MIM WEBINARS

AN IN-MEMORY COMPUTING SERIES

Next Talk: 21/March/2022, 4-5:30pm CET

NEUROSIM: A BENCHMARK FRAMEWORK OF COMPUTE-IN-MEMORY HARDWARE ACCELERATORS FROM DEVICES/CIRCUITS TO ARCHITECTURES/ALGORITHMS

Prof. Shimeng Yu, Georgia Institute of Technology

Compute-in-memory (CIM) is a new paradigm for machine learning hardware acceleration. DNN+NeuroSim is an integrated framework to benchmark CIM accelerators for deep neural network (DNN), with hierarchical design options from device-level, to circuit-level and up to algorithm-level. NeuroSim is a C++ based circuit-level macro model, which can achieve fast early-stage pre-RTL design exploration (compared to a full SPICE simulation). It takes design parameters including memory types (includes SRAM, RRAM, PCM, MRAM and FeFET), non-ideal device parameters, transistor technology nodes (from 130 nm to 7nm), memory array size, training dataset and traces to estimate the area, latency, dynamic energy, leakage power. A python wrapper is developed to interface NeuroSim with deep learning platforms Pytorch, to support flexible network topologies including VGG, DenseNet and ResNet for CIFAR/ImageNet classification. It supports weight/activation/gradient/error quantization in algorithm, and takes non-ideal properties of synaptic devices and peripheral circuits, in order to estimate training/inference accuracy. The framework is open-sourced and publicly available on GitHub https://github.com/neurosim/. NeuroSim’s user community is growing. This talk aims for a broader education to the community, and help the researchers to use/modify the code more flexibly for their own research purposes. More information about the event and the speaker: https://www.ict.tuwien.ac.at/staff/taherinejad/MiM/next.html