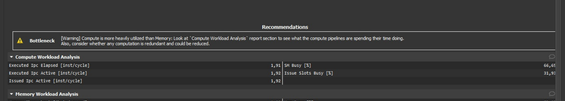
Page: Details | Launch: 6 - 117 -som_fbr$2413 | Add Baseline | Apply Rules | Save as PDF

Current 117 -som_fbr$2413 (128000, 1, 1) Time: 1,64 nsecond Cycles: 2.492.201 Regs: 47 GPU: GeForce GTX 1060 6GB SM Frequency: 1,52 Ghz CC: 6.1

GPU Speed Of Light

SOL SM [%]	66,49	Duration [second]	1,64
SOL Memory [%]	93,87	Elapsed Cycles [cycle]	2.492.201
SOL TEX [%]	36,37	SM Frequency [Ghz]	1,52
SOL L2 [%]	93,87	Memory Frequency [Ghz]	7,51
SOL FB [%]	8,60		

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.

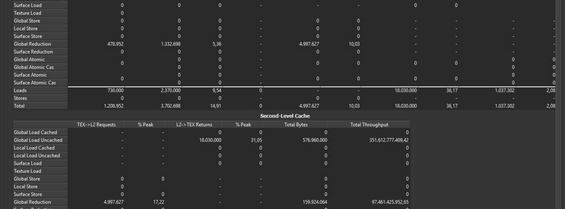
Compute Workload Analysis

Executed Ipc Elapsed [inst/cycle]	1,93	SM Busy [%]	66,49
Executed Ipc active [inst/cycle]	1,93	Issue Slots Busy [%]	31,93
Issued Ipc active [inst/cycle]	1,93		

Memory Workload Analysis

Memory Throughput [Gbyte/second]	1,42	Mem Busy [%]	93,87
L1 Hit Rate [%]	0	Max Bandwidth [%]	65,13
L2 Hit Rate [%]	99,77	Mem Pipes Busy [%]	11,81

Memory Chart



Shared Memory

	Instructions	Requests	% Peak	Bank Conflicts
Shared Load	1.441.744	1.402,016	5,65	0
Shared Store	1.050,000	970.144	1,94	0
Shared Atomic	0	0	-	0
Total	2.491,744	2.380,960	5,58	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	720,000	0	0	0	0	-	0	0	0	0
Global Load Uncached	0	2.370,000	9,54	-	-	-	18.030,000	36,17	-	-
Local Load Cached	0	0	0	0	0	0	0	0	0	0
Local Load Uncached	0	0	0	0	-	-	-	-	1.037,302	2,08
Surface Load	0	0	0	-	-	-	0	0	-	-
Texture Load	0	0	0	-	-	-	-	-	-	-
Global Store	0	0	0	-	0	0	-	-	-	-
Local Store	0	0	0	-	0	0	-	-	-	-
Surface Store	0	0	0	-	0	0	-	-	-	-
Global Reduction	478,952	1.332,688	5,36	-	4.997,627	10,03	-	-	-	-
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	-	-	0	0
Surface Atomic Cas	-	0	0	-	0	0	-	-	0	0
Loads	720,000	2.370,000	9,54	0	0	0	18.030,000	36,17	1.037,302	2,08
Stores	0	0	0	-	0	0	-	-	-	-
Total	1.200,952	3.702,688	14,91	0	4.997,627	10,03	18.030,000	36,17	1.037,302	2,08

Second-Level Cache

	TEX->L2 Requests	% Peak	L2->TEX Returns	% Peak	Total Bytes	Total Throughput
Global Load Cached	-	-	-	-	0	0
Global Load Uncached	-	-	18.030,000	31,05	576.960,000	351.612,777,426,62
Local Load Cached	-	-	0	0	0	0
Local Load Uncached	-	-	0	0	0	0
Surface Load	-	-	0	0	0	0
Texture Load	-	-	-	-	0	0
Global Store	0	0	-	-	0	0
Local Store	0	0	-	-	0	0
Surface Store	0	0	-	-	0	0
Global Reduction	4.997,627	17,22	-	-	159.924,064	97.461.425,952,65
Surface Reduction	0	0	-	-	0	0
Global Atomic	0	0	0	0	0	0
Global Atomic Cas	-	-	-	-	-	-
Surface Atomic	0	0	0	0	0	0
Surface Atomic Cas	-	-	-	-	-	-
Loads	-	-	18.030,000	31,05	576.960,000	351.612,777,426,62
Stores	0	0	-	-	0	0
Total	4.997,627	17,22	18.030,000	31,05	736.884,064	449.074,203.382,27

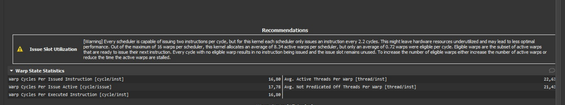
Device Memory

	L2->FB Sectors	% Peak	Bytes	Throughput
Load	50.284	0,54	1.609,088	980.615,468,62
Store	3.641	0,06	180.512	110.038.787,65
Total	53.925	0,60	1.789,600	1.090.623.659,27

Scheduler Statistics

Active Warps Per Scheduler [warp/cycle]	8,34	Instructions Per Active Issue Slot [inst/issue]	1,40
Eligible Warps Per Scheduler [warp/cycle]	0,72	No Eligible [%]	64,79
Issued Warps Per Scheduler [issue/cycle]	0,93	One or More Eligible [%]	45,42

Warps Per Scheduler



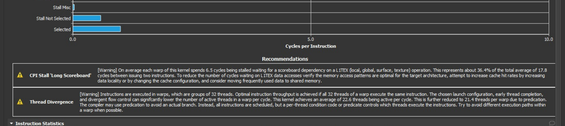
Recommendations

Issue Slot Utilization [Warning] Every scheduler is capable of issuing two instructions per cycle, but for this kernel each scheduler issues an instruction every 2,3 cycles. This might have 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 8,34 active warps per scheduler, but only an average of 0,72 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]	16,98	avg. Active Threads Per warp [thread/inst] 32,0
warp Cycles Per Issue Active [cycle/issue]	17,78	avg. Not Predicated Off Threads Per warp [thread/inst] 21,41
warp Cycles Per Executed Instruction [cycle/inst]	16,98	

Warp State (All Cycles)



Recommendations

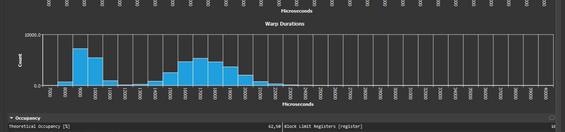
CPI Stall 'Long Scoreboard' [Warning] On average each warp of the kernel spends 5,4 cycles being stalled waiting for a scoreboard dependency on a L1TEX (local, global, surface, texture) operation. This represents about 36,4% of the total average of 17,8 cycles between issuing two instructions. To reduce the number of cycles waiting on L1TEX data accesses verify the memory access patterns are optimal for the target architecture, attempt to increase cache hit rates by increasing data locality or by changing the cache configuration, and consider moving frequently used data to shared memory.

Thread Divergence [Warning] Instructions are executed in warps, which are groups of 32 threads. Optimal instruction throughput is achieved if all 32 threads of a warp execute the same instruction. The chosen launch configuration, early thread completion, and divergent flow control can significantly lower the number of active threads in a warp per cycle. This kernel achieves an average of 22,6 threads being active per cycle. This is further reduced to 21,4 threads per warp due to predication. The compiler may use predication to avoid an actual branch. Instead, all instructions are scheduled, but a per-thread condition code or predicate controls which threads execute the instructions. Try to avoid different execution paths within a warp when possible.

Instruction Statistics

Executed Instructions [inst]	47.376,774	avg. Executed Instructions Per Scheduler [inst]	1.189.469,46
Issued Instructions [inst]	47.384,020	avg. Issued Instructions Per Scheduler [inst]	1.189.639,30

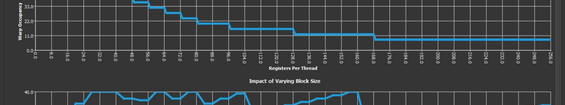
Executed Instruction Mix



Launch Statistics

Grid Size	10.000	Registers Per Thread [register/thread]	47
Block Size	128	Static Shared Memory Per Block [Kbyte/block]	0
Threads [thread]	1.280.000	Dynamic Shared Memory Per Block [Kbyte/block]	72
Waves Per SM	100	Shared Memory Configuration Size [Kbyte]	48

Block Durations



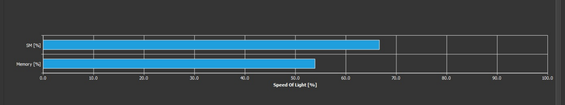
Warp Durations

Occupancy

Theoretical Occupancy [%]	62,50	Block Limit Registers [register]	10
Theoretical Active Warps Per SM [warp/cycle]	40	Block Limit Local Mem [byte]	16
Achieved Occupancy [%]	52,18	Block Limit Warps [warp]	16
Achieved Active Warps Per SM [warp/cycle]	33,40	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