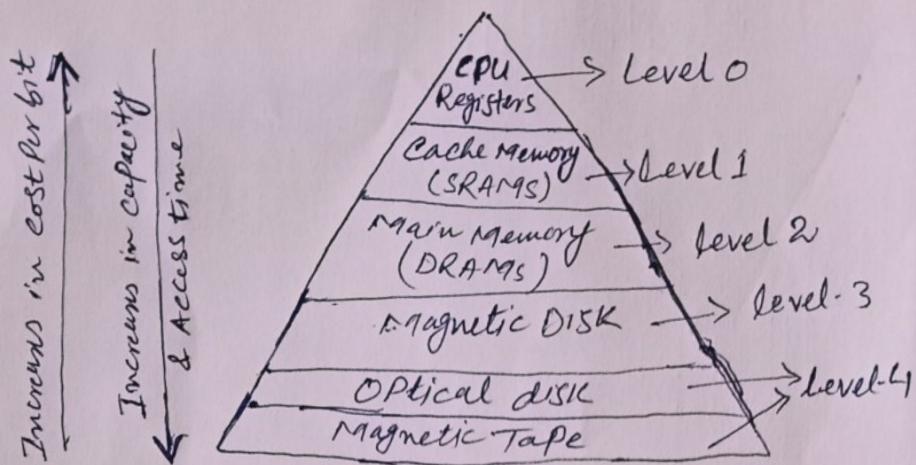


Memory Hierarchy

For economical as well as technical consideration, a computer uses several memory devices of different technology as shown in figure below



General purpose registers of CPU are also storage device. They hold data temporarily while the CPU executes instructions.

We can infer the following characteristics of Memory hierarchy design from above figure:

1. Capacity:- It is the global volume of information the memory can store. As we move from top to bottom in the hierarchy, the capacity increases.
2. Access Time:- It is the time interval between the read/write request and availability of the data. As we move from top to bottom in the hierarchy, the access time increases.
3. Cost per bit:

As we move from bottom to top in the hierarchy, the cost per bit increases i.e internal memory is costlier than external memory.

Real and Virtual Memory

The main memory actually available in a computer system is called real or physical memory. The term 'virtual memory' refers to something which appears to be present but actually it is not. The virtual memory technique allows users to use more memory for a program than the real memory of the computer. A programmer can write a program which requires more memory space than the capacity of main memory. Such a program is executed by virtual memory technique. The program is stored in the secondary memory. The memory management unit (MMU) transfers the currently needed part of the program from the secondary memory to the main memory for execution. This part of the program is executed by the processor. After execution, this part of the program is sent back to the secondary memory together with intermediate results. Thereafter, the CPU takes another part of the program for execution. Thus the main memory always keeps only currently needed part of the program. This type of 'to and fro' movement of instructions and data between main memory and secondary memory is called swapping. Thus, a program requiring more memory space than the capacity of the main memory can be executed using swapping technique. This concept is known as virtual memory technique. The memory space needed by the program is virtual memory.

Memory Management Unit (MMU):-

MMU is a hardware unit which is used in a multiuser, multiprogramming or multitasking system. It is placed in between the processor and the main memory. Modern microprocessor contains MMU on chip itself. In a multiuser system user should not interfere another user. This feature is provided by MMU. This ~~feature~~ feature is called protection.

Cache Performance :

When the processor needs to write or read a location in the main memory, it first checks for a corresponding entry in the cache.

- * If the processor finds that the memory location is in the cache, a "cache hit" has occurred and data is read from cache.
- * If the processor does not find the memory location in the cache, a "cache miss" has occurred. For a cache miss, the cache allocates a new entry and copies in data from main memory, then the request is fulfilled from the contents of the cache.

The performance of the cache memory is frequently measured in terms of a quantity called "Hit ratio".

$$\text{Hit ratio} = \frac{\text{hit}}{(\text{hit} + \text{miss})} = \frac{\text{no. of hits}}{\text{total accesses}}$$

We can improve cache performance using higher cache block size, higher associativity and reduce miss rate.