In his exalted name

Parallel Processing

Winter 2015

Assignment No. 4

Total Points: 100

Due Date: Friday, January 8, 2016

Problem Statement

In this assignment you are supposed to develop a CUDA program that implements two parallel versions of matrix multiplication. One of them should use one-dimensional thread addressing and other one should use two-dimensional thread addressing. Compare your code to the serial version from previous homework and compute speedups. It is recommended that you follow the following guidelines:

1. Write a simple host (sequential) version of the program in C.
2. Have a naïve GPU kernel by modify your implementation in #1 so that each thread computes one cell in the result matrix. So far you only use one thread block but the size of the matrix will be limited to the number of threads allowed in a thread block.
3. As a solution to the limitation introduced in the 2nd implementation: give each thread more work.

- For a detailed step-to-step explanation refer to this document.

Deliverables

Executable files should get 4 command-line parameters as input arguments. Dimensions of operands and number of preferable threads per block.

Ex. The command below means that the matrices dimensions are (10*20) and (20*30) and the preferable number of threads per block is 512.

```
$ ./CUDAMatrixMultiplication 10 20 30 512
```

Report files have a brief description of parallelization methods you used, the results and your analysis of them. The result section has time and speed-up tables for input sizes (size of both matrices – in bytes), (100KB, 1MB, 10MB, 100MB and 1GB) and different thread numbers.

Ex. Speedup table (numbers are fake!)

Average time of 5 runs

Input size	1MB	10MB	100MB	1GB
#Threads				
512	1.81	1.85	1.93	1.94
1024	3.70	3.68	3.67	3.71
2048	6.21	6.92	7.21	7.76

Submission

Upload your source code, executable file(s) along with your report PDF in an archive file to our course webpage, named in the following format:

[Parallel Programming] [HW4] <First Name> <Last Name> - <Student ID>

Ex. [Parallel Programming] [HW4] Ahmad Siavashi – 94131100

- *It's **extremely important** that you explain the architecture of your target GPU in details.*
- *Feel free to ask your questions. There is a forum in the course page which you can use for this purpose.*
- *This video here explains how to profile & debug your CUDA programs (Explaining Nsight Visual Profiler).*

The deadline is Friday, January 8, 2016, 11:55 PM. There is a delay penalty of -5% per day.

Good Luck ☺

The template for assignments documents was inspired by those of Mr. AmirHossein Sojoodi at Shiraz University.