BETR Solid State Technology and Devices Seminar

Friday, October 13, 2023

Edge Intelligence with Neuromorphic Computing: From Algorithms to Hardware Design

Priya Panda

Assistant Professor
Department of Electrical Engineering
Yale University

Abstract: Spiking Neural Networks (SNNs) have emerged as an alternative to deep learning especially for edge computing due to their huge energy efficiency benefits on neuromorphic hardware. In this presentation, I will discuss the roadmap of current activities in the SNN algorithm and hardware design space, especially in reference to compute-in-memory accelerators. In the first half, I will talk about the importance of temporal dimension in SNNs which unlock unique behavior such as, robustness and privacy and bring in huge benefits in terms of latency, energy, and accuracy in different applications like video segmentation, human activity recognition, event sensing among others. In the second half, I will delve into the hardware perspective of SNNs when implemented on compute-in-memory (CiM) accelerators with our recently proposed SpikeSim benchmarking tool. It turns out that the multiple timestep computation in SNNs can lead to extra memory overhead and exacerbates the effect of CiM non-idealities that annuls all the compute-sparsity related advantages. I will highlight some techniques such as, input-aware early time-step exit and temporally evolving batch normalization to reduce the overhead. I will discuss an algorithm-hardware co-search methodology that explores the design space of CiM hardware and the neural network topology together with a search-based optimization to yield best performance-energy efficiency tradeoffs. Finally, I will discuss a future landscape of CiM device-circuit-system co-design for SNNs that includes investigating emerging devices for reducing on-chip memory for neuronal computations and efficient dot-product implementations.

Bio: Dr. Priya Panda is an assistant professor in the electrical engineering department at Yale University, USA. She received her B.E. and Master's degree from BITS, Pilani, India in 2013 and her Ph.D. from Purdue University, USA in 2019. During her PhD, she interned in Intel Labs where she developed large scale spiking neural network algorithms for benchmarking the Loihi chip. She is the recipient of the 2019 Amazon Research Award, 2022 Google Research Scholar Award, 2022 DARPA Riser Award, 2023 NSF CAREER Award, 2023 DARPA Young Faculty Award. She has also received the 2022 ISLPED Best Paper Award and 2022 IEEE Brain Community Best Paper Award. Her research interests lie in Neuromorphic Computing, Spiking Neural Networks, and In-Memory Computing.