Programming Assignment 7

ECE 759, Prof. TW Huang

Sai Tadinada

GitHub link to programming tasks:
https://github.com/phantom3012/repo759/tree/main/HW07

1 Question 1

1.a

`task1.cu` can be found at https://github.com/phantom3012/repo759/blob/main/HW07/task1.cu

1.b

Scaling analysis plot for task1 is shown as in Figure 1:

1.c

The best possible value of `block_dim` for $n = 2^{14}$ is 32.

1.d

The performance of the code is better when using `float` data type compared to `double` data type. This is because `float` data type requires less memory and hence, the number of operations required to perform the matrix multiplication is less compared to `double` data type.

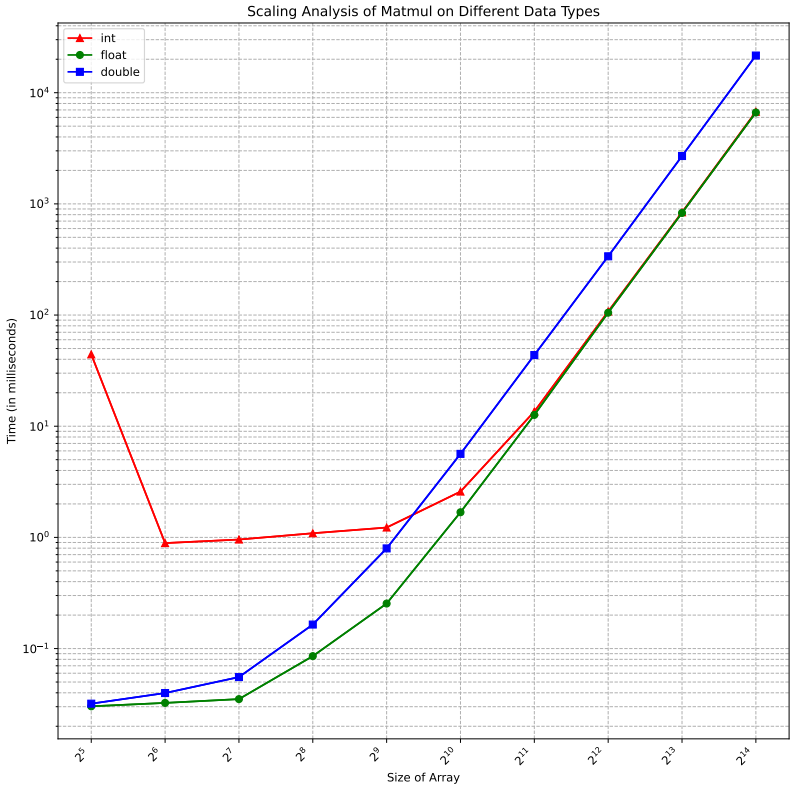

Figure 1: Scaling Analysis for task1 (matrix multiplication)

1.e

In the naive implementation, the runtime for the matmul function is 52469.1ms as compared to the tiled implementation which results in a time of 21599.4 ms. The tiled implementation is faster by a factor of 2.43 simply due to the fact that the memory access pattern is more efficient in the tiled implementation.

1.f

The CPU implementation times out after 10 minutes whereas the GPU implementation takes 21599.4 ms. The GPU implementation is faster and more efficient because of the exploitation of parallelism in the GPU.

2 Question 2

2.a

`task2.cu` can be found at https://github.com/phantom3012/repo759/blob/main/HW07/task2.cu

2.b

Scaling analysis plot for task2 is shown below:

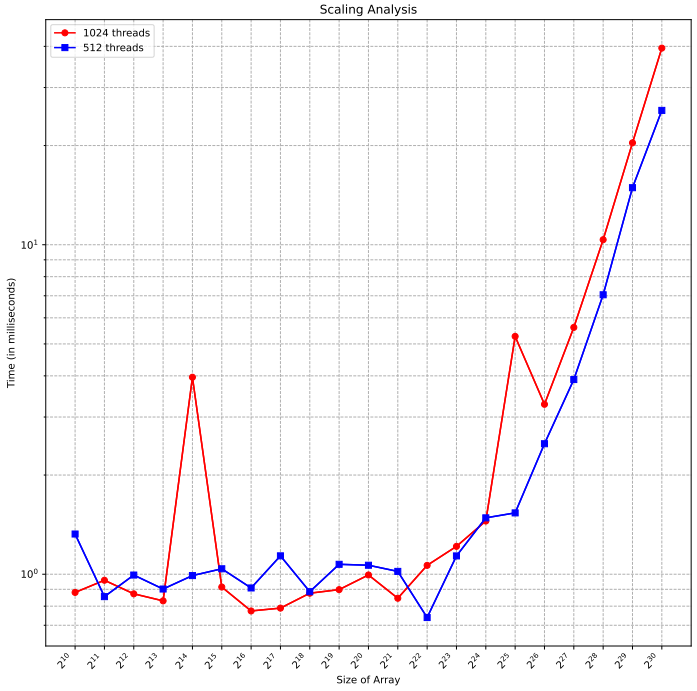

Figure 2: Scaling Analysis for task2 (reduction)