Fall 2011   EE 7700-1

Performance Optimization of Applications on GPUs

Professor: **J. (Ram) Ramanujam,** 345 EE Bldg., 578-5628 Email: jxr@lsu.edu
Time: 5:10pm - 6:30pm MW, 149 EE
Reference: *Class notes and recent papers from the literature*
Prerequisites: Graduate standing

Description (EE 7700-1)

The course deals with the study of optimizations aimed at improving performance of science and engineering applications on GPUs and heterogeneous systems consisting of CPUs and GPUs. Basic ideas of parallel computing will be covered first. The course will then discuss the two GPU programming models, CUDA and OpenCL. Next, we will discuss higher-level programming approaches for GPUs such as PyCUDA, HiCUDA, HMPP and Accelerator. In addition, we will discuss optimization strategies, application performance optimization and the use of compiler tools such as PLUTO for automatic code generation. Program transformations that enhance data locality and optimize memory usage will get a lot of coverage. In addition, compilation starting from high-level specifications or for domain-specific languages will be discussed.

Course Outline

1. **Introduction**
   o Parallel computing
   o Multicore processor architectures, GPGPUs
   o CPUs versus GPUs
   o Brief Overview of Programming Models for CPUs and GPUs
   o Brief Overview of Optimizations for CPUs and GPUs

2. **GPU Architectures and programming**
   o GPU architectures
   o General-purpose computation on GPUs
   o Programming models and idioms
   o GPU programming models/environments:
      ➤ OpenCL
      ➤ CUDA
      ➤ PyCUDA
      ➤ HiCUDA
      ➤ PGI Accelerator
      ➤ CAPS/HMPP
   o Code examples on GPUs
   o Examples of CPU vs. GPU performance

3. **Optimizations and tuning for GPUs**
   o Performance characterization
   o Performance profiling
   o Application case studies
   o Optimizing memory accesses
   o Multi-level parallelism exploitation
   o Performance models and empirical search
   o Compiler-driven tuning
   o Examples of application optimization
   o Software managed memory hierarchies
   o Optimizations in HMPP, Accelerator, PLUTO

4. **Optimizations and tuning for heterogeneous CPU-GPU systems**
   o Performance characterization
   o Application case studies
   o Examples of application optimization