Department of Computer Science and Engineering
National Sun Yat-sen University
Second Semester of 2021 PhD Qualifying Exam

Subject： Computer Architecture

Problem 1: True or False (10 points)

_____ (1) Increasing the block size of a cache is likely to take advantage of spatial locality.

_____ (2) The miss rate may go up if the block size is made very large

_____ (3) Increasing the associativity decreases miss rate due to lower compulsory miss

_____ (4) In the virtual memory system, only the write-through policy is practical because of the cheaper implementation cost.

_____ (5) Cache hit can be implied the TLB hit.

_____ (6) If the cycle time and the CPI both increase by 10% and the number of instruction decreases by 20%, then the execution time will remain the same.

_____ (7) RAID system aim to increase the storage space in the storage system.

_____ (8) The Von Neumann model divides the computer hardware into three subsystems.

_____ (9) Cache coherence problem is usually happened in the single-core system.

_____ (10) To ensure the inclusive property of memory hierarchy system, we need to move the data in the upper-level memory to the lower level memory.

※ Note that you need to explain the reason if you think any above description is incorrect.

Problem 2: (30 points)

You are designing a write buffer between a write-through L1 cache and a write-back L2 cache. The write-data bus of L2 cache is 16-byte wide and can perform a write to an independent cache address every 4 processor cycles.

A. (10%) How many bytes should each write buffer entry be?

B. (10%) What speedup could be expected in the steady state by using a merging write buffer instead of a non-merging buffer when zeroing memory by the execution of 64-bit stores if all other instructions could be issued in parallel with the stores and the blocks are present in the L2 cache?

C. (10%) What would the effect of possible L1 misses be on the number of required write buffer entries for systems with blocking and nonblocking caches?

Problem 3: (40 points)

You are a system engineer. Today, you need to design a memory hierarchical system, including one CPU, one or two level cache, one main memory, and one hard disk. Currently, you have the following design policies for the cache-level design.

Policy 1: With only L1 cache by using direct mapping strategy

Policy 2: With only L1 cache by using 2-way associativity

Policy 3: With direct mapping L1 and L2 caches

Policy 4: With direct mapping L1 cache and 2-way associative L2 cache

Policy 5: With 2-way associative L1 cache and direct mapping L2 cache

Policy 6: With 2-way associative L1 and L2 caches

Assume that there is one embedded TLB in each cache level and one embedded page table in the main memory (*i.e.*, we do not need extra memory to store the TLB and page table). On the other hand, the specifications of this memory hierarchical system are

- This is a 32-bit machine.
- The base CPI is 1.0 and the clock rate is 5GHz.
- Each cache block is a single-word block.
- The L1 cache can contain 8KB data.
- The L2 cache can contain 16KB data.
- The page size is 2^{12} bytes.
- In the TLB and page table, we need to involve extra one dirty bit to implement the write-back policy; one reference bit to approximate the LRU replacement policy; one valid bit to judge the data hit/miss.
- The number of the TLB entries in L1 and L2 caches are 10 and 20 respectively. Besides, the number of page table entry is 30. To reduce the miss rate, the fully associative policy is adopted to implement the TLB and page table.

A. (10%) Please determine the number of bits required in the page table, TLB in the L1 cache, and TLB in the L2 cache.

B. (10%) Please determine the number of bits required if
   a. L1 cache is implemented by using direct mapping strategy.
   b. L1 cache is implemented by using 2-way associative mapping strategy.
   c. L2 cache is implemented by using direct mapping strategy.
   d. L2 cache is implemented by using 2-way associative mapping strategy.

C. (20%) Without considering the data transferring time, we assume the access time of L2 cache is 5 ns including all the miss handling; access time of the main memory is 100 ns including all the miss handling; access time of the hard disk is 1 us including all the miss handling. According to the data transference time between each memory level, we ignore the data transference time between the L1 and L2 cache and the data transference time between the main memory and the lowest level cache and disk are both 50 ns including all the miss handling. Besides, the miss rate of the L1 cache and the embedded TLB are both 2%; the miss rate of the L2 cache and the embedded TLB are both 0.5%; the miss rate of the main memory is 0.1%. During the manufacturing, we need to spend 0.01 USD to handle one bit in each kind of memory. Please provide a design suggestion, including how many cache level you suggest and what kind of mapping strategy for each cache level you suggest, to your customer by considering the system performance and the manufacturing simultaneously cost. You need to explain your reasons.

Problem 4: (20 points)

By following your decided design policy in the previous problem, you aim to further improve the bandwidth between the cache and the main memory in the memory system. Now, you have the following three design policies. Besides, the specification of this memory system is the same as the previous problem.

Policy 1: One-word-width memory organization

Policy 2: Five-word-width memory organization

Policy 3: Interleaved memory organization with five memory banks

A. (10%) You need to evaluate the three design polices and find the proper design policy. We assume that the access timing overhead in wider bus is 25%. Please determine which design policy can help us to have the best memory performance if we do not consider the manufacturing cost.

B. (10%) We assume that we need to spend extra 10 USD to implement the Policy 2 and 5 USD to implement the Policy 3. Please evaluate each design policy again if you need to consider the manufacturing cost. Then, please determine the most cost-efficient design policy.