Big AI for Small Devices

Abstract: As artificial intelligence (AI) transforms industries, state-of-the-art models have exploded in size and capability. However, deploying these models on resource-constrained edge devices remains a significant challenge. Smartphones, wearables, and IoT sensors face stringent limitations on computing, memory, power, and communication, creating a big gap between demanding AI models and edge hardware capabilities that hinders the deployment of intelligence. In this talk, we will re-examine techniques to bridge this gap and embed big AI on small devices. We will begin by discussing how the properties of various hardware platforms impact the design strategies of efficient deep neural network (DNN) models, such as quantization and pruning. Next, we will discuss techniques aimed at reducing the inference and training costs of distributed collaborative edge AI systems. Finally, we will delve into the underlying design philosophies and their evolution toward efficient, scalable, robust, and secure edge computing systems.

Biography: Hai (Helen) Li is the Marie Foote Reel E'46 Distinguished Professor and Department Chair of the Electrical and Computer Engineering Department at Duke University. She received her B.S. and M.S. degrees from Tsinghua University, and her Ph.D. degree from Purdue University. Her research interests include neuromorphic circuits and systems for brain-inspired computing, machine learning acceleration and trustworthy AI, conventional and emerging memory design and architecture, and software and hardware co-design. Dr. Li served/serves as the Associate Editor-in-Chief and Associate Editor for multiple IEEE and ACM journals. She was the General Chair or Technical Program Chair of multiple IEEE/ACM conferences and the Technical Program Committee member of over 30 international conference series. Dr. Li has received many awards, including the IEEE Edward J. McCluskey Technical Achievement Award, Ten Year Retrospective Influential Paper Award from ICCAD, TUM-IAS Hans Fischer Fellowship from Germany, ELATE Fellowship, nine best paper awards, and another ten best paper nominations from IEEE/ACM. Dr. Li is a fellow of IEEE, ACM, and NAI.