**ECEN 5593 – Advanced Computer Architecture**

**CUDA Vector Addition**

**Name**: Adarsh Anand Kulkarni

1. Execution time and memory transfer time for different data sizes:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Input Size | GPU Execution Time | Memory Transfer Time | Overall Execution Time (Memory + GPU Execution) | Percentage Execution Time over Memory Transfer Time |
| 1000 | 0.045000 (ms) | 0.033000+0.020000 = 0.053 ms | 0.113000 (ms) | 46.90265 % |
| 10000 | 0.039000 (ms) | 0.069000 + 0.059000 = 0.128 ms | 0.184000 (ms) | 69.56522 % |
| 100000 | 0.083000 (ms) | 0.429000 + 0.425000 = 0.854 ms | 0.953000 (ms) | 89.61175 % |
| 1000000 | 0.237000 (ms) | 2.903000 + 3.421000 = 6.324 ms | 6.582000 (ms) | 96.08022 % |
| 1500000 | 0.253000 (ms) | 4.195000 + 4.999000 = 9.194 ms | 9.470000 (ms) | 97.08553 % |
| 2000000 | 0.270000 (ms) | 5.450000 + 6.554000 = 12.004 ms | 12.296000 (ms) | 97.62524 % |

1. What is the approximate execution Input Size that is better to run on a CPU versus GPU? (Restated- what is the execution time where GPUs start to have better performance than CPUs to execute vector addition). To solve this and get full credit, you must implement sequential loop-based vector addition, and time the execution and present the results in the following table:

|  |  |  |
| --- | --- | --- |
| Input Size | GPU Overall Execution Time | CPU Overall Execution Time |
| 1000 | 0.111 (ms) | 0.001 (ms) |
| 10000 | 0.171 (ms) | 0.036 (ms) |
| 100000 | 0.907 (ms) | 0.34 (ms) |
| 1000000 | 5.091 (ms) | 3.6750 (ms) |
| 1500000 | 7.44 (ms) | 5.762 (ms) |
| 2000000 | 9.429 (ms) | 7.666 (ms) |