GPU Teaching Kit: Project Guidelines

# Purpose of the Project

The purpose of the project is to apply data-parallelism and CUDA concepts to a more substantial piece of code than in the module labs. This could take many forms, including:

* Thorough performance analysis and improvement of existing GPU code or architectures
* Implement a GPU version of some existing computationally-intensive CPU code.
* Reproduce some existing GPU research work
* Do novel GPU research

At the end of the day, the goal is to demonstrate thorough command of GPU programming and data-parallel concepts in an open-ended problem.

# Project Outline

Broadly, a successful application-parallelization project might take the following steps.

|  |  |
| --- | --- |
| Broad Outline | Concrete Example |
| Choose an application. | Dense Matrix-Matrix Multiply |
| Determine what part of the application is taking the majority of the time. |  |
| Determine one or more data-parallel approaches to solving the problem. | Assign one output to each thread in a gather-style approach. |
| Create multiple implementations of the approach. | One naïve version, one version with shared memory tiling, one version with register tiling. |
| Measure the performance and execution characteristics of the implementations for various parameters | Record memory transfer time, kernel time, utilization, FLOPS, etc. |
| Relate results to course concepts | Performance may be impacted by utilization, shared-memory accesses, memory coalescing, and control divergence. |

This approach would be modified according to the exact goals of the project. For example, many GPU research projects have specific tasks and evaluations methods, and those would be the relevant targets for implementations.

# Grading Rubric

The final project and the competition will be graded based on the following criteria:

* Demonstration of Functionality and Performance (50%):
  + Produces correct results
  + Achieves good speedup relative to base code and/or competitors
  + Optimizations applied
* Presentation (20%):
  + Preparedness and organization
  + Quality of oral delivery
  + Quality of visual aid
  + Members demonstrate equal participation and understanding of the project
  + Satisfying answers provided for questions asked
* Report (20%):
  + Covers all the required sections
  + Technically sound
  + Clear, organized, and well-written
  + Shows good in-depth analysis of the problem, the approach, and the results
* Code quality (10%):
  + Coding style is neat
  + Code is well documented

While the presentation and report each cover 20% of your grade, it is essential that they are done well to receive a good functionality score. For example, if you applied an interesting optimization that you would like to receive functionality credit for, it is essential that you explain it well in your presentation and/or report.