

令和 6 年 6 月 17 日現在

機関番号：12605

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K11804

研究課題名（和文）IoT環境での機械学習に向けたベクトルレジスタ共有型SoC FPGAの実現

研究課題名（英文）Implementation of Shared Vector Register SoC FPGA for Machine Learning under IoT Environment

研究代表者

中條 拓伯（Nakajo, Hironori）

東京農工大学・工学（系）研究科（研究院）・教授

研究者番号：80217736

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究においては、RISC-Vのベクトル拡張を内部プロセッサとして実装し、プロセッサとアクセラレータ間においては、プロセッサ内のベクトルレジスタをデュアルポート化し、プロセッサがメモリからロードしたデータや演算処理を行ったディスティネーションレジスタを直接アクセラレータに高速に供給し、さらには演算結果を効率よく収集する方式SHAVER（SHARED VECTOR REGISTER）を提案し、設計・実装を行い、その有効性を検証した。AIアプリケーションとしてはディープニューラルネットワーク（DNN）で用いられる積和演算の高速化のためのアクセラレータをFPGA上での実装を行った。

研究成果の学術的意義や社会的意義

本研究は、高速なアクセラレータを有効に活用するために必要となる高速データ転送機構を提案したものであり、今後のアクセラレータへのデータ供給や結果の収集において有効なものとなり、特にAIアプリケーションにおいて学習や推論に要する処理時間を大幅に短縮できるものである。GPUを用いた高速化とは別に、FPGAによるハードウェアアクセラレーションにおいて有効となり、リコンフィギュラブルシステムにおいて学術的な意義は大きい。本機構によりエッジにおけるAI処理の高速化に寄与し、スマート農業や自動運転においても有効な手法であり、社会的意義は高いものである。

研究成果の概要（英文）：In this research, we have implemented the RISC-V vector extension as an internal processor. Between the processor and accelerator, we have built the dual-ported vector register in the processor to supply the data loaded by the processor from memory and the destination register where the calculation is executed directly to the accelerator at high speed. We also propose and have designed a method called SHARED VECTOR REGISTER (SHAVER) to efficiently collect the calculation results. We have verified its effectiveness. As an AI application, we have implemented an accelerator on an FPGA to speed up the convolution calculations used in deep neural networks (DNNs).

研究分野：計算機工学

キーワード：ハードウェアアクセラレーション RISC-V FPGA SoC ベクトル拡張 ベクトルレジスタ

1. 研究開始当初の背景

【IoT 技術と AI エッジコンピューティング】

IoT (Internet of Things) 技術によるディープニューラルネットワーク (DNN) などの人工知能では、サーバやクラウドにおける学習処理と、IoT デバイスに近いエッジ側における推論処理と切り離した形で実現されてきた。その推論処理をエッジ側で高速化するアクセラレーションに GPU を用いる手法に対し、消費電力の観点から書き換え可能 LSI である FPGA (Field Programmable Gate Array) を利用する研究が注目され、エッジ側の推論専用システムもベンチャにより開発が進められている。GPU、FPGA のいずれの場合もホスト CPU も必要となることから、エッジで利用するためにはコンパクト化が障害となっている。また、エッジ側で学習処理も行うことができれば、山間部の農地など、低バンド幅のネットワーク環境で画像、動画、音声などの膨大なデータをクラウドに転送する必要はなくなり、転送時間の削減と学習時間の大幅な短縮につながる。

【SoC FPGA の台頭とデータ転送ボトルネック】

FPGA の容量は年々増加し、動作速度も向上する中、Xilinx 社の Zynq や Altera 社を吸収した Intel 社の Agilix などといった、CPU と FPGA を 1 つのダイに搭載した SoC (System on Chip) が登場してきた。現在の SoC においては、データを外部メモリから SoC 内部メモリに DMA 転送し、そのデータを元に FPGA 内に実装した演算回路による計算処理を行い、内部メモリに結果を格納し、その後外部メモリに DMA 転送するといったメモリアクセスを中心としたデータ転送に依存している。しかしながら、FPGA 内部メモリのビット長や内部バスバンド幅がボトルネックとなり、ハードウェア的に膨大な演算器を並べても、効率よくデータを供給できないことから、ハードウェアによる高速化が十分に得られず、FPGA による高速化の実現例の多くは、省電力化のメリットしか優位性が得られていないのが実情である。重みと活性化関数の出力を二値化した CNN (Convolutional Neural Network) を FPGA で実装した研究もあるが、速度やメモリ容量の問題は解決できたものの、認識精度が低下する問題が生じている。

2. 研究の目的

本研究では、SoC FPGA 内における CPU と FPGA に実装された高速演算回路群との間で、その CPU の ベクトルレジスタを FIFO 化してデータ授受に利用し、CPU-アクセラレータ間の効率的なデータ転送を実現する方式を提案する。この提案方式により、今後の SoC FPGA におけるプロセッサとアクセラレータ間のデータ転送の新たな方式として装備し、学習処理、推論処理を同時にサポートする SoC FPGA プラットフォームとして位置づけられるものと考える。

3. 研究の方法

(1) デュアルポートレジスタの実装とレジスタ経由による遅延の評価

本研究では、RISC-V を内部プロセッサとして実装し、プロセッサとアクセラレータ間においては、従来のメモリを介したデータ転送ではなく、プロセッサ内のレジスタをデュアルポート化し、プロセッサがメモリからロードしたデータや演算処理を行ったディスティネーションレジスタを直接 FPGA に高速に供給し、さらには演算結果を効率よく収集する方式を設計・実装する。IoT デバイスからのセンサデータを想定して、画像データなどの大量データをアクセラレータに転送する時間を計測する。DMA 機能を用いて、外部メモリから内部メモリへのデータ転送に要する時間と比較し、レジスタを経由した場合のデータ転送性能の低遅延性を明らかにする。

(2) データ転送速度向上のためのベクトルレジスタ実装と転送速度の評価

次に、FPGA 高速計算のための 演算ロジックに接続するデータのビット幅を拡張する。そのためには、最近の汎用プロセッサに備わる ベクトルレジスタを利用する。ベクトルレジスタは複数ポートを持ち、SoC 内の CPU と、実装した高速演算ロジックの間で共有され、多数のデータを同時に演算ロジックに供給し、演算結果を一括して収集する方式を提案する。実装した改良 RISC-V を用いて、ベクトルレジスタを用いたデータ転送速度を計測し、(1) と同様に DMA 機能によるデータ転送時間と比較することで、提案方式の優位性を明らかにする。

(3) アクセラレータ機能の実装と機械学習パフォーマンスの評価

ディープニューラルネットワークにおける畳み込み演算において、ベクトルレジスタにより多数のデータを同時に供給できたとしても、そのベクトルレジスタにおいて同期を取る必要があり、アクセラレータによって読み出しが完了する前に次のデータで上書きすることはできず、

CPUは待機することとなる。そこで、ベクトルレジスタの一部をFIFO化し、CPUから連続してデータを書き込み、アクセラレータは、FIFO型レジスタがEmptyになるまで連続して読み出し、アクセラレーション処理を継続する機構を実装し評価する。

上記(1)～(3)においては、拡張RISC-Vプロセッサをベースとするgem5を用いたシミュレータおよび実機ボードにより検証する。その結果より、今後のSoC FPGAにおけるプロセッサとアクセラレータにおける高バンド幅のデータ共有機構の標準となるよう仕様を策定する。

4. 研究成果

本研究においては、(1)デュアルポートレジスタの実装とレジスタ経由による遅延の評価、(2)データ転送速度向上のためのベクトルレジスタ実装と転送速度の評価、(3)アクセラレータ機能の実装と機械学習パフォーマンスの評価、の3つの研究項目を挙げている。そのために、RISC-Vを内部プロセッサとして実装し、プロセッサとアクセラレータ間においては、プロセッサ内のレジスタをデュアルポート化し、プロセッサがメモリからロードしたデータや演算処理を行ったディスティネーションレジスタを直接FPGAに高速に供給し、さらには演算結果を効率よく収集する方式を設計・実装する。

初年度にあたる2021年度では、上記研究項目の中の(1)において、RISC-Vのマイクロアーキテクチャ設計を中心に概念設計、詳細設計を行い、命令を限定した形でFPGA実装を行った。

(2)については、RISC-VのFPGA実装の設計・実装を行い、動作確認を行った。シミュレータについても開発を進め、LLVMコンパイラの出力する実行コードが動作することを確認した。

評価に用いるAIアプリケーションとしてはディープニューラルネットワーク(DNN)をターゲットとしており、演算処理を簡略化するアルゴリズムを考案したり、ノイズを含む学習データから効率的かつ認識精度を維持した学習アルゴリズムを考案し、その成果はそれぞれ英語論文としてまとめて発表し、採録された。

RISC-Vにおいて同時マルチスレッド実行機構を付加したプロセッサを実装し評価した結果を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において発表し、ベクトル実装についても、2022年6月開催の同研究会において発表を行った。

ベースとなるRISC-VプロセッサとSMT機能の融合を図り、その成果を合同研究発表会(ETNET2024)において発表した。

さらに、SMT機構を付加したRISC-VプロセッサにP拡張を組み込んだ新たなアーキテクチャについて国際会議に投稿し、RISC-Vにメッセージパッシング機構を行う拡張命令を提案、実装し、スレッド間通信を高速に行う機構を提案し、2024年6月開催のHotSPAにおいて発表した。

5. 主な発表論文等

〔雑誌論文〕 計9件（うち査読付論文 4件/うち国際共著 2件/うちオープンアクセス 3件）

1. 著者名 田中友章, 東良輔, 田中清史, 長名保範, 三好健文, 多田十兵衛, 中條拓伯	4. 巻 vol. 122, no. 60, RECONF2022-5
2. 論文標題 ハードウェアアクセラレーションのためのベクトルレジスタ共有機構	5. 発行年 2022年
3. 雑誌名 信学技報 RECONF	6. 最初と最後の頁 26-31
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 -

1. 著者名 Hidetaro Tanaka, Tomoaki Tanaka, Ryosuke Higashi, Tsutomu Sekibe, Shuichi Takada, and Hironori Nakajo	4. 巻 -
2. 論文標題 Implementation of a RISC- V SMT Core in an AI processor	5. 発行年 2022年
3. 雑誌名 Proceedings of the 11th International Symposium on Information and Communication Technology (SoICT 2022)	6. 最初と最後の頁 15-22
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3568562.3568634	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 東良輔, 田中友章, 田中秀太郎, 加藤倫也, 田中清史, 長名保範, 三好健文, 多田十兵衛, 中條拓伯	4. 巻 2023-ARC-252
2. 論文標題 シミュレーションによるベクトルレジスタ共有機構の評価	5. 発行年 2023年
3. 雑誌名 情処研報 システム・アーキテクチャ (ARC)	6. 最初と最後の頁 1-7
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 -

1. 著者名 Cai Jingyong, Takemoto Masashi, Qiu Yuming, Nakajo Hironori	4. 巻 11
2. 論文標題 Trigonometric Inference Providing Learning in Deep Neural Networks	5. 発行年 2021年
3. 雑誌名 Applied Sciences	6. 最初と最後の頁 6704 ~ 6704
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/app11156704	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 Nguyen Quoc, Shikina Tomoaki, Teruya Daichi, Hotta Seiji, Han Huy-Dung, Nakajo Hironori	4. 巻 11
2. 論文標題 Leveraging Expert Knowledge for Label Noise Mitigation in Machine Learning	5. 発行年 2021年
3. 雑誌名 Applied Sciences	6. 最初と最後の頁 11040 ~ 11040
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/app112211040	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 田中秀太郎, 田中友章, 長岡慶太, 東良輔, 関部勉, 高田周一, 中條拓伯	4. 巻 121
2. 論文標題 仮想エンジンアーキテクチャにおけるRISC-V同時マルチスレッディング(SMT)コアの実現	5. 発行年 2022年
3. 雑誌名 信学技報 RECONF2021-65	6. 最初と最後の頁 43 ~ 48
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 -

1. 著者名 前田依莉子, 照屋大地, 中條拓伯	4. 巻 121
2. 論文標題 再構成可能仮想アクセラレータ(ReVA)の実現に向けたキャッシュコヒーレントな相互接続規格の検討	5. 発行年 2022年
3. 雑誌名 信学技報 RECONF2021-65	6. 最初と最後の頁 132 ~ 137
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 -

1. 著者名 田中友章, 東良輔, 田中清史, 長名保範, 三好健文, 多田十兵衛, 中條拓伯	4. 巻 -
2. 論文標題 ハードウェアアクセラレーションのためのベクトルレジスタ共有機構	5. 発行年 2022年
3. 雑誌名 信学技報	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 -

1. 著者名 Tomoaki Tanaka, Ryosuke Higashi, Hidetaro Tanaka, Takefumi Miyoshi, Yasunori Osana, Jubee Tada, Kiyofumi Tanaka, Hironori Nakajo	4. 巻 -
2. 論文標題 Shared Vector Register of RISC-V for the Future Hardware Acceleration	5. 発行年 2022年
3. 雑誌名 Sixth Workshop on Computer Architecture Research with RISC-V (CARRV 2022)	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 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 -

〔学会発表〕 計7件 (うち招待講演 0件 / うち国際学会 2件)

1. 発表者名 田中友章
2. 発表標題 ハードウェアアクセラレーションのためのベクトルレジスタ共有機構
3. 学会等名 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 (RECONF)
4. 発表年 2022年

1. 発表者名 Hidetaro Tanaka
2. 発表標題 Implementation of a RISC- V SMT Core in an AI processor
3. 学会等名 The 11th International Symposium on Information and Communication Technology (SoICT 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 東良輔
2. 発表標題 シミュレーションによるベクトルレジスタ共有機構の評価
3. 学会等名 情報処理学会 システム・アーキテクチャ研究会 (ARC)
4. 発表年 2023年

1. 発表者名 田中秀太郎, 田中友章, 長岡慶太, 東良輔, 関部勉, 高田周一, 中條拓伯
2. 発表標題 仮想エンジンアーキテクチャにおけるRISC-V同時マルチスレッディング(SMT)コアの実現
3. 学会等名 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
4. 発表年 2022年

1. 発表者名 前田依莉子, 照屋大地, 中條拓伯
2. 発表標題 再構成可能仮想アクセラレータ(ReVA)の実現に向けたキャッシュコヒーレントな相互接続規格の検討
3. 学会等名 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
4. 発表年 2022年

1. 発表者名 田中友章, 東良輔, 田中清史, 長名保範, 三好健文, 多田十兵衛, 中條拓伯
2. 発表標題 ハードウェアアクセラレーションのためのベクトルレジスタ共有機構
3. 学会等名 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
4. 発表年 2022年

1. 発表者名 Tomoaki Tanaka, Ryosuke Higashi, Hidetaro Tanaka, Takefumi Miyoshi, Yasunori Osana, Jubee Tada, Kiyofumi Tanaka, Hironori Nakajo
2. 発表標題 Shared Vector Register of RISC-V for the Future Hardware Acceleration
3. 学会等名 Sixth Workshop on Computer Architecture Research with RISC-V (CARRV 2022) (国際学会)
4. 発表年 2022年

〔図書〕 計1件

1. 著者名 サラ・L・ハリス (著), デイビッド・ハリス (著), 天野英晴, 鈴木貢, 中條拓伯, 永松礼夫 (翻訳)	4. 発行年 2022年
2. 出版社 星雲社	5. 総ページ数 334
3. 書名 デジタル回路設計とコンピュータアーキテクチャ [RISC-V版]	

〔産業財産権〕

〔その他〕

東京農工大学 中條研究室 ホームページ https://www.nj.cs.tuat.ac.jp/
--

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
---------	---------