

Fujitsu High Performance CPU for the Post-K Computer

August 21st, 2018

Toshio Yoshida

FUJITSU LIMITED

Key Message

- A64FX is the new Fujitsu-designed Arm processor
 - It is used in the post-K computer
- A64FX is the first processor of the Armv8-A SVE architecture
 - Fujitsu, as a lead partner, collaborated closely with Arm on the development of SVE
- A64FX achieves high performance in HPC and AI areas
 - Our own microarchitecture maximizes the capability of SVE

■ Fujitsu Processor Development

■ A64FX

- Overview

- Microarchitecture

- Performance

- Power Management

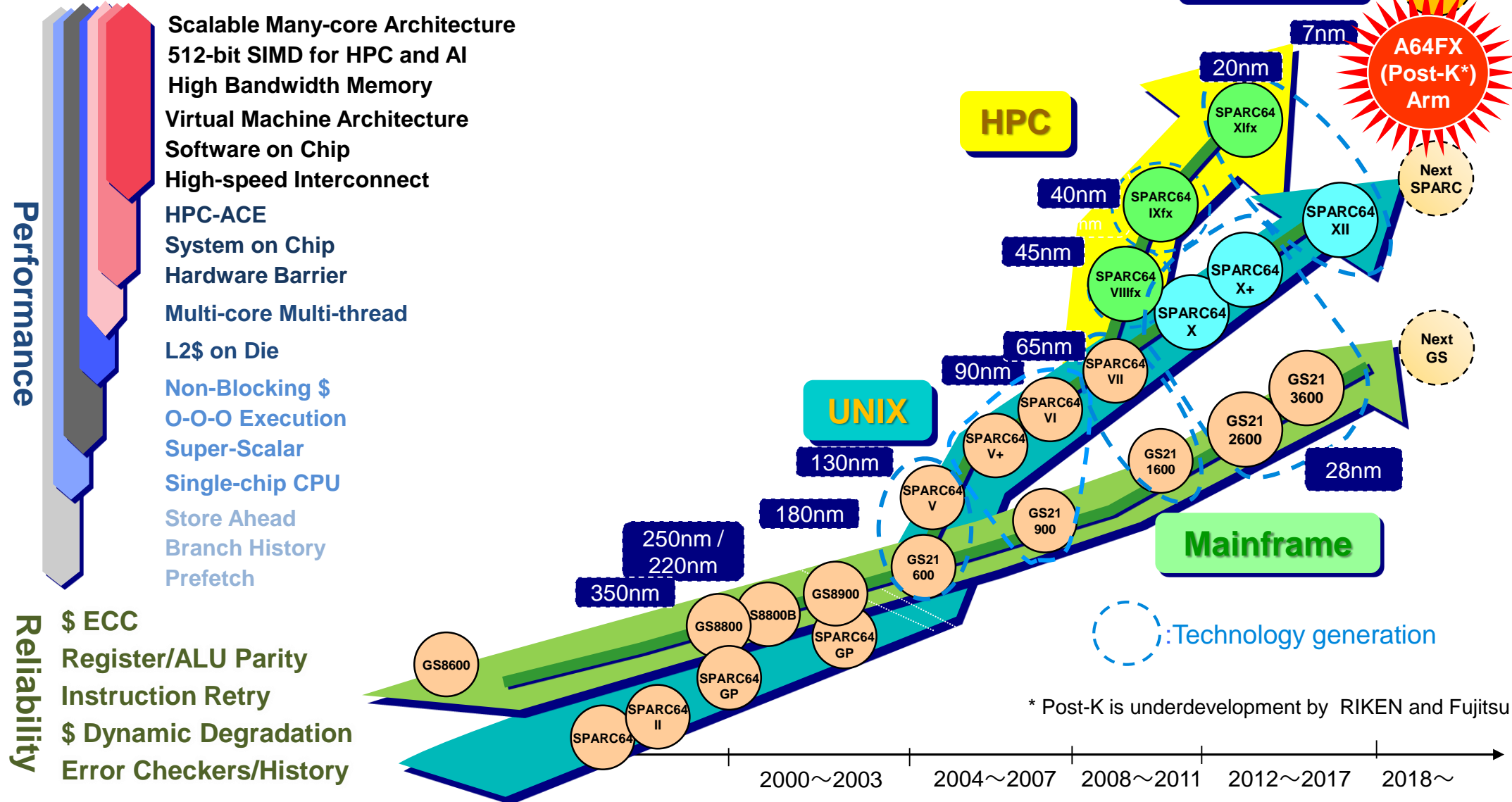
- RAS

■ Software Development

■ Summary

Fujitsu Processor Development

Persistent Evolution > 60 years



DNA of Fujitsu Processors

- A64FX inherits DNA from Fujitsu technologies used in the mainframes, UNIX and HPC servers



High reliability

Stability
Integrity
Continuity



High speed & flexibility

Thread performance
Software on Chip
Large SMP



High performance-per-watt

Execution and memory throughput
Low power
Massively parallel



CPU w/ extremely high throughput

High performance
Massively parallel
Low power
Stability and integrity

A64FX Designed for HPC/AI

A64FX = CPU with extremely high throughput

1. High Performance

HPC/AI apps. >> General purpose CPU
Various data types
(FP64/32/16, INT64/32/16/8)

2. High Throughput

Vector : 512-bit wide SIMD x 2 pipes /core
Memory : HBM2 (extremely high B/W)
Scalable : 48 cores, Tofu interconnect

3. High Efficiency

Performance
(D|S|H)GEMM >90%
Stream Triad >80%
Perf-per-watt >> General purpose CPU

4. Standard

Binary compatibility with
Armv8.2-A + SVE + SBSA* level3

*Arm's "Server Base System Architecture"

A64FX Chip Overview

Architecture Features

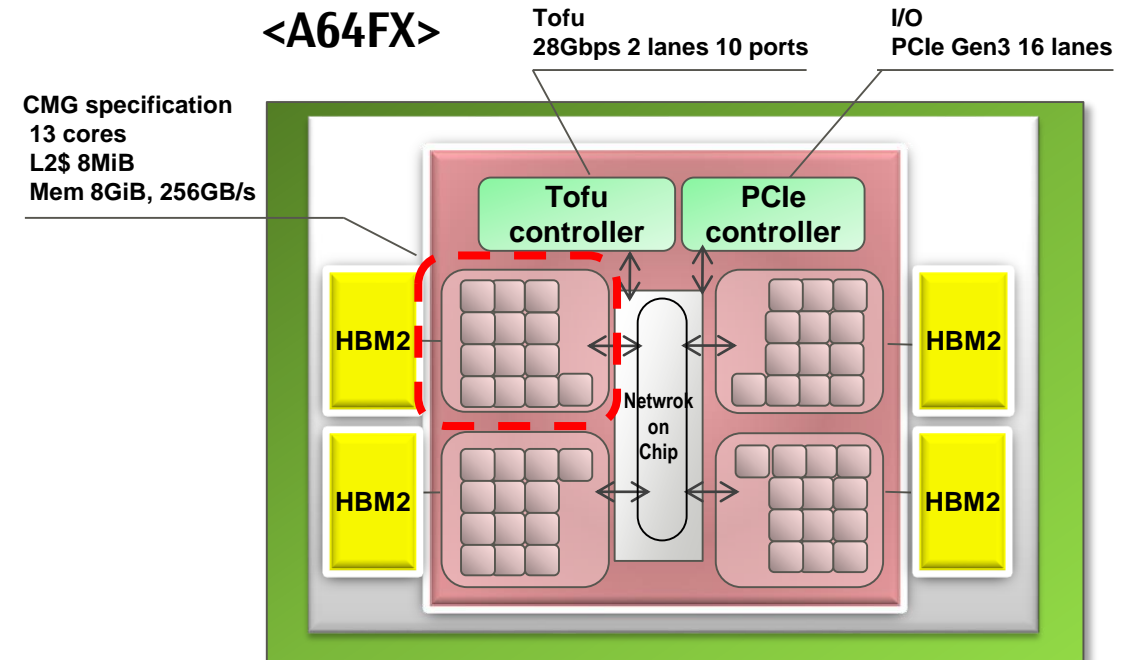
- Armv8.2-A (AArch64 only)
- SVE 512-bit wide SIMD
- 48 computing cores + 4 assistant cores*
*All the cores are identical
- HBM2 32GiB
- Tofu 6D Mesh/Torus
28Gbps x 2 lanes x 10 ports
- PCIe Gen3 16 lanes

7nm FinFET

- 8,786M transistors
- 594 package signal pins

Peak Performance (Efficiency)

- >2.7TFLOPS (>90%@DGEMM)
- Memory B/W 1024GB/s (>80%@Stream Triad)



	A64FX (Post-K)	SPARC64 Xlfx (PRIMEHPC FX100)
ISA (Base)	Armv8.2-A	SPARC-V9
ISA (Extension)	SVE	HPC-ACE2
Process Node	7nm	20nm
Peak Performance	>2.7TFLOPS	1.1TFLOPS
SIMD	512-bit	256-bit
# of Cores	48+4	32+2
Memory	HBM2	HMC
Memory Peak B/W	1024GB/s	240GB/s x2 (in/out)

A64FX Features

- Collaboration with Arm to develop and optimize SVE for a wide range of applications
 - FP16 and INT16/8 dot product are introduced for AI applications

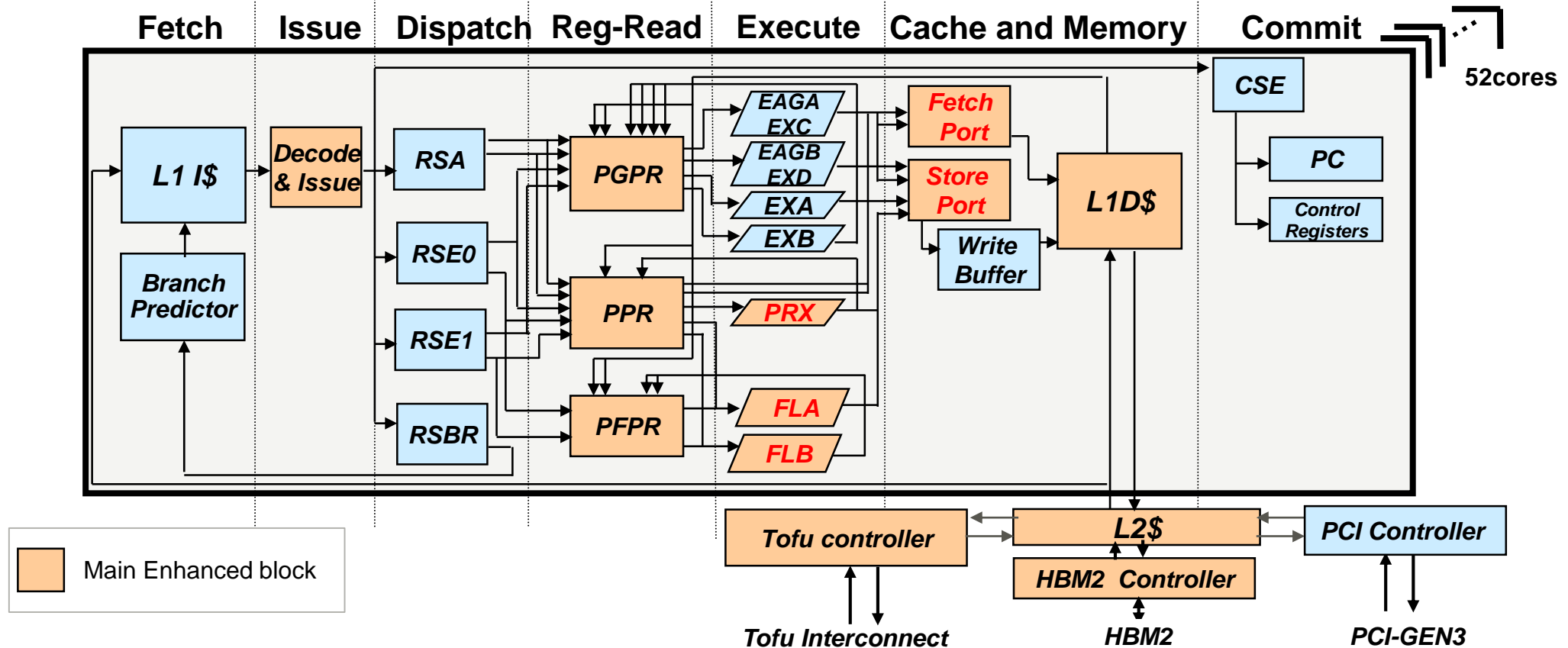
	A64FX (Post-K)	SPARC64 Xlfx (PRIMEHPC FX100)	SPAR64 VIIIfx (K computer)
ISA	Armv8.2-A + SVE	SPARC-V9 + HPC-ACE2	SPARC-V9 + HPC-ACE
SIMD Width	512-bit	256-bit	128-bit
Four-operand FMA	✓ Enhanced	✓	✓
Gather/Scatter	✓ Enhanced	✓	
Predicated Operations	✓ Enhanced	✓	✓
Math. Acceleration	✓ Further enhanced	✓ Enhanced	✓
Compress	✓ Enhanced	✓	
First Fault Load	✓ New		
FP16	✓ New		
INT16/ INT8 Dot Product	✓ New		
HW Barrier* / Sector Cache*	✓ Further enhanced	✓ Enhanced	✓

* Utilizing AArch64 implementation-defined system registers

A64FX Core Pipeline

■ A64FX enhances and inherits superior features of SPARC64

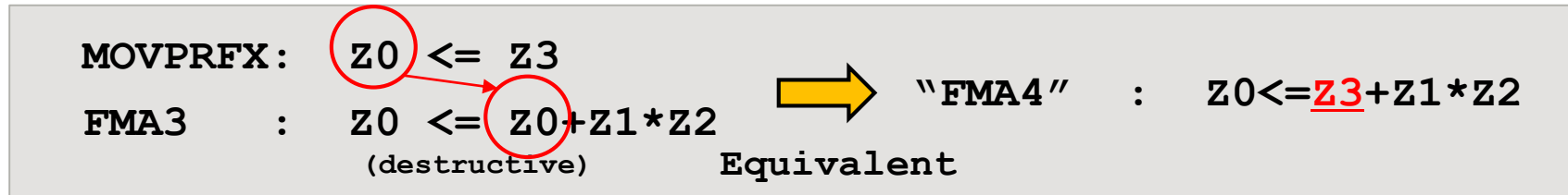
- Inherits superscalar, out-of-order, branch prediction, etc.
- Enhances SIMD and predicate operations
 - 2x 512-bit wide SIMD FMA + Predicate Operation + 4x ALU (shared w/ 2x AGEN)
 - 2x 512-bit wide SIMD load or 512-bit wide SIMD store



Four-operand FMA with Prefix Instruction

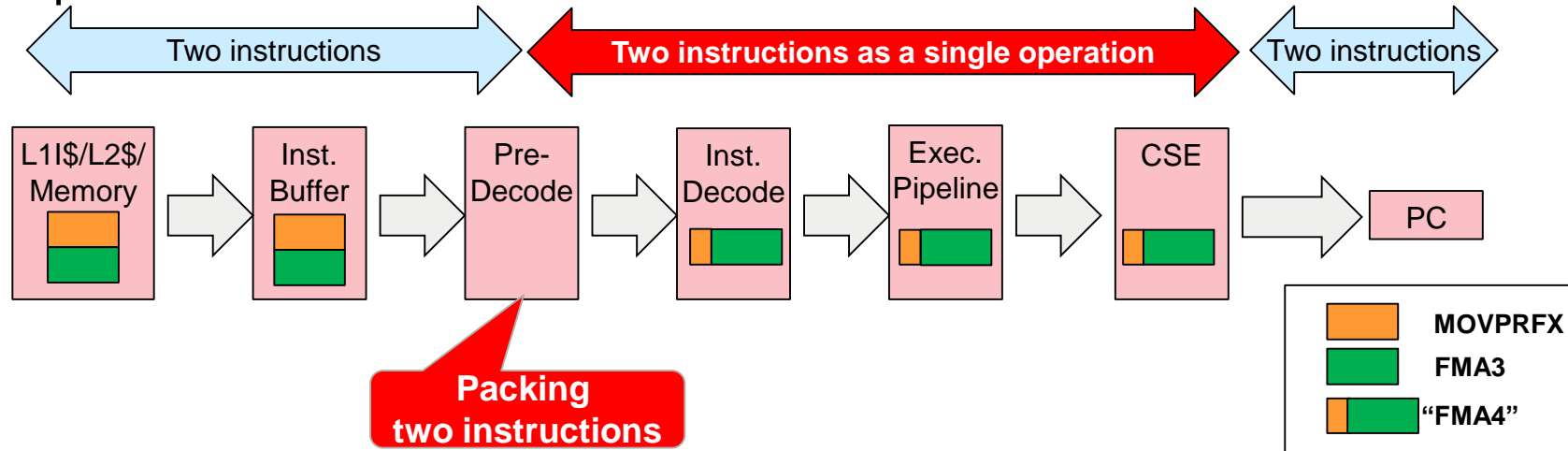
■ MOVPRFX as a prefix instruction

- For SVE, four-operand “FMA4” requires a prefix instruction (MOVPRFX) followed by destructive 3-operand FMA3



■ A64FX implementation for MOVPRFX

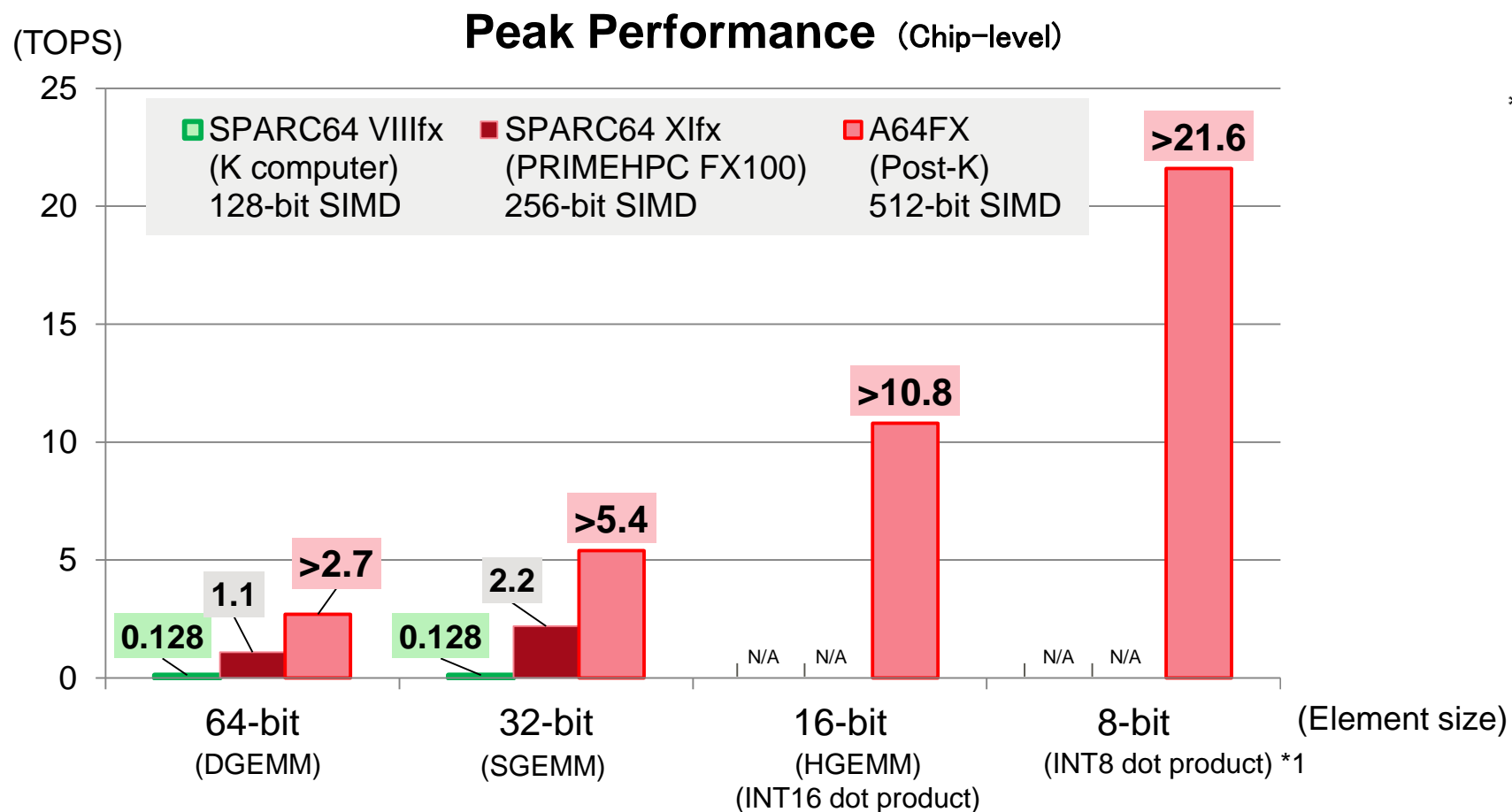
- A64FX hides the overhead of its main pipeline by packing MOVPRFX and the following instruction into a single operation



Execution Unit

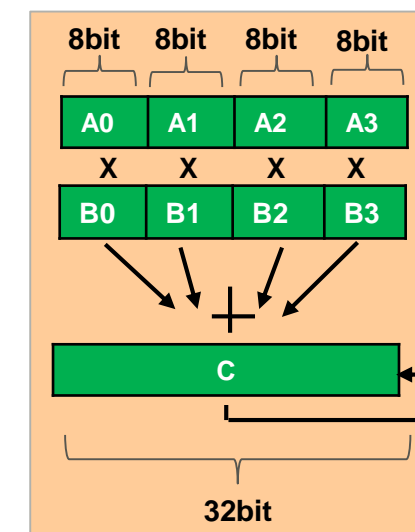
Extremely high throughput

- 512-bit wide SIMD x 2 Pipelines x 48 Cores
- >90% execution efficiency in (D|S|H)GEMM and INT16/8 dot product



*1 INT8 dot product

$$C = \sum (A_i \times B_i) + C$$

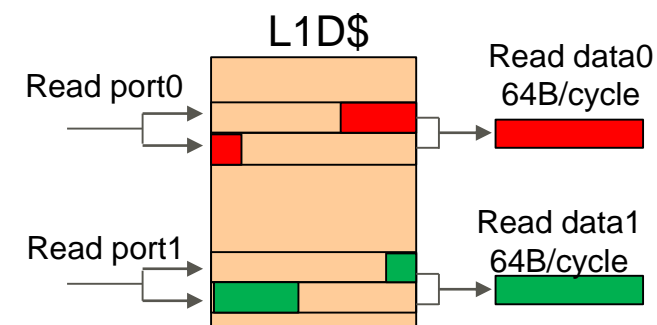


Level 1 Cache

■ L1 cache throughput maximizes core performance

■ Sustained throughput for 512-bit wide SIMD load

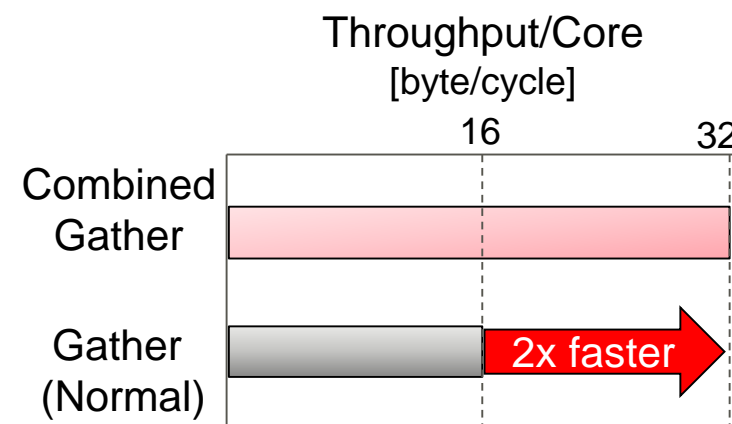
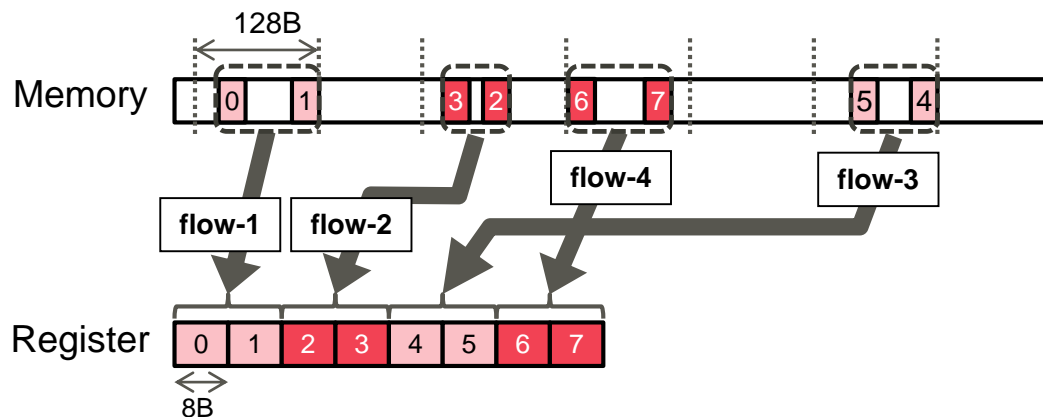
- An unaligned SIMD load crossing cache line keeps the same throughput



■ “Combined Gather” mechanism increasing gather throughput

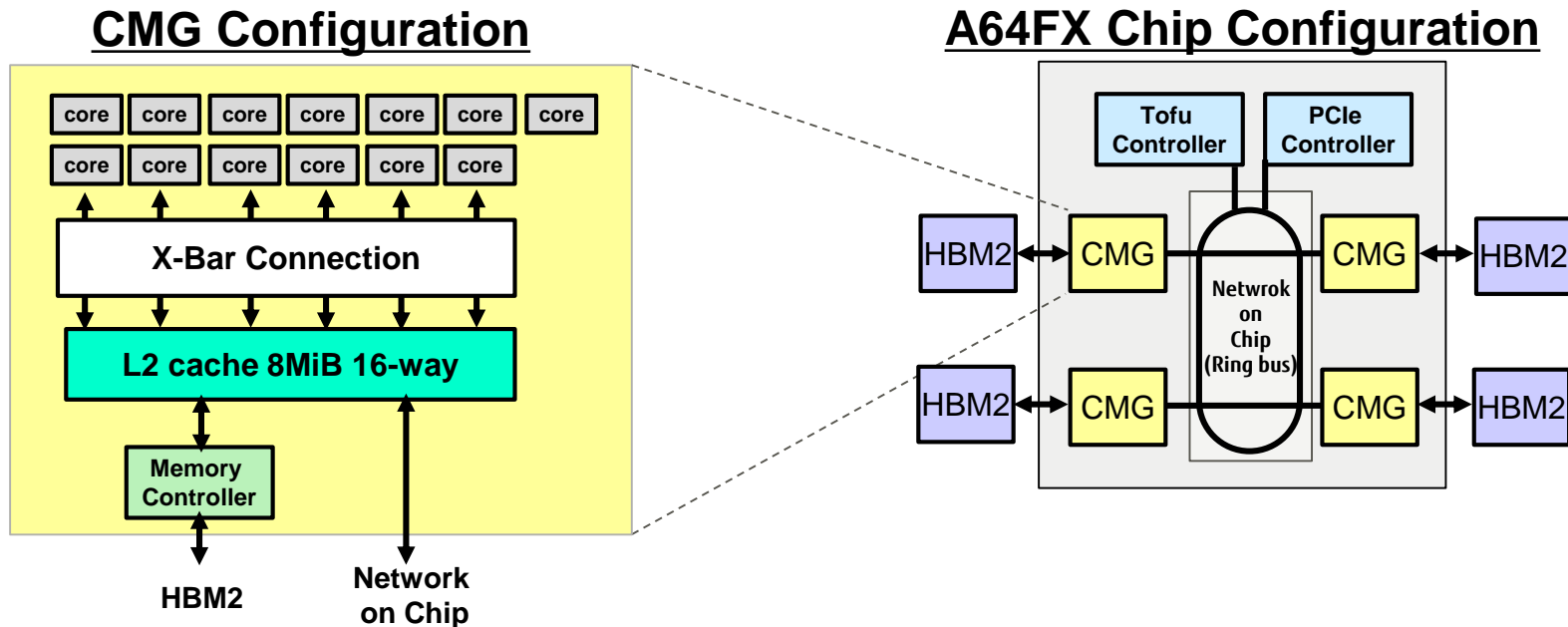
- Gather processing is important for real HPC applications
- A64FX introduces “Combined Gather” mechanism enabling to return up to two consecutive elements in a “128-byte aligned block” simultaneously

< Combined Gather >



Many-Core Architecture

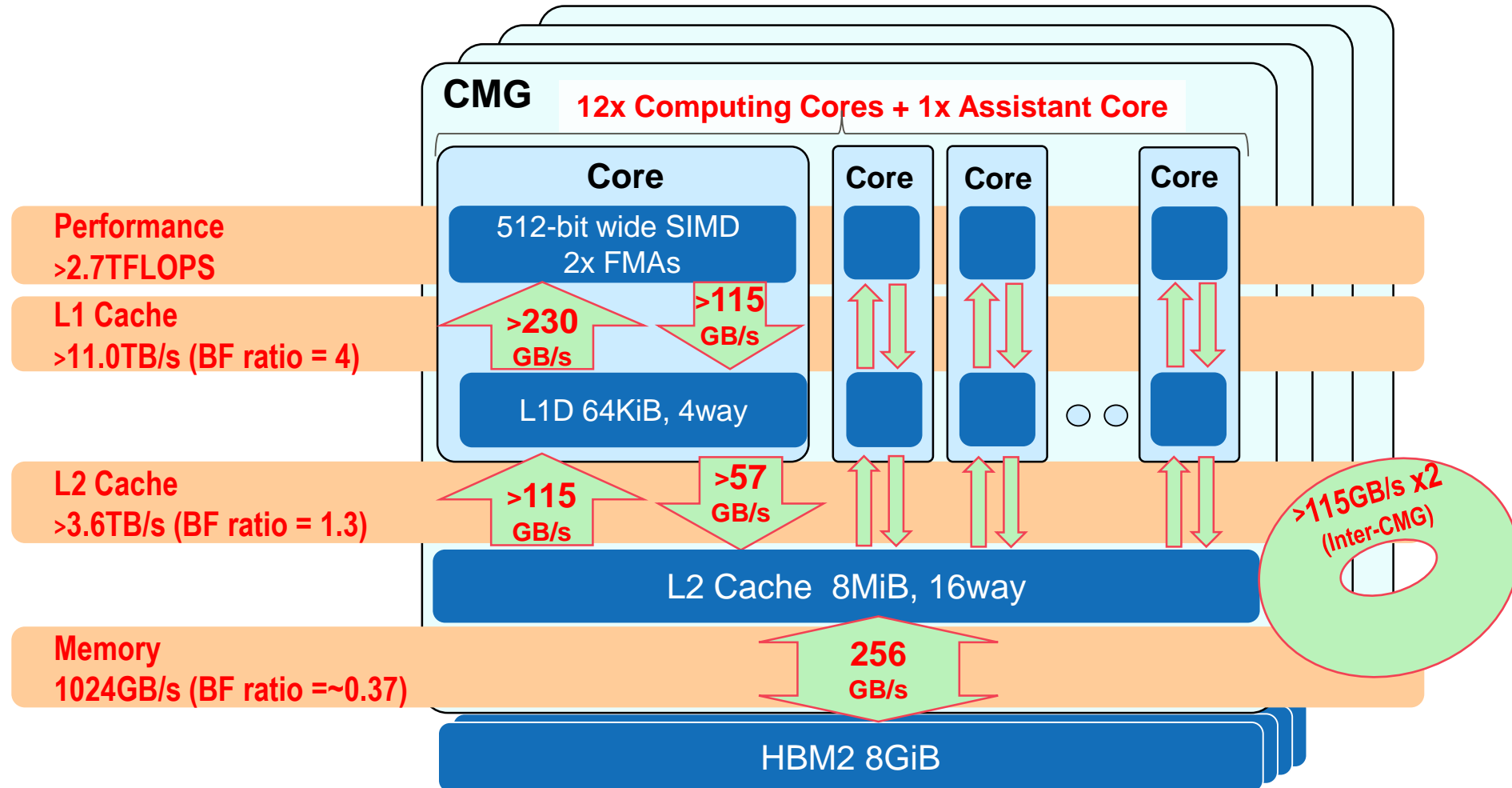
- A64FX consists of four CMGs (Core Memory Group)
 - A CMG consists of 13 cores, an L2 cache and a memory controller
 - One out of 13 cores is an assistant core which handles daemon, I/O, etc.
 - Four CMGs keep cache coherency by ccNUMA with on-chip directory
 - X-bar connection in a CMG maximizes high efficiency for throughput of the L2 cache
 - Process binding in a CMG allows linear scalability up to 48 cores
- On-chip-network with a wide ring bus secures I/O performance



High Bandwidth

■ Extremely high bandwidth in caches and memory

- A64FX has out-of-order mechanisms in cores, caches and memory controllers. It maximizes the capability of each layer's bandwidth

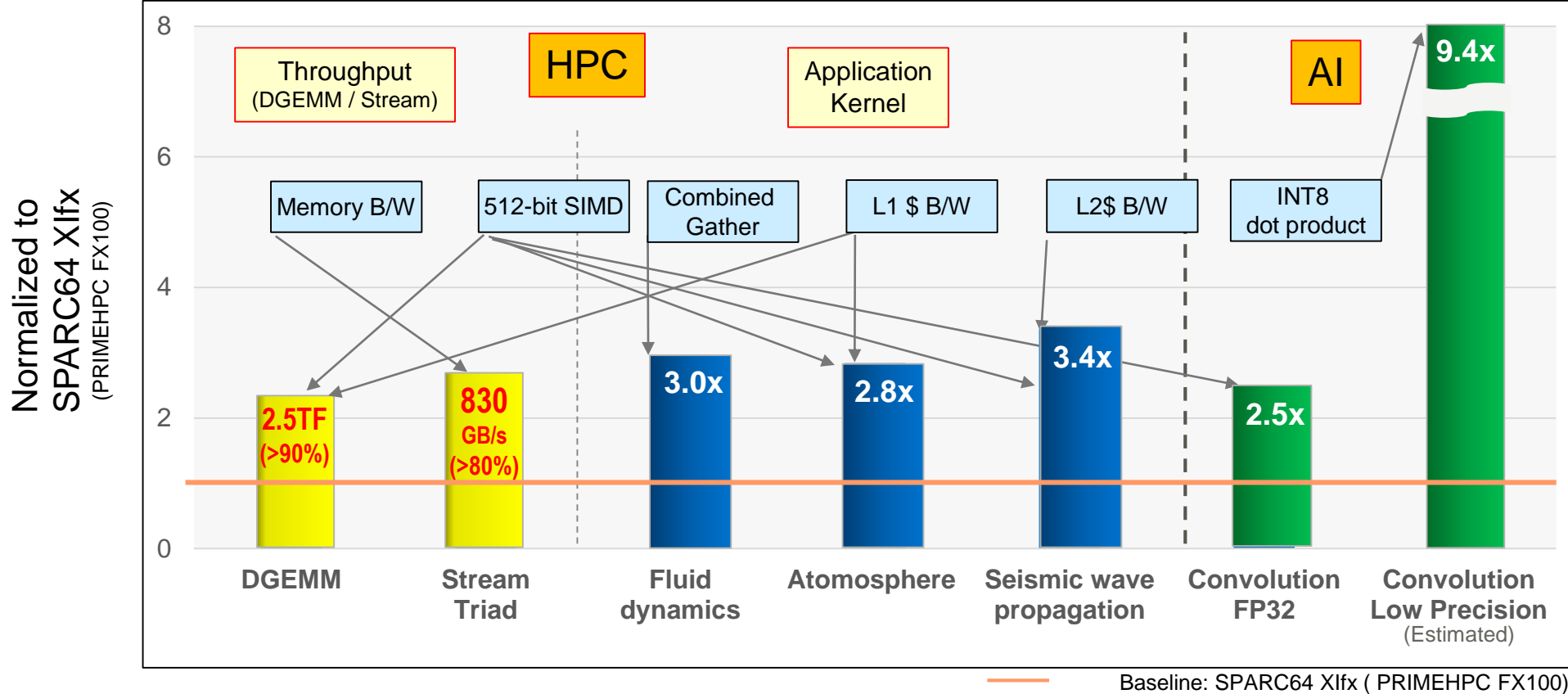


Performance

■ A64FX boosts performance up by microarchitectural enhancements, 512-bit wide SIMD, HBM2 and process technology

- > 2.5x faster in HPC/AI benchmarks than SPARC64 Xlfx (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)



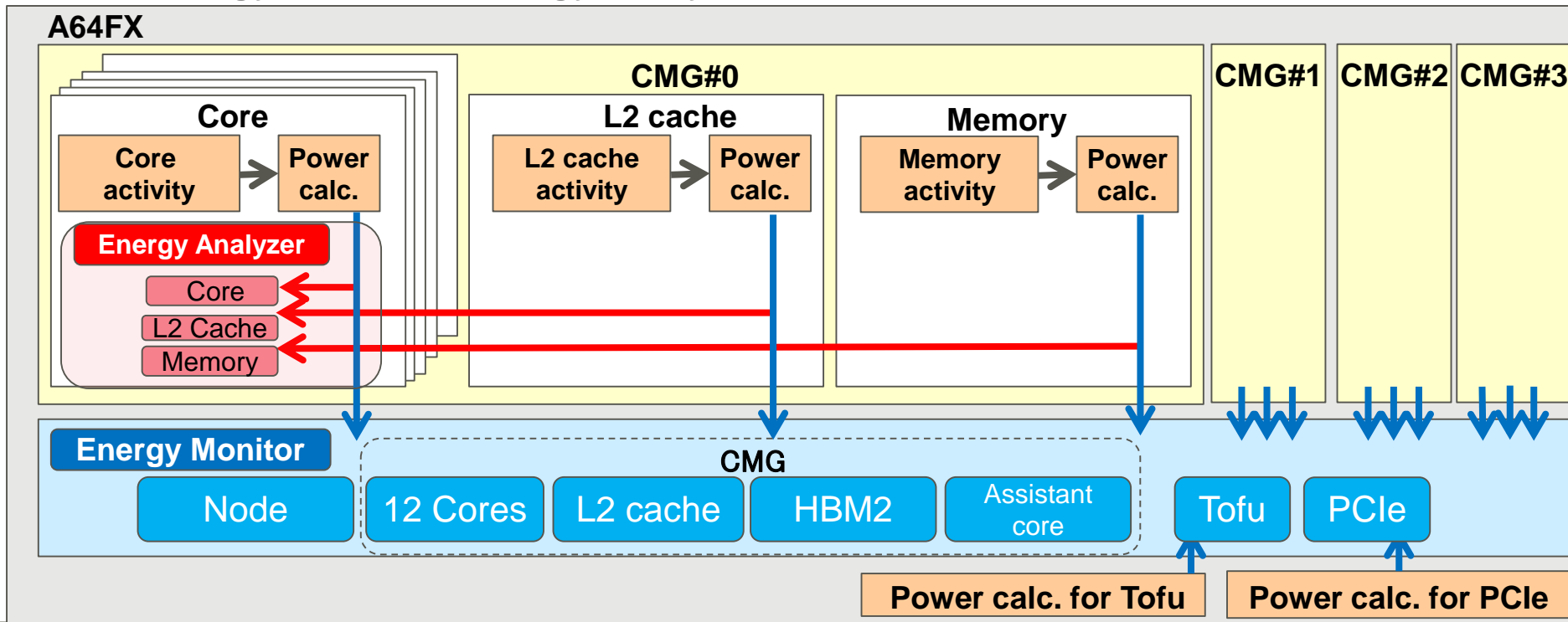
Power Management

■ “Energy monitor” / “Energy analyzer” for activity-based power estimation

- ✓ Energy monitor (per chip) : Node power via Power API* (~msec) *Sandia National Laboratory
 - Average power estimation of a node, CMG (cores, an L2 cache, a memory) etc.
- ✓ Energy analyzer (per core) : Power profiler via PAPI** (~nsec) ** Performance Application Programming Interface
 - Fine grained power analysis of a core, an L2 cache and a memory

→ Enabling chip-level power monitoring and detailed power analysis of applications

<A64FX Energy monitor/ Energy analyzer>



Power Management (Cont.)

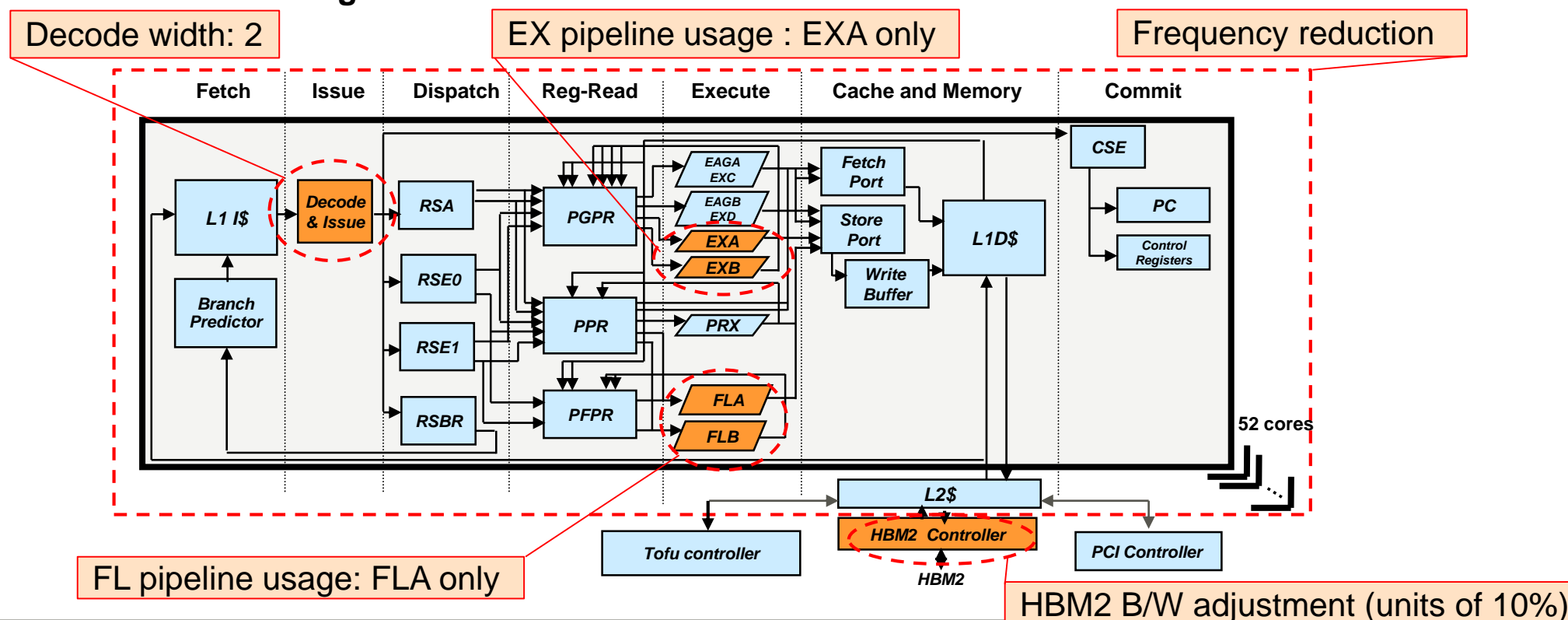
■ “Power knob” for power optimization

■ A64FX provides power management function called “Power Knob”

- Applications can change hardware configurations for power optimization

→ Power knobs and Energy monitor/analyzer will help users to optimize power consumption of their applications

<A64FX Power Knob Diagram>



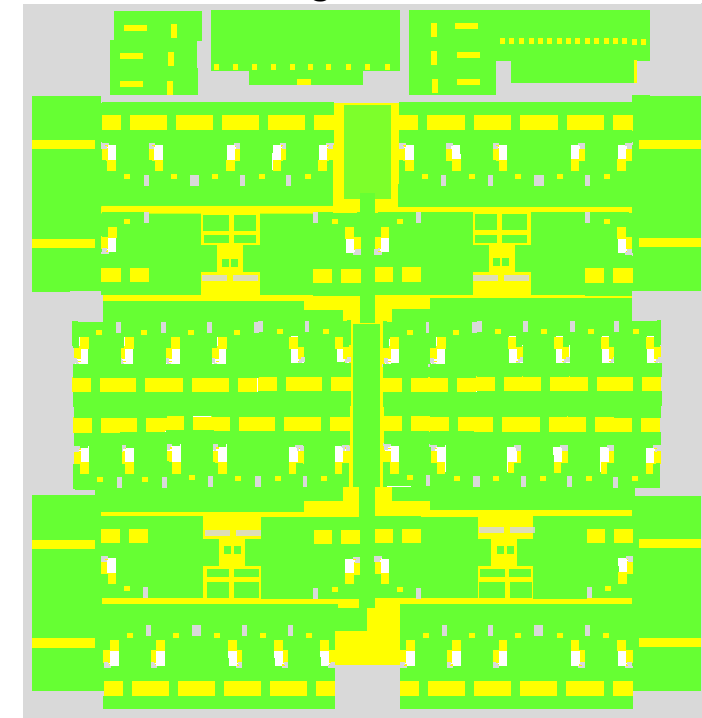
Fujitsu Mission Critical Technologies

- Large systems require extensive RAS capability of CPU and interconnect
- A64FX has a mainframe class RAS for integrity and stability.
It contributes to very low CPU failure rate and high system stability
 - ✓ ECC or duplication for all caches
 - ✓ Parity check for execution units
 - ✓ Hardware instruction retry
 - ✓ Hardware lane recovery for Tofu links
 - ✓ ~128,400 error checkers in total

<A64FX RAS Mechanism>

Units	Error Detection and Correction
Cache (Tag)	ECC, Duplicate & Parity
Cache (Data)	ECC, Parity
Register	ECC (INT), Parity(Others)
Execution Unit	Parity, Residue
Core	Hardware Instruction Retry
Tofu	Hardware Lane Recovery

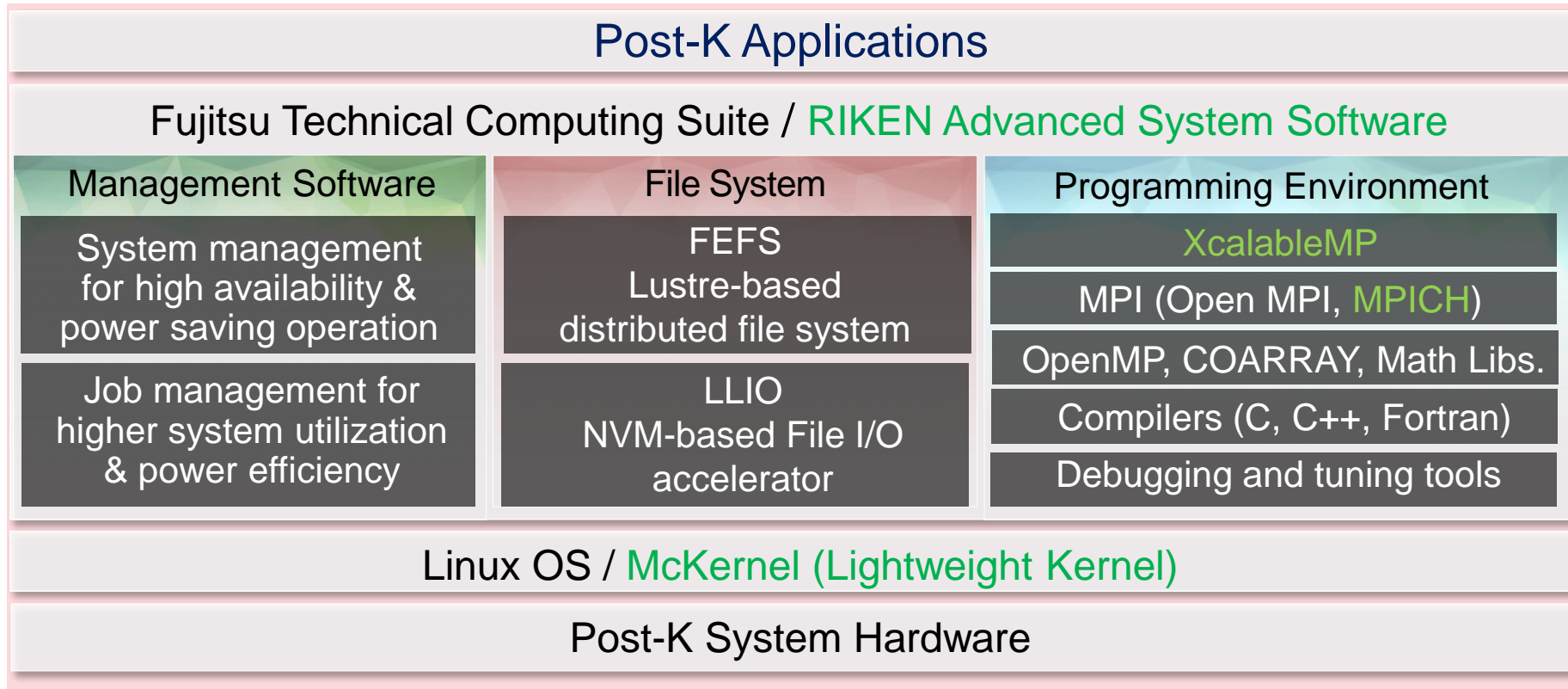
<A64FX RAS Diagram>



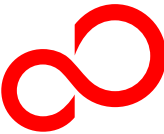
Green: 1 bit error Correctable
Yellow: 1 bit error Detectable
Gray : 1 bit error harmless

Software Development

- RIKEN and Fujitsu are developing software stacks for the post-K computer
 - Fujitsu compilers are optimized for the microarchitecture, maximizing SVE and HBM2 performance
- We collaboratively work with RIKEN / Linaro / OSS communities / ISVs and contribute to Arm HPC ecosystem



- A64FX is the first processor of the Armv8-A SVE architecture.
It is used for the post-K computer
- Fujitsu's proven microarchitecture achieves high performance in HPC and AI areas
- Fujitsu collaboratively works with partners and continuously contributes to Arm ecosystem
- We will continue to develop Arm processors



FUJITSU

shaping tomorrow with you

■ A64FX

- RSA: Reservation station for address generation
- RSE: Reservation station for execution
- RSBR: Reservation station for branch
- PGPR: Physical general-purpose register
- PFPR: Physical floating-point register
- PPR: Physical predicate register
- CSE: Commit stack entry
- EAG: Effective address generator
- EX : Integer execution unit
- FL : Floating-point execution unit
- PRX : Predicate execution unit
- Tofu: Torus-Fusion