	Result	Size	Time	Cycles	GPU	SM Frequency	Process	Attributes
Current	563 - reduce_argmin	(4, 1, 1)x(1024, 1, 1)	13.12 us	7,631	0 - Tesla T4	581.62 Mhz	[12442] exe	

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

Compare | Tools | View | Export

GPU Speed Of Light Throughput

All

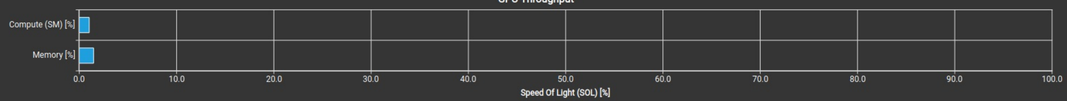
High-level overview of the throughput for compute and memory resources of the GPU. For each unit, the throughput reports the achieved percentage of utilization with respect to the theoretical maximum. Breakdowns show the throughput for each individual sub-metric of Compute and Memory to clearly identify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

Compute (SM) Throughput [%]	1.02	Duration [us]	13.12
Memory Throughput [%]	1.47	Elapsed Cycles [cycle]	7,631
L1/TEX Cache Throughput [%]	14.89	SM Active Cycles [cycle]	519.83
L2 Cache Throughput [%]	0.57	SM Frequency [Mhz]	581.62
DRAM Throughput [%]	1.47	DRAM Frequency [Ghz]	4.90

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 e Launch Statistics for more details.

Roofline Analysis The ratio of peak float (fp32) to double (fp64) performance on this device is 32:1. The kernel achieved 0% of this device's fp32 peak performance and 0% of its fp64 peak performance. See the Kernel Profiling Guide for more details on roofline analysis.

GPU Throughput



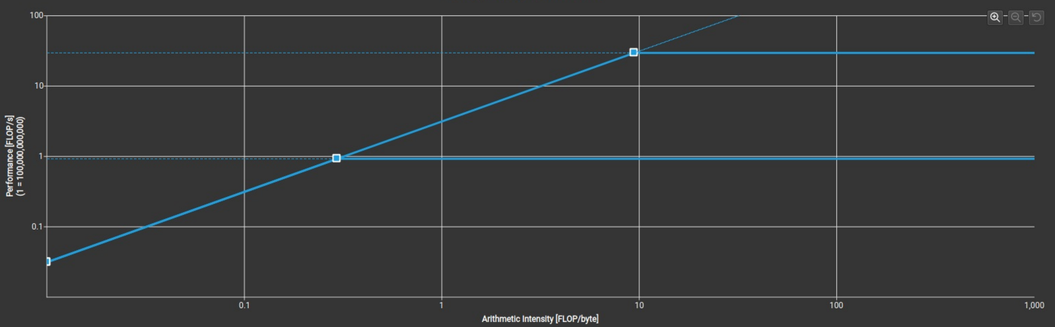
Compute Throughput Breakdown

SM: Inst Executed Pipe Lsu [%]	1.02
SM: Issue Active [%]	0.53
SM: Inst Executed [%]	0.46
SM: Mio Inst Issued [%]	0.40
SM: Pipe Alu Cycles Active [%]	0.26
SM: Mio Pq Write Cycles Active [%]	0.21
SM: Inst Executed Pipe Adu [%]	0.19
SM: Mio Pq Read Cycles Active [%]	0.18
SM: Pipe Fma Cycles Active [%]	0.15
SM: Mio2rf Writeback Active [%]	0.15
SM: Inst Executed Pipe Uniform [%]	0.11
SM: Inst Executed Pipe Cbu Pred On Any [%]	0.10
SM: Inst Executed Pipe Tex [%]	0.01
SM: Memory Throughput Internal Activity [%]	0
SM: Pipe Tensor Cycles Active [%]	0
SM: Pipe Shared Cycles Active [%]	0
SM: Pipe Fp64 Cycles Active [%]	0
IDC: Request Cycles Active [%]	0
SM: Instruction Throughput Internal Activity [%]	0
SM: Inst Executed Pipe Xu [%]	0
SM: Inst Executed Pipe Ipa [%]	0
SM: Inst Executed Pipe Fp16 [%]	0

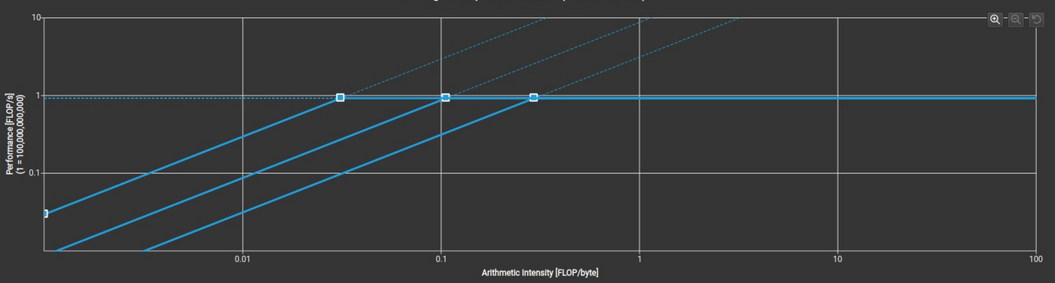
Memory Throughput Breakdown

DRAM: Cycles Active [%]	1.47
DRAM: Dram Sectors [%]	1.06
L1: Lsuin Requests [%]	1.02
L1: Data Pipe Lsu Wavefronts [%]	0.67
L2: T Sectors [%]	0.57
L1: M Xbar2l1tex Read Sectors [%]	0.51
L2: Lts2xbar Cycles Active [%]	0.51
L2: D Sectors Fill Device [%]	0.44
L1: Lsu Writeback Active [%]	0.29
L2: D Sectors [%]	0.25
GPU: Compute Memory Access Throughput Internal Activity [%]	0.21
L2: T Tag Requests [%]	0.18
L1: M L1tex2xbar Req Cycles Active [%]	0.14
L1: Data Bank Writes [%]	0.13
L2: Xbar2lts Cycles Active [%]	0.13
L1: Data Bank Reads [%]	0.11
L1: Texin Sm2tex Req Cycles Active [%]	0.06
L1: F Wavefronts [%]	0.04
L2: D Atomic Input Cycles Active [%]	0.01
L1: Tex Writeback Active [%]	0
L1: Data Pipe Tex Wavefronts [%]	0
L2: D Sectors Fill Sysmem [%]	0
GPU: Compute Memory Request Throughput Internal Activity [%]	0

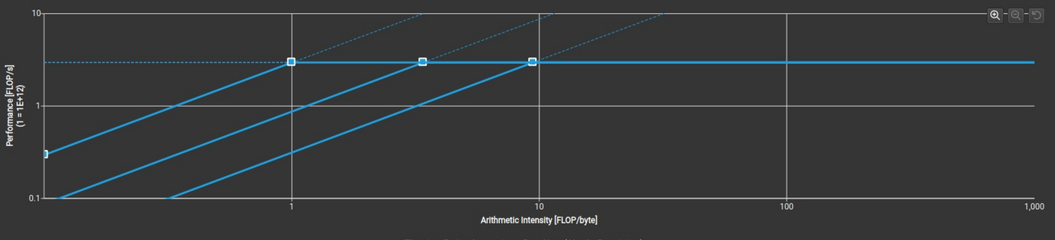
Floating Point Operations Roofline



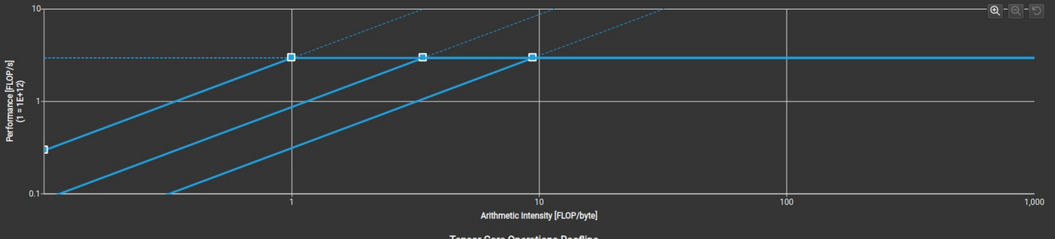
Floating Point Operations Roofline (Double Precision)



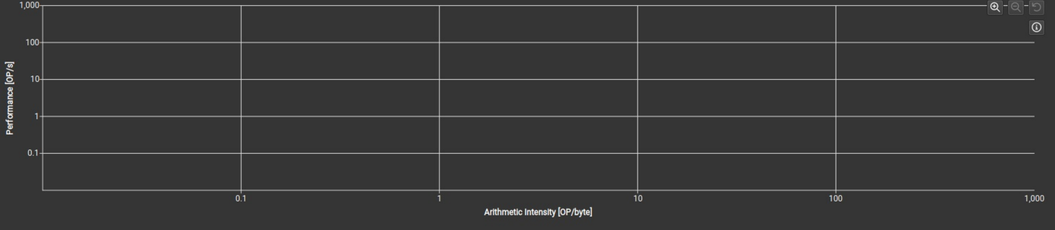
Floating Point Operations Roofline (Half Precision)



Floating Point Operations Roofline (Single Precision)



Tensor Core Operations Roofline



	# Operations	# Operations / Cycle	# Operations / s	Peak % ?eak Operations / Cycle	Peak Operations / s
Src:fp16,bf16,tf32 Dst:fp32	0	0	0	40,960	23,677.50
Src:fp16 Dst:fp16	0	0	0	40,960	23,677.50
Src:int1	0	0	0	81,920	47,355.00
Src:int4	0	0	0	81,920	47,355.00
Src:int8	0	0	0	81,920	47,355.00

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]	519.83	Average L1 Active Cycles [cycle]	519.83
Average L2 Active Cycles [cycle]	536.28	Average SMSP Active Cycles [cycle]	263.82
Average DRAM Active Cycles [cycle]	945	Total SM Elapsed Cycles [cycle]	303,368
Total L1 Elapsed Cycles [cycle]	303,368	Total L2 Elapsed Cycles [cycle]	356,928
Total SMSP Elapsed Cycles [cycle]	1,213,472	Total DRAM Elapsed Cycles [cycle]	514,048

SMs Workload Imbalance
Est. Speedup: 6.30%
One or more SMs have a much lower number of active cycles than the average number of active cycles. Maximum instance value is 91.91% above the average, while the minimum instance value is 100.00% below the average.

L1 Slices Workload Imbalance
Est. Speedup: 6.30%
One or more L1 Slices have a much lower number of active cycles than the average number of active cycles. Maximum instance value is 91.91% above the average, while the minimum instance value is 100.00% below the average.

Workload Distribution

	Average	Min	Max	Sum
SM Active Cycles	519.83	0	6,427	20,793
SMSP Active Cycles	263.82	0	6,409	42,212
L1 Active Cycles	519.83	0	6,427	20,793
L2 Active Cycles	536.28	398	1,319	17,161
DRAM Active Cycles	945	864	1,064	7,560