

情報担体とその集積のための材料・デバイス・システム
2021 年度採択研究代表者

2022 年度
年次報告書

小菅 敦丈

東京大学 大学院工学系研究科
講師

デバイス・システム協調による超低電圧布線論理型 AI プロセッサ

研究成果の概要

Society5.0 実現のためには実空間で超低電力に動作する情報処理システムが必要不可欠である。従来 AI プロセッサの課題は、高い電力を消費するメモリアクセスであった。本研究ではメモリアクセスを無くし、人間の脳と同じように、データが演算素子間をダイレクトに流れ情報処理される布線論理型プロセッサを研究している。

2021 年度は上記課題を解決し世界に先駆けて大規模 AI の布線論理型プロセッサを実現するために、2 つの要素技術:(A)省ニューロンで省シナプスなニューラルネットワークアルゴリズム(Non-linear neural network NNN)、(B)NNN を省面積に処理するシフトレジスタ型ニューロンセルアレーの研究を進めた。2022 年度において、これら成果を学会発表した。プロセッサ分野における最高峰の学会の 1 つである HotChips 2022[1]、および設計自動化手法について発表した[2]。

基本技術について開発を進め一定の成果が得られている一方、実際の応用に対してどのくらいの性能を発揮できるかに関し検証が進んでいない。2022 年度はこれまで開発した NNN 技術とシフトレジスタ型ニューロンセルアレーを発展させ、実アプリケーションへの適用を目指しさらなる大規模化技術を開発した。

そこで(C)音声コマンド認識 AI 技術を題材に、実応用における性能検証を目標に布線論理型プロセッサ開発を進めた。新たに開発したアルゴリズム-回路協調最適化技術により、40nm CMOS プロセス 3mm 角の 1 チップにおいて、16 層の深層ニューラルネットワークを布線論理方式で実装することに成功した。従来 AI プロセッサと比較し、1/2552 もの消費電力削減を実現した。35 種識別可能な音声コマンド AI を、乾電池 1 本で 2.2 年にわたり連続動作させることが可能となった。スマートフォン、ドローン、自動車内エンタメ機器制御、AR/VR 機器へ応用が期待される。回路分野の最高峰学会である VLSI シンポジウムで発表した[3]。

さらに(D)ラスタースキャン型ニューロンセルアレー技術を開発した。音声コマンド認識や大規模画像認識などの実用的なアプリケーションにおいては、より大きなネットワークサイズが必要となる。大きなネットワークサイズはチップサイズの増大を招き、複数チップ実装が必要となる。チップ間通信の消費電力が課題となるため、チップ面積の増大を抑えつつ多数のフィルタを実装可能な新たな布線論理型アーキテクチャが必要となる。そこで本研究では、ラスタースキャン型ニューロンセルアレー技術を開発した。フィルタの展開実装を最小限にとどめ時分割で回路素子を再利用することで、シフトレジスタ型ニューロンセルアレーよりも 5 倍以上省面積化することを可能にした。本成果は回路とシステム分野における最高峰の学会である ISCAS'23[4]にて口頭発表を行った。

【代表的な原著論文情報】

[1] Y.-C. Hsu *et al.*, *IEEE Hot Chips 34 Symposium (HCS)*, Aug. 2022.

[2] R. Sumikawa *et al.*, *IEEE ASP-DAC'23*, Jan. 2023.

[3] A. Kosuge *et al.*, *IEEE Symposium on VLSI Circuits*, June 2023 (to be presented).

[4] D. Li *et al.*, *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2023.