



Programming Assignment 4 Multiple GPUs programming

Assigned Date

2/15/2018

Due Date

2/22/2015

Overview

- Task: Calculate the matrix equation results by using multiple GPUs
- What should be done:
 - The equations should follow this style: $M1 * M2 + M3 * M4 + M5 * M6 + \dots$
 - There should be at least 8 matrices (i.e. $M1 * M2 + \dots + M7 * M8$) and each matrix should be at least have $1000 * 1000$ elements.
 - Compare the performance by using different number of GPUs and answer this question: if you use 2 GPUs instead of 1 GPU, will the throughput be doubled? Why?
 - Don't forget to change "`--gres=gpu:X`" in the command "`srun --gres=gpu:1 ./PA4`". "X" means how many GPUs will be assigned for the task in Cubix.
 - Include checks for invalid input, Chapter 3.3 cuda by example book
 - Check if allocate threads and blocks more than maximum
 - Timing -- Add statements to time the execution of the code using CUDA events or nvprof, both for the host-only (CPU) computation and with the device (GPU) computation, and display results.
 - Compute and graph the appropriate metrics (runtime, speed-up factor, throughput...).

Deliverables

- Two parts:
 - Report:

- Results: multiple timings of runs of various sizes
 - Appropriate graphs
- Code **ONLY GITHUB, don't use any library:**
 - Sequential C part
 - Cuda part
 - Same repository and output comparing results
- Have a pdf of your report emailed to Fred Harris, Lee Barford, and Rui Wu.