

近藤 正章

電気通信大学大学院情報システム学研究科・准教授

ポストペタスケールシステムのための電力マネージメントフレームワークの開発

§1. 研究実施体制

(1)「システムソフトウェア」グループ

- ① 研究代表者: 近藤 正章 (電気通信大学大学院情報システム学研究科、准教授)
- ② 研究項目
 - ・電力制約適応型システムのための基盤システムソフトウェアの要件調査と初期設計
 - ・電力制約適応型ジョブ管理技術構築に向けた大規模システムの実行ログ解析

(2)「最適化コード生成」グループ

- ① 主たる共同研究者: 三吉 郁夫 (富士通株式会社次世代テクニカルコンピューティング開発本部、マネージャー)
- ② 研究項目
 - ・既存アプリケーションプログラムの挙動傾向調査

(3)「電力性能予測技術」グループ

- ① 主たる共同研究者: 井上 弘士 (九州大学大学院システム情報科学研究院、准教授)
- ② 研究項目
 - ・実機を対象とした電力測定環境の構築
 - ・HPC アプリケーションの電力特性解析
 - ・電力性能挙動予測フレームワークの基本設計

(4)「システムアーキテクチャ」グループ

- ① 主たる共同研究者: 三輪 忍 (東京大学大学院情報理工学系研究科、助教)
- ② 研究項目
 - ・アーキテクチャトレンドの調査

・電力性能の予備調査

§ 2. 研究実施内容

本年度は、既存システムの調査や基礎データの収集、ソフトウェアの初期設計、性能・電力測定環境の構築を目的に据え、各研究グループで以下の研究を実施した。

2.1 システムソフトウェアグループグループ

電力制御の中核を担う基盤ソフトウェアとして、複数ノードに跨った統一的な電力管理を行える電力制御 API や、その指示等に基づきハードウェアの電力性能ノブを制御するミドルウェアが必要となる。その設計と開発に向け、既存 API や OS が持つ電力制御機構として、各種デバイスの電源管理インタフェースの標準である ACPI と Linux の電源管理手法、また Intel Sandy Bridge プロセッサが持つ RAPL について調査した。現状、これらでは多数ノードを持つシステムや複数ユーザ利用を想定した電力制御には対応できていない。将来的に複数ノードを前提とした電力制御 API や、そのための各種ソフトウェア開発が必要になるため、開発ソフトウェアスタックを 1) 電力制御 API ライブラリ、2) ジョブ内電力管理ミドルウェア、3) ジョブスケジューラとして整理し、その機能要件について検討した。

また、大規模システム運用時におけるジョブ間での電力資源配分最適化や、それを実現するジョブスケジューラ開発に向け、九州大学情報基盤研究開発センターのジョブ実行ログと各ジョブの電力使用履歴を入手した。現在、その解析を進めているところである。

2.2 最適化コード生成グループ

研究の第一ステップとして、電力最適化の対象となるアプリケーションプログラムの基本的な電力消費傾向と、電力最適化の際に手段となる既存の電力計測技術について調査を行った。

アプリの電力消費傾向調査では、GPU を使用しない x86 クラス上で CPU コアやメモリの消費電力変動幅(メモリは推定値)を観測した。HPC Challenge Benchmark を用いた観測では、DGEMM 計測区間の消費電力が目立って高く、この条件での消費電力がシステム全体の最大消費電力に直結することが確認された。また、STREAM や RandomAccess、FFT 等の計測区間でそれぞれに電力消費傾向が異なることも確認された。SPEC CPU2006 を用いた観測では、各ベンチマークについて消費電力の平均値と最大値を調査したところ、平均して 1.45 倍の差が計測された。これらより、アプリの種類に応じて電力消費を最適化する余地があることが確認された。

電力計測技術の調査では、Intel 社の Sandy Bridge プロセッサに搭載されている電力計測および制御のための RAPL を対象に、アプリケーションプログラムから当該機能を利用することについて検討を進めた。その結果、オープンソースソフトウェア PAPI を用いることによって、アプリケーションプログラムから走行中の消費電力変動を遅滞なく計測できることを確認した。また、性能分析ツール TAU 上でも、電力計測を行えるように機能拡張が可能であることを確認した。

今後の課題としては、アプリケーションプログラムで観測された電力消費状況を可視化し、アプリケーション開発者が電力最適化の着眼点を理解／検討しやすくする手法を具体化すると共に、実際にプログラムへ電力消費の制御ポイントを埋め込んでいく技術の開発が必要である。

2.3 電力性能予測技術グループ

まず、HPC アプリケーションの電力性能特性を解析するため、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を対象とした消費電力測定環境を構築した。これにより、1)システム全系ならびに複数計算ノード単位での消費電力を実測、2)各計算ノードにおけるプロセッサと主記憶(DRAM)の消費電力値の取得、が可能となった。具体的には、インテル社が提供するRAPLをベースとし、アプリケーションコード中に指示文を挿入することで、周期的に各消費電力値を取得するAPIを開発した。また、測定結果を自動でグラフ化するための環境を整備した。また、実際のHPCアプリケーションとして電磁流体シミュレーションならびにフラグメント分子軌道法計算を対象とし、消費電力特性の解析を行った。その結果、電磁流体シミュレーションにおいては、消費電力バジェットを十分に活用できていない場合や、消費電力キャッピングを施した場合でも大幅な性能低下が見られない場合が存在することが明らかになった。これらの解析結果は、現在実施している「消費電力制約下での性能チューニング」に活用することができる。

次に、既存のインターコネクト・シミュレータをベースとした電力性能挙動予測フレームワークの基本設計を行った。具体的には、インターコネクト・シミュレータにおいて、プログラムコード中の計算を行う部分を抽象化し、実行時間モデルに置き換えることで、通信のみならず計算による時間経過も反映可能とするシミュレーション方式を決定した。今後は、1)インターコネクト・シミュレータでの消費電力推定機能の実装、2)インターコネクト・シミュレータをベースとしたシステム性能推定ツールの開発、3)当該ツールにおける消費電力推定機能のサポート、を実施する予定である。

2.4 システムアーキテクチャグループ

電力制約適応型システムのあるべきシステムアーキテクチャの提示と、アプリの特徴と電力制約に応じた制御による電力性能の大幅な改善への第一歩として、1)アーキテクチャトレンドの調査、2)電力性能の予備調査を行った。

アーキテクチャトレンドの調査として、本プロジェクトの根幹をなす電力性能ノブについて現状調査と将来性の検討を行った。現在利用可能なノブには、パワー・ゲーティング、DVFS、クロック・ゲーティングなどがある。ただし、これらすべてがシステムソフトウェアから利用できるというわけではなく、特にメモリやネットワークについては、現状ハードウェア制御であり、ソフトウェアによるモード遷移手段は用意されていない。今後は、ソフトウェアによるモード遷移のあり方を検討していく必要がある。また、ネットワークの省電力モードであるLPI(Low Power Idle)は、スパコンのインターコネクトにはまだ取り入られていないが、将来実現可能性が高いノブとして今後も検討していく予定である。

電力性能の予備調査では、制御アルゴリズム検討の初期段階として、既存ハードウェア環境下で、電力制約とアプリの特徴に応じて制御を行った際の電力性能を評価した。特に、CPUとGPUへのタスク配分と、各デバイスの動作周波数と電源電圧をどのように調整すれば電力性能を最大化できるか調査した。評価の結果、CPUのみの場合と比べ、オフロード率と周波数を最適化することで最大3倍の電力性能の向上が達成できることを確認した。