	Result	Size	Time	Cycles	GPU	SM Frequency	Process	Attributes
Current	589 - EdgeIteratorAlgorithmKernel	(3000000, 1, 1)x(24, 1, 1)	11,46 ms	22.007.354	0 - NVIDIA GeForce RTX 4060 Laptop GPU	1,92 Ghz	[22516] main_v2_1.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 [%]	23,86	Duration [ms]	11,46
Memory Throughput [%]	23,86	Elapsed Cycles [cycle]	22.007.354
L1/TEX Cache Throughput [%]	24,01	SM Active Cycles [cycle]	22.040.219,92
L2 Cache Throughput [%]	8,05	SM Frequency [Ghz]	1,92
DRAM Throughput [%]	18,44	DRAM Frequency [Ghz]	7,99

Latency Issue This workload exhibits low compute throughput and memory bandwidth utilization relative to the peak performance of this device. Achieved compute throughput and/or memory bandwidth below 60.0% of peak typically indicate latency issues. Look at Scheduler Statistics and Warp State Statistics for potential reasons.

Key Performance Indicators

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	3.000.000	Function Cache Configuration	CachePreferNone
Registers Per Thread [register/thread]	20	Static Shared Memory Per Block [byte/block]	0
Block Size	24	Dynamic Shared Memory Per Block [Kbyte/block]	1,54
Threads [thread]	72.000.000	Driver Shared Memory Per Block [Kbyte/block]	1,02
Waves Per SM	5.208,33	Shared Memory Configuration Size [Kbyte]	65,54
Uses Green Context	0	Stack Size	1.024
# SMs [SM]	24	# TPCs	12
Enabled TPC IDs	all	-	-

Block Size
Est. Speedup: 25.00%

Threads are executed in groups of 32 threads called warps. This kernel launch is configured to execute 24 threads per block. Consequently, some threads in a warp are masked off and those hardware resources are unused. Try changing the number of threads per block to be a multiple of 32 threads. Between 128 and 256 threads per block is a good initial range for experimentation. Use smaller thread blocks rather than one large thread block per multiprocessor if latency affects performance. This is particularly beneficial to kernels that frequently call __syncthreads(). See the Hardware Model description for more details on launch configurations.

Key Performance Indicators

Occupancy

% Occupancy Graphs

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 [%]	50	Block Limit Registers [block]	84
Theoretical Active Warps per SM [warp]	24	Block Limit Shared Mem [block]	25
Achieved Occupancy [%]	47,45	Block Limit Warps [block]	48
Achieved Active Warps Per SM [warp]	22,77	Block Limit SM [block]	24

Theoretical Occupancy
Est. Local Speedup: 50.00%

The 6.00 theoretical warps per scheduler this kernel can issue according to its occupancy are below the hardware maximum of 12. This kernel's theoretical occupancy (50.0%) is limited by the number of blocks that can fit on the SM.

Key Performance Indicators

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]	22.040.219,92	Average L1 Active Cycles [cycle]	22.040.219,92
Average L2 Active Cycles [cycle]	20.127.723,31	Average SMSP Active Cycles [cycle]	22.039.555,30
Average DRAM Active Cycles [cycle]	16.892.220	Total SM Elapsed Cycles [cycle]	532.343.856
Total L1 Elapsed Cycles [cycle]	532.343.856	Total L2 Elapsed Cycles [cycle]	319.034.000
Total SMSP Elapsed Cycles [cycle]	2.129.375.424	Total DRAM Elapsed Cycles [cycle]	366.469.120

Missing Roofline and Memory Charts? Profile again with the *detailed* metric set to collect all necessary metrics.
Also consider collecting all available sections with the *full* metric set.