Introduction to CUDA Parallel Programming
Homework Assignment 1

B07901069 電機四 劉奇聖

March 9, 2022

1 README

This file is `report.pdf`. `src/` folder contains the source code. `result/` folder contains the execution results.

In `src/` folder, executing `make` to compile the program.

The program usage is `./MatAdd <gpu_id> <block_size>`. For example, `./MatAdd 0 16` will use gpu id 0 and set the block size as (16, 16). The `MatAdd` program will do matrix addition and outputs the processing time for GPU and GPU Gflops.

In `result/` folder, there are `block_4/`, `block_8/`, `block_10/`, `block_16/`, `block_20/`, `block_32/`, folders, corresponding to the execution results for block size (4, 4), (8, 8), (10, 10), (16, 16), (20, 20), (32, 32), respectively. In each folder, there are ten text files, each one is one execution result with that block size.

2 Result

I run the program on `twqcd80` and use gpu 0. As mentioned in README section, for each block size, I executed 10 times to get a more precise result. Below are the average GPU processing times and average Gflops.

Block Size	Processing Time (ms)	Gflops
(4, 4)	27.4467778	4.5488349
(8, 8)	14.4338048	8.6571818
(10, 10)	17.274083	7.3854252
(16, 16)	20.7544382	6.9630054
(20, 20)	20.3018115	7.170773
(32, 32)	21.0736962	7.0816665

3 Discussion

From the result, we can see that the optimal block size is (8, 8). We can conclude that neither smaller nor bigger block size will result in better performance. To obtain the optimal block size, we need to try multiple block sizes.