

Lg Throttle Stalls

Est. Speedup: 50.57%

On average, each warp of this workload spends 18.4 cycles being stalled waiting for the L1 instruction queue for local and global (LG) memory operations to be not full. Typically, this stall occurs only when executing local or global memory instructions extremely frequently. Avoid redundant global memory accesses. Try to avoid using thread-local memory by checking if dynamically indexed arrays are declared in local scope, of if the kernel has excessive register pressure causing by spills. If applicable, consider combining multiple lower-width memory operations into fewer wider memory operations and try interleaving memory operations and math instructions. This stall type represents about 63.0% of the total average of 29.3 cycles between issuing two instructions.

L1TEX Global Load Access P

Est. Speedup: 29.14%

The memory access pattern for global loads from L1TEX might not be optimal. On average, only 22.6 of the 32 bytes transmitted per sector are utilized by each thread. This could possibly be caused by a stride between threads. Check the Source Counters section for uncoalesced global loads.

•

 \odot

Uncoalesced Global Accesse

Est. Speedup: 27.27%

This kernel has uncoalesced global accesses resulting in a total of 4609584 excessive sectors (28% of the total 16746864 sectors). Check the L2 Theoretical Sectors Global Excessive table for the primary source locations. The CUDA Programming Guide has additional information on reducing uncoalesced device memory accesses.

