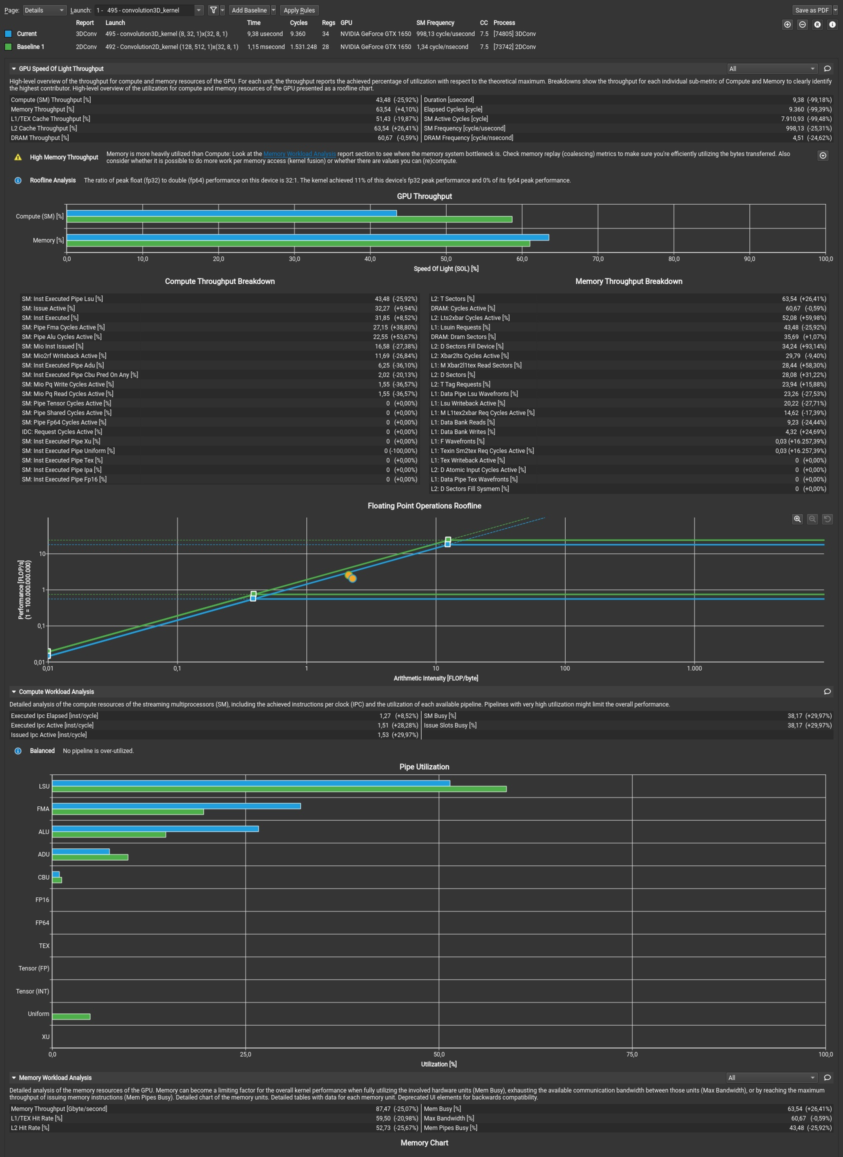


	Report	Launch	Time	Cycles	Regs	GPU	SM Frequency	CC	Process
🟦 **Current**	3DConv	495 - convolution3D_kernel (8, 32, 1)x(32, 8, 1)	9,38 usecond	9.360	34	NVIDIA GeForce GTX 1650	998,13 cycle/usecond	7.5	[74805] 3DConv
🟩 **Baseline 1**	2DConv	492 - Convolution2D_kernel (128, 512, 1)x(32, 8, 1)	1,15 msecond	1.531.248	28	NVIDIA GeForce GTX 1650	1,34 cycle/nsecond	7.5	[73742] 2DConv

GPU Speed Of Light Throughput

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 [%]	43,48 (-25,92%)	Duration [usecond]	9,38 (-99,18%)
Memory Throughput [%]	63,54 (+4,10%)	Elapsed Cycles [cycle]	9.360 (-99,39%)
L1/TEX Cache Throughput [%]	51,43 (-19,87%)	SM Active Cycles [cycle]	7.910,93 (-99,48%)
L2 Cache Throughput [%]	63,54 (+26,41%)	SM Frequency [cycle/usecond]	998,13 (-25,31%)
DRAM Throughput [%]	60,67 (-0,59%)	DRAM Frequency [cycle/nsecond]	4,51 (-24,62%)

⚠ **High Memory Throughput** Memory is more heavily utilized than Compute: Look at the Memory Workload Analysis report section to see where the memory system bottleneck is. Check memory replay (coalescing) metrics to make sure you're efficiently utilizing the bytes transferred. Also consider whether it is possible to do more work per memory access (kernel fusion) or whether there are values you can (re)compute.

ⓘ **Roofline Analysis** The ratio of peak float (fp32) to double (fp64) performance on this device is 32:1. The kernel achieved 11% of this device's fp32 peak performance and 0% of its fp64 peak performance.

Compute Throughput Breakdown

SM: Inst Executed Pipe Lsu [%]	43,48 (-25,92%)
SM: Issue Active [%]	32,27 (+9,94%)
SM: Inst Executed [%]	31,85 (+8,52%)
SM: Pipe Fma Cycles Active [%]	27,15 (+38,80%)
SM: Pipe Alu Cycles Active [%]	22,55 (+53,67%)
SM: Mio Inst Issued [%]	16,58 (-27,38%)
SM: Mio2rf Writeback Active [%]	11,69 (-26,84%)
SM: Inst Executed Pipe Adu [%]	6,25 (-36,10%)
SM: Inst Executed Pipe Cbu Pred On Any [%]	2,02 (-20,13%)
SM: Mio Pq Write Cycles Active [%]	1,55 (-36,57%)
SM: Mio Pq Read Cycles Active [%]	1,55 (-36,57%)
SM: Pipe Tensor Cycles Active [%]	0 (+0,00%)
SM: Pipe Shared Cycles Active [%]	0 (+0,00%)
SM: Pipe Fp64 Cycles Active [%]	0 (+0,00%)
IDC: Request Cycles Active [%]	0 (+0,00%)
SM: Inst Executed Pipe Xu [%]	0 (+0,00%)
SM: Inst Executed Pipe Uniform [%]	0 (-100,00%)
SM: Inst Executed Pipe Tex [%]	0 (+0,00%)
SM: Inst Executed Pipe Ipa [%]	0 (+0,00%)
SM: Inst Executed Pipe Fp16 [%]	0 (+0,00%)

Memory Throughput Breakdown

L2: T Sectors [%]	63,54 (+26,41%)
DRAM: Cycles Active [%]	60,67 (-0,59%)
L2: Lts2xbar Cycles Active [%]	52,08 (+59,98%)
L1: Lsuin Requests [%]	43,48 (-25,92%)
DRAM: Dram Sectors [%]	35,69 (+1,07%)
L2: D Sectors Fill Device [%]	34,24 (+93,14%)
L2: Xbar2lts Cycles Active [%]	29,79 (-9,40%)
L1: M Xbar2l1tex Read Sectors [%]	28,44 (+58,30%)
L2: D Sectors [%]	28,08 (+31,22%)
L2: T Tag Requests [%]	23,94 (+15,88%)
L1: Data Pipe Lsu Wavefronts [%]	23,26 (-27,53%)
L1: Lsu Writeback Active [%]	20,22 (-27,71%)
L1: M L1tex2xbar Req Cycles Active [%]	14,62 (-17,39%)
L1: Data Bank Reads [%]	9,23 (-24,44%)
L1: Data Bank Writes [%]	4,32 (+24,69%)
L1: F Wavefronts [%]	0,03 (+16.257,39%)
L1: Texin Sm2tex Req Cycles Active [%]	0,03 (+16.257,39%)
L1: Tex Writeback Active [%]	0 (+0,00%)
L2: D Atomic Input Cycles Active [%]	0 (+0,00%)
L1: Data Pipe Tex Wavefronts [%]	0 (+0,00%)
L2: D Sectors Fill Sysmem [%]	0 (+0,00%)

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 limit the overall performance.

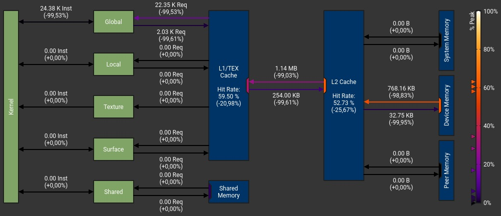
Executed Ipc Elapsed [inst/cycle]	1,27 (+8,52%)	SM Busy [%]	38,17 (+29,97%)
Executed Ipc Active [inst/cycle]	1,51 (+28,28%)	Issue Slots Busy [%]	38,17 (+29,97%)
Issued Ipc Active [inst/cycle]	1,53 (+29,97%)		

ⓘ **Balanced** No pipeline is over-utilized.

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. Deprecated UI elements for backwards compatibility.

Memory Throughput [Gbyte/second]	87,47 (-25,07%)	Mem Busy [%]	63,54 (+26,41%)
L1/TEX Hit Rate [%]	59,50 (-20,98%)	Max Bandwidth [%]	60,67 (-0,59%)
L2 Hit Rate [%]	52,73 (-25,67%)	Mem Pipes Busy [%]	43,48 (-25,92%)



Shared Memory

	Instructions	Requests	Wavefronts	% Peak	Bank Conflicts
Shared Load	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
Shared Load Matrix	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
Shared Store	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
Shared Atomic	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
Other	-	-	4.096 (-99,61%)	3,13 (-36,10%)	0 (+0,00%)
Total	0 (+0,00%)	0 (+0,00%)	4.096 (-99,61%)	3,13 (-36,10%)	0 (+0,00%)

L1/TEX Cache

	Instructions	Requests	Wavefronts	% Peak	Sectors	Sectors/Req	Hit Rate	Bytes	Sector Misses to L2	% Peak to L2	Returns to SM	% Peak to SM
Local Load	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	37.279 (-99,03%)	28,45 (+58,35%)	22.400 (-99,55%)	17,10 (-25,89%)
Global Load	22.352 (-99,53%)	22.352 (-99,53%)	22.352 (-99,55%)	17,06 (-25,58%)	103.530 (-99,53%)	4,63 (-0,64%)	64,14 (-22,24%)	3.312.960 (-99,53%)				
Surface Load	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
Texture Load	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)		
Global Store	2.032 (-99,61%)	2.032 (-99,61%)	2.032 (-99,61%)	1,55 (-36,57%)	8.128 (-99,61%)	4 (+0,00%)	0 (+0,00%)	260.096 (-99,61%)	8.128 (-99,61%)	6,20 (-36,57%)	-	-
Local Store	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)			-	-
Surface Store	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	-	-
Global Reduction	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	-	-
Surface Reduction	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	-	-
Global Atomic ALU	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	see above	see above
Global Atomic CAS	0 (+0,00%)											
Surface Atomic ALU	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	see above	see above
Surface Atomic CAS	0 (+0,00%)											
Loads	22.352 (-99,53%)	22.352 (-99,53%)	22.352 (-99,55%)	17,06 (-25,58%)	103.530 (-99,53%)	4,63 (-0,64%)	64,14 (-22,24%)	3.312.960 (-99,53%)	37.279 (-99,03%)	28,45 (+58,35%)	22.400 (-99,55%)	17,10 (-25,89%)
Stores	2.032 (-99,61%)	2.032 (-99,61%)	2.032 (-99,61%)	1,55 (-36,57%)	8.128 (-99,61%)	4 (+0,00%)	0 (+0,00%)	260.096 (-99,61%)	8.128 (-99,61%)	6,20 (-36,57%)	-	-
Total	24.384 (-99,53%)	24.384 (-99,53%)	24.384 (-99,55%)	18,61 (-26,64%)	111.658 (-99,54%)	4,58 (-0,35%)	59,47 (-21,03%)	3.573.056 (-99,54%)	45.407 (-99,24%)	34,66 (+24,89%)	22.400 (-99,55%)	17,10 (-25,89%)

L2 Cache

	Requests	Sectors	Sectors/Req	% Peak	Hit Rate	Bytes	Throughput	Sector Misses to Device	Sector Misses to System	Sector Misses to Peer
L1/TEX Load	15.056 (-99,22%)	37.228 (-99,03%)	2,47 (+23,30%)	51,85 (+59,29%)	38,94 (-22,60%)	1.191.296 (-99,03%)	127.058.020.477,82 (+18,09%)	24.576 (-98,83%)	0 (+0,00%)	0 (+0,00%)
L1/TEX Store	2.032 (-99,61%)	8.128 (-99,61%)	4 (+0,00%)	11,32 (-36,10%)	100 (+0,00%)	260.096 (-99,61%)	27.740.614.334,47 (-52,63%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
L1/TEX Atomic ALU	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
L1/TEX Atomic CAS	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
L1/TEX Reduction	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
L1/TEX Total	17.145 (-99,30%)	45.364 (-99,24%)	2,65 (+8,75%)	63,18 (+25,71%)	49,17 (-26,40%)	1.451.648 (-99,24%)	154.825.938.566,55 (-6,81%)	24.576 (-98,83%)	0 (+0,00%)	0 (+0,00%)
GPU Total	17.270 (-99,29%)	45.619 (-99,23%)	2,64 (+8,59%)	63,54 (+26,41%)	49,35 (-26,11%)	1.459.808 (-99,23%)	155.696.245.733,79 (-6,29%)	24.581 (-98,83%)	0 (+0,00%)	0 (+0,00%)

Device Memory

	Sectors	% Peak	Bytes	Throughput
Load	24.581 (-98,83%)	58,19 (+89,96%)	786.592 (-98,83%)	83.894.197.952,22 (+43,19%)
Store	1.048 (-99,95%)	2,48 (-91,84%)	33.536 (-99,95%)	3.576.791.808,87 (-93,85%)
Total	25.629 (-99,39%)	60,67 (-0,59%)	820.128 (-99,39%)	87.470.989.761,09 (-25,07%)

L1/TEX Cache (Deprecated)

	Instructions	L1/TEX Requests	% Peak	Hit Rate	L2 Requests	% Peak	L2 Returns	% Peak	L1/TEX Returns	% Peak
Global Load Cached	22.352 (-99,53%)	22.352 (-99,53%)	17,06 (-22,48%)	64,14 (-22,24%)	-	-	37.279 (-99,03%)	28,45 (+58,35%)	22.400 (-99,55%)	17,10 (-25,89%)
Global Load Uncached	-	-	-	-	-	-				
Local Load Cached	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	-	-				
Local Load Uncached	-	-	-	-	-	-				
Surface Load	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	-	-	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
Texture Load	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	-	-	0 (+0,00%)	0 (+0,00%)		
Global Store	2.032 (-99,61%)	2.032 (-99,61%)	1,55 (-36,57%)	0 (+0,00%)	8.128 (-99,61%)	6,20 (-36,57%)	-	-	-	-
Local Store	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)			-	-	-	-
Surface Store	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	-	-	-	-
Global Reduction	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	-	-	-	-
Surface Reduction	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	-	-	-	-
Global Atomic	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	see above	see above
Global Atomic Cas	0 (+0,00%)									
Surface Atomic	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	see above	see above
Surface Atomic Cas	0 (+0,00%)									
Loads	22.352 (-99,53%)	22.352 (-99,53%)	17,06 (-22,48%)	64,14 (-22,24%)	-	-	37.279 (-99,03%)	28,45 (+58,35%)	22.400 (-99,55%)	17,10 (-25,89%)
Stores	2.032 (-99,61%)	2.032 (-99,61%)	1,55 (-36,57%)	0 (+0,00%)	8.128 (-99,61%)	6,20 (-36,57%)	-	-	-	-
Total	24.384 (-99,53%)	24.384 (-99,53%)	18,61 (-23,89%)	59,47 (-21,03%)	8.128 (-99,61%)	6,20 (-36,57%)	37.279 (-99,03%)	28,45 (+58,35%)	22.400 (-99,55%)	17,10 (-25,89%)

L2 Cache (Deprecated)

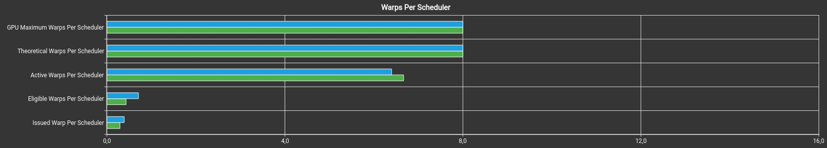
	L2 Requests	% Peak	L2 Returns	% Peak	Total Bytes	Total Throughput
Global Load Cached	-	-	37.279 (-99,03%)	28,45 (+58,35%)	1.192.928 (-99,03%)	127.232.081.911,26 (+18,26%)
Global Load Uncached	-	-			-	-
Local Load Cached	-	-			-	-
Local Load Uncached	-	-			-	-
Surface Load	-	-	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
Texture Load	-	-	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
Global Store	8.128 (-99,61%)	6,20 (-36,57%)	-	-	260.096 (-99,61%)	27.740.614.334,47 (-52,63%)
Local Store			-	-		
Surface Store	0 (+0,00%)	0 (+0,00%)	-	-	0 (+0,00%)	0 (+0,00%)
Global Reduction	0 (+0,00%)	0 (+0,00%)	-	-	0 (+0,00%)	0 (+0,00%)
Surface Reduction	0 (+0,00%)	0 (+0,00%)	-	-	0 (+0,00%)	0 (+0,00%)
Global Atomic	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
Global Atomic Cas						
Surface Atomic	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)	0 (+0,00%)
Surface Atomic Cas						
Loads	-	-	37.279 (-99,03%)	28,45 (+58,35%)	1.192.928 (-99,03%)	127.232.081.911,26 (+18,26%)
Stores	8.128 (-99,61%)	6,20 (-36,57%)	-	-	260.096 (-99,61%)	27.740.614.334,47 (-52,63%)
Total	8.128 (-99,61%)	6,20 (-36,57%)	37.279 (-99,03%)	28,45 (+58,35%)	1.453.024 (-99,24%)	154.972.696.245,73 (-6,73%)

Scheduler Statistics

Summary of the activity of the schedulers issuing instructions. Each scheduler maintains a pool of warps that it can issue instructions for. 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 skipped issue slots indicates poor latency hiding.

Active Warps Per Scheduler [warp]	6,40 (-4,04%)	No Eligible [%]	61,20 (-13,38%)
Eligible Warps Per Scheduler [warp]	0,71 (+64,54%)	One or More Eligible [%]	38,80 (+32,22%)
Issued Warp Per Scheduler	0,39 (+32,22%)		

⚠ **Issue Slot Utilization** Every scheduler is capable of issuing one instruction per cycle, but for this kernel each scheduler only issues an instruction every 2.6 cycles. This might leave hardware resources underutilized and may lead to less optimal performance. Out of the maximum of 8 warps per scheduler, this kernel allocates an average of 6.40 active warps per scheduler, but only an average of 0.71 warps were eligible per cycle. Eligible warps are the subset of active warps that are ready to issue their next instruction. Every cycle with no eligible warp results in no instruction being issued and the issue slot remains unused. To increase the number of eligible warps, reduce the time the active warps are stalled by inspecting the top stall reasons on the Warp State Statistics and Source Counters sections.



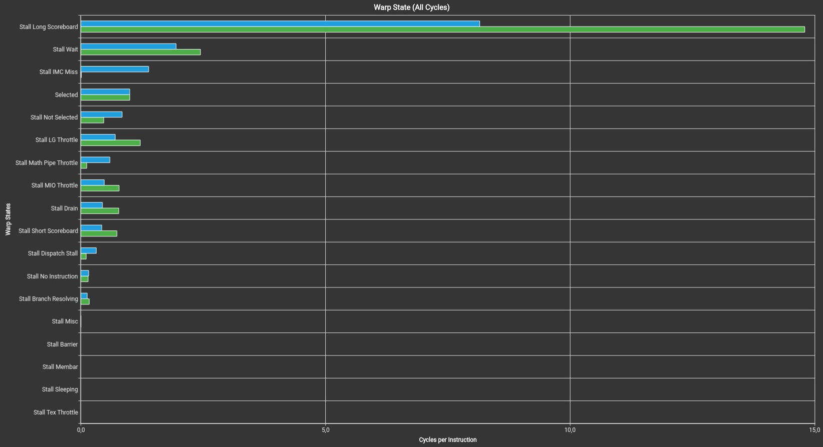
Warp State Statistics

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 per instruction define the latency between two consecutive instructions. The higher the value, the more warp parallelism is required to hide this latency. For each warp state, the chart shows the average number of cycles spent in that state per issued instruction. Stalls are not always impacting the overall performance nor are they completely avoidable. Only focus on stall reasons if the schedulers fail to issue every cycle. When executing a kernel with mixed library and user code, these metrics show the combined values.

Warp Cycles Per Issued Instruction [cycle]	16,49 (-27,42%)	Avg. Active Threads Per Warp	31,80 (-0,58%)
Warp Cycles Per Executed Instruction [cycle]	16,71 (-26,47%)	Avg. Not Predicated Off Threads Per Warp	31,42 (+0,30%)

⚠ **long_scoreboard** On average, each warp of this kernel spends 8.2 cycles being stalled waiting for a scoreboard dependency on a L1TEX (local, global, surface, texture) operation. This represents about 49.4% of the total average of 16.5 cycles between issuing two instructions. To reduce the number of cycles waiting on L1TEX data accesses verify the memory access patterns are optimal for the target architecture, attempt to increase cache hit rates by increasing data locality or by changing the cache configuration, and consider moving frequently used data to shared memory.

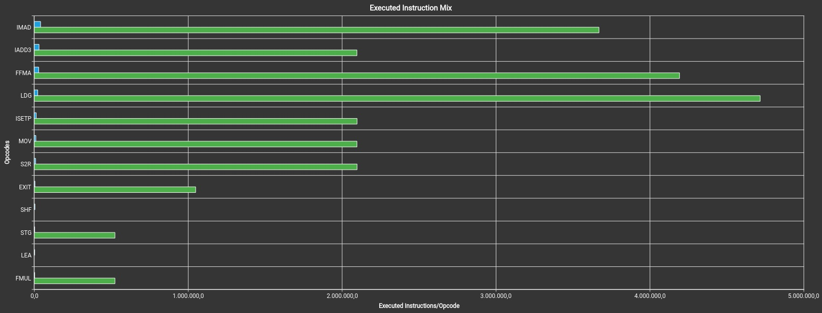
ⓘ **Warp Stall** Check the Source Counters section for the top stall locations in your source based on sampling data.



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]	166.896 (-99,34%)	Avg. Executed Instructions Per Scheduler [inst]	2.980,29 (-99,34%)
Issued Instructions [inst]	169.108 (-99,33%)	Avg. Issued Instructions Per Scheduler [inst]	3.019,79 (-99,33%)



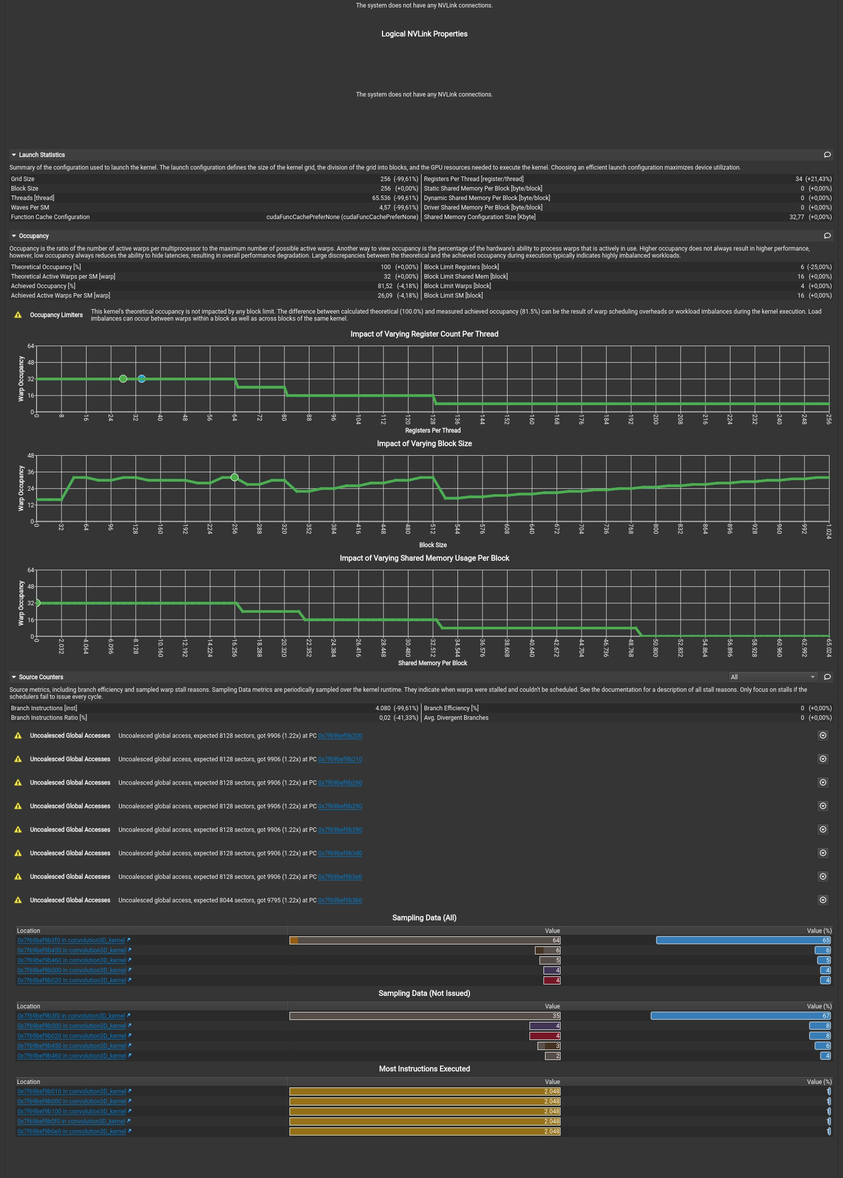
NVLink

All

High-level summary of NVLink utilization. It shows the total received and transmitted (sent) memory, as well as the overall link peak utilization. NVLink Topology. Detailed tables with properties for each NVLink.

Physical Links	0 (+0,00%)	Logical Links	0 (+0,00%)

NVLink Topology



The system does not have any NVLink connections.

Logical NVLink Properties

The system does not have any NVLink 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.

Grid Size	256 (-99,61%)	Registers Per Thread [register/thread]	34 (+21,43%)
Block Size	256 (+0,00%)	Static Shared Memory Per Block [byte/block]	0 (+0,00%)
Threads [thread]	65.536 (-99,61%)	Dynamic Shared Memory Per Block [byte/block]	0 (+0,00%)
Waves Per SM	4,57 (-99,61%)	Driver Shared Memory Per Block [byte/block]	0 (+0,00%)
Function Cache Configuration	cudaFuncCachePreferNone (cudaFuncCachePreferNone)	Shared Memory Configuration Size [Kbyte]	32,77 (+0,00%)

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 (+0,00%)	Block Limit Registers [block]	6 (-25,00%)
Theoretical Active Warps per SM [warp]	32 (+0,00%)	Block Limit Shared Mem [block]	16 (+0,00%)
Achieved Occupancy [%]	81,52 (-4,18%)	Block Limit Warps [block]	4 (+0,00%)
Achieved Active Warps Per SM [warp]	26,09 (-4,18%)	Block Limit SM [block]	16 (+0,00%)

⚠ **Occupancy Limiters** This kernel's theoretical occupancy is not impacted by any block limit. The difference between calculated theoretical (100.0%) and measured achieved occupancy (81.5%) 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.

Impact of Varying Register Count Per Thread

Impact of Varying Block Size

Impact of Varying Shared Memory Usage Per Block

Source Counters

Source metrics, including branch efficiency and sampled warp stall reasons. Sampling Data 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]	4.080 (-99,61%)	Branch Efficiency [%]	0 (+0,00%)
Branch Instructions Ratio [%]	0,02 (-41,33%)	Avg. Divergent Branches	0 (+0,00%)

⚠ **Uncoalesced Global Accesses** Uncoalesced global access, expected 8128 sectors, got 9906 (1.22x) at PC 0x7f69bef9b200

⚠ **Uncoalesced Global Accesses** Uncoalesced global access, expected 8128 sectors, got 9906 (1.22x) at PC 0x7f69bef9b210

⚠ **Uncoalesced Global Accesses** Uncoalesced global access, expected 8128 sectors, got 9906 (1.22x) at PC 0x7f69bef9b260

⚠ **Uncoalesced Global Accesses** Uncoalesced global access, expected 8128 sectors, got 9906 (1.22x) at PC 0x7f69bef9b290

⚠ **Uncoalesced Global Accesses** Uncoalesced global access, expected 8128 sectors, got 9906 (1.22x) at PC 0x7f69bef9b390

⚠ **Uncoalesced Global Accesses** Uncoalesced global access, expected 8128 sectors, got 9906 (1.22x) at PC 0x7f69bef9b3d0

⚠ **Uncoalesced Global Accesses** Uncoalesced global access, expected 8128 sectors, got 9906 (1.22x) at PC 0x7f69bef9b3e0

⚠ **Uncoalesced Global Accesses** Uncoalesced global access, expected 8044 sectors, got 9795 (1.22x) at PC 0x7f69bef9b3b0

Sampling Data (All)

Location	Value	Value (%)
0x7f69bef9b3f0 in convolution3D_kernel ↗	64	65
0x7f69bef9b450 in convolution3D_kernel ↗	6	6
0x7f69bef9b460 in convolution3D_kernel ↗	5	5
0x7f69bef9b000 in convolution3D_kernel ↗	4	4
0x7f69bef9b020 in convolution3D_kernel ↗	4	4

Sampling Data (Not Issued)

Location	Value	Value (%)
0x7f69bef9b3f0 in convolution3D_kernel ↗	35	67
0x7f69bef9b000 in convolution3D_kernel ↗	4	8
0x7f69bef9b020 in convolution3D_kernel ↗	4	8
0x7f69bef9b450 in convolution3D_kernel ↗	3	6
0x7f69bef9b460 in convolution3D_kernel ↗	2	4

Most Instructions Executed

Location	Value	Value (%)
0x7f69bef9b010 in convolution3D_kernel ↗	2.048	1
0x7f69bef9b000 in convolution3D_kernel ↗	2.048	1
0x7f69bef9b100 in convolution3D_kernel ↗	2.048	1
0x7f69bef9b0f0 in convolution3D_kernel ↗	2.048	1
0x7f69bef9b0e0 in convolution3D_kernel ↗	2.048	1