FUJITSU

shaping tomorrow with you

Post-K Supercomputer with Fujitsu's Original CPU, A64FX Powered by Arm ISA

Toshiyuki Shimizu

Nov. 15th, 2018

Post-K is under development, information in these slides is subject to change without notice

Copyright 2018 FUJITSU LIMITED

Agenda

FUJITSU

- Background
  - K computer and Post-K
  - Post-K goals and approaches
- Fujitsu new Arm CPU A64FX
- Post-K system
  - System software
  - Configuration
- Performance discussion
- Summary and development status

Copyright 2018 FUJITSU LIMITED

Agenda

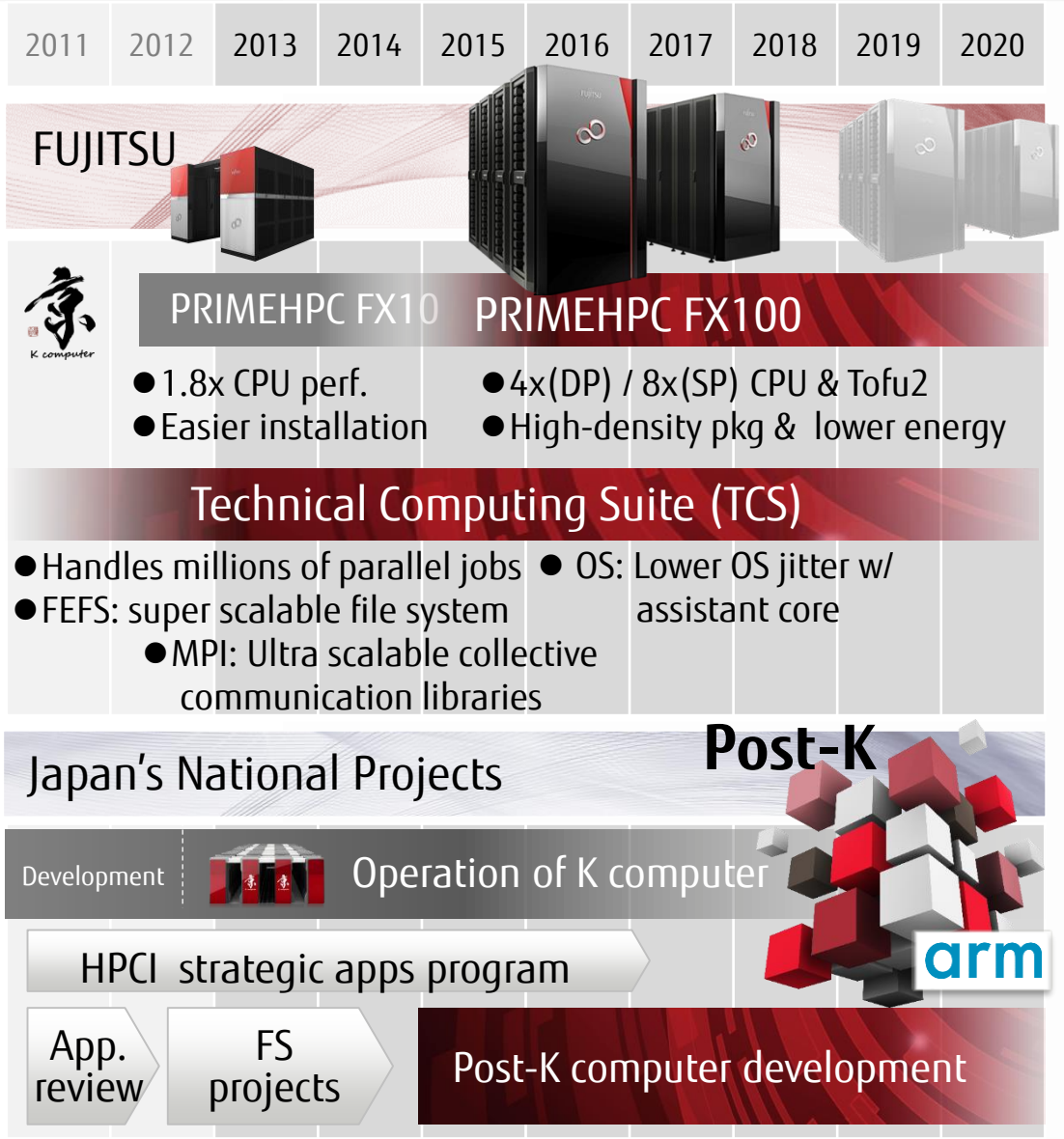
FUJITSU

- Background
  - K computer and Post-K
  - Post-K goals and approaches
- Fujitsu new Arm CPU A64FX
- Post-K system
  - System software
  - Configuration
- Performance discussion
- Summary and development status

Copyright 2018 FUJITSU LIMITED

K computer, PRIMEHPC, and Post-K

FUJITSU



K computer, PRIMEHPC FX

- Many applications are running and being developed for science and various industries
- System software TCS supports hardware with newly developed technologies

Post-K

- RIKEN and Fujitsu are working together for Post-K
- OS is running and design verification is proceeding as scheduled

Copyright 2018 FUJITSU LIMITED

Post-K goals and approaches

FUJITSU

- Post-K goals
  - High application performance and good power efficiency
  - Good usability and better accessibility for users
  - Keeping application compatibility while advancing from predecessors
- Our approaches
  - Develops high performance and scalable, custom CPU cores
    - 【Performance】 Wider SIMD & mathematical acc. primitives, high memory BW
    - 【Scalability】 Scalable interconnect “Tofu”
    - 【Power efficiency】 The best device tech, power control functions, optimal resources
  - Adopts Arm ISA, binary compatibility
  - Maintains performance balance and supports advanced features

Copyright 2018 FUJITSU LIMITED

Agenda

FUJITSU

- Background
  - K computer and Post-K
  - Post-K goals and approaches
- Fujitsu new Arm CPU A64FX
- Post-K system
  - System software
  - Configuration
- Performance discussion
- Summary and development status

Copyright 2018 FUJITSU LIMITED

Fujitsu new Arm CPU: A64FX

- A64FX approach
- Specs and technologies
- Tofu interconnect D
- Power management
- A64FX summary

Copyright 2018 FUJITSU LIMITED

A64FX: Approach

FUJITSU

- High performance Arm CPU
  - Develops original CPU core supporting Arm SVE (Scalable Vector Extension)
- Targeting high performance servers
  - Floating point calculations
  - High memory bandwidth
  - Low power consumption
  - Scalable performance and configuration
- Extension for emerging applications
  - Half precision (FP16) & INT16/8 partial dot product support

High BW/Calc. is the key for Real apps.

Copyright 2018 FUJITSU LIMITED

A64FX: Specs & technologies

FUJITSU

- CPU generations and key parameters

CPU	SPARC64 VIIIfx	SPARC64 IXfx	SPARC64 XIfx		A64FX
1st system w/ CPU	K	FX10	FX100		Post-K
Si tech. (nm)	45	40	20		7
Core perf. (GFLOPS)	16	14.8	34	SVE	57~
# of cores	8	16	32		48
Chip perf. (TFLOPS)	0.13	0.24	1.1	CMG	2.7~
Memory BW (GB/s)	64	85	480	HBM	1024
B/F (Bytes/FLOP)	0.5	0.4	0.4		0.4

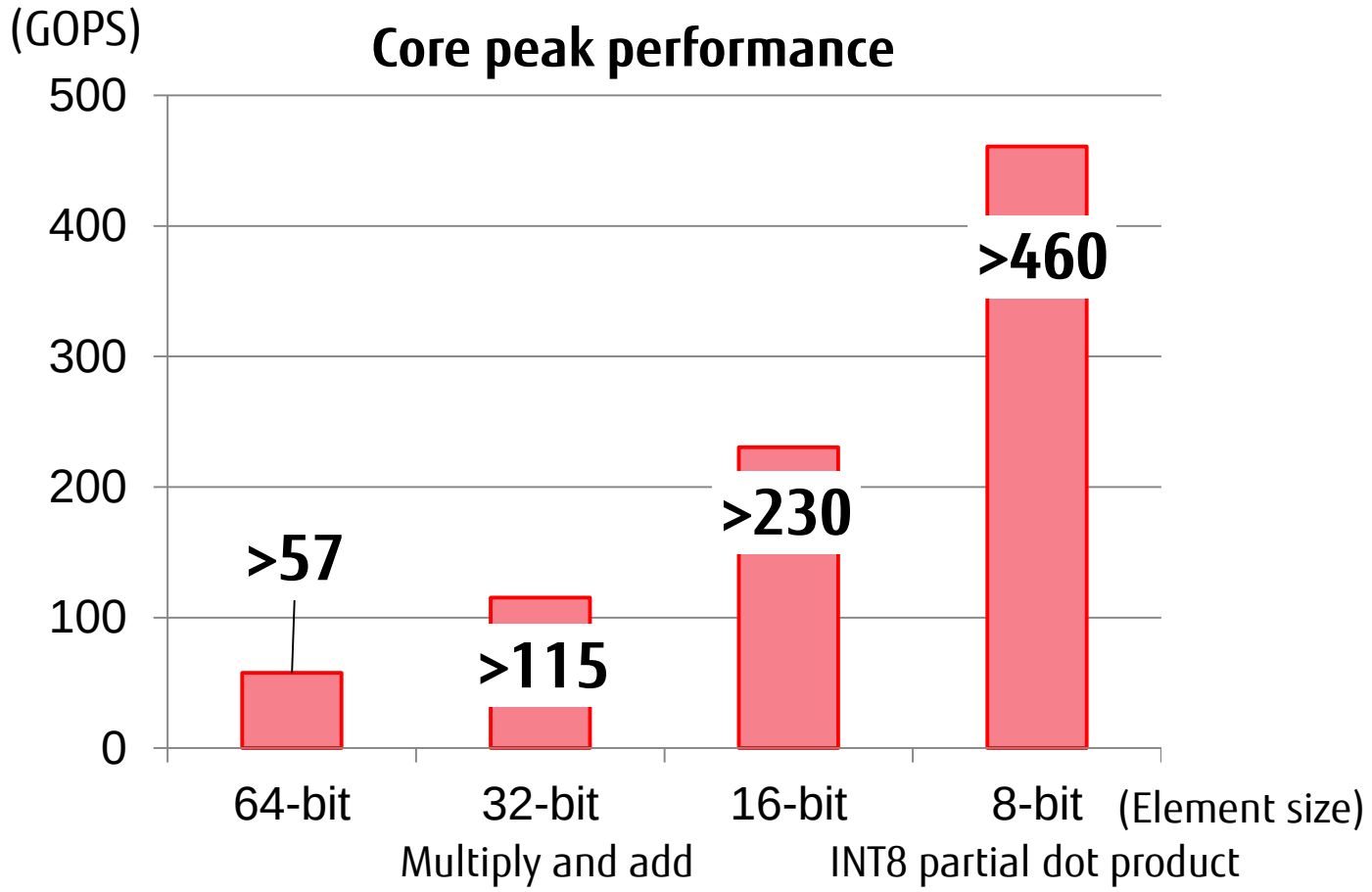
- Combination of technologies
  - SVE increases core performance
  - CMG is scalable architecture to increase # of cores
  - HBM enables high bandwidth

Copyright 2018 FUJITSU LIMITED

A64FX technologies: Core performance

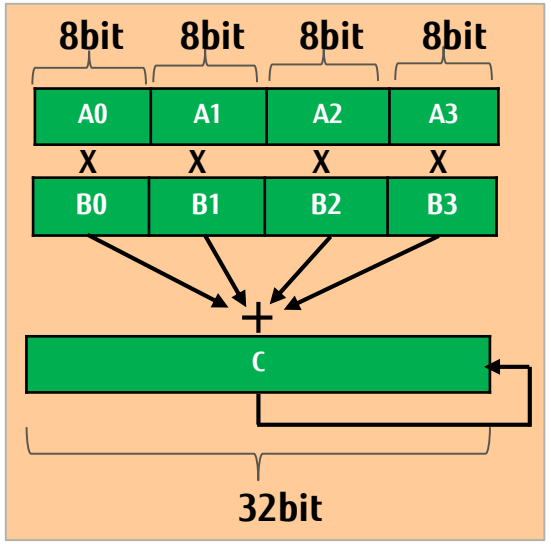
FUJITSU

- High calc. throughput of Fujitsu original CPU core w/ SVE
  - 512-bit wide SIMD x 2 pipelines and new integer functions



INT8 partial dot product

$C = \Sigma\ (Ai \times Bi) + C$



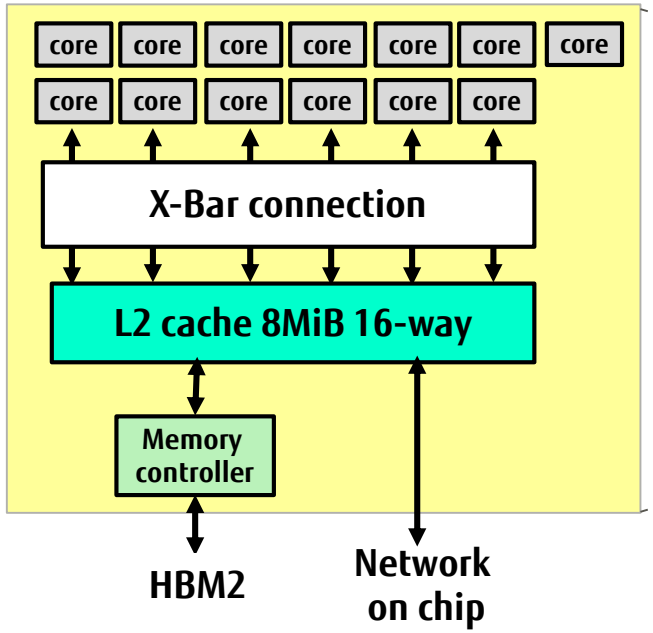
Copyright 2018 FUJITSU LIMITED

A64FX technologies: Scalable architecture

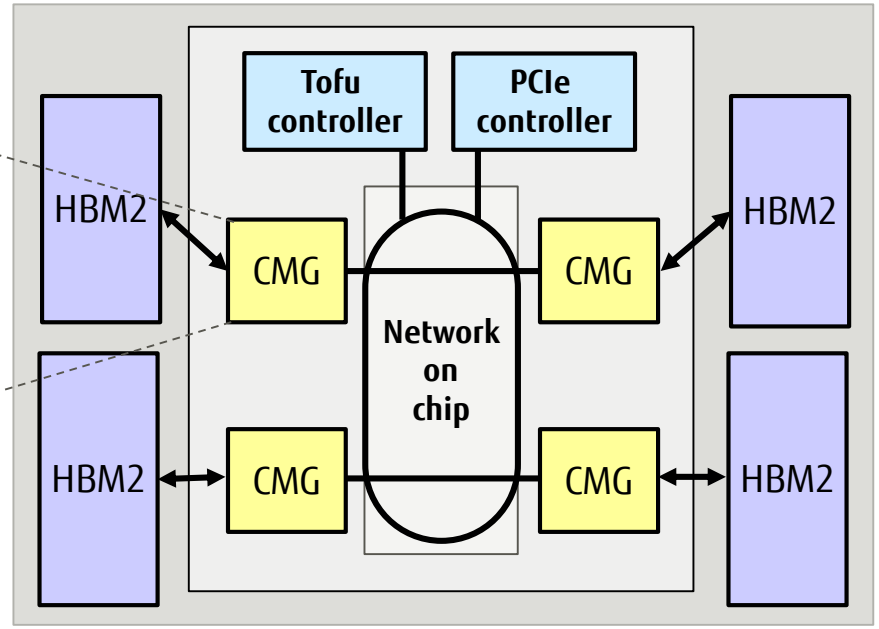
FUJITSU

- Core Memory Group (CMG)
  - 12 compute cores and an assistant core for OS daemon, I/O, etc.
  - Shared L2 cache
  - Dedicated memory controller
- Four CMGs maintain cache coherence w/ on-chip directory
  - Threads binding within a CMG allows linear speed of up to 48 compute cores

CMG configuration



A64FX package configuration

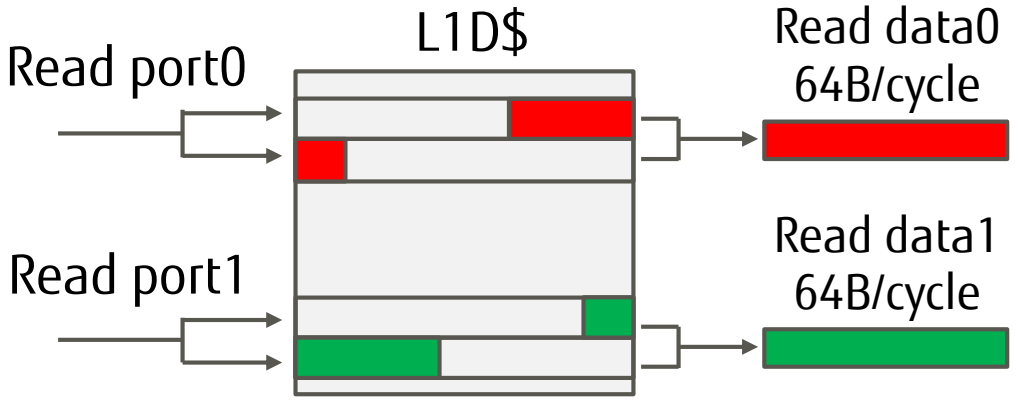


Copyright 2018 FUJITSU LIMITED

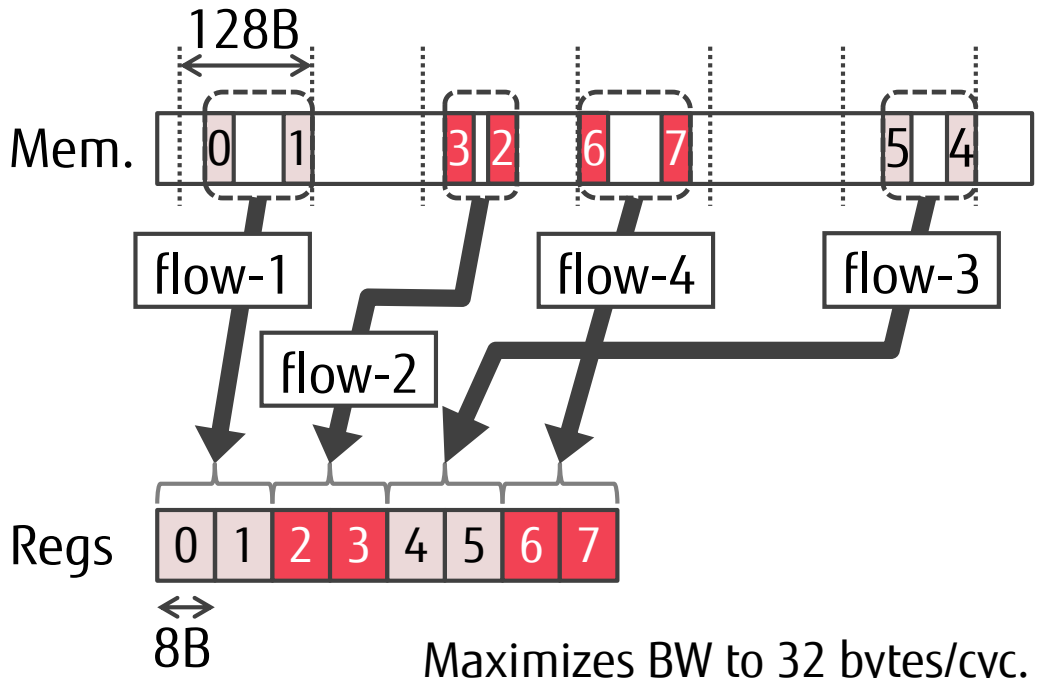
A64FX: L1D cache uncompromised BW

FUJITSU

- 128B/cycle sustained BW even for unaligned SIMD load



- "Combined Gather" doubles gather (indirect) load's data throughput, when target elements are within a "128-byte aligned block" for a pair of two regs, even & odd



Maximizes BW to 32 bytes/cyc.

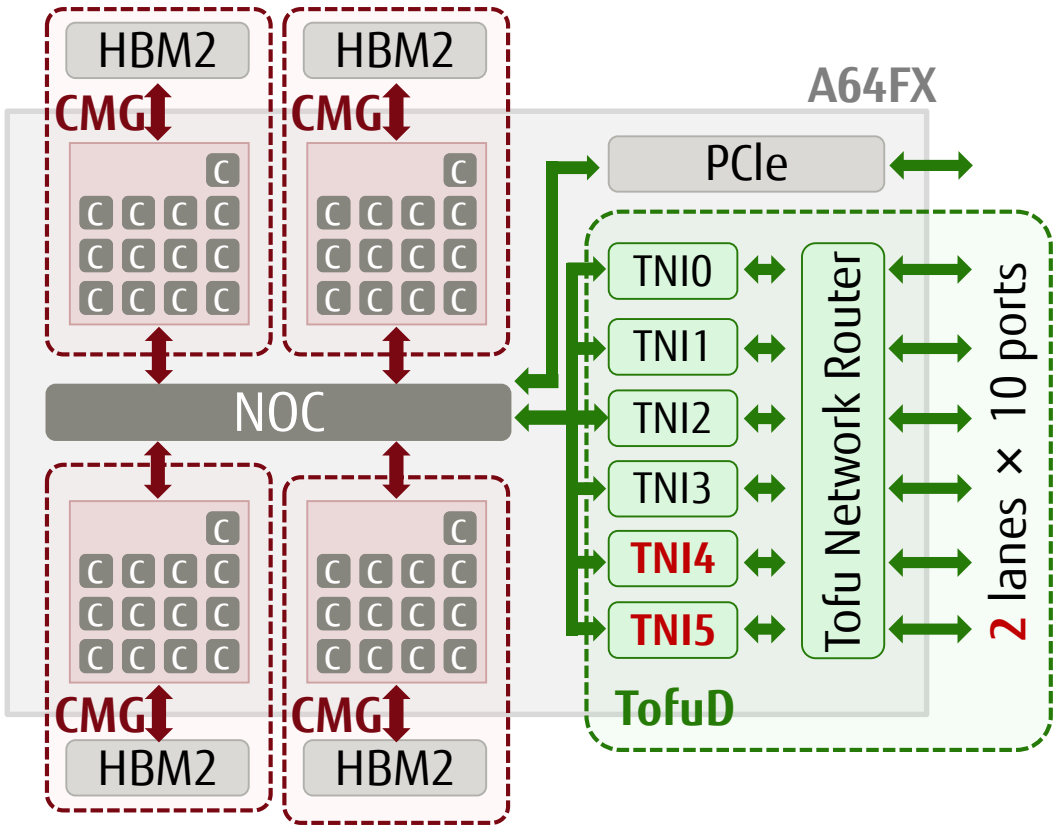
Copyright 2018 FUJITSU LIMITED

A64FX: Tofu interconnect D

FUJITSU

- Integrated w/ rich resources
  - Increased TNIs achieves higher injection BW & flexible comm. patterns
  - Increased barrier resources allow flexible collective comm. algorithms
- Memory bypassing achieves low latency
  - Direct descriptor & cache injection

	TofuD spec
Data rate	28.05 Gbps
Link bandwidth	6.8 GB/s
Injection bandwidth	40.8 GB/s
	Measured
Put throughput	6.35 GB/s
PingPong latency	0.49~0.54 μs



Copyright 2018 FUJITSU LIMITED

A64FX: Power monitor and analyzer

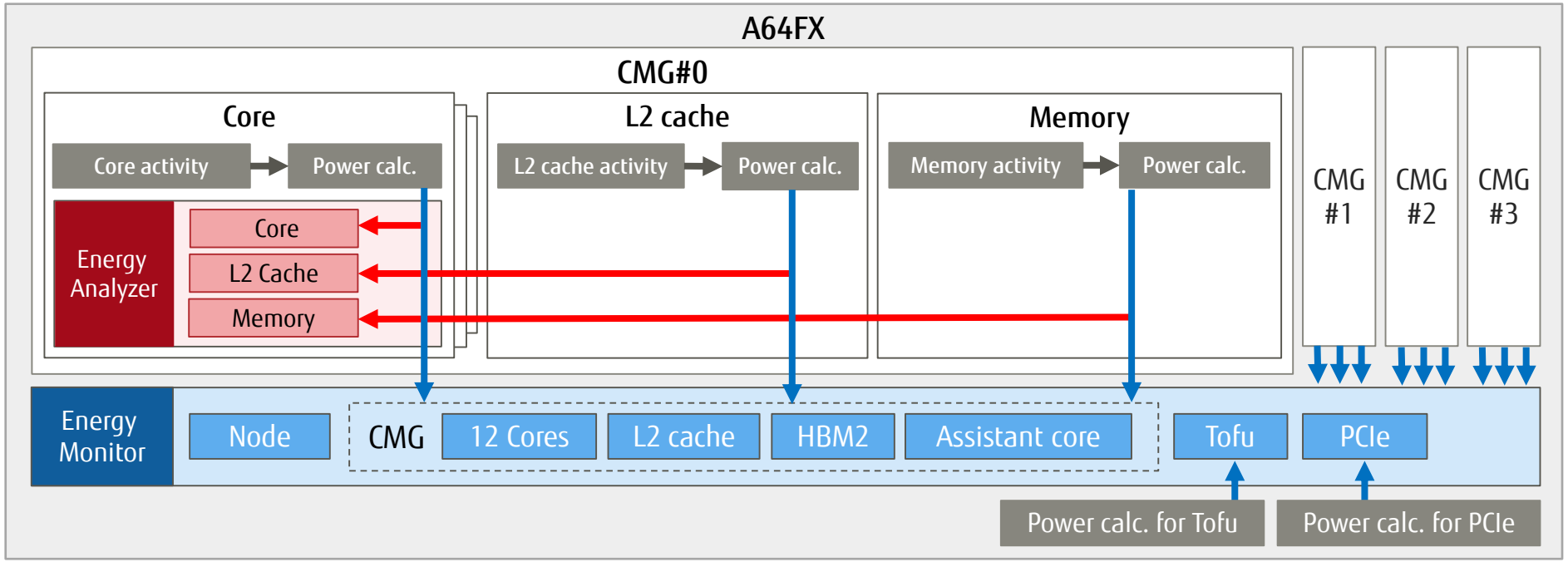
FUJITSU

- Energy monitor (per chip)
  - Node power via Power API*1 (~msec)
  - Averaged power of a node, CMG (cores, an L2 cache, a memory) etc.

*1: Sandia National Laboratory

- Energy analyzer (per core)
  - Power profiler via PAPI*2 (~nsec)
  - Fine grained power analysis of a core, an L2 cache and a memory

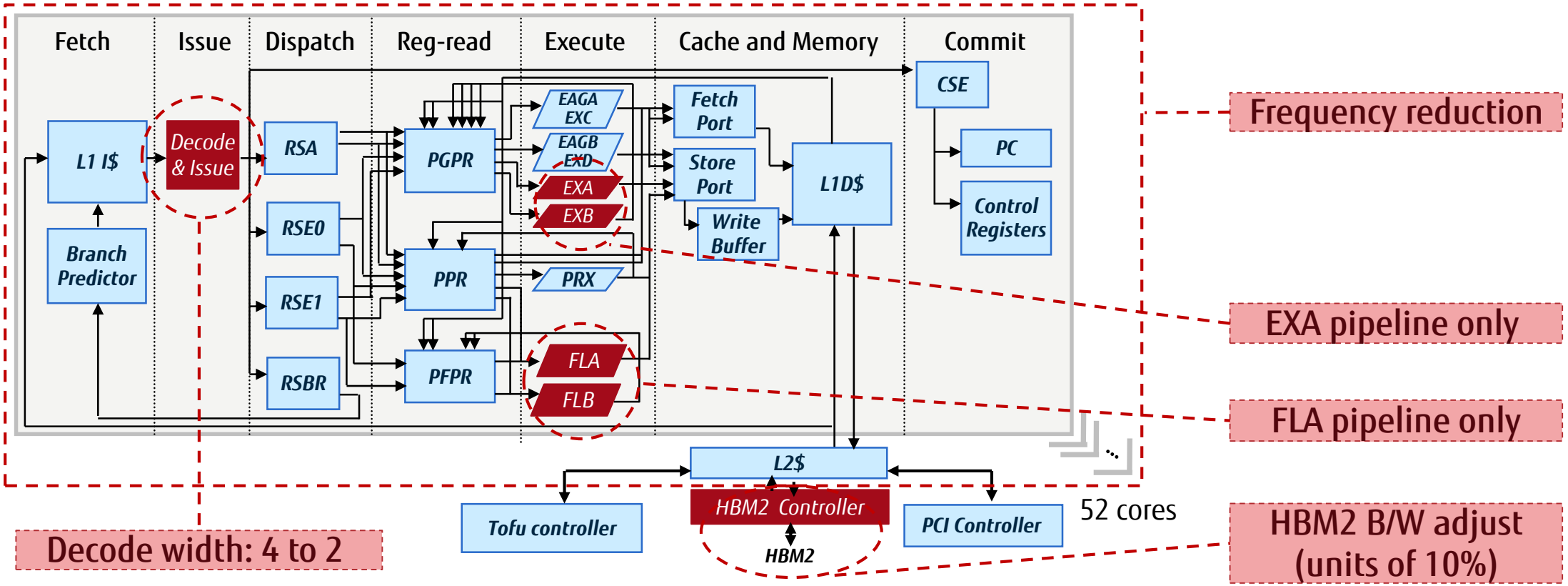
*2: Performance Application Programming Interface



Copyright 2018 FUJITSU LIMITED

A64FX: Power Knobs to reduce power consumption

- "Power knob" limits units' activity via user APIs
- Performance/W would be optimized by utilizing Power knobs, Energy monitor & analyzer



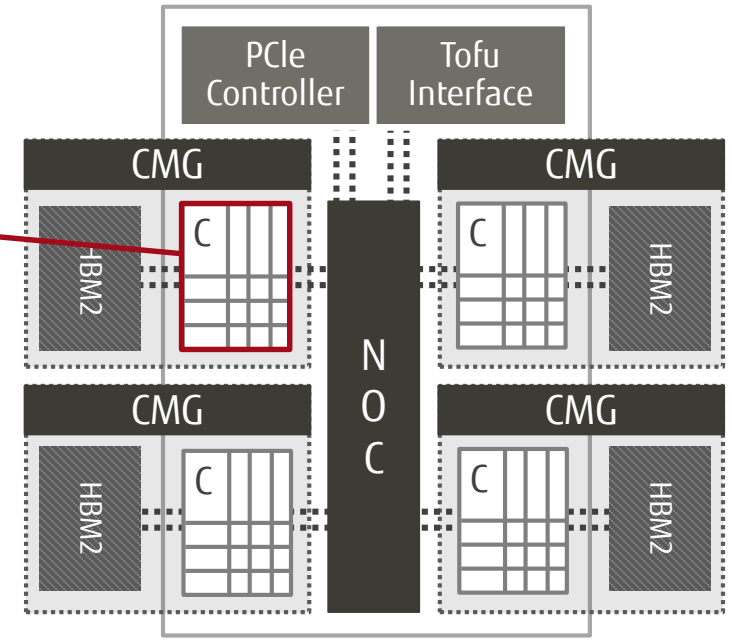
Copyright 2018 FUJITSU LIMITED

A64FX: Summary

FUJITSU

- Arm SVE, high performance and efficiency
  - DP performance >2.7 TFLOPS, >90%@DGEMM
  - Memory BW 1024 GB/s, >80%@STREAM Triad

CMG
12x compute cores
1x assistant core



CMG : Core Memory Group NOC : Network on Chip

	A64FX
ISA (Base, extension)	Armv8.2-A, SVE
Process technology	7 nm
Peak DP performance	>2.7 TFLOPS
SIMD width	512-bit
# of compute cores	48
Memory capacity	32 GiB (HBM2 x4)
Memory peak bandwidth	1024 GB/s
PCIe	Gen3 16 lanes
High speed interconnect	TofuD integrated

Copyright 2018 FUJITSU LIMITED

Agenda

FUJITSU

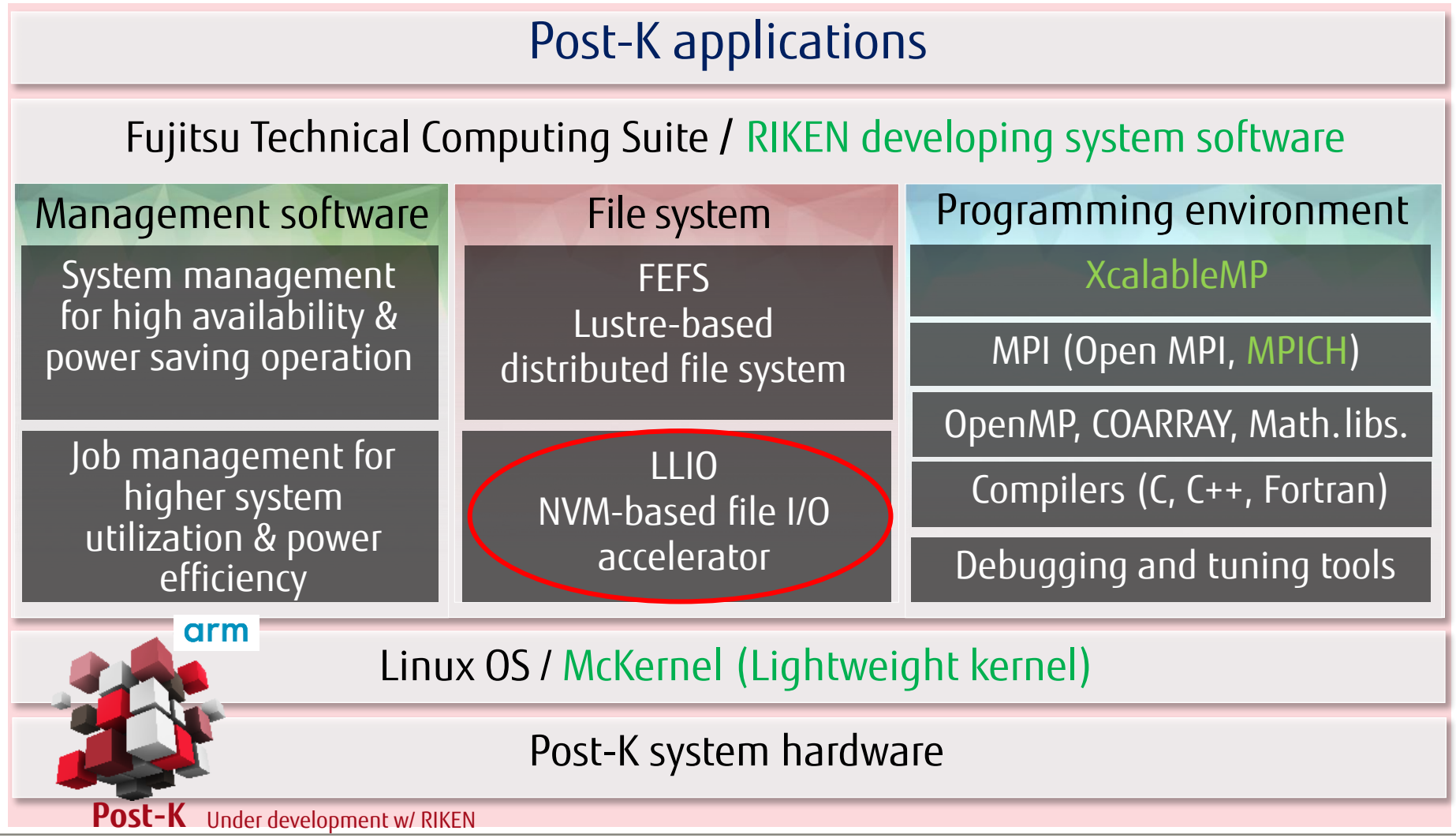
- Background
  - K computer and Post-K
  - Post-K goals and approaches
- Fujitsu new Arm CPU A64FX
- Post-K system
  - System software
  - Configuration
- Performance discussion
- Summary and development status

Copyright 2018 FUJITSU LIMITED

Software development: System software

FUJITSU

- RIKEN and Fujitsu are developing software stacks for Post-K



Copyright 2018 FUJITSU LIMITED

FEFS vs Lustre, LLIO vs Burst buffer

FUJITSU

- Post-K supports superior performance file system, FEFS + LLIO

	FEFS	Lustre
Storage network	Tofu, InfiniBand, Ether, OPA	InfiniBand, Ether, OPA
Injection BW per node	40.8 GB/s	12.5 GB/s (IB EDR)
MDS scalability	Yes	Yes (ver. 2.4~)
Interoperability	x86, Arm (A64FX)	x86, Power, Arm?
	FEFS + LLIO*	**Lustre + Burst buffer****
Local temporary file acc.	Yes	No
Cache file acc.	Yes	Yes
Shared file within a job.	Yes	Yes
Explicit function call	Not required (by mount point)	Required (Source modification)

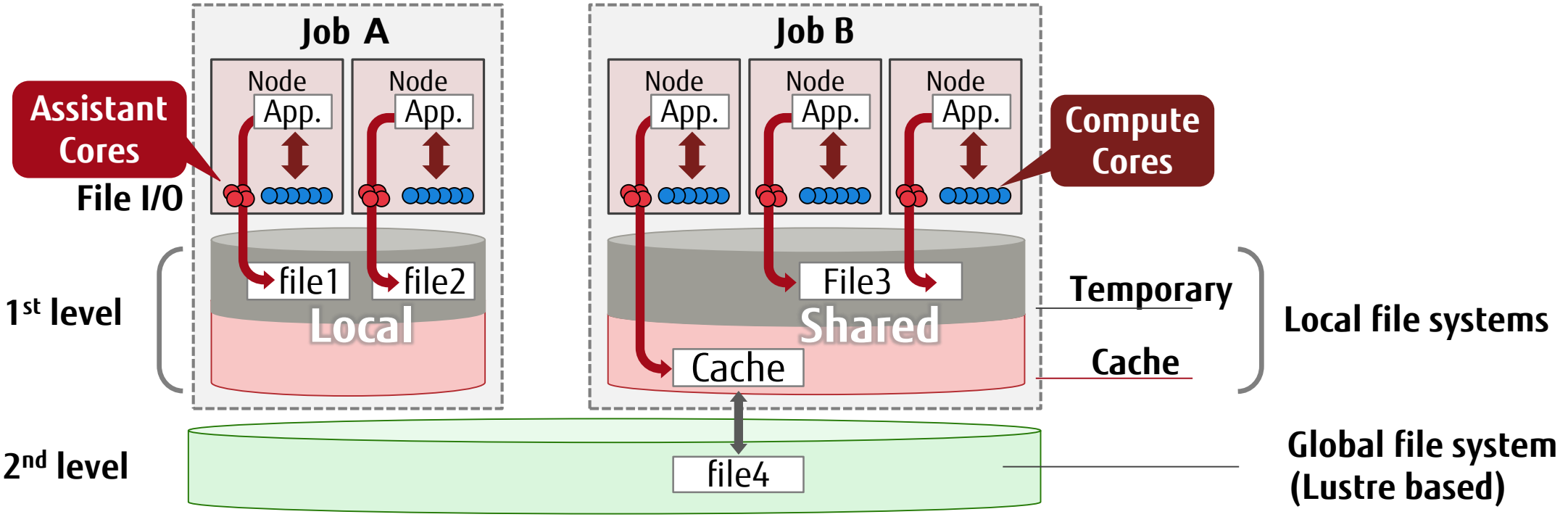
*Lightweight Layered IO-Accelerator **DDN IME

Copyright 2018 FUJITSU LIMITED

LLIO (Lightweight Layered IO-accelerator) overview

FUJITSU

- Maximizing application file I/O performance
  - Easy access to User Data: File Cache of Global File System
  - Higher Data Access Performance: Temporary Local FS (in a process)
  - Higher Data Sharing Performance: Temporary Shared FS (among processes)

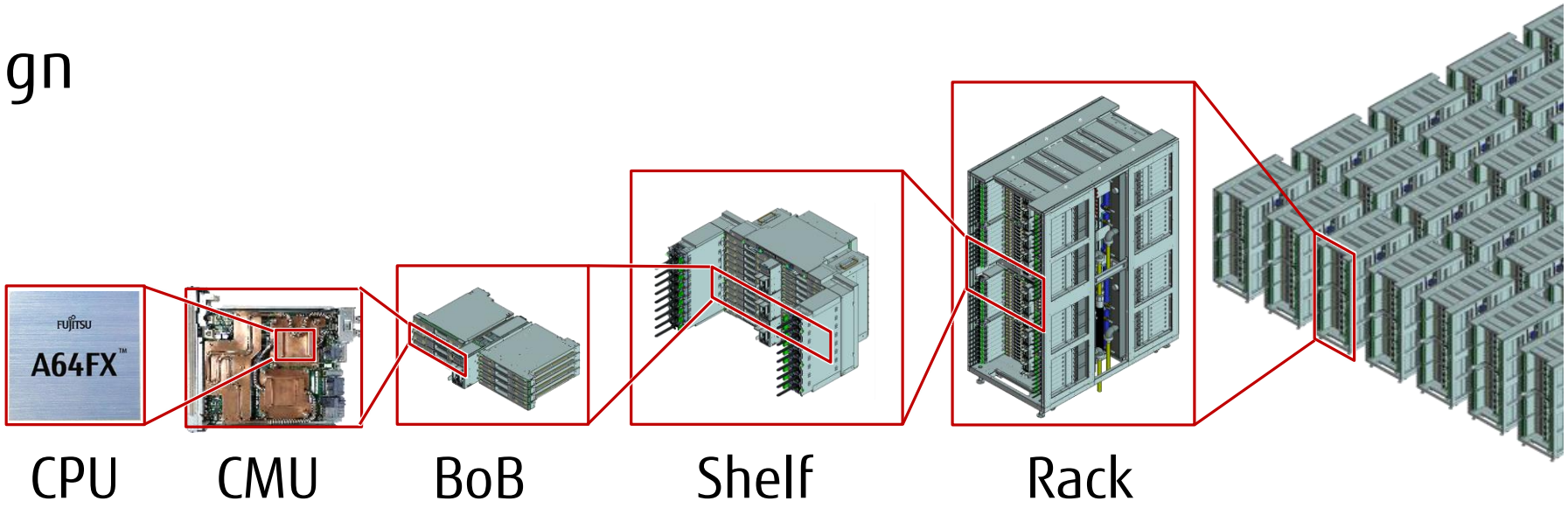


Copyright 2018 FUJITSU LIMITED

Post-K system configuration

FUJITSU

Scalable design

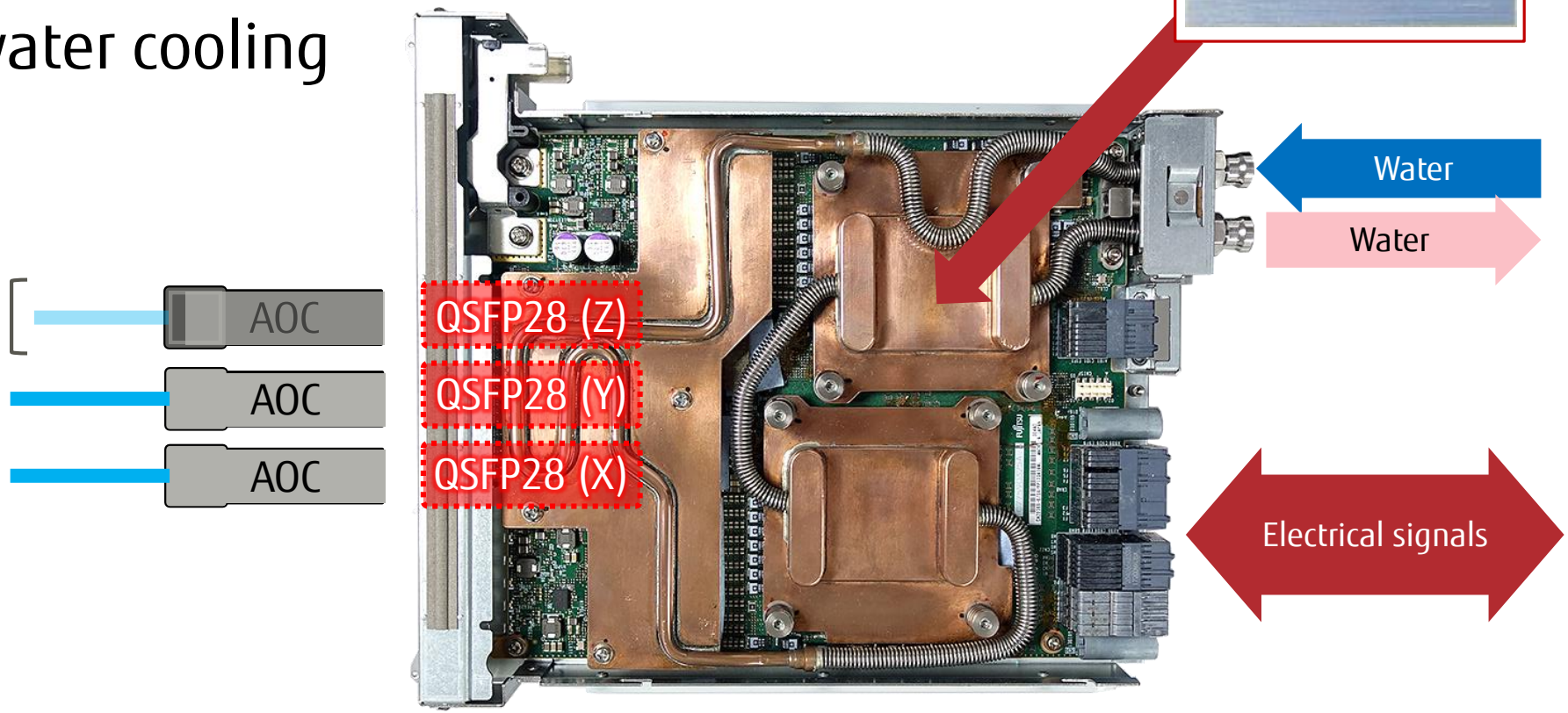


Unit	# of nodes	Description
CPU	1	Single socket node with HBM2 & Tofu interconnect D
CMU	2	CPU Memory Unit: 2x CPU
BoB	16	Bunch of Blades: 8x CMU
Shelf	48	3x BoB
Rack	384	8x Shelf

Copyright 2018 FUJITSU LIMITED

CMU: CPU Memory Unit

- A64FX CPU x2
- QSFP28 x3 for Active Optical Cables
- Single-side blind mate of signals & water
- ~100% direct water cooling



Copyright 2018 FUJITSU LIMITED

Agenda

FUJITSU

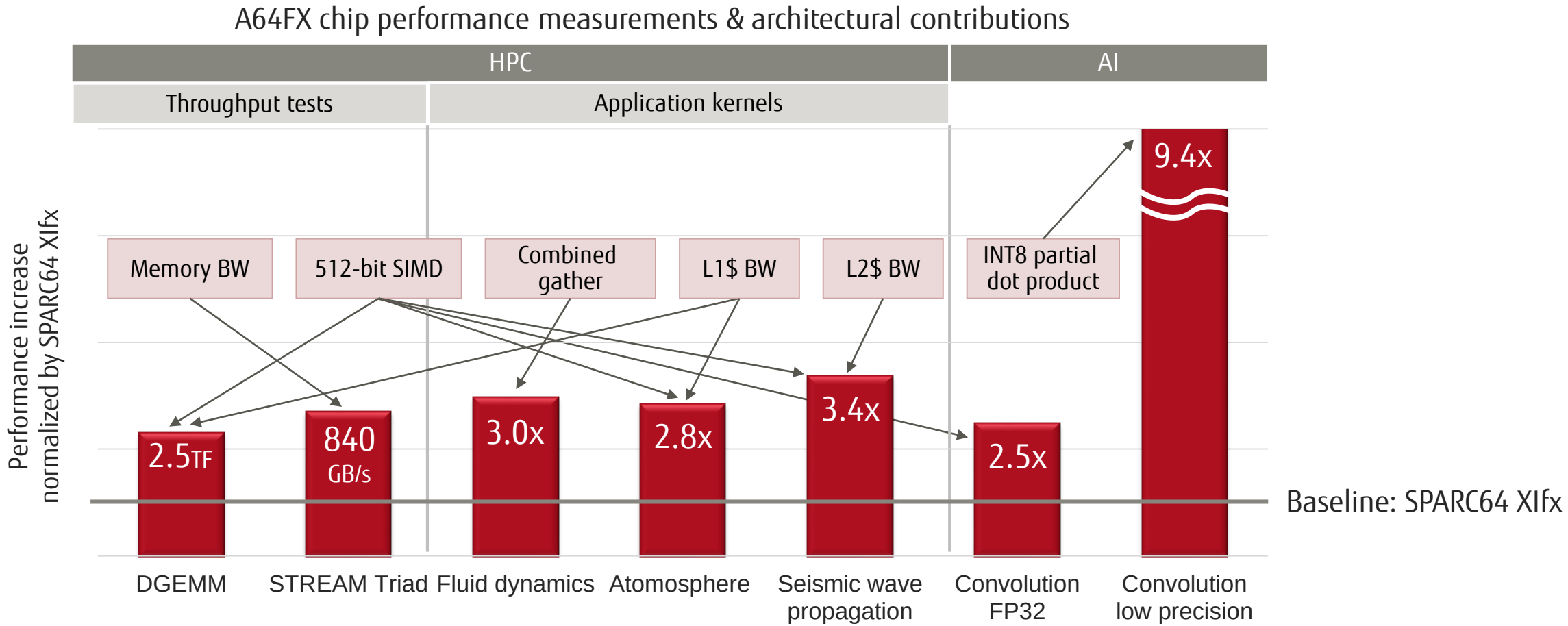
- Background
  - K computer and Post-K
  - Post-K goals and approaches
- Fujitsu new Arm CPU A64FX
- Post-K system
  - System software
  - Configuration
- Performance discussion
- Summary and development status

Copyright 2018 FUJITSU LIMITED

Preliminary performance evaluation results

FUJITSU

- Over 2.5x faster in HPC & AI benchmarks than SPARC64 XIfx
  - Arm SVE, Fujitsu's microarchitecture, and high bandwidth



Copyright 2018 FUJITSU LIMITED

Agenda

FUJITSU

- Background
  - K computer and Post-K
  - Post-K goals and approaches
- Fujitsu new Arm CPU A64FX
- Post-K system
  - System software
  - Configuration
- Performance discussion
- Summary and development status

Copyright 2018 FUJITSU LIMITED

Summary of Post-K

FUJITSU

- High application performance and good power efficiency
  - High memory bandwidth & wider SIMD
- Arm SVE support
  - Scalable vector extension, state of art Arm instruction set architecture
  - Binary compatibility advances and expands the power of ecosystem by incorporating HPC technologies and applications
- Optimizing for new supercomputer standards
  - Low power consumption in many aspects
  - Being focusing on the existing applications and emerging applications

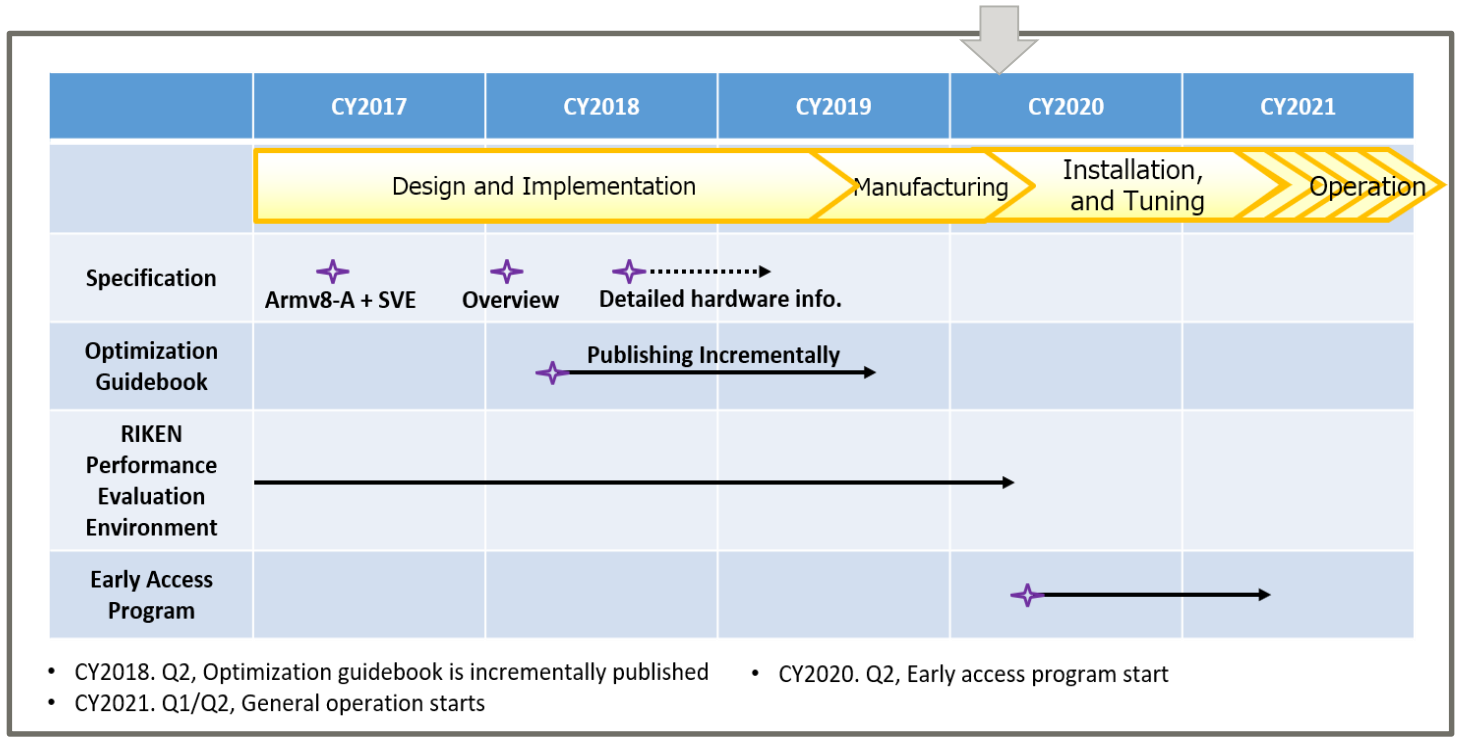
Post-K inheriting the strong and proven microarchitecture with HPC system software will be more powerful with Arm and its ecosystem

Copyright 2018 FUJITSU LIMITED

System development status: Post-K

FUJITSU

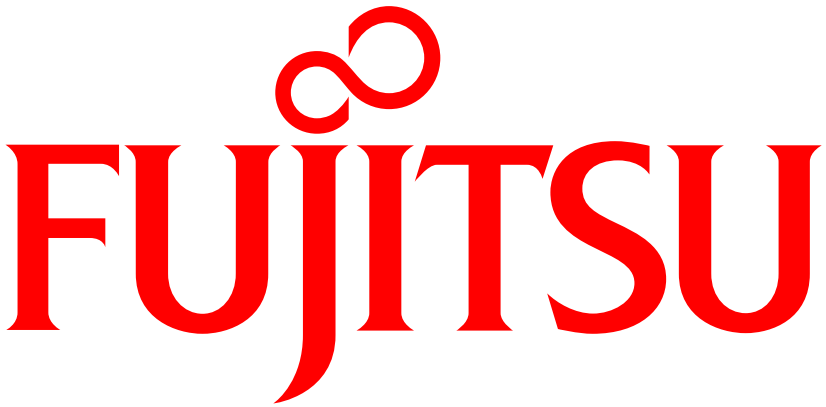
- Fujitsu is currently developing Post-K, successor to K computer, with RIKEN
- OS running and design verification is underway as scheduled
- RIKEN announced Post-K early access program to begin around Q2/CY2020



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

Post-K prototype rack

Copyright 2018 FUJITSU LIMITED



shaping tomorrow with you