News Release
2024.10.11

NEDO（国立研究開発法人新エネルギー・産業技術総合開発機構）
国立大学法人東北大学
株式会社アイシン

大容量 MRAM を搭載したエッジ領域向け「CMOS／スピントロニクス融合 AI 半導体」により従来比 10 倍以上の電力効率をシステム動作シミュレーションで確認

NEDOは「省エネAI半導体及びシステムに関する技術開発事業」（以下、本事業）において、エッジ領域に適した半導体デバイスの早期実現を目指して、開発を進めています。このたび、国立大学法人東北大学と株式会社アイシンは、大容量MRAMを搭載したエッジ領域向け「CMOS／スピントロニクス融合AI半導体」により従来比10倍以上の電力効率をシステム動作シミュレーションで確認しました。磁気抵抗メモリ（MRAM）の不揮発性と広バス帯域の特性を活用し、大容量のMRAMを搭載して外付けメモリの合理的な内蔵化を図ることにより、動作時および待機時電力の大幅低減、起動時間の短縮が可能になります。RTLでのシステム動作シミュレーションの検証では、従来比で電力効率10倍以上、起動時間10分の1以下の改善効果を確認しました。今後は、車載やサーベイランス（監視）システムなどへの応用技術開発を進めます。

また、本事業の成果について、2024年10月15日から10月18日まで幕張メッセで開催される「CEATEC2024」のNEDOブースに展示します。

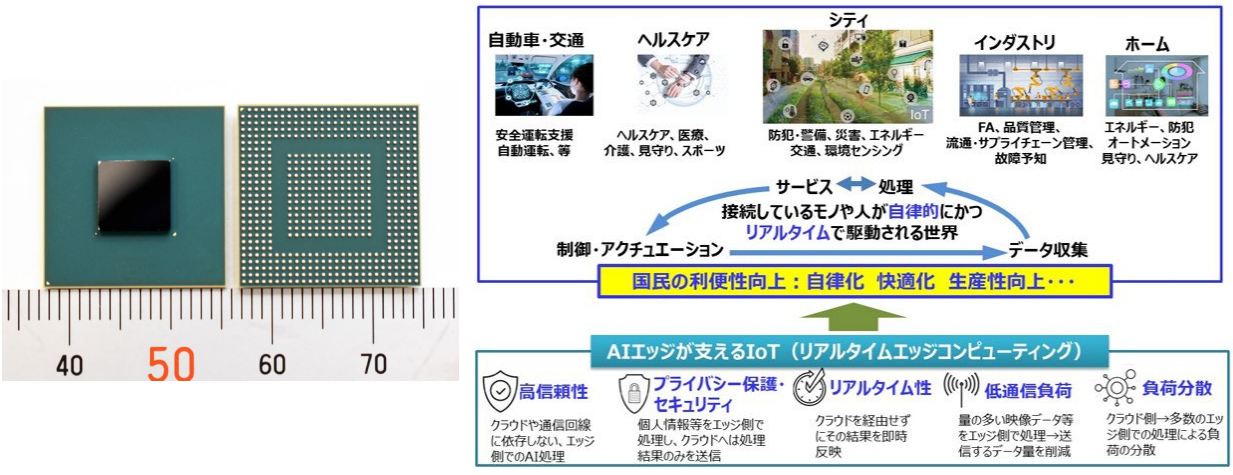

図1　設計した実証チップと本事業が目指す多様な社会実装

1. 概要

近年、情報処理に用いるデバイスの高度化、AIなどを用いる、さまざまな産業の創出とその基礎となるビッグデータの活用や、5Gなどの情報通信技術・インフラ整備により、ネットワーク上のデータ量が爆発的に増加しています。さらに利用環境としてエッジ領域はクラウド領域と異なり、情報処理に用いられる電力や、サイズ、利用環境などにさまざまな制約があるため、エッジ領域用途に適したデバイスの早期実現が

重要と考えられます。

このような背景の下、NEDOでは、本事業※1を実施し、この一環として東北大学、アイシン、日本電気株式会社と共同で、CMOS／スピントロニクス融合技術※2によるAI処理半導体の設計効率化と実証およびその応用技術に関する研究開発を進めています。

本事業において、東北大学はCMOS／スピントロニクス融合技術を今後のエッジ向けAI半導体の基盤となる内部メモリとして有効に活用するべく、MRAM※3を用いた自動設計環境の構築とそれに基づくAIアクセラレータ※4を開発し、アイシンはこれとアプリケーションプロセッサ※5ほか周辺IPを統合した実証チップ（図1）のシステム設計を行うとともに、RTLでのシステム動作検証において、従来比で電力効率10倍以上および起動時間10分の1以下の改善効果を確認しました。

2. 今回の成果

(1)エッジAI用プロセッサの共通課題をCMOS／スピントロニクス融合技術であるMRAMにより解消

現在、小規模から大規模まで多くのAIシステムでは、アプリケーションプロセッサを内蔵したチップを用いたシステムが採用されています。このようなチップでは、BOOT用外付けFLASHメモリ、外付けメモリ（DRAM）および内蔵メモリ（SRAM）を備える構成が標準的で、ファーメモリ・コンピューティング構造※6となっています。

BOOT用外付けFLASHメモリは、他のメモリに比べバス帯域が狭く、ランダムアクセスができないため、起動時に外付けFLASHメモリの内容を内蔵SRAMや外付けDRAMにコピーするプロセスが必須であり、小規模なAIエッジシステムを構築する際にも長い起動時間がかかるという課題を抱えていました。

今回設計した実証チップでは、内部メモリと重みメモリ※7にMRAMを用いることでニアメモリ・コンピューティング構造※8を実現し、外付けFLASHメモリのバス帯域不足の解消や、アプリケーションプロセッサ上のソフトウエアが起動する際に必要とされていた多くのプロセスが削減でき、外付けデバイスの初期化時間の削減も可能になります（図2）。

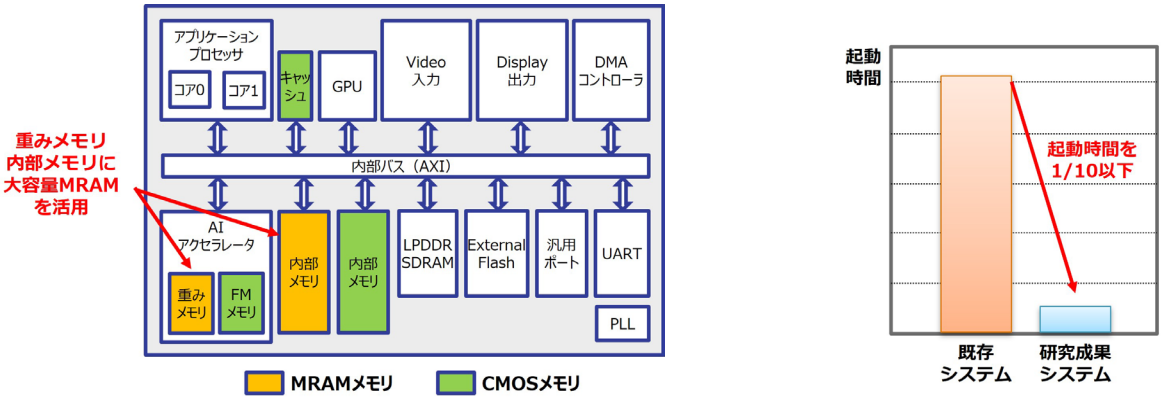

図2　設計した実証チップのブロック図とシステム動作シミュレーションによる起動時間短縮の効果

(2)低消費電力AIアクセラレータの開発

今回設計した実証チップには、東北大学国際集積エレクトロニクス研究開発センターが開発・設計してきた低消費電力AIアクセラレータを搭載しています。本アクセラレータでは、主要メモリにMRAMを適用することで、従来のSRAMと比較して面積効率と電力効率のどちらにも良い特徴があるため、待機電力や動作電力を大幅に削減できます。さらに、不揮発化により重みメモリへのロード時間を削減でき、AI処理システム

全体の高速起動が実現可能になります。

(3)実証チップのアーキテクチャ開発と回路設計

今回アイシンは実証チップのアーキテクチャ設計を行い、アプリケーションプロセッサのBOOT用途とメインメモリ用途を兼ねた内蔵メモリとして世界で初めて大容量MRAMを採用し、チップに内蔵しました。内蔵メモリにMRAMを採用することにより、起動時間の短縮と外付けメモリ容量の削減が可能になり、チップの小面積化および低消費電力化が図れます。

アイシンは東北大学と連携してRTLでのシステム動作シミュレーション検証を行い、従来比で電力効率10倍以上および起動時間10分の1以下の改善効果を確認しました。本実証チップは台湾積体電路製造(TSMC社)のMRAM混載に対応した次世代16nm FinFETプロセスのPDK(Process Design Kit)を用いて設計しました。アプリケーションプロセッサにはArm® Cortex®-A53デュアルコアを採用しています。

3. 今後の予定

NEDOと各機関は連携して、電力効率の詳細評価およびシステム検証を進めます。本研究成果をいち早く実用化につなげるため、車載機器やサーベイランス機器への応用を計画しています。本技術の確立によるエッジAI処理での電力効率改善とその早期社会実装を通して、二酸化炭素(CO_2)排出量の削減に貢献します。

なお、本事業の成果については、2024年10月15日から10月18日まで幕張メッセで開催される「CEATEC2024」のNEDOブース※9にて展示する予定です。

【注釈】

※1 本事業

事 業 名:省エネAI半導体及びシステムに関する技術開発事業/AIエッジコンピューティングの産業応用加速のための設計技術開発/CMOS/スピントロニクス融合技術によるAI処理半導体の設計効率化と実証、及び、その応用技術に関する研究開発

事業期間:2022年度～2024年度

事業概要:省エネAI半導体及びシステムに関する技術開発事業 https://www.nedo.go.jp/activities/ZZJP_100254.html

※2 CMOS/スピントロニクス融合技術

技術開発が進むCMOS技術において極めて重要な問題となっている待機時電力をスピントロニクスの不揮発性を活用して大幅に削減し、加えてスピントロニクスの高い面積効率を生かし、スピントロニクス素子の微細化を通して電力効率10倍以上の改善を実現可能とする技術です。

※3 MRAM

MRAM (Magnetoresistive Random Access Memory) は、磁化の方向で情報を記憶する不揮発性メモリです。1ns程度の高速な磁化反転速度により高速動作が可能であるとともに、原子移動がないために書き換え耐性が高く、他の不揮発性メモリにはない優位性を有しています。

※4 AIアクセラレータ

AIアクセラレータとは、AIの計算処理を高速化するために設計されたハードウエアのことを指します。従来のCPUやGPUよりも高速にAIの計算を行い、AIアプリケーションにおけるコストを大幅に低減します。

※5 アプリケーションプロセッサ

アプリケーションプロセッサとは、スマートフォンやタブレット端末などに内蔵されているマイクロプロセッサの一つで、コンピュータ機能においてCPUとしてオペレーティングシステム(OS)やアプリの実行を担当するものです。

※6 ファーメモリ・コンピューティング構造

演算回路とプログラムおよびデータが格納されているメモリを離れた位置に配置する構造のことで、大容量メモリや複数の異なるメモリ配置が可能となり、また使用するメモリを複数のモジュールで共有することができシステムの共通化が可能になります。しかしこの構造の要になるメモリと演算器をつなぐバス（配線）が、高性能化と低消費電力化を阻むボトルネックになってしまう欠点を抱えています。

※7 重みメモリ

重みメモリとは、ニューラルネットワークにおいて入力値の重要性、貢献度を数値化して表したものを格納するメモリのことです。 ニューラルネットワークでは重みとバイアスを調整することで学習を進めます。バイアスは入力値を一定の範囲に偏らせるために用いるもので、重みはその入力値ごとに決められ、その入力値の価値を決めるもので推論や学習に使用されます。

※8 ニアメモリ・コンピューティング構造

演算回路とプログラムおよびデータが格納されているメモリを極限まで近傍に配置する構造のことで、メモリアクセスがシステム性能上のボトルネックとなることを解消できます。イン・メモリ・コンピューティング（In Memory Computing : IMC)構造とも言います。

※9 NEDOブース

「CEATEC2024」への出展

https://www.nedo.go.jp/events/IT_100107.html