RIKEN launches international initiative with Fujitsu and NVIDIA for "FugakuNEXT" development

– Building the next-generation AI-HPC platform to solve complex social challenges through computational science –


Satoshi Matsuoka, Director
Riken Center for Computational Science

Aim of the Next-Generation Computing Infrastructure, Built on the Legacy of “Fugaku”

- **The value and national significance of the flagship system**
  - Ensuring global leadership in computational science through a Zetta-scale HPC environment
  - Leveraging economies of scale for deployment, operation, and community engagement
  - Japan’s most powerful research platform for AI development, where computation is decisive

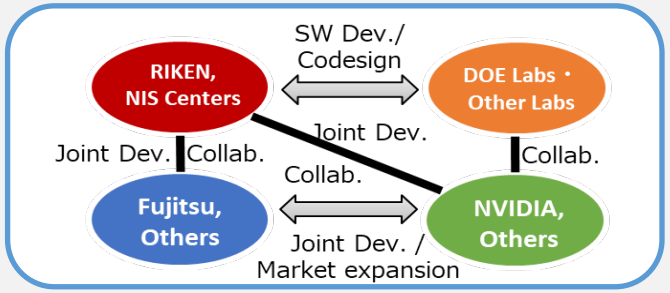
Zetta-scale computing resources to drive value creation by expanding computational frontiers

- **Development of “FugakuNEXT” to enhance industrial competitiveness**
  - Built on the ARM architecture of “Fugaku” while pursuing further advancements in system design
  - Pioneering computation-driven problem solving, incl. “AI for Science” and quantum comp. integration
  - Introducing GPUs as accelerators under a Japan–U.S. joint development framework, modernizing Japan’s applications with active use of AI to accelerate social implementation of research outcome
  - Establishing a “Made with Japan” devel. framework in alignment with Japan’s semiconductor strategy

Promoting the advancement and succession of competitive domestic technologies, securing their strategic indispensability in IT industry, and driving their expansion into global markets

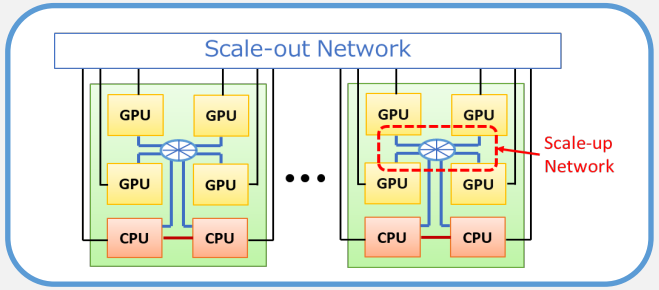
“FugakuNEXT” Development Strategy for Enhancing Scientific and Industrial Competitiveness

Made with Japan



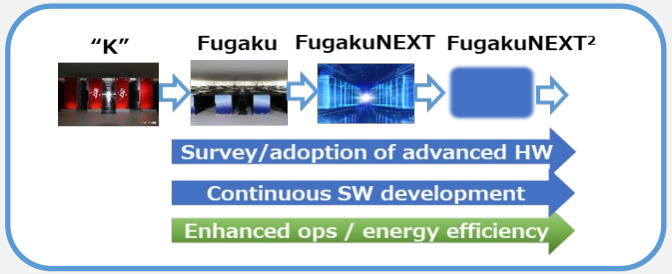
- Building competitive systems and advancing domestic techs through global collaboration, driving market expansion in Japan’s semi industry
- Talent development through international collaboration for ensuring sovereignty in IT techs

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

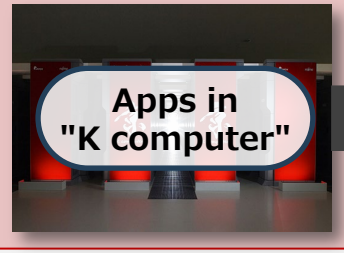


- Building ecosystem-ready systems with sustained SW development
- Application modernization for future advanced systems and establishing a supporting framework
- Achieving energy efficiency by advancing operational techs

“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

Road to Achieving up to 100x Application Performance

	"K computer" (2011-)	Supercomputer "Fugaku" (2020-)	"FugakuNEXT" (2030-)
Application Perf.	 Apps in "K computer"	100x	 Apps in "Fugaku"	App Perf. **Up to 100x** Speed up	 Apps in "FugakuNEXT"
Hardware Perf. (FP64/Mem. BW)	10.6PFLOPS 5.64PB/s	Vendor contributions Approx.40x → **442.0PFLOPS 163 PB/s**	Vendor contributions **Up to 6x**	**>2.6EFLOPS >800PB/s**
Software Perf. (incl. algorithm advancement)	1 (Baseline: "K" computer)	Led by RIKEN Approx. 3x → **3**	Led by RIKEN **Up to 20x**	**Up to 60**
		• Surrogate models / PINNs (10-20x) • Mixed-precision arithmetic • Emulation of FP64 (e.g., Ozaki scheme) } (2-8x)	
AI Hardware Perf. (FP16/FP8)	Unsupported	New → **2.0 EFLOPS**	**Up to 300x** Vendor contribution	**> 600 EFLOPS** (with FP8 sparsity)