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時代のメモリ階層の
深化に対応するソフトウェア技術」

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：遠藤　敏夫
（東京工業大学学術国際情報センター、
准教授）

§1 研究実施の概要

(1)実施概要

今後のコンピュータアーキテクチャにおいては、メモリウォール問題の悪化のために、高バンド幅と大容量の間のトレードオフの影響がより激しくなる。本チームの目的は、単一メモリ階層のみの利用ではより困難となる、科学技術計算・シミュレーションの高速化かつ大規模化の両立を、アプリケーションプログラマの手間を抑制しつつ実現することである。そのためにはアルゴリズム層・システムソフトウェア/ツール層・アーキテクチャ層にまたがったコデザインが必要であり、特にチームとしてはシステムソフトウェア/ツール層と一部のアルゴリズム層を中心に研究を行う。システムソフトウェア層においては、深化するメモリ階層を効率的に利用するランタイムライブラリ(遠藤グループ・緑川グループ)、動的メモリプロファイラやチューニングツール(遠藤グループ)についての研究を推進している。アルゴリズム層については、主にステンシル計算を中心とし局所性大幅向上技術の研究を行っている。現在までの結果として、グループ間連携を通し GPU デバイスメモリ・ホストメモリ・Flash SSD からなるメモリ階層(ノードあたり 1TB 級)を、局所性向上技術とランタイム技術により、ステンシル計算から効率的に利用可能であることを実証した。

(2)顕著な成果

＜優れた基礎研究としての成果＞

1.

概要:

本チームのメモリ階層活用技術を、ステンシルベースの実アプリケーションである都市気流シミュレーションコードに、ポストペタ丸山直也チームとの協働で適用することにより以下の成果を得た。このアプリは多数 GPU 環境向けに実装されたもので、速度性能は高いが、問題領域サイズは限定された GPU デバイスメモリ容量を超えられなかった。これに対し、(1) 局所性向上のために時間ブロッキング手法によりコード書き換え、(2) メモリ階層ランタイムHHRTとリンク、を施すことにより、これまでの 4 倍の問題サイズの際に約 80%の性能が維持されることを示した。本成果により情報処理学会コンピュータサイエンス領域奨励賞を受賞した。

2.

概要:

メモリ階層活用ランタイム技術のチーム内統合により、GPU デバイスメモリ・ホストメモリ・Flash SSD からなるメモリ階層(ノードあたり 1TB 級)を、MPI/CUDA で記述されたユーザプログラムから透過的に利用可能とした。これにより実現可能な問題サイズは最下層のメモリ容量にのみ制限されることとなり、ステンシル計算においてその技術を実証した。本成果の論文は IEEE Cluster 2016 会議に採択され、Best paper nominee (4 件)に選出された。残りの研究期間において、TSUBAME3.0 スーパーコンピュータ上で超大規模実証実験を予定している。

＜科学技術イノベーションに大きく寄与する成果＞

1.

概要:

当チームの研究により得られた知見として、現状の計算機アーキテクチャパラメータにおいては、ノードあたり 1GB/s 以上のアクセス速度を持つストレージデバイスであれば、科学技術計算(特にステンシルベースのもの)のためのホストメモリ容量の拡張として、現実的な性能オーバヘッド内で利用可能であると判明した。それを活用し、遠藤が主要な仕様策定メンバーとして参加し、平成 29 年 8 月に稼働開始した東京工業大学 TSUBAME3.0 スーパーコンピュータにおいては、ノードあたり Read 2.7GB/s, Write 1.8GB/s, 容量 2TB の高速大規模 SSD がローカルストレージとして搭載された。これを活用し、システム全体の理論最大では問題規模

1PB 級のシミュレーション(理研・京コンピュータ全体の DRAM 容量と同等)が実現可能なシステムが、学術・産業界に広く提供されることとなった。

2.
概要:
本チームで開発するメモリプロファイラツールExanaは、目的に応じて解析詳細レベルと解析時間のトレードオフを容易に調整可能な特徴を持つ。Exana はすでに多数研究グループによってアプリケーションソフトウェアのチューニングに用いられている。理化学研究所計算科学研究機構と連携し、次々世代スパコンのアーキテクチャ検討に重要な重点アプリやミニアプリの実際のチューニングに用いられた。またポストペタ藤澤克樹チームと協働で、ビッグデータベンチマークとして重要性の増す Graph500 ランキングの BFS 処理のチューニングを行った。ツールチェーンのほとんどが、github 上でオープンソースとして公開されている。

3.
概要:
TB 級の容量と GB/s 級の接続速度を持つ Flash デバイスを、DRAM と併用して用いることにより、ステンシル計算の規模を大幅に拡張することに成功した。デバイスの特性を考慮したアクセス手法と、時間ブロッキング手法による局所性向上により実用的な速度を得られた。この成果について米国 Flash Memory Summit などで発表を行ったところ、不揮発メモリの応用を拡張する試みとして、国内外の企業や研究者から興味を持たれ、技術詳細についての情報交換を行い、当チームの主催する MemoryPlus ワークショップで招待講演を依頼するきっかけともなった。

§2 研究実施体制

(1)研究チームの体制について

① 遠藤グループ (旧佐藤グループを平成 27 年度より遠藤グループに併合)
研究参加者

氏名	所属	役職	参加時期
遠藤　敏夫	東京工業大学	准教授	H24.10～
鯉渕　道紘	国立情報学研究所	准教授	H24.10～
額田　彰	東京工業大学	特任准教授	H28.4～
佐藤　仁	産業技術総合研究所	研究員	H24.10～
佐藤　真平	東京工業大学	助教	H27.4～ (H26.10～H27.3 は佐藤グループ)
佐藤　幸紀	東京工業大学	特任講師	H27.4～ (H24.10～H27.3 は佐藤グループ)
田邊　昇	東京工業大学	研究員	H28.10～
前澤　朋子	東京工業大学	研究員	H26.4～
幸　朋矢	東京工業大学	技術支援員	H27.5～
松宮　遼	東京工業大学	研究補助員(学生D2)	H28.4～
伊藤　祐貴	東京工業大学	研究補助員(学生M1)	H29.4～
金　光浩	東京工業大学	研究員	H24.10～H27.11

Irina Demeshko	東京工業大学	研究補助員(学生 D4)	H24.10～H25.3
岩渕　圭太	東京工業大学	研究補助員(学生 D3)	H24.10～H29.3
星野　哲也	東京工業大学	研究補助員(学生 D2)	H24.10～H27.12
河村　知輝	東京工業大学	研究補助員(学生 M2)	H24.10～H26.3
高崎　祐樹	東京工業大学	研究補助員(学生 M2)	H25.4～H27.3
佐々木 尚人	東京工業大学	研究補助員(学生 M2)	H26.4～H28.3
辻田　裕紀	東京工業大学	研究補助員(学生 M2)	H26.4～H28.3
都筑　一希	東京工業大学	研究補助員(学生 M2)	H26.4～H28.3
黒田　勝汰	東京工業大学	研究補助員(学生 M2)	H28.4～H29.3

研究項目

- メモリ階層対応ランタイムの研究開発とプログラミングモデル・アーキテクチャ統合
- メモリ階層対応ダイナミックコンパイレーション技術の研究開発(H27 年度より)

②(旧)佐藤グループ(グループ活動は平成 26 年度末迄)
研究参加者

氏名	所属	役職	参加時期
佐藤　幸紀	北陸先端科学技術大学院大学	助教	H24.10～H27.3
田中　清史	北陸先端科学技術大学院大学	准教授	H24.10～H27.3
請園　智玲	北陸先端科学技術大学院大学	助教	H24.10～H27.3
松原　裕貴	北陸先端科学技術大学院大学	産学官連携研究員	H25.4～H27.2
久保　幸	北陸先端科学技術大学院大学	研究補助員	H24.12～H27.3
西條　晶彦	北陸先端科学技術大学院大学	RA (学生・D5)	H24.12～H27.3
Faisal, Faiz Al	北陸先端科学技術大学院大学	LA (学生・M2)	H26.1～H27.3
佐藤　真平	北陸先端科学技術大学院大学	産学官連携研究員	H26.10～H27.3

研究項目

- メモリ階層対応ダイナミックコンパイレーション技術の研究開発

③緑川グループ
研究参加者

氏名	所属	役職	参加時期
緑川　博子	成蹊大学	助教	H24.10～
甲斐　宗徳	成蹊大学	教授	H24.11～
丹 英之	アルファシステムズ（株）	派遣社員	H25.4～ (H26.9 雇用形態変更)
北川　健司	アルファシステムズ（株）	派遣社員	H28.5～
大浦　陽	成蹊大学　理工学研究科	研究補助員（学生・M2）	H28.4～
白澤　卓磨	成蹊大学　理工学研究科	研究補助員（学生・M2）	H28.4～
柴山　悠	成蹊大学　理工学研究科	研究補助員（学生・M1）	H29.4～
外口　美幸	成蹊大学	時間給事務、技術補助	H25.4～H28.3
岩井田　匡俊	成蹊大学　理工学研究科	研究補助員（学生・M2）	H25.4～H27.3
直木　三華	成蹊大学　理工学研究科	研究補助員（学生・M2）	H24.11～H26.3
鈴木　裕一郎	成蹊大学　理工学研究科	研究補助員（学生・M2）	H24.11～H25.3
古尾谷　歩	成蹊大学　理工学研究科	研究補助員（学生・M2）	H24.11～H25.3

研究項目

・　大容量、高性能を実現する多種多階層型メモリ構成技術と管理手法の研究

(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

当チームの構成メンバーの多くはシステムソフトウェア・ツール層の研究者であり、アプリケーション・プログラミングモデル・アーキテクチャ層を中心として、研究チーム外と様々な形態で連携を行っている。

遠藤グループ(遠藤)

- CREST 丸山直也チーム(主に青木尊之グループ)から流体計算・結晶シミュレーションなどの大規模並列アプリケーションの提供を受け、本チームのメモリ階層対応・データ移動削減技術との統合を行うことにより、計算規模の大規模化や更なる高性能化を実際に果たした。この成果に関する論文により、情報処理学会 2015 年度 コンピュータサイエンス領域奨励賞を受賞した。
- CREST 丸山直也チームと、彼らの持つプログラミングフレームワーク技術と、本チームのメモリ階層活用技術・局所性向上アルゴリズムの統合について協働で進めている。計算機の特性をユーザプログラムから意識することなく、メモリ階層の活用により高性能・大規模なシミュレーションを可能とした。
- CREST 藤澤克樹チームと協働で数理最適化ソルバーの高性能化・問題大規模化に取り組んでいる。そのために本グループのメモリ階層間データ移動削減技術やスケー

ラブルな通信技術を活用している。この成果に関連して、国際会議 IEEE SC や IEEE IPDPS などにおいて複数の共著論文がある。

- CREST 藤澤克樹チームおよびビッグデータ領域 CREST 松岡聡チームと、深いメモリ階層を用いた計算機アーキテクチャなどについて密な情報交換を行っている。
- 遠藤は東京工業大学のスパコン TSUBAME シリーズの運用・開発チームの中核メンバーの一人として、新スパコン TSUBAME3.0 のメモリ階層への本チームの研究成果のフィードバックを行った。それにより該当システムの各ノードには GB/s 級のアクセス速度の高速 SSD が搭載された。
- 産業技術総合研究所と東京工業大学の連携により平成 29 年 2 月に開設された「実社会ビッグデータ活用オープンイノベーションラボラトリ」(RWBC-OIL)の中心メンバーの一人として参画し、メモリ階層活用を含む高性能計算機アーキテクチャのビッグデータ応用技術についての共同研究を行っている。
- 不揮発メモリの応用技術に関する国際会議 IEEE Non-Volatile Memory Systems and Applications Symposium (NVMSA2017)において、遠藤は Program co-chair を務め、国内の研究者 10 名近くを PC メンバーとして推薦を行うなど、不揮発メモリ研究分野そのものへの貢献および我が国のプレゼンスを高めた。次年度の NVMSA2018 ははじめて国内で開催されることとなり、遠藤は General co-chair に内定している。

遠藤グループ(佐藤)

- メモリ依存プロファイラおよび並列性見積もりツールである DiscoPoP を開発しているドイツ Ali Jannesari 博士 (Technical University of Darmstadt) と密に情報交換を行っている。DiscoPoP は LLVM をベースとする静的なプロファイリングコード挿入を基盤としており、Exana と補完する役割や機能を持つため、Exana の開発や機能強化にフィードバックを行っている。
- メモリアクセスパターン解析機能については、類似の機能を持つ Sigma ツールを開発していた米国 Cray 社の Luiz DeRose 博士とメモリアクセスパターン解析に求められる要件等の情報交換を実施している。
- 理化学研究所計算科学研究機構ソフトウェア技術チーム (チームヘッド南一生) と連携し、Exana にて理研の重点アプリや各種ミニアプリ (Fiber, Mantevo) のプロファイルを取得し、チューニングに実応用することに取り組んだ。同時に、コード最適化計画を効果的に作成するために、実アプリケーションの手動のチューニングが性能に寄与する事例の蓄積を行った。
- 九州大学の CREST 藤澤克樹チームとは、グラフ解析処理のベンチマーク Graph500 を題材にメモリアロケーションやメモリ局所性向上を狙うチューニング手法の検討及び実装を共同で行った。
- 東北大学とは文部科学省 HPCI FS　将来の HPCI システムのあり方の調査研究 「高メモリバンド幅アプリケーションに適した HPCI システムのあり方の調査研究」(東北大学チーム) の活動においては CREST プロジェクトで開発した Exana を利用して実アプリケーションの要求する命令レベルの BF 値(Byte per FLOP ratio)が有用な指標となりうるとの知見を得て、実アプリケーションを評価する際に利用した。HPCI FS に関しては、理化学研究所・東京工業大学チームと連携しアプリケーション解析を目的とした Exana の利用やその実用性について議論した。
- 東京工業大学の宮崎純教授とデータベースの高速化のために Exana を利用することの現実性などの情報を交換し、データベースシステムの高速化で頻繁に論じられるデータ構造、データレイアウトの概念は HPC 分野の SoA、AoS(Structure of Arrays, Array of Structures)と共通であり、適切なデータ構造にチューニングする支援を行うツールは有用との知見を得ている。
- X86 バイナリ変換ツール mcSema や、LLVM 中間言語を対象とした polyhedral コンパイラ Polly といった広く利用されているオープンソースソフトウェアに関して、研

究の過程で発見・作成したバグ報告やパッチ提供を行うことにより、コミュニティへの貢献を行っている。

緑川グループ

- CREST 建部グループの大山チームとは、情報交換、議論のためのミーティング、研究発表会を継続して行っている。大山チームの関わるネットワーク広域ファイル、OS、システムソフトウェアの知識・経験と、緑川グループが行っている高速 FlashSSD 利用による主メモリ拡張研究、非同期入出力による高速アクセス手法と性能、分散共有メモリシステムにおけるネットワークページ転送における効率化などは、お互いに重なる研究トピックがあり、違う視点からの質疑、議論、提案により、新たな研究ヒントとなることも多い。また、緑川が国際ワークショップなどで得た不揮発性メモリの最新動向なども情報提供している。昨年の大山らの論文には緑川への謝辞を頂いている。
- 開発中の不揮発性メモリの種類、性能は多岐にわたり、企業間の思惑や開発競争が絡むため、Flash Summit や Non-Volatile Memories Workshop(NVMW)などの国際的な関連ワークショップなどに参加、発表し、直接、担当者と議論、質疑することで得られる情報が多い。緑川は関連ワークショップに積極的に参加し、日本では得られない最新かつ詳細な実情報を得るだけでなく、NVM、ストレージなどの多数の研究者、メーカー開発者ら（HGST 社、株式会社　東芝、ソニー株式会社、中央大学竹内健博士など）と知り合いとなり、帰国後、メモリプラスワークショップを企画する上で非常に役立った。2014 年に主催したメモリプラスワークショップでは、メモリを核として、メモリデバイス、ストレージ、OS カーネル、システムソフトウェア、関連アプリケーションなどの研究者（菅野伸一（東芝）、追川修一（筑波大学）、吉田雅徳（株式会社　日立製作所）、レモアルダミアン（HGST）など他 3 件）に招待講演を依頼し、企業、大学を含む多様な参加者の事前アンケート（所属、業種、興味分野）などにより、活発な議論を行った。今後のメモリを核とする大きなうねりの中で、各分野のネットワーク形成にも役立つと考えている。
- 2010 年から NVMW を主催する UCSD の Non-Volatile Systems Laboratory (NVSL), Center for Magnetic Recording Research(CMRR)には東芝などの開発者も在籍し、企業主導の Flash Summit とは異なり、アカデミック分野からのユニークなレベルの高い発表も多い。NVM2014 で、高速ストレージを念頭としたカーネルドライバ性能に関するポスター発表していた Le Moal Damien 博士(HGST)とは、帰国後、数回のミーティングを持ち、OS カーネルの問題点、高速ストレージドライバ、カーネルモジュール開発、HGST の開発する最新 MRAM に関する情報などを得た。一方、緑川の Flash 利用による主メモリ拡張という研究トピックにも非常に興味を持って頂き、緑川は HGST ジャパン藤沢研究所で技術講演を行った。
- NVM2015 においても、上記 Le Moal Damien 博士と継続して議論をおこなった。さらに Lawrence Livermore 国立研究所の Brian C. Van Essen 博士が、緑川の Flash 利用による主メモリ拡張の発表に興味を持って頂き、博士の開発した独自の mmap カーネルモジュール di-mmap にも我々の研究と関わりが深かったため、会議中に多くの議論をおこなった。帰国後、di-mmap のソースプログラムを得て、我々の性能評価にも用いており、今後も議論を持ちたいと考えている。
- 2014 年に続き、2016 年に第二回メモリプラスワークショップを開催し、成瀬彰(NVIDIA)、池井満　(Intel)、大島成夫（東芝）の各氏による招待講演などを行った。これらの講演においては、開催時期から直近で利用可能となる、積層メモリを含んだアクセラレータ・メニーコアプロセッサ(成瀬氏・池井氏)や、Flash Summit 2016 でキーノート講演で発表されたばかりの三次元 Flash(大島氏)など、タイムリーな内容であり、聴衆から強い関心を得、活発な議論を行った。

§３ 研究実施内容及び成果

３．１ メモリ階層対応ランタイムの研究開発とプログラミングモデル・アーキテクチャ統合（東京工業大学 遠藤グループ）

(1)研究実施内容及び成果

本項目では、次世代メモリ技術を含めた多階層のメモリを持つ計算機システムを対象として、科学技術計算・シミュレーションなどの計算大規模化・高性能化のために、メモリ階層の効率的利用を可能とするランタイムライブラリおよび、アルゴリズムの局所性向上技術の研究開発を行う。これらの技術統合・検証を、深いメモリ階層を持つ大規模スーパーコンピュータ上において行う。

上記のねらいで述べた計算大規模化と高性能化の両立のためには、高位メモリのバンド幅の高さと、低位メモリの容量の大きさを効率的にアプリケーションから活用可能である必要がある。このために、主に下記の研究を実施する。

[a] 階層対応ランタイムの設計・実装・評価
[b] アルゴリズムの局所性向上技術の研究
[c] 大規模スーパーコンピュータにおける技術統合
[d] 将来のメモリデバイスの大規模演算への影響の解析

これらの項目や他グループの研究項目は完全に独立しているものではなく、密接に関連している。主要プラットフォームとして、GPU アクセラレータを多数搭載した東京工業大学 TSUBAME2.5 スーパーコンピュータおよびチームで導入した計算機を用いており、残り研究期間においては平成29年８月に稼働開始したTSUBAME3.0スーパーコンピュータを大規模実証実験のために用いる予定である。

[a] 階層対応ランタイムの設計・実装・評価

アプリケーションが、高位メモリのバンド幅の高さと低位メモリの容量の大きさを効率的に、かつプログラミングコストの低い形で活用可能とするためには、アプリケーション開発者がデータのメモリ階層間移動を（すでに記述済のものを除いて）追加記述しなくてよい、かつ低位メモリの低速さの影響を大幅に抑制する、ことが必要となる。この実現に向けて下記を実現するようなメモリ階層対応ランタイムライブラリの概念設計・検討を行った：データのメモリ階層間移動(スワップに相当)をランタイムライブラリに行わせることによりアプリケーション開発者から隠ぺいする。一方、低位メモリの低速さの影響抑制のためにはアプリケーションの局所性向上が必要であり、本項目ではコード書き換えを前提としている（詳細は次項目[b]で記述)。その際のプログラミングコストを軽減可能であるようなランタイムライブラリである必要がある。上記に基づく階層対応ランタイムとして「MPI 版」と「高レベル版」の研究開発を行ってきた。

[a-1]階層対応ランタイム MPI 版

対象環境・アプリケーションとしては、多数の NVIDIA GPU・多数ノードからなる計算機環境上で、MPI および CUDA で記述されたアプリケーションに注目する。これらが容易に利用可能なメモリ量は、通常 GPU デバイスメモリ容量内に限定されるところを、シームレスに拡張するランタイムライブラリである HHRT（Hybrid Hierarchical RunTime)を設計・開発した。現在の CUDA ではページ単位のスワップ機能やメモリ保護機能は（少なくともユーザレベルからフック可能な形では）提供されておらず、この制限のもとスワップを可能とするために、HHRT は以下のような設計となっている。

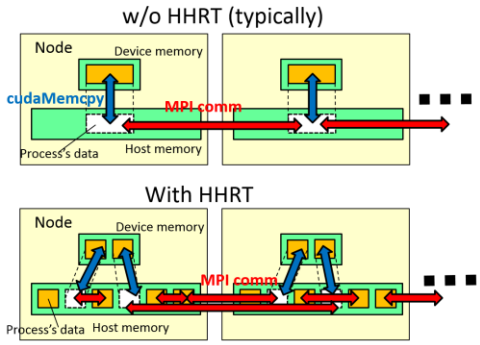


図 A-1: 典型的なマルチ GPU アプリケーションの実行モデル(上)と、HHRT を用いた実行モデル(下)

HHRT をリンクするアプリケーションは、複数(例えば m 個)の MPI プロセスが単一の GPU を共有する。そして m 個のプロセスの合計で、使用メモリ量はデバイスメモリ容量を超えうるとする。同時に m 個のプロセスはデバイスメモリを利用することはできないため、「プロセス単位のスワップ」機能を HHRT が提供する。

この HHRT の利用により利用可能メモリサイズは拡大するが、その利用のみでは多くの場合データのメモリ階層間移動コストにより性能が低下する(GPU 搭載計算機においては 8GB/s 程度の PCI-Express 通信コストを被る)。この点については次項目[b]で述べる。

中間報告以降には、適用可能な問題規模のさらなる拡大のために、データのスワップアウト先として、ホストメモリに加え高速 Flash SSD を利用する拡張を、緑川グループの技術を活用して行った。GB/s 級の高速 m.2 SSD を用いた実証実験を通して、GPU デバイスメモリのみならずホストメモリ容量を超えるステンシル計算が実現可能であることを示した。この成果により、IEEE Cluster 2016 会議において、Best paper nominee(4 本)に選出された。

しかしながら、上位メモリ階層の 20 倍以上(100GB 以上)の問題規模において、性能維持が困難となる、または実行に失敗することが判明した(図 A-2 の TB+HHRT(2016)に相当)ため、その詳細な原因解析を行った。その結果デバイスメモリとホストメモリの双方において、メモリ容量を圧迫する要因があり、大規模メモリの効率的な利用を阻むことが分かった。HHRT モデルにおいては問題規模の拡大のために、多数プロセスを 1 ノードに共存させるという方法を取る。このときに、(1) 各プロセスは暗黙的にデバイスメモリの一部を占有する(K20X GPU の場合はプロセスあたり 70MB 程度)。数十プロセスが起動すると、デバイスメモリの大半を消費してしまう。(2) 各プロセスが MPI 通信を行うときに、現在の HHRT の実装では通信対象バッファをピンダウンし、HHRT スワップの対象外としている。これがホストメモリを圧迫する。それぞれについて、(1) プロセスが長期間スリープする際には、CUDA コンテクストを強制的に破棄して、デバイスメモリ消費を抑える、(2)MPI 通信対象バッファも SSD へスワップアウト可能とし、実際のデータ転送は SSD から部分的に読み書きすることとし、ホストメモリ消費を抑える、という対処を取った。その結果、問題サイズが 100GB を超える場合の性能は改善され、実現可能な計算規模は SSD の容量で規定されるという理想的な状況となった(図 A-2 TB+HHRT(2017)に相当)。ただし 100GB 超の場合には依然として小規模実行の速度の半分以下となり、このコスト解析を行っている。

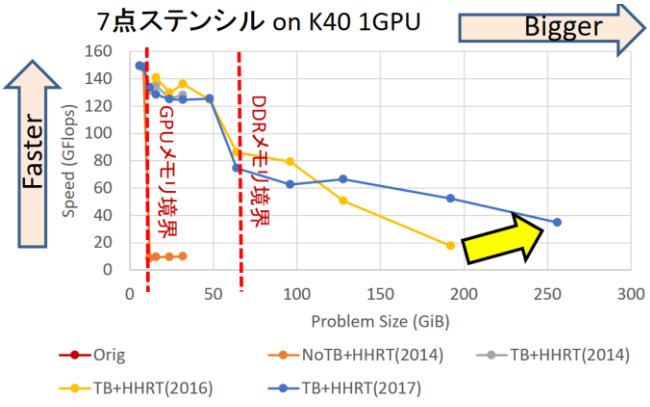


図 A-2: HHRT を用いたステンシル計算の問題規模と性能

この HHRT の機構を、プログラミングコストを少ない形で活用可能とするために、ポストペタ CREST 旧丸山直也チームとの連携のもと、ステンシル向け並列化フレームワーク[Shimokawabe et al. SC14]とHHRT、および局所性向上技術である時間ブロッキングの統合を行い、ユーザプログラムへの変更を必要とせずに高性能・大規模なステンシル計算の実現を行った。フレームワークと組み合わせた場合に現在利用可能なのは GPU デバイスメモリとホストメモリの2階層であり、大規模統合実験に向けて、Flash メモリを利用可能とする改良を準備中である。

[a-2]階層対応ランタイム高レベル版

上記の HHRT は、CUDA および MPI という、比較的低レベルのプログラミングモデルで記述されたアプリケーションを対象としている。また、HHRT におけるメモリ階層間スワップ単位はプロセスであるが、大規模数値解析などにおいては、より細粒度なスワップが望ましいことを実証してきた。

以上の議論から、より高レベルなプログラミングレイヤにおいて大規模シミュレーションを可能とするアプローチとして、PGAS ランタイムシステムとメモリ階層活用技術の統合を行っている。現在、UPC++などの多くの PGAS システムから利用されている通信レイヤである GASNet を拡張した vGASNet システムを設計・開発している。MPI（特に MPI-1)と異なり、PGAS では片方向通信モデルであるため、そこで現れる通信進行関数などにおいて、Flash デバイス I/O などを行うこととした。さらに、このような環境でボトルネックとなる Flash デバイスへのアクセスを軽減するために、ホストメモリの一部をキャッシュとして用いる機構、さらには多数ノードから同一メモリ領域にアクセスが発生した場合のボトルネックを軽減するためのスケーラブルな協調キャッシュ機構を実装した。このシステム上で、藤澤克樹チームの半正定値計画問題ソルバーSDPARA の、求解可能な問題規模をホストメモリ容量を超えて拡大することを計画している。

[b] アルゴリズムの局所性向上技術の研究

[b-1] ステンシル計算

局所性向上・通信削減によりメモリ階層を効率利用するアルゴリズムの研究として、ステンシル演算の時間ブロッキングに関する研究を、GPU スパコンを主プラットフォームとして取り組んだ。基本的には時間ブロッキングとは、ステンシル対象領域のうち一部領域（上位メモリに収まる範囲）の計算にとりかかったら、その一部領域について複数時間ステップの計算を一気に行う手法である。これは局所性向上手法としてかつてより知られた方法であるが、キャッシュ効率向上よりもデバイスメモリ・ホストメモリ間通信を削減することを主眼に取り組んでいる。

時間ブロッキングを冗長計算を導入せずに行う場合、部分領域計算の間の依存解決のために計算結果を再利用する必要がある。その再利用のためのバッファまでもが GPU メモリ容量を超えても良いよう、ホストメモリへ退避する手法を提案した。7 点ステンシルベンチマークに手法を導入し評価した結果、前年度よりさらに性能が向上し、5%以下のオーバーヘッドで GPU メモリ容量の 10 倍以上の規模の 7 点ステンシル計算を実現した。本研究項目について 2014 年 7 月に行われた GTC Japan 2014 においてポスター発表をおこない、NVIDIA Award を受賞している。

さらに、時間ブロッキング導入に伴うプログラミングコストを低減する手法の一つとして、polyhedral コンパイラの一つである Polly への拡張を行った。Polly にはすでにブロッキングによるループ変換機構が組み込まれているが、時間ブロッキングで発生する依存関係を考慮した斜め方向のブロッキング変換を行わないことが判明した。これに対し、時間ブロッキング用の二次元および三次元のループスケジューリング変換機能を追加した。変換可能なループは簡単なものに限られるものの、Xeon および Xeon Phi 上での実験により、手動で時間ブロッキングを導入した場合に近い性能が得られることが分かった。

[b-2] 密行列計算

ステンシル計算以外にも、密行列演算の局所性向上・通信削減技術の研究開発を、CREST 藤澤克樹チームと協働で推進している。これまで、半正定値計画問題ソルバー SDPARA の大規模問題対応・高性能化について、コレスキー分解部の改良を行い、TSUBAME2.5 の 4000GPU を用い、行列サイズ 230 万の大規模密行列計算において 1.7PFlops を実現した。さらなる改善のために、GPU-CPU 間の通信量大幅削減をねらいとしてタイル単位のデータドリブン実行方式を採用した。さらにタイル単位のジョブのスケジューリング方式、メモリ階層間のスワップ方式の比較検討を行った。これらにより上述の実装に比べ、最大 25%の性能向上を果たした。

以上の成果は GPU デバイスメモリとホストメモリの 2 階層を利用するものであるが、[a-2]に述べたメモリ階層活用ランタイム高レベル版の利用により、TSUBAME3.0 などの Flash SSD を用いて、さらなる問題規模拡大を予定している。

[b-3] 深層学習計算

畳み込みニューラルネットワーク(CNN) による深層学習は計算量が非常に大きく、GPU などによる高速化は必須となっている。しかし、GPU デバイスメモリ容量の問題により、ネットワークや画像サイズが大きい CNN の演算はハードルが高い。そのため、このような深層学習演算についてもメモリ階層活用を可能とするため、out of core cuDNN(ooc cuDNN) ライブラリの設計と実装を行った。ooc cuDNN は、デファクトスタンダードな深層学習ライブラリの cuDNN を拡張したものであり、デバイスメモリを超えるサイズのテンソルを入出力にとることができる。計算の分割を行う際にはその分割サイズの決定が性能に大きく影響するため、性能モデルおよび探索ヒューリスティクスに基づいたアルゴリズムを提案した。さらにメモリ階層間のデータ移動を削減するため、複合された計算を行う API を提案した。以上により、16GB のメモリを持つ NVIDIA P100 GPU 上で、60 GB 以上のメモリを必要とする CNN の計算が、小規模問題に比べ 13 %程度の速度低下で実行できることを示した(図 A-3)。

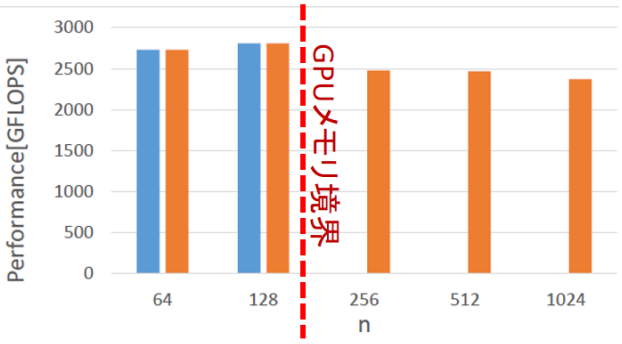


図 A-3: ooc_cuDNN を用いて VGG16 の計算を P100 GPU で行った性能。n は学習バッチサイズを示し、大きいほど大容量データを用いる

現在このライブラリについて CREST プロジェクト「社会インフラ映像処理のための高速・省資源深層学習アルゴリズム基盤」(代表:篠田浩一)などで活用すべく準備を進めている。

[c] 大規模スーパーコンピュータにおける技術統合

[a]のランタイム技術と[b]の局所性向上技術を、大規模実アプリに適用しスーパーコンピュータ上で統合を行った。この項目は本チーム発足以来のねらいであった、大規模・高性能シミュレーションの実現に直接寄与するものであり、まず中間報告時点での技術統合について述べる。

CREST 丸山直也チーム青木グループにて開発された都市気流シミュレーションを取り上げた。これは都市部におけるビル風を含む気流の計算を Lattice-Boltzmann 法に基づき行うアプリケーションであり、CUDA および MPI で約 15,000 行で記述されている。このアプリについて、HHRT 上での実行を仮定し、時間ブロッキングを組み込むリファクタリングを行った。ソースコードへの変更量を調査したところ、時間ブロッキングを組み込むために 148 行、さらにデータ移動を抑制するための最適化のために 1,021 行の変更量であった。図 A-4 に示すように GPU メモリ容量の 4 倍の問題サイズの際に約 80%の性能が維持されており、問題規模・性能・生産性の維持が実証された。さらに、TSUBAME2.5 スパコンの多数 GPU 利用時においても、64GPU で 57 倍という良好なウィークスケーラビリティを得ている。

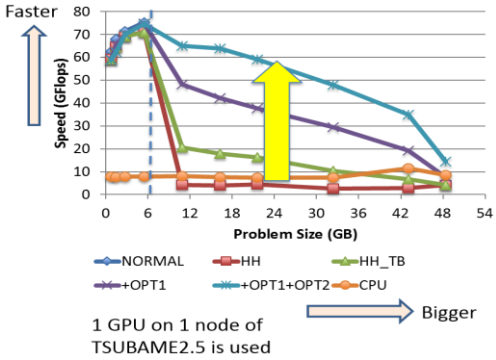


図 A-4: 都市気流アプリに時間ブロッキングを組み込み HHRT 上で実行した場合の性能

この時点の課題としては、Flash メモリ活用技術との未統合である点や、最適化のためのコード変更量が比較的多い点などが挙げられる。残り研究期間においては、これらの点についての改良を行い、TSUBAME3.0 上での大規模実証実験を実現する計画である。コード変更量の課題については、[a-1]で示したステンシルフレームワークと HHRT の統合により困難を避ける。当該フレームワークにおいてはすでに局所性向上技術である時間ブロッキングが組み込まれており、ユーザプログラムから透過的に効率的なメモリ階層活用が可能である。Flash への対応については、HHRT 自身の対応は[a-1]で述べたように終了しているものの、実装上の理由により、ステンシルフレームワークがホストメモリに確保した領域を Flash へスワップできない。この点の改良を急ぎ、TSUBAME3.0 上での大規模実証実験を通し、PB 級の規模と 1PB/s に近い高性能に迫る。

本成果により、本チームの目標である、エクサスケール時代における O(10PB)の大規模・O(10PB/s)の高性能を両立するシミュレーションの実現性について重要な知見を得られると考える。

[d] 将来のメモリデバイスの大規模演算への影響の解析

現在のアーキテクチャ構成要素を考慮したソフトウェア設計および技術開発に加え、近い将来利用可能と期待されるメモリデバイスが、大規模シミュレーション・ビッグデータ解析などの性能に与える影響の解析・モデル化について研究を推進した。

Intel 3D Xpoint などのアドレッサブルなメインメモリとして利用可能な不揮発メモリについて予告されている。そのような DRAM より大規模だが遅延も悪化すると考えられるメモリを想定し、種々の計算プログラムの速度性能がどのように影響を受けるか研究を行った。そのために、パフォーマンスカウンタをベースとした性能推定ツールを整備したが、その推定スループットはサイクルアキュレートシミュレータ MARSSx86 より約一万倍、軽量とされる推定ツール Quartz よりも十倍近く実行することができる。Graph 500 や疎行列計算を主な対象とし、大遅延時の性能特性を評価した。疎行列においては、多数の非零パターンにおいて、CSR, MSR, JAD, ELL を始めとする複数フォーマットを用いて実験を行った。非零パターンだけでなく、疎行列フォーマットおよび前処理法が大きく遅延による影響を左右することが判明している。

[開発ソフトウェア]

公開済:

HHRT (Hybrid Hierarchical RunTime)

- 機能:MPI および CUDA のラッパライブラリとして、その上で稼働するユーザアプリケーションの確保したデータを暗黙的に上位メモリ階層から下位メモリ階層にスワップする。こ

れにより、アプリケーションの利用可能なメモリ容量を仮想的に増大させる。
- 特徴：HHRT ヘッダファイルのインクルード、ライブラリのリンクにより、ユーザアプリから容易に利用可能である。またスワップタイミングは MPI 通信時に限定され、多くの場合に OS スワップを用いるよりも高性能である。
- 前提とするマシン： NVIDIA GPU を搭載した計算機クラスタにおいて最も効果を得られるが、GPU を持たないマシンであってもホストメモリと Flash のようなメモリ階層を持てば、問題規模拡大のために利用可能である。
- 潜在的ユーザ：主に GPU クラスタ上で動作するアプリを持っており、問題規模拡大を望むユーザ。

公開予定：

vGASNet
- 機能：多数計算ノードにまたがったホストメモリ・Flash 等からなるメモリ階層を Global Address Space として見せるライブラリである。PGAS システムの低位通信レイヤとして用いられる GASNet ライブラリと互換 API を持つ。
- 特徴：ホストメモリ容量を超えた、Flash SSD 群の合計容量を単一アドレス空間として利用可能とする。SSD へのアクセスオーバヘッドを低減するため、ホストメモリの一部をキャッシュとして用いる機能、スケーラビリティ向上のための協調キャッシュ機能を備える。
- 前提とするマシン：各ノードに GB/s 級の高速 SSD を備える計算機クラスタを想定する。
- 潜在的ユーザ：PGAS モデルにより記述されたアプリを持っており、問題規模拡大を望むユーザ。

ooc_cuDNN
- 機能：深層学習におけるテンソルへのカーネル処理(畳込み、アクティベート関数適用)を行うライブラリであり、GPU デバイスメモリ容量を超える入出力データに対応する。NVIDIA 社 cuDNN ライブラリと基本的に互換 API を持つ。
- 特徴：各関数内部で計算を分割し、それぞれの部分計算はオリジナルの cuDNN に任せることにより高速演算を実現する。分割サイズの決定については詳細な性能モデルに基づく。データのメモリ階層間移動を抑制するための複合された計算の API を持つ。
- 前提とするマシン：NVIDIA GPU を搭載した計算機。
- 潜在的ユーザ：TensorFlow や Chainer などの、cuDNN により加速された深層学習フレームワークに組み込まれることを想定している。潜在的には、それにより問題規模拡大が可能となったフレームワークのユーザが想定される。

3．2 メモリ階層対応ダイナミックコンパイレーション技術の研究開発
((平成 26 年度末迄)JAIST 佐藤グループ、(平成 27 年度以降)遠藤グループ)

(1) 研究実施内容及び成果

本項目のねらいは下記の通りである。ヘテロジーニアスで多階層なメモリアーキテクチャを持つシステムにおいて高い性能効率を得るためには、既存のキャッシュメモリやコンパイラでは考慮されていないメモリ階層間の特性の違いを意識した局所性の意識する必要がある。そこで、メモリ階層や異種メモリのパラメータの相違をアプリケーションのデータ参照局所性に最大限かつ生産的にマッピングすることを目的としたメモリ階層対応ダイナミックコンパイレーション技術を研究開発し、メモリ階層チューニングを行うダイナミックコンパイラツールチェインとして確立する。

上記のねらいに向けて、メモリ階層や異種メモリのパラメータの相違をアプリケーションのデータ参照局所性に最大限マッピングするメモリ階層対応コンパイレーション技術を研究開発し、メモリ階層チューニングを自動／半自動で行うコンパイラツールチェインとして確立することを実施する。具体的には、アプリケーションプログラムに由来するデータ局所性を

システムアーキテクチャやプロセッサアーキテクチャに由来するメモリ階層に写像する変換を支援する上で非常に重要となるメモリ局所性プロファイラ、更に、想定するメモリ階層における性能予測を行いコード最適化のプランを作成するメモリモデルを用いたコード最適化計画、データのプレースメント・レイアウトの最適化を自動で行うコード最適化機構をそれぞれ開発し、メモリ階層に関係する最適化をユーザから透過的に行う手法を提供する。

メモリ局所性プロファイラの要素機能の研究開発

メモリ局所性プロファイラに関しては、実行バイナリコードからメモリ参照に関する情報を静的に抽出し、アプリケーションの実行時に出現するメモリを介した動的なデータ依存関係をデータメモリ局所性情報として抽出するプロファイラの実装を行い、逐次実行ベンチマークプログラムを用いて評価を行った。更に、メモリ局所性プロファイラを他のグループやチームが利用する利便性を強化するために関数およびループ階層の各ノードを単位とした区間ごとの実行時間の透過的計測やそれらの間のデータ依存関係を可視化するツールを作成し、インタラクティブに興味領域を設定することを可能とするインタフェースとして整備した。メモリを介するデータ依存関係に加えて、アプリケーションのメモリアクセスパターンをモデル化しメモリ局所性情報としてメモリモデルを用いたコード最適化計画に入力することを目的とした階層的メモリアクセスパターン記述形式を策定し、プログラムのメモリアクセストレースより階層的にメモリアクセスパターンを抽出するツールの実装を行った。更に、アプリケーションの要求する命令レベルの BF 値(Byte per FLOP ratio)を実行時に測定する機構を実装し、HPCI FS 東北大学チームのアプリケーションにて BF 値を評価する際に利用した。

メモリ階層性能シミュレータのプロトタイプの実装

メモリ階層性能シミュレータに関しては、実行バイナリコードを入力として実行時に L1 キャッシュから L3 キャッシュの挙動をシミュレーション可能な実行駆動型キャッシュシュミレータの実装を行った。本シミュレータは任意の Way 数やキャッシュライン幅などのパラメータに対応するメモリ性能の見積りを行う機能を有し、現状で x86 ベースの CPU や京に実装されている SPARC 64 のメモリ階層についてのメモリ性能を見積ることができる。また、既存のアクセラレータにおけるメモリ階層の詳細構造や特性に関する調査を実施し、調査の一環としてアクセラレータにおけるメモリ階層構造をアプリケーション特性に応じて再構成可能デバイスを用いてカスタマイズするという自由度が与えられた場合の性能を評価した。

メモリモデルを用いたコード最適化計画の設計及び実装

メモリモデルを用いたコード最適化計画作成については、始めにメモリ局所性プロファイラとメモリ性能シミュレータの機能強化および連携強化を実施した。その成果としてアプリケーション実行時に抽出したメモリアクセスパターンやメモリ階層性能シミュレータの結果を入力として、統合的にメモリ性能を見積る仕組みが完成した。次に、メモリ階層を考慮した性能モデルを策定することに取り組んだ。本モデルは、ルーフラインモデルをベースとし、平成 25 年度までに実装した BF 値解析機構やメモリ階層性能シミュレータと連携し、各種パラメータを変化させた際のアプリケーション実行性能を予測する。本モデルを利用して、データレイアウトやアクセスパターンに起因するキャッシュ競合の原因となる箇所の特定が可能なことを確認した。

加えて、これらの機能に関しても Exana ツールとしてとりまとめ、そのプロトタイプをチーム内やチーム外部の深く性能チューニングに携わっている研究チームに提供し、協力して検証およびチューニングへの実応用への課題を探った：①特に、理化学研究所の計算科学研究機構ソフトウェア技術チーム（チームヘッド南一生）と連携し、Exana にて理研の重点アプリや各種ミニアプリ（Fiber, Mantevo）のプロファイルを取得し、チューニングに実応用することに取り組んだ。同時に、コード最適化計画を効果的に作成するために、実アプリケーションの手動のチューニングが性能に寄与する事例の蓄積を行った。

②チーム内連携によりステンシルコードの代表である姫野ベンチマークに高度なループ変

換の例としてテンポラルブロッキングの適用を行ったり、CREST 藤澤克樹チームと連携してグラフ解析処理のベンチマークGraph500を題材にメモリアロケーションやメモリ局所性向上を狙うチューニング手法の適応を行った。

単一命令セット環境におけるコード変換機構

コード変換機構は以下のような4つのステージP-E-T-Sから構成される設計とした。Pは Profile を示し、アプリケーションの性能情報や依存関係をランタイムに抽出するステージである。E は Estimate を示し、Profile 結果を利用し性能を推定しどのような変換を行うかの戦略を立てる。TはTranslateを示し、戦略に基づきバイナリコードを変換するステージである。SはSwitchを示し、コード実行をオリジナルのものからTステージで変換したものに切り替えるステージである。PおよびEステージはExanaツールとして実装しているメモリ階層性能モデルそのものであり、メモリモデルの実装と連動して進めている。計画当初においては本項目を動的バイナリ変換技術に基づき実現する予定であったが、ループ変換を実施した際のソフトウェア的な等価性を担保する手法の開発など解決するべき課題が多いことが明らかとなったことなどから、LLVM中間言語向けpolyhedralコンパイラであるPollyを活用する方針に転換し、次項目で述べるように自動/半自動チューニングにより焦点を当てて研究を推進した。

フィードバック駆動型チューニング機構

平成26年度からは実アプリケーションのプロファイリング結果をソースコードのコード最適化にフィードバックするフィードバック駆動型ソースコード変換に基づく最適化の方式を検討し、その自動化のための概念設計を実施している。具体的には、コード最適化計画を立てる上で鍵となるメモリ階層を考慮した性能モデルやアプリケーションのメモリ局所性特性をExana にて実際に取得できることの検証を行っている。更に、テンポラルブロッキングされたステンシルコードを題材にループブロッキングを実施する際のブロックサイズや、各種ループ変換の実施の有無など最適化のパラメータを探索し、それらを反映する性能モデルの開発に向けた研究を行った。これらの解析結果において検出された問題箇所はコンパイラの生成するデバッグ情報に基づきソースコードの位置に対応付けされ、アプリケーション開発者にフィードバックする拡張を行った。より具体的には、Exana 内部に競合キャッシュミス検出機構 C2Sim を組み込み、競合ミスの原因となる配列の名前等をユーザに提示する。この過程では同一配列内での競合ミスか、異なる配列どうしのアクセスに起因する競合ミスかの判定も可能である。

この拡張により、性能チューニングのためにソースコードを修正するべき箇所の同定と見込まれるキャッシュヒット向上率の導出が可能となり、チューニングの生産性を向上させることが可能となることを確認した。例えば、上述した競合キャッシュミスを引き起こす配列を知ったユーザは、その原因の配列に注目し、配列内パディングもしくは配列間パディングのふさわしいコード変更を行うことができる。このようなフィードバック型チューニングにより、ステンシル計算等で実際にキャッシュヒット率を改善できることを確認した。

更に、これまで取り組んできたソースコードを起点とするコード最適化に加えて、LLVMやPollyにおける最適化機構にIRレベルでフィードバックすることにより最適化された実行バイナリコードを再度生成するという最適化プロセスの実装についても取り組んだ。具体的には、LLVM-IRレベルでタイリングサイズを調整する機能を拡張し、タイリングパラメータの最良値を探索する機構であるPATT (Polyhedral compiler based Auto Tile size opTimizer)を開発し評価を進めた。

関連するレイヤとのコデザインによるシステムの性能チューニングとして、スパコンクラスのマシンでの動作や実稼働しているアプリケーションでの利用にも対応できる実用性を備えるツールとなることを目標に Exana ツールの改善を進めた。具体的には、普及が進みつつあるメニーコア型CPUにおける有効性を評価するため、XeonPhi (Knight Lnading)の環境へ Exana ツールを移植し動作の確認を行うと同時にマルチスレッドプログラムに向けた

Exana ツールの改良を行った。XeonPhi 上で動作するコードのキャッシュ競合プロファイリングを Exana を用いて行った結果、一般的な CPU と比べてメモリ階層構造やキャッシュの構造が相違していることに由来する異なる特性が観測されることも確認した。

[開発ソフトウェア]
公開済:
Exana
- 機能:ユーザプログラムの挙動解析を透過的に行う解析ツールであり、実行バイナリを入力とする。内部にキャッシュシミュレータを備えており、メモリアクセス・キャッシュミスの挙動を詳細に解析可能である。またプログラムの実行時に出現するループおよび関数呼び出しの階層構造抽出(LCCT)も可能であり、実行のボトルネックやデータ依存関係を俯瞰的に把握することができる。さらに LCCT の解析結果を直感的に利用可能な、可視化ツールの公開を行っている。
- 特徴:キャッシュシミュレータには競合ミス解析器 C2Sim を含んでおり、競合ミスの発生およびその原因となるデータ領域やアクセス命令をソースコード上の情報として提示することができる。解析内容と解析時間のトレードオフのために、各機能のオン/オフや、プロファイリングのサンプリングも可能である。各種言語(C/C++, Fortran)、コンパイラ環境、MPI 環境、マルチスレッド環境に対応し、共有ライブラリ、動的・静的リンク、バッチジョブ、子プロセスのフォーク、再帰のあるプログラムも対応可能である。
- 前提とするマシン:バイナリ変換機構の一部が Intel Pintool に依存しており、現時点では Intel プロセッサ搭載マシンである必要がある。Cray XC30, SGI Altix UV, TSUBAME 2.5 等のスパコン環境や汎用 x86Linux クラスタで動作実績がある。
- 潜在的ユーザ:アプリケーションのメモリレイアウト等の最適化が可能か知りたいソフトウェア開発者。

3.3 大容量、高性能を実現する多種多階層型メモリ構成技術と管理手法の研究 (成蹊大学緑川グループ)

(1) 研究実施内容及び成果

本研究では、多様化かつ多階層化する記憶構造を持つノードから構成される次世代高性能コンピュータシステムにおいて、ノード内外の記憶を含む統一的な多階層メモリモデルを構築し、メモリアクセス局所性を生かしたデータの配置、移動を行い、並列処理性能のスケーラビリティ向上を目指す。また、ローカルメモリを超える大容量データ処理性能を最大限に引き出すことを可能にすることを目的とする。

対象とするメモリ階層としては下記を想定する。
- 垂直方向記憶 メモリ階層デバイス間(キャッシュ、DRAM, NVM, SSD, HDD)の連携と効率的利用
- 水平方向記憶 ネットワークによる多ノード連携、局所・遠隔メモリ、分散共有メモリ利用

垂直方向記憶研究では、Flash SSD を用い、主記憶の拡張メモリとして利用するため、時間的、空間的局所性を利用したアルゴリズムを導入し、非同期多重 IO による方式と自動最適化機構により、高性能計算で広く用いられているステンシル計算には実用上十分利用できるような環境を実現した。

水平方向メモリ拡張研究では、マルチノード・マルチスレッド並列処理向けシステムとして、マルチスレッド対応の分散共有メモリシステムのプロトタイプを設計実装し、初期稼働実験を終えている。一方、単一ノードからの主メモリ拡張のための遠隔メモリ利用(リモートページング)については、シングルスレッド応用プログラムにおいて非常に効果の高かったアクセス局所性に応じた実行時ページサイズ可変機構を、マルチスレッド応用プログラム向けに、設計、実装した。

単一ノード内多階層メモリ間のスワップ方式の試作と事前評価実験

平成 24-25 年度において、最も広く利用可能である NVM として、PCIe バス接続型 Flash SSD を用い、主記憶(DRAM)の拡張としてプログラムから用いることを目的に、(1)fast swap kernel 利用による malloc 方式、(2) file system 利用による mmap 方式 を用い、ステンシル計算を中心に、DRAM 主記憶サイズを超える問題サイズに対して適用する性能評価を行った。また DRAM-flash 間の 1000 倍程度もある大きな latency ギャップを埋めるために、空間的、時間的ブロック化によるデータアクセス局所性を高めるアルゴリズム(図 C-1)を DRAM, Flash 向けに新たに設計・導入して効果を得た。これにより、応用プログラムのデータアクセス局所性を高めることにより、Flash が主メモリ拡張として十分利用可能であることを明らかにした。最新 NVMe デバイス利用時には、DRAM のみ利用時に比べ 80%～90%の性能を実現している。

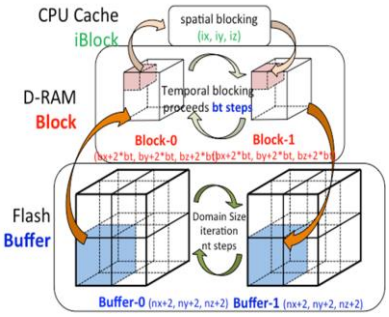


図 C-1 L3-Cache, DRAM, FlashSSD 多階層利用
時間・空間ブロッキングによるステンシル計算

依然、DRAM に比べ 1000 倍程度のアクセス遅延を持つフラッシュであるが、RAID0 構成によりスループット向上させ、スレッド並列による非同期 IO と、さらに後述の最適ブロックサイズを用いることで、mmap を用いる手法に比べ、高性能かつ非常に安定した動作が可能であることを示した。さらに、1 ノード用アルゴリズムを拡張し、各ノードに SSD を備えたコンピュータクラスタ向けに、ノードに分散するメモリの総容量を超える大規模データを扱うためのステンシルアルゴリズムを開発した。また、マルチノード向けの最適パラメタ(ノード間, ノード内多層レイヤにおけるブロックサイズ)の決定手法を提案、実クラスタで効果を得た。

上記のような局所性向上技術を用いた Flash 向けステンシル計算について、MultiMem-Stencil と呼ばれるパッケージとして公開予定である。

静的情報利用によるデータ配置と限定的メモリ同期方式の設計・実装

平成26年度においては、Linux Kernel実装による非同期入出力を用い、Flash SSDをブロックデバイスとして直接入出力することにより、ファイルシステムレイヤオーバヘッド、kernelによるページキャッシュ操作をスキップし、マルチスレッドによる多重IOによる(3)aio方式を新たに設計し、従来のmmap, mallocに比べ高い効果を得た。(図C-2)

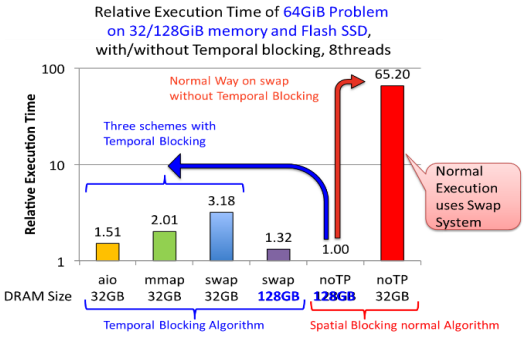


図 C-2 ステンシル計算 3 手法の相対性能

ただし、aio方式では、IOデータサイズ、オフセットなどがSSDブロックサイズにアラインされる必要がある。このため、aio方式に適した新たなデータレイアウト、マルチコア間の計算work-share方式などの最適化を行った。mmap方式にも最適化を行い、この結果、両方式とも、最適化前の約半分の時間で処理可能となった。(図C-3) さらに、**NUMAシステム向けの最適化**として、データアクセスと計算コアのaffinity制御など導入し効果を得た(図C-4)。

平成26年度開発のaioアルゴリズムでは、64GiBの主メモリ(DRAM)とFlashSSD利用により、主メモリサイズの16倍サイズの問題(ステンシル計算)を行った場合、十分なDRAMのみを用いて処理した場合の性能(実効MFlops値)に比べ、SMPシステムでは13%(図C-5)、NUMAシステムで20%(図C-6)の低下で収まることを明らかにした。

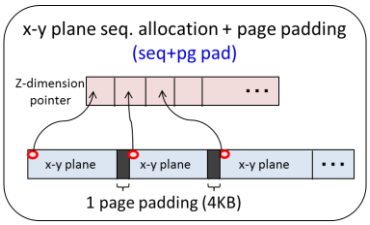


図 C-3 最適化手法（データレイアウト(左図 aio)，ブロック形状，コア間 work-share 方式）により実行時間が 55%(aio),59%(mmap)に減少（右図）

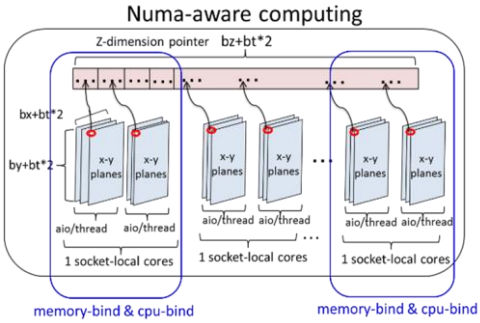


図 C-4　NUMA 向け最適化：55%性能向上

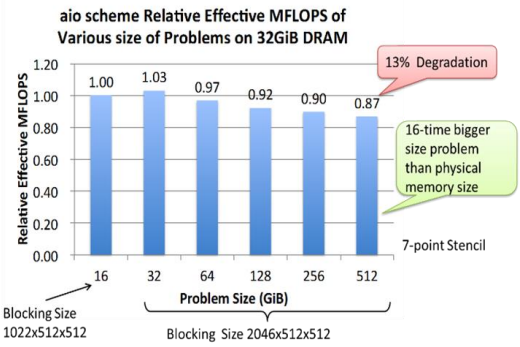


図 C-5　aio 手法（SMP）ステンシル計算相

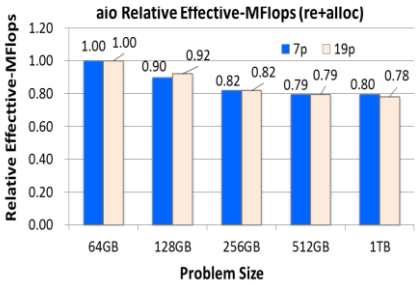
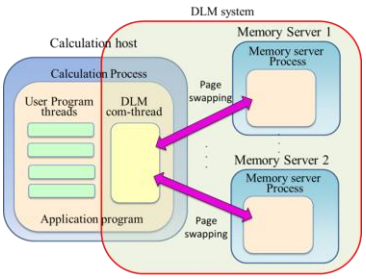


図 C-6　aio 手法（NUMA）ステンシル計算

さらに、時間ブロック化による冗長計算を削減するFlash SSD向けステンシルアルゴリズムを開発中であり、mmap方式とaio方式における効果を解析した。

実行時情報利用によるデータ配置と限定的メモリ同期方式の設計・実装

上述の Flash 向けステンシル処理最適化の知識を基に、問題サイズや用いるハードウェアに応じて、誰でも容易に Flash SSD を用いた主メモリサイズを超えるステンシル計算を高性能で処理できるように、実行時自動パラメータ設定機構 Blk-Tune を新たに開発した。この機構は、実行時に、用いる計算システムのハードウェア情報（Flash SSD サイズ、DRAM メモリサイズ、キャッシュサイズ、コア数、ソケット数など）を取得し、これに合わせて自動的に各メモリ階層（Flash SSD, DRAM, L3-Cache）における適切なデータサイズ、形状を計算し、コア、ソケット数に合わせた affinity 設定と work-share 設定を行い、実行する。これによりユーザは、(1)処理したい問題サイズ(2)時間ステップ数(3) Flash SSD のパス名の 3 つを指定すれば、実行システムの各メモリ階層サイズ、CPU コア、ソケット数適切なパラメータが自動設定され、合わせた実行ができる。



Blk-Tune の開発当初は冗長計算なしテンポラルブロッキングアルゴリズムを主な対象としていたが、平成 28 年度にはさらに、用いる CPU の計算性能とフラッシュの IO 性能の指標を与えることで、冗長計算有のテンポラルブロッキングアルゴリズムにも適用可能とした。

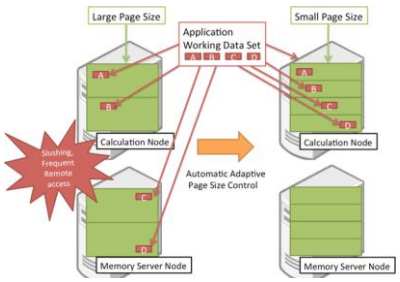


図 C-8　ページサイズ変更とスラッシング回避の原理

遠隔メモリアクセスを利用した水平方向へのメモリ拡張研究としては、平成 26 年度に，分散大容量メモリシステム DLM（図 C-7）に，自動適応型ページサイズ制御（AAPC）（図 C-8）（応用プログラムの各ループのワーキングデータセットサイズに応じ、実行時にページサイズを変更し、ページフェッチ slashing を回避する）を、新たに導入し動作確認した。

中間報告時点の実装では、マルチスレッドプログラムへの対応に課題があったが、その後動的に生成・消滅するユーザスレッドに対応する遠隔ページフェッチ機能、ページ交換プロトコルの改良を行った。具体的には、計算ノードの遠隔ページ受信時に複数ユーザスレッドによる当該ページの不正アクセスを防止するためのサスペンド機能においてユーザスレッド消滅変動に対し強固な実装とした。さらに、遠隔メモリサーバとのデータ送受信を行う通信スレッドに加え、受信スレッドを新たに増設し、２つの DLM システムスレッドによるページ交換プロトコルを複数実装し、その効果、安定性、性能を調査した。

多数ノードに分散するメモリを有効利用するためのメモリ資源割当処理方式の設計・実装

垂直方向のメモリ階層利用では、平成27年度に、**aio方式のマルチノード拡張**として、SSD搭載クラスタ向け大規模ステンシル計算アルゴリズムを開発した。最新4ノードサーバ(バス接続SSD, FDR-InfiniBand)と, TSUBAME2.5(SATA-SSD, QDRx2－IB)における性能調査では、ノード数　x　ローカルSSD容量の問題サイズ(1TiB－4TiB)のステンシル計算が、バス接続型SSDでは95%、SATA接続SSDでは90%の並列効率で処理できる(図C-9,C-10)。したがって利用するノード数(SSD数)を増やすことでほとんど性能劣化なくSSDを主メモリ拡張として用い、クラスタ全体の主メモリ総量を超える大規模ステンシル計算を行うことができる。

これまでは高速大容量SSD搭載クラスタが利用できず、例えばTSUBAME2.5ではローカルSSD容量が主メモリの2倍程度であり、速度も数百MB/sに限られていた。平成29年8月に稼働開始したTSUBAME3.0など、大容量SSD搭載クラスタにおいて、各ノード搭載SSD容量　×　ノード数　分を大容量メモリとして利用する性能実験も行う予定である。また、前述の自動パラメータ設定システムのように、与えられた問題サイズと、実行システムの構成(ノード数、メモリ容量、SSD容量)などから、各ノードへのデータ分割方式(ブロック分割)についても、最善手法を提案するシステムを構築することを考えている。

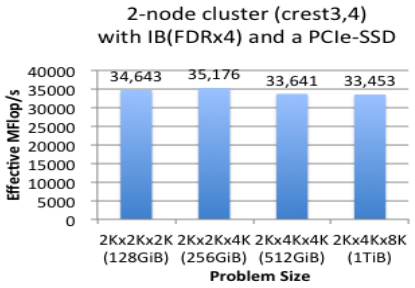


図 C-9　2 ノードサーバにおける実効性能

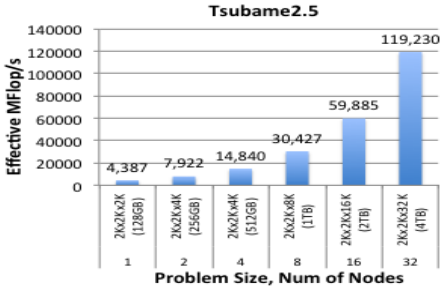


図 C-10　TSUBAME 複数(1-32)ノードにおける実効性能

さらに、マルチスレッド対応型**ソフトウェア分散共有メモリ Multi-SMS**(図 C-11)を設計し、実装をおこなった。これにより、各ノードでマルチスレッドプログラム(pthread、OpenMP)が動作するような共有メモリプログラミングがクラスタ上で可能となる．2 ノードサーバ，TSUBAME2.5 において、マルチノード・マルチスレッドによるステンシル計算の動作確認を行った。マルチスレッドのアプリケーションの際に課題があり、その点への実装の対応を行った。プログラミングインタフェース面において、以前に開発したデータ分散メモリ API (MpC)を移植する作業を進めた。すなわち、マルチノードへのデータ分散配置を行う API を用い、過去に開発した C トランスレータとの連携により、データのノード分散配置より容易にユーザが利用できる環境を実現した。

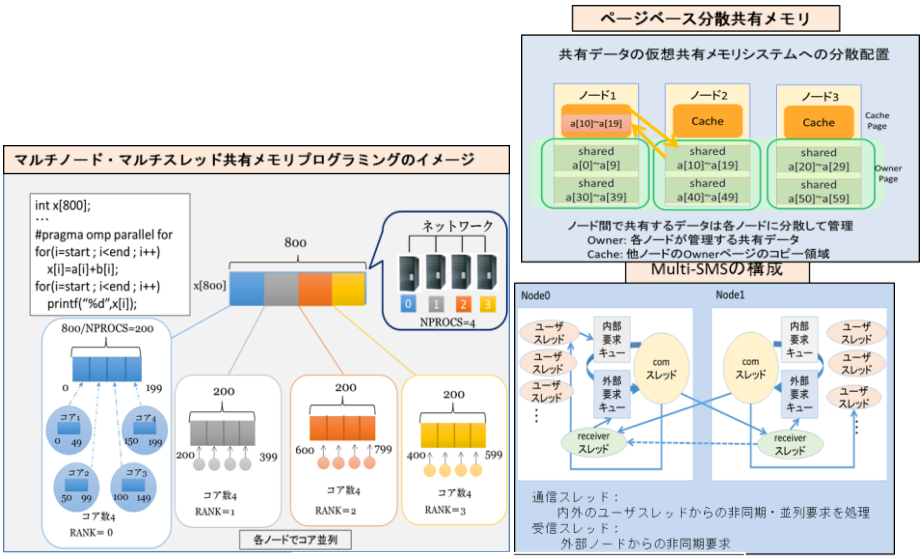


図 C-11 分散共有メモリ：Multi-SMS

[開発ソフトウェア]

公開予定:

Multi-Mem Stencil Series

- 機能:SSD を利用することにより大規模演算が可能であり、かつ時間ブロッキングにより速度効率を高めたステンシル計算のパッケージである。計算内容は実アプリではない単純なステンシルであるが、メモリ階層の性能バランスを評価するために利用可能である。
- 特徴:基本版(SMP/NUMA 版, mmap-swap/aio 版 4種)、自動パラメータ設定つき aio 版、マルチノード版からなる。
- 前提とするマシン: SSD を備える一台もしくは複数台のマシン。aio 機能を用いる場合はそれに対応した Linux バージョンであること。
- 潜在的ユーザ:SSD を含む大規模演算に興味があり、メモリ階層の性能のバランスを評価したいユーザ。

DLM

- 機能:遠隔ページング用ライブラリ。他計算ノードのメモリを利用することにより既存アプリケーションの問題規模を拡大することができる。
- 特徴:マルチスレッドのアプリケーションに対応する。ページサイズ可変機構の有無を切り替えることができる。
- 前提とするマシン: 高速ネットワークで接続された複数台のマシン。Intel プロセッサ・Linux OS で動作確認されている。
- 潜在的ユーザ:既存アプリケーションの問題規模拡大を行いたいユーザ。

§4 成果発表等

(1)原著論文発表 (国内(和文)誌 2 件、国際 (欧文) 誌 28 件)

遠藤グループ(平成 27 年度以降は遠藤主担当分)

[A-1] Katsuki Fujisawa, Toshio Endo, Hitoshi Sato, Makoto Yamashita, Satoshi Matsuoka, Maho Nakata. High-Performance General Solver for Extremely Large-scale Semidefinite Programming Problems. In Proceedings of IEEE/ACM International Conference for High Performance

Computing, Networking, Storage and Analysis (SC12), Saltlake City, November 2012 (DOI: 10.1109/SC.2012.67).
[A-2] Guanghao Jin, Toshio Endo, Satoshi Matsuoka. A Multi-level Optimization Method for Stencil Computation on the Domain that is Bigger than Memory Capacity of GPU . In Proceedings of The Third International Workshop on Accelerators and Hybrid Exascale Systems (AsHES), in conjunction with IEEE IPDPS 2013, Boston, May 2013.
[A-3] Guanghao Jin, Toshio Endo, Satoshi Matsuoka. A Parallel Optimization Method for Stencil Computation on the Domain that is Bigger than Memory Capacity of GPUs . In Proceedings of IEEE Cluster 2013, Indianapolis, USA, September 2013. (DOI 10.1109/CLUSTER.2013.6702633)
[A-4] Katsuki Fujisawa, Toshio Endo, Yuichiro Yasui, Hitoshi Sato, Naoki Matsuzawa, Satoshi Matsuoka, Hayato Waki. Peta-scale General Solver for Semidefinite Programming Problems with over Two Million Constraints. In Proceedings of the International Conference on Parallel and Distributed Processing Symposium 2014 (IPDPS2014), pp.1171-1180, Phoenix, USA, May 22, 2014. [DOI:10.1109/IPDPS.2014.121]
[A-5] Toshio Endo, Guanghao Jin. Software Technologies Coping with Memory Hierarchy of GPGPU Clusters for Stencil Computations. In Proceedings of IEEE Cluster Computing (CLUSTER2014), pp.132-139, Madrid, September 25, 2014. [DOI:10.1109/CLUSTER.2014.6968747]
[A-6] Guanghao Jin, James Lin, Toshio Endo. Efficient Utilization of Memory Hierarchy to Enable the Computation on Bigger Domains for Stencil Computation in CPU-GPU Based Systems. In Proceedings of IEEE International Conference on High Performance Computing and Applications (ICHPCA-2014) Bhubaneswar, December, 2014.
[A-7] 高嵜祐樹, 遠藤敏夫, 松岡聡. GPU クラスタにおける大規模都市気流シミュレーションの最適化と性能モデル. 情報処理学会ハイパフォーマンスコンピューティングと計算科学シンポジウム (HPCS2015), 2015 年 5 月.
[A-8] Kazuki Tsuzuku, Toshio Endo. Power Capping of CPU-GPU Heterogeneous Systems Using Power and Performance Models. In Proceedings of International Conference on Smart Cities and Green ICT Systems (SMARTGREENS2015), 8pages, Lisbon, May 2015.
[A-9] Yuki Tsujita, Toshio Endo. Data Driven Scheduling Approach for the Multi-node Multi-GPU Cholesky Decomposition. In Proceedings of Workshop on Job Scheduling Strategies for Parallel Processing (JSSPP), in conjunction with IPDPS 2015, 15pages, Hyderabad, May 2015.
[A-10] Naoto Sasaki, Kento Sato, Toshio Endo, Satoshi Matsuoka. Exploration of Lossy Compression for Application-level Checkpoint/Restart. In Proceedings of IEEE International Conference on Parallel and Distributed Processing Symposium 2015 (IPDPS2015), Hyderabad, May 2015.
[A-11] Yuki Tsujita, Toshio Endo, Katsuki Fujisawa. The Scalable Petascale Data-Driven Approach for the Cholesky Factorization with Multiple GPUs. In Proceedings of First International Workshop on Extreme Scale Programming Models and Middleware (ESPM2 2015), in conjunction with IEEE/ACM SC15, Austin, November 15, 2015.
[A-12] Toshio Endo, Yuki Takasaki, Satoshi Matsuoka. Realizing Extremely Large-Scale Stencil Applications on GPU Supercomputers . In Proceedings of The 21st IEEE International Conference on Parallel and Distributed Systems (ICPADS 2015), pp. 625-632, Melbourne, December, 2015. [DOI: 10.1109/ICPADS.2015.84]
[A-13] 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 . In proceedings of the ACM International Conference on Computing Frontiers

(CF'16), pp. 274-281, 2016.
[A-14] Toshio Endo. Realizing Out-of-Core Stencil Computations using Multi-Tier Memory Hierarchy on GPGPU Clusters . In Proceedings of IEEE Cluster Computing (CLUSTER2016), pp. 21-29, Taipei, Sep 2016. [DOI: 10.1109/CLUSTER.2016.61]
[A-15] Ryo Matsumiya, Toshio Endo. PGAS Communication Runtime for Extreme Large Data Computation. In Proceedings of Second International Workshop on Extreme Scale Programming Models and Middleware (ESPM2), in conjunction with IEEE/ACM SC16, 8pages, Saltlake City, November 18, 2016. [DOI: 10.1109/ESPM2.2016.007]
[A-16] Satoshi Imamura, Keitaro Oka, Yuichiro Yasui, Yuichi Inadomi, Katsuki Fujisawa, Toshio Endo, Koji Ueno, Keiichiro Fukazawa, Nozomi Hata, Yuta Kakibuka, Koji Inoue, Takatsugu Ono. Evaluating the Impacts of Code-Level Performance Tunings on Power Efficiency. In Proceedings of IEEE International Conference on Big Data (BigData 2016), 6pages, Dec 2016.

旧佐藤グループ（平成 27 年度以降は遠藤グループ佐藤主担当分）
[B-1] Yukinori Sato, Yasushi Inoguchi and Tadao Nakamura. Evaluating Reconfigurable Dataflow Computing Using the Himeno Benchmark. In Proceedings of 2012 International Conference on ReConFigurable Computing and FPGAs (ReConFig2012). Cancun, Dec. 2012. (DOI: 10.1109/ReConFig.2012.6416746).
[B-2] Yukinori Sato, Yasushi Inoguchi and Tadao Nakamura. Whole Program Data Dependence Profiling to Unveil Parallel Regions in the Dynamic Execution. In Proceedings of 2012 IEEE International Symposium on Workload Characterization (IISWC 2012). La Jolla, Nov. 2012. (DOI: 10.1109/IISWC.2012.6402902).
[B-3] 佐藤幸紀. メモリ階層対応ダイナミックコンパイレーション機構の動作原理とコードプロファイリング. 情報処理学会第第 55 回プログラミング・シンポジウム予稿集, p.25-32, 伊東市, 2014 年 1 月. ［概要による査読］
[B-4] Yuichiro Yasui, Katsuki Fujisawa, Yukinori Sato. Fast & Energy-Efficient Breadth-First Search on a Single NUMA System. International Supercomputing Conference 2014 (ISC'14), Lecture Notes in Computer Science Volume 8488, pp. 365-381, 2014. (DOI: 10.1007/978-3-319-07518-1_23)
[B-5] Yuki Matsubara and Yukinori Sato. Online memory access pattern analysis on an application profiling tool. Proceedings of 2014 Second International Symposium on Computing and Networking, pp.602-604, 2014. (DOI:10.1109/CANDAR.2014.86)
[B-6] Yukinori Sato, Yasushi Inoguchi, Tadao Nakamura. Identifying Program Loop Nesting Structures during Execution of Machine Code. IEICE Transaction on Information and Systems, Vol.E97-D, No.9, pp.2371-2385, 2014. (DOI:10.1587/transinf.2013EDP7455)
[B-7] Yukinori Sato, Shimpei Sato, Toshio Endo. Exana: An Execution-driven Application Analysis Tool for Assisting Productive Performance Tuning. In Proceedings of The Second Workshop on Software Engineering for Parallel Systems (SEPS), in conjunction with ACM SPLASH 2015, Pittsburgh, October 27, 2015.
[B-8] Shimpei Sato, Yukinori Sato, Toshio Endo. Investigating Potential Performance Benefits of Memory Layout Optimization based on Roofline Model. In Proceedings of The Second Workshop on Software Engineering for Parallel Systems (SEPS), in conjunction with ACM SPLASH 2015, Pittsburgh, October 27, 2015.
[B-9] Yukinori Sato, Tomoya Yuki and Toshio Endo. ExanaDBT: A Dynamic Compilation System for Transparent Polyhedral Optimizations at Runtime. In Proceedings of ACM International Conference on Computing Frontiers 2017, 10pages, Siena, May 2017.
[B-10] Yukinori Sato and Toshio Endo. An Accurate Simulator of Cache-line Conflicts to Exploit the Underlying Cache Performance. In Proceedings of 23rd International

European Conference on Parallel and Distributed Computing (Euro-par 2017), Santiago, Spain, August 2017.

緑川グループ

[C-1] Hiroko Midorikawa, Hideyuki Tan and Toshio Endo, An Evaluation of the Potential of Flash SSD as Large and Slow Memory for Stencil Computations, Proceedings of the 2014 International Conference on High Performance Computing and Simulation (IEEE HPCS2014) (ISBN 978-1-4799-5311-0), pp.268-277, (2014.7)

[C-2] Hiroko Midorikawa, Hideyuki Tan, Locality-Aware Stencil Computations using Flash SSDs as Main Memory Extension, Proceeding of IEEE/ACM International Symp. on Cluster, Cloud and the Grid Computing CCGrid2015, Shenzhen, China, pp.1163-1168, (DOI 10.1109/CCGrid.2015.126), (2015.5/6)

[C-3] Hiroko Midorikawa: ″Blk-Tune: Blocking Parameter Auto-Tuning to Minimize Input-Output Traffic for Flash-based Out-of-Core Stencil Computations″, The Eleventh International Workshop on Automatic Performance Tuning iWAPT2016, in conjunction with IPDPS2016, proc. of IPDPW2016, pp.1516-1526, 2016

[C-4] Hiroko Midorikawa, Hideyuki Tan: ″Evaluation of Flash-based Out-of-core Stencil Computation Algorithms for SSD-Equipped Clusters″, The 22nd IEEE International Conference on Parallel and Distributed Systems ICPADS2016 , pp.1031-1040, 2016

(2)その他の著作物(総説、書籍など)

遠藤グループ

[A-17] 藤澤克樹, 遠藤敏夫. 大規模半正定値計画問題に対する内点法アルゴリズムの高速計算 . 東京工業大学学術国際情報センター, TSUBAME e-Science Journal, No.7, 2012 年 12 月, pp. 2--6.

[A-18] Katsuki Fujisawa, Toyotaro Suzumura, Hitoshi Sato, Koji Ueno, Yuichiro Yasui, Keita Iwabuchi, Toshio Endo. Advanced Computing & Optimization Infrastructure for Extremely Large-Scale Graphs on Post Peta-Scale Supercomputers. Fujisawa, Katsuki, Shinano, Yuji, and Waki, Hayato (eds.), Optimization in the Real World - Toward Solving Real-World Optimization Problems -, Series of Mathematics for Industry, Springer, pp. 1-13, 2016. [DOI: 10.1007/978-4-431-55420-2_1]

(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 4 件、国際会議 3 件)

遠藤グループ

[A-17] 遠藤 敏夫. TSUBAME2.0/2.5 スーパーコンピュータとポストペタスケール時代に向けた課題. 日本学術会 議 電気電子工学委員会 URSI 分科会 無線通信システム信号処理小委員会 (URSI-C) 第 22 期 第 8 回公開研究会, 沖縄, 2013 年 9 月 27 日.

[A-18] 遠藤敏夫. ポストペタスケール時代に向けた格子系アルゴリズムの局所性向上. 自動チューニング研究会 第 5 回 自動チューニング技術の現状と応用に関するシンポジウム, 東京, 2013 年 12 月 25 日.

[A-19] Toshio Endo. Software Technology that Deals with Deeper Memory Hierarchy in Post-petascale Era, The Japanese Extreme Big Data Projects Workshop, Fukuoka, Japan, February 26th 2014.

*[A-20] Toshio Endo. Harnessing Memory Hierarchy towards Extreme Fast and Big Simulations. 2015 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing. Taipei, Feb 27, 2015.

[A-21] Toshio Endo. Harnessing Multi-tier Memory Hierarchy of GPU, Host and Flash. 2016 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing, Taipei, February 20, 2016.
[A-22] 遠藤敏夫． NVMe Flash SSD 利用による 科学技術計算の 規模と演算速度の両立, 第 7 回ビッグデータ基盤研究会， 東京, 2017 年 3 月 29 日.

旧佐藤グループ
[B-11] 佐藤幸紀. バイナリ変換による実行駆動型アプリケーションプロファイリングとそのチューニングへの応用. 自動チューニング研究会オープンアカデミックセッション, 東京大学, 2013 年 10 月 16 日.

② 口頭発表 （国内会議 51 件、国際会議 13 件）

遠藤グループ(平成 27 年度以降は遠藤主担当分)
[A-23] 遠藤 敏夫． ポストペタスケール時代のメモリ階層の深化に対応するソフトウェア技術. JST ERATO 湊離散構造処理系プロジェクトセミナー， 札幌, 2012 年 12 月.
[A-24] 金 光浩, 遠藤 敏夫, 松岡 聡． GPU メモリ容量を超える問題規模に対応する高性能ステンシル計算法 ． ハイパフォーマンスコンピューティングとアーキテクチャの評価に 関する北海道ワークショップ(HOKKE-20), 情報処理学会研究報告, 2012-ARC-194/HPC-137, 6 pages, 2012 年 12 月.
[A-25] 野村 哲弘, 遠藤 敏夫, 松岡 聡． TSUBAME2.0 における Multi-rail InfiniBand ネットワークの性能評価 ． ハイパフォーマンスコンピューティングとアーキテクチャの評価に 関する北海道ワークショップ(HOKKE-20), 情報処理学会研究報告, 2012-ARC-194/HPC-137, 2012 年 12 月.
[A-26] 遠藤敏夫. 並列プログラムをメモリ利用階層利用可能とするランタイム. 2013 年並列/分散/協調処理に関する『北九州』サマー・ワークショップ(SWoPP 北九州 2013), 情報処理学会研究報告 2013-HPC-140, 北九州国際会議場, 2013 年 8 月 2 日.
[A-27] 遠藤 敏夫, 佐藤 幸紀, 緑川 博子. JST CREST ポストペタ領域遠藤チーム ポストペタスケール時代のメモリ階層の深化に対応するソフトウェア技術. HPC ワークショップ金沢 2013, 金沢, 2013 年 10 月 29 日.
[A-28] Jin Guanghao, Endo Toshio, Matsuoka Satoshi . Multi-level Temporal Blocking for Stencil Computation for Memory Hierarchy on TSUBAME2.5. IPSJ SIGHPC, SIGHPC Technical report 2014-HPC-143, Nanao, Japan, March 4th 2014.
[A-29] Guanghao Jin, Mohamed Wahib, Naoya Maruyama, Toshio Endo, Satoshi Matsuoka. Locality Optimizations for Stencil Computations: Algorithms and Implementations, Workshop on Programming Abstractions for Data Locality (PADAL 2014), Lugano, April 28, 2014.
[A-30] 遠藤 敏夫． ポストペタスケール時代に向けた格子系アルゴリズムの局所性向上手法 ． 日本計算工学会 第 19 回計算工学講演会， オーガナイズドセッション 19， 広島, 2014 年 6 月 12 日.
[A-31] 佐々木尚人, 佐藤賢斗, 遠藤敏夫, 松岡聡． 実アプリケーションにおけるウェーブレット変換を用いたチェックポイントデータの非可逆圧縮手法． 2014 年並列／分散／協調処理に関する『新潟』サマー・ワークショップ (SWoPP 新潟 2014), 2014-HPC-145 No.7, 新潟, 2014 年 7 月 28 日.
[A-32] 都筑一希, 遠藤敏夫． CPU・GPU 混載ノードにおける電力・性能モデルを用いたパワーキャッピング手法． 2014 年並列／分散／協調処理に関する『新潟』サマー・ワークショップ (SWoPP 新潟 2014), 2014-HPC-145 No.10, 新潟, 2014 年 7 月 28 日.
[A-33] 辻田裕紀, 遠藤敏夫． マルチノード GPU 上のコレスキー分解へのデータドリブン型アルゴリズムの適用 ． 2014 年並列／分散／協調処理に関する『新潟』サマー・ワークショップ (SWoPP 新潟 2014), 2014-HPC-145 No.46, 新潟, 2014 年 7 月 30 日.

[A-34] 遠藤敏夫. 格子系アルゴリズムの局所性向上と HHRT ライブラリ上の実装. メモリプラスワークショップ － メモリとファイルストレージと OS と－, JAIST 品川サテライトオフィス. 2014 年 9 月 17 日.
[A-35] 高嵜 祐樹, 遠藤 敏夫, 松岡 聡. GPU クラスタ上の実ステンシルアプリケーションの大規模化に向けた局所性向上の評価 . 情報処理学会ハイパフォーマンス研究会, 2014-HPC-146 No.23, 沖縄, 2014 年 10 月 3 日.
[A-36] Toshio Endo. Software Technology that Deals with Deeper Memory Hierarchy in Post-petascale Era. 2014 ATIP Workshop: Japanese Research Toward Next-Generation Extreme Computing, New Orleans, November 17, 2014.
[A-37] Toshio Endo. Locality Improvement of Stencil Computations for Big Simulations. JST/CREST International Symposium on Post Petascale System Software (ISP2S2), Kobe, December 4, 2014.
[A-38] Tianqi Xu, Jin Guanghao, Endo Toshio, Matsuoka Satoshi. Efficient Utilization of Multi-level Memory System for Stencil Computation, IPSJ SIG Technical Report, 2014-HPC-147 No.10, 7 pages, 小樽, 2014 年 12 月 10 日.
[A-39] 高嵜 祐樹, 遠藤 敏夫, 松岡 聡. GPU 搭載システムにおける都市気流シミュレーションの大規模化と性能モデル . 情報処理学会ハイパフォーマンス研究会, 2015-HPC-148 No.13, 別府, 2015 年 3 月 3 日.
[A-40] Toshio Endo, Satoshi Matsuoka. Realizing Extremely Large-Scale Stencil Applications on GPU Supercomputers with a Memory Hierarchy Management Runtime Library. Workshop on Programming Abstractions for Data Locality (PADAL 2015), Berkeley, June 25, 2015.
[A-41] Yuki Tsujita, Toshio Endo, Katsuki Fujisawa. The Scalable Petascale Data-Driven Approach for the Cholesky Factorization with multiple GPUs . 2nd Annual Meeting on Advanced Computing System and Infrastructure (ACSI2016), 2016 年 1 月 19 日.
[A-42] 遠藤 敏夫. 大規模・高性能演算のための多階層メモリの活用 . 情報処理学会研究報告, 2015-HPC-153 No.14, 7pages, 2016 年 3 月 2 日.
[A-43] 下川辺 隆史, 遠藤 敏夫, 青木 尊之. GPU デバイスメモリを超える計算を可能とするためのステンシル計算フレームワークの拡張とその性能評価 . 日本計算工学会 第 21 回計算工学講演会, B-5-3, 新潟, 2016 年 6 月 1 日.
[A-44] 松岡 聡, 天野 英晴, 中島 研吾, 井上 弘士, 工藤 知宏, 丸山 直也, 田浦 健次朗, 岩下 武史, 片桐 孝洋, 塙敏博, 遠藤 敏夫. ポストムーア時代における FLOPS から BYTES への変革 . 並列/分散/協調処理に関するサマーワークショップ(SWoPP2016), 情報処理学会研究報告, 2016-HPC-155 No.32, 2016 年 8 月 10 日.
[A-45] 松宮 遼, 遠藤 敏夫. Flash SSD を含む多階層メモリを活用する PGAS ランタイムシステム . 並列/分散/協調処理に関するサマーワークショップ(SWoPP2016), 情報処理学会研究報告, 2016-HPC-155 No.31, 2016 年 8 月 9 日.
[A-46] 遠藤敏夫. GDDR・DDR・Flash の多階層メモリを利用するランタイムライブラリと大規模ステンシルへの応用, 第2回メモリプラスワークショップ, 東京, 2016 年 8 月 31 日.
[A-47] 黒田 勝汰, 遠藤 敏夫, 松岡 聡. ディレクティブによる時空間ブロッキングの自動適用 . 情報処理学会研究報告, 2016-HPC-157 No.18, 2016 年 12 月 22 日.
[A-48] 田邊 昇, 遠藤 敏夫. 中遅延大容量メモリ階層出現のインパクトと新たな対応に関する初期検討 . 情報処理学会研究報告, 2016-HPC-157 No.11, 2016 年 12 月 22 日.
[A-49] 田邊 昇, 遠藤 敏夫. 疎行列系アプリケーション性能の主記憶遅延増加の影響評価 . 情報処理学会研究報告, 2017-HPC-158 No.15, 2017 年 3 月 9 日.
[A-50] 伊藤祐貴, 松宮遼, 遠藤敏夫. メモリ階層の利用によってGPUメモリ容量を超える深層学習手法 . The 1st. cross-disciplinary Workshop on Computing Systems, Infrastructures, and Programming (xSIG 2017), 東京, 2017 年 4 月.
[A-51] 松宮 遼, 遠藤 敏夫. vGASNet: メモリ階層深化に向けたスケーラブルな低レイヤ通信ライブラリ . 並列/分散/協調処理に関するサマーワークショップ(SWoPP2017), 情報処理

学会研究報告, 2017-HPC-160 No.7, 2017 年 7 月 26 日
[A-52] 松岡 聡, 遠藤 敏夫, 額田 彰, 三浦 信一, 野村 哲弘, 佐藤 仁, 實本 英之, Drozd Aleksandr. HPC とビッグデータ・AI を融合するグリーン・クラウドスパコン TSUBAME3.0 の概要 . 並列/分散/協調処理に関するサマーワークショップ(SWoPP2017), 情報処理学会研究報告, 2017-HPC-160 No.29, 2017 年 7 月 28 日.
[A-53] 田邊 昇, 遠藤 敏夫. Intel Xeon Phi における主記憶遅延増加の影響評価 . 並列/分散/協調処理に関するサマーワークショップ(SWoPP2017), 情報処理学会研究報告, 2017-HPC-160 No.12, 2017 年 7 月 26 日.
[A-54] 伊藤 祐貴, 松宮 遼, 遠藤 敏夫. ooc_cuDNN: GPU 計算機のメモリ階層を利用した大規模深層学習ライブラリの開発 . 並列/分散/協調処理に関するサマーワークショップ(SWoPP2017), 情報処理学会研究報告, 2017-HPC-160 No.38, 2017 年 7 月 28 日.

旧佐藤グループ (平成 27 年度以降は遠藤グループ佐藤主担当分)

[B-12] Yukinori Sato, Hiroko Midorikawa, and Toshio Endo. Identifying working data set of particular loop iterations for dynamic performance tuning. In 6th Workshop on Architectural and Microarchitectural Support for Binary Translation (AMAS-BT2013). Held in conjunction with the 40th Int'l Symposium on Computer Architecture (ISCA-40), Tel-Aviv, Israel, pp. 1-6, Jun. 24, 2013.
[B-13] 佐藤幸紀. バイナリ変換による透過的なループ構造解析とコード実行時の区間実行時間の計測. 2013 年並列/分散/協調処理に関する『北九州』サマー・ワークショップ(SWoPP 北九州 2013). 情報処理学会研究報告 2013-HPC-140, pp.1-8, 2013 年 7 月 31 日.
[B-14] 松原裕貴, 佐藤幸紀. メモリトレース解析によるアクセスパターンのモデル化. 2013 年並列/分散/協調処理に関する『北九州』サマー・ワークショップ(SWoPP 北九州 2013). 情報処理学会研究報告 2013-HPC-140, pp. 1-6, 2013 年 8 月 2 日.
[B-15] Yukinori Sato. Architecting Dynamic Compilation Mechanisms for Transparent Performance Tuning of Data Locality in Memory Subsystem. The International Workshop on Innovative Architecture for FutureGeneration High-Performance Processors and Systems (IWIA) 2014, Hawaii, USA, Mar. 19, 2014.
[B-16] 松原裕貴, 佐藤幸紀. テンポラルブロッキングを適用したステンシルコードにおける階層的メモリアクセスパターン解析. 2014 年並列／分散／協調処理に関する『新潟』サマー・ワークショップ(SWoPP 新潟 2014). 朱鷺メッセ新潟コンベンションセンター. 2014 年 7 月 28 日.
[B-17] 佐藤幸紀. Exana ツールによるメモリアクセスプロファイリング. メモリプラスワークショップ--メモリとファイルストレージと OS と . JAIST 品川サテライトオフィス. 2014 年 9 月 17 日.
[B-18] Yukinori Sato. Exana: An Application Profiling and Optimization Infrastructure for Accelerating Systems with Deeper Memory Hierarchy. JST/CREST International Symposium on Post Petascale System Software (ISP2S2), December 2014.
[B-19] 佐藤 幸紀 , 遠藤 敏夫. 実行駆動型キャッシュシミュレーションおよびメモリ参照特性解析におけるオーバーヘッドの評価. 2015 年並列／分散／協調処理に関する『別府』サマー・ワークショップ(SWoPP2015). システムアーキテクチャ研究会報告(ARC), 2015-ARC-216(31), pp. 1-7.
[B-20] 佐藤 真平, 佐藤 幸紀, 遠藤 敏夫. ルーフラインモデルによる性能幅推定とステンシル計算コードにおけるメモリレイアウト最適化による性能最大化. 2015 年並列／分散／協調処理に関する『別府』サマー・ワークショップ(SWoPP2015). システムアーキテクチャ研究会報告(ARC), 2015-ARC-216(32), pp. 1-6.
[B-21] 佐藤真平, 佐藤幸紀, 遠藤敏夫. テンポラルブロッキングを適用したステンシル計算コードの SIMD 化とルーフラインモデルを用いた性能解析. 第 151 回ハイパフォーマンスコンピューティング研究発表会. 2015 年 10 月.
[B-22] Shimpei Sato, Yukinori Sato, Toshio Endo. A Cache-aware Temporal Blocking Method for 3D Stencil Computation . 3rd International Workshop on High-Performance Stencil

Computations (HiStencils 2016), In conjunction with HiPEAC 2016, Prague, January 18, 2016.
[B-23] Yukinori Sato, Toshio Endo. Dynamic Compilation for Transparent Data Locality Analysis and Memory Subsystem Tuning . The International Workshop on Architectural and Micro-Architectural Support for Dynamic Optimization (AMAS-DO), In conjunction with CGO 2016, Barcelona, March 13, 2016.
[B-24] 佐藤 真平, 佐藤 幸紀, 遠藤 敏夫. ステンシル計算コードの性能とメモリレイアウトの関係性について . 並列/分散/協調処理に関するサマーワークショップ(SWoPP2016), 情報処理学会研究報告, 2016-HPC-155 No.37, 2016 年 8 月 10 日.
[B-25] 佐藤幸紀. 動的バイナリ変換によるメモリ階層性能プロファイリングと透過的メモリ階層チューニング, 第2回メモリプラスワークショップ, 東京, 2016 年 8 月 31 日.
[B-26] 佐藤幸紀, 幸朋矢, 遠藤敏夫. 透過的メモリ階層チューニングのための動的バイナリ変換機構の設計と開発 . 情報処理学会研究報告, 2016-ARC-216 No.35, 2017 年 1 月 25 日.
[B-27] 幸 朋矢, 佐藤 幸紀, 遠藤 敏夫. Polyhedral コンパイラを用いたタイリングパラメータ自動調整ツールのメニーコア環境での評価 . 並列/分散/協調処理に関するサマーワークショップ(SWoPP2017), 情報処理学会研究報告, 2017-HPC-160 No.34, 2017 年 7 月 28 日.

緑川グループ
[C-5] 緑川博子, 丹英之:"メモリサイズを超えるデータ処理を目的としたバス接続型 SSD の性能評価", 情報処理学会、ハイパフォーマンス研究会 Vol.2013-HPC-140, No.44, pp.1-6, (2013, 8/2)
[C-6] 丹英之,緑川博子, フラッシュ向け Linux スワップシステムの評価, 電子情報通信学会, コンピュータシステム研究会 Vol.113, No.282, pp.61-66, (2013, 11)
[C-7] 丹英之,緑川博子:"フラッシュ SSD をメモリセマンティクス API で用いるための予備調査", ハイパフォーマンスコンピューティングと計算科学シンポジウム HPCS2014, HPCS2014 論文集,(2014, 1)
[C-8] 緑川博子, 丹英之:"大規模ステンシル計算のための Flash SSD 向けテンポラルブロキングの性能評価", 情報処理学会、ハイパフォーマンス研究会 Vol.2014-HPC-145, No.22, pp.1-9, (2014, 7/29)
[C-9] H.Midorikawa, "Using a Flash as Large and Slow Memory for Stencil Computations", Flash Memory Summit 2014, Santa Clara, (2014, 8)
[C-10] 丹英之,緑川博子:"ブロックデバイス非同期 I/O によるフラッシュストレージを用いたステンシル計算の性能評価",電子情報通信学会、コンピュータシステム研究会 CPSY2014-52, pp.31-36, (2014, 10/10)
[C-11] Hiroko Midorikawa, "An Evaluation of Flash SSDs as Main Memory Extension for Stencil Computation", JST/CREST International Symposium on Post Peta-scale System Software (ISP2S2), (2014, 12)
[C-12] Hiroko Midorikawa, "Using Flash SSDs as Main Memory Extension with a Locality-aware Algorithm ", Non-Volatile Memories Workshop 2015, (2015, 3)
[C-13] 丹英之,緑川博子:"SSD 搭載クラスタを用いた大規模ステンシル計算のための out-of-core アルゴリズム", ハイパフォーマンス研究会 Vol.2015-HPC-149, No.4, pp.1-7, (2015, 6/25)
[C-14] 緑川博子, 丹英之:"フラッシュ SSD を用いた out-of-core ステンシル計算の性能向上手法とその効果", 電子情報通信学会、コンピュータシステム研究会 信学技報 CPSY2015-41, pp.241-246, (2015, 8/6)
[C-15] Hiroko Midorikawa: Minimizing Flash I/O Traffic with Explicit I/Os for Efficient Out-of-Core Algorithms, Non-Volatile Memories Workshop 2016 (2016/3/9)
[C-16] 緑川博子, 丹英之: Flash を用いた out-of-core ステンシル計算のための最適ブロッキングパラメタ自動チューニングシステム, 情報処理学会, ハイパフォーマンスコンピューティング

研究会(HPC) (2016/8/1)
[C-17] 緑川博子. Flash 利用による out-of-core ステンシルアルゴリズムとブロックサイズ自動チューニングシステム, 第2回メモリプラスワークショップ, 東京, 2016 年 8 月 31 日.
[C-18] 白澤卓磨, 緑川博子, 甲斐宗徳: マルチノードマルチコア向け分散共有メモリにおけるデータ分散配置 API の導入, 第 15 回情報科学技術フォーラム FIT2016 (2016/9/9)
[C-19] 大浦陽, 緑川博子, 甲斐宗徳: 遠隔メモリ利用による Out-Of-Core OpenMP プログラムの性能評価実験, 第 15 回情報科学技術フォーラム FIT2016 (2016/9/9)
[C-20] 緑川 博子, 北川 健司, 大浦 陽: マルチスレッドプログラム向け遠隔メモリサーバにおけるページ交換プロトコルの評価実験, 並列/分散/協調処理に関するサマーワークショップ (SWoPP2017), 情報処理学会研究報告, 2017-HPC-160 No.36 (2017/7/28)

③ ポスター発表 (国内会議 21 件、国際会議 29 件)

遠藤グループ(平成 27 年度以降は遠藤主担当分)
[A-55] Toshio Endo. Memory Hierarchy Aware Software Stack. IEEE/ACM International Conference for High Performance Computing, Networking, Storage and Analysis (SC12), Saltlake City, November 2012 (展示会場の東工大ブースにて研究内容に関するポスター展示)
[A-56] Keisuke Fukuda, Naoya Maruyama, Toshio Endo, Miquel Pericas, Satoshi Matsuoka. Fast Multipole Method on a Heterogeneous Dynamic Task Scheduling Engine, GPU Technology Conference (GTC), poster session, San Jose, March 2013.
[A-57] Katsuki Fujisawa, Toshio Endo, Hitoshi Sato, Yuichiro Yasui, Naoki Matsuzawa, Hayato Waki. Peta-Scale General Solver for Semidefinite Programming Problems with over Two Million Constraints, International Supercomputing Conference 2013(ISC'13), Leipzig, Germany, June 17th 2013.
[A-58] Katsuki Fujisawa, Toshio Endo, Hitoshi Sato, Yuichiro Yasui, Naoki Matsuzawa, Hayato Waki. Peta-Scale General Solver for Semidefinite Programming Problems with Over Two Million Constraints, IEEE/ACM SC13, poster session, Denver, November 19th 2013.
[A-59] Toshio Endo, Guanghao Jin, Satoshi Matsuoka. Dealing with Deeper Memory Hierarchy. The International Conference for High Performance Computing, Networking, Storage and Analysis (SC13), Denver, 18-21 Nov. 2013. (SC 展示会場の東工大ブースにて研究内容に関するポスター展示)
[A-60] Guanghao Jin, Tomoki Kawamura, Naoya Maruyama, Toshio Endo, Satoshi Matsuoka. Optimization methods for efficient utilization of memory hierarchy on GPU cluster. GPU Technology Conference, poster session, San Jose, March 24th 2014.
[A-61] Guanghao Jin, Toshio Endo and Satoshi Matsuoka. Efficient Utilization of Memory Hierarchy on GPU Clusters: Optimization Methods and Performance Models. HPC in Asia poster session, held with ISC'14, Leipzig, June 2014.
[A-62] Naoto Sasaki, Kento Sato, Toshio Endo and Satoshi Matsuoka. Exploration of Application-level Lossy Compression for Fast Checkpoint/Restart. HPC in Asia poster session, held with ISC'14, Leipzig, June 2014.
[A-63] Guanghao Jin, Toshio Endo. Data Management and Loop Controlling to Surpass Memory Capacity of GPU in OpenACC Framework. GTC Technology Conference Japan, 東京, 2014 年 7 月 16 日.
[A-64] 高嵜 祐樹, 遠藤 敏夫, 松岡 聡. GPU クラスタ上での実アプリケーションに対するテンポラルブロッキングの実装と性能評価. GTC Technology Conference Japan, 東京, 2014 年 7 月 16 日.
[A-65] Toshio Endo, Yukinori Sato, Hiroko Midorikawa. Software Technology that Deals with Deeper Memory Hierarchy in Post-petascale Era. The 2014 International Conference for High

Performance Computing, Networking, Storage, and Analysis (SC '14), 17-20 November 2014. (SC 展示会場の JST-CREST ブースにて研究内容に関するポスター展示)
[A-66] Toshio Endo. Dealing with Deeper Memory Hierarchy. The 2014 International Conference for High Performance Computing, Networking, Storage, and Analysis (SC '14), 17-20 November 2014. (SC 展示会場の東工大ブースにて研究内容に関するポスター展示)
[A-67] Toshio Endo, Yukinori Sato, Hiroko Midorikawa. Software Technology that Deals with Deeper Memory Hierarchy in Post-petascale Era. JST/CREST International Symposium on Post Petascale System Software (ISP2S2), poster session, Kobe, December 2, 2014.
[A-68] Guanghao Jin, Toshio Endo. The Efficient Utilization of Memory Hierarchy on GPU Clusters. JST/CREST International Symposium on Post Petascale System Software (ISP2S2), poster session, Kobe, December 2, 2014.
[A-69] Kazuki Tsuzuku, Toshio Endo. Power Capping of CPU-GPU Heterogeneous Systems using Power and Performance Models. GPU Technology Conference (GTC 2015), poster session, San Jose, March, 2015.
[A-70] 辻田裕紀, 遠藤敏夫. マルチノード・マルチ GPU 上のコレスキー分解に対するデータドリブン型アルゴリズム手法. 情報処理学会ハイパフォーマンスコンピューティングと計算科学シンポジウム (HPCS2015), ポスターセッション, 東京, 2015 年 5 月 19 日.
[A-71] 遠藤敏夫. 異種プロセッサマシンのメモリ階層を活用する HHRT ライブラリの実装. 情報処理学会ハイパフォーマンスコンピューティングと計算科学シンポジウム (HPCS2015), ポスターセッション, 東京, 2015 年 5 月 19 日.
[A-72] Guanghao Jin, Toshio Endo. Efficient Utilization of GPU Cluster Resource for Stencil Computation. 情報処理学会ハイパフォーマンスコンピューティングと計算科学シンポジウム (HPCS2015), ポスターセッション, 東京, 2015 年 5 月 19 日.
[A-73] Guanghao Jin, Toshio Endo. High Productive Framework to Enable Stencil Computation on Bigger Domains on TSUBAME2.5 , GPU Technology Conference Japan (GTC Japan), poster session, Tokyo, September 18, 2015.
[A-74] Kazuki Tsuzuku, Toshio Endo. Online Power Capping of CPU-GPU Heterogeneous Systems, GPU Technology Conference Japan (GTC Japan), poster session, Tokyo, September 18, 2015.
[A-75] Toshio Endo, Yukinori Sato, Hiroko Midorikawa. Software Technology that Deals with Deeper Memory Hierarchy in Post-petascale Era. The 2015 International Conference for High Performance Computing, Networking, Storage, and Analysis (SC '15), November 2015. (SC 展示会場の JST-CREST ブースにて研究内容に関するポスター展示)
[A-76] Toshio Endo. Dealing with Deeper Memory Hierarchy. The 2015 International Conference for High Performance Computing, Networking, Storage, and Analysis (SC '15), November 2015. (SC 展示会場の東工大ブースにて研究内容に関するポスター展示)
[A-77] 松宮遼, 遠藤敏夫, 大山恵弘. 深化する記憶装置階層のための大規模データ処理基盤の提案 . 第 57 回情報処理学会プログラミング・シンポジウム, ポスターセッション, 伊東, 2016 年 1 月.
[A-78] Toshio Endo, Yukinori Sato, Hiroko Midorikawa. Software Technology that Deals with Deeper Memory Hierarchy in Post-petascale Era. The 2016 International Conference for High Performance Computing, Networking, Storage, and Analysis (SC '16), November 2016. (SC 展示会場の JST-CREST ブースにて研究内容に関するポスター展示)
[A-79] Toshio Endo. Dealing with Deeper Memory Hierarchy. The 2016 International Conference for High Performance Computing, Networking, Storage, and Analysis (SC '16), November 2016. (SC 展示会場の東工大ブースにて研究内容に関するポスター展示)
[A-80] Takashi Shimokawabe, Toshio Endo, Naoyuki Onodera, Takayuki Aoki. Performance Evaluation of Wind Simulation Based on a GPU-computing Framework to Realize Large-scale Stencil Computations Beyond Device Memory Capacity. The 7th AICS International Symposium,

Poster session, Kobe, Feb 2017.
[A-81] 伊藤祐貴, 松宮遼, 遠藤敏夫. メモリ階層の利用によってGPUメモリ容量を超える深層学習手法 . The 1st. cross-disciplinary Workshop on Computing Systems, Infrastructures, and Programming (xSIG 2017), ポスターセッション, 東京, 2017 年 4 月.
[A-82] 松宮遼, 遠藤敏夫. Flash SSD を活用する PGAS フレームワークに対する協調キャッシングの導入 . The 1st. cross-disciplinary Workshop on Computing Systems, Infrastructures, and Programming (xSIG 2017), ポスターセッション, 東京, 2017 年 4 月.

旧佐藤グループ (平成 27 年度以降は遠藤グループ佐藤主担当分)

[B-28] Tomoaki Ukezono, Yukinori Sato, Kiyofumi Tanaka. An Analysis for Deeper Memory Hierarchy in HPCS. IEEE/ACM International Conference for High Performance Computing, Networking, Storage and Analysis (SC12), Saltlake City, 12-15 November 2012. (展示会場の JAIST ブースにて研究内容に関するポスター展示)
[B-29] Yuki Matsubara, Yukinori Sato. Dynamic Compilation and Memory Access Analysis Tools for Accelerating Systems with Deeper Memory Hierarchy. The International Conference for High Performance Computing, Networking, Storage and Analysis (SC13), Denver, 18-21 Nov. 2013. (SC 展示会場の JAIST ブースにて研究内容に関するポスター展示)
[B-30] 松原裕貴, 佐藤幸紀. メモリ階層の深化に対応するメモリアクセス解析ツール. 2014 年ハイパフォーマンスコンピューティングと計算科学シンポジウム(HPCS2014), 東京, 2014 年 1 月 7 日.
[B-31] Shimpei Sato, Akihiko Saijo, Yukinori Sato. Profiling B/F Ratios and Cache Behaviors within Loop and Call Nests in the Actual Program Execution. 2014 ATIP Workshop: Japanese Research toward Next-Generation Extreme Computing. Ernest N. Morial Convention Center. Nov. 17.
[B-32] Shimpei Sato, Yuki Matsubara, Akihiko Saijo, and Yukinori Sato, An Application Profiling Toolchain for Accelerating Systems with Deeper Memory Hierarchy. The 2014 International Conference for High Performance Computing, Networking, Storage, and Analysis (SC '14), 17-20 November 2014. (SC 展示会場の JAIST ブースにて研究内容に関するポスター展示)
[B-33] Shimpei Sato, Akihiko Saijo, Yukinori Sato. A Profiling Tool set for measuring B/F Ratios and Cache Behaviors from Actual Applications. JST/CREST International Symposium on Post Petascale System Software (ISP2S2), December 2014.
[B-34] 佐藤幸紀, 佐藤真平:メモリ階層性能シミュレータを用いた CPU 単体性能チューニング, ハイパフォーマンスコンピューティングと計算科学シンポジウム. (HPCS 2015) 論文集 2015, p. 100, 2015 年 5 月.
[B-35] 佐藤幸紀, 佐藤真平, 遠藤敏夫. CPU 性能チューニングを支援するアプリケーション解析ツール Exana のデモ, 萌芽的コンピュータシステム研究展示会(CEATEC2015 併設), 2015 年 10 月.
[B-36] Yukinori Sato and Toshio Endo. Consolidating memory locality information obtained from static and dynamic analysis of code for performance tuning in source code . 2nd Annual Meeting on Advanced Computing System and Infrastructure (ACSI2016), ポスターセッション, 福岡, 2016 年 1 月 19 日.

緑川グループ

[C-21] 古尾谷歩, 緑川博子:"リモートページングのためのページサイズ自動調整機構 ーループ文におけるワーキングセット推定と選択的制御の導入ー ", ハイパフォーマンスコンピューティングと計算科学シンポジウム HPCS2013, HPCS2013 論文集,(2013,1)
[C-22] 鈴木悠一郎, 岩井田匡俊, 緑川博子, 甲斐宗徳, "マルチノード・マルチスレッドプログラム向け並列ランタイムシステムの設計",SACSIS2013, pp.135-136, (2013,5)
[C-23] 岩井田匡俊, 鈴木悠一郎, 緑川博子:"InfiniBand を用いた遠隔メモリアクセスの性能",

SACSIS2013, pp.113-114, (2013,5)
[C-24] Hiroko Midorikawa, Yuichiro Suzuki, Masatoshi Iwaida, User-level Remote Memory Paging for Multithreaded Applications, Proceeding of 13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid2013), Delft, Netherlands, May 2013, pp.196-197, 2013-5 (DOI 10.1109/CCGrid.2013.63)
[C-25] H.Midorikawa, "DLM: Remote Memory Paging for Efficient Use of Memory Resource on Clusters", JST/CREST International Symposium on Post Peta-scale System Software , P16 Poster, Kobe Japan, (2014.12)
[C-26] Hiroko Midorikawa, Hideyuki Tan, Locality-Aware Stencil Computations using Flash SSDs as Main Memory Extension, Proceeding of IEEE/ACM International Symp. on Cluster, Cloud and the Grid Computing CCGrid2015, Shenzhen, China, pp.1163-1168, (DOI 10.1109/CCGrid.2015.126), (2015.5/6)
[C-27] 緑川博子,岩井田匡俊:"マルチスレッド対応型分散ン共有メモリシステムの設計と実装", ハイパフォーマンスコンピューティングと計算科学シンポジウム HPCS2015, 東大,(2015, 5/19)
[C-28] 丹英之,緑川博子:"非同期I/Oを用いたFlash SSDによるメインメモリ拡張のための性能調査", ハイパフォーマンスコンピューティングと計算科学シンポジウム HPCS2015, 東大, (2015,5/19)
[C-29] Hiroko Midorikawa: Locality-Aware Stencil Computations using Flash SSDs as Main Memory Extension, IEEE/ACM International Symp. on Cluster, Cloud and the Grid Computing CCGrid2015 (2015/5/3)
[C-30] Hiroko Midorikawa: Blk-Tune: Blocking Parameter Auto-Tuning for Flash-based Out-of-Core Stencil Computations, ACM The 25th International Symposium on High Performance Parallel and Distributed Computing HPDC'16 (2016/5/31)
[C-31] 大浦陽, 緑川博子: 遠隔メモリ利用による Out-of-Core マルチスレッドプログラム向け自動適応型ページサイズ可変機構の動作実験, ハイパフォーマンスコンピューティングと計算科学シンポジウム HPCS2016 (2016/6/6)
[C-32] 白澤卓磨, 緑川博子: マルチノードマルチコア向け分散共有メモリにおけるデータ分散配置機能稼働実験, ハイパフォーマンスコンピューティングと計算科学シンポジウム HPCS2016 (2016/6/6)
[C-33] Hiroko Midorikawa, Hideyuki Tan: A Highly Efficient I/O-based Out-of-Core Stencil Algorithm with Globally Optimized Temporal Blocking, IEEE Non-Volatile Memory Systems and Applications Symposium (NVMSA) (2017/8/16)

(4)知財出願

①国内出願 (0 件)

②海外出願 (0 件)

(5)受賞・報道等

①受賞

2014 年 7 月 16 日に東京で開催された GPU Technology Conference Japan 2014 において、本チームの成果を発表したポスター"Data Management and Loop Controlling to Surpass Memory Capacity of GPU in OpenACC Framework" により、金光浩(遠藤グループ研究員)および遠藤が NVIDIA Award を受賞。

＊2015 年 3 月 3 日の情報処理学会ハイパフォーマンスコンピューティング研究会(大分)において発表された本チームの成果論文「GPU 搭載システムにおける都市気流シミュレーションの

大規模化と性能モデル」により、高嵜 祐樹(遠藤グループメンバー、平成 27 年 3 月卒業)が、情報処理学会 2015 年度 コンピュータサイエンス領域奨励賞を受賞。

2016 年 9 月に台北で開催された国際会議 IEEE Cluster Computing において発表された本チームの成果に関する論文” Realizing Out-of-Core Stencil Computations using Multi-Tier Memory Hierarchy on GPGPU Clusters”が、best paper nominee (4 件)に選出。

2017 年 4 月に東京で開催された情報処理学会 cross-disciplinary Workshop on Computing Systems, Infrastructures, and Programming (xSIG 2017)における当チームの成果発表「メモリ階層の利用によって GPU メモリ容量を超える深層学習手法」により、伊藤祐貴(遠藤グループメンバー)が Outstanding B4 Student Award を受賞。

② マスコミ(新聞・TV等)報道(プレス発表をした場合にはその概要もお書き下さい。)
2017 年 2 月 17 日:東京工業大学プレスリリース「東工大のスパコン TSUBAME3.0 が今夏稼働開始―半精度演算性能 47.2 ペタフロップス、人工知能分野における需要急増へ対応―」において、TSUBAME3.0 スパコンの導入ベンダーおよびアーキテクチャ(メモリ・ストレージ階層を含む)の発表がなされた。

(6)成果展開事例

①実用化に向けての展開

- 遠藤グループ・緑川グループを中心とする研究により得られた、異種 DRAM やフラッシュメモリを含むメモリアーキテクチャに関する知見を、東京工業大学で平成 29 年 8 月に稼働開始した TSUBAME3.0 スパコンのアーキテクチャへフィードバックを行った。TSUBAME3.0 の各計算ノードには 2TB の容量と GB/s を超えるアクセス速度の NVMe Flash SSD を備え、通常のローカルスクラッチ領域としての利用に加え、当チームのランタイムライブラリ等により、ホストメモリの拡張として利用可能である。
- 遠藤グループで開発したランタイムライブラリ HHRT について github で公開中である。旧丸山直也チームの文脈で研究開発されたステンシルフレームワークとの統合を、東京大学下川辺隆司准教授らと協働で推進している。
- 遠藤グループで開発中の大規模深層学習ライブラリ ooc_cuDNN について、CREST プロジェクト「社会インフラ映像処理のための高速・省資源深層学習アルゴリズム基盤」(代表:篠田浩一)や、企業との共同研究における活用に向けて、議論を行っている。
- 佐藤を中心に開発したメモリプロファイラ Exana について、github で公開中であるとともに、理化学研究所や藤澤克樹チームなどのアプリケーションのメモリアクセス解析・最適化に活用されている。

②社会還元的な展開活動

・不揮発メモリの応用技術に関する国際会議 IEEE Non-Volatile Memory Systems and Applications Symposium (NVMSA2017)において、遠藤は Program co-chair を務め、不揮発メモリ研究分野への貢献を行った。次年度の NVMSA2018 ははじめて国内で開催されることとなり、遠藤は General co-chair に内定している。

- チームの研究成果について、Supercomputing(SC)国際会議・展示会において、JST-CREST ブース、東京工業大学ブース、JAIST ブースにおいて出展した。東京工業大学ブースにおいては、他の出展内容を含めて例年 300 名前後の来訪者があった。

§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 10 月 28 日・29 日	HPC ワークショップ金沢 2013	ホテル日航金沢	30 名	学術交流：CREST 遠藤チームの 3 グループがそれぞれ研究成果を発表
2014 年 9 月 17 日	メモリプラスワークショップ －メモリとファイルストレージと OS と－	JAIST 品川サテライトオフィス Room D-E	84 名	不揮発メモリを含むメモリ階層を中心とした研究分野における学術交流：メモリを核としてメモリデバイス、ファイル、システムソフトウェア、OS に関し、企業、大学などからの招待講演、本チームの研究紹介、研究パネル、参加者の事前アンケート、技術討論を行った
2015 年 2 月 25 日	第29回 J-BEANS セミナー（JAIST と公益財団法人北陸先端科学技術大学院大学支援財団が共催し地域にも開放するセミナー）	JAIST	30 名	佐藤が「スパコンの性能向上への挑戦とダイナミックコンパイレーション技術の研究開発」という講演を行い CREST の成果を地域に向けて広報
2016 年 8 月 31 日	第 2 回　メモリプラスワークショップ	東京工業大学 キャンパスイノベーションセンター(田町)	50 名	最新メモリに関わる招待講演3件と CREST グループ成果発表3件を含むワークショップ

ほか、チーム内遠隔ミーティング(非公開)を、月に 2 回のペースで開催している。

§６ 最後に

研究期間を通して、メモリ階層活用システムソフトウェア・ツールチェーンに関する研究開発を遂行した。特に、実際のステンシルベースの実アプリケーションにおいて技術統合を行い、高位メモリ階層の高性能と低位メモリ階層の大容量を活用するという当初からの目標を実現できた。本稿執筆時点では、TSUBAME3.0 スーパーコンピュータの導入時期の関係で大規模実証実験を行えていないものの、1PB/s に迫る速度性能と 1PB に近い計算規模をめざして最終的な実証実験を行う。

当チームの構成メンバーは当初より、システムソフトウェア分野の研究者が多かった。アプリ分野を含めたコデザインのために、藤澤克樹チーム、丸山直也チームなどとの連携を行い、共著論文を多数発刊するなど、連携は成功したと考えている。

チーム内の研究項目については、それぞれ成果発表を行い、多くのケースでソフトウェアの公開までこぎつけた。一方でさらなる技術統合を行いたい面がある。たとえば Polyhedral コンパイラを中心としたキャッシュ最適化と、DRAM よりも低位メモリの階層活用の統合などである。この点などについては、残り研究期間で推進しつつ、当プロジェクトの後継となるプロジェクトの立ち上げを計画し

ている。

ポストペタ・エクサスケール時代に向けて、国内外の大規模システムの計画に注意を払う必要は引き続きあるものの、メモリ階層の効率的な利用を通して 10PB/s 級の性能と 10PB 級の計算規模の両立実現に向けての道筋は整備されたと考えている。海外ではたとえば Hewlett-Packard 社により巨大なメモリプールを光ネットワークなどの超高帯幅で結合する The Machine プロジェクトが進んでいる。また不揮発メモリの活用技術については米国・中国などが先行しており、国内ではデバイスそのものの研究が多かった傾向にあったが、システムレベル・アプリレベルでの研究も増えつつあう。以上のような動向を踏まえつつ、超大規模・高性能なスーパーコンピューティング・ビッグデータ処理を低プログラミングコストで実現するための技術開発をさらに発展させたい。