

近似3次元スタッキング・シナプス記憶による低消費電力AIシステム

准教授 Dang Nam Khanh



データ分割による近似計算で消費電力を劇的に改善する

関連発明：ニューラルネットワークプロセッサ(特願2024-047372)

概要

○本技術の主要なアイデアは、データ・メモリを複数のシリコン層に分けて扱うことである。ここで、各層はビットのサブセットを格納する。各ビットのサブセットは、アンダーヴォルテージまたはパワーゲーティングのいずれかになる。その結果、動的量子化シナリオ（パワーゲーティング）またはノイズの多いデータ（アンダーボルテージ）となる。

○人工知能では、ノイズの多いデータや量子化された情報を使って計算を実行することが許容される。そのため、本技術における研究では、低消費電力技術の副作用は重要度の低いビットにのみ影響する。従って、全体的には計算性能を維持しながら、消費電力を劇的に削減することが可能となる。

実用化の可能性

本プロジェクトの成果として、以下のような実用化が可能である。

- 超低消費電力AIアプリケーション
- 電力消費量の調整レベルが要求される電力調整可能なデバイス

UBICからのメッセージ

本技術はAI処理における近似計算の許容度の性質を利用して、データを複数レイヤに分割格納することで、消費電力を大幅に改善するものです。近年のAI処理では、膨大な計算量による消費電力が課題となっていますが、その計算の特徴を逆に使って、一定の精度を維持しながら電力消費を劇的に抑える手法といえます。今後増大していくAI処理での活用が期待されます

研究概要図

