Graduate School Fair 2024
[Poster ID: 06] Power-efficient AI Computing with Approximate Spiking Neural Networks

Department of Computer Science and Engineering, CE Division

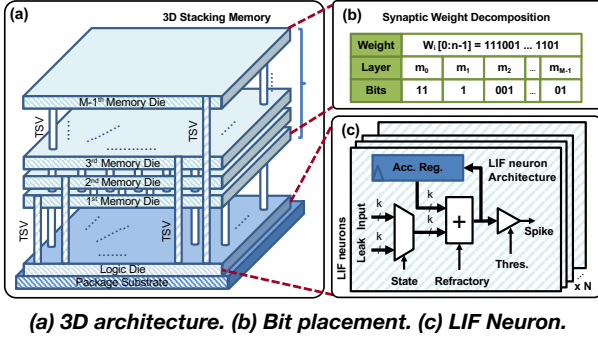*Ryoji Kobayashi, Hariprakash Subbaiah Ravi, Khanh N. Dang*

Abstract

Spiking Neural Networks (SNNs) are characterized by their low-power operation and noise resilience. This research aims to leverage these properties by applying the approximation technique to the neuron cell and memory system. As a result, with the proper approximation, energy consumption is reduced by up to 55.9% while having a 1.6% accuracy loss when combining approximate neuron cells and approximate memory.
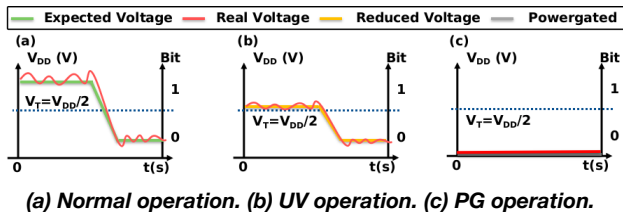
Methodology

The idea is to apply the Undervolting (UV) and Power-gating (PG) techniques to the memory system and approximate the arithmetic operator in the neuron cell. Here, the synaptic weights are distributed to different layers of 3D stacking memory, and an adder in the Leaky integrate-and-fire (LIF) neuron is replaced with an approximate adder. Since the supply voltage can be controlled layer-wise, the bit errors resulting from low voltage operation are restricted in LSBs. Therefore, combining both realizes a low-power neuromorphic system.
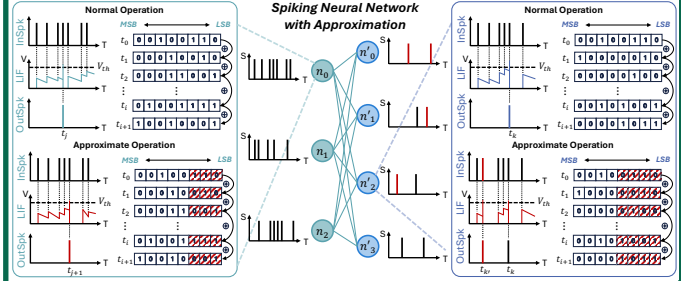
Overview Hardware Architecture

(a) 3D architecture. (b) Bit placement. (c) LIF Neuron.

Side effects of voltage scaling techniques

(a) Normal operation. (b) UV operation. (c) PG operation.

Reducing the voltage of memory cells may induce bit-flips, and power-gating makes bits zeroed.
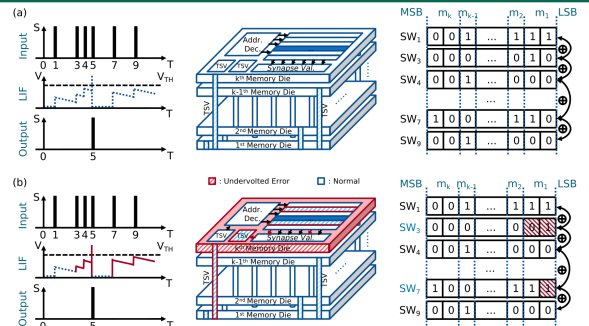
SNN with approximate neuron cells
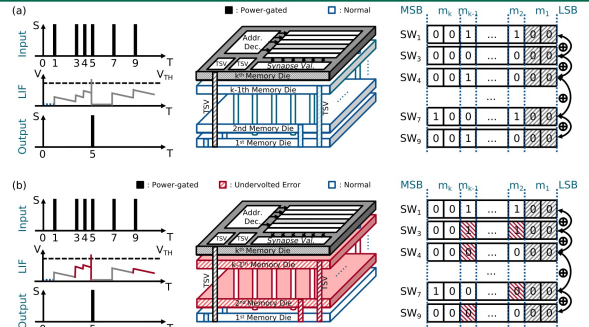

Comparison of neuron behaviors in SNN.

Approximation operations may affect the spike timing of neurons.

UV synaptic operations

(a) Normal synaptic operation. (b) UV synaptic operation.
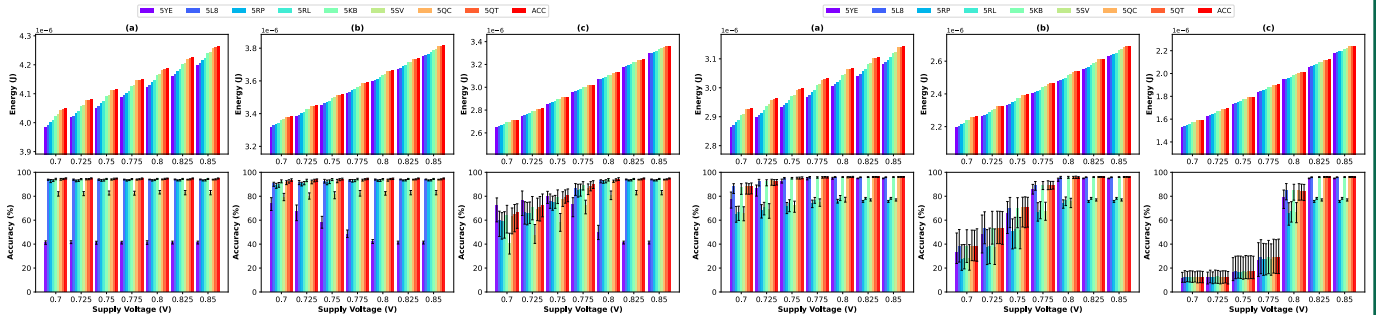
PG & UV synaptic operations

(a) PG synaptic operation. (b) PG & UV synaptic operation.

Experimental Results: Accuracy, Energy

MNIST dataset - Three perception layers of SNN [784 48 10]
(ACC: Accurate adder, others: Approximate adder)

Accuracy vs Energy with UV:
(a) UV one layer. (b) UV two layers. (c) UV three layers.

Accuracy vs Energy with UV & PG:
(a) UV one layer & PG one layer. (b) UV two layers & PG one layer.
(c) UV three layers & PG one layer.

Although the accuracy is maintained while the low voltage region is limited in LSBs, it drastically decreases when MSBs begin to be affected.