Page: Details | Result: 0 - 23 - nqfaf | Add Baseline | Apply Rules | Occupancy Calculator | Source Comparison | Save as PDF

	Result	Time	Cycles	Regs	GPU	SM Frequency	CC	Process
Current	23 - nqfaf (9248, 1, 1)x(64, 1, 1)	650,76 msecond	886.574.386	38	0 - NVIDIA GeForce GTX 1650 Ti	n/a	7.5	[5892] NQueensFAF.exe

Memory Workload Analysis

Memory Chart

Detailed analysis of the memory resources of the GPU. Memory can become a limiting factor for the overall kernel performance when fully utilizing the involved hardware units (Mem Busy), exhausting the available communication bandwidth between those units (Max Bandwidth), or by reaching the maximum throughput of issuing memory instructions (Mem Pipes Busy). Detailed chart of the memory units. Detailed tables with data for each memory unit.

Memory Throughput [Mbyte/second]	540,99	Mem Busy [%]	43,58
L1/TEX Hit Rate [%]	66,67	Max Bandwidth [%]	49,56
L2 Hit Rate [%]	31,72	Mem Pipes Busy [%]	49,56

Memory L2 Compression The optional metric lts__average_gcomp_input_sector_success_rate.pct could not be found. Collecting it as an additional metric could enable the rule to provide more guidance.

L1TEX Global Load Access Pattern The memory access pattern for global loads in L1TEX might not be optimal. On average, this kernel accesses 4.0 bytes per thread per memory request; but the address pattern, possibly caused by the stride between threads, results in 16.0 sectors per request, or 16.0*32 = 512.0 bytes of cache data transfers per request. The optimal thread address pattern for 4.0 byte accesses would result in 4.0*32 = 128.0 bytes of cache data transfers per request, to maximize L1TEX cache performance. Check the Source Counters section for uncoalesced global loads.

L1TEX Global Store Access Pattern The memory access pattern for global stores in L1TEX might not be optimal. On average, this kernel accesses 4.0 bytes per thread per memory request; but the address pattern, possibly caused by the stride between threads, results in 8.0 sectors per request, or 8.0*32 = 256.0 bytes of cache data transfers per request. The optimal thread address pattern for 4.0 byte accesses would result in 4.0*32 = 128.0 bytes of cache data transfers per request, to maximize L1TEX cache performance. Check the Source Counters section for uncoalesced global stores.

Shared Load Bank Conflicts
Est. Local Speedup: 35.64% The memory access pattern for shared loads might not be optimal and causes on average a 1.6 - way bank conflict across all 2387040385 shared load requests.This results in 1321652921 bank conflicts, which represent 35.64% of the overall 3708693306 wavefronts for shared loads. Check the Source Counters section for uncoalesced shared loads.

Shared Store Bank Conflicts
Est. Local Speedup: 54.23% The memory access pattern for shared stores might not be optimal and causes on average a 2.2 - way bank conflict across all 1118570877 shared store requests.This results in 1331988457 bank conflicts, which represent 54.23% of the overall 2456006910 wavefronts for shared stores. Check the Source Counters section for uncoalesced shared stores.

Source Counters

Source metrics, including branch efficiency and sampled warp stall reasons. Warp Stall Sampling metrics are periodically sampled over the kernel runtime. They indicate when warps were stalled and couldn't be scheduled. See the documentation for a description of all stall reasons. Only focus on stalls if the schedulers fail to issue every cycle.

Branch Instructions [inst]	7.012.056.594	Branch Efficiency [%]	76,21
Branch Instructions Ratio [%]	0,16	Avg. Divergent Branches	17.469.039,47

Uncoalesced Global Accesses
Est. Speedup: 2.53% This kernel has uncoalesced global accesses resulting in a total of 887808 excessive sectors (67% of the total 1331712 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.

L2 Theoretical Sectors Global Excessive

Location	Value	Value (%)
0x70143b390 in nqfaf	221.952	25
0x70143b380 in nqfaf	221.952	25
0x70143b370 in nqfaf	221.952	25
0x70143b360 in nqfaf	221.952	25
0x70143c6d0 in nqfaf	0	0

Uncoalesced Shared Accesses
Est. Speedup: 38.80% This kernel has uncoalesced shared accesses resulting in a total of 2236905871 excessive wavefronts (39% of the total 5742517133 wavefronts). Check the L1 Wavefronts Shared Excessive table for the primary source locations. The CUDA Best Practices Guide has an example on optimizing shared memory accesses.

L1 Wavefronts Shared Excessive

Location	Value	Value (%)
0x70143c150 in nqfaf	566.617.332	25
0x70143c460 in nqfaf	557.602.257	25
0x70143c560 in nqfaf	556.625.164	25
0x70143c250 in nqfaf	555.754.665	25
0x70143c4a0 in nqfaf	247.833	0

Warp Stall Sampling (All Samples)

Location	Value	Value (%)
0x70143c2c0 in nqfaf	256.249	7
0x70143c660 in nqfaf	249.474	7
0x70143c350 in nqfaf	249.075	7
0x70143c5d0 in nqfaf	230.304	6
0x70143c600 in nqfaf	184.033	5

Most Instructions Executed

Location	Value	Value (%)
0x70143c5f0 in nqfaf	1.180.268.006	3
0x70143c2e0 in nqfaf	1.131.043.104	3
0x70143c360 in nqfaf	596.941.996	1
0x70143c350 in nqfaf	596.941.996	1
0x70143c340 in nqfaf	596.941.996	1

To customize your report even further, you might want to learn about custom sections and writing your own rules.
You might also want to consider adding individual metrics.