Post-K Development

Yutaka Ishikawa
Project Leader, Flagship 2020
RIKEN Center for Computational Science

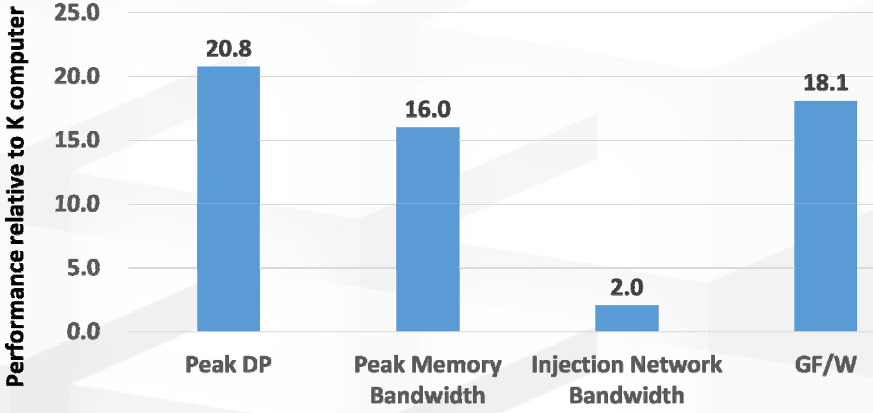
Post-K

- A Post-K prototype machine was built in Summer 2018. Since then, Fujitsu has been testing and evaluating the machine.
- Ten racks of Post-K achieve almost the same performance of K computer (864 racks)

X 10 =

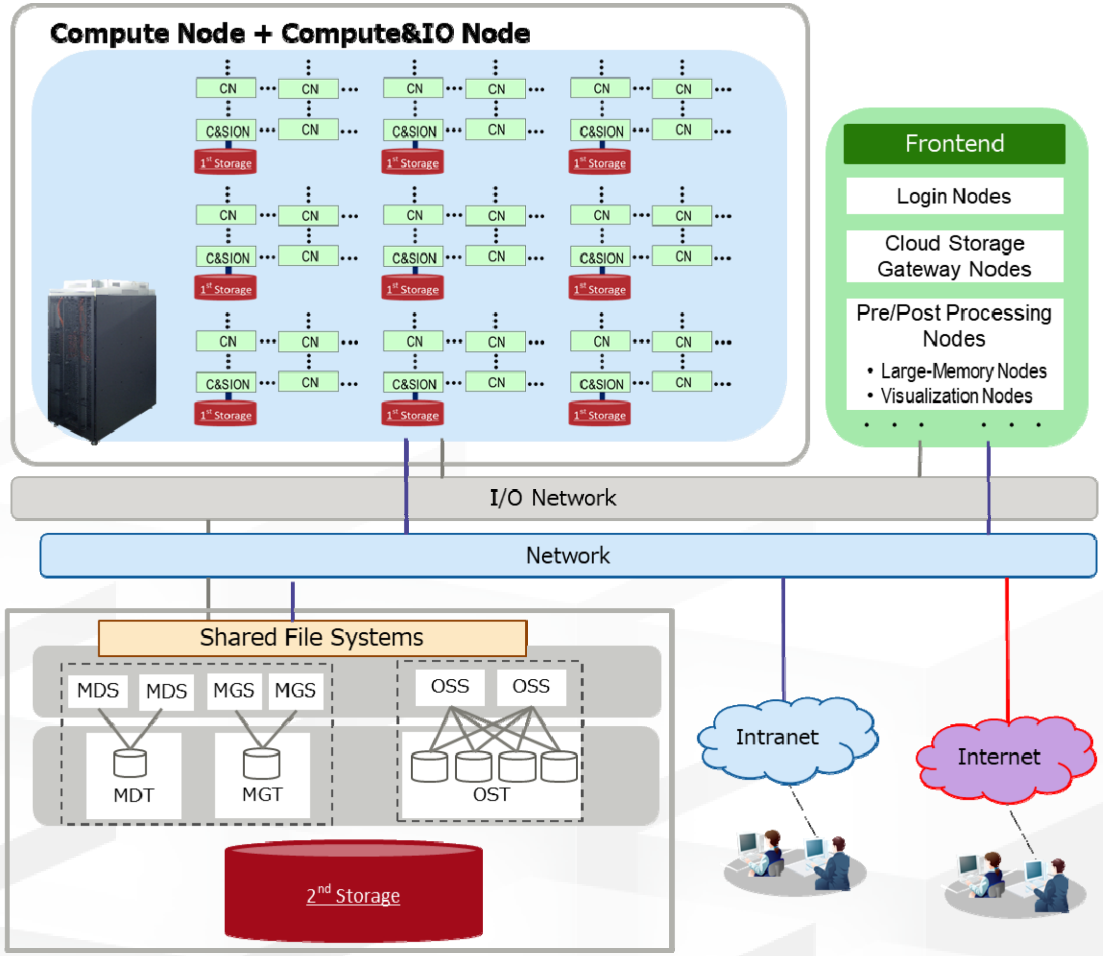
©RIKEN

		Post-K	K
	CPU Architecture	A64FX (Armv8.2-A SVE +Fujitsu Extension)	SPARC64 VIIIfx
Node	Cores	48	8
	Peak DP performance	2.7+ TF	0.128 TF
	Main Memory	32 GiB	16 GiB
	Peak Memory Bandwidth	1024 GB/s	64 GB/s
	Peak Network Performance	40.8 GB/s	20 GB/s
Rack	Nodes	384	102
	Peak DP performance	1+ PF	< 0.013PF
	Process Technology	7 nm FinFET	45 nm



An Overview of Post-K Hardware

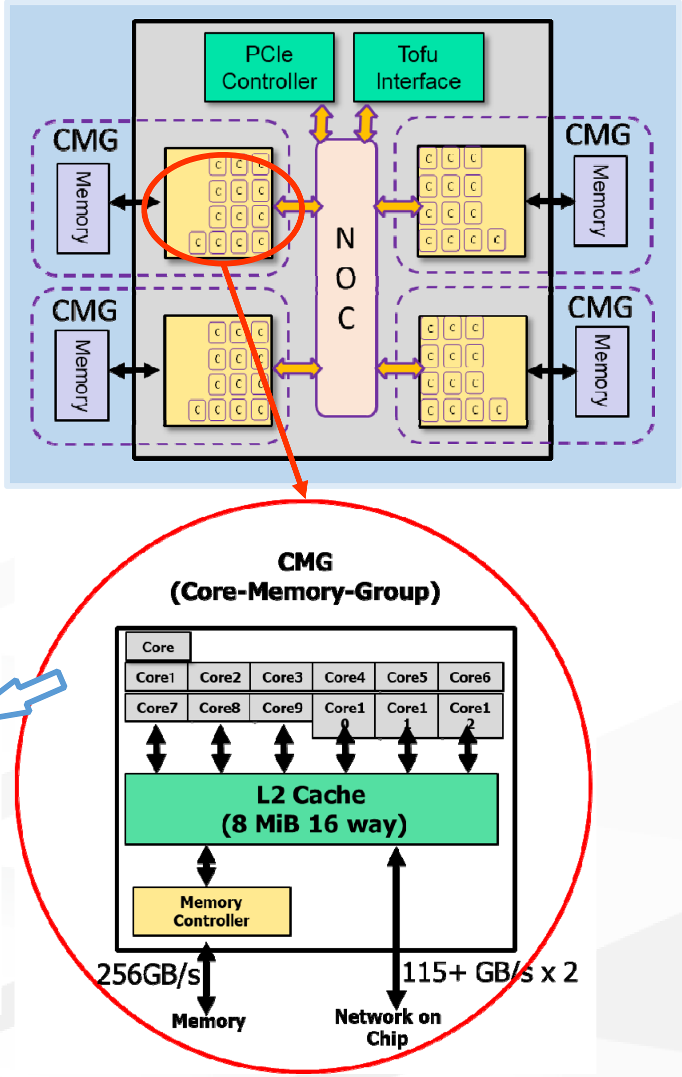
- **150k+ node**
- **Two types of nodes**
  - Compute Node and Compute & I/O Node connected by Fujitsu TofuD, 6D mesh/torus Interconnect
- **3-level hierarchical storage system**
  - 1st Layer
    - One of 16 compute nodes, called Compute & Storage I/O Node, has SSD about 1.6 TB
    - Services
      - ~ Cache for global file system
      - ~ Temporary file systems
        - Local file system for compute node
        - Shared file system for a job
  - 2nd Layer
    - Fujitsu FEFS: Lustre-based global file system
  - 3rd Layer
    - Cloud storage services

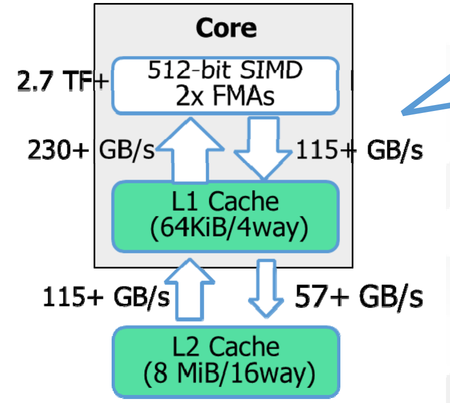


CPU A64FX

Architecture	Armv8.2-A SVE (512 bit SIMD)
Core	48 cores for compute and 2/4 for OS activities
	DP: 2.7+ TF, SP: 5.4+ TF, HP: 10.8+ TF
Cache L1	64 KiB, 4 way, 230+ GB/s(load), 115+ GB/s (store)
Cache L2	CMG: 8 MiB, 16way Node: 3.6+ TB/s Core: 115+ GB/s (load), 57+ GB/s (store)
Memory	HBM2 32 GiB, 1024 GB/s
Interconnect	TofuD (28 Gbps x 2 lane x 10 port)
I/O	PCIe Gen3 x 16 lane
Technology	7nm FinFET

Courtesy of FUJITSU LIMITED





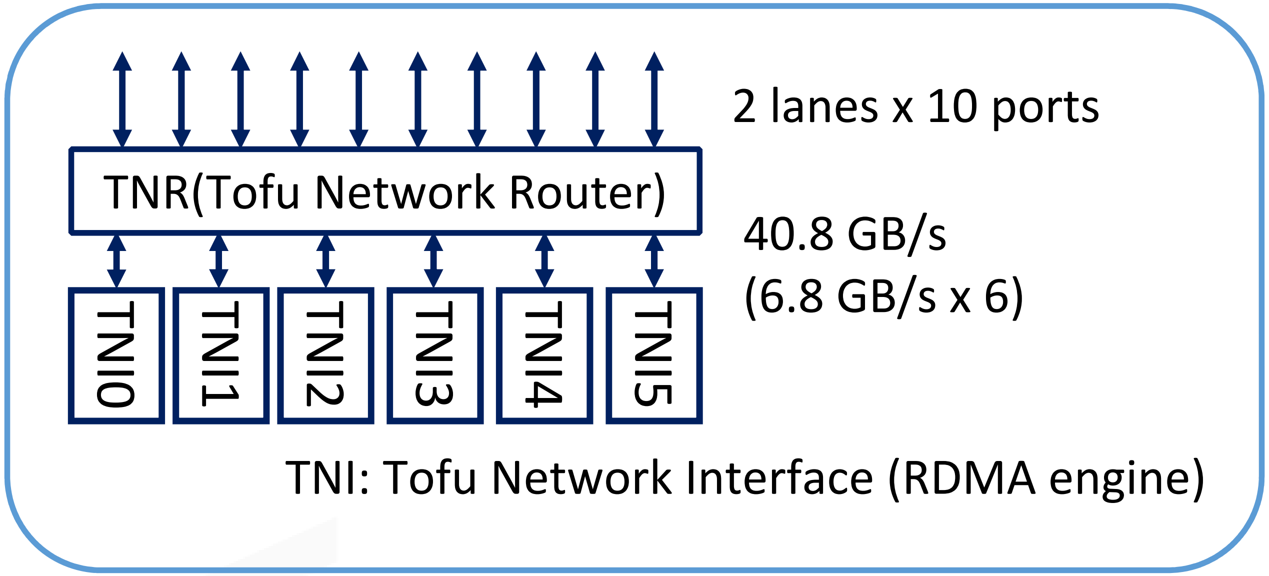
Performance

Stream triad: 830+ GB/s

Dgemm: 2.5+ TF (90+% efficiency)

ref. Toshio Yoshida, “Fujitsu High Performance CPU for the Post-K Computer,” IEEE Hot Chips: A Symposium on High Performance Chips, San Jose, August 21, 2018.

TofuD Interconnect



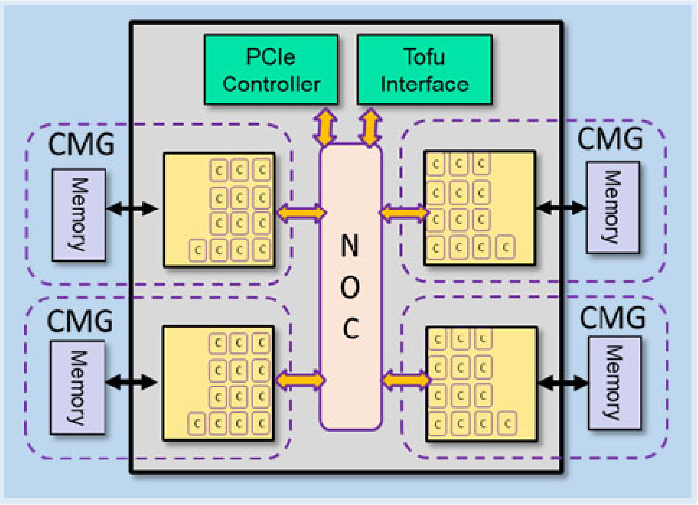
2 lanes x 10 ports

TNR(Tofu Network Router)

40.8 GB/s
(6.8 GB/s x 6)

TNI0 TNI1 TNI2 TNI3 TNI4 TNI5

TNI: Tofu Network Interface (RDMA engine)



- 6 RDMA Engines
- Hardware barrier support
- Network offloading capability

8B Put latency	**0.49 – 0.54 usec**
1MiB Put throughput	**6.35 GB/s**

rf. Yuichiro Ajima, et al. , “The Tofu Interconnect D,” IEEE Cluster 2018, 2018.

Post-K Programming Environment

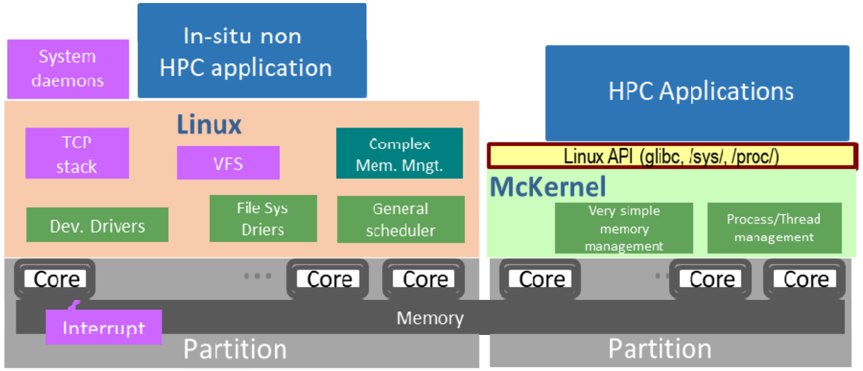
- **Programing Languages and Compilers provided by Fujitsu**
  - Fortran2008 & Fortran2018 subset
  - C11 & GNU and Clang extensions
  - C++14 & C++17 subset and GNU and Clang extensions
  - OpenMP 4.5 & OpenMP 5.0 subset
  - Java
  - GCC, LLVM, and Arm compiler will be also available
- **Parallel Programming Language & Domain Specific Library provided by RIKEN**
  - XcalableMP
  - FDPS (Framework for Developing Particle Simulator)
- **Process/Thread Library provided by RIKEN**
  - PiP (Process in Process)
- **Script Languages provided by Fujitsu**
  - E.g., Python+NumPy, SciPy
- **Communication Libraries**
  - MPI 3.1 & MPI4.0 subset
    - Fujitsu MPI (Based on Open MPI), Riken MPI (Based on MPICH)
  - Low-level Communication Libraries
    - uTofu (Fujitsu), LLC(RIKEN)
- **File I/O Libraries provided by RIKEN**
  - pnetCDF, DTF, FTAR
- **Math Libraries**
  - BLAS, LAPACK, ScaLAPACK, SSL II (Fujitsu)
  - EigenEXA, Batched BLAS (RIKEN)
- **Programming Tools provided by Fujitsu**
  - Profiler, Debugger, GUI

Other Software

- **Batch Job System (Fujitsu)**
  - Technical Computing Suite
    - Successor of K's batch job system

- **Operating System on Compute Nodes**
  - Linux (Fujitsu)
  - McKernel, Light-weight Kernel (RIKEN)
    - Executes the same binary of Linux without any recompilation
    - One of advantages is that McKernel provides much larger page sizes
      - ~ Applications, accessing a huge memory area randomly, may benefit
    - User may select one of McKernel configurations without rebooting

- **Other User-Land**
  - A Linux distribution
- **Open Source Management Tools**
  - Spack/EasyBuild



		McKernel Default 4K	McKernel Default 64K	Linux
.text		4K	64K	64K
.data		64K,2M,32M, 1G	2M, 512M	2M
.bss		64K,2M,32M, 1G	2M, 512M	2M
Stack		64K,2M,32M, 1G	2M, 512M	2M
malloc		64K,2M,32M, 1G	2M, 512M	2M
thread stack		64K,2M,32M, 1G	2M, 512M	2M
Shared memory	System V IPC	64K,2M,32M, 1G	2M, 512M	64K
	POSIX	4K	64K	64K
	XPMEM	64K,2M,32M, 1G	2M, 512M	64K

Concluding Remarks

- **Post-K board, CMU, is displayed in the poster session room**
- **Poster presentations**

□ Programming Environments

[50] Dynamic Multitasking in Upcoming XcalableMP 2.0

□ System Software

[53] Prototype Implementation of MPICH and Data Transfer Framework for Post-K Supercomputer

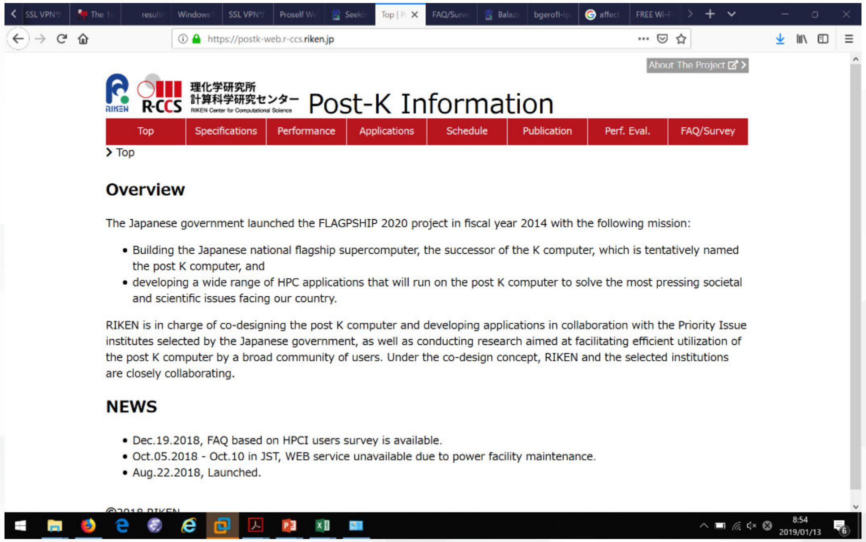
[54] Operating System and Runtime Enhancements for the Post-K Computer

[55] Enhancing MPI-IO with Topology-Awareness at the K computer

[56] Development of Scientific Numerical Libraries on post-K computer

- **Post-K Information is available**

https://postk-web.r-ccs.riken.jp/



Post-K Information

Overview

The Japanese government launched the FLAGPSHIP 2020 project in fiscal year 2014 with the following mission:

- Building the Japanese national flagship supercomputer, the successor of the K computer, which is tentatively named the post K computer, and
- developing a wide range of HPC applications that will run on the post K computer to solve the most pressing societal and scientific issues facing our country.

RIKEN is in charge of co-designing the post K computer and developing applications in collaboration with the Priority Issue institutes selected by the Japanese government, as well as conducting research aimed at facilitating efficient utilization of the post K computer by a broad community of users. Under the co-design concept, RIKEN and the selected institutions are closely collaborating.

NEWS

- Dec.19.2018, FAQ based on HPCI users survey is available.
- Oct.05.2018 - Oct.10 in JST, WEB service unavailable due to power facility maintenance.
- Aug.22.2018, Launched.

©2018 RIKEN