



東京工業大学 学術国際情報センターの紹介 TSUBAMEスパコンシリーズを中心に

東京工業大学
遠藤 敏夫

2024年6月27日（木）

PCクラスタワークショップ in すずかけ台2024

東京工業大学の情報インフラ一般を担当

- 情報支援部門：情報技術による教育研究支援
 - 認証・ネットワーク分野
 - 情報基盤活用分野
 - 情報セキュリティ分野；東工大CERTと連携
- 先端研究部門：高性能計算等の先端的研究
 - 高性能計算システム分野
 - 高性能計算先端応用分野
 - 大規模データ情報処理分野

} TSUBAMEスパコンシリーズ等の
運用・設計・研究
- HPCIおよびJHPCNに参加
- 国内外の研究機関・企業と連携・共同研究

学術国際情報センターの場所

- 大岡山キャンパスが主拠点（キャンパス内移転中）
- TSUBAME関連教員の一部(2名)がすすかけ台に2022~2023に移転

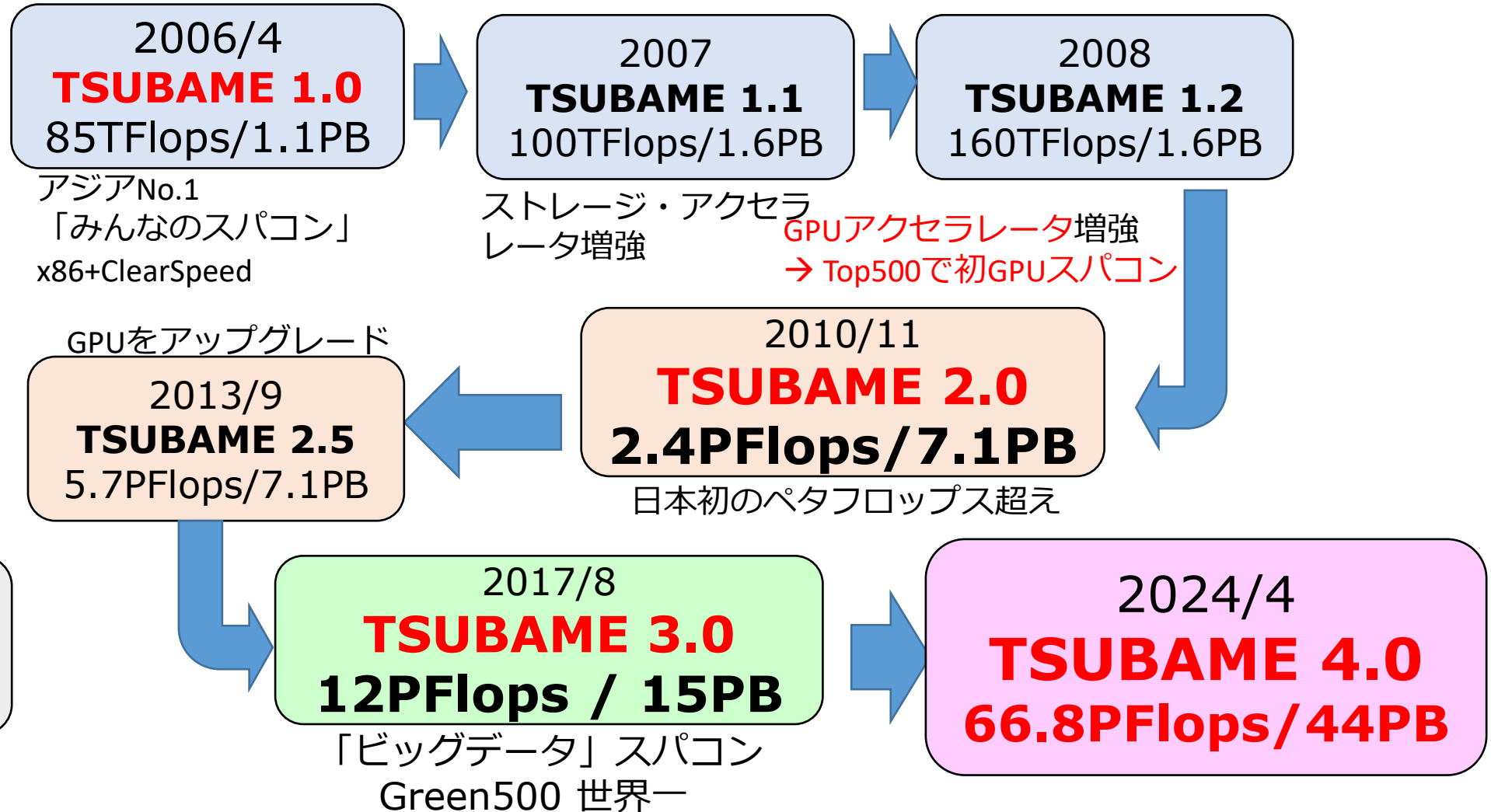
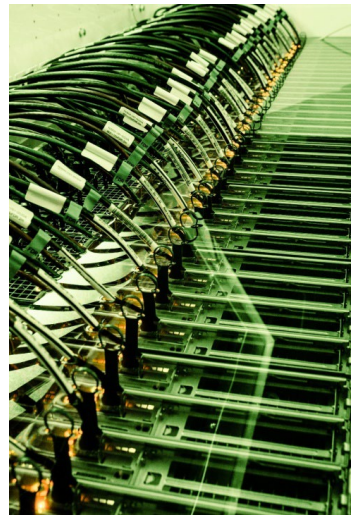


東工大TSUBAMEスパコンの概要



- 学術国際情報センターで運用するスーパーコンピュータ
- ➔ 現在はTSUBAME4.0 (2024年4月～)
- 「**みんなのスパコン**」をキャッチフレーズに、コモディティ技術を用いた使いやすさと、アクセラレータなどの先進的な取り組みを両立してきた
 - 省エネ性・設備面でも注目される
- 多分野・学内外（学生も）・産業利用含むユーザ
 - アクティブユーザ1000～1500人

東工大TSUBAMEスパコンの歴史



TSUBAME1.0~1.2スーパーコンピュータ (2006~)

Voltaire ISR9288
Infiniband 10Gbps

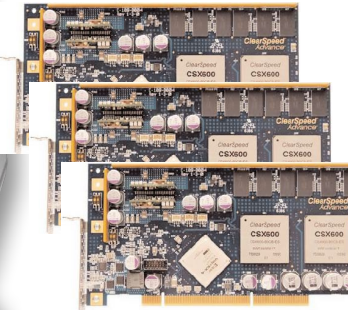


SunFire X4600
16 Opteron
cores
x 655nodes

Integrated
by NEC



SunBlade X6250
8 Xeon cores x 90nodes
(tsubasa cluster)



ClearSpeed X620
x ~~360~~ PCI-X boards
648



NVIDIA Tesla
S1070
4GPUs x 170Units

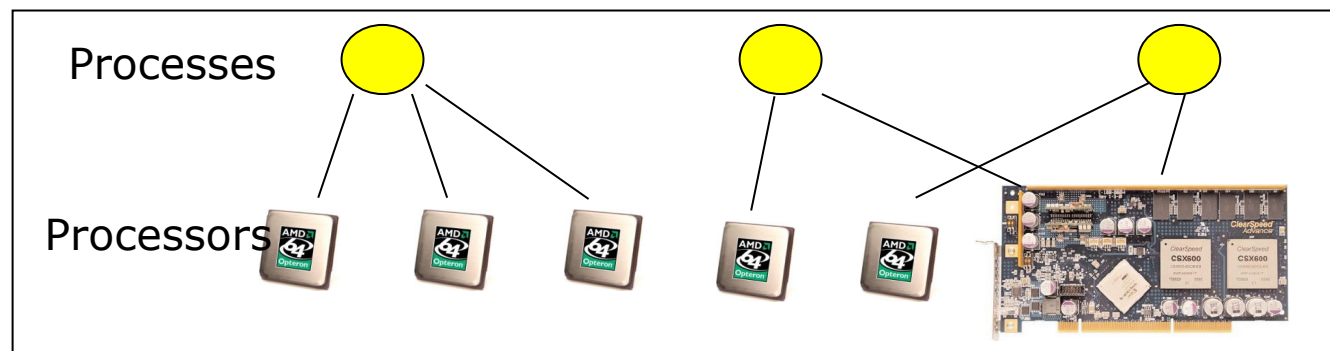
世界初のGPUスパコンとして
Top500ランキングに
(2008, TSUBAME1.2)

T1の時代：多数アクセラレータの利用の模索

- ClearSpeed BLAS, CUDAなどはあった
- しかし多数のアクセラレータを連携させるには？
 - Linpackベンチマークを題材に
 - MPIと各種APIの併用
 - CS, GPU, CPUのヘテロプロセッサへの負荷分散・・・



Example of mapping of DGEMM computation



ハードウェア増強・ソフトウェア改良によりLinpack7期連続性能向上!

	Jun06	Nov06	Jun07	Nov07	Jun08	Nov08	Jun09	Nov09
LinpackSpeed	38.18 (TF)	47.38	48.88	56.43	67.70	77.48	87.01	
Rank	7	9	14	16	24	29	41	56

TSUBAME2.0~2.5スーパーコンピュータ (2010~)

TSUBAME2.0: A GPU-centric Green 2.4 Petaflops Supercomputer

Tsubame 2.0: "Tiny" footprint, very power efficient

- Floorspace less than 200m² (2,100 ft²)
- Top-class power efficient machine on the Green 500

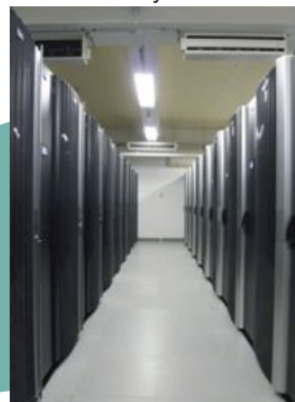
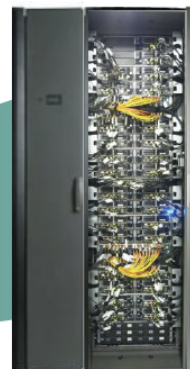
System
(42 Racks)
1408 GPU Compute Nodes,
34 Nehalem "Fat Memory" Nodes

Rack
(8 Node Chassis)

Node Chassis
(4 Compute Nodes)

Compute Node
(2 CPUs, 3 GPUs)

Chip
(CPU, GPU)



CPU(Westmere EP)
76.8 GFLOPS

GPUs(Tesla M2050)
515 GFLOPS
3 GB

1.6 TFLOPS
55 GB/103 GB

6.7 TFLOPS
220 GB/412 GB

53.6 TFLOPS
1.7 TB/3.2 TB

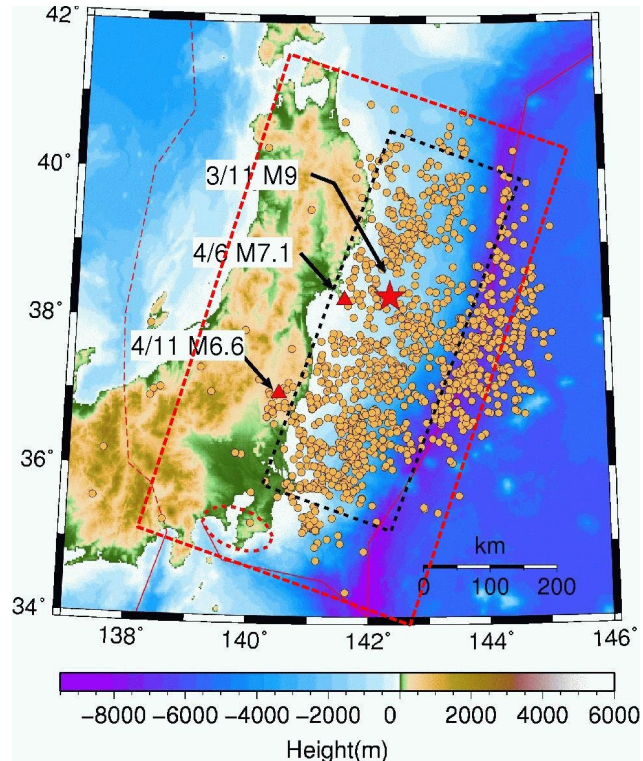
2.4 PFLOPS
80 TB

Integrated by NEC Corporation

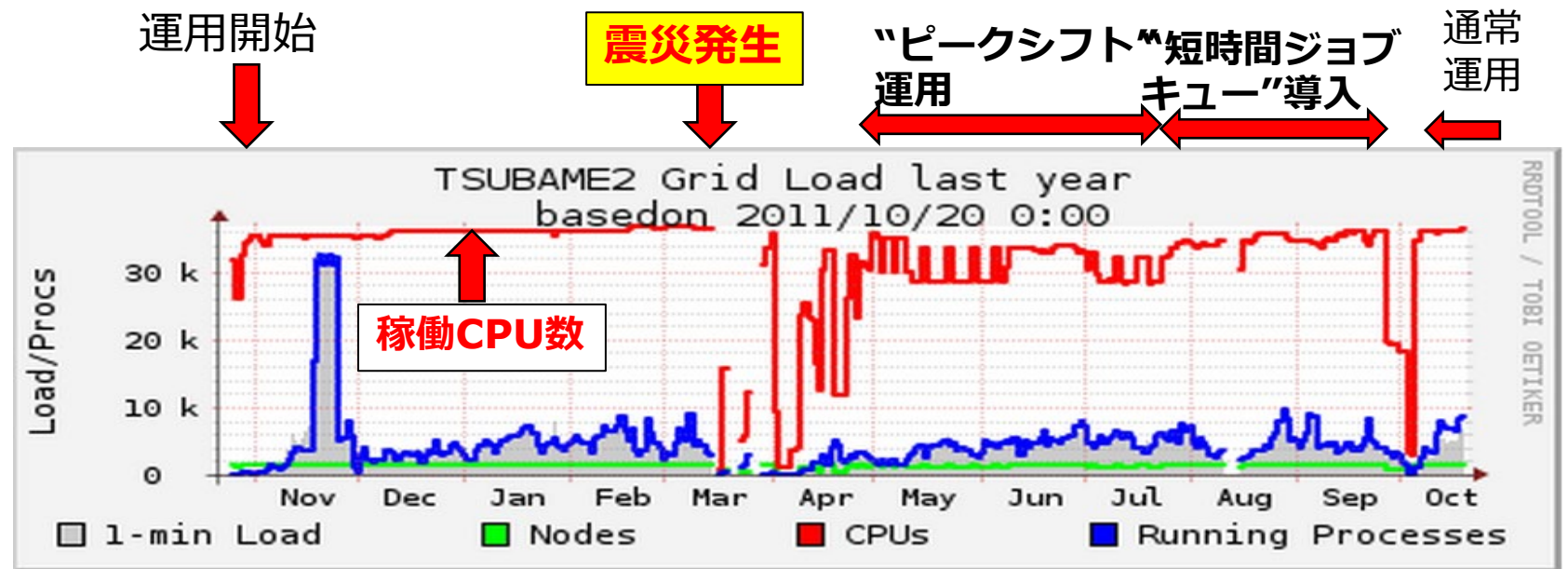
- NVIDIA Tesla M2050 4224枚
(のち、K20Xへアップグレード)
- 国内初のペタフロップス超え
- Top500 4位
- ペタスケールシミュレーション
など国際的成果多数

T2の時代：電力危機への対応にせまられる

2011.3.11 東日本大震災



- 東電原発事故による関東地方の電力不足のため、運用を大きく変更
 - 三月の計画停電
 - 政府・大学による節電要請
- TSUBAME2はキャンパスの約10%の電力 (1~1.7MW)を消費



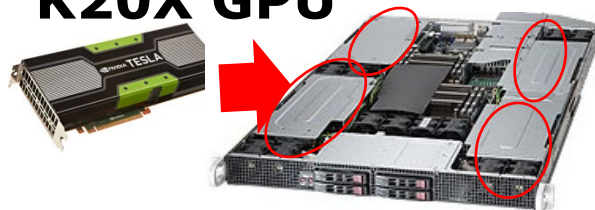
ピークシフト運用では、朝夕の電源ON/OFFを自動化

TSUBAME-KFC: ウルトラグリーン・スパコンテストベッド (2013)

液浸冷却＋大気冷却＋高密度スパコン技術
を統合した、コンテナ型研究設備

4GPU搭載計算サーバ群

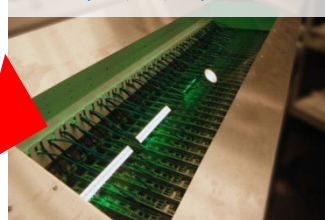
K20X GPU



NEC LX104Re-1G改×40台

液浸サーバラック

熱はプロセッサチップから油へ



熱交換器

熱は油から水へ



蒸散熱
自然大気中へ

合計理論性能

217TFlops (倍精度)

645TFlops (単精度)

その後K80 GPUへ置き換え、性能倍増



コンテナ型研究設備
20フィートコンテナ(16m²)

冷却塔:
熱は水から
自然大気へ

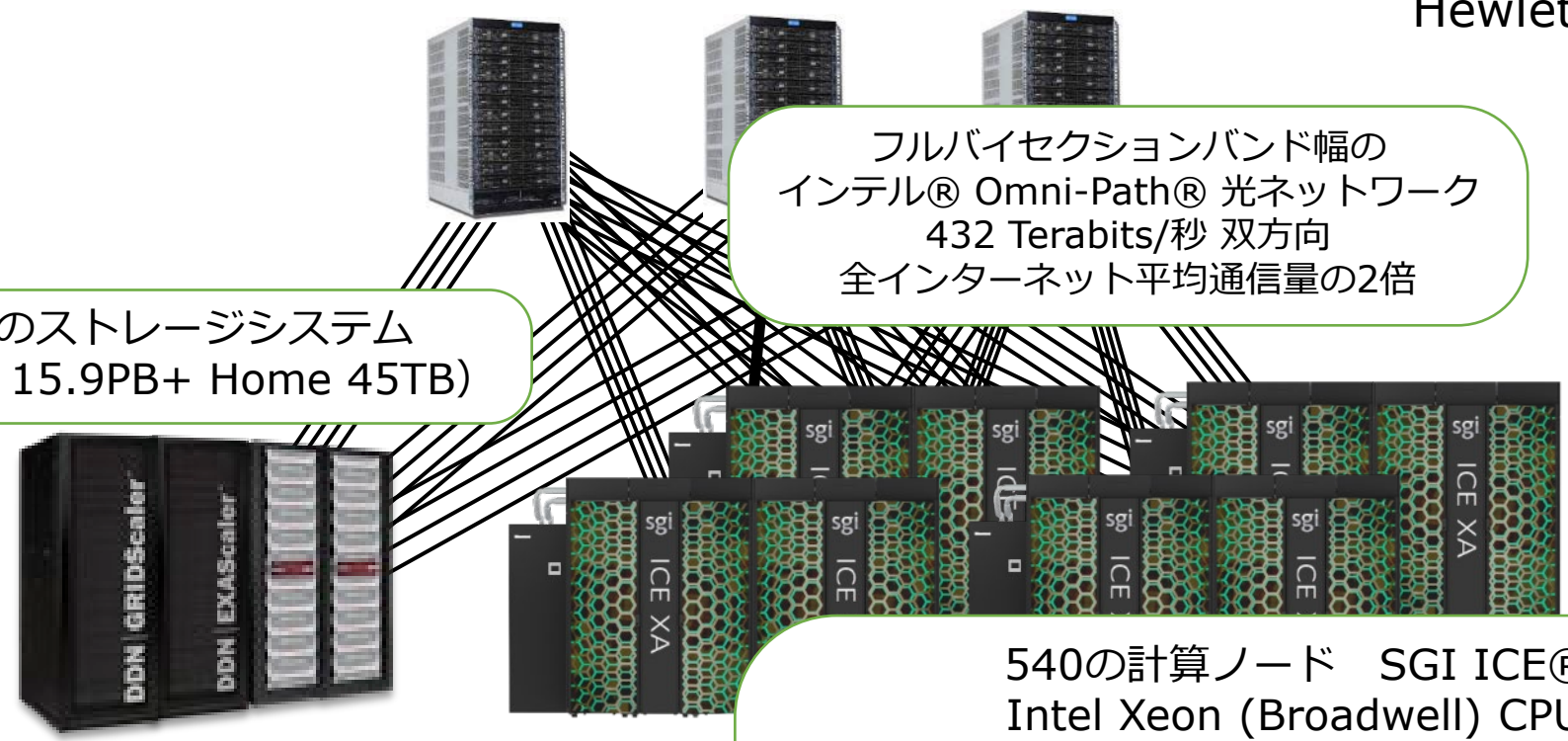


2019年5月に撤去
駐車場スペース返却

- 世界トップクラスの電力性能比, 4.5GFlops/Watt
→ 省エネランキング Green500 世界一
- 実測PUE 1.09

TSUBAME3.0スーパーコンピュータ (2017~)

Integrated by
Hewlett Packard (HPE)



DDNのストレージシステム
(並列FS 15.9PB+ Home 45TB)

フルバイセクションバンド幅の
インテル® Omni-Path® 光ネットワーク
432 Terabits/秒 双方向
全インターネット平均通信量の2倍

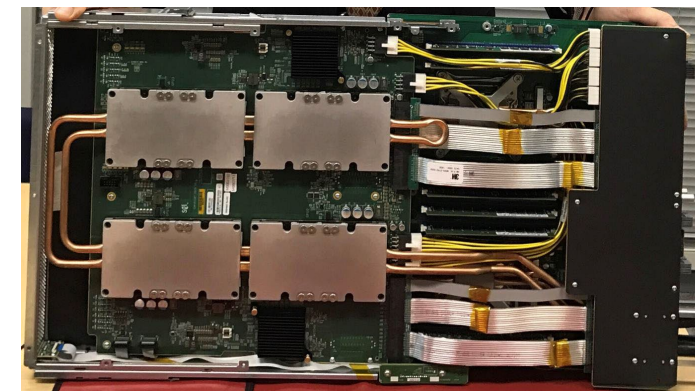
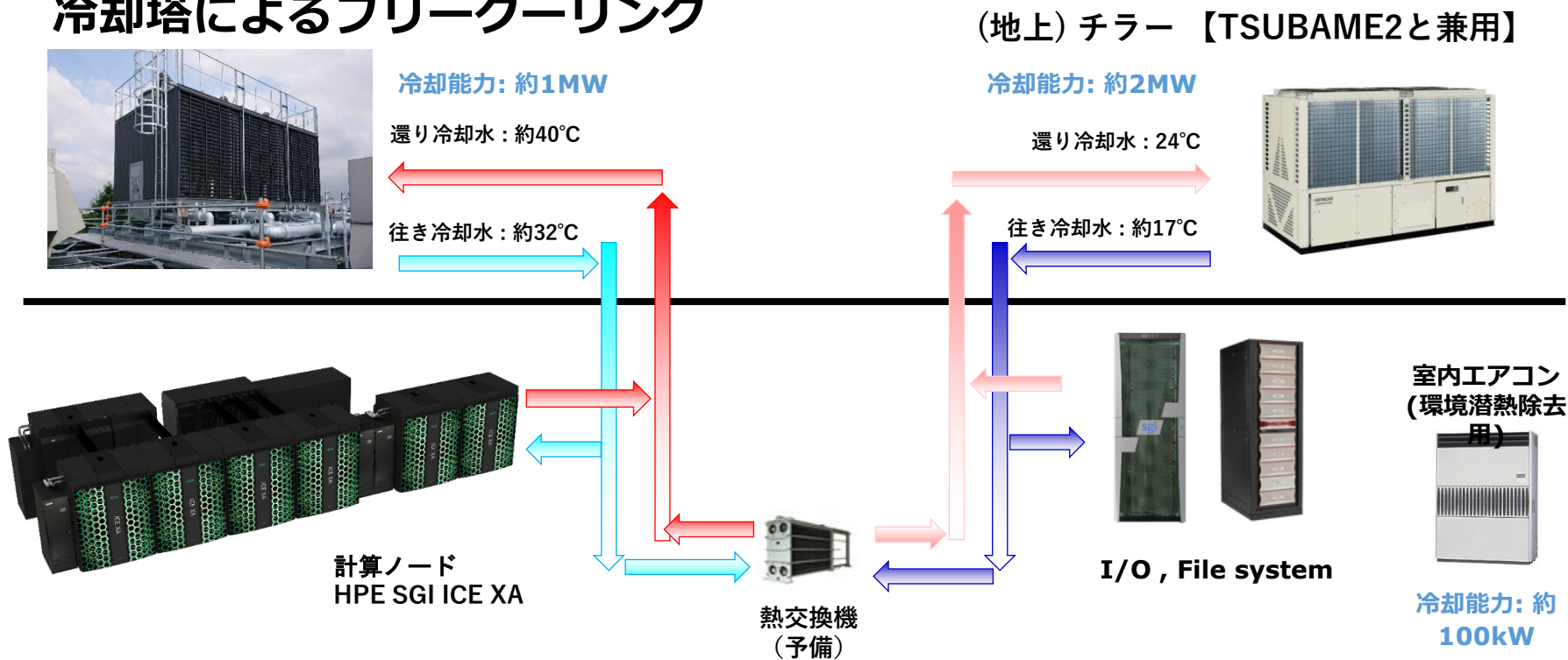
- ユーザには学内・学外研究者・産業利用を含み、アクティブユーザ数1,400
- 東工大では200近くの研究室が利用
- 深層学習ユーザが大幅増加

540の計算ノード SGI ICE® XA
Intel Xeon (Broadwell) CPU×2
+ NVIDIA P100 GPU×4
256GBメモリ、2TBのNVMe対応インテル®SSD
12PFlops (倍精度), 47PFlops (FP16)

TSUBAME3の省エネ冷却技術

- TSUBAME-KFCの経験を経て
 - 液体冷却方式と、その液温の重要性
 - しかし、液浸方式は大変

冷却塔によるフリークーリング



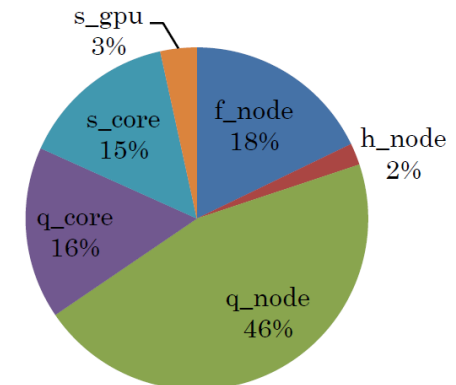
専用設計のノード:
直接液体冷却の
ための水パイプ有

Green500世界一 (プロトタイプシステムの計測時)

※OpenOnDemandがメジャーになる少し前だった



-
- The diagram illustrates a multi-processor system architecture. It features three main columns of components:
- Left Column (Green Boxes):** Four GPU units stacked vertically.
 - Middle Column (Blue Boxes):** Two CPU units stacked vertically. The top CPU is labeled "14core" and the bottom CPU is labeled "7core".
 - Right Column (Blue Boxes):** Four OPA/HFI units stacked vertically. The top two are labeled "H" and "Q", and the bottom two are labeled "G1" and "C4".
- Horizontal dashed lines connect the components across the columns, indicating data flow or communication paths. Specifically, there are connections between the GPUs and CPUs, and between the CPUs and OPA/HFI units.



拡張した形でTSUBAME4.0へ引き継ぎ

TSUBAME4.0

2024/4/1すずかけ台で稼働開始



計算ノード：

HPE Cray XD665 ×**240**台

共有ストレージ：

HPE Cray ClusterStor E1000

総演算性能：

- **66.8 PFlops** (倍精度)
- **952 PFlops** (半精度)

総容量：

- 44 PByte (ハードディスクベース)
- 327 TByte (SSDベース)

30ラックに格納

- 計算ノード 23ラック
 - ストレージ・管理 7ラック
- 東工大すずかけ台キャンパス
G4-A棟に設置

構築業者：

Hewlett Packard Enterprise



G4-A棟



マシン室
210m²

フリーアクセス
フロア

電気室

2024/10、東工大＋医科歯科大 → 東京科学大学へ
学術国際情報センターも改組予定
これからも、先進的な情報インフラと高性能計算技術の
提供により、国内外の研究・教育を前進

Advancing science and human wellbeing



Institute of
SCIENCE TOKYO

Science Tokyo will be established on October 1, 2024,
following the merger between Tokyo Tech and TMDU.

Science Tokyo introductory website
<https://www.isct.ac.jp>

