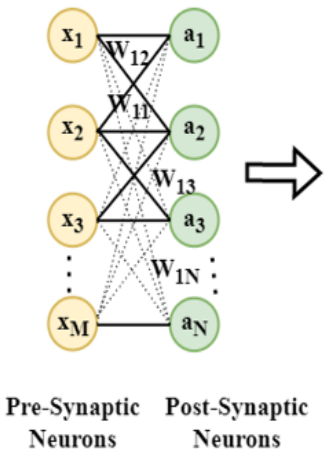
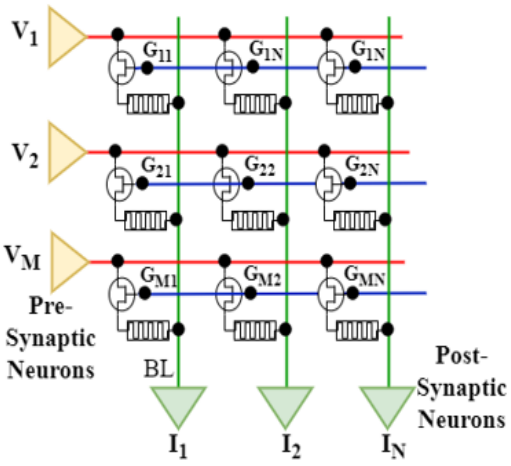
Master Thesis: Fault-Tolerant Techniques for Computation in Memory for Deep Learning Application

Neuron Connection



CiM for NN Acceleration



Mapping of Neural Networks (NNs) on CiM architecture:

Weight (W): Conductance (G),

Pre-synaptic(x): input to the crossbar (V),

post-synaptic (a): output of crossbar (I)

$$I_j = \sum_{i=1}^{N} V_i \, Gij$$

Matrix vector multiplication (MVM) is one of the most frequently performed operations in neural network (NN) hardware accelerators. The computation-in-memory (CiM) architecture is a promising alternative to conventional processor-centric computing architecture, which can efficiently perform MVM operations. However, the memristive devices used in CiM architecture suffer from different defects or faults. This can negatively impact the inference accuracy of deep learning applications. Thus, it is necessary to have fault-tolerant strategies such as error correction coding (ECC), fault-aware training, sensitivity analysis, circuit level and architecture level strategy, etc.

We want to develop an efficient fault-tolerant strategy (in terms of computation, memory, area, latency, energy, etc.) for CiM architecture that can enhance the accuracy of NNs under faulty scenarios. We are not limited to CiM only; we can also extend the idea for a digital hardware-based accelerator (GPU, TPU, etc.) for NN applications.

Skills required for the thesis

- Basic understanding of computer architecture, digital hardware logic design
- Background in deep learning and its implementation using Pytorch
- Programming skills: Python, Verilog, C/C++

Skills acquired within the thesis

- Opportunity to contribute to the cutting edge of fault-tolerant NN accelerators
- Gain experience in developing digital logic design using Verilog HDL
- Gain experience in evaluating NN performance for fault-free and faulty scenario
- Potential to publish findings in a top-tier Design/Test/EDA conference or journal

References

1. A. Shafiee et al., "ISAAC: A Convolutional Neural Network Accelerator with In-Situ Analog Arithmetic in Crossbars," 2016 ACM/IEEE 43rd ISCA. (**Basics of CiM**)
2. https://arxiv.org/pdf/2209.12260.pdf (**Quantized NN on CiM**)
3. F. Su, C., "Testability and Dependability of AI Hardware: Survey, Trends, Challenges, and Perspectives," in IEEE Design & Test (**Fault Tolerant Techniques**)
4. https://www.youtube.com/watch?v=yYwczGl4AS8&ab_channel=ShimengYu (**NeuroSim**)

Contact

Surendra Hemaram (Research Scholar, ITEC-CDNC, KIT): surendra.hemaram@kit.edu and Shaan Mangadahalli (**shanmukha.siddaramu@kit.edu**).

This work needs to be done in **English**. I am also available to deliver a short presentation online or in-person to explain the topic in detail.

Karlsruhe Institute of Technology

Prof. Dr. Mehdi B. Tahoori
ITEC, CDNC, KIT
Haid-und-Neu-Str. 7, Karlsruhe 76131

Karlsruhe Institute of Technology