

Power-aware Neuromorphic Architecture with Partial Voltage Scaling 3D Stacking Synaptic Memory

Ngo-Doanh NGUYEN, Ryoji KOBAYASHI, Khanh N. DANG, Abderazek Ben ABDALLAH

Supervisor: Khanh N. DANG

Introduction · 導入

Spiking Neural Networks (SNNs) mimic the biological neural systems, which have low power and high performance. This research aims to leverage the low-power advantage of spiking neural networks using 3-D hardware architectures and low-power design techniques.

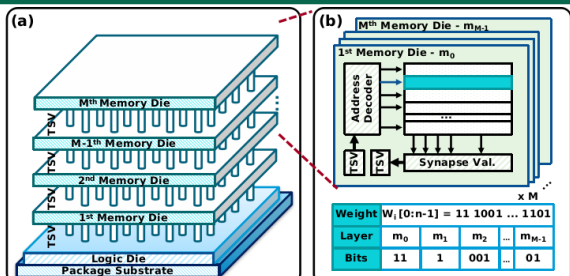
スパイクング・ニューラル・ネットワーク (SNNs) は、低消費電力かつ高性能である生物学的神経システムを模倣しています。本研究では、3次元ハードウェア・アーキテクチャと低消費電力設計技術を用いて、スパイクング・ニューラル・ネットワークの低消費電力の利点を活用することを目指します。

Methodology · 方法論

The idea is to distinguish the critical levels of synaptic weights and then isolate them using stacking layers of 3-D architectures. As a result, the side effects of low-power techniques only affect low-important bits. Hence, it dramatically reduces power consumption while maintaining its performance.

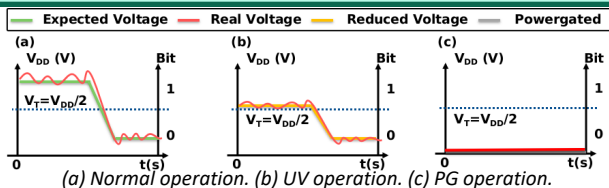
シナプスの重みの重要度を区別し、3次元構造の層を使ってそれらを分離するといった方法を用います。その結果、低消費電力技術に伴う問題は、重要性の低いビットにしか影響しません。したがって、性能を維持しつつ、消費電力を劇的に削減することができます。

Overview Hardware Architecture · 概要アーキテクチャ



(a) 3次元構造。 (b) ビット配置。

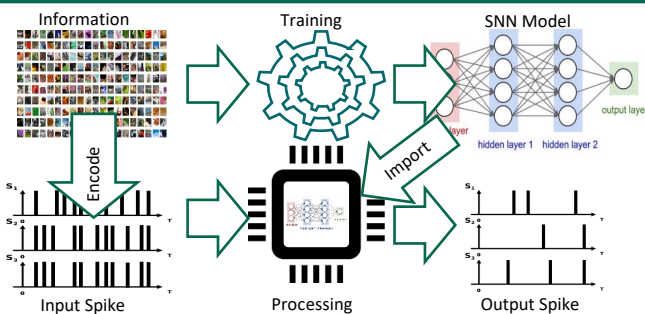
Side effect of low-power techniques · 低電力技術の副作用



Reducing voltage may flip bit and powergating makes bit zeroed.

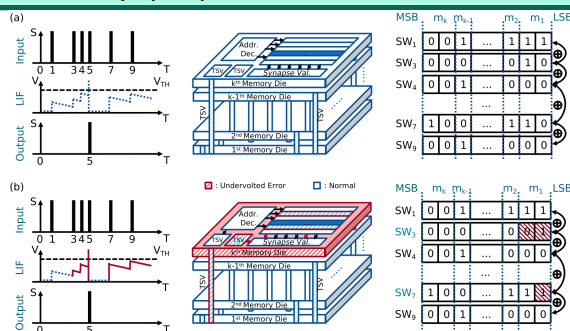
低電圧化はビットの値を反転させる可能性があり、パワーゲーティングは各ビットの値をゼロにしてしまいます。

Implementation Flow · 導入の流れ



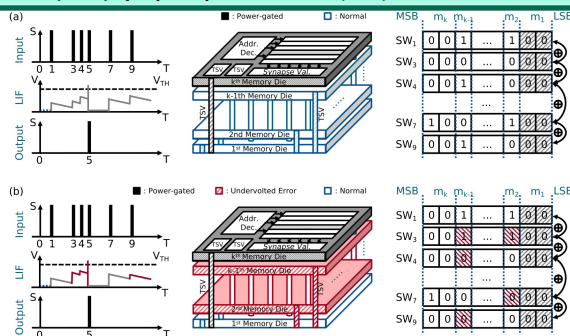
ソフトウェア上で重みを学習後、ハードウェア上書き込みます

UV synaptic operations · UVのシナプスの操作



(a) 通常動作。 (b) UV動作

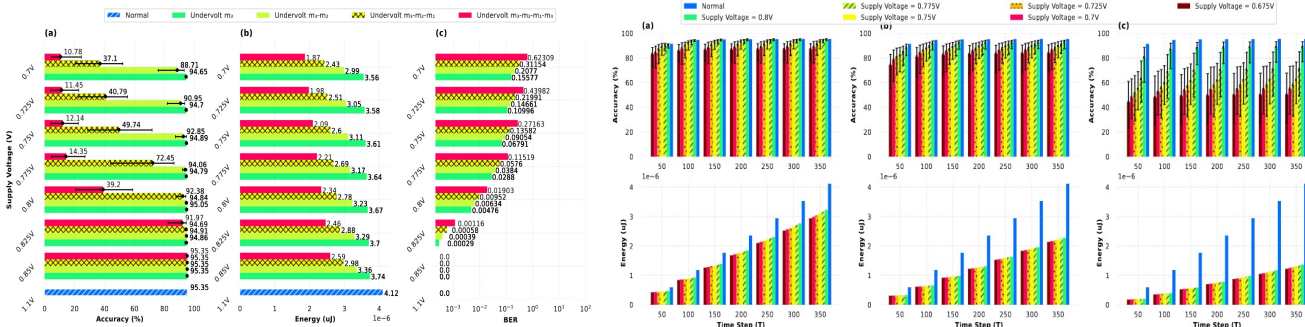
PG (+ UV) synaptic operations · PG(+UV)のシナプスの操作



(a) PG動作。 (b) PG & UV動作。

Experimental Results: Accuracy, Energy, and Bit-Error-Rate · 実験結果: 精度、エネルギー、ビットエラー率

MNIST dataset - Three perception layers of SNN [784 48 10] · MNIST データセット · SNN の3つの認識層 [784 48 10]



(a) 精度。 (b) 消費エネルギー。 (c) ビットエラー率。

Accuracy vs Energy with UV & PG:
(a) UV 2 layers. (b) PG 1 layer & UV 2 layers. (c) PG 2 layers & UV 2 layers.

精度 vs エネルギー · UV & PG:
(a) UV 2層。 (b) PG 1層 & UV 2層。 (c) PG 2層 & UV 2層。