ECE4700J Computer Architecture

Summer 2022

HW #4

Due: **2:59am (Beijing Time) July 4th, 2022**

*Please submit a single **PDF** file on Canvas*

Q1 (30%): Consider the following program and cache behaviors.

Data Reads per 1000 instructions	Date Write per 1000 Instructions	Instruction Cache Miss Rate	Data Cache Miss Rate	Block Size (bytes)
250	100	0.30%	2%	64

(1) (15%) Suppose a CPU with a write-through, write-allocate cache achieves a CPI of 2. What are the read and write bandwidths (measured by bytes per cycle) between RAM and the cache? (Assume each miss generates a request for one block.)

(2) (15%) For a write-back, write-allocate cache, assuming 30% of replaced data cache blocks are dirty, what are the read and write bandwidths needed for a CPI of 2?

Q2 (10%): Based on what you have learned in this course and in VE370, please describe on your own words at least 3 differences between **Cache** and **Virtual Memory**.

Q3 (20%): Whereas larger caches have lower miss rates, they also tend to have longer hit times. Assume a direct-mapped 8 KB cache has 0.22 ns hit time and miss rate m1; also assume a 4-way associative 64 KB cache has 0.52 ns hit time and a miss rate m2.

(1) (10%) If the miss penalty is 100 ns, when would it be advantageous to use the smaller cache to reduce the overall memory access time?

(2) (10%) Repeat part (1) for miss penalties of 10 and 1000 cycles. Conclude when it might be advantageous to use a smaller cache.

Q4 (10%): You are investigating the possible benefits of a way predicting L1 cache. Assume that a 64 KB four-way set associative single-banked L1 data cache is the cycle time limiter. This cache has an AMAT of 1.69ns. The clock cycle time of the system is 0.5ns. For an alternative cache organization, you are considering a **way-predicted cache** modeled as **a 64 KB direct mapped cache** with 80% prediction accuracy. The 64KB direct mapped cache has an access time of 0.86ns, and a miss rate of 0.33% along with miss penalty of 20 cycles. Unless stated otherwise, assume that a mispredicted way access that hits in the cache takes one more cycle. What is the average memory access time of the way-predicted cache?

Q5 (10%): Use your own words to explain the advantage and disadvantage does a Harvard L1 cache have over a unified L1 cache?

Q6 (20%): Consider a two-level memory hierarchy made of L1 and L2 data caches. Assume that both caches use write-back policy on write hit and both have the same block size. List the actions taken in response to the following events:
(1) (10%) An L1 cache miss when the caches are organized in an inclusive hierarchy:

Example Answer: When there is L1 cache miss, it will first access L2 cache, then if hit, XXX, if miss XXX. In both hit and miss, if storing an L1 evicted block in L2 causes a block to be evicted from L2, then XXX.

(2) (10%) An L1 cache miss when the caches are organized in an exclusive hierarchy. The style of the answers should be similar to (1).