Department of Electrical Engineering
香港城市大學
City University of Hong Kong

University of Pittsburgh

CityU Architecture Lab for Arithmetic and Security (CALAS) Seminar Series

Efficient Programming on Heterogeneous Accelerators for Sustainable Computing

Prof. Peipei ZHOU, Assistant Professor
Department of Electrical and Computer Engineering, University of Pittsburgh

Abstract:

There is a growing call for increasingly agile computational power for edge and cloud infrastructure to serve the computationally complex needs of ubiquitous computing devices. One important challenge is addressing the holistic environmental impacts of these next-generation computing systems. A life-cycle view of sustainability for computing systems is necessary to reduce environmental impacts such as greenhouse gas emissions from these computing systems in different phases: manufacturing, operational, and disposal/recycling. My research investigates how to efficiently program and map widely used workloads on heterogeneous accelerators and seamlessly integrate them with existing computing systems towards sustainable computing.

In this talk, I will first discuss how new mapping solutions, i.e., composing heterogeneous accelerators within system-on-chip with both FPGAs and AI tensor cores, achieve orders of magnitude energy efficiency gains when compared to monolithic accelerator mapping designs for various applications, including deep learning, security, and others. Then, I will apply such novel mapping solutions to show how design space explorations are performed when composing heterogeneous accelerators in latency-through tradeoff analysis. I will further discuss how such mapping and scheduling can be applied to other computing systems, such as GPUs, to improve energy efficiency and, therefore, reduce the operational carbon cost. Finally, I will introduce the REFRESH FPGA chiplets, explain why REFRESH chiplets help reduce the embodied carbon cost, and discuss the challenges and opportunities.

Biography:

Prof. Peipei ZHOU is a tenure-track assistant professor in the Department of Electrical Computer Engineering at the University of Pittsburgh. She received her Ph.D. in Computer Science (2019) and M.S. in Electrical and Computer Engineering (2014) from UCLA, and her B.S. in Electrical and Computer Engineering (2012) from Southeast University. Her research investigates architecture, programming abstraction, and design automation tools for reconfigurable computing and heterogeneous computing. She has published 30 papers in IEEE/ACM computer system and design automation conferences and journals including FPGA, FCCM, DAC, ICCAD, ISPASS, TCAD, TODAES, TECS, IEEE Micro, etc. Her work has won the 2019 IEEE TCAD Donald O. Pederson Best Paper Award. Other awards include the 2023 ACM/IEEE IGSC Best Viewpoint Paper Finalist, the 2018 IEEE ISPASS Best Paper Nominee, and the 2018 IEEE/ACM ICCAD Best Paper Nominee.

18 Mar 2024 (Mon); 9:00am - 11:00am; P1402; https://cityu.zoom.us/j/96742093029