

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 17 日現在

機関番号：12102

研究種目：新学術領域研究（研究領域提案型）

研究期間：2010～2014

課題番号：22104003

研究課題名（和文）大規模並列環境における数値計算アルゴリズム

研究課題名（英文）Numerical Computation Algorithms for Large-scale Parallel Environment

研究代表者

高橋 大介 (Takahashi, Daisuke)

筑波大学・システム情報系・教授

研究者番号：00292714

交付決定額（研究期間全体）：（直接経費） 17,300,000 円

研究成果の概要（和文）：ベタフロップスを超える性能を持つ大規模並列環境における数値計算アルゴリズムとして、高速フーリエ変換（FFT）、GPUによる4倍・8倍精度BLAS、GPUによる3倍・4倍精度浮動小数点演算、GPUにおける疎行列ベクトル積、CPU+GPU環境上で動作する実対称固有値ソルバ、GPU環境下での固有値ソルバ開発と既存ソルバとの性能評価、第3世代NVIDIA GPUを用いた高性能固有値ソルバの開発、Block Krylovアルゴリズムによる連立一次方程式の求解高速化、そしてBlock Krylov部分空間反復法の精度向上について研究を行うと共に、性能チューニング手法について検討した。

研究成果の概要（英文）：Research was conducted on the use of large-scale parallel environments, whose performance exceeded 10 petaflops, for numerical computation algorithms. These included the fast Fourier transform (FFT), quadruple- and octuple-precision basic linear algebra subprograms (BLAS) on graphics processing units (GPUs), triple- and quadruple-precision floating point operations on GPUs, sparse matrix-vector multiplication on GPUs, and a real symmetric eigenvalue solver operating in a CPU + GPU environment. An evaluation was also carried out of the performance of a newly developed eigenvalue solver and an existing solver in a GPU environment, a high-performance eigenvalue solver employing third-generation NVIDIA GPUs, solution acceleration of simultaneous linear algebraic equations using the Block Krylov algorithm, accuracy improvement of Block Krylov partial space iterative methods, and performance-tuning methods.

研究分野：計算機科学

キーワード：大規模並列環境 数値計算アルゴリズム

1. 研究開始当初の背景

(1) 2010年3月の時点で、1PFlopsを超える理論ピーク性能を持つスーパーコンピュータが5システム出現していた。次世代、次世代のスーパーコンピュータとしては、マルチコアCPUに加えてGPUなどのアクセラレータを搭載した計算ノードを数千～数万台以上接続したものが主流になることが予想されていた。

(2) このようなマルチコアCPU(+マルチGPU)から構成されるスーパーコンピュータにおいては、プロセッサコア数の増加や演算性能あたりのメモリバンド幅の不足などにより、高い実行効率を得ることが困難になりつつあった。

(3) したがって、今後計算科学においてランドチャレンジを行うためには、これまでに提案されてきた並列数値計算アルゴリズムや性能チューニング手法を用いるだけでは不十分であることが予想された。

2. 研究の目的

(1) 本研究の目的としては、ペタフロップスを超える性能を持つ次世代のスーパーコンピュータに向けた数値計算アルゴリズムや性能チューニング手法について研究を行うことである。

(2) これまでの研究成果を十分に活用し、研究期間内に以下の点について明らかにする。
 ・ペタスケール計算環境に向けた並列数値計算アルゴリズムを実現する。
 ・現在利用可能なマルチコア超並列クラスタおよびGPUクラスタにおいて性能評価を行い、提案する並列数値計算アルゴリズムの高速性を実証する。
 ・これらの結果より、ペタスケールおよびエクサスケール計算環境に適したアルゴリズムおよび最適化手法についての知見を得る。

3. 研究の方法

(1) 高速フーリエ変換、固有値計算、疎行列連立一次方程式の反復解法について、ペタフロップス級の次世代スーパーコンピュータに向けた数値計算ライブラリおよび数値アルゴリズムについて研究を行う。

(2) これらの研究においては、本計画班内だけではなく領域内の計算物質科学の研究者と連携し、実アプリケーションプログラムをマルチコア大規模並列コンピュータで実行した場合に、性能向上の妨げになる部分を見だし、新たなアルゴリズムや性能チューニング手法の構築を行う。

(3) さらに、アクセラレータ(GPU)向けの性能チューニング手法についても検討する。

4. 研究成果

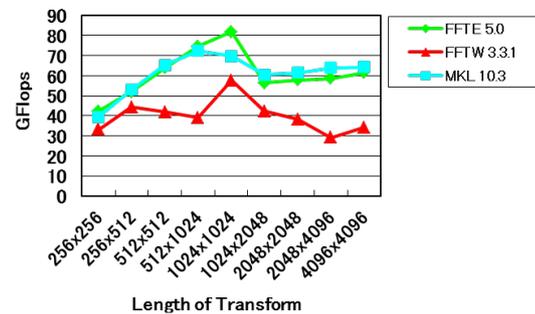
(1) Intel AVX 命令を用いた並列二次元 FFT の実現と評価

浮動小数点演算をより高速に処理するために、最近のプロセッサでは Intel Xeon の SSE ,SSE2 ,SSE3 ,SSSE3 ,SSE4 ,AVX や Motorola PowerPC の AltiVec ,そして Fujitsu SPARC64 Viiifx の HPC-ACE など、Short Vector SIMD 命令を搭載しているものが多い。

しかし、これらの Short Vector SIMD 命令を使ったとしても、最近のプロセッサのデータ供給能力は、キャッシュに頼っているのが現状であり、メモリアクセスの最適化もあわせて行う必要がある。

そこで、AVX 命令を用いて FFT カーネル部分の性能を向上させるとともに、ブロック二次元 FFT アルゴリズムを用いることで、データがキャッシュに収まらない場合にも高い性能を維持する FFT ライブラリ FFT を実装した。

以下の図に Intel Xeon E5-2670 (Sandy Bridge-EP 2.6 GHz, 20 MB L3 cache, 2 CPUs, 16 cores, 128 GB DDR3-SDRAM) における並列二次元 FFT の性能を示す。FFTW に比べて FFTE が高い性能を示していることが分かる。



(2) GPU による 3 倍・4 倍精度 BLAS

GPU による 4 倍精度 BLAS を実現し、NVIDIA Tesla C2050 において性能評価を行った結果、Intel Core i7 920 での同一処理と比べ、4 倍精度 AXPY が約 9.5 倍、8 倍精度 AXPY が約 19 倍高速化された。さらに、3 倍精度数を倍精度数と単精度数に分けて格納する Double+Single 型 3 倍精度型 (D+S 型) および D+S 型 3 倍精度演算 (D+S 型演算) を提案し、GPU により 3 倍精度の BLAS ルーチンを実装して、その性能を Tesla C2050 で評価した。その結果、Tesla C2050 では 3 倍精度 AXPY が CUBLAS の倍精度 AXPY の約 1.57 倍の実行時間、3 倍精度 GEMV が倍精度 GEMV の約 1.69 倍の実行時間となり、それぞれ 4 倍精度ルーチンよりも高速な性能を示した。

(3) GPU における格納形式自動選択による疎行列ベクトル積の高速化

科学技術計算に用いられる連立一次方程式では、疎行列を対象とするものが多い。この場合、反復法により解を求めることになり、疎行列ベクトル積が何度も呼び出されるこ

とになるため、この疎行列ベクトル積の高速化が求められる。疎行列の格納形式には様々なものが提案されているが、対象となる疎行列によって各格納形式での疎行列ベクトル積の性能に差が生じる。疎行列ベクトル積は演算量よりもメモリアクセス量が多い演算であるため、GPU の性能を發揮させるためには最適な格納形式を選択することが重要である。

そこで本研究では、疎行列の格納形式を自動選択することで疎行列ベクトル積の高速化を行った。NVIDIA Tesla C2050 による性能評価の結果、今回用いた疎行列のうち多くの疎行列において最適な格納形式を選択し、CRS形式のまま CUSPARSE を用いて実行する場合と同等、またはそれよりも高速であることが分かった。また、特に反復回数が少ない時でも本研究で提案した自動選択が有効であることが分かった。

(4) SYMV kernel on ASPEN-K2

GPU-CUDA 版固有値ソルバ Eigen-sg を開発した。Eigen-sg の計算で最も負荷の高い部分はハウスホルダー三重対角化であり、さらにその詳細分析を進めると対称行列ベクトル積操作ということが明らかになる。GPU 上の BLAS を CUDA-BLAS で置き換えた最も簡単な GPGPU 化がこれまでの実装であった。また、独自の性能改善を行った ASPEN.K2 を取り込むことにより性能向上を進めてきたが、行列ベクトル積 SYMV に対して独自の実装方式である atomic アルゴリズムと L+U アルゴリズムを併用した実装方式を提案した。著名な GPGPU 数値計算ソフトウェアである MAGMA と比較して 1.3 から 1.5 倍の高速化を達成している。

(5) Block Krylov 部分空間反復法の精度向上

複数の右辺ベクトルをもつ連立一次方程式の数値解法として、Block Krylov 部分空間反復法がある。Block Krylov 部分空間反復法の反復過程において、残差のノルムが激しく振動する場合は精度劣化が生じ、近似解が目的精度まで到達しないことがある。双共役残差型アプローチに基づく反復解法である Block BiCR 法は、Block BiCG 法と比較して残差ノルムの履歴が滑らかになる傾向にある。Block BiCR 法では、係数行列に関する行列積の回数を減少させるため、行列積を漸化式によって求める箇所が存在する。この漸化式計算が近似解の精度に影響し、Block BiCR 法で得られる近似解の精度が劣化することがある。

本研究では、Block BiCR 法で生成される近似解の精度向上を目指し、同法の修正版を構築した。真の相対残差が小さい場合は、高精度の近似解が得られていることを示す。修正 Block BiCR 法を用いることで、残差ノルムの振動を抑えることができ、かつ他の解法より

も高精度の近似解を得ることができた。

5. 主な発表論文等

(研究代表者、研究分担者および連携研究者には下線)

[雑誌論文](計16件)

Shusaku Saito, Hiroto Tadano, and Akira Imakura, Development of the Block BiCGSTAB(l) method for solving linear systems with multiple right hand sides, JSIAM Letters, Vol. 6, pp. 65-68, 2014. 査読有
DOI:10.14495/jsiaml.6.65

Hiroto Tadano, Youichi Ishikawa, and Akira Imakura, Improvement of the accuracy of the approximate solution of the Block BiCR method, JSIAM Letters, Vol. 6, pp. 61-64, 2014. 査読有
DOI:10.14495/jsiaml.6.61

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. 査読有
DOI:10.1109/CSE.2013.36

Akira Imakura, Lei Du, and Hiroto Tadano, A weighted Block GMRES method for solving linear systems with multiple right hand sides, JSIAM Letters, Vol. 5, pp. 65-68, 2013. 査読有
DOI:10.14495/jsiaml.5.65

山崎育朗, 今倉暁, 多田野寛人, 櫻井鉄也, 残差最小性に基づく Krylov 部分空間反復解法に対する疎行列用直接解法を用いた前処理のパラメータ推定, 日本応用数学会論文誌, Vol. 23, pp. 381-404, 2013. 査読有
<http://ci.nii.ac.jp/naid/110009636317>

Tetsuya Sakurai, Yasunori Futamura, and Hiroto Tadano, Efficient parameter estimation and implementation of a contour integral-based eigensolver, Journal of Algorithms & Computational Technology, Vol. 7, pp. 249-269, 2013. 査読有
<http://connection.ebscohost.com/c/articles/89860149/efficient-parameter-estimation-implementation-contour-integral-based-eigensolver>

Ichitaro Yamazaki, Hiroto Tadano,

Tetsuya Sakurai, and Tsutomu Ikegami, Performance comparison of parallel eigensolvers based on a contour integral method and a Lanczos method, *Parallel Computing*, Vol. 39, pp. 280-290, 2013. 査読有
DOI:10.1016/j.parco.2012.04.001

Toshiyuki Imamura, Susumu Yamada, and Masahiko Machida, A High Performance SYMV Kernel on a Fermi-core GPU, Proc. 10th International Meeting on High-Performance Computing for Computational Science (VECPAR 2012), *Lecture Notes in Computer Science*, Vol. 7851, pp. 59-71, Springer-Verlag, 2013. 査読有
DOI:10.1007/978-3-642-38718-0_9

Y. Nakamura, K.-I. Ishikawa, Y. Kuramashi, T. Sakurai, and H. Tadano, Modified Block BiCGSTAB for Lattice QCD, *Computer Physics Communications*, Vol. 183, pp. 34-37, 2012. 査読有
DOI:10.1016/j.cpc.2011.08.010

Michihiro Naito, Hiroto Tadano, and Tetsuya Sakurai, A modified Block IDR(s) method for computing high accuracy solutions, *JSIAM Letters*, Vol. 4, pp. 25-28, 2012. 査読有
DOI:10.14495/jsiaml.4.25

Daisuke Takahashi, An Implementation of Parallel 2-D FFT Using Intel AVX Instructions on Multi-Core Processors, Proc. 12th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2012), Part II, *Lecture Notes in Computer Science*, Vol. 7440, pp. 197-205, Springer-Verlag, 2012. 査読有
DOI:10.1007/978-3-642-33065-0_21

Y. Nakamura, K.-I. Ishikawa, Y. Kuramashi, T. Sakurai, and H. Tadano, Modified block BiCGSTAB for lattice QCD, *Computer Physics Communications*, Vol. 183, pp. 34-37, 2012. 査読有
DOI:10.1016/j.cpc.2011.08.010

Daichi Mukunoki and Daisuke Takahashi, Implementation and Evaluation of Triple Precision BLAS Subroutines on GPUs, Proc. 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW 2012), The 13th Workshop

on Parallel and Distributed Scientific and Engineering Computing (PDSEC-12), pp. 1378-1386, 2012. 査読有
DOI:10.1109/IPDPSW.2012.175

Daichi Mukunoki and Daisuke Takahashi, Implementation and Evaluation of Quadruple Precision BLAS Functions on GPUs, Proc. 10th International Conference on Applied Parallel and Scientific Computing (PARA 2010), Part I, *Lecture Notes in Computer Science*, Vol. 7133, pp. 249-259, Springer-Verlag, 2012. 査読有
DOI:10.1007/978-3-642-28151-8_25

Takatoshi Nakayama and Daisuke Takahashi, Implementation of Multiple-Precision Floating-Point Arithmetic Library for GPU Computing, Proc. 23rd IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS 2011), pp. 343-349, 2011. 査読有
DOI:10.2316/P.2011.757-041

Yuji Kubota and Daisuke Takahashi, Optimization of Sparse Matrix-Vector Multiplication by Auto Selecting Storage Schemes on GPU, Proc. 11th International Conference on Computational Science and Its Applications (ICCSA 2011), Part II, *Lecture Notes in Computer Science*, Vol. 6783, pp. 547-561, Springer-Verlag, 2011. 査読有
DOI:10.1007/978-3-642-21887-3_42

[学会発表](計4件)

Hiroshi Maeda and Daisuke Takahashi, Performance Evaluation of Sparse Matrix-Vector Multiplication Using GPU/MIC Cluster, 2015 SIAM Conference on Computational Science and Engineering, 2015年3月14日, Salt Palace Convention Center (Salt Lake City, Utah, USA).

今村俊幸, 椋木大地, 山田進, 町田昌彦, CUDABLAS等の選択による最速GPU固有値ソルバーの性能評価, 情報処理学会第148回ハイパフォーマンスコンピューティング研究発表会, 2015年3月2日, 花菱ホテル(別府市).

今村俊幸, 椋木大地, 山田進, 町田昌彦, CUDA-xSYMVの実装と評価, 情報処理学会第146回ハイパフォーマンスコンピュー

ティング研究発表会 2014 年 10 月 3 日，
沖縄産業支援センター（那覇市）。

前田広志，高橋大介，GPU/MIC クラスタ
における疎行列ベクトル積の性能評価，
情報処理学会第 144 回ハイパフォーマンス
コンピューティング研究発表会，2014
年 5 月 26 日，独立行政法人海洋研究開発
機構横浜研究所三好記念講堂（横浜市）。

6．研究組織

(1)研究代表者

高橋 大介 (TAKAHASHI, Daisuke)
筑波大学・システム情報系・教授
研究者番号：00292714

(2)研究分担者

今村 俊幸 (IMAMURA, Toshiyuki)
理化学研究所・計算科学研究機構・チーム
リーダー
研究者番号：60361838

多田野 寛人 (TADANO, Hiroto)
筑波大学・システム情報系・助教
研究者番号：50507845

(3)連携研究者

佐藤 三久 (SATO, Mitsuhsa)
筑波大学・システム情報系・教授
研究者番号：60333481

朴 泰祐 (BOKU, Taisuke)
筑波大学・システム情報系・教授
研究者番号：90209346

櫻井 鉄也 (SAKURAI, Tetsuya)
筑波大学・システム情報系・教授
研究者番号：60187086