THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY | SCHOOL OF ENGINEERING | HKUST - ACCESS AI Chip Center for Emerging Smart Systems | HKSTP

HKUST Engineering x HKSTP Distinguished Speaker Series

Prof. Onur Mutlu

Professor of Computer Science, ETH Zurich

Onur Mutlu is a Professor of Computer Science at ETH Zurich. He is also a faculty member at Carnegie Mellon University, where he previously held the Strecker Early Career Professorship. His current broader research interests are in computer architecture, systems, hardware security, and bioinformatics.

Moderator: Prof. Tim Cheng, Dean, HKUST School of Engineering

7TH OCT 2020 | 5:00PM (HK TIME) | VIA ZOOM

REGISTRATION

HTTPS://BIT.LY/3HMH2TA

Explore Job Opportunities at HKUST ACCESS

Intelligent Architectures for Intelligent Machines

Computing is bottlenecked by data. Large amounts of application data overwhelm storage capability, communication capability, and computation capability of the modern machines we design today. As a result, many key applications' performance, efficiency and scalability are bottlenecked by data movement.

We describe three major shortcomings of modern architectures in terms of 1) dealing with data, 2) taking advantage of the vast amounts of data, and 3) exploiting different semantic properties of application data. We argue that an intelligent architecture should be designed to handle data well. We show that handling data well requires designing architectures based on three key principles: 1) data-centric, 2) data-driven, 3) data-aware. We give several examples for how to exploit each of these principles to design a much more efficient and high performance computing system.

We will especially discuss recent research that aims to fundamentally reduce memory latency and energy, and practically enable computation close to data, with at least two promising novel directions: 1) performing massively-parallel bulk operations in memory by exploiting the analog operational properties of memory, with low-cost changes, 2) exploiting the logic layer in 3D-stacked memory technology in various ways to accelerate important data-intensive applications. We discuss how to enable adoption of such fundamentally more intelligent architectures, which we believe are key to efficiency, performance, and sustainability.

We conclude with some guiding principles for future computing architecture and system designs. Throughout the talk, we will point out how open source hardware can enable innovation in and adoption of the paradigms we introduce.