SCIENTIFIC CONFERENCE
Russian Supercomputing Days

September 21-22, 2020

HPC: Where We Are Today And A Look Into The Future

Jack Dongarra
University of Tennessee
Oak Ridge National Laboratory
University of Manchester

State of Supercomputing in 2020

- Pflops (> 10^{15} Flop/s) computing fully established with all 500 systems.
- Three technology architecture possibilities or “swim lanes” are thriving.
  - Commodity (e.g. Intel)
  - Commodity + accelerator (e.g. GPUs) (144 systems; 134 NVIDIA, 6 Intel Phi + 4)
  - Lightweight cores (e.g. IBM BG, Xeon Phi, TaihuLight, ARM (3 systems))
- China: Top consumer and top producer overall.
- Interest in supercomputing is now worldwide, and growing in many new markets (~50% of Top500 computers are in industry).
- Intel processors largest share, 94%; followed by AMD, 2%.
- Exascale (10^{18} Flop/s) projects exist in many countries and regions.

TOP500 Highlights

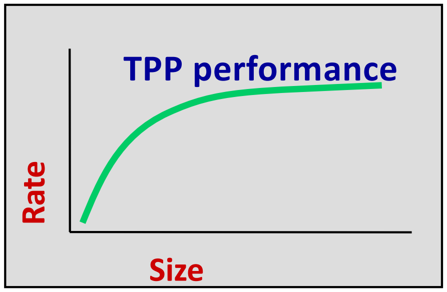
- Japanese's Fugaku is the new #1 in the TOP500
  - It measured at over 1 Exaflop on the HPL-AI benchmark which uses reduced precision arithmetic
- TOP10 has four new systems
- Overall turn-over in the Top500 is at a record low
  - Only 51 system dropped off, has been as high as 300
- TOP100 Research System and Commercial Systems show very different markets

TOP500 The List. Since 1993

H. Meuer, H. Simon, E. Strohmaier, & JD

- Listing of the 500 most powerful Computers in the World
- Yardstick: Rmax from LINPACK MPP

  Ax=b, dense problem



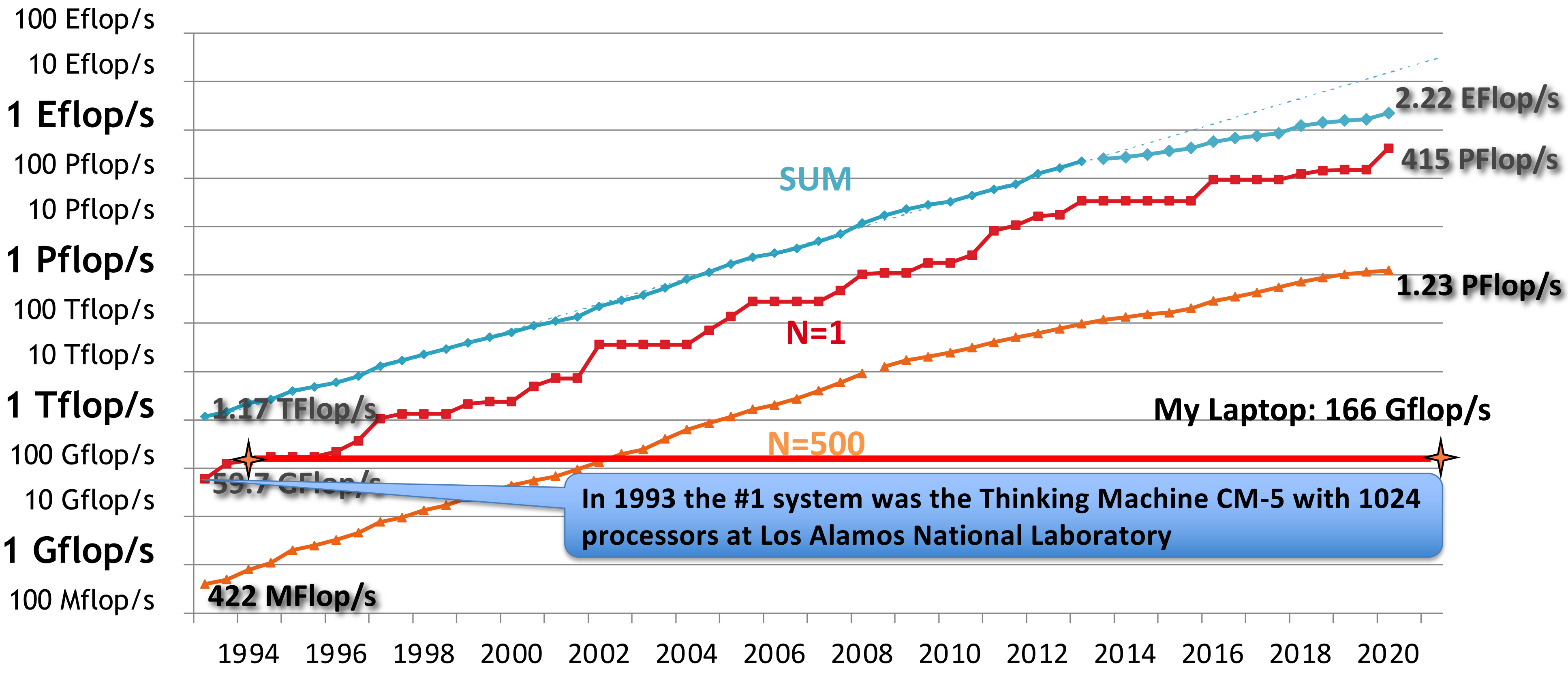
- Updated twice a year

  SC'xy in the States in November

  Meeting in Germany in June

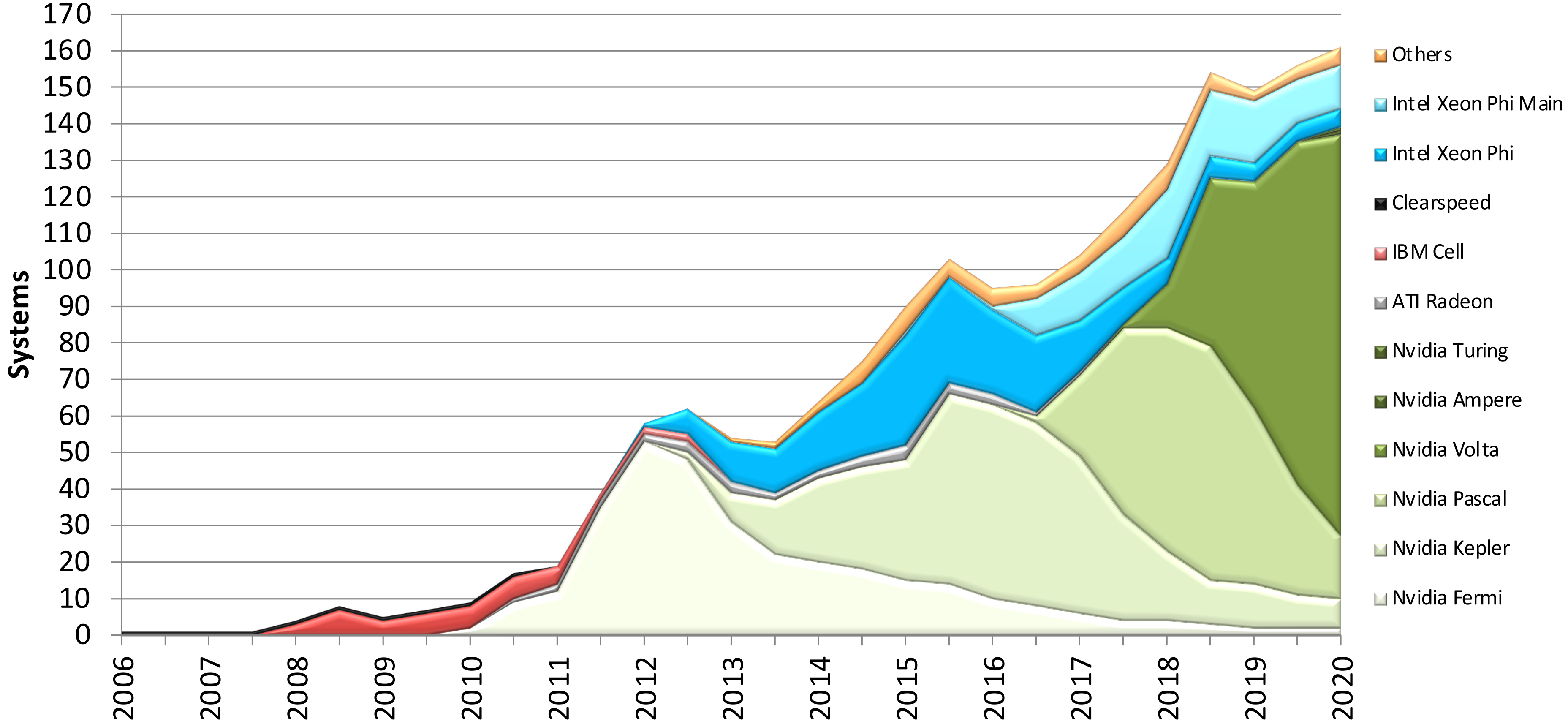
- All data available from **www.top500.org**

PERFORMANCE DEVELOPMENT



In 1993 the #1 system was the Thinking Machine CM-5 with 1024 processors at Los Alamos National Laboratory

ACCELERATORS – NVIDIA DOMINATES W/134

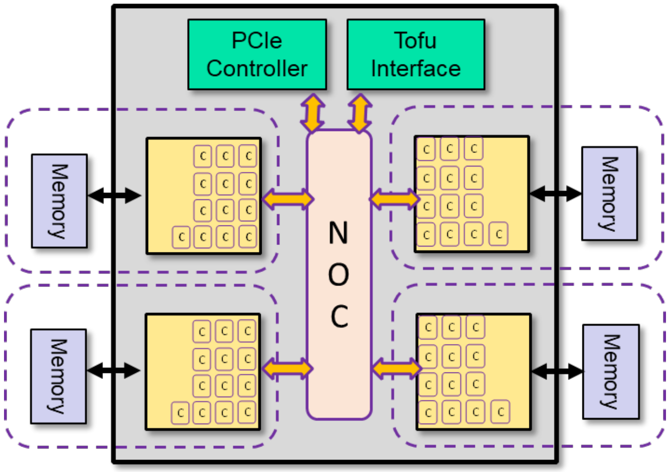


June 2020: The TOP 10 Systems (43% of the Total Performance of Top500)

Rank	Site	Computer	Country	Cores	Rmax [Pflops]	% of Peak	Power [MW]	GFlops/ Watt
1	RIKEN Center for Computational Science	Fugaku, ARM A64FX (48C, 2.2 GHz), Tofu D Interconnect	Japan	7,299,072	415.	81	28.3	14.7
2	DOE / OS Oak Ridge Nat Lab	Summit, IBM Power 9 (22C, 3.0 GHz), Nvidia GV100 (80C), Mellonox EDR	USA	2,397,824	149.	74	10.1	14.7
3	DOE / NNSA L Livermore Nat Lab	Sierra, IBM Power 9 (22C, 3.1 GHz), Nvidia GV100 (80C), Mellonox EDR	USA	1,572,480	94.6	75	7.44	12.7
4	National Super Computer Center in Wuxi	Sunway TaihuLight, SW26010 (260C) + Custom	China	10,649,000	93.0	74	15.4	6.05
5	National Super Computer Center in Guangzhou	Tianhe-2A NUDT, Xeon (12C) + MATRIX-2000 + Custom	China	4,981,760	61.4	61	18.5	3.32
6	Eni S.p.A	HPC5, Dell EMC PowerEdge C4140, Xeon (24C, 2.1 GHz) + Nvidia V100 (80C), Mellonax HDR	Italy	669,760	35.5	69	2.25	15.8
7	NVIDIA Corporation	Selene, Nvidia DGX AMD (64C, 2.25 GHz) + Nvidia A100 (108C), Mellanox HDR	USA	277,760	27.6	80	1.34	20.6
8	Texas Advanced Computing Center / U of Texas	Frontera, Dell C6420, Xeon Platinum, 8280 28C 2.7 GHz, Mellanox HDR	USA	448,448	23.5	61		
9	CINECA	Marconi-100, IBM Power System AC922, P9 (16C, 3 GHz) + Nvidia V100 (80C), Mellonox EDR	Italy	347,776	21.6	74	1.98	10.9
10	Swiss CSCS	Piz Daint, Cray XC50, Xeon (12C) + Nvidia P100 (56C) + Custom	Swiss	387,872	21.2	78	2.38	8.90

Fugaku's Fujitsu A64fx Processor is…

- A Many-Core ARM CPU…
  - 48 compute cores + 2 or 4 assistant (OS) cores
  - New core design
  - Near Xeon-Class Integer performance core
  - ARM V8 --- 64bit ARM ecosystem
  - Interconnect Tofu-D
  - 3.4 TFLOP/s Peak 64-bit performance

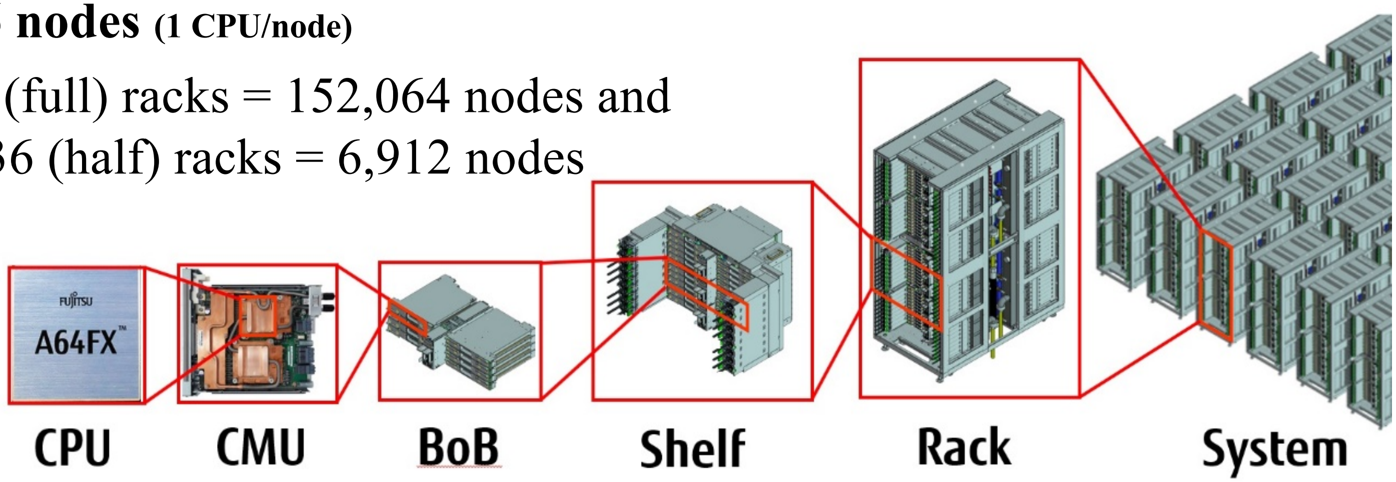


- …but also an accelerated GPU-like processor
  - SVE 512 bit x 2 vector extensions (ARM & Fujitsu)
    - Integer (1, 2, 4, 8 bytes) + Float (16, 32, 64 bytes)
  - Cache + memory localization (sector cache)
  - HBM2 on package memory – Massive Mem BW (Bytes/DPF ~0.4)
    - Streaming memory access, strided access, scatter/gather etc.
  - Intra-chip barrier synch. and other memory enhancing features

http://bit.ly/fugaku-report

Fugaku Total System Config & Performance

- **Total # Nodes: 158,976 nodes (1 CPU/node)**
  - 384 nodes/rack x 396 (full) racks = 152,064 nodes and
    192 nodes/rack x 36 (half) racks = 6,912 nodes



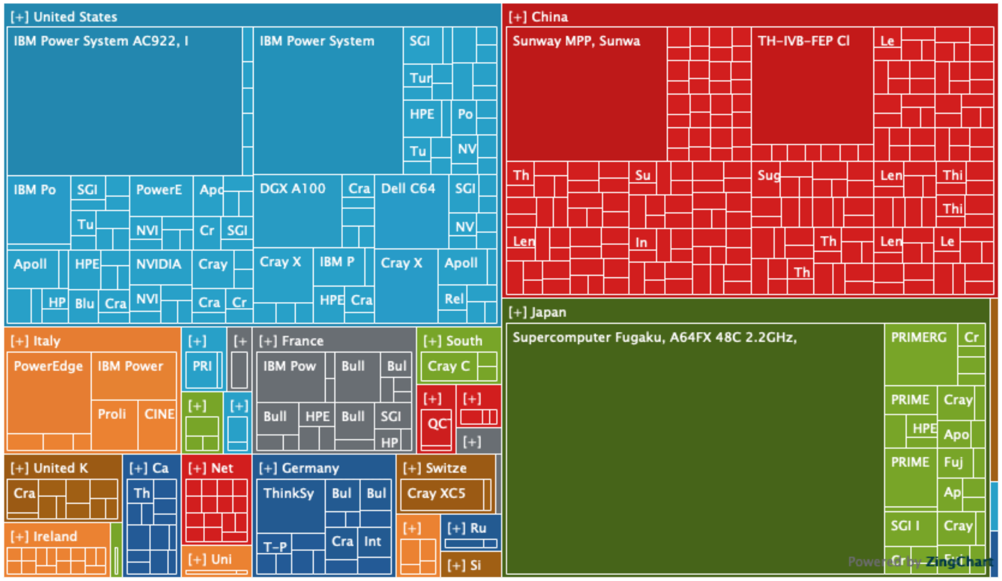
Footprint: 1,920 m^2

- **Theoretical Peak Compute Performances**
  - Normal Mode (CPU Frequency 2GHz)
    - **64 bit** Double Precision FP: **488 Petaflops**
    - **32 bit** Single Precision FP: **977 Petaflops**
    - **16 bit** Half Precision FP (AI training): **1.95 Exaflops**
    - **8 bit Integer** (AI Inference): **3.90 Exaops**
- **Theoretical Peak Memory BW: 163 Petabytes/s**

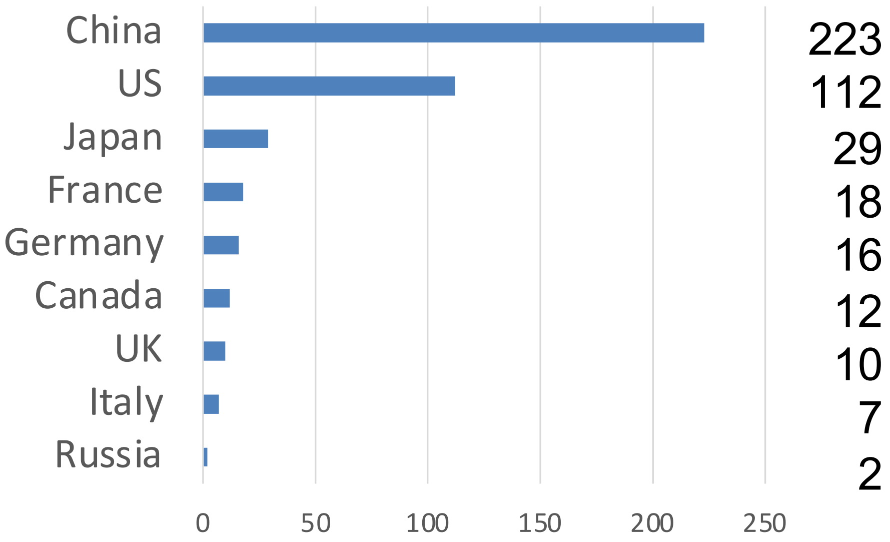
Fugaku represents 19% of all the other Top500 systems.

http://bit.ly/fugaku-report

COUNTRIES SHARE



Count of Number of Systems in Country



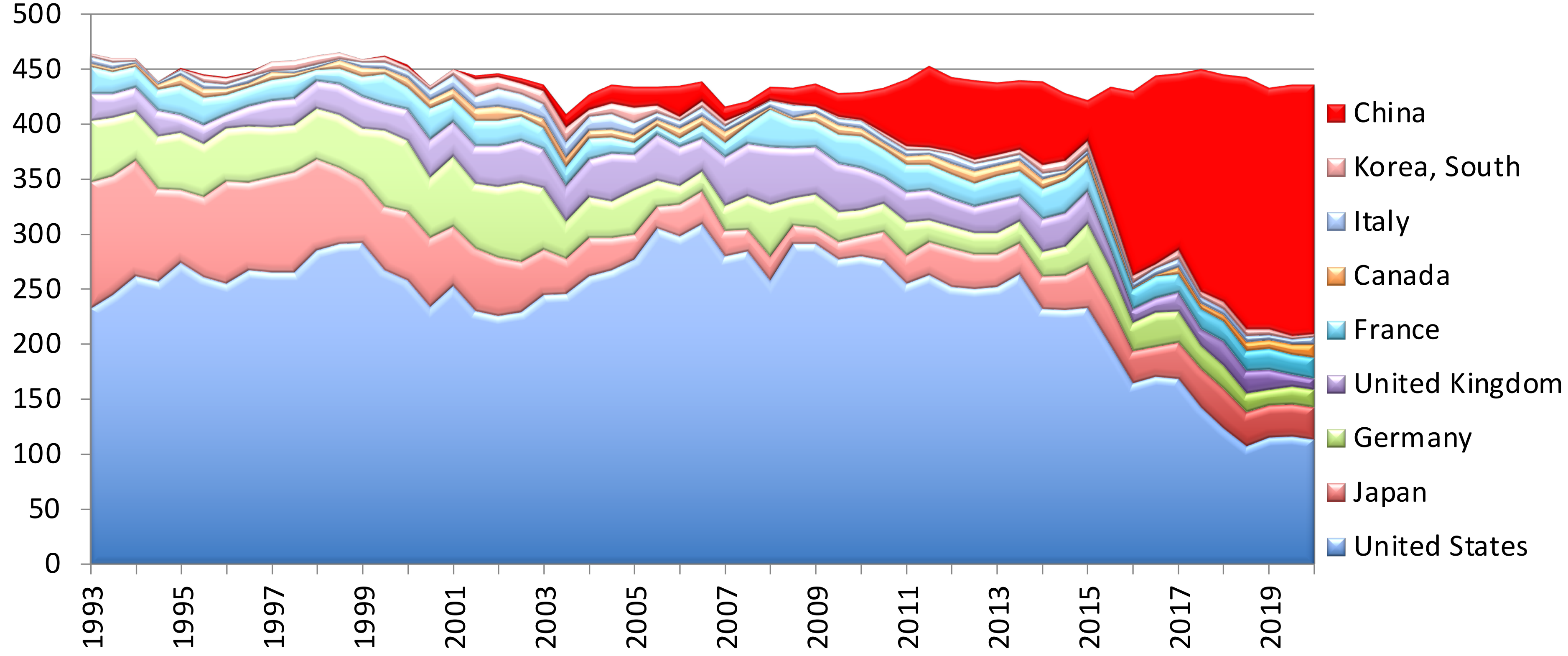
In terms of number of systems: China has 42% of the systems
US has 23% of the systems
Japan has 5.8% of the systems

In terms of performance: US has 28%
China has 26%
Japan has 23%

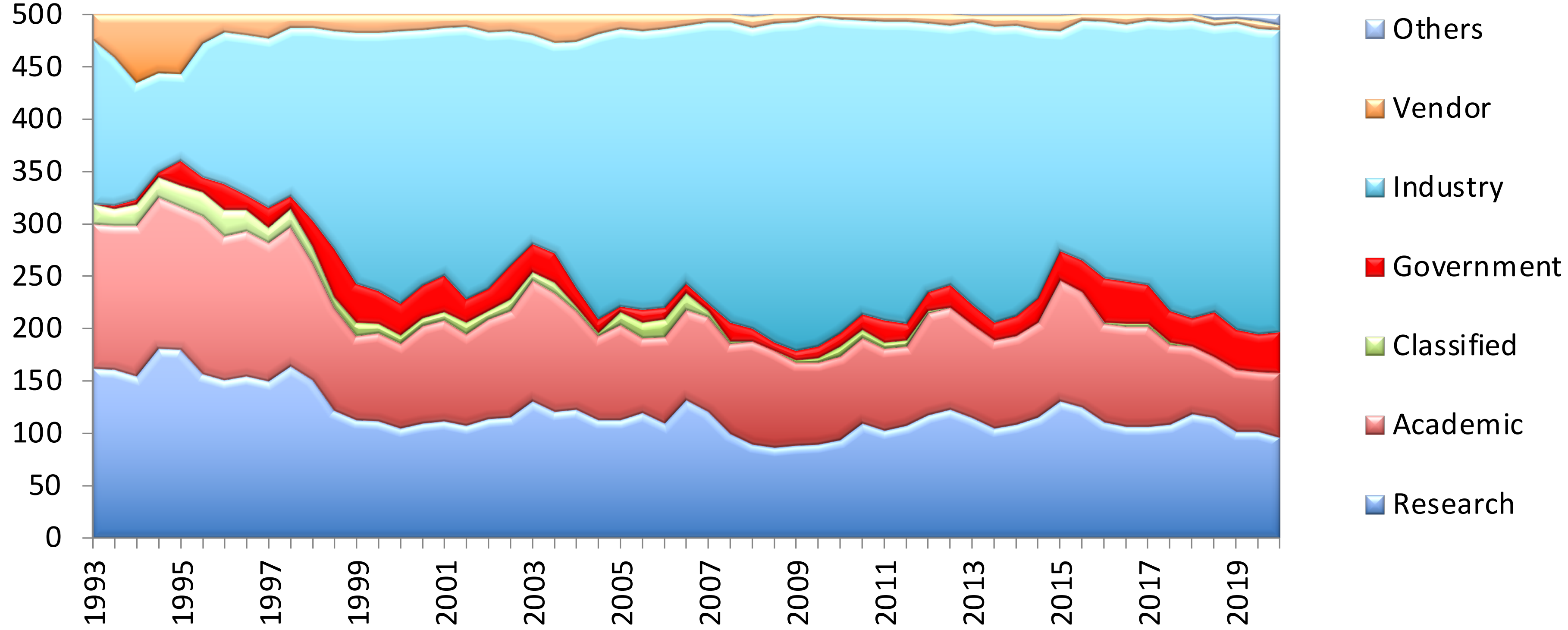
TWO RUSSIAN SYSTEMS ON TOP500

Rank	Name	Computer	Site	Manufacturer	Year	Segment	Total Cores	Accelerator/Co-Processor Cores	LINPACK Rmax [TFlop/s]	Rpeak [TFlop/s]	Accelerator/Co-Processor	Processor Generation
36	Christofari	NVIDIA DGX-2, Xeon Platinum 8168 24C 2.7GHz, Mellanox InfiniBand EDR, NVIDIA Tesla V100	SberCloud	Nvidia DGX-2	2019	Industry	99600	96000	6669	8790	NVIDIA Tesla V100	Xeon Platinum
132	Lomonosov 2	T-Platform A-Class Cluster, Xeon E5-2697v3 14C 2.6GHz,Intel Xeon Gold 6126, Infiniband FDR, Nvidia K40m/P-100	Moscow State University - Research Computing Center	T-Platforms A-Class Cluster	2014	Academic	64384	40960	2478	4945	NVIDIA Tesla K40m	Intel Xeon E5 (Haswell)

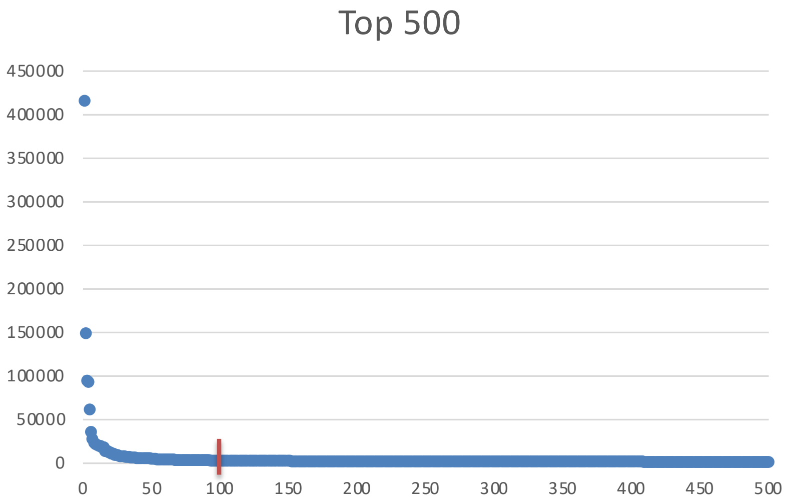
COUNTRIES BY COUNT

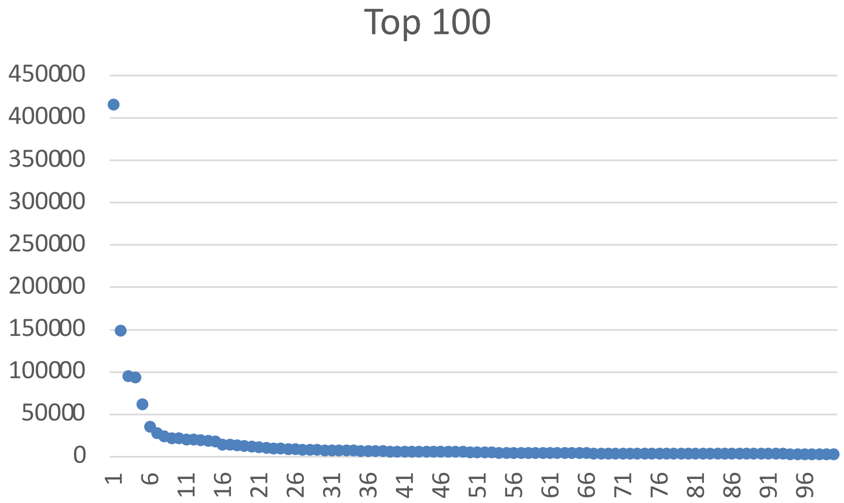


MARKET SEGMENTS



Performance Distribution

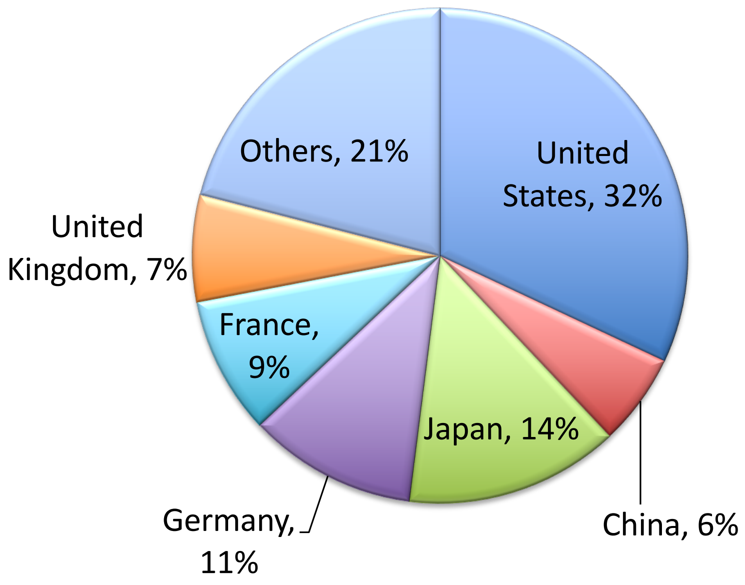




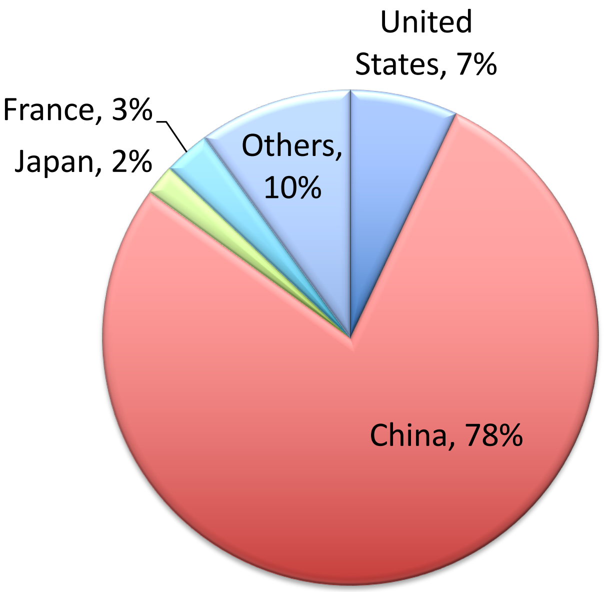
For all 500 systems
Tpeak_{500} = 1.23 Pflop/s

For top 100 systems
Tpeak_{100} = 2.8 Pflop/s
(36 systems use GPUs)
68% of the total performance of Top500 in Top100

COUNTRIES / SYSTEM SHARE FOR TOP100

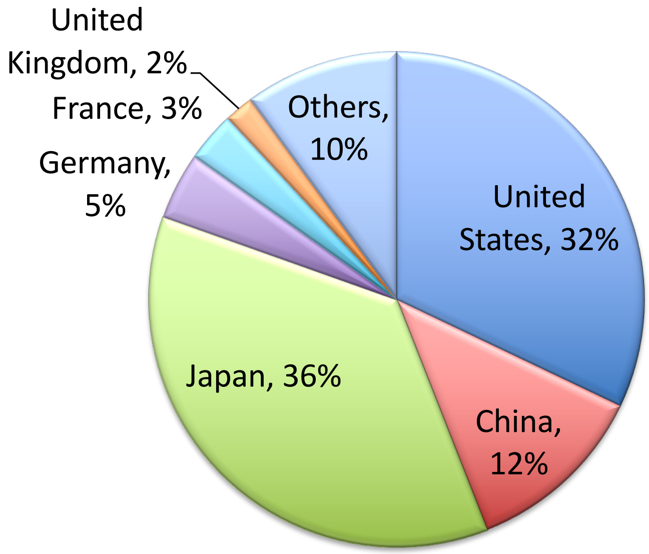


Research

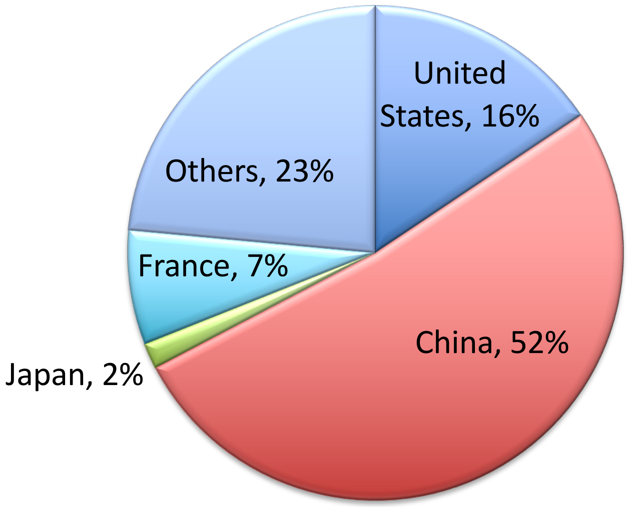


Commercial

COUNTRIES / PERFORMANCE SHARE FOR TOP100



Research

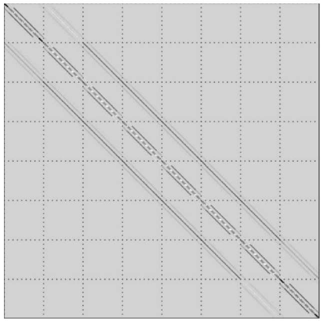


Commercial

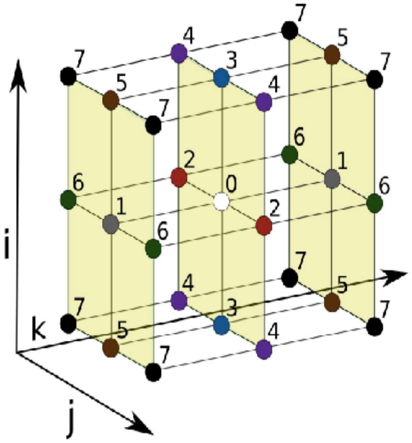
hpcg-benchmark.org

HPCG Results; The Other Benchmark

- 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).

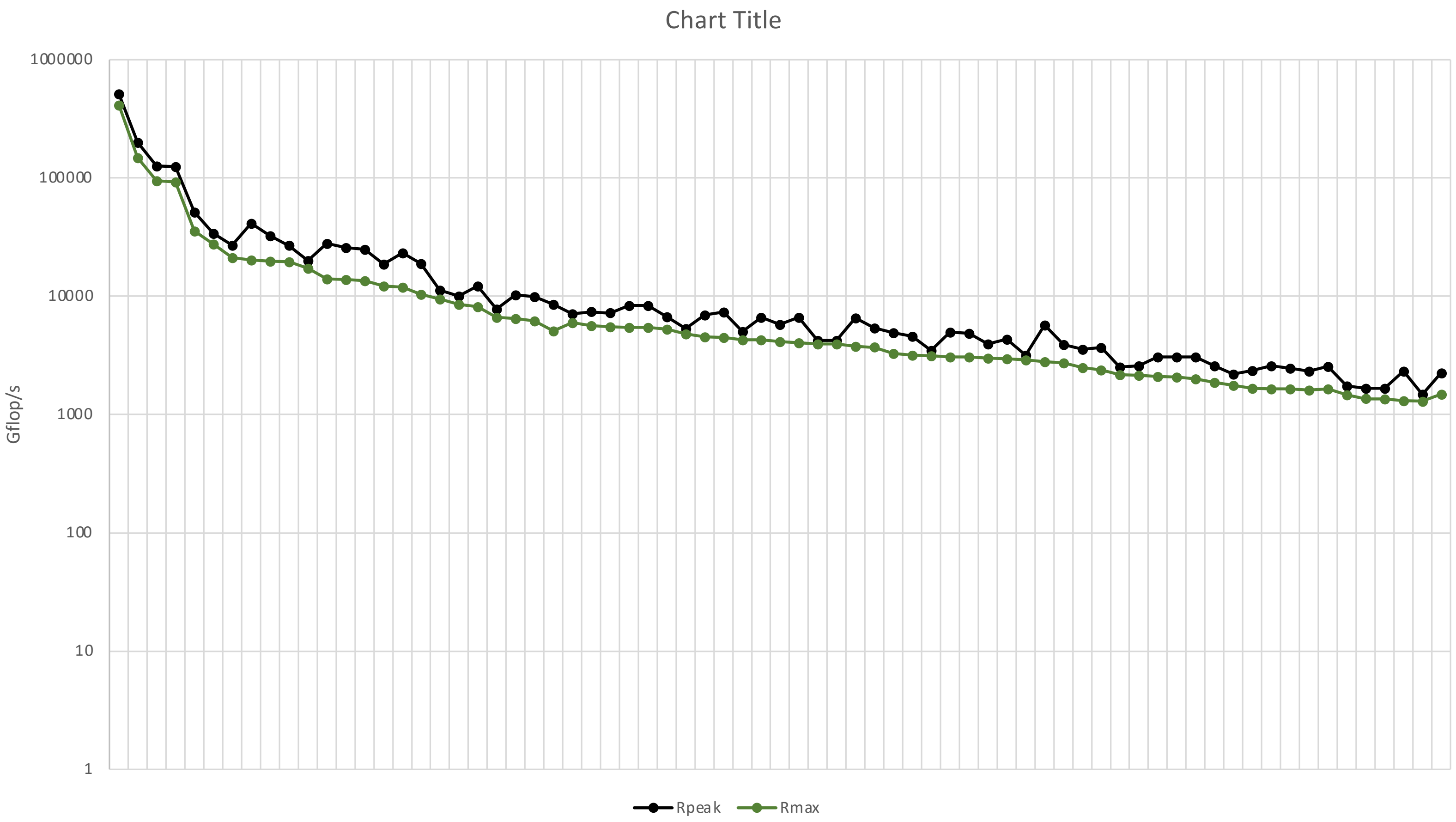


27-point stencil operator

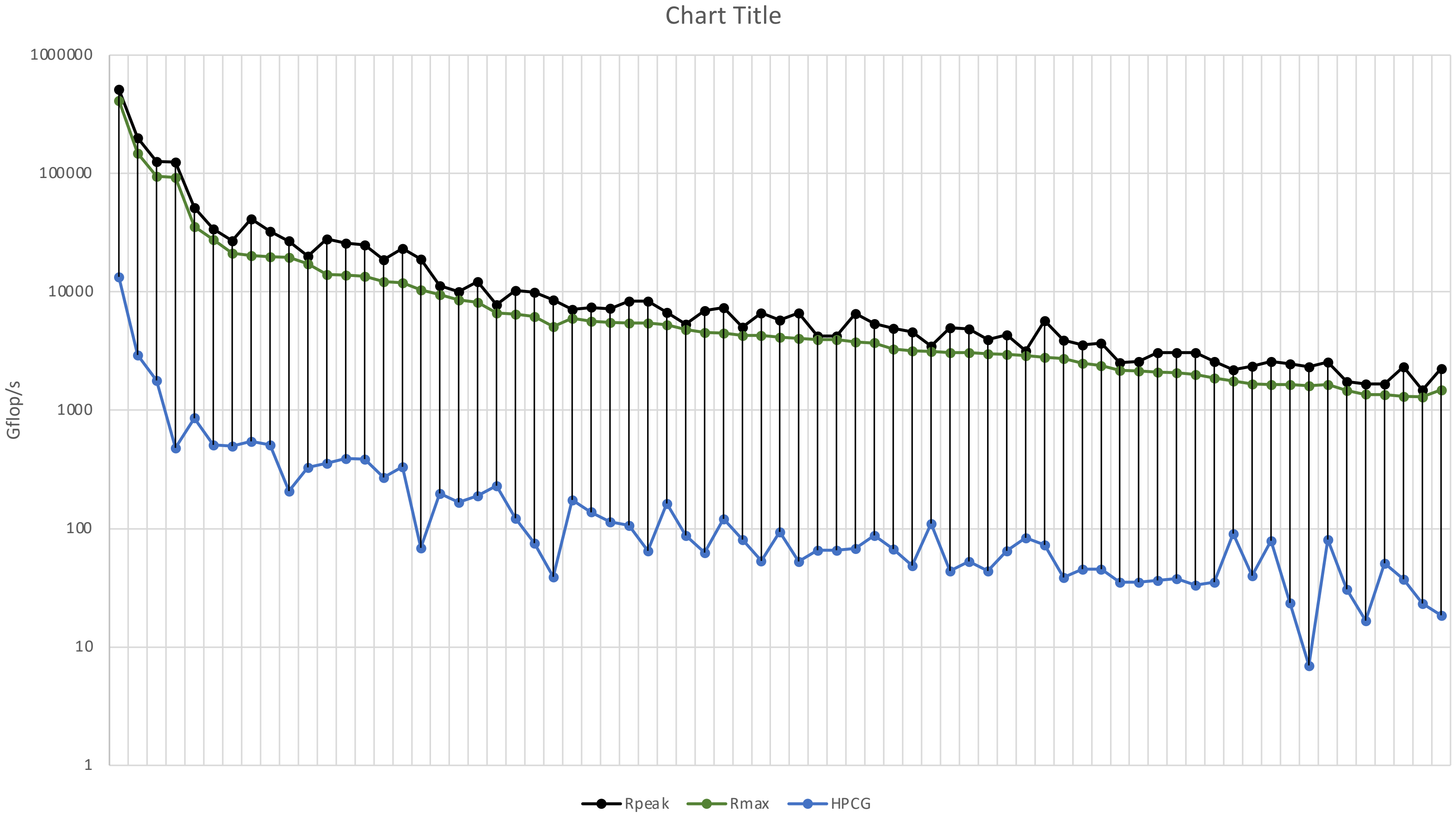
HPCG Benchmark June 2020

Rank	Site	Computer	Cores	HPL Rmax (Pflop/s)	TOP500 Rank	HPCG (Pflop/s)	Fraction of Peak
1	RIKEN Center for Computational Science Japan	**Fugaku**, Fujitsu A64FX, Tofu	7,299,072	415.53	1	13.4	2.5%
2	DOE/SC/ORNL USA	**Summit**, AC922, IBM POWER9 22C 3.7GHz, Dual-rail Mellanox FDR, NVIDIA Volta V100, IBM	2,414,592	143.50	2	2.926	1.5%
3	DOE/NNSA/LLNL USA	**Sierra**, S922LC, IBM POWER9 20C 3.1 GHz, Mellanox EDR, NVIDIA Volta V100, IBM	1,572,480	94.64	3	1.796	1.4%
4	Eni S.p.A. Italy	**HPC5**, PowerEdge, C4140, Xeon Gold 6252 24C 2.1 GHz, Mellanox HDR, NVIDIA Volta V100	669,760	35.45	6	0.860	2.4%
5	DOE/NNSA/LANL/SNL USA	**Trinity,** Cray XC40, Intel Xeon E5-2698 v3 16C 2.3GHz, Aries, Cray	979,072	20.16	11	0.546	1.3%
6	NVIDIA USA	**Selene**, DGX SuperPOD, AMD EPYC 7742 64C 2.25 GHz, Mellanox HDR, NVIDIA Ampere A100	277,760	27.58	7	0.5093	1.8%
7	Natl. Inst. Adv. Industrial Sci. and Tech. (AIST) Japan	**ABCI**, PRIMERGY CX2570M4, Intel Xeon Gold 6148 20C 2.4GHz, Infiniband EDR, NVIDIA Tesla V100, Fujitsu	391,680	16.86	12	0.5089	1.7%
8	Swiss National Supercomputing Centre (CSCS) Switzerland	**Piz Daint**, Cray XC50, Intel Xeon E5-2690v3 12C 2.6GHz, Cray Aries, NVIDIA Tesla P100 16GB, Cray	387,872	19.88	10	0.497	1.8%
9	National Supercomputing Center in Wuxi China	**Sunway** TaihuLight, Sunway MPP, SW26010 260C 1.45GHz, Sunway, NRCPC	10,649,600	93.01	4	0.481	0.4%
10	Korea Institute of Science and Technology Information Republic of Korea	**Nurion**, CS500, Intel Xeon Phi 7250 68C 563584C 1.4GHz, Intel Omni-Path, Intel Xeon Phi 7250, Cray	570,020	13.93	18	0.391	1.5%

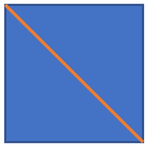
Comparison Peak, HPL, and HPCG: June 2020



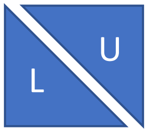
Comparison Peak, HPL, and HPCG: June 2020



HPL-AI Benchmark Utilizing 16-bit Arithmetic



1. Generate random linear system Ax=b
2. Represent the matrix A in low precision (16-bit floating point)
3. Factor A in lower precision into LU by Gaussian elimination



4. Compute approximate solution with LU factors in low precision
5. Perform up to 50 iterations of refinement, e.g., GMRES to get accuracy up to 64-bit floating point
    a. Use LU factors for preconditioning
6. Validate the answer is correct: scaled residual small $\frac{||Ax-b||}{||A||||x||+||b||} \times \frac{1}{n\epsilon} \leq O(10)$
7. Compute performance rate as $\frac{2}{3} \times \frac{n^3}{\text{time}}$

HPL-AI Benchmark List: #1 and #2

Rank	Site	Computer	Cores	HPL Rmax (Eflop/s)	TOP500 Rank	HPL-AI (Eflop/s)	Speedup
1	RIKEN Center for Computational Science Japan	**Fugaku**, Fujitsu A64FX, Tofu D	7,299,072	0.416	1	1.42	3.42x
2	DOE/SC/ORNL USA	**Summit**, AC922 IBM POWER9, IB Dual-rail FDR, NVIDIA Volta V100	2,414,592	0.144	2	0.551	3.83x

Department of Energy's Roadmap to Exascale Systems

An impressive, productive lineup of *accelerated node* systems supporting DOE's mission

Pre-Exascale Systems [Aggregate Linpack (Rmax) = 323 PF]

First U.S. Exascale Systems

2012	**2016**	**2018**	**2020**	**2021-2023**

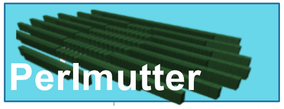
ORNL
Cray/AMD/NVIDIA

ANL
IBM BG/Q

ANL
Cray/Intel KNL

LBNL
Cray/Intel Xeon/KNL

ORNL
IBM/NVIDIA



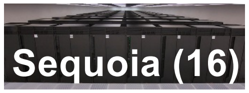
LBNL
Cray/AMD/NVIDIA

$1.8 B,
Just for 3 Hardware systems at Exascale

FRONTIER

ORNL
AMD/Cray

ANL
Intel/Cray



LLNL
IBM BG/Q

LANL/SNL
Cray/Intel Xeon/KNL

LLNL
IBM/NVIDIA

CROSSROADS

LANL/SNL
TBD

LLNL
AMD/Cray

Frontier System Overview

FRONTIER

OAK RIDGE National Laboratory | OAK RIDGE LEADERSHIP COMPUTING FACILITY

System Specs	Titan 2012	Summit 2018	Frontier 2021
Peak Performance	27 PF	200 PF	>1.5 EF
Footprint	200 cabinets	256	More than 100 cabinets (~7,300 square feet)
Node	1 AMD Opteron CPU 1 NVIDIA K20X Kepler GPU	2 IBM POWER9™ CPUs 6 NVIDIA Volta GPUs	1 HPC and AI Optimized AMD EPYC CPU 4 Purpose Built AMD Radeon Instinct GPU Coherent memory across the node High-bandwidth GPU-CPU link
CPU-GPU Interconnect	PCI Gen2	NVLINK Coherent memory across the node	AMD Infinity Fabric Coherent memory across the node Multiple slingshot NICs providing 100 GB/s network bandwidth
System Interconnect	Gemini	2x Mellanox EDR 100Gb/s InfiniBand Non-blocking Fat-Tree	Multiple Cray Slingshot NICs providing 100 GB/s network bandwidth. Slingshot dragonfly network which provides adaptive routing, congestion management and quality of service.
Storage	32 PB Lustre Filesystem 1 TB/s	250 PB, 2.5 TB/s, Spectrum Scale using GPFS™ technology	2-4x performance and capacity of Summit's I/O subsystem. Frontier will have near node storage like Summit.

Chinese plans for Exascale in 2020-2021

- Three separate developments in HPC; "Anything but from the US"
- Wuxi
  - Upgrade the ShenWei O(100) Pflops
- National University for Defense Technology
  - Tianhe-2A O(100) Pflops will be Chinese ARM processor + accelerator
- Sugon - CAS ICT
  - X86 + accelerator based; collaboration with AMD

China’s Plans

- 2020: Shandong Jinangnan institute
  - Developed by Sunway
  - Peak 1 Eflop/s
- 2021: Tianjin
  - Developed by National University for Defense Technology
  - Peak 1 Eflop/s
- 2022: Shenzhen Dawning
  - Developed by Sugon
  - Peak 2 Eflop/s

07

- There is a proposal (no decision on this) for two 10-Eflop/s systems in the next 5 year plan (21-25)

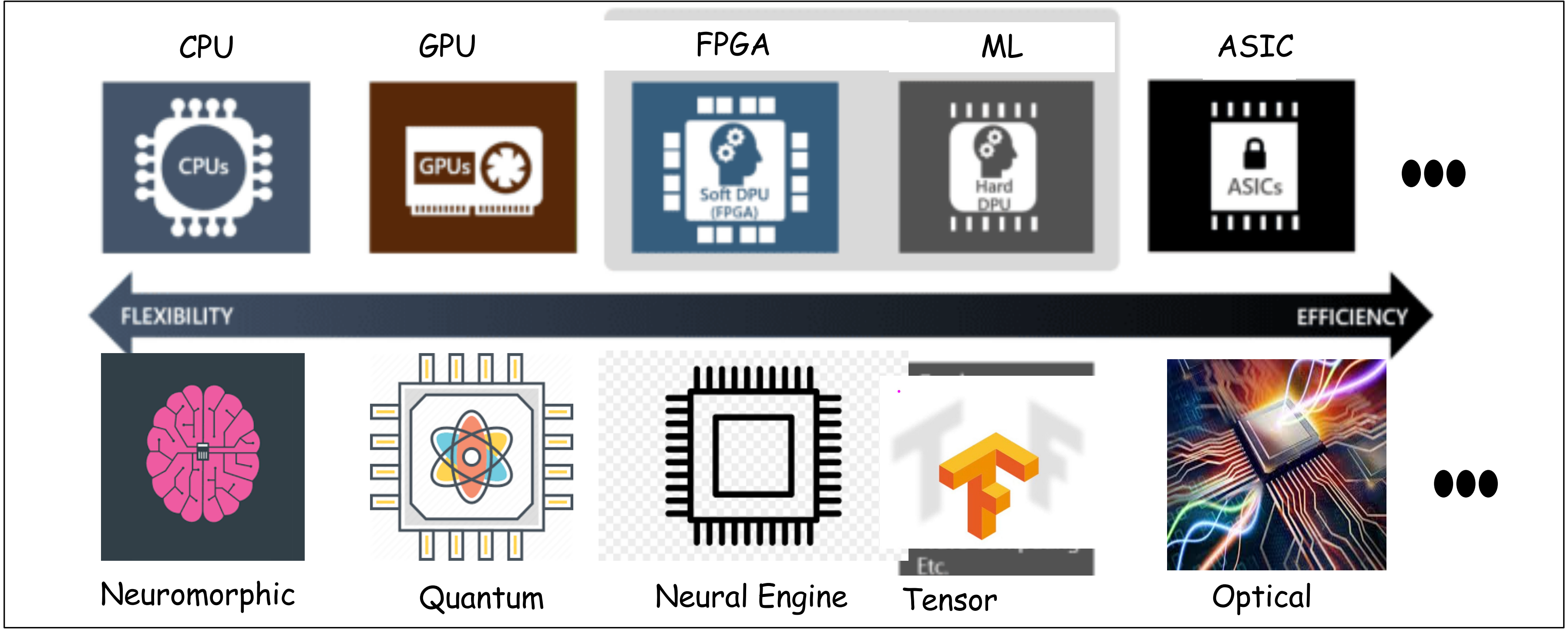
Going Forward What Will Systems Look Like?

- HPC will have extreme heterogeneity and build custom systems for each important application.
- **See this today with Apple iPhone X**
  - iPhone X
    - **A12 processor, 7nm,**
      - 4+2 core 64-bit ARM CPU
      - 4 cores GPU
      - 8 cores Neural Engine
    - **Accelerometer/Gyroscope**
    - **Compass**
    - **Barometric Press sensor**
    - **Audio Codec**
    - **NFC Controller**
    - **Touch Display**

- **2 camera modules**
- **IR camera**
- **Floor Illuminator**
- **Dot projector**
- **Light sensor**
- **RF chipset**
  - LTE modem
  - Baseband processor
  - RF transceiver
- **MCU**

Future HPC Systems Will be Customized…

- You will be able to dial up what you need in your computer for your application mix …



2020 TOP500 Highlights

- Fugaku is the new #1 on the TOP500, 19% of whole list!
- It measured at over 1 Exaflop on the HPL-AI in reduced precision
- TOP10 has four new systems
- Overall turn-over in the list is at a record low
  - Only 51 systems, has been as high as 300
- Research System and Commercial Systems show very different markets