EECS 117

HW3 CUDA

3/11/2018

Hikaru Kasai (ID: 50765935)

Jay Patel (ID: 77742251)

HW3 Answers to Questions

**Part 1: Naive Parallel Reduction**

**a) Result**

=== Running 5 trials of naive ... ===

**\*\*\* Trial 1 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute naive GPU reduction kernel\_Part\_0: 0.003348 secs

Effective bandwidth: 10.02 GB/s

Time to execute naive CPU reduction: 0.116106 secs

SUCCESS: GPU: 41.958740 CPU: 41.958740

**\*\*\* Trial 2 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute naive GPU reduction kernel\_Part\_0: 0.003347 secs

Effective bandwidth: 10.03 GB/s

Time to execute naive CPU reduction: 0.116079 secs

SUCCESS: GPU: 41.958740 CPU: 41.958740

**\*\*\* Trial 3 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute naive GPU reduction kernel\_Part\_0: 0.003349 secs

Effective bandwidth: 10.02 GB/s

Time to execute naive CPU reduction: 0.116099 secs

SUCCESS: GPU: 41.946426 CPU: 41.946423

**\*\*\* Trial 4 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute naive GPU reduction kernel\_Part\_0: 0.003346 secs

Effective bandwidth: 10.03 GB/s

Time to execute naive CPU reduction: 0.116096 secs

SUCCESS: GPU: 41.946426 CPU: 41.946423

**\*\*\* Trial 5 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute naive GPU reduction kernel\_Part\_0: 0.003356 secs

Effective bandwidth: 10.00 GB/s

Time to execute naive CPU reduction: 0.116101 secs

SUCCESS: GPU: 41.938293 CPU: 41.938297

=== Done! ===

**b) Explain how the effective bandwidth is being calculated.**

The effective bandwidth is calculated by dividing the total (input size \* dtype) by the execution time for kernel0. This represents the total throughput by the GPU.

BW = (N \* sizeof(dtype))/time of execution

**Part 2: Strided Access by Consecutive Threads**

**a) Result**

=== Running 5 trials of stride ... ===

**\*\*\* Trial 1 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute strided index GPU reduction kernel\_Part\_2: 0.002217 secs

Effective bandwidth: 15.14 GB/s

Time to execute naive CPU reduction: 0.116158 secs

SUCCESS: GPU: 41.936550 CPU: 41.936550

**\*\*\* Trial 2 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute strided index GPU reduction kernel\_Part\_2: 0.002217 secs

Effective bandwidth: 15.14 GB/s

Time to execute naive CPU reduction: 0.116089 secs

SUCCESS: GPU: 41.936550 CPU: 41.936550

**\*\*\* Trial 3 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute strided index GPU reduction kernel\_Part\_2: 0.002216 secs

Effective bandwidth: 15.14 GB/s

Time to execute naive CPU reduction: 0.116092 secs

SUCCESS: GPU: 41.957001 CPU: 41.957005

**\*\*\* Trial 4 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute strided index GPU reduction kernel\_Part\_2: 0.002218 secs

Effective bandwidth: 15.13 GB/s

Time to execute naive CPU reduction: 0.116079 secs

SUCCESS: GPU: 41.957001 CPU: 41.957005

**\*\*\* Trial 5 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute strided index GPU reduction kernel\_Part\_2: 0.002236 secs

Effective bandwidth: 15.01 GB/s

Time to execute naive CPU reduction: 0.116002 secs

SUCCESS: GPU: 41.959137 CPU: 41.959137

=== Done! ===

**b) How much faster than the initial code is this version?**

The average GPU execution time for Naive is 0.003349 s, and the average for Stride is 0.00221 s. Therefore Stride is 0.0033492/0.002208 = 1.51 times faster than Naive.

The average bandwidth for Naive is 10.02 GB/s and Stride is 15.11 GB/s. Stride has 15.11/10.02 = 1.51 times more bandwidth than Naive.

**Part 3: Sequential Access by Consecutive Threads**

**a) Result**

=== Running 5 trials of sequential ... ===

**\*\*\* Trial 1 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute sequential index GPU reduction kernel: 0.001766 secs

Effective bandwidth: 19.00 GB/s

Time to execute naive CPU reduction: 0.116103 secs

SUCCESS: GPU: 41.940788 CPU: 41.940784

**\*\*\* Trial 2 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute sequential index GPU reduction kernel: 0.001767 secs

Effective bandwidth: 18.99 GB/s

Time to execute naive CPU reduction: 0.116453 secs

SUCCESS: GPU: 41.940788 CPU: 41.940784

**\*\*\* Trial 3 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute sequential index GPU reduction kernel: 0.00177 secs

Effective bandwidth: 18.96 GB/s

Time to execute naive CPU reduction: 0.116467 secs

SUCCESS: GPU: 41.940910 CPU: 41.940907

**\*\*\* Trial 4 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute sequential index GPU reduction kernel: 0.001769 secs

Effective bandwidth: 18.97 GB/s

Time to execute naive CPU reduction: 0.116349 secs

SUCCESS: GPU: 41.940910 CPU: 41.940907

**\*\*\* Trial 5 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute sequential index GPU reduction kernel: 0.001757 secs

Effective bandwidth: 19.10 GB/s

Time to execute naive CPU reduction: 0.099955 secs

SUCCESS: GPU: 41.957352 CPU: 41.957352

=== Done! ===

**b) Bandwidth improvement**

The average bandwidth is 19.004 GB/s. Improvement over stride is 19.004/15.11 = 1.26x. Improvement over naive is 19.004/10.002 = 1.9x.

**Part 4: First Add Before Reduce**

**a) Result**

=== Running 5 trials of first\_add ... ===

**\*\*\* Trial 1 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute first add GPU reduction kernel: 0.000963 secs

Effective bandwidth: 34.84 GB/s

Time to execute naive CPU reduction: 0.116016 secs

SUCCESS: GPU: 41.940533 CPU: 41.940533

**\*\*\* Trial 2 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute first add GPU reduction kernel: 0.000981 secs

Effective bandwidth: 34.20 GB/s

Time to execute naive CPU reduction: 0.116118 secs

SUCCESS: GPU: 41.940533 CPU: 41.940533

**\*\*\* Trial 3 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute first add GPU reduction kernel: 0.000981 secs

Effective bandwidth: 34.20 GB/s

Time to execute naive CPU reduction: 0.116057 secs

SUCCESS: GPU: 41.941795 CPU: 41.941795

**\*\*\* Trial 4 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute first add GPU reduction kernel: 0.000981 secs

Effective bandwidth: 34.20 GB/s

Time to execute naive CPU reduction: 0.116054 secs

SUCCESS: GPU: 41.941795 CPU: 41.941795

**\*\*\* Trial 5 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute first add GPU reduction kernel: 0.000981 secs

Effective bandwidth: 34.20 GB/s

Time to execute naive CPU reduction: 0.099568 secs

SUCCESS: GPU: 41.941795 CPU: 41.941795

=== Done! ===

**b) Bandwidth improvement**

The average bandwidth is 35.13 GB/s. Compared to sequential, improvement in bandwidth is 35.13/19.004 = 1.85x greater throughput. Compared to stride, it is 35.13/15.11 = 2.32x greater. Compared to naive, it is 35.13/10.004 = 3.51x greater.

**Part 5: Unroll the Last Warp**

**a) Result**

=== Running 5 trials of unroll ... ===

**\*\*\* Trial 1 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute unrolled GPU reduction kernel: 0.000619 secs

Effective bandwidth: 54.21 GB/s

Time to execute naive CPU reduction: 0.11625 secs

SUCCESS: GPU: 41.937187 CPU: 41.937187

**\*\*\* Trial 2 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute unrolled GPU reduction kernel: 0.000619 secs

Effective bandwidth: 54.21 GB/s

Time to execute naive CPU reduction: 0.116036 secs

SUCCESS: GPU: 41.937569 CPU: 41.937569

**\*\*\* Trial 3 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute unrolled GPU reduction kernel: 0.000597 secs

Effective bandwidth: 56.21 GB/s

Time to execute naive CPU reduction: 0.11602 secs

SUCCESS: GPU: 41.955223 CPU: 41.955223

**\*\*\* Trial 4 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute unrolled GPU reduction kernel: 0.000617 secs

Effective bandwidth: 54.38 GB/s

Time to execute naive CPU reduction: 0.116032 secs

SUCCESS: GPU: 41.954903 CPU: 41.954906

**\*\*\* Trial 5 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute unrolled GPU reduction kernel: 0.00061 secs

Effective bandwidth: 55.01 GB/s

Time to execute naive CPU reduction: 0.116043 secs

SUCCESS: GPU: 41.942497 CPU: 41.942497

=== Done! ===

**b) Bandwidth improvement**

The average bandwidth is 55.804 GB/s. Compared to first add, improvement in bandwidth is 55.804/35.128 = 1.58x greater throughput. Compared to sequential, it is 55.804/19.004 = 2.93x greater. Compared to stride, it is 55.804/15.112 = 3.69x greater. Compared to naive, it is 55.804/10.004 = 5.57x greater.

**Part 6: Algorithm Cascading**

**a) Result**

=== Running 5 trials of multiple ... ===

**\*\*\* Trial 1 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute multiple add GPU reduction kernel: 0.000343 secs

Effective bandwidth: 97.83 GB/s

Time to execute naive CPU reduction: 0.11605 secs

SUCCESS: GPU: 41.947502 CPU: 41.947506

**\*\*\* Trial 2 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute multiple add GPU reduction kernel: 0.000328 secs

Effective bandwidth: 102.30 GB/s

Time to execute naive CPU reduction: 0.116021 secs

SUCCESS: GPU: 41.938148 CPU: 41.938145

**\*\*\* Trial 3 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute multiple add GPU reduction kernel: 0.000328 secs

Effective bandwidth: 102.30 GB/s

Time to execute naive CPU reduction: 0.117042 secs

SUCCESS: GPU: 41.935520 CPU: 41.935520

**\*\*\* Trial 4 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute multiple add GPU reduction kernel: 0.000332 secs

Effective bandwidth: 101.07 GB/s

Time to execute naive CPU reduction: 0.116085 secs

SUCCESS: GPU: 41.948006 CPU: 41.948006

**\*\*\* Trial 5 \*\*\***

N: 8388608

Timer: gettimeofday

Timer resolution: ~ 1 us (?)

Time to execute multiple add GPU reduction kernel: 0.000327 secs

Effective bandwidth: 102.61 GB/s

Time to execute naive CPU reduction: 0.116072 secs

SUCCESS: GPU: 41.941360 CPU: 41.941357

=== Done! ===

**b) Bandwidth improvement**

The average bandwidth is 102.07 GB/s. Compared to unroll, it is 102.07/55.804 = 1.83x bandwidth improvement. Compared to sequential, improvement in bandwidth is 102.07/19.004 = 5.37x greater throughput. In comparison to stride, it is 107.07/15.11 = 7.09x greater. Compared to naive, it is 107.07/10.004 = 10.7x greater. Overall, the final improvement is more than 10x greater bandwidth which aligns with the discussion we had during lecture on HW3 hints.