Analysis Report

ColorToGrayscale(float*, float*, int, int)

Duration	5.727 µs
Grid Size	[21,31,1]
Block Size	[16,16,1]
Registers/Thread	12
Shared Memory/Block	0 B
Shared Memory Executed	0 B
Shared Memory Bank Size	4 B

[0] Tesla P100-PCIE-16GB

GPU UUID	GPU-ebadc2f2-0e1a-33a1-db44-1c2de22b5985
Compute Capability	6.0
Max. Threads per Block	1024
Max. Threads per Multiprocessor	2048
Max. Shared Memory per Block	48 KiB
Max. Shared Memory per Multiprocessor	64 KiB
Max. Registers per Block	65536
Max. Registers per Multiprocessor	65536
Max. Grid Dimensions	[2147483647, 65535, 65535]
Max. Block Dimensions	[1024, 1024, 64]
Max. Warps per Multiprocessor	64
Max. Blocks per Multiprocessor	32
Half Precision FLOP/s	9.523 TeraFLOP/s
Single Precision FLOP/s	9.523 TeraFLOP/s
Double Precision FLOP/s	4.761 TeraFLOP/s
Number of Multiprocessors	56
Multiprocessor Clock Rate	1.329 GHz
Concurrent Kernel	true
Max IPC	3
Threads per Warp	32
Global Memory Bandwidth	732.16 GB/s
Global Memory Size	15.899 GiB
Constant Memory Size	64 KiB
L2 Cache Size	4 MiB
Memcpy Engines	2
PCIe Generation	3
PCIe Link Rate	8 Gbit/s
PCIe Link Width	16

1. Compute, Bandwidth, or Latency Bound

The first step in analyzing an individual kernel is to determine if the performance of the kernel is bounded by computation, memory bandwidth, or instruction/memory latency. Unfortunately, the device executing this kernel can not provide the profile data needed for this analysis.

2. Instruction and Memory Latency

Instruction and memory latency limit the performance of a kernel when the GPU does not have enough work to keep busy. Unfortunately, the device executing this kernel can not provide the profile data needed for this analysis.

3. Compute Resources

GPU compute resources limit the performance of a kernel when those resources are insufficient or poorly utilized. Unfortunately, the device executing this kernel can not provide the profile data needed for this analysis.

4. Memory Bandwidth

Memory bandwidth limits the performance of a kernel when one or more memories in the GPU cannot provide data at the rate requested by the kernel. Unfortunately, the device executing this kernel can not provide the profile data needed for this analysis.