

令和 3 年 6 月 21 日現在

機関番号：21602

研究種目：基盤研究(C) (一般)

研究期間：2018～2020

課題番号：18K11221

研究課題名(和文) 低消費エネルギーな非同期式畳み込みニューラルネットワーク回路のFPGA実装

研究課題名(英文) FPGA implementation of low energy asynchronous convolutional neural network circuits

研究代表者

齋藤 寛 (Saito, Hiroshi)

会津大学・コンピュータ理工学部・上級准教授

研究者番号：50361671

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：本研究では、Field Programmable Gate Array (FPGA)を対象に、画像分類を行う畳み込みニューラルネットワークを低消費エネルギーな非同期式回路として実現し、深層機械学習に対する非同期式回路の有用性を明らかにすることを目的に研究を行った。また、量子化した二値化ニューラルネットワーク回路も非同期式回路として設計した。二値化ニューラルネットワークに関しては、同期式回路と比較して消費エネルギーを最もよい場合で約半分にすることができた。この他に、性能改善を目的に、配置制約を用いて非同期式回路をFPGAに設計するための設計手法を提案した。

研究成果の学術的意義や社会的意義

深層学習を実現するにあたり、GPUを用いた場合、電力消費やコストが大きくなる。一方、CPUを用いた場合、性能が問題となる。こうした問題に対し、深層学習を行う回路をFPGAに実現する手法が提案されている。しかし、こうした回路は、クロック信号を基に回路全体を制御する同期式回路として実現されているため、クロック周りの消費電力が大きい。本研究で実現した非同期式回路によって、消費エネルギーを削減することで、電力要求が厳しいアプリケーションでの使用が期待できる。

研究成果の概要(英文)：In this research, to clarify the usefulness of asynchronous circuits for deep machine learning, we designed low energy convolutional neural network circuits for image classification on Field Programmable Gate Arrays (FPGAs) as asynchronous circuits. Also, we designed quantized binarized neural network circuits as asynchronous circuits. Compared to synchronous counterparts, the energy consumption of the binarized neural network circuits was reduced to half in the best case. Furthermore, to improve performance, we proposed a design method to implement asynchronous circuits on FPGAs using placement constraints.

研究分野：計算機システム

キーワード：非同期式回路 FPGA 畳み込みニューラルネットワーク

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

画像認識は、深層機械学習を用いることで認識精度の向上を図っている。画像認識ではリアルタイム性が求められるため、撮影した画像の推論を利用者に近いエッジ側で行うことが注目されている。こうしたことを実現するエッジ向け Internet-of-Things (IoT) デバイスは、電気が確保できない場所での使用も見込まれるため、低消費エネルギー化が課題となる。

深層機械学習では、畳み込みニューラルネットワークが用いられる。この計算をマイクロプロセッサで行うと、処理時間が長い、電力効率が悪いといったことから、Graphics Processing Unit (GPU) や Field Programmable Gate Array (FPGA) を用いた研究が盛んに行われている。特に FPGA は、用途に応じて回路構成を変更することができるため、用途毎に性能や電力効率の面で最適なものが実現できる。

畳み込みニューラルネットワーク回路を FPGA に実現するために様々な研究が行われているが、これらはクロック信号にて回路全体を制御する同期式回路として実現されている。しかし、高周波なクロック信号を広い範囲で分配すると、クロック信号における消費電力が問題となる。同期式回路とは異なり、ローカルなハンドシェーク信号やタイミング信号にて必要な回路部品を制御する非同期式回路は、低消費電力な回路を実現することができる。しかしながら、市販の FPGA は同期式回路を想定しているため、非同期式回路を実現するためには様々な工夫が求められる。

### 2. 研究の目的

本研究では、推論を行うエッジ向け IoT デバイスを対象とした、低消費エネルギーな非同期式畳み込みニューラルネットワーク回路を FPGA 上に実現する。実現した回路に対して、消費エネルギーを評価し、同期式回路と比較することで、非同期式回路の有効性を明らかにする。

### 3. 研究の方法

低消費エネルギーな非同期式畳み込みニューラルネットワーク回路を FPGA に実現するにあたって、以下の方法で研究を行った。

#### (1) パイプライン化された非同期式回路をFPGAに実現するための設計支援環境の構築

市販のFPGAは、同期式回路を想定している。設計支援ツールも同様で、非同期式回路に特有な設計制約の生成、タイミング検証、タイミング違反時の遅延調整は行わない。そのため申請者は、[1]をベースに、非同期式回路を市販FPGA上に実現するための設計支援環境を開発してきた。開発した設計支援環境は、パイプライン回路に対応できていない。畳み込みニューラルネットワークのFPGA実装は、スループット向上のために、パイプライン回路として実現するのが一般的である。そのため、これまでに開発した設計支援環境を拡張し、パイプライン回路を扱えるようにした。

#### (2) 非同期式畳み込みニューラルネットワーク回路の設計と評価

本項目では、FPGAを対象に同期式回路として実現された畳み込みニューラルネットワーク回路を基に、非同期式畳み込みニューラルネットワーク回路を設計し評価した。畳み込みニューラルネットワークモデルは、手書き数字を分類するLeNetを用いた。初めに、TensorFlowやPythonを用いて実現したニューラルネットワークモデルから、同期式回路を設計した。次に、[2]の手法を基に、同期式回路モデルから束データ方式による非同期式回路モデルを生成した。束データ

方式による非同期式回路では、データパス回路は同期式回路と同じものを用いる。制御回路は、パイプラインステージ毎に非同期式制御モジュールを割り当てることで実現する。図1は、生成した非同期式回路モデルを表す。最後に、生成した非同期式回路モデル

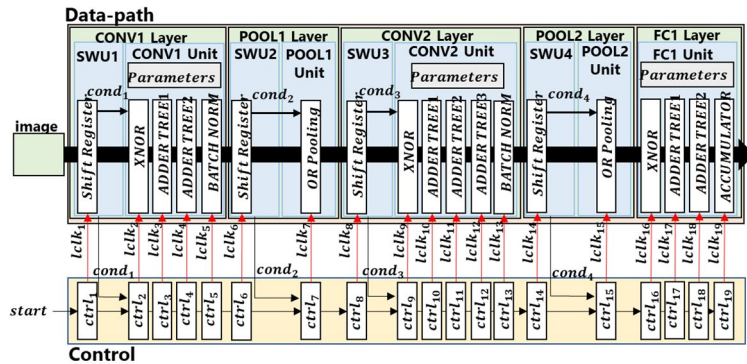


図 1 非同期式畳み込みニューラルネットワーク回路モデル

に対し、FPGAベンダが提供する設計支援ツールと(1)で実現した設計支援環境を用いることで、FPGA上に非同期式畳み込みニューラルネットワークを実現した。実現後、実行時間、回路面積、消費電力、および消費エネルギーを評価し、同期式畳み込みニューラルネットワーク回路と比較することで、非同期式回路の効果を確認した。

### (3) 非同期式二値化畳み込みニューラルネットワーク回路の設計と評価

本項目では、畳み込みニューラルネットワークの重みや活性化関数の出力を二値で表現した、二値化畳み込みニューラルネットワークを非同期式回路として設計し評価した。設計手法、および評価手法は(2)と同じ手法を用いている。

### (4) 配置制約による性能最適化

(3)で設計した非同期式二値化畳み込みニューラルネットワーク回路は、同期式回路と比べ消費電力は大幅に改善できたが、性能は同期式回路より悪化している。市販のFPGA設計支援ツールは、同期式回路を想定しており、非同期式回路に最適な配置配線を実現することができないためである。その結果、非同期式回路による消費エネルギーの削減効果が小さくなる。

本項目では、非同期式回路の性能を改善するため、配置制約を用いて設計を行った。配置制約は、以下の2つの方針で生成した。

制御論理を FPGA の同一論理ブロック内に配置する。この方法によって、制御論理内の配線遅延を短くする。

申請者らが考案した[3]のフロアプラン手法を、パイプライン回路向けに拡張する。非同期式回路の性能を最適化しようフロアプランを行い、リソース間の配線遅延を短くする。また、畳み込みニューラルネットワークの層やチャンネル毎に配置制約を生成し、消費エネルギーの面で効果があるものを探索する。

## 4. 研究成果

3で述べた研究の方法(1)から(4)までの成果は以下のとおりである。

### (1) パイプライン化された非同期式回路をFPGAに実現するための設計支援環境の構築

パイプライン制御モデルを検討し、[1]で開発した設計支援ツールセットでパイプライン回路を扱えるよう拡張した。拡張にて、非同期式パイプライン回路に対する制約生成、タイミング検証、及び遅延調整を自動で行うことが可能となった。

### (2) 非同期式畳み込みニューラルネットワーク回路の設計と評価

同期式回路として設計したLeNetより、非同期式回路を設計した。同期式回路は、タイミング

違反がなく、かつクロックサイクルタイムが最速となるものを用いた。設計した非同期式畳み込みニューラルネットワーク回路は、同期式回路とほぼ同じ性能を達成し、動的消費電力を2.4%削減することができた。そのため、実行時間と動的消費電力の積から求めた消費エネルギーは2.3%の削減となった。この成果は、[4]として口頭発表を行った。

### (3) 非同期式二値化畳み込みニューラルネットワーク回路の設計と評価

(2)と同様に、同期式回路として設計した二値化畳み込みニューラルネットワークから非同期式回路を設計した。LeNetやVGG-16を対象に設計を行ったが、最も良いケースで動的消費電力を60.5%削減することができた。しかし、1枚の画像分類に要する時間は21.5%増加したため、消費エネルギーは52%の削減に留まった。この成果は、[5]にまとめた。なお、同期式回路モデルから非同期式回路モデルに変換する手法を提案し、[6]で発表した。

### (4) 配置制約による性能最適化

性能改善を目的に、非同期式回路の制御回路やレジスタに対して配置制約を与え合成する手法を提案した。この成果は[7]で発表した。

## 参考文献

- [1] K. Takizawa et.al., “A Design Support Tool Set for Asynchronous Circuits with Bundled-data Implementation on FPGAs”, Proc. FPL 2014, pp.1-4, 2014.
- [2] J. Furushima et.al., “Design of an Asynchronous Processor with Bundled-data Implementation on a Commercial Field Programmable Gate Array”, Informatica 40 (2016), Vol. 40 Issue 4, pp.399-408, 2016.
- [3] H. Saito and N. Hamada, “A floorplan method for asynchronous circuits with bundled-data implementation on FPGAs”, Proc. ISCAS 2010, pp.925-928, 2010.
- [4] H. Kato and H. Saito, “Design of Asynchronous CNN Circuits on Commercial FPGA from Synchronous CNN Circuits”, Proc. MCSoc 2019, pp.61-67, 2019.
- [5] H. Kato, “Binarized Neural Network Circuits on FPGAs Based on Asynchronous Circuits with Bundled-data Implementation”, Master Thesis, The University of Aizu, 2021.
- [6] S. Semba et.al., “Optimization Methods During RTL Conversion from Synchronous RTL Models to Asynchronous RTL Models”, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E103.A, No.12, pp.1417-1426, Dec. 2020.
- [7] T. Otake and H. Saito, “A Design Method for Designing Asynchronous Circuits on Commercial FPGAs Using Placement Constraints”, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E103.A, No.12, pp.1427-1436, Dec. 2020.

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 S. Semba and H. Saito	4. 巻 E102-A, Issue 7
2. 論文標題 Conversion from Synchronous RTL Models to Asynchronous RTL Models	5. 発行年 2019年
3. 雑誌名 IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences	6. 最初と最後の頁 904--913
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transfun.E102.A.904	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 OTAKE Tatsuki, SAITO Hiroshi	4. 巻 E103.A
2. 論文標題 A Design Method for Designing Asynchronous Circuits on Commercial FPGAs Using Placement Constraints	5. 発行年 2020年
3. 雑誌名 IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences	6. 最初と最後の頁 1427 ~ 1436
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transfun.2020VLP0006	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 SEMBA Shogo, SAITO Hiroshi, TATSUOKA Masato, FUJIMURA Katsuya	4. 巻 E103.A
2. 論文標題 Optimization Methods during RTL Conversion from Synchronous RTL Models to Asynchronous RTL Models	5. 発行年 2020年
3. 雑誌名 IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences	6. 最初と最後の頁 1417 ~ 1426
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transfun.2020VLP0004	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6件（うち招待講演 0件/うち国際学会 6件）

1. 発表者名 S. Semba and H. Saito
2. 発表標題 Comparison of RTL Conversion and GL Conversion from Synchronous Circuits to Asynchronous Circuits
3. 学会等名 IEEE International Symposium on Circuits and Systems (国際学会)
4. 発表年 2019年

1 . 発表者名 S. Semba and H. Saito
2 . 発表標題 A Study on the Optimization of Asynchronous Circuits During RTL Conversion from Synchronous Circuits
3 . 学会等名 Workshop on Synthesis And System Integration of Mixed Information technologies ( 国際学会 )
4 . 発表年 2019年

1 . 発表者名 T. Otake and H. Saito
2 . 発表標題 Design of Asynchronous Circuits on Commercial FPGAs Using Placement Constraints
3 . 学会等名 Workshop on Synthesis And System Integration of Mixed Information technologies ( 国際学会 )
4 . 発表年 2019年

1 . 発表者名 H. Kato and H. Saito
2 . 発表標題 Design of Asynchronous CNN Circuits on Commercial FPGA from Synchronous CNN Circuits
3 . 学会等名 IEEE International Symposium on Embedded Multicore/Many-core Systems-on-Chip ( 国際学会 )
4 . 発表年 2019年

1 . 発表者名 T. Kudo and H. Saito
2 . 発表標題 Comparison of Pipelined Asynchronous Circuits Designed for FPGA
3 . 学会等名 3rd International Conference on Applications in Information Technology ( 国際学会 )
4 . 発表年 2018年

1. 発表者名 S. Semba and H. Saito
2. 発表標題 Study on an RTL Conversion Method from Pipelined Synchronous RTL Models into Asynchronous RTL Models
3. 学会等名 Workshop on Synthesis And System Integration of Mixed Information technologies (国際学会)
4. 発表年 2021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	富岡 洋一  (Tomioka Yoichi)  (10574072)	会津大学・コンピュータ理工学部・上級准教授    (21602)	

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