Memory mangement

Revya Naik V
Assistant Professor
Computer science & engineering dept
RGUKT Basar

Semiconductor main memory

- Semiconductor main memory is used to store program and data. The main memory(primary) is directly accessible by processor.
- To store digital information integrated circuit memory used like large integrated circuit memory, semiconductor memory, memory chip and transistor memory.
- The main memory two type RAM and ROM. Random access memory two types SRAM, DRAM. The read only memory are PROM, EPROM, EEPROM etc.

Semiconductor main memory

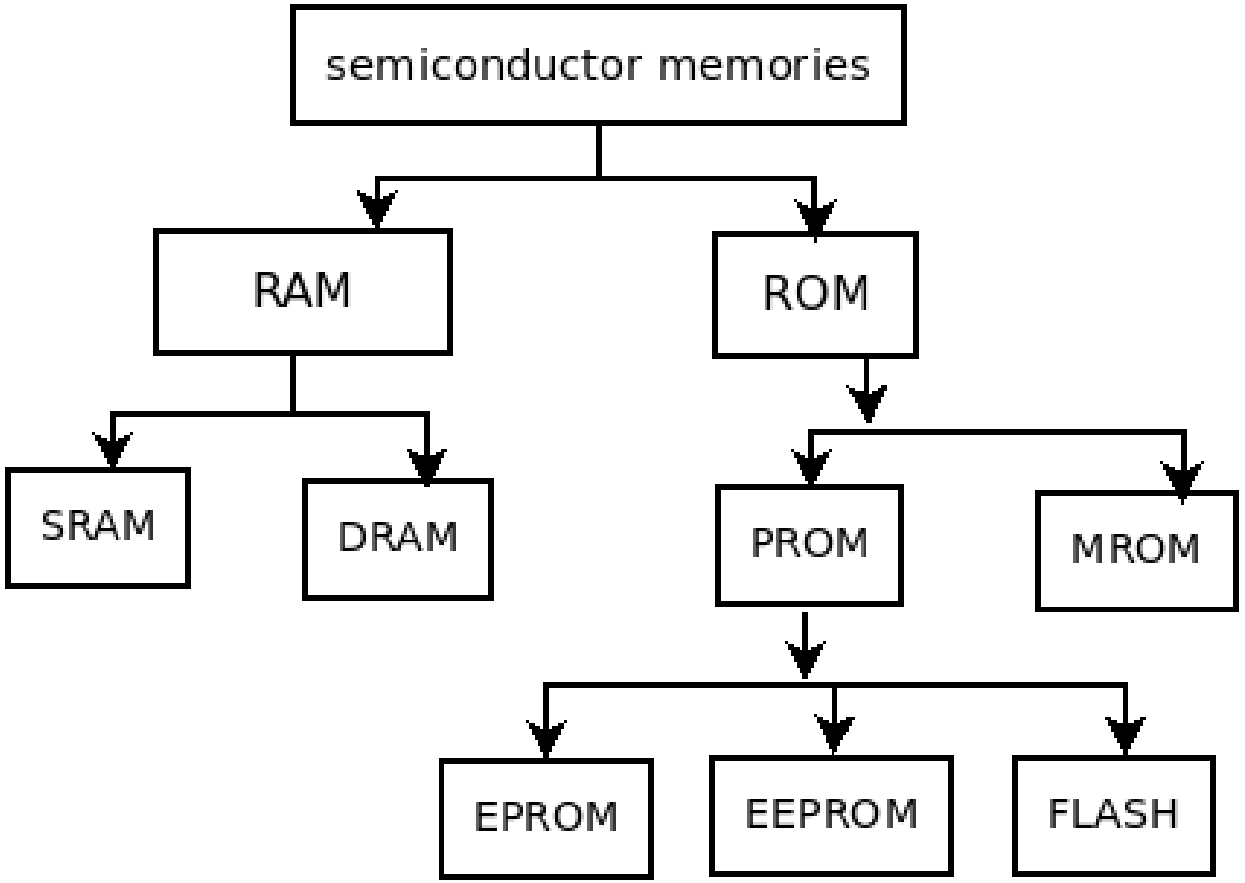


Figure 1 Semiconductor main memory

Random access memory(RAM)

- RAM used for reading and wirting data in any order (as it is required by the processor). Processor may access variavles on a randomly from memory.
- Dynamic RAM is a volatile memory. DRAM uses a capacitor to store each bit of data, and the level of charge on each capacitor determines wherther that bit is a logical 1 or 0, Dynamic refreshing.
- Static RAM is a volatile memory. In SRAM the data does not need to be refreshed dynamically. It also support memory read and memory write.

Read only memory

- ROM used where the data is written once and then not changed. It is a non-volatile memory and data stored permanently, even when the power is removed. Program stored in ROM like Basic Input-output system(BIOS).
- Programmable read only memory in which data can be write only once and. These memories re bought in a blank format and they are programmed using a special PROM programmer.
- EPROM memory can be programmed and then erased at a later time by ultravillet light.
- EEPROM: Data can be wrtten to it and it can be erased using an electirecal voltage.
- Flash memory data can be written to it and it can be erased. In erase memory data can be rewritten. Ex Memory cards.

Memory hierarchy

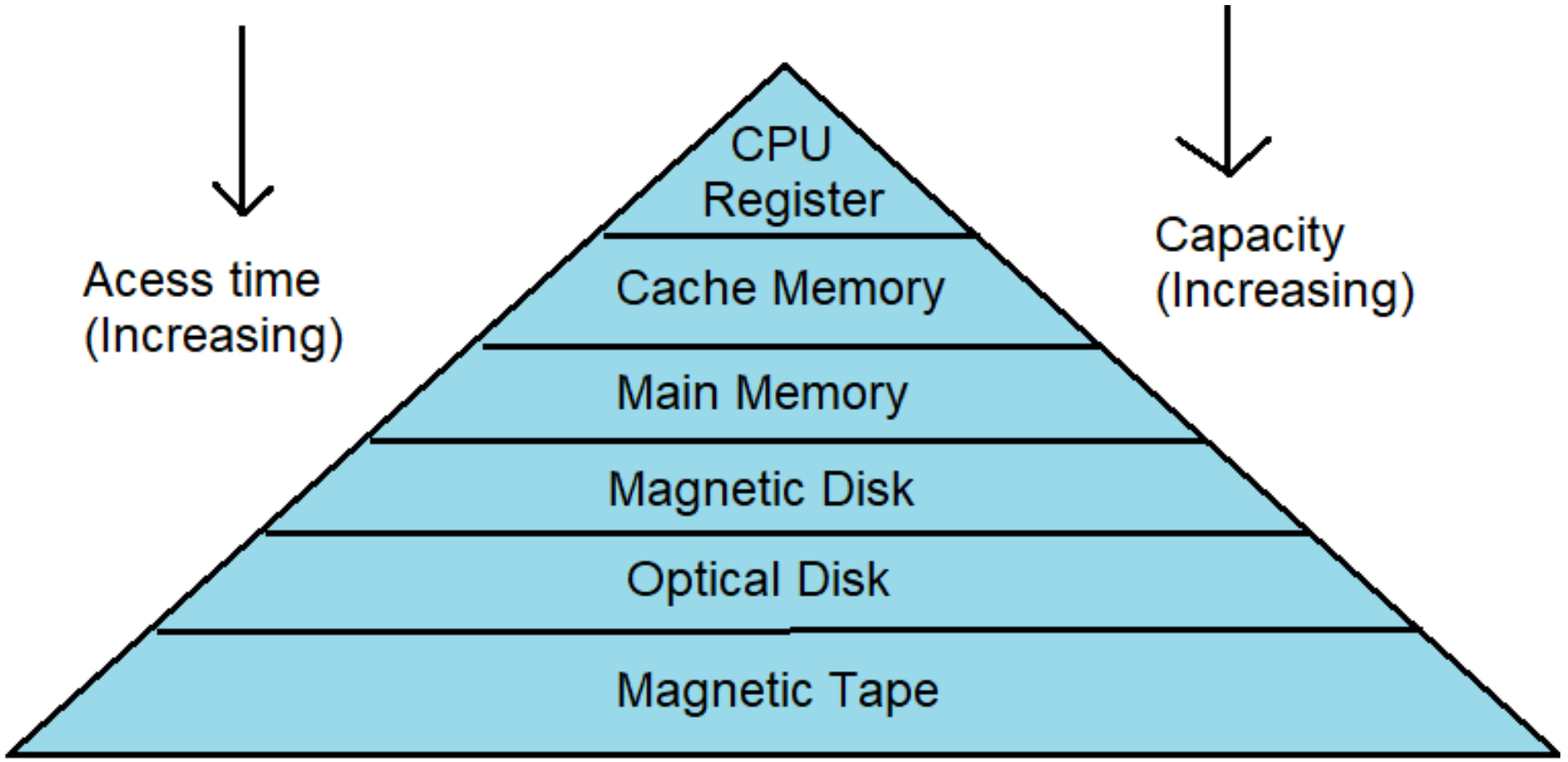
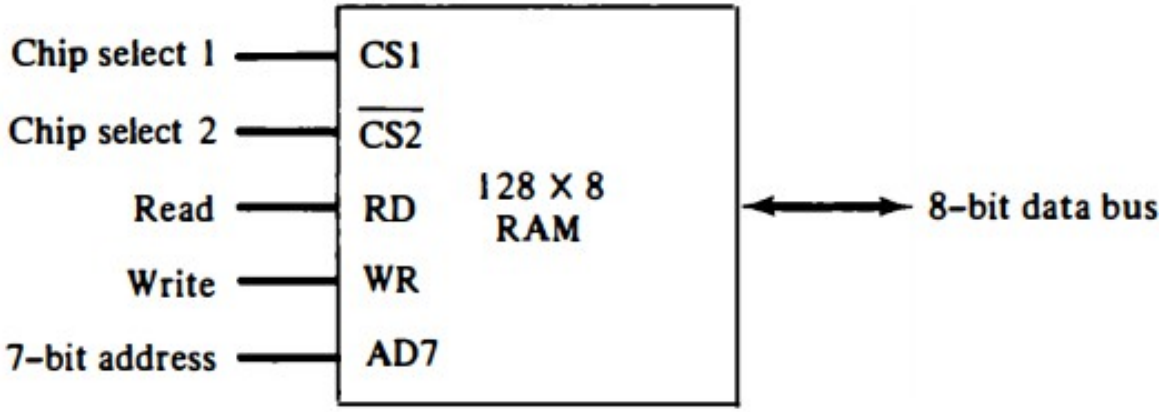


Fig:- Memory Hierarchy

RAM chip

Figure 2 Typical RAM chip.



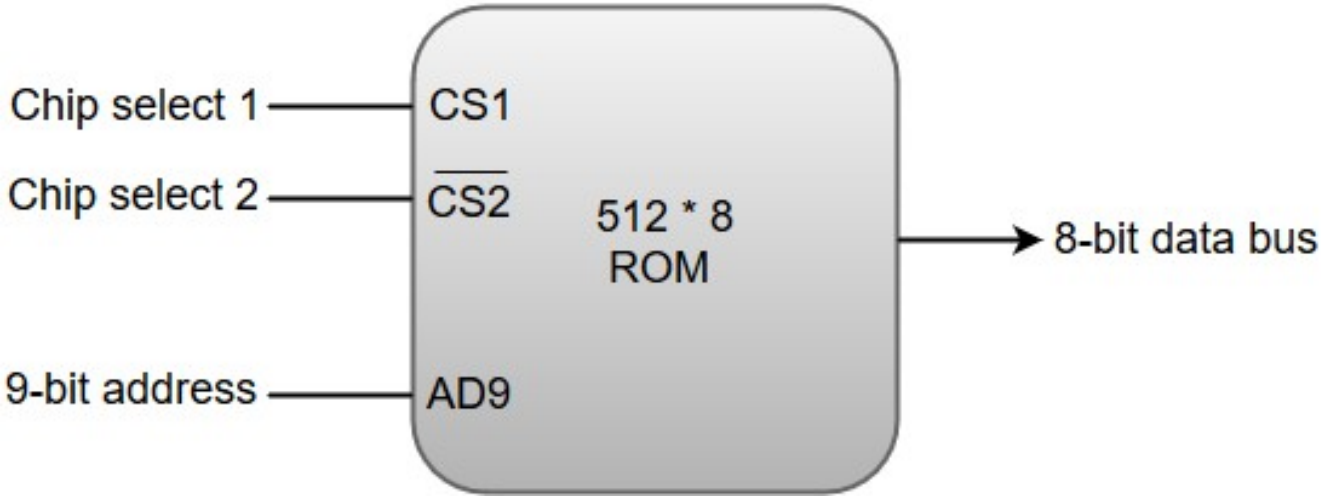
(a) Block diagram

CS1	$\overline{CS2}$	RD	WR	Memory function	State of data bus
0	0	×	×	Inhibit	High-impedance
0	1	×	×	Inhibit	High-impedance
1	0	0	0	Inhibit	High-impedance
1	0	0	1	Write	Input data to RAM
1	0	1	×	Read	Output data from RAM
1	1	×	×	Inhibit	High-impedance

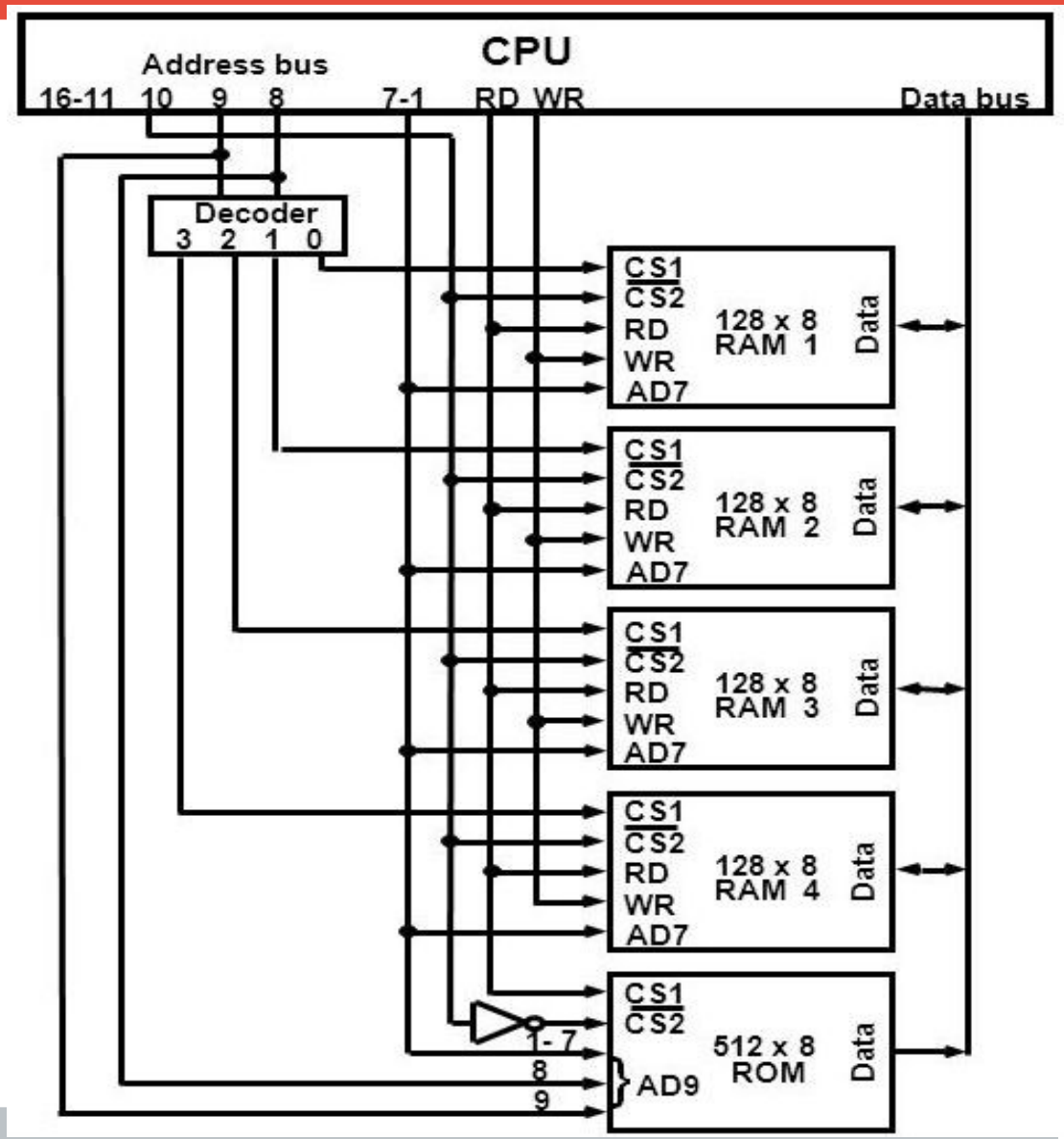
(b) Function table

ROM chip

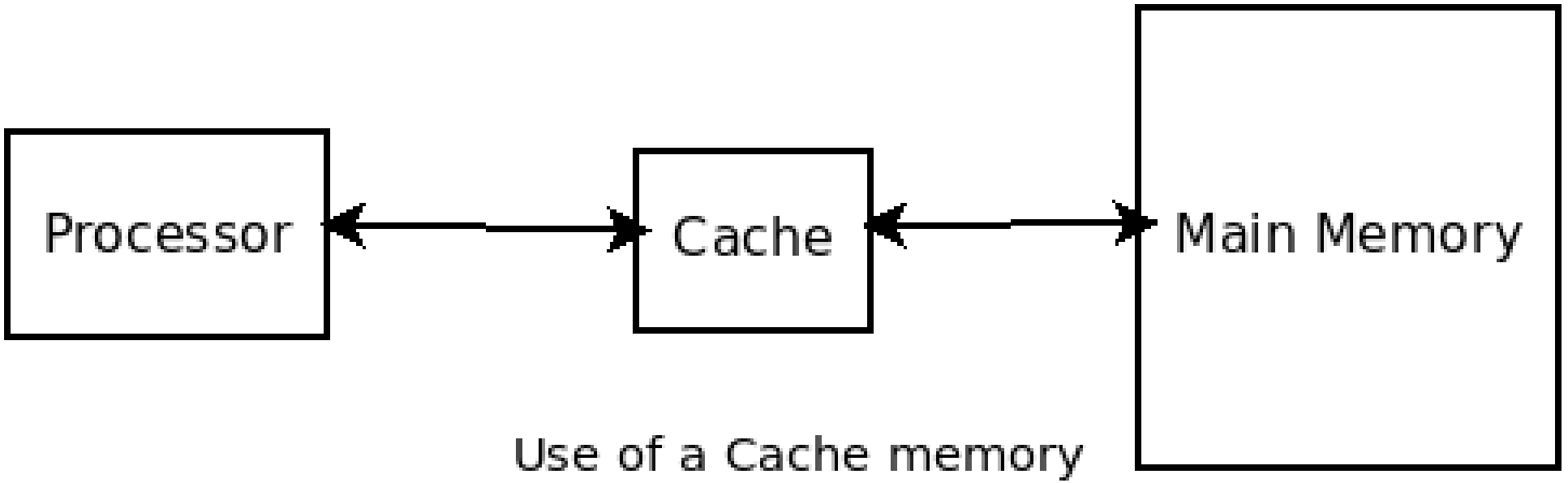
Typical ROM chip:



Memory connection to CPU



Cache memory



Use of a Cache memory

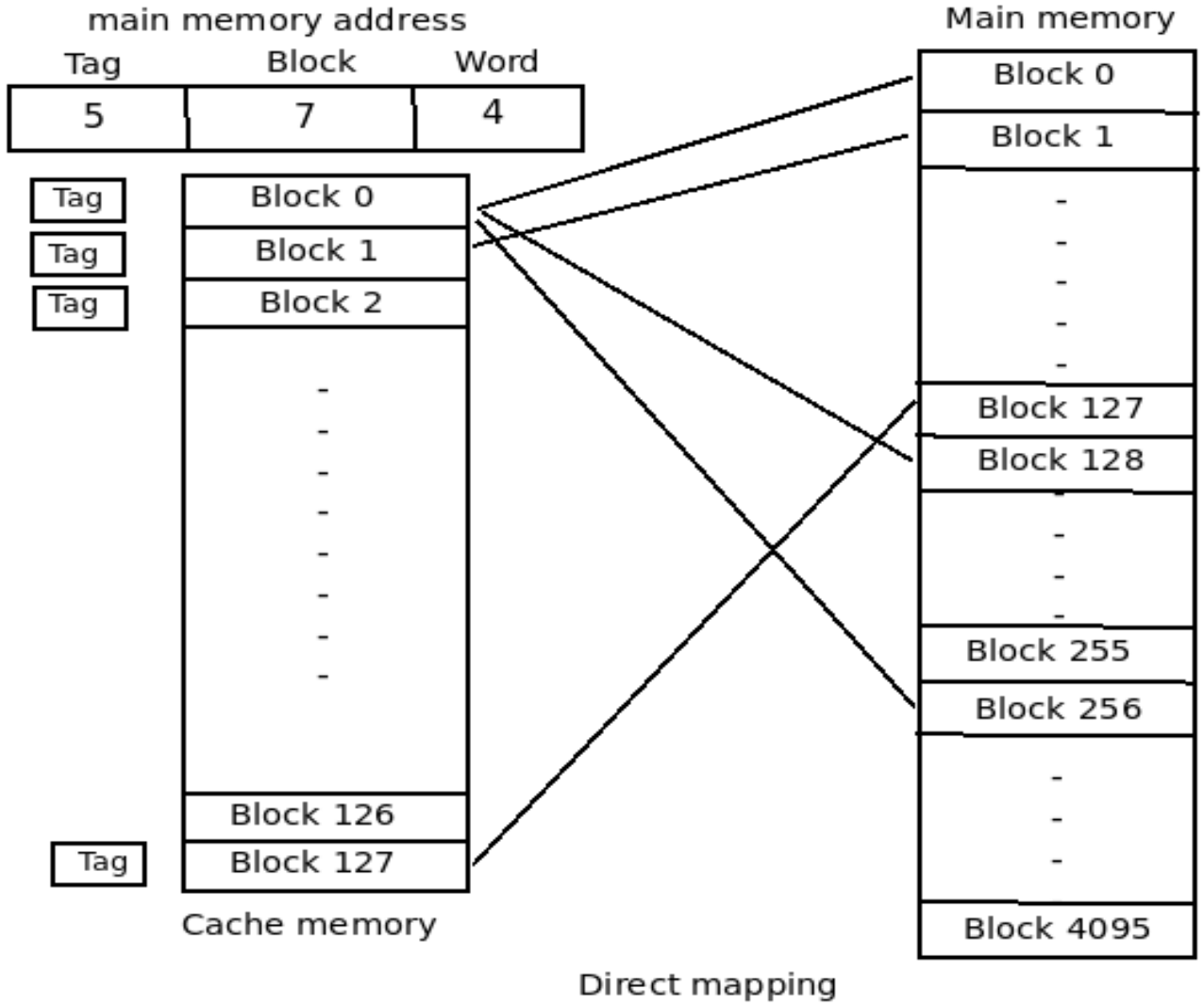
Cache memory

- Fast cache memory which essentially makes the main memory appear to the processor to be faster.
- Instructions like loop, nested loops or functions call repeatedly each other.many instructions in localized areas of the program are executed repeatedly during some time period. This is referred to as locality of reference.
- Write through protocal-cache location and main memory location are updated simultaneously.
- Update only cache location and to mark it as updatd with an associated flag bit, often called the dirty or modified bit.
- To free cache memory for new block, block is removed and copied in main memory is called write back or copy back.

Direct mapping function

- It is simplest way to determines cache locations in which to store memory blocks
- Block j of the main memory maps onto block j modulo 128 of the cache.
- Cache consisting of 128 blocks of 16 words each, total of 2048(2K) words.
- Main memory address bus size is 16 bits, has 64K words, which we will view of as 4K blocks of 16 words each.
- Main memory blocks 0, 128, 256,..... is loaded in the cache block 0, blocks 1,129,....are stored in cache block 1. It is not flexible.

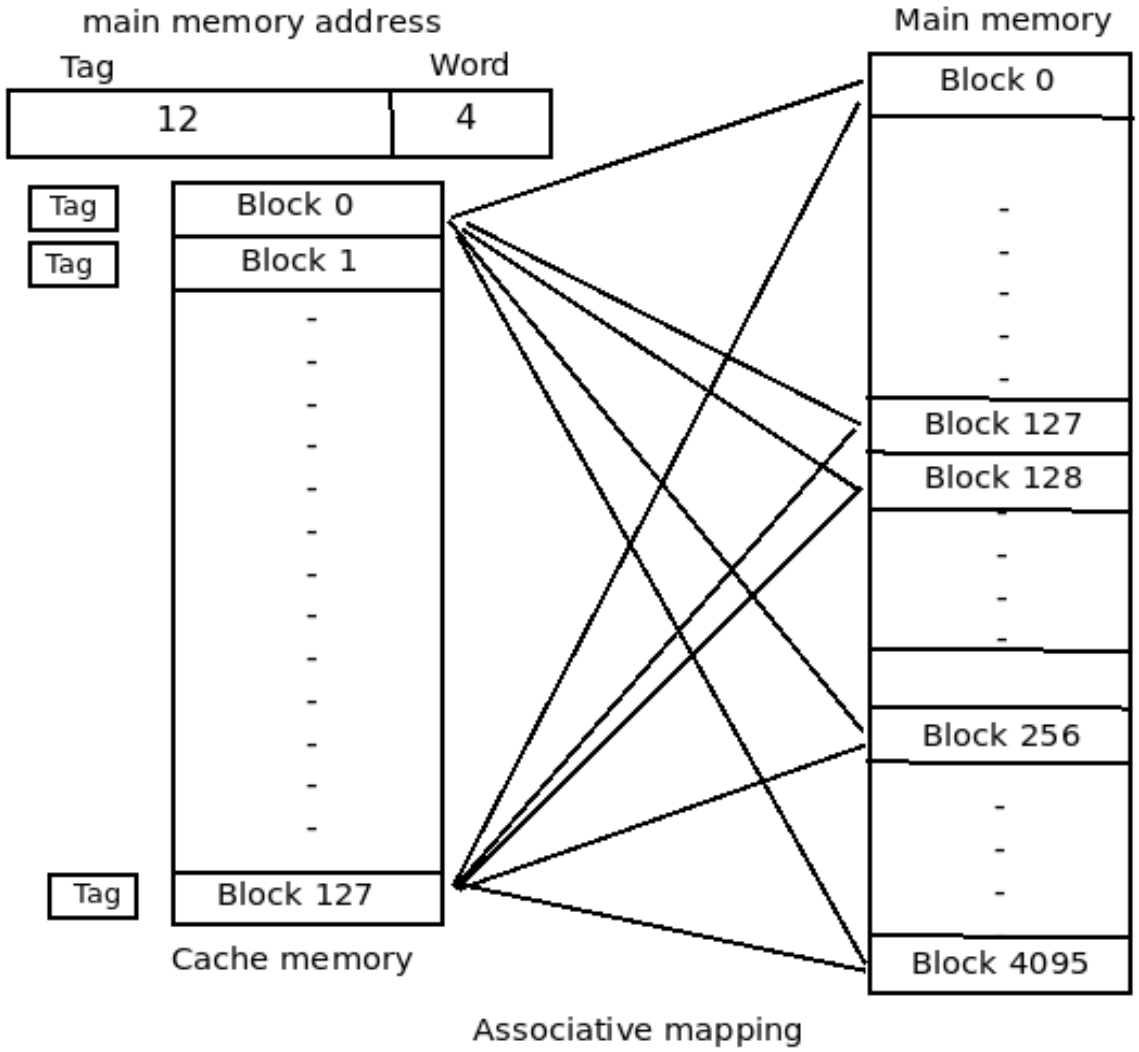
Direct mapping function



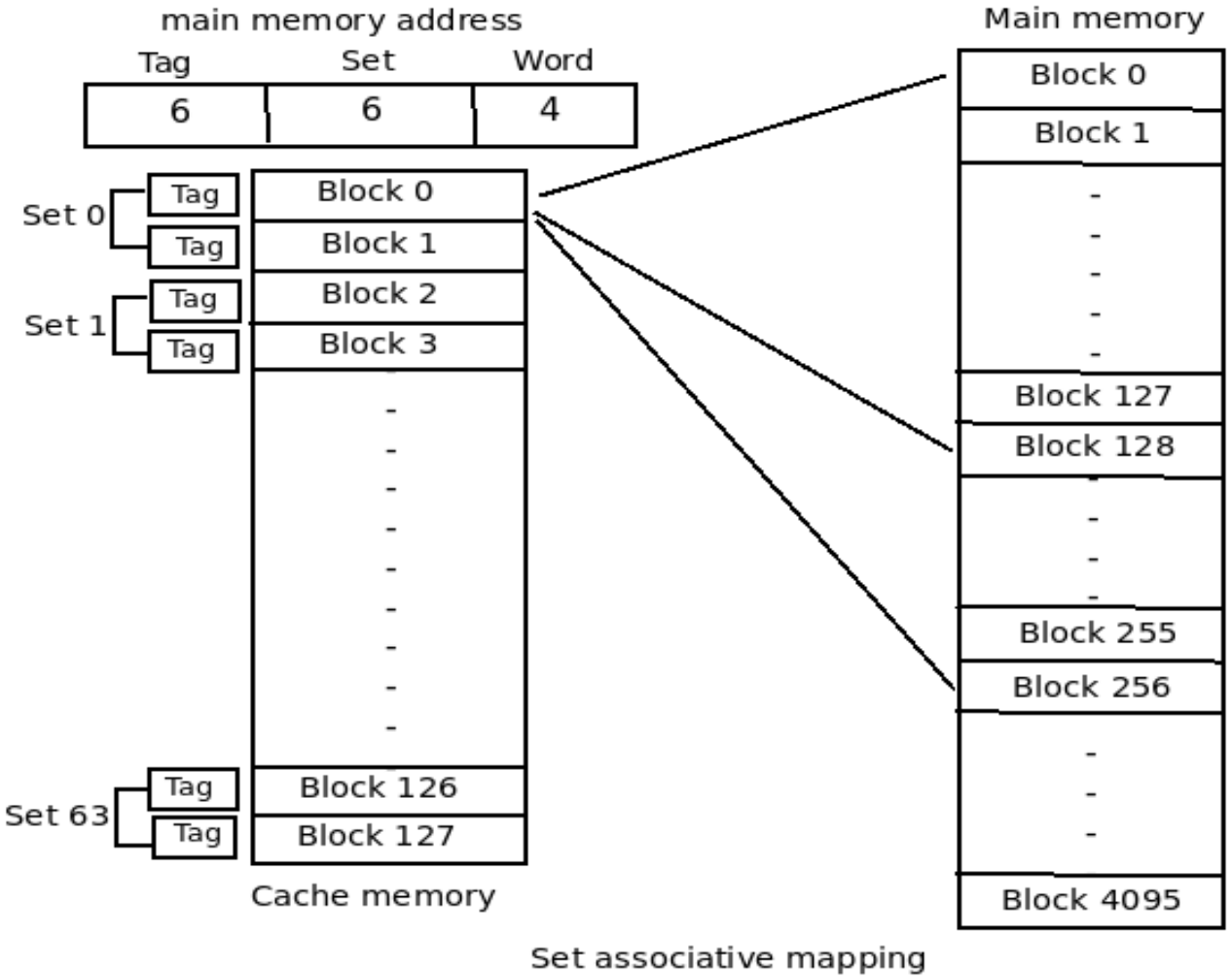
Associative & set associative mapping

- In Associative map main memory block can be placed in any cache memory block.
- The cost of an associative cache is higher than cost of a direct mapping because of the need to search all 128 tags
- Set associative mapping is combination of both direct and associative mapping.
- Memory blocks 0,64, ..., map into cache set 0.

Associative mapping



Set associative mapping



Set associative mapping

Replacement algorithms

- In direct mapping, the position of each block is predetermined; hence, no replacement strategy exists.
- In associative and set-associative mapping there exists some flexibility. When a new block is to be brought into the cache and all the positions are full, the cache controller must decide which old block to overwrite.
- In locality of reference property, programs usually stay in localized areas for reasonable periods of time, there is a high probability that the blocks that have been referenced recently will be referenced again soon.
- To overwrite the one that has gone the longest time without being referenced. This block is called the least recently used(LRU) block and technique is called LRU.
- Other replacement algorithms are FIFO and randomly choose the block to be overwritten.

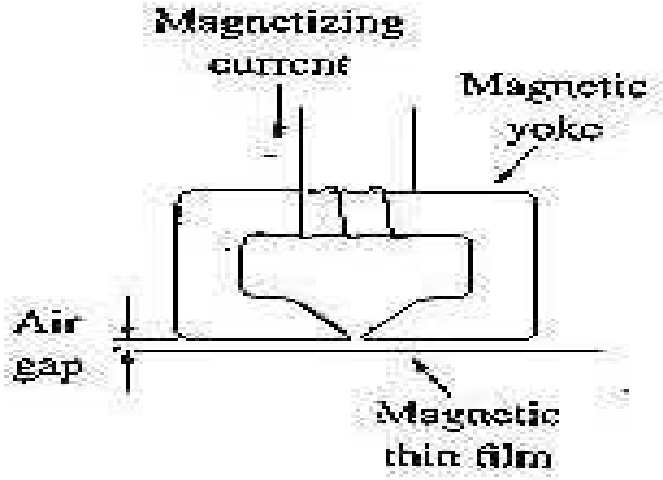
Auxilliary memory

- Main memory is a semiconductor memory and it is volatile. The devices need to store large information permanently are either magnetic or optical devices.
- Magnetic devices are magnetic disk(Hard disk), floppy disk, magnetic tape.
- Optical devices are CD-ROM, DVD etc.

Magnetic disk

Read/Write head

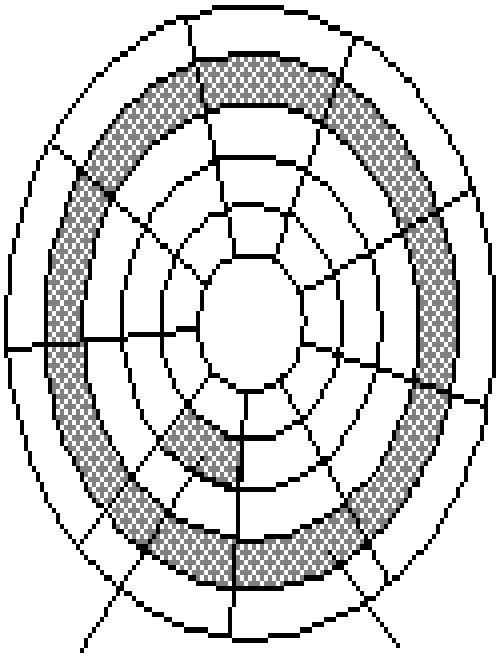
The head is a relatively small device capable of reading from or writing to a portion of the platter rotating beneath it.

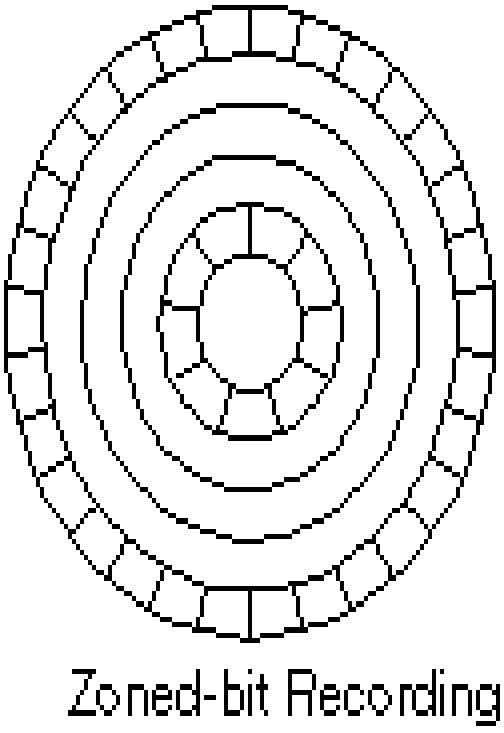


Magnetic disk

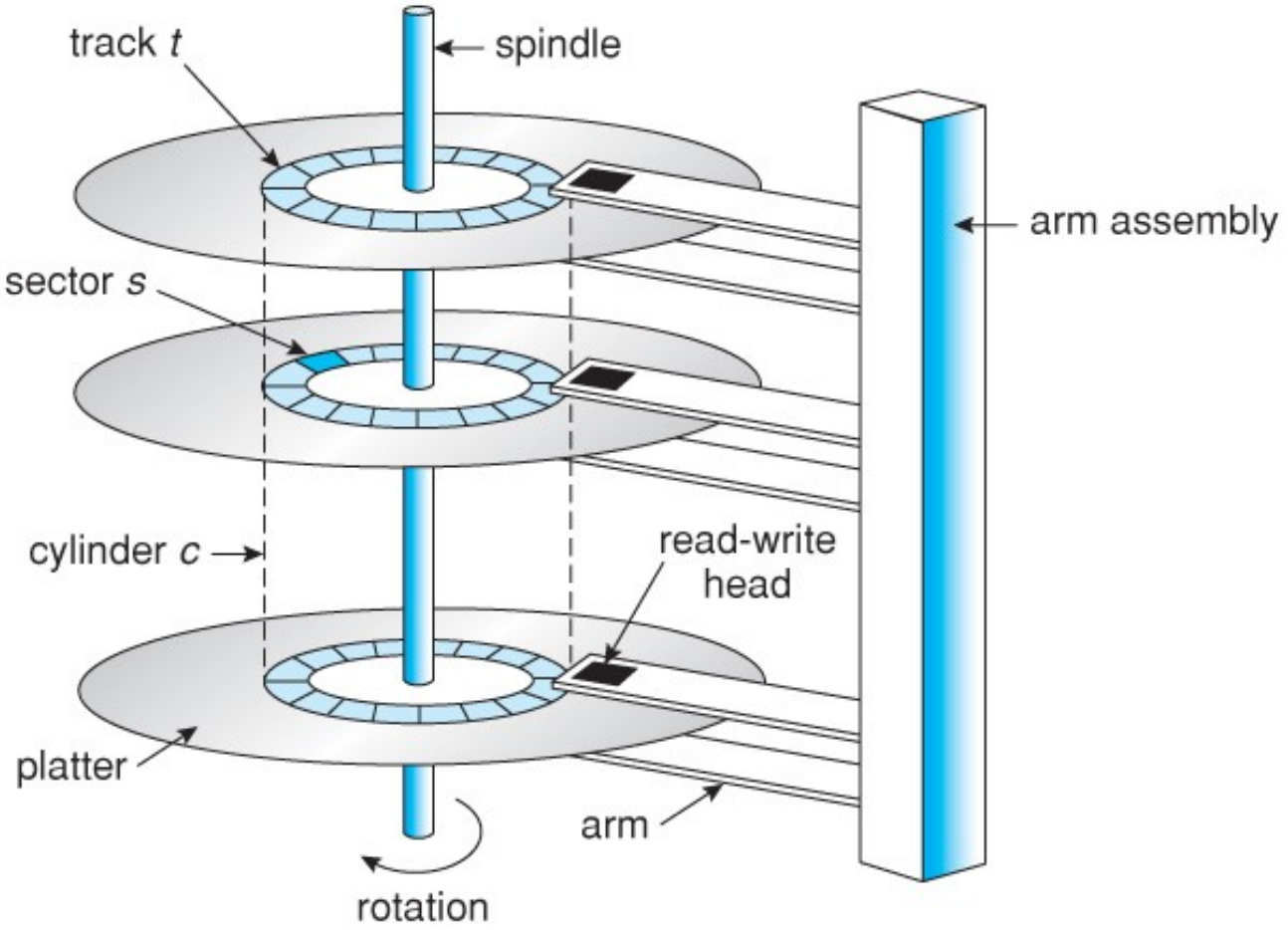
- The data on the disk are organized in a concentric set of rings called track. Each track has the same width as the head adjacent tracks are separated by track gaps.
- Data density is more in inner most track and data density is less in outer most track, so there is wastage of space in outer tracks.
- To increase the capacity the concept of zone is used instead of sectors. Each track is divided in zone of equal length and fix amount of data is stored in each zone. More no of zones in outer track, less in inner track.

Magnetic disks





Disk mechanism



Magnetic disk

- Seek time is time required to move the read/write head to the proper track.
- Once the track is selected, the disk controller waits until the appropriate sector rotates to line up with head. The time takes to reach the beginning of the desired sector is known as rotational delay or rotational latency
- Access time=seek time+rotational latency
- Trasfer time: transfer time to or from the disk depends on the rotational speed of disk and it estimated as T=b/rN
- T= transfer time, b is no of bytes to be transferred, N is no of bytes on a track, r=rotational speed in revolultion per second.
- .avg access time Ta=Ts+1/2r+b/rN
- Disk has 8 data recording surfaces with 4096 tracks per surface. Tracks are divided into 256 sectors and each sector contains 512 bytes recorded serially then total capacity.