Introduction to CUDA Parallel Programing
Homework 1: Result and Discussion

Kuan-Chia Chiu

March 14, 2018

1 Result

The Output file is attached in the folder "hw1".

```
Set GPU with device ID = 0
Matrix Addition: C = A + B
Enter the size of the matrices: 6400
Enter the number of threads per block: 4
The number of blocks is 1600
Input time for GPU: 60.439102 (ms)
Processing time for GPU: 6.088672 (ms)
GPU Gflops: 20.181740
Output time for GPU: 92.432930 (ms)
Total time for GPU: 158.960693 (ms)
Processing time for CPU: 1747.233032 (ms)
CPU Gflops: 0.070328
Speed up of GPU = 10.991604
Check result:
norm(h_C - h_D)=0.000000000000000e+00

Set GPU with device ID = 0
Matrix Addition: C = A + B
Enter the size of the matrices: 6400
Enter the number of threads per block: 8
The number of blocks is 800
Input time for GPU: 60.804192 (ms)
Processing time for GPU: 3.650624 (ms)
GPU Gflops: 33.659999
Output time for GPU: 90.415779 (ms)
Total time for GPU: 154.870605 (ms)
Processing time for CPU: 1811.071899 (ms)
CPU Gflops: 0.067849
Speed up of GPU = 11.694098
```

```
Check result:
norm(h_C - h_D)=0.000000000000000e+00

Set GPU with device ID = 0
Matrix Addition: C = A + B
Enter the size of the matrices: 6400
Enter the number of threads per block: 10
The number of blocks is 640
Input time for GPU: 60.498302 (ms)
Processing time for GPU: 3.742944 (ms)
GPU Gflops: 32.829772
Output time for GPU: 92.419197 (ms)
Total time for GPU: 156.660446 (ms)
Processing time for CPU: 1748.491699 (ms)
CPU Gflops: 0.070278
Speed up of GPU = 11.161029
Check result:
norm(h_C - h_D)=0.000000000000000e+00

Set GPU with device ID = 0
Matrix Addition: C = A + B
Enter the size of the matrices: 6400
Enter the number of threads per block: 16
The number of blocks is 400
Input time for GPU: 60.979649 (ms)
Processing time for GPU: 3.368288 (ms)
GPU Gflops: 36.481441
Output time for GPU: 90.505501 (ms)
Total time for GPU: 154.853439 (ms)
Processing time for CPU: 1803.688354 (ms)
CPU Gflops: 0.068127
Speed up of GPU = 11.647713
Check result:
norm(h_C - h_D)=0.000000000000000e+00

Set GPU with device ID = 0
Matrix Addition: C = A + B
Enter the size of the matrices: 6400
Enter the number of threads per block: 20
The number of blocks is 320
Input time for GPU: 60.824287 (ms)
Processing time for GPU: 3.428640 (ms)
GPU Gflops: 35.839284
Output time for GPU: 90.901154 (ms)
Total time for GPU: 155.154083 (ms)
Processing time for CPU: 1810.045776 (ms)
```

```
CPU Gflops: 0.067888
Speed up of GPU = 11.666118
Check result:
norm(h_C - h_D)=0.000000000000000e+00

Set GPU with device ID = 0
Matrix Addition: C = A + B
Enter the size of the matrices: 6400
Enter the number of threads per block: 32
The number of blocks is 200
Input time for GPU: 60.520607 (ms)
Processing time for GPU: 3.505504 (ms)
GPU Gflops: 35.053448
Output time for GPU: 92.343933 (ms)
Total time for GPU: 156.370041 (ms)
Processing time for CPU: 1752.097290 (ms)
CPU Gflops: 0.070133
Speed up of GPU = 11.204814
Check result:
norm(h_C - h_D)=0.000000000000000e+00
```

2 Discussion

From the result, I made some summary here:
As Fig. 1 shown, the computation time for GPU is faster than CPU even using different block size.
We can find the GPU Gflops varying when the block size changing in Fig. 2. Here the optimized value is near the ”16” for block size.
From Fig. 3, it is found the Gflops for CPU does not vary so much, and the value is much smaller than the value for GPU.
In Fig. 4, I checked the input and output time. The output time is around 150% larger than input time. It suggests if we want to increasing the speed getting the result from graphics cards, we could start to reduce the memory usage of output data first. Then, we could reduce input data as well. The computing time is much shorter than data transferring.

Figure 1: The total time for GPU and CPU

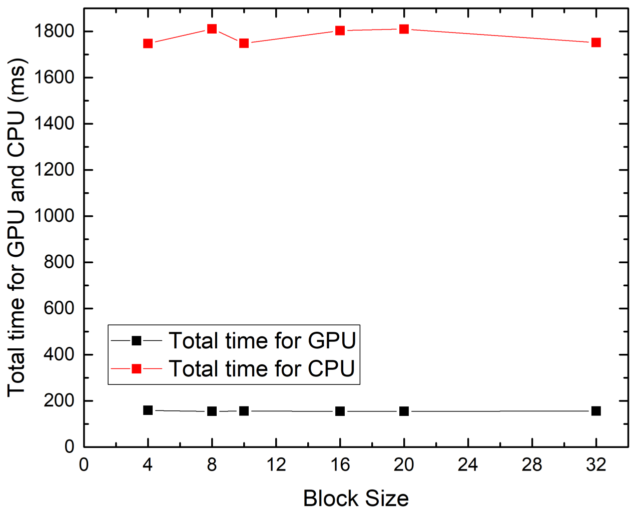


Figure 2: Gflops of GPU in different block size

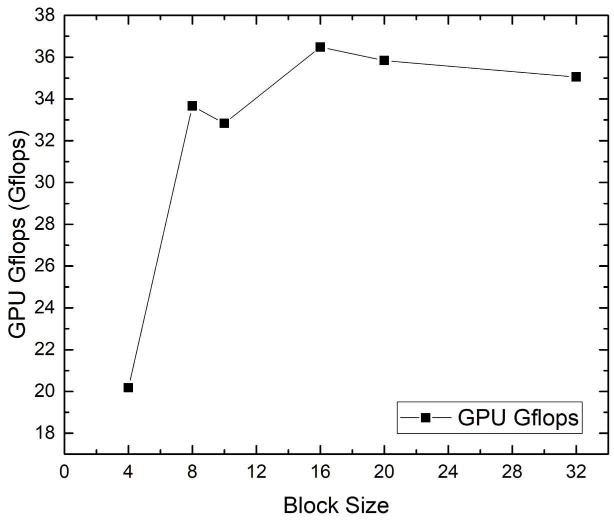


Figure 3: Gflops of CPU in different block size

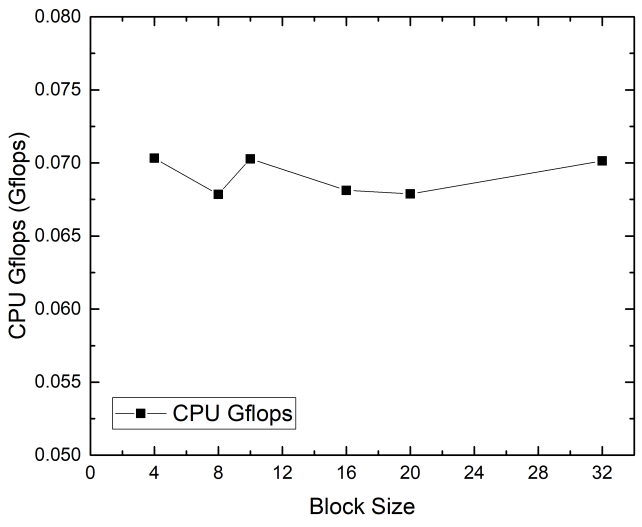
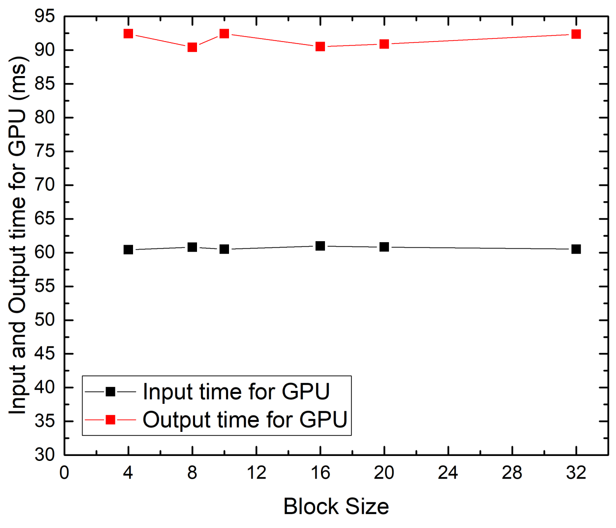


Figure 4: Comparison of input and output time


Input and Output time for GPU (ms)
Block Size
Input time for GPU
Output time for GPU