MODULE 4: MEMORY SYSTEM

BASIC CONCEPTS

• Maximum size of memory that can be used in any computer is determined by addressing mode.

Address	Memory Locations
16 Bit	$2^{16} = 64$ K
32 Bit	$2^{32} = 4$G (Giga)
40 Bit	$2^{40} = 1$T (Tera)

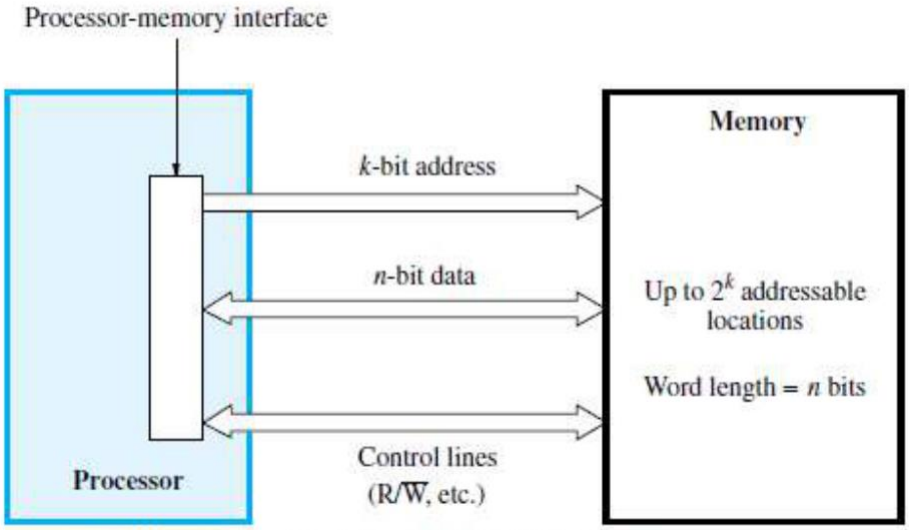

Figure 8.1 Connection of the memory to the processor.

• If MAR is k-bits long then
→ memory may contain upto 2^K addressable-locations
• If MDR is n-bits long, then
→ n-bits of data are transferred between the memory and processor.
• The data-transfer takes place over the processor-bus (Figure 8.1).
• The processor-bus has
1) Address-Line
2) Data-line &
3) Control-Line (R/W", MFC – Memory Function Completed).
• The Control-Line is used for coordinating data-transfer.
• The processor reads the data from the memory by
→ loading the address of the required memory-location into MAR and
→ setting the R/W" line to 1.
• The memory responds by
→ placing the data from the addressed-location onto the data-lines and
→ confirms this action by asserting MFC signal.
• Upon receipt of MFC signal, the processor loads the data from the data-lines into MDR.
• The processor writes the data into the memory-location by
→ loading the address of this location into MAR &
→ setting the R/W" line to 0.
• **Memory Access Time:** It is the time that elapses between
→ initiation of an operation &
→ completion of that operation.
• **Memory Cycle Time:** It is the minimum time delay that required between the initiation of the two successive memory-operations.

RAM (Random Access Memory)

- In RAM, any location can be accessed for a Read/Write-operation in fixed amount of time,

Cache Memory

- It is a small, fast memory that is inserted between
  - → larger slower main-memory and
  - → processor.
- It holds the currently active segments of a program and their data.

- The address generated by the processor is referred to as a **virtual/logical address**.
- The virtual-address-space is mapped onto the physical-memory where data are actually stored.
- The mapping-function is implemented by MMU. (MMU = memory management unit).
- Only the active portion of the address-space is mapped into locations in the physical-memory.
- The remaining virtual-addresses are mapped onto the bulk storage devices such as magnetic disk.
- As the active portion of the virtual-address-space changes during program execution, the MMU
  - → changes the mapping-function &
  - → transfers the data between disk and memory.
- During every memory-cycle, MMU determines whether the addressed-page is in the memory. If the page is in the memory.
  - Then, the proper word is accessed and execution proceeds.
  - Otherwise, a page containing desired word is transferred from disk to memory.

- Memory can be classified as follows:
  1) RAM which can be further classified as follows:
     i) Static RAM
     ii) Dynamic RAM (DRAM) which can be further classified as synchronous & asynchronous DRAM.
  2) ROM which can be further classified as follows:
     i) PROM
     ii) EPROM
     iii) EEPROM &
     iv) Flash Memory which can be further classified as Flash Cards & Flash Drives.

SEMI CONDUCTOR RAM MEMORIES

INTERNAL ORGANIZATION OF MEMORY-CHIPS

- Memory-cells are organized in the form of array (Figure 8.2).
- Each cell is capable of storing 1-bit of information.
- Each row of cells forms a memory-word.
- All cells of a row are connected to a common line called as **Word-Line**.
- The cells in each column are connected to **Sense/Write** circuit by 2-bit-lines.
- The Sense/Write circuits are connected to data-input or output lines of the chip.
- During a write-operation, the sense/write circuit
  - → receive input information &
  - → store input info in the cells of the selected word.

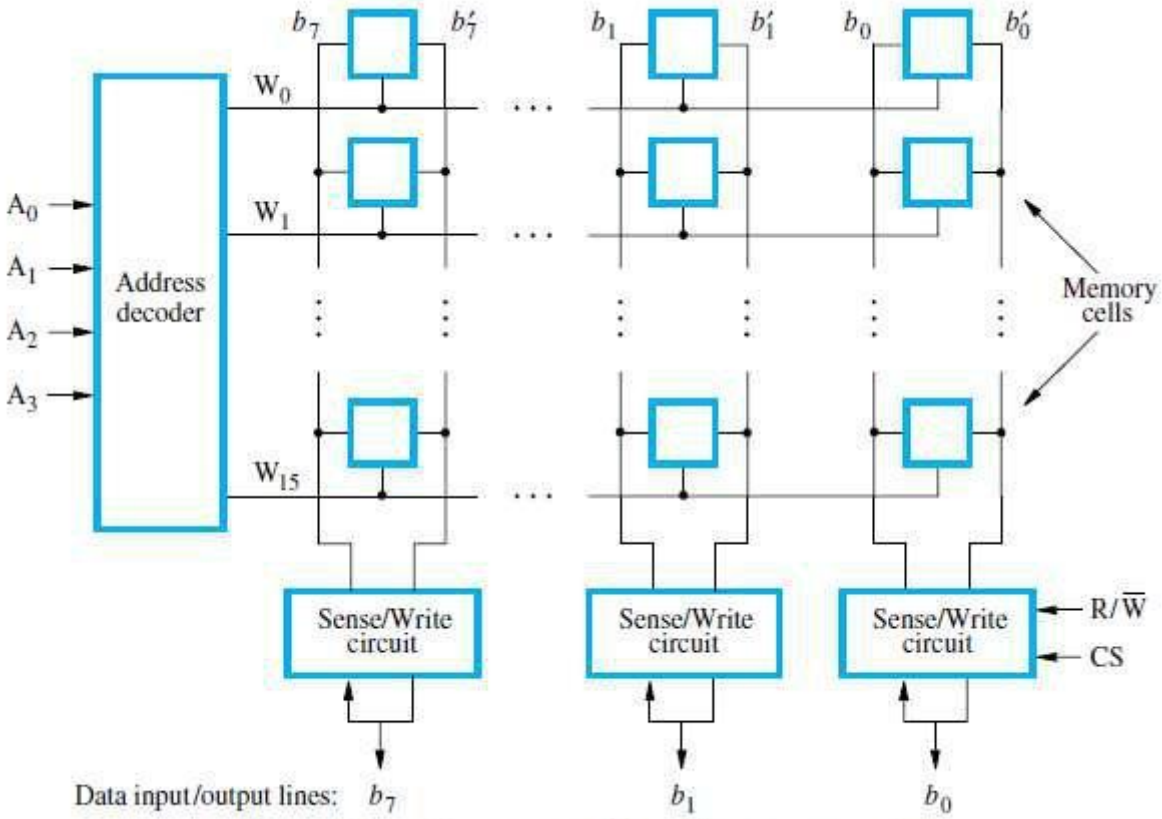

Figure 8.2 Organization of bit cells in a memory chip.

- The data-input and data-output of each Sense/Write circuit are connected to a single bidirectional data-line.
- Data-line can be connected to a data-bus of the computer.
- Following 2 control lines are also used:
  - **1) R/W'** → Specifies the required operation.
  - **2) CS'** → Chip Select input selects a given chip in the multi-chip memory-system.

Bit Organization	Requirement of external connection for address, data and control lines
128 (16x8)	14
(1024) 128x8(1k)	19

STATIC RAM (OR MEMORY)

• Memories consist of circuits capable of retaining their state as long as power is applied are known.

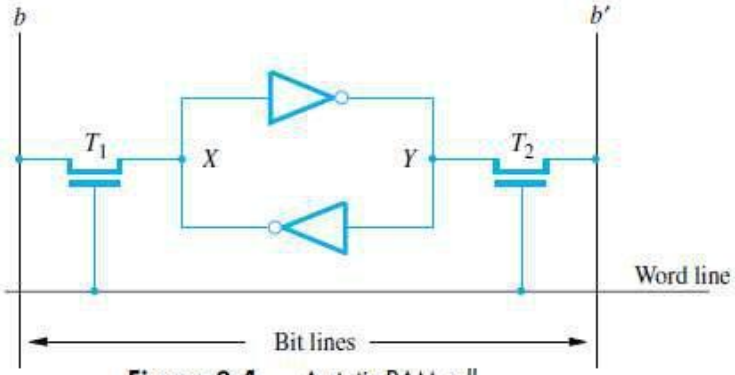

Figure 8.4 A static RAM cell.

• Two inverters are cross connected to form a latch (Figure 8.4).
• The latch is connected to 2-bit-lines by transistors T_1 and T_2.
• The transistors act as switches that can be opened/closed under the control of the word-line.
• When the word-line is at ground level, the transistors are turned off and the latch retain its state.

Read Operation

• To read the state of the cell, the word-line is activated to close switches T_1 and T_2.
• If the cell is in state 1, the signal on bit-line b is high and the signal on the bit-line b" is low.
• Thus, b and b" are complement of each other.
• Sense/Write circuit
  → monitors the state of b & b" and
  → sets the output accordingly.

• The state of the cell is set by
  → placing the appropriate value on bit-line b and its complement on b" and
  → then activating the word-line. This forces the cell into the corresponding state.
• The required signal on the bit-lines is generated by Sense/Write circuit.

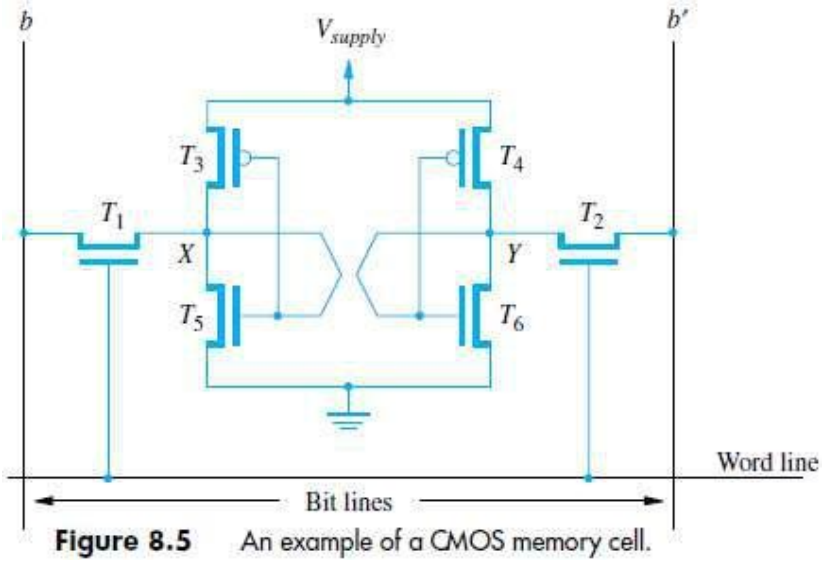

Figure 8.5 An example of a CMOS memory cell.

CMOS Cell

• Transistor pairs (T_3, T_5) and (T_4, T_6) form the inverters in the latch (Figure 8.5).
• In state 1, the voltage at point X is high by having T_5, T_6 ON and T_4, T_5 are OFF.
• Thus, T_1 and T_2 returned ON (Closed), bit-line b and b" will have high and low signals respectively.
• **Advantages:**
  1) It has low power consumption „." the current flows in the cell only when the cell is active.
  2) Static RAM"s can be accessed quickly. It access time is few nanoseconds.
• **Disadvantage:** SRAMs are said to be volatile memories „." their contents are lost when power is interrupted.

ASYNCHRONOUS DRAM

- Less expensive RAMs can be implemented if simple cells are used.
- Such cells cannot retain their state indefinitely. Hence they are called **Dynamic RAM (DRAM).**
- The information stored in a dynamic memory-cell in the form of a charge on a capacitor.
- This charge can be maintained only for tens of milliseconds.
- The contents must be periodically refreshed by restoring this capacitor charge to its full value.

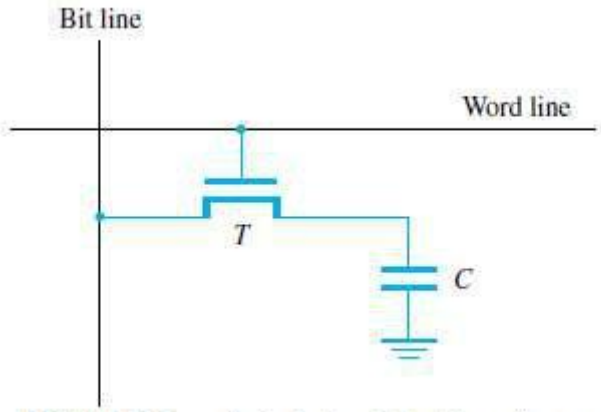

Figure 8.6 A single-transistor dynamic memory cell.

- In order to store information in the cell, the transistor T is turned „ON" (Figure 8.6).
- The appropriate voltage is applied to the bit-line which charges the capacitor.
- After the transistor is turned off, the capacitor begins to discharge.
- Hence, info. stored in cell can be retrieved correctly before threshold value of capacitor drops down.
- During a read-operation,
  - → transistor is turned „ON"
  - → a sense amplifier detects whether the charge on the capacitor is above the threshold value.
    - If (charge on capacitor) > (threshold value) → Bit-line will have logic value „1".
    - If (charge on capacitor) < (threshold value) → Bit-line will set to logic value „0".

ASYNCHRONOUS DRAM DESCRIPTION

- The 4 bit cells in each row are divided into 512 groups of 8 (Figure 5.7).
- 21 bit address is needed to access a byte in the memory. 21 bit is divided as follows:
  1) 12 address bits are needed to select a row. i.e.
     A_{8-0} → specifies row-address of a byte.
  2) 9 bits are needed to specify a group of 8 bits in the selected row.
     i.e. A_{20-9} → specifies column-address of a byte.

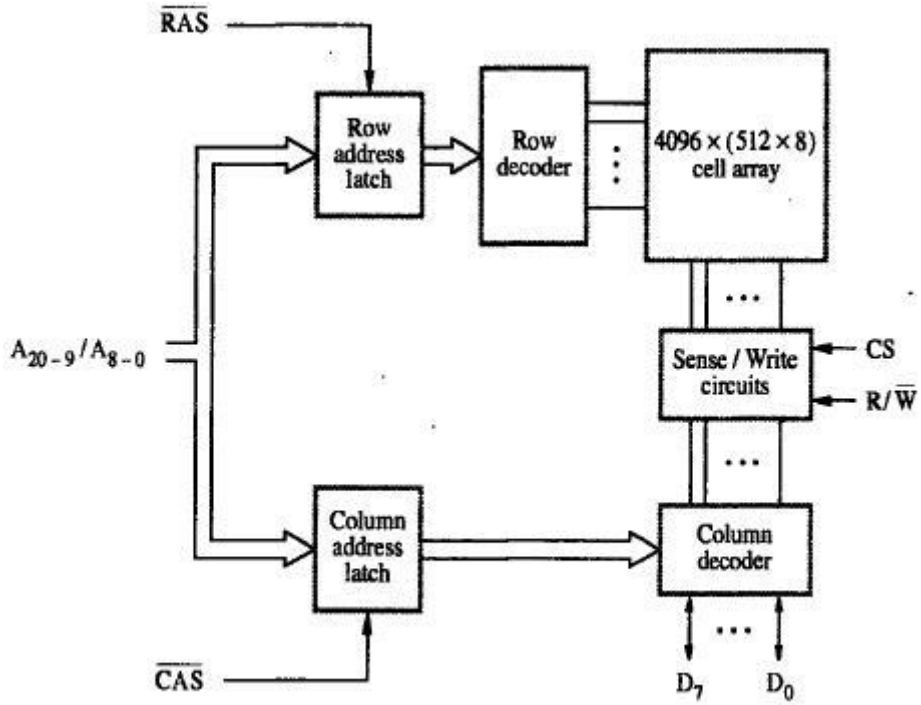

Figure 5.7 Internal organization of a 2M × 8 dynamic memory chip

- During Read/Write-operation,
  → row-address is applied first.
  → row-address is loaded into row-latch in response to a signal pulse on **RAS'** input of chip.
  (RAS = Row-address Strobe CAS = Column-address Strobe)
- When a Read-operation is initiated, all cells on the selected row are read and refreshed.
- Shortly after the row-address is loaded, the column-address is
  → applied to the address pins &
  → loaded into **CAS'.**
- The information in the latch is decoded.
- The appropriate group of 8 Sense/Write circuits is selected.
  R/W'=1(read-operation) → Output values of selected circuits are transferred to data-lines D_0-D_7.
  R/W'=0(write-operation) → Information on D_0-D_7 are transferred to the selected circuits.
- RAS" & CAS" are active-low so that they cause latching of address when they change from high to low.
- To ensure that the contents of DRAMs are maintained, each row of cells is accessed periodically.
- A special memory-circuit provides the necessary control signals RAS" & CAS" that govern the timing.
- The processor must take into account the delay in the response of the memory.

  - Transferring the bytes in sequential order is achieved by applying the consecutive sequence of column-address under the control of successive CAS" signals.
  - This scheme allows transferring a block of data at a faster rate.
  - The block of transfer capability is called as *fast page mode***.**

READ ONLY MEMORY (ROM)

- Both SRAM and DRAM chips are volatile, i.e. They lose the stored information if power is turned off.
- Many application requires non-volatile memory which retains the stored information if power is turned off.
- For ex:
  OS software has to be loaded from disk to memory i.e. it requires non-volatile memory.
- Non-volatile memory is used in embedded system.
- Since the normal operation involves only reading of stored data, a memory of this type is called ROM.
  - **At Logic value '0'** → Transistor(T) is connected to the ground point (P).
    Transistor switch is closed & voltage on bit-line nearly drops to zero (Figure 8.11).
  - **At Logic value '1'** → Transistor switch is open.
    The bit-line remains at high voltage.

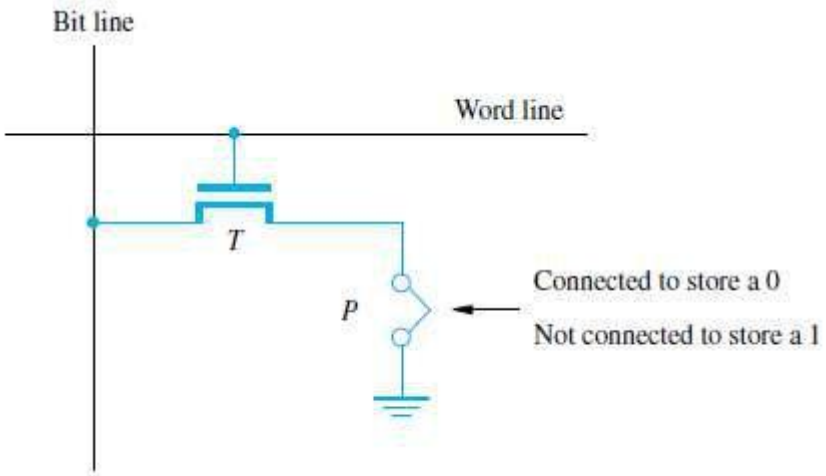

Figure 8.11 A ROM cell.

- To read the state of the cell, the word-line is activated.
- A Sense circuit at the end of the bit-line generates the proper output value.

TYPES OF ROM

- Different types of non-volatile memory are
  1) PROM
  2) EPROM
  3) EEPROM &
  4) Flash Memory (Flash Cards & Flash Drives)

PROM (PROGRAMMABLE ROM)

- PROM allows the data to be loaded by the user.
- Programmability is achieved by inserting a „fuse" at point P in a ROM cell.
- Before PROM is programmed, the memory contains all 0"s.
- User can insert 1"s at required location by burning-out fuse using high current-pulse.
- This process is irreversible.
- **Advantages:**
  1) It provides flexibility.
  2) It is faster.
  3) It is less expensive because they can be programmed directly by the user.

EPROM (ERASABLE REPROGRAMMABLE ROM)

- EPROM allows
  - → stored data to be erased and
  - → new data to be loaded.
- In cell, a connection to ground is always made at „P" and a special transistor is used.
- The transistor has the ability to function as
  - → a normal transistor or
  - → a disabled transistor that is always turned „off".
- Transistor can be programmed to behave as a permanently open switch, by injecting charge into it.
- Erasure requires dissipating the charges trapped in the transistor of memory-cells.

- **Advantages:**
  1) It provides flexibility during the development-phase of digital-system.
  2) It is capable of retaining the stored information for a long time.
- **Disadvantages:**
  1) The chip must be physically removed from the circuit for reprogramming.
  2) The entire contents need to be erased by UV light.

EEPROM (ELECTRICALLY ERASABLE ROM)

- **Advantages:**
  1) It can be both programmed and erased electrically.
  2) It allows the erasing of all cell contents selectively.
- **Disadvantage:** It requires different voltage for erasing, writing and reading the stored data.

FLASH MEMORY

- In EEPROM, it is possible to read & write the contents of a single cell.
- In Flash device, it is possible to read contents of a single cell & write entire contents of a block.
- Prior to writing, the previous contents of the block are erased.

- Single flash chips cannot provide sufficient storage capacity for embedded-system.
- **Advantages:**
  1) Flash drives have greater density which leads to higher capacity & low cost per bit.
  2) It requires single power supply voltage & consumes less power.
- There are 2 methods for implementing larger memory: 1) Flash Cards & 2) Flash Drives

1) Flash Cards

- One way of constructing larger module is to mount flash-chips on a small card.
- Such flash-card have standard interface.
- The card is simply plugged into a conveniently accessible slot.
- Memory-size of the card can be 8, 32 or 64MB.
- Eg: A minute of music can be stored in 1MB of memory. Hence 64MB flash cards can store an hour of music.

2) Flash Drives

- Larger flash memory can be developed by replacing the hard disk-drive.
- The flash drives are designed to fully emulate the hard disk.
- The flash drives are solid state electronic devices that have no movable parts.

1) hey have shorter seek & access time which results in faster response.
2) hey have low power consumption. .". they are attractive for battery driven application.
3) They are insensitive to vibration.

1) The capacity of flash drive (<1GB) is less than hard disk (>1GB).
2) It leads to higher cost per bit.
3) Flash memory will weaken after it has been written a number of times (typically at least 1 million times).

CACHE MEMORIES

• The effectiveness of cache mechanism is based on the property of „**Locality of Reference'.**

Locality of Reference

• Many instructions in the localized areas of program are executed repeatedly during some time period
• Remainder of the program is accessed relatively infrequently (Figure 8.15).
• There are 2 types:

➢ The recently executed instructions are likely to be executed again very soon.

2) Spatial

➢ Instructions in close proximity to recently executed instruction are also likely to be executed soon.

• If active segment of program is placed in cache-memory, then total execution time can be reduced.
• **Block** refers to the set of contiguous address locations of some size.
• The cache-line is used to refer to the cache-block.

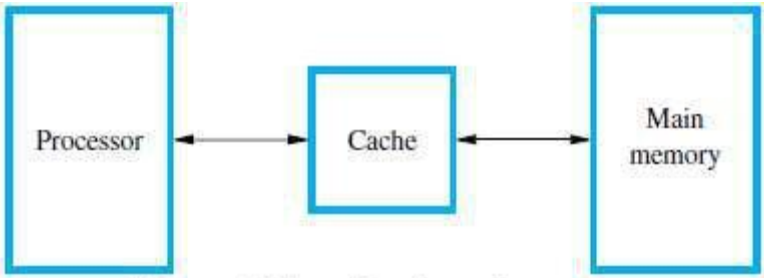

Figure 8.15 Use of a cache memory.

• The Cache-memory stores a reasonable number of blocks at a given time.
• This number of blocks is small compared to the total number of blocks available in main-memory.
• Correspondence b/w main-memory-block & cache-memory-block is specified by mapping-function.
• Cache control hardware decides which block should be removed to create space for the new block.
• The collection of rule for making this decision is called the **Replacement Algorithm.**
• The cache control-circuit determines whether the requested-word currently exists in the cache.
• The write-operation is done in 2 ways: 1) Write-through protocol & 2) Write-back protocol.

➢ Here the cache-location and the main-memory-locations are updated simultaneously.

Write-Back Protocol

➢ This technique is to
→ update only the cache-location &
→ mark the cache-location with associated flag bit called **Dirty/Modified Bit.**
➢ The word in memory will be updated later, when the marked-block is removed from cache.

During Read-operation

• If the requested-word currently not exists in the cache, then **read-miss** will occur.
• To overcome the read miss, *Load–through/Early restart protocol* is used.

➢ The block of words that contains the requested-word is copied from the memory into cache.
➢ After entire block is loaded into cache, the requested-word is forwarded to processor.

During Write-operation

• If the requested-word not exists in the cache, then **write-miss** will occur.
1) If **Write Through Protocol** is used, the information is written directly into main-memory.
2) If **Write Back Protocol** is used,
→ then block containing the addressed word is first brought into the cache &
→ then the desired word in the cache is over-written with the new information.

VIRTUAL MEMORY

• It refers to a technique that automatically move program/data blocks into the main-memory when they are required for execution (Figure 8.24).
• The address generated by the processor is referred to as a **virtual/logical address**.
• The virtual-address is translated into physical-address by **MMU** (Memory Management Unit).
• During every memory-cycle, MMU determines whether the addressed-word is in the memory.
    If the word is in memory.
        Then, the word is accessed and execution proceeds.
        Otherwise, a page containing desired word is transferred from disk to memory.
• Using DMA scheme, transfer of data between disk and memory is performed.

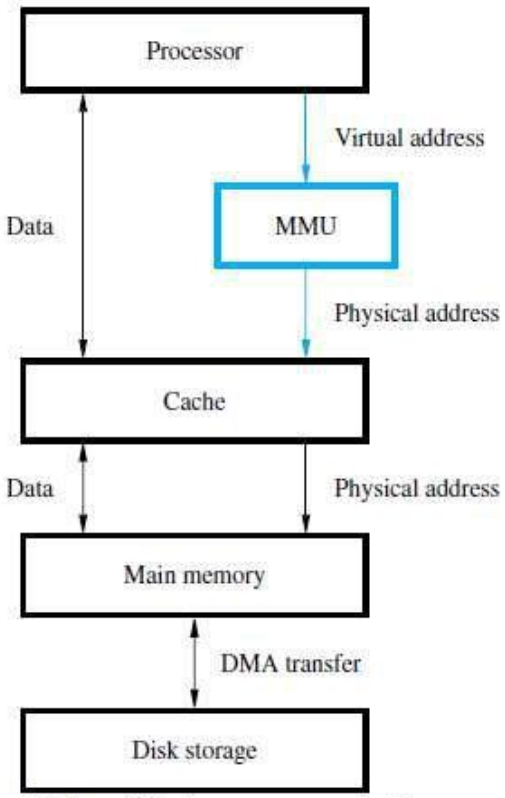

Figure 8.24 Virtual memory organization.

SECONDARY-STORAGE

• The semi-conductor memories do not provide all the storage capability.
• The secondary-storage devices provide larger storage requirements.
• Some of the secondary-storage devices are:
    1) Magnetic Disk
    2) Optical Disk &
    3) Magnetic Tapes.

MAGNETIC DISK

• Magnetic Disk system consists of one or more **disk** mounted on a common **spindle**.
• A **thin magnetic film** is deposited on each disk (Figure 8.27).
• Disk is placed in a **rotary-drive** so that magnetized surfaces move in close proximity to R/W heads.
• Each **R/W head** consists of 1) Magnetic Yoke & 2) Magnetizing-Coil.
• Digital information is stored on magnetic film by applying current pulse to the magnetizing-coil.
• Only changes in the magnetic field under the head can be sensed during the Read-operation.
• Therefore, if the binary states 0 & 1 are represented by two opposite states,
    then a voltage is induced in the head only at 0-1 and at 1-0 transition in the bit stream.
• A consecutive of 0"s & 1"s are determined by using the clock.
• **Manchester Encoding** technique is used to combine the clocking information with data.

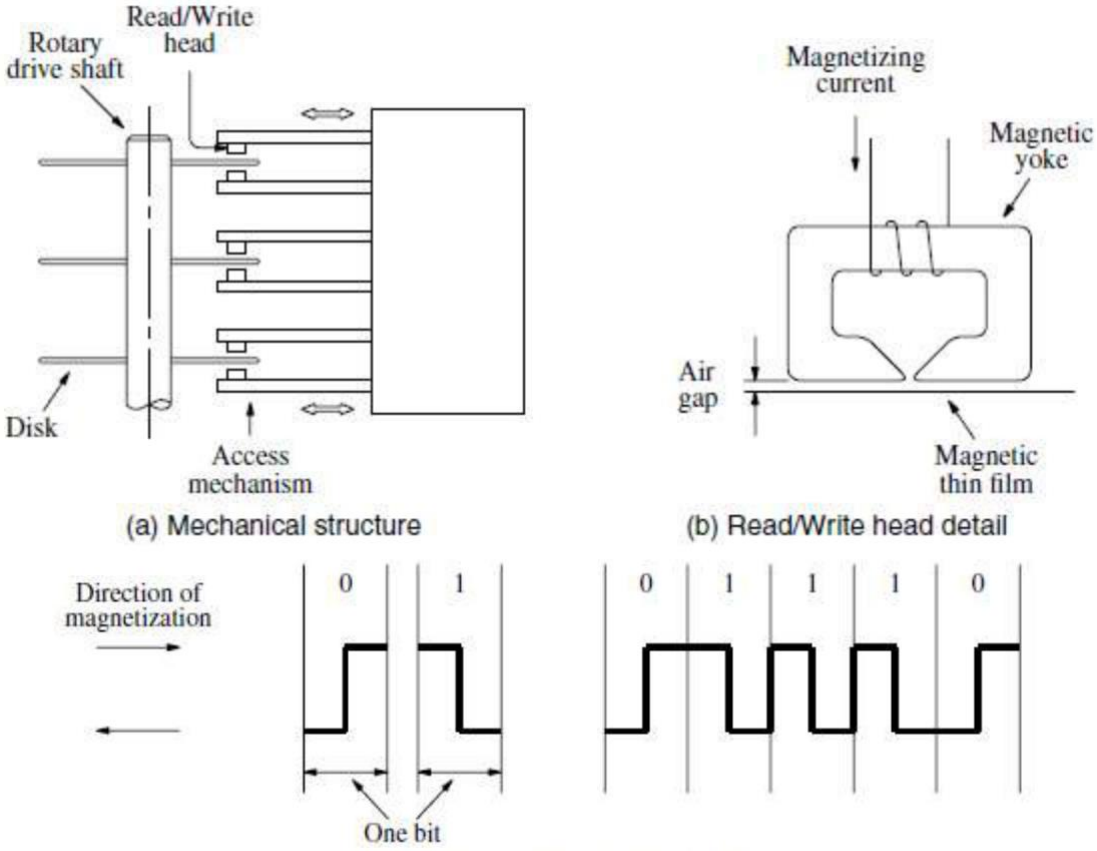

Figure 8.27 Magnetic disk principles.

- R/W heads are maintained at small distance from disk-surfaces in order to achieve high bit densities.
- When disk is moving at their steady state, the air pressure develops b/w disk-surfaces & head.

- The flexible spring connection between head and its arm mounting permits the head to fly at the desired distance away from the surface.

Winchester technology

- Read/Write heads are placed in a sealed, air–filtered enclosure called the Winchester Technology. The read/write heads can operate closure to magnetic track surfaces because the dust particles which are a problem in unsealed assemblies are absent.

Advantages

- It has a larger capacity for a given physical size.
- The data intensity is high because
  the storage medium is not exposed to contaminating elements.
- The read/write heads of a disk system are movable.
- The disk system has 3 parts:
  1) Disk Platter (Usually called Disk)
  2) Disk-drive (spins the disk & moves Read/write heads)
  3) Disk Controller (controls the operation of the system.)

ORGANIZATION & ACCESSING OF DATA ON A DISK

- Each surface is divided into concentric **Tracks** (Figure 8.28).
- Each track is divided into **Sectors.**
- The set of corresponding tracks on all surfaces of a stack of disk form a **Logical Cylinder***.*
- The data are accessed by specifying *the surface number, track number and the sector number.*
- The Read/Write-operation start at sector boundaries.
- Data bits are stored serially on each track.

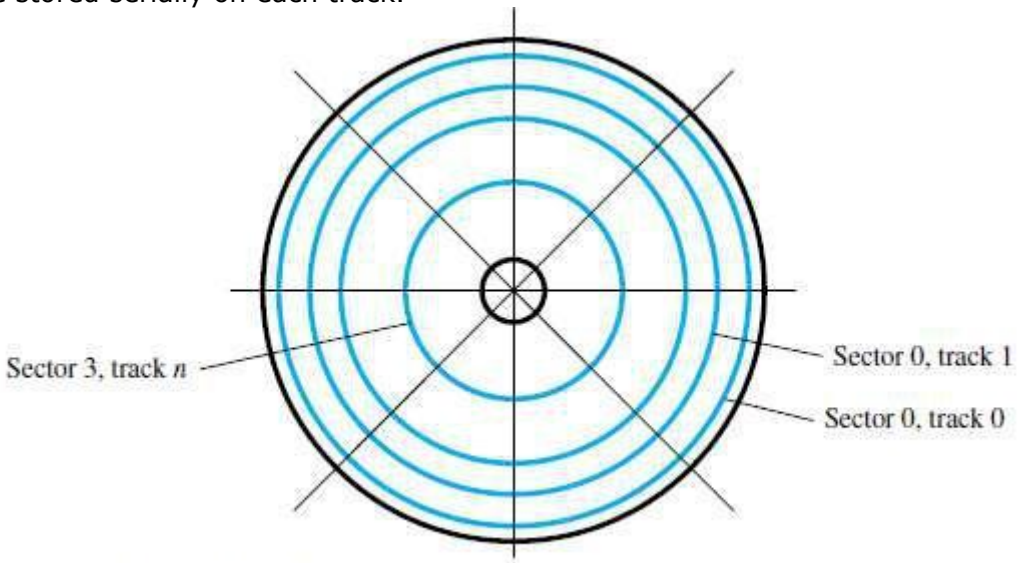

Figure 8.28 Organization of one surface of a disk.

- Each sector usually contains 512 bytes.
- **Sector Header** --> contains identification information.

- **ECC** (Error checking code)- is used to detect and correct errors.
- An unformatted disk has no information on its tracks.
- The formatting process divides the disk physically into tracks and sectors.
- The formatting process may discover some defective sectors on all tracks.
- **Disk Controller** keeps a record of various defects.
- The disk is divided into logical partitions:
  1) Primary partition
  2) Secondary partition
- Each track has same number of sectors. So, all tracks have same storage capacity.
- Thus, the stored information is packed more densely on inner track than on outer track.

Access Time

- There are 2 components involved in the time-delay:

  1) Seek time: Time required to move the read/write head to the proper track.

  2) Latency/Rotational Delay: The amount of time that elapses after head is positioned over the correct track until the starting position of the addressed sector passes under the R/W head.

  Seek time + Latency = Disk access time

  Typical Disk

  - One inch disk-weight = 1 ounce, size -> comparable to match book
    Capacity -> 1GB
  - Inch disk has the following
    parameter Recording
    surface=20 Tracks=15000
    tracks/surface Sectors=400.
    Each sector stores 512 bytes of data

    Capacity of formatted disk=20x15000x400x512=60x109 =60GB
    Seek time=3ms
    Platter rotation=10000 rev/min
    Latency=3ms
    Internet transfer rate=34MB/s

DATA BUFFER/CACHE

- A disk-drive that incorporates the required SCSI circuit is referred as **SCSI Drive**.
- The SCSI can transfer data at higher rate than the disk tracks.
- A data buffer can be used to deal with the possible difference in transfer rate b/w disk and SCSI bus
- The buffer is a semiconductor memory.
- The buffer can also provide cache mechanism for the disk.

i.e. when a read request arrives at the disk, then controller first check if the data is available in the cache/buffer.

If data is available in cache.

Then, the data can be accessed & placed on SCSI bus.

Otherwise, the data will be retrieved from the disk.

DISK CONTROLLER

- The disk controller acts as interface between disk-drive and system-bus (Figure 8.13).
- The disk controller uses DMA scheme to transfer data between disk and memory.
- When the OS initiates the transfer by issuing R/W" request, the controllers register will load the following information:

1) Memory Address: Address of first memory-location of the block of words involved in the transfer.

2) Disk Address: Location of the sector containing the beginning of the desired block of words.

3) Word Count: Number of words in the block to be transferred.

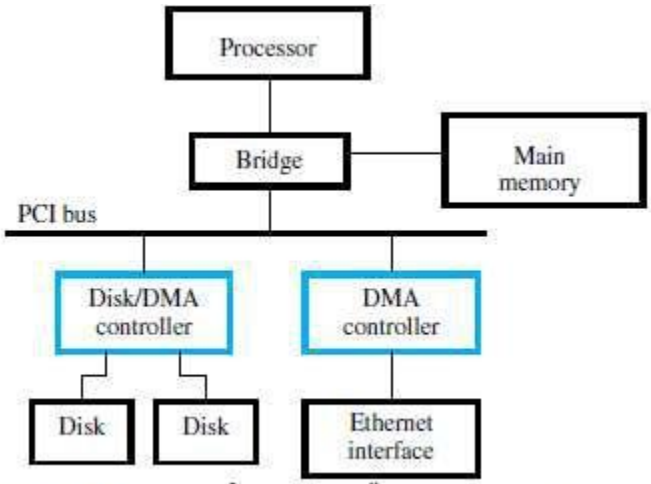

Figure 8.13 Use of DMA controllers in a computer system.

- The disk-address issued by the OS is a logical address.
- The corresponding physical-address on the disk may be different.
- The controller's major functions are:

1) Seek - Causes disk-drive to move the R/W head from its current position to desired track.

2) Read - Initiates a Read-operation, starting at address specified in the disk-address register. Data read serially from the disk are assembled into words and placed into the data buffer for transfer to the main-memory.

3) Write - transfers data to the disk.

4) Error Checking - Computes the error correcting code (ECC) value for the data read from a given sector and compares it with the corresponding ECC value read from the disk.

In case of a mismatch, it corrects the error if possible;

Otherwise, it raises an interrupt to inform the OS that an error has occurred.