RIKEN R-CCS HIGH PERFORMANCE BIG DATA RESEARCH TEAM

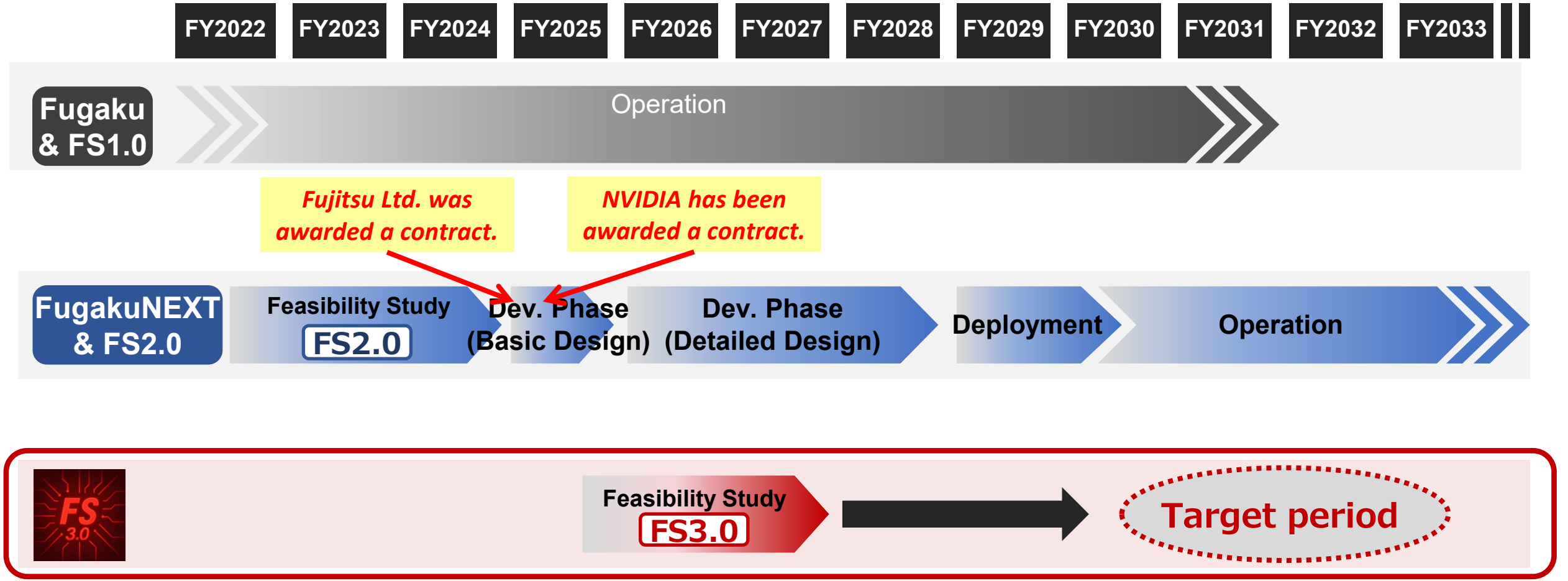
FS3.0 Web site

FS3.0*: Feasibility Study on the future HPCI
(Next-Generation Computing Infrastructure)

* Commissioned project funded by the Ministry of Education, Culture, Sports, Science and Technology (MEXT)

FS3.0 = FugakuNEXT-NEXT

Expected Timeline of FugakuNEXT and FS3.0



HPCI: High Performance Computing Infrastructure

- HPCI is a shared computational infrastructure that connects the flagship system “Fugaku” with supercomputers and storage systems at universities and research institutions across Japan via a high-speed network
- It enables integrated use of the world’s leading supercomputers with diverse characteristics, thereby supporting a wide range of user needs.

High-Performance Computing Infrastructure (HPCI)

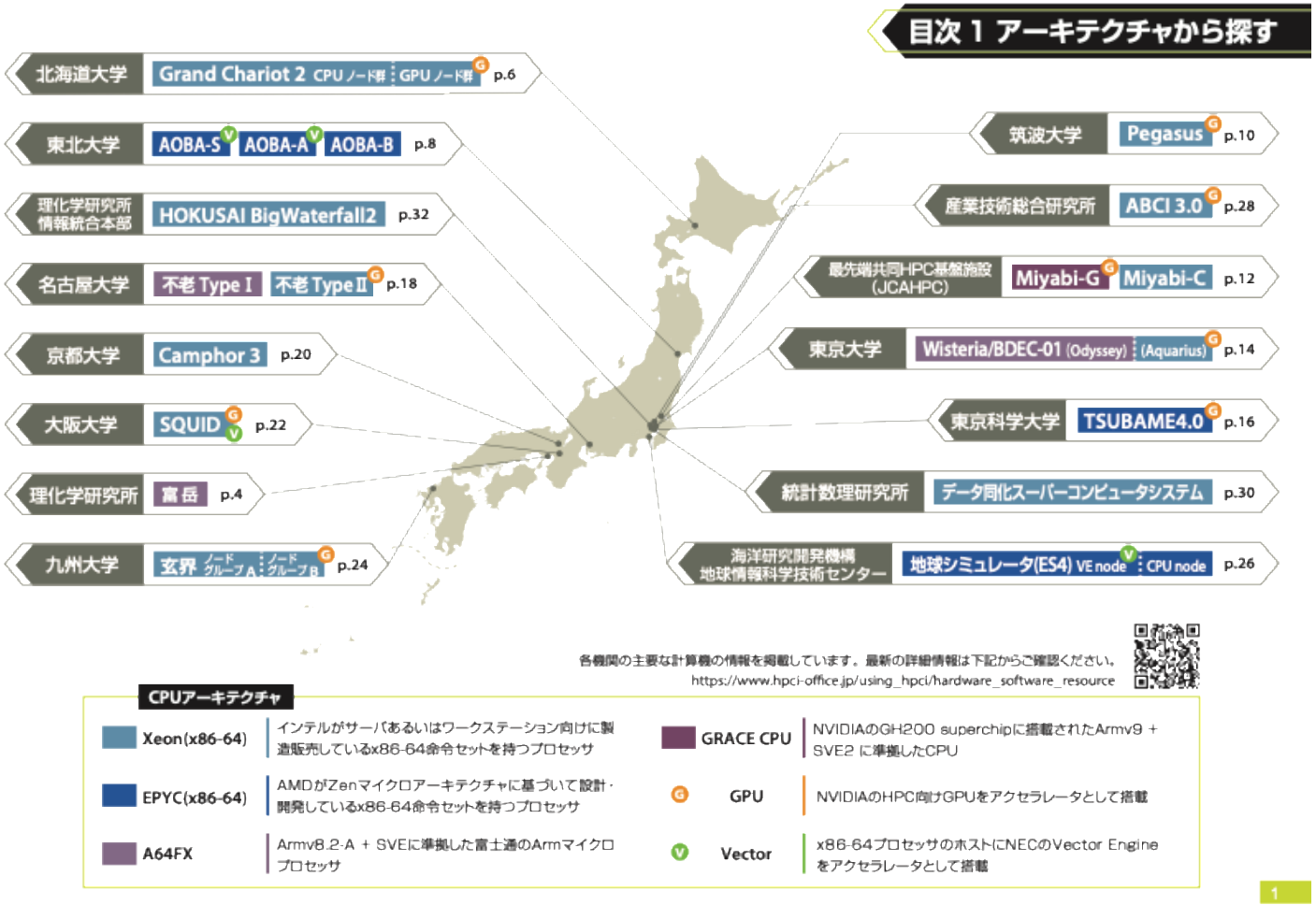
HPCI is a large-scale shared computing infrastructure established by Japan's Ministry of Education, Culture, Sports, Science and Technology (MEXT).

HPCI is a large-scale shared computing infrastructure established by Japan's Ministry of Education, Culture, Sports, Science and Technology (MEXT). Centered on the flagship **system Fugaku**, it connects world-class supercomputers and storage systems located at national research institutes and universities via a high-speed network, providing an **advanced** shared computing environment that meets diverse user needs.

HPCI System Configuration

- RIKEN Center for Computational Science (R-CCS)
- National Institute of Informatics (NII)
- Institute for Solid State Physics, The University of Tokyo
- Cyberscience Center, Tohoku University
- Center for Computational Sciences, University of Tsukuba,
- Advanced Institute for Computational Science (AICS), RIKEN
- Information Technology Center, The University of Tokyo
- Information Media Center, Hiroshima University
- Information Technology Center, Nagoya University
- Academic Center for Computing and Media Studies, Kyoto University
- Cybermedia Center, Osaka University
- Research Institute for Information Technology, Kyushu University
- Earth Simulator Center, JAMSTEC
- Institute of Statistical Mathematics
- Information Technology Center, The University of Tokyo (Kashiwa)
- Information Technology Center, Hokkaido University

HPCI計算資源ハンドブック:
https://fugaku100kei.jp/download_file/aabde610-655a-4780-acff-462f43b62d4b/9



FS3.0: Overview (Lead organization: RIKEN R-CCS)

Background

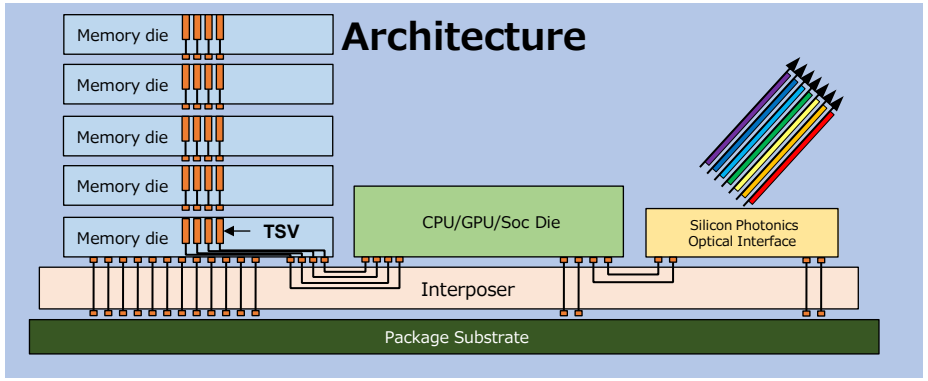
- The development of the next flagship system that will succeed the supercomputer Fugaku (hereinafter referred to as “FugakuNEXT”) and the enhancement of the HPCI (High Performance Computing Infrastructure for Innovation), which integrates diverse computational resources across Japan, have been progressing in parallel.
- In the FugakuNEXT era, the rapid advancement of generative AI will significantly increase the importance of AI for Science, and the computational resources required to support such workloads are becoming increasingly diverse.
- To meet these emerging requirements, the FS3.0 project conducts feasibility studies on computer architecture, system software, and next-generation application development, and proposes procurement/development plans for HPCI systems in the FugakuNEXT era and beyond. (FS1.0: Feasibility Study for Fugaku; FS2.0: Feasibility Study for FugakuNEXT; FS3.0: Feasibility Study for Systems Beyond FugakuNEXT.)

Feasibility Study Areas

- **Architecture Feasibility Study**
  - In collaboration with participating vendors, we investigate advanced technologies such as next-generation semiconductor technologies, 2.5D/3D/3.5D stacked packaging, and chip-to-chip optical interconnects, and explore a range of possible system configurations for HPCI systems in the FugakuNEXT era.
- **System Software Feasibility Study**
  - Building on the software ecosystem inherited from Fugaku and FugakuNEXT, we will define Japan’s system-software development positioning within the international community and examine system software that supports AI for Science in addition to traditional large-scale simulations.
- **Application Feasibility Study**
  - We investigate trends in innovative application use and development enabled by generative AI—such as automated computational science experiments and automatic code generation—across diverse scientific domains, and clarify the system requirements needed to enable such next-generation HPC applications.

Applications
enabled by AI and HPC computational resources

System software
bridging applications and next-generation architecture



FS3.0 Project Organization

FS3.0 Steering Group
RIKEN
GL: Kento Sato, GSL: Kazuhiko Komatsu, AD: Masaaki Kondo
Project Management Office
GL: Yoji Shimada, GSL: Naoki Nishi,
Mem: Tokuko Masui, Naoko Otani, AD: Hiroki Iwasaki

GL/GSL: Group Leaders/Group Sub-Leaders
SGL/SGSL: Subgroup Leaders/Subgroup Sub-Leaders
AD: Advisor

Architecture Group
Tohoku University / Science Tokyo / RIKEN
GL: Kazuhiko Komatsu@Tohoku Univ.,
GSL: Ryohei Kobayashi@Science Tokyo, Masayuki Sato@Tohoku Univ.,
Shigeki Tomishima@RIKEN AD: Kentaro Sano@RIKEN

- **Architecture Sub-G1**: **Intel**, SGL: Joshua Fryman
- **Architecture Sub-G2**: **NVIDIA**, SGL: Dan Ernst
- **Architecture Sub-G3**: **HPE**, SGL: Fumiki Negishi
- **Architecture Sub-G4**: **Fujitsu**, SGL: Naoki Shinjo
- **Architecture Sub-G5**: **NEC**, SGL: Shintaro Momose
- **Architecture Sub-G6**: **AMD**, SGL: Keiji Yoshida

System Software Group
Osaka University
GL: Keichi Takahashi, GSL: Arata Endo,
AD: Kento Sato@RIKEN

- **HPC Programming & Communication Env. Sub-G**: **University of Tsukuba**, SGL: Norihisa Fujita@Univ. of Tsukuba, AD: Hitoshi Murai@RIKEN, Miwako Tsuji@RIKEN, Takeshi Nanri@Kyushu Univ.
- **Advanced Accelerator Programming Env. Sub-G**: **Meiji University**, SGL: Takaaki Miyajima
- **Numerical Library Sub-G**: **Hokkaido University**, SGL: Takeshi Fukaya , AD: Toshiyuki Imamura
- **Storage and File Systems Sub-G**: **University of Tsukuba**, SGL: Osamu Tatebe

Application Group
RIKEN
SGL: Rio Yokota,
GSL: Mohamed Wahib, Jens Domke,
AD: Takeshi Iwashita, Takeshi Shimokawabe

- **Next Gen. App. Needs Sub-G**: **University of Tokyo**, SGL: Yohei Miki , SGSL: Takashi Shimokawabe
- **Next Gen. App. Dev. Methodology Sub-G**: **Nagoya University**, SGL: Tetsuya Hoshino

Research Group/Sub-Group

Operation Technology Advisory Group
RIKEN
AD: Keiji Yamamoto, Atsuko Takefusa

Quantum HPC Hybrid Advisory Group
Tohoku University
AD: Kazuhiko Komatsu

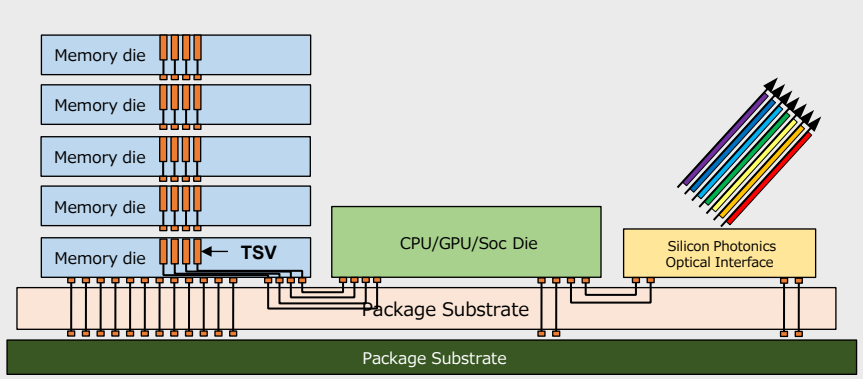
Strategic Advisory Group

V1.5 (Plan From April 22, 2026)

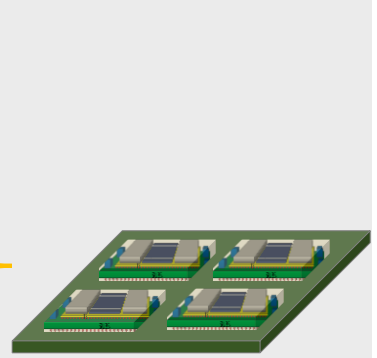
Direction of architectural research

- **Study of Architectures for Advancing Next-Generation HPC and AI Research**
- **Node Architecture**
  - **Heterogeneous systems**: Explore heterogeneous node designs that combine domestic and international CPUs, GPUs, FPGAs, and AI accelerators, in addition to the FugakuNEXT architecture (domestic Arm CPU + GPU).
  - **Next-generation memory hierarchies**: Study configurations featuring large-capacity, high-bandwidth memory using HBM and CXL, along with hierarchical memory-access management.
  - **Adoption of 2.5D/3D/3.5D packaging and chiplet technologies**: Investigate the potential for high-density integration to alleviate data-movement bottlenecks and enable low-power operation (Collaborators: Tadahiro Kuroda, Atsutake Kosuge).
  - **New computing principles**: Assess the applicability of high-bandwidth, low-latency, and energy-efficient computing and communication technologies such as optical computing (e.g., silicon photonics) (Collaborator: Jun Shiomi).
- **System Architecture**
  - **Large-scale HPC/AI systems**: Supporting HPC simulations, AI processing, and their convergence at the scale of several thousand to ten thousand nodes, emphasizing high performance and efficiency through tightly coupled and homogeneous system architectures.
  - **Small- to medium-scale AI-utilization systems**: Consider flexible cluster or distributed environments capable of deploying large numbers of AI agents or distributed inference workloads, enabling diverse loosely coupled and heterogeneous system configurations, including cloud-like operational models.

Node Architecture



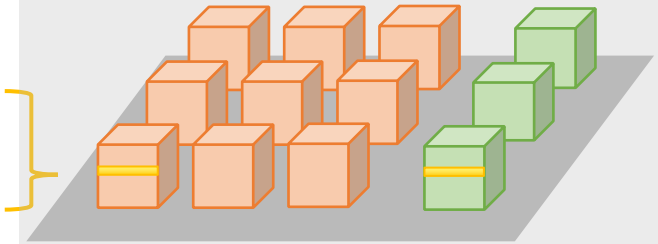
Example of Processing Element Configuration



Example of Mounting on Substrate

System Architecture

For large-scale HPC and AI experiments
Tightly coupled and homogeneous system architecture



For AI utilization
Loosely coupled and heterogeneous system

Planned Deliverables: Continuous FS Platform for future HPCI

- **Continuous FS Platform for future HPCI**
  - Ensure long-term preservation and utilization of FS outcomes, including results and know-how from **HPCI-RB**/**HPCI-CB**
  - (1) **HPCI-RB**: HPCI Reference Blueprint
    - Reference describing system requirements for future HPCI systems based on feasibility studies
    - **TS ref. (Tech. Spec. Reference) :** Provide references for architecture and system software roadmaps
    - **PEC ref. (Performance Evaluation Criteria Reference):** Provide effective performance to be achieved by the target applications on systems proposed in TS ref.
  - (2) **HPCI-CB**: HPCI Continuous Benchmarking Env.
    - Enables continuous performance evaluation of target applications and performance prediction in future HW