Programming Assignment 2
Matrix Multiplication

Assigned Date
2/1/2018

Due Date
2/6/2015

Overview
- Task: Multiply two N*N matrices using CUDA
- What should be done:
    - Dynamic size for the matrices: dynamically allocated memory and add keyboard input statements to specify N
    - Should be LARGE matrices (N at least 1000)
    - Should be able to handle a matrix that is _more than_ your video card maxThreadsPerBlock (page 66, cuda by example book)
    - Different CUDA grid/block structures and sizes – Add keyboard statements to input different values for numbers of threads in a block and number of blocks in a grid
    - 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, 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.