

ECE 18-742 - Parallel Computer Architecture

Fall 2012 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: HanBin Yoon, hanbinyoon@cmu.edu, CIC 4th Floor

Office hours: Mondays and Fridays 3:15-4:15pm

Web Page: <https://www.ece.cmu.edu/~ece742>

Course Hours: Mondays, Wednesdays, and Fridays, 4:30pm to 6:20pm – The schedule will be very dynamic, determined on a weekly basis. There can be zero to three lectures or discussions any given week. When there are no lectures or discussions, students are expected to work on the literature survey and the research project.

Course Location: Doherty Hall 1112

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 fundamentals, 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. **Since this is an advanced graduate course, the students are expected to discover and critically think on their own – structured lab assignments and homework will not be available, and feedback will be provided to the student only upon request.**

Course Prerequisites

At least one undergraduate computer architecture course (18-447 or equivalent)

At least one graduate-level computer architecture course (15/18-740 or equivalent)

You should have done very well in both 18-447 and 15/18-740, or equivalents.

You should be ready to explore on your own and dedicate at least 20 hours per week to doing research in this course.

You should have the spirit, excitement and dedication for deep exploration of a topic in computer architecture.

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.**

Who Should Not Take 18-742?

If you are expecting structured lab and homework assignments and continuous feedback on your progress, this course will not meet your needs. An important goal of this course is to enable you to become an

independent researcher and a strong critical thinker, so you should be in the mindset of exploring, thinking critically, and determining your work schedule on your own with only occasional guidance.

Textbooks and Reading/Research Material

Most reading material will be in the form of research papers and articles referenced in class (see the below paragraph for more). The students are expected to dig out other material that would help them understand the concepts better. You are encouraged to talk with me as you dig up papers and sources on your own.

This is an advanced graduate-level course. You will need to do your own research, find many of the readings on your own, 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. However, they are not required.

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)
- 25% - Literature Survey and Its Presentation to Class
- 20% - Exam (Could be an oral or take-home exam)
- 15% - Reviews, Class Participation, Quizzes, Assignments
- 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; in some cases, 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.*

More details on the logistics of the research project will be provided in lectures and the project handout.

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 (in some cases, 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 papers 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.

You will be evaluated on the quality, clarity and the insightfulness of your presentation (and paper, if one is required).

More details on the logistics of the literature survey will be provided in lectures and handouts.

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 homework. However, there could be homework to make some concepts (e.g., dynamic dataflow, coherence, consistency) 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.

All the dates indicated below are tentative and are subject to change.

Please see the actual lecture schedule from Spring 2011 to get a good grasp on the topics (this site also includes links to lectures from Spring 2011):

<http://www.ece.cmu.edu/~ece742/2011spring/doku.php>

<http://www.ece.cmu.edu/~ece742/2011spring/doku.php?id=lectures>

| Session | Date | Tentative Topic | Assignments/Events |
|---------|-----------------------------|--|--|
| | Aug 27-31 | CSD Immigration course (no class) | |
| | Sep 3 | Labor Day (no class) | |
| 1 | Sep 5 | Introduction and Logistics | |
| 2 | Sep 7 | Basics of Parallel Computing | HPCA 2013 Deadline |
| 3 | Sep 10 | Multiprocessors | |
| 4 | Sep 12-14 | Multi-core I (Organization, trends, alternatives, tradeoffs) | |
| 5 | Sep 17 | Multi-core II (Symmetric, bottlenecks) | |
| 6 | Sep 19-21 | Multi-core III (Asymmetric, opportunities) | Project proposal due / Literature survey papers due |
| 7 | Sep 24 | Multithreading I | |
| 8 | Sep 26-28 | Multithreading II | |
| 9 | Oct 1-3 | Resource management I | |
| 10 | Oct 5 | Resource management II | |
| 11 | Oct 8 | Interconnection networks I | |
| 12 | Oct 10-12 | Interconnection networks II | |
| 13 | Oct 15 | Cache coherence | |
| 14 | Oct 17 | Synchronization | Project milestone I due |
| | Oct 19 | Mid-semester break (no class) | |
| 15 | Oct 22 | Speculation I (Speculative threading, Transactional Memory) | |
| 16 | Oct 24-26 | Speculation II (Speculative threading, Transactional Memory) | |
| | Oct 29 | EXAM (tentative – will very likely move to another date) | |
| 17 | Oct 31-Nov 2 | Dataflow | |
| 18 | Nov 5 | Memory consistency | |
| 19 | Nov 7-9 | Data parallelism, GPUs, systolic arrays I | |
| 20 | Nov 12 | Data parallelism, GPUs, systolic arrays II | |
| 21 | Nov 14-16 | Recent Research Survey Presentations I | Project milestone II due |
| 22 | Nov 19 | Recent Research Survey Presentations II | |
| | Nov 21-23 | Thanksgiving (no class) | ISCA 2013 Deadline |
| 23 | Nov 26 | Recent Research Survey Presentations III | |
| 24 | Nov 28-30 | Parallel Architecture Case Studies I | |
| 25 | Dec 3 | Parallel Architecture Case Studies II / Project preparation | MICRO 2012 |
| 26 | Dec 5-7 | Case Studies III / Project preparation -- Last day of classes | MICRO 2012 |
| | FINAL EXAM TIME (tentative) | Project talks and poster session | Project talk due Project poster due Project report due (final exam time) |