EFFICIENT MACHINE LEARNING: ALGORITHMS-CIRCUITS-DEVICES CO-DESIGN

Prof. Hai (Helen) Li, Duke University

Following technology advances in high-performance computation systems and fast growth of data acquisition, machine learning, especially deep neural networks (DNNs), made remarkable success in many research areas and applications. Such a success, to a great extent, is enabled by developing large-scale network models that learn from a huge volume of data. The deployment of such a big model, however, is both computation-intensive and memory-intensive. Though the research on hardware acceleration for neural networks has been extensively studied, the progress of hardware development still falls far behind the upscaling of DNN models at the software level. The holistic co-design across algorithm, circuit, and device levels emerges more important for execution acceleration, energy efficiency, and design flexibility. In this presentation, we will present our studies on how to optimize the training process for sparse and low-precision network models for general platforms. We will also discuss the memristor-based computing engine designs optimized for DNN inference and training. More information about the event and the speaker:

https://www.ict.tuwien.ac.at/staff/taherinejad/MiM/next.html