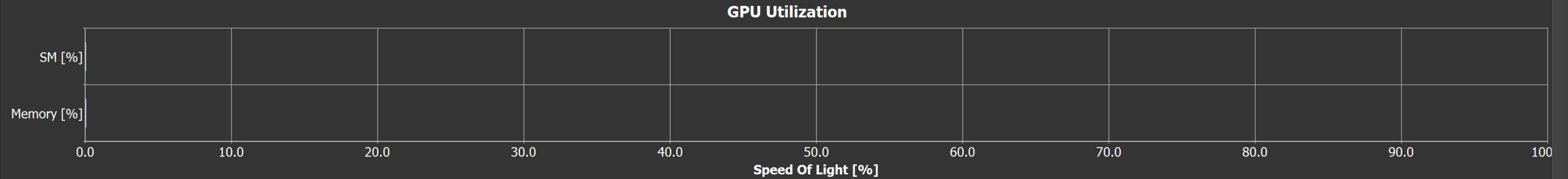


GPU Speed Of LightAll

High-level overview of the utilization for compute and memory resources of the GPU. For each unit, the Speed Of Light (SOL) reports the achieved percentage of utilization with respect to the theoretical maximum.

SOL SM [%]	0.04	Duration [usecond]	47.39
SOL Memory [%]	0.06	Elapsed Cycles [cycle]	49,758
SOL TEX [%]	1.92	SM Active Cycles [cycle]	608,448
SOL L2 [%]	0.06	SM Frequency [cycle/nsecond]	1.05
SOL FB [%]	0.03	Memory Frequency [cycle/usecond]	744.0



SOL SM Breakdown		SOL Memory Breakdown	
SOL SM: Pipe Alu Cycles Active [%]	0.04	SOL L2: Xbar2lts Cycles Active [%]	0.06
SOL SM: Issue Active [%]	0.04	SOL L2: Lts2xbar Cycles Active [%]	0.05
SOL SM: Inst Executed [%]	0.04	SOL GPU: Dram Throughput [%]	0.03
SOL SM: Inst Executed Pipe Cbu Pred On Any [%]	0.03	SOL L1: Data Pipe Lsu Wavefronts [%]	0.02
SOL SM: Inst Executed Pipe Lsu [%]	0.02	SOL L1: Lsuin Requests [%]	0.02
SOL SM: Pipe Fma Cycles Active [%]	0.01	SOL L1: M L1tex2xbar Req Cycles Active [%]	0.02
SOL SM: Mio Inst Issued [%]	0.01	SOL L2: T Sectors [%]	0.01
SOL SM: Mio2rf Writeback Active [%]	0.00	SOL L2: T Tag Requests [%]	0.01
SOL SM: Mio Pq Write Cycles Active [%]	0.00	SOL L1: Lsu Writeback Active [%]	0.01
SOL SM: Mio Pq Read Cycles Active [%]	0.00	SOL L2: D Sectors [%]	0.01
SOL SM: Inst Executed Pipe Xu [%]	0.00	SOL L1: Texin Sm2tex Req Cycles Active [%]	0.01
SOL SM: Inst Executed Pipe Tex [%]	0.00	SOL L1: F Wavefronts [%]	0.01
SOL IDC: Request Cycles Active [%]	0	SOL L1: M Xbar2l1tex Read Sectors [%]	0.00
SOL SM: Inst Executed Pipe Adu [%]	0	SOL L1: Data Bank Writes [%]	0.00
SOL SM: Inst Executed Pipe Fp16 [%]	0	SOL L1: Data Bank Reads [%]	0.00
SOL SM: Inst Executed Pipe Ipa [%]	0	SOL L2: D Atomic Input Cycles Active [%]	0.00
SOL SM: Pipe Fp64 Cycles Active [%]	0	SOL L2: D Sectors Fill Device [%]	0.00
SOL SM: Pipe Shared Cycles Active [%]	0	SOL L2: D Sectors Fill Sysmem [%]	0.00
SOL SM: Pipe Tensor Cycles Active [%]	0	SOL L1: Data Pipe Tex Wavefronts [%]	0
		SOL L1: Tex Writeback Active [%]	0

Recommendations

Bottleneck

High-level bottleneck detection

Apply

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	1	Registers Per Thread [register/thread]
Block Size	1	Static Shared Memory Per Block [byte/block]
Threads [thread]	1	Dynamic Shared Memory Per Block [byte/block]
Waves Per SM	0.00	Shared Memory Configuration Size [byte]

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 [%]	50	Block Limit Registers [block]
Theoretical Active Warps per SM [warp/cycle]	32	Block Limit Shared Mem [block]
Achieved Occupancy [%]	1.56	Block Limit Warps [block]
Achieved Active Warps Per SM [warp]	1	Block Limit SM [block]