COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

DISSERTATION DEFENSE

AKSHITHA SRIRAMAN

Enabling Hyperscale Web Services

Friday, July 16, 2021
5:00 pm
Virtual (Passcode: 532409)

ABSTRACT: Current hardware and software systems were conceived at a time when we had scarce compute and memory resources, limited quantity of data and users, and easy hardware performance scaling due to Moore's Law. These assumptions are not true today. Today, emerging web services require data centers that scale to hundreds of thousands of servers, i.e., hyperscale, to efficiently process requests from billions of users. In this new era of hyperscale computing, we can no longer afford to build each layer of the systems stack separately. Instead, we must rethink the synergy between the software and hardware worlds from the ground up.

This dissertation rethinks the bridge between the software and hardware worlds to enable the hyperscale web services of tomorrow. In this dissertation, I first contribute *µSuite*, a benchmark suite of modern web services that enables studying web service software and hardware behaviors. Next, I propose *µTune*, a software threading framework that is aware of the overheads induced by the underlying hardware's constraints. Finally, I introduce *SoftSKU*, *µNotify*, and *Accelerometer* – my proposals to answer the question of: How should we build data center hardware for emerging software paradigms in the post-Moore era?

This dissertation's contributions have improved the performance of real-world hyperscale web services that currently serve billions of users, saving millions of dollars and meaningfully reducing the global carbon footprint. Additionally, some of the hardware design proposals have influenced real-world processor architectures.

CHAIR: Prof. Thomas Wenisch