CMPS 297S/396AA: GPU COMPUTING ASSIGNMENT 2

In this assignment, you will implement a simple matrix-matrix multiplication kernel. The kernel takes a matrix $\mathbb A$ of size $\mathbb M \times \mathbb K$ and a matrix $\mathbb B$ of size $\mathbb K \times \mathbb N$, and produces a matrix $\mathbb C$. Do not perform any memory optimizations. You will do this in the next assignment.

Instructions

- 1. Place the files provided with this assignment in a single directory. The files are:
 - main.cu: contains setup and sequential code
 - kernel.cu: where you will implement your code (you should only modify this file)
 - common.h: for shared declarations across main.cu and kernel.cu
 - timer.h: to assist with timing
 - Makefile: used for compilation
- 2. Edit kernel.cu where TODO is indicated to implement the following:
 - Allocate device memory
 - Copy data from the host to the device
 - Configure and invoke the CUDA kernel
 - Copy the results from the device to the host
 - Free device memory
 - Perform the computation in the kernel
- 3. Compile your code by running: make
- 4. Test your code by running: . / mm
 - If you are using the HPC cluster, do not forget to use the submission system. Do not run on the head node!
 - For testing on different matrix sizes, you can provide your own values for matrix dimensions as follows: ./mm <M> <N> <K> (example: ./mm 256 512 128)

Submission

Submit your modified kernel.cu file via Moodle by the due date. Do not submit any other files or compressed folders.