卒業論文

電力制約下における蓄電池を用いた高性能計算システムの性能向上

03-120601　**酒井 崇至**

指導教員　中村 宏 教授

2014 **年** 2 **月**

東京大学工学部計数工学科システム情報工学コース

概要

近年、コンピュータの消費電力の増大が大きな問題となっており、コンピュータの性能の指標として単なる実行速度だけではなく、消費電力あたりの実行速度（電力対性能）が重要視されるようになってきている。特にスーパーコンピュータのような今日の高性能計算システムでは数メガワットもの電力を消費しており、物理的制約からこれ以上の電力供給力の向上は困難である。このような背景により、予め決められた消費電力の制約下での実行速度の最大化が、今後の高性能計算システムの性能向上の鍵となっている。

そこで、本論文では蓄電池を用いた高性能計算システムの性能を向上手法を提案する。現在の高性能計算システムには、停電時にもシステムへの電力供給を続けられるように UPS（無停電電源装置）が搭載されている。それを非停電時にも積極的に充放電を行い、アプリケーションの中の電力対性能が上がりにくい部分から上がりやすい部分へ時間方向に電力を融通することによって、電力制約下における性能を向上させることができる。今回はこの手法を CPU-GPU ハイブリッド構成の計算ノードを用いて 3 種類のベンチマークで性能評価実験を行い、この手法を用いない場合に比べて平均？ % の性能向上が実現できることを示し、その有用性を確認した。

目次

第 1 章

序論

現代社会においてコンピュータの担う役割はかつてないほど大きくなっており、我々の生活に欠くことのできない存在となっている。より高性能なコンピュータを作るべく、これまで多くの研究者がコンピュータ技術の発展に貢献し、Moore の法則 [1] の示す通りチップの集積度が指数関数的に増すと共にコンピュータの性能も向上し続けている。

近年、コンピュータの性能向上の妨げとなっている要因の一つが消費電力の増大である。一般に、高速な演算を行うためには大きな電力を消費しなければならず、数年前までは性能向上と共に消費電力も増加し続けてきた。ところがスーパーコンピュータなどの HPC 領域においては既に供給できる限界に近い電力を消費しており、物理的な電力供給能力によってコンピュータの性能が制限されている。そのため、与えられた電力制約の中でいかに処理能力を向上させるかが現在の大きな課題となっている。

この課題を解決するため、プロセッサやメモリの動作速度を動的に制御する DVFS という技術が開発され、現在の多くのコンピュータに搭載されている。この技術は性能のクリティカルパス上にないモジュールの動作速度を落とすことにより、性能低下を防ぎつつ消費電力を下げるというものであり、この技術を HPC 領域に応用することによる、電力制約下での性能向上が期待されている。

また、現在のデータセンターやスーパーコンピュータなどの大規模高性能計算システムにおいては、BCM(事業継続マネジメント) の観点から、地震や火事などの災害による停電時にも継続してコンピュータを稼働させられるように自家発電設備や蓄電池が搭載されているケースが多くなってきた。ただ、現状ではそれらの設備はあくまで緊急時のための予備電源としてのみ見なされており、平常時には使用されていない。そのため、それらの新たな電力資源を有効活用して電力対性能を向上させることができると提案されている [2] が、まだこの可能性が示唆されてから日が浅く、未開拓の領域が多く残されている。

そこで本論文では、蓄電池が搭載された高性能計算システムにおいて非停電時にも積極的に蓄電池の充放電を行うことによって、電力制約下での性能向上手法を提案する。HPC 領域において蓄電池を用いた電力制約下における性能向上手法はいまだ提案されておらず、本稿において初めての試みである。

本手法では、Tapasya Patki らの研究 [3] の対象となっているような、厳しい電力制約のた

めに全てのモジュールを常に最高動作速度で動作させることはできないようなシステムを対象とする。まずアプリケーションのテスト実行時のプロファイルデータからアプリケーションの電力対性能グラフの時間推移を予測する。そして消費電力を減らしても性能が下がりにくい部分を見つけて充電し、逆に消費電力を増やすと大きく性能が上がる部分で放電することにより、電力制約下における性能向上を目指す。

以降、2 章では本論文に関する技術や研究を紹介し、3 章では解くべき問題の定義と、提案手法の核となる論理を説明する。4 章では 3 章での手法の有用性を確認するための実験方法について述べる。5 章で実験結果を示し、6 章でその結果について考察した後、7 章で結論と今後の課題を述べる。

第 2 章

研究の背景

本章ではまず提案手法の核となる技術である DVFS、及び DVFS を用いた既存の電力削減手法について説明する。そして、対象とする蓄電池を含んだシステムの電力供給システムについて説明した後、蓄電池と DVFS の両方を用いた電力削減手法の関連研究を紹介する。

2.1 DVFS

2.1.1 DVFS とは

基本的に、プロセッサやメモリはある一定の周波数で動作するように設計されている。動作周波数が高いほど処理能力も高くなるが、同様に消費電力も大きくなる。そのため、プロセッサやメモリを省電力化する最も単純な手法の一つとして、動作周波数を低くするというものがある。かつてのプロセッサやメモリは設計時に決められた一つの動作周波数でしか動作することはできなかったが、現在では一つのプロセッサやメモリが複数の動作周波数をサポートしており、演算中であっても瞬時に動作周波数を切り替えられるようになった。この技術を用いて動的に動作周波数を切り替え、処理速度と消費電力を変化させることによって省電力化を行う手法を DVFS(Dynamic Voltage and Frequency Scaling) と呼ぶ。

図 2.1 に、あるサーバプロセッサにおいて DVFS を用いたときの電力削減のグラフを示す [4]。動作周波数を低くすることにより処理できる最大負荷 (Computer load) は下がるが、電力を削減することもできている。つまり、処理できる負荷であれば低い動作周波数の方が消費電力を少なくすることができる。図 2.1 の DVS savings の推移を見れば分かるように、この例では DVFS を用いることにより最大で 20% ほどの電力削減が行えることになる。

2.1.2 DVFS を用いたコンピュータの既存の省電力化手法

プログラム実行時、メモリやネットワークなどのプロセッサ以外のモジュールがボトルネックとなっているときには、プロセッサはビジーループとなり、処理を行わず電力だけを消費している時間の割合が高くなる。そのためこのような状況ではプロセッサ自体の処理能力を落としてもシステム全体の処理能力はあまり下がらないため、プロセッサを低い動作周波数に切り

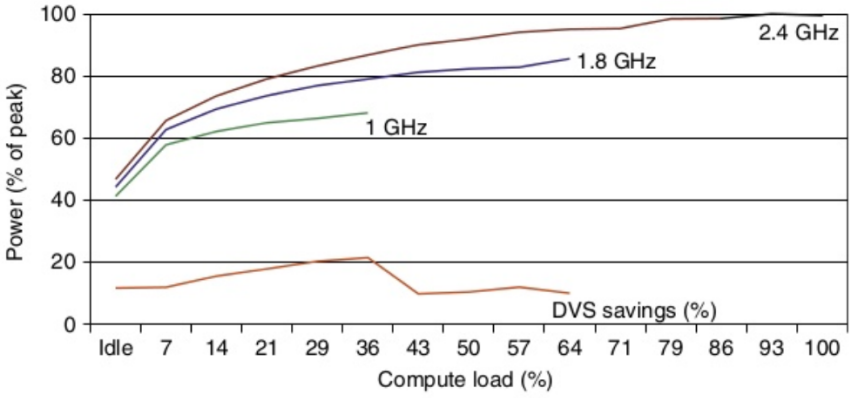

図 2.1. DVFS による電力削減 (AMD Opetron microprocessor) 文献 [4] Figure 1.12 より

替えることで性能低下を防ぎつつ省電力化を行ってきた。

同様に、メモリがボトルネックとなっていない状態ではメモリの動作周波数を落とすことで電力を削減することができる [5]。

また、近年では複数のプロセッサを搭載したマルチプロセッサシステムが増えてきた。マルチプロセッサシステムは複数のプロセッサで並列に処理を行うことで高速化をはかっている。しかし、ひとつずつ順番に処理を行うことが必要なプログラムではひとつのプロセッサのみが処理を行っており、その他のプロセッサはほとんど処理を行っておらず、無駄な消費電力が発生していた。そのような状況では、処理を行っているひとつのプロセッサのみを高い周波数で動作させ、その他のプロセッサの動作周波数を落とすことで消費電力を削減している。

DVFS という技術は登場してからまだ日が浅く、DVFS を用いた電力削減や電力対性能向上の研究は現在盛んに行われている。例えば、2008 年の研究では、複数プロセッサの組み込みシステムにおいてナノ秒単位で DVFS 制御を行うことにより既存の DVFS 制御からさらに 20% もの電力削減が行えるとされている [6]。2011 年の研究によると、メモリのバンド幅の使用率を用いてメモリの DVFS 制御を行うことにより、システム全体のエネルギーの 2.4% を削減できる [5]。これ以外にも多くの研究がなされているが、それでもまだまだ多くの課題が残されているのが現状である。

2.1.3　プロセッサとメモリの DVFS と組み合わせた電力削減の関連研究

一般に、プログラム実行時はプロセッサかメモリのどちらかの処理能力がシステム全体のボトルネックとなっていることが多く、このときボトルネックとなっていないモジュールでは処理能力が必要以上に高い状態となっており、電力が無駄に消費されている。そのため、それぞれのモジュール間での処理能力の差をなくすことが、無駄な電力消費を減らす上で重要である。

この問題を解決するため、プロセッサとメモリの DVFS を同時に用いることによって、そ

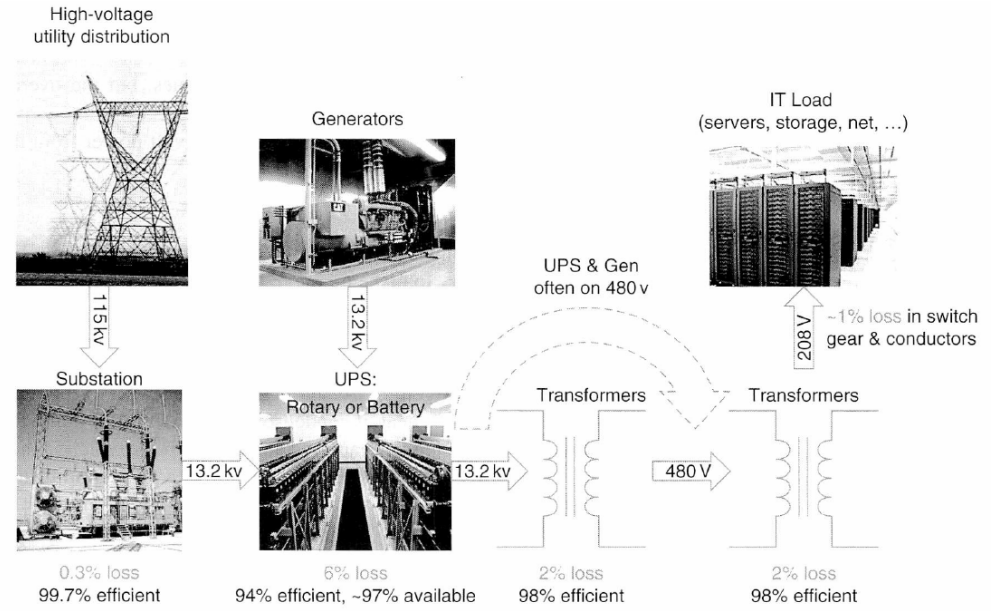

図 2.2. データセンタにおける電源設備 文献 [4] Figure 6.9 より

れぞれの DVFS を別々に行う場合よりもさらに電力対性能の向上を目指した手法が存在する [7]。この手法では、5 ミリ秒おきにプロセッサとメモリの処理能力の両方を監視して、一方のモジュールの処理能力が足りないときにはそのモジュールに電力を融通することによって処理能力の偏りをなくし、与えられた性能制約を満たしつつ省電力化を行っている。

2.2 蓄電池を含む電力供給システム

スーパーコンピュータやデータセンタなどの大規模高性能計算システムにおいては、高い信頼性が要求されるため、一瞬たりとも電圧低下や電力供給停止は許されない。そのため、停電や機器の故障によって電力会社からの電力供給が受けられない時にも、コンピュータへの電力供給を継続するためにいくつかの冗長電源設備が用意されている。現在のデータセンタの一般的な電源設備は図 2.2 のようになっている。

電力会社からの電力供給が停止した場合には、UPS(無停電電源装置) が電力供給を行い、同時に自家発電設備が起動する。数分後、自家発電設備が完全に起動して電力供給が可能になると、自家発電設備から電力供給が行われるようになる。電力会社からの電力供給が再開すると自家発電設備は停止し、電力会社からの電力を使用するようになる。

ここで UPS は 3 つの役割を担っている。一つ目は、コンピュータへの供給電圧を安定させること。二つ目は、停電時に自家発電設備からの電力供給が始まるまでの間、電力を供給すること。三つ目は、停電復帰後に自家発電設備から電力会社に電力供給元を切り替えるとき、一時的に電力供給を行うことである。

UPS 単体がシステム全体に電力を供給し続けられる時間は数分〜30 分程度である場合が多

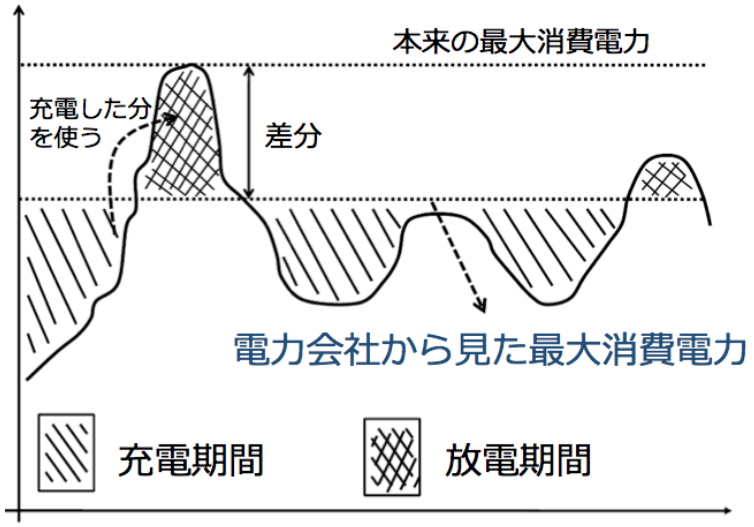

図 2.3. 蓄電池を用いた電力ピークカット手法

い。現在の多くの UPS では電源として蓄電池が使用されているが、充放電が行われるのは基本的に停電時のみであり、今のところ平常時に積極的に充放電を行うような使い方はなされていない。

2.3　データセンタにおける蓄電池を用いたピーク電力削減手法

前節 2.2 で述べたように、今までは平常時に積極的に UPS の蓄電池から充放電を行うことはなかったが、2011 年に発表された論文 [2] において、UPS からの充放電を用いたデータセンタの電力ピークカット手法が提案された。本稿の提案手法と大きく関わる内容であるので、ここで詳しく紹介する。

データセンタにおいてはコンピュータでの消費電力や冷却にかかる電力コストは全体の運用コストの 10〜30% に上り、サービス向上のために電力コストの削減が必要とされている。電力会社との契約料金はピーク時の電力に大きく影響される。そのためピーク電力を削減すべく、この論文では UPS の中の蓄電池を用いた電力ピークカット手法を提案している。

データセンタの 1 日の電力需要の推移は、統計や過去の研究によってある程度予測ができるようになっている。その電力需要曲線から最適な蓄電池の充放電計画を立て、電力会社から引き込む電力の最大値を低く抑えることがこの紹介論文の主旨である (図 2.3)。

紹介論文において対象とする問題をまとめて言葉で表現すると以下のようになる。

- 目的
  - minimize (一日の最大消費電力)
- 与えられる情報
  - 一日の電力推移グラフ
- 制御対象

  - バッテリーをいつ、どれだけ充放電するか
- 制約条件
  - 一日の放電時間・回数
  - バッテリーの残量

この紹介論文の研究以前にも、電力会社からのピーク電力を削減するためにプロセッサの動作速度を変更する手法 [8, 9, 10, 11, 12] や、負荷を時間的もしくは空間的に分散させる手法 [13, 14] が提案されてきた。しかし、これらの手法を適用すると処理速度の低下が必ず起こってしまうことが問題であった。紹介論文における提案手法は、UPS に含まれるバッテリーという既存設備を用いることで、この性能低下を起こさずに電力ピークカットを実現できることを示している。

この手法で実際に用いられているアルゴリズムは図で表現すると図 2.4 のようになり、言葉で表現すると以下のようになる。

1. 一番高いピークが、二番目に高いピークと同じ高さになるように放電を計画 (制約条件を満たせば、次のステップへ)
2. 二番目のピークより高いピーク全てが、三番目に高いピークと同じ高さになるように放電を計画 (制約条件を満たせば、次のステップへ)
3. ・・・(制約条件を満たさなくなるまで繰り返し)

この紹介論文は、他にもバッテリーの電力を使用することによる停電時の信頼性低下や、充放電頻度に対するバッテリーの寿命低下も考慮に入れて充放電計画を立てることによって、データセンタの事業継続性を保ちつつ電力コストを削減できるとしている。

Power Capping の先行研究 [15].

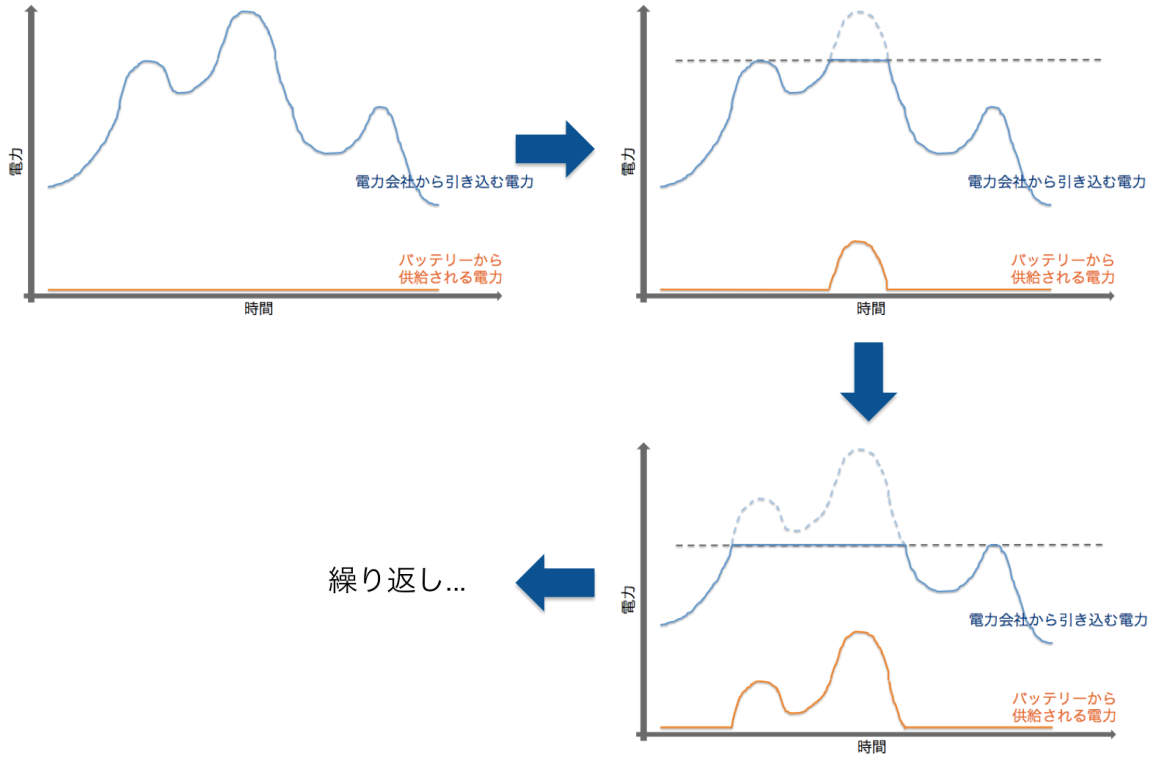

図 2.4. 紹介論文 [2] における UPS を用いた電力ピークカットアルゴリズム

第 3 章

蓄電池を用いた高速化手法

3.1 **フェーズ間の電力融通手法の提案**

3.2 **フェーズの要件**

3.3 **フェーズの求め方**

3.4 **電力融通問題の定式化**

3.5 **電力融通問題の解法**

第 4 章

実験

4.1 実験の目的

4.2 実験方法

4.3 結果

4.4 考察

第 5 章

結論

謝辞

参考文献

[1] Gordon E. Moore. Cramming more components onto integrated circuits. *Electonics*, pages 114–117, April 1965.

[2] Sriram Govindan, Anand Sivasubramaniam, and Bhuvan Urgaonkar. Benefits and limitations of tapping into stored energy for datacenters. *SIGARCH Comput. Archit. News*, 39(3):341–352, June 2011.

[3] Tapasya Patki, David K. Lowenthal, Barry Rountree, Martin Schulz, and Bronis R. de Supinski. Exploring hardware overprovisioning in power-constrained, high performance computing. In *Proceedings of the 27th International ACM Conference on International Conference on Supercomputing*, ICS '13, pages 173–182, New York, NY, USA, 2013. ACM.

[4] John L. Hennessy and David A. Patterson. *Computer Architecture, Fifth Edition: A Quantitative Approach.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 5th edition, 2011.

[5] Howard David, Chris Fallin, Eugene Gorbatov, Ulf R. Hanebutte, and Onur Mutlu. Memory power management via dynamic voltage/frequency scaling. In *Proceedings of the 8th ACM International Conference on Autonomic Computing*, ICAC '11, pages 31–40, New York, NY, USA, 2011. ACM.

[6] Wonyoung Kim, M.S. Gupta, Gu-Yeon Wei, and D. Brooks. System level analysis of fast, per-core dvfs using on-chip switching regulators. In *High Performance Computer Architecture, 2008. HPCA 2008. IEEE 14th International Symposium on*, pages 123–134, 2008.

[7] Qingyuan Deng, D. Meisner, A. Bhattacharjee, T.F. Wenisch, and R. Bianchini. Coscale: Coordinating cpu and memory system dvfs in server systems. In *Microarchitecture (MICRO), 2012 45th Annual IEEE/ACM International Symposium on*, pages 143–154, 2012.

[8] Yiyu Chen, Amitayu Das, Wubi Qin, Anand Sivasubramaniam, Qian Wang, and Natarajan Gautam. Managing server energy and operational costs in hosting centers. In *Proceedings of the 2005 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '05, pages 303–314,

New York, NY, USA, 2005. ACM.

[9] Canturk Isci, Alper Buyuktosunoglu, Chen-Yong Cher, Pradip Bose, and Margaret Martonosi. An analysis of efficient multi-core global power management policies: Maximizing performance for a given power budget. In *Proceedings of the 39th Annual IEEE/ACM International Symposium on Microarchitecture*, MICRO 39, pages 347–358, Washington, DC, USA, 2006. IEEE Computer Society.

[10] Ramya Raghavendra, Parthasarathy Ranganathan, Vanish Talwar, Zhikui Wang, and Xiaoyun Zhu. No "power" struggles: Coordinated multi-level power management for the data center. *SIGARCH Comput. Archit. News*, 36(1):48–59, March 2008.

[11] L. Ramos and R. Bianchini. C-oracle: Predictive thermal management for data centers. In *High Performance Computer Architecture, 2008. HPCA 2008. IEEE 14th International Symposium on*, pages 111–122, Feb 2008.

[12] Xiaorui Wang and Ming Chen. Cluster-level feedback power control for performance optimization. In *High Performance Computer Architecture, 2008. HPCA 2008. IEEE 14th International Symposium on*, pages 101–110, Feb 2008.

[13] L. Ganesh, J. Liu, S. Nath, G. Reeves, and F. Zhao. Unleash stranded power in data centers with rackpacker. In *Proceedings of the Workshop on Energy-Effcient Design (WEED)*, 2009.

[14] Justin Moore, Jeff Chase, Parthasarathy Ranganathan, and Ratnesh Sharma. Making scheduling "cool": Temperature-aware workload placement in data centers. In *Proceedings of the Annual Conference on USENIX Annual Technical Conference*, ATEC '05, pages 5–5, Berkeley, CA, USA, 2005. USENIX Association.

[15] Xiaobo Fan, Wolf-Dietrich Weber, and Luiz Andre Barroso. Power provisioning for a warehouse-sized computer. *SIGARCH Comput. Archit. News*, 35(2):13–23, June 2007.

発表文献

[1] 組込み研究会

16

付録 A