DD2360 Assignment 1

Group 35

Shiyuan Shan

Exercise 1

1.1 Main differences between CPU and GPU:

|  |  |
| --- | --- |
| CPU | GPU |
| A few cores for general-purpose tasks | Many smaller specialized cores for repetitive computationally intensive tasks |
| Allows task parallelism | Specialized for data parallelism |
| Handles complex control flows like branches and solves corresponding hazards | Focuses mainly on uniform control flow to lower latency in control divergence |
| Suits tasks for general purposes | Suits specific tasks that are repetitive and computationally intensive like rendering |

1.2 Rank:

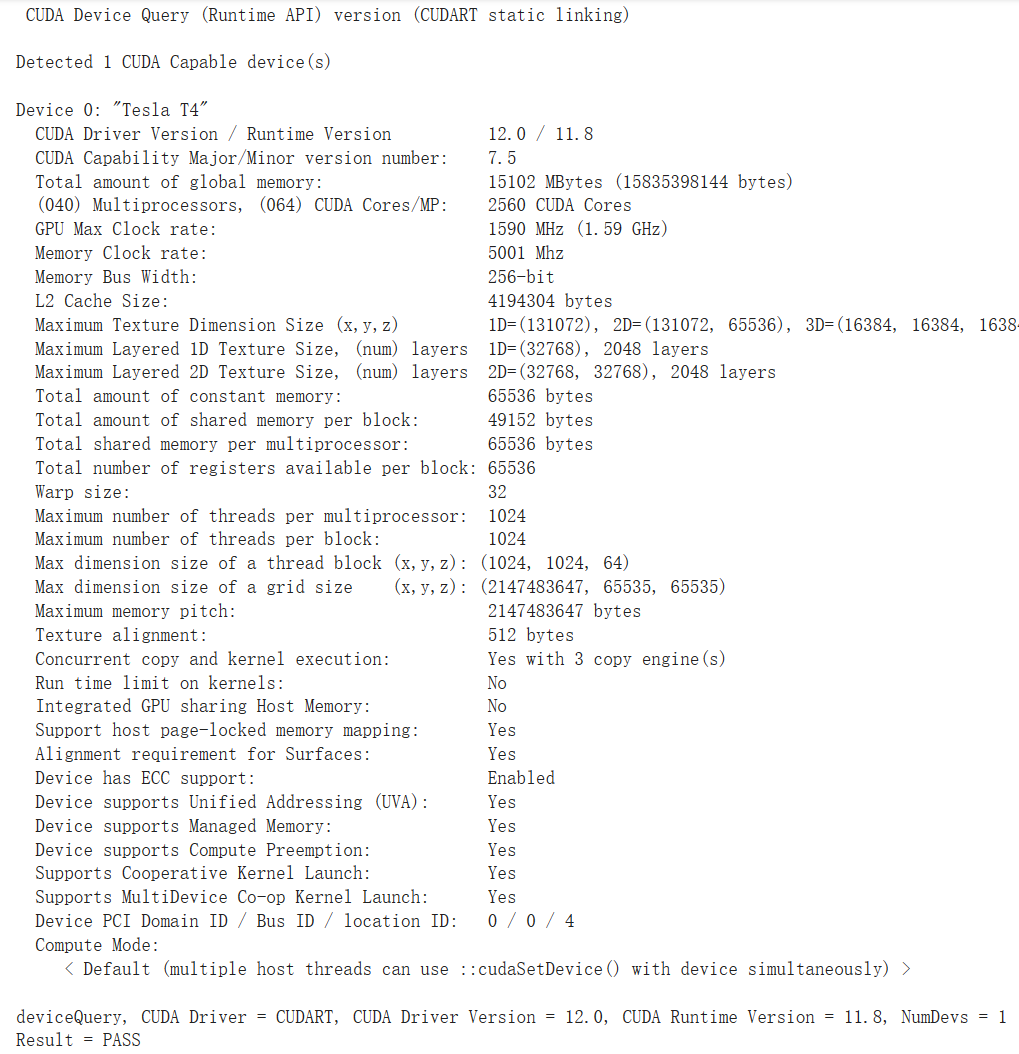
|  |  |  |
| --- | --- | --- |
| Rank | Name | GPU model |
| 1 | Frontier | AMD Instinct MI250X |
| 2 | Fugaku |  |
| 3 | LUMI | AMD Instinct MI250X |
| 4 | Leonardo | NVIDIA A100 |
| 5 | Summit | NVIDIA Volta GV100 |
| 6 | Sierra | NVIDIA Volta GV100 |
| 7 | Sunway TaihuLight |  |
| 8 | Perlmutter | NVIDIA A100 |
| 9 | Selene | NVIDIA A100 |
| 10 | Tianhe-2A |  |

1.3 Power effieciency:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rank | Name | Rmax(PFlops/s) | Power(kW) | Power efficiency (PFlop/s/kW) |
| 1 | Frontier | 1194.00 | 22703 | 0.052592 |
| 2 | Fugaku | 442.01 | 29899 | 0.014783 |
| 3 | LUMI | 309.10 | 6016 | 0.05138 |
| 4 | Leonardo | 238.70 | 7404 | 0.032239 |
| 5 | Summit | 148.60 | 10096 | 0.014719 |
| 6 | Sierra | 94.64 | 7438 | 0.012724 |
| 7 | Sunway TaihuLight | 93.01 | 15371 | 0.006051 |
| 8 | Perlmutter | 70.87 | 2589 | 0.027374 |
| 9 | Selene | 63.46 | 2649 | 0.023956 |
| 10 | Tianhe-2A | 61.44 | 18482 | 0.003324 |

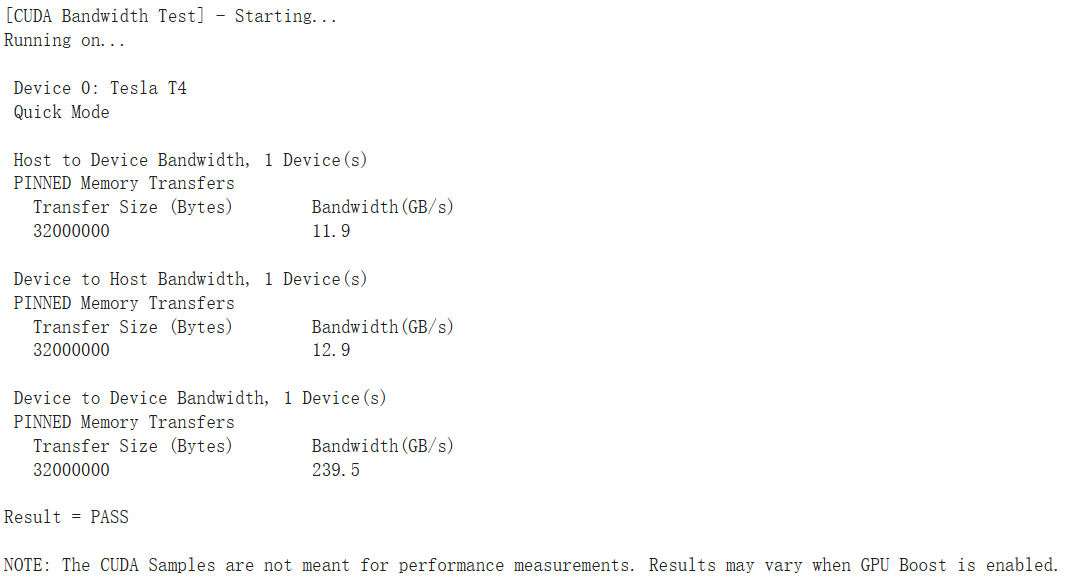
Exercise 2

2.1



2.2 The compute capability is Version 7.5

2.3



2.4 Memory bandwidth = clock rate \* bus width = 5001MHz \* 256bit = 160GB/s

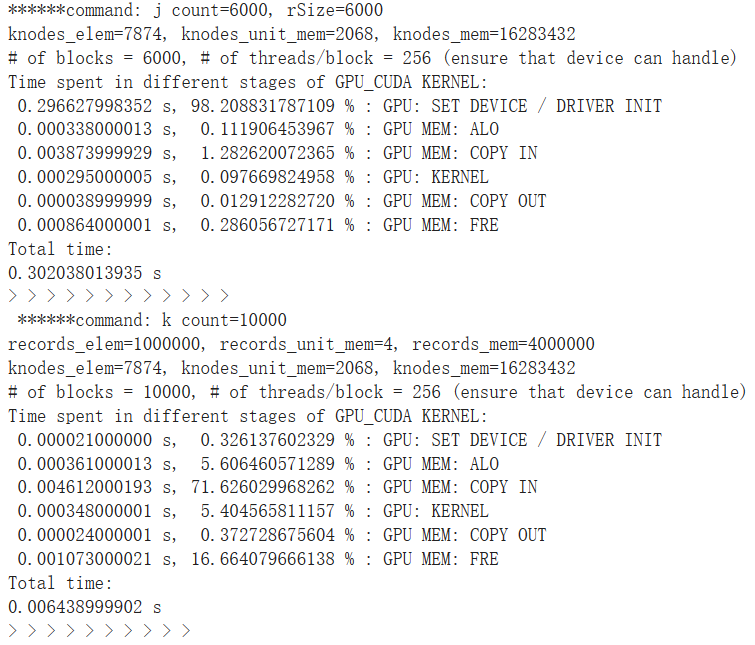
With DDR, bandwidth gets doubled = 320GB/s

Not quite consistent with results 239.5GB/s

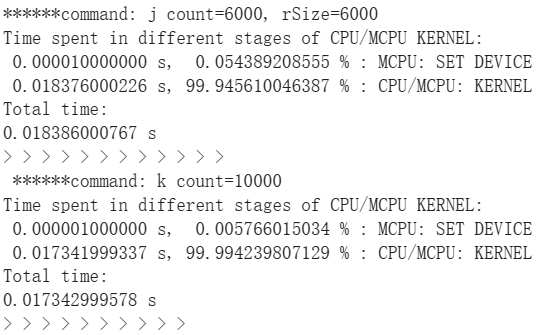
Exercise 3

3.1 All CUDA\_FLAG related code is deleted in the makefile.

3.2 b+tree:

Cuda execution total time: 

OpenMP execution total time:



lavaMD:

Cuda execution total time: 0.451678991318 s

OpenMP execution total time: 2.641376018524 s

3.3 b+tree: for j count = 6000 and rSize = 6000, CPU out-performs GPU, but for k count = 10000, GPU out-performs CPU.

lavaMD: GPU out-performs CPU much.

Discussion: I do observe a great performance advantage of GPU in most bench marks, while in the rest few cases it seems to be the opposite case. This probably is due to the difference between algorithms and data sizes. GPU out-performs CPU when it comes to parallelizable computations, but it requires extra data transfer and parallelizable computation demands. Thus if the data size is not big enough or the algorithm needs sequential calculations, it is sensible that GPU performs worse than CPU does.

Exercise 4

Don’t have PDC account yet