

滝沢 寛之

東北大学大学院情報科学研究科
准教授

進化的アプローチによる超並列複合システム向け開発環境の創出

§ 1. 研究実施体制

(1) 滝沢グループ

- ① 研究代表者: 滝沢 寛之 (東北大学大学院情報科学研究科、准教授)
- ② 研究項目
 - ・システムの複合性を扱うためのプログラミングインタフェースの設計
 - ・HPC アプリケーションのためのリファクタリングツールの開発

(2) 高橋グループ

- ① 主たる共同研究者: 高橋 大介 (筑波大学システム情報系、教授)
- ② 研究項目
 - ・超並列複合システムに向けた高速フーリエ変換 (FFT) ライブラリおよび代数的多重格子法ライブラリの開発
 - ・混合精度計算に対応した基本線形計算ライブラリの基本設計

(3) 須田グループ

- ① 主たる共同研究者: 須田 礼仁 (東京大学情報理工学系研究科、教授)
- ② 研究項目
 - ・ドメイン特化型ツールによる超並列複合システムの階層的抽象化に関する研究

(4) 江川グループ

- ① 主たる共同研究者: 江川 隆輔 (東北大学サイバーサイエンスセンター、准教授)
- ② 研究項目

- ・HPC のためのリファクタリングカタログの設計
- ・超並列複合システム向けのアプリケーション開発手法の確立

§ 2. 研究実施の概要

現在、多くの高性能計算アプリケーションは特定のシステムを想定し、その想定システム上で高性能を達成できるように開発・最適化されている。このため、そのアプリケーションを他のシステム上で実行する場合には、高い性能を達成できない恐れがある。すなわち、性能可搬性が低いことが問題となっており、既存のアプリケーション資産を新システムに移行する場合にも同様の問題が発生する。このため本研究では、既存のアプリケーション資産の性能可搬性を高めることで、その資産の将来のシステムへの円滑な移行を実現する。そのためには性能可搬性の阻害要因を大規模なアプリケーションコードから分離する必要がある、書き直しの困難な規模のアプリケーション資産を対象に考える場合には、コードを段階的に再構成することで分離を実現する必要がある。この分離作業を、本研究では HPC リファクタリングと呼ぶ。

本研究の大きな目標の一つは、事例に基づいて性能可搬性の阻害要因を体系的にまとめ、それらを切り離す HPC リファクタリングの方法論をまとめることである。平成 25 年度には、多様なアプリケーションの性能を評価し、性能可搬性を低下させる要因になっているコードパターンの事例集をまとめてきた。先行して集計したベクトルシステム向けの事例に加えて、スカラシステム向けの事例も集計しており、事例集の公開に向けて準備を進めている。

また、HPC リファクタリングによってアプリケーションコードから切り離された情報は、特定のシステムで性能を出すための情報と言い換えることができ、単に切り離すだけでは性能が低下する。このため、各システムで性能を出すための情報をアプリケーションコードとは別の存在として抽象化し、システムに合わせて使い分けることで、それらのシステム間の性能可搬性を実現することを考えている。この抽象化のための最も一般的な方法は、汎用性の高い計算を数値ライブラリとして実装することである。平成 25 年度には、単体 GPU および GPU クラスタをターゲットとして、これらの数値計算ライブラリにおけるアルゴリズムおよび最適化手法の検討を行うと共に実装を行い、単体 GPU 向けの疎行列ベクトル積ライブラリを公開した。また、通信削減アルゴリズムや集団通信の通信回数削減など、ポストペタ時代に必要となる高性能並列化の要素技術を検討して評価を行った。

上記のように抽象化する仕組みを用意しても、全てを抽象化することはできないため、アプリケーションコードの修正が求められる。そのようなコード修正は一般化が難しく、特定のシステムやアプリケーション向けの修正となる傾向にある。そのようなコード変換も何らかの形でアプリケーションコードから切り離し、再適用できる仕組みがなければ、HPC リファクタリングされたコードで元のコードと同じ性能を出すことはできない。そこで本研究では平成 25 年度に、そのようなシステム特有、アプリケーション特有のコード修正を、ユーザー定義変換として定義するための仕組みを構築してきた。コンパイラの内部表現として用いられる抽象構文木(AST)を外部に公開し、それに対する変換規則を定義することで、ユーザー定義のコード変換を実現することを考えた。一般的に木構造データ

を表現するために広く用いられている道具の一つとして、XML がある。この XML を使えば、AST をテキスト形式で表現することが可能であり、すでに確立されている XML 関連の様々な技術を使うことができるため、XML に基づくユーザー定義変換の基盤を試作した。また、その使用例を通じて本アプローチの可能性と限界を探った。

§ 3. 成果発表等

(3-1) 原著論文発表

論文詳細情報(国内)

- A-1 伊野文彦, 萩原兼一, “GPU アクセラレータとその研究動向”, *Medical Imaging Technology*, Vol. 31, No. 3, pp. 147-152, 2013. (DOI: 10.11409/mit.31.147)
- A-2 平澤将一, 滝沢寛之, 小林広明, “ソフトウェア進化のための自動性能追跡システム”, *情報処理学会論文誌コンピューティングシステム(ACS)*, Vol. 6 No. 4, pp. 96-104, 2013. (NAID: 110009616697)

論文詳細情報(国際)

- A-3 Fumihiko Ino, Kentaro Shigeoka, Tomohiro Okuyama, Masaya Motokubota, and Kenichi Hagihara, “A Parallel Scheme for Accelerating Parameter Sweep Applications on a GPU,” *Concurrency and Computation: Practice and Experience*, Vol. 26, No. 2, pp. 516-531, 2014. (DOI: 10.1002/cpe.3016)
- A-4 Hiroyuki Takizawa, Makoto Sugawara, Shoichi Hirasawa, Isaac Gelado, Hiroaki Kobayashi, and Wen-mei W. Hwu, “clMPI: An OpenCL Extension for Interoperation with the Message Passing Interface,” *the IEEE 27th International Symposium on Parallel & Distributed Processing Workshops (IPDPSW2013)*, pp. 1138-1148, 2013. (DOI: 10.1109/IPDPSW.2013.183)
- A-5 Makoto Sugawara, Shoichi Hirasawa, Kazuhiko Komatsu, Hiroyuki Takizawa and Hiroaki Kobayashi, “A Comparison of Performance Tunabilities between OpenCL and OpenACC,” *the IEEE 7th International Symposium on Embedded Multicore SoCs (MCSoc-13)*, pp. 147-152, 2013. (DOI: 10.1109/MCSoc.2013.31)
- A-6 Fumihiko Ino, Shinta Nakagawa, and Kenichi Hagihara, “GPU-Chariot: A Programming Framework for Stream Applications Running on Multi-GPU Systems,” *IEICE Transactions on Information and Systems*, Vol. 96-D, No. 12, pp. 2604-2613, 2013. (DOI: 10.1587/transinf.E96.D.2604)
- B-1 Daichi Mukunoki and Daisuke Takahashi, “Optimization of Sparse Matrix-vector Multiplication for CRS Format on NVIDIA Kepler Architecture GPUs,” *Proc. 13th International Conference on Computational Science and Its Applications (ICCSA)*

- 2013), Part V, Lecture Notes in Computer Science, Vol. 7975, pp. 211-223, 2013. (DOI: 10.1007/978-3-642-39640-3_15)
- B-2 Daisuke Takahashi, "Implementation of Parallel 1-D FFT on GPU Clusters," Proc. 2013 IEEE 16th International Conference on Computational Science and Engineering (CSE 2013), pp. 174-180, 2013. (DOI: 10.1109/CSE.2013.36)
- B-3 Takaaki Hiragushi and Daisuke Takahashi, "Efficient Hybrid Breadth-First Search on GPUs," LNCS Algorithms and Architectures for Parallel Processing (ICA3PP 2013), Vol. 8286, pp. 40-50, 2013. (DOI: 10.1007/978-3-319-03889-6_5)
- C-1 Ayumu Tomiyama, Reiji Suda, "Automatic Parameter Optimization for Edit Distance Algorithm on GPU," LNCS High Performance Computing for Computational Science - VECPAR 2012, Vol. 7851, pp. 420-434, 2013. (DOI: 10.1007/978-3-642-38718-0_38)
- C-2 Kamil Rocki, Reiji Suda, "High Performance GPU Accelerated Local Optimization in TSP," Third Workshop on Parallel Computing and Optimization (PCO'13) in conjunction with 27th IEEE International Parallel & Distributed Processing Symposium (IPDPS), pp. 1788-1796, 2013. (DOI: 10.1109/IPDPSW.2013.227)
- C-3 Cheng Luo and Reiji Suda, "An Efficient Task Partitioning and Scheduling Method for Symmetric Multiple GPU Architecture," the 11th International Symposium on Parallel and Distributed Processing with Applications (ISPA2013), pp. 1133-1142, 2013. (DOI: 10.1109/TrustCom.2013.137)
- C-4 Tian Xiaochen, Kamil Rocki, Reiji Suda, "Register Level Sort Algorithm on Multi-Core SIMD Processors," IA³ Workshop on Irregular Applications: Architectures & Algorithms, The International Conference for High-Performance Computing, Networking, Storage, and Analysis (SC13), pp. 9:1-9:8, 2013. (DOI: 10.1145/2535753.2535762)
- C-5 Kamil Rocki, Martin Burtscher, and Reiji Suda, "The Future of Accelerator Programming: Abstraction, Performance or Can We Have Both?," the 29th Symposium on Applied Computing, Gyeongju (SAC 2014), 2014. (to appear)