	Result	Size	Time	Cycles	GPU	SM Frequency	Process	Attributes
Current	588 - EdgeIteratorAlgorithmKernel	(12, 1, 1)x(256, 1, 1)	17,47 us	33.426	0 - NVIDIA GeForce RTX 4060 Laptop GPU	1,91 Ghz	[8244] main_v1_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 [%]	1,63	Duration [us]	17,47
Memory Throughput [%]	19,27	Elapsed Cycles [cycle]	33.426
L1/TEX Cache Throughput [%]	8,02	SM Active Cycles [cycle]	13.847,75
L2 Cache Throughput [%]	7,18	SM Frequency [Ghz]	1,91
DRAM Throughput [%]	19,27	DRAM Frequency [Ghz]	7,97

Small Grid This kernel grid is too small to fill the available resources on this device, resulting in only 0.1 full waves across all SMs. Look at Launch Statistics for more details.

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	12	Function Cache Configuration	CachePreferNone
Registers Per Thread [register/thread]	18	Static Shared Memory Per Block [byte/block]	0
Block Size	256	Dynamic Shared Memory Per Block [byte/block]	0
Threads [thread]	3.072	Driver Shared Memory Per Block [Kbyte/block]	1,02
Waves Per SM	0,08	Shared Memory Configuration Size [Kbyte]	16,38
Uses Green Context	0	Stack Size	1.024
# SMs [SM]	24	# TPCs	12
Enabled TPC IDs	all	-	-

Small Grid
Est. Speedup: 50.00%
The grid for this launch is configured to execute only 12 blocks, which is less than the GPU's 24 multiprocessors. This can underutilize some multiprocessors. If you do not intend to execute this kernel concurrently with other workloads, consider reducing the block size to have at least one block per multiprocessor or increase the size of the grid to fully utilize the available hardware resources. 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 [%]	100	Block Limit Registers [block]	10
Theoretical Active Warps per SM [warp]	48	Block Limit Shared Mem [block]	16
Achieved Occupancy [%]	14,89	Block Limit Warps [block]	6
Achieved Active Warps Per SM [warp]	7,15	Block Limit SM [block]	24

Achieved Occupancy
Est. Local Speedup: 85.11%
The difference between calculated theoretical (100.0%) and measured achieved occupancy (14.9%) can be the result of warp scheduling overheads or workload imbalances during the kernel execution. Load imbalances can occur between warps within a block as well as across blocks of the same kernel. See the CUDA Best Practices Guide for more details on optimizing occupancy.

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]	13.847,75	Average L1 Active Cycles [cycle]	13.847,75
Average L2 Active Cycles [cycle]	22.901,19	Average SMSP Active Cycles [cycle]	13.224,26
Average DRAM Active Cycles [cycle]	26.832	Total SM Elapsed Cycles [cycle]	749.872
Total L1 Elapsed Cycles [cycle]	749.872	Total L2 Elapsed Cycles [cycle]	484.368
Total SMSP Elapsed Cycles [cycle]	2.999.488	Total DRAM Elapsed Cycles [cycle]	557.056

SMs Workload Imbalance
Est. Speedup: 23.19%
One or more SMs have a much lower number of active cycles than the average number of active cycles. Maximum instance value is 52.33% above the average, while the minimum instance value is 100.00% below the average.

Key Performance Indicators

SMSPs Workload Imbalance
Est. Speedup: 23.04%
One or more SMSPs have a much lower number of active cycles than the average number of active cycles. Maximum instance value is 54.44% above the average, while the minimum instance value is 100.00% below the average.

Key Performance Indicators

L1 Slices Workload Imbalance
Est. Speedup: 23.19%
One or more L1 Slices have a much lower number of active cycles than the average number of active cycles. Maximum instance value is 52.33% above the average, while the minimum instance value is 100.00% below the average.

Key Performance Indicators

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.