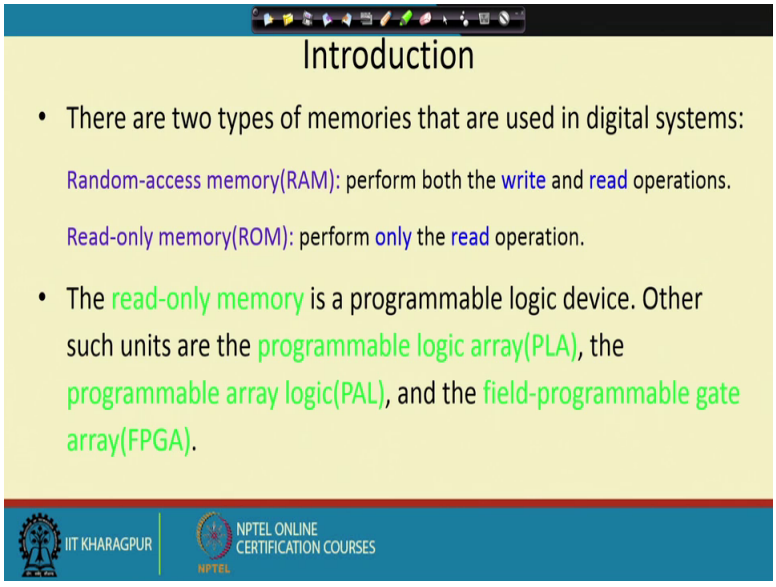
Digital Circuits
Prof. Santanu Chattopadhyay
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture – 41
Memory

A semiconductor memories: so, these are one of the a very important component like if you look into any digital system, because we have seen that if you want store some information in a digital system, then one possible storage that we have seen at the registers. But you see registers are quiet costly because for making one register one register so, you will need a number of flip lops and some control logic like that.

So, if you are willing to store large amount of information, and then you want to access those contents, then we need some mechanism by which it can be done in a much better fashion, not via some circuitry to axis individual flip lops. So, the semiconductor memory design so, this will tell us like how this these things are done and where can what are the functionalities of this memory and there are many memory chips available so different types. So, what are the general control lines that you should expect in a memory chip and how is the memory organized in a system, so we will be looking into these aspects.

(Refer Slide Time: 01:24)



So, there if you look into this memory there are 2 broad types of memories if that are used in digital system, one is known as random access memory or RAM and another is read only memory or ROM so, in random access memory so, you can perform both write and read operation, and in read only memory so, you can perform only the read operation.

So, though this name random access memory is somewhat miss leading so, it means that there is some as if there is some other memory which is not random access which is sequential axis. This is true, but sequential axis memories are there, but they are no more used in the computer systems now, but the both random access memory or RAM and read only memory ROM so, they are they act they are actually random access so, if there is no order in which you have to axis the locations in this chips.

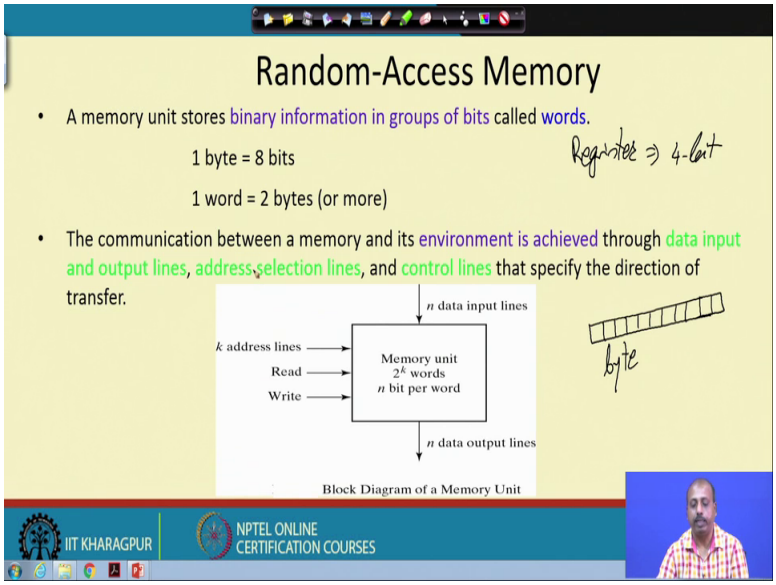
So, you can tell the location address and the memory will be able to access that particular location. So, the basic difference between a RAM and a ROM is that the in case of RAM so, you can do both write and read operation so, either that is you can modify the content of a location in the RAM or you can read the content of a location in the RAM. Whereas, for read only memory so, it is you can only to read operations so, normally if the information is not going to change. For example, say the some part of the operating system which is the memory resident part of the operating system, which is not going to change which is commonly known as the basic input output sub system ok.

So, those or the monitor program so, they are actually kept in the ROM because they are not going to change. Similarly if you are thinking about some other application where say some filtering up application where you have got a set of filter coefficient. So, this filter coefficients may be kept in a ROM and then we can do the operation so, the filter coefficients are not going to change, so, they are they can be kept in a ROM.

So, this read only memory is a programmable logic device so, this is programmable logic device means you can program it, and you can put some content there. So, random access memory so, you can the user can change the content at any point of time, whereas, this programmable logic device based read only memory so, you can change the content, but only I have with some restriction so, you cannot change the content while the system is in operation; so, either you have to take out the chip or you have to stop the system functionality, and do some modification to the programmable device.

So, apart from this read only memory which can act as programmable logic device so, there are other programmable logic devices like programmable logic array, we have seen some part in our discussion previously the programmable array logic or pal and filled programmable get array is or FPGAs. So, will be looking into this devices also to some extent.

(Refer Slide Time: 04:21)



So, to start with will be looking into the random access memory, a memory unit it stores the binary information in groups of bits called words. So, as we know that in case of in case of a register so, if I say that it is a 4 bit register, so, 4 bit register then we know that this 4 bits may be contained in four flip flops. So, the same concept we have in this random access memory, so, we have got a few locations of this memory chip which are access together ok.

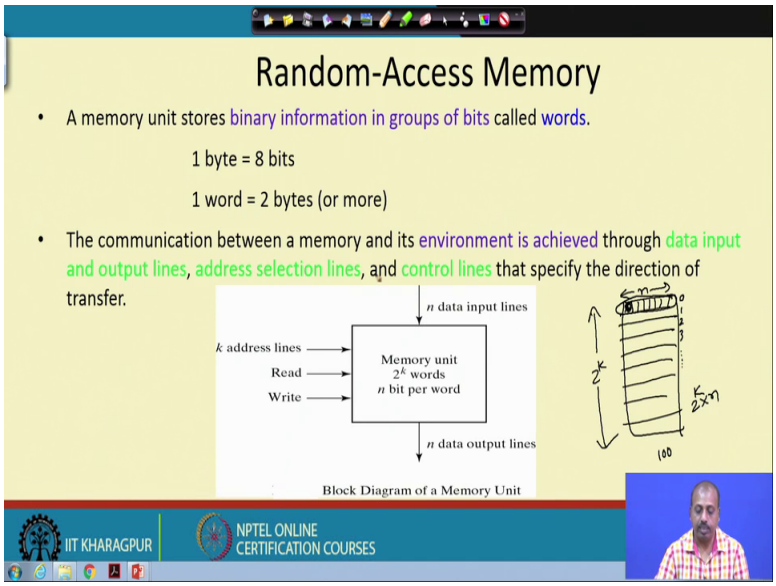
So, these are this may be there may be 8 such bits which are access together so, that constitutes a byte. So, this 8 bit will constitute a byte and depending upon the memory design in one memory access, so, you may get the content of one you may get the content of one byte or may be more than one byte. So, this is determined by this lines like if you say if you see that I have got a memory unit like this, that has got n data input lines and n data output lines; so, if you are accessing a particular location in this memory so, it will be you will be getting data in terms of n bits.

So, if you trying to modify something so, you have to give n bit data here which will be retain at some a set of n bit locations or if you are trying to get the content of this memory for a particular location so, the you will get this n bit data. So, this individual bits of memory so, they are accessed in a group so, that will be call that will call a word.

So, so, that is known as the word size of the memory. So, word size is normally equal 2 bytes or 4bytes like that so, there is no harden first rule like what should be word size, but word size is never less than one byte of course. So, so you can in your computer system so, you may have different word sizes.

So, communication between a memory and its environment is achieved through data input and output lines address selection lines and control lines so, you can think about this memory as if it is having 2 to the power k words and each word is n bit.

(Refer Slide Time: 06:46)



So, you can you can think that as if my memory is like this so, it has got 2 to the power k such words and each location is consisting of n bits so, this side you have got n and this side you have got 2 to the power k.

So, if you can think of this memory as a collection of cells, and number of cells is equal to 2 to the power k by n. So, this is individual cells you have got 2 to the power k by n. But the point is you cannot access a single cell alone so, if you trying to if you are trying to access a single cell so, you have to access this entire location so, the I can so, there are

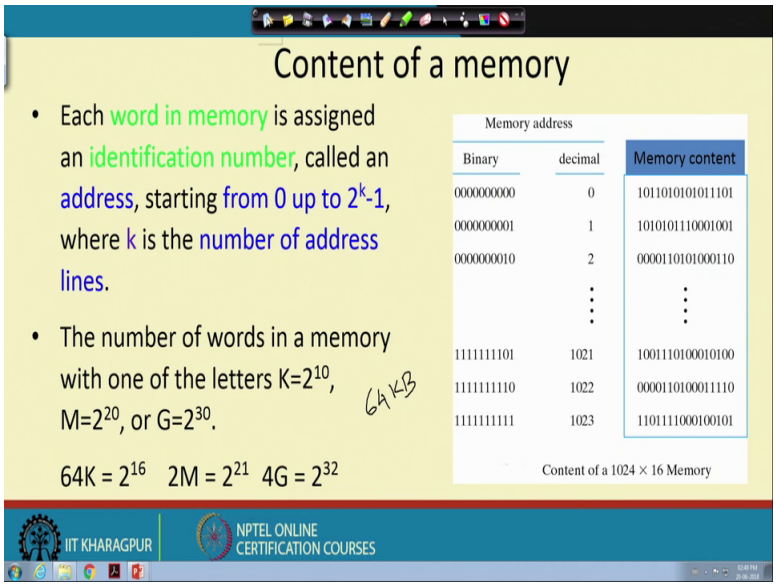
some so, every location has got an address so, if this may be address 0, this may be 1, this may be 2 this may be 3 so, it goes like this.

So, for example, if I ask for the content of location 100 so, it will give me and n bit pattern that that was that is the content of that location. So, we have got these memory unit that consist of 2 to the power k words, and each word is n bit quite.

So, this k address lines had to be given to tell which location we are trying to access so, then this n bit data input and output lines are there, if you are doing a read operation then the content of that address location will be available on this n bit data output line. On the other hand if you are doing a write operation so, whatever value is available on this n data input lines will be stored in the corresponding location.

So, these control lines the read and write so, they specify the direction of the transfer. So, this is the idea of this basic random access memory.

(Refer Slide Time: 08:32)



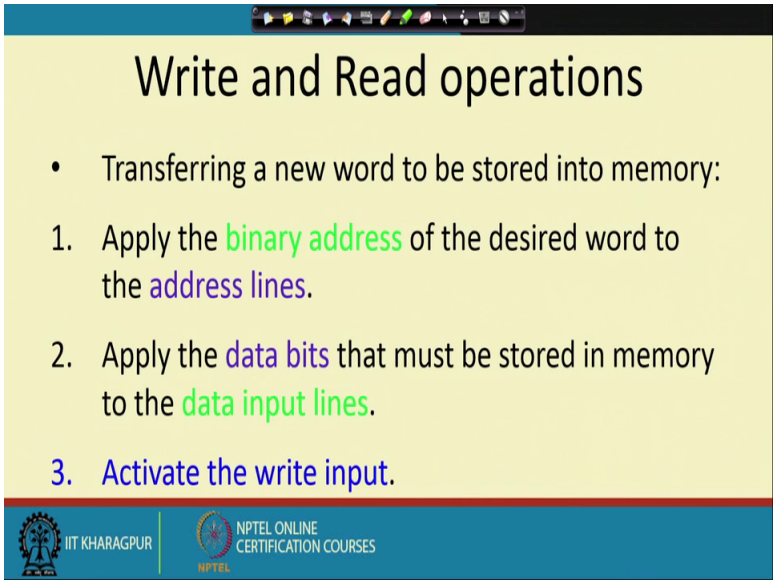
So, each word in memory is assigned an identification number called an address. So, these are the addresses that the each locations so, this is address 0 this is address one this is the address 2 like that so, this is the up to address 1023.So, in this particular case we have got 1024 locations 1 k 1 kilo memory locations one kilo memory address so, if the k equal to 10 in this case so, the address will go from 0 2 2 to the power k minus 1 so, address go from 0 to 1, 0 to 3 k is the number of address lines and number of words in a

memory we with one letter that is K right it in the K which is kilo which is 2 to the power 10 mega m 2 to the power 20 or giga 2 to the power 30 like that.

So, 64 k that is equal to 2 to the power 16 so, k is 2 to the power 10 and this 64 is 2 to the power 6 so, that gives us 2 to the power 16 similarly 2 mega it is 2 to the power 21. Now if you write a whether it is byte or word etcetera so, that will be written here like if I write like 64 KB so, if I write like 64 KB. So that means, individual locations are having 8 bits ok.

So, this way we can tell the number of locations in the memory chip.

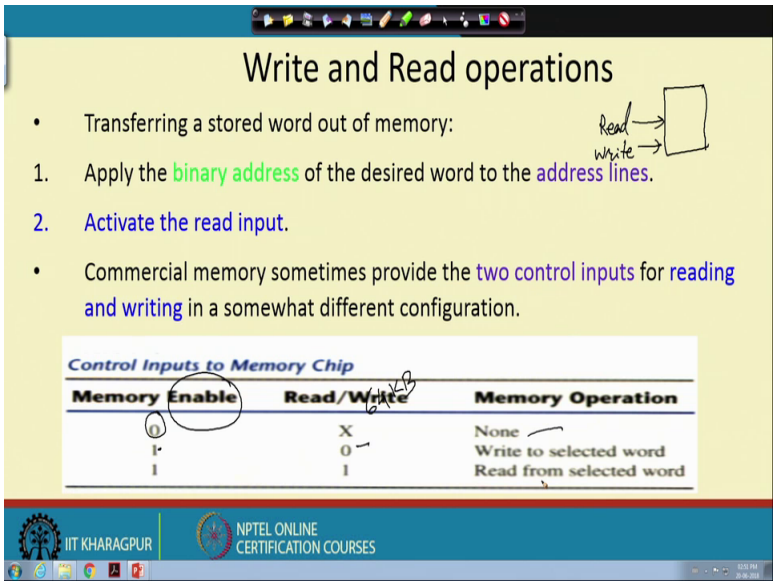
(Refer Slide Time: 09:55)



So, so, if you try to do write and read operations so, first of all will looking into the write operation, which is to transfer in new word to be stored into the memory. So, for that what we have to do is that, we have to apply the binary address of the desired word, to the address lines and apply the data bits that must be stored in the memory to the data input lines and you have to tell the right inputs.

So, we have to so, if you look into this diagram for writing, we have to give the address lines we have to give the data value and we have to keep the rights control signal so, read control signal should not be activated. So, that is how this write operation will take place

(Refer Slide Time: 10:42)



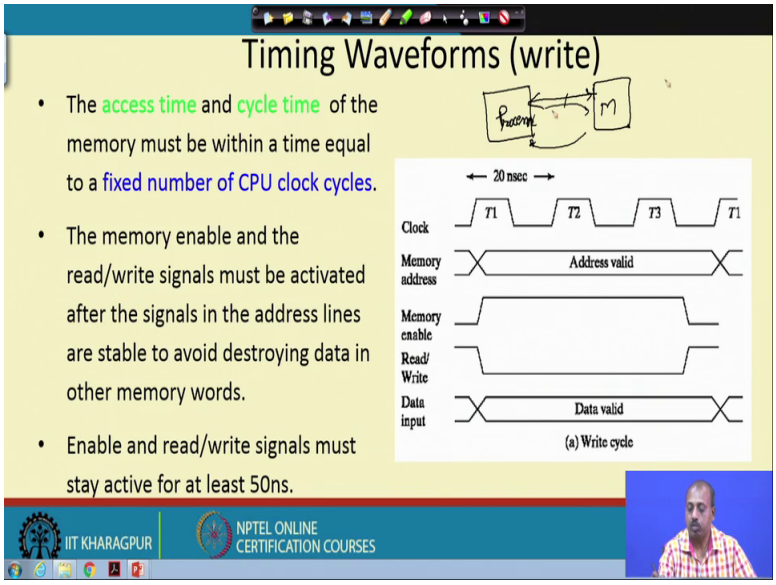
Similarly, we can do this read operation which is basically transferring a stored word out of memory so, here also it is similar so, apply first the binary address of the desired word to the address lines, activate the read input, and so then the content will be available on the a data output lines. So, commercial memory is they sometimes provides 2 control input so, reading and writing in a somewhat different organization. So, basically the problem that we have with the configuration that we have taken like say we have said that we have got a memory chip memory like this and it has got 2 controls one is read another is write.

Now, in this situation how do you tell that what a which operation your trying to do. So, it may be the situation that I do not want do a either read neither read nor write operation so, I do not want do any of those operations ok. So, somehow they should be a mechanism by which I should be able to distinguish between the situation that say I do not want to do any of the operation. So, this one possibility is like this memory chips they come with an enable input. So, with this enable line is 0 then what ever be this read write value so, no operation will be done. If enable line is 1, then if the read write bit is 0 so, that is a write operation and if enable is 1 and read write is 1 that is the read operation.

So, in many processors we see that it is written has read write bar so, so, the control lines will written has read write bar. So, with this enable is one and read write bar is equal to

0; that means, we are trying to do a write operation similarly if you this line is one; that means, you are trying to do a read operation. So, this is the organization of this control signals memory read write operations and all.

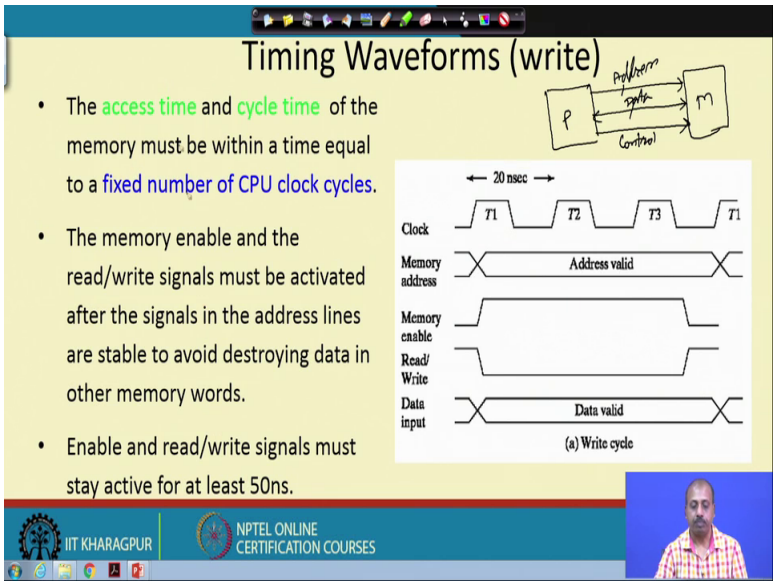
(Refer Slide Time: 12:44)



So, typically this memory access can be represented by means of a timing diagram so, what happens is that this memory chips they are connected to some processor for this which will be doing the operation. So, in our from our school days so, we know that if I have got a processor then this processor and memory if they are connected by means so, so they are connected by means of some lines and then whenever processors needs to write something on to the memory so, it will send the data to the memory.

Similarly, when it wants to read so, it gets the value there. And this line that I have shown here so, this is not a single line so, we will have the situations like this.

(Refer Slide Time: 13:34)



So, this lines they can be divided into 3 D at least 3 different set of lines so, this is this side is the processor this side is the memory. The first set of lines they are the address lines, then we have got the data lines these are the data lines and there are some control lines these are the control lines so, with basically read write or this enable so, they who has the control lines.

So, this processor how does it operate is that so, processor operate are some clock so, it so, this suppose this is the clock signal, where the duration of this clock signal period this 20 nanosecond so, this is just an example. Now this so, at the at T 1 so, this the processor will put the address on to the address bus so, address will be valid for the entire time T 1 T 2 T 3 or the three cycles and then after putting the address on to the address bus this memory enable signal is made high.

Then this read write bar line so, this is made low because this is write operation so, this read write bar line is made low, but before that so, data is put on to the data bus. So, you see that these 2 points. So, they occur simultaneously that is the memory address is made a valid address is put on to the address bus, valid data is put on to the data bus after that the memory signal memory enable signal is given, and this read write control is made low so, that this is the write operation then the memory will be doing the operation

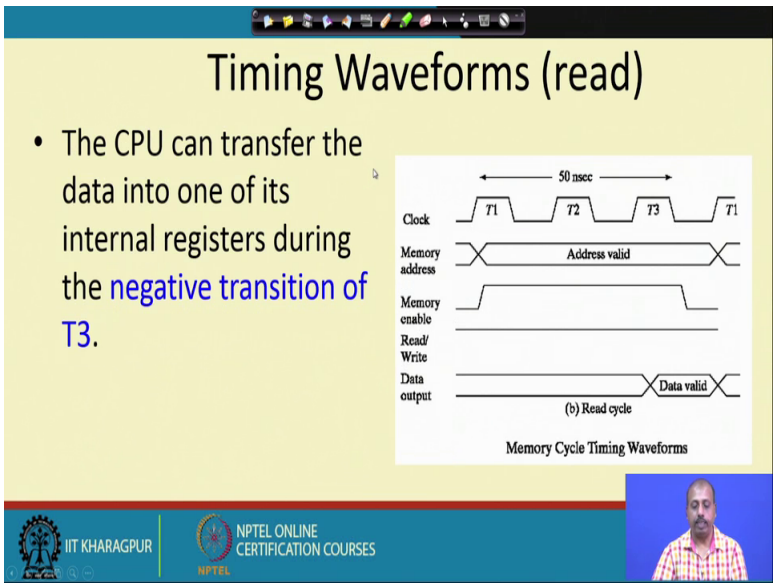
So, how much time this memory takes to respond with the data so, that is given by the cycle kind. So, this is this memory at the access time and cycle time of the memory must

be within a time equal to the fixed number of CPU clock cycles. So, access time means how much time it takes to access the memory so, for how much time this read write this address input then this read write control input should be valid. So, that memory will understand it and cycle time means after we have put the address data and say control line after how much time the operation will be done by the memories there is a cycle time.

So, this memory enable and read write signal must be activated after the after the signal in the address lines are stable, to avoid destroying data in other memory words. Because otherwise what will happen is that if this signals are given before they become valid so, they are having some wrong values. And if you given enable and this read write signals before that, then the memory will access some other location and that locations content will be lost, that location content will get destroy.

So, this is the problem so, it is always advisable that we first put this make this address and data lines valid, and then only we give this control signals. So, this enable and read write signals must be the must stay active for at least 50 nanoseconds. So, this is say typical situation like if your clock CPU processor clock is of period 20 nanosecond, then this read write signals are expected to be active for about 50 nanosecond.

(Refer Slide Time: 16:48)

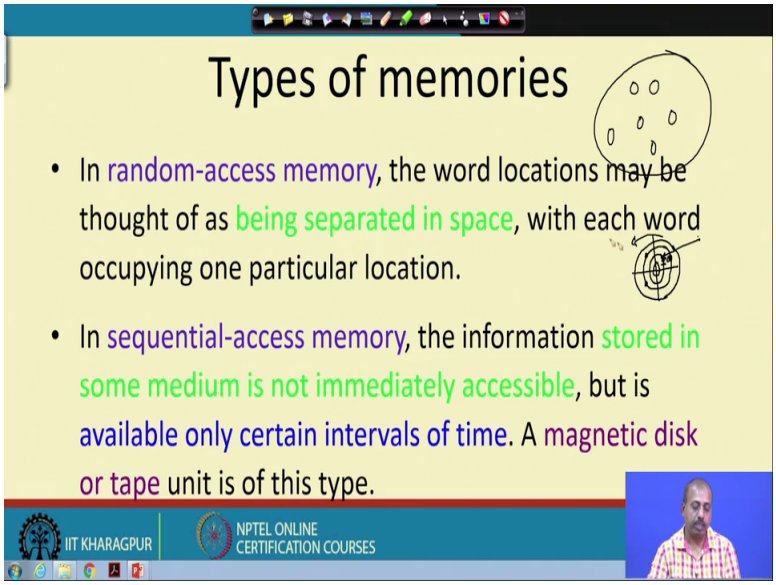


So, next will be looking into the read operation so, in case of read operation also the CPU can transfer data into one of its internal registers using this negative transition or T

3. So, here also the operation is like this the address valid address is put, memory enable signal is made active and this read write bar line is made active so, this is made high so, this is read operation is done. So, so after 50 nanosecond so, if I the memory cycle kind is 50 nanosecond so, after 50 nanosecond the data is available on to the data output so, this data valid comes here.

So, processor should take the data at this point so, it has a started the write operation at this point, but it should take the value at this point. So, this negative transition of T 3 so, this falling age of T 3 so, this should be used for storing the data into the CPU register because at that time only the data is valid so, data is definitely valid at this point of time so, this is the way this read operations will be done.

(Refer Slide Time: 17:50)



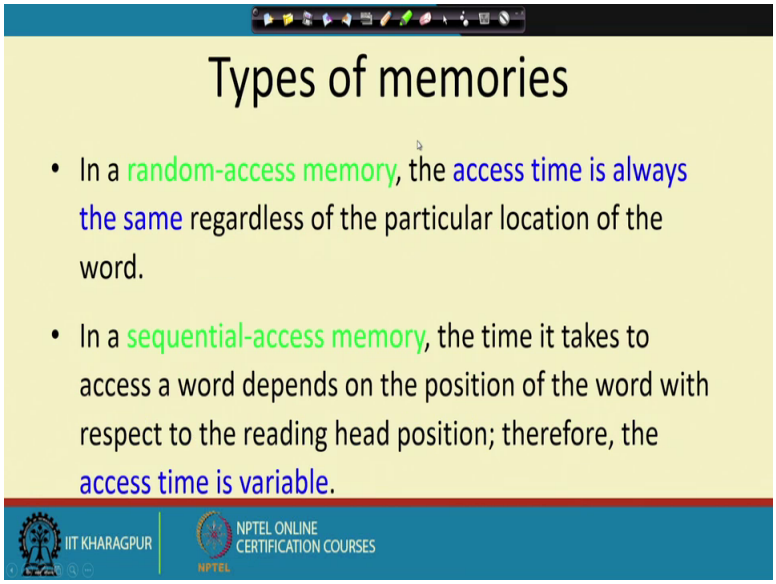
So, if you will going back to the types of memories so, in random access memory the world locations may thought has been separated in space with each word occupying one particular locations. So, as we can think of it logically as if we have got a in case of this random access memory so, we have got a set of locations, so all these are locations, and you can access any of them randomly so, there is no fixed ordering in which you have to access them. On the other hand this sequential access memory information is stored in some medium is not immediately accessible, but is available only certain intervals of time.

So, basically this is typically this is true for magnetic this for tape type of devices, so, they are what happens is that we have got these type of disk and there are some tracks unit there are some tracks which are further divided into sectors like this and there is a read write head. So, there is a read write head so, a suppose this is a read write head.

Now, so, this disk rotates and accordingly whichever comes under this read write head so, that value is read at that point so, so, since this disk is rotating so, after you have access say this location, then only you can access this location if I assume that the disk is rotating like this. So, you cannot access randomly so, it cannot be the case that first you access this then you access this then you access this so, that cannot happen. Because if you disk is rotating like this if the and initial read write head is at this point then it will this content will be read first then this content then this content.

So, that is a sequential access so, this semiconductor memory they are mostly random access type, and this secondary storage like say disk and all so, they are of this sequential access type.

(Refer Slide Time: 19:54)

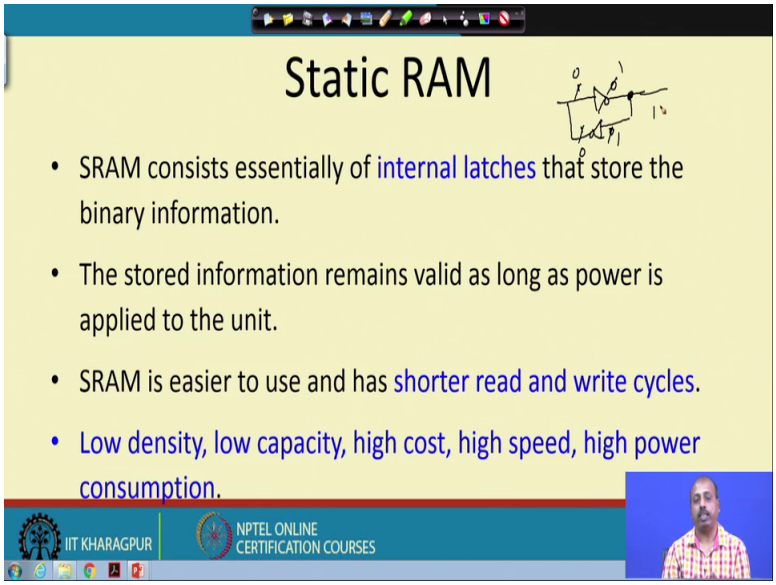


So, in random access memory access time is always same regardless of the particular location of the word so, it does not matter which location your accessing so, location accessing location 100 or location 200, they takes same amount of time.

Whereas in a sequential access memory the time takes a 2 access a word depends on the position on the word with respect to the reading head position. As I was telling previously that since the disk is rotating, and if the word that you are trying to access is close to the read write head then it will be access must faster compare to the situation where when this actual data point is will be come much later so, it is not adjacent to the current head position.

So, in that case it will take time for the disk to rotate and the head get align to that location. So, this is the problem so, all the sequential access and the access time becomes variable so, in case of random access memory. So, depending upon the technology we can immediately find out what is the access time, but in case if sequential access memory. So, it depends on the location that your trying to access accordingly the access time will vary.

(Refer Slide Time: 21:02)

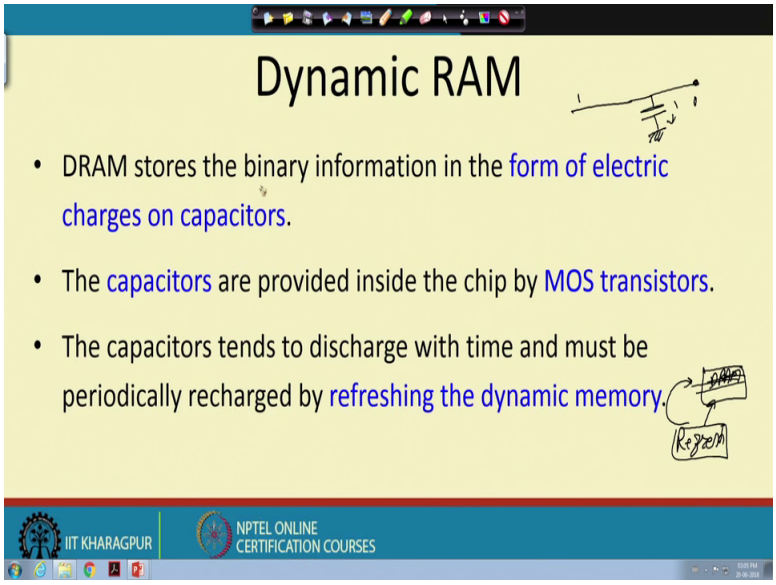


So, static RAM or s RAM so, this is one type of random access memory, it consist of essence essentially it has got latches that store the binary information. The stored information remains valid as long as power is applied to the unit in put so, SRAM is easier to use and have shorter read and write cycles, low density, low capacity, high cost, high speed high power consumption. So these are the features of this SRAM. So, this low density means par unit area how many such how many memory cells you can put.

So, low density means so, you cannot make the density high so, you cannot put large number of memory cells in a short in a small area. Low capacity so, if the density is low then naturally capacity will also be low because how many such cells, you can put into the chip so, that way the capacity will be low cost is high.

Cost is high because you are putting less cells there so, the cost is going to be high. But the advantage that you get is the high speed so, speed of operation is high the access time is faster, power consumption is also high. So, these are the problems with static ram, but still they are used because of this the speed high speed of property.

(Refer Slide Time: 22:29)



There is another variant of this RAM which is known has dynamic RAM or DRAM so, it stores the binary information in the form of electric charges on capacitors. So, in case of static RAM what happens is that so, conceptually a static RAM is like this so, we have got 2 inverters which are connected back to back so, which are connected back to back.

So, once this is say equal to once so, this is equal to 0 since this is equal to 0 so, this is equal to one as a result, this point will always remain has 0. On the other hand if this value is sense 0 then this will be 1 so, this will be one and this will be 0 and that way this point will always remain at one. So, this static RAM so, very simplistic structure so, that will be that can be formed by connecting to back to back inverters, but that way the cost of the system goes high there are other constructions of this SRAM so, will not going to that.

But in case of dynamic RAM what is done is that, you do not have such back to back inverters rather we have we take a capacitor. So, the capacitor is charged and this capacitor is actually holding the charge. So, if you what trying to store a 1 so, this capacitor is charge, and so, naturally so, it will be so, if you sense the value here so, you will get a 1. On the other hand after some time of course, we cannot make a perfect capacitor so, this capacitor will leak so, after some time the voltage level will come down as the charge leaks from this capacitor so, it will go towards 0.

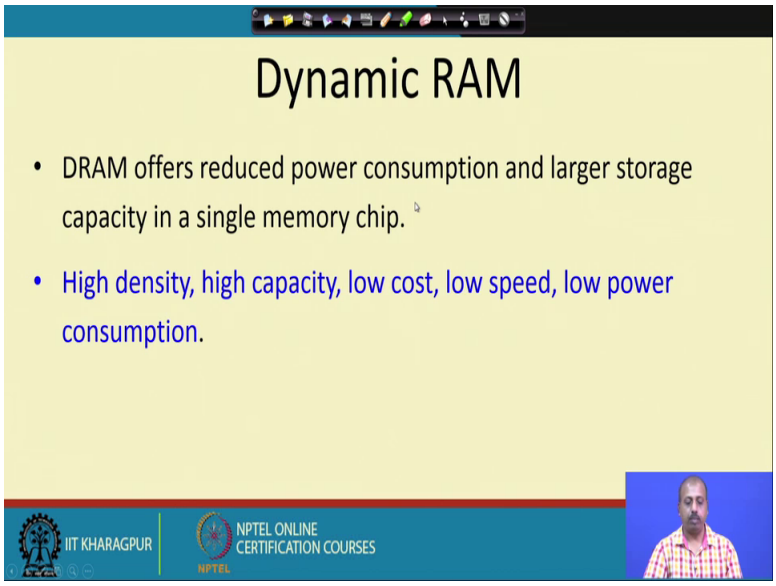
So, this is the so, what is what is required is that we need some sort of refreshing. So, this dram or dynamic RAM its stores binary information in the form of electric charges on capacitors, the capacitors are provided inside the chip b by means of MOS transistors. So, we have seen CMOS logic so, the MOS transistor we have seen there so, they are the gate of heat reacts act as the capacitor. So, this so, this gate value the gate part of heat acts as a capacitor to hold the bit pattern bit value.

And the capacitors as the capacitor they try tend to discharge with time so, periodically we must recharge the dynamic memory. So, this a compared to static memory the which does not require any extra circuitry for do a for holding the value so, in case of dynamic memory periodically you have to read the content of the memory and write it again so, that way we should have a refreshing circuitry.

So, it is so, apart from this if this is your D RAM chip, then along with that I should have a refreshing circuit. So, what this refreshing circuitry does is that, it reads the content of individual locations and write the same value there. So, as since it is recent periodically so, these capacitors will get recharge and then the value will be restored to their original one.

So, this is the idea of this dynamic RAM. So, of course, we do not go into the circuit part.

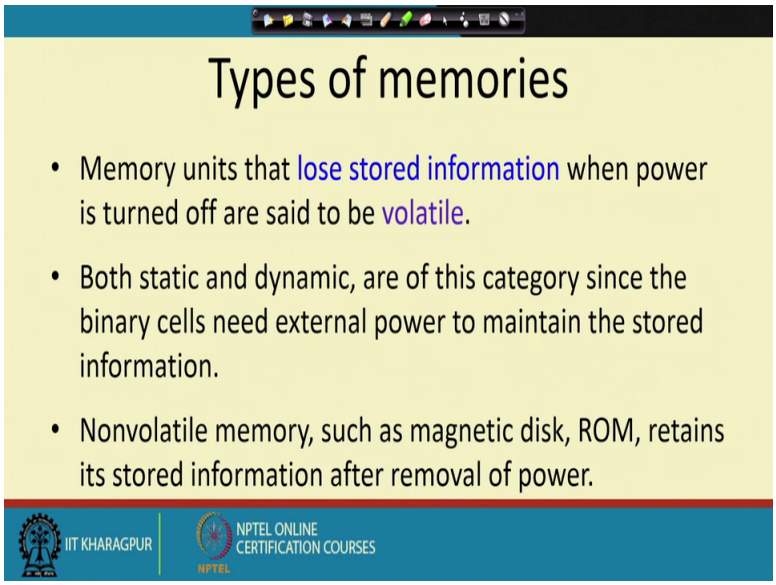
(Refer Slide Time: 25:44)



So, in case of dynamic RAM so, it offers reduced power consumption so, that is one of the largest advantage that we have with D RAM. Storage capacity is high because only a single capacitor is needed for making a dram chip sorry a dram cell, so, that way this where as in case of your SRAM cell so, standard SRAM cell it required 6 transistors whereas, standard dram cell it is requiring only a single transistor.

So, that the way capacity is much higher and high density high capacity, low cost, low speed, low power consumption. So, this low speed because it requires this refreshing periodic refreshing so, speed is not that high as this a SRAM; SRAM configuration, but it is still use because very much because of this high density and high capacity and low power consumption.

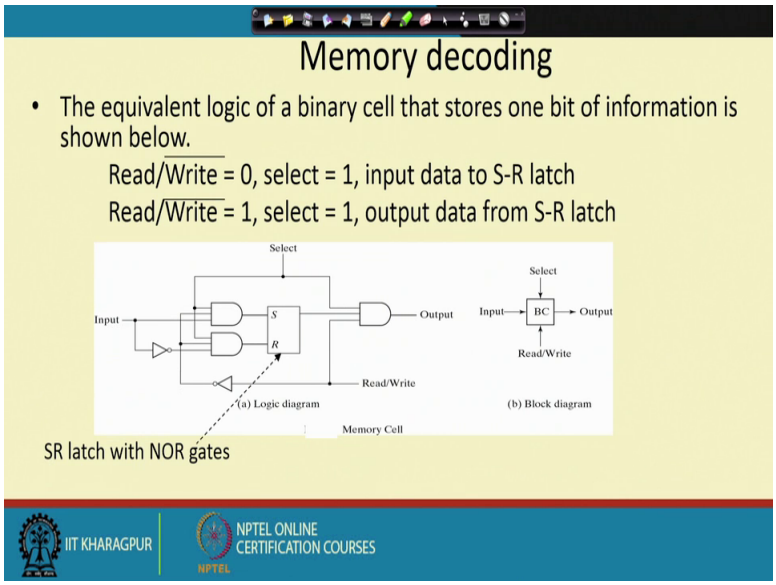
(Refer Slide Time: 26:44)



So, memory units that loose stored information when power is turned off is are said to be volatile. So, all this chip static RAM or dynamic RAM that I am talking about so, if you with the power is turned off then the content is lost so, this type of memory is there known as volatile memory. Both static and dynamic are of this category since the binary cells need external power to maintain the stored information.

There are nonvolatile memory is like magnetic disk and ROM, that retains its stored information even after removal of power. So, in a ROM so, the information is stored permanently so, you can you can switch of the ROM and, but still get the content there, when later on also the content will be there, but for RAM it will get destroyed.

(Refer Slide Time: 27:34)



Next we start with memory decoding; so memory decoding process. So, this will try to figure out like how this addresses will be selecting a particular location, and then. So, here the we can logically we can think of as if there is an SR flip lop and then this SR flip lop is getting selected by means of this read write lines. So, this is just a logical diagram. So, ideally it does not happen like this so, that is a there are separate codes on memory design. So, that will be talking about this cell design part, but for our course so, will be looking into this logical diagram and try to understand the philosophy of operation.