Tingqi (Ting) Wang

Problem 1: Matrix Multiplication

The following is the output from running both versions of matrix multiplication on the HPC. Where p1 is the unoptimized implementation using global memory only and p2 is the block matrix multiplication using shared memory.

```
========================================
SLURM_JOB_ID = 17594137
SLURM_JOB_NODELIST = d13-05
TMPDIR = /tmp/SLURM_17594137
========================================
p1 time is 14351860.000000 ns
c[451][451]=2048.000000
p2 time is 7419125.000000 ns
c[451][451]=2048.000000
```

From the results above, we can see that the runtime of the two implementations of matrix multiplication is:

Unoptimized version (p1): 14351860 ns
Optimized version (p2): 7419125 ns

Through using shared memory and block matrix multiplication, the runtime of p2 is almost 2 times faster than p1. By utilizing the shared memory in the kernel, there are less fetches to global memory that have higher access time than shared memory, thus reducing the memory access time and the overall kernel runtime.