

# FUJITSU Supercomputer PRIMEHPC FX1000/FX700

2020年8月27日  
富士通株式会社  
プラットフォーム開発本部  
プリンシパルエンジニア  
清水 俊幸

# 国家プロジェクトと PRIMEHPC シリーズ



## スーパーコンピュータ「京」



© RIKEN

## PRIMEHPC FX10



## PRIMEHPC FX100



## スーパーコンピュータ 「富岳」

© RIKEN



- スーパーコンピュータ「京」は 7年間の稼働ののち、2019年8月に運用を停止
- スーパーコンピュータ「富岳」は2020年5月に出荷を完了
  - ISC2020においてTOP500, HPCG, HPL-AI, Graph500の世界1位を同時獲得(史上初)
  - 2021年度からの共用開始に向けてソフトウェアの整備中
- 富岳テクノロジーPRIMEHPC FX1000, FX700の出荷を開始

- PSI (Petascale System Interconnect) から「富岳」へ
- TOP500 & HPCG 概観
- PRIMEHPC FX1000 と FX700
- まとめ

# 要素技術の継続的な開発と導入



## スーパーコンピュータ「京」



TOP 500  
The List.  
効率No.1 (91%, 2008)

FX1



© JAXA



© RIKEN

PRIMEHPC FX10



2012

PRIMEHPC FX100



2015

スーパーコンピュータ  
「富岳」

© RIKEN

arm



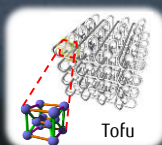
PRIMEHPC  
FX1000/FX700



PSI: 2005~2007

VISIMPACT: ハイブリッド並列  
高機能スイッチ: HWリダクション

Tofu & TBI: Tofu barrier interface



Tofu

Feasibility  
Study



2012.6~2014.3  
レイテンシコア  
の高度化研究

CMG: Core memory group  
AC: Assistant core



SPARC64VIIIfx

# FX1からFX1000への進化

## ■ CPU generations and key parameters

CPU	SPARC64 VII	SPARC64 VIIIfx	SPARC64 IXfx	SPARC64 XIfx	A64FX
1 <sup>st</sup> system w/ CPU	FX1	K	FX10	FX100	Fugaku/FX1000
Si tech. (nm)	65	45	40	20	7
Core perf. (GFLOPS)	10	16	14.8	34	57~
Cores (Assistant core)	4	8	16	32(2)	48(4)
Chip perf. (TFLOPS)	0.04	0.13	0.24	1.1	2.7~
Memory BW (GB/s)	40	64	85	480	1024
B/F (Bytes/FLOP)	1.0	0.5	0.4	0.4	0.4
Inter-core barrier	VISIMPACT ✓	✓	✓	✓	✓
Interconnect	外付けIB SW	Tofu	Tofu	Tofu2内蔵	TofuD内蔵
HW reduction	高機能SW 外付けSW	Tofuに内蔵	Tofuに内蔵	Tofu2に内蔵	TofuDに内蔵

# Fujitsu Arm CPU A64FX



## ■ ArmアーキをHPC向けに拡張

- HPC向け拡張(SVE\*)の策定に当社は“Lead partner”として貢献
- SVEをサポートする高性能コアを独自開発し、世界で初めて採用

## ■ 高いアプリ性能を実現するアーキテクチャ

- 512bit SIMD
- HBM2 x4 搭載による高いメモリバンド幅をCMG\* 構成によって有効利用

## ■ AI向け演算機能を強化

- FP16、INT16/8をサポート

\* SVE: Scalable Vector Extension、CMG: Core Memory Group



### 仕様

命令セットアーキテクチャ	Armv8.2-A SVE (512-bit SIMD)
コア数	48 コア (12コア x 4 CMG)
アシスタントコア数	計算ノード: 2コア IO兼計算ノード: 4コア
メモリ	32GiB (HBM2:8GiB×4)
製造プロセス	7nm FinFET
倍精度演算性能	3.3792 TFLOPS
メモリバンド幅	1,024 GB/s
スレッド並列サポート	コア間HWバリア
スケーラビリティ	TofuインターコネクトD内蔵

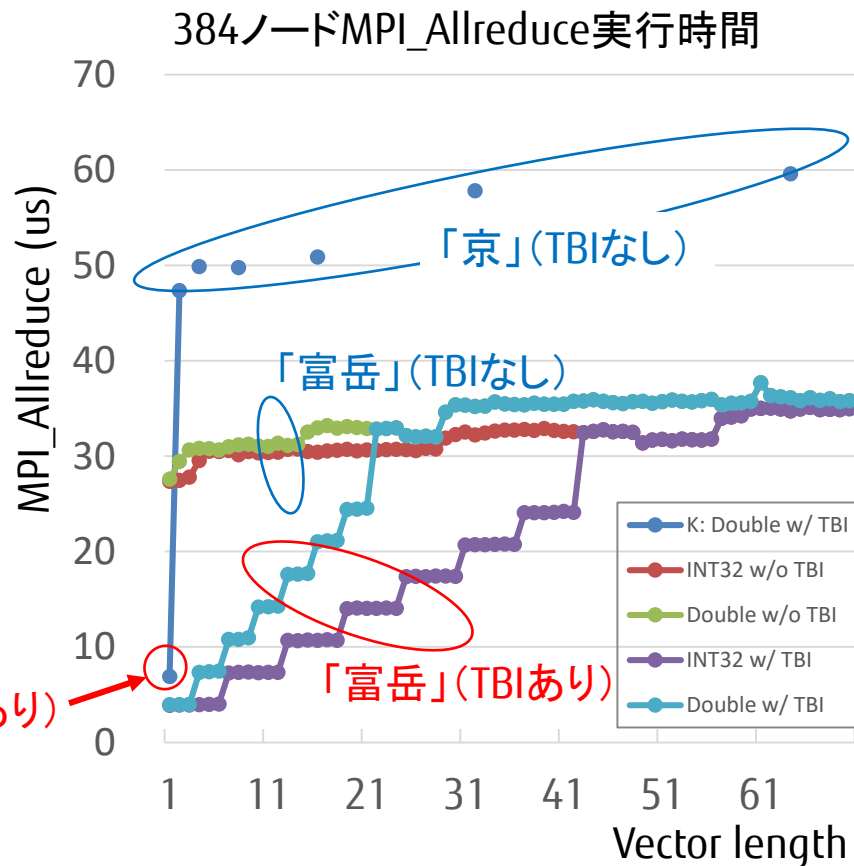
## ■ ハードリダクション(TBI)の要素数

	「京」～	「富岳」	「富岳」での活用例
浮動小数点	1	3	Complex+Double
整数	1	6	3x MAXLOC

## ■ MPI

- 要素数に応じ最適アルゴリズムを選択
  - TBI, Recursive doubling, Reduce & bcast
- TBI複数回呼び出し(「京」では1回)

「京」(TBIあり)





## ■ OSジッタ低減

### ■ ノイズ削減

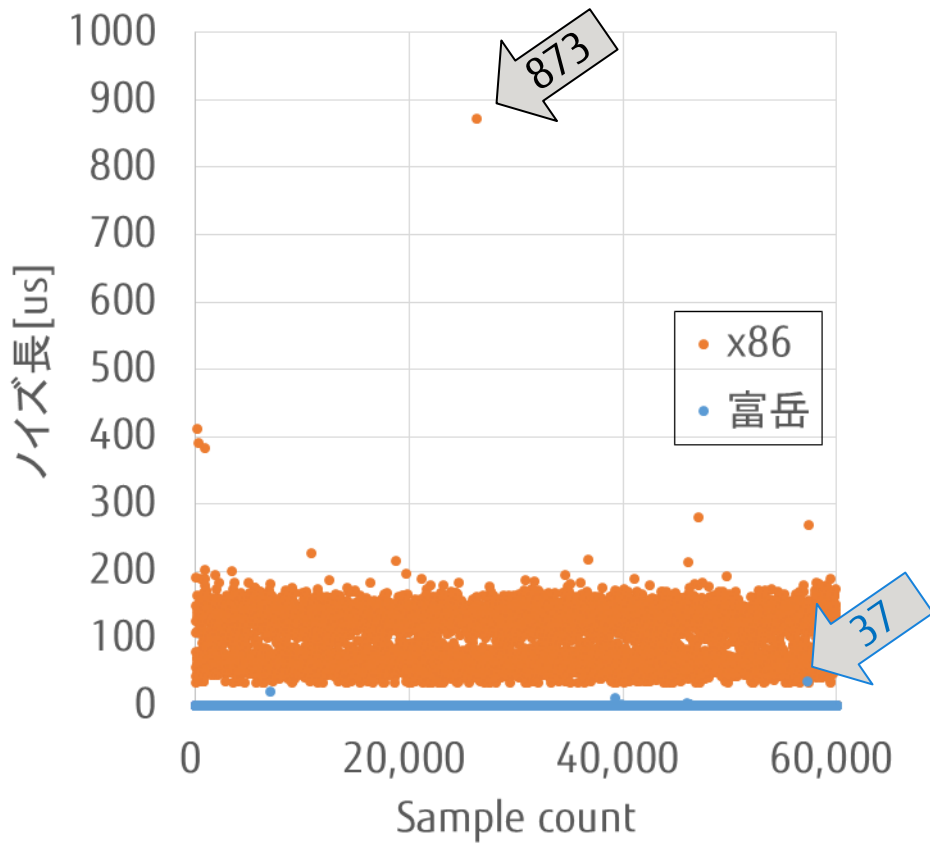
- LinuxのTicklessモードを採用し、不要なデーモンやサービスを停止

### ■ アシスタントコアの活用

- OS動作(割込/デーモン等)をアシスタントコアにオフロード

## ■ OSジッタ測定値

	平均ノイズ率	最大ノイズ長
「富岳」	7.1E-07	37 $\mu$ s
「京」	6.6E-5	85 $\mu$ s
x86	3.7E-03	873 $\mu$ s





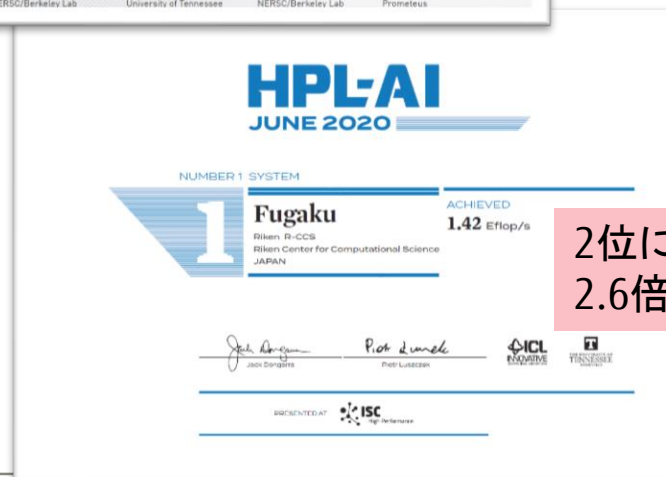
# 「富岳」の世界記録(4冠) 2020.6.22



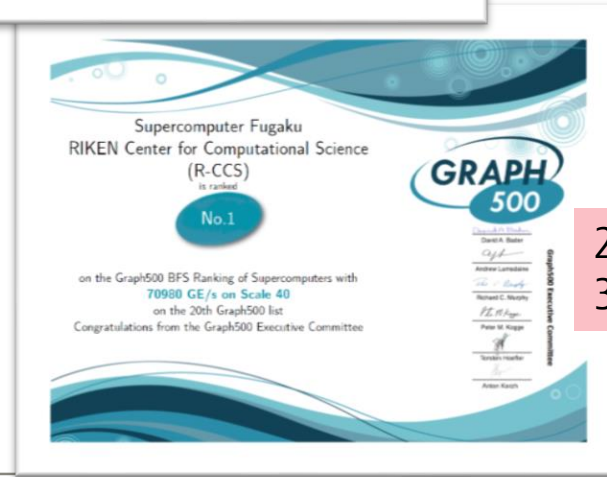
2位に対して  
2.8倍の性能



2位に対して  
4.6倍の性能



2位に対して  
2.6倍の性能

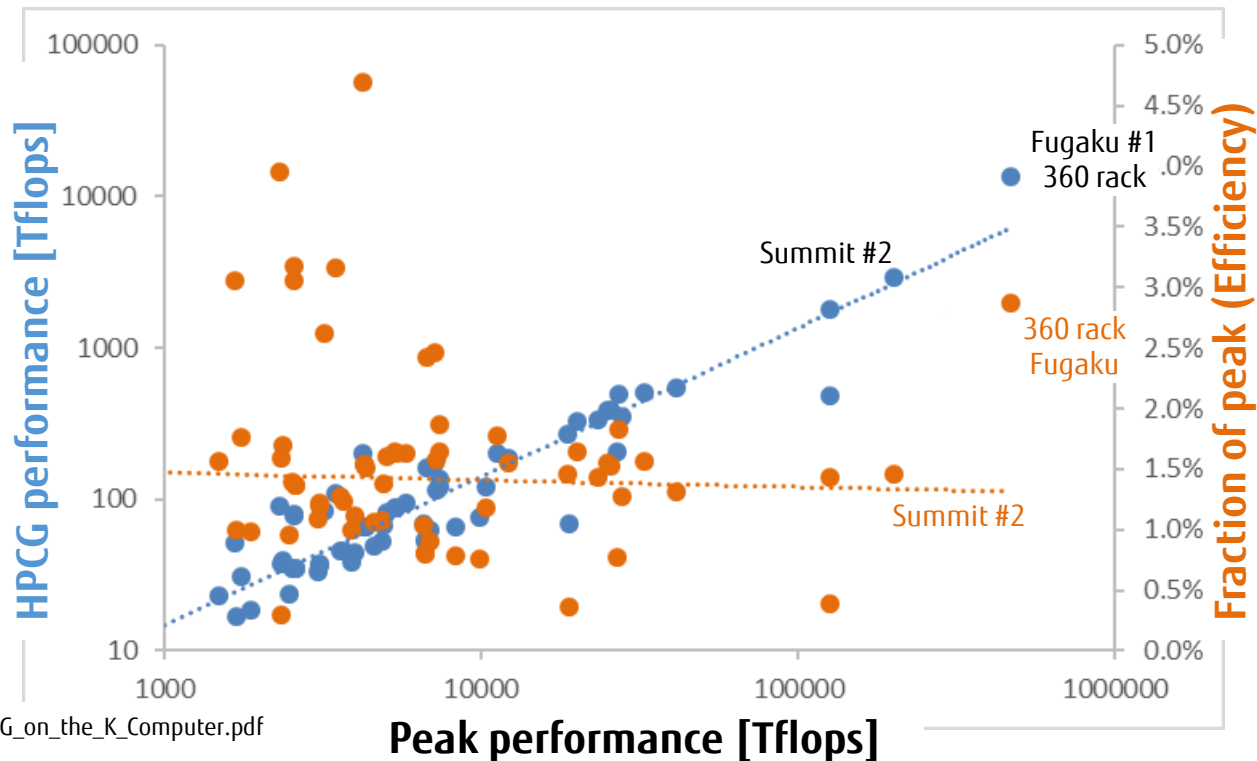


2位に対して  
3.0倍の性能

# HPCG (June 2020)の性能と効率

■「富岳」の演算効率は約 3% で他に比べ高い

- Symmetric Gauss-Seidelループ分割最適化[\*]
- Tofu 6方向同時通信による袖領域の通信
- Tofuのリダクション通信のハードウェアサポート



[\*] [http://www.hpcg-benchmark.org/downloads/sc16/HPCG\\_on\\_the\\_K\\_Computer.pdf](http://www.hpcg-benchmark.org/downloads/sc16/HPCG_on_the_K_Computer.pdf)

ランク	システム名	アクセラレータ無	コア数	演算性能 [TFlop/s]	電力当たり性能 [GFlops/W]
1	MN-3	-	2080	1621	21.108
2	Selene	-	272800	27580	20.518
3	NA-1	-	1271040	1303	18.433
▶ 4	<b>A64FX prototype</b>	✓	36864	1999.5	<b>16.876</b>
5	AiMOS	-	130000	8339	16.285
6	HPC5	-	669760	35450	15.740
7	Satori	-	23040	1464	15.574
8	Summit	-	2414592	148600	14.719
▶ 9	<b>Supercomputer Fugaku</b>	✓	7299072	415530	<b>14.665</b>
10	Marconi-100	-	347776	21640	14.661
:			:		
▶ 39	<b>Sunway TaihuLight</b>	✓	10649600	93015	<b>6.051</b>

× 2.8

アクセラレータ無しシステムでの次点(Sunway TaihuLight)の**2.4~2.8倍**の電力当たり性能

# スパコンのランキングについて

- TOP500: Linpackベンチマーク( $Ax = b$ を解く)性能を順位付け
- HPCG: 実用に近い疎行列を解く性能を順位付け
- Green500: TOP500のスパコンの電力あたり性能を順位付け
  - #1はPFN(Preferred Networks)のMN-3で、21.1GF/W
- Graph500: グラフ問題を解く性能を順位付け
- HPL-AI: 16ビット演算を活用したLinpack実行性能を評価

TOP500 順位	コンピュータ名	TOP500 (TF)	(効率)	HPCG (TF)	(効率)	Green500 (GF/W)	Graph500 (GSTEPS)	HPL-AI (PF)
1	Fugaku	415,530	80.9%	13,366	2.86%	14.665	70,980	1,421
2	Summit	148,600	74.0%	2,925	1.46%	14.719	7,666	550
3	Sierra	94,640	75.3%	1,795	1.43%	12.723	67	-
4	Sunway TaihuLight	93,014	74.2%	480	0.38%	6.051	23,756	-
204	A64FX prototype	1,999	84.7%	-	-	16.876	-	-
(20※)	「京」コンピュータ	10,510	93.2%	602	5.34%	0.83	31,302	-

赤字が2020年6月時点での1位の性能、緑字は前回の1位 ※「京」コンピュータは2019年8月に撤去のため参考

# 「富岳」性能予測(理研様まとめ)

<https://postk-web.r-ccs.riken.jp/perf.html>

■ GENESIS, NICAM+LETKFで  
100倍超を  
達成見込み

## □ Performance Targets

- ✓ 100 times faster than K for some applications (tuning included)
- ✓ 30 to 40 MW power consumption

## □ Peak Performance

	PostK	K
Peak DP (double precision)	400+ Pflops (34x +)	11.3 Pflops*
Peak SP (single precision)	800+ Pflops (70x +)	11.3 Pflops
Peak HP (half precision)	1600+ Pflops (141x +)	--
Total memory bandwidth	150+ PB/sec (29x +)	5,184TB/sec

\* Reported in TOP500 (including I/O nodes)

## □ Geometric Mean of Performance Speedup of the 9 Target Applications over the K-Computer

37x +

## □ Predicted Performance of 9 Target Applications As of 2019/05/14

Area	Priority Issue	Performance Speedup over K	Application	Brief description
Health and longevity	1. Innovative computing infrastructure for drug discovery	125x +	GENESIS	MD for proteins
	2. Personalized and preventive medicine using big data	8x +	Genomon	Genome processing (Genome alignment)
Disaster prevention and Environment	3. Integrated simulation systems induced by earthquake and tsunami	45x +	GAMERA	Earthquake simulator (FEM in unstructured & structured grid)
	4. Meteorological and global environmental prediction using big data	120x +	NICAM+LETKF	Weather prediction system using Big data (structured grid stencil & ensemble Kalman filter)
Energy issue	5. New technologies for energy creation, conversion / storage, and use	40x +	NTChem	Molecular electronic simulation (structure calculation)
	6. Accelerated development of innovative clean energy systems	35x +	Adventure	Computational Mechanics System for Large Scale Analysis and Design (unstructured grid)
Industrial competitiveness enhancement	7. Creation of new functional devices and high-performance materials	30x +	RSDFIT	Ab-initio simulation (density functional theory)
	8. Development of innovative design and production processes	25x +	FFB	Large Eddy Simulation (unstructured grid)
Basic science	9. Elucidation of the fundamental laws and evolution of the universe	25x +	LQCD	Lattice QCD simulation (structured grid Monte Carlo)

## ■ 「富岳」の3つの目標

- アプリケーション性能
- 省電力
- 使いやすさ

## ■ 状況

- アプリケーション性能
  - ・ 重点課題アプリ (GENESIS, NICAM+LETKF) は、「京」の100倍性能に目途
  - ・ ベンチマークランキングで4冠
- 省電力
  - ・ Green500で上位 (アクセラレータ無しのシステムでは1位)
- 使いやすさ
  - ・ Arm Linux (Red Hat Enterprise Linux) サポート

「富岳」の頂は見たが、  
すそ野は十分に広がっているか？

長いスパコン開発経験、PSIから取り組んできた要素技術開発、「富岳」でのコデザインなど  
**超並列システム(フラッグシップマシン)の高効率アーキテクチャ技術・開発の結実**

# TOP500 (June 2020)のスパコンのインターコネクト



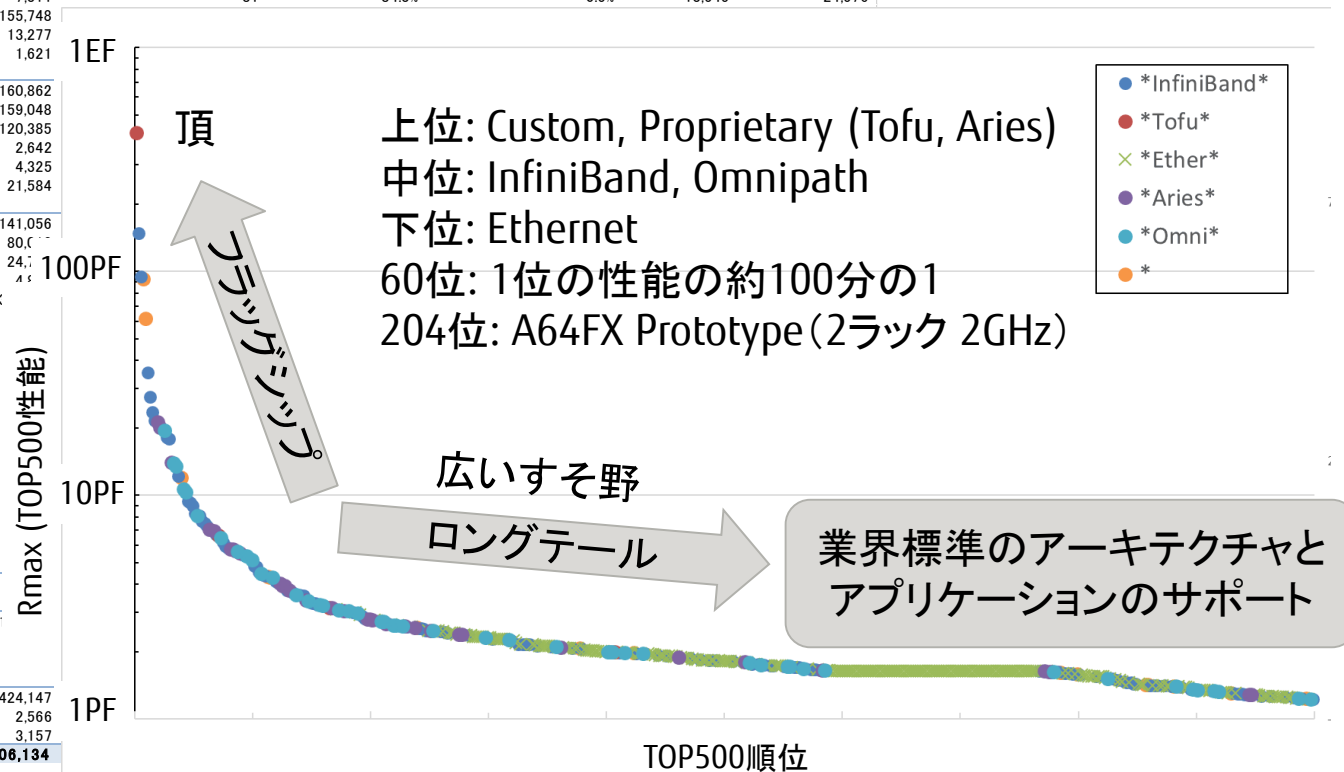
行ラベル	↓	↑	カウント / 合計 / Rmax [TFlop/s]	合計 / HPCG [TFlop/s]	平均 / Efficiency	平均 / HPCG Fraction	平均 / Socks	最大値 / Socks	
Custom Interconnect									
Sunway			1	93,015	481	74.2%	0.4%	40,960	40,960
TH Express-2			2	63,516	38	64.2%	0.6%	18,688	35,584
Custom Interconnect			3	7,311	81	84.3%	0.5%	13,545	24,576
Aries interconnect			34	155,748	1,977	73.0%	0.5%	6,616	18,832
Bull BX1 1.2			2	13,277	371	53.6%	1.5%	4,542	8,256
RoCEv2/MN-Core DirectConnect			1	1,621		41.3%	0.0%	80	80
Gigabit Ethernet									
10G Ethernet			95	160,862		50.5%	0.0%	6,544	165,200
25G Ethernet			80	159,048		50.7%	0.0%	1,888	8,300
40G Ethernet			73	120,385		77.8%	0.0%	2,880	2,880
100Gb 6D-Torus			2	2,642		60.3%	0.0%	1,880	2,000
200Gb 6D-Torus			1	4,325		70.5%	0.0%	1,024	1,024
100G Ethernet			11	21,584		52.4%	0.0%	126	160
Infiniband									
Infiniband EDR			48	141,056	851	60.4%	0.2%	2,371	26,000
Infiniband FDR			36	80,013	722	68.8%	0.5%	4,742	24,111
Mellanox InfiniBand HDR			2	24,782	18	64.2%	0.5%	8,346	16,016
Infiniband QDR			3	4,880	51	66.0%	1.0%	8,421	9,648
Dual-rail Mellanox EDR Infiniband			8	311,871	4,721	72.9%	0.4%	2,913	9,216
Mellanox HDR Infiniband			4	76,714	1,370	70.2%	0.8%	3,312	6,368
Infiniband FDR14			1	2,814		78.6%	0.0%	6,144	6,144
4xEDR Infiniband			2	3,159	91	72.6%	2.0%	3,225	5,130
Mellanox EDR InfiniBand/ParTec ParaStz			1	6,178	75	62.5%	0.8%	4,770	4,770
Mellanox InfiniBand HDR100			5	24,800		68.7%	0.0%	2,943	4,608
Infiniband HDR			7	14,545	44	56.2%	0.1%	2,066	4,560
EDR Infiniband/ 8D Hypercube			1	4,851		66.8%	0.0%	4,200	4,200
Infiniband EDR/HDR			2	8,156	45	63.6%	0.6%	3,231	4,158
InfiniBand HDR100			6	14,523		62.6%	0.0%	2,098	4,032
56Gb 3D-Torus			1	1,428		74.4%	0.0%	4,000	4,000
Mellanox InfiniBand EDR			12	62,946	974	68.7%	1.1%	1,020	4,000
Infiniband EDR/FDR			1	1,228	22	46.5%	0.8%	2,080	2,080
Infiniband			1	1,757		65.7%	0.0%	1,740	1,740
Myrinet									
Myrinet 10G			1	1,975		29.9%	0.0%	5,600	5,600
OmniPath									
Intel Omni-Path			46	180,770	1,752	65.5%	0.3%	2,982	12,744
Dual-rail Intel Omni-Path			1	3,082		62.0%	0.0%	2,232	2,232
Intel Omni-Path 100G			1	1,403		45.7%	0.0%	192	192
Proprietary Network									
Tofu interconnect D			3	424,147	13,597	83.5%	1.9%	51,712	152,064
Proprietary			1	2,566		54.6%	0.0%	14,336	14,336
Tofu interconnect 2			1	3,157	110	90.7%	3.2%	3,443	3,443
総計			500	2,206,134	27,391	62.2%	0.2%	4,226	165,200

\* 291位 Sugon TC6000, AMD Epyc 7501 32C 2GHz, AMD Vega 20, Infiniband HDR のRpeakについては9,404 TFとして計算



# TOP500 (June 2020)のスパコンのインターコネクト

行ラベル	↓	↑	カウント / 合計 / Rmax [TFlop/s]	合計 / HPCG [TFlop/s]	平均 / Efficiency	平均 / HPCG Fraction	平均 / Socks	最大値 / Socks
<b>Custom Interconnect</b>								
Sunway			1	93,015	481	74.2%	0.4%	40,960
TH Express-2			2	63,516	38	64.2%	0.6%	18,688
Custom Interconnect			3	7,311	81	84.3%	0.5%	13,545
<b>Aries interconnect</b>								
Bull BX1 1.2			2	13,277				24,576
RoCEv2/MN-Core DirectConnect			1	1,621				
<b>Gigabit Ethernet</b>								
10G Ethernet			95	160,862				
25G Ethernet			80	159,048				
40G Ethernet			73	120,385				
100Gb 6D-Torus			2	2,642				
200Gb 6D-Torus			1	4,325				
100G Ethernet			11	21,584				
<b>Infiniband</b>								
Infiniband EDR			48	141,056				
Infiniband FDR			36	80,000				
Mellanox InfiniBand HDR			2	24,000				
Infiniband QDR			3	1,000				
Dual-rail Mellanox EDR Infiniband			8	1,000				
Mellanox HDR Infiniband			4	1,000				
Infiniband FDR14			1	1,000				
4xEDR Infiniband			2	1,000				
Mellanox EDR InfiniBand/ParTec ParaStx			1	1,000				
Mellanox InfiniBand HDR100			5	1,000				
Infiniband HDR			7	1,000				
EDR Infiniband/ 8D Hypercube			1	1,000				
Infiniband EDR/HDR			2	1,000				
Infiniband HDR100			6	1,000				
56Gb 3D-Torus			1	1,000				
Mellanox InfiniBand EDR			12	1,000				
Infiniband EDR/FDR			1	1,000				
Infiniband			1	1,000				
<b>Myrinet</b>								
Myrinet 10G			1	1,000				
<b>Omnipath</b>								
Intel Omni-Path			46	1,000				
Dual-rail Intel Omni-Path			1	1,000				
Intel Omni-Path 100G			1	1,000				
<b>Proprietary Network</b>								
Tofu interconnect D			3	424,147				
Proprietary			1	2,566				
Tofu interconnect 2			1	3,157				
総計			500	2,208,134				



# PRIMEHPCシリーズ FX1000とFX700

## ■ 「富岳」に採用したCPU「A64FX」を搭載した 2 種類のスパコン製品をラインアップ



FX1000

FX700

### コンセプト

「富岳」の開発を通じて培った技術を適用。  
高性能、高拡張性、高信頼性に加え、世界最高レベルの超低消費電力を実現。  
最大で1.3EFLOPS以上のシステムを構成可能で大規模ユーザー向けに最適なシステム。

スタンダード技術をベースとした高性能Armサーバ。  
小規模から導入可能なフォームファクタと空冷方式を採用。A64FXの高パフォーマンスの手軽な導入や、「富岳」向けアプリケーション開発に最適なシステム。  
オープンなソフトウェアスタックをサポート。

### CPU

A64FX 384ノード/ラック

A64FX 8ノード/シャーシ

### インターコネクト

TofuインターコネクトD

**InfiniBand EDR/HDR100**

### 冷却方式

水冷

**空冷**

### ラック/シャーシ

専用ラック

2Uラックマウントタイプ・シャーシ

### ソフトウェアスタック

FUJITSU Software Technical Computing Suite

**FUJITSU Software Compiler Package, OpenHPC,  
Bright Cluster Manager**

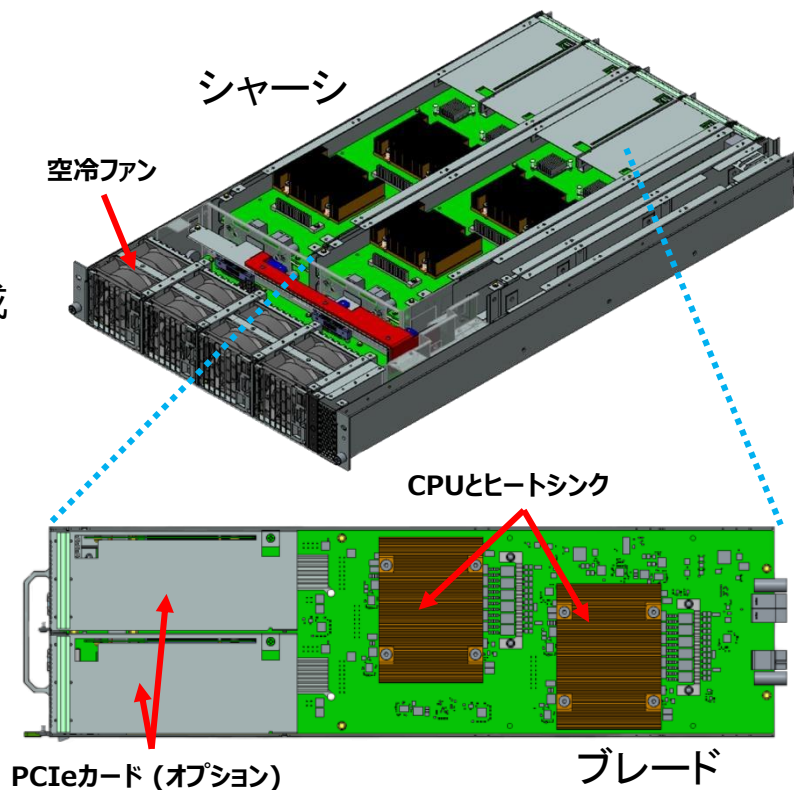
## ■ スタンダード技術をベースとしたデザイン

### ■ 導入しやすいフォームファクタと空冷方式

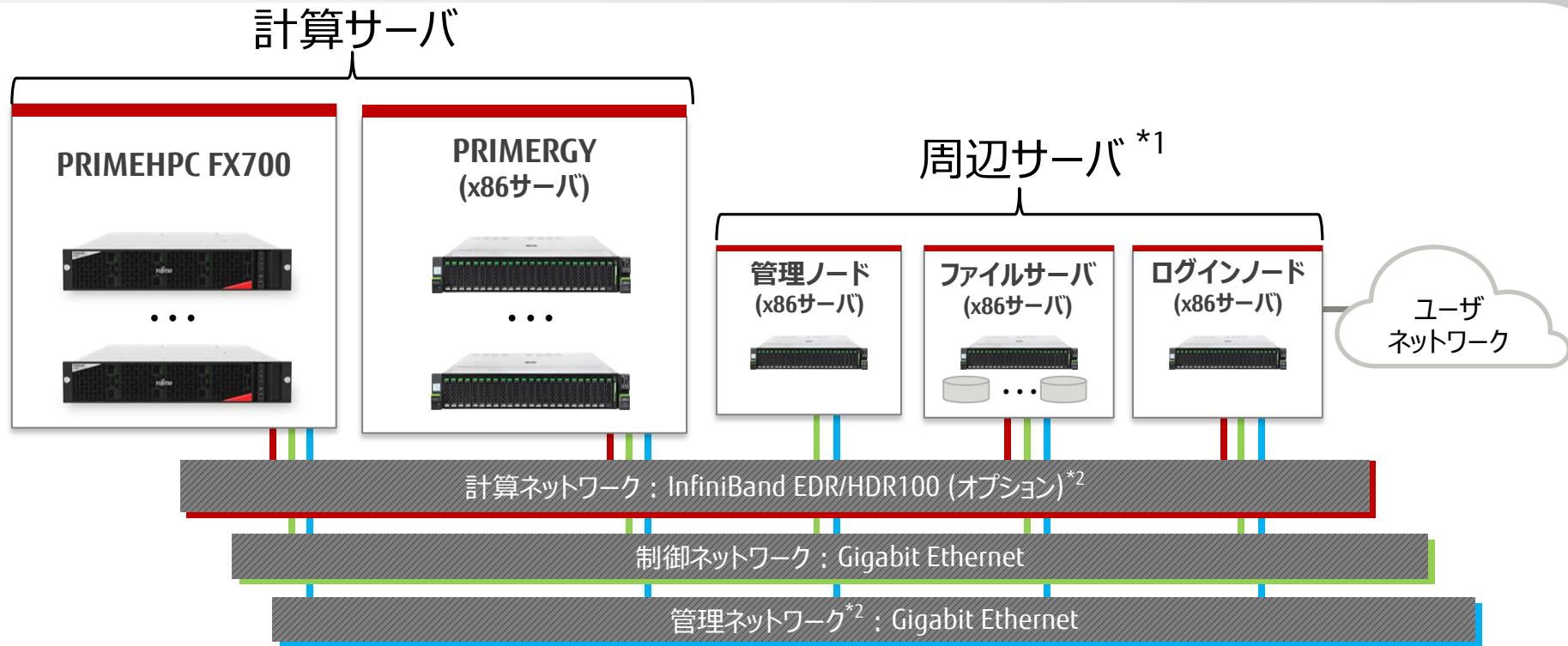
- 19インチラックに搭載可能な2U筐体
- シャーシあたり最大4ブレードを搭載
- 1ブレードに2ノード実装し、シャーシに最大8ノード構成
- 冷却に空冷方式を採用
- 水冷設備がない環境に容易に導入可能

### ■ 業界標準のインターコネクト、OSを採用

- 計算ノード間インターコネクトにInfiniBandを採用
- OSはRedHat Enterprise Linuxを採用
- 高いソフトウェア移行性を実現



# 全体システム構成

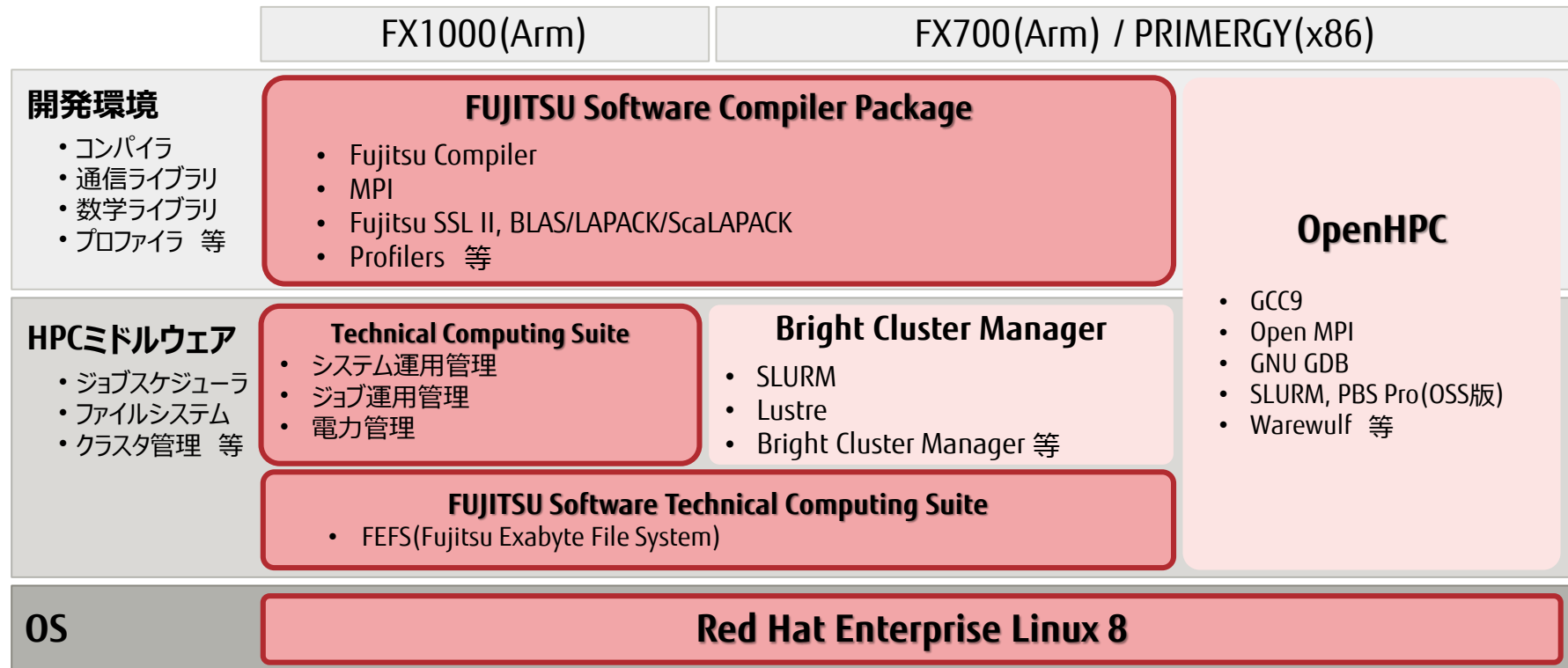


(\*1) 周辺サーバの必要台数・構成・スペックは、ソフトウェアスタックの種類やシステム規模等により異なる

(\*2) PRIMEHPC FX700ではInfiniBand EDR/HDR100はオプション

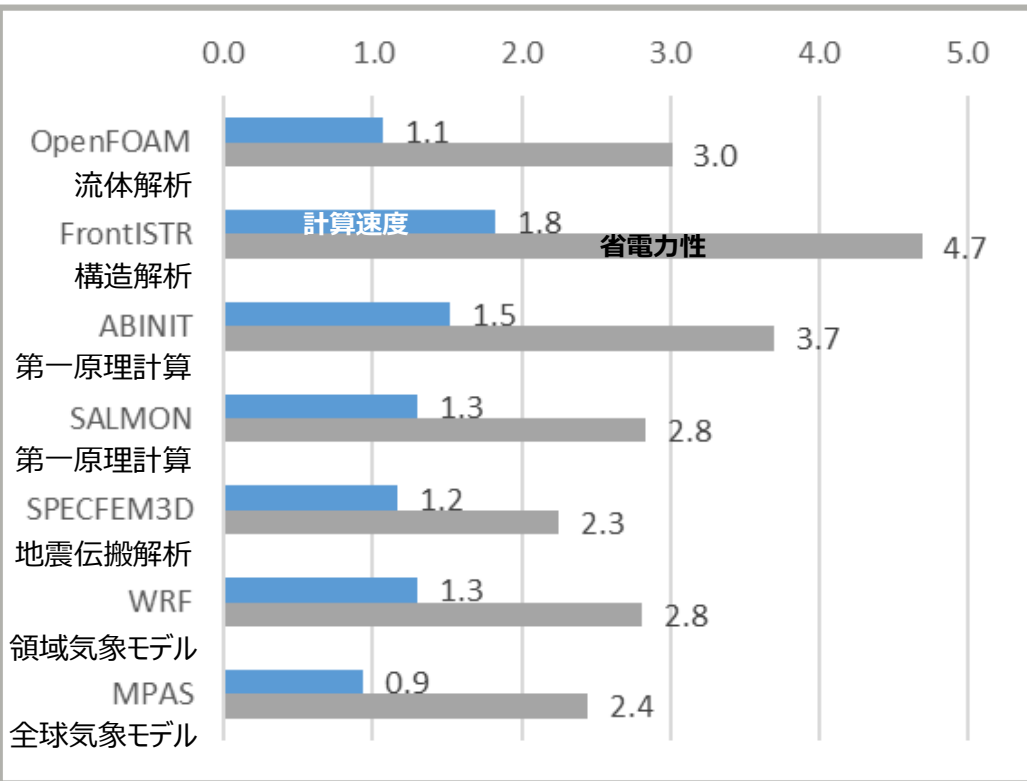
InfiniBand EDR/HDR未選択の場合、計算ネットワークは管理ネットワーク(Gigabit Ethernet)と兼用

## ■ A64FXの性能を引き出す開発環境を提供



# A64FXの計算速度と省電力性（OSS実アプリ）

## x86 CPU x2に対するA64FX x1の性能倍率



■ 同一コア数での比較で、A64FX<sup>\*</sup>は最新のx86 CPU<sup>\*</sup>と比較して、

■ 計算速度で～1.8倍

■ 省電力性で～4.7倍

\* A64FX : 48コア×1CPU(2.2GHz)、  
x86 CPU : 24コア×2CPU(2.9GHz)

■ 高い性能倍率を以下により実現

■ HPC向けに強化したマイクロアーキテクチャ

■ 省エネルギーを意識した設計と実装

# 商用アプリケーションの拡大

- ISV各社と協同し、商用アプリのFX1000対応に取り組んでいます。  
バイナリー互換のFX700と「富岳」にも対応します。

近日、提供予定

研究開発中

(2020年6月)

## エンジニアリング分野 (構造解析、流体解析、エレクトロニクス)

LS-DYNA

(by Ansys, Inc.)

Poynting

(by Fujitsu Limited)

化学分野\*

Amber

Gaussian16

(by Gaussian, Inc.)

ADVENTURECluster

(by Allied Engineering Co.)

 **CONVERGE**  
CFD SOFTWARE

(by Convergent Science)

Marc

(by MSC Software Ltd.)

VASP

\*協力: オーストラリア国立大学

Altair Radioss™

(by Altair Engineering, Inc.)

**HELYX®**

(by ENGYS Ltd. & VINAS Co., Ltd.)

scFLOW

(by Software Cradle Co., Ltd.)

VPS (PAM-CRASH)

(by ESI Group)

Ansys Fluent

(by Ansys, Inc.)

**JMAG®**

(by JSOL Corporation)

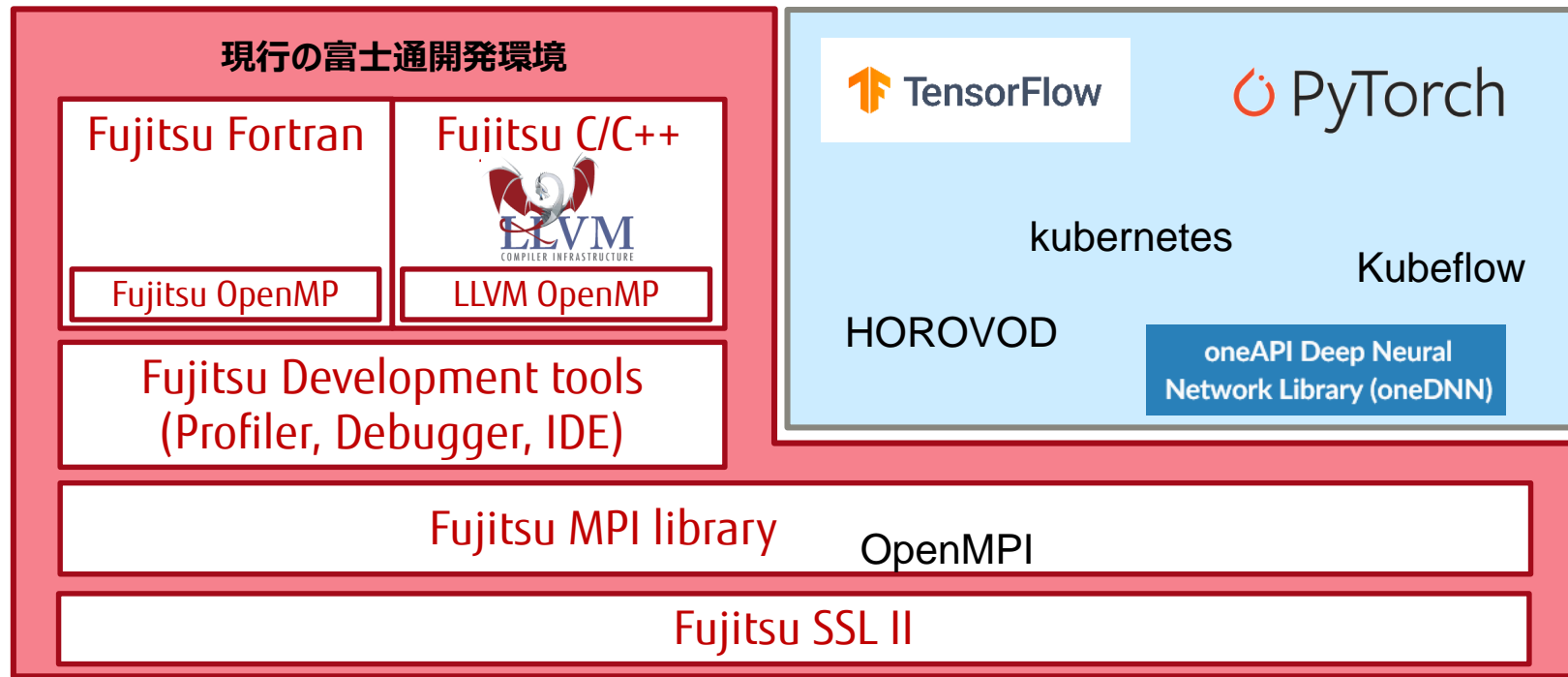
Simcenter STAR-CCM+

(by Siemens Industry Software Inc.)

\*\*スライド内の全てのアプリケーション名は、各ベンダーの商標または登録商標です。



## ■ OSSを活用したHPC × AI環境の実現に向け取り組み中



TensorFlow, the TensorFlow logo and any related marks are trademarks of Google Inc.

# スーパーコンピュータが支える社会課題の解決

～ スパコンを活用したイノベーションが期待される分野例 ～

SDGsの  
目標



具体的分野

ライフサイエンス

エネルギー

ものづくり

防災・環境

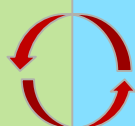
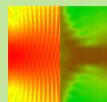
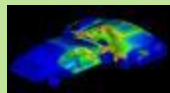
将来

現在

過去

## 【シミュレーション】

- 流体シミュレーション
- 衝突シミュレーション
- 材料シミュレーション
- 



## 【データ解析】

- AI(Deep Learning)
- ビッグデータ
- 



～ スーパーコンピュータ ～

シミュレーションとデータ解析の両輪で、スパコンは社会課題の解決やDXを支えるインフラとして利用が拡大

- 富士通が40年に渡って培った、ハードウェアからソフトウェアにいたるスパコン技術を全て結集することで、巨大システム「富岳」の実現に貢献
- 「富岳」の高いアプリケーション性能を目指した設計と実装により、多くのベンチマークランキングで首位を獲得
- 「富岳」の世界最高速の性能は、高いシミュレーション・AI処理能力が必要となる、Society 5.0の実現に貢献できると期待
  - COVID-19対策に向けた先行利用の取り組み、等
- 「富岳」で採用した技術を広く、グローバルに展開 (PRIMEHPC、A64FX)

富士通はスパコンの生み出す成果を通じて、豊かで夢のある未来を世界中の人々に提供することを目指します。デジタル時代を迎え、高いシミュレーション・AI処理能力を有するスパコンは、より一層、様々な分野での活用が広がると考えています。

## ■ 社外公開サイトのご案内

- 公開サイトにて PRIMEHPC の製品概要・仕様等掲載しております。

日本語 URL: <https://www.fujitsu.com/jp/supercomputer/>

英語 URL: <https://www.fujitsu.com/supercomputer/>

- ダウンロードコンテンツ (PDF) URL: <https://www.fujitsu.com/jp/supercomputer/downloads/>

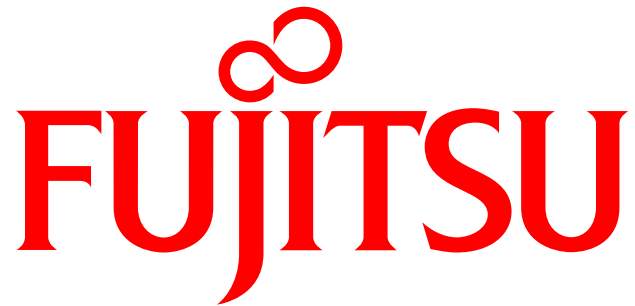
- ◆ データシート【日・英】

- ◆ ホワイトペーパー【日・英】

- ハードウェア
- ソフトウェア

こちらからもアクセスできます⇒





shaping tomorrow with you