Reading: CHAPTER 5 of the attached book is on Caches – 5.1-5.3

Assignment Details

The goal of this homework assignment is to explore power profiling, compare the warp schedulers and Cache conscious behavior of the applications.

Configuration files are in the location: https://github.com/gpgpu-sim/gpgpu-sim_distribution/tree/master/configs/tested-cfgs

For the 6 GPU configurations

1. Enable different warp schedulers in the configs/GTX480/gpgpusim.config file
   Example: https://github.com/gpgpu-sim/gpgpu-sim_distribution/blob/master/configs/tested-cfgs/SM2_GTX480/gpgpusim.config#L140

   Two Level Scheduler with active and pending pools (-gpgpu_scheduler two_level_active:6:0:1)

   Loose round robbin scheduler(-gpgpu_scheduler lrr)

   Greedy then oldest scheduler(-gpgpu_scheduler gto)

2. Enable the power profiling by setting -power_simulation_enabled 1

   Make sure you copy all the files the configs/GTX480/* into your Benchmarks directory before running each application.

Benchmarks

For this assignment, we will make use of 20 ISPASS2009/Rodinia/CUDA toolkit benchmarks. Choose applications of your choice from these standard benchmark suites.

Note: output statistics are generated at the end of each kernel. Some benchmarks may have multiple kernel, so be sure to use the results from the last kernel only.

Questions

Each question must be answered with supporting data

1. What is the runtime for each configuration?

   Example: Plot showing the IPC on Y-axis and application name on X-axis, legend: different warp schedulers.

2. Which warp scheduler has the best cache hit rate? Why?

Example: Plot showing the following on Y-axis and application name on X-axis, legend: different warp schedulers.

(i) L1D miss rates

(ii) L2 miss rates

3. Categorize the applications w.r.t the L1D and L2 Cache hit rates. What changes do you observe w.r.t L1D and L2 cache hit rates when the L1D cache size is increased from 32KB to 8MB? (use warp scheduler GTO)

4. What percentage of power is consumed by Execution units, DRAM, Register Files in each application run? Do you notice any correlation between the L1D cache hit rates observed in Question 3 and the Power consumption between different applications?

Submission

Please submit typed document answering the above questions along with justification data on google classroom or github.