
HARDWARE/SOFTWARE CO-DESIGN FOR AI SYSTEMS

Dr. Yiran Chen, Professor in the Department of Electrical and Computer Engineering Duke University

Abstract: The rapid growth of modern neural network (NN) models' scale generates ever-increasing demands for high computing power of artificial intelligence (AI) systems. Many specialized computing devices have been also deployed in the AI systems, forming a truly application-driven heterogeneous computing platform. This talk discusses the importance of hardware/software co-design in AI system designs. We first use resistive memory based NN accelerators to illustrate the design philosophy of heterogeneous AI computing systems, and then present several hardware friendly efficient NN model design techniques. We also briefly introduce the automation of the co-design flow, e.g., neural architecture search. A research roadmap of our group in the relevant topics is given at the end of the talk.

Biography: Yiran Chen received B.S (1998) and M.S. (2001) from Tsinghua University and Ph.D. (2005) from Purdue University. After five years in industry, he joined University of Pittsburgh in 2010 as Assistant Professor and then promoted to Associate Professor with tenure in 2014, held Bicentennial Alumni Faculty Fellow. He is now the Professor of the Department of Electrical and Computer Engineering at Duke University and serving as the director of NSF Industry–University Cooperative Research Center (IUCRC) for Alternative Sustainable and Intelligent Computing (ASIC) and the co-director of Duke Center for Computational Evolutionary Intelligence (CEI). His group focuses on the research of new memory and storage systems, machine learning and neuromorphic computing, and mobile computing systems. Dr. Chen has published 1 book and more than 400 technical publications and has been granted 96 US patents. He serves or served the associate editor of more than ten international academic transactions/journals and served on the technical and organization committees of more than 60 international conferences. He is now serving as the Editor-in-Chief of the IEEE Circuits and Systems Magazine. He received seven best paper awards, one best poster award, and fourteen best paper nominations from international conferences and workshops. He is a distinguished lecturer of IEEE CEDA and listed in the HPCA Hall of Fame. He is a Fellow of the ACM and IEEE.

