LOGISTICS AND INTRODUCTION

Mahdi Nazm Bojnordi
Assistant Professor
School of Computing
University of Utah
Advanced Computer Architecture

Basics of Computer Systems: CPU, Memory, Storage, IO, etc.

Processor/Memory Performance Optimization: ILP, TLP, AMAT, etc.

Today/Future Concerns: Power Wall, Energy-efficiency, Security, etc.
Logistics

Course organization and rules
Instructor

- 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
  - Emerging 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
CS/ECE 7810: Advanced Computer Architecture

Course Information
- **Time:** Mon/Wed 11:50-01:00 PM
- **Location:** WEB 2470
- **Instructor:** Mahin Sayadi, email: lastname@cs.utah.edu, office hours: email me for appointment, N5 5418
- **Pre-Requisite:** CS/ECE 6810
- **Textbook:** 'Memory Systems: Cache, DRAM, Disk', Jacob et al.
- **Textbook:** 'Principles and Practices of Interconnection Networks', Dally and Towles.
- **Textbook:** 'Parallel Computer Architecture', Culier, Singh, Gupta.
- **Canvas** is the main venue for class announcements, homework assignments, and discussions.
- **Description:** This course is based on advanced topics in computer architecture, including cache energy innovations, memory system optimizations, interconnection networks, cache coherence protocols, and emerging computation models.
- **Expectation:** In addition to homework assignment and final exam, students are expected to present a conference paper related to their course project in April. A project presentation is expected for each group of students in the last two classes and final project reports are due in May. Important dates are listed below.

Important Policies and University Support

Please refer to the [College of Engineering Guidelines](#) for disabilities, add, drop, appeals, etc. Notice that we have zero tolerance for cheating; as a result, please read the [Policy Statement on Academic Misconduct](#) carefully. Also, you should be aware of the [SoC Policies and Guidelines](#).

Class rosters are provided to the instructor with the student's legal name as well as 'Preferred first name' (if previously entered by you in the Student Profile section of your CIS account). While CIS refers to this as merely a preference, I will honor you by referring to you with the name and pronoun that feels best for you in class, on papers, exams, group projects, etc. Please advise me of any name or pronoun changes (and please update CIS) so I can help create a learning environment in which you, your name, and your pronoun will be respected.

The University of Utah values the safety of all campus community members. To report suspicious activity or to request a courtesy escort, call campus police at 801-585-COPS (801-585-2677). You will receive important emergency alerts and safety messages regarding campus safety via text message. For more information regarding safety and to view available training resources, visit [SAFEU](#). Also, consider installing the SafeUT app on your phones.

Grading

The following items will be considered for evaluating the performance of students.

<table>
<thead>
<tr>
<th>Fraction</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Project</td>
<td>50% Creative, simulation based projects done by groups of 2/3 students.</td>
</tr>
<tr>
<td>Homework Assignments</td>
<td>20% A homework assignment will be posted in Canvas.</td>
</tr>
<tr>
<td>Paper Presentation</td>
<td>10% Every student presents a recent publication related to their course project.</td>
</tr>
<tr>
<td>Final Exam</td>
<td>20%</td>
</tr>
</tbody>
</table>

Important Dates

All of the submissions must be made through Canvas.

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>02/03</td>
<td>Project group composition.</td>
</tr>
<tr>
<td>02/10</td>
<td>Project proposal.</td>
</tr>
<tr>
<td>02/24</td>
<td>The homework assignment will be posted.</td>
</tr>
</tbody>
</table>
Course Expectation

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

<table>
<thead>
<tr>
<th>Task</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 people by Feb. 2
- Choose your topic by Feb. 10
- Prepare for an in-class presentation in April
- Prepare a conference-style report by end of 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. 25
- Conferences such as ISCA, MICRO, ASPLOS, HPCA

A homework assignment will be posted on Feb. 24
- Due on Mar. 4 (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.

Important Policies

Please refer to the College of Engineering Guidelines for disabilities, add, drop, appeals, etc. Notice that we have zero tolerance for cheating; as a result, please read the Policy Statement on Academic Misconduct, carefully. Also, you should be aware of the SoC Policies and Guidelines.

Class rosters are provided to the instructor with the student's legal name as well as "Preferred first name" (if previously entered by you in the Student Profile section of your CIS account). While CIS refers to this as merely a preference, I will honor you by referring to you with the name and pronoun that feels best for you in class, on papers, exams, group projects, etc. Please advise me of any name or pronoun changes (and please update CIS) so I can help create a learning environment in which you, your name, and your pronoun will be respected.
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?
The importance of energy efficient computing
Energy and Power Trends

- Power consumption is increasing significantly

(data source: ITRS, DarkSilicon’11)
CPU Power Consumption

- Major power consumption issues

**Peak Power/Power Density**
- Heat
  - Packaging, cooling, component spacing
- Switching noise
  - Decoupling capacitors

**Average Power**
- Battery life
  - Bulkier battery
- Utility costs
  - Probability, cannot run your business!
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)*
The High Cost of Data Movement

- Data movement is the primary contributor to energy dissipation in nanometer ICs.

Source: NVidia
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?
Problem: Energy Efficiency

- Unconventional solutions are needed!
  - Hardware
  - Software

Solar powered dresses
(source: www.ecochunk.com)

Harvesting motion energy
(source: www.ecouterre.com)
“People who are really serious about software should make their own hardware.”

— Alan Kay
Goal: enable energy and bandwidth efficient data movement between memory and the processor cores.

1. Energy efficient data encoding for large on-die cache
2. Bandwidth and Energy Efficient Interface
3. Efficient In-Package Memory Systems
4. Non-von Neumann Computing In Memory Modules with Emerging Technologies
Emerging Technologies

- High bandwidth memory

  Off-chip Memory
  
  Lower Bandwidth
  Lower Costs

  3D Stacked Memory
  
  Higher Bandwidth
  Higher Costs
Emerging Non-volatile Memories

- Use resistive states to represent info.
  - Can we build non-von Neumann machines?
    - In-Memory and In-situ computers