



DISSERTATION DEFENSE



Vidushi Goyal

Building User-Driven Devices

Thursday, October 13, 2022

9:00 – 11:00am

Hybrid Event – 3725 BBB

[Zoom](#) – Passcode: 311094

ABSTRACT: Edge devices like smartphones and wearables have become an integral part of our daily routines. Their ubiquitous and portable nature allows them to operate in any sort of environment. They can be deployed in the wild or at home without requiring a constant power source plugged into them. However, the small form factor and resource-constrained nature of edge devices limits their computation capabilities, and thus significantly impacts the efficiency of tasks on edge devices. The application efficiency directly affects the quality of user-experience; thus, these shortcomings of the edge device impact the user-experience as well.

In this dissertation, we develop solutions to address these limitations of edge devices to enhance the performance and energy efficiency for a wide range of user edge applications. Our proposed solutions are either low-cost alternatives that can replace expensive silicon, or re-purpose the already in-use silicon to extract extreme efficiencies, thus lowering the total cost of ownership of edge devices. Our solutions are driven by two key strategies: 1) cross-component optimizations across the system and 2) leveraging user information and preferences in the hardware. We first study edge applications with tiny microcontrollers and sensors. We propose an automated framework to find optimal sensing rates for power-intensive sensors, governed by low-power sensors based on users' sensing capabilities. This elongates battery life with minimal impact on the perceivable user-experience. In our second solution, we customize the ML-based image recognition application for each user by creating small and accurate user-specific ML models on the edge device. This significantly lowers computation demands and memory footprint without impacting user accuracy. In the next work, we leverage the user history and profile information to decompose the monolithic recommendation model into separate user and item models. The user model processes user information in a lightweight manner on edge, and its computation is reused by the item model, processing 100s of items at the data-center. Finally, we build a low-cost system-in-package based interconnect architecture using the 2.5D stacking interposer technology, which can replace the expensive system-on-chip. The architecture exposes high-bandwidth links of the interposer to the applications, enhancing performance and energy efficiency.

CO-CHAIRS: Prof. Valeria Bertacco and Prof. Reetuparna Das