	Result	Size	Time	Cycles	GPU	SM Frequency	Process	Attributes
Current	601 - I1CacheTest	(4096, 1, 1)x(256, 1, 1)	412.18 ms	78,65,25,067	0 - NVIDIA GeForce RTX 4060 Laptop GPU	1.91 Ghz	[25684] GPU-CUDA-PERFORMANCE-PROFILER.exe	

Summary | Details | Source | Context | Comments | Raw | Session

Compare | Tools | View | Export

GPU Speed Of Light Throughput

GPU Throughput Chart

High-level overview of the throughput for compute and memory resources of the GPU. For each unit, the throughput reports the achieved percentage of utilization with respect to the theoretical maximum. Breakdowns show the throughput for each individual sub-metric of Compute and Memory to clearly identify the highest contributor.

Compute (SM) Throughput [%]	73.87	Duration [ms]	412.18
Memory Throughput [%]	11.37	Elapsed Cycles [cycle]	78,65,25,067
L1/TEX Cache Throughput [%]	11.40	SM Active Cycles [cycle]	78,52,18,757.29
L2 Cache Throughput [%]	0.74	SM Frequency [Ghz]	1.91
DRAM Throughput [%]	0.28	DRAM Frequency [Ghz]	5.96

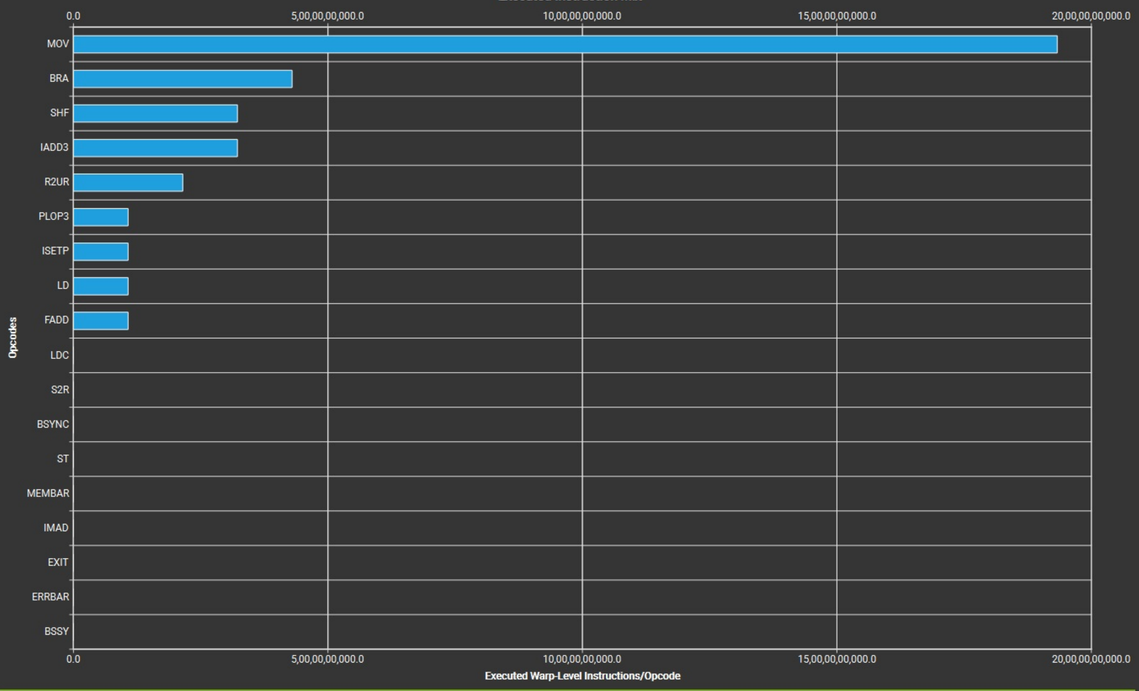
High Compute Throughput Compute is more heavily utilized than Memory: Look at the Compute Workload Analysis section to see what the compute pipelines are spending their time doing. Also, consider whether any computation is redundant and could be reduced or moved to look-up tables.

Instruction Statistics

Statistics of the executed low-level assembly instructions (SASS). The instruction mix provides insight into the types and frequency of the executed instructions. A narrow mix of instruction types implies a dependency on few instruction pipelines, while others remain unused. Using multiple pipelines allows hiding latencies and enables parallel execution. Note that 'Instructions/Opcode' and 'Executed Instructions' are measured differently and can diverge if cycles are spent in system calls.

Executed Instructions [inst]	36,51,08,00,744	Avg. Executed Instructions Per Scheduler [inst]	38,03,20,841.08
Issued Instructions [inst]	36,51,08,48,595	Avg. Issued Instructions Per Scheduler [inst]	38,03,21,339.53

Executed Instruction Mix



Launch Statistics

Summary of the configuration used to launch the kernel. The launch configuration defines the size of the kernel grid, the division of the grid into blocks, and the GPU resources needed to execute the kernel. Choosing an efficient launch configuration maximizes device utilization.

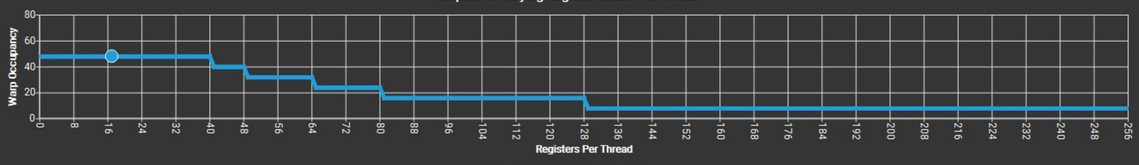
Grid Size	4,096	Function Cache Configuration	CachePreferNone
Registers Per Thread [register/thread]	17	Static Shared Memory Per Block [byte/block]	0
Block Size	256	Dynamic Shared Memory Per Block [byte/block]	0
Threads [thread]	10,48,576	Driver Shared Memory Per Block [Kbyte/block]	1.02
Waves Per SM	28.44	Shared Memory Configuration Size [Kbyte]	16.38
Uses Green Context	0	# SMs [SM]	24

Occupancy

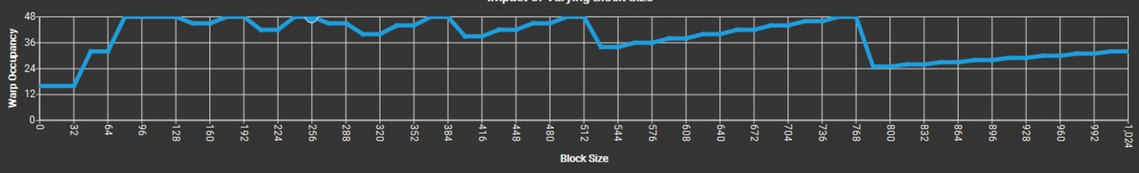
Occupancy is the ratio of the number of active warps per multiprocessor to the maximum number of possible active warps. Another way to view occupancy is the percentage of the hardware's ability to process warps that is actively in use. Higher occupancy does not always result in higher performance, however, low occupancy always reduces the ability to hide latencies, resulting in overall performance degradation. Large discrepancies between the theoretical and the achieved occupancy during execution typically indicates highly imbalanced workloads.

Theoretical Occupancy [%]	100	Block Limit Registers [block]	10
Theoretical Active Warps per SM [warp]	48	Block Limit Shared Mem [block]	16
Achieved Occupancy [%]	99.97	Block Limit Warps [block]	6
Achieved Active Warps Per SM [warp]	47.99	Block Limit SM [block]	24

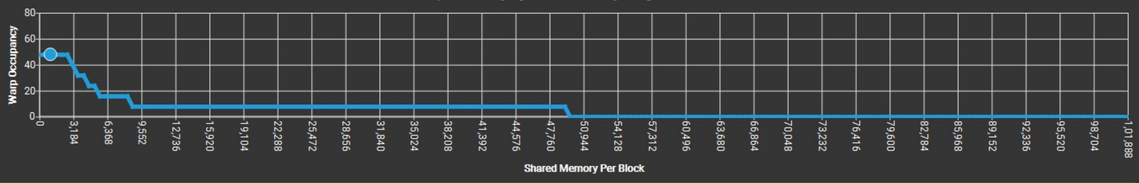
Impact of Varying Register Count Per Thread



Impact of Varying Block Size



Impact of Varying Shared Memory Usage Per Block



GPU and Memory Workload Distribution

Analysis of workload distribution in active cycles of SM, SMP, SMSP, L1 & L2 caches, and DRAM

Average SM Active Cycles [cycle]	78,52,18,757.29	Average L1 Active Cycles [cycle]	78,52,18,757.29
Average L2 Active Cycles [cycle]	14,97,55,393.44	Average SMSP Active Cycles [cycle]	78,78,81,839.85
Average DRAM Active Cycles [cycle]	68,25,920	Total SM Elapsed Cycles [cycle]	18,89,77,10,312
Total L1 Elapsed Cycles [cycle]	18,89,77,10,312	Total L2 Elapsed Cycles [cycle]	11,69,76,90,144
Total SMSP Elapsed Cycles [cycle]	75,59,08,41,248	Total DRAM Elapsed Cycles [cycle]	9,82,37,89,056

L2 Slices Workload Imbalance Est. Speedup: 5.97% One or more L2 Slices have a much higher number of active cycles than the average number of active cycles. Maximum instance value is 29.14% above the average, while the minimum instance value is 2.83% below the average.

Workload Distribution

	Average	Min	Max	Sum
SM Active Cycles	78,52,18,757.29	78,50,62,256	78,53,77,625	18,84,52,50,175
SMSP Active Cycles	78,78,81,839.85	78,77,19,193	78,81,10,693	75,63,66,56,626
L1 Active Cycles	78,52,18,757.29	78,50,62,256	78,53,77,625	18,84,52,50,175
L2 Active Cycles	14,97,55,393.44	14,55,14,710	21,13,50,213	2,39,60,86,295
DRAM Active Cycles	68,25,920	68,20,912	68,32,496	2,73,03,680

Source Counters

Source metrics, including branch efficiency and sampled warp stall reasons. Warp Stall Sampling metrics are periodically sampled over the kernel runtime. They indicate when warps were stalled and couldn't be scheduled. See the documentation for a description of all stall reasons. Only focus on stalls if the schedulers fail to issue every cycle.

Branch Instructions [inst]	4,29,52,72,896	Branch Efficiency [%]	100.00
Branch Instructions Ratio [%]	0.12	Avg. Divergent Branches	341.33

Warp Stall Sampling (All Samples)

Location	Value	Value (%)
HelloWorld.cu:11 (0xb00c60f60 in I1...	10,56,228	29
HelloWorld.cu:11 (0xb00c60f50 in I1...	1,99,019	5
HelloWorld.cu:11 (0xb00c60f40 in I1...	1,80,562	5
HelloWorld.cu:10 (0xb00c60f90 in I1...	1,09,939	3
HelloWorld.cu:10 (0xb00c60fc0 in I1...	1,03,260	3

Most Instructions Executed

Location	Value	Value (%)
HelloWorld.cu:10 (0xb00c60e40 in I1...	1,07,38,07,360	3
HelloWorld.cu:10 (0xb00c60e30 in I1...	1,07,38,07,360	3
HelloWorld.cu:10 (0xb00c60e20 in I1...	1,07,38,07,360	3
HelloWorld.cu:10 (0xb00c60e10 in I1...	1,07,38,07,360	3
HelloWorld.cu:10 (0xb00c60e00 in I1...	1,07,38,07,360	3

To customize your report even further, you might want to learn about custom sections and writing your own rules.
You might also want to consider adding individual metrics.