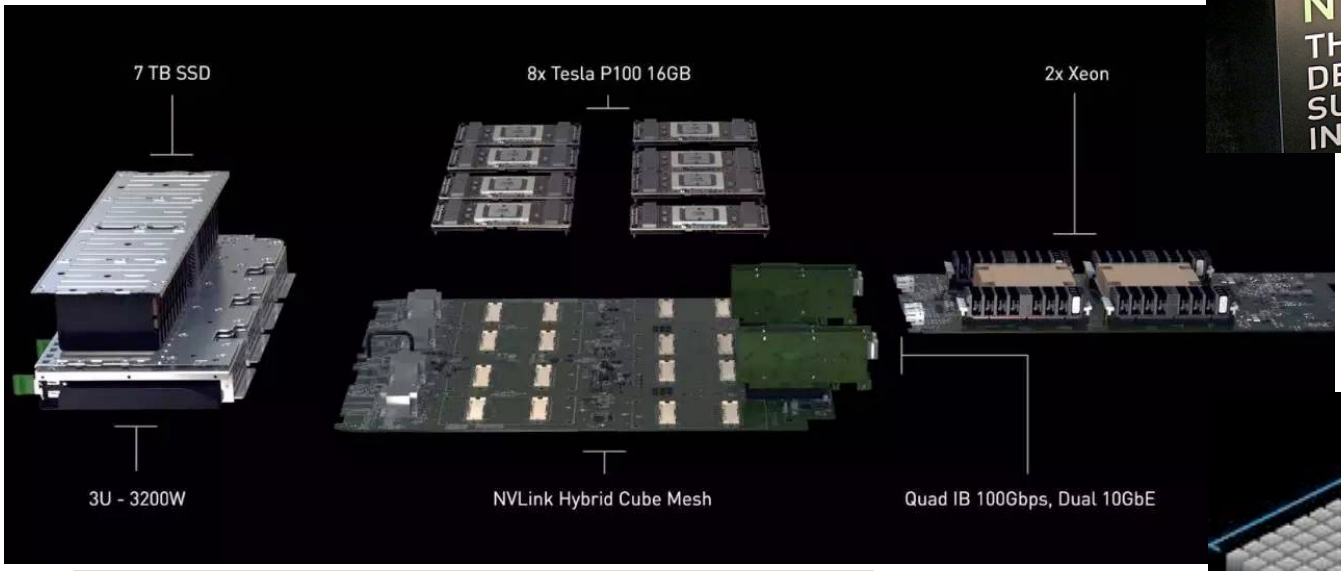
NVidia DGX A100 node



NVidia DGX-1 SaturnV: before Selene

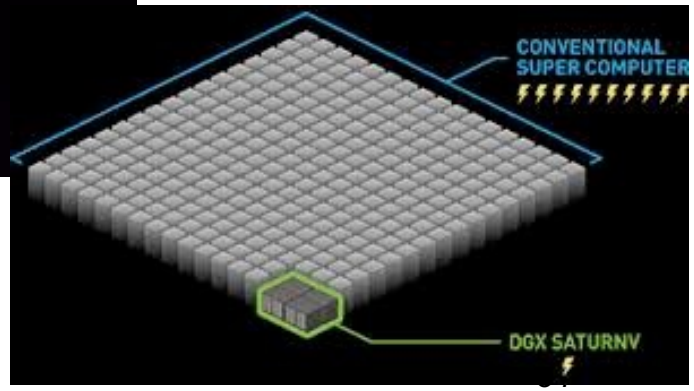


$149,000



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



Overview of Tianhe-2A

(Tianhe-2 #1 in June'13 TOP500)

10

TOP 500 The List.

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

Nov'22

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/

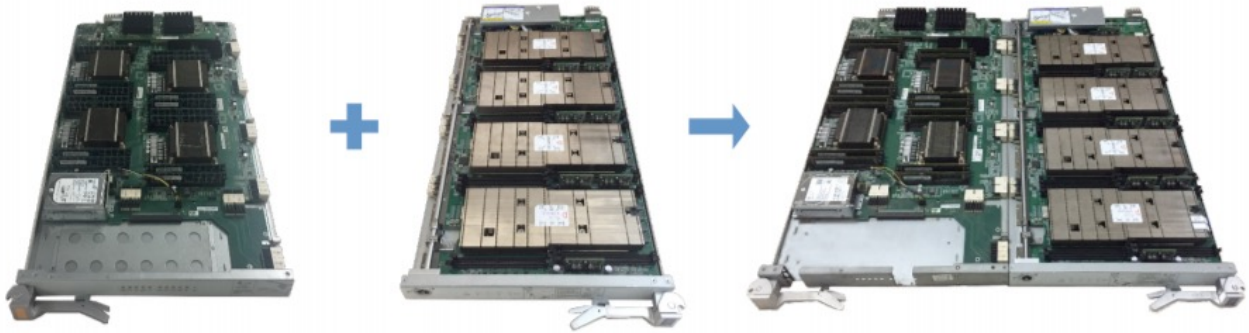
Overview of Tianhe-2A

Compute nodes

Heterogeneous Compute Blades

- Compute blade = Xeon part + Matrix-2000 part

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

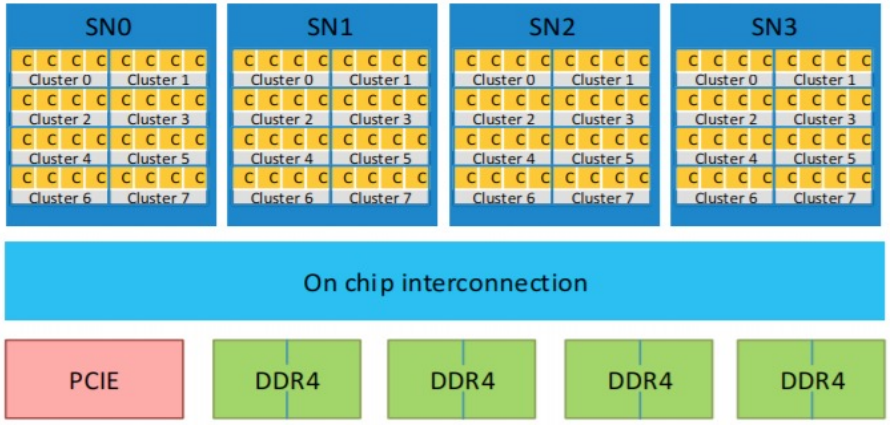


- Use the Matrix-2000 part to replace the KNC part

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/

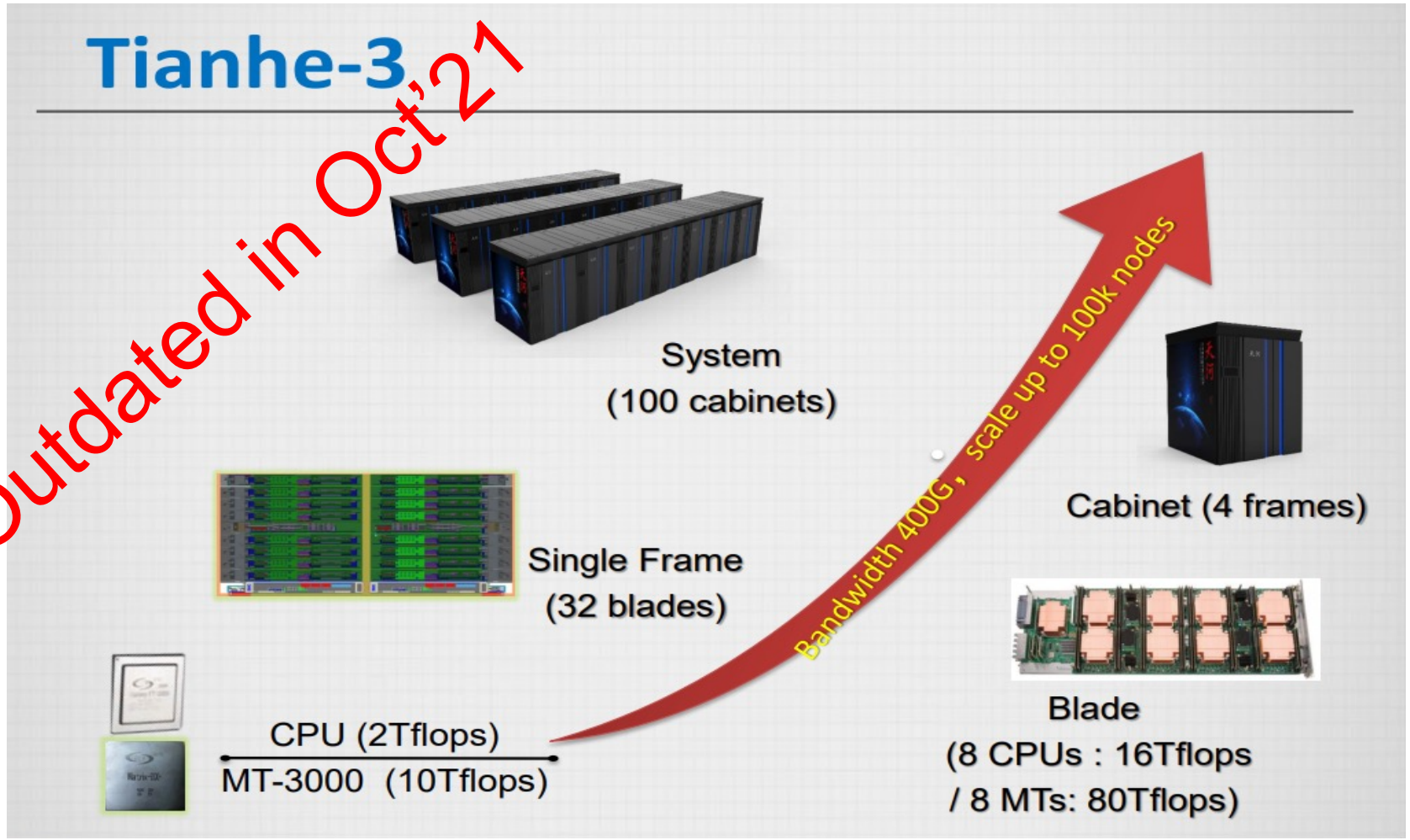
Phytium-2000+

Next: Tianhe-3

with ~~Fujitsu A64FX~~ ARM~~-SVE~~ + Matrix-~~3000~~ accelerators

2000+

https://www.hpcwire.com/2021/11/24/three-chinese-exascale-systems-detailed-at-sc21-two-operational-and-one-delayed/



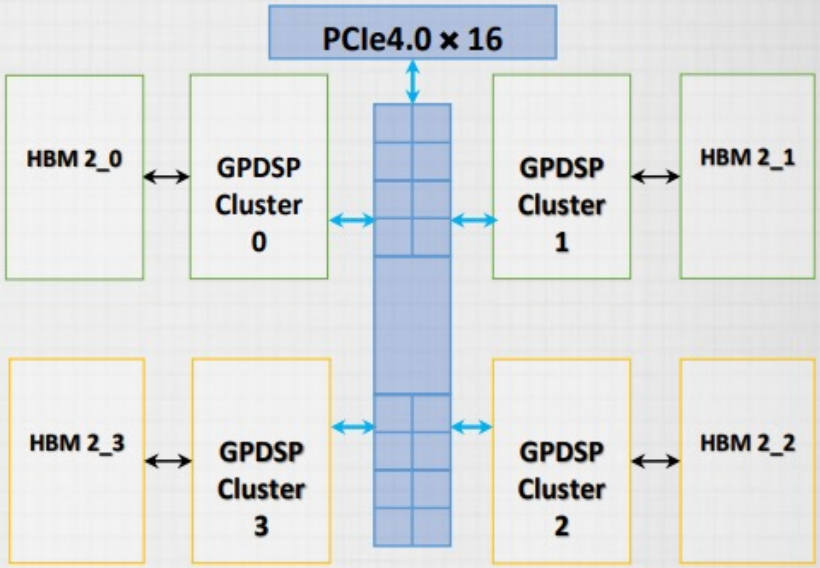
https://www.nextplatform.com/2019/05/02/china-fleshes-out-exascale-design-for-tianhe-3/

NUDT Matrix-3000 accelerator

National University of Defense Technology (NUDT)

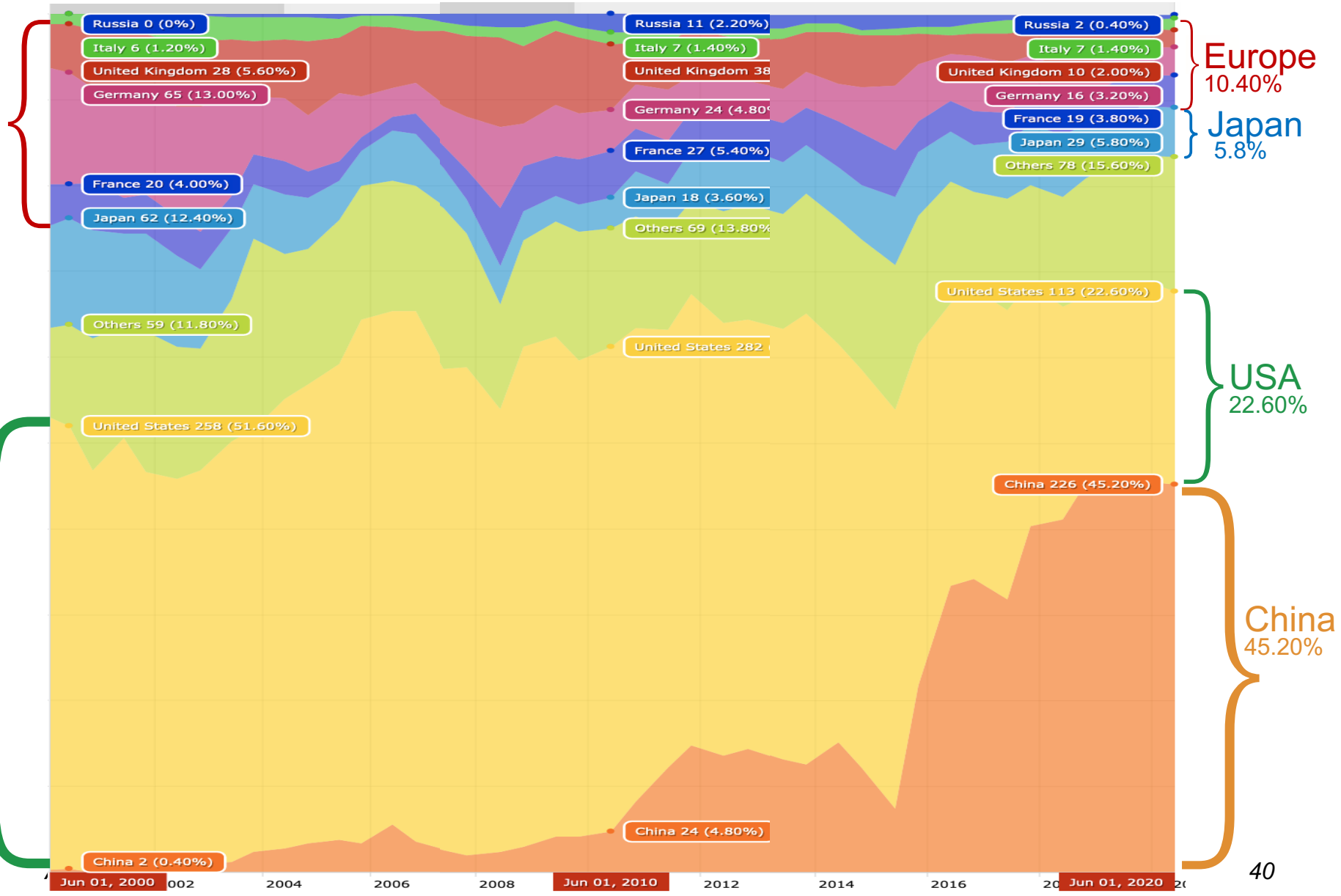
Matrix-3000

- GPDSP
- Cores>=96, > 10 Tflops
- HBM2
- PCIe Gen4
- Support half precision

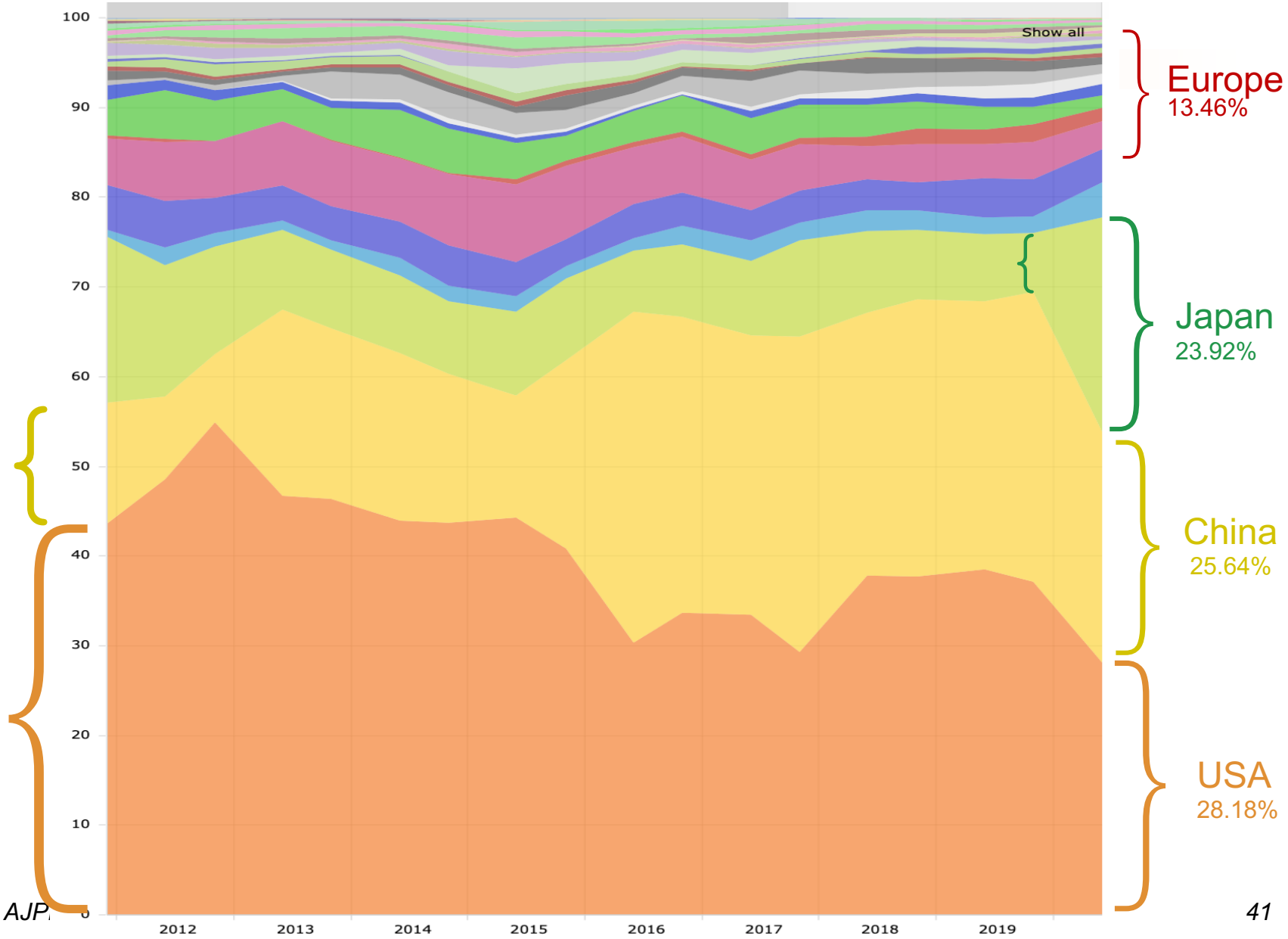


https://www.nextplatform.com/2019/05/02/china-fleshes-out-exascale-design-for-tianhe-3/

Country distribution over 20 years:
from 2000 to 2020, #systems

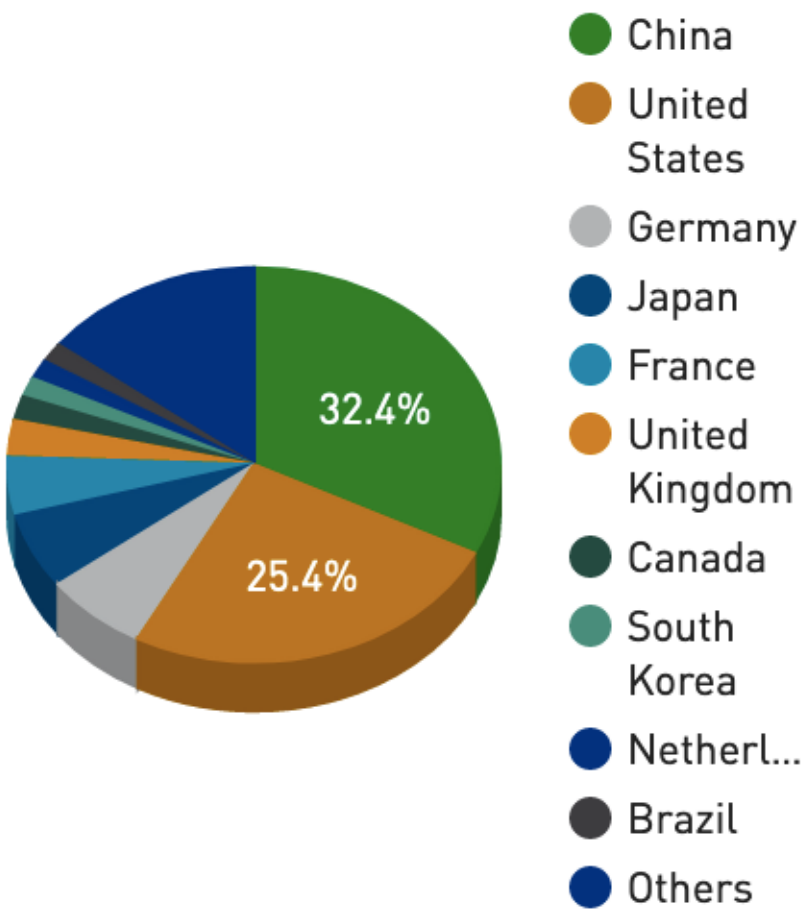


Country distribution over 20 years:
from 2000 to 2020, aggregate performance

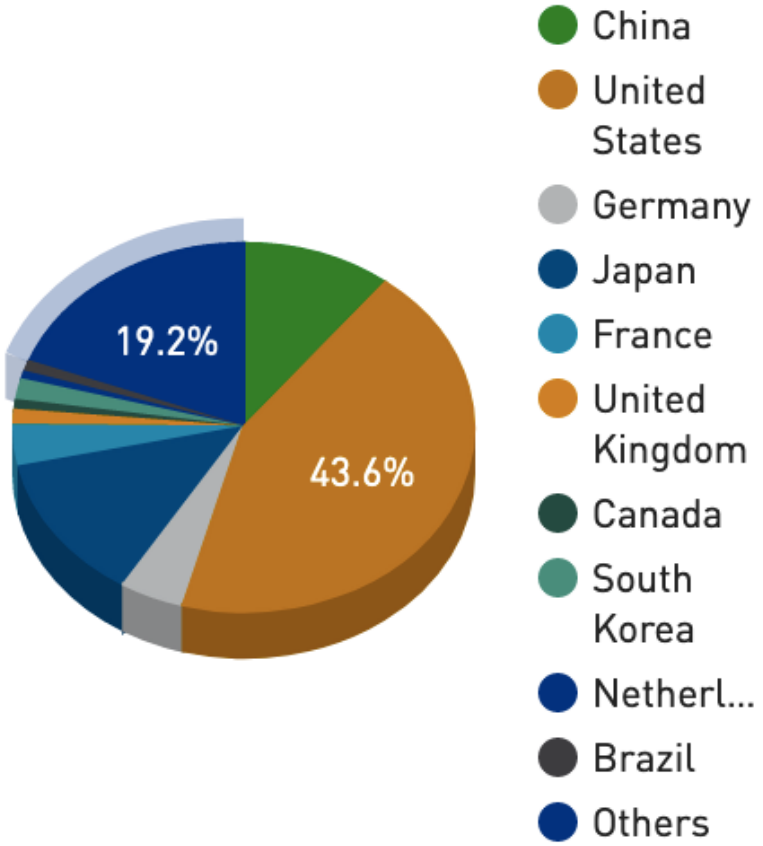


Country distribution in Nov'22

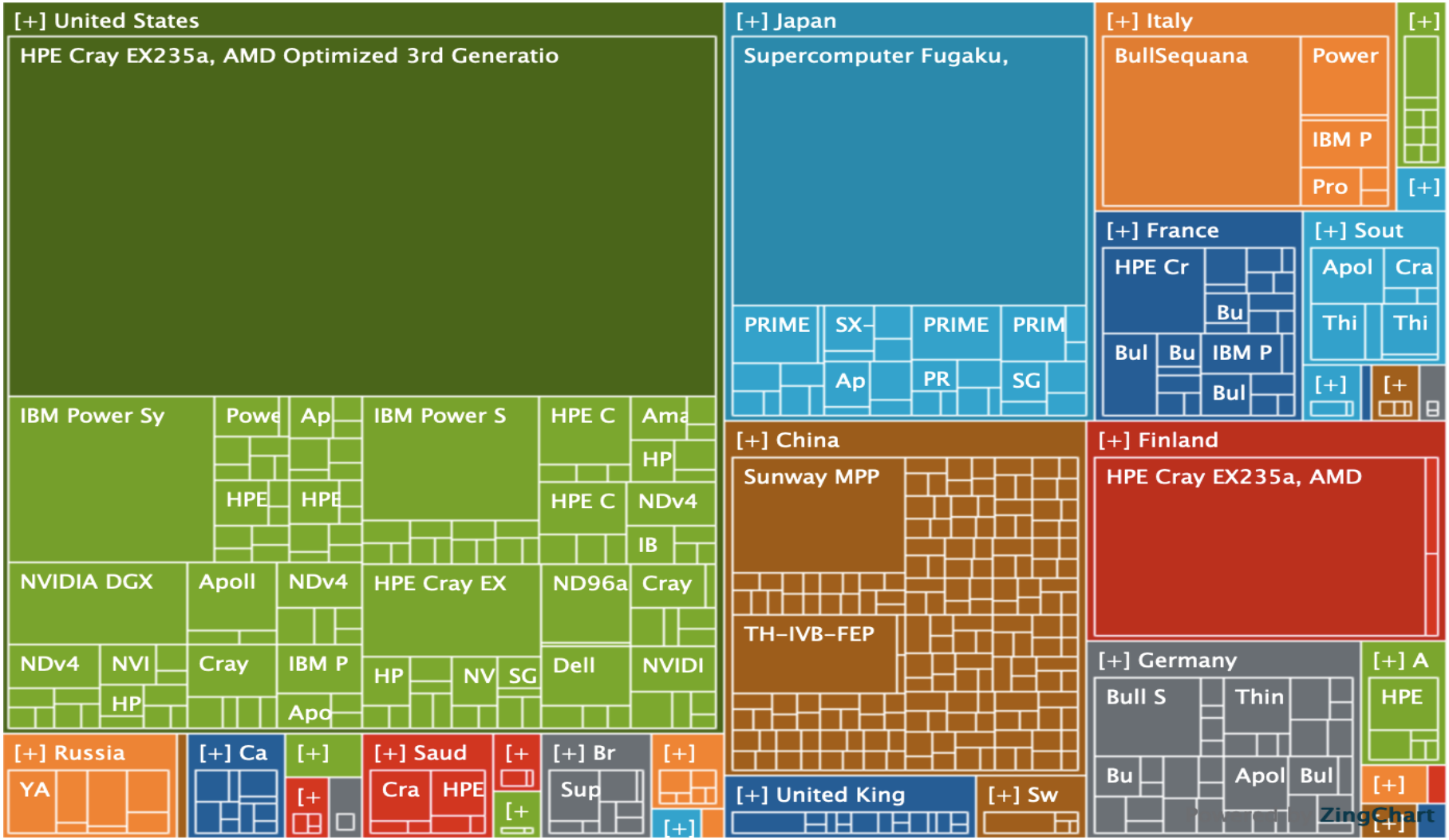
Countries System Share



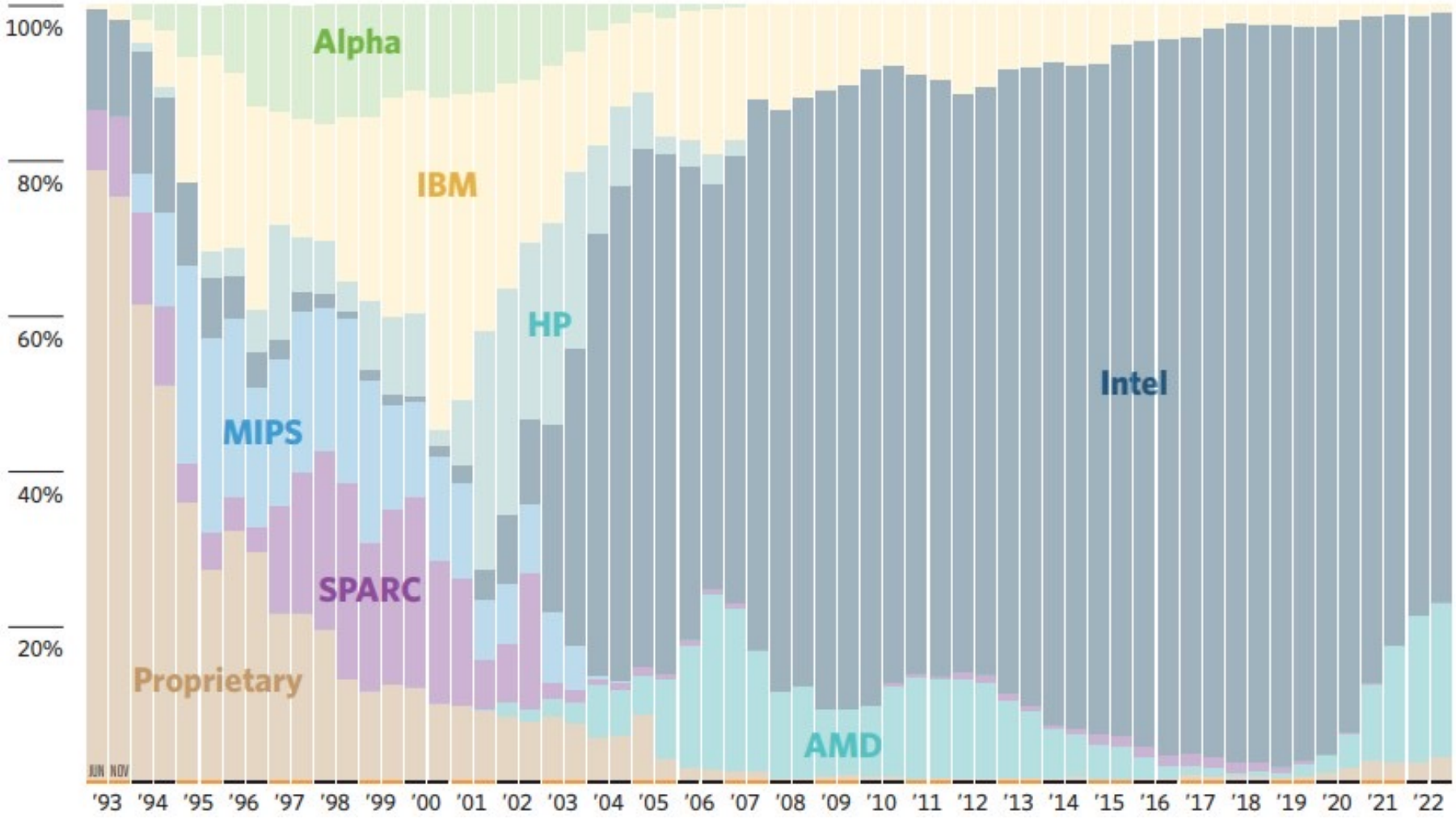
Countries Performance Share



Country distribution in Nov'22:
#systems & performance

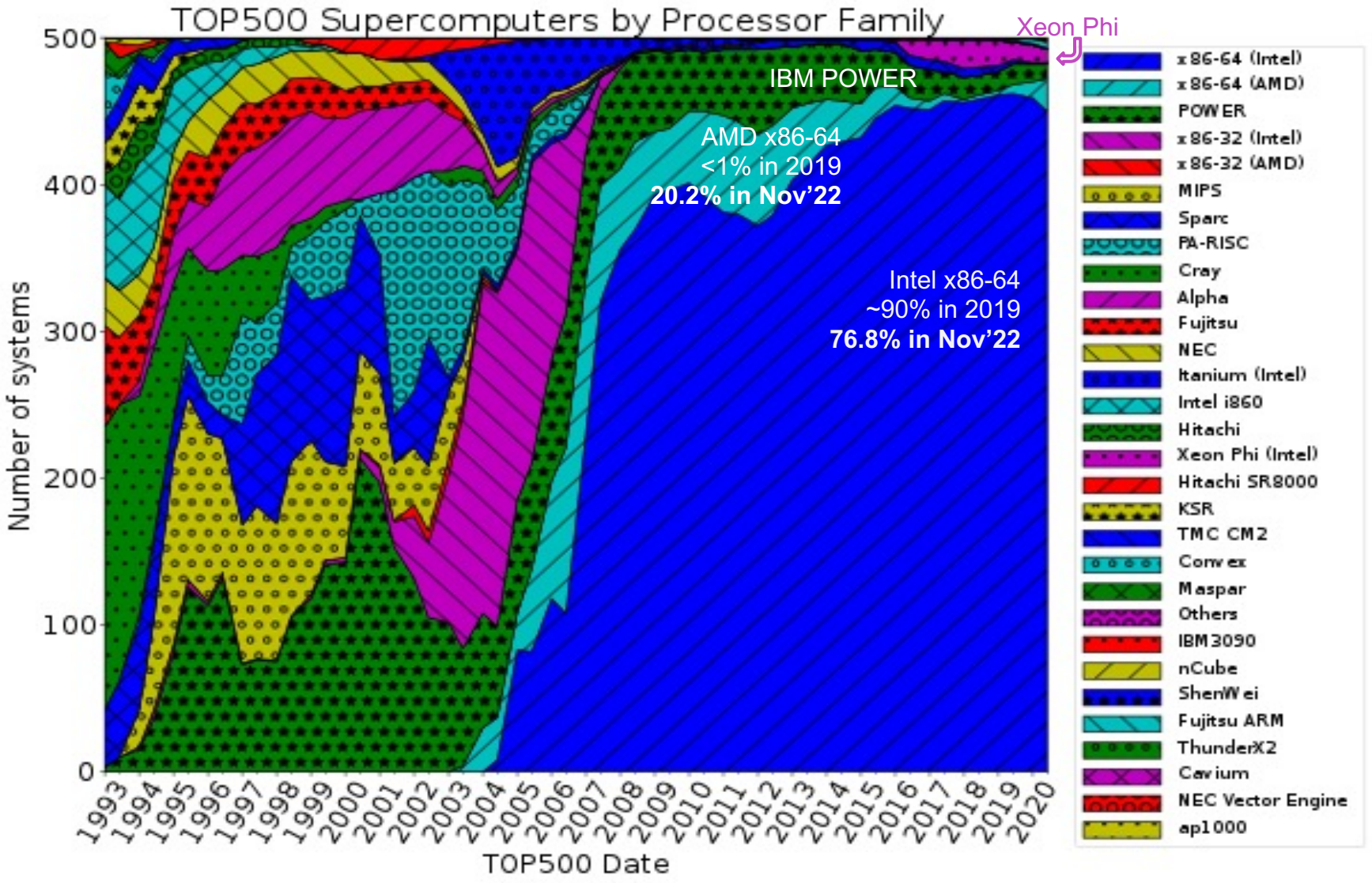


PU chip technology
1993 to 2022



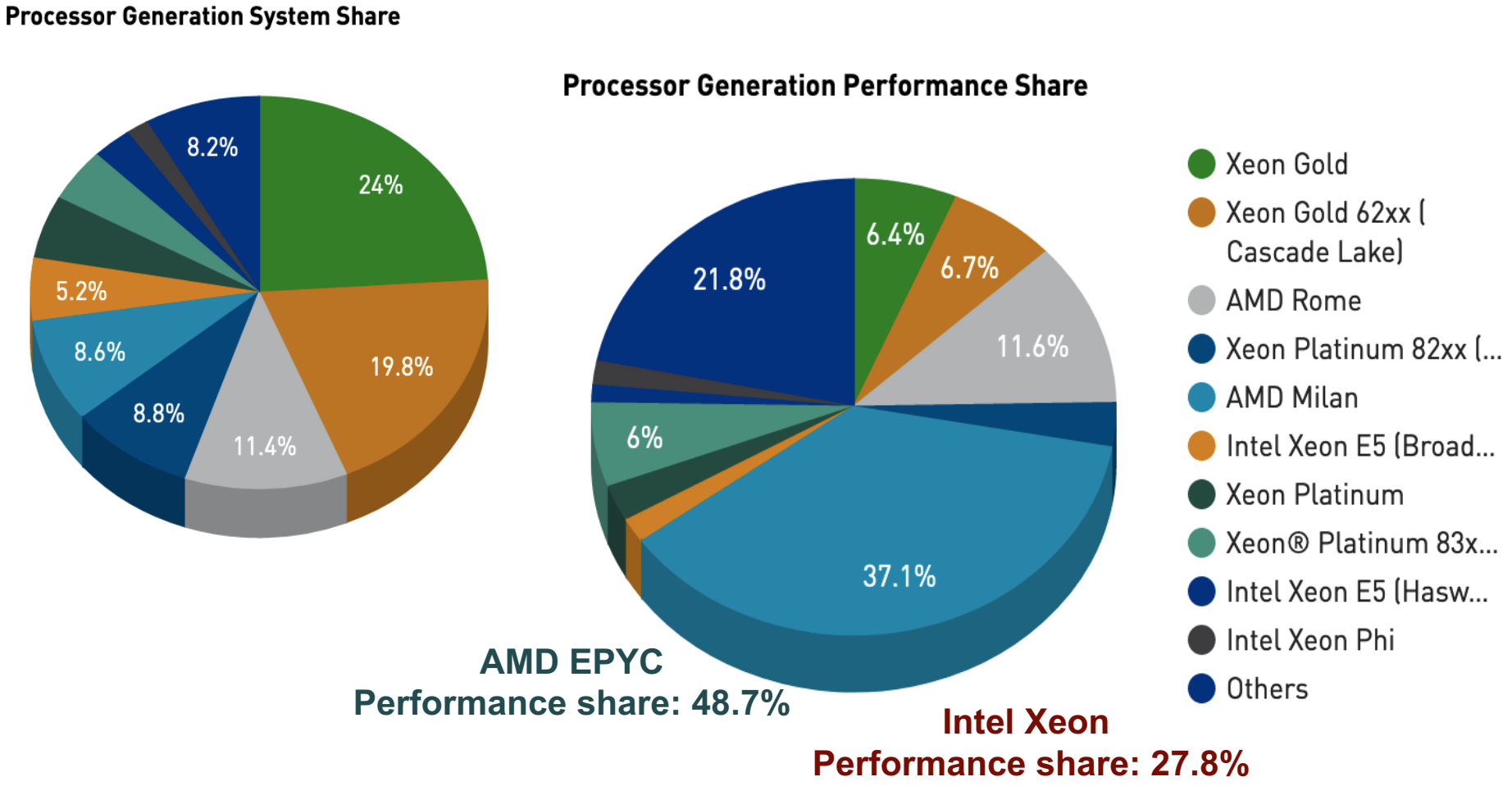
https://www.nextplatform.com/2022/11/14/testing-the-mettle-of-the-top-supercomputing-iron-on-earth/

Top processor families
1993 to 2020



https://en.wikipedia.org/wiki/TOP500

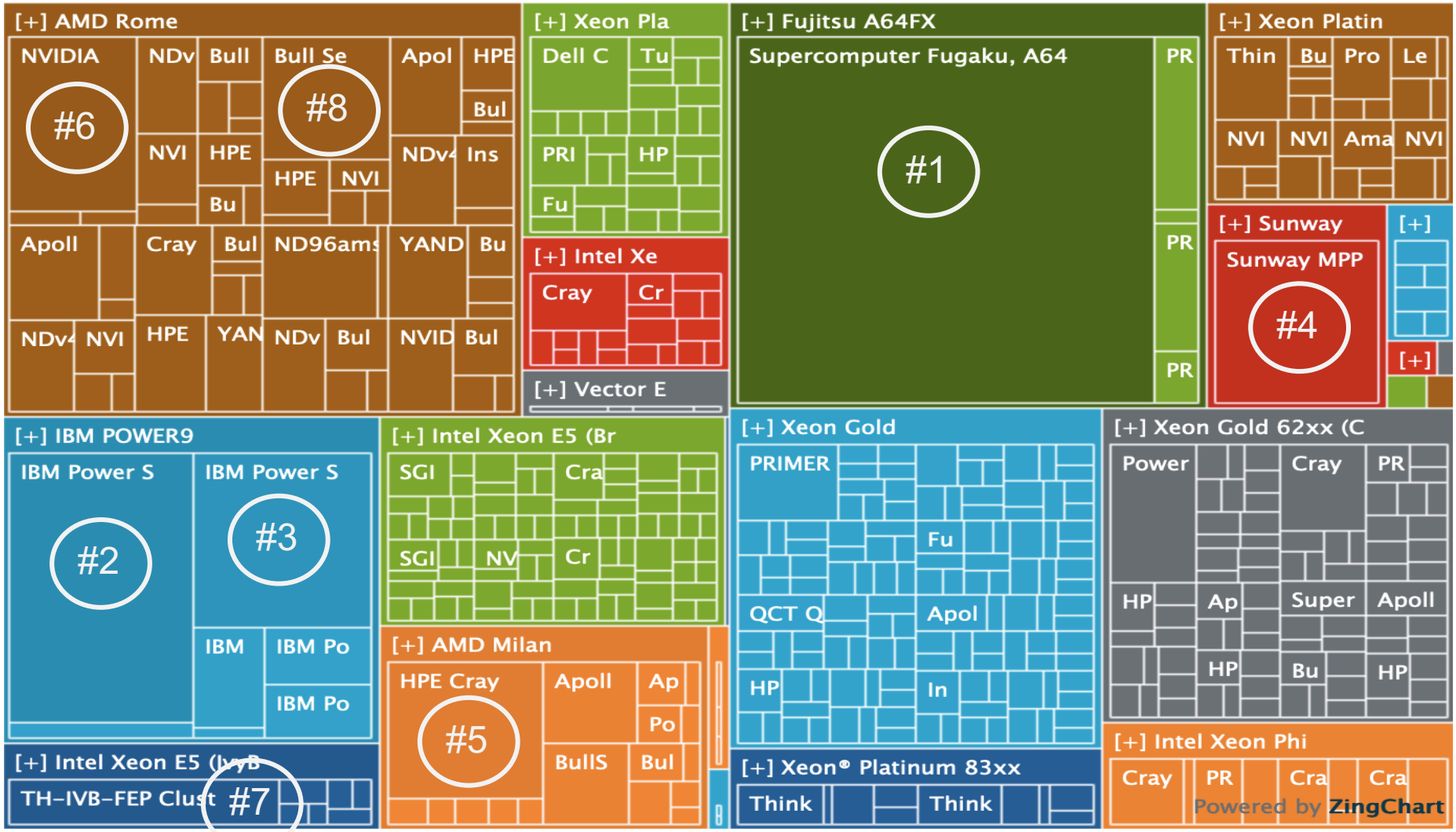
Processor generation share
Nov'22



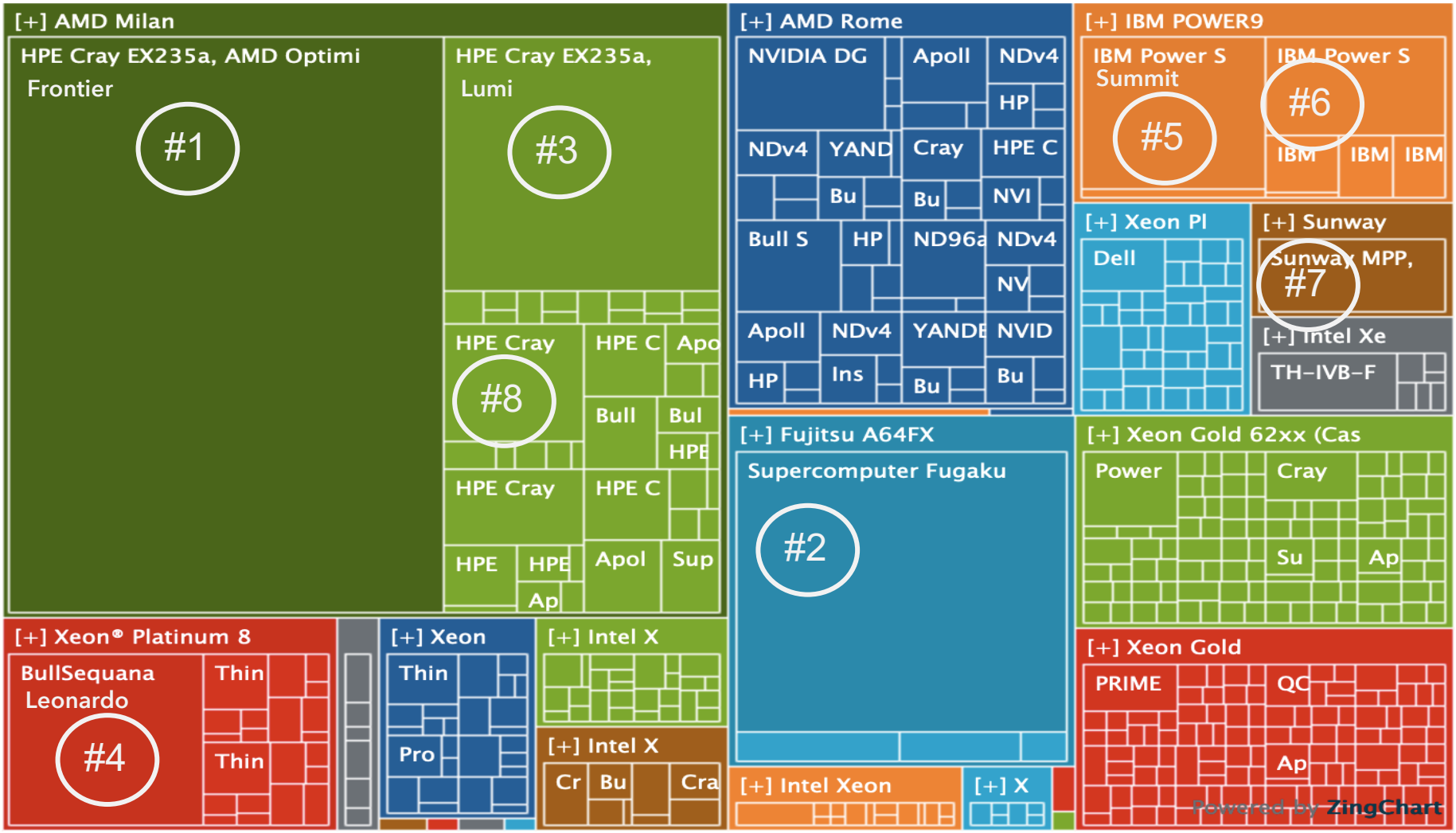
AMD EPYC
Performance share: 48.7%

Intel Xeon
Performance share: 27.8%

Processor distribution
Nov'21

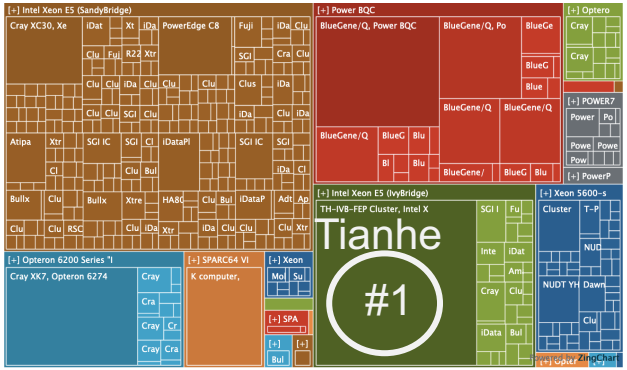


Processor distribution
Nov'22



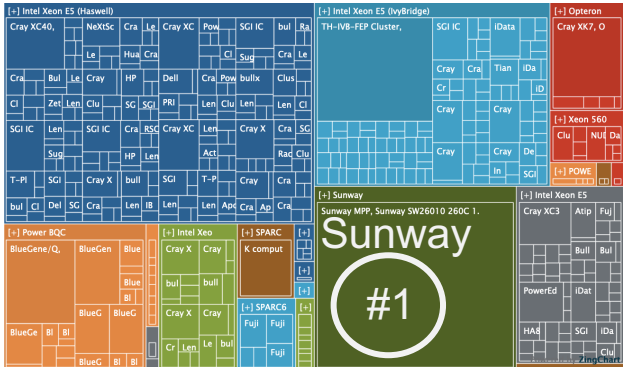
Architecture family of key #1's:
Nov'13 to Nov'22

Nov'13



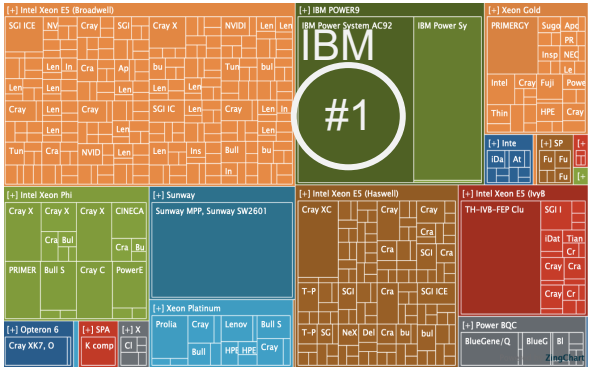
with Xeon Phi

Jun'16



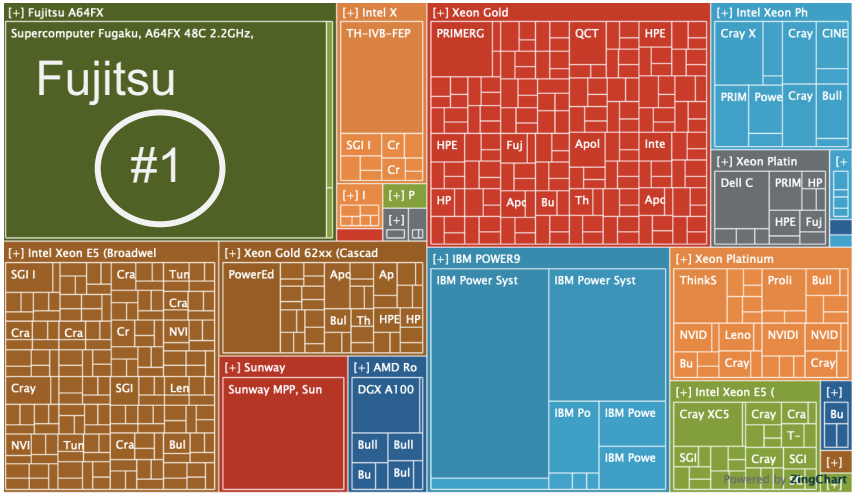
without any accelerator

Jun'18



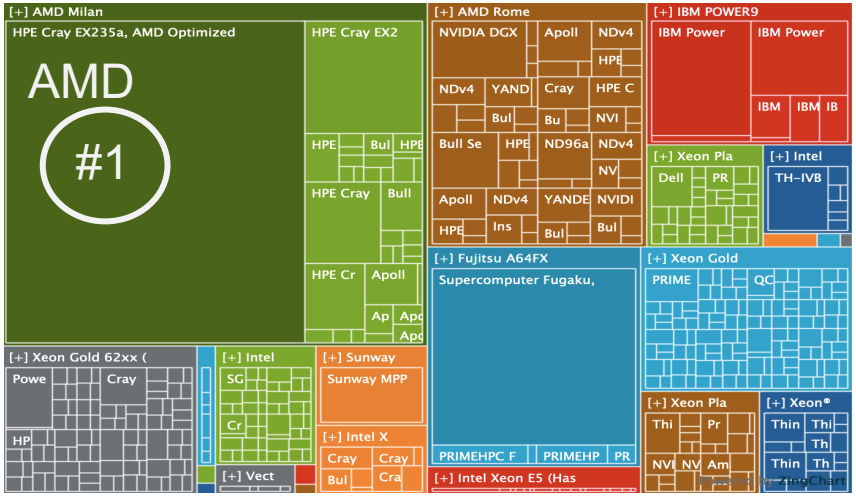
with Nvidia Volta

Jun'20



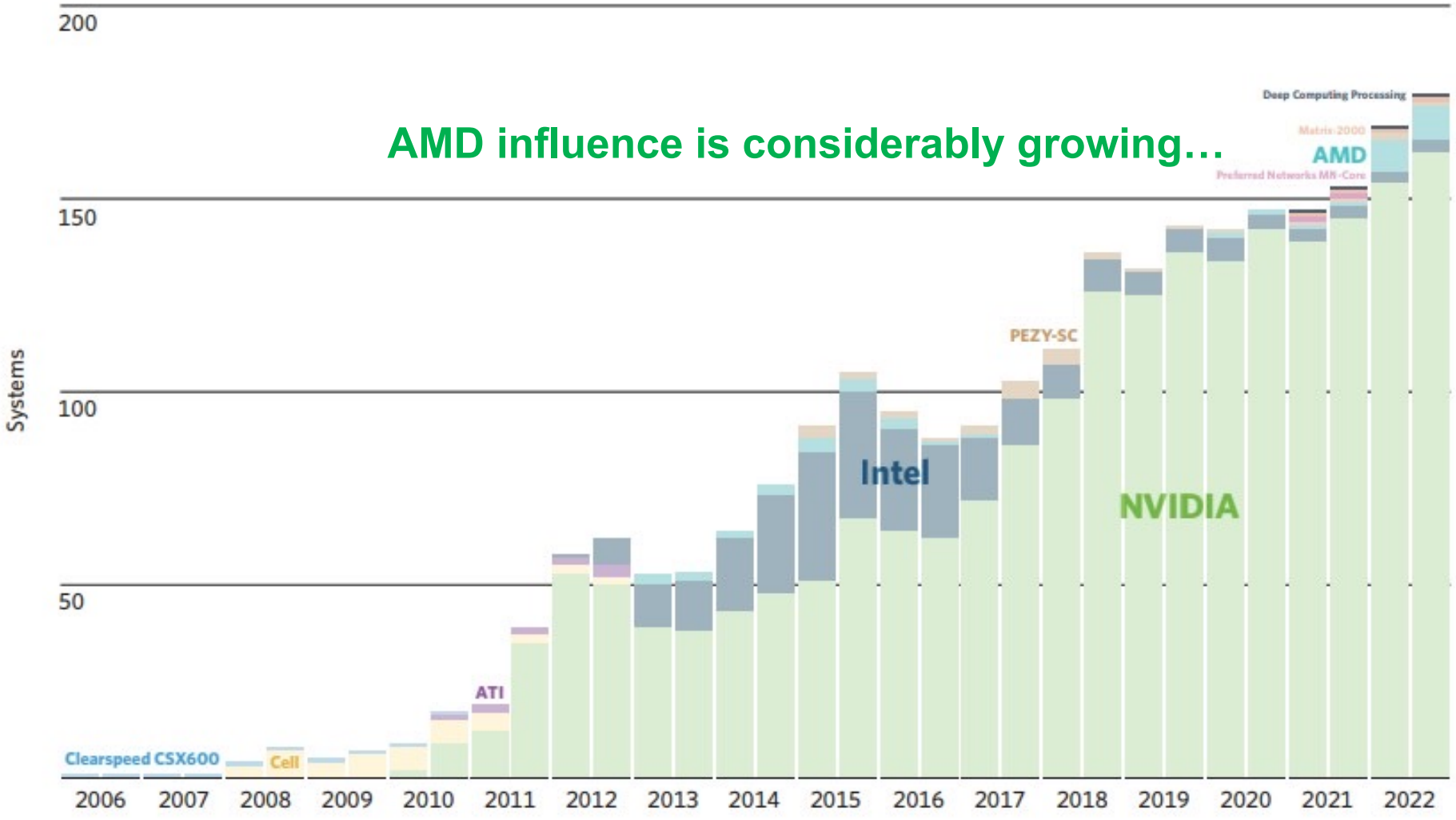
without any accelerator

Jun'22



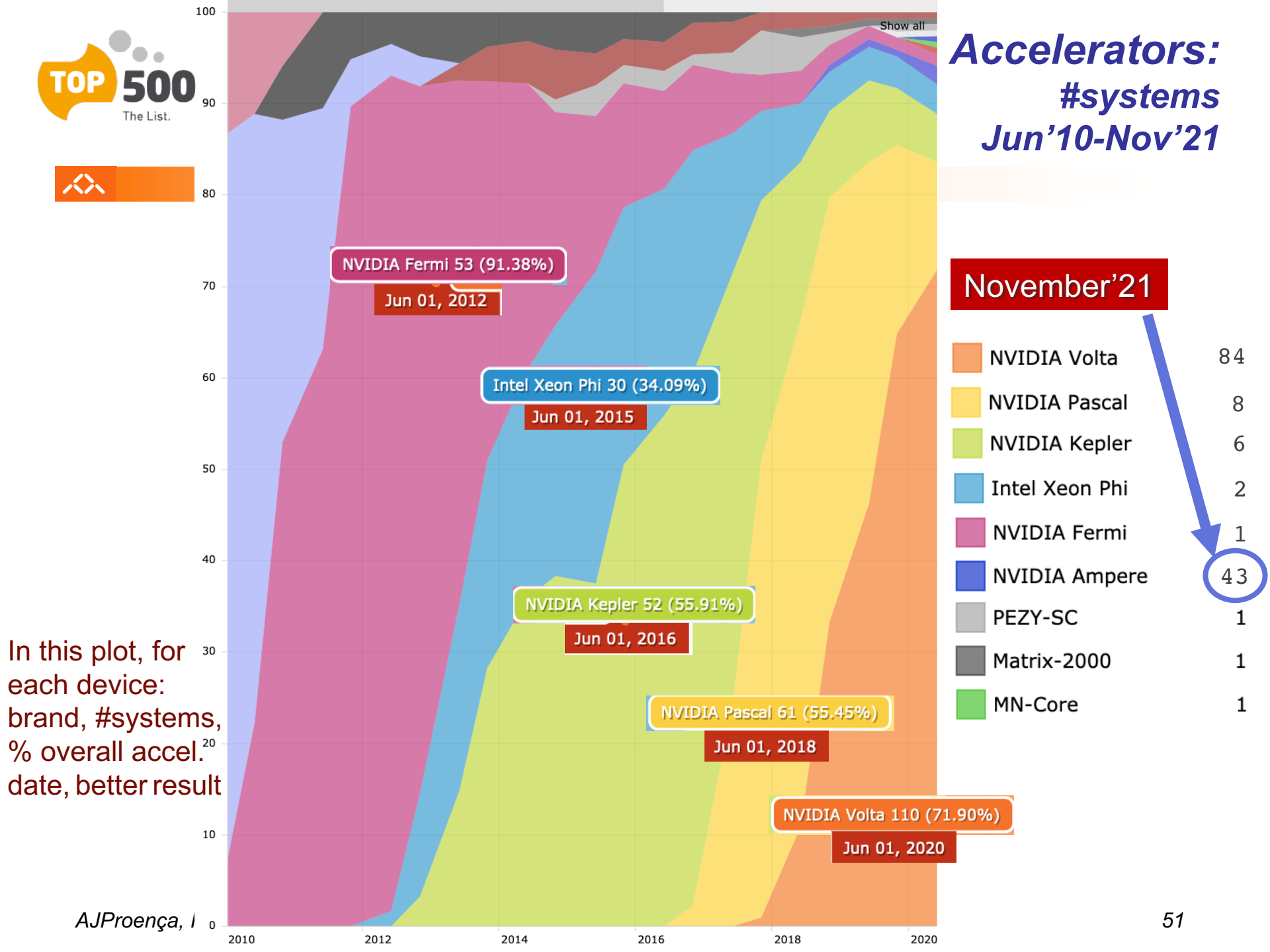
with AMD MI250x

Accelerator families
2006 to 2022



https://www.nextplatform.com/2022/11/14/testing-the-mettle-of-the-top-supercomputing-iron-on-earth/

Accelerators: #systems Jun'10-Nov'21



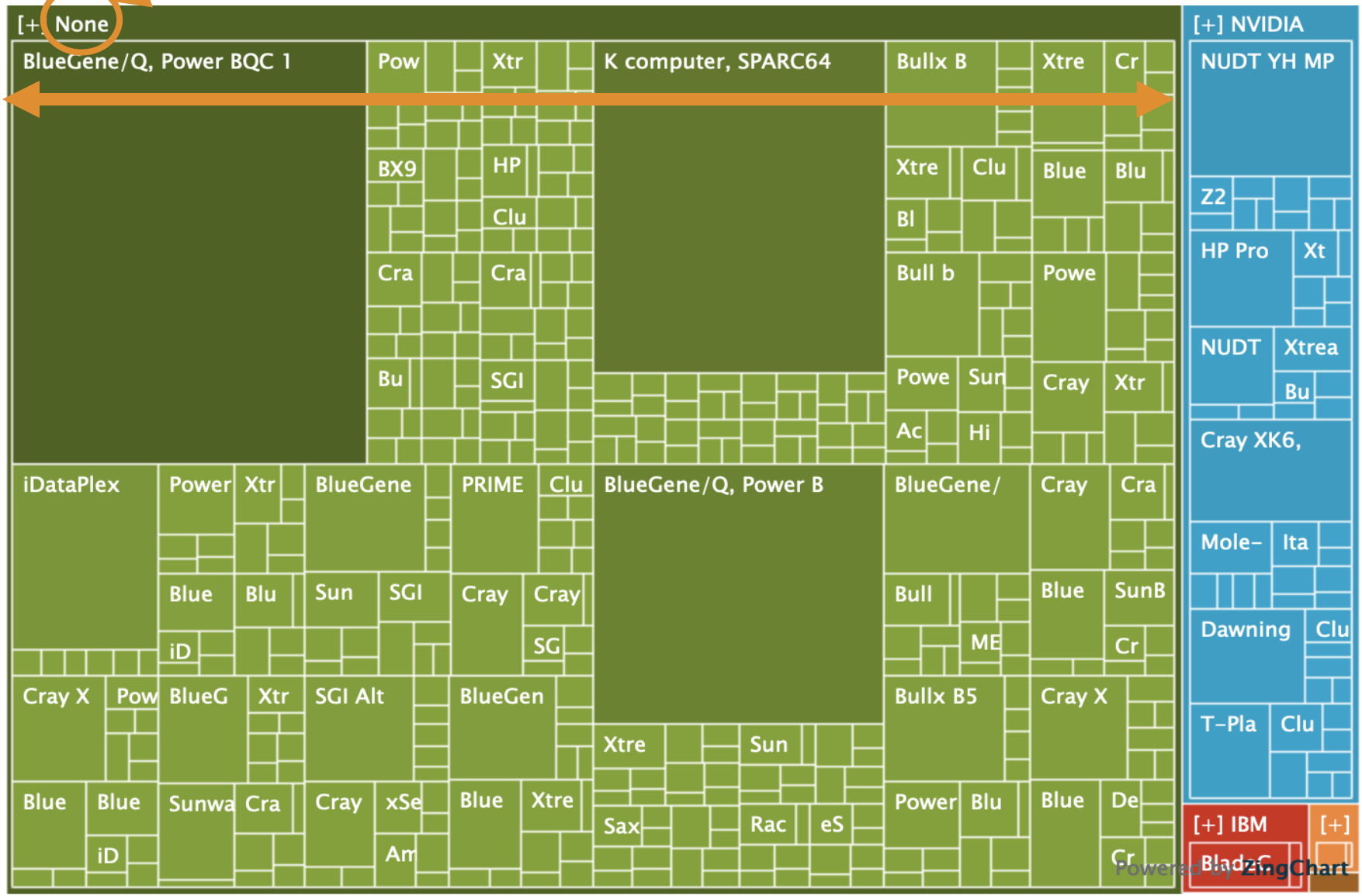
November'21

NVIDIA Volta	84
NVIDIA Pascal	8
NVIDIA Kepler	6
Intel Xeon Phi	2
NVIDIA Fermi	1
NVIDIA Ampere	43
PEZY-SC	1
Matrix-2000	1
MN-Core	1

In this plot, for each device: brand, #systems, % overall accel. date, better result

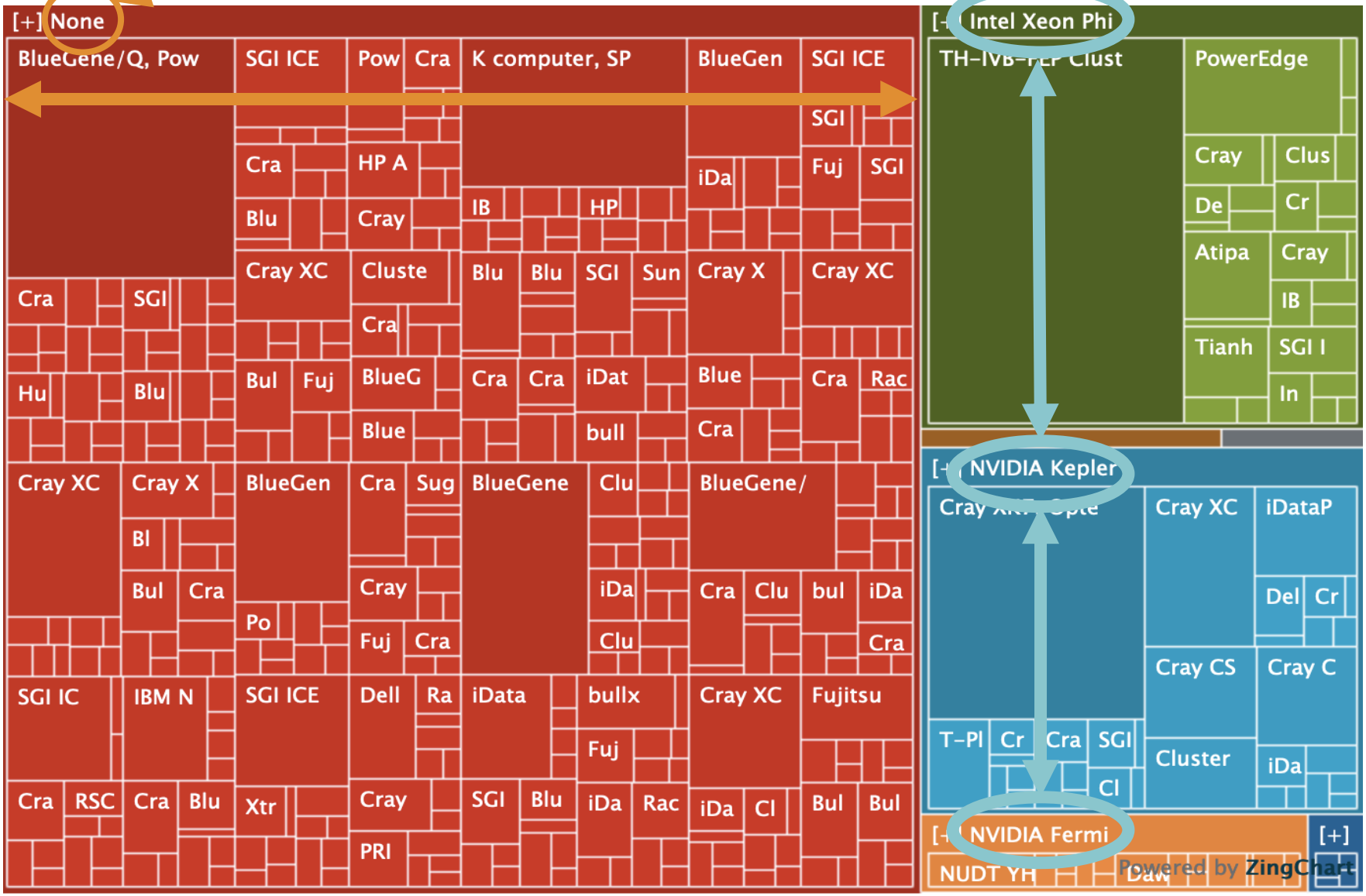
Accelerator family distribution

Jun'12

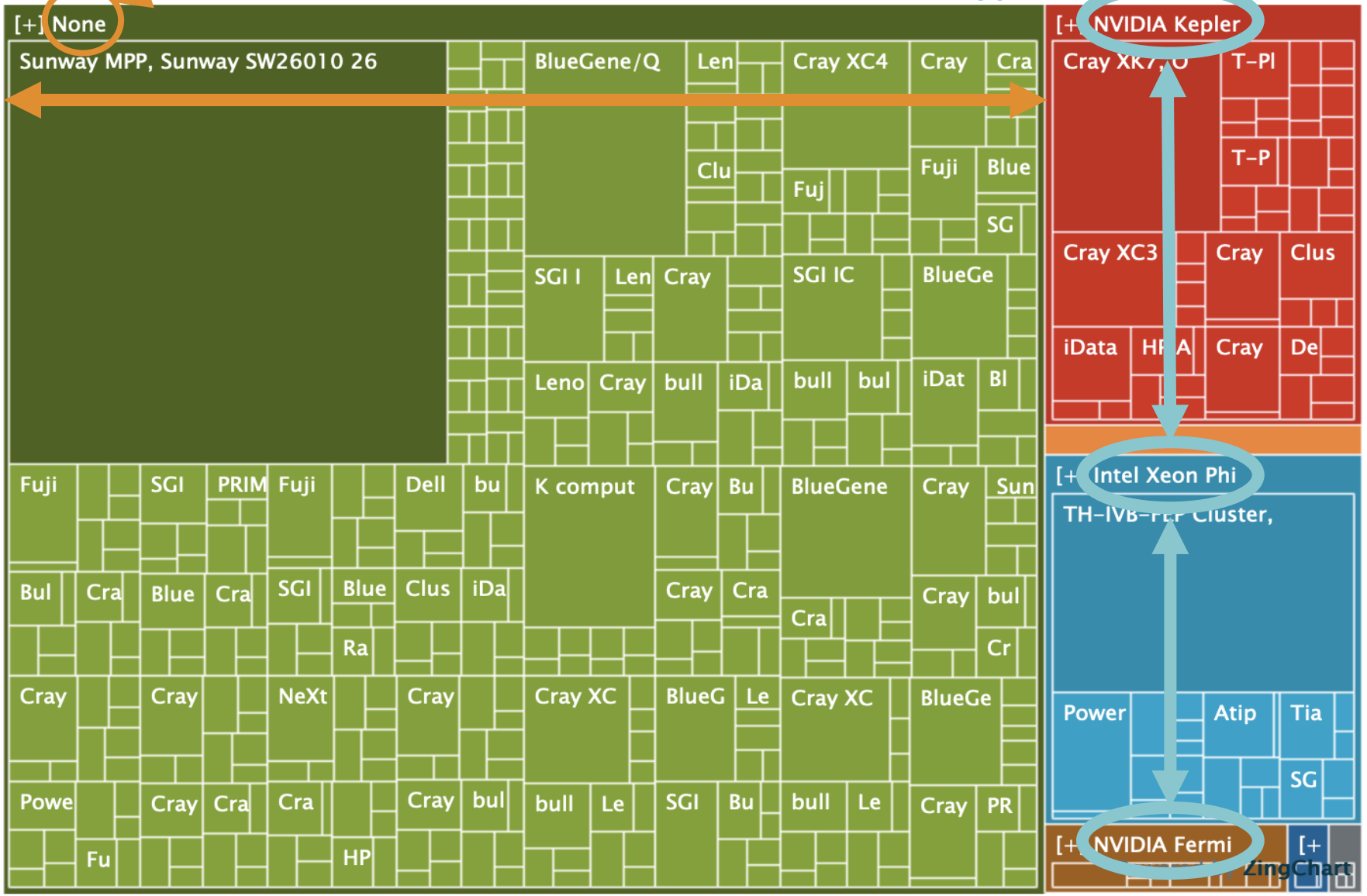


Accelerator family distribution

Jun'15

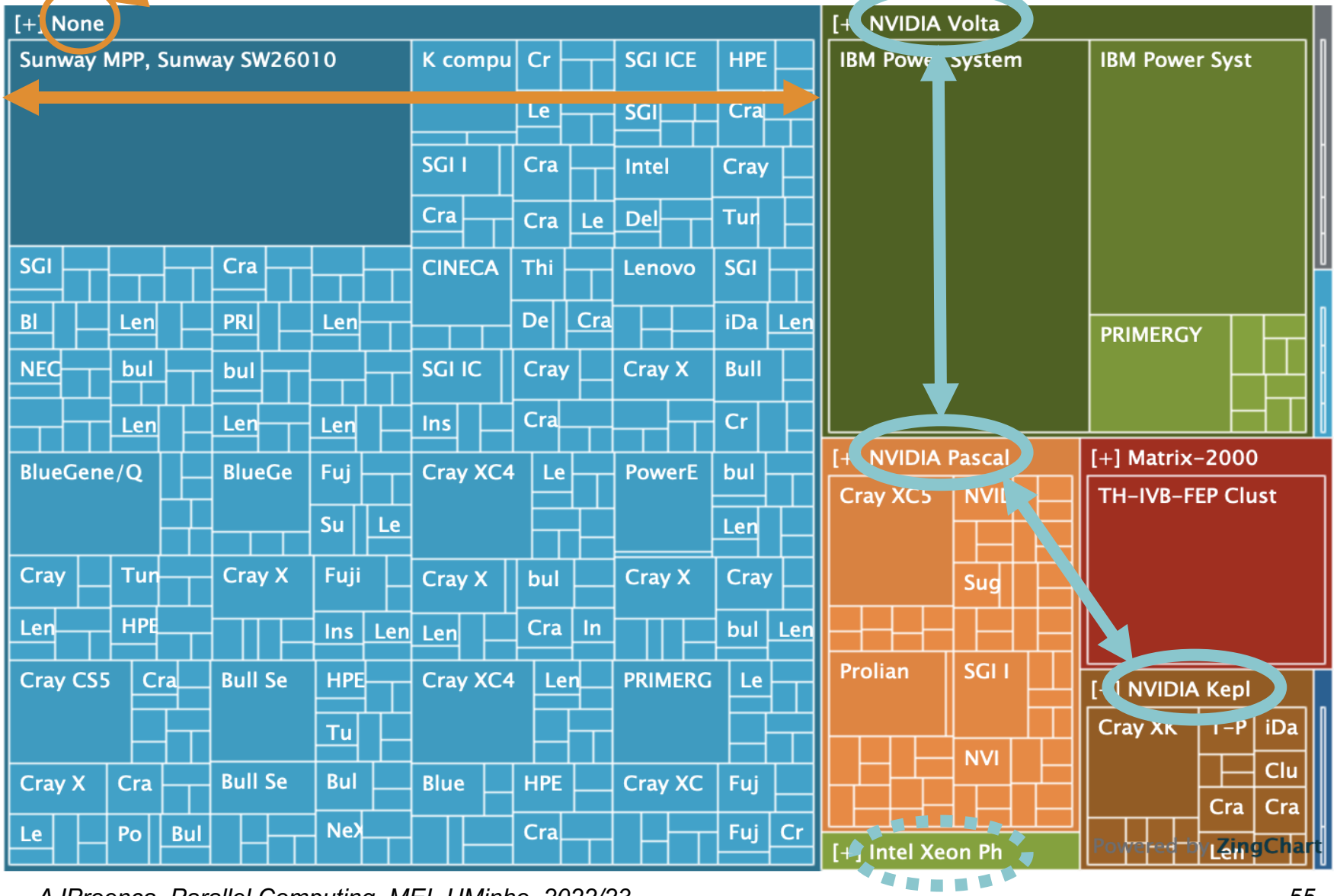


Accelerator family distribution Jun'16



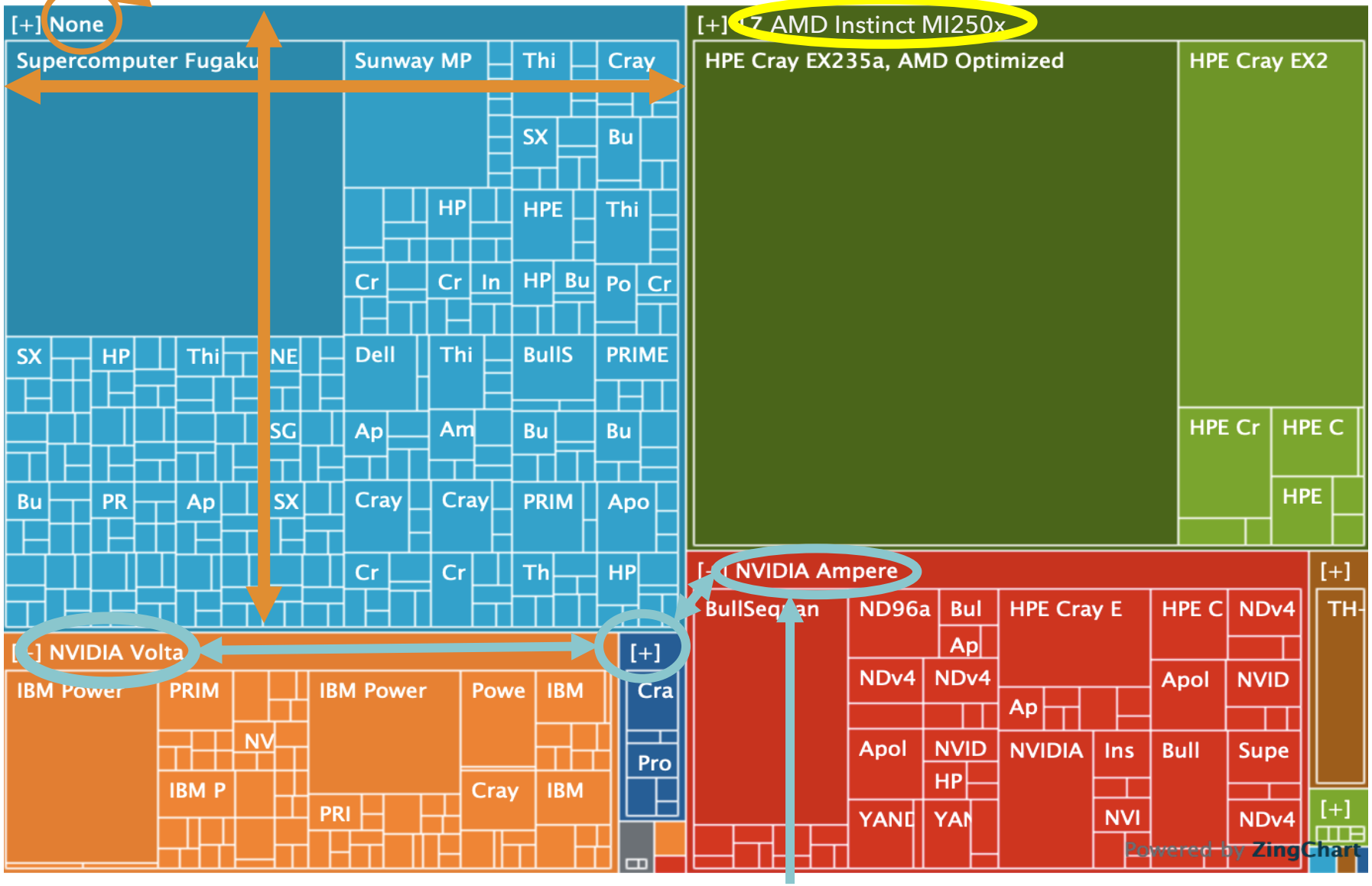
Accelerator family distribution
Jun'18

NVIDIA

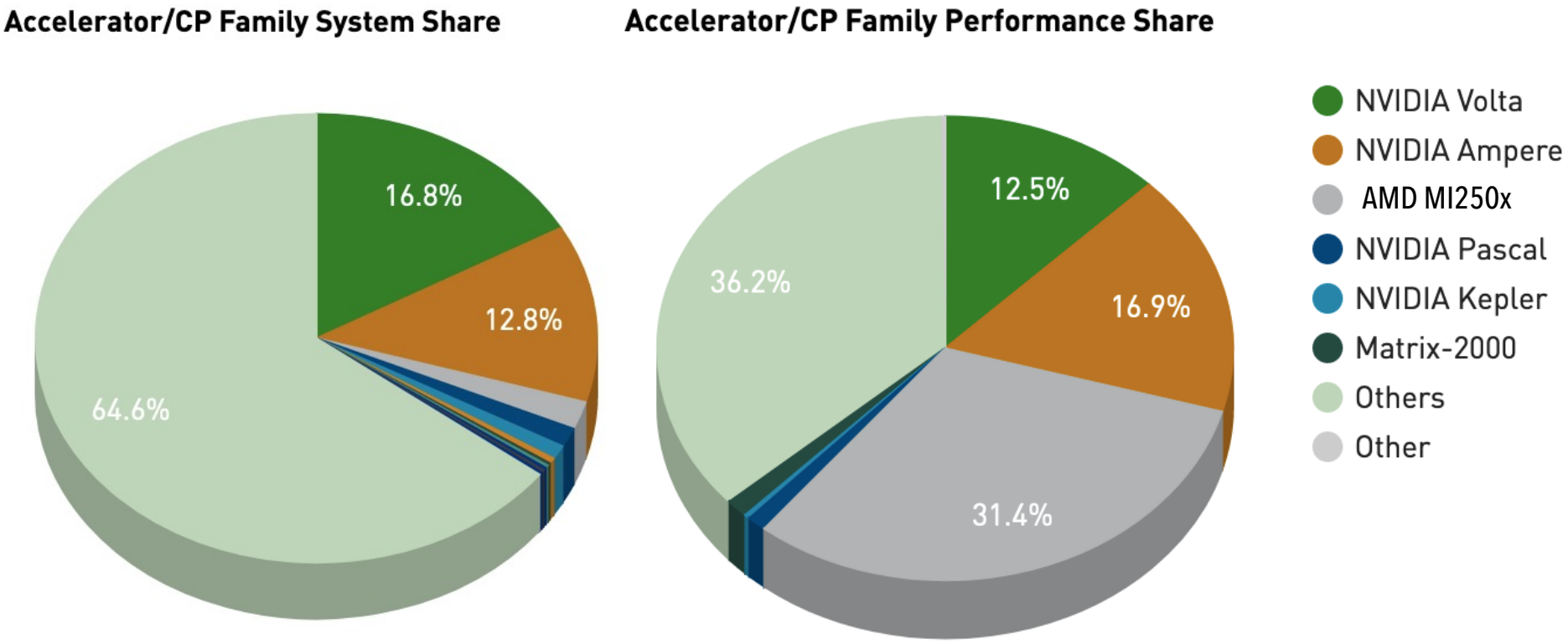


Accelerator family distribution Nov'21

Accelerator family distribution Nov'22



Accelerator family distribution
Nov'22



Outline

1. TOP500
   a) TOP10 lists from Nov'17 to Nov'22
   b) Analysis of some relevant systems and architectures
   c) PU chip technology evolution in the past 25 years and since last year
   d) Evolution of the accelerators since they were available
   e) Country distribution over the past 25 years

2. GREEN500
   a) TOP5 list from Nov'19 to Nov'22
   b) Analysis of some relevant systems and architectures

3. HPCG
   a) HPCG *vs.* HPL: an overview
   b) TOP10 in Nov'22

4. HPL-AI
   a) High-performance Linpack (HPL) and artificial intelligence (AI) workloads

The Green500 list

The list ranks computers in terms of energy efficiency, typically measured as LINPACK FLOPS per watt.

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

The GREEN 500

Top Green500 systems

Nov'19

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

~Ghyoukou

Top Green500 systems
Nov'20

Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	Power (kW)	Power Efficiency (GFlops/watts)
1	170	**NVIDIA DGX SuperPOD** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia NVIDIA Corporation United States	19,840	2,356.0	90	26.195
2	330	**MN-3** - MN-Core Server, Xeon Platinum 8260M 24C 2.4GHz, Preferred Networks MN-Core, MN-Core DirectConnect, Preferred Networks Preferred Networks Japan	1,664	1,652.9	65	26.039
3	7	**JUWELS Booster Module** - Bull Sequana XH2000 , AMD EPYC 7402 24C 2.8GHz, NVIDIA A100, Mellanox HDR InfiniBand/ParTec ParaStation ClusterSuite, Atos Forschungszentrum Juelich (FZJ) Germany	449,280	44,120.0	1,764	25.008
4	146	**Spartan2** - Bull Sequana XH2000 , AMD EPYC 7402 24C 2.8GHz, NVIDIA A100, Mellanox HDR Infiniband, Atos Atos France	23,040	2,566.0	106	24.262
5	5	**Selene** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia NVIDIA Corporation United States	555,520	63,460.0	2,646	23.983
6	239	**A64FX prototype** - Fujitsu A64FX, Fujitsu A64FX 48C 2GHz, Tofu interconnect D, Fujitsu	36,864	1,999.5	118	16.876

~Selene

#1 in Jun'20

~Fugaku

Top Green500 systems
Nov'21

Rank	TOP500 Rank	System	Cores	Rmax (TFlop/s)	Power (kW)	Power Efficiency (GFlops/watts)
1	301	**MN-3** - MN-Core Server, Xeon Platinum 8260M 24C 2.4GHz, Preferred Networks MN-Core, MN-Core DirectConnect, Preferred Networks Preferred Networks Japan	1,664	2,181.2	55	39.379
2	291	**SSC-21 Scalable Module** - Apollo 6500 Gen10 plus, AMD EPYC 7543 32C 2.8GHz, NVIDIA A100 80GB, Infiniband HDR200, HPE Samsung Electronics South Korea	16,704	2,274.1	103	33.983
3	295	**Tethys** - NVIDIA DGX A100 Liquid Cooled Prototype, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100 80GB, Infiniband HDR, Nvidia NVIDIA Corporation United States	19,840	2,255.0	72	31.538
4	280	**Wilkes-3** - PowerEdge XE8545, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 80GB, Infiniband HDR200 dual rail, DELL EMC University of Cambridge United Kingdom	26,880	2,287.0	74	30.797
5	30	**HiPerGator AI** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Infiniband HDR, Nvidia University of Florida United States	138,880	17,200.0	583	29.521
6	403	**Snellius Phase 1 GPU** - ThinkSystem SD650-N V2, Xeon Platinum 8360Y 36C 2.4GHz, NVIDIA A100 SXM4 40 GB, Infiniband HDR, Lenovo SURF Netherlands	6,480	1,818.0	63	29.046

#1 in Jun'20

All based on EPYC + A100

Xeon + A100

Additional comments:

- from #7 to #11, #13 to# 21, … all based on EPYC/Xeon + A100
- #12: NA-IT1, follow-up of NA-1 with PEZY-SC3

12	435	**NA-IT1** - ZettaScaler3.0, AMD EPYC 7702P 64C 1.5GHz, PEZY-SC3, Infiniband EDR, PEZY Computing / Exascaler Inc. NA Simulation Japan

Top Green500 systems
Nov'22

Rank	TOP500 Rank	System	Cores	Rmax (PFlop/s)	Power (kW)	Energy Efficiency (GFlops/watts)
1	405	**Henri** - Lenovo ThinkSystem SR670 V2, Intel Xeon Platinum 8362 2800Mhz (32C), NVIDIA H100 80GB PCIe, Infiniband HDR, Lenovo Flatiron Institute United States	5,920	2.04	31	65.091
2	32	**Frontier TDS** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	120,832	19.20	309	62.684
3	11	**Adastra** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Grand Equipement National de Calcul Intensif - Centre Informatique National de l'Enseignement Suprieur (GENCI-CINES) France	319,072	46.10	921	58.0[illegible]1
4	15	**Setonix – GPU** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Pawsey Supercomputing Centre, Kensington, Western Australia Australia	181,248	27.16	477	56.9[illegible]3
5	68	**Dardel GPU** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE KTH - Royal Institute of Technology Sweden	52,864	8.26	146	56.491
6	1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,730,112	1,102.00	21,100	52.227
7	3	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, [illegible] HPE EuroHPC/CSC Finland	2,220,288	309.10	6,016	51.382

3rd Gen Xeon + NVidia Hopper

All remaining are based on AMD EPYC Milan Trento + MI250x

EuroHPC

The MN-3 system

#1 in Jun'20 Green500

2 330

The GREEN 500

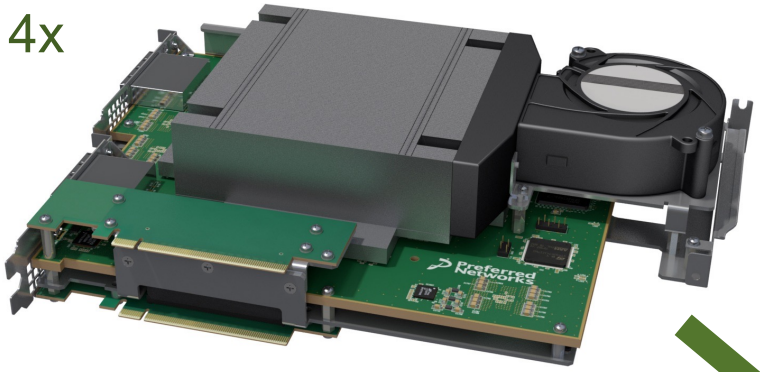
MN-3 - MN-Core Server, Xeon Platinum 8260M 24C 2.4GHz, Preferred Networks MN-Core, MN-Core DirectConnect, Preferred Networks

Preferred Networks Japan

Nov'20

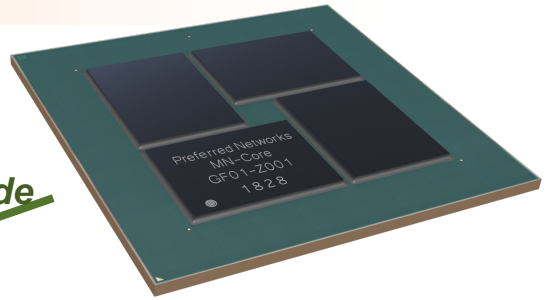
2x  + 4x

Xeon 8260M 24c

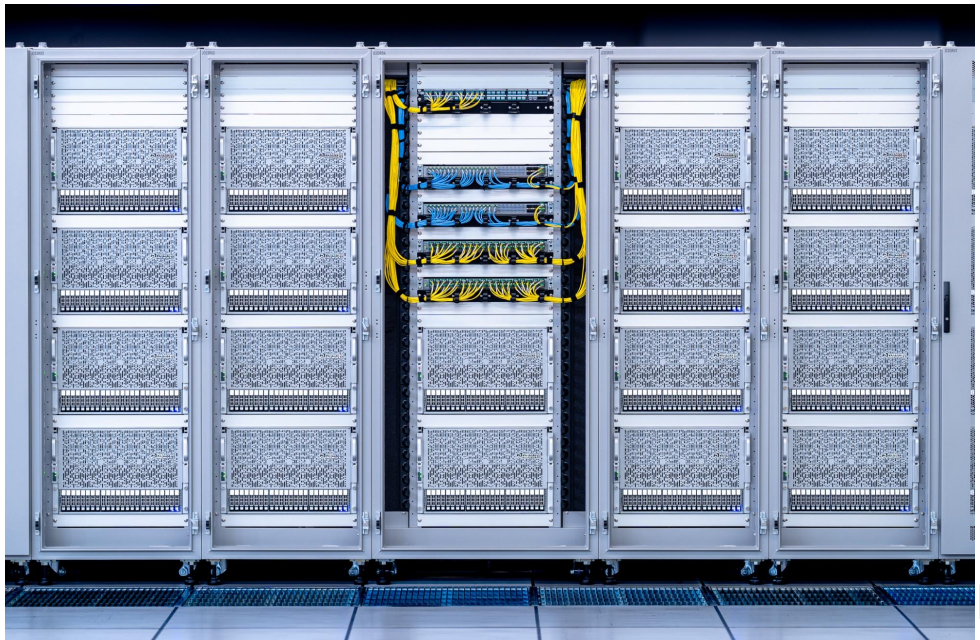


MN-Core Board

inside

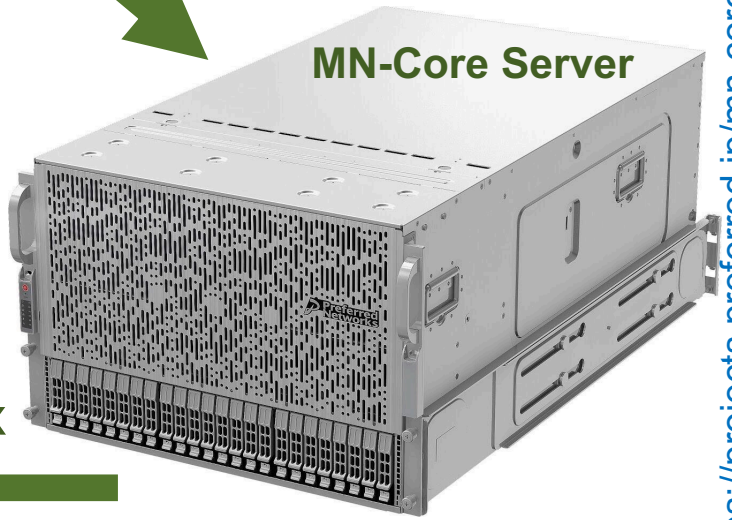


MN-Core accelerator



MN-3 *(partial view)*

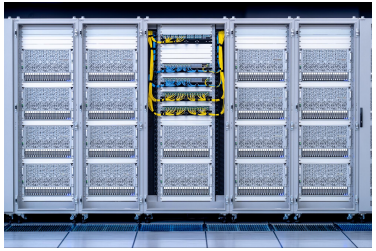
40x

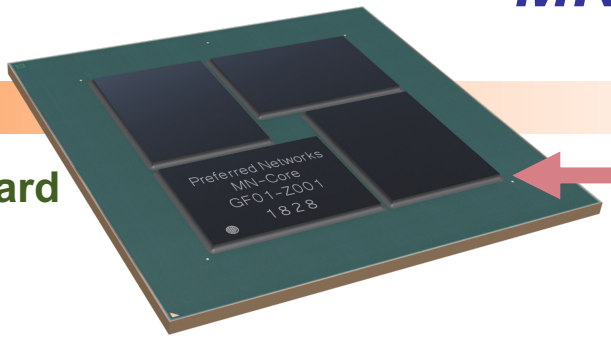


MN-Core Server

https://projects.preferred.jp/mn-core/en/

MN-Core architecture

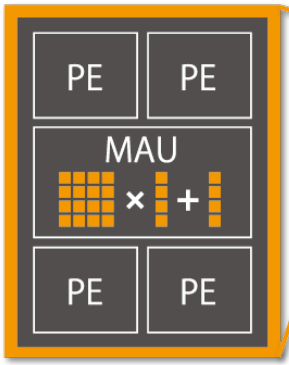




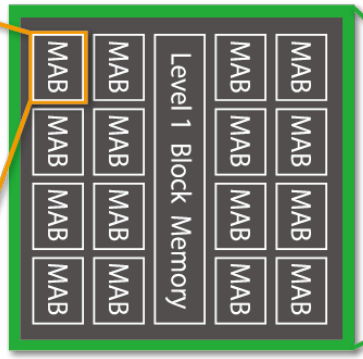
MN-Core Board

4 dies
each die

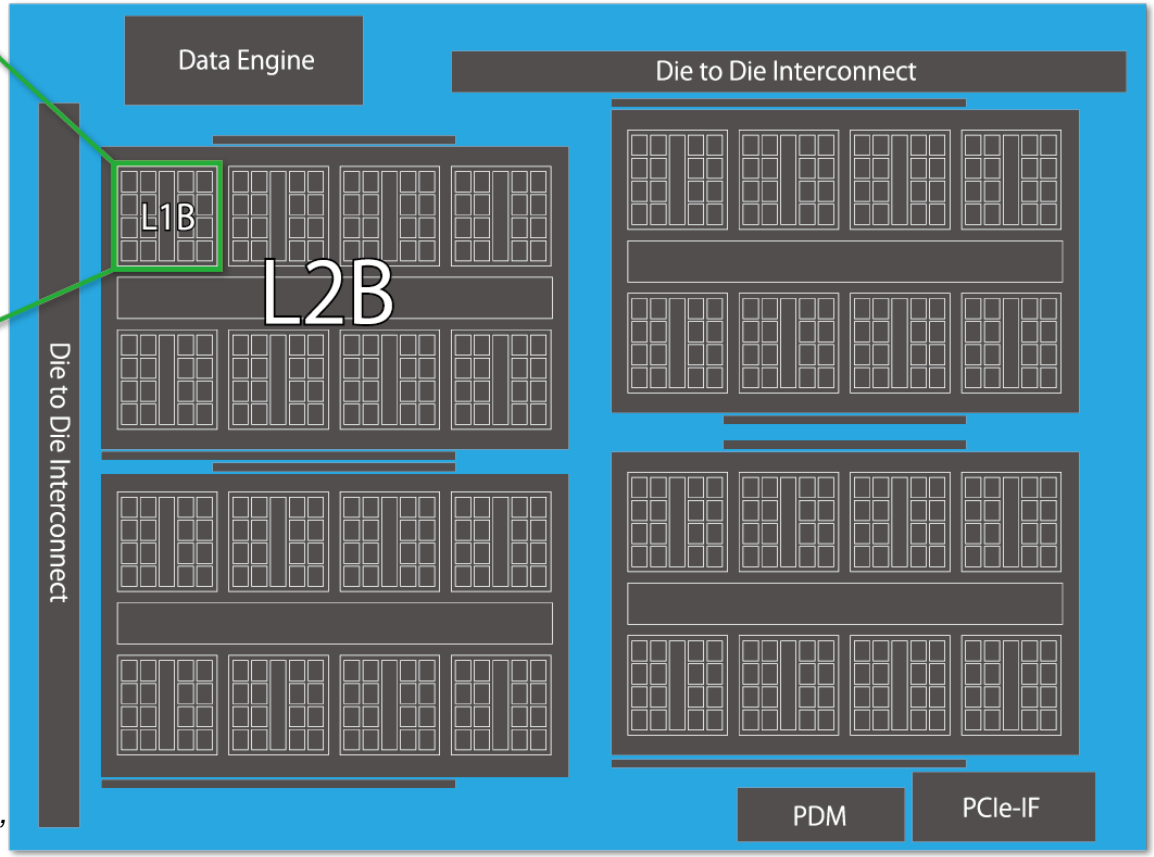
MAB



L1B



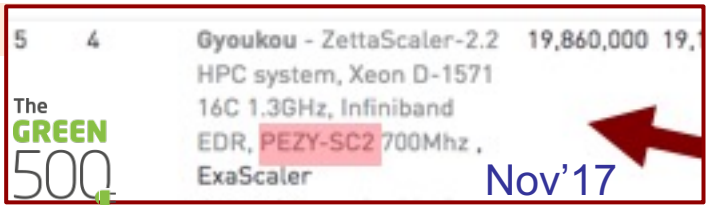
Chip



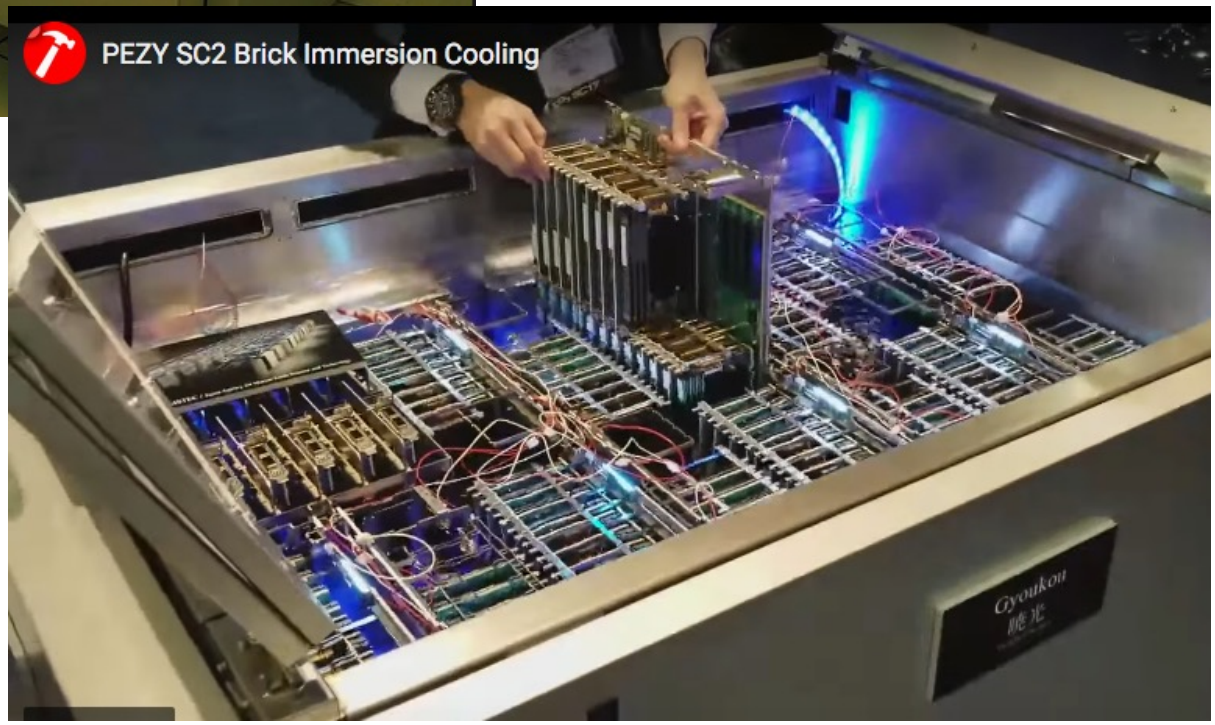
MAU: Matrix Arithmetic Unit
PE: Processing Element
MAB: Matrix Arithmetic Block
L1B: Level 1 Block w/ 16 MABs
L2B: Level 2 Block w/ 128 MABs
1 chip/die with 512 MABs
1 MN-Core package w/ 2048 MABs

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

Gyoukou ZettaScaler-2.2

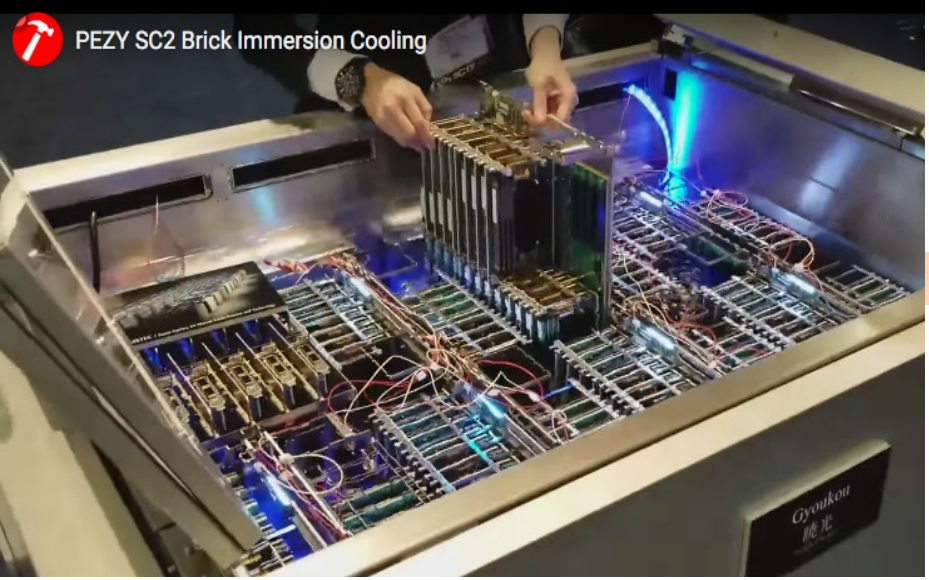


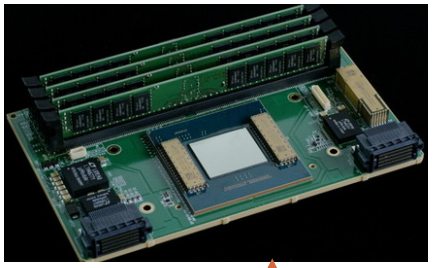
NA-1, #3 in Jun'20 is similar to Gyoukou, but w/less cores



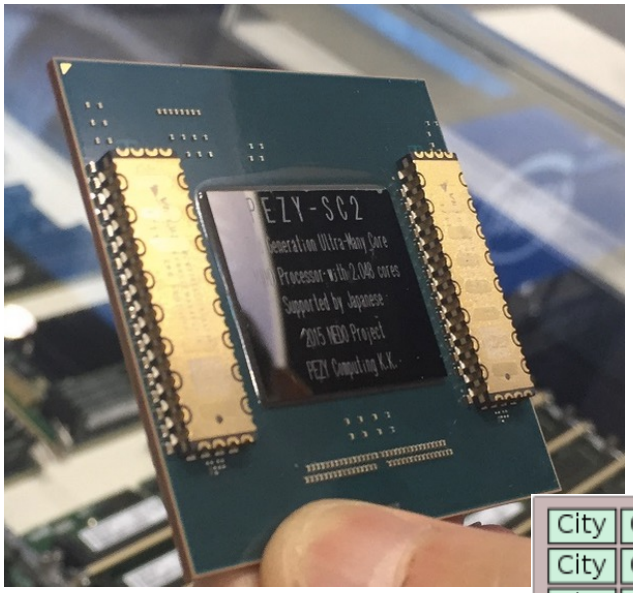
20 immersion tanks
each tank 16 bricks
each brick 32 PEZY
each PEZY
~2K 8-way SMT cores
=>
each tank ~1M cores

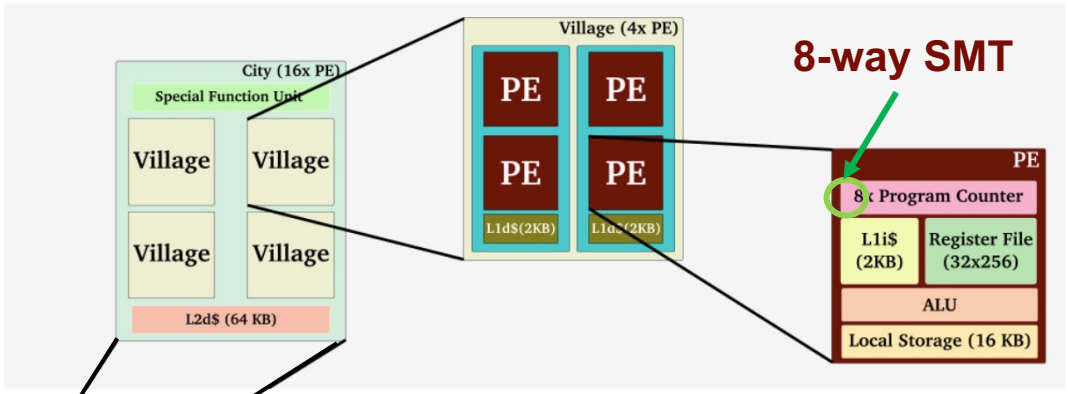
Gyoukou
ZettaScaler-2.2





PEZY-SC2 in ZettaScaler-2.2

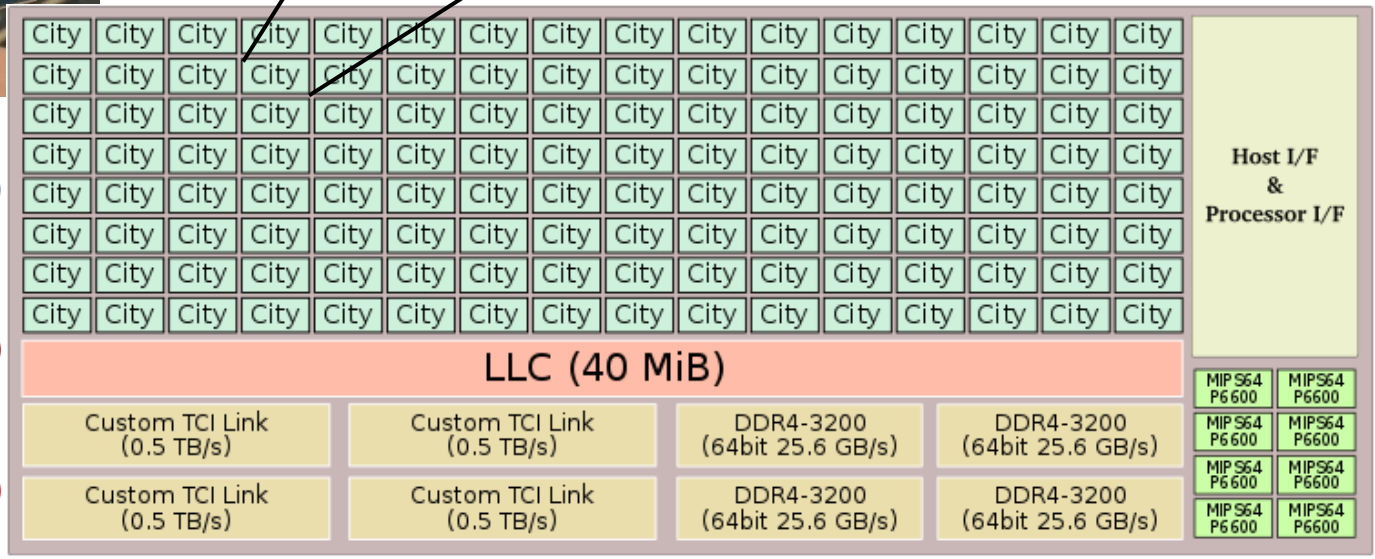


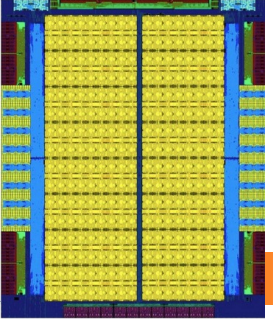


PEZY-SC2: **2 048 cores**
+ 8x MIPS cores *(2017)*

PEZY-SC3: 8 192 cores
(due 2019, arrived 2021…)

PEZY-SC4: 16 384 cores
(due in 2020, but…)





ZettaScaler-3.0 with PEZY-SC3:
estimation in 2019

1 tank => 40 nodes x **[1 AMD Epyc** (64-core) + **4 PEZY-SC3** (8192-core)**]**
1 tank has **1 312 320 cores**

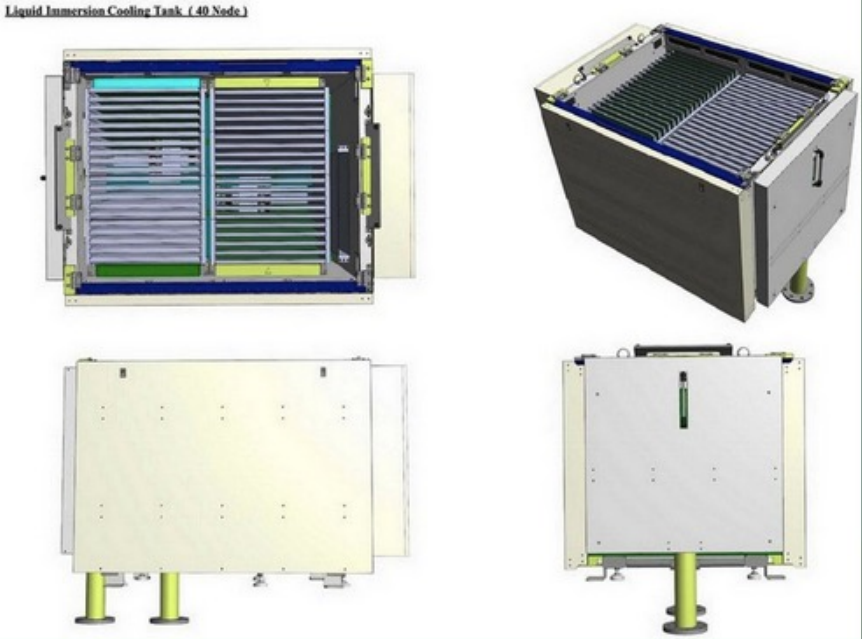
Estimated ZettaScaler-3.0 Specs

tanks configuration will provide about 100 PetaFLOPS (Rmax) and only consumes 4MW with the system cost of around $100M

ZettaScaler-3.0 single tank will have 40 nodes, 40 AMD EPYC2 (64 core), 160 of PEZY-SC3 with 48DC power

Single tank will have 3.2 PetaFLOPS (Rpeak) and 2.4 PetaFLOPS (Rmax) of DP performance

System power efficiency will be 30 GFLOPS/W or so and single tank requires 100kW range power



Outline

1. TOP500
   a) TOP10 lists from Nov'17 to Nov'22
   b) Analysis of some relevant systems and architectures
   c) Country distribution over the past 25 years
   d) PU chip technology evolution in the past 25 years and since last year
   e) Evolution of the accelerators since they were available

2. GREEN500
   a) TOP5 list from Nov'19 to Nov'22
   b) Analysis of some relevant systems and architectures

3. HPCG
   a) HPCG *vs.* HPL: an overview
   b) TOP10 in Nov'22

4. HPL-AI
   a) High-performance Linpack (HPL) and artificial intelligence (AI) workloads

HPCG benchmark

HPCG is a self-contained C++ program with MPI and OpenMP support 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 (part of Gauss-Seidel smoother)

HPCG benchmark

From Wikipedia, the free encyclopedia

The **HPCG** (**high performance conjugate gradient**) **benchmark** is a supercomputing benchmark test proposed by Michael Heroux from Sandia National Laboratories, and Jack Dongarra and Piotr Luszczek from the University of Tennessee.[1][2] It is intended to model the data access patterns of real-world applications such as sparse matrix calculations, thus testing the effect of limitations of the memory subsystem and internal interconnect of the supercomputer on its computing performance.[3] Because it is internally I/O bound, HPCG testing generally achieves only a tiny fraction of the peak FLOPS of the computer.[4]

HPCG is intended to complement benchmarks such as the LINPACK benchmarks that put relatively little stress on the internal interconnect.[5] The source of the HPCG benchmark is available on GitHub.[6]

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 systems
Nov'22

June 2022 HPCG Results

Rank	Site	Computer	Cores	HPL Rmax (Pflop/s)	TOP500 Rank	HPCG (Pflop/s)	Fraction of Peak
1	RIKEN Center for Computational Science **Japan**	**Supercomputer Fugaku** — A64FX 48C 2.2GHz, Tofu interconnect D	7,630,848	442.01	2	16.00	3.0%
2	DOE/SC/Oak Ridge National Laboratory **United States**	**Summit** — IBM POWER9 22C 3.07GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Volta GV100	2,414,592	148.60	4	2.926	1.5%
3	EuroHPC/CSC **Finland**	**LUMI** — AMD Optimized 3rd Generation EPYC 64C 2GHz, Slingshot-11, AMD Instinct MI250X	1,110,144	151.90	3	1.936	0.9%
4	DOE/SC/LBNL/NERSC **United States**	**Perlmutter** — AMD EPYC 7763 64C 2.45GHz, Slingshot-10, NVIDIA A100 SXM4 40 GB	761,856	70.87	7	1.905	2.0%
5	DOE/NNSA/LLNL **United States**	**Sierra** — IBM POWER9 22C 3.1GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Volta GV100	1,572,480	94.64	5	1.796	1.4%
6	NVIDIA Corporation **United States**	**Selene** — AMD EPYC 7742 64C 2.25GHz, Mellanox HDR Infiniband, NVIDIA A100	555,520	63.46	8	1.623	2.0%
7	Forschungszentrum Juelich (FZJ) **Germany**	**JUWELS Booster Module** — AMD EPYC 7402 24C 2.8GHz, Mellanox HDR InfiniBand/ParTec ParaStation ClusterSuite, NVIDIA A100	449,280	44.12	11	1.275	1.8%

5.5x

NVIDIA advantage?

Outline

1. TOP500
   a) TOP10 lists from Nov'17 to Nov'22
   b) Analysis of some relevant systems and architectures
   c) Country distribution over the past 25 years
   d) PU chip technology evolution in the past 25 years and since last year
   e) Evolution of the accelerators since they were available
2. GREEN500
   a) TOP10 lists from Nov'19 to Nov'22
   b) Analysis of some relevant systems
3. HPCG
   a) HPCG *vs.* HPL: an overview
   b) TOP10 in Nov'22
4. **HPL-AI**
   a) High-performance Linpack (HPL) and artificial intelligence (AI) workloads

HPL-MxP

HPL-AI Mixed-Precision Benchmark

HPL-MXP MIXED-PRECISION BENCHMARK

https://hpl-mxp.org/

The HPL-MxP benchmark seeks to highlight the emerging convergence of high-performance computing (HPC) and artificial intelligence (AI) workloads. While traditional HPC focused on simulation runs for modeling phenomena in physics, chemistry, biology, and so on, the mathematical models that drive these computations require, for the most part, 64-bit accuracy. On the other hand, the machine learning methods that fuel advances in AI achieve desired results at 32-bit and even lower floating-point precision formats. This lesser demand for accuracy fueled a resurgence of interest in new hardware platforms that deliver a mix of unprecedented performance levels and energy savings to achieve the classification and recognition fidelity afforded by higher-accuracy formats.

HPL-MxP strives to unite these two realms by delivering a blend of modern algorithms and contemporary hardware while simultaneously connecting to the solver formulation of the decades-old HPL framework of benchmarking the largest supercomputing installations in the world. The solver method of choice is a combination of LU factorization and iterative refinement performed afterwards to bring the solution back to 64-bit accuracy. The innovation of HPL-MxP lies in dropping the requirement of 64-bit computation throughout the entire solution process and instead opting for low-precision (likely 16-bit) accuracy for LU, and a sophisticated iteration to recover the accuracy lost in factorization. The iterative method guaranteed to be numerically stable is the generalized minimal residual method (GMRES), which uses application of the L and U factors to serve as a preconditioner. The combination of these algorithms is demonstrably sufficient for high accuracy and may be implemented in a way that takes advantage of the current and upcoming devices for accelerating AI workloads.

HPL-MxP

HPL-AI List
Nov'22

June 2022

Rank	Site	Computer	Cores	HPL-AI (Eflop/s)	TOP500 Rank
1	DOE/SC/ORNL, USA	Frontier	8,730,112	6.861	1
2	RIKEN, Japan	Fugaku	7,630,848	2.000	2
3	DOE/SC/ORNL, USA	Summit	2,414,592	1.411	4
4	NVIDIA, USA	Selene	555,520	0.630	8
5	DOE/SC/LBNL, USA	Perlmutter	761,856	0.590	7
6	FZJ, Germany	JUWELS BM	449,280	0.470	11
7	University of Florida, USA	HiPerGator	138,880	0.170	34

https://hpl-mxp.org/results.md

TOP500 Overview

1. TOP500

a) TOP10 lists from Nov'17 to Nov'22
b) Analysis of some relevant systems and architectures
c) Country distribution over the past 25 years
d) PU chip technology evolution in the past 25 years and since last year
e) Evolution of the accelerators since they were available

2. GREEN500

a) TOP10 lists from Nov'19 to Nov'22
b) Analysis of some relevant systems

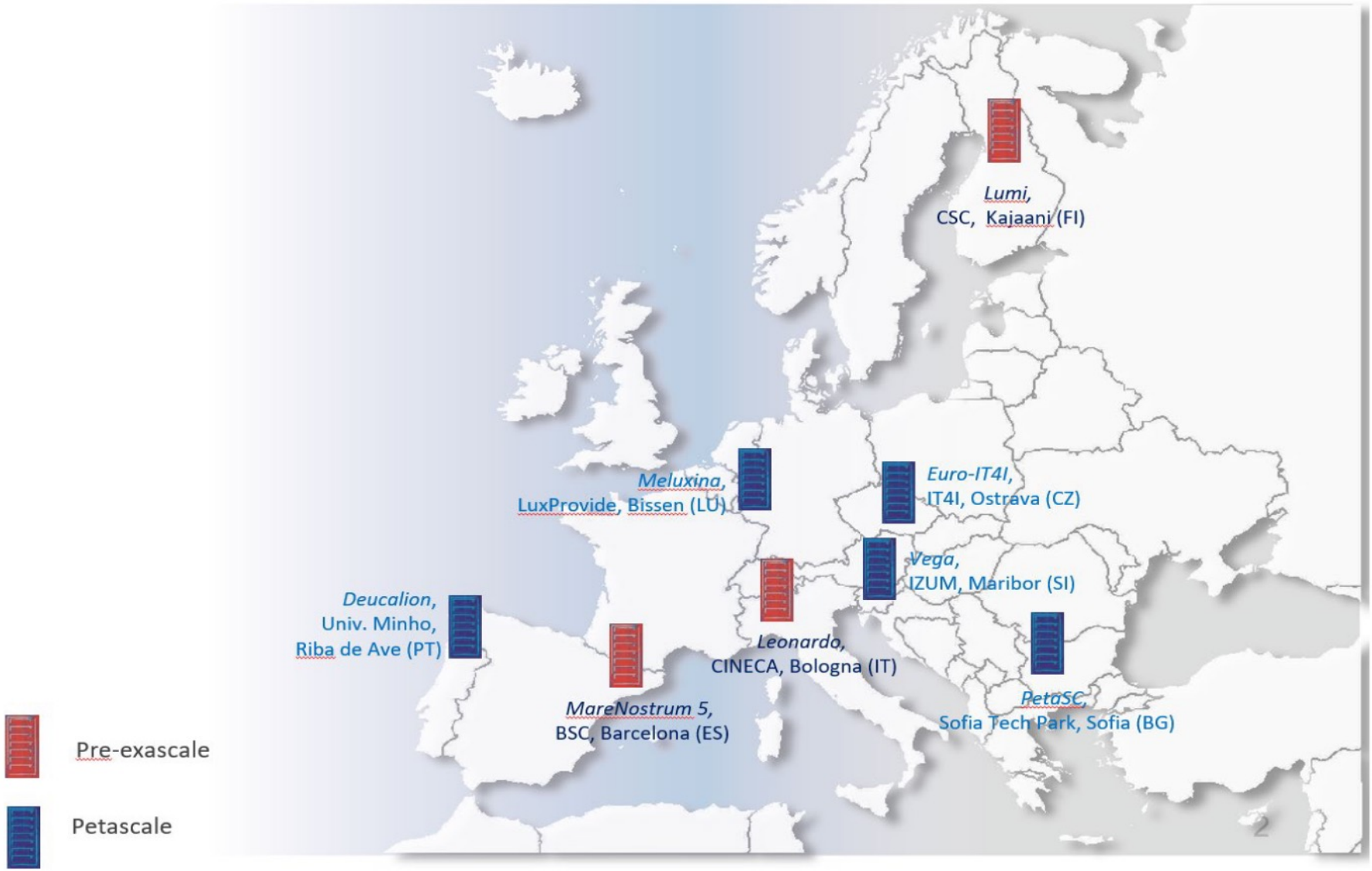
3. HPCG

a) HPCG *vs.* HPL: an overview
b) TOP10 in Nov'22

4. HPL-AI

a) High-performance Linpack (HPL) and artificial intelligence (AI) workloads

EuroHPC supercomputers:
their locations



https://eurohpc-ju.europa.eu/sites/default/files/2021-10/EuroHPC%20Systems%20Report-Sep2021.pdf

MACC

https://macc.fccn.pt/

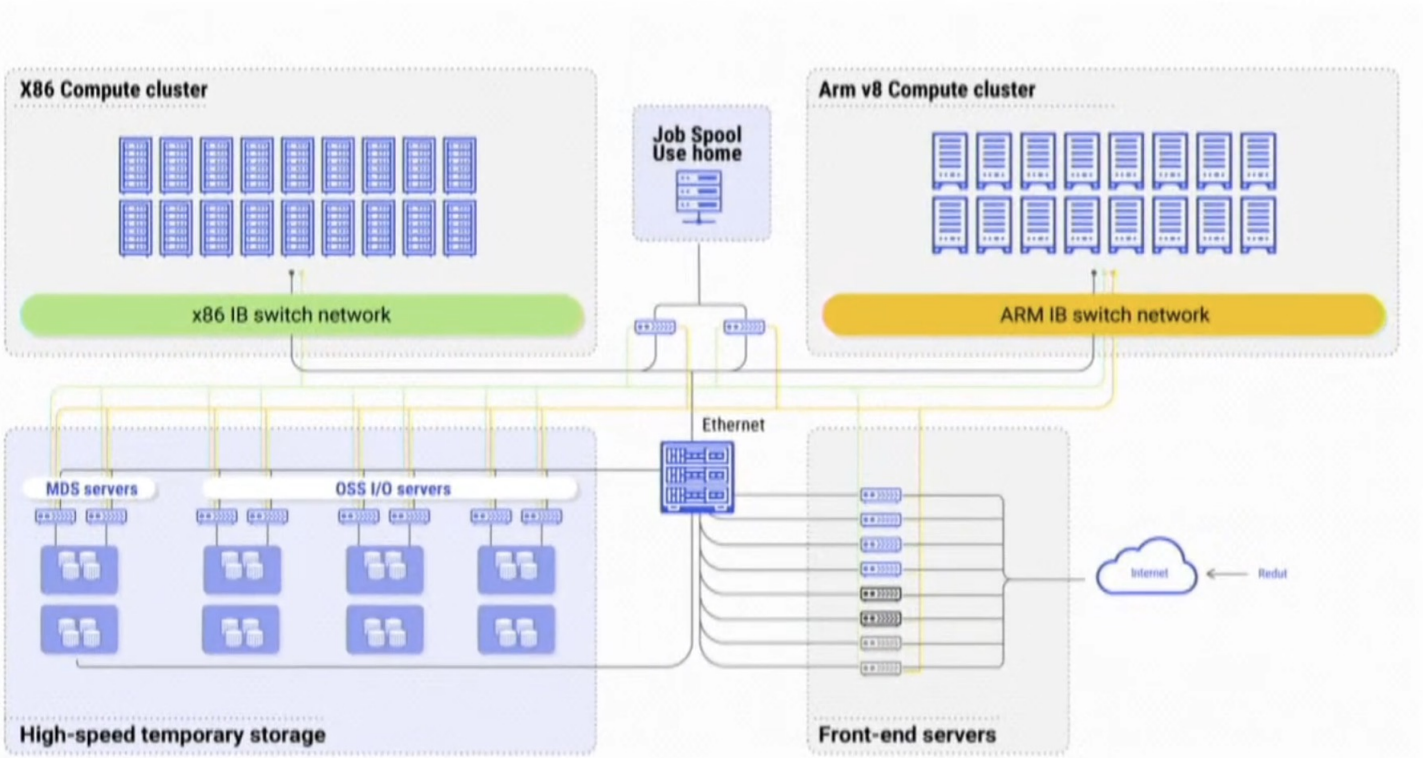
MACC Minho Advanced Computing Center

Advancing knowledge discovery for

Minho Advanced Computing Center is empowering all partners to discover, innovate, and

The Deucalion in MACC

DEUCALION OVERALL ARCHITECTURE



AMD X86 CLUSTER
EPYC Rome 7742 2.25GHz
500 DUAL NODES
64000 CORES
128TB RAM

NVIDIA A100 CLUSTER
33 DUAL + QUAD NODES
16 TB RAM + 5TB HBM

FUJITSU ARM CLUSTER
Fujitsu A64FX 2.0GHz
1632 NODES
78336 CORES
52TB HBM RAM

DDN HSS
10 PB SSD+HDD