ECE/CS 674 – Heterogeneous Computing

Electrical and Computer Engineering Dept.
and Computer Science Dept.
Colorado State University
November 18, 2008

Prerequisite: Any one of: ECE 550, ECE 554, CS551, CS570, CS575.
Course Credits: 3

Instructor: Professor H. J. Siegel
Office: Engineering B115
Phone: 491-7982
Email: HJ@ColoState.edu

Description

In a heterogeneous computing environment, a suite of different machines is interconnected to provide a variety of computational capabilities to execute collections of application tasks that have diverse requirements. The execution times of a task will vary from one machine to the next, and tasks will compete for machines in the suite. There are many types of heterogeneous systems, including parallel, distributed, clusters, and grids. They can be found in industrial, laboratory, government, academic, and military settings. Such systems may be used in production, computing center, embedded, or real-time environments. An important research problem for heterogeneous computing is how to assign computation and communication resources to tasks and to schedule the order of their execution to maximize some performance criterion, a process known as mapping. Factors that must be considered include machine and network loading, how well the execution needs of a task match the computational capabilities of a machine, any inter-task communications, operating constraints, and the performance criterion to be optimized.

An overview of the field of heterogeneous computing will be given. Dynamic and static heuristics for mapping tasks to resources in a heterogeneous system will be presented. Tasks that involve priorities, deadlines, and alternate versions of different worths to the user will be considered. The design of resource allocations that are robust against uncertainties will be studied. Open problems in the field of heterogeneous computing will be discussed.

This course is intended for ECE and CS graduate students who want to learn about the ways in which a collection of heterogeneous machines can be used to execute a single large application task or a set of independent or interrelated application tasks in a way that will optimize some performance criterion.

Course Objectives

This course will enable the student to:

- understand the potential advantages of using heterogeneous computing systems
- analyze some of the factors that must be considered when designing resource management systems for heterogeneous environments
- be familiar with a variety of *dynamic* (on-line, real-time) and *static* (off-line) techniques for assigning resources to tasks and scheduling their execution to optimize some performance criterion
- evaluate the effectiveness of resource allocation schemes for heterogeneous systems
- formulate a performance metric appropriate for the goals of a given computing environment
- be aware of the open research problems in heterogeneous computing that are important areas for future research and development.

Course Materials

This course will be taught using technical papers from the literature.

There will be evening exams.

Course Policies (subject to change)

1. **Projects**
   There will be two research projects, both of which may involve programming and simulation studies. Students will work in pairs on these projects.

2. **Exams**
   a. There will be three exams, each covering approximately one-third of the course; there will be no comprehensive final (i.e., no final exam that covers the entire course).
   b. Exams will be given in the evening (the third exam may be given in the final exam time slot).

3. **Grading**
   a. Your final course grade will be based on your projects and examinations. Each of the two projects will be worth 20%, and each of the three exams will be worth 20%.
   b. Your letter course grade will be determined from the total points that you obtain from your projects and tests, and will be based on a combination of a relative and an absolute scale. You determine your own grade by your performance on these items.

Course Outline

	Topic	Weeks
1.	Class policy, introduction to heterogeneous computing, automatic heterogeneous computing, open problems	1
2.	Static mapping of applications composed of communicating tasks	1
3.	Static mapping techniques for independent tasks	2
4.	Assignment and description of course projects	0.5
5.	Dynamic mapping techniques for independent tasks	1
6.	Mapping of tasks with priorities, deadlines, and versions	1
7.	Scheduling communications in overloaded networks	1
8.	Project 1 presentations by students	1
9.	On-line use of off-line derived mappings	0.5
10.	Classification scheme for resource management techniques	0.5
11.	Collective quality of service measures	0.5
12.	Advanced current topics, such as robustness	3
13.	Project 2 presentations by students	1
14.	Exams	1