Near and In-Memory Computing for AI: From Devices, Circuits to Architecture

The AI revolution is straining traditional computing systems, with data movement between memory and processors consuming excessive energy and limiting performance. Near and In-Memory Computing (NIMC) offers a breakthrough by enabling computation directly within or closer to the memory, unlocking orders-of-magnitude gains in speed and efficiency for AI workloads. This talk delves into how NIMC technologies—from resistive RAM (ReRAM) and Magnetic RAM (MRAM) to computational DRAM—are reshaping AI hardware.