Post-K: A Game Changing Supercomputer for Convergence of HPC and Big Data / AI

Satoshi Matsuoka
Director, Riken Center for Computational Science /
Professor, Tokyo Institute of Technology

SCAsia 2019 SG-JP Session
Singaore
20190312

Apr 1 2018 Became Director of Riken-CCS:
Science, of Computing, by Computing, and for Computing

Riken Center for Computational Science (R-CCS)
World Leading HPC Research, active collaborations w/Universities, national labs, & Industry

Sci. of Computing

Foundational research on computing in high performance for K, Post-K, and beyond towards the "Post-Moore" era, including future high performance architectures, new computing and programming models, system software, large scale systems modeling, big data analytics, and scalable artificial intelligence / machine learning

Sci. by Computing

Breakthrough Science & Technology using high performance computing capabilities of K, Post-K and beyond to address the issues of high public concern, in areas such as life sciences, climate & environment, disaster prediction & prevention, advanced manufacturing, applications of machine learning for Society 5.0.

High Resolution, High Fidelity Analysis & Simulation

Mutual Synergy

Sci. for Computing

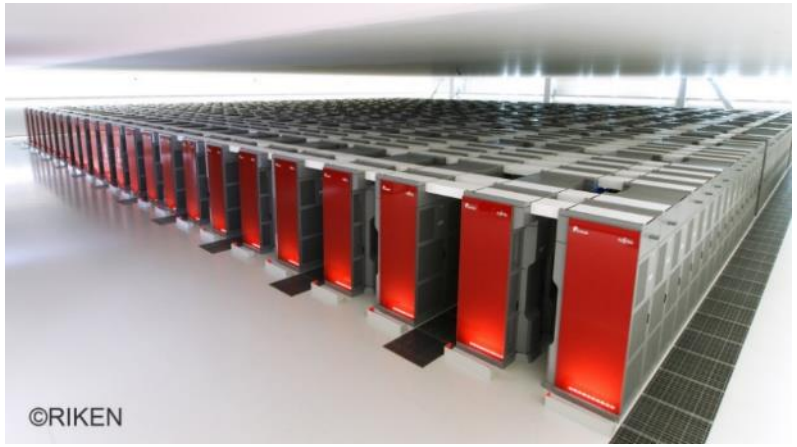
Novel Future High Performance Computing Architectures & Algorithms

New Materials & Electronic Devices e.g., Photonics, Neuromorphics, Quantum, Reconfigurable

K computer

Specifications

- Massively parallel, general purpose supercomputer
- No. of nodes : 88,128
- Peak speed: 11.28 Petaflops
- Memory: 1.27 PB
- Network: 6-dim mesh-torus (Tofu)



Top 500 ranking

LINPACK measures the speed and efficiency of linear equation calculations
Real applications require more complex computations.

- No.1 in Jun. & Nov. 2011
- No.10 in Nov. 2017

Graph 500 ranking

"Big Data" supercomputer ranking
Measures the ability of data-intensive loads

- No.1 in Nov. 2017

HPCG ranking

Measures the speed and efficiency of solving linear equation using HPCG
Better correlate to actual applications

- No. 1 in Nov. 2017

K computer achieved balance of processor speed, memory, and network.
high performance for wide areas of science.

Post-K: The Game Changer

1. **Heritage of the K-Computer, HP in simulation via extensive Co-Design**
- High performance: up to x100 performance of K in real applications
- Multitudes of Scientific Breakthroughs via Post-K application programs
- Simultaneous high performance and ease-of-programming

2. New Technology Innovations of Post-K

- **High Performance, esp. via high memory BW**
  Performance boost by "factors" c.f. mainstream CPUs in many HPC & Society5.0 apps
- **Very Green e.g. extreme power efficiency**
  Ultra Power efficient design & various power control knobs
- **Arm Global Ecosystem & SVE contribution**
  Top CPU in ARM Ecosystem of 21 billion chips/year, SVE co-design and world's first implementation by Fujitsu
- **High Perf. on Society5.0 apps incl. AI**
  Architectural features for high perf on Society 5.0 apps based on Big Data, AI/ML, CAE/EDA, Blockchain security, etc.

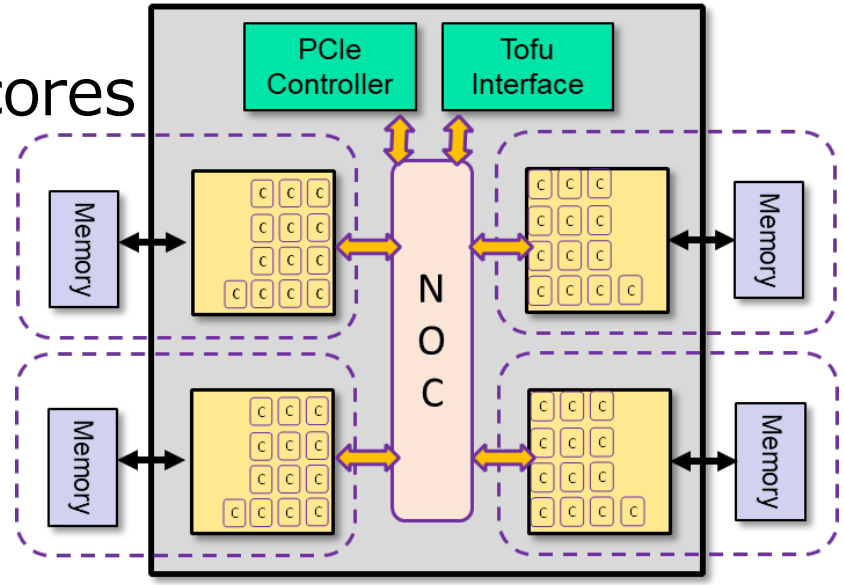
Global leadership not just in the machine & apps, but as cutting edge IT

ARM: Massive ecosystem from embedded to HPC

Technology not just limited to Post-K, but into societal IT infrastructures e.g. Clouds

Post K Processor is…

- **an Many-Core ARM CPU…**
  - 48 compute cores + 2 or 4 assistant (OS) cores
  - Brand new core design
  - Near Xeon-Class Integer performance core
  - ARM V8 --- 64bit ARM ecosystem
  - Tofu 3 + PCIe 3 external connection



- **…but also a GPU-like processor**
  - SVE 512 bit vector extensions (ARM & Fujitsu)
    - Integer (1, 2, 4, 8 bytes) + Float (16, 32, 64 bytes)
  - Cache + scratchpad local memory (sector cache)
  - Multi-stack 3D memory – Massive Mem BW (Bytes/DPF ~0.4)
    - Streaming memory access, strided access, scatter/gather etc.
  - Intra-chip barrier synch. and other memory enhancing features

- **GPU-like High performance in HPC, AI/Big Data, Auto Driving…**

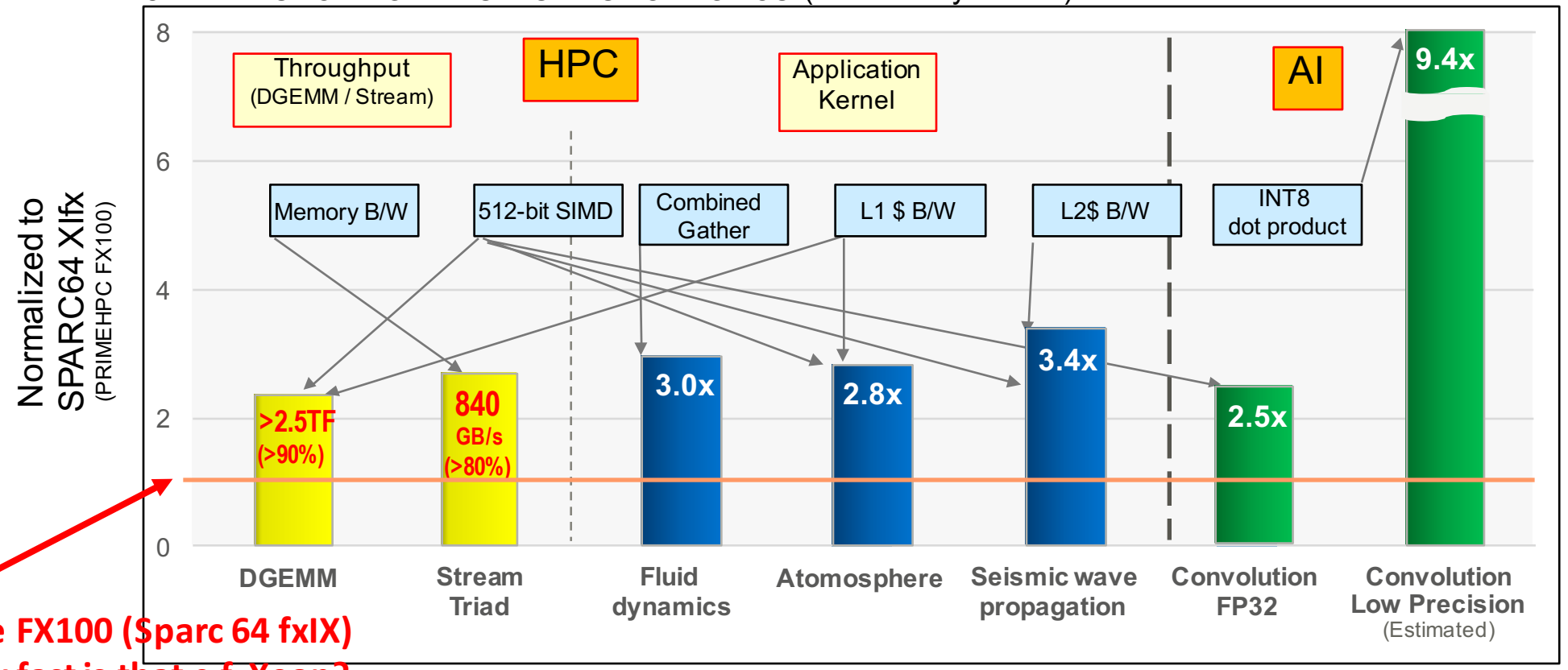
Post-K A64fx A0 (ES) performance

	Performance / CPU					Machine Performance (HPC)		
	Peak TF (DFP)	Peak Mem. BW	Stream Triad	Theoretical B/F	DGEMM Efficiency	Linpack Efficiency	GF/W	Network BW Per Chip
Post-K A64fx (A0 Eng. Sample)	2.764/ 3.072	1024GB/s	840GB/s	0.37/ 0.33	94 %	87.7 %	>15	TOFU-D 40.8GB/s (6.8x 6)
Intel KNL	3.0464	600GB/s	490GB/s	0.20	66%	54.4 %	4.9	12.5 GB/s
Intel Skylake	1.6128	127.8GB/s	97 GB/s	0.08	80 %	66.7 %	4.5	6.2GB/s
NVIDIA V100 (DGX-2)	7.8	900 GB/s	855GB/s	0.12		76 %	15.113	160GB/s 6.2GB/s

Performance

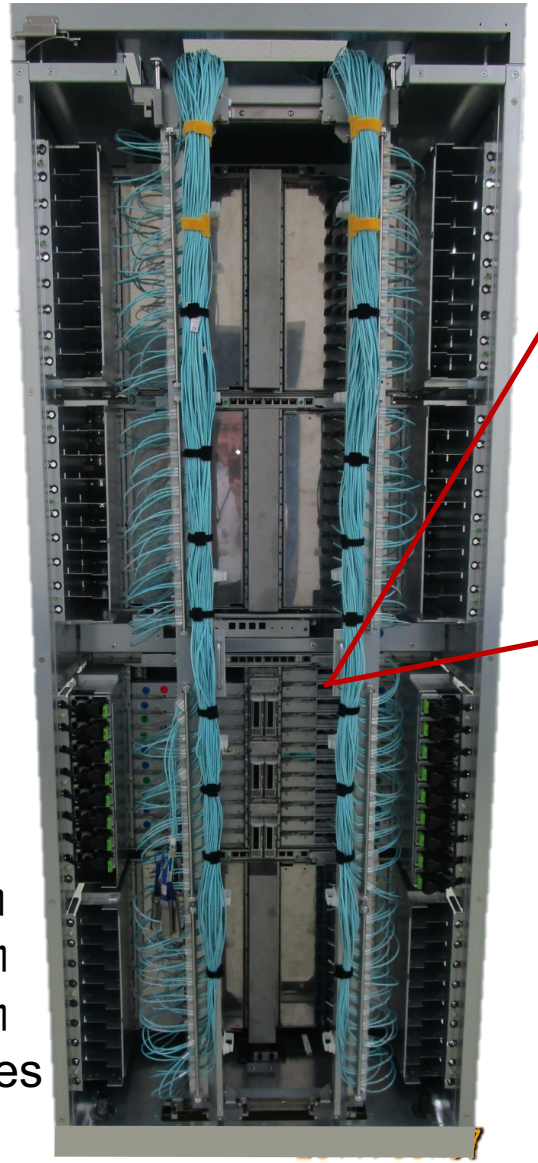
- A64FX boosts performance up by microarchitectural enhancements, 512-bit wide SIMD, HBM2 and process technology
  - > 2.5x faster in HPC/AI benchmarks than SPARC64 XIfx (Fujitsu's previous HPC CPU)
    - The results are based on the Fujitsu compiler optimized for our microarchitecture and SVE

A64FX Benchmark Kernel Performance (Preliminary results)

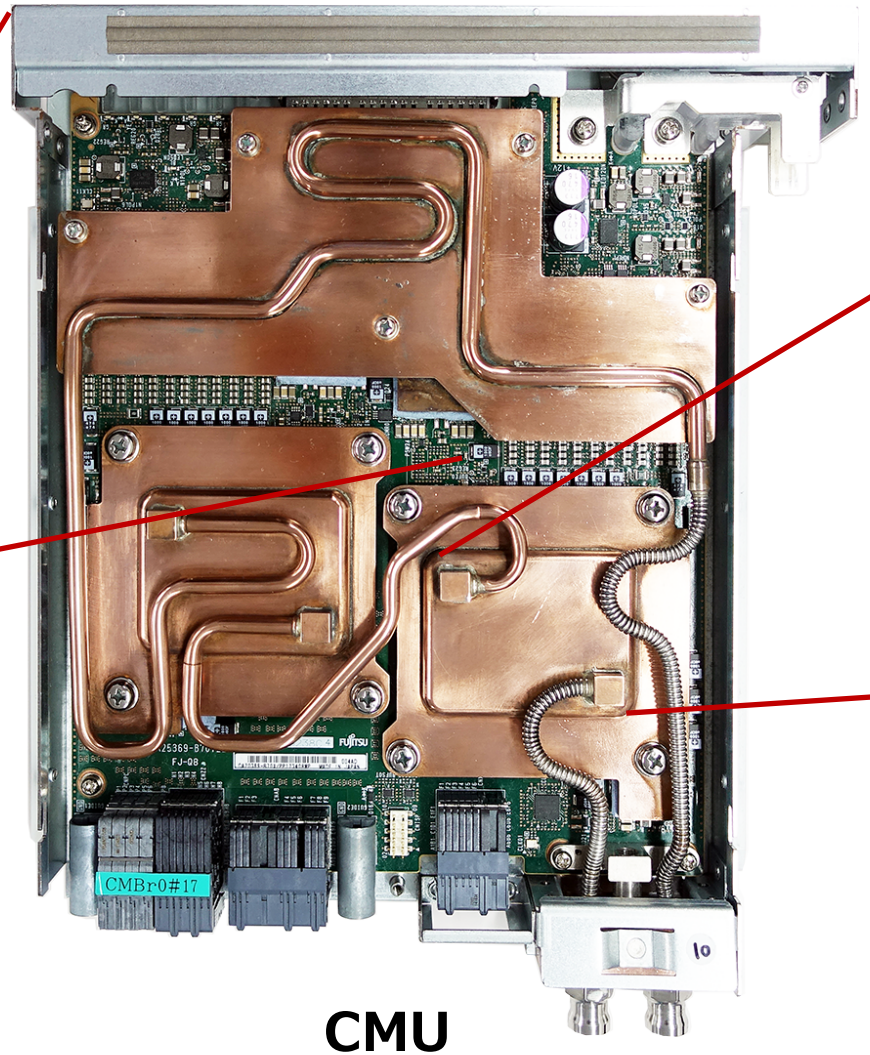


Baseline FX100 (Sparc 64 fxIX)
But how fast is that c.f. Xeon?

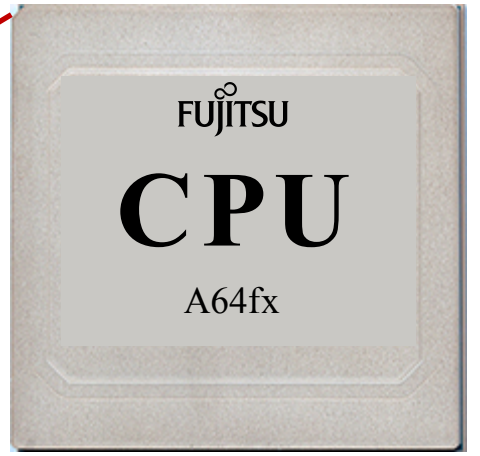
Post-K Chassis, PCB (w/DLC), and CPU Package



W 800mm
D1400mm
H2000mm
384 nodes



CMU

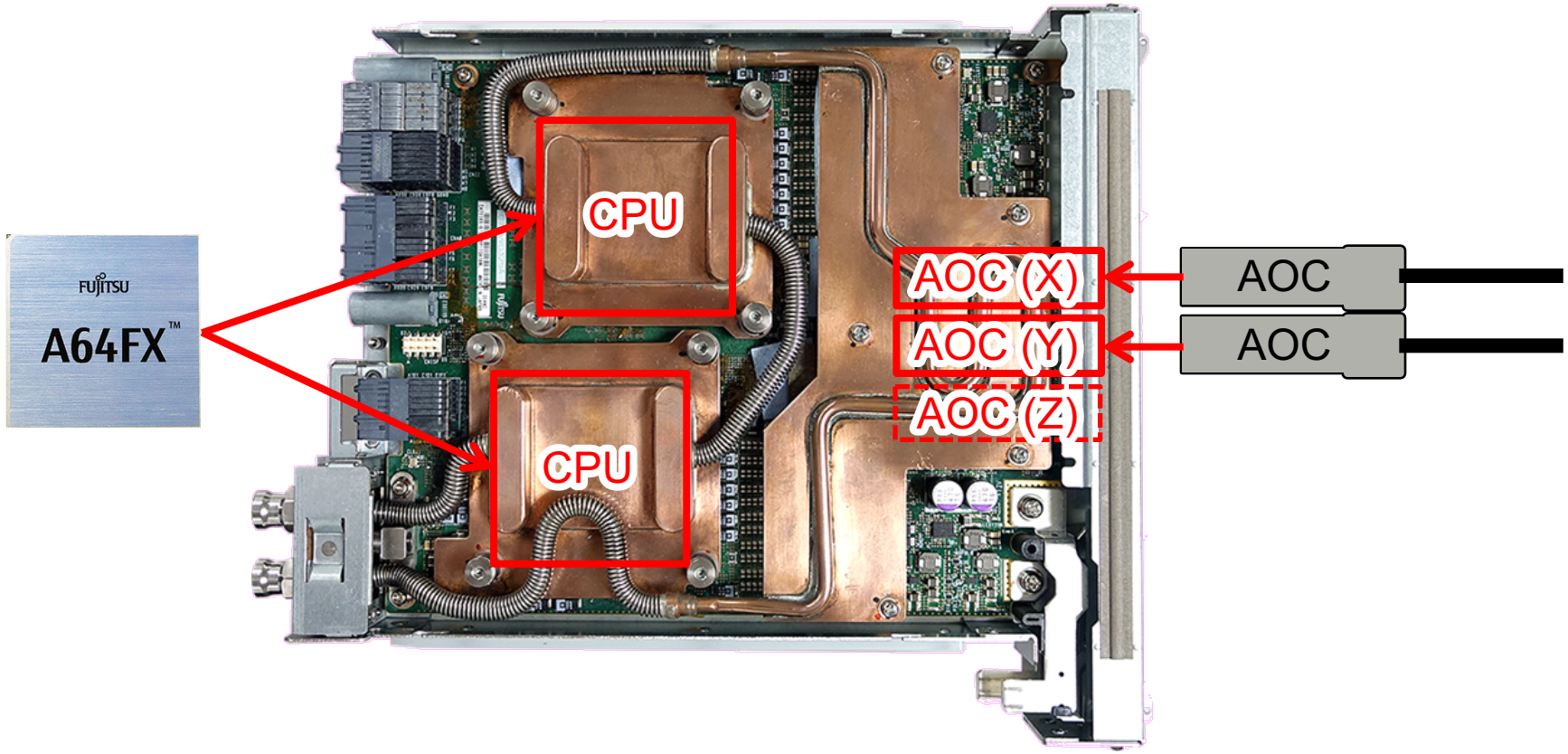


CPU Package

A0 Chip Booted in June
Undergoing Tests

Packaging – CPU Memory Unit of Post-K

- Two CPUs connected with C-axis
  - X × Y × Z × A × B × C = 1 × 1 × 1 × 1 × 1 × 2
- Two or three active optical cable (AOC) cages on the board
  - Each cable bundles two lanes of signals from each of the two CPUs



Put Latencies

FUJITSU

- 8B Put transfer between nodes on the same board
  - The low-latency features were used

	Communication settings	Latency
Tofu1	Descriptor on main memory	1.15 µs
	Direct Descriptor	0.91 µs
Tofu2	Cache injection OFF	0.87 µs
	Cache injection ON	0.71 µs
TofuD	To/From far CMGs	0.54 µs
	To/From near CMGs	0.49 µs

0.20 µs

0.22 µs

- Tofu2 reduced the Put latency by 0.20 µs from that of Tofu1
  - The cache injection feature contributed to this reduction
- TofuD reduced the Put latency by 0.22 µs from that of Tofu2

Injection Rates per Node

FUJITSU

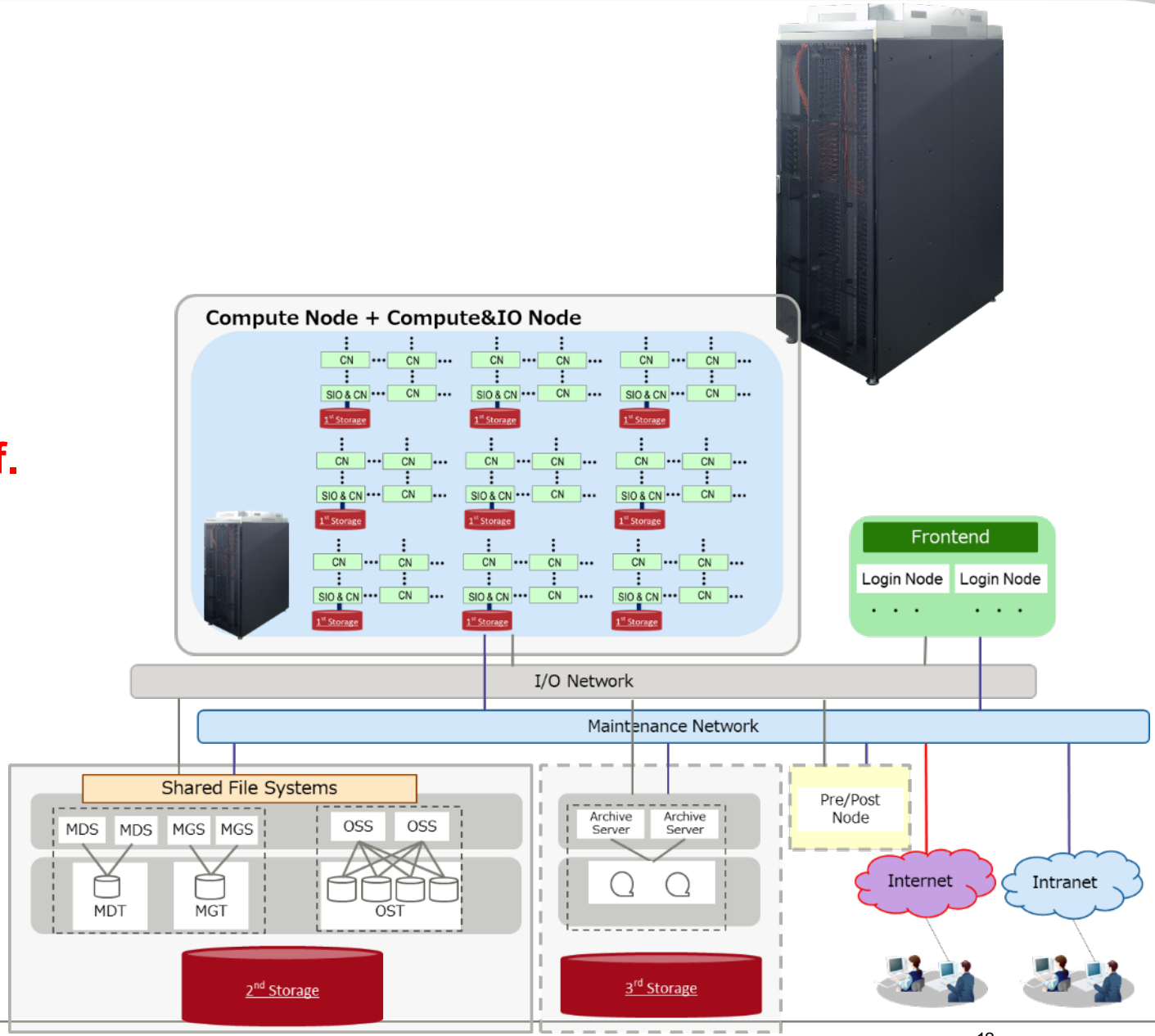
- Simultaneous Put transfers to multiple nearest-neighbor nodes
  - Tofu1 and Tofu2 used 4 TNIs, and TofuD used 6 TNIs

	Injection rate	Efficiency
Tofu1 (K)	15.0 GB/s	77 %
Tofu1 (FX10)	17.6 GB/s	88 %
Tofu2	45.8 GB/s	92 %
TofuD	38.1 GB/s	93 %

- The injection rate of TofuD was approximately 83% that of Tofu2
- The efficiencies of Tofu1 were lower than 90%
  - Because of a bottleneck in the bus that connects CPU and ICC
- The efficiencies of Tofu2 and TofuD exceeded 90 %
  - Integration into the processor chip removed the bottleneck

Overview of Post-K System & Storage

- **Compute Node, Compute + I/O Node connected by TOFU-D**
- **3-level hierarchical storage**
  - 1st Layer: GFS Cache + Temp FS
  - 2nd Layer: Lustre-based GFS
  - 3rd Layer: Off-site Cloud Storage
- **Full Machine Spec**
  - **>150,000 nodes, ~8 million High Perf. Arm v8.2 Cores**
  - **> 400 racks**
  - **~40 MegaWatts Machine+IDC PUE ~ 1.1 High Pressure DLC**
  - **~= 15~30 million state-of-the art competing CPU Cores for HPC workloads (both dense and sparse problems)**



Post-K Programming Environment

FUJITSU

- **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
- **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)

GCC and LLVM will be also available

- Script Languages provided by Linux distributor
  - E.g., Python+NumPy, SciPy
- Communication Libraries
  - MPI 3.1 & MPI4.0 subset
    - Open MPI base (Fujitsu), MPICH (RIKEN)
  - Low-level Communication Libraries
    - uTofu (Fujitsu), LLC(RIKEN)
- File I/O Libraries provided by RIKEN
  - Lustre
  - pnetCDF, DTF, FTAR
- Math Libraries
  - BLAS, LAPACK, ScaLAPACK, SSL II (Fujitsu)
  - EigenEXA, Batched BLAS (RIKEN)
- Programming Tools provided by Fujitsu
  - Profiler, Debugger, GUI
- NEW: Containers (Singularity) and other Cloud APIs
- NEW: AI software stacks (w/ARM)

“Post-K” Chronology

FUJITSU

(Disclaimer: below includes speculative schedules and subject to change)

- 1H2019 “Post-K” manufacturing budget approval by the Diet, actual manufacturing commences
- Apr 2019 R-CCS lead research activities on next-gen architectures will commence => whitepaper to be written by Winter
- Aug 2019 End of K-Computer operations
- 4Q2019~1Q2020 “Post-K” installation starts
- 1H2020 “Post-K” preproduction operation starts
- 2020~2021 “Post-K” production operation starts (hopefully)
- And of course we move on…

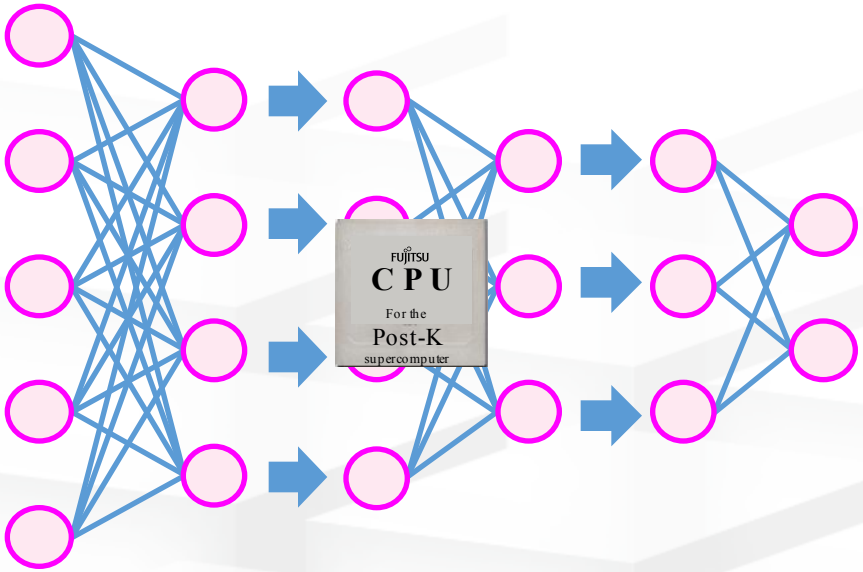
Watch for announcements on “Post-K” technology commercialization by Fujitsu and its partner vendors RSN

Massive Scale Deep Learning on Post-K

Post-K Processor

- High perf FP16&Int8
- **High mem BW for convolution**
- **Built-in scalable Tofu network**

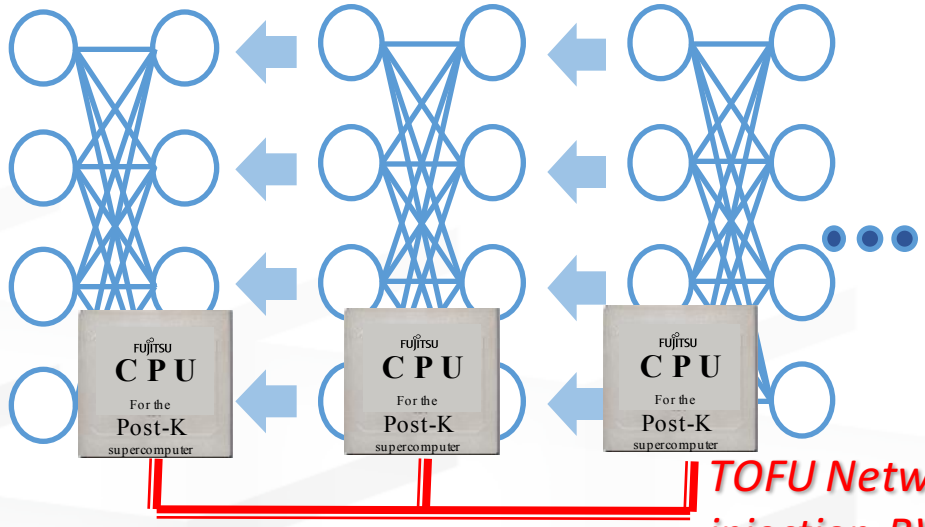
High Performance DNN Convolution



Low Precision ALU + High Memory Bandwidth + Advanced Combining of Convolution Algorithms (FFT+Winograd+GEMM)

Unprecedened DL scalability

High Performance and Ultra-Scalable Network for massive scaling model & data parallelism



Unprecedented Scalability of Data/

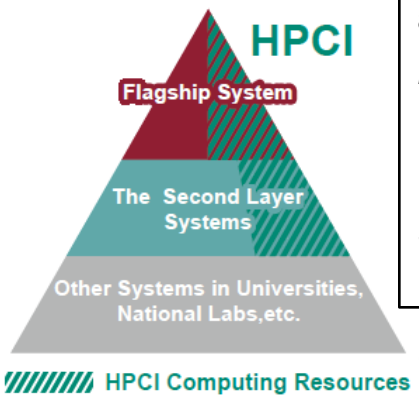
Update of Japanese HPCI

March 12, 2019

Research Organization for Information Science & Technology

What is *HPCI* ?

- World's top class computing resources are open to the world-wide HPC communities



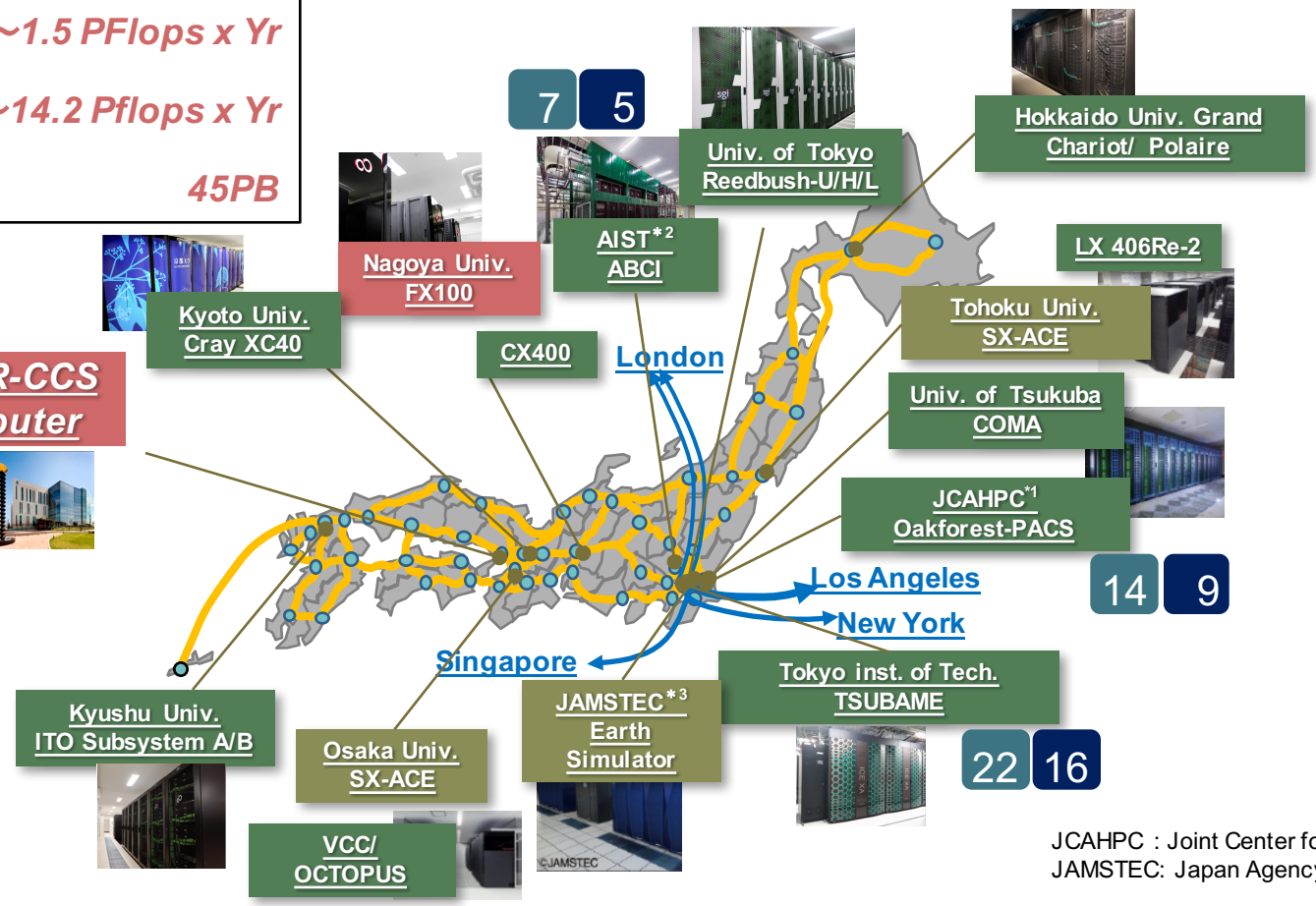
Computing Resources Allocated at Public Calls in FY 2019

K computer
~1.5 PFlops x Yr

Others in total
~14.2 Pflops x Yr

Shared Storage
45PB

Service ends in Aug., 2019



CPU type
- : SPARC
- : SX
- : Xeon

SINET-5
- : 100Gbps
- : Node

International line
- : 100Gbps
- : 10Gbps

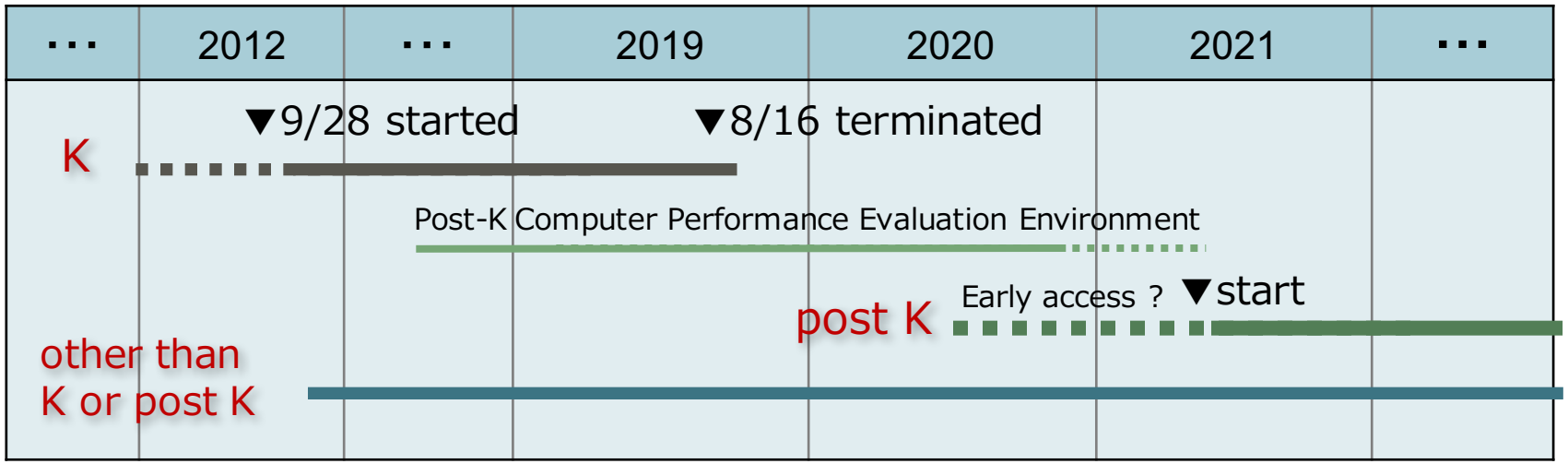
Nov. 2018 Ranking
- n : TOP500
- m : HPCG

JCAHPC : Joint Center for Advanced High Performance Computing
JAMSTEC: Japan Agency for Marine-Earth Science and Technology

As of Jan. 2019

Duration after “K” and before “Post-K”

- HPCI in the absence of the flagship computer
  - To make up the loss, the other computing centers in HPCI provide a larger amount of resource *(14.2 Pflops x Year)*
  - The next ***general and junior researcher promotion projects call*** will be open in Sep. 2019
- Towards post-K
  - Call for project proposals using ***Post-K Computer Performance Evaluation Environment*** is open now
  - The JP government is considering on an early-access project to the post-K computer



Project Categories and Timing of Public Calls for Proposals (FY2019)

Project Categories (Period)		Duration	K computer*1 [Max resources/year]	Other HPCI computers
General Projects (One year)		One year	until Aug. 16, 2019	once a year
	General Trial Use (Six months)	Half-year in max	year-roun… [50,00…]	---
Junior Researcher Promotion Projects		One year		once a year
Industrial Use projects	Non-proprietary Use	One y…		once a year
	Industrial Trial Use*3	Depend on each resource	year-round [Up to 50,000 NH]	year-round *3
	Proprietary Use (usage fee is charged)	One year in max	year-round [Up to 6 million NH]	year-round *3
HPCI Shared Storage		One year in max	year-round [Up to 1.5PB]	

Next call will start in Sep. 2019

New !

*1 : The operation of K computer will stop in FY 2019.
*2 : Period A (from April 1 to March 31) Period B (from October 1 to September 30)
*3 : Some Resource-Providers do not accept projects in these categories

Periodic calls
Calls opening year-round

Important note:
In the case where applicants belong to foreign companies...
proposals must be submitted jointly with project representatives who belong to corporations registered in Japan.

Open call for project proposals using

Post-K Computer Performance Evaluation Environment

- Who are the targets?
  - Potential users of post-K computer, desiring to port, develop, or optimize programs towards the post-K computer
- What can you do using the Environment?
  - Approximate performance of programs is attainable
  - The Environment is mainly consist of:
    - "Processor simulators"
    - "Performance Estimation Tools"
    - "Compilers" for the post-K computer (Fortran, C/C++)
- Project proposals now calling
  - The project period is **up to 6 months**
  - The call is **open throughout the year**
  - Further information is found in HPCI Portal
  - RIST provides technical supports

http://www.hpci-office.jp/folders/english