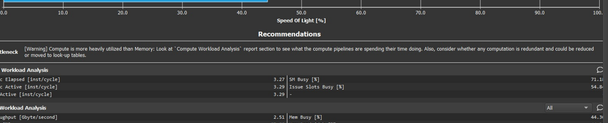
Page: Details | Process: All | Launch: 2 - 643 - cachedRegsAssignmentKernel | Add Baseline | Apply Rules | Save as PDF

Current 643 - cachedRegsAssignmentKernel (7864320, 4, 1) Time: 2.63 second Cycles: 3,701,056,221 Regs: 49 GPU: NVIDIA GeForce GTX 1060 SM Frequency: 1.41 Ghz CC: 6.1 Process: [13272] k-means.exe

GPU Speed Of Light

SOL SM [%]	71.16	Duration [second]	2.6
SOL Memory [%]	44.30	Elapsed Cycles [cycle]	3,70,10,91,221
SOL TEX [%]	6.81	SM Active Cycles [cycle]	3,70,05,03,552.50
SOL L2 [%]	3.16	SM Frequency [Ghz]	1.4
SOL FB [%]	1.35	Memory Frequency [Ghz]	7.7

GPU Utilization



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 or moved to look-up tables.

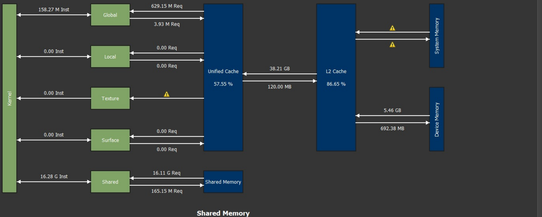
Compute Workload Analysis

Executed Ipc Elapsed [inst/cycle]	3.27	SM Busy [%]	71.1
Executed Ipc Active [inst/cycle]	3.28	Issue Slots Busy [%]	54.3
Issued Ipc Active [inst/cycle]	3.29	-	-

Memory Workload Analysis

Memory Throughput [Byte/second]	2.51	Mem Busy [%]	44.3
L1 Hit Rate [%]	57.55	Max Bandwidth [%]	44.3
L2 Hit Rate [%]	91.57	Mem Pipes Busy [%]	45.4

Memory Chart



Shared Memory

	Instructions	Requests	% Peak	Bank Conflicts
Shared Load	16,11,20,25,600	16,11,20,25,528	43.85	0
Shared Store	16,51,50,720	16,51,50,720	0.45	0
Shared Atomic	0	0	-	-
Total	16,27,71,76,320	16,27,71,76,248	44.30	0

First-Level (Unified) Cache

	Instructions	SM->TEX Requests	% Peak	Hit Rate	TEX->L2 Requests	% Peak	L2->TEX Returns	% Peak	TEX->SM Returns	% Peak
Global Load Cached	15,72,86,400	62,91,45,600	1.71	57.55	-	-	1,28,20,38,528	1.73		
Global Load Uncached		0	0		-	-	0	0		
Local Load Cached	0	0	0	0	-	-	0	0	62,91,45,600	0.85
Local Load Uncached		0	0	-	-	-	0	0		
Surface Load	0	0	0	-	-	-	0	0		
Texture Load	0									
Global Store	9,83,040	39,32,160	0.01	-	39,32,160	0.01	-	-	-	-
Local Store	0	0	0	-	0	0	-	-	-	-
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
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	15,72,86,400	62,91,45,600	1.71	57.55	-	-	1,28,20,38,528	1.73	62,91,45,600	0.85
Stores	9,83,040	39,32,160	0.01	-	39,32,160	0.01	-	-	-	-
Total	15,82,69,440	63,30,77,760	1.72	57.55	39,32,160	0.01	1,28,20,38,528	1.73	62,91,45,600	0.85

Second-Level (L2) Cache

	TEX->L2 Requests	% Peak	L2->TEX Returns	% Peak	Total Bytes	Total Throughput
Global Load Cached	-	-	1,28,20,38,528	1.48	41,00,52,32,896	15,60,17,47,119.25
Global Load Uncached	-	-	0	0	0	0
Local Load Cached	-	-	0	0	0	0
Local Load Uncached	-	-	0	0	0	0
Surface Load	-	-	0	0	0	0
Texture Load						
Global Store	39,32,160	0.01	-	-	12,58,29,120	4,78,52,357.49
Local Store	0	0	-	-	0	0
Surface Store	0	0	-	-	0	0
Global Reduction	0	0	-	-	0	0
Surface Reduction	0	0	-	-	0	0
Global Atomic						
Global Atomic Cas	0	0	0	0	0	0
Surface Atomic						
Surface Atomic Cas	0	0	0	0	0	0
Loads	-	-	1,28,20,38,528	1.48	41,00,52,32,896	15,60,17,47,119.25
Stores	39,32,160	0.01	-	-	12,58,29,120	4,78,52,357.49
Total	39,32,160	0.01	1,28,20,38,528	1.48	41,15,10,62,016	15,64,95,99,476.73

Device Memory (FB)

	L2<->FB Sectors	% Peak	Bytes	Throughput
Load	18,13,39,025	1.20	5,80,48,48,800	2,21,11,41,291.76
Store	2,26,87,786	0.15	72,60,09,152	27,60,90,644.57
Total	20,60,26,811	1.35	6,59,28,57,952	2,50,72,39,942.33

Scheduler Statistics

Active Warps Per Scheduler [warp/cycle]	4.00	Instructions Per Active Issue Slot [inst/issue]	1.37
Eligible Warps Per Scheduler [warp/cycle]	1.16	No Eligible [%]	30.8
Issued Warp Per Scheduler [issue/cycle]	0.69	One or More Eligible [%]	69.1

Warps Per Scheduler

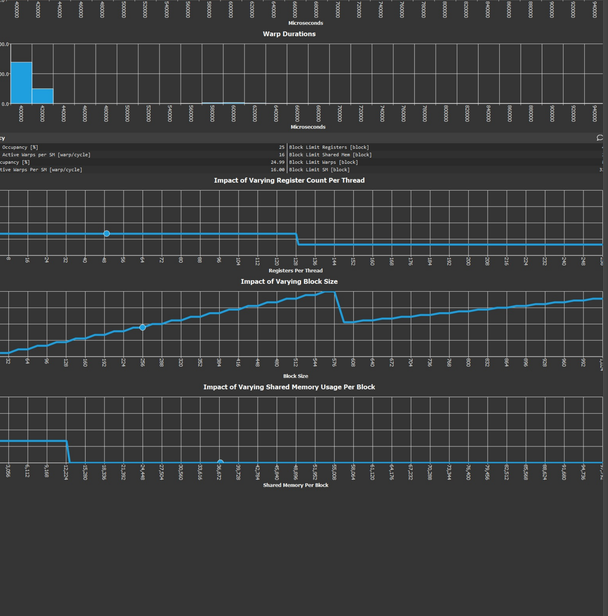
Recommendations

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 1.4 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 4.00 active warps per scheduler, but only an average of 1.16 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]	4.29	Avg. Active Threads Per Warp [thread/inst]	32.00
Warp Cycles Per Issue Active [cycle/issue]	5.16	Avg. Not Predicated Off Threads Per Warp [thread/inst]	31.81
Warp Cycles Per Executed Instruction [cycle/inst]	4.29	-	-

Warp State (All Cycles)



Recommendations

CPI Stall 'Wait' [Warning] On average each warp of this kernel spends 1.7 cycles being stalled on a fixed latency execution dependency. This represents about 32.4% of the total average of 5.1 cycles between issuing two instructions. Typically, this stall reason should be very low and only shows up as a top contributor in already highly optimized kernels. If possible, try to further increase the number of active warps to hide the corresponding instruction latencies.

Instruction Statistics

Executed Instructions [inst]	1,20,89,49,49,609	Avg. Executed Instructions Per Scheduler [inst]	3,02,24,24,202.65
Issued Instructions [inst]	1,20,89,79,34,689	Avg. Issued Instructions Per Scheduler [inst]	3,02,24,48,365.0

Executed Instruction Mix

Launch Statistics

Grid Size	1,22,880	Registers Per Thread [register/thread]	49
Block Size	256	Static Shared Memory Per Block [byte/block]	0
Threads [thread]	3,14,57,280	Dynamic Shared Memory Per Block [byte/block]	36.86
Waves Per SM	6,144	Shared Memory Configuration Size [Byte]	73.7

Block Durations

Warp Durations

Occupancy

Theoretical Occupancy [%]	25	Block Limit Registers [block]	5
Theoretical Active Warps per SM [warp/cycle]	16	Block Limit Shared Mem [block]	5
Achieved Occupancy [%]	24.99	Block Limit Warps [block]	5
Achieved Active Warps Per SM [warp/cycle]	16.00	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