EE 7722
GPU Microarchitecture

Where/When
http://www.ece.lsu.edu/gp/

Who
David M. Koppelman, Room 345 ERAD
(225) 578-5482, koppel@ece.lsu.edu
Office Hours: Monday–Friday: 14:00–15:00.

Prerequisites
EE 4720 or equivalent. By Topic: Computer architecture and digital logic. Also, students must also have a familiarity with machine language and C++ programming.

Topics
Overview—Accelerators (e.g., NVIDIA GPU, Xeon Phi) and their role in typical systems. Characteristics of workloads. Survey of GPU and manycore programming APIs and systems, such as CUDA, OpenCL, OpenMP, OpenACC. A brief history.

Parallelism and Other Basic Concepts—Threads, Speedup. Latency, Throughput. Performance Limiters.

CPU Execution Features—Role of dynamic scheduling, caches, branch prediction. Execution efficiency for certain workloads.

Execution Architecture—API thread organizations. Many-thread (NVIDIA GPU) v. many-core (Xeon Phi) organization. Hardware organization, functional units. Scheduling, instruction execution, latency hiding.


Instruction Sets—Memory access instructions, address spaces, locking. Predication, explicit reconvergence.


Text
Papers and other references.

Grading
35% Midterm Exam • 35% Final Exam • 30% Homework and Projects
Final exam weight may be increased for a student who shows significant improvement on the final exam.

Late assignment penalty: 10% per day late deducted. Missed-midterm-exam policy: at instructor’s discretion either a makeup exam, use final exam grade for midterm grade (i.e., 70% final exam weight), or use of zero for midterm grade. Daily attendance: optional, however students are responsible for all material, instructions, and notices presented in class.