**CSE 490/590 – Quiz 2 – Summer 2025 Solution**

**[Question 1] (6 Points)**

Assume the presence of the following memories in a MIPS system:
i. L1 cache
ii. Main Memory
iii. L3 cache
iv. Registers
v. SSDs
vi. L2 cache
vii. Hard drive

Show the memory hierarchy and order them in terms of (State Increasing or decreasing order):

a. Speed

Increasing Speed:
Hard drive → SSDs → Main Memory → L3 Cache → L2 Cache → L1 Cache → Registers

b. Cost-per-byte

Increasing Cost:
Hard drive → SSDs → Main Memory → L3 Cache → L2 Cache → L1 Cache → Registers

c. Memory capacity (size)

Increasing Size:
Registers → L1 Cache → L2 Cache → L3 Cache → Main Memory → SSDs → Hard drive

**[Question 2] (6 Points)**

Write-through and write-back are approaches used when there is a cache hit.
Explain how write-through and write-back work. Highlight the pros and cons of each.

Write-back: Handles writes by updating only to the block in the cache. The modified cache block is written to main memory only when it is replaced.

Advantages:

(i) Low latency

(ii) High throughput for write-intensive applications

Disadvantages: There is a risk of losing data if the cache is overwritten without backing up in the lower memory. Since such overwrites cannot be done, writes either take 2 cycles or use a write buffer, meaning either increased latency or increased complexity/overhead.

Write-through: Handles writes by updating both the block in the cache and the block in main memory.

Advantages:
(i) Easier to implement
(ii) Read misses never result in writes to the lower level
(iii) Data coherency/consistency

Disadvantages: Higher latency from multiple writes to the lower-level memory.

**[Question 3] (8 Points)**

a. If a direct-mapped cache has a hit rate of 80%, a hit time of 3 ns, and a miss penalty of 110 ns, what is the AMAT (Average Memory Access Time)?

AMAT = Hit Time + Miss Rate * Miss Penalty
= 3 + (0.2 * 110)
= 25ns

b. If an L2 cache is added to (a) with a hit time of 9 ns and a hit rate of 60%, what is the new AMAT? Miss penalty stays the same.

AMAT = L1 Hit Time + L1 Miss Rate * (L2 Hit Time + L2 Miss Rate * Miss Penalty)
= 3 + (0.2 * (9 + 0.4 * 110))
= 13.6ns

c. If replacing the L1 cache with a 2-way set associative increases the hit rate to 90%, but increases the hit time to 4 ns, what is the new AMAT? Miss penalty stays the same. Assume (b) does not apply. (there is no L2 cache)

AMAT = Hit Time + Miss Rate * Miss Penalty
= 4 + (0.1 * 110)
= 15ns

**[Question 4] [10 Points]**

The following are parameters of a processor:

- Instruction cache miss rate: 5%
- Data cache miss rate: 9%
- Miss penalty: 150 cycles
- 20% of the instructions to be executed are Load and Store Instructions.

Assume the CPI with ideal cache (no misses) is 2.

a. Compute the actual CPI.

We have two sets of caches here, the instruction cache and the data cache. The instruction cache contains instructions fetched from the main memory and the data cache contains data fetched from the main memory. Before executing a given instruction, the CPU would check if the instruction is in the instruction cache and would also check if the data used in the instruction is in the data cache.

Accounting for misses on both, we can write the miss penalty per instruction as follows:
Miss Penalty Per Instruction = (Instr Cache Miss Rate * Miss Penalty) + (Percentage Load/Store Instrs * Data Cache Miss Rate * Miss Penalty)

Miss Penalty Per Instruction = (0.05 * 150) + (0.20 * 0.09 * 150) = 10.2
Actual CPI = Ideal CPI + Miss Penalty Per Instruction = 2 + 10.2 = 12.2

b. Compute the percentage of execution time wasted due to cache misses.

The percentage of execution time wasted due to cache misses can be defined as the number of clock cycles used to resolve cache misses over the total number of cache cycles taken.
In this case the number of clock cycles taken to resolve the cache misses is equivalent to the miss penalty = 10.2
The total CPI is equivalent to the actual CPI from part (a) = 12.2

Therefore, the percentage of execution cycles wasted to take care of the misses would be 10.2/12.2= 0.8360 = 83.60%

**[Question 5] (10 Points)**
Assuming 32-bit memory addresses, how many bits are associated with the tag, index, and offset of the following configurations for a byte-addressable direct mapped cache memory?

a. 64 blocks, 4 bytes per block
Offset bits = log(4) = 2
Index bits = log(64) = 6
Tag bits = 32-2-6 = 24

b. 128 blocks, 8 bytes per block
Offset bits = log(8) = 3
Index bits = log(128) = 7
Tag bits = 32-3-7 = 22

**[Question 6] (12 Points)**

Consider a direct-mapped cache of size 4 Bytes. Each block in the cache can hold only 1 word (here 1 word = 1 Byte).

a) Fill in the missing cache blocks at each step according to the address reference, and specify whether it is a hit or a miss. The first step has been done for you. **Recreate the tables below on paper (recommended) or use the editor to recreate and finish the tables.**

Address references are (in order): 4, 3, 1, 4, 0, 7

Reference: 0x4

| Tag | Cache Content |
|---|---|
| 01b | Mem(4) |
| | |
| | |
| | |

hit / miss

0x3

| Tag | Cache Content |
|---|---|
| 01b | Mem(4) |
| | |
| | |
| 00b | Mem(3) |

hit / miss

0x1

| Tag | Cache Content |
|---|---|
| 01b | Mem(4) |
| 00b | Mem(1) |
| | |
| 00b | Mem(3) |

hit / miss

0x4

| Tag | Cache Content |
|---|---|
| 01b | Mem(4) |
| 00b | Mem(1) |
| | |
| 00b | Mem(3) |

hit / miss

0x0

| Tag | Cache Content |
|---|---|
| 00b | Mem(0) |
| 00b | Mem(1) |
| | |
| 00b | Mem(3) |

hit / miss

0x7

| Tag | Cache Content |
|---|---|
| 01b | Mem(4) |
| 00b | Mem(1) |
| | |
| 01b | Mem(7) |

hit / miss

Here, the upper 2 bits of the reference are the tag bits.
(e.g. If reference is 0x7 → 0111b, 01b is the tag.)

The lower two bits (11b in the case of 0x7) tell us the position of the mapped content.
(0x4 → lower two bits are 00b, this gets mapped to the first block with its tag, 01b.)

b) Calculate the miss rate for the above (Part a.)

Miss rate = no. of misses / total references
= 5 / 6
= 0.83 OR 83.33%.