

令和 6 年 6 月 10 日現在

機関番号：32613

研究種目：基盤研究(C) (一般)

研究期間：2018～2023

課題番号：18K11340

研究課題名(和文)大規模並列計算機用高精度演算環境の構築と演算精度自動チューニングの研究

研究課題名(英文) Research on construction of high-precision computing environment for massively parallel computers and automatic tuning of computing accuracy

研究代表者

田中 輝雄 (Tanaka, Teruo)

工学院大学・情報学部(情報工学部)・教授

研究者番号：90622837

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：大規模計算機環境で必要とされる高精度算術演算の研究である。(1)現在開発中の2つのツールの拡張を行った。倍精度演算ライブラリDD-AVXにAVX-512の基本演算機能を追加した。また、GMPコード生成ツールにMPIサポートを追加した。(2)条件数が多いなど収束性の悪い問題に対して、高精度並列計算アルゴリズムに切り替えた場合の挙動を解析した。(3)GPUの強力な単精度計算能力を引き出すため、混合精度計算への拡張を行った。特に、ディレクティブベースの記述に本手法を適用してみた結果、ある程度の性能を引き出すことがわかったが、設定パラメータが規定値ではないため、最適化が必要であることがわかった。

研究成果の学術的意義や社会的意義

より大規模かつ複雑な問題に対処するために高精度演算が重要となる。しかしながら、計算機の専門外の研究者が高精度演算を使いこなすためには、専門知識、あるいは、研究に関係ないコストが必要となる。そのために、前者に対しては、高精度計算の精度に関する調査を行った。また、後者については、倍々精度演算ライブラリDD-AVX、GMP生成ツールXev-GMPの機能拡張を行い、さらに、ディレクティブベース記述でのGPU利用上でのチューニングなどを行ってきた。これらにより、情報学以外の一般工学、社会学などの広い分野の研究者が余分なコストを抑え、高精度演算を利用可能となる道筋になると考えている。

研究成果の概要(英文)：This is a study of high-precision arithmetic operations required in a large-scale computing environment. (1) Extensions were made to two tools currently under development. The basic arithmetic functions of AVX-512 were added to the double-precision arithmetic library DD-AVX. MPI support was also added to the GMP code generation tool. (2) We analyzed the behavior of high-precision parallel computation algorithms when switching to high-precision computation for problems with poor convergence, such as a large number of conditions. (3) In order to bring out the powerful single-precision computing capability of GPUs, we extended the scope to mixed-precision computation. In particular, we tried to use the method in directive-based descriptions, and the results showed that it can bring out some performance, but the setting parameters are not specified values and need to be optimized.

研究分野：高性能計算

キーワード：高精度計算 高性能計算 倍々精度演算 大規模並列処理 自動チューニング 混合精度計算

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様式 C-19、F-19-1 (共通)

1. 研究開始当初の背景

物理・化学分野などの自然科学の分野、あるいは、社会科学の分野全般において、対象が大規模化、複雑化することにより、それに応じてシミュレーションモデルも大規模化・複雑化している。そのため、従来から扱われている単精度、倍精度データでは、現象を正確に表すことができない問題が生じている。そのため、2つの倍精度データを組み合わせた倍々精度データ、あるいは、それ以上の高精度データが使われだしている。しかしながら、それらに対して、コンピュータ上ではソフトウェアでの対応が必要であり、一般に使いづらく性能を引き出すことが難しい。そこで、一般（つまり、情報学以外を専門とする）研究者、技術者でも利用できるようにツール化を行い、さらに、それらのツールの高性能化を図っていく必要がある。この課題に対して、我々はこれまでに、倍々精度演算ライブラリDD-AVX、GMP生成ツールXev-GMPを開発してきた[1][2]。

2. 研究の目的

本研究では、大規模並列計算機用高精度演算環境の構築として、①倍々精度演算ライブラリDD-AVXのAVX-512への対応、②GMP生成ツールXev-GMPのMPI並列記述変換への対応、③多倍長データ通信処理の改善を行う。さらに、演算精度自動チューニングを実現するための基礎的な研究として、④多倍長計算を用いた疎行列向け反復処理アルゴリズムの開発を行い、⑤自動チューニングへの適用を試みる。

3. 研究の方法

(a) 高並列計算機用高精度演算環境の構築

- ①倍々精度演算ライブラリのAVX-512への対応：SSE/AVX系のアクセラレータはSIMD演算処理機構であるAVX-512に統一されつつあり、その有効利用をはかる。開発はAVX2に対応した倍々精度演算ライブラリDD-AVXをベースに推進する。
- ②GMP向けMPIコード自動生成：現在、GMP生成ツールXev-GMPにおいてMPI対応を進めており、基本的な動作の確認は済んでいる。さらに、構造体への対応、データ（変数）ごとの精度の指定などの課題について、開発を行う。
- ③大規模並列計算機用高精度演算環境での通信性能の改善：大規模並列計算機用高精度演算環境においては、通信性能が全体の性能に大きく関わる。通信量削減方式としては、Matrix Power Kernel技術やアルゴリズムの修正などによる通信データ数の削減やMPI通信に代わるRemote DMAやremote operationなどの通信プロトコルの軽量化の研究を行う。通信量削減方式については、転送データ量で通信プロトコルを自動的に切り替える方式を検討する。

(b) 演算精度自動チューニングの研究

- ④高精度並列計算アルゴリズムの開発：上記(a)①～③にて構築した高並列計算機用高精度演算環境上で、多倍長データを用いた高精度演算による(疎)行列を用いた反復計算アルゴリズムを開発する。本研究では、開発する混合精度演算環境を用いて、種々の条件の悪い疎行列問題について混合精度演算の特性評価とアルゴリズムの開発を行う。
- ⑤自動チューニング技術の適用：高並列計算機用高精度演算環境において、対象とする大規模問題ごとにデータの精度によりアルゴリズムの挙動が大きく異なる。計算量（実用的な収束性の確保）やユーザの要求する解の精度などにより要求される演算精度自体を評価パラメタとして、自動チューニング技術の適用を研究する。そのために、高精度計算環境の構築において、データ（変数）ごとに精度設定が記述できるようにすることが重要となる。

以上により、最終的には、一般ユーザ(研究者、技術者)に対して、使い勝手の良い多倍長反復計算ライブラリの提供を目指す。

なお、本研究のプログラムの開発の一部は、研究メンバそれぞれの大学の研究室の学生の協力も得て行なう。研究で得られた成果は、順次、開発したソフトウェアを公開する形で、広く利用促進をはかる。また、適宜、国内、国際学会での発表を行い、論文としてまとめる。

4. 研究成果

本研究の課題について、それぞれ研究実績を項目ごとにまとめる。

(a) 高並列計算機用高精度演算環境の構築：

①倍々精度演算ライブラリのAVX-512への対応：

- ・ Xeon Phi上のAVX-512上に我々が開発した倍々精度演算ライブラリDD-AVXを実装し、基本演算機能を実現した。
- ・ Xeon Phi上に実装した倍々精度演算ライブラリDDAVXに対して、C++を用いて複数の混合精度での対応を共通関数化により、簡素化することを試みた。なお、このライブラリは、Intelの上位マルチコアCPUにも適用可能である。
- ・ Intel Xeon 56コアの環境で評価を行い、その特性を明らかにした。

②GMP向けMPIコード自動生成に関して：

- ・ GMP生成ツールXev-GMPにMPI対応を実現した。

③大規模並列計算機用高精度演算環境における通信量削減方式の実現に関して

- ・ 倍々精度に対応する前に通常精度にてまずは検討を進めた。
- ・ RDMA通信およびセマフォによる同期処理の削減によるデータ通信手法を実現した。
- ・ 実現したRDMA通信およびセマフォによる同期処理の削減によるデータ通信機能に対して、名古屋大学のFX100を用いて評価を行ない、その有効性を確認した。

(b) 演算精度自動チューニングの研究

④高精度並列計算アルゴリズムの開発：

- ・ 条件数が大きいなどの収束性が悪い問題に対し、高精度計算の精度を切り替えたときの挙動について分析を行なった。
- ・ これまでは疎行列-ベクトル積(SpMV)を中心に研究を進めてきたが、より計算量の大きな行列-行列積(GEMM)に範囲を広げ、高精度演算での特にキャッシュの効果について分析した。
- ・ ベクトルデータの格納方式であるAoSとSoAについて比較分析を行い、従来、SoAのほうがすぐれていると考えられていたが、キャッシュに入りきらないサイズでは性能に差がなく、使い勝手の良いAoSでも実用的であることを示した。
- ・ GPUを用いた計算において、従来から主流であったCUDAでの記述方法に対し、近年、ディレクティブベースの記述方法が実現されており、性能面から評価を行った。ディレクティブベースでも性能を引き出すことができていること、ただし、既定値で設定されているパラメタでは最適化が十分でなく、パラメタの最適化が必要であることがわかった。
- ・ GPUを用いた計算において、従来からCUDAでの記述が主流であったが、近年、ディレクティブベースの記述方法が実現されており、性能面から評価を行った。その結果、複数の計算機環境での性能特性が異なることを示した。

- ・近年、特にGPUの高速な低精度演算の活用により注目されている混合精度演算について、スーパーコンピュータ向けベンチマークプログラムであるHPL-MxPを題材にして検討を行った。まず、計算量が n^3 オーダーとなり、処理時間の大半を占める単精度LU分解について、ディレクティブベースの記述方法であるOpenACCを用いてGPUの能力をどこまで引き出せるかの検討を行なった。結果として、単にOpenACCの記述適用だけでは性能を引き出すことはできず、基本的な高速化手法により、CPUでの並列実行よりも高い性能を実現できた。

⑤自動チューニング技術の適用について

- ・高精度演算への適用の前に、まずは高精度演算にて必要になると考える複数性能パラメタの同時最適組合せ推定を実現した。
- ・高精度演算への適用するために、開発ツールの実用レベルへの適用化を進め、複数性能パラメタの同時最適組合せ推定に対して、実環境での実行時間などの評価パラメタの実測のブレにも柔軟に対応できるように改良を行った。また、機械学習など、今後、半精度を含め、混合精度演算が重要となると考える分野に対する適用を試み、開発した手法の有効性を示すことができた。
- ・高精度演算への適用するために、まずは、機械学習のプログラムの並列環境での実装を行い、名古屋大学の最新スーパーコンピュータ不老Type-IIの高GPU環境を用いて、並列化により逐次処理に比べて約23倍の高速化を実現した。
- ・複数個所の同時性能パラメタチューニング手法を提案した。また、機械学習など、半精度を含めた混合精度演算が重要となる分野に対する適用を試みた。

[1] 菱沼利彰, 藤井昭宏, 田中 輝雄, 長谷川秀彦, AVX2を用いた倍精度BCRS形式疎行列と倍々精度ベクトル積の高速化, 情報処理学会論文誌 コンピューティングシステム, 48, 2014年9月.

[2] Toshiaki Hishinuma, Takuma Sakakibara, Akihiro Fujii, Teruo Tanaka, Shiichi Hirasawa, Xev-GMP: Automatic code generation for GMP multiple-precision code from C code, 19th IEEE International Conference on Computational Science and Engineering (CSE 2016), 2016.8.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計33件（うち招待講演 1件 / うち国際学会 15件）

1. 発表者名 Teruo Tanaka
2. 発表標題 Acceleration Techniques for Software Auto-Tuning to Hyperparameters on Machine Learning Software
3. 学会等名 2024 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing (ATAT on HPSC) (招待講演) (国際学会)
4. 発表年 2024年

1. 発表者名 高山智礼, 久木田仁, 矢島雄河, 藤井昭宏, 田中輝雄
2. 発表標題 OpenACCを用いた単精度LU 分解のGPU 並列化
3. 学会等名 情報処理学会全国大会
4. 発表年 2024年

1. 発表者名 Yuga Yajima, Akihiro Fujii, Teruo Tanaka
2. 発表標題 Job level parallel search in software auto-tuning
3. 学会等名 HPCAsia2024 (Nagoya) Poster Session (2024) (国際学会)
4. 発表年 2024年

1. 発表者名 Kota Yoshimoto, Akihiro Fujii, Teruo Tanaka
2. 発表標題 Fast adjacent communication among with RDMA, MPI-RMA, and Double buffering
3. 学会等名 HPCAsia2024 (Nagoya) Poster Session (2024) (国際学会)
4. 発表年 2024年

1. 発表者名 田中輝雄
2. 発表標題 機械学習ソフトウェアへのソフトウェア自動チューニングの適用
3. 学会等名 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 (ATOS30)
4. 発表年 2023年

1. 発表者名 Sorataro Fujika, Yuga Yajima, Teruo Tanaka, Akihiro Fujii, Yuka Kato, Satoshi Ohshima, Takahiro Katagiri
2. 発表標題 Parallelization of Automatic Tuning for Hyperparameter Optimization of Pedestrian Route Prediction Applications using Machine Learning
3. 学会等名 HPC Asia '23: Proceedings of the International Conference on High Performance Computing in Asia-Pacific Region (国際学会)
4. 発表年 2023年

1. 発表者名 Xuan Yang, Sorataro Fujika, Yuga Yajima, Akihiro Fujii, Teruo Tanaka, Kazutoshi Akita, Norimichi Ukita
2. 発表標題 Applying Automatic Tuning to Hyper-parameter Optimization of Machine Learning Programs for Super-Resolution (Poster)
3. 学会等名 HPC Asia '23: Proceedings of the International Conference on High Performance Computing in Asia-Pacific Region (国際学会)
4. 発表年 2023年

1. 発表者名 寺田洋人, 慈道亮人, 菱沼利彰, 藤井昭宏, 田中輝雄
2. 発表標題 複数のGPU向けプログラミングモデルを用いた 倍々精度疎行列ベクトル積の特性分析
3. 学会等名 第85回情報処理学会全国大会
4. 発表年 2023年

1. 発表者名 慈道亮人, 寺田洋人, 菱沼利彰, 藤井昭宏, 田中輝雄
2. 発表標題 倍々精度SpMVにおけるSIMD利用時のデータレイアウトによる性能分析
3. 学会等名 第85回情報処理学会全国大会
4. 発表年 2023年

1. 発表者名 吉本 航太、藤井 昭宏、田中 輝雄
2. 発表標題 FX1000におけるダブルバッファリングとRDMAを用いた隣接通信の性能評価
3. 学会等名 第26回 計算工学講演会
4. 発表年 2022年

1. 発表者名 福永晋司, 山浦朴人, 菱沼利彰, 藤井昭宏, 田中輝雄
2. 発表標題 倍々精度演算におけるSIMD命令利用時のデータレイアウトによる性能差分析
3. 学会等名 情報処理学会第84回全国大会
4. 発表年 2022年

1. 発表者名 山浦朴人, 福永晋司, 菱沼利彰, 藤井昭宏, 田中輝雄
2. 発表標題 ディレクティブベースのGPUプログラミングモデルを用いた倍々精度演算の性能評価
3. 学会等名 情報処理学会第84回全国大会
4. 発表年 2022年

1. 発表者名 森戸建太郎, 矢島雄河, 藤家空太郎, 楊 暄, 藤井昭宏, 田中輝雄
2. 発表標題 ソフトウェア自動チューニングツール DSICE の構築
3. 学会等名 情報処理学会第 8 4 回全国大会
4. 発表年 2022年

1. 発表者名 矢島雄河, 森戸建太郎, 藤家空太郎, 楊 暄, 藤井昭宏, 田中輝雄
2. 発表標題 ソフトウェア自動チューニングにおける反復1次元d-spline探索法の複数カ所同時適用
3. 学会等名 情報処理学会第 8 4 回全国大会
4. 発表年 2022年

1. 発表者名 Kota Yoshimoto, Akihiro Fujii, Teruo Tanaka
2. 発表標題 RDMA with Double Buffering for Adjacent Communication
3. 学会等名 The International Conference on High Performance Computing in Asia-Pacific Region (HPC Asia 2022), poster (国際学会)
4. 発表年 2022年

1. 発表者名 Sorataro Fujika, Teruo Tanaka, Akihiro Fujii, Yuka Kato, Satoshi Ohshima, Takahiro Katagiri
2. 発表標題 Parallelization of Automatic Tuning by Executing Machine Learning Programs in Multiple Jobs
3. 学会等名 The International Conference on High Performance Computing in Asia-Pacific Region (HPC Asia 2022), poster (国際学会)
4. 発表年 2022年

1. 発表者名 Kota Yoshimoto, Akihiro Fujii, Teruo Tanaka
2. 発表標題 Fast Adjacent Communication with RDMA and Double Buffering
3. 学会等名 The 20th International Symposium on Advanced Technology (ISAT-20), Lightning Talk session (国際学会)
4. 発表年 2021年

1. 発表者名 Sorataro Fujika, Teruo Tanaka, Akihiro Fujii, Yuka Kato, Satoshi Ohshima, Takahiro Katagiri
2. 発表標題 Parallelization of Automatic Tuning using Multiple Jobs for Machine Learning Programs
3. 学会等名 The 20th International Symposium on Advanced Technology (ISAT-20), Lightning Talk session (国際学会)
4. 発表年 2021年

1. 発表者名 藤家空太郎, 多部田敏樹, 藤井昭宏, 田中輝雄, 加藤由花, 大島聡史, 片桐孝洋
2. 発表標題 GPUクラスタを用いて並列化した自動チューニングの機械学習プログラムへの適用と安定性の検証
3. 学会等名 情報処理学会, HPC研究会論文集, Vol.2020-HPC-178
4. 発表年 2021年

1. 発表者名 多部田敏樹, 藤家空太郎, 藤井昭宏, 田中輝雄, 加藤由花, 大島聡史, 片桐孝洋
2. 発表標題 マルチGPU環境における機械学習ハイパーパラメータの自動チューニング(1)
3. 学会等名 情報処理学会第83回全国大会
4. 発表年 2021年

1. 発表者名 藤家空太郎, 多部田敏樹, 藤井昭宏, 田中輝雄, 加藤由花, 大島聡史, 片桐孝洋
2. 発表標題 マルチGPU環境における機械学習ハイパーパラメータの自動チューニング(2)
3. 学会等名 情報処理学会第83回全国大会
4. 発表年 2021年

1. 発表者名 土肥樹, 菱沼利彰, 田中輝雄, 藤井昭宏
2. 発表標題 Intel Xeonにプロセッサ上における AVX512を用いた倍々精度演算の性能評価
3. 学会等名 応用数学学会年会
4. 発表年 2020年

1. 発表者名 多部田敏樹, 藤井昭宏, 田中輝雄, 滝沢寛之
2. 発表標題 複数のパラメータを同時に推定する手法の実装と機械学習ソフトウェアへの適用
3. 学会等名 計算工学講演会論文集, Vol.25
4. 発表年 2020年

1. 発表者名 Toshiki Tabeta, Naoto Seki, Akihiro Fujii, Teruo Tanaka, Hiroyuki Takizawa
2. 発表標題 An Optimization Technology of Software Auto-Tuning Applied to Machine Learning Software
3. 学会等名 International Conference Series on HPC Technologies in Asia Pacific Region(HPC Asia 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 Naoto Seki, Toshiki Tabeta, Akihiro Fujii, Teruo Tanaka
2. 発表標題 Stable Automatic Tuning Method for Performance Fluctuation
3. 学会等名 New Approaches for Software Auto-Tuning and Accuracy Assurance, Part II of II, The SIAM Conference on Parallel Processing for Scientific Computing, (PP2020) (国際学会)
4. 発表年 2020年

1. 発表者名 范 谷瑛, 関 直人, 多部田 敏樹, 藤井 昭宏, 田中 輝雄
2. 発表標題 ソフトウェア自動チューニングにおける反復2次元d-Spline探索法の提案と評価
3. 学会等名 研究報告ハイパフォーマンスコンピューティング (HPC)
4. 発表年 2019年

1. 発表者名 Naoto Seki, Toshiki Tabeta, Akihiro Fujii, Teruo Tanaka
2. 発表標題 Approach and Application for Performance Parameter Search on AT
3. 学会等名 ITBL Booth Poster, Super Computing 19(SC19) (国際学会)
4. 発表年 2019年

1. 発表者名 関直人, 范谷瑛, 多部田敏樹, 藤井昭宏, 田中輝雄
2. 発表標題 性能パラメータ推定における評価対象プログラムの実行時間の揺らぎに対応した自動チューニング手法の提案
3. 学会等名 情報処理学会, HPC研究会, Vol.2019-HPC-169, No.9, pp.1-8 (2019.5)
4. 発表年 2019年

1. 発表者名 Teruo Tanaka, Fan Guuing, Akihiro Fujii, Takahiro Katagiri
2. 発表標題 Enhancement of Performance Parameter Search Method for Multiple Parameter Estimation
3. 学会等名 2019 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing (ATAT in HPSC 2019) (国際学会)
4. 発表年 2019年

1. 発表者名 多部田敏樹, 田中輝雄, 藤井昭宏, 関 直人, 范 谷瑛
2. 発表標題 ソフトウェア自動チューニングにおける複数性能パラメータを同時推定する手法の効率化
3. 学会等名 第81回情報処理学会全国大会
4. 発表年 2019年

1. 発表者名 出蔵英真, 藤井昭宏, 田中輝雄
2. 発表標題 共役勾配法におけるダブルバッファリング利用したRDMA通信の性能評価
3. 学会等名 第81回情報処理学会全国大会
4. 発表年 2019年

1. 発表者名 Masaki Suwa, Akihiro Fujii, Teruo Tanaka, Hidehiko Hasegawa
2. 発表標題 Strategy of Precision Switching for Mixed Precision Iterative Method
3. 学会等名 10th International Workshop on Parallel Matrix Algorithms and Applications (PMAA 18) (国際学会)
4. 発表年 2018年

1. 発表者名 Naoto Seki , Guuing Fan, Akihiro Fujii, Teruo Tanaka
2. 発表標題 Two Approaches for Performance Parameter Search on AT in the simulation of the one-dimensional motion of mass points
3. 学会等名 Super Computing 2018(SC18), ITBL booth, Poster (国際学会)
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	長谷川 秀彦 (Hasegawa Hidehiko) (20164824)	筑波大学・図書館情報メディア系・教授 (12102)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------