

Advanced Computer Architecture (ACA) GPU (CUDA) Vector Reduction

Vijoy Sunil Kumar

Files

GPU implementation: vectorReduce.cu (Normal GPU implementation)
vectorReduce_atomic.cu (Atomic ADD support)
vectorReduce_noif.cu (Without the if statement in the kernel)

Usage

Makefile changes to build vectorReduce.cu

```
# Add source files here
EXECUTABLE      := vectorReduce
# Cuda source files (compiled with cudacc)
CUFILES_sm_11   := vectorReduce.cu

# C/C++ source files (compiled with gcc / g++)
CCFILES         := \

#####
# Rules and targets

include ../../common/common.mk
```

Makefile changes to build vectorReduce_atomic.cu

```
# Add source files here
EXECUTABLE      := vectorReduce_atomic
# Cuda source files (compiled with cudacc)
CUFILES_sm_11   := vectorReduce_atomic.cu

# C/C++ source files (compiled with gcc / g++)
CCFILES         := \

#####
# Rules and targets

include ../../common/common.mk
```

Makefile changes to build vectorReduce_noif.cu

```
# Add source files here
EXECUTABLE      := vectorReduce_noif
# Cuda source files (compiled with cudacc)
CUFILES_sm_11   := vectorReduce_noif.cu

# C/C++ source files (compiled with gcc / g++)
CCFILES         := \

#####
# Rules and targets

include ../../common/common.mk
```

Running vectorReduce

`./vectorReduce -size 1000 -block 32`

Running vectorReduce_atomic

`./vectorReduce_atomic -size 1000 -block 32`

Running vectorReduce_noif

`./vectorReduce_noif -size 1024 -block 32`

[Item 2] - Provide the execution time and memory transfer time for the following data sizes:

All time values are in milliseconds

Input Size	Blocksize	GPU Execution Time	Memory Transfer Time	CPU Time (to add partial sums)	Overall Execution Time (Memory + GPU Execution)
1000	32	0.040	0.068	0.005	0.103
10000	32	0.047	0.098	0.048	0.126
100000	32	0.115	0.393	0.476	0.349
1000000	32	0.650	2.923	4.770	2.210
2000000	32	1.053	5.660	9.540	4.070
1000	64	0.040	0.067	0.004	0.103
10000	64	0.156	0.103	0.048	0.240
100000	64	0.116	0.379	0.474	0.342
1000000	64	0.630	2.939	4.765	2.210
2000000	64	1.056	5.616	9.564	4.034
1000	128	0.040	0.068	0.005	0.102
10000	128	0.044	0.101	0.047	0.125
100000	128	0.116	0.377	0.474	0.342
1000000	128	0.647	3.072	4.753	2.358
2000000	128	1.079	5.686	9.562	4.090
1000	256	0.040	0.070	0.004	0.105
10000	256	0.044	0.101	0.047	0.125
100000	256	0.209	0.379	0.474	0.438
1000000	256	0.679	2.910	4.766	2.235
2000000	256	1.059	5.680	9.566	4.071

Screenshots

Vector Reduction run on GPU with different array sizes and block size = 32

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 1000 -block 32
Vector reduction: size 1000
CPU Reduction time: 0.005000 (ms)
CPU to GPU Transfer Time: 0.025000 (ms)
GPU Execution time: 0.040000 (ms)
GPU to CPU Transfer Time: 0.043000 (ms)
Overall Execution Time (Memory + GPU): 0.103000 (ms)
CPU Partial Sums Execution time: 0.005000 (ms)
CPU results : 508.125519
GPU results : 508.125549
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 10000 -block 32
Vector reduction: size 10000
CPU Reduction time: 0.047000 (ms)
CPU to GPU Transfer Time: 0.040000 (ms)
GPU Execution time: 0.047000 (ms)
GPU to CPU Transfer Time: 0.058000 (ms)
Overall Execution Time (Memory + GPU): 0.126000 (ms)
CPU Partial Sums Execution time: 0.048000 (ms)
CPU results : 4971.319336
GPU results : 4971.321777
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 100000 -block 32
Vector reduction: size 100000
CPU Reduction time: 0.470000 (ms)
CPU to GPU Transfer Time: 0.181000 (ms)
GPU Execution time: 0.115000 (ms)
GPU to CPU Transfer Time: 0.212000 (ms)
Overall Execution Time (Memory + GPU): 0.349000 (ms)
CPU Partial Sums Execution time: 0.476000 (ms)
CPU results : 49986.671875
GPU results : 49986.851562
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ █
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 1000000 -block 32
Vector reduction: size 1000000
CPU Reduction time: 4.720000 (ms)
CPU to GPU Transfer Time: 1.386000 (ms)
GPU Execution time: 0.650000 (ms)
GPU to CPU Transfer Time: 1.537000 (ms)
Overall Execution Time (Memory + GPU): 2.211000 (ms)
CPU Partial Sums Execution time: 4.770000 (ms)
CPU results : 500003.750000
GPU results : 500006.281250
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 2000000 -block 32
Vector reduction: size 2000000
CPU Reduction time: 9.454000 (ms)
CPU to GPU Transfer Time: 2.672000 (ms)
GPU Execution time: 1.053000 (ms)
GPU to CPU Transfer Time: 2.988000 (ms)
Overall Execution Time (Memory + GPU): 4.070000 (ms)
CPU Partial Sums Execution time: 9.549001 (ms)
CPU results : 999969.375000
GPU results : 999936.750000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ █
```

Vector Reduction run on GPU with different array sizes and block size = 64

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 1000 -block 64
Vector reduction: size 1000
CPU Reduction time: 0.004000 (ms)
CPU to GPU Transfer Time: 0.025000 (ms)
GPU Execution time: 0.040000 (ms)
GPU to CPU Transfer Time: 0.042000 (ms)
Overall Execution Time (Memory + GPU): 0.103000 (ms)
CPU Partial Sums Execution time: 0.004000 (ms)
CPU results : 508.125519
GPU results : 508.125549
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 10000 -block 64
Vector reduction: size 10000
CPU Reduction time: 0.047000 (ms)
CPU to GPU Transfer Time: 0.040000 (ms)
GPU Execution time: 0.156000 (ms)
GPU to CPU Transfer Time: 0.063000 (ms)
Overall Execution Time (Memory + GPU): 0.240000 (ms)
CPU Partial Sums Execution time: 0.048000 (ms)
CPU results : 4971.319336
GPU results : 4971.321777
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 100000 -block 64
Vector reduction: size 100000
CPU Reduction time: 0.469000 (ms)
CPU to GPU Transfer Time: 0.174000 (ms)
GPU Execution time: 0.116000 (ms)
GPU to CPU Transfer Time: 0.205000 (ms)
Overall Execution Time (Memory + GPU): 0.342000 (ms)
CPU Partial Sums Execution time: 0.474000 (ms)
CPU results : 49986.671875
GPU results : 49986.851562
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ █
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 1000000 -block 64
Vector reduction: size 1000000
CPU Reduction time: 4.717000 (ms)
CPU to GPU Transfer Time: 1.384000 (ms)
GPU Execution time: 0.630000 (ms)
GPU to CPU Transfer Time: 1.555000 (ms)
Overall Execution Time (Memory + GPU): 2.210000 (ms)
CPU Partial Sums Execution time: 4.765000 (ms)
CPU results : 500003.750000
GPU results : 500006.281250
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 2000000 -block 64
Vector reduction: size 2000000
CPU Reduction time: 9.470000 (ms)
CPU to GPU Transfer Time: 2.667000 (ms)
GPU Execution time: 1.056000 (ms)
GPU to CPU Transfer Time: 2.949000 (ms)
Overall Execution Time (Memory + GPU): 4.034000 (ms)
CPU Partial Sums Execution time: 9.564000 (ms)
CPU results : 999969.375000
GPU results : 999936.750000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ █
```


Vector Reduction run on GPU with different array sizes and block size = 128

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 1000 -block 128
Vector reduction: size 1000
CPU Reduction time: 0.005000 (ms)
CPU to GPU Transfer Time: 0.026000 (ms)
GPU Execution time: 0.040000 (ms)
GPU to CPU Transfer Time: 0.042000 (ms)
Overall Execution Time (Memory + GPU): 0.102000 (ms)
CPU Partial Sums Execution time: 0.005000 (ms)
CPU results : 508.125519
GPU results : 508.125549
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 10000 -block 128
Vector reduction: size 10000
CPU Reduction time: 0.047000 (ms)
CPU to GPU Transfer Time: 0.041000 (ms)
GPU Execution time: 0.044000 (ms)
GPU to CPU Transfer Time: 0.060000 (ms)
Overall Execution Time (Memory + GPU): 0.125000 (ms)
CPU Partial Sums Execution time: 0.047000 (ms)
CPU results : 4971.319336
GPU results : 4971.321777
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 100000 -block 128
Vector reduction: size 100000
CPU Reduction time: 0.469000 (ms)
CPU to GPU Transfer Time: 0.173000 (ms)
GPU Execution time: 0.116000 (ms)
GPU to CPU Transfer Time: 0.204000 (ms)
Overall Execution Time (Memory + GPU): 0.342000 (ms)
CPU Partial Sums Execution time: 0.474000 (ms)
CPU results : 49986.671875
GPU results : 49986.851562
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ █
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 1000000 -block 128
Vector reduction: size 1000000
CPU Reduction time: 4.704000 (ms)
CPU to GPU Transfer Time: 1.387000 (ms)
GPU Execution time: 0.647000 (ms)
GPU to CPU Transfer Time: 1.685000 (ms)
Overall Execution Time (Memory + GPU): 2.358000 (ms)
CPU Partial Sums Execution time: 4.753000 (ms)
CPU results : 500003.750000
GPU results : 500006.281250
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 2000000 -block 128
Vector reduction: size 2000000
CPU Reduction time: 9.467000 (ms)
CPU to GPU Transfer Time: 2.699000 (ms)
GPU Execution time: 1.079000 (ms)
GPU to CPU Transfer Time: 2.987000 (ms)
Overall Execution Time (Memory + GPU): 4.090000 (ms)
CPU Partial Sums Execution time: 9.562000 (ms)
CPU results : 999969.375000
GPU results : 999936.750000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ █
```

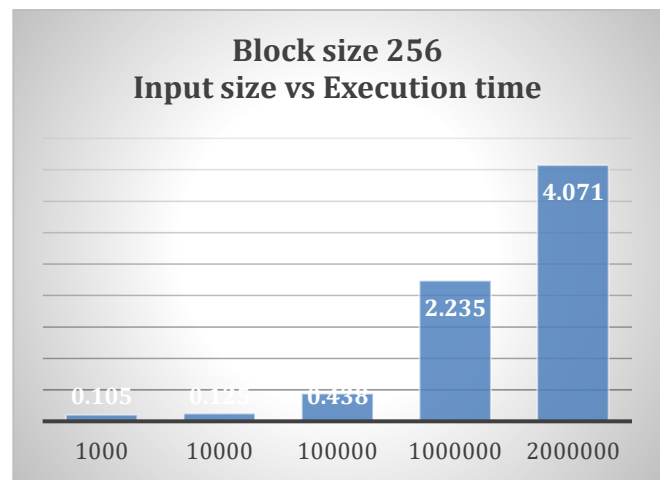
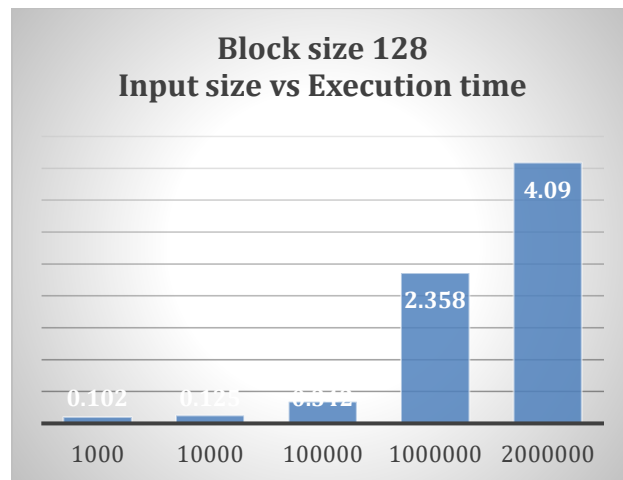
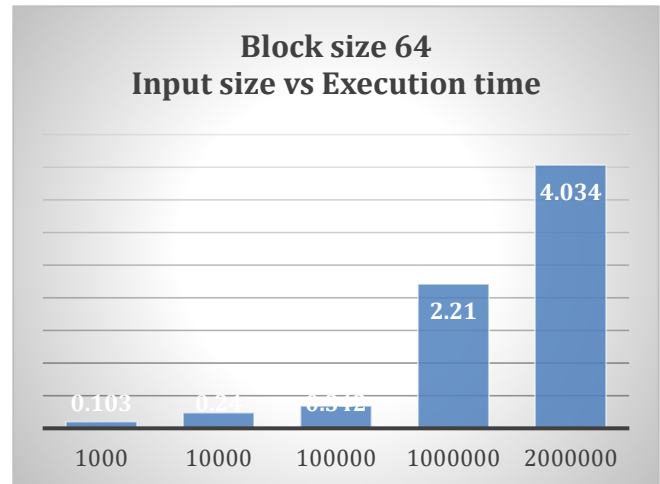
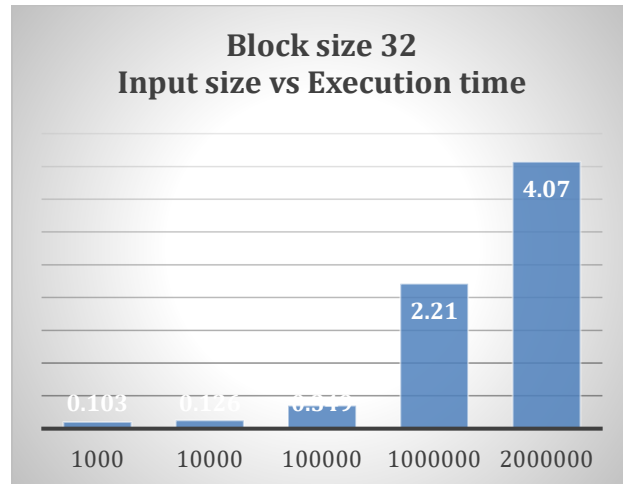
Vector Reduction run on GPU with different array sizes and block size = 256

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduce -size 1000 -block 256
Vector reduction: size 1000
CPU Reduction time: 0.004000 (ms)
CPU to GPU Transfer Time: 0.026000 (ms)
GPU Execution time: 0.040000 (ms)
GPU to CPU Transfer Time: 0.044000 (ms)
Overall Execution Time (Memory + GPU): 0.105000 (ms)
CPU Partial Sums Execution time: 0.004000 (ms)
CPU results : 508.125519
GPU results : 508.125549
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduce -size 10000 -block 256
Vector reduction: size 10000
CPU Reduction time: 0.047000 (ms)
CPU to GPU Transfer Time: 0.041000 (ms)
GPU Execution time: 0.044000 (ms)
GPU to CPU Transfer Time: 0.060000 (ms)
Overall Execution Time (Memory + GPU): 0.125000 (ms)
CPU Partial Sums Execution time: 0.047000 (ms)
CPU results : 4971.319336
GPU results : 4971.321777
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduce -size 100000 -block 256
Vector reduction: size 100000
CPU Reduction time: 0.469000 (ms)
CPU to GPU Transfer Time: 0.171000 (ms)
GPU Execution time: 0.209000 (ms)
GPU to CPU Transfer Time: 0.208000 (ms)
Overall Execution Time (Memory + GPU): 0.438000 (ms)
CPU Partial Sums Execution time: 0.474000 (ms)
CPU results : 49986.671875
GPU results : 49986.851562
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduce -size 1000000 -block 256
Vector reduction: size 1000000
CPU Reduction time: 4.717000 (ms)
CPU to GPU Transfer Time: 1.378000 (ms)
GPU Execution time: 0.679000 (ms)
GPU to CPU Transfer Time: 1.532000 (ms)
Overall Execution Time (Memory + GPU): 2.235000 (ms)
CPU Partial Sums Execution time: 4.766000 (ms)
CPU results : 500003.750000
GPU results : 500006.281250
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduce -size 2000000 -block 256
Vector reduction: size 2000000
CPU Reduction time: 9.471000 (ms)
CPU to GPU Transfer Time: 2.694000 (ms)
GPU Execution time: 1.059000 (ms)
GPU to CPU Transfer Time: 2.986000 (ms)
Overall Execution Time (Memory + GPU): 4.071000 (ms)
CPU Partial Sums Execution time: 9.566000 (ms)
CPU results : 999969.375000
GPU results : 999936.750000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$
```

[Item 3] – Graph the result using Excel or any plotting tool, clearly showing the overall execution time versus Input Size and Block Size.

Execution time vs Input size and Block size



[Item 4] – You must MODIFY the CUDA kernel to perform the following:

Kernel 1: Similar to the regular original CUDA reduction kernel, except each block should use the atomic support operations to add their local_block_sum to the collective total. In this case, only one thread of the block should perform the atomicAdd:

```
atomicAdd(d_Total, local_block_sum);
```

In the case of the reduction, the resulting answer will be a single value and require no CPU computation, but still needs to be transferred back to the CPU. Compare the performance to the previous item (2), include all memory transfer times:

Input Size	Blocksize	Previous total execution (CPU+GPU)	Total execution (atomic support in GPU)	Speedup %
1000	32	0.108	0.114	NA
10000	32	0.174	0.139	20.1
100000	32	0.825	0.393	52.3
1000000	32	6.98	2.185	68.6
2000000	32	13.61	4.049	70.2
1000	64	0.107	0.114	NA
10000	64	0.288	0.138	52.0
100000	64	0.816	0.393	51.8
1000000	64	6.975	2.328	66.6
2000000	64	13.598	4.021	70.4
1000	128	0.107	0.112	NA
10000	128	0.172	0.138	19.7
100000	128	0.816	0.392	51.9
1000000	128	7.111	2.194	69.1
2000000	128	13.652	4.044	70.3
1000	256	0.109	0.115	NA
10000	256	0.172	0.138	19.7
100000	256	0.912	0.419	54.0
1000000	256	7.001	2.197	68.8
2000000	256	13.637	3.980	70.8

Screenshots

Atomic Vector Reduction run on GPU with different array sizes and block size = 32

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReductionAtomic -size 1000 -block 32
Vector reduction: size 1000
CPU Reduction time: 0.001000 (ms)
CPU to GPU Transfer Time: 0.035000 (ms)
GPU Execution time: 0.040000 (ms)
GPU to CPU Transfer Time: 0.017000 (ms)
Overall Execution Time (Memory + GPU): 0.114000 (ms)
CPU results : 1000
GPU results : 1000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReductionAtomic -size 10000 -block 32
Vector reduction: size 10000
CPU Reduction time: 0.010000 (ms)
CPU to GPU Transfer Time: 0.056000 (ms)
GPU Execution time: 0.044000 (ms)
GPU to CPU Transfer Time: 0.016000 (ms)
Overall Execution Time (Memory + GPU): 0.139000 (ms)
CPU results : 10000
GPU results : 10000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReductionAtomic -size 100000 -block 32
Vector reduction: size 100000
CPU Reduction time: 0.168000 (ms)
CPU to GPU Transfer Time: 0.268000 (ms)
GPU Execution time: 0.086000 (ms)
GPU to CPU Transfer Time: 0.016000 (ms)
Overall Execution Time (Memory + GPU): 0.393000 (ms)
CPU results : 100000
GPU results : 100000
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReductionAtomic -size 1000000 -block 32
Vector reduction: size 1000000
CPU Reduction time: 0.954000 (ms)
CPU to GPU Transfer Time: 1.611000 (ms)
GPU Execution time: 0.532000 (ms)
GPU to CPU Transfer Time: 0.016000 (ms)
Overall Execution Time (Memory + GPU): 2.185000 (ms)
CPU results : 1000000
GPU results : 1000000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReductionAtomic -size 2000000 -block 32
Vector reduction: size 2000000
CPU Reduction time: 2.021000 (ms)
CPU to GPU Transfer Time: 2.936000 (ms)
GPU Execution time: 1.062000 (ms)
GPU to CPU Transfer Time: 0.021000 (ms)
Overall Execution Time (Memory + GPU): 4.049000 (ms)
CPU results : 2000000
GPU results : 2000000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$
```

Atomic Vector Reduction run on GPU with different array sizes and block size = 64

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 1000 -block 64
Vector reduction: size 1000
CPU Reduction time: 0.001000 (ms)
CPU to GPU Transfer Time: 0.035000 (ms)
GPU Execution time: 0.039000 (ms)
GPU to CPU Transfer Time: 0.017000 (ms)
Overall Execution Time (Memory + GPU): 0.114000 (ms)
CPU results : 1000
GPU results : 1000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 10000 -block 64
Vector reduction: size 10000
CPU Reduction time: 0.009000 (ms)
CPU to GPU Transfer Time: 0.056000 (ms)
GPU Execution time: 0.044000 (ms)
GPU to CPU Transfer Time: 0.016000 (ms)
Overall Execution Time (Memory + GPU): 0.138000 (ms)
CPU results : 10000
GPU results : 10000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 100000 -block 64
Vector reduction: size 100000
CPU Reduction time: 0.174000 (ms)
CPU to GPU Transfer Time: 0.267000 (ms)
GPU Execution time: 0.087000 (ms)
GPU to CPU Transfer Time: 0.016000 (ms)
Overall Execution Time (Memory + GPU): 0.393000 (ms)
CPU results : 100000
GPU results : 100000
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 1000000 -block 64
Vector reduction: size 1000000
CPU Reduction time: 0.955000 (ms)
CPU to GPU Transfer Time: 1.639000 (ms)
GPU Execution time: 0.643000 (ms)
GPU to CPU Transfer Time: 0.021000 (ms)
Overall Execution Time (Memory + GPU): 2.328000 (ms)
CPU results : 1000000
GPU results : 1000000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 2000000 -block 64
Vector reduction: size 2000000
CPU Reduction time: 2.025000 (ms)
CPU to GPU Transfer Time: 2.923000 (ms)
GPU Execution time: 1.047000 (ms)
GPU to CPU Transfer Time: 0.021000 (ms)
Overall Execution Time (Memory + GPU): 4.021000 (ms)
CPU results : 2000000
GPU results : 2000000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$
```


Atomic Vector Reduction run on GPU with different array sizes and block size = 128

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 1000 -block 128
Vector reduction: size 1000
CPU Reduction time: 0.001000 (ms)
CPU to GPU Transfer Time: 0.034000 (ms)
GPU Execution time: 0.039000 (ms)
GPU to CPU Transfer Time: 0.016000 (ms)
Overall Execution Time (Memory + GPU): 0.112000 (ms)
CPU results : 1000
GPU results : 1000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 10000 -block 128
Vector reduction: size 10000
CPU Reduction time: 0.010000 (ms)
CPU to GPU Transfer Time: 0.056000 (ms)
GPU Execution time: 0.044000 (ms)
GPU to CPU Transfer Time: 0.139000 (ms)
Overall Execution Time (Memory + GPU): 0.261000 (ms)
CPU results : 10000
GPU results : 10000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 10000 -block 128
Vector reduction: size 10000
CPU Reduction time: 0.010000 (ms)
CPU to GPU Transfer Time: 0.056000 (ms)
GPU Execution time: 0.044000 (ms)
GPU to CPU Transfer Time: 0.016000 (ms)
Overall Execution Time (Memory + GPU): 0.138000 (ms)
CPU results : 10000
GPU results : 10000
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 1000000 -block 128
Vector reduction: size 1000000
CPU Reduction time: 0.954000 (ms)
CPU to GPU Transfer Time: 1.613000 (ms)
GPU Execution time: 0.532000 (ms)
GPU to CPU Transfer Time: 0.021000 (ms)
Overall Execution Time (Memory + GPU): 2.194000 (ms)
CPU results : 1000000
GPU results : 1000000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 2000000 -block 128
Vector reduction: size 2000000
CPU Reduction time: 2.026000 (ms)
CPU to GPU Transfer Time: 2.924000 (ms)
GPU Execution time: 1.071000 (ms)
GPU to CPU Transfer Time: 0.021000 (ms)
Overall Execution Time (Memory + GPU): 4.044000 (ms)
CPU results : 2000000
GPU results : 2000000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$
```

Atomic Vector Reduction run on GPU with different array sizes and block size = 256

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 1000 -block 256
Vector reduction: size 1000
CPU Reduction time: 0.001000 (ms)
CPU to GPU Transfer Time: 0.036000 (ms)
GPU Execution time: 0.040000 (ms)
GPU to CPU Transfer Time: 0.016000 (ms)
Overall Execution Time (Memory + GPU): 0.115000 (ms)
CPU results : 1000
GPU results : 1000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 10000 -block 256
Vector reduction: size 10000
CPU Reduction time: 0.009000 (ms)
CPU to GPU Transfer Time: 0.055000 (ms)
GPU Execution time: 0.044000 (ms)
GPU to CPU Transfer Time: 0.016000 (ms)
Overall Execution Time (Memory + GPU): 0.138000 (ms)
CPU results : 10000
GPU results : 10000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 100000 -block 256
Vector reduction: size 100000
CPU Reduction time: 0.169000 (ms)
CPU to GPU Transfer Time: 0.267000 (ms)
GPU Execution time: 0.086000 (ms)
GPU to CPU Transfer Time: 0.045000 (ms)
Overall Execution Time (Memory + GPU): 0.419000 (ms)
CPU results : 100000
GPU results : 100000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ █
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 1000000 -block 256
Vector reduction: size 1000000
CPU Reduction time: 0.959000 (ms)
CPU to GPU Transfer Time: 1.624000 (ms)
GPU Execution time: 0.532000 (ms)
GPU to CPU Transfer Time: 0.016000 (ms)
Overall Execution Time (Memory + GPU): 2.197000 (ms)
CPU results : 1000000
GPU results : 1000000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_atomic -size 2000000 -block 256
Vector reduction: size 2000000
CPU Reduction time: 2.011000 (ms)
CPU to GPU Transfer Time: 2.901000 (ms)
GPU Execution time: 1.028000 (ms)
GPU to CPU Transfer Time: 0.021000 (ms)
Overall Execution Time (Memory + GPU): 3.980000 (ms)
CPU results : 2000000
GPU results : 2000000
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ █
```


[Item 5] – You must examine the overhead of performing the runtime check within the kernel of

```

if (globalid < N) {
... // work
}

```

at runtime for values that are a multiple of the blocksize. In the previous experiments, you examined values 1000, 10000, etc. For this experiment, you will perform the following : 1024, 4096, 16384, 262144, 1048576. The difference is that these are multiples of the blocksize, so there is no need to include the if check, the work is always execute. The goal of this task is to determine the overhead that the if statement incurs for kernel execution. Note, in the case of N=1000 and blocksize of 32, the if statement overhead is incurred for the first 31 blocks [Block 0-30] even though the computation (work) is always entered, but the if statement matters only to the last block [block 31] where some of the threads of the last block execute and some do not.

Input Size	GPU Overall Execution Time (blocksize=32) with the if statement present	GPU Overall Execution Time (blocksize=32) without the if statement present	Percentage different in performance
1024	0.104	0.104	NA
4096	0.113	0.112	0.7
16384	0.141	0.140	0.7
262144	0.844	0.835	1.06
1048576	2.478	2.439	1.5

Screenshots

Vector Reduction with If statement run with different array sizes

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 1024 -block 32
Vector reduction: size 1024
CPU Reduction time: 0.011000 (ms)
CPU to GPU Transfer Time: 0.026000 (ms)
GPU Execution time: 0.039000 (ms)
GPU to CPU Transfer Time: 0.017000 (ms)
Overall Execution Time (Memory + GPU): 0.104000 (ms)
CPU Partial Sums Execution time: 0.011000 (ms)
CPU results : 518.912537
GPU results : 518.912659
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 4096 -block 32
Vector reduction: size 4096
CPU Reduction time: 0.019000 (ms)
CPU to GPU Transfer Time: 0.033000 (ms)
GPU Execution time: 0.041000 (ms)
GPU to CPU Transfer Time: 0.017000 (ms)
Overall Execution Time (Memory + GPU): 0.113000 (ms)
CPU Partial Sums Execution time: 0.020000 (ms)
CPU results : 2049.713867
GPU results : 2049.712402
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 16384 -block 32
Vector reduction: size 16384
CPU Reduction time: 0.077000 (ms)
CPU to GPU Transfer Time: 0.052000 (ms)
GPU Execution time: 0.047000 (ms)
GPU to CPU Transfer Time: 0.018000 (ms)
Overall Execution Time (Memory + GPU): 0.141000 (ms)
CPU Partial Sums Execution time: 0.078000 (ms)
CPU results : 8143.619141
GPU results : 8143.619629
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 262144 -block 32
Vector reduction: size 262144
CPU Reduction time: 1.237000 (ms)
CPU to GPU Transfer Time: 0.427000 (ms)
GPU Execution time: 0.342000 (ms)
GPU to CPU Transfer Time: 0.052000 (ms)
Overall Execution Time (Memory + GPU): 0.844000 (ms)
CPU Partial Sums Execution time: 1.250000 (ms)
CPU results : 131072.468750
GPU results : 131074.328125
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction -size 1048576 -block 32
Vector reduction: size 1048576
CPU Reduction time: 4.946000 (ms)
CPU to GPU Transfer Time: 1.503000 (ms)
GPU Execution time: 0.777000 (ms)
GPU to CPU Transfer Time: 0.173000 (ms)
Overall Execution Time (Memory + GPU): 2.478000 (ms)
CPU Partial Sums Execution time: 5.034000 (ms)
CPU results : 524281.968750
GPU results : 524285.343750
```

Vector Reduction with NO If statement run with different array sizes

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_noif -size 1024 -block 32
Vector reduction: size 1024
CPU Reduction time: 0.005000 (ms)
CPU to GPU Transfer Time: 0.026000 (ms)
GPU Execution time: 0.040000 (ms)
GPU to CPU Transfer Time: 0.042000 (ms)
Overall Execution Time (Memory + GPU): 0.103000 (ms)
CPU Partial Sums Execution time: 0.006000 (ms)
CPU results : 518.912537
GPU results : 518.912659
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_noif -size 4096 -block 32
Vector reduction: size 4096
CPU Reduction time: 0.019000 (ms)
CPU to GPU Transfer Time: 0.032000 (ms)
GPU Execution time: 0.041000 (ms)
GPU to CPU Transfer Time: 0.017000 (ms)
Overall Execution Time (Memory + GPU): 0.113000 (ms)
CPU Partial Sums Execution time: 0.019000 (ms)
CPU results : 2049.713867
GPU results : 2049.712402
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_noif -size 16384 -block 32
Vector reduction: size 16384
CPU Reduction time: 0.077000 (ms)
CPU to GPU Transfer Time: 0.052000 (ms)
GPU Execution time: 0.047000 (ms)
GPU to CPU Transfer Time: 0.018000 (ms)
Overall Execution Time (Memory + GPU): 0.140000 (ms)
CPU Partial Sums Execution time: 0.078000 (ms)
CPU results : 8143.619141
GPU results : 8143.619629
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_noif -size 262144 -block 32
Vector reduction: size 262144
CPU Reduction time: 1.249000 (ms)
CPU to GPU Transfer Time: 0.403000 (ms)
GPU Execution time: 0.342000 (ms)
GPU to CPU Transfer Time: 0.459000 (ms)
Overall Execution Time (Memory + GPU): 0.822000 (ms)
CPU Partial Sums Execution time: 1.262000 (ms)
CPU results : 131072.468750
GPU results : 131074.328125
```

```
visu3975@ecee-gpu5:~/NVIDIA_GPU_Computing_SDK/C/bin/linux/release$ ./vectorReduction_noif -size 1048576 -block 32
Vector reduction: size 1048576
CPU Reduction time: 4.946000 (ms)
CPU to GPU Transfer Time: 1.499000 (ms)
GPU Execution time: 0.698000 (ms)
GPU to CPU Transfer Time: 0.247000 (ms)
Overall Execution Time (Memory + GPU): 2.466000 (ms)
CPU Partial Sums Execution time: 5.005000 (ms)
CPU results : 524281.968750
GPU results : 524285.343750
```