

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(1,2), 伊奈拓也(1,2), 工藤 周平(1,2), 廣田 悠輔(1,2), 椋木 大地(1,2), Yiyu Tan(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
- (2) 計算科学研究センター FS2020 開発プロジェクト アーキテクチャ開発チーム

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(以下「富岳」)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 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な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

一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 について、FX100 を富岳の評価環境と捉え、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。

小規模な線形カーネルの高速化のために、BatchedBLAS インターフェイスをはじめとしたバッチインターフェイス環境の実践プラットフォームとして、ベンダーBLAS とスレッド並列化との相性や実際の BatchedBLAS ライブラリの動作確認を実施する。

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行列向け固有値カーネルの実測を行う。

さらに、既存開発ソフトウェアの互換性や移植性の確認作業のために、K20m ならびに Skylake-SP を利用する。

3. 結果

1) バッチ型数値計算カーネルの開発

N 回の FFT 計算で生じる集団通信を一度にまとめてバッチ実行することで通信回数を減らし大規模並列計算時の同期コストを削減することを目指した。しかし、HOKUSAI (BWMPC) 32 ノード利用して集団通信の同期コストの低減効果を検証したが並列数や非同期通信の動作が期待したものではなく、当初想定した効果は確認できなかった

2) 小規模カーネルの実装と動作確認

ヤコビ回転カーネルの性能改善版の評価環境として HOKUSAI Big Waterfall システムを利用した。ヤコビ型固有値計算、特異値計算どちらについても、FX100, Skylake-SP の両 CPU で高い計算性能を示すことが確認された。

3) 既開発ソフトウェアの互換性確認

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 ライブラリの移植確認を実施した。

4. まとめ

本課題では従来から進めてきた富岳向けの評価機としての FX100 を利用している。大規模並列数学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とともに、バッチ型の新しい実行形態についての成果も得られている。

5. 今後の計画・展望

2019年度 利用報告書

小規模カーネルの真の需要は深層学習やデータ分析分野にあるため、当該分野への応用を開始する。また、「富岳」の本格稼働前の準備期間中に、重要度が高い「富岳」向け大規模並列数学ソフトウェアを継続して移植・開発する。

6. 利用がなかった場合の理由

2019 年度 利用研究成果リスト

【会議の予稿集】

[1] Toshiyuki Imamura, Masaaki Aoki, Mitsuo Yokokawa: “Batched 3D-distributed FFT kernels towards practical DNS codes”, International Conference on Parallel Computing, Prague, Czech, September 2019、Proc. of ParCo2019 (to appear)

【口頭発表】

[2] S. Kudo, and T. Imamura: “A level-3 BLAS like kernel of the Jacobi rotations for the Jacobi’s eigenvalue algorithms,” ParNum2019, Dubrovnik, Oct. 2019.