NOTE: this is the SOLUTION to Quiz 6.

The correct answers are indicated for each question, with explanations as needed.

Dr. Manikas

1 4 / 4 points

A memory size of 1 KB = ___ bytes

- 2^{50}
- 2^{20}
- ✓ 2^{10}
- 2^{30}
- 2^{40}

Feedback

General Feedback

- 2^{10} = 1 K (kilo)
- 2^{20} = 1 M (mega)
- 2^{30} = 1 G (giga)
- 2^{40} = 1 T (tera)

- 2^{50} = 1 P (peta)

2 4 / 4 points

We are given a memory system with a 2-level hierarchy:

- Primary level access time = 5 ns
- Secondary level access time = 60 ns
- Hit rate = 80%

What is the **average access time** for this memory system?

✓ 16 ns

Feedback

General Feedback

For a two-level hierarchy, the average access time is given by:

$$t_a = ht_p + (1 - h)\, t_s$$

We are given

- Primary level access time t_p = 5 ns
- Secondary level access time t_s = 60 ns
- Hit rate = 80%, so h = 0.8

Then, average access time t_a is:

$$t_a = (0.8)\,(5\,ns) + (1 - 0.8)\,(60\,ns)$$
$$= (0.8)\,(5\,ns) + (0.2)\,(60\,ns)$$
$$= 4\,ns\ +\ 12\,ns\ =\ 16\,ns$$

3 4 / 4 points

We are given a memory system with the following parameters:

- Main memory size is 4 GB
- Cache size is 64 MB
- Block size is 64 bytes

If the cache is **associative,** what are the number of bits for each field below?

TAG ✓ 26 BYTE ✓ 6

Feedback

General Feedback

An **associative** cache has two fields: tag (to identify main memory block) and byte (to identify byte in a block).

The main memory size = 4 GB = $2^2 2^{30}$ B = 2^{32} B

The block size is 64 B = 2^6 B, so the main memory is divided into $2^{32}/2^6 = 2^{26}$ blocks.

Thus, the memory address fields and corresponding number of bits are:

Tag	Byte
26	6

4 4 / 4 points

We are given a memory system with the following parameters:

- Main memory size is 16 GB
- Cache size is 2 MB
- Block size is 64 bytes

If the cache is **direct-mapped,** what are the fields in the memory address?

TAG ✓ 13 GROUP ✓ 15 BYTE ✓ 6

Feedback

General Feedback

For a **direct-mapped** cache, the main memory is viewed as table. The number of table rows = number of cache blocks, or **groups**

The main memory size = 16 GB = $2^4 2^{30}$ B = 2^{34} B

The block size is 64 B = 2^6 B, so the main memory is divided into $2^{34}/2^6 = 2^{28}$ blocks.

The cache size = 2 MB = $2^1 2^{20}$ B = 2^{21} B = $2^{21}/2^6 = 2^{15}$ blocks, or 2^{15} **groups (rows)**

The number of table columns = **tags** = # main memory blocks / # cache lines = $2^{28}/2^{15} = 2^{13}$

The resulting memory address fields are:

Tag	Group	Byte
13	15	6

5 4 / 4 points

We are given a memory system with a direct-mapped cache. This cache has the following memory address fields:

Tag	Group	Byte
16	8	4

If we change the cache to **4-way set-associative,** what are the fields in the memory address?

TAG ✓ 18 SET ✓ 6 BYTE ✓ 4

Feedback

General Feedback

Recall that for a **direct-mapped** cache, the main memory is viewed as table. For this cache:

- Tag field = 16 bits, so number of table columns = 2^{16}
- Group field = 8 bits, so number of table rows = 2^8
- Byte field = 4 bits; this will be the same field size for the set-associative cache

A **4-way set associative** cache expands on the direct-mapped scheme by also viewing the cache as a table. Here, 4 ways = 4 cache table columns.

Therefore, the cache now has $2^8/2^2 = 2^6$ cache rows, or **sets**. Recall that the number of cache rows = number of main memory rows.

Since the number of rows is divided by 4, the number of main memory columns must be multiplied by the same amount. Therefore, number of main memory columns, or **tags** = $(2^{16})(2^2) = 2^{18}$.

The resulting memory address fields are:

Tag	Set	Byte
18	6	4

6 4 / 4 points

We are given a memory system with separate caches for instructions (I-cache) and data (D-cache). The system has the following specifications:

- Base CPI = 2
- Load and stores are 40% of instructions
- I-cache miss rate = 4%
- D-cache miss rate = 4%
- Miss penalty = 100 cycles

What is the **actual CPI** for this system if we consider the effects of cache misses?

✓ 7.6

Feedback

General Feedback

First, look at the effects of the **I-cache**

CPI added due to cache misses = (% items accessed)(miss rate)(miss penalty)

% items accessed = 1, since each instruction fetch accesses the I-cache

So $CPI_{I\text{-cache}} = (1)(0.04)(100) = 4$

Next, look at the effects of the **D-cache**

Here, % items accessed = 0.40, since D-cache is only accessed 40% of the time (loads and stores)

So $CPI_{D\text{-cache}} = (0.40)(0.04)(100) = 1.6$

Actual CPI = base CPI + $CPI_{I\text{-}cache}$ + $CPI_{D\text{-}cache}$ = 2 + 4 + 1.6 = **7.6**