

令和 4 年 5 月 18 日現在

機関番号：12601

研究種目：若手研究

研究期間：2019～2021

課題番号：19K15051

研究課題名（和文）深層学習用途に最適化した不揮発性半導体メモリシステムの研究

研究課題名（英文）Research on non-volatile memory system for deep learning applications

研究代表者

松井 千尋（Matsui, Chihiro）

東京大学・大学院工学系研究科（工学部）・特任助教

研究者番号：80823484

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：不揮発性半導体メモリの性能・信頼性・電力のトレードオフを解決するため、複数種類の不揮発性半導体メモリを活用したメモリシステムを協調設計することに取り組んだ。高速・高信頼・リアルタイムにデータを収集・蓄積・解析するため、不揮発性半導体メモリのデバイス構成、データマネジメント手法、データ変調手法を提案し、高速ネットワークで接続され階層化されたクラウドデータセンタからエッジデバイスまでの不揮発性半導体メモリシステム全体をまとめた。

研究成果の学術的意義や社会的意義

データが生成される近くでデータ解析を行い、データに基づいた意思決定を行うエッジコンピューティングが進んでいる。車やロボットに搭載されデータを収集する画像センサ、温度センサなどのエッジデバイス自身が、取得した画像データやセンサデータを大量に保存し、リアルタイムに処理する。本研究は特に半導体工場をモチーフとして、高速ネットワークにより半導体工場の装置やサーバが複数階層を構築するときに、各階層に必要とされるレイテンシを達成するための不揮発性半導体メモリシステム構成を明らかにするなど、スマートファクトリーの実現への貢献も大きい。

研究成果の概要（英文）：This research project works on co-designing memory system utilizing several types of non-volatile memories to resolve the trade-off among performance, reliability, and power of memories. To collect, store, and analyze data with high speed, high reliability, and real-time, memory device configuration, data management algorithm, and data modulation algorithm of non-volatile memory system have been proposed. The entire non-volatile memory system from cloud data center to edge devices, connected by hierarchical high-speed networks, is integrated.

研究分野：不揮発性半導体メモリシステム

キーワード：不揮発性半導体メモリ メモリシステム 深層学習 ネットワーク データストレージ

1. 研究開始当初の背景

IoT などの応用では、実世界の情報を収集・応答するエッジと、エッジで集めたデータを蓄積・解析して実世界にフィードバックするクラウドデータセンタが強調する、サイバーフィジカルシステム（CPS）が重要になる。IoT の用途である自動運転車、セキュリティカメラを使った見守り・セキュリティシステムなどの深層学習を用いたシステムは、ミリ秒以下のリアルタイム応答を必要とする。次世代高速無線通信の登場によりネットワーク、データプロセッシングは高速化が進む一方、現在の主要な大容量不揮発性半導体メモリであるフラッシュメモリはアクセス時間がミリ秒以上と遅く、CPS システム全体の中でメモリがボトルネックとなる。高速不揮発性半導体メモリである MRAM、ReRAM、PRAM などを使うことは極めて重要であるが、性能・信頼性・電力の問題のため普及が進んでいないという問題があった。

2. 研究の目的

不揮発性半導体メモリの本質的な課題である性能・信頼性・電力のトレードオフを解決する、電子デバイス・回路システムを統合した解決手法を構築する。エッジデバイスとクラウドデータセンタで行われる深層学習アプリケーションに対して、最適な不揮発性半導体メモリシステムを構築する。エッジデバイスでは低電力・低コストが求められ、クラウドデータセンタでは強力なコンピューティングリソースを持つ。そこで、クラウドデータセンタでは大量のデータを用いて深層学習を行い、その学習モデルを用いてエッジデバイスで推論を行う。つまり、クラウドデータセンタでは高速な不揮発性半導体メモリ、エッジデバイスでは低消費電力な不揮発性半導体メモリが必要となる。本研究では不揮発性半導体メモリのデバイス構成、データマネジメント手法、データ変調手法の工夫により、深層学習アプリケーションに適した不揮発性半導体メモリシステムを明らかにする。

3. 研究の方法

エッジデバイスおよびクラウドデータセンタ向けの不揮発性半導体メモリシステムエミュレータを構築する。入力するアプリケーションワークロードを変更し、深層学習を行うエッジデバイス向けおよびクラウドデータセンタ向けのエミュレータとして用いる。不揮発性半導体メモリをモデル化し、メモリセルの書き換え時間・読み出し時間を変更することで、高速および低速な不揮発性メモリの動作や、メモリセルに生じるエラーを低減するペリファイ書き込みを模擬する。また、アプリケーションワークロードによって不揮発性半導体メモリセルに生じるエラーを評価する。発生したエラーを訂正するため、さまざまなエラー率に対応するエラー訂正符号（ECC）も実装した。さまざまな特性を持つストレージワークロードを用いて、これらシステム全体のデータ処理性能、消費エネルギー、エラーを評価し、メモリシステムのボトルネックの原因を明らかにする。

4. 研究成果

不揮発性半導体メモリの性能・信頼性・電力のトレードオフを解決するため、複数種類の不揮発性半導体メモリを活用したメモリシステムを協調設計する研究に取り組んだ。従来のネットワークではネットワークのレイテンシがボトルネックだったが、高速ネットワークはネットワークのレイテンシが短くなるため、不揮発性半導体メモリの書き込み・読み出しレイテンシがボトルネックとなる。クラウドからエッジまでが高速ネットワークで接続されたメモリシステムの階層で深層学習を行う場合に、各ネットワーク階層のエッジマシン・エッジサーバ・クラウドサーバで扱うデータの性質を考慮し、それぞれ異なる不揮発性半導体メモリを用いたメモリシステムモデルを設計した。エッジデバイスには 3D-TLC NAND 型フラッシュメモリ、エッジサーバには 3D-SLC NAND 型フラッシュメモリ、クラウドサーバには 3 次元積層した抵抗変化型メモリをそれぞれ用いる。

エッジデバイス向けの 3D-TLC NAND 型フラッシュメモリシステムでは、イメージセンサで取得した画像を高信頼に保存し、空間的局所性および色のパターンを持つ画像の認識精度を高めるデータ変調技術を提案した。画像データの (Green, Blue, Red) をそれぞれ 3D-TLC NAND 型フラッシュメモリの Upper/Middle/Lower ページに保存し、行ごとに最適なデータ変調を行うことでメモリの信頼性が向上し、その結果画像認識精度が向上した。

エッジサーバ向けの高速・大容量の 3D-SLC NAND 型フラッシュメモリシステムでは、メモリのエラー訂正符号の種類によって、訂正能力の低い ECC ではネットワークボトルネック、訂正能力の高い ECC ではメモリボトルネックとなる。高い訂正能力を持つ ECC を適用する場合、複数のメモリチップを並列動作することでメモリによる遅延を低減できた。

クラウドサーバでは深層学習の共通モデルを更新しエッジサーバへ共通モデルを分配するため、高速な抵抗変化型メモリを用いることで不揮発性半導体メモリシステムの遅延をネットワーク遅延より短縮できる。さらに 3 次元積層した抵抗変化型メモリシステムでは、上層にプロセスばらつきがある場合、ネットワークボトルネックとなることを明らかにした。書き換えやデータ保持によって不揮発性半導体メモリセルに生じたエラーを訂正するため、ECC 遅延を含めたメモリシステム全体の遅延の評価を行い、高速ネットワークによる遅延と同等のメモリ遅延が実現できた。

以上により、エッジデバイスからエッジサーバ・クラウドサーバまでシステム・ネットワーク全体を包括する不揮発性半導体メモリシステムを確立する基礎を築いた。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1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 Chihiro Matsui and Ken Takeuchi	4. 巻 60
2. 論文標題 Non-volatile memory system design of edge server and cloud centralized server for multiple-tier 5G network	5. 発行年 2021年
3. 雑誌名 Japanese Journal of Applied Physics	6. 最初と最後の頁 SBBB05 ~ SBBB05
掲載論文のDOI（デジタルオブジェクト識別子） 10.35848/1347-4065/abe7fc	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3件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 松井千尋、竹内健
2. 発表標題 階層型5Gネットワークとヘテロジニアス不揮発性メモリシステムの協調設計
3. 学会等名 第82回応用物理学会秋季学術講演会
4. 発表年 2021年

1. 発表者名 Chihiro Matsui and Ken Takeuchi
2. 発表標題 SLC Flash & ReRAM Heterogeneous Memory System with Multi-Tier 5G Network & Device Co-Design for Smart Manufacturing
3. 学会等名 International Conference on Solid State Devices and Materials (国際学会)
4. 発表年 2020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
---------------------------	-----------------------	----

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