EE 7722
GPU Microarchitecture

```
        S2R R11, SR_TID.X;          }
        ISETP.GT.U32.AND P0, PT, R11, 0x3, PT;
   @!P0 SHL R10, R11.reuse, 0x4;
   @!P0 IADD R8.CC, R10, R2;
        MOV32I R3, 32@hi(d_app);
{  @!P0 SHR R0, R11, 0x1c;
        LDG.E R25, [R2];            }
{  @!P0 IADD.X R9, R0, R3  SLOT 0;
        S2R R0, SR_CTAID.X  SLOT 1;         }
   @!P0 LDG.E R4, [R8+0x4];
   @!P0 LDG.E R5, [R8+0x8];
   @!P0 LDG.E R6, [R8+0xc];
   @!P0 LDG.E R7, [R8+0x10];
        XMAD R9, R0.reuse, c[0x0] [0x8], R11;
        XMAD.MRG R11, R0.reuse, c[0x0] [0x8].H1, RZ;
{       XMAD.PSL.CBCC R24, R0.H1, R11.H1, R9  SLOT 0;
        LDG.E R0, [R2+0x44]  SLOT 1;        }
   @!P0 STS.128 [R10], R4;
        BAR.SYNC 0x0;
        ISETP.GE.AND P0, PT, R24, R0, PT;
    @P0 EXIT;
        LDS.U.128 R20, [RZ];
        LDS.U.128 R16, [0x10];
        LDS.U.128 R12, [0x20];
        LDS.U.128 R8, [0x30];
```

Where/When

145 EE Building, Mon. Wed. Fri. 11:30–12:20 **Spring 2017**
`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 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 Related Concepts—Threads, Speedup. Latency, Throughput. Performance Limiters.

CPU Execution Features—Dynamic scheduling, caches, branch prediction.

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

Storage Hierarchy, Synchronization—Address spaces. Scratchpad stores and cache hierarchy. Basic data access strategies. Synchronization, atomic operations. Basic reduction strategies.

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

Basic Algorithms and Techniques—Matrix multiplication. Reduction techniques. Stencil calculations. Sorting.

Many Thread v. Many Core v. CPU, Research Directions—Workload studies. Warps v. vectors. Scratchpad v. cache. Many threads v. prefetch. Mixed CPU/GPU chips. Low-energy execution.

Text

Papers and other references.

Grading

35% Midterm Exam • 35% Final Exam • 30% Homework and Projects
Midterm exam grade may be replaced with final exam grade if it is higher.

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.

A.D.A.

Louisiana State University is committed to providing reasonable accommodations for all persons with disabilities. This syllabus is available in alternate formats upon request. Any student with a documented disability needing academic adjustments is requested to speak with Disability Services and the instructor, as early in the semester as possible. All discussions will remain confidential. Please contact Disability Services in 115 Johnston Hall, 225-578-5919 or at `http://www.lsu.edu/disability`.