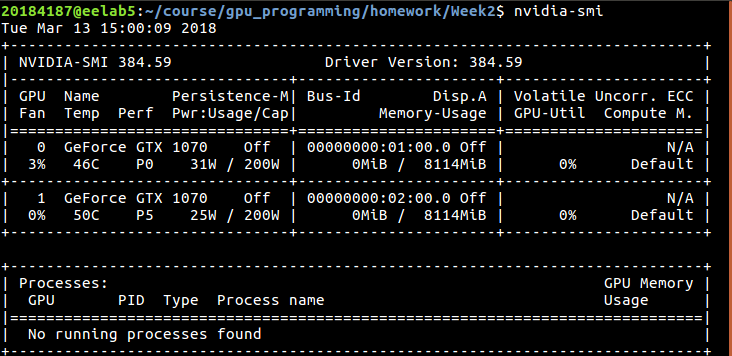
**EE817 GPU Programming and Its Applications Spring 2018**

**Name:** Dinh Vu **Student ID number:** 20184187

**Homework 1**

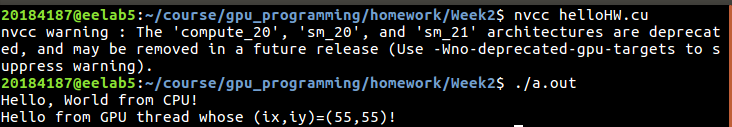
The computer used in this homework contains NVIDIA GeForce 1070 based on Pascal GP104 architecture.



**Figure 1.** Graphic card information

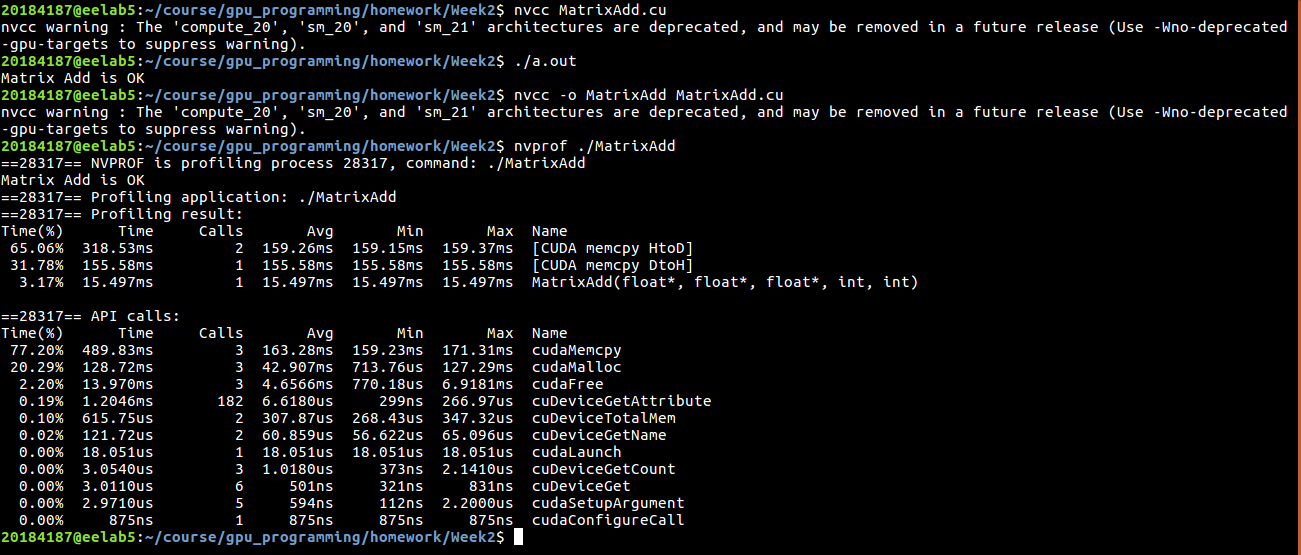
1. **Problem 1**

My result in problem 1 is shown in Figure 2 below.



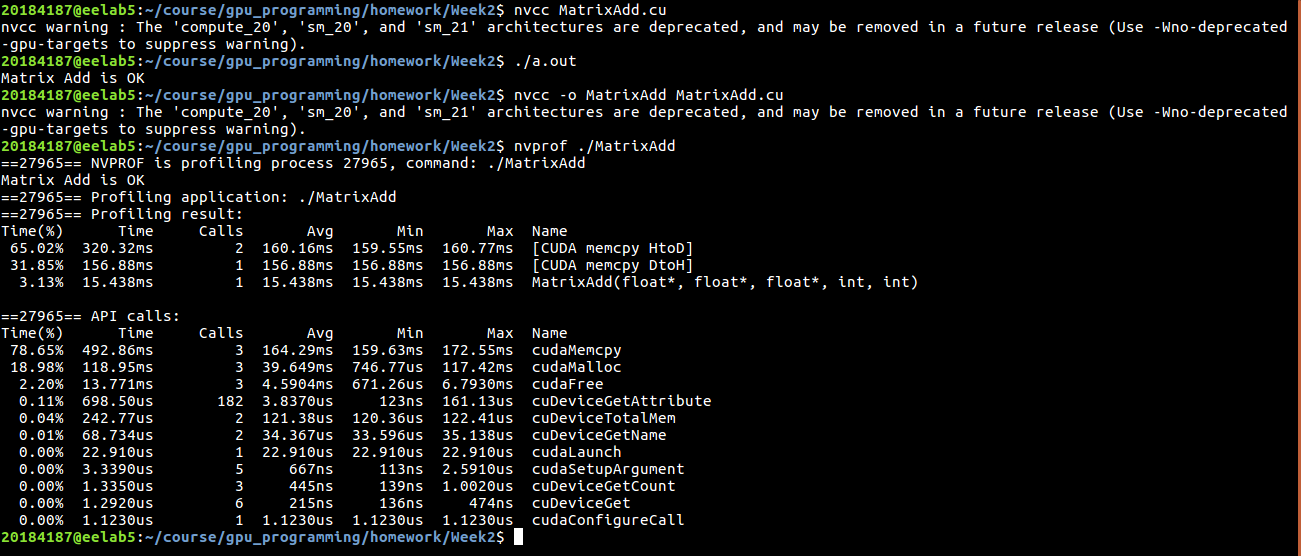
**Figure 2.** The result of problem 1

1. **Problem 2**
2. **Block 32×32**



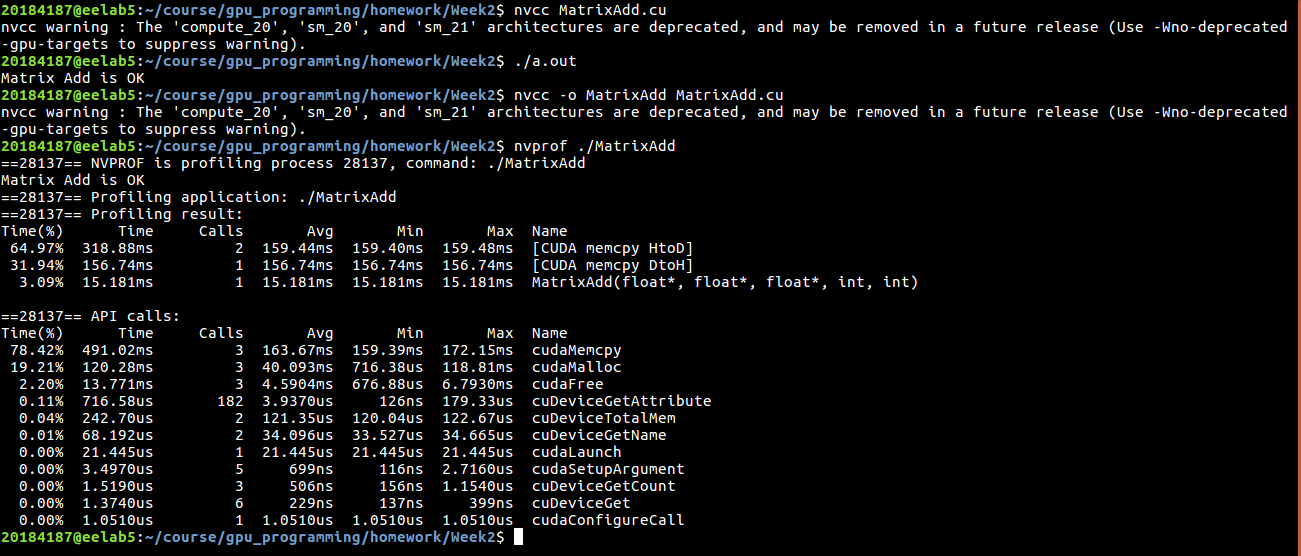
**Figure 3.** The result of problem 2 with block 32×32

1. **Block 8×8**



**Figure 4.** The result of problem 2 with block 8×8

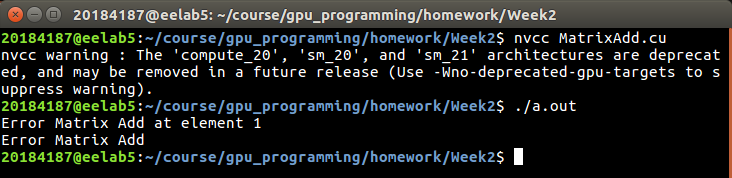
1. **Block 16×16**



**Figure 5.** The result of problem 2 with block 16×16

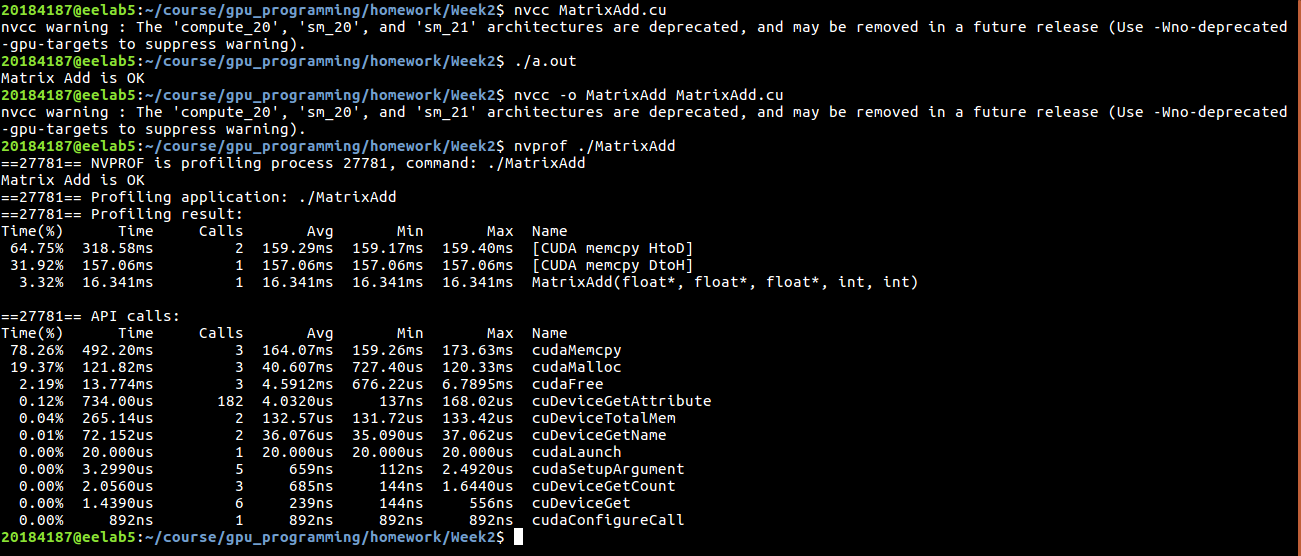
1. **Block 64×64**

In this case, the total number of threads per block is 64×64 = 4096 threads, greater than the maximum thread block size in Pascal architecture. Hence, it cannot be executed and the result is display in Figure 2.4.



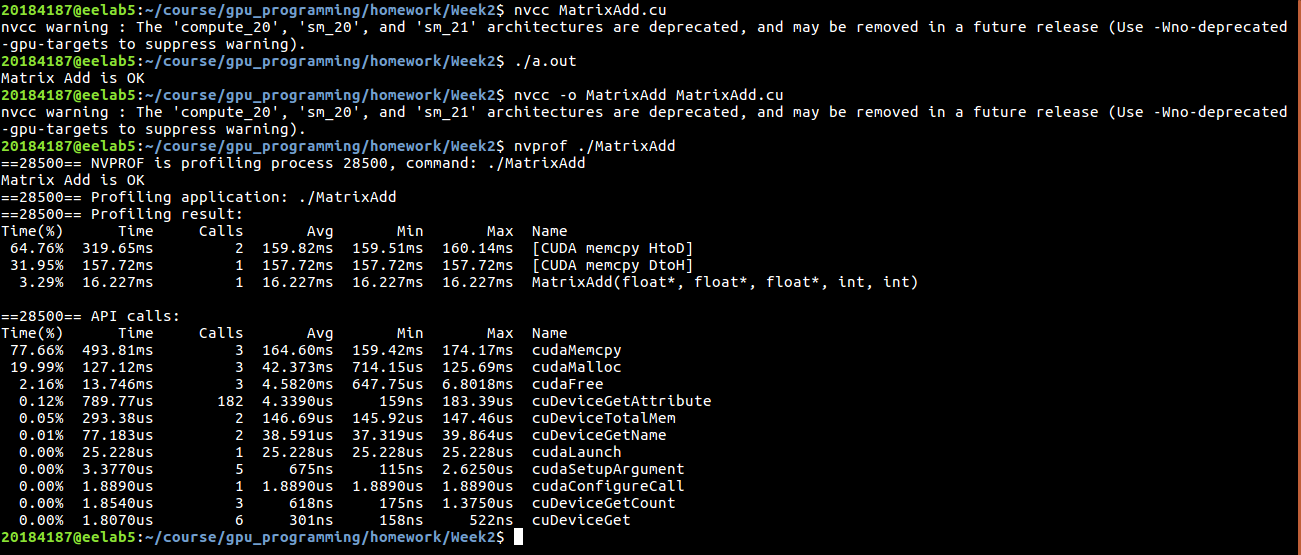
**Figure 6.** The result of problem 2 with block 64×64

1. **Block 8×16**



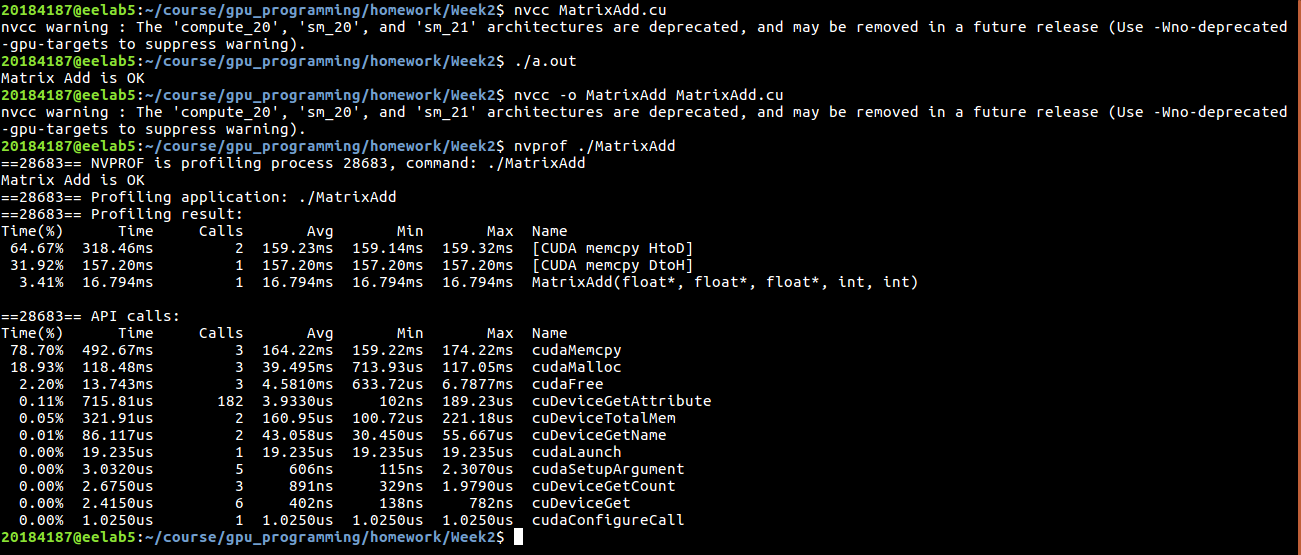
**Figure 7.** The result of problem 2 with block 8×16

1. **Block 8×32**



**Figure 8.** The result of problem 2 with block 8×32

1. **Block 8×64**



**Figure 9.** The result of problem 2 with block 8×64

* 1. **Comparison between results and explanation**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Block** | **Grid** | **Execution time of MatrixAdd function (ms)** |
| 1 | 8×8 | 2000×2000 | 15.438 |
| 2 | 16×16 | 1000×1000 | 15.181 |
| 3 | 32×32 | 500×500 | 15.479 |
| 5 | 8×16 | 2000×1000 | 16.341 |
| 6 | 8×32 | 2000×500 | 16.227 |
| 7 | 8×64 | 2000×250 | 16.794 |

**Table 1.** Comparison of execution time between the different block dimensions

The execution time of MatrixAdd function for block 16×16 is smallest and for block 8×64 is greatest.

The execution time of MatrixAdd function depends on the execution time in each SM including read/write data from registers, caches and memory. The 2-D thread matrix of a block is stored in memory of the SM as 1-D array following row order. So, it means that if block.x is greater, address pointer move faster then the processing time in SM reduce.

However, the execution time of MatrixAdd function for block 32×32 is greater than for block 16×16 because Pascal GPU GP104 have 256-bit memory bus, so transferring a block 32×32 need more time than block 16×16.