CS/EE 217 Lab 2
Tiled Matrix Multiplication
Due Mon. Oct 26 at 8pm

1) Download the lab2-starter zip file from the class website. Unzip lab2-starter into your sdk projects directory or working linux directory.

2) Edit the source files kernel.cu and main.cu to complete the functionality of the matrix multiplication on the device. The two matrices could be any size, but we will not test your code with an output matrix size exceeding 64,000 elements.

3) There are three modes of operation similar to those in the first lab. The difference in this lab. is that you will support these modes using a Tiled implementation.

4) Answer the following questions:

a-In your kernel implementation, how many threads can be simultaneously executing? Assume a GeForce GTX 280 GPU which has 30 streaming multiprocessors.

b-Use nvcc –ptxas-options="-v" to report the resource usage of your implementation. Note that the compilation will fail but you will still get a report of the relevant information. Experiment with the NviDia visual profiler, which is part of the CUDA toolkit, and use it to further understand the resource usage. In particular, report your branch divergence behavior and whether your memory accesses are coalesced.

c-Compare the performance of the The Tiled Matrix multiplication to the simple matrix multiplication as you increase the size of the matrices and for different tile sizes. Explain any trends that you see.

Grading:

Please upload your zipped directory (after cleaning up executables and any unnecessary files) to iLearn. Your submission will be graded on the following aspects.

Correctness and performance (50%)
1. Produces correct results
2. Shared memory is used correctly (tiling) to improve performance

Report (50%)
Answers to the questions above (10 points for a, and 20 points for each of b and c).