

Only academic

# HPC Infrastructure in Japan

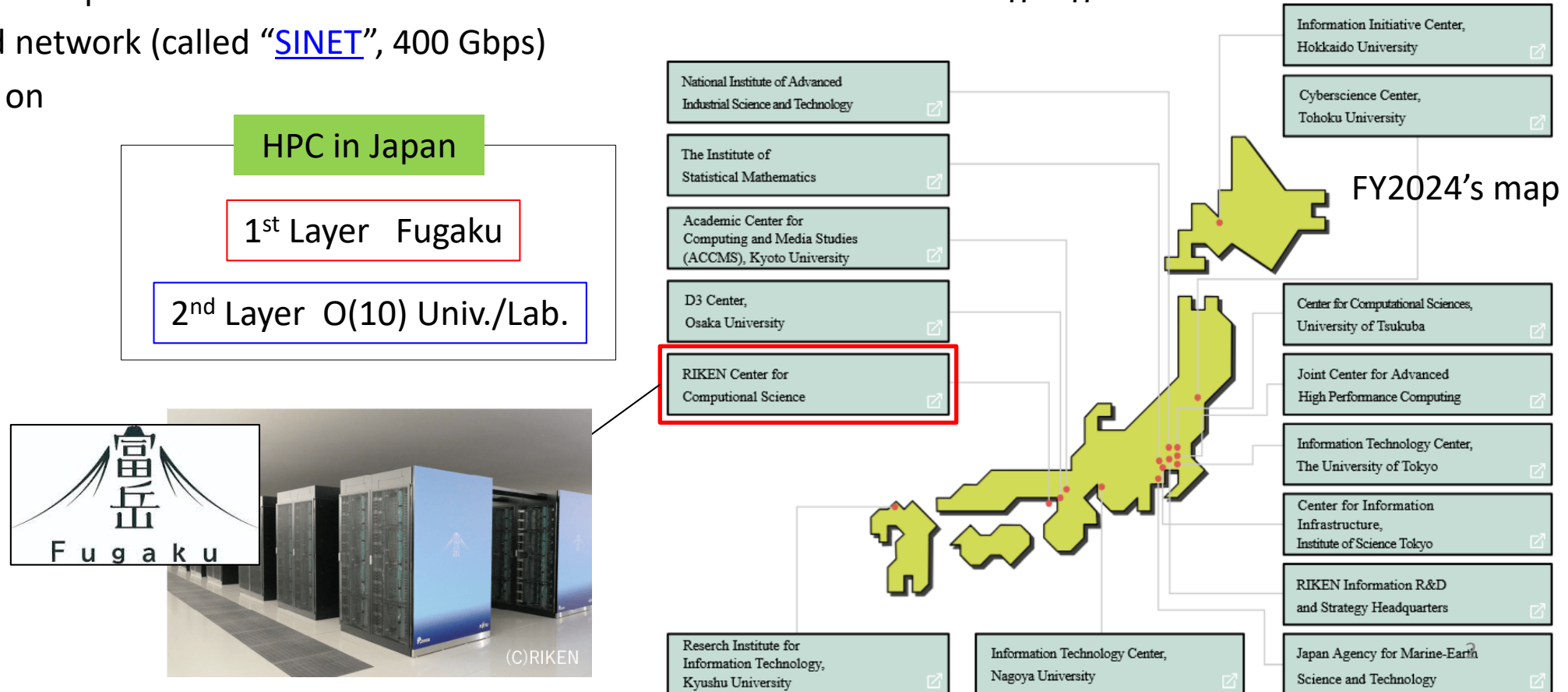
Junichi Tanaka

International Center for Elementary Particle Physics (ICEPP), The University of Tokyo

**“HEP/HPC Strategy Meeting – All Regions” at CERN**

# HPC (“Supercomputer”) for Academic in Japan

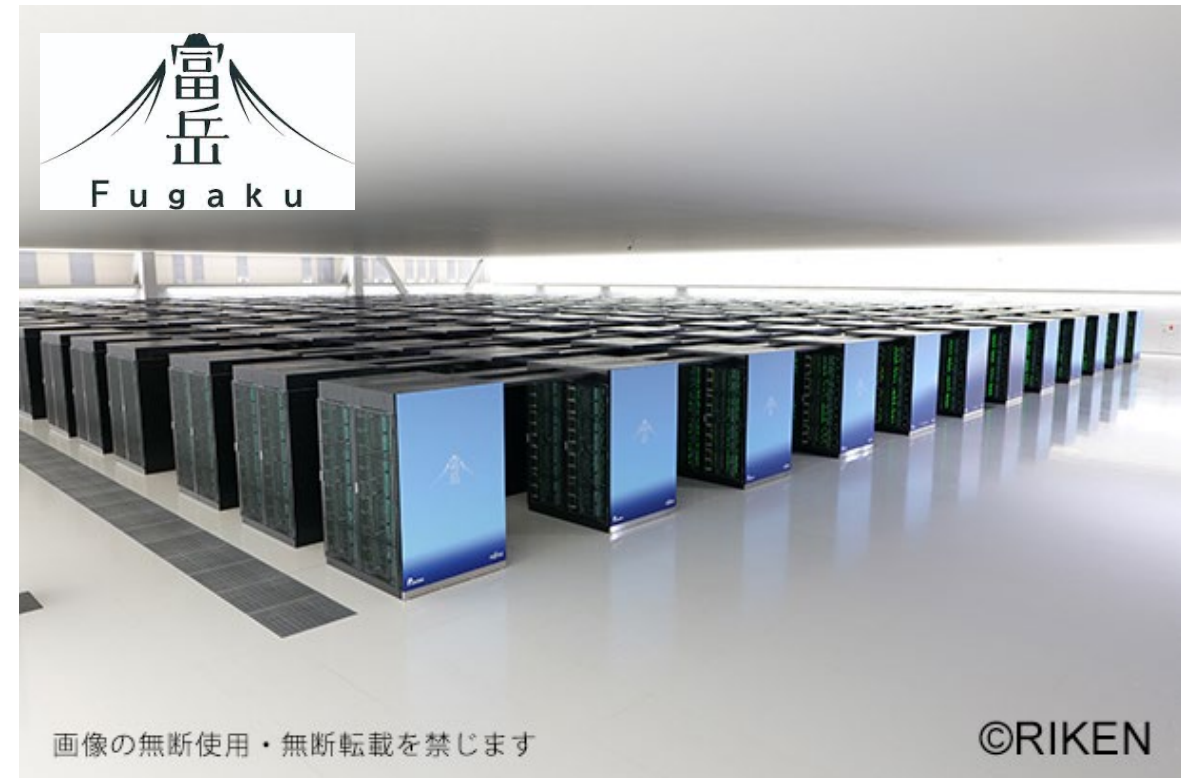
- No clear definition of the HPC, the so-called supercomputer
- Some universities and research centers in Japan have HPC, whose spec is **roughly 5%-10% of Fugaku**.
- One of services is called **High-Performance Computing Infrastructure: HPCI** ( <https://www.hpci-office.jp/en> ), which provides researchers with computing and storage resources:
  - Major supercomputers of academic research centers and universities including Fugaku
  - High-speed network (called “[SINET](#)”, 400 Gbps)
  - Single-sign on



# Fugaku (1<sup>st</sup> Layer)

- Fugaku began operation in March 2021 as a public use.
  - The **1<sup>st</sup> ranking** in TOP500 (June 2020 to Nov 2021): 442 Peta Flops in TOP500
  - Japan's national flagship supercomputer (US \$0.7 billion with 155JPY/\$)
- Processor **A64FX** (an original processor developed with Fujitsu)
  - Instruction set: **Armv8.2-A SVE** (Scalable Vector Extension) 512bit
- 158,976 nodes (~400 racks)
  - 48 cores/node
  - Memory on Chip: 32 GB/node
- Performance
  - **488 Peta Flops (theoretical peak)** for 64bit
  - Fugaku is faster than “Kei” (the previous flagship HPC in Japan) by ~40 but the energy consumption (30MW) is only by ~2.
- **Still the 6<sup>th</sup> ranking in TOP500 (Nov 2024).**  
**The 1<sup>st</sup> ranking in HPCG (Nov 2024).**

<https://www.r-ccs.riken.jp/en/fugaku/about/>



# HPCs in the 2<sup>nd</sup> Layer

Most of them use x86 (Xeon or EPYC) with NVIDIA GPU (H100, A100, V100 etc.)

Top500  
(show only <300 ranking)  
Ranking (FLOPS)

## HPCI 2<sup>nd</sup> Layer System Roadmap (as 2020 Nov.)

Up to **80 PF (theoretical peak)** system

76 (17PF)

228 (4.3PF)

58 (22PF), 224 (4.4PF)

36 (40PF)

173 (6.6PF), 214 (4.9PF)

167 (6.7PF)

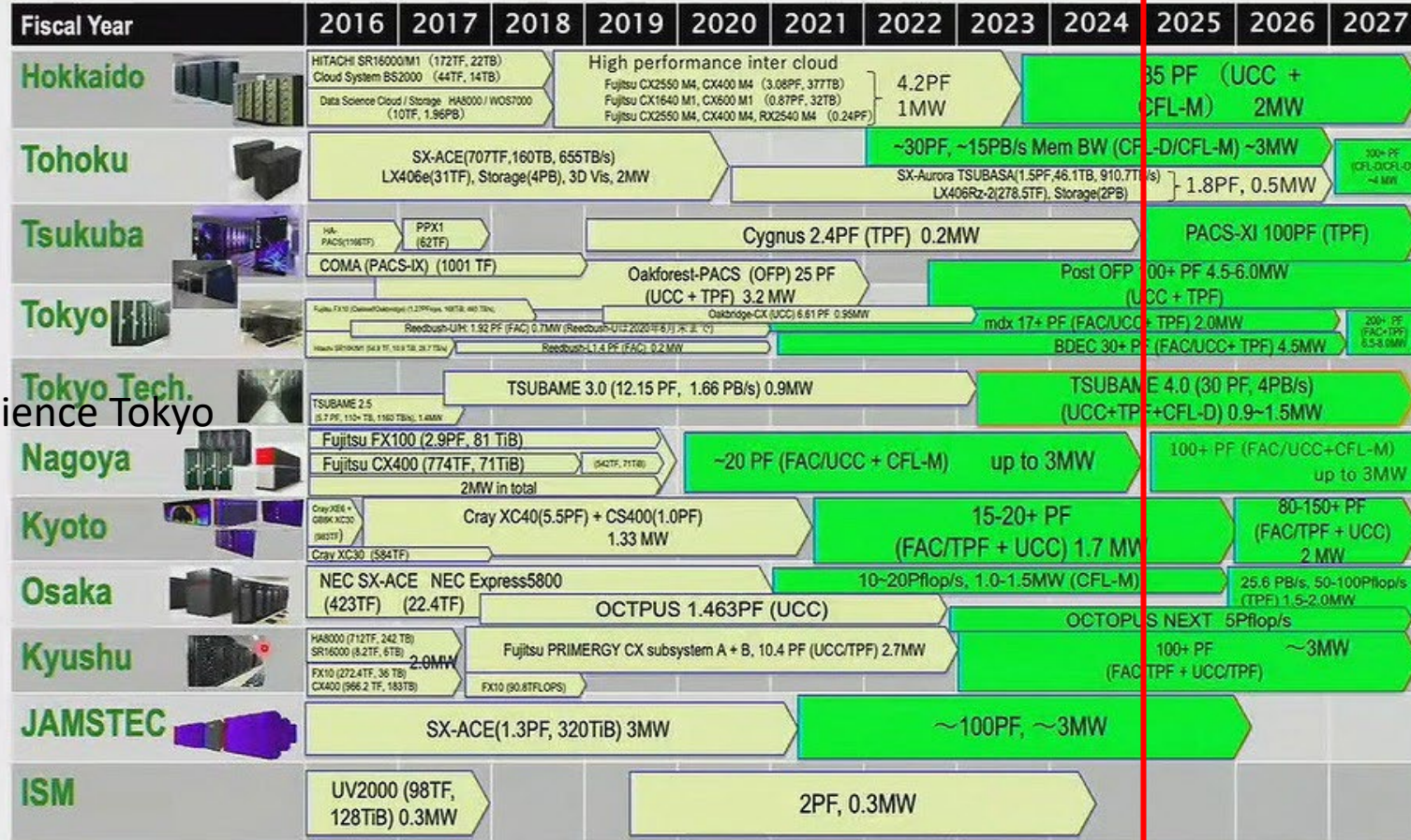
183 (6.1PF)

178 (6.2PF)

119 (10PF)

28 (47PF)

Science Tokyo



Vector engine  
SX-Aurora TSUBASA

A64FX

A64FX

SX-Aurora TSUBASA

2021/03/09 JCAHPC (Joint center of Tokyo and Tsukuba Universities)


Power is the maximum consumption including cooling

NVIDIA Grace CPU+H100

Slide (HPCI forum) <https://fugaku100kei.jp/events/20210309/>

# Fugaku's strategy: Application-First

- It is important that the computing platform is easy for users to use, and that it has the perspective of leading and contributing to research and development in a wide range of scientific, technological, and industrial fields. → **Application-First**
  - According to Prof. Satoshi Matsuoka (Director, RIKEN Center for Computational Science), he said “the value of Fugaku lies not just in its hardware, but in all of its software solutions, including applications.”
- Fugaku's CPU is based on ARM, so Fugaku can support many standard applications.  
Fugaku has been used in many researches including studies related to Covid-19.
  - 2021: Gordon Bell Special Prize for High Performance Computing-Based Covid-19 Research
  - 2022: Gordon Bell Prize: Design of Laser-Based Electron Accelerators
- >~95% is continuously used: See its [monitor page](#).
- “Virtual Fugaku” is also available now.
  - We can use the same applications like GENESIS, a molecular dynamics software used in drug discovery and materials development, at AWS. AWS has an instance that is highly compatible with Fugaku's processor, the c7g series, which is equipped with the ARM-based AWS Graviton3 processor. See [AWS's blog](#).



RedHat8,  
OpenHPC, oneAPI,  
python, PyTorch,  
TensorFlow, Julia,  
Ruby, KVM etc.

# Post Fugaku

- The discussion of post Fugaku was started around 2020 (or earlier?), which is before the operation of Fugaku as a public use.
- There was a gap period between Kei and Fugaku, where there was no flagship supercomputer in Japan.
  - Kei stopped in Aug 2019 and Fugaku started in March 2021.
  - We should not have such a gap again.
- RIKEN announced that they start the development of the **FugakuNEXT** (the official code name) this month (Jan 2025).
  - [https://www.riken.jp/pr/news/2025/20250122\\_1/index.html](https://www.riken.jp/pr/news/2025/20250122_1/index.html) (Japanese, I cannot find English page.)
- The FugakuNEXT might begin operation in **at least 2030**.
  - Mentioned in “Final report on next-generation computing infrastructure” (MEXT (Ministry of Education, Culture, Sports, Science and Technology) in Japan)

# FugakuNEXT (1)

- **Performance and Functionality Guidelines**
  - Uses accelerator-units such as GPUs in addition to CPUs.
  - Provides a computing environment with significantly improved power performance.
  - Achieves simulations 5 to 10 times faster than Fugaku.
  - Realizes the world's highest level environment for AI learning and inference (effective performance: 50EFLOPS or more).
- **Development Policy**
  - Technology innovation
    - Development of technologies that significantly accelerate AI, resulting in tens to hundreds of times improvements in AI application performance
    - Adoption of high-bandwidth and heterogeneous node architecture, and advanced memory technology
    - System design that meets the demand for new computing resources that are expected to develop in the future, such as "AI for Science"
  - Sustainability and continuity
    - Building a system and continuously improving the software environment, that are highly compatible with standards and existing ecosystems
    - Realizing R&D environments that realize a sustainable and continuous system construction and operation environment
    - Realizing energy savings by advancing operation technology that further evolves the efforts made on Fugaku
  - Made with Japan
    - Advancing domestic technologies that have appealing power to the world, ensuring strategic indispensability in the information industry by promoting technology succession, and expanding into the global market
    - Promoting projects through international collaboration, such as collaboration between domestic and international technologies and human resources

## FugakuNEXT (2)

- Quantum computers are not yet at the stage where they can be widely provided as a practical computing platform. However, given that the research and demonstration of hybrid calculations with supercomputers are progressing, one will consider connecting them to HPC and providing an environment for using them. This is the case for the existing HPCs including Fugaku.

- Target system performance

	CPU	Accelerator
# of nodes	>3400	
FP64 Vector Perf.	>48PFLOPS	>3.0EFLOPS
FP16/BF16 Matrix Op. Perf.	>1.5EFLOPS	>150EFLOPS
FP8 Matrix Op. Perf.	>3.0EFLOPS	>300EFLOPS
FP8 Matrix Op. Perf. with sparsity	-	>600EFLOPS
Main memory size	>10PiB	>10PiB
Man memory bandwidth	>7PB/s	>800PB/s
Total power consumption	<40MW (computing nodes and storage)	



# High energy physics and Fugaku/FugakuNEXT/Other HPCs in Japan

- HEP computing resource is completely independent of the HPC in Japan.
  - ATLAS, Belle II, Super Kamiokade etc.
- In Japanese HPC community, HEP is not identified so much (my personal opinion). HEP is just one of users. So, what we can do is to buy their resources by following their standard procedure. (In HPC, the “Particle physics” means Lattice QCD calculation etc.)
- ICEPP (ATLAS Tier2 in Japan) has used some of them (not Fugaku but x86) as external compute nodes for the ATLAS.
  - We also know that A64FX (Fugaku chip) is not optimized for the ATLAS simulation job. A64FX is slower than other ARM chips by a factor of about 3-4. See [Murakami-san's slide](#). (We also tested it using the ATLAS standard simulation jobs and the conclusion was the same.)

# Summary

- HPC for academic in Japan has two layers.
  - We have one flagship supercomputer. This is Fugaku now.
  - We have O(10) supercomputers, which are 5-10% computing powers of the flagship one.
  - HPCI service is available for researchers, for example, single-sign on to the HPC system.
- Fugaku began operation in March 2021 and has been used for many researches.
  - “Application-First”: many standard applications are supported in the supercomputer.
  - >~95% is continuously used.
- FugakuNEXT’s development has been just started.
  - 5-10 faster than Fugaku
  - Accelerator-units (such as GPUs) will be adopted.
  - “Made with Japan” (like A64FX) is also an important policy.
  - Plan to operate in at least 2030.