

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.

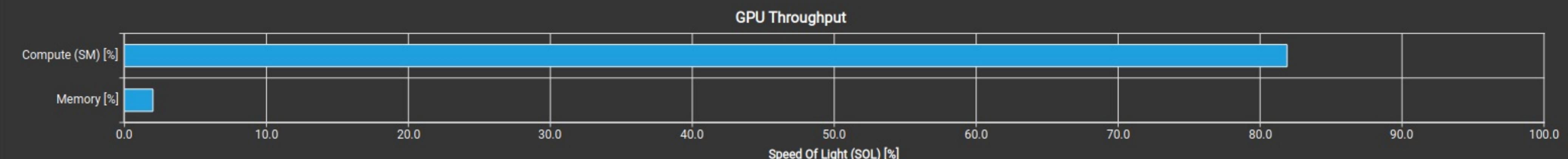
Compute (SM) Throughput [%]	81.90	Duration [ms]	2.16
Memory Throughput [%]	2.01	Elapsed Cycles [cycle]	1,264,962
L1/TEX Cache Throughput [%]	1.56	SM Active Cycles [cycle]	1,257,861.20
L2 Cache Throughput [%]	0.67	SM Frequency [Mhz]	584.99
DRAM Throughput [%]	2.01	DRAM Frequency [Ghz]	4.99

High Throughput

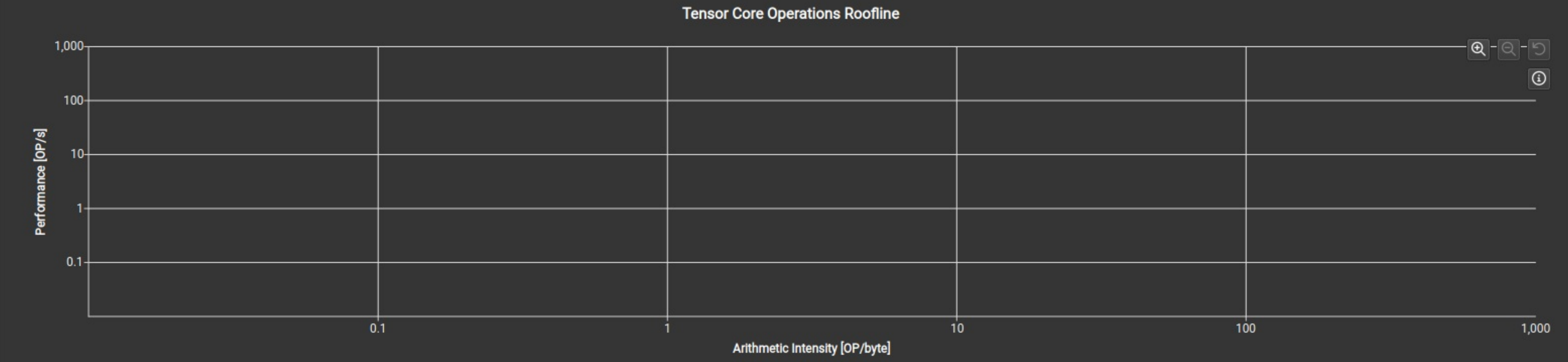
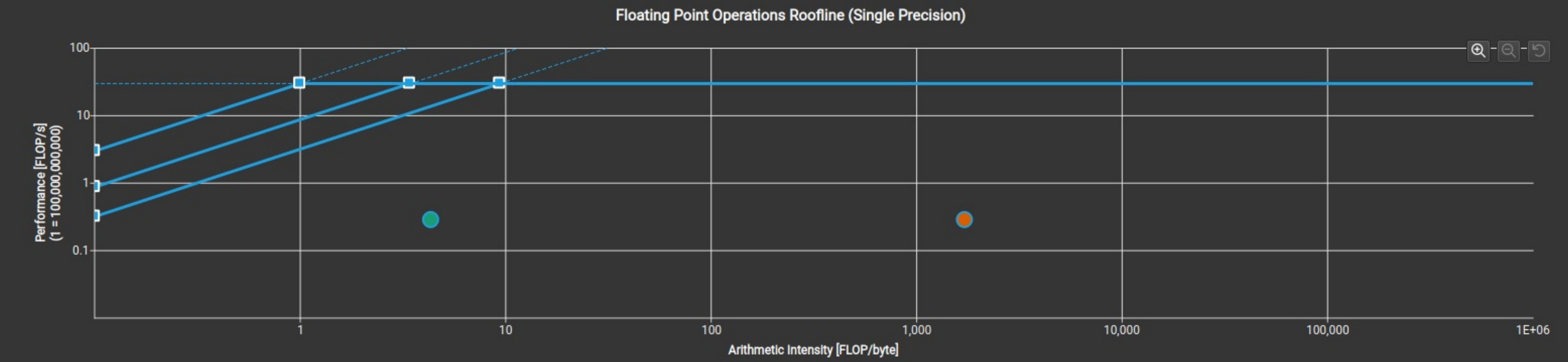
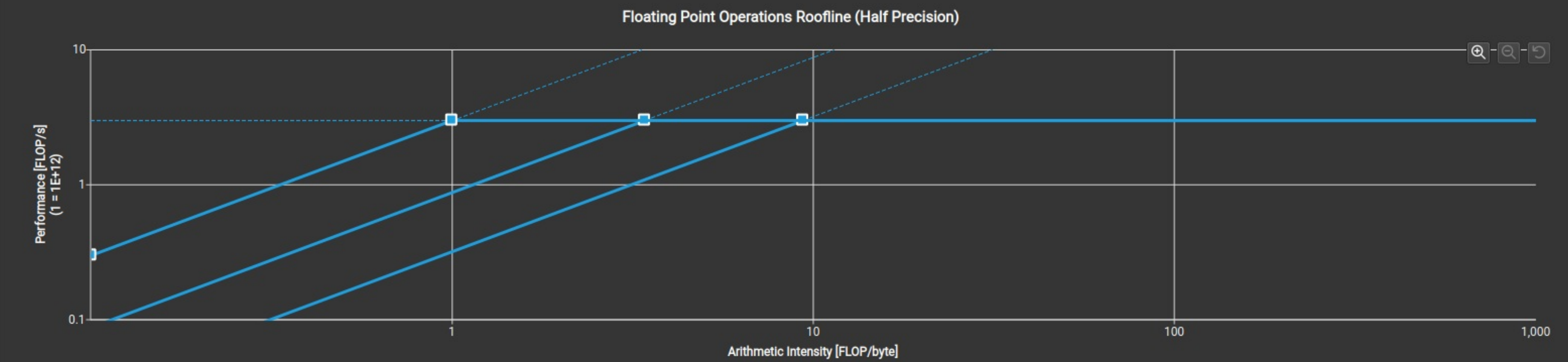
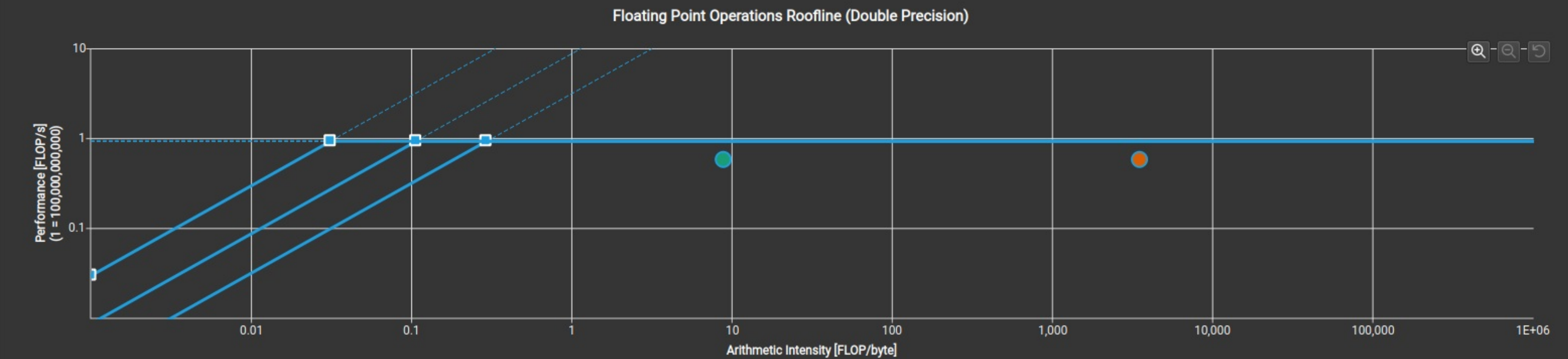
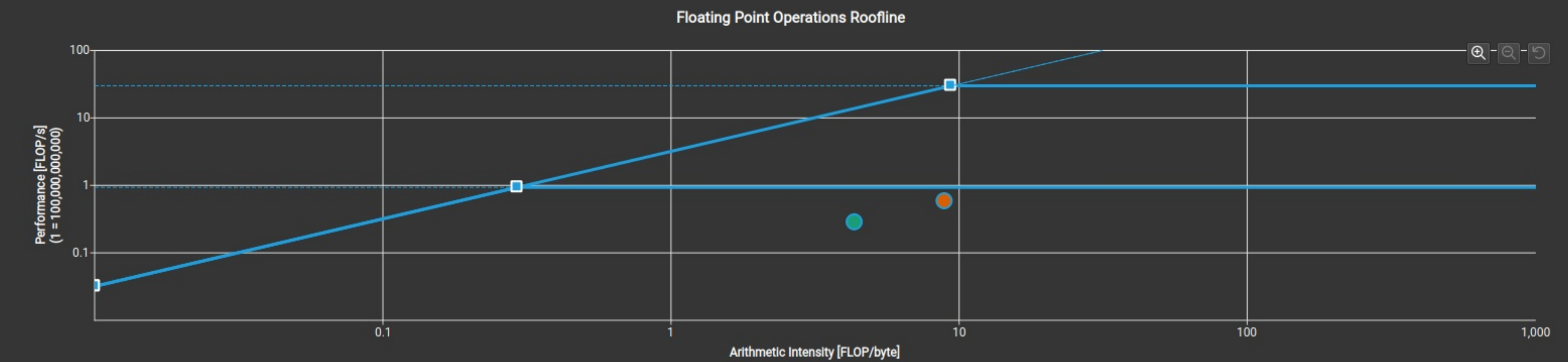
The kernel is utilizing greater than 80.0% of the available compute or memory performance of the device. To further improve performance, work will likely need to be shifted from the most utilized to another unit. Start by analyzing workloads in the [Compute Workload Analysis](#) section.

FP64/32 Utilization

The ratio of peak float (fp32) to double (fp64) performance on this device is 32:1. The kernel achieved close to 1% of this device's fp32 peak performance and 61% of its fp64 peak performance. If [Compute Workload Analysis](#) determines that this kernel is fp64 bound, consider using 32-bit precision floating point operations to improve its performance. See the [Kernel Profiling Guide](#) for more details on roofline analysis.



Compute Throughput Breakdown				Memory Throughput Breakdown			
SM: Pipe Fp64 Cycles Active [%]		81.90		DRAM: Cycles Active [%]		2.01	
SM: Pipe Alu Cycles Active [%]		43.49		DRAM: Dram Sectors [%]		1.46	
SM: Issue Active [%]		34.16		L1: M L1tex2xbar Req Cycles Active [%]		0.78	
SM: Inst Executed [%]		34.16		L2: T Sectors [%]		0.67	
SM: Mio Pq Write Cycles Active [%]		26.91		L2: Xbar2lts Cycles Active [%]		0.67	
SM: Inst Executed Pipe Adu [%]		19.71		L1: Lsuin Requests [%]		0.39	
IDC: Request Cycles Active [%]		11.31		L2: D Sectors [%]		0.32	
SM: Pipe Fma Cycles Active [%]		10.22		L1: Data Pipe Lsu Wavefronts [%]		0.28	
SM: Mio Inst Issued [%]		10.11		GPU: Compute Memory Access Throughput Internal Activity [%]		0.17	
SM: Mio Pq Read Cycles Active [%]		5.92		L2: T Tag Requests [%]		0.17	
SM: Inst Executed Pipe Cbu Pred On Any [%]		4.60		L1: Data Bank Reads [%]		0.10	
SM: Mio2rf Writeback Active [%]		2.89		L1: Data Bank Writes [%]		0.10	
SM: Inst Executed Pipe Lsu [%]		0.39		L1: Lsu Writeback Active [%]		0.06	
SM: Inst Executed Pipe Xu [%]		0.13		L2: Lts2xbar Cycles Active [%]		0.00	
SM: Pipe Tensor Cycles Active [%]		0		L1: Texin Sm2tex Req Cycles Active [%]		0.00	
SM: Inst Executed Pipe Fp16 [%]		0		L1: F Wavefronts [%]		0.00	
SM: Pipe Shared Cycles Active [%]		0		L2: D Sectors Fill Device [%]		0.00	
SM: Memory Throughput Internal Activity [%]		0		L1: M Xbar2l1tex Read Sectors [%]		0	
SM: Instruction Throughput Internal Activity [%]		0		L1: Tex Writeback Active [%]		0	
SM: Inst Executed Pipe Uniform [%]		0		L2: D Atomic Input Cycles Active [%]		0	
SM: Inst Executed Pipe Tex [%]		0		L2: D Sectors Fill Sysmem [%]		0	
SM: Inst Executed Pipe lpa [%]		0		L1: Data Pipe Tex Wavefronts [%]		0	
				GPU: Compute Memory Request Throughput Internal Activity [%]		0	



	# Operations	# Operations / Cycle	# Operations / s	Peak %	Peak Operations / Cycle	Peak Operations / s
Src:fp16,bf16,tf32 Dst:fp32	0	0	0	0	40,960	23,948.21
Src:fp16 Dst:fp16	0	0	0	0	40,960	23,948.21
Src:int1	0	0	0	0	81,920	47,896.41
Src:int4	0	0	0	0	81,920	47,896.41
Src:int8	0	0	0	0	81,920	47,896.41

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]	1,257,861.20	Average L1 Active Cycles [cycle]	1,257,861.20
Average L2 Active Cycles [cycle]	110,532.94	Average SMSP Active Cycles [cycle]	1,255,956.49
Average DRAM Active Cycles [cycle]	216,228.50	Total SM Elapsed Cycles [cycle]	50,571,128
Total L1 Elapsed Cycles [cycle]	50,571,128	Total L2 Elapsed Cycles [cycle]	59,161,024
Total SMSP Elapsed Cycles [cycle]	202,284,512	Total DRAM Elapsed Cycles [cycle]	86,263,808

Workload Distribution

	Average	Min	Max	Sum
SM Active Cycles	1,257,861.20	1,254,404	1,262,086	50,314,448
SMSP Active Cycles	1,255,956.49	1,249,981	1,261,988	200,953,039
L1 Active Cycles	1,257,861.20	1,254,404	1,262,086	50,314,448
L2 Active Cycles	110,532.94	109,877	111,386	3,537,054
DRAM Active Cycles	216,228.50	215,640	216,880	1,729,828