

# Introduction to Parallel Computer Architecture

## Vector Dot Product

Instructor: Prof. Naga Kandasamy  
ECE Department  
Drexel University

November 25, 2015

The assignment is due December 7, 2015, 11:59 pm. You can work on this assignment in a team of up to two people.

1. **(20 points)** Given two  $n$ -element vectors  $\mathbf{a}$  and  $\mathbf{b}$ , 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  $i^{\text{th}}$  elements of vectors  $\mathbf{a}$  and  $\mathbf{b}$ , 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:

- **(15 points)** 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. The CUDA source files for this question are available in a zip file on BBLearn.
- **(5 points)** 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.

Submit all the files needed to run your code on BBLearn as a single zip file.