

論文内容の要旨

申請者氏名 ZHU GUANGXIAN

The demand for energy-efficient, high-performance computing in data-intensive applications like machine learning necessitates novel computing architectures. This paper contributes to power-efficient computing by leveraging new principles such as stochastic computing, innovative neural network topologies, and superconducting devices. Firstly, we introduce an ultra-compact calculation unit with temporal-spatial reconfigurability based on a Bisection Neural Network (BNN) topology. This design allows flexible partitioning of processing elements, achieving spatial reconfigurability by adjusting unit shapes and locations. Temporal reconfigurability is achieved through stochastic computing logic by adjusting bit-stream lengths, enhancing accuracy. Experimental results show our unit outperforms state-of-the-art approximate units in energy efficiency. Secondly, we present SuperSIM, a benchmarking framework for neural networks using superconducting Josephson devices, focusing on Adiabatic Quantum Flux Parametron based Processing-in-Memory architectures. SuperSIM offers detailed simulations and assessments, supporting various AQFP memory types, clocking methods, and both single and multi-bit designs. It provides precise measurements of energy consumption, delay, and area. Case studies examine the impacts of data precision, crossbar size, operating frequency, and clocking schemes, validating SuperSIM's effectiveness for advancing AQFP PIM chip development. These contributions advance energy-efficient computing architectures and neural network accelerators by introducing temporal and spatial reconfigurability and providing tools for optimizing superconducting technologies, paving the way for more efficient and scalable computing systems.

論文審査結果の要旨

申請者氏名 ZHU GUANGXIAN

機械学習のようなデータ集約型アプリケーションにおけるエネルギー効率の高い高性能コンピューティングの需要に応えるには、新しいコンピューティングアーキテクチャが必要である。本論文は、確率的コンピューティング、革新的ニューラルネットワークトポロジ、超伝導デバイスなどの新しい原理を活用して、電力効率の高いコンピューティングに貢献している。第1に、二分ニューラルネットワーク(BNN)トポロジに基づく、時間的・空間的再構成可能性を備えた超小型計算ユニットを提案している。本設計により、処理要素を柔軟に分割し、ユニットの形状と場所を調整することで空間的な再構成を可能としている。時間的再構成可能性は、ビットストリームの長さを調整して精度を高める確率的コンピューティングロジックにより実現している。実験結果は、本ユニットがエネルギー効率の点で最先端の近似ユニットよりも優れていることを示している。第2に、断熱量子フラックスパラメロンベースのメモリ内処理アーキテクチャに焦点を当てた、超伝導ジョセフソンデバイスを使用したニューラルネットワークのベンチマークフレームワークSuperSIMを提案している。SuperSIMは、様々なAQFPメモリタイプ、クロッキング方法、シングルビットとマルチビットの両方の設計をサポートする詳細なシミュレーションと評価を提供し、エネルギー消費、遅延、面積を正確に測定できる。ケーススタディでは、データ精度、クロスバーサイズ、動作周波数、クロッキングスキームの影響を調べ、AQFP PIMチップ開発の進歩に対するSuperSIMの有効性を検証している。以上の貢献は、時間的・空間的再構成可能性を導入し、超伝導技術を最適化するツールを提供することを通じて、エネルギー効率の高いコンピューティングアーキテクチャとニューラルネットワークアクセラレータが進歩し、より効率的でスケーラブルなコンピューティングシステムに至る道を開いている。

以上、本論文は学術上、實際上寄与するところが少なくない。よって、本論文は博士(工学)の学位論文として価値あるものと認める。