UCF | Electrical and Computer Engineering

Spring 2018 Seminar Series

Architectural Challenges and Innovation for Accelerating Big Data Analytics

THURSDAY APRIL 5, 2018
10:30 AM – HEC 450

Big data applications are the driving force behind the current revolutions in computing architecture and systems design. For instance, modern server systems deploy non-volatile memory-based storage devices to provide lower data transfer latency from storage to compute nodes. Graphics processing units (GPUs) are also widely employed as primary accelerators to handle huge computation demands from a wide range of data-intensive applications. As demands for high-performance data processing platforms increases in Big Data era, hardware architecture and systems design for efficient data use and data transfer is becoming more critical.

In this talk, I will present the architectural challenges and innovation for improving the performance of big data applications. First, I will discuss the intelligent storage systems architecture, which can accelerate data computation and reduce data movement cost. As modern flash memory-based storage devices equip general-purpose embedded processors, storage devices have potential to work as active compute systems using this computation resources. I will present the dynamic near data processing framework, which opportunistically enables computation near data in the modern storage systems. Second, I will talk about the hardware support for efficient data use in GPU memory hierarchy. I will reveal GPU memory hierarchy does not work effectively for data analytics applications. Due to the burst data requests from many concurrent threads, GPU cache suffers significant cache contention and premature data eviction. I will present the per-load cache management scheme that improves the low cache utilization of GPU. Then, I will conclude with my future research interest for near data accelerator architecture for exascale data processing.

Gunjae Koo

University of Southern California

Gunjae Koo is a Ph.D. candidate in the Department of Electrical Engineering at the University of Southern California. He received his M.S. and B.S. degrees in Electrical Engineering and Computer Science from Seoul National University, South Korea. His research interest is in computer systems architecture for data processing acceleration and spans intelligent storage systems, memory systems, parallel processor architecture, and FPGA. He has published 11 papers, including four top-tier computer architecture conference papers at ISCA, MICRO, and HPCA. Prior to starting the Ph.D. study, he worked as a senior research engineer on system-on-chip design for storage devices and digital TV platforms at LG Electronics for eight years. He also worked on memory controller architecture for multi-core server processors at Intel.