"Algorithm-Architecture-Hardware Co-Design in Computing Systems: From Chip Multicore to the Cloud"

Wenkai Guan

Ph.D. Candidate – ECE Dept., Marquette University

Thursday, February 16th, 2023
2:00 – 3:00 p.m. Olin 202
Sponsor: ECE Dept., and the Ph.D. Club

Abstract: The computational demands for training deep learning models doubled every three months recently. However, according to Moore's Law, the computational power available only doubled every two years. To bridge this demand-supply gap while optimizing energy consumption and carbon footprint, through my research, I propose a novel algorithm-architecture-hardware co-design cross-layer approach for computing systems: from chip multicore to the cloud. In this approach, I focused on multi-objective scheduling algorithms. Research results demonstrated that this approach outperforms state-of-the-art schedulers from industry and academia by 41.98% in energy delay product (EDP), 38.65% in energy usage, and 10.2% in job completion time. This is only the first step towards the way we must rethink and redesign energy-efficient computing systems - from chip multicore to the cloud - to support emerging Artificial Intelligence (AI) and Machine Learning (ML) applications.

Bio: Wenkai Guan received B.S. and M.S. degrees from the Wuhan University of Technology in China and the Marquette University, respectively. He was a research assistant at the Huazhong University of Science and Technology in China. He is a Ph.D. candidate at Marquette University in the Department of Electrical and Computer Engineering. His research passions include machine learning and datacenter, cloud computing, and multicore embedded systems. Mr. Guan won the Best Research Video Award at Design Automation Conference (DAC) 2021 and was the recipient of the DAC Young Fellowship in 2021 and 2022. Mr. Guan works with the Ph.D. Club.