LOGISTICS AND INTRODUCTION

Mahdi Nazm Bojnordi

Assistant Professor
School of Computing
University of Utah
Computer System Architecture

- Computer systems are everywhere.
- What are the current and emerging challenges?
Logistics

Course organization and rules
Mahdi Nazm Bojnordi
Assistant Professor, School of Computing
PhD degree in Electrical Engineering (2016)
Worked in industry for four years (before PhD)
Research in Computer Architecture
Energy-efficient computing
Novel memory technologies
Office Hours
Please email me for appointment
MEB 3418
This Course

- Prerequisite
  - CS/ECE 6810: Computer Architecture

- Advanced topics in computer architecture
  - cache energy innovations
  - memory system optimizations
  - interconnection networks
  - cache coherence protocols
  - emerging computation models
Resources

- Recommended books and references
  - “Memory Systems: Cache, DRAM, Disk”, Jacob et al
  - “Principles and Practices of Interconnection Networks”, Dally and Towles
  - “Parallel Computer Architecture”, Culler, Singh, Gupta

- Class webpage
  - [http://cs.utah.edu/~bojnordi/teaching.html](http://cs.utah.edu/~bojnordi/teaching.html)
Course Expectation

- Use Canvas for all of your submissions
  - No scanned handwritten documents please!

- Grading
  - Up to 10% extra points for insightful questions during project presentations.

<table>
<thead>
<tr>
<th></th>
<th>Fraction</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project</td>
<td>50%</td>
<td>One simulation-based project</td>
</tr>
<tr>
<td>Homework</td>
<td>20%</td>
<td>One homework assignment</td>
</tr>
<tr>
<td>Paper presentation</td>
<td>10%</td>
<td>One in class paper presentation</td>
</tr>
<tr>
<td>Final</td>
<td>20%</td>
<td></td>
</tr>
</tbody>
</table>
Course Project

- A creative, simulation-based project on
  - Memory system optimization (SRAM, DRAM, RRAM, etc.)
  - Data movement optimizations (Off/On–chip interfaces)
  - Hardware accelerators (GPU, FPGA, ASIC)
  - ...

- Form a group of 2-3 people by Feb. 1
- Choose your topic by Feb. 8
- Prepare for an in-class presentation in April
- Prepare a conference-style report by early May
Every student presents a paper in class

- A related work on your course project is recommended
- Three main components must be included
  - The goal and key idea
  - Strengths and weaknesses
  - Future work

Email me your paper by Mar. 29
- Conferences such as ISCA, MICRO, ASPLOS, HPCA

A homework assignment will be posted on Feb. 27
- Due on Mar. 8 (11:59PM)
Academic Integrity

- Do NOT cheat!!
  - Disciplinary hearings are no fun
  - Please read the Policy Statement on Academic Misconduct, carefully.
  - We have no tolerance for cheating

- Also, read the College of Engineering Guidelines for disabilities, add, drop, appeals, etc.

- For more information, please refer to the important policies on the class webpage.
About You ...

- Are you working in a research area?
- Do you know programming languages?
  - C/C++
- Do you know any hardware description languages?
  - Verilog
- Are you familiar with simulators?
Energy-efficient Computing

The importance of energy efficient computing
Energy and Power Trends

- Power consumption is increasing significantly

(data source: ITRS, DarkSilicon’11)
New Challenges

- Excessive energy consumption
  - More energy-efficient architectures are needed

200M wearable devices will be sold in 2019 *(source: IDC forecast)*
New Challenges

- Power delivery and cooling systems
  - More energy-efficient architectures are required

Facebook datacenter at edge of the Arctic circle *(source: CNET, 2013)*

Microsoft underwater datacenter *(source: NYTimes, 2016)*
Data movement is the primary contributor to energy dissipation in nanometer ICs.
By 2020, the energy cost of moving data across the memory hierarchy will be orders of magnitude higher than the cost of performing a floating point operation.

Possible Solutions

- How to minimize data movement energy?
An Example Optimization
Cache Architecture

- Physical cache structure
Cache Banking

- Divide cache into multiple identical arrays
- Use part of the address bits to select the bank
- Remaining banks consume no dynamic power