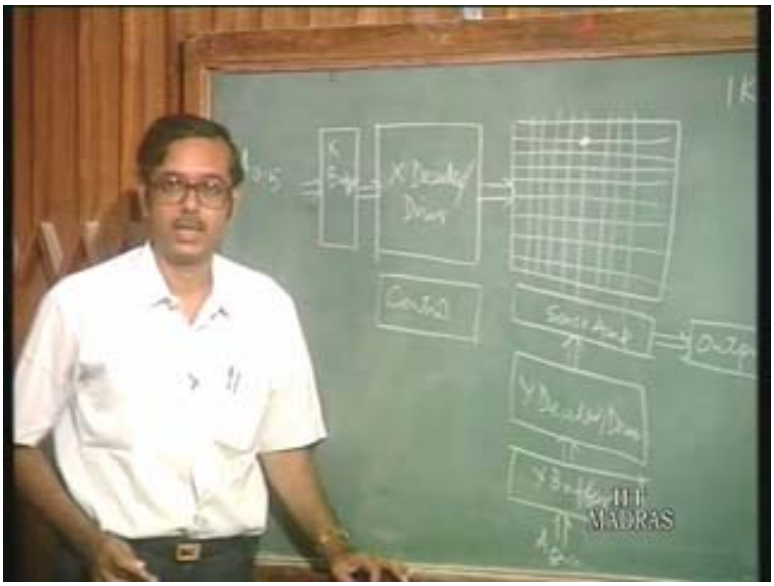
Digital Integrated Circuits
Dr. Amitava Dasgupta
Department of Electrical Engineering
Indian Institute of Technology, Madras
Lecture No # 32
Semiconductor Memories;
ECL RAM

Today's class we shall discuss the realization of semi-conductor memories. We have already seen in the last class, the block diagram of a memory chip. You have the decoder x, we have the buffer here, x buffer then x decoder and driver which is followed by the memory array, cell array and here you have the y buffer then y decoder driver and the sense amplifier, you have the control circuit and you have the output buffer. So we have seen that if you have say for example 2 to the power n memory cells, you are going to have n address lines.

This is of course assuming that the memory is configured as one bit memory that is if you have say 1 k into one bit memory, 1 k is 2 to the power 10 so you have 10 address lines in order to select the memory. That is the 10 address lines are divided into 5 address lines going to the x buffer and 5 address lines going to the y buffer. This one may be A is 0 to 5 and this one may be A 6 to 10 and this goes to the decoders. At the output of the decoders if you have n input lines, you have 2 to the power n output lines. Here again you have 2 to the power n address output lines. Here you have an array, so you have the output of the x decoder, you have 2 to the power n lines this way and you have again number of columns. So you have number of rows and number of columns.

(Refer Slide Time: 05:44)



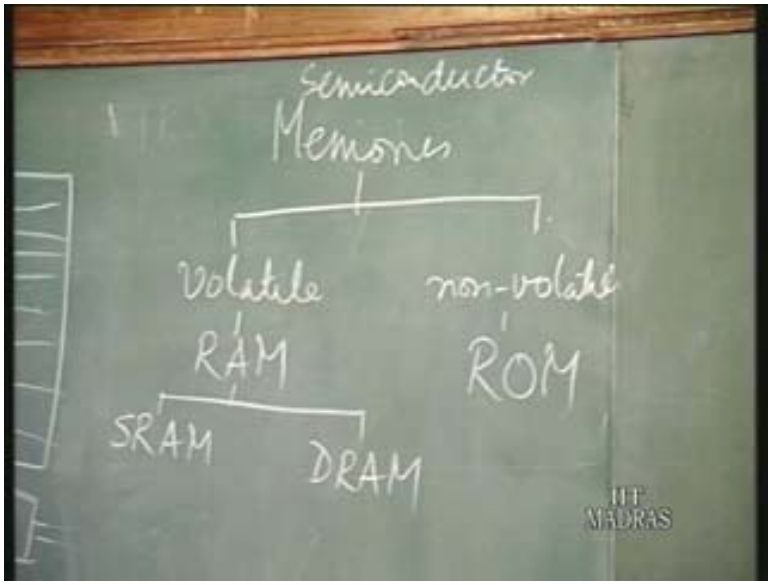
At the intersection of the rows and columns you have the memory cells. Basically what you do is this x address decoder selects one of the rows that is one of the rows is selected, one row goes high and then each column has the information corresponding to the cells located at that particular row. Then what we do is you have a number of sense amplifiers. The number of sense amplifiers is equal to the number of columns and then finally this y decoder driver selects one sense amplifier and you have the output corresponding to the particular sense amplifier which is selected.

Basically you are selecting one row with the help of the x decoder driver and you are selecting one column with the help of the y decoder driver and basically you are selecting the cell at the intersection of that particular row and that particular column. Again it is an example, if you have 1 k into one bit memory, you require 10 address lines, so the x 5 and y and you have 32 rows and 32 columns, 2 to the power 5 and 2 to the power 5. So you are selecting one particular cell here. Now of course if you have 1 k into 16 bit memory say, so you will have 16 k cells. You have 2 to the power 10 into 2 to the power 4 cells. So you basically have 16 such memory chips in parallel, so you will have 16 times of the sense amplifiers and you have 16 output buffers.

Basically if you have so many of them side by side, you will have 16 times the number of columns. So it is going to be a wide memory. So now if you want to make it a square memory, usually you want to make the array as square array. You remove 2 address lines here and put it in the x buffer. So you have 7 address lines here, 3 address lines here say and then basically you have 16 of them so 3, 2 to the power 3 into 2 to the power 4, so you have 2 to the power 7 columns and here you have 7 of them, so you have 2 to the power 7 rows. So you can make it a square memory. So this is about the memory organization.

Now you can have different types of memories depending on the type of cell which you have, the way you are storing the information. Memories as such semi-conductor memories can be basically classified into 2 types, one is volatile and the other is non-volatile. Volatile means the memory information is lost if the power supply is not there. You require the power supply for retaining the memory. Nonvolatile is even with without any power supply, memory is retained. So examples of this is the different types of ROM which actually do not require any power supply to retain this whatever is stored which is inherently stored there. We shall come to that, here in a volatile you have the RAM's and this RAM's again can be classified into the static RAM and the dynamic RAM. Again they are different because in the mechanism of storage of information. In a static RAM, the information is going to be retained as long as the power supply is there.

(Refer Slide Time: 08:17)



But in a dynamic RAM the mechanism of storage is by charging a capacitance but what happens is if after some time the capacitance gets discharged because there are leakage paths and the charge may flow out through the leakage paths and whatever is stored in the memory may be lost. So you have to do periodic refreshing of memory. So that is a dynamic RAM, so we shall study all this different types of memories. Look at how their cells look like and how you can do the major operation that is reading from a memory as well as writing into a memory cell.

First we shall take up the static RAM and see the different types of static RAM's which you can have, just like we have seen different logic families that is basically the same logic can be realized using different types of circuits. You have seen bipolar logic families, MOS logic families similarly for static RAM also you can have different types of static RAM. For example bipolar static RAM's, MOS static RAM's, CMOS static RAM, bi CMOS static RAM. So you have the different types and of course the circuitry is different. So we shall go through that and similarly for the other memories.

Basically for all these memories, the architecture or the organization is like this which is similar. In fact all these memories are what are called random access memories in the sense that random access memory is such that the access time for each memory location irrespective of the memory location is the same. That is if you want to access any memory location whether it is here or here or here doesn't matter. The access time is

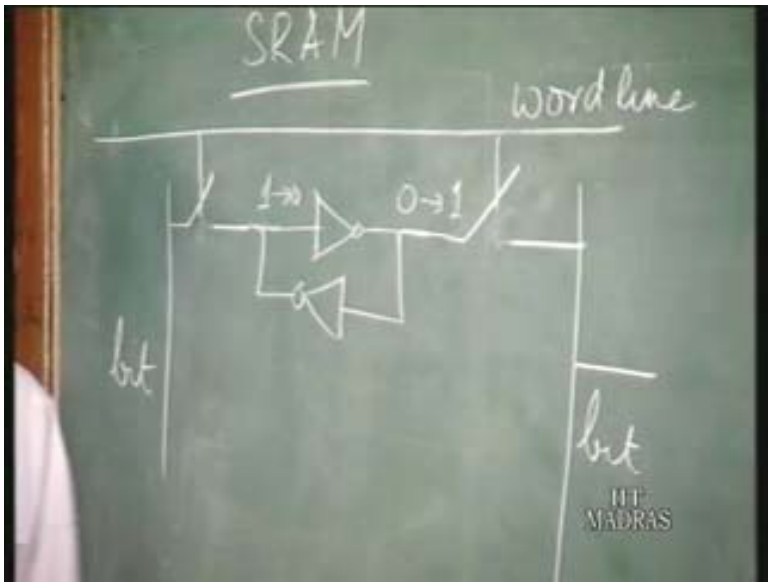
going to be the same. This is as opposed to the sequential access memory where you have to access in a sequence.

For a practical example take a cassette tape, if you have different things stored in that cassette tape depending on where you are located now. If you want to locate another memory location, the access time is going to be different for different locations.

So here these are all random access that is the access times are the same irrespective of which memory location you want to access. Of course you also have semi-conductor sequential access memories but they are not much in use. So we will not actually discuss that. The static RAM cell consists of two inverter connected back to back. So this is what is used to store information in a static RAM, so the two inverters. Here you have logic one, this is an inverter so here you have logic 0, this is fed back so this is a perfectly stable configuration (Refer Slide Time: 12:25).

Once you have anything here, this will be retained as long as the power supply is available for these inverters. Now suppose you change this to 0, this will automatically change to 1 and again this is the perfectly stable configuration. This is the basic structure of a static RAM cell, 2 inverters connected back to back. Now what do you have is in order to access this cell, you have 2 switches which is connected like this and this switches are actually activated by this word line. When this line is high, we can say that this switch goes to the closed configuration and when this line is low this switches are open and this end is connected to the bit line. This line is called the word line and these lines are called the bit lines. So one line is called a bit and the other line is called a bit bar.

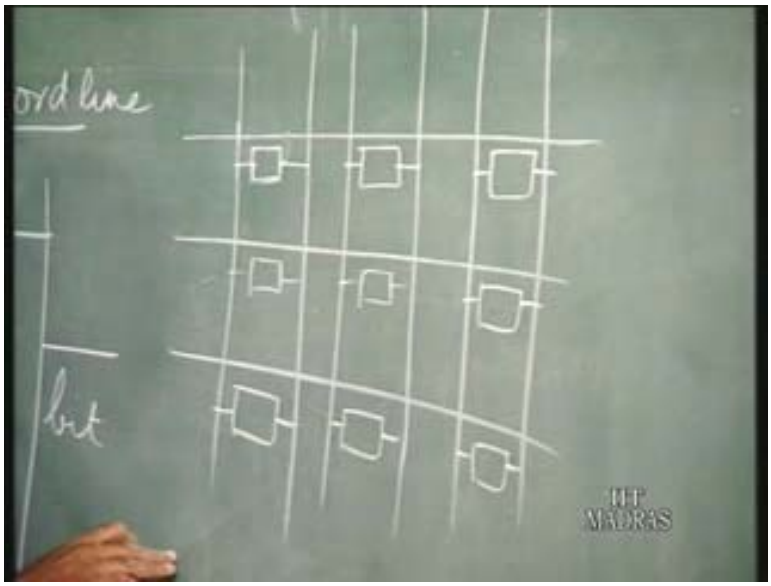
(Refer Slide Time: 13:53)



So this word line is the output of the x decoder. So you have such cells, if this is the memory cell, so this is the word line say so many word lines and with each word line you have the bit and bit bar lines and you have cell connected here. This is a bit and bit bar line, this corresponds to one column so here you have another cell. Here you will have another cell and similarly you have cells connected everywhere. This is a word line, there are three word line shown here and there are three columns shown here.

Depending on which word line is high say for example if this word line is high, all the cells connected on this particular row will be connected to the bit and bit bar lines. In all the other rows, the cells will be disconnected from the bit and bit bar lines. So in this column if this particular row is high, you have information corresponding to this particular cell that is again going back to this. So if this particular row is high, the information on what is stored here is available on the bit and bit bar lines because this switch is closed.

(Refer Slide Time: 15:39)

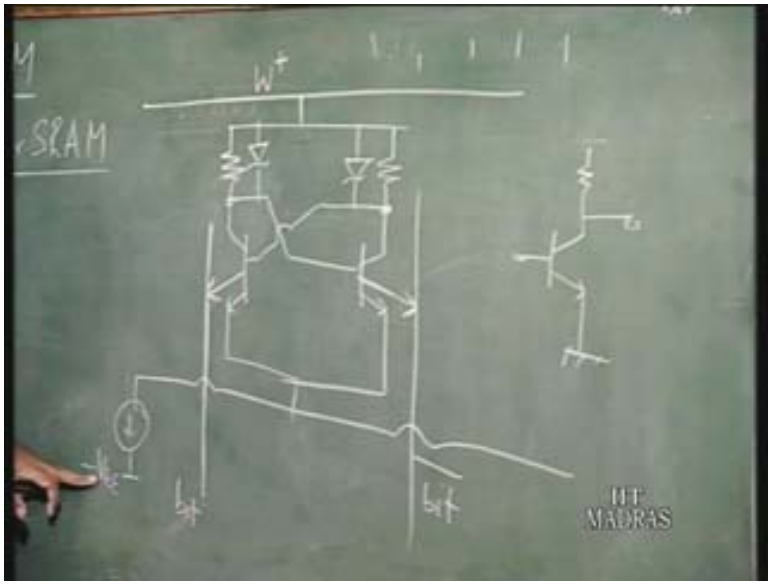


So this information is as now, if you want to read the cell, you have to have some mechanism to interpret whatever is stored in this particular cell. If you want to write into a cell basically again you select that particular row and then one particular column you have to energize and you must be able to force one end, this point if you want to change from one to zero you must make this point go to ground, force it to ground so that this point becomes zero. So automatically this goes to one and like so force a particular value into that cell.

So that is how you can read and write into a particular cell, so this is in general. Now we shall look at specifically different types of static RAM memories. Firstly we shall

take up the bipolar static RAM. So the bipolar static RAM cell los something like this. So again you have the word line. You have actually two lines like this, w plus. Here the cell is something like this, bit and this goes here. So you can recognize the bipolar inverter here. So you have 2 inverters bipolar inverters, so this is a current source here, this is going to minus V_{EE}.

(Refer Slide Time: 20:43)

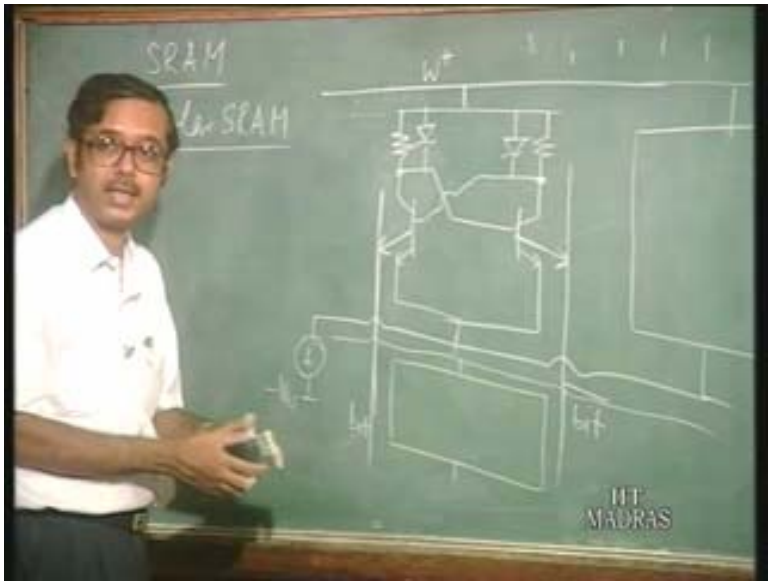


Basically this bipolar memory is like an ECL memory, it is an ECL cell. If you see this, this is a bipolar inverter that is transistor with a resistance in the collector side and the collector output is going to be the inverted form of the base input and this collector output is connected to the base of this transistor. If you consider this as a bipolar inverter something like this, so this is the input, this is the output. So the output of this is fed as input here and the output of first inverter is fed as input to the second inverter, just like the cell which we had discussed, two inverters connected back to back.

The output of one inverter is connected as the input to the other. The other thing is this transistor is a multi-emitter transistor it has two emitters also this one and this emitter is connected to the bit line and this say is connected to bit bar line. So you have the word line here, you have this cell here. This is an emitter coupled structure, you can see here both the emitters are shorted and this goes to a current source just like in an ECL. This goes to a current source which goes to a negative power supply. So this is the bipolar cell so in fact you are going to have similar cells like this and all of them are connected

like this. This corresponds to one row, this row you have large number of cells like this.

(Refer Slide Time: 21:22)



Similarly on this column you are going to have other cells or there are different word lines to energize these different cells in this particular column. This is one particular cell now let us see how this memory works. For a memory you basically have three states, you can say or 3 conditions. One is the holding state where you just retain whatever is stored in the memory, the same information is retained. Basically you are not reading or writing, you are just holding whatever is in the memory.

In this case what is done is in the holding state this w plus line goes low and the bit and bit bar line are both high that is in the holding state. Basically it means that when you are not selecting a particular cell for reading or writing purpose, the w plus should go low, b should be low and bit and bit bar line should both be high. Obviously it means

that if you are selecting for reading or writing purposes, it has to be in a different condition. We will come to that what is going to be for w plus and bit and bit bar lines.

Basically what happens is if the w plus goes low, now what you have here is a current source. So the voltage in this line so which is actually called the w minus line, which is the common emitter point here. So what happens is if the w plus line goes up, this w minus potential is also going to go up or if the w plus potential goes down, w minus potential goes down. So this is a current source. Basically the voltage at this point can be anything it doesn't matter. So now what happens is here you see, you have two cross coupled inverters.

Usually what is going to happen is one of these transistors is going to be on and the other is going to be off say suppose this $T_{1.}$ and $T_{2.}$, $T_{1.}$ is on and $T_{2.}$ is off. So $T_{1.}$ is on means what is going to happen is there is going to be a collector current flowing and if there is a collector current flowing, the voltage at the base of $T_{2.}$ is going to be w plus that is what is a voltage at w plus minus the drop across this resistance. This is off and since this is off, there is no current flowing across this resistance.
The base voltage of $T_{1.}$ is going to be the voltage at w one and since the emitter voltages are constant so the difference in the base emitter voltages of $T_{1.}$ and $T_{2.}$ is going to be the drop across this resistance, whatever is the $I_{R.}$ drop across this resistance. Suppose you have 0.3 volts drop, here this short key diode actually limits the drop across this resistance. So this drop, when this is conducting fully is going to be around 0.3 volts. So the base emitter voltage of T two is going to be 0.3 volts less than the base emitter voltage of $T_{1.}$ and since when $T_{1.}$ is conducting say the voltage is say 0.8 volts, $V_{BE.}$ of $T_{1.}$ is 0.8 volts. So $V_{BE.}$ of $T_{2.}$ is going to be around 0.5 volts. So obviously when $V_{BE.}$ is 0.5 volts, it is below the cut in voltage so this is going to be off. So the collector current is going to be zero so there is no drop across this (Refer Slide Time: 26:11).

Here you have the voltage between w plus and w minus, if you take this path is going to be basically equal to $V_{BE.}$ of $T_{1.}$. So this is true when $T_{1.}$ is conducting, if $T_{2.}$ is conducting you will have a similar argument, you can show that $T_{1.}$ is going to be off and again the difference between w plus and w minus is the base emitter drop of $T_{2.}$ which is around 0.8 volts. So what happens is this is constant. Now if you raise w plus voltage, w minus is also going to follow with a difference of 0.8 volts. Since this is the current source, if the voltage across the current source changes, it doesn't make any difference. In the ideal current source, the current is going to be independent of the voltage drop across it.

The same current is flowing, this current which is flowing depends on the current which is flowing through this transistor. Basically if w plus is increased, w minus follows, it goes up. W plus is reduced w minus follows, it goes down. So now in this condition if you make w plus low, w minus is also low. You are raising the bit and the bit bar lines, this voltage is made high, bit and bit bar both are made high. So what is

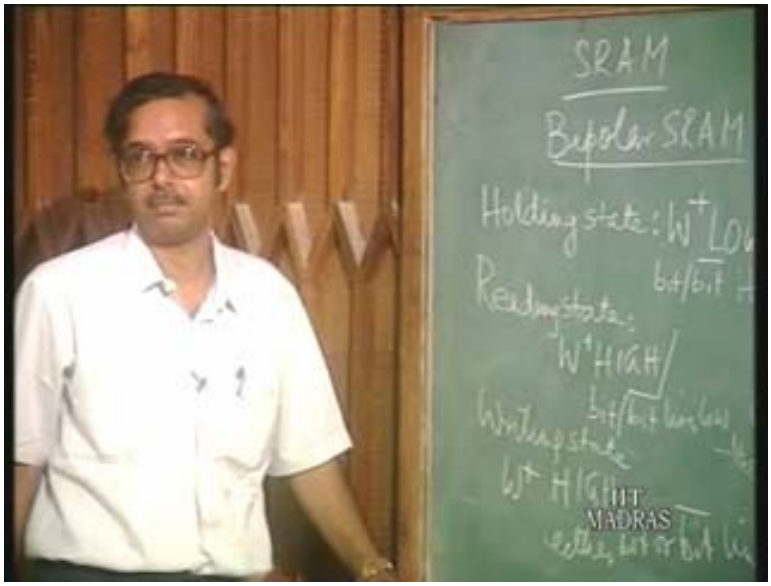
happening? This emitter voltage is going to be low so if you consider the two emitters of any transistor $T_{1.}$ or $T_{2.}$, this emitter which is called the holding emitter which is connected to w minus that voltage is going to be much less compared to the emitter voltage of this other emitter which is called the bit line emitter. You have made this voltage high, you have made w minus low.

If you consider any of this transistors $T_{1.}$ and $T_{2.}$, what is going to happen is since this emitter point, the emitter voltage of the holding emitter is less than the bit line emitter, it is the holding emitter which is going to conduct and the bit line emitter is not going to conduct. The base emitter junction is going to be much less and that is going to be off. So basically there is no current which is going to flow along the bit and bit bar lines. So in the holding state what happens is you reduce w plus which means that you reduce w minus also and you are reducing the voltage of the holding emitter and you are raising the voltage of the bit line emitter. So the holding base emitter junction is going to be much higher than that of the bit line base emitter junction. So this is going to conduct and the other bit line base emitter junction does not conduct. That is the holding state.

Now suppose you want to read the state of the memory, whether a one or zero has been stored. You have to follow a convention that is suppose you say that when $T_{1.}$ is on what you have stored is a one, if $T_{2.}$ is on what we have stored is a zero. You can have these two condition. When $T_{1.}$ is on obviously $T_{2.}$ is off and when $T_{2.}$ is on, obviously $T_{1.}$ is off. Now you want to read the state, basically you want to see which of these transistors is on. Then what you do is basically you have to transfer control from the holding emitters to the bit line emitters. So what you do is you make w plus high and bit and bit bar lines low. Basically what you are doing is one of the rows, one of the w plus line goes high and one of the bit and bit bar columns is going to go low. At the intersection of that row and column, only in one cell the word line is high and the bit and bit bar line are low. In the other cells, you do not have this particular condition.

Now what happens is if you are making the word line high, you are raising this. If you are making w plus high, you are also raising w minus potential again because this drop is almost equal to $V_{BE.}$ of a transistor, the difference between w plus and w minus. So when this goes up this is also going to pull it along, the potential. So this emitter point voltage goes up, the bit and bit bar lines potential goes down. So what happens? Suppose $T_{2.}$ was on that means the base voltage of $T_{2.}$ we have seen was around 0.3 volts greater than the base voltage of $T_{1.}$. So the emitter of $T_{2.}$ connected to the bit line is going to conduct whereas $T_{1.}$ is not going to conduct. So basically the control is transferred from the holding emitters to the bit line emitters.

(Refer Slide Time: 36:38)



So you are going to get a differential current flowing in the two bit lines. One of the emitter is going to conduct, one of the bit line is going to conduct, current is going to flow. The other current is not going to flow. This difference can be sensed by a sense amplifier. We shall come to that, in fact we shall discuss the sense amplifier when we take up bi CMOS memories. Basically because in bi CMOS also you use bi polar sensing so that is going to be similar. Here what you are doing is you are going to sense the difference in the two currents, the currents flowing in the two bit lines.

If $T_{2.}$ is on, current is going to flow in this line whereas in the other line there is no current flowing and of course if $T_{1.}$ was on, it's going to be different. Reading state w plus goes high and bit and bit bar lines go high. Now writing state; again what you do is w plus goes high and one of the bit line is going to go low. So either bit or bit bar lines go low. In all other columns, the bit lines are high. So in all other columns the bit lines are high means that the transfer is still with the holding emitters. In one particular row, the w plus line goes high and in one of particular column one of the bit lines goes low.

So what is happening is by raising the w minus line, you are basically transferring control through the bit line emitters and you are reducing the potential in one particular bit line. So what you are doing is since you reduce the potential there so that base emitter junction, the voltage is going to be much more. $V_{BE.}$ of the transistor is going to be much more because the emitter voltage is forcefully reduced. So that transistor will turn on. If that transistor turned on there will be a collector current flowing and if that collector current is flowing so it is going to turn off the other transistor. Also because the bit line or the emitter voltage is going to be made much high on the other transistor. So that other transistor the VBE is going to be pretty low and so it is going to be turned off. So by doing this, we can write information into a particular cell that is depending

on which line is made low, we can either make $T_{1.}$ conduct or we can make $T_{2.}$ conduct. So this is how a bi polar static RAM cell works.

Again as you can see this particular cell functions by adjusting the voltages on the bit and bit bar lines. That is by adjusting the word line voltages and the bit and bit bar line voltages, you are selecting a particular cell. You do not have special switches actually to select a particular cell. Now for this bi polar memory, the advantages are of course that it is going to be quite high speed that is always there. The reason is that for this cell to operate the voltage swings on the bit lines are quite low compared to as a CMOS type of memory. The swing on the bit line voltages, you require much larger swings to be able to sense that. Here basically you are sensing the current which is flowing here. So the bit and bit bar line voltages they need not vary much, you are just sensing the current.

The potentials on this line is not varying by much. In fact the voltage variations on each of these lines for the entire different operations will be within one volt or so, less than one volt. It does not require to go beyond that because the VBE of a bi polar transistor is just point eight volts. To switch on and switch off a bi polar transistor, you require a very small voltage variation. So that is why you require less time to sense whatever is stored in that particular cell. Also because the current on this line is driven by bi polar transistors and because of the large current drive capability of bi polar transistors this is going to be much faster. The bit lines are driven by bi polar transistors. So that is for this reasons, this memory cell or this type of bi polar static RAM is going to be faster compared to MOS type memories but there is a basic disadvantage here.

The disadvantage is that it is very difficult to make large size memories. The reason is like this say in the holding state. If you are selecting a particular cell, all the other cells I mean are in the holding state. In fact even if you are selecting a cell there is going to be a current which is flowing and this current say suppose cell is in the holding state, it has been found that you require at least of the order of few micro amperes of current to flow through the transistors to keep it on. Suppose $T_{1.}$ is on, at least a few micro amperes of current must be flowing through this transistor to retain it in on condition because you have to switch the other one off also. If you say that the holding current for a particular cell is say 2 micro amperes and if you think of a one mega bit memory, it actually requires two amperes of current flowing through this memory, just flowing through the cells which is the very large amount of current. You see that it is very difficult because of the large power dissipation problem to have large size ECL type memories like this.

The advantage is of course higher speed but the disadvantage is you cannot go for very large size memories. As you can in the case of say MOS type of memories because the static power dissipation is very much less. So that is the basic disadvantage. Another disadvantage with this is that the control signals that is for reading, writing, holding

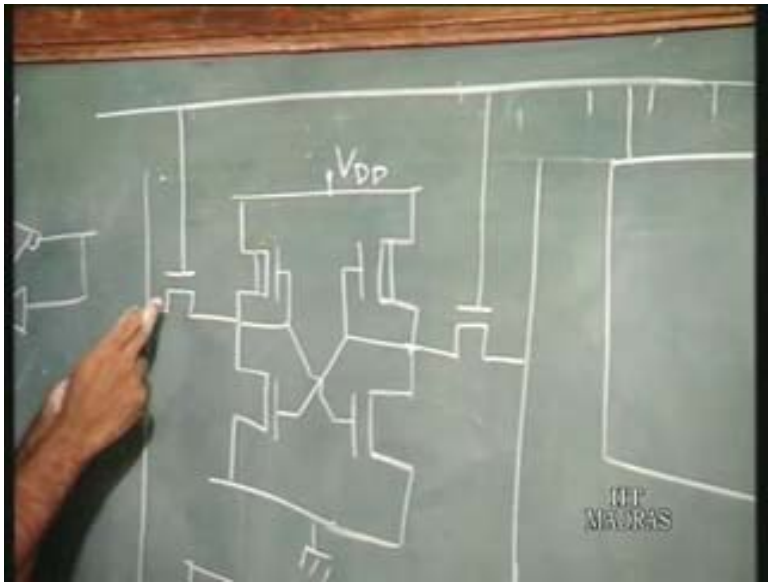
operation they are much more complicated here compared to the CMOS memories, MOS and type of memories.

Here you have to generate the signals and they have to be very much precise because the variation in voltages are very small from one state to the other. So the control signals must be generated or much more complicated, generation of these control signals when you compare with MOS type of memories. This is the comparison of the bi polar type memories with CMOS memories. So obviously nowadays the trend is to have the memory chips having a large density, large sizes. So for that reason one has to go to MOS type of memories because also the density we can achieve in MOS type of memories is much larger primarily because the power dissipation is going to be much less because you do not have this static power dissipation. As also MOS transistors can occupy much less area compared to bipolar transistors and say suppose you have a one mega bit memory, you actually have one million such cells. You must realize that you have one million such cells.

So the major portion of the memory chip is taken up by the cell array and so if the cell size can be reduced then you can have smaller area. So that is another constraint. So MOS type memories are more suited for larger density memories whereas bi polar is not very much suited for that but of course if you want a small memory but very high speed memory for example cache memory in the computers, you have a small density memory but which you want very high speed because the computer is going to interact with that memory much more regularly than the other memory. So you can have that made of bi polar memories. Now just talk about the MOS memory cell. Again the MOS memory cell has the similar structure, the same structure as we have seen the two inverters connected back to back. Now this can be NMOS that is NMOS inverter you have say a depletion mode NMOS.

This is a depletion mode transistor so in that case this is an inverter structure, the two inverters. So the output of this inverter is connected to the input here and the output of this inverter is connected to the input here. So this is a cell and then you have pass transistors which act as switches to connect them to the bit line. this is the word line so on this row you will have large number of such similar cells, so this is on this row and on this column also you will have large number of such cells. This is a cell. Now this cell is going to be selected when this particular word line goes high that is x decoder output goes high for this.

(Refer Slide Time: 46:11)



So these two transistors turn on and this particular cell is connected to the bit lines. Since only one word line can be high at a time, so in each column only one cell is connected to the bit lines. So you have information corresponding to that on the bit lines and then of course the help of a y decoder, you choose one particular column and that is the output which you get. Then basically choosing one particular cell because you have already chosen one particular row and now with the help of the y decoder, you choose the output of one particular column. So here again this is an inverter, so if the input of an inverter is high, the output is going to be low and if you have one here which is the output of this particular inverter, so this input is high.

So the output is going to be low and so this is the perfect stable configuration. But of course the NMOS, the problem is a static power dissipation so almost all MOS memories nowadays use CMOS because in a CMOS inverter you know that static power dissipation is not there. You see most of the memory cell they are not in changing state for most of the time. So there is no power dissipation as such. There is power dissipation only when there is a change of state of the memory cells otherwise there is no power dissipation as such in the cells. This is an inverter, the output of this inverter is fed as a input here. The output of this inverter is fed as input here, output of this goes as input here to the other inverter. There are two inverters and the output of one goes as input to the other. This is a CMOS memory cell and we can select again the particular cell in the same way be making one word line go high and then selecting one of the different columns. An information is stored by making one of the inverter output to go high and the other is output is low.

So you can say that if this is the bit line and this is the bit bar line when this 1 is stored, here 0 is stored and again we can force this memory cell to go to a particular state by forcing some certain voltages on the bit and bit bar line and so controlling the state of

this particular cell that is basically the writing operation and while reading you just read the information stored in that particular cell.

Now in the next class we shall take up again some aspects of design of this CMOS memory and then we shall see how you can combine the advantages of both bi polar static RAM and a CMOS static RAM, in a bi CMOS static RAM which combines the advantages of static RAM of being a low power dissipation memory. So you can have high density of memory, at the same time you get the advantages of bi polar that is a high speed memory. So if you want high density as well as high speed, bi CMOS can be used where you derive the advantages of the two types of memories.