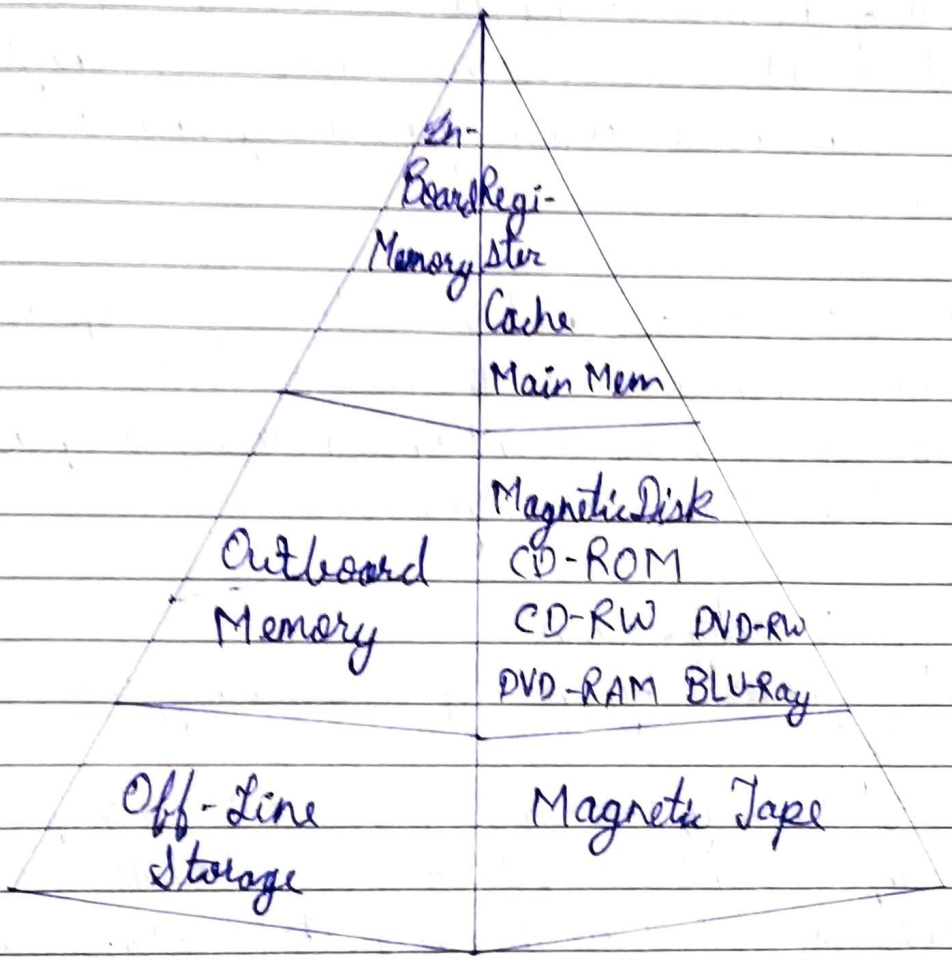
Varun Khadayate AO16 17-10-20

# COA Assignment

Q1. The memory hierarchy in a computer system mainly includes different storage devices. Most of the computers were inbuilt with extra storage to run more powerfully beyond the main memory capacity.




The memory hierarchy characteristics mainly include the following:-

- Performance: Previously, the designing of a computer was done without memory hierarchy, and the speed gap among the main memory as well as the CPU registers enhances because of the huge disparity in access time, which will cause lower system performance. So, the enhancement was mandatory. The enhancement of this was designed in the memory hierarchy model due to the systems performance increase.

- Ability: The ability of the memory hierarchy is the total amount of data the memory can store. Because whenever we shift from top to bottom inside the memory hierarchy model due to the system performance increase.

- Access Time: The access time in the memory hierarchy is the interval of the time among the data availability as well as request to read or write. Because whenever we shift from top to bottom inside the memory hierarchy, then the access time will increase.

- Cost Per Bit: When we shift from bottom to top in memory hierarchy, then the cost for each bit will increase which means an internal memory is expensive compared to external memory.

Memory Hierarchy Design

The memory hierarchy in computers mainly include the following.

- Registers: Usually, the registers is a SRAM in the processor of the computer which is used for holding the data word which is typically 64 or 128 bits. The program counter register is the most important as well as found in all processors. Most of the processor use a status word register as well as an accumulator. Usually, computers like complex instruction set computers have so many registers for accepting main memory.

- Cache Memory: It can also be found in the processor, however rarely it may be another IC which is seperated into levels. The cache holds the chunk of data which are frequently used from main memory.

- Main Memory: The main memory in the computer is nothing but, the memory unit in the CPU that communicates directly. It is the main storage of the computer. Its made of both RAM & ROM.

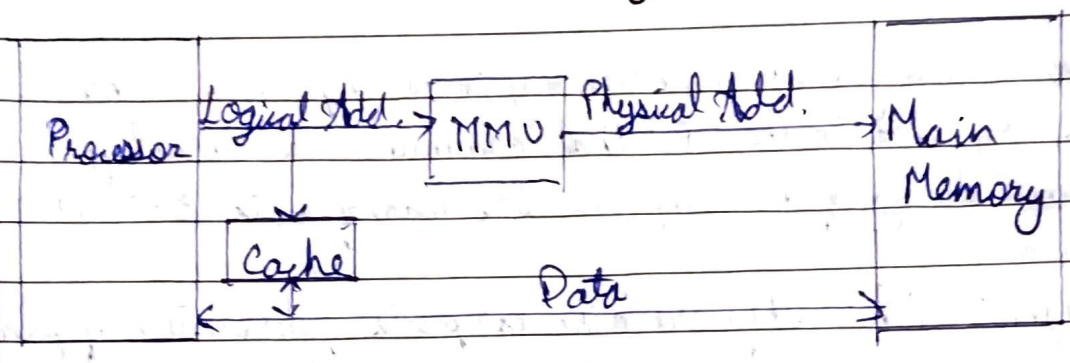
- Magnetic Disks: They are circular plates fabricated of plastic otherwise metal by magnetized material. The tracks in the computer are nothing but bits which are stored within the magnetized plane in spots next to concentric circles. These are usually seperated into sections named as sectors.

- Magnetic Tape: Its a normal ~~tape~~ magnetic recording tape designed with a slender magnitizable covering on an extendable plastic film of thin strip.

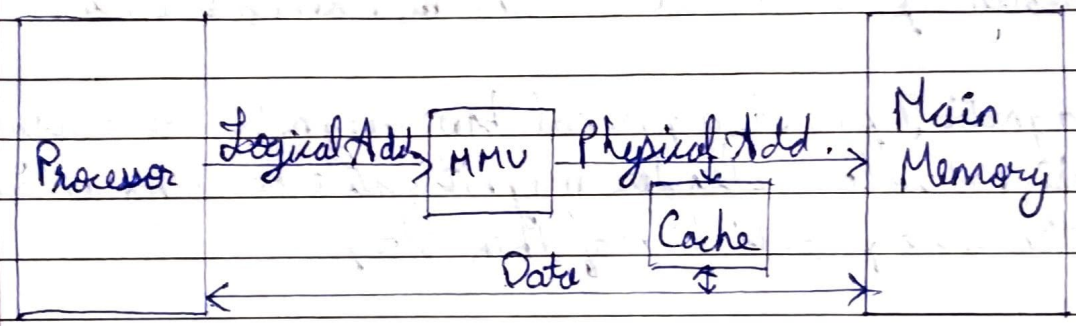
Q2. Elements of Cache Memory Design are:

1. Addressing: Its of 2 types:-

- Logical: Cache is placed between processor & MMU. It stores data using virtual memory address.




- Physical: Cache is placed between MMU & main memory. Stores data using physical address




2. Cache size: The size of cache has such that the avg cost/bit is close to main memory and avg. access time is close to that of cache.

3. Mapping Function: There are 3 types of mapping:

- Direct Mapping: Cache blocks of main memory maps to only one possible cache line. Its easily implemented using main memory address.

- Associative Mapping: The main memory block can hold into any line of cache and the CCL memory address is interpreted simply as tag & word. Tag identified a block of memory.

- Set Associative Mapping: It has strength of both direct & associative. Each word maps into all a cache lines in a specific set.

4. Replacement Algorithm
- Direct: Each block only maps to one line & no choice is possible so the line is replaced entirely.

- Associative & Set Associative: There are 4 commonly used algos:
  - LRU
  - LFU
  - FIFO
  - Random.

5. Block Size

6. No. of caches.

Q3. Cache memory is a very high speed memory used to speed up & synchronize with high speed of CPU. Acts as a buffer between RAM & CPU.
CPU cache is devided into 3 levels:
- L1: Fastest memory present in computer system. L1 is smallest memory.
- L2: Slower & larger than L1 & L3
- L3: Slower & larger than L2.

Q4. There are 3 types of mapping:

- Direct Mapping : A particular block of main memory can map only to a particular line of cache.

- Associative Mapping: A block of main memory can map to any line of the cache that is freely available at that moment. More flexible than direct mapping. Here, all line are freely available and thus, any block of main memory can map to any line of cache.

- Set-Associative Mapping : Line are grouped into sets where each set contains k-no. of lines. A particular block of main memory can link to only one set of cache. However, within a set, a block can map only any line that is freely available.

Q5. For associative mapping, if all the cache line have been occupied, then the existing blocks will have to be replaced. The commonly used algorithms are:

- LRU : It is most effective algorithm. It replaces the block which has been in cache for the longest time.

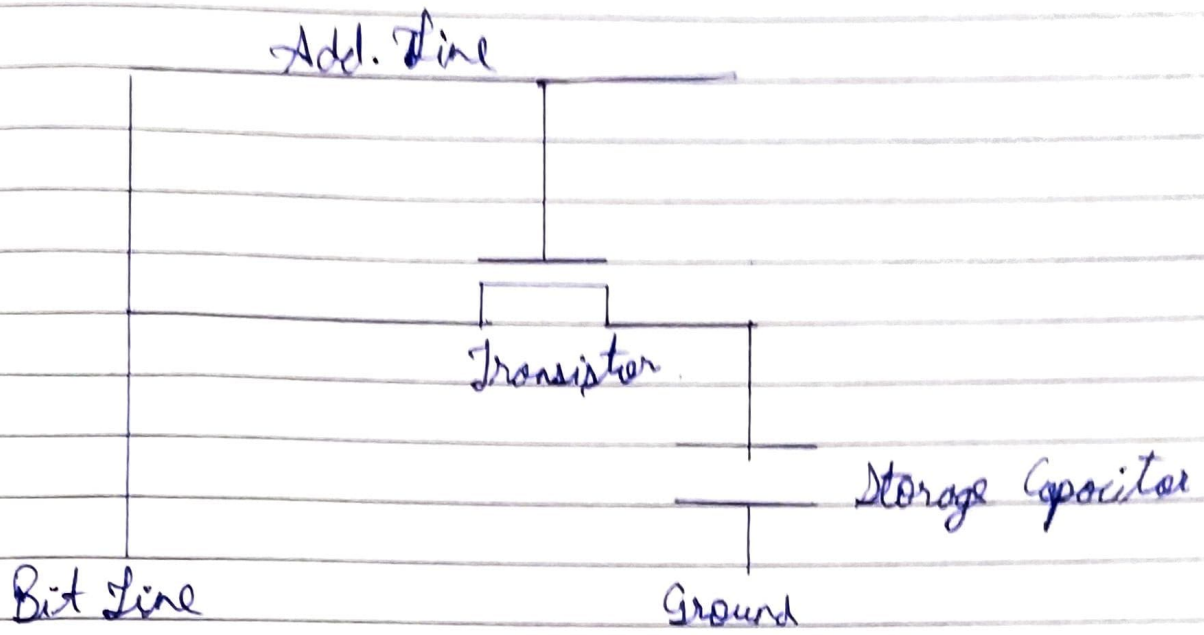
- FIFO: Replaces the block which has been in cache for longest time.

- LFU: Replaces block used for least time.

- Random : Block is randomly selected & a new block is replaced in cache.

Q6. Dynamic Ram (DRAM)

In D-Ram, bits are stored as charge in capacitors. It needs regular refreshing even when powered since charge leaks




Static Ram (SRAM)

Bits are stored as on/off switches. There is no leakage of charges and no refresh is needed when powered.