IN1006 Systems Architecture 2023\_2024

Tutorial 05: Memory Hierarchy

Exercises

1. What is a memory hierarchy and why do we need one?
2. Describe the two principles of locality that allow (data) memory caches to operate effectively.
3. What is the small fast memory that acts as a local buffer for the main memory called?
4. In the context of a cache, explain: cache hit and cache miss?
5. Explain the operation of a direct mapped cache
6. Explain what happens during a cache miss
7. Describe the differences between combined and Harvard cache architectures
8. What is a Write Strategy? Give two examples.
9. Suppose a computer using direct mapped cache has 232 byte of byte-addressable main memory, and a cache of 1024 blocks, where each cache block contains 32 bytes.
10. How many blocks of main memory are there?
11. What is the format of a memory address as seen by the cache, i.e., what are the sizes of the tag, block, and offset fields?
12. To which cache block will the memory address 0x000063FA map?

**10** Suppose we have a computer that uses a memory address word size of 8 bits. This computer has a 16-byte cache with 4 bytes per block. The computer accesses a number of memory locations throughout the course of running a program. Suppose this computer uses direct-mapped cache. The format of a memory address as seen by the cache is given below:

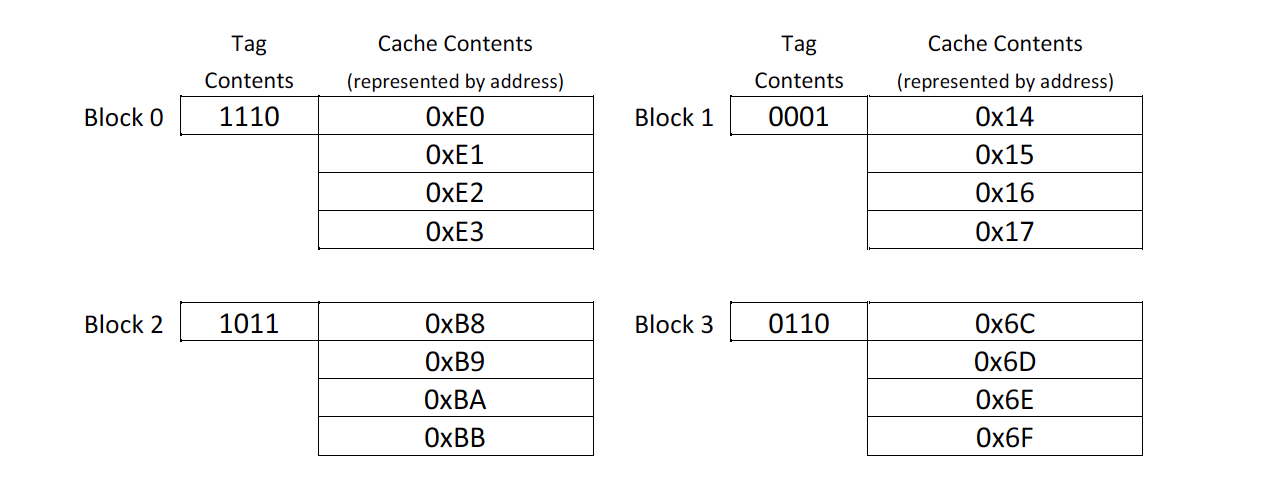
Tag: 4 bits

Block: 2 bits

Offset: 2 bits

The system accesses memory addresses in this exact order: 0x6E, 0xB9, 0x17, 0xE0, 0x4E, 0x4F, 0x50, 0x91, 0xA8, 0xA9, 0xAB, 0xAD, 0x93, and 0x94. The memory addresses of the first four accesses have been loaded into the cache blocks as shown below. (The contents of the tag are shown in binary and the cache “contents” are simply the address stored at that cache location.)

1. What is the hit ratio for the entire memory reference sequence given above, assuming we count the first four accesses as misses?
2. What memory blocks will be in the cache after the last address has been accessed?
3. What is the effective access time EAT if AccessC=10ns and AccessMM=200ns?

**