Homework 6 Report

1 Test Environment

I tested my code on a Google Compute Instance with NVIDIA Tesla V100 GPU, with CUDA 11.1.

2 Experiments and Results

2.1 CG using CSR

Max iteration: 20,000 Tolerance: 1e-6
Number of blocks: (m + threads - 1) / threads

Threads Per Block	Time (seconds)	norm of (b - A * x)
32	8.132622	4.43E-05
64	7.928369	4.43E-05
128	7.867733	4.43E-05
256	8.048368	4.43E-05
512	6.907522	4.43E-05
1024	10.923411	4.43E-05

Table 1.CG using CSR

2.2 CG using ELL (row-major with shared memory)

The skeleton code converts CSR to ELL by storing the first row, the second row, ..., the last row into the 1-D array.
Max iteration: 20,000 Tolerance: 1e-6
Number of blocks: m

Threads Per Block	Time (seconds)	norm of (b - A * x)
32	4.029764	6.45E-05
64	3.77639	5.27E-05
128	4.523906	4.93E-05
256	6.740506	4.93E-05
512	12.004819	4.93E-05
1024	25.205187	4.93E-05

Table 2.CG using ELL (row-major with shared memory)

2.3 CG using ELL (column-major without shared memory)

Most ELL format uses column-major format. This is also the way I converted CSR to ELL for homework5. It stores the first column, second column, ..., the last column into the 1-D array.
Max iteration: 20,000 Tolerance: 1e-6
Number of blocks: m

Threads Per Block	Time (seconds)	norm of (b - A * x)
32	3.718985	4.43E-05
64	3.690877	4.43E-05
128	3.677182	4.43E-05
256	3.682065	4.43E-05
512	3.685367	4.43E-05
1024	4.24437	4.43E-05

Table 3.CG using ELL (column-major without shared memory)

2.4 CG using ELL (column-major with shared memory)

This one tests column-major ELL with shared memory.
Max iteration: 20,000 Tolerance: 1e-6
Number of blocks: m

Threads Per Block	Time (seconds)	norm of (b - A * x)
32	9.118011	6.45E-05
64	9.165521	5.27E-05
128	9.534499	4.93E-05
256	10.45859	4.93E-05
512	13.044246	4.93E-05
1024	24.3799	4.93E-05

Table 4.CG using ELL (column-major with shared memory)

2.5 Larger Max Iterations

This tests the error for CSR and ELL vs. different max iteration.

Max Iteration	CSR Error	ELL Error
13000	9.891570e-04	9.891570e-04
14000	2.109754e-03	2.109754e-03
16000	4.004684e-04	4.004684e-04
18000	1.920288e-04	1.920288e-04
20000	4.431995e-05	4.431995e-05
22000	1.261112e-04	1.261112e-04

Table 5. Error vs. Max Iteration

2.6 Performance Comparison

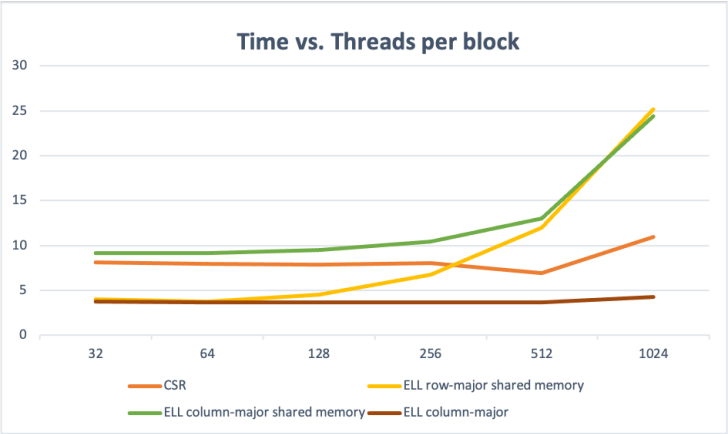

Figure 1. Performance Comparison

3 Findings

- ELL (column-major without shared memory) takes 1/2 time than CSR. This is because ELL uses padding to store the non-zero elements to store non-zero elements contiguously. By allowing coalesced memory access, it outperforms CSR. The trade-off is that ELL will end up taking more space to store the sparse matrix than CSR.
- By using row-major format ELL with shared memory, it performs slightly worse than column-major format ELL without shared memory. However, it performs better than column-major format with shared memory.
- When max iteration is 20,000, CG has the lowest error for CSR and ELL.