ECE 18-742 - Parallel Computer Architecture
Spring 2011 Syllabus

Course Contacts and Logistics

Instructor: Onur Mutlu, onur@cmu.edu, Office: HH (Hamerschlag Hall)-A305, 412-268-1186
http://www.ece.cmu.edu/~omutlu
Office hours: Wednesdays 2:30-3:30pm, or by appointment
Teaching Assistant: Chris Craik, ccraik@ece.cmu.edu, HH A-300 Level
Office hours: TBD
Web Page: https://www.ece.cmu.edu/~ece742
Course Hours: Mondays, Wednesdays, and Fridays, 4:30pm to 6:20pm – There will usually be two lectures per week or two lectures and one discussion per week
Course Location: SH 208

Course Overview

Parallel computers are now almost everywhere in our lives. This course is designed to build a strong understanding of the fundamentals of the architecture of parallel computers and the tradeoffs made in their design, but will also touch on how the machines are used. We will examine how architectures are designed to exploit different types of parallelism. We will study, for example, multi-core architectures, parallel memory systems, vector architectures, dataflow machines, and interconnection networks. The focus will be on f*undamentals*, *tradeoffs in parallel architecture design*, and *cutting-edge research*. The course will be very research oriented: you will conduct 1) a literature survey of very recent papers on a topic and 2) an open-ended research project. The goal of your research project will be to substantially advance the state of the art in computer architecture. You will be expected to deliver technical presentations of both your survey and project findings to the entire class.

Course Prerequisites

At least one undergraduate computer architecture course (18-447 or equivalent)
At least one graduate-level computer architecture course (18-741, 15/18-740 or equivalent)
You should have done well in 18-741 or 15/18-740.

Who Should Take 18-742?

18-742 is an advanced graduate course and is best suited for graduate and fifth-year IMB students who have a very strong research interest and graduate-level background in computer architecture. Like other graduate-level courses, **feedback on performance will be given only upon request by a student**.

Textbooks and Reading/Research Material

This is an advanced graduate-level course. I encourage you to do your own research, consult multiple sources, question assumptions and statements, and talk with me and the TAs whenever you have questions. Lectures will serve as the main source of information and they will provide the required references to textbooks or other reading material (such as research articles). A good source of information on all covered topics is the **research articles** that introduced or built upon the covered topic. These articles are usually published in top conferences (such as ISCA, MICRO, ASPLOS, HPCA) or journals (such as IEEE or ACM Transactions). I strongly encourage you to dig out the original source of the covered topics as well as the research that builds upon it. This will help you become a successful and well-read researcher in computer architecture/systems. When in doubt, ask questions.

The following books could be useful as supplements to lectures:
1. “Parallel Computer Architecture: A Hardware/Software Approach” by Culler, Singh, and Gupta.
2. “Principles and Practices of Interconnection Networks” by Dally and Towles.
3. “Readings in Computer Architecture” by Hill, Jouppi, Sohi.

The following websites contain links to original articles as well as articles that refer to them:
1. Google and Google Scholar
2. ACM Digital Library
3. IEEE Explore

Other reading material will be distributed in class and/or will be available on the website electronically.

Assignments/Grading

The course will be graded on a curve scale. The tentative breakdown of grades is given below:
40% - Research Project (detailed breakdown to follow in project handouts)
20% - Literature Survey and Its Presentation to Class
20% - Exam (Likely an oral or take-home exam)
10% - Reviews, Review Presentation, and Class Participation
5% - Quizzes, Homeworks
5% - The teaching team’s evaluation of your performance

Research Project

This course is a hands-on research oriented course. You (in groups of two or three) are expected to propose, conduct, and experimentally evaluate a 3-month long research project whose goal is to advance the state-of-the-art and/or current understanding in parallel computer architecture or a related subject. The topic of the project is flexible, but it must be approved by me. You will have to find your own project partners. *This is your chance to explore in depth a parallel computer architecture topic that interests you and perhaps even publish your innovation in a top computer architecture conference. I strongly encourage you to start thinking about your project topic as early as possible and interacting with me and the teaching assistants to crystallize it over time.*

Literature Survey and Class Presentation

A major assignment in the course is a survey of recent research on a topic and papers approved by me. In groups of two or three, you are expected to deliver a survey presentation and possibly write a short survey paper that 1) describes the problem solved and key ideas suggested in at least three recent research paper and 2) critically evaluates the papers and proposed ideas. You will prepare a presentation to class on your survey. The presentation should be similar to a lecture and should not assume the audience knows the problem beforehand.

Reviews and Homework

Much of research is understanding and critically evaluating published and unpublished work (including your own). ***This course will be heavily reading and critical analysis oriented.***

I will assign weekly readings. You will be expected to write reviews describing the problem solved, key ideas, and strengths/weaknesses of the reading from a technical perspective. You will be asked to present the paper and describe your evaluation in some class meetings. You will be evaluated on the quality of your critical evaluations and participation in discussions.

Since this is a research-oriented class, I do not intend to assign regular homeworks. However, there will be homeworks assigned to make some concepts (e.g., dynamic dataflow) concrete.

Late Policy

No late assignments will be accepted unless valid prior arrangements are made well in advance.

Academic Honesty: ECE Department and Course Policy on Cheating

Simply put, cheating is submitting work that is not your own; material handed in for grading must be the product of individual effort; anything else constitutes cheating. **Cheating** in any form or shape will result in a **failing grade** for the course. **No exceptions will be made.** Students are referred to the University Policy about Cheating and Plagiarism Policy**:**
http://www.ece.cmu.edu/student/integrity.html

Compile-Time Course Schedule (Heavily Subject to Change at Run Time)

The course will be ***very dynamically scheduled***, so take the below as only a rough outline (and a very rough schedule) of the topics I intend to cover.

Session	Date	Tentative Topic	Assignments
1	Jan 10	Introduction and Entrance Exam (Quizzes 0 and 1)	
2	Jan 12-14	Basics of Parallel Computing	
	Jan 17	**MLK Day (no class)**	
3	Jan 19-21	Multiprocessors	
4	Jan 24	Multi-core I (Organization, trends, alternatives, tradeoffs)	
5	Jan 26-28	Multi-core II (Symmetric, bottlenecks)	Project proposal due
6	Jan 31	Multi-core III (Asymmetric, opportunities)	
7	Feb 2-4	Multithreading I	
8	Feb 7	Multithreading II	
9	Feb 9-11	Resource management I	
10	Feb 14	**HPCA 2011 – read assigned papers from conference**	
11	Feb 16-18	**HPCA 2011 – read assigned papers from conference**	Project milestone I due
12	Feb 21	Resource management II	
13	Feb 23-25	Interconnection networks I	
14	Feb 28	Interconnection networks II	
15	Mar 2-4	Cache coherence	
	Mar 7	**ASPLOS 2011 and Spring Break (no class)**	
	Mar 9-11	**ASPLOS 2011 and Spring Break (no class)**	
16	Mar 14	Synchronization	Project milestone II due
17	Mar 16-18	Speculation I (Speculative threading, Transactional Memory)	
	Mar 21	**EXAM**	
18	Mar 23-25	Speculation II (Speculative threading, Transactional Memory)	
19	Mar 28	Dataflow	
20	Mar 30-Apr 1	Memory consistency	
21	Apr 4	Data parallelism, GPUs, systolic arrays I	Project milestone III due
22	Apr 6-8	Data parallelism, GPUs, systolic arrays II	
23	Apr 11	Recent Research Survey Presentations I	
24	Apr 13-15	Recent Research Survey Presentations II	
25	Apr 18	Recent Research Survey Presentations III	
26	Apr 20-22	Parallel Architecture Case Studies I	
27	Apr 25	Parallel Architecture Case Studies II	
28	Apr 27-29	Parallel Architecture Case Studies II **-- Last day of classes**	
	FINAL EXAM TIME (tentative)	Project talks and poster session	Project report due at the same time