

超低電力 AI ハードウェアを実現する高精度不揮発性アナログメモリ技術を開発

■ 発表のポイント

- 株式会社フローディア(以下、フローディア)と日本電気株式会社、国立大学法人九州工業大学(以下、九州工業大学)は共同で、超低消費電力で AI 推論が可能な次世代ハードウェアである不揮発性アナログ・コンピューティング・イン・メモリ(nvACiM)で必須となる、AI モデルパラメータの高精度書き分け・長時間保持技術を開発しました。
- これまでの nvACiM 向けメモリでは書き込み速度のばらつきが大きく、個々のメモリ素子に設定するパラメータのレベル数(ビット精度)を大きくできない、また、設定したパラメータが時間とともに変化してしまうという問題があり、推論精度の低下を生じていました。
- 今回、SONOS 型フラッシュメモリを用い、複数のステップに分けて目標のパラメータ値に応じた最適な電圧で書き込みを行うことで、1つのメモリ素子に多値パラメータを高精度に設定することが可能となりました。また、特殊な熱処理により SONOS 型メモリ素子の電荷蓄積部のバンド構造を工夫するとともに、抜けやすい電子を書込みシークエンス中に除去することで、長時間のパラメータ保持を実現しました。
- AI のエッジ応用では推論を行う半導体メモリの低電力化が急務です。本技術により nvAiCM の演算精度が向上することでその応用が広がり、ロボットやドローン、自動車といったエッジ AI 応用機器の高性能化や省エネ、小型化への貢献が期待されます。

■ 背景

大規模言語処理や画像認識など、GPU を用いた AI 技術は我々の生活を大きく変化させています。しかし、GPU を用いた推論はデジタル計算のため、性能向上にはプロセッサの演算速度向上とともに、演算に用いる入力データと特徴を識別するパラメータ(重みともいう)をプロセッサとメモリ間で逐次高速でやり取りする必要があり、大きな電力を消費します。不揮発性メモリを用いたアナログ・コンピューティング・イン・メモリ(nvACiM)技術は、あらかじめ複数のメモリ素子に任意のパラメータを設定し、並列にアナログデータを入力して積和演算を行うことで、超低電力な AI 処理が可能となります(図 1)。

しかし、これまで nvACiM 向けに検討されてきたメモリではメモリ素子間の書き込み速度のばらつきが大きく、個々のメモリ素子に設定できるパラメータのレベル数(ビット精度)に限られるという問題がありました。また、時間とともに設定したパラメータの値が変化してしまうという問題もあり、これらは演算精度の低下を招いていました。このため、各メモリ素子に多値パラメータを設定する高精度な書き分けとその長時間保持技術の開発が必要となっていました。

■ 発表内容

(1) SONOS メモリの適用

今回、nvACiM として、ファウンダリで認証済のフローディア SONOS 型フラッシュメモリ技術を用いました。SONOS 型メモリはシリコン窒化膜中に不連続に存在するトラップ（電子捕獲準位）に電子を蓄積することで情報の記憶を行います。このため原理的に高い精度でパラメータ値を設定することが可能です。また、メモリ素子間の書込み速度ばらつきが小さく、パラメータ値のばらつきを抑制しやすい、蓄積した電子の保持特性が高いため設定したパラメータ値を長時間維持しやすい、という特長があります。しかし、これまでの他機関での報告では、設定できるパラメータ精度は 8 値(3bit 精度)にとどまっていた。

(2) パラメータ高精度書き分け技術

今回、SONOS 型メモリを用いた nvACiM で高いビット精度を実現するため、書込みを複数ステップに分け、夫々のステップでパラメータの値に応じた最適な電圧を用いて書込みと検証を繰り返し、パラメータのレベル数を倍々に増やしていく、iterative multi-step programming sequence 手法を開発しました。シミュレーションにより、4~1024nA の範囲で 256 レベル(8bit 精度)のパラメータ電流が設定できることを確認しました。

(3) パラメータ長時間保持技術

SONOS 型メモリは他のメモリに比べ優れた電荷保持特性を有するものの、nvACiM ではパラメータレベル間の電流差が小さいため、更なる特性向上が必要です。今回、電荷蓄積部の形成工程に特殊な熱処理を追加しバンド構造を工夫することで、書込み電圧の上昇を抑えつつ電子の漏洩を抑制する band control thermal treatment (BCT)技術を開発しました。また、SONOS 型メモリではシリコン窒化膜中の浅いトラップに蓄積した電子が離脱するとパラメータ値が変化してしまうため、(2)のパラメータ書き分けシーケンス中に浅いトラップに蓄積した電子を除去する shallow trap charge reduction (STR) process 技術を導入しました。その結果、パラメータの長時間保持が可能となりました。

(4) nvACiM テストチップを用いた検証

(2)(3)の技術を実装し、効果を検証しました。2~64 nA 間に 2 nA 刻みで 32 値(5 ビット精度)のパラメータを設定できました。各パラメータ値の分布の標準偏差は目標に対し 0.78%です(図 2)。これらの値は、従来、浮遊ゲート型フラッシュメモリで報告されている 3~100nA 間での 3nA 刻み書き分け、標準偏差 1.3%を凌ぐもので、極低消費電力での推論の実現につながります。さらに、高精度な積和演算が実現可能なことを確認しました。

本成果は国立研究開発法人新エネルギー・産業技術総合開発機構(NEDO)の委託業務(JPNP16007)の結果得られたものです。



本技術は 2026 年 5 月 10 日~13 日にベルギー、ルーベン市で開催の International Memory Workshop (IMW) 2026 で発表されました。

学会情報

学会名：IEEE International Memory Workshop (IMW) 2026

会期：2026 年 5 月 10 日~13 日

題名：Enhancing SONOS-type Flash Memory for Nonvolatile Analog Computing-in-Memory via Precise Multi-level Weight Control and Improved Retention

著者名：T. Kobayashi, Y. Taniguchi, H. Ikegaya, T. Tamatsu, M. Yamaguchi, K. Yanagisawa, Y. Uji, T. Yabe, S. Yoshida, Y. Kawashima, T. Shimozato, H. Kobori, T. Kato, F. Owada, S. Noda, H. Yoneda, K. Okuyama, T. Hosomi, K. Oto, S. Koshina, T. Morie

URL：<https://ewh.ieee.org/soc/eds/imw/>

本プレスリリースに関する問い合わせ先

フローディア：<https://floadia.com/jp/#contact>

九州工業大学：ニューロモルフィック AI ハードウェア研究センター 特別教授 森江 隆
(morie@brain.kyutech.ac.jp)

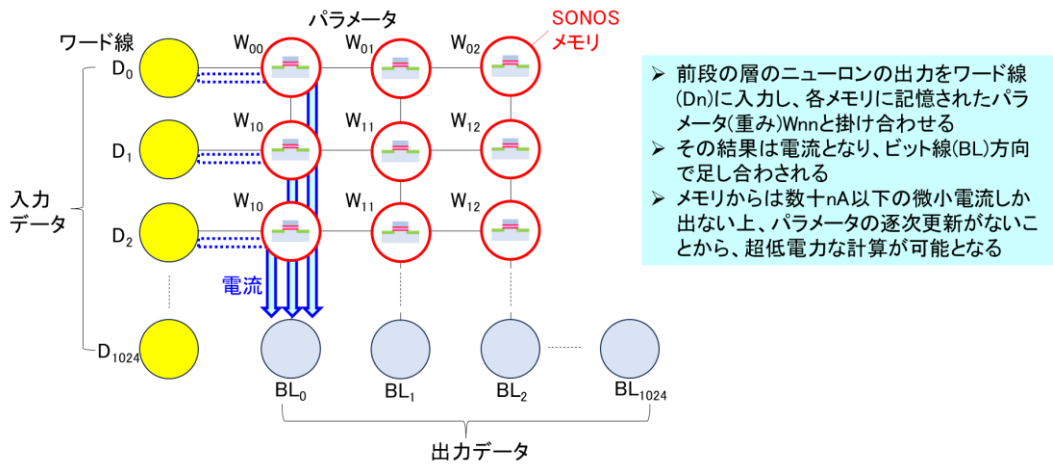


図 1 nvACiM の構成と計算

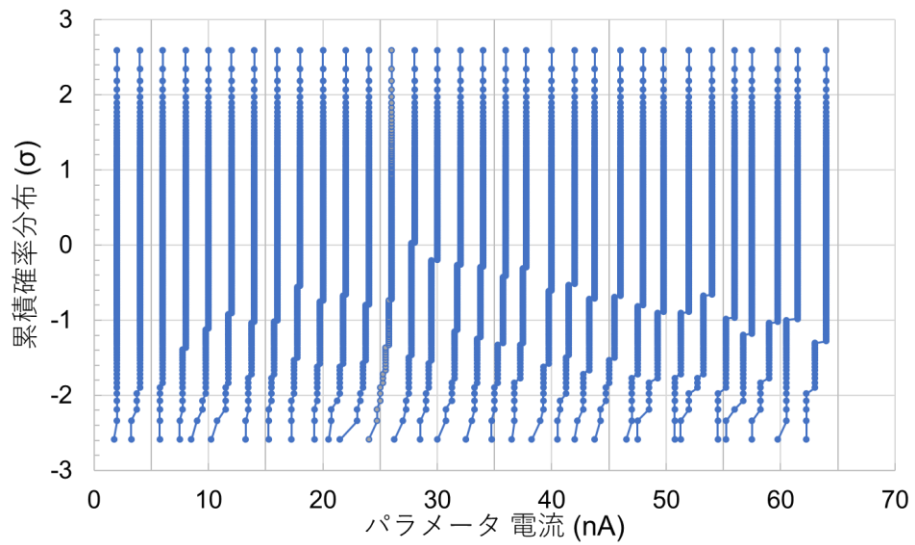


図 2 2nA 刻み 32 値パラメータ書き分け結果