

令和 2 年 6 月 12 日現在

機関番号：13903

研究種目：挑戦的研究（萌芽）

研究期間：2017～2019

課題番号：17K19971

研究課題名（和文）精度の自動調整機能を備えるApproximate Computing基盤

研究課題名（英文）An Accuracy-Adaptive Approximate Computing Platform

研究代表者

津邑 公暁（TSUMURA, Tomoaki）

名古屋工業大学・工学（系）研究科（研究院）・教授

研究者番号：00335233

交付決定額（研究期間全体）：（直接経費） 4,800,000円

研究成果の概要（和文）：本研究では、Approximate Computing（近似計算）のための統一的な計算基盤の実現を目指した。画像処理や機械学習といった、大規模データを入力とする処理に対する効果を確認した。特に畳み込みニューラルネットワーク（CNN）に対しては、大幅な計算量の削減が期待できることを示した。また計算基盤のハードウェアについても、その構成および実装の検討を行った。スーパースcalarプロセッサをベースとした設計を行い、その速度および消費電力について評価した。また、消費電力の削減手法についても設計し、シミュレーションによりその効果を確認した。

研究成果の学術的意義や社会的意義

本申請研究は、Approximate Computing（近似計算）の統一的適用手法の発見、および、可用性の高い近似計算基盤の実現を目指して研究を行った。今後の発展が期待されている科学技術分野は、その多くが、ビッグデータマイニング、機械学習、コンピュータビジョンなど、大規模データの高速度処理が必要とされるものであることから、計算量自体を削減することに対する要求は大きく、これまで検討すらされていなかったこの「統一的な近似計算基盤」に関して検討を行ったことは、広きにわたる学術分野・産業の発展につながる事が期待できる。

研究成果の概要（英文）：In this study, we aimed to design a common platform for Approximate Computing. We confirmed the availability of this platform on large-scale data processing, such as image processing and machine learning. Especially for convolutional neural networks (CNNs), we show that the computational complexity can be significantly reduced with our method. In addition, we also studied the configuration and implementation of the hardware of the computing platform. We designed it based on a super-scalar processor and evaluated its speed and power consumption. In addition, we proposed a method to reduce power consumption of the platform and confirmed its efficacy by simulation.

研究分野：プロセッサアーキテクチャ

キーワード：計算再利用 Approximate Computing 近似計算 畳み込みニューラルネットワーク 機械学習

1. 研究開始当初の背景

情報システムで扱うべきデジタルデータは、画像や音声など旧来のメディアデータはもちろんのこと、RFID やセンサにより収集されるデータ、ソーシャルメディアを介して発信されるデータなどが加わり、爆発的な増大を見せている。今後も、いわゆる「超可観測」が進行することで、扱うべきデータ量はさらに超大規模化していくと考えられる。これに伴い、計算性能の向上はもちろんのこと、その計算形態自体の再考が迫られている。

特にこのような大規模データを対象とする代表的アプリケーションである、信号処理やマイニング、機械学習などにおいては、個々の計算結果の高信頼性と高精度を必ずしも強く要求しない場合が多く、必要最低限の精度で、いかに早く結果を出力するか、あるいはより多くの演算を行うかが重要となる。

このような背景から、出力の精度に重大な影響を及ぼさない範囲で多少の誤差を許容しつつ、適切に近似的な計算を行うことで処理を高速化・省電力化する、Approximate Computing というパラダイムが期待されている[a]。しかし、計算近似化が出力に与える影響や、許容される出力誤差量は、各アプリケーションの性質に強く依存するため、開発者は個別のアプリケーションに対して試行錯誤しながら計算近似度を調整する必要がある、このチューニングの困難さと生産性の低さが大きな障害となっていた。

[a] R. Nair: Technical Perspective: Big Data Needs Approximate Computing, Communications of the ACM, Vol.58, No.1, p.104 (2015)

2. 研究の目的

そこで本研究では、プログラマが出力誤差の許容度を指定するだけで、動的に計算近似度を調整し、指定された誤差の範囲内で最大のパフォーマンスを引き出す、可用性の高い Approximate Computing 計算基盤の実現を目指す。この計算基盤は、以下の2つから構成される。

【ソフトウェア】要求出力精度および計算近似度を容易に指定可能なプログラミングフレームワーク

【ハードウェア】Approximate Computing エンジン、およびそれを利用するための拡張命令セットを備えるプロセッサ、およびそのFPGA実装

今後の発展によりさまざまな応用が期待されているビッグデータマイニングや機械学習など、大規模データの高速度処理を必要とする多くの科学技術分野に対し、この計算基盤を提供することで、幅広い学術分野・産業の発展促進に寄与することが、本申請研究の最終目的である。

3. 研究の方法

研究代表者は、計算再利用の概念が Approximate Computing. と親和性が高いことを発見し、本研究では、この計算再利用をベースとした、可用性の高い Approximate Computing. 計算基盤を実現することで、幅広いアプリケーションに対して Approximate Computing. の適用およびチューニングを容易かつ効果的に行える環境を実現する。

既存の計算基盤を、本研究により実現する Approximate Computing. 計算基盤で置き換えることで、大規模データの処理を必要とするあらゆるアプリケーションの高速化・省電力化・生産性向上に資することができる。特に今後の発展によりさまざまな応用が期待されている、大規模データマイニングや機械学習などは、Approximate Computing. を必要としている分野であり、これらの発展に資することで、さまざまな科学技術分野・産業の発展を促進することが期待できる。本研究で実現する計算基盤は以下のように段階的に拡張実装する。

- i. Approximate Computing. 対象区間および計算近似度をプログラマに指定させる実装
- ii. 計算近似度ではなく出力誤差の許容度をプログラマに指定させ、計算近似度を自動的に調整する実装
- iii. Approximate Computing. が有効となる対象区間も自動的に抽出し、プログラマの指示なしに半自動的に Approximate Computing. を適用可能な実装

まず初年度では、Approximate Computing. 対象区間およびその区間に適用する計算近似度をプログラマに指定させるためのディレクティブを含む記述仕様を設計し、それを解釈・実行するコンパイラおよびプロセッサを設計・実装する(i)。連想検索エンジンについては、最大性能の確認のため、まずはハードウェアでの実装を検討し、既存のプロセッサシミュレータ上に設計・実装する。また、そのハードウェアエンジンを利用するための拡張命令セットを設計する。

Approximate Computing. 対象として指定できる処理区間は、まずはプログラム中の関数とし、プログラマは必要に応じて、対象区間を関数として切り出して定義した上で、その関数に対してディレクティブを付加する。ディレクティブには、解析が容易なプラグマ形式を採用する。その上で、プラグマで指定されたパラ

メータ(入力近似度)を解釈し、拡張命令へと変換するコンパイラを設計・実装する。

4. 研究成果

計算再利用のアクセラレーション機構を備えるプロセッサ(以下,自動メモ化プロセッサと呼ぶ)に関するさまざまな検討を行った。

(1)

自動メモ化プロセッサが再利用の対象区間と見なす関数は,一般にプログラム中の様々なアドレスから呼びだされる。そのため,関数が呼びだされるまでの動作により,同一の関数でも入力の特徴や周期性などの特徴が異なると考えられる。そこで,関数の復帰アドレスの違いが,関数の再利用に対しどのような影響を及ぼすかについて調査した。そしてその結果に基づき,自動メモ化プロセッサの性能向上に向けた手法を検討した。

自動メモ化プロセッサが再利用の対象区間と見なす関数は,復帰アドレスによって入力の規則性や周期性などの特徴が異なるであろうという予測に基づき,復帰アドレス別の再利用率について調査した結果,予測通り復帰アドレスによって再利用率が大きく異なることが確認できた。また,復帰アドレスごとに計算再利用の適否を判定し,再利用成功率が高いと判定される復帰アドレスからの呼出時に限って再利用テストを行うことでより多くの実行サイクル数が削減できることを確認した。一方で,再利用成功率が低い復帰アドレスからの呼出しを計算再利用の対象から早期に除外することで,再利用テスト失敗に伴う再利用表検索オーバーヘッドを抑制できることも確認した。(文献)

(2)

画像認識において,Convolutional Neural Network (CNN) と呼ばれるニューラルネットワークが高い認識精度を示し,顔認識や文字認識,自動運転における歩行者認識などのアプリケーションに広く利用されている。一方で,CNNの学習には膨大な計算資源と時間コストが必要なことが知られており,この計算コストの大きさが問題となっている。近年では認識精度を向上させるために,CNNの規模を増大させる傾向にあり,これに伴って,必要な計算時間や消費エネルギーも更に増大を続けている。この問題に対し,CNNに含まれる計算を近似することで計算量を削減する研究が盛んに行われている。しかし,近似方法によっては元のデータが持つ情報が大きく損なわれ,CNNの認識精度が大きく低下してしまう。このような認識精度の低下は,アプリケーションによっては無視できない場合も多い。そこで本研究では,CNNの認識精度の低下を抑えつつ,学習に要する時間,および,推論に要する時間の両方を削減する方法について検討する。学習時間削減のために,カーネルをテンプレート化するという新しい視点を提案し,この効果を評価する。また推論時間削減のために,CNN計算に対する計算再利用の親和性について議論・評価した。

提案したカーネルテンプレート化により,認識精度を損うことなく,学習対象となるパラメータ数を大きく削減できることを確認した。また,CNN計算における乗算については,繰り返し出現する乗数・被乗数の組合せが数多く存在することを確認し,計算再利用がCNN計算に対して高い親和性を持つことが確認できた。更に,カーネルテンプレート化の適用により,より多くの乗数・被乗数の組合せが繰り返し出現することが確認でき,CNNが,計算再利用に基づいた近似計算が非常に有効に働くアプリケーションであることが確認できた。(文献)

(3)

自動メモ化プロセッサは,実行自体を一部省略することから,実行ユニットで消費される動的電力が低減するため,このアイデアはプロセッサの消費電力削減にも効果がある可能性がある。そこで本稿では,自動メモ化プロセッサの詳細な電力評価を行い,計算再利用が持つ,プロセッサ省電力化に対するポテンシャルについて検証した。その上で,より電力効率を向上させるために,メモ化専用ユニットの消費エネルギー削減の方法についても検討した。

消費エネルギー評価した結果,計算再利用のための専用ユニットが新たに追加されているにもかかわらず,一部のプログラムにおいて総消費エネルギーの減少が確認できた。専用ユニットの中でも,CAMでの構成を想定しているInTblでの消費電力が大きいことを確認した。この結果を受け,InTblの一部をRAMで構成する手法について検討し,予備評価によりその手法の有効性を確認した。(文献)

<引用文献>

松山且樹, 藤井政圭, 津邑公暁, 中島康彦: “自動メモ化プロセッサにおける復帰アドレス別の再利用率調査とその応用”, 情処研報, Vol.2017-ARC-227, No.19, pp.1-6 (Jul. 2017)

井内悠太, 津邑公暁: “カーネルテンプレート化と計算再利用によるCNNの計算量削減に関する検討”, 情処研報, Vol.2018-ARC-232, No.11, pp.1-7 (Jul. 2018)

Yuta INOUCHI, Hayato YAMAKI, Shinobu MIWA, Tomoaki TSUMURA: “Functionally-Predefined Kernel: a Way to Reduce CNN Computation”, Proc. IEEE Pacific Rim Conf. on Communications, Computers and Signal Processing (PacRim 2019), 6 pages (Aug. 2019) **received Best Paper Award for Computers Track (1/27 = 3.7%)**

武石 隆太郎, 津邑公暁, 中島康彦: “計算再利用によるプロセッサ省電力化の検討”, 情報研報 (SWoPP2019), Vol.2019-ARC-237, No.6, pp.1-7 (Jul. 2019)

武石 隆太郎, 津邑公暁, 中島康彦: “入力検索機構の改良による自動メモ化プロセッサの省電力化”, The 4th cross-disciplinary Workshop on Computing Systems, Infrastructures, and Programming (xSIG 2020), «the paper is available only to the attendees» (Jul. 2020) **Best Master's Student Award 受賞**

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 6件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 KAWAMURA Shinji, TSUMURA Tomoaki	4. 巻 E101.D
2. 論文標題 Hardware Accelerated Marking for Mark & Sweep Garbage Collection	5. 発行年 2018年
3. 雑誌名 IEICE Transactions on Information and Systems	6. 最初と最後の頁 1107-1115
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transinf.2017EDP7163	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 Futamase Yuki, Hayashi Masaki, Tajimi Tomoki, Shioya Ryota, Goshima Masahiro, Tsumura Tomoaki	4. 巻 -
2. 論文標題 An Analysis and a Solution of False Conflicts for Hardware Transactional Memory	5. 発行年 2018年
3. 雑誌名 Proc. 25th IEEE Int'l Conf. on Electronics Circuits and Systems (ICECS 2018)	6. 最初と最後の頁 529-532
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ICECS.2018.8617977	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 Tajimi Tomoki, Hayashi Masaki, Futamase Yuki, Shioya Ryota, Goshima Masahiro, Tsumura Tomoaki	4. 巻 -
2. 論文標題 Isolation-Safe Speculative Access Control for Hardware Transactional Memory	5. 発行年 2018年
3. 雑誌名 Proc. 25th IEEE Int'l Conf. on Electronics Circuits and Systems (ICECS 2018)	6. 最初と最後の頁 517-520
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ICECS.2018.8618020	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 Norishige FUKUSHIMA, Yoshihiro MAEDA, Yuki KAWASAKI, Masahiro NAKAMURA, Tomoaki TSUMURA, Kenjiro SUGIMOTO, Sei-ichiro KAMATA	4. 巻 -
2. 論文標題 Efficient Computational Scheduling of Box and Gaussian FIR Filtering for CPU Microarchitecture	5. 発行年 2018年
3. 雑誌名 Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2018)	6. 最初と最後の頁 875-879
掲載論文のDOI（デジタルオブジェクト識別子） 10.23919/APSIPA.2018.8659674	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 Shinji KAWAMURA, Tomoaki TSUMURA	4. 巻 E101-D
2. 論文標題 Hardware Accelerated Marking for Mark & Sweep Garbage Collection	5. 発行年 2018年
3. 雑誌名 IEICE Trans. on Information and Systems	6. 最初と最後の頁 1107-1105
掲載論文のDOI (デジタルオブジェクト識別子) 10.1587/transinf.2017EDP7163	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 Tomoki TAJIMI, Anju HIROTA, Ryota SHIOYA, Masahiro GOSHIMA, Tomoaki TSUMURA	4. 巻 1
2. 論文標題 Initial Study of a Phase-Aware Scheduling for Hardware Transactional Memory	5. 発行年 2017年
3. 雑誌名 Proc. IEEE Pacific Rim Conf. on Communications, Computers and Signal Processing	6. 最初と最後の頁 1-6
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/PACRIM.2017.8121912	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20件 (うち招待講演 1件 / うち国際学会 5件)

1. 発表者名 津邑 公暁
2. 発表標題 Computation Reuse と Approximate Computing
3. 学会等名 電子情報通信学会 ソサイティ大会 (招待講演)
4. 発表年 2018年

1. 発表者名 Tomoki TAJIMI, Masaki HAYASHI, Yuki FUTAMASE, Ryota SHIOYA, Masahiro GOSHIMA, Tomoaki TSUMURA
2. 発表標題 Isolation-Safe Speculative Access Control for Hardware Transactional Memory
3. 学会等名 25th IEEE Int'l Conf. on Electronics Circuits and Systems (ICECS 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Yuki FUTAMASE, Masaki HAYASHI, Tomoki TAJIMI, Ryota SHIOYA, Masahiro GOSHIMA, Tomoaki TSUMURA
2. 発表標題 An Analysis and a Solution of False Conflicts for Hardware Transactional Memory
3. 学会等名 25th IEEE Int'l Conf. on Electronics Circuits and Systems (ICECS 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Norishige FUKUSHIMA, Yoshihiro MAEDA, Yuki KAWASAKI, Masahiro NAKAMURA, Tomoaki TSUMURA, Kenjiro SUGIMOTO, Sei-ichiro KAMATA
2. 発表標題 Efficient Computational Scheduling of Box and Gaussian FIR Filtering for CPU Microarchitecture
3. 学会等名 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Tomoki Tajimi, Yuki Futamase, Masaki Hayashi, Ryota Shioya, Masahiro Goshima, Tomoaki Tsumura
2. 発表標題 Speculatively Granting Conflicting Accesses on Hardware Transactional Memory
3. 学会等名 ACM Student Research Competition (SRC), held in conjunction with MICRO-51 (国際学会)
4. 発表年 2018年

1. 発表者名 林 昌樹, 二間瀬 悠希, 多治見 知紀, 塩谷 亮太, 五島 正裕, 津邑 公暁
2. 発表標題 トランザクショナルメモリとロックを併用する並行性制御手法
3. 学会等名 The 2nd cross-disciplinary Workshop on Computing Systems, Infrastructures, and Programming (xSIG 2018)
4. 発表年 2018年

1. 発表者名 横手 宥則, 三輪 忍, 井内 悠太, 津邑 公暁, 八巻 隼人, 本多 弘樹
2. 発表標題 学習済み重みを利用した畳み込みニューラルネットワークの学習法の初期検討
3. 学会等名 電子情報通信学会 総合大会
4. 発表年 2018年

1. 発表者名 飯田 凌大, 津邑 公暁
2. 発表標題 軽量な一貫性検証によるソフトウェアトランザクショナルメモリの並列性向上に関する検討
3. 学会等名 2018年並列 / 分散 / 協調処理に関する『熊本』サマー・ワークショップ (SWoPP2018)
4. 発表年 2018年

1. 発表者名 井内 悠太, 津邑 公暁
2. 発表標題 カーネルテンプレート化と計算再利用によるCNNの計算量削減に関する検討
3. 学会等名 2018年並列 / 分散 / 協調処理に関する『熊本』サマー・ワークショップ (SWoPP2018)
4. 発表年 2018年

1. 発表者名 多治見 知紀, 林 昌樹, 二間瀬 悠希, 塩谷 亮太, 五島 正裕, 津邑 公暁
2. 発表標題 競合アクセスを投機的に許可するトランザクショナルメモリの検討
3. 学会等名 HotSPA2018 (Hot SPring Annual meeting 2018)
4. 発表年 2018年

1. 発表者名 二間瀬 悠希, 林 昌樹, 多治見 知紀, 塩谷 亮太, 五島 正裕, 津邑 公暁
2. 発表標題 トランザクショナルメモリにおける競合誤検出の影響調査とその改善手法
3. 学会等名 HotSPA2018 (Hot SPring Annual meeting 2018)
4. 発表年 2018年

1. 発表者名 進藤 智司, 松井 優樹, 八巻 隼人, 津邑 公暁, 三輪 忍
2. 発表標題 カーネルの類似性に基づく近似計算を行うCNNアクセラレータの検討
3. 学会等名 第222回システム・アーキテクチャ・第183回システムとLSIの設計技術・第47回組込みシステム合同研究発表会 (ETNET2018)
4. 発表年 2017年

1. 発表者名 古橋 一輝, 津邑 公暁
2. 発表標題 高抽象度言語とオートチューニング機能を持つ動画処理環境
3. 学会等名 第222回システム・アーキテクチャ・第183回システムとLSIの設計技術・第47回組込みシステム合同研究発表会 (ETNET2018)
4. 発表年 2017年

1. 発表者名 林 昌樹, 二間瀬 悠希, 多治見 知紀, 塩谷 亮太, 五島 正裕, 津邑 公暁
2. 発表標題 最適な並行性制御を適用するコード生成手法の検討
3. 学会等名 第222回システム・アーキテクチャ・第183回システムとLSIの設計技術・第47回組込みシステム合同研究発表会 (ETNET2018)
4. 発表年 2017年

1. 発表者名 松井 優樹, 三輪 忍, 進藤 智司, 津邑 公暁, 八巻 隼人, 本多 弘樹
2. 発表標題 CNN計算の省メモリ化のためのカーネル・クラスタリング手法の検討
3. 学会等名 第222回システム・アーキテクチャ・第183回システムとLSIの設計技術・第47回組込みシステム合同研究発表会 (ETNET2018)
4. 発表年 2017年

1. 発表者名 進藤 智司, 松井 優樹, 八巻 隼人, 津邑 公暁, 三輪 忍
2. 発表標題 高電力効率なCNNアクセラレータ実現に向けたカーネルクラスタリングの応用の検討
3. 学会等名 2017年並列 / 分散 / 協調処理に関する『秋田』サマー・ワークショップ (SWoPP2017)
4. 発表年 2017年

1. 発表者名 多治見 知紀, 廣田 杏珠, 塩谷 亮太, 五島 正裕, 津邑 公暁
2. 発表標題 実行フェーズを考慮したトランザクショナルメモリのスケジューリング手法
3. 学会等名 2017年並列 / 分散 / 協調処理に関する『秋田』サマー・ワークショップ (SWoPP2017)
4. 発表年 2017年

1. 発表者名 早川 慎一郎, 河村 慎二, 津邑 公暁
2. 発表標題 ハードウェア支援型GCの消費エネルギー評価
3. 学会等名 2017年並列 / 分散 / 協調処理に関する『秋田』サマー・ワークショップ (SWoPP2017)
4. 発表年 2017年

1. 発表者名 松山 且樹, 藤井 政圭, 津邑 公暁, 中島 康彦
2. 発表標題 自動メモ化プロセッサにおける復帰アドレス別の再利用率調査とその応用
3. 学会等名 2017年並列 / 分散 / 協調処理に関する『秋田』サマー・ワークショップ (SWoPP2017)
4. 発表年 2017年

1. 発表者名 Tomoki TAJIMI, Anju HIROTA, Ryota SHIOYA, Masahiro GOSHIMA, Tomoaki TSUMURA
2. 発表標題 Initial Study of a Phase-Aware Scheduling for Hardware Transactional Memory
3. 学会等名 IEEE Pacific Rim Conf. on Communications, Computers and Signal Processing (PacRim 2017) (国際学会)
4. 発表年 2017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