

課題名(タイトル):

大規模・高並列・高性能な数値計算ソフトウェアライブラリの研究開発

利用者氏名:

○今村 俊幸(1), 工藤 周平(1), 廣田 悠輔(1), 椋木 大地(1), Yiyu Tan(1), Doru Thom Popovich(1), Franz Frachetti(1)

理研における所属研究室名:

(1) 計算科学研究センター 大規模並列数値計算技術研究チーム

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

本課題ではポスト京コンピュータおよびエクサスケールコンピュータに向けた、大規模・高並列・高性能な数値計算ソフトウェアライブラリの実現に向けた研究開発を行うことを目的とする。計算科学で用いられるアプリケーションの多くは、連立一次方程式や固有値計算、高速フーリエ変換、非線形方程式などの様々な数値計算ソフトウェアライブラリの上に構築されている。申請チームで業務として推進している京コンピュータでの高度利用化プロジェクトと比べてさらなる大規模化・高並列化が次世代のスーパーコンピュータにおいて進むことが明確になっている。これらのソフトウェアライブラリが十分な性能を発揮できるようにするためには、アルゴリズムや実装方法の見直しが不可欠である。本研究では京コンピュータをベースとする FX10 の後継機種である FX100 と GPU クラスタ、さらに Intel の最新 CPU である SkyLake-SP を活用して、次世代のスーパーコンピュータに向けた数値計算ソフトウェアライブラリに必要となるアルゴリズムの選定・評価、プログラムの実装技術に関する基礎研究を行う

2. 具体的な利用内容、計算方法

昨年度からの継続として、研究チームが開発する高性能固有値ソルバ EigenExa や関連ソフトウェア Kevd を FX100 が具備する HPC-ACE2 ならびに Intel AVX512 による SIMD 化を行い、ポスト京コンピュータで採用される Arm8-SVE に向けた移植準備と知見集積を行う。

また、今年度から新たに、小規模な線形カーネルの高速化のために、BatchedBLAS インターフェイスへの対応をはじめ、ベンダーBLAS とスレッド並列化との相性や実際の BatchedBLAS ライブラリの性能評価を実施する。

ベンダー提供の BLAS カーネルで性能不足になるような場合に対処する、BLAS+X 構想の一例として FX100 と SkyLake-SP を対象とした小規模な三重対角化ルーチンの実測を行う。

既存開発ソフトウェアの互換性や移植性の評価についても、SkyLake-SP を中心に実施した。

3. 結果

HPC-ACE2 や AVX2, AVX-512 での SIMD 化手順を踏襲して、ARM-SVE 化への移植方法の一部をまとめることができた。

小規模カーネルの実装と高速化については、Batched BLAS と BLAS+X の考え方のもと生産性を考慮した開発方針などを見出すに至った。本テーマと本実験については ISC18 ならびに HPCAsia2019 での対外発表を実施済み。

チーム開発物の新規プラットフォームの移植は SkyLake-SP への EigenExa 対応、ならびに CUDA9 以降の環境対する ASPEN.K2 の移植に成功している。

4. まとめ

本課題では一昨年来からの FX100 での大規模並列数値ソフトウェアの移植・性能評価作業に引き続き、小規模カーネルに特化した線形計算カーネルの研究を開始し一定の成果を得た。小規模ならびに大規模数値計算ソフトウェアの成果は今後、ポスト「京」開発プロジェクトと連携し開発に応用されていく。

5. 今後の計画・展望

SkyLake-SP での詳細な性能評価ならびに、継続的なソフトウェアの移植・保守を進める。また、来年度中の「京」コンピュータ停止に伴う、開発環境等の Hokusai システムへの移行を速やかに進め、小規模カーネルの高性能化から深層学習やデータ分析分野への応用を開始する。また、ポスト「京」に向けて大規模並列数値ソフトウェアの新規開発も継続して実施する。

平成 30 年度 利用研究成果リスト

【会議の予稿集】

[1] 工藤 周平, 今村 俊幸: 三重対角化におけるメニーコア環境に適した同期手法: 研究報告ハイパフォーマンスコンピューティング(HPC), Vol. 2018-HPC-165, No. 33, Page 1 - 8, 2018 年 07 月 23 日 (第 165 回 HPC 研究発表会 (SWoPP2018), 熊本市国際交流会館, 2018 年 8 月 1 日)

[2] Shuhei Kudo, Toshiyuki Imamura: Cache-efficient implementation and batching of tridiagonalization on manycore CPUs: Proceedings of the International Conference on High Performance Computing in Asia-Pacific Region (HPC Asia 2019), Pages 71-80, 2019

[3] Yiyu Tan¹, Toshiyuki Imamura¹: Performance Evaluation of a Toolkit for Sparse Tensor Decomposition, Proceedings of the 27th International Symposium on High-Performance Parallel and Distributed Computing (HPDC18), Pages 5-6, 2018

【ポスター発表】

[1] Yusuke Hirota, Daichi Mukunoki, and Toshiyuki Imamura: Automatic Generation of Full-Set Batched BLAS. Research Poster, International Supercomputing Conference (ISC'18), 2018 年 6 月 26 日, ISC HPC