FUJITSU

shaping tomorrow with you

Fujitsu's Architectures and Collaborations for Weather Prediction and Climate Research

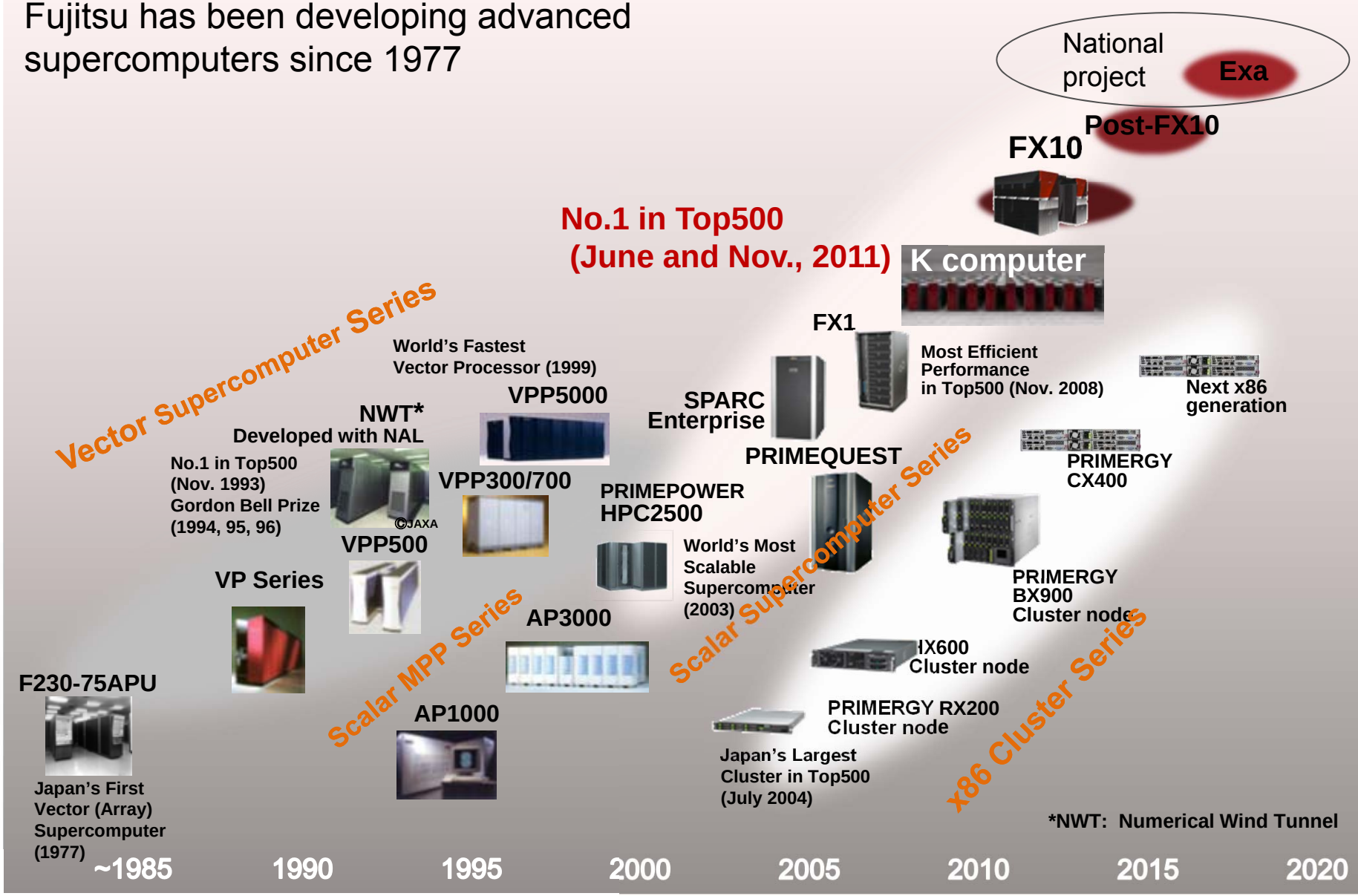
Ross Nobes

Fujitsu Laboratories of Europe

Fujitsu HPC - Past, Present and Future

FUJITSU

Fujitsu has been developing advanced supercomputers since 1977



Fujitsu's Approach to the HPC Market

FUJITSU

SPARC64-Based Supercomputers

X86 PRIMERGY HPC solutions

Powerful Workstations

The K computer

FUJITSU

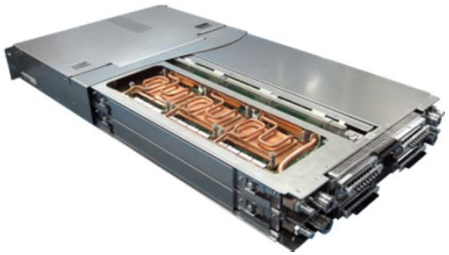
June 2011	No.1 in TOP500 List (ISC11)
November 2011	Consecutive No. 1 in TOP500 List (SC11)
November 2011	ACM Gordon Bell Prize Peak-Performance (SC11)
November 2012	No.1 in Three HPC Challenge Award Benchmarks (SC12)
November 2012	ACM Gordon Bell Prize (SC12)
November 2013	No.1 in Class 1 and 2 of the HPC Challenge Awards (SC13)
June 2014	No.1 in Graph 500 “Big Data” Supercomputer Ranking (ISC14)

Build on the success of the K computer

Evolution of PRIMEHPC

FUJITSU

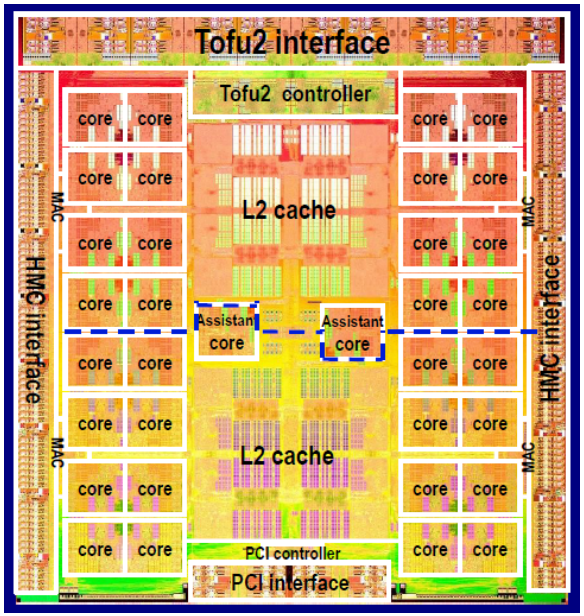
	K computer	PRIMEHPC FX10	Post-FX10
CPU	SPARC64 VIIIfx	SPARC64 IXfx	SPARC64 XIfx
Peak perf.	128 GFLOPS	236.5 GFLOPS	1 TFLOPS ~
# of cores	8	16	32 + 2
Memory	DDR3 SDRAM	←	HMC
Interconnect	Tofu Interconnect	←	Tofu Interconnect 2
System size	11 PFLOPS	Max. 23 PFLOPS	Max. 100 PFLOPS
Link BW	5 GB/s x bidirectional	←	12.5 GB/s x bidirectional

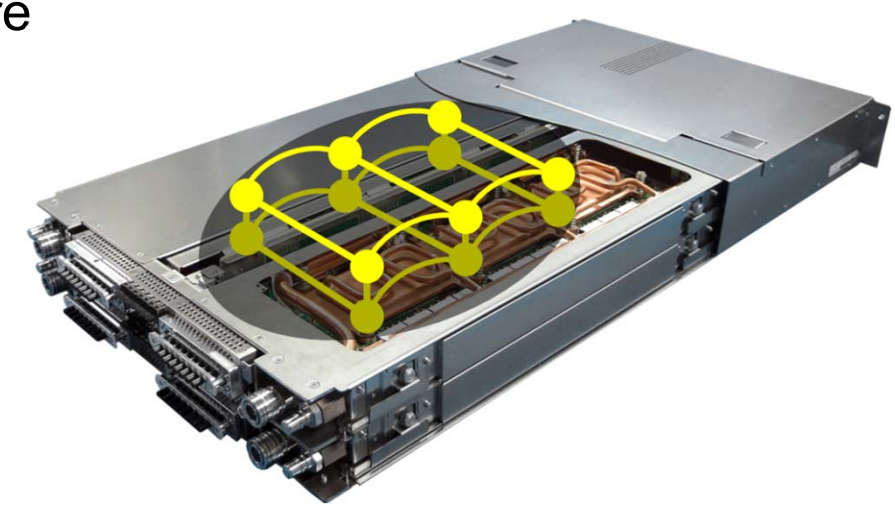


Continuity in Architecture for Compatibility

FUJITSU

- Upwards compatible CPU:
  - Binary-compatible with the K computer & PRIMEHPC FX10
  - Good byte/flop balance
- New features:
  - New instructions
  - Improved micro architecture
- For distributed parallel executions:
  - Compatible interconnect architecture
  - Improved interconnect bandwidth





32 + 2 Core SPARC64 XIfx

FUJITSU

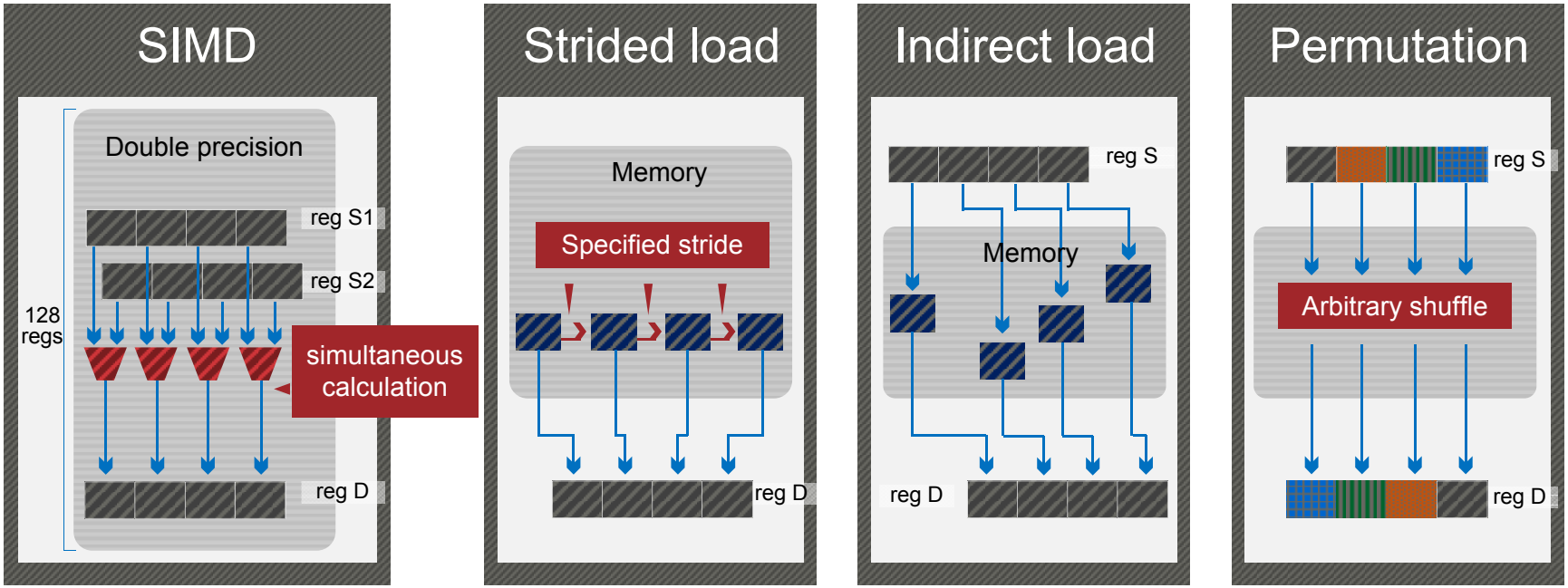
- Rich micro architecture improves single thread performance
- 2 additional, Assistant-cores for avoiding OS jitter and non-blocking MPI functions

		K	FX10	Post-FX10
Peak FP performance		128 GF	236.5 GF	1-TF class
Core config.	Execution unit	FMA × 2	FMA × 2	FMA × 2
	SIMD	128 bit	128 bit	256 bit wide
	Dual SP mode	NA	NA	2x DP performance
	Integer SIMD	NA	NA	Support
	Single thread performance enhancement	-	-	Improved OOO execution, better branch prediction, larger cache

Flexible SIMD Operations

FUJITSU

- New 256-bit wide SIMD functions enable versatile operations
  - Four double-precision calculations
  - Strided load/store, Indirect (list) load/store, Permutation, Concatenation



Hybrid Memory Cube (HMC)

FUJITSU

- The increased arithmetic performance of the processor needs higher memory bandwidth
  - The required memory bandwidth can almost be met by HMC (480 GB/s)
  - Interconnect is boosted to 12.5 GB/s x 2 (bi-directional) with optical link

Peak performance/node	K	FX10	post-FX10
DP performance (Gflops)	128	236.5	Over 1TF
Memory Bandwidth (GB/s)	64	85	480
Interconnect Link Bandwidth (GB/s)	5	5	12.5

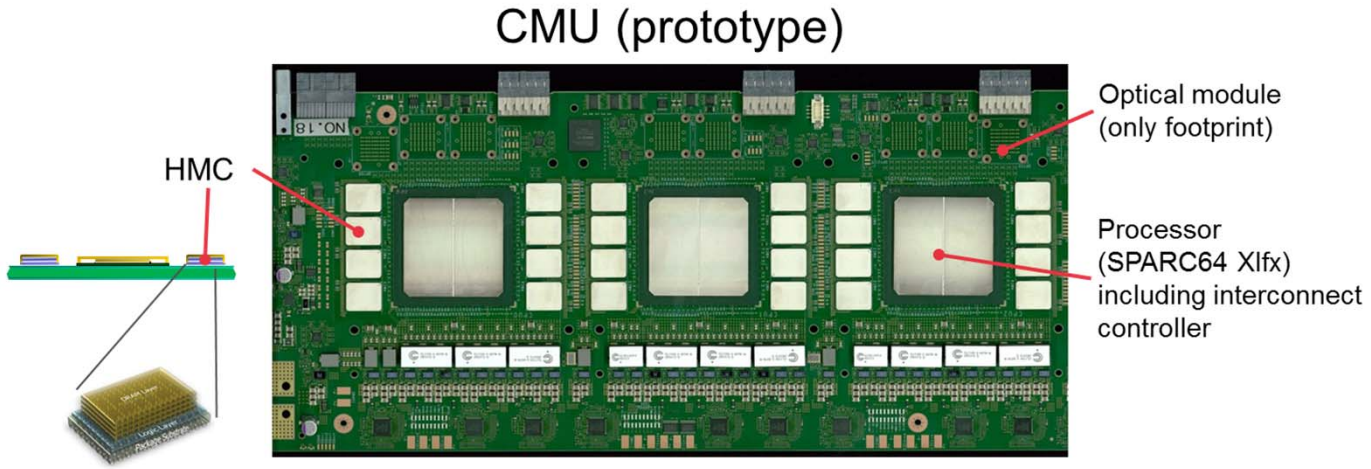
HMC (2)

FUJITSU

Amount per processor	Capacity	Memory BW
HMC x8	32 GB	**480 GB/s**
DDR4-DIMM x8	**32~128 GB**	154 GB/s
GDDR5 x16	8 GB	320 GB/s

HSSD was adopted for main memory since its bandwidth is three times more than DDR4

- HMC can deliver the required bandwidth for high performance multi-core processors

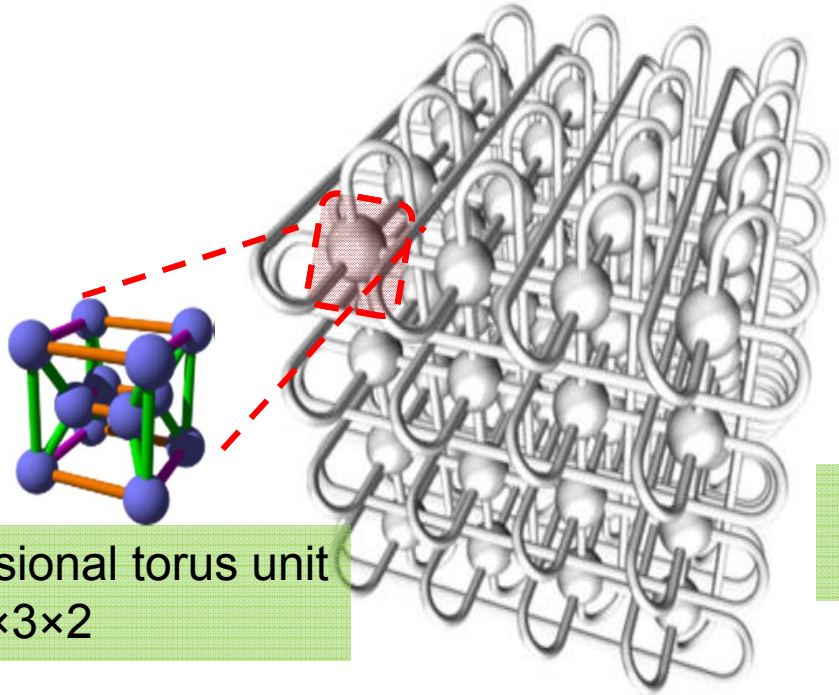


Tofu2 Interconnect

FUJITSU

- Successor to Tofu Interconnect
  - Highly scalable, 6-dimensional mesh/torus topology
  - Logical 3D, 2D or 1D torus network from the user's point of view
  - Increased link bandwidth by 2.5 times to 100 Gbps

Scalable three-dimensional torus



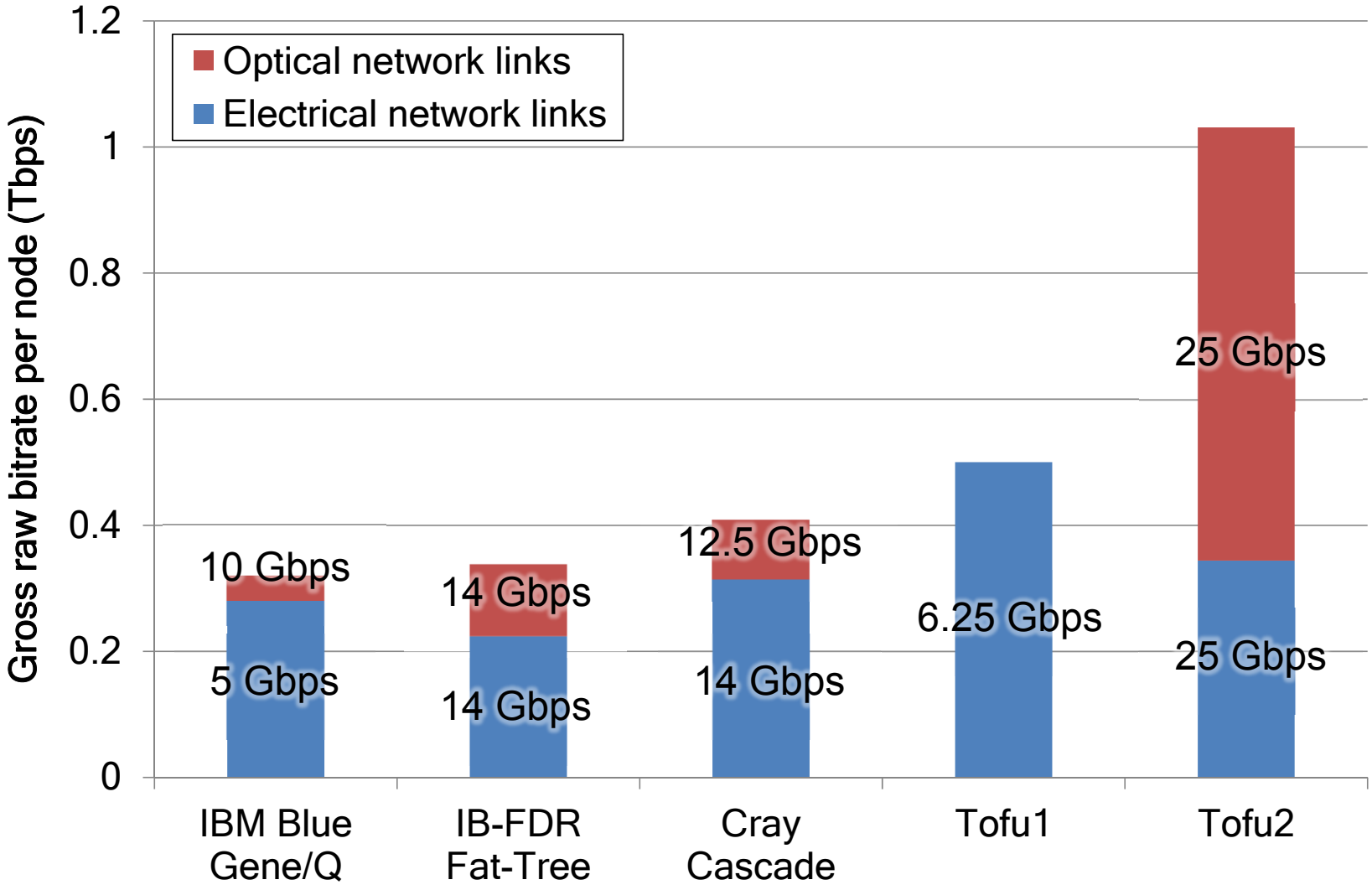
Three-dimensional torus unit
2×3×2

Well-balanced shape
available

A Next Generation Interconnect

FUJITSU

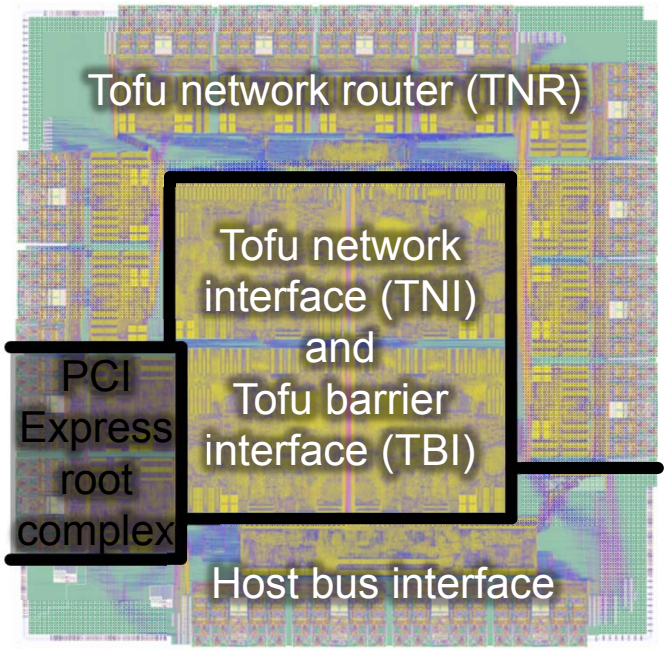
- Optical-dominant: 2/3 of network links are optical



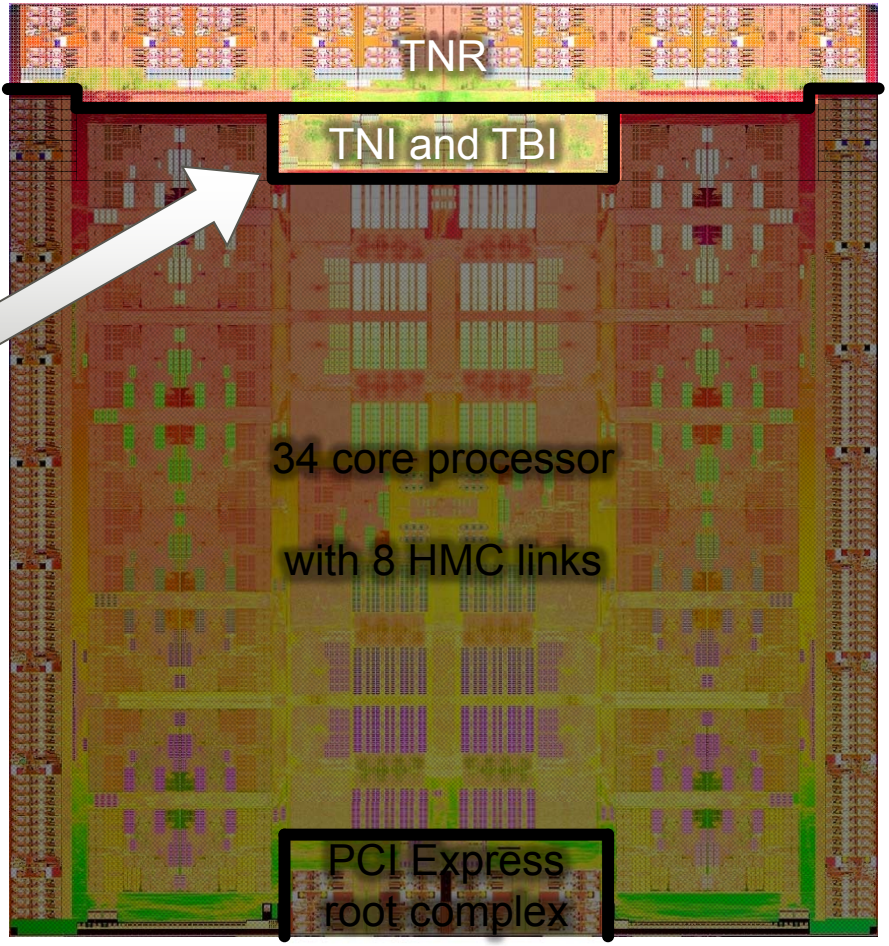
Reduction in the Chip Area Size

FUJITSU

- Process technology shrinks from 65 to 20 nm
- System-on-chip integration eliminates the host bus interface
- Chip area shrinks to 1/3 size



Tofu1 < 300mm^2
InterConnect Controller (65nm)

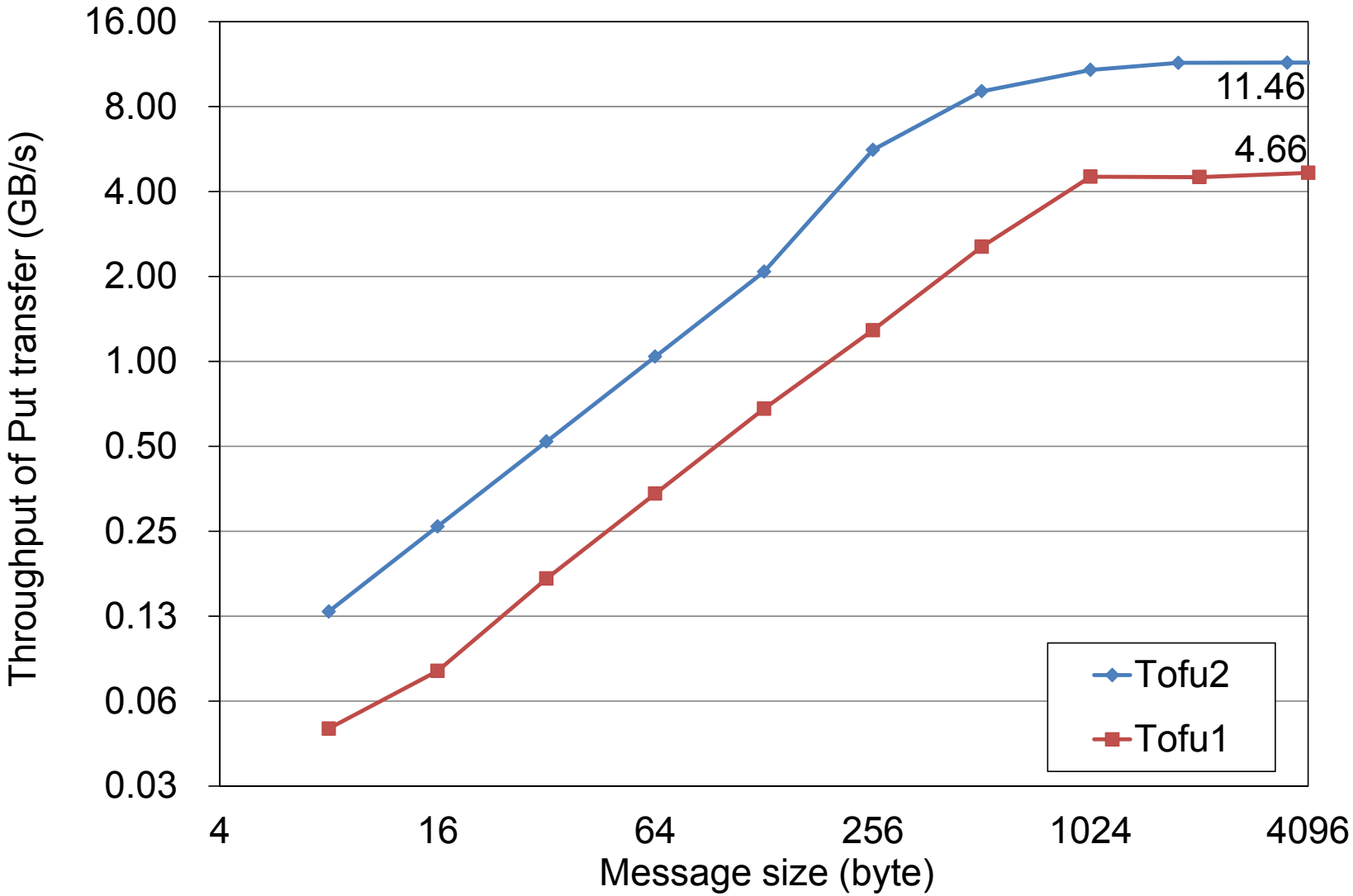


Tofu2 < 100mm^2 – SPARC64^{TM} XIfx (20nm)

Throughput of Single Put Transfer

FUJITSU

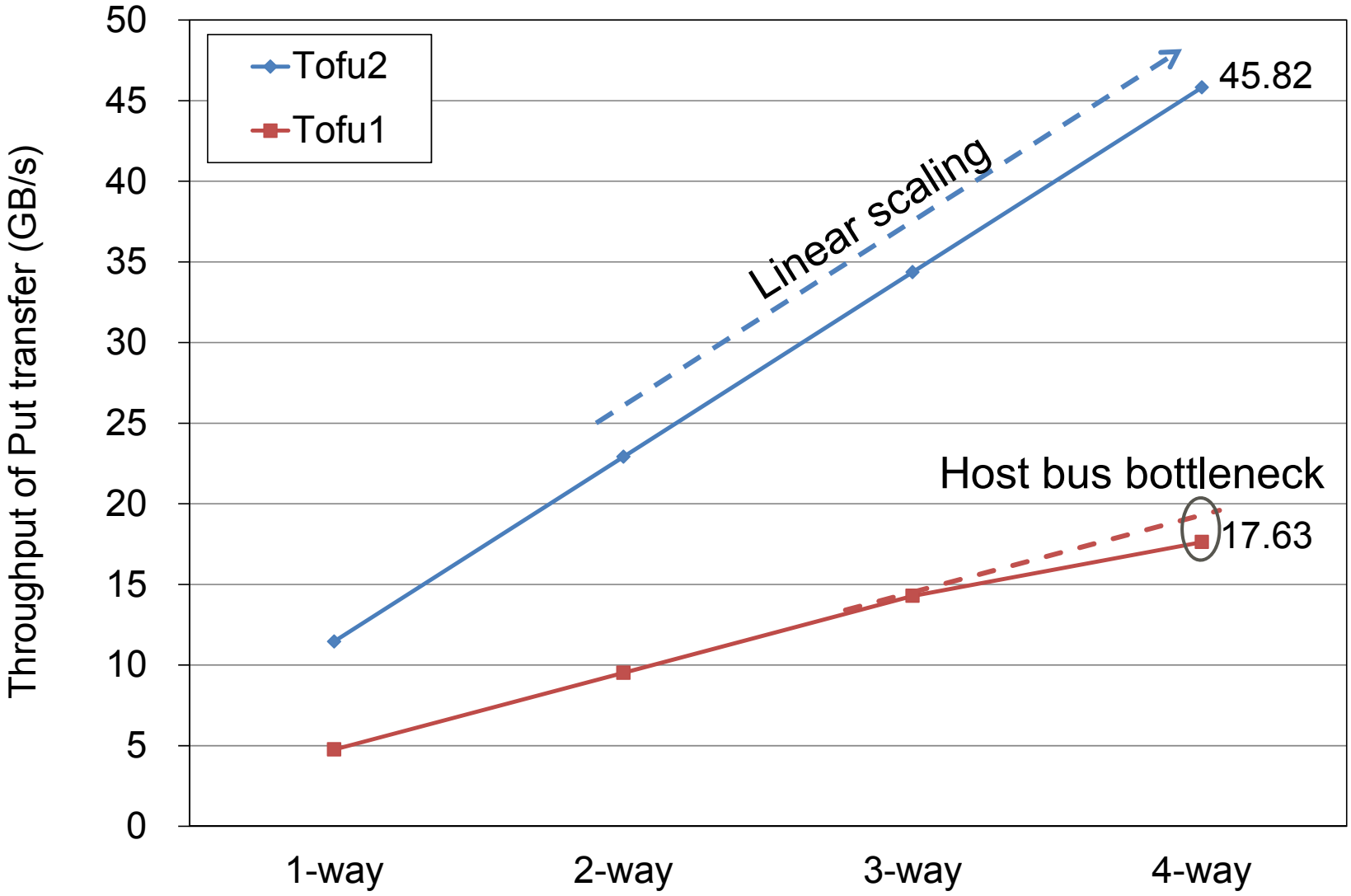
- Achieved 11.46 GB/s of throughput which is 92% efficiency



Throughput of Concurrent Put Transfers

FUJITSU

- Linear increase in throughput without the host bus bottleneck



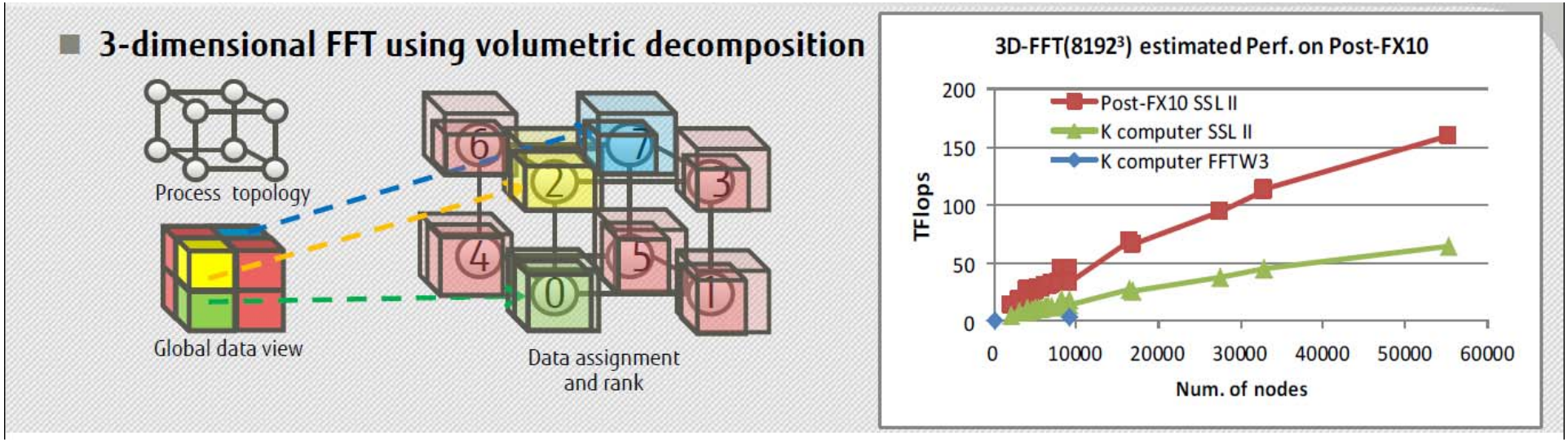
Communication Latencies

FUJITSU

Pattern	Method	Latency
One-way	Put 8-byte to memory	0.87 µsec
	Put 8-byte to cache	0.71 µsec
Round-trip	Put 8-byte ping-pong by CPU	1.42 µsec
	Put 8-byte ping-pong by session	1.41 µsec
	Atomic RMW 8-byte	1.53 µsec

Communication Intensive Applications

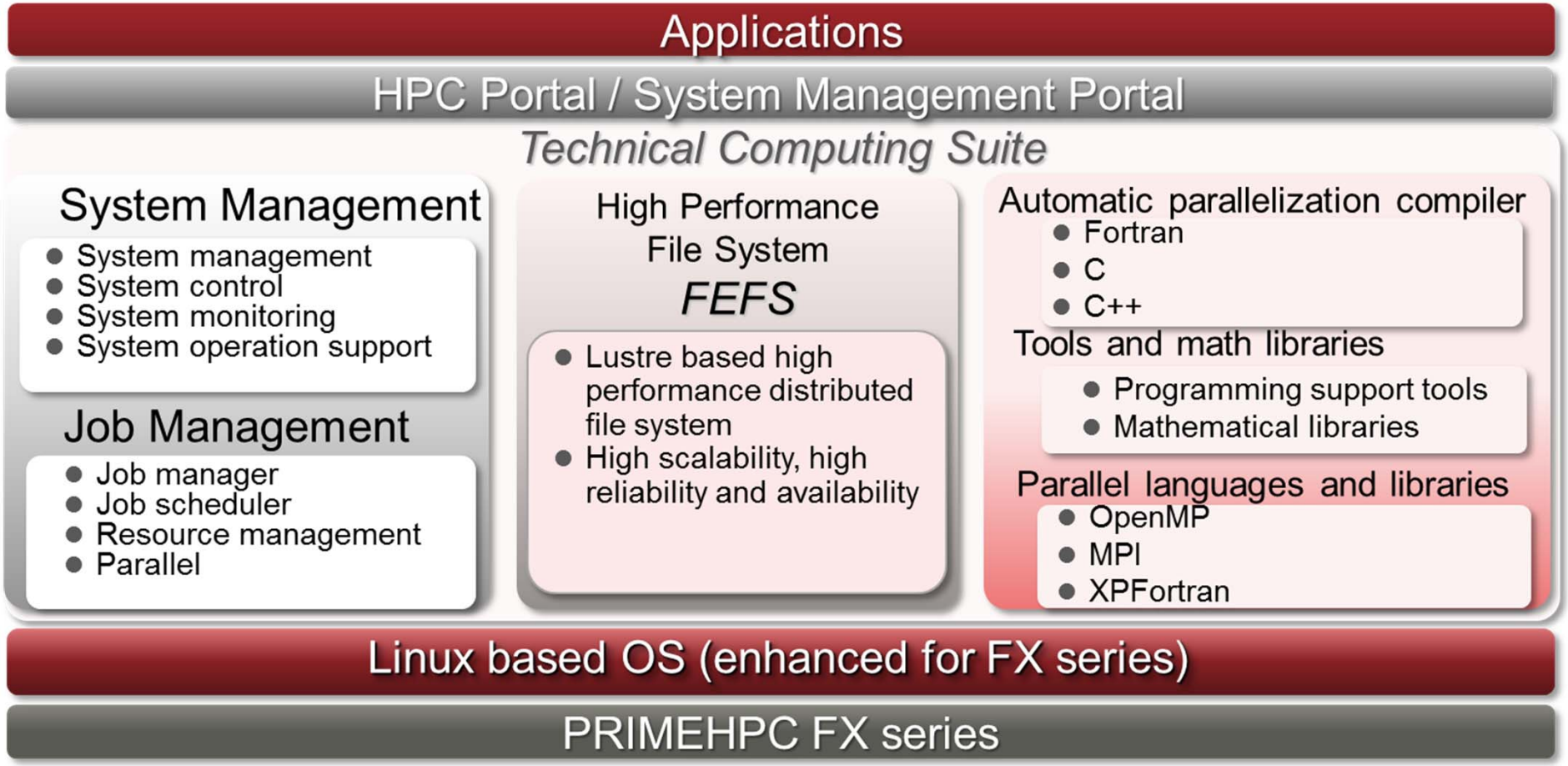
FUJITSU



- Fujitsu's math library provides 3D-FFT functionality with improved scalability for massively parallel execution
  - Significantly improved interconnect bandwidth
  - Optimised process configuration enabled by Tofu interconnect and job manager
  - Optimised MPI library provides high-performance collective communications

Enhanced Software Stack for Post-FX10

FUJITSU



Enduring Programming Model

FUJITSU

The MPI + (OpenMP or automatic-parallelization) programming model will stay important.

- VISIMPACT (Virtual Single Processor by Integrated Multi-core Parallel Architecture)
  - Automated multi-thread parallelization
    - Inherits vectorization technology
  - High performance hardware thread barrier
  - Easy hybrid parallelization reduces number of processes
    - Reducing communication overhead
    - Saving memory usage

POST-FX10

2011
FUJITSU Supercomputer PRIMEHPC FX10

2011 #1 on Top500
K computer (*)

(*) The K computer is a supercomputer jointly developed by RIKEN and Fujitsu.

2009
FX1

2000
PRIMEPOWER

1999
VPP5000

1993 #1 on Top500
NUMERICAL WIND TUNNEL

1974
FACOM-M190

2

Smaller, Faster, More Efficient

FUJITSU

- Highly integrated components with high-density packaging.
- Performance of 1-chassis corresponds to approx. 1-cabinet of K computer.

Efficient in space, time, and power

1 Fujitsu designed SPARC64™ XIfx

- 32 + 2 core CPU
- Tofu2 integrated

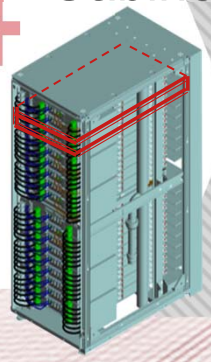
2 CPU Memory Board

- Three CPUs
- 3 x 8 HMCs (Hybrid Memory Cube)
- 8 optical modules (25Gbpsx12ch)

3 Chassis (12 CPUs)

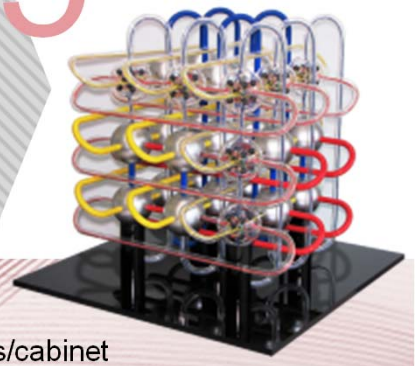
- 1 CPU/1 node
- 12 nodes/2U chassis
- Water cooled

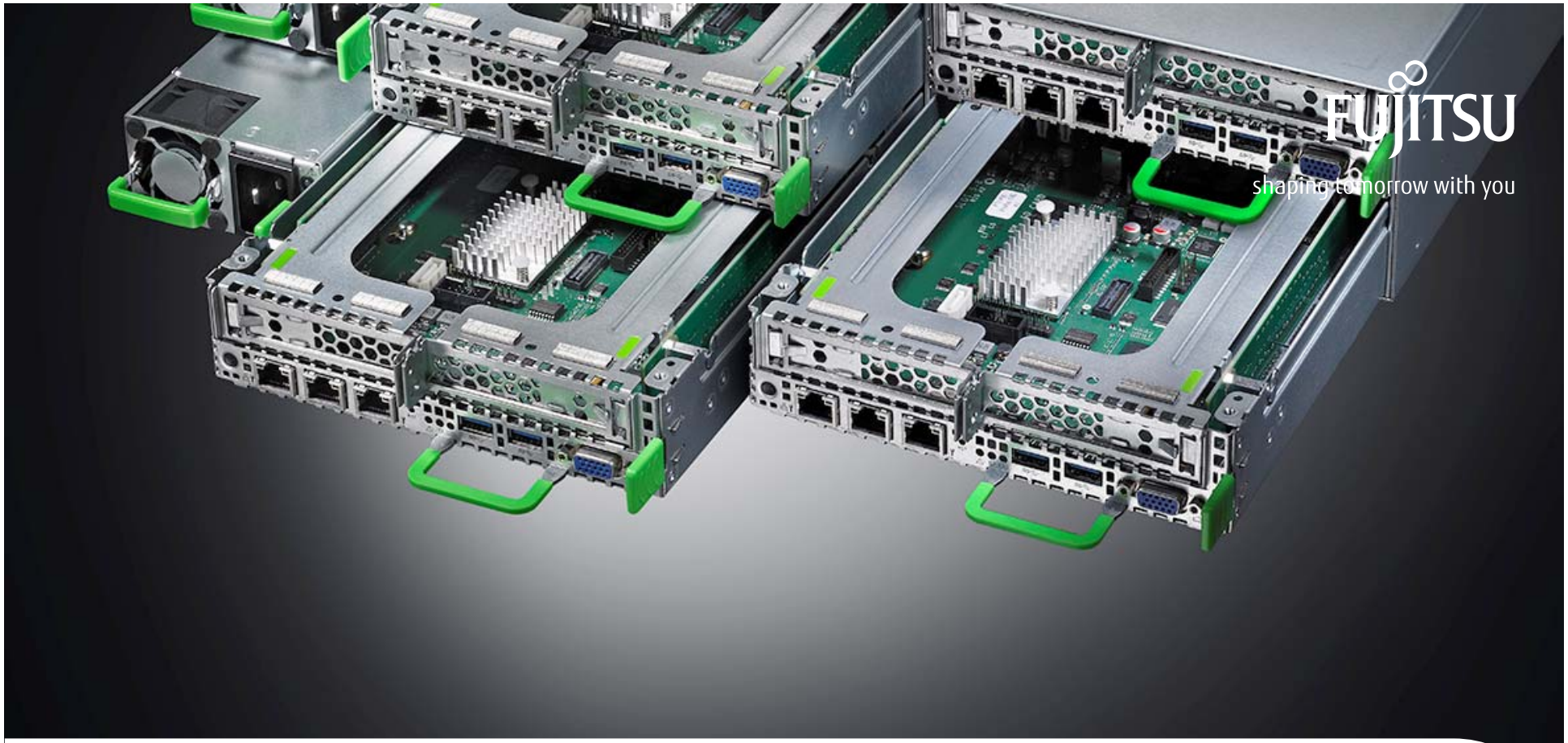
4 Cabinet



- Over 200 nodes/cabinet

5 Tofu 2





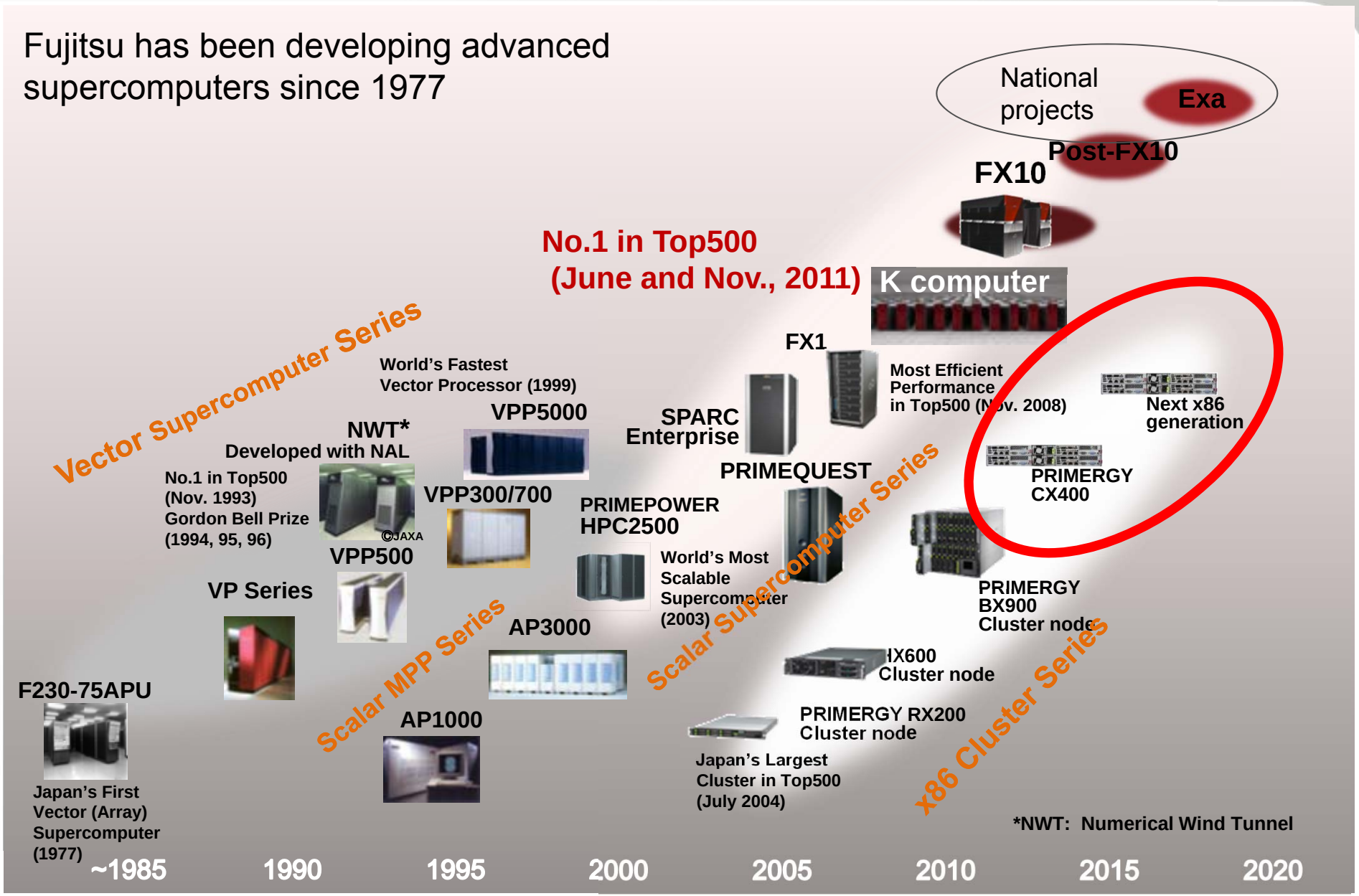
Scale-Out Smart for HPC and Cloud Computing

Fujitsu Server
PRIMERGY CX400 M1 (CX2550 M1 / CX2570 M1)

PRIMERGY x86 Cluster Series

FUJITSU

Fujitsu has been developing advanced supercomputers since 1977



Fujitsu PRIMERGY Portfolio

FUJITSU

TX Family	RX Family	BX Family	CX Family

CX400 **CX420**

Chassis

CX2550 **CX2570**

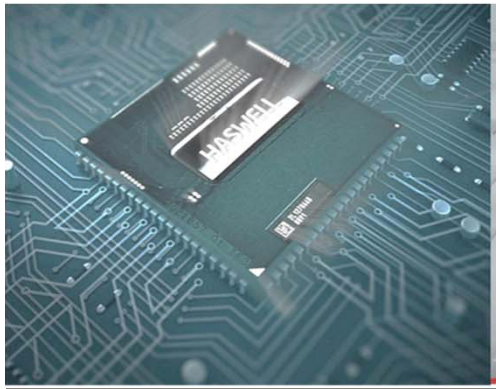
Server Nodes

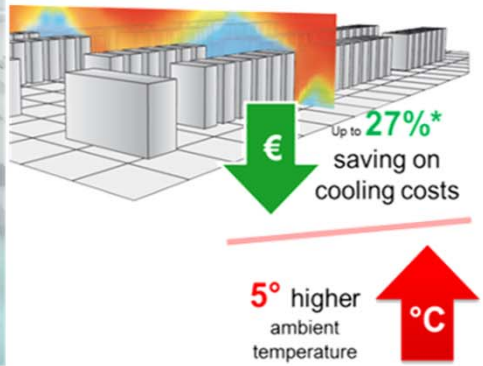
Products

Fujitsu PRIMERGY CX400 M1

Feature Overview

- 4 dual-socket nodes in 2U density
- Up to 24 storage drives
- Choice of server nodes
  - Dual socket servers with Intel® Xeon® processor E5-2600 v3 product family (Haswell)
  - Optionally up to two GPGPU or co-processor cards
- DDR4 memory technology
- Cool-safe® Advanced Thermal Design enables operation in a higher ambient temperature





Fujitsu PRIMERGY CX2550 M1

FUJITSU

Feature Overview

- Highest performance & density
  - Condensed half-width-1U server
  - Up to 4x CX2550 M1 into a CX400 M1 2U chassis
  - Intel Xeon E5-2600 v3 product family, 16 DIMMs per server node with up to 1,024 GB DDR4 memory
- High reliability & low complexity
  - Variable local storage: 6x 2.5” drives per node, 24 in total
  - Support for up to 2x PCIe SSD per node for fast caching
  - Hot-plug for server nodes, power supplies and disk drives enable enhanced availability and easy serviceability

Fujitsu PRIMERGY CX2570 M1

FUJITSU

Feature Overview

- HPC optimization
  - Intel Xeon E5-2600 v3 product family, 16 DIMMs per server node with up to 1,024 GB DDR4 memory
  - Optional two high-end GPGPU/co-processor cards (Nvidia Tesla, Grid or Intel Xeon Phi)
- High reliability & low complexity
  - Variable local storage: 6x 2.5” drives per node, 12 in total
  - Support for up to 2x PCIe SSD per node for fast caching
  - Hot-plug for server nodes, power supplies and disk drives enable enhanced availability and easy serviceability

Extreme Weather

FUJITSU

- Strong interest within Wales in impacts of extreme weather events
- Fujitsu has established collaborations in this area
  - HPC Wales studentships
  - Knowledge Transfer Partnership with Swansea University

HPC Wales – Fujitsu PhD Studentships

FUJITSU

- 20 PhD studentships in Welsh Government priority sectors including “Energy and Environment”

Cardiff University	**Dr Michaela Bray**
Extreme weather events in Wales: Weather modelling to flood inundation modelling Unified Model linked to river-estuary numerical flow model	

Bangor University	**Dr Reza Hashemi**
Simulating the impacts of climate change on estuarine dynamics Integrated catchment-to-coast model	

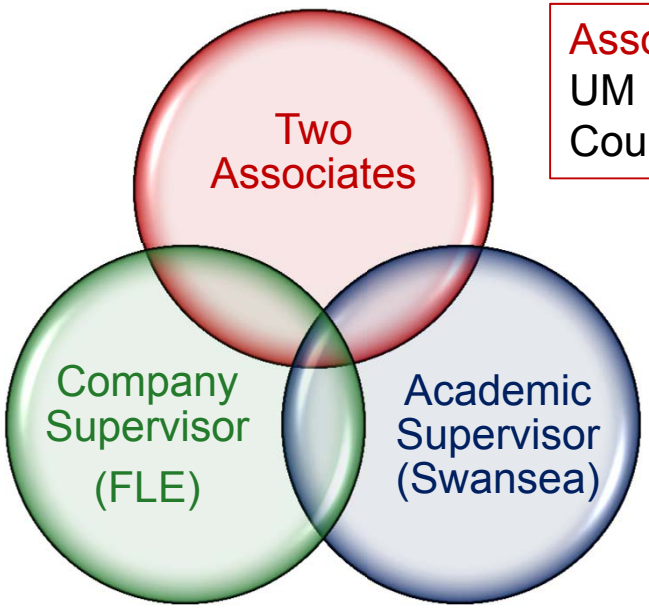
Bangor University	**Dr Simon Creer**
Biogeochemical analysis of the effects of drought on CO_2 release from peat bogs and fens	

Knowledge Transfer Partnership

FUJITSU

- UK Government program to promote skills uptake
- Partnership between Swansea University and Fujitsu
- Funding from Welsh Government including overseas visits

- Met Office Unified Model performance
- Modelling of extreme weather and flood risk



Associates
UM Optimisation
Coupling with hydrology codes

Company
Improved application performance
New solutions/services

Knowledge Base
Engineering, Swansea

Collaboration with NCI

FUJITSU

NCI

PROVIDING AUSTRALIAN RESEARCHERS WITH WORLD-CLASS HIGH-END COMPUTING SERVICES

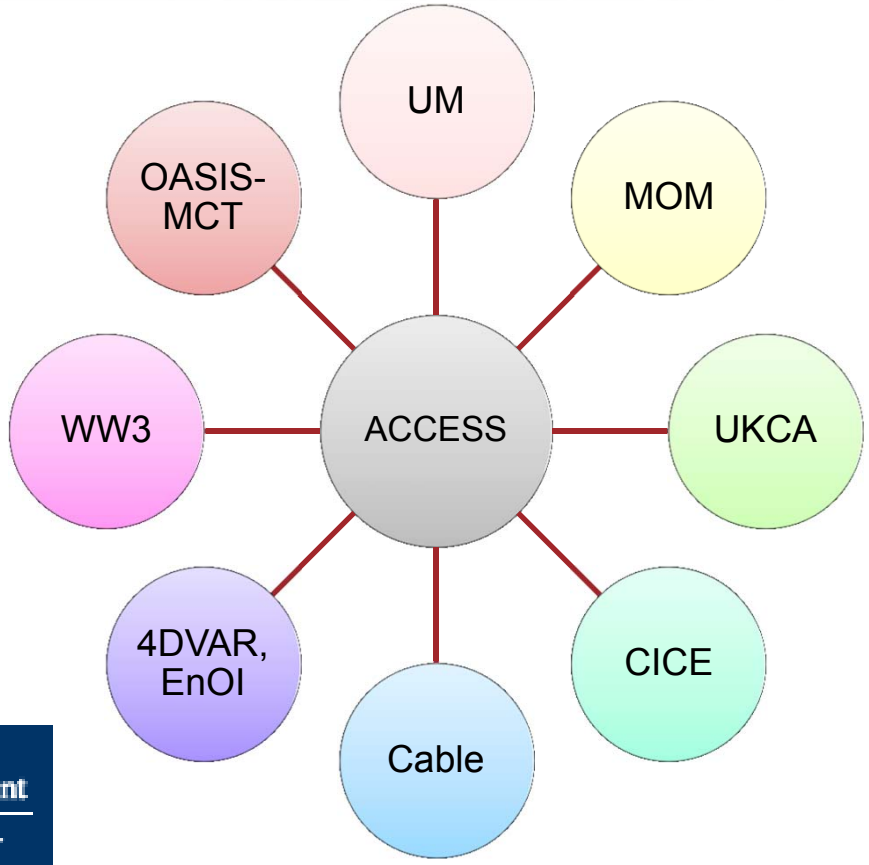
- Planned team of 10:
  4 funded positions
  + 2 in-kind (NCI) + 2 in-kind (BoM) + 2 in-kind (Fujitsu)

- Contribute to the development of NG-ACCESS

Next Generation Australian Community Climate and Earth-System Simulator (NG-ACCESS) – A Roadmap 2014-2019

ACCESS

FUJITSU



Australian Government
Bureau of Meteorology

Met Office

CSIRO

NCI

ARC CENTRE OF EXCELLENCE FOR CLIMATE SYSTEM SCIENCE

Collaborative Activities

FUJITSU

- Understanding of scalability bottlenecks
- IO server optimisation
- Improved use of thread (OpenMP) parallelism

Activities within the ACCESS Optimisation Project	
Atmosphere	• Scaling and optimisation of 2-3 global configurations of UM • Benchmark configurations of UM provided by the Bureau
Ocean	• Scaling and optimisation of two global resolution configurations of the MOM ocean model • Coupled MOM-CICE utilising the OASIS3-MCT coupler
Coupled Climate	• Performance characteristics of ACCESS-CM 1.3 and then introducing a 0.25 degree global ocean into ACCESS-CM 1.4
Data Assimilation	• Collect performance and scalability data for the various data assimilation codes used in ACCESS Parallel Suites • Initial work will focus on scalability of 4DVAR

Summary

FUJITSU

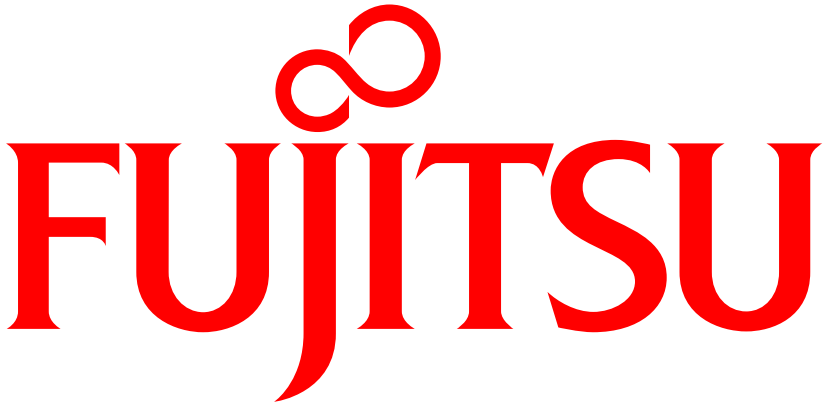
- Fujitsu is committed to developing high-end HPC platforms
  - Fujitsu selected as RIKEN's partner in FLAGSHIP 2020 project to develop an exascale-class supercomputer
  - Post-FX10 is a step on the way to exascale computing

October 1, 2014

RIKEN Selects Fujitsu to Develop New Supercomputer

Oct. 1 — Following an open bidding process, Fujitsu Ltd. has been selected to work with RIKEN to develop the basic design for Japan's next-generation supercomputer.

- Fujitsu also offers x86 cluster solutions across the range from departmental to petascale
- Fujitsu is promoting collaborative research and development programmes with key partners and customers



shaping tomorrow with you