

富士通HPCの過去、現在、未来 ～これまでの20年と未来に向けて～

2021.12.8

富士通株式会社 ICTシステム研究所

中島 耕太



これまでの20年

- Myrinet (1994)
 - NIC上のプロセッサでプロトコル処理をオフロード
 - さまざまなプロトコル(AM/FM/PM)がMyrinet上で開発
- SCoreプロジェクト (RWCP)
 - SCore: PC cluster operating system
 - PM: zero-copy data transfer library
- PCクラスタの実装:
 - RWC PC Cluster II (*)
 - Pentium Pro 200MHz /Myrinet/128 nodes
 - Score Cluster III
 - Pentium III 933MHz/Myrinet 2000/512 nodes



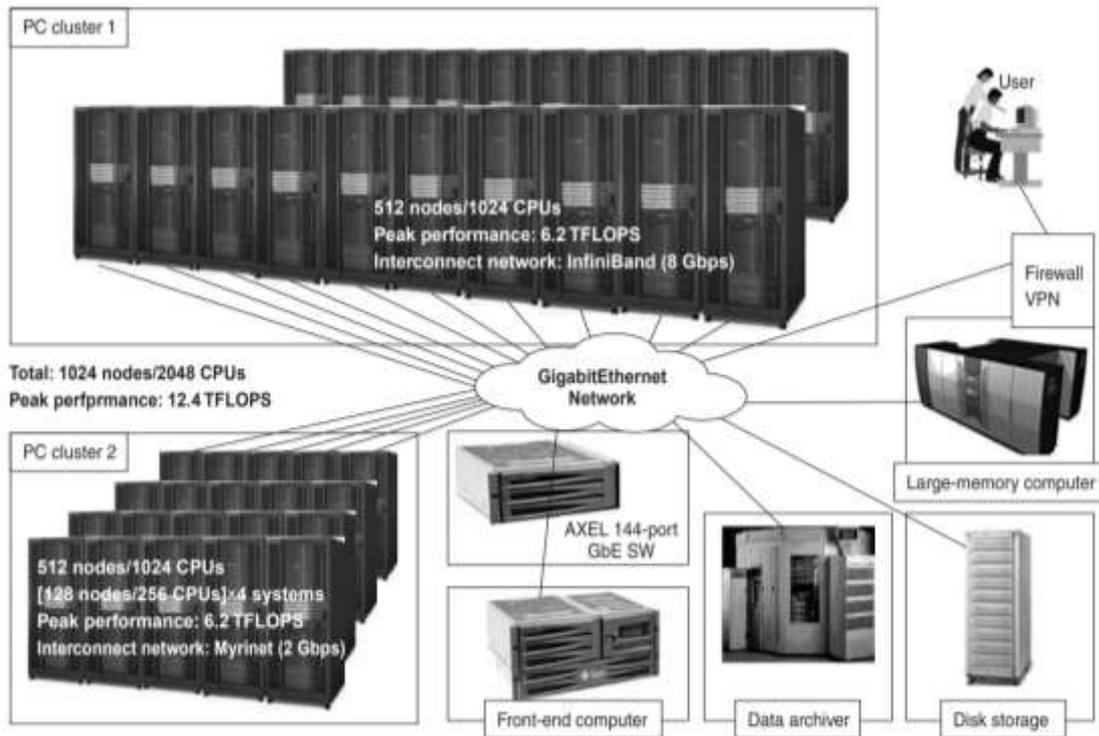
(*) H. Tezuka, A. Hori, Franci O'Carroll, Y. Ishikawa:
RWC PC Cluster II and its performance,
IPJS SIG Technical Report, 1997-ARC-128

Myrinetの登場によりネットワークボトルネックを解消、PCクラスタの実用化に弾み

Riken Super Combined Cluster (RSCC)

8.728TFlops, 7th/Top500 (2004.6)

- 共同研究センタースパコンとして日本で最初のPCクラスタ
- 当時日本最速のPCクラスタ
 - 1,024ノード構成
 - 理論性能: 12.5TFlops
- InfiniBandが使われた当時最大のPCクラスタ
- OSSベース: Linux/SCore

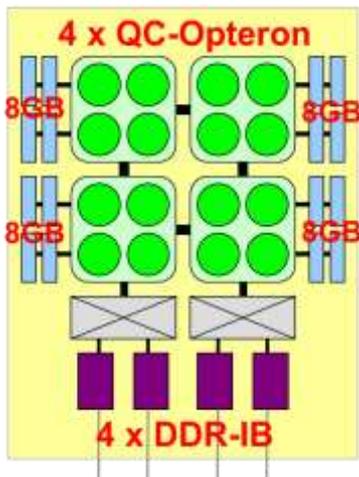


T2K Open Supercomputer/KYODAI

50.51TFlops, 35th/Top500 (2008.6)

○ Open supercomputer project

○ 筑波大、東大、京大による仕様の共通化



- ✓ 4 x Quad Core-Opteron server
- ✓ memory: 32GB
- ✓ Network: 4x DDR-IB



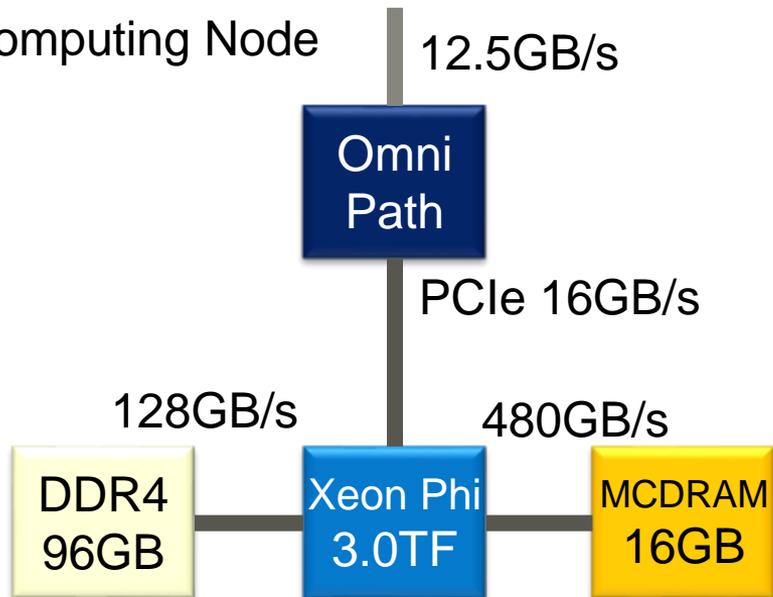
<https://pr.fujitsu.com/jp/news/2008/06/2.html>

Opteronクラスタが広がり、高いシステム性能を達成

13.55PFlops, 6th/Top500 (2016.11)

- メニーコアアーキテクチャによる本格的なスーパーコンピュータ

Computing Node



- 全体構成: 8,208ノード
- Intel Xeon Phi 68core 1.4GHz
- 16GB(MCDRAM)+96GB(DDR4)
- OmniPathによるFBB構成
- 論理ピーク性能: 25PFlops

「京」を抜き日本一性能を達成

○Top500

Rank	System	Rmax	Rpeak
1	Sunway TaihuLight	93,014.6	125,435.9
2	Tianhe-2 (MilkyWay-2)	33,862.7	54,902.4
3	Titan	17,590.0	27,112.5
4	Sequoia	17,173.2	20,132.7
5	Cori	14,014.7	27,880.7
6	Oakforest-PACS	13,554.6	24,913.5
7	K computer	10,510.0	11,280.4

○HPCG

Rank	System	HPCG
1	K computer Fujitsu	0.6027
2	Tianhe-2 NUDT	0.5801
3	Oakforest-PACS Fujitsu	0.3855
4	Sunway TaihuLight NRCPC	0.3712
5	Cori Cray	0.3554

System	N	(A) Elapsed time	(B) # of nodes	(A) x (B)	Perf/node
Cori	6,984,960	4.5 hours	9,304	41,868 hours = 4.8 years	1.53TF
Oakforest-PACS	9,938,880	13.4 hours	8,208	109,987hours = 12.6 years	1.66TF

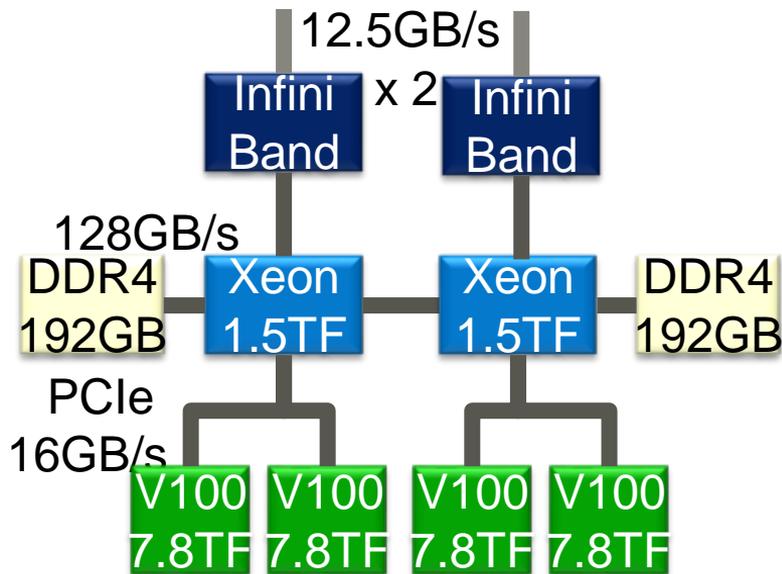
x2.6
8.5%

高い安定性による長時間動作によってHPL性能を向上

AI Bridging Cloud Infrastructure (ABCI)

19.88PFlops, 5th/Top500 (2018.6)

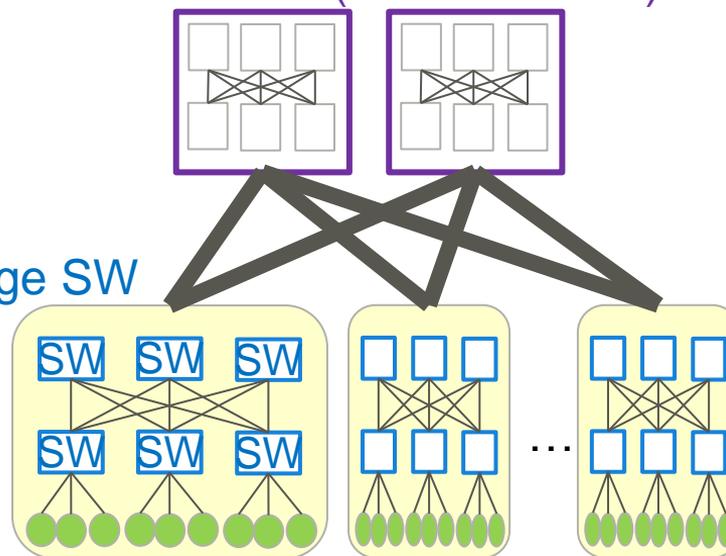
Computing Node



Fabrics

Director SW (Cluster of SW)

Edge SW



AIワークロードに特化したスーパーコンピュータ

A64FX向けのAIライブラリの開発

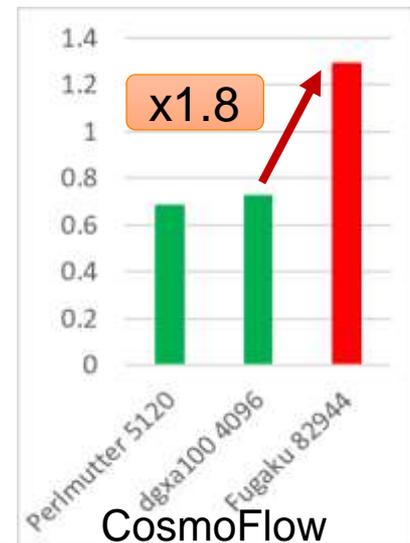
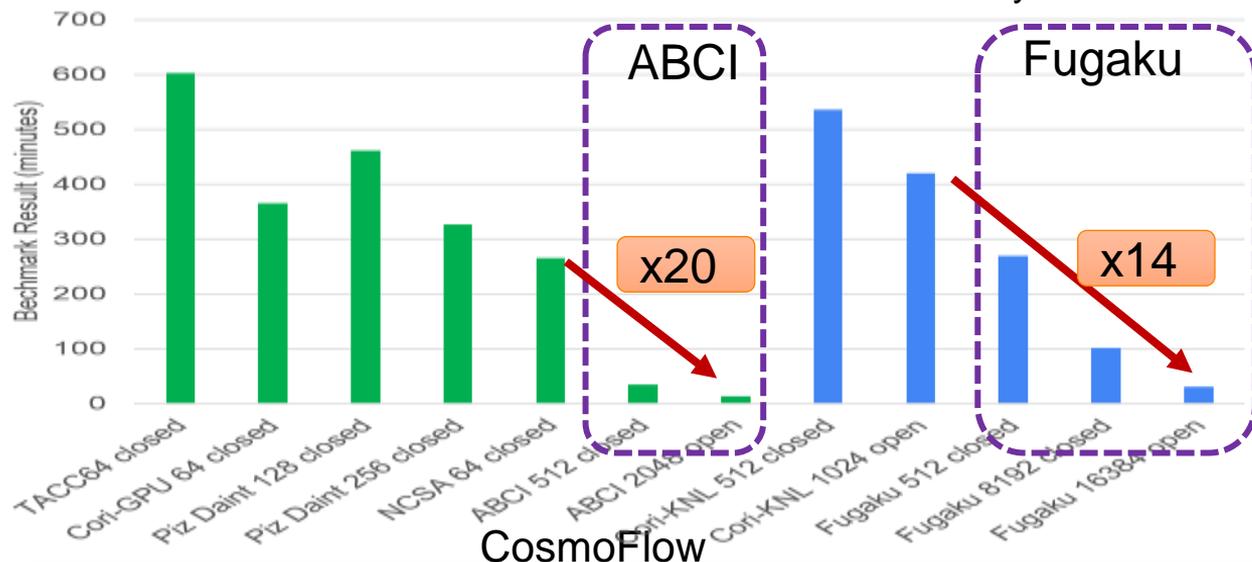
○ A64FX向けAIライブラリを開発、MLPerf HPCを計測

○ 2020年 Strong Scale

○ 2021年 Weak Scale

GPU system

CPU system



シミュレーションとAIを融合したアプリケーションでの利用を期待

未来に向けて

コンピューティングが社会へ浸透する時代に

DXではAI利用により
計算量が爆発的に増大

AIの計算需要
30万倍

「京」コンピュータ全系で



- CPUのメニーコア化、アクセラレータの活用と進化が進む
- 一方でシステム性能バランスは演算性能に重心を置くように

	RSCC(2004)	T2K Kyoto(2008)	Oakforest-PACS(2016)	ABCI(2018)
# of core	2	16	68	- (GPU)
Flops	12.2GFlops	147.2GFlops	3046.4GFlops	32.1TFlops
Memory BW	4.3GB/s (B/F: 0.35)	42.6GB/s (B/F 0.30)	480GB/s(*) (B/F 0.16)	3600GB/s (B/F 0.11)
Network BW	1.0GB/s (B/F: 0.08)	8.0GB/s (B/F: 0.05)	12.5GB/s (B/F: 0.004)	25.0GB/s (B/F: 0.001)

(*) Xeon Phiのメモリバンド幅は実測値。

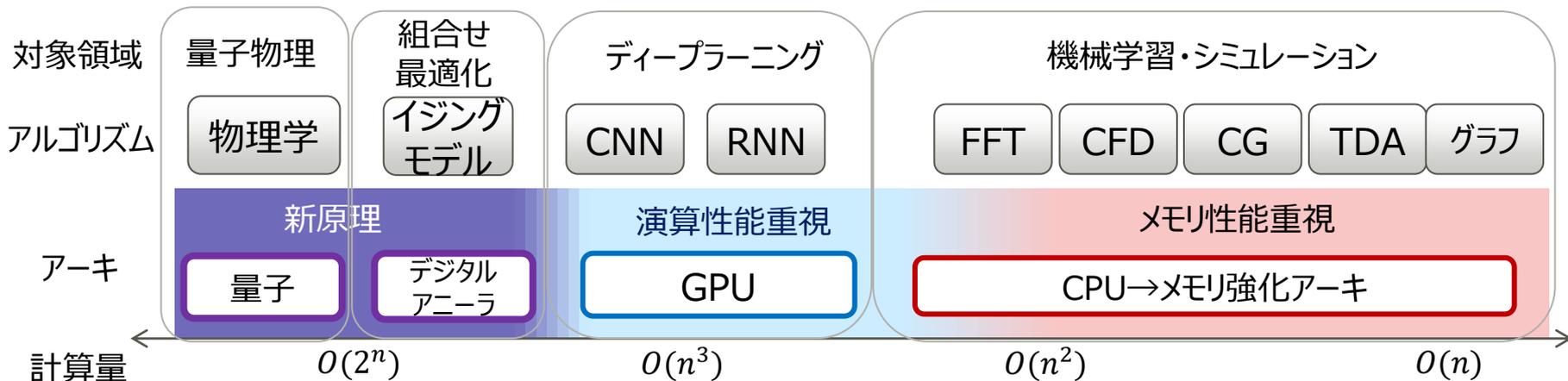
性能は向上したが、使いやすさの点では課題が累積

2030年のコンピューティング像

- 「コンピューティング」が新しい価値を生む源泉として社会に広がる
- コンピューティングの進化：並列化と専用化（アクセラレータ）
- 社会への浸透：ありとあらゆるところでのコンピューティング利用

コンピューティングの進化の方向性

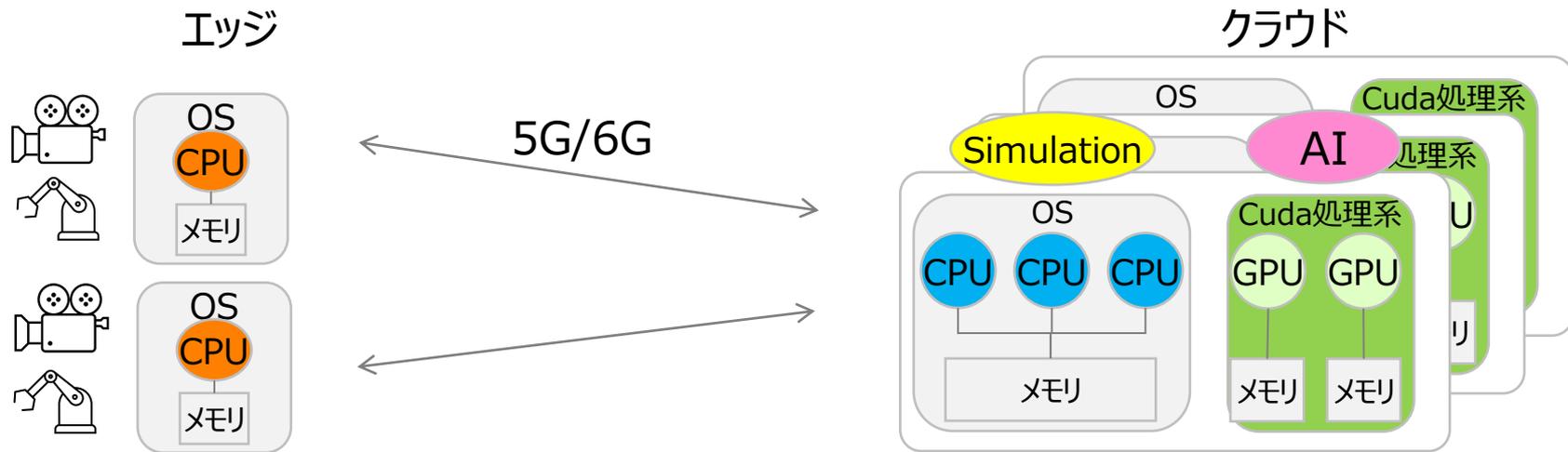
- ディープラーニング：演算中心でGPU活用が重要
- 機械学習・シミュレーション：多くのワークロードはメモリ性能が重要
- 新原理によるコンピューティングも今後広がる可能性大



ワークロードの多様性が拡大、様々なコンピューティング技術を組み合わせた最適化が必須に

社会への浸透とリアルタイム・エッジコンピューティング

- リアル世界との接点であるため、リアルタイム性が求められる
- AIやシミュレーションを使った認識・制御・判断で高度化、計算負荷は高まる



ヘビーな計算負荷をスケジューリングしつつ
システム全体での協調動作によってリアルタイム性を保証する制御系が必要

- これまでの20年
 - PCクラスタの発展がHPC全体を牽引
 - システム全体での使いやすさは大きな課題
- 今後に向けて
 - 「コンピューティング」が新しい価値を生む源泉として社会に広がる
 - 並列化と専用化: ヘテロ環境の使いこなしとユーザへの性能提供が課題

**ユーザにとって使いやすい、かつ、高性能を提供できる
ソフトウェアスタック設計と普及によって、社会全体に新しい価値を提供**

Thank you

