Rethinking Memory System Design for Data-Intensive Computing

Onur Mutlu

Carnegie Mellon University, United States

Abstract. The memory system is a fundamental performance and energy bottleneck in almost all computing systems. Recent system design, application, and technology trends that require more capacity, bandwidth, efficiency, and predictability out of the memory system make it an even more important system bottleneck. At the same time, DRAM and flash technologies are experiencing difficult technology scaling challenges that make the maintenance and enhancement of their capacity, energy-efficiency, and reliability significantly more costly with conventional techniques.
In this talk, we examine some promising research and design directions to overcome challenges posed by memory scaling. Specifically, we discuss three key solution directions: 1) enabling new memory architectures, functions, interfaces, and better integration of the memory and the rest of the system, 2) designing a memory system that intelligently employs multiple memory technologies and coordinates memory and storage management using non-volatile memory technologies, 3) providing predictable performance and QoS to applications sharing the memory/storage system. If time permits, we may also briefly describe our ongoing related work in combating scaling challenges of NAND flash memory.

Biography

Onur Mutlu is the Dr. William D. and Nancy W. Strecker Early Career Professor in ECE (and, by courtesy, CSD) at Carnegie Mellon University. His research interests are in computer architecture and systems, especially in the interactions between languages, operating systems, compilers, and microarchitecture. He was previously a researcher in the Computer Architecture Group at Microsoft Research (from 2006 to 2009) and a Research Fellow at the University of Texas at Austin (from 2007 to 2009). Before that, he was a member of the HPS Research Group at the University of Texas at Austin, where he received his PhD in 2006. His PhD dissertation was on efficient runahead execution processors. He received his BS degrees in computer engineering and psychology from the University of Michigan, Ann Arbor, in 2000 and his MS degree in electrical and computer engineering from UT-Austin in 2002. He worked at Intel Corporation during the summers of 2001-2003 and at Advanced Micro Devices during the summers of 2004-2005. He received several honors for his research, including the University of Texas George H. Mitchell Award for Excellence in Graduate Research in 2005, Microsoft Gold Star Award in 2008, NSF CAREER Award in 2010, ASPLOS 2010 Best Paper Award, VTS 2010 Best Paper Award, ICCD 2012 Best Paper Award, RTAS 2014 Best Paper Award, 2011 IEEE Computer Society TCCA Young Computer Architect Award, 2012 Intel Early Career Faculty Honor Program Award, 2012 Carnegie Mellon College of Engineering George Tallman Ladd Research Award, and several "computer architecture top pick" paper recognitions by the IEEE Micro magazine.