MEMORY ORGANIZATION

MEMORY UNIT

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)

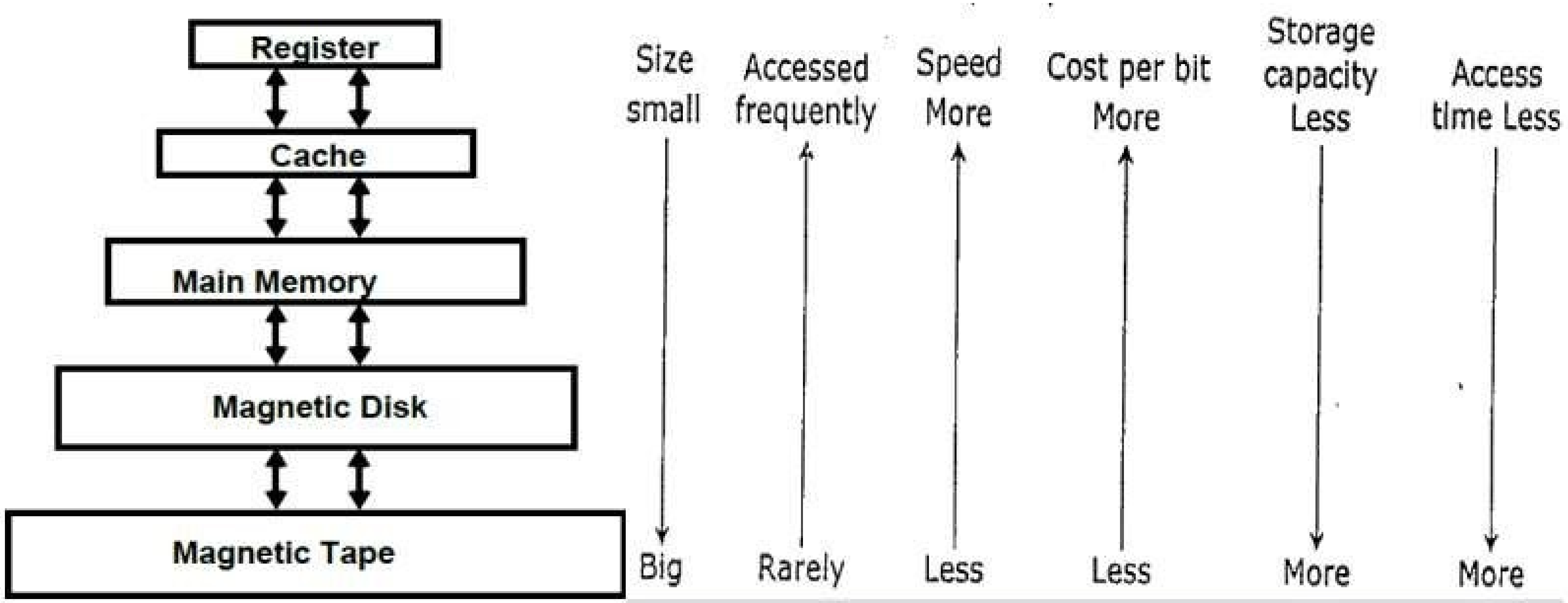
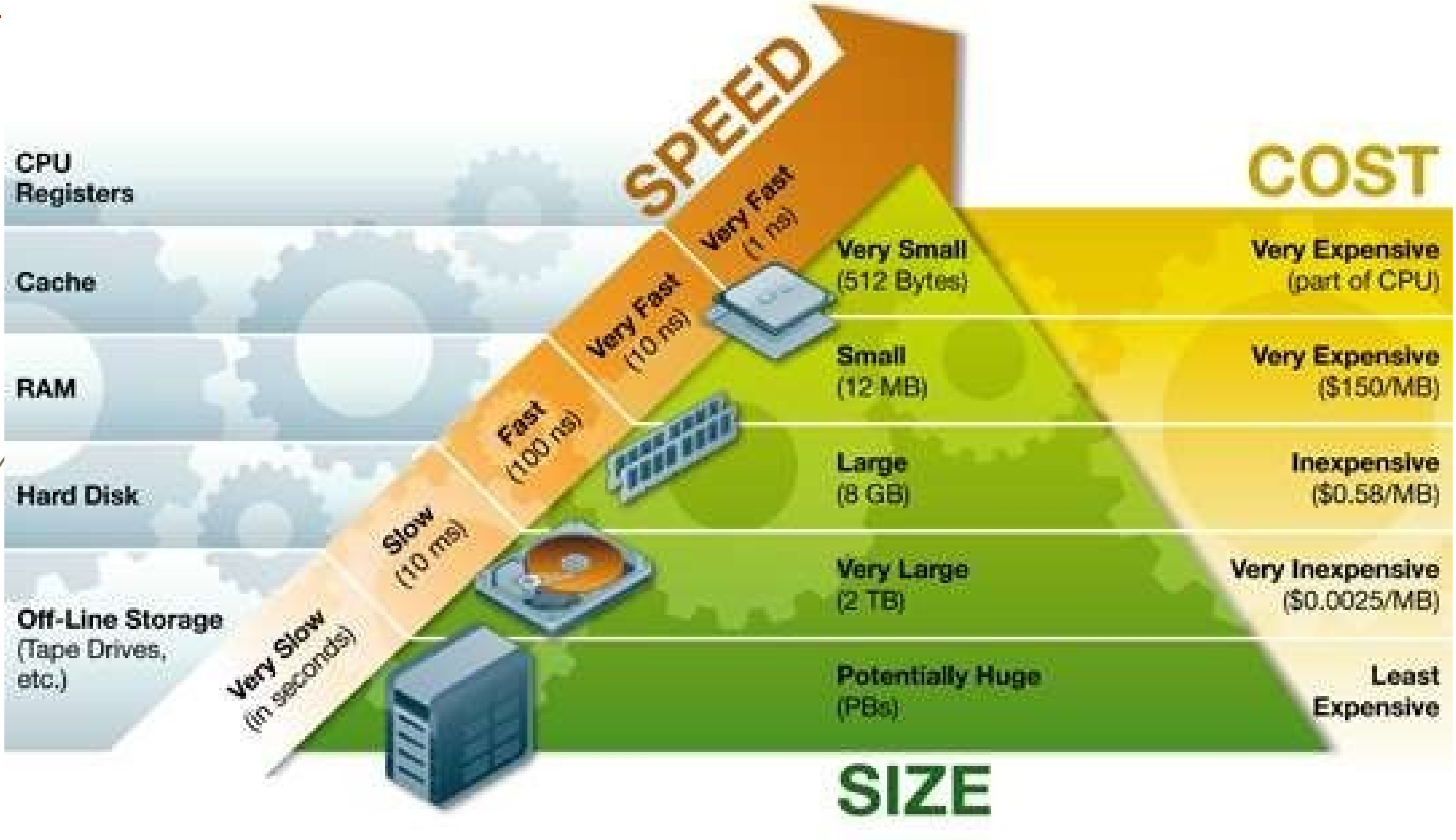


Figure: Memory Hierarchy


SPEED
COST
SIZE
CPU Registers
Cache
RAM
Hard Disk
Off-Line Storage
(Tape Drives, etc.)
Very Fast
(1 ns)
Very Fast
(10 ns)
Fast
(100 ns)
Slow
(10 ms)
Very Slow
(in seconds)
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

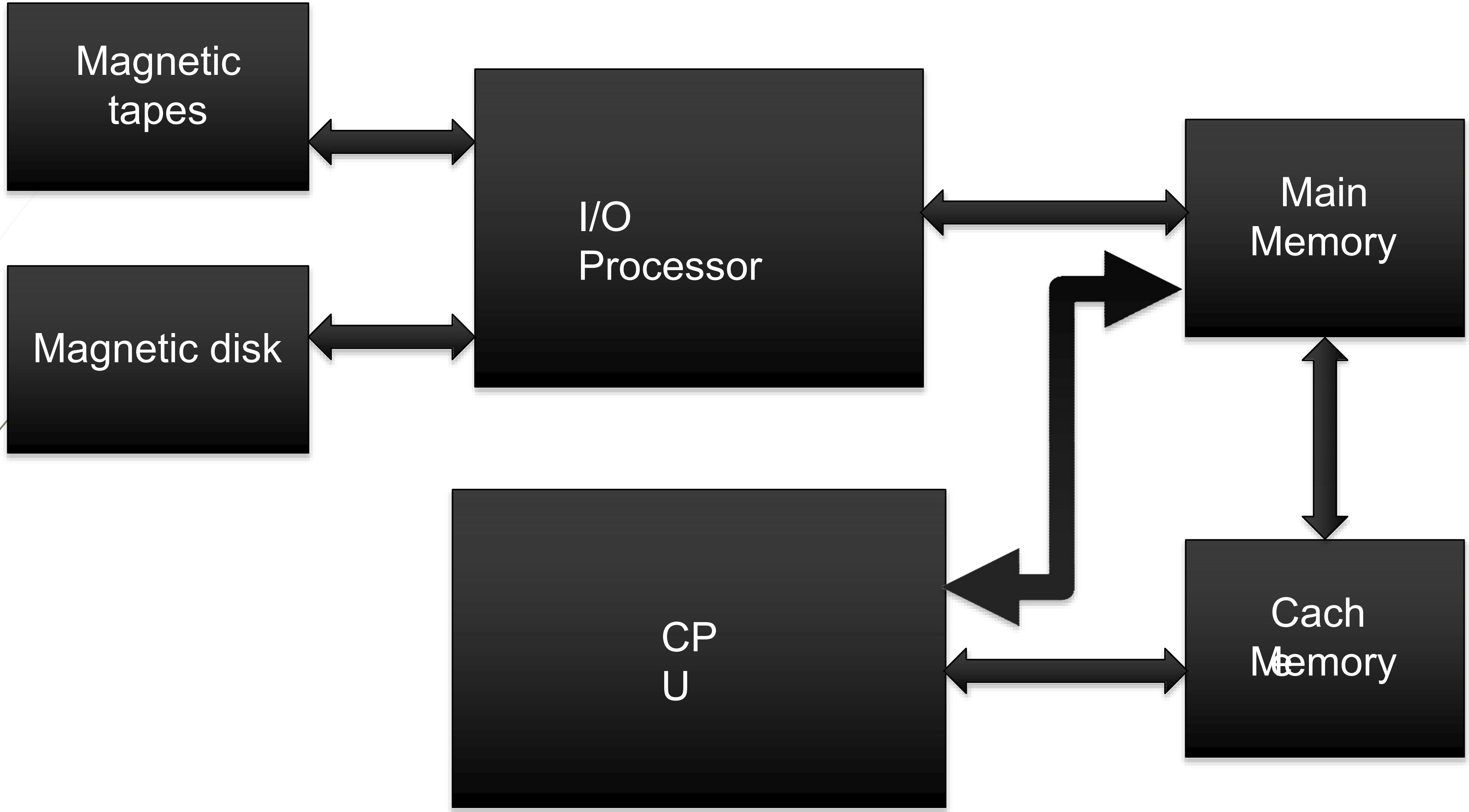


Fig: Memory Hierarchy in a Computer System

MAIN

Main memory

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

RAM

- Volatile memory.
- Store the 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

- Permanent memory
- Store the information required for computer operations
- Types of ROM (PROM,EPROM,EEPROM)

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() MEMORY(SECONDARY

- The most common auxilliary device used in the computer system is magnetic disk and magnetic tape.
- Store large amount of data permanantly.
- Portable

- **Types of auxilliary memory**
- Magnetic disk
- Magnetic tape

Magnetic disk

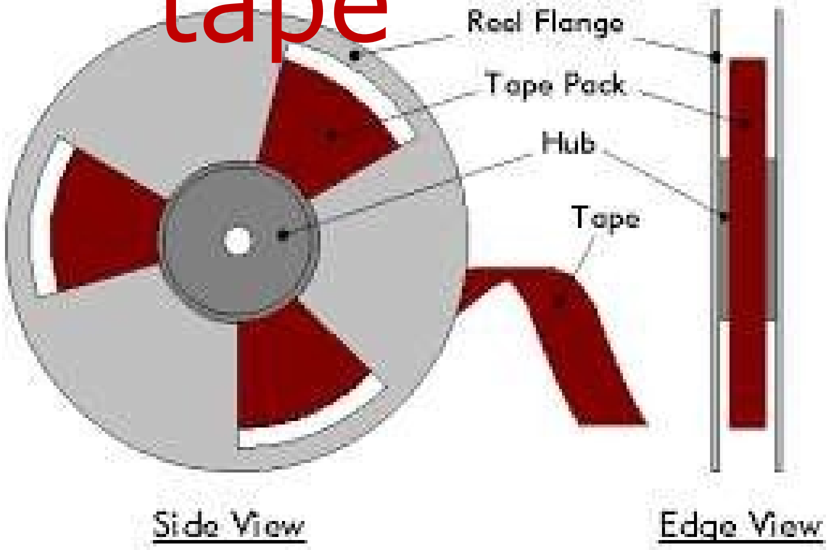
- circular plate ,made of metal or plastic coated with magnetize material. High speed of rotation
- Bits are store in a concentric circle called tracks.
- Division of tracks are called sectors.

Magnetic disk

Magnetic tape

- Sequential access memory used for storing,backup,audio,video data etc .
- Highly reliable memory.
- Slower for read write operation.

Magnetic tape



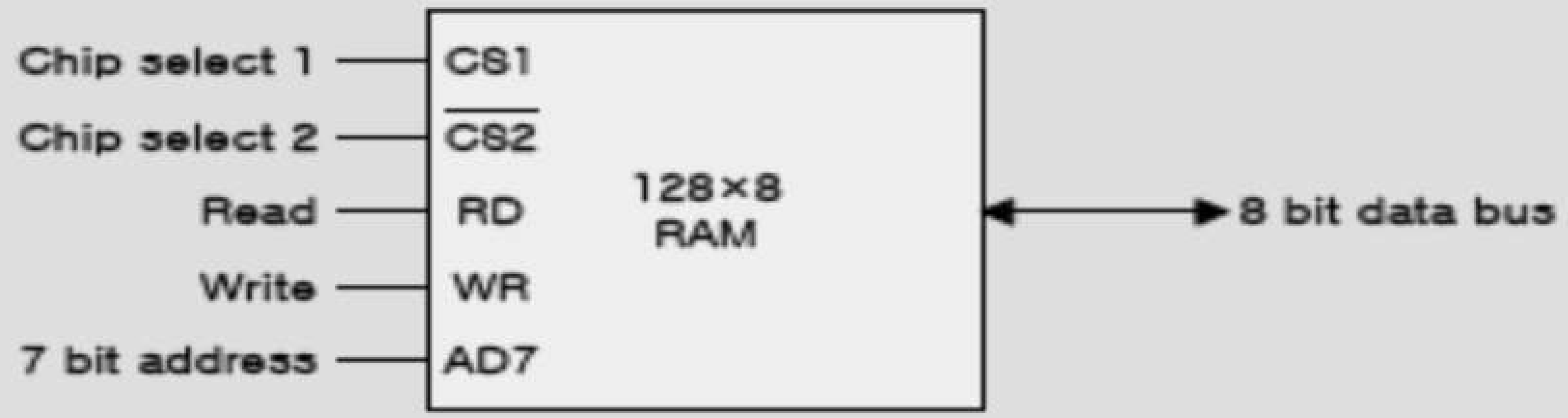
RAM AND ROM CHIP

RAM AND ROM CHIPS

RAM chip:

- Used for communication with the CPU if one or more control inputs.
- requires 7-bit address and an 8-bit bidirectional data bus.
- chip select (CS) are for enabling the chip.

Fig. RAM chip



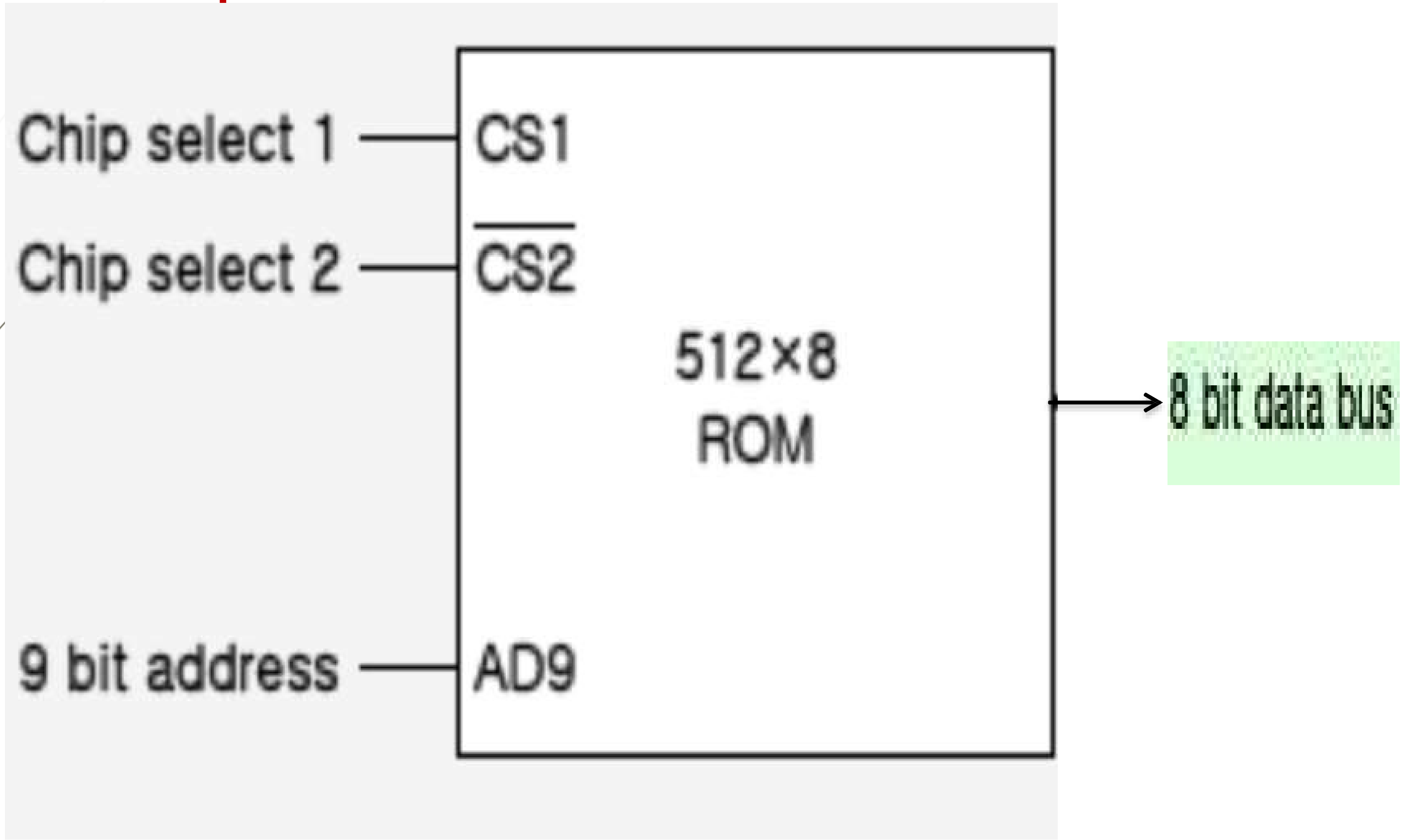
(a) Block diagram

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


Chip select 1
Chip select 2
9 bit address
CS1
$\overline{CS2}$
AD9
512×8
ROM
8 bit data bus

- 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

- 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)

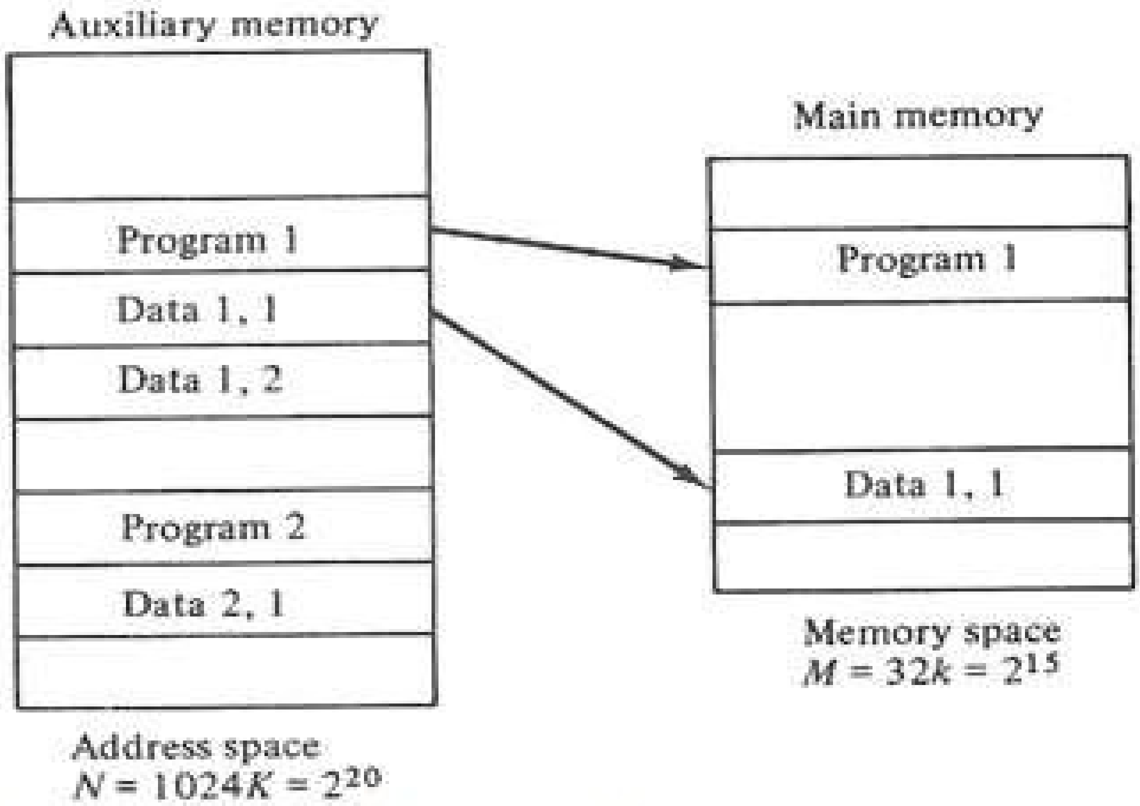


Fig: Relation between address and memory space in a virtual memory system

- 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.

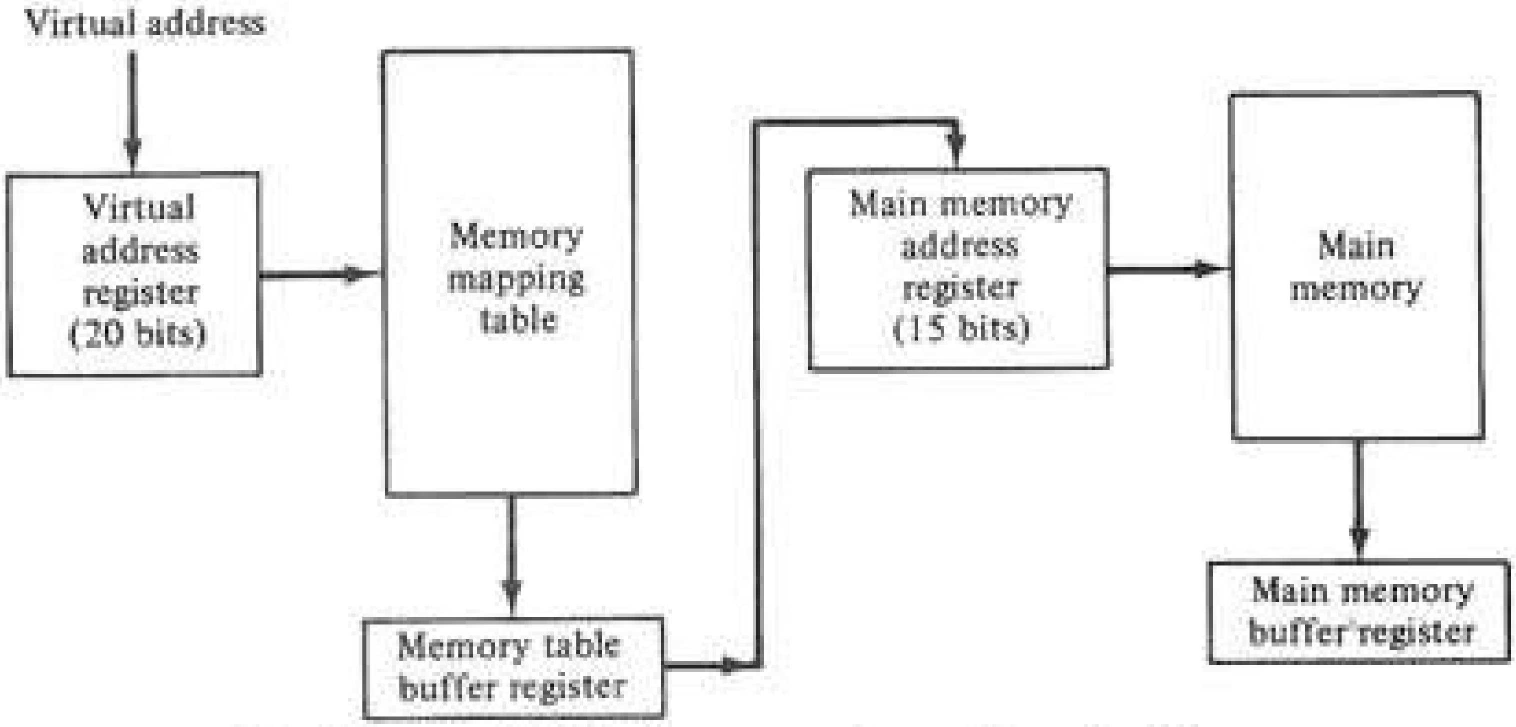


Fig: Memory table for mapping a virtual address

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.

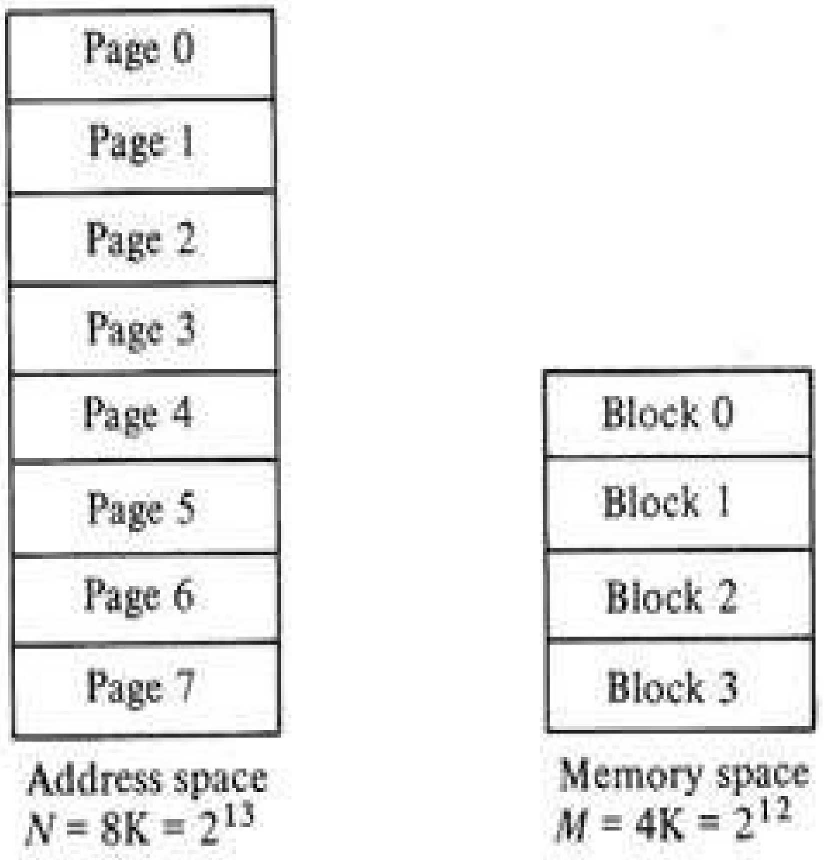
Example: consider computer with address space = 8K and memory space = 4K.

If we spit both spaces into

groups of 1k words we get

8 pages and 4 blocks.

blocks.



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?

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.

- When a page fault occurs in a virtual memory system, it signifies that the page referenced by the program is not in main memory. A new page is then transferred from auxiliary memory to main memory. If main memory is full, it would be necessary to remove a page from a memory block to make a room for a new page. The policy for choosing pages to remove is determined from the replacement algorithm that is used.
- GOAL: try to remove the page least likely to be referenced by in the immediate future.

Memory Management Hardware

Introduction

Collection of hardware and software procedures for managing various program.

Basic component of MMU

- Sharing common program by multiple user
- Protection of information unauthorized

Segmented Page Mapping

- Length of each segment allowed to grow and contract according need of program execution.
- Way of specifying the length of a segment by associating with it a no. of equal sized page
- Consider diagram below

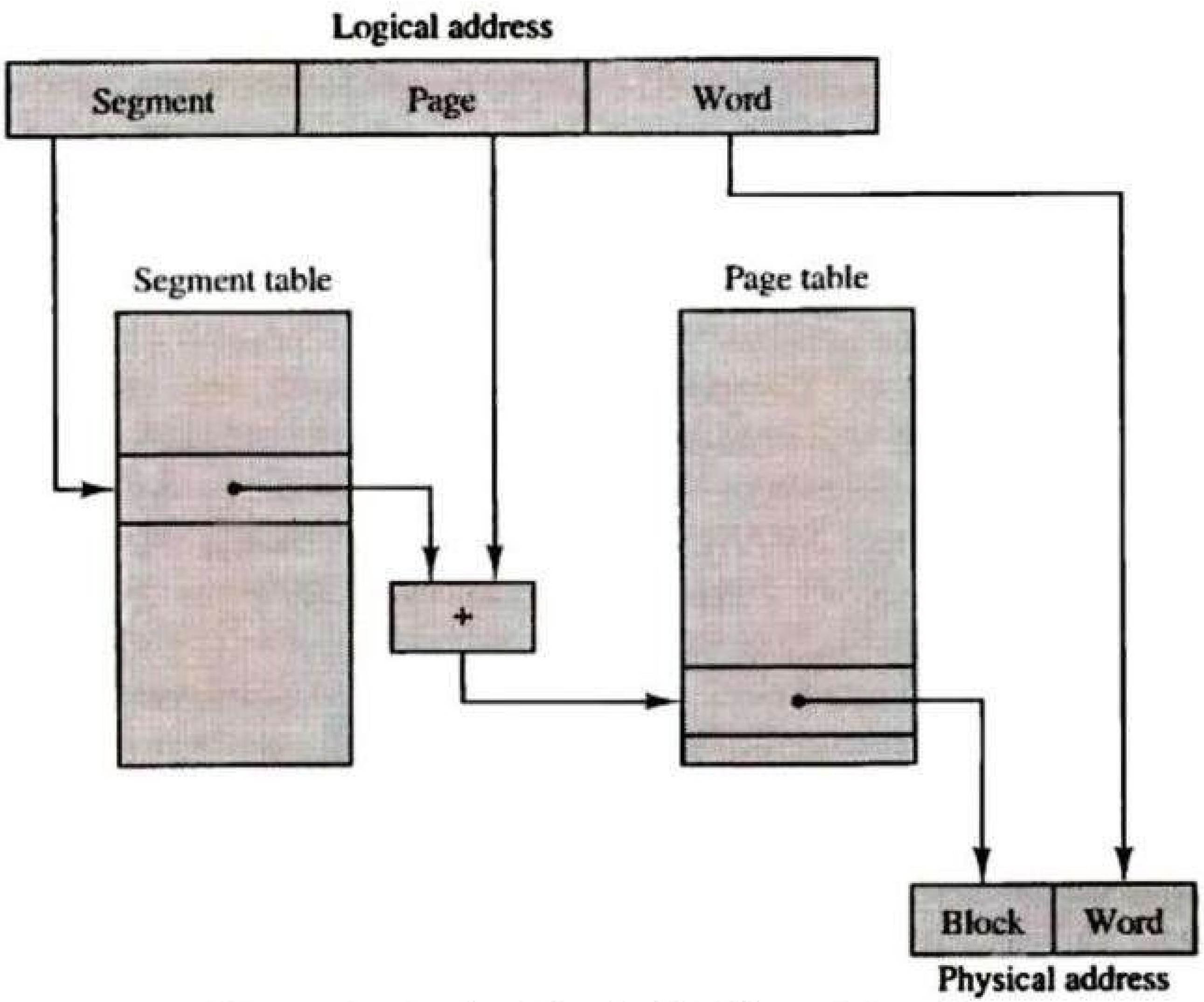


Figure: Logical to Physical Address Mapping

In above fig:

Consider

Logical Address=Segment+page+word

segment specifies segment no.

Page field specifies page within the segment

Word field specifies specific word within the page

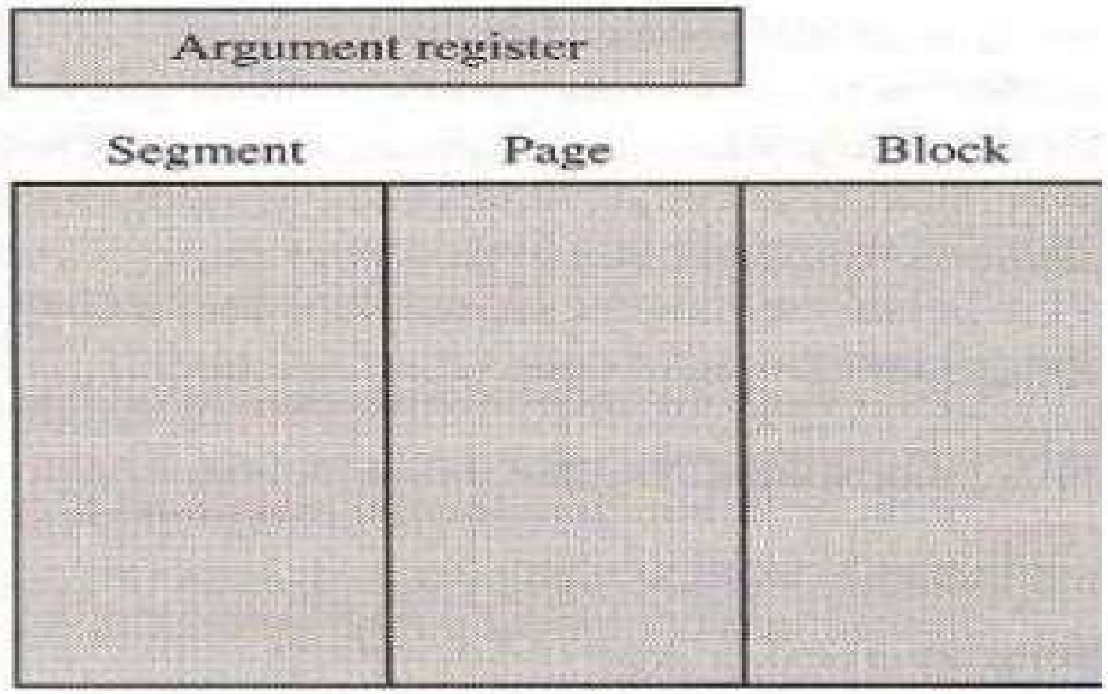


Fig: Associative memory Translation Look-aside Buffer (TLB)

This is a fast associative memory (TLB) and holds most recently referenced entries.
(Alternatively we could store above two tables: segment table and page table, in two separate small memories which really increases the CPU access time)