October 8-12, 2018
047004- Memory Systems: Fundamentals, Recent Research, Challenges, Opportunities
Prof. Onur Mutlu, ETH Zurich

General Information

Instructor: Prof. Onur Mutlu

Teaching assistant: Roman Kaplan

Prerequisite: See below

Lectures: 26 Hours

Academic points: 2pts

Course tentative schedule:

- 10:00-12:00 - Part I
- 13:00-14:30 - Part II
- 15:00-17:00 - Part III
- Friday the (12th) 9:30-12:30

Final Examination: Monday, Oct. 15, 10:00-12:00. Written examination on the course material.

Course Registration:

- Undergraduate students: registration to the course is through UG https://ug3.technion.ac.il/rishum/ -> [Details to follow]
- Graduate students: contact Danit - danitc@ef.technion.ac.il to register
- Other: contact Liraz liraz@tce.technion.ac.il

Course Abstract and Outline

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, performance, and reliability significantly more costly with conventional techniques. In fact, recent reliability issues with DRAM, such as the RowHammer problem, are already threatening system security and predictability. We are at the challenging intersection where issues in memory reliability and performance are tightly coupled with not only system cost and energy efficiency but also system security.

In this course, we first provide a comprehensive overview of memory systems, taking an approach that covers both fundamentals and recent research. We will first introduce fundamental principles and ideas, covering DRAM and emerging memory technologies as well as many architectural concepts and ideas related to memory organization, memory control, processing-in-memory, and memory latency / energy / bandwidth / reliability / security / QoS. We will discuss major challenges facing modern memory systems (and the computing platforms we currently design around the memory system) in the presence of greatly increasing demand for data and its fast analysis. We will examine some promising research and design directions to overcome these challenges.

On the research-related part of course (sprinkled across topical lectures), we will discuss at least the following key research topics in detail, focusing on both open problems and potential solution directions:

1. Fundamental issues in memory reliability and security and how to enable fundamentally secure, reliable, safe architectures
2. Enabling data-centric and hence fundamentally energy-efficient architectures that are capable of performing computation near data
3. Reducing both latency and energy consumption by tackling the fixed-latency/energy mindset
4. Enabling emerging memory technologies
5. Enabling predictable and QoS-aware memory systems

If time permits, we will also discuss research challenges and opportunities in enabling emerging NVM (non-volatile memory) technologies and scaling NAND flash memory and SSDs (solid state drives) into the future.

Topics

Many of the topics that will be covered will be be based on the following graduate-level course at ETH Zurich: https://safari.ethz.ch/architecture/fall2017/doku.php?id=schedule

The following are topics the course will discuss in varying detail:

1. Fundamentals, Memory Hierarchy, Caches
2. Cache Management and Memory Parallelism
3. Main Memory and DRAM Fundamentals
4. DRAM Operation, Memory Control & Memory Latency
5. Low-Latency DRAM and Processing In Memory
6. Emerging Memory Technologies
7. Memory Interference and Quality of Service
8. Multi-Core Cache Management (if time permits)
9. Latency Tolerance and Prefetching (if time permits)
10. Multiprocessors, Consistency, Coherence (if time permits)
11. Interconnects: Fundamentals and Recent Research (if time permits)
12. NAND Flash Memory (if time permits)
13. Potpourri of Recent Research Papers

Schedule

Day 1 (Monday)

- 10:00-12:00 - Fundamentals, Memory Hierarchy, Caches
- 13:00-14:30 - Cache Management, Memory Parallelism
- 15:00-17:00 – Main Memory and DRAM: Fundamentals, Reliability, Security and Safety

Day 2 (Tuesday)

- 10:00-12:00 - DRAM Operation, Memory Control & Memory Latency
- 13:00-14:30 - Low-Latency DRAM and Processing In Memory
- 15:00-17:00 – Emerging Memory Technologies

Day 3 (Wednesday)

- 10:00-12:00 - NAND Flash Memory
- 13:00-14:30 - Memory Interference and Quality of Service (part 1)
- 15:00-17:00 – Memory Interference and Quality of Service (part 2)

Day 4 (Thursday)

- 10:00-12:00 - Latency Tolerance and Prefetching
- 13:00-14:30 - Multiprocessor Basics and Issues
- 15:00-17:00 – Interconnects

Day 5 (Friday)

- 9:30-12:30 - Potpourri of Promising Recent Research Papers

Prerequisites

044262 | Logic Design and Introduction to Computers **or**

044252 | Digital Systems and Computer Structure **or** 234262 | Logic Design

Optional & recommended background courses -

046267 | Computer Architecture **or** 234267 | Computer Architecture

A rigorous computer architecture and digital design course similar to:
https://safari.ethz.ch/digitaltechnik/spring2018/doku.php?id=schedule

Interest in computer architecture research and practice.

Assignments

The course will consists of: lectures, discussion, paper reviews, programming assignments (tentative), exam

Bio

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 William D. and Nancy W. Strecker Early Career Professorship. His current broader research interests are in computer architecture, hardware security, computing systems, and bioinformatics. He is especially interested in interactions across domains and between applications, system software, compilers, and microarchitecture, with a major current focus on memory and storage systems. A variety of techniques he, along with his group and collaborators, have invented over the years have influenced industry and have been employed in commercial microprocessors and memory/storage systems. He obtained his PhD and MS in ECE from the University of Texas at Austin and BS degrees in Computer Engineering and Psychology from the University of Michigan, Ann Arbor. His industrial experience spans starting the Computer Architecture Group at Microsoft Research (2006-2009), and various product and research positions at Intel Corporation, Advanced Micro Devices, VMware, and Google. He received the inaugural IEEE Computer Society Young Computer Architect Award, the inaugural Intel Early Career Faculty Award, faculty

partnership awards from various companies, a healthy number of best paper or "Top Pick" paper recognitions at various computer systems and architecture venues, and the ACM Fellow recognition "for contributions to computer architecture research, especially in memory systems." His computer architecture course lectures and materials are freely available on YouTube, and his research group makes software artifacts freely available online.

For more information, please see his webpage at http://people.inf.ethz.ch/omutlu/