

Reedbush-HにおけるGPUの電力ばらつき

三輪 忍

電気通信大学大学院情報理工学研究所

1. はじめに

近年の世界トップレベルのスーパーコンピュータの消費電力は運用上の限界に達しており、システムの省電力化が次世代スーパーコンピュータ開発における重要な課題となっている。例えば、2018年6月時点において世界最速のスーパーコンピュータである Summit¹は **8.8MW** の電力を消費して **200ペタフロップス²**のピーク演算性能を実現しているのに対し、現在米国において実施されている次世代スーパーコンピュータの開発プロジェクト (Exascale Computing Project) の開発目標は **20-30MW** の消費電力で**エクサフロップス³**級の演算性能を有するシステムとなっている。すなわち、最先端機と次世代機の間には**電力性能⁴**に換算すると 1.5~2.2 倍の開きが存在しており、この事実は、近年のスーパーコンピュータ開発の傾向である、消費電力の増加を抑制しつつ高い演算性能を提供することの難しさを表していると言える。

このような背景から、多くのスーパーコンピュータセンターでは**ヘテロジニアスシステム**が運用されている。ヘテロジニアスシステムは、演算装置として CPU (Central Processing Unit) 以外に大量の GPU (Graphic Processing Unit) を搭載することによって、データレベル並列性が高いアプリケーションに対して少ない消費電力で高い演算性能を提供する。実際、上述の Summit を始め、東京大学情報基盤センターにて運用中の Reedbush-H/L など、多くのスーパーコンピュータが GPU を搭載している。ヘテロジニアスシステムは厳しい電力制約下において高演算性能を実現するための重要な解と位置づけられており、事実、世界トップ 10 のスーパーコンピュータの半数がヘテロジニアスシステム⁵である。

ところで、スーパーコンピュータを構成する CPU には電力ばらつきが存在することが知られており、近年、ばらつきの影響を考慮したスーパーコンピュータの電力管理手法に関する研究が盛んに行われている。スーパーコンピュータの各計算ノードには同じ製品仕様の CPU が利用されることが多いが、これらの CPU は、理論上、同じアプリケーションに対して同じ電力性能を示すことが期待される。しかしながら、実際のスーパーコンピュータにおいては、LSI の製造ばらつき⁶やマシンルーム内の温度ばらつき⁷などの影響により、各 CPU の電力性能が異なる (B. Rountree, et al. 2012; D. Hackenberg 2014)。そこで、CPU 毎の電力性能を考慮しつつジョブを計算ノードに適切に配置するジョブスケジューラなど、CPU の電力ばらつきを利用し

¹ 米国オークリッジ国立研究所に設置。

² ペタフロップス=1秒間に 10^{15} 回の浮動小数点演算を処理可能な性能。

³ エクサフロップス=1秒間に 10^{18} 回の浮動小数点演算を処理可能な性能。

⁴ 電力当たりの性能。1ワット当たりのフロップス値 (FLOPS/W) が用いられることが多い。

⁵ 2018年6月の TOP500 ランキング (<https://www.top500.org/>) による。

⁶ 製造工程において発生する LSI の不均質性。

⁷ マシンルーム内の温度は一定の温度以下となるように管理されているが、空調の位置やラック内の位置によってコンピュータの周辺温度は異なる。また、LSI の静的電力はチップ温度に依存するため、同じ処理を行う場合でも、チップの周辺温度が異なる場合は消費電力が異なることが報告されている (池淵大輔ら 2010)。

システム構成		ノード構成	
項目	詳細	項目	詳細
総理論演算性能	1418.2 TFlops	CPU	2x Intel Xeon E5-2695v4, 18C, 2.1 GHz
総ノード数	120		
総主記憶容量	30 TB	メモリ	256 GB DDR4-2400
ネットワーク	Full-bisection Fat Tree, InfiniBand FDR 4x	GPU	2x NVIDIA Tesla P100, 300W TDP, 56 SM, 16 GB Memory, CPU-GPU 間 : PCI Express Gen3 x16 (16 GB/s) GPU-GPU 間 : NVLINK 2 brick (20 GB/s x2)
ファイルシステム	並列ファイルシステム (5.04 PB, 145.2 GB/s) +高速ファイルキャッシュ (209 TB, 436.2 GB/s)		

第 1 図: Reedbush-H のハードウェア構成。

てスーパーコンピュータの電力性能を改善する手法が開発されている (Y. Inadomi, et al. 2015; E. Tottoni, et al. 2015; B. Acun, et al. 2016; R. Sakamoto, et al. 2017)。

一方、ヘテロジニアスシステムには同じ製品仕様の GPU が大量に搭載されていることから、ヘテロジニアスシステムを構成する GPU にも電力ばらつきが存在することが予想される。GPU は CPU と同様のプロセス技術を用いて製造されることから、GPU にも CPU と同様の製造ばらつきが存在する可能性が高い。また、ヘテロジニアスシステムを設置したマシンルームの冷却設備はホモジニアスシステムを設置したマシンルームの冷却設備と大きな違いはないことから、前者のマシンルーム内にも温度ばらつきが存在する可能性が高い。以上のように、ヘテロジニアスシステムを構成する GPU には電力ばらつきが存在する可能性が極めて高いものの、GPU の電力ばらつきに関する報告はほとんどない。仮に GPU にも CPU と同程度の電力ばらつきが存在するのであれば、GPU のばらつきを考慮してジョブやハードウェアを制御することによって、ヘテロジニアスシステムの電力性能を大幅に改善できる可能性がある。

このような背景から、2018 年 1 月、および、同年 7 月に実施した計 2 回の大規模 HPC チャレンジにおいて、我々の研究グループは Reedbush-H における GPU の電力ばらつき調査を実施した。より具体的には、Reedbush-H に搭載された計 240 台の GPU において実アプリケーションを含む計 6 種類のアプリケーションを実行した時の GPU の消費電力やアプリケーション性能などを測定し、分析を行った。その結果、1) Reedbush-H の GPU において有意なばらつきが存在すること (平均消費電力の差は最大 19.2%, 消費エネルギーの差は最大 20.2%), 2) アプリケーションによってばらつきの大きさは異なること、3) アプリケーションによって最も電力性能の良い (悪い) GPU は異なること、が確認できた。本稿では、その詳細について報告する。

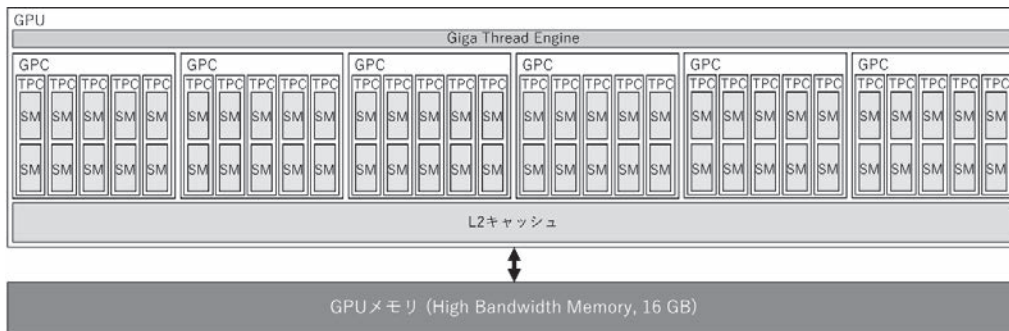
本稿の構成は以下の通りである。まず次章では、実験環境である Reedbush-H のアーキテクチャについて述べる。続く 3 章では今回行った実験の方法について説明し、4 章では実験結果を示す。5 章で関連研究についてまとめた後、最後 6 章でまとめと今後の展望を述べる。

2. Reedbush-H

Reedbush-H は東京大学情報基盤センターにて 2017 年 3 月に運用を開始したスーパーコンピュータであり、同センターが初めて運用するヘテロジニアスシステムでもある。Reedbush-H の

項目	詳細
OS	Red Hat Enterprise Linux 7
コンパイラ	GNU, PGI, NVCC
通信ライブラリ	Intel MPI, SGI MPT, Open MPI, MVAPICH2, Mellanox HPC-X, GPUDirect for RDMA (OpenMPI, MVAPICH2-GDR)
その他のライブラリ	FFTW, GNU Scientific Library, NetCDF, Parallel netCDF, Xabclib, ppOpen-HPC, ppOpen-AT, MassiveThreads, OpenJDK, cuBLAS, cuFFT, cuSPARSE, MAGMA, OpenCV, ITK, Theano, Anaconda, ROOT, TensorFlow
アプリケーション	Torch, Caffe, Chainer, GEANT4

第 2 図: Reedbush-H のソフトウェア構成。



第 3 図: GP100 アーキテクチャ。

ハードウェア構成 (坂井朱美ら 2017) を第 1 図に示す。Reedbush-H は計 120 台の計算ノードから構成されており、ノードあたり 2 本の 4x FDR InfiniBand リンクを用いてフルバイセクション Fat Tree ネットワークを構築している。各計算ノードは、Intel 社の Xeon E5-2695v4 を 2 基と 256GB の DDR4 メモリを搭載する。また、各計算ノードには、CPU に加えて、演算アクセラレータとして NVIDIA 社の Tesla P100 が 2 基搭載されている。したがって、Reedbush-H には計 240 台の Tesla P100 が搭載されており、ばらつきの調査に十分な数の GPU を備えている。

Reedbush-H のソフトウェア構成を第 2 図に示す。Reedbush-H では OS として Red Hat 社の Enterprise Linux 7 を採用しており、GNU, PGI, NVCC などのコンパイラが使用可能である。また、ノード間通信のためのライブラリとして Intel MPI, Open MPI, MVAPICH2 などの各種 MPI ライブラリを Reedbush-H はサポートする。さらに、Reedbush-H では Root, Theano, TensorFlow などのデータ解析用/ディープラーニング用のソフトウェアが豊富に用意されており、高性能計算環境を必要とするデータサイエンティストや機械学習エンジニアの期待に応えるシステム構成となっている。

Tesla P100 で採用されている GP100 のアーキテクチャを第 3 図に示す。GP100 は、1 つの Giga Thread Engine と 6 つの GPC (Graphics Processing Cluster)、および、1 つの L2 キャッシュによって構成されている。各 GPC は 5 つの TPC (Texture Processing Cluster) からなり、各 TPC は 2 つの SM (Streaming Multiprocessor) からなる。各 SM には単精度浮動小数点演算を実行可能な演算コアである CUDA コアが 64 個搭載されており、各 CUDA コアが 1 つのスレッドを実

プログラム名	詳細
SMC	ホスト/GPU メモリ間のデータコピー (from CUDA SDK 8.0.44)
QMCPACK	量子モンテカルロ計算 (from CORAL-2)
MTEST	SAXPY を行う自作プログラム
HACC	宇宙論的 N 体シミュレーション (from CORAL-2)
MM	行列積演算 (from CUDA SDK 8.0.44)
MM-CUBLAS	cuBLAS ライブラリを使用した行列積演算 (from CUDA SDK 8.0.44)

第 4 図: ベンチマークプログラム。

行する。したがって、GP100 の総 SM 数は 60 個、総 CUDA コア数は 3,840 個である⁸。また、GP100 の GPU チップはメモリコントローラを介して GPU メモリと接続される。GPU メモリには 16GB の HBM2 (High Bandwidth Memory) が使用されており、これにより 732GB/s の広メモリ帯域幅を提供する。

ここで強調しておきたいことは、GP100 の GPU と GPU メモリ (具体的には HBM2 を構成する各 DRAM チップ) は別チップであり、それぞれのチップは異なる半導体プロセスを用いて製造される点である。そのため、GPU と GPU メモリでは製造ばらつきの傾向が異なることが予想される。後述するように、Tesla P100 の電力ばらつきの傾向はアプリケーションによって大きく異なるが、これは、アプリケーションによって GPU と GPU メモリの使用率が異なっており、アプリケーションを実行した際にそれぞれの製造ばらつきから受ける影響が異なるためと我々は考えている。

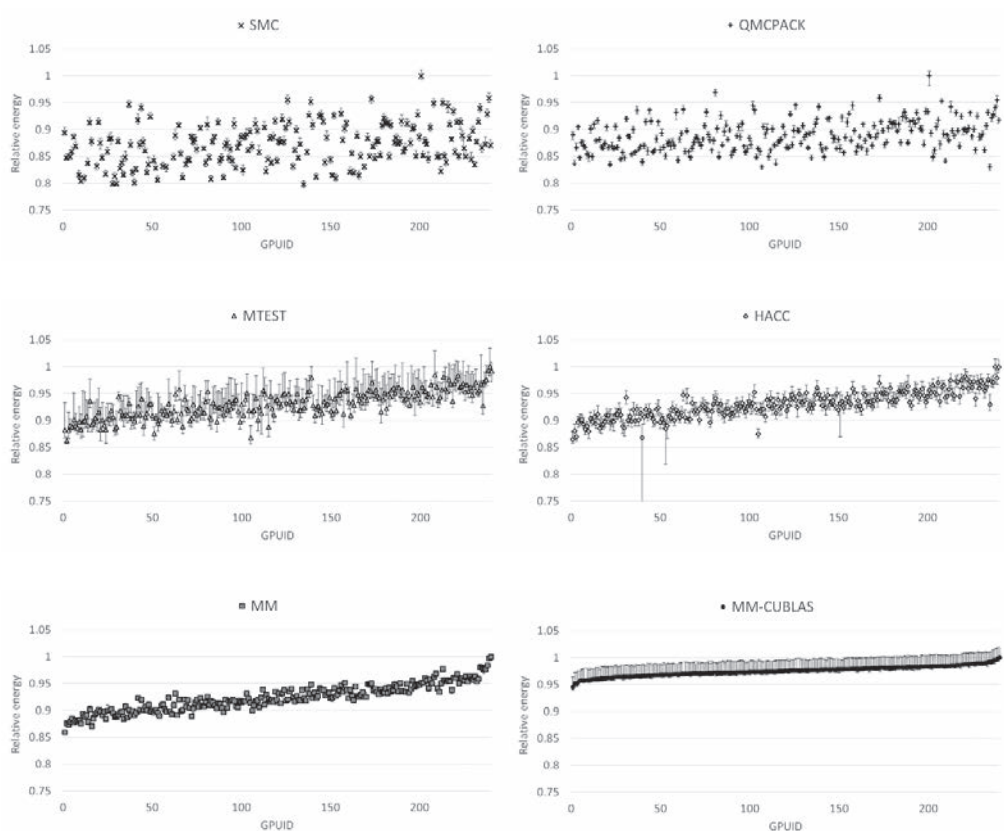
3. 実験方法

実験に使用したベンチマークプログラムを第 4 図に示す。計 6 本のプログラムを実験に使用した。6 本のプログラムのうちの 3 本 (MM, MM-CUBLAS, SMC) は GPGPU において頻出する基本的な演算を行うプログラムであり、CUDA SDK 8.0.44 から選択した。残りの 3 本のうちの 2 本 (HACC, QMCPACK) は CORAL-2 ベンチマークプログラムから選択した実 HPC アプリケーションである。残りの 1 本 (MTEST) はメモリアクセス性能をテストするために作成した自作プログラムである。

同一アプリケーションを実行した時の GPU の電力ばらつきを確認するため、MM, MM-CUBLAS, SMC, MTEST の 4 本のプログラムについてはシングル GPU 実行を行った。すなわち、Reedbush-H の 240 台の GPU の中から 1 つを選んでプログラムを実行するという実験を全 240 台の GPU に対して繰り返した。シングル GPU 実験においては、各プログラムを各 GPU において 20-50 回実行した。一方、HACC, QMCPACK については、実際の利用状況における GPU の電力ばらつきを確認するため、全 240 台の GPU を使用してマルチ GPU 実行を行った。マルチ GPU 実験における各プログラムの実行回数は 20-30 回とした。

今回測定した項目は、アプリケーションの実行時間、アプリケーションを実行中の GPU の平均消費電力と平均温度の 3 つである。前述のように、製造ばらつき以外にマシンルーム内の温

⁸ GP100 を搭載した GPU ボードである Tesla P100 では、歩留まりを改善するために 60 個の SM のうちの 4 個が無効化されており、エンドユーザが使用可能な SM は 56 個に制限されている。



第 5 図: Reedbush-H の各 GPU の消費エネルギー。

度ばらつきも GPU の電力ばらつきの原因と考えられることから、アプリケーション実行中の GPU の平均温度についても測定を行った。GPU の平均消費電力と平均温度は、nvidia-smi コマンドによって 1 ミリ秒毎の GPU の消費電力と温度を測定し、得られた値を平均化することによって求めた。

4. 実験結果

Reedbush-H の各 GPU の消費エネルギーを第 5 図に示す。グラフの横軸は GPU の ID を表しており、Reedbush-H に搭載された計 240 台の GPU を、MM-CUBLAS を実行した際の平均消費エネルギーの少ない順に並べてある。縦軸は各 GPU の消費エネルギーを表しており、プログラム毎に消費エネルギーが最大となる GPU の平均消費エネルギー値で正規化してある。6 枚のグラフは 6 本のプログラムに対応しており、グラフ上の各点は当該プログラムを当該 GPU 上で複数回実行した時の消費エネルギーの平均値である。各点の上下に延びたひげは、複数回測定した消費エネルギーの最大値/最小値を表す。

グラフより、我々が当初予想した通り、Reedbush-H の GPU においてばらつきが存在することが確認できる。例えば、SMC を実行した時に最も平均消費エネルギーが少なかった GPU (ID は 134) は、最も平均消費エネルギーが多い GPU (ID は 200) の 79.8% の平均消費エネルギーであった。この GPU 間の消費エネルギーの差は、SMC や QMCPACK などいくつかのプログラムでは、1

	SMC	MM	MM-CUBLAS
IPC	0.75	1.28	1.87
GPU メモリアクセス率	20.4%	2.08%	2.46%

第 6 図： SMC, MM, MM-CUBLAS の IPC と GPU メモリアクセス率。

	SMC	QMCPACK	MTEST	HACC	MM	MM-CUBLAS
相関係数	0.19	0.16	0.21	0.30	0.25	0.073

第 7 図： GPU 毎の平均消費エネルギーと平均 GPU 温度との間の相関係数。

つの GPU でプログラムを複数回実行した時の試行間の消費エネルギー差（すなわち、ひげの長さ）よりも大きい。したがって、Reedbush-H の GPU 間には、有意な消費エネルギーの差が存在すると言える。

また、上記のグラフからは、GPU 間の消費エネルギーの差がプログラムによって異なることも確認できる。例えば、SMC を実行した時の GPU 間の平均消費エネルギーの差は最大 20.2% であるのに対し、MM-CUBLAS を実行した時のそれは最大 5.5% に過ぎない。このように、平均消費エネルギーのばらつきの程度はプログラムによって異なる。

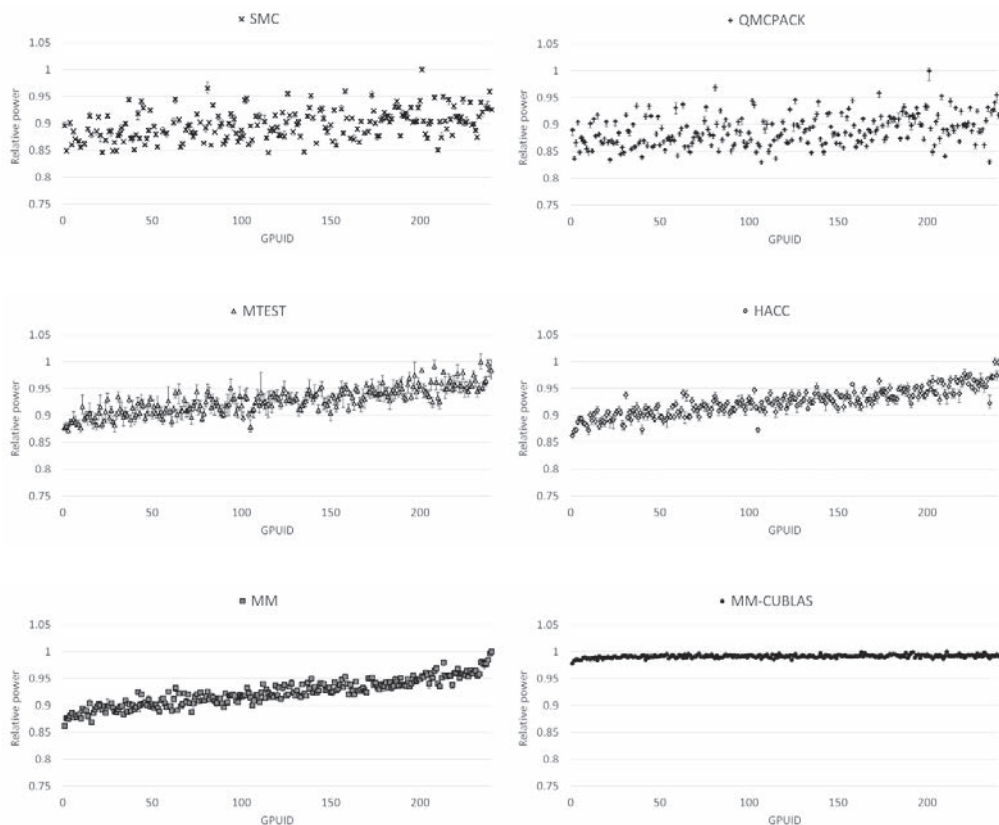
興味深いことに、消費エネルギーのばらつきの傾向もプログラムによって異なることが今回の実験では確認された。例えば、SMC を実行した時に最も消費エネルギーの少ない GPU は ID が 134 番の GPU であったのに対し、MM-CUBLAS を実行した時のそれは ID が 0 番の GPU、また、QMCPACK を実行した時のそれは 234 番の GPU であった。プログラムによって消費エネルギーの少ない GPU が異なることから、GPU の電力管理手法を開発する際は、GPU 毎のエネルギー効率だけでなく、プログラム毎のエネルギー効率にも注意する必要がある。

SMC, MM, MM-CUBLAS の IPC と GPU メモリアクセス率を第 6 図に示す。GPU メモリアクセス率は GPU メモリへのリードおよびライトの総数を全実行命令数で割った値である。アプリケーションの IPC, GPU メモリへのリード/ライト数, 実行命令数は nvprof を用いて計測した。表より、消費エネルギーのばらつきの傾向が類似している MM と MM-CUBLAS は計算インテンシブなアプリケーションであるのに対し、これらのアプリケーションとは消費エネルギーのばらつきの傾向が大きく異なる SMC はメモリインテンシブなアプリケーションであることがわかる。

前述のように Tesla P100 の GPU チップと GPU メモリは別チップであることから、それぞれのチップには異なる傾向の製造ばらつきが存在すると推測される。すなわち、計算インテンシブなアプリケーションである MM と MM-CUBLAS は GPU チップの製造ばらつきの影響を強く受けるのに対し、メモリインテンシブなアプリケーションである SMC は GPU メモリの製造ばらつきの影響を強く受ける。その結果として、これらのアプリケーション間では消費エネルギーのばらつきの傾向が異なると考えられる。この点に関しては今後詳しく調査する予定である。

各アプリケーションにおける GPU 毎の平均消費エネルギーと平均 GPU 温度との間の相関係数⁹を第 7 図にまとめる。表に示すように相関係数はいずれのアプリケーションにおいても 0.30 以下となっており、GPU の平均消費エネルギーと平均 GPU 温度との間の相関は弱いことが確認

⁹ 2 つの確率変数間の線形関係の強弱を表す。-1 から 1 までの値を取り、0 に近いほど相関が弱い。



第 8 図: Reedbush-H の各 GPU の平均消費電力。

	Avg. Power (CV)	Exec. Time (CV)	Energy (CV)	Temperature (CV)
SMC	51.3 W (0.031)	94.3 sec (0.031)	4.83 kJ (0.044)	35.3 °C (0.051)
QMCPACK	53.6 W (0.032)	125.8 sec (0.0004)	6.75 kJ (0.032)	37.6 °C (0.050)
MTEST	77.7 W (0.025)	8.92 sec (0.0049)	0.694 kJ (0.026)	37.0 °C (0.051)
HACC	130.4 W (0.027)	77.3 sec (0.0039)	10.1 kJ (0.027)	41.6 °C (0.047)
MM	212.0 W (0.027)	121.8 sec (0.0011)	25.8 kJ (0.027)	54.4 °C (0.031)
MM-CUBLAS	294.2 W (0.0030)	78.1 sec (0.0083)	23.0 kJ (0.0092)	58.8 °C (0.028)

第 9 図: 各アプリケーションの消費電力, 実行時間, 消費エネルギー, GPU 温度。

できた。後述するように今回の実験では平均 GPU 温度のばらつきは大きかったものの、予想に反して GPU 温度が GPU の平均消費エネルギーに与えた影響は小さかったと言える。

各アプリケーションを実行中の各 GPU の平均消費電力を第 8 図に示す。第 8 図を第 5 図と比較するとわかるように、平均消費電力のばらつきの傾向は消費エネルギーのばらつきの傾向と類似している。すなわち、第 5 図に示した GPU 間の消費エネルギーのばらつきは、アプリケーション実行時の GPU 間の平均消費電力のばらつきに起因すると言える¹⁰。

¹⁰ GPU アプリケーションの消費エネルギーは GPU の平均消費電力と実行時間の積で与えられる。

各アプリケーションの平均消費電力、平均実行時間、平均消費エネルギー、平均 GPU 温度を第 9 図にまとめる。表の括弧内の数値は GPU 間の変動係数を表している。表より、平均実行時間の変動係数は平均消費電力、あるいは、平均消費エネルギーの変動係数よりも小さく、したがって、平均実行時間に関する GPU 間のばらつきは小さいことがわかる。また、平均 GPU 温度の変動係数は 0.028 から 0.051 とやや大きいものの、第 7 図に示したように GPU 温度のばらつきが GPU の消費エネルギーに与える影響は小さい。

同表の平均消費電力の列に着目すると、Tesla P100 の TDP (Thermal Design Power) は 300W であるのに対し、ほとんどのアプリケーションの平均消費電力は 300W に満たないことがわかる。例えば、SMC は平均 51.3W の電力しか消費しておらず、したがって、SMC の実行中は 248.7W の使用可能な電力を余している。ヘテロジニアスシステムの電力性能向上のためにはこのようなアプリケーションにおける電力資源の利用法を改善する必要がある。

5. 関連研究

前述のように、スーパーコンピュータの CPU のばらつきに関してはいくつかの調査が既に行われており、ばらつきを考慮した電力管理手法の開発が進められている。以下、CPU のばらつきの研究事例についてまとめる。

B. Rountree らは、米国ローレンスリバモア国立研究所に設置された 2 種類の PC クラスタ (Zinfandel: 32 ノード, Merlot: 137 ノード) の各ノード上で NAS Parallel ベンチマークを実行し、アプリケーションの実行時間と CPU の消費電力を計測した (B. Rountree, et al. 2012)。その結果、各ノードの CPU の消費電力には有意な差 (最大 10W) が存在することを確認した。また、CPU に電力制約を課した場合、上記の電力ばらつきは動作周波数のばらつきに転嫁され、その結果としてアプリケーションの実行時間にばらつき (最も高速な CPU と最も低速な CPU との実行時間の差は 1 割程度) が生じることが報告されている。

CPU 単体ではなく計算ノード全体の電力ばらつきに関する報告もある。D. Hackenberg は 2 種類のスーパーコンピュータ (SuperMUC, Taurus. どちらも GPU は非搭載) の各ノード上でアプリケーションを実行した結果、ノード間に最大 41W の消費電力差があったことを報告している (D. Hackenberg 2014)。また、ノードの消費電力の 3/4 は CPU によるものであり、上記の電力ばらつきは CPU の電力ばらつきに依るところが大きいことも併せて報告している。

CPU の電力ばらつきはターボブースト¹¹の効果を左右する。B. Acun らの調査によると、CPU のターボブーストを有効化したスーパーコンピュータにおいては、CPU の電力ばらつきによって最大 16% のアプリケーション性能差が生じる (B. Acun, et al. 2016)。この問題への対策として、彼らは、CPU の性能差を考慮して実行中のタスクを再配置するロードバランサーなどを提案している。

ジョブのスケジューリング時に CPU の電力ばらつきを考慮する手法についても研究が進められている。R. Sakamoto らは、電力制約下にあるスーパーコンピュータにおいて、ノード間の電力ばらつきを考慮してジョブをノードに割り当てるスケジューラを開発した (R. Sakamoto, et al. 2017)。E. Totoni らは、CPU 内のコアの電力ばらつきを考慮してジョブを割り当てるス

¹¹ Intel 社の CPU で用いられている、負荷が低い時に動作周波数を定格以上の値に変更することで CPU 性能を向上させる技術。同様の技術は、AMD や IBM など他社の CPU においても採用されている。

ケジューラを開発した (E. Totoni, et al. 2015)。これらの方式は、CPU の電力ばらつきが存在する環境において、電力効率の良いジョブの割り当て法を提供する。

Y. Inadomi らは、電力制約下にあるスーパーコンピュータ上でマルチノードジョブを実行する状況を想定し、ジョブ性能を最大化する各ノードの電力割当法を開発した (Y. Inadomi, et al. 2015)。彼らの方法では、スーパーコンピュータの設置時にテストプログラムを実行することによって各計算ノードの CPU とメモリの電力を計測し、それらの電力ばらつきを記録しておく。また、ジョブの実行前には、アプリケーションを 1 ノード上で実行し、CPU とメモリの消費電力値を取得する。そして、ジョブの実行時には、テストプログラムによって得られた電力ばらつきに関する情報、および、アプリケーションごとの CPU とメモリの消費電力値から各計算ノードの CPU とメモリの消費電力を推定し、電力制約を満たす最適な電力配分を決定する。九州大学が有するスーパーコンピュータ HA8000 (計算ノード数は 960) を用いて実験を行った結果、提案手法によってジョブ性能が最大 5.4 倍改善したと報告されている。

以上のように、CPU や GPU 非搭載の計算ノードの電力ばらつきに関する報告は存在するものの、GPU や GPU 搭載ノードの電力ばらつきに関する報告はほとんどない。GPU の電力ばらつきについては未解明な部分がまだ多いことから、今回、大規模 HPC チャレンジを利用して Reedbush-H の GPU の電力ばらつきに関する実験を実施した。

6. まとめと今後の展望

本稿では、2018 年 1 月、および、同年 7 月の大規模 HPC チャレンジにて行った、Reedbush-H における GPU の電力ばらつきの実験結果について報告した。今回の実験では、Reedbush-H に搭載された 240 台の Tesla P100 間には有意なばらつき (平均消費電力の差は最大 19.2%, 消費エネルギーの差は最大 20.2%) が存在しており、ばらつきの大きさや傾向はアプリケーションによって異なることが確認できた。

今後は、ばらつきの大きさや傾向を決定するアプリケーションの特徴について、さらなる解析を進める予定である。また、ばらつきを考慮した GPU の電力管理手法を開発することによって、Reedbush-H を始めとするヘテロジニアスシステムの高電力効率な運用法を確立したいと考えている。

謝 辞

本研究は、東京大学情報基盤センター「Reedbush-H 大規模 HPC チャレンジ」における採択課題「Reedbush-H における GPU のばらつきに関する研究」および「メモリ・アクセス頻度が GPU のばらつきに与える影響の調査」によって行われた。SMC, MM, MM-CUBLAS の実験に関しては、電気通信大学の田所英行君に大いに協力していただいた。また、本研究を進めるにあたり、電気通信大学の本多弘樹教授、八巻隼人助教、および、東京大学の埴敏博准教授、坂本龍一助教、有間英志特任助教には非常に有意義なご意見をいただいた。ここに感謝の意を表する。

参 考 文 献

- P. Messina and S. Lee, “The U. S. Exascale Computing Project,” SC16 (Birds of a Feather), 2016
- B. Rountree, et al., “Beyond DVFS: A First Look at Performance Under a Hardware-Enforced

- Power Bound,” IPDPS Workshops (HPPAC), pp. 947-953, 2012
- D. Hackenberg, “Node Power Consumption Variability,” Energy-Efficient HPC WG Workshop, 2014
- 池淵大輔ら, 『細粒度パワーゲーティングを適用した汎用マイクロプロセッサ Geysers-1』, 情報処理学会研究報告 2010-ARC-187/2010-EMB-15, No. 1, pp. 1-6, 2010
- Y. Inadomi, et al., “Analyzing and Mitigating the Impact of Manufacturing Variability in Power-Constrained Supercomputing,” SC15, No. 78, pp. 1-12, 2015
- E. Totoni, et al., “Scheduling for HPC Systems with Process Variation Heterogeneity,” PPL Technical Report, 10 pages, 2015
- B. Acun, et al., “Variation Among Processors Under Turbo Boost in HPC Systems,” ICS, No. 6, pp. 1-13, 2016
- R. Sakamoto, et al., “Production Hardware Overprovisioning: Real-world Performance Optimization using an Extensible Power-aware Resource Management Framework,” IPDPS, pp. 957-966, 2017
- 坂井朱美ら, 『Reedbush スーパーコンピュータシステムの運用』, 大学 ICT 推進協議会, 2017
- NVIDIA, “CUDA Toolkit Documentation,” <https://docs.nvidia.com/cuda/>
- CORAL Collaboration, “CORAL-2 Benchmarks,” <https://asc.llnl.gov/coral-2-benchmarks/>