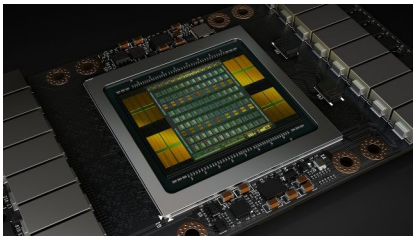
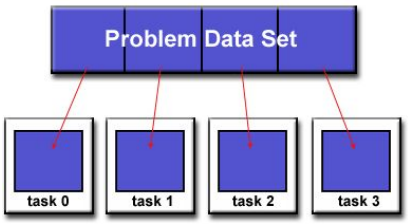
EECS 498/598 - Applied Parallel Programming with GPUs - Fall 2021





The goal of this class is to teach parallel computing and developing applications for massively parallel processors (e.g. GPUs). Self-driving cars, machine learning and augmented reality are examples of applications involving parallel computing. The class focuses on computational thinking, forms of parallelism, programming models, mapping computations to parallel hardware, efficient data structures, paradigms for efficient parallel algorithms, and application case studies.

The course will cover popular programming interface for graphics processors (CUDA for NVIDIA processors), internal architecture of graphics processors and how it impacts performance, and implementations of parallel algorithms on graphics processors. The curriculum will be delivered in ~29 lectures. The class has heavy programming components, including six hands-on assignments and a final project.

Prerequisites: EECS 281, EECS 370

Instructor: Reetuparna Das (www.eecs.umich.edu/~reetudas)

Time and Location: Lect. MW 10:30-12PM (1690 Beyster); Dis F 10:30-11:30 AM (NAME 138)

Textbook: D. Kirk and W. Hwu, Programming Massively Parallel Processors: A Hands-on Approach, Third Edition

Programming Assignments

There will be six programming assignments, which will evolve successively requiring more sophisticated programming skills. The initial assignments aim for acquisition of basic GPU/CUDA programming skills, the software interface, and understanding of the architecture of the device. The later assignments will focus on efficient parallel algorithms that utilize shared memory and synchronization, minimize path divergence, data layout and decomposition, full utilization of shared memory resources and global bandwidth through bank conflict avoidance and memory coalescing. The programming assignments can be done in groups of two. List of assignments (subject to adjustments):

Assignment 1 - Parallel Vector Addition
Assignment 2 - Parallel Matrix Multiplication
Assignment 3 - Tiled Parallel Matrix Multiplication
Assignment 4 - Parallel Reduction
Assignment 5 - Parallel Scan
Assignment 6 - Tiled Parallel Convolution

Final Project

In the final quarter of the semester, students will work on a larger, more complex project in groups of four. The final project will bring together all the learnings of the class and would ideally focus on a real-world compute intensive problems (e.g., image similarity search, markov chain convergence, optical flow, sparse system solver, cryptocurrency mining, convolutional Neural Nets or other deep learning.) The project topic needs approval from the instructor during the project proposal phase. The final project will be graded based on the quality of solution, a paper report, and a presentation/demo.

Grading

Programming Assignments:	55%
Final Project:	15%
Midterm:	15%
Final:	15%

UG Requirements

4 credits, ULCS, Flex Tech Elec.,

Honor Code

All students (including LS&A and Engineering) are required to observe the Engineering Honor Code in all assignments and exams. A copy of the honor code can be found at http://ossa.engin.umich.edu/honor-council/. Please make sure that you clearly understand what constitutes cheating. If you are not sure in any specific case, you should ask the teaching staff. The University takes honor code violations seriously, and penalties can be severe. You are not allowed to share your code with anyone other than your partners. You are not allowed to make use of project or homework solutions by others, including solutions from previous semesters. Make sure that you do not upload your code on github public repositories, as this also constitutes violation of the honor code.