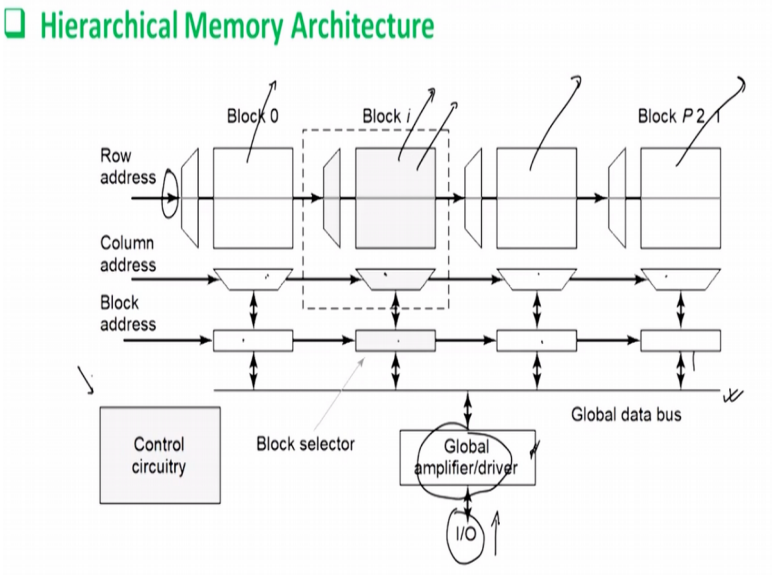
CMOS Digital VLSI Design
Prof. Sudeb Dasgupta
Department of Electronics and Communication Engineering
Indian Institute of Technology – Roorkee

Module No # 08
Lecture No # 40
Concept of Memory and its Designing – II

Hello everybody and welcome to the NPTEL online certification course of the CMOS digital VLSI design. We will start with the second module of memory and its design and in the first module we have seen the very the basic architecture of the memory and what is the basic parameters for example how do you define read access time and write access time. We also seen how does what are the various peripherals which is generally there.

And let me therefore come to this module and explain to you as I was discussing earlier that we will always keep generally the high design must larger than the weight so try to keep the height larger than the weight. So that data decoders are relatively easy to handle the bit of data but column decoders are made for selecting appropriate words as I discussed with you and row decoders are selection of the bit line and bit line bar.

(Refer Slide Time: 01:19)



Now let me come to the hierarchical memory architecture and therefore which means that if I have a memory generally then this how it looks like so this is the block so this is the block 1 block 0 which you see of block 1 this is the memory block which is available to you right and

you have a row address which is this one which gets access to a row recorder across the whole blocks of memory.

You will have a also column decoders or column address decoders is this and this and they will be actually responsible for selecting the appropriate beater bar line and this will be finally given by the block address decoder here right and therefore there will be global data bus which is fed here so the data can be written or read from this global data bus you will have a global amplifier or driver which is responsible for amplifying signal and then finally (()) (02:17) to the final IO.

So this is from where you will get inputs available to you this will amplify your signal send it to a global data bus. If you want to write the data say in the say Ith block then you can actually select that particular block using the block address and then use the column address to select the particular bit line and bit bar line and similarly the row address to the row challenge and then you can actually store a bit of data from that particular.

You also have control circuitry here which controls the flow of the signal between the IO pads and the block global diagram. Now the idea behind this is that the global data bus which you see is if you look very closely is effectively isolated from the memory block because you do have a block address and a column address decoders which help you to decode the address.

Now generally there is a problem that you require that your noise margin that your margin should be high and the reason is that is this global data bus is connected to IO which is external data therefore this global data bus is prone to large amount of noise right. So these noise origin might be right from a VDD difference change in power line noise or it might be even a 1 / F noise is it might be even a process noise so and hence so forth as a result what will happen is that at certain point of time if your block address and the column address are both in a own state your block Ith block will be connected to the global data bus.

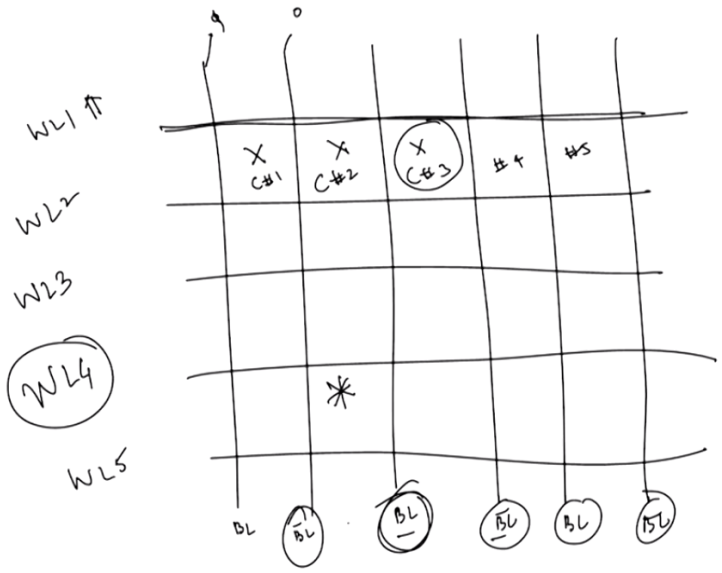
And therefore even a small finite change in the noise margin as small change or a small input in the noise in a global data bus will actually change the noise or change the will insert noise into the system and a one might be actually read as a 0 or vice-versa right. So we have seen in our previous discussion that there two noise margins which is high noise margins and low noise

margins and depending on whether you are storing a bit 1 or 0 or reading 1 and 0 from the point you should be very careful about that.

For example if you are reading 1 or you are reading 0 and then suddenly there is noise inserted into the system that 0 might be even data is 1 provided the noise voltages approximately with DD / 2. So you should be very careful about these noise getting separated out at the very initial stages and people use is hierarchical architecture to that okay let us look at the memory code we will not be looking at the peripherals at this stage may be but we will be looking at the memory code where 1 bit of information will be stored here.

The general memory code though not shown I will just show it you is basically made up of sorry just to make it for you.

(Refer Slide Time: 05:10)



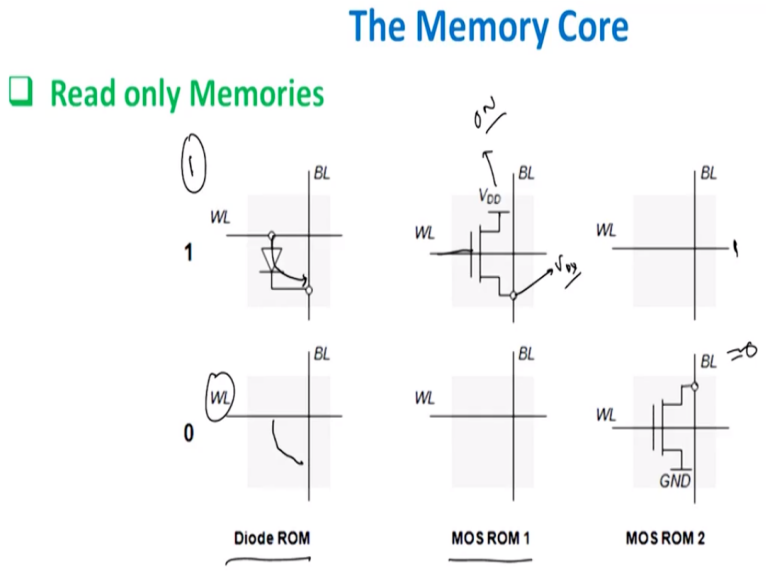
So it is made up of large number of word lines so these are your word lines WL1, WL2 and then it shared bit line architecture. So you will have such sharing so say this is bit line bar then you have bit line and bit line bar so if you ones stored if you have 1 coming here or 0 1 here coming here and 0 will be coming here and so and hence so forth and these are the memory code cell.

So this is cell 1, cell 2, cell 3, cell 4, cell 5 and so and hence so forth now if you want to access for example cell 3 then you keep your word line WL1 high and therefore this will be all high I will show it you in bit later slide and then you make it bit and bit bar line work for you then this

cell will be selected right. So this bit and bit bar line this need to be selected and then you can select this 3 similarly if you want to select 5 number cell then you select WL1 and then select these 2 bit lines.

Similarly if I want to select this line right then select WL4 and select this bit line and bit line bar right. So this is basically a IIF columns and rows by selecting appropriate columns and rows you will be able to select a particular cell for writing or reading purposes.

(Refer Slide Time: 07:01)



Now if you look so let us not compare that same architecture with what you are seeing in front of you. Now let us suppose if you have diode here word line and bit line you are selected suppose your work line is 1 right then this diode will be on and this 1 will be written on to a bit line here. So that is a most simplest method of writing but this basically a wrong which means that once a one has been written you cannot rewrite it one and one word again once it is done it is done.

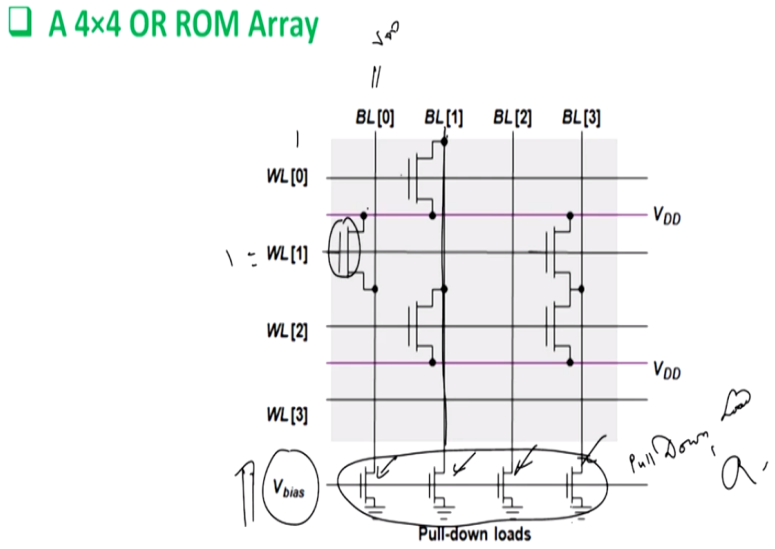
So once you have diode connection between these WL and bit line which you shown here which generally do not have anything else to. So and I have no connection between these 2 as 0 will be stored right similarly if you have word line. So similarly this is basically a diode based you also have a MOS base so this is basically diode based ROM you can also have a MOS based ROM where you have MOSFET which is get here right and your word line becomes high this MOSFET will become on and when it becomes on this VDD connected to this point.

And therefore bit line goes to VDD and it goes high and it goes to higher value available to you. If there is no MOSFET available to you then we say 0 to be the obvious thing another way of looking at is that if you have work line high then bit line will be I think 0 because this is grounded here and if you do not write anything here you will have 1 in this case 1 it will be there.

So there are three major compositions of read only memory one is basically a diode ROM right another basically a MOS ROM 1 and the third one is basically MOS ROM 2 in the second case which is MOS ROM 1 depending on the ON state or OFF state of the MOSFET you output bit line will be either charge to VDD or to ground generally to VDD for storing 1 and there is no MOSFET in between it will be storing 0 there where as the reverse will be 2 in case of MOSFET 2.

Where in if you have word line high bit line will be storing in ground 0 and other case it will be storing on 1. So this is the basic memory code which you see as far as ROM is concerned let us look at a 4 / 4 or R based ROM array right pretty simple and easy to understand.

(Refer Slide Time: 09:20)



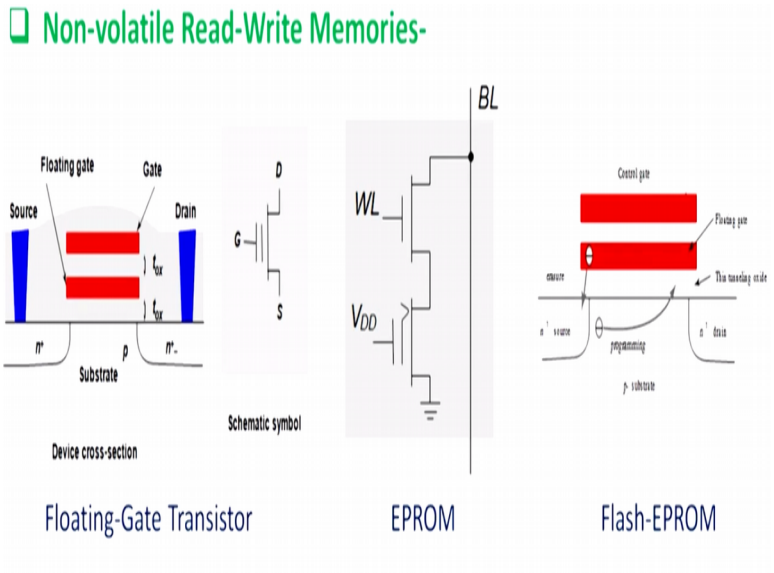
Say for example your WL1 is high right it is 1 so this switches on to the data this switches on this 1 and therefore this VDD is written here. So this is equals to VDD right this is equals to VDD similarly if your WL0 is 1 and this also 1 and then your VDD is again written here and we

reduce again written here and therefore there will be VDD available at this particular point and so on and hence so forth right.

So these are basically these are which you see here these are known as pull down loads these are known as pull down loads why they are pull down loads? Because they are responsible for pulling down the voltages to ground once the data as actually been written there right and this basically a ROM array 4 cross 4 because 4 cross 4 because this is there are 4 word line 0, 1, 2, 3 and therefore bit line 0, 1, 2, 3 and depending on that this thing the V bars which you see here will be responsible for.

So therefore there will be these transistors here we will act as a resistive transistors so they will act as a resistances and therefore we you were B bar goes high right these get switched on and therefore the act down as a pull down resistances for all practical purposes this is 4 / 4 and therefore bit line will be equals to 1 bit line will be again equals to 1 bit line 2 is nothing is there so it will be 0 and bit line 3 will have again 2 available to you right and this is how a 4 / 4 R based ROM will be there.

(Refer Slide Time: 10:53)

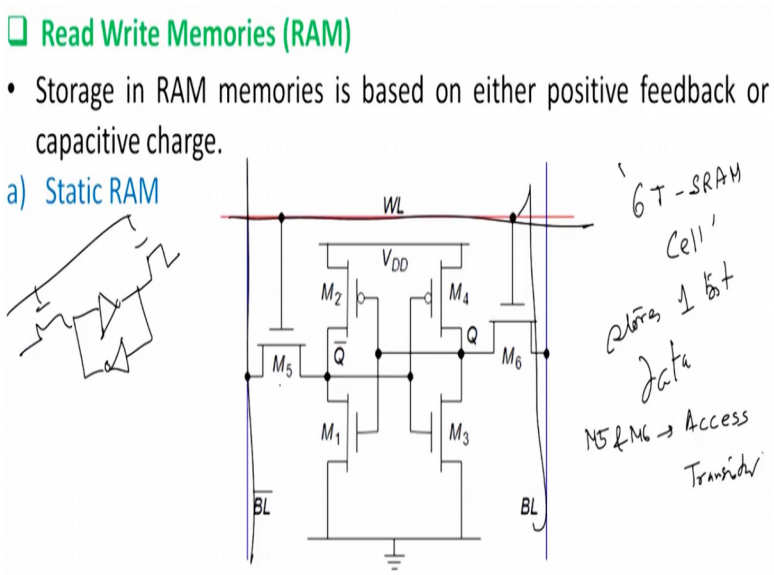


I am going to details of how it works out because that is quite a subject by itself but if you look at non-volatile right memory these are the few examples here the left one is basically a floating gate transistor then you have a EPROM which is erasable programmable read only memory and then you have a Flash EPROM available to you right. So these are three types of Non-volatile

read write memory which is available to you right and depending on the distance between the floating gate and the original gate we can actually lower or higher the and data and therefore in this you can actually store a data.

In this case once you actually remove the power supply all the charges are removed and therefore you do not have any data being stored during the VDD = 0 case.

(Refer Slide Time: 11:40)



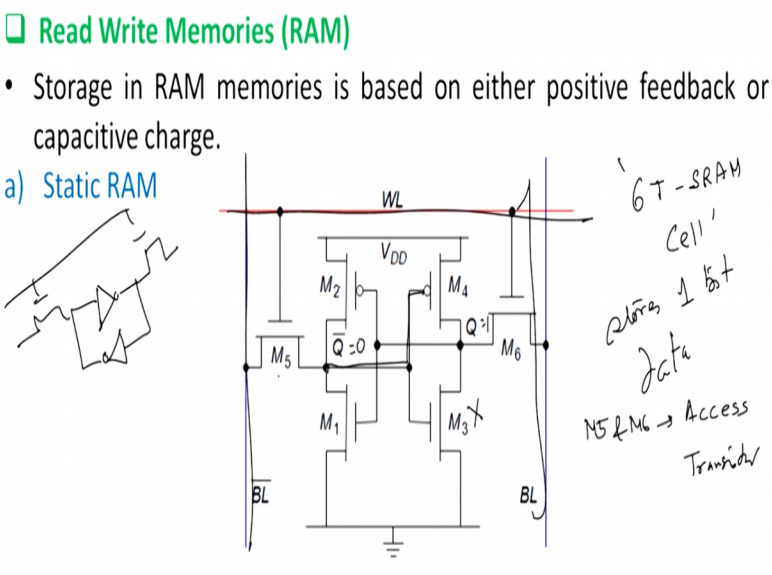
Let me come to an statics this is basically known as the sixth transistor SRAM cell right and this 6 turns to SRAM cell actually stores 1 bit of data. So this is known as 60 why 60 because 1 so M1, M2, M3 and M4, M5 and 6 if you look at it carefully this M1, M2, M3 and M4 they form a back to back connected inverter and then you have an access transistor which is available here.

So this is the word line which you see in front of you this is the word line and this is the bit which you see this is the bit line. So M5 and M6 M5 and M6 are referred to as access transistors right they are referred to as access transistors and M1, M2, M3 and M4 are part of the CMOS inverter pair which you see. So if you make your work line high both M5 and M6 which is on right.

And therefore your internal cell which is the internal value of Q and Q bar can be written on bit and bit line bar respectively. So by simply choosing which word line you want to make high you can make all the cells high and then depending upon which one of the cells you want to choose

you can take typical values of bit or bit line bar. This is also known as static random access memory or SRAM and it is known as 6TSRAM cell right.

(Refer Slide Time: 13:25)



Before we move forward I need to explain to you just two things and these are so I have got a inverter right and I have got an inverter here. So this by itself we can understand is a memory so this by itself is a memory what does it do is something like this I have a memory here and then I have cell here. So whatever suppose I am storing one here this will be 0 this 0 will go again here as a 1 and there this is a close loop available to you and therefore when you word line goes low suppose it is 0 and this is also 0 then this cell is totally insulated from the external peripherals and therefore you can store the information for the last period of time.

Please understand this is basically a very high impedance load right this typically this node and this node and Z in is very almost infinity Z in is infinitely high it is very high bit as node as it is discussed with you earlier because CMOS input side will be the gate side and as a result the chart storage is very proper and there is no leakage available to you in storage starts for longer duration of time.

But the problem here is when you make work line high right and you make word line high here when this cell becomes transparent and that time you can read or write the cell we define two important terms here one is known as the pull up ratio another cell ratio right pull up ratio is the

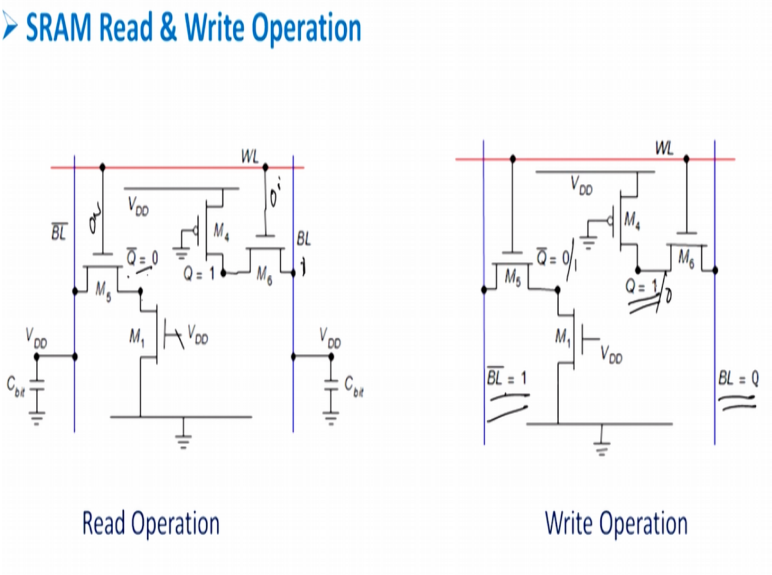
ratio of W / L of access to W / L of pull up which is PMOS right and cell ratio is W / L of access to W / L of pull down which is NMOS.

So we define cell ratio as this one and we define PR = this much so we have a we have a cell ratio and we have a pull up ratio these two are quiet important understanding the understanding the whole issue of working principles. Let me come to the exam read and write operation so read operation is very simple and straight forward. So you see you make your word line high this becomes on right and this also becomes on say you have storing a bit one here and therefore bit 0 then you please understand M4 will be on because you are grounding M4 and as a result when this becomes on this VDD to this much and this one appears as the bit line bar here.

Since this is on this is VDD and this is also on therefore you have pull down available to ground and therefore Q bar = 0 so Q bar appears to bit line bar. This is the very simple straight forward way of looking at it when you want to read a cell therefore it is quiet easy that out of M1, M2, M3 and M4. M1 and M4 are switched on right and M4 gate terminal is always grounded in this case it is grounded because your Q = 1 and Q bar = 0 since the Q bar = 0 M4 gate will always be grounded.

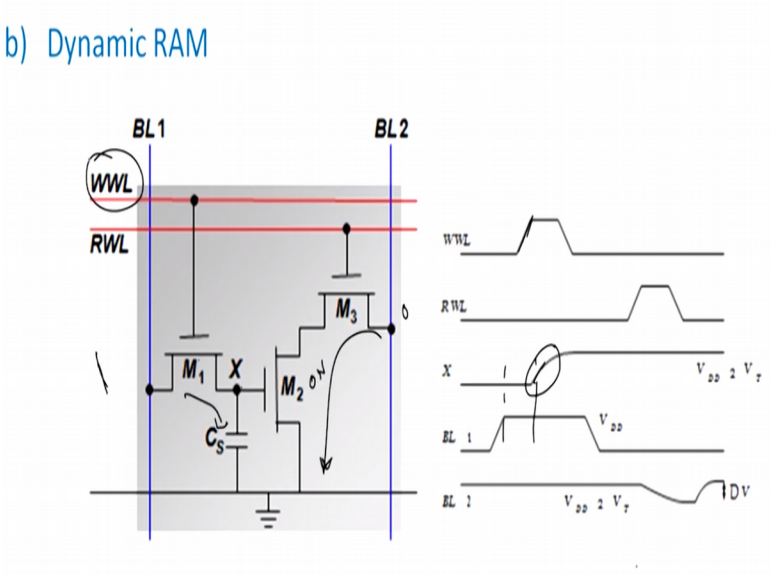
I think I made myself clear see M4 is connected to Q bar getting my point so when this Q bar = 0 M4 is on and when your Q bar = 0 M3 is cut off because it is an NMOS so it will not switched similarly M1 will be switched on because why it will be switched on because Q = 1 let us suppose but the M1 will be switched on and so it will pull this voltage to ground and therefore Q bar will goes to 0.

(Refer Slide Time: 16:55)



So exactly the same concept applies here that this will go to 0 and this Q bar will go to 0 and Q will be equals to 1 this is for the read operation. If I want to go for write operation then simple it is simple as well simple when you want to write it say Q = 1 you want to write the new you want to make it line = 0 and bit line bar = 1 so what it will do is this overwrite this 0 to 1 and this will overwrite 1 to 0 and then you go for right operation. So write operation is when this comes it to VDD and read operation is when we connected to directly to the ground here.

(Refer Slide Time: 17:28)



There is another technique static random access memories static why because you do not have to refresh it once you have stored the data it remains there for a long duration of time. However there is another type of RAM which is known as dynamic RAM or DRAM's this cell is basically

1T1C cell 1 transistor 1 capacitor. So M1 is basically the transistor which you see in front of you and CS is basically the capacitor across which you are storing the data.

So whenever your bit line 1 is high and your WWL is high so whenever your WWL is high this is high the X value starts to rise right actually will rising why because this is high M1 is on bit line is suppose is one so when bit line is 1 X will start to rise right and so you see this rising of X but it rises out of sometime this is because of the delay which you see in front of you. Then X is rising right it stores data at CS and it goes to high value approximately = VDD.

Now as it goes to high value M2 switches on have an M2 switches on this gives a low conducting path for a 3 and therefore initially which is 1 and which is 0 this is 0 then this will actually geos to 0 grounded when you make a RWL high. So when you make WWL your high write high then you actually you are able to store 1 from bit line 1 to X mode by this thing and when you want to make the bit line 2 high sorry when you want to make a bit line too low then you actually ground it across M3 and M2.

So this is how your actually see basically dynamic RAM but the problem here is since the capacitance itself which is voltage across the capacitance which is actually be leaking because the capacitor leaks you need to refresh this time and again in the dynamic RAM right. So let me recapitulate what we have done in this two modules of memory we have not gone for peripherals but just giving an very brief idea about the memory coal.

(Refer Slide Time: 19:42)

Recapitulation

- The type of memory unit that us preferable for a given application is a function of required memory size, the time it takes to access the stored data, the access pattern, the application and the system requirements.
- RWM belong to the class of volatile memory while ROM belongs to the category of non-volatile memories.
- EPROM and E2PROM are the memories which offer both read and write functionality even that are of non-volatile in nature.
- RAM can be classified as Static RAM and Dynamic RAM based on the storage concept.

We have also look into the various applications and the system requirement we have also looked into EPROM and E2PROM and then we looked into the various elements of static random access memory and dynamic Random access memory we define what is cell ratio and pull up ratio and we have also seen how can you read and write in the cell in a memory right. So this comes to an end for the memory design we will take up in the next time thank you very much.