

温水冷却の効率に関する定量的評価の試み

庄司 文由、野中 丈士

理化学研究所計算科学研究センター

埴 敏博

東京大学情報基盤センター/最先端共同 HPC 基盤施設

1. はじめに

近年、温水冷却技術は、HPC システムの冷却にかかるエネルギー効率を改善するために、多くの HPC センターおよびデータセンターで採用されている。温水冷却においては、CPU 等の冷水温度を高く設定することで、外気による自然な冷却が促されるため、冷凍機等を駆動するための電力が削減できる[1][2]。その一方で、CPU 等のシリコンから構成される半導体は、高温で動作させればさせるほど、漏れ電流の増加により、消費電力が増加することが知られている。また、近年の CPU は、駆動温度が一定の水準を超えると、上記の消費電力の増加や故障率の上昇を抑制するために自動的にクロック周波数や電圧を下げる機構 (Dynamic Voltage and Frequency Scaling, DVFS) を備えている[3][4][5]。

温水冷却の効果を正しく評価するためには、冷却のための電力を削減できるというポジティブな点に加え、上記のような消費電力の増加や、クロック周波数の低下による演算性能の悪化のようなネガティブな点も等しく考慮に入れる必要がある。演算性能が低下すれば、その分より長く計算することにより、HPC システムが消費するエネルギーの総量が増加する可能性があるためである。特にクロック周波数の低下による演算性能への影響は、大規模なジョブでより深刻になると予想される。クロック周波数は CPU 毎に独立にコントロールされるが、CPU (プロセス) 間で同期を取る場合、最も遅い CPU (プロセス) に律速されるからである。

以上を踏まえ、本研究では、単体および複数の CPU を使う様々なジョブを、異なる冷水温度で実行し、冷却設備および HPC システムの消費電力の増加と演算性能の低下を定量的に分析する。左記の分析に基づき、効率的な施設運用の実現に向けた運用手順の確立を目指す。

本稿では、今回の大規模 HPC チャレンジによりこれまでに得られた成果の概要を紹介する。なお、本研究の成果の一部は、ISC2020 Research Poster[6]および IEEE Cluster2020 paper[7]として採録されているので、詳細についてはそれらを参照いただきたい。

2. 実験方法

冷水温度の変更には施設側のオペレーションが必要で、温度が安定するまでに時間がかかるため、今回の実験では、それぞれの冷水温度において、同じベンチマークプログラムのセットを実行する方法を採用した。ベンチマークプログラムとしては、演算インテンシブなものとして LINPACK[8]、メモリアクセスインテンシブなものとして HPCG[9]、ウィークスケーリングを評価するものとして GeofEM をそれぞれ用いた。LINPACK (シングルノード及びマルチノード版) と HPCG は Intel Parallel Studio XE (Ver. 2019.5.075) に配布されているものを用いた。

最初に計算ノード毎の演算性能や消費電力等の特性のばらつきを評価するためにシングルノ

ードの LINPACK を実行した。その後、それぞれのベンチマークプログラムについて、複数のサイズのマルチノードジョブを実行した。それぞれのケースで、演算性能、クロック周波数、消費電力等を測定した。用いた冷水温度は、9℃、12℃、18℃である。ちなみに、25℃のケースも試みたが、大多数ノードでの LINPACK 実行中に冷却システムが追いつかなくなる事象に遭遇したため測定を断念した。通常運用時の設定温度は 12℃とのことである。測定した冷水温度、ノードサイズ、測定回数を表 1 に示す。

	ノードサイズ	9℃	12℃	18℃
LINPACK	1	5120	4096	4097
MP-LINPACK/HPCG/GeoFEM	128	5/6/6	7/7/7	5/6/6
	256	5/5/5	6/6/6	5/5/5
	512	5/5/5	6/6/6	5/5/5
	1024	5/5/5	6/6/6	5/5/5
	2048	5/5/5	6/6/6	5/5/5
	4096	5/5/5	6/6/6	5/5/7

表 1 ベンチマーク毎の各冷水温度におけるノードサイズと測定回数

3. 結果

最初に冷水温度の変化に対する、冷却設備の効率の変化を見る。

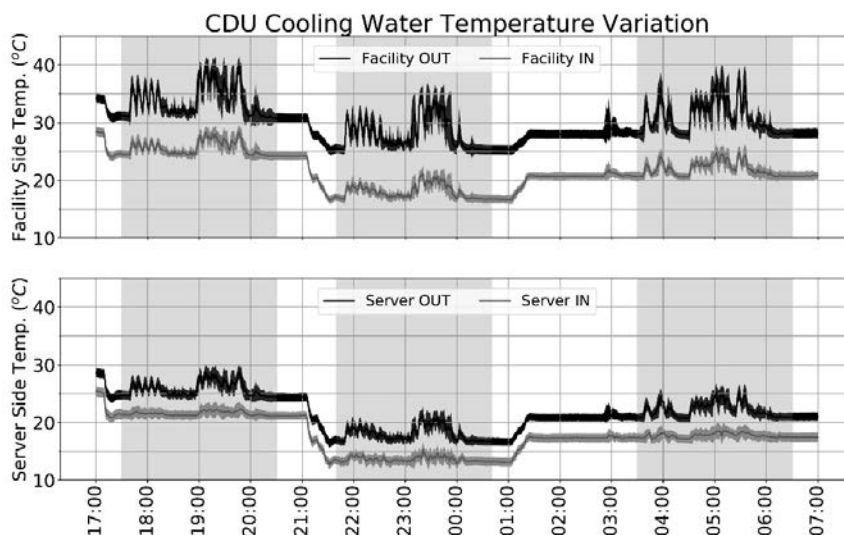


図 1 冷却設備側 (上) と HPC システム側 (下) の冷水温度の変化

図 1 の上段は冷却設備側、下段は HPC システム側の、inlet と outlet の冷水温度の推移を示している。それぞれの温度のケースについて、同じベンチマークプログラムのセットを実行したが、時間にすると概ね 3 時間であった。図 1 の濃い色の背景になっている部分はその時間帯を示している。したがって、この 3 時間において、冷却設備および HPC システムが消費したエネルギーを用いて冷却設備の効率の変化を評価する。

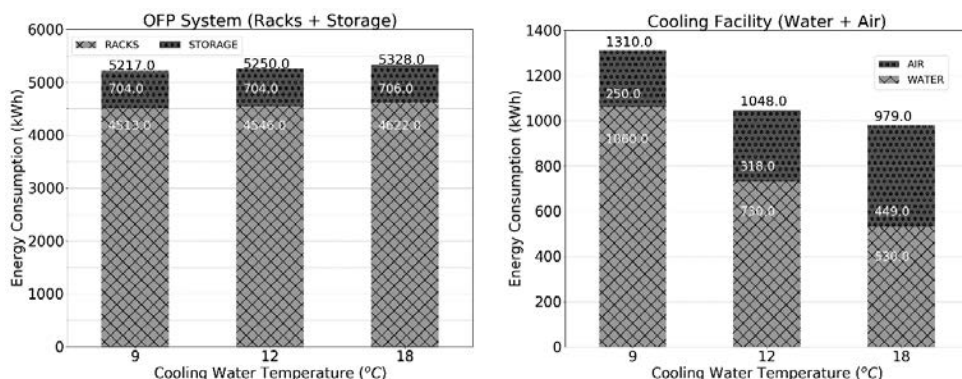


図 2 冷水温度の変化に対する HPC システム (左) および冷却システム (右) の消費エネルギーの変化

図 2 に冷水温度の変化に対する HPC システム (左) および冷却システム (右) の消費エネルギーの変化を示す。冷水温度が高くなれば、より外気によって自然に冷却されやすくなるので、冷却にかかるエネルギーは減少する。その一方で、HPC システムは冷水温度が上昇するとリーク電流が増加するため、消費電力および消費エネルギーが増加すると見込まれるが、今回の実験においては、3つの温度において、消費エネルギーはほぼ変わらなかった。これは後述する DVFS によって、消費電力および消費エネルギーの増加が抑制されたためと考えられる。

今回は、冷却システムのエネルギー効率を評価するための指標として、COP (Coefficient of Performance) と PUE (Power Usage Effectiveness) を用いた。本稿における COP は、冷却に消費したエネルギーに対し、どれだけの熱エネルギーを除去できたかを示すものであり、大きいほど効率が良いことを示す。OakForestPACS が消費したエネルギーがすべて熱に変換され、それが冷却システムですべて冷却されたと仮定し、以下の式で COP を定義する。

$$COP = \frac{OFP\ System\ (Racks)}{Cooling\ Facility\ (Water)}$$

一方、PUE は冷却システムも含めたサイト全体の消費エネルギーに対する IT システムの消費エネルギーの比である。冷却等にかかるエネルギーがゼロのような理想的なケースでは 1 となり、大きくなるほど効率が悪いことを示す。本稿では以下の式で定義する。

$$PUE = \frac{OFP\ System + Cooling\ Facility}{OFP\ System}$$

それぞれの指標について、図 2 に示した数値で計算したものが、図 3 である。

PUE および COP 共に、冷水温度が高いほど効率が良くなることを示している。特に COP については、9°C と 18°C のケースを比較すると 2 倍以上効率が改善されたことがわかる。PUE も 18°C のケースでは 1.2 を下回るスコアになっており、9°C のケースと比較して大きく改善されることがわかった。これらの成果の詳細については [6] を参照されたい。

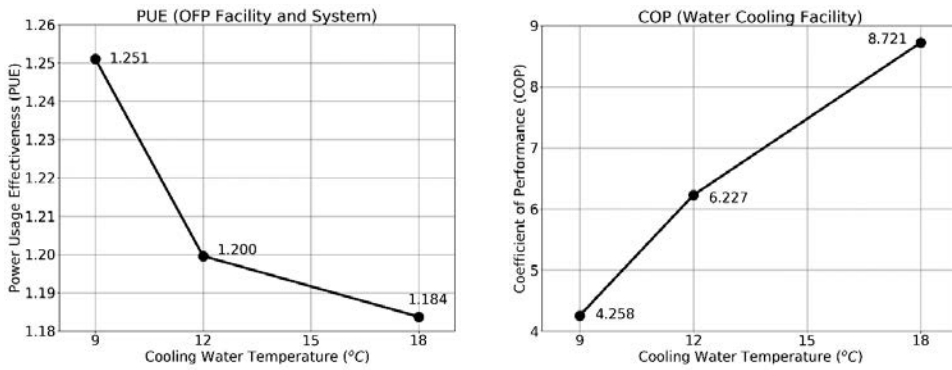


図 3 冷水温度の変化に対する PUE および COP の変化

次に、冷水温度の変化によるシングルノード性能への影響を見る。

まず 12°C のケースで LINPACK を実行した際の実行性能の分布 (図 4) と演算性能と CPU の温度の相関 (図 5) を見る。

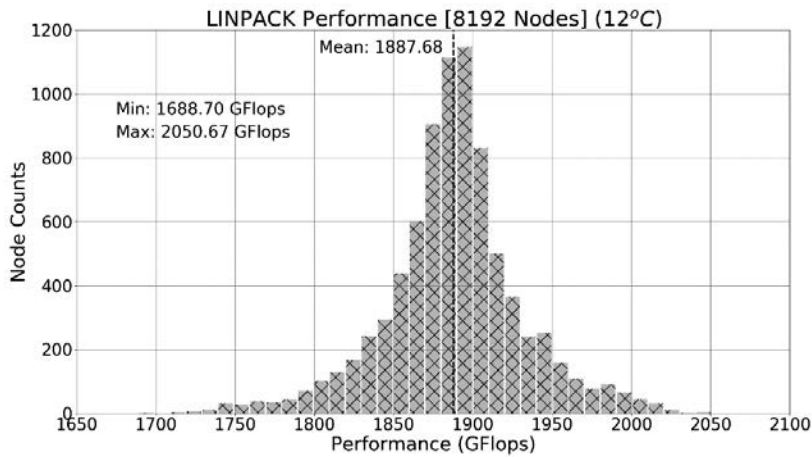


図 4 冷水温度 12°C のケースの LINPACK 性能の分布

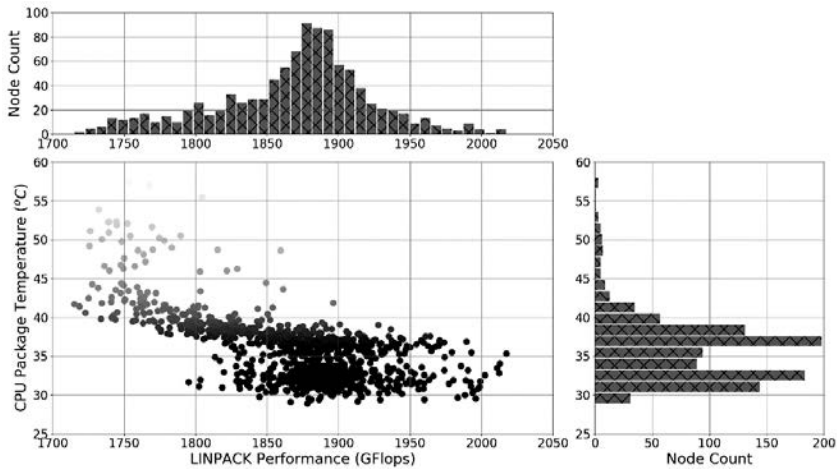


図 5 冷水温度 12°C のケースの LINPACK 性能と CPU の温度の分布

図 4 に示す LINPACK 性能の分布については、通常運用時の 12℃ のケースで既に平均値に対し 20% 近くの幅で性能のばらつきがあることがわかる。図 5 からは、CPU 毎に LINPACK 性能に加えて、温度も広い範囲で分布していることがわかる。全体としては温度が低い CPU ほど高い演算性能を出す傾向にあるが、少数ながらも 45℃ 以上で性能が低いグループも存在しており、マルチノードジョブの際にこれらのノードが含まれると全体の性能に影響が出ることが懸念される。

次に、9℃ と 18℃ のケースを LINPACK 性能とエネルギー消費の観点で比較する。

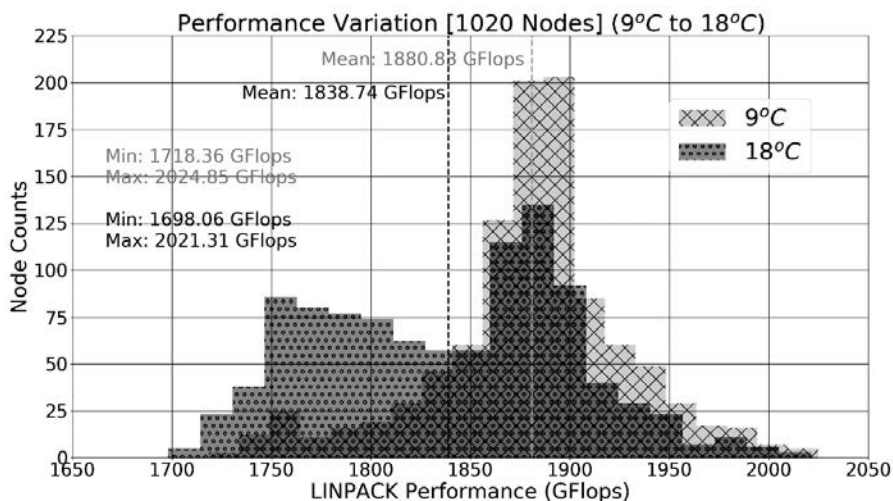


図 6 9℃ と 18℃ のケースの比較 (LINPACK 性能)

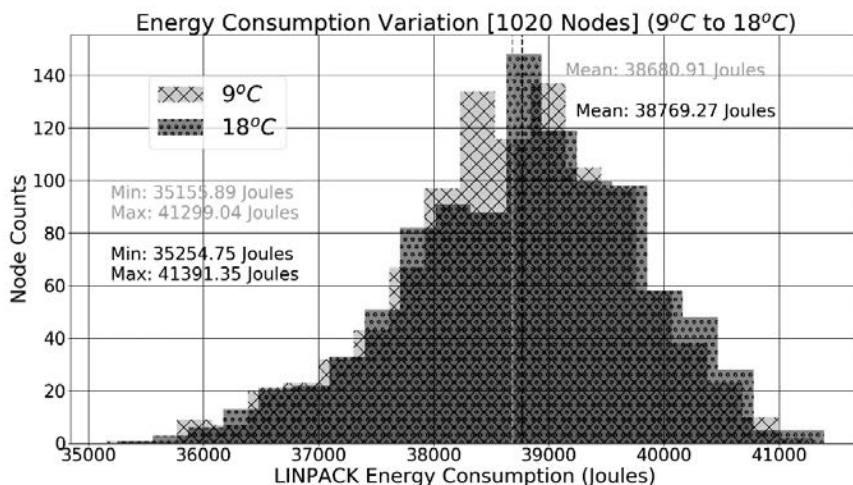


図 7 9℃ と 18℃ のケースの比較 (エネルギー消費)

図 6 は冷水温度 9℃ と 18℃ のケースを LINPACK 性能の観点で比較したグラフである。18℃ のケースでは、9℃ のケースに比べ、明らかにパフォーマンスが低い CPU が増えており、平均値および最低値共に 9℃ のケースよりも低くなっている。この結果は冷水温度に対する感受性が CPU 毎に大きく異なることを示しており、特に最低値が下がることにより、マルチノードジョブの性

能に対するインパクトはより大きくなることを示唆する。同期を取るケースなど、一番遅いノードの性能に律速されるからである。一方で、エネルギー消費の分布は9℃と18℃でほとんど変わらない。これは DVFS がうまく機能し、エネルギー消費を抑えた結果だと考えることができる。つまり、高い冷水温度では、DVFS によりクロック周波数や駆動電圧が抑制されやすくなることで、演算性能は悪化（つまり演算時間は増加）するものの、エネルギー消費の増加は抑制できるということだと考えられる。これらの成果の詳細については[7]を参照されたい。

最後に、マルチノードでの性能の評価を行う予定だったが、実行時の設定が不十分だったため、特定のノード数を用いるジョブに対し、同じノード群を用いた実行ができておらず、異なる冷水温度の間で比較できるデータが十分に得られていなかった。上述のとおり、今回の LINPACK のシングルノード性能の評価結果は、冷水温度を上げることによる性能低下の影響は、マルチノードジョブでより深刻になることを示唆しており、今後機会があればマルチノードジョブに対する詳細かつ定量的な評価を行いたい。

5. まとめ

温水冷却技術を正しく評価するためのひとつの方法として、冷水温度を変化させつつ、特定のベンチマークを実行し、その際の HPC システムの消費エネルギーおよび演算性能と冷却システムの消費エネルギーから、全体のエネルギー効率を評価する方法を提案し、OakForestPACS を用いて実証実験を行った。

冷却システムのエネルギー効率については、冷水温度が高い方がより高い効率で冷却できること、特に9℃と18℃のケースを比較すると、倍以上の効率が実現できることがわかった。また PUE についても 5%以上改善された。

シングルノードにおける LINPACK の実行により、各 CPU の演算性能と駆動温度の分布を調査した。通常運用時の冷水温度 12℃のケースでも、演算性能、駆動温度共に広く分布しており、演算性能については平均値に対し 20%近くの幅でばらつきがあることがわかった。冷水温度 9℃と18℃のケースを比較すると、演算性能が劣化する CPU が増加し、18℃のケースでは平均値、最低値共に低下することがわかった。この結果は、冷水温度を上げることによるマルチノードジョブの性能に対するインパクトはより大きくなることを示唆している。

今回の実験により、冷水温度を上げることにより、冷却システムの消費エネルギーを抑制できるというメリットがある一方で、演算性能の低下というデメリットがあることについて、定量的に評価することができた。今後、評価の範囲をマルチノードジョブにも拡大し、冷水温度の変化に対する、性能の劣化（とそれを補完するための追加のエネルギー消費）を考慮したエネルギー効率の評価手法を確立したい。それができれば、冷水温度をパラメータとして、エネルギーやターンアラウンドタイム等それぞれの運用者が着目する指標を最適化した運用が可能となる。今後その目標に向けて研究を進めていきたい。

6. 謝辞

本研究は、大規模 HPC チャレンジという制度がなければ到底実現できなかった。このような機会を与えていただいたことに深く感謝致します。また、今回の実証実験を行う上で、JCAHPC の関係者、運用主体の富士通株式会社の皆様に大変なご苦勞をおかけしました。お詫びと共に深く感

謝申し上げます。GeoFEM のコードの提供および OFP 上での実行に関するサポートについて、東京大学情報基盤センター 中島教授および Intel 株式会社 堀越氏に多大なご協力をいただきました。

参 考 文 献

- [1] S. Zimmermann, et.al., “Aquasar: A hot water cooled data center with direct energy reuse,” Energy Volume 43, Issue 1, July 2012, Pages 237-245.
- [2] H. Huber et al., “Case Study: LRZ Liquid Cooling, Energy Management, Contract Specialities,” 2012 SC Companion: High Performance Computing, Networking Storage and Analysis, Salt Lake City, UT, 2012, pp. 962-992, doi: 10.1109/SC.Companion.2012.123.
- [3] Intel, “Intel Xeon Phi processor: Your path to deeper insight.” [Online]. Available: <https://www.intel.com/content/www/us/en/processors/xeon/xeon-phi-processor-product-brief.html>
- [4] K. D. Stroup and P. Peltz, “Measuring and mitigating processor performance inconsistencies,” in Cray User Group Conference (CUG 2019), 2019. [Online]. Available: <https://cug.org/proceedings/cug2019proceedings/includes/files/pap124s2-file1.pdf>
- [5] R.Schöne, T.Ilsche, M.Bielert, A.Gocht and D.Hackenberg, “Energy efficiency features of the Intel Skylake-SP processor and their impact on performance,” arXiv preprint arXiv:1905.12468, 2019.
- [6] J. Nonaka, T. Hanawa and F. Shoji, “Analysis on the Impact of the Cooling Water Temperature on the HPC System and Facility - Case Study: Oakforest-PACS (OFP) System and Facility,” ISC2020, Frankfurt, Germany, 2020. Available: <https://2020.isc-program.com/presentation/?id=post131&sess=sess325>.
- [7] J. Nonaka, T. Hanawa and F. Shoji, “Analysis of Cooling Water Temperature Impact on Computing Performance and Energy Consumption,” 2020 IEEE International Conference on Cluster Computing (CLUSTER), Kobe, Japan, 2020, pp. 169-175, doi: 10.1109/CLUSTER49012.2020.00027.
- [8] <http://www.netlib.org/benchmark/hpl/>
- [9] <https://www.hpcg-benchmark.org/>