

# 次世代のスパコン設備にむけて (主に冷却・電力計測)

野村 哲弘

東京科学大学 情報基盤センター



# スパコンにとっての施設・設備

- スパコンを設置するための「器」
  - 床面積・高さ
  - 耐荷重
  - 電力供給 (電圧・容量)
  - 冷却供給 (空冷/液冷・冷媒温度・流量・熱容量)
  - 配線・配管スペース (フリーアクセスフロア・天吊り等)・ネットワークへの接続
  - モニタリング・運転最適化の余地
    - 搬出入経路等その他の事項、細かい工夫…
- 上記はシステムへの制約条件、もしくはシステムからの要求
- いくつかの選択肢からCapEx(導入コスト)とOpEx(運用コスト)を鑑みながら最適なものを選ぶ
  - とはいえ、外部制約が多いため、たいていの項目は選べるほど自由度はない
  - 「カーボンニュートラル」と聞くとものすごく高尚なものに聞こえるが、少なくとも消費エネルギー削減は電気代の話、すなわちOpExの話として捉えればよい

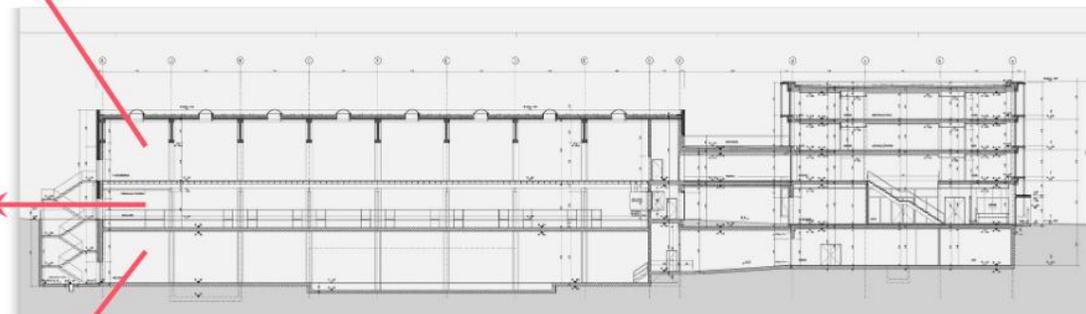
# 将来を見据えた設備を作るのは難しい(世界共通見解)

- スイス国立スパコンセンターの施設系発表(2019)より

**Flexible infrastructure,  
since we can only predict that technology will change**



Flexible Infrastructure:  
• power/cooling: 12 MW,  
• upgradable to 25 MW  
Free cooling with water from Lugano lake  
Current Power Usage Effectiveness (PUE) 1.2



- 施設はフレキシブルに、将来何かが変わることはわかってもどう変わるかはわからない
- 実際、2025/3に訪問した際にこのスライドの更新版が使われていたが、デマンド増加に伴い一部数字が書き換わっていた(しかもそのスライドも「古いから数字は信用するな」とのこと)
- 施設の寿命はスパコンの寿命より長い、使い続ける必要

# 冷却技術のトレンド

- パッケージエアコンなどの空冷はもはやメインの冷却手段になりえない
  - 計算機の熱密度上昇に伴い、熱を取り切れない
- 何らかの形での液冷(incl. 水冷)を行う
  - 直接水冷
    - 水冷ヘッド等で熱源(プロセッサ等)のそばまで冷却水を配管
  - 間接水冷
    - 計算機自体は空冷とするが、その空気は水を使って局所的に冷やす(リアドア冷却等)
    - 施設の観点(対システムへのインタフェース)では、直接水冷と冷却水を供給することに変わりはない
- 液浸冷却
  - 計算機自体を絶縁性の液体に浸して冷却する

# 液浸冷却っていけるの？

- ファンを除去、空冷もほぼ不要
  - PUE 1.04 など
  - システム全体では、対応が難しい機器は残る (例: ストレージ機器・ネットワーク機器)
- TSUBAME-KFC(オイル)、PEZY-SC系(フロリナート)
  - 従来の冷媒ではビニール・ゴムなどの部品が劣化、高価、、 → 新しい冷媒の開発？
    - ENEOS 冷却用オイル  
[https://www.eneos.co.jp/information/upload\\_pdf/20221227\\_01\\_01\\_1170836.pdf](https://www.eneos.co.jp/information/upload_pdf/20221227_01_01_1170836.pdf)
    - PFAS・PFOA規制の影響でフロリナート系が全滅
- 冷媒によって冷却方法はさまざま
  - 二相式 (液化⇔気化)
    - 冷媒の漏出・喪失防止のための機構が必要 (環境・経済性)
  - 一相式 (液体のまま)
    - メンテナンス時・メンテナンス後部材の冷媒汚染
- シャーシ単体を覆うラックマウントタイプのものも開発
- Intelがリファレンスデザイン
  - <https://pc.watch.impress.co.jp/docs/news/1411121.html>

冷媒自体の  
環境性能も考慮が必要

最新プロセッサの消費電力(TDP)は急上昇  
→液浸冷却の急速な普及？

NVIDIA A100→H100:  
400W→700W

AMD EPYC Milan→Genoa:  
280W→360W

Intel IceLake→Sapphire Rapids:  
270W→350W

メンテナンス性・実装密度(上に積めず、横に広がる)を考慮すると時期尚早  
→ 通常の水冷ブロック式冷却等が現実解

# では、現実的な冷却法は？

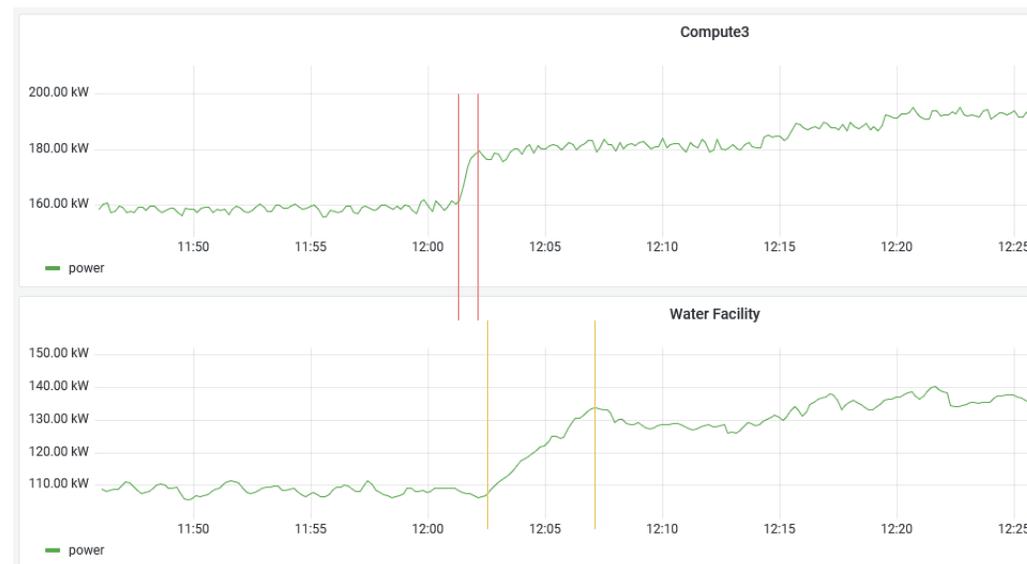
|    | 技術         | 設置条件      | 導入費用 | 管理費用      | 消費電力        | 実績 | 冷却水温度         | その他   |
|----|------------|-----------|------|-----------|-------------|----|---------------|---|
| 水冷 | チラー        | ○         | ▲    | ×<br>(保守) | ×           | ◎  | 7度~25度?       | 水温変化の追従性が低い<br>冷却過多による故障防止に注意が必要                                      |
|    | チラー+冷却塔    | ×<br>(空間) | ○    | ▲         | ○           | ○  |               | 制御が複雑化するため、大規模システムにおいては運転要員の待機が必要                                     |
|    | チルドタワー     | ▲<br>(重量) | ×    | ×<br>(保守) | ▲           | ×  |               | チラー + 冷却塔の一体型<br>設備費用が高い  |
|    | 冷却塔のみ      | ○         | ◎    | ○         | ◎<br>但し水が必要 | ▲  | 夏季32度<br>他24度 | TSUBAME3, ABCIで実証済み<br>(夏場猛暑において無停止)<br>湿球温度に大きく影響 DC建屋のコストは低い(温湿度維持) |
| 空冷 | PAC        | ▲<br>(空間) | ○    | ○         | ×           | ○  |               | 既存技術  |
|    | 水冷式        | ○         | ▲    | ○         | ▲           | ○  |               | 中規模消費電力に対応可能  |
|    | 外気導入       | ×         | ○    | ×         | ○           | ×  |               | 温湿度の維持が難しい<br>塩害防止フィルターが高価  |
|    | 間接外気冷房型PAC | ○         | ▲    | ○         | ▲           | ▲  |               |   |

# 高温冷却水(32°C)による水冷

- 冷却塔のみによるフリークーリングの実現の利点
  - 冷却手法を1種に絞ることでコスト低減・空間制約充足
  - チラーにおける圧縮機の削減による冷却エネルギー削減
  - 運転の簡略化(冷却過多の心配がなく、冷却可能温度域が広い)
- よく言われている欠点・懸念点
  - 運転温度上昇によるリーク電力増加
    - FinFETの採用により、リーク電力は減少している
  - 運転温度上昇による故障率の増加
    - 定量比較は難しいが、冷却CapEx, OpExの低下を上回る影響か??
  - 異常気象時の冷却能力不足
    - チラーほかでも事情は変わらない
  - 計算ノード以外も32°Cで冷やせる?(ストレージ・ネットワーク機器・管理ノード類等)
- 導入事例・製品化事例
  - TSUBAME3.0 (東京工業大学 (現: Science Tokyo))
  - ABCIシリーズ(産業技術総合研究所)
  - ASPIRE 2A (NSCC Singapore)
  - HPE Cray-EX シリーズ
  - Lenovo Neptune シリーズ
  - Supermicro Rack-Scale Liquid Cooling Solutions



- 消費電力やジョブのデータから将来の冷却需要を先読みして、冷却装置の運転等に反映できないか?
  - 計算機の消費電力に比べて冷却施設の消費電力は1~5分程度の遅れがある(例: TSUBAME4.0)
- 消費電力量に制約がある・制約を設ける場合、消費電力のリアルタイムモニタリングが必要に
  - 契約電力超過のペナルティは30分毎の累計値で判断→30分未満で問題解決が必要
  - 標準化されたプロトコルで電力情報を出力する機材が欲しい
    - Web UIからスクレイピングした5分遅れの30分累積値を加工して… とかだと困る
- 特にNFS(富岳・ポスト富岳)では、NIS(大学スパコン等)に比べて大きなジョブの割合が多く、ジョブ毎の消費電力の変化が急峻になりがち
  - アクセラレータの採用でこの傾向が強まるか?
- 実際に東京科学大(TSUBAME4.0)と東大/JCAHPC(Miyabi)にネットワークにつながる電力計を設置してみた
  - 右のグラフも設置した電力計のデータ



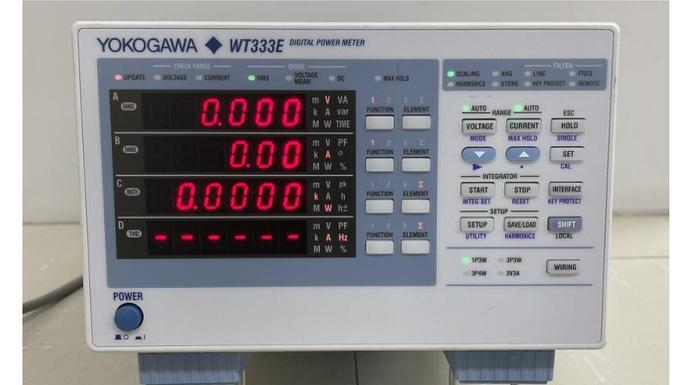
[TSUBAME4 消費電力情報](#)

2025/6/23 12:00付近

(グランドチャレンジ大規模計算の実行開始タイミング) 7

# 実際に設置した機器

- 東京大学(JCAHPC Miyabi)
  - 横河電機デジタルパワーメーター WT333E-C1-D/C7 9台
  - Miyabiの各分電盤計9か所に設置
- 東京科学大学(TSUBAME4.0)
  - シュナイダーエレクトリック PM8240 + Socomec 192T9675 4式
  - TSUBAMEの計算機向け分電盤3か所、冷却機器向け分電盤1か所に設置
- いずれの機器もMODBUS-TCPプロトコルでリアルタイム(1秒毎)の積算電力データを高精度に取得可能
  - EEHPC Power Measurement Methodology Level3 に適合
  - 電気料金算定のためのメータと同レベルの情報
- 長期データの蓄積を継続して実施、運転への反映は今後の課題
  - NISマシンはジョブが細かく、イベント時以外の消費電力変動は均されてしまいがち



# ここまでのまとめ

- 施設・設備面での検討事項の紹介
  - 冷却手法に関する検討
  - 消費電力モニタリング手法(機器)の紹介
- 
- カーボンニュートラル実現に向けた各種手法の検討結果
  - ポスト富岳における実際の建屋の検討  
は、次の講演で