Assignment #2: The Big Dot II

The dot product of two vectors $a = (a_0, a_1, ..., a_{n-1})$ and $b = (b_0, b_1, ..., b_{n-1})$, written $a \cdot b$, is simply the sum of the component-by-component products:

$$a \cdot b = \sum_{i=0}^{n-1} a_i \times b_i$$

Dot products are used extensively in computing and have a wide range of applications. For instance, in 3D graphics (n = 3), we often make use of the fact that $a \cdot b = |a||b|cos\theta$, where $|\ |$ denotes vector length and θ is the angle between the two vectors.

In this assignment, you are expected to:

1. Write a CUDA code to compute in parallel the dot product of two random single precision floating-point vectors with size N = 1<<24;
2. Write two kernel functions for the dot product computation on GPU:
   - I) kernel1: use **shared memory** and **parallel reduction** to calculate partial sum on each thread block. (Add up all the partial sums on CPU after transferring all the partial sums back to host from device)
   - 2) kernel2: use **shared memory**, **parallel reduction**, and **atomic function or atomic lock** to perform the entire computation on GPU. (Transfer the final dot product result back to host from device)
3. Compare the time it takes for kernel1 and kernel 2. (Use **cudaEventRecord()** for the timing.)
4. Turn in your source code on Canvas with a readme.txt to explain whatever I need to know to run your code successfully.