

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 9 日現在

機関番号：11301

研究種目：挑戦的萌芽研究

研究期間：2014～2016

課題番号：26540017

研究課題名(和文)メモリオール問題を突破するバンド幅圧縮ハードウェア基盤技術の創出

研究課題名(英文)Fundamental Technologies for Bandwidth Compression Hardware to Overcome Memory-Wall Problem

研究代表者

佐野 健太郎 (Sano, Kentaro)

東北大学・情報科学研究科・准教授

研究者番号：00323048

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：計算性能の低下をもたらすメモリオール問題を解決するために、可逆データ圧縮をリアルタイムに行うバンド幅圧縮ハードウェアを提案し、実応用問題に適用可能な基盤技術を創出した。数値データの連続性を利用した予測に基づく圧縮アルゴリズムと複数のデータストリームを扱うための符号化方式、およびそのハードウェアを設計した後に、FPGAによる実装を行った。ベンチマーク問題として格子ボルツマン法に基づく流体計算を用い、実際に動作する高性能ストリーム計算システムを構築した。計算途中データのDDR3メモリへの読書きを圧縮したところ、正しい計算が行えることと、データストリームの実効バンド幅を向上できることを確認できた。

研究成果の概要(英文)：We developed technologies for bandwidth compression hardware that can be applied to numerical computing problems in order to address the memory-wall problem which degrades the computational performance. We have proposed a lossless data-compression algorithm based on prediction using continuity in numerical data and an encoding format to handle multiple data streams with compression. Then we have designed and implemented the bandwidth compression hardware by using FPGA. We built a high-performance stream computing system which performs fluid dynamics simulation with the lattice Boltzmann method with and without the bandwidth compression. We demonstrate that the system performs correct computation with bandwidth compression for DDR3 memories. We also made sure that the bandwidth compression actually enhances the sustained memory bandwidth for internal computing core, resulting in increasing performance by alleviating the memory bottleneck.

研究分野：計算機科学

キーワード：バンド幅圧縮 ハードウェア 高性能計算 FPGA リンコンフィギャラブル 可逆データ圧縮

1. 研究開始当初の背景

高性能計算を実現するためのスーパーコンピュータ(SC)には、高い電力性能比が求められている。しかしながら、汎用のメニーコアプロセッサをネットワークで接続した巨大並列機という現在の SC アーキテクチャでは、電力あたり実効性能へのブレークスルーは困難である。これに対し、計算問題毎にハードウェアそのものを最適な形に変え性能向上を実現するリコンフィギャラブル計算(RC)が、低電力高性能計算のプラットフォームとして有望視されている。特に、近年回路再構成可能デバイス FPGA は、その大規模化・高性能化・低電力化に加え浮動小数点演算器を搭載する等のアーキテクチャの改良により、ビットや整数から浮動小数点に渡る様々な演算に基づく計算処理に対し、メニーコアプロセッサや GPU 等のアクセラレータを上回る性能が期待されている。

RC による高性能計算の鍵は、問題毎に特化したデータ移動である。アプリケーション毎に求められるデータ移動に最適化されたメモリシステムやネットワークを構築できれば、そこがボトルネックとならずに計算性能を引き出すことができる。しかしながら、メモリやネットワークのバンド幅そのものは大変貴重な資源であり、それが不足する場合には性能向上の妨げとなる。バンド幅不足の問題に対し、アプリごとに専用ハードウェアを構築できるという RC の特長を利用した解決方法が求められる。

2. 研究の目的

本研究の目的は、低電力・高性能計算機として有望視されているリコンフィギャラブル高性能計算に必要な、メモリ・ネットワークの実効性能向上のためのバンド幅圧縮ハードウェア基盤技術の創出である。メモリやネットワークのバンド幅制約を高スループットデータ圧縮ハードウェア技術により克服し、従来はオフチップバンド幅が限界を与えるような高性能計算問題に対してもオンチップの演算性能をさらに引き出せるような高効率システムを構築可能とする。計算のためのデータストリームに対しバンド幅圧縮アルゴリズム、およびそのハードウェアモジュールを提案・開発し、回路再構成デバイスである FPGA を用いた試作システムにより、実用問題についてその有効性を実証する。

3. 研究の方法

本研究では、対象とする浮動小数点数値データストリームの特徴を調査した上で、その可逆圧縮アルゴリズムの設計と、データ圧縮を高スループットで実現する小面積バンド幅圧縮ハードウェア IP ライブラリの開発を行う。FPGA を用いてバンド幅圧縮ハードウェア

と数値計算 IP コアを搭載した実験システムを実装し、実際の計算に対して提案方式の有効性を実証する。

4. 研究成果

初年度では、偏微分方程式を解く計算問題をリコンフィギャラブル高性能計算の対象と位置付け、特にその中でも浮動小数点データによる数値流体力学計算に関して研究を実施した。まず、流体計算のデータについてその空間的連続性を評価したところ、小さな誤差の下で局所的に 1~3 次の多項式近似が可能であることが分かった。この知見に基づき、多項式による予測に基づいた圧縮アルゴリズムと符号化方式を、ハードウェア化を念頭に置きながら設計した。

特に、少ないハードウェアリソースで実装が可能でかつ高いスループットを実現するために、多項式予測計算の整数化と圧縮データの固定長変換回路の単純化を提案し、これらの方式に対する圧縮率の評価を行った。その結果、工夫を行う前の性能と殆ど変わらない圧縮が可能であり、単精度浮動小数点データの場合には3分の1程度のバンド幅圧縮が期待できることが明らかとなった。

次に、圧縮ハードウェアの要求仕様検討と基本設計を行った。対象とする問題では複数の変数を更新しながら計算を進めることから、独立した複数のデータストリームを圧縮し、それを単一の送信路に符号化する必要がある。このための符号化方式とハードウェアの基本設計を行い、回路面積と動作周波数を評価した。

2 年度目では、前年度に設計したバンド幅圧縮ハードウェアの改良と試作実装を行い、実機上で動作するシステムを構築して、高性能計算に対する本手法の性能評価の準備を行った。まず、対象とする数値流体計算にバンド幅圧縮ハードウェアを適用するために、流体シミュレーションのストリームハードウェアを設計し、FPGA 上に実装した。次に、そのストリームハードウェアの複数入出力チャンネルのそれぞれを並行に圧縮・展開するようなバンド幅圧縮ハードウェアを実装した。また、圧縮ハードウェア、流体計算ストリームハードウェア、DMA、性能計測用カウンタなどを制御するためのソフトウェアとして、圧縮制御ライブラリを開発した。これらを用いた実験により、バンド幅圧縮を用いる場合でも正しい流体計算が行えることを確認した。

最終年度では、前年度に準備したシステムを改良して、バンド幅圧縮ハードウェアの性能評価と、それを適用した場合の高性能計算アプリケーションの性能評価を行った。まず、これまでの設計と比べて回路面積と動作周波数を改善するために、圧縮後のデータ符号化方式を改良し、それを実現するハードウェアの設計と実装を行った。この新しい方式で

は、あるビット幅を単位として量子化を行う従来と異なり、予測ワードと実際のワードとの残差に3種類のビット幅のみを与える。例えば、4,8,32などである。この幅はパラメータ化されており、自由に選ぶことができる。このような粗い量子化により符号を短縮できる他、符号化ハードウェアを小さく出来る。特に予測精度が高く残差長が概ね短くなるようなデータの場合には、最小幅を適切に設定することにより圧縮率向上・回路面積縮小、動作周波数向上を実現できる。

計算に必要な複数変数のために複数チャンネル化した上で、以上の方式をFPGA上に実装した流体シミュレーションのストリーム計算ハードウェアへ適用した。その結果、計算途中データの圧縮により実効メモリ帯域が向上し、メモリ帯域が制約となる条件において1.7倍高い計算性能を達成することが確認された。また、複数のFPGAを相互に接続した拡張可能なシステムに対し、ネットワーク帯域向上の評価を行った。16個のFPGAを数珠つなぎにして実装する流体シミュレーションストリーム計算機に対して、提案するデータ圧縮ハードウェアはSFP+接続の10Gbps x 2の帯域を2.25倍に拡張でき、それに伴い計算性能の低下が抑えられる見通しが得られた。

以上要するに、提案する方式を実用的な計算問題に適用することにより、メモリやネットワークの実効帯域を拡張して計算性能を向上できることが実証できた。主な成果は以下の通りである。

- (1) ハードウェア実装向けの可逆データ圧縮アルゴリズム
- (2) 浮動小数点数値データ圧縮ハードウェアIPコア
- (3) 複数圧縮データストリームの符号化方式とハードウェアIPコア
- (4) FPGAによる数値流体計算ハードウェアおよびメモリバンド幅圧縮ハードウェアの試作実装と性能評価
- (5) 複数FPGA間のSFP+リンクに対するバンド幅圧縮の性能評価

これらの成果は、国際会議や国際論文誌において論文や講演の形で発表が行われた。今後の展望としては、流体計算以外のアプリケーションに対する応用と性能評価、浮動小数点以外のデータ形式への適用、大規模計算への応用などが挙げられる。

5. 主な発表論文等

〔雑誌論文〕(計21件)

Kentaro Sano and Satoru Yamamoto, "FPGA-based Scalable and Power-Efficient Fluid Simulation using Floating-Point DSP Blocks," *IEEE Transactions on Parallel and*

Distributed Systems (TPDS), 査読有, 印刷中, DOI: 10.1109/TPDS.2017.2691770, 2017.

Tomohiro Ueno, Kentaro Sano, and Satoru Yamamoto, "Memory Bandwidth Compressor for FPGA-based High-Performance Custom Stream Computation," *ACM Transactions on Reconfigurable Technology and Systems (TRETS)*, 査読有, vol.10, no.3, Article No.18, DOI:10.1145/3053688, 2017.

Kohei Nagasu, Kentaro Sano, Fumiya Kono, and Naohito Nakasato, "FPGA-based Tsunami Simulation: Performance Comparison with GPUs, and Roofline Model for Scalability Analysis," *Journal of Parallel and Distributed Computing*, 査読有, vol.106, pp.153-169, DOI: 10.1016/j.jpdc.2016.12.015, 2017.

Kentaro Sano, Tomohiro Ueno, Daichi Tanaka, and Satoru Yamamoto, "High-Performance Fluid Simulation using Multiple FPGAs with Bandwidth-Compressed Links," *Proceedings of Second International Workshop on Heterogeneous High-Performance Reconfigurable Computing (H2RC'16) in conjunction with SC16*, 査読有, URL: https://h2rc.cse.sc.edu/papers/paper_16.pdf, 2016.

Daichi Tanaka, Tomoya Ueno, Kentaro Sano, and Satoru Yamamoto, "Performance Evaluation of Inter-Cube Data Exchange in FPGA-based Fluid Simulation with Building Cube Method," *Proceedings of the International Symposium on Highly-Efficient Accelerators and Reconfigurable Technologies (HEART)*, 査読有, pp.57-60, 2016.

上野 知洋, 佐野 健太郎, 山本 悟, "予測残差長の量子化を用いた可逆データ圧縮ハードウェアの性能評価," *ハイパフォーマンスコンピューティングと計算科学シンポジウム(HPCS2016)論文集*, 査読有, pp.88-96, 2016.

Kohei Nagasu, Kentaro Sano, Fumiya Kono, and Naohito Nakasato, "Tsunami Simulation Accelerator Exploiting Fine and Coarse-Grain Parallelism with FPGA," *Proceedings of the*

International Conference on Parallel Computational Fluid Dynamics (ParCFD2016), 査読有, pp.60-61, 2016.

Tomohiro Ueno, Kentaro Sano, and Satoru Yamamoto, "Performance Improvement of FPGA-Based Stream Computing by Bandwidth Compressor," Proceeding of IEEE Symposium on Low-Power and High-Speed Chips (COOL Chips XIX), 査読有, poster paper#21, 2016.

Kohei Nagasu, Kentaro Sano, Fumiya Kono, Naohito Nakasato, "Performance and Power Evaluation of FPGA-based Tsunami Simulator using Floating-Point DSPs," Proceeding of IEEE Symposium on Low-Power and High-Speed Chips (COOL Chips XIX), 査読有, poster paper#17, 2016.

上野 知洋, 佐野 健太郎, 山本 悟, "複数FPGAによる並列流体計算と帯域圧縮通信による性能向上," 第30回数値流体力学シンポジウム講演論文集, 査読無, paper no. D10-4 (CDROM), 2016.

長州 航平, 佐野 健太郎, "浮動小数点DSPブロックを搭載したFPGAによる津波シミュレーションアクセラレータの性能評価," 第30回数値流体力学シンポジウム講演論文集, 査読無, paper no. D10-3 (CDROM), 2016.

長州 航平, 佐野 健太郎, "津波シミュレーション専用計算機のルーファインモデルと次世代FPGAの性能評価," 第214回システム・アーキテクチャ研究発表会 情報処理学会研究報告, 査読無, vol.2016-ARC-222, no.13, 2016.

長州 航平, 佐野 健太郎, 河野 郁也, 中里 直人, "ストリーム計算ハードウェアコンパイラSPGenを用いたFPGA津波シミュレータの開発," 電子情報通信学会リコンフィギャラブルシステム研究会 信学技法, 査読無, vol.115, no.400, pp.131-136, 2016.

Tomohiro Ueno, Kentaro Sano, and Satoru Yamamoto, "Area Efficient Memory Bandwidth Compressor for Stream Computation on FPGAs," Proceeding of IEEE Symposium on Low-Power and High-Speed Chips (COOL Chips XVIII), 査読有, poster#21, 2015.

長州 航平, 佐野 健太郎, 河野 郁也, 中

里 直人, "FPGAによる津波シミュレーション専用計算機の開発," 第29回数値流体力学シンポジウム講演論文集, 査読無, paper no. B09-3 (CDROM), 2015.

上野 知洋, 佐野 健太郎, 山本 悟, "メモリ帯域圧縮ハードウェアを用いた数値計算の高性能化," 第151回ハイパフォーマンスコンピューティング研究発表会 情報処理学会研究報告, 査読無, vol.HPC151, no.10, paper no.010, 2015.

佐野 健太郎, 河野 郁也, 中里直人, Alexander Vazhenin, Stanislav Sedukhin, "FPGAによる津波シミュレーションの専用ストリーム計算ハードウェアと性能評価," 第149回ハイパフォーマンスコンピューティング研究発表会 情報処理学会研究報告, 査読無, vol.HPC149, no.5, paper no.005, 2015.

上野 友也, 上野 知洋, 佐野 健太郎, 山本 悟, "ビルディングキューブ法に基づく流体専用計算機のキューブ間データ交換に関する一考察," 電子情報通信学会リコンフィギャラブルシステム研究会 信学技法, 査読無, vol.114, no.426, pp.79-84, 2015.

上野 知洋, 伊藤 涼, 佐野 健太郎, 山本 悟, "予測残差長の偏りを利用した小面積帯域圧縮ハードウェアの提案," 電子情報通信学会リコンフィギャラブルシステム研究会 信学技法, 査読無, vol.114, no.426, pp.73-78, 2015.

Tomohiro Ueno, Ryo Ito, Kentaro Sano, and Satoru Yamamoto, "Bandwidth Compression of Multiple Numerical Data-Streams for High Performance Custom Computing," Proceedings of the International Conference on Application-specific Systems, Architectures and Processors (ASAP), 査読有, pp.190-191, DOI: 10.1109/ASAP.2014.6868660, 2014.

21 上野 知洋, 伊藤 涼, 佐野 健太郎, 山本 悟, "複数ストリームのための帯域圧縮ハードウェアの実装と評価," 電子情報通信学会リコンフィギャラブルシステム研究会 信学技法, Vol.113, No.418, pp.131-136, Jan, 2014.

{学会発表}(計18件)

上野 知洋, "複数FPGAによる並列流体計算と帯域圧縮通信による性能向上," 第30

回数値流体力学シンポジウム, 2016年12月14日, タワーホール船橋(東京都).

長州 航平, "浮動小数点 DSP ブロックを搭載した FPGA による津波シミュレーションアクセラレータの性能評価," 第30回数値流体力学シンポジウム, 2016年12月14日, タワーホール船橋(東京都).

Kentaro Sano, "High-Performance Fluid Simulation using Multiple FPGAs with Bandwidth-Compressed Links," Second International Workshop on Heterogeneous High-Performance Reconfigurable Computing (H2RC'16) in conjunction with SC16, 2016年11月14日, ソルトレイク市(米国).

長州 航平, "津波シミュレーション専用計算機のルーラインモデルと次世代 FPGA の性能評価," 第214回システム・アーキテクチャ研究発表会 2016年10月6日, 幕張メッセ(千葉市).

Daichi Tanaka, "Performance Evaluation of Inter-Cube Data Exchange in FPGA-based Fluid Simulation with Building Cube Method," International Symposium on Highly-Efficient Accelerators and Reconfigurable Technologies (HEART), 2016年7月26日, 香港(中国).

上野 知洋, "予測残差長の量子化を用いた可逆データ圧縮ハードウェアの性能評価," 2016年7月7日, 仙台市東北大学.

Kohei Nagasu, "Tsunami Simulation Accelerator Exploiting Fine and Coarse-Grain Parallelism with FPGA," International Conference on Parallel Computational Fluid Dynamics (ParCFD2016), 2016年5月12日, 神戸国際会議場(神戸市).

Tomohiro Ueno "Performance Improvement of FPGA-Based Stream Computing by Bandwidth Compressor," IEEE Symposium on Low-Power and High-Speed Chips (COOL Chips XIX), 2016年4月21日, 情報文化センター(横浜市).

Kohei Nagasu, "Performance and Power Evaluation of FPGA-based Tsunami Simulator using Floating-Point DSPs," IEEE Symposium on Low-Power and High-Speed Chips (COOL Chips XIX), 2016年4月21日, 横浜市情報文化センター.

長州 航平, "ストリーム計算ハードウェアコンパイラ SPGen を用いた FPGA 津波シミュレータの開発," 電子情報通信学会リコンフィギャラブルシステム研究会, 2016年1月20日, 慶應義塾大学(横浜市).

長州 航平, "FPGA による津波シミュレーション専用計算機の開発," 第29回数値流体力学シンポジウム, 2015年12月17日, 九州大学(福岡市).

上野 知洋, "メモリ帯域圧縮ハードウェアを用いた数値計算の高性能化," 第151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9月30日, 沖縄産業支援センター(那覇市).

佐野 健太郎, "FPGA による津波シミュレーションの専用ストリーム計算ハードウェアと性能評価," 第149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6月26日, 工学院大学(東京都).

Tomohiro Ueno, "Area Efficient Memory Bandwidth Compressor for Stream Computation on FPGAs," IEEE Symposium on Low-Power and High-Speed Chips (COOL Chips XVIII), 2015年4月14日, 情報文化センター(横浜市).

上野 友也, "ビルディングキューブ法に基づく流体専用計算機のキューブ間データ交換に関する一考察," 電子情報通信学会リコンフィギャラブルシステム研究会, 2015年1月29日, 慶應義塾大学(横浜市).

上野 知洋, "予測残差長の偏りを利用した小面積帯域圧縮ハードウェアの提案," 電子情報通信学会リコンフィギャラブルシステム研究会, 2015年1月29日, 慶應義塾大学(横浜市).

Tomohiro Ueno, "Bandwidth Compression of Multiple Numerical Data-Streams for High Performance Custom Computing," International Conference on Application-specific Systems, Architectures and Processors (ASAP), 2014年6月19日, チューリッヒ(スイス).

上野 知洋, "複数ストリームのための帯域圧縮ハードウェアの実装と評価," 電子情報通信学会リコンフィギャラブルシステム研究会, 2014年1月27日, 慶應義塾大学(横浜市).

6 . 研究組織

(1)研究代表者

佐野 健太郎 (SANO, KENTARO)
東北大学・大学院情報科学研究科・准教授
研究者番号：00323048

(2)研究分担者

なし

(3)連携研究者

なし

(4)研究協力者

なし