CSUN | CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Tech Series

DR JEFF SOFTWARE
INDIE APP DEVELOPER

Rev 12-30-24

HPC/Super

High Performance & Super Computers

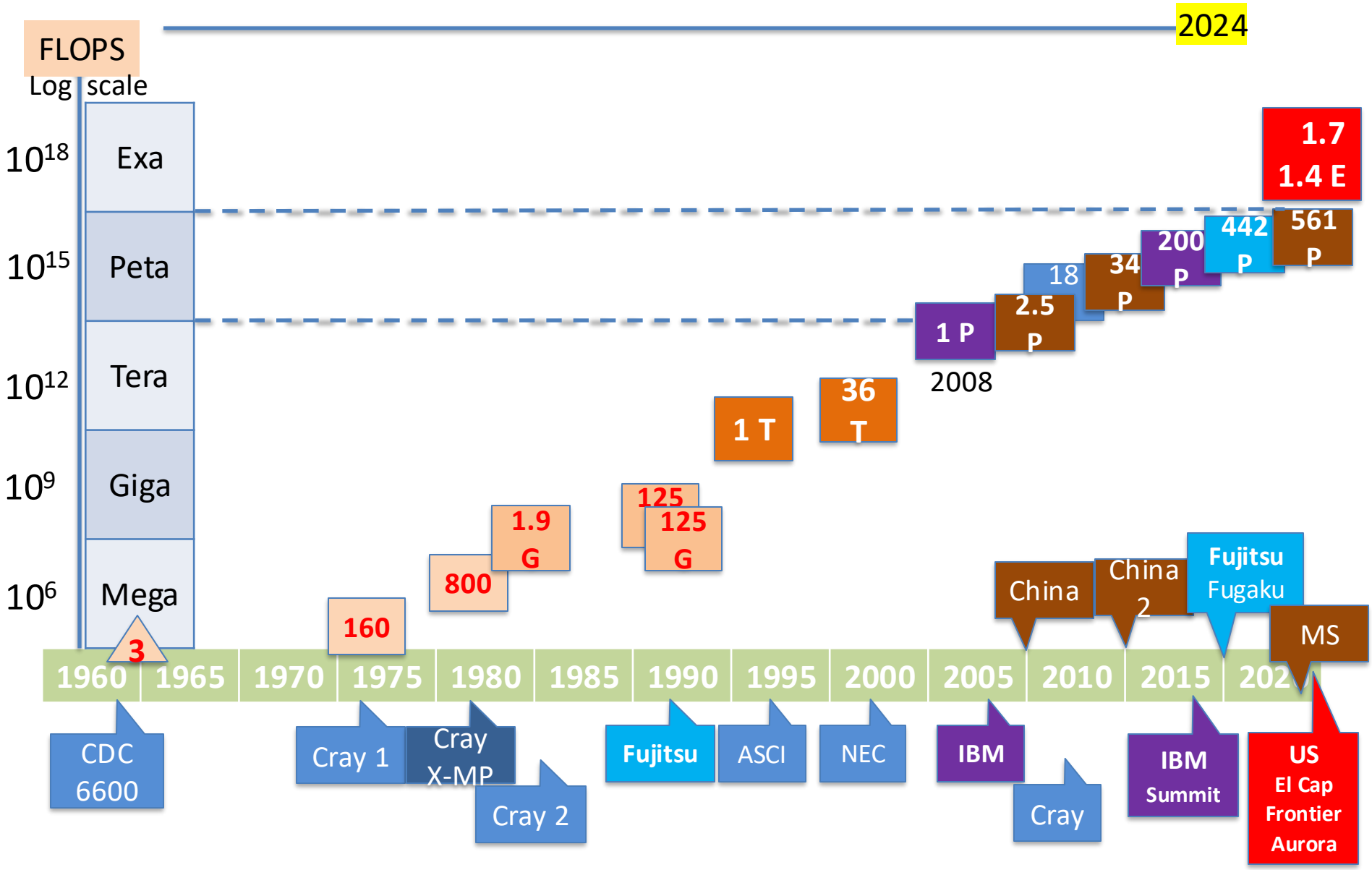
By
Dr Jeff Drobman

website → drjeffsoftware.com/classroom.html

email → jeffrey.drobman@csun.edu

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

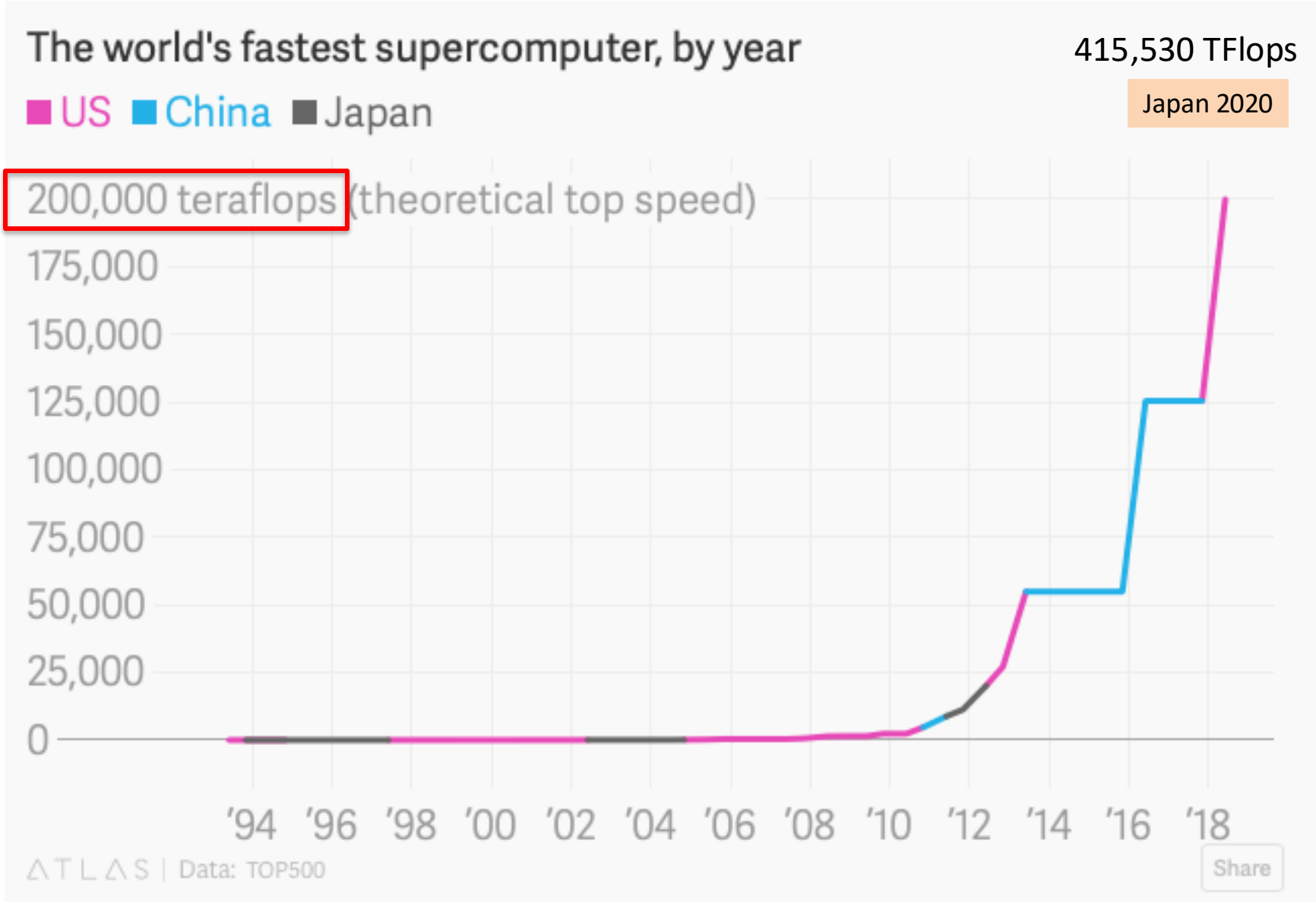
Supercomputer Timeline



CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Supercomputers

DR JEFF SOFTWARE
INDIE APP DEVELOPER



CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Top 10 Supercomputers

2024

DR JEFF SOFTWARE
INDIE APP DEVELOPER

Rank	System	Cores	(PFlop/s)	(PFlop/s)	(kW)
1	**El Capitan** - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS, HPE DOE/NNSA/LLNL United States	11,039,616	1,742.00	2,746.38	29,581
2	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Cray OS, HPE DOE/SC/Oak Ridge National Laboratory United States	9,066,176	1,353.00	2,055.72	24,607
3	**Aurora** - HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11, Intel DOE/SC/Argonne National Laboratory United States	9,264,128	1,012.00	1,980.01	38,698
4	**Eagle** - Microsoft NDv5, Xeon Platinum 8480C 48C 2GHz, NVIDIA H100, NVIDIA Infiniband NDR, Microsoft Azure Microsoft Azure United States	2,073,600	561.20	846.84	
5	**HPC6** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, RHEL 8.9, HPE Eni S.p.A.	3,143,520	477.90	606.97	8,461

LLNL

ORNL

ANL

Top 10 Supercomputers

2024

6	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	[illegible]	29,899
7	**Alps** - HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11, HPE Cray OS, HPE Swiss National Supercomputing Centre (CSCS) Switzerland	2,121,600	434.90	574.84	7,124
8	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,752,704	379.70	531.51	7,107
9	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, EVIDEN EuroHPC/CINECA Italy	1,824,768	241.20	306.31	7,494
10	**Tuolumne** - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS, HPE DOE/NNSA/LLNL United States	1,161,216	208.10	288.88	3,387

Top Supercomputers

Giga (10^9) → Tera (10^12) → **Peta** (10^15)

Summit and Sierra remain in the top two spots. Both are IBM-built supercomputers employing Power9 CPUs and NVIDIA Tesla V100 GPUs. Oak Ridge National Laboratory's Summit system holds top honors with an HPL result of 148.6 petaflops. The second-ranked Sierra system at Lawrence Livermore National Laboratory comes in at 94.6 petaflops.

100 **P**flops

Close behind at number three is the Sunway TaihuLight supercomputer, with an HPL mark of 93.0 petaflops. TaihuLight was developed by China's National Research Center of Parallel Computer Engineering & Technology (NRCPC) and is installed at the National Supercomputing Center in Wuxi. It is powered exclusively by Sunway's SW26010 processors.

Tianhe-2A (Milky Way-2A), a system developed by China's National University of Defense Technology (NUDT) and deployed at the National Supercomputer Center in Guangzhou, China, holds the number four spot with 61.4 petaflops. It is powered by Intel Xeon CPUs and Matrix-2000 accelerators."

Top500 List: Top 2

06/2020 Highlights

The new top system, Fugaku, turned in a High Performance Linpack (HPL) result of 415.5 petaflops, besting the now second-place Summit system by a factor of 2.8x. Fugaku, is powered by Fujitsu's 48-core A64FX SoC, becoming the first number one system on the list to be powered by ARM processors. In single or further reduced precision, which are often used in machine learning and AI applications, Fugaku's peak performance is over 1,000 petaflops (1 exaflops). The new system is installed at RIKEN Center for Computational Science (R-CCS) in Kobe, Japan.

The most energy-efficient system on the Green500 is the MN-3, based on a new server from Preferred Networks. It achieved a record 21.1 gigaflops/watt during its 1.62 petaflops performance run. The system derives its superior power efficiency from the MN-Core chip, an accelerator optimized for matrix arithmetic. It is ranked number 395 in the TOP500 list.

In second position is the new NVIDIA Selene supercomputer, a DGX A100 SuperPOD powered by the new A100 GPUs. It occupies position seven on the TOP500.

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Top500 List: Top 10

DR JEFF SOFTWARE
INDIE APP DEVELOPER

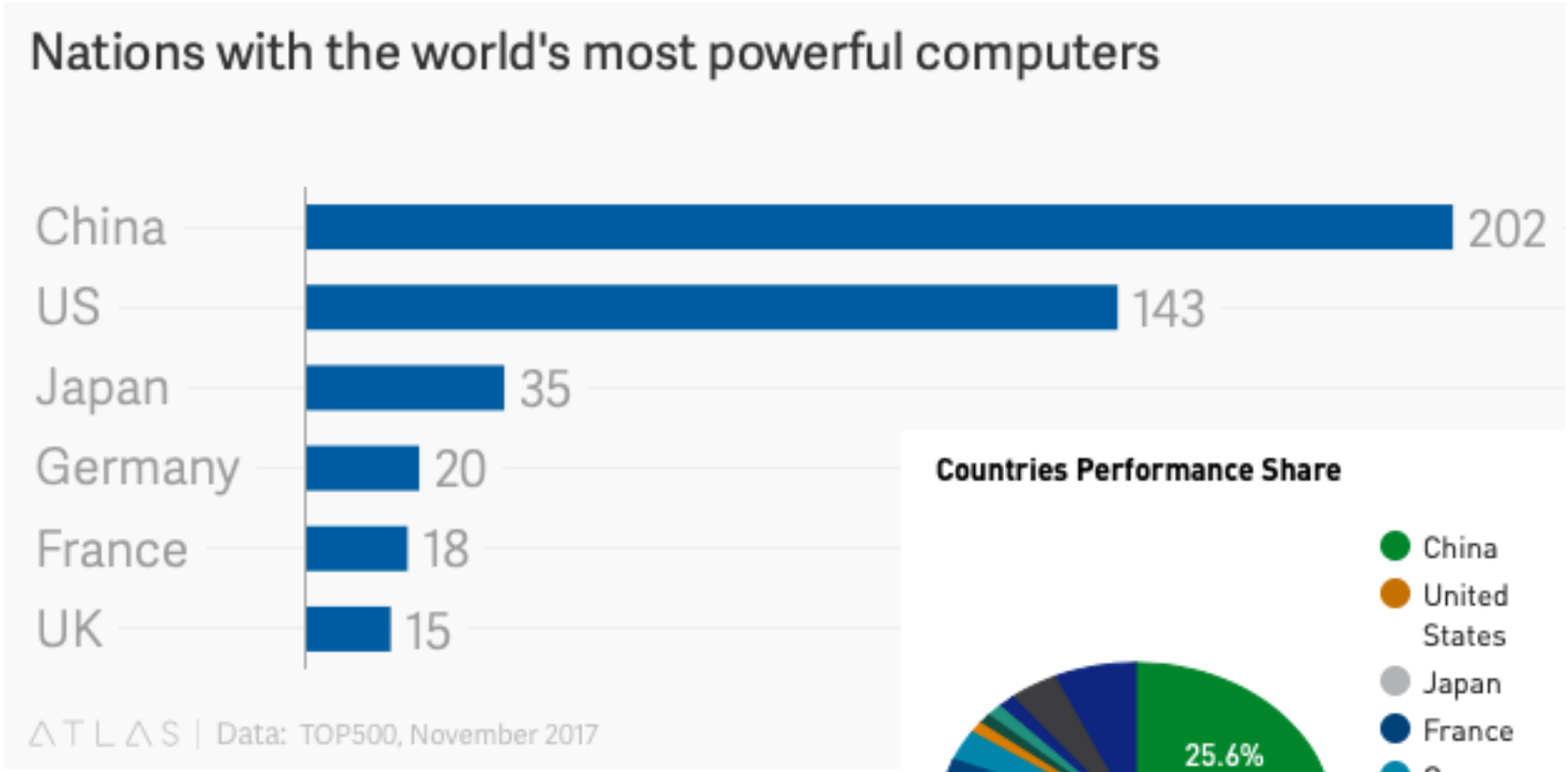
June 2020

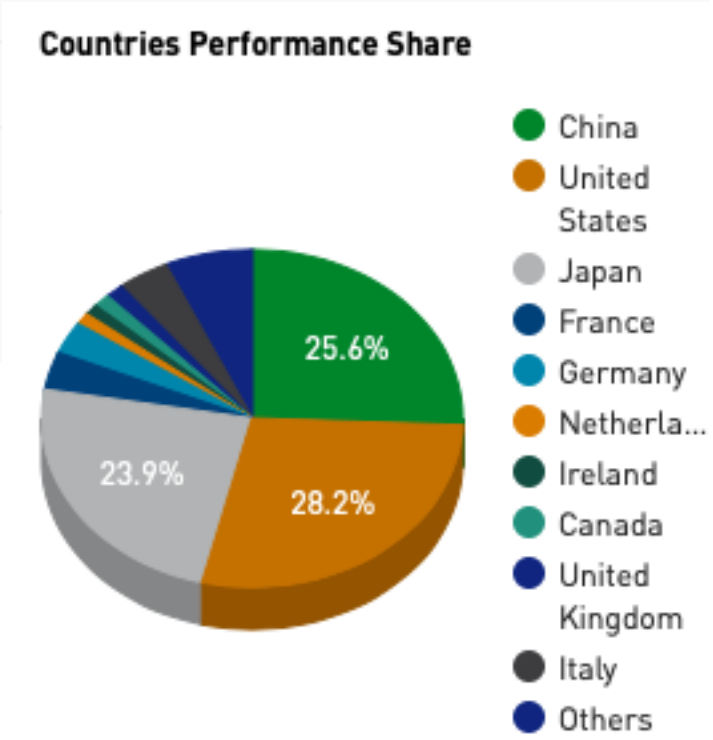
1	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu	ARM
2	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM	IBM
3	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox	IBM
4	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC	
5	**Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000, NUDT	Intel Xeon
6	**HPC5** - PowerEdge C4140, Xeon Gold 6252 24C 2.1GHz, NVIDIA Tesla V100, Mellanox HDR Infiniband, Dell EMC	Intel Xeon
7	**Selene** - DGX A100 SuperPOD, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia	AMD Epyc
8	**Frontera** - Dell C6420, Xeon Platinum 8280 28C 2.7GHz, Mellanox InfiniBand HDR, Dell EMC	Intel Xeon
9	**Marconi-100** - IBM Power System AC922, IBM POWER9 16C 3GHz, Nvidia Volta V100, Dual-rail Mellanox EDR Infiniband, IBM	IBM
10	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100, Cray/HPE	Intel Xeon

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Supercomputers

DR JEFF SOFTWARE
INDIE APP DEVELOPER





Supercomputers

2019

QUARTZ EMAILS EDITIONS BECOM

For the first time in five years, the world's fastest computer is no longer in China.

Yesterday (June 8), the US Department of Energy's Oak Ridge National Laboratory announced the top speeds of its Summit supercomputing machine, which nearly laps the previous record-holder, China's Sunway TaihuLight. The Summit's theoretical peak speed is 200 petaflops, or 200,000 teraflops. To put that in human terms, approximately 6.3 billion people would all have to make a calculation at the same time, every second, for an entire year, to match what Summit can do in just one second. (Another way to see it: if you want to go toe-to-toe with Summit yourself, settle in. You'll be making a calculation every single second for the next 6.3 billion years.)

Supercomputing technology has been improving rapidly in recent years. Just over a decade ago, the world hadn't yet built a machine that could crack even a single petaflop (or 1,000 teraflops). Now, in just a year, we've gone from 125 petaflops to 200.

Super Applications

Frontier's upgraded hardware is the main factor behind that improvement. But hardware alone doesn't do scientists that much good if they don't have software that can harness the machine's new oomph. That's why an initiative called the Exascale Computing Project (ECP)—which brings together the Department of Energy and its National Nuclear Security Administration, along with industry partners—has sponsored 24 initial science-coding projects alongside the supercomputers' development.

Those software initiatives can't just take old code—meant to simulate, say, the emergence of sudden severe weather—plop it onto Frontier and say, "It made an okay forecast at lightning speed instead of almost lightning speed!" To get a more accurate result, they need an amped-up and optimized set of codes. "We're not going to cheat here and get the same

Super Applications

Astronomy

Another early Frontier project called ExaStar, led by Daniel Kasen of Lawrence Berkeley National Laboratory, will investigate a different cosmic mystery. This endeavor will simulate supernovae—the end-of-life explosions of massive stars that, in their extremity, produce heavy elements. Scientists have a rough idea of how supernovae play out, but no one actually knows the whole-cow version of these explosions or how heavy elements get made within them.

In the past, most supernova simulations simplified the situation by assuming stars were spherically symmetric or by using simplified physics. With exascale computers, scientists can make more detailed three-dimensional models. And rather than just running the code for one explosion, they can do whole suites, including different kinds of stars and different physics ideas, exploring which parameters produce what astronomers actually see in the sky.

Supercomputers

Machines

El Capitan

The **El Capitan computer** refers to a **supercomputer** being developed by the **U.S. Department of Energy (DOE)** and is set to be installed at **Lawrence Livermore National Laboratory (LLNL)**.

Purpose:

•El Capitan is designed for tasks such as **nuclear stockpile management** to ensure the safety and reliability of the U.S. nuclear arsenal without the need for physical testing.

•It will also contribute to broader scientific research, including advanced simulations in health, materials science, and physics.

Performance:

•It is expected to achieve **2+ exaFLOPS** of performance, meaning it can perform over two quintillion calculations per second.

•This makes it one of the most powerful supercomputers globally, surpassing other exascale systems like *Frontier* at Oak Ridge National Laboratory.

CSUN
CALIFORNIA
STATE UNIVERSITY
NORTHRIDGE

El Capitan

DR JEFF
SOFTWARE
INDIE APP DEVELOPER

The **El Capitan computer** refers to a **supercomputer** being developed by the **U.S. Department of Energy (DOE)** and is set to be installed at **Lawrence Livermore National Laboratory (LLNL)**.

Technology:

- Built in collaboration with **Hewlett Packard Enterprise (HPE)** and using **AMD's EPYC processors and Instinct GPUs**, it will leverage cutting-edge hardware for its unprecedented computing power.
- It incorporates advanced AI and machine learning capabilities.

Deployment Timeline:

- Originally scheduled for 2023, delays pushed its expected operational date to **2024**.

Aurora Super

intel newsroom

You have probably heard about Argonne National Laboratory, which will be deploying more than 60,000 Max Series GPUs and 20,000 Max Series CPUs to power the Aurora supercomputer this year. Aurora is expected to become the world's first supercomputer with 2 exaflops of peak performance. Deployment is going well, with Intel collaborating closely on testing and development. Argonne expects the system to be accessible to early researchers by the third quarter of 2023.

Lawrence Livermore National Laboratories (LLNL) and Sandia National Laboratories are installing thousands of nodes of 4th Gen Intel Xeons in their CTS-2 systems – the supercomputing workhorse of the Department of Energy (DOE). LLNL's Intel Xeon-powered predecessor, JADE, recently contributed to the breakthrough in fusion energy, helping to design the optimal package for laser induction.

Los Alamos National Laboratory (LANL), another DOE research center, is installing more than 10,000 Max Series CPUs for its Crossroads supercomputer, which will power national security and wildfire research.

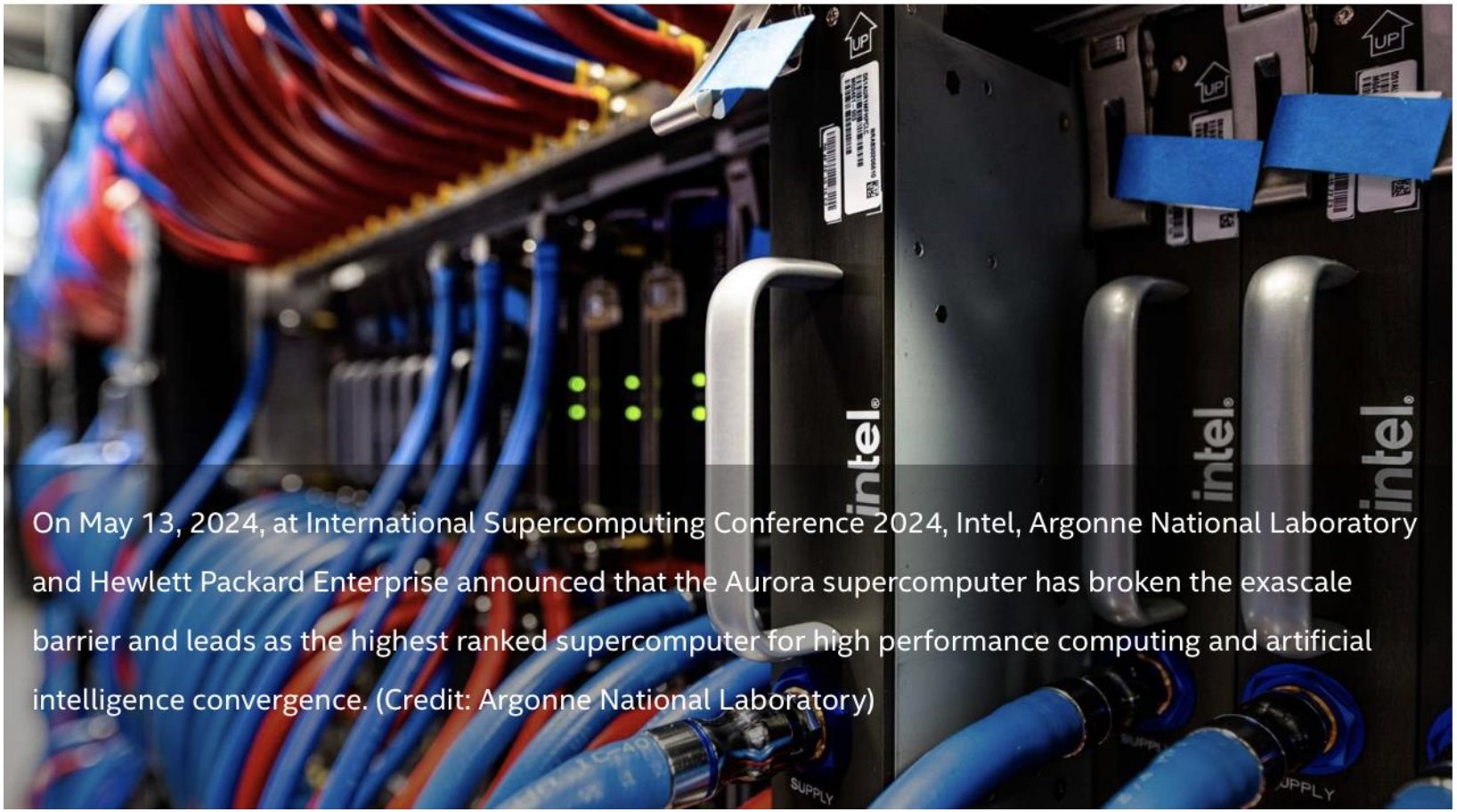
CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

DR JEFF SOFTWARE
INDIE APP DEVELOPER

Aurora Super

intel newsroom

May 13, 2024



On May 13, 2024, at International Supercomputing Conference 2024, Intel, Argonne National Laboratory and Hewlett Packard Enterprise announced that the Aurora supercomputer has broken the exascale barrier and leads as the highest ranked supercomputer for high performance computing and artificial intelligence convergence. (Credit: Argonne National Laboratory)

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

DR JEFF SOFTWARE
INDIE APP DEVELOPER

Aurora Super

intel newsroom

May 2024

What's New: At ISC High Performance 2024, Intel announced in collaboration with Argonne National Laboratory and Hewlett Packard Enterprise (HPE) that the Aurora supercomputer has broken the exascale barrier at 1.012 exaflops and is the fastest AI system in the world dedicated to AI for open science, achieving 10.6 AI exaflops. Intel will also detail the crucial role of open ecosystems in driving AI-accelerated high performance computing (HPC).

> *"The Aurora supercomputer surpassing exascale will allow it to pave the road to tomorrow's discoveries. From understanding climate patterns to unraveling the mysteries of the universe, supercomputers serve as a compass guiding us toward solving truly difficult scientific challenges that may improve humanity."*
>
> –Ogi Brkic, Intel vice president and general manager of Data Center AI Solutions

Aurora Super

intel newsroom

May 2024

How AI is Optimized: At the heart of the Aurora supercomputer is the Intel Data Center GPU Max Series. The Intel X^{e} GPU architecture is foundational to the Max Series, featuring specialized hardware like matrix and vector compute blocks optimized for both AI and HPC tasks. The Intel X^{e} architecture's design that delivers unparalleled compute performance is the reason the Aurora supercomputer secured the top spot in the high-performance LINPACK-mixed precision (HPL-MxP) benchmark – which best highlights the importance of AI workloads in HPC.

The X^{e} architecture's parallel processing capabilities excel in managing the intricate matrix-vector operations inherent in neural network AI computation. These compute cores are pivotal in accelerating matrix operations crucial for deep learning models. Complemented by Intel's suite of software tools, including Intel® oneAPI DPC++/C++ Compiler, a rich set of performance libraries, and optimized AI frameworks and tools, the X^{e} architecture fosters an open ecosystem for developers that is characterized by flexibility and scalability across various devices and form factors.

Aurora Super

intel newsroom

May 2024

Aurora Supercomputer's Details: The Aurora supercomputer is an expansive system with 166 racks, 10,624 compute blades, 21,248 Intel® Xeon® CPU Max Series processors and 63,744 Intel® Data Center GPU Max Series units, making it one of the world's largest GPU clusters. Aurora also includes the largest open, Ethernet-based supercomputing interconnect on a single system of 84,992 HPE slingshot fabric endpoints. Aurora supercomputer came in second on the high-performance LINPACK (HPL) benchmark but broke the exascale barrier at 1.012 exaflops utilizing 9,234 nodes, only 87% of the system. Aurora supercomputer also secured the third spot on the high-performance conjugate gradient (HPCG) benchmark at 5,612 teraflops per second (TF/s) with 39% of the machine. This benchmark aims to assess more realistic scenarios providing insights into communication and memory access patterns, which are important factors in real-world HPC applications. It complements benchmarks like LINPACK by offering a comprehensive view of a system's

Aurora Super

intel newsroom

May 2024

What's Next: New supercomputers being deployed with Intel Xeon CPU Max Series and Intel Data Center GPU Max Series technologies underscore Intel's goal to advance HPC and AI. Systems include Euro-Mediterranean Centre on Climate Change's (CMCC) Cassandra to accelerate climate change modeling; Italian National Agency for New Technologies, Energy and Sustainable Economic Development's (ENEA) CRESCO 8 to enable breakthroughs in fusion energy; Texas Advanced Computing Center (TACC), which is in full production to enable data analysis in biology to supersonic turbulence flows and atomistic simulations on a wide range of materials; as well as United Kingdom Atomic Energy Authority (UKAEA) to solve memory-bound problems that underpin the design of future fusion powerplants.

The result from the mixed-precision AI benchmark will be foundational for Intel's next-generation GPU for AI and HPC, code-named Falcon Shores. Falcon Shores will leverage the next-generation Intel Xe architecture with the best of Intel® Gaudi®. This integration enables a unified programming interface.

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

New Top Super: Frontier

DR JEFF SOFTWARE
INDIE APP DEVELOPER

Quora 273 606 Search Quora

Currently, as of the time of writing this (August 2022), the world's fastest computer is the HP Frontier supercomputer, owned by Oak Ridge National Laboratory.

Supercomputers

Quora

Franklin Veaux · Follow
Professional Writer · Aug 23

How much RAM does the world's fastest computer have?

Currently, as of the time of writing this (August 2022), the world's fastest computer is the HP Frontier supercomputer, owned by Oak Ridge National Laboratory.

Frontier has 4,849,664 gigabytes (4,000 terabytes) of RAM and and 47,360 terabytes of NAND flash storage.

Frontier

SCIENTIFIC
AMERICAN.

New Exascale Supercomputer Can Do a Quintillion Calculations a Second

New "exascale" supercomputers will bring breakthroughs in science. But the technology also exists to study nuclear weapons

By Sarah Scoles on February 9, 2023

Frontier can process seven times faster and hold four times more information in memory than its predecessors. It is made up of nearly 10,000 CPUs, or central processing units—which perform instructions for the computer and are generally made of integrated circuits—and almost 38,000 GPUs, or graphics processing units. GPUs were created to quickly and smoothly display visual content in gaming. But they have been reappropriated for scientific computing, in part because they're good at processing information in parallel.

Frontier

SCIENTIFIC
AMERICAN®

Inside Frontier, the two kinds of processors are linked. The GPUs do repetitive algebraic math in parallel. "That frees the CPUs to direct tasks faster and more efficiently," Kothe says. "You could say it's a match made in supercomputing heaven." By breaking scientific problems into a billion or more tiny pieces, Frontier allows its processors to each eat their own small bite of the problem. Then, Kothe says, "it reassembles the results into the final answer. You could compare each CPU to a crew chief in a factory and the GPUs to workers on the front line."

The 9,472 different nodes in the supercomputer—each essentially its own not-so-super computer—are also all connected in such a way that they can pass information quickly from one place to another. Importantly, though, Frontier doesn't just run faster than machines of yore: it also has more memory and so can run bigger simulations and hold tons of information in the same place it's processing those data. That's like keeping all the acrylics with you while you're trying to do a paint-by-numbers project rather than having to go retrieve each color as needed from the other side of the table.

Fugaku Supercomputer

Quora

The Fugaku in Japan became the fastest supercomputer in the world last year, at one point able to hit 2.0 ExaFLOPS with its 158,979 48-core ARMv8.2 CPU's.

So, this works out to 7,630,848 individual ARM cores to "brute force" just about any task imaginable.

The Fugaku can do a documented **442 FP64 PetaFLOPS**.

The "miniaturized" version of this might be this NVIDIA DGX A100 SuperPOD with 24 DGX modules each featuring sixteen A100 GPUs and a pair of 64-core EPYC CPUs.

This "miniature" supercomputer puts out a peak of **3.74 FP64 PetaFLOPS** and uses about 268,000 watts. At a cost of over nine million dollars, it wouldn't even be 1% as fast as Japan's behemoth.

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Frontier (vs Fugaku)

DR JEFF SOFTWARE
INDIE APP DEVELOPER

Rank (previous)	Rmax Rpeak (PetaFLOPS)	Name	Model	CPU cores	Accelerator (e.g. GPU) cores	Interconnect
1 –	1,102.00 1,685.65	Frontier	HPE Cray EX235a	591,872 (9,248 × 64-core Optimized 3rd Generation EPYC 64C @2.0 GHz)	36,992 × 220 AMD Instinct MI250X	Slingshot-11
2 –	442.010 537.212	Fugaku	Supercomputer Fugaku	7,630,848 (158,976 × 48-core Fujitsu A64FX @2.2 GHz)	0	Tofu interconnect D

Leapfrog: Fugaku

2020

48 core SoC

The world's new fastest supercomputer is named Fugaku, powered by Fujitsu's 48-core A64FX SoC. It is installed at the RIKEN Center for Computational Science (R-CCS) in Kobe, Japan. It beat the previous performance champion, the US Department of Energy's Summit system -- which topped the Top500 list twice in a row -- by a factor of 2.8x, the non-profit Top500 organization announced on June 22, 2020.

2.8x Summit

Supercomputer Fugaku
Located: RIKEN Center for Computational Science (R-CCS) in Kobe, Japan

Processor: A64FX 48C 2.2GHz
Cores: 7,299,072
Memory: 4,866,048 GB
Interconnect: Tofu interconnect D

415 peta FLOPS

Linpack performance: 415,530 TFlop/s
Power consumption: 28.3 MW

28.3 MegaWatts!

Leapfrog: Fugaku

2020

Fugaku

28.3 MegaWatts!

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Supercomputer Pix

DR JEFF SOFTWARE
INDIE APP DEVELOPER

Quora

Andrew McGregor · Follow
Performance Measurement Lead at Fastly (company) (2019–present) · Upvot
Gaurav Saxena, MSc High Performance Computing, University of Edinburgh ·

What does a super computer look like?

From the outside:

Looks like a big industrial building with a lot of power and cooling… because that's what it is. Depending on how we look at it, that's one machine, 17 machines, 33 machines, or a classified number that might be a few hundred thousand (I didn't look it up).

Supercomputer Pix

Quora

Andrew McGregor · Follow
Performance Measurement Lead at Fastly (company) (2019–present) · Upvot
Gaurav Saxena, MSc High Performance Computing, University of Edinburgh

What does a super computer look like?

Each tray in the racks on the right is a very big PC-like server and some hard drives.

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

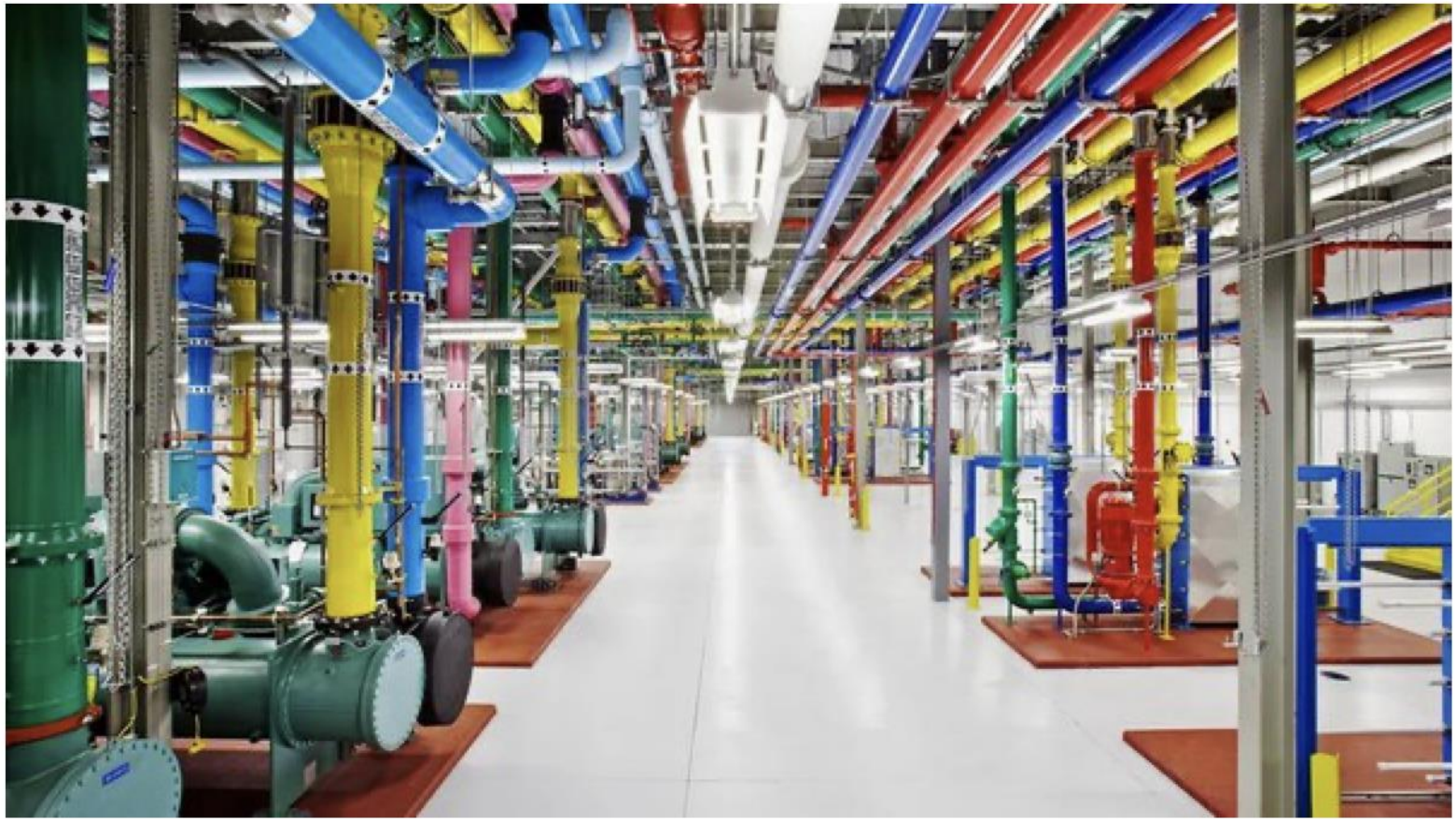
Supercomputer Pix

DR JEFF SOFTWARE
INDIE APP DEVELOPER

Quora

Andrew McGregor · Follow
Performance Measurement Lead at Fastly (company) (2019–present) · Upvot
Gaurav Saxena, MSc High Performance Computing, University of Edinburgh ·

What does a super computer look like?



Cooling. For scale, this space is triple-height, so those big green things on the left are about shoulder-high to a tall man.

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Supercomputer Pix

DR JEFF SOFTWARE
INDIE APP DEVELOPER

Quora

Andrew McGregor · Follow
Performance Measurement Lead at Fastly (company) (2019–present) · Upvot
Gaurav Saxena, MSc High Performance Computing, University of Edinburgh ·

What does a super computer look like?

That's the network interface. All the yellow stuff? Fiber optic cable, and trays to carry it.

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Argonne NL *Polaris*

News

ADVANCED MICRO DEVICES (AMD) (*111.32 -0.08*) *I*

AMD's EPYC Processors Picked for Argonne National Laboratory's Polaris Supercomputer

11:07 AM EDT, 08/30/2021 (MT Newswires) -- Advanced Micro Devices (AMD) said Monday the US Department of Energy's Argonne National Laboratory will use AMD's EPYC processors in its new Polaris supercomputer. Built by Hewlett Packard Enterprise ... (MT Newswires 11:07 AM ET 08/30/2021)

Argonne National Laboratory

From Wikipedia, the free encyclopedia

Argonne National Laboratory is a science and engineering research national laboratory operated by UChicago Argonne LLC for the United States Department of Energy. The facility is located in Lemont, Illinois, outside of Chicago, and is the largest national laboratory by size and scope in the Midwest.

Argonne had its beginnings in the Metallurgical Laboratory of the University of Chicago, formed in part to carry out Enrico Fermi's work on nuclear reactors for the Manhattan Project during World War II. After the war, it was designated as the first

Argonne NL *Intrepid*

The IBM Blue Gene/P supercomputer "Intrepid" at Argonne National Laboratory runs 164,000 processor cores using normal data center air conditioning, grouped in 40 racks/cabinets connected by a high-speed 3D torus network.[1][2]

IBM Summit Super

Summit

summit

Sponsors	U.S. Department of Energy
Operators	IBM
Architecture	9,216 POWER9 22-core CPUs 27,648 NVIDIA Tesla V100 GPUs[1]
Power	13 MW[2]
Operating system	Red Hat Enterprise Linux (RHEL)[3][4]
Storage	250 PB
Speed	200 petaFLOPS (peak)
Purpose	Scientific research
Web site	www.olcf.ornl.gov/olcf-resources/compute-systems/summit/

Summit components

IBM Summit Super

Summit (supercomputer)

From Wikipedia, the free encyclopedia

Summit or **OLCF-4** is a supercomputer developed by IBM for use at Oak Ridge National Laboratory, capable of 200 petaFLOPS, making it the second fastest supercomputer in the world (it held the number 1 position from November 2018 to June 2020.[5][6]) Its current LINPACK benchmark is clocked at 148.6 petaFLOPS.[7] As of November 2019, the supercomputer is also the 5th most energy efficient in the world with a measured power efficiency of 14.668 gigaFLOPS/watt.[8] Summit was the first supercomputer to reach exaflop (a quintillion operations per second) speed, achieving 1.88 exaflops during a genomic analysis and is expected to reach 3.3 exaflops using mixed-precision calculations.[9]

History [edit]

The United States Department of Energy awarded a $325 million contract in November 2014 to IBM, NVIDIA and Mellanox. The effort resulted in construction of Summit and Sierra. Summit is tasked with civilian scientific research and is located at the Oak Ridge National Laboratory in Tennessee. Sierra is designed for nuclear weapons simulations and is located at the Lawrence Livermore National Laboratory in California.[10] Summit is estimated to cover the space of about 55 meters and require 219 kilometers of cabling.[11] Researchers will utilize Summit for diverse fields such as cosmology, medicine and climatology.[12]

In 2015, the project called Collaboration of Oak Ridge, Argonne and Lawrence Livermore (CORAL) included a third supercomputer named Aurora and was planned for installation at Argonne National Laboratory.[13] By 2018, Aurora was re-engineered with completion anticipated in 2021 as an exascale computing project along with Frontier and El Capitan to be completed shortly thereafter.[14]

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

IBM Summit Super

- IBM Power9 CPU
- Nvidia Tesla GPU

Each one of its 4,608 nodes (with 2 IBM POWER9 CPUs and 6 Nvidia Tesla GPUs in each node[17]) has over 600 GB of coherent memory (96 GB HBM2 plus 512 GB DDR4 SDRAM) which is addressable by all CPUs and GPUs plus 800 GB of non-volatile RAM that can be used as a burst buffer or as extended memory.[18] The POWER9 CPUs and Nvidia Volta GPUs are connected using NVIDIA's high speed NVLink. This allows for a heterogeneous computing model.[19] To provide a high rate of data throughput, the nodes will be connected in a non-blocking fat-tree topology using a dual-rail Mellanox EDR InfiniBand interconnect for both storage and inter-process communications traffic which delivers both 200Gb/s bandwidth between nodes and in-network computing acceleration for communications frameworks such as MPI a

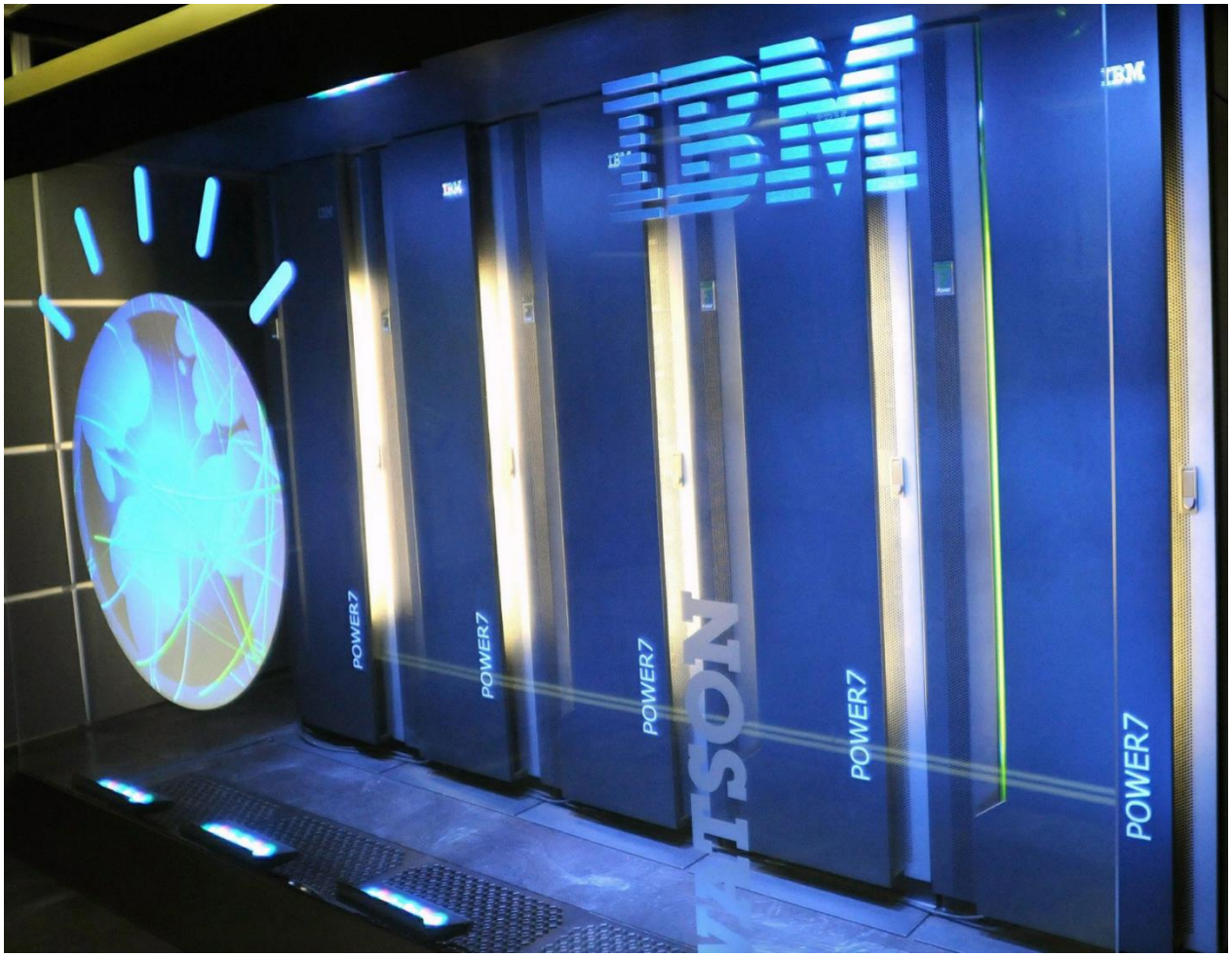
Summit components

- DDR4 SDRAM

POWER9 wafer with TOP500 certificates for Summit and Sierra

Supercomputer IBM Watson

2011



Supercomputer IBM Watson

Update Aug 19, 2021

The Watson they built was a roomsize supercomputer with thousands of processors running millions of lines of code. Its storage disks were filled with digitized reference works, Wikipedia entries and electronic books. Computing intelligence is a brute force affair, and the hulking machine required 85,000 watts of power. The human brain, by contrast, runs on the equivalent of 20 watts.

IBM says it has 40,000 Watson customers across 20 industries worldwide, more than double the number four years ago. Watson products and services are being used 140 million times a month, compared with a monthly rate of about 10 million two years ago, IBM says. Some of the big customers are in health, like Anthem, a large insurer, which uses Watson Assistant to automate customer inquiries.

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

New $100M Beast

12-30-24

SCIENCE OF SUCCESS

The Giant Supercomputer Built to Transform an Entire Country—and Paid For by Ozempic

The world's latest AI machine is powered by the success of two products: Nvidia's chips and Novo Nordisk's weight-loss drugs

Gefion

Jensen Huang, Nadia Carlsten and King Frederik X of Denmark ceremonially plug in Gefion. EPA-EFE/SHUTTERSTOCK/ SHUTTERSTOCK

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

New $100M Beast

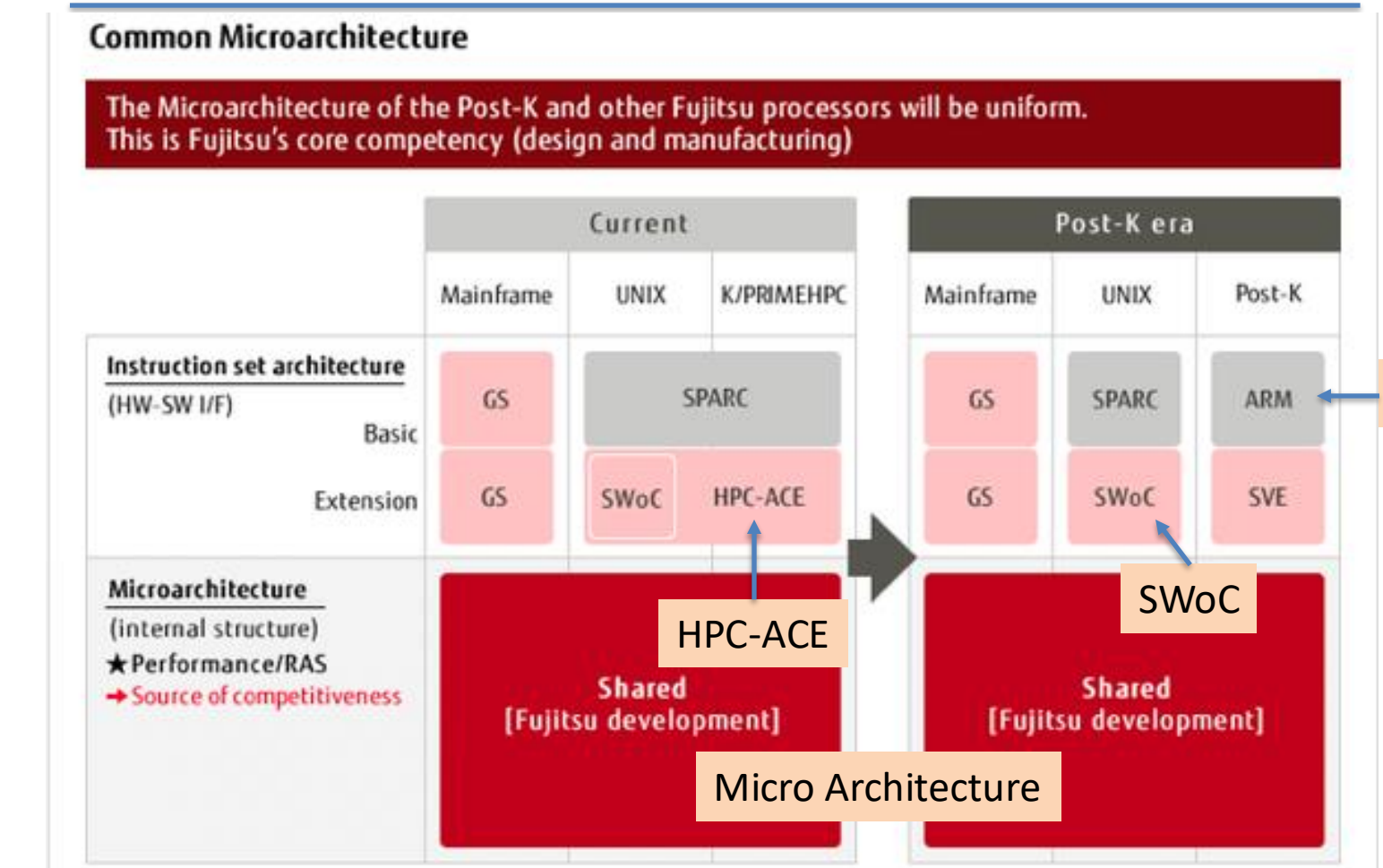
"In time, you'll discover that it's not a data center," Huang said at the supercomputer's unveiling. "It's a factory of intelligence."

A model of the Gefion supercomputer at Copenhagen's airport. The actual machine is bigger than a basketball court. PHOTO: MADS CLAUS RASMUSSEN/ZUMA PRESS

Whatever you call it, Gefion is a beast. It is bigger than a basketball court. It weighs more than 30 tons. It took six months to manufacture and install. It also required an investment of $100 million.

That funding came from a public-private initiative between the Novo Nordisk Foundation and the state-owned Export and Investment Fund of Denmark.

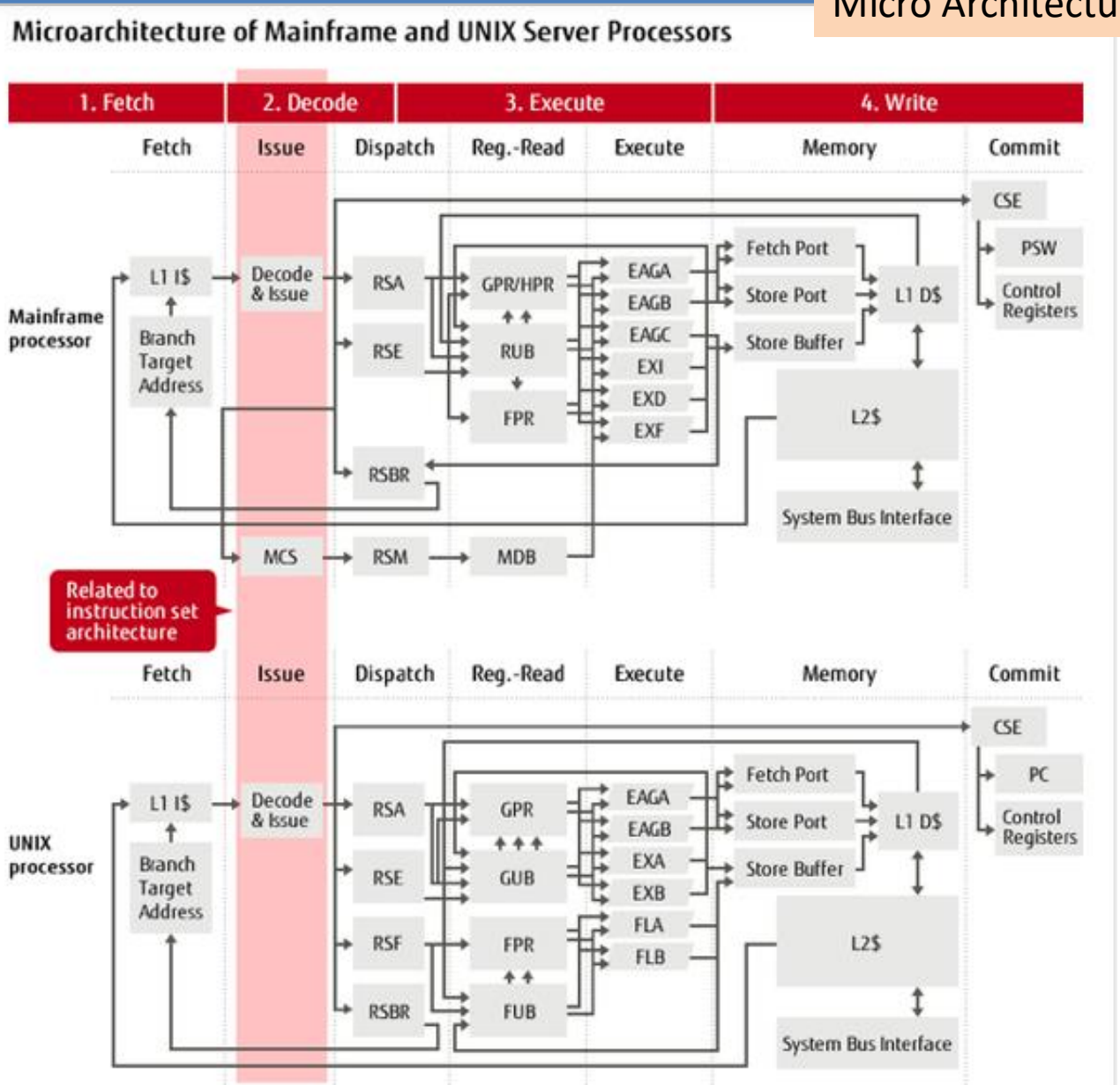
Fujitsu Sparc SoC



Notice how they're sliding in an ARM in place of SPARC, but they're using a common microarchitecture underneath.

Fujitsu Sparc SoC

Micro Architecture



CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Supercomputing Pi

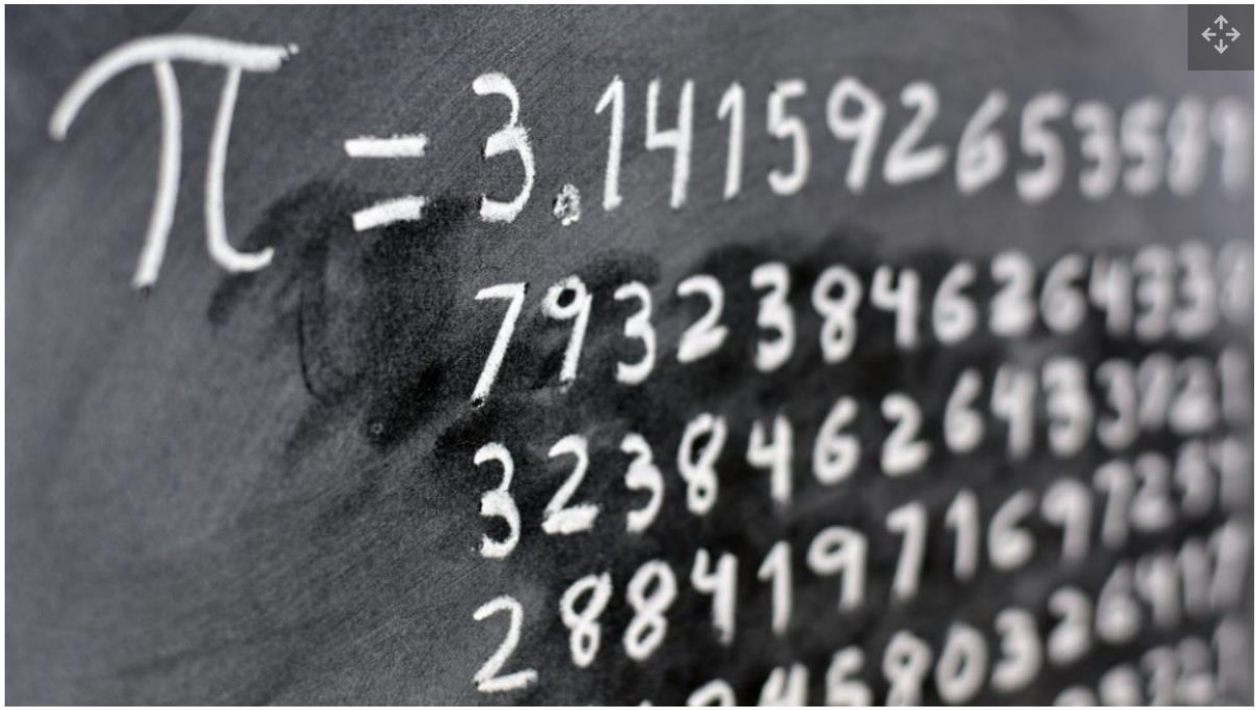
DR JEFF SOFTWARE
INDIE APP DEVELOPER

Aug 18, 2021

Pi calculated to a record-breaking 62.8 trillion digits

By Harry Baker - Staff Writer 3 days ago

Supercomputer took 108 days to run the calculations.



Pi is an irrational number, meaning it has an infinite number of decimal points. (Image credit: Shutterstock)

Supercomputing Pi

A Supercomputer Just Calculated Pi to a Record 62.8 Trillion Digits

Aug 18, 2021

Caroline Delbert
Wed, August 18, 2021, 10:25 AM · 4 min read

Knowing more digits of pi isn't particularly important for mathematics.

But calculating the value of pi to high precision has long been used as a benchmark to test the processing power of computers. In 2019, a Google cloud computing system calculated the constant's value to more than 31 trillion decimal places, and in 2020, Timothy Mullican of Huntsville, Alabama, founder of a nonprofit called North Alabama Charitable Computing, calculated 50 trillion decimal places, using his personal computer, according to Guinness World Records.

The DAViS team not only broke Mullican's record but also did so in roughly a third of the time —

taking just 108 days and 9 hours, compared with Mullican's 303 days — even though they used the same algorithm to run the calculations.

Supercomputers vs Clusters

Lawrence Stewart, I've been building and programming supercomputers since 2004

Answered 4h ago

My understanding is that Fugaku operates as a cluster, but is not built that way.

What I mean is that the software treats the machine as independent servers interconnected by a communications system (a cluster) but the communications system is very tightly integrated with the processors and hardware, so it is not built out of independent boxes wired together.

The SC series machines from my dear departed SiCortex were like that too. Software thought is was a cluster, but the hardware was highly integrated.

I am aware of any very large HPC systems that are not, in software at least, clusters. The hardware is often much more highly integrated because it is cheaper that way, and you can build a better interconnect than by plugging cards into the PCI.

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Tesla AI (Dojo SC)

DR JEFF SOFTWARE
INDIE APP DEVELOPER

Chinese QC/Optical

Quora

Search Quora

On the other hand, this is a rather limited "quantum computer." And one can imagine that it took an army of graduate students to keep all the optics tweaked up.

Jeff Drobman

Just now

hmmm. seems to me this is an "optical computer", not a quantum one, and is not programmable, so not universal.

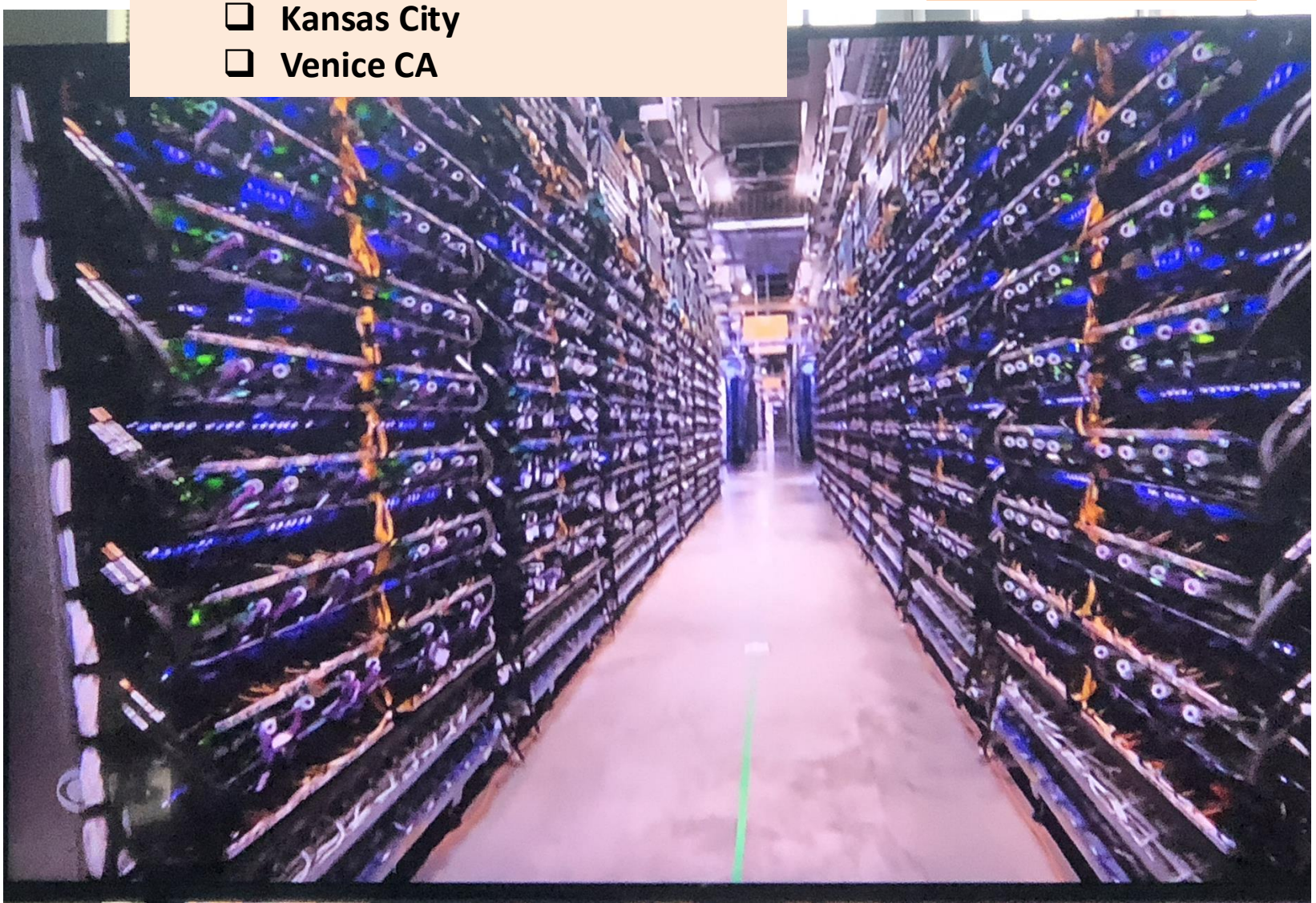
Section

Data Centers: Server Farms /Clusters

Google

- Google has several **Data Centers**
  - **Portland, OR**
  - **Kansas City**
  - **Venice CA**

- 70,000+ servers



Key Facts

Key Facts

- The Citadel located in Nevada is the largest server farm on the globe covering 7.2 million square feet. Its reliance on renewable energy sources sets it apart from several of its competitors.
- Switch SuperNAP is the third largest. It covers 3.5 million square feet and caters to tech giants such as Amazon, Google, and Microsoft.
- Utah Data Center covers 1.5 million square feet and caters to government agencies such as the NSA.

Top 10

Summary of The Largest Server Farms Ever

Rank	Farm
#1	The Citadel – 7.2 million sq. ft.
#2	Range International Information Group – 6.3 million sq. ft.
#3	Switch SuperNAP – 3.5 million sq. ft.
#4	DFT Data Center – 1.6 million sq. ft
#5	Utah Data Center 1.5 million sq. ft
#6	Microsoft Data Center – 1.2 million sq. ft.
#7	Lakeside Technology Center – 1.1 million sq. ft.
#8	Tulip Data Center – 1 million sq. ft.
#9	QTS Metro Data Center – 990,000 sq. ft.
#10	Next Generation Data Europe – 750,000 sq. ft.

Top Server Farm

1. The Citadel – 7.2 million sq. ft.

This farm, located in Tahoe Reno, Nevada, is set to take the title of the largest server farm ever built, boasting an impressive 7.2 million square feet of space.

Although much of its operations remain secretive, it's known that the data center processes massive quantities of electronic communications, ranging from emails to phone calls, for surveillance and intelligence analysis purposes. The enormous size of the plant allows for sophisticated systems and backup generators to keep the facility running continuously.

But, the size isn't the only thing that sets the Citadel apart. The data center will also be powered exclusively by renewable energy sources. Located near the Tesla Gigafactory, the Citadel's sustainability focus fits well within the industrial park's push for green technology.

Other Top Farms

3. Switch SuperNAP – 3.5 million sq. ft.

This farm in Las Vegas, Nevada, is one of the world's most extensive data center facilities. Spanning a massive 3.5 million square feet, the facility hosts server farms for some of the biggest companies on the internet, including Amazon, Google, and Microsoft.

What sets it apart from other data centers is its incredible power capacity — with 430 megawatts at its disposal, it can handle more than ten times the amount of energy an average data center uses!

5. Utah Data Center – 1.5 million sq. ft

This data center sprawls over 1.5 million square feet of space and is designed to han exabytes of data. It serves as a hub for server farms, where information is collected a stored by various government agencies, including the NSA.

6. Microsoft Data Center – 1.2 million sq. ft.

In West Des Moines, Iowa, Microsoft Data Center is a massive facility covering 1.2 million square feet of commercial space. Once completed, the total data center space owned by Microsoft will reach 6.3 million square feet. Like many data centers, this facility consists of large server farms with enormous storage and processing power for Microsoft's services and products.

Clusters

If you could join computers together, is that like having one supercomputer?

Jeff Drobman
Lecturer at California State University, Northridge (2016–present) · Just now ·

computers have been "joined together" ever since (and before) Ethernet was invented 50 years ago to form a "LAN" network. and one could argue ever since the Internet was first launched as a WAN called ARPANET in 1969 (although access was limited to research then). but that is networking with a "client-server" model.

a "supercomputer" today is a tightly coupled "cluster" of 10,000 to 80,000 "computers" — typically a multi-chip module from AMD called "Epyc" or from Intel called "Xeon". they consume MW of power and cost upwards of $10,000,000. they take weeks if not months to construct. (HPE has been doing the construction for the US' latest supers at national labs.)

Section

HPC

Epyc

AMD

March 2021

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Epyc

AMD

March 2021

CSUN
CALIFORNIA
STATE UNIVERSITY
NORTHRIDGE

AMD Epyc 7003

DR JEFF SOFTWARE
INDIE APP DEVELOPER

3-15-21

News Story

AMD Rolls Out Epyc 7003 Series Processors; Shares Rise

3:15 PM 3/15/2021 - MT Newswires

03:15 PM EDT, 03/15/2021 (MT Newswires) -- Advanced Micro Devices (AMD) said Monday that it has rolled out its new Eypc 7003 series CPUs.Shares of the Santa Clara, California-based company were up 1.8% in late-day trading.Price: 82.34, Change: +1.29, Percent Change: +1.59

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Epyc 7003

AMD

March 2021

AMD EPYC™ 7003 FAMILY

AVAILABLE TODAY

CORES	MODEL
64 CORES	7763 7713/P
56 CORES	7663
48 CORES	7643
32 CORES	75F3 7543/P 7513
28 CORES	7453
24 CORES	74F3 7443/P 7413
16 CORES	73F3 7343 7313/P
8 CORES	72F3

AMD EPYC

ALL-IN FEATURE SET

8-Channels of DDR4-3200
4TB Memory Capacity
128 Lanes PCIe® 4
Infinity Guard Security Features
Socket Compatible

Epyc Zen

AMD

March 2021

OUR "ZEN" JOURNEY

"ZEN" / "ZEN+"

- ~52% IPC
- 4-core complex
- 8MB L3 per complex
- 14nm/12nm
- Simultaneous multithread
- SEV

"ZEN 2"

- ~15% IPC
- 4-core complex
- 16MB L3 per complex
- 7nm
- Chiplet design
- FP-256
- SEV-ES

"ZEN 3"

- ~19% IPC
- New core layout
- New cache topology
- 7nm
- Doubled INT8 throughput
- New security features: Shadow Stack & SEV-SNP

2017 — 2020

Exascale HPC

https://insidehpc.com/2012/03/the-international-race-to-exascale/

The International Race to Exascale

March 5, 2012 by Rich Brueckner

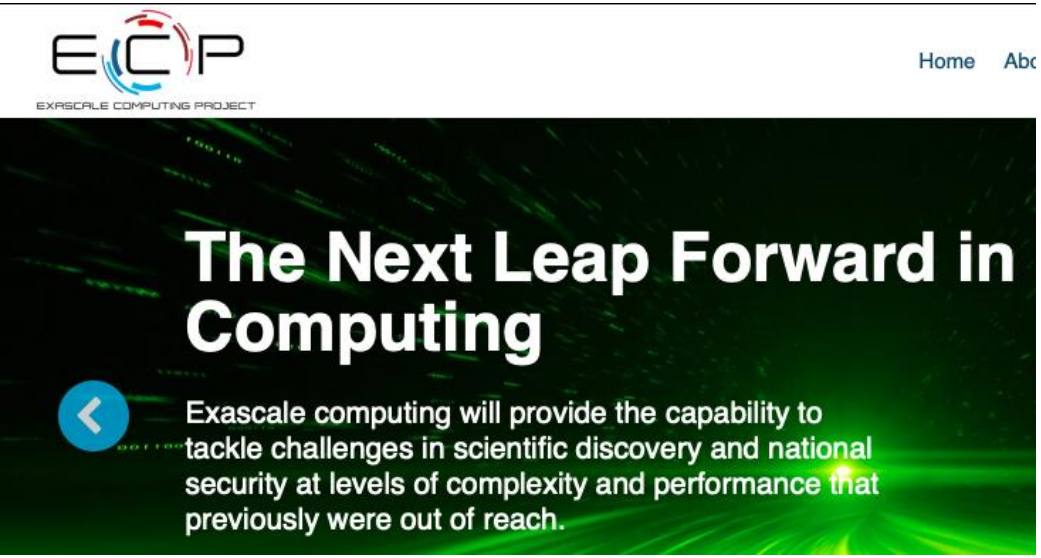
In this special feature from The Exascale Report, Mike Bernh the race while Russia continues its silent and steady march.

Two of the most interesting competitors in the exascale race are also the most quiet. The government of India has committed close to $ 1 billion (USD) toward an advanced supercomputing program with exascale written all over it. The actual details of the plan have not been disclosed, not even to many of the country's top scientists. This is the largest research program ever funded in India, and yet it is being handled in deep secrecy.

Exa = 10^18

Exascale HPC

https://www.exascaleproject.org



Daisy Marie, Web Developer
Answered 8h ago

EuroHPC: Europe/Sweden

The EuroHPC Joint Undertaking is an important initiative in Europe and for Swedish research. The JU's aim is to acquire by 2020 world-class precursor to exascale supercomputers – machines that are capable of 10/17 calculations per second – for Europe's scientific, industrial and public users.

Near Exa = 10^17

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

DR JEFF SOFTWARE
INDIE APP DEVELOPER

PEZY-SC2

Supercomputer Chip

PEZY-SC2 (**PEZY Super Computer 2**) is a third generation many-core microprocessor developed by PEZY and introduced in early 2017. This chip, which operates at 1 GHz, incorporates 2,048 cores dissipating 180 W. The PEZY-SC2 powers the ZettaScaler-2.x series of supercomputers.

PEZY-SC2

General Info

Designer	PEZY
Manufacturer	TSMC
Model Number	PEZY-SC2
Market	Supercomputer
Introduction	2015 (announced) 2017 (launched)

General Specs

Family	PEZY-SCx
Frequency	1,000 MHz

Microarchitecture

Process	16 nm
Technology	CMOS
Die	620 mm²
Cores	2,048
Threads	16,384

Electrical

Power dissipation	180 W
power dissipation (average)	130 W
V_{core}	0.8 V

Introduced by PEZY along with their second-generation ZettaScaler-2.0 supercomputer series, the SC2 incorporates 2,048 cores along with 8-way SMT support for a total of 16,384 threads, twice as many cores as its predecessor. The PEZY-SC2 powers many of the top Green500 most efficient supercomputers with upward of 14 GFLOPS/watt in performance.

CSUN
CALIFORNIA STATE UNIVERSITY NORTHRIDGE

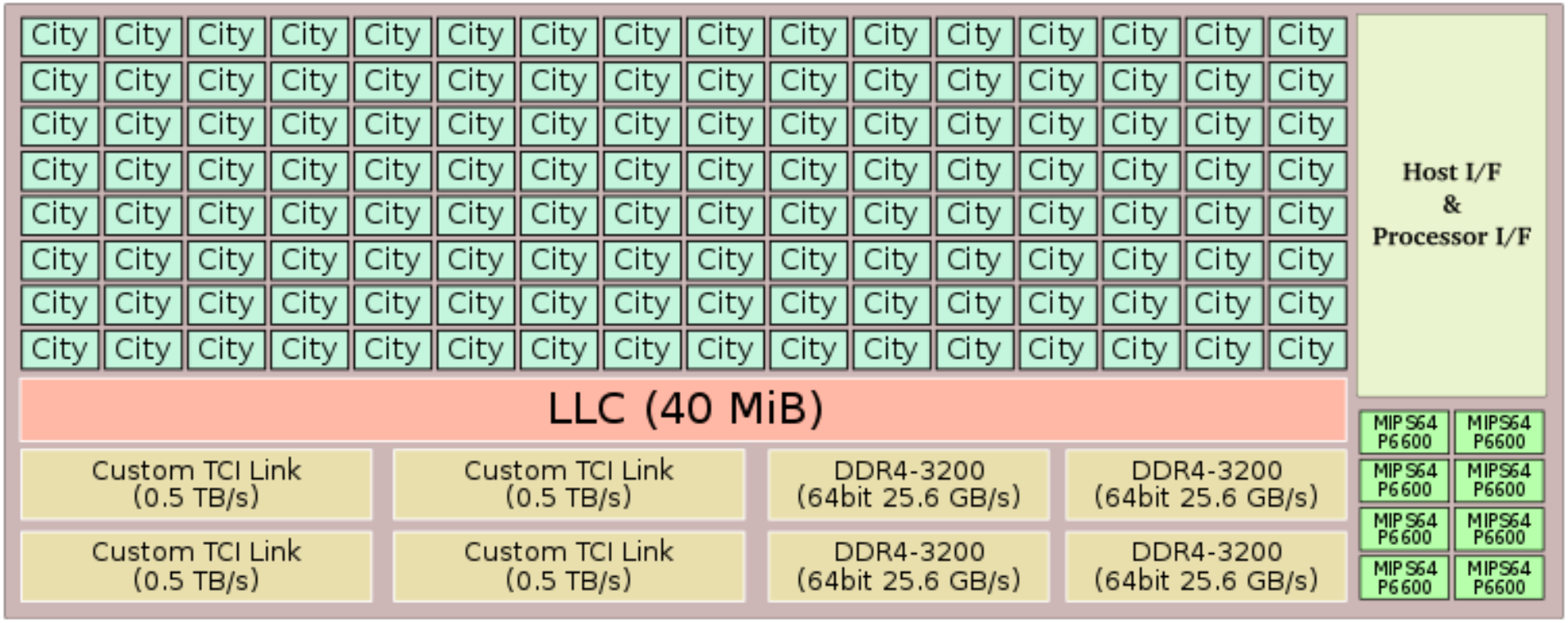
PEZY-SC2

Supercomputer Chip

In attempt to increase adaptability in the field of deep learning and AI as well as to increase throughput for specialized workloads, the PEZY-SC2 introduced support for 16-bit half precision floating point support. At 1 GHz, the SC2 can peak at 16.4 TFLOPS for half precision.

Architecture [edit]

Managing the tiny PEZY cores are six P-Class P6600 MIPS (MIPS64R6) processors. Previously, the PEZY-SC relied on two lightweight ARM926 cores that proved to be too much of a performance bottleneck. The SC2 got rid of the four "Prefecture" units that incorporated 256 cities along with 2 MiB of L3 cache. Instead, the SC2 now has 40 MiB of shared last level cache shared not only by all the cities, but also by the MIPS cores. In order to improve performance further, the MIPS cores and the PEZY cores now share the same address space, reducing data transfer overhead. It's worth noting that the use of powerful MIPS cores mean they no longer require to rely on an external Intel Xeon E5 host processor.



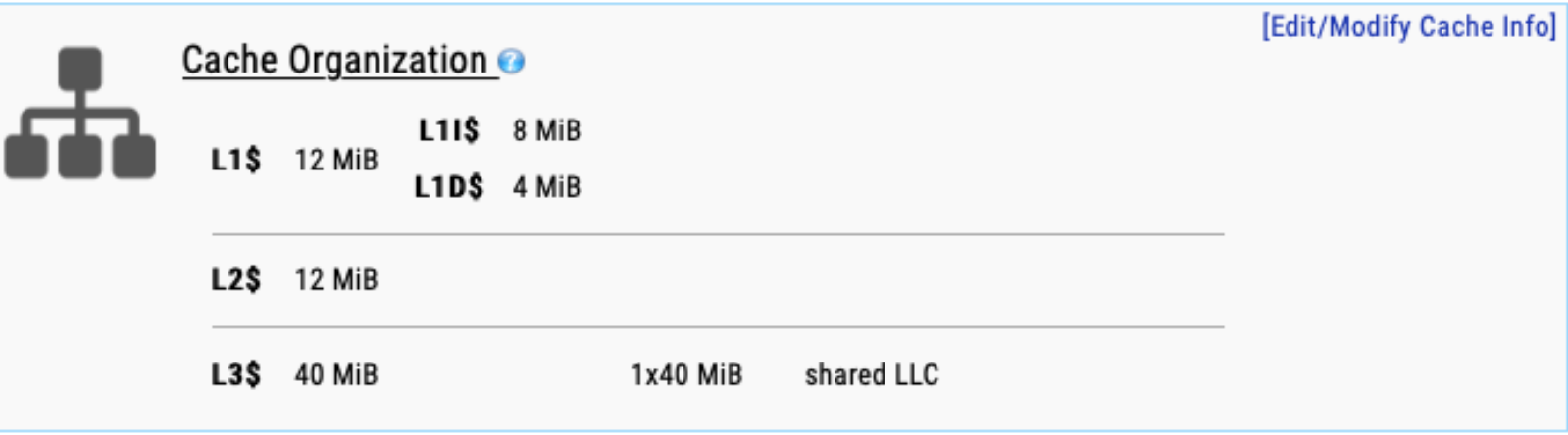
PEZY-SC2

Supercomputer Chip

Cache [edit]

The SC2 a hexa-core MIPS P6600 process which has its own separate cache:

The SC2 integrates a multi-level cache hierarchy:



Additionally, there is another 40 MiB consisting of 20 KiB per PE of scratch pad memory. This was increased from 16 KiB in the Pezy-SC.

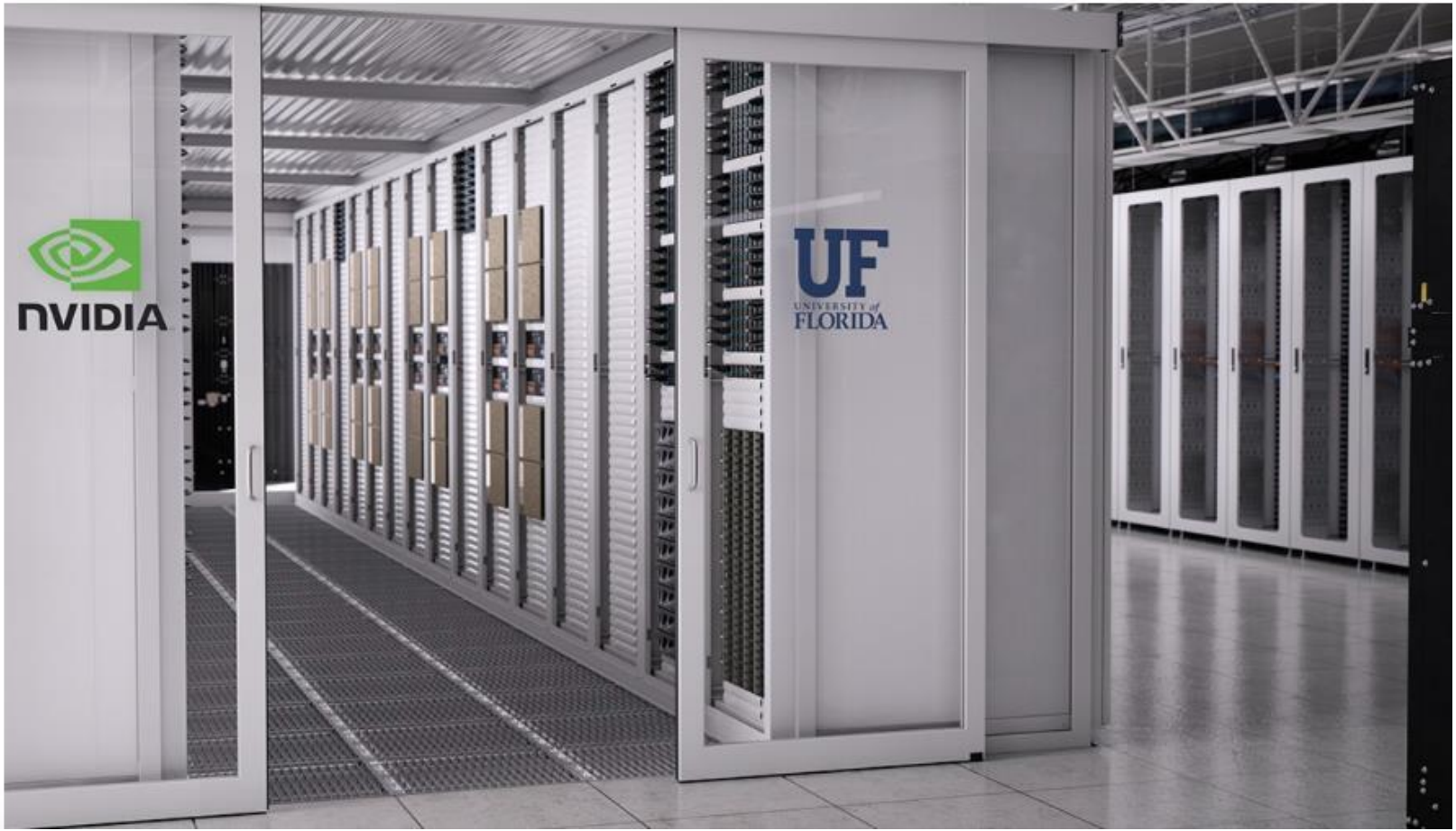
GPU-Based AI

U of Florida

University of Florida, NVIDIA to Build Fastest AI Supercomputer in Academia

Effort will infuse AI throughout UF's curriculum and help address wide range of challenges across the state.

July 21, 2020 by BRIAN CAULFIELD



GPU-Based AI

U of Florida

Working closely with NVIDIA, UF will boost the capabilities of its existing supercomputer, HiPerGator, with the recently announced NVIDIA DGX SuperPOD architecture. The system will be up and running by early 2021, just a few weeks after it's delivered.

This gives faculty and students within and beyond UF the tools to apply AI across a multitude of areas to address major challenges such as rising seas, aging populations, data security, personalized medicine, urban transportation and food insecurity. UF expects to create 30,000 AI-enabled graduates by 2030.

GPU-Based AI

U of Florida

The University of Florida and NVIDIA Tuesday unveiled a plan to build the world's fastest AI supercomputer in academia, delivering 700 petaflops of AI performance.

The effort is anchored by a $50 million gift: $25 million from alumnus and NVIDIA co-founder Chris Malachowsky and $25 million in hardware, software, training and services from NVIDIA.

"We've created a replicable, powerful model of public-private cooperation for everyone's benefit," said Malachowsky, who serves as an NVIDIA Fellow, in an online event featuring leaders from both the UF and NVIDIA.

UF will invest an additional $20 million to create an AI-centric supercomputing and data center.