GAYATRI VIDYA PARISHAD COLLEGE OF ENGINEERING (Autonomous)
Approved by AICTE, New Delhi and Affiliated to JNTU-Kakinada
Re-accredited by NAAC with "A" Grade with a CGPA of 3.47/4.00
Madhurawada, Visakhapatnam - 530 048.

Department of Computer Science and Engineering

COMPUTER ORGANIZATION
(19EC11D4)

Ms. N SANTOSHI
Assistant Professor
ECE Department

UNIT V
MEMORY SYSTEM DESIGN

Memory Unit

- An essential component in any general purpose computer since it is needed to store programs and data.
- memory unit that communicates directly with the CPU = main memory
- devices that provide backup storage = auxiliary memory.
- Auxiliary memory devices are used to store system programs, large data files and other backup information. Only programs and data currently needed by the processor reside in main memory. All other information is stored in main memory and transferred to main memory when needed.

Cache Memory

- Memory that lies in between main memory and CPU

  Holds those parts of the program and data that are most heavily used

- increases the overall processing speed of the computer by providing frequently required data to the CPU at a faster speed.

Main Memory

- Memory unit that communicates directly with CPU
- Programs and data currently needed by the processor reside here
- Also known as primary memory
- RAM and ROM

Auxiliary Memory

- Made of devices that provide backup storage
- Magnetic tapes, Magnetic disks
- At the bottom of the hierarchy are the relatively slow magnetic tapes used to store removable files whereas at the top level, magnetic disks used as backup storage

Memory Types

Sequential Access Memory

- A class of data storage device that read their data in sequence
- Are usually a form of magnetic memory
- Typically used for secondary storage in general-purpose computers due to their higher density, resistance to wear and non-volatility
- Eg: hard disk, CD-ROMs, magnetic tapes etc

Random Access Memory

- Is a form of computer data storage
- Allows stored data to be accessed in any order
- Associated with volatile types of memory
- Type: SRAM and DRAM

Memory Hierarchy

- To obtain the highest possible access speed while minimizing the total cost of the memory system
- Consists of all storage device in a computer system (auxiliary, cache, main , high speed registers and processing logic)

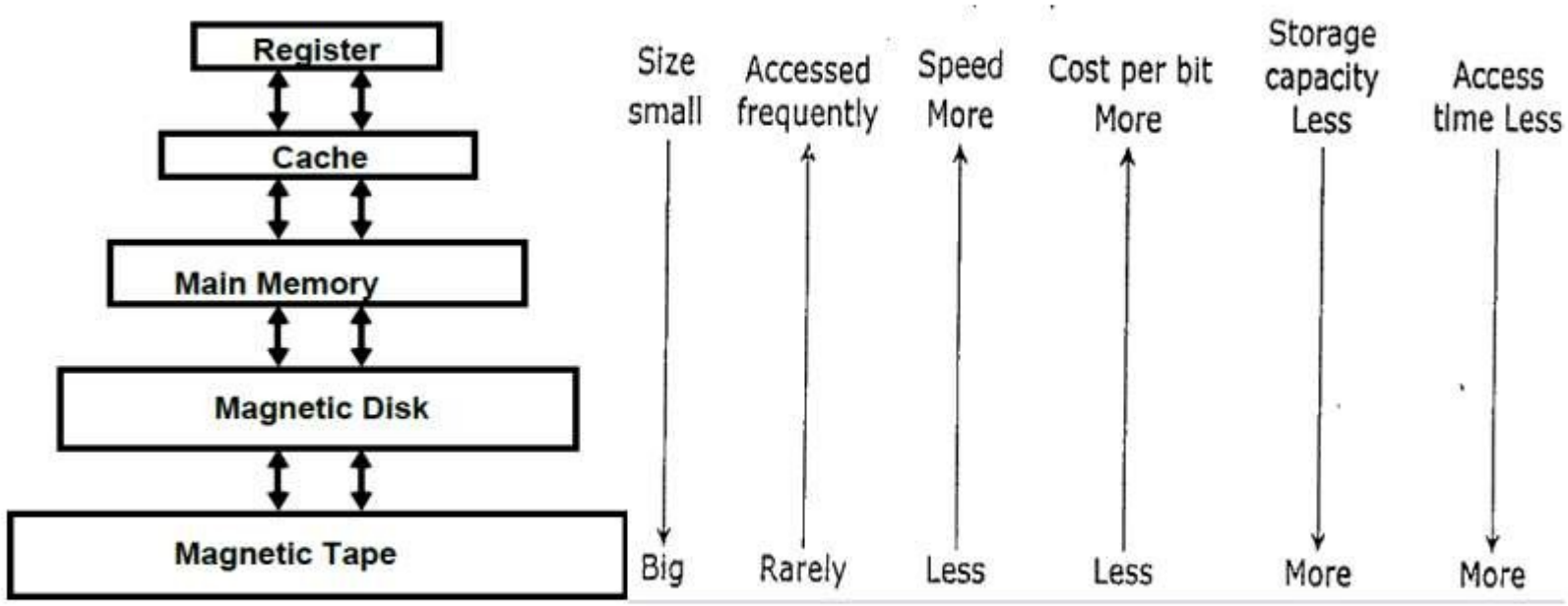
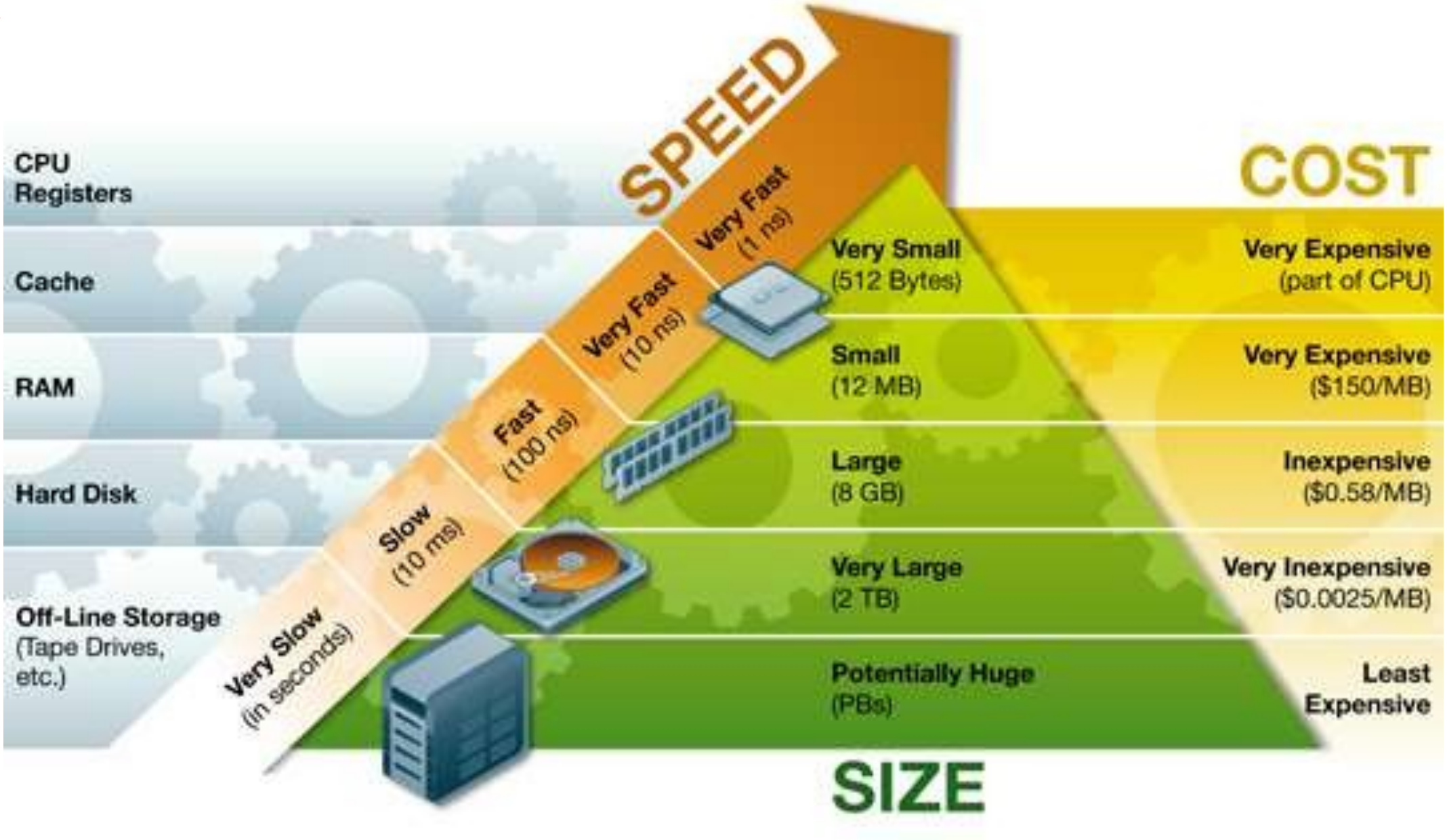


Figure: Memory Hierarchy


CPU
Registers
Cache
RAM
Hard Disk
Off-Line Storage
(Tape Drives,
etc.)
SPEED
Very Fast
(1 ns)
Very Fast
(10 ns)
Fast
(100 ns)
Slow
(10 ms)
Very Slow
(in seconds)
COST
Very Small
(512 Bytes)
Small
(12 MB)
Large
(8 GB)
Very Large
(2 TB)
Potentially Huge
(PBs)
Very Expensive
(part of CPU)
Very Expensive
($150/MB)
Inexpensive
($0.58/MB)
Very Inexpensive
($0.0025/MB)
Least
Expensive
SIZE

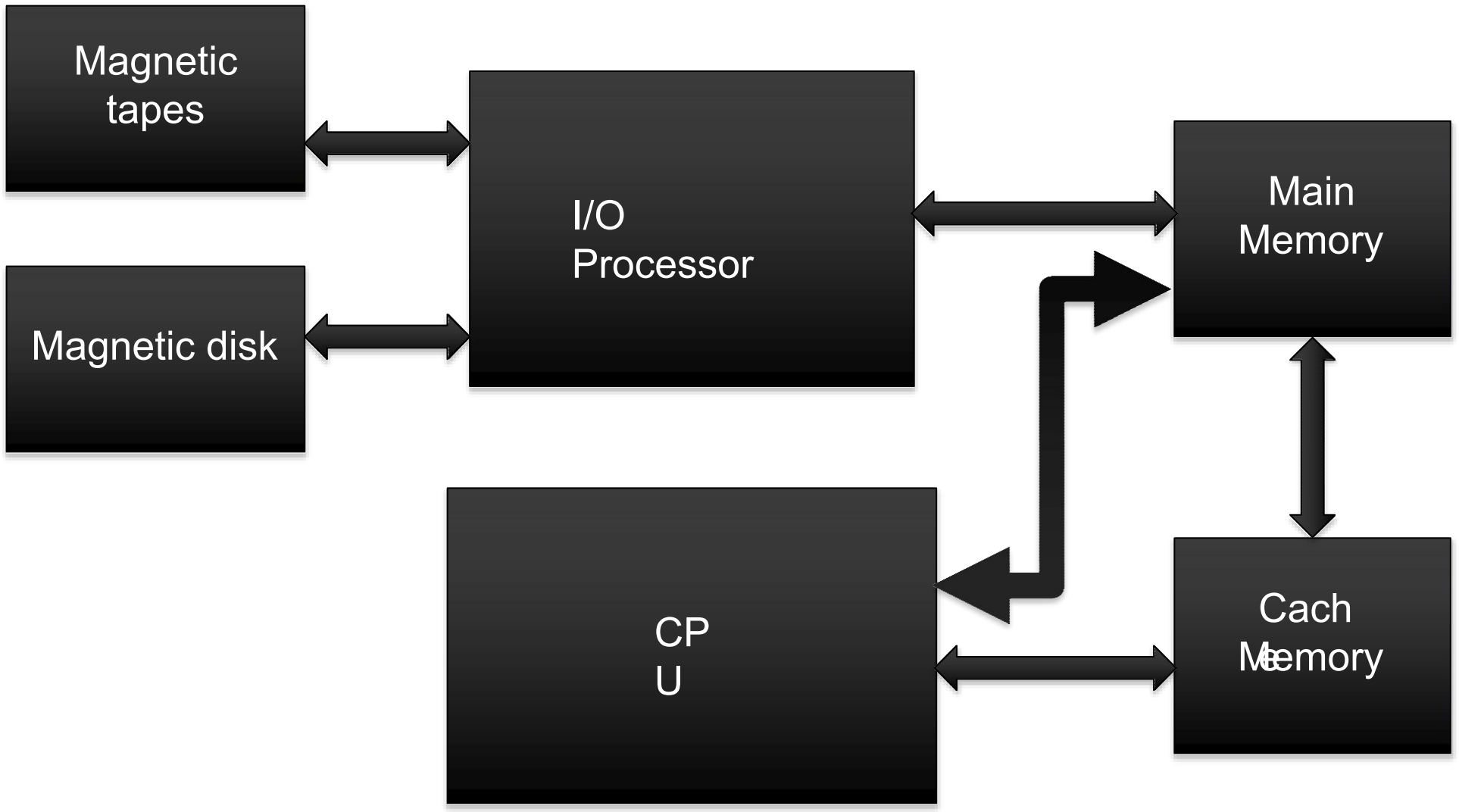


Fig: Memory Hierarchy in a Computer System

MAIN MEMORY

Main

- Basic memory of the computer
- Temporary memory except rom
- Faster for read write operation
- Expensive internal memory so not portable.

RAM

- Volatile memory.
- Stores information required during processing

Two types of random Static Ram(SRAM) and Dynamic Ram(DRAM)

TYPES OF RAM

SRAM

- Does not lose its content until computer is turned off
- Information is stored in form of voltage
- faster

DRAM

- Loses its content after few seconds
- Information is stored in the form of charge
- slower

ROM

Non volatile
Bootstrap
loader

- Permanent memory
- Stores information required for computer operations
- Types of rom(PROM,EPROM,EEPROM)

Types of ROM

- PROM:
  - Programmable ROM
  - User can store programs only once.
  - User can make micro code program can be made that are needed mostly.
  - The process of making program in PROM is called 'Burning'.
  - Example: CD-R
- EPROM:
  - Erasable PROM
  - Information can be removed by ultra violet rays.
  - Information can be re-write after removing previous information.
  - It is cheaper than PROM because it is re-useable.
  - Example: CD(RW)
- EEPROM:
  - Electrically Erasable PROM
  - Information can be removed by electric signals.
  - It is the simplest way to store info in ROM.
  - Now it is used to store BIOS in Memory.
  - Example :Pen Drive

Cache Memory: is a semi-conductor (buffer)memory which lies between main memory and processor. It is mainly used for increasing the speed of RAM.

Boot Strap loader

- Initial program whose function is to start the computer operating system after the power is turned on. and it is stored in the rom portion of the main memory .
- Computer start up :starting the execution of initial program after computer is turned on
- Boot strap loader loads the portion of disk to main memory and control is then transferred to os.

Auxiliary memory (SECONDARY MEMORY)

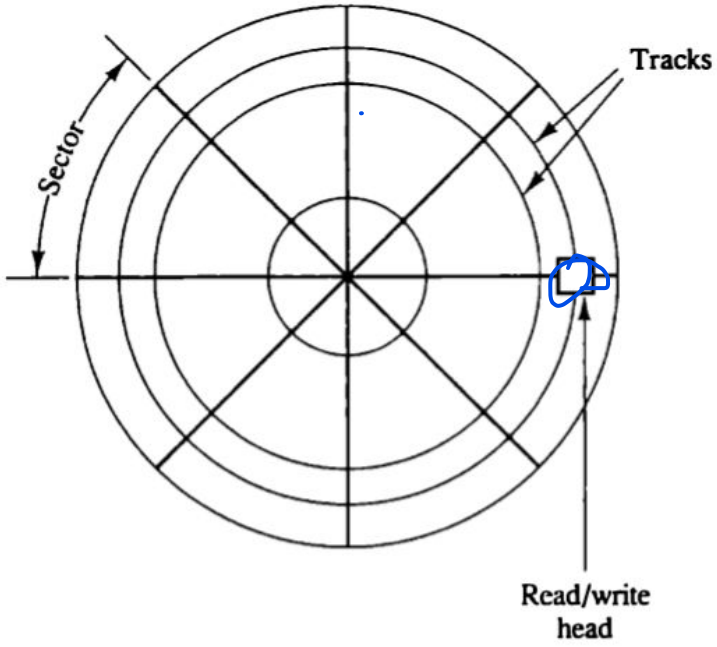
- The most common auxilliary device used in the computer system is magnetic disk and magnetic tape.
- Store large amount of data permanently.
- Portable
- The important characteristics of any device are its access mode, access time, transfer rate, capacity, and cost.

Types of auxilliary memory

- Magnetic disk
- Magnetic tape

Magnetic disk

- circular plate ,made of metal or plastic coated with magnetized material.High speed of rotation
- Bits are store in a concentric circle called tracks.
- Division of tracks are called sectors.



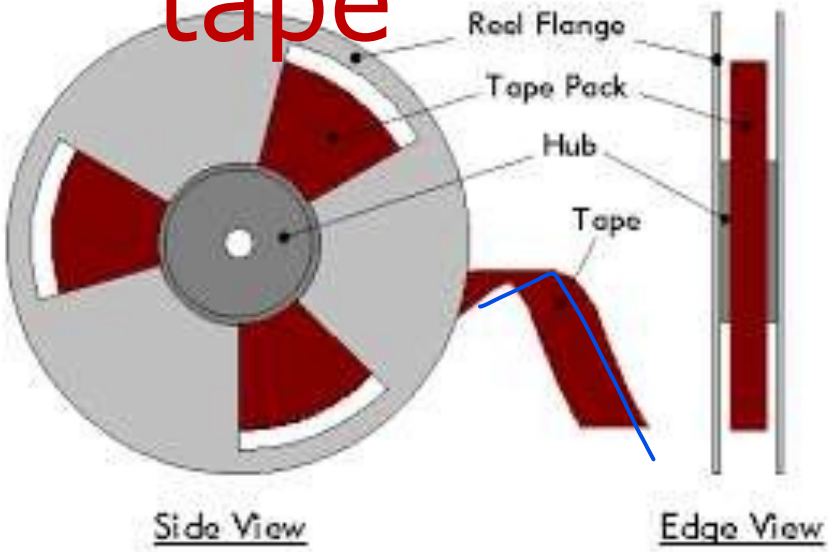
Magnetic disk

- A track in a given sector near the circumference is longer than a track near the center of the disk.
- If bits are recorded with equal density, some tracks will contain more recorded bits than others.
- To make all the records in a sector of equal length, some disks use a variable recording density with higher density on tracks near the center than on tracks near the circumference. This equalizes the number of bits on all tracks of a given sector.

Magnetic tape

- Sequential access memory used for storing,backup,audio,video data etc .
- Highly reliable memory.
- Slower for read write operation.
- It is a strip of plastic coated with a magnetic recording medium.
- Bits are recorded as magnetic spots on the tape along several tracks.
- Magnetic tape units can be stopped, started to move forward or in reverse, or can be rewound.

Magnetic tape



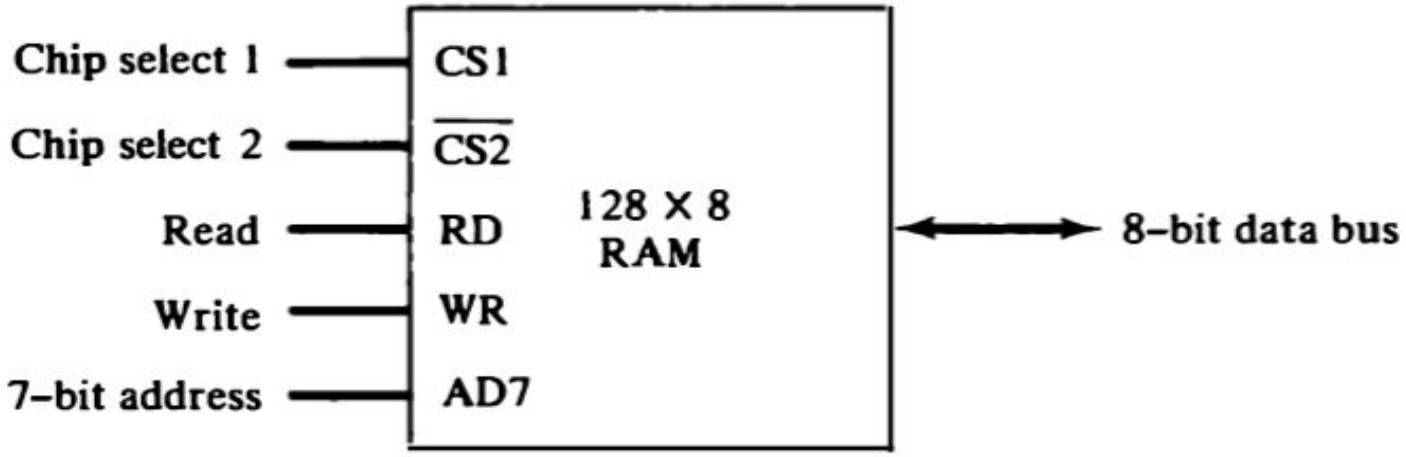
RAM AND ROM CHIP

RAM AND ROM CHIPS

RAM chip: The capacity of the memory is 128 words of eight bits (one byte) per word.

- requires 7-bit address and an 8-bit bidirectional data bus.
- chip select (CS) are for enabling the chip.

Fig. RAM chip



(a) Block diagram

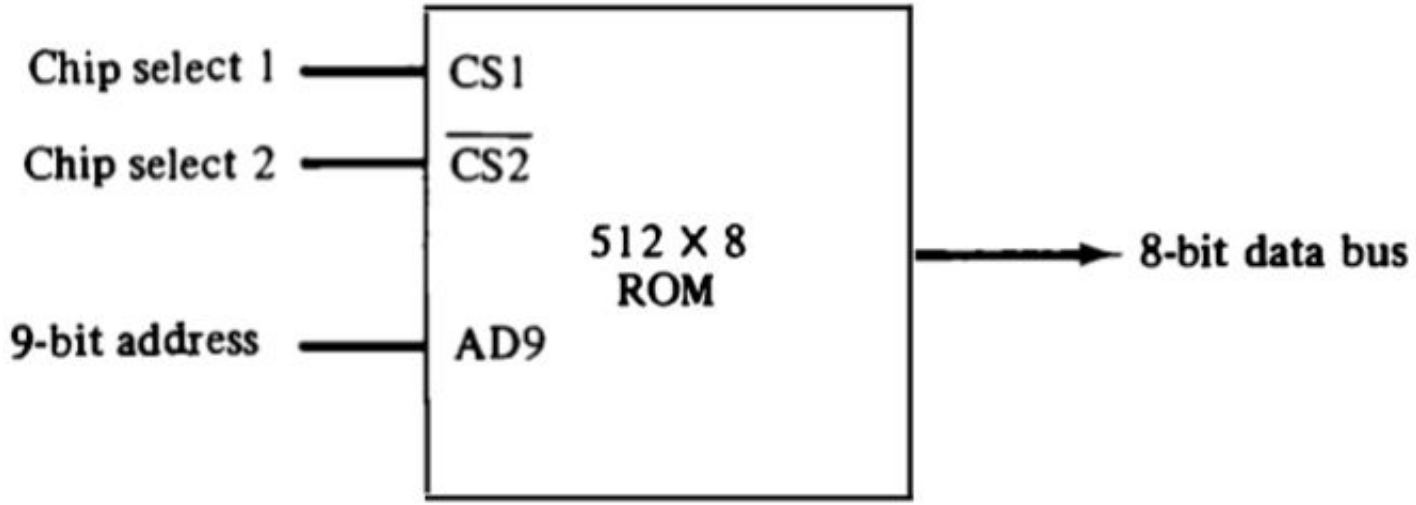
CS1	$\overline{CS2}$	RD	WR	Memory function	State of data bus
0	0	×	×	Inhibit	High-impedance
0	1	×	×	Inhibit	High-impedance
1	0	0	0	Inhibit	High-impedance
1	0	0	1	Write	Input data to RAM
1	0	1	×	Read	Output data from RAM
1	1	×	×	Inhibit	High-impedance

(b) Function table

- when CS1=1 and (CS2)'=0, the unit in operation.
- High impedance state indicates open circuit.
- When CS1=1 and (CS2)'=0, the memory is places in a R/W mode.
- When the RD input is enabled, the content of the selected byte is placed into the data bus.

ROM chip

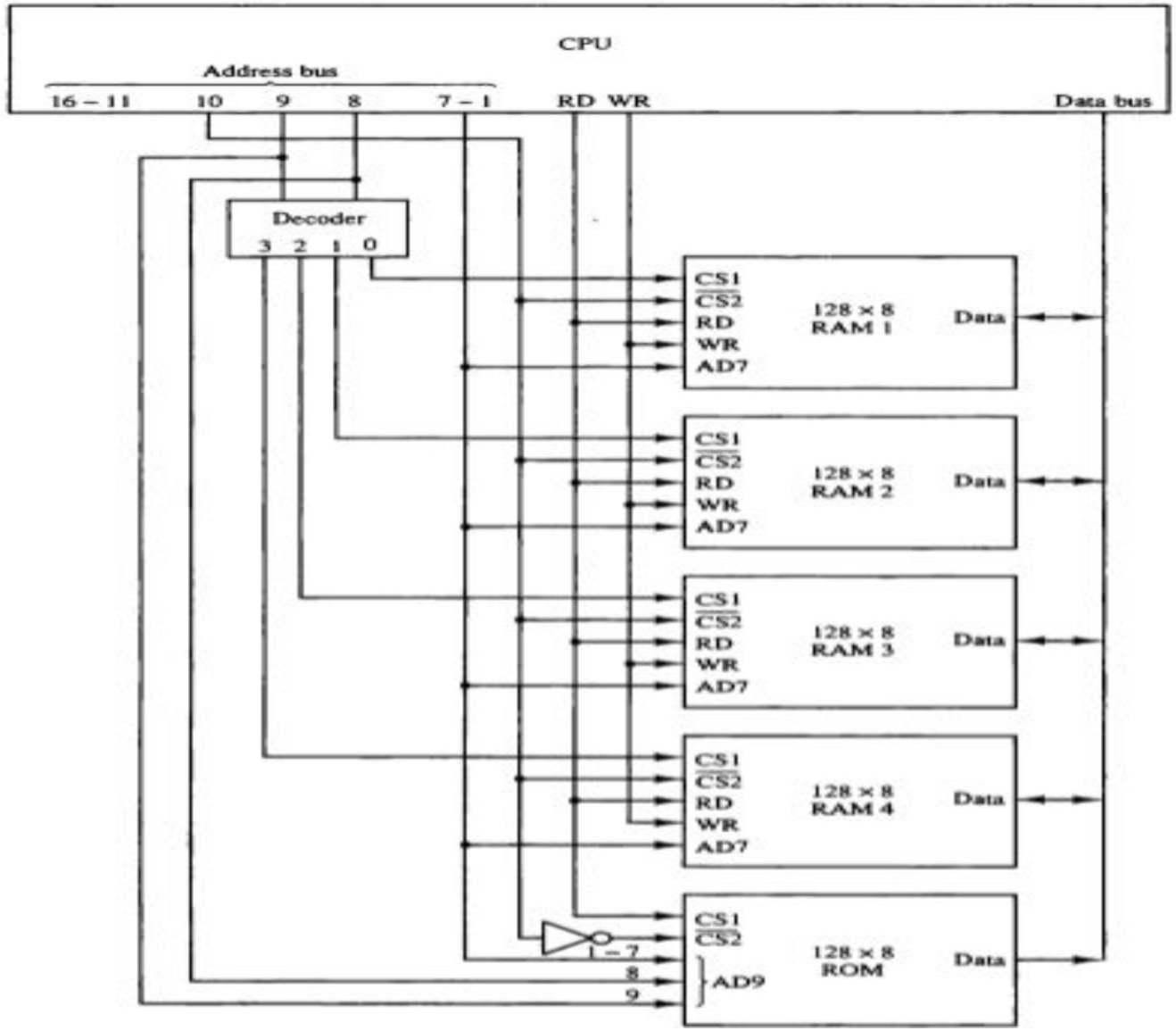
- ROM SIZE : 512 Bytes
- Requires 9 address lines



Memory address map for microcomputer

- To demonstrate with a particular example, assume that a computer system needs 512 bytes of RAM and 512 bytes of ROM.
- The RAM and ROM chips are of size 128 bytes and 512 bytes respectively.

Component	Hexadecimal address	Address bus									
		10	9	8	7	6	5	4	3	2	1
RAM 1	0000–007F	0	0	0	x	x	x	x	x	x	x
RAM 2	0080–00FF	0	0	1	x	x	x	x	x	x	x
RAM 3	0100–017F	0	1	0	x	x	x	x	x	x	x
RAM 4	0180–01FF	0	1	1	x	x	x	x	x	x	x
ROM	0200–03FF	1	x	x	x	x	x	x	x	x	x


CPU
Address bus
16 – 11
10
9
8
7 – 1
RD
WR
Data bus
Decoder
3
2
1
0
CS1
CS2
RD
WR
AD7
128 × 8
RAM 1
Data
CS1
CS2
RD
WR
AD7
128 × 8
RAM 2
Data
CS1
CS2
RD
WR
AD7
128 × 8
RAM 3
Data
CS1
CS2
RD
WR
AD7
128 × 8
RAM 4
Data
CS1
CS2
1 – 7
8
9
AD9
128 × 8
ROM
Data

- A ROM chip is unidirectional.
- 9 address lines to address 512 bytes.
- chip select CS1=1 and (CS2)'=0 for the unit to operate.
- Otherwise, the data bus in a high-impedance state.

VIRTUAL MEMORY

- Attempts to optimize the use of the main memory(the high speed portion) with the hard disk (the lower speed portion).
  - Technique for using the secondary storage to extend the apparent limited size of the physical memory beyond its physical size .
  - Implemented since the available physical memory will not be enough to host all the program.

Address space and Memory space

- An address used by the programmer is virtual memory , the set of such address is called address space.
- An address in main memory is location ,the set of such location is called memory space.
- Example: consider main memory :32k words(k=1024)=2^15 and auxiliary memory 1024k words=2^20(to address 15 bits of physical memory and 20 bits of virtual memory is required)

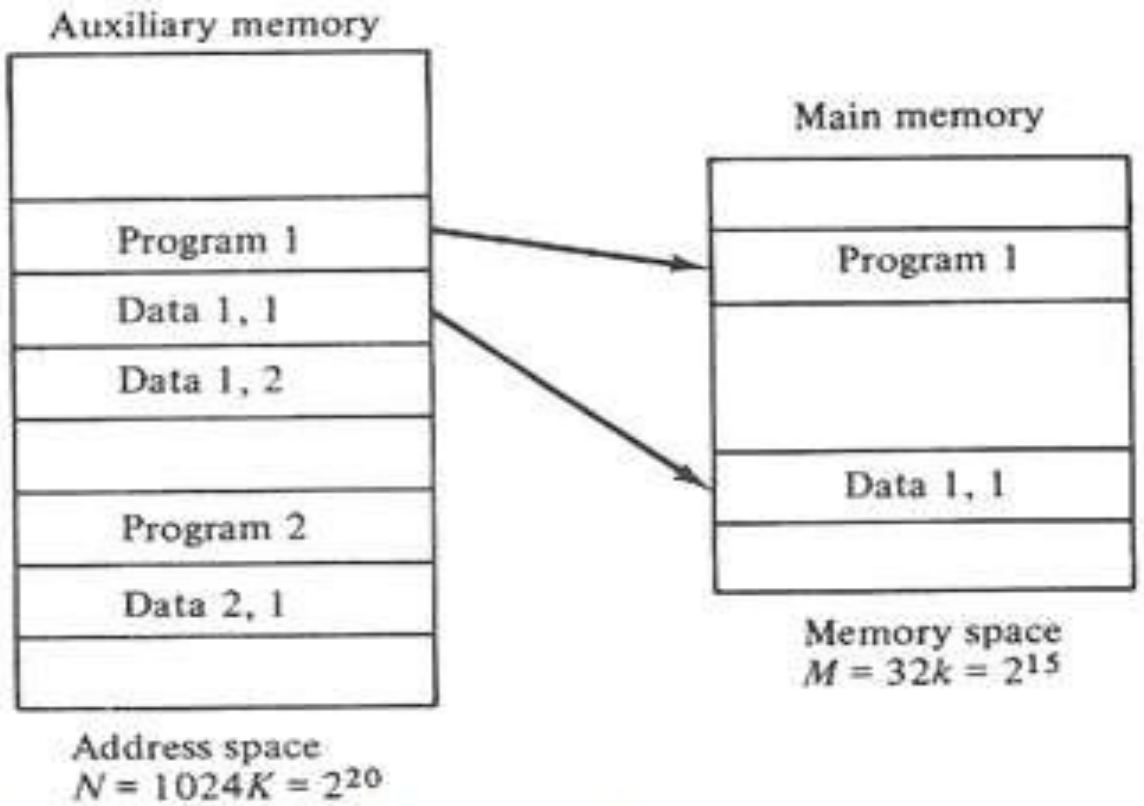


Fig: Relation between address and memory space in a virtual memory system

- Here auxiliary memory has the capacity of storing information equivalent to 32 main memories.
- Address space N = 1024K
- Memory space M = 32K
- In multiprogram computer system, programs and data are transferred to and from auxiliary memory and main memory based on the demands imposed by the CPU.

- In our example we have 20-bit address of an instruction (to refer 20-bit virtual address) but physical memory addresses are specified with 15-bits. So a table is needed to map a virtual address of 20-bits to a physical address of 15-bits.
- Mapping is a dynamic operation, which means that every address is translated immediately as a word is referenced by CPU.

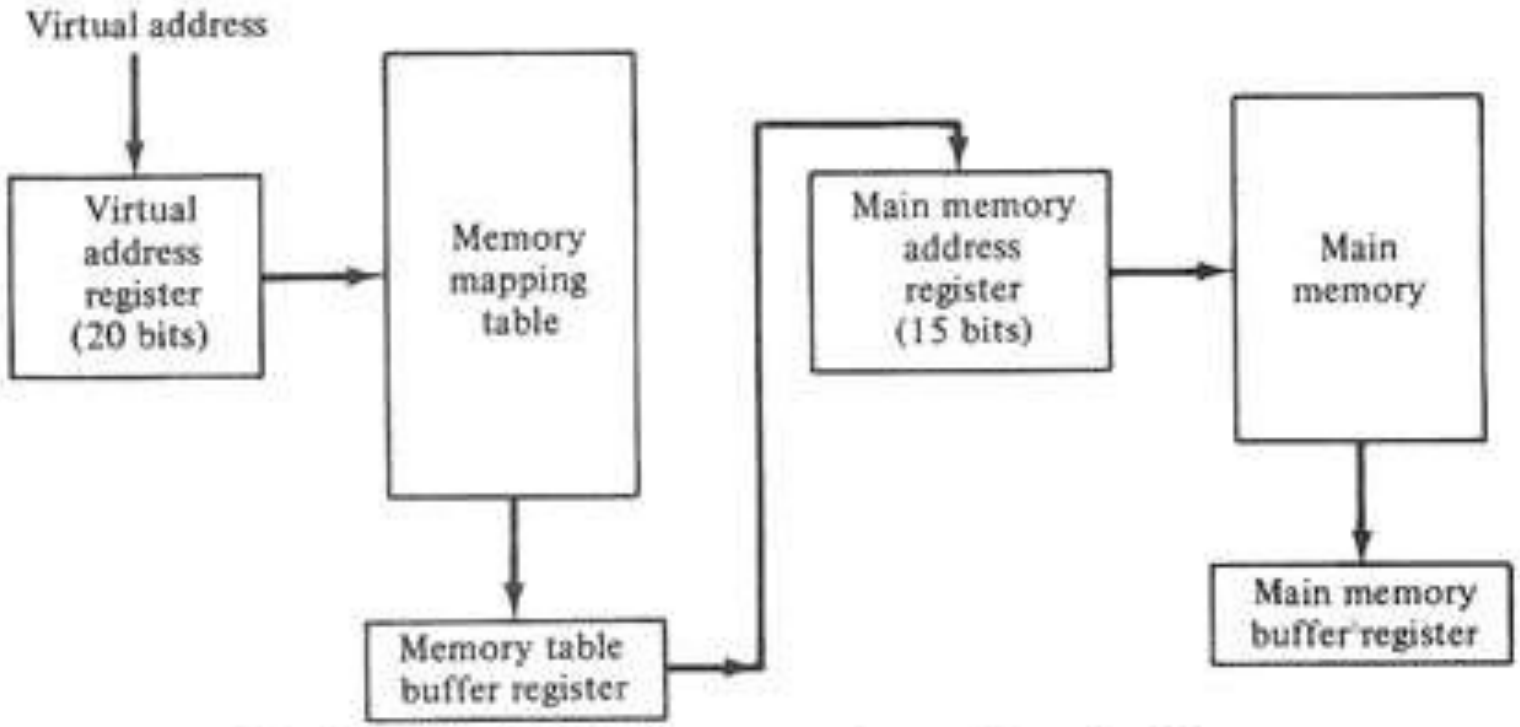


Fig: Memory table for mapping a virtual address

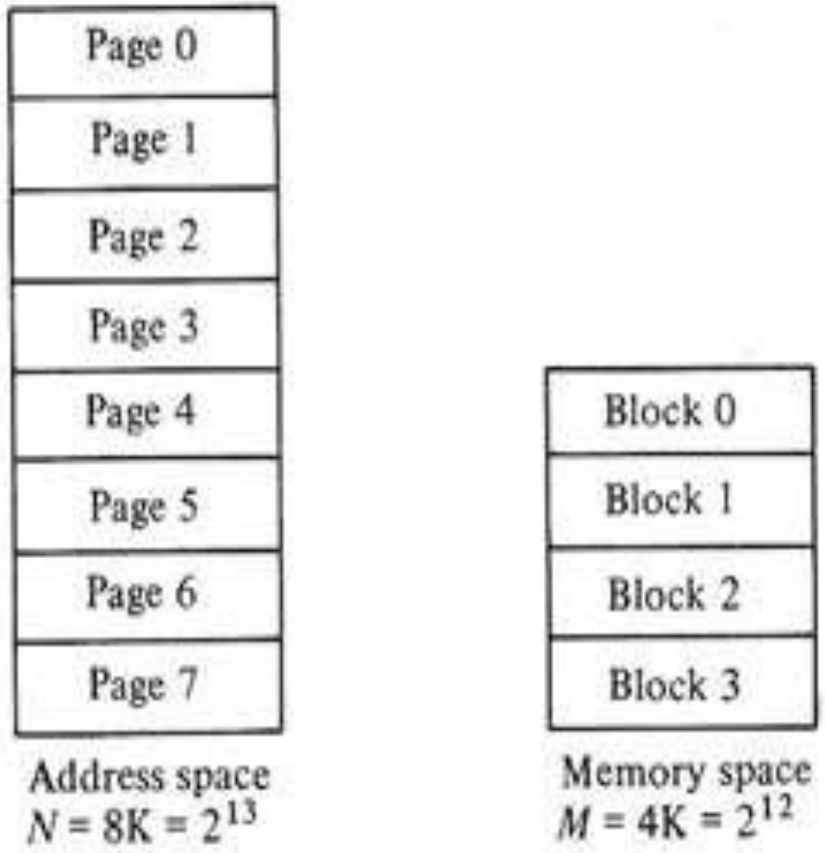
Address Mapping using Pages

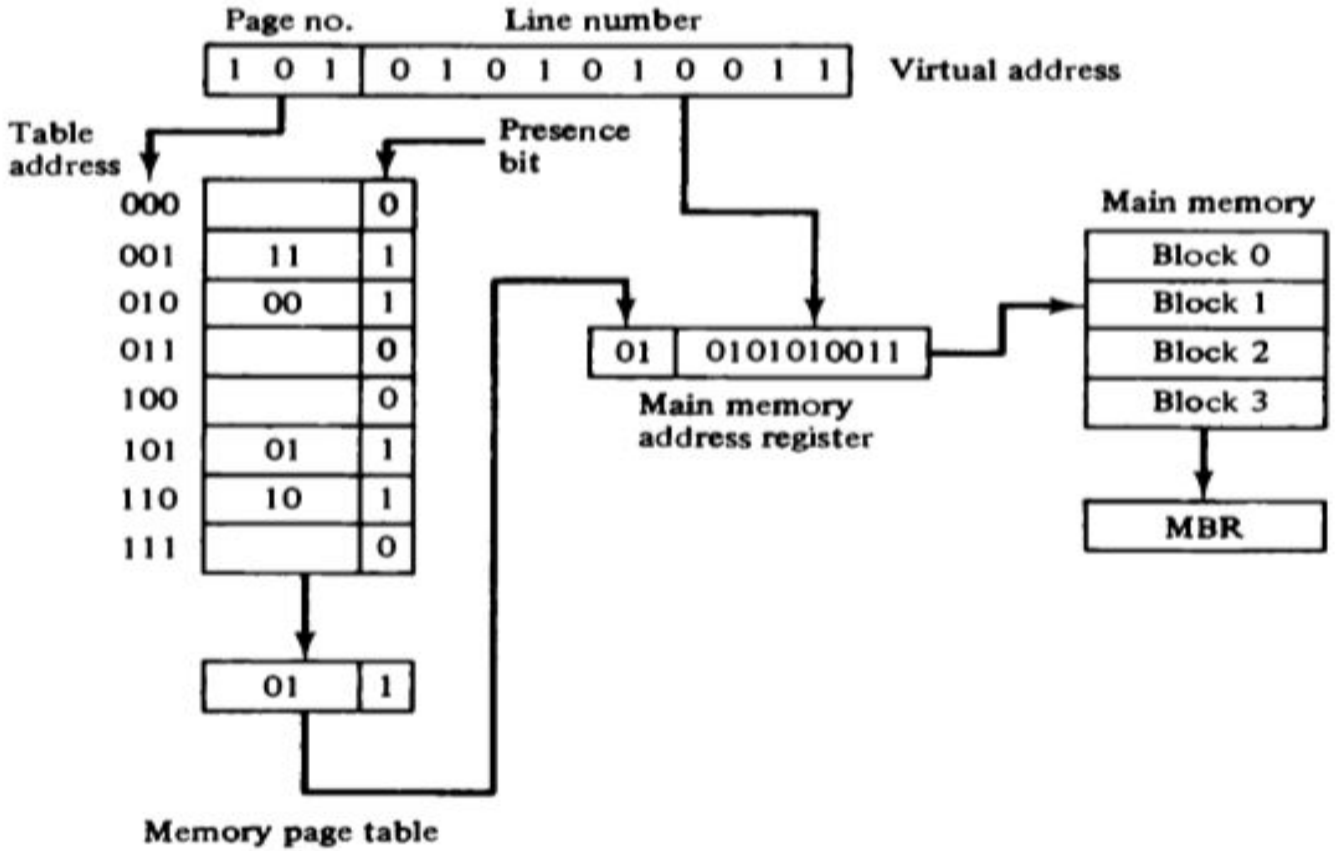
- **Blocks** (or page frame): Blocks are the groups of equal size which are broken down from physical memory and ranges from 64 to 4096words each.
- **Pages**: refers to a portion of subdivided virtual memory having same size as blocks i.e. groups of address space.

Exam ple: consider computer with address space = 8K and memory space = 4K.

If we split both spaces into groups of 1k words we get 8 pages and 4 blocks.

blocks.




Page no.
Line number
1 0 1
0 1 0 1 0 1 0 0 1 1
Virtual address
Table address
Presence bit
000
001 11 1
010 00 1
011 0
100 0
101 01 1
110 10 1
111 0
01 1
Memory page table
01 0101010011
Main memory address register
Main memory
Block 0
Block 1
Block 2
Block 3
MBR

Page Replacement

A virtual memory system is a combination of hardware and software techniques. A memory management software system handles:

- Which page in main memory should be removed to make room for a new page?
- When a new page is to be transferred from auxiliary memory to main memory?
- Where the page is to be placed in main memory?

Page fault

- When a program starts execution, one or more pages are transferred into main memory and the page table is set to indicate their position. The program is executed from main memory until it attempts to reference a page that is still in auxiliary memory. This condition is called page fault.
- When page fault occurs

1. The execution of the present program is suspended until the required page is brought into main memory.
2. It signifies that the page referenced by the CPU is not in main memory.

- GOAL: try to remove the page least likely to be referenced by in the immediate future.

.

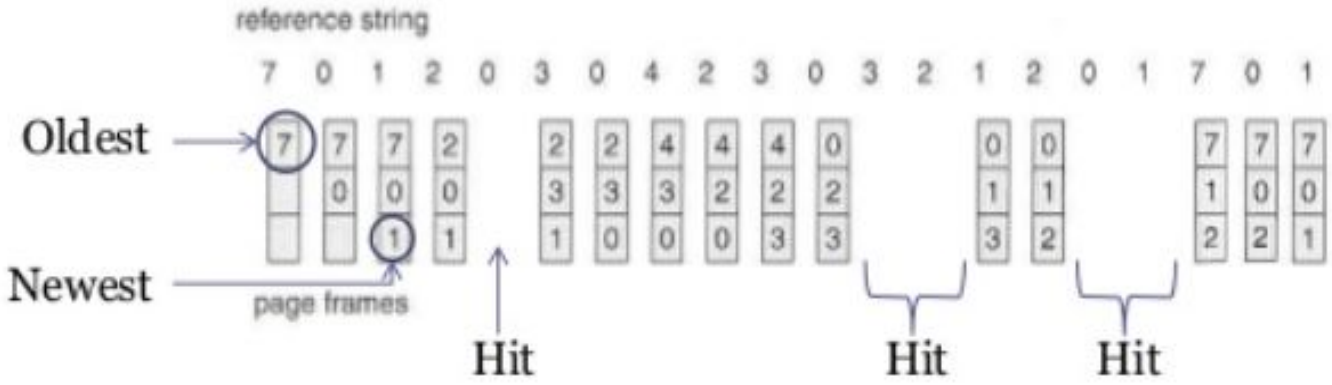
There are numerous page replacement algorithms, two of which are:

1. First-in First-out (FIFO): replaces a page that has been in memory longest time.
2. Least Recently Used (LRU): assumes that least recently used page is the better candidate for removal than the least recently loaded page.

FIFO

- The FIFO algorithm selects for replacement the page that has been in memory the longest time. Each time a page is loaded into memory, its identification number is pushed into a FIFO stack.
- FIFO will be full whenever memory has no more empty blocks.
- When a new page must be loaded, the page least recently brought in is removed. The page to be removed is easily determined because its identification number is at the top of the FIFO stack.

FIFO example



Advantages:

- FIFO is easy to understand.
- It is very easy to implement.

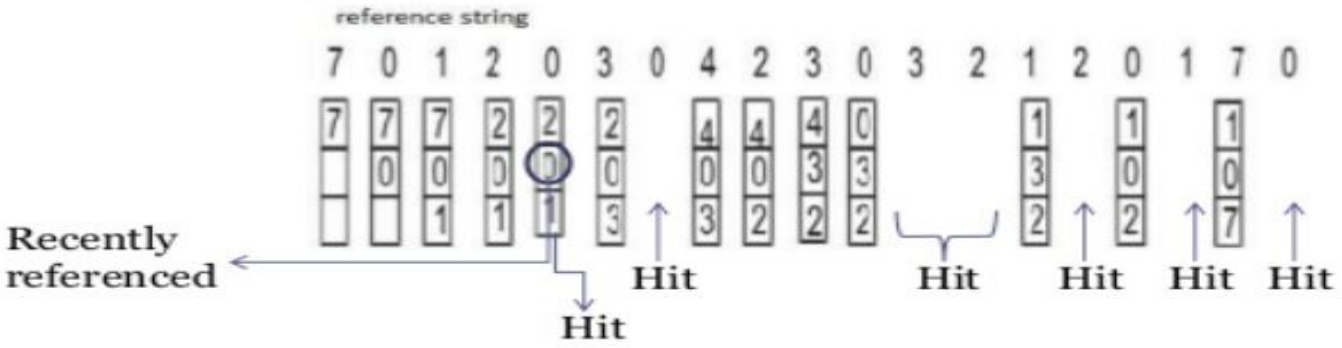
Disadvantage:

- The oldest block in the memory may often used.

LRU

- assumes that the least recently used page is a better candidate for removal than the least recently loaded page as in FIFO.
- The LRU algorithm can be implemented by associating a counter with every page that is in main memory. When a page is referenced, its associated counter is set to zero.
- At fixed intervals of time, the counters associated with all pages presently in memory are incremented by 1.
- The least recently used page is the page with the highest count.
- The counters are often called aging registers, as their count indicates their age, that is, how long ago their associated pages have been referenced.

LRU example



Advantages:

- It is quiet efficient.

Disadvantages:

- Implementation is difficult.

Cache Memory

- If the active portions of the program and data are placed in a fast small memory, the average memory access time can be reduced, thus reducing the total execution time of the program. Such a fast small memory is referred to as a cache memory.
- It is placed between the CPU and main memory.
- The cache memory access time is less than the access time of main memory by a factor of 5 to 10.
- The performance of cache memory is frequently measured in terms of a quantity called hit ratio.

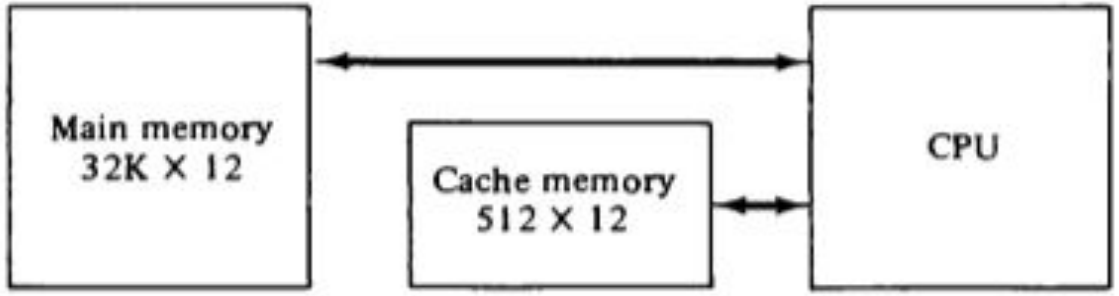
$$\frac{\text{Number of cache hits}}{(\text{Number of cache hits} + \text{Number of cache misses})} = \text{Cache hit ratio}$$

Cache Mapping

The transformation of data from main memory to cache memory is referred to as a mapping process. Three types of mapping procedures

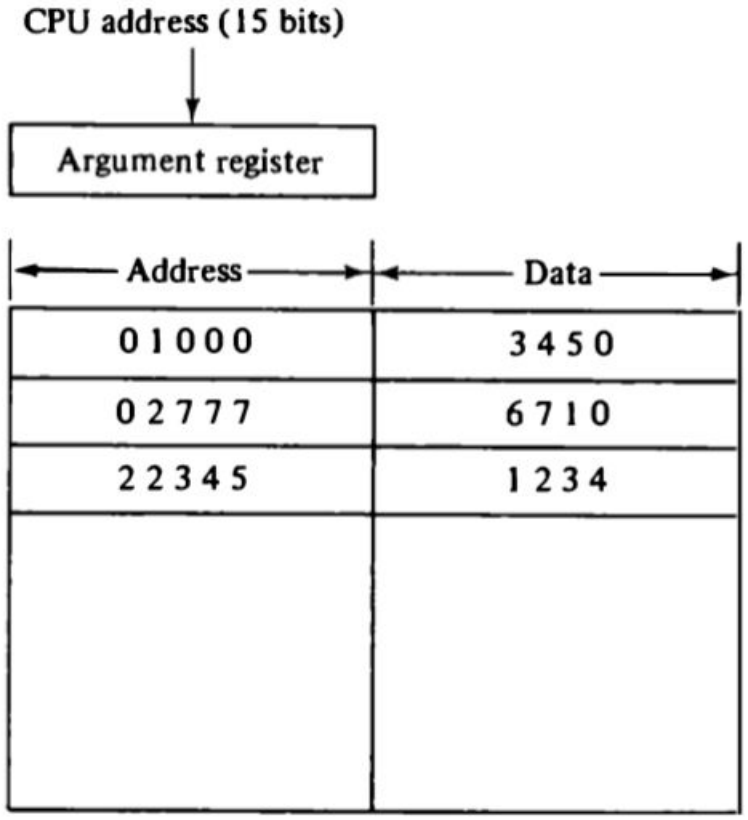
1. Associative mapping
2. Direct mapping
3. Set-associative mapping

Cache memory mapping techniques



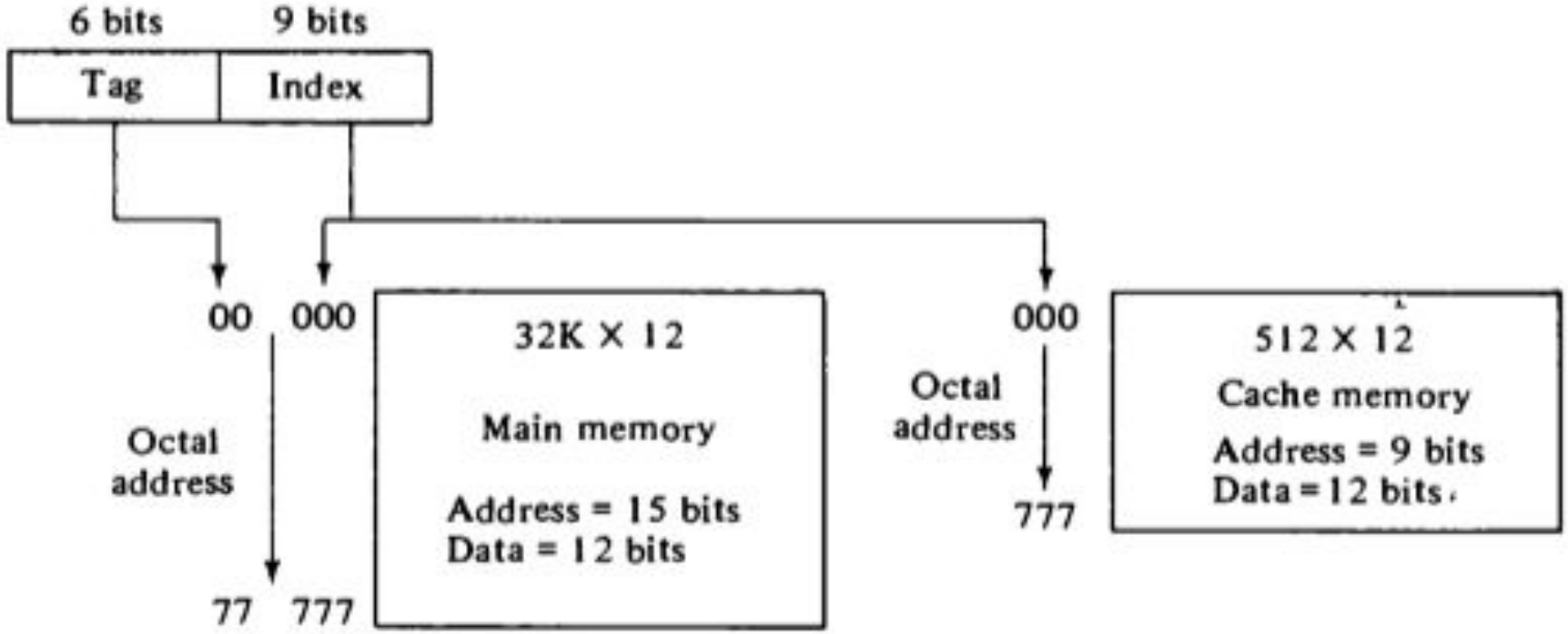
Associative Mapping

Associative mapping cache (all numbers in octal).

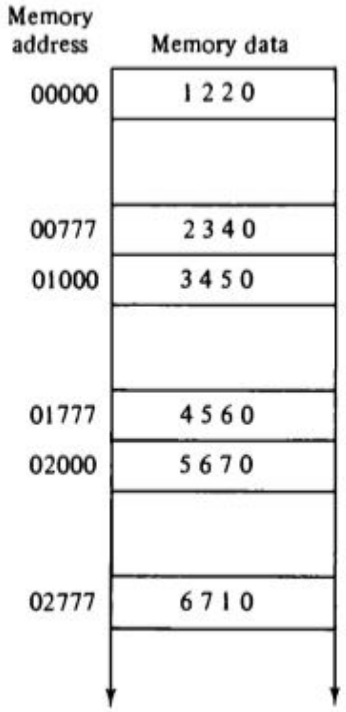


Direct Mapping

The n-bit memory address is divided into two fields: k bits for the index field and n - k bits for the tag field.

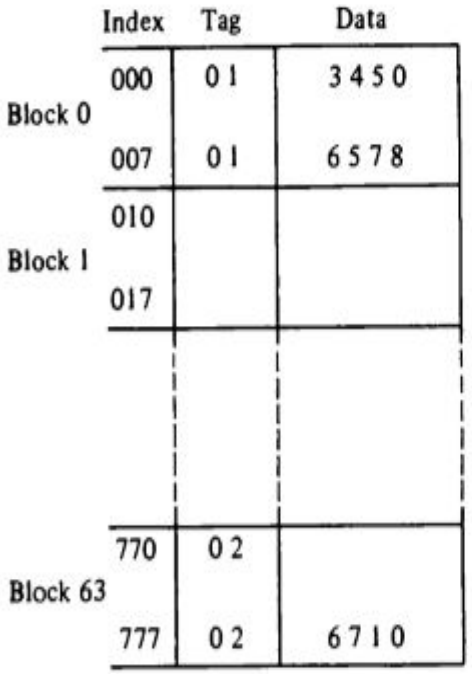


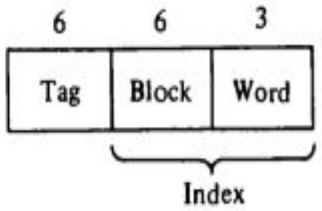
Direct Mapping



(a) Main memory

(b) Cache memory





Set Associative Mapping

Disadvantage of direct mapping is that two words with the same index in their address but with different tag values cannot reside in cache memory at the same time.

Index	Tag	Data	Tag	Data
000	01	3450	02	5670
777	02	6710	00	2340

THANK YOU!