Keynote Speech I
9:40–10:40, Monday, October 24, 2022

Hardware/Software Codesign for Machine Learning Acceleration with Silicon Photonics

Sudeep Pasricha
Department of Electrical and Computer Engineering
Colorado State University

Abstract

The massive data deluge from mobile, IoT, and edge devices, together with powerful innovations in data science and hardware processing, have established machine learning (ML) as the cornerstone of modern medical, automotive, industrial automation, and consumer electronics domains. Domain-specific ML accelerators such as Google's TPU and Apple's Bionic, now dominate CPUs and GPUs for energy-efficient ML processing. However, the evolution of these electronic accelerators is facing fundamental limits due to the slowdown of Moore's law and the reliance on metal wires, which already severely bottleneck computational performance today. Silicon photonics represents a promising post-Moore technological alternative to overcome these limitations. Not only can photonic interconnects fabricated in CMOS-compatible processes provide near speed of light transfers at the chip-scale, but photonic devices can now also perform computations entirely in the optical domain. In this talk, I will present my vision of how silicon photonics can drive an entirely new class of sustainable ML hardware accelerators that can provide orders of magnitude energy improvements over today's accelerators. I will discuss new directions in hardware/software codesign for ML acceleration with silicon photonics, with multi-objective goals related to power and energy minimization, variation tolerance, fault resilience, and secure computing.

Biography

Sudeep Pasricha received the B.E. degree in Electronics and Communication Engineering from Delhi Institute of Technology, India, in 2000, after which he spent several years working for STMicroelectronics, India/France, and Conexant, USA. He received his Ph.D. degree in Computer Science from the University of California, Irvine in 2008. He joined Colorado State University (CSU) in 2008 where he is currently a Walter Scott Jr. College of Engineering Professor in the Department of Electrical and Computer Engineering. He is a former University Distinguished Monfort Professor and Rockwell-Anderson Professor. He is currently also Chair of Computer Engineering and Director of the Embedded, High Performance, and Intelligent Computing (EPIC) Laboratory at CSU. His research focuses on the design of innovative software algorithms, hardware architectures, and hardware-software co-design techniques for energy-efficient, fault-tolerant, real-time, and secure computing, with applications to embedded, IoT, and cyber-physical systems. Prof. Pasricha has published more than 250 papers in peer-reviewed journals and conferences that have received seven best paper awards and six best paper nominations. He has filed for multiple patents and co-authored several books and book chapters. He has also given several keynotes, invited talks, and tutorials. His contributions have been recognized with various awards, including the George T. Abell Outstanding Research Faculty Award, IEEE-CS/TCVLSI Mid-Career Research Achievement Award, IEEE/TCSC Award for Excellence for a Mid-Career Researcher, AFOSR Young Investigator Award, ACM Technical Leadership Award, and ACM SIGDA Distinguished Service Award. He is currently the Vice Chair of ACM SIGDA and the Steering Committee Chair for the IEEE Transactions on Sustainable Computing. He is also a Senior Associate Editor for the ACM Journal of Emerging Technologies in Computing, and an Associate Editor with several ACM and IEEE journals. He has served as General Chair and Technical Program Chair of 12 conferences, Steering and Organizing Committee Member of 40+ conferences, and Technical Program Committee Member of 100+ conferences. He is an IEEE Senior Member, an ACM Distinguished Member, and an ACM Distinguished Speaker.