Parallel Computer Architecture
CUDA Programming Assignment
Vector Dot Product

Instructor: Prof. Naga Kandasamy

ECE Department

Drexel University

March 4, 2018

The assignment is due March 14, 2018 by 11:59 pm via BBLearn. You may work on the problem in a team of up to two people. The submitted code must be your own. Please do not copy code from other students/teams or from online sources. Violation of this policy will result in a score of zero for the entire assignment.

(20 points) Given two n-element vectors **a** and **b**, their dot product $\mathbf{a} \cdot \mathbf{b}$ is given by

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i,$$

where a_i and b_i denote the ith elements of vectors **a** and **b**, respectively.

The program provided to you accepts n as an argument. It creates two randomly initialized vectors and computes their dot product using both the CPU and the GPU. The solution provided by the GPU is compared to that generated by the CPU. Answer the following questions:

- Edit the `compute_on_device()` function in `vector_dot_product.cu` to complete the functionality of vector dot product on the GPU. You may add multiple kernels to the `vector_dot_product_kernel.cu` file to achieve this functionality. Use the GPU memory hierarchy judiciously to achieve the best speedup that you can.

- Upload all files needed to run your code as a single zip file via BBLearn. Also, submit a brief report describing: (1) the design of your kernel using code or pseudocode to clarify the discussion; (2) the speedup achieved over the serial version for vectors of 10^5, 10^6, and 10^7 elements; and (3) sensitivity of the kernel to thread-block size in terms of the execution time.