COLLEGE OF ENGINEERING
COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

Dissertation Defense

Abraham Lamesgin Addisie

Domain-specific Architectures for Data-intensive Applications

Wednesday, May 6, 2020
2:00 pm – 4:00 pm
bluejeans.com/378816302

ABSTRACT: Graphs' versatile ability to represent diverse relationships, make them effective for a wide range of applications. For instance, search engines use graph-based applications to provide high-quality search results. Medical centers use them to aid in patient diagnosis. Most recently, graphs are also being employed to support the management of viral pandemics. Looking forward, they are showing promise of being critical in unlocking several other opportunities, including combating the spread of fake content in social networks, detecting and preventing fraudulent online transactions in a timely fashion, and in ensuring collision avoidance in autonomous vehicle navigation. Unfortunately, all these applications require more computational power than what can be provided by conventional computing systems. The key reason is that graph applications present large working sets that fail to fit in the small on-chip storage of existing computing systems, while at the same time they access data in seemingly random patterns, thus cannot draw benefit from traditional on-chip storage.

In this dissertation, we set out to address the performance limitations of existing computing systems so to enable emerging graph applications like those described above. To achieve this, we identified three key strategies: 1) specializing memory architecture, 2) processing data near its storage, and 3) message coalescing in the network. Based on these strategies, this dissertation develops several solutions: OMEGA, which employs specialized on-chip storage units with co-located specialized compute engines to accelerate the computation; MessageFusion, which coalesces messages in the interconnect; and Centaur, providing an architecture that optimizes the processing of infrequently-accessed data. Overall, these solutions provide 2.2x in performance improvements with negligible hardware overheads across a wide range of applications. Finally, we demonstrate the applicability of our strategies to other data-intensive domains by exploring an acceleration solution for MapReduce applications, which achieves a 4.0x performance speedup with negligible area and power overheads.

Chair: Prof. Valeria Bertacco