UNIT 4

MEMORY SYSTEM

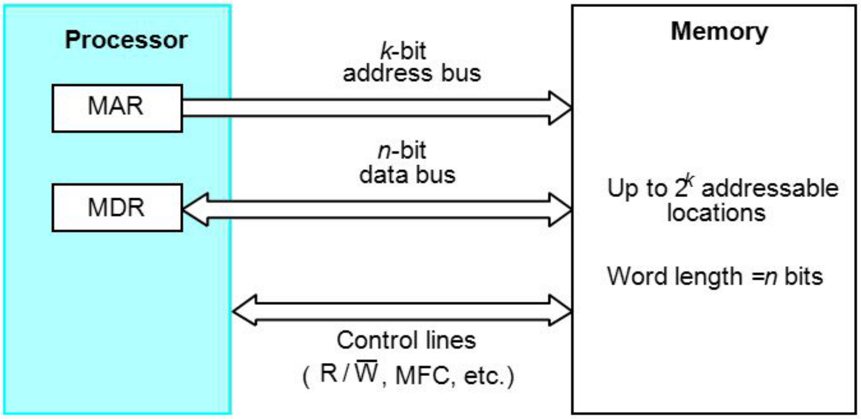
Basic Concepts

The maximum size of the memory that can be used in any computer is determined by the addressing scheme.

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

- Most modern computers are byte addressable
- Form the system standpoint, we can view the memory unit as a block box
- Data transfer between the memory and processor takes place through the use of two processor registers, MAR and MDR

Fig: Connection of Memory to Processor:



- If MAR is k bits long and MDR is n bits long, then the memory may contain up to 2^K addressable locations and the n-bits of data are transferred between the memory and processor. This transfer takes place over the processor bus.
- The processor bus has,
  - AddressLine
  - Data Line
  - Control Line (R/W, MFC – Memory FunctionCompleted)

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 MFCsignal.
- Upon receipt of MFC signal, the processor loads the data onto the data lines into MDRregister.

- The processor writes the data into the memory location by loading the address of this location into MAR and loading the data into MDR sets the R/W line to0.

Memory Access Time→ It is the time that elapses between the intiation of an Operation and the completion of that operation.

Memory Cycle Time→ It is the minimum time delay that required betweenthe initiation of the two successive memory operations.

RAM (Random Access Memory):

In RAM, if any location that can be accessed for a Read/Write operation in fixed amount of time, it is independent of the location's address.

Techniques to increase the effective size and speed of the memory

Cache Memory: Used to increase effective speed.

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

Virtual memory: Used to increase effective size.

- The address generated by the processor does not directly specify the physical locations in the memory. 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 a special memory control circuit is often called the memory managementunit.
- Only the active portion of the address space is mapped into locations in the physicalmemory.
- The remaining virtual addresses are mapped onto the bulk storage devices used, which are usually magnetic disk.
- As the active portion of the virtual address space changes during program execution, the memory management unit changes the mapping function and transfers the data between disk andmemory.
- Thus, during every memory cycle, an address processing mechanism determines whether the addressed in function is in the physical memoryunit.
  If it is, then the proper word is accessed and execution proceeds.
- If it is not, a page of words containing the desired word is transferred from diskto memory.
- This page displaces some page in the memory that is currentlyinactive.

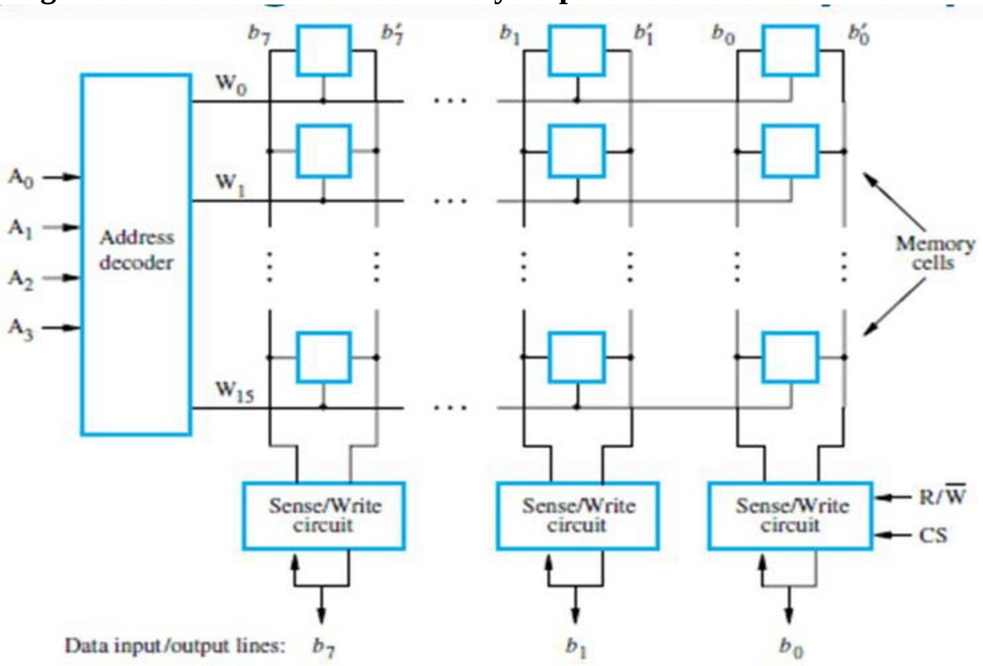
Semi Conductor Ram Memories:

- Semi-Conductor memories are available is a wide range of speeds.
- Their cycle time ranges from 100ns to 10ns

INTERNAL ORGANIZATION OF MEMORY CHIPS:

- Memory cells are usually organized in the form of array, in which each cell is capable of storing one bit of information.
- Each row of cells constitutes a memory word and all cells of a row are connected to a common line called as **wordline**.
- The cells in each column are connected to Sense / Write circuit by two 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 and store it in the cells of the selected word.

Fig: Organization of bit cells in a memory chip



Organization of bit cells in a memory chip.

- The data input and data output of each senses / write circuit are connected to a single bidirectional data line that can be connected to a data bus of the computer

R / W →Specifies the required operation.

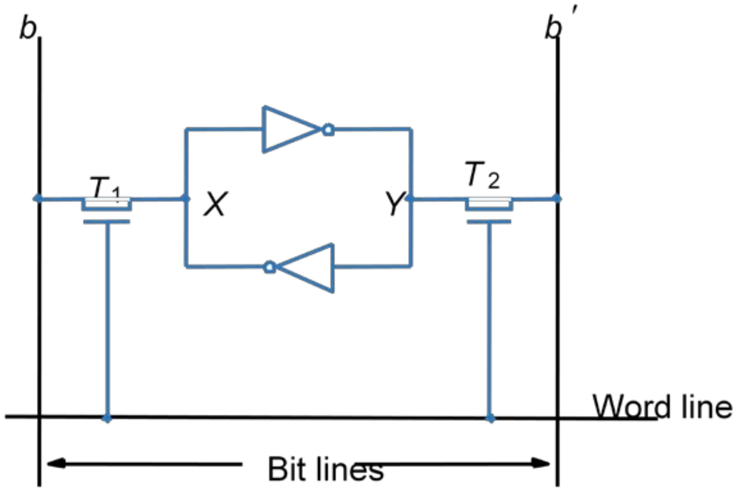
CS →**Chip Select** input selects a given chip in the multi-chip memorysystem

Static Memories:

Memories that consists of circuits capable of retaining their state as long as power is applied are known as **static memory.**

Static RAM

Fig: Static RAM cell



- Two inverters are cross connected to form a latch
- The latch is connected to two bit lines by transistors T_1 and T_2.
- These 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 itsstate.

Read Operation:

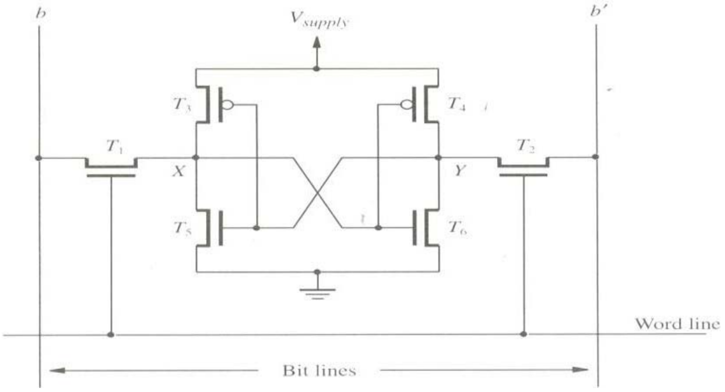
- In order to read the state of the SRAM cell, the word line is activated to close switches T_1 and T_2.
- If the cell is in state 1, the signal on bit line b is high and the signal on the bit line b' is low. Thus b and b' are complement of each other.
- Sense/write circuit at the end of the bit line monitors the state of b and b" and set the output accordingly.

Write Operation:

- 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 are generated by Sense / Write circuit.

CMOS:

Fig: CMOS cell (Complementary Metal oxide Semi Conductor)



- Transistor pairs (T_3, T_5) and (T_4, T_6) form the inverters in the latch.
- 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_2returned ON(Closed), bit line b and b' will have high and low signals respectively.
- The CMOS requires 5V (in older version) or 3.3.V (in new version) of power supply voltage.
- The continuous power is needed for the cell to retain itsstate

Merit :

- It has low power consumption because the current flows in the cell only when the cell is being activated accessed.
- Static RAM's can be accessed quickly. It access time is few nano seconds.

Demerit:

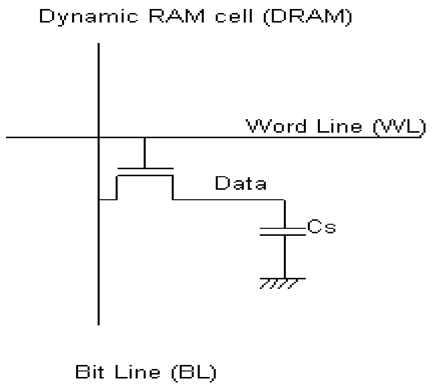
- SRAM‟s are said to be volatile memories because their contents are lost when the power is interrupted.

Asynchronous DRAMS:-

- Less expensive RAM‟s can be implemented if simplex calls are used such cells cannot retain their state indefinitely. Hence they are called **Dynamic RAM's(DRAM).**
- The information stored in a dynamic memory cell in the form of a charge ona capacitor and this charge can be maintained only for tens ofMilliseconds.
- The contents must be periodically refreshed by restoring by restoring this capacitor charge to its fullvalue.

DRAM:

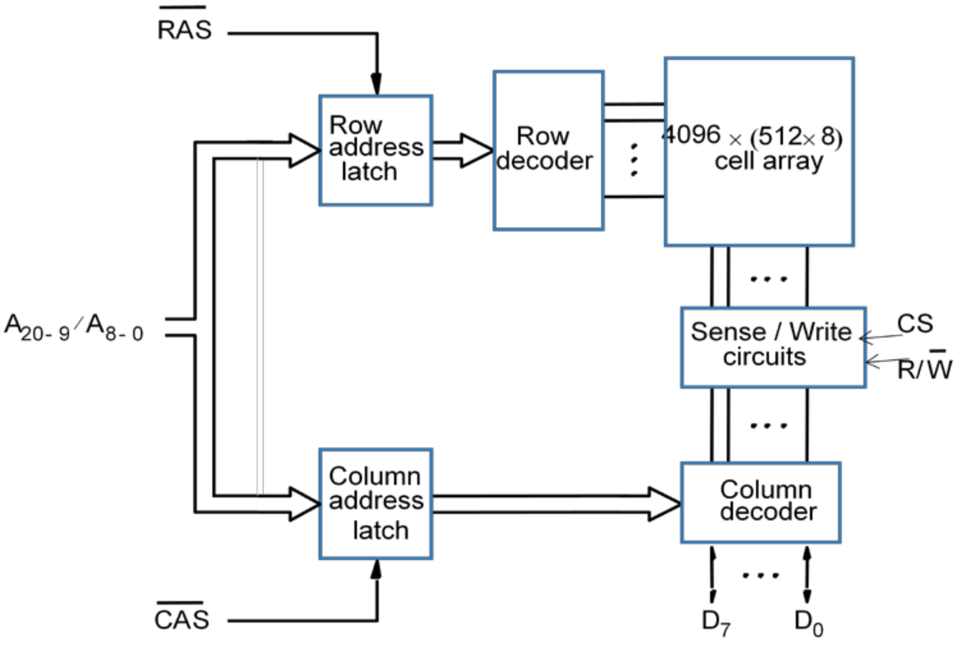
Fig: A single transistor dynamic Memory cell



- In order to store information in the cell, the transistor T is turned „on‟&the appropriate voltage is applied to the bit line, which charges thecapacitor.
- After the transistor is turned off, the capacitor begins to discharge which is caused by the capacitor’s own leakage resistance.
- Hence the information stored in the cell can be retrieved correctly beforethe threshold value of the capacitor dropsdown.
- During a read operation, the transistor is turned “on‟ & a sense amplifier connected to the bit line detects whether the charge on the capacitor is abovethe thresholdvalue.

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‟.

Internal organization of a 2M X 8 dynamic Memory chip.



DESCRIPTION:

- The 4 in each row bit cells are divided into 512 groups of 8.
- 21 bit address is needed to access a byte in the memory(12 bit→ To select a row,

  9bits→Specify the group of 8 bits in the selected row).

  A_{8-0}→ Column address of a byte.

  A_{20-9}→ Row address of a byte.
- During Read/ Write operation, the row address is applied first. It is loaded intothe row address latch in response to a signal pulse on **Row Address Strobe (RAS)** input of the chip.
- 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 **Column AddressStrobe(CAS).**
- The information in this latch is decoded and the appropriate group of 8 Sense/Write circuits are selected.
- R/W =1(read operation)→The output values of the selected circuits are transferred to the data lines D0 -D7.
- R/W =0(write operation)→The information on D0 - D7 are transferred to the selectedcircuits.
- RAS and CAS are active low so that they cause the latching of address when they change from high to low. This is because they are indicated by RAS &CAS.
- To ensure that the contents of a DRAM „s are maintained, each row of cells must be accessed periodically. Refresh operation usually perform this function automatically.
- A specialized memory controller circuit provides the necessary controlsignals RAS &CAS, that govern the timing.
- The processor must take into account the delay in the response of the memory. Such memories are referred to as **Asynchronous DRAM's.**

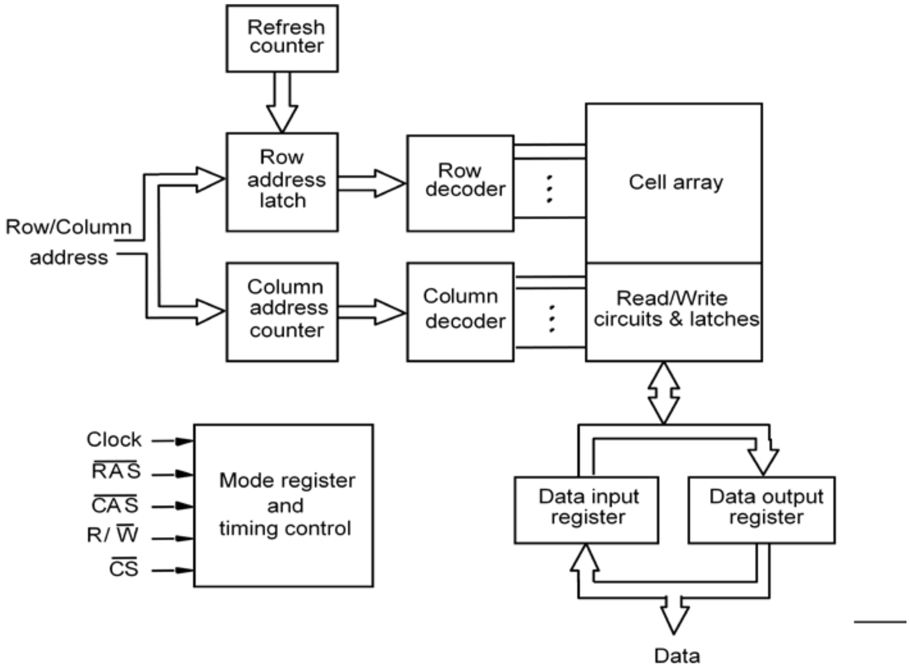
Fast Page Mode:

- 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 PageMode.**

Synchronous DRAM:

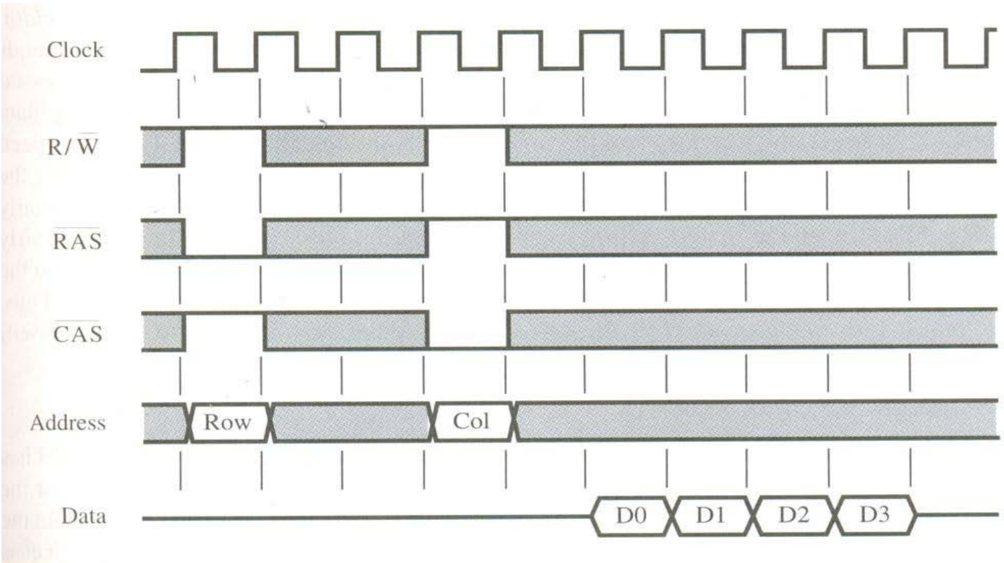
- Here the operations e directly synchronized with clock signal.
- The address and data connections are buffered by means of registers.
- The output of each sense amplifier is connected to a latch.
- A Read operation causes the contents of all cells in the selected row to be loaded in these latches.

Fig: Synchronous DRAM



- Data held in the latches that correspond to the selected columns are transferred into the data output register, thus becoming available on the data output pins.

Fig: Timing Diagram →Burst Read of Length 4 in an SDRAM



- First ,the row address is latched under control of RAS signal.
- The memory typically takes 2 or 3 clock cycles to activate the selected row.
- Then the column address is latched under the control of CAS signal.
- After a delay of one clock cycle ,the first set of data bits is placed on the data lines. The SDRAM automatically increments the column address to access the next 3 sets of bits in the selected row, which are placed on the data lines in the next 3 clock cycles.

Latency & Bandwidth:

- A good indication of performance is given by two parameters .They are,
  - Latency
  - Bandwidth

Latency:

- It refers to the amount of time it takes to transfer a word of data to or from the memory.
- For a transfer of single word, the latency provides the complete indicationof memoryperformance.
- For a block transfer, the latency denote the time it takes to transfer the firstword ofdata.

Bandwidth:

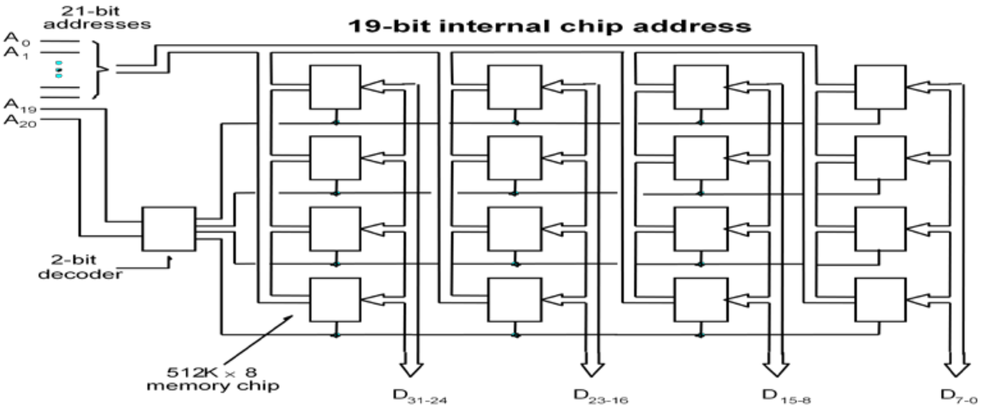
- It is defined as the number of bits or bytes that can be transferred in one second.
- Bandwidth mainly depends upon the speed of access to the stored data & onthe number of bits that can be accessed inparallel.

Double Data Rate SDRAM (DDR-SDRAM)

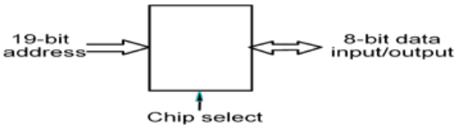
- The standard SDRAM performs all actions on the rising edge of the clock signal.
- The double data rate SDRAM transfer data on both the edges(loadingedge, Trailing edge).The Bandwidth of DDR-SDRAM is doubled for long burst transfer.
- To make it possible to access the data at high rate , the cell array is organized into two banks.
- Each bank can be accessed separately.
- Consecutive words of a given block are stored in different banks.
- Such interleaving of words allows simultaneous access to two words that are transferred on successive edge of the clock.

Larger Memories:

Static Memories:



512K × 8 memory chip



- *Implement a memory unit of 2M words of 32 bits each.*
- *Use 512x8 static memory chips.*
- *Each column consists of 4 chips.*
- *Each chip implements one byte position.*
- *A chip is selected by setting its chip select control line to 1.*
- *Selected chip places its data on the data output line, outputs of other chips are in high impedance state.*
- *21 bits to address a 32-bit word.*
- *High order 2 bits are needed to select the row, by activating the*
- *four Chip Select signals.*
- *19- bits are used to access specific byte locations inside the selected chip.*

Dynamic Memory System:

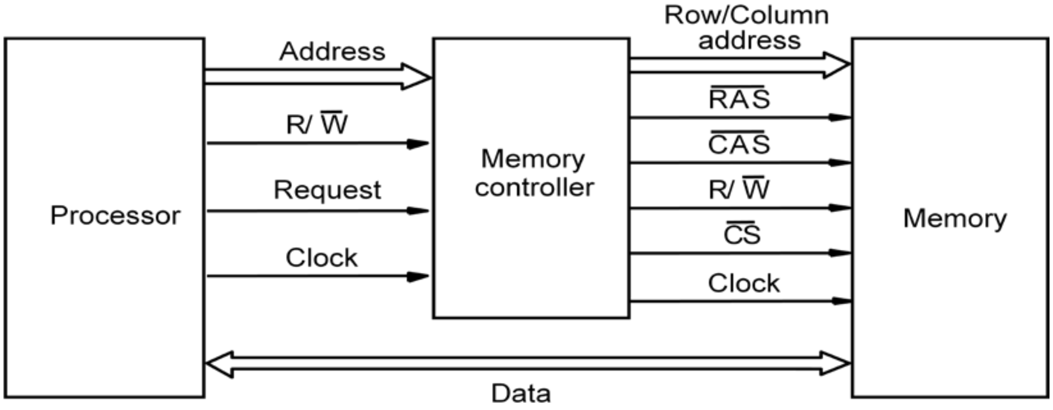
- The physical implementation is done in the form of Memory Modules.
- If a large memory is built by placing DRAM chips directly on the main system Printed circuit board that contains the processor, often referred to as Mother board ;it will occupy large amount of space on the board.
- These packaging considerations have led to the development of larger memory units known as SIMM‟s & DIMM‟s.

  SIMM-Single Inline memory Module
  DIMM-Dual Inline memory Module
- SIMM & DIMM consists of several memory chips on a separate small boardthat plugs vertically into single socket on themotherboard.

MEMORY SYSTEM CONSIDERATION:

- To reduce the number of pins, the dynamic memory chips use multiplexed addressinputs.
- The address is divided into two parts. They are,
  - **High Order Address Bit** (Select a row in cell array & it is providedfirst and latched into memory chips under the control of RASsignal).
  - **Low Order Address Bit** (Selects a column and they are provided onsame address pins and latched using CASsignals).
- The Multiplexing of address bit is usually done by **Memory Controller Circuit.**

Fig: Use of Memory Controller

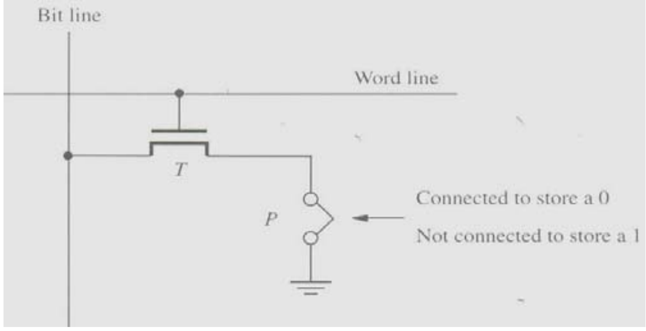


- The Controller accepts a complete address & R/W signal from the processor, under the control of a Request signal which indicates that a memory access operation isneeded.
- The Controller then forwards the row & column portions of the address tothe memory and generates RAS &CAS signals. It also sends R/W &CS signals to the memory.
- The CS signal is usually active low, hence it is shown as CS.

Read Only Memory:

- Both SRAM and DRAM chips are volatile, which means that they lose the stored information if power is turnedoff.
- Many applications require Non-volatile memory (which retains the stored information if power is turnedoff).
- Eg: Operating System software has to be loaded from disk to memory which requires the program that boots the Operating System ie. 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.

Fig: ROM cell



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.

At Logic value ‘1’ → Transistor switch is open.

The bit line remains at high voltage.

- 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 outputvalue.

Types of ROM:

- Different types of non-volatile memoryare,
  - PROM
  - EPROM
  - EEPROM
  - Flash Memory

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 it is programmed, the memory contains all 0’s
- The user can insert 1’s at the required location by burning out the fuse at these locations using high-current pulse.
- This process is irreversible.

Merit:

- It provides flexibility.
- It is faster.
- It is less expensive because they can be programmed directly by the user.

EPROM:-Erasable reprogrammable ROM:

- EPROM allows the stored data to be erased and new data to be loaded.
- In an EPROM cell, a connection to ground is always made at “P‟ and a special transistor is used, which has the ability to function either as a normal transistor or as a disabled transistor that is always turned ”off‟.
- This transistor can be programmed to behave as a permanently open switch,by injecting charge into it that becomes trappedinside.
- Erasure requires dissipating the charges trapped in the transistor of memory cells. This can be done by exposing the chip to ultra-violet light.

Merits:

It provides flexibility during the development phase of digital system.

- It is capable of retaining the stored information for a longtime.

Demerits:

- The chip must be physically removed from the circuit for reprogramming and its entire contents are erased by UVlight.

EEPROM:-Electrically Erasable ROM:

Merits:

1) It can be both programmed and erased electrically.
2) It allows the erasing of all cell contents selectively.

Demerits:

- 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 the contents of a single cell but it is only possible to write the entire contents of a block.
- Prior to writing, the previous contents of the block are erased.

  Eg. In MP3 player, the flash memory stores the data that represents sound.
- Single flash chips cannot provide sufficient storage capacity for embedded system application.
- There are 2 methods for implementing larger memory modules consisting of number of chips. They are,
  - ➢ FlashCards
  - ➢ FlashDrives.

Merits:

- Flash drives have greater density which leads to higher capacity & low costper bit.
- It requires single power supply voltage & consumes less power in theiroperation.

Flash Cards:

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

Flash Drives:

- Larger flash memory module 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.

Merits:

- They have shorter seek and access time which results in faster response.
- They have low power consumption which makes them attractive for battery driven application.
- They are insensitive to vibration.

Demerit:

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

SPEED,SIZE COST:

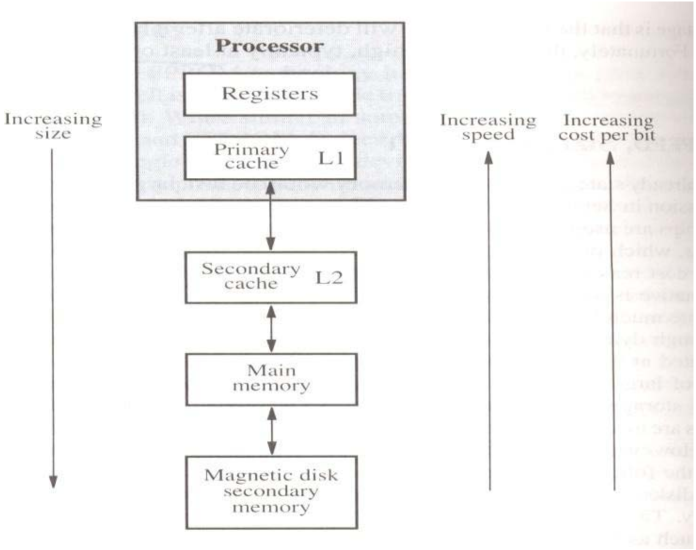
Characteristics	**SRAM**	**DRAM**	**Magnetis Disk**
Speed	Very Fast	Slower	Much slower than DRAM
Size	Large	Small	Small
Cost	Expensive	Less Expensive	Low price

Magnetic Disk:

- A huge amount of cost effective storage can be provided by magneticdisk;The main memory can be built with DRAM which leaves SRAM‟s to be used in smaller units where speed is ofessence.

Memory	**Speed**	**Size**	**Cost**
Registers	Very high	Lower	Very Lower
Primary cache	High	Lower	Low
Secondary cache	Low	Low	Low
Main memory	Lower than Seconadry cache	High	High
Secondary Memory	Very low	Very High	Very High

Fig: Memory Hierarchy



- Fastest access is to the data held in processor registers. Registers are at the top of the memory hierarchy.
- Relatively small amount of memory that can be implemented on the processor chip. This is processor cache.
- Two levels of cache. Level 1 (L1) cache is on the processor chip. Level 2 (L2) cache is in between main memory and processor.
- Next level is main memory, implemented as SIMMs. Much larger, but much slower than cache memory.
- Next level is magnetic disks. Huge amount of inexpensive storage.
- Speed of memory access is critical, the idea is to bring instructions and data that will be used in the near future as close to the processor as possible

.CACHE MEMORIES

- The effectiveness of cache mechanism is based on the property of "**Locality of reference".**

Locality of Reference:

- Analysis of large number of program shows that reference to memory at any given interval of time tend to be confined to few localized area in memory. This is known as locality of reference.
- **Cache Memory definition:** If the active portion of program and data are placed in fast memory, then average execution time of the program can be reduced. Such fast memory is called cache memory.
- It is placed in between the main memory and the CPU.

.

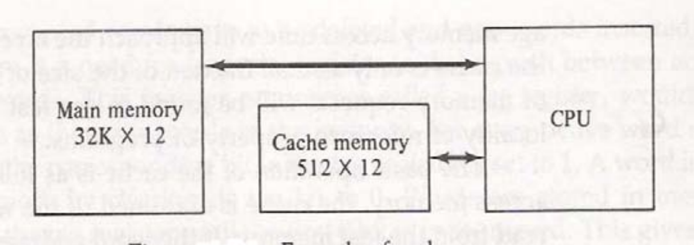


Figure Example of cache memory.

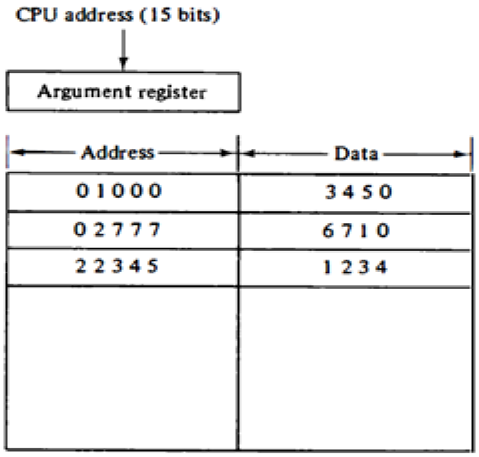
- When the CPU need to access the memory it first search in cache. If word is Found , it is read.
- If the word is not found, it is read from main memory and a block of data is transferred from main memory to cache which contain the current word.
- If the word is found in cache, it is said hit. If the word is not found, it is called miss.
- Performance of cache is measured in terms of hit ratio which ratio of total hit to total memory access by CPU.
- The Cache memory stores a reasonable number of blocks at a given time but this number is small compared to the total number of blocks available in Main Memory.
- The transformation of data from main memory to cache is known as mapping process. Three types of mapping procedures are:
  - Associative Mapping
  - Direct Mapping
  - Set-Associative Mapping

.

- The Cache control hardware decide that which block should be removed tocreate space for the new block that contains the referenced word during miss. 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.

Associative mapping:

- Fastest and most flexible cache organization uses associative memory.
- Here our cache memory is constructed by associative memories.
- Let us assume that cache memory is able to store data in the octal format and able to store three words at a time
- Our cache is divided into two parts address part (15 bits),data part(12 bits)
- CPU referred(15 bit) Address is placed in argument register and memory is searched for matching address.
- If address is found corresponding data is read.
- If address is not found, it is read from main memory and transferred to cache.
- If the cache is full, an address- word pair must be displaced.
- Various algorithms are used to determine which pair to displace. Some of them are FIFO (First In First Out), LRU (Least Recently Used) etc.



Address	Data
01000	3450
02777	6710
22345	1234

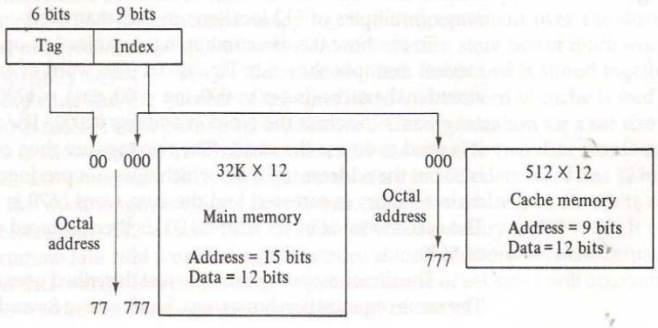
Disadvantage:

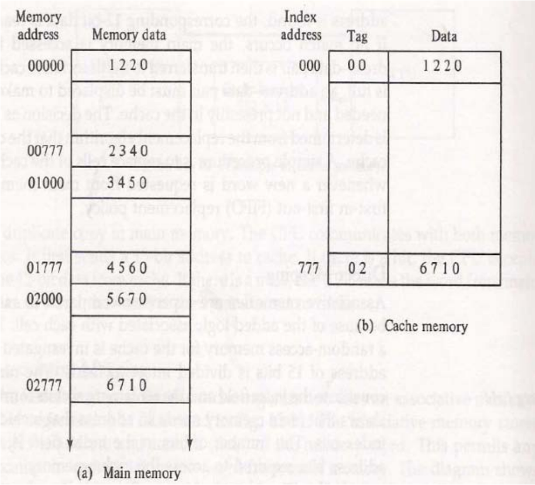
- Here our cache is constructed using associative memories which are expensive**.**

Direct Mapping:

- In direct Mapping Instead of using associative memories to construct cache SRAM devices are used.
- CPU address is divided into two fields tag(6 bits) and index(9 bits).
- Index field is required to access cache memory and total address is used to access main memory.
- If there are 2^k words in cache and 2^n words in main memory, then n bit memory address is divided into two parts. k bits for index field and (n-k) bits for tag field.

Figure Addressing relationships between main and cache memories.





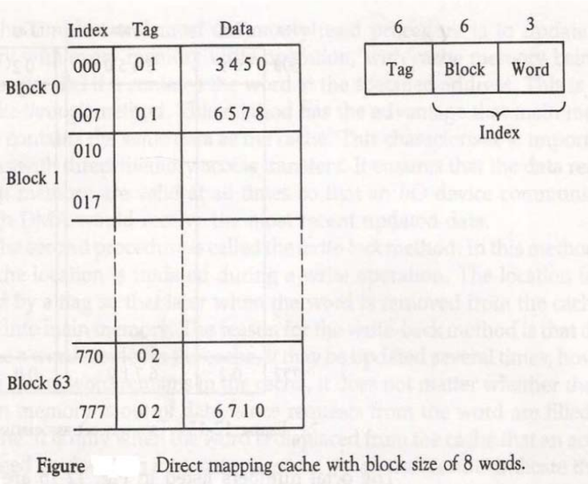


Figure Direct mapping cache with block size of 8 words.

Fig1(a) fig1(b)

- When CPU generates memory request, index field is used to access the cache.
- Tag field of the CPU address is compared with the tag in the word read. If the tag match, there is hit.
- If the tag does not match, word is read from main memory and updated in cache.
- This examplefig1(a) use the block size of 1.
- The same organization can be implemented for block size 8fig 1(b)
- The index field is divided into two parts: block field and word field.
- In 512 word cache there are 64 blocks of 8 words each(64*8=512).
- Block is specified with 6 bit field and word within block with 3 bit field.
- Every time miss occur, entire block of 8 word is transferred from main memory to cache.

Disadvantage

- In direct mapping two words with same index in their address but different tag values can't reside simultaneously in memory.

Set Associative mapping:

- In direct mapping two words with same index in their address but different tag values can't reside simultaneously in memory.
- In this mapping, each data word is stored together with its tag and number of tag-data items in one word of the cache is said to form set. Here in our example we have 2 sets so it is called 2-way set associative mapping.
- In general, a set associative cache of set size k will accommodate k words of main memory in each word of cache.

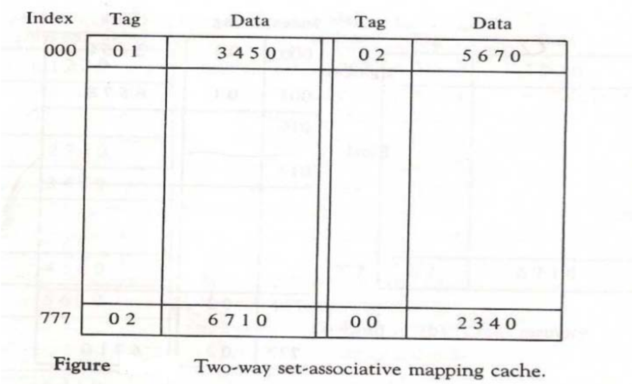


Figure Two-way set-associative mapping cache.

When a miss occur and the set is full, one of the tag data item is replaced with new value using various algorithm

Writing into cache :

Writing in to cache can be done in two ways:

- ○ Write through
- ○ Write Back

- In **write through,** whenever write operation is performed in cache memory, main memory is also updated in parallel with the cache.
- In **write back**, only cache is updated and marked by the flag. When the word is removed from cache, flag is checked if it is set the corresponding address in main memory is updated

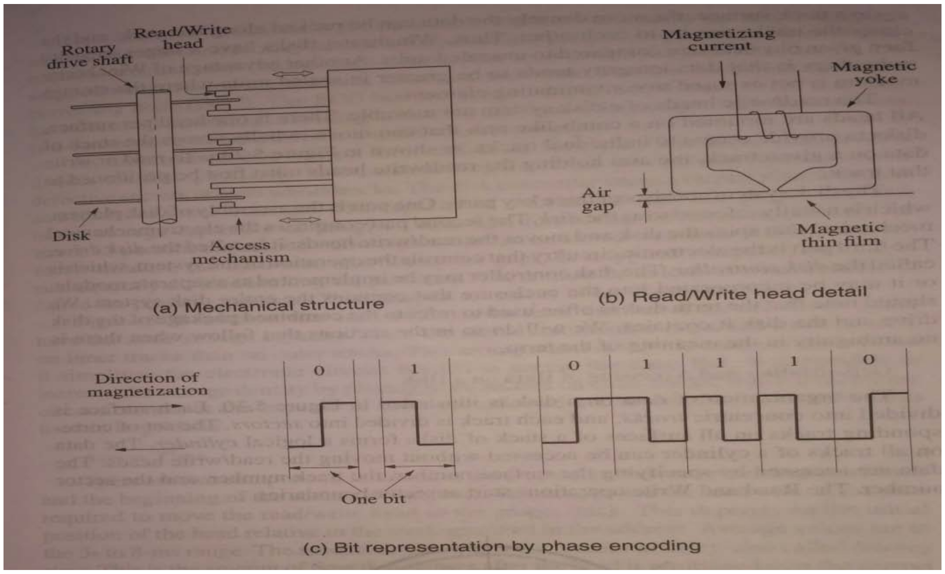
Secondary Storage Devices:

- 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,
  - ➢ Magnetic Disk
  - ➢ Optical Disk
  - ➢ 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, usually on bothsides.
- The disks are placed in a rotary drive so that the magnetized surfaces move in close proximity to read /write heads. Each head consists of **magnetic yoke & magnetizing coil**.
- Digital information can be stored on the magnetic film by applying the current Pulse of suitable polarity 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 of magnetization, a voltage is induced in the head only at 0-1 and at 1-0 transition in the bit stream.
- A consecutive “0‟s&”1‟s are determined by using the clock which is mainly used for synchronization.
- Manchester Encoding (it is a synchronous clock **encoding** technique used to **encode** the clock and data of a synchronous bit stream.) is the technique to combine the clocking information with data.
- The Manchester Encoding describes that how the self-clocking schemeis implemented.
- The Read/Write heads must be maintained at a very small distance from the moving disk surfaces in order to achieve high bit densities.
- When the disk are moving at their steady state, the air pressure develops between the disk surfaces & the head forces the head away from the surface.
- The flexible spring connection between head and its arm mounting permits the head to fly at the desired distance away from the surface.

Mechanical Structure



(a) Mechanical structure

(b) Read/Write head detail

(c) Bit representation by phase encoding

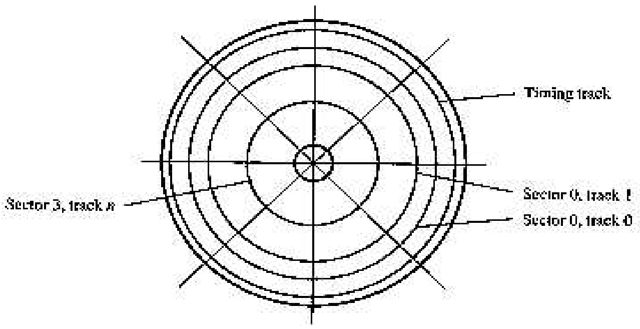
Wanchester Technology:

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

Merits:

- It have 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. They are,
  - **Disk Platter**(Usually called Disk)
  - **Disk Drive(**spins the disk & moves Read/write heads)
  - **Disk Controller**(controls the operation of the system.)

Fig: Organizing& accessing the data on disk



(a) Organization of one surface of a disk

- Each surface is divided into concentric **tracks**.
- 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.
- Each sector usually contains 512 bytes.

Sector header -> contains identification information.
It helps to find the desired sector on the selected track.

ECC (Error checking code)- 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 and this process may discover some defective sectors on all tracks.
- The disk controller keeps a record of such defects.
- The disk is divided into logical partitions. They are,
  - Primary partition
  - Secondary partition
- In the diag, 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 between **receiving an address** and the **beginning of the actual data transfer**. They are,
  - Seek time
  - Rotational delay /Latency

Seek time – Time required to move the read/write head to the proper track.
Latency – The amount of time that elapses after the head is positioned over the correct track until the starting position of the addressed sector passes under the read/write head.

Seek time + Latency = Disk access time

Typical disk

One inch disk- weight=1 ounce,
size -> comparable to match book Capacity -> 1GB
Recording surface=20
Tracks=15000 tracks/surface
Sectors=400.
Each sector stores 512 bytes of data
Capacity of formatted disk=20x15000x400x512=60x10^9 =60GB
Seek time=3ms
Platter rotation=10000 rev/min
Latency=3ms
Internet transfer rate=34MB/s

Data Buffer / cache

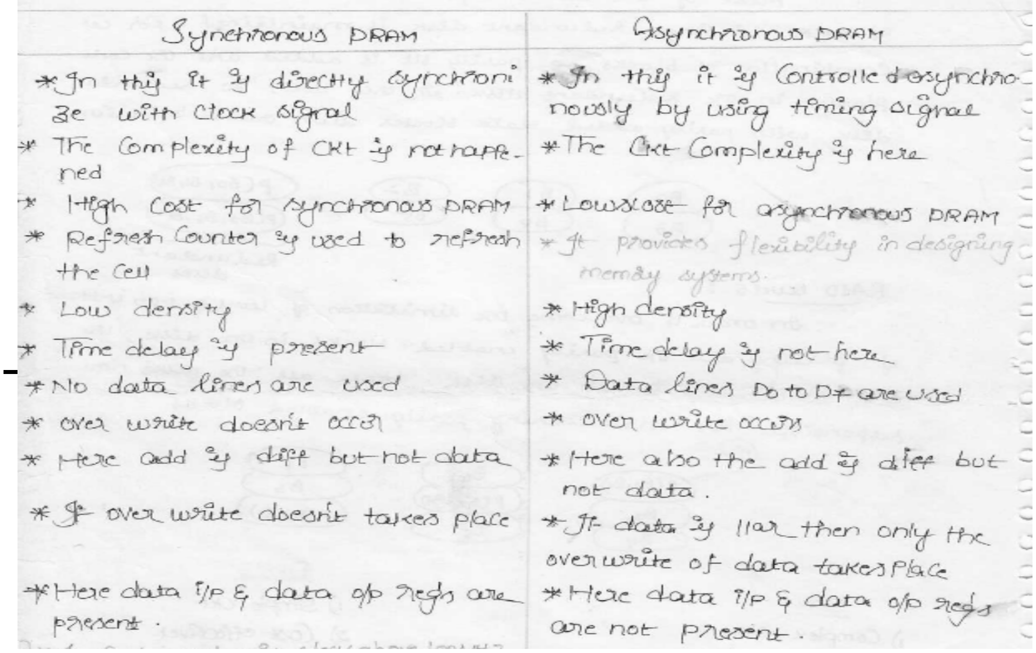
- A disk drive that incorporates the required SCSI circuit is referred as SCSI drive.
- The SCSI can transfer data at higher rate than the disk tracks.
- An efficient method to deal with the possible difference in transfer rate between disk and SCSI bus is accomplished by including a data buffer.
  This buffer is a semiconductor memory.
- The data buffer can also provide cache mechanism for the disk (ie) when are ad request arrives at the disk, then controller first check if the data is available in the cache(buffer).
- If the data is available in the cache, it can be accessed and placed on SCSI bus .If it is not available then the data will be retrieved from the disk.

Disk Controller

The disk controller acts as interface between disk drive and system bus.

- The disk controller uses DMA scheme to transfer data between disk and main memory.
- When the OS initiates the transfer by issuing Read/Write request, the controllers register will load the following information. They are,
- Main memory address(address of first main memory location of the block of words involved in the transfer)
- Disk address(The location of the sector containing the beginning of the desired block of words)

Difference between Synchronous DRAM &Asynchronous DRAM



Synchronous DRAM	Asynchronous DRAM
* In this it is directly synchronize with clock signal	* In this it is Controlled asynchronously by using timing signal
* The Complexity of Ckt is not happened	* The Ckt Complexity is here
* High Cost for synchronous DRAM	* Low cost for asynchronous DRAM
* Refresh Counter is used to refresh the Cell	* It provides flexibility in designing memory systems.
* Low density	* High density
* Time delay is present	* Time delay is not here
* No data lines are used	* Data lines D0 to D7 are used
* Over write doesn't occur	* Over write occurs
* Here add is diff but not data	* Here also the add is diff but not data.
* If over write doesn't takes place	* If data is llar then only the overwrite of data takes place
* Here data i/p & data o/p regs are present.	* Here data i/p & data o/p regs are not present.

Difference between SRAM & DRAM

- **Static RAMs (SRAMs):**
  - SRAM constructed using transistors and latches. It is faster than DRAM. Hence it is used for construction of cache memory.
  - Consist of circuits that are capable of retaining their state as long as the power is applied.
  - Volatile memories, because their contents are lost when power is interrupted.
  - Access times of static RAMs are in the range of few nanoseconds.
  - However, the cost is usually high.

- **Dynamic RAMs (DRAMs):**
  - DRAM is constructed using capacitors and Transistors. It is used to construct main memory, since it is little bit slow in operation than SRAM.
  - Do not retain their state indefinitely.
  - Contents must be periodically refreshed.
  - Contents may be refreshed while accessing them for reading.

Cache Coherence problem:

- A bit called as “valid bit” is provided for each block
- If the block contains valid data, then the bit is set to 1, else it is 0.
- Valid bits are set to 0, when the power is just turned on.
- When a block is loaded into the cache for the first time, the valid bit is set to 1.
- Data transfers between main memory and disk occur directly bypassing the cache.
- When the data on a disk changes, the main memory block is also updated.
- However, if the data is also resident in the cache, then the valid bit is set to 0.
- What happens if the data in the disk and main memory changes and the write-back protocol is being used?
- In this case, the data in the cache may also have changed and is indicated by the dirty bit.
- The copies of the data in the cache, and the main memory are different. This is called the cache coherence problem.
- One option is to force a write-back before the main memory is updated from the disk.