

## Home Assignment 4 (15 points)

---

### Description

Using CUDA C to complete the following coding requirement to deepen your understanding. Some of your results/code will be used for your final project. Due date of the assignment: **Nov. 21th, 11:00pm, Eastern Time.**

### Coding the Following Question (9 points)

- Write CUDA C based codes to conduct the following multiple matrix multiplication operations. Given  $N$  matrixes  $A_1, \dots, A_N$ , and  $N$  matrix multiplication  $B_1, \dots, B_N$ , the  $N$  matrixes  $C_1, \dots, C_N$  could be obtained by the following equations.

$$\begin{aligned} C_1 &= A_1 * B_1 \\ C_2 &= A_2 * B_2 \\ &\vdots \\ C_N &= A_N * B_N \end{aligned} \tag{1}$$

where  $A_1, \dots, A_N$  and  $B_1, \dots, B_N$  have the same shape, and the matrixes  $A$  and  $B$  are randomly generated.

- Students MUST use CUDA C to conduct the operation. Other coding environments/libraries such as PyCuda, MATLAB, CuPy are NOT allow to use. However, students could use cuBLAS library to complete the assignment as cuBLAS is one of the most important libraries in CUDA.
- The CUDA run-time should be report in the assignment.
- Please run your codes under the following ONE of the three conditions. i)  $A \in R^{500 \times 500}$ ,  $B \in R^{500 \times 400}$ ,  $N > 100$ ; ii)  $A \in R^{50 \times 20}$ ,  $B \in R^{20 \times 50}$ ,  $N > 5000$ ; iii)  $A \in R^{6 \times 4000}$ ,  $B \in R^{4000 \times 9}$ ,  $N > 1000$ ;

### Performance Analysis (6 points)

- (1 point) Analysis the Performance of your Assignment 3 Part II. This time, write a Python or Matlab code of the Assignment 3 Part II, and then compare the running time between the your CUDA code and your Python/MATLAB based code.

- (5 points) Analysis the Performance of your code in this assignment. This time, write a Python or Matlab code, and then compare the running time between the your CUDA code and your Python/MATLAB based code. Please try your best to improve your CUDA code to get more speed improvement. If the performance improvement is not quite significant, some points will be cut.

## Requirement

- All the provided codes will be run by Instructor based on the following two environment. i) Google Colab. If students use Google Colab, please indicate it in the assignment so I will copy your code and run it in the Google Colab. ii) otherwise, all the codes will be run in the Visual Studio 2019/2017 environment.
- If the codes could not successfully run in the above environments, at least 80 percent of points will be cut off. Please make sure your code could be run in the environment, no additional files will be allow to upload or submit after the due date.
- If students use libraries such as PyCuda, CuPy to conduct the assignment, all the points will be cut off.
- No late submission policy.
- If similar codes have been found, the assignment will be marked as Zero and will report to the University.
- The matrixes  $A$  and  $B$  should be randomly generated to make sure the codes are reproducibly.