UNIT:4

MEMORY ORGANIZATION

Topics To be Covered....

4.1 Memory Hierarchy

4.2 Memory Classification

4.3 RAM,ROM,PROM,EPROM

4.4 Main Memory

4.5Auxiliary Memory

4.6 Associative Memory

4.7 Cache Memory

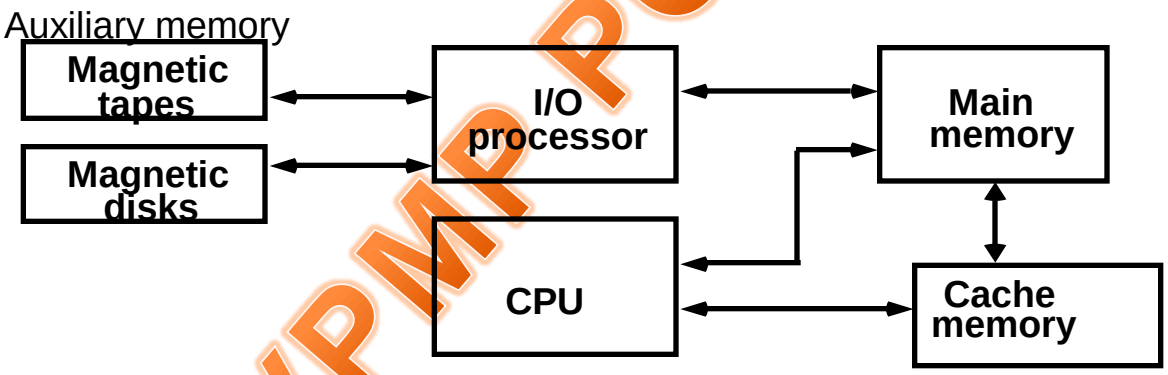
4.8 Virtual Memory

4.1 MEMORY HIERARCHY

- **GOAL :** To obtain the **highest possible access speed** while **minimizing the total cost** of the memory system.

- The memory unit that communicate directly with CPU is called **Main Memory**.

- Devices that provides backup storage are called **Auxiliary Memory**.

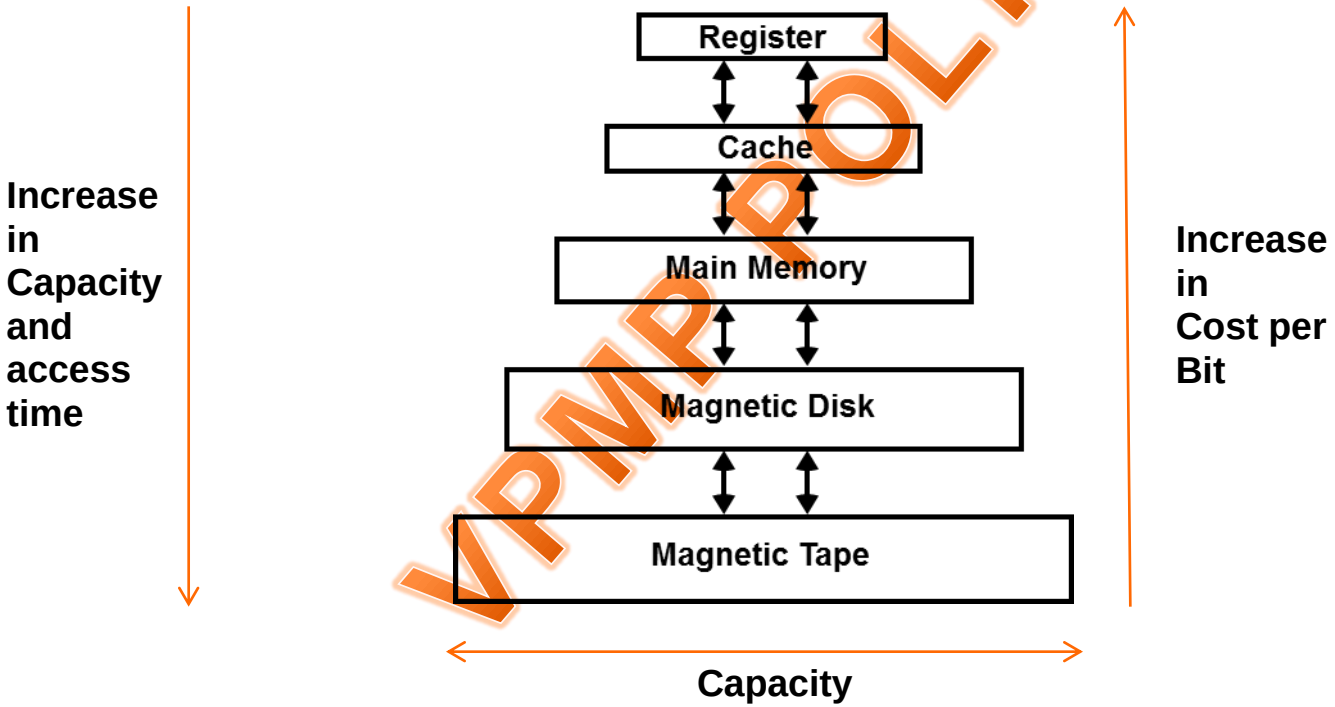


CONT…

- Only program and data **currently needed** by the processor reside in main memory.
- The **Main Memory communicate directly with CPU** and with **Auxiliary Memory Devices through an I/O Processor**.
- Cache is used to increase the speed of processing by making current programs and data available to CPU.
- Auxiliary Memory→ Large storage capacity, Relatively Inexpensive, Low access speed compared to main memory.
- Cache Memory→ Very small, Relatively expensive, very high access speed.

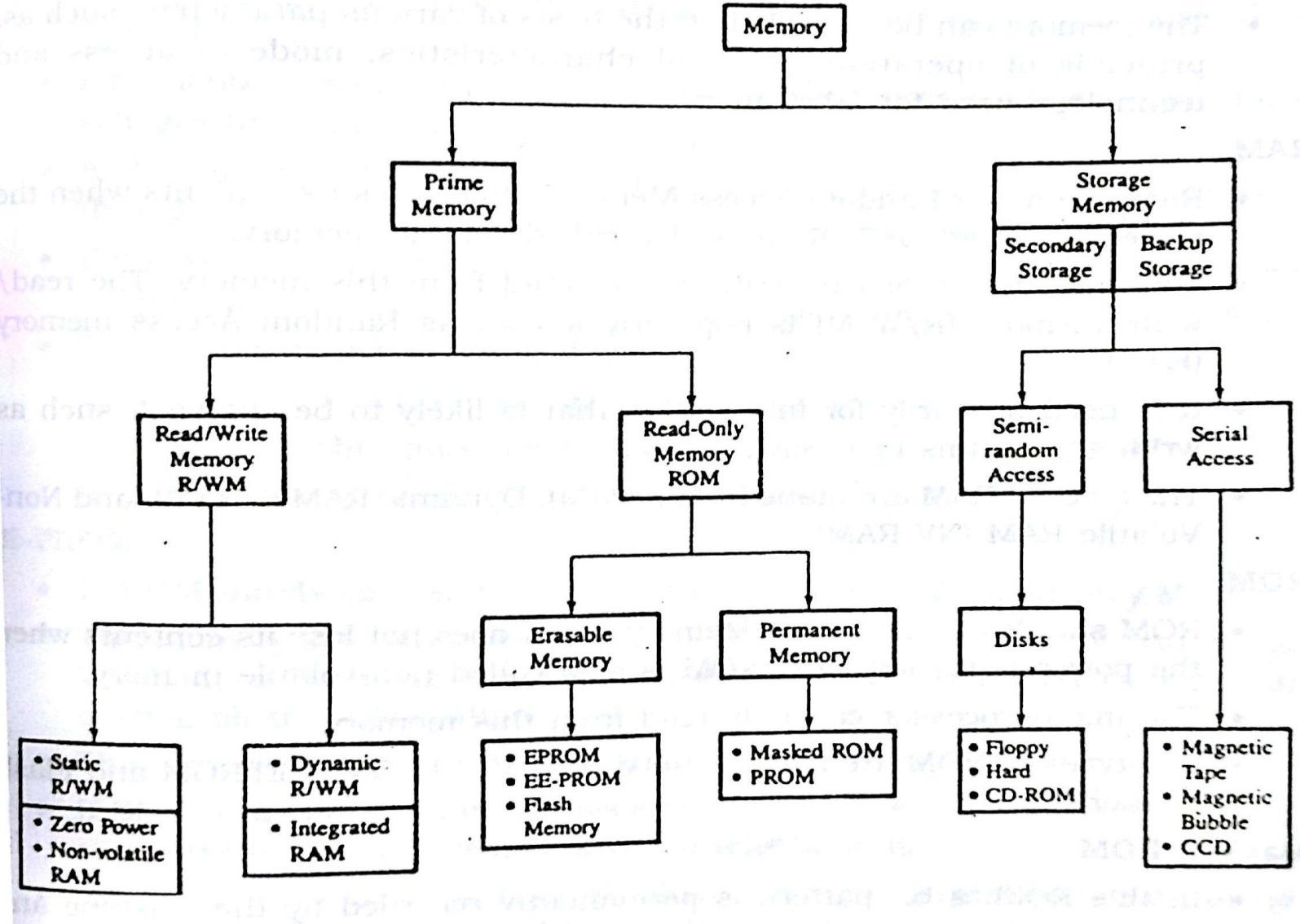
CONT…

- **Advantage of Memory Hierarchy:**
  - ✓ Increase capacity
  - ✓ Improve average access time
  - ✓ Decrease cost/bit
  - ✓ Decrease frequency of accesses to slow memory



Four Level Memory Hierarchy

4.2 MEMORY CLASSIFICATION



CONT…

- The memory is made of registers.
- Each registers has a group of flip flops that store bits of information.
- These flip flops are called memory cells.
- The number of bits stored in a register is called memory word.
- Memory chips are available in various word sizes.
- The user can use this memory to hold program and store data.
- In a memory chips, all registers are arranged in a sequence and identified by binary numbers called memory addresses.

4.3 RAM, ROM,PROM ,EPROM

RAM

- Random Access Memory
- RAM losses its contents when the power is turned OFF. So, RAM is called volatile memory.
- User save data in memory but they may not be store permanently.
- It is called due to Random selection of Memory Location.
- The microprocessor can write into or read from this memory.
- Types of RAM are:
  - ✓ Static RAM (SRAM)
  - ✓ Dynamic RAM(DRAM)

CONT…

ROM

- ROM does not loss its contents when the power is turned OFF. So, ROM is called Non volatile memory.
- The microprocessor can only read from this memory.
- It is programmable once.
- Types of ROM are:
  - ✓ PROM
  - ✓ EPROM
  - ✓ EEPROM
  - ✓ Masked ROM
  - ✓ Flash Memory

CONT…

PROM

- Programmable Read only memory
- **PROM Programmer program** in a blank chip.
- Once you program it, You can never change.
- It maintains **large storage media** but **can not** offer the **erase** feature.

- **Write Data once and Read many times**.
- It has **poly silicon wires** arranged in a matrix.
- These wires can be functionally viewed as diodes or fuses.

CONT…

EPROM

- Erasable Programmable Read only memory
- This memory stores a bit by charging the gate of **field effect transistor.**
- Information is stored by **EPROM Programmer**, which apply high voltage to charge the gate.

- User can **delete the data of EPROM** through pass on **Ultraviolet Light** and the chip can be reprogrammed.
- **Chip can be used many times.**
- **Erasing Process takes 15 to 20 minutes.**

CONT…

EEPROM

- Electrically Erasable Programmable Read only memory
- Functionally similar to EPROM**.**

- Information can be **changed by using electrical signals at the register level** instead of **all the information**.

- **Entire chip** can be erased in **10ms**.
- This memory is **expensive** compared to EPROM

CONT…

FLASH MEMORY

- It is aVariation of EEPROM.
- It erase data at the block level .

MASKED ROM

- It Programmed by the **Integrated Circuit Manufacture**.
- Use in **Network OS and Server OS.**
- Bit pattern is **permanently** recorded by the **masking**.

4.4 MAIN MEMORY (RAM, PRIMARY MEMORY)

- Random Access Memory
- Types of RAM are:
  - ✓ Static RAM (SRAM)
  - ✓ Dynamic RAM(DRAM)

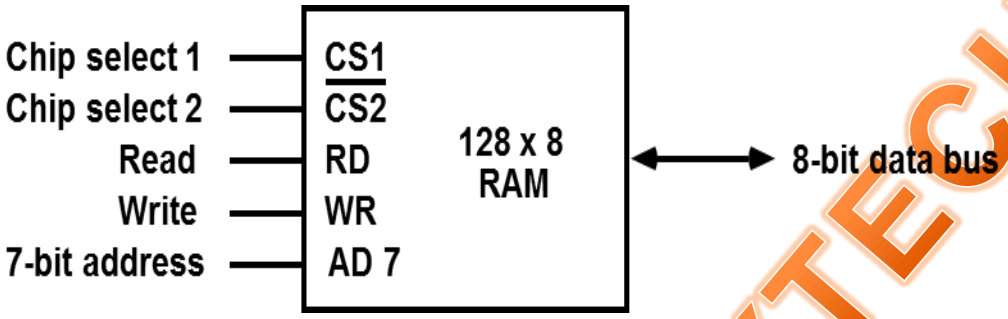
CONT…

- **Static RAM**
  - ✓ It consists of flip flops that stores the binary information.
  - ✓ It is used in implementing the cache.
  - ✓ It is expensive, fast, low density, high power in operation.
  - ✓ Easy to use and shorter Read and Write cycles.

CONT…

- **Dynamic RAM**
  - ✓ It is used in implementing the main memory.
  - ✓ It stores the binary information in the form of electric charges.
  - ✓ It offers reduced power consumption and large storage capacity.

CONT…



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

Typical RAM chip

CONT…

- It has **bidirectional data bus** that allow transfer of data from Memory to CPU during Read or From CPU to Memory during write operation.
- High Impedance state behaves like an open circuit.
- Each word is of eight bit.
- It requires 7-bit address.
- When the chip is selected, the two binary states in R/W specify the two operations of Read or Write.
- $\overline{CS2}$ →bar indicates this input is enabled when it is equal to 0.

READ ONLY MEMORY

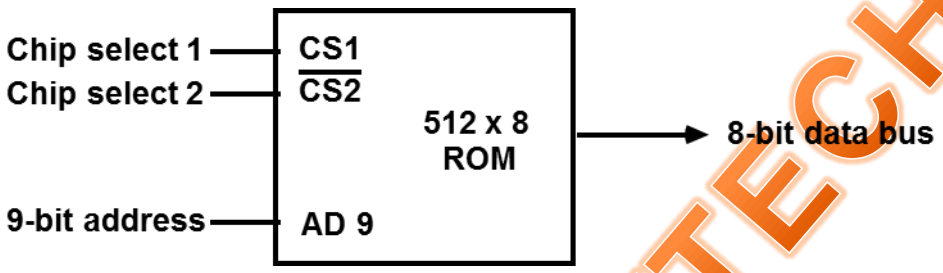
- ROM does not loss its contents when the power is turned OFF. So, ROM is called Non volatile memory.

- The microprocessor can only read from this memory.
- It is programmable once.
- Types of ROM are:

  - ✓ PROM
  - ✓ EPROM
  - ✓ EEPROM
  - ✓ Masked ROM
  - ✓ Flash Memory

CONT…

ROM Chip



- The two chip select inputs must be CS1=1 and CS2=0 for the unit to operate.
- Otherwise the data bus is in High Impedance State.
- When the chip is enabled by the two select inputs, the byte selected by the address lines appears on the data bus.

BOOTSTRAP LOADER

- It is stored in ROM.
- When power is turned, the hardware of the computer sets the program counter to the first address of the bootstrap loader.

- The bootstrap program loads a portion of the operating system from disk to main memory and control is then transferred to the operating system.

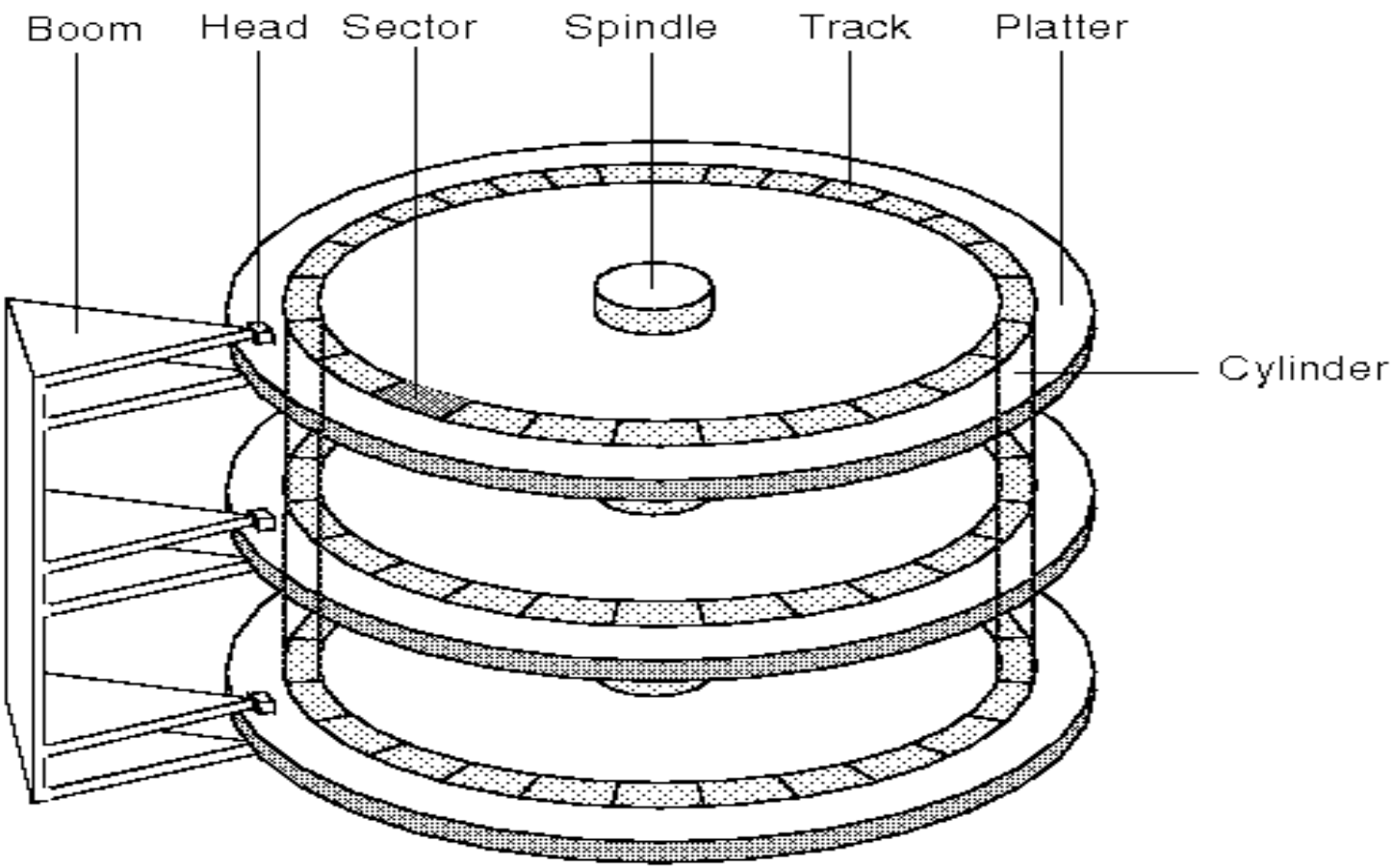
4.5 Auxiliary memory

Example: Magnetic Disk, Tape

- **Magnetic Disk** (Hard Disk, Floppy Disk)
- It is a **circular Plate** constructed of **metal or plastic** coated **magnetized material**.
- Both sides of disks are used.
- **Several disks** may be stacked on **one spindle**, with Read-Write head available on each surface.
- All sides rotate together at high speed.

CONT…

CONT…



CONT…

- Bits are stored in magnetized surface along concentric circles are called **Tracks**.
- Tracks are divided into sections called **Sectors**.
- Some unit use a **single Read/Write head** for each **disk Surface**.
- Some unit use a **separate Read/Write heads for each track** in each surface.
- A disk system is addressed by **address bits** that specify the **Disk number,Disk surface,Sector number and the Track** within the sector.
- Storage capacity of disk depends on **bits per inch of track** and **Track per inch of surface**

CONT…

How To Read/Write?

- First Read/Write heads are positioned in the specified **Track**.
- The system has to wait until the disk reaches the **specified sector** under Read/Write head.
- Information **transfer is very fast** once sector has been reached.

Auxiliary memory- Floppy Disk

CONT…

- A **disk drive with removable disk** is called a **Floppy Disk**.
- It is made of **plastic coated magnetic material**.
- Type of Floppy:
    1. **5 ¼ Inch**: Capable of storing between between 100k and 1.2 MB of Data.
    2. **3 ½ Inch**: Small size, Larger storage capacity,Common size is 1.44 MB
- It has slower access than Hard Disk
- It has Less storage capacity.
- It is Less Expensive
- It is Portable

Auxiliary memory- Magnetic Tap

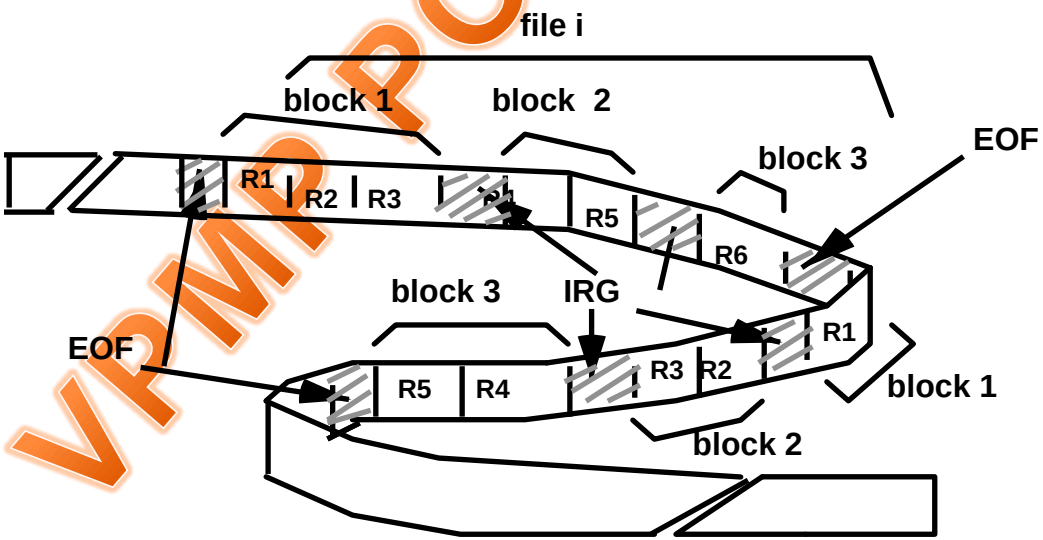
CONT…

- It consists of **magnetic,** coating on a thin plastic strip.
- Example: Recording Tap(used for audio, video, general purpose digital data storage using a computer)

- It is a Secondary Storage Device.
- It consists of Variety of Length(600 to 3000 ft)
- It is packaged in one of three ways:
    1. Open Reel
    2. Cartridge
    3. Cassette

CONT…

- RW/WR heads are mounted on each track so that data can be recorded and read as a sequence of characters.
- **Information is stored in blocks referred to as records**.
- **Each record** on tape has an identification bit pattern **at the beginning and end.**



Auxiliary memory- Performance parameter

- Access Time: Time Required **to reach a storage location** in memory and **obtain its content**.
- Transfer Time: Time required **to transfer data to or from the device**.
- Transfer Rate: **No of words/characters** that device can transfer per **second**.

4.6 Associative memory

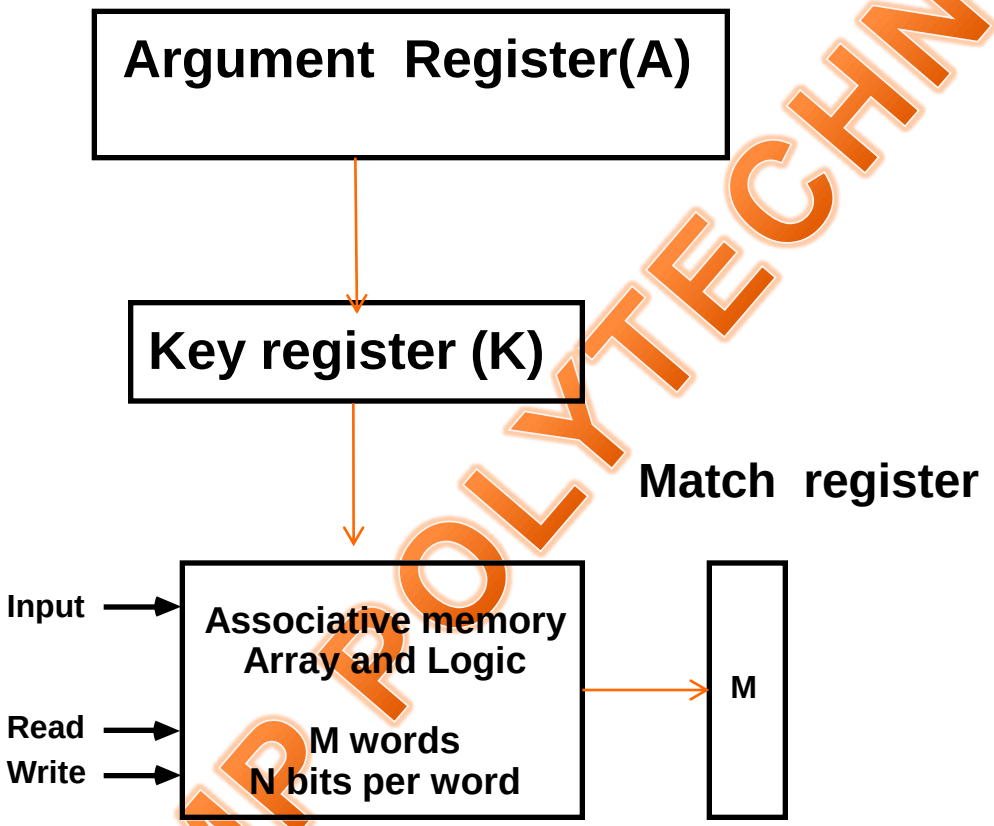
- For Searching Process:
  - Choose a sequence of **addresses**
  - **Read the content** of memory at each address
  - **Compare** the information with searched items.
- A memory unit accessed by content is called an **Associative Memory or CAM(Content Addressable Memory)**

CONT…

- When a **word is written** in an associative memory, no address is given.
- The memory is **capable of finding an empty unused location** to store the word.
- When the word is to be read from memory, the content of the word is specified.

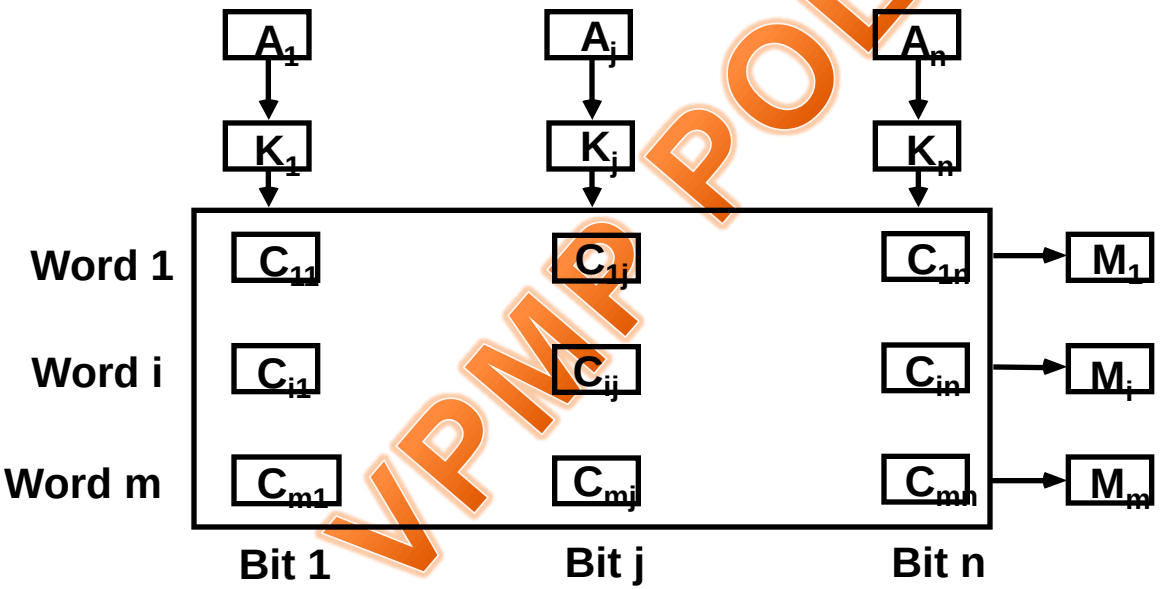
- Associative memory is **more Expensive than a RAM**.

CONT…



Block Diagram of Associative Memory

- The argument register A and key register have n bits.
- Match register M has m bits.
- Example: A 101 111100
  - K **111** 000000
  - Word1 100 111100 No match
  - Word2 101 000011 Match Occur



ORGANIZATION OF CAM

4.7 Cache Memory

- Cache memory is a **small, temporary fast memory**.
- It is placed between **CPU and Memory.**
- It runs at **speeds** similar to CPU Registers.
- **Access Time** is **less than** the access time of main memory.

CONT…

- Operation of Cache Memory:

1. When the CPU needs to access memory, the cache is examined.
2. If the word is found in the cache, It is read from that.
3. If the word is **not found**, **the main memory** is accessed to read the word and **block of words** are transferred from **main memory to cache memory** for future references.

CONT…

- **Performance Parameter of Cache Memory**

1. **Hit**: When the CPU refers to memory and **find the word in cache** , it is said to produce a HIT.
2. **Miss**: The requested data is **not found** in the cache memory, then it counts as a MISS.
3. **Hit Rate**: Performance of cache memory is measured in terms of quality, is called Hit Rate(Hit Ratio)

Hit Rate = No. of Hits / Total CPU references to Cache memory

Hit Rate= No. of Hits / (Hits + Misses)

CONT…

- **Miss Rate:** The percentage of memory accesses not found in a given level of memory.

  Miss Rate=1- Hit Rate

- **Hit Time:** The **time required to access the requested information** in a given level of memory.
- **Miss Penalty:** The time required to process a miss, which include replacing a block in memory plus the time required to deliver requested data to the processor.

CONT…

WRITING INTO CACHE

- **Read Operation**: When CPU finds a word in a Cache Memory, Main Memory is not involved.
- **Write Operation**: When CPU finds a word in a Cache Memory, Main Memory is involved.

Write Through Method:

- ✓ When Cache memory is update, the main memory is also updated in Parallel.
- ✓ Main Memory always content same data as cache.

CONT…

Write Back Method:

- Only the cache location is updated during a write.
- The location is then marked by flag.
- So, when the word is removed from cache, it is copied into memory.
- When the word is resides in the cache, it may be updated several times.

MEMORY AND CACHE MAPPING

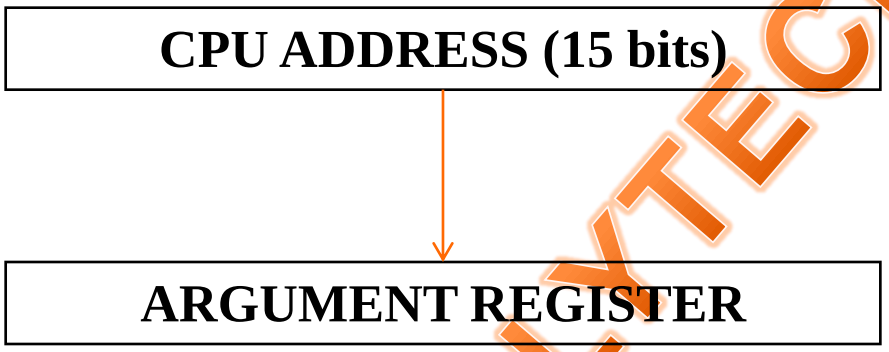
- The transformation of data from main memory to cache memory is referred to as a **mapping process.**
- Types of Mapping:
    1) **Associative Mapping**
    2) **Direct Mapping**
    3) **Set-associative Mapping**

MEMORY AND CACHE MAPPING

- Example

- CPU First sends 15 bit address to cache.($32 * 1024 = 2^{15}$)
- If There is a **hit**, CPU accepts 12 bit data from cache.
- If there is **miss**, the CPU reads the word from main memory and the word is transferred to cache.

MEMORY AND CACHE MAPPING - ASSOCIATIVE MAPPLING -

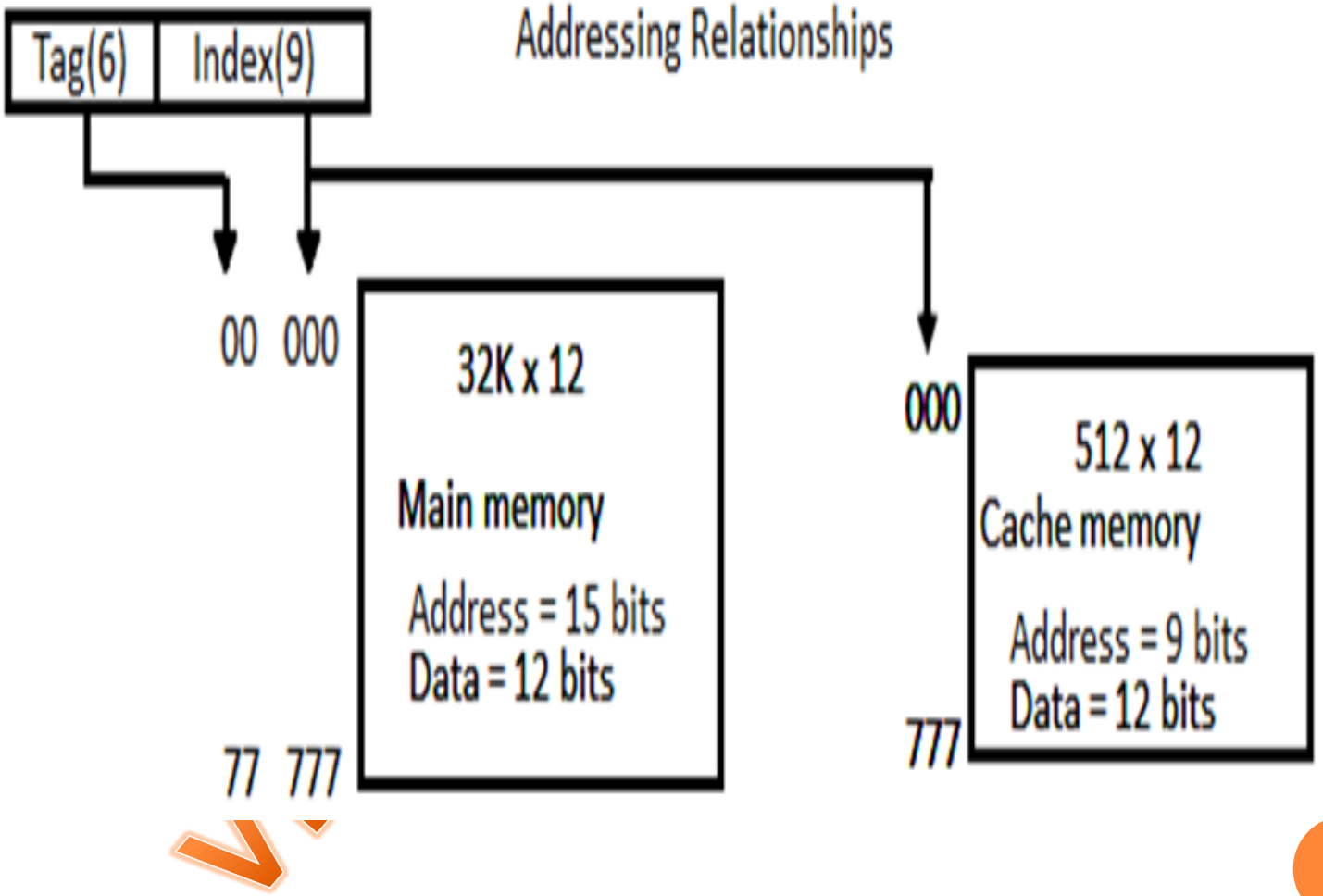


Address	Data
01000	3450
02777	6710
22345	1234
…	..
….	..

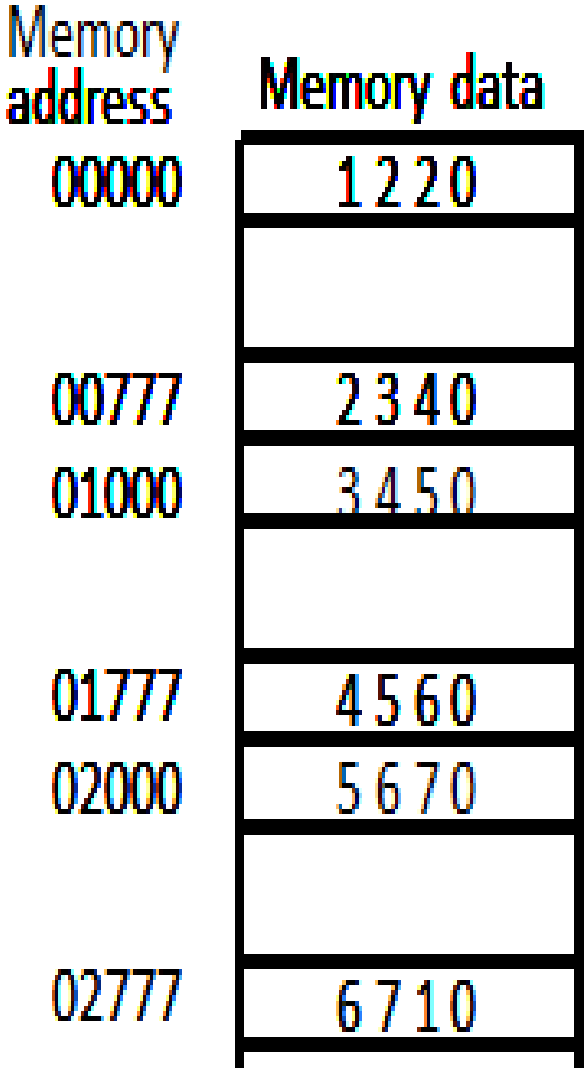
MEMORY AND CACHE MAPPING - ASSOCIATIVE MAPPLING -

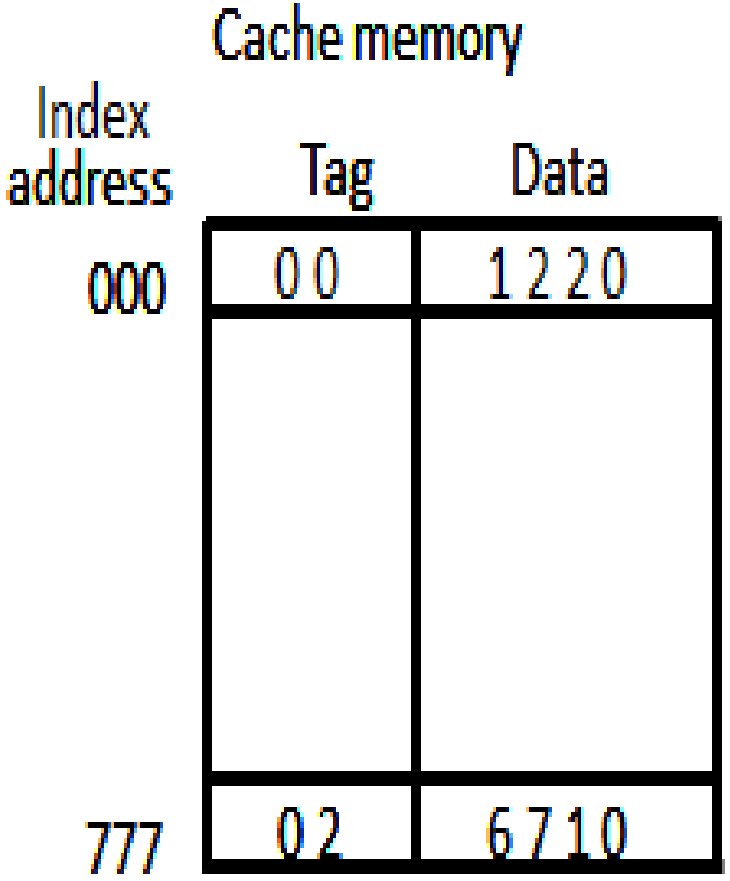
- The associative mapping stores both the **address and data** of the memory word.
- This permits **any location** in cache to store any word from main memory.
- **15 bits address is shown as a five digit octal number.**
- CPU Address of 15 bits is placed in **Argument Register.**
- Then associative memory is searched for a matching address.
- If the address is **found**, the **12 bit data** is read and sent to CPU.
- If **no match** occur, The main memory is accessed for the word.

MEMORY AND CACHE MAPPING - DIRECT MAPPING -



MEMORY AND CACHE MAPPING - DIRECT MAPPING -





MEMORY AND CACHE MAPPING - DIRECT MAPPING -

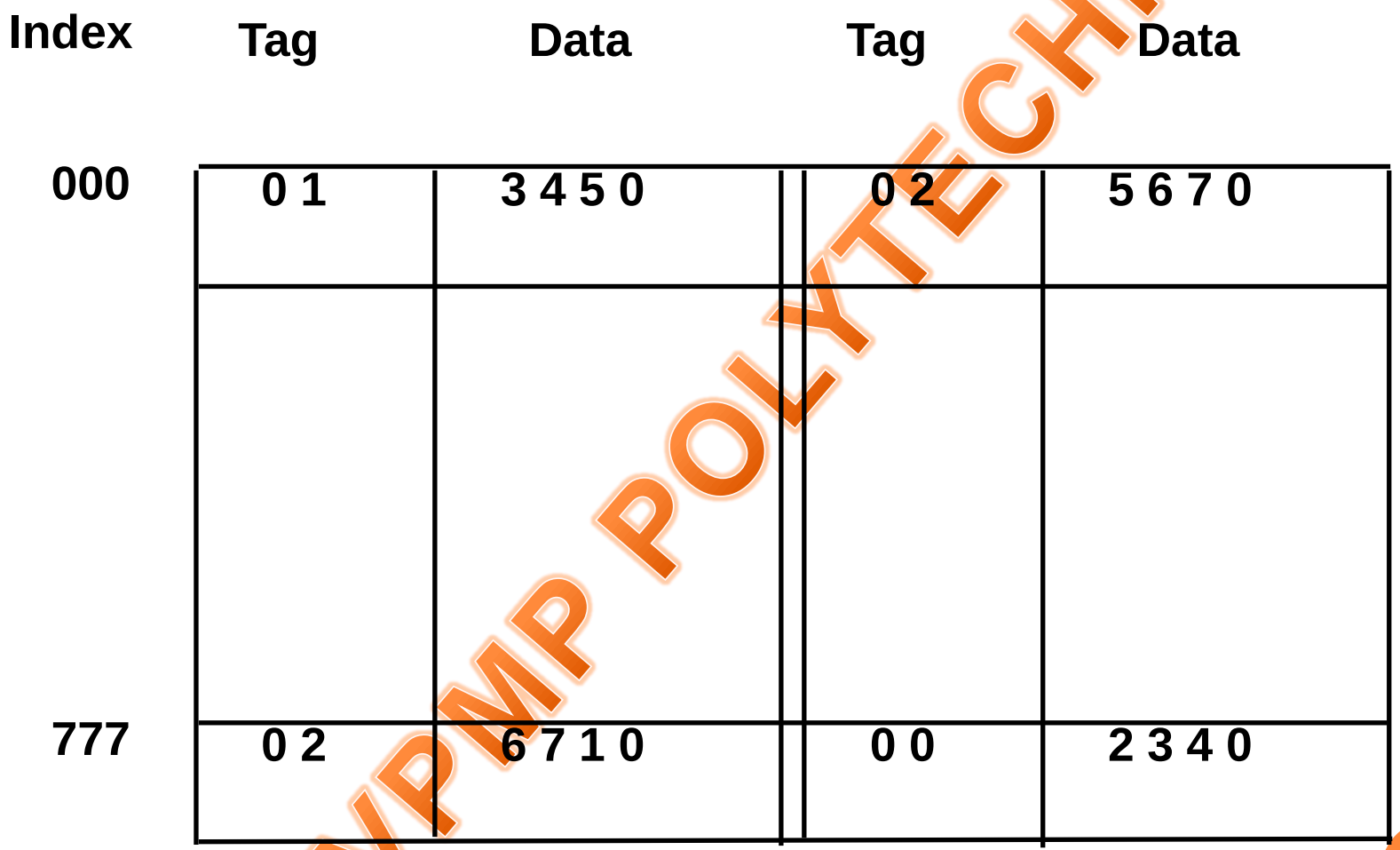
- Each word in cache consists of the **data word** and its **associated tag**.
- When the CPU generates a memory request, the **index field** is used as a **address to access the cache**.
- The **tag field** of the CPU address is compared with the **tag in the word read from the cache**.
- If the **two tags match**, there is a **HIT** and the desired data is in cache.
- If the **two tags do not match**, there is a **MISS** and desired word is read from main memory and then stored in the cache with new tag.

MEMORY AND CACHE MAPPING - DIRECT MAPPING

Disadvantage:

Two or more words with same index but different tags are accessed repeatedly.

MEMORY AND CACHE MAPPING - SET ASSOCIATIVE MAPPING



MEMORY AND CACHE MAPPING - SET ASSOCIATIVE MAPPING

- Each word of cache can **store two or more words of memory under the same index.**
- Each data word is **stored** together **with its tag**.
- **Each index refers to two data words and their Associated tags.**

MEMORY AND CACHE MAPPING - SET ASSOCIATIVE MAPPING

- Each tag require six bits and each data word has 12 bits.
- So, the word length is 2*(6+12)=36 bits.
- The size of the cache memory is 512*36.
- The word at addresses 01000 and 02000 of Main Memory are stored in cache memory at index address 000.
- Similarly, the words at addresses 02777 and 00777 are stored in cache at address 777.
- When the CPU generates a memory request, the **index value** of the address is used to access the **cache.**

MEMORY AND CACHE MAPPING - SET ASSOCIATIVE MAPPING

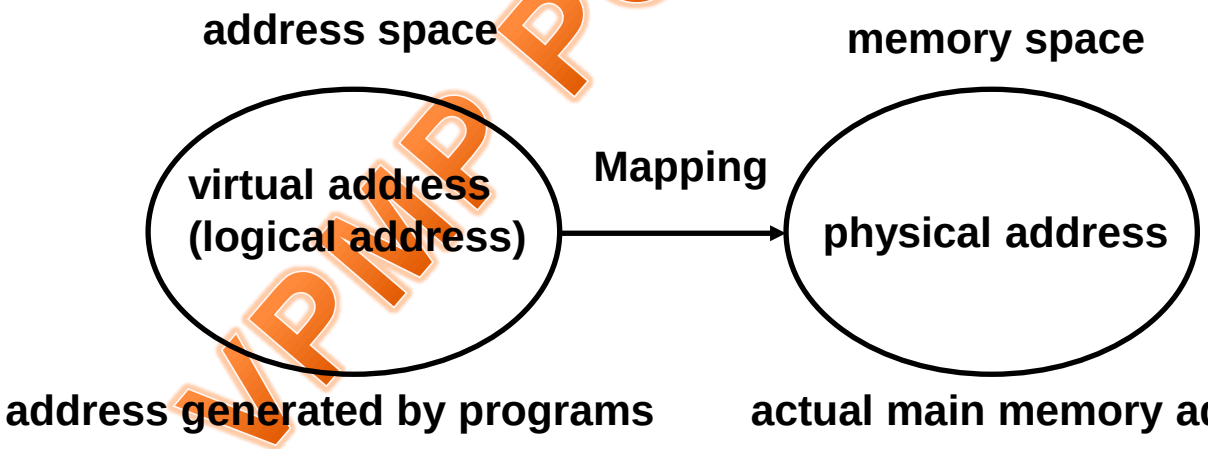
- The **tag field of the CPU address** is then compared with **both tags** in the cache.

- When a **miss** occurs and **cache is full** then, it is necessary **to replace one of the tag data items with a new value.**

4.8 VIRTUAL MEMORY

- It is a concept used in a large computer system.

- Give the **programmer** the illusion that the **system has a very large memory**, even though the **computer actually has a relatively small main memory.**



CONT…

- **Virtual Address**: An address used by a programmer.
- **Address Space**: Set of Virtual Address.
- **Physical Address**: An address in Main Memory.
- **Memory Space:** Set of Physical Address.

CONT…

- Example:
- Main Memory: 32K→32 * 1024=2^5*2^{10} = 2^{15} bits

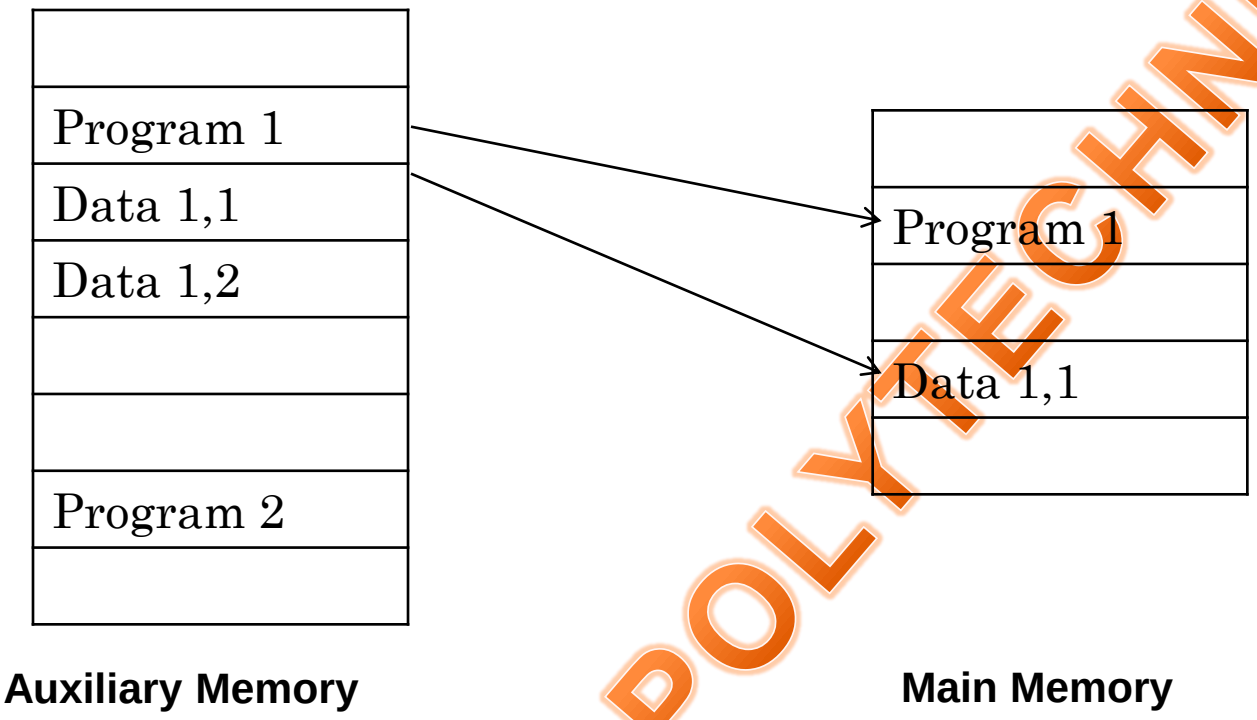
Physical Address = 15 bits

- Auxiliary Memory: 1024 K→1024 *1024=2^{20} bits

Virtual Address= 20 bits

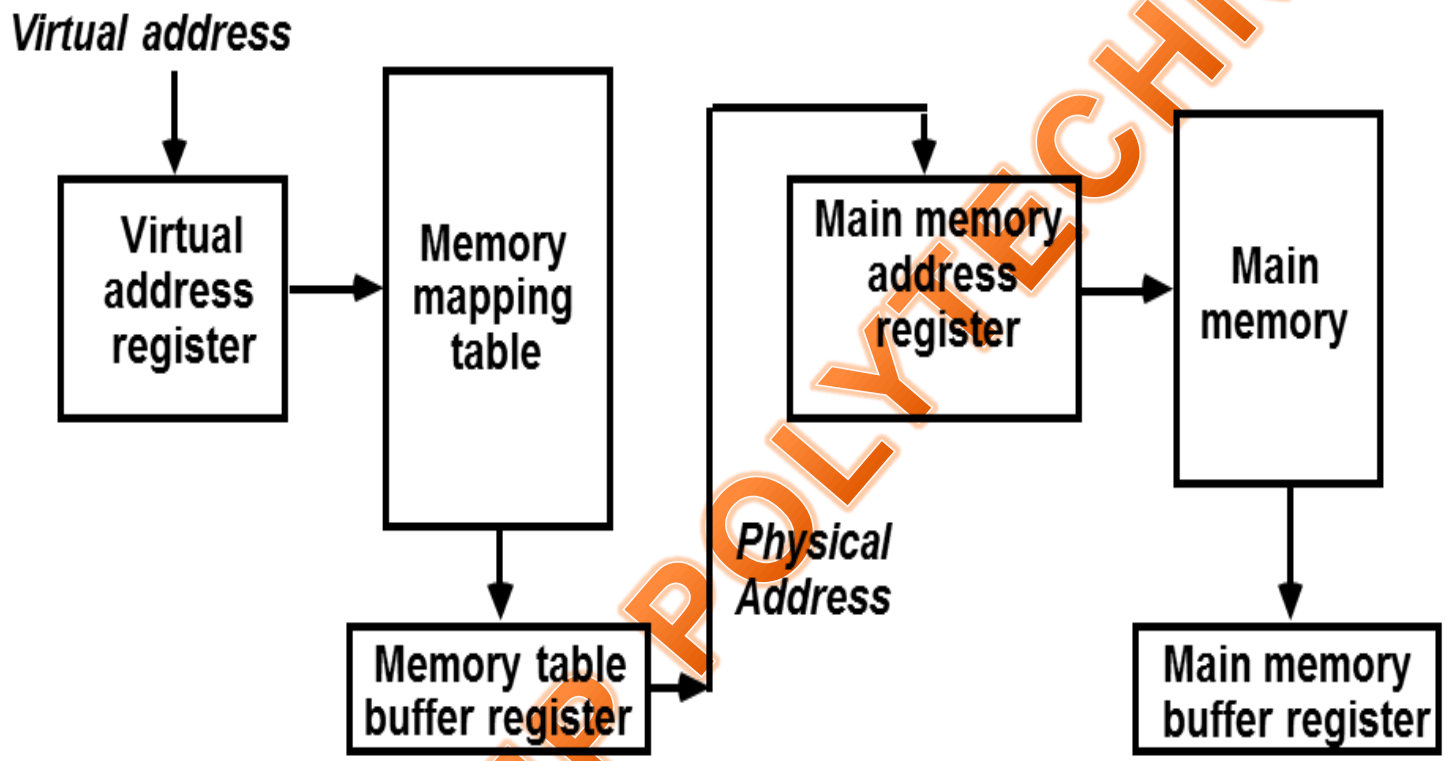
- In a multiprogramming Computer System, **programs and data are transferred to and from Main Memory and Auxiliary Memory.**

CONT…



- **Address field** of **instruction code** is 20 bits and **physical memory** is 15 bits. So, a **table** is needed to map a virtual address of 20 bits to a physical address of 15 bits.

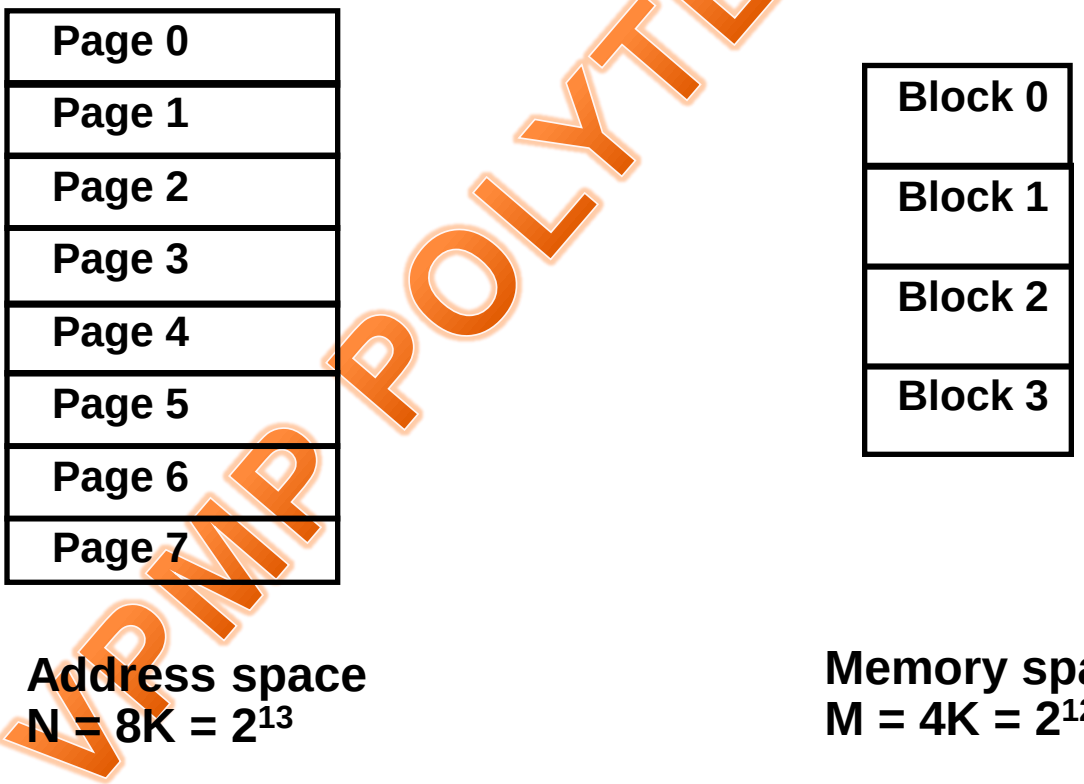
CONT…



CONT…

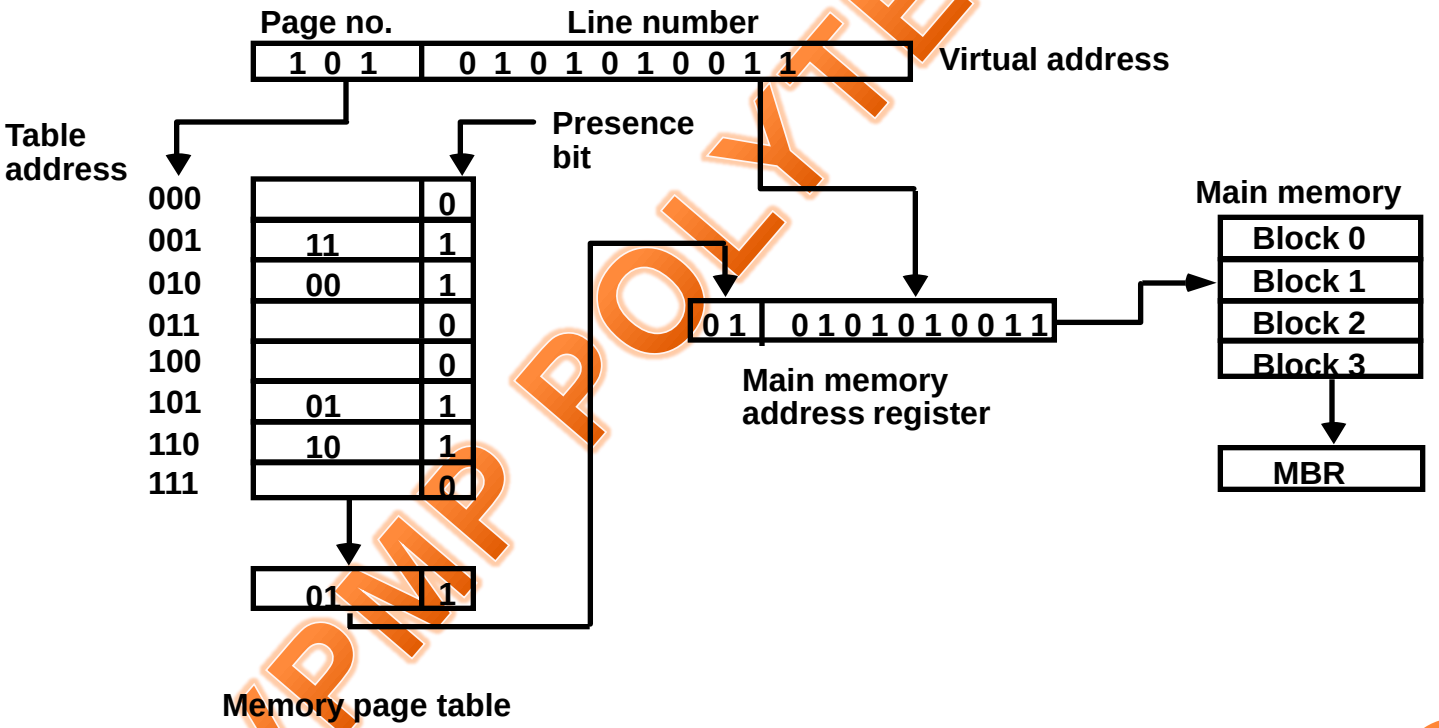
Address Space and Memory Space are each divided into fixed size group of words called blocks or pages

1K words group



CONT…

Organization of memory Mapping Table in a paged system



CONT…

- The table shows that pages 1,2,5,6 are available in memory.
- **Three High order bits of virtual address** specify a **page number** and also an address for the memory page table.
- The content of the word in the memory page table at the page number address is read out into **memory table buffer register**.
- If the presence bit is 1, the block number is trasferred to the two high order bits of Main Memory Address Register.