

戦略的創造研究推進事業 CREST  
研究領域「ポストペタスケール高性能計算に資する  
システムソフトウェア技術の創出」  
研究課題「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」

## 研究終了報告書

研究期間 平成24年10月～平成30年 3月

研究代表者:近藤 正章  
(東京大学大学院情報理工学系研究  
科、准教授)

## § 1 研究実施の概要

### (1) 実施概要

限られた電力資源を各アプリケーションに、また各ハードウェア要素に適応的に配分することで性能や電力効率を向上させる「電力制約適応型システム」をコンセプトに、電力資源を真に有効利用できる電力マネージメントフレームワークの開発を目的として研究を行ってきた。システムソフトウェア(東大)グループがジョブ内・ジョブ間電力制御のための基盤システムソフトウェアを、最適化コード生成(富士通)グループがノブ制御最適化コード生成技術を、電力性能予測技術(九大)グループがプロファイリングやシミュレーション等により大規模システムの電力性能特性を予測する技術を、システムアーキテクチャ(電通大)グループが将来的なノブを含めたノブ制御アルゴリズムを担当し、研究開発を実施した。特に、電力制約下のアプリケーション実行性能最適化、複数ジョブ間の電力資源管理ツールの開発、電力性能挙動予測環境構築、電力観測・制御 API 検討の4つを研究の柱として位置づけ、4研究グループが密接に協力しつつ、電力マネージメントフレームワークの構築を目指してきた。柱毎に中核となるシステムソフトウェアやツール等の開発を推進し、それらのソフトウェアを公開すること、さらに従来設計のシステムと比べ、最大利用可能電力に対する実効性能比を、最大で5倍程度に向上させることを目標として研究開発を実施してきた。

電力制約下のアプリケーション性能最適化として、各ジョブの電力配分最適化プロセスを自動化する電力性能最適化パッケージを、九大グループで開発したノブ制御ライブラリをベースに、東大グループが中心となり開発・公開した。これは、1)オリジナルコードにプロファイル/電力ノブ制御用 API 呼び出しを挿入するためのコード解析、2)得られたコードによるプロファイリング、3)その情報を利用し電力分配最適化アルゴリズムにより各制御ポイントでの電力配分決定、4)電力配分に従い API により電力ノブを制御しつつ実行を行うものである。現在は富士通グループの動的最適化手法や、電通大グループで開発したいくつかのノブ制御アルゴリズムを統合しツールを拡張中である。簡単な最適化例では、最大で約2倍の性能向上が得られることを確認している。

電力資源管理を行うスケジューラとして、資源管理ツールとして広く利用されている Slurm をベースに電力使用状況やアプリケーション特性に応じた電力資源管理を行うためのスケジューラを東大グループが中心となり開発・公開した。種々の電力資源管理・電力制約適応型ジョブスケジューリングが実装できるようにプラグインベースのインタフェースを実装した。ジョブの要求に応じて動的に電力予算を管理する技術や、空調の電力も含めた電力制約下でのジョブ間の電力配分最適化と、それに伴ったジョブスケジューリング手法も開発した。従来のスケジューリングに比べてシステム全体のジョブ実行スループットを1.4倍に向上することに成功した。

電力性能予測ツールに関しては、性能ばらつきを考慮した電力制約下での性能推定フレームワークと、インターコネクトの電力性能をシミュレートするためのツールを開発した。特にインターコネクト・シミュレータに関しては、米国ローレンス・リバモア国立研究所(LLNL)で開発中の TraceR へと移植中であり、今後当該ツールを公開予定である。

ノブ制御最適化アルゴリズムに関しては、九大グループで大規模グラフ解析ベンチマークの Graph500 を対象に CPU と DRAM の電力配分最適化の検討を行ったほか、電通大グループではインターコネクトの余剰電力を CPU に配分し性能向上を図る手法や、主記憶デバイスである DIMM の待機電力削減のためのメモリホットプラグ機能を将来的な電力ノブとして DIMM の動的オン・オフ制御に応用することを検討した。

各グループの研究成果を利用することで、ジョブ内のノブ制御最適化により、電力制約下での性能を1.2～5.4倍程度向上させることに成功した。また、最適化パッケージや電力資源管理機能付き Slurm のプロトタイプを公開したほか、様々な関連ツールを公開予定である。

研究成果を広く展開するべく、他の研究グループと積極的に連携を図りつつ研究を実施した。グラフアプリケーションの電力性能最適化を同 CREST の藤澤チームや遠藤チームと連携して行った。また、電力資源管理機能付きの Slurm 開発やスケジューリングアルゴリズム、インターコネクト・シミュレータの開発は LLNL との国際連携体制のもとで実施した成果である。

## (2) 顕著な成果

### <優れた基礎研究としての成果>

#### 1.

##### 概要:

CPU には製造時の消費電力特性にばらつきがあり、電力制約下では同一ハードウェア構成にて同一アプリケーションを実行した場合でも、計算ノード間に電力性能特性ばらつきが生じる。このばらつきを考慮した電力制御の必要性を世界で初めて指摘し、これを考慮した電力性能制御アルゴリズムを開発した。これにより、均一に電力制約を適用する従来法に比べて電力制約下での性能を最大 5.4 倍向上させることに成功した。この成果は、高性能計算分野における最高峰の国際会議の一つである SC2015 に採択された。

#### 2.

##### 概要:

システム全体の電力供給と使用状況や、アプリケーションの特性に応じたジョブスケジューリング、ジョブ間電力分配を行う電力管理機能付きの Slurm リソースマネージャを開発し、世界で初めて 1000 ノード規模の大規模システムでオーバプロビジョンドシステムの性能を解析した他、いくつかのスケジューリングアルゴリズムを開発して評価を行った。これらの成果は、高性能計算分野で歴史のある国際会議である CCGrid2016 や IPDPS2017 に採択された。

#### 3.

##### 概要:

アプリケーション実行時の CPU/主記憶/ネットワークなど各ハードウェア要素の電力余剰を検出し、より電力を効率的に利用できるハードウェアへと電力を分配するパワーシフティング技術を開発した。特にスケルトンコードを用いてネットワークの電力使用状況を高速プロファイリング(通常のプロファイリング方法と比較して最大 11.5 倍高速化)し、それをもとに CPU へと電力を分配する手法では、最大で 1.23 倍の性能向上が得られることがわかっている。本成果も、高性能計算分野の最高峰の国際会議の一つである SC2015 に採択された。

### <科学技術イノベーションに大きく寄与する成果>

#### 1.

##### 概要:

複数ジョブが存在する高性能計算機環境では、システム全体の電力供給と使用状況や、アプリケーションの電力性能特性に応じたジョブスケジューリング、またジョブ間での電力分配など、ハードウェア資源と同様に電力資源をも管理するシステムソフトウェアが必須となる。そこで、世界でも最も利用されているリソースマネージャの一つである Slurm を拡張する形で電力管理ツールを実装中であり、プロトタイプ版を公開した。

#### 2.

##### 概要:

本研究で得られた成果を国際的に展開するべく、米国ローレンス・リバモア国立研究所やアリゾナ大学との国際連携体制を構築し、積極的に技術交流や議論をしながら研究を遂行している。計算ノード間の電力性能特性ばらつきを考慮した電力性能最適化手法、インターコネクトの電力性能推定ツール、電力管理機能付き Slurm を用いたスケジューリングアルゴリズムの性能評価などに関して協力して研究を実施した。共同で論文も執筆し、SC2015 や IPDPS2017 に採択された実績を有する。

#### 3.

##### 概要:

本研究プロジェクトで得られた知見や成果を、ポスト京の開発へと展開するべく、「ポスト京プロジェクト」の電力制御ワーキンググループへ参加し、電力制約下でのアプリケーション最適化やシステムの電力管理、電力制御 API などに関して意見交換や知見・アイデアの提供を行っている。

## § 2 研究実施体制

### (1) 研究チームの体制について

#### ① 「システムソフトウェア」グループ

研究参加者

| 氏名    | 所属                      | 役職   | 参加時期         |
|-------|-------------------------|------|--------------|
| 近藤 正章 | 東京大学大学院情報理工学系研究科        | 准教授  | H24.10～      |
| 中田 尚  | 同上                      | 助教   | H27.3～H28.9  |
| 有間 英志 | 同上                      | 特任助教 | H28.4～       |
| 中村 宏  | 同上                      | 教授   | H27.3～       |
| 和田 康孝 | 明星大学                    | 准教授  | H24.10～      |
| 本多 弘樹 | 電気通信大学大学院<br>情報システム学研究科 | 教授   | H24.10～H27.2 |
| 齋藤 翔太 | 同上                      | M2   | H24.10～H25.3 |
| 橋本 崇浩 | 同上                      | M2   | H24.10～H25.3 |
| 宮部 創一 | 同上                      | M2   | H24.10～H26.3 |
| 池田 賢祐 | 同上                      | M1～2 | H24.10～H26.3 |
| 松本 洋平 | 同上                      | M1～2 | H24.10～H26.3 |
| 佐々木 沢 | 同上                      | M1～2 | H25.4～H27.3  |
| 藤原 祐太 | 同上                      | M1～2 | H25.4～H27.3  |
| 松村 正隆 | 同上                      | M1～2 | H25.4～H27.3  |
| 斎藤 純  | 同上                      | M1～2 | H25.12～H27.3 |
| 林 信吾  | 同上                      | M2   | H27.3～H27.3  |
| 高田 遼  | 同上                      | M1～2 | H27.4～H29.3  |
| 東 耕平  | 同上                      | M1～2 | H28.4～       |

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開発
- ・ 電力制約適応型動的ノブ調整ミドルウェアの開発
- ・ 電力制約適応型ジョブ管理技術

#### ② 「最適化コード生成」グループ

研究参加者

| 氏名    | 所属                           | 役職     | 参加時期    |
|-------|------------------------------|--------|---------|
| 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 | 部長     | H24.10～ |
| 三輪 英樹 | 同上                           | マネージャー | H24.10～ |
| 向井 優太 | 同上                           | -      | H26.4～  |
| 竹内 剛  | 同上                           | -      | H29.4～  |

研究項目

- ・ 電力制約適応型システム向けのアプリケーション特徴抽出
- ・ 電力性能ノブ制御最適化コード生成技術

### ③「電力性能予測技術」グループ

#### 研究参加者

| 氏名     | 所属                     | 役職    | 参加時期         |
|--------|------------------------|-------|--------------|
| 井上 弘士  | 九州大学大学院システム<br>情報科学研究院 | 教授    | H24.10～      |
| 小野 貴継  | 同上                     | 助教    | H27.4～       |
| 佐々木 広  | 同上                     | 学術研究員 | H25.3～H26.4  |
| 深沢 圭一郎 | 京都大学学術情報メデ<br>ィアセンター   | 准教授   | H24.10～      |
| 上田 将嗣  | 九州大学情報基盤課              | 主任    | H24.10～      |
| 花田 高彬  | 九州大学大学院システム<br>情報科学研究院 | D2～3  | H24.10～H26.3 |
| 上野 伸也  | 同上                     | D2～3  | H24.10～H26.3 |
| 石田 浩貴  | 同上                     | M1～2  | H28.4～       |

#### 研究項目

- ・ 電力性能挙動プロファイリング技術とプロファイルに基づくプログラムチューニング
- ・ 電力性能挙動予測シミュレータ開発／見積もり技術開発
- ・ 統合評価環境作成と評価

### ④「システムアーキテクチャ」グループ

#### 研究参加者

| 氏名     | 所属                      | 役職    | 参加時期         |
|--------|-------------------------|-------|--------------|
| 三輪 忍   | 電気通信大学大学院<br>情報理工学研究科   | 准教授   | H24.10～      |
| 本多 弘樹  | 同上                      | 教授    | H27.3～       |
| 八巻 隼人  | 同上                      | 助教    | H28.4～       |
| 中村 宏   | 東京大学大学院情報理<br>工学系研究科    | 教授    | H24.10～H27.2 |
| 中田 尚   | 同上                      | 助教    | H24.10～H27.2 |
| 金 均東   | 同上                      | D3    | H24.10～H25.3 |
| 和 遠    | 同上                      | D3    | H24.10～H26.3 |
| 薦田 登志矢 | 同上                      | D2～3  | H24.10～H26.3 |
| 有間 英志  | 同上                      | M2～D2 | H24.10～H27.2 |
| 岩澤 直弘  | 同上                      | M2    | H24.10～H25.3 |
| 岡本 和也  | 同上                      | M2    | H24.10～H25.3 |
| 井上 聖等  | 同上                      | M1～2  | H24.10～H25.9 |
| 會田 翔   | 同上                      | M1～2  | H24.10～H26.3 |
| 重松 拓也  | 同上                      | M1～2  | H24.10～H26.3 |
| 林 信吾   | 同上                      | M1～2  | H25.4～H27.2  |
| 田中 維人  | 同上                      | M1～2  | H25.4～H27.2  |
| 早川 耕太  | 同上                      | M1    | H25.4～H25.9  |
| 石川 雄介  | 電気通信大学大学院<br>情報システム学研究科 | M1～2  | H27.3～H28.3  |
| 岩崎 一輝  | 同上                      | M1～2  | H27.3～H28.3  |

|       |                       |      |             |
|-------|-----------------------|------|-------------|
| 藤田 匠  | 同上                    | M1～2 | H27.3～H28.3 |
| 大場 百香 | 同上                    | M1～2 | H27.4～H29.3 |
| 石原 雄也 | 同上                    | M1～2 | H27.4～H29.3 |
| 澁谷 俊憲 | 同上                    | M1～2 | H27.4～H29.3 |
| 山添 高広 | 同上                    | D1～3 | H27.4～      |
| 吉田 昌弘 | 同上                    | M1～2 | H27.4～      |
| 西郷 雄斗 | 同上                    | M1～2 | H27.10～     |
| 愛甲 達也 | 電気通信大学大学院<br>情報理工学研究科 | M1～2 | H28.4～      |
| 高德 真晴 | 同上                    | M1～2 | H28.4～      |
| 松井 優樹 | 同上                    | M1～2 | H28.4～      |
| 松尾 駿  | 同上                    | M1～2 | H28.4～      |
| 三須 雅仁 | 同上                    | M1～2 | H28.4～      |

#### 研究項目

- ・ 電力性能ノブの探索
- ・ 電力性能ノブのモデリング
- ・ 電力性能ノブ制御アルゴリズムの検討

#### (2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

- ① 東京工業大学学術国際情報センター
  - ・ 連携内容：電力モニタリング・制御 API を連携して検討した。
- ② 米国ローレンス・リバモア国立研究所
  - ・ 連携内容：電力モニタリング・制御 API、電力性能挙動プロファイリングに基づくプログラムチューニング、消費電力可視化技術に関して連携して検討・評価を行った。
- ③ 米国アリゾナ大学
  - ・ 連携内容：電力性能挙動プロファイリングに基づくプログラムチューニング技術に関して連携して検討・評価を行った。
- ④ 当 CREST 領域の藤澤チーム／遠藤チーム
  - ・ 連携内容：大規模グラフアプリケーションを対象にした電力性能最適化手法の開発を行い、ソースコードレベルのチューニングが電力性能に与える影響の評価や解析を行った。

### § 3 研究実施内容及び成果

#### 3. 1 システムソフトウェア(東京大学 近藤グループ)

##### (1)研究実施内容及び成果

##### ① 研究実施内容

電力制約適応型システムでは、必要なハードウェア資源に必要な分だけ電力資源を割り当てるための最適化が必要であり、アプリケーションの特徴に応じて電力資源をスケジューリングするための各種システムソフトウェア、電力性能最適化を意識したアプリケーションプログラムの開発支援、電力観測・制御 API、システム全体の電力供給や使用状況などの運用状況に応じた電力資源スケジューリングが必須となる。そこで、システムが持つ電力性能ノブを適切に抽象化しつつ、プログラムコードからの指示等に基づきノブを制御するための基盤システムソフトウェア、および電力制約に動的に適応できる電力制御ミドルウェア、電力供給の変動に追従しつつタスクの優先度等と電力資源量に応じたジョブ管理を行う技術等を開発し、標準的な電力観測・制御 API を意識したシステムソフトウェアを設計した。具体的には、電力性能最適化のためのアプリケーションプログラムの開発支援を行いつつ、自動・半自動でノブ制御までを行う最適化パッケージを開発しフリーソフトウェアとして公開した。また、電力資源量に応じてジョブスケジューリングやジョブ間の電力配分を行うスケジューラを、世界的にも広く利用されている Slurm リソース管理ツールの拡張モジュールとして開発し、フリーソフトウェアとして公開した。後者に関しては、世界で初めて 1000 ノード規模の大規模システムに開発したリソース管理ツールをインストールし、オーバプロビジョンドシステムの性能を解析した他、いくつかのスケジューリング戦略を開発して評価を行い、その有効性を示した(CCGrid2016、IPDPS2017 にて成果を発表)。

##### ②研究実施方法

アプリケーション内での電力性能ノブ最適化技術として、アプリケーション開発支援のためのアプリケーション特徴抽出技術と、自動・半自動で最適化コード生成を行う技術、そして基盤ソフトウェアとして各種ツールやライブラリを統合した最適化パッケージの開発を行った。また、システムの電力資源管理として、ジョブ管理技術と動的ノブ調整技術、そしてそれらを実装するための電力管理機能付きリソースマネージャの開発を行った。以降、各項目に整理しつつそれらの研究実施方法について述べる。

アプリケーションの特徴抽出技術は、既存の性能解析ツールとして広く利用されている Tuning and Analysis Utilities (TAU) を拡張して開発を行った。最適化コード生成のための基盤ソフトウェアについては、TAU/PDT (Program Database Toolkit) を利用し、電力性能ノブの制御ポイントを半自動的なコードの書き換えにより埋め込み、プロファイリング情報を利用しつつノブ制御を最適化する技術について開発した。これらと、電力性能挙動予測技術(九大)グループが開発した Intel 社が提供する RAPL (Running Average Power Limit)を使用した電力ノブ制御ライブラリ等をパッケージ化することで、電力性能ノブ最適化を行うための基盤ソフトウェアである電力性能最適化パッケージを開発した。本パッケージには静的・動的にノブ制御最適化を行うためのアルゴリズム等を統合するため、最適化コード生成(富士通)グループ、電力性能挙動予測技術(九大)グループ、システムアーキテクチャ(電通大)グループと協力しつつ開発を行った。

動的ノブ調整技術・ジョブ管理技術の開発では、複数ジョブ間の電力資源管理ツールとして、システムの電力制約や各ジョブの電力需要に応じてジョブ間での電力配分を調整する動的ノブ制御ミドルウェアと、電力資源量に応じてジョブスケジューリングを行うスケジューラを開発した。世界的にも広く利用されている Slurm リソース管理ツールへの拡張として実装することで、ジョブスケジューラ・電力資源配分管理を含めた汎用の資源管理ツールとして公開することを目指す。

なお、これらの研究の実施には、将来の大規模計算機センターでの運用ポリシーを考慮したシステムソフトウェア開発を行うために、九州大学情報基盤研究開発センターと連携するほか、電力観測・制御用 API に関しては高電力効率スパコン TSUBAME の開発・運用経

験を持つ東京工業大学学術国際情報センターと連携し、また、本プロジェクトで開発する各種ソフトウェアを広く普及させるために米国ローレンス・リバモア国立研究所との国際連携体制を構築し、定期的に議論を行いつつ研究を実施した。

### ③ 研究成果とその位置づけ

#### 1) アプリケーションの特徴抽出技術

自動最適化は有用である反面、ユーザ自身が電力を考慮してアプリケーションのファインチューニングを行う場合も多い。電力性能最適化では最大消費電力と実行時間のトレードオフを考える必要があり、時間方向の平均値を見るだけでは不十分である。また電力性能を最適化するためには、アプリケーションプログラムのどの部分をどう処置すべきかが重要である。そのため、消費電力を単に時系列分析するだけでなく、アプリケーションプログラムの実行位置と対応付けが同時に必要である。その補助の一つとして、プロファイリングツールにおいて通常のパフォーマンスカウンタ等の情報に加え、消費電力に関する情報も併せて取得し、可視化してプログラマに提示する環境を構築した。既存の性能解析ツールとして広く利用されている TAU をベースに検討したところ、まずは消費電力情報取得機能について、観測オーバーヘッドと電力測定の精度が課題だとわかったため、それらについて改良を行った。その結果、TAU オリジナルの実装に比べて低オーバーヘッドで消費電力情報を記録できるようになったが、それでもプロファイリングを行わない場合と比較して 10～20%程度実行時間が増加する。パフォーマンスカウンタなどの性能情報は測定対象区間ののみが取得できるので、実行時間の増加は大きな問題とならないが、消費電力は単位時間毎の電力値を取得するため、実行時間の変動は電力値の精度に大きく影響する。そこで、電力情報の精度を高めるべく、RAPL (Running Average Power Limit)を用いた低オーバーヘッドな電力取得環境と組み合わせる方式を開発した。これにより、Open Trace Format Version2 (OTF2) 形式で性能情報と共に高精度で電力情報を保存することができ、既存の可視化ツール等を利用してユーザが電力最適化を行う際の補助となる環境を開発することができた。

#### 2) 基盤システムソフトウェア(最適化パッケージ)の開発

電力制約適応型システムにおいて、電力性能ノブを適切に制御しつつ電力制約下で性能を最大化することが重要となるが、一般のユーザがそのための最適化を行うのは容易ではない。そのための基盤システムソフトウェアとして、自動・半自動で電力性能ノブの制御ポイント埋め込み、プロファイリング、ノブ制御最適化を行うアプリケーション内電力性能ノブ最適化ツールを開発した。

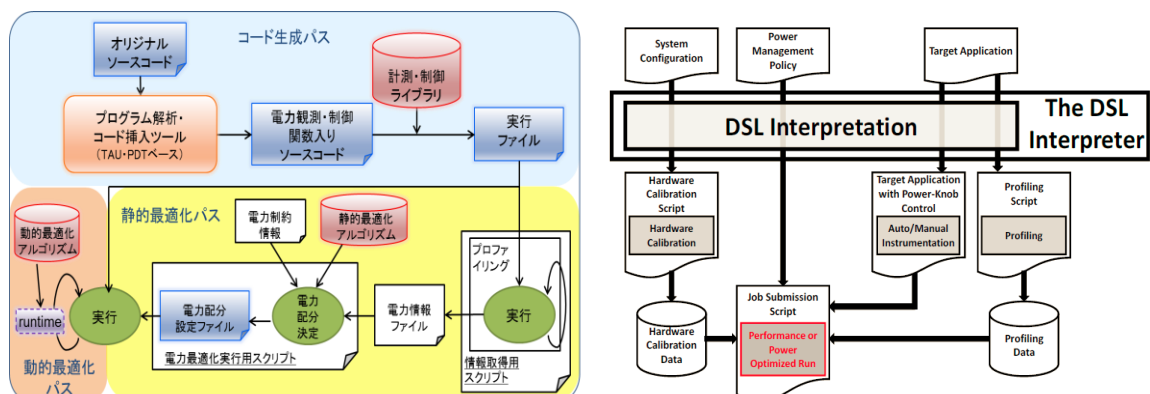


図 1: 電力性能最適化パッケージの概要(左)と DSL によるコード変換フロー(右)

本ツールは、電力性能ノブの制御ポイント埋め込みは TAU/PDT の機能を利用して元のプログラムに対しコード解析を行い、電力ノブを制御する可能性のあるポイント

に API 呼び出しの関数埋め込みを行う。電力性能予測技術グループで開発された電力観測・制御ライブラリ等を統合し、プロファイリングを行うことで、各制御ポイントの電力ノブ設定を決定し、最終的に電力ノブ制御が最適化されたコードを実行する。これらの機能を統合したものを最適化パッケージとして公開することを目的に開発を行ってきた。

図 1(左)に本パッケージの概要を示す。本パッケージは以下の手順で自動・半自動でアプリケーションコードの最適化を行うものである。

1. オリジナルなアプリケーションプログラムに対し、コード解析して電力ノブを制御する可能性のあるポイントに API 呼び出しの関数を埋め込む。
2. 複数ノブ設定下でプロファイリング実行を行い、API を利用して電力性能情報を取得する。
3. 得られた情報をもとに、電力分配最適化アルゴリズムにより各制御ポイントでの電力分配を決定し、それを設定ファイルとして保存する。
4. 実行時には、電力分配設定ファイルを参照し、1. で埋め込まれた API により電力ノブを制御しつつ実行を行う。

本パッケージ上に、電力性能予測技術グループで開発された最適化アルゴリズムを例として実装し、機能および有用性の検証を行った。

電力分配アルゴリズムは様々なものが考えられ、システム毎、さらにはユーザ毎に運用ポリシーやアプリケーションの特徴に合わせて異なるアルゴリズムを用いて最適化を行うべきと考えられる。そこで、本フレームワークの汎用性を高めるべく、上記 3. の電力分配最適化アルゴリズムや、コード埋め込みからプロファイリング、最適化コードの実行を含む最適化フロー制御を記述するための Domain Specific Language (DSL) を定義し、様々なシステム・アルゴリズムに対応できるようにフレームワークを拡張した。図 1(右)に DSL によるコード変換フローを示す。

DSL への入力のうち、システム仕様はシステム管理者より、またアプリケーションの特徴情報はユーザより与えられると仮定する。最適化のためのポリシー（図中の Power Management Policy）として、利用情報の選択や、電力モデル・電力配分の手続きが DSL により記述され、アプリケーション情報・プロファイリング結果をもとに実際に電力ノブ制御の設定が決定される。最終的に、アプリケーションコード、あるいはジョブスクリプトへとその電力ノブ設定が埋め込まれる。

電力分配決定のケーススタディとして、アプリケーションの実効ピーク電力や平均電力を推定して電力キャップを設定するケース、ユーザの指定する許容性能低下率を満たすように電力キャップを設定するケースを想定して、DSL を用いて配分アルゴリズムを指定した。本フレームワークにより自動で電力ノブ設定のコードを生成し、16 ノードの計算機システム上で電力評価を行った結果、所望の電力設定でアプリケーションが実行されることを確認している。

様々なシステム上で恒久的に利用可能な電力制御を行うシステムソフトウェアを開発するためには、電力観測・制御のための API 標準化は重要な課題である。これまで、東京工業大学学術国際情報センターと連携して、システム全体の電力制御に必要な API の検討を行ってきたが、米国サンディア国立研究所(SNL)が標準 API である PowerAPI を公開し、現在は徐々に普及し始めている。そこで、本最適化パッケージの出力コードやシステム仕様の指定に PowerAPI を用いることで、本フレームワークの汎用性をより高めることを目指し、開発を継続中である。なお、本パッケージのプロトタイプ版は GitHub で公開している。

### 3) 動的ノブ調整技術、および 5) ジョブ管理技術

電力制約適応型システムにおける、リアルタイム電力監視・動的ノブ制御を行うミドルウェアの開発を行った。ミドルウェアに必要な要件調査の結果、各ノードで動作するアプリケーションの電力消費状況は変動するため、その変動に適切に対応できること、外部要因で電

力制約が変化した場合にも適切に制御できること、運用ポリシーに依存して電力制約による性能低下の影響を抑えるべき高優先度のジョブが存在すること、などがミドルウェア開発上の要件と考えられることがわかった。そこで、これらの要求に対応すべく、性能低下の許容範囲をジョブ毎に設定し、ジョブの特性に応じて電力キャップを動的に決定する適応的電力キャッピング手法と、キャッピングの結果余剰となった電力で他のジョブを動作させるためのミドルウェアを開発した。32 ノードのシステムで開発したミドルウェアを動作させたところ、各ジョブの性能低下率を許容範囲に留め、かつシステム全体の電力制約を満たしつつ、より多くのジョブを動じに実行できることがわかった。本ミドルウェアにより、最適化なし(No Cap)に比べ、動的制御(Adaptive)により、ジョブの平均ターンアラウンドタイムを約 16.27%まで短縮できるという評価結果を得ている(図 2)。なお、本成果は CCgrid2016 で発表した。

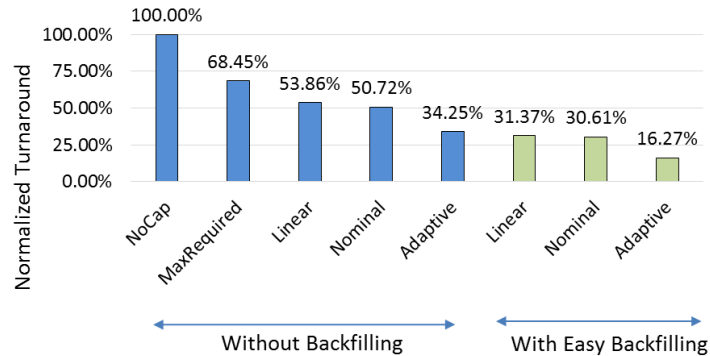


図 2: 平均ターンアラウンドタイム

#### 4) ジョブ管理技術と電力資源管理ツール

ポストペタスケールシステムが設置されるような計算機センターでは、システムを分割して複数ジョブが同時に実行されることが通常である。そのような複数ジョブが存在する環境では、システム全体の電力供給と使用状況やアプリケーションの優先度に応じた動的電力スケジューリングを行うことが必要である。そのためのミドルウェア・システムソフトウェアを、クラスタシステムの資源管理ツールとして広く利用されている Slurm リソースマネージャをベースに開発し、そのプロトタイプ版をオープンソフトウェアとして公開している (<https://github.com/pompp>)。

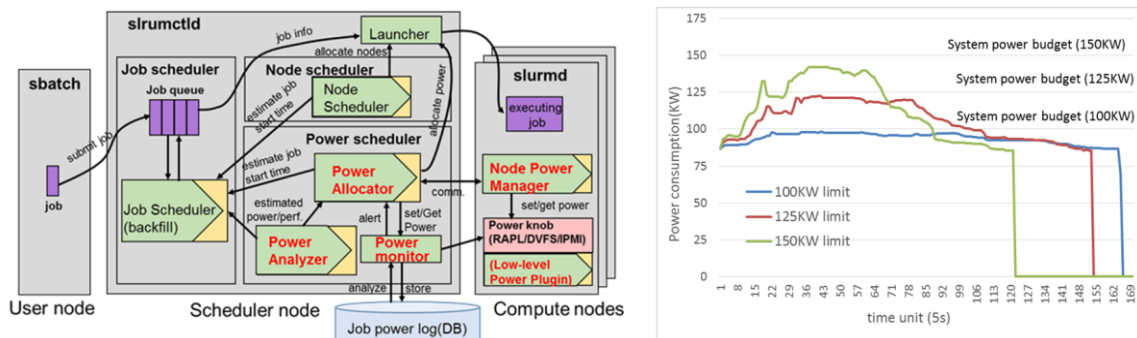


図 3: 開発した Slurm ベースの資源管理ツール(左)と電力制約設定時の実効電力

図 3 に開発した電力制御機能付き Slurm の概要を示す。オリジナルの Slurm に対してジョブの必要電力や制約下での性能を予測するモジュール(Power Analyzer)、ジョブの最適な電力割り当て決定モジュール(Power Allocator)、各ノードの電力収集と電力制約の設定モジュール(Power Monitor)、各ノードで動作し実際に電力のモニタリングや電力ノブの制御を行うモジュール(Node Power Manager)を追加している。これらはプラグインベースのインタフェースを持ち、機能拡張が容易になるよ

う設計を行った。本ツールにより、実際に電力制約内でジョブがスケジューリングできていることを確認した(図3右)。

開発した電力資源管理機能付き Slurm を九州大学情報基盤研究開発センターの HA8000 システムにインストールし、電力制約を考慮して資源管理を行った際の電力・性能特性について解析した。具体的には、実際の HPC システムを用いて様々なジョブセットを実行した場合の性能や実行中の消費電力について評価した。

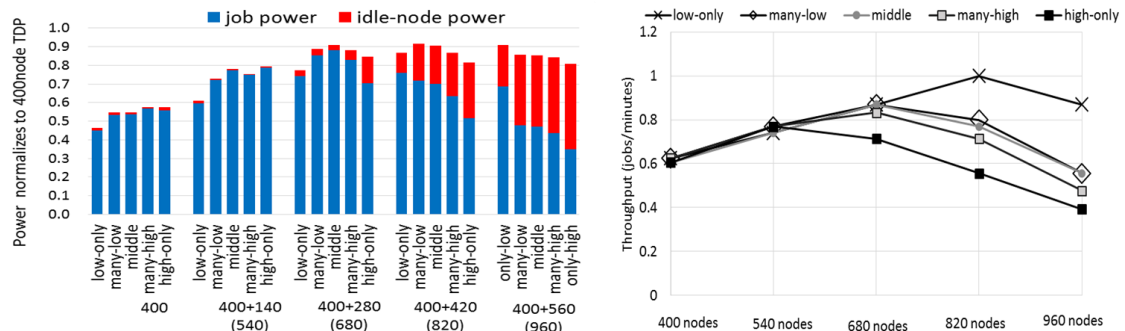


図4: 電力使用率(左)とシステムスループットの評価結果

評価においては、様々なジョブの電力特性をエミュレーションすることができるよう、行列計算のサイズを変更してキャッシュミス率を調整可能なジョブセットを作成した。本ジョブを利用して、HA8000 の 965 ノードを用いて消費エネルギーや平均消費電力、ジョブの実行時間等の評価を行った。評価ではオーバプロビジョニング環境を想定し、システム全体の電力制約値は 400 ノードの TDP 分の電力であると仮定し、システム全体のノード数の仮定を 400, 540, 680, 820, 960 と変更した場合を仮定する。ジョブの到着タイミングには理研 RICC のログを用い、その先頭から 600 ジョブを実行すると想定する。さらに、ジョブ特性の違いによる性能・電力消費特性の違いを解析するため、評価では 5 つのジョブミックスとして、85W, 70W, 55W の電力キャップで動作させるジョブの割合を、85W:70W:55W = 1:0:0 (high-only)、3:2:1 (many-high)、1:1:1 (middle)、1:2:3 (many-low)、0:0:1 (low-only) と変化させ評価した。図4 左は、システムの電力制約にチアする電力使用率を示しており、ジョブが割り当てられたノードとアイドルノードの電力内訳も示している。図より、ノード数を 680 台まで増やした場合は電力利用率が向上することが分かる。一方でそれ以上にノード数を増やすと逆に電力利用率が低下することも分かる。基準の 400 ノードに対し、オーバプロビジョニングとして追加するノード数が少ない場合は、ノードの利用率は高い。これは計算ノードに対し電力が十分確保できるためである。一方で追加ノードが多い場合、アイドルノード時の消費電力電力も増えることにより、ジョブ実行に利用できる電力が減少してしまう。

ジョブスループットはノード数が増えるにつれ向上することが図4 右よりわかる。オーバプロビジョニングで追加ノード数だけより多くのジョブが実行できるためである。一方、960 ノードになるとスループットは大きく低下している。さらに、低下具合はジョブミックスによって大きく異なることが確認できる。CPU インテンシブジョブが多い場合は各ノードのジョブ実行電力が大きくなるため、追加ノードが少ない場合であっても電力資源が枯渇しやすい傾向にあり、追加ノードは少ないほうが好ましい。一方で low-only のように個々のジョブの消費電力が少ないメモリインテンシブなジョブが多い場合、電力資源が枯渇しにくいいため多くのノードを追加してもそれらノードを使い切ることができ性能が向上する傾向にある。つまり、追加ノード数は多いほうが良いことが分かる。本評価より、オーバプロビジョニングシステムの設計には、アイドルノードの電力を考慮することが重要であることがわかった。

電力制約適応型システムにおけるジョブ管理技術として、空調に必要な電力も考慮し、空調も含めた合計電力を制約としてシステムの電力配分を最適化するためのジョブスケジューリング技術の構築も行った。

一般的に、大規模 HPC システムでは空調に必要な電力も無視できず、計算ノードの電力に匹敵するほどの電力を消費するとの報告もある。空調は各ノードのシャーシ内の温度が十分に低く保たれるようにその出力が決定される。そこで、シャーシ内温度の最大値が低くなるように、すなわちホットスポットの温度が低くなるように各ノードの電力を制御することができれば空調の電力を削減することができ、その分の電力を計算ノードに配分することが可能となる。そこで、ノード間の熱の拡散も考慮し、ホットスポットになりやすいノードに対しては消費電力が低いジョブを、逆にノード内シャーシの温度が低く保たれやすいノードに対しては、高い電力を必要とするジョブを割り当てることで、システム全体としてホットスポットの温度を低く保ちつつ、電力制御を行った場合の各ジョブの性能への影響も抑えることができると考えられる。

そこで、ノードのホットスポットへの成りやすさと、ジョブの性能電力特性をもとに、ノードへのジョブ割り当てについて整数線形計画法を用いて最適化しつつ、ジョブスケジューリングを行う手法を開発した。シミュレーションにより、Interpid システムのジョブ実行ログをもとにして、大規模システムを想定したスケジューリングの評価を行った。評価結果を図 5 に示す。

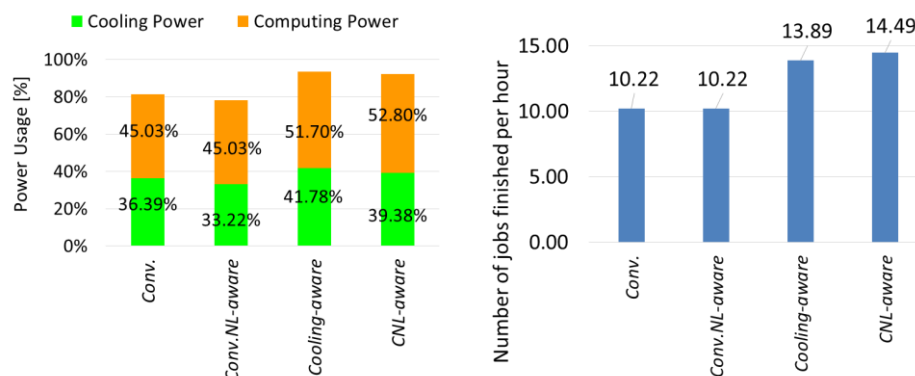


図 5: 空調を考慮したジョブスケジューリングの評価結果

図 5 左は空調電力と計算ノード電力の内訳を、合計電力制約に対する電力比で示したものである。Conv. は空調と計算ノード間で電力分配をせず、またノードへのジョブ割り当ても最適化しない従来スケジューリングを、Conv.NL-aware はノードへのジョブ割り当てのみを最適化した場合を、Cooling-aware は空調と計算ノード間での電力分配のみを考慮した場合を、そして、CNL-aware がノードへのジョブ割り当て、および空調と計算ノード間で電力分配を考慮した提案手法を表している。図より、空調と計算ノード間での電力分配を考慮することで、計算ノードで利用可能な電力資源が増加し、計算ノードで多くの電力が消費できることがわかる。また、ノードへのジョブ割り当てを最適化することで、ホットスポットの温度が低下し、空調の電力が低下するとともに、さらに多くの電力を計算ノードに割り当てることができている。

計算ノードに多くの電力資源を割けることで、結果的としてシステム全体の性能を向上させることが可能となる。図 5 右は各スケジューリング手法におけるシステムのスループット(時間あたりのジョブ実行数)を示したものである。提案手法である CNL-aware が最も高いスループットを達成できており、Conv. に比べておよそ 1.4 倍の高スループットを達成している。本年度は、従来型の資源管理手法に比べて 1.5 倍近くのスループット向上を目標に研究を実施したが、その目標をほぼ達成したと考えられる。昨年度までに、他の電力制約下でのスケジューリング最適化手法も検討しており、それらと組み合わせることでさらに高いスループット

ットを達成可能である。なお、本成果は IPDPS2017 で発表した。

その他、複数コアを持つプロセッサ内のネットワークの消費電力制御技術や消費エネルギー削減手法も開発した。

### ★開発したソフトウェア(公開済み)

#### 電力性能最適化パッケージ

- ・機能：自動でのノブ制御ポイント埋め込み、プロファイリング、電力性能最適化
- ・特徴：自動・半自動でのコード最適化機能を持ち、電力性能最適化を静的に行うことが可能
- ・実行可能なマシン：電力性能ノブを持つ様々なシステム
- ・潜在的なユーザ：電力制約下でのスパコンアプリ最適化を行うユーザ

#### 電力制約適応型システム向けのリソースマネージャ

- ・持っている機能：電力制約内でのジョブスケジューリング、計算ノードの電力監視と制御
- ・特徴：広く利用されている Srun に統合されているため導入・利用が容易
- ・実行可能なマシン：Intel 系の CPU を搭載する HPC 計算機システム
- ・潜在的なユーザ：大規模 HPC 計算機システムを運用する管理者

### ★エクサスケールへの貢献度

- 電力性能最適化パッケージの開発と公開：エクサスケールシステムでは利用可能な消費電力に制限がある中で、いかにアプリケーションの性能を引き出すかが重要となる。その最適化のための戦略の開発やツール群を実装して公開したことは、将来的なエクサスケールシステムで必須となるであろう電力を意識した最適化の一つの道筋を示したと考えられ、貢献度は大きいと考えられる。
- 電力制約適応型システム向けのリソースマネージャの開発と公開：エクサスケールシステムでは消費電力制約内でシステムを運用することが求められるため、電力資源管理が可能な資源管理ツールは必須となる。それを開発し、大規模システムで検証・評価もしており、さらに海外の他研究グループでも電力制約下でのジョブスケジューリング最適化に利用されていることから今後の学術的イノベーションも期待される。これらの点で、本ツールの開発と公開の貢献度は非常に大きい。

## 3.2 最適化コード生成(富士通株式会社 三吉グループ)

### (1)研究実施内容及び成果

#### ① 研究実施内容

ポストペタスケール HPC システムの開発では消費電力が主要な設計制約の 1 つとなるが、そのような制約に対する対応としてハードウェア・オーバプロビジョニングというアプローチが提案されている。ハードウェア・オーバプロビジョニングの考え方で構築されたシステムでは、システムを構成する各コンポーネント(CPU、GPU、メモリ、インターコネクト等)がすべて最大負荷時電力で動作した場合にそれらの合計が電力制約を超過してしまうため、電力性能ノブを制御することで実効電力を制約以下に抑えるための仕組みが必要となる。その際、当然のことながら、性能の最大化も求められるため、本質的には電力あたり性能の最適化が重要課題となる。

電力性能ノブとしては CPU の動作周波数を変更する方法(Dynamic Frequency Scaling, DFS)やスレッド並列処理の使用コア数を変更する方法(Dynamic Concurrency Throttling, DCT)などがあり、これらを適切に用いることで電力あたり性能を改善できることが知られている。しかし従来研究では、ジョブ実行の全体を通じて CPU 動作周波数を選択/固定し、その結果の実行時間を収集して、良い結果となった実行条件(CPU 動作周波数)を活用す

る程度に留まっていた。しかしこの方法では、i)1 回のジョブ実行時間が長い場合に最適解が求まるまでの時間が長く掛かる、ii)ジョブ実行の内容(アプリケーションとしての計算内容)が同一でない場合の扱いが自明ではない、という問題があり、広く実用に供されるには大きな支障となっていた。

そこで本研究では、1 つのジョブの実行最中に電力性能ノブを最適化することで、これら問題点の解決を目指した。具体的には、アプリケーション開発者が性能確認用に設定した性能評価区間をそれぞれ対象として、電力あたり性能の評価と電力性能ノブの制御を行い、実行性能の低下を一定範囲に抑えつつ、単一ジョブの実行最中に電力あたり性能を最適化する手法を開発した。この手法を用いることにより、ジョブ毎の計算内容の違いを懸念することなく、短時間で、電力あたり性能を最適化できることが示された。

## ② 研究実施方法

上述の研究実施内容を進めるにあたり、i)アプリケーション特徴抽出技術の開発、ii)電力性能ノブ制御最適化コード生成技術の開発、の 2 つの研究項目を当初設定し、研究を推進した。

前者のアプリケーション特徴抽出技術については、アプリケーション毎の電力消費傾向の調査、Intel 社による RAPL 機能等の電力計測技術の調査、パフォーマンスカウンタを用いた電力あたり性能の推定手法の検討/評価/改良などを行い、上記最適化手法の実現に必要な要素技術を整備した。

一方、後者の電力性能ノブ制御最適化コード生成技術に関しては、上記の電力消費傾向調査の結果、コンパイル時よりも実行時のほうが最適化の自由度が高いと判断し、RAPL/DFS/DCT 等の電力制御技術の調査、実行時電力性能最適化ライブラリの試作/評価/改良/応用検討などを行った。

このようにして開発した最適化手法は、近藤グループ主導の電力性能最適化フレームワークへ統合することを目指して、技術的検討を継続している。また、これらの研究実施においては、自機関の評価用設備を用いるだけでなく、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も利用し、評価用データの採取/分析を行った。

## ③ 研究成果とその位置づけ

試作した実行時電力性能最適化ライブラリを用い、HPCG ベンチマークを対象として Xeon E5-2698v3 (2.3GHz、16コア)×1 ソケットで評価した結果を以下に示す。CPU 動作周波数を引き下げると共に、スレッド並列処理の使用コア数を抑制することによって、消費電力は 27～30%削減され、電力あたり性能は 24～26%改善されたことが示されている。また、RAPL 機能による電力計測では最適化対象区間を短くできないという問題の対策として、パフォーマンスカウンタ値からの推定で最適化を行った場合(図中 Optimized w/o RAPL のケース)でも遜色ない結果が得られたことも示されている。なお、これらの最適化結果は、実行性能の低下を 10%以内に抑えることも要件としつつ得られたものであり、利用者の不利益が一定範囲に収められるよう配慮された上での結果である。

また、ポスト「京」の開発に向けて整備されたミニアプリの 1 つである NICAM-DC-MINI を対象として Xeon E5-2680 (2.7GHz、8コア)×2 ソケット×10 ノードで評価した結果を以下に示す。この評価では、評価環境の運用上の制約によってスレッド並列処理の使用コア数しか最適化できなかったが、平均で 5.1 コア/ソケットを使用することにより、実行性能の低下を 3%に抑えつつ、消費電力を平均 13%削減できたことが示されている。

これらの結果より、特にメモリ律速傾向のアプリケーションでは、単一ジョブの実行最中に電力性能ノブを制御し、実行性能の低下を一定範囲に抑えつつ消費電力を削減して、電力あたり性能を改善することが可能であると示された。すなわち、エクサスケールの HPC システム開発において、ハードウェア・オーバプロビジョニングのアプローチに実用性があると示されたと言える。

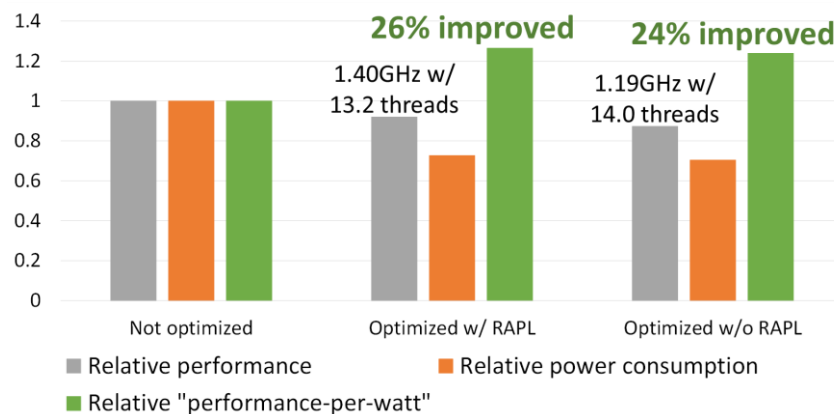


図 6: HPCG ベンチマークの電力あたり性能最適化結果

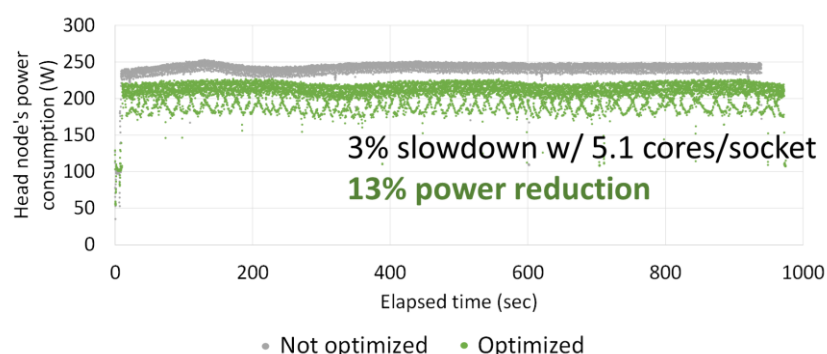


図 7: ミニアプリ NICAM-DC-MINI の電力あたり性能最適化結果

なお類似の研究としては、Intel 社の主導する GEOPM プロジェクトがあり、RAPL 機能の設定をノード毎、性能評価区間毎に実行時に調整することで、並列プログラムのノード毎の実行進捗バラつきを平準化し性能改善した事例が報告されている。しかしこの報告は、ジョブ全体で一定の電力バジェットを各ノードに適切に再配分するものであって、電力バジェット自体を最適化するものではないと考えられる。

### 3.3 電力性能予測技術(九州大学 井上グループ)

#### (1)研究実施内容及び成果

##### ① 研究実施内容

適応的な電力制御のためには、アプリケーションの電力性能特性の抽出が必須である。しかし、運用中の大型計算機をプロファイリング用途等に利用することは現実的ではない。さらに、メモリサイズや計算時間の制限から、大規模アプリケーションを小規模システム上でそのまま実行させることも難しい。加えて、新たな電力性能ノブを利用する場合は、その効果を見積もる手段が必要となる。そこで、大規模システム上での電力性能挙動を小規模システムで予測するための技術、ならびに各種ツール開発を行った。具体的には、電力性能挙動プロファイリング技術、電力制約下におけるシステム性能予測技術、インターコネクション・ネットワーク消費電力推定技術を開発した。加えて、米国ローレンス・リバモア国立研究所ならびに米国アリゾナ大学と連携して、本電力プロファイリング環境を利用した大規模スーパーコンピュータの電力特性を解析した。その結果、電力制約下では性能ばらつき (LSI 製造ばらつきに起因) が問題になることが分かった。そこで、この問題を解決するための電力資源配分技術を開発し、その有効性を示した (2015 年 SC にて発表)。また、大規模グラフ解析など新しいアプリケーションを対象とした電力性能特性解析を行うなど、他 CREST グループと連携した研究も促進した。

## ② 研究実施方法

電力性能挙動プロファイリング技術では、単一(または少数)の実機計算ノードを用いて、アプリケーション実行時における電力性能の詳細な挙動をプロファイリングするための技術ならびにライブラリを開発した。また、Intel 社が提供する電力制御 API を用いた電力キャッピング適用環境等を構築し、プログラム実行における電力制御を可能にした。その後、大規模スーパーコンピュータへ適用するための各種改良を加えた。その結果、約 2,000 基の計算ノードから構成されるスーパーコンピュータでの消費電力測定ならびに電力制御(電力キャッピングの適用や動作周波数の設定)が可能となり、本 CREST プロジェクトにおける実験環境を整備した。このような環境は世界的にもまれであり、本 CREST プロジェクトでの技術開発に大きく貢献した。

インターコネクション・ネットワークの電力予測に関しては、過去に開発されたインターコネクト・シミュレータ NSIM を改造して NsimPower と呼ぶインターコネクト部の電力シミュレーション環境を構築し、電力モデルや電力推定法の妥当性に関する検証を行った。その後、ツールの公開とその普及を鑑み、米国ローレンス・リバモア国立研究所と連携して同研究所で開発されたインターコネクト・シミュレータ TraceR に電力推定機能を追加拡張した。一方、計算機ノードを対象としたシステムレベルの電力性能に関しては、電力制約時に生じる LSI 製造ばらつきを考慮した推定法を考案し、その有効性を評価した。

## ③ 研究成果とその位置づけ

### 1) 電力測定・制御環境の構築

HPC アプリケーションの電力性能特性を解析するため、まず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。Intel 社が提供する RAPL (Running Average Power Limit) の使用を前提とし、CPU や DRAM に電力キャッピングを施すための API を開発実装した(ただし、大規模スーパーコンピュータにおいては、DRAM 電力キャッピングは無効)。また、現状では特定のアプリケーションに限定されるが、プログラムのソースコード部分とその実行における消費電力値の対応付けが可能な可視化ツールを開発し、当該アプリケーション実行時の電力消費の詳細を分析可能とした。電力観測・制御用のプロファイリング環境は、MPI と OpenMP のハイブリッド並列化されたプログラムの消費電力の測定および電力制約が可能ないように開発しており、またライブラリ化がされている。これを用いることで、並列化されたアプリケーションプログラムの消費電力の最大値が与えられた電力制約値を超えないように制御できることも確認している。

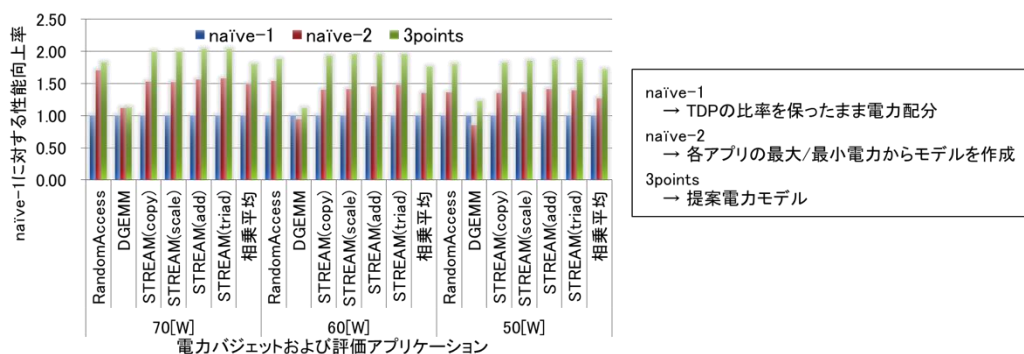


図 8: CPU と DRAM の電力配分による性能解析結果

上記で開発した電力観測・制御環境を用い、いくつかのアプリケーションを例に電力性能特性解析を実施した。対象としたアプリケーションプログラムは MHD (Magneto Hydro Dynamic) シミュレーションや FMO(Fragment Molecular Orbital method)などが挙げられる。計算ノードが使用できる総電力バジェット量が与えられた際に、アプリケーション毎、関数毎、

特性の異なるフェーズ毎など様々な粒度で CPU と DRAM の電力バジェットを変更し性能を測定した。その結果、CPU と DRAM の TDP(Thermal Design Power)に基づき電力バジェット配分を決定するナイーブな方法に対し、CPU ならびにメモリへの負荷を考慮して適切な電力バジェット配分を施すことで、MHD の場合で最大 2.1 倍(フェーズ単位で電力バジェット配分を変更)、FMO の場合で 1.25 倍(粗粒度関数単位で電力バジェット配分を変更)の性能向上を達成できることを明かにした(図 8)。

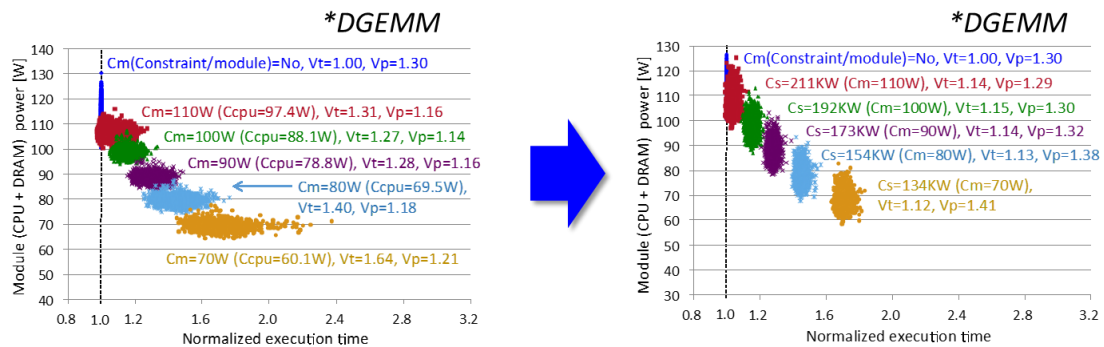


図 9: 性能ばらつきと電力分配最適化

実機を用いた電力性能特性の詳細を解析した結果、同一ハードウェア構成にて同一アプリケーションを実行した場合でも、計算ノード間には電力性能特性のばらつきが生じる(同一電力バジェットで実行した場合でもノード間で性能ばらつきが生じる)ことが分かった(図 9 左)。また、この性能ばらつきは、電力制約が厳しくなるにつれ、より顕著になる傾向があることが明らかになった。そこで、このような電力性能特性ばらつきを考慮した電力性能制手法を開発した(図 10)。具体的には、少数計算ノードを用いた全計算ノード電力性能特性予測技術を開発した。本技術では、全計算ノードにおいてマイクロベンチマークを用いた消費電力測定をシステム起動時などに実施し、電力性能特性テーブルを作成する。そして、実行対象アプリが定まった時点で少数の計算ノードで消費電力プロファイリングを行い、電力性能特性テーブルに基づき、実行ターゲットとなるアプリケーションが与えられた際の計算ノード間特性ばらつきを推定する。この結果に基づき、各計算ノードに配分すべき消費電力、または、各計算ノードの動作周波数を自動で決定し、電力制約時における計算ノード間の性能ばらつきを抑制する(図 9 右)。このような機能を搭載したライブラリを開発し、実 HPC アプリケーションを用いた定量的評価を行った結果、最大で 5.4 倍の性能向上を達成した(図 11)。なお、本最適化技術は、米国ローレンス・リバモア国立研究所、および米国アリゾナ大学と連携して開発したものであり、本成果の内容を 2015 年の SC にて発表した。

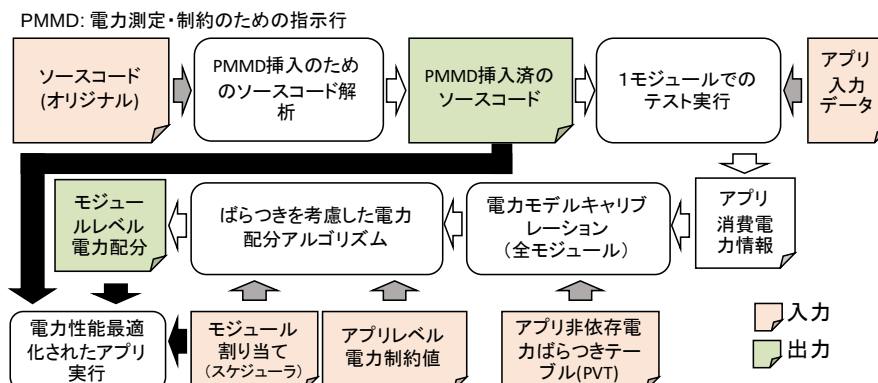


図 10: ばらつきを考慮した電力性能制御手法

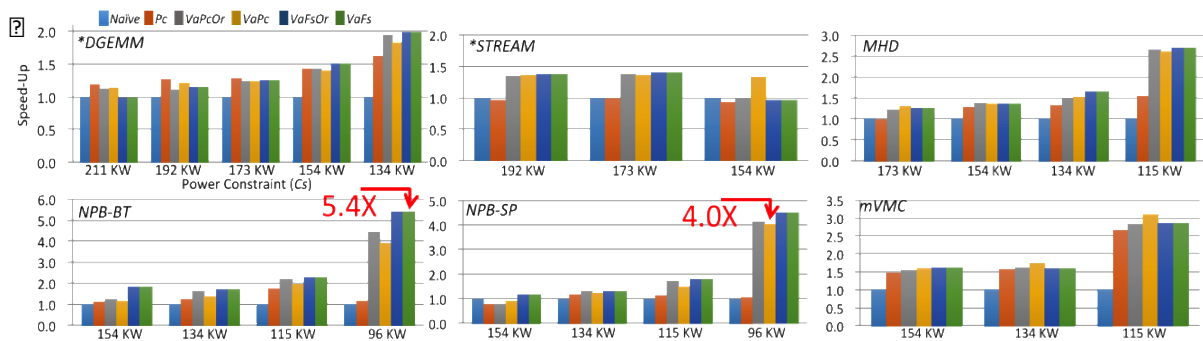


図 11: 性能評価結果

### 3) LSI 製造ばらつきを考慮した電力性能推定技術の開発

大規模スーパーコンピュータに対し電力制約を施した際、HPC アプリケーションの実効性能がどの程度低下するかを推定するためのフレームワーク(図 12-A)を構築した。本手法では、小規模クラスタを用いてアプリ実行を行うことで消費電力-動作周波数および動作周波数-正規化実行時間に関する相関モデルを構築し(図 12-B)、これらと非制約時の実行時間の実測値とを組み合わせることで電力制約時の実行時間を推定する。1,920 モジュール(各モジュールに 1 個の CPU と DRAM が含まれる)を有する大規模並列実行を対象とした評価を行った結果、モデルに基づく簡便な推定手法であるにもかかわらず、平均誤差 10~15%の精度で推定できることが明らかとなった(図 13)。

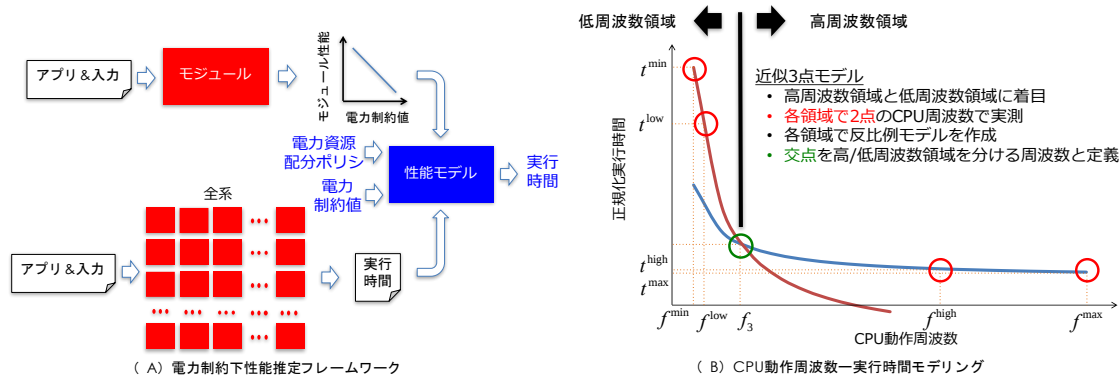


図 12: 電力制約下での性能推定法

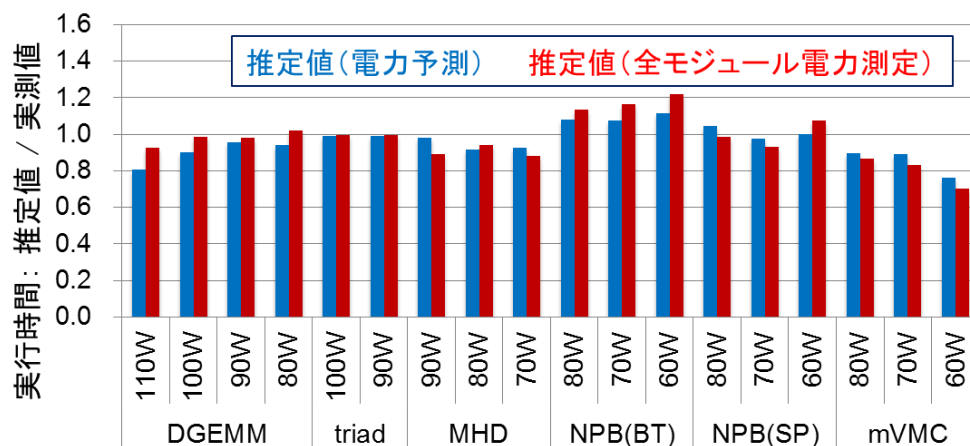


図 13: 電力制約下での性能推定結果

#### 4) インターコネクション・ネットワーク消費電力推定ツールの開発

インターコネクション・ネットワーク(以降ネットワーク)においてはリンクが大きな電力を消費することから将来は低電力モードをサポートするリンク(以降 On/Off リンク)がネットワークに採用されることが期待されている。通常のリンクが通信の有無に関わらず常に一定の電力を消費するのに対し、On/Off リンクは非通信時に低電力モード(通常時の 10%~60%)に遷移することによって電力を削減する。アプリケーションの実行中であっても通信を行っていない状態のリンクが HPC システムには多数存在することから、On/Off リンクを利用することでネットワークの電力を大幅に削減できる。On/Off リンクを用いたネットワークの性能と消費電力はネットワーク・トポロジやアプリケーションの通信パターンに強く影響されることから、高性能かつ省電力なネットワークを設計するためには上記の影響を正確に見積もる必要がある。大規模ネットワークの性能や消費電力を見積もる手段として、一般にはインターコネクト・シミュレータが利用されている。そのひとつとして、並列離散事象シミュレーションにより高いスケーラビリティを実現する TraceR が挙げられる(米国ローレンス・リバモア国立研究所が公開している)。TraceR はプログラムの通信トレースを入力として、その通信性能をシミュレーションすることが可能である。さらに、複数のジョブを同時に実行した際の性能を見積もることが可能であり、HPC システムの実利用を想定してシミュレーションすることができる。しかしながら、On/Off リンク機能を実装しておらず、On/Off リンク適用時の消費電力や通信性能をシミュレーションできないという課題があった。

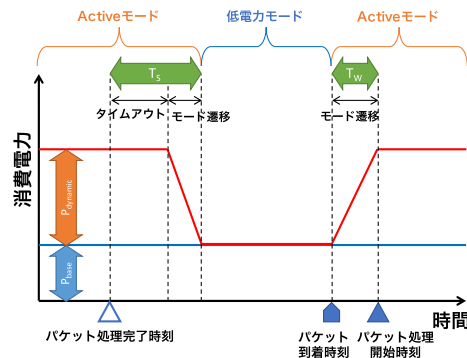


図 14: On/Off リンクの概念

そこで我々は、米国ローレンス・リバモア国立研究所と連携し、TraceR をベースに On/Off リンクのシミュレーション機能を有するインターコネクト・シミュレータ(TraceRP)を開発した。TraceRP が想定するネットワークの低電力化技術の概要を図 14 に示す。パケット処理後、一定時間( $T_s$ )パケットがない場合、当該リンクは低電力モードへと遷移する。このとき、低電力モードでは消費する電力が ACTIVE モード時よりも  $P_{dynamic}$  だけ削減される。低電力中に処理するパケットが生じると、当該リンクは低電力モードから  $T_w$  時間をかけて ACTIVE モードへと遷移する。TraceRP は低電力モードの利用によって生じるインターコネクト電力の時間的な変化を見積もることが可能なだけでなく、低電力モードから ACTIVE モードへと遷移する際に生じる通信遅延も含めた通信性能も見積もることが可能である。

Fat-tree、Torus、Dragonfly の 3 つのトポロジを対象に 8,192 プロセスを実行して消費電力と性能を調査した。評価には HPC システムで利用されるアプリケーションを想定した 5 つの通信性能評価プログラムを用いた。 $T_s=600[\mu s]$ 、 $T_s=17[\mu s]$ 、ポートあたりの  $P_{base}=2.08[W]$ 、 $P_{dynamic}=1.36[W]$ として評価した。評価の結果を図 15 に示す。左縦軸はリンクの正規化消費電力であり、右縦軸は正規化実行時間であり、横軸にはベンチマーク・プログラムを示している。Fat-tree では最大約 27%、Torus では 33%、Dragonfly では 26%の消費電力削減効果があることが分かる。また、各トポロジにおいて性能が低下しており、特に Torus では最大 4%の性能低下が生じることが分かった。

TraceRPを用いることで、様々なトポロジに対してプログラムの性能およびネットワークのリ

リンクの消費電力を推定可能になる。また、On/Off リンクにおいて消費電力と性能を決定する重要なパラメータである  $T_s$  の探索も実施可能であり、ネットワーク設計者による適切な  $T_s$  の設定をサポートすることができる。

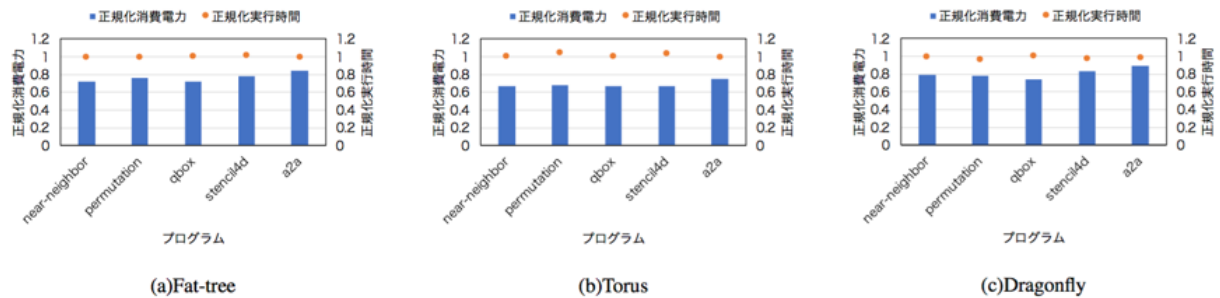


図 15: 低電力モード適用時の消費電力と性能

#### 5) 大規模グラフアプリケーションを対象とした電力性能最適化

大規模グラフ解析で主要な処理となる幅優先探索に着目し、実機を用いた電力特性解析を行った (CREST 藤澤グループならびに遠藤グループと連携)。具体的には、CPU と DRAM への個別電力制御が可能な単一計算ノードを評価用プラットフォームとし、大規模グラフ解析ベンチマークとして知られる Graph500 を用いた解析を実施した。特に、問題サイズ依存性とスレッド数依存性を調査し、消費電力の時間的変動を解析した。その結果、Graph500 の性能を決める幅優先探索では、1) DRAM の消費電力は入力となる問題サイズが一定の範囲で増加すること、2) 実行に要するスレッド数により DRAM 消費電力が大きく異なること、3) グラフ構築処理においては電力消費の振る舞いに関して幾つかのフェーズが存在すること、4) グラフ探索では全体を通して一つのフェーズとなっていること、などが明らかになった。また、電力制約下での Graph500 の性能について評価した結果、5) 最大の性能となる CPU/DRAM への電力資源配分は探索対象グラフのサイズにより異なり、グラフサイズが小さい場合には CPU へ、グラフサイズが大きい場合には DRAM へより多くの電力資源を配分する必要があることが分かった。これらの他にも、大規模スーパーコンピュータを用いた Graph500 の電力性能特性の解析や、ソースコードレベル性能チューニングが電力特性に与える影響の解析も行った。

#### ★公開済みソフトウェア: 電力測定・制御ライブラリ

- 機能: ソースコードにてディレクティブを挿入するだけで、プログラム実行時の CPU と DRAM の消費電力測定ならびに電力制約値設定が可能
- 特徴: ソースコード内で容易に利用可能
- 実行可能環境: Intel 社が提供する RAPL が利用可能な環境
- 潜在的なユーザ: スパコン設計者、電力最適化を実施するスパコンアプリ開発者
- その他: 東大グループが開発した最適化パッケージに統合し公開

#### ★公開予定ソフトウェア: インターコネクト消費電力測定ツール (TraceRP)

- 機能: 様々なトポロジを想定したインターコネクト消費電力推定
- 特徴: 既存シミュレータ TraceR と連携して動作
- 実行可能環境: TraceR が実行可能な環境
- 潜在的なユーザ: インターコネクト設計者、電力最適化を実施するスパコンアプリ開発者
- その他: LLNL が Github で公開している TraceR の拡張版として、LLNL の TraceR の同一リポジトリに TraceRP を公開

★エクサスケールへの貢献度

- 電力測定・制御ライブラリの公開:エクサスケールシステムを設計する上で最も厳しい制約となる消費電力を測定・制御するためのベースとなるライブラリ実装し、他グループの開発ツールや評価に利用された点で、貢献度は非常に大きい。

3. 4 システムアーキテクチャ(電気通信大学大学 三輪グループ)

(1)研究実施内容及び成果

① 研究実施内容

1) 電力性能ノブの探索

スーパーコンピュータ内の各ハードウェア・コンポーネントで既に利用可能な電力性能ノブならびに将来利用可能になるであろう電力性能ノブを調査した。各種の文献調査ならびに実際にスーパーコンピュータの運用に携わっている方への聞き取り調査をもとに、現在のスーパーコンピュータで利用可能な電力性能ノブを明らかにした。また、デバイス、アーキテクチャ、ハイパフォーマンスコンピューティング等の国際会議への参加を通して将来的な実用化が見込まれる電力性能ノブに関する情報のキャッチアップに努めた。特に後述する On/Off リンクとメモリホットプラグは、既に一部のハードウェアやソフトウェアでサポートが開始されていることから将来のスーパーコンピュータで利用可能になる可能性が高い電力性能ノブと考え、これらの技術動向調査に重点的に取り組んだ。

2) 電力性能ノブのモデリング

各種の既存電力性能ノブのモデリングを実施するとともに、将来的に実用化が見込まれる電力性能ノブのモデリングも行った。既存電力性能ノブとしては、スーパーコンピュータ内で最も消費電力の大きなハードウェアである CPU と GPU を制御する技術である DVFS に着目し、CPU と GPU それぞれについて電力性能モデルを構築した。ネットワークについては、将来的な実用化が見込まれている電力性能ノブである On/Off リンクに着目し、On/Off リンクを有するネットワークの電力性能モデルの開発を行った。さらにメモリについては、やはり将来的な実用化が見込まれている電力性能ノブであるメモリホットプラグに着目し、メモリホットプラグを利用可能なシステムの電力性能モデルを開発した。

3) 電力性能ノブ制御アルゴリズムの検討

既存の電力性能ノブに加えて将来的な電力性能ノブの制御アルゴリズムの開発を行った。既存の電力性能ノブの制御アルゴリズムとしては、CPU と GPU の DVFS およびタスクマッピングを利用することでジョブの電力性能を最適化する手法、ならびに、UPS 等の蓄電池を有するシステムにおいて CPU の DVFS を利用して電力制約下におけるジョブ性能を向上する手法を開発した。将来的な電力性能ノブの制御アルゴリズムとしては、On/Off リンクを有するネットワークで発生した余剰電力を利用して CPU ブースティングを行う手法、ならびに、メモリホットプラグを利用してジョブ実行中のノードのメモリ電力を削減する手法を開発した。また、近藤グループが開発した最適化フレームワークに上記の手法を統合するため、On/Offリンクを有するネットワークの余剰電力をプロファイリングする際のプロファイリング用コードを自動生成するコンパイラの開発を H29 年 8 月現在行っており、上記開発は H30 年 3 月までには完了する予定である。

② 研究実施方法

電力制約下で、電力性能ノブを通してハードウェアとしてのシステム特性をソフトウェアから制御・再構成できるという状況を想定し、種々のアプリケーションに対して高性能な計算能力を提供できるシステムアーキテクチャを提示した。システムソフトウェアと連携し適応的電力制御を行える真に有用なアーキテクチャを探索しつつ、その知見をポストペタスケールシステム開発に繋げることを目指し、将来的なハードウェアアーキテクチャも見据えた電力性能ノブの探索、電力性能ノブのモデリング、電力性能ノブ制御アルゴリズムの開発の研究項目を実施した。

電力性能ノブの探索では、技術トレンドから導かれるデバイスやハードウェア実装技術の動向を元に、ポストペタスケールシステムとして可能性のあるシステム構成を列挙し、それらに対して従来から利用可能な電力性能ノブの他、将来的に利用可能となるであろう電力性能ノブの候補とその制御範囲を調査した。

電力性能ノブのモデリングでは、上記ノブ探索で得られた多様な電力性能ノブの抽象化とモデリングを行い、ソフトウェアに見せるべきシステムアーキテクチャを検討した。電力性能ノブが実現する電力・性能のトレードオフ関係、電力性能ノブ制御のオーバーヘッドとそれに依存する制御の時間粒度、実装方式に依存するノブの空間粒度の観点からアーキテクチャの検討を行った。

電力性能ノブ制御アルゴリズムの開発においては、電力と性能がアプリケーションの実行状況に依存することを考慮し、電力性能ノブの効果を最大化すべく、静的最適化と動的最適化を相補的に融合し、ソフトウェア階層と協調する手法を開発した。このアルゴリズム開発は、電力性能ノブ最適化技術の「基盤システムソフトウェアの開発」で開発中の最適化パッケージへの融合を目指し、他グループと密な連携をとりながら実施した。

### ③ 研究成果とその位置づけ

#### 1) 電力性能ノブの探索

現在のスーパーコンピュータにおいて利用可能な電力性能ノブの調査を行った。既に利用可能な電力性能ノブとしては、CPU における DVFS、クロックゲーティング、パワーゲーティング、GPU における DVFS、メモリにおけるリクエストのスケジューリングなどがあり、それぞれは制御の空間粒度も時間粒度も電力削減量も異なることが調査により明らかとなった。また最近では制御用ソフトウェアの開発も進んでおり、Intel 社の RAPL (Running Average Power Limit) に代表されるように、ユーザは簡略化されたインタフェースを通して上記電力性能ノブの制御が可能になっていることがわかった。スーパーコンピュータにおける電力性能制御の必要性が増すにつれて制御を行うユーザ側の負担を軽減するソフトウェアの重要性も増すと予想されるため、このようなソフトウェアの技術動向には今後も注意していく必要がある。

スーパーコンピュータにおいて将来利用可能になるであろう電力性能ノブの調査を行った。CPU に関しては、近年ではコア単位の DVFS やパワーゲーティング制御が可能な CPU が登場し始めており、今後はコアあるいはコア内のユニットなどの細かい空間/時間粒度における電力制御が可能になると期待できる。メモリに関しては、現在ハードウェアによる自動制御が行われている DRAM の電力を、今後はソフトウェアが直接制御できるようになると期待される。そのような技術の 1 つとしてメモリホットプラグがあり、メモリホットプラグでは OS の指示により DIMM の電源遮断と復帰を実現する。メモリホットプラグは既に Linux カーネル 3.9 から正式にサポートされており、今後ハードウェアとファームウェアのサポートが進めばスーパーコンピュータにおいても利用可能になる見込みが大きい。ネットワークに関しては、通信を行っていないリンクを省電力モードに変更することでリンクの消費電力を削減する技術 (On/Off リンク) の実用化が期待される。イーサネットにおいては Energy Efficient Ethernet (EEE) という名称で上記の技術は既に規格化されており、今後は上記の規格に適合したイーサネットを採用したスーパーコンピュータが開発される可能性が高い。また、On/Off リンクはイーサネットにおいて既にその有用性が実証されていることから、今後は InfiniBand などの他のネットワークにおいても規格化が進むと期待される。

以上の調査結果をもとに、本研究では、On/Off リンクとメモリホットプラグという 2 つの新たな電力性能ノブに着目し、これらの電力性能ノブを採用することによるスーパーコンピュータの電力と性能への影響を検討した。On/Off リンクとメモリホットプラグに着目した理由は、スーパーコンピュータの電力制約が厳しくなるにつれてこれまであまり注目されてこなかったネットワークとメモリの電力制御の重要性が増すと考えられたこと、そしてこれらの電力性能ノブが実用化の段階に入りつつあったこと、の 2 点である。

On/Off リンクについては、EEE 対応の 10G イーサネットを用いて On/Off リンク評価用の

プロトタイプシステムを世界で初めて構築し、上記プロトタイプシステムを用いて On/Off リンクが並列アプリケーションの性能と電力に与える影響に関する実証実験を世界で初めて実施した。プロトタイプシステムは、8 ノードが 1 台のネットワークスイッチに接続された構成とした。このシステム上で NAS Parallel Benchmark (NPB) を実行した結果、EEE 機能を使用した場合にはリンクを省電力モードから通常モードへと復帰する際に生じる通信遅延によるアプリケーションの性能低下が起きると予想されたが、いずれの問題サイズにおいてもそのような性能低下は確認できなかった。また、EEE 機能を使用することにより、アプリケーション実行中のネットワークスイッチの消費電力を平均 4.67W、最大 5.44W 削減できることが確認できた。またノード数を 16 台に増やした場合も EEE 機能の使用による性能低下は確認できなかったことから、On/Off リンクを大規模計算環境に適用した場合にも期待の持てる結果であった。この成果は情報処理学会 ACS 論文誌 Vol. 7 No. 4 に掲載済みである。

メモリホットプラグについては、九州大学情報基盤センターで運用中のスーパーコンピュータ CX400 のジョブログを解析し、スーパーコンピュータにおいてメモリホットプラグが利用可能になったと仮定した場合の省電力効果の見積もりを行った。解析の結果、Linux 3.11 以降でサポートされているメモリホットプラグの制御アルゴリズムを利用することでノードあたり 4W 程度の電力を削減できることを世界で初めて確認した(図 FIXME の ACPI)。さらに、上記の制御アルゴリズムの電力削減量はノードの稼働率に強く影響を受けるが、後述する我々が開発した制御アルゴリズムを使用することで電力削減効果をノードあたり 8W 程度にまで改善できることを初めて示した(図 16)。

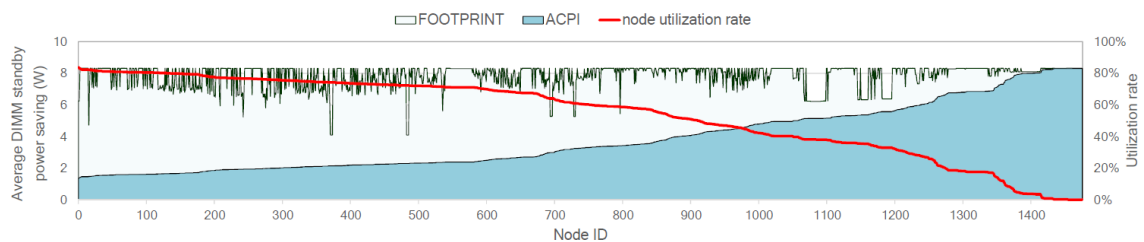


図 16: CX400 におけるメモリホットプラグの効果 (2013 年 1 月から 6 月までに実行されたジョブのログをもとに試算。各ノードは 16x 8GB DDR3 メモリを有し、各メモリはアイドル時に Self-refresh + register-off モードに遷移する (0.13W/GB の電力消費) と仮定。)

## 2) 電力性能ノブのモデリング

CPU と GPU を同時に使用してアプリケーションを実行するヘテロジニアスな実行環境において、CPU と GPU の周波数と各デバイスに割り振るタスク量から各デバイスで実行されるタスクの性能と消費電力を見積もるモデルを開発した。各モデルは、CPU と GPU それぞれ単体でアプリケーションを実行した時の各周波数における電力と性能が既知であると仮定し、タスクの分配率が与えられた時の各デバイスの稼働率をもとにそれぞれの電力と性能を予測する。CPU と GPU それぞれ単体でアプリケーションを実行した時の各周波数における電力と性能はプロファイルにより取得する。Rodinia Benchmark Suite の 4 つのアプリケーション (Bfs, Hotspot, Kmeans, Particlefilter) を用いて評価した結果、開発したモデルは CPU の電力と性能の予測ならびに GPU の電力と性能の予測それぞれにおいて良好な精度を示すことが確認できた。本成果は ICCD'13 にて発表した。

On/Off リンクを有するネットワークの性能モデルを開発した。開発したモデルは、On/Off リンクを有しないネットワーク上でアプリケーションを実行した時の性能、ならびにアプリケーションの平均通信間隔と通信回数を入力とし、上記アプリケーションを On/Off リンクを有するネットワーク上で実行した時の性能を出力とする。4 ノードが EEE 対応のネットワークスイッチによって接続されたシステムと NPB を用いて評価した結果、開発したモデルは、EEE 機能を使用していない状態のアプリケーション性能と通信プロファイルから EEE 機能を使用した時の性能を高い精度で予測できることが確認できた。本成果は Springer 社 Journal of Computer Science - Research and Development, Vol. 29, Issue 3-4 に掲載済みである。

リンクオフスレッシュホールドを有する On/Off リンクの電力モデリングを行った。On/Off リンクの通信遅延によるアプリケーションの性能低下を抑えるためにはリンクオフスレッシュホールド(リンクが省電力モードに遷移するまでのタイムアウト時間)が必要だが、On/Off リンクの既存電力モデルはこれを考慮しておらず、リンクオフスレッシュホールドを有する On/Off リンクの消費電力を正確に見積もることができないという問題があった。この問題に対して、既存モデルを拡張して入力変数にリンクオフスレッシュホールドを導入し、リンクオフスレッシュホールドを有する On/Off リンクの電力見積もりが可能なモデルを新たに提案した。EEE 対応のスイッチを介して接続された 2 ノードのシステムにおいて、リンクオフスレッシュホールドならびにリンク負荷を変更してピンポン通信を行った結果、提案モデルは既存モデルの誤差を大幅 (28.6%から 8.6%) に改善できた。この研究成果は、第 155 回 HPC 研究会 (SWoPP 松本) にて発表した。

メモリホットプラグを利用可能なシステムの電力性能モデルを開発した。開発したモデルは、システム構成に依存するいくつかのパラメータ (使用メモリバンド幅、周波数に依存するメモリの消費電力量など)、ならびにアプリケーション内の LLC (Last Level Cache) ミスする命令とそれ以外の命令の比率をもとに、電源投入する DIMM 数を変更した時のアプリケーションの電力性能を予測する。8 コアの Xeon CPU と 12x 4GB DDR3 メモリを搭載したサーバ、ならびに LLC ミスする命令の割合を変更可能なシミュレーションなアプリケーションを用いて評価を行った結果、開発したモデルは平均 2.64%、最大 9.02% と高い精度で電源投入 DIMM 数を変更した時のノード電力を予測できることを確認した。本成果は第 143 回 HPC 研究会にて発表した。

### 3) 電力性能ノブ制御アルゴリズムの検討

CPU と GPU を搭載したヘテロジニアスな実行環境において、各デバイスの周波数と各デバイスに割り振るタスク量を調整することによって最適な電力性能でアプリケーションを実行するアルゴリズムを開発した。開発したアルゴリズムは、前述の CPU/GPU の電力性能モデルを用いてアプリケーションを各周波数と各タスク配分比率で実行した際の電力性能を予測し、電力性能が最適となる周波数とタスク配分比率で上記アプリケーションの実行を行う。CPU と GPU をそれぞれ 1 つ搭載した実行環境において Rodinia Benchmark Suite を用いて評価を行った結果、提案アルゴリズムは理想的な周波数セッティングと理想的なタスク配分比率による実行時の電力性能とほぼ同等の電力性能を達成した。本成果は前述の CPU/GPU の電力性能モデルとともに ICCD'13 にて発表した。

蓄電池と CPU DVFS を利用した電力制約下での性能向上では、蓄電池を用いて時間方向に電力を融通し、電力制約下における性能向上を実現するパワーシフティング手法を開発した。提案手法では、停電時のために設置されている UPS (無停電電源装置) 内の蓄電池と周波数制御を併用し、電力を投入しても性能が上がりにくいフェーズから、電力を投入することで性能が大きく上がるフェーズへ電力を融通することで高い性能を達成する。評価の結果、提案手法を用いることで、従来の周波数制御を用いた電力制約制御手法に対して、CPU アプリケーションで平均 4.5%、GPU アプリケーションで平均 17.1% の性能向上が実現できることを示した。この成果は第 143 回 HPC 研究会にて発表した。

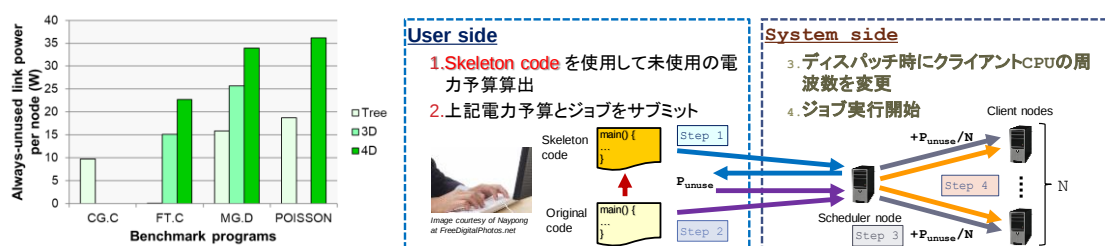


図 17: On/Off リンクを有するネットワークにおける未使用電力予算 (左) と Profile-based Power Shifting の概要 (右)

On/Offリンクを有するネットワークにおける余剰電力を用いて CPU ブースティングを行うことにより、電力制約下におけるアプリケーション性能を改善するアルゴリズムを開発した。図 17 左に示すように On/Offリンクを有するネットワークには未使用の電力予算が存在すること、ならびに上記電力予算はアプリケーションやトポロジによって異なることを我々のグループは突き止めた。そこで、この余剰電力予算をプロファイリングによって求め、プロファイリング結果をもとに、ジョブのローンチ時に、スケジューラが電力制約の範囲内でクライアント CPU の周波数をブースティングすることによりアプリケーション性能を改善する手法 (Profile-based Power Shifting, 図 17 右)を提案した。大規模計算機のシミュレータである SimGrid に独自の改良を施したシミュレータと NPB を用いて評価した結果、提案手法は与えられた電力制約下において最大 1.23 倍の性能向上を達成できることを明らかにした。本成果は SC15 にてフルペーパー発表した。また本研究を通して得た知見をフィードバックし、井上チームの電力性能予測シミュレータ TraceRP の開発に協力した。

メモリホットプラグを用いてアプリケーション実行中の DIMM の待機電力を削減するアルゴリズムを開発した。開発したアルゴリズムは、アプリケーションの使用メモリ量の監視とメモリの制御を行うデーモンプロセスを各ノードに配置し、上記デーモンプロセスがメモリ容量不足を検出すると電源遮断していたメモリの電源を投入することで容量不足を回避する。

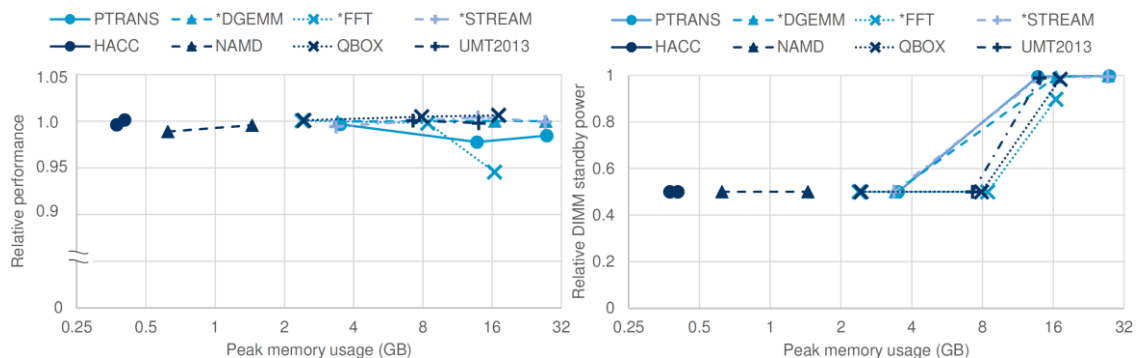


図 18: Footprint-based DIMM Hotplug のジョブ性能 (左) と DIMM 待機電力削減量 (右)。

Xeon CPU を 2 基搭載したノード 1 台と HPC Challenge Benchmark を用いて評価した結果、少ないメモリ量で実行しても多くのアプリケーションは性能低下しないことを確認した。本成果は SC15 にてポスター発表を行い、ベストポスター賞候補にノミネートされた。

さらに上記のメモリホットプラグ制御アルゴリズムの改良を行った。上記の制御アルゴリズムは各ノードにおけるメモリ使用量の監視と DIMM の電源状態の管理をデーモンプロセスが定期的に行っており、この制御オーバーヘッドによりアプリケーション性能が低下することが問題となっていた。そこでメモリ使用量の監視と DIMM の電源状態の管理を OS のページ管理システムに組み込むことにより制御オーバーヘッドを削減した手法 (Footprint-based DIMM Hotplug) を開発した。最大 4 ノードからなる並列計算機システムに上記の改良を施した OS をインストールし、HPC Challenge Benchmark の一部と CORAL Benchmark の一部を用いて評価した結果、提案手法はほとんどのアプリケーションと問題サイズの組み合わせにおいて性能低下が見られなかった(図 18 左)。また、メモリフットプリントの小さなジョブに対しては DIMM の待機電力の 50% を削減できることを確認した(図 18 右)。

#### ★公開予定ソフトウェア: ネットワーク電力プロファイリング用コードを生成するコンパイラ

- 機能: C または C++ で記述されたプログラムを通信パターンのみを再現するプログラムの実行バイナリに変換
- 特徴: 任意の ISA の実行バイナリを生成可能
- 実行可能環境: LLVM 4.0.0 が実行可能な環境
- 潜在的なユーザ: ネットワーク電力や通信パターンのプロファイリングを行うユーザ

★公開予定ソフトウェア: SimGrid-3.11 用 On/Offリンク・シミュレーション・パッチ

- 機能: On/Offリンクのシミュレーション機能を有しない SimGrid-3.11 に On/Offリンクのシミュレーション機能を付加
- 特徴: On/Offリンクの性能見積もりが可能, リンクの消費電力トレースを生成可能
- 実行可能環境: SimGrid-3.11 が実行可能な環境
- 潜在的なユーザ: インターコネクト研究者, 開発者

★公開予定ソフトウェア: Linux Kernel 4.1.34 用 Footprint-based DIMM Hotplug パッチ

- 機能: ジョブ実行中の DIMM 待機電力削減機能を有しない Linux Kernel 4.1.34 に同機能を付加
- 特徴: メモリフットプリントの小さなジョブの実行において DIMM の待機電力を半減することが可能
- 実行可能環境: Linux Kernel 4.1.34 のビルドとインストールが可能な環境
- 潜在的なユーザ: HPC システム運用者, 一般の Linux ユーザ

★エクサスケールへの貢献度

- 新規ノブの開拓: 厳しい電力制約のもとでエクサスケールを実現するため、On/Offリンクとメモリホットプラグという2つの新規電力性能ノブの制御方法を確立し有効性を示すことができた点において、エクサスケールシステムのハードウェア開発に資する。

## § 4 成果発表等

(1)原著論文発表 (国内(和文)誌 5件、国際(欧文)誌 19件)

1. Hiroshi Sasaki, Satoshi Imamura, and Koji Inoue, “Coordinated Power-Performance Optimization in Manycores”, The 22nd International Conference on Parallel Architectures and Compilation Techniques (PACT2013), pp.51-62, 2013.
2. Toshiya Komoda, Shingo Hayashi, Takashi Nakada, Shinobu Miwa and Hiroshi Nakamura, “Power Capping of CPU-GPU Heterogeneous Systems through Coordinating DVFS and Task Mapping”, The 31st IEEE International Conference on Computer Design (ICCD’13), pp.349-356, 2013.
3. 本田宏明, 稲富雄一, 眞木淳, “SIMD 演算の効率的利用に向けたアレイジョブ型 Hartree-Fock 法計算”, 情報処理学会論文誌コンピューティングシステム, Vol.7, No.4, pp.15-24, 2014.
4. Shinobu Miwa, Sho Aita and Hiroshi Nakamura, “Performance Estimation for High Performance Computing Systems with Energy Efficient Ethernet Technology”, Journal of Computer Science – Research and Development, Vol.29, Issue 3-4, pp.161-169, Springer, 2014.
5. 三輪忍, 會田翔, 安島雄一郎, 清水俊幸, 安里彰, 中村宏, “実 HPC 環境における EEE の電力/性能評価”, 情報処理学会論文誌コンピューティングシステム, Vol.7, No.4, pp.67-83, 2014.
6. Yuki Abe, Hiroshi Sasaki, Shinpei Kato, Koji Inoue, Masato Eda, and Martin Peres, “Power and Performance Characterization and Modeling of GPU-accelerated Systems”, The 28th IEEE International Parallel & Distributed Processing Symposium, pp.113-122, May 2014.
7. Keiichiro Fukazawa, Tomonori Tsuchida, Kyohei Yoshida, Aruta Uehara, Masakazu Kuze, Masatsugu Ueda, Yuichi Inadomi, Koji Inoue and Mutsumi Aoyagi, “Performance and Power Consumption Evaluation of MHD Simulation for Magnetosphere on Parallel Computer System with CPU Power Capping”, Extreme Green & Energy Efficiency in

- Large Scale Distributed Systems, May 2014.
8. Shinobu Miwa and Charles R. Lefurgy, “Evaluation of Core Hopping on POWER7”, ACM SIGMETRICS Performance Evaluation Review, Special Issue on Greenmetrics 2014, pp.11-16, 2014.
  9. Satoshi Imamura, Hiroshi Sasaki, Koji Inoue, and Dimitrios S. Nikolopoulos, “Power-Capped DVFS and Thread Allocation with ANN Models on Modern NUMA Systems”, International Conference on Computer Design (ICCD2014), pp.324-331, 2014.
  10. 黄巍, 岩澤直弘, カオタン, 和遠, 近藤正章, 中村宏 (東大), “エネルギー効率を考慮した電力制約下でのスループット指向ジョブスケジューリング”, 2015 年ハイパフォーマンส์コンピューティングと計算科学シンポジウム, 2015.
  11. 大坂隼平, 和田康孝, 近藤正章, 三吉郁夫, 本多弘樹, “電力性能特性ばらつきを考慮した MPI 並列アプリケーションの性能最適化”, 2015 年ハイパフォーマンส์コンピューティングと計算科学シンポジウム, 2015.
  12. Yuan He, Masaaki Kondo, Takashi Nakada, Hiroshi Sasaki, Shinobu Miwa, and Hiroshi Nakamura, “Runtime Multi-Optimizations for Energy Efficient On-chip Interconnections”, The 33rd IEEE International Conference on Computer Design (ICCD2015) (Accepted as poster presentation), 2013 (to appear).
  13. Shinobu Miwa and Hiroshi Nakamura, “Profile-Based Power Shifting in Interconnection Networks with On/Off Links”, The International Conference for High Performance Computing, Networking, Storage and Analysis (SC15), 2015.
  14. Yuichi Inadomi, Tapasya Patki, Koji Inoue, Mutsumi Aoyagi, Barry Rountree, Martin Schulz, David Lowenthal, Yasutaka Wada, Keiichiro Fukazawa, Masatsugu Ueda, Masaaki Kondo, Ikuo Miyoshi, “Analyzing and Mitigating the Impact of Manufacturing Variability in Power-Constrained Supercomputing”, The International Conference for High Performance Computing, Networking, Storage and Analysis (SC15), 2015.
  15. Yoshihiro Tanaka, Keitaro Oka, Takatsugu Ono, Koji Inoue, “Accuracy Analysis of Machine Learning-Based Performance Modeling for Microprocessors,” In Proc. of the 2016 Fourth International Japan-Egypt Conference on Electronics, Communications and Computers (JEC-ECC '16), pp.87-90, May 2016.
  16. Thang Cao, Yuan. He, Masaaki. Kondo, “Demand-Aware Power Management for Power-Constrained HPC Systems”, The 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid2016), pp.21-31, May 2016.
  17. Yuan He and Masaaki Kondo, “Opportunistic Circuit-Switching for Energy Efficient On-Chip Networks”, The 24th IFIP/IEEE International Conference on Very Large Scale Integration (VLSI-SoC 2016), Sep. 2016.
  18. S. Imamura, Y. Yasui, K. Inoue, T. Ono, H. Sasaki, K. Fujisawa, “Power-Efficient Breadth-First Search with DRAM Row Buffer Locality-Aware Address Mapping”, 1st High Performance Graph Data Management and Processing workshop (HPGDMP), Nov. 2016.
  19. S. Imamura, K. Oka, Y. Yasui, Y. Inadomi, K. Fujisawa, T. Endo, K. Ueno, K. Fukazawa, N. Hata, Y. Kakibuka, K. Inoue, T. Ono, “Evaluating the Impacts of Code-Level Performance Tunings on Power Efficiency”, IEEE International Conf. on Big Data, pp.362-369, Dec. 2016.
  20. R. Sakamoto, T. Cao, M. Kondo, K. Inoue, M. Ueda, T. Patki, D. Ellsworth, B. Rountree, and M. Schulz, “Production Hardware Overprovisioning: Real-world Performance Optimization using an Extensible Power-aware Resource Management Framework”, 31st IEEE International Parallel & Distributed Processing Symposium (IPDPS 2017). May 2017.

21. T. Cao, W. Huang, Y. He and M. Kondo, “Cooling-Aware Job Scheduling and Node Allocation for Overprovisioned HPC Systems”, 31st IEEE International Parallel & Distributed Processing Symposium (IPDPS 2017). May 2017.
22. Teruo Tanimoto, Takatsugu Ono, Koji Inoue, Hiroshi Sasaki, “Enhanced Dependence Graph Model for Critical Path Analysis on Modern Out-of-Order Processors”, IEEE Computer Architecture Letters, Vol. PP, No 99, Mar. 2017.
23. Teruo Tanimoto, Takatsugu Ono, and Koji Inoue, “Dependence Graph Model for Accurate Critical Path Analysis on Out-of-Order Processors”, Journal of Information Processing, Vol. 25, pp.983–992, Dec. 2017.
24. Teruo Tanimoto, Takatsugu Ono and Koji Inoue, “CPCI Stack: Metric for Accurate Bottleneck Analysis on OoO Microprocessors”, The Fifth International Symposium on Computing and Networking, pp. 166–172, Nov. 2017.
25. Yasutaka Wada, Yuan He, Thang Cao, Masaaki Kondo, “A Power Management Framework with Simple DSL for Automatic Power-Performance Optimization on Power-Constrained HPC Systems”, SupercomputingAsia 2018 (SCA18), Mar. 2018 (to appear).
26. Ryuichi Sakamoto, Tapasya Patki, Thang Cao, Masaaki Kondo, Koji Inoue, Masatsugu Ueda, Daniel Ellsworth, Barry Rountree, and Martin Schulz, “Analyzing Resource Trade-offs in Hardware Overprovisioned Supercomputers”, 32nd IEEE International Parallel & Distributed Processing Symposium (IPDPS2018), May 2018 (to appear).

(2)その他の著作物(総説、書籍など)

1. 稲富雄一、井上弘士, “並列分子科学計算プログラムの電力制約下における性能最適化”, 日本コンピュータ化学会学会誌(Journal of Computer Chemistry, Japan), Vol. 14, No. 6, pp.201–202, 2015.
2. 近藤正章, “HPC システムのための電力管理技術への取り組み”, 日本シミュレーション学会学会誌「シミュレーション」, Vol. 36, No. 2, pp.85–88, 2017.

(3)国際学会発表及び主要な国内学会発表

① 招待講演 (国内会議 3件、国際会議18件)

1. 近藤正章, “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”, ポストペタスケールコンピューティングワークショップ, 東京都目黒区, 2012年8月31日.
2. Masaaki. Kondo and Hiroshi Nakamura, “Power Management Framework for Post-Petascale Supercomputing”, LLNL Seminar, Livermore CA, March 27, 2013.
3. Koji Inoue, “PSI Project: Performance Prediction and Analysis of Supercomputers”, LLNL Seminar, Livermore CA, March 27, 2013.
4. 近藤正章, “エクサスケールで顕在化する Power Wall 問題 ～現状と今後の打開策～”, サイエンティフィック・システム研究会科学技術計算分科会, 神戸, 2013年10月23日.
5. K. Inoue, M. Schulz, M. Kondo, H. Nakamura, T. Gamblin, and B. Rountree, “Power Management and Optimization toward Trans-Petascale Supercomputing”, Workshop on International Cooperation on System Software for Trans-Petascale, Leipzig, June 19, 2013.
6. Masaaki Kondo, “Power Wall: A Design Challenge to Exascale Systems”, Extreme Performance Computational Science French-Japanese Conference, Tokyo, April 14, 2014.
7. Hiroshi Nakamura, Masaaki Kondo, Koji Inoue, Martin Schulz, Todd Gamblin, Barry

- Rountree, Toshio Endo, Akira Nukada, Satoshi Matsuoka, “Power Management and Optimization toward Extreme Scale Supercomputing”, Workshop on International Cooperation for Extreme-Scale Computing, Leipzig, June 22, 2014.
8. Koji Inoue, “CPU-Memory Power Shifting on High Performance Computing,” 14th International Forum on Embedded MPSoC and Multicore (MPSoC), Margaux, July 9, 2014.
  9. Masaaki Kondo, “Power Management Framework for Post-Petascale Supercomputers”, 2014 ATIP Workshop: Japanese Research Toward Next-Generation Extreme Computing, New Orleans, Nov. 17, 2014.
  10. K. Inoue, Y. Abe, Y. Inadomi, K. Isaacs, T. Gamblin, and M. Schulz, “NsimPower: A Power Analyzing Environment for Large Scale Interconnection Network”, JST CREST International Symposium on Post Petascale System Software, Kobe, Dec 3, 2014.
  11. Shinobu Miwa and Hiroshi Nakamura, “Power Shifting between Networks and CPUs in HPC System”, JST CREST International Symposium on Post Petascale System Software, Kobe, Dec 3, 2014.
  12. Y. Inadomi, Y. Wada, M. Kondo, K. Inoue, B. L. Rountree, and M. Schulz, “Power-Performance Optimization for Power-Constrained Supercomputer Systems”, JST CREST International Symposium on Post Petascale System Software, Kobe, Dec 3, 2014.
  13. Masaaki Kondo, “Power Management Framework for Extreme-Scale Computing”, The workshop on Big Data and Extreme-scale Computing (BDEC), Barcelona, Jan. 29, 2015.
  14. Koji Inoue, “NsimPower: Interconnect Simulator for Power and Performance Prediction,” International Forum on MPSoC for Software-defined Hardware (MPSoC), Ventura, July 15, 2015.
  15. 近藤正章, “エクサスケールシステム向け電力制御技術の開発”, STE シミュレーション研究会, 京都, 2015 年 9 月 29 日.
  16. Koji Inoue, Yuichi Inadomi, and Takatsugu Ono, “Challenges in Power Constrained High Performance Computing”, Organized Session for Green HPC, Annual Meeting on Advanced Computing System and Infrastructure, Fukuoka, Jan. 2016.
  17. Satoshi Matsuoka, Hideharu Amano, Kengo Nakajima, Koji Inoue, Tomohiro Kudoh, Naoya Maruyama, Kenjiro Taura, Takeshi Iwashita, Takahiro Katagiri, Toshihiro Hanawa, Toshio Endo, “From FLOPS to BYTES: Disruptive Change in High-Performance Computing towards the Post-Moore Era”, Invited, International Conference on Computing Frontiers (CF’16), pp. 274-281, May 2016.
  18. Koji Inoue, “Impact of Manufacturing Variability in Power Constrained Supercomputing”, International Forum on MPSoC for Software-defined Hardware (MPSoC), Nara, July 11, 2016.
  19. Masaaki Kondo, “Energy Efficient Network-on-Chip with Runtime Optimization Selection for MPSoCs”, International Forum on MPSoC for Software-defined Hardware (MPSoC), Nara, July 11, 2016.
  20. Masaaki Kondo, “Research and Development of Power Management Framework for Exascale Computing”, The 35th JSST Annual Conference International Conference on Simulation Technology (JSST2016), Kyoto, Oct. 28, 2016.
  21. Masaaki Kondo, “Energy Efficient Network-on-Chips with Opportunistic Circuit-Switching for MPSoCs”, International Forum on MPSoC for Software-defined Hardware (MPSoC), Annecy, July 3, 2017.

② 口頭発表 (国内会議42件、国際会議 0件)

1. 三輪忍, 會田翔, 安島雄一郎, 清水俊幸, 安里彰, 中村宏, “FX10 におけるインタコネクト・コントローラの省電力化手法の初期検討”, 情報処理学会計算機アーキテクチャ研究会／ハイパフォーマンส์コンピューティング研究会, 北海道札幌市, 2012 年 12 月 13 日.
2. 近藤正章, “電力マネージメントフレームワークの構築と標準化”, 第 8 回戦略的高性能計算システム開発に関するワークショップ, 東京都千代田区, 2013 年 2 月 2 日.
3. 松本洋平, 近藤正章, 和田康孝, 本多弘樹, “GPU における細粒度パワーゲーティング向けスレッド発行制御手法の検討”, 情報処理学会計算機アーキテクチャ研究会, 和歌山県田辺市, 2013 年 3 月 26 日.
4. 稲富雄一, 梅田宏明, 辻美和子, 村井均, 南一生, 横川三津夫, 佐藤三久, 青柳睦, “耐障害性を考慮した並列 FMO プログラムの実装,” 日本コンピュータ化学会年会講演予稿集, pp.37-38, 東京, 2013 年 5 月 30 日.
5. 會田翔, 三輪忍, 中村宏, “電力制約下における CPU とネットワークの電力制御協調手法”, 情報処理学会ハイパフォーマンส์コンピューティング研究会 2013-HPC-140, 北九州, 2013 年 7 月 31 日.
6. 深沢圭一郎, 岡慶太郎, “電磁流体コードを用いた Xeon Phi の性能評価”, 情報処理学会ハイパフォーマンส์コンピューティング研究会 2013-HPC-141, 那覇, 2013 年 9 月 30 日.
7. 吉田匡兵, 佐々木広, 深沢圭一郎, 稲富雄一, 上田将嗣, 井上弘士, 青柳睦, “CPU と主記憶への電力バジェット配分を考慮した HPC アプリケーションの性能評価”, 情報処理学会ハイパフォーマンส์コンピューティング研究会 2013-HPC-141, 那覇, 2013 年 10 月 1 日.
8. カオタン, 和田康孝, 近藤正章, 本多弘樹, “RAPL インタフェースを用いた HPC システムの消費電力モデリングと電力評価”, 情報処理学会ハイパフォーマンส์コンピューティング研究会 2013-HPC-141, 那覇, 2013 年 10 月 1 日.
9. 藤原祐太, 松本洋平, 和田康孝, 近藤正章, 本多弘樹, “使用コア数最適化と DVFS を用いた GPU の省電力化手法の検討”, 情報処理学会計算機アーキテクチャ研究会 & ハイパフォーマンส์コンピューティング研究会 2013-ARC-207&2013-HPC-142, 札幌, 2013 年 12 月 16 日.
10. 松本洋平, 和田康孝, 近藤正章, 本多弘樹, “GPU における走行時パワーゲーティング向けスレッドブロック割り当ておよびワープ発行制御手法”, 情報処理学会計算機アーキテクチャ研究会 & ハイパフォーマンส์コンピューティング研究会 2013-ARC-207&2013-HPC-142, 札幌, 2013 年 12 月 16 日.
11. 稲富雄一, 吉田匡兵, 深沢圭一郎, 上田将嗣, 青柳睦, 井上弘士, “電力指向型次世代スーパーコンピュータを想定した HPC アプリケーションの性能最適化～量子化学計算の場合～”, 情報処理学会計算機アーキテクチャ研究会 & ハイパフォーマンส์コンピューティング研究会 2013-ARC-207&2013-HPC-142, 札幌, 2013 年 12 月 17 日.
12. 和田康孝, カオタン, 近藤正章, 本多弘樹, “HPC システムにおける電力・性能可視化ツールに関する比較検討”, 情報処理学会計算機アーキテクチャ研究会 & ハイパフォーマンส์コンピューティング研究会 2013-ARC-207&2013-HPC-142, 札幌, 2013 年 12 月 17 日.
13. 上原有太, 稲富雄一, 井上弘士, “電力制約下での性能向上を目的とした大規模並列計算機システム向け稼働計算ノード数決定法の提案,” 情報処理学会火の国シンポジウム, 大分, 2014 年 3 月 4 日.
14. 會田翔, 三輪忍, 中村宏, “ロードバランスを考慮した電力制約下における CPU の DVFS 制御”, 情報処理学会ハイパフォーマンส์コンピューティング研究会 2014-HPC-143, 七尾, 2014 年 3 月 4 日.
15. 米澤亮太, 會田翔, 三輪忍, 中村宏, “物理メモリの増減による電力制約下での

- HPC システムの性能向上”, 情報処理学会ハイパフォーマンズコンピューティング研究会 2014-HPC-143, 七尾, 2014 年 3 月 4 日.
16. 酒井崇至, 薦田登志矢, 三輪忍, 中村宏, “電力制約下における蓄電池を用いた HPC システムの性能向上”, 情報処理学会ハイパフォーマンズコンピューティング研究会 2014-HPC-143, 七尾, 2014 年 3 月 4 日.
  17. 阿部祐希, 稗田拓路, 稲富雄一, 柴村英智, 眞木淳, 小林泰三, 深沢圭一郎, 井上弘士, “大規模インターコネクトを対象とした消費電力解析環境 NsimPower の構築”, 情報処理学会ハイパフォーマンズコンピューティング研究会 2013-HPC-145, 新潟, 2014 年 7 月 28 日.
  18. 和田康孝, 稲富雄一, 井上弘士, 近藤正章, 本多弘樹, “高性能計算環境向け電力配分自動最適化のためのコンパイラ環境の構築”, 情報処理学会ハイパフォーマンズコンピューティング研究会 2014-HPC-145, 新潟, 2014 年 7 月 28 日.
  19. 三輪忍, 塩谷亮太, 佐々木広, “ダーク・シリコン時代のプロセッサ・アーキテクチャに関する初期検討”, 情報処理学会計算機アーキテクチャ研究会 2014-ARC-211, 新潟, 2014 年 7 月 28 日.
  20. 黄巍, 岩澤直弘, カオタン, 和遠, 近藤正章, 中村宏, “エネルギー効率を考慮した電力制約下でのスループット指向ジョブスケジューリング手法”, 情報処理学会ハイパフォーマンズコンピューティング研究会 2013-HPC-146, 那覇, 2014 年 10 月 2 日.
  21. 稲富雄一, 和田康孝, 深沢圭一郎, 青柳睦, 近藤正章, 三吉郁夫, 井上弘士, “MPI 並列アプリケーションの電力性能最適化”, 情報処理学会ハイパフォーマンズコンピューティング研究会 2013-HPC-146, 那覇, 2014 年 10 月 2 日.
  22. 深沢圭一郎, 青柳睦, 津秦伴紀, 吉田匡平, 稲富雄一, 井上弘士, “CPU と DRAM の消費電力制限下での MHD シミュレーションコードの電力特性”, 情報処理学会ハイパフォーマンズコンピューティング研究会 2013-HPC-146, 那覇, 2014 年 10 月 2 日.
  23. 三輪忍, 塩谷亮太, 佐々木広, “回路資源の投入により電力効率を改善するプロセッサ・アーキテクチャ”, 情報処理学会計算機アーキテクチャ研究会 2014-ARC-212, 別府, 2014 年 10 月 6 日.
  24. 稲富雄一, 和田康孝, 深沢圭一郎, 青柳睦, 近藤正章, 三吉郁夫, 井上弘士, “電力性能特性ばらつきを考慮した MPI 並列アプリケーションの性能最適化”, 情報処理学会計算機アーキテクチャ研究会 & ハイパフォーマンズコンピューティング研究会 2013-ARC-213 & 2013-HPC-147, 小樽, 2014 年 12 月 9 日.
  25. 大坂隼平, 和田康孝, 近藤正章, 三吉郁夫, 本多弘樹, “HPC システムにおける性能プロファイリングツールの電力測定精度の評価”, 情報処理学会計算機アーキテクチャ研究会 & ハイパフォーマンズコンピューティング研究会 2013-ARC-213 & 2013-HPC-147, 小樽, 2014 年 12 月 9 日.
  26. 佐々木沢, 和田康孝, 近藤正章, 本多弘樹, “次世代 3 次元実装メモリのメモリネットワーク構成に関する初期検討”, 情報処理学会計算機アーキテクチャ研究会 & ハイパフォーマンズコンピューティング研究会 2013-ARC-213 & 2013-HPC-147, 小樽, 2014 年 12 月 9 日.
  27. 稲富雄一, 井上弘士, 和田康孝, 深沢圭一郎, 上田将嗣, 近藤正章, 三吉郁夫, Barry Rountree, Martin Schulz, Tapasya Patki, David Lowenthal, 青柳睦, “電力制約スーパーコンピューティングにおける製造ばらつき問題とその対策～大規模計算機システムを対象とした電力バジェット配分法の提案～”, 情報処理学会ハイパフォーマンズコンピューティング研究会 2015-HPC-150, 別府, 2015 年 8 月 5 日.
  28. 坂本龍一, カオタン, 和遠, 近藤正章, 深沢圭一郎, 上田将嗣, 稲富雄一, 井上弘士, “電力制約を考慮した資源管理を行うリソースマネージャの実装と評価”, 情報処理学会ハイパフォーマンズコンピューティング研究会, 2015-HPC-151, 那覇, 2015 年 9 月 30 日.

29. 坂本龍一, カオタン, 和遠, 近藤正章, 深沢圭一郎, 上田将嗣, 稲富雄一, 井上弘士, “電力制約を考慮した資源管理ツールによる HPC システムの電力性能解析”, 情報処理学会ハイパフォーマンส์コンピューティング研究会, 2016-HPC-155, 松本, 2016 年 8 月 9 日.
30. 谷本輝夫, 佐々木広, 小野貴継, 井上弘士, “アウトオブオーダー命令実行の依存グラフ表現に関する考察”, 情報処理学会計算機アーキテクチャ研究会, 2016-ARC-221, 松本, 2016 年 8 月 8 日.
31. 垣深悠太, 安井雄一郎, 小野貴継, 稲富雄一, 藤澤克樹, 井上弘士, “CPU と DRAM への電力バジェット配分を考慮した Graph500 の性能評価”, 情報処理学会ハイパフォーマンส์コンピューティング研究会, 2016-HPC-155, 松本, 2016 年 8 月 9 日.
32. 稲富雄一, 垣深悠太, 小野貴継, 井上弘士, “電力制約型スーパーコンピュータにおける性能モデリング”, 情報処理学会ハイパフォーマンส์コンピューティング研究会, 2016-HPC-155, 松本, 2016 年 8 月 9 日.
33. 今村智史, 安井雄一郎, 稲富雄一, 藤澤克樹, 井上弘士, 小野貴継, “コードレベル性能最適化が電力効率に与える影響の分析”, 情報処理学会ハイパフォーマンส์コンピューティング研究会, 2016-HPC-155, 松本, 2016 年 8 月 9 日.
34. 西郷雄斗, 三輪忍, 八巻隼人, 本多弘樹, “リンクオフスレッシュホールドを有する ON/OFF リンクの電力見積手法の初期検討”, 情報処理学会ハイパフォーマンส์コンピューティング研究会, 2016-HPC-155, 松本, 2016 年 8 月 9 日.
35. 石原雅也, 三輪忍, 八巻隼人, 本多弘樹, “メモリホットプラグを用いたメインメモリの省電力化に関する初期検討”, 情報処理学会ハイパフォーマンส์コンピューティング研究会, 2016-HPC-155, 松本, 2016 年 8 月 9 日.
36. 澁谷俊憲, 三輪忍, 塩谷亮太, 佐々木広, 八巻隼人, 本多弘樹, “ヘテロジニアス・プロセッサの設計探索手法の初期検討”, 情報処理学会計算機アーキテクチャ研究会, 2016-ARC-221, 松本, 2016 年 8 月.
37. 大場百香, 三輪忍, 進藤智司, 津邑公暁, 八巻隼人, 本多弘樹, “再構成可能なニューラルネットワークアクセラレータの提案と性能分析”, 情報処理学会計算機アーキテクチャ研究会, 2016-ARC-221, 松本, 2016 年 8 月 10 日.
38. 石原雅也, 三輪忍, 八巻隼人, 本多弘樹, “ジョブ実行中の計算ノードにおける DIMM 待機電力削減手法の実装と評価”, 情報処理学会ハイパフォーマンส์コンピューティング研究会, 2017-HPC-158, 熱海, 2017 年 2 月.
39. 大場百香, 三輪忍, 進藤智司, 津邑公暁, 八巻隼人, 本多弘樹, “マルチコアニューラルネットワークアクセラレータにおけるデータ転送のブロードキャスト化”, 情報処理学会計算機アーキテクチャ研究会, 2017-ARC-225, 久米島, 2017 年 3 月 9 日.
40. 坂本龍一, カオタン, 近藤正章, 井上弘士, 上田将嗣, Tapasya Patki, Daniel Ellsworth, Barry Rountree, Martin Schulz, “オーバープロビジョニング環境での大規模 HPC システムの電力と性能評価”, 情報処理学会ハイパフォーマンส์コンピューティング研究会, 2017-HPC-160, 秋田, 2017 年 7 月 29 日.
41. 小野貴継, 垣深悠太, 三輪忍, 井上弘士, “電力性能推定を目的としたインターコネクタ・シミュレータ TraceRP の開発”, 情報処理学会ハイパフォーマンส์コンピューティング研究会, HPC-161, 函館, 2017 年 9 月 (発表予定).
42. 三吉郁夫, 三輪忍, 井上弘士, 近藤正章, “DFS/DCT 制御による電力あたり性能の実行時最適化”, 情報処理学会ハイパフォーマンส์コンピューティング研究会, HPC-163, 松山, 2018 年 2 月 28 日.

③ ポスター発表 (国内会議 6 件、国際会議 9 件)

1. Y. Abe, H. Sasaki, S. Kato, K. Inoue, M. Edahiro, and M. Peres, “Power and

- Performance of GPU-accelerated Systems: A Closer Look,” IEEE International Symposium on Workload Characterization, Portland, Sep. 23, 2013.
2. 稲富雄一, 深沢圭一郎, 井上弘士, “電力制約下における分子積分プログラムの性能最適化”, 日本コンピュータ化学会 2014 春季年会, 2014 年 5 月.
  3. 稲富雄一, 深沢圭一郎, 井上弘士, 青柳睦, “並列 FMO プログラム OpenFMO の電力性能最適化”, コンピュータ化学会 2014 秋季年会, 2014 年 10 月.
  4. 岡慶太郎, 井上優良, 井上弘士, “ソースコード移植を伴わない GPU 性能推定手法に関する研究”, 情報処理学会計算機アーキテクチャ研究会ポスターセッション, 別府, 2014 年 10 月 6 日.
  5. 稲富雄一, “スパコンの消費電力問題とアプリケーションプログラムの電力性能最適化,” 第2回 CUTE シンポジウム「コンピュータ化学」, 2014 年 10 月.
  6. T. Cao, M. Kondo, Y. Wada, H. Honda, “Modeling Power Usage of HPC Systems by RAPL Interface”, HPC in Asia Posters at ISC’14, Leipzig, June 26, 2014.
  7. Masaaki Kondo, Thang CAO, Yuan He, Yasutaka Wada, Hiroki Honda, Ikuo Miyoshi, Yuichi Inadomi, Keiichiro Fukazawa, Koji Inoue, Shinobu Miwa, Hiroshi Nakamura, “Power Management Framework for Post-Petascale Supercomputers”, JST CREST International Symposium on Post Petascale System Software (Poster Session), Kobe, Dec. 2, 2014.
  8. K. Inoue, Y. Abe, Y. Inadomi, K. Isaacs, T. Gamblin, and M. Schulz, “NsimPower: Large Scale Interconnection Network Simulator for Power/Performance Analysis”, JST CREST International Symposium on Post Petascale System Software (Poster Session), Kobe, Dec. 2, 2014.
  9. Y. Inadomi, Y. Wada, M. Kondo, K. Inoue, B. L. Rountree, and M. Schulz, “Power-Performance Optimization for Power-Constrained Supercomputer Systems,” JST CREST International Symposium on Post Petascale System Software (Poster Session), Kobe, Dec. 2, 2014.
  10. Thang Cao, Huang Wei, Naohiro Iwasawa, Yuan He, Masaaki Kondo, “Job Scheduling and Resource Management for Power-Constrained Supercomputers”, JST CREST International Symposium on Post Petascale System Software (Poster Session), Kobe, Dec. 2, 2014.
  11. K. Oka, W. Jia, M. Martonosi, and K. Inoue, “Characterization and Cross-Platform Analysis of High-Throughput Accelerators”, International Symposium on Performance Analysis of Systems and Software (Poster), Philadelphia, Mar. 30 2015.
  12. S. Miwa, and H. Honda, “Memory Hotplug for Energy Savings of HPC systems”, The International Conference for High Performance Computing, Networking, Storage and Analysis (SC’15, poster), 2015.
  13. M. Ohba, S. Miwa, S. Shindo, T. Tsumura, H. Yamaki, and H. Honda, Initial Study of Reconfigurable Neural Network Accelerators, The 7th International Workshop on Advances in Networking and Computing (poster presentation), pp.707-709, (Nov 2016).
  14. Yasutaka Wada, Yuan He, Thang Cao, Masaaki Kondo, “The PomPP Framework: From Simple DSL to Sophisticated Power Management for HPC Systems”, International Conference on High Performance Computing in Asia-Pacific Region (HPCAsia2018), Jan. 2018.
  15. Ikuo Miyoshi, Shinobu Miwa, Koji Inoue, Masaaki Kondo, “Run-Time DFS/DCT Optimization for Power-Constrained HPC Systems”, International Conference on High Performance Computing in Asia-Pacific Region (HPCAsia2018), Jan. 2018.

(4)知財出願

①国内出願 (1 件)

1. 発明の名称:「情報処理装置、消費電力計測プログラムおよび消費電力計測方法」、  
発明者:三吉 郁夫、出願人:富士通株式会社、出願日:2015 年 4 月 16 日、出願番  
号:2015-084113

②海外出願 (0件)

③その他の知的財産権

(5)受賞・報道等

①受賞

- ・地球電磁気・地球惑星圏学会大林奨励賞, 深沢圭一郎, 2014 年 11 月 2 日.
- ・情報処理学会 システム・アーキテクチャ研究会若手奨励賞, 谷本輝夫, 2016 年.
- ・電子情報通信学会 コンピュータシステム研究専門委員会 2016 年度研究会優秀若手講  
演賞, 谷本輝夫, 2017 年.

②マスコミ(新聞・TV等)報道

③その他

(6)成果展開事例

①実用化に向けての展開

- ・ 本研究課題の成果や知見をポスト「京」開発事業へと展開するべく、理研内部の打ち合  
わせ等に参加して意見交換を行っている。

②社会還元的な展開活動

- ・ 本研究成果の一部をインターネット(URL:  
<http://www.hal.ipc.i.u-tokyo.ac.jp/research/pompp/>)で公開し、一般に情報提供している。

## § 5 研究期間中の活動

### 5. 1 主なワークショップ、シンポジウム、アウトリーチ等の活動

| 年月日                 | 名称                          | 場所                                        | 参加人数        | 概要                      |
|---------------------|-----------------------------|-------------------------------------------|-------------|-------------------------|
| 2013 年<br>7 月 24 日  | 研究総括、アドバイザー<br>サイトビジット      | 電気通信大学                                    | 13 人        | 研究進捗報告ミーティング            |
| 2014 年<br>9 月 19 日  | システム情報談話会                   | 東京大学                                      | 50 人        | 学術交流                    |
| 2015 年<br>6 月 24 日  | 研究総括、アドバイザー<br>サイトビジット      | 東京大学                                      | 14 人        | 研究進捗確認                  |
| 2016 年<br>3 月 30 日  | 国際連携<br>研究打ち合わせ             | Lawrence<br>Livermore<br>National<br>Lab. | 13 人        | 技術交流と国際連携に<br>関するミーティング |
| 2016 年<br>11 月 13 日 | SC' 16「ポストペタスケ<br>ール高性能計算に資 | Salt Palace<br>Convention                 | 100 人<br>程度 | 成果発表、および<br>学術交流        |

|                     |                                                        |                        |         |        |
|---------------------|--------------------------------------------------------|------------------------|---------|--------|
|                     | するシステムソフトウェア技術の創出」研究領域研究展示                             | Center, Salt Lake City |         |        |
| 2016 年<br>12 月 13 日 | 「ポストペタスケール高性能計算に資するシステムソフトウェア技術の創出」研究領域 28 年度公開ワークショップ | 秋葉原コンベンションホール          | 150 人程度 | 成果発表   |
| 2017 年<br>7 月 11 日  | 研究総括、アドバイザーサイトビジット                                     | 東京大学                   | 13 人    | 研究進捗確認 |

## § 6 最後に

当研究課題を遂行にあたり、研究領域の戦略目標にも掲げられている基盤ソフトウェアの開発や当初に設定した電力制約下での性能最適化の数値目標など、マイルストーンに関しては十分に達成でき、さらには SC や IPDPS をはじめとして PACT、ICCD など、当分野におけるメジャーな国際会議で複数の論文発表の機会を得ることができたことから、本プロジェクトでは非常に良好な研究成果が得られたと自負している。

さらに、技術的な面のみならず、国際連携などを通じて国際的な研究者のネットワークも構築することができた。米国ローレンス・リバモア国立研究所との共同での論文執筆や、学生や若手研究者の派遣といった実質的なコラボレーションを成功させてきたことも本プロジェクトの大きな成果であると考えている。

近年の関連学会等の論文発表動向等を鑑みるに、超並列高性能計算システムにおける消費電力制御や消費電力を考慮したアプリケーション開発の重要性は高まるばかりである。ポストペタスケールシステムを見据えたシステムソフトウェアの創出を目的とする本研究領域において、電力制御を主たる研究課題としてとらえつつ、電力マネジメントフレームワークの開発を目指す本プロジェクトの成果は、ポストペタスケールシステム設計へのフィードバックなどを通して学術的なイノベーションに大きく貢献すると期待している。実際に、研究代表者は「ポスト京プロジェクト」の電力制御ワーキンググループ等へ参加し、システムの電力管理や電力制御などに関して本研究プロジェクトで得られた知見や成果をもとに、意見交換やアイデアの提供を行っている。このことから、本研究プロジェクトの社会的・学術的な貢献度は大きかったと考えている。

本プロジェクトの成果は、技術的なアドバイスを始め、国際連携の場の提供や API 等に関する議論、「ポスト京プロジェクト」参加により発言の機会を頂いたことなど、研究統括やアドバイザーの先生方からのサポートなしにはなし得なかったものであり、厚くお礼を申し上げたい。

最後に、当研究領域のアドバイザーを務められておられた故青柳睦先生には、本プロジェクトの立ち上げ当初から、技術面・精神面で厚いサポートを頂いた。九州大学の大規模高性能計算機システム上での電力解析・電力制御評価に基づいて電力性能特性のばらつきを考慮した電力制御の必要性を世界で初めて指摘でき、大規模システム上に開発した電力資源管理機能付き Slurm をインストールし世界初となる大規模システム上でのオーバープロビジョンドシステムの性能解析が行えたのも、青柳睦先生のご尽力があったからこそ実現できたものである。この場をお借りし、心より感謝申し上げます。