Page: Default | Driver: | Kernel: | Grid: | Block: | Stream: | Device: | Context: | Process: | Kernel Filter: | Exclude API Calls: | Range: | Replay Mode: | Metrics: | Section: | Baseline: | Add Baseline | Apply Baseline | Clear Baselines | Manage Baselines | Save as New | Save as Default

	Name	Arch	Time	Cycles	Regs	SPs	FP Instructions	IMC Frequency	CC	Process
Current	sram_32x3	490 - device_compute(int [32], int, int[32], int, int)	1.12 µs	1,120	24	32	16	1000 MHz GPU clock(s)	7.5	[0]d1h02 cmwbalson
Total in Session	sram_32x3	490 - device_compute(int [32], int, int[32], int, int)	1.12 µs	1,120	24	32	16	1000 MHz GPU clock(s)	7.5	[0]d1h02 cmwbalson
Total in Device	sram_32x3	490 - device_compute(int [32], int, int[32], int, int)	1.12 µs	1,120	24	32	16	1000 MHz GPU clock(s)	7.5	[0]d1h02 cmwbalson
Total/Device	sram_32x3	490 - device_compute(int [32], int, int[32], int, int)	1.12 µs	1,120	24	32	16	1000 MHz GPU clock(s)	7.5	[0]d1h02 cmwbalson

GPU Speed Of Light Throughput

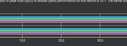
This section provides a high-level overview of the utilization for compute and memory resources of the GPU. For each unit, the throughput reports the achieved percentage of utilization with respect to the theoretical maximum. Breakdowns show the throughput for each individual sub-metric of Compute and Memory to identify the busiest units. Utilization of other units that are not shown in the breakdowns, might need some of the throughput reported for the busiest unit, causing it to be the limiter of kernel performance.

Compute (SM) Throughput [%]	0.08
Memory Throughput [%]	0.08
L1/TEX Cache Throughput [%]	0.08
L2 Cache Throughput [%]	0.07
DRAM Throughput [%]	0.07

Latency bound This kernel 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, warp state and memory statistics for potential reasons.

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 Roofline chart for more details on roofline analysis.

GPU Throughput



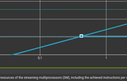
Compute Throughput Breakdown

Int Unit Executed Pipe cycles [%]	0.08
SM Issue Active [%]	0.07
Int Unit Executed Pipe [%]	0.07
SM Cycle Issued [%]	0.07
SM Pipe Fma Cycles Active [%]	0.06
SM Pipe Fmaheavy Cycles Active [%]	0.06
SM Pipe Fmaheavy Cycles Active [%]	0.06
SM Int Executed Pipe Mio [%]	0.05
SM Int Executed Pipe Mio [%]	0.05
SM Int Executed Pipe Lsu [%]	0.04
SM Int Executed Pipe Lsu [%]	0.04
SM Mio Pq Read Cycles Active [%]	0.04
SM Mio Pq Write Cycles Active [%]	0.04
SM Int Executed Pipe Uniform [%]	0.04
SM Int Executed Pipe Uniform [%]	0.04
SM Pipe Shared Cycles Active [%]	0
SM Pipe Tex0 Cycles Active [%]	0
SM Pipe Tex1 Cycles Active [%]	0
SM Int Executed Pipe Tex [%]	0
SM Int Executed Pipe Tex [%]	0
SM Int Executed Pipe Tex [%]	0

Memory Throughput Breakdown

L1/Tex Hit Latency [cycle]	31.06
L1/Tex Miss Latency [cycle]	31.06
L2 Hit Latency [cycle]	31.06
L2 Miss Latency [cycle]	31.06
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0
L1/Tex Sectors Miss Pipe Lsu Mem Global Op Ld [%]	0

Floating Point Operations Roofline



Compute Workload Analysis

Detailed analysis of the compute resources of the streaming multiprocessors (SM), including the achieved instructions per clock (IPC) and the utilization of each available pipeline. Pipelines with very high utilization might be the limiting factor for kernel performance.

Executed Ipc Elapsed [inst/cycle]	1.24	(-97.0%, +41.2%)
Executed Ipc Active [inst/cycle]	1.24	(-97.0%, +41.2%)
Issued Ipc Active [inst/cycle]	1.24	(-97.0%, +41.2%)
Warps Per Scheduler [warp]	0.02	(-99.9%, +0.0%)
Eligible Warps Per Scheduler [warp]	0.01	(-99.9%, +0.0%)
No Eligible [%]	99.39	(+99.3%, +41.1%)
One Warp Eligible [%]	0.61	(-99.3%, -41.1%)
Two Warps Eligible [%]	0.00	(-100.0%, -100.0%)
Three Warps Eligible [%]	0.00	(-100.0%, -100.0%)
Four Warps Eligible [%]	0.00	(-100.0%, -100.0%)
Single Issue [%]	0.61	(-99.3%, -41.1%)
Dual Issue [%]	0.00	(-100.0%, -100.0%)
Issue Slots Busy [%]	31.06	(-97.0%, +41.2%)
Issued Instructions Per Scheduler	0.31	(-97.0%, +41.2%)
Issue Slot Utilization [%]	31.06	(-97.0%, +41.2%)
Issued Warp Per Scheduler	0.31	(-97.0%, +41.2%)
No Issue [%]	68.94	(+97.0%, -41.2%)
Average Active Threads Per Warp	32	
Average Thread Rewinds Per Warp	0.00	

Recommendation 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 Roofline chart for more details on roofline analysis.

Pipe Utilization

Memory Workload Analysis

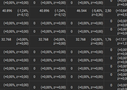
Detailed analysis of the memory resources of the GPU. Memory can become a limiting factor for the overall kernel performance when fully utilizing the involved hardware units (Mem Busy), exhausting the available communication bandwidth between those units (Max Bandwidth), or by reaching the maximum throughput of issuing memory instructions (Mem Pipes Busy). Detailed chart of the memory units. Detailed tables with data for each memory unit.

Memory Throughput [Gbyte/second]	27.40	(-73.63%, +0.00%)
Mem Busy [%]	0.08	(-99.9%, +0.0%)
Max Bandwidth [%]	31.86	(-73.63%, +0.0%)
Mem Pipes Busy [%]	0.08	(-99.9%, +0.0%)

L2 Store Access Pattern The memory access pattern for stores from L2 to DRAM is 32 bytes per transaction. The granularity of L2 to DRAM transfers is also 32 bytes. That is a compression ratio of one between sectors per L2 request. However, this kernel only accesses an average of 1.0 sectors out of the possible 4 sectors per cache line. Check the Source Counters section for uncoalesced stores and try to minimize how many cache lines need to be accessed per memory request.

L2 Load Access Pattern The memory access pattern for loads from L2 to L1TEX is 32 bytes per transaction. The granularity of L2 to L1TEX transfers is also 32 bytes. That is a compression ratio of one between sectors per L2 request. However, this kernel only accesses an average of 1.0 sectors out of the possible 4 sectors per cache line. Check the Source Counters section for uncoalesced loads and try to minimize how many cache lines need to be accessed per memory request.

Memory Chart



Shared Memory

	Instructions	Requests	Wavefronts	L1 Peak	Bank Conflicts
Shared Load	294,912 (+0.0%, +0.0%)	294,912 (+0.0%, +0.0%)	294,912 (+0.0%, +0.0%)	2.00 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)
Shared Load Matrix	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)
Shared Store	41,304 (-12.0%, -12.0%)	41,304 (-12.0%, -12.0%)	41,314 (-12.0%, -12.0%)	1.25 (-12.0%, -12.0%)	0 (+0.0%, +0.0%)
Shared Atomic	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)
Other			79,436 (-1.04%, +0.0%)	0.54 (-1.04%, +0.0%)	0 (+0.0%, +0.0%)
Total	336,221 (-2.14%, +0.0%)	336,221 (-2.14%, +0.0%)	411,662 (-2.14%, +0.0%)	2.24 (-0.0%, +0.0%)	0 (+0.0%, +0.0%)

L1TEX Cache

	Instructions	Requests	Wavefronts	L1 Peak	Sectors	Sectors/Req	Hit Rate	Bytes	Sector Misses to L2	% Peak to L2	Returns to Sm	L1 Peak
Local Load	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	
Global Load	49,896 (-2.52%, +0.0%)	49,896 (-2.52%, +0.0%)	49,904 (-2.52%, +0.0%)	46.584 (-3.64%, -2.50%)	147,504 (-3.64%, +0.0%)	2.96 (-1.12%, +0.0%)	0.00 (-100.0%, +0.0%)	4,913.408 (-3.64%, +0.0%)	105,904 (-13.62%, +0.0%)	5.68 (-10.48%, +0.0%)	47,274 (-0.0%, -2.5%)	3.54 (-0.0%, +0.0%)
Surface Load	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	
Texture Load	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	
Global Store	32,768 (+0.0%, +0.0%)	32,768 (+0.0%, +0.0%)	32,768 (+0.0%, +0.0%)	32.768 (+0.0%, +0.0%)	1,048,576 (+0.0%, +0.0%)	32.00 (+0.0%, +0.0%)	100.00% (+0.0%, +0.0%)	33,554,432 (+0.0%, +0.0%)	32,768 (+0.0%, +0.0%)	1.70% (+0.0%, +0.0%)	1,572,274 (+0.0%, +0.0%)	
Local Store	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)			
Surface Store	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	
Global Reduction	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	
Surface Reduction	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	
Global Atomic	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)			see above
Global Atomic CAS	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)			
Surface Atomic ALU	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)			
Surface Atomic CAS	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)			
Stores	32,768 (+0.0%, +0.0%)	32,768 (+0.0%, +0.0%)	32,768 (+0.0%, +0.0%)	32.768 (+0.0%, +0.0%)	1,048,576 (+0.0%, +0.0%)	32.00 (+0.0%, +0.0%)	100.00% (+0.0%, +0.0%)	33,554,432 (+0.0%, +0.0%)	32,768 (+0.0%, +0.0%)	1.70% (+0.0%, +0.0%)	1,572,274 (+0.0%, +0.0%)	
Atomics & Reductions	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	
Total	79,664 (-0.41%, +0.0%)	79,664 (-0.41%, +0.0%)	79,672 (-0.41%, +0.0%)	79.312 (-0.45%, -0.0%)	1,196,080 (-3.14%, +0.0%)	15.01 (-2.72%, +0.0%)	41.38% (-58.62%, +0.0%)	5,961.984 (-3.62%, +0.0%)	138,152 (-13.42%, +0.0%)	7.44 (-10.48%, +0.0%)	47,274 (-0.0%, -2.5%)	3.54 (-0.0%, +0.0%)

L2 Cache

	Requests	Sectors	Sectors/Req	L1 Peak	Hit Rate	Bytes	Throughput	Sector Misses to Device	Sector Misses to System	Sector Misses to Peer
L1TEX Load	155,456 (-11.64%, +0.0%)	155,904 (-11.62%, +0.0%)	1.00 (+0.0%, +0.0%)	1.24 (-0.0%, +0.0%)	0.00 (-100.0%, +0.0%)	5,399,936 (-11.62%, +0.0%)	40.0% (-59.0%, +0.0%)	32,768 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)
L1TEX Store	32,768 (+0.0%, +0.0%)	32,768 (+0.0%, +0.0%)	1.00 (+0.0%, +0.0%)	1.02 (-0.0%, +0.0%)	0.00 (-100.0%, +0.0%)	1,048,576 (+0.0%, +0.0%)	15.1% (-0.0%, +0.0%)	32,768 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)
L1TEX Atomic ALU	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)
L1TEX Atomic CAS	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)
L1TEX Reduction	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)
L1TEX Total	188,159 (-3.93%, +0.0%)	188,672 (-3.92%, +0.0%)	1.00 (+0.0%, +0.0%)	0.79 (-0.0%, +0.0%)	0.00 (-100.0%, +0.0%)	6,450,176 (-3.92%, +0.0%)	19.2% (-0.0%, +0.0%)	64,160,434.134 % (-0.0%, +0.0%)	32,768 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)
EDS Total							19.2% (-0.0%, +0.0%)		32,768 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)
CPS Total	155,542 (-3.92%, +0.0%)	155,904 (-3.92%, +0.0%)	1.00 (+0.0%, +0.0%)	1.24 (-0.0%, +0.0%)	0.00 (-100.0%, +0.0%)	4,989,184 (-3.92%, +0.0%)	44.57% (-0.0%, +0.0%)	32,768 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)	0 (+0.0%, +0.0%)

Device Memory

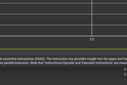
	Requests	Sectors	Bytes	Throughput
Load	32,773 (+0.0%, +0.0%)	1,049,056 (+0.0%, +0.0%)	1,049,056 (+0.0%, +0.0%)	15.107 GTexel/s (+0.0%, +0.0%)
Store	64 (-99.0%, +41.2%)	2,048 (-99.0%, +41.2%)	65,600 (-99.0%, +41.2%)	29.697 GB/s (-99.0%, +41.2%)
Total	32,837 (-0.0%, +0.0%)	1,051,104 (-0.0%, +0.0%)	1,114,656 (-0.0%, +0.0%)	15.804 GTexel/s (-0.0%, +0.0%)

Scheduler Statistics

Summary of the activity of the schedulers issuing instructions. Each SM contains multiple schedulers which issue instructions to the various functional units. The upper bound of warps in the pool (Theoretical Warps) is limited by the launch configuration. On every cycle each scheduler checks the state of the allocated warps in the pool (Active Warps). Active warps that are not stalled (Eligible Warps) are ready to issue their next instruction. From the set of eligible warps the scheduler selects a single warp from which to issue one or more instructions (Issued Warp). On cycles with no eligible warps, the issue slot is skipped and no instruction is issued. Having many eligible warps allows hiding latency between issues.

Active Warps Per Scheduler [warp]	0.02	(-99.9%, +0.0%)
Eligible Warps Per Scheduler [warp]	0.01	(-99.9%, +0.0%)
Issued Warps Per Scheduler	0.31	(-97.0%, +41.2%)

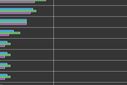
Issue Slot Utilization Every scheduler is capable of issuing one instruction per cycle, but for this kernel each scheduler only issues an instruction every 3.2 cycles. This might have hardware resources underutilized and may lead to less optimal performance. Out of the maximum of 8 schedulers, 0.31 schedulers issue an instruction every cycle. Compute Workload section will show what kind of instructions are issued. Look at Scheduler Statistics and Warp State Statistics for potential reasons.



Warp State Statistics

Analysis of the states in which all warps spent cycles during the kernel execution. The warp states describe a warp's readiness or inability to issue its next instruction. The warp cycles on the instruction define the latency between two consecutive instructions. The higher the value, the more warp parallelism is required to hide this latency. Stalled warps not eligible to issue cycles do not contribute to performance. In most cases the goal is to decrease the number of samples in the non-issue states and increase the number of issue cycles, where stalled warps do not impact the overall performance but are they completely unusable. Long focus on stall reasons if the kernel has a low issue rate or if the issue cycles are very infrequent.

Warp Cycles Per Issued Instruction	31.27	(-19.9%, +1.2%)
Warp Cycles Per Executed Instruction	31.27	(-19.9%, +1.2%)
Avg. Active Threads Per Warp	31.34	(-0.0%, +0.0%)
Avg. Not Predicated Off Threads Per Warp	30.85	(-0.0%, +0.0%)

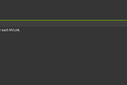


Instruction Statistics

Statistics of the executed low level assembly instructions (SASS). The instruction mix provides insight into the type 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 change if cycles are spent in system calls.

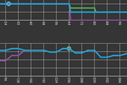
Executed Instructions [inst]	19,201.31	(-0.0%, +0.0%)
Issued Instructions [inst]	2,304.589	(-0.0%, +0.0%)
Avg. Instructions Executed Per Scheduler [inst]	19,238.24	(-0.0%, +0.0%)
Avg. Issued Instructions Per Scheduler [inst]	2,311.27	(-0.0%, +0.0%)

Executed Instruction Mix



WGMA Topology

WGMA Topology diagram shows logical WGMA connections with transmit/receive Bandwidth.



The systems does not have any WGMA connections.

WGMA Tables

Detailed tables with properties for each WGMA.

Logical WGMA Properties

The systems does not have any WGMA connections.

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.

Block Size	256	(-99.9%, +0.0%)
Grid Size	1,024	(-99.9%, +0.0%)
Threads [thread]	262,144	(-99.9%, +0.0%)
Waves Per SM	0.31	(-99.9%, +0.0%)
Function Cache Configuration	cudaFuncCachePreferNone	

Registers Per Thread [register/thread]	24	(-0.0%, +0.0%)
Shared Memory Configuration Size [Kbyte]	164.00	(-0.0%, +0.0%)
Dynamic Shared Memory Per Block [Byte/block]	0	(-100.0%, +0.0%)
Static Shared Memory Per Block [Byte/block]	0	(-100.0%, +0.0%)
Driver Memory Configuration Size [Kbyte]	32.77	(-0.0%, +0.0%)

Occupancy

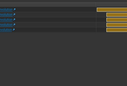
Occupancy is the ratio of the number of active warps per multiprocessor to the maximum number of possible active warps. Another way to view occupancy is the percentage of the hardware's ability to process warps that is actively in use. Higher occupancy does not always result in higher performance, but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

Theoretical Occupancy [%]	50	(-50.0%, +0.0%)
Achieved Occupancy [%]	16.13	(-67.74%, +0.0%)
Achieved Active Warps Per SM [warp]	516.17	(-67.74%, +0.0%)
Achieved Active Warps Per SM [warp]	27.62	(-100.0%, +0.0%)

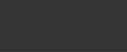
Occupancy Limiters This kernel's theoretical occupancy is not impacted by any block limit. The difference between calculated theoretical (50.0%) and achieved theoretical (50.0%) may be due to the rounding of block resources. Check the Occupancy section for more details on optimizing occupancy.

Block Limit SM	16	
Block Limit Registers [block]	10	(-37.5%, +0.0%)
Block Limit Shared Mem [block]	10	(-37.5%, +0.0%)
Block Limit Warps [block]	5	(-68.75%, +0.0%)
Block Limit Sm [block]	16	(-0.0%, +0.0%)

Impact of Varying Block Size



Impact of Varying Shared Memory Usage Per Block



Source Counters

Source counters provide a variety of low level performance indicators. Warp Stall Sampling metrics are periodically sampled over the kernel runtime. They indicate when warps were stalled and could not be scheduled. See the documentation for a description of all stall reasons. Only focus on those reasons that matter for your kernel.

Branch Instructions [inst]	141,601	(-75.0%, +2.5%)
Branch Instructions Ratio [%]	0.04	(-0.0%, +0.0%)
Branch Efficiency [%]	99.73	(+0.0%, +0.0%)
Avg. This Warp Exited Loop [inst]	1,024.00	(-0.0%, +0.0%)

Uncoalesced Global Accesses This kernel has uncoalesced global accesses resulting in a total of 1048064 excessive sectors (50% of the total 1,1758 sectors). Check the L2 Theoretical Sectors Global Excessive table for the primary source locations. The CUDA Programming Guide gives further details about optimizing memory accesses.

L2 Theoretical Sectors Global Excessive

Location	Value	Value [%]
L1_TEX(1,TA,0) in device_compute(int *	524,288	50.00
L1_TEX(1,TA,1) in device_compute(int *	524,288	50.00

Warp Stall Sampling (All Cycles)

Location	Value	Value [%]
L1_TEX(1,TA,1) in device_compute(int *	1,024	100.00
L1_TEX(1,TA,0) in device_compute(int *	0	0.00

Warp Stall Sampling (Not-Issued Cycles)

Location	Value	Value [%]
L1_TEX(1,TA,1) in device_compute(int *	1,024	100.00
L1_TEX(1,TA,0) in device_compute(int *	0	0.00

Most Instructions Executed

Location	Value	Value [%]
L1_TEX(1,TA,1) in device_compute(int *	19,456	50.00
L1_TEX(1,TA,0) in device_compute(int *	19,456	50.00
L1_TEX(1,TA,1) in device_compute(int *	19,456	50.00
L1_TEX(1,TA,0) in device_compute(int *	19,456	50.00