FUJITSU

shaping tomorrow with you

FUJITSU Supercomputer PRIMEHPC FX100

Hardware and Software Overview

FUJITSU Supercomputers

- Fujitsu has been developing supercomputers nearly 40 years, and will continue its development to deliver the best application performance

Exascale

K computer

Peak performance: 11.28 petaflops

PRIMEHPC FX10

Peak performance: up to 23.2 petaflops

PRIMEHPC FX100

Peak performance: over 100 petaflops

PRIMEHPC FX100 Design Concept

FUJITSU

Designed to be a massively parallel supercomputer system

- High performance for a wide range of real applications

Inherited the K computer features

- General purpose CPU architecture for application productivity
- 6D mesh/torus topology, hardware barrier synchronization, sector cache, etc.

Introducing new technologies for Exascale computing

- HPC-ACE2 : Wide SIMD enhancements
- Assistant cores : Dedicated cores for non-calculation operation
- HMC : Leading-edge memory technology

SPARC64™ XIfx

FUJITSU

Over 1 TF high performance processor

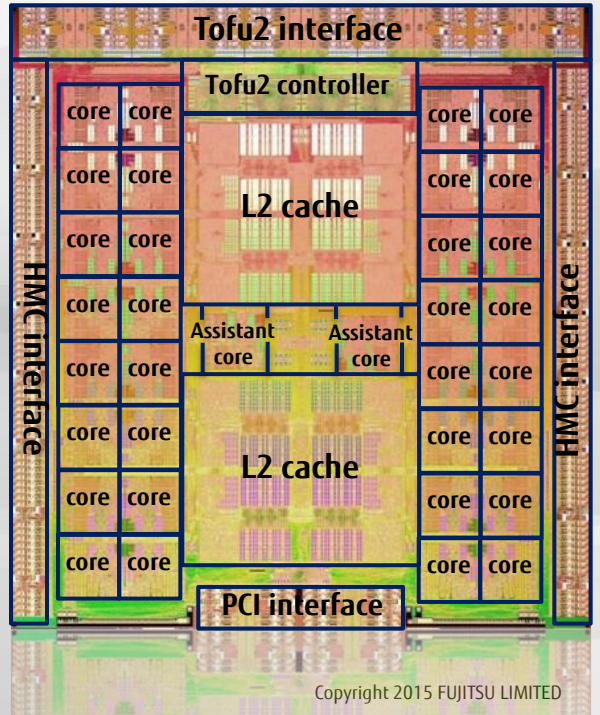
- 32 compute cores
- 2 assistant cores: Offloading non-calculation operations
  ⟶ Daemons, IOs, non-blocking MPI functions, etc.

HPC-ACE2: ISA enhancements

- Two 256-bit wide SIMD units per core
- Various SIMD instructions
  (stride load/store, indirect load/store, permutation, etc.)

HMC support

- 480GB/s/node of theoretical memory throughput



Tofu Interconnect 2

FUJITSU

Enhanced Tofu interconnect

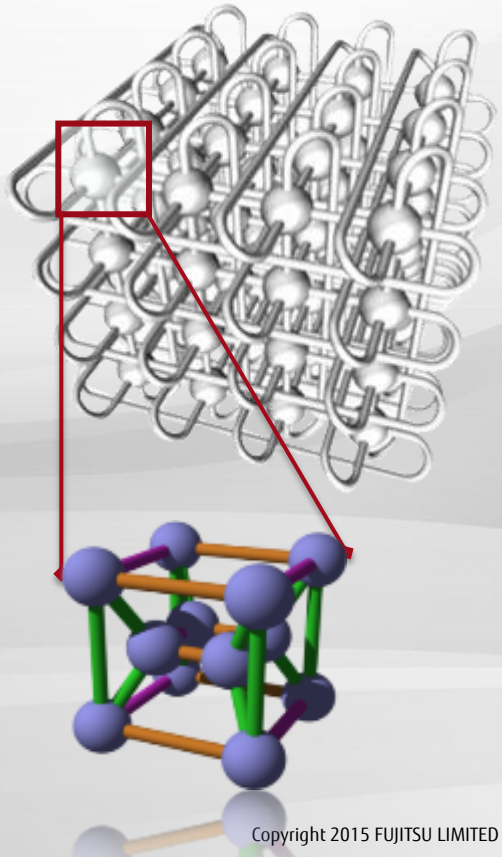
- Highly scalable, 6-dimensional mesh/torus topology
- Increased link bandwidth by 2.5 times to 12.5GB/s
- Added atomic memory operations

CPU-integrated interconnect controller

- Reduced communication latency
- Improved packaging density and energy efficiency

Optical cable connection between chassis

- Enable flexible installation



Technical Computing Suite

FUJITSU

Enhanced software stack developed by Fujitsu

Applications

Technical Computing Suite

Management software

- System management
- Job management

High Performance File System FEFS

- Lustre-based distributed file system (enhanced for FX100)

Programming Environment

- MPI, OpenMP, COARRAY
- Compilers(C,C++,Fortran) Mathematical libraries
- Debugging and tuning tools

Linux-based OS enhanced for FX100

PRIMEHPC FX100

The Evolution of FUJITSU Software

Future

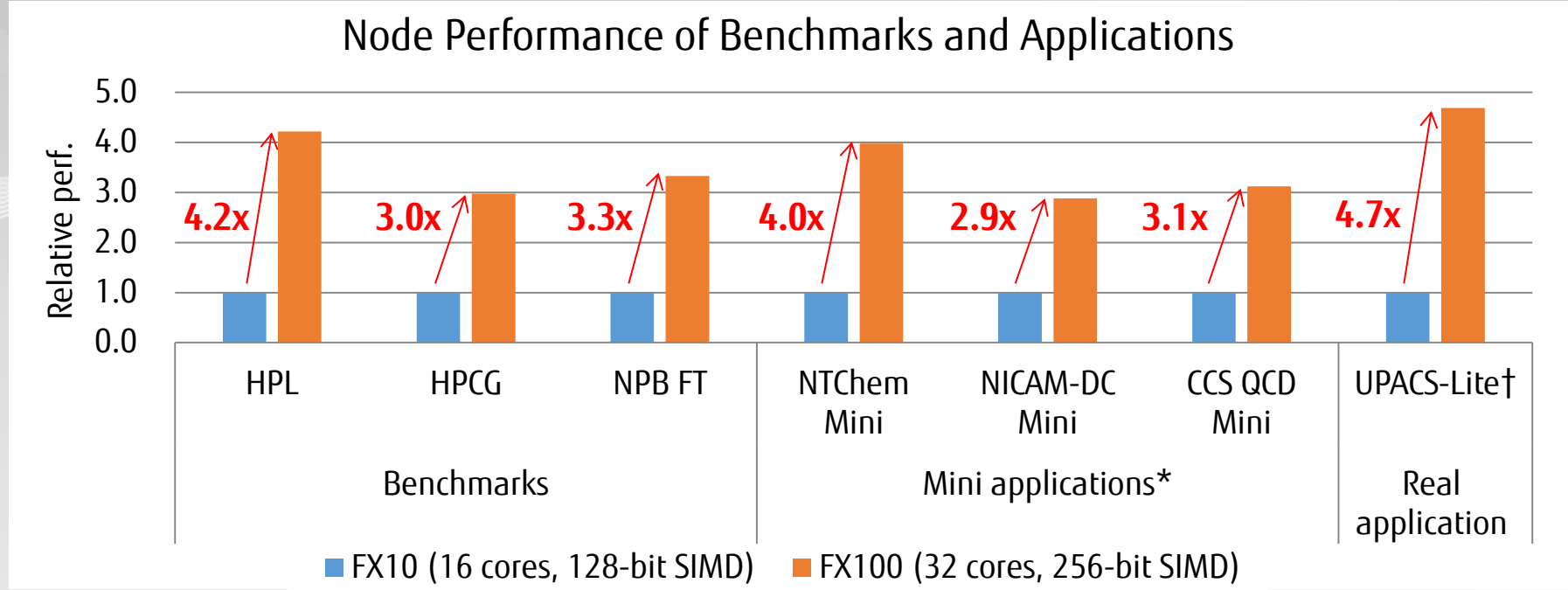
2011 (K computer)	2015 (PRIMEHPC FX100)	
Fortran, C, C++ with sophisticated optimization	COARRAY in Fortran 2008, C++11 with advanced vectorization for wide SIMDs	**Optimization strategy based on application characteristics**
Scalable MPI over 100k procs	Asynchronous MPI comm. for low-latency and scalability	**Scalable MPI over 1M procs**
Large-scale job scheduler (over 80k nodes)	Flexible job allocation for high throughput computing	**Power saving functions**

FX100 Performance and the Effect of the New Technologies

The Performance Improvement of FX100

FUJITSU

- FX100 greatly improves the performance of various types of programs

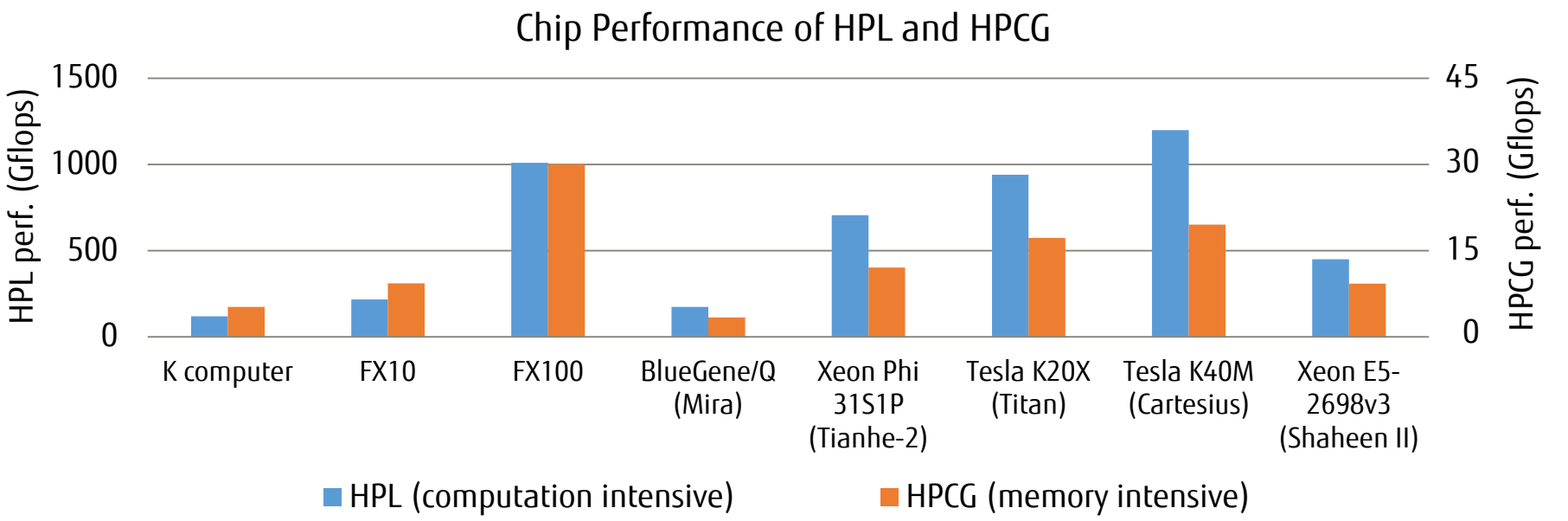


* Fiber miniapp suite developed by RIKEN † Calculation of compressible fluid dynamics. Developed by JAXA

Balanced Enhancement of FLOPS and Memory

FUJITSU

- Over 1 TFLOPS and 480 GB/s memory bandwidth per chip
- PRIMEHPC series show high performance for both HPL and HPCG

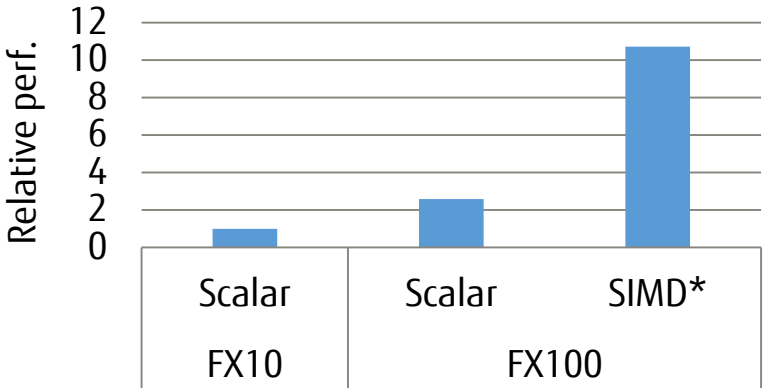


Loop Vectorization by New SIMD Instructions

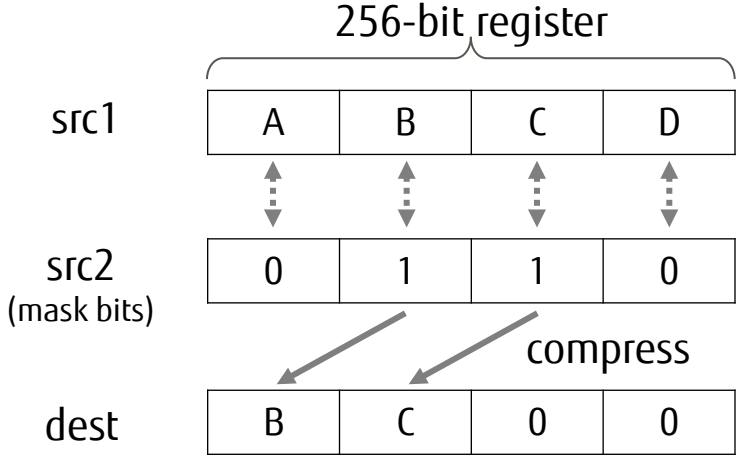
FUJITSU

- Vectorizing complex loops is a key to get higher performance
- FX100 introduces new SIMD instructions, such as non-continuous memory accesses, integer calculations, permutation, compression, etc.

The Effect of SIMD Compression (NPB EP)



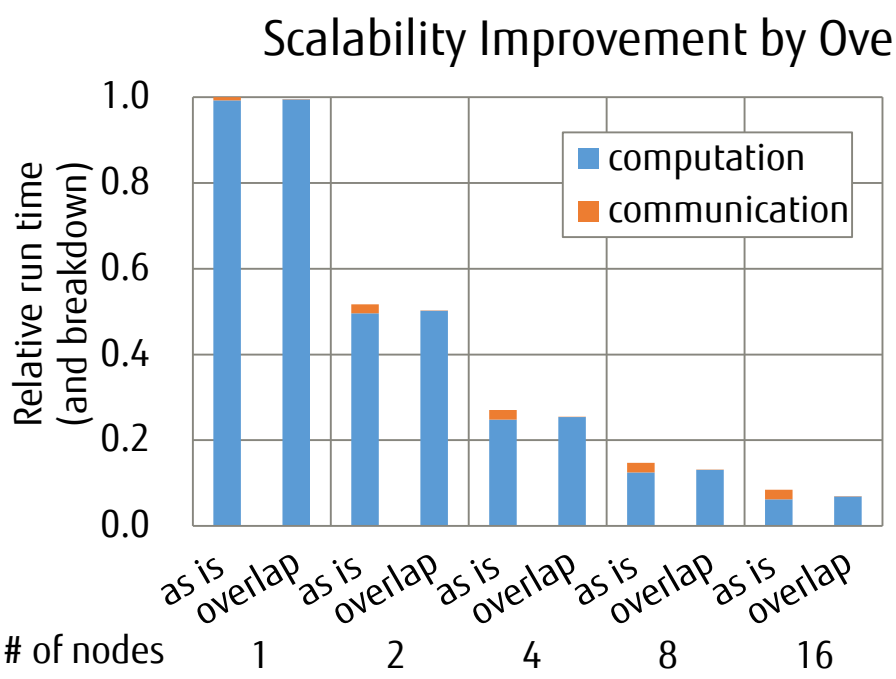
* With a loop fission to promote vectorization

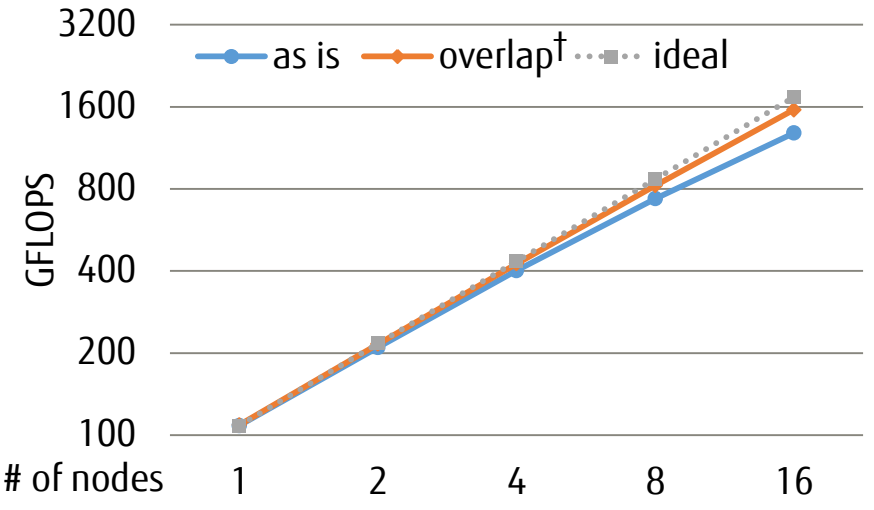


Better Scalability by Comp. & Comm. Overlap

FUJITSU

- By offloading MPI processing to assistant cores, non-blocking communications are performed simultaneously with computation



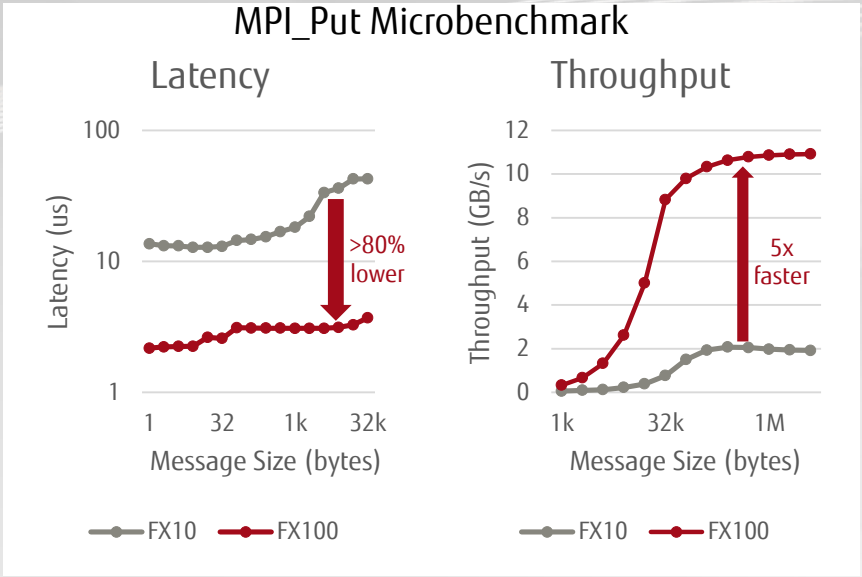


* A stencil code solving the Poisson's equation solution
† Halo exchanges are overlapped

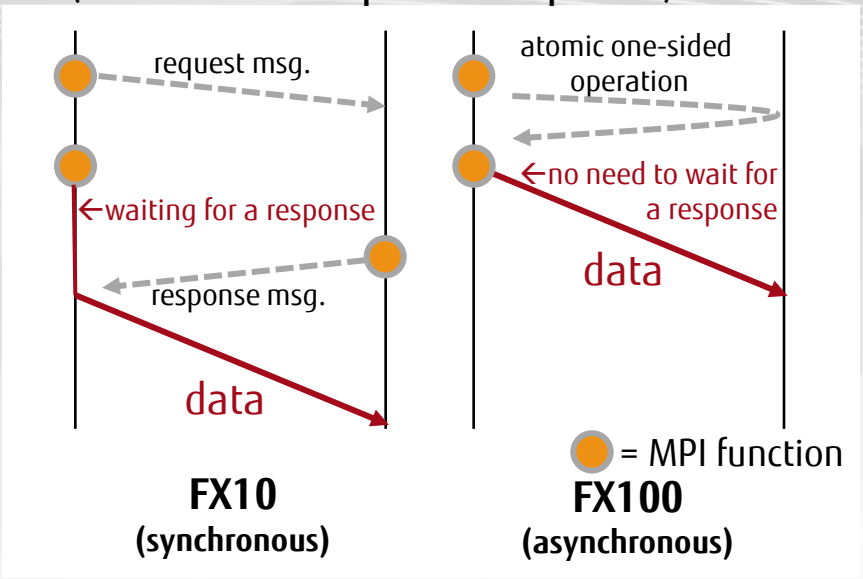
MPI Remote Memory Access Performance

FUJITSU

- Fujitsu MPI now supports MPI-3.0, including RMA functions!



- Almost all FX100's RMA functions start transfer asynchronously (no remote response required)

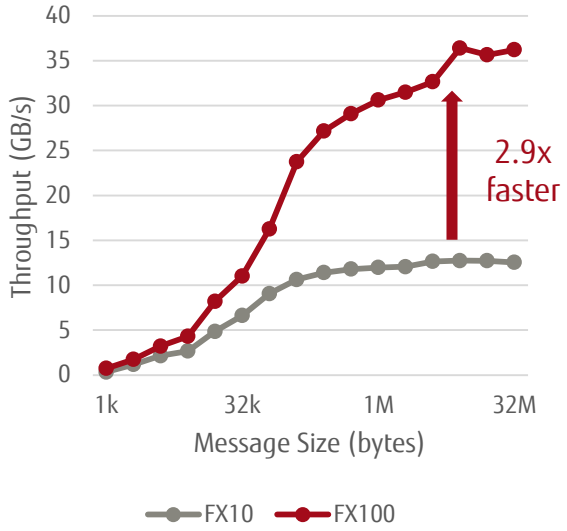


MPI Collective Communication Performance

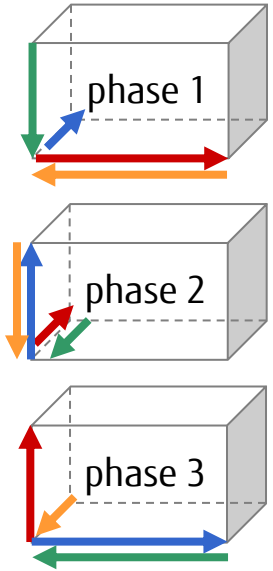
FUJITSU

- Fujitsu MPI provides high-bandwidth collective functions optimized for Tofu

MPI_Allgather Throughput (36 nodes)



Algorithm



Why are Fujitsu MPI's collectives so fast?

- ✓ High bandwidth of the Tofu interconnect 2 (peak 12.5GB/s per network engine)
- ✓ Driving 4 network engines in parallel
- ✓ Low latency communication protocol thanks to RDMA

Summary

FUJITSU

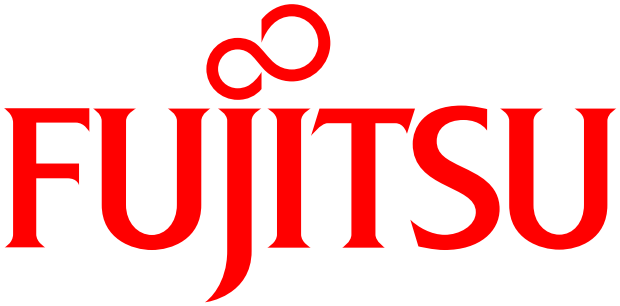
- FX100 achieves high performance of various applications by the new technologies and inherited features
- This evolution is continuing to the next generations

Exascale

K computer

PRIMEHPC FX10

PRIMEHPC FX100



shaping tomorrow with you