

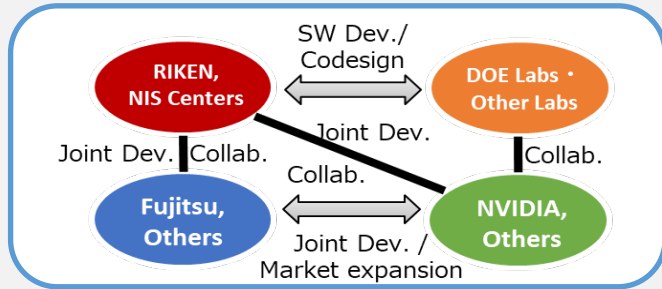
FugakuNEXT Project

Masaaki Kondo

**Division Director
Next-Generation HPC Infrastructure Development Division
RIKEN Center for Computational Science**

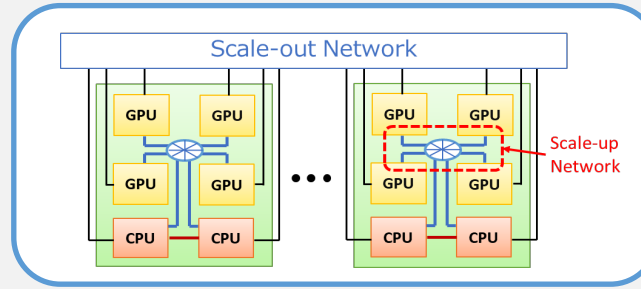
“FugakuNEXT” Development Strategy for Enhancing Scientific and Industrial Competitiveness

Made with Japan



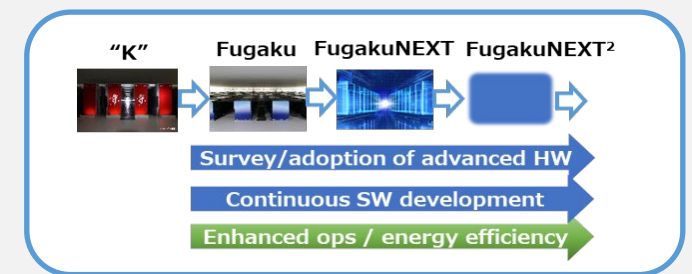
- Building competitive systems and advancing domestic techs through global collaboration, driving market expansion in Japan’s semi industry
- Talent development through international collaboration for ensuring sovereignty in IT techs

Technological Innovation



- Collaborative development of tightly integrated high-performance CPU–GPU architectures, advanced memory technologies, etc.
- Driving AI and HPC innovation for up to 100x application performance gains

Sustainability / Continuity



- Building ecosystem-ready systems with sustained SW development
- Application modernization for future advanced systems and establishing a supporting framework
- Achieving energy efficiency by advancing operational techs

“FugakuNEXT” Ecosystem to Strengthen Japan’s Semiconductor and IT Infrastructure

- Expanding Computational Frontiers through the Development of Next-Generation AI-HPC Platforms and Advancing Science with “AI for Science”
- Ensuring Japan’s Sovereignty in Advanced AI Technologies and Computing Infrastructure
- Sustained R&D through Continuous Semiconductor Innovation and Secure Computing Resources 2

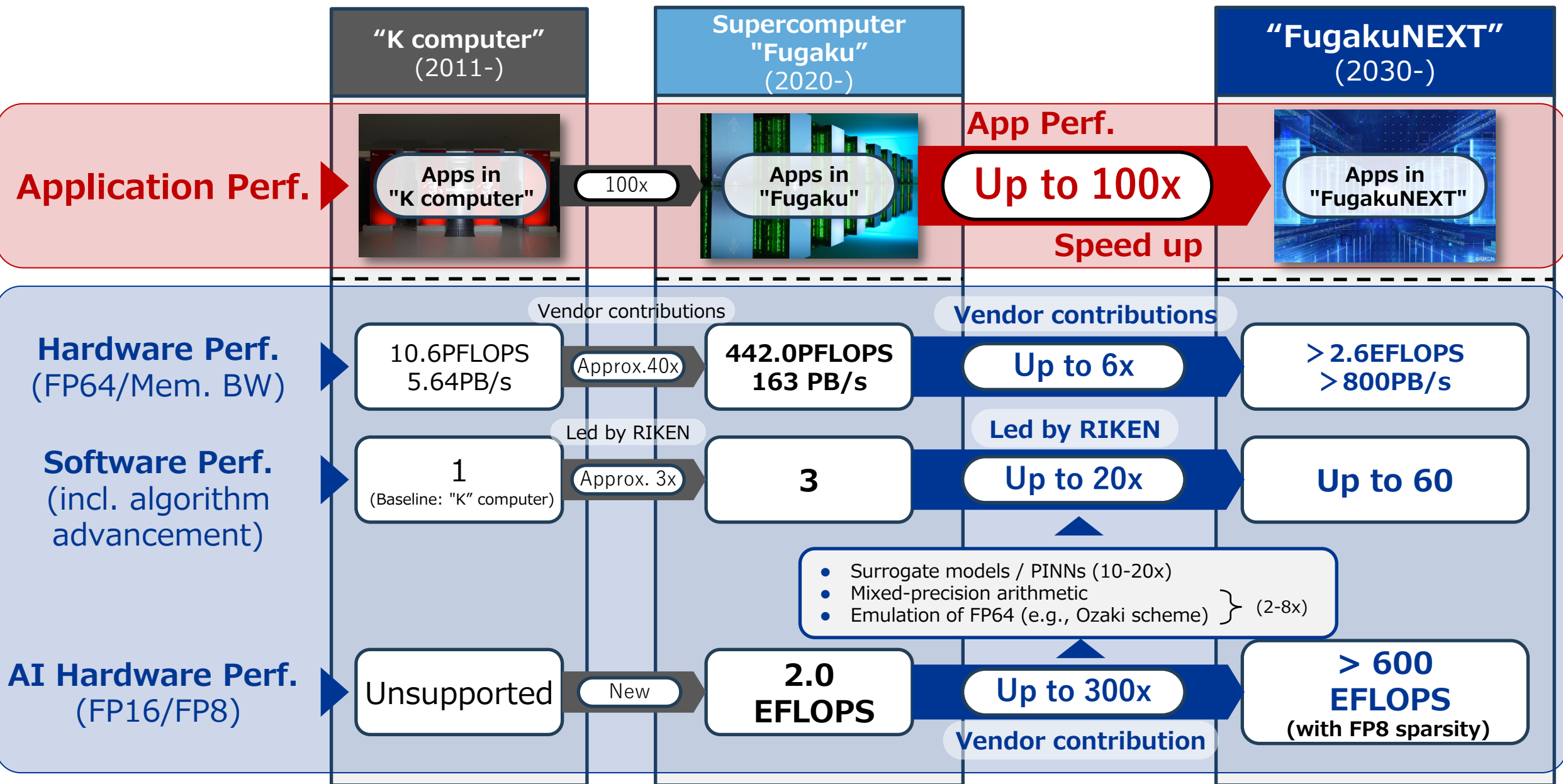
Targeted Performance of FugakuNEXT

- Performance targets by “MEXT HPCI Plan and Promotion Committee”
 - 5 to 10x higher effective performance than Fugaku for existing HPC applications
 - 50 EFLOPS or higher for effective AI perf. with Zetta-scale peak in mind
 - overall application speedup of several tens of times by combining simulation and AI

Towards these goals, vendor procurement was conducted under the following specs

Items	CPU _s	GPU _s	Fugaku	Improvements
Total node count	≥ 3,400		158,976	
Theoretical peak perf. of FP64 vector	≥ 48 PFLOPS	≥ 3.0 EFLOPS	488-537 PFLOPS	5.7x~
Theoretical peak perf. of FP16/BF16 matrix	≥ 1.5 EFLOPS	≥ 150 EFLOPS	1.95-2.15 EFLOPS	70.5x~
Theoretical peak perf. of FP8 matrix	≥ 3.0 ELOPS	≥ 300 EFLOP	—	
(Sparsity considered FP8 matrix)	—	≥ 600 EFLOPS	—	
Main memory capacity	≥ 10 PiB	≥ 10 PiB	4.85 PiB	4.1x~
Main memory bandwidth	≥ 7 PB/s	≥ 800 PB/s	163 PB/s	4.9x~
System power consumption	≤ 40 MW (compute node and storage)		about 30 MW	

Road to Achieving up to 100x Application Performance

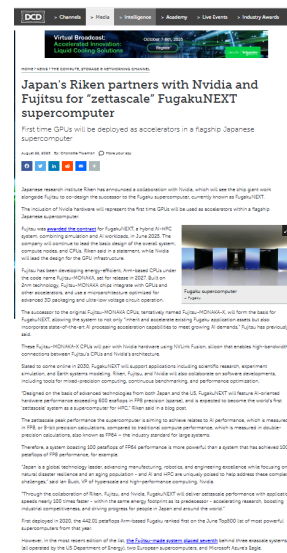


Launch of FugakuNEXT Initiative: Press Conf. in Aug. 2025

RIKEN announced the launch of the internationally coordinated development of “FugakuNEXT” with Fujitsu and NVIDIA. This has been positively received by many media outlets.

Media outlets	#
Online and newspaper articles	39
Television	7
Total	46

Summary for Aug 22-28, 2025



“Made with Japan” Development Framework

International collaboration on FugakuNEXT system-SW development
 Co-design of AI-HPC systems (including feedback to CPU/GPU design)
 FugakuNEXT testbed construction and co-design tool deployment
 AI-enabled facilities and operations infrastructure

RIKEN

Design and development of MONAKA-X
 World-leading AI-HPC perf as a standalone CPU
 Planned fabrication in 2029 using a 1.4 nm process
 tight integration with NVIDIA GPU

FUJITSU

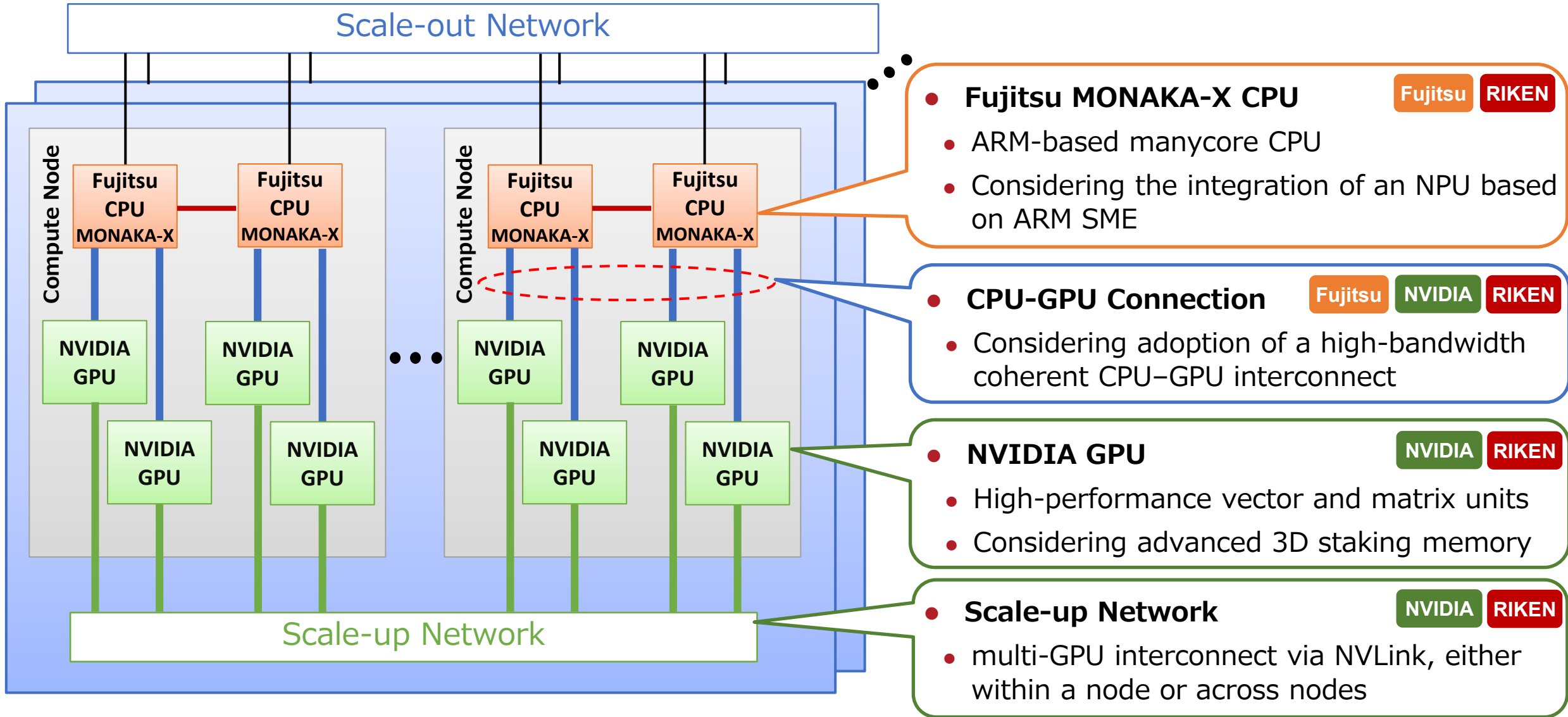
Design and devel. of GPUs for AI-HPC apps
 Users support including GPU porting
 high-BW coherent connection with CPUs
 Exploration of advanced technologies for new
 memory and higher power efficiency

NVIDIA

Also with 4-way MoU among ANL, Fujitsu, NVIDIA, and RIKEN to advance AI and HPC

- Advancing globally compelling technologies by leveraging each party’s strengths
- Designing a global-standard system in the era of AI/HPC convergence

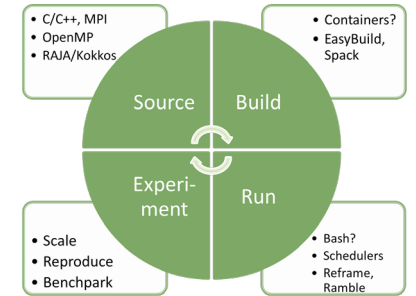
Overview of “FugakuNEXT” Architecture



Detailed architecture configuration will be determined in basic design phase

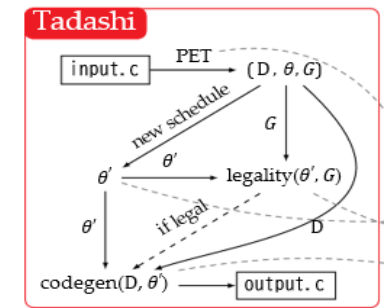
- **Co-design with wide applications**

- Benchmarking a wide variety of apps with continuous verification through CI/CD/CB framework
 - Leveraged by Benchmark effort



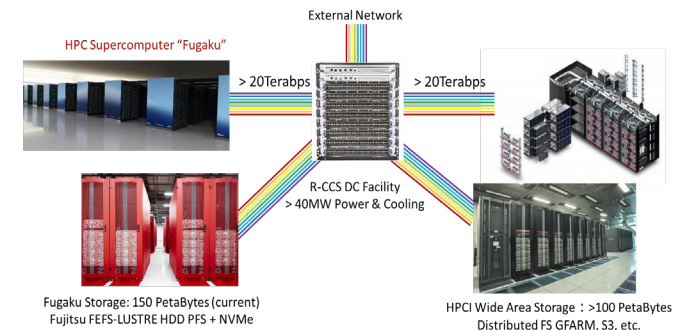
- **AI-driven application development/coding/optimization**

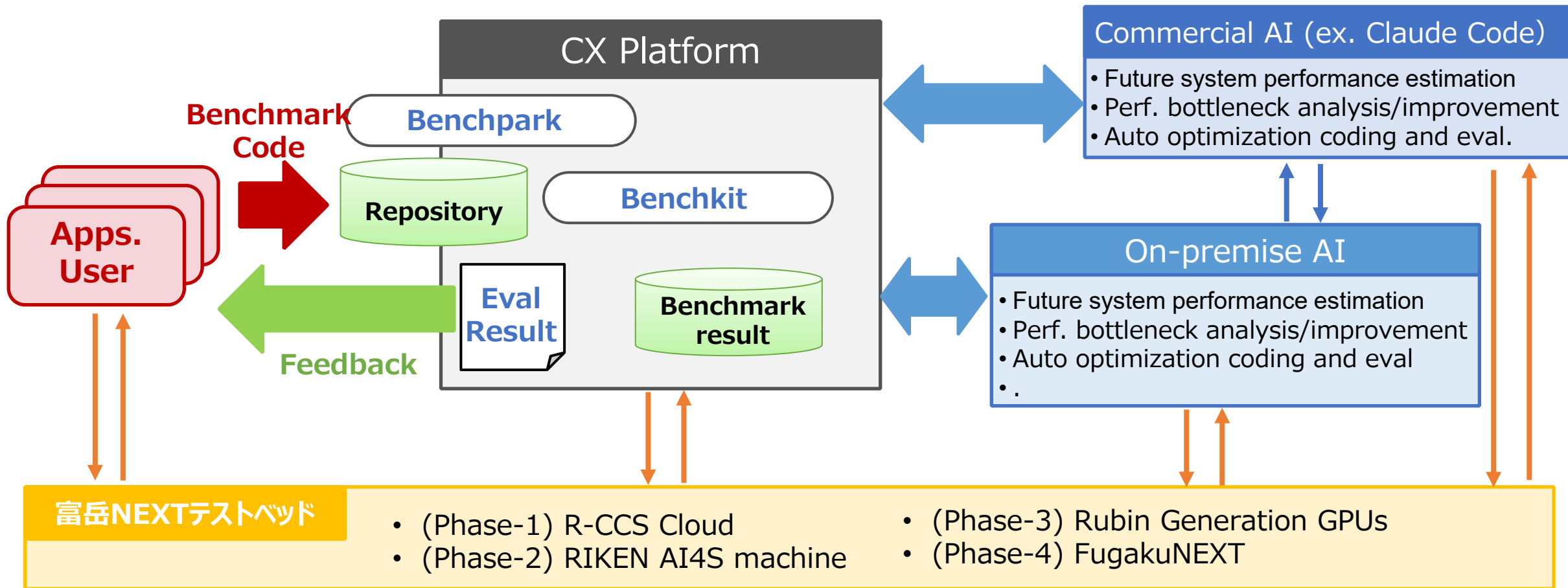
- Deployment of a coding AI and Chatbot
- Aggressive agentic AI support



- **FugakuNEXT testbeds**

- Deploying proxy machines for FugakuNEXT
- Leveraging new GPU machines installed in R-CCS and other supercomputers
- Broaden our user base, even to international users

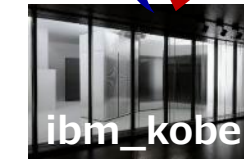




In collaboration with domestic/international institutions, develop an AI-driven co-design and coding environment (Prototype planned for release around mid-FY2026)

Expected Schedule of FugakuNEXT Project

- Goal: starting operation in 2030
- To avoid a loss in available computing resources, Fugaku and FugakuNEXT will operate in parallel
- Construction of a new computer bldg. has begun at R-CCS Kobe



Planned Schedule for FugakuNEXT and New Computer Building Development

