Homework 5

(Due March 10, 2020)

1. What is the major contribution of the Layer Performance Matching?

2. We have given five case studies in Lecture 9. How these case studies related to I/O applications? Can you give one or two good examples for data transfer matching?

3. What is memory stall time? What is the advantage of the Layered Performance Matching method? Why LPM can improve memory performance (in terms of memory stall time) by more than one hundred times?

4. The following figure shows a cycle-accurate pure^2 (pure-pure) miss example. In this example, there are 4 data accesses. Base on this figure, please calculate
   a. What is the average miss penalty of L1 misses?
   b. What is the average pure miss penalty of L1 pure misses?
   c. What is the average miss penalty of L2 misses?
   d. What is the average pure miss penalty of L2 pure misses?
   e. How many pure^2 misses are there in this figure? How many pure^2 miss cycles does each pure^2 miss have?

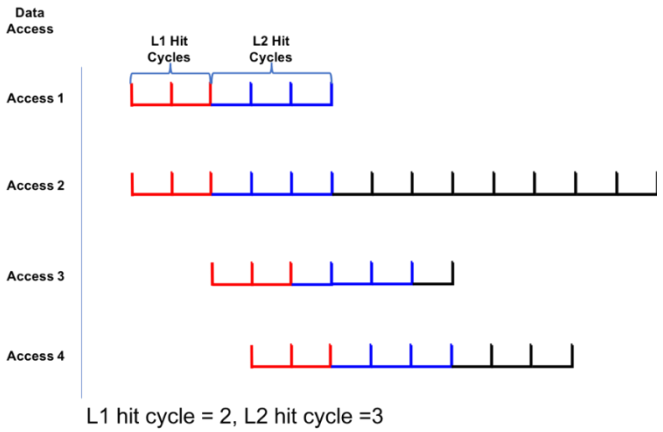

Figure 1. Cycle-accurate C-AMAT Example

5. Suppose there is no sequential portion, the program can be perfectly parallelized. The performance bottleneck is data access delay, and the data access delay is $ɯ_c$. Please calculate the speedup of the three models of parallel speedup assuming the data access delay cannot be parallelized, and the number processors is p = 20.
   a) Amdahl's law
   b) Gustafson's law
   c) Sun/Ni's law, assuming G(p)=p/2

6. Please give a short answer to the following questions.
   a) Why victim cache can improve cache performance?
   b) What is the motivation and idea behind the skewed associative caches?

7. Please give a multicore false sharing example.

8. Each miss is allocated on an MSHR (Miss Status Holding Register) entry before a request to service that miss is sent to memory at each level of a memory hierarchy. So, the miss number can be calculated via MSHR allocation. Based on your knowledge of pure miss, please describe how to use MSHR to detect pure miss of L2.

Challenging Homework (could be part of your term project)

C-1. What is your thought on overlapping based latency hiding? How C-AMAT can help?
C-2. Do you have any thought for any of the challenging questions given in Lecture 12?