

The tests of the histogram using global memory, and shared memory were approximately 6.721us, and 53.314us for the histogram kernels. In the shared memory, the truncate kernel for shared memory took an extra 2.529us for a total of 55843us.

These results are surprising as the shared memory is significantly slower than the global memory kernel. This may be caused by read/write collisions, or the use of 2x the amount of `atomicAdd()` in the shared kernel vs the global kernel.

My shared kernel runs in a near similar time to the reference executable provided for us, which ran at 53475us on my computer.

The results of this lab suggest that it may be preferable to write simpler kernels that use only global memory, rather than waste time writing kernels that use shared memory.

I had difficulty in this lab as in my global memory kernel, there was an error where the first time `atomicAdd` was called, it would return a garbage result instead of the expected result. The error has since been fixed, however, I don't know which of my changes caused it to start working correctly.

Pictures of NSIGHT analysis are provided below.



CUDA Summary

CUDA Devices [All](#)

	Device ID	Device Name	Contexts	Device %	Host to Device (Bytes)	Device to Host (Bytes)
1	[0]	GPU 0 - GeForce GTX 750 Ti	1	0.00 %	2,016,384	16,384

CUDA Contexts

	Total	No Context	1	
CUDA Device ID		-	0	
Runtime API Calls Summary All				
# Calls	18	0	18	
# Errors	0	0	0	
% Time	65.52	0.00	65.52	
Driver API Calls Summary All				
# Calls	87	87	0	
# Errors	0	0	0	
% Time	0.31	0.31	0.00	
Launches Summary All				
# Launches	1	0	1	
% Device Time	0.00	0.00	0.00	
Memory Copies All				
H to D # Copies	2	0	2	
H to D # Bytes	2,016,384	0	2,016,384	
H to D % Time	0.1	0.0	0.1	
D to H # Copies	1	0	1	
D to H # Bytes	16,384	0	16,384	
D to H % Time	0.0	0.0	0.0	
D to D # Copies	0	0	0	
D to D # Bytes	0	0	0	
D to D % Time	0.0	0.0	0.0	

Top Device Functions By Total Time [Summary](#) | [All](#)

	Name	Launches	Device %	Total (μs)	Min (μs)	Avg (μs)	Max (μs)
1	histogram_kernel	1	0.00	6.721	6.721	6.721	6.721

CUDA Summary						
Device ID	Device Name	Contexts	Device %	Host to Device (Bytes)	Device to Host (Bytes)	
1	[0] GPU 0 - GeForce GTX 750 Ti	1	0.02 %	2,016,384	16,384	

CUDA Contexts

	Total	No Context	1	
CUDA Device ID		-	0	
▲ Runtime API Calls Summary All				
# Calls	24	0	24	
# Errors	0	0	0	
% Time	67.28	0.00	67.28	
▲ Driver API Calls Summary All				
# Calls	87	87	0	
# Errors	0	0	0	
% Time	0.29	0.29	0.00	
▲ Launches Summary All				
# Launches	2	0	2	
% Device Time	0.02	0.00	0.02	
▲ Memory Copies All				
H to D # Copies	2	0	2	
H to D # Bytes	2,016,384	0	2,016,384	
H to D % Time	0.1	0.0	0.1	
D to H # Copies	1	0	1	
D to H # Bytes	16,384	0	16,384	
D to H % Time	0.0	0.0	0.0	
D to D # Copies	0	0	0	
D to D # Bytes	0	0	0	
D to D % Time	0.0	0.0	0.0	

Top Device Functions By Total Time [Summary](#) | [All](#)

	Name	Launches	Device %	Total (μs)	Min (μs)	Avg (μs)	Max (μs)
1	histogram_kernel_shared	1	0.02	53.314	53.314	53.314	53.314
2	truncate	1	0.00	2.529	2.529	2.529	2.529



CUDA Summary

	Device ID	Device Name	Contexts	Device %	Host to Device (Bytes)	Device to Host (Bytes)
1	[0]	GPU 0 - GeForce GTX 750 Ti	1	0.02 %	2,000,000	16,384

CUDA Contexts

	Total	No Context	1	
CUDA Device ID		-	0	
Runtime API Calls Summary All				
# Calls	24	0	24	
# Errors	0	0	0	
% Time	60.48	0.00	60.48	
Driver API Calls Summary All				
# Calls	87	87	0	
# Errors	0	0	0	
% Time	0.30	0.30	0.00	
Launches Summary All				
# Launches	2	0	2	
% Device Time	0.02	0.00	0.02	
Memory Copies All				
H to D # Copies	1	0	1	
H to D # Bytes	2,000,000	0	2,000,000	
H to D % Time	0.1	0.0	0.1	
D to H # Copies	1	0	1	
D to H # Bytes	16,384	0	16,384	
D to H % Time	0.0	0.0	0.0	
D to D # Copies	0	0	0	
D to D # Bytes	0	0	0	
D to D % Time	0.0	0.0	0.0	

Top Device Functions By Total Time [Summary](#) | [All](#)

	Name	Launches	Device %	Total (μs)	Min (μs)	Avg (μs)	Max (μs)
1	histogram_kernel	1	0.02	51.587	51.587	51.587	51.587
2	convert_kernel	1	0.00	1.888	1.888	1.888	1.888