

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 [%]	79.77	Duration [ms]	2.22
Memory Throughput [%]	10.98	Elapsed Cycles [cycle]	1,299,137
L1/TEX Cache Throughput [%]	11.83	SM Active Cycles [cycle]	1,292,207.82
L2 Cache Throughput [%]	0.65	SM Frequency [Mhz]	584.98
DRAM Throughput [%]	1.95	DRAM Frequency [Ghz]	4.99

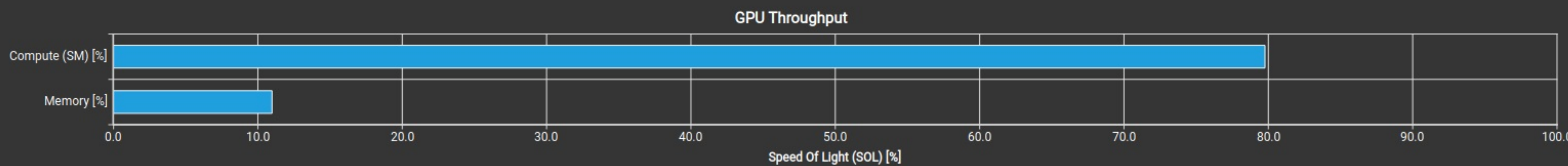
High Compute Throughput

Compute is more heavily utilized than Memory: Look at the [Compute Workload Analysis](#) section to see what the compute pipelines are spending their time doing. Also, consider whether any computation is redundant and could be reduced or moved to look-up tables.

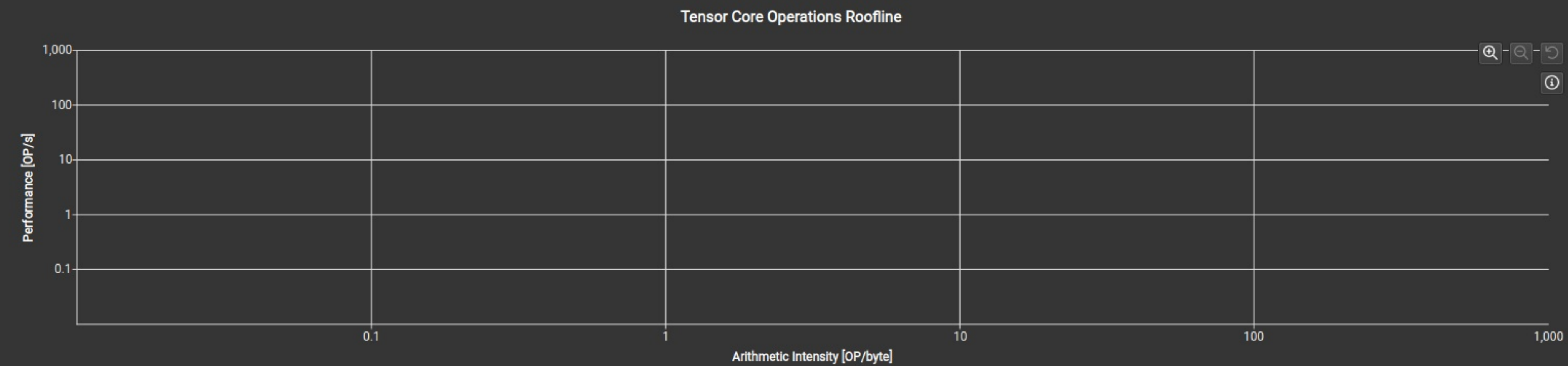
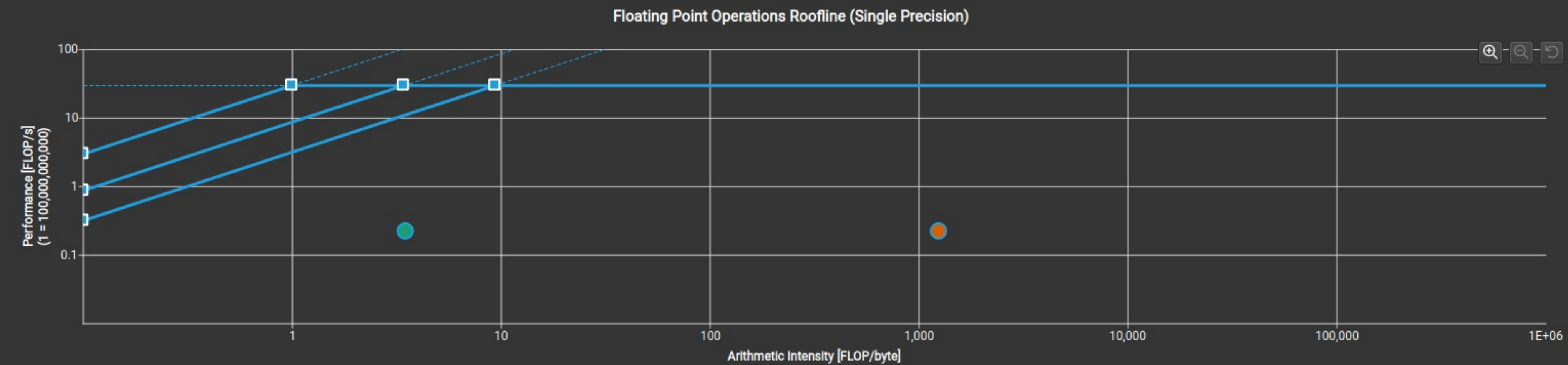
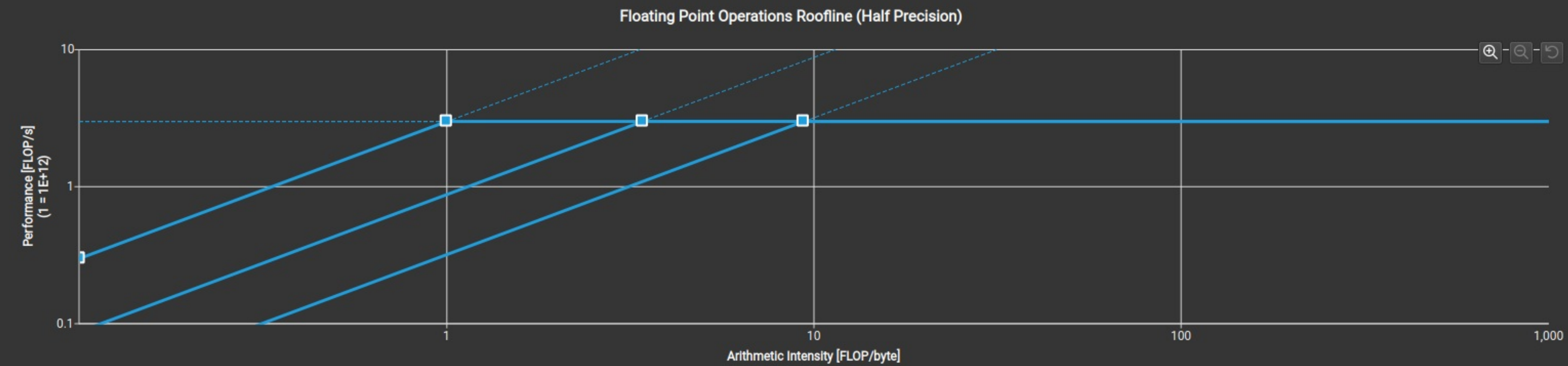
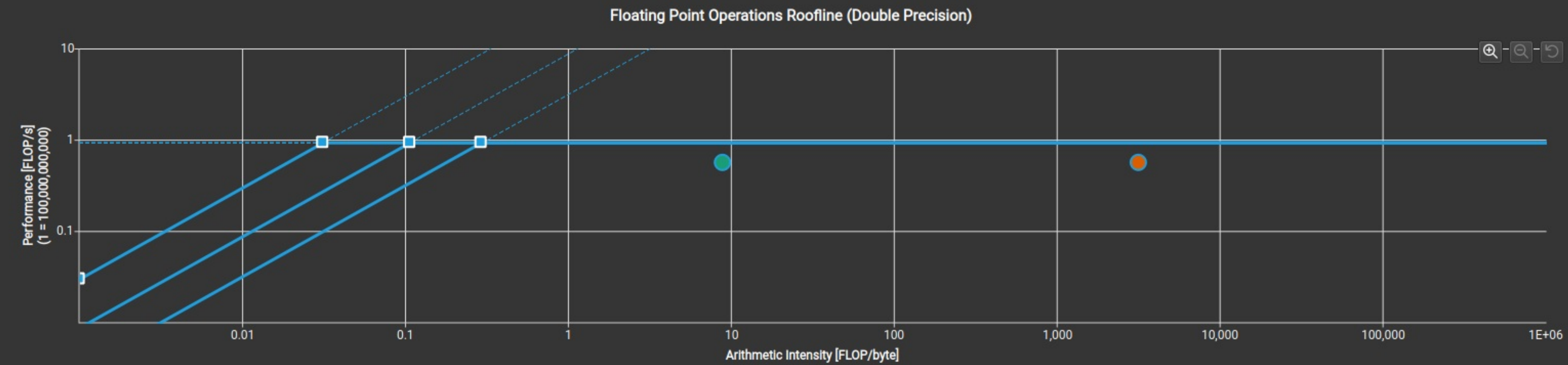
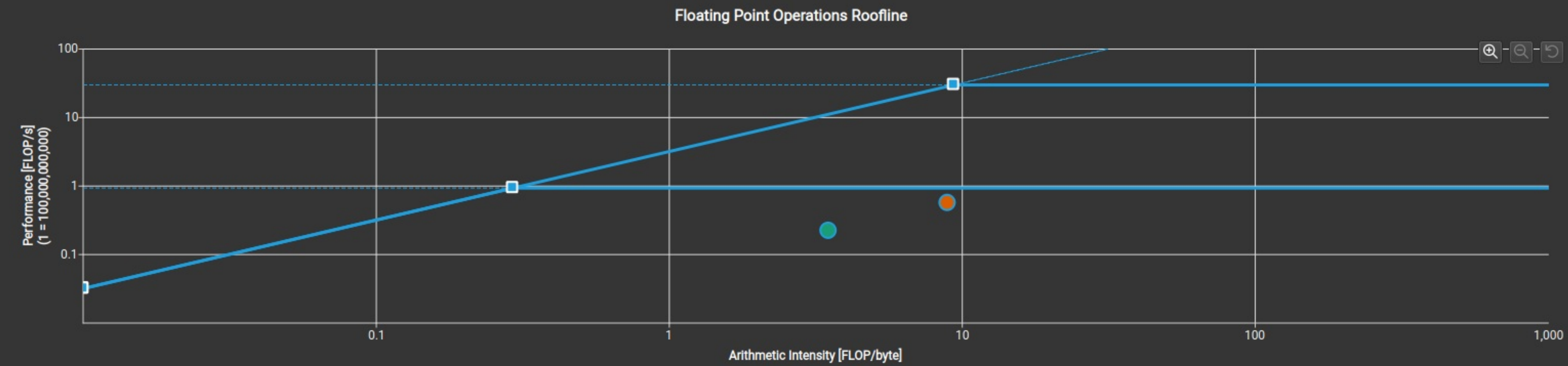
FP64/32 Utilization

Est. Local Speedup: 69.58%

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 60% 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 [%]	79.77	L1: Lsuin Requests [%]	10.98		
SM: Pipe Alu Cycles Active [%]	48.39	L1: Data Pipe Lsu Wavefronts [%]	5.91		
SM: Issue Active [%]	43.22	L1: Lsu Writeback Active [%]	2.62		
SM: Inst Executed [%]	43.21	DRAM: Cycles Active [%]	1.95		
SM: Mio Pq Write Cycles Active [%]	31.11	DRAM: Dram Sectors [%]	1.42		
SM: Inst Executed Pipe Adu [%]	23.39	L1: M L1tex2xbar Req Cycles Active [%]	0.76		
IDC: Request Cycles Active [%]	15.77	L1: Data Bank Reads [%]	0.69		
SM: Pipe Fma Cycles Active [%]	15.19	L2: T Sectors [%]	0.65		
SM: Mio Inst Issued [%]	14.78	L2: Xbar2lts Cycles Active [%]	0.65		
SM: Inst Executed Pipe Lsu [%]	10.98	L1: Data Bank Writes [%]	0.40		
SM: Mio Pq Read Cycles Active [%]	10.66	L2: D Sectors [%]	0.31		
SM: Inst Executed Pipe Cbu Pred On Any [%]	8.52	GPU: Compute Memory Access Throughput Internal Activity [%]	0.16		
SM: Mio2rf Writeback Active [%]	5.28	L2: T Tag Requests [%]	0.16		
SM: Inst Executed Pipe Xu [%]	0.13	L2: Lts2xbar Cycles Active [%]	0.00		
SM: Inst Executed Pipe Uniform [%]	0.03	L1: Texin Sm2tex Req Cycles Active [%]	0.00		
SM: Pipe Tensor Cycles Active [%]	0	L1: F Wavefronts [%]	0.00		
SM: Inst Executed Pipe Fp16 [%]	0	L2: D Sectors Fill Device [%]	0.00		
SM: Pipe Shared Cycles Active [%]	0	L2: D Atomic Input Cycles Active [%]	0		
SM: Memory Throughput Internal Activity [%]	0	L1: Tex Writeback Active [%]	0		
SM: Instruction Throughput Internal Activity [%]	0	L2: D Sectors Fill System [%]	0		
SM: Inst Executed Pipe Tex [%]	0	L1: M Xbar2l1tex Read Sectors [%]	0		
SM: Inst Executed Pipe Ipa [%]	0	GPU: Compute Memory Request Throughput Internal Activity [%]	0		
		L1: Data Pipe Tex Wavefronts [%]	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	23,941.79
Src:fp16 Dst:fp16	0	0	0	0	40,960	23,941.79
Src:int1	0	0	0	0	81,920	47,883.58
Src:int4	0	0	0	0	81,920	47,883.58
Src:int8	0	0	0	0	81,920	47,883.58

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]	1,292,207.82	Average L1 Active Cycles [cycle]	1,292,207.82
Average L2 Active Cycles [cycle]	118,524.22	Average SMSP Active Cycles [cycle]	1,290,404.96
Average DRAM Active Cycles [cycle]	216,113	Total SM Elapsed Cycles [cycle]	51,923,760
Total L1 Elapsed Cycles [cycle]	51,923,760	Total L2 Elapsed Cycles [cycle]	60,759,360
Total SMSP Elapsed Cycles [cycle]	207,695,040	Total DRAM Elapsed Cycles [cycle]	88,623,104

Workload Distribution				
	Average	Min	Max	Sum
SM Active Cycles	1,292,207.82	1,288,718	1,295,878	51,688,313
SMSP Active Cycles	1,290,404.96	1,284,714	1,295,798	206,464,794
L1 Active Cycles	1,292,207.82	1,288,718	1,295,878	51,688,313
L2 Active Cycles	118,524.22	109,709	369,123	3,792,775
DRAM Active Cycles	216,113	215,532	216,540	1,728,904