

# Feasibility Studies on Next-Generation Supercomputing Infrastructures

Takahiro Katagiri (Information Technology Center, Nagoya University, E-mail: katagiri@cc.nagoya-u.ac.jp / RIKEN R-CCS, Visiting Researcher)

## Overview of Study (System Research by RIKEN)

### Project Overview

The next-generation computational infrastructure is expected to become a platform for realizing SDGs and Society 5.0 by **providing advanced digital twins** that will bring "Research DX" in the science. Aiming to realize a versatile computing infrastructure that can **execute entire workflow by making full use of wide range of computational methods, simulation techniques, and BigData** at scale, we conduct a holistic investigation on architecture, system software and library technologies through co-design with applications.

As a basic principle of system design, we **practice the "FLOPS to Byte" concept** from architecture development to algorithm or application design to **streamline data transfer and computation under power constraints**, while taking necessary computing accuracy into consideration. Under the **ALL JAPAN team composition**, we will investigate system configurations and elementary technologies which improve effective performance of the next-generation computing infrastructure.



### Object of Investigation

#### Research on Architecture

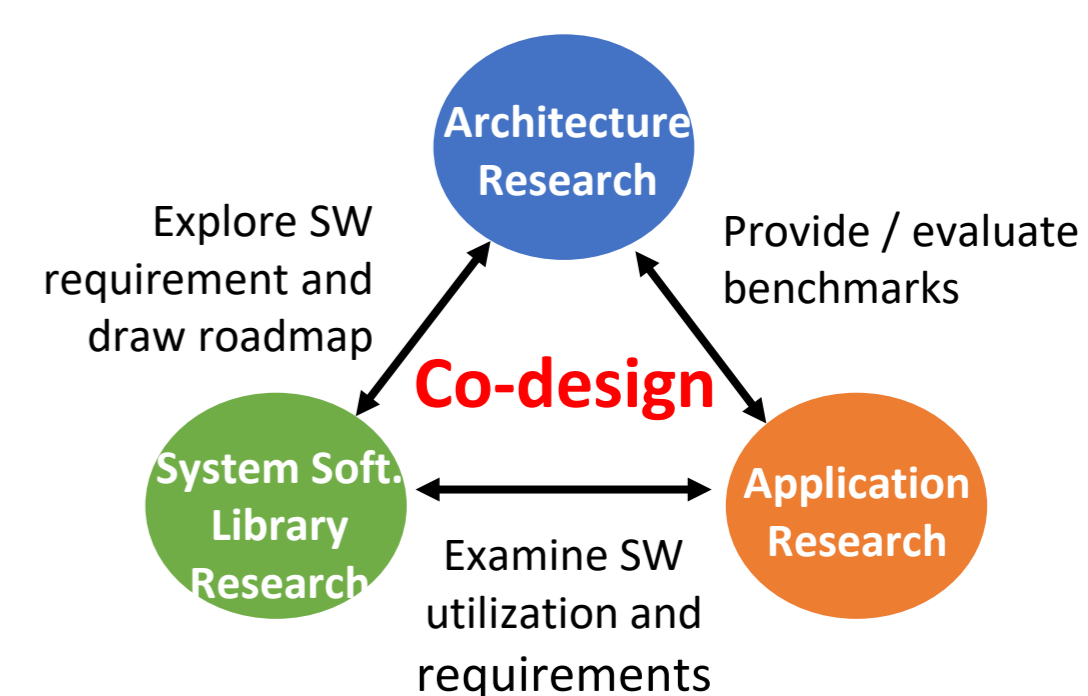
- Investigating technological possibilities (such as 3D stacked mem, accelerators, chip-to-chip direct optical link) and performance of the entire system or its components based on trends in semiconductor and packaging technologies
- Predicting future system performance based on performance analysis of benchmark sets provided by Application Research Group, and feeding back to next-generation application development

#### Research on System Software and Library

- Drawing roadmap for future system software development in Japan, specially considering data utilization enhancement, integration of AI technology with first-principles simulation, real-time data processing, and assurance of high security

#### Research on Applications

- Building a broad benchmark set to evaluate multiple architecture choices while considering improvements in algorithms and parameters of application based on the results of architectural evaluations and exploratory "what-if" performance analysis
- Investigating what classes of algorithms are expected to evolve significantly for future systems



### Investigation Schedule

	2022 Q3	2022 Q4	2023 Q1	2023 Q2	2023 Q3	2023 Q4	2024 Q1
Architecture	Explore device/architecture technology	Examine existing SW and its utilization		Performance estimation with benchmarks	Identify requirement of SW development	Architecture study	Draw roadmap
Application	Examine existing apps and benchmark design			Perf. analysis by benchmark evaluation		Study algorithm improvement	

## Overview of System Software and Library Research

- Investigate technological trend of system software and draw R&D roadmap based on it
  - Holistic studies on 8 system SW areas, cross-cutting technologies (security, auto-tuning, etc.) and platforming
  - Prioritize SW / tools to develop domestically by considering usage on flagship and 1-tier machines, development cost, proficiency of existing ecosystem, and other factors
- Study new technological area in system software for industrialization
  - Investigate new areas of development to encourage use in a wide range of applications
  - Providing a platformed digital twin environment for industrial use, promotion of data utilization, fusion of machine learning & first-principles simulation, advanced large-scale real-time data processing, high security, etc.
- ALL JAPAN team organization with industry-academia collaboration
  - Expect future ripple effect on HPCI supercomputer centers

