CS121 Problem Set 5

Instructions: This problem set is due before 7:30pm on April 26, 2018. Submit a hardcopy of your solutions to the TA, or email your solutions to shtcshw@163.com.

1) Indicate whether the following statements are true or false. Justify your answers.

   a. CUDA uses an SIMD programming model in which all threads executing concurrently in a grid perform the same instruction.

   b. Each CUDA thread has its own registers; therefore, the number of registers available to each thread is constant, independent of the number of threads.

   c. In CUDA, thread blocks are the largest threading units within which threads can share data and synchronize with each other.

2) You need to write a kernel that operates on an image of size 400 x 900 pixels. You would like to assign one thread to each pixel, and you want your thread blocks to be square. Suppose your device supports up to 1024 threads per block, up to 2048 threads per SM, and up to 16 thread blocks per SM.

   a. What size thread blocks should you use to maximize the number of threads on each SM?

   b. For the block size from part a, how many idle threads will you have?

3) Consider the vector addition example from the first CUDA lecture. Suppose that instead of processing one element, each thread processes 2 elements. Give the mapping from thread/block indices to vector index. Do the same for each thread processing 4 or 8 vector elements.

4) Design a CUDA program to multiply an $n \times n$ matrix A by an $n \times 1$ vector B to produce another $n \times 1$ vector C, where $C[i] = \sum_{j=1}^{n} A[i][j] \cdot B[j]$. Use one thread to produce each value in C.