PURDUE UNIVERSITY | Elmore Family School of Electrical and Computer Engineering

AI at the Edge: Nanotechnology-Inspired Artificial Intelligence Hardware

Haitong Li
Stanford University

Tuesday, March 1, 2022
1:30 PM • MSEE 239

Zoom Info: https://purdue-edu.zoom.us/j/93328074360 ~ Meeting ID: 933 2807 4360

Abstract

Ubiquitous artificial intelligence (AI) promises to empower broad edge applications, from health-monitoring and augmented-reality wearables to autonomous robots. To fulfill the big promises, the future electronics must deliver unprecedented energy efficiency with new functionalities, enabling real-time adaptation and lifelong learning at the edge. Such growing demands simply cannot be met by isolated advancements in materials, devices, integrated circuits, and architectures, and requires cross-layer design for the next-generation AI hardware.

In this talk, I will present my work on nanotechnology-inspired AI hardware, realized by exposing and connecting the unique properties of emerging nanotechnologies at device and circuit levels all the way to the diverse AI model characteristics. I will illustrate the cross-layer design, integration, and creation of new compute kernels and chip architectures. First, I will present SAPIENS, the first integrated chip that enables on-chip, one-shot learning with scarce and never-before-seen data, built with 65,536 RRAMs (resistive switching memories) and mixed-signal silicon CMOS. Next, towards emerging AI paradigms beyond neural networks, I will present the first experimental demonstration of three-dimensional (3D) nanokernels for hyper-dimensional (HD) computing, seamlessly translating the intrinsic device physics and the native 3D connectivity of multi-layer 3D memories into HD compute kernels. The cross-layer design methodology, grounded on my device-to-system modeling work, has also been applied to other case studies.

I will conclude by outlining my vision for future nanotechnology-inspired AI hardware, bridging the worlds of nanoelectronics and AI, to materialize energy-efficient, real-time machine intelligence at the edge.

Bio

Haitong Li is a Ph.D. candidate in Electrical Engineering at Stanford University, supervised by Professor H.-S. Philip Wong. He received M.S. in Electrical Engineering from Stanford University in 2017, and B.S. in Microelectronics from Peking University in 2015. His current research interests are in the broad area of nanotechnology-inspired artificial intelligence hardware. His research results include the first integrated chip for one-shot learning with memory-augmented neural networks, the first experimental demonstration of 3D nanokernels for emerging hyper-dimensional (HD) computing, and physics-based, hierarchical resistive RAM (RRAM) models widely used by academia and industry communities. Haitong serves as a committee member for the IEEE Electron Devices Society (EDS) VLSI Technology and Circuits Committee, and has held Research Intern positions at Arm Research and Facebook Reality Labs. He is a recipient of 2019 IEEE EDS PhD Student Fellowship and 2016 IEEE EDS Masters Student Fellowship.

Host: Professor Joerg Appenzeller ~ appenzeller@purdue.edu ~ 765-494-1076