ECE 7110: High-Performance Computing on GPUs (Spring 2018)

This course aims to provide students with knowledge and hands-on experience in developing applications on Graphics Processing Units (GPUs) which feature massive parallel computing capability and tremendous data delivering rate. In general, we refer to a processor as massively parallel if it has the ability to complete more than 64 arithmetic operations per clock cycle. Sufficiently exploiting the potential of GPUs demands in-depth knowledge about the processor hardware feature and parallel algorithm design principles, as well as the parallelism models of GPU architecture. All of them will be covered in this course. The target audiences of this course are students who want to develop applications for these GPUs, for fun, profit, and for their research, as well as those who want to understand the feature of GPUs in order to propose techniques to further enhance their architectures.

The course will involve a number of programming projects with steadily growing complexities. All programming assignments will involve programming a massively parallel system in CUDA, which is a popular commercial language extension for C/C++ for GPU programming. Assignments involve tasks such as matrix operation, vector reduction, prefix-scan, radix sorting, graph computing and machine learning. Through the entire semester, students are expected to work on a large and complex, sometimes publishable, project in groups.

Potential Students: Graduate students, Junior and Senior Undergraduates

Book: CUDA By Examples: An Introduction to General Purpose GPU Programming by Jackson Sanders and Edward Kandrot

Prerequisites/Co-requisites: Data structure C/C++ (EECE 3220), Programming (EECE 2160), Microprocessor I (EECE 3170), and Computer Architecture (EECE 4820). Or equivalent courses from other departments.

Time: Monday 6:30pm - 9:20pm

Instructor: Prof. Hang Liu (http://hang-liu.com)

Tentative Syllabus

Time	Course Topics	Homework
1/24 (Week 1)	Introduction to GPU and CUDA	
1/31 (Week 2)	Basics in CUDA	Homework 1 Due
2/7 (Week 3)	CUDA Parallel Execution Model	
2/14 (Week 4)	CUDA Memory Model	Homework 2 Due
2/21 (Week 5)	Continued CUDA Memory Model	
2/28 (Week 6)	Matrix Multiplication (MM) on GPUs	
3/7 (Week 7)	Continue MM on GPUs	Midterm Project Report Due
3/14 (Week 8)	**Midterm presentation**	
3/21 (Week 9)	Reduction Tree on GPUs	Homework 3 Due
3/28 (Week 10)	Atomic Operation/Voting/Sync	
4/4 (Week 11)	Continue	
4/11 (Week 12)	Graph Traversal on GPUs	
4/18 (Week 13)	Continue Graph Traversal on GPUs	Final Project Report Due
4/25 (Week 14)	**Final Project Presentation**	

Sample projects:

1. Machine Learning:
    a. Deep Artificial Network
    b. K-means
    c. Recommendation System
    d. Belief Propagation
2. Linear Algebra:
    a. Conjugate Gradient Solver
    b. Fast Fourier Transform
    c. Sparse Matrix-Matrix Multiplication
3. Graph Algorithms:
    a. Weakly Connected Components
    b. PageRank
    c. Graph Coloring
    d. Single Source Shortest Path