様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ５ 月 ２３ 日現在

機関番号：８２４０１
研究種目：若手研究
研究期間：2019～2022
課題番号：１９Ｋ２０２８６

研究課題名（和文）超並列計算環境のための高精度かつ再現性のある行列計算ライブラリの開発

研究課題名（英文）Development of accurate and reproducible matrix computation library for massively parallel environments

研究代表者

椋木　大地（Mukunoki, Daichi）

国立研究開発法人理化学研究所・計算科学研究センター・研究員

研究者番号：９０７４２２８９

交付決定額（研究期間全体）：（直接経費）　　1,800,000円

研究成果の概要（和文）：本研究ではスーパーコンピュータ等に採用される超並列アーキテクチャを対象に，計算精度の高精度化と計算結果の再現性を保証可能な，CPUとGPUに対応した基本線形代数演算ライブラリ（いわゆるBasic Linear Algebra Subprograms，BLAS）の開発を行った．本研究では主として尾崎スキームに着目し，高精度かつ再現可能なBLASの高性能実装を開発し，疎行列反復ソルバーへの応用を示した．さらに応用として，低精度演算器（Tensor Cores）を用いた単精度/倍精度行列積の実装，そして単精度/倍精度行列積によるbinary128型4倍精度行列積の実装を提案した．

研究成果の学術的意義や社会的意義
CPUおよびGPUにおいて高精度かつ計算結果の再現が可能なBLASルーチンを実現し，疎行列ソルバーへの応用を示した．既存手法と比べて性能および実装が容易であり，応用数理分野での応用も期待できる．またAI向け低精度演算器を単精度・倍精度の行列計算に応用可能であることを示した．今後のハードウェアデザインへのインパクトも期待できる．

研究成果の概要（英文）：In this study, we developed the Basic Linear Algebra Subprograms (BLAS) for massively parallel architectures, which is accurate and can ensure reproducibility of computation results among different environments. Focusing mainly on the Ozaki scheme, we have developed a high-performance implementation of accurate and reproducible BLAS routines, and demonstrated its application to sparse iterative solvers on CPUs and GPUs. As further applications, we proposed an implementation of a single/double precision matrix multiplications using low-precision arithmetic units (Tensor Cores) and a binary128 matrix multiplication using single/double precision matrix multiplications.

研究分野：高性能計算

キーワード：高精度　再現性　行列計算　疎行列反復法

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

(1) 高精度計算：有限桁による浮動小数点演算は丸め誤差により計算結果の精度が悪化しうる．悪条件問題の計算においては倍精度演算でも精度不足に陥ることがある．また近年急速に性能が向上した主に AI 向けの FP16 等の低精度演算器は，そのままでは精度不足のため汎用計算において活用が難しいことから，高精度演算技術による低精度演算器の活用可能性が検討できる．

(2) 再現可能計算：浮動小数点演算は丸め誤差により演算順序によって計算結果が変わりうる．並列計算では演算順序が非決定的であることが多い．この問題はコードのデバッグ，コードの異なるシステムへのポーティング，シミュレーション等の品質保証において問題となりうる．再現可能計算法についてはさまざまな方法が存在するが，性能と実装コストの問題があり研究の余地が大きい．

(3) 超並列アーキテクチャ：GPU 等のメニーコアや大規模分散並列環境における高性能プログラミングは，性能を引き出すことの難しさ，スケーラビリティといった多くの挑戦がある．

本研究では上記のような背景から，科学技術計算の基礎となる基本的なベクトル・行列計算を対象に，計算の高精度化，計算結果の再現性の保証を実現し，超並列計算機環境で高性能を実現する実用的な数値計算ライブラリ実装手法の開発を試みる．

２．研究の目的

本研究では計算結果の高精度化および再現性の保証を実現し，最先端の超並列計算機環境において高性能を実現できる BLAS（Basic Linear Algebra Subprograms）ルーチン，および疎行列ベクトル積の開発を行う．これらは科学技術計算の基本カーネルである．超並列計算機環境として共有メモリ並列（CPU/GPU 等のメニーコア）および分散メモリ並列の両方を対象とする．また単・倍精度に加え，半精度（FP16）等の低精度演算器の活用も視野に入れる．高性能とは並列化効率とハードウェアの理論ピーク性能に対する実行効率の点で従来のベンダー実装と同等の効率を達成できることを意味する．本研究はライブラリ開発という実学的なテーマであるが，計算科学の観点で提案する実装手法の有効性や性能を理論的に議論する．また既存手法の多くが開発コストの問題で実用化に至っていない現状を踏まえ，開発コストに着目した議論を行うとともに，アプリケーションへの応用を示すことを目的とする．

３．研究の方法

高精度と再現性はどちらも丸め誤差に起因する問題であり，両者に同時に対処できる手法が存在する．また複数の手法を組み合わせることも検討する．本研究では高精度行列積計算法である尾崎スキーム（Ozaki et al. 2012）を中心として，再現可能計算法の ExBLAS スキーム（Iakymchuk et al. 2014），高精度演算法の DotK スキーム（Ogita et al. 2005），丸め誤差自動分析法である CADNA スキーム（Jezequel et al. 2008）にスポットをあてて研究を進める．

４．研究成果

(1) 2019 年度

尾崎スキームを用いた CPU･GPU 向けの高精度かつ再現可能な BLAS の基本ルーチンを開発した．

ベクトルバッキング，バッチ BLAS の活用などの実装最適化手法を示した．本手法は ExBLAS 等の既存手法と比べて実装が容易，階層的なソフトウェア開発が可能である．この成果は国際学会（PPAM2019）において査読付論文を発表した．また国際会議（RuSCDays2019）においてポスター発表し，Best Research Poster Award を受賞した．さらに尾崎スキームに基づく疎行列ベクトル積を実装し，上記 BLAS の内積と組み合わせることで，高精度かつ再現可能な疎行列反復ソルバー（CG 法）を開発した．また尾崎スキームを拡張し，FP16/32 の混合精度ハードウェアである Tensor Cores を活用した，単精度・倍精度の行列積あるいはその高精度版・再現可能版実装を開発し，国際学会（ISC2020）において査読付論文を投稿した．Tensor Cores を用いて倍精度行列積の計算を可能にした研究は本研究が初であり，低精度演算器の汎用計算への応用可能性を示した．CADNA スキームについては，共同研究を進めているソルボンヌ大学との共同研究で，内積ベースの演算において高コストな確率的演算をスキップする新しい手法を開発し，共著者として参加した論文を国際学会（NSV20）に投稿した．線形計算コードの精度検証（それは再現性の実現と関連する）の高速化が期待できる．さらに計算結果の精度を担保しながら数値計算に用いられる演算精度を最適化して計算の高速化，省電力化を実現する方法の研究を開始した．既存技術の組み合わせで実現可能なことを整理し，FPGA の活用等を検討した．本研究の応用と位置付けられる．この研究に関して国際会議（SC19）において査読付ポスター発表を行った．

(2) 2020 年度

昨年度採択された論文について国際会議（ISC2020）において発表を行い論文が出版された．同内容に関連して国際会（CSE21）においても口頭発表を行なった． ISC2020 の提案手法の電力性能を評価し，これに関して共著者として執筆に参加した論文が国 際会議 IPDPS2021 に採択された．また，ISC2020 の提案手法を応用した CPU における単精度・倍精度行列積を用いた binary128（IEEE 4 倍精度）型行列に対する行列積実装を開発し，日本応用数理学会 2020 年度年会において口頭発表した．昨年度に開発を開始した尾崎スキームに基づく高精度かつ再現性のある疎行列反復ソルバーに関する論文を国際会議(HPC Asia2021)に投稿し採択され，口頭発表を実施し，論文が出版された．この研究では ExBLAS スキームに基づく同等の実装との性能比較を実施し優位性が示された．さらに，昨年度に共著者として参加し投稿した NSV20 論文が採択され出版された．

(3) 2021 年度

日本応用数理学会 2020 年度年会で発表した尾崎スキームに基づく binary128(IEEE 4 倍精度)行列積について実装をさらに高速化し，国際会議（ISC2021）におけるポスター発表，国際会議（ICPP-2021）における査読付論文発表を行った．ISC2021 においては Research Poster Award を受賞した．また，尾崎スキームと DotK スキーム(Dot2)を組み合わせることで，無限精度の内積・疎行列ベクトル積を高速化する方法を開発した．CPU・GPU において実装を行い，昨年度開発した再現可能な疎行列反復解法に適用して，その高速化が可能であることを確認した．さらに，尾崎スキームに基づく高精度・再現可能 BLAS，および疎行列反復ソルバーをテンプレート化し，コードの混合精度対応化を実施した．また，CADNA スキームに基づく疎行列反復解法の精度検証について検討を行った．

(4) 2022 年度

尾崎スキームと DotK スキーム(Dot2)を組み合わせる無限精度内積の高速化法について，

CPU/GPU 実装のさらなる高速化を行い，国際会議におけるポスター発表(ISC2022)，査読付論文発表（PPAM2022）を行った．ISC2022 においては Research Poster Award 2nd Place Winner を受賞した．

本研究の総括として，まず主目的であった尾崎スキームによる高精度かつ再現可能な BLAS の高性能実装を開発し，DotK スキームと組み合わせたさらなる高速化法を提案した．さらに疎行列ソルバーへの適用によって応用が示された．また既存の ExBLAS スキーム等による優位性も示された.さらに低精度演算器の活用として，尾崎スキームの拡張による Tensor Cores を用いた単精度・倍精度行列積の実装，そして単精度・倍精度行列積による binary128 型 4 倍精度行列積の実装を提案した．概ね良い成果が得られたと総括する．

一方，大規模分散並列環境向けの実装や CADNA スキームに関する研究は予定した進捗が得られなかったが，本研究を基課題とする科学研究費助成事業 国際共同研究加速基金(国際共同研究強化(A))#20KK0259 が採択され，2022 年度より実施の運びとなった．本研究の継続課題を引き継いでさらに発展させることが期待される．

５．主な発表論文等

〔雑誌論文〕　計7件（うち査読付論文　6件／うち国際共著　4件／うちオープンアクセス　2件）

１．著者名 Mukunoki Daichi、Ozaki Katsuhisa、Ogita Takeshi、Imamura Toshiyuki	４．巻 --
２．論文標題 Accurate Matrix Multiplication on Binary128 Format Accelerated by Ozaki Scheme	５．発行年 2021年
３．雑誌名 Proc. The 50th International Conference on Parallel Processing (ICPP-2021)	６．最初と最後の頁 1-11
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3472456.3472493	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Mukunoki Daichi、Ozaki Katsuhisa、Ogita Takeshi、Imamura Toshiyuki	４．巻 12151
２．論文標題 DGEMM Using Tensor Cores, and Its Accurate and Reproducible Versions	５．発行年 2020年
３．雑誌名 Proc. ISC High Performance 2020, Lecture Notes in Computer Science	６．最初と最後の頁 230～248
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-50743-5_12	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Jezequel Fabienne、Graillat Stef、Mukunoki Daichi、Imamura Toshiyuki、Iakymchuk Roman	４．巻 12549
２．論文標題 Can We Avoid Rounding-Error Estimation in HPC Codes and Still Get Trustworthy Results?	５．発行年 2020年
３．雑誌名 Proc. 13th International Workshop on Numerical Software Verification 2020 (NSV 20), Lecture Notes in Computer Science	６．最初と最後の頁 163～177
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-63618-0_10	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

１．著者名 Mukunoki Daichi、Ozaki Katsuhisa、Ogita Takeshi、Iakymchuk Roman	４．巻 -
２．論文標題 Conjugate Gradient Solvers with High Accuracy and Bit-wise Reproducibility between CPU and GPU using Ozaki scheme	５．発行年 2021年
３．雑誌名 Proc. The International Conference on High Performance Computing in Asia-Pacific Region (HPCAsia 2021)	６．最初と最後の頁 100～109
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3432261.3432270	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Jens Domke, Emil Vatai, Aleksandr Drozd, Peng Chen, Yosuke Oyama, Lingqi Zhang, Shweta Salaria, Daichi Mukunoki, Artur Podobas, Mohamed Wahib, Satoshi Matsuoka	4．巻 -
2．論文標題 Matrix Engines for High Performance Computing: A Paragon of Performance or Grasping at Straws?	5．発行年 2021年
3．雑誌名 Proc. 35th IEEE International Parallel & Distributed Processing Symposium (IPDPS 2021)	6．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/IPDPS49936.2021.00114	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1．著者名 Daichi Mukunoki, Takeshi Ogita, Katsuhisa Ozaki	4．巻 12043
2．論文標題 Reproducible BLAS Routines with Tunable Accuracy Using Ozaki Scheme for Many-core Architectures	5．発行年 2020年
3．雑誌名 13th International Conference on Parallel Processing and Applied Mathematics (PPAM2019), Lecture Notes in Computer Science	6．最初と最後の頁 516-527
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-43229-4_44	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1．著者名 Toshiyuki Imamura, Daichi Mukunoki, Fabienne Jezequel, Stef Graillat, Roman Iakymchuk	4．巻 -
2．論文標題 Numerical Reproducibility based on Minimal-Precision Validation	5．発行年 2019年
3．雑誌名 Computational Reproducibility at Exascale Workshop (CRE2019)	6．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕　計22件（うち招待講演　0件／うち国際学会　17件）

1．発表者名 Daichi Mukunoki, Katsuhisa Ozaki, Takeshi Ogita, Toshiyuki Imamura
2．発表標題 Accurate Matrix Multiplication on Binary128 using Ozaki Scheme
3．学会等名 ISC High Performance (ISC 2021), research poster session（国際学会）
4．発表年 2021年

１．発表者名
Roman Iakymchuk, Daichi Mukunoki, Takeshi Ogita, Katsuhisa Ozaki, Stef Graillat

２．発表標題
Accurate and Reproducible Conjugate Gradient in Hybrid Parallel Environments

３．学会等名
ISC High Performance (ISC 2021), research poster session（国際学会）

４．発表年
2021年

１．発表者名
Daichi Mukunoki, Katsuhisa Ozaki, Takeshi Ogita, Toshiyuki Imamura, Roman Iakymchuk

２．発表標題
High-Precision, Accurate, and Reproducible Linear Algebra Operations using Ozaki Scheme

３．学会等名
3rd R-CCS International Symposium（国際学会）

４．発表年
2021年

１．発表者名
Daichi Mukunoki, Katsuhisa Ozaki, Takeshi Ogita, Toshiyuki Imamura, Roman Iakymchuk

２．発表標題
Impact and Contribution of Ozaki scheme in High Performance Computing

３．学会等名
International Workshop on Reliable Computing and Computer-Assisted Proofs (ReCAP 2022)（国際学会）

４．発表年
2022年

１．発表者名
椋木大地

２．発表標題
精度自動チューニングに向けた基盤技術の検討

３．学会等名
第13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 (ATTA2021)

４．発表年
2021年

１．発表者名
Daichi Mukunoki, Katsuhisa Ozaki, Takeshi Ogita, Toshiyuki Imamura

２．発表標題
DGEMM using Tensor Cores

３．学会等名
SIAM Conference on Computational Science and Engineering (CSE21)（国際学会）

４．発表年
2021年

１．発表者名
Daichi Mukunoki

２．発表標題
DGEMM using Tensor Cores and OzBLAS

３．学会等名
11th Joint Laboratory for Extreme Scale Computing (JLESC) Workshop（国際学会）

４．発表年
2020年

１．発表者名
椋木大地，尾崎克久，荻田武史

２．発表標題
binary128に対する尾崎スキーム行列積

３．学会等名
第4回精度保証付き数値計算の実問題への応用研究集会（NVR 2020）

４．発表年
2020年

１．発表者名
Daichi Mukunoki, Katsuhisa Ozaki, Takeshi Ogita, Toshiyuki Imamura

２．発表標題
DGEMM using Tensor Cores

３．学会等名
SIAM Conference on Computational Science and Engineering (CSE21)（国際学会）

４．発表年
2021年

１．発表者名
椋木大地，尾崎克久，荻田武史

２．発表標題
尾崎スキームを用いたbinary128による4倍精度行列積

３．学会等名
日本応用数理学会2020年度年会

４．発表年
2020年

１．発表者名
Daichi Mukunoki, Katsuhisa Ozaki, Takeshi Ogita, Toshiyuki Imamura

２．発表標題
Accurate DGEMM using Tensor Cores

３．学会等名
HPC Asia 2020 (poster session) (国際学会)

４．発表年
2020年

１．発表者名
Roman Iakymchuk, Fabienne Jezequel, Stef Graillat, Daichi Mukunoki, Toshiyuki Imamura, Yiyu Tan, Atsushi Koshiba, Jens Huthmann, Kentaro Sano, Norihisa Fujita, Taisuke Boku

２．発表標題
Optimizing Precision for High-Performance, Robust, and Energy-Efficient Computations

３．学会等名
HPC Asia 2020 (poster session) (国際学会)

４．発表年
2020年

１．発表者名
Daichi Mukunoki, Toshiyuki Imamura, Yiyu Tan, Atsushi Koshiba, Jens Huthmann, Kentaro Sano, Fabienne Jezequel, Stef Graillat, Roman Iakymchuk, Norihisa Fujita, Taisuke Boku

２．発表標題
Minimal-Precision Computing for High-Performance, Energy-Efficient, and Reliable Computations

３．学会等名
SC19 (research poster session) (国際学会)

４．発表年
2019年

１．発表者名
Daichi Mukunoki, Toshiyuki Imamura, Yiyu Tan, Atsushi Koshiba, Jens Huthmann, Kentaro Sano, Fabienne Jezequel, Stef Graillat, Roman Iakymchuk, Norihisa Fujita, Taisuke Boku

２．発表標題
Minimal-Precision Computing for High-Performance, Energy-Efficient, and Reliable Computations

３．学会等名
France-Japan-Germany trilateral workshop: Convergence of HPC and Data Science for Future Extreme Scale Intelligent Applications (poster presentation) (国際学会)

４．発表年
2019年

１．発表者名
Daichi Mukunoki, Takeshi Ogita, Katsuhisa Ozaki:

２．発表標題
Accurate and Reproducible Linear Algebra Operations for Many-core Architectures

３．学会等名
Russian Supercomputing Days 2019 (RuSCDays 2019) (poster session) (国際学会)

４．発表年
2019年

１．発表者名
Daichi Mukunoki

２．発表標題
Minimal-Precision Computing for High-Performance, Energy-Efficient, and Reliable Computations

３．学会等名
SIAM Conference on Parallel Processing for Scientific Computing (PP20) (国際学会)

４．発表年
2020年

１．発表者名
Daichi Mukunoki

２．発表標題
Accurate BLAS implementations: OzBLAS and BLAS-DOT2

３．学会等名
Workshop on Largescale Parallel Numerical Computing Technology (LSPANC 2020 January) (国際学会)

４．発表年
2020年

１．発表者名
Daichi Mukunoki

２．発表標題
Minimal-Precision Computing for High-Performance, Energy-Efficient, and Reliable Computations

３．学会等名
Sapporo Winter HPC Seminar 2020（国際学会）

４．発表年
2020年

１．発表者名
椋木大地，荻田武史，尾崎克久

２．発表標題
尾崎スキームによる高精度BLAS実装「OzBLAS」とその応用

３．学会等名
第3回 精度保証付き数値計算の実問題への応用研究集会（NVR 2019）

４．発表年
2019年

１．発表者名
Daichi Mukunoki, Takeshi Ogita, Katsuhisa Ozaki

２．発表標題
Accurate and Reproducible CG Method on GPUs

３．学会等名
European Numerical Mathematics and Advanced Applications Conference 2019 (ENUMATH2019)（国際学会）

４．発表年
2019年

１．発表者名
Daichi Mukunoki

２．発表標題
High-Performance Implementations of Accurate and Reproducible BLAS Routines on GPUs

３．学会等名
Workshop on Largescale Parallel Numerical Computing Technology (LSPANC 2019 June)（国際学会）

４．発表年
2019年

１．発表者名 椋木大地
２．発表標題 尾崎スキームに基づく高精度かつ再現性のあるBLASルーチンの実装と自動チューニングの適用
３．学会等名 第22回AT研究会オープンアカデミックセッション（ATOS22）
４．発表年 2019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

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			
フランス	Sorbonne University			