

CSCI-GA.3033-012
Graphics Processing Units (GPUs): Architecture and Programming
Programming Assignment 1 – Due Th Feb 23rd, 2012

1. In the course web page you found a link for the nvidia GPUs available at school. Make a quick search and gather the following information about the GPU (i.e. the individual GPU not the whole system):

- Number of SMs
- Number of CUDA cores per SM
- Frequency
- Memory
- Number of transistors
- Release year

2. Write the matrix multiplication program that we have seen in class and run it on the school's GPUs with the following configurations (assume $W \times W$ matrices):

- Host-only version (i.e. just pure C)
- 1×1 block and $W \times W$ threads

Repeat the above for $W = 8, 16, 32, 64, 128, 256, 512,$ and 1024 .

Each time you do an experiment record the amount of time taken by the program from start to finish. Feel free to use any method you like to record the time (e.g. `clock()`). However, be consistent and use the same method for all experiments.

Draw the following graphs:

X-axis: W values

Y axis: Speedup obtained relative to the serial version (i.e. $\text{time_of_serial} / \text{time_of_GPU_version}$).

What can you say about the graph you obtained? Draw all kind of conclusions you can get.