GPU Speed Of Light

SM. SM [%]	34.96	Duration [msecond]	6.13
SM. Memory [%]	15.37	Elapsed Cycles [cycle]	52,03,113.50
Executed Ipc Active [instruction]	0.20	SM Frequency [Ghz]	848.78
SOL L2 [%]	15.37	Memory Frequency [Ghz]	6.68
SOL FB [%]	1.56		

GPU Utilization

Speed Of Light [%]

Recommendations

Bottleneck [Warning] Compute is more heavily utilized than Memory: Look at 'Compute Workload Analysis' report section to see what the compute pipelines are spending their time doing. Also consider whether any computation is redundant and could be reduced.

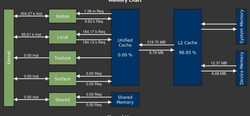
Compute Workload Analysis

Executed Ipc Elapsed [inst/cycle]	0.11	SM Busy [%]	34.98
Executed Ipc Active [inst/cycle]	0.20	Issue Slots Busy [%]	3.03
Issued Ipc Active [inst/cycle]	0.20		

Memory Workload Analysis

Memory Throughput [Gbyte/second]	3.04	Mem Busy [%]	15.17
L1 Hit Rate [%]	0.00	Max Bandwidth [%]	13.93
L2 Hit Rate [%]	96.93	Mem Pipes Busy [%]	0.98

Memory Chart



Shared Memory

	Instructions	Requests	% Peak	Bank Conflicts
Shared Load	0	0	0	0
Shared Store	0	0	0	0
Shared Atomic	0	0	0	0
Total	0	0	0	0

First-Level Cache

	Instructions	SM->TEX Requests	% Peak	Hit Rate	TEX->L2 Requests	% Peak	L2->TEX Returns	% Peak	TEX->SM Returns	% Peak
Global Load Cached	4,03,053	0	0	-	-	-	0	0	0	0
Global Load Uncached	-	13,87,764	2.77	-	-	-	1,02,93,587	6.18	-	-
Local Load Cached	42,872	1,84,168	0.37	0.00	-	-	1,84,168	0.11	28,36,284	1.70
Local Load Uncached	-	0	0	-	-	-	0	0	-	-
Surface Load	0	0	0	-	-	-	0	0	-	-
Texture Load	-	-	-	-	-	-	-	-	-	-
Global Store	1,216	9,616	0.02	-	38,400	0.02	-	-	-	-
Local Store	42,725	1,84,125	0.37	-	1,84,125	0.11	-	-	-	-
Surface Store	0	0	0	-	0	0	-	-	-	-
Global Reduction	0	0	0	-	0	0	-	-	-	-
Surface Reduction	0	0	0	-	0	0	-	-	-	-
Global Atomic	0	0	0	-	-	-	0	0	0	0
Global Atomic Cas	0	0	0	-	0	0	0	0	0	0
Surface Atomic	-	0	0	-	-	-	-	-	0	0
Surface Atomic Cas	0	0	0	-	0	0	0	0	0	0
Loads	4,45,925	15,71,932	3.13	0.00	-	-	1,04,77,553	6.29	28,36,284	1.70
Stores	43,949	1,93,741	0.39	-	2,22,525	0.13	-	-	-	-
Total	4,89,874	17,65,673	3.52	0.00	2,22,525	0.13	1,04,77,553	6.29	28,36,284	1.70

Second-Level Cache

	TEX->L2 Requests	% Peak	L2->TEX Returns	% Peak	Total Bytes	Total Throughput
Global Load Cached	-	-	0	0	0	0
Global Load Uncached	-	-	1,02,93,587	13.69	32,93,88,384	53,73,28,48,253.00
Local Load Cached	-	-	1,84,166	0.24	58,93,312	96,13,71,015.73
Local Load Uncached	-	-	0	0	0	0
Surface Load	-	-	0	0	0	0
Texture Load	-	-	-	-	-	-
Global Store	38,400	0.05	-	-	12,28,800	20,04,53,107.55
Local Store	1,84,125	-	-	-	58,92,000	96,11,56,990.28
Surface Store	0	0	-	-	0	0
Global Reduction	0	0	-	-	0	0
Surface Reduction	0	0	-	-	0	0
Global Atomic	0	0	0	0	0	0
Global Atomic Cas	0	0	0	0	0	0
Surface Atomic	-	-	0	0	0	0
Surface Atomic Cas	0	0	0	0	0	0
Loads	-	-	1,04,77,553	13.93	33,52,82,496	54,69,42,20,108.77
Stores	2,22,525	0.05	-	-	71,20,800	1,16,16,10,097.83
Total	2,22,525	0.05	1,04,77,553	13.93	34,24,02,496	55,85,58,30,366.56

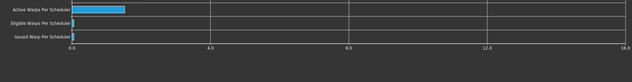
Device Memory

	L2->FB Sectors	% Peak	Bytes	Throughput
Load	4,09,376	1.06	1,29,72,032	2,11,61,16,434.65
Store	2,18,607	0.54	70,00,024	1,14,17,23,056.60
Total	6,24,243	1.54	1,99,72,056	3,25,88,40,295.25

Scheduler Statistics

Active Warps Per Scheduler [warp/cycle]	1.52	Instructions Per Active Issue Slot [inst/issue]	1.00
Eligible Warps Per Scheduler [warp/cycle]	0.05	No Eligible [%]	95.04
Issued Warp Per Scheduler [issue/cycle]	0.05	One or More Eligible [%]	5.19

Warps Per Scheduler



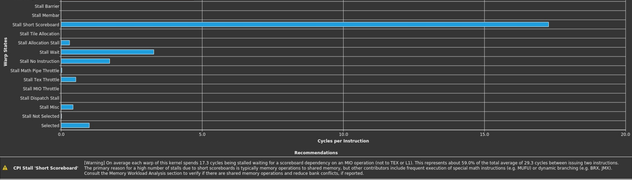
Recommendations

Issue Slot Utilization [Warning] Every scheduler is capable of issuing two instructions per cycle, but for this kernel each scheduler only issues an instruction every 19.3 cycles. This might leave hardware resources underutilized and may lead to less optimal performance. Out of the maximum of 16 warps per scheduler, this kernel allocates an average of 1.52 active warps per scheduler, but only an average of 0.05 warps were eligible per cycle. Eligible warps are the subset of active warps that are ready to issue their next instruction. Every cycle with no eligible warp results in no instruction being issued and the issue slot remains unused. To increase the number of eligible warps either increase the number of active warps or reduce the time the active warps are stalled.

Warp State Statistics

Warp Cycles Per Issued Instruction [cycle/inst]	27.38	Avg. Active Threads Per Warp [thread/inst]	25.36
Warp Cycles Per Issue Active [cycle/issue]	29.27	Avg. Not Predicated Off Threads Per Warp [thread/inst]	24.87
Warp Cycles Per Executed Instruction [cycle/inst]	27.38		

Warp State (All Cycles)



Cycles per instruction

Recommendations

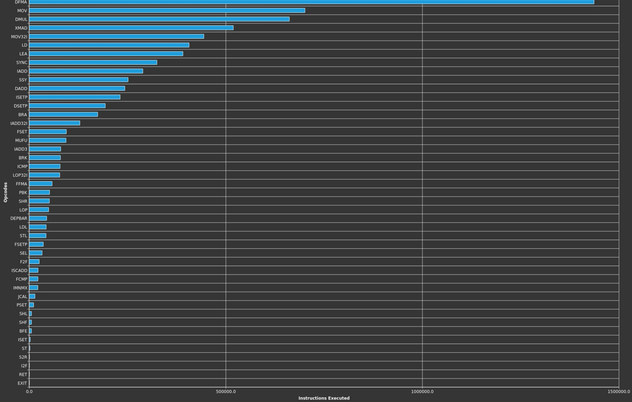
CPI Stall 'Short Scoreboard' [Warning] On average each warp of this kernel spends 17.5 cycles being stalled waiting for a scoreboard dependency on an MIO operation (not to TEX or L1). This represents about 59.0% of the total average of 29.3 cycles between issuing two instructions. The primary reason for a high number of stalls due to short scoreboards is typically memory operations to shared memory, but other contributors include frequent execution of special math instructions (e.g. MUFU) or dynamic branching (e.g. BRX, JMX). Consult the Memory Workload Analysis section to verify if there are shared memory operations and reduce bank conflicts, if reported.

CPI Stall 'Barrier' Warp stall analysis for 'Barrier' issues	Apply
CPI Stall 'Dispatch Stall' Warp stall analysis for 'Dispatch Stall' issues	Apply
CPI Stall 'Drain' Warp stall analysis for 'Drain' issues	Apply
CPI Stall 'IMC Miss' Warp stall analysis for 'Immediate constant cache (IMC)' issues	Apply
CPI Stall 'LG Throttle' Warp stall analysis for 'LG Throttle' issues	Apply
CPI Stall 'Long Scoreboard' Warp stall analysis for 'Long Scoreboard' issues	Apply
CPI Stall 'Math Pipe Throttle' Warp stall analysis for 'Math Pipe Throttle' issues	Apply
CPI Stall 'Membar' Warp stall analysis for 'Membar' issues	Apply
CPI Stall 'MIO Throttle' Warp stall analysis for 'MIO Throttle' issues	Apply
CPI Stall 'Misc' Warp stall analysis for 'Misc' issues	Apply
CPI Stall 'No Instructions' Warp stall analysis for 'No Instructions' issues	Apply
CPI Stall 'Not Selected' Warp stall analysis for 'Not Selected' issues	Apply
CPI Stall 'Sleeping' Warp stall analysis for 'Sleeping' issues	Apply
CPI Stall 'TEX Throttle' Warp stall analysis for 'TEX Throttle' issues	Apply
CPI Stall 'Wait' Warp stall analysis for 'Wait' issues	Apply
Thread Divergence Warp and thread control flow analysis	Apply

Instruction Statistics

Executed Instructions [inst]	95,02,607	Avg. Executed Instructions Per Scheduler [inst]	1,42,228.23
Issued Instructions [inst]	95,06,785	Avg. Issued Instructions Per Scheduler [inst]	1,42,293.52

Executed Instruction Mix

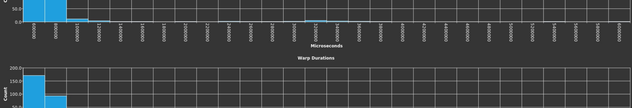


Instructions Executed

Launch Statistics

Grid Size	381	Registers Per Thread [register/thread]	255
Block Size	32	Static Shared Memory Per Block [byte/block]	0
Threads [thread]	9,632	Dynamic Shared Memory Per Block [byte/block]	0
Waves Per SM	2.36	Shared Memory Configuration Size [Kbyte]	48

Block Durations



Warp Durations

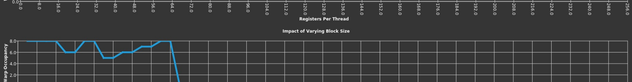
Recommendations

Launch Configuration Kernel launch configuration analysis | Apply

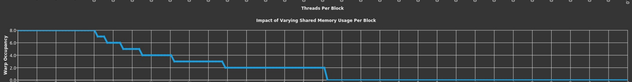
Occupancy

Theoretical Occupancy [%]	12.50	Block Limit Registers [register]	1
Theoretical Active Warps per SM [warp/cycle]	4	Block Limit Local Mem [Byte]	non
Achieved Occupancy [%]	1.78	Block Limit Warps [warp]	64
Achieved Active Warps Per SM [warp/cycle]	4.98	Block Limit SM [block]	32

Impact of Varying Register Count Per Thread



Impact of Varying Block Size



Impact of Varying Shared Memory Usage Per Block