Vol.2018-HPC-167 No.18
2018/12/18

MHD シミュレーションコードを利用した CPU 電力キャッピング下でのスーパーコンピュータシステム ITO の消費電力特性評価

深沢圭一郎[†1]　南里豪志[†2]　本田宏明[†3]

概要：九州大学スーパーコンピュータシステム ITO では，Skylake 世代の Xeon を搭載しており，Intel RAPL を利用した CPU 電力キャッピングが可能である．そこで，実アプリケーションである MHD シミュレーションを利用し，ITO の CPU 電力にキャッピングをかけた状態での消費電力特性を評価した．Skylake 世代の Xeon では，CPU 動作周波数がダイナミックに変化し，消費電力もそれに伴い変化しているが，キャッピングをかけることで，周波数の変動はあるが，消費電力の変動は少なくなっていた．またキャッピングされた CPU 電力と比べ，DRAM の消費電力はそれほど低下しないため，B/F 値の変化による計算性能変化も見えた．本研究ではこれらの詳細な性能，計測データを示し，その結果を議論する．

キーワード：消費電力，電力キャッピング，MHD シミュレーション，性能評価

Power Consumption Evaluation of Supercomputer System ITO with MHD Simulation Code under Power Capping

KEIICHIRO FUKAZAWA[†1]　TAKESHI NANRI[†2]　HIROAKI HONDA[†3]

Abstract: The supercomputer system ITO consists of Skylake Xeon processors so that the Intel RAPL can be used to cap the CPU power consumption on the system. Using the RAPL, we constrain the CPU power of ITO system and run the MHD simulation code on it. The Skylake Xeon has dynamic variation of CPU frequency and its power consumption changes with it, however under the power capping, the variation of power consumption decreases, on the other hand the variation of CPU frequency still appears. The power consumption of DRAM is not so decreased under CPU power capping then the calculation performance gain due to the B/F value is appeared. In this study we will evaluate the detail of performance and measurement results and discuss them.

Keywords: Power consumption, Power capping, MHD simulation, Performance evaluation

1. はじめに

エクサフロップス級スーパーコンピュータシステム（エクサ級スパコン）を開発する上で，システムの消費電力が最大の問題となっている[1, 2]．エクサ級スパコンで利用可能な電力は 20 MW 程度と予測されており[2]，50 Flops/W の電力性能効率が求められている．これは現在の Green500 の Top 1 と比べても 3.9 倍程度の電力性能が必要となり[3]，この数年内での達成が難しい状況にある．一方で，計算機センターにとっても消費電力の増大に伴い電源容量が限界に達し，また電力料金が運用コストの大部分を占めるようになっており，運用の面からも消費電力の削減は重大な問題となっている．これら電力問題を解決するためには，前述のようなハードウェア単体の電力性能だけではなく，ミドルウェア，更にはアプリケーションレベルでの電力性能最適化が重要なアプローチと考えられる．つまり，現在のハードウェアやアプリケーションの最適化手法ではこれからのスパコンを開発・運用していくには不十分と考えられる．このような状況を考えると，今後アプリケーション開発者は電力性能最適化を行うために，自分のアプリケーションがどのような消費電力特性を持っているか理解しておく必要がある．

現在，アプリケーションの消費電力をコントロールするいくつかの手法が提案されている．例えば，Adagio Runtime [4] は DVFS (Dynamic Voltage and Frequency Scaling)を利用し，計算性能をほとんど下げずに消費電力を削減している (UMT2K や ParaDiS を利用)．Sandy Bridge 世代以降の Intel CPU では RAPL (Running Average Power Limit) が利用でき，Rountree らは RAPL を用いた電力制限下でのベンチマークアプリケーションの性能測定を行っている[5]．また，我々も RAPL を利用し，CPU と DRAM への電力供給バランスの最適化について報告している[6]

本研究では，九州大学情報基盤研究開発センターのスーパーコンピュータシステム ITO が搭載している最新の Xeon において，電磁流体力学（MagnetoHydroDynamic: MHD）シミュレーションコードを利用し，消費電力の評価を行っ

†1 京都大学・学術情報メディアセンター
Academic Center for Computing and Media Studies, Kyoto University
†2 九州大学 情報基盤研究開発センター
Research Institute for Information Technology Kyushu University
†3 株式会社ハイドロ総合技術研究所
Hydro Technology Institute Co., Ltd.

た．MHD シミュレーションコードは流体シミュレーションコードの 1 種であり，今回の研究結果は一般の流体計算にも適用できると考えられる．

本研究報告の構成は以下の通りである．第 2 章では，スーパーコンピュータシステム ITO について説明し，第 3 章では MHD シミュレーションコードについて説明をする．第 4 章で消費電力測定の結果を述べ，その結果を第 5 章で議論し，最後に研究のまとめをする．

2. スーパーコンピュータシステム ITO

スーパーコンピュータシステム ITO は, 2017 年度に九州大学情報基盤研究開発センターに導入された計算機システムであり，Skylake 世代の Xeon を搭載している．システム高性能詳細は表 1 の通りだが，多数の CPU 計算ノードが接続されたサブシステム A（2,000 ノード）と 1 ノード当たり 4GPU が搭載されたサブシステム B（128 ノード）があり，本研究ではシステム A のみを利用した．Skylake 世代の Xeon は Sandy Bridge 世代以降の Xeon のため，RAPL が利用できる．そこで本研究では，Inadomi らが開発した RAPL を利用するインターフェースである RIC を利用し，消費電力の測定，CPU 消費電力の制限をかけた場合の電力性能の評価を行っている[7].

3. MHD シミュレーションコード

宇宙空間は真空と思われているが，その 99 %はプラズマで満たされている．プラズマとは電離した気体のことであり，帯電している電子とイオンが分かれて存在する状態である．宇宙空間，特に我々の暮らす太陽系においては太陽から太陽風と呼ばれるプラズマの風が常時吹き出しており，太陽系全体にそのプラズマが充満している．このようなプラズマの振る舞いを記述する方程式として Vlasov-Maxwell 方程式がある．これは，無衝突 Boltzmann 方程式と Maxwell 方程式から成る．Vlasov（無衝突 Boltzmann）方程式は以下の形をとる．

$$\frac{\partial f_s}{\partial t}+\mathbf{v}\frac{\partial f_s}{\partial \mathbf{r}}+\frac{q_s}{m_s}(\mathbf{E}+\mathbf{v}\times\mathbf{B})\frac{\partial f_s}{\partial \mathbf{v}}=0 \qquad (1)$$

ここで **E**，**B**，**r** と **v** はそれぞれ電場，磁場，距離，速度を表す．また，$f_s(\mathbf{r}, \mathbf{v}_s, t)$は位置－速度位相空間における分布関数であり，$s$はイオンや電子など種類を示す．$q_s$ は電荷を m_s は質量を表す.

しかしながら，Vlasov 方程式は多くの成分からなる非線形方程式であり，計算機システムを用いても解くことが非常に難しい．そこで，Vlasov 方程式のモーメントをとることで求められる電磁流体力学（MHD）方程式が，グローバルなプラズマ構造を調べるときには使用されている．MHD 方程式は以下のようになる．

表 1　ITO サブシステム A の諸元
Table 1　Subsystem A of ITO

機種名	Fujitsu PRIMERGY CX2550/CX2560 M4	
計算ノード	CPU	Intel Xeon Gold 6154 (Skylake-SP) × 2 /node
	コア数	18 cores /CPU
	周波数	3.0 GHz (Turbo 3.7 GHz)
	理論性能	3,5 TFlops /node (**倍精度**)
	メモリ	DDR4 192 GB /node
	Bandwidth	255．9 GB/s /node
	B/F	0.074
総ノード数	2,000 nodes	
総理論性能	6.91 PFlops	
ノード間接続	InfiniBand EDR 4x (100Gbps)	

$$\begin{aligned}
\frac{\partial \rho}{\partial t}&=-\nabla\cdot(\mathbf{v}\rho)\\
\frac{\partial \mathbf{v}}{\partial t}&=-(\mathbf{v}\cdot\nabla)\mathbf{v}-\frac{1}{\rho}\nabla p+\frac{1}{\rho}\mathbf{J}\times\mathbf{B}\\
\frac{\partial p}{\partial t}&=-(\mathbf{v}\cdot\nabla)p-\gamma p\nabla\cdot\mathbf{v}\\
\frac{\partial \mathbf{B}}{\partial t}&=\nabla\times(\mathbf{v}\times\mathbf{B})
\end{aligned} \qquad (2)$$

上から，連続の式，運動方程式，圧力変化の式（エネルギーの式），最後が磁場の誘導方程式となる．簡単に言えば，電磁場を考慮した流体力学方程式と呼べる．詳しい導出方法は参考文献を参照されたい[8].

MHD 方程式を解く数値計算法としては，Modified Leap Frog（MLF）法[9，10]という計算法を使用する．これは最初の 1 回を two step Lax-Wendroff 法で解き，続く（l － 1）回を Leap Frog 法で解き，その一連の手続きを繰り返す．l の値は数値的に安定の範囲で大きい方が望ましいので，本手法で採用する 2 次精度の中心空間差分では，数値精度の線形計算と予備的シミュレーションから l = 8 に選んでいる．

本評価では，OpenMP と MPI によるハイブリッド並列を使用する．プロセス並列化手法としては 3 次元空間を分割する領域分割法を用いる．

4. 電力性能評価

4.1 全ノード利用時の消費電力性能

前述のように Skylake Xeon では RAPL が利用できるため，本研究では RAPL を利用する RIC という電力測定・キャッピング関数を用いて ITO の電力測定を行った．まず ITO の全ノード（2,000 ノード）を利用し，システム全体の消費電力を測定した．

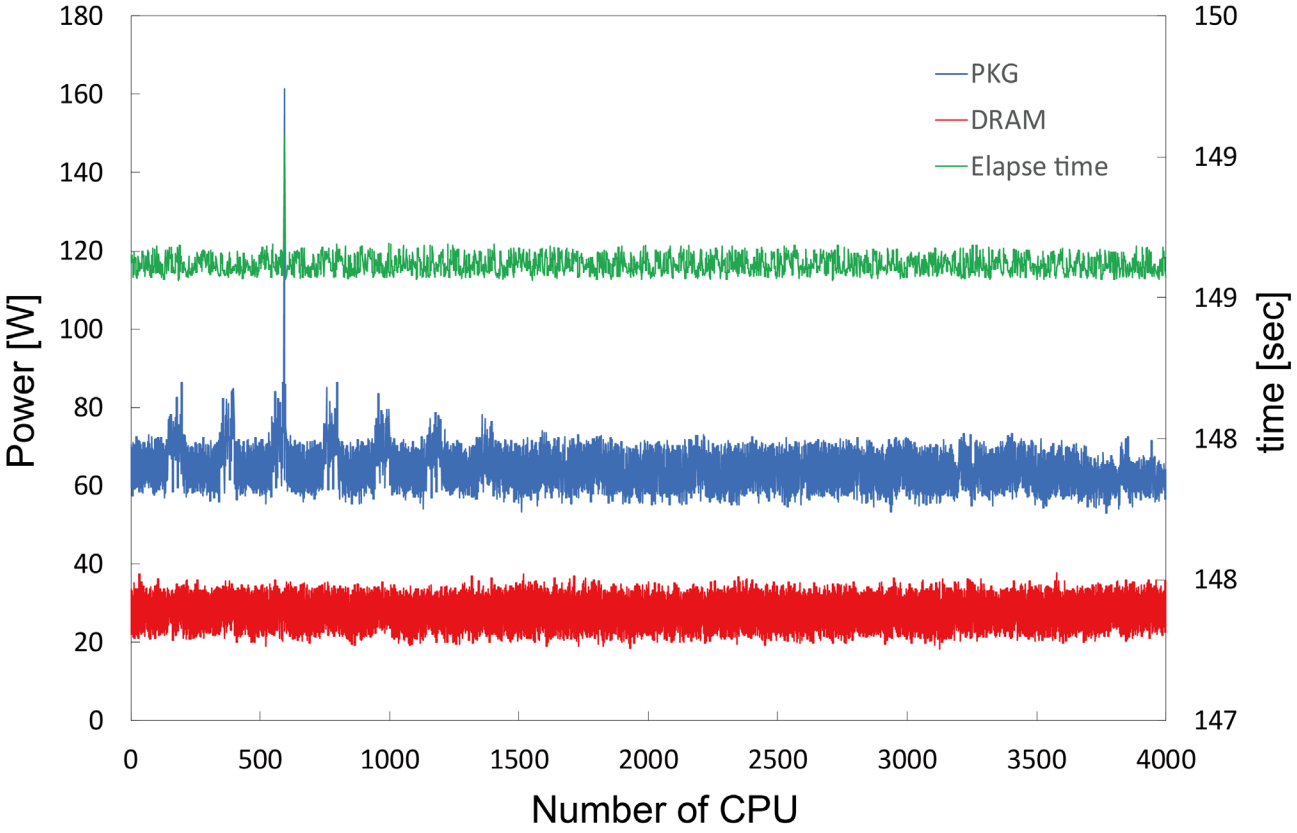

図 1　ITO 全ノードを利用した MHD シミュレーションコードの電力性能
Figure 1　Power performance of MHD code with all nodes of ITO

表 2　ITO 全ノードを利用した MHD シミュレーションコード実行時の消費電力特性
Table 2　Power consumption characters of MHD simulation code on ITO

	経過時間 [秒]	CPU 消費電力 [W]	DRAM 消費電力 [W]
平均	149.556	64.639	27.762
最大	150.968	86.372	37.738
最小	148.562	52.882	18.273

ITO の全ノード上で MHD シミュレーションコードを動作させた結果を図 1 に示す. ここでは 1 ソケット毎の MHD シミュレーションコード実行時間における平均 CPU 電力，平均 DRAM 電力，そして計算時間を示している．また，表 2 に図 1 の測定結果から最大・最小値と平均値を載せている. CPU の消費電力は 0～1,500 番でのこぎり状のぶれが見えており，その後はぶれが少し小さくなっている．しかしながら，最低でも 10 W 程度のぶれが常に現れる結果となった．平均は 64 W 程度であり，ぶれが約 15 %存在することになる．DRAM の消費電力は CPU の際に見えた特徴的なぶれ構造は見えないが, ぶれの幅が CPU と同程度ある結果となった．平均 DRAM 消費電力は 27 W であり，最大最小ともに約 10 W のぶれがあり, ぶれが平均消費電力の 37%をしめている．Skylake Xeon では，CPU 自体が周波数をダイナミックに変更する機能があるため，CPU の消費電力も大きく変化することはこれまでの測定結果から分かっていたが[11]，DRAM 消費電力がこれほど大きく変化するとは考えられていなかった．

CPU と DRAM の消費電力がばらつく一方で，計算時間はそれほどばらつきが見られない．MHD シミュレーションコードでは集団通信が含まれておらず，全体での同期は行われないため，単純に各ノードの計算時間にはぶれが少ないことが分かる．ITO の計算ノードには良い電力性能を示すノード（計算性能は他のノードと同様）があるなど電力特性がノード毎に異なることが報告されているが[12]，計算時間におけるぶれの少なさは，この特性を示していると考えられる．

全ノードを利用した合計の消費電力は CPU で 258 kW，DRAM で 111 kW となり，合計で 369 kW となった．その他の電力を消費する機器についての情報は無いが 100 kW 程度の消費電力と想定すると，MHD シミュレーションコードは ITO の全ノードを利用すると，500 kW 程度を消費

すると見積もられる．ベンチマークとしては最大規模の消費電力となる Linpack を計測すると ITO では 1,312.8 kW の消費電力となっている（4.5 PFlops 達成時）．実アプリケーションである MHD シミュレーションコードを実行すると，その約 40 %の電力を消費していることとなり，実運用上と設計電力には差があることが分かる．

4.2 CPU 消費電力制限下での電力特性

次に，CPU 消費電力に制限をかけ，MHD シミュレーションコードの性能，消費電力がどのように変化するか調べた．RAPL では機能的には DRAM への電力制限も可能だが，ITO では制限ができなかったため，CPU への消費電力制限のみを行っている．また，消費電力制限をかける場合は，2,000 ノードではなく，16 ノードを利用し，性能測定を行った．ITO に搭載された Skylake Xeon は，最大 CPU 消費電力が 413 W という設定であったため，CPU 消費電力の制限を 400 W から 10 W ずつ下げていき，60 W まで消費電力の制限をかけた．

図 2 に CPU 消費電力制限下での CPU と DRAM の消費電力を示す．ここでは，MHD シミュレーションコード実行時間中に 16 ノード（32CPU）において，CPU と DRAM が消費した最大・最小値と平均値を示している．これまでの研究で，Skylake Xeon ではアプリケーション実行中に大きく消費電力の振る舞いが変わることが分かっているため，図 1 の 2,000 ノードの測定結果と異なり，MHD シミュレーションコード実行中の消費電力変化をここでは見ている．MHD シミュレーションコードが基本的に CPU で約 150 W 消費することから，CPU 消費電力の制限が 150 W になるまでは，消費電力に変化はあまりない．少し最低 CPU 消費電力にバタつきが見えるが，それほど大きくはない．150 W 以下の CPU 消費電力に制限がかかると，実際に消費する電力もその制限値を取るような変化を示している．この結果，制限が無い場合には CPU 消費電力の大きなばらつきがあったが，CPU 消費電力に制限がかかると，制限値に律され，ぶれが見えなくなる．一方で DRAM の消費電力は，CPU 消費電力制限下においてもほとんど変化がない．最大 DRAM 消費電力だけが，少し減少しているようにも見える．

次に，CPU の周波数，MHD シミュレーションコードの実行時間を図 3 に示す．図 2 と同様に最大・最小と平均値を示している．CPU 周波数は直接的に消費電力に関連しているため，基本的には図 2 の CPU 消費電力と同じように，約 150 W の CPU 消費電力制限から，周波数の低下が見られる．しかしながら，消費電力では見えなくなったぶれが制限下でも残っており，CPU の電力性能の良いものと悪いものの差が出ていると考えられる．60 W の制限では，最大と最小で 1 GHz 程度の差がある一方で，消費電力は差が無いことから，低消費電力下では CPU 電力性能の差が大きく

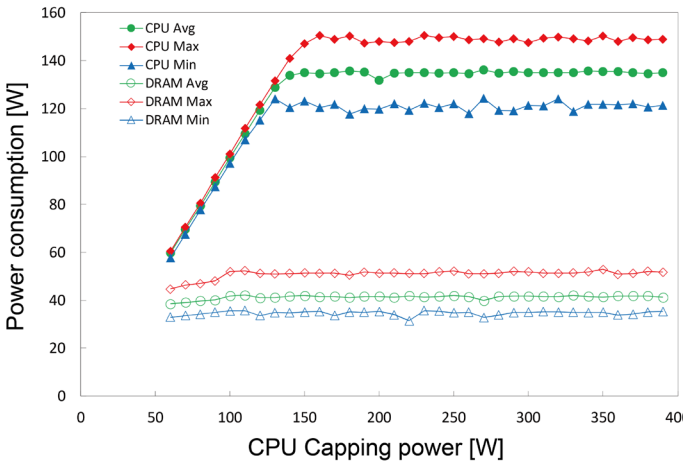

図 2　CPU 消費電力制限下における MHD シミュレーションコードの CPU と DRAM 消費電力

Figure 2　Power consumption of MHD simulation code under the CPU power capping

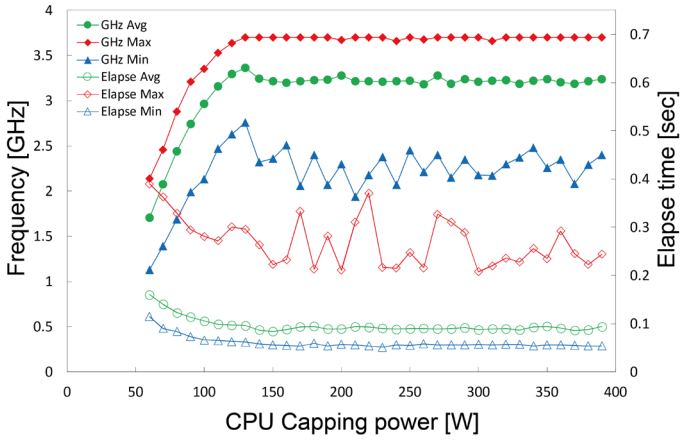

図 3　CPU 消費電力制限下における MHD シミュレーションコードの実行時間と CPU 周波数

Figure 3　Elapse time of MHD simulation and CPU frequency under CPU power capping

現れるように思われる．また，最低 CPU 周波数は，CPU 消費電力制限が効いていない間でも大きく変動している．この変動が CPU 最低消費電力に見えた小さなぶれの理由と思われるが，CPU の周波数を低くする条件は，CPU 温度など様々な要素が関係するため，一様になりにくいことがぶれの理由と考えられる．

MHD シミュレーションコードの実行時間は，CPU 消費電力より CPU 周波数の影響を強く受けているように見え，CPU 消費電力制限が無い間でも，最大計算時間が大きく変動している．最大計算時間は CPU 消費電力制限下において，大きく増加している一方で，最小と平均計算時間は緩やかに増加している．計算時間も CPU 消費電力と異なり，CPU 電力制限下でもばらつきは消えない．

Skylake Xeon では，CPU 消費電力に制限をかけない場合は，消費電力のばらつきが見え，計算機システム運用の面からは，扱いが難しいが，CPU 消費電力に制限をかけるとばらつきが消え，電力のコントロールが容易になる一方で計算性能にはばらつきが残り，最低 CPU 周波数への影響からか，計算時間の増加が大きく見られ，計算機利用者にとっては良くない効果が大きい．

5. 消費電力制限の計算性能への影響

図 3 にあるように，CPU 消費電力を制限すると，MHD シミュレーションコードの計算時間は遅くなり，アプリケーションを実行するユーザの面からは有用な点がない．しかしながら，電源容量の問題，季節による電力需要の問題，更には災害時における電力供給の制限など，スパコンセンターとして消費電力を制限せざるを得ないことは現実に起きている．そこで，これまでの CPU 消費電力制限下での計算性能がどのように変化しているかを調べ，消費電力制限下で最大の計算性能を出すことを考えることは重要である．

図 4 に CPU 消費電力と周波数，更に計算時間が，消費電力制限が無いときと比べて，どれだけ変化しているかを示した．簡単のために平均値を利用している．CPU の消費電力，周波数は消費電力制限の値に従い，ある程度一定の割合で減少していくことが分かるが，計算時間は 130～110 W 辺りだけでは，増加せずに一定に近くなっている．CPU 消費電力が減少し，周波数も下がると CPU の Flops 性能が減少する．一方で図 2 にあるように DRAM の消費電力は下がっておらず，DRAM のバンド幅は変化していないと考えられる．これにより，計算機が持つ B/F 値に改善が見られ，比較的高い B/F を必要とする MHD シミュレーションコードでは，計算性能が下がらなかったと考えられる．しかしながら，100 W 以上の消費電力制限をかけると計算性能は劣化していくことが分かり，ある程度の CPU と DRAM の消費電力バランスが必要と想定される．

そこで図 4 では CPU 消費電力を DRAM 消費電力で割った値（C/D index）を計算し，載せている．これによると C/D Index が 2.6 までは計算性能がそれほど下がらないということが分かる．ここから，電力制限が必要な場合は，C/D Index が 2.6 を下回らないような制限であれば，計算性能への影響はほとんど無く，また消費電力自体は削減が可能と考えられる．この C/D Index はアプリケーションの B/F 値と計算機の B/F 値が関連していると容易に想像できるが，計算機システムで計算性能を調べる際に，この Index も調べ，把握しておくことが今後重要になるかもしれない．

6. まとめ

本研究では，九州大学のスーパーコンピュータシステム

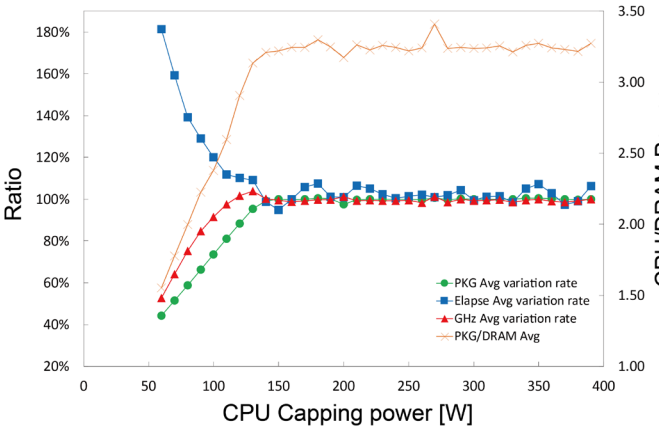

図 4　CPU 消費電力制限に対する CPU・DRAM 消費電力と周波数の変化率と CPU/DRAM 消費電力の関係

Figure 4　Relation between the variation rate of CPU and DRAM power consumptions and CPU frequency and CPU/DRAM power index

ITO を利用して，CPU 消費電力制限下での宇宙プラズマを解く MHD シミュレーションコードの消費電力測定を行った．消費電力はエクサスケール級の計算機だけで無く，スパコンセンターでも重要な課題となっている．CPU 消費電力制限を行わない場合に 2,000 ノードを利用した結果では，ノード毎で CPU と DRAM 消費電力に大きなばらつきが見える一方，計算時間にはそれほどのばらつきは見えなかった．2,000 ノード利用した全体の CPU と DRAM の消費電力は 369 kW となり，ネットワークなどを含むと 500 kW 程度の消費電力と想定される．これは ITO での Linpack 測定時に消費した電力の約 40 %となる．

CPU 消費電力に制限をかけた場合では，CPU 消費電力に現れていたばらつきが無くなったが，CPU 周波数や計算性能はばらつきが現れたままとなった．特に最低周波数に大きなばらつきが見え，最大計算時間にもその影響が大きく出ている．CPU 消費電力に制限をかけた場合も DRAM 消費電力は余り変化が無く，これにより計算機システムの B/F 値が改善し，ある制限区間では計算性能の劣化が少なくなった．このように CPU 消費電力制限下でも計算性能が劣化しないことがあるため，アプリケーション毎にその条件を調べておくことが，電力制限が当たり前の運用では，ユーザにとって重要である．

謝辞　本研究は，九州大学情報基盤研究開発センター平成 29/30 年度先端的計算科学研究プロジェクトの支援による．

参考文献

[1] P. M. Kogge, et al., Exa Scale Computing Study: Technology Challenges in Achieving Exascale Systems, in Exascale Computing Study Report, 2008.

(http://users.ece.gatech.edu/mrichard/ExascaleComputingStudyReports/exascale_final_report_100208.pdf)
[2] P. M. Kogge, and T. J. Dysart, "Using the TOP500 to trace and project technology and architecture trends," High Performance Computing, Networking, Storage and Analysis (SC), 2011 International Conference for , pp.1,11, 12-18 Nov. 2011.
[3] The Green 500 Site. (http://www.green500.org/)
[4] B. Rountree, D. K. Lowenthal, B. de Supinski, M. Schulz, V. W. Freeh, and T. Bletsch, “Adagio: Making DVS practical for complex HPC applications”, Proceedings of the 23rd international conference on Supercomputing, June 08-12, 2009, Yorktown Heights, NY, USA doi:10.1145/1542275.1542340.
[5] B. Rountree, D. Ahn, B. R. de Supinski, D. K. Lowenthal, and M. Schulz, 2012. “Beyond DVFS: A first look at performance under a hardware-enforced power bound”, Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2012 IEEE 26th International, pp.947,953, 21-25 May 2012, doi: 10.1109/IPDPSW.2012.116.
[6] K. Fukazawa, M. Ueda, Y. Inadomi, M. Aoyagi, T. Umeda, K. Inoue, "Performance Analysis of CPU and DRAM Power Constrained Systems with Magnetohydrodynamic Simulation Code", HPCC2018, 2018.
[7] Inadomi. Y., et al., “Analyzing and Mitigating the Impact of Manufacturing Variability in Power-Constrained Supercomputing”, Technical Paper, SC'15, Austin (USA).
[8] F. F. Chen, 1974. Introduction to Plasma Physics. Plenum Press, NY.
[9] T. Ogino, R. J. Walker, M. Ashour-Abdalla, A global magnetohydrodynamic simulation of the magnetopause when the interplanetary magnetic field is northward, IEEE Trans. Plasma Sci.20, 817.828, 1992.
[10] Fukazawa, K., T. Ogino, and R.J. Walker, "The Configuration and Dynamics of the Jovian Magnetosphere", J. Geophys. Res., 111, A10207, 2006.
[11] 深沢圭一郎，南里豪志，本田宏明，スーパーコンピュータシステム ITO における MHD シミュレーションコードの計算性能・消費電力評価，研究報告ハイパフォーマンスコンピューティング（HPC）,2018-HPC-166(4),1-7 (2018-09-20) , 2188-8841.
[12] L. Li , K. Fukazawa , H. Nakashima , T. Nanri, A Node Level Performance/Power Efficiency Aware Resource Management Technique，研究報告ハイパフォーマンスコンピューティング（HPC）,2018-HPC-166(3),1-7 (2018-09-20) , 2188-8841.