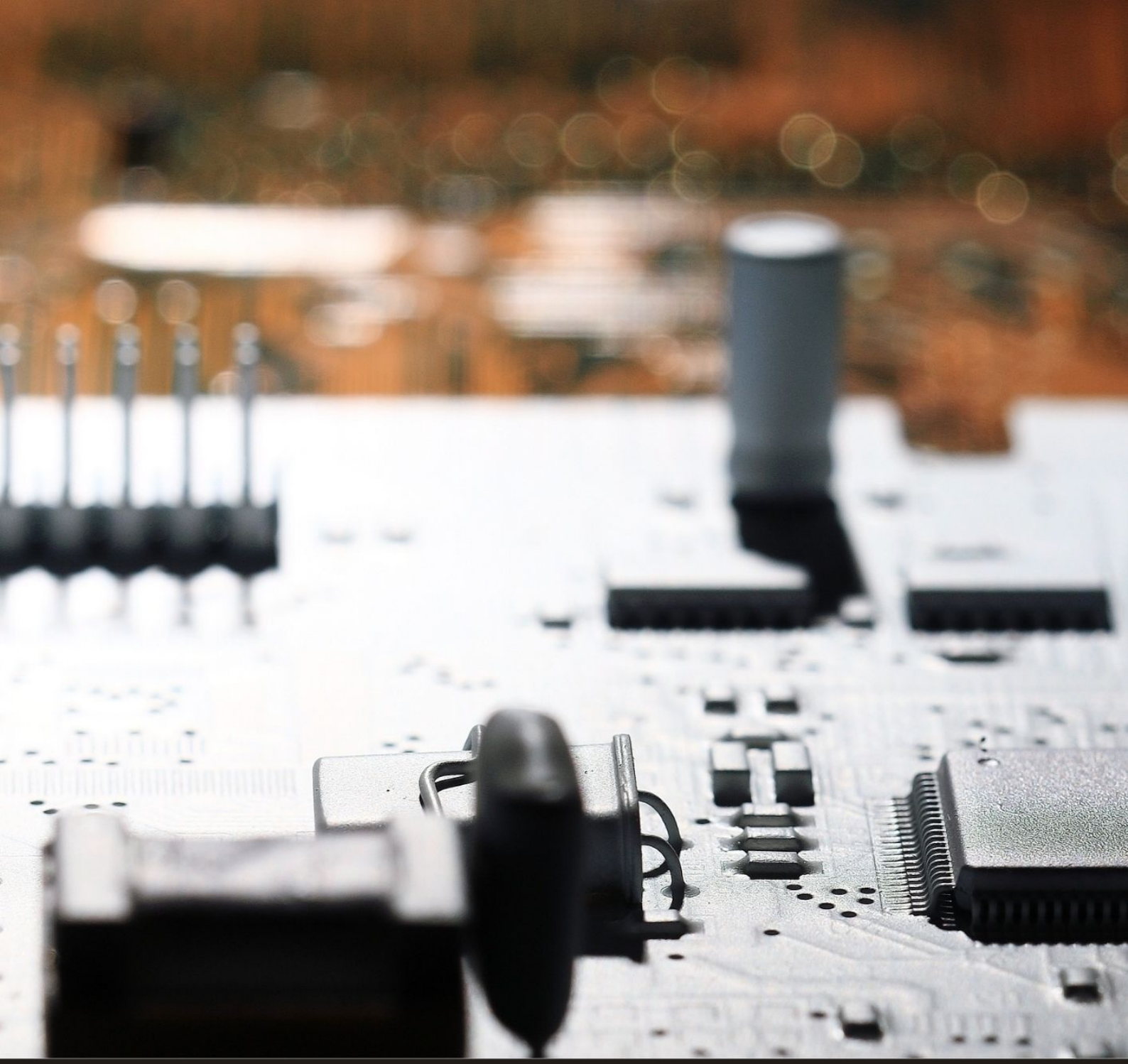


# MICROELECTRONICS: DEVICES TO CIRCUITS

**Prof.Sudeb Dasgupta**
**Electrical Engineering**
**IIT Roorkee**

# INDEX

## *Week 7*



## *Week 8*



## *Week 9*



## *Week 10*



## *Week 11*

## *Week 12*

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−01**
**Bipolar Junction Transistor: Physical Structure and Modes of Operation**

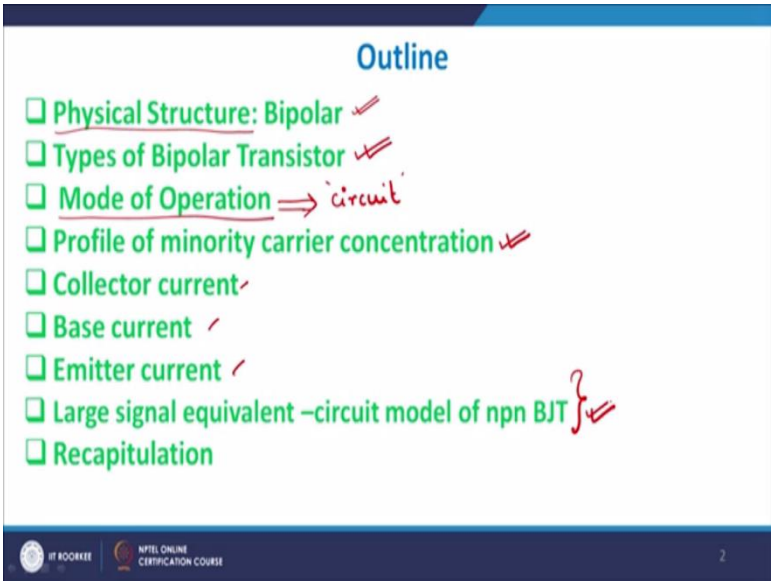
Hello everybody, welcome to the NPTEL online certification course on Microelectronics: Devices to Circuits. This is basically a 30 hour lecture which is being sponsored by NPTEL and delivered by me. I am Professor S. Dasgupta at IIT Roorkee in the department of Electronics and Communication Engineering.

The prime motivation for this 30 hour lecture which was supposed to be delivered to students and stack holders who are interested in this area is to give you a full knowledge right from devices till circuit level, realization of those devices which means that the first few lectures or the first few weeks will actually carry forward the concept of devices the various devices which are currently used in the area of microelectronics. Starting with bipolar technology and then a short description of MOSFET and CMOS technology after that we will be actually entering into the usage of these devices for both analog and digital applications.

So this course primarily therefore deals with the application area of these devices in analog and digital domain. To understand and to appreciate its usage we need to understand the physics of the devices in a proper fashion and therefore, the first 5 weeks or 4 weeks of this lecture plan or this lecture will primarily deal with the devices and then we will be the understanding therefore the basic device characteristics we will use those device characteristics for realizing certain circuits for all applications.

Now, so the first thing which the first lecture which starts today we will be dealing with bipolar technology which is bipolar junction transistor technology and in this lecture today we will be discussing its physical structure and modes of operation. So today's lecture is termed as bipolar junction transistor or BJT and its physical structure and modes of operation. So we will understand how a bipolar technology or a bipolar transistor looks like and what are the various modes of operation of this device.

(Refer Slide Time: 02:53)



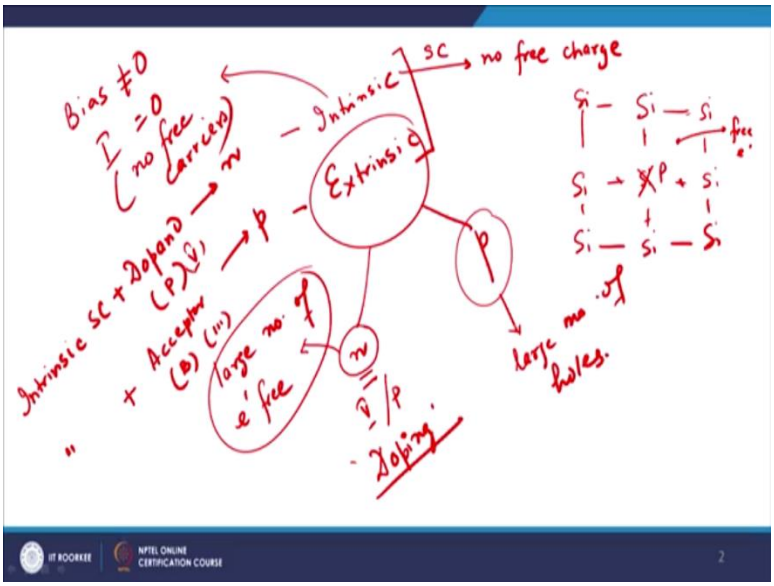

So what we intend to do in the next part of this work is, we will so the outline of my talk is something like this that we will be actually looking into a bipolar technology right. What is a bipolar technology and how is it different from unipolar technologies? We will look into the physical structure of the bipolar technology which means that how does a bipolar transistor looks like physically, various types of bipolar technology or transistors we will be having a look and we will be seeing on what basis the differentiation is done between two bipolar transistors.

We will be then focusing into its mode of operation because this is quite important in the terms of circuit implementation. So when we do a circuit implementation we actually require to know its mode of operation. So we should know under what bias input bias will my transistor work at what stage or at what mode right. So this topic will give you an idea about the working principle with respect to the biases of voltages and currents given to this transistors.

We will be looking into the minority current carrier profiling in this case because that will help us to find out the total current which is there in a bipolar technology. We will be looking into emitter current, base current, and the collector current and its interdependency based on certain rules applied to a bipolar technology we will be also looking at the circuit model for npn transistor which means that if I have an npn BJT or a bipolar and I want to use it in a circuit analysis what modifications or what equivalent circuit models can we have for this BJT so that I can just plug and play in a circuit. So that we will be doing it at the last part of our talk.

So basically large signal equivalent we will be looking into circuit model large signal equivalent which means that once we know what are the currents and voltages of the bipolar transistor is there can I therefore convert that into known devices and known factors so that it is very useful when we want to translate this device into a circuit that we will be doing when we do a last signal equivalent circuit model of BJT. So you should be very careful about the whole flow of this bipolar technology.

(Refer Slide Time: 05:31)




So before we even start with bipolar technology let me give you an idea about what is a basically a brief idea about PN junction because that is where we are starting with. If you remember interestingly in a very basic thing which was being remembering we have two types of transistors or two types of semiconductor devices we have a extrinsic semiconductor and we have got intrinsic semiconductors. I will just go about briefly to give an idea about because this is a starting were with. What is an intrinsic semiconductor? Intrinsic semiconductor is a one in which we do not have any impurity and there are no free charge carriers there are no free charge carriers.

So therefore, in an intrinsic semiconductor for example a silicon if we apply a bias since there are no free charge there will be no current flow and since there are no charge there are no free carrier there will be no current flow. So therefore, in an intrinsic semiconductor for bias equals to even nonzero bias your current will be still equals to 0 as there are no free carriers right.

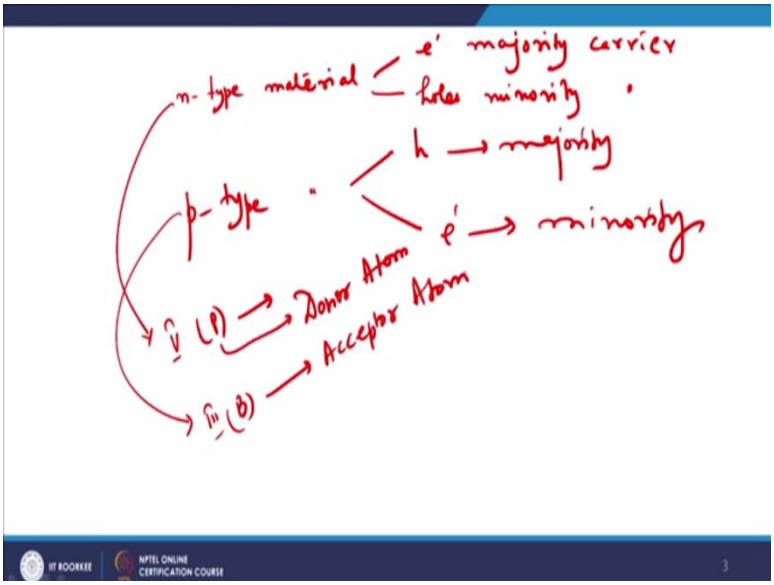
Now, let us come to extrinsic, extrinsic as you must be knowing from your previous knowledge we have two types of semiconductor, one is known as n type semiconductor and we have a p type semiconductor, what is a n type semiconductor? An n type semiconductor is a one in which we have doped silicon with group five element for example phosphorus and therefore an n type material or n type semiconductor we will have large number of large number of electrons as free carriers.

So what is a extrinsic n type material n type material will a large number of free electrons as free charge carriers from where they are coming they are coming from the donor atoms which are primarily phosphorus in nature, so if I have a silicon here right and I have got silicon here and I have got silicon here right all of must be knowing there are covalently bonded with respect to each other right and so on and so forth, so I have a silicon here I have a silicon here right.

Now if this phosphorus or this silicon is replaced by a phosphorus which is group 5 then all the four electrons of phosphorus will group with this, with this, with this and with this, but its extra electron will be actually free so this will be acting as a free electron right so this what is known as we must be knowing also is basically known as doping, right. So what we do? We take an intrinsic semiconductor we add a dopant to it dopant basically in this case may be phosphorus which is group 5 element and we convert them into n type semiconductor, right.

We similarly take an intrinsic semiconductor add may be accepted type group 3, group 3 element which is for example boron so this group 5 and this is group 3 and we convert this into p type semiconductor available to us so which means that n type semiconductor will have large number of free electrons whereas a p type semiconductor will have large number of large number of holes as a free charge carriers right so this will be large number of holes.

(Refer Slide Time: 08:56)




So and therefore let us comeback to an original issue to statement therefore that I have a n type material, I have a p type material, n type will have electrons as very large quantity and therefore electrons will be referred to as. So what we will say is in an n type material, I will be discussing or I will be knowing that electrons will be referred to as a majority carriers a majority carrier right and holes we will see why, holes will be referred to as minority carriers.

Similarly, in a p type material, in a p type material right holes will be referred to as majority current carriers right where as electrons will be considered as minority carriers. Which means that, and we will not discuss this in detail because that will be a part of another course, but what happens actually is in an n type material which is basically an electron having majority carriers you will still have few amount of holes still available to you right.

So maybe there are 100 electrons and may be 3 or 4 holes right may be that will be the ratio. Similarly, for a p type material you might have 100 holes and may be 5 or 6 electrons, but they are still there right there are in small percentage but still there and they make a difference when we are discussing certain issues related to bipolar technology.

So, with this understanding, the basic knowledge which you possibly, already, maybe aware of or if you are then I will refer that you please consult one of the standard books in devices, which will give you an idea about all this things, but for our purposes as far as this course is concerned,

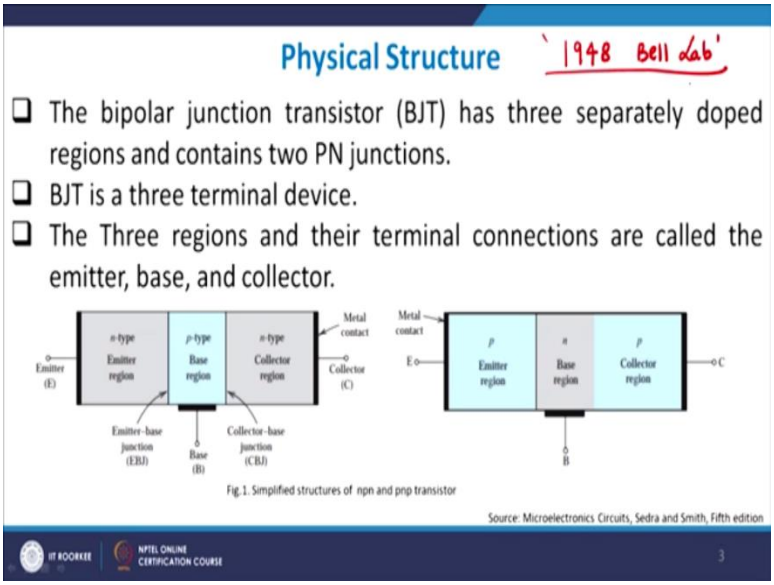
this much knowledge will be at this stage sufficient for moving forward, that I have two types of materials, I have a intrinsic silicon, I have extrinsic silicon.

An extrinsic silicon can be of two types, n type and p type. n type is a silicon which have got electron as the majority current carrier and hole as the minority current carriers, and p type is semiconductor which is holes as the majority carrier and electrons as the minority current carrier, right. Now please understand, I have still not applied any bias or I have not applied any temperature variation, nothing. So I have just doped it which means my dopant species have been just added, right.

Now in n type material, when we add to this group 5 element for example of a phosphorus, this phosphorus is known as a donor atom, why? Because it is donating one extra electron to the silicon, whereas when we make a p type material, when we add our group 3, for example boron, it is known as an acceptor atom, why? Because it is accepting one electron from the system and a hole is left there and therefore we define it to be an acceptor atom.

So a donor atom will donate electron and make this species intrinsic silicon to an extrinsic silicon n type and the acceptor atom when it is being doped on to intrinsic silicon makes it a p type material, right. So we are clear about these two basic philosophy or basic concept as far as this course is concerned, right.

(Refer Slide Time: 12:06)



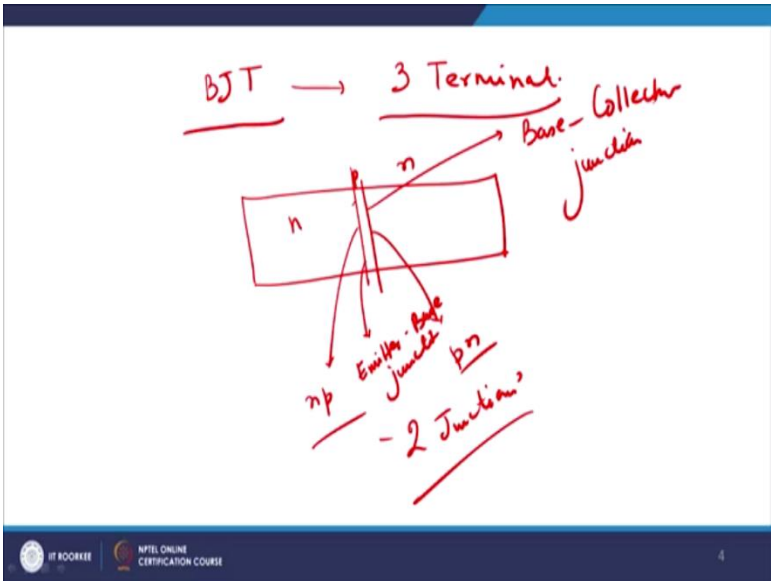
With this understanding, let me come to the first understanding of a bipolar technology. Well, BJT as the name suggest that is the bipolar junction transistor, first came into picture in 1948, right and it was done in Bell Labs at that point of time, so 1948 in Bell Labs in U.S, it was developed and it was very well known to us and the first BJT or bipolar technology which was physically done was by Bardeen and Shockley, they were the first one to propose this concept and 1948 was the time and it was physically available to all of us.

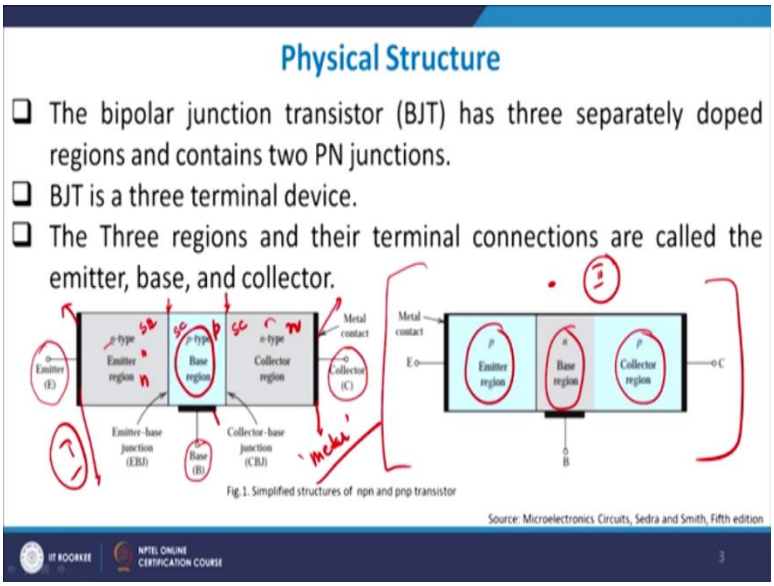
Let us look at the physical structure of the bipolar technology and then see how things move along. Now if you look at the chronological aspects of bipolar technology, in 1950s and 60s, a bipolar technology was the main stay for all your digital and analog applications. So, we had this bipolar technology act as an inverter and also act as an amplifier, this bipolar technology was acting as a logic for all logic. For example TTL, right Transistor Transistor logic, so all your transistor logic used bipolar technology or a BJT for doing all your digital logic design.

Similarly, for the purpose of amplification, voltage amplification, current amplification we use this bipolar technology and in 1950s and 60s and so on and so for. So it was a very standard technology at that point of time, later on MOSFET and CMOS came into picture, this was relegated as a second level device, but still in many applications BJT find still is used, one of the application is BiCMOS technology right. For making rad hard design, for example radiation hardens design we use a BiCMOS technology. So it is known as a BiCMOS technology where we use a bipolar technology, bipolar plus CMOS technology together right. They have been used quite often in all these domain analysis.

I will go in to the physical structure of a bipolar and then we will see why is it known as a bipolar junction transistor right and we will see what is the basic physical reasoning or understand the basic physical working of this bipolar technology. Now, BJT is primarily, if you look very carefully has got all these three regions. So, it is basically a three terminal device.

(Refer Slide Time: 14:45)




So, a bipolar technology is basically a bipolar, a BJT bipolar transistor is basically a three terminal device which means that there are three terminals associated with bipolar technology, right. What are these three terminals? These three terminals are referred to as, first terminal is known as emitter terminal, this one is known as emitter terminal, we have a second terminal which is known as base terminal right. And then we have a third terminal which is known as collector terminal fine. So, this is basically a semiconductor, this is semiconductor and this is also a semiconductor. But difference is that you have an n type semiconductor for majority current carrier as electrons, the emitter is basically an n type semiconductor and the collector is

also an n type semiconductor and sandwich between the two is basically your p type base region, right.

So therefore this is referred to as an npn transistor, right. So I have got npn which means that, so there are three regions emitter, base and collector. Emitter and collector will have the same doping, same type of doping, not necessarily the level of doping, but same type of doping and base will be sandwich between emitter and collector, right, so these are the three separate region are there. Now what we can see is that, therefore if you look very carefully you will have two basically np junction available here, right. What are these two pn junctions? The first one is this junction and the second is this junction, right.

So if you go back to the original diagram of an npn transistor which we are aware of, so there is an npn, this is basically your np junction and this your pn junction. So we have got two junctions, which is there, there are two junctions, right and they play a critical role in determining the properties of this bipolar technology, right. So there are two junctions here, this is known as emitter base junction, right and this is known as base collector junction.

So, I have got two pn junctions, connected back to back, one is known as emitter base junction and other is known as base collector junction. And these two junctions are just auditory pn junction, as we know for sure which is available to us.

Now, the structure is something like this, its dual structure if you see on this side, on the right side of the power point presentation, this side-this side is basically the n type semiconductor which is Emitter, has been replaced by p type, the base which was initially p type has been replaced by n type, the collector which was initially n type has been replaced by a p type, right.

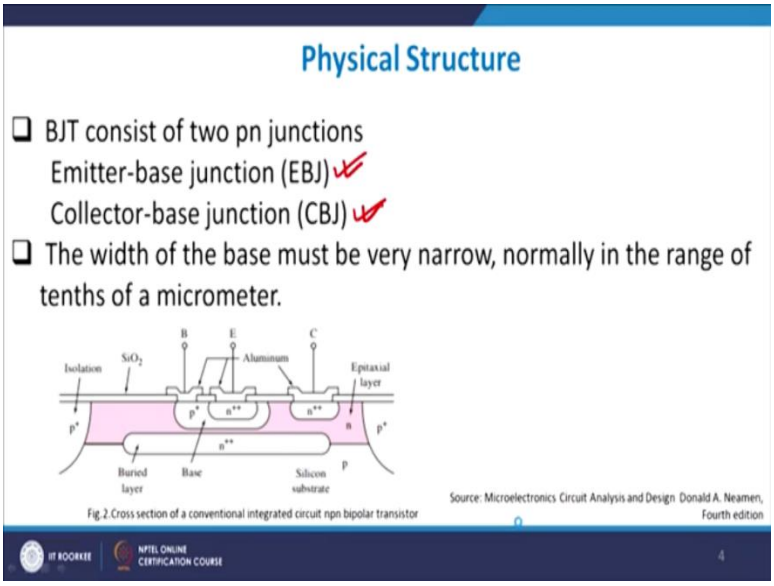
So the p one is basically blue one and the n type is basically shown by brown or light grey. So what we have done is that the second structure you see here is just the dual of the first structure. Which means that structurally exactly the same and doping also is just identically complimentary with respect to each other. I will discuss that later on as we move along.

So, I can have two types of transistor therefore one is known as npn, right The first one which you will see on the left hand side of the screen is npn, right And the second one is pnp . So depending upon where you have sandwiched your base layer, right and what is the doping of the

base layer, you can actually make a point whether it is a npn or a pnp transistor. But, we have two types of transistors that is for sure, right.

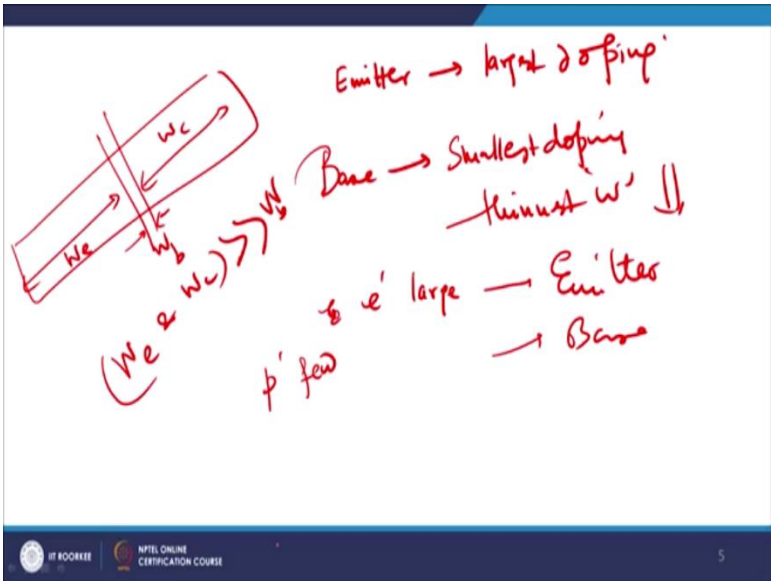
Please understand also one more important point shown in this diagram, this black material, which you see this, this and this, these are all metal contacts, so these are metal, right. So, I have metal contact, you a need metal contact to form a Schottky barrier between the semiconductor and the metal, right. So there will be Schottky contact here at emitter region here, so there will be Schottky contact here and there will be Schottky contact at this point as well, right. So there will be therefore there are two junctions, three terminals and there are two types of material, p type and n type which you see in front of you.

(Refer Slide Time: 19:23)




As I discussed with you in the previous slide, we have got two types of junction available one is known as emitter base junction, another is known as collector base junction. Now, what we see is that this is quite interesting that I will discuss it later also that it has been seen or it has been predicted or it has been proposed that the emitter region should have the largest doping concentration among the three regions.

(Refer Slide Time: 19:50)




So if you look at the emitter, emitter should have the largest doping and we will discuss it why is it like that, but this is what is there, right, Base will have the smallest doping or the lightest doping, right and will be thinnest so its width is also very low. Width means the disk thickness. So if I have got this I will make this as w has to be very low. So this is your $w_e$, this is your $w_b$ and this is your $w_c$.

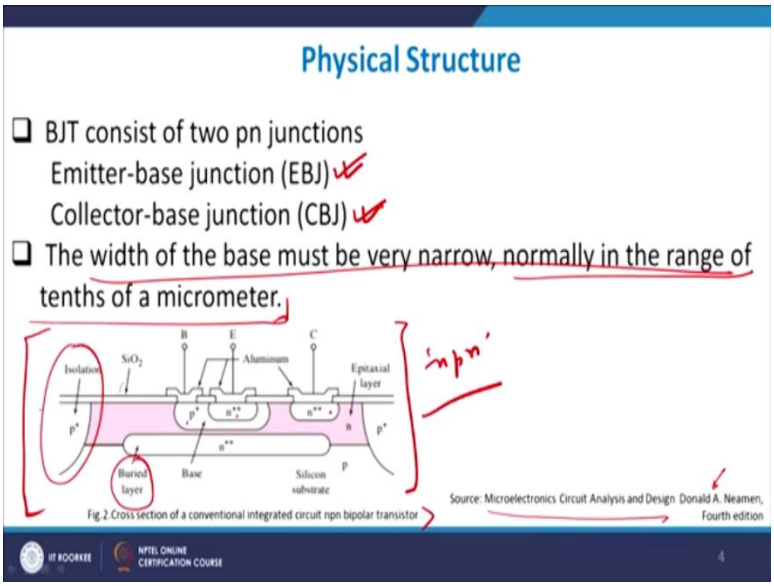
So what you try to do is that $w_e$ and $w_c$ should be much greater that $w_b$, this is a general trend which we follow. Which means that, the base should be approximately 1000 of your emitter length, right and it should be very very small and very very lowly doped as compare to the emitter. So if you doped emitter very heavily you will have large number of free electrons available in the emitter side.

Whereas if you doped the base region very lightly as I discussed with you, the number of free holes available in an npn transistor in the base side will be very very low, right. So please get the situation from this point of view, I have still not apply the bias. Please understand once again, but what we are doing here is that my emitter is heavily doped so electrons are very large in number, right and this is for emitter, right and base is basically p type first of all but not only p type it is very lowly doped so there are very few holes available there, right.

Moreover, since the width is very small the overall number of holes in the base region is also very small, fine. So that is the reason why we do this, we will be clearing it off as we move along

in this course, but structurally why therefore emitter should be the largest in area, largest in dimension with highest doping, right and base should be very thin or narrow with very very light doping about 100 doping of the emitter side and the collector should be moderately doped, right and these things you have to be very cautious when we are selecting the bipolar technology as far as this study is concerned.

(Refer Slide Time: 22:07)




As I discussed with you therefore that the width of the base should be very narrow, right normally in the range of tenths of a micrometre and that is quite important and therefore the width of the base, right determines to what extent you will get the output current right. Just to give you a brief inside into this as we move along, but primarily the width of the base is a deciding factor as far as the total output current is available to you.

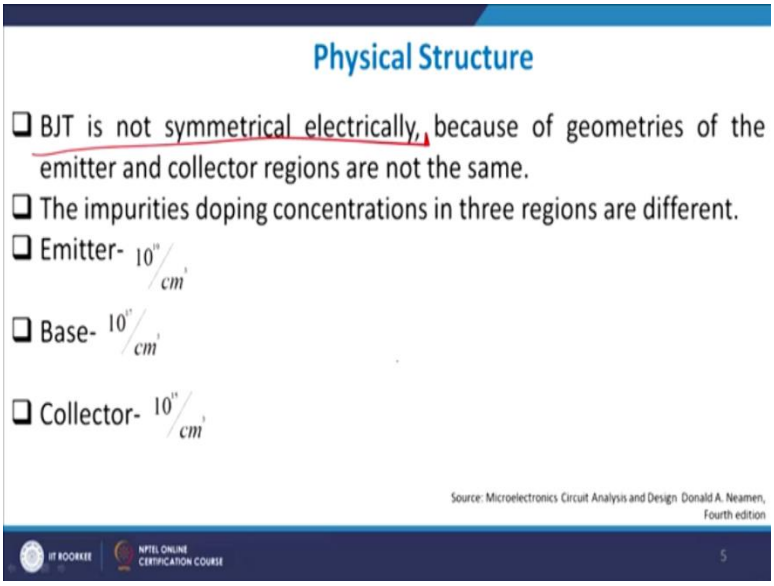
What you see in front of you which is this one is basically a cross section of a conventional npn bipolar transistor, right This discussion can be found in a book which is Microelectronic circuit analysis and design by Donald Neaman in fourth edition but may be later editions are also available and which they are actually giving you the cross section.

So, if you look very carefully this is basically your emitter, this is your base and this happens to be a collector, right and this is a collector which you see in front of you. This is basically an npn transistor, which is there. This buried layers, why do you put this buried layers in silicon dioxide and so on and so forth is that we do not want any leakage current, that means if there are no

buried layer or if there are no silicon dioxide which is relatively high dielectric material, I would expect to see some amount of current flowing through this areas as well and therefore the offstate leakage will be large which will effectively means the power dissipation will be large, right.

So, we put this buried material here so on and so forth. We do also isolation here, a $p^+$ isolation to distinguish between two BJTs. So if I draw or if I design two bipolar technologies this isolation will help me to electrically neutralize or electrically isolate these two bipolar transistors, right and this is what we generally grow over here and sort up of twin tub process we use it in this case for the physical structure.

(Refer Slide Time: 24:19)

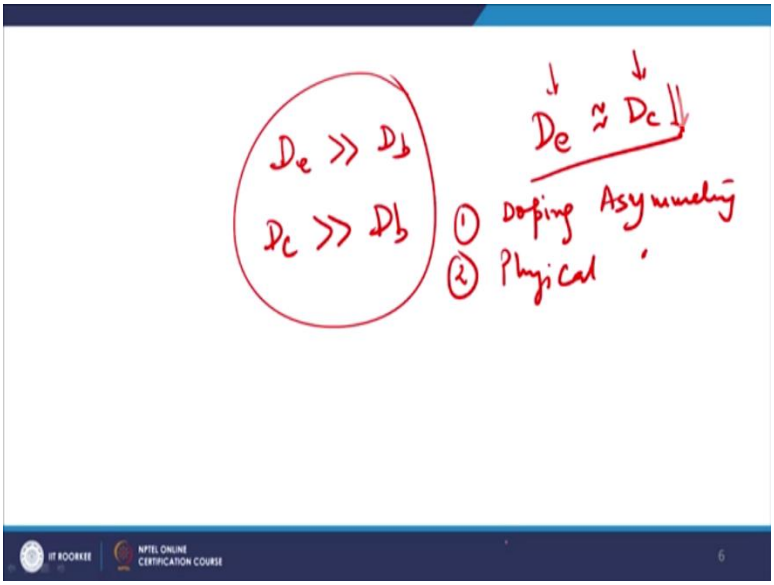



So what we have learned till now is two things, the first thing is that as you can see the first point here that BJT is not symmetrical electrically, why? Because though it looks symmetrical it is not, because your emitter side is having a larger concentration of electron as compared to electron on the collector side, for an npn transistor. Not only that, the emitter and collector regions are also not equal in size, emitter is quite large in dimensions and collector is relatively small in dimension as compared to emitter

As I discussed with you earlier that the doping concentration should be very very low in the base side, it is relatively lower as compared to emitter, one hundred of the thousands of the doping concentration and the collector should be at a much lower concentration of dopant species. So,

the dopant species generally if you keep, even if you keep the emitter equal to collector there is no problem.

(Refer Slide Time: 25:15)



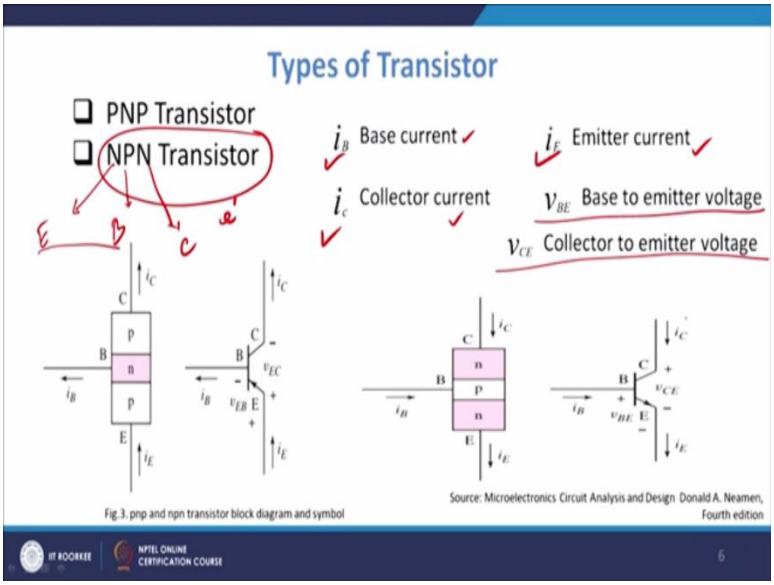

So, in terms of dopant species if you want to look, then typically your doping of your emitter should be much larger as compared to your doping of your base, right and doping of your collector should also be much larger as compared to doping of your base, right. $D_e$ can almost be equal to $D_c$ right. So doping of emitter and Collector can be approximately equal, even if you keep the collector doping slightly lower it does not matter to a larger extent, but this is pretty important, that you need to keep the doping concentration of emitter and collector much larger as compared to base.

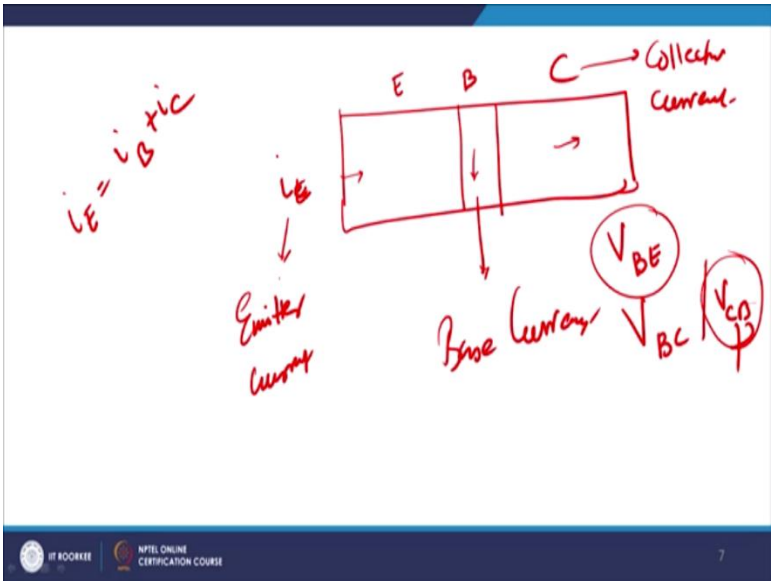
So this is the first thing, the second thing is as I discussed with you, first thing is doping, right. So this is a doping asymmetry, there is also a physical asymmetry. What is a physical asymmetry? That the length of the dimension of emitter is much larger as compared to that of the collector side, right.

So therefore what I wanted to stress, finally is that emitter has got the largest number of electrons available to it, the collector has got relatively less amount of electrons and base has got a very very small quantity of hole within the small region of operation because of very low doping and very thin width of the doped, right. With these three basic understanding, let me therefore

explain to you the various type of transistors which are available here and we will discuss that later on as we move along.

(Refer Slide Time: 26:40)




Now we have two types of transistor as I have discussed with you, one is known as the npn transistor, npn as the real name suggest, this is emitter, right this is base, and this is collector, so I have a pn junction emitter base junction. I have got also a base collector pn junction, right and now you see carefully in an npn transistor the majority current carriers are actually electrons, right.

So an npn transistor electrons of the majority current carriers, why? Because the emitter side is basically an n type semiconductor, the n type semiconductor majority current carriers is obviously an electron and therefore electrons are the most largest amount of current carrier here and therefore they result in a larger amount of charge carries and therefore electrons is the majority current carrier here, right.

Similarly, if you just find the dual of npn that is pnp , right. So pnp emitter is having a p type semiconductor, base is basically an n type semiconductor and the collector is a p type semiconductor exactly the same doping profile as we did in the previous case, that emitter of the pnp should be heavily dope so we have a large number of holes are available on the emitter side right.

You have very very small amount of electron are available on the base side and you have negligibly small amount of holes relatively large amount of holes, but on the p side, collector side right. So I have an npn and pnp transistor which is which is there with me on in all practical purposes right and now since its a three terminal device what we do is, since it's a three terminal device which I have just shown here right.

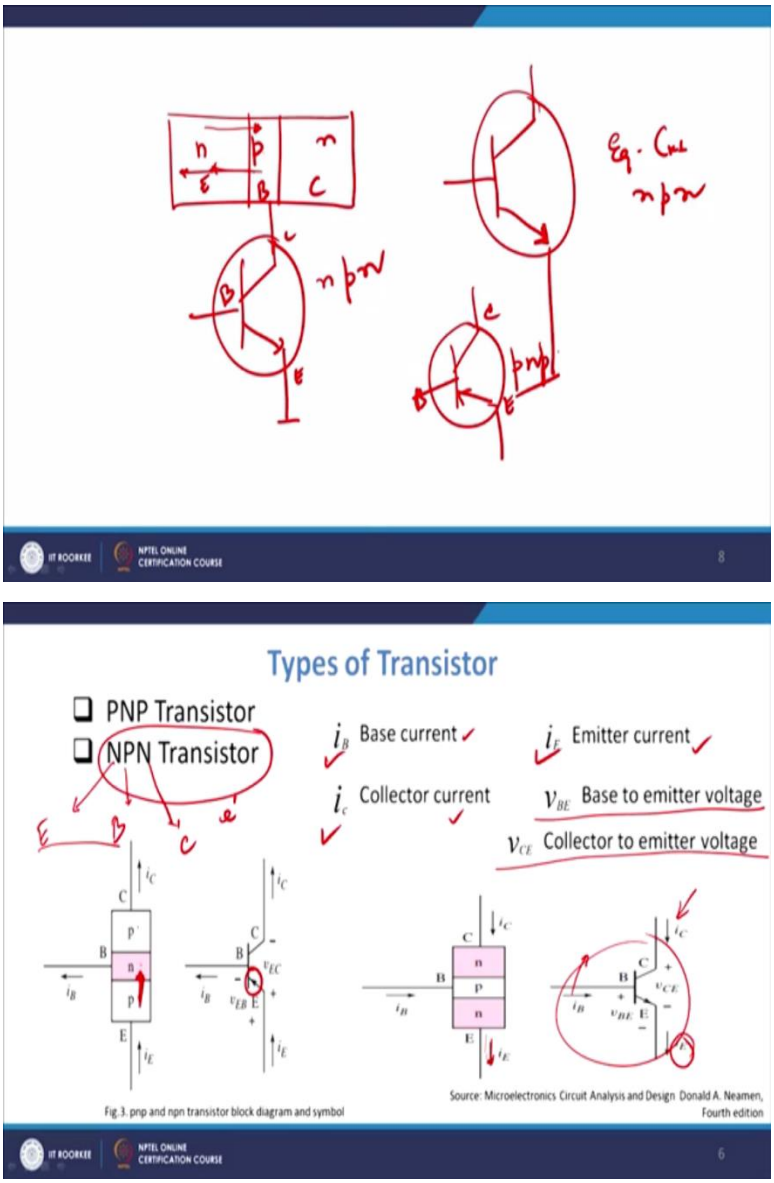
May be this one so I have an emitter, base, and a collector right .We have three currents and two voltage associated with this three terminal device, what are the three currents? One is known as the emitter current $i_e$ right known as emitter current, we have got the another current known as the base current and we have third current which is known as the collector current.

So there are three currents, emitter current, base current and collector current and there are two voltages, voltage between base and emitter, and voltage between base and collector and all we can $V_{CB}$ as well right. So we have two voltages $V_{CB}$ and $V_{BE}$, $V_{BE}$ is also determined as base emitter voltage this one is known as base emitter voltage and $V_{CB}$ is collector to base voltage and there two three currents emitter current, right, base current and we have got collector current right and of course need not to say by Kirchhoff law this will always hold good that emitter current will be always equal to base plus collector current right this will always hold good for practical purposes right.

So that is what I was say that we have got $i_b$, $i_c$, $i_e$ right $i_b$ is the base current, $i_c$ is the collector current and $i_e$ is the emitter current $V_{BE}$ is base to emitter voltage and $V_{CE}$ is collector to emitter

voltage, what is the collector emitter voltage? We will see that later on, but $V_{BE}$ and $V_{CB}$ collector to base right. Now you see if you look quite interestingly here and this what the type of transistors and how we show its functionality, we will just show it to you.

(Refer Slide Time: 30: 26)




Say for example you have got a npn transistor, npn, so this is your emitter, this is your base, this is your collector. How we show it in circuit or how we represent it in circuit in this manner. Base is always shown by a straight line, right? And then we write down like this as the two edges right. If you want to show the lower one as the emitter so, you define from electron from holes point of view right.

So, if you look very carefully here for example in this case for example in this it will be easier to handle it. Since, electrons are moving from base to emitter, sorry, from emitter to base right? So the conventional current will flow from base to emitter right. Is it okay? So you see what I am trying to say is that you have an npn right.

Electrons will be moving in this direction right? Which is the majority current carrier. Which means that the holes will be moving in base to emitter side. So, we always measure the hole current or the conventional current and therefore the arrow head is always away from the base. So, you see because the holes are moving in this direction therefore I will always get an arrow head like this. So this the equivalent diagram or equivalent circuit diagram of an npn transistor.

As you can see here for an npn transistor this how it looks like. So in npn transistor I will have emitter current right. Flowing outwards just flowing outwards base current and collector current. So $i_e$ will be equals to $i_b + i_c$ right. So there is no violation of Kirchhoff's law as such and we are able to obtain this variations in a much better manner right and we are able to achieve this profiling in much better.

Now if you want to convert this into an npn to pnp then you see pnp and on the left hand side of your this you have. Then you see what we will do the emitter will blow holes into the electrons right into base side.

So, the direction of holes or the direction of movement of holes is from emitter to base and that is the reason the direction of arrow is from emitter to base so base is a straight line collector and emitter are shown by straight lines, curved lines the direction of the arrow in the emitter is actually the direction of the holes flowing in the emitter and you should be able to handle the separation.

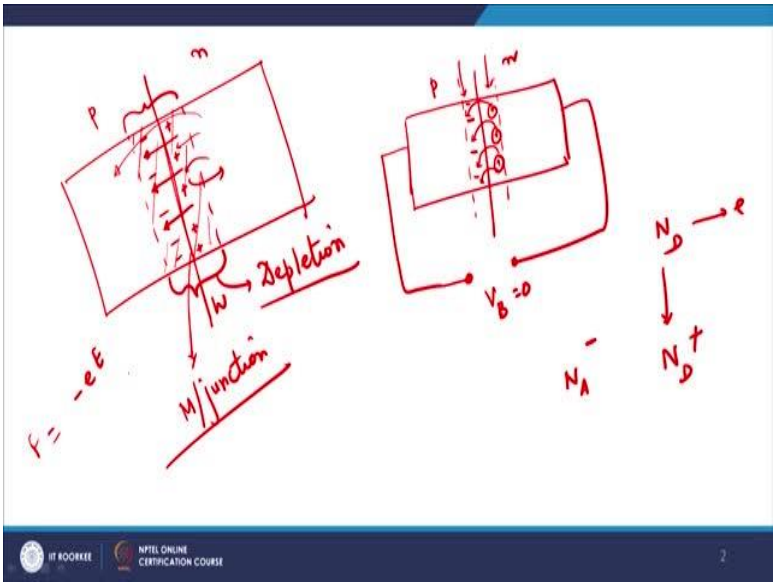
So therefore if I want to make an npn transistor like npn transistor then this is the this is my NPN right. This is my emitter, base and collector. If I want to make a pnp transistor right. Then, this is my emitter, base and collector so this is a pnp, so there is a two transistor pnp and npn this gives you an idea about the basic structure of an npn transistor and pnp transistors right. We will take care of the next section in the next class. Thank you very much!

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−02**
**Bipolar Junction Transistor: Modes of Operation-I**

Welcome back to the NPTEL online certification course on Microelectronics: Devices to Circuits. In our previous lecture we had understood the basic structure and the doping profile of a BJT. We have also understood two types of BJT which is available to us. One is known as NPN and other one is known as PNP. In the first case which is NPN, electron is the majority current carrier, in the second case holes are the majority current carriers.

What we will learn in this module or in this lecture is the working principles of BJT, right. And we will see why is it therefore termed as a bipolar technology. Let us may be it will be easy if we start with an NPN transistor and then we can take up the case of a PNP transistor. So before we move forward and we discuss about NPN and PNP transistors let us just recapitulate what we know about PN junction or PN junction diode or simple PN junction right.

(Refer Time Slide: 1:26)




So if you have P type material and N type material here and you do not apply any bias, let us suppose the bias is 0, so I am just keeping it open circuit over here. So $V_B$ is equals to 0 then what will happen is that free electronics from here will be jumping to this place they will be diffusing, right, because there will be diffusion at this place. Why there will be diffusion?

Because you have a large free electron concentration here and you have got very low free electron concentration here therefore electrons will be jumping from N side to P side as a result what will happen is, on the N side you will be left with what positive ions which are nothing but your donor species. And what happens on the P side, these electrons will recombine with holes and you will have negative charges, effective negative charges will be available to you, right.

So, what is the big picture therefore, that even if you do not apply any bias because of a diffusion between P type and N type material you will have a region near its junction, this is known as the known as metallurgical junction, right? Please understand this is known as a metallurgical junction, this junction, PN junction is basically known as a metallurgical junction.

If you look at this junction so initially therefore what will happen is that electrons will be moving and this side you will have fixed charges which are positive in nature and you will have fixed charges which are negative in nature. Why, because electrons have left from the from the negative side as they went they left behind their donor atoms there and therefore you will have this positive charge. So, what happens is that? There will be electric field which will be directed from N side to P side inside this region, right.

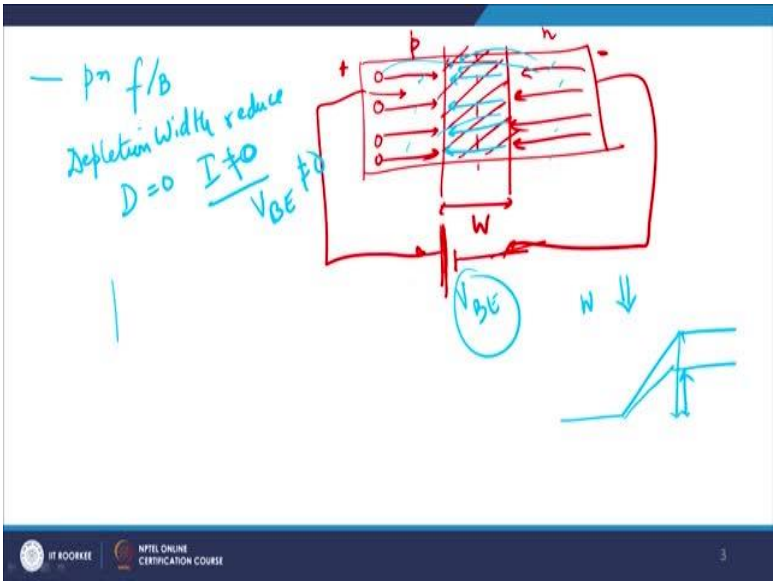
First of all, therefore, this region which you see is referred to as the depletion region, right. So this is the depletion region, so why it is known as depletion region? Because this is the region which is basically depleted of any free charge carriers, so please understand there are charges but these are not free charges, they are fixed charges, right? Why fixed charges? Because electrons have left their position leaving their donor atoms ionised donor atoms.

So, if I have a donor atom $N_D$, right, it gives one electron and it converts itself to $N_D^+$ right, accepts that atom right, gives one holes there and becomes negative atom available to us. Now, So the idea therefore is that since this is and therefore this region which you see in front of you this region is primarily devoid of any free charge carriers and therefore it is known as the depletion region which means that under zero, even under zero bias condition the PN junction is basically having a depletion width which is W let us suppose right, this will always be there you like it or not at even at zero bias there will be the junction will be deployed or devoid of any free charge carriers.

Now why after sometimes this movement stops so the free electron from N type to P type. So what is happening is, therefore, for electron to move from N type material to P type material now it has to cross this depletion region, agreed? But as just as it enters this depletion region it sees an electric field which is directed this side which means that, because if you remember it is -E into E right, negative why (mine) because electron is in negative charge therefore, the force on the electron is -E into E and therefore electron trying to enter this side will be pushed back on this side right.

So, beyond a particular time when the electric field becomes so large that any extra electron coming from N type and trying to enter into the depletion width to go towards the P side will be pushed by the electric field within the depletion region and force it to move outside the depletion region and throw it towards the N side. So what is the overall picture therefore what we get?

(Refer Time Slide: 5:44)




That overall picture therefore is under zero bias condition your there will be a fixed width of a PN junction and that fixed width is basically my depletion width W right and therefore, this will be devoid of any, this part will be devoid of any free charge carriers. Right if this is the PN junction diode and we have not applied bias here. Now, let us see what will happen if we forward bias this PN junction as we very well know forward bias basically means the P type is connected to a positive terminal and N type is connected to negative terminal fine. Let us look from an N type material.

Now what will happen is that or from the P type material either way you can have a look from the P type material. So if you have a positive polarity battery connected to the P type material here, the holes which are majority current carriers here will see an electric field in this direction because this is positive and this is negative, so there will be electric fields. So these holes will be pushed towards this region, right? It will be pushed.

Similarly electrons will be pushed in this region right? They will be pushed near the depletion edge as it enters. So please understand, within the depletion region the electric field is in this direction so I will make the highlighter pen slightly change the colour so that you are able to appreciate what I am trying to say. The electric field within the depletion region is in this direction whereas outside it is this one right, this is the direction.

Therefore, once the hole enters a depletion region, let us suppose this is the hole, it enters the depletion region by virtue of this electric field. It will be again thrown, it will be at least, these electric fields which are marked in blue will try to throw the hole towards the P side and any electron entering from N side will be thrown towards N side. But now, please understand, you have an external bias which is say $V_{BE}$ let us suppose base emitter or whatever $V_{BE}$ or $V_{XE}$ whatever, which will help the holes and electrons to cross this barrier right.

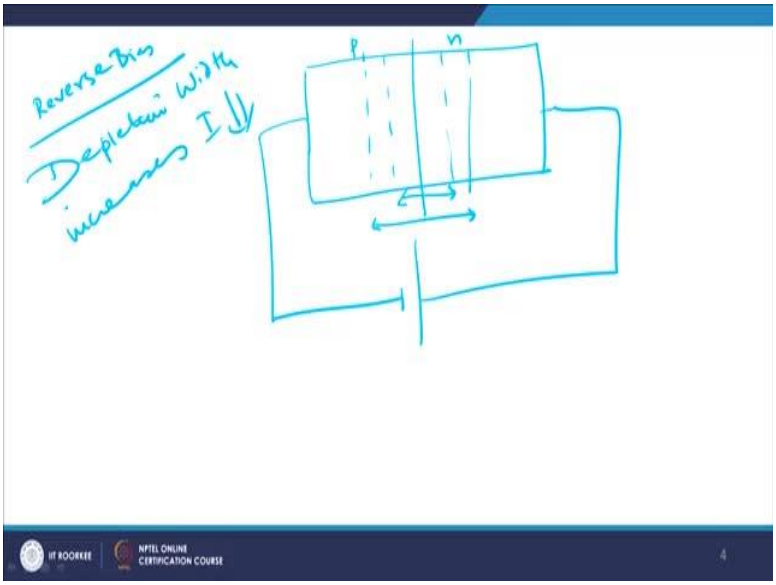
So, if so if I am able to see that so extra electrons will be crossing this barrier or extra holes will be crossing this barrier and once it reaches on this side right it tries to remove some of the electrons available here and therefore what happens similarly electrons moving from this side to this side as it starts to move the depletion width actually therefore starts to reduce, right. So as you make the PN junction more and more forward bias your depletion width starts to become lower and lower this can be also understood in this manner.

So if you look at the depletion width it primarily means the holes has to acquire an extra energy to cross this barrier so for an hole it is basically a barrier right. So if you are able to provide a bias which is $V_{BE}$ what happens to the holes is that it lowers the effective barrier drastically, so this was the initial barrier length, now this becomes the initial barrier.

Now it is easier for holes to cross the barrier right so we should look it in this manner so we come to one major conclusion out of it is that when your PN junction is forward biased right, the depletion width will reduced, a time might come when the depletion width will go to 0 and you will have some current flowing because of the external bias for $V_{BE}$ not equal to 0.

So, there will be some current not equals to 0 for $V_{BE}$ not equal to 0. This is the condition when you do forward this is known as forward bias.

(Refer Time Slide: 9:40)



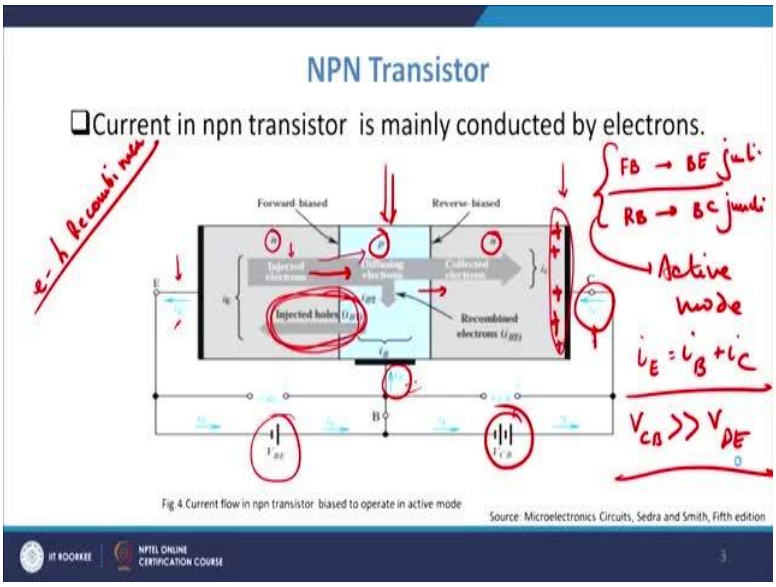

Now, let us look at the fact what will happen if you go to reverse biased, which means that you just convert this whole thing and may be the next I can help in this slide. Let me say that I have this and I have got this PN, I had a depletion region already formed side now I had connected P type to a negative and N side to a positive side. So I just reversed the polarity of the battery right.

Now, I leave an exercise to you to and you can easily appreciate what will happen now is that this depletion width which was initially this do this will now become larger, will become larger. The new value of the depletion width will be this one, which means that under reverse bias condition, reverse bias condition the depletion width, depletion width increases, width increases, right, and as a result the current will be further lowered or there will be no current flow.

So, please understand therefore so from a basic study of PN junction diode theory or PN junction theory, when you forward bias a junction more is the current, when you reverse bias a junction less is the current. Why, because the forward bias will give you a lower depletion width and reversed bias will give you a higher depletion width. With this knowledge you have gained as far as a simple PN junction is there let me come back to NPN transistor explain to you how a NPN transistor works.

So, look at the fact, let us remember our previous lectures basic principles that the emitter will have a very heavily doped, your base will be relatively very less doped, it will be very thin also and your collector will be relatively less doped but it will be same doping as that of the emitter.

(Refer Time Slide: 11:29)




So this is an NPN if you look very carefully this is N, N type semiconductor this is again an N type semiconductor and this is P type semiconductor. What we have done is we have forward biased the base emitter junction. And what we have also done is we have reverse biased the base collector junction right this is just for information at this stage, this is also known as the active mode of operation. We will discuss that in subsequent slides, but just to give you a functionality of this NPN transistor, we forward biased the base emitter and reverse biased the base collector and see how the current flows.

If you look very carefully therefore, when you forward bias the base emitter junction right, you reduce the depletion width on the base emitter junction region as we discussed just few minutes back, right and you reduce it. When you reduce it you force large number of electrons from the emitter side and they will be injected on to the base side. So this is the injected electron which is flowing right, because you have forward biased N and P junction here, right. As it enters the base region this is what is the critical aspect of a bipolar technology bipolar junction transistor works.

As it enters the depletion region here right, please understand, the base region is typically very having very small number of holes because it is slowly doped and it is very thin. So

what will happen is majority of the electrons will actually cross the base very fast because it has very high velocity, but some of the electrons will actually combine with some of the holes present here. So there will be an electron hole recombination right, so the holes will be lost because electrons will eat into the holes and as a result the holes will be lost, right.

In order to make the same number of holes as it was previously you see there will be a current that is known as $I_B$ which will be flowing into the base, so please understand this blue coloured current source which we are showing to you is basically the conventional current these are not the electronic current right. So, therefore, if you look at the base here which is this one if you look at the base here then in the base the direction of the current is basically into the base and the reason why it is into the base is that the holes are now.

So, let us suppose 10 electrons recombine with 10 holes, so 10 holes were lost in order to therefore make again the 10 holes available to me 10 holes from the base side which is being fed by $V_{BE}$, what $V_{BE}$ this $V_{BE}$ will enter into the base side and therefore, the blue coloured arrow is going inside the base right and this is the base current, very-very small current because why base current is very-very small, because the number of holes are very-very small, right. So the probability of recombination is also very-very small and therefore I would not expect large base current to be available to me.

Now, what has happened therefore is, those electrons which did not recombine they went and towards the base collector junction. But please understand base collector was reverse biased right so when the electrons come here what does it sees? It sees a negative potential here, sorry it sees a positive potential here, I am sorry, because of this $V_{CB}$ and therefore, they will rush through the depletion region, right and go towards the collector side. If there would not have been any positive charge here or positive potential here electrons could not have cross the base collector barrier because base collector barrier please understand is large as compared to base emitter because it is reversed biased.

So, if there would not have been any positive potential near the collector end I would not expected to see the electrons to cross the barrier because you are not providing that much amount of energy to electrons to cross the barrier and that is reason what will happen the electrons would not flow. But since you have provided a $V_{CB}$, a very large potential as compared to $V_{BE}$ and that too reverse biasing the base collector, the electrons will see a very large potential on the collector side and as a result what will happen it will force itself through the depletion region and reach the collector side right.

And as a result there will be a collector current as you can see it is again pointing inwards why pointing inwards? You very well understand why is it pointing inwards because the electrons moving outwards therefore hole current is moving inwards. So, I have $I_B$ inwards $I_C$ inwards and $I_E$ also moving outwards right, $I_E$ outwards because electron currents is moving inwards right, so this is what is basic structure of base current looks like or the basic idea looks like.

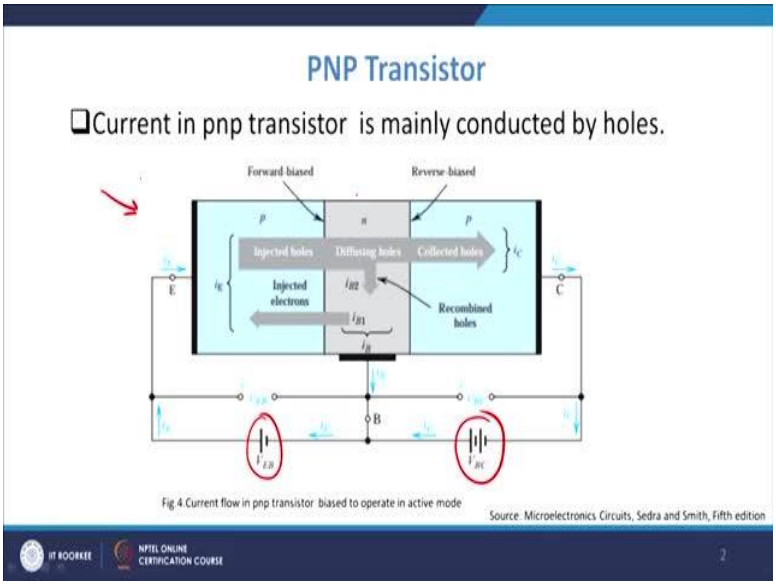
Now, if you look at another issue here which is quite an interesting issue is that when you apply a $V_{BE}$ base emitter you are basically forward biasing this base emitter junction right. Base emitter junction you are forward biasing it. Now, when you are forward biasing it, you are also reducing width of the emitter base junction layer as we have already discussed this point. So, just as it is easier for the electrons to move towards the base side it is also easy for the holes to move from base towards the N side.

So, you see you will also have one current contribution because of the injected hole from the base side right, so you have injected electrons this is primarily diffusion and then you have drift and these are injected holes from base towards the collector side. However, as $I_{B\,1}$, $I_{B\,1}$ will be very-very small because as I discussed with you small base current is there and therefore $I_{B\,1}$ will be relatively very small which is there with you, right. So, I have three currents therefore, $I_E$ which is here right, $I_B$ and $I_C$ and therefore, by Kirchhoff's current law $I_E$ will be equal to $I_B + I_C$, right.

And obviously $V_{CB}$ should be much larger than $V_{BE}$, fine. And this is the basic fundamental working which means that if 100 electrons started from the emitter there maybe 90 electrons reached towards the collector 10 were recombined in the base and therefore by changing the biasing of the base emitter junction I can change the total amount of current flowing through the collector side, right.

I can also therefore change the collector current by changing the base current because $I_E = I_B + I_C$ so if I want to make $I_C$ larger I reduce my $I_B$ because $I_E$ is fixed. So changing one current or getting the final current high or low by changing one of the parameters of the other current is the prime example of a BJT. That is the basic example of a BJT which I just wanted to give you an idea.

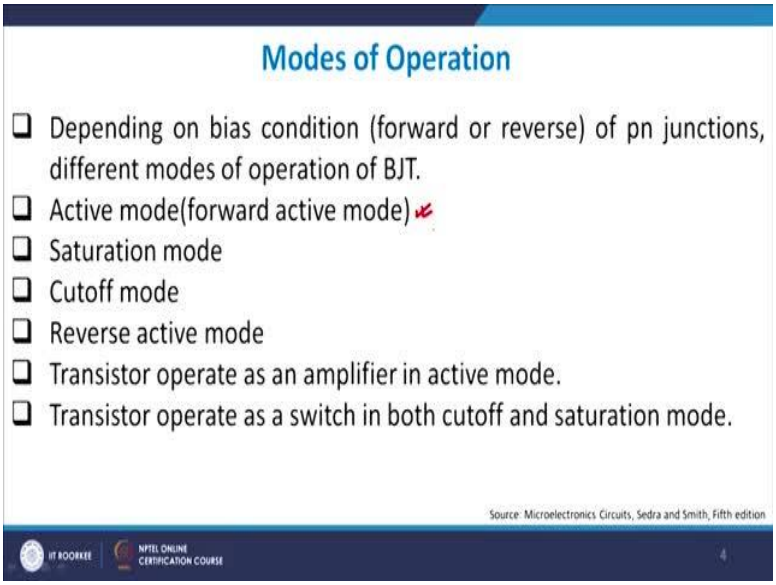
(Refer Time Slide: 18:54)




The same concept the same thing exactly the same thing can be applied to holes also. Only in this case the biasing of the battery will change and this is therefore forward biasing it because it is P and N type and this is $V_{BC}$ again reverse biasing it. Understanding part is exactly the same as the previous case so I am not going into detail of this you please work it out yourself from the book or you can yourself do it, it is pretty easy and pretty simple to understand. Concept remains that same the base emitter will be still forward biased, depletion width will be small. Base collector will be reversed biased it will be still large and so on and so forth.

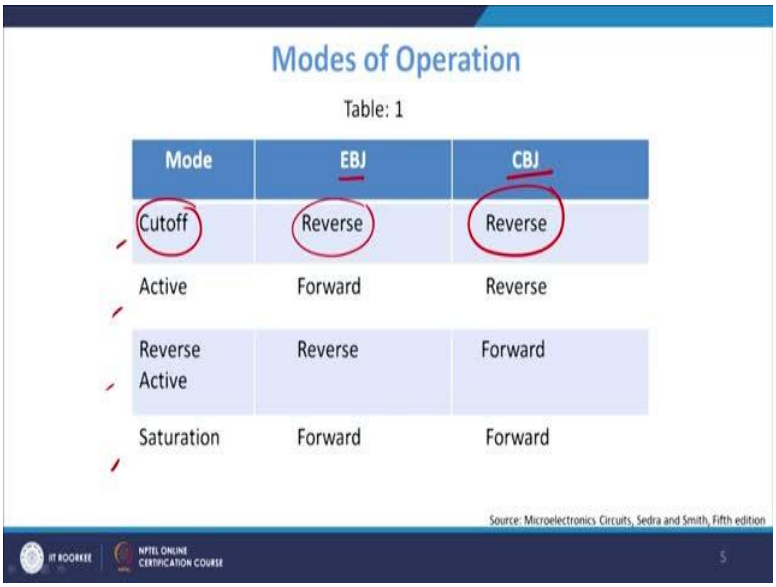
In this case since the base is made up of N type material the recombination on the base side will be relatively small and therefore, the base current will be less relatively very small right. And in the second case as you increase the value of $V_{BC}$ or $V_{EB}$ you will be actually able to control the current flowing through the collector side. So, these are the few important aspects which one should be able to handle as far as this is concerned.

(Refer Slide Time 20:02)




Now, therefore the various modes of operation are given as these are the various modes of operation right, depending upon the bias condition very simple straight forward. So there are two biases right, so I can have $2^2$, so I can have four sort of modes right.

(Refer Time Slide: 20:16)

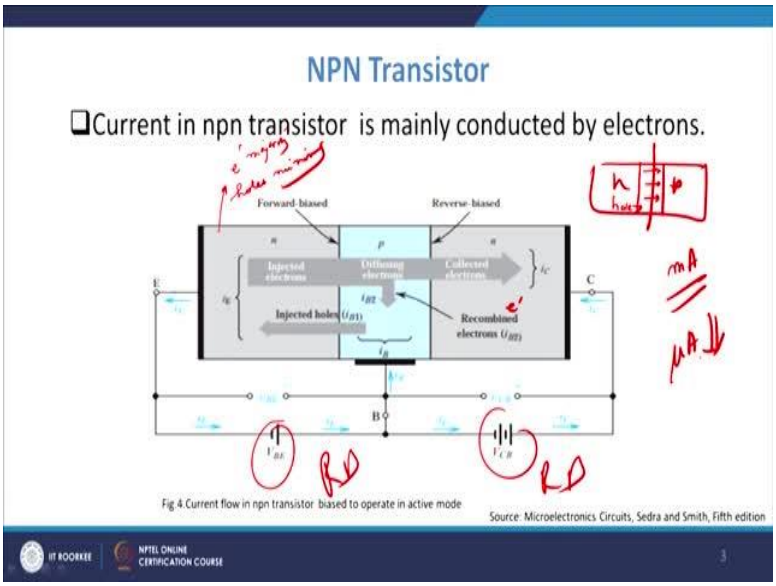



Is it okay? Why, because there are two bias conditions, right and I can therefore, have and each one of them can be reversed, right because why two biases two battery sources and each battery source can be either forward biased or reversed biased. So, I can have therefore four conditions available to me forward-forward, forward-reverse, reverse-reverse and reverse-forward. So, each one of them will give rise to each methodology. We will come to it one by one and we will see how it works out. I think you can appreciate these points.

The first thing is if both emitter base junction this is EBJ and CBJ is the collector base junction. So if EBJ is reversed and CBJ is reversed then, of course, you can understand it will be cut-off why. Why it will be cut-off? The reason being is both a reverse biased then emitter base junction depletion width is also very large, right and collector based junction width is also very large.

Both the widths are very large, so there is no question or no probability of any electron coming from the emitter side and entering into the base side, right and therefore, there will be very-very small current or almost zero current in the region, then we define that to be as cut-off, right. Why cut-off, because you are applying a bias but you are not getting any output current, then we define that to be as cut-off mode of operation. I just missed one small point in the previous discussion which I will just take care of it by this understanding.

(Refer Time Slide: 21:52)




That is something like this, see in an N type semiconductor which you see in front of you this one N type. Electrons are the majority carriers right, this we have already seen majority carriers but we forgot one small thing that the holes are also there but these are minority, very small in numbers. So on emitter side you will have electrons as well as holes, electrons will be very large in numbers and therefore, we were discussing the current flow in a BJT from electron point of view but there will be also holes available here, please keep this in mind.

But, for example, in this case even if there are few holes inside this N type semiconductor and you forward bias this emitter base junction which is available. Please understand for a minority current carrier it will not be a hill but a slope. Why it will be a slope? Because this is

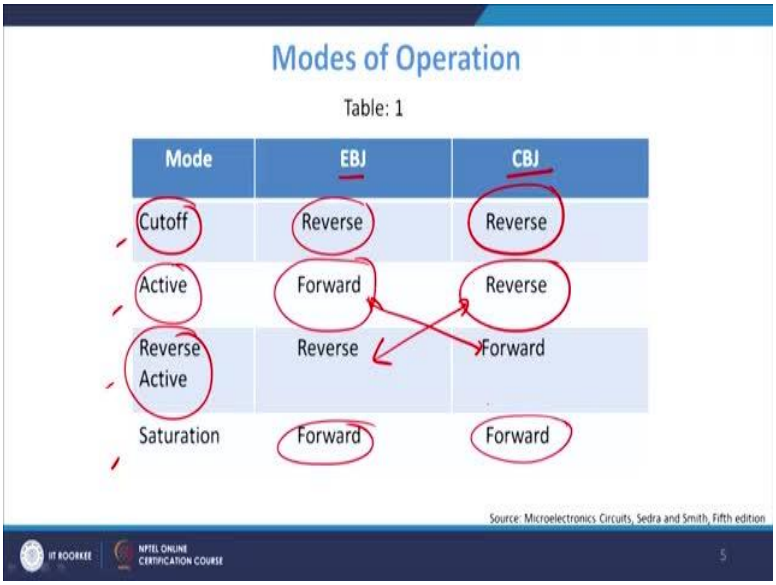
emitter base junction so if it is N, basically it is a PN junction. It is an N and P so this is therefore the electric, the electric field will be, these are electrons, electrons will move the side and electric field will be directed in this direction.

So, any hole which enter somehow other than the depletion region can be dragged, so for hole it is not a big hump, for hole it is basically a small slope. Similarly, electrons on this side will be also a small slope as you move. So please understand though majority current carriers do not contribute to the current when both are cut-off which means that when both $V_{BE}$ and $V_{CB}$ are reverse biased RB-RB then we define it to be cut-off, but please understand this is by virtue of the fact by which cut-off because majority of the current carriers are very low or almost zero.

But please understand there might be a current because of the minority current carrier because for minority current carrier it is not a hill but a hump going downwards, right. But since, that current is very-very low may be ten or hundred orders lower than the actual current, so if you have milliampere current flowing for the forward bias or on current this might be of the order of microampere or much smaller than that, we generally neglect that current and say that okay the device is off.

But please keep in mind that as a second order effect, you generally have these currents which are there with us by virtue of minority current carriers, they are always there that makes our life slightly difficult. So, we understood cut-off so cut-off is what when both my emitter base as well as collector base junctions are reverse biased and you therefore, understand why they are reverse biased.

(Refer Time Slide: 24:40)



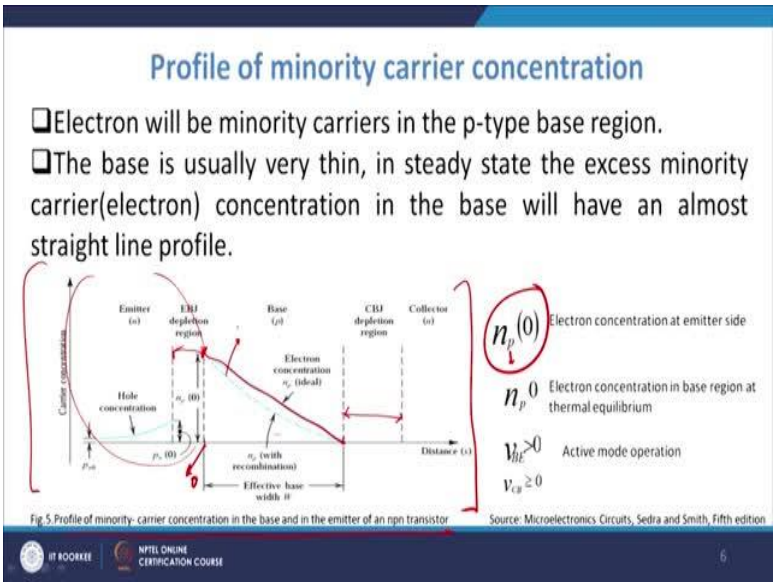

The second is basically my active mode of operation and this is the most actively use mode of operation generally used in amplifiers to a larger extent, where your emitter base junction is forward biased but your base collector junction is reverse biased, please understand the idea behind it. When I have emitter base junction as forward biased you are allowing larger number of emitter atoms which is electrons, emitter free charge carriers to enter into the base region to collector region, right. So currents are very-very large so when you reverse biased a collector region you actually allow large amount of, you allow acceptance of large number of electrons from the emitter side, right.

So, therefore, in the active region emitter base is always forward biased and collector base is always reverse biased, right. Let us look at reverse active, reverse active is when you just reverse these two. That means your emitter base junction is now forward biased and you collector base junction is, oh sorry, your emitter base junction is now reverse biased which means that you do not allow large number of emitter currents to flow but then you make your forward biased collector base junction.

The last one is saturation in which case both emitter as well as collector are forward biased, in this case you have a large amount of current which flows through the because both your collector base junction and emitter base junctions are forward biased. As a result you allow a large number of amount of current to flow. So there are four modes of operation, in general as we move on we will see that I can use transistor as a switch, right, wherein I can move from cut-off to active, active to saturation.

I can also use transistor as an amplifier when I moved from active to reverse active and so on and so forth. So just by varying the biases, right applied biases I can let it works sometimes as a switch from ON to OFF state and vice versa and I can also let it work as an amplifier. So we will see as we move along what these operations are and how these operations looks like, right. We will not give details of this one but I will give a brief idea about where we are at this stage.

(Refer Time Slide: 26:50)




As I was discussing with you that there will always be minority current carriers which will be there apart from majority current carriers in a particular region. Now if you look at this plot, see the plot which you see in front this plot, this is basically a profile of the minority carrier concentration in the base and the emitter of an NPN transistor, so this is my emitter side, this is my emitter base depletion junction. This is the emitter base depletion junction, this is my base region and then this is my collector base depletion junction here.

So, your NPN transistor base is basically P type right. So, what does this graph show is that in ideal condition the electrons concentration which electron is basically a minority current here on the base side will fall linearly from $N_{P0}$, $N_{P0}$ if you look very carefully $N_{P0}$ means electrons on the P side at zero means this is zero this is zero. So from there it will be maximum because it is nearest to the emitter side and then it will fall to zero near the base at the emitter side. Why, because at the collector side all the electrons are pulled by the virtue of the reverse bias of the base collector junction.

So at the emitter side you have large number of minority current carriers because electrons are being fed from the emitter side. So your NP0 is typically very high. But as you move towards the collector side, since collector is drawing all the electrons away from it I would expect to see per unit volume electron available there to be relatively very low, right. So that is the reason we see a linear, almost a straight line drop of electron concentration from a very high value to almost zero value near this thing.

But we will not go into details of it but with recombination because actually this is with assumption it is the only combination, if there is recombination it is almost a parabolic sort of a non-linear profile which you see in front of you. Not at this stage not very important from the point of view of understanding the physics of the device. But please understand two things that the doping concentration of my minority current carriers in the base region is a function of my width effective base width and so on and so forth, right.

So, this is what the effective base width looks like, may be in the next turn when we come back we will discuss about the various concentration, electron diffusion current and so on and so forth. What we have done in today's lecture is given you an idea about, this thing about how the minority current carrier behave in a P type base region. We have also come to a point how we have, how a PN junction works and therefore, how it can be translated into an NPN and PNP transistor and its functionalities right.

So, we understood all these things, when we meet tomorrow or next time when we meet I will be discussing some mathematics about the current flow in a transistor which is carried off both drift and diffusion and then see how we can optimize the functioning of an NPN and PNP BJT. So, this is what we have learned today and we will hope that you have understood a part of this work as far as this course is concerned. Thank you very much!

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−03**
**Bipolar Junction Transistor: Modes of Operation-II**

Hello everybody and welcome to the NPTEL online certification course on Microelectronics: Devices and Circuits. What we will start today's lecture by just recapitulating what we did in the previous turn. We had seen that given a NPN transistor where emitter is basically an N type, base is P type and collector is again N type, base width is very-very small, emitter doping concentration is the highest, base doping concentration is the lowest, and collector is moderately doped.

We have also seen the reason why the base is relatively lower doped. And the reason was that with low doping we will have less number of majority carriers there and therefore the electron-hole pair recombination there will be smaller. And as a result most of the electrons which started from the emitter side will reach towards the collector side and we will get large current on the collector side. So that was the prime reason why the base width was effectively very thin not only that the base doping concentration was also very small.

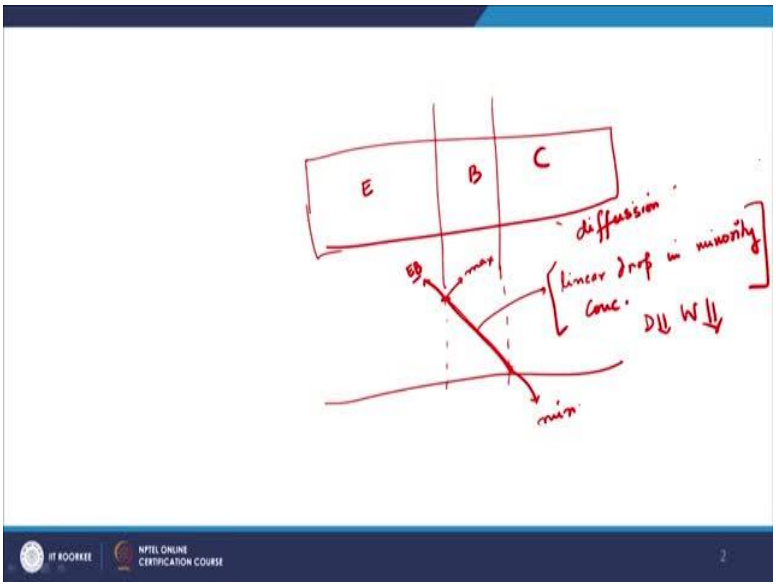
With these two things in mind we have also seen yesterday in the previous discussion or in the previous lecture that the base current itself consists of two components. The first component is primarily the current which is the base current which is coming out because of recombination of holes. So when electrons recombine with holes to replenish those holes you have to insert a current that is one part of the current, which is basically a current which is being supplied by the forward bias base to emitter junction.

There is also a current which is basically a current which is primarily due to the diffusion of the minority current carriers in this case electrons from hole to the from base side to the emitter side because that is though it is forward biased but

for that it can easily move through the region. Now we have also seen in the previous turn that with this basic concept there are two types of transistors available one is known as NPN and we have also a PNP transistor. We have also seen the various modes of operation and how various modes of operation work under what condition. For amplification purposes we have learned your emitter base junction should always be forward biased, and your base collector should be reversed biased.

Similarly if you want to do in a cut-off mode then both junction's emitter base and collector base should be reversed biased. If you want to move into saturation then emitter base should be emitter base forward biased and base collector should also be forward biased and so on and so forth. So we did have two biases and two types of combinations so they were effectively four modes of operations of BJT which we have dealt in the previous turn.

(Refer Time Slide: 3:42)

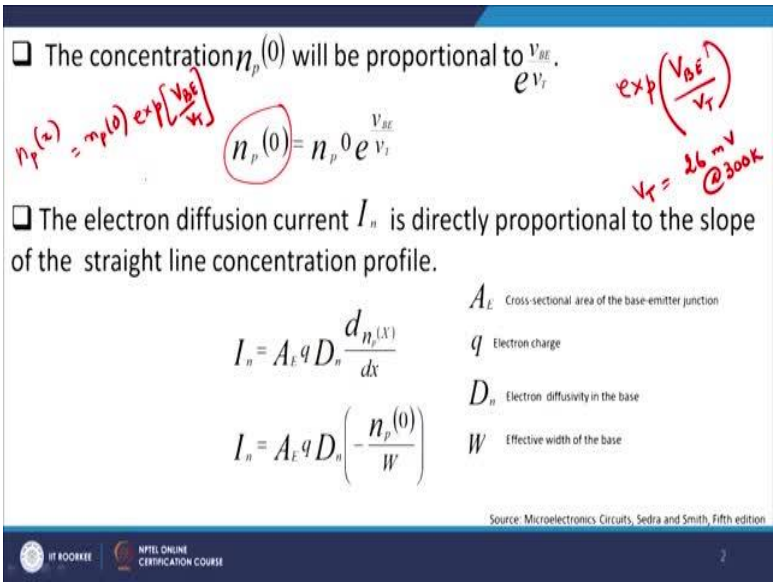



We have also seen that in the previous turn that if you have a therefore NPN transistor right. And then if you want to plot the minority current carrier concentration in the base region then it was shown to you that in the base region if I plot a graph then in the base region I would expect to see the minority

current carrier concentration to fall almost linearly right. And this linear falling is primarily due to so this is basically a linear drop in minority concentration, right and this is minority concentration.

This is primarily to do with a thought that your doping was low and your width of the depletion region is also low. Under these two criteria we saw that it is almost like a straight line which is just falling between point. And as you can see so this is the emitter base collector this is basically your EB junction emitter base junction your concentration of minority current carriers maximum here right. And it becomes linearly to become almost minimum at this point we have seen why this profiling is being maintained to a larger extent. As you can see this will be primarily a diffusion phenomena right and therefore we will depend upon the concentration gradient of the charge particles which is basically here electrons within the base region.

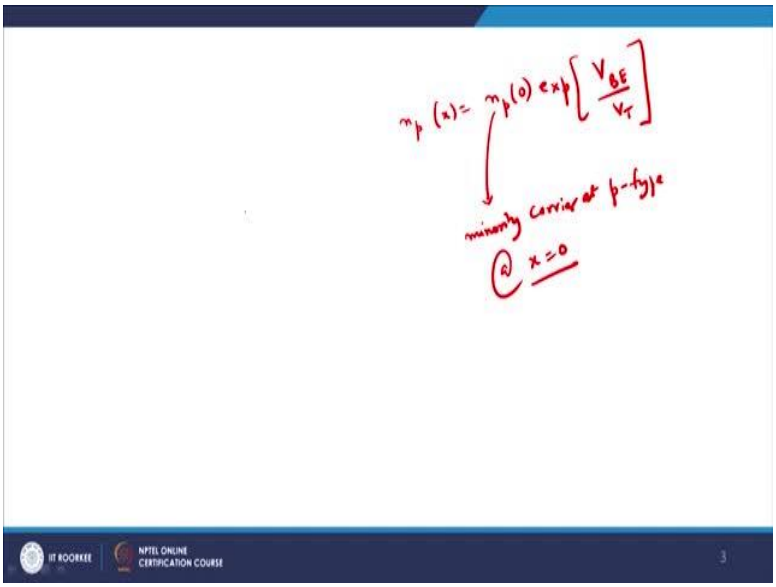
(Refer Time Slide: 5:14)




Today what we will be looking into is have a look into our basic fundamental principles which were left yesterday. If you look very carefully then, the minority carrier concentration will be proportional to e to the power or exponential $V_{BE}$ base to emitter voltage and divided by the thermal equivalent

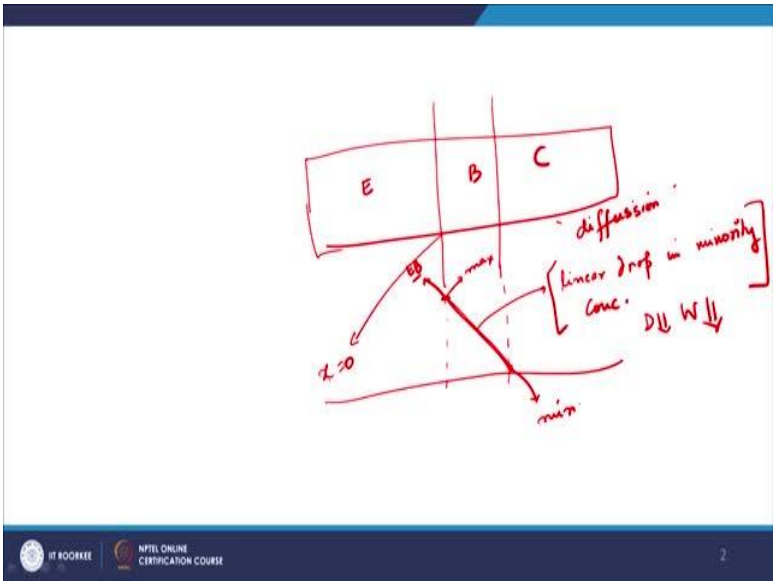
voltage which is effectively equals to 26 millivolts. So $V_T$ is equal to 26 millivolts at 300 Kelvin right 26 millivolts at 300 Kelvin which means that as $V_{BE}$ increases based emitter voltage increases the value of $n_P(0)$, $n_P0$ is it should not be $n_P0$ it is basically the formula is $n_PX$ at any point is equals to $n_P0$, e to the power exponential $V_{BE}$ by $V_T$, please make a correction here. That it is $n_PX$ will be equals to $n_P0$ exponential $V_{BE}$.

(Refer Time Slide: 6:17)



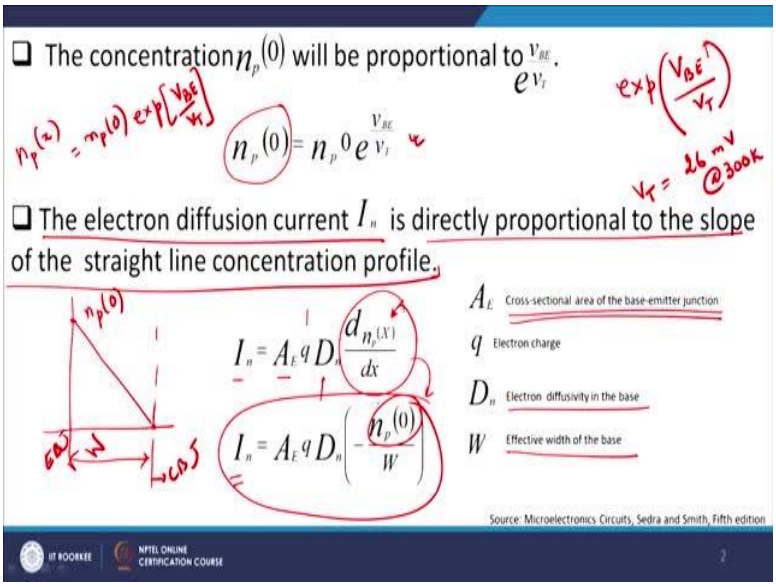

Now what does it tell me about this thing. So if it is $n_PX$ equals to X is the distance from the base region $n_P0$ exponential $V_{BE}$ minus by $V_T$ as you can see from this concept or this basic idea $n_P$ means n suffix p means this is basically minority carriers minority carrier concentration, carrier concentration at P type base at X equals to 0.

(Refer Time Slide: 6:54)




So X equals to 0 primarily means if you look at X equals to 0. It basically means this point. This corresponds to X equals to 0.

(Refer Time Slide: 7:00)




And let us see if we proceed forward it gives me an exponentially falling down, which means that though it is basically shown as exponential we approximate it as a linear fall concentration of electron as you can see. Therefore, if my $V_{BE}$ would have been larger this is where the basic fundamental principles comes if

$V_{BE}$ would have been larger I would have expected to see the value of $n_pX$ to be also large right is it okay? For the same value of X which you see.

Because $V_{BE}$ is typically large then I would expect to see $n_pX$ to be also large because it depends upon the value of any particular point. Now the point which I wanted to make overall expression therefore is that if so therefore we define in this manner. Now electron now so, so you see so therefore what will happen is for the minority current carrier it will be basically a diffusion current which will be flowing in the base region. From a region of high concentration on the emitter base junction to region of low concentration or almost zero concentration on the base collector junction.

Now obviously this current will depend upon the difference of the concentration at the interface between emitter base and base collector. So as you can see here the electron diffusion current $I_n$ is directly proportional to the slope of this straight line of the concentration profile which means that if concentration profile is assumed to be straight I can very well find the value current $I_n$ to be equals to $A_E$ multiplied by Q multiplied by $D_n$ multiplied by $d_P$, $dn_pX/dx$ right.

This is the standard formula for finding out the current where $A_E$ is the cross-sectional area of the base emitter junction right, it is the cross-sectional area of the base emitter junction, q is the electronic charge, $D_n$ is the electron diffusivity in the base and W is the effective width of the base and W is the effective width of the base. So you see $I_n$ therefore if you take this equation, right and then this comes out to be minus $n_p(0)$ by W and the reason being very simple because we have assumed till now that the fall in concentration or the drop in the concentration will be almost a linear drop right, we have already assumed that it will be a linear drop.

So if it is a linear drop then it is very easy to find the slope of a linear graph provided you know this is basically your $n_p0$ which means the concentration of

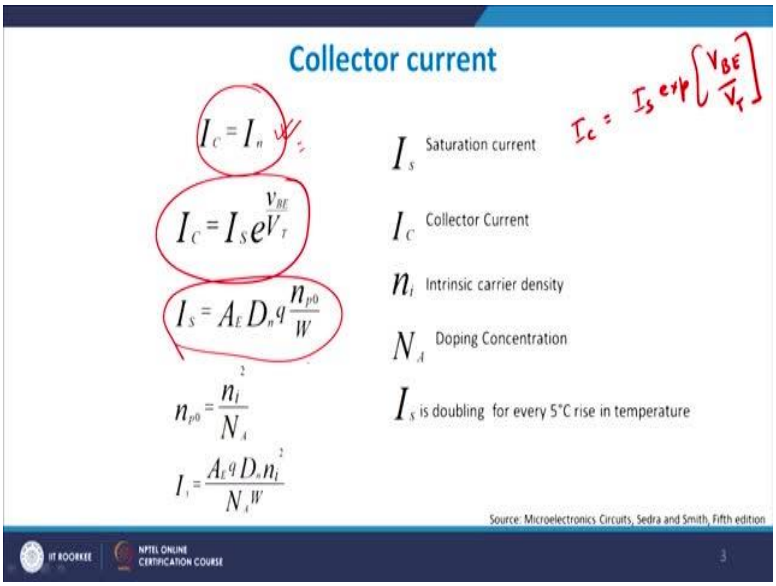
minority current carrier at the edge of the emitter base junction. So this is my emitter base junction, and this is my collector base junction right, this is my EBJ and this is my CBJ. Then this minus 0 assuming that you have reached zero value I get $n_p0$ and this is basically your W width of the base region right.

And that will give you the value of $I_n$ in this case. Please understand one more important point here that in no way, right the depletion width on the emitter base or the collector base junction is influencing your electron concentration in the region of the base minority current carriers contribution. Because they will be nonetheless available at those particular points and therefore it will not affect the overall concentration of the minority profile in this case.

Right so this is what we have learned so we have learned this basic fundamental principle that therefore the current will depend upon the area $A_E$ of the emitter and it will also depend upon the minority current carrier concentration at the emitter base junction, right. And we have assumed at this stage that the minority current carrier concentration on the collector base junction is approximately equals to 0. If it has got some finite value obviously this will be a lower this will be a lowered value and therefore I will get a lower current, right.

So higher the difference or higher the gradient between the emitter base and collector base in terms of concentration gradient I would expect to see a larger current to flow in this region. Right so we have learned therefore two or three important points out of all these discussion. And the first important point is that the diffusion current is independent obviously of the applied bias, it only depends upon the area, the diffusion constant and most importantly on the concentration gradient which you can see from this equation in front of you this equation right.

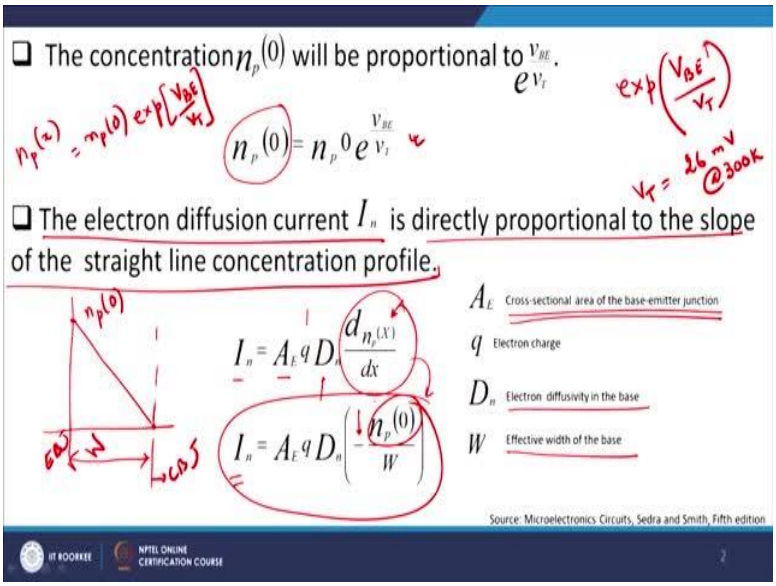
(Refer Time Slide: 11:56)




With this knowledge or with this idea we come to the next idea which is basically my collector current now assuming that you did have all the emitter charge carriers which were electrons and there was zero recombination in the base, I would expect to see all the electronic charge reach the collector side, right. So I can safely assume in that case that my electron current $I_n$ will be equals to the collector current, right. So contribution in the hole is relatively small because it is a minority current carrier. And therefore the major contribution is from the majority current carrier which is basically the electrons.

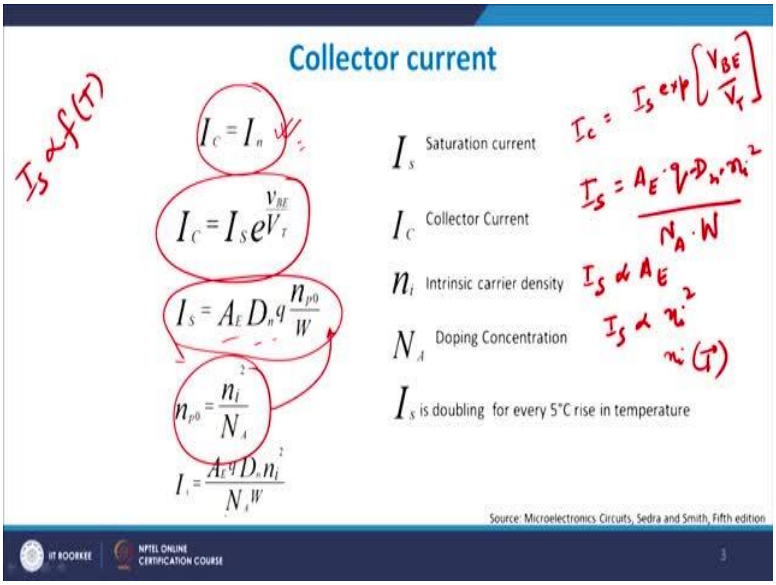
And therefore I safely write down $I_C$ to be equals to $I_n$, right and that is known as the basic equation which you see. Now $I_C$ is the collector current can be written as $I_S$e to the power $V_{BE}$ by $V_{TE}$ so $I_C$ equal to $I_S$ exponential $V_{BE}$ $V_T$, right. Where $V_{BE}$ is the base to emitter voltage and $V_T$ is the thermal equivalent voltage. $I_S$ is defined as the saturation current, right. Very-very small current but it is the saturation current. Now, so I get $I_S$ there therefore what we can write down $I_S$ to be equals to this much $I_S$ is the saturation current $A_E$ into the $D_n$ into q into $n_p0$ by W right.

(Refer Time Slide: 13:23)




If you look at the previous slide you see you have a negative sign here. It is a negative sign which you see. And the reason is that the concentration starts to fall down as you move from X equals to 0 to X equals to W. So there is a decrement in the concentration and as a result there is a negative value which is been shown here.

(Refer Time Slide: 13:43)



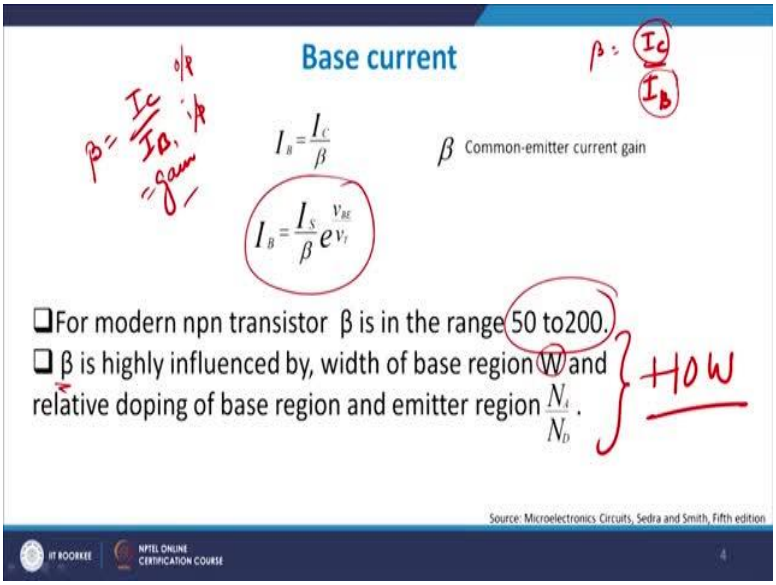

Now as you can see here $I_S$ is given by this formula because I just replaced $I_S$ by $A_E$, $D_n q$ $n_p 0$ by W, again assuming 0 charge carriers on the collector base

junction. Now from my earlier basic semiconductor physics theory I get $n_p$ 0 to be equals to $n_i^2$ square by $N_A$. When $n_i$ is intrinsic current carrier concentration and $N_A$ is basically my acceptor ion concentration. And therefore if you take $n_p0$ and place it in this equation I get $I_S$ to be equals to $A_Eq$, $I_S$ to be equal $A_E$ multiplied by q multiplied by $D_n$ multiplied by $n_i$ square divided by $N_A$ into W, W is the, W is the width of the base region.

Which means that the $I_S$ saturation current depends on the numbers of intrinsic carriers which actually doubles for every 5 to 10 degree rise in temperature, right. So you would expect to see a very large change in the value of saturation current in this case, second important point which you should be very careful and you can go through it you can see it that as the area of the emitter the cross-sectional area of the emitter increases your short-circuit your IS also increases. So $I_S$ is therefore is directly proportional to $A_E$ right and $I_S$ is also directly proportional to $n_i$ square.

Since $n_i$ is directly dependent on is a function of temperature I can safely assume that $I_S$ is therefore also a function of temperature right higher the temperature higher is the value of $I_S$ which is saturation current in a standard NPN or a PNP transistor.

(Refer Time Slide: 15:38)




Now my base current therefore $I_B$ is equals to $I_C$ by β and therefore we define β to be as a common emitter current gain. Because if you look at β, β is basically $I_C$ by $I_B$, right under the condition that you have a common emitter configuration so it is basically output by input and therefore this is also referred to as a current gain equation, right. Now, therefore again as I discussed with you just now $I_B$ will be equal to $I_S$ by β into e to the power $V_{BE}/V_T$.

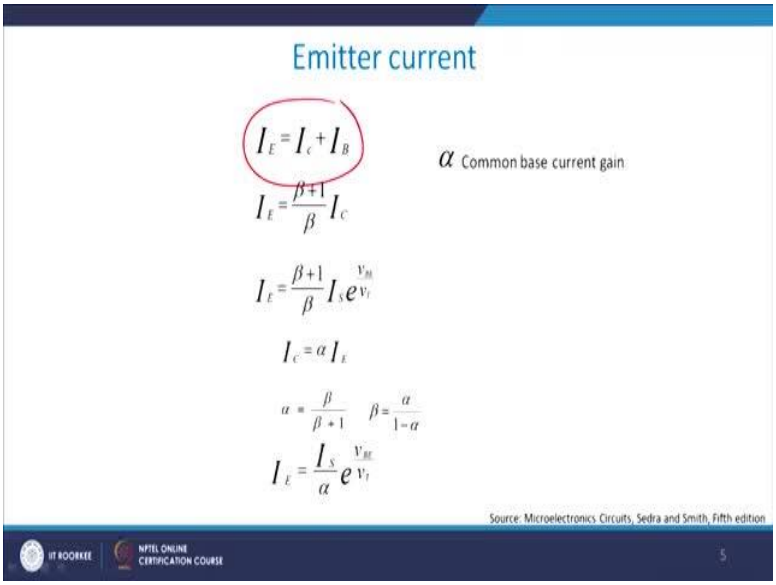
Typically current model NPN transistors have β in the range of 550 to 200, right. The typical range of β is between, is typically between 50 and 200, right. And this range depends upon many factors apart from the material of the device it also depends upon the type of doping that has been done on the device, the level of doping which has been done on the device, right.

Now if β is high β is basically as you can see is basically $I_C$ by $I_B$ right. So if β is typically of the order of 500, 200, 300 so on and so forth. It primarily means that your collector current orders may be two orders or three orders higher than the base current and this is expected also and the reason behind this is that all the carrier concentration which was available to you from the emitter side, right

goes towards the collector side, right and therefore recombination of electrons or holes on the base region is very-very minimally low.

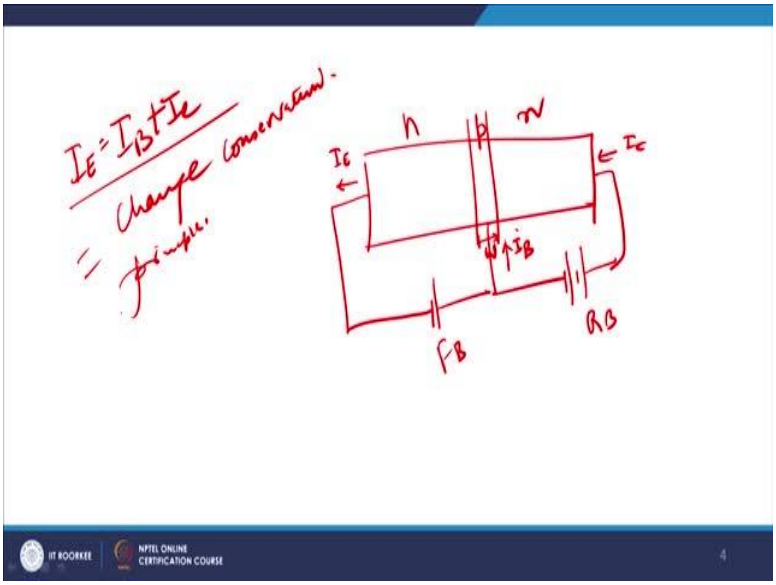
When it is low your β value actually becomes typically large, because $I_B$ value reduces drastically. So as you can see β is highly influenced by the base width and related doping of the base region and the emitter right so it depends upon $N_A$ upon $N_D$, right. I suppose you will be able to find out why and how, right. I will leave it as an exercise to you to please understand why is it like that. That why β depends upon the base width region, region of the base region of the base. If it is very thick then β is relatively small if it is very thin then β is relative larger and so on and so forth, right. I give it as an exercise to you find out the value.

(Refer Time Slide: 18:05)



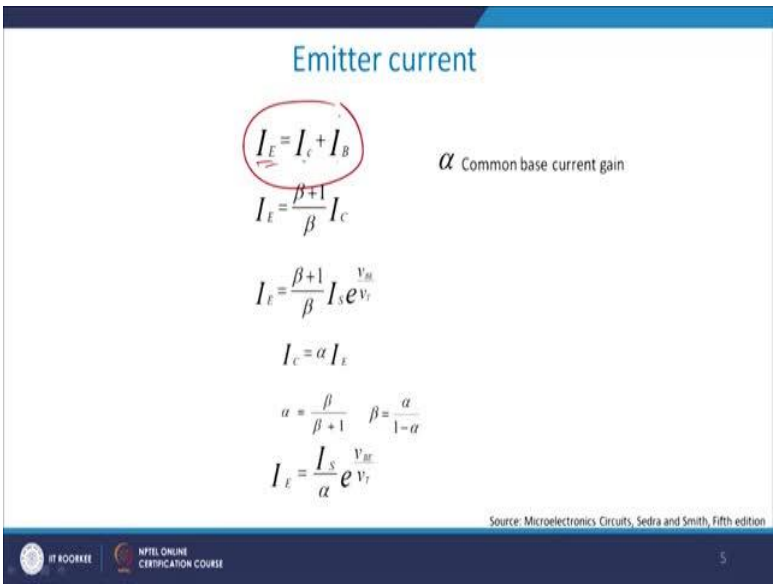

The Kirchhoff cannot be violated all of us know that over the years we have seen that. It cannot be violated in more sense than another. And therefore I can write down $I_E$ to be equals to $I_C$ + $I_B$, right. If you go back to your previous slides as far as this course is concerned then you will see that, or maybe let me draw for you.

(Refer Time Slide: 18:20)



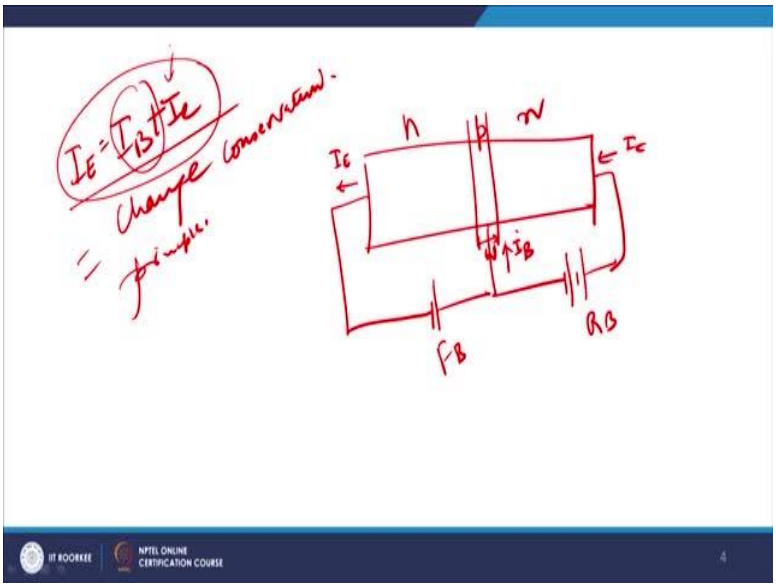

That if I have got a NPN or a PNP, if I have an NPN or a PNP then what I get from here is that this is forward biased and this is reverse biased and this is forward biased, right. And this is the width region where W is the width of the base region which is oppositely biased as N and P region. Now if you look carefully so it is an NPN. So if its emitter current is $I_E$ this will be $I_B$, $I_B$ and this will be $I_C$. So I can safely write down $I_E$ to be equals to $I_B + I_C$ and this is basically charge conservation principle.

(Refer Time Slide: 19:28)



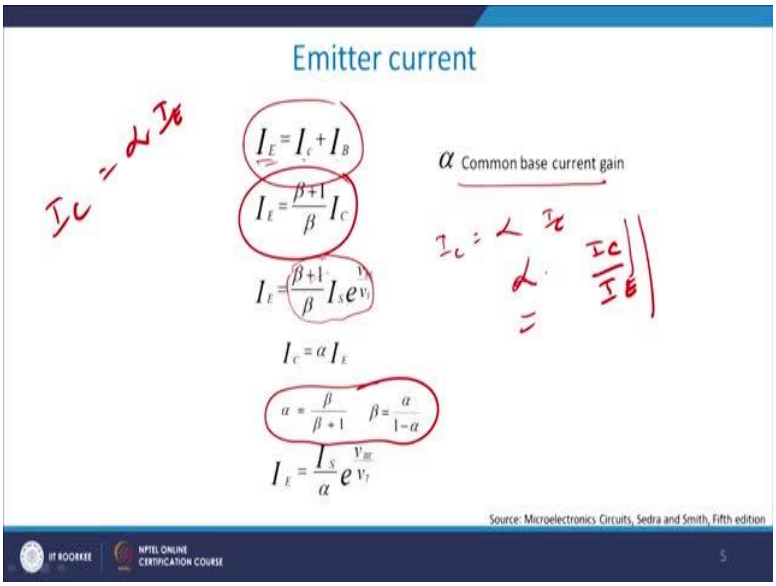
So this is basically a charge conservation principle, the charges is always conserved irrespective of the type of device you take irrespective of the material of the device you take. So $I_E$ is equal to $I_C$ + $I_B$ because you will be following the Kirchhoff's law and since you will be following the Kirchhoff's law the total emitter current which is entering into the system must be exactly equal to the total.

(Refer Time Slide: 19:50)




So if you look at very carefully $I_E$ is equal to $I_B$ plus $I_C$ and therefore total emitter current in the system is sum of base current and collector current, right. And that is how you manipulate. So you try to therefore maybe control the flow of the collector current by increasing or decreasing the value of the base current right by changing the value of base current you can actually make $I_C$ or $I_E$ almost close to each other.

(Refer Time Slide: 20:17)



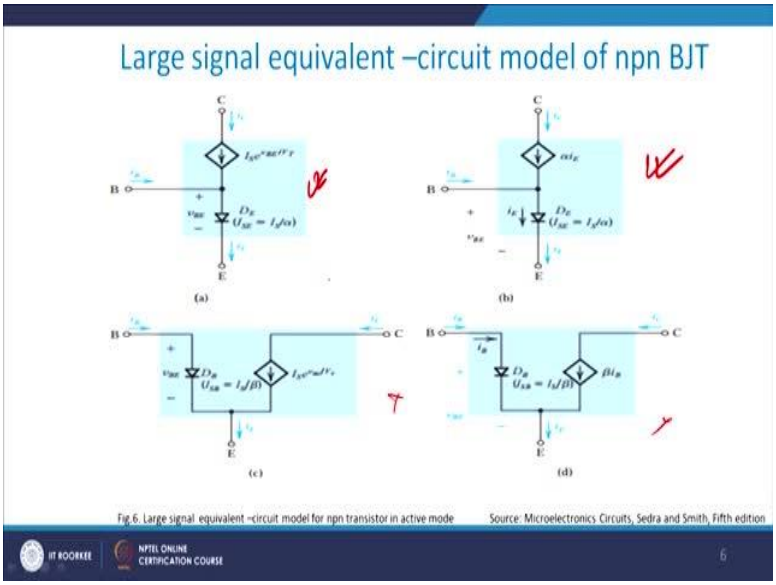

And that is the major study area which people are working. So $I_C$ +, $I_C$ + β we know that $I_E$ is equal to β + 1 upon β into $I_C$ well that is the standard formula which people use, right. Then $I_E$ will be therefore equals to β + 1 by β into $I_S$e to the power $V_{BE}$ by $V_T$. And from here I get $I_C$ equals to α times $I_E$. So all this $I_C$ part which is basically your this part right is $I_C$, right, I get $I_C$ to be equals to therefore α times $I_E$ which is a known fact. α therefore is equals to $I_C$ by $I_E$, right, and α is therefore referred to as common emitter current gain. Sorry α is known as common base current gain β is common emitter current gain.

And as you can see α depends upon the value of $I_C$ and $I_E$ which is understandably so because $I_E$ is basically the total current being entering into the system and $I_C$ is the total current exiting the system. So if you solve it, I get $I_C$ is equals to α times $I_E$ I get $I_C$ to be equals to α times $I_E$. Which means that and this is quite simple and straightforward so α will be very close to 1. If it is 1 then $I_C$ will be equal to $I_E$. In that case we are assuming that all the charge carriers which were initially originating from the electron or from the emitter side will actually terminate on the collector side and therefore $I_C$ is equal to α times $I_E$, right. Two important equations to take care of α is equal to β upon β + 1 and β equals to α upon 1 - α. These two important equations are required for solving

problems, right. And that is quite interesting problems solving which you can do using these two questions. Let me come to the last part of our talk and that is basically your large signal equivalent circuit model for NPN BJT.

Well, you see what we try to do is that when we have a device primarily in its initial stages of fabrication and testing we want to convert that device into an appropriate machine or a simple machine so that we can actually record its current versus voltage characteristics in a much better manner, right. And therefore we need to do a circuit level implementation of the device.

(Refer Time Slide: 23:07)

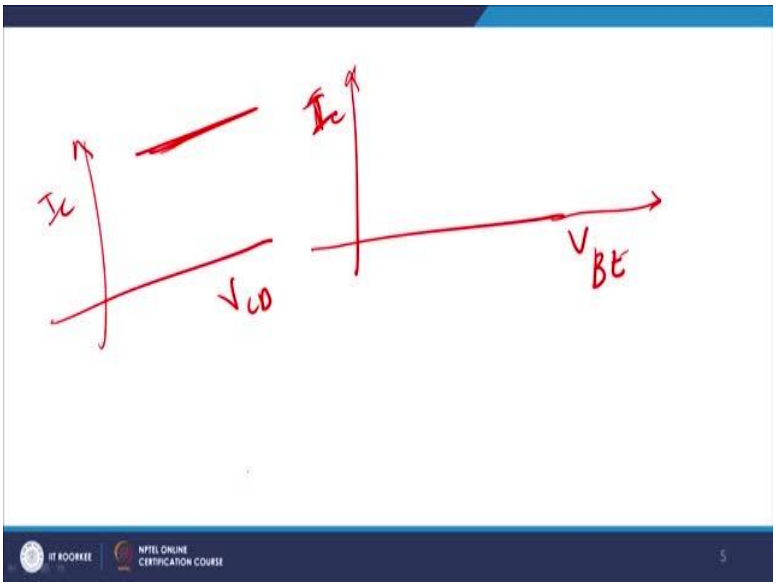



So this is basically a large signal equivalent circuit model of an NPN BJT. Which means that if you take a BJT right and place it in a large circuit domain then we can replace that BJT by these large signal equivalents, these large signal equivalents. This one as well as this one you cannot do this or this, but these you can surely present as an amplifier let us see what is what. As you can see if base current is the emitter current and base is this terminal. $I_C$ is the collector current and C is the collector profile and E is the emitter terminal and $I_E$ is the emitter profile or emitter voltage or current which is available here. Now you see if you look very carefully that whatever voltage source was there

on the source side right we were able to give that voltage source and draw current. This current has same into the collector current so collector current is approximately same as the base current the emitter current the only problem is if the base width is large or the doping is relatively high then this assumption does not hold good, right. As long as the base width is small and your dimensions or your values are very very small in terms of potential, you will not be able to do this problem in a proper fashion.

Now, you see here if you look at the first picture or the first diagram here of the schematic here you will see that I have replaced the collector side by current source, right and that's true also.

(Refer Time Slide: 24:42)



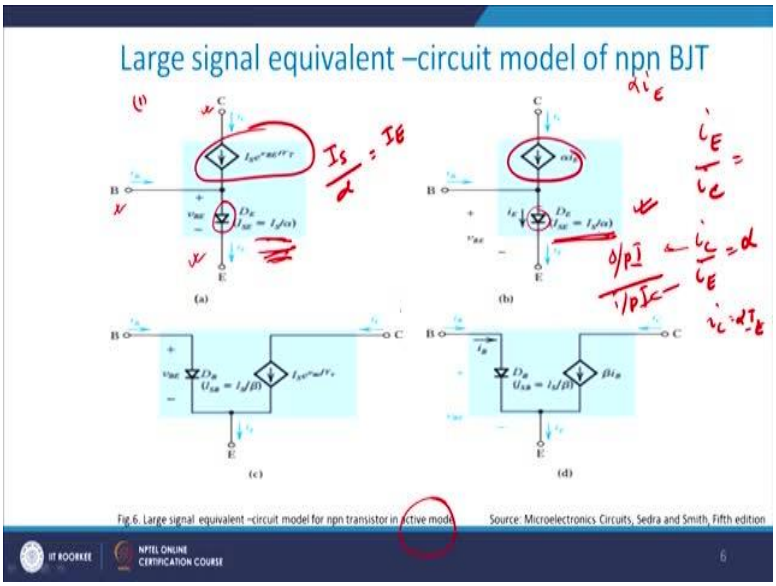

So, if you look very carefully once if you plot $V_{BE}$ versus maybe if I plot if I plot $V_{BE}$ versus $I_C$ collector current most of the time it will be difficult for him or her to explain to me this $I_C$ versus $V_{BE}$ of any typical bipolar technology which you see. If that is the case we assume that we assume that that the base side potential is always fix it does not vary and therefore it does not vary to larger extent and therefore the collector starts to behave like a constant current

source which means that even if you change the reverse bias of a collector base junction too much the current does not change too much.

So as I discussed in the previous turn it starts to behave like an ideal current source, right so if you plot $I_C$ versus $V_{CB}$, I will almost get something like this a constant current source, right. Which means that I can replace the collector side by a constant current source if it is an ideal constant current source my output impedance will be infinitely large right and I will get almost fixed value of current even if I change the value of $V_{CB}$ too much, right.

And therefore, you see this part which is the collector part is basically a controlled current source, right. So it is basically a controlled current source, who is controlling it the value of $V_{BE}$ base to emitter current, right, that is controlling the collector current we have already discussed point and $V_T$ is final equivalent voltage.

(Refer Time Slide: 26:28)




Now onto the emitter side this is the base and this is the collector on the emitter side we have a reverse biased PN junction why we have already seen this why. The reason being that at this stage since emitter base junction when you are using in an active mode because we are using this in the active mode emitter

base junction will always be reversed biased, right because it will be reverse biasing the emitter base junction.

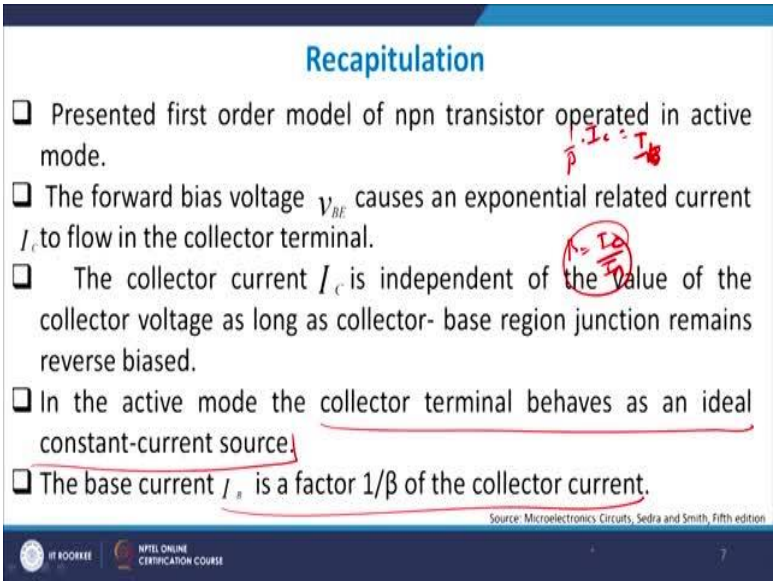
So once you reverse bias it I replace it by a reverse biased PN junction as you can see here and the current flowing through it is given as $I_C$ is equal to $I_S$ by α, right and that makes sense also where $I_S$ is basically the saturation current divided by α is basically your $I_E$. And that is the reason this is therefore the large signal equivalent model of a NPN transistor BJT transistor.

Now the same thing if you look at this point which is this one you can see that it again at the basically collector base emitter and this model is again $I_S$ by α as you can see same thing emitter current is available to me but now I can replace this by α times $I_E$ because remember $I_E$ by $I_C$ is equals to sorry $I_C$ by $I_E$ is equals to α output current divided by input current is basically referred to as α.

And therefore I can refer $I_C$ as α times $I_E$ this is what people have done here and shown it here. This remaining exactly the same as the previous case, we will get this profiling done in a much more detailed manner. This takes care of approximately the whole large signal model equivalent profile of the model.

The diagram number C and D as you see you can do it by yourself is basically the common base configuration and therefore the base sorry it is not common base it is basically the base side is connected to input source and the collector side is connected to output source, right and from there I can find the value of C and D equations in a much more detail manner but typically what we have learned is basically large signal model of a NPN transistor.

(Refer Time Slide: 29:00)




So let me recapitulate what we have learned till now, we have presented a first order equivalent NPN model operated in the active mode. Active mode is when emitter base junction is forward biased and collector base junction is reverse biased. Any forward bias profile or entity will always cause an exponential change in the value of current and therefore the collector current $I_C$ is independent of the value of the collector voltage as long as the collector base junction remains reversed bias.

So as long as it is reversed bias I do not worry what is happening on the other side, on the other side means on this side of the collector. And we end up having therefore almost a steep profile in this thing. In active mode the collector current terminal behaves like as an ideal current source we discussed why. And the base current $I_B$ is a factor of 1 by β of $I_C$ so basically 1 by β of $I_C$ is equals to 1 by $I_E$ or 1 by β.

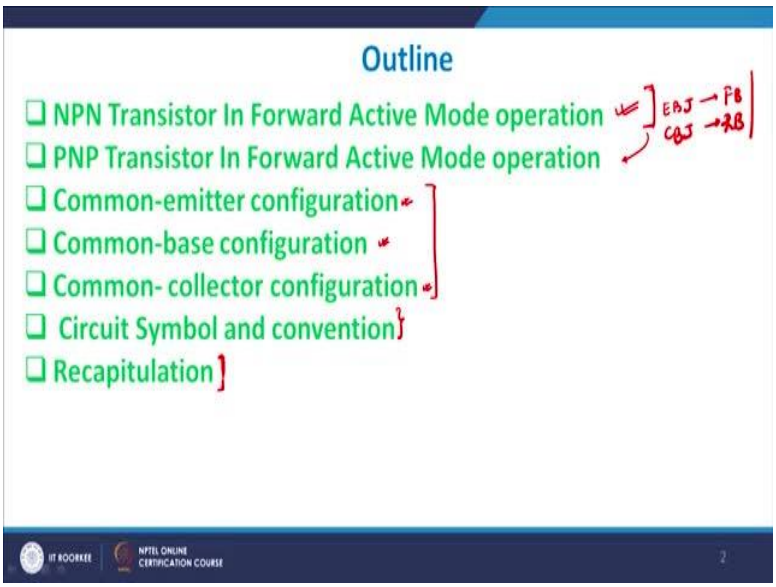
So I can safely write down it β to be equals to $I_C$ by $I_B$ for $I_B$ is a base current, right. It will be a very small quantity but nonetheless the base current will be there which is available to you right. So this takes care of all your discussion here, I thank you all for your patient hearing. Thank you very much!

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−04**
**BJT: Operation in Active Mode Circuit Symbol and Conventions - I**

Hello everybody and welcome to the NPTEL online certification course on Microelectronics: Devices and Circuits. In our previous lecture we have understood the basic functionality of a transistor and why is it basically working as it should be. We have also understood the various numerical formulae's used for finding out the current flowing through the circuitry.

We have also gathered a position or we have also understood or appreciated the role of minority current carrier's in the NPN or a PNP junction. We have also seen that it can work in four different modes of operation for the purpose of switching as well as for amplification purposes.

(Refer Time Slide: 1:17)



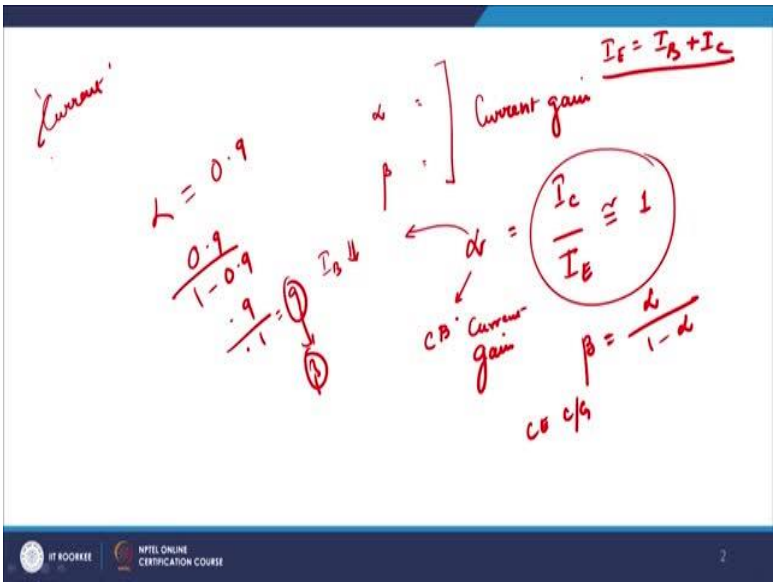

In this lecture we will be looking into its operation so the name of this lecture or the title of this lecture is Operation in the active mode, circuit symbols and conventions. So that is the lecture topic and what we will be looking into in this module or in this lecture is basically NPN in forward active mode. So once you have understood NPN in forward active mode we will be easily able to understand PNP in forward active mode, right.

Forward active if you remember is when your emitter base junction is forward biased and your collector base junction is basically reverse biased, right. So this is your forward active

mode junction which you see. We will also look into three configurations of working of this transistor which is basically common emitter, common base and common collector.

And then we will see some of the conventions and circuit symbols used and then we will recapitulate what we have learned across this module. Now if you remember in the previous discussion or in the previous understanding we had encountered two important terms one was basically known α and other one was basically known as β, right.

(Refer Time Slide: 2:30)




And both were if you look very carefully both were actually current gains, right, current gain. α was defined as $I_C$ by $I_E$ so this is basically your common base current gain current gain. And if you look very carefully the numerical value will be very close to one. And the reason being that $I_C$ is approximately equals to $I_E$ because $I_B$ is very very small assuming that we already have this Kirchhoff's this is already there. I would say this to be very close to 1 right α to be very close to 1.

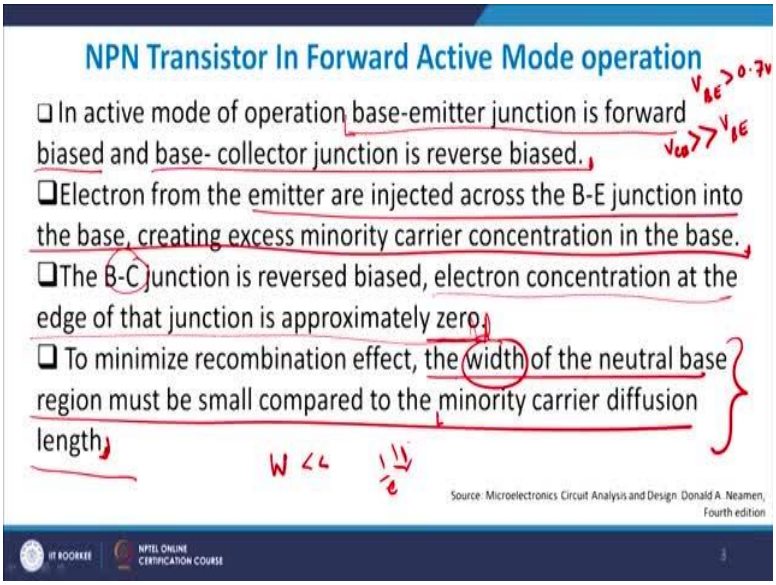
Typically β is known as common emitter current gain and it is referred to as α upon α - 1 so that is what we get as β here so it is basically α multiplied by 1 - α. So if you take α to be equals to 0.9 approximately then I get 0.9 by 1 - 0.9, so 0.9 by 0.1, I get approximately equals to 9, right. Is this clear conceptually so I get 9 so this β which you see in common emitter current gain.

So you see we will discuss this as we move along. But primarily you see that we have tried to find out in our previous discussion the current gain because of BJT. Now let me show to you

the transistor action how we can achieve a high gain or a high amplification in this case that will be the main focus of our topic in this case.

So let me explain to you how this first topic which is basically NPN transistor in forward active configuration or bias.

(Refer Time Slide: 4:43)




As I discussed with you in the previous turn in active mode of operation, right base emitter junction base emitter junction is always forward biased and the base collector junction is always reverse biased. You have to ensure this base collector junction should be forward biased. Now, one important point which you should note down as far as designing of a BJT is concerned. If you remember the previous talk that the $I_S$ value the saturation current value was depending upon the cross-sectional area of the emitter.

So if you have got two transistors both NPN or both PNP of exactly the same doping concentration but one having a large emitter area as compared to the first case then the second will have a larger current also as compared to the first case so this was one of the techniques by which people use to increase the level of current in a structurally reconfiguring our NPN or a PNP transistor, right.

Similarly you should also be careful the first point which is to be noted is that your base emitter junction should be forward biased which primarily means that your emitter base junction of an NPN silicon transistor should be at least the forward bias should be at least more at 0.7 volts, right for a silicon. And the reverse bias should be much larger than the forward biased please understand this point also. So you have to be very cautious about two

things that is $V_{BE}$ base to emitter junction should be at least greater than 0.7 so the requirement therefore is that $V_{BE}$ should be much greater than 0.7 volts greater than equal to.

And $V_{CB}$ collector to base should be much larger than $V_{BE}$. So the reverse bias is typically very large as compared to the forward bias itself and that you should be carefully looking into that properly. So now let us go to the second point therefore the electrons from the emitter are injected across the base emitter junction into the base and thereby creating an excess minority current carrier concentration in the base.

We have already discussed this point earlier and we also saw that that the minority carrier concentration will be falling linearly between the emitter base and the collector base junction. As you can see therefore the electron concentration at the edge of that junction is approximately which junction collector base junction is approximately equal to 0 we have also seen this to be true right. Now to ensure that that the minimum number of electrons gets recombined in the base region, right, the width of the base region should be very small as compared to the total length of the dimension of the device.

I will explain to you what does it mean and I will explain this term in a much more detail manner. It says that that the width of the neutral base region neutral base region why? Because you see you had initially P type for an NPN transistor you should have P type now you have minority current carrier concentration N inserting into it trying to make it neutral. The base width W this width of the neutral base region must be very small as compared to this minority carrier diffusion length.

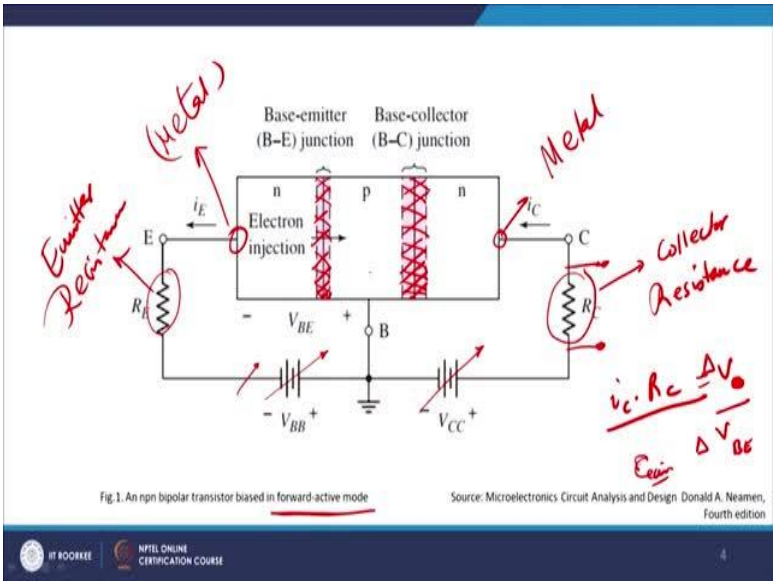
What is a minority carrier diffusion length we define diffusion length as the distance at which because of recombination the value of the minority current or the charge carriers 1 /$e^{th}$ of its maximum value. So if you entered at this stage and you went to a distance of 1 centimetre and you saw that there was a drop in the number of charge carriers by virtue of recombination. And we saw that 1 /$e^{th}$ of the charge has actually 1 /$e^{th}$, so 1 /$e^{th}$ of the charge has actually fallen down then we define that length to be diffusion length.

So if your width base width is even smaller than that then by the time the electron starts moving I would expect to see that no large diffusion has taken place or no large recombination has taken place sorry. So what we will expect to see therefore is that try to keep the width of the base much smaller than the active diffusion length of the minority current carrier's right and that is what this will minimise your recombination effect. Then the

obvious question asked is that why should we have the recombination in the first place. Well the reason for that is then then we have a control, we have a current controlled device here by controlling the base current I can control the collector current, right. Because $I_E$ is equal to $I_B$ + $I_C$ and therefore if I want to reduce my $I_C$ I need to increase my $I_B$ large make it large.

And as a result effectively $V_{BE}$ controls the value of $I_C$ by making the $I_B$ value vary from point to point or from value to value right and that is what one should be looking into in a detail manner. Let us look at therefore NPN transistor in the forward active mode configuration.

(Refer Time Slide: 9:57)



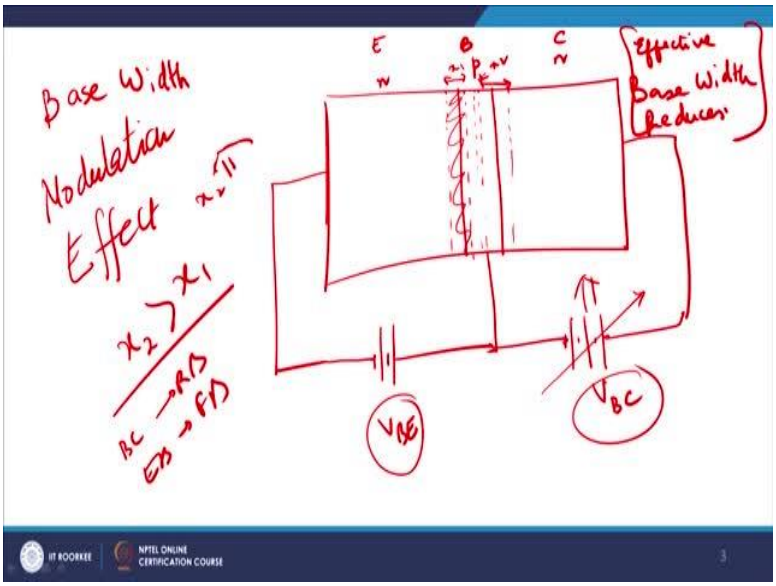

And we define this and we generally therefore put arrow head here and arrow head here assuming means implying that these are variable power sources or variable power supply which is there with you. This $R_E$ and $R_C$ are basically the resistances this takes care of the Schottky resistance at the contact because the contact with this side and this side will have metal contact right we had discussed this point earlier also. This will be metal this will be metal also so both are metal.

So metal contact will have some resistances plus the wire will have some resistances plus any resistance which you want to put it in the external world in order to restrict the current or get a voltage output you put those resistance. You add up all those things and we get a collector resistance here, collector resistance here and we also get emitter resistance here, right. Now what will happen is once this current flows here and $I_C$ multiplied by $R_C$ that is the voltage drop across this $R_C$ that is the output voltage $V_O$, right and you will be giving a small voltage

here input voltage is $V_{BE}$ so output voltage by $\partial V_{BE}$ $\partial V_0$ will be a gain for a transistor. So this is a very simple basic form of a transistor as an amplifier. As long as you are able to sustain large current across the collector end that multiplied by the collector resistance will give you a large voltage drop. This voltage drop will therefore change depending upon the value of $V_{BE}$.

So $V_{BE}$ is therefore your input voltage and therefore output voltage by input voltage will give you the current. If you can see very clearly your base emitter junction this pink side is basically my depletion this one is a depletion width is very small because it is forward biased and it is relatively large in dimension because it is reverse biased, right. And therefore base collector junction will have larger width and therefore larger effective bias whereas emitter base junction will have a relatively lower depletion width therefore a large current will be flowing from the base and emitter junction. Now at this stage what we have not done or what we have understood therefore is two important points I should take care here before we move forward is one is what is known as base width modulation effect.

(Refer Time Slide: 12:43)




So this is known as and I had explained to you what is known as base width modulation effect again a second order effect in a BJT. If you have a base which is something like this , right then if you look very carefully if it is a forward active mode I am assuming this to be as NPN and this is emitter, base, collector then we bias it such that this emitter base junction to be forward biased and this base collector junction to be as a reverse biased so this is reverse biased. So I defined let it suppose $V_{BC}$ and $V_{BE}$.

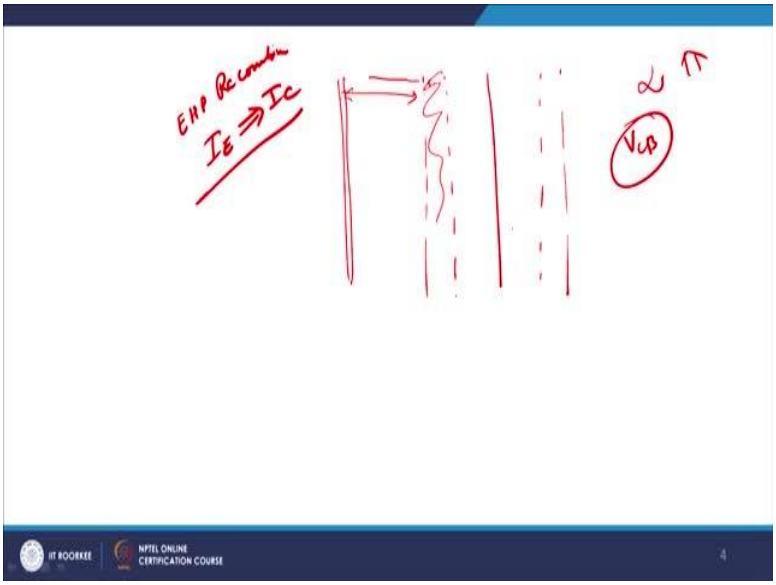
Now you see because of $V_{BE}$ there will be a thin depletion region here but since it is forward biased it will be very thin in nature dimensions. Whereas because of $V_{BC}$ I will expect to see a larger depletion width here. Fine is it okay? So I will see a larger so if this is say $X_1$ right and this $X_2$ then obviously $X_2$ will be larger than $X_1$ because this is reverse biasing because your base collector is reversed biased and your emitter base is basically forward biased fine.

With this knowledge let us do one thing let us fix up $V_{BE}$, right and increase the value of $V_{BC}$ we go on increasing the value of $V_{BC}$, as you go on increasing the value of $V_{BC}$ what will happen? That this depletion width will go on increasing because you know very well that in a PN junction diode or in a simple PN junction if we increase the reverse bias the depletion width becomes larger and larger. So what will happen is the depletion width will become larger and larger, right.

So $X_2$ will become still further more and more, when $X_2$ becomes more and more you could easily appreciate that the effective base width reduces, effective base width reduces fine. As it reduces now you allow a larger amount of current to flow through it because now the number of minority current carrier's there or the number of holes there is still smaller. Is this concept clear therefore? I will explain once again what I was trying to say because this concept will be important to understand why the collector current is not behaving as an ideal current source.

What will happen therefore is that as you increase the value of $V_{BC}$ base to collector voltage since it is a reverse biased base to collector voltage as you make the base to collector voltage go higher and higher, the depletion width on the base to collector region the depletion width will also increase as a result what will happen is that the depletion width on the base to collector junction will eat away into the active area of the base.

(Refer Time Slide: 16:12)



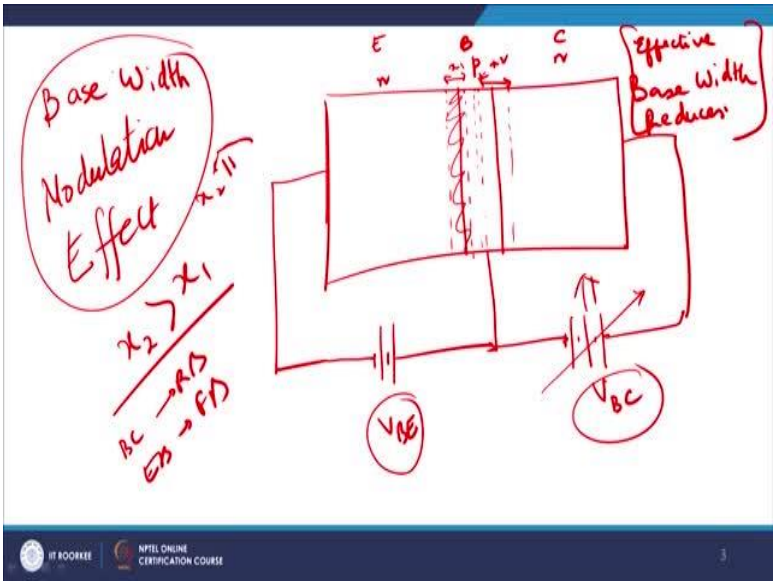

So your base was something like this now initially your collector base junction was reverse biased like this. Now increase the bias and this has shifted this line has shifted to this part and it has become something like this and this is something like this. So which means that this much amount of base is gone therefore and the effective base width is reduced to this much amount only, right.

As a result the EHP electron hole pair recombination will also be reduced right and more of $I_E$ will be allowed to go to $I_C$. And therefore your α value will increase and as a result the current will not remain constant but current will change with respect to $V_{CB}$. Please understand this point very carefully. Till now we were all explaining that as $V_{CB}$ becomes larger and larger or as $V_{CB}$ is increasing as long as it is reverse biased collector base junction is reverse biased independent of its value the current will always remain constant and what we were therefore doing was that we were actually assuming therefore it is behaving like a constant current source. The collector is behaving as a constant current source and if you remember the last talk last part you will see that we had removed the collector and replaced it by constant current source, right. $I_C$ was equal to α times $I_E$ and we have shown that this is the constant current source.

But in reality you will see that that depending on the value of $V_{CB}$, $I_C$ will change so higher the value of $V_{CB}$ I would expect to see a higher change. Not because of the large electric field but because of our reduction in the effective base width which effectively means recombination of electron hole pair will be minimised and therefore more of emitter current

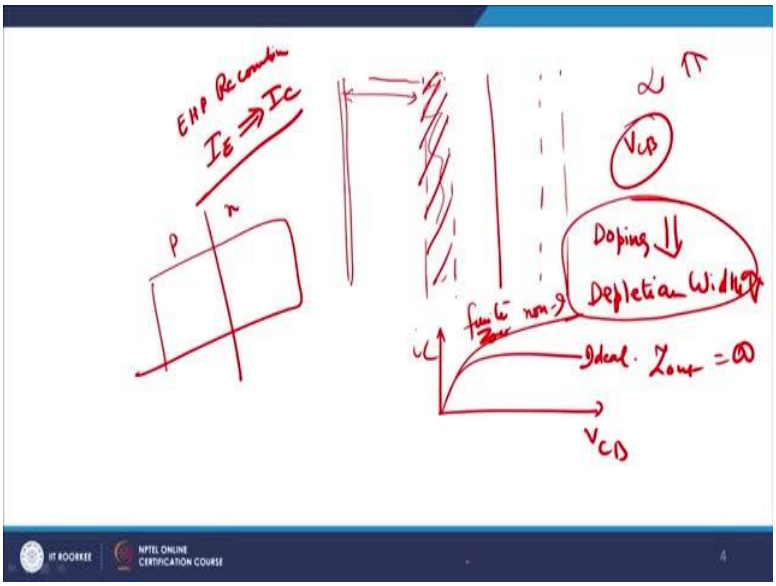
will reach towards the collector side and at the collector side they will register a current, right.

(Refer Time Slide: 18:16)




And that is what is known as is also known as Base width modulation effect, right. So we do a small change in our understanding and say that no in second order effect if base width is very very small and we are varying our $V_{CB}$ to a larger extent we would expect to see a larger change in the effective base width which will result in the large change in the output current, right. So this is what we have learned and we have understood as far as this course is concerned.

(Refer Time Slide: 18:48)

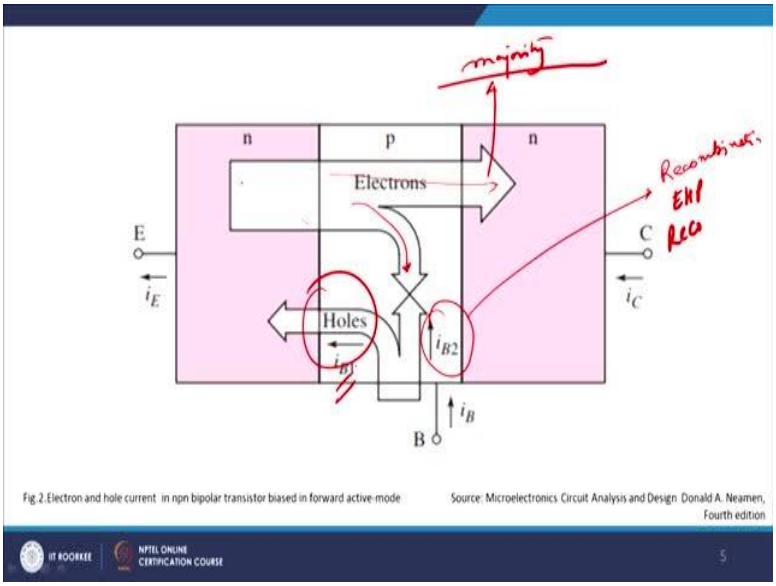


One more thing which I wanted to point out and may be of interest to most of you is that if you remember that if I have a P and N, right and if my doping concentration is low then the depletion width is large on that side so just keep this in mind that if the doping is low right and you apply a bias in a PN junction then the depletion width depletion width will be larger in that semiconductor in which the doping is low fine. So please be careful that the lower the doping more will be the depletion width.

So you see this also adds to the effective base width modulation effect why? Because intrinsically we have discussed the point that the base will have a much-much lower doping right so it is very-very low doped. When it is very-very low doped you automatically end up a larger depletion width on the base side, right. So the variation on the base side will be much larger as compared to the collector side and as a result your emitter base junction or the effective base width will be further minimised in the case of this one, right. As a result you will have increase in current. So if you plot collector current versus $V_{CB}$ ideally I should get something like this but if you see it will be something like this.

So this is the ideal current source whereas this is the non-ideal current source, right. Therefore this will have $Z_{out}$ equals to infinity but this will have finite $Z_{out}$, $Z_{out}$ is the impedance output impedance. So current source ideal current source will have infinite output impedance whereas a non-ideal current source will have a relatively lower infinite current source. So coming back to the therefore the basic issue once again we saw that for active base emitter is forward and the base collector was reverse biased.

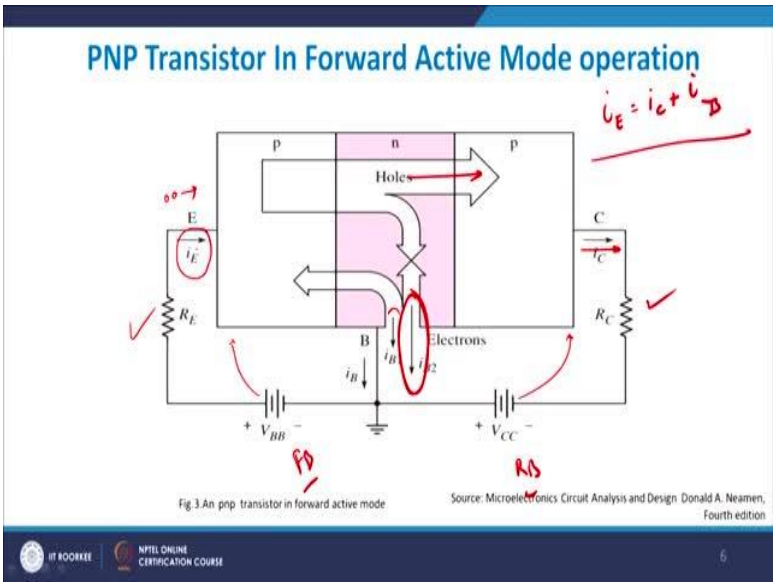
(Refer Time Slide: 20:51)



We again came back to the same position that if you have an NPN here this is basically the thicker arrow thicker arrow is basically my majority. So majority current carriers are inserted very fast and you get a large amount of electrons coming here, this is what is known as the recombination. So $I_{B2}$ is basically because of recombination between electron-hole so this is basically my recombination current right which is because of electron-hole pair getting recombined and therefore the current which is flowing is basically $I_{B2}$ because of that and please also understand that since this is forward biased holes can easily punch through this emitter base junction this one and will result in some current $I_{B1}$.

So there are two phases of current $I_{B1}$ and $I_{B2}$, $I_{B1}$ is primarily because of holes and $I_{B2}$ is primarily because of recombination of electron hole pair in the base region. We understood this point earlier also and we have also got the principle current. Let us see how we can look into we have look at the NPN, can we look into PNP as well which is basically a NPN and PNP in active forward active mode.

(Refer Time Slide: 22:02)

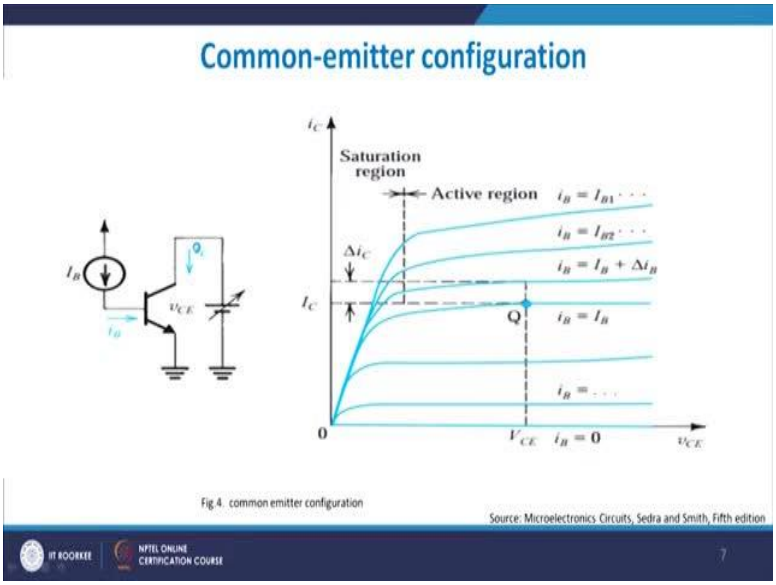



The same concept as in the previous case the only thing is the polarity of the battery will now be changed or reversed and you will have $V_{BB}$ in which the positive end is connected to P type and $V_{CC}$ the negative end is connected to P type. So this will be reversed bias and this will be forward biased so we are keeping the same concept exactly the same as in the previous case and trying to formulate a policy for finding out the current.

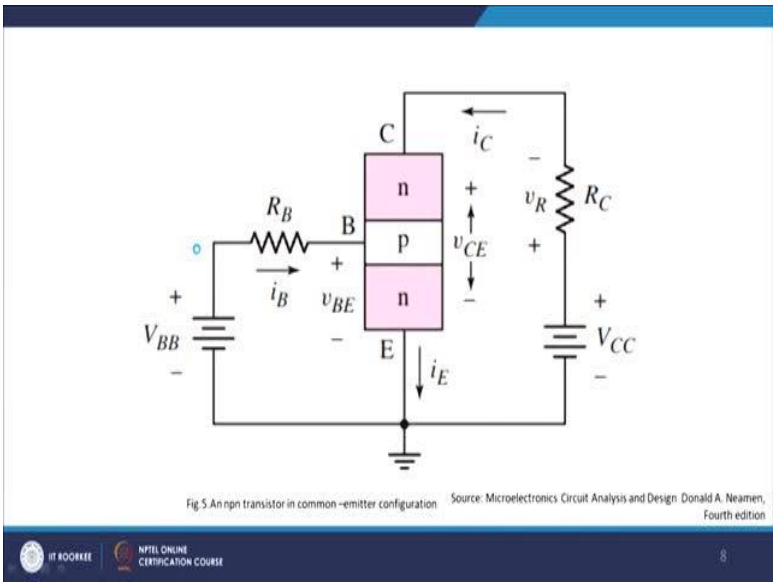
Here also $I_E$ will be equals to $I_C + I_B$ and this will be the only thing is because hole current is basically the conventional current therefore we see that the direction of the arrow head

actually shows me the direction of the holes flow of the holes. So this arrow head which you see here the arrow head which you see here this one is basically the direction of the hole, right. And therefore this shows the holes will be moving in this direction as a result the holes so holes and electrons will combine here electrons and holes will be combined to recruit that many number of electrons, electrons will be moving in this direction so holes will be coming out in this direction so you see $I_{B2}$ to be that electron and the rest of the electrons holes move in this direction and therefore $I_C$ is always in the outward direction. So $I_E$ is equal to $I_B + I_C$, $R_C$ $R_B$ are the resistances which you see in a typical operation.

(Refer Time Slide: 23:24)




(Refer Time Slide: 23:25)

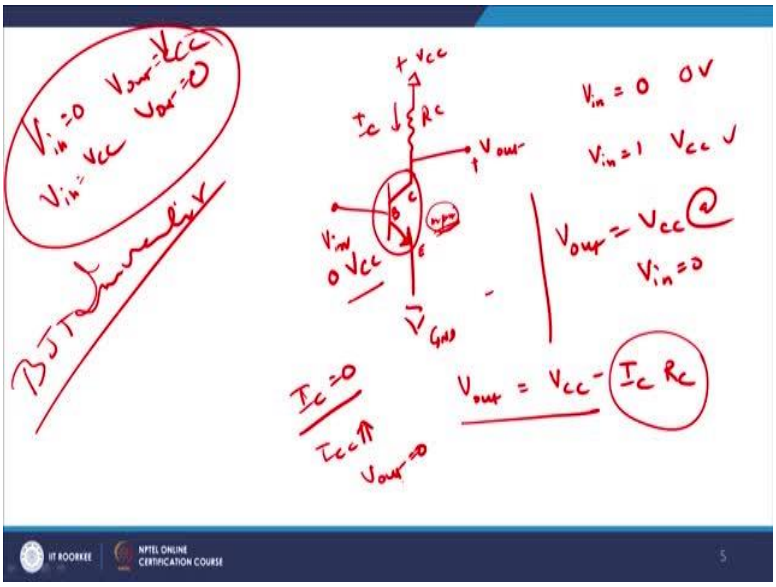


Now, let me explain to you maybe, just a minute, I will explain to you the common emitter configuration of the BJT because we will now enter into so now we will be seeing maybe in next two lectures or one lecture is that given therefore a transistor can we find out the current flowing through any of the nodes of a transistor provided we know the bias given to that particular node, right.

So what we do, we give a particular set of bias to the transistor any bias by giving forward or negative bias that means can we therefore take a transistor fit into all the four modes available to us and checkout how the currents will flow with respect to the variations in the biases in those transistors. So that we will learn in the next three to four slides in our understanding of this basic concept here.

We will be also one by one as we move along we will deal one by one in a detailed manner and we will see how it works as an inverter and all those things. Before we move forward therefore let me give you a very simple brief understanding of a BJT as an inverter.

(Refer Time Slide: 24:31)




That means if I have got a BJT here, a bipolar transistor, right and I apply a bias here and this is my $V_{CC}$ and this is my ground this is $R_C$ here and I take the output from here $V_{out}$ and $V_{out}$ to ground I take the output so this is positive negative and this is my input I give here. And this is my base and this is basically if you look at the arrowhead since arrow is pointing from base to base to so this is my emitter if you look very carefully this is my emitter, base, collector.

I am only doing it for the first time I will not be doing it again as we move along but just for your information say for the first time base to emitter the arrowhead is looking which means that the direction of the holes are in this direction which means the electrons are moving from emitter to base and therefore this is basically an NPN transistor right this is an NPN transistor which you see NPN transistor that is there with me.

And input side is always on the base side so you see most practically almost 60-70 percent of the case or 80 percent of the cases you give the signal or you give the input to the base side only, right. And try to take the output either from the collector or from the emitter in this case we are taking from the collector and let us see how it works out we will see one by one as you move along. Let us do one thing therefore that let me keep my $V_{in}$ to be so let us look at the digital part of it.

So if my $V_{in}$ is 0 or my V in is 1, 1 means let us suppose $V_{CC}$ or large value and 0 means 0 voltage fine this is the condition which I am maintaining. Now when my V in is equals to 0 then my base emitter junction is cut-off right and therefore no current is flowing no $I_E$ is flowing and therefore no current will be flowing through the circuitry and my $V_{out}$ will be therefore equals to $V_{CC}$ at when at $V_{in}$ is equal to 0 so when my $V_{in}$ is equal to 0 my $V_{out}$ equals to $V_{CC}$.

Now let me take $V_{in}$ equals to a very high value let me switch it ON which makes it which is transistor is ON this transistor when it becomes ON this potential will go to ground and therefore when my $V_{in}$ equals to $V_{CC}$ my $V_{out}$ equals to 0. And therefore this is how inverter BJT inverter is designed. Very simple straight forward from very first principles of circuit theory. We can predict the value of the output voltage with respect to the input voltage which you see and very straight forward way of looking at it. Also if you want the mathematical derivation I can write $V_{out}$ to be equals to $V_{CC}$ - $I_C$ into $R_C$, $I_C$ is the current flowing through this domain $I_C$ into $R_C$. Now if $I_C$ equals to 0 right I get $V_{out}$ equals to $V_{CC}$ and if your $I_C$ is very-very large quantity then this will be very as close to $V_{CC}$ and therefore my $V_{out}$ will be equals to 0 fine.

And therefore I can safely say that depending on the value of the collector current my structure my device which is NPN transistor can be switched on or off. So basic switch nothing very extraordinary difficult. Simply a basic switch which will switch from on state to off state and vice versa as I shown it here in this diagram. In the next class what we have learned therefore is the common BJT in a forward and reverse mode we have also looked into

the concept of base width modulation effect and we also looked into BJT as an inverter we will go into details if required in the later stages.

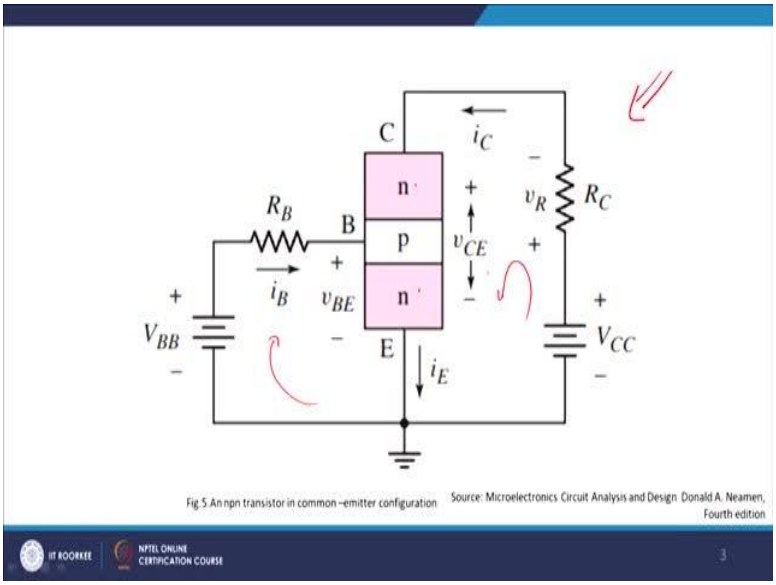
And in the next lecture we will be looking into BJT common emitter configuration working principles. With that we finish this lecture thank you very much!

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−05**
**BJT: Operation in Active Mode Circuit Symbol and Conventions - II**

Hello everybody and welcome to the NPTEL online certification on Microelectronics: Devices to Circuits. We continue with the lecture which we had done in the previous turn and that is Operation in active mode. We will look at circuit symbols and certain conventions. In our previous discussion or interactions we had seen that we were able to understand the basic fundamental principles for functioning of a BJT.

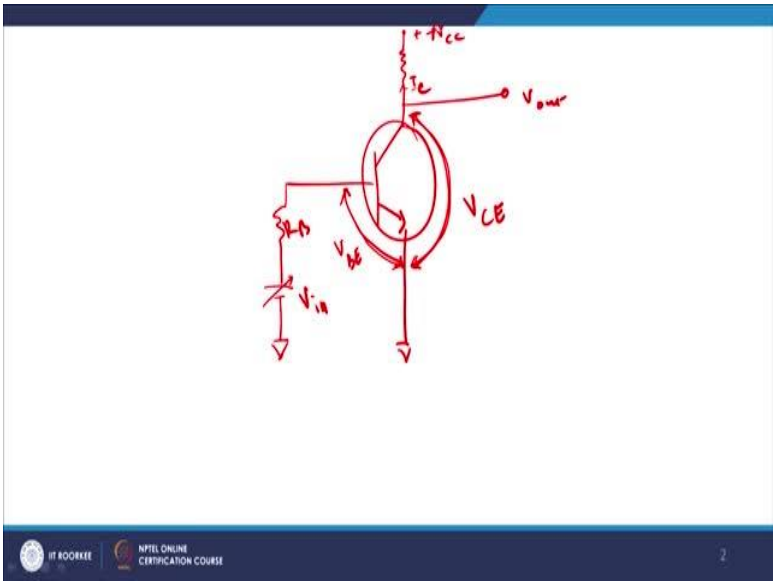
We have also seen how a BJT works in a forward active mode. What we will be looking today is basically the common emitter configuration. What we will be looking today is basically common emitter configuration of a BJT and also known as emitter grounded configuration or common emitter configuration, right and what we do in this case is that if you and that is what you can see on this diagram here on this diagram here.

(Refer Time Slide: 1:18)




I will make it a much better diagram from the point of view of understanding which is this one, right. So I have an NPN transistor, I have a BJT and I have my base emitter junction forward biased, right and my collector base junction to be as reverse biased this one right.

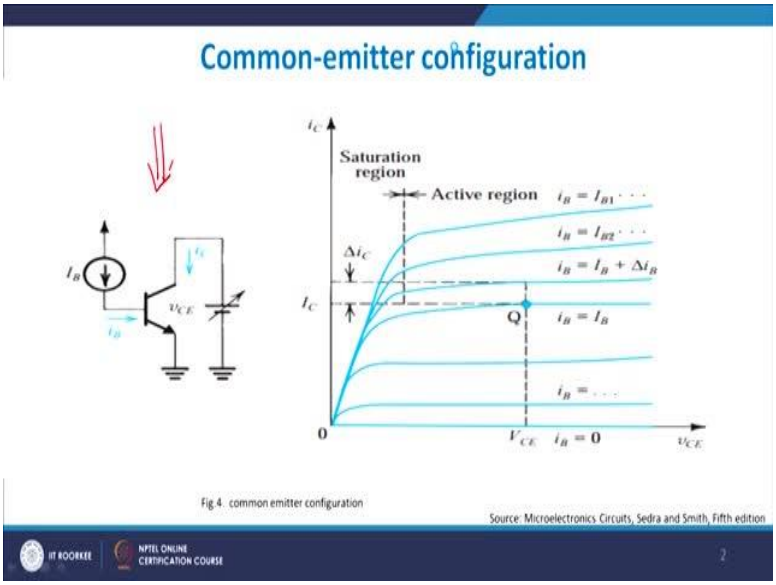
(Refer Time Slide: 1:37)




So let me just draw for you from the circuit point of view, right and let us see a common emitter configuration works in a sense in this case. So I have an RC here I have a base resistance $R_B$ here and since this is NPN so I can just do like this, right and we can vary this one to a larger extent and maybe I can reverse maybe I can fix the value of voltage $+V_{CC}$ here. And we will try to take the value of $V_{out}$ from this place particular place, right.

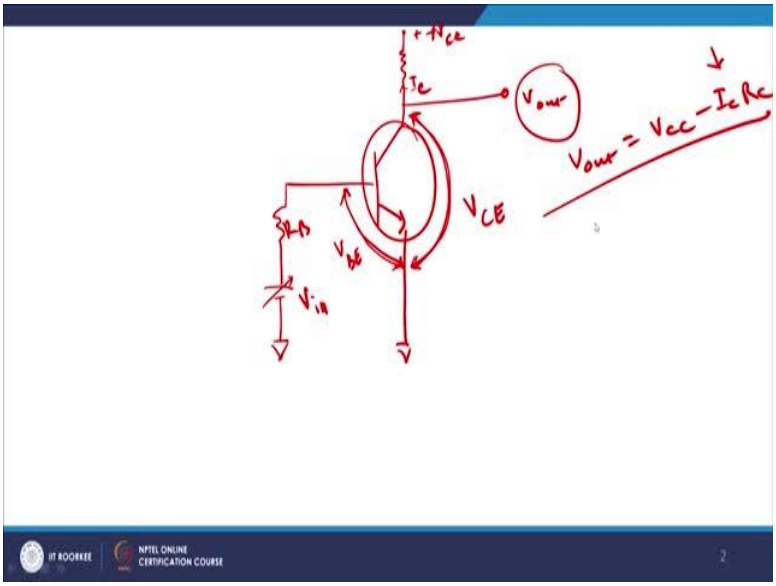
And this is $R_B$ and this is my $V_{in}$ and it is varying and I want to take out the output from this end and this is the collector current which is flowing here $I_C$. Now if you be so this is my $V_{BE}$ base to emitter equals to input voltage and this is also referred to as $V_{CE}$ collector to emitter voltage, right. So I have got $V_{CE}$ here I have got $V_{BE}$ here and I have got + $V_{CC}$ which is ruling here.

(Refer Time Slide: 2:47)



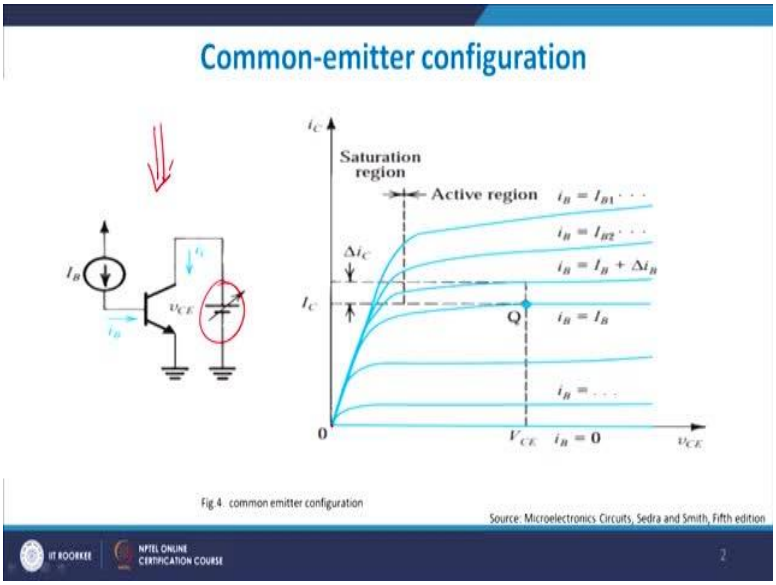

If you go back to the previous slide therefore you see that if you plot $I_C$ versus $V_{CE}$.

(Refer Time Slide: 2:56)



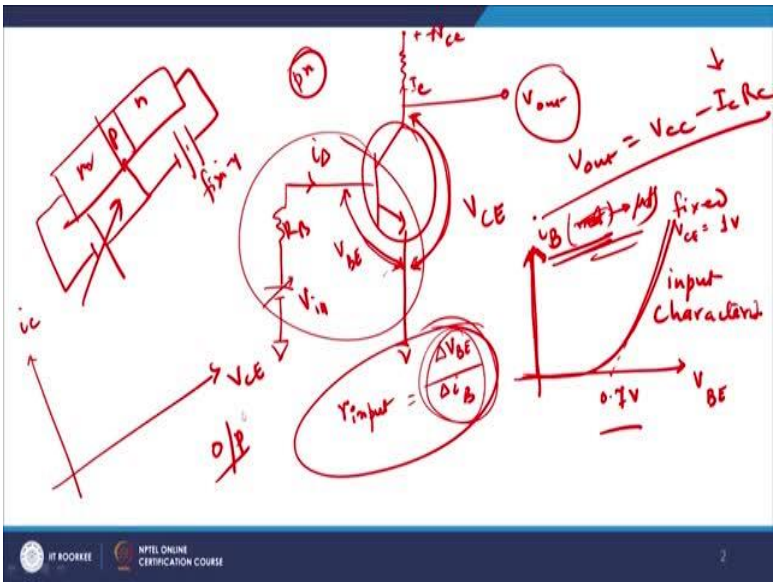

So let me plot IC which is the collector current versus $V_{CE}$ which is this potential here right. This potential here is nothing but the value of $V_{out}$ if you look very carefully this is nothing but the $V_{out}$ which you see from here. So $V_{out}$ if you find out will be equals to $V_{CC}$ - $I_C$ $R_C$. So if you can find out how $I_C$ varies with $V_{CE}$, I can predict how my $V_{out}$ will vary with $V_{CC}$. And this is straightforward linear relationship is there which is available with us.

(Refer Time Slide: 3:25)




Now if you look very carefully here, when $V_{CE}$ was equals to 0, right I had 0 current which is expected also the collector current was 0. And if $V_{CE}$ is equals to 0 the collector to emitter voltage is 0 then I would expect to see 0 current available to me.

(Refer Time Slide: 3:48)




Quite interestingly let me also show to you if I want to plot $I_B$ versus $V_{BE}$ versus $I_B$, this is also known as input characteristics. So base current $I_B$ versus $V_{BE}$, right for a fixed value of $V_{CE}$, right. So my $V_{CE}$ is fixed $V_{CE}$ I vary $V_{BE}$ and try to take the value of $I_B$, right. So if you look very carefully for a fixed value of $V_{CE}$ this starts so this starts to behave like a PN junction diode so in a forward bias condition.

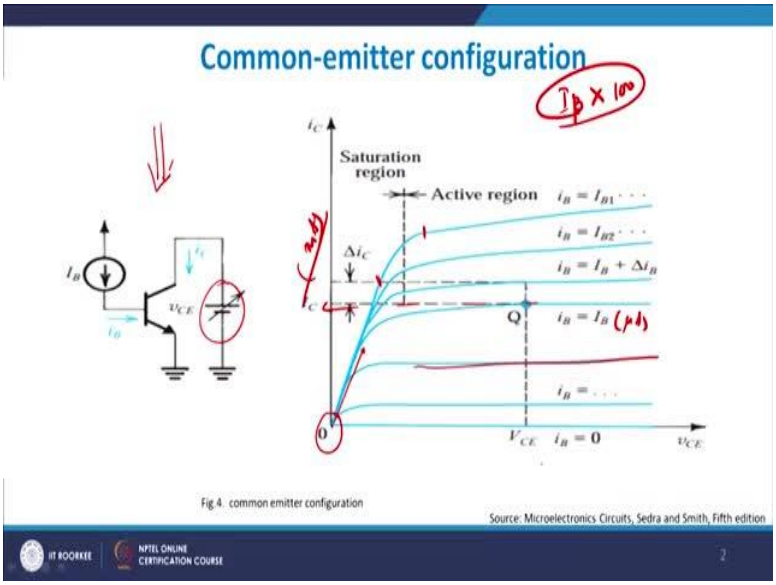
So forward bias condition you very well know that it looks something like this for a fixed value of $V_C$ so $V_{CE}$ is equal to say 1 volt or 5 volts or whatever. And if you revert it backward it will be approximately equal to 0.7 volts which you see, right. And this will be your current which will be flowing because that is how the current in a PN junction diode forward biased characteristics looks like, right. If you go on varying the value of $V_{CE}$ make the $V_{CE}$ larger I would expect to see a larger current to flow in this case, right.

$V_{CE}$ larger means large current will start to flow with the same value of so it will not be like this it will be something like this, right. It will be something like this which you will be seeing, fine. So this results in an understanding of input characteristics. Now if you want to find out the forward resistance of the PN junction diode then resistance will be $\partial V_{BE}$ upon $\partial I_B$ will be defined as input resistance of the $R_{input}$ of the NPN transistor.

So it means that the $V_{CE}$ is fixed so if you look very carefully my NPN this I am forward biasing it and this I am reverse biasing it, right. And I am fixing this one but now I am varying this one varying this one and I will get a profile something like this I will get a profile. As a result what will happen is that this profile will lead to a fact that this this implies that that if I want to find out the resistance offered $\partial V_{BE}/\partial I_B$ and it will be order of few ohms because is an on state resistance of the diode and it will be of the order of few ohms which is there this current will be of the order of few milliamperes to microamperes range of current will be there.

Very small current will be there microamperes range current will be there and that is what is what you get from this. So this is known as the input characteristics of the bipolar transistor right input characteristics of a transistor. Now if you want to plot therefore $I_C$ versus $V_{CE}$ also known as the output characteristic which is already plotted here in this region.

(Refer Time Slide: 6:45)




And I can explain to you what it happens. So for a fixed value of $I_B$ your $I_B$ is fix how $I_B$ is fixed because your $V_{BE}$ is fixed. So what we have done now is that rather than changing $V_{CE}$ or $V_{CB}$ we are changing we are fixing $V_{BE}$ and therefore my $I_B$ is fixed when my $I_B$ is fixed I vary $V_{CE}$ and try to get the value of $I_C$. So I try to get the value of try to get the variation of collector current with respect to collector to base voltage or voltage which is available with me.

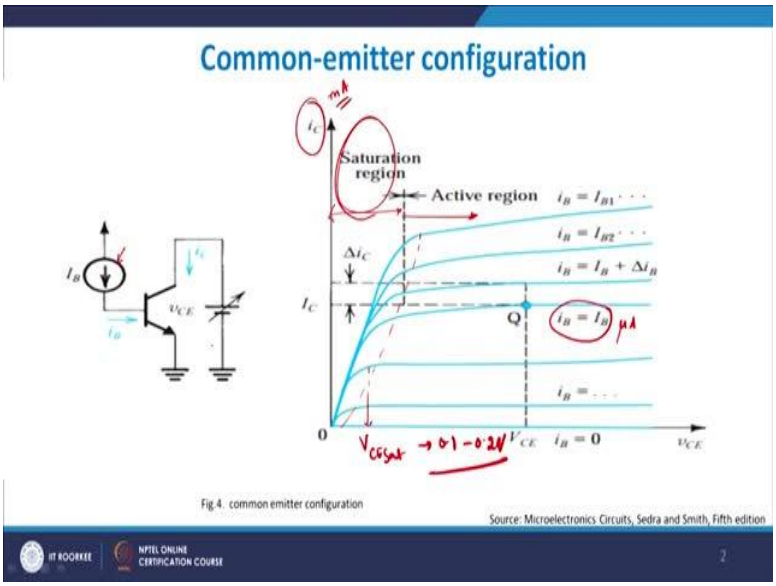
Now when my therefore $V_{CE}$ equals to 0, right my collector current will also be equal to 0. And the reason is that when my $V_{CE}$ is equal to 0 there is no electric field on the collector side to accept the electrons which are coming from the emitter side. So even if there is a base current or even if there is a emitter current available all the emitter current is either going to the base because you do not have any potential on the collector side to accept the charge carriers and therefore the current is 0 and that is the reason the current is 0 for $V_{CE}$ equals to 0.

As you start to increase your $V_{CE}$ the current starts to rise and therefore you see a linear increase in the value of current, right there is a linear increase in the current. As you make your $V_{CE}$ larger and larger there is a linear region and then you enter into a non-linear region of operation of the $V_{CE}$ current increases because more and more charge carriers from the emitter side passing through the base side is collected by the collector. And a time comes when you will approximately get a straight line here for a fixed value of $I_B$.

Currently meaning that it starts to behave like an ideal current source and therefore the collector current is almost independent of value of voltage across the collector to emitter voltage and it has got a fixed value with respect to the output current. So what we do is now if we go on increasing the current $I_B$ current by changing the value of $V_{BE}$ of course the collector current will also change, right. And the collector current will change by how much. So this will be in microamperes, right $I_B$ whereas this will be in milliamperes the collector current will be in milliamperes because you multiply $I_B$ with a factor known as which is approximately equals to the 100 multiply.

Why do you multiply because the simple reason that your emitter current is equal to your collector current, right? Your base current is just very small because of recombination. So whatever is left is basically collector current and therefore the collector current $I_C$ is of the order of few milliamps where the base current is of the order of few microamps. Now if you look very carefully the region which you see in front of you.

(Refer Time Slide: 9:42)

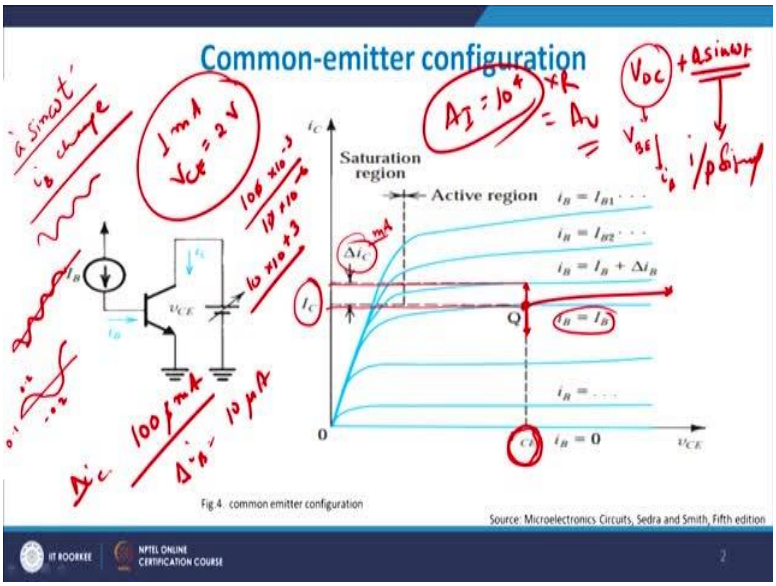



I will just show it to you by discussion here, this region from 0 to approximately this region $V_{CE\ sat}$ this is also referred to as line this is known as this point where the transistor actually saturates is known as $V_{CE\ sat}$. It is typically 0.1 to 0.2 volts is the maximum value of voltage at which the saturation tendency comes into picture. And this region is actually known as the saturation region. This region starting from here to here and beyond that is defined as my active region. Beyond that is defined as my active beyond the $V_{CE\ sat}$ part. This is my active region for the NPN transistor forward active mode so this is the active mode is basically the region.

Now you see in a common emitter configurations let us suppose write in a common emitter configurations this is my emitter now for a fixed base current I get a fixed for example so we have understood two things therefore I have a common emitter configuration I have a saturation region I have an active region, right. Saturation region is the region where the output current will be a linear or non-linear variation depending upon the value of $V_{CE}$.

$V_{CE\ sat}$ is the saturated collector to emitter voltage beyond which the current will be independent of $V_{CE}$ behaving like a constant current source approximately 0.12 volts will be the approximate value of the current technology. Also if you look very carefully that you have this $I_B$ value of the order of few microamperes whereas this collector current is of the order of few milliamperes. And that you need to find out or you need to know this basic fundamental principle.

(Refer Time Slide: 11:28)




Now let us look at one small point before we move forward that in fact evenly, even without doing any drastic mathematical formulation I can show to you that we are actually indeed doing a small amount of amplification here. And how the amplification is done I can just show it to you roughly by this diagram only.

You see you generally apply the signal to the base side right. Now for achieving amplification in a common emitter configuration for any configuration. The device should be initially biased in the active region of the operation of the device. If it is not, your amplification will be a non-linear amplification, right. In the emitter in the active region only

you will automatically get a very linear $A_V$. So $A_V$ or the voltage gain will not be a function of the input voltage or the output voltage.

It will be fixed if you bias your device in active region therefore what you generally, what generally people do is they apply first of all a DC bias. A DC $V_{BE}$ and a DC $V_{CE}$ such that the device is biased in this Q point this Q so this is the value of $V_{CE}$ and this is the value of $I_C$. So I know the value of $I_C$ as 1 milliamp and $V_{CE}$ as say 2 volts, right. I just take up a circuit. This value of $V_{CE}$ and this value of 1 milliamp we get and then we bias it by an external DC source.

Now this DC source will not change across the experiment it will still remain the same it will just help you to bias the device in an region or place the device in a region where amplification is possible and not only possible but it will give you a linear amplification independent of the input voltage. Once you have fixed the DC bias super imposed on that you will have to give your AC bias, right. So what you do is you first insert a DC bias $V_{DC}$ and then add to it *ASinwT*. This is your input signal fine.

So what will happen I will tell you what will happen? So I have a $V_{DC}$ because of $V_{DC}$ which is basically $V_{BE}$, right I will have a fixed value of $I_B$. So let us suppose this is the value of $I_B$ which I am getting. This is the value of $I_B$ which I am inserting and I am fixing the value of here. Corresponding to this my $V_{CE}$ value is already fixed at say 1 volt or 2 volt. Now what I do now I give a signal which is input signal which is *ASinwT* let us suppose.

As a result my current IB, right will also change sinusoidally it will change sinusoidally, right. So I have a DC bias and there will be an AC bias overriding over it which is basically my sinusoidal signal input signal so I will have a sinusoidal input signal. Which means that the current will once it will go from nominal value to a high value and then come down go to the bare minimum low value then again it goes up and again it comes down, goes up, comes down so on and so forth.

Which means that $I_B$ so if your nominal value was say suppose 0.1 and your peak to peak variation is say 0.4 then say 0.1 it goes to 0.2 it again goes to 0.1 and again comes to minus 0.2 let us suppose and then goes up. Let us suppose this is how the variation is. So the peak to peak variation is plus 0.4 and so on and so forth. Which means that this Q point by virtue of the signal being inserted into the base to emitter junction of the NPN common emitter

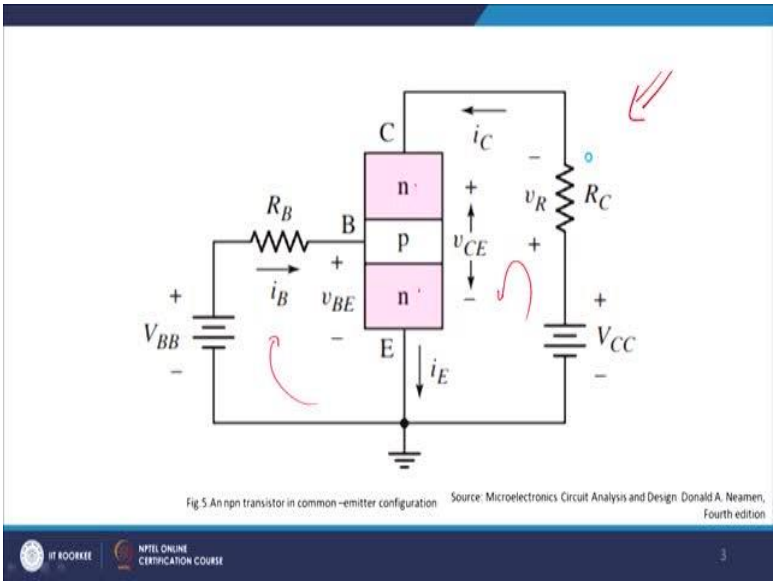
configurations this Q point will shift by 0.2 volts at the top and then go minus 0.2 volt at the bottom in current sorry.

So the base current will go so the Q point will go top and then it will again go at the bottom again it will go at top again it will go at the bottom by application of a principle. But please understand this will give rise to a change in the current which is so much $\partial$ $I_C$ and this is in milliamperes because we just now learnt it. Which means that a microampere change in the current results in a millampere change in the output current.

So the input current changes by let us say 10 micro amp and the output current changes by say 100 millamp. So $\partial$ $I_B$ is this much and $\partial$ $I_C$ is this much so if I want to find out the current gain I will get 100 into $10^{-3}$ divided by 10 into $10^{-6}$. So this cancels out 10 into $10^{+3}$, right.

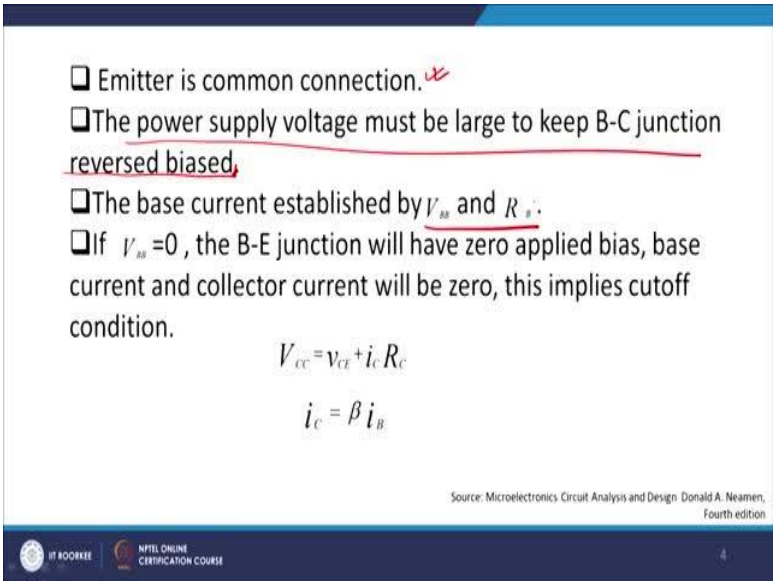
So automatically I get a current gain $A_I$ current to be equals to approximated to the power four time increase in current which means that for a small change of approximately 10 microampere peak to peak change if I get a 100 millampere peak to peak change I get 10 to the power four times $A_I$. This if you multiply with an external resistance R is basically my voltage gain or voltage change which you see fine. So this is what you get as a common emitter mode configuration basic amplification which you see as far as this basic understanding is concerned right. So this is the common emitter configuration as far as the previous discussion which we had as far as concern.

(Refer Time Slide: 17:14)



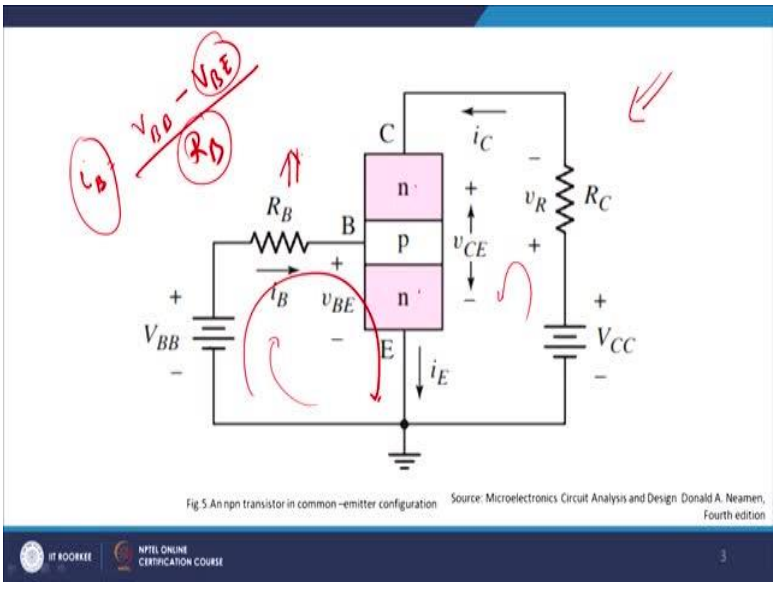

This is the active common emitter mode configuration active source forward active mode configuration with the schematic shown in this figure.

(Refer Time Slide: 17:27)




So as I was discussing with you therefore the emitter is common connection between the ground, right. Now I think this is clear to you that the power supply has to be must be large to keep the base collector junction reverse biased, right is it okay? Now the base current is established by $V_{BB}$ and $R_B$ why?

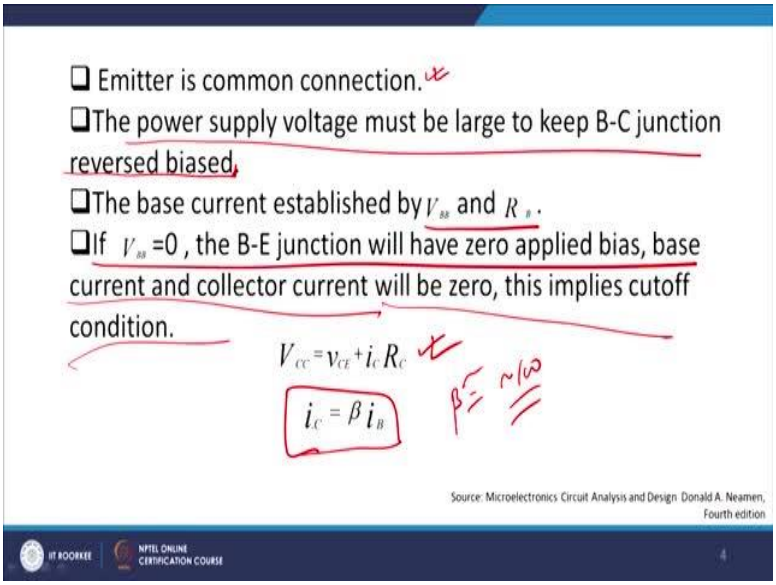
(Refer Time Slide: 17:47)



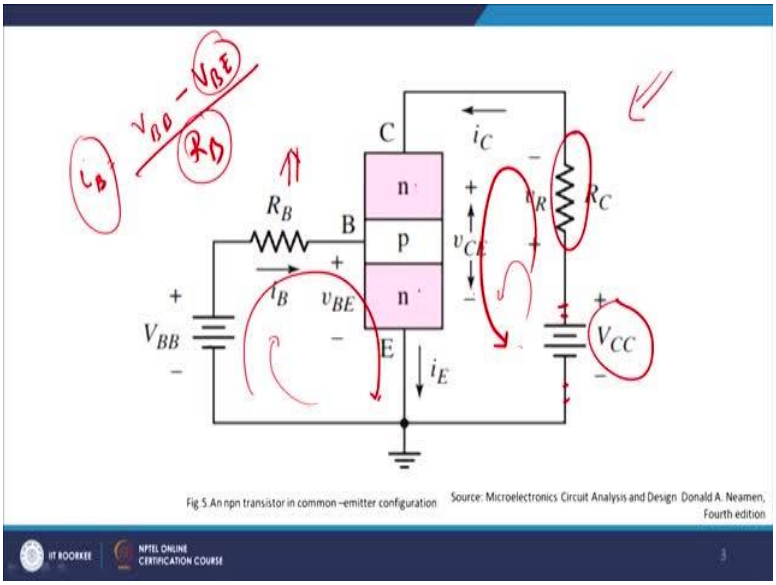

If you look at $V_{BB}$ this is my $V_{BB}$ and this is my $R_B$ right. So I get $I_B$ to be equals to $V_{BB}$, right, if you go from this side to this side $V_{BB}$ - $V_{BE}$ right base to emitter junction divided by $R_B$. This is 0.7 this has to be of the order of few kilo-ohms for this to be in the order to few

microamp. So this has to be of the order of few large value of voltage has to be $R_B$ has to be typically very large otherwise the base will short out.

(Refer Time Slide: 18:22)




(Refer Time Slide: 19:05)

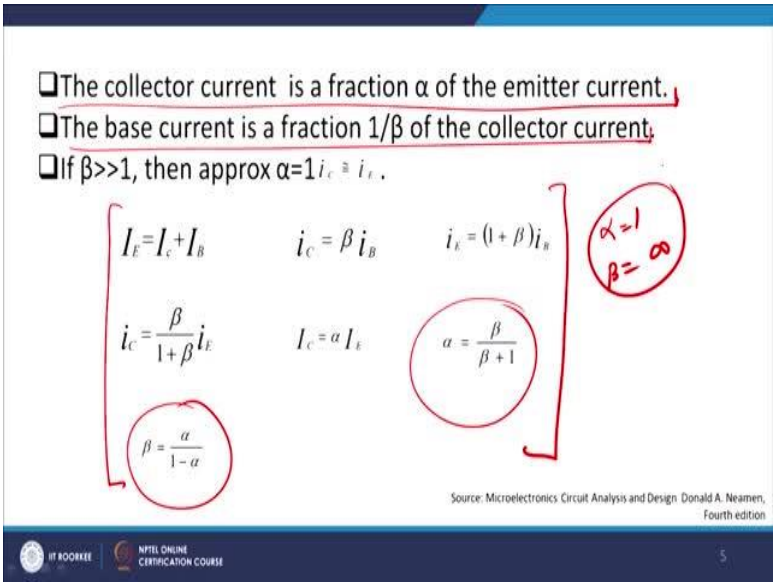



So what we get from here is that yes if $V_{BB}$ is equals to 0 the base emitter junction will have 0 applied bias right. Why it will have 0 applied bias because with $V_{BB}$ equals to 0 effectively your base emitter current if I am assuming it to be ideal one is 0. So 0 minus 0 is 0 and I get no current available to us and therefore your base current and the collector current will all be 0.

And this implies in cut-off condition so which means that if your $V_{BE}$ or $V_{BB}$ is 0 you are not able switch on the transistor and therefore you reached to cut-off condition. In that case we write down $V_{CC}$, $V_{CC}$ is what may be I use the previous slide. $V_{CC}$ is this one $V_{CC}$ is equals to $V_{CE}$ + $I_C$ $R_C$ we have discussed this point. $V_{CE}$ so if you go by this loop $V_{CC}$ is this is equals to $I_C$ $R_C$ + $V_{CE}$. $I_C$ $R_C$ is the voltage drop here + $V_{CE}$ and that is what is given by this value.

And we know that $I_C$ is equals to β times $I_B$ right β is of the order of few hundreds of the order of few hundred only and therefore I get $I_C$ equals to β times $I_B$. And therefore $I_B$ is of the order of few microamperes I would expect to see $I_C$ of the order of few millamperes in this case.

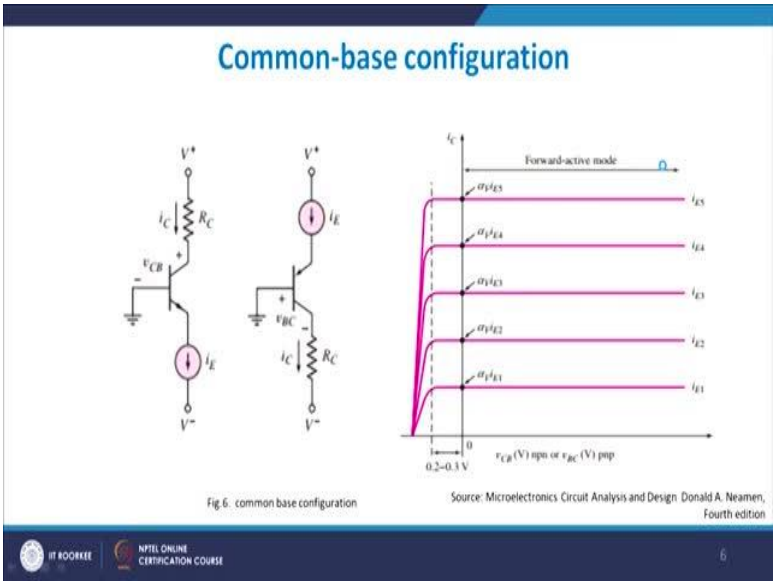
(Refer Time Slide: 19:43)




Right and then I was discussing the collector current is therefore is a fraction α of the emitter current. And for α is very close to 1 so if emitter current is 1 millamp you will expect to see collector current very close to 1 millamp because α is very close to 1, right. And the base current is 1 by β of the collector current. This we have already seen just now in this case and these are the few important parameters found in the book also.

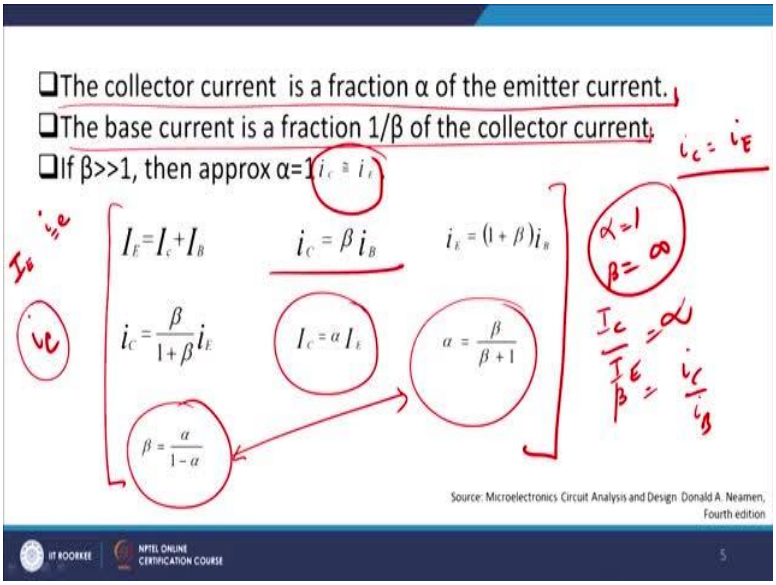
Out of which these two are quite important parameters. I will not do a formal derivation here. I will leave an exercise to you that α is equals to β upon β + 1 and β equals to α upon 1 - α. And therefore for α equals to 1 I get β equals to infinity and β equals to infinity. So infinity it will be infinitely high value of β you will get.

(Refer Time Slide: 20:40)




And this is what a typical configuration.

(Refer Time Slide: 20:44)




So let me therefore at this stage let me finish off with common emitter configuration understanding. When we meet next time we will start with common base and common collector and that will take care of the effective values here. So what we have learned this lecture is common emitter configuration module and how we can calculate the value of current and basic amplification in a common emitter configuration and we have also learned how I can act as a switch also.

So when β is much larger than 1 then I get approximately α equals to 1 and I get $I_C$ equals to $I_E$ which you can see from here. I get $I_C$ equals to $I_E$ and therefore the collector current is approximately equals to the emitter current. Now the very important certain very important

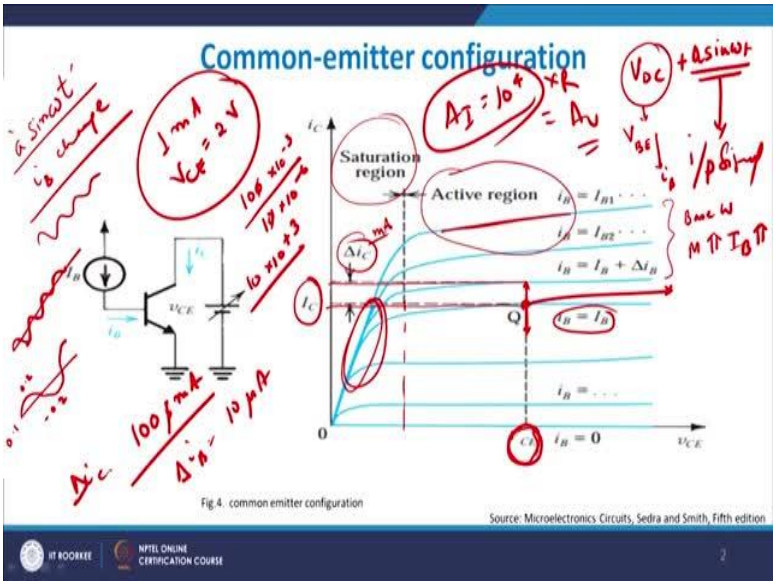
formulae's which is shown in front of in this slide is one is α is how related to β and how β and α are related to each other is like this.

And standard books which you will find you will get the methodology for how you got this α and β from basic Kirchoff's law. And we have also understood that $I_C$ is equals to α times $I_E$ and therefore $I_C$ by $I_E$ is basically equals to α which we have already seen we have seen β to be equals to $I_C$ by $I_B$ for all practical purposes.

So this is what we get from it. Just wanted to make one small important this thing observation before we move forward. That when we refer as capital I there is a difference between capital I and small I, right. Now capital I primarily refers to as DC current so whenever you have DC currents or DC biases we refer this to be as capital I, and small i is generally referred to the case when you have got AC currents.

Similarly if you have got subscripts as capital and subscript as small then when subscript is capital you are going to get DC bias and small means you are doing a transient analysis, right. So you should be very careful when you use these, so if I have got I and C which is this case it primarily means that I am using an AC analysis small signal AC analysis but I am trying to find out the total current flowing through the circuitry. So I have got therefore this notation is quite important in terms of understanding this basic principle.

(Refer Time Slide: 23:15)

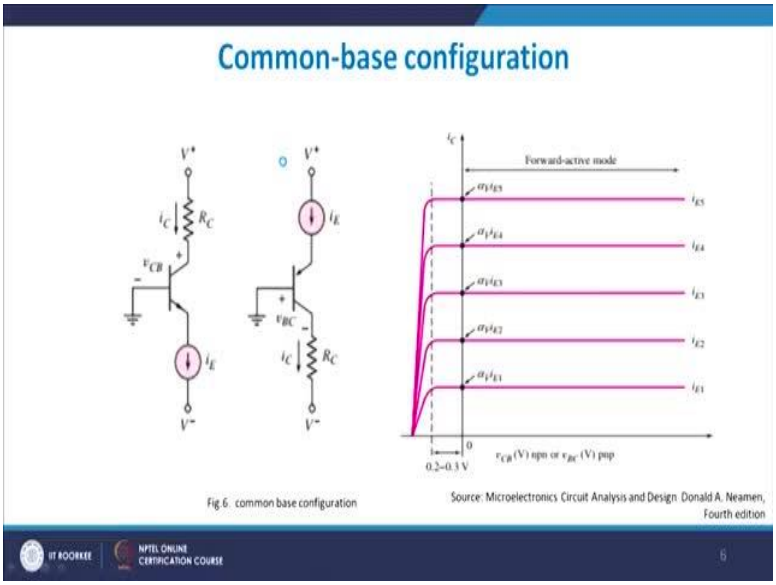



What we also learned in the previous slides or previous discussion is two important points that if you look at this graph once again if you see then at least for the higher values of $I_B$ your base width modulation is very-very large, right. So your base width modulation is quite large for large values of $I_B$. So when $I_B$ is low you generally do not see or even your $V_{BE}$ is low you do not see large amount of base width modulation. But you see a large amount of base width modulation when you have larger value of $I_B$ effectively with you.

The second thing which you can see is that in the saturation region, right, the resistance offered is typically relatively low whereas in the active region the resistance offered is relatively very high, right. Resistance offered by the device is very-very high. And that has to do of course with understanding that during the saturation region which is the just before the $V_{CE\ sat}$ we end up having the device is just going to the ON state.

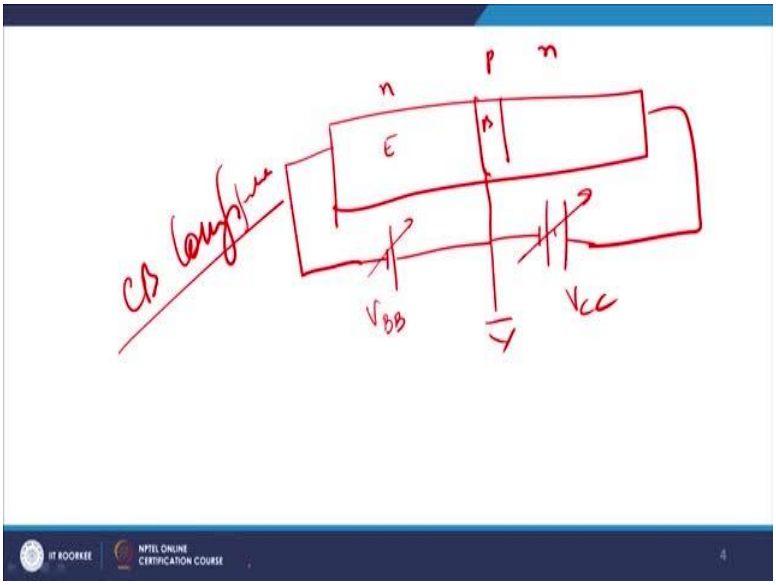
Please understand various books and various sources give you different regions some of the books also give you this region to be as the saturation region and this region to be as the active region. But we will follow the same trend for all our subsequent lectures. And we will assume that the active region is the region where you should bias the device in order to use it as an amplifier, right. And this is the active biasing which you should do in this case.

(Refer Time Slide: 24:49)




We come to the next important configuration and that is known as common base configurations. And in common base configurations maybe I will just show you here, maybe I can just show you here. Okay I do not have it in that sense but I can show you here in a common base configuration how it works out.

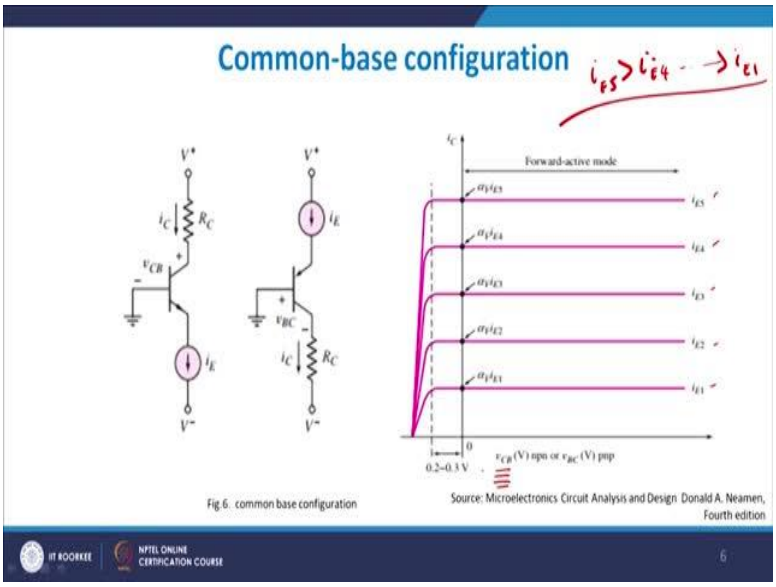
(Refer Time Slide: 24:59)




So common base means the base is common between the emitter and collector. So if I have an NPN and this I have been doing for quite often now that this emitter base will therefore be forward biased right and therefore this will be also this will be reverse biased in the sense. And therefore this will be this will be like this, right.

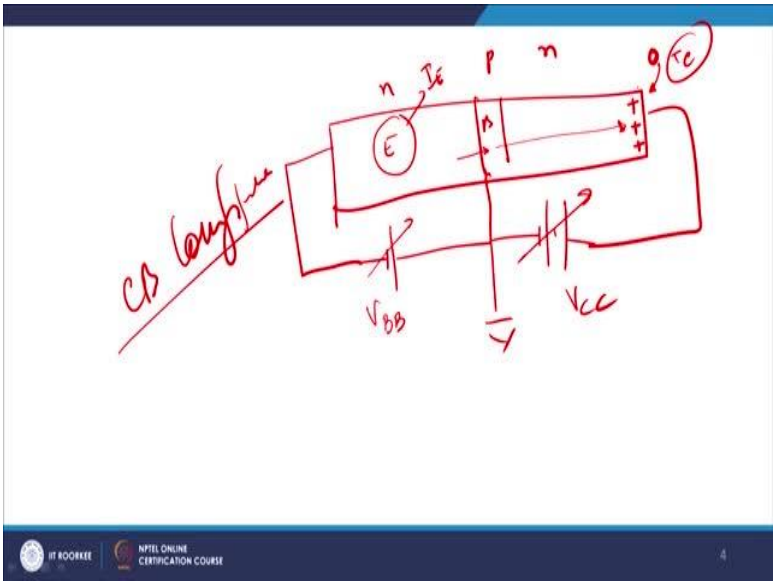
So this is the common base common base CB configuration common base configuration which you see. And as you can see we have already discussed this time and often. This is $V_{BB}$ let us suppose and this is $V_{CC}$ and you can vary both of them and so on and so forth. And you can get you can get large amount of characteristics out of it from this characteristics.

(Refer Time Slide: 25:47)

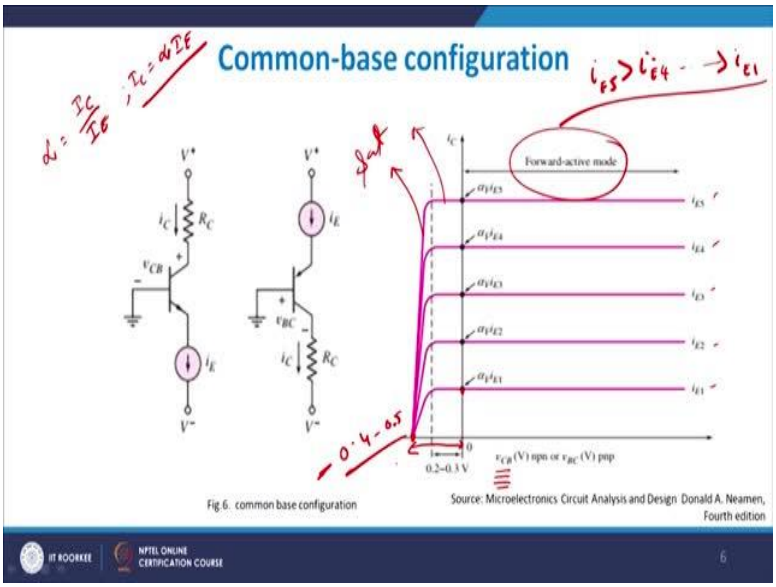



Common base configuration as the name suggests when we plot therefore for NPN $V_{CB}$ CB $V_{CB}$ collector to base versus $I_C$, right. Now you see quite interestingly even when I think all of you can understand this graph this is for a fixed emitter current so for varying emitter current, right $I_{E1}$ greater $I_{E1}$, $I_{E2}$, $I_{E3}$, $I_{E4}$, $I_{E5}$. Such that $I_{E5}$ is greater than $I_{E4}$ and so on and so forth and the last one is $I_{E1}$. With this condition or with this idea what we see is that as VCB is positive it primarily means, it is positive means what? It positive means what positive means I will show you it positive means this is positive.

(Refer Time Slide: 26:46)




(Refer Time Slide: 26:52)




So I have a biased positive so when it is 0 bias even when it is 0 biased. You will see that you still have large amount of current flowing through the device. And the reason being very simple that even at 0 bias you have applied such a large electric filed because of the $V_{BB}$ that electrons which were predominantly my emitter current $I_E$, right passed through the base region very fast and they cross through this region very fast and reach the collector side.

Though we did not applied any potential positive potential on the collector side on the $I_C$ side collector side but still current from the emitter side is reaching towards the collector side in a very fast pace manner. And the reason is that though a bias was not applied but the velocity is so strong that it can easily cross the base region and reach the collector side.

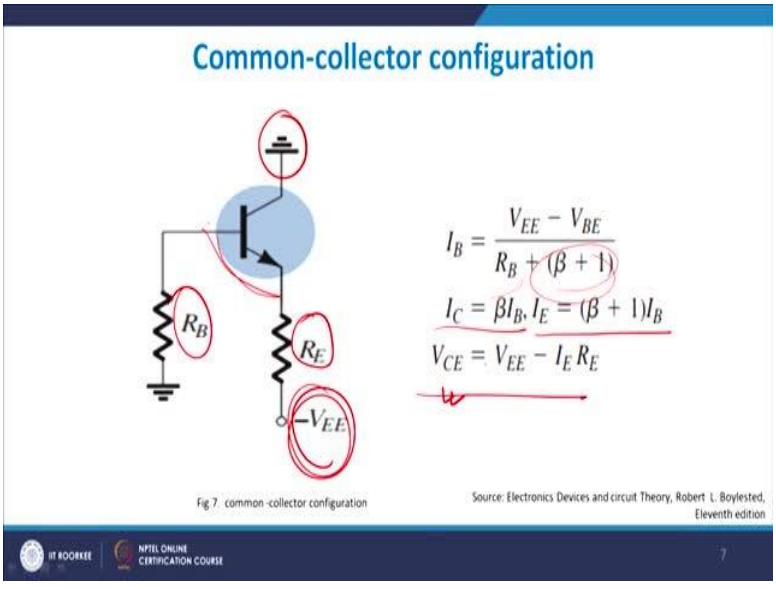
And therefore if you see very carefully that you will always have a current which is equal to α into $I_E$ because you remember α was equal to was defined as $I_C$ by $I_E$. So $I_C$ will be nothing but α times $I_E$. So I get α times $I_E$. Similarly $I_{E1}$, $I_{E2}$, $I_{E3}$, $I_{E4}$, $I_{E5}$. Now if you increase $V_{CB}$ it is almost constant independent that we have already discussed why is it like that.

But then it means that you have to make your $V_{CB}$ go in forward bias in order to stop the current flow and I think you can understand why. So you have to give a negative potential here in order to repel the electrons to move in this direction and not reach here. Is this concept clear? Which means that you minimum have to give this to be approximately equal to 0.4 to 0.5 with a negative sign so that it is in the cut-off mode.

So therefore this is known as forward active mode as I have written here. This is the forward active mode and this is the cut-off mode which you see. This is the cut-off mode, right. We define this to be as the saturation mode and this to be as the cut-off so you can do cut-off here. So this much amount of minimum amount of $V_{CB}$ has to be provided in order for the device to be in cut-off.

So if you want to switch ON the device you have to bias your device to go from VCB of - 0.5 volts to approximately 0 volts for our practical purposes. And this this though it has been shown in a much greater sense but the drop is very-very drastic it is a very fast drop which you see here. And it is almost a straight line curve for a common base configuration.

(Refer Time Slide: 29:28)




But what you will see here α is very close to 1 so you do not get any amplification out of common base configuration because the amount of emitter current that you are getting is

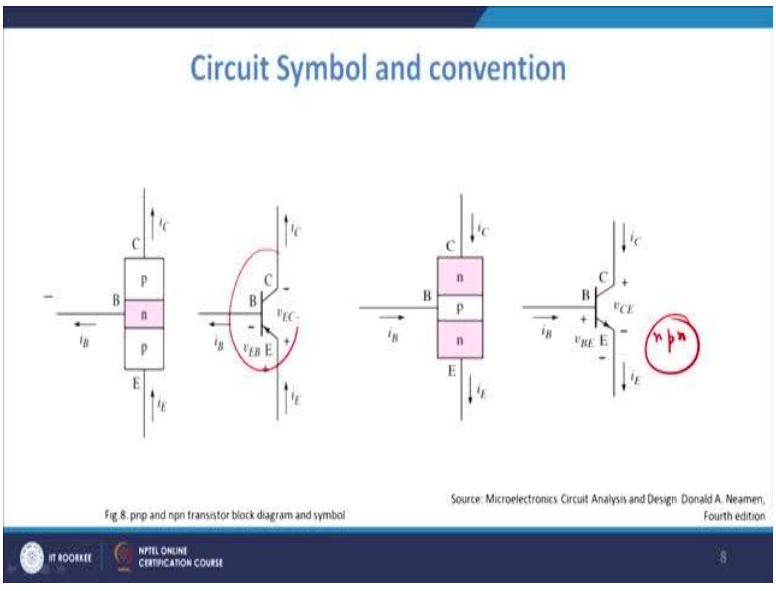
approximately same as the collector base configuration, right. And that is the basic idea of a common base collector common base configurations.

We do have a common collector configurations which is there with us. In which the collector is common or collector grounded configuration. In which the collector is grounded or it is common between the base and the emitter. $R_B$ is the base resistance and $R_E$ is the emitter resistance and we apply minus $V_{EE}$ here in order to bias the device into the active region of operation.

If you look very carefully here $I_B$ the base current will be equals to $V_{EE}$. $V_{EE}$ is this much minus $V_{BE}$ which is the voltage drop across this part divided by $R_B$ which is the resistance here plus β plus 1 which is β plus 1 is the effective value of which you see. $I_C$ equals to β times $I_B$ $I_E$ is equals to β plus 1 $I_B$ and $V_{CE}$ equals to $V_{EE}$ minus $I_E$ $R_E$. So this will give you but common collector is not very seldom used. It is generally used for impedance matching but generally not used for any other purposes. Most commonly used is common emitter configuration device and that is the most commonly used configuration here.

We end today's topic by doing two things. By seeing the circuit symbols we have already seen this. But arrow pointed outwards which is this one.

(Refer Time Slide: 30:58)

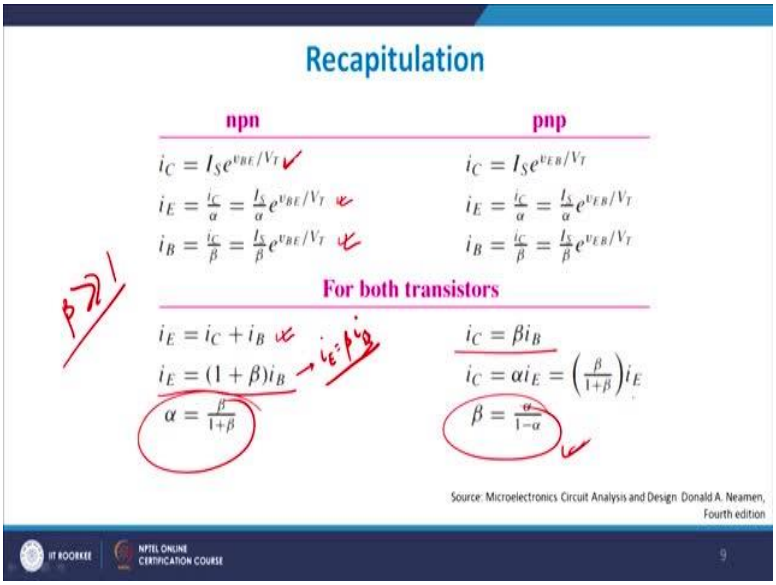



Primarily means it is the direction of the holes so holes move from base to emitter, which primarily means that the electrons are moving inwards and therefore this is basically an NPN transistor, right. And since it is a PNP transistor since holes are moving from emitter towards

base side the region is shown in this manner, right. So you will have to convention in the sign convention you just have to worry about the direction of the arrow, right.

The arrow head direction gives you the direction of the holes and that is what is shown in this diagram. Let me recapitulate for both the transistors and that will take care of approximately all the understanding of common emitter common base common collector.

(Refer Time Slide: 31:41)




If you look at the collector current it is given by this and for one you can see other we can always learn. $I_E$ is equals to $I_C$ by $\alpha$ I get this $I_B$ equals to $I_C$ by $\beta$ I get this for both the transistors I will get automatically $I_E$ equals to $I_C + I_B$ because Kirchhoff's current law cannot be violated. $I_C$ equals to $\beta$ times $I_B$ right. $I_E$ equals to actually equal to $1 + \beta$ but $\beta$ is since it is very-very large as compared to 1. I can approximate this as $I_E$ equals to $\beta$ times $I_B$, right.

$\alpha$ equals to $\beta$ upon $\beta + 1$ and $\beta$ equals to $\alpha$ upon $1 - \alpha$ and this is how you show the values of it. So which this let me thank you and give you an idea about what we did in the previous lecture series. That we have understood therefore the various configurations of working of a PNP and NPN transistors and which one is the best.

So the best option available to us is the common emitter for voltage amplification purposes. You can use common emitter as a switch as well, we can also common base as a switch for practical purposes. When we meet next time we will do certain other applications of BJT and we will do Ebers-Moll model for BJT. Thank you very much!

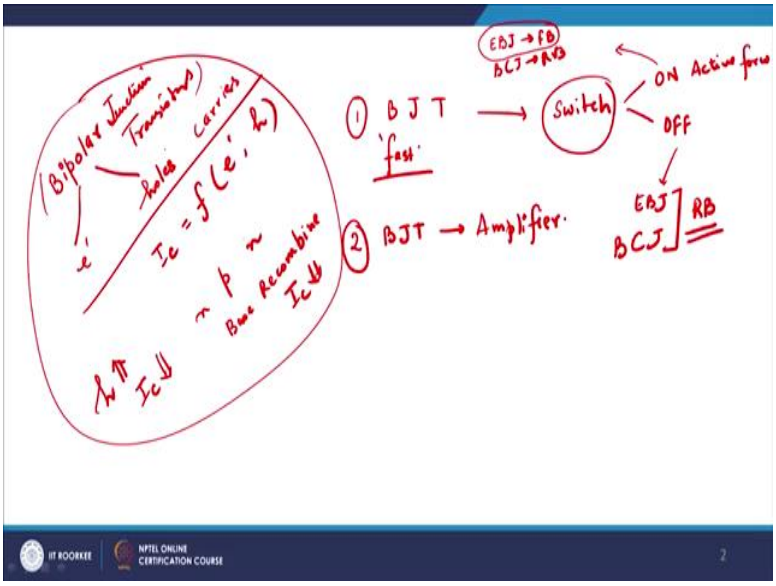
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−06**
**BJT as an Amplifier, Small Circuit Model-I**

Hello everybody and welcome to the online certification course on Microelectronics Devices and Circuits. Now in this lecture module we will be actually looking into the BJT as an amplifier and then we will also be looking at BJT from small signal model point of view. Let me recapitulate what we did in our previous lectures. We had looked into bipolar transistor, the structure of that both NPN and PNP.

What are the relative doping concentrations of the emitter base and collector and what should be the size of each of the regions of the BJT for the most optimized functioning of the BJT? We have also seen the various electrical characteristics of the BJT and we have tried to ascertain the factor that how can you bias 'C' or emitter base and base collector junction. And we have seen that depending on that there are four modes of operation.

So I can use BJT in analog domain for the purpose of amplification when you are actually biasing it in the forward active mode, right? You can also make it into cut-off provided both your emitter base and base collector junctions are reverse biased, right? So you learn two things, very important things in our by the previous discussion which we had.

(Refer Slide Time: 1:52)




And these two things, the first thing is that BJT can be used as a, of course as a switch, right? So we can go from ON to OFF state, right? How can we go from ON to OFF state? This is

again active forward, right? Active forward state, so what will happen in this state? Your EBJ emitter based junction will be forward biased and base collector junction will be always reverse biased, right? So you have here ON state, what is the OFF state? Both EBJ emitter based junction and base collector junction both will be reverse biased in that case it will be off state, right?

So I can use BJT as a switch for sure and therefore whenever we require a heavy switching action BJT can be used and it is a very fast switch, understand it to be very fast. The only problem with BJT switch is that when you go from on to off state and you want to change the biasing, we actually have to change the biasing of the EBJ, from EBJ from FB to RB so EBJ has to be changed.

And therefore we have to be very cautious that how fast this change is. The more fast this change is more higher frequency, at a much higher frequency the switching can occur, right? So that is what we need to look into. We have also learned second thing is that since BJT is able to convert small currents into large currents it can be also used as an amplifier, right? We have seen that in our previous discussion.

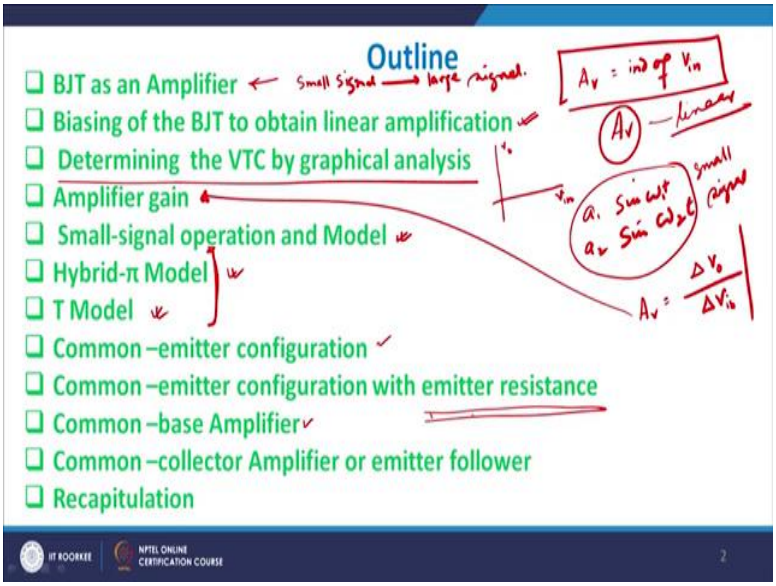
So two things are clear at this stage that a bipolar transistor can be used as a switch as well as an amplifier. One thing which we have left in our previous discussion is the name bipolar, why is it known as a bipolar junction transistor? And the reason is very simple because in both NPN and PNP we use both electrons as well as holes for as carriers, right? They use both electrons and holes.

So both electron and hole current is responsible for the total collector current. So the total collective current which you see is actually a function of the number of electrons as well as number of holes. Why number of holes in an NPN transistor? For example, why number of holes? Because if the number of electrons are high of course your emitter current will be high but if your number of holes are also high then the probability of recombination in the base region, right? Recombination will also be high, right? And as a result your collector current will be smaller now, which means that if your hole concentration is larger, large as compared to the previous case I would expect to see my collector current to slightly drop-down as compared to the previous case. And therefore bipolar technology or bipolar transistor as the name suggests will have contribution from both the carriers, electrons as well as holes, right?

And therefore we define that to be as a bipolar junction transistor and that was the name which was given to it way back in 50s when it was actually found out, right? Fine. So we have understood the origin of BJT, we have also understood that BJT can be working as a switch. We have understood the working principles of an NPN and PNP transistor. We have also understood the biasing criteria for the BJT to work fine.

So let us now therefore switch our attention to BJT as an amplifier that is the first thing which we will be learning. So let us just have an outline of the talk which we will be going through in this series of lectures. We will be looking at BJT as an amplifier first of all, so I can, so given a small signal.

(Refer Slide Time: 5:42)




Given a small signal peak-to-peak how can I increase it to a large signal, right? So that is what we will be learning in this lecture that given a peak-to-peak small should I get a BJT as an amplifier? I can get a large this thing. And we are also looking therefore about the biasing criteria of the BJT, right? Because this is quite important, very-very important section.

And the reason is you can have amplification at any one point of time but ideally your amplification should be independent. A voltage amplification should be actually independent of the input voltage, right? So where you bias your input voltage or where is your input voltage should not determine the value of your voltage gain ($A_v$) right? So please be careful. So therefore if the BJT has to be used ideally as an amplifier then it does not matter to me perse that whether you are giving a signal $A_1 \, Sin \, (wt)$ or you are giving a signal $A_2 \, Sin \, (w2t)$ and so on and so forth.

As long as these signals are small signals, right? And we will see that later on I would expect to see that I would get almost a linear amplification. Linear amplification primarily means that there is a linear profiling between the amplification and the input voltages, right? Now to do that we should therefore learn which is the second case that how will you bias your circuit in a manner, right?

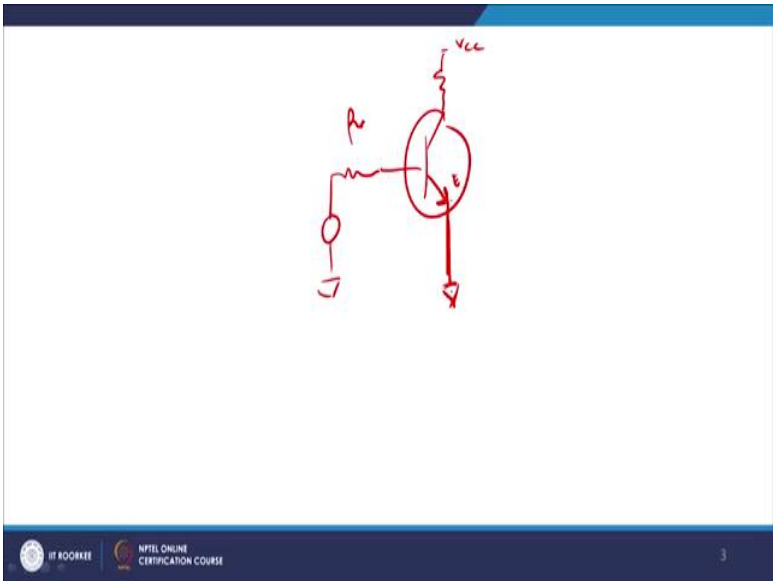
And to obtain a linear amplification, right? Then we will therefore learn that how to therefore determine the voltage transfer characteristics by graphical analysis, right? What is the meaning of voltage transfer characteristics? As we will see it is output voltage on the y-axis and $V_{out}$ and on the input axis, on the x-axis we have $V_{in}$. So how does $V_{out}$ vs $V_{in}$ vary? Right?

So if you vary $V_{in}$, how does $V_{out}$ vs $V_{in}$ vary, right? And we therefore need to find out $A_v$ as $\partial V_{out}$ upon $\partial V_{in}$, right? That is what we get as a voltage gain or the amplifier gain. That is what I was talking about in the fourth part of our lecture that amplifier gain will be discussed and on what factors does the amplifier gain depend we will be looking into it. Now in most of the cases these BJT or bipolar transistors as a name does not in the sense works standalone, right?

You have to use this in circuit analysis, so you have to sort of take this BJT and plug it and play it over a circuit for example, right? So what people have done over the years is rather than therefore every time discussing the device behavior if we are able to have a circuit representation of a BJT, right? A circuit representation then every time we encounter a BJT we replace it by its equivalent circuit that is what we will be learning when we discuss Small signal model, Hybrid-π model and T model.

So these are the two models which we will be learning, in these two models we will be able to therefore given any BJT in any configuration 'C', common emitter, common base or common collector I will be able to give you a corresponding circuit model for that which can be plugged and played within any of the BJT configurations for better understanding. We will be again looking into we will be revisiting again the common emitter, right? Common base and common collector amplifier and we will be looking with how does it vary when you have an emitter resistance being applied here, right? So please remember yesterday's talk when we were discussing common emitter, there was no emitter resistance.

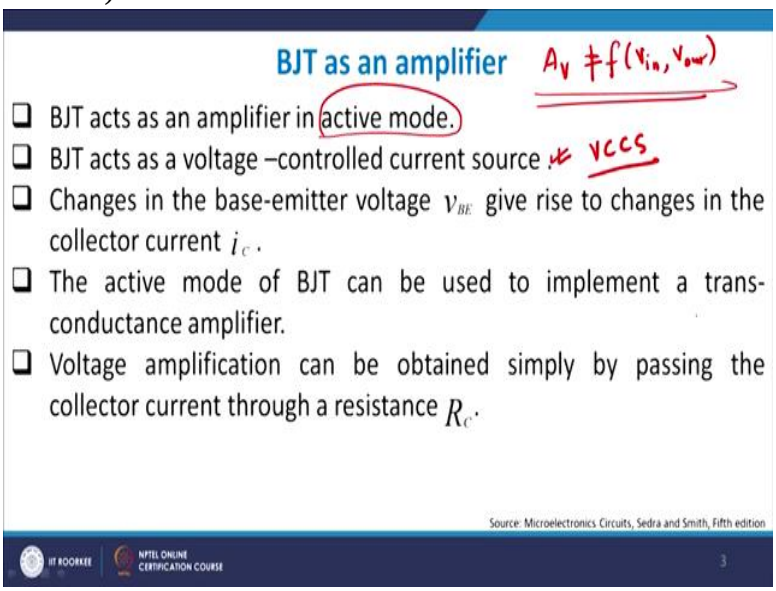
(Refer Slide Time: 9:46)




So what we did in our previous turn was that we were something like this that it was grounded, right? And this was $R_C$ and this was $V_{cc}$ and this was $R_B$, right? And this was applying a base and so on and so forth. So this was my BJT and I... I... my emitter which was this one was totally grounded we did not apply any potential here, right? Or there was no resistance between the ground and the emitter node of the BJT, right?

We will like to see how inserting a particular resistance in the BJT in that particular emitter node, how does it influence the overall capability of the common emitter configuration? So these things we will be watching along as we and then finally we will recapitulate what we studied across this lecture series, right?

(Refer Slide Time: 10:31)

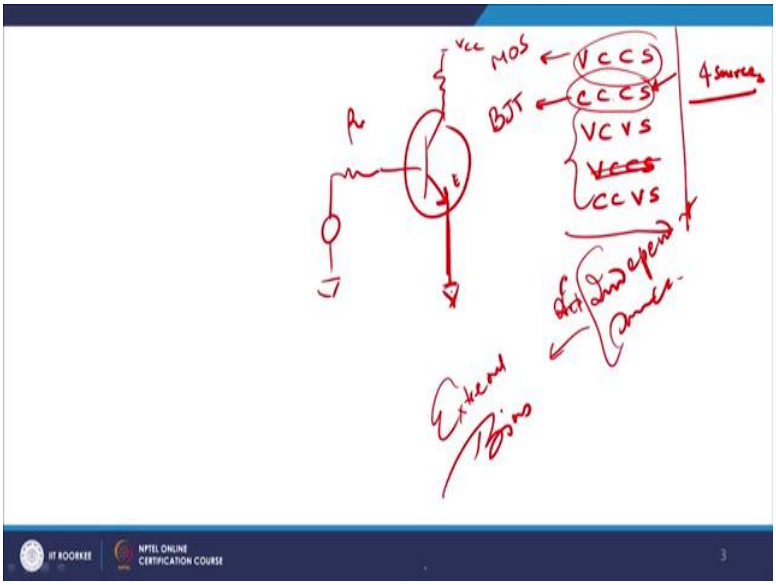


So this is what we will be looking into, the first thing will be that BJT is an amplifier. Well as I discussed with you the BJT will act as an amplifier only in the active mode, right? It will also act as amplifier in other mode but that amplification will not be a linear amplification and therefore my $\partial V_{out}$, $\partial V_{in}$ which is basically the voltage gain might be a function of $V_{in}$, right or $V_{out}$. In a sense that if $V_{in}$ therefore varies my $A_v$ will vary and therefore that is not a good idea.

In reality you should be very careful that voltage gain should not be a function of $V_{in}$ or even $V_{out}$. It is not a $V_{out}$ but surely not a $V_{in}$, right? So this criteria has to be held very confirm or in a very careful manner that a voltage gain cannot be should not be a function of $V_{in}$ and $V_{out}$, right? And that is quite important observation to take care of that typically they should be holding no functionality in terms of $V_{in}$ and $V_{out}$.

Now as we have seen therefore that BJT therefore acts as a voltage control current source, why? This is basically voltage control current source. So if you remember we have four types of sources.

(Refer Slide Time: 11:58)

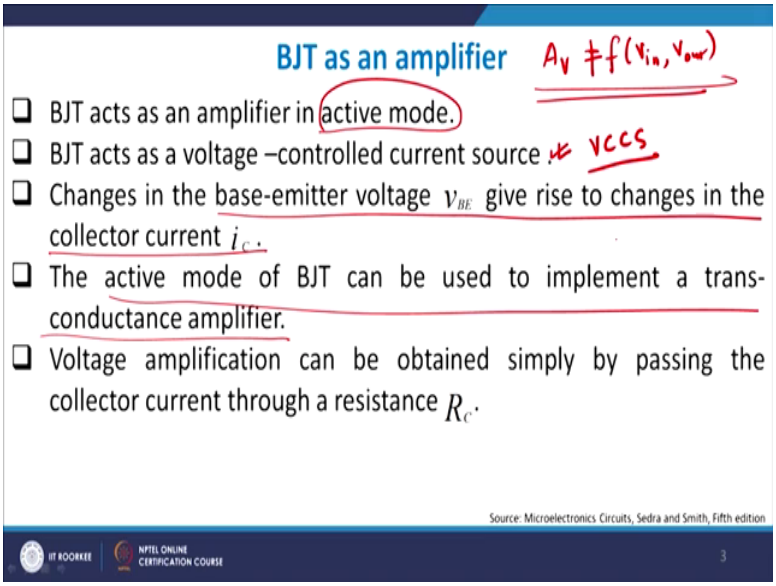



One is known as voltage control current source, current control current source, right you have a voltage control voltage source and you have a current control voltage source, right? So there four types of sources which is generally available to any electronic designer, right? An example of current control current source is basically a BJT right? A MOSFET for example is a voltage control current source, so a MOS device we generally get, think about these to yourselves and you can get it.

One example of BJT is also a voltage control current source just like BJT and we will see that later on but primarily BJT is a current control current source and therefore it is basically a, because the current in the base side of a BJT will determine how much amount of current there is in the collector side. Therefore we define that to be as a CCCS domain as far as BJT is concerned, right?

So these three things you should be very careful that these are four sources, right? And these four sources are all dependent sources sorry independent sources, but each source is having an external bias, right? External bias, so all of the sources are having external bias but they are of independent nature, right? So this is what we will be learning as we move along.

(Refer Slide Time: 13:45)

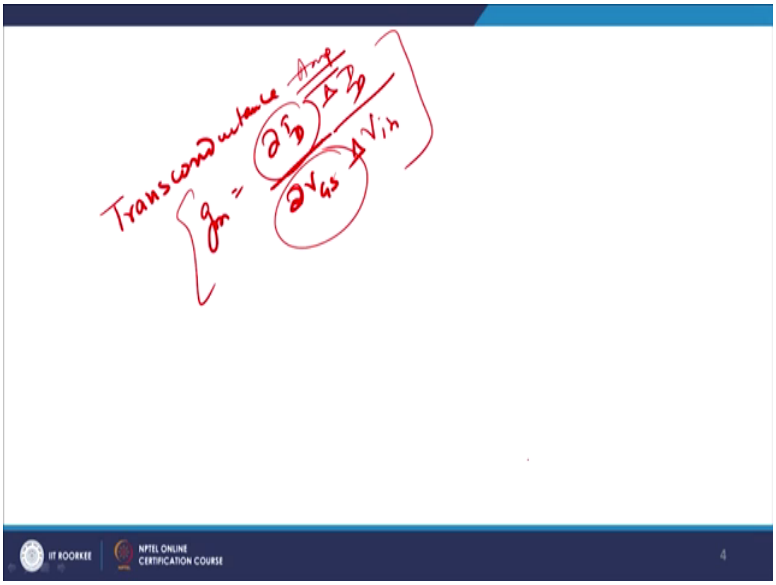



And therefore what we have learned just now therefore is that a change in the base emitter voltage which is this one gives rise to a change in the collector current $i_C$. I think this has been already discussed to a larger extent so for example if you are using a silicon transistor and your $V_{BE}$ is less than 0.7 you can be amply sure that the emitter base junction is cut-off and therefore there is no current appreciable amount of current flowing through the emitter base junction.

So your device will be in OFF state in that case because you are not allowing any of the transistors to electrons to flow in case of NPN transistor, the majority current carriers you are not allowing it to flow. But once your base emitter voltage exceeds 0.7 in case of silicon which is the cut-in voltage I would expect to see an exponential rise in the current by varying a $V_{BE}$, right?

So if you vary $V_{BE}$ near the knee point of the PN junction then I would expect to see a large amount of current flowing through the collector side once the input voltage is larger than the cut-in voltage or 0.7 in case of silicon. As I discuss with you therefore the active mode of BJT can be used to implement a trans-conductance amplifier, forget about this word trans-conductance we will be looking into it in a detail manner.

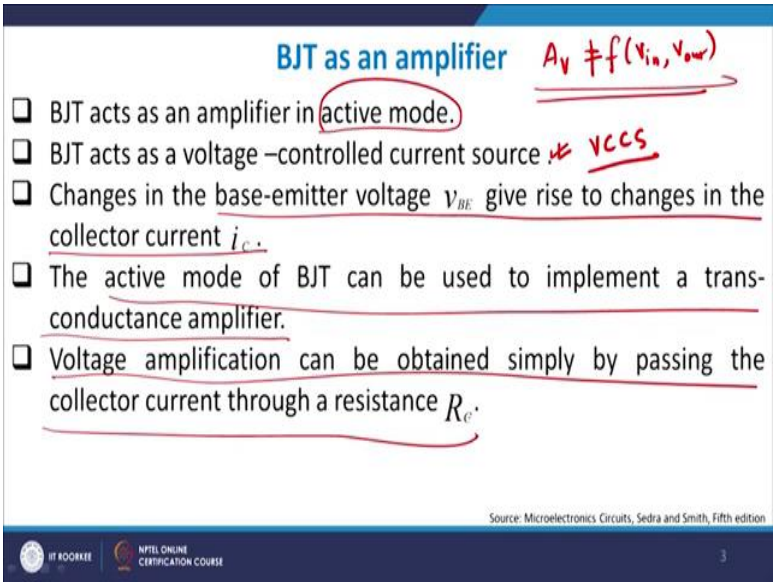
(Refer Slide Time: 15:12)




But you know you must be remembering what is transconductance? Transconductance is defined as $g_m$ is defined as $\partial I_D$, $\partial V_{GS}$, right? Is $\partial I_D$, $\partial V_{GS}$ which means that it is basically a rate of change of drain current with the rate of change of gate to source voltage, right? That is what is $g_m$. Now therefore what is a transconductance amplifier? Transconductance amplifier is an amplifier which amplifies the output current because of the change in the input voltage.

So this is your input voltage, right? So there is a $\partial$ of $V_{in}$ here because of which there will be a change in the output current $\partial I_D$ and division of that is basically my $g_m$ Transconductance, right? And therefore it is also referred to as a transconductance amplifier. So we refer to this as a transconductance amplifier and that is quite an interesting topic which people have been dealing with when they are dealing with these BJTs or bipolar technology.

(Refer Slide Time: 16:21)



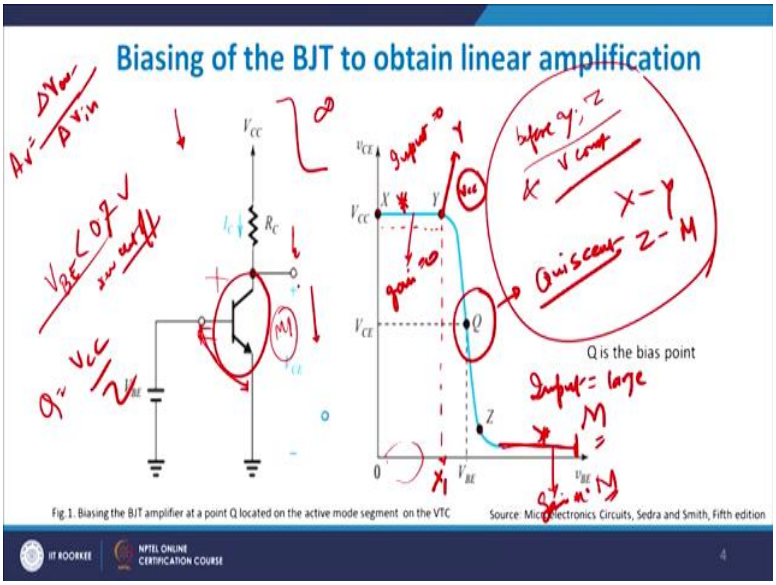

Now therefore so that is what I was saying that, so the last point is that a voltage amplification, right can be obtained simply by passing the collector current through a series resistor ($R_c$). So you have $i_c$ available with you because of a collector current is available with you. Now if the collector current you are forcing it to pass it through a resistance $R_c$ then $i_c$ $R_c$ is basically the drop of the output voltage, right?

Now even if your input voltage is small and your output voltage is large your $A_v$ will be therefore typically very high of the order of maybe 100, 200, 300, and so on and so forth, right also referred to as a voltage gain in this case. So two things we take away from this slide; BJT should be in active mode if you want it to act as an amplifier, changes in the value of $V_{BE}$ will give rise to a change in the value of $i_c$.

The third thing is that it is also referred to as a transconductance amplifier because the input is basically a voltage and the output is basically a current and the amplification can be thought of as the output current passing through the resistance $R_c$. The voltage drop across that resistance is defined as my output voltage, right? And the division of output voltage to input voltage is therefore referred to as a voltage gain, right?

Okay, let me come to the next slide and show to you how biasing of a BJT can be used to obtain a linear amplification, right? And we will see step-by-step each one of them in a much more detail manner.

(Refer Slide Time: 17:57)




Now if you look at this left hand side figure which is this one, right? I have an NPN transistor which is marked by this red circle let us suppose, right? And what we get is that we bias it by an external source $V_{BE}$ and we have $I_c$ and $V_{CE}$, right? Now the obvious question how do you physically amplify? When you amplify any signal, how do you do that physically? Well, first of all you apply a DC biased, right?

You apply a DC bias onto the device itself which you are supposed to amplify. Once the DC bias is applied you need to superimpose the small signal model over the DC bias in order to sustain a output voltage variation, right? So I therefore mark my DC level and superimpose with that I have an AC cycle available with me that multiplied by ($A_v$) Voltage gain gives me the output for any BJT.

Now that we will come later on, let us see, what are the functionalities of linear amplification? Now if you look at this figure here when $V_{BE}$ is less than 0.6 or 0.7 volts which is on this side, somewhere here, right? The device is cut-off, right? The device is cut-off, so when the device is cut-off automatically all the $V_{cc}$ will fall across the output voltage and that is what is happening.

For the output voltage you have got $V_{cc}$ and remains $V_{cc}$ up till a point $Y$ where what is happening? Where you will have the other Institute amplification to be zero in that case. So the amplification will start after a point $Y$ but from the point of view operation we say that till $Y$ or till let us suppose $Y$ on this axis and say $X_1$ on the $V_{BE}$ scale, right? I can safely assume that the device is in the OFF state and therefore output is at $V_{cc}$, agreed?

When my therefore voltage crosses 0.7, right? It switches on my transistor here as a result the voltage at this point starts to fall to $V_{dd}$, sorry at zero bias and then something happens here, right? We define a point which is known as the Q-point or the bias point where you need to sustain or you need to keep your device there in order to ensure that the device is in linear amplification mode and I will explain to you why is it like that.

See if you look very carefully, *Y* and *Z* are the two points at the two extremes after which the voltage $V_{CE}$ is almost independent of $V_{BE}$, it is a straight line. Before *Y* and after *Z* actually, after *Z*, I have not drawn but after *Z* it will look something like this. So almost a straight line, right? In both the cases before *Y* and *Z*, right? Before *Y* and *Z* we expect to see that the voltage is constant, right?

The voltage is constant and then after *Y*, the output voltage starts to fall down and you know the reason why because after your input voltage has crossed the value of your ON condition this tries to switch off this *M1*. As it tries to switch off this *M1*, more and more current will flow and you will expect to see this to fall down and when the time will come and this goes off or almost very near to zero then this will result in this point which is the starting point to be actually elevated to high-value voltage that is what is happening here.

So as you make your $Y_1$ move slightly higher there is a certain drop in the voltage and the reason is in that case the device is in the on state and therefore when the device is in the on state it pulls all the extra charge carriers towards its center living its neutral charges that did not happen and therefore what will happen is at point *Y*, I would expect to see that the output voltage is latched to $V_{CC}$. Now if you input an active voltage in fact it is $V_{CE}$ which is latched to a high value of $V_{CC}$. Now you see we define the Q-point quiescent point, Q is referred to as quiescent point or a point which is highly stable point, right? So these two are the definitions of the Q-point which is seen in front of you. And which means that this Q-point by application of an external bias can move up and down but within the linear regions of amplification process, right?

And that you should be very careful as far as designing is concerned that you cannot let the Q move anywhere it likes but across the only across the small signal domain, right? As such that you get a large amplification in this area. Now if you see very carefully as the voltage exceeds *Z*, right? Input voltage exceeds *Z* that ensures that this device is basically your cut-off, so $V_{BE}$ is very-very large.

Very-very large primarily means that your cut-off has taken place and the voltage actually goes down almost to zero value, right? What is left for you to understand therefore is how does Q-point behave which is this point. This is also referred to as a quiescent point, how does it behave and how does it help us in averting any problem? See if you look very carefully, if your device is biased here or here, you are either in input stage zero, so this is the input stage zero or you have got here input, a large input, fine.

So a large input and a small input I can show it on the same graph and this is how it is done. Before we show you how it is done, we will just comment on the fact that the points between *Y* and *Z*, right are so small spaced between each other that they seem to be merged with each other, right? But there are sufficient numbers of points to evaluate beyond *Z* and *Y* that this region is highly amplification zone and this region are the basically a digital zone.

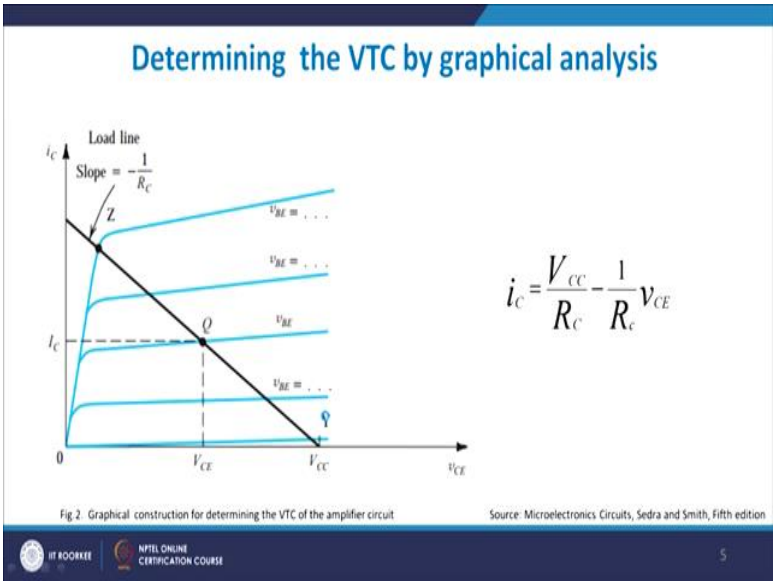
So if you want to work in digital electronics or digital domain you need to work in the region *X*, *Y*, and *Z* to this much point say this is *M*, *Z* to *M*, right and so on and so forth, you do not have to go for all the details which is available in the literature, open literature for this case then. Now typical value of Q is such that, Q is somewhere around $V_{CC}$ by 2, right? So whatever is the input voltage or $V_{dd}$, I am using or whatever clock voltage I just have to half it whether we talk about quiescent characteristics or a low-power characteristics.

Now what will happen is that as you... as you... for example $V_{CE}$ is equal to Q-point as you can see and the *X*-axis is $V_{BE}$, right? Now let me say that therefore if you want to work in digital domain *X*, *Y* and *Z*, *M* are the most logical digital principles because it is 1, 0 and 0, 1 here. And this slope gives rise to a constant gain, are you able to get the picture? Because $A_v$ will be equals to $\partial V_{out}$ by $\partial V_{in}$, right?

So it is $\partial V_{out}$ by $\partial V_{in}$ which means that my output voltage will depend upon the input voltage as such, right? Secondly, what about the gain at *X*, *Y* and *Z*, *M*? Because gain is change in the output voltage upon change in the input voltage. So though your input voltage is changing, output voltage is constant and therefore your gain at this point is exactly equals to gain at this point, right?

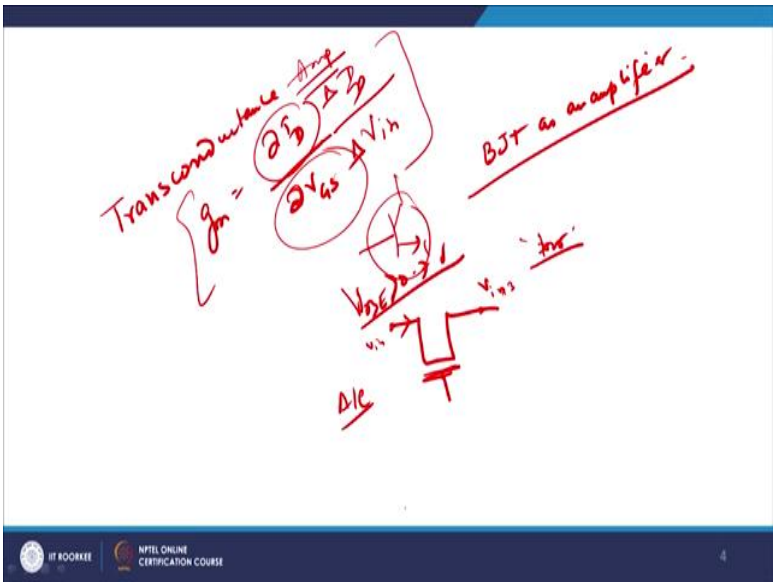
Fine, so gains are exactly equal in both the cases, but gain shows a drastic change in the middle region depending upon how steep the slope is. So if your slope is very-very steep, right something like this, you will have even infinity gain shown to you, right? And that is how people work with inverters for starting application purposes, right?

(Refer Slide Time: 27:31)




We now therefore come to determining the VTC by graphical method, right? How will you find out the voltage transfer characteristics graphically or looking at a picture at a graphical point of view, right? And that is what is required and I will discuss that in this section in this lecture, right? Let us take one important point, let us take BJT switch or the BJT transistor as an amplifier, right?

(Refer Slide Time: 28:05)

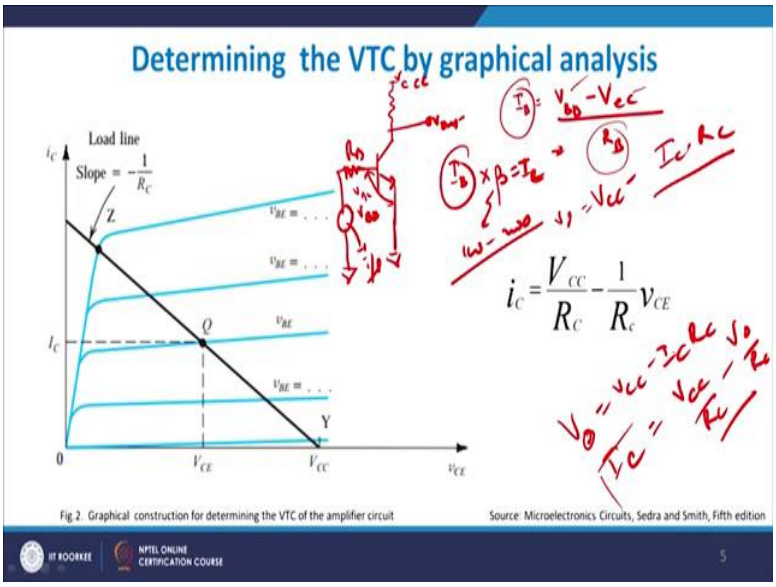



BJT as an amplifier let us take, right? So what we do is, we try to provide the candidate enough database in order to appreciate the working principle. Let us see how it works out. So if I have a diode connected load for that matter any load, right? What I am trying to tell you is that, in this case as you switch on the value of input voltage and one of them maybe AC, you

expect to see a large amount of current flowing here but since it is normally on-off state there will be no current flowing through the inverter or in this matter CMOS and therefore you require a person from inside to switch it on, right?

How will you do that? By simply making this BJT work at typically at $V_{BE}$ greater than 0.7 volts. When you make it $V_{BE}$ greater than 0.7 volts you switch on this base emitter junction and as a result output is available to you.

(Refer Slide Time: 29:12)



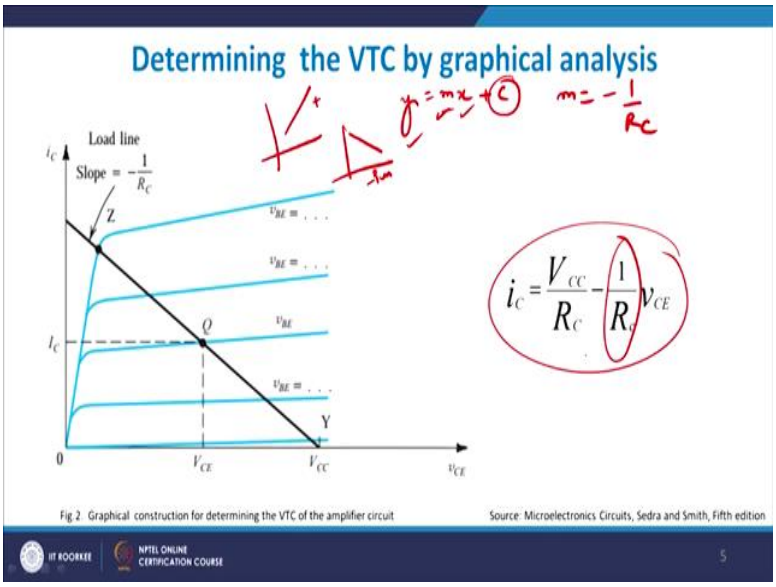

So what was the concept is that in a common emitter mode configuration, let me draw for you the common emitter mode configuration design, right? And this is without the collector emitter configuration $V_{cc}$, right? You had $R_B$ and then you had $V_{BB}$ and then you had this and then on the emitter side you had this, this is your output and this is your input which is basically a sinusoidal input.

So what has happened is that because of this $V_{BB}$, I will have a current, this $V_{BB}$ is here, so what will happen is there will be a base current $I_B$ which will be formed by doing $V_{BB}$ minus $V_{ec}$, so this is the breakdown voltage. So we write down $I_B$ to be equals to this divided by $R_B$ where $R_B$ is the base resistance here, fine. So you put the value of $V_{cc}$, you put the value of $V_{BB}$, put the value of $R_B$ and you get $I_B$, right? You get $I_B$.

Now once you get $I_B$ you need to multiply that with beta in order to get $I_C$, right? And this β is with the order of few 100s to 200 value and I get $i_c$, right? I get $i_c$ here the collector current. Now this collector current multiplied by $R_C$ gives you the voltage drop across the resistance,

right? So that is what you get as that so what is output? Output is $V_{cc}$ minus $I_C$ $R_C$ is equals to $V_{out}$, fine. So I get $V_{out}$ to be equals to $V_{cc}$ minus $I_C$ $R_C$, right? And I get this quantity here and therefore I get $I_C$ to be equals to $V_0$, right? So $V_{cc}$ by $R_C$ and I will get minus $V_0$ by $R_C$, so I get $I_C$ to be like this, right?

(Refer Slide Time: 31:31)




Which means that if you look very carefully or even if you look at this particular point I can safely say that this is the final content that the collector current is only depending on the value of $V_{cc}$ by $R_C$ and it depends upon minus 1 by $R_C$ $V_C$ which means that if you compare with the fact $Y$ equals to $m_x$ plus $c$, $m$ is basically your 1 by $R_C$, right? So why with the negative sign because the slope is negative, right?

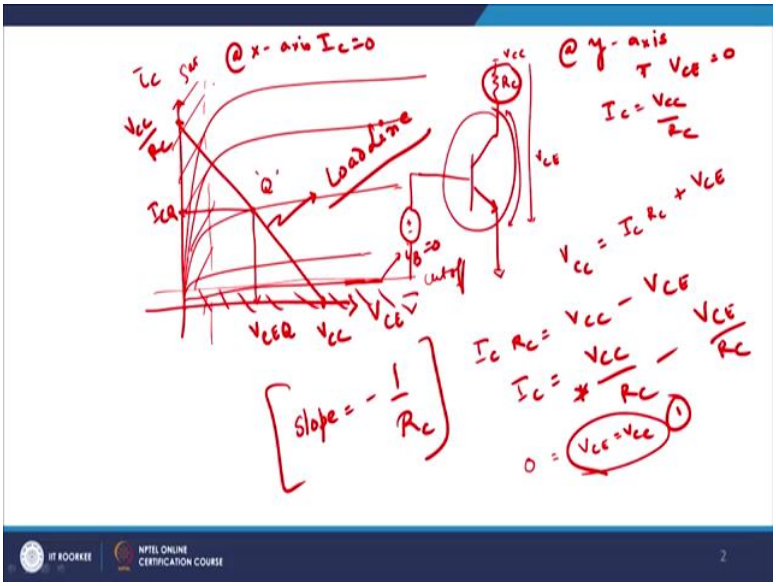
So if I have something like this, this positive slope, if I have something like this it is a negative slope. Since this line is something like this in the reverse direction you get a negative slope and therefore m is fixed at minus 1 by $R_C$, right? Now therefore if you know m you know $XY$, I can get the value of $C$, right? So this is a standard methodology which people use.

Now what has happened is over the years and typically in the early years we used to find out the VTC or the voltage transfer characteristics by graphical method, fine. So we will do that in the next turn, today we have learned the basic idea of a short-circuit or a small signal model next time we will be learning how to find out the voltage transfer characteristics by graphical analysis, a detailed analysis will be followed, okay. Thank you very much.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−07**
**BJT as an Amplifier, Small Circuit Model-II**

Hello everybody and welcome to the online NPTEL online certification course on microelectronics: Devices and Circuits. In our previous discussion on our lecture and the previous module we have understood how your BJT acts as a switch as well as an amplifier and where should we bias our device, so that the device works as an amplifier. In this slide we are continuing with the same thing which we have left in the previous one.

(Refer Slide Time: 1:05)




That we see that in a common emitter mode configuration, if you have a common emitter mode configuration BJT, right? Then you have $R_C$ here, right? You have $V_{CC}$ and then you will have, this so for a common emitter mode configuration you will have this and then there will be a bias given here and this bias will bias the device at a particular, appropriate Q-point, let's see how do, I do that. Assuming, so if you go back to this table this is basically your $V_{CE}$, so I get, $V_{CC} = I_C R_C + V_{CE}$, right?

So you see, $V_{CC} = I_C R_C + V_{CE}$,, therefore if you want to make $I_C R_C$ one side I get $I_C R_C = V_{CC} - V_{CE}$, so as I discussed with you $I_C$ will be there for equals to $V_{CC}$ by $R_C$ minus $V_{CE}$ by $R_C$. So now if you plot a graph of, let us say on this side you have got $I_C$, right? And on this side you have got $V_{CE}$, right? Now putting the same equation in this one if you look at the *x*-axis your $I_C$ will be equals to zero in that case.

So when $I_C$ equals to zero, I get $V_{CE}$ or $V_{CE}$ to be equals to $V_{CC}$ so what I can point out is, that this is the $V_{CC}$ point, right? And then on the y-axis your $V_{CE}$ will be equal to zero. So, at *x*-axis, right? $I_C$ equals to zero, place it into this equation and you get $V_{CE}$ equals to $V_{CC}$ this is the first thing you get and you place a point here. The second point is at *y*-axis, I get $V_{CE}$ to be equal to zero.

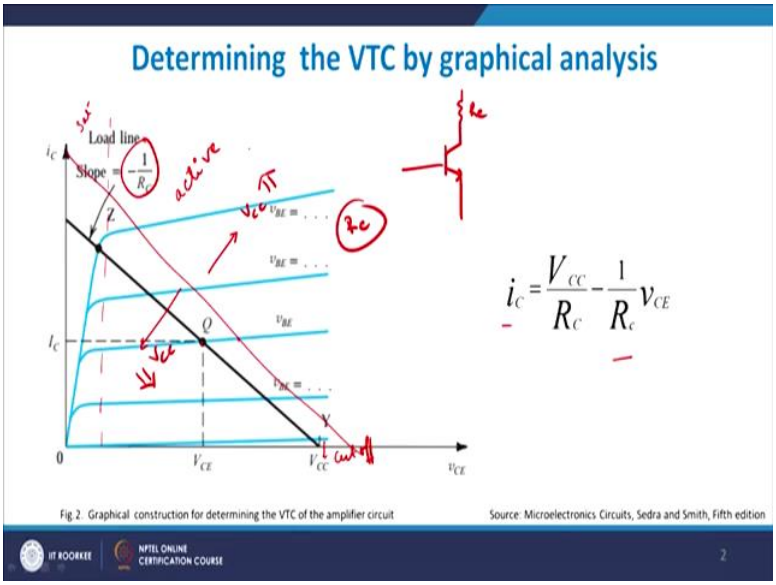
As I get $V_{CE}$ equal to be zero, I safely write down this to be as $I_C$ to be equals to $V_{CC}$ by $R_C$, so I put a point here which is basically $V_{CC}$ by $R_C$. Now simply join these two lines and we define this to be as the load line. I will explain to you why it is known as load line. So it's a load line, whose slope is incidentally equals to -1 by *R*, fine, which means that if the resistance 1 by $R_C$ actually this $R_C$ is collector resistance. So depending upon the collector resistance the slope of this line will change, right?

Now wherever it cuts the *I-V* characteristics of the device, we define that to be as the Q-point those are the possible points where you will get the Q-point. So if you have *I vs V* characteristics here that's a good idea to plot it first of all like this and I get something like this then, I get this to be possibly a Q-point because within the active region this Q-point seems to be further away from saturation as well as from cut-off.

So this is my cut-off point, this is my active and this is saturation. So how did you find cut-off? Cut-off is when your $I_B$ is zero. For any point in which $I_B$ is less than zero, it is cut-off. So this is your cut-off region. So your device is cut-off there is no idea, this one is saturation we have already discussed yesterday and this is the active region, so you should bias your device using a DC bias, external DC bias such that the Q-point is somewhere lying here. So the Q-point is this much.

Corresponding to this Q-point you will have $V_{CE}$ in this direction and $I_C$ in this direction, so this $V_{CE}$ and so we write $V_{CEQ}$ and $I_{CQ}$. So $I_{CQ}$ is the quiescent collector current and we define this to be as the collector to emitter quiescent voltage, right? And once you have found out, now what you do? With this value of a voltage and current, if you give in the input side using a DC bias I superimpose my AC signal over this DC bias, right?

(Refer Slide Time: 5:41)



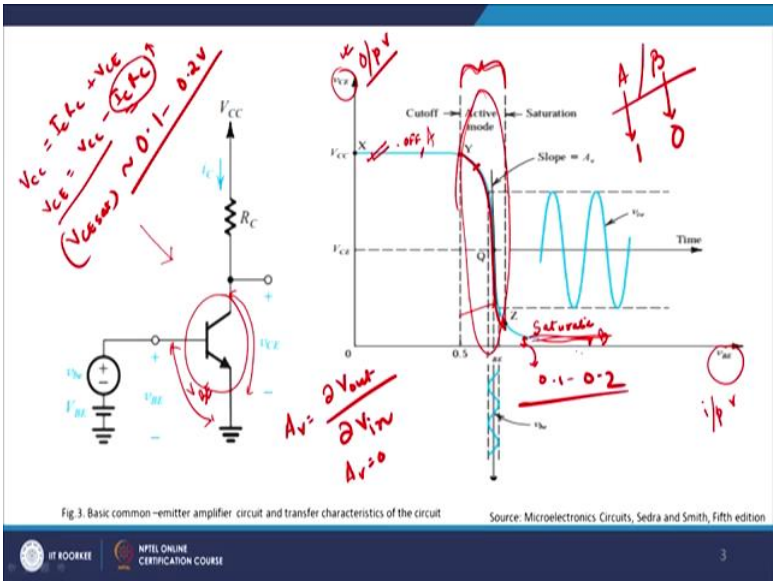

And that is what they're trying to do over the period of time that we are trying to find out our DC signal, so as you can see from here the Q-point is somewhere, well so you see this is your saturation region, this is your active region by the definition we are handling and below this you have got cut-off, right? And this is the Q-point which you see for various values of $I_B$, right?

It is actually $V_{BE}$ because $I_B$ will depend upon also the value of $V_{BE}$. Now $I_C$ is by this and therefore the slope is equals to minus 1 by $R_C$ and therefore this is known as load line but you go on changing the load here, right? You go on changing the load here and you automatically get… So if the $R_C$ value is now smaller the slope will change drastically, right? The slope will be something like this now. It might be something like this, right?

So if you make your $V_{CC}$ larger and larger the load line will move to the right. If you make it smaller and smaller it will move to the left. So it will move like this, if $V_{CC}$ is made high, right? And it will move to this side if $V_{CC}$ is made lower, right? And depending on the value of $R_C$ the slope of the load line will go on changing, fine. So we have therefore found out a graphical method by which we can predict what should be the Q-point or the bias point where my device should be sitting, so that in the active region so that a proper amplification is available to me, right? That is very-very important here I wanted to tell you.

Ahh.. This is the, therefore this is the graphical method of determining the VTC the voltage transfer characteristics. And this gives me a quite a nice idea about or a decent idea about where should the bias point or the Q-point should be kept in general.

(Refer Slide Time: 7:32)




With the same concept we have learned till now, let me show you a proper amplification here. Now as I discussed with you that this is the common emitter mode configuration transistor which you see, this is my NPN transistor. $V_{CE}$ is this voltage is $V_{CE}$, this is $V_{BE}$ which you see, right? And this is your $V_{CE}$, right? Now if you, as I discussed with you for low values of $V_{BE}$ typically your, the device will be cut-off because base to emitted junction will not be forward biased for silicon it is 0.7.

So anything below 0.7 cut-off, cut-off means this is zero. Zero means all $V_{CC}$ appears across the output voltage and this is what you see here. So this part is that discussion which you will see. Similarly if your input voltage $V_{BE}$ is very large as compared to the cut-in voltage of the base emitter junction which is 0.7 for silicon, I would expect to see the transistor to get on, so I will get from off-state, so this is cut-off state and I will get on-state, right?

This is the on-state, right? And I get saturation here, right? I get saturation. So saturation is there and the output voltage will certainly fall off. Why it will fall off because if you remember, $V_{CC} = I_C R_C + V_{CE}$, right? So if you take it this side and do sort of try to find out the value of $V_{CE} = V_{CC} - I_C R_C$ . So when your $I_C$ is zero, you get $V_{CE}$ equals to $V_{CC}$ and that's what you get here.

When you have large values of $I_C$ this quantity becomes large and what you get is $V_{CESat}$ or the saturating value of $V_{CE}$ which is approximately 0.1 to 0.2 volts typical value, so this value is approximately equals to 0.1 to 0.2 in saturation, fine? Now you see if you bias it here point '*A*' let us suppose or you bias it here let me say here, right?

Point '*A*' and point '*B*' if you bias it obviously point '*A*' and point '*B*' will give you a DC bias, external DC output voltage. Point '*A*' will give you output 1 and point '*B*' will give you output 0 as you can see from the definition of output here but let's concentrate on the part where this active region is there or active mode is there, right active mode? In the active mode if you see the voltage is drastically falling $V_{CE}$ with the small change in the value of $V_{BE}$.

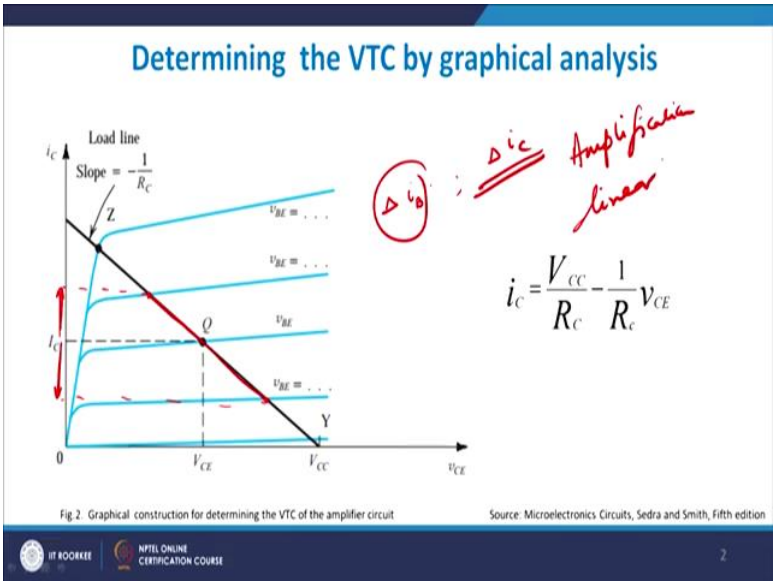
So this is basically $V_{CE}$ vs $V_{BE}$, so this is the output voltage and this is the input voltage. So this is my input voltage, right? And this is my output voltage so you see, so I define my voltage gain $A_V$ to be equals to $\partial V_{out}$ to $\partial V_{in}$, right? This we have already discussed, so if you look at point '*A*' my $\partial V_{out}$ is 0, so my $A_V$ is zero because there is no change in the output voltage is constant, independent of the input voltage.

Similarly at point '*B*' again my output is constant, independent of input voltage. So in both the cases my overall gain ($A_V$) is zero, voltage gain is zero but the voltage gain in the active mode which is this one is pretty high because you see a very small change in input gives you a large change in the output, right? But the change is what? Negative change. So I get a negative gain sort of concept which means that for an increase in value of input, I get a decrease in the value of output, right?

And the decrease is not equal but it is a very large decrease, so you get a large voltage gain. So higher the slope of this curve from '*Y*' to '*Z*' through '*Q*', more will be the gain or more will be the slope and more will be the gain of the transistor, right? So if you want to operate the device as an active device then please operate within this region which is dotted here and if you don't want to operate and let it work as a digital world you operate in these two positions.

So when you use TTL logic, you use this right transistor-transistor logic. When you want to use it as an amplifier you use this, right? So we have made one thing very clear here that for sure therefore, that if you have a BJT working in the active mode of operation and I am biasing my device in the active mode I am surely getting a large level of amplification to me, right? And is amplification will be linear amplification, right? And I will also find, will show, I would like to convey to you why is it like that also.

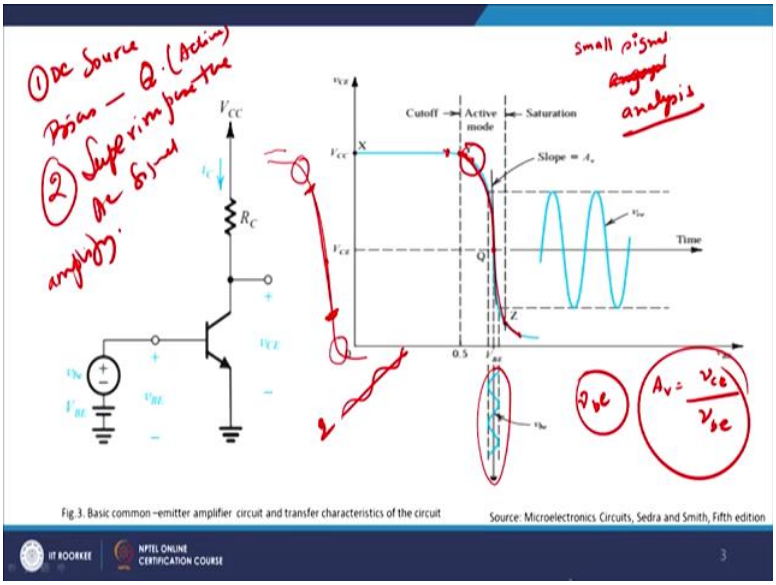
(Refer Slide Time: 12:12)




The reason is something like this. The reason is that for equal change in the value of base current, I should get an equal change in the value of a collector current, that is what linearity means. That if I do similar changes in $I_B$, I should expect to see a similar change in $I_C$ also, right? So you see if you change $I_B$ by this much amount, right?

Such that the Q-point shifts from here to here from this point and then goes from here to here, right? But the change in $I_C$ is linear, so this distance is exactly equals to this distance, right? And therefore we safely say that the amplification is linear in nature, so the amplification which you find out amplification is primarily linear in dimension, sort of linear in nature, right? But the reason being that you are able to, because the rate of change is almost constant for both the cases, fine. And that makes our life easier as far as understanding the graphical analysis is concerned. So you see I gave an input, so my input will be given, so…so… let me therefore recapitulated, so what you do?

(Refer Slide Time: 13:26)




You first insert the DC source, right? You bias the device where, so that it is in the Q-point in the active region for amplification purposes, right? And then superimpose the AC signal, right? Which you want to amplify, right? So you superimpose the AC signal which you want to amplify. So what is the AC signal here?

The blue one if you see here, this is my input AC signal and we are giving it to as $V_{BE}$, so I am giving as $V_{BE}$ base to emitter voltage change, so I have a DC bias, say 2 volts and 2 volts DC bias is kept superimposed on that I have my AC signal, small signal AC signal, fine. One important point which I should point out here is that there is a limit of the input AC signal.

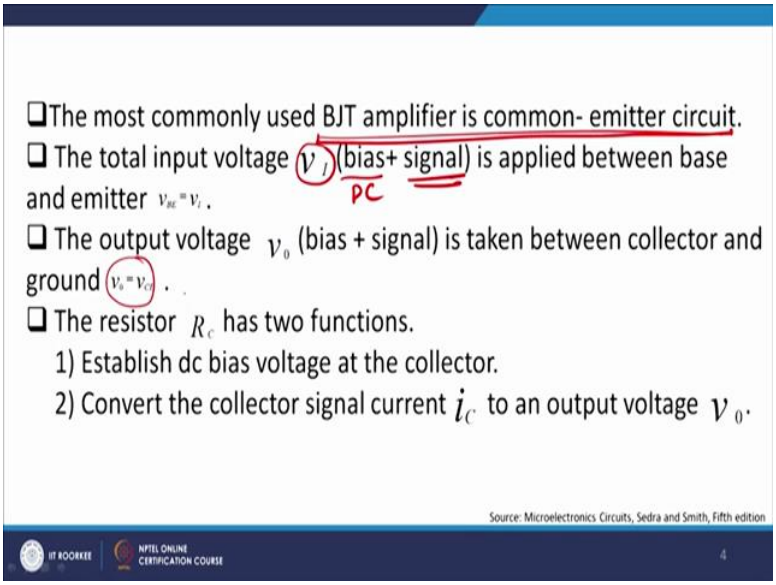
You just cannot give any AC signal you want and it will work fine for the BJT for example, specially the peak-to-peak value, right? If the peak-to-peak value, suppose you are biasing Q-point here and your peak value is so large that it makes this Q-point go somewhere here in the non-linear region or even here or on the negative side get somewhere on the non-linear region or even here, then you enter into non-linear amplification domain.

Although your amplification fall down, not only that it will depend upon the value of input not a good idea. So you see therefore, we always consider this to be as a small signal analysis, right? What is the meaning of small signal analysis? It means that, that your input signal is so small peak-to-peak that I can sufficiently state that the amplification is linear because if the input signals are too large then the Q-point will move so much away from its basic $V_{CE}$ value which is kept here, right?

That it might even go up to this much point, here. Here if you see it is something like this, it is coming like this and then it fell like this and then it became linear and then it became like this, so if you are somewhere working here or here, it is basically non-linear and as a result you will get a non-linear amplification. So you need to bias the Q-point such that you are just restricting your peak-to-peak between this point and this point, right?

So you should be very careful how you are biasing you device, depending on that will get the output. So depending on that you are getting the output here which is 180 degree phase shifted as well as amplified by a factor which is equal to value known to you. So what will be the small signal gain which you will get? $A_v$ will be equal to $V_{CE}$, right? Divide by $V_{BE}$ typically that will be the voltage gain which you will see out of the signal, right?

(Refer Slide Time: 16:31)

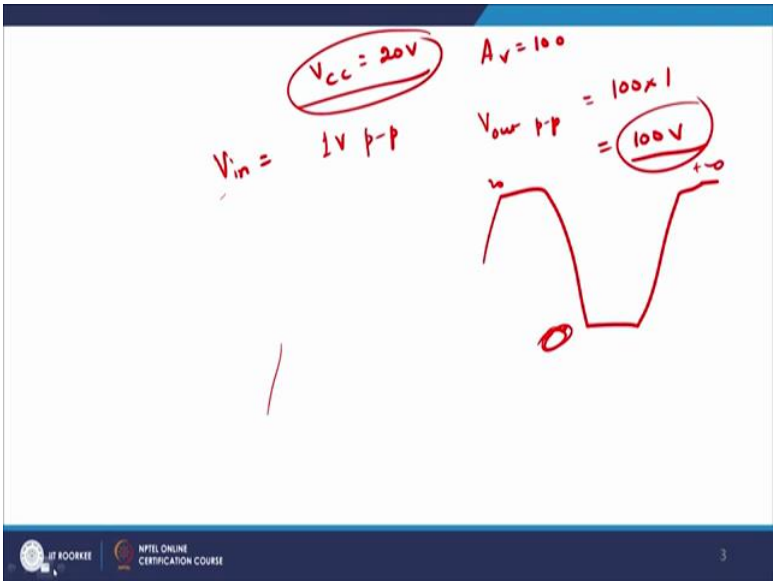



So with this we have understood the basic amplification process of a BJT, so as I discussed with you earlier also, in the earlier module as well as in this module the most commonly used BJT amplifier is the common emitter amplifier, right? So common emitter amplifier or common emitter circuit is the most commonly used amplifier design. Now, my input signal as I discussed with you, will always be a sum of the DC bias plus the signal which you applied, right?

So, $V_{in}$ is the input signal will be a sum of the AC signal plus the DC bias. So DC bias will fix up the Q-point and AC signal will be a signal which will be given, which will shift the Q-point so that you are able to amplify the signal between input and output. Now the output range is $V_{CE}$ as I discussed with you, right? And it cannot go beyond a particular region. $V_{CE}$

should also be limited by certain things. Just as I should make it a point that $V_{CE}$ cannot be more than of course $V_{CC}$ I will give you an example.

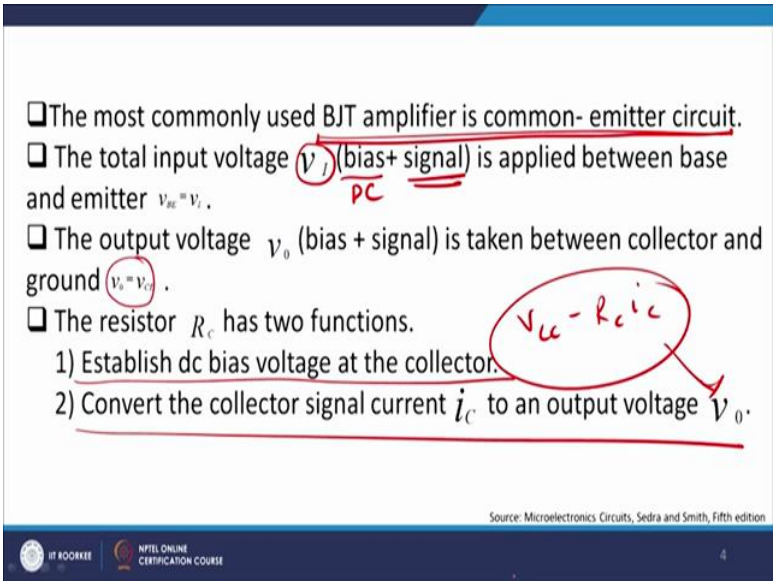
(Refer Slide Time: 17:36)




Let us suppose you have an input signal which is 1 volt peak-to-peak, right? You have $V_{in}$ which is 1 whole peak-to-peak your $A_v$ is 100, right? And your $V_{CC}$ is let us say, 20 volts for a BJT, right? So what will be your $V_{out}$ peak-to-peak? It will be 100 multiplied by 1 which is equals to 100 volts, fine. So your output is 100 volts but your $V_{CC}$ is only 20 volts, so you see that in no way we can get 100 volts output.

So what will happen is that it will go up to 20 volts, it will then clip back here, because this is +20 this is -20 whatever +20 and so on and so forth, because zero it will go to zero and not go further down, right? So you see therefore that the output of a BJT will depend not only on the gain, small signal gain of the amplifier but it will also depend upon the $V_{CC}$ which you are using or the amount of bias you are using to get those functions.

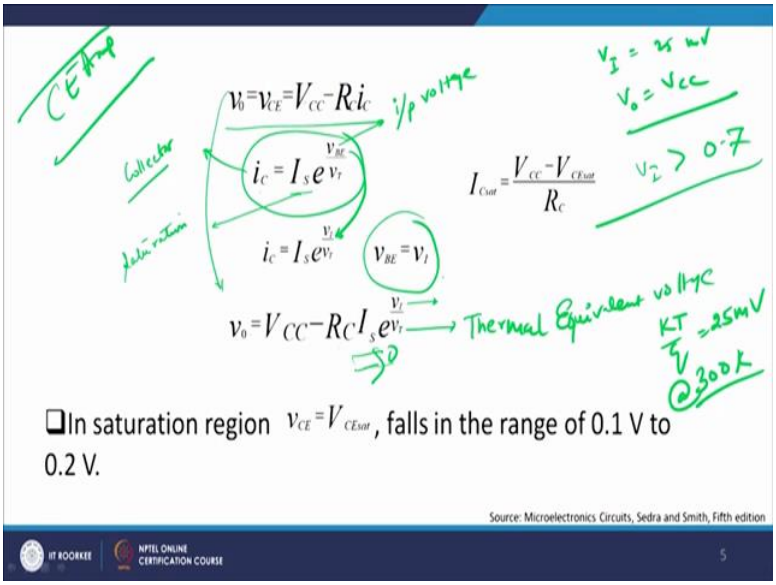
(Refer Slide Time: 18:54)




The resistor $R_c$ has two functions. We have already discussed this point $R_c$ will help you to establish a DC bias, so higher is the value of $R_c$ slope will be lower but with a negative sign, right? And lower the value of $R_C$ higher will be the slope with a negative sign. The slope is from the side, it is not like this, it is like this. So you will always understand that the slope is negative in the sense.

And second is that more important point is that, $I_C$ multiplied by $R_C$ is the output voltage, now it is the voltage across the value of, so this is the reason why we convert collector signal current to, so $I_CRc$, right? So $V_{CC}$ minus $i_cR_c$, is the output voltage. So $R_C$ has two components or two issues, one is to establish a DC biased and the collector and the second portion is to convert a signal current, collector signal current $Ic$ to its output value voltage $V_0$.This gives me a basic idea about the basic amplification process of BJT.

(Refer Slide Time: 19:51)

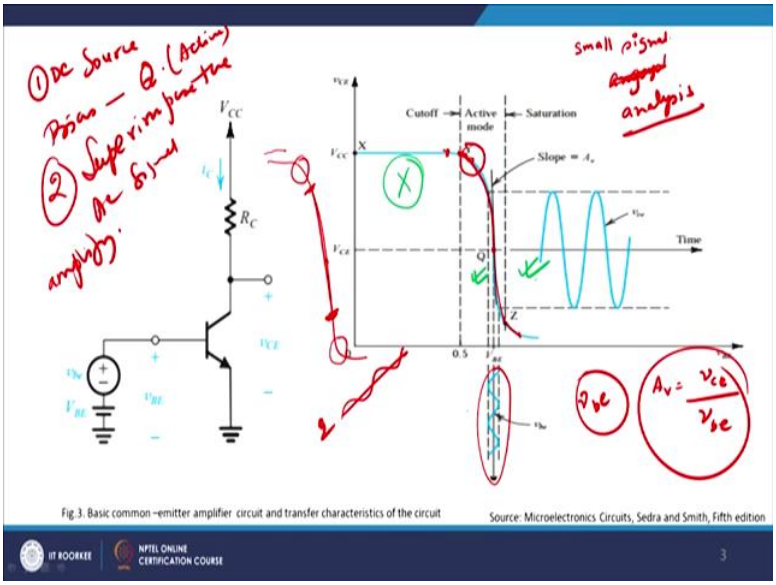



Let me give you the mathematical treatment here. So if you look very carefully $V_0$ which is in front of you, right? $V_0$ is there, right? If you look at $V_0$, this $V_0$ is the output voltage and is equals to obviously $V_{CE}$ must be equal to $V_{CC}$ minus $I_C R_C$, remember? We have already discussed this point. My yesterday's talk or previous modules talks we have seen that $I_C$ was equal to $I_s e^{\frac{V_{BE}}{V_T}}$. $I_S$ is the saturation current this is the saturation current which you see, this is the saturation current, right? This is the collector current and $V_{BE}$ is the input voltage in the case of common emitter. In case of common emitter amplifier and therefore I get $I_{Csat}$ to be equals to $V_{CC}$ minus $V_{CEsat}$ by $I_C R_C$ and I think this is pretty clear. So $V_{CC}$ minus $V_{CEsat}$ by $R_C$ gives you $I_{Csat}$.

Now since $I_C$ therefore is equal to $I_S$, so $V_{BE}$ is now replaced by $V_I$, we have just discussed this point why? Because I can safely say that the total voltage is a sum of the DC bias which you are giving and the AC bias of the signal which you want to amplify, right? So these are the two important voltage sources which you are trying to do. So I get $V_0$ therefore this statement it goes to $V_{CC}$ minus $R_C$ times $I_C$, $I_C$ can be written therefore $I_s e^{\frac{V_i}{V_T}}$. right? Is it okay? Now therefore you see very interestingly that the output voltage $V_0$ has an exponential dependence on the value of your input voltage $V_{in}$, right? Which means that if $V_{in}$ increases even by a small amount, I would expect to see a drop in $V_0$ by a large amount because this is an exponential drop which you see. $V_T$ is the thermal equivalent voltage which is given as $KT$ by $q$ and this equals to 25 mVs at 300 K.

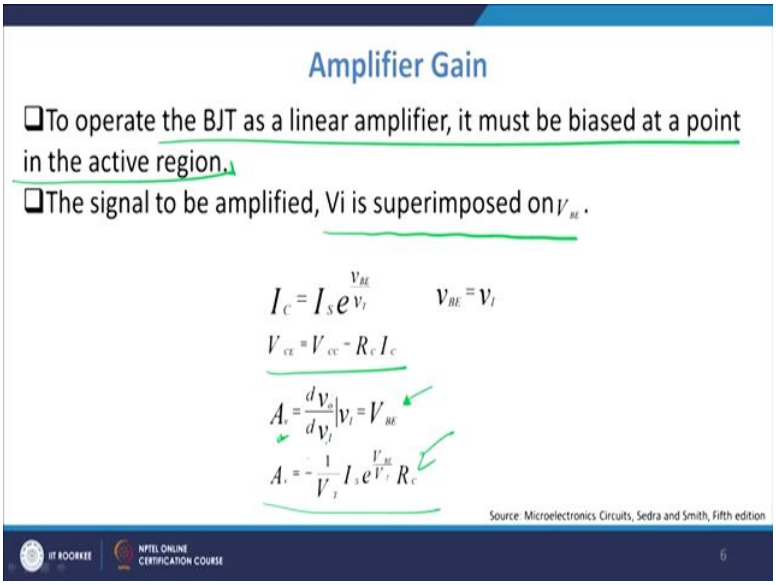
So at 300 K, $V_T$ equals to 25 mVs, right? So if your $V_I$ is let us suppose is, so if my $V_I$ is 25 mVs, right? So $e^1$ will be equals to 0 and therefore I get $V_0$ equals to $V_{CC}$. Remember as long as $V_I$ crosses 0.7, I will expect to see $V_0$ to be equals to $V_{CC}$ because this will effectively $g_o$ to zero, if not exactly go to zero, right, in the saturation region, which is the saturation region after the active region expands.

(Refer Slide Time: 22:50)




So, I have got if you look carefully I have got this thing that I have got cut-off region which is this one, right? This is the cut-off region then we have the amplification region with the active mode and I have got a saturation region here. So cut-off region, active mode and saturation it will vary depending on the value of $V_{BE}$ which you are choosing. So depending on the value of $V_{BE}$ we can think of active and saturation region, right?

(Refer Slide Time: 23:12)

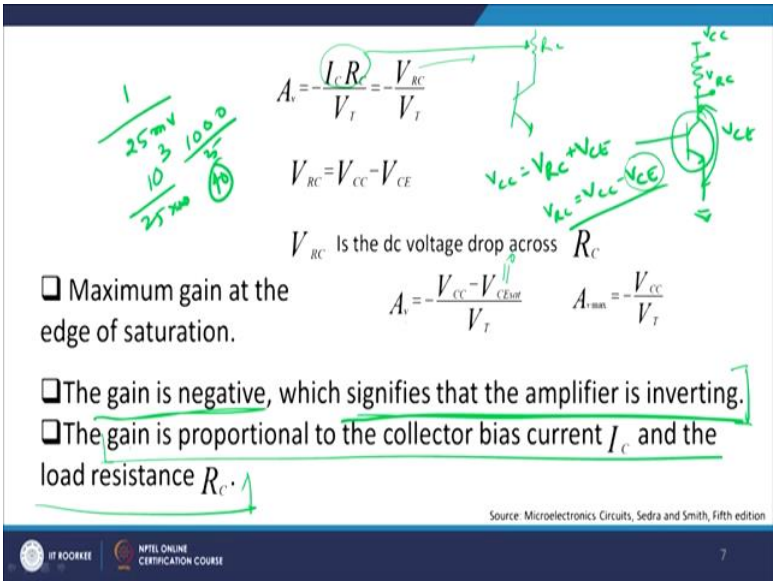



With this knowledge we have gained till now, let me see as what we are looking into. So we are looking into the fact that therefore to operate the BJT as a linear amplifier it must be biased at a point in the active region, we have discussed this point already time and again that we need to bias it, amplifier. And then $V_I$ is therefore superimposed on $V_{BE}$. We again discussed this point in a much more detail manner.

So I get $V_{CE} = V_{CC} - I_C R_C$, so I get $A_v$ to be equal to $\partial V_0 \, \partial V_I$ which means the rate of change of output voltage with respect to change in the input voltage for a fixed $V_I$ comes out to be $V_{BE}$, right? So $A_v$ is equals to $V_{BE}$. $A_v$ also comes out to be equal to -1 by $V_T$, $Is\, e^{\frac{V_B}{I_{RC}}}\, I_C$. Please do it yourself you'll find this answer very straightforward and simple, right?

And I will therefore get a negative sign expected also because of slope is negative and therefore I should get a negative gain. Negative gain primarily means with increasing voltage, input voltage the output voltage is decreasing and that is the reason we define that to be as a negative voltage.

(Refer Slide Time: 24:13)




Therefore I get $A_v$ to be equals to minus, why did you get $A_v$ equals to this thing whole quantity. Now this quantity is basically $I_C$, so I get $I_C R_C$ by $V_T$. So this is basically the voltage drop across the resistance at the top, right? This is the resistance *Rc*, so this voltage is basically the drop at this resistance *Rc* and that is equals to minus $V_{RC}$ by $V_T$, right? Now $V_{RC}$ is equal to $V_{CC}$ by minus $V_{CE}$, remember why?

Because you had a resistance here and you had a, this thing therefore we get $V_{CC}$ is this one, right? And so $V_{RC}$ is this voltage, between these two points and this is $V_{CE}$, right? So I can write down $V_{CC}$ to be equals to $V_{RC}$ plus $V_{CE}$, fine. And therefore we can write down $V_{RC}$ to be equals to $V_{CC}$ minus $V_{CE}$, right? So in saturation this $V_{CEsat}$ will be typically very small and $V_{RC}$ will be approximately equals to $V_{CC}$ and you get a very high gain, right?

So, But then you have to ensure that you don't go into saturation because then $V_{CE}$ sat will be typically very small. So where do you say? You stay within the active region of the operation of the device. So therefore I get $A_v$ to be equals to minus $V_{CC}$ minus $V_{CEsat}$ by $V_T$ and therefore the maximum value of $A_{Vmax}$ is $V_{CEsat}$, so if I assume this to be equal to 0 approximately I get - $V_{CC}$ by $V_T$ as the $A_{Vmax}$.

So you see even if I use $V_{CC}$ of 1 volt and if 25 mVs you are getting here, so I will get 1 by $25 \times 10^{-3}$, so I get $10^{+3}$ here, so this is basically thousand by 25 which is basically 40, right? So even if you don't do much manipulation and incidences, what you get is primarily that the voltage gain of a common emitter mode configuration at *T* equals to 300 K happens to be maximum value attainable is basically 40.

But then this is assumption that $V_{CEsat}$ equals to 0 which is never. So actually I will get a drop in the value of $V_{CE}$ in the $A_V$ value. In reality $A_V$ will be much smaller than 40 even, right? And that you have to take care of when you're actually designing a system. As I discuss with you therefore, the gain is negative, right? The gain is negative which signifies that the amplifier is inverting amplifier.

So I have an inverting amplifier available with me and the gain is negative this thing. Further the gain is proportional to the collector bias current $I_C$ and a load resistance $R_C$. This we have already seen and you know the reason why the gain is like the collector current and basic reason is there. So before we finish this module of lecture let me recapitulate what we did. We had looked into BJT as an amplifier.

We have also understood the various parameters which we need to fix from DC domain in order to bias it in the active region of operation. You cannot bias it in a saturation or in the cut-off region, you have to bias it only in the active region. When you bias it in the active region you automatically get a very large gain. We also saw that theoretically speaking the maximum gain available to us in a common emitter mode configuration is minus 40 times, right?

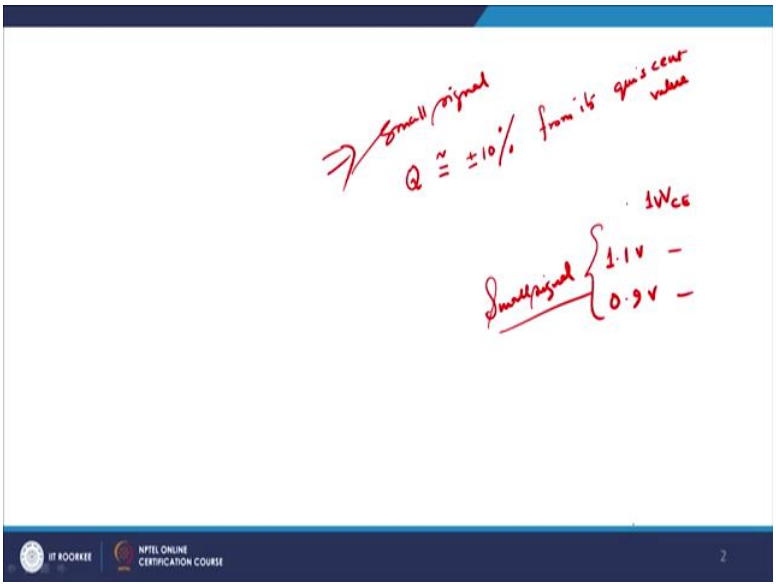
Provided you are giving a $V_{CC}$ of 1 volt, I get $V_{CC}$ by $V_{TS}$ the maximum theoretical limit which you get in the voltage gain. In reality, obviously we will not get because $V_{CEsat}$ will never be equal to 0 it will have some finite value and therefore your $A_{Vmax}$ will reduce drastically in that case, right? So this takes care of the basic understanding of our design and in the next turn we will be looking into small single models BJT, right? Thank you very much.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−08**
**BJT: Small Signal Circuit Model-I**

Welcome to the NPTEL online certification course on Microelectronics: Devices and Circuits. This module we will be taking care of a Small Signal Circuit Model for BJT. The obvious question asked is, why should we require this? Well, BJT as a stand-alone device has got no usage or application, so this has to be integrated onto a circuit for example BJT as an amplifier, BJT as a switch, BJT as a current source and so on and so forth.

Therefore, if you have understood how BJT works which we have already done in our previous interactions, we will understand or we will together learn how this BJT the device can be broken down into sets of active and passive elements, so that it is very easily integrated with the circuit analysis. And it is easier with circuits and it is easy to do the analysis in circuits in a much detail manner.

(Refer Slide Time: 1:45)




Why small signal? We will discuss why small signal now. What is the meaning of small signal? Let me first make it clear to you. A small signal primarily means that the Q-point does excursion of approximately plus minus 10 percent from its quiescent value, right? So if you have biased your device then we define this to be as a small signal, right? So if you have biased your device say at 1 volt, $V_{CE}$, right? 1 volt then if the input signal changes it from

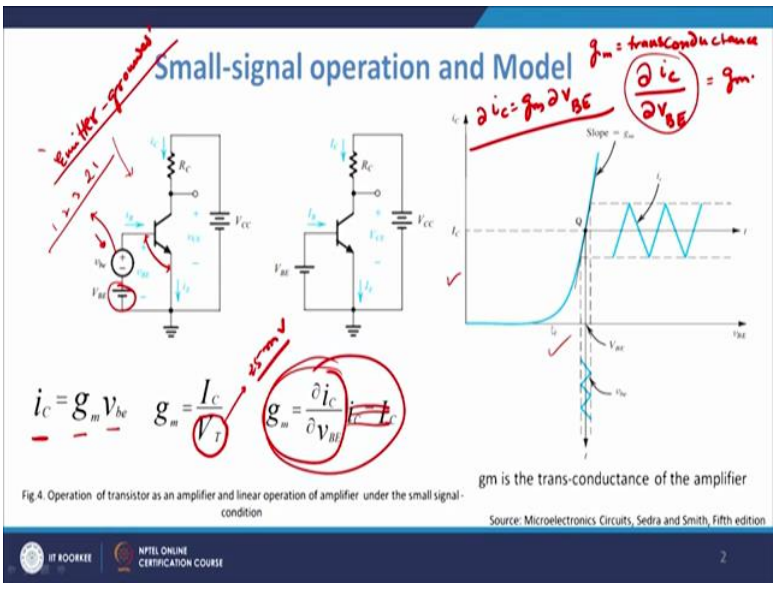
1.1 volt to 0.9 volt peak-to-peak, I would expect to see that, we can define this to be as a small signal.

This is a rather not a very decent definition but nonetheless works fine for all circuit applications assuming that this much amount of variation in the voltage, input voltage will not force the device to go into either the saturation or cut-off region. So we have biased Q somewhere in the middle of active region, a 10 percent high rise or fall will result in approximately a linear change in the value of your amplification, right?

So you are still in the linear region of amplification as far as Q point is concerned. So two things one is a small signal circuit model primarily means that I am assuming that my input signal is basically small signal which means that is just a plus minus 10 percent of quiescent value and the second point which I am assuming is that there is no nonlinearity into the picture by virtue of a large signal, right?

So and the third point is that obviously we require to have a corresponding circuit elements being inserted, so that whenever we plug those BJT and use it as a switch or any device we can easily do that by doing small amount of manipulation in terms of circuit application. So let's see what the issues are.

(Refer Slide Time: 3:54)




So let me come to that point in a detailed manner. I will take one by one as we move along. If you see clearly again this is, as you can see this circuit is basically a BJT in an emitter grounded, so this is basically your emitter grounded configuration or even a common emitter configuration, emitter grounded configuration or common configuration. And what you see

from here is that $V_{BE}$ is the applied AC bias which you are giving and this $V_{BE}$ is responsible for biasing it in the appropriate Q-point.

We define therefore $I_C$, the output current to be equals to $g_m$ times $V_{BE}$. So if you understand $V_{BE}$ is the voltage which you have applied here, right? And this $V_{BE}$ is varying voltage because that is what the input voltage is all about. So one time it is 1 volt, 2 volt, 3 volt then again comes to 2, 1, so on and so forth. So it is a rising and falling maybe a sine wave maybe a pulse waveform or maybe any other wave forms available to you.

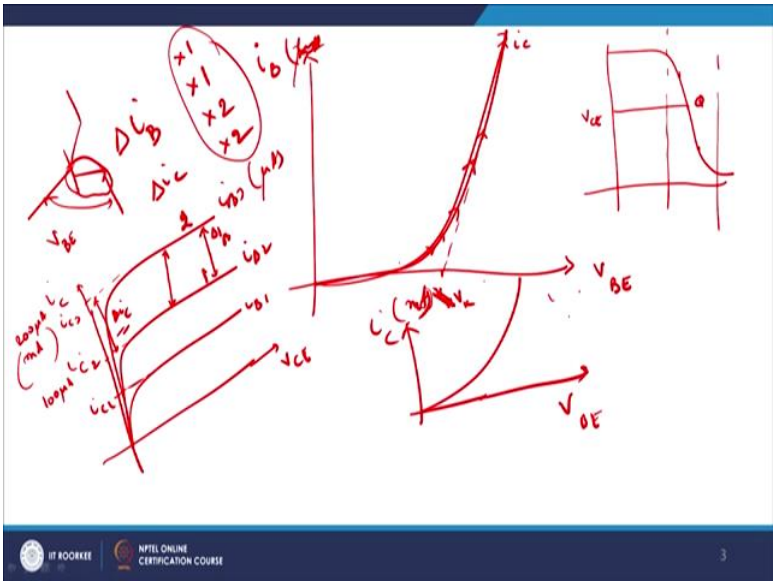
And therefore this base to emitter voltage which you give that converts into a corresponding current on the collector side, how? By mathematically multiplying with a value which is $g_m$. $g_m$ is referred to as a transconductance, right? And it is defined as, in this case as $\partial_{i_c}, \partial_{V_{BE}}$, right? So it is defined as rate of change of collector current with the base of voltage. So once you get this to be as $g_m$, right?

I simply get $\partial_{i_c}$ to be equal to $g_m$ times $\partial V_{BE}$, so higher the value of $g_m$ of a transistor, I would expect to see a larger amplification appearing to me, right? Because higher the value of $g_m$ will primarily mean that the transistor is more sensitive to variations in the input voltage, right? So $g_m$ is basically a notation or a parameter which gives me how sensitive your BJT is all about.

So if your BJT is quite sensitive, I would expect to see a large value of transconductance available to me. So I define $g_m = \frac{i_c}{V_{BE}}$. So if you see it is $i_c$ by $V_{BE}$ and this is replaced by $V_T$ for the lowest case because this is thermal energy which you get which is 25 mVs at 300 K, right? You can also write down, therefore $g_m = \frac{\partial i_c}{\partial V_{BE}}$ for the fixed value of $I_C$, right?

So at a fixed value of $I_C$ what is the change in this thing? But you need not forget about all these things, this you just forget. $g_m$ is basically $\partial i_c,\ \partial V_{BE}$ which you see, right? Now if you plot $I_C$ vs $V_{BE}$, this versus this, let's see why and how we get $I_C$ vs $V_{BE}$ in this graph.

(Refer Slide Time: 7:07)




So for common emitter configuration if you plot say $V_{BE}$, right? And you plot say, maybe an $i_B$, let us suppose, right? Let us see how you plot $i_B$. If you plot $i_B$ it will look something like this, right? We have discussed this in our previous discussion, this is the threshold voltage or the knee voltage for the PN junction diode which is the 0.7, right? So for $i_B$ if you have a common emitter configuration in this case then, this is the value of $V_{BE}$, right?

And when this $V_{BE}$ crosses 0.7, right? This fire this junction, switches on this PN junction and as a result the current starts to rise, right? Below that particular point the current is very -very low or very small and drops very drastically. So $i_B$ with $V_{BE}$ is in this manner. We have also seen from our previous discussion that if you bias your device in such a manner that your Q-point is here, so this is your active region.

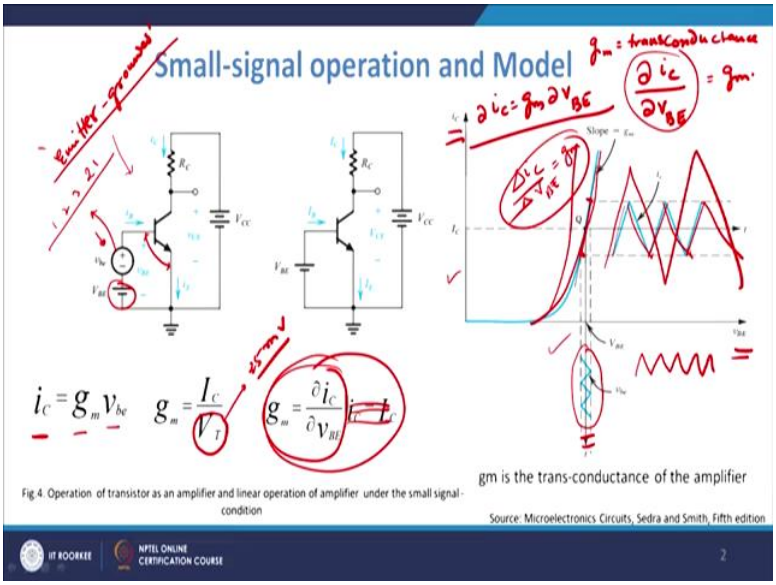
If you bias your device in the Q-point here $V_{CE}$ let us suppose, right? Then I can safely say that for the same change in the value of $i_B$, I would expect to see same change in the value of $i_C$. So remember when we are plotting the graphs, right? For various values of $i_B$ when we are plotting $I_C$ vs $V_{CE}$, right? The common emitter configuration, for various values of $i_B$.

So it is $i_{B1}$, $i_{B2}$, $i_{B3}$, so you see this is $i_{B1}$, this is $i_{C1}$, this is $i_{B2}$, $i_{C2}$, this is $i_{B3}$ , $i_{C3}$, so this $i_B$ is of the order of few µA and this is of the order of few mA, so thousand times increases the approximately between base current….. but..but please understand that for the same increase, so if this is say $i_{B2}$ is say 1 micron, this is 2 micron let us suppose, then this current will be actually if suppose 100 times larger, so I get 100 microampere (µA) here or 0.1 mA and I get 200 µA.

So you see for the same change in $\partial i_B$, right? I get the same change in $\partial i_C$, so $\partial i_B$, right? And $\partial i_C$ if you want to find out the rate of change is always the same. So for the same increment in $\partial i_B$, I get equal increments in $\partial i_C$, right? The increment themselves might be different obviously 100 times larger so quantitatively they are different but I am 100 percent sure that if I increase the base current by once my collector current will also increase by once.

If I increase the base current twice my collector current will also increase by twice, why? Because I am in the linear region of operations, right? And therefore I wanted to tell you is, that this is $i_B$ vs $V_{BE}$ exactly the same graph is for $i_c$ vs $V_{CE}$. So if you plot $i_C$ vs $V_{BE}$, you will get the same curve again, right? Only thing will be that this $i_C$ will be mA which was here it was in µA or may be mA as well but this will be more on mA, right?

(Refer Slide Time: 10:49)



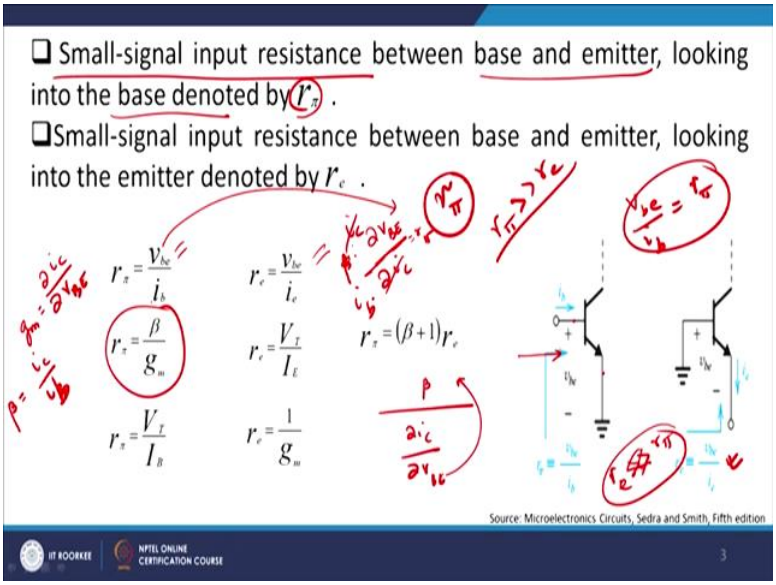

So let me show you therefore if you plot $i_C$ vs $V_{BE}$ here and then in input side if you vary this is your input curve $V_{BE}$ is varying, it is a triangular waveform which you have given, this triangular waveform again small signal assuming so that, the Q is just shifting from this side and coming to this side. So the maximum value is Q is this and this. So we are still in the linear region of operation of the transistor. I get this to be the output, right? Triangular output waveform will be there, right?

Now, the slope of this curve which you see, this curve, the slope of this curve, sorry, the slope of this curve, right? Which is $\partial i_C, \partial V_{BE}$, this is basically your $g_m$ transconductance. So if this slope is not steeper, I would expect to see a much larger variation in the output waveform, for the same change in the value of input waveform, fine. So that's the small

signal model and in the small signal model therefore we have learned one important point and that important characteristics is transconductance ($g_m$) and $g_m$ is therefore defined as the transconductance which in this case happens to be $\partial i_C, \partial V_{BE}$, right?

And it gives me an idea about sensitivity. So if I know the base to emitter voltage I can predict the value of the collector current directly by multiplying $g_m$ with the rate of $V_{BE}$, right? And that's quite an interesting phenomena which is there with us.

(Refer Slide Time: 12:21)




And now we define another resistance known as the small signal input resistance and it is referred to as $r_\pi$, right? And it is nothing but the resistance looking from the base side where between base and emitter, right? So looking from the base side, if you're looking from the base side the resistance offered by this is basically $V_{be}$ by $i_B$. So $V_{be}$ is the voltage which you see between these two points and $i_B$ is the current by virtue of this voltage.

So $V_{be}$ by $i_B$ is equals to $r_\pi$, right? So it is basically the forward biased on resistance of a BJT, right? It is an on resistance which means that the base emitter junction is forward biased it is getting on and therefore the resistance offered is defined as $r_\pi$, right? Sorry! But plus understand, this you are looking from the base side, right? So you are looking from this base side and you are trying to find out.

So your emitter is grounded, please understand. Another is $r_e$ which is basically the base to emitter resistance looking from emitter side. So you give a potential on the emitter side and you ground the base side and then looking at that, you will get the value of $r_e$ to be equals to

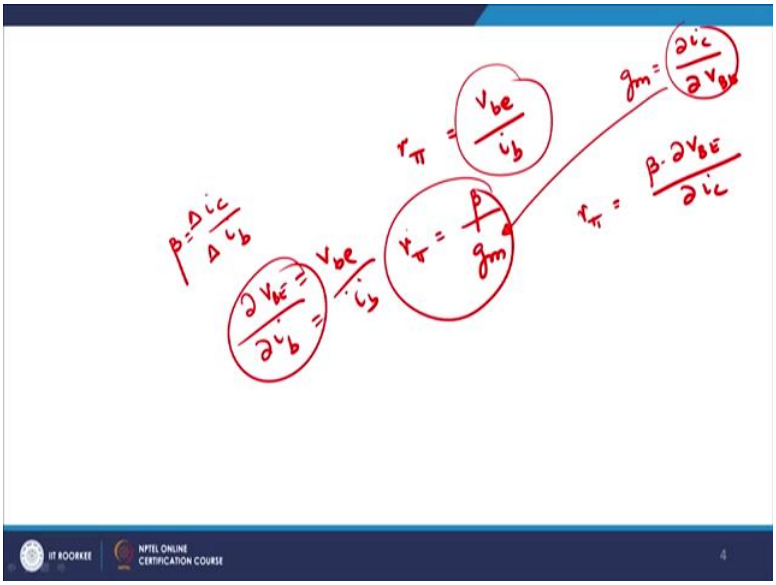
$V_{be}$ by $i_e$, right? So it is $V_{be}$ by $i_e$, right? Obviously as you can find out $i_e$ is much-much larger as compared to $i_B$. So I would expect to see $r_e$ to be much-much smaller as compared to $r_\pi$.

So $r_\pi$ will be relatively larger as compared to $r_e$, fine. So I get $r_\pi$ is equals to therefore $V_{be}$ by $i_B$ this is also equals to $V_{be}$ by $r_e$, right? So $r_\pi$ can be written as, so what we can write down is that, it is equal to $\beta/g_m$, right? How $\beta/g_m$ we will just show it to you. If you remember $g_m$, what was its definition? It was, $\frac{\partial i_c}{\partial V_{BE}}$, right? So if you put $\frac{\partial i_c}{\partial V_{BE}}$ here, so I get what?

So I get $\beta$ divided by $\partial i_c$ by $\partial V_{BE}$, so this $V_{BE}$ goes to the numerator I get $\partial V_{BE}$ at the top, right? Divided by into $\beta$ divided by $\partial i_c$, this you get, right? This you get. So this $V_{be}$ is coming from here, so now you see remember the definition of $\beta$ from a previous discussion, $\beta$ was equal to your $i_c$ by $i_b$, remember. So if you put $i_c$ by $i_b$ here I get $i_c$, so I replace $\beta$ and make it $i_c$ by $i_b$, right?

This $i_c$ , $i_c$ gets cancelled out I get $V_{BE}$ by $i_b$, so this is equals to $r_\pi$. So $r_\pi$ equals to $\beta$ by $g_m$ also we can write.

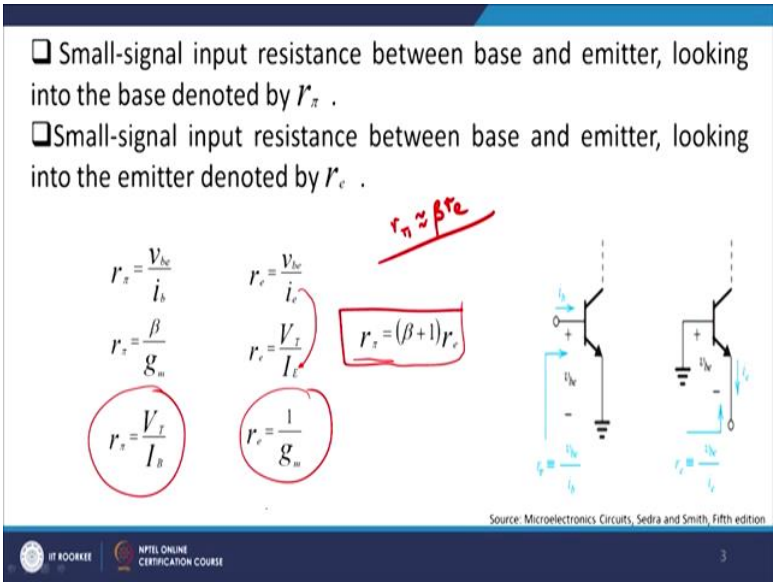
(Refer Slide Time: 15:52)




So I will do it once again if you want to do. $r_\pi$ is given as $V_{be}$ by $i_b$, remember. So we also know $r_\pi$ has been also given as beta by $g_m$. We know $g_m$ to be equals to what? $g_m$ to be equals to $\partial i_c$, $\partial V_{BE}$. So if you put the value of $g_m$ I get $r_\pi$. If you at place this equation here, I get $\beta$ times $\partial V_{BE}$ divided by $\partial i_c$, right? Now we know the definition of $\beta$ as the current gain in common base configuration and therefore I can write $\beta$ to be equals to $i_c$ by $i_b$.

So $\partial i_c$, $\partial i_b$, let me write down. Place this value of $\beta$ here and therefore this $\partial i_c$ and $\partial i_c$ gets cancelled out, I am left with $\partial V_{BE}$ upon $\partial i_b$, right? And that's what you are getting here, fine. Since it is a small signal therefore $\partial V_{BE}$ equals to $V_{be}$ and $\partial i_b$ equals to $i_b$ and therefore this is valid assumption which we get. So I get $\beta$ by $g_m$ as the $r_\pi$ value, fine. So $\beta$ is relatively very high value and therefore $r_\pi$ will be very-very high, so this is what we get from our discussion as far as this is there.

(Refer Slide Time: 17:16)



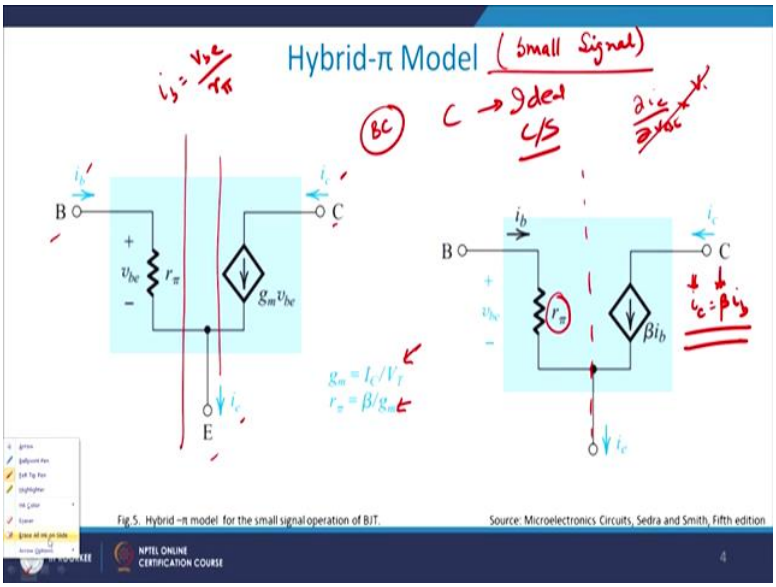

Similarly $r_e$ therefore can be written as $i_e$ is the emitter current and $V_T$ is the base to emitter voltage at 300 mV we will write it as $V_T$ thermal equivalent voltage. Therefore I can write down that $r_\pi$ could be written as $\beta$ plus 1 $r_e$, right? And you can do it yourself to find out this value that means $r_\pi$ is approximately $\beta$ times larger as compared to $r_e$, right?

So…So… Which means that, and it is expected also, if you look very carefully, the reason being, the base is, so looking from the base side the resistance is very large as compared to looking from the emitter side and you can understand why is it? Because base has got relatively low doping, right? And the area is also very small. So in both will result in very - very high resistivity. Resistance will be typically very high and that's the reason $r_\pi$ is much larger than beta re approximately 100 to 200 times larger, right? And that the reason why you get it.

Therefore $r_\pi$ can also be written as $V_T$ by $I_B$ in case you want to do it and $r_e$ can be written as 1 by $g_m$ as well, as you mean $\beta$ is approximately equals to 1, I can get to be approximately these. These approximations can be taken care of or it can be handled easily. So we have

understood therefore that for small signal model if you're looking at the base side, emitter side resistance is not equal from the base at resistance is much higher as compared to that from the emitter side.

(Refer Slide Time: 18:40)




We come to the models, various models of BJT which is available. Small signal model is, these are again small signal models, these are all small signal model which is available, right? And they are quite interesting and very easy to handle from circuit point of view once we know the basic fundamental principles of working of these devices. Look at, so we have already learned that BJT is basically a three terminal device.

That everybody knows that BJT is obviously a three terminal device, so what are those three terminals? Base, collector and emitter, right? So the arrow head which you see, the blue arrow here, the blue arrow here, the blue arrow here shows you the direction of the conventional current, right? And therefore we say that looking from the base side the amount of resistance offered is $r_\pi$ and across it we have a $V_{BE}$ as the voltage drop available to me.

So therefore $i_b$ can be directly written as $V_{BE}$ divided by $r_\pi$, right? $V_{BE}$ is always fixed approximately 0.7 is supposed the approximately voltage. If I give you therefore the $r_\pi$ value you can predict the value of $i_b$ directly, right? Now this is therefore the base emitter junction, right? Let's look at the base collector junction. We have already discussed that base collector junction, the collector starts to behave like an ideal current source, right?

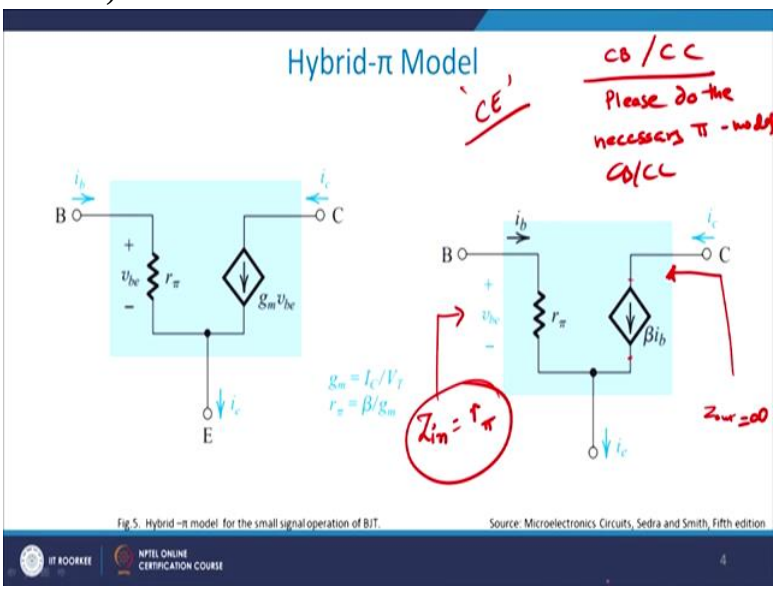
So it gives us ideal current source. How much is the value of current? Will be given by $g_m$ times $V_{BE}$. Remember $g_m$ was equal to $\partial i_c$, $\partial V_{BE}$. So $\partial i_c$, $\partial V_{BE}$ multiplied by $V_{BE}$, so $V_{BE}$ gets

cancelled out and I get the current available to me, right? So this is the current source with me and this is the direction, the direction of the arrowhead of the current tells me that the current is directed inwards which means, it is basically an NPN transistor and therefore it is directed inwards and therefore you see this directed inwards whereas emitted current is directed outwards, right?

So by the previous definition we know that gm is equal to $I_c$ by $V_T$ which is available here and this is $\beta$ by $g_m$ is $r_\pi$, right? So it's completely analogous to see on the right-hand side, this one, right? It is also a hybrid model, hybrid-π model. Now, it looks like π, so that is the reason it is hybrid-π model. Moreover it is made up of $r_\pi$ and therefore it is a π model. So you remember, $r_\pi$ is the resistance offered by the base side, base emitter junction and therefore we have $r_\pi$ as the resistance here.

Whereas on the collector side it is the current source, wherein current source $i_c$ will be equal to β times $i_b$, right? So if you have a large value of $i_b$, I will get a large value of $i_c$ and I can therefore have a constant current source towards the collector side, right? Therefore this model will give you a very first instance the input impedance and output impedance of the device, right?

(Refer Slide Time: 21:50)




So if you look at the BJT therefore the output impedance is basically mine, $Z_{out}$ if you want to find out, is looking from the collector side this is the output. How much amount of resistance is offered? Of course you can see, from the output side if you look very carefully it is being

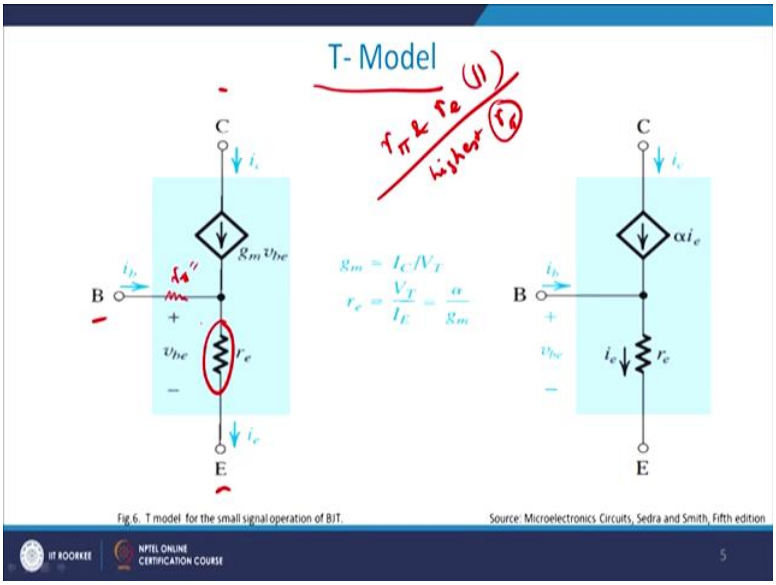
terminated across a constant current source and therefore $Z_{out}$ will be approximately equal to infinity in idealized condition.

So if you want to follow a hybridized pi model, then $Z_{out}$ will be equal to infinity and $Z_{out}$ will be equal to infinity primarily meaning, the reason being that, we are looking into the current source here, an ideal current source, right? And assuming it to be an ideal current source, we are assuming that the $Z_{out}$ will be the infinity. So whereas looking from the input side $Z_{in}$, input impedance will be approximately equals to $r_\pi$, fine.

So this is for the common emitter configuration, right? I leave as an exercise for you to find out common base and common collector, please do it, right? Please do the necessary, right? π model for *CB* and *CC*, common base and common collector, right? In common base, you will see it will be re which will be coming into picture as the input impedance, right? Therefore just be able to handle this hybrid-π model in a much better manner in this case, right? So this is one model.

The second model which is most prevalent but most of the time we use of hybrid-π model for most practical purposes but sometimes we also use what is known as a T model, right? This is a transmission model or T model which you see, this is the T model which is there.

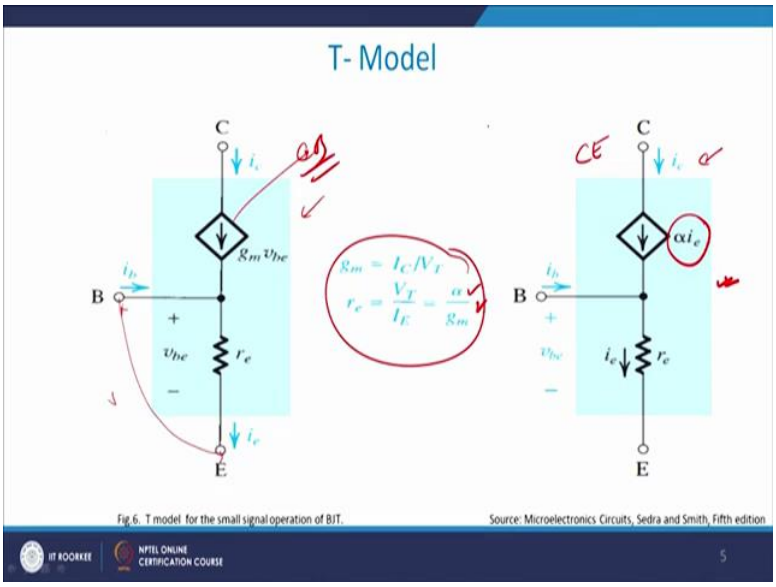
(Refer Slide Time: 23:31)




And the T model, still again three terminal device so base, collector, and emitter. We have also seen that the resistance between emitter and base is $r_e$ looking from emitter side. Looking from the base side, it would have been here $r_\pi$ but since we are looking from the emitter side at this case we are giving $r_e$, you can also have $r_\pi$ here but they will be in parallel.

So being parallel it will be always less than the least, remember? And therefore this has been neglected. So you see $r_\pi$ and $r_e$ will be in parallel. Since they will be in parallel the maximum, the highest value will get neglected. So highest is nothing but $r_\pi$ and therefore $r_\pi$ is not shown in this diagram, right?

(Refer Slide Time: 24:16)




So therefore understand that why $r_\pi$ is not here or here or here, because it is so large that it does not play a major role, $r_e$ plays a major role. So I get $r_e$ here, the voltage across $r_e$ is basically my $V_{BE}$, base to emitter voltage and this is the applied voltage which you see the current direction is of course, the conventional current direction as we have been discussing quite often.

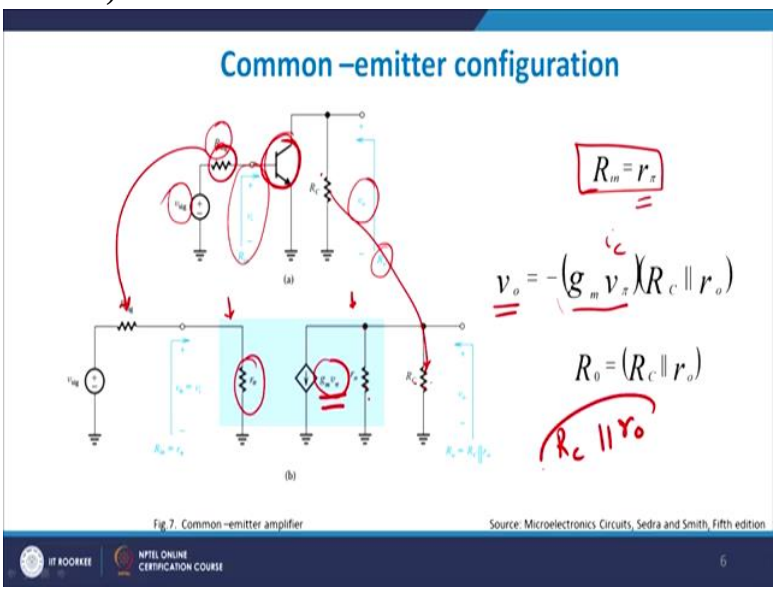
And therefore, the collector current will be equal to $g_m$ times $v_{be}$, right? We have again discussed this point as the input point. So... So… you see, what I'm trying to tell you, is your base emitter junction or your base emitter voltage, right? $v_{be}$ starts to change the value of your current flowing through collector, right? And that is basically therefore a control current source with you, right?

So basically, it is a current control current source because $i_b$ will change the value of $i_c$ and therefore it is basically known as the current control current source. So I get from here that it is $\alpha$ by $g_m$ is the value of $r_e$ which we get. $\alpha$ by $g_m$ you can prove it yourself to get the value of $r_e$. $\alpha$ is basically the common emitter current gain, right? Common emitter current gain will mean that you will have $r_e$ approximately just to $\alpha$ by $g_m$, right?

And you can get the value of $\alpha$ and $g_m$, it is typically very low because, $\alpha$ will be hardly of the order of 0.99 or 0.98 and this $g_m$ will be typically large because sensitivity is very large, so your $r_e$ will be typically very small quantity. So your resistance looking from the emitter side will be very small, right? And that makes it go to a very small state or a very small region.

And therefore as you can see on the right-hand side if you look very carefully from the common base configuration I get $r_e$ I get $\alpha$ times $i_e$ as the collector current and therefore, for a common emitter configuration, so this is common emitter consideration *CE* and this is for the *CB* configuration and therefore I can get the value of this thing. So I will recommend that for both hybrid-π model as well as, T model for all the three configuration *CB*, *CC* and *CE* which means common base, common collector and common emitter please try to find out the value and the model itself please try to get it from any standard books which are available in the market, right? Right?

(Refer Slide Time: 26:34)




Let me therefore now come to the basic, we will finish with this common emitter configuration. Let me come to the common emitter configuration now and explain to you what common emitter configuration how it works, as far as, how we are inserting the small signal model in a common emitter configuration, right? So this is the application of the small signal model, right? So what is the application?

The application is something like this that I have a signal input resistance, I have an input signal, I have BJT which is there, $R_C$ is the resistance, so if you look $r_n$ is basically, which is this one, is the input resistance looking from the base side and output resistance looking from

the collector side is $R_o$. So what we are doing, we will replace this. So if you look out of the blue box.

This blue box, out if you look, this $R_C$ is this $R_C$, so this $R_C$ is this $R_C$, right? This $R_{sig}$ which you see is this $R_{sig}$, right? Signal. So looking from this side this can be broken down into an $r_\pi$, right? Because $r_\pi$ is already known to you and this will be $g_m$ times $r_\pi$, right? And $R_o$ will be the output impedance offered by the device which is there for the device itself, not for the circuit point of view.

So therefore I get $R_0$ always parallel to my current source here, right? Howsoever large or small we don't know but it will be the value which will be there $R_0$ here. So we get $R_{in}$ to be equals to $r_\pi$, no need to understand why is it, very simple, because looking from this side from the base side I only get $r_\pi$ as my impedance, so $r_\pi$ is the $Z_{in}$ value. $V_0$ is $g_m$ times $V_\pi$, $V_\pi$ is nothing but the voltage across this $V_{BE}$ or $V_\pi$ whatever you want to name, it is basically the output current.

This multiplied by $R_c$, so this $R_c$ and $R_o$ in parallel, right? So the affective resistance seen by the current is basically $R_c$ parallel to $R_o$, this if you multiply with $i_c$, I will get $V_0$, right? And I get $V_0$ here. So $R_0$ is nothing but $R_c$ parallel to $r_0$. So the output impedance is basically $R_c$ parallel to $r_0$, right? So this is the output impedance which you see. So for a common emitter configuration we are able to see the output impedance, right?

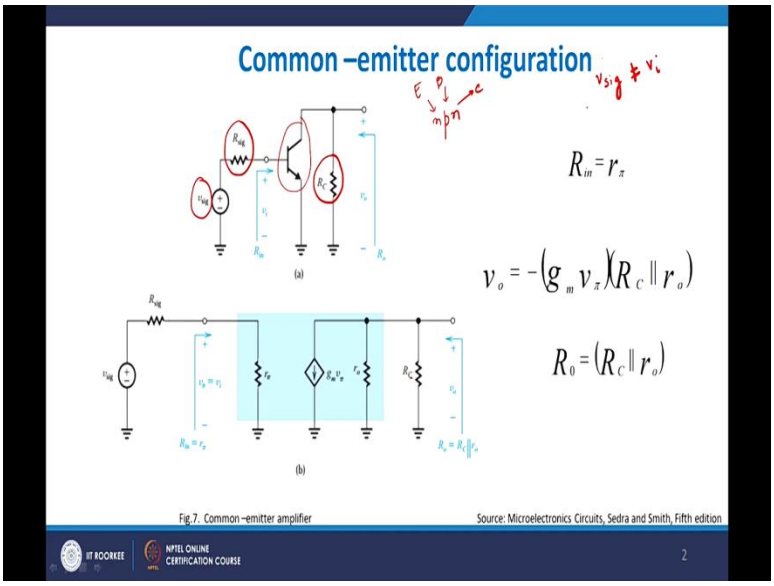
Okay, with this knowledge, with this basic knowledge will stop here today and what we have learned is basically the small signal input model for the BJT and in the next turn we will look into how we can apply further for the common base as well as common character configuration design. So we have learned how to bias a device, DC bias, super impose AC signal over it and we have also learned how to use those for the purpose of common emitter configuration using the small signal model, fine. I hope you have understood the issues taken up in this lecture, thank you for your patient hearing.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication**
**Indian Institute of Technology Roorkee**
**Lecture−09**
**BJT: Small Signal Circuit Model-II**

Hello and welcome to NPTEL Online Certification course on Microelectronics: Devices to Circuits. In our previous lecture, we had discussed with you the π model and the T model. We have also discussed with you the condition for small signal model and we have seen that the requirement for small signal model has been that, if your input signal is small enough so that your bias point does not change too much and you are in the linear region of operation of the device where amplification is almost independent of the input voltage, we define that to be as smallest signal. We have also understood that if I have a BJT in a common emitter configuration then the amplification is typically very large, of the order of few hundreds in Voltage Amplification.

Whereas common collector does not give me voltage amplification directly, right? We also understood about the relationship between $\alpha$ and $\beta$, where $\beta$ is basically the current gain and $\alpha$ is also a current gain in common emitter configuration and $\beta$ is common current gain in common emitter configuration and $\alpha$ is the current gain in common base configuration. We have seen the functionality and the operation of BJT. What we will do today is, have a look at small signal circuit model we will extend that to *CE*, *CB* and *CC* which means common emitter, common base and common collector.

(Refer Slide Time: 2:07)

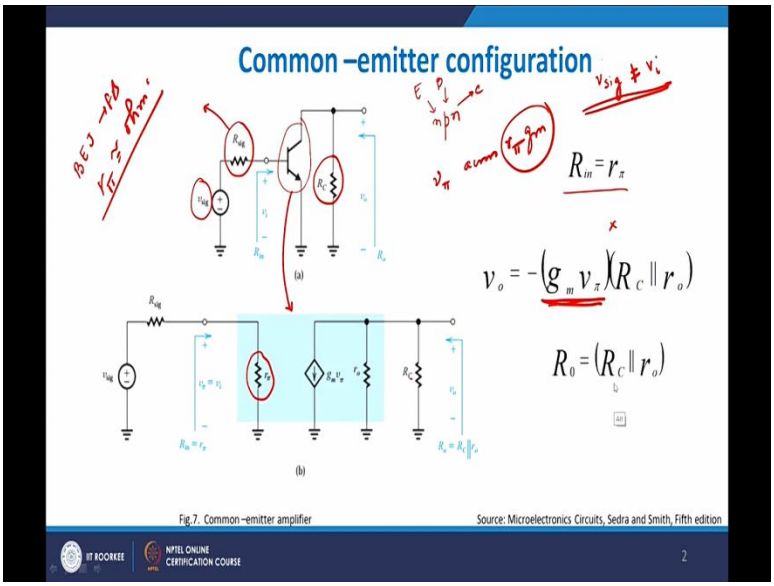



So, just to continue with yesterday's talk, we will come to this point that if you have a common emitter configuration which you see then, as I discussed with you in the yesterday's discussion or slide that, you do have a BJT here which is an NPN transistor looking at the direction of the arrow, this is basically an NPN right? So I have a p-type base, an n-type emitter and we have got a collector which is n-type, there is also a resistance '$R_c$' which primarily is a resistance which is seen at the collector side right? And you have an input resistance also referred to as '$R_{sig}$' here and we define '$R_{in}$' as the input resistance or input impedance and '$R_0$' as the output resistance or impedance and '$V_0$'as the output voltage whereas '$V_{sig}$' is my input voltage, input voltage.

Now you see that the two aspects which should be very clear at this stage, $V_{sig}$ is not equal to '$V_i$' right, so your input voltage to the BJT will be slightly less as compared to the actual signal voltage and the reason being the presence of this $R_{signal}$ right? Because these are all voltage sources right, and these voltage sources will have a finite output impedance and as a result, this output impedance will, there will be some drop across that output impedance and therefore the actual potential applied to the base emitter junction of a BJT will be relatively smaller, smaller than what is expected from here, right?

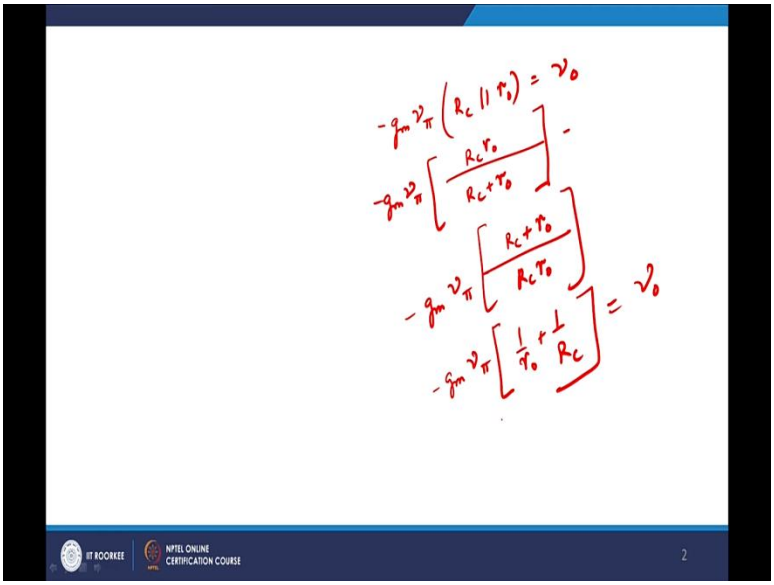
Now so we have therefore if you so this BJT is broken up into equivalent small signal model as you can see the blue box here, where you have $r_\pi$, $r_\pi$ is basically nothing but the value of $R_{in}$ and the reason being $r_\pi$ is nothing but the resistance between the base emitter junction, right so when the base emitter junction is forward biased this $r_\pi$ will be of the order of few ohms right to the order of ohms right, because it is a forward bias device characteristics.

(Refer Slide Time: 4:08)



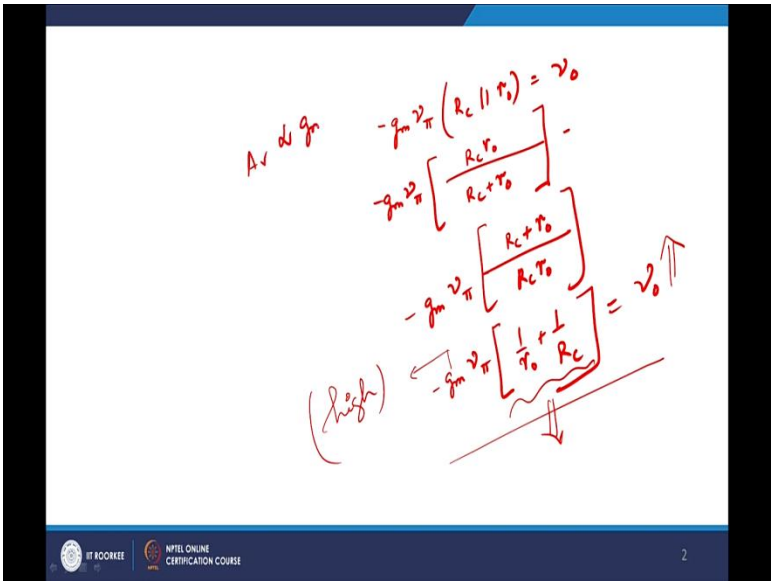

Now so….so…. the voltage across this $r_\pi$ is the voltage which is basically $V_{be}$ happens to be $r_\pi$ so $V_\pi$ is the voltage across across $r_\pi$ right? So this is sort of an input voltage which is visible to us and that when multiplied by transconductance ($g_m$) gives me the output current, right? And that is the count, so $g_m$ times $r_\pi$ or $V_\pi$ is my output current, that if you multiply with the effective resistance seen from the output sign. What is the effective resistance? $R_C$ parallel to $R_0$, right? $R_C$ parallel to $R_0$ and therefore I get $V_0$ equals to minus $g_m$ times $V_i$ into $R_C$ parallel to $R_0$.

(Refer Slide Time: 5:05)



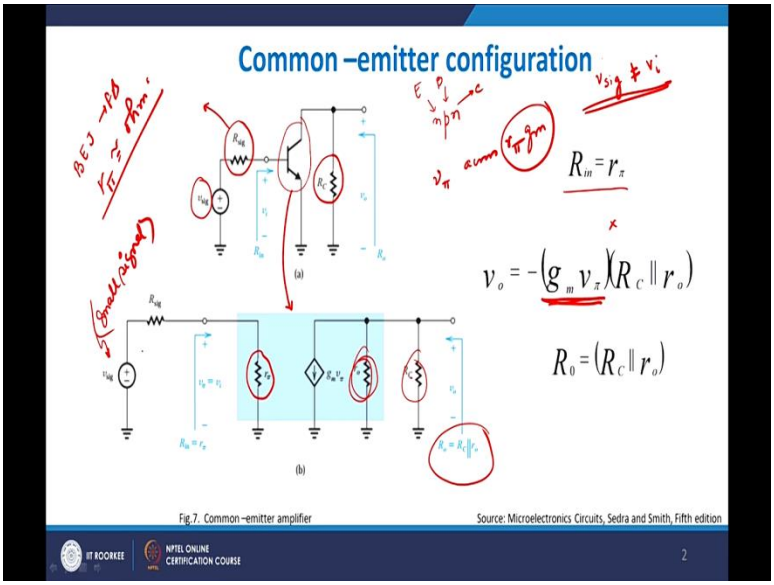

So if you look very carefully or understand it, then it is basically minus $g_m$ times $V_\pi$, right? $R_C$ parallel to $R_0$, this is the value of your $V_0$, right? See if you if you break it down $V_\pi$ this will be $R_C R_0$ upon $R_C$ plus $R_0$, right? This will be your so, basically the 1 upon this so minus $g_m V_\pi$ into $R_0$, $R_0$, fine? So you see therefore if you break it up, I get 1 by $R_0$ plus 1 by $R_C$, right? This is equals to your $V_0$ for our practical purposes.

(Refer Slide Time: 6:28)



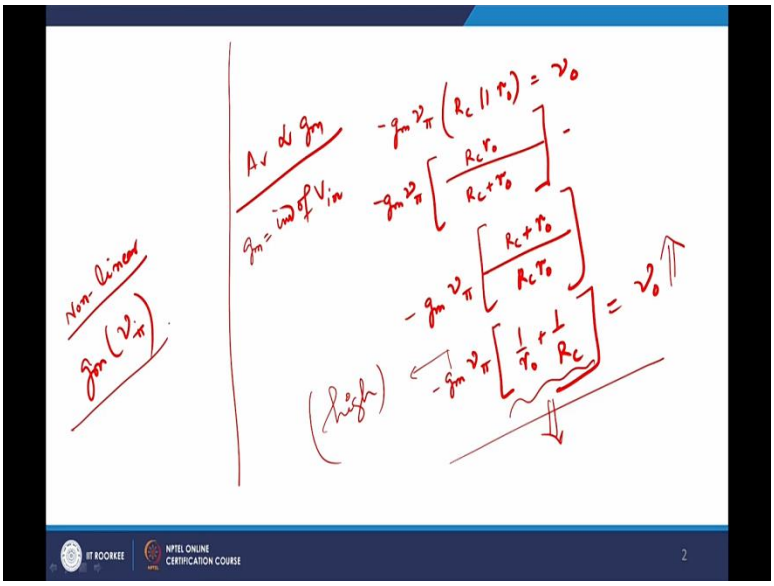

Now you see where $R_0$ at $R_C$ is the external potential, $R_C$ is the external bias which you have given and $R_0$ is basically the, the output impedance of the BJT, right? So the parallel combination of $R_0$ and $R_C$ gives you the net output impedance seen by the device and therefore I have written here that $R_0$ is equal to $R_C$ parallel $R_0$, right? And that gives you a fair idea about the type of voltages which you will get. Since $g_m$ is relatively high right, so $g_m$ is high. Even if this quantity is low quantity right, even if it is low I will get $V_0$ which quiet high, right?

(Refer Slide Time: 7:12)



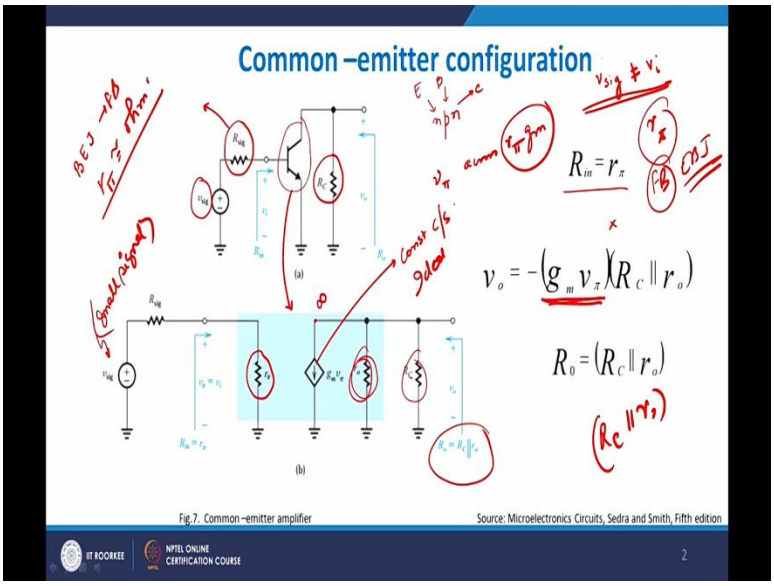

So therefore, we can safely assume that that the gain is directly proportional to $g_m$ which is an understood phenomena, that it is actually depending on the value of $g_m$, so higher the sensitivity is for a BJT to work, more will be the gain and therefore more will be the output voltage with the same amount of input voltage but please keep in mind that the $V_{sig}$ which you are using the signal input signal voltage is basically a small signal, right? Please keep this in mind for all practical purposes. This is basically small signal right and therefore these are all small signal models.

(Refer Slide Time: 7:43)




Now if your signal is large then you entail a non-linearity into your $g_m$, so assuming $g_m$ to be constant will be an over assumption, right? So one important point when you do a non-linear analysis or a non-linear profiling which is which we are not doing here, non-linear. Then $g_m$ will be actually a function of $V_{in}$, in this case $V_\pi$, right? But in reality or at least for the all understanding purposes basic purposes, we assume that $g_m$ is sort of independent of independent of $V_{in}$, right? In reality $g_m$ might be also a function of $V_\pi$ and therefore you will entail a non-linearity into the whole system. So…so… therefore we have understood that you do have a problem of nonlinearity also coming into picture, but if we sustain a small value of input voltage there will be no nonlinearity and I would expect to see a gain which is totally independent of $V_{in}$ and linear gain will be available.

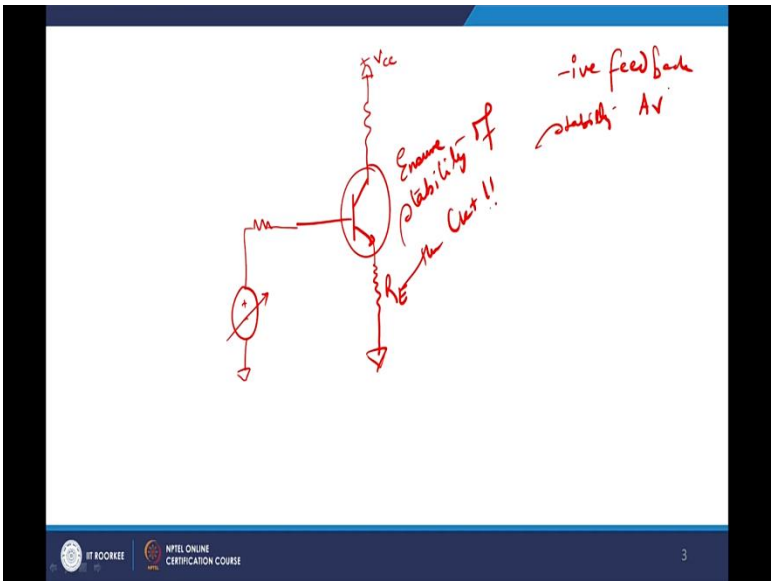
(Refer Slide Time: 8:26)




Two things we therefore, we take out from this slide is that my input impedance is $R_\pi$ and output impedance is basically $R_C$ parallel to $R_0$ right? Now so your input impedance is basically $R_\pi$ which is nothing but the forward biased EBJ right, Emitter Based Junction. The resistance offered by that is basically my EBJ, right? Second thing is, there is also assumption that this is behaving like a constant current source right, constant current source and we are also assuming it to be ideal. So the voltage, so the impedance offered by this is infinity. But since this is parallel of course therefore this does not play into the picture. I think it is clear to you right?

So, what we have therefore assumed is that the current source is basically an ideal current source as a result, the output impedance is infinity and, Since it is a constant current source and the output impedance is infinity, when it is parallel to $R_0$ and $R_C$ right, so that vanishes off because it is quite high a quantity, right. So in parallel combination that does not stay to a large extent right, obviously because in parallel you remember it is less than the least in the series and it is larger than the largest. In this case it is less than the least and therefore we just neglect that particular case, right? Okay.

(Refer Slide Time: 9:57)



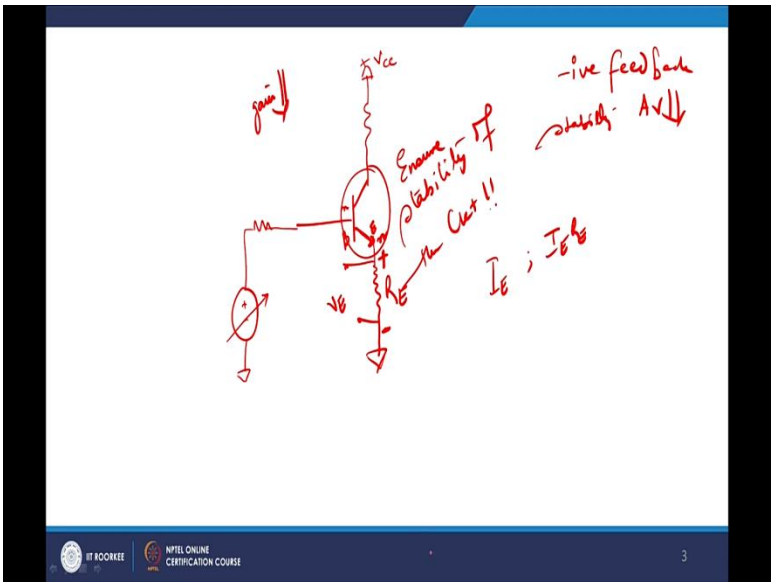

We now move into the next slide and see what happens with emitter resistance. Let me see why do you at all needed emitter resistance, very important issue is there when we are having an emitter. Till now if you remember we were actually looking from the point of view or the fact that we had a BJT right? And there was a resistance here, fine and you had a bias here and this bias was varying and you did it like this, like this and then we have a plus $V_{CC}$ and then we are doing like this sorry, we are doing like this, this is $R_E$.

Now what we are trying to tell you is, that no let me do small, a change here and the change is that we try to do some change in emitter configuration and what is the change, change is that we try to put a resistance here, why we will see that later on. So let me put a resistance $R_E$ here right, emitter resistance, all right. Now see, the reason we want to do it is to ensure what is known as stability, ensure stability of the circuit and I will explain that to you how I get it.

See if you remember from your basic class tenth even undergraduate days that or you can understand that whenever you have a negative feedback right? Whenever you have a negative feedback, you automatically have a more stable system, right. The (pay) the price you pay for it is here, gain starts to fall down, right?

So if I do a negative feedback in voltage domain which means that the output feedback, the output voltage is fed into the input voltage, input side with a 180 degree phase shifted value then, I can safely assume that my overall gain will reduce because obviously there will be principle of superposition can be applied and the two voltages will be added algebraically as a result the overall voltage bit falls down but then, we also ensure that by doing a negative feedback that the overall gain is more stabilized which means that the gain with respect to change in the input is much smaller.

(Refer Slide Time: 12:13)



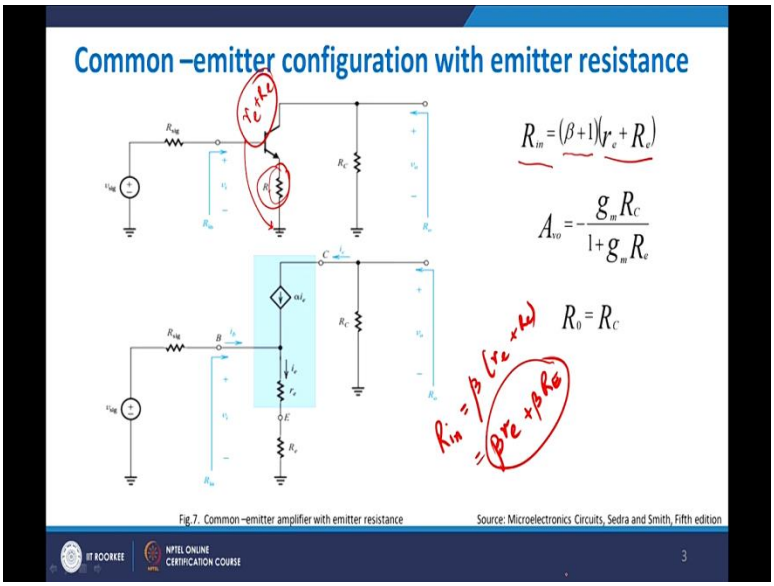

So, negative feedbacks actually gives you a stable system that is what we have learnt and we know all of us possibly know this basic concept. Now if this $R_E$ is in series to this external emitter resistance, emitter resistance is applied to, to the side of emitter then when a current $I_E$ flows through it then there is a voltage drop which is $I_E$ times $R_E$ in which this is positive and this is negative, right? So the voltage across this which is let us suppose we assume it to be as $V_E$ appears because of $I_E$, $R_E$.

This $V$ will start to reverse bias my this base emitter junction right, as it starts to reverse bias it is because you are applying positive side to n-type and maybe a negative side to p-type right, or even a positive bias to the n-side and as a result what is happening is that the emitter is getting

more and more positive which means it is trying to reverse bias the emitter base junction and therefore majority current carrier is reducing and you have a large amount of minority carrier which appears.

Therefore, when you apply $R_E$ your gain starts to fall down, right? And it starts to go down and down. So when you put an emitter resistance $R_E$ in series to the BJT we end up having a reduced gain right, a reduced gain but a more stabilized, stabilized circuit. So I have a reduced gain but more stabilized is with me. So that is the reason why we do, we bias with emitter resistance.

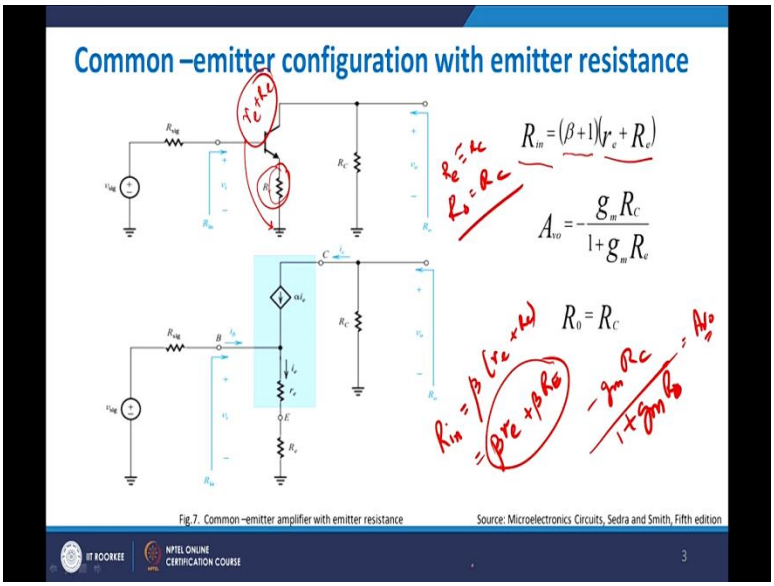
(Refer Slide Time: 13:49)




Now if you look very carefully therefore, in this case we have an $R_E$ here which is basically this one and therefore $R_{in}$ is given as $\beta$ plus 1, $R_E$ plus $r_e$ , plus $R_E$ right, where capital $R_E$ is basically the resistance, external resistance and $r_e$ is the internal resistance. So this so if you go from this point to this point, it is basically $R_E$ plus $R_E$ which you see in front of you right, that you multiplied with $\beta$ plus 1 you get total $R_{in}$ value right, because that is the basic definition of a $\beta$ now. $\beta$ will be very much larger than 1 and therefore I can safely write down $R_{in}$ to be equals to $\beta$ times $r_e$ plus $R_E$ which appears as $\beta$ $r_e$ plus $\beta$ times $R_E$, right?

So if your $\beta$ is typically very high let us say, it is about 150-200 then the resistances offered by the BJT right and the circuitry gets amplified drastically, right? And the input (temp) and the input resistance becomes very-very large because you can see here, you are multiplying $\beta$ to both $r_e$, small $r_e$ and capital $r_e$ and as a result you will have a large amount of voltage across this terminal right, not only that you will also have to reduce the gain of the MOS device which you are using, other FET device of the MOSFET, sorry BJT which you are using.

(Refer Slide Time: 15:27)




Now therefore I get $A_{v0}$ to be equals to minus $g_m$ times $R_C$ divided by 1 plus $g_m$ $R_E$, right? So it is basically gain is equals to minus $g_m$ $R_C$ divided by 1 plus $g_m$ times $R_0$, fine. And this your voltage gain $A_{v0}$. We also assume for all practical purposes that $R_E$ is so large as compared to $R_C$ that $R_E$ is $R_C$ is approximately equal to $R_E$ and as a result $R_0$ equals to $R_C$, right? So the output impedance or output resistance seen by the device is approximately equal to the collector resistance offered by the BJT, right. These two things we will be assuming for all practical purposes.

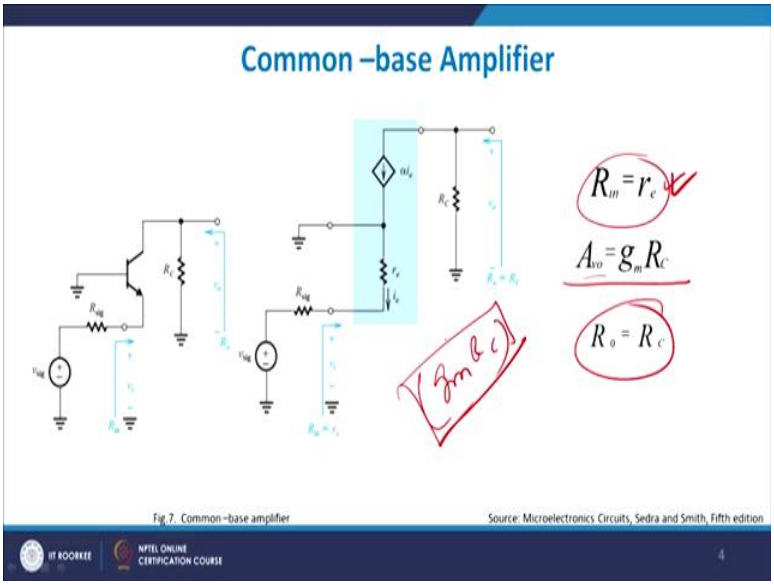
If this is true we can look at the left hand side figure, the down figure here, then we see that I have got $V_{sig}$, I have got $R_{sig}$ which is the signal specifications which terminates at this particular point and you have $\alpha$ times $I_E$ as the collector current source right, assuming that the emitter current is fully biased and only depends upon the input bias and not on anything else. I can safely

say that the collector current here is equals to $\alpha$ times $I_E$, right? The $I_E$ also flows to the below side so $I_E$ times $R_E$ is basically the voltage drop which you see on the below side of the channel, right. And that is what you get from all of these regions, right. Okay.

We now come to the common base amplifier and the common base amplifier you will see that my base is grounded right, so the previous was basically your base emitter grounded, now we are entered into base grounded. So when the base is grounded I get my, I am applying signal to the input side on the emitter side and let us see how it works out. The problem with this is technically is when you apply a signal on the emitter side right, when you apply a signal on the emitter side, you end up having a large parasitic capacitances being extracted even at relatively small voltages, right and as a result what will happen is that the speed of the operation will get reduced, right.

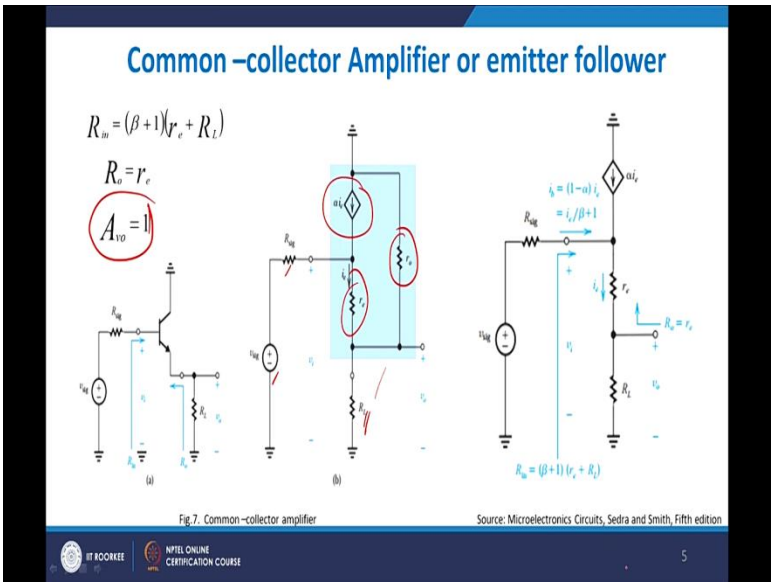
So one has to be very cautious when we design a common base amplifier in a small confined domain. Now with this knowledge we can safely right down or we can actually give an instruction that $R_{in}$, is equal to $R_E$, no doubt because of the previous discussion we can just remove it.

(Refer Slide Time: 18:00)




So if you see $R_{in}$ is equal to $R_E$, which is given here, so I get $R_{in}$ equals to $R_E$. That is what you get $R_{in}$ equals to $r_e$ why, because input resistance right, input resistance which you feed here is exactly equals to $R_E$. $R_E$ is the output impedance of a BJT at this stage. So I get $A_{v0}$ equals to $g_m$ times $R_C$ and $R_0$ equals to $R_C$, right? So you can understand that, your $R_0$ is typically very large quantity and therefore $g_m$ times $R_C$ is very large and that is the reason you do have a large amplification in their case, in their this thing that is compared to other experiments or other this thing because the same reason is that your $g_m$ is relatively independent of the applied currents, right.

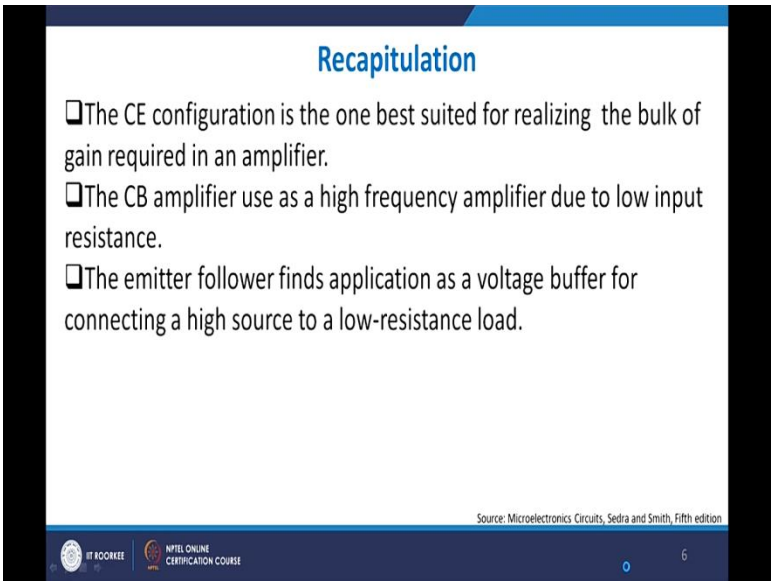
(Refer Slide Time: 19:01)




Common Collector Amplifier as the name suggests is also referred to as Common emitter or emitter follower amplifier and is given by this formula. In this case please understand the voltage gain is approximately equals to zero and that is what we all understand that common collector is not a very good amplifier or it cannot amplify small signals to large values, right and that is quite important that they cannot do it and since they cannot do it, they are used for some other purposes but not for amplification purposes, right.

And that is what we come to know right, because $A_{v0}$ equals to 1 so, it is basically phase change which is looking into and nothing else and is given by $\alpha$ times $I_E$ which is basically the actual current source which is available to us. $R_L$ is already there, $R_{signal}$ and $V_{signal}$ is also available. Output impedance of current and resistances are also available with me and gives me quite interesting results as far as common collector or a common this is given.

(Refer Slide Time: 19:55)

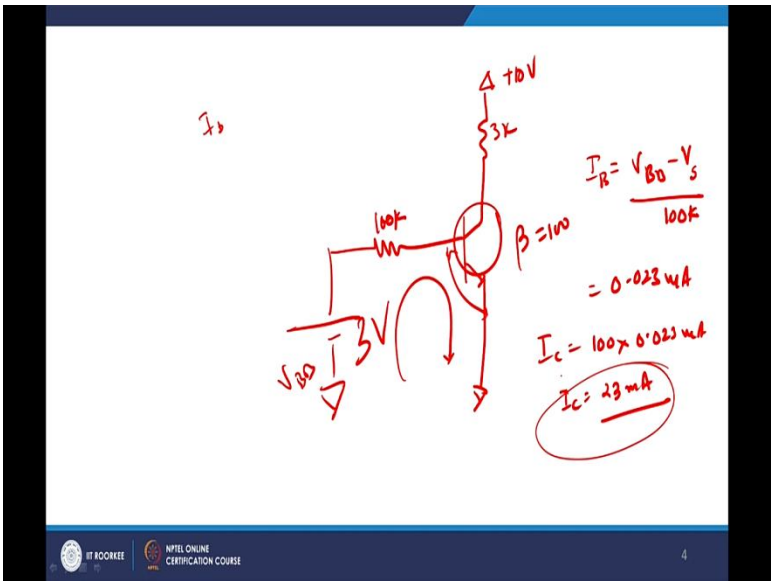



So let me recapitulate what we did. Common emitter configuration is the most easiest and the most high gain, can be obtained from common emitter configuration. The common base amplifier user high frequency amplified it with low input signal, right? Input signal is quite low and therefore you use a high amplifier design. The emitter follower finds application as a voltage buffer right, for connecting high source and passes it onto right source.

So if you do a common collector, you are actually, though you amplification will be low voltage to voltage but your transits will be very-very fast which means that if you want to compare the output impedance of a particular system with your system. On the collector side if you put the resistance is small and therefore you can easily match the low input impedance and their design. So the emitter follower finds application as voltage buffer for connecting a higher source to a load resistance source, right?

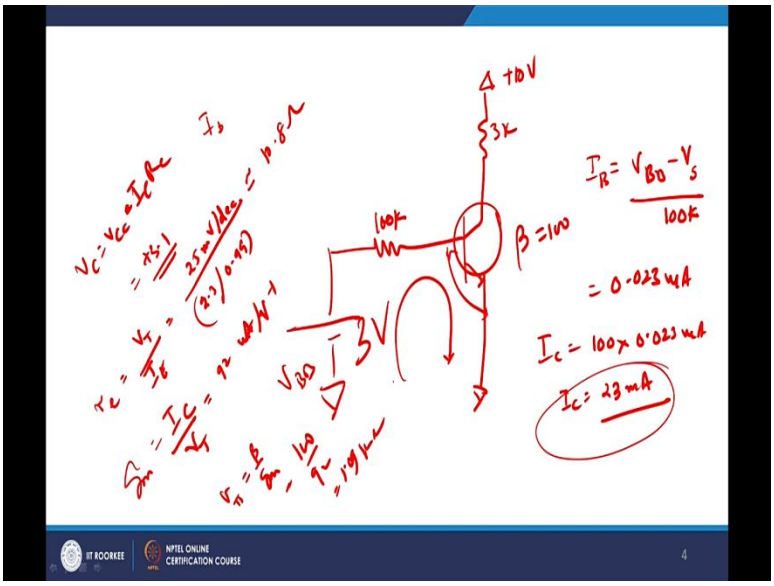
So emitter follower generally used, typically used for the purpose of buffering or the purpose of connecting between two points through a network right and that is what is emitter follower is all about. It is basically we require it the power it to be low and it is excessively low right and that that is what is very important. With this we have almost finished this section but before we move forward let me give you a brief questionnaire or a question maybe, right?

(Refer Slide Time: 21:27)




So I have a question, so I have a transistor whose beta is equals to 100 right and applied voltage is 10 volts so I get $I_C$, so I need to find out first of all beta, this is 3 volts, so we need to find out $I_B$ first. So to do that in this loop, let me find out that this is 0.7, right? And you are given 3 volts, so what you get from here is that $I_B$ will be equals to $V_{BB}$ minus $V_s$ right? This just drop divided by r which is 100 k and if you solve it comes out to 0.023 mA, right?

(Refer Slide Time: 23:12)




So the amount of current which flows is typically very small. So therefore $I_C$ will be equals to 100 times 0.023 mAs so $I_C$ will be there for approximately 23 mAs, fine. So typically this much amount of current will be required for all practical purposes.

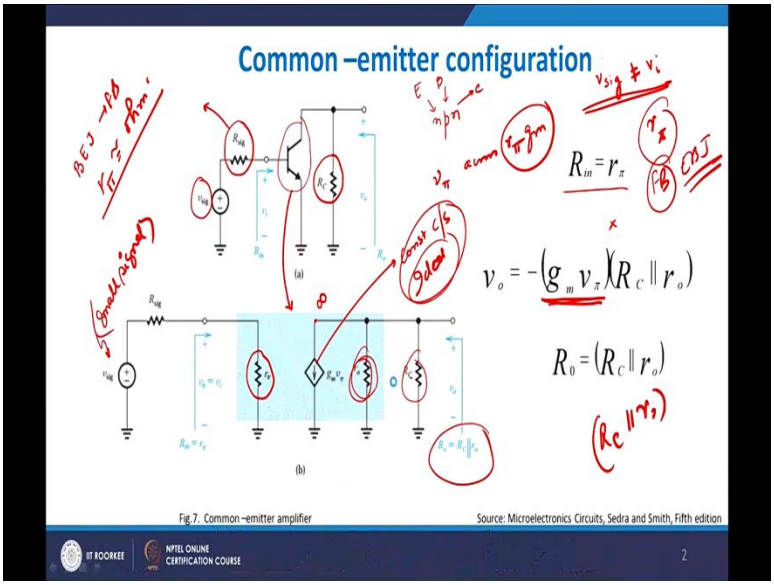
Now, if you again go back and explain the relative slides here we can (show), we can safely show that $V_C$ is equals to $V_{CC}$ minus $I_C R_C$, right? In this case we find out this to be as plus 3.1 volt. Now if you find out the value of $R_E$ it is $V_T$ by $I_E$. $V_T$ is 25 mV/decade divided by $V_T$, $V_T$ is divided by $I_E$ actually. $I_E$ will be $V_T$ divided by 2.3 divided by 0.99 and this comes out to be approximately close to 10.8 Ω, which means that 10.8 Ω is the is the emitter resistance offered by the emitter to the input system , right?

Transconductance ($g_m$) can be written as $I_C$ by $V_T$ and therefore this can be written as 92 mA per voltage minus, right? $R_\pi$ $\beta$ by $g_m$, we get 100 by 92, fine. So this is how we do a small signal analysis as far as this course is concerned and gives you an idea about how a small signal analysis works even for a single transistor or for a BJT. A transistor wherein BJT starts to find out new methodologies to destabilize itself, right?

So therefore these are the few areas in which people are working and quite interesting work is going on. Now, so, what we did was, we used a common emitter configuration for the best configuration and we and then what we saw was that for if your gain is typically very large, bulk gain is very-very large and your actual gain or the device gain is relatively small we can still have that as the best suited amplifier.

The common base is used for high frequency due to low input resistance is used for high frequency amplifier, right? The emitter follower finds applications of voltage buffer for connecting a high source to a low resistance path, right. And this we have done for emitter follower, we have done and we do for common base as well, right?

(Refer Slide Time: 25:42)



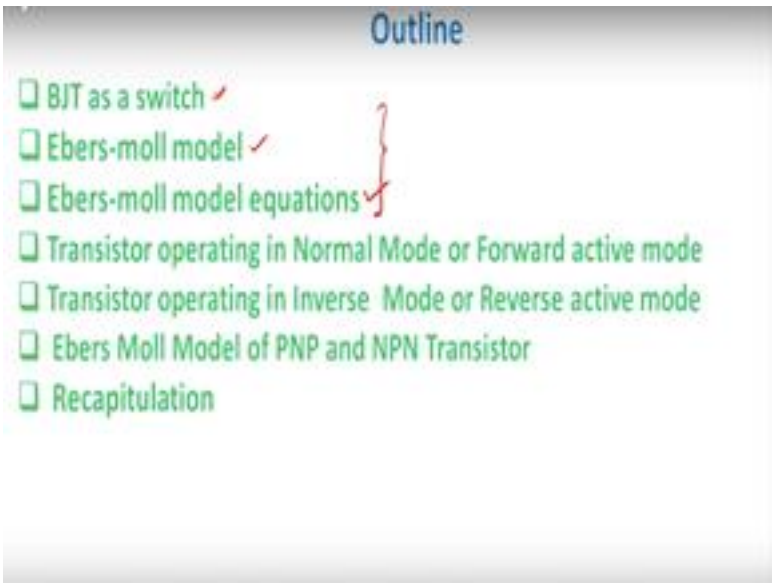

We will therefore in this stage, recapitulate what we did, we did collector common collector emitter base configuration, we also looked into the fact that given a result or given this thing which one is the best form of configuration and for that common emitter is the best form of configuration as far as the results are confirmed. Thank you very much for your patient hearing. We will come back to you next time with more information.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−10**
**BJT as a Switch and Ebers Moll Model**

Hello, everybody and welcome to the NPTEL online course on Microelectronics: Devices and Circuits. We start today's lecture with understanding of BJT as a switch and then conclude with Ebers Moll model of BJT. I will also give you brief introduction of Gummel Poon model, which is generally used in a SPICE, for SPICE purposes, which is a circuit simulator but before we move forward let me recapitulate what we did earlier. We had a look into the various small signal models available to us, out of which we had conclusively looked into common emitter, common base and common collective configuration.

We have also seen in common emitter, if you put an emitter resistance in series to the emitter port it increases the stability and decreases the gain. So that's what we have done in the previous lectures. We will start today to look into BJT as a switch and then maybe look into Ebers moll model at a later time, right.

(Refer Slide Time: 1:35)

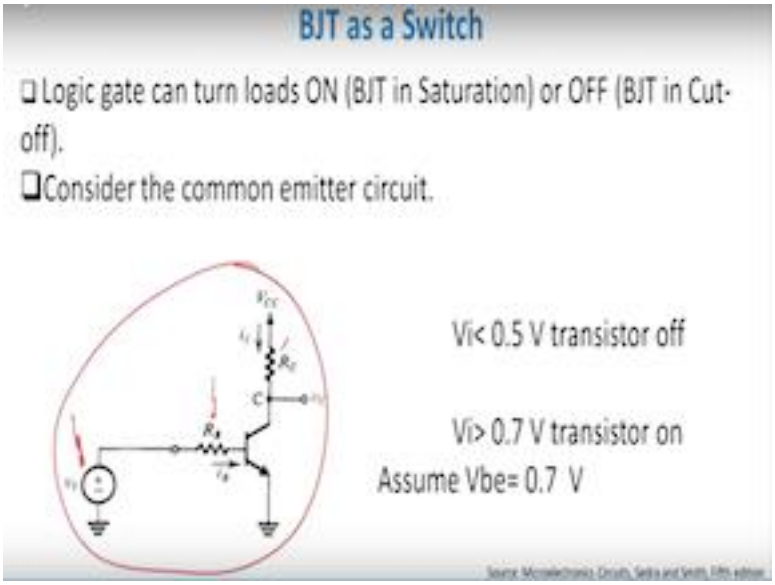


So just to, the outline of this presentation or the set of presentation will be; that we will be looking at BJT as a switch so it can be used BJT for the purpose of switch, right?, and the requirement is that whenever we have switch networks can we use BJT to do that, right, or can we use BJT to take care of those switch networks. We look at Ebers Moll model and its equations so, again as I discussed with you that small signal model and it is basically a circuit simulation model and therefore, we will be able to directly port these equations into circuits for the purpose of circuit simulation.

Then we would be looking at normal mode of operation of transistor which we have already seen in the inverse mode and then we will be putting the Ebers Moll model for NPN as well as PNP transistor, the recapitulate the whole thing once again, which means that we will be looking at into these detailed analysis once again. Now we will be looking first of all the BJT switch right, that is the first assignment, the first part in this idea. And they have been used for all your transistor-transistor logic, which is basically TTL. So let's see how a BJT can work as a switch.

Now, so we remember we all just now discussed and we saw that when the base emitter voltage goes beyond 0.7, which is basically for Silicon then we would expect to see it become forward biased and as a result it will emit large number of electrons if it is an NPN transistor and the device will be in on state, right, and it will be in off state provided the base emitter voltage is lower than 0.7 as such. So if you are able to make a transition of the base emitter voltage from 0.7 upwards to below, I can move from saturation to cut-off and vice-versa.

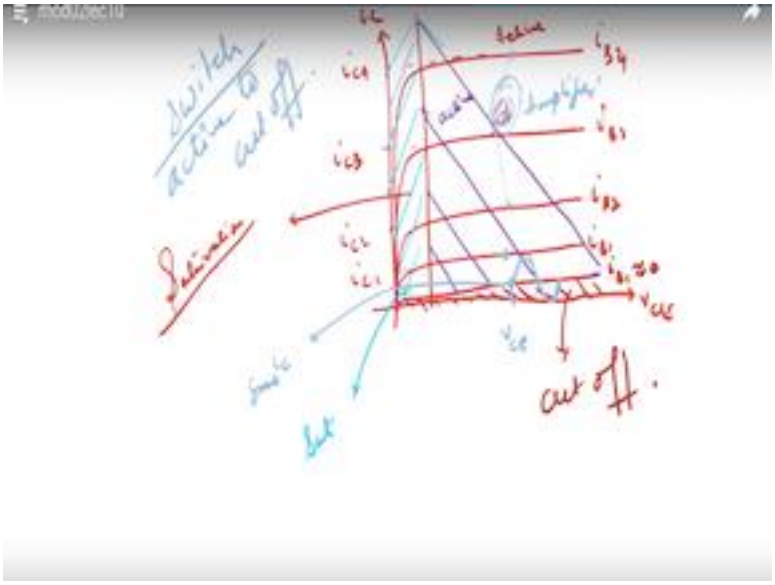
(Refer Slide Time: 4:56)




Consider the circuit which you see in front of you, this circuit, right? Where I am giving an input voltage $V_i$, right. So this means that we are giving a input voltage that is $V_i$ and this will be a sum of DC and AC bias. DC bias will be helpful for choosing the Q-point and AC bias will be actually giving you the switching characteristics. So whenever my, this $R_B$ and $R_C$, this $R_B$ and this $R_C$ are the resistance which is there in the base side and the collected side of this BJT and there so that excessive current does not flow through BJT and burnt it out.

So if let us suppose that this gets shorted $R_B$ was shorted, then if your source which is the voltage source, a DC voltage source because of some problem or other gives you a large current suddenly then the chances are that the BJT will burnout and therefore in order to ensure that it does not burn out we give a $R_B$ value here. So this $R_B$ $R_C$ and the $\beta$ of the transistor fixes the, fixes up the Q-point of the transistor so it gives me the fixed value of $V_{CE}$ and $I_C$, right? It gives me that value, where should you bias so that it acts as a switch that we need find out how do you bias it.

(Refer Slide Time: 5:01)




So if you plot the common emitter configuration for various values of $V_C$ vs $I_C$ if you plot, right. Then this is for the various value of $I_B$, I think, right. So this is $i_{B1}$, $i_{B2}$ and so on and so forth, $i_{B4}$, $i_{B3}$ and this is basically $i_C$ and this is your $V_{CE}$. So this will becomes your $i_{C4}$, right. This becomes your $i_{C3}$ this becomes your $i_{C2}$ and this becomes your $i_{C1}$. So this is your saturation region almost and this is your active region. Active region, this is your saturation region. So suppose you have a current source which is somewhere here.

So typically, for, to be used as switch, if you bias it somewhere here, let us suppose, which we generally do if you are using as an amplifier. Suppose use a Q-point. Then you, so if you look at the Q-point then, I will just give you the reason, so this is your let me show you using different color maybe and I will just show you the different color combination and may be let me use this, now this part, right, this part, which you see as I discussed with you is basically your saturation, right. Now this part, right is your, sorry this part is your active, this is active and below this part, below this part, this is all cut-off and this $I_B$ is approximately equals to 0 or very close to 0. So if you bias your device in such a manner that you end up having, your Q-point is somewhere here then you can understand that even if I give small signal input I will not be able to cross the cut-off point. So the positive half cycle is surely no, no region, negative half cycle also $i_B$ will just come from $i_{B4}$ to $i_{B2}$.

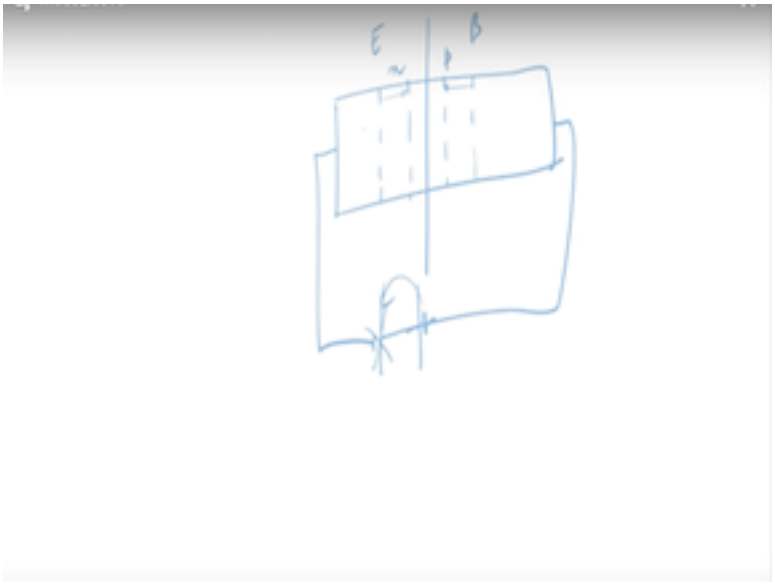
So if you bias your device at this Q-point, you will never be able to work it as a switch but then you can work it as a good amplifier. So the good idea is that if you want to use it as a amplifier bias is by using external DC sources in such a manner that it is somewhere in the middle of the active region, not very far from the saturation is cut-off. You want to use it as a switch then bias it somewhere maybe here or may be somewhere here and then in the positive half cycle agreed there nothing will happen; but the negative half cycle you will enter into cut off, right. So when your input is moving in the negative half cycle, the output is also going to the negative half cycle then you enter the, into cut-off limit, right.

So you can do cut-off but to do that you have to your $V_{CE}$ maybe, typically large but then you have to ensure your $i_C$ to be small. So this $i_C$ is small $i_C$. This small $i_C$ can only be done if you have $i_B$ small. How do you do that? How can you make your $I_B$ go very-very small, well very simple and straightforward; if you look at this point make, your $R_B$ very large? So when you make your $R_B$ very large you ensure that your, $i_B$ falls down drastically, right and then as a result you will have a much, much lower value of $i_B$ available with you.

As a result you will be able to go into cut-off. Now, the idea therefore is that when you go from therefore active to cut-off, right, without going into saturation we define that to be as a switch, right. So we define switch to be from active to cut-off, and then vice-versa, so from the cut-off to active again you want to go to it. Now you will, obvious question ask is can BJT be used for high frequency application which means that can I use this switch when the input frequency is typically very high? The answer is it is difficult; it's pretty difficult and the reason being that how does one go from on to off state?

We go via changing the emitter base junction from forward to reverse bias and then from reverse to forward. Now there is always a finite amount of time which has to be given to the diode, PN junction diode in this case such that you allow for the systems or the charge carriers to come back to its original position; this is known as reverse recovery time, right. We will not discuss any further than this case but I will give you a brief idea what I was talking about.

(Refer Slide Time: 10:08)

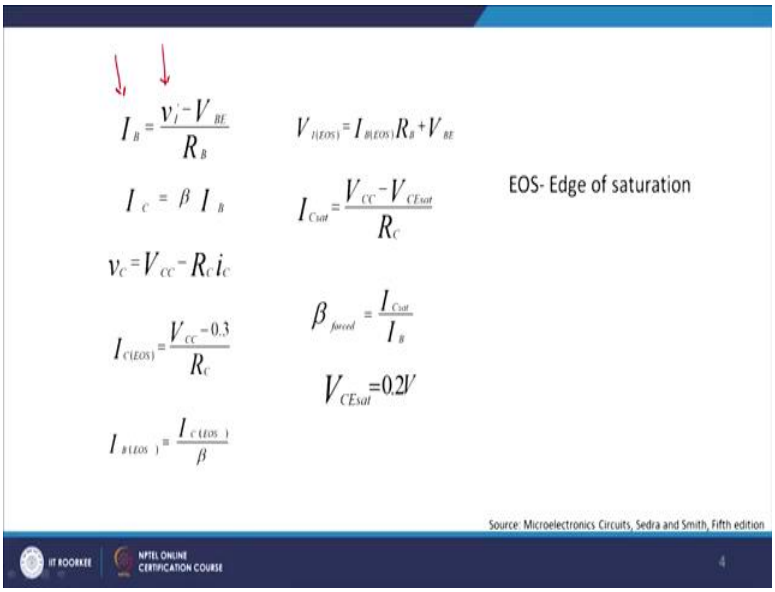



So let us suppose I have PN junction, right, so assume this to be as emitter and this to be as base so I have got NPN so this is N and this is P let us suppose. I have a depletion region here, now what I do is I first of all forward bias it, so I do like this, as I do it current flows in this no problem, then what I do is I have to reverse bias it, what I do? I just remove this part and I add something like this... Once I reverse bias it this again, the current stops because the depletion thickness becomes quite large and there is no current.

Then we again go from reverse to forward, right. Once I do it then this has to go down; this limit has to go down. How it can go down? Provided the charge carriers from both sides move across the boundary metrological boundary and thereby form the, remove the depletion region. This requires a finite amount of time and that's the reason you have to give sort of a minimum cooling period for the diode so that it basically moves from on to off state and off to on state.

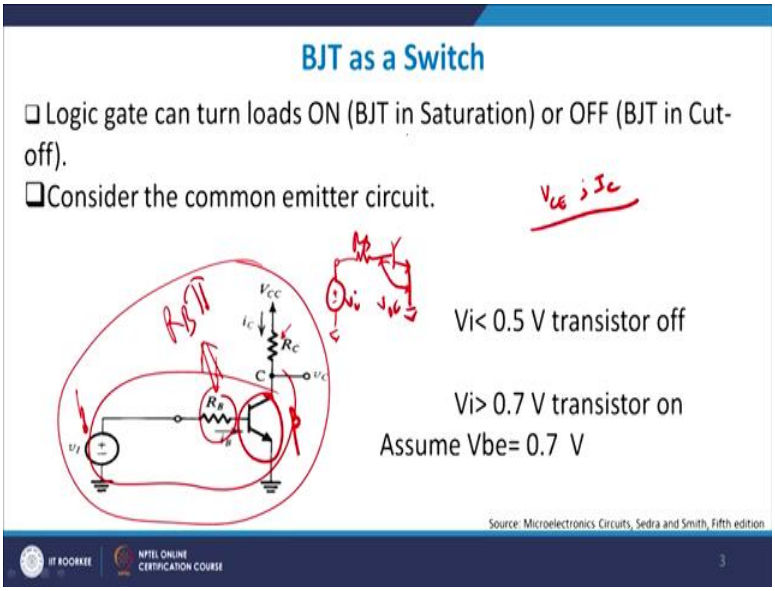
So…so… typically BJT cannot for this simple reason, there are other reasons also it is one of the reason why you cannot work it at a very, very high frequency, right. At typically gigahertz or terahertz, you can-not work with BJT; the reason is you do have problem that a minimum amount of time has to be given to the transistor for the BJT for the recovery of the time, right so that is reason is always a limit to the frequency at which the BJT can work as a switch.

(Refer Slide Time: 11:51)



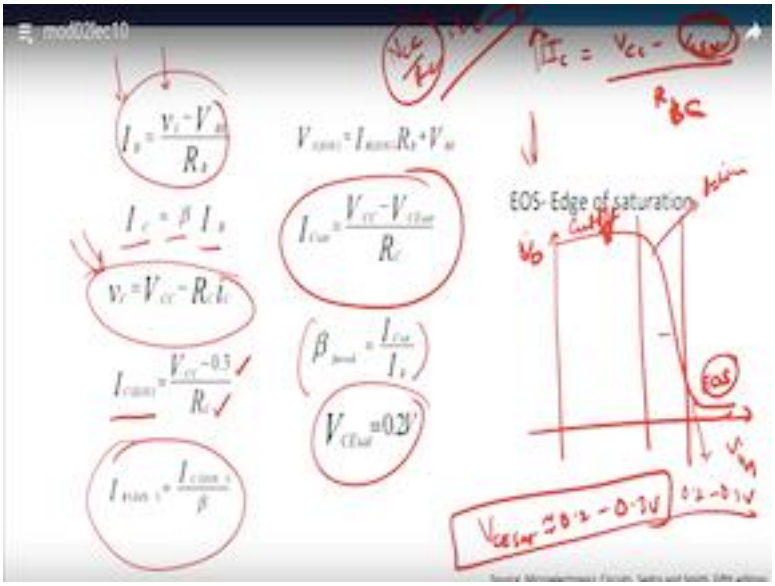

So if you look at this figure or in this equation you can see that the $I_B$, the base current can be given as $V_I$ , $V_I$ is the input voltage minus $V_{BE}$, minus $V_{BE}$ why?

(Refer Slide Time: 11:56)




Because this is the, so I am concentrating at this place, right. So it is something like this now that I have $R_B$, right and then this, right and then, you have got $V_I$. So $V_I$ minus $V_{BE}$, right, divided by $R_B$ is your $I_B$.

(Refer Slide Time: 12:17)




So this is your $I_B$, right? So $I_C$ will be equal to $\beta$ times $I_B$; a straightforward way of looking at it and therefore, if you multiply this $I_C$ with $R_C$ and then $V_C$ we will be equals to $V_{CC}$ minus $I_C R_C$ and I will get from this equation the value of $V_C$ which is the collector voltage available with me. Therefore as you can remember $I_{CSAT}$ from here

also you can see, $I_{CSAT}$ will be equals to $V_{CC}$ minus $V_{CESAT}$ by $R_C$. Now I will ask you, where is this *SAT* or from where I can get the *SAT* value?

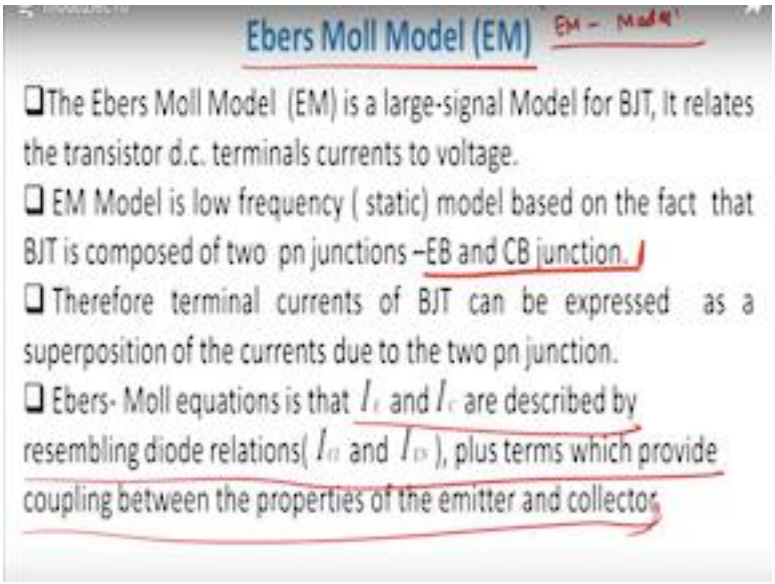
Remember, if you remember the voltage transfer characteristics which we are studying right, and we were studying like this and then it was going like this and we defined this to be my active; this was active, this was cutoff and this was, sort of edge of saturation. We started with saturation from here and this was *EOS* or the edge of saturation where you have a negative cut-off with you. So this is your input and this is your output, right. So this was active region, this is the active region and this is the edge of saturation.

So, I get this is approximately 0.2 to 0.3 voltage $V_{CESAT}$. So $V_{CESAT}$ is approximately equals to 0.2 to 0.3 volts, fine. With this knowledge I say that $I_{CSAT}$, which is the saturated collector current will be equals to $V_{CC}$, right, minus $V_{CESAT}$ by $R_C$, fine, which means that it is basically $V_{CC}$ minus $V_{CESAT}$ minus $R_C$. $V_{CESAT}$ will be typically very small, 0.3 maybe. So I get $V_{CC}$, so if you look at edge of saturation minus 0.3 divided by $R_C$, right, I get $V_{CC}$ minus 0.3 divided by $R_C$ and therefore, I get $V_{CESAT}$ approximately if I say 0.2 and I get $\beta$ forced is equal to $I_{CSAT}$ by $I_B$ and $I_B$ equals to $I_C$ by $\beta$, so that is very straightforward way of looking at it.

But I wanted to make it clear write one important point here, that therefore $I_C$ if you remember is nothing but $V_{CC}$ minus $V_{CSAT}$ remember divided by $R_B$, base current, right. $R_B$, no $R_C$ I think, sorry! $R_C$, right..? So the obvious question asked is how can you maximize $I_C$ because then only you get large currents, is that if you make this one as close to zero as possible, right?

So typically the maximum value of $I_C$ will be $V_{CC}$ by $R_C$, fine, that is the maximum value. So $I_{CMAX}$ can be written as $V_{CC}$ by $R_C$, right, and that's the reason you always get, for example in this case; when you have to have a switch you have to ensure that the current actually moves from, the $I_C$ value moves from a relatively high value to relatively low-values when we move from active region to edge of saturation. In the edge of saturation the value will be relatively very, very small, right. So this is for you to keep in mind as far as switching is concerned or switching understanding switching is concerned.

(Refer Slide Time: 15:44)



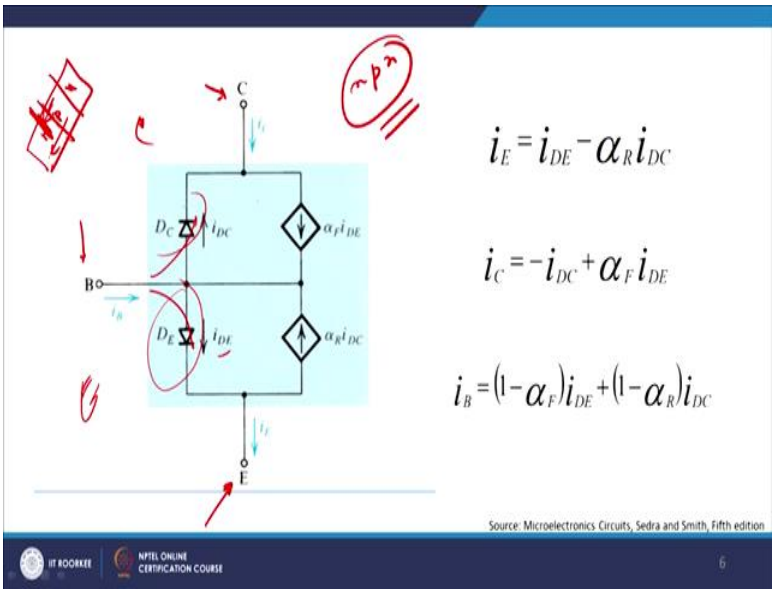

We now come to what is an important terminology that we use in the bipolar technology and that is known as the Ebers Moll model, right. This is known as Ebers Moll Model also referred to as *EMM* or *EM* model, right, by two scientist Ebers and Moll. Ebers Moll model is a large signal model for BJT unlike the previous models which is T and π, this basically is a large signal which means that we don't restrict ourselves to small changes in input but we are looking into large changes in the input and then try to predict the output voltage or current from this *EM* model.

This *EM* model is basically a low frequency. So it is not even high frequency model, it is a low frequency model and what it does is that it takes care of BJT as a two junction device. So it's an EB and CB device. So I have got two junctions emitter base and collector base junction. Right? Therefore what the idea of Ebers Moll was that therefore, everything can be broken down into set of two currents, one current flowing through emitter base another flowing through collector and base, right. These two things and also two other component, which depends on the emitter base junction and collector base junction.

Therefore told that the total terminal currents $I_C$, $I_B$ and $I_E$ will be a weighted sum or superposition of these two-junction currents *EB* junction and *CB* junction, so that's what we are saying the Ebers Moll equation is that, that $I_E$ and $I_C$ are described by resembling diode relationship plus, the coupling between the properties of emitter and

collector, right. So these are the reasons why we do an Ebers Moll model as such. Let us see how Ebers Moll model looks like, right.

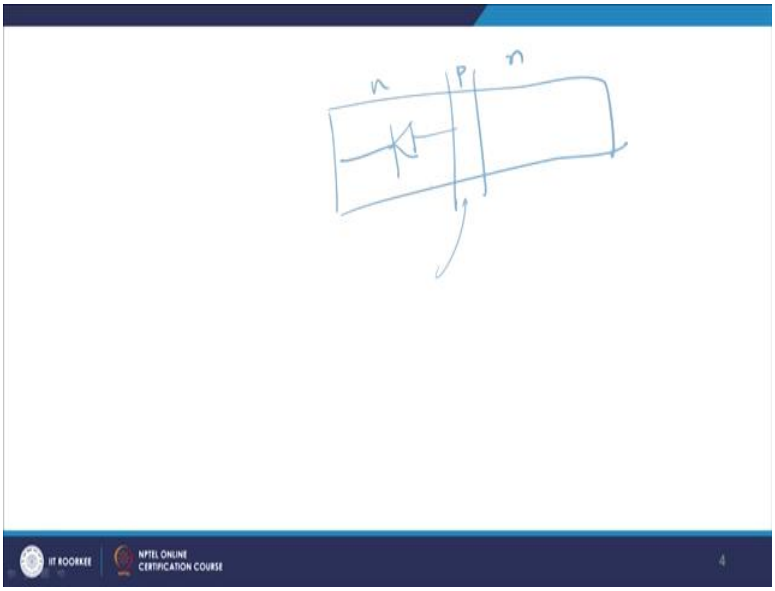
(Refer Slide Time: 17:44)




Now look at the, so this is the, it is basically a three terminal device, I have a base, I have a collector and I have an emitter here, *E*, is the emitter here and. So it is basically again as I discussed with you, it's a three terminal device. Now I have current $I_B$ flowing into the base collector $I_e$ and $I_e$, so it is basically an NPN transistor design, which we are trying to do. Now as you enter this point, this side is emitter, right and this is collector. So this is emitter and this is collector, right.

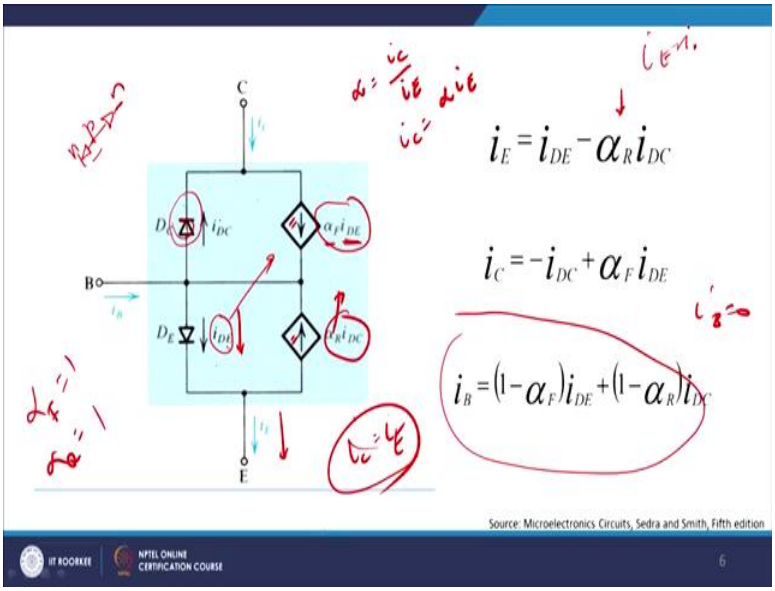
Now if you look very carefully this is basically emitter base junction, this one and this is collector base junction, right. Emitter base junction is defined by a diode, right, whose one side, p-side is connected to base, because it is NPN, so this is base, so this will be diode equation, right. As a result you will get a *DE*. So why do we get $i_{DE}$ here, I will just show you, why we get it.

(Refer Slide Time: 18:47)




So if you plot a NPN and if you look from this side from the base side, right. Then it will be, if you want to forward bias it, it will be something like this, right, and that's the reason we are showing you that by this statement here.

(Refer Slide Time: 19:04)



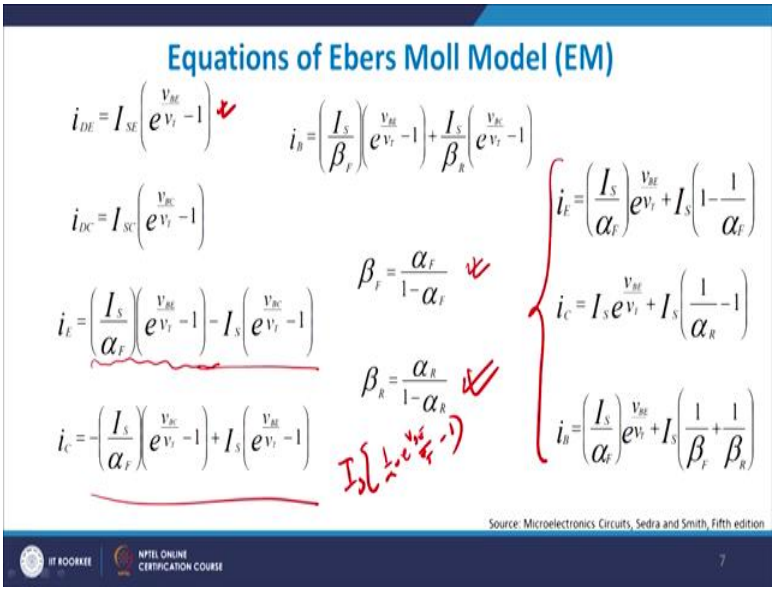

So this is the reason we see $i_{DE}$ here and base to collector is shown in this manner, because it is an NPN so the diode will be in this direction and it will be in this direction, that is what we are seeing, right. So both P are connected towards base and both N connected toward collector and emitter.

What is $\alpha_F$ into $i_{DE}$, right. $\alpha_F$ into $i_{DE}$, $\alpha_F$ into $i_{DE}$ is emitted current, so $i_{DE}$ is this one emitter current that you see, if you multiply this with $\alpha$ you get. So remember $\alpha$ was equal to $i_{CE}$ by $i_E$, fine. So I can define $i_C$ to be equals to $\alpha_{iE}$, that's what you have done. So $\alpha_F$ forward current multiply $i_{DE}$, right, so this you multiply with this one you get this one and similarly $\alpha$ into $i_{DC}$ will give you the value of collector current, right.

So therefore see, therefore you see the direction of the collector current is also and the emitter current is also shown in this manner and this manner, right and you can predict the reason why is it like that for all practical purpose. So I get from here, $i_E$ which is looking from this side $i_E$ equals to $i_{DE}$ which is, so $i_E$ is this side, $i_{DE}$ is this side and since $i_{DC}$ is the opposite direction I get minus $\alpha_R$ times $i_{DC}$, you see $\alpha_R$ is basically your this thing, reverse current again, right. And $i_C$ will be equals to minus $i_{DC}$ plus $\alpha_F$ into $i_{DE}$, minus $i_{DC}$ because the direction of current is reverse.

Therefore I can finally find out since $i_E$ is equals to $i_B$ plus $i_C$, $i_B$ will be 1 minus $\alpha_F$ $i_{DE}$ plus 1 minus $\alpha_R$ into $i_{DC}$ fine, so this is the value of $i_B$ which we get as such. So therefore if you say, $\alpha_F$ and $\alpha_R$ both are equals to 1, then $i_B$ equals to 0. And therefore $i_C$ equals to $i_E$, fine. But this is not true because $\alpha_F$ which is a forward gain and $\alpha_R$ which is the reverse gain will never be equal to each other, they will be a large difference between the two and therefore, this inequality that $i_B$ is equal to 0 will not hold good, right. And that is almost a sure shot sort of a network. So using that network which we have just now saw, we can write down the basic equations of Ebers Molls Model.

(Refer Slide Time: 21:45)

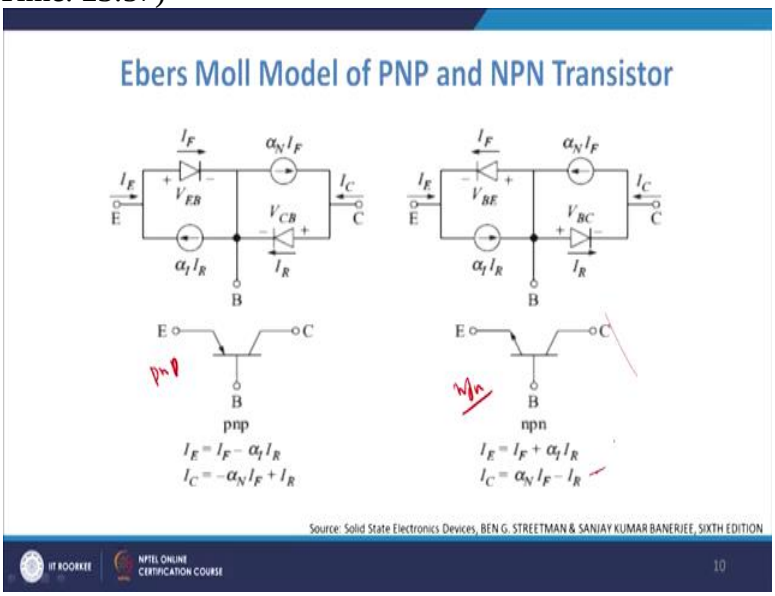



The first equation is of course as we have discussed is this one, which is $i_{DE}$ equals to $I_{SE}$. $V_{BE}$ $e^{\frac{V_{BE}}{V_T}-1}$. $i_{DC}$ is equals to $I_{SE}$ $e^{\frac{V_{BC}}{V_T}-1}$. Then $i_E$, if you look very closely is this is basically your $i_{DE}$, right. And therefore I get this minus this, right. And then I get this minus this we get, then $\beta_F$, forward and resistance and negative resistance $\beta_R$ and then we get from these equations the values of $i_E$, $i_B$ and $i_C$, right.

And we see that, if we look very carefully here that base to collector junction is always in a forward active region. $V_{BC}$ will be quite large and negative, right. As a result this will all vanish. So I will be only left with minus plus $I_S$, fine. So $I_S$ by $\alpha$ plus $I_S$ so I will get $I_S$ common, so I will get $I_S$ common, Right I get 1 by α multiplied by $e^{\frac{V_{BE}}{V_T}-1}$. Right, we can get something like this for our understanding purposes. Same thing happens if we take negative value for your Ebers Moll Model.

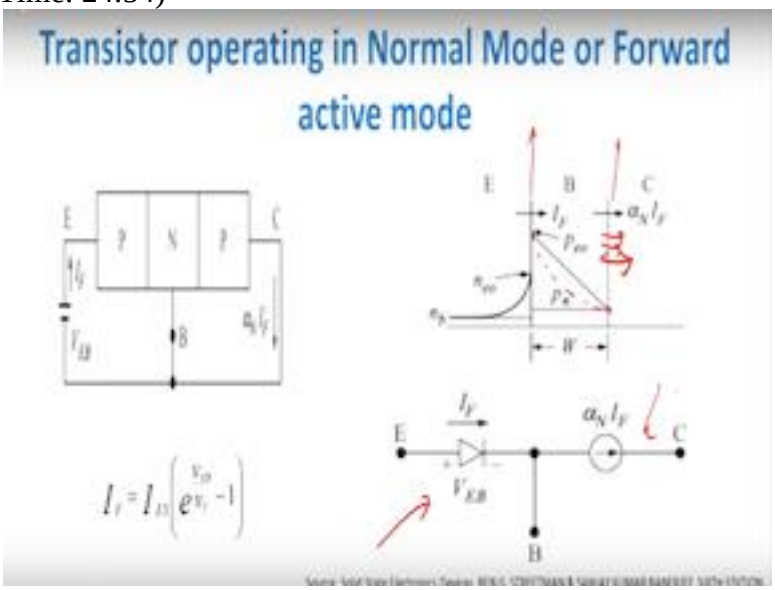
Ok, so we have therefore understood the basic fundamental principles of the Ebers Moll model. We have also understood how Ebers Moll model work. We have also understood how does a BJT, Bipolar Junction Transistor can be removed and we can replace it by its corresponding model file and provided we are able to sustain these equations in the model, right. And we are able to achieve a sustained value of these model files. Before we move forward to inversion this thing, let me give you a Ebers Moll model for NPN and PNP transistor. Ebers Moll model for NPN and PNP, this is basically a PNP, right.

(Refer Slide Time: 23:57)




This is basically an NPN, right. NPN is relatively easy to handle, so I can do it, $I_E$ is equals to $I_F$ plus $\alpha_I I_R$, right. Similarly $I_C$ can be written as $\alpha_N I_F$ minus, because it is a reverse bias, minus $I_R$, right, because in negative side there is no gain which is available at this transistor at NPN but for PNP you do have a gain available and therefore this comes out to be like this, right, and I get $\alpha_{IF}$ plus $\alpha_{IR}$ as the value of the base collector current, right.

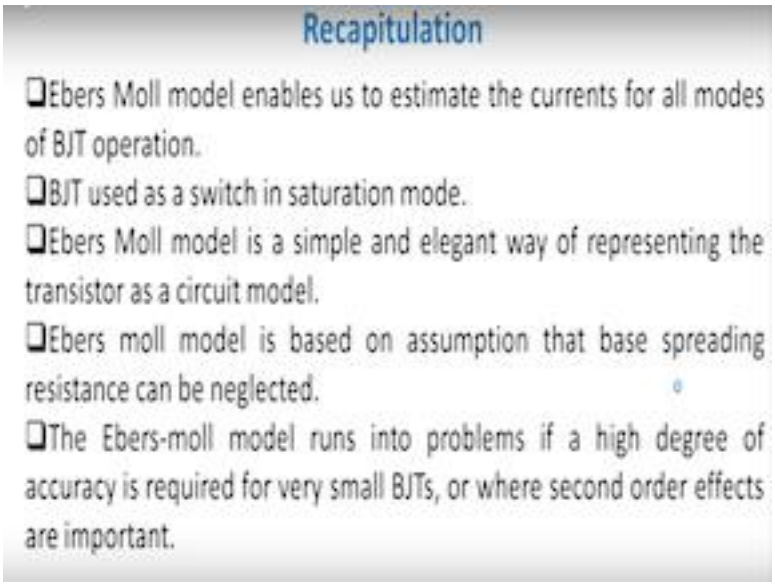
(Refer Slide Time: 24:54)



Now we have already discussed transistor operating in normal mode or forward active mode. Just to give you a brief insight what will happen in reality we will see that, as I discussed with you that minority current carrier will actually go down linearly from emitter base to collector base and it will go to such a low value at almost near this thing will go to zero because, the collector will be removing all the electrons from the junction. As a result will go to zero.

So that is quite difficult but since you do have recombination in the base side therefore this will ideally will not be a straight line but will be slightly bend in this manner and as a result because of the recombination, the result you will have a bending parameter there. So this, if you see at this model, this one, it is emitter base collector. So in the emitter base collector this $V_{EB}$ is nothing but the base to emitter forward resistance. $I_F$ is the forward current and $\alpha_N$ into $I_F$ is basically the collector current, which is to be developed or which has been developed, right.

(Refer Slide Time: 25:56)



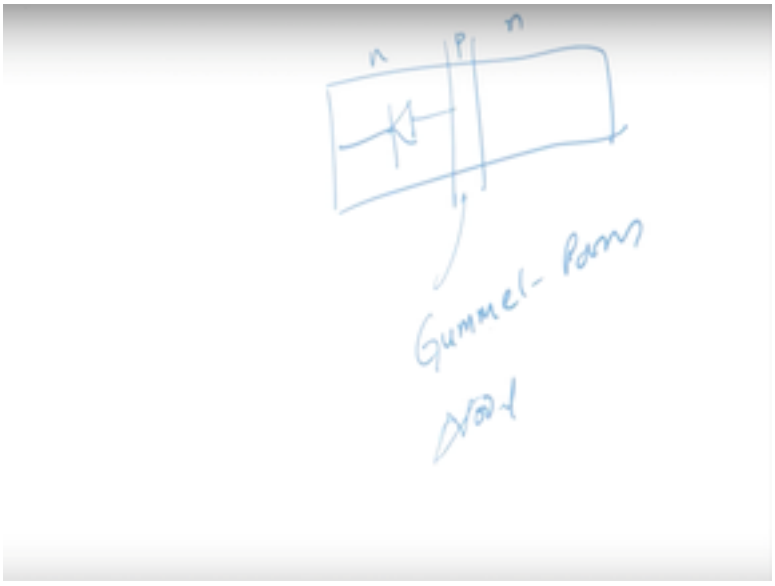

With this idea let me recapitulate what we have done in this module. We looked into Ebers Moll model and tried to estimate the currents for all modes of operations of BJT. Please ensure that the current will be typically very high in active region, almost zero in the lap region. It will be moderately good near the other regions, right. So in the saturation region it will be typically very, relatively high but fall very fast, active region it will be typically very high but remain in that high position for quite a long

time until and unless an external force is pulling the node down and cut-off is region where the effective current is approximately equals to zero.

Please take my mind reads effective resistance and approximate resistance goes to zero. Real resistance might not even go to zero at any point of time. So as I discussed with you Ebers Moll model is an elegant way of expressing a circuit, right, and then what we have neglected in Ebers Moll is, we have neglected the effective base with model, right. So please keep this in mind when we do the next time, we will be using the effective base model which means that the, we remember when we have reverse passing the collector base junction.

The effective base width was reducing, right. So it was going down, and as a result there was certain issues which were formed down, which were not addressed by the Ebers Moll model, right, and therefore all your second order effects and everything which is emanating from Ebers Moll model need to be questioned and we need to look into those facts in a detailed fashion.

(Refer Slide Time: 27:36)




So Ebers Moll model actually fails when you want to have a second-order effects coming into picture, right. So Ebers Moll model does not work very good in that case when you have second order, in that case what works is basically my Gummel Poon model. You should look at Gummel Poons model and this is much more robust model

at high frequency, at a very high frequency domain, right. And this is what we have learned from our previous understanding of statements, right. Ok, with this let me also therefore, we have therefore taken care of two important points in this lecture.

The first point is that we have looked into transistors, the transistors BJT as a switch, so I can switch active to cut-off and cut-off to active, vice-versa by fast movement of an external peripheral source. We have also seen the Ebers Moll model and how can it be integrated with circuit analysis. We also looked into basic Gummel Poon. Gummel Poon will come to later on if time permits but we have looked into those facts and we were able to sustain a much better design for a BJT base design, okay. Thank you very much, thank you.

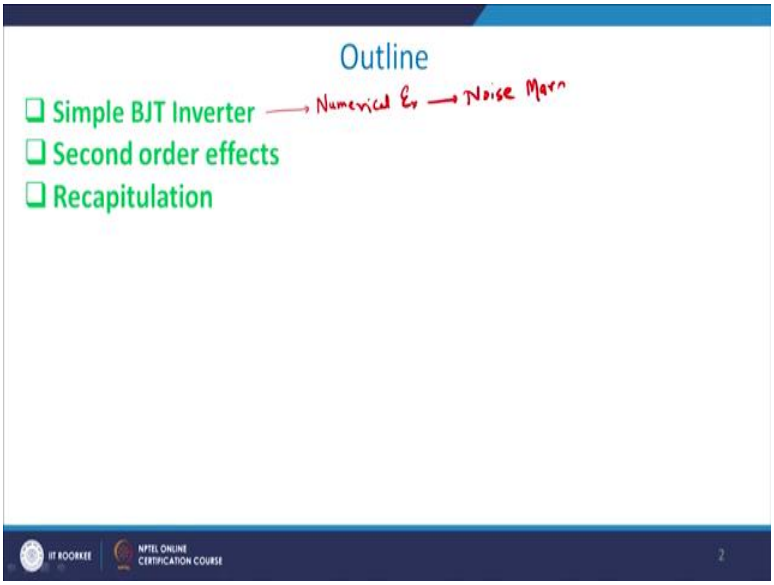
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−11**
**Simple BJT Inverter and Second Order Effects**

Hello everybody, and welcome to the NPTEL online certification course on Microelectronics: Devices to Circuits. In our previous interactions, we have actually seen the functioning of BJT, various modes of operation of a bipolar transistor and then we have seen BJT as an amplifier, right? So given a small input signal how can I actually achieve amplification in a voltage domain as well as in current domain of course. We have seen that in our previous discussion.

We have also seen the three modes of operation of BJT which is common emitter, common base and common collector. And we have appreciated what is the utility of these three modes of operations of bipolar transistor. What we will be doing as the last part of the bipolar before we move to CMOS technology is look to BJT as an inverter, because that is the basic logic which we tried to design initially. So we will look at BJT inverter and then we will look into the various second order effects of an inverter right? Second order effects of bipolar transistors, right?

One we have already dealt earlier but in details we will be dealing it now, what is the early effect? But before we move ahead with that, let me first discuss with you a simple BJT inverter, right? So the topic of this slide, the topic of this section is basically your BJT, simple BJT inverter and second order effects, right? So this is the topic of this lecture series talk. So what will we be covering here? We will be covering, covering basic simple BJT inverter. So given a BJT inverter, can we find out the inverter characteristics of a BJT and use it for logic inversion purposes? That is what we will be seeing here.

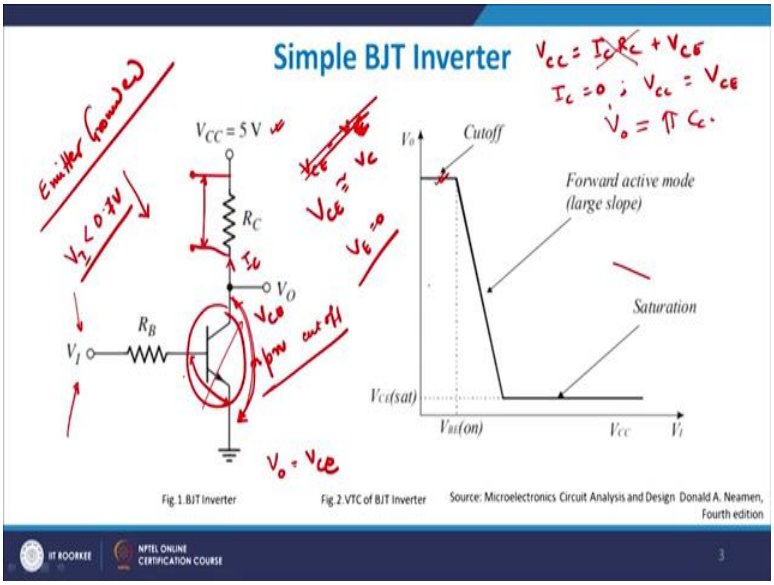
(Refer Slide Time: 2:23)




Maybe we will solve a numerical example in this case. I will show you a numerical example right? Numerical example. And we will also concentrate on what are known as noise margins for BJT, right? So, we will also discuss this in detail when we go to CMOS logic. But for BJT logic we will also understand what is noise margin and how do you define noise margin in a BJT. And this will be through a numerical example setup here. Then subsequently we will be going for other second order effects of BJT where we will be looking at base widening, then we will be looking at early effect and so on and so forth. And then how does it influence therefore the *I-V* characteristics of a BJT, right?

That will take care of approximately the major understanding portion of BJT and therefore knowing this, you can actually approach BJT from analog point of view when you use it as an amplifier and you can also approach it from a digital point of view when you use it as an inverter, right? So this is what we will be... So that is the reason or the logic for having this bipolar transistor as the first, first part of this new lecture series on NPTEL.

(Refer Slide Time: 3:35)




With the knowledge you have gained now, let me just show you a simple BJT inverter. As you can see, the simple BJT inverter is this which you can see here. And this is a BJT which is an NPN transistor, so I have an NPN transistor here whose emitter is grounded. So this is basically an emitter, emitter grounded inverter, inverter. And we apply a voltage input to the base through a base resistance $R_B$ and we try to extract the output voltage from a collector of the BJT. And you apply your voltage $V_{CC}$ which is equal to 5 Volts here, right?

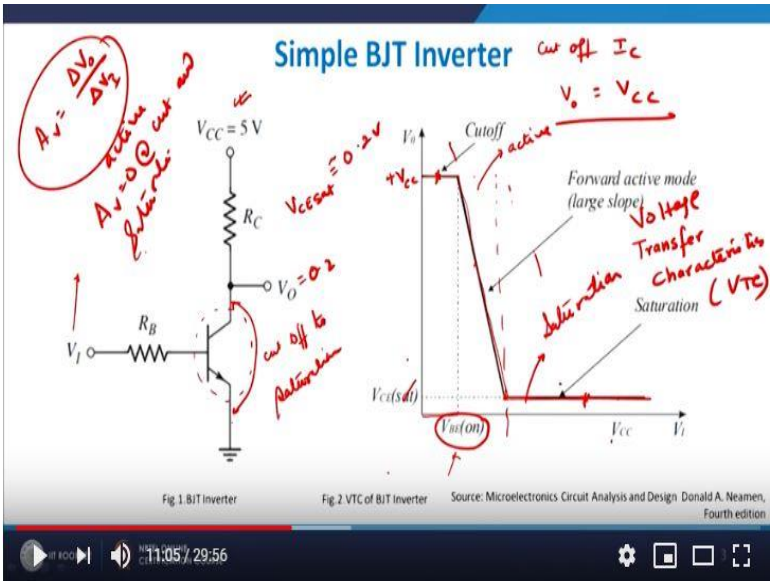
So this is the basic principle. We have already seen it earlier, now we will look into the details from the voltage transfer characteristics point of view. So as you can see therefore, when your input is typically very low, right? It is much smaller than 0.1 then we can, sorry, when $V_I$ is much smaller than 0.7 Volts which is the turn-on voltage for this base emitter junction, right? You would expect to see that, this, this is cut off. NPN will be in cut-off and as a result, what will happen is that this $V_O$ will be exactly equal to 5 Volts, right?

Why? Because if you... we had discussed this earlier also that we can write down $V_{CC}$ to be equal to $I_C R_C$ plus $V_{CE}$, right? This is $V_{CE}$. This one is $V_{CE}$, right? And, this is, $I_C$ is flowing, let us suppose, here. So, $I_C R_C$ is the voltage drop across this resistor. This, this is $I_C R_C$. This voltage drop plus this voltage drop must be equal to $V_{CC}$. But you see, since your emitter is not through a resistor, it is directly coupled to the ground, $V_{CE}$ can also be written as simply $V_C$.

So $V_{CE}$ is nothing but the collector, your sorry $V_E$, I am sorry $V_E$. I am sorry, I am sorry. $V_{CE}$ can be exactly written as $V_C$ because $V_E$ is actually equals to 0, right, because emitter is grounded. But keeping this in mind therefore, when your device is in the cut-off state, when this is cut-off, NPN is cut off $I_C$ is equal to 0. $I_C$ equals to 0 implies that your $V_{CC}$ is equals to $V_{CE}$, right? Because $I_C$ equal to 0 means this is 0 and therefore $V_{CC}$ equals to $I_{CE,}$ right?

And therefore, you can safely assume that what you can see from all this discussion is that if you try to find out the value of $V_{out}$, $V_{out}$ will be equal to, will be nothing but this $V_{CE}$, right? $V_{out}$ is equal to $V_{CC}$, right? Now, when my $I_C$ equals to 0, right? $V_{CC}$ equals to $I_CR_C$ plus $V_{CE}$. When my $I_C$ equals to 0, this goes off, right? And therefore, the voltage across this thing is pretty large, right? It is pretty large because all the voltage drop will be across. So this is cut-off, sorry so this is cut-off and therefore, I will get $V_0$ to be equal to a high value, equal to $V_{CC}$ and you get a large value here. So when the device is in the cut-off mode, your voltages are all going to 0.

(Refer Slide Time: 6:58)




And as a result what will happen is, that therefore when your device is in cut-off mode, right? $I_C$ equals to 0 implying that your output voltage is approximately equal to $V_{CC}$. So if you are using a plus 5 Volt $V_{CC}$, I will get this to be as equal to plus 5 Volts, right? Plus $V_{CC}$. Now as you go on increasing the value of $V_I$ from a low value to a high value, you switch on the device drastically. And when you switch it on, what happens is that this becomes from cut-off to, cut-off to, it goes to saturation.

Why saturation? Because the voltages are not large enough in order to ensure that these have moved fully into. When, this is when it moves to cut-off, this $V_{CE}$ can be written as $V_{CE}$ can be written as $V_{CESat}$, saturated $V_{CE}$ which is approximately equal to 0.2 Volts which means that this is the value of voltage which you get as 0.2 Volts. At $V_{out}$ equals to 0.2. So we have understood one basic fundamental principle that whenever my input voltage is low, my output voltage is high and vice-versa. So I can use it as an inverter.

As we have also discussed an earlier part that below $V_{BE}$. So you see if you look at this curve and this is $V_I$ versus $V_O$. So this is basically your voltage transfer characteristics, right? This is known as *VTC* or characteristics. And this is known as voltage transfer characteristics, voltage transfer also referred to as *VTC*, right? So if you look at the *VTC*, right, it is just behaving like an inverter. Because when your input is $V_{BE}$ (on) till that much point, till that much point your device is cut-off and the voltage is latched to $V_{CC}$.

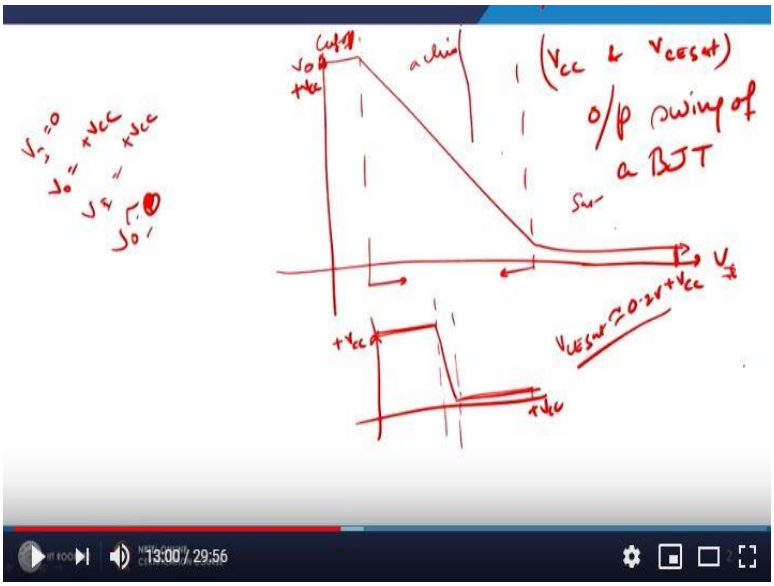
Beyond $V_{BE}$ (on), the devices switches and the voltage starts to fall down. It falls down to how much value? Equals to $V_{CEsat}$ and in this region $V_{CEsat}$. After this if you even increase the value of $V_I$, it does not affect because the device is fully on. So you see, this region, this region is defined as my active region. And this region is defined as my saturation region, right? So I have got cut-off region, active region and saturation region. The active region is a region where the voltage $V_0$ is a function of $V_I$ and therefore there is a finite $A_V$ or the voltage gain because $A_V$ will be defined as $\partial V_0 \, \partial V_I$.

Therefore, I will have a finite gain in the active region, right? But I will have $A_V$ equals to 0 at cut-off and saturation. At these two places, my voltage gain will be 0. So if you want to bias my device as an amplifier, you bias it in the active region of the device. If you want to bias it as a switch then make your active region as small as possible and try to shift from cut-off from here to here, right? And you can get a digital 1 and digital 0 in the output and input side, right?

So, basic BJT therefore, we had a problem that the BJT is though it is pretty fast in terms of switching, but the problem is that it has got a high power dissipation because of high leakages. Not only that, BJT also has a problem that to move it from saturation to cut-off, right, you have to actually remove large number of charge carriers from the saturation region, right? To make it to go into cut-off.

So, there is some dead-time also associated with the BJT switching action. Which means that the BJT has to wait minimum amount of that much amount of time before it crosses from saturation to cut-off, right? And therefore the speeds are relatively low for BJT when it acts as a switch, right? So the maximum frequency operations are relatively low in this case as compared to the previous cases, as compared to CMOS technologies per se.

(Refer Slide Time: 11:08)



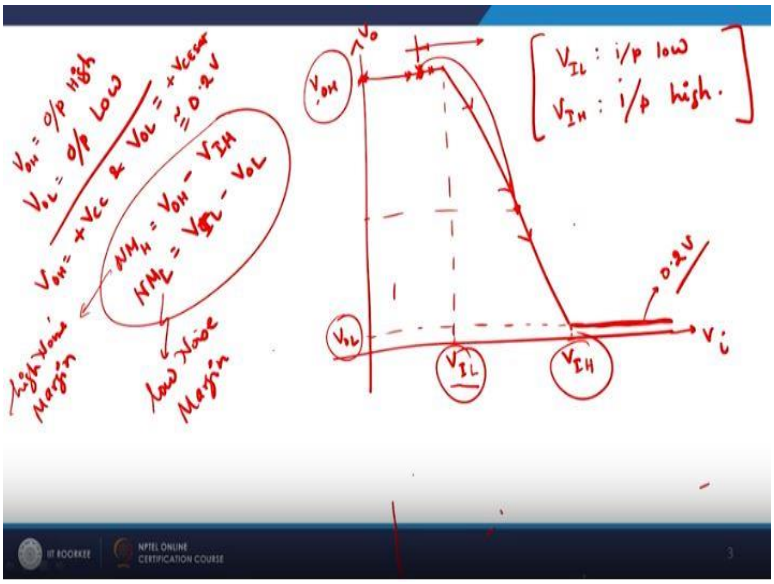

As you can see here, therefore, if I want to therefore make my transistor work in the saturation region, the active region then try to make the active region as large as possible and then this is my active region, right? And this is my active region and this is my saturation and this is my active, right? And this is cut-off, fine? And this is $V_O$ versus $V_I$. Now if you want to reduce the active region, you need to bring these two points close to each other. Which means that if I want to reduce it, it is something like this, right?

So your active region has drastically reduced here. So the gain which you get is that your cut-off region is now very large and your saturation region is also very large. Because if you remember, this will be actually equal to $+V_{CC}$. This point is $+V_{CC}$. Output will be equal to $+V_{CC}$ approximately if your input is equal to 0. So when your input, so when $V_I$ equals to 0, $V_{out}$ will be equal to $+V_{CC}$ and that is what is written here. And when $V_I$ equals to $+V_{CC}$, $V_O$ will be equal to 0, approximately equal to 0.

So at this point when it is $+V_{CC}$, when $V_I$ equals to $+V_{CC}$. Let us suppose here $+V_{CC}$, then output actually falls to 0 value or very close to 0 value; typically equals to $V_{CEsat}$ actually which is approximately equal to 0.2 Volts. Fine? So we have learnt one important point that though BJT is a good thing but a good switching action but the point is that it cannot go below $V_{CEsat}$ in the lower dimension and the highest one, it can of course go to $V_{CC}$.

So the swing is restricted between $V_{CC}$ and $V_{CEsat}$. This is the output swing, output swing of a bipolar junction transistor, right? So this is what you have learnt, what you have learnt here. We will define some important terms here and that will be making it much easier for you to understand or appreciate that.

(Refer Slide Time: 13:08)




Let us suppose I have got a curve like this and I give like this and I give it like this and then you go like this, right? So we draw this curve, remember $V_{BE}$ (on) and then we get this and then we get this and we of course get this into consideration. We define this to be as $V_{OH}$. So this point is, $V_{OH}$ is $V_{OH}$ implies output high, right? And this is $V_{OL}$. So $V_{OL}$ is output low. So $V_{OH}$ is output high and $V_{OL}$ is output low. In our case, $V_O$ in our case, $V_{OH}$ is equal to $+V_{CC}$ and $V_{OL}$ is equal to plus $V_{CEsat}$, approximately therefore equals to 0.2 Volts. Fine? Is it okay?

We also define two terms in the input side because this is $V_I$ and this is $V_O$ on the input side. And this is when $V_{IH}$ and $V_{IL}$. This is $V_{IL}$, this is $V_{IL}$. This is $V_{IL}$, right? $V_{IL}$ and this is $V_{IH}$. So $V_{IL}$ and $V_{IH}$, why? Because $V_{IL}$ means $V_{IL}$ is meaning that input is low. And $V_{IH}$ is meaning that input is high. So I get four conditions or four biases. $V_{OH}$, $V_{OL}$, $V_I$ so I get $V_{OH}$ right?

$V_{OH}$, $V_{OL}$, $V_{IL}$ and $V_{IH}$. $V_{IL}$ and $V_{IH}$ are input high input low. Input low primarily means that that is the maximum value, right? Of your input in the low condition when the output will be actually registered as one. And $V_{IH}$ is the, maximum, the minimum value in the highest condition when the output will be read as 0, right? So you have to be very careful about what is the meaning of $V_{IL}$ and $V_{IH}$.

So $V_{IL}$ is input low. Input low means this is the maximum value of input at which you will actually get a high. If you exceed, if you exceed this value of $V_{IL}$, your output will start to drop down. And that is what is happening here. See? Right? Similarly, $V_{IH}$ is the minimum value at which your output will be low. After this, if you look very carefully it might fall down or it might remain constant, right? So this is what you have learnt. So typically it is 0.2 for a silicon diode $V_{CEsat}$ is equal 0.2. We define a new term here and that is how you define your... we define this to be as a noise margin and that is basically my noise margin static noise margin we define, high noise margin.

It is defined as $V_{OH}$ - $V_{IH}$. And $N_{ML}$ is $V_{O,}$ $V_{IL}$ - $V_{OL}$. Right? $V_{IL}$ - $V_{OL}$. So the difference between the voltage of $V_{OH}$, $V_{OH}$ here and $V_{IH}$ here is defined as my high noise margin. So this is defined as high noise margin (noise margin) and this is referred to as a low noise margin. What does a high and low noise margin mean?

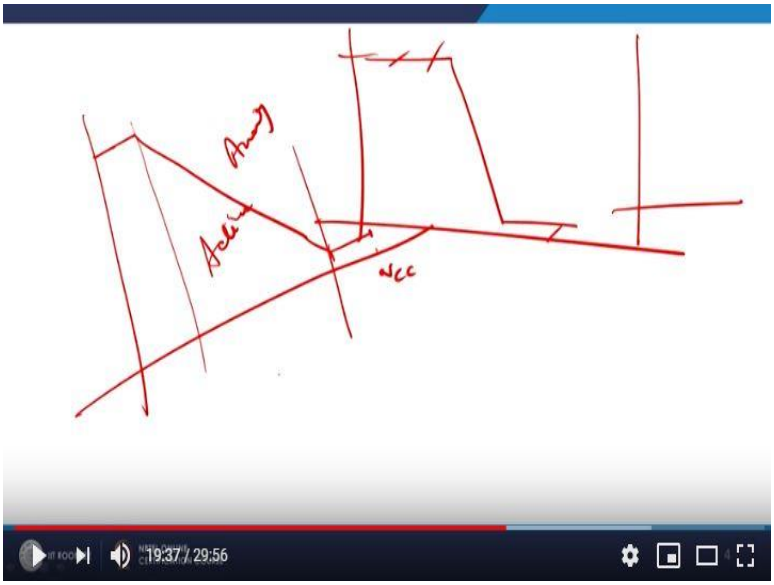
This means that, we will discuss also when we will go to CMOS but make it clear as well, so we can come over it. High noise margin means that, let us suppose we have 1 being entered into the BJT. So my input is 1 right? Now, let us suppose that there is noise in the system. If the noise becomes very large, a 1 can be actually seen as a 0 by the input side, right? Or it can change from 1 to 0. So which means that let us suppose you had biased your device somewhere here, right?

And your input cycle was slightly like this, no problem. So what is happening? The maximum value of the Q-point goes here to here, no problem. Your output is still high. But if you make a very large input, because of noise of some other means, this Q-point will shift from here to somewhere here. And therefore the output starts to fall down. And therefore you are not able to store an information of 1 in the output size because the input is actually going from high to low. Right? Sorry low to high.

So you had an input which was low, right? You inserted, you add noise to it, right? And the (minimum) maximum noise which can be added to the signal without changing the output is defined as my noise margin. So we define the high noise margin for 1 and we define the low noise margin for 0. Fine? So higher this value, better your design is because better it will be rejecting the noise.

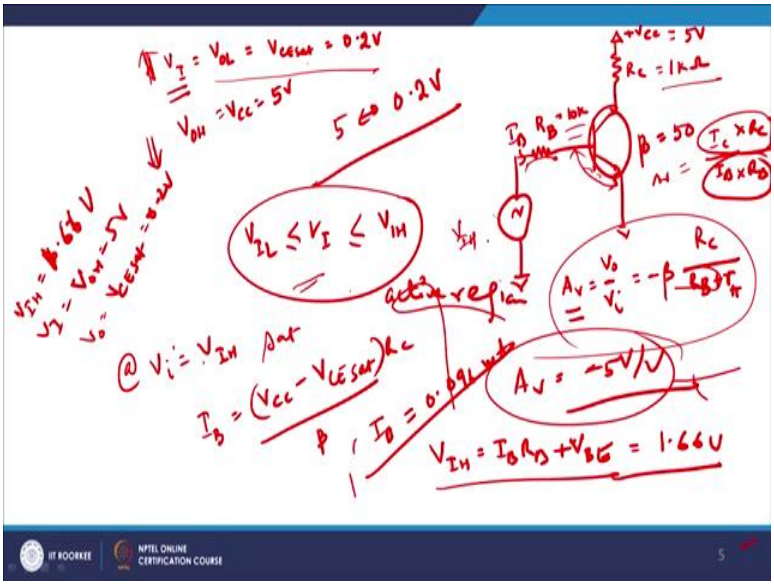
So if your noise margin is typically, say, 1.5 Volts it means that till 1.5 Volts you give me any data, it will always be read as output 1. Right? And if the $N_{ML}$ is also very high, it means that you give me a value of input data equal to the $N_{ML}$ value and it will be read as output 0. So we have two noise margins therefore as I discussed with you. We have got a high noise margin and we have got a low noise margin, right? And we try to keep both the noise margins as high as possible.

(Refer Slide Time: 19:04)



So if you want to use it for the purpose of digital logic then you keep your noise margins relatively high, right? You may keep this relatively high. If you want to use it as an amplifier, then keep this low but increase the value of your... So this is your $V_{CC}$. Increase the value of this transition active region. Then again use it in an analog domain. But if you want to use it in a digital domain, keep this one as low as possible and try to achieve equal values of input and output noise margins.

(Refer Slide Time: 19:36)




I will take up an example and show it to you how I can follow a noise margin. Say I have got a, I will give you a basic fundamental principle. I have an $R_C$ here, right? I have got resistance here. And I have got $R_B$ here. I am applying an input signal. It has already been biased by a DC bias. This is $R_B$, $R_C$ and this is (v) $+V_{CC}$. And this is the transistor $\beta$ whose $\beta$ is exactly equal to, let us suppose, the $\beta$ value for this is exactly equal to 50 let us suppose. $V_{CC}$ is equal to 5 Volts. $R_C$ is approximately equal to 1 kilo ohm and $R_B$ is equal to, let us suppose, 10 kilo ohms.

So let us see how does my noise margin and all these things come out. So if you see, $V_I$ is equal to $V_{OL}$ equals to $V_{CEsat}$ equals to 0.2 Volts. So when you apply $V_I$ input, right? And I assume my output to be low. So my input is high (input is high), my output is low, equals to $V_{CEsat}$ equals to 0.2 right? Similarly, when I apply $V_{OH}$, input is high right? So in this case input is high, in this case input is low. And then I get $V_{OH}$ equals to $V_{CC}$ equals to 5 Volts. Right?

So when you input is low your output is $V_{OH}$ and that is almost equal to $V_{CC}$. And when your input is high output is low, almost equal to $V_{CEsat}$. So the swing is from 5 to 0.2 Volts. This is the swing in the voltage domain which you see when you have such type of a system available with you. Now assume, let us (let us) do one thing that I assume that my $V_{in}$ or input is between $V_{IL}$ input low as well as between $V_{IH}$, right? So I am in the active region of operation.

So in the active region, I can safely write down $A_V$ to be equal to $V_0$ by $V_I$, output voltage by input voltage which is minus beta times $R_C$ by $R_B + R_\pi$. We will discuss this time and again. But if you can understand $R_C$ by $R_B$ is the ratio of the collector current to the base current right? So it is the ratio of the collector current and the base current. $\beta$ is equal to $I_C$ by $I_B$. So if you multiply it I get a voltage gain $A_V$ and because $\beta$ is basically output current by input current, right?

If you multiply this with $R_C$ and this if you multiply with $R_B$, I get output voltage upon input voltage which is exactly equal to $A_V$ value. Fine? Is it okay? So with this concept, if you put the value of $R_C$ as 1 kilo ohm, $R_B$ as 10 kilo ohm and you also put $\beta$ to be equal to 50, I get $A_V$ to be approximately equal to -5 Volts per Volts. Right? Which means that it is basically negative because obviously you are entering into a (negative) 180 degree phase shift, output is always 180 degree phase shift with respect to input and therefore you get $A_V$ equals to - 5 Volts by 5 Volts right?

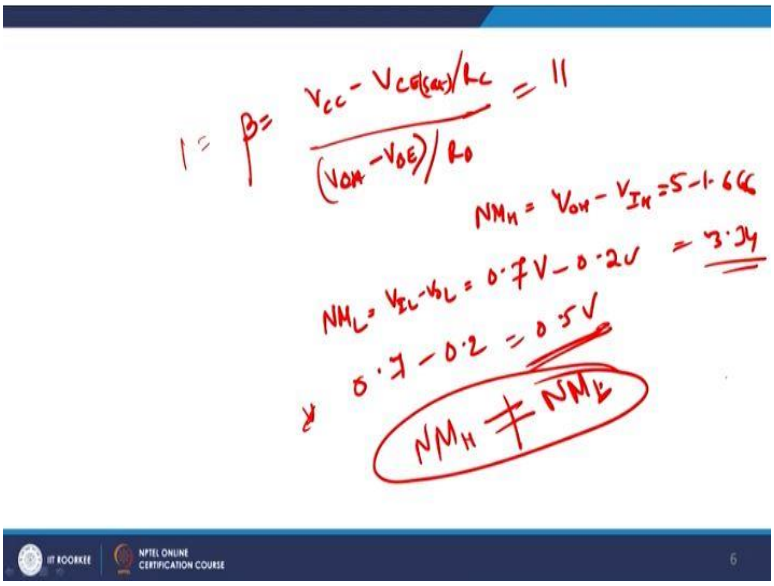
Now within the active region therefore, within the active region, I get $A_V$ equals to this much. Now at when your $V_I$ is equal to $V_{IH}$, because that is what you will get, I will enter into saturation region and I can safely write down $I_B$ to be equals to $V_{CC}$ - $V_{CEsat}$, right, divided by $\beta$ divided by $R_C$. Because if there would not have been $\beta$ then it should have been $I_C$.

But you put a $\beta$ here and then directly convert into $I_B$. So if you solve it, I get $I_B$ to be approximately equal to 0.096 milliamps. So it is the order of microampere right? Now, if I can write down $V_{IH}$ equal to $I_BR_B$. $I_BR_B$ means $I_B$ times $R_B + V_{BE}$. Means this voltage which is $V_{IH}$, let us suppose DC bias is exactly equal to the voltage here and the voltage this one. So if you add these two together, I should get the third voltage.

And therefore I get $V_{IH}$ equals to $I_BR_B + V_{BE}$, right? Put the value of $I_BR_B$ and $V_{BE}$, I get 1.66 Volts. So input high is 1.66 which we follow right? Now, if this is true, I can write down $V_{IH}$ therefore to be equal to 6.66 ahh,,, 1.66 Volts, right? So therefore $V_I$ equals to $V_{OH}$ equals to

5 Volts, right? And $V_O$ will be equal to $V_{CEsat}$ and that is equal to 0.2 Volts. Fine? So I get these many definitions available to me, right?

(Refer Slide Time: 24:57)



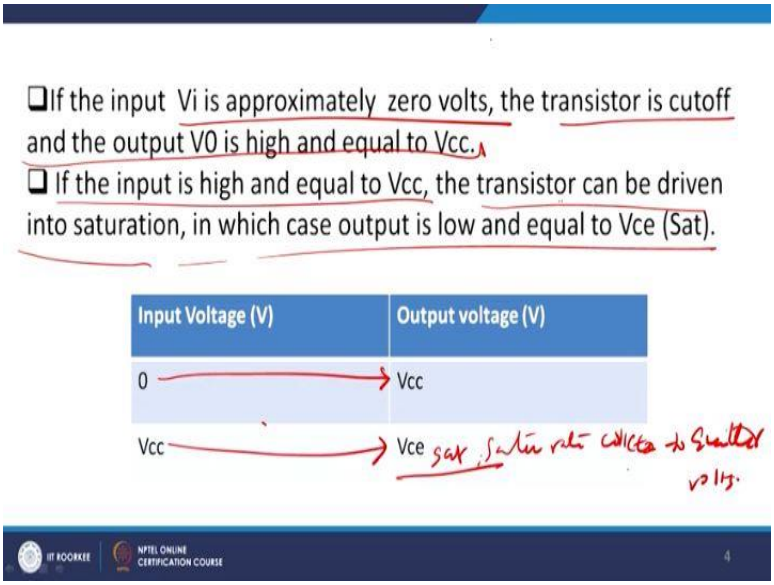

Now, if I assume these definitions are equal then from here, I can get the value of $\beta$. $\beta$ is the current gain to be exactly equal to $V_{CC}$ - $V_{CEsat}$ right, divided by $V_{CC}$ - $V_{BE}$ or $V_{OH}$ - $V_{BE}$. This divided by $R_C$, this divided by $R_B$. If they are exactly equal, $\beta$ will be equal to 1, right? And that is what you get. But in clear (clearly) this is equal to 11, which you get. Therefore $N_{MH}$ will be equal to $V_{OH}$ - $V_{IH}$ and this comes out to be 5 minus 1.666 and that is equal to 3.34.

In the second case when I find $N_{ML}$, it is $V_{IL}$ - $V_{OL}$. So what is the value of $V_{IL}$? 0.7 Volts - 0.2 Volts and that is equal to 0.5 Volts, right? So we end up having a very interesting discussion on bipolar transistor, that if you don't do any manipulation in terms of doping or circuit changes, you would expect to see $\beta$ to be approximately of the order of 10 to 15. Once you have that $\beta$ value, you can achieve a current gain of approximately 10 to 15 within the chip right? Or within the design which you have (which you have) made.

So let me find out $N_{MH}$ and $N_{ML}$. And that comes out to be 3.34 and if you find out $N_{ML}$, it is $V_{IL}$ minus $V_{OL}$. So $V_{IL}$ is 0.7 - 0.2 and this is 0.5 Volts. But you see, your $N_{MH}$ is not equal to L. Which means that there is a problem. That means high to low noise margin is not equal to low noise margin. Which means that 1 can be transmitted very easily even with a noise figure of very high noise figure. Whereas 0 will be stopped or inverted if my output noise is typically small.

So in reality, we should have $N_{MH}$ equal to $N_{ML}$. But in practical, it says as you can see in this bipolar transistor, $N_{ML}$ is not equal to $N_{MH}$ right? And that is a problem area which people face as far as designing is concerned. So somehow or other you have to make $N_{ML}$ equal to $N_{MH}$ right? And that is what is very important, right? Okay, so with knowledge or this basic criteria, we have therefore understood the basic concept of a simple BJT inverter, how does it works and what are the various features of a BJT inverter.

(Refer Slide Time: 27:46)




Therefore, as I discussed with you, if the inverter $V_I$ is approximately 0 Volts input, the transistor is cut off and the output voltage $V_0$ is high and equals to $V_{CC}$. So when you input is low, output is high. Now if the input is high and equal to $V_{CC}$, the transistor can be driven to saturation in which case the output is low and equals to $V_{CEsat}$. So the output is either equal to $V_{CC}$ or equal to $V_{CEsat}$.

Two extremes which are available for this particular design right? And you can actually switch from one to the other very fast in this case. Right. Now if you therefore see, that takes care of our understanding of your transistor action and you can therefore see that when input is equal to 0, the output is equal to $+V_{CC}$.

When the input voltage is $+V_{CC}$, the output voltage is $V_{CEsat}$ also known as saturated collector to emitter voltage, right? So we will do the subsequent talks the next time. We will look into the

second order effects in a detailed manner in the next turn and we will go into the details of each one of them as we move along.

So what have you learnt? BJT as an inverter, what is a noise margin, how we calculate the noise margin from the device characteristics and how noise margin is related to the overall sensitivity and robustness of the inverter design, right? With these words, thank you very much for your patient hearing. Thank you.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−12**
**BJT Second Order Effects - I**

Hello everybody and welcome to the NPTEL online certification course of Microelectronics: Devices to Circuits. In our previous interactions, we have actually looked into BJT as an inverter and we have understood how BJT's inverter can be made using a simple common emitter or emitter grounded BJT. We also seen that the noise margins are not equal. So, basically $N_{ML}$ and $N_{MH}$ are not equal, right? And therefore though it, though therefore a BJT will be a good acceptor of one, it might not be a very good acceptor of zero or receptor of zero right?

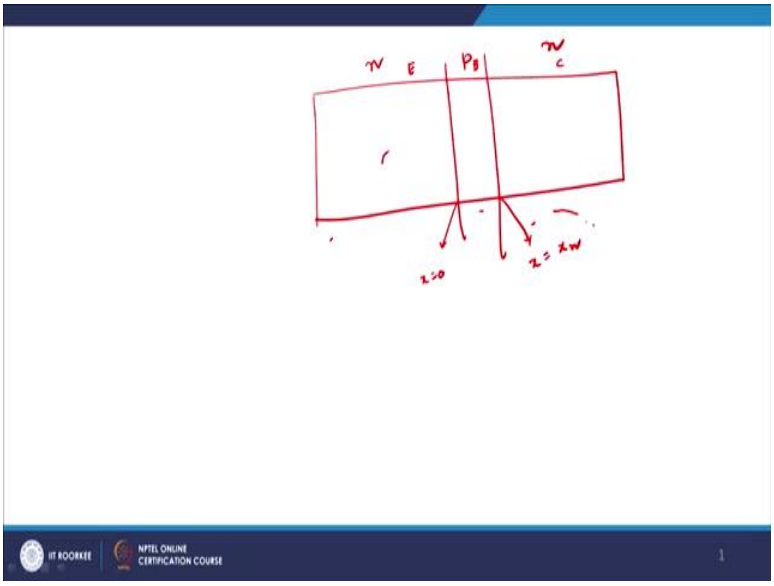
So if your low noise margins are typically very low, then even a small noise when you are inserting zero will result in a change in the value of an output voltage. And that is the problem in any of your BJT that your noise margins are relatively not very high and therefore you generally get a noisy reception in the output side. First thing.

Second thing is that typically as I discussed with you, since you have to if you want to use it as a switch, you have to move from saturation to cut-off and vice-versa. And therefore removal of charge carriers from saturation will take some finite amount of time, right? And therefore the time taken to switch from cut-off to saturation and back to cut-off, there is a finite time and therefore you will always have a limitation at which the BJT can work in terms of frequency limitations.

What we will be now looking is basically into bipolar technology's second order effects. Which means that the first order effects were primarily the drift diffusion phenomena which the bipolar technology works and we have also seen the current equation there. We will be now looking at some second order effects, which means that those effects which generally don't rise as such but may become prevalent under certain conditions, right? So under certain conditions they might show their influence on the current voltage characteristics of the device. Otherwise they are quite silent and they don't, they are not so much receptive to variations.

So what we will be looking therefore is, one is the drift in the base region and what people have done now, we would just like to show you what we have done here is that if you look at this equation $I_{NX}$, which is this one, right? This one.

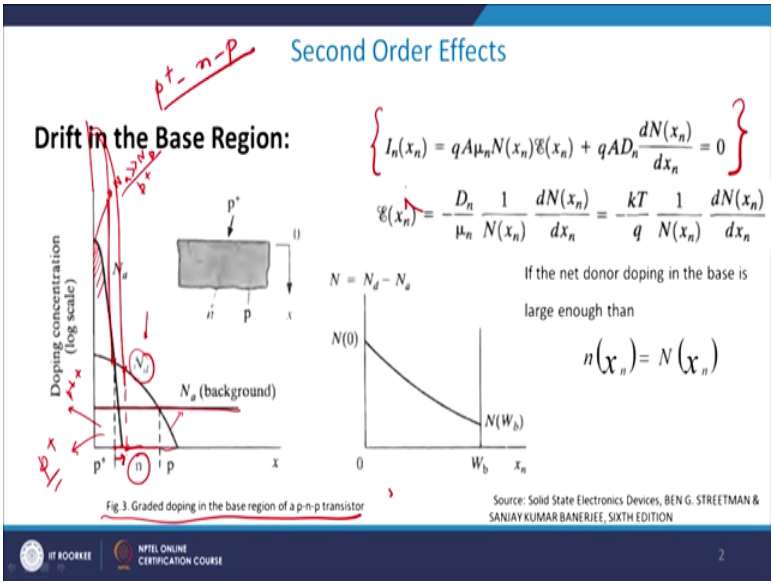
(Refer Slide Time: 2:56)




The first part, so this is the total current because of the charge carriers which is there in the region. So I have got NPN transistor right? And within that, as I discussed with you the P region. So I have got NPN. N has got the largest doping, highest area. P has got the minimum doping and minimum area and N has got the, or collector has got the relatively moderate doping.

So what I am trying to tell you is that from this point, let us suppose this accounts to X equals to zero to this point which accounts X equals to say $X_n$. Then we know that the very, how is the electron profile which is basically the majority current carried in the P type base varies, right? So this is what I wanted to discuss with you in this case.

(Refer Slide Time: 3:48)




So if you go back to the slide, you can see that. So if you look at this figure 3 which you see, this is basically known as a graded doping and you can see here that what we try to do is that we first make an acceptor grading.

So acceptor background is there. So this background is basically acceptor which means that all of it is basically made as P type, first of all, right? We make it all P type. The whole of it, we make it as P type. And then what we do is that, we then dope it with donor atoms. So this is the donor atom which we are doping with and then we ensure that the donor atoms exceed the acceptor atoms at certain regions which is between this point and this point. And therefore these two points are basically the edge of the base.

So this is my base for a $P^+$, for a $P^+$ NP BJT. Since your $N_d$ is larger than $N_a$ at this point and at this point, right? So they are larger in the sense that they are more in value. So I would expect to see a larger donor concentration and therefore this will become more N type as compared to P and that is the reason you get P type here. In this region of course, $N_a$ is therefore larger than $N_d$. And therefore this is basically a P type.

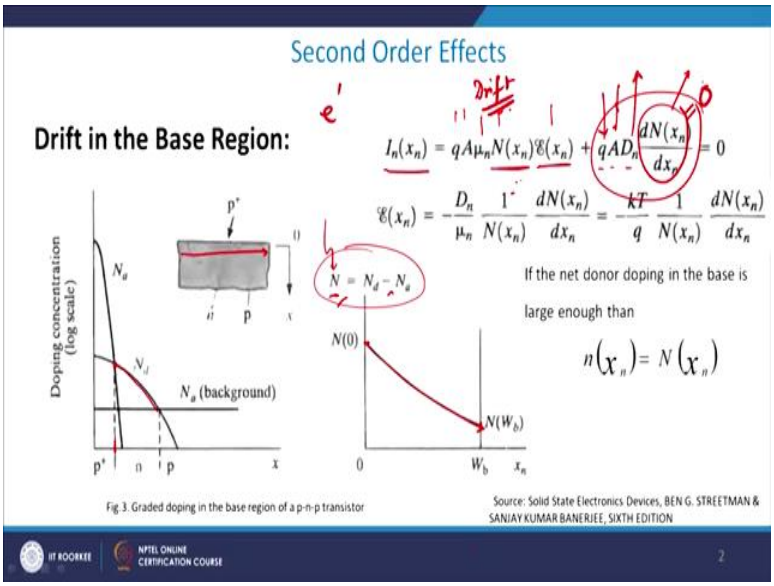
Since it is very very large, as you can see it is quite large as compared to $N_d$, we refer to this as a $P^+$ region because the doping is very very high, right? Similarly, when you reach to this side, your acceptor ion, your dopant $N_d$, the donor ion actually reduces much below the value of your acceptor ion concentration. And as a result this becomes a P type. So I got a $P^+$ NP.

So how do you do it? This is also known as a graded junction doping or a graded doping profile. And what we do here is therefore in the graded doping profile, we vary we can control the width of my base as well as of the emitter as well as the collector by simply changing the dopant profile, right? So, doping profile right? That is very simple and easy to do action. You can also change the doping.

For example if you wanted to make this one more $P^+$ type, so you want to make it $P^{++}$ then you simply increase the acceptor ion concentration in this manner. You wanted to shift this line from this side to this side then you need to cut somewhere here. So when you cut the $N_a$ here, the starting value comes out here. So I can do all sorts of manipulation and I am able to achieve a very straight forward methodology of having $P^+ N$.

Now let us suppose your electrons are the majority current carrier, holds are the majority current carriers. Therefore electrons which are there are basically your minority current carriers.

(Refer Slide Time: 6:43)




What we are assuming is that let us suppose the effective charge carrier at a particular place is nothing but $N_d$ - $N_a$ right? So you had $N_d$, let us suppose $10^{15}$ and $N_d$ was $10^{19}$ and Na was $10^{15}$ or $10^{16}$. Then you subtract both of them and that is the effective $N_a$ or the effective number of dopant species.

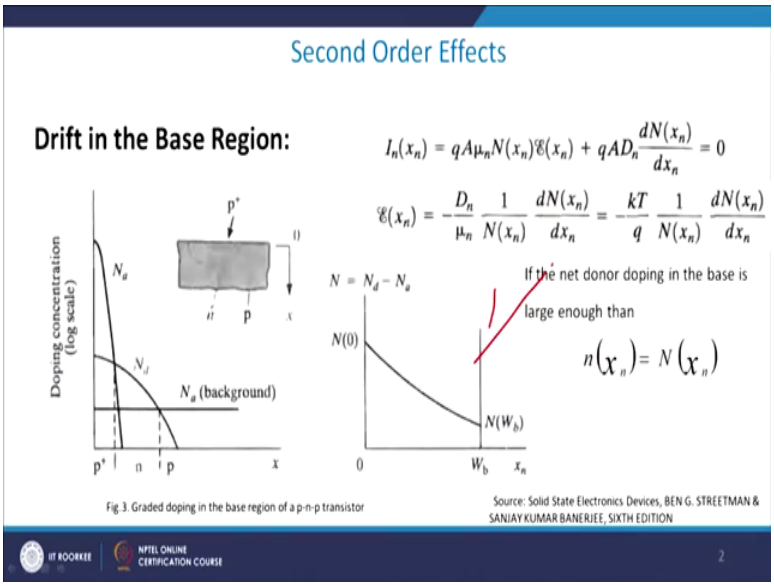
So if we have 100 % ionization, I would expect to see that that many number of free charge carriers, if $N_d > N_a$, free charge carrier is electrons. If $N_a > N_d$, free charge carriers are holes. We will be of course available to you. Now to understand this, let us suppose I have electrons as the majority current carriers. Then I define the current $I_n$ at any point $X_n$ within the circuitry, within the BJT. So this is the value this is the X variation, this is the X variation, is equals to q times A $qA\mu_n N(X_n)E(X_n)$.

I will explain to you what these terms are. N is the charge density, the dopant charge density which you see, E is the electric field from base to the, from the emitter to the base region, E electric field at that particular point, Q is the electronic charge, A is the area of the cross section of the emitter or the base, and $\mu_n$ is basically the mobility of the charge carriers, right? So this is basically by virtue of your sort of a drift, a drift current which you see. Right?

Similarly, you will also add a current which is by virtue of diffusion because, please understand that the doping within the base region is not fixed right? It varies because as you can see here, $N_d$ is varying like this. So I would expect to see that at 0 width, at 0 or at when the base starts, the doping concentration is maximum and then it starts to fall down in this fashion, right? And that is what is shown here that I have got maximum width here, maximum number of doping concentration and then it starts to fall down as we move from the edge of the base to the, edge of the emitter to the edge of the collector.

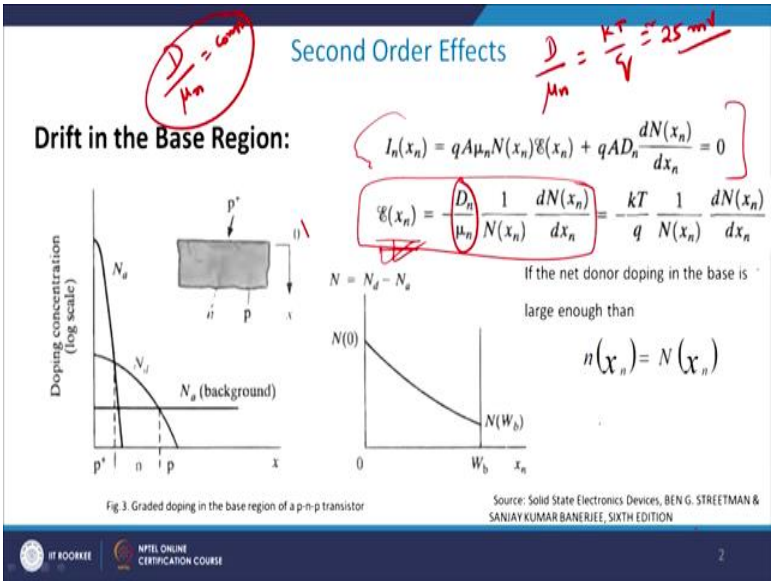
And therefore, we have a diffusion term which is given as $qAD_n\ dN(X_n)\ d\ (X_n)$. Which means that it is the charge again, A is the area right? D is the diffusion constant for electron or a diffusion coefficient for electron. And $dN(X_n)\ d\ (X_n)$ is basically meaning the number, how the number of charge carriers are varying within this point. Now please understand if the doping concentration within the base side would have been constant, my this term would have been equals to 0. Constant means independent of the distance the doping is always fixed. So if it is 0, the only contribution would not have, would have been only from the drift component and not from the diffusion component.

(Refer Slide Time: 9:33)




And, all of you can therefore understand that therefore obviously the drift component will have always a component which is primarily proportional to the area as well as to the effective number of charge carriers or number of charge, doping species in that particular point.

(Refer Slide Time: 9:40)



Similarly from Einstein's equation or from otherwise, we can find out that E ($X_n$) means electric field at any particular point ($X_n$) is given by this formula, the same formula which we get right? And we can from there we can get the value.

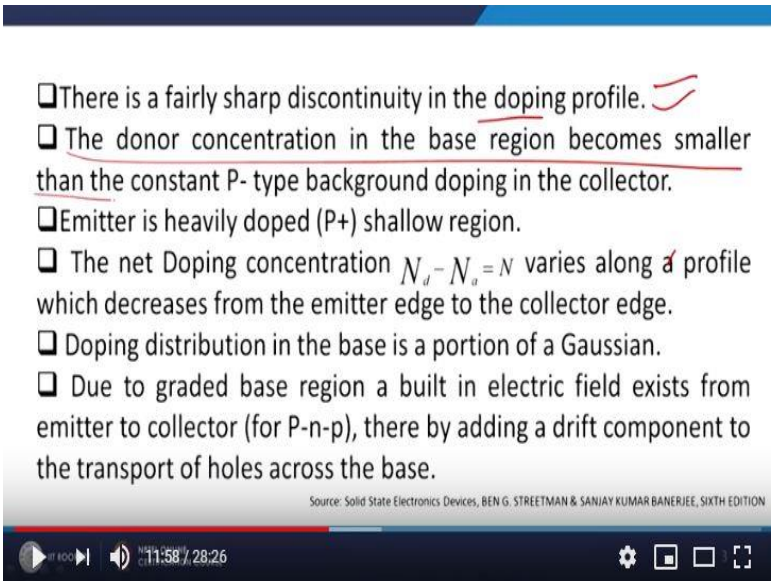
So this is very simple and straight forward. From this equation only, from this equation only I can get this very easily, right? Provided you make the drift current equals to 0. Then I get $D_n/\mu_n$, $D_n/\mu_n$ into $1/N_x$ dN ($X_n$) is equal to – kT/q, right? 1/N ($X_n$) and d ($X_n$). Because why it is kT/q because $D/\mu_n$, $D/\mu_n$ is equals to kT/q. kT/q is constant at 300 K, approximately 26 millivolt or 25 millivolt. So $D/\mu_n$ is always a constant quantity, right? At 300 K, it is a constant quantity.

And therefore we can safely write down this $D/\mu_n$ to be a constant and that is equal to – kT/q minus equals to – kT/q upon 1/N ($X_n$) dN (nx) dx. Therefore the total electric field at any particular point E ($X_n$) will be a strong function of how the diffusion species is varying with respect to space right? That is very important which you see here. It will also depend upon the number of dopant species available at a particular point $X_n$ within the network.

If that is very large, the electric field, the absolute value of the electric field will be fall, right? Will be falling down. So we have understood therefore that in the base region if there is a small amount of change in the value of $N_d$, I will have the drift in the base region. Therefore there will be a drift carrier, there will be a drift, electron will be drifted through the base region right? And that is quite an interesting phenomena which people see.

So as you can see there is a fairly sharp discontinuity in the doping profile, right? Your doping profile shows a sharp discontinuity. Which means that since the doping, and the reason being that since there is a certain change at the interface between $N_d$ and $N_a$, right? And therefore there is a sharp change.

(Refer Slide Time: 11:58)




So as you can see the donor concentration in the base region becomes smaller than the constant P-type background doping in the collector. That we have already discussed in our previous slide, we have already discussed that the donor concentration in the base region, right?

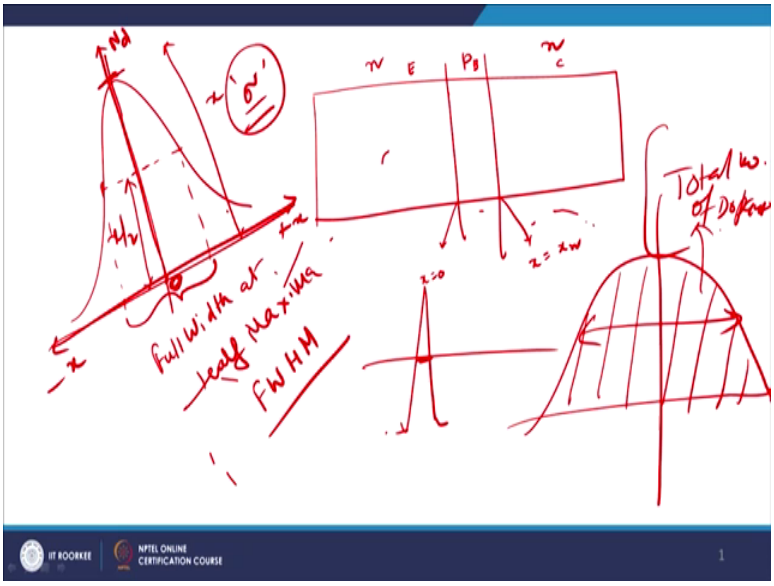
So if you remember the previous discussion was, the previous discussion if you look, this is $P^+NP$ right? This is $P^+$ N and P. So what I am trying to tell you is that the donor concentration, the base region becomes smaller as compared to the collector concentration on the collector side.

So on the collector side you still have P but on the base side, since donor is slightly higher, you get N region available to you. And as you can see emitter is heavily doped shallow region, it is a heavily doped region. Why? Because the acceptor concentration is much much larger as compared to the donor concentration there. Now the net doping concentration $N_d$ - $N_a$ varies along a profile which decreases from emitter edge to the collector edge. This we again discussed. Why? The reason being very simple that if you look very carefully, the $N_d$ is falling from this edge to this edge, right?

And since it is $N_d$ - $N_a$, right? So since Na is constant at this stage, I would expect that the profiling here will also be between this point and this point, right? Parallel to this line, right. Fine? And that is what you get. That it is varying, it is maximum at the emitter edge and minimum at the collector edge, right? Doping distribution in the base is basically can be simulated using Gaussian.

Gaussian functions or a Gaussian doping concentration is something that looks like this.

(Refer Slide Time: 13:28)




It is known as inverted dumbbell shaped doping profile. The idea here is this is known as this is maximum, right? And this is say, this is X. Then, this is approximately X/2 and we define this width, right? We define this width as full width at half maxima. Right?

So we define full width at half maxima, FWHM right? Now if you make a FW. So this is basically the distance and this is the doping concentration. So this is $N_d$ and this is the X, right? So this is –X , +X , 0. So 0 you have the highest doping and as you move from highest doping to left and right, it falls down. Now if you want to make the doping very very sharp, you just have to make your FWHM very small. Right? You have to make FW... If you want to make the doping a very constant doping, do something like this; spread out. Right? Spread out. And then your FWHM becomes large.

So what is the advantage of a Gaussian profile? The advantage of a Gaussian profile is that it helps you to give a profile or it helps you to control the profile from maybe a constant profile, constant doping profile to a sharp peak profile by simply changing the value of your FWHM or full width of half maxima, also referred to as sigma in most of the cases right? Sigma in most of the cases and that is what you refer to as a standard deviation but technically it is full width of half maxima. And therefore you can see that automatically what you get from here is, that simply

by changing the value of FWHM, either you can reach a peak value of doping concentration or you can actually reduce the doping concentration drastically, right? And so and so forth.

So this is what we get. The area under this curve is basically the total number of dopant species total number of dopant species. Dopant or acceptor whatever you want to do, right? So that is the reason, Gaussian profile is most preferred profile in a design because you can control it very, in a much better manner. So due to a graded base region, a built in electric field exists from emitter to collector, thereby adding a drift component to the transport of holes across the base.

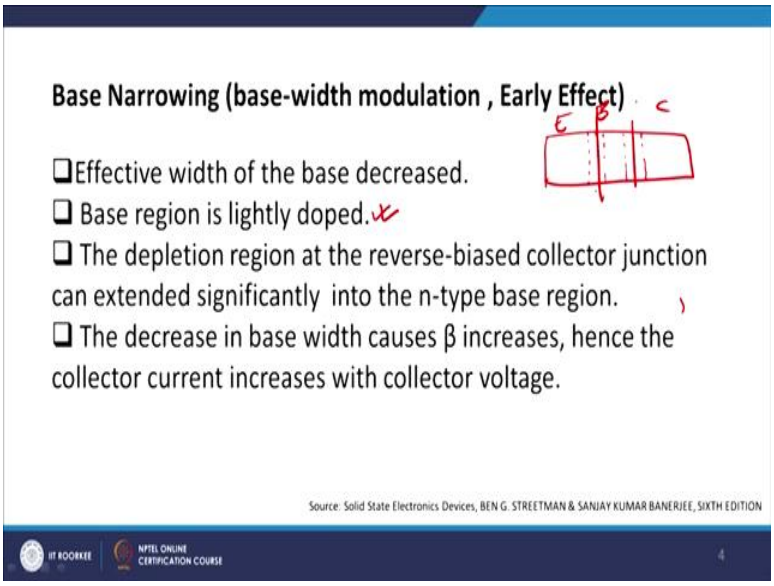
This is very very important right? So you see, what I was talking about in the previous case that what is happening is, if you just look at it, that because of the graded base region, there is an electric field from emitter to collector, right? For a PNP, right?

And therefore... so if there is a electric field for example from this direction to this direction. So if you have holes entering the base region, right, there will be that will be always drifted by the strong electric field towards to collector side, right? And that is the reason drift in the base region right?

So that is the reason what we get. So the advantage of it is that you get a larger number of charge carriers on the collector side because of this drift in the base region. Right? We come to the next second order effect and that is basically known as early effect or what is also known as base width modulation effect. We have already studied this but I will just go it a bit fast in this case so that you understand the issues in a much more detailed manner. Base width narrowing effect or early effect is to do with what?

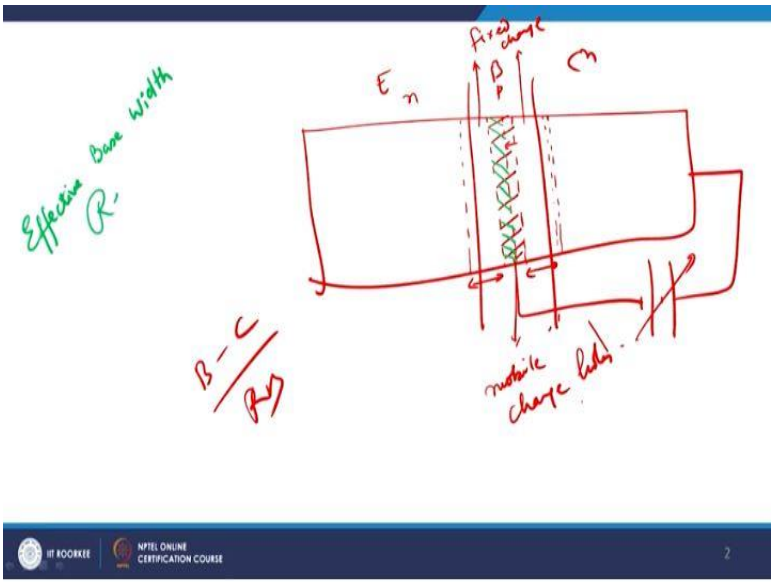
That if you remember, your base is lightly doped as I have written here, this is lightly doped.

(Refer Slide Time: 17:15)




And your, what happens is that if your base is lightly doped, so I have got emitter base and emitter... Sorry this is emitter base and collector. So emitter base junction will have a depletion region like this and collector base will have something like this. But as I have told you previously also that the depletion thickness will be larger towards that region whose doping concentration is low. Right?

(Refer Slide Time: 17:48)




So what I am trying to tell you therefore is that if I have a PN junction or maybe an NPN transistor which is a BJT, and let us suppose this is emitter base collector and since base is relatively lowly doped, I would expect to see the doping concentration something like this. So

you will have it like this; they are not equal. Similarly this will have like this and you will have like this. But since emitter base junction will be forward biased, so this total distance will be smaller than this total distance, clear? So what is the effective base width now, is this much.

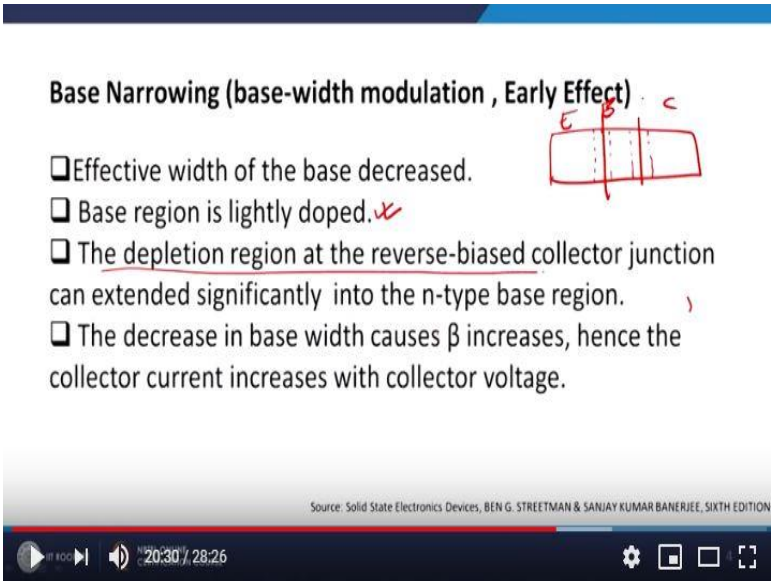
This is the effective base width where doping is there. Otherwise these are all fixed charge carriers. These are all fixed charge, fix charge. This is the only place where you have a mobile charge mobile charge and those are holes, and those are holes. And these are fixed charge here so basically they don't contribute to the overall current. Now, as I discussed with you when you are active forward mode, your base collector is always reversed biased, right? So when you reverse bias it, that means if this is N P L, let us suppose, if you reverse bias it then this will be something like this, right?

So when you, once you reverse bias it, this depletion thickness will further increase as we had discussed earlier. That means reverse bias will increase the depletion thickness. So the effective depletion thickness increases this side. So what happens to the effective base width? What happens to your effective base width? It reduces actually, if you see very (carefully) clearly. So your effective base width now becomes this much. Fine? It has reduced.

When it has reduced, so this is basically known as therefore, base width reduction or effective base width reduction. So once it reduces, now you have lesser number of carriers in the base region and therefore the electrons moving from emitter to the base side, the recombination will be smaller and larger number of electrons will be reaching to the collector side and you will be reaching a very large value of $\alpha$, very close to one which is 0.99 maybe. Right?

And if your this base region would have been quite thick, your recombination would have been higher and your collector current would have been much smaller as compared to your emitter current. So with this fundamental understanding, we just now come to the...

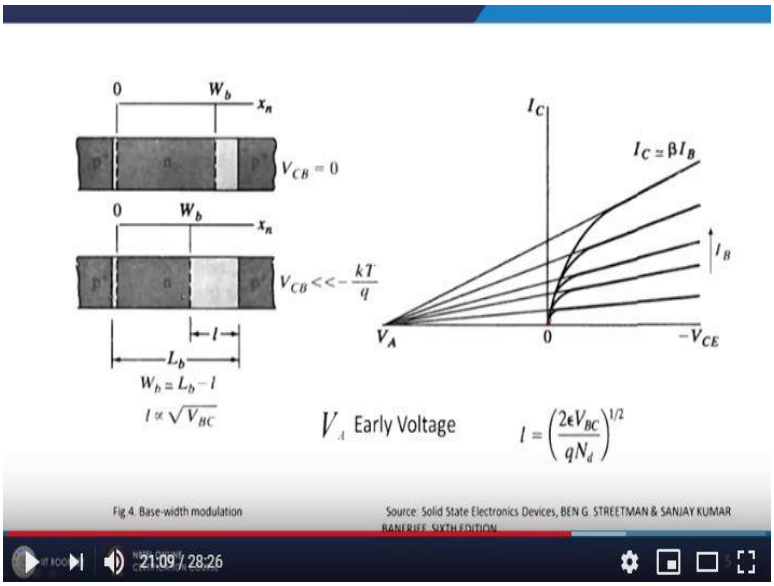
(Refer Slide Time: 20:29)




So the depletion thickness at the reverse collector junction can be extended significantly to the n type-base region. We have discussed this point.

And the decrease in the base width causes β to increase and α to increase as well. Hence the collector current increases with collector voltage. Right? So (what do) please understand, till now we were assuming that the collector current is always constant independent of the value of $V_{CB}$, but now, what we are seeing is since now beyond a particular higher value of $V_{CB}$, your effective base width is reducing. I would expect to see an increase in the current. So that is what we are seeing also.

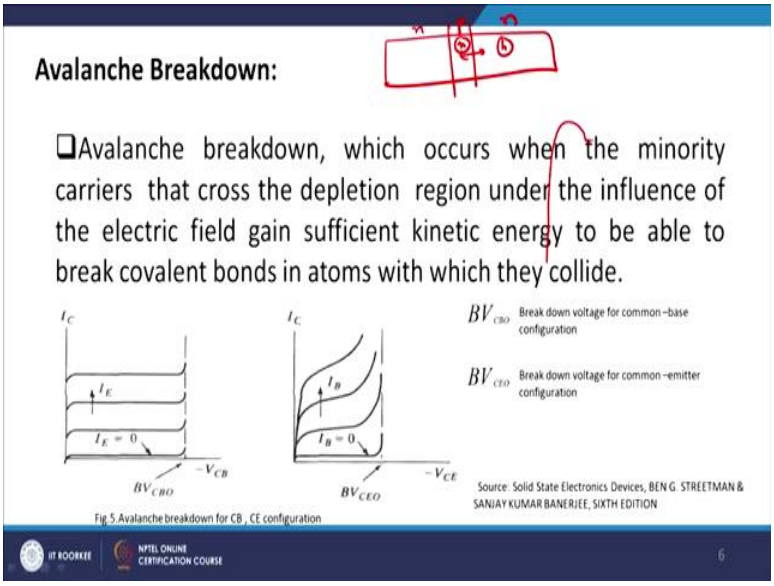
(Refer Slide Time: 21:09)




So as you can see here, the current is increasing, it is saturating. There would not have been an early effect, it would not have been something like this, right? It would have been something like this, constant. But then what is happening? At a higher value of $V_{CE}$ or $V_{CB}$, your this profiling is becoming high, right? Which means that, what is happening is that this value is becoming high and therefore $I_C$ becomes large.

Now the way how to find the voltage if this happens is, that if you back it all of your extrapolate it or interpolate it backwards all the line, interpolate it backwards, the place at where it meets is defined as my early voltage. Right? It is known as early voltage. An early voltage therefore is that voltage at which they will start behaving like a fixed current source, right? And this is what the value of current is.

(Refer Slide Time: 22:01)

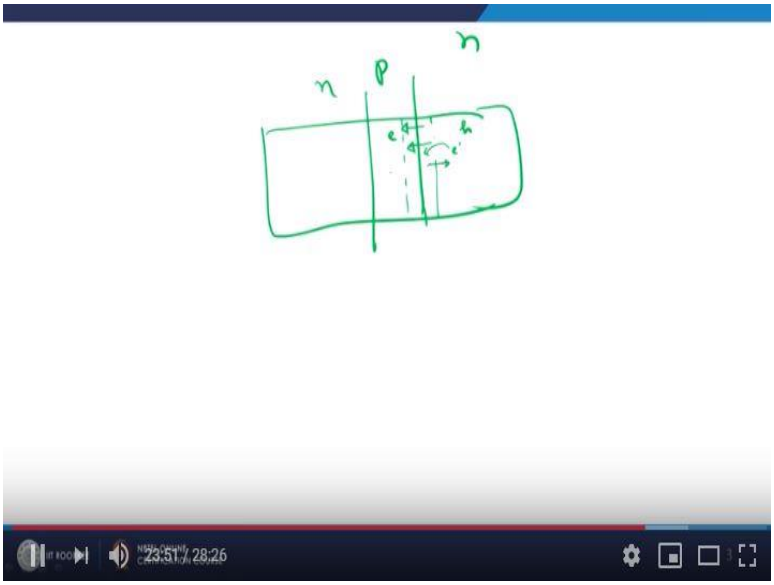



Let me come to the third second order effect and that is the avalanche breakdown. And I think it is very simple and straightforward. Avalanche breakdown is, remember your base collector is always reversed bias, right? We had been talking about for this long time that base collector will always be reversed bias. So when the base collector is reversed bias, for, so you will have minority current carriers (hole) holes here. And you will have electrons as a minority current carriers here, right?

Now when you reverse bias it, electron from N-type and holes from P-type won't be able to move because the depletion thickness is very large. But for the minority current carriers which is hole on the N side and electrons on the P side, it is more of a hill. It is going down. It is not a up hill, it is down hill, right? So therefore I will expect to see a large amount of current because of minority current carrier. But since the number of charge carrier, minority current carriers are very low, therefore the charge associated with it or the current associated will be very very small.

Yes, it will be small of course but let us see what happens.

(Refer Slide Time: 23:18)




Now this…Once they start moving, so I have this junction, this is emitter based junction and I have got NPN and electrons are available here and holes are available here, right? So the electric field internally which is there in the depletion region between P and N because it is reversed bias so I have a depletion region here, right? (And I would expect) so the internal electric field will be from this side to this side.

So any hole, any electron trying to enter from this side to this side, electron is a majority current carrier in N type, will be forced to go this side. Similarly any hole entering from this side to this side will be again forced to go this side. So any majority current carrier contribution will be almost 0. Anything which will be coming will be directly coming from emitter.

But for these holes and electrons which are minority current carriers in collector and base respectively, if the hole enters here it will be dragged by this electric field within the depletion region to reach this point. So if you go on increasing the value of your reverse bias collector junction, right? If you go on increasing it drastically, so these electrons and holes will gain large amount of energy and when they pass through this depletion region, they will ionize and form electron and hole pairs.

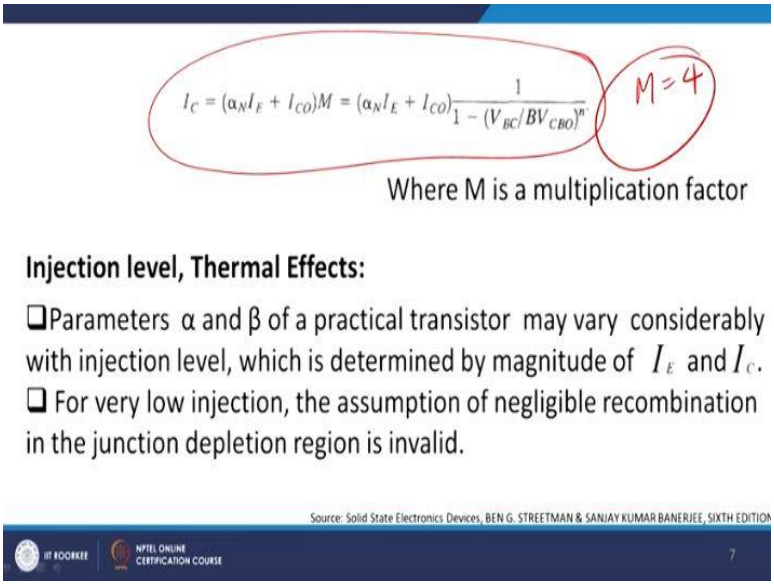
So what will happen is, beyond a particular point, you might have a suddenly large increase in the value of electron, right? And that is what is known as an avalanche breakdown.

But due to this sudden electron charge, you will have a large electron current and therefore if you plot $I_C$ versus $V_{CB}$, it is this constant we have discussed this point, constant current. But as you go on making it more and more negative biased and make it larger, suddenly a time comes when the current suddenly increases, almost drastically like this.

And the reason is that you now have large amount of ionization and large amount of electron-hole pair formation because of which there is a large current available to you, right? These are primarily, electron-hole pair formation is taking place. You can also have another thing that electrons become move very fast and they ionize the atom. So the exterior most electron from the (kinetic is so) the kinetic energy of the incident electron is so large that it directs all its energy to this peripheral atom, electron and the electron comes out of the atom, right? And it becomes a free electron.

In any case you will have a very large quantity of free electrons which will be available and therefore you could expect to see a sudden increase in the current, right? This breakdown voltage is defined as $V_{CBO}$ when it is a common base configuration. Also referred as $B_{CEO}$, BV suffix CEO when it is a common emitter configuration, which is there, right?

(Refer Slide Time: 26:06)



So we have these two types of configurations available. This is the net current which one sees because of the multiplication factor, because of M.

M is the multiplication factor primarily meanings that for each inonization how many number of electrons are formed. So if one electron comes and form four electrons in the multiplication factor M is equals to 4, right? And you will have four number times of electrons and hole current available right?

So with this, we have understood the basic three mechanisms, basic three mechanisms of breakdown. When we meet next time, we will be actually looking in to the next two mechanisms a second order effects. As you can see therefore, these are special cases as I discussed with you. Why special? Because only and only when, when your collector base or a base collector junction becomes larger than this breakdown voltage, then only you will expect to see a very large increase in the current.

Otherwise the current will be almost constant or remain as it is. So therefore these effects are known as second order effects. What we have learnt? Avalanche breakdown we have learnt, one. We have also learnt your drift of charge carriers within the base region a second order effects and we have also learnt about the early effect or effective base width modulation.

When we meet next time, we will be actually looking into the other facts, as far as this dimension goes and we will see what are the other facts. We will look in thermal effects and we will be looking into the base resistor or emitter crowding effect. So these two effects we will look and that will ensure that we will be finishing off BJT with the understanding of these few fundamental things. Thank you very much.

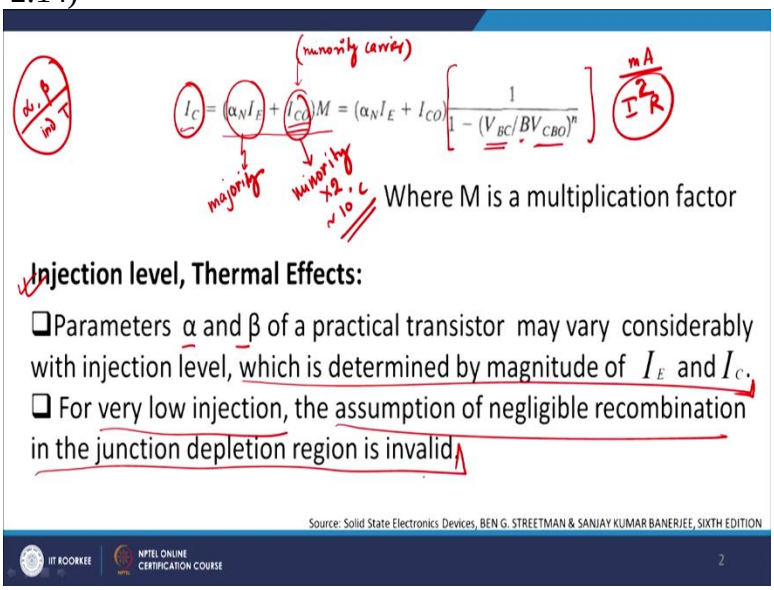
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−13**
**BJT: Second Order Effects - II**

Hello and welcome to the online certification course on Microelectronics: Devices to Circuits. In our previous module we have learnt the second-order effects which are prevalent in a bipolar junction transistor. What are these second-order effects? These are those effects which normally don't show their presence but under certain conditions of electrical parameters or structural parameters, their effects are pronounced and their effects are relatively seen or influence the electrical characteristics of the device. One of the phenomena which we looked closely was basically your, the concept of base width modulation effect also known as early effect and associated with that, we found out a voltage known as early voltage.

This makes the early effect phenomena makes the device behave as a non-ideal current source in the saturation region or in the active region, as we say. And we have also seen that there is a multiplication of charge carriers under effectively large reverse bias near the depletion region which results in a large change in the value of your collector current. We will start from where we have left in the previous term and see what the other second-order effects are and also look into the typical cross-section of a bipolar transistor and then we will recapitulate this whole discussion.

So typically, we would expect to see that this will be the last sort of presentation on bipolar technology. After this, we will move to CMOS technology because that will be the next structure we will see that.

(Refer Slide Time: 2:14)




So as we left the previous discussion, the collector current $I_C$ was equals to given as α N times $I_E$ where $I_{CO}$ is basically my reverse bias sort of a minority current carrier concentration. So this is basically minority carrier concentration, right? And what we are predicting is that the total collector current which is flowing through a circuitry will always be a sum of the majority current carrier contribution plus the minority carrier contribution. And this factor M is given by this term which is 1/1 - BC, base collector junction and... Base collector junction with a capital B to the power N where N is a basically a, an integer value.

Now depending on the value of N which is basically a positive integer, we would expect to see $I_C$ to be large or small right? We also see the fact that there is a contribution, this is basically your majority carrier contribution, right? This is a majority carrier contribution and this is primarily the minority carrier contribution, minority, right? Now please understand that with rise in temperature, you would expect to see minority carrier contribution almost double for every ten degree rise in temperature right?

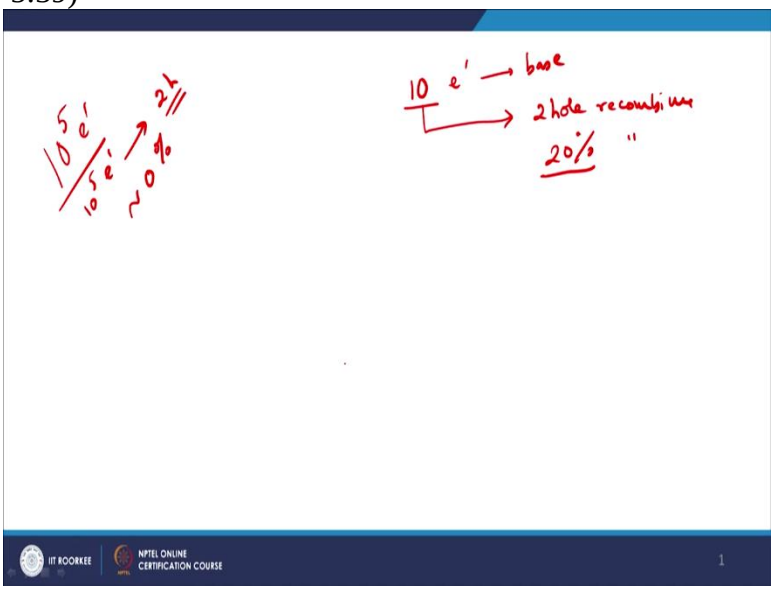
Or approximately five to ten degree rise in temperature, which means that the collector current doubles itself due to minority current carriers if you raise the temperature. We will come to that when we discuss other effects in our subsequent slides. We have also considered for the time being that α and β which is basically your current gain in common emitter and common base

configurations, these are primarily independent of independent of temperature. In reality, not true. And that is where our second issue is coming into picture.

Then what is the effect of thermal effect? So what is thermal effect primarily means that please understand, this BJT is working in an environment where applied voltages are very large. Not only that, your collector current is in the order of millivolts. So if your collector current is in the order of millivolts milliampere sorry, then I square into R which is basically the power dissipation which you expect to see from a device is typically very high, right? As a result, the on chip temperature for a bipolar technology based silicon design will be relatively high. So we cannot neglect the effects of thermal variations or effects of thermal changes onto the device characteristics.

And till now we were assuming that $\alpha$ and $\beta$ were independent of temperature. But now we will see that the $\alpha$ and $\beta$ are depending on the value of temperature. Not only that, they also depend on the magnitude of $I_E$ and $I_C$. Right? We have, this we have already seen, right? Now at very low levels of injection. What does it mean? Low level of injection primarily means that when your bias is so small that even in an NPN transistor, if the number of electrons being injected onto the base side is relatively very small, then we define that to be as a very low level of injection, right? Now the assumption of negligible recombination in the depletion region is invalid. I will explain to you what does it mean.

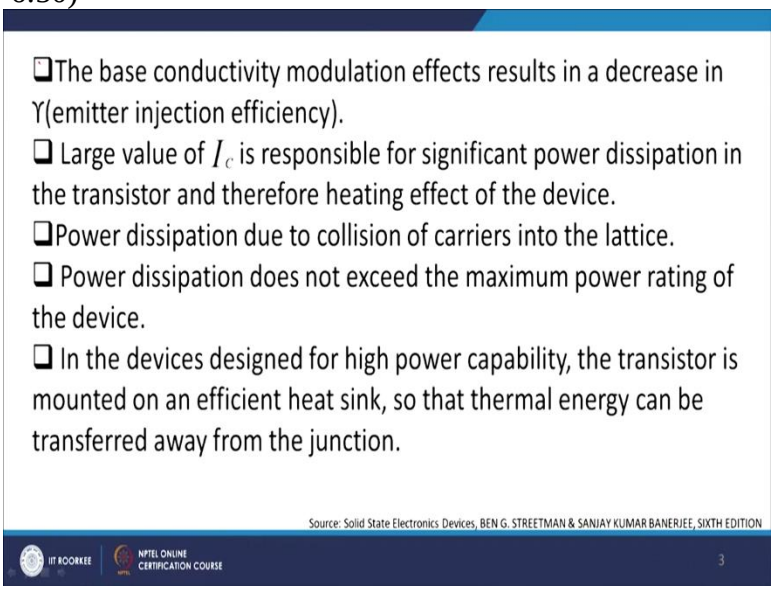
(Refer Slide Time: 5:39)



It means that that let us suppose you are injecting, you are injecting let us suppose in one case, which is low level of injection, you are injecting say 10 carriers right? 10 electrons are being injected onto the base side. Right? Whereas for a large injection, we will have let us suppose $10^5$ electrons being injected. So if you have 10 electrons, then with respect to 10, even if there are 2 holes in the base side which is recombining right, you still have a 20 % right recombination. Fine?

Whereas if you have $10^5$ electrons being inserted and you only have two (elect) two holes being combining more you can understand this negligibly small percentage of recombination, or almost 0 percent recombination. That is what it is written here, the assumption of negligible recombination in the junction depletion region is invalid. And the reason is that we cannot assume that the recombination current is very very low. Right? It will be very very high as compared to your base injection current and that is what is an important issue.

(Refer Slide Time: 6:50)




Now, we have also seen that this we have been discussed earlier but we will make it, we will discuss at this point which you see, this point if you see very carefully for the base conductivity modulation effect results in a decrease in the value of $\gamma$ which is emitter injection efficiency. What does it mean that? I will explain to you what does it mean. It means that let us suppose you did have a base width modulation, which means that you forward biased your base emitter and

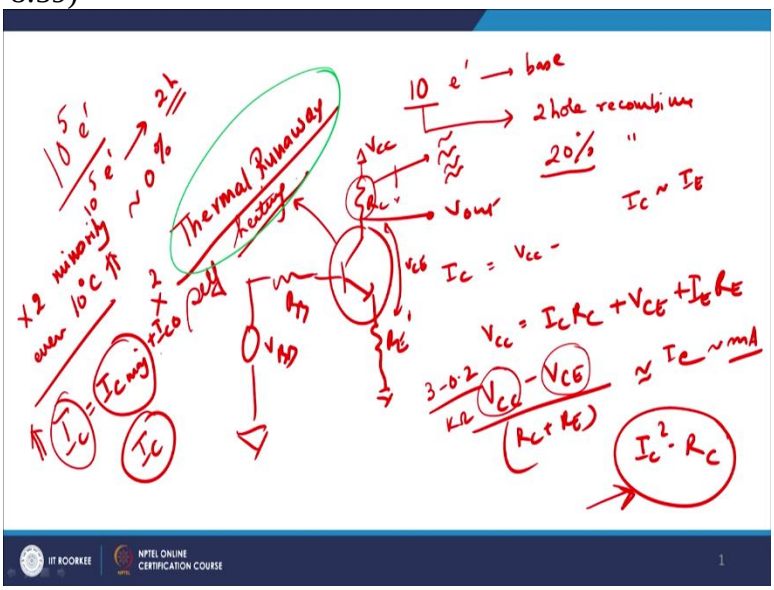
you reverse bias your base collector and you go on increasing the reverse bias of the base collector.

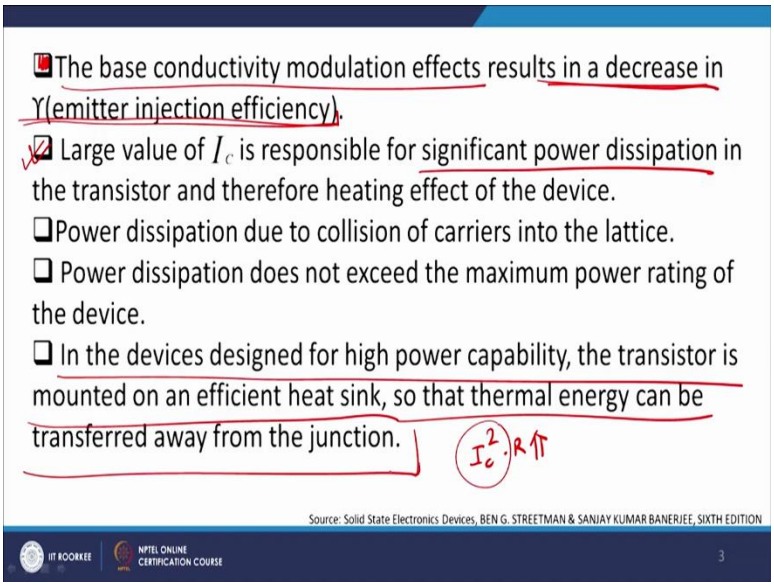
As a result, the effective base width is decreasing. Now once it is decreasing which means that the effective number of holes available is actually getting reduced and therefore its probability of recombination with electrons is also getting reduced. Right? So what we have learnt from this is that if the base conductivity is there, it will reduce the, it will reduce or decrease the emitter injection efficiency which means that it reduces the efficiency by which the emitter will be able to inject its carrier onto the base side because if it is very small, all the electrons will not be combining with holes, as a result the base the emitter efficiency reduces.

Now we are 100 percent sure that when you are doing common emitter configuration for example, your collector currents are very very high, right? Right? And they are very very high in the sense, they are typically very high. And as a result, there will be significant power dissipation in a transistor, right? And therefore as a result, there will be self-heating of the device itself. The reason is that so large $I_C$, so $I_C$ square $I_C$ square into R which is basically my power dissipation, this will be typically very high because your $I_{Cs}$ are very very high or collector currents are very very high.

And you see it is basically almost like a parabolic increase in the power dissipation. So even if you double your collector current, you are actually making the power dissipation increase by four times which is quite large power dissipation. Now I will just give you an idea what is known as thermal runaway. Right? Basic thermal runaway, concept of basic thermal runaway. Right?

(Refer Slide Time: 8:59)




So, we will discuss one basic concept which is known as thermal runaway. Now you see that suppose you have a common emitter base configuration, emitter grounded configuration with emitter bias here and you had a base collector junction here and you had a $R_C$ here and $R_B$ and $R_E$ right? And then you appropriately bias it with $V_{CC}$, this is with $V_{BB}$ and so on and so forth. Right? And you have an NPN transistor, this is your $V_{out}$. Now you see, in this case your $I_C$, $I_C$ can be given as, it can be given as $V_{CC}$ -, so if you solve it from this end to this end, I get

assuming that $I_E$ is the collector, so $V_{CC}$ will be equals to $I_C R_C$ right + $V_{CE}$. Which is $V_{CE}$? This is $V_{CE}$. Right?

$V_{CE}$ + $I_E R_E$, right? Assuming that the base current is very small, I can assume simply that $I_C$ is approximately equals to $I_E$ and therefore I can write down $V_{CC}$ - $V_{CE}/R_C$ + $R_E$ is effectively equals to $I_C$. Right? Approximately equals to $I_C$. $V_{CE}$ is typically 0.2 or 0.3, this will be around 3 volts. So 3 - 0.2 divided by a few kilo ohms, some kilo ohms. So this will be of the order of a few milliamps right? So it will be of the order of a few milliamps, collector current.

Now what happens is that, this $I^2$, $I_C^2$ into $R_C$ will result in power dissipation across this resistance, so there will be a power dissipating, right? As a result this transistor will start get heated away, right? And there will be heating. So what is also known as self-heating. We will explain to you what this self-heating is. So as the temperature rises, please understand, that by my previous discussion just now, my minority current carriers almost doubles for every ten degree rise in temperature. For minority, not for majority, please understand.

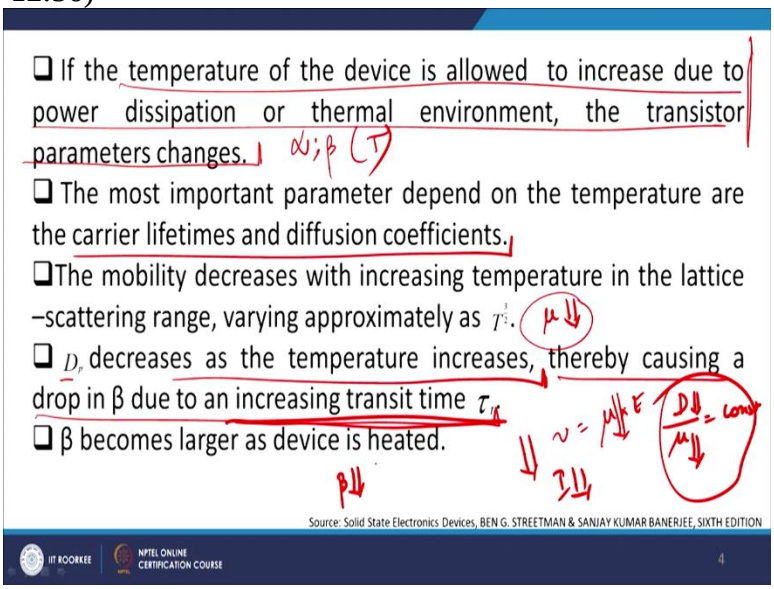
For minority carrier doubles for every ten degree centigrade rise in temperature right? So if a (temp), so you have increase in power dissipation which results in increase in temperature. This results, this increase in temperature results in increase in minority carriers. Now $I_C$ overall current will be made up of $I_C$ majority + $I_{CO}$ or $I_{CO}$ minority. Now though this is constant, but this every time doubles itself which primarily means that you're $I_C$ will again increase, with temperature. As $I_C$ increases, $I_C^2$ into $R_C$ will also increase.

So more heating effect, more temperature and more $I_C$ zero. So this is known as, so a time will come when $I_C$ will be so large that it will destroy the BJT itself. So, there will, this phenomenon is known as thermal runaway, right? So this phenomenon is known as basically known as, known as thermal runaway. So this is thermal runaway or self-heating issues, right? And therefore this thermal runaway gives you a large amount. So in the devices designed for high-power application, the transistor is mounted on an efficient heat sink for the reason I just now discuss with you, so that the thermal energy can be transferred away from the junction right?

So as long as you are able to shift the thermal energy away from the base collector or base emitter junction, we would expect to see no rise in the minority current carriers because of

temperature because you are able to shift the important or the major part of our power dissipation or temperature rise away from that. So if you have a heatsink which is working as a very good heatsink, that will help you to remove the power very fast away from the collector, emitter base junction.

(Refer Slide Time: 12:56)



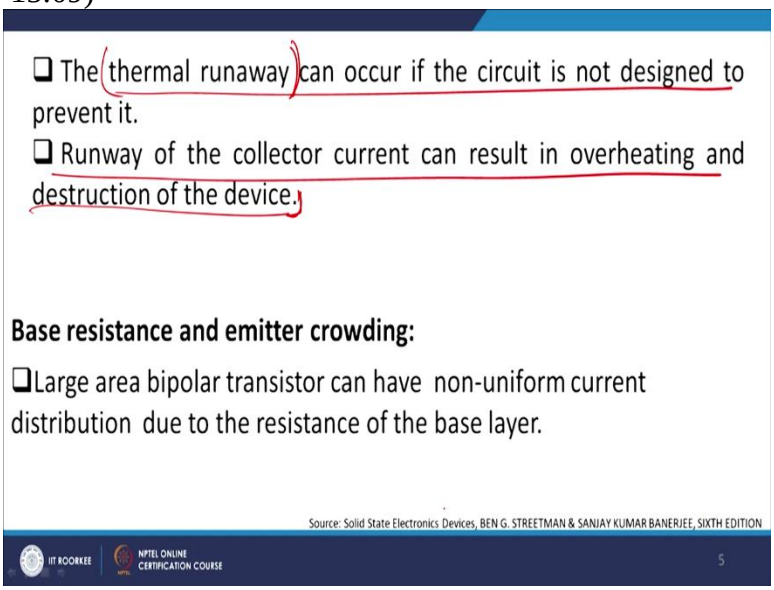

Ok, This is what I was saying that if you allow therefore, if the temperature of the device is allowed to increase due to power dissipation of thermal environment, the transistor parameter changes. So your α and β are a function of therefore temperature. Right? And current gain increases. So as we discussed we will not go in the details of it but the most important parameters, the carrier lifetime and diffusion coefficients, also we have seen that as the temperature increases the mobility actually falls down. Right?

And because this is very simple that as the temperature increases, the lattice itself starts to move across its mean position right because of high thermal energy. As a result, electron travelling across the lattice gets scattered by collision with these lattice atoms. As a result, the mobility falls down. So higher the temperature, the mobility starts to fall down and this results in a lower mobility and therefore when you talk of drift, your drift will be equals to, velocity of the carrier will be μ into E. So as a result, when μ falls down, velocity falls down and therefore the current is reduced drastically. Right?

The diffusion coefficient of holes and electrons also decrease as the temperature increases, thereby causing a drop in β due to increasing transit time. I think, I will make it clearer to you what do I mean to say by that. It means that as you as your temperature rises, right, the diffusion coefficients actually starts to fall down. So when that happens, you have because it primarily means that see D/μ is constant, Einstein's relation. So when D starts to drop down, μ also has to go down to make this ratio constant.

So when mobility drops down, it increases the transit time, right? And when it increases the time, the β actually starts to reduce. The β is basically the current gain. Why current gain starts to reduce? Because now, since the transit time is lowered, then within the same amount of time, lesser amount will be (charged) charge will be transferring from emitter to base. So, your overall collector current will reduce and therefore the gain will reduce, right? And therefore, and that is the reason why you have reduction due to increase in transit time as the temperature increases right? And this is one of the problem areas which people face in this thing.

(Refer Slide Time: 15:09)




Now, thermal runaway is a standard problem area and if circuit is not designed properly, you will have a thermal runaway which will make the device non-functional and will burn up within a very short period of time. As I discussed with you, runaway of the collector current can be due to overheating and the destruction of the device right. And this is what leads to fatal destruction. Let me come to the third second order effect and that is base resistance, emitter crowding. Now

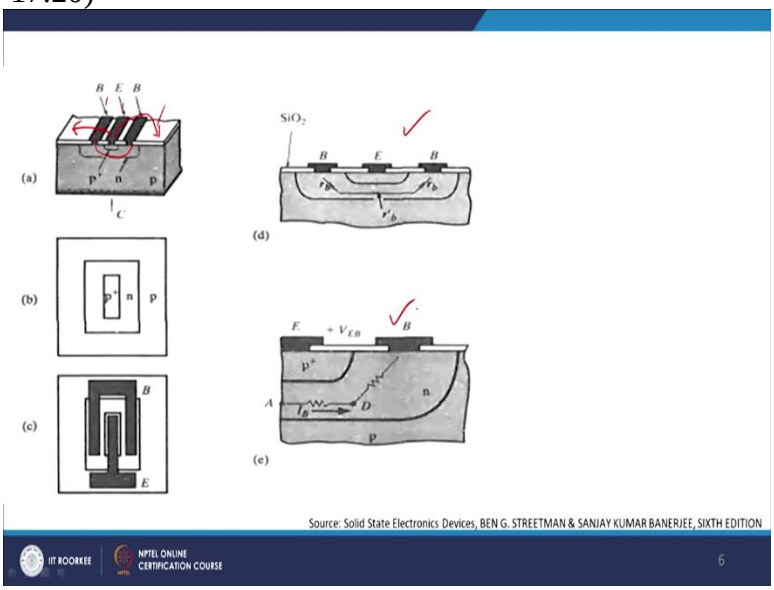
what happens is that we are assuming that the cross-sectional area of the NPN transistor, emitter base and base collector junctions are relatively small, right?

And therefore the flux of electric field is almost constant across that emitter base or base collector junction right? That is true even if you have very small junction areas there. But if the junction areas are relatively large, typically very large, then you do have a non-uniform current density which means that the flux lines are not uniform across the interface. Right? We will not go into details of why is it like that but that is the reality. You don't get it. And therefore the collector current is not equal across the interface through the base layer.

This is basically known as base resistance or base resistance increases in that case, right? So these are the three major second order effects or second order phenomena which people have seen across the board and its removal is quite important. Early voltage or base width modulation effect can be removed by using the doping concentration on the base side, much much higher, relatively higher as compared to collector, this is one thing.

The second thing is, how do you actually go for thermal runaway removal, well there are standard stabilization techniques by which we can actually have circuit level manipulations to reduce the thermal runaway, or almost remove the thermal runaway. And for the base splitting and emitted following, we require to make the emitting base junction and base collector junction area, cross-sectional area as small as possible.

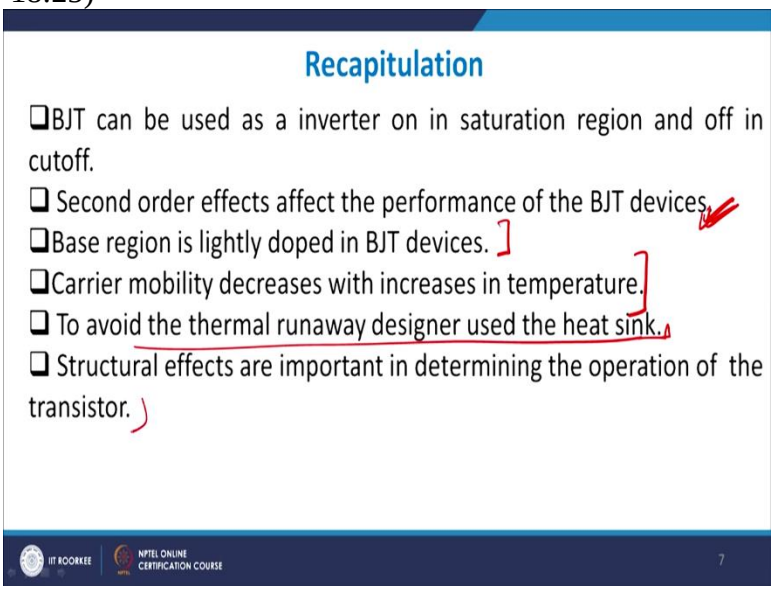
(Refer Slide Time: 17:20)




Now let me just finish of this bipolar transistor lecture by giving you a effective cross-section of a bipolar device. This is your emitter base junction. If you look very carefully, it is base, emitter and this is also your, so this is your collector, this is your base. So you have got two bases connected back-to-back. So this one base and therefore emitter base collector and emitter base collector, so you have got 2 transistors in series, sort of. And their layout looks something like this and this right? So we will not go into further detail than this, it is available in a standard textbooks for understanding purposes and it gives you a detailed analysis about the BJT as such. Right?

We therefore recapitulate the bipolar technology in a detailed manner which we have already seen. We have seen therefore that the bipolar technology can be used as an inverter to make it saturation, to go to saturation and cut-off.

(Refer Slide Time: 18:25)




Second order affects the performance or the behaviour of the device right and it degrades the behaviour of the device. It degrades it, right? Base region is always lightly doped. Carrier mobility decreases with increase in temperature, we have seen that. We, as we discussed earlier that in order to remove thermal runaway, designer has to use heat sink and proper heat sink design is an important one. And structural effects are important in determining the operation of the device. The structural effects primarily mean, how what is the cross-sectional area of the base emitter and base collector junction, right? And these are important.

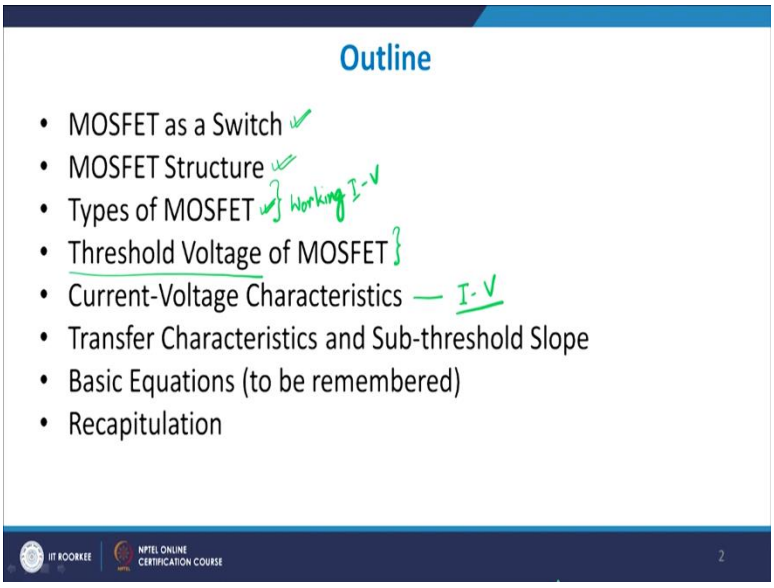
So this takes care of our understanding of typically the first part of our, of the lecture series. And this was still bipolar transistor, we will and it takes care of basic understanding of bipolar, the parameters of bipolar transistor, various configurations of bipolar transistor, how do you calculate current and therefore what are the second order effects prevalent in a bipolar transistor and that has been discussed in these set of modules. In the next module or the next lecture, we will be going through or going ahead with a CMOS technology or MOSFET technology. Right? With these words let me thank you for your patient hearing. Thank you.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−14**
**MOS Transistor Basics-I**

Welcome to the NPTEL online course on Microelectronics: Devices to Circuits. In our previous interactions and modules, we have seen the concepts of bipolar technology or BJT, bipolar junction transistor and we also saw the various characteristics, the structure of the transistor, the types of transistor, the modes of operation of the transistor and then we saw some of its circuits implication in terms of common emitter, common base and common collector. We also saw the second order effects of bipolar transistor which were prevalent at certain conditions of bias or structure and that took care of our basic understanding of bipolar transistors right.

In our subsequent slides or subsequent interactions, we will be looking into what is known as a CMOS technology, or a complementary metal oxide semiconductor technology. So to understand that or CMOS technology, we need to first of all understand the basic concept of a transistor or a metal oxide semiconductor field effect transistor which is basically also acronym-ed as MOSFET, M-O-S-F-E-T. Right? So let us start with this lecture on MOS transistor basics one. So the lecture topic is basically MOS transistor and we start with basics one right?

(Refer Slide Time: 1:59)



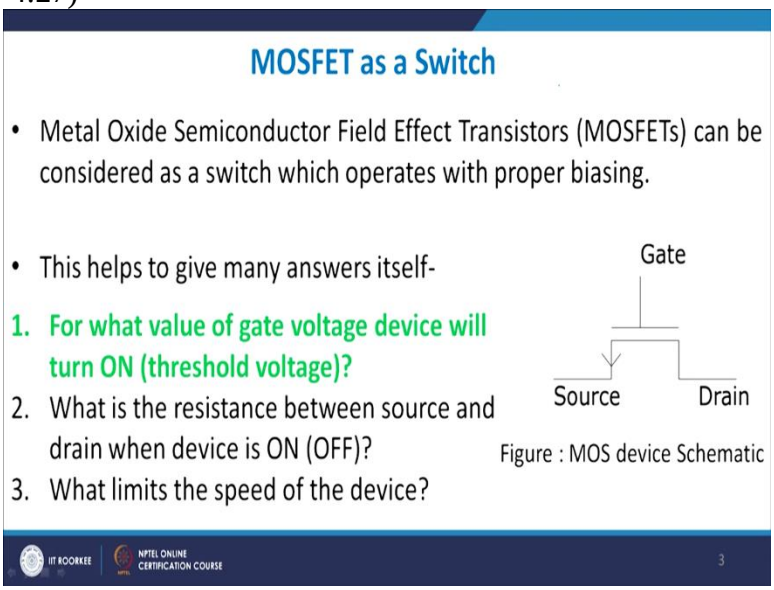
The outline of the talk is that we will be looking at MOSFET as a switch first of all, right? We will be looking at a switch, MOSFET as a switch, right? And then we will be also looking at the MOSFET structure, right? So the idea here is that we need to first of all look into the fact that why we migrated from a bipolar technology which is basically a technology in which both electrons and holes are majority current carriers and results in overall current, to a technology which is basically a unipolar or only one of the electrons, either electron or holes are the majority current carriers and they are responsible for the current. Right?

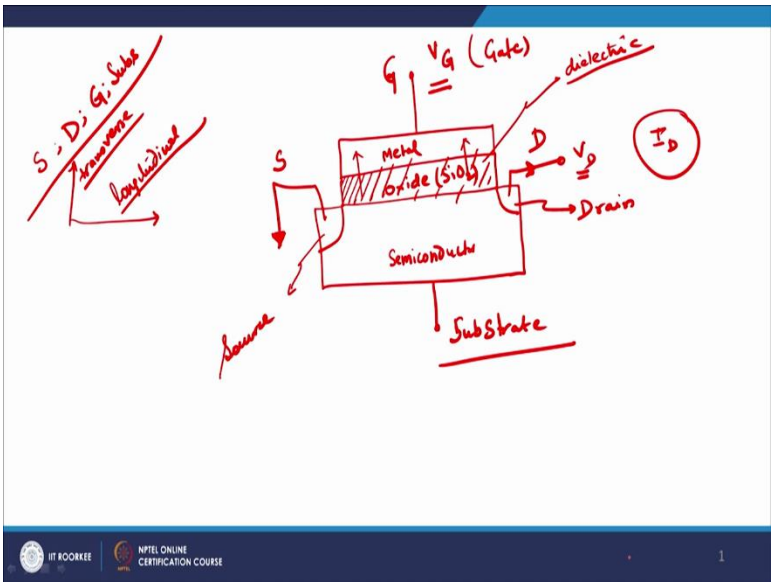
And after that, we will be looking at types of MOSFET. So once you have understood the structure, we will be (understood) understanding the types of MOSFETs and their working behaviour. So how they are working actually in terms of real IV characteristics, right? Then we will be looking into one important property of MOSFET known as the threshold voltage of MOSFET right? A very important property from the point of view of both, analog as well as digital electronics and we will see later on that this voltage is quite critical to the operation of a MOS device. We will be also looking into the current versus voltage.

So current versus voltage characteristics, which means that given a set of biases, can I predict the output current of a MOSFET? Right? Now these set of biases can be a variable quantity, can be a fixed quantity. So I can have a DC analysis in which possibly I fix the bias voltages at various points of the MOSFET and then see how much current the MOSFET is carrying, right? Or I can also have something like this that I vary the voltage and see how much variation in the current is visible to me.

We will be also looking at transfer characteristics and sub threshold slope. This is quite important from the point of view of switching. So, whenever we want to use it for high switching purposes or the purpose of high-frequency applications we need to understand what is the meaning of sub threshold slope and what is transfer characteristics. So after that we will be looking into the basic equations which will be helpful for finding out the current and then we will recapitulate the MOS device as such right. So this is the basic outline of the talk.

(Refer Slide Time: 4:27)




Let me come to you as the basic idea here. So before I even go to the MOSFET as a switch, let me give you a basic diagram of a MOS device. You must be aware from your other sources but just to give a continuity in whole idea, this is how a MOSFET primarily looks like. Right? Since I am making it as a free hand drawing, it is exactly not like this. So I have got metal here. This is a metal right? I have an oxide here which is primarily silicon dioxide in most of the cases, silicon dioxide and I have got a semiconductor here, right? I have a semiconductor here.

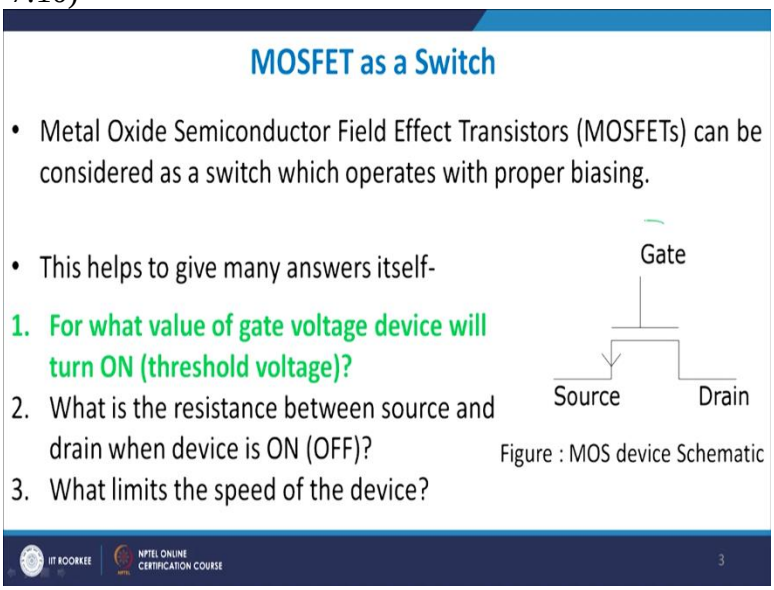
And this is basically my, let us suppose, source, this is termed as a source. Right? And we have a drain. So this is your drain. So I have a source here, I have a source and a drain here and I have a

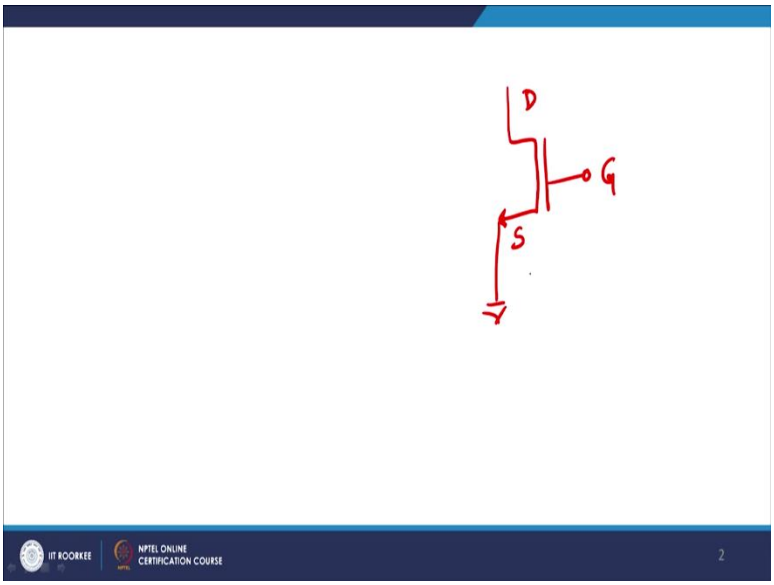
semiconductor here and this side is referred to as VG or also referred to as gate. So I have got, so basically if you look very carefully, it is basically a three terminal device at this stage. One is source, another is drain, another is gate, right? You also have fourth terminal which if possible we will come later on and this is known as substrate terminal.

So typically a MOS device has got four terminal device- source, drain, gate and substrate, right? It is a four terminal device. Typically, for all practical purposes, my source is grounded. That is the reason I have shown source to be grounded here which means that the potential on the source side is approximately equals to zero or exactly equals to zero. We give a bias on the gate side and drain side and then we try to find out the current through the drain side right. So $I_D$ is the current through the drain side which means that the current is flowing through the drain side. Right?

Now you see, since oxide is there, this is oxide layer right? So oxide is basically a dielectric. You remember, it is basically a dielectric and as a result, you won't expect to see any current flowing in the transverse direction. So we refer this as longitudinal direction, longitudinal right? And we refer to this as transverse direction. This is a typical parlance which we follow or you can call it this one vertical and this one horizontal. But in typical device physics, we convert this to be as longitudinal and this to be as transversal.. So, perpendicular to the silicon interface is basically my transversal and this is the longitudinal direction which you see.

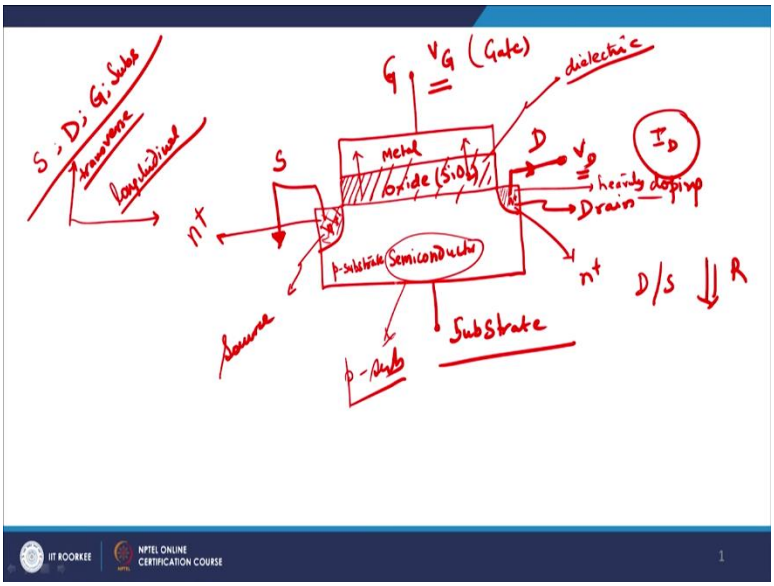
(Refer Slide Time: 7:10)






Now if you come back to, come back to its, to the operation, the way it is being transferred or written is something like this. So the methodology which we write is gate, source and drain right? So the method, so what people follow is that, if you have a transistor, then we refer to the transistor as this. This means that, this is your gate. Right? Gate. And we refer to this as source and drain. So the arrowhead right refers to the direction of the movement of holes, not electrons. Gate is there and drain is there. Now you see, we will come back to the previous slide here and see what works, how a device is turned on.

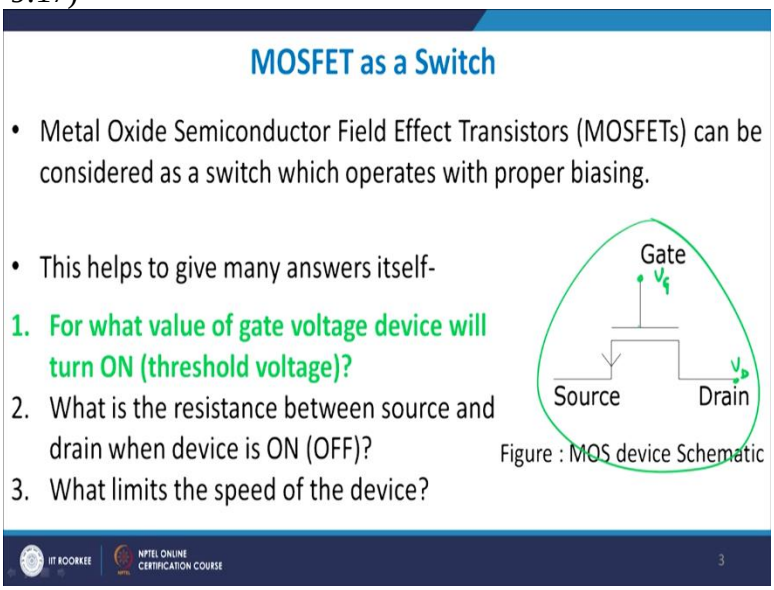
(Refer Slide Time: 7:54)




Generally, this is basically made $N^+$ and $N^+$, which means that this is practically very heavily doped, it is heavily doped right? Both sides, source and drain are very heavily doped. Heavily doped, therefore means that the number of electrons here inside the drain region and the source region are very very large as compared to the nearby regions, right? So they are very heavily doped. So therefore, since they are very heavily doped from a basic device physics, we will understand that the resistance offered by the drain and source is very very low.

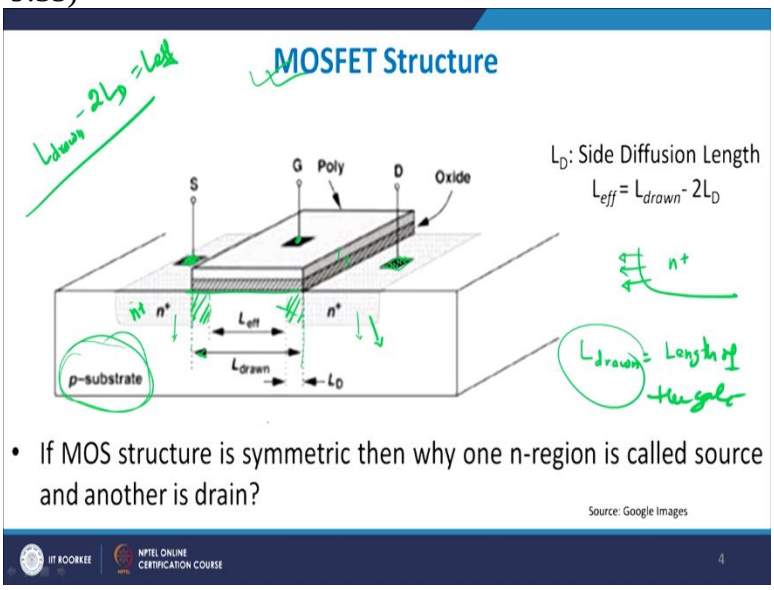
So there is no resistance or there is no potential drop inside source and drain because now you have very large number of electrons available. So the resistance is fallen down and as a result, there is no potential drop within the source or drain. But surely, you can apply a bias from an external world by applying a $V_D$ voltage and we will see how it works out, right? Let us see. So if I have this one as $N^+$ right, I have this as $N^+$. Right? The substrate which is basically my semiconductor, I use it as P type substrate. Right? I use a P type substrate, which means that typically this semiconductor is basically P type right? It is a P type substrate and I have got a P type substrate at this, at this point.

(Refer Slide Time: 9:17)




So you see this drain, you will apply a potential here, here you apply a potential $V_D$, here you apply a signal or a potential $V_G$ and generally source is grounded right. So 1, 2, 3 that will, we are not looking at substrate nowadays at this point. But we will see how it works out.

(Refer Slide Time: 9:35)




Now if you look from the basic concept, we will, if you look at the MOSFET structure now, it looks exactly like what you see in front of you. Right? So I have a P type substrate. This is a substrate. Right? This is my source. This is, this black color which you see is the metal contact. So a metal contact through which you are contacting the external world. This is my

semiconductor, heavily doped. Semiconductor, heavily doped. And then again, this black contact is my metal contact which is also metal contact here. This is my oxide layer, this is the oxide layer which you see and then the white one is basically my metal layer, right? Also referred to as poly layer and this is basically my gate junction.

So I have my gate junction, source junction, drain junction and I have got a P type substrate at this particular point. Right? Is it okay? Let us see, how we define channel length before we move forward any further. Now generally what happens is that whenever you draw, whenever you put $N^+$, for example, you have put $N^+$ region here as well as $N^+$ region here, right? Is it okay? Now obviously, because of temperature variations, because of large density potential gradient, because of large gradient of the charge carriers, some of these $N^+$ regions will shift to the left.

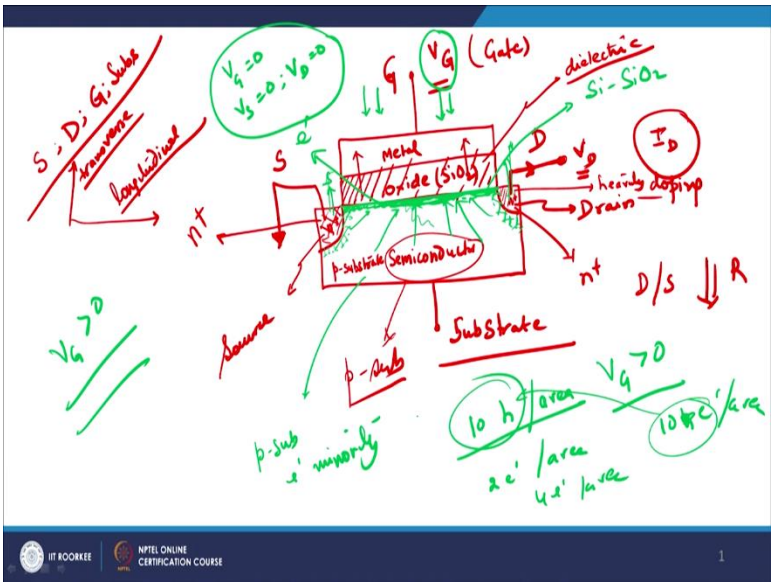
So you wanted to, you diffused it up to this much point or you doped to this much point. After some amount of time, this will actually start shifting on this direction because of sudden diffusion and as a result, you initially wanted it to be till here but you ended up till here, right? Right? Is it okay? You ended up here. Similarly, you wanted it till here, till source region, till the edge of the gate, but you ended up somewhere here. So we define 2 quantities here.

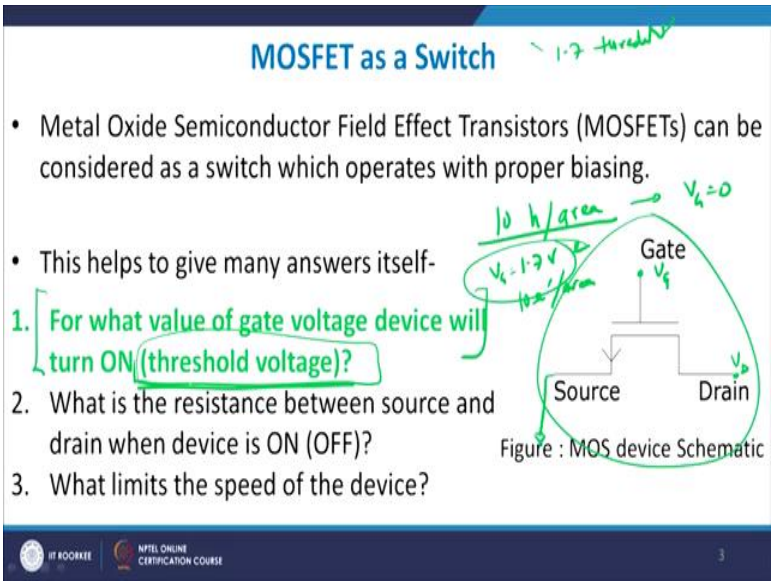
Physical drawn length which is L drawn. L drawn is a physical drawn length. Physical drawn length is nothing but the length of the gate. So if a gate length has 20 nanometer then this basically, this one is basically a twenty nanometer from this end to this end. But what has happened in the meantime is because of doping been laterally shifted, because of temperature variations and all these things, the effective channel length has actually been reduced from L effective to L drawn from L drawn which is basically my drawn length, minus two times $L_D$. And why two times? Because one on this side and one on this side. Right?

So your effective length is always smaller than your drawn length or a physical length. Fine? And this has to do with the doping or the diffusion being laterally moving across the frame. So you had a diffusion here or you had a large number of charge carriers here. It started to diffuse laterally in this direction because there is a gradient of charge. It stops at a particular point. When the excess amount of, assuming that these two are perfectly symmetrical under all circumstances, I assume that my L effective will be equals to L drawn minus $2L_D$. Right? L effective will be equals to L drawn minus 2L effective right?

And is this what you get what you get from it. So from this basic idea, let me see how can I work it as a switch, does it work as a switch.

(Refer Slide Time: 13:00)




Now you see, in our previous discussion when we were taking up this basic concept of MOS device, if you look very carefully, if you do not apply any gate voltage right so I have P type substrate here. Right? So P type substrate will have electrons as minority current carriers, we have already discussed this point. Howsoever small, they will have some amount of electron

available here. Also, this source and this source will have large amount of free electrons because it is a source and it is a drain. As a result, this forms a PN junction diode. So under no bias condition, there will be a depletion region here, assuming that my $V_G$ equals to 0, $V_S$ is of course equals to 0 and $V_D$ equals to 0 and $V_{substrate}$ is of course equals 0.

For all these three conditions of these things, I would sufficiently assume that the depletion region is exactly equal on the left and right side which means that the doping concentration on the source and drain are exactly equal. If the doping concentrations are same, you will see the depletion region to be exactly equal. Fine? So we have what is the current situation therefore? That this is basically P type substrate, this is P type substrate. I have got a depletion region here which is basically devoid of any free charge carriers, similarly I have a depletion region here devoid of free charge carriers and I have got a high $N^+$ region here and a high doping $N^+$ region here known as drain.

Now what I do is, this is quite interesting and this is where we can actually do quite amount of change. What we try to do is, we try to make our $V_G$ now greater than 0 right? And that is quite important. As you make $V_G$ greater than 0, there will be electric field in this direction, of course because of course the as reason suggest and as a result, large number of electrons from here, from source right, from drain and from substrate, minority current carriers, will all rush towards the interface, will all rush to the interface.

So please understand, that whenever my gate voltage therefore rises up, at a certain gate voltage, large number of charge carriers come near the silicon-silicon dioxide. So this is, this interface which you see in front of you is basically silicon -silicon dioxide interface. So what will happen is there will be a large number of electrons which are already stationed within the source and drain region and because of this external voltage applied on the gate side, a positive voltage, $V_G$ greater than 0, for $V_G$ greater than 0 right, they will start populating the electrons near the silicon-silicon dioxide interface.

So you have a large density of electrons here. Right? Large density of electrons. So this will be typically large density of electrons. Now these electrons are coming from where? Coming from source, and drain and the substrate right? So initially those P type, now let us suppose I am

throwing as, I am increasing the gate voltage and I am making it more and more electrons, more and more are coming here. Right? If the number of electrons per unit volume you can take or per unit area near the silicon-silicon dioxide interface, is exactly equal to the number of holes per unit area initially kept right, so suppose initially we had ten number of holes per unit area, now what we will do?

We will now increase the gate voltage and they will make you say, say there are two electrons per unit area. Right? Increase further. I get four electrons per unit area. Right? I go on increasing further till I reached to ten electrons per unit area. Agreed? Once this happens, it was happening earlier also, these ten electrons will have a chance of recombining with these ten holes and as a result, there will be no holes available with me. Now if we increase the gate voltage slightly ahead, larger than this, you will have excess of electrons there, rather than holes because whatever hole was there has been actually accommodated by virtue of recombination and as a result, that area is devoid of any free charge carriers.

Free charge carriers, they are not devoid of charge carriers but they are devoid of free charge carriers. Now what we do is that we still further increase the value of $V_G$. Then you will have now the population of electron getting more and more because recombination has ensured that the hole strength is almost zero. So, if there were ten holes available near silicon-silicon dioxide, electrons, ten electrons goes there and recombine with holes and the holes finish, right? Now if you go and increase the value of $V_G$, you will be having the excess of electrons, right?

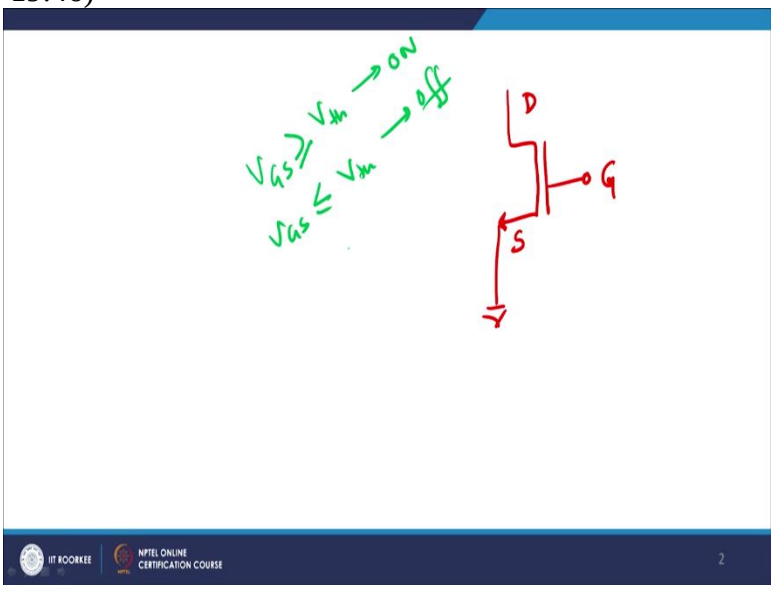
So we define a new term known as threshold voltage where we define, this, if you look at which is defined as threshold voltage, right? How do we define threshold voltage? That gate voltage at which the number of electrons per unit area or volume is exactly equals to as many number of holes were there per unit volume earlier, we define that voltage, gate voltage to be as threshold voltage. So there were let us suppose ten holes per unit area near silicon-silicon dioxide interface right, when your $V_G$ was equals to zero, now you apply $V_G$ equals to 1.7 volts and the number of electrons are ten electron per unit area near silicon-silicon dioxide, then we define $V_G$ equals to 1.7 as the threshold voltage of the device.

So how do you define threshold voltage? Threshold voltage of the device is that voltage at which it becomes as much N type or P type as it was earlier P type or N type respectively. Fine? You

got the point. And that is very very important. And therefore quite critically, the threshold voltage is also referred to as the On voltage of the MOS device. So this is the voltage when the device gets on, switched on. Right? Because you know, a large number of free charge carriers which can take part in conduction.

So On means it does not necessarily mean a current but it surely means that you will have large number of charge carriers near silicon-silicon dioxide. Now you have to apply a drain voltage in order to sweep the current and form a, sweep the charge carriers form the current. Right? So that is what I am saying? In green what is written here. So what value of gate voltage will it turn on? So the value is basically my threshold voltage, right? At threshold voltage, it will turn on the device. So if I want to switch my MOS device from on to off stage and vice versa, I just need to put the gate voltage above threshold voltage and I am, so what I do is I do something like this.

(Refer Slide Time: 19:46)

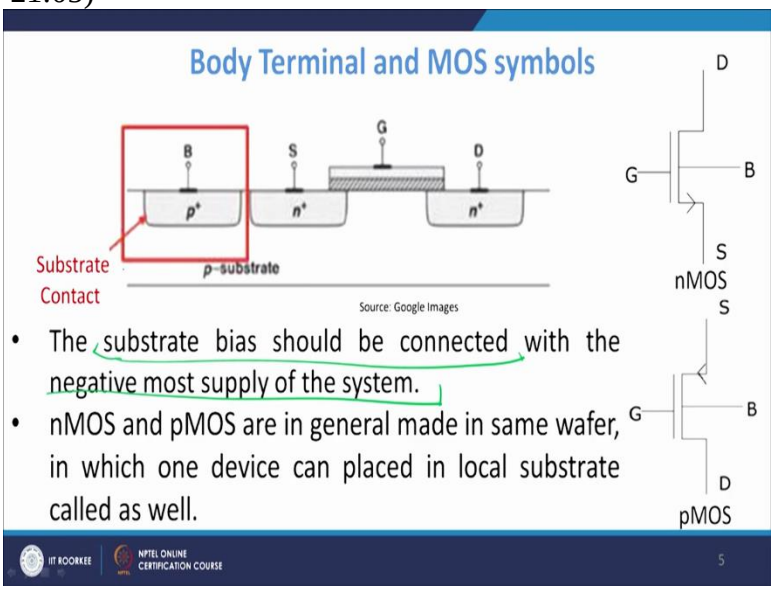



That I apply V gate to source voltage as greater than equals to threshold voltage of the device and I switch it on. If I apply $V_{GS}$ less than, equal to threshold voltage, I define it to be as a Off state. Fine? So threshold voltage is quite an important term in analog as well as in digital because it gives me an idea about how it works out. Na. Now so what is the reason between source and drain when the device is on? Well, the resistance will be very small because now you have large number of free charge carriers between source and drain.

And therefore, the resistance offered by the channel will be very small whereas when the device is in the off state, the resistance offered by the channel will be relatively very high. Right? And that is also one of the methodology by which you will know whether the device is switched on or off. So if the resistance of the channel, offered by the channel is very very high, pretty high, then you can assume it to be as a as a your if it is very high, which means that there is the drain and source are removed from each other, it is off state.

If it is low, it is in on state, right. So it is now understood and we have understood the basic feature of two terminal device. Right?

(Refer Slide Time: 21:05)

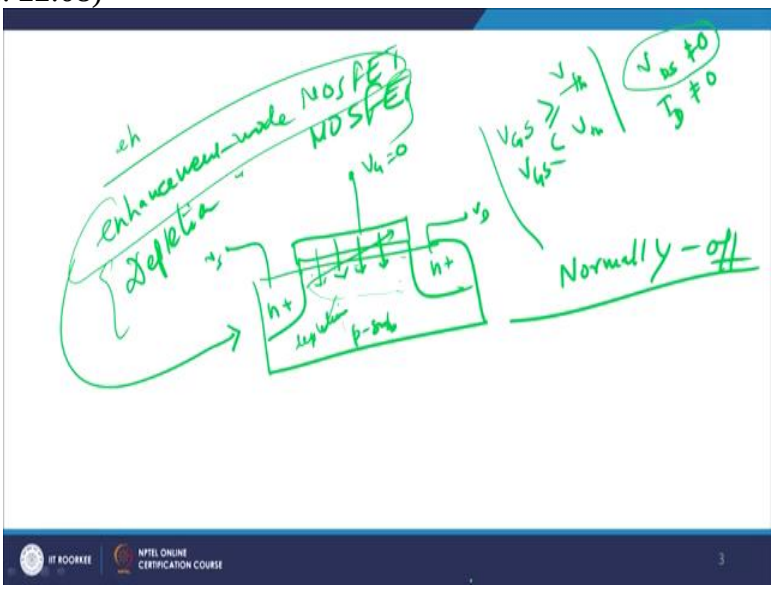



Typically the rule of thumb is, that the substrate should be always connected to the most negative terminal of the supply. So negative most terminal of the system should be connected to the substrate bias. And we will see later on why, but you can find it out yourself from many other books because once you do that, when you keep it to the negative most terminal of the battery, you don't therefore manipulate the threshold voltage once it is already been fixed by the external biases. Right? And if therefore since nMOS and pMOS are made from the same wafer, one device can be placed in local substrate called a well.

So I have a well structure, this is the well which you see in front of you. In the well, we have an $N^+$ region. So this is my nMOS and your substrate contact here which is there with me and this

is the fourth contact which is generally used for, for the forthcoming for the purpose of contacts here, right? Let me therefore come to the basic issue of of…, so let me understand, let me explain to you therefore what are the various, what are the various issues involved in MOSFET.

(Refer Slide Time: 22:08)




The first issue which we have seen is that if my gate voltage is greater than equals to threshold, the device is on and if it is less than equals to threshold, it is off, right? And it is true also that the devices are on and off but for the current to flow, you require to have $V_{DS}$. So $V_{DS}$ if it should be non-equals to 0, then only $I_D$ will be equals to non-equals to 0. Right? If $V_{DS}$ is 0, then you will not be able to have any current flow between, though the device will be in on because $V_G$ is greater than $V_{TH}$ but since $V_{DS}$ is not equals to 0 or $V_{DS}$ is equal to 0, let us suppose that $I_D$ equals to 0, primarily meaning that the device is not in the on state.

Although, it is not at least carrying the current from point A to point B. Right? So this is what we get. Let me come to the type of MOSFETs here now. And there are 2 types of MOSFETs which is with us. And one is known as an enhancement, sorry enhancement mode MOSFET. Right? And we have got a depletion mode MOSFET. Right? We will first discuss enhancement mode and then we will come to depletion mode. Fine? Let us look at enhancement mode MOSFET.

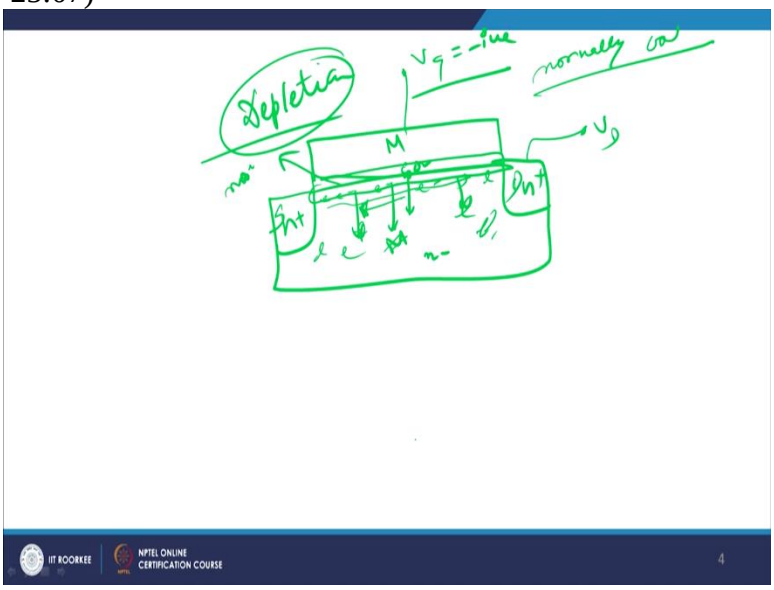
In the enhancement mode MOSFET, as the name suggests, you have to give an external potential on the gate side to enhance the number of charge carriers near the silicon-silicon dioxide

interface in order to form the channel, which means that if I have an $N^+$ $N^+$ region here and I have got a hole and, sorry, I have got a dielectric and a, sorry, a gate here and drain here right, drain here and this is source right and then I join these two together for understanding purposes, then if we do not apply any gate bias, $V_G$ equals to 0, this is P type substrate, nothing will happen and the reason is since there is no electric field in this direction there is no electric field, right?

There is no electric field. Then no charge carriers will be moving towards silicon-silicon dioxide interface. Right? So there will be no charge carriers. Right? And therefore this will be all this will be all depletion region. All will be depletion region. Right? All will be depletion region. Fine? Now what happens, we go on increasing $V_G$ and we are pulling the electrons towards the interface and thereby enhancing the charge carriers because it causes threshold voltage, you say that the device is fully on.

So what is an enhancement mode MOSFET? Enhancement mode MOSFET is a device which enhances the charge carriers near silicon-silicon dioxide thereby making it on, right and therefore it is also referred to as a normally off device. It is a normally off device means, normally off means, if we did not apply any bias, it will be off state, right because there is no channel charge formation taking place.

(Refer Slide Time: 25:07)




Let me come to therefore the depletion mode MOSFET. In the depletion mode MOSFET, I assume that these are all $N^+$, exactly the same as the previous case, right? And everything else

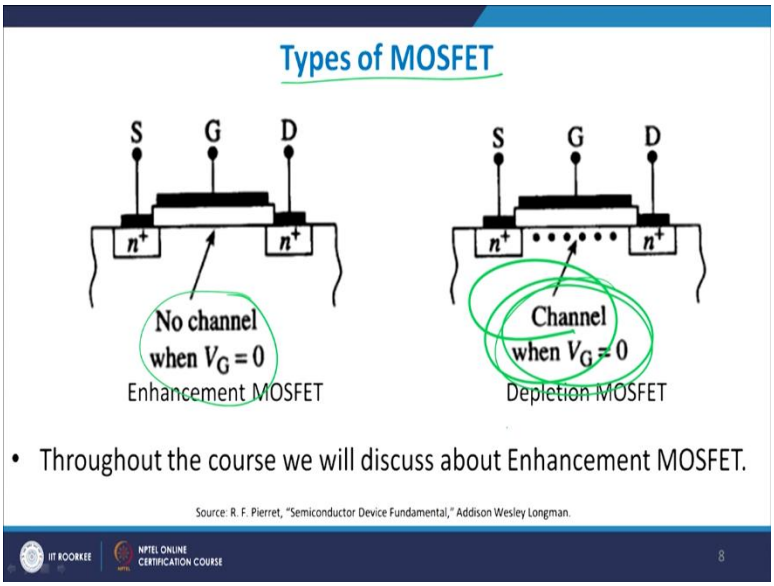
remains the same. I have an oxide layer, I have a, sorry, a metal layer, I have a silicon dioxide and a metal here. Right? Now, so we are discussing depletion mode design. What happens in this case is that, in this case let us suppose this $N^+$ $N^+$ and rather than substrate to be P type, let me make it N type. Then all will be electrons here.

All will be electrons because it is a majority current carriers. Now if we apply a drain bias $V_D$, all the electrons will be moving in this direction and therefore there will be a current $I_D$ flowing in this direction. Right? This is known as depletion mode. Why? And the reason is, this is also known as normally on. Normally on, why? Because even when you do not apply any gate bias, $V_G$, there will be large number of charge carriers between source and drain. Right? Effectively, the number of carriers will be very large.

And if we do not apply any gate bias, then all these charge carriers are already near Silicon-silicon dioxide interface and therefore they are helping you to form a current between source and drain. So how will you stop it? You have to give a gate voltage sources, such that, so it is an N type, that it should be negative in dimensions or negative in nature. If it is negative, all the electrons will be pushed backwards and therefore this whole thing will be left to the main issue right. So that will be there. Which means that if I apply a gate voltage which is negative, it will push all the electrons downwards and all the holes will come up and it will be going into off state.

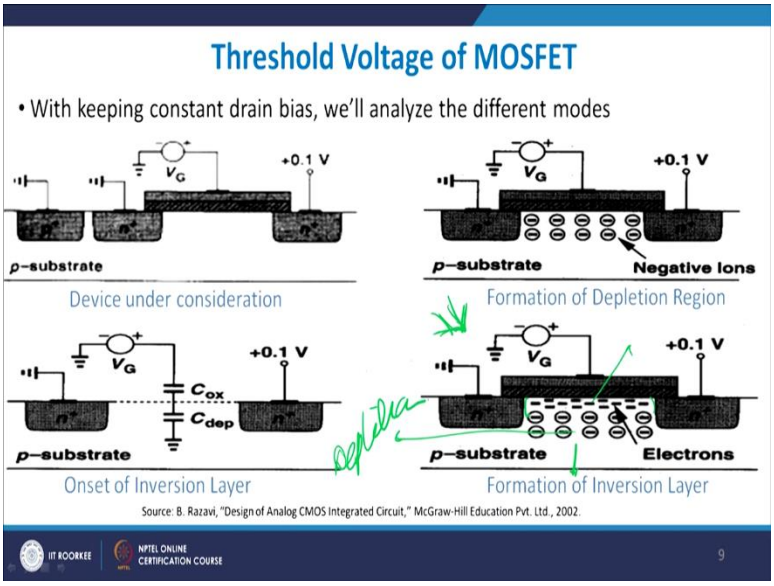
So I have got normally on and normally off device. Normally on device is depletion mode, normally off device is basically your enhancement mode structure, right? So this we have learned and this is what you see.

(Refer Slide Time: 27:02)




No channel when $V_G$ equals to 0 for enhancement mode MOSFET. For depletion mode MOSFET, when $V_G$ equals to 0, you still have some channel being formed here right? And that is quite an interesting phenomena as far as types of MOSFETs are concerned.

(Refer Slide Time: 27:14)



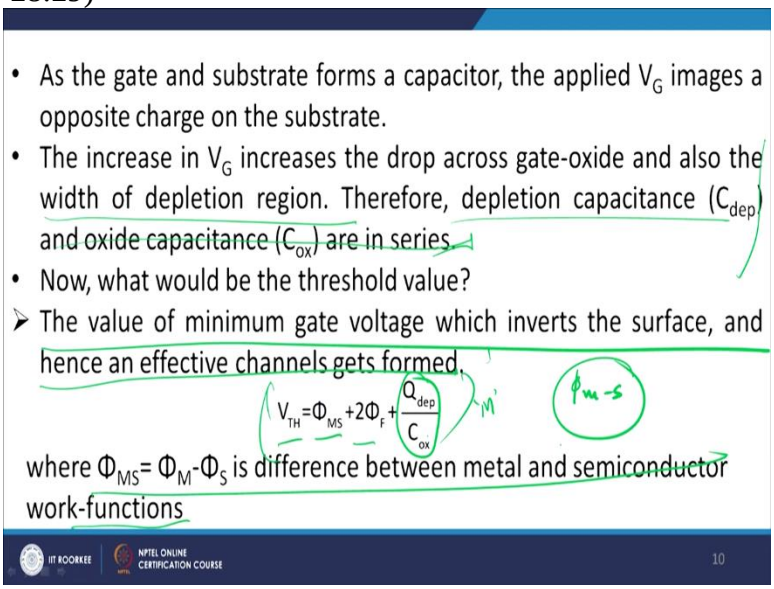

We have also seen therefore how do you define threshold voltage of the MOSFET and we see that threshold voltage is defined as that gate to source voltage at which the substrate becomes as

much N type as it was initially P type or becomes as much P type as it was initially N type, either of the 2 cases right. And you are able to solve the problem of MOSFET relatively easily. As you can see here from the fourth diagram here, fourth diagram right, the fourth diagram, as my gate voltage becomes more and more positive, it pulls electrons near.

So these are the electrons. And all my negative donor concentrations have been shifted downwards. Why? Because these are atoms right? And atoms have typically large, very small mobility and large scattering. Right? And as a result, these atoms do not get enough time to move towards silicon-silicon dioxide interface. Before even these starts to move, the electrons are very fast and they move towards the silicon-silicon dioxide interface. So that is the reason.

So you have an inversion layer here and you have a depletion layer here. Right? And that is the reason why you get a normally on device. Right?

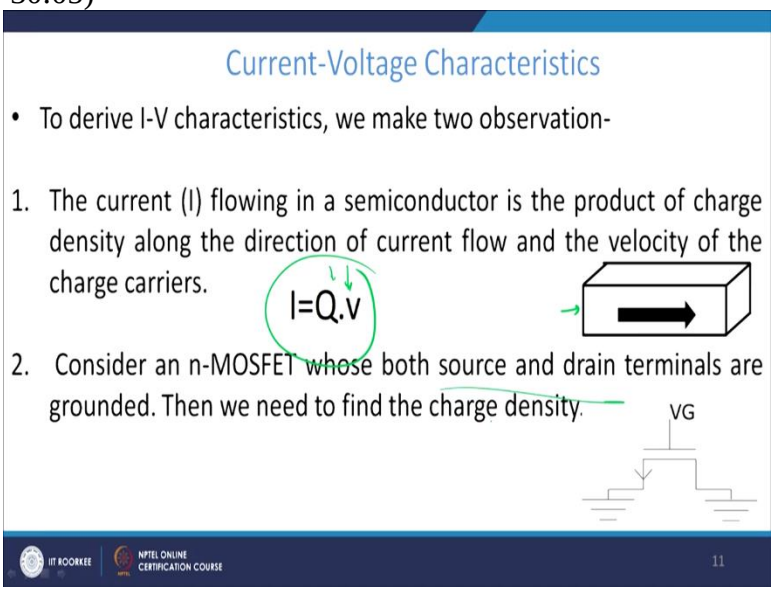
(Refer Slide Time: 28:29)




That is what I was writing here, that as we increase the value of $V_G$, the drop across the gate oxide and also the width of the depletion region increases. Therefore depletion capacitance and oxide capacitance are in series. Right? So we define threshold voltage as therefore $V_{TH}$ equals to $Phi_{MS}$ plus 2 $Phi_F$ by Q depletion by $C_{ox}$, where $Phi_{MS}$ is the metal semiconductor what function, right? $Phi_F$ is the Fermi potential and Q depletion upon C oxide is basically the potential which is basically the charge. Charge by potential is basically my potential which is available to me.

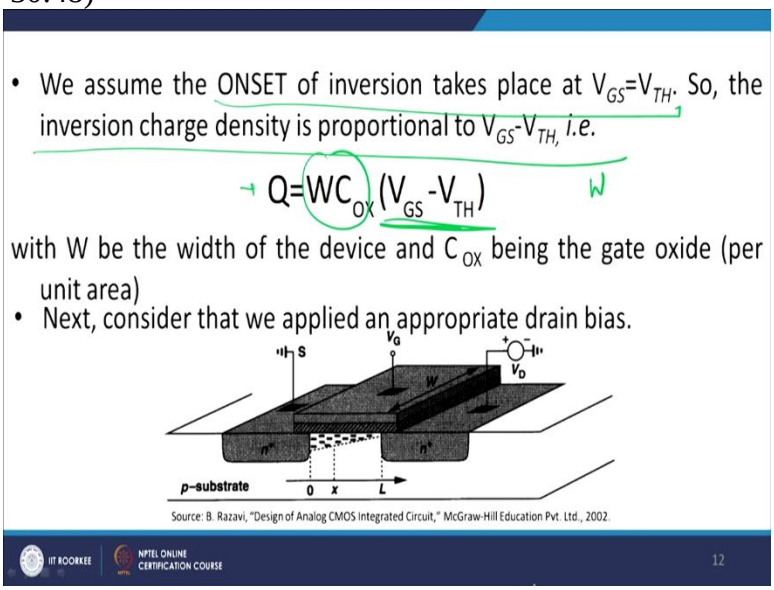
Now, so the minimum value of the gate voltage which inverts the surface and hence an effective channel gets formed, is defined as a threshold voltage. Fine? $Phi_{MS}$ is the work function difference between metal and semiconductor and that is placed here, right? So you will have typical, some value of Phi and M. What is $Phi_M$ because metal and semiconductors do not have the same bandgap and the same intrinsic carrier concentrations, so they are shifted in space domain by a large quantity. Right, so if we can find out, if we can bring somehow or the other, these two together in one domain, we will be able to distinguish or understand $Phi_{MS}$ which is basically the difference of $Phi_M$ and $Phi_S$, where $Phi_M$ is basically the work function of metal and $Phi_S$ is the work function of silicon right?

(Refer Slide Time: 30:05)




So with this, let me come to the current voltage characteristics of MOS device, right? The current voltage characteristics of the MOS device can be looked into the fact that if we apply a voltage here, then current will be equals to charge multiplied by the velocity of the charge carriers, Q into V. Suppose I assume that both my source and drain are grounded, then we require a charge density, which means that if my source and drain are grounded, either of the two are grounded, there will be no chance of movement of charge carriers in the longitudinal direction and you will not be able to explain the charge carrier formation, right? So you require to have some other technique to do it. When you have, so the onset of inversion takes place when $V_{GS}$ equals to $V_{TH}$ right?

(Refer Slide Time: 30:48)

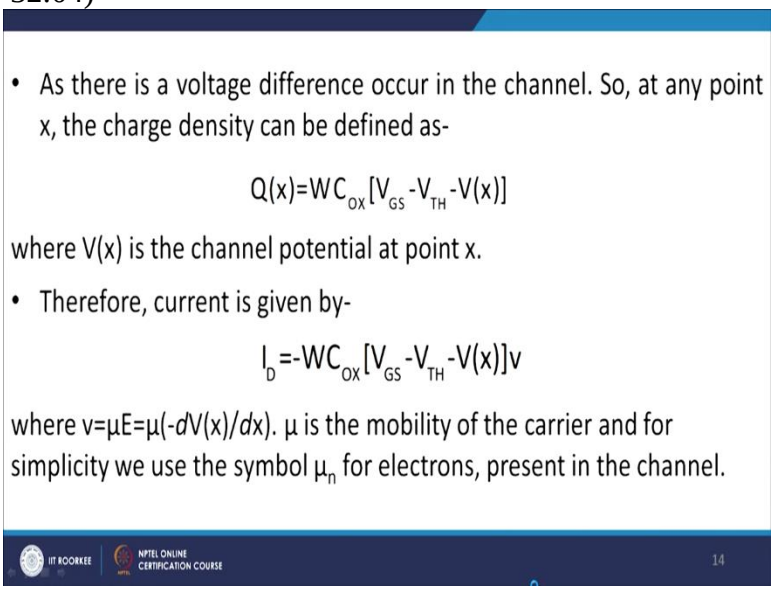


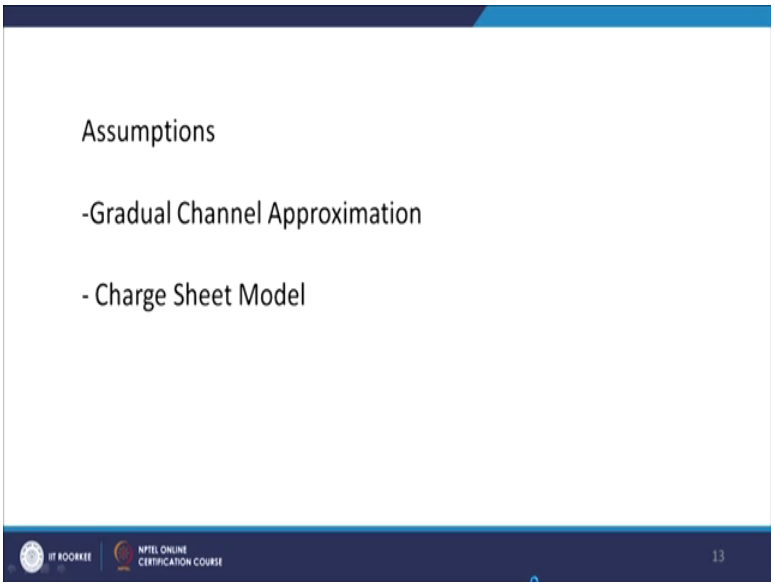

So the inversion charge density is proportional to $V_{GS}$ - $V_{TH}$. Why is it like that? The reason is, say threshold voltage is the gate voltage at which it becomes on right? Now if your $V_{GS}$, which is gate to source voltage is higher than the threshold voltage, the amount by which it is higher right? If that amount is typically very large, that will allow you to basically make the, make it much better. The device to be much better shape, right? So what we do is, at the onset of inversion, $V_{GS}$ equals to $V_{TH}$.

But as the student proceeds further, he sees that the total inversion charge density, right, is directly proportional to $V_{GS}$ - $V_{TH}$. If that is value is larger, you will be able to do anything of this which what is what? Therefore we can say that charge Q is basically W into $C_{ox}$ into $V_{GS}$ minus $V_{TH}$ right? And this is the charge there. W is the width and C oxide is the gate oxide capacitance per unit area and W is the width here. Right? Let us suppose we have applied now a appropriate drain bias.

So I have a drain here. Drain is applied with $V_D$, source is grounded and on the gate side, we again have not applied at this stage anything.

(Refer Slide Time: 32:04)



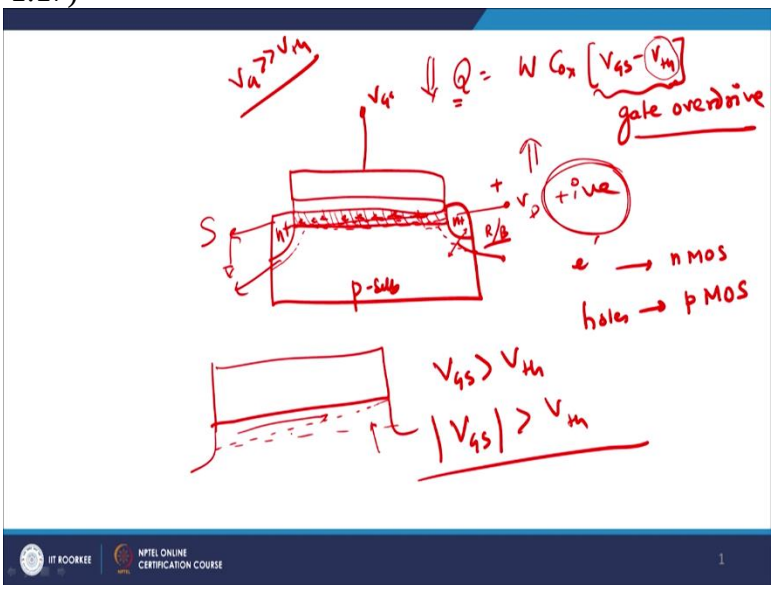




So what are the assumptions we will look? We will look at the gradual channel approximation and charge sheet model and this will be taken care of in the next lecture. Thank you very much.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−15**
**MOS Transistor Basics-II**

Alright, welcome to the NPTEL online certification course on Microelectronics: Devices to Circuits. We will be covering in this module as basic MOS transistor basics part 2. So in part 1, what we had looked into was how a MOS, what is basically a MOS device or a MOSFET and we have seen that in a MOSFET, we have it is basically a four terminal device but typically even if it is in three terminal, the source is generally grounded. We apply a drain bias, we also apply a gate bias and when the gate bias is above a threshold voltage, the channel is formed and the if you apply a drain bias, a current will flow.

(Refer Slide Time: 1:17)




We have also seen that the width, that the total charge which is available is equals to W into $C_{OX}$ into $V_{GS}$ -$V_{TH}$ which is, this is known as gate overdrive, right? So higher is the gate overdrive, larger is the charge accumulated which makes sense also because higher the value of $V_{GS}$, we have to subtract the threshold because that much amount of minimum voltage will be required to form the channel and excess of that will give you large amount of charge carriers.

We have also seen that when you have, when you have a gate voltage larger than the threshold voltage and your drain bias is 0, means you do not apply any drain bias, let us suppose I have $N^+$, $N^+$ and I have P type substrate right? I apply a gate bias here and this gate bias is getting larger than the threshold voltage of the device, so I told to you that there will be a charge carriers which will be coming here, right? So you have a large number of electrons which will be free electrons, which will be available here and there will be a depletion thickness here.

And there will be also a depletion region here. Fine? We have discussed this point earlier also that there will be a depletion region here and here. You will ask me from where these electrons are coming? They are being either supplied by the source side which is generally grounded or by drain which is generally, we apply a positive bias in case of an N channel MOSFET. Now if the charge carriers here are electron, we define this to be as an NMOSFET right? And if it is a holes, which is basically there, then we define that to be a pMOS right? PMOSFET or a pMOS.

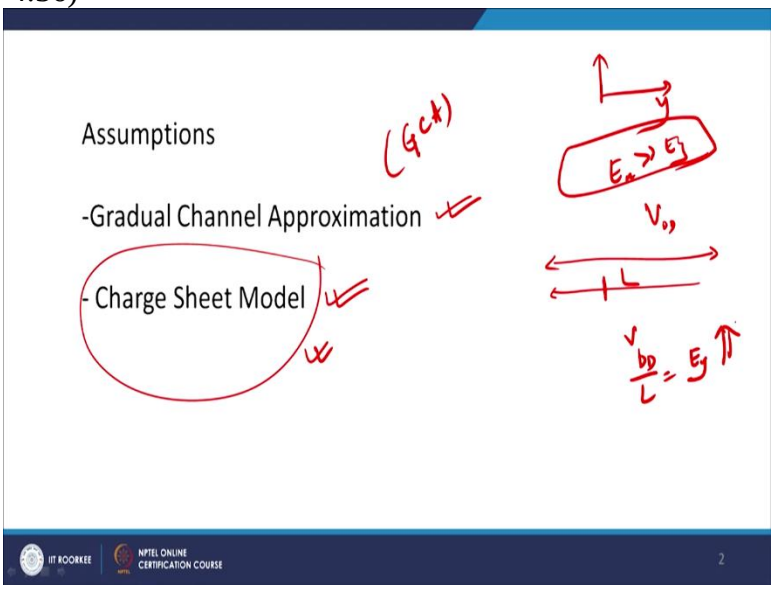
So in case of N MOSFET as I discussed with you, gate voltage should be greater than threshold to achieve it whereas when you are applying a pMOS, gate voltage, mod of that should be greater than equals to threshold voltage of the device. Right? Which means that this is what we get from the basic configuration or idea. So but you see that when you do not apply any gate bias, this is perfectly like this. Right? It is exactly like straight line which means that the number of charge carriers per unit area is almost or the total number of charge carrier is almost constant as we move from source to drain.

It is perfectly symmetrical in nature. But as you go on increasing the positive bias, let us suppose you apply a positive bias and you go on increasing it, then you are actually increasing the depletion thickness because this is, this drain to substrate bias here is basically reverse bias and (there), because you are apply a positive voltage here. So when you go on increasing the positive voltage, the depletion thickness here becomes more and more. So it is becoming more and more. As a result, the thickness of the charge carriers right from source to drain, which was initially like this right, now becomes like this. Right? And this becomes like this.

So your depletion eats away into the channel and you have a reduced, and you have carriers which are just moving towards the source drain region. We will see its implications later on but

that is what we see from here. So what assumptions we take as far as this MOS device is concerned.

(Refer Slide Time: 4:36)



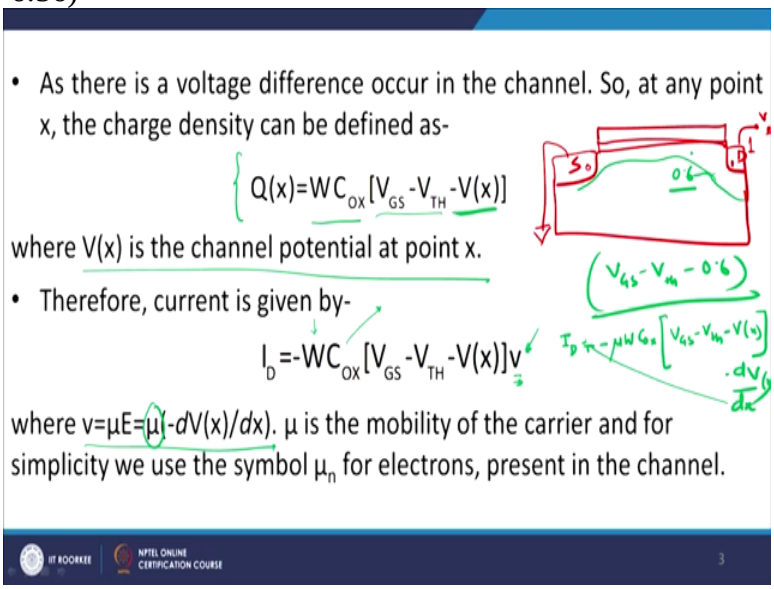

We assume that it is a, first of all we assume that it is a gradual channel approximation and then we assume that charge sheet model. What is gradual channel approximation? Gradual channel approximation tells me that the electric field in the transverse direction is slightly larger as compared to that in the longitudinal direction, right? So that is one. So basically a 1D Poisson equation will be enough to solve the gradual channel approximation based design. So all your long channel devices, MOS devices, you can apply gradual channel approximation wherein, we tell you that the electric field along transverse direction is relatively larger as compared to that along the longitudinal direction.

So if this is X and this is Y, then we say EX is relatively larger as compared to EY, right? And we can do one-dimensional Poisson equation solution to get these values. We also come to what is known as the charge sheet model which basically tells me that we assume that it is basically a sheet of charge near a silicon-silicon dioxide interface and therefore what is referred to as a charge sheet model. Right? Which means that it is basically a model which is assuming that it is a sheet of charge and therefore rather than per unit volume, it is basically per unit area that the charge will be collected with.

So we have two approximation, major approximations and the, one is known as gradual channel approximation, also referred to as GCA. Right and we have a charge sheet model here. Now please understand that you might not have gradual channel approximation when your channel dimensions are very very small. So if your lengths are very very small right, so it is okay to have if your length is very large, electric fields might be low. But if you reduce your channel lengths to very low values, then with the same value of $V_{DD}$, $V_{DD}$ by L which is electric field in let us suppose the Y direction, is very high.

Then my basic assumption that EY, EX is much larger as compared to EY will not hold good and therefore, both will be approximately equal to each other and therefore GCA will not hold good. So GCAs are always holding good or always good when your channel lengths are typically very large in dimensions, right? Fine. And that is what we define as channel length saturation.

(Refer Slide Time: 6:56)




As there is a voltage, now what I do is, we have seen that when we do not apply any drain bias, the length of my or the thickness of my charge sheet of electron which is in this case the charge carriers are almost constant, is exactly constant from source to drain side, this is a thin layer which you see. Now if I apply a bias, let us see how does the whole scenario change. Now at any point, so let us suppose I apply a bias $V_{DS}$, so I have got this, I have got this, I have got a drain, I have got a source, $V_G$ right, I have got a this and this.

I apply a drain bias here right, $V_{DS}$. What it will do is, since you apply 1 volts, so there will be a potential profile from 0 to 1. So somewhere here will be 0.5, here will be 0.6, here will be 0.2, here will be 0.4, so on and so forth. As a result, at this edge, you will have the most reverse bias. Is it okay? And as you move towards the source end, the reverse bias starts to reduce because the value of voltage starts to fall down right? And therefore reverse bias increases and therefore the depletion thickness also goes on reducing.

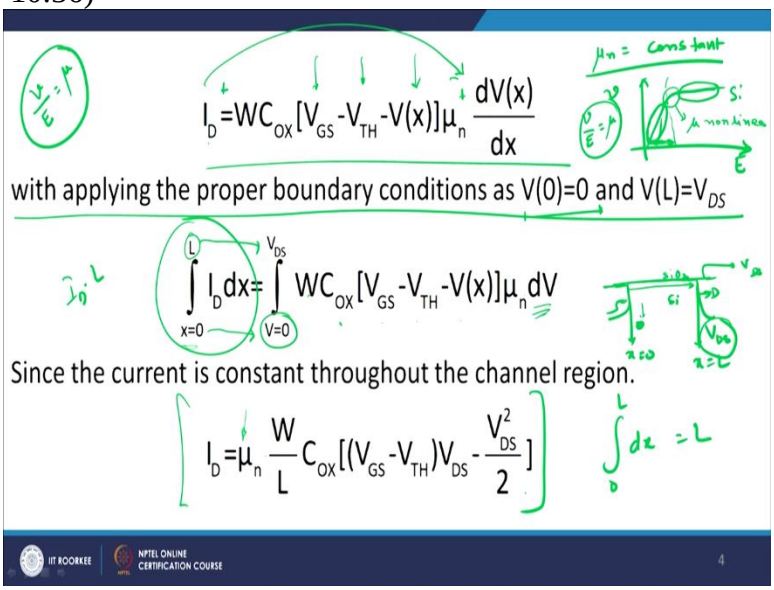
So you will have the maximum voltage thickness potential depletion thickness near the drain end and as a result I will just draw for you. So you will have the maximum voltage thickness, depletion thickness near the drain end and you will have let us see. You will have sorry, so you will have what? You will have largest depletion thickness here and then as you move towards the drain end, it will become thinner and thinner. So you will have a largest depletion region here because the voltage is maximum here.

Then we define Q of X. Q of X is minding the charge at any point X is equals to W into $C_{OX}$ right for obvious reasons, into $V_{GS}$ - $V_{TH}$ - $V_X$. Why minus $V_X$? Because this is the voltage which is formed because of the charge carriers. So that has to be subtracted. So where $V_X$ is the channel potential at any point X. So let us suppose that this point X, you have 0.6, 0.6. So you have to subtract $V_{GS}$ - $V_{TH}$ - 0.6. Why 0.6? Because 0.6 is just required to form the depletion region.

So subtract that to get the charge right? And therefore the current is given by minus W. Minus W into C oxide into voltage, applied voltage is V. Why minus W? Because of the simple reason that if the reason being that it is an electron. So the charge will be opposite in nature to the flow of electrons right. Alright, C oxide is the oxide capacitance per unit area of the device. Right? And as you can see here from this discussion that as the value of $V_X$ goes on reducing, the current goes on increasing and so on and so forth.

So as you make a V larger, where V is the velocity of the charge carriers and is given as μ into E where μ is the mobility of the charge carriers and electric field is minus $DV_X/DX$, where V is the voltage at any particular point at this point. So I can replace this by what? I can replace this by $I_D$ equals to this μ which you see from here, can be brought forward. So I can write down minus μ W $C_{OX}$ right $V_{GS}$ - $V_{TH}$ - $V_X$ right into $DV_X/DX$. Fine? Then what do you do? Then let us see how you go about it. You can then transfer DX on this side and let us see how it works out.

(Refer Slide Time: 10:36)




Now what you do? So you get $I_D$ equals to therefore minus this thing. So this there was a minus sign here initially and there was a minus sign here, so there, that is the reason this become plus now. Right? So you see $I_D$ therefore is equals to W into $C_{OX}$ which is the oxide capacitance per unit area multiplied by $V_{GS}$, Gate to source voltage minus the threshold voltage of the device, $V_{TH}$ minus $V_X$ right because $V_X$ is the channel potential multiplied by mobility or the charge carriers and we write $DV_X/DX$. Now we are assuming here, for all practical purposes, mobility of charge carriers to be constant. Right?

Now this is a bit, slightly over assumptions but for, at this stage of your understanding, assuming mobility to be constant is the right approach to find out the current. Now as you can see from this discussion, that how do you define mobility is basically velocity per unit electric field. That is how you define mobility right? So constant primarily means that and if you plot for silicon, if you plot for silicon velocity versus electric field for silicon, it looks something like this. Right? So this is the region where V by E is constant.

Similarly here is the region where V by E is constant. Right? Somewhere here, you have a non-linear mobility profile. So here it is non-linear. Right? But let us assume we are not going to that high electric field. You are restricting yourself to low electric fields and velocities are small and therefore I can safely assume that V by E is constant which is equals to the mobility. You also

should know at this stage before we move forward that mobility can also be of 2 types. One is known as the bulk mobility, another is known as surface mobility.

What is surface mobility? See, it is pretty simple to understand. Bulk mobility is when you have silicon right and you apply a bias, silicon is doped with say N type material, phosphorus, and you apply a bias and electron is moving through the bulk of silicon right. It is scattered by the atoms of silicon and so on and so forth. And it reaches to the drain end, and that is known as the bulk mobility. That velocity by electric field will be the bulk mobility. What is surface mobility?

Well, in reality when you talk about MOSFET, as I discussed with you in your previous discussions that in a MOSFET, the electrons will be moving if it is drain source and you apply a drain bias $V_{DS}$, then the electrons will be moving through this region right where this is basically your silicon dioxide and this is your silicon. So they will be moving across silicon-silicon dioxide interface. This interface is not very good because you will have large amount of traps, there will be some oxygen atoms which will be dangling and so on and so forth.

So there will be heavy scattering, unlike when the electron move through the bulk, as it moves through the surface, the scattering is very heavy. So when the electrons are very close to the interface of silicon-silicon dioxide, the mobility still falls down lower and therefore the current falls down because current is directly proportional to the mobility of the charge carriers right, as you can see from here. It is directly proportional to mobility. And the mobility falls down.

When the mobility falls down, the current also falls down but that is a second-order effect which will come later on. At this stage, it is sufficient to know that mobility is constant and this is how you can find the total current going through it. Now by applying the proper boundary conditions, so this is basically a differential equation and we require one, two boundary conditions to solve it. The first boundary condition is that we assume that at X equals to 0, $V_0$ equals to 0 and $V_L$ is equals to L which means that at this point, this is X equals to 0 right and this is X equals to L.

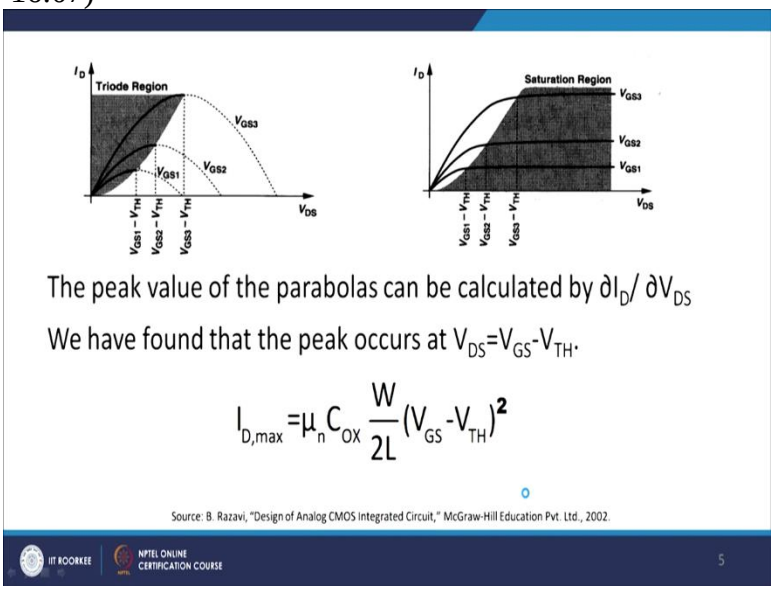
At X equals to 0, the potential is assumed to be 0 and at this point, it is assumed to be $V_{DS}$. Right? Obviously because I am assuming that the potential drops within the source and the drain region is almost equals to 0. Because it is heavily doped, it is having $N^+$ $N^+$ region typically within it. It has got $N^+$ region typically within it and therefore I would not expect to see a large

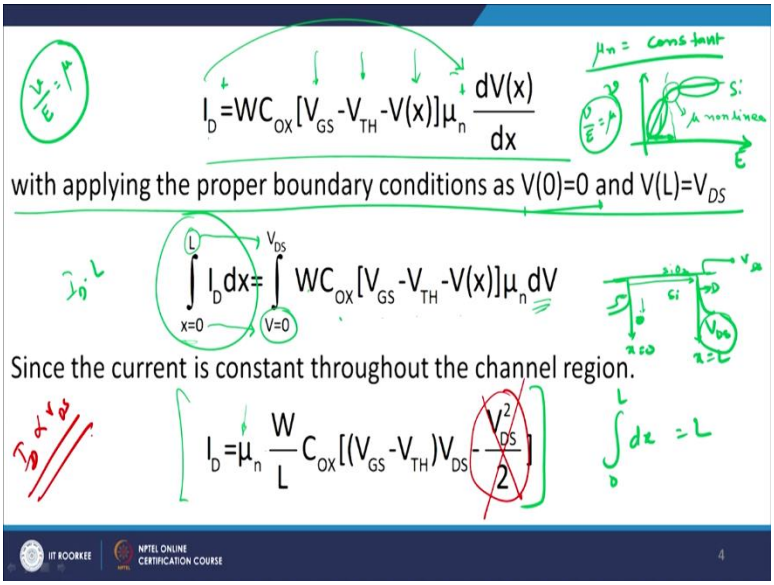
potential drop in the source and drain region right. And therefore it is safe to assume that at X equal to 0, you will have a potential equals to 0 and at X equals to L, you will have potential equals to $V_{DS}$ or drain to source voltage, whatever drain to source voltage you have applied.

Now what you do, you may take $D_X$, cross multiply. So $I_D$ $D_X$ and you integrate it from 0 to L right, where L is the length of the transistor. And similarly, V is from 0, so when X equals to 0, V was equals to 0. When X equals to L, V was equals to $V_{DS}$ and you get W into $C_{OX}V_{GS}$ - $V_{TH}$ - $V_X$ $\mu_n$ into $D_V$. Fine. So when you integrate this one, this is straightforward. This becomes, so when you integrate this left-hand side, this becomes $I_D$ into L right? Because 0 to L is nothing, so integral $D_X$ 0 to L is nothing but L right and therefore that L is going to the denominator.

So I get W by L and this μ is already there, I bring it mu outside. This is $C_{OX}$ and then what I do is, this μ L also comes in the front side because it is constant one independent of the value of $V_{DS}$ and then if you solve integral $D_V$ if you do with respect to $V_{GS}$ - $V_{TH}$ - $V_X$ and from 0 to $V_{DS}$ if you solve this, I get this equation here, right and it is equals to $I_D$ equals to $\mu_n$ W by L $C_{OX}$ $V_{GS}$ - $V_{TH}$ into $V_{DS}$ into $V_{DS}^2$ by 2. Fine.

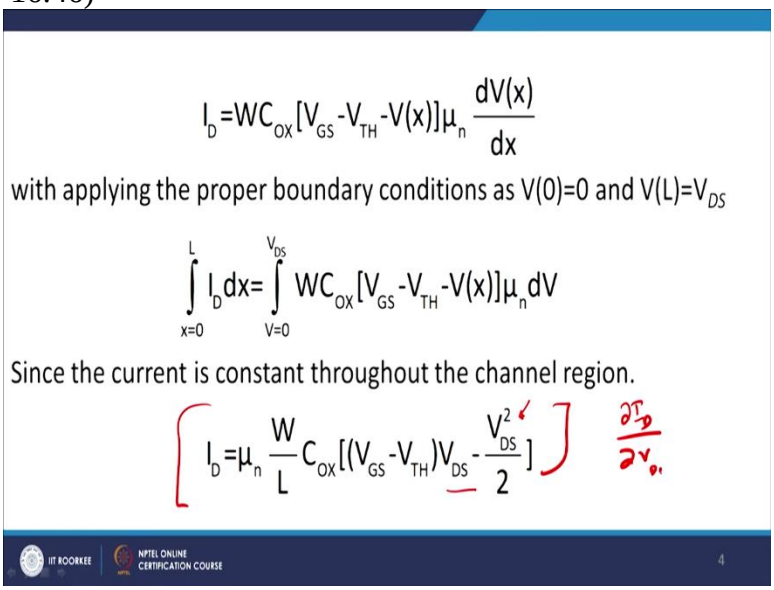
(Refer Slide Time: 16:07)

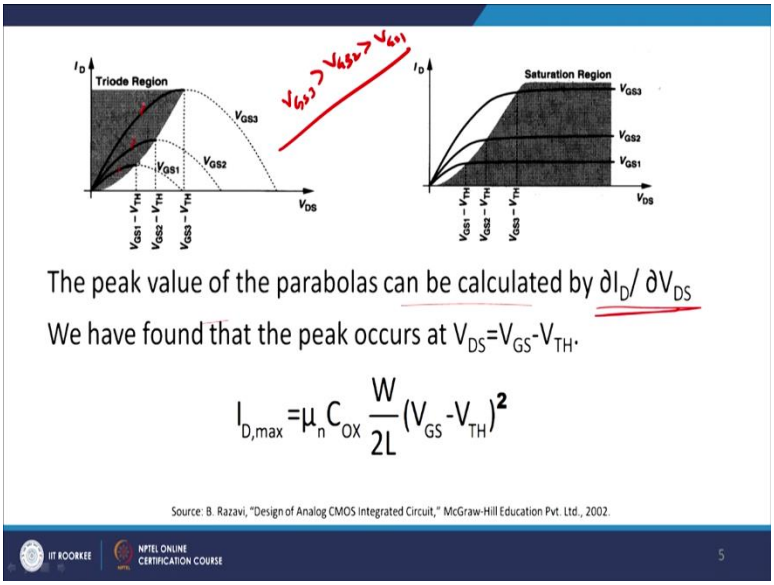



And this is the value of the current which you see in front of you. Now you see here, quite interestingly that now the drain current is not a linear function of $I_{DS}$ right. It would have been so, had had this term would not have been there. Then I would have experienced that $I_D$ is almost proportional to $V_{DS}$ and therefore there would have been almost a linear increase.

(Refer Slide Time: 16:46)



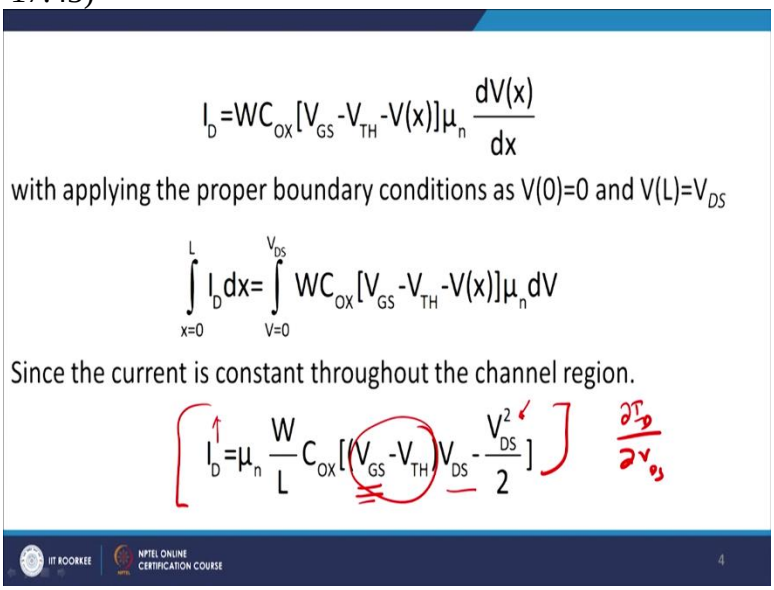

But since there is a presence right, since there is a presence and you do have a $V_{DS}$ term coming into picture right, you do have a $V_{DS}$ term coming into picture which is also $V_{DS}$ here also you have got right and you have $V_{DS}$ here also and this is $V_{DS}^2$ mind you.

So what I can safely say is that there will be a non-linear profile at smaller values of $V_{DS}$. Let us see how it works out. Now, so if you look at this basically the previous discussion, this is basically a parabolic equation and then if you find out a $\partial I_D$ $\partial V_{DS}$, then you can find the peak of the parabola right. And this is what is, it can be calculated $\partial I_D$ $\partial V_{DS}$. So once you differentiate it

for various values of $V_{GS}$, now if you plot $I_D$ versus $V_{DS}$ right, for various values of increasing $V_{GS}$, so $V_{GS3}$ is greater than $V_{GS2}$ is greater than $V_{GS1}$ right.

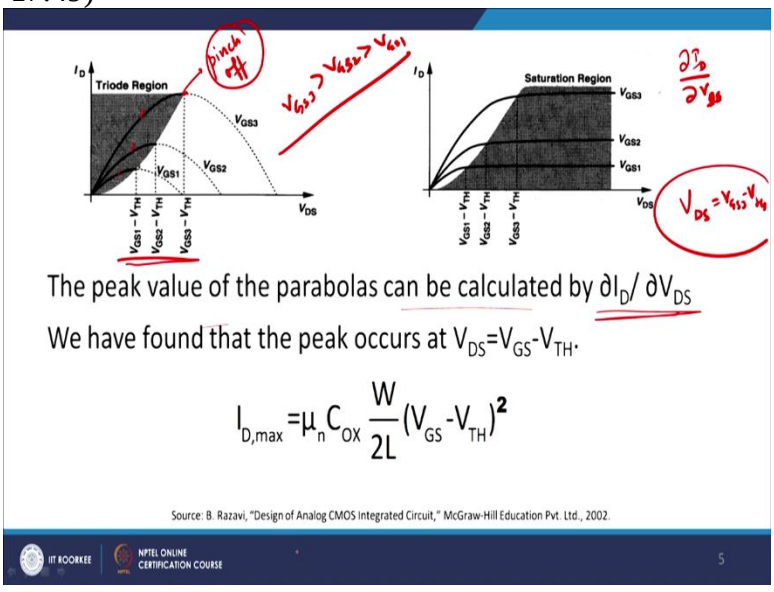
If this is true, which means that the gate voltage is, as the gate voltage increases, you see the currents are also increasing and this is very simple because your over drives are increasing, see.

(Refer Slide Time: 17:43)




If $V_{GS}$ is higher and higher, this quantity becomes higher and higher and therefore the current also increases.

(Refer Slide Time: 17:49)



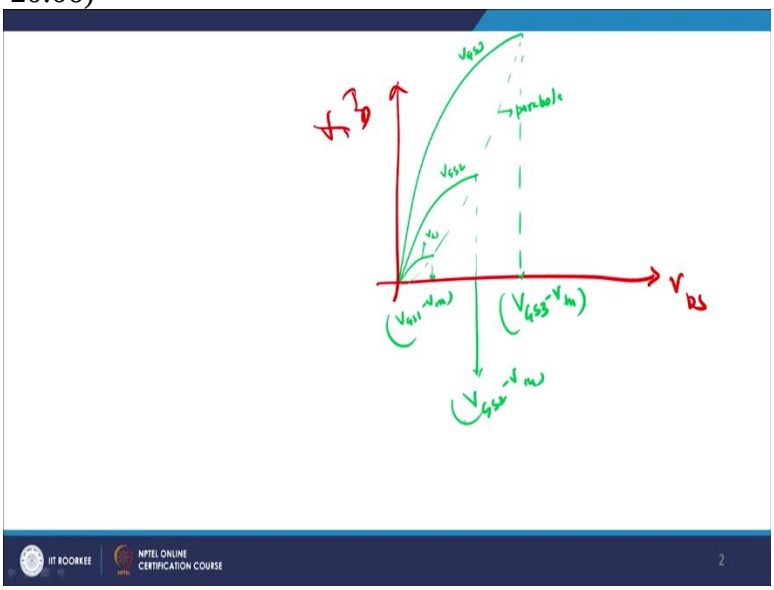
As you can see, therefore as the gate voltage increases, your, this value also increases. Not only that, your $\partial I_D$ $\partial V_{GS}$ $\partial I_D$ $\partial V_{GS}$ will, well $\partial I_D$ $\partial V_{DS}$, sorry, will get drain to source voltage will also show a shift towards right for a higher gate voltage, which means that I will explain to you physically why is it. Mathematically, let us understand first. That once the drain to source, once the gate voltage is higher, you require a larger drain voltage to sort of, if you look very carefully from this region, this is the maximum value of current.

After this, either the current will fall, according to this equation the current will fall or it will remain constant. It will never increase for sure. Which means that charge carriers have attained its constant value. If this is true, then let us see why is it like physically. Physically, when you have reached the maximum value here, it primarily means that higher the value of $V_{GS}$, more is the transverse direction electric field, larger will be the number of carriers and therefore to deplete it, you require a larger drain to source voltage.

You are getting my point? And therefore this pinch off, this is also referred to as a pinch off point let us suppose. This is pinch off point let us, let me give a term here. This pinch off point goes on increasing right as the value of $V_{GS}$ goes on increasing. So higher the value of $V_{GS}$, more will be the pinch off point, which means that if you see, if you look very carefully, this is the pinch off point which you get. That when you have $V_{GS3}$ the pinch off, the value of $V_{DS}$ right must be equals to $V_{GS3}$ -$V_{TH}$ at which it becomes maximum.

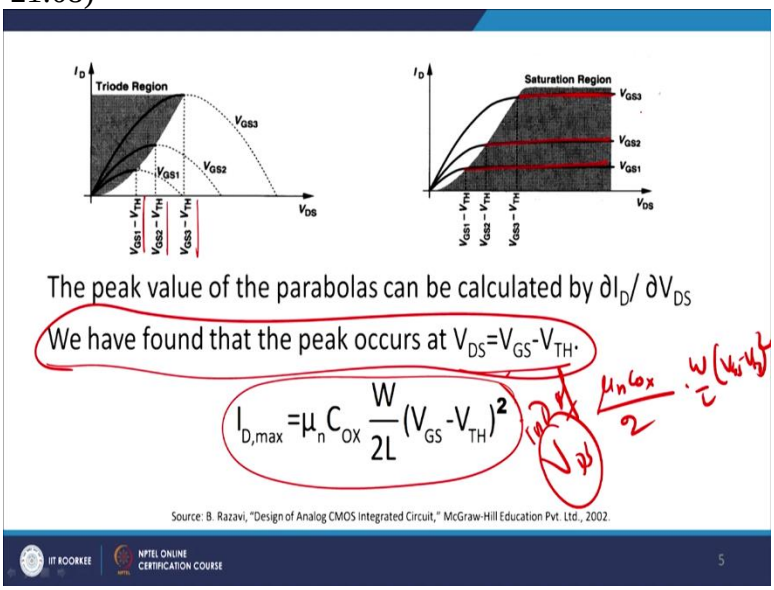
And the reason is very simple. I suppose you have understood by this time why is it very simple. The reason is being simple is that as you make the gate voltage higher and higher, you have now larger number of carriers available, that is the reason the current is higher but then, you therefore have to give a larger drain to source voltage to reverse bias the region and therefore make it depleted. And therefore the value of $V_{DS}$ is also higher.

(Refer Slide Time: 20:06)




So what we finally get from this overall understanding is that if you plot $I_D$ versus $V_{DS}$ right at least till the point when you have reached it, you will get something like this. We will discuss this point, let me not draw at this stage, anything think like this but let me draw it. So let me draw like this and we keep it like this right? As we move $V_{GS}$ higher right and when we move $V_{GS}$ lower, it will be like. So if you plot it, this will be parabola. This will be $V_{GS3}$ - $V_{DS}$, this will be $V_{GS2}$ - $V_{TH}$ and this will be $V_{GS1}$ - $V_{TH}$, where this is the value of $V_{GS1}$, $V_{GS2}$ and this is $V_{GS3}$. Fine. So higher the overdrive, more is the pinch off value of voltage which you see.

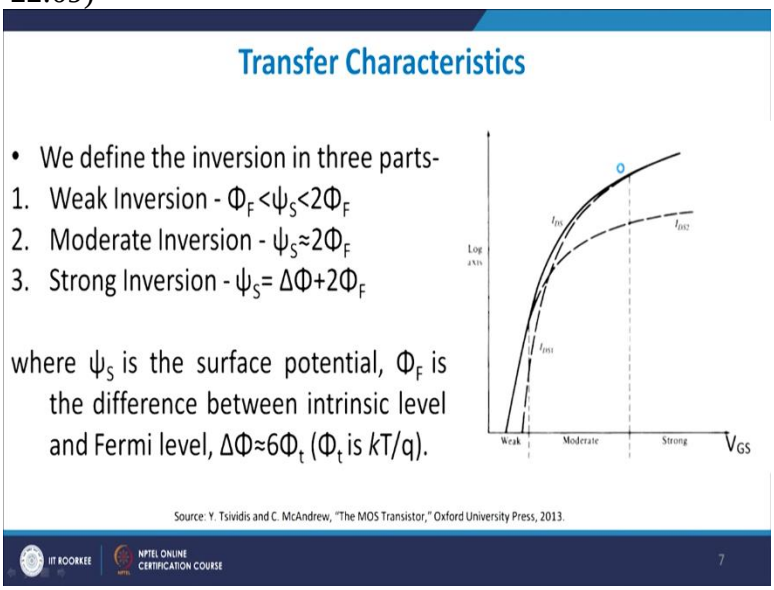
(Refer Slide Time: 21:08)

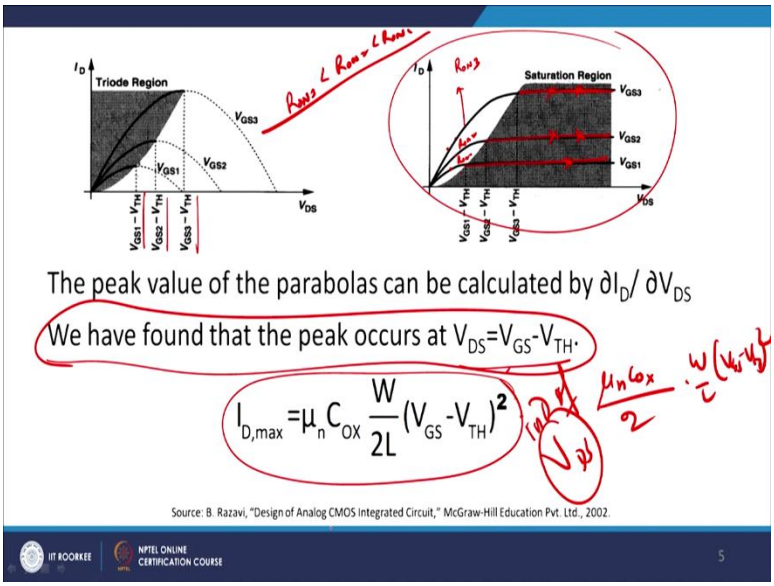



And that is what you see here at this particular point that so as you can see from all this discussion, this is the pinch off point, alright. So now we have found out that the peak occurs at $V_{DS}$ equals to $V_{GS}$ - $V_{TH}$. Right? So the peak occurs at $V_{GS}$ - $V_{TH}$. Beyond this particular point right, if your peak occurs at this, you just feed this value into the previous equation and you automatically get $I_D$ to be equals to this much. So $I_D$ Max is $\mu_n$ C oxide right by 2 into W by L into $(V_{GS} - V_{TH})^2$.

And as you can see here, this is actually independent of $V_{DS}$. So this is independent of $V_{DS}$. As you can see therefore this region is defined as the saturation region. So we have saturated. So current is constant independent of the value of $V_{DS}$ which you see. Right? And therefore it is constant for all larger values of $V_{DS}$.

(Refer Slide Time: 22:09)



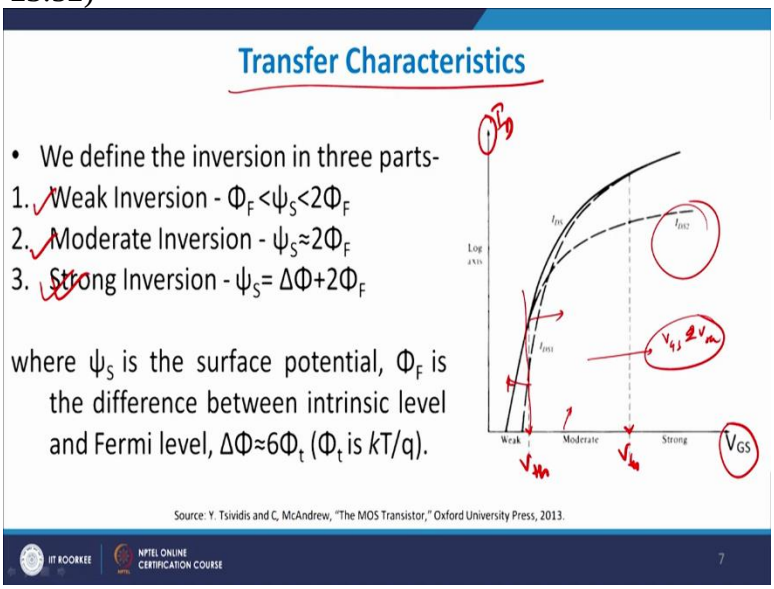




With this knowledge, let me come to three important characteristics of so this gives me, so what we have discussed till now is that we have discussed the IV characteristics of the device. And we have seen that if current versus voltage is plotted for a fixed value of $V_G$ right, if this is what is plotted, then we get almost like this profile with us right. Then we almost get saturation region, this is the saturation region and this is the linear region, and so on and so forth. So at this point, the device behaves as a constant current source, this point it starts to behave like a resistor right and the value of resistance will go on changing as you move from this to this to this.

So higher the overdrive, lower is the resistance. I think it is clear to you why is it like that because higher the value of the overdrive, more will be the charge, more will be the charge, less will be the resistance right. And therefore this, suppose this is on resistance 3, this is $R_{on2}$ and this is $R_{on1}$, then $R_{on3}$ will be smaller than $R_{on2}$ smaller than $R_{on1}$. Right? Because the slope is going on increasing as we move ahead and therefore the resistance is falling down because it is inverse of the slope which we are trying to find out.

With this knowledge we have actually understood the IV, or the current versus voltage characteristics of the device and we have understood the basic fundamental principles of current versus voltage. Now what we will do is that we will fix the value of $V_{DS}$ and we will vary $V_{GS}$ and see how does my current change, right? So that is what we are next, we are going to do and that is known as the transfer characteristics right. And that is known as transfer characteristics.

(Refer Slide Time: 23:52)




So what is the meaning of transfer? It is basically $I_D$. This is $I_D$ versus $V_{GS}$. We will not go into details of this one. I will make it very simple for you to understand and we will discuss further than this. See as I discussed with you that when the gate voltage is much smaller than threshold voltage, the device will be in the off state because the threshold voltage because there are no large number of charge carriers available in the device and as a result device will be considered off state.

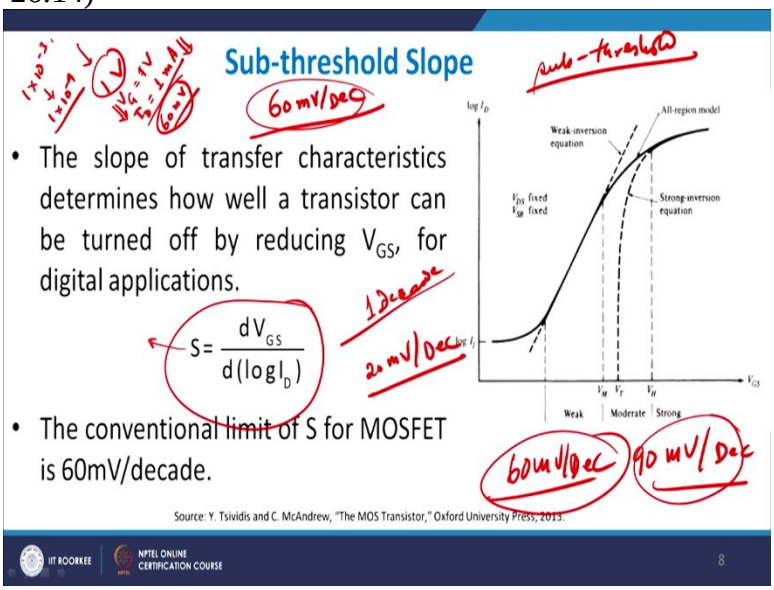
When the gate voltage crosses the threshold voltage, you will have large number of charge carriers in the silicon-silicon dioxide interface and as a result, you will have an on state available with you. Now, typically, if you look at this graph here, we define this to be as whenever the gate voltage, so this is my threshold voltage. Whenever my gate voltage crosses the threshold voltage $V_{TH}$ right, I can get, sorry, this is my threshold voltage, I am sorry, this is my threshold voltage. Whenever my gate voltage crosses the threshold voltage on the right-hand side, I get a moderate inversion.

Whenever my gate voltage crosses twice $V_{TH}$ or large values, I get strong inversion and anything smaller than this, we get a anything gate voltage smaller than the threshold voltage, we defined that to be as a weak inversion. So we have three points of notation. One is known as the weak inversion, one is known as moderate inversion and we have a strong inversion. Strong inversion primarily means that, inversion means it has become, it is just becoming N type and it was initially P type.

Strong inversion means, if it is becoming twice the, twice, so let us suppose you had a P type substrate where the number of holes were hundred right. Now what you do? You go on adding, you apply a negative bias, so electrons are being pulled, they are recombining and then holes are fully removed, fully depleted. So there are no holes now. We have only depletion available. Now what you do? You still go on increasing the gate voltage, more number of electrons are pulled there and it crosses equal to hundred. Then we define it to be inversion.

So now what has happened is, as many holes were initially present, that many number of electrons are now present. Right? If it is almost double of that electron, we define that to be as a strong inversion, right? So I have a weak inversion, moderate inversion and strong inversion.

(Refer Slide Time: 26:14)




We define a, sorry, we define a new term here and that is known as a sub-threshold slope. A very important term in both analog as well as in digital world, especially in digital world where switching is to be done. See what happens is that though we presume that our gate voltage is equal to threshold voltage is a transition point between on and off state, in reality, this is not true which means that even when the gate voltage is even just smaller than threshold voltage, the device is not fully off right? The device is not fully off which means that there are certain, still some charge carriers left which has not been removed.

And therefore we define that region to be as the sub-threshold region. Means just smaller than threshold. Right? In that case, we define a term known as S, also known as sub-threshold slope as the amount of gate voltage required to change the drain current by at least one decade which means that, let us suppose your threshold voltage was 1 volt, right? You applied the gate voltage and the gate voltage is approximately equals to 1 volt. Now what subthreshold slope tells me is that at 1, when the gate voltage was 1 volt let us suppose right, when the gate 1 volt, the drain current which was flowing was let us suppose 1 milliamp, just for information sake.

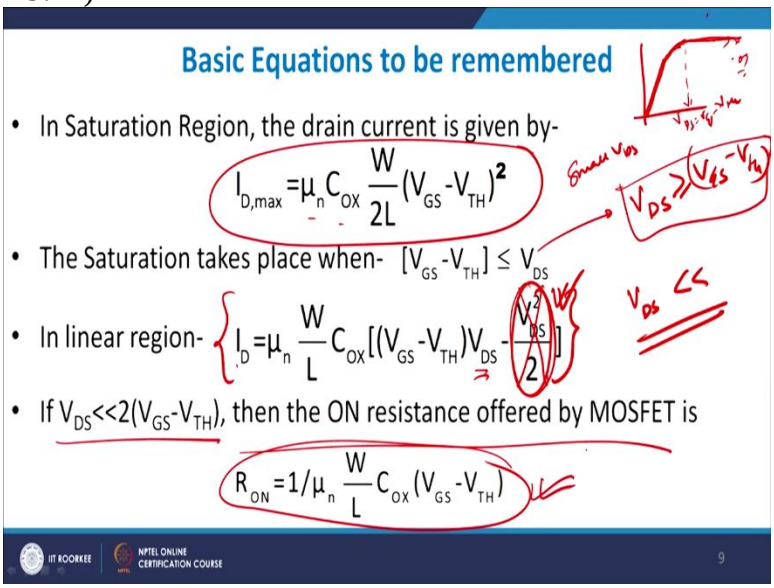
Now what I do? I decrease the gate voltage by say 60 millivolts, right? And if then the $I_D$ value, if this gate voltage reduces by 60 millivolts, if $I_D$ drops by one decade, one decade means from one milli, one milli is 1 into $10^{-3}$ and it drops to 1 into $10^{-4}$, 1 decade drop in current and to do that if I require that let us suppose 60 millivolts of gate voltage change, then we define

60 millivolt to be as the sub-threshold slope. And it is given, and the unit is given as 60 millivolt per decade.

So for one, for every one decade fall in the current, drain current, the amount of gate voltage which is supposed to be changed, in order to bring that change of the drain current is defined as my sub-threshold slope. Smaller the value, faster is your change. I hope you understand. Let us suppose, I have a device which gives me 20 millivolt per decade, it means that I have to only apply a 20 millivolt gate voltage change in order to 1 decade drop. So switching is faster.

So sub-threshold slope is basically means that that it is switching is faster. Lower the value, lower is the switching. So if sub-threshold slope is 90 millivolt per decade and I have a design which is basically 60 millivolt per decade, then 90 millivolt per decade then 60 millivolt per decade is a kind of much faster switching from on to off and off to on as compared to 90 millivolt per decade. Right? The conventional limit for sub-threshold slope is 60 millivolt per decade. That is the conventional limit. At T equals to 300 Kelvin, we define this to be as a conventional voltage.

(Refer Slide Time: 29:42)




With this knowledge, let me recapitulate the basic equations which is to, which you have to remember from next time onwards to make it easier for you to understand, the first thing in the saturation region, the drain current is given by $I_{DMax}$ and that is a maximum current, is given as

$\mu_n$ C oxide W by 2L into $(V_{GS} - V_{TH})^2$ which means that if your $V_{GS} - V_{TH}$ is doubled, your current will be becoming almost 4 times with you and that is one of the major but here if you see clearly, it is independent of $V_{DS}$.

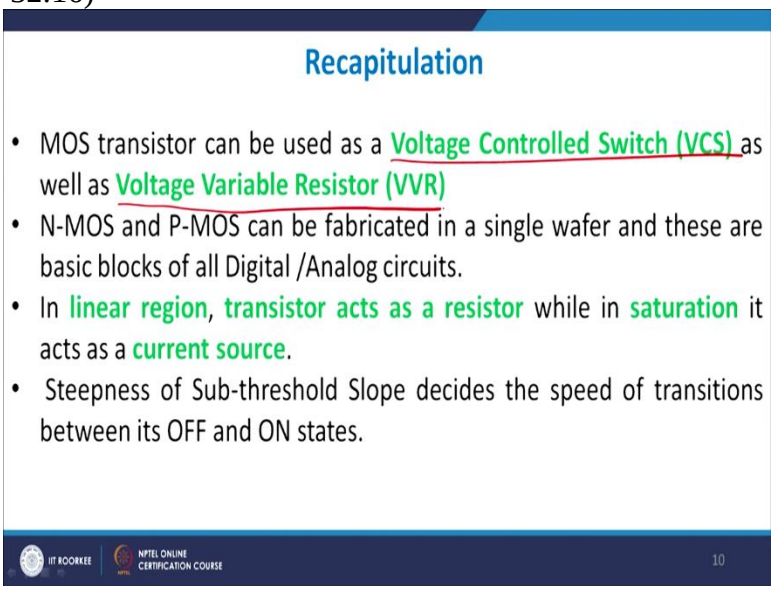
So $I_D$ is independent of $V_{DS}$ primarily meaning that $I_D$ is almost constant. Right? It is almost constant independent of $V_{DS}$. What is the condition for saturation? The condition for saturation is this one which is $V_{DS}$ should be greater than equals to $V_{GS} - V_{TH}$. Very very important term from digital as well as analog point of view and we should know what is the condition for the onset of saturation. For onset of saturation, $V_{DS}$ should be greater than equals to $V_{GS} - V_{TH}$, which means that drain to source voltage should be larger than the overdrive of the device.

In the linear region, we define this as my drain current source. As I told to you, if your and let me say that $V_{DS}$ is very very small, a very small value, then $V_{DS}$ square will be still smaller and I can neglect this and then I can have $I_D$ proportional to $V_{DS}$. That is the linear region of operation, right? So whenever you have linear region of operation is this one right when your $V_{DS}$ is very very small. In that case, $V_{DS}^2$ will be also must so is $V_{DS}$ is 0.2, 0.2 square is 0.04, right? So 0.04 is very small quantity, I can afford to neglect it.

But anything larger than 1, for example 0.3, say you use 0.3, 0.3 into 0.9, it will be also very small. So for small values of $V_{DS}$, small $V_{DS}$, this is almost linear because $I_D$ is directly proportional to $V_{DS}$. As you move your $V_{DS}$ slightly higher than 1 and makes it, this non-linearity comes into picture. And you do have a non-linear region here. Till how much point? Till $V_{DS}$ equals to $V_{GS} - V_{TH}$, the onset of saturation. After this, it becomes straight, almost straight line.

So I will, I have a saturation, non-linear and exactly linear in dimensions, right? When $V_{DS}$ is much smaller than 2 times $V_{GS} - V_{TH}$, the on resistance offered by the device is given by this formula that R on his equals to 1 by $\mu_n W$ by L $C_{OX}$ $(V_{GS} - V_{TH})$. And this gives you a small $R_{on}$ in this case. So with this, let me recapitulate what we did.

(Refer Slide Time: 32:16)




We have also understood that it is basically a voltage controlled switch, MOS device. Who is controlling? The gate voltage is controlling it as a switch. So I can apply a gate voltage rather than threshold to switch it on, less than threshold to switch it off. It is also acting as a voltage variable resistor because if you see, as I discussed with you, for a lower value of $V_{GS}$ in the linear region, when the lower value of $V_{GS}$, $R_{on}$ is large and as you increase the value of $V_{GS}$, $R_{on}$ reduces. So I can use it as a voltage variable resistor as in JFET.

Now typically, nMOSFETs and pMOSFETs can be fabricated in a single wafer and these are the basic building blocks of digital and analog circuits which is a very simple one. Of course, as I discussed with you, in linear region, the transistor acts as a resistor while in saturation it acts as a current source. So therefore MOSFETs you can see even without understanding anything, it can act as a resistor as well as a current source right and therefore the lower the, what is the sub-threshold slope? Sub-threshold slope is defined as the amount of change in the gate voltage to change one decade of drain current right. Lower the value, better the speed of the device is in switching right.

So with this, we have finished with the basic concepts of MOS device and we have understood how a MOS device works. We have also understood the basic functionality of a MOS structure and its uses as a current source as well as a voltage variable resistor. We have understood three

regions of operation, linear, non-linear and saturation. We have also appreciated what is the value of $V_{DS}$ at which you will have onset of saturation. Right? And this will be quite interesting and important for our subsequent studies. I thank you for your patient hearing. We will take up the next issue of MOS devices in the later courses. Thank you!

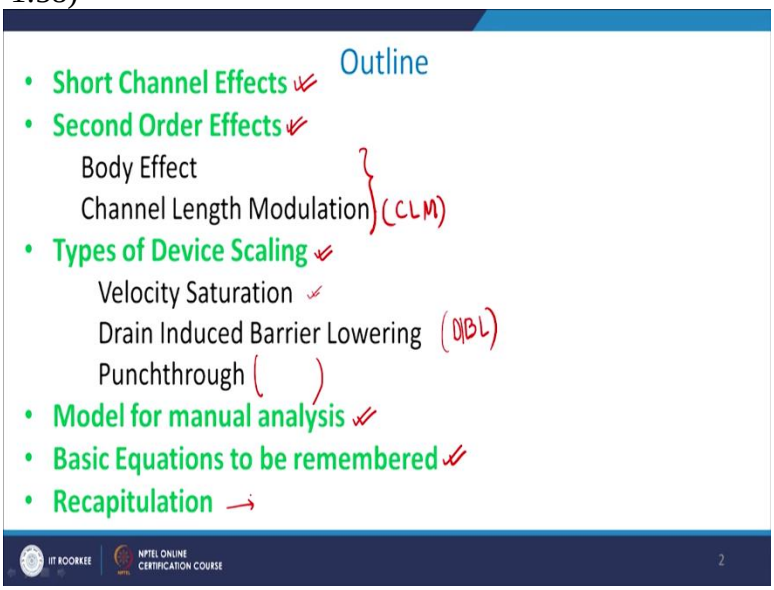
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture–16**
**MOS Transistor Basics – III**

Hello everybody and welcome again to the NPTEL online certification course on Microelectronics: Devices to Circuits. As we have, in our previous interactions, we have understood basic BJT which we have already done quite a lot and then we started with CMOS technology and to understand that, we understood what is the MOSFET technology and we understood therefore in the previous 2 slides or previous 2 interactions or modules, what is a MOSFET really, the structure of MOSFET and the functionality of MOSFET.

So first order, we saw that the MOSFET therefore can be the recapitulated as a current source when working in the saturation region, it can work as a voltage variable resistor when it is working in the linear region of operation. These two things we have already seen as a device. We have also seen that the MOSFET can be used as a switch when your gate voltage swings above and below the threshold voltage of the device. We have also seen that there are two types of devices, enhancement mode MOSFET and depletion mode MOSFET.

In one case, the threshold voltage is positive and in the other case, the threshold voltage is negative. Enhancement mode is also known as normally off-device and depletion mode is also referred to as normally on-device. What we will be looking subsequently is the following thing right and the topic of today is transistor basics. We will be going into the part three part.

(Refer Slide Time: 1:58)



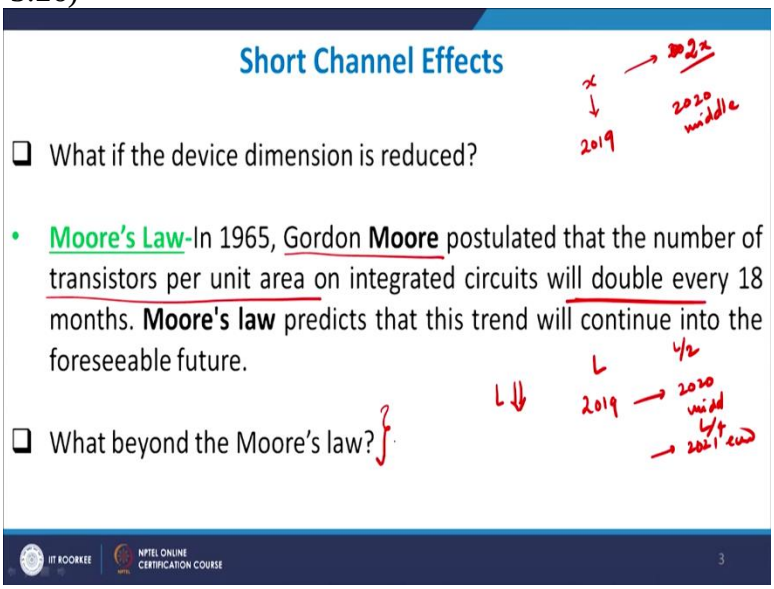

Within part 3, the outline of my this topic will be, we will be looking at various short channel effects, right. We will be looking at second order effects. Short channel effect primarily means that as you reduce the dimensions of the device, what second-order phenomena or effects comes into picture which is detrimental to the performance of the device, that is basically a short channel effect. We will be also looking at body effect as channel length modulation, this is CLM, right?

And this happens at extremely low dimensions and how does it influence the output electrical characteristics of the device, we will be looking into that. We will be also looking at the type of device scaling. Primarily we will be looking into two types of device scaling. One is electric field scaling, another is the voltage scaling. And then we will see its relative advantages and disadvantages. Within short channel effects, we will be looking into velocity saturation effects, right?

We will be actually looking into drain induced barrier lowering, also referred to as DIBL and we will be looking at punch-through which is basically a hot carrier effect phenomena. Then, once the MOSFET is, MOSFET as a device is clear to me, we will like always this MOSFET to be used in circuit simulation purposes. So can we have techniques of doing some amount of circuit simulation using, so can we have therefore some electrical equivalent models of MOSFET which can be used for circuit purposes?

We will also look at basic equation we should, we need to remember and then recapitulate the whole talk, right.

(Refer Slide Time: 3:26)




Let us look at the various short channel effects. Now why short channel effect came into picture? See, in 1965, Gordon Moore, he was a scientist at IBM and he predicted a very important trend, very famously known as Moore's Law which is very famous within the semiconductor industry as well as in academia. He stated that for every one and a half year growth of semiconductor industry, you would expect to see that the number of transistors per unit area will be doubling right which means that if the number of transistors per unit area is X today, then after so if it is at, this is 2019 let us suppose and then at middle of 2020, right, I would expect to see this the number of devices, active devices per unit area to be equals to 2X right.

And this was given by Gordon Moore who stated that the number of active devices will increase and why will it increase? To increase the functionality of the chip and to enhance the functionality of the chip. So those companies or those industries or those people who are able to sustain this Moore's Law have been able to increase the functionality of the chip and therefore the profitability is there when it follows Moore's Law. Easier said than done because when you want to reduce the, when you want to increase the number of charge carriers or number of active devices per unit area, the only option available to you is that you reduce the dimension of the device.

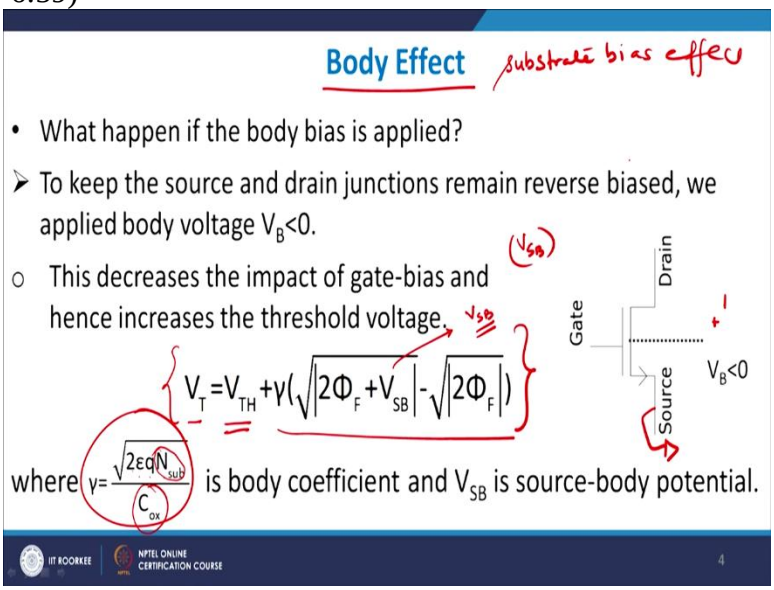
Because your silicon area is always fixed right and you cannot increase it drastically beyond a particular point because of fabrication limitations. So the only option available to you, if you want to follow Moore's law is that you want to reduce the dimension of the device right. So how do you reduce the dimension of the device? For example, in a MOSFET, how do you reduce dimension of the device? You simply channel length, you lower it. So if today your channel length in 2019 is L, then in the middle of 2020, 2020 middle I would expect to see this to be as L by 2.

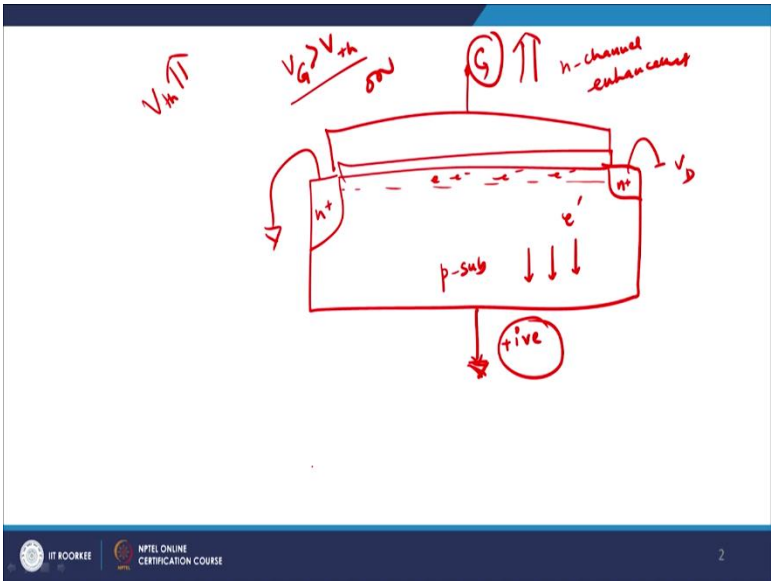
Similarly, at 2021 end, 2021 end, this should be equals to L by 4 and so on and so forth. So you see for every, for example, 2 to 3 years of growth approximately, you are almost quartersizing your device, right? And so the idea was therefore that in case of a FET for example, when you are reducing the channel length, you are also making the source and drain come closer to each other right. Now theoretically speaking, therefore if you continue with Moore's Law, a time will come when that length and drain will touch each other or will be close to each other and they will be shorting.

Once the shorting is there, it will not work as a FET anymore. So therefore people are trying to find out alternatives to your bulk MOSFETs so that your performance, electrical performances do not get degraded at low dimensions. So that is the reason, I was saying what is beyond Moore's Law right and I would recommend that you can have a look at large number of research papers as well as large number of videos on the YouTube and a large number of open literature which actually gives you an idea what happens beyond Moore's Law.

Now therefore, if these dimensions are reducing, then there are certain things which are known as short channel effects which comes into picture. So that is what is we are referring to. That is once the dimensions are reduced because of Moore's Law, you have short channel effects coming into picture.

(Refer Slide Time: 6:59)




The first effect which I wanted to stress in front of you is basically the body effect right? Or also known as substrate bias effect, substrate bias effect or also known as body effect. We will come to this as we discuss. Let me explain to you in a very simple terms and then we can move forward, right? So if I have a p type substrate right and we have discussed this earlier also, we have p type substrate and I have got $N^+$ source, $N^+$ drain and I have got silicon dioxide metal and then you have gate, source is grounded right.

And then the drain is, you are applying a voltage $V_D$. Now in the substrate side, we generally ground it. For most practical purposes, we ground it. But let us suppose we do not ground it and

we start giving let us suppose a positive bias. A positive bias if you give and if it is an N channel enhancement model MOSFET right, then we can explain to you, as we have discussed this point $V_G$ should be greater than equals to $V_{TH}$ for the device to be on, which means that there will be large number of charge carriers here.

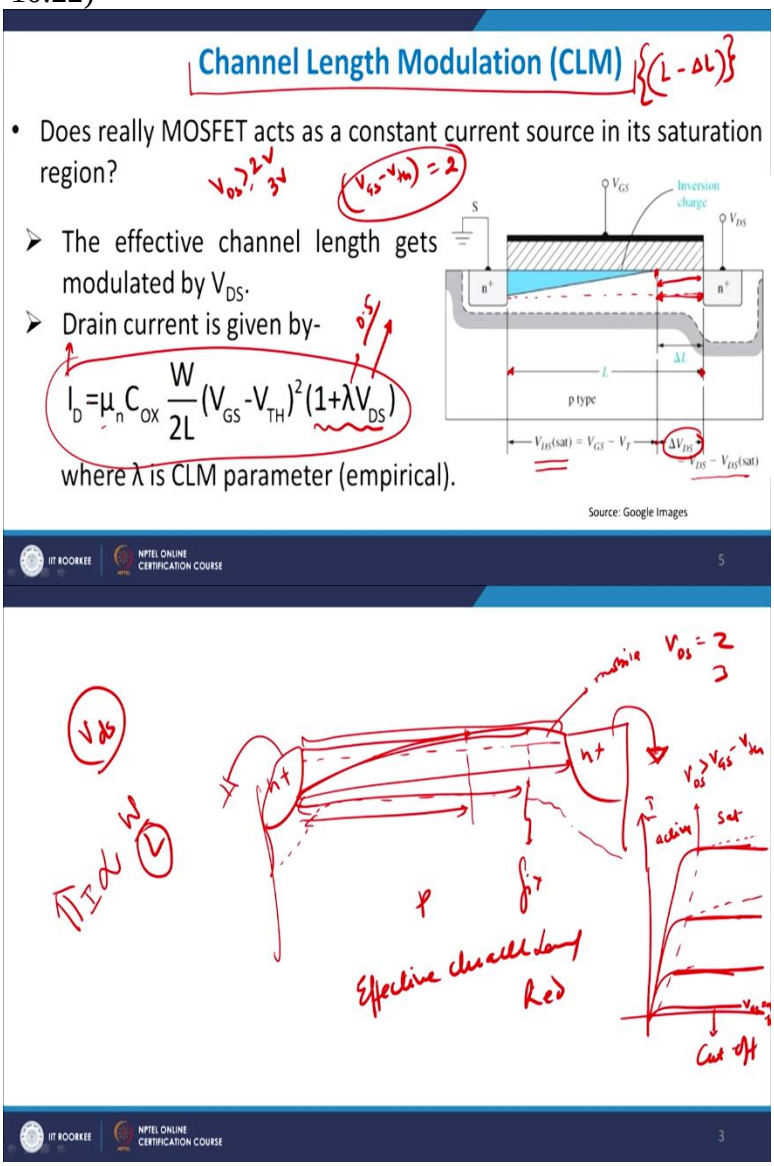
What are the charge carriers? Electrons, free electrons right, will be here provided $V_G$ is greater than $V_{TH}$, threshold voltage. But now, you apply the positive bias on the substrate which means that there is this positive bias will be pulling these electrons towards itself. These electrons will be pulled here. As a result, your gate voltage has to further increase in order to form the channel. So what happens to threshold voltage? The threshold voltage rises. This is known as the substrate bias effect.

That if the substrate which was ordinarily grounded right, you apply a positive bias in case of an N channel MOSFET, you end up having a threshold voltage increased. Not a good idea but that is what is happening. Now so the definition is something like this that when your $V_{TH}$, $V_{TH}$ is the value of threshold voltage without body bias effect and this is with body bias. And this is with $V_{SB}$ is basically my substrate source, source to bulk potential right.

And you have to give the source to bulk potential, we do not have to worry about its sign, we do not have to worry about sign. If it is source to bulk, source is generally grounded right. Bulk will be given, so if we give bulk positive dimension, positive, then source to bulk will be positive. If it is negative, source to bulk will be negative in dimensions or in nature. $\gamma$ is referred to as a body coefficient parameter given by $2\varepsilon$ Q into $N_{sub}$. $N_{sub}$ is the substrate concentration of the doping concentration and $C_{ox}$ is basically the oxide capacitance per unit area.

And $Phi_F$ is basically my Fermi potential, right. So we have understood therefore that this $V_{SB}$, substrate voltage will have an adverse impact on the threshold voltage. So you actually started with a low threshold voltage and you might end up having a large threshold voltage which means that you require a larger gate voltage to switch it on, which means that you have a larger power dissipation for a particular device, which means that if there is a short channel effect, because of body bias, n channel enhancement MOSFET you give a substrate bias greater than 0, you might end up having a potential or having a threshold voltage which is larger than expected. This is the first thing.

(Refer Slide Time: 10:22)




The second short channel effect is basically by channel length modulation effect, right. And it is pretty simple and easy to explain but we will still nonetheless explain each one of them individually before moving forward. So we have understood what is basically a body bias or a back gate bias. Let me therefore give you an idea about what is channel length modulation effect. See, in our previous discussion, when we were discussing devices we came to the fact that whenever you have a channel, channel basically in our case means that it is basically having free charge carriers connecting source to drain in at least as much it is n type as it was initially p type.

That is what is basically known as a strong inversion phenomena. And the strong inversion therefore, you will have direct connectivity path available to you. Now what you do is that, on

the drain bias if you do not give any voltage then the substrate, then the depletion region on the substrate or the depletion region on the channel, near the channel this is $N^+$, $N^+$ and this is P. So if you do not give any bias, this is grounded, this is also grounded. Then the depletion thickness will look something like this, it will go something like this and it will be something like this.

It will be symmetrical in dimensions right. Now if you go on increasing it, a time will come when you will have inverted layer here. So this will be mobile charge carrier, this will be mobile and you will have fixed charge carriers here, fixed here. Again we have still grounded our drain. Now what I do is, I do something else and that is that I increase the value of your $V_{DS}$. So I go on increasing the value of $V_{DS}$ right. And let us see how does it influence the overall current carrying capability of the MOSFET.

If you remember yesterday's talk or the previous lecture's talk, you will see that when we were discussing saturation, we saw that this is parabolic in nature and this was happening at $V_{DS}$ greater than $V_{GS}$ - $V_{TH}$, this was saturation right. This was active right. And this was cut off. What was cut off? Whenever my $V_{GS}$ equals to 0, I define this to be as cut off or equal to threshold voltage, we define the lower value to be cut off. So $V_{GS}$ equals to $V_{TH}$ with cut-off.

Now you see, at saturation also referred to as active in some of the textbooks. If you look very carefully, this is almost a straight line which primarily meaning that the current is constant independent of the voltage right. And it starts to behave like constant current source. But in reality, we will see that the channel modulation effect does not allow you to do that. Let us see why. So as you make your drain voltage higher and higher, the depletion thickness here becomes larger and larger and it starts to eat away into the channel.

And as it starts to eat away into the channel, you see the inversion charge is only restricted up to this much point. However, if you did not have applied any bias and your drain voltage was 0, you are inversion layer would have been something like this. Right? All filled up with free electrons but now since we have apply a drain bias which is quite large, the depletion region itself eats away into the inversion layer and your effective inversion charge carriers are shifted from this region.

So the channel length was which was initially till this to this point, from this to this point has actually reduced by a factor ΔL. Therefore the effective channel length is L - ΔL. Right. And therefore your, if you define channel length as that length at which the inversion charge is available, then by that definition, you will automatically have a reduced channel length which is given by L - ΔL at this particular point. Right. Which means that therefore, if I give a potential $V_{DS\ sat}$, $V_{DS\ sat}$ is the saturation drain to source voltage, then $V_{GS}$ - $V_{TH}$ is the voltage drop till this much point. Right?

And the extra $V_{DSsat}$ which makes it saturation falls across this region. So this $\Delta V_{DS}$ which is $V_{DS}$ - $V_{DS\ sat}$, is the voltage which is required here. For example let us suppose, to form the, this is 2 volts. So $V_{GS}$ -$V_{TH}$ to require to form is 2 volts, right. So I require that $V_{DS}$ should be greater than 2 volts for saturation. Fine. That is okay. But then if I give 3 volts $V_{DS}$, then out of the 3 volts, 2 volt actually goes to form this channel right, this channel.

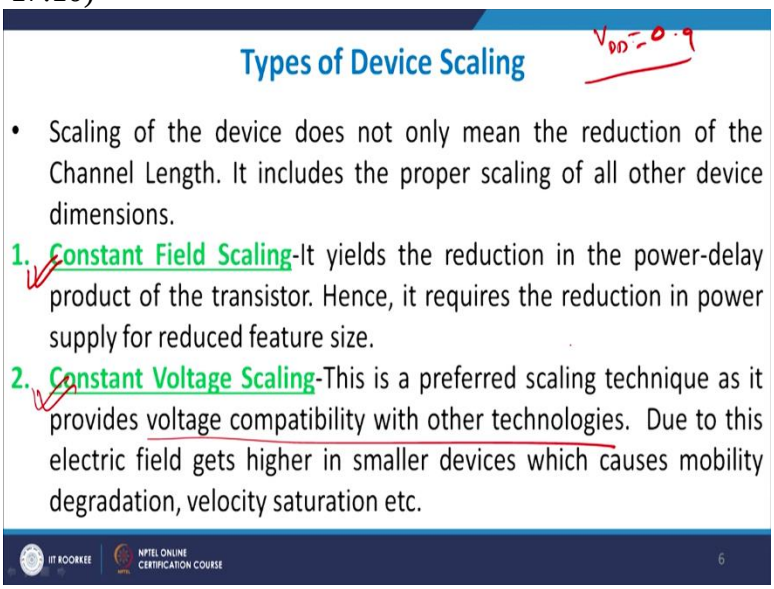
And the rest 1 volt appears across this particular region. Fair. And that is known as $\Delta V_{DS}$ or which is given by $V_{DS}$ - $V_{DS\ sat}$ right? As a result, if you look very carefully therefore that therefore what I will see here is, so I had, so now my depletion, so my inversion layer was somewhere here when we apply $V_{DS}$ say equals to 2 volts. You increase it to 3 volts, the inversion layer further comes down. It was initially here, now it is here, which means that the effective channel length is actually reducing.

It was initially, initially it was this much right. Now it became this much right. And now further increasing $V_{DS}$ became this much. Fine, which means therefore that current which is proportional to W by L right, so L is discussing with increasing $V_{DS}$ which means that the current is rising and therefore assuming that current is constant in the saturation region is a wrong assumption and therefore the current will actually show an increase. So I will show an increase.

This increase is primarily because of an increase in $V_{DS}$ and reduction in the effective channel length. That is what is shown here and therefore the drain current is approximated as this value. Remember, if you remember the previous talk, it is $\mu_n C_{ox}$ W by 2L into $(V_{GS} - V_{TH})^2$. Even extra term here which is $1 + \lambda V_{DS}$, right? $\lambda$ is defined as the CLM parameter or channel length modulation parameter. It is an empirical constant varying from 0 to 1 and $V_{DS}$ is the applied voltage which you see.

So if λ is equal to say 0.5, then for every increase in $V_{DS}$, I would see that 50 percent of $V_{DS}$ will be responsible for increase in the current at higher values of drain voltage. And therefore, from the IV characteristics of the MOSFET, what thing becomes clear is that the current will increase with increasing voltage. Right? The current will go on increasing with increase in voltage; otherwise we are not allowing to do that because we are making it fixed value of voltage, right? So that is what is known as channel length modulation effect.

(Refer Slide Time: 17:10)




Now coming back to therefore to the basic concept of device scaling, we have two types of device scaling. The scaling will be one is known as the constant field scaling or electric field scaling and we have constant voltage scaling which is available with us. Now scaling does not only mean scaling the channel length. When you scale, when you do a structural scaling, when you scale up or scale down, you scale up everything. So you scale down channel length, you also scale down the oxide thickness, you scale down the metal contact, you scale down the base, the source contact and the drain contact.

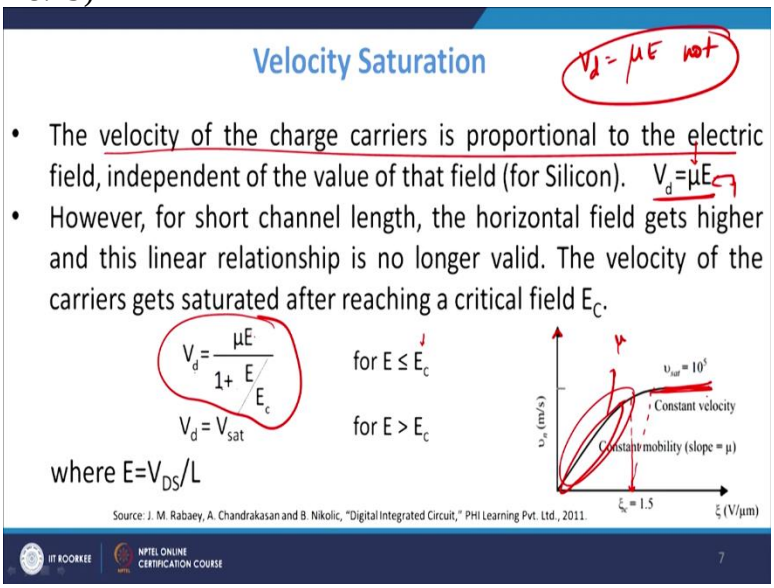
So everything has been scaled down or scaled up, right. So it is not just this channel length which you are doing it together. Now the constant field scaling primarily means that in this case we assume, what we do is that as we scale down our dimensions, we also take into consideration the electric field remains constant. Right? To do that, you have to scale down your voltages in a

straightaway fashion so that V by L is constant. Right? V by channel length or V by dimensions is almost constant. What is constant voltage scaling?

Well, this is a preferred scaling technique as it provides voltage compatibility with other technologies. Due to this the cost you pay for it is that you start, in the 1st case whereas when you lower your channel, you also lower your $V_{DD}$. In this case, you do not do that. You keep your voltage constant right. So when you keep your voltage constant then the electric field enhancement is there, mobility reduction is there and so on and so forth. But then, it is compatible with technology.

So if you want to work with a 90 nanometer technology, if I use a $V_{DD}$ of 0.9 volt, when I do not have to do any scaling and it will work, give me an optimized result available to me. Right? So for a single amount of scaling, this gives me a very well, a well-defined result as far as this device is concerned.

(Refer Slide Time: 18:49)



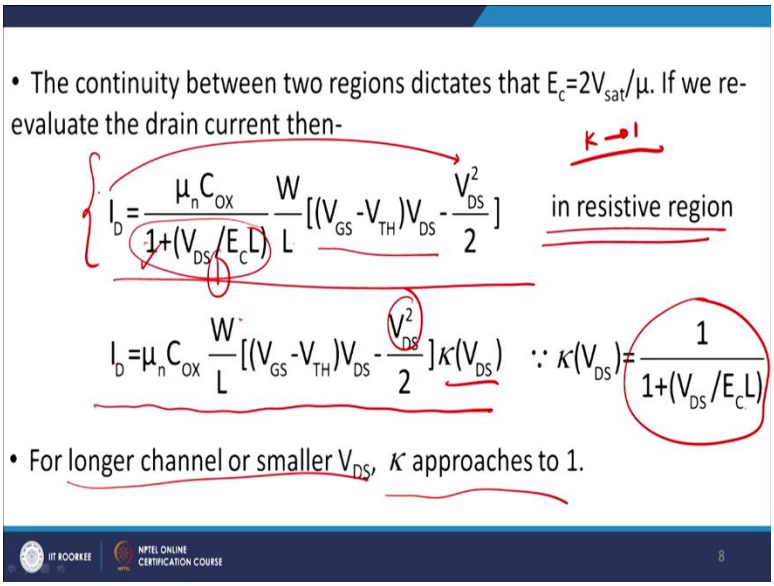

We come to the velocity saturation but as we have discussed earlier, that the velocity of the charge carriers is proportional to the electric field. So $V_{Ddrift}$, this $V_D$ is known as the drift velocity, is equals to μ, mobility of the charge carriers multiplied by the electric field. Right? However for a short channel device, the electric field in the longitudinal direction and the horizontal direction and transverse direction are both equally high. Right. And therefore this linear relationship of $V_D$ equals to μ times E is not valid anymore, is not valid.

Right? Why? Because now the electric field is not only in this direction very strong, it is also very strong in this direction. So under the influence of these 2, you will have a new value of $V_D$ and that value of $V_D$ is given by this formula which you see in front of you, given as μ into E upon 1 - E less than $E_C$, where $E_C$ is the critical electric field. So critical electric field is defined as that electric field after which the mobility is constant or after which the velocity is constant. So if you plot electric field velocity versus electric field for silicon, it looks something like this.

So this much point, since V versus E is always constant or maybe this much point right, if this is almost linear profile, so mobility is constant. Beyond this, the mobility starts to decrease and beyond a particular point, the mobility is actually very low or very small. And $V_{sat}$ is typically very large. $V_{sat}$ is the saturation velocity which you see. Now therefore, you see that your $V_D$ here is equals to $V_{sat}$ right? And that is understandably so that the applied drain voltage is almost near the saturation voltage and we get a very large value and the increase in velocity in case of, so rather than the pinch off of the depletion region, I would actually get a pinch off in the velocity of the charge carriers.

That will result in almost a saturated drain current for a MOS device for short channel effects because the electric fields are very very high right.

(Refer Slide Time: 20:55)

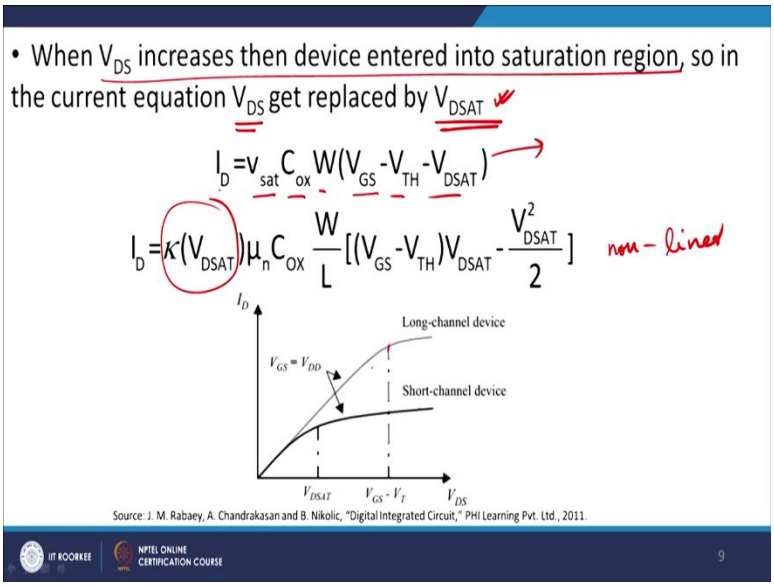



Now if you plot it, then we get, if we evaluate it, we get something like this that the drain current $I_D$ is given by this basic formula, if you remember yesterday's talk, it was basically 1 + $V_{DS}$ by $E_C L$ and this will give you the value, W by L is coming here and $V_{GS}$ - $V_{TH}$ into $V_{DS}$ and $V_{DS}^2$. So this is for the resistive region. And therefore you see, $I_D$ has a non-linear dependency on $V_{DS}$ because it is a square term which you see in front of you. Similarly, if you look at this curve, again, $I_D$ has got a non-linear dependency on $V_{DS}$ right? Where K $V_{DS}$ is given by this one formula.

So this formula is nothing but this whole thing. Right? So I define K to be equals to 1 by 1 plus this thing and 1 + $V_{DS}$ by $E_C$ into L. $E_C$ is a critical electric field, L is the channel length right. Now for, actually what is happening is for long channel lengths, K will approach to 1. So 1 plus 1 is equal to 2 and therefore you reach the conventional equation which we have already studied earlier right?

That $\mu_n$ C oxide by 2 WL into $(V_{GS} - V_{TH})^2$, right. So that is what we reach if provided K approaches to 1 when you have a longer $V_{DS}$, when you have a longer channel length or a smaller $V_{DS}$. And in those cases, we reach to this value in this case.

(Refer Slide Time: 22:16)

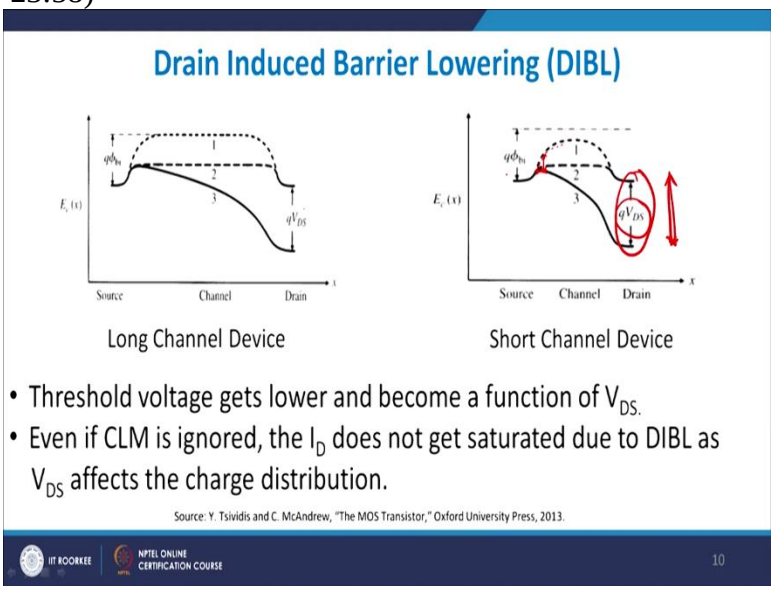



When $V_{DS}$ increases therefore as you can see the device enters into the saturation region. We have already discussed this point. So what we do is that we replace $V_{DS}$ by $V_{DS\ sat}$. Right? So whenever you are sure that the device has entered into saturation region, then the drain to source voltage which you apply is rather than writing it as $V_{DS}$, we write it as $V_{D\ sat}$. Right? And that is $V_{D\ sat}$. $V_{D\ sat}$ means saturation velocity drain voltage.

So I get $I_D$ equals to $V_{sat}$ into C oxide W into $V_{GS}$ - $V_{TH}$ - $I_{D\ sat}$. So this $V_{sat}$ is again broken down into K $V_{D\ sat}$ and therefore I get this into $V_{DS}$. So this is for the non-linear region right, for the non-linear region and this is for linear region of the operation of the device. Right. Which comes to the final expression therefore that if we have a long channel device and if you have a short tunnel device, automatically the short channel, so long channel device saturation will take place at a much higher $V_{D\ sat}$ whereas for a short channel device, it will take at a much lower $V_{Dsat}$.

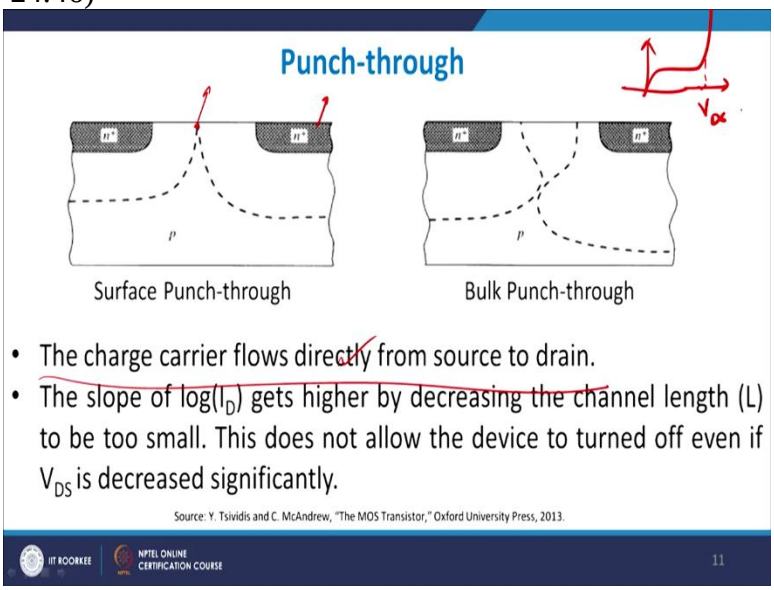
Because short channel, you require lesser amount of drain voltage to pinch off the carriers and therefore it is much smaller as compared to the first case. We come to another important topic and that is basically drain induced barrier lowering. At this stage, we will not go any further than saying that initially when drain and source are very far from each other, drain has got no influence over the source but what is happening is, as you go on bringing drain and source closer to each other, then the change in the drain voltage will also start influencing the source side.

(Refer Slide Time: 23:58)




So as you can see here that as you make your drain voltage higher, the channel length, the conduction band diagram, Q into $V_{DS}$ there is a drop of voltage, right? Because into electron, you multiply. So there is a drop in the voltage, which means that you actually saw a larger current but now you actually see a dip. So there is always a decrease in the hump. As a result what happens is that that it affects and lowers the threshold voltage. The reason being, your built-in voltage is reduced. Right? Because of DIBL or DIBL effect. As a result, your charge carriers are more and more and current becomes more and more right and it affects the charge distribution model for this case.

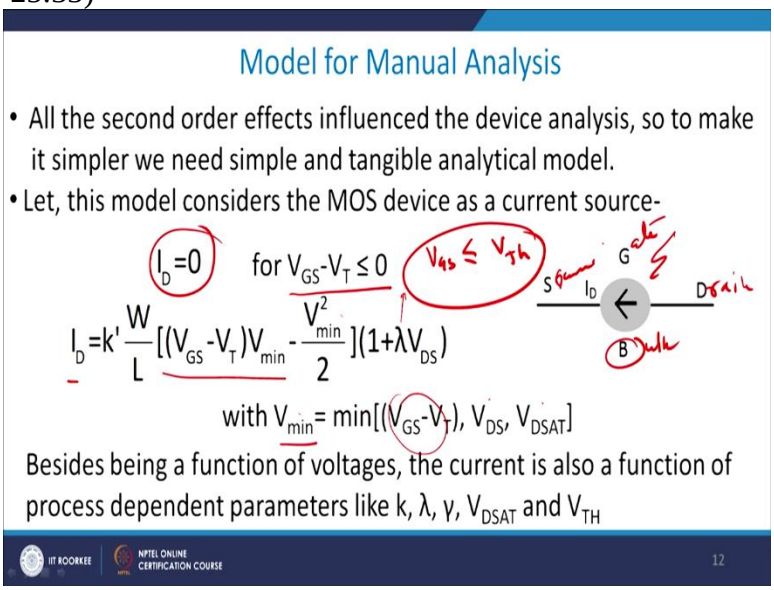
(Refer Slide Time: 24:46)




We come to the last part that is basically your punch through. What happens in a punch through is that if you go on increasing, as you can see, as you go on increasing the drain voltage, this depletion region touches each other, right? And this condition is known as punch through right. So charge carriers directly flow from source to drain. So you do not require the channel right? The drain voltage is so high that any electron can be pulled away from the source and it can pass through the depletion thickness and reach the channel and that is known as the punch through phenomena.

For most practical purposes, we do not use punch through phenomena in any case right. So what we do is that if you increase the value of $V_{DS}$ significantly then this will happen. But in most of the cases, so therefore if you plot $I_D$ versus $V_D$ then it will be something like this and then suddenly it becomes large for $V_{DS}$ somewhere here, it becomes large. We do not operate generally in this one but this is known as punch through which you see.

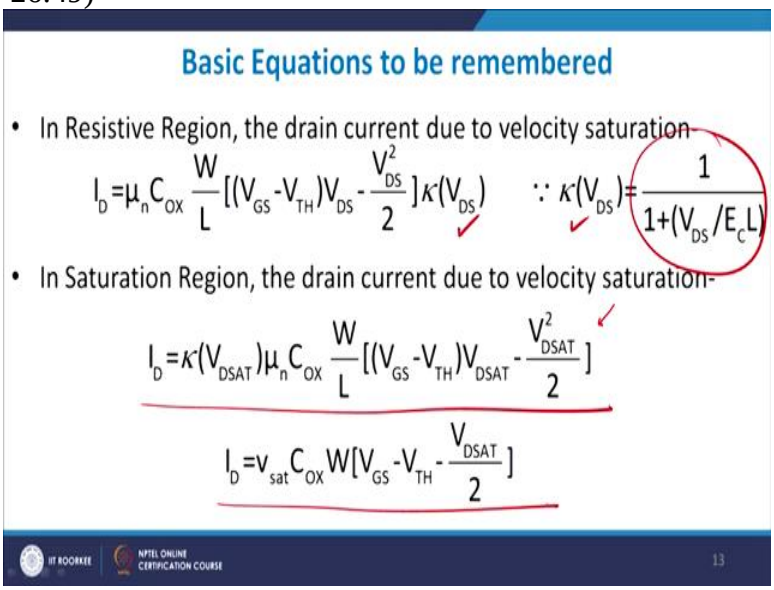
(Refer Slide Time: 25:33)




Now therefore the model which you apply for manual analysis which means that the model which you will be or second-order effects can be taken care of by a simple tangible analytical model and the model is that when $V_{GS}$ - $V_{TH}$ is less than 0 or $V_{GS}$ is less than equals to $V_{TH}$ for an n channel enhancement mode MOSFET, the current through the MOSFET is 0 and whenever my input voltage is larger than threshold voltage, then I get $I_D$ to be equals to K' W by L $V_{GS}$ - $V_{TH}$ $V_{min}$ - $V_{min}^2$ by 2 into 1 + $\lambda V_{DS}$ right.

$V_{min}$ is basically $V_{DS}$, minimum value of $V_{DS}$ and $\lambda$ is the CLM parameter which you see and $V_{DS}$ is the drain to source voltage which you see. So V min is basically minimum of the overdrive plus $V_{DS}$ and $V_{D\,sat}$ right? Out of the 3, whichever is the really the smallest one, we assume that to be equals to value of $V_{min}$ but typically, it is $V_{DS}$ which comes into picture. And it is replaced by a current source as you can see with a current $I_D$ and the source, drain, bulk and a gate. So this is bulk right, this is drain, and this is gate and this is source and this is the drain current which is flowing through the device.

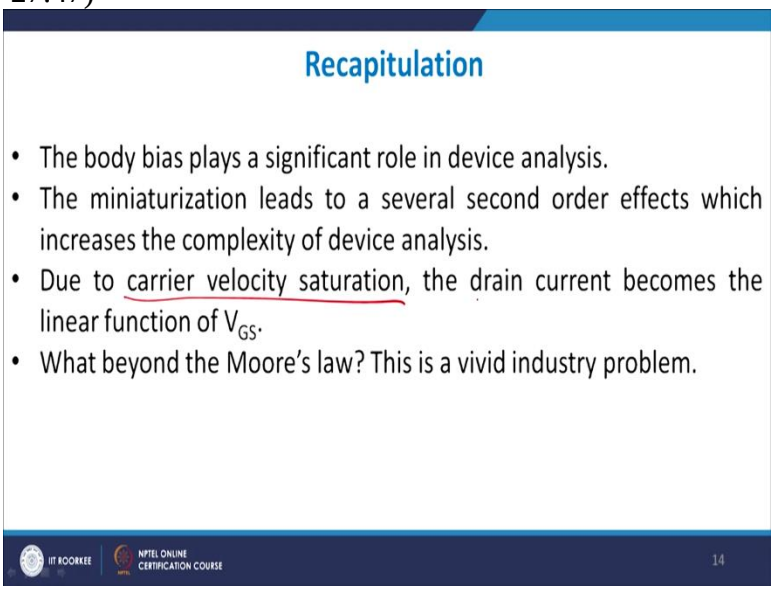
(Refer Slide Time: 26:49)




So let me see what basic equations you need to remember as far as this lecture is concerned. The basic equation you need to remember is this one where K $V_{DS}$ is given by this. This is also known as velocity saturation equation. Right? In the saturation region, as I discussed with you, the drain current is due to velocity saturation, which is therefore given by $V_{DS\ sat}$ and $I_D$ equals to $V_{sat}$ into C oxide and W by L $V_{GS}$ -$V_{TH}$ - $V_{DS\ sat}$ by 2. So once you know the value of $V_{DSsat}$, you can find the value of $I_D$ in a much more detailed manner, right?

And that gives you an idea about the various basic equations you need to follow. So equations are exactly the same as you follow for a bulk MOSFET, small changes but when you do a short channel effect it is primarily because of velocity saturation. Whereas for long channel effect, it is primarily to do with pinch off at the drain side, right. And therefore these 2 are totally different issues which you need to find out. Let me therefore recapitulate today's lecture in detailed manner which we have collected.

(Refer Slide Time: 27:47)




We have understood what is basically a MOSFET technology, we have also understood that the body bias, which is the third terminal body bias, fourth terminal body bias plays an important role in determining the change in the threshold voltage of the device. Now, we have also understood that one of the important effects of miniaturization is the drain and source coming closer to each other and therefore you will have drain induced barrier lowering which means that the voltage change on the drain side will influence or will reduce the barrier on the source side because source side if you remember is basically a depletion width is there.

Now if I am able to somehow or other reduce the depletion width by giving an external bias and this case happens to be the bias given to the drain side and the drain is very close to the source side, so any variation in the drain side is replicated on the source aside. Then we are able to reduce the dimensions or we reduce the barrier on the drain side and thereby changing the threshold voltage of the device. Now the third point is due to carrier velocity saturation, I get the drain current is almost a linear function of $V_{GS}$ right.

So it is $(V_{GS} - V_{TH})^2$ and therefore it is a linear function of $V_{GS}$ right? Well. Beyond Moore's law, we do not know what will happen. There are many alternatives which are coming out but these alternatives need to be tested in terms of its performance analysis and more importantly, these structures should be able to be, can be fabricated in the current foundries which are available to us, right. So this is the basic idea or the basic concept or the MOS device second-

order effects, right. So we have dealt with the basic concept and we have understood the basic idea of this phenomenon. With these, let me thank you for your patient hearing and I will be open for any questioning. Thank you!

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−17**
**MOS Parasitics and SPICE Model**

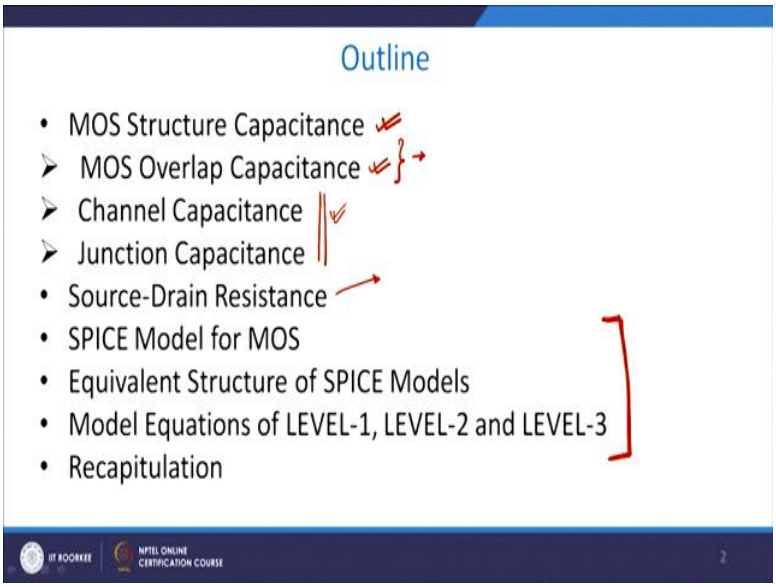
(Refer Slide Time: 0:26)


Hello and welcome to the NPTEL online course on Microelectronics: Devices to Circuit, today we will take up what is known as the MOS parasitics and SPICE model, the name of the today's module is MOS parasitics and SPICE module or SPICE model. In our previous discussion or interaction we had seen the second-order effects which are related to MOS devices, which primary means that what do you mean by drain induced barrier lowering, punch through. We have also seen the substrate effect or a body bias effect. We have also understood what is the meaning of velocity saturation, under what limits of operation do we have second-order effects coming into picture, what is the reason why the second-order effects coming into picture, what are short channel effects, how does threshold voltage gets affected by the short channel effects.

Why this is important to study? Because we move ahead in our course we will see that these influence the output electrical characteristics of the device, and as a result, the circuit get influenced by these variations. Now another important variation of a device or important parameter of the device is its parasitic, which means that those parasitic resistances and capacitances which the device has already inbuilt into it, but they appeared to you at certain frequencies of operation, thereby limiting the usage of the MOS devices at those frequencies.

So in order to incorporate all those things and understand the basic concepts of MOS device and therefore give you a basic feel of basic SPICE models which are available to us, let me explain to you therefore the outline of this current talk as far as MOS device is concerned.

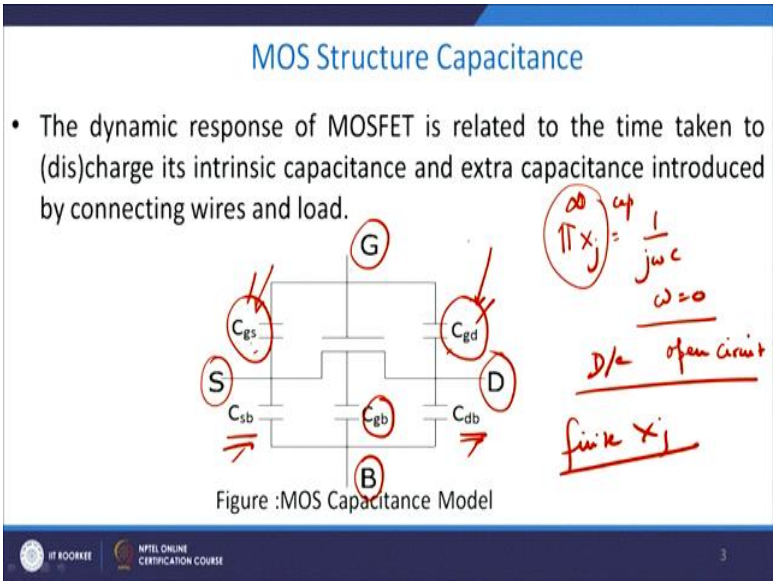
(Refer Slide Time: 2:24)

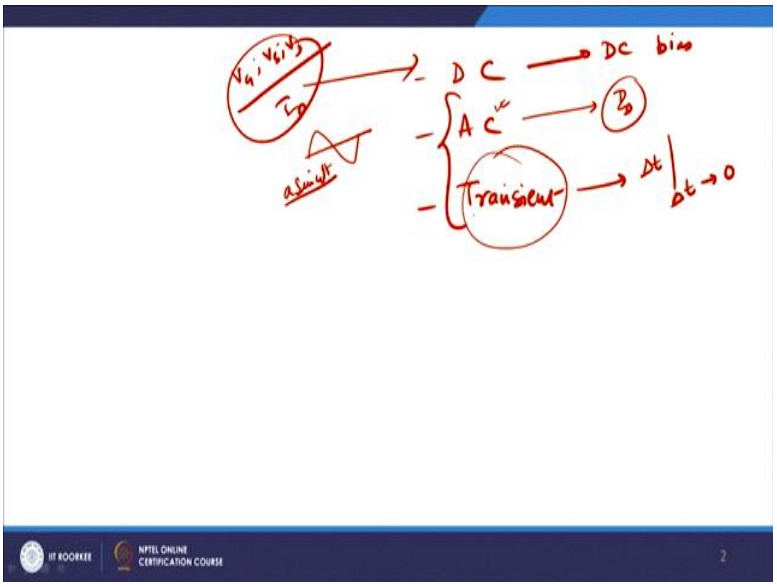


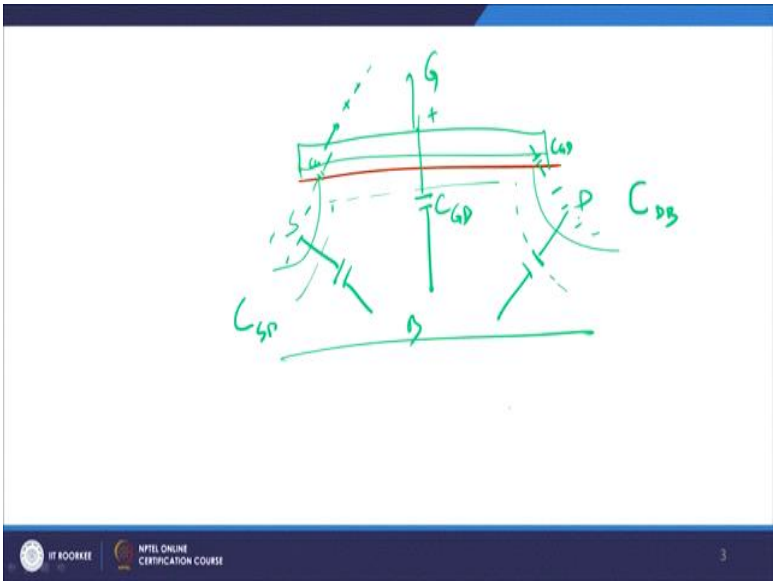

We will look at the MOS structural capacitance, which means that we will look into the fact that how a MOSFET structure, the very fact that it is structurally designed will give me various capacitance values associated with it, those things will be clear to you. We will look at MOS therefore overlap capacitances, various overlap capacitances, we will be looking at junction capacitance and channel capacitance right, these capacitances which is basically MOS overlap is primarily because of problems with fabrication and so on and so forth, but channel capacitance and junction capacitances will always arise even with one of the best fabrication facilities.

The meaning of source-drain resistance should be clear to you as a designer and then we will be entering into what is known as the SPICE model for MOS devices right, SPICE is basically a circuit simulator, which is being used quite often by large number of people across the world, most of the versions of SPICE are open source and can be downloaded from the Net and can be used by you and therefore understanding that within the SPICE how does the MOS work, what is the equivalent circuit model for a MOS device within the SPICE should be understood. And that is what we will be taking up at these MOS devices, equivalent circuit models and then we will look at level 1, 2 and 3 of model equations of MOS device right and see how, so each level will actually refine the drain-current characteristics, so you should be able to understand the various level and how it works in terms of its working principles.

(Refer Slide Time: 4:03)






Let me now come to MOS, MOS structure and its capacitance. Now please understand when you are working with a DC bias or when you have applied a fixed value of voltage either to the gauge, drain or source terminal then of course w is equal to 0, and therefore or the frequency operation is 0, and $X_J$ if you remember for a capacitance is 1 upon JwC and therefore $X_J$ is infinitely large for w equals to 0. So for a DC bias, for a DC I can afford to basically let them behave like open circuit right, like an open circuit. So when you are dealing with DC biases any capacitances within any other networks right, any electrical network can be made as open circuit because the concerned impedances are infinitely large and therefore you open it right.

When you open it fine then you do not have any other place to go and therefore there will be a current flow and this current flow will depend upon the network parameters. Whereas when you have a variable in the frequency which means you are working in the AC domain and the frequency of the AC signal is varied from some X to Y value it is varied, then this $X_J$ which is the capacitive impedance right and assuming it to be infinity when you are applying a DC that does not hold good.

So you will have now a finite right, $X_J$ or a finite impedance and as a result, we have to replace this by sort of a resistive element which is basically frequency dependent resistive element and that is the problem area for MOS transistor, so when MOS device you are working in a DC domain and try to find out the I V characteristics, well does not matter because you are buying the value of DC and for all DC this is open circuitry and all the discussion which we had for the previous few interactions will hold good.

But let us suppose that you are taking AC analysis or you are doing a transient analysis right, there are three types of analysis which we generally follow: one is known as the DC analysis, we have a AC analysis and we have got transient analysis right. DC is when you apply a DC biases, so which means that is not a time varying bias, it is a fixed bias and you give a bias, and try to find out a current at those fixed biases. So you fix the value of $V_G$, $V_S$, $V_D$ and then try to find out the value of $I_D$ and that is what is known as the DC bias.

What is AC bias or AC analysis? AC analysis means you give an input signal, which is not DC but suppose a sine wave, let us suppose a sinwT and for these values of a sinwT, you try to find out the $I_D$ value at the output side right and you try to find out what the $I_D$ value is that is what is AC analysis.

What is transient analysis? Transient analysis is when you do the analysis for ΔT and T tending to, ΔT tending to 0, which means that for very short interval of time you try to find out the value of voltage or current provided the applied biases are varying from, you know in a very, very fast sense and you are able to calculate the variations in the value of drain current for those small values right. So for all these, for last these, for these two we will actually will be doing a large amount of capacitive analysis.

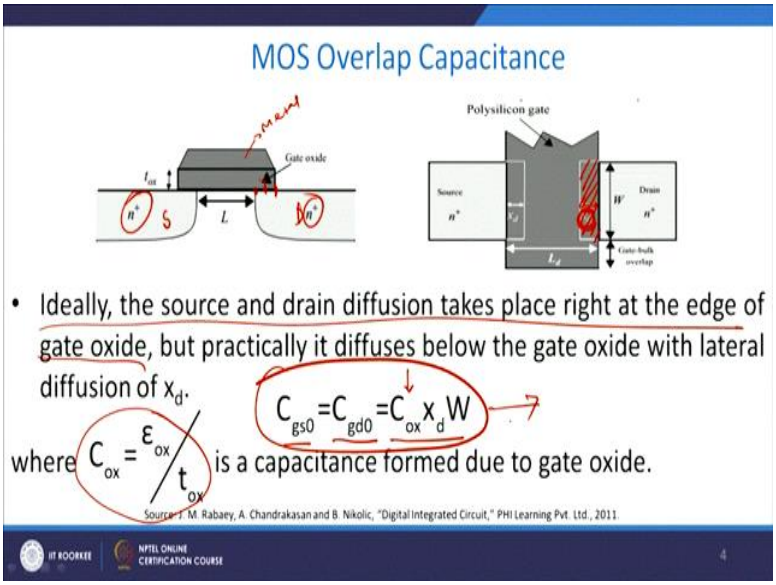
So you see here that, if you look at the MOS device right and if you have the, these are basically known as this and this are basically $C_{GS}$ gate to source and $C_{GD}$, so you have gate terminal here, you have bulk, you have drain and you have source, so this a four terminal device. Now as we, so if you go back to your previous discussion, from a previous discussion knowledge you will understand that if you have a, sorry let me help you out in the sense that, let me say that you have a device and your device is something like this that your source is here and drain is here, so I have a source, drain right. I have a gate here and then I have a metal gate here right and I have bulk here and this is gate.

So you see this is high charge, this is also negatively charged capacitance, this is negatively charged, so which means that between these two I can have a $C_{GS}$, C gate to source, similarly between this two I will have $C_{GD}$ gate to drain, these are the overlap capacitances right. You also have a capacitance which is, so this is $C_{GS}$ right and this is $C_{GD}$, gate to source and gate to drain, you also have a drain to bulk which means that let me see, so drain right, you will have a depletion region here and you will have a depletion region here right, so you will always have a drain to bulk capacitance here and source to bulk capacitance here. So this is

$C_{DB}$ and this is $C_{SB}$ or BS right, whatever you want to say, so this is $C_{SB}$ and $C_{DB}$ which you see.

So this is overlap capacitances, these are structural capacitances, which is available to you. You also have gate to bulk capacitances, what is gate to bulk? Because you will have gate potential here and bulk potential here, so the capacitance available at this point is given by $C_{GB}$, so effectively there are five capacitances which influence the AC or the transient analysis of a MOS device right and these 5 capacitances are mentioned in this figure.

(Refer Slide Time: 9:33)



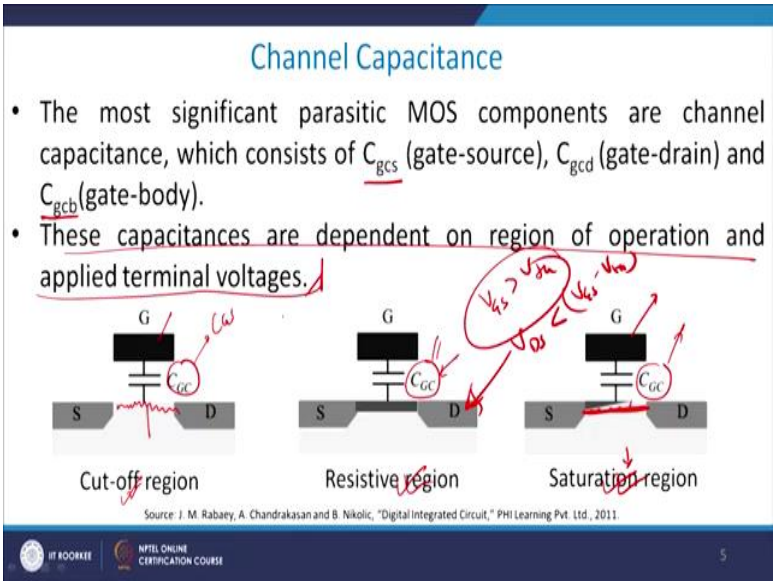

As I discussed with you earlier, so when you have source and drain and they are overlapping through the gate oxide, so this the metal gate right, this is the metal gate and this is the gate oxide. This is channel length and this is the t oxide value. As I discussed with you ideally, the source and drain takes place at the edge of the gate oxide, ideally this should be at this edge right, but practically there is some diffusion right, why? Because see you do this doping here $N^+$ region and $N^+$ region by a diffusion phenomena.

So you have a diffusion phenomenna and diffusion is, if you remember is by the by very nature, you cannot stop at a particular point it will have some amount of out diffusivity, so the diffusion will result in what? The diffusion layer being extended towards the each other right, and as a result, you are supposed to stop somewhere here, you added up, you ended up somewhere here. Fine, and as a result, you will have some capacitance which is available at this particular point.

So if there is lateral diffusion and because of this, you have an extra $X_D$ as the diffusion, lateral diffusion length, then we define $C_{GSO}$ and $C_{GDO}$ as C oxide into $X_D$ into W, I suppose you can understand why. Because $X_D$ is this, W is basically this, this dimension W, so W into $X_D$ is the area under the overlap region, this multiplied by area per unit capacitance of C oxide which is over it will give you the value of $C_{GSO}$ and $C_{GDO}$ right, $C_{GSO}$ and $C_{GDO}$ means gate to drain and gate to source voltage and this is known as gate to bulk overlap and this is gate sorry, and this is gate to drain overlap and this is gate to source overlap right.

C oxide is given by Epsilon oxide by t oxide which you see and therefore this is the gate oxide with lateral diffusion when $X_D$ is available. So, as you can see higher the value of lateral diffusion more will be the value of $X_D$ and as a result your $C_{GSO}$ and $C_{GDO}$ will be also larger in that case, which means that their influence over the transient analysis will also be pretty heavy in nature or pretty large in number.

(Refer Slide Time: 11:35)




Now let us come to the channel capacitances, now channel capacitance as I told to you will have again two components primarily. We have understood $C_{GS}$, so it is basically gate to source, so it is gate to sorry, this is $C_G$ gate, channel source, you have gate drain, so gate drain and then gate body, so this is gate to channel, this is gate to channel and this is also gate to channel for three regions of operation for cut-off, for resistive region and for saturation region right.

As we discussed cut-off means when gate voltage is almost 0, gate voltage is almost less than threshold voltage. At that point of time you do not have any inversion layer available with

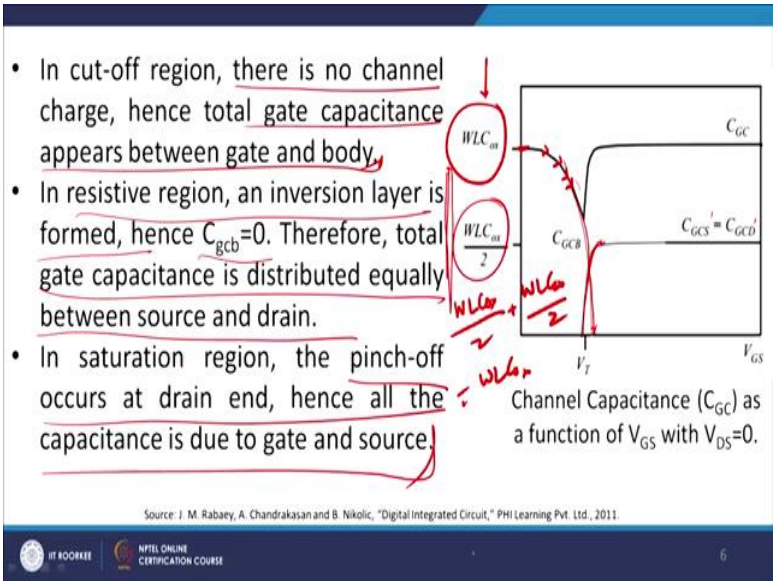
you right, you do not have any inversion layer at this particular point right, but you might have some charges here and, so this charge and this charge separated by an oxide layer will give rise to $C_{GC}$, $C_{GC}$ which is basically your $C_{GC}$ this one and then you get $C_{GC}$ because when you go to resistive region and your gate voltage is just greater than $V_{GS}$ but $V_{DS}$ is less than $V_{GS}$ -$V_{TH}$, this is this region on operation right.

So you will have thin-layer available to me, but it is still not inverted and fully inverted. In that case you will have $C_{GC}$ and now when you have saturation when drain to source voltage is so large that it eats away into the channel and therefore your affected channel reduces, you will have a new value $C_{GC}$. Now what people have done over the years is just for the sake of making it easy to understand they have divided this $C_{GC}$ into two parts, one is towards the source end, one is towards the drain end and then they say that if the doping concentrations in the source and the source and the drain is almost the same I would expect to see that the $C_{GC}$ on the source side and $C_{GC}$ on the drain side to be almost equal to each other.

One important point which you should take care of is that when you have inversions, strong inversion, then in saturation region, for example you have a large number of charge carriers here right and they act as a screening, sort of a screening charge screening mechanism, so they tend to screen the charge of the gate from the bulk region right, so that is quite interesting and therefore you will see that the gate to bulk actually falls down, capacitance falls down at a very, very high value of $V_{DS}$ because the inversion layer which is there in the middle between at the silicon-silicon dioxide interface, they tend to screen out the gate and the bulk charge and therefore you do have a drop in the value of $C_{GS}$ right.

As I discussed with you therefore these capacitances are dependent on the region of operation and also of the applied potential terminals right, so how much potential you are applying at various points, they are depending on that and then whether your operating in the cut-off region, whether you are operating in the resistive region or whether you operate in saturation region will also be important right.

(Refer Slide Time: 14:35)




Okay, in the cut-off region as I was discussing there is no channel charge right, hence the total gate capacitance appears between gate and body, I thought this is perfectly clear and very easy to understand but when you do not apply any, when you cut-off there is no inversion layer and gate to body and gate to channel are exactly the same because channel in any case you do not have any free charge carriers right.

Now if you therefore plot VGS versus capacitance here and we will see how it works out, when your gate voltage is very, very low or in a cut-off region you WLCox, Cox is basically oxide capacitance per unit area multiplied by area gives you the total capacitance, so the total capacitance is something like this. As you go on increasing the value of gate voltage, since more and more charge carriers starts to form, the screening takes place and the capacitance starts to fall down, so you have a reduction in capacitance at this particular point.

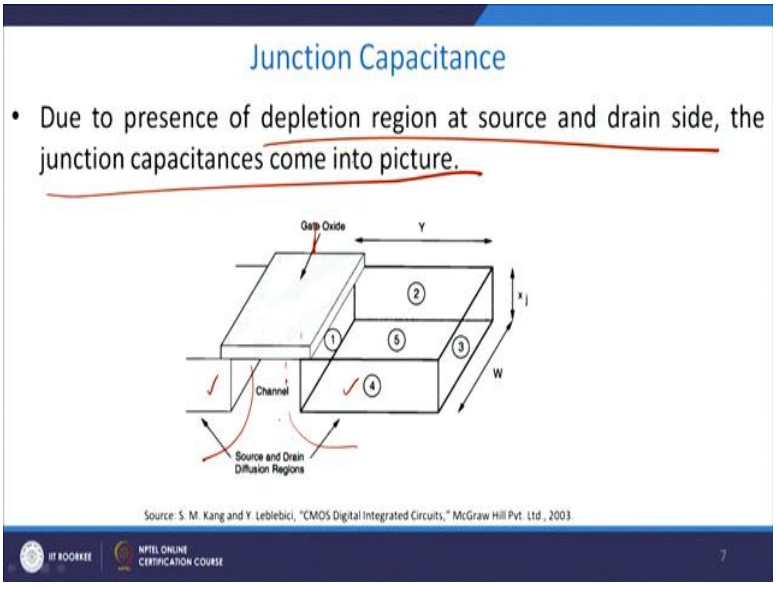
In the resistive region, an inversion layer is formed right, hence $C_{GCB}$ or bulk is 0. Therefore the total gate capacitance is distributed equally between source and drain right. So let us suppose you had 10 femtofarad as a total gate to bulk, bulk was to 0 and therefore this breaks down into 5 femtofarad, one goes to the source side another goes to the drain side and that is what is basically found out in this region, that you will have equal distribution between source and drain right. You will have equal distribution between source and drain and this will result in equal charging on the source and drain side right and they will be almost equal to each other.

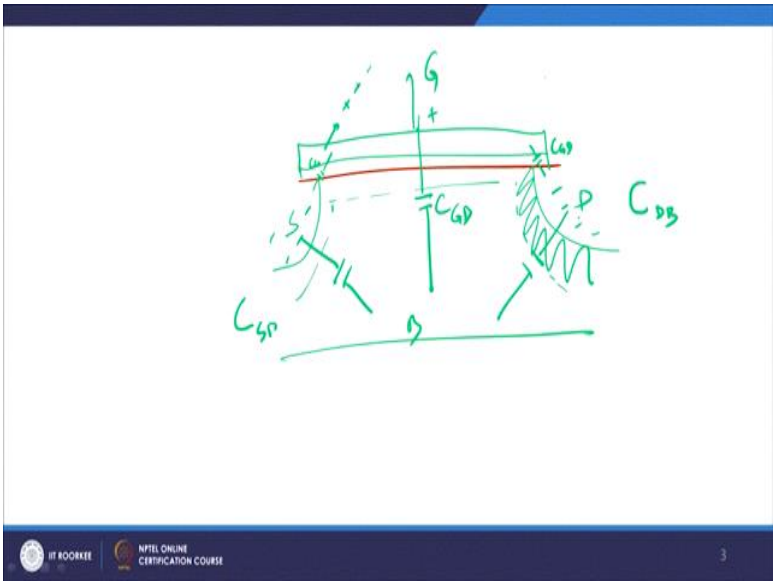
Now in saturation, since the pinch off occurs at the drain end, hence all the capacitance is due to gate and source, I think this is pretty clear because when you pinch off on the drain side and there are no charge carriers there because it all depleted, all depleted of free charge carriers, so again, the all, it is being shared by the gate and the source, so it goes, if you look very carefully the total charge we will see later on. But, if you look very carefully below $V_T$ the contribution is gate to channel bulk as I discussed with you, but it falls down, see its falls down and it goes something like this right, because screening effect is there.

Beyond $V_T$ gate to channel for source and gate to channel for drain it looks something like this, it becomes larger and larger right, but as the gate voltage becomes larger and almost fixed, for a fixed value of a gate voltage, I get that the gate to source voltage and gate to, gate to channel for source contribution and drain contribution is equal and therefore you see it is given as WLCox by 2, so half, so half is supported by, so you are initially starting with WLC oxide, you ended, in saturation region you ended having the maximum value of each one of them to be WLCox by 2. Since they are in parallel, therefore I can safely write down plus WLCox by 2 as the total capacitance which comes out to be WLC oxide.

So you see, which was initially the value, which was here? So which means that when the devices were switched off, the whole capacitance was taken care of by gate to channel, as it switches on and goes to saturation the same capacitance get distributed between the source end and the drain end right. So that is what we have learned from all discussion so far, as far as this device is concerned.

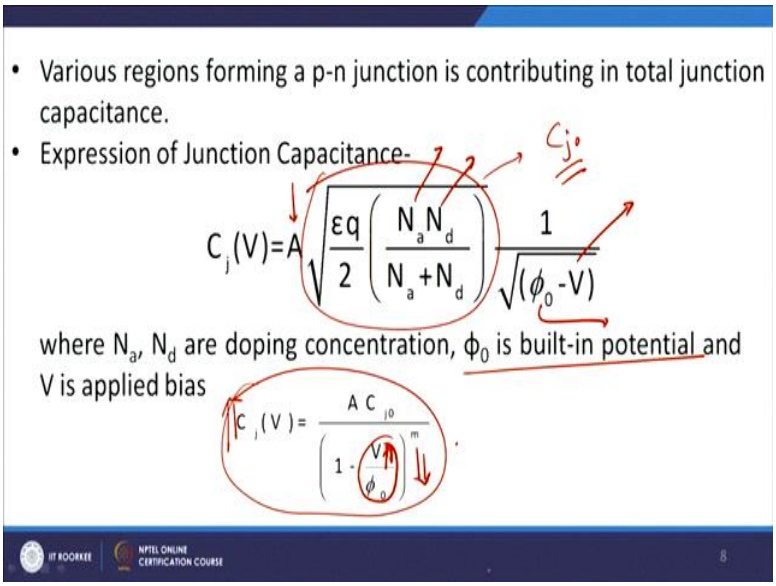
(Refer Slide Time: 17:53)






Now what is the junction capacitance? Now due to the presence of depletion region at the source and drain end, the junction capacitance comes into picture; this is very straightforward and simple way of looking at it. So, you remember there was a depletion region here, this will give rise to junction capacitance, so I think that is very fair story which you see. So if you look very carefully this is the source and this is the drain diffusion region and you have a gate region here and so you will have a depletion region somewhere here and a depletion region somewhere here, this will give rise to a junction capacitance.

(Refer Slide Time: 18:22)




Of course, maybe I can show you to is, so it is basically a p-n junction which is formed, a p-n junction depletion region right, so if you remember your basic p-n junction theory, we can come to know that the expression for the junction capacitance $C_J$ is given as A times epsilon

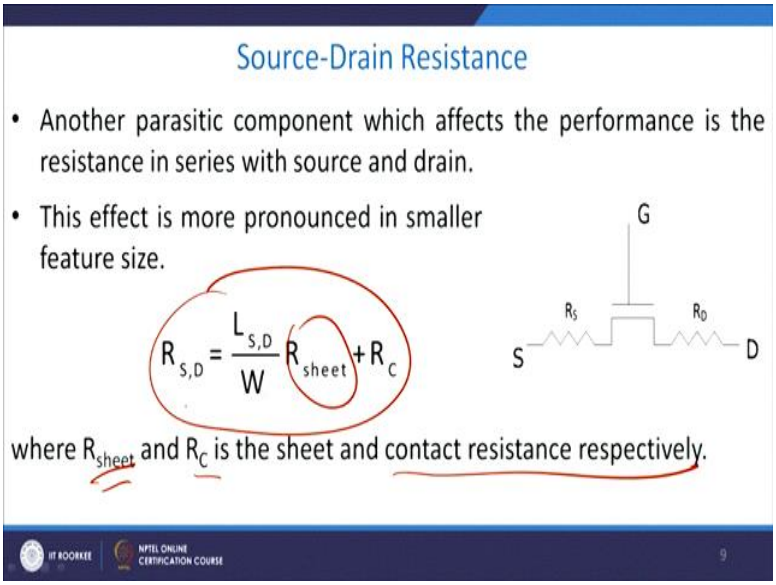
Q by 2, Na, Nd upon Na + Nd root over upon 1 upon this value, where A is the area of the cross-section between the source and the drain side, Na, Nd the acceptor and the donor concentration and V is the applied voltage which you see and $Phi_0$ is basically known as built-in voltage right and V is the applied bias.

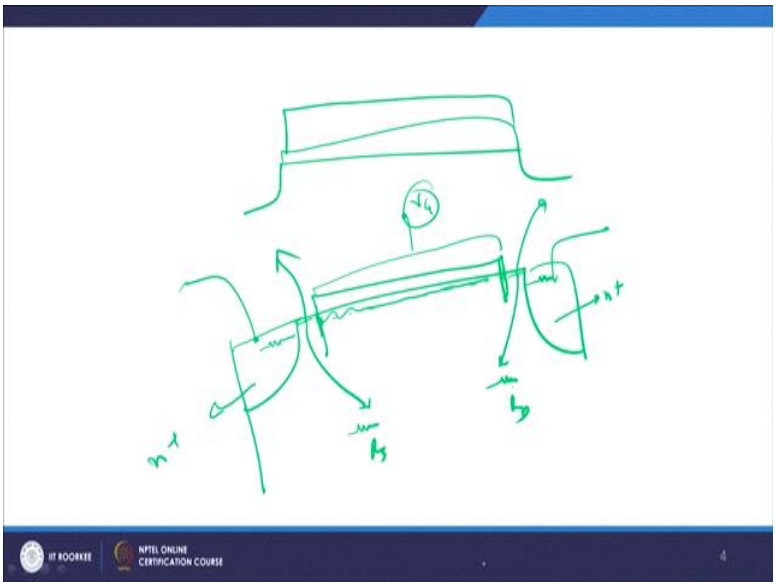
So $C_{JV}$ is basically A times, this factor is basically my $C_{J0}$, $C_{J0}$, which means that under 0 bias the junction capacitance, why? Because even if you do not apply any bias right and you have a differential in doping concentration, there will be always a depletion region formed, howsoever big or small right, that capacitance because of that is basically my $C_{J0}$. Now that gets effected by an external bias when we apply a voltage V, so you can very well see that when you apply a voltage V if you go on increasing this V, this quantity becomes larger and larger, so 1 upon this quantity goes on reducing right, and as a result this whole quantity becomes larger and larger.

So when you increase the value of voltage right and you reverse bias it the depletion thickness becomes larger and larger and you end up having a larger capacitance, which is there with you in case of a p-n junction diode theory as well as for the MOSFET theory right.

Let me come to source and drain resistance, very important part in a parasitic part, so we learned four capacitances with us, actually five, one is gate to source, gate to drain overlap, one is channel, bulk to source, bulk to drain channel, another is gate to channel right. But the point is that all of them do not show their effectiveness at each bias right, as the bias changes the weight of each of the capacitance goes on changing right and therefore it is fair to say that the capacitances are not equal and they are not, they do not show their effect equally for all the conditions of operation of the device, right. So this is what you learned from till now and this is basic learning from a capacitance point of view and these are parasitic capacitance mind you, and therefore they restrict your high-frequency operation of MOS device.

(Refer Slide Time: 20:50)




Now what happens is that when you have a MOS device and you do have a source and gate region and you have a source and drain region available with you and this is your source and, let us suppose these are not overlap devices and I have got source and drain, now then what happens is that or let me make it something like this, something like this, so you will have electric field or channel will be very largely from below these two points, which means that these two points will be still left.

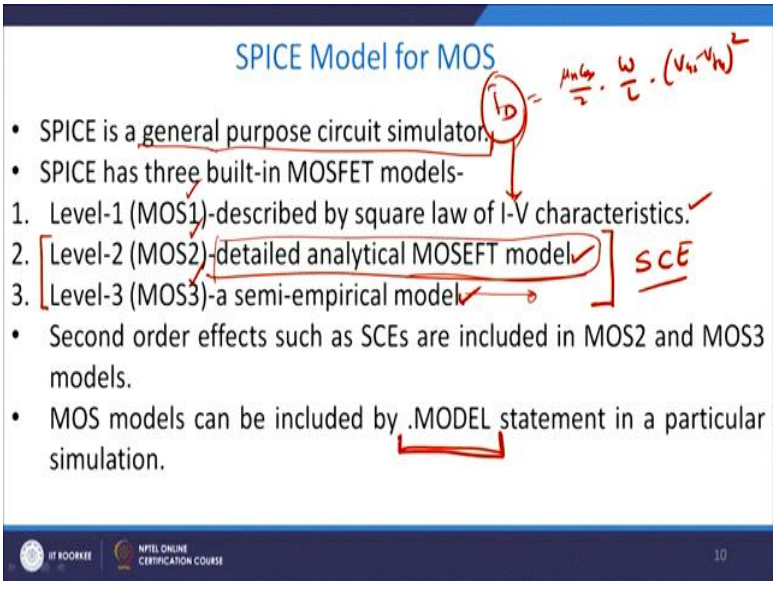
So this is known as source resistance, this is also source resistance, drain resistance $R_D$ and this is $R_S$, this assuming that you will have a contact here and assuming that you will also have a contact here, there will be some finite resistance here, some finite resistance here, even though this is $N^+$ and this is $N^+$, I can safely assume that there will be some finite resistance

howsoever small it is. That plus the $C_B$ resistance gives you the value of $R_S$, this plus this series resistance on the drain side gives you the value of this thing and the reason you get $R_S$, $R_D$ is that prima facia if you look very carefully here is the point where you apply gate voltage.

So the electric field is very large in the transverse direction between these two points, so the number of charge carriers are very large here, but the number of charge carriers here and here are very small because you do not have transverse direction electric field here and therefore the resistance offered by this arm and this arm is relatively very high. And that the reason you will see that you have $R_S$ $R_D$ which is the basically given by this thing.

So $R_S$ $R_D$ is given by $L_S$ $R_D$ and divided by W into R sheet + $R_C$, where $R_C$ is defined as the contact resistances which is basically the resistances between the contact and the semiconductor and R sheet is the sheet resistance of the device that multiplied by $L_{SD}$ by W gives you the total value of $R_{SD}$. You need not worry from where we are getting this formula or how this formula will be used. At this stage, it will be enough for you to understand that there will be some finite resistance between source and drain when the devices working either in saturation or in the non-linear region of operation.

(Refer Slide Time: 23:09)




Let me come to the last part of our talk this time and this is basically the SPICE model for MOS devices and SPICE is basically a general purpose circuit simulator right. It is basically a general-purpose circuit simulator, it was built way back and it is still being used by many

colleges and in many industries also SPICE is being used by, for the purpose of circuit simulation purposes.

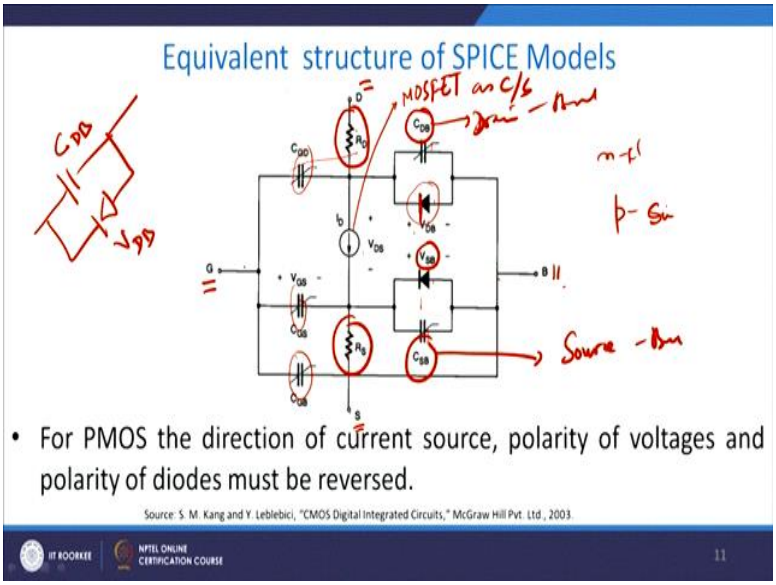
So it is basically a plug and play sort of a schematic entry, which means that you have these devices available with you, you just have to click it and place it on the schematic, schematic is the place, we will discuss this later on, but schematic is the place where you draw the circuit perse. So when you draw, drag-and-drop you actually are picking up the library of the device right, so all the equations related to the device are pulled from the library and placed with the device on the schematic and that is what is an important issue which you should keep in mind.

There are three models or the three levels of models which are available to us and this is level 1, level 2 and level 3 also referred to as MOS1, MOS2 and MOS3 respectively. The MOS1 is basically is on the I-V characteristics of the MOS device, so remember that square law term, where you had got $V_{GS}$ - $V_{TH}$ whole square, so $\mu_n$, Coxide by 2 into W by L into $(V_{GS} - V_{TH})^2$, this is the square law for $I_D$, this is used as first law or MOS1.

A much more detailed analytical model of MOSFET is when used is MOS2 right, MOS2 model or level 2 model right. Another model which is being extracted from experimental data, also referred to semi-empirical model is used as model 3 or MOS3, now $2^{nd}$ and $3^{rd}$, level 2 and level 3 takes care of short channel effects, level 1 is basically a long channel effect and therefore long channel model, which means you have applied gradual channel approximations and you do not have any short channel effect in level 1, but in level 2 and level 3 you have a large amount of short channel effects here right.

The MOS model, when you write a statement in the SPICE, the MOS model is included by putting a .MODEL statement in a particular simulation, so when you put .MODEL and then nMOS, then you write down its values and so on and so forth, you effectively bringing or the model file associated with a particular MOS device onto the schematic level and that is what we are actually looking into as far as SPICE model for MOS is concerned.

(Refer Slide Time: 25:55)




Now if you look at the equivalent circuit of a SPICE model, which means that how a SPICE looks at the MOS device then this is how it looks like its MOS device right and it looks something like this that you will have a drain resistance remember $R_D$, you also have a source resistance $R_S$. Again you have got four terminals: gate, drain, you have bulk and you have source right, so there are four terminals out of which these two takes care of $R_S$ $R_D$, this $I_D$ is nothing but MOSFET behaving like a current source as current source remember.

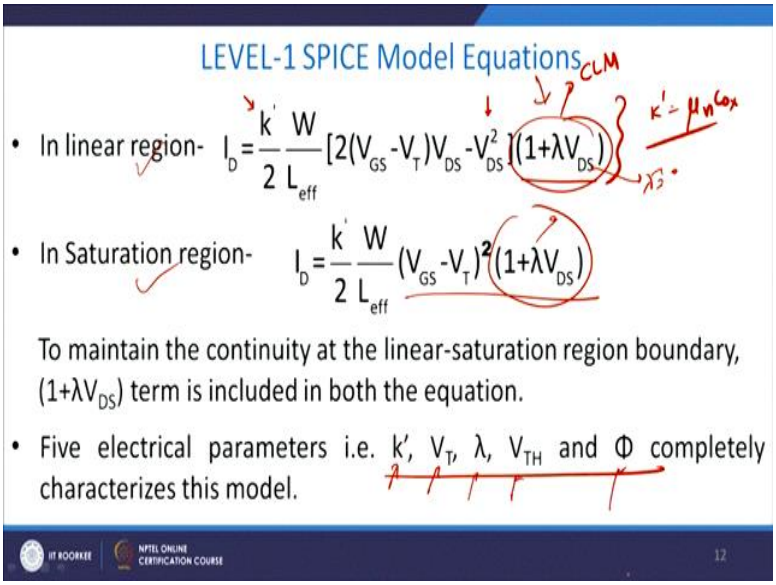
MOSFET is actually a current source, it is a basically a voltage controlled current source and therefore the voltage, where voltage, the gate side voltage is my voltage control current source and this gives you that resistance available here. If you remember $C_{GD}$ gate to drain overlap, you have gate to bulk overlap and gate to source, so gate to source right, gate to bulk right, this is bulk and you have got gate to drain, so gate to drain overlap here right. If you look back on this side bulk and the gate side, remember you were saying that as I was discussing with you, you will always have a junction capacitance between the substrate and the base and that is what is happening right. You also I have between source and bulk sorry, this is source and bulk capacitance, source and bulk capacitance and this is basically a drain and bulk capacitance right.

Remember, please understand that it is basically a p-n junction diode. If you remembered it is basically a p-n junction diode and p-n junction will have P side towards the bulk side, if it is an N channel MOSFET right, if it is an N channel MOSFET because the bulk side is basically P type or P substrate and therefore you will have P side coming to bulk side and you have this towards the N side, so this model of this and this takes care of the $V_{DB}$. This takes

care of the reverse bias drain to bulk and this takes care of the reverse bias source to bulk right and therefore you see that you do have two diodes and you have two capacitances, which are variable capacitances in these cases and this is drain current and the drain voltage which is in front of you.

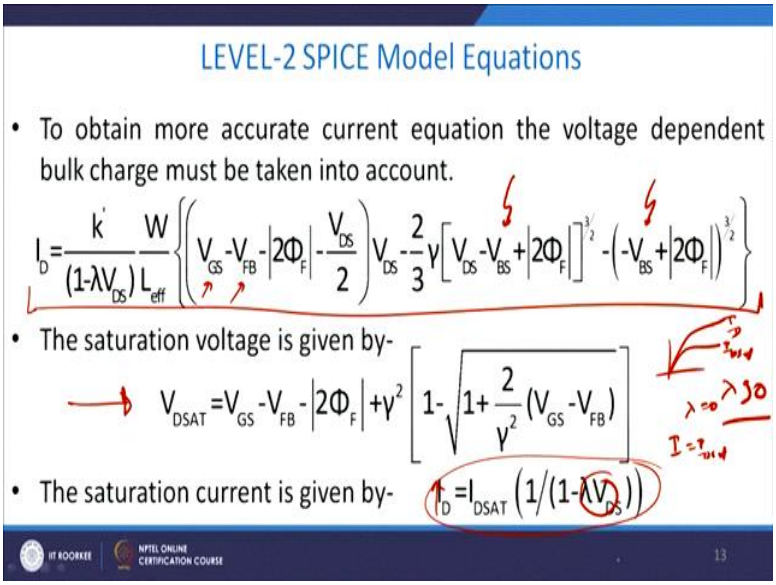
Of course, if you change the polarity or the type of bias the polarity of the diodes will also be reversed right. So for PMOS the direction of the current source, polarity of the voltages and the polarity of diodes must be reversed right.

(Refer Slide Time: 28:29)




Let me come to model 1 SPICE model equations again, it is a linear model, we have already discuss this point, it is $1 + \lambda V_{DS}$, where λ is basically my CLM parameter and it also depends upon the value of $V_{DS}$ K', K' is basically $\mu_n$, K' in this case is basically $\mu_n$ Coxide, also referred to as a process trans-conductance parameter. In saturation we have already discussed it is K' by 2 W by L effective this much value, where λ is the now this thing. Now to include the continuity between linear and saturation region this $1 + \lambda V_{DS}$ is kept common right, in the linear region, maybe it will be true that λ will be approximately equals to 0 and therefore this will not play a role right. But this will play a role in the saturation region, so we have got five electrical characteristics and this completely characterises the model. So if I give you the value of this K', $V_T$, λ, $V_{TH}$ and Phi, you will automatically be able to formulate this level 1 SPICE model in a much better manner right.

(Refer Slide Time: 29:36)

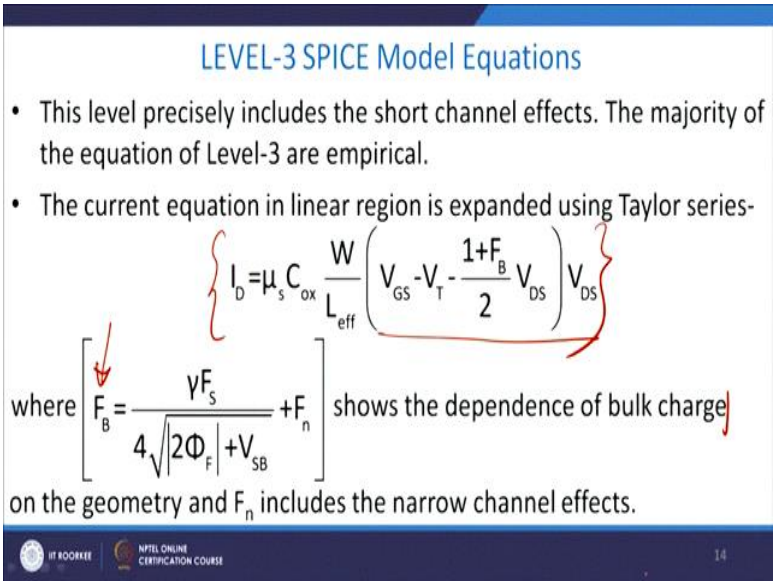



Let me come to SPICE level 2 model, where in this case, the voltage dependent bulk charge must also be taken into account right, because please understand the bulk charge varies with the applied voltages and depending on that people have found out a new equation given by this, big equation which is used in SPICE, do not worry but more importantly that you can see here is that you have $V_{BS}$ coming into picture, which means that the substrate effect is being taken care of in SPICE level 2 model right. And you do have gate to source voltage, flat band voltage and so on and so forth.

The saturation voltage is given by this formula and $I_D$ equals to $I_{D\ SAT}$ upon **1 - λ$V_{DS}$**, **1 - λ$V_{DS}$** equals to the saturation current. So if you see if **λ** equals to 0 that I get ID equals to $I_{D\ SAT}$ right, if **λ** is high-value, relatively high-value this quantity is high, 1 minus that quantity will be low and therefore $I_D$ will be relatively high as compared to $I_{D\ SAT}$. And that is the reason you will always get a current which is something like this, so this is $I_D$ and this is $I_{DSAT}$, $I_D$ will be always larger than $I_{D\ SAT}$ for **λ** greater than 0, right and that is how it behaves like as a non-ideal current source.

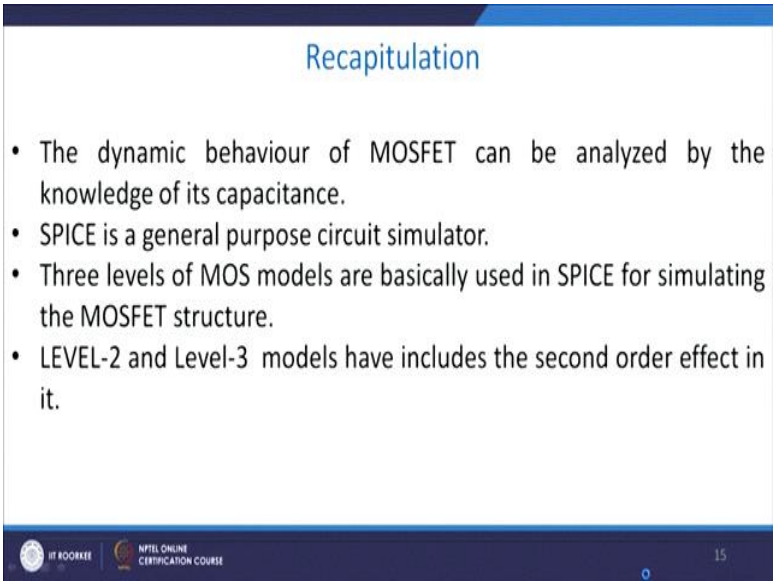
Level 3 is basically empirical model, which means that I extract I-V from my data from experimental model and when try to fit the linear and nonlinear equations with those experimental model, we change it and see which is the line of best fit and that equation is basically known as the level 3 SPICE model.

(Refer Slide Time: 31:00)




And it is given by this formula here, $I_D$ equals to $\mu_S$Coxide, W by L, this $F_B$ is basically my geometry dependent idea and that varies with the level which is there. Most of the time we do not use level 3, most of the time for our practical purposes level I and level II is the most seldom used SPICE model file for our practical applications.

(Refer Slide Time: 31:25)




We come to the last part of our talk and let me recapitulate what we have done till now. We have understood that we have three types of analysis DC, AC and transient. When you do DC analysis capacitance do not play a role because they are open circuited, when they play a role when you have a AC analysis or transient analysis coming into picture. SPICE has been quite useful as a general purpose schematic simulator for a long time and we should therefore

know what MOS model files are used in the SPICE in order to simulate the current versus voltage characteristics. We saw there were three basic level files, there are many but for your understanding at the basics level, you have level 1, level 2, level 3, level 3 is much more of an empirical model, based on empirical model.

Level 1 is primarily a large your gradual channel approximation which means your lengths are typically large, level 2 takes care of $V_{SB}$ which is a source to bulk and therefore substrate bulk mechanism and therefore short channel effects into consideration.

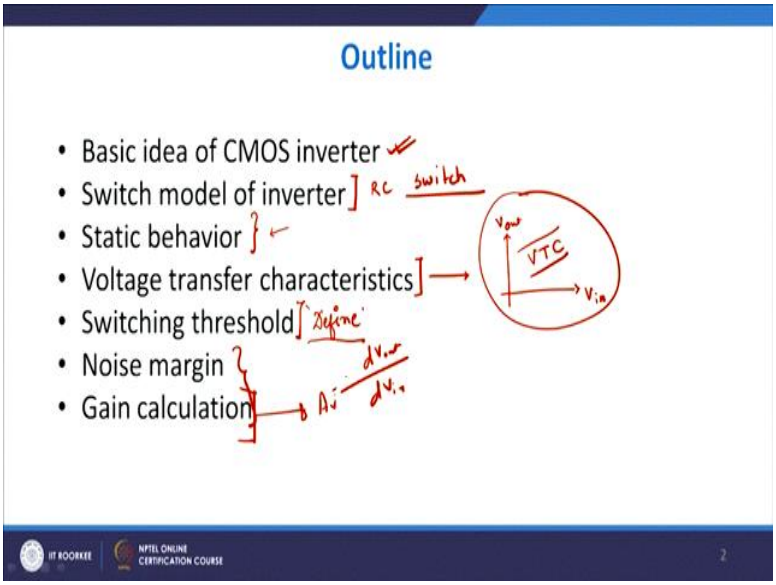
So level 2 and 3 takes care of short channel effects, level 1 may not take care of it. So with this, we have understood the basic fundamental principles of, basic ideas of SPICE model files at least and understood the requirement of capacitances and parasitic capacitances in this formula. I hope I have made much clear. Thank you very much and for your patient hearing.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−18**
**CMOS Inverter Basics - I**

Hello everybody, and welcome again to the NPTEL online certification course in Microelectronics: Devices and Circuits. We will start with today's lecture on CMOS Inverter Basics Part 1, so that will be the lecture module which is supposed to be delivered now, the CMOS Inverter Basics 1. In our previous interactions we have understood the basic principle or a functioning of a MOSFET, MOSFET is the current source and MOSFET is a switch. We also understood the basic concept of threshold voltage, which means that the gate voltage above which for an N channel depletion in enhanced mode MOSFET, the device is in the on state and below which it goes to the off state.

We have also understood the region of operation of the MOS device in saturation as well as in the cut-off region and we have therefore also appreciated on the second-order effects which are available in MOSFETs which come out because of certain, under certain conditions of structural or electrical inputs. What we will be doing is we will be stepping off from a device perspective to the circuit prospectives and we will start with today's lecture on CMOS which is CMOS technology, also known as complementary metal oxide semiconductor technology right, so it is basically C stands for complementary, complementary metal oxide. We will see why it is known as complementary as we move along.

(Refer Slide Time: 2:04)




So let us just give you the basic outline which we are supposed to deal with it within this interaction. The first thing is basically the basic idea of a CMOS inverter right, so we will see what is the meaning of an inverter because this is the basic building block for any digital or even analog VLSI design. So for any of the chip or any of the circuit which you want to design the prime notification or the prime building block is basically a CMOS inverter, so we will understand what is the basic idea of CMOS inverter, how it looks like, how it functions.

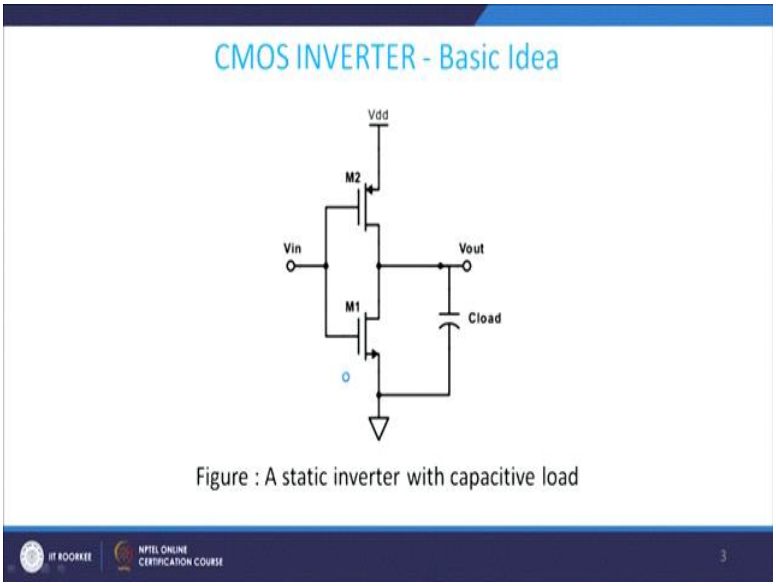
We will look into the switch model of inverter also known as RC switch model, so we will be taking these resistances, so this MOS devices as resistances, how these MOS devices tend to store a charge in the capacitance. X is a current source and then in a cut-off, X is open circuit and therefore we will be looking at the static behaviour of the device, which means that static behaviour primary means that you give a DC bias to the gate side of the MOS transistor and then see how does a current flow in the output of a MOS transistor, that is known as static right. Whereas dynamic behaviour, you see behaviour is when you start giving a voltage and you vary the voltage from low to high value and that is known as dynamic behaviour.

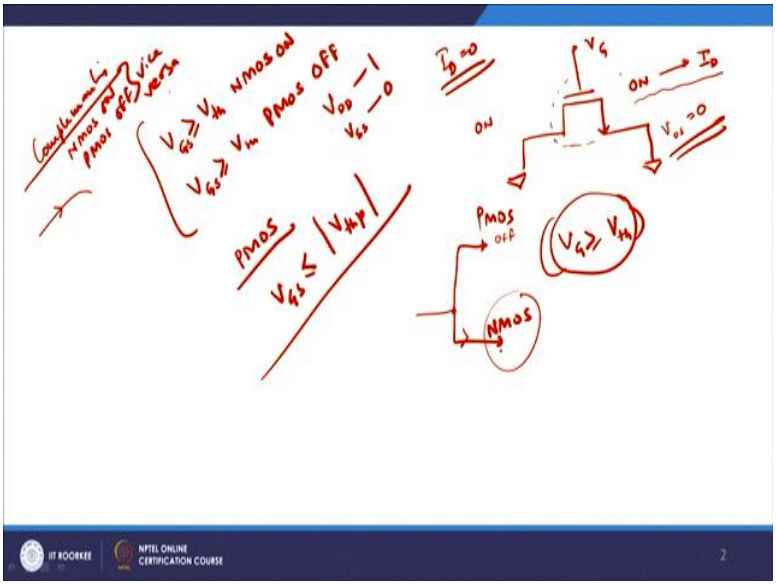
We will also look at voltage transfer characteristics, voltage transfer characteristics is basically the transfer or the profile between $V_{out}$ and $V_{in}$, so if I vary $V_{in}$, how does my $V_{out}$ vary in a CMOS inverter is defined as a voltage transfer characteristics right and it gives us any idea about how switching or how is the MOSFET is getting switched from on to off and off to on and how is it helping me to charge and discharge a battery or external capacitor in the output volt.

With this when we have understood what is a VTC and we have seen how VTC works, we will be actually coming to what is known as a switching threshold. We will define switching threshold with perspective to digital logic as well as to analog logic and see what is the meaning of threshold and what factors switching threshold depend.

Finally, we will be ending with noise margin and gain calculation. And noise margin primarily, basically we will understand that how much amount of noise can be inserted onto CMOS inverter without flipping of the data. So if 1 is there, it is still read as 1, so let me see how much amount on noise can I introduce into it, and that much amount of noise and inverter will easily reject and the data will not be corrupted. So we will be actually looking at noise margins of the inverter and then finally, from analog perspective or view we will be actually doing a gain calculation, which means that we will try to find out ∂V out right, ∂V in, that is what we find out $A_V$. So ∂V out, ∂V in we will try to find out and this is basically your voltage gain which we will be seeing as we move along.

(Refer Slide Time: 5:06)






So this is the outline or the basic outline of the talk and let me come therefore to a CMOS, basic CMOS inverter. Before we go to basic CMOS let me give you the brief idea of what we have done earlier so that it becomes easier if you do appreciate this point, if you remember in my NMOS transistor, if I had a gate voltage with me $V_G$, if my source was grounded and even if your drain is grounded, then $V_G$ should be greater than equals to $V_{TH}$ of this device, for this device to be on. It primarily means that for the device to be in the inverted state or in the on state my $V_G$ should be greater than or equal to threshold voltage of the device.

In that case only my device will be in on state, but in the on state it is not necessarily that there will be an $I_D$. As you can see since your drain is grounded you do not have any chance of any drain current flowing right and therefore $I_D$ can only happen, provided you have a $V_{DS}$ right. But at this stage $V_{DS}$ equals to 0 and therefore I think $I_D$ equals to 0 and therefore even if $I_D$ is equal to 0, the devices is in the on state.

So whenever, let me recapitulate for a n channel device when $V_{GS}$, $V_G$ is greater than equals to $V_{TH}$, NMOS is on right and when let us look at PMOS, whenever $V_G$ is greater than $V_{TH}$ PMOS will be always off, because for PMOS I require that, this condition should be $V_{GS}$ gate to source, a gate to source and for PMOS the condition is that my $V_{GS}$ should be less than equals to $|V_{THP}|$ right, that is very, very important.

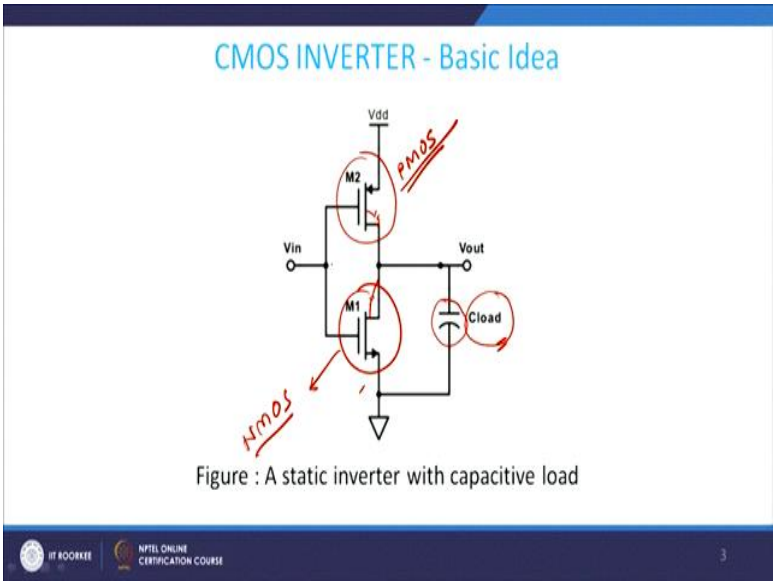
So you see for a same input signal right, if my NMOS device is on, and I give the same input to my NMOS, PMOS device and NMOS device, let us suppose this is my PMOS and this is my NMOS and my NMOS is switched on by the virtue of the fact that this gate voltage at this point is larger than threshold voltage of the device, this will automatically ensure that my

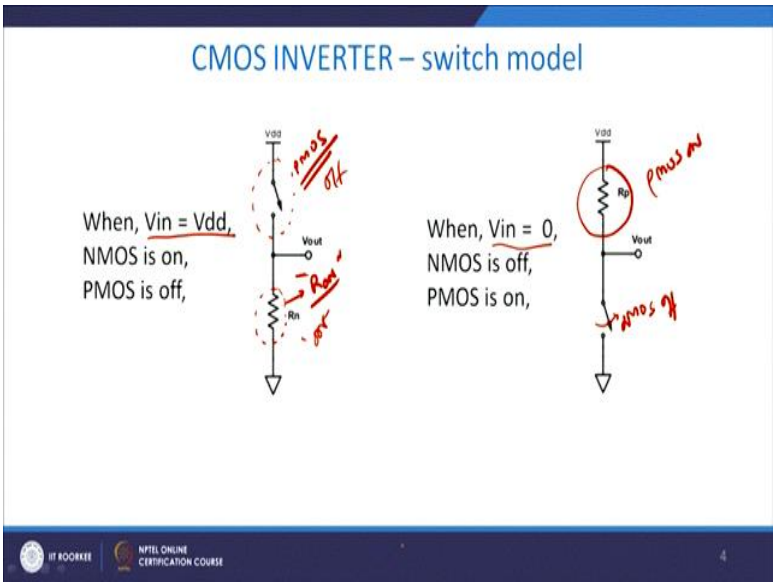
PMOS will be off, why? Because PMOS you require a lower threshold voltage or threshold voltage to below in order to switch on the device.

So the primary difference between NMOS and PMOS is the very basic fact that for the same amount of gate voltage given to NMOS and PMOS, if means like in digital logic, if which means that if $V_{DD}$ corresponds to 1 and $V_{SS}$ or V ground corresponds to 0, then when apply $V_{DD}$ to both the gate terminal of NMOS and PMOS, the NMOS is switched on and the PMOS goes to cut-off mode and whenever $V_{SS}$ equals to 0 or my input is equal to $V_{SS}$, then my PMOS switches on and NMOS switches off.

And hence the name complementary, which means that both NMOS and PMOS will never be on together at, they will be on for a small duration of time, but for most of the voltage range I will always get this to be as a complementary switch, complementary, the complementary right, what is reason of complementary? Because both NMOS when it is on, PMOS is off right and vice versa, this will be vice versa right. So this is very, very important idea, which you should always keep in mind, that you always have a switching of threshold taking place in this manner.

(Refer Slide Time: 8:26)



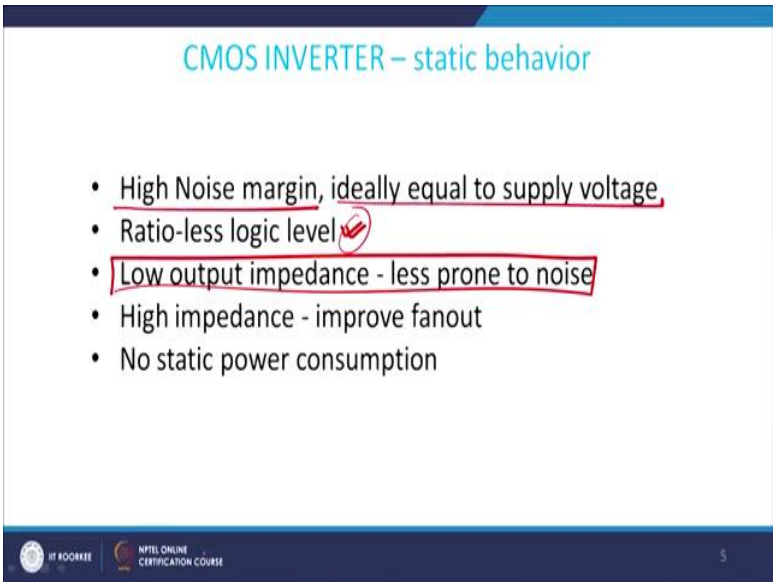



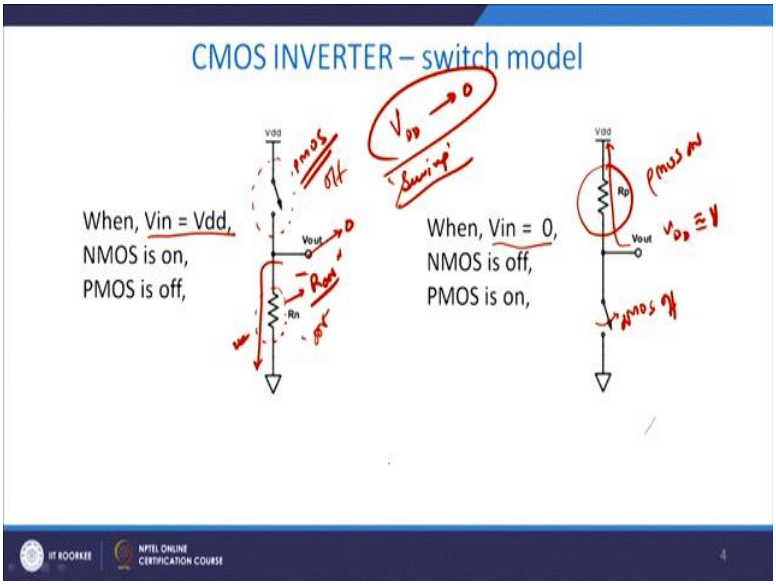

Now to come into that fact as you can see here, this M2 is basically my PMOS, this is my PMOS, this is my PMOS and this is my NMOS, this is my NMOS and this my PMOS which you see here. Now whenever my, we obviously have a load capacitance here, this load capacitance is nothing but the overall capacitance of this part here right or this part between these two points and so on and so forth. So all the capacitances in this PMOS and NMOS if you add together, I get C load available to you, also C load can be the load capacitance which you are putting in the external volt or the load capacitance subsequence stage of the CMOS inverter right, that you plug it and place it here right. So therefore it basically means that the inverter at one point of time will be charging the C load and another point discharging let me see how it works out.

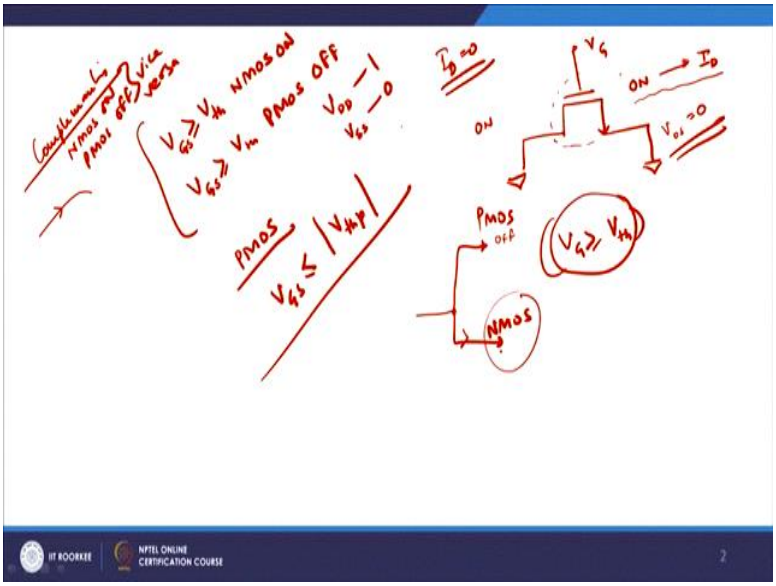
So when V in maybe I just show in the next slide right, now when V in equals to $V_{DD}$ right, when V in equals to $V_{DD}$, when my input voltage is equals to $V_{DD}$ at that point of time, my NMOS is switched on right, but my PMOS is switched off because my input is high, so as you can see in the switch model, if you look here this my PMOS, PMOS is basically open circuited whereas NMOS is closed circuited with resistance equals to R on right. R on is the transient resistance between source and read of the MOS device, and that is the reason it is complementary, means both are not together on.

Similarly, when V in equals to 0, my PMOS becomes on and therefore I have a resistance offered in this manner and I have my switch model shows me an NMOS to be off, so this NMOS is off, whereas PMOS is on right. In this case, the PMOS is off, and NMOS is on, so that is pretty simple and straightforward way of looking at inverter switch model.

(Refer Slide Time: 10:39)






If you therefore come to the basic fact that the idea here is therefore if you, let us go back to the slide, if you look very carefully when your NMOS is switched on right, this V out will have a path and this will be grounded here, so the output voltage will be 0. When my PMOS is on and my NMOS is off, then this will be, voltage will be plugged to $V_{DD}$ and the output will be equals to $V_{DD}$ which is effectively equals to 1volt, which means that this inverter allows you to do a peak to peak swing, peak to peak basically means it is actually swinging from $V_{DD}$ output to 0 output, such a large range it is allowing it to do.

So we have large swings associated with CMOS inverter switch model right and the reason being that you have now large amount of, you allow the voltages or the output to go till $V_{DD}$ when my PMOS is on and go to ground when my NMOS is on right and that is the reason you get such type of circuit and that is the reason you therefore have a very high noise margin

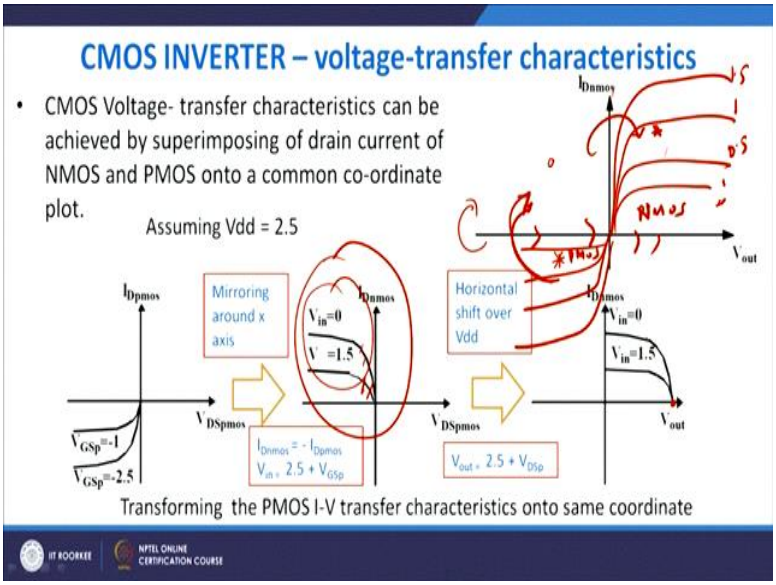
available with you and ideally equal to the supply voltage. So whatever is supply voltage you get noise margin which means that a static CMOS inverter right, since it is able to pull your output voltage to logic one which is equals to $V_{DD}$ right, it is utilising the full support from $V_{DD}$ to ground rail, so the whole range is being utilised by the input level right.

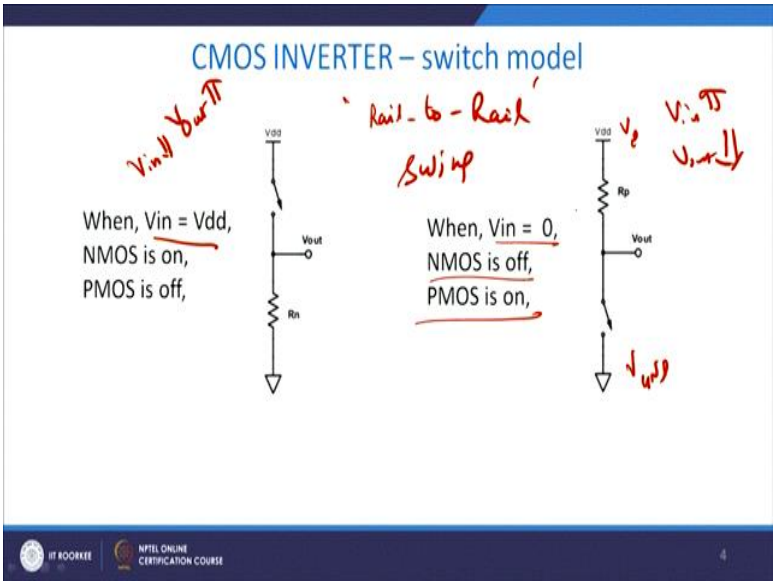
This is also referred to as a ratio-less logic level, which means that independent of the dimensions of these devices, W by L ratio of PMOS or NMOS I am 100 percent sure that the threshold voltage will always remain the same it is, which means that once you have fabricated your PMOS and NMOS and attached it as a CMOS inverter using a twin tub process, you will see that, we will see that your logic is basically a ratio-less logic, which means that, I do not have to change my W by L of my PMOS or NMOS in order to have different sets of behaviour.

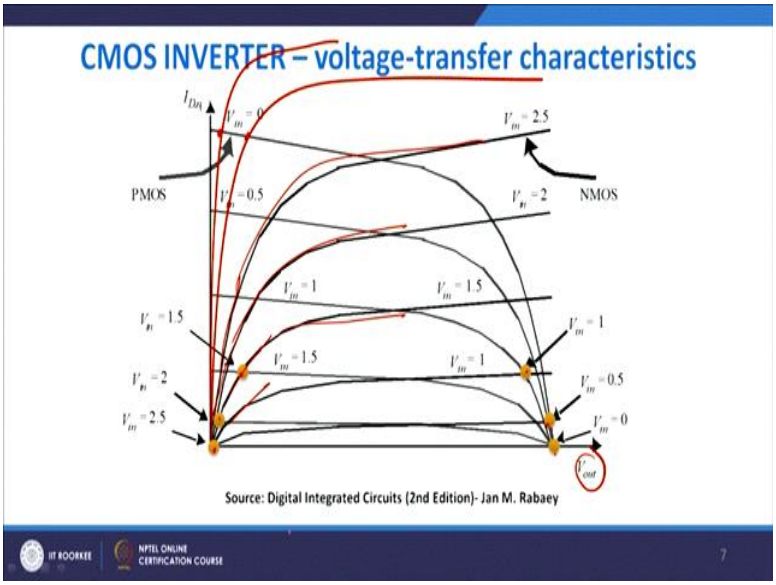
So for a same set of behaviour as a switch the same W by L which you are using earlier, you can use it and gives you a very good result here. So if you go back to your previous case, the ratio-less logic is there, which means that it is independent of W by L. Since the output impedance is the low impedance node, it is less prone to noise and this is a very standard way of looking at it, that whenever an output source is connected to high impedance you have less noise immunity. When it is connected to low impedance node the noise immunity is higher and therefore you have less noise at the output node.

The impedance is high, the output impedance, the output the input impedance is high because I am inserting my signal onto the gate side of those devices. Now gate is always terminating through oxide layer or a dielectic layer and as a result, if you look from the gate side the resistance offered by the device to input signals are typically very large right and therefore it is high impedance and therefore improve fanout. There is no static power consumption or even if it is there it is very small static power consumption is there between CMOS right.

(Refer Slide Time: 14:26)



So we have therefore understood one basic idea, let me clarify before we move forward therefore, that two things are very, very important as far as this structure is concerned and the two things which are very important are the first thing is that this CMOS inverter gives you a rail to rail swing fine, what does it mean rail to rail swing? It gives me ride from $V_{DD}$ to $V_{GND}$ the whole signal can be used as directly right. So when my V in is lower, when my V in is low right, my V out is high, high and when my V in is high, V out is low, fine and exactly it behaves like a inverter, it may mix a behaviour of the inverter right and so on and so forth.

So if you look very carefully when V in equals to 0 NMOS is off and PMOS is on right and we are able to achieve a much better design in this case as compared to the previous cases. With the knowledge therefore you have gained till now, let us formulate the voltage transfer characteristics of CMOS, voltage transfer basically means on the x-axis you will have V out and on the y-axis you will have V in, so if I vary V in, how does my V out vary? That supposed to be seen in the VTC of a CMOS inverter.

So what do we do is, it is very simple and straightforward, we try to find out the drain current characteristics of both PMOS and NMOS and then invert it to get the values right. If you look very carefully NMOS has got a profiling in this manner right and if you look very carefully NMOS is always on the this quadrant right, this quadrant it is there. In PMOS, PMOS will be always in the negative and the current will be also negative, so this is my PMOS quadrant device right, this is my NMOS quadrant device right. Then what you do is that you fold it across this spirit in such that you fold it in this manner right, once you fold in this and then you fold to this, let us see how it works out.

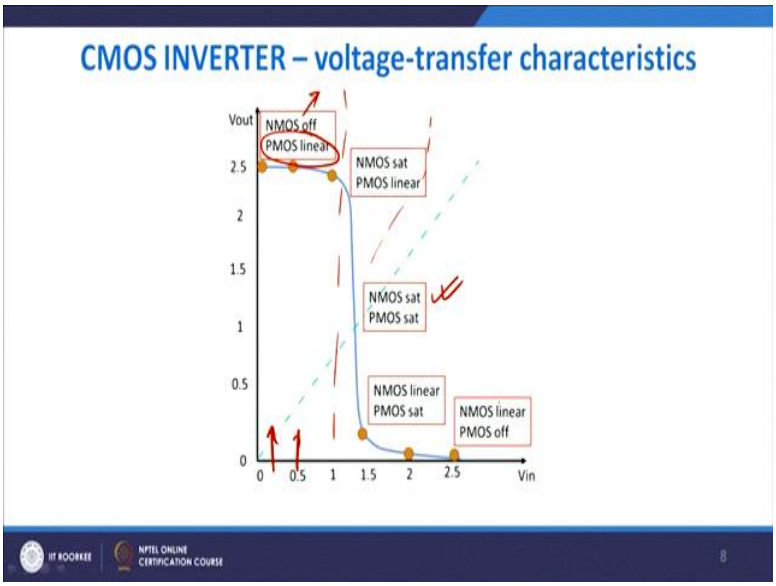
So I have got something like this, this is NMOS right and from here I get PMOS is fine. Now if you fold this to this side and then fold this to this side, you finally get something which is very close to the voltage transfer, where they cut they will get voltage transfer characteristics. Now this is NMOS, V in 0, we get to 1.5 and if you solve it on right hand side, I get something like this. So this is 1.5, 1 volt, this is 0.5 and so on and so forth right.

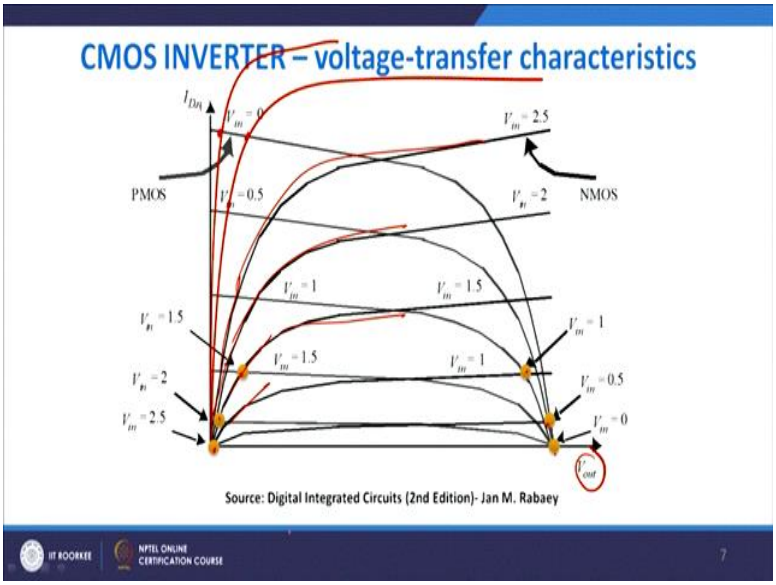
When you have PMOS is you are in third quadrant, you push it across this x-axis, go to this point and then push it across the y-axis to go to this point right which is this one, so I get this. Now if I shift it to the right inside, I get something like this, fine and therefore what we get from all this discussion is, therefore if my this module or this module and this module right, they superimpose on each other, I will start getting a set of points, a set of intersection points in this case right, a set of intersection points.

So how do I do? This is my NMOS, so as you can see it is I versus V and this is the curve which you get right, this is a curve which you get, fine and this is a curve which you get. Now when you plot PMOS versus NMOS, PMOS $I_{DN}$ versus or $I_{DP}$ versus V out, I get something like this right, this is the profiling which you get right. This is the profiling which you get, now wherever these two cut each other, for example, it cuts each other at this point, at this point and at this point, similarly at this point, this point and this point, so these are the points where the currents from both NMOS and PMOS will be equal and they will be sufficiently important to run the at the CMOS inverter right.

So with this knowledge, we can safely therefore say, safely say that these points which are not available to you will still, so if you have an NMOS which goes something like this, you will still have one point here, if it goes something like this then you still have another point here and this will be all your PMOS points. NMOS points will be directly scaled at this particularly limit, so that is what you get out of these differential equations.

(Refer Slide Time: 19:10)



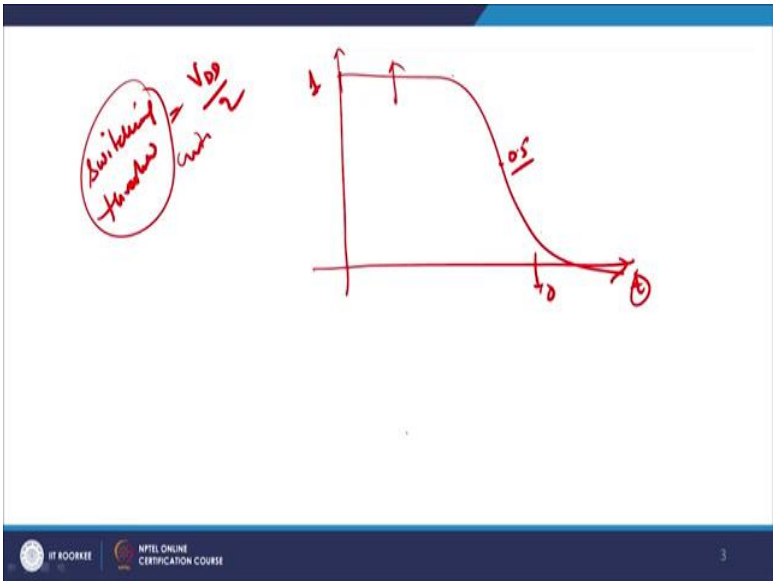


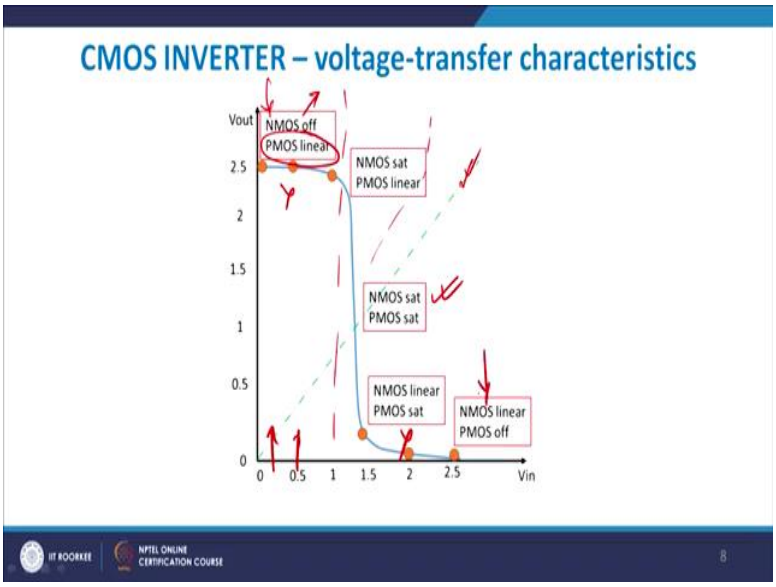

Once you have done that, now you are in a position to explain the various factors which are there with us right. Let me therefore come to the previous case, let me therefore explain to you what the issues are. So now we have done what? We have done the voltage transfer characteristics we have drawn, anything larger than the voltage transfer characteristics will drive PMOS to heavy saturation as well as if you do heavy on the N side, N side will be driven into cut-off region right and that is what you get from here.

Now if you look at CMOS inverter, you see when my input was low, which is this one right, my NMOS was off and my PMOS was in the linear region of operation. So let us be very clear about this particular point, that during the point when your input was very, very low less than 1 volt, then we saw that NMOS was off, of course, because is low than threshold voltage and PMOS is in the linear region of operation right.

If you increase your V in and you reach to a region where your input is there, but the output is not there, then we define NMOS to be saturated and PMOS to be in the non-linear region, so the first was NMOS off, the second is NMOS saturation and PMOS in the linear region and then what we do is, in the third one both NMOS and PMOS in saturation and then just the reverse happens here. Now NMOS is in the linear and PMOS is in saturation, here NMOS is off, so here NMOS will be on and linear and this is linear and off. So you will have off PMOS available with you, so all the five notations which are there at a voltage transfer characteristics of a CMOS appears here as well right, that is very, very important to be careful about.

(Refer Slide Time: 21:12)




Now how I define a switching threshold? A very important point switching threshold and it is something like this that if you have threshold then it is something like this, suppose let us suppose this is how I get right, this is time and this is your profile. Then we say that switching threshold is that value of input voltage at which you would expect to see a change of state in the output side, typically the ideal value of switching threshold is, switching threshold is basically $V_{DD}$ by 2. So if you apply $V_{DD}$ of 1 volt typically switching threshold critical can be written as 0.5, so if your input voltage is just below 0.5, please read it as 0, if it is greater than 0.5 please read it as 1, right.

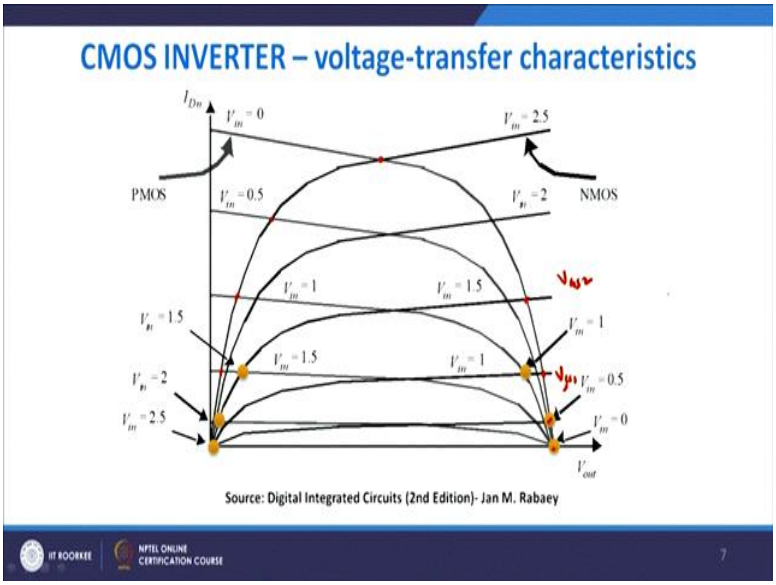
So we have understood what is the difference between the switching threshold and positive switching threshold. In this case to be switching threshold right, so we have understood

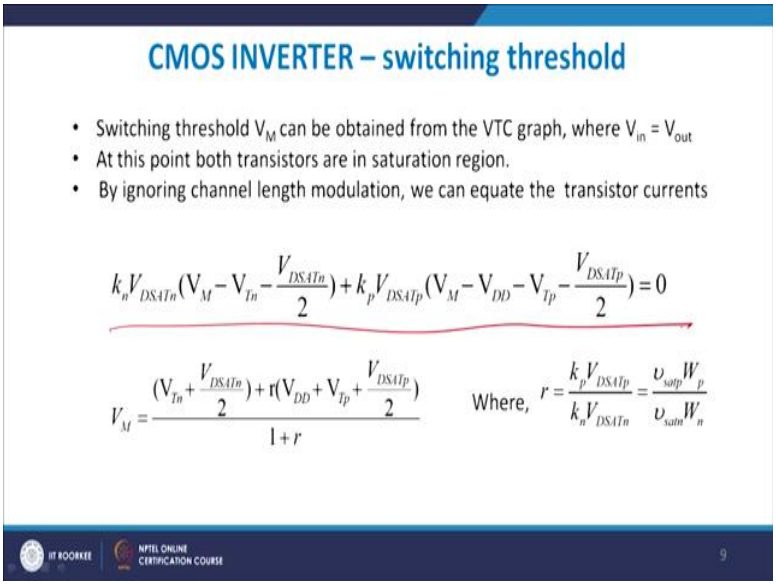
CMOS inverter, working principle, CMOS switching threshold, what is the meaning of switching threshold? And CMOS inverter noise analysis, you have also understood basic fundamental principles that dV out and dV in should be 1, which means the amplification should be 1 for finding out the values of other variables.

Now let me explain to you the basic, as I discussed with you the MOSFET inverter principles and therefore they are divided into five regions, the middle region which is the green one which you see here is basically the region which is highly unstable, because a small shift in the input will result the Q point to shift either to the top or to the bottom. Top, bottom means this point or at this point right.

So you have to be very careful when dealing with analog or mix signal blocks because they force output to either go to the $V_{DD}$ or go to ground right, whereas your job in this stage is yes, you need to go to high $V_{DD}$ and low $V_{DD}$, but then you have to be very cautious in the terms that you are stable also in that value of voltages right. So when $V_{DD}$ goes to NMOS of linear region and in my this case NMOS to be in the linear region, whereas here PMOS was linear region, so this is PMOS linear region, this is NMOS linear region right and it works fine for any of a circuitry which you see.

(Refer Slide Time: 23:33)

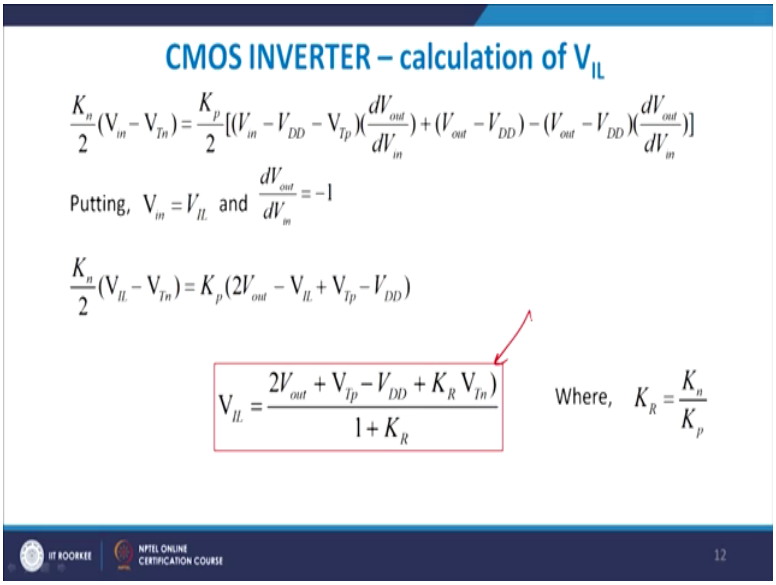





As I discussed with you just beforehand that when you plot this graph, so what I was trying to tell you is that this is being cut, so this is NMOS and this is PMOS, they are cut at various points, the first point is this one NMOS and PMOS, then you this point available with you, then you have got, so this point, this point and so on and so forth and on this side you will have this, and then you will have this, this and so on and so forth right. And therefore for various gate voltages, V in is gate voltages, I start getting a new value of input voltage and this input voltage are strong functions of the output voltage.

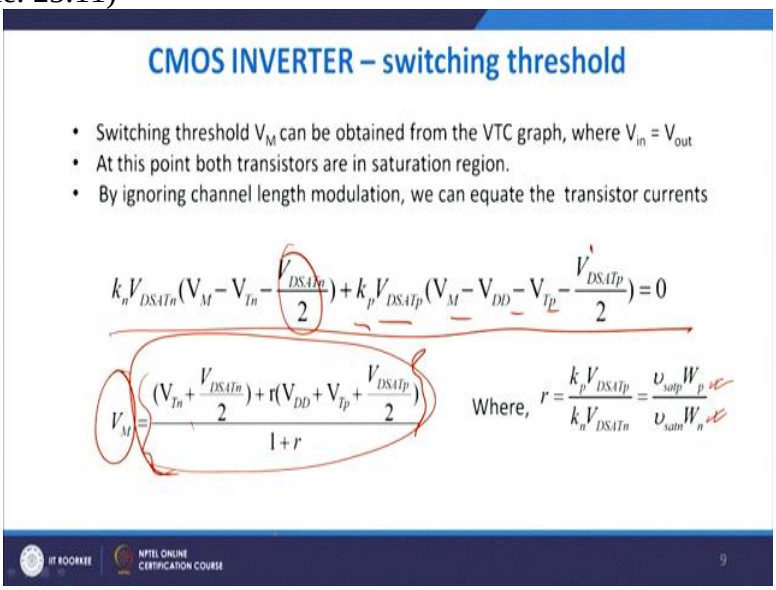
As you can see here, therefore this is for V, so for the same $V_{GS}$, so this is my $V_{GS\ 1}$ right, $V_{GS\ 1}$, $V_{GS\ 2}$ and so on and so forth and these voltages are same for both NMOS and PMOS right. Now we have understood therefore the various regions of operation, let me see how we obtained switching threshold from the VTC of the graph right, Now the VTC is given of switching threshold. Please understand your devices will be in linear region of operation, whenever you are in either region A or region C or region E of the major issue or maybe I will just show you.

(Refer Slide Time: 25:02)




Okay, I will just tell you that, so what I will show to you is that my $V_{IL}$ is nothing but this whole quantity which you see in front of you. How did you find out $V_{IL}$ and we will show you that, doing that but let me come to therefore….

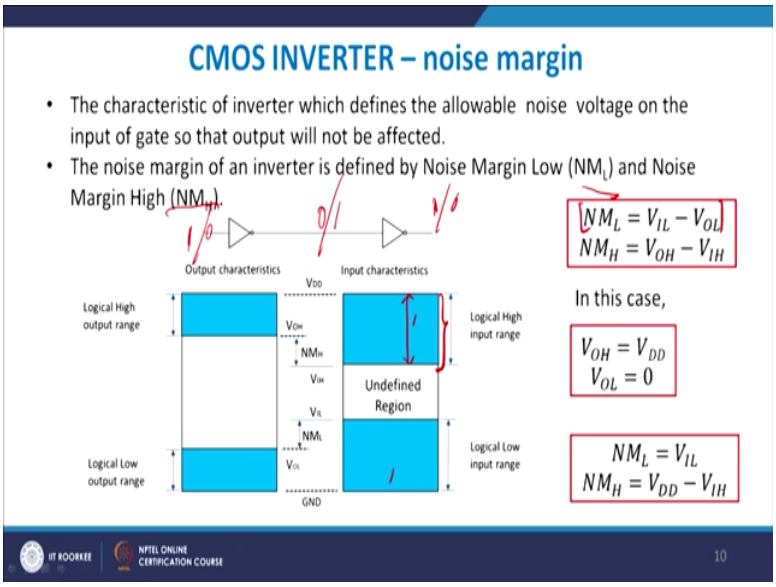
(Refer Slide Time: 25:11)




So understand that VM is given by this quantity, where r is basically the ratio of $K_P$ to $K_n$, $V_{D\ SAT\ P}$, to $V_{D\ SAT\ n}$ and that is also equals to $V_{SAT\ P}$ WP and $V_{SAT\ n}$ W n, so these are the four parameters which are there and we should appreciate the fact that all the transistors if you want to find out both the transistors in saturation right and therefore we say that $K_n V_{D\ SAT\ n}$ into $V_M - V_{TH}$, $V_M$ is nothing but the applied gate voltage, minus $V_{TH}$ - $V_{D\ SAT\ n}$ by 2 right, so this $V_{D\ SAT\ n}$ by 2 is a fixed quantity available with you.

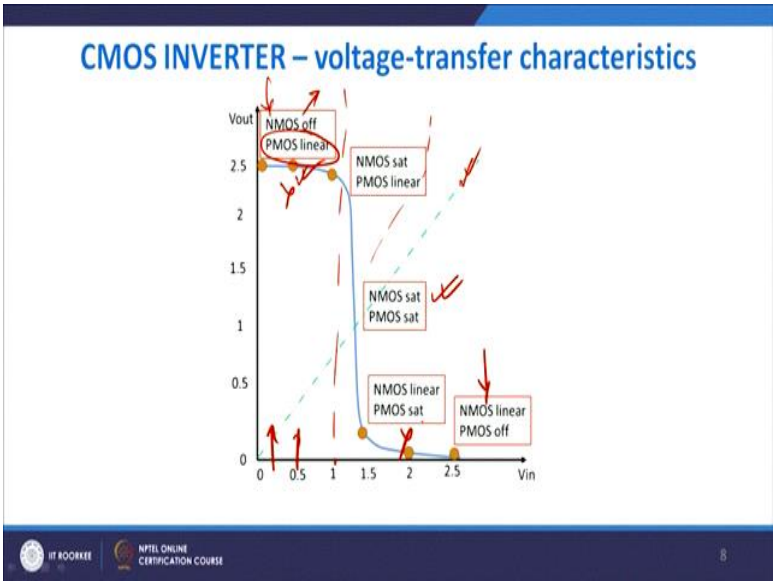
Similarly this, this are all fixed quantities and from here you can obtain the value of $V_M$ or switching threshold right and the given by this formula for the switching threshold $V_{T n} + 2 V_{D SAT n} + R$ and so on and so forth. So as you can see higher the value of $V_{D SAT}$, higher will be my switching threshold.

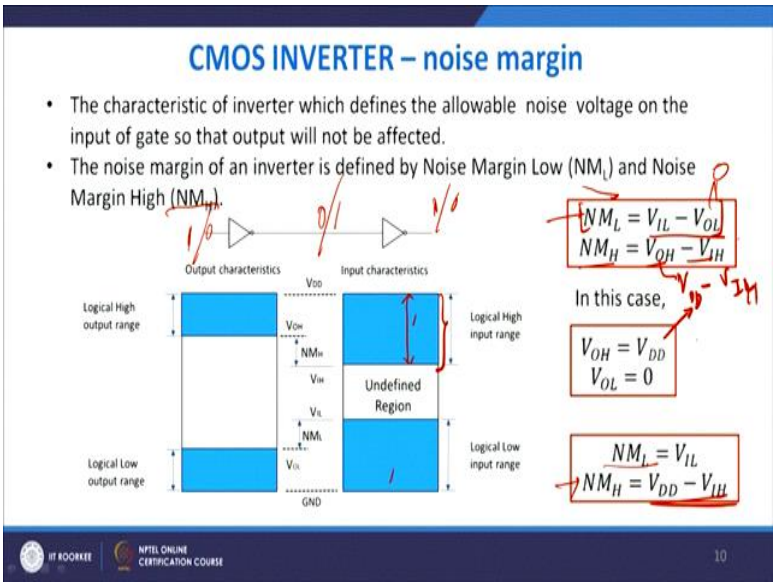
(Refer Slide Time: 26:13)




Now let me come to the CMOS inverter, inverter is very simple and straightforward, if there are two inverters connected in this manner right, they connected in this manner then if you apply a 1 voltage here, here I will get 0, I will get also 1 here but let us suppose by mistake I emit 0, then this will be 1, this will be 0. So this is my VDD, output characteristics input and this is logical high range as I discussed with you, this is logical high, which means that any particle or issue with voltage range in this value will be unstable and therefore it will not be available to you in the literature. Whereas if you take care of low noise margin we assume that $N_{ML}$ equals to $N_{MH}$ which means the low noise margin $N_{ML}$ is exactly equals to $N_{MH}$ which is high noise margin. And you do have this and this region effectively available to you. Now we define therefore the $N_{ML}$ is equals to $V_{IL}$ - $V_{OL}$.

(Refer Slide Time: 27:22)



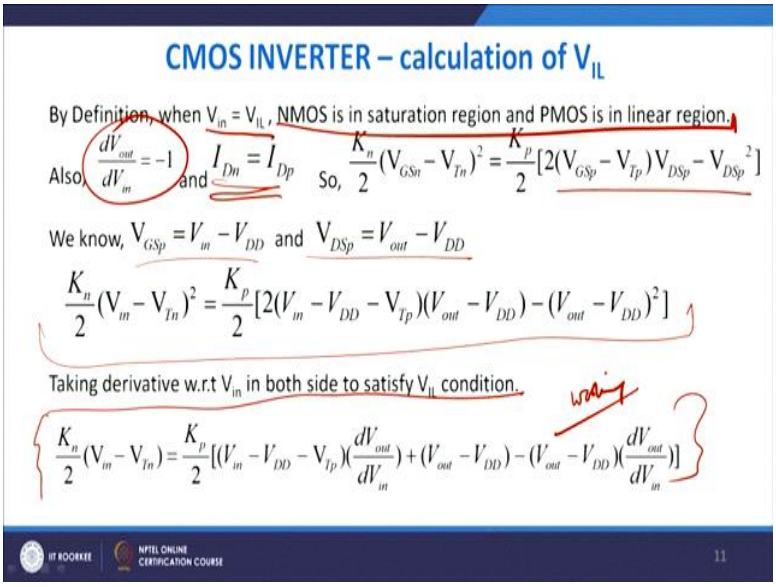

What is $V_{IL}$? Right, let us see what is $V_{IL}$. Right, let me explain to you what is $V_{IL}$. $V_{IL}$ is basically meaning, as the name suggest it is $V_{IL}$ is V input low right. Whenever my input is low, so if you look very carefully here when my input is low, output is high, when my input is high output is low. So when my input is low, I would expect to see that my output is high and similarly when my input is high my output is low.

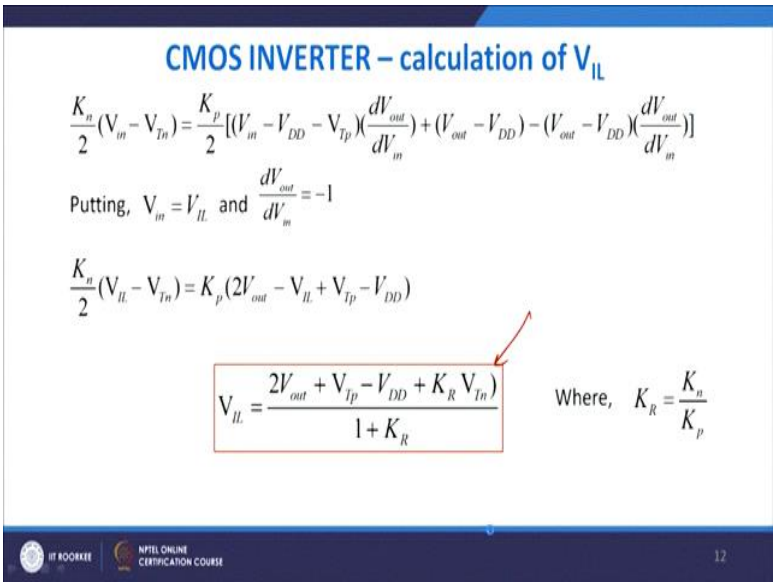
With this knowledge we define, so we define low noise margin $N_{ML}$ as $V_{IL}$ - $V_{OL}$ and this is input low minus output low and $N_{MH}$ basically is output high minus input high right and therefore this $N_{MH}$ takes care of the higher values, one higher probability of values. Whereas $N_{ML}$ takes care of zeros and lower probability values, right and so on and so forth. Now, as I discussed with you since $V_{OH}$ equals to $V_{DD}$, so I get, so $N_{MH}$ will $V_{DD}$, $V_{DD}$ - $V_{IH}$, $V_{IH}$ is

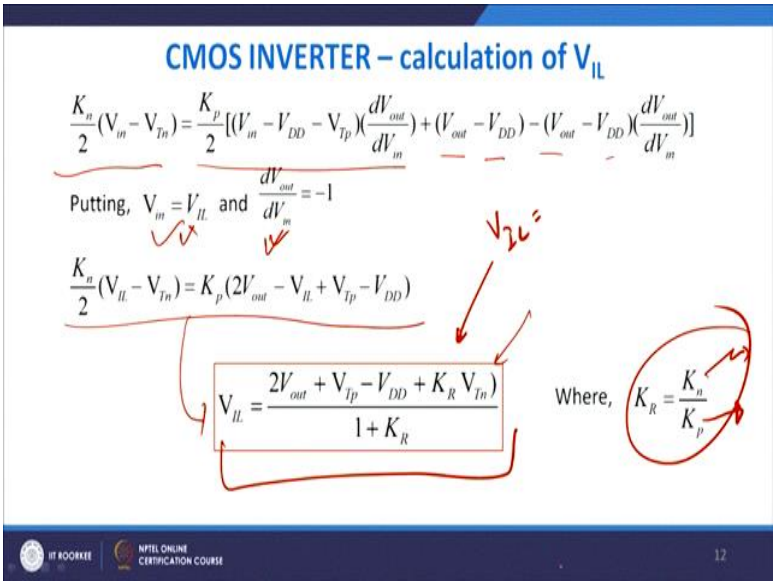
basically equals to V, let us suppose $V_{IH}$. So my $N_{MH}$ is equals to $V_{DD}$ - $V_{IH}$ and my $N_{ML}$ is just equals to $V_{IL}$ because $V_{OL}$ indicates 0.

So I get $N_{ML}$ equals to $V_{OL}$, so I get $N_{ML}$ equals to 0 or equals to $V_{OL}$ and the second case, I get $N_{MH}$ equals to $V_{DD}$ - $V_{IH}$, and that is what we have understood or given process and idea.

(Refer Slide Time: 28:55)





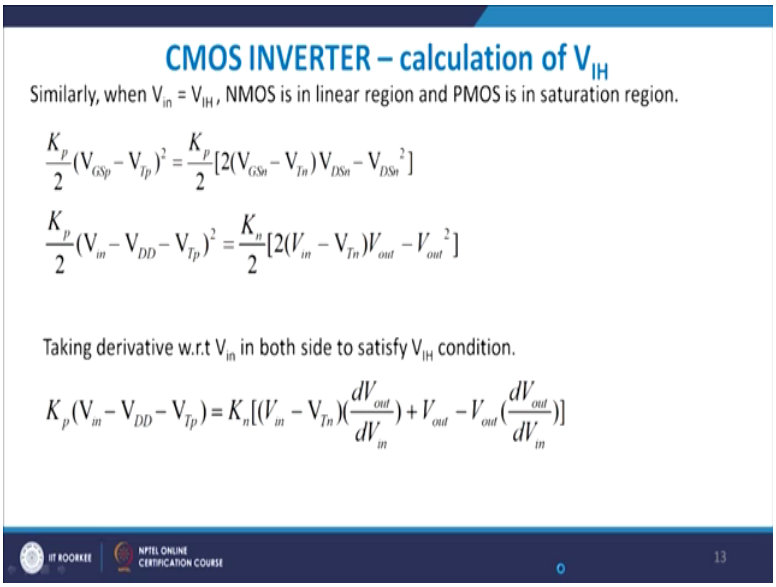

So how to calculate $V_{IL}$? Well, by definition when $V_{IN}$ equals to $V_{IL}$, NMOS transistor is in saturation and PMOS is in linear region right, so just the onset of saturation for NMOS and for PMOS it is a linear region of operation and we also know and this is quite interesting. I like you to appreciate this point that dV out, dV in is equals to always minus 1. So you have to be very cautious that when you are calculating $V_{IL}$, you have to assume that my input and output are exactly equal, if they are not you have to first make it equal in order to find out the value of current, which is in this case $I_{D\,n}$ equals to $I_{DP}$ right.

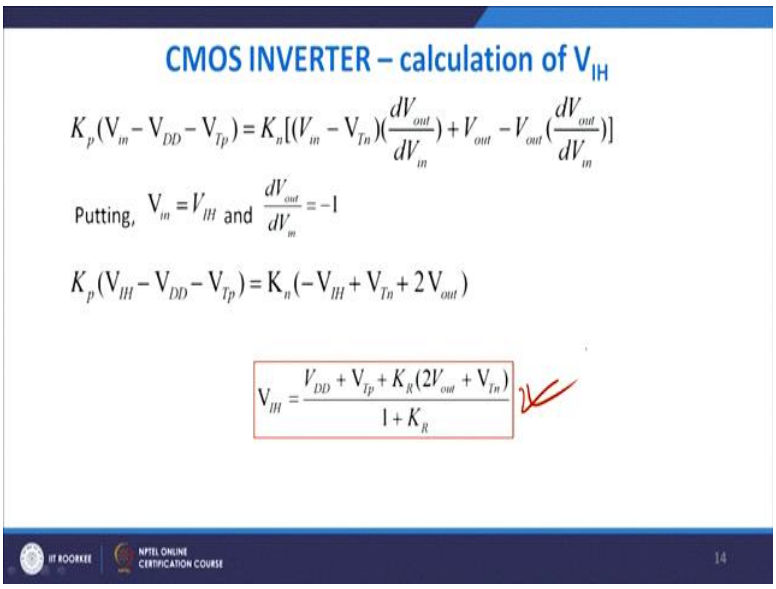
So therefore, in saturation mode I know $K_n$ by 2, $V_{GSn}$ - $V_{Sn}$ whole square must be equals to $K_P$ by 2, then this is a non-linear region of operation. If you remember, I think it is clear to you. So in this problem or in the solution to the problem is that you do have a fixed value of gate voltages and your applied voltages but the input signal will tend to change it slightly whenever the output signal is above, when the output signal is above $V_{DSP}$. So we know $V_{GSP}$ is equal to $V_{IN}$ - $V_{DD}$ and $V_{DSP}$ is equal to $V_{out}$ - $V_{DD}$. If you solve it I get this big equation, it is available to me. If you take derivative with respect to $V_{IN}$ then both sides will satisfy $V_{IL}$ conditions, so I get this equals to this, into this plus this plus this. So I get this as the workable solution for this $V_{IL}$ input low.

So we have learned two things, what is input low? Input low is the voltage at which the output is high right. CMOS inverter calculate $V_{IL}$, $V_{IL}$ is basically this point. So what I say is that this equals to $K_P$ by 2, this whole thing plus this minus this minus this minus this dV out dV in right and if I place dV out, dV in equals to minus 1, $V_{IN}$ equals to be $V_{IL}$, I get a big equation like this, which can be actually downtrodden to $V_{IL}$ to be equals to this much. So relatively $K_R$ is basically $K_n$ by $K_P$ parameter for n and P. But if these are okay, if these are

fine, the values of parameters are extracted fine manner, you can say that $V_{IL}$ is given by this equation, fine and it depends on the value of $K_R$, it also depends upon the value of V out and $V_{DD}$.

(Refer Slide Time: 31:40)

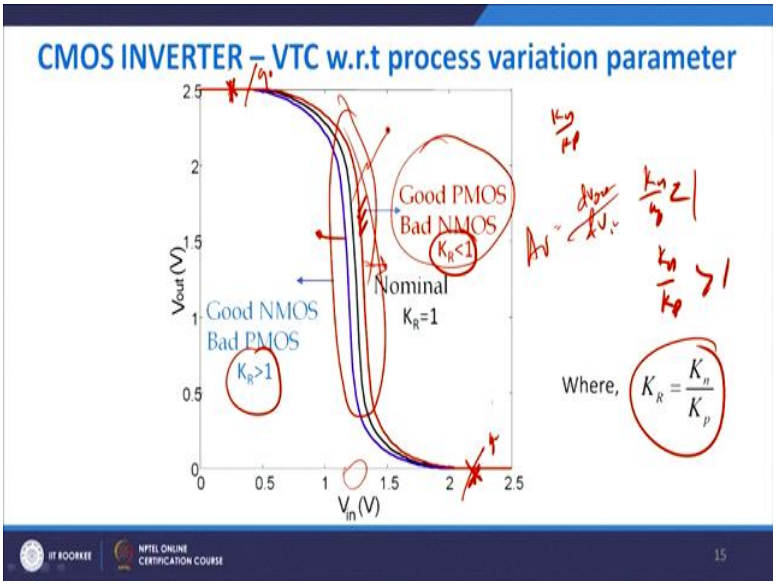


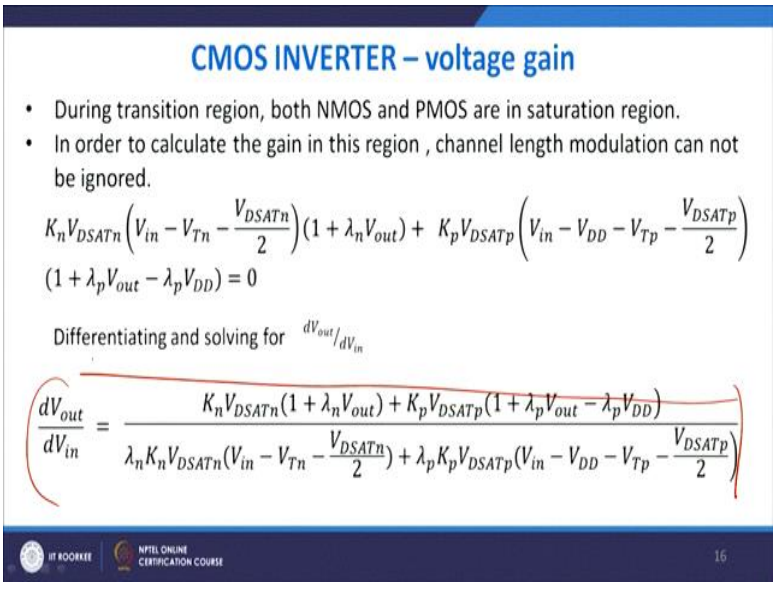




If you want to calculate the value of $V_{IH}$, then $V_{IH}$ is input high, which means that whenever my input is high and output is low anything larger than that it is not expected from you. So if I solve it that grade is exactly the same, $V_{GSP}$ - $V_{TP}$ whole square is equal to this whole quantity. Similarly, for a β value or a breakdown value I get this into consideration and if you take the derivative of both sides, I get $V_{IH}$ as some value $V_{IN}$ equal to $V_{IH}$ and dV out equals to dV in. If you solve it I get $V_{IH}$ equals to this quantity which you see in front of you. This is

the value here, but this would not have come out if you did not do the derivation properly in your previous case.

(Refer Slide Time: 32:32)



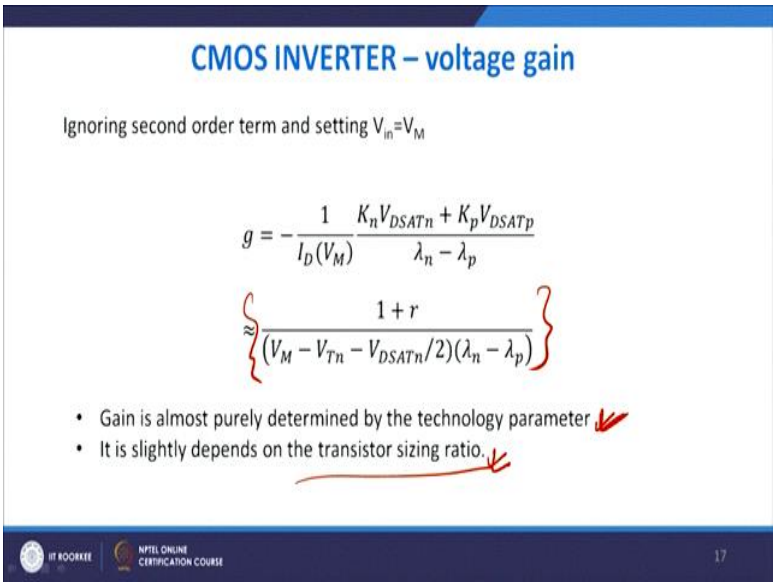

The voltage transfer characteristics varies in this manner that for nominal when $K_R$ equals to 1, I get this, the black one is the nominal 1 and it is almost giving a fall but if you take good NMOS or good PMOS and bad NMOS then the transfer characteristics shifts to your right, whereas if you take good NMOS and bad PMOS shifts to the left, good, bad variably means that when you say good PMOS, does it have, I say it is good if and only if it has got a pull up capability to VDD.

Similarly an NMOS can be said to be good provided it is able to push down or pull down the voltage to ground right, so none of them we are able to do and therefore I can safely say that

this is a good PMOS right and therefore shifted to the right with high switching threshold and this is good NMOS and lower switching threshold right. So this is $K_R$ less than 1 and $K_R$ greater than 1, $K_R$ is basically $K_n$ by $K_P$, so when the $K_n$ by $K_P$ is greater than 1 I shift to the left and when this is less than 1, I shift to the right. And the nominal value is $K_R$ equals 1 means $K_n$ equals to $K_P$, which primary means that PMOS and NMOS have equal pulling capacitor.

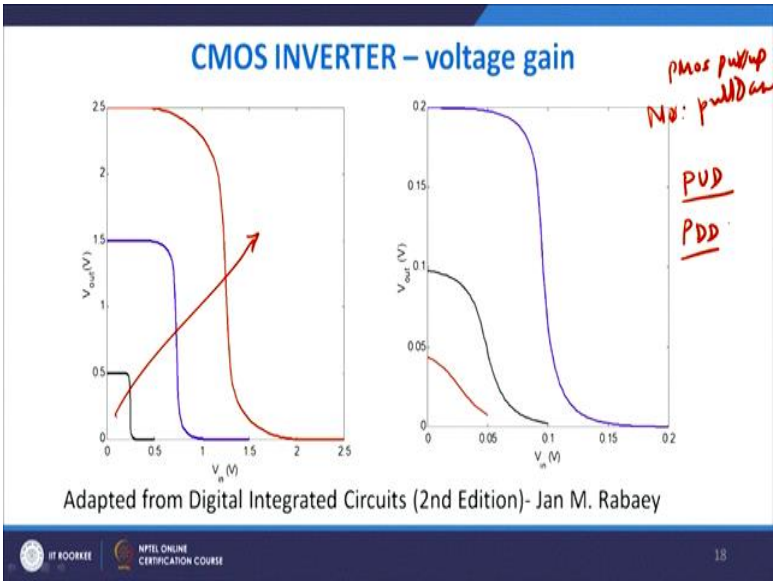
From the same graph, see then, If I say, from this graph, you can see that I can switch from this point to this point and I can go from logic 1 to logic 0 right, this is output logic 0 and this is output logic 1 which you see. But then somewhere in the middle, obviously the gain here will be 0 and the gain here will also be equals to 0. But somewhere in the middle your gain will be a strong function of your $V_{IN}$ because if you remember $A_V$ equals to dV out right, dV in, and that is what you get here. So dV out dV in right down and you get this big expression which is important in front of you.

(Refer Slide Time: 34:32)




Now if you ignore the second-order term and putting $V_{IN}$ equals to $V_M$, $V_M$ is switching threshold, then I get g or the gain to be equals to this big quantity which you see in front of you and it is given by this quantity which you see here. It depends upon all these factors and so on and so forth, so therefore gain is purely determined by the technology parameter and it slightly depends upon the transistor sizing ratio right. So it depends very, very small on the sizing ratio and it is mostly determined by the technology parameters and not by the structural parameters of the device right.

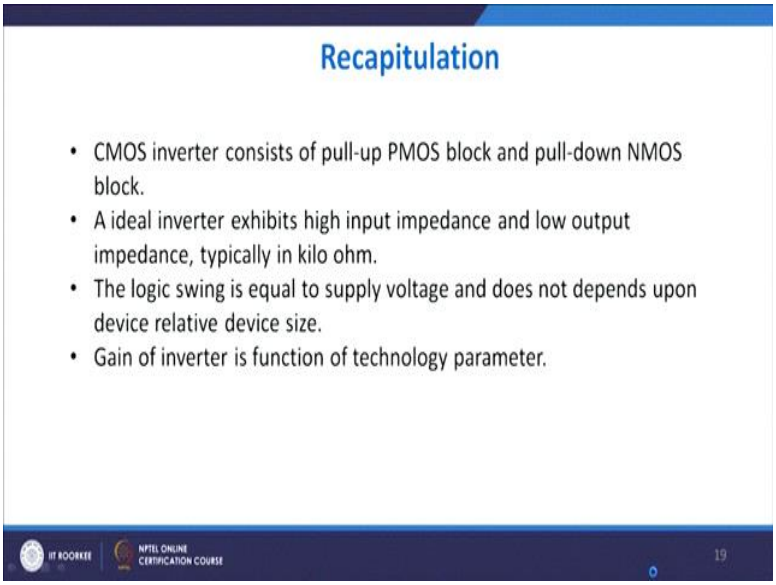
(Refer Slide Time: 35:09)




As you can see here, therefore for varying values of $V_{DD}$ I would expect to see, so which means that the $V_{DD}$ is moving in this direction, higher the value of $V_{DD}$, higher will be the switching threshold and lower the value of $V_{DD}$ lower will be the switching threshold. I hope you have understood why is it. Similarly, if you do a plotting of the graph between various values of $V_{IN}$, then automatically those values of $V_{IN}$ which has got higher values of V out will be shifted to the right and vice versa will be for the blue and black, black and red onto the left and that gives me good idea. It gives me therefore the, by making my inverter pull up and pull down, so why we define?

So my PMOS is referred to as pull up and my NMOS is referred to as pull down fine, pull up and pull down, why pull up? Because PMOS is pulling the voltage to the $V_{DD}$ and NMOS is pulling that to ground and therefore there is a pull up device and, so we define the PMOS to be as pull of device and below one we define as pull down device and we get this substrate configuration here.

(Refer Slide Time: 36:28)



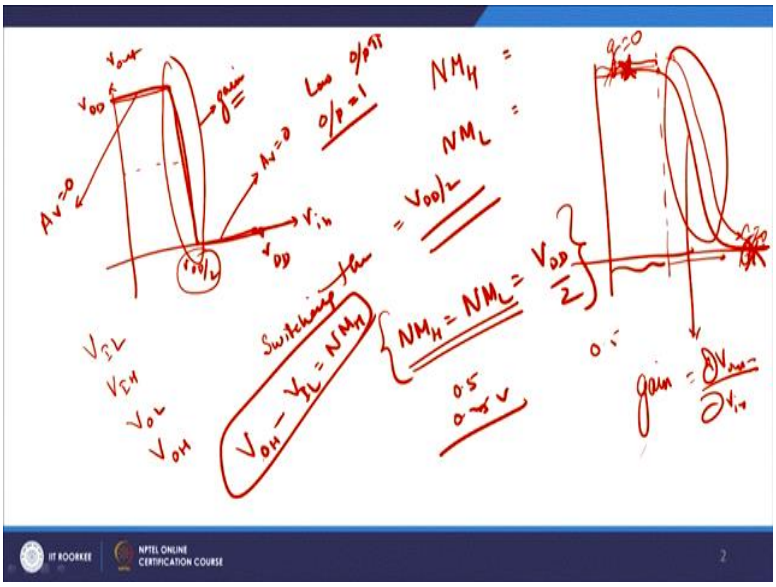

So let me recapitulate what we did. CMOS consists of PMOS block and NMOS block, an ideal inverter has got a high impedance and 0 output impedance. The logic swing is equal to supply voltage, it does not depend upon the device relative size and the gain the inverter is mostly a function of the technology.

I hope you have understood what we have discussed in this lecture; we will come back with another series of lecture on the CMOS basic inverter part II right. Thank you, thanks a lot, thank you very much.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−19**
**CMOS Inverter Basics - II**

Welcome to the NPTEL online course on Microelectronics: Devices to Circuits. We start today's lecture titled as CMOS inverter basics part III, in our previous module we have learned how an inverter works, what is the meaning of inverter which is basically a CMOS inverter and what you mean by $V_{IL}$, $V_{IH}$, $N_{ML}$ and $N_{MH}$. So just to refresh your memory let me give you what the idea is.

(Refer Slide Time: 1:00)




So, generally we define two types of noise margin $N_{MH}$ and $N_{ML}$, so $N_{MH}$ is defined as high noise margin and it gives me an impression that high noise margin primary means that when your input is low right, and your output is high, then at that point of time that means if your output is equals to 1, then how much amount of input right, in the input side how much amount of noise can be given even without changing the output from 0 to 1 right, from 1 to 0.

So let us suppose I have a VTC in this manner right, we have already discussed this point in our previous slide in this manner, then if you look very carefully then typically this much amount of input voltage, even if I give my output voltage will still, so if this is the input voltage I give my output will still be high. But yes, if I across this value and go to this side, the output will fall drastically to this value and output will be equals to 0. So high noise margin is defined as that value or that voltage in the input side, maximum noise voltage

which can be given, so that my output does not change from 1 to 0, right and therefore higher the value of $N_{MH}$ or high noise margin better the design is.

The ideal value of your design is something like this, it is like this, it is the ideal value, which you see. This is $V_{DD}$ right, this is your $V_{DD}$ by 2 and this is also $V_{DD}$, so this is your V out versus V in and this is the profile which you get for all practical purposes, you get the profile something like this, which means that I will expect to see a switching at somewhere around $V_{DD}$ by 2. So we define the switching threshold as or switching threshold as $V_{DD}$ by 2 for ideal case right.

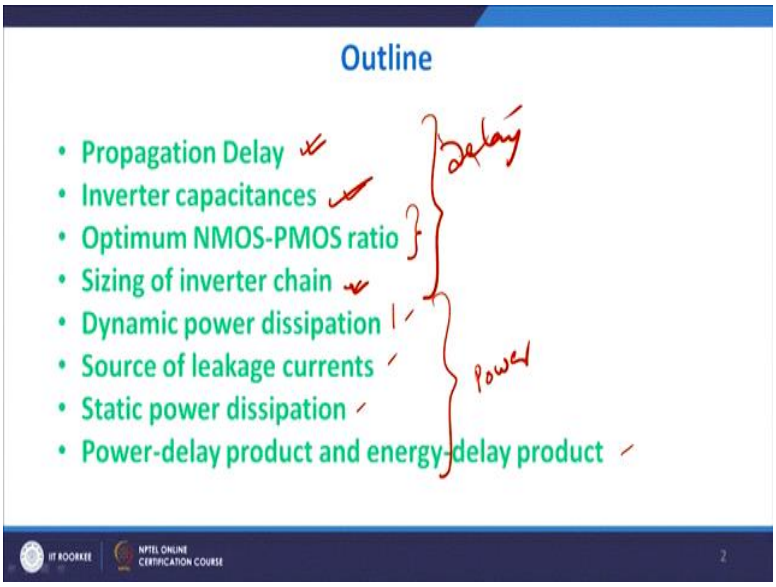
So if you, for ideal case and therefore if you look very closely in for ideal case $N_{MH}$ equals to $N_{ML}$ equals to $V_{DD}$ by 2, which means that the high noise margin and low noise margin are both equals to $V_{DD}$ by 2. Now if you want to lower your $V_{DD}$ for whatever reasons, your noise margin will also be lowering itself automatically and therefore its property of rejection of noise will be compromised, right?

For example if your $V_{DD}$ is 1 volt then $N_{ML}$, $N_{MH}$ equals to 0.5, if it is now 0.5 then this will be 0.25, 0.25 volt, it primarily means that if a noise comes whose equivalent value of voltage is 0.25, I would expect to see output going from 1 to 0, right, so that is what basically the idea is? We also saw in the previous discussion that for digital applications, we generally put it our we bias your device somewhere here or here right. Whereas for analog applications we need to bias it here and at this place only I will get a gain which is given as $\partial$ of V out, $\partial$ of Vin, why? Because at this stage if you look very carefully at this stage my $\partial$ V out is equal to 0 here, $\partial$ V out is also equals to 0 here.

So gain will be 0 here, gain will be 0 here right, so whenever you have this bias right, this and this the voltage gain is always equal to 0 and therefore not used for analog applications, whereas this point is the point where you get a fast change in the output for a small change in the input and this is the point where you define your gain to be there. And if you want to bias, it is an analog device, you need to bias it somewhere in this, this region right. The problem with this region is that it is so unstable that for a small change in the input I will see a large change in the output and that makes it slightly unstable in design, from design or design aspect point of view.

Now, so this is what we have learned, we have also learned how to calculate $V_{IL}$ input low, $V_{IH}$ input high, similarly $V_{OL}$ output low and $V_{OH}$ output high right and similarly we also learned that if you subtract the $V_{OH}$ from $V_{IL,}$ I automatically get $N_{MH}$ and so on and so forth right, we have already learned all these things.

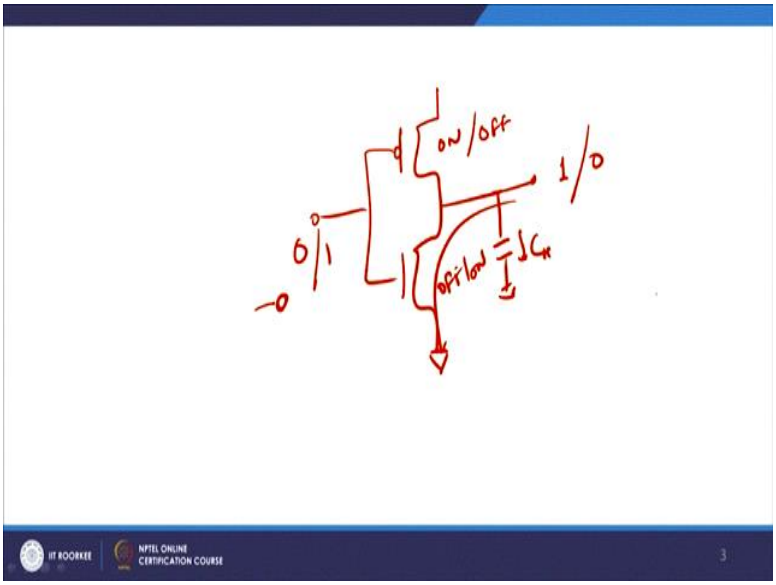
(Refer Slide Time: 5:30)

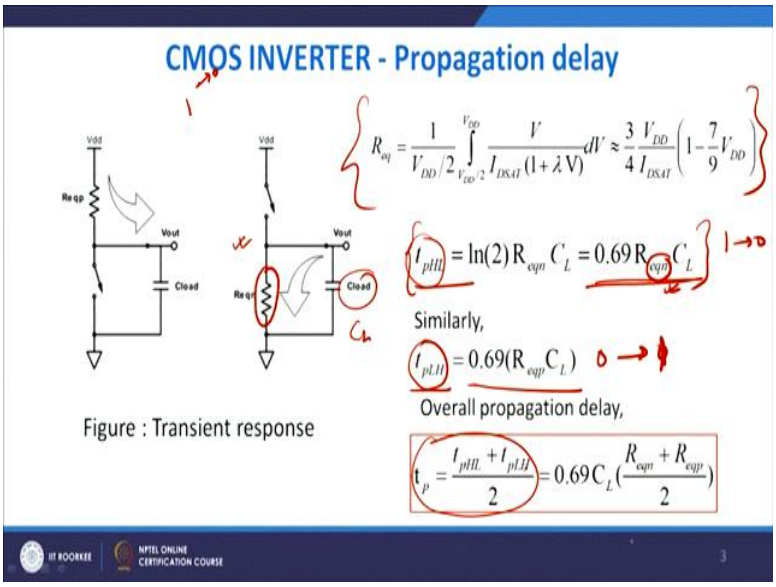



So let me explain to you what are the various things which we will doing today. We will be actually looking into what is known as propagation delay, we will look into inverter capacitances, we will have a look at what is the optimised NMOS by PMOS ratio so that I get the minimum delay position available to me. We will look into the sizing of inverter chain and then start with the power dissipation, dynamic power dissipation, leakage power dissipation, static power dissipation and power delay product.

So the 1st half of our lecture will be concentrating on the delay and the 2nd half will be concentrating on power, so this will be power, where this will be delay. So given a inverter, given an inverter can I optimize its power and delay? So that will be the major motivation behind this module and that will be the major motivation for this module or this lecture series.

(Refer Slide Time: 6:40)




Let me come to how, what is inverter delay. Now as I discussed with you in earlier slide or the earlier presentation that when my PMOS is switched on and NMOS is on right, my output actually falls to 0 right. So we have already discussed this in our previous discussion, in our previous slides that suppose I have got PMOS right, I have got a PMOS here and NMOS here right and this is my PMOS and this is my gate. So I give a, let us suppose I initially add 0 here, so when it is 0, this is on, this is off and this capacitance charges to 1 and therefore output goes to 1, right, this $C_{HS}$ charging.

In the 2nd half cycling 0 to 1, this goes to off, right, this goes to on and therefore this charge finds the discharging path to ground and therefore this 1 goes to 0, this was there. We defined now therefore two delays, one is defined as the propagation delay low to high, high to low

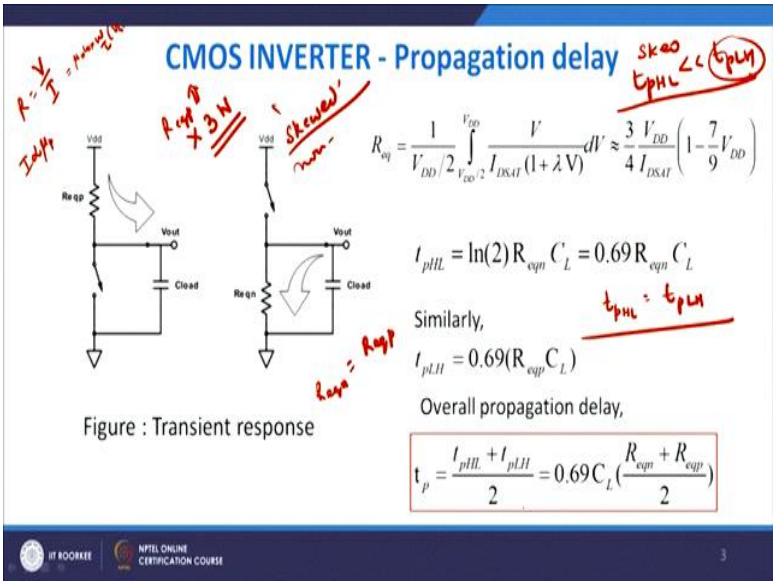
and we also defined low to high right. Now what is high to low? When the output goes from 90 percent of high to 90 percent of low, we define that to as a propagation delay of high to low, right.

So $t_{PHL}$ is basically defined as $t_P$ is propagation delay and HL primarily meaning high to low. High to low means when the output goes from 1 to 0, so that is what we are defining. We are defining, we are not deriving it here, but primarily this is what the equation is 0.69 R equivalent into $C_L$, now what is R equivalent? For example in this case, this is the high to low transition, R equivalent is nothing but the resistance offered by the PMOS, NMOS inverter in its on state, in its on state ideally it should be 0.

But in reality it is never 0 because there will be some transit resistance available to you, as a result there will be always, $t_{PHL}$ will be always have some finite value equals to 0.69 R equivalent into $C_L$, where $C_L$ is basically the load capacitances. So this C load is basically your load capacitance, right and R equivalent is given by this formula which you might have a look into it, but not very critical at this stage of time, more important is this one.

Similarly, when you are discussing output low to high, then we get $t_{PLH}$ to be equal to low to high, again, the formula is 0.69 R equivalent P into $C_L$. So you see it is R equivalent N here because output is going from 1 to 0 right, 1 to 0 and it is equivalent P, so output is going from 0 to 1 right, $t_P$ high to low. Overall propagation delay in the formula is something like this, that your $t_{PHL}$ + $t_{PLH}$ by 2, so we average of these two we take out and we therefore, so this is common, $C_L$ is common you take it outside and then R equivalent P + R equivalent N by 2.

(Refer Slide Time: 9:14)



Now the idea is that maybe we can discuss it here itself, or maybe we will see later on, but if you look at the idea here and let us see what the idea is. That, sorry, see if you want that $t_{PHL}$, if you do not want, if you want this to be true, but $t_{PHL}$ equals to $t_{PLH,}$ which means that high to low propagation delay is exactly equals to low to high propagation delay. Then what should you do? Then you should actually make your R equivalent N exactly equals to R equivalent to P because $C_L$ in any case is equal and that will make you equal propagation delay.

But the problem is that holes, mobility of charge carriers, so what is? It is given as voltage by, so resistance will be voltage by current, applied voltage by current right. Now current in a fat device if you remember by our previous discussion is equals to μn C oxide, W by L $(V_{GS} - V_{TH})^2$ something, which means that the current is directly proportional to the mobility of the charge carriers. In case of therefore PMOS with the majority current carriers are holes, holes have got a much lower mobility as compared to electrons and therefore if you do not do any manipulation in the device structure or the circuit, the current by this.....

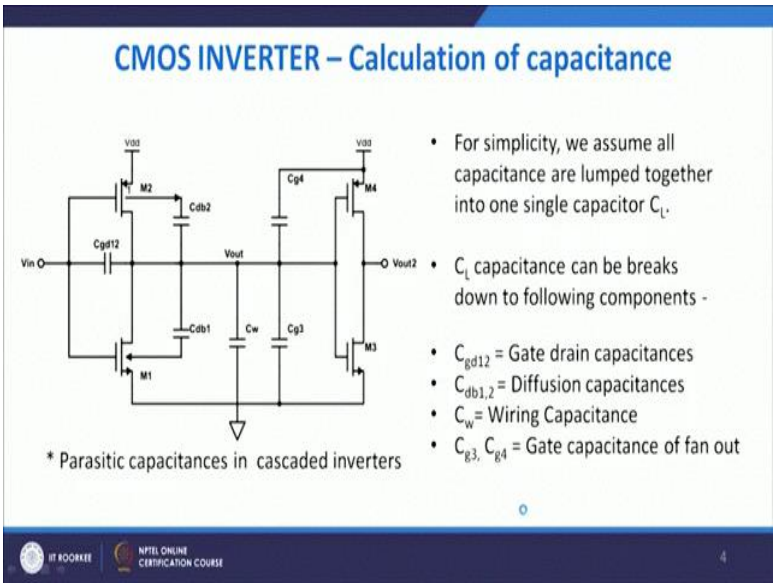
So if the aspect ratio of the device is same NMOS and PMOS is same, W by L ratio is same, voltage also same, then since the mobility of the charge carriers of hole is approximately 2 to 3 times smaller as compared to that of electrons, my current will be also 2 to 3 times smaller and therefore my resistance will be 2 to 3 times larger because it is inversely proportional to the current, right. So this is the problem area which you will face that if you do not do any manipulation, the value of the resistance will still remain the same right, so this is the problem area and the resistance will be different and therefore $t_{PHL}$ will not be equals to $t_{PLH}$.

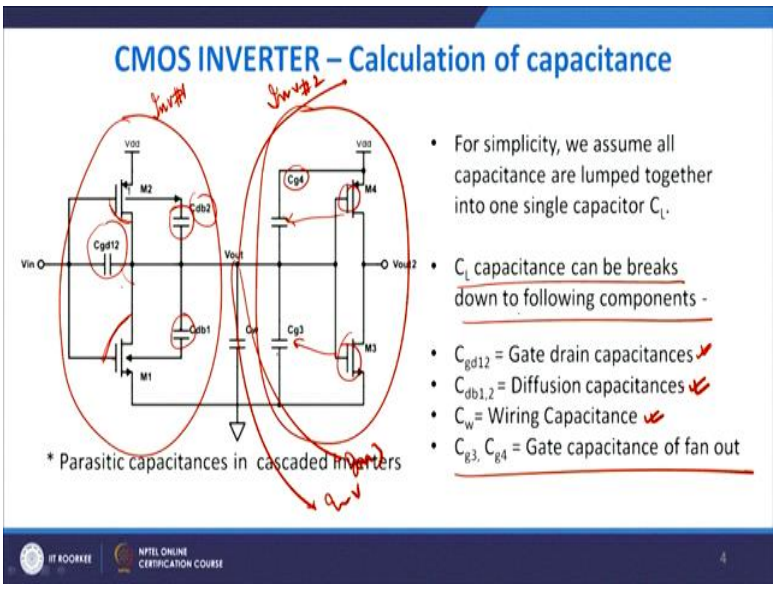
If you want to make them equal then the process which you should follow is something like this: that try to make the, since your resistance is large, R equivalent P is large you need to reduce it by how many times? By approximately 3 times, how can you do that? If you make the width of my PMOS 3 times larger right, then the area actually becomes large and the resistance falls to one-third of its value, to the same value as PMOS. So therefore if you want the $t_{PHL}$ it must be equals to $t_{PLH}$, you have to simply make the PMOS width approximately 3 times larger as compared to NMOS and you automatically get the same values of $t_{PHL}$ equals to $t_{PLH}$, right.

So that is what I want you to say that, if your propagation delay is 1, then you to make this thing, this is also known as a skewed transistor right, skewed transistor. When you have skewed $t_{PHL}$ equals to $t_{PLH}$, non-skewed if you have got then $t_{PHL}$ will be smaller as compared

to $t_{PLH}$. So if you do not have any skew, then $t_{PHL}$ high to low will be smaller as compared to $t_{PLH}$ because this has got a higher resistance and therefore higher current right, so this is just for information sake, a quite important one.

(Refer Slide Time: 12:34)



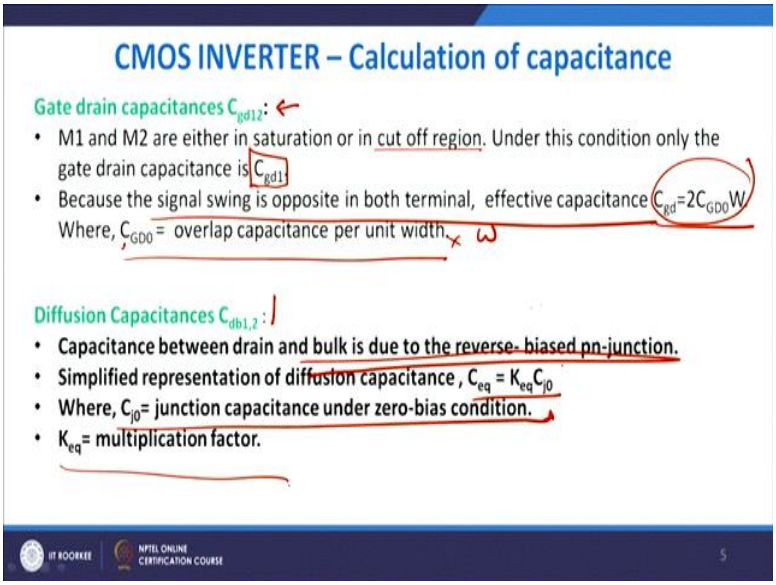

Now with this we are going into another transient analysis, we have done the DC analysis of CMOS inverter, we also wish to have a look into the transient analysis and to understand the transient analysis, we actually have to look into the various capacitive models available within the CMOS inverter. Now, what we, but people generally have done over the years is that all the inverters, all the inverters they have taken all the capacitance together, lumped it together and thrown in the output side and then they have assumed that the inverter itself is capacitance free.

So the inverter capacitance is free, so the inverter is free of any capacitance and all the capacitance will be lumped together and it has been thrown in the output side as $C_L$ or whatever. Now, but if you break it down, then you say this is one inverter, this is one inverter and driving another inverter here, so let us suppose this is inverter 1, this is inverter 2 and it is driving it. So you see this is drain to bulk capacitance, you have gate to drain, you have actually drain to bulk here and then drain to bulk here, drain to bulk for NMOS and drain to bulk for PMOS, this is gate to drain, this capacitance gate to drain right and this is also gate to drain. So these two gate to drain, sorry sorry……, these two gate to drain together form $C_{GD\ 12}$.

Similarly, you will have a gate capacitance here, so only gate capacitances right and this is gate capacitance is falling down, so this gate capacitance is appearing at this particular point, this is gate capacitance is appearing at this point, right. So when you lump this, this, this, this, this and this, together, we get the load capacitance, so that is what I was saying that $C_L$, the load capacitance can be broken down into following components, $C_{GD\ 1\ 2}$, which is this one, gate to drain capacitance, $C_{DB}$ drain to bulk of 1 2, this is your diffusion capacitance, sorry depletion capacitances and this is your diffusion capacitances. $C_W$ is basically the wiring capacitance, so any wire just like your $R_{NC}$ element can be broken down, there will be a wiring capacitance there.

You will have $C_G$ 3 and 4 the gate capacitance of fan out, so this is converted to inverter 2, this is inverter 3 here, inverter 4 here, so this is inverter 3 and inverter 4, then inverter 3 and 4 will have $C_G$ 5, 6, $C_G$ 7, 8 so on and so forth, all will be added together to form the value of $C_L$. So $C_L$ consists of the previous stage diffusion and depletion capacitance and the next stage gate capacitance plus the wiring capacitance, that takes care of approximately all the values of the inverter capacitances.

(Refer Slide Time: 15:26)




Now how do we calculate therefore this $C_{GD}$, which is gate to drain capacitance, right? How do we calculate $C_{GD}$, which is gate to drain capacitance here? Now when M1 and M2 are either in saturation or in cut-off, now then you only have the $C_{GD}$ 1, gate to drain 1, right, why is it true? Because if it is cut-off, then obviously there is no channel formation taking place, when the saturation, the channel is formed, but it screens, it screens of the depletion region from the gate region, right?

Only, that is the reason I am saying only gate to drain capacitance is $C_{GD}$ 1 right, so $C_{GD}$ 1 is only available to you gate to drain, whenever you want to cut-off on the saturation region. Now, since the signal swing is opposite in both the terminals, $C_{GD}$ is equal to 2 $C_{GD}$ 0 into W, where $C_{GD}$ 0 is overlap capacitance per unit width, so you see why we are multiplying it by 2 W or 2 into W? Because since $C_{GD}$ 0, this $C_{GD}$ 0 is per unit width to multiply with width, you get the total capacitance and since you have 2 devices connected to the same input, we multiply that by 2.

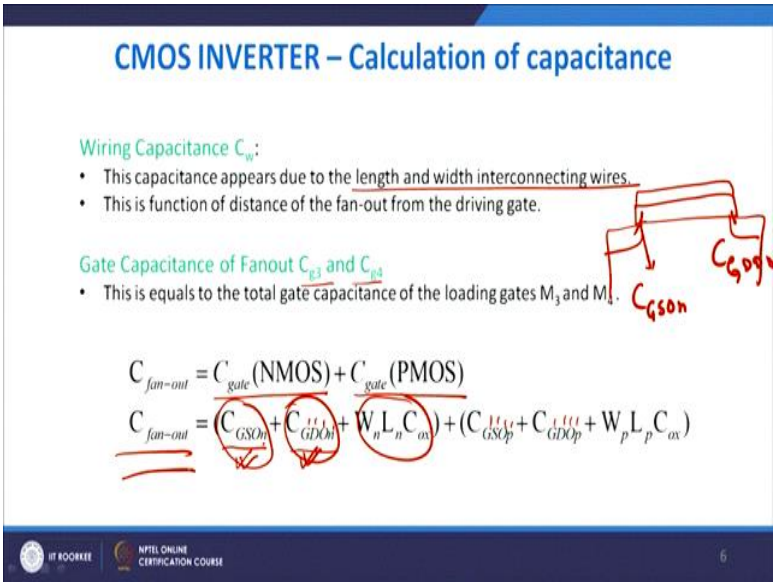
So overall $C_{GD}$ happens to be equal to 2 times $C_{GD}$ 0, right, and the signal swing is opposite, so when this is positive for PMOS, it is negative for NMOS and vice versa, right, and so both will come out, effective capacitances will be just double of that because they will be in parallel to each other.

Now there are 2, so this was basically your drift, your gate to drain capacitance, we also have drain to bulk right, 1 and 2 and this is primarily because as we discussed in our previous term p-n junction reverse bias right because of the p-n junction reverse bias we will always have a

capacitance between drain and bulk and it is given as C equivalent equals to K equivalent $C_J0$, where $C_J0$ is basically the junction capacitance in zero-bias condition and K is the multiplication factor, K multiplication factor depends upon how much amount of bias you have given in the input side.

So the first one is gate to drain, the second one is diffusion one, the gate to drain is a depletion one, this is a diffusion one right. Diffusion primarily occurs because when p-n junction is reverse biased, the depletion region will be formed, which depletion region will be depending on the type of bias you have given and level of the minority and majority current carriers available with you.

(Refer Slide Time: 17:41)



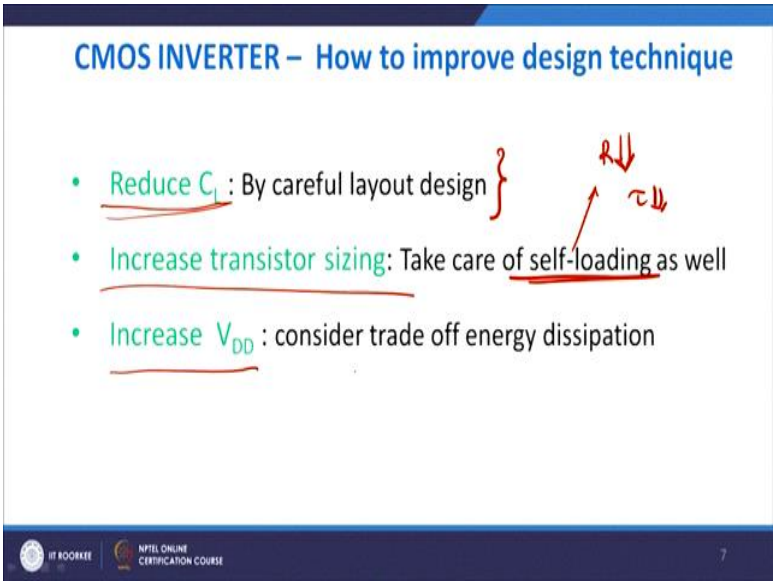

Now to calculate wiring capacitance as I told to you is due to length and width of interconnecting wires and also it depends upon how many fan outs are there. So your fan out is 4, then there will be 4 wires typically which will be emanating from the driver device onto the driven device and that will add up to the wire capacitance there, so larger the length more will be the capacitor.

The gate capacitance $C_G$ 3 and $C_G$ 4 comes out from the gate capacitance of the subsequent stage and therefore fan out is equal to C gate into NMOS into C gate into PMOS, I think, very simple and straightforward way of looking at it, C gate can be again broken down into two parts $C_{GS}$ ON and $C_{GD}$. So you see the channel can be broken into two parts, one contribution between gate and source, so that is this one in the on state and the overlap capacitances, and then gate to drain overlap capacitance, so this is O means overlap, gate to

source overlap capacitance for N channel and gate to drain for overlap for N channel, this is gate to source overlap P channel, gate to drain overlap P channel, right?

So this is basically the overlap channel, this is also the overlap channel which you see and this one is primarily the gate oxide, so when you have plotted, when you have actually drawn this, so let me say you have drawn something like this and you have this thing, then it is something like this that, this overlap here drain side and source side, this is basically your, so this is actually your $C_{GS}$ ON and this is your $C_{GD}$ n right, D n, D0 N, so that plus W n into L n means, width of NMOS into length of NMOS. Since C oxide is the oxide per unit area, multiply with that you get the total oxide, you add those 2 oxides and you get the total fan out oxide which is available with you, and it is quite a large sum which you see.

(Refer Slide Time: 19:36)




Now, so how to design a very good inverter? If you want to improve the speed, of course the best way to do that is reduce $C_L$ and this can be done by layouts, so either by layout or by critical sizing of the transistors and you can reduce the value of $C_L$. Similarly, the idea was is, if you remember $t_{PHL}$ and $t_{PLH}$ was depending on the value of R equivalent N, which means the resistance offered by the transistor. In the 1st case when you are reducing $C_L$ right, this can be the, so if you reduce $C_L$ obviously your $\tau$ reduces.

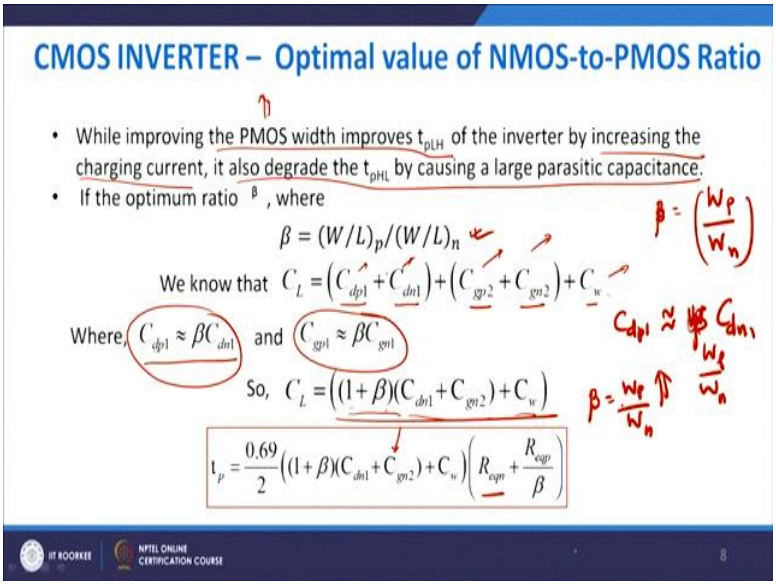
Similarly, if you increase a transistors size, well, that will make your R actually go low right, R go low, but when R goes low obviously your Tau reduces and you can work at much faster pace but then be very careful about what is known as self-loading. Right, what is the meaning

of less self-loading is? That you are increasing your W right, fine, why you are increasing your W?

So that the area under the gate goes on increasing and you automatically have a smaller resistance available to you, but then what happens is as we go on increasing the value of R, W you also tend to increase the capacitance of the gate. So gate has to drive in a much better manner in order to invert the channel and therefore your actually the first point is neglected and you will not able to reduce the value of $C_L$, your $C_L$ value starts to rise again because the W has increased drastically, this is known as self-loading right.

Why should you increase $V_{DD}$? Very important that if you increase $V_{DD}$ for the same amount of charges the current will be large, you got the point. So if the current is large, $I_D$ is large, then your charging and discharging process can be done much faster right because for the same amount of time more charge will be collected or discharged if $I_D$ is typically very large and $\tau$ therefore reduces. So therefore if you want to increase the speed 3 things, reduce capacitive loading, increase the transistor sizing by increasing the W ratio, W width and then increase $V_{DD}$. But if you increase $V_{DD}$ you will have also higher power dissipation in a much larger manner and so power dissipation will be much larger in that case, right?

(Refer Slide Time: 21:54)




Ok, let me see the optimal value for NMOS to PMOS ratio and now as I discussed with you, while improving the PMOS width right, improving the PMOS width improves $t_{PLH}$ low to high, if you increase PMOS width $t_{PLH}$ will go on reducing by increasing the charging current as I discussed with you, but it also degrades $t_{PHL}$ by causing a large parasitic capacitances,

why? Because as you go on increasing the PMOS width right, the PMOS itself might be doing very good, but then it will start loading your external load capacitance.

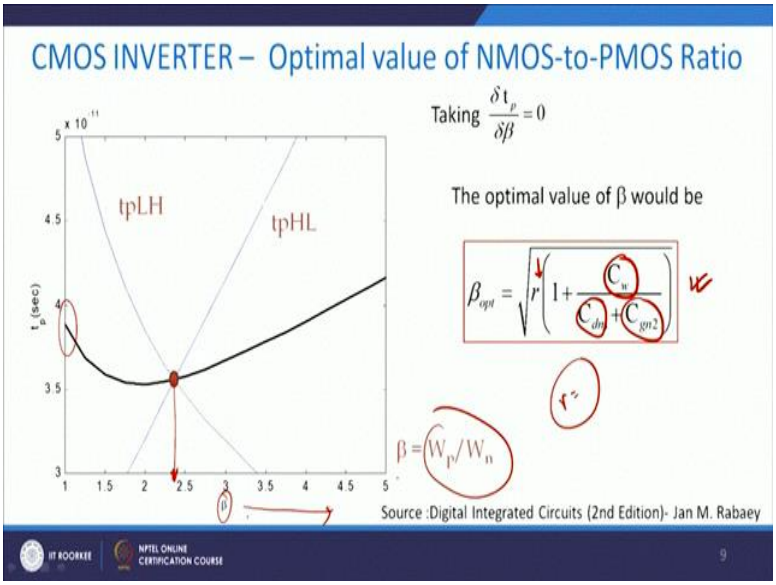
So the load capacitance will be, parasitic capacitance will be larger and larger. So at one hand, you are actually reducing the value of R equivalent P but in the same instance of time your capacitance is going on increasing, so we do not know exactly whether it will be always win-win situation, so you need to optimize it. Now if the optimising ratio β is given W by L of P upon W by L of N, we know that CL is given by dpl 1, dn2 plus gp 2 gn 2 plus wire, so this is your gate of the second fan out transistor and this is your dp, dn of the first transistor and this is wire transistor, where C dp 1 is β times C dn 1.

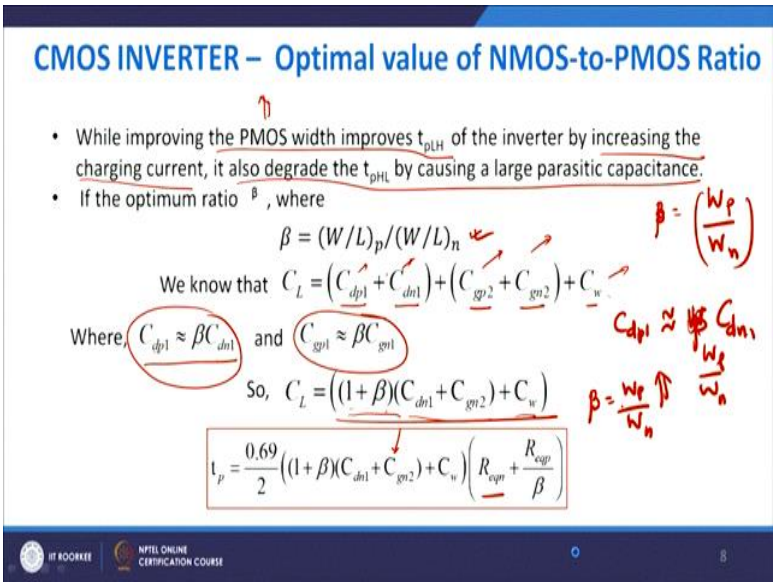
Why? Because the ratio, if you look at the β ratio, β ratio is given as, so if L is constant, it is nothing but $W_P$ by Wn right, so if you want to find out C dp, C dp1 will be approximately equals to W n times right or maybe β times C dn 1 right. So this is nothing but we develop $W_P$ by Wn right and therefore you can automatically say that this holds good, which means that the output capacitance will be approximately equals to β times C dn 1 and C gp will be approximately equal to C gn1 right.

So this is drain, so this is C dp 1 means this is basically your depletion capacitance of PMOS 1 will be β times depletion capacitance of 1 because width has increased by β times. Similarly gate capacitance of the 2nd case will be approximately β times because you see capacitance is directly proportional to the width and therefore you directly multiply, so if you look at $C_L$ and if you just put this into this formula, I get 1 + β times C dn 1 $C_{gn}$ + $C_{gn}$ 2.

Now if you $t_P$ if you find out 0.69 by 2, 1 + β because remember, it was $C_L$, so this is $C_L$, this is $C_L$, this $C_L$ comes here and into R equivalent plus R equivalent P by β, you understand why it is by β, because you actually multiplied this β ratio by this thing, which means that beta is equals to $W_P$ by Wn, so higher the value of β when you add it, you will divide it by β right, and therefore you get $t_p$ equals to R equivalent n plus R equivalent Q by β.

(Refer Slide Time: 24:46)

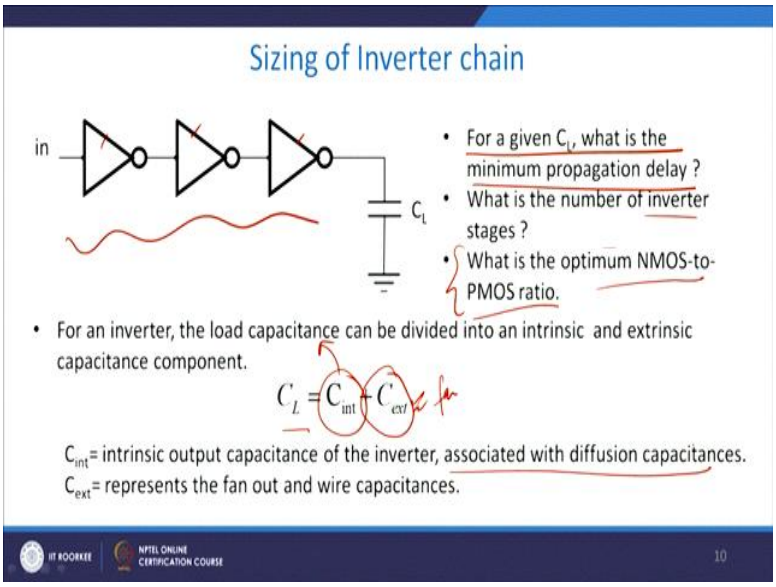



Now if you differentiate $\partial\tau_P$ with respect to β, I get equals to 0, the optimal value of β comes out to be this value, where β equals to $W_P$ by Wn and R if you look is basically the ratio which you see, R is basically the ratio which you will get this ratio, R is basically your, R equivalent P by R equivalent n is the ratio of R right. So for a fixed value of R, right, wiring capacitance it depends upon the diffusion capacitance and the gate capacitance of the second one and if you plot the propagation delay with respect to β, we see that somewhere near 2.5 to 2.5 of β I will expect to see $t_P$ equals to 0 and this is also true from my previous understanding that if I am able to sustain the width of my PMOS 3 times more as compared to NMOS I automatically get a very reduced profile.

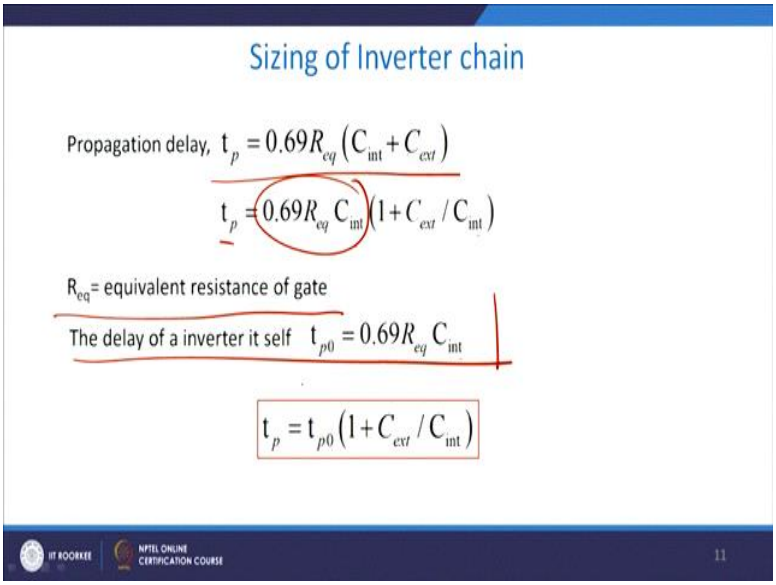
Therefore why, think about it. Why with therefore larger beta more than 2.5, I actually see a reduction in the value of $t_P$, for a low value of β is very simple, for low value of β since my width my PMOS is smaller as compared to NMOS, relatively smaller, it is not 3 times, it is less than 3 times, therefore the mobility is so large, so small that the current is very small and therefore resistance is high and that is the reason you get a larger $t_p$. And you minimize somewhere around 2.5, beyond 2.5 your actually $t_{PHL}$, your $t_{PHL}$ starts to go grow, high to low starts, goes high, and as a result the overall gain starts or overall delay starts to become larger. So if you are biasing your device, please bias it in a manner such that it is approximately equals to 2.5 to 3 times, right?

(Refer Slide Time: 26:35)



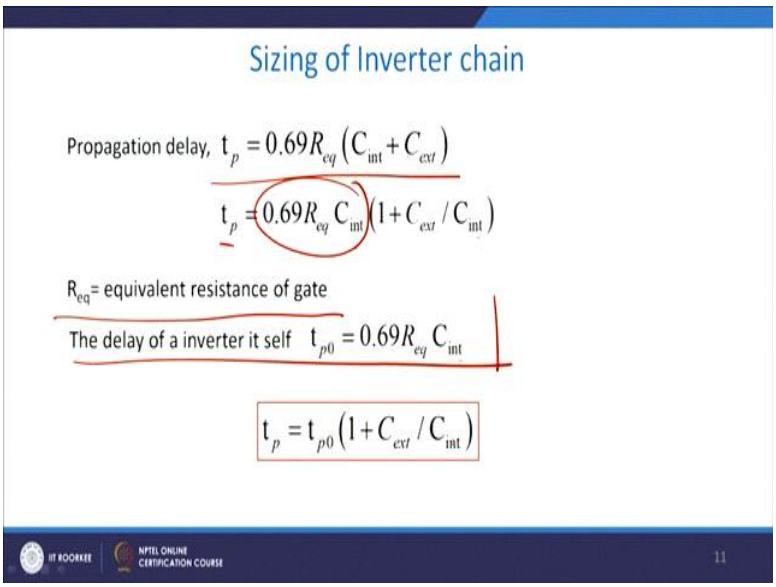

Now, let me come to the sizing of the inverter, the idea is for a given $C_L$ or for a given load capacitance what is the minimum propagation delay? Right, How? If you have a buffer? Means you have a large 1, 2, 3 inverters right and therefore you need to find out what are the optimum number of stages and what is optimum NMOS to PMOS ratio. As we have discussed already that a load capacitance can be distributed by internal capacitance or intrinsic capacitance and external capacitance, so this is primarily because of fan out, this because of fan out and this is because of intrinsic output capacitance of the inverter associated with the diffusion capacitance and so on and so forth, so I get $C_L$ equals to $C_{int}$ plus $C_{ext}$.

(Refer Slide Time: 27:18)




So propagation delay tp will be equals to 0.69 R equivalent to C internal plus C external we have seen this point. If C internal as output, then I get tp equals to 1 plus C internal by C external, where R equivalent is basically the equivalent resistance of the gate, whatever gate your trying to use. So the delay of the inverter itself is given by this value that tp0 equals to 0.69 R equivalent to C internal and therefore we replace this by tp0, tp0, you understand that the inverter itself will have some intrinsic delay and that is given by 0.69 R equivalent to C internal and I get tp0 C external by C internal.

(Refer Slide Time: 27:53)



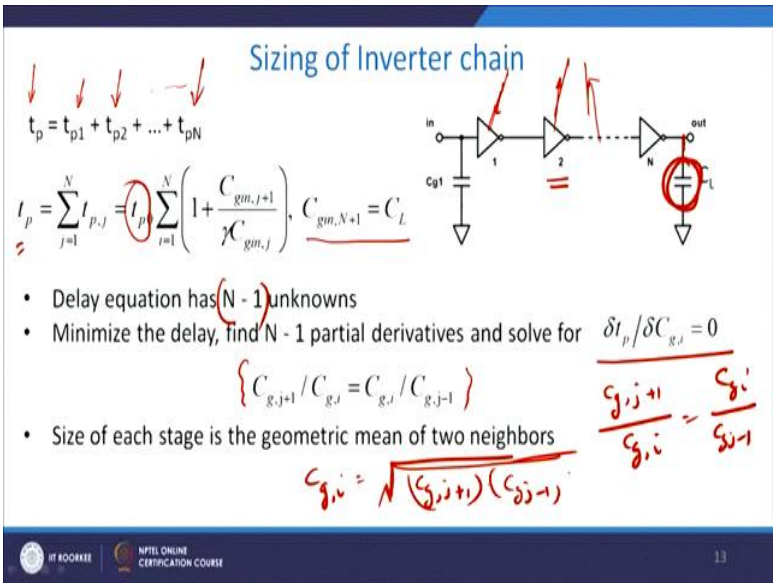

Now therefore sizing up an inverter will reduce its delay because its $R_S$ getting reduced but this is same instance of time, its input capacitance also increases, so we are not very sure

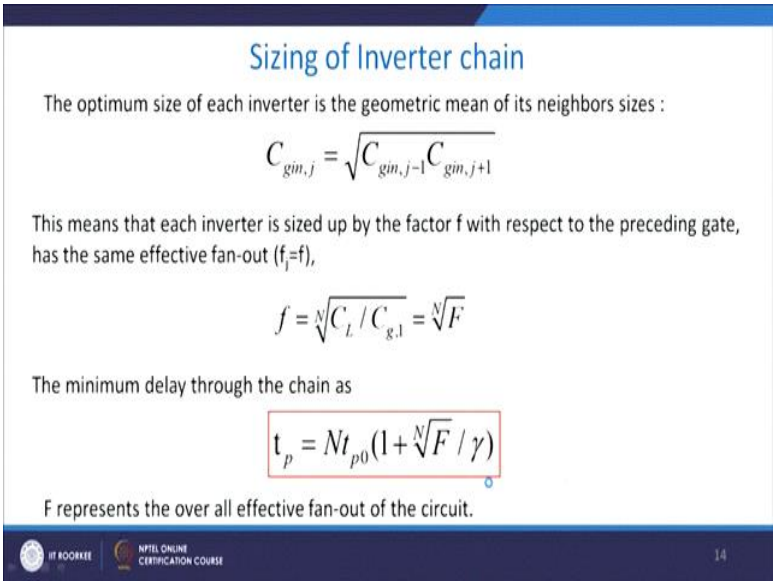
whether what is happening, so there must be some optimal value of increase or the sizing of the inverter which will ensure me a minimum delay.

Now let us suppose my the input gate capacitance and intrinsic output capacitance obviously are the function of gate size, so therefore I say that C intrinsic is γ times Cg, where γ is basically a proportionality factor depending upon the function of technology, so whatever technology you want to use you can use and therefore I can safely write down tp to be equals to tp0, 1 plus f by γ right, what is f? f is C external by C internal.

So by, this is C external by C internal is f right, f is C external divided by γ if you write down, then I get tp equals to this, this, which means that the delay of the inverter is a function of external load and its internal capacitance, intrinsic capacitance. So if you take an inverter, right, then the delay of the inverter is basically ratio between the external loads $C_L$, load capacitance and its intrinsic capacitances, what is intrinsic capacitance? Again diffusion capacitances, gate to drain depletion capacitances, overlap capacitances, so on and so forth. So it depends upon the ratio of your output capacitance to input capacitance for a basic inverter.

(Refer Slide Time: 29:29)






So let us see how we can size an inverter chain. So what we do is that tp which we say is the total delay is equals to tp1 for the 1st gate, 2nd gate so on and so forth till nth gate, for nth gate you go on adding all the delays together, so tp therefore will be given as this quantity, summation tpJ, J equals to 1 to n, whereas tp0 is the intrinsic delay and summation 1 to n, if you do it we just saw that, if you take for example this inverter, 2nd inverter, then 1 plus $C_{g\,in}$ means input gate capacitance of this one divided by $\gamma$ times $C_{g\,in}$ of the previous one, this one, right and this we get.

So I get $C_{g\,in}$ N +1 equals to $C_L$, which means that the last capacitor you are loading with load capacitor $C_L$, fine. So the delay therefore has got N - 1 unknowns because there are N - 1 transistors available to it. We need to solve therefore $\partial$tp, $\partial C_{g\,i}$ and equate it to 0 and if you do that, we get something like this into our consideration that $C_J$, $C_{g\,J}$ + 1 upon $C_{g\,i}$ equals to, so I get $C_{g\,J}$ + 1 divided by $C_{g\,i}$ equals to $C_{g\,i}$ divided by $C_{g\,J}$ - 1. So if you take $C_g$, if you want to find out $C_{g\,i}$ it is nothing but square root of $C_{g\,J}$ + 1 multiplied by $C_{g\,J}$ - 1, right.

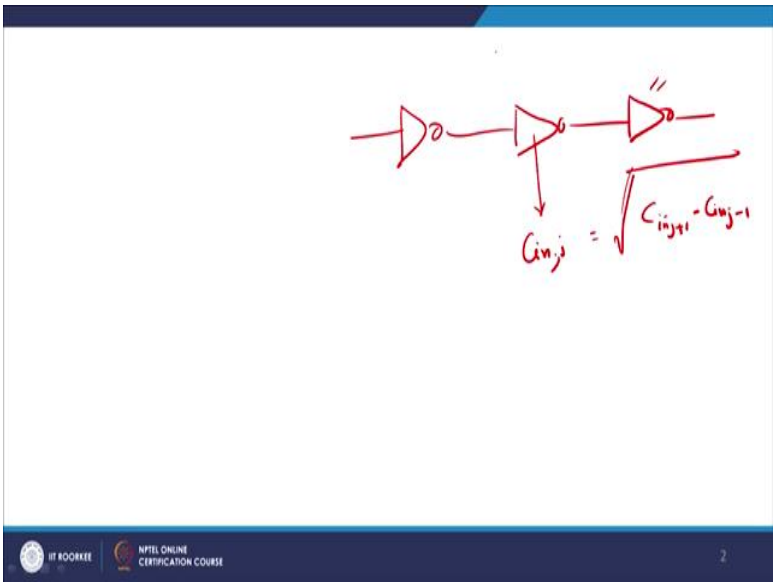
So if you want to find out the input gate capacitance of 2nd one then find the geometric mean of the 1st and 3rd one right, with that geometric means you see square root of this gate capacitances. If you are able to fix the value of second one such that it is the square root of the two subsequent ones, we automatically get a reduced factor. We will take care of the next profile in the next subsequent lecture. Thank you very much!!!!!!!

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture – 20**
**CMOS Inverter Basics - III**

Hello everybody and welcome to the NPTEL online certification course on Microelectronics: Devices to Circuit, we start today with CMOS inverter basics part 4, what we will be doing in this case is we will be looking into the from the time when we left in the previous case, we saw that in the previous interaction that for an optimized design in terms of reduced delay reduced delay between primary input and output.

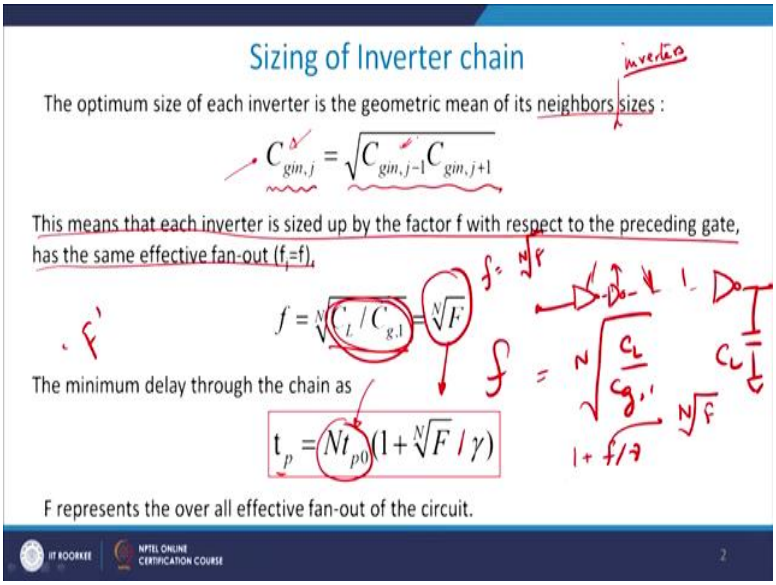
The gate capacitance of the middle transistor or a middle inverter in a chain of inverter should be in a geometric progression of the subsequent and the president inverter.

(Refer Slide Time: 01:18)




So if we have three inverters, so what I wanted to say was that if we have got two inverters so you have got one inverter, two and then three here, right, then if you want that the delay should be minimized, then try to keep the input capacitance of this inverter, right, J let us suppose, it is J to be equals to square root of Cin J plus 1, this 1 multiplied by C in J minus 1 which is this one, right. So we have seen that and therefore, we… how we got it? We have derived in the previous lecture this basic concept that if you want to do this you have to do these basic concepts.

(Refer Slide Time: 02:02)




So let me start from there and show to you that therefore, the optimum size of each inverter $J^{th}$ inverter is the geometric mean of its neighbouring inverters or neighbouring sizes, right, it should be neighbouring inverters as well, right. So what we are telling you is that $J^{th}$ inverter will be basically if you are able to fix its value of input capacitance to be equals to this, then possibly you will get a reduced delay.

Now, this means that each inverter is sized up by a factor of widths F with respect to the preceding gate, has the same effective manner which means that you see, so what I was saying was that you had inverters like this chain of inverters and it ended up at $n^{th}$ inverter, then the last one you had a load capacitance $C_L$ and this was V out.

So, if each of the inverters gives you a, this should be geometric mean of this thing. Similarly, this should be geometric mean of this and this, then what we finally get is that if you take an overall chain do a cross multiplication, then you get $C_L$ by $c_g1$ square root of $n^{th}$ square root $n^{th}$ root, this will be the function f, fine? I think it is clear to all of you, why?

Because as I discussed with you that as you move from the first inverter to the last inverter in a chain of inverters then you have to progressively size it up as you go from lowest to highest value, but how you will is… the rate at which you will be resizing it up will determine, whether you are optimizing a design or not?

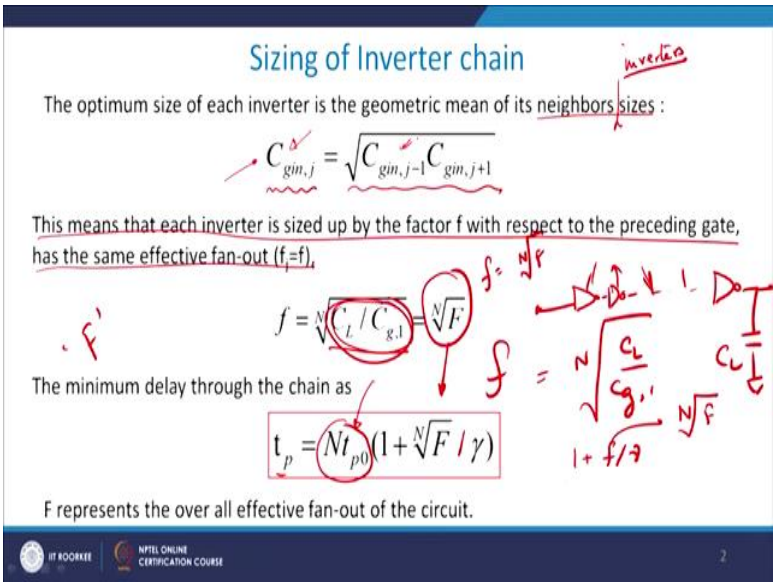
Now, if your last inverter was terminating into a load capacitance whose values equals to CL and the first inverter was basically equals to $c_{gi}$ input capacitance, then $C_L$ by $c_{gi}$ or $c_g1$ $n^{th}$ root of that will be your f factor which is basically your sizing of factor f. So if this is capital

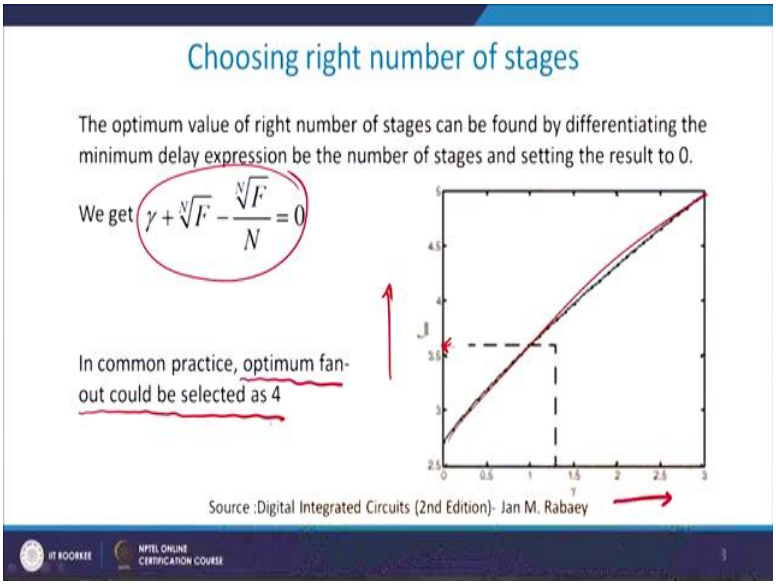
F which we term, let us suppose the ratio is termed as capital F also referred to as fan out, electrical fan out, then we refer to small f to be equals to $n^{th}$ root of capital F, which you see in front of it, right.

Once you have known this you just have to feed it into your original equation of chain, so you get tp equals to Ntp0 because see if each inverter has an intrinsic gain an intrinsic delay of tp0 and if you have n such chain, obviously that delay will be obviously larger than Ntp0. And therefore, you see Ntp0 coming here as a quantity.

You also have therefore, this quantity therefore coming up here, divided by γ from where they got this? From the previous, our previous understanding we got this, that if this is so I was saying it is F by γ if you remember. Now this f is nothing but a $n^{th}$ root of f, where capital F is given by this quantity and therefore, I get this by γ to be equals to tp, where tp is the overall delay on this thing. So F represents the effective fan-out, right and we get effective fan out.

(Refer Slide Time: 05:12)






Now what you do is that you need to differentiate the previous equation which equation this equation tp with respect to N and make it equal to 0 and then minimize it and then if it say equal to 0 I get this equation into consideration that γ plus something minus $n^{th}$ root of F by N equals to 0.

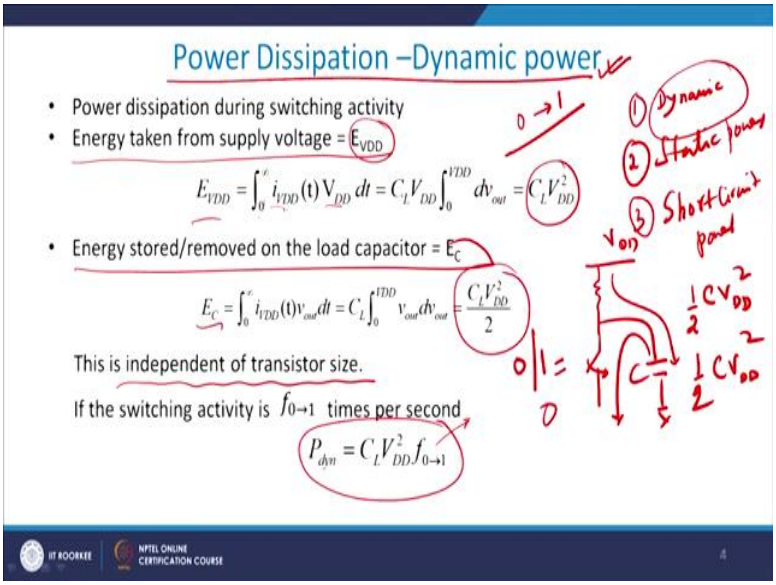
Now this typically means that it has been shown that therefore, if you plot F or if you plot γ on the x-axis and you plot the delay on the y-axis typically you get a curve something like this you will get a curve something like this, somewhere around if you fix γ equals to 1.5 you will get a delay of approximately 3.5 maximum delay which you sent approximately 4.

So in common practice it is always advisable to keep the optimal fan out as equals to 4, , right because this is what you get so approximately 3.5 to 4 you get if your γ is goes to 1.5, gamma is basically your factor which gives you the output ratio of your capacitances.

Now with this knowledge we therefore, with this knowledge we therefore tell that the optimal fan out should be approximately equals to 4 which means that if a single transistor is there and you want to optimize the output you just have to have fan out of 4 available with you and that will give you the best results as far as this design is concerned, right.

Let me now come to an important topic, so we have understood what the delay is all about, how to size a transistor to get the optimal delay. Now a very important property of inverter is that it is basically switching from ON to OFF state or 0 to 1 state in the output side. So since it is switching back and forth from a high value to a low value and vice versa, there is always an energy which is being dissipated or a power which is being dissipated due to the switching action and that power is basically known as dynamic power, right.

(Refer Slide Time: 07:28)




Typically a transistor or an inverter will have three types of power, one is known as dynamic power, right, the other is known as basically a static power, this is static power and third one is actually referred to as a short circuit power, we will explain each one of them individually, but let us first understand dynamic power, right.

Dynamic as the name suggest is basically the power which the CMOS dissipates or draws from $V_{DD}$ rail when you do have a input which is varying from 0 to 1 and 1 to 0, so you do have switching characteristics which is available with you. If you see then as all of you are aware of that energy taken from a supply, suppose $EV_{DD}$ is the energy taken from the supply $V_{DD}$ must be equals to 0 to infinity, I current multiplied by voltage, V into I is basically the power which you get and if you integrate from 0 to T in time domain, so what will happen is this if you if you break down I and V, I get $C_L$ times $V_{DD}$ into integral 0 to $V_{DD}$ dV out, right, if you solve it, I get $C_L V_{DD}$ square, right.

So if you remember from our basic inverter, so I had this, right and I had this, and then this. So every time you are actually charging this capacitor, you are taking it half $CV_{DD}$ square where C is this capacitance, power from the $V_{DD}$ rail, right. In the next half cycle when this was closed the same used to go here and you just remove this half $CV_{DD}$ square, so you take half.

So in one cycle you take half $CV_{DD}$ square and you throw it in the next cycle to the output side, so when you have input equals to 0 your capacitor charges to half $CV_{DD}$ square and

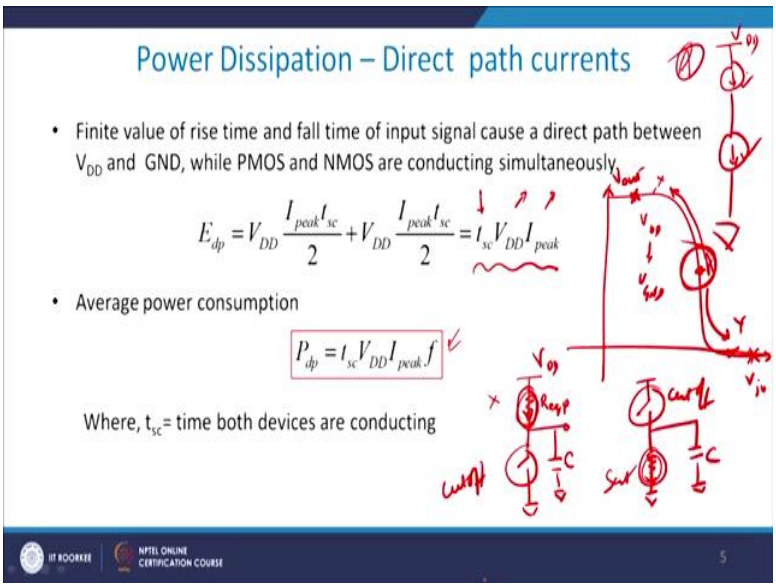
energy is stored is half $CV_{DD}$ square, from where does it come? It comes from the $V_{DD}$ rail, right which is the power deal.

In the next half cycle when input is equals to 1 and output goes to 0 then the charge accumulated on to this is dissipated across the ground and therefore, it goes to 0 and therefore, I get the total power dissipated is equal to half plus half is basically $C_LV_{DD}$ square over two cycle of ranges.

Now energy stored, removed in a capacitor is given by as I discussed with you is $C_LV_{DD}$ square by 2 because half it is there, this is independent of the transistor size. So please understand that your effectively the charging or discharging of the capacitor is independent of the size of the transistor, so it can be any size which you can choose and you can actually have a size which is available with you.

Now, if switching activity is 0 to 1, so you see an important point is that only when your output goes from 0 to 1, right that is the time when you are accepting power from the VDD rail. So please understand the dynamic power, which if you are not dissipating you are getting it from the $V_{DD}$ rail is only happening when your output is going from 0 to 1 because that is the point when your capacitor is getting charged through PMOS and you are drawing power from the $V_{DD}$ rail and that is the reason we say P dynamic to be equals to $C_LV_{DD}$ square multiplied by the frequency of 0 to 1, right what is the frequency of 0 to 1 which is there with me.

(Refer Slide Time: 10:50)



Now that is known as that is as the dynamic power dissipation. Now what happens is that we have discussed this point earlier also, that when you draw the voltage transfer characteristics, right this is what you get, this is your V in, right, and this is your V out, at this stage NMOS is cut off and PMOS is switched on, and at this stage PMOS is cut off and NMOS is switched on and that is the reason grounded.

And therefore, you do not have a direct path between $V_{DD}$ and ground, so if you at this point if you, let us suppose this point X at point X if you want to find out what is the output characteristics it looks something like this, it looks like this and then this is open and this is ch, right, so this NMOS is opened and this is basically by R equivalent P, this is $V_{DD}$ and this is my C capacitance available with me.

So which means that this is saturated and this is cut off, what happens at Y which is this point? Y will be this will be open and then I will represent it by a resistance here, and there will be capacitance here C, so this will be then saturated and this will be cut off, fine, but these are the two extremes where you are seeing it, somewhere in the middle, somewhere here I possibly will be seeing that both will be acting as a current source, both will be saturated, somewhere at this point say A at point A.

As you can see since two current sources in series is basically intrinsically an unstable situation, obviously you do not stay here for a quite long time, a small change in the input will result this A shifting to this point or to this point, but then if you are working at A then you do have a direct path between the $V_{DD}$ and the ground rail, please understand this properly that whenever you, this is somewhere when both the devices are in saturated state at point A, at this point A when you in saturated state both are in on state, on state primarily meaning is that it is basically behaving like a current source in a saturated state and as a result you will have large amount of current flow.

So even I can represent this by a current source both of them right those output impedance is typically very large. Now this will, then so there is a short circuit path between $V_{DD}$ and ground, right, and as a result you will see a large shift current flowing that current,,,, flow of current is given by this formula. So $V_{DD}$ is the applied voltage I peak is the peak current multiplied by how much amount of time this current is flowing is tsc a short-circuit time and therefore, it is given by energy Edp, right.

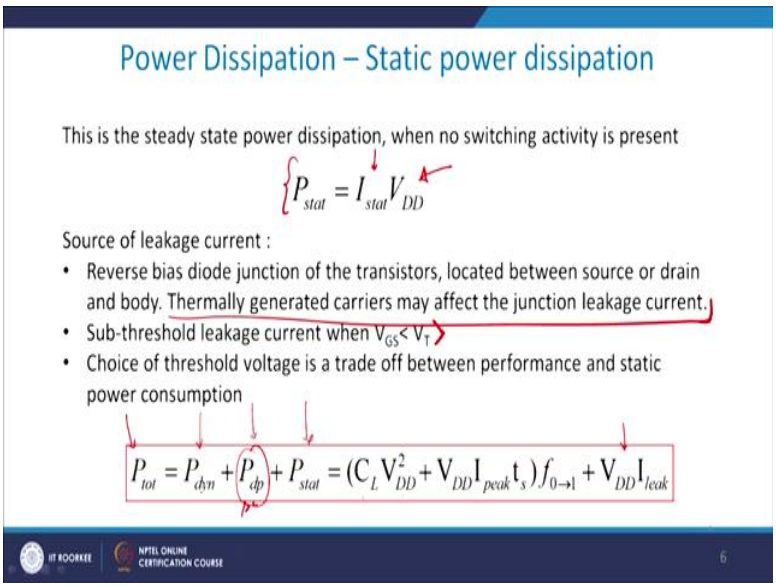
So Pdp will be multiplied by frequency tsc $V_{DD}$ I peak into f so if you go once from a high to low and then you go to low to high, then you are actually traversing through short circuit path to twice and therefore, it is depending on the value of this thing tsc is the time when both the devices are conducting, so these are the two time when they are conducting.

So we have discussed basically the dynamic power we have also understood what is the short circuit power, electrical static power, static power or the steady-state power is defined as that power when you are not in the dynamic position which means that your input is not varying but it is fixed or even your device may be in the cut off state and you expect that there should not be any current flow through the device and I would expect to see that the power dissipation because of should be equals to zero, but it is not and the reason is something like this.

The reason is that whenever you assume that the device is off by saying that the gate voltage is falling below threshold, but if you go to literature available literatures, you will see that not necessarily the device is actually off when your gate voltage falls below threshold. So if your threshold voltage is say 1 volt and you are at 0.9 volt, then as per our understanding it should be off, but in reality there is some sub threshold current still available with there.

And so though you think your device has been switched off, in reality the device is still on and that gives you so small current internally, right that is what is known as a sub threshold leakage when your $V_{GS}$ is less than $V_{Th}$, right.

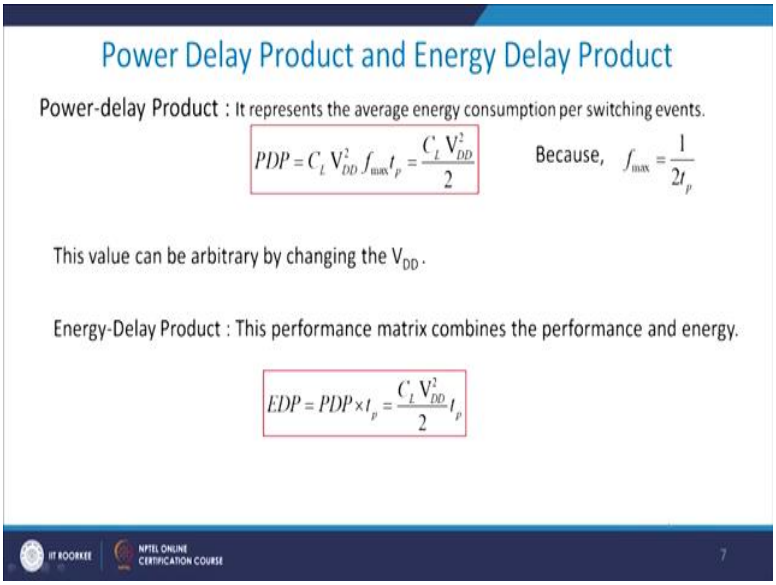
(Refer Slide Time: 15:34)

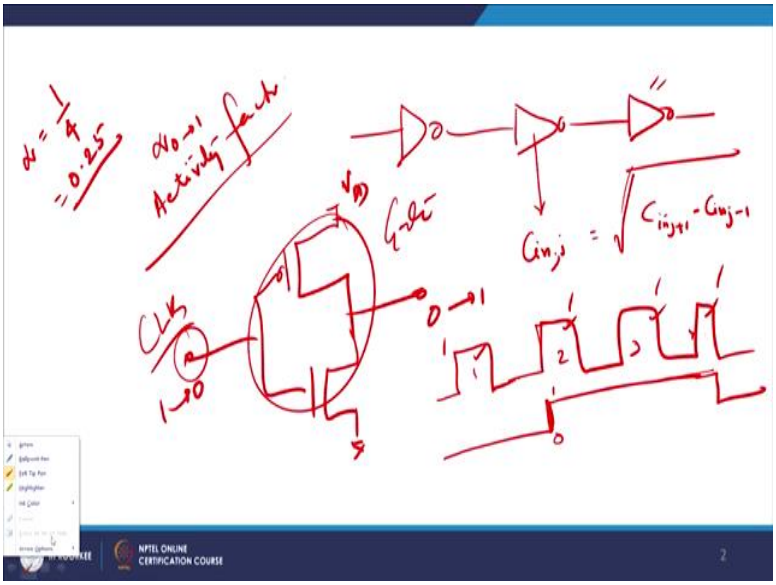


Similarly you will have thermally generated carriers which will affect your junction leakage, remember your base to the source and drain is reverse bias junction you remember and therefore, the depletion thickness is quite large there, but for minority current carriers it is a basically not a hill, but a slope so if the temperature increases by even 10 degree I would expect to see almost doubling of the value of your minority current carriers which might result in a large current.

So these currents which is basically the virtue of reverse by a saturation current is primarily because of these reasons, so I define $P_{stat}$ as the static power to be equals to $I_{stat}$ the static current multiplied by $V_{DD}$. So we define P total the total power which is with us as equals to dynamic power plus your static power plus your this is short-circuit power, right, it should be SC short-circuit power. So I refer to as $C_L V_{DD}$ square plus $V_{DD}$ into Ise into tsc into f 0 to 1 and this is $V_{DD}$ into $I_{leak}$.

(Refer Slide Time: 16:52)



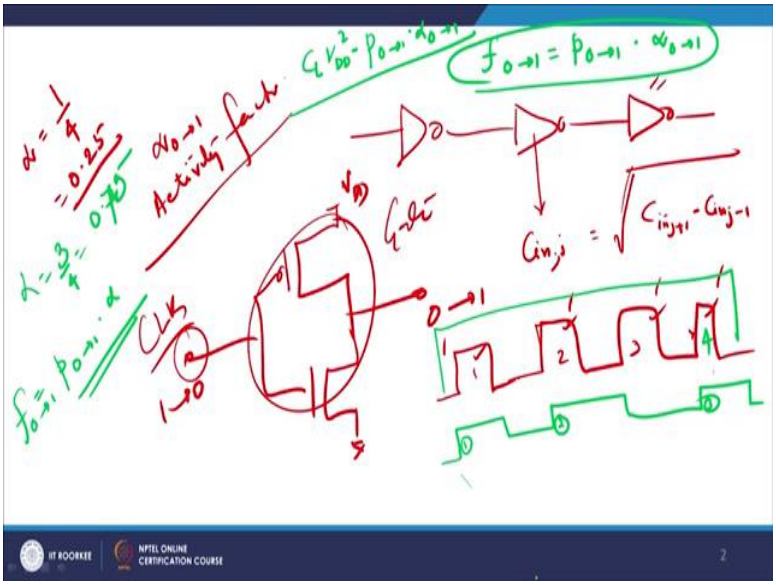



I will explain one important term which is well it is not here, but I will explain it later on maybe I will explain it here. I define a term known as activity factor α, α 0 to 1 also referred to as activity factor. Let us suppose I have an inverter, the inverter will be easy to explain and I have an inverter here and my output is here, input is here, so I just check out how many 0 to 1 transitions are available.

So let me let my input be a clock, so I have four clock cycles 1 2 3 4, right. Now, I have a structure which is not an inverter, right which is not an inverter which is something, some gate, some r bit gate and that r bit gate if you look very carefully has got this output which means that for every 4 clock cycle this is the clock I am giving let us suppose, and this is some gate, very complex gate, my output is showing a 0 to 1 transition only in the second clock cycle, and it is then 1 to 0 in the fourth clock cycle.

So there is only, so for every 4 clock cycle there is only one 0 to 1 transition, then we define α to be equals to 1 by 4 that is equals to 0.25, fine have you understood?

(Refer Slide Time: 18:28)

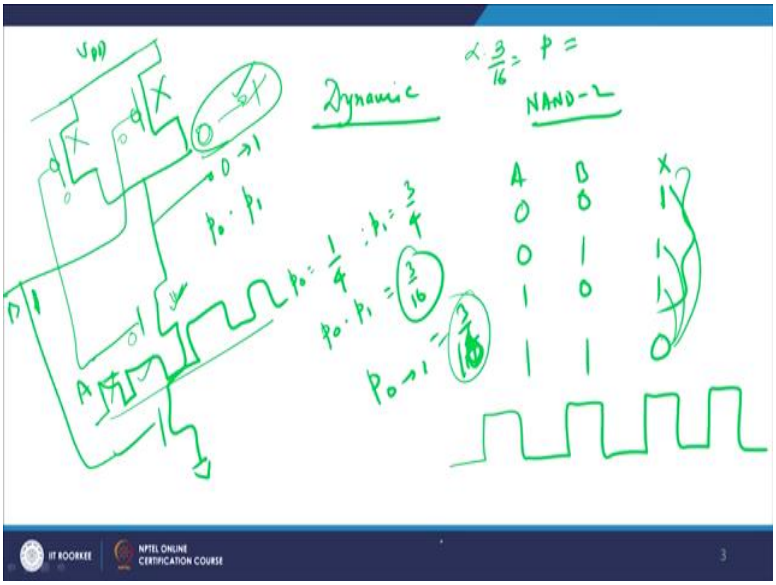



Similarly, if let us suppose, let me just erase this one, right and let me just put it like this that you do have the same clock cycle there are 4 clock cycles, but now what has happened is that rather than one 0 to 1, let us suppose you have one something like this, so there are 4 clock cycles here 4, 1 2 3 4, now you have one 0 to 1 here, you have two 0 to 1 here, you have a third 0 to 1 here, then your activity factor alpha is basically 3 by 4 and it is equals to 0.75.

You will ask me where it is required? Well the frequency which you see f 0 to 1 can be written as probability of 0 to 1 multiplied by α, α is the probability vector. So f 0 to 1 is the frequency of 0 to 1 transitions that is written as probability of 0 to 1 transition multiplied by α 0 to 1 transition, so this is a typical formula which you use, so I can have therefore, the dynamic power dissipation to be equals to $C_L V_{DD}$ square multiplied by probability 0 to 1 multiplied by α 0 to 1, right.

So this is defined as the α is defined is my activity factor for all practical purposes and this gives me an idea if your design is having a… so the gate structure is such that you have a large value of α, then you would expect to see a larger power dissipation, whereas if your gate value has got a lower value of alpha you would expect to see a lower dynamic power dissipation for this case, right.

(Refer Slide Time: 20:25)




One thing which you should be therefore careful about is that whenever you are planning to draw a design or dynamic power dissipation it is not only important that how your design is but what type of gate the design is. For example, I will give you a brief idea, let us take you take NAND-2 logic let us suppose you take. So I have got A, B and NAND-2 logic let us suppose X is the NAND-2 logic, so 0 0 1 1 0 1 0 1, so NAND gate means 0 0 will give you 0 output and 1 1 will get it, you will get 1 1 0, right which means that I will get this, this, this to be a power consuming cycle, whereas 1 to 0 will not be a power consuming cycle, right.

So if you want to find the probability of 0 to 1 you need to find out what is the probability of 0 multiplied by probability of 1, so probability of 0 here is basically P of 0 is 1 by 4 because there are 4 and 1 and probability of 1 is basically 3 by 4 here, 3 by 4. So simply multiply P 0 by P 1 and I get 3 by 16 as the probability of 0 to 1, so probability of 0 to 1 is basically 3 by 16, right.

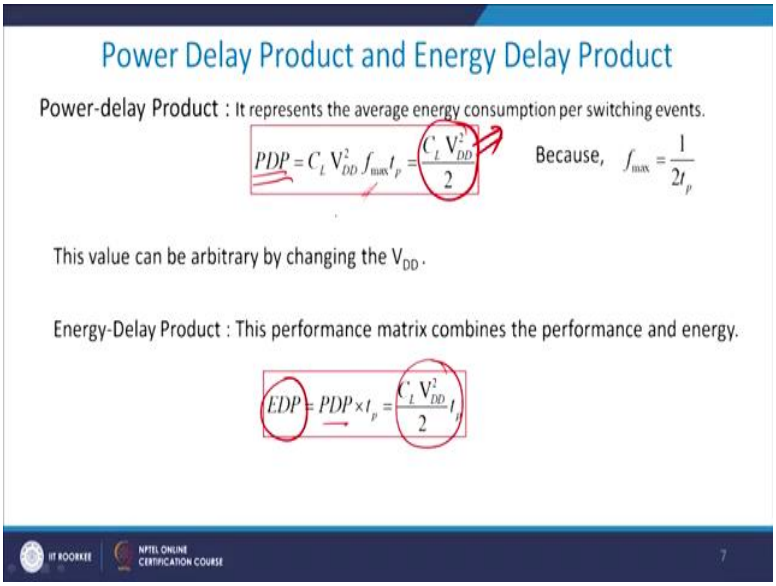
Now in a NAND-2 gate remember therefore, if I give a clock cycle let me say I have a NAND-2 gate, NAND-2 look something like this, right, I have a this is NAND-2 logic, so this is A and let us suppose this is B, so I have got this NAND-2 logic and this is my $V_{DD}$ here and this is a NAND-2 logic. So when it is 0 0 output is 1, 0 0 when it is 0 where 1 output is 1, 1 0 output is 1, 1 1 output is 0 and I get the NAND-2 logic here.

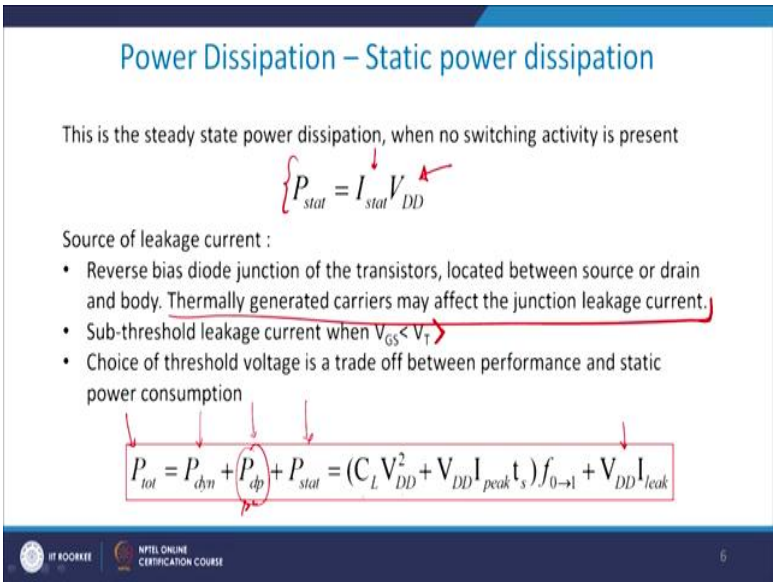
So what is the probability is 3 by 16 here, what is alpha? So if I have now an input cycle which is something like this on the A and let us suppose A and it is something like this on the

A, and B is latch to say 1, then you can find the probability how many transitions are there in the output side from 0 to 1, right.

So if you have 1 here, it means that this is cut off and if you A is giving if A is initially 0 so this is on output will be 0, as it goes high as it goes out this cut offs this switches on, right and since this is already one as this switches on this goes from 0 to 1, so this is 1 so with 1 clock cycle I get one 0 to 1. Similarly, the next clock cycle is get one 0 to 1, so on and so forth so that α multiplied by 3 by 16 will give you the overall probability or the frequency and therefore, that will give you the output characteristics available to you, right, this gives you the frequency of operation of the device.

(Refer Slide Time: 23:15)



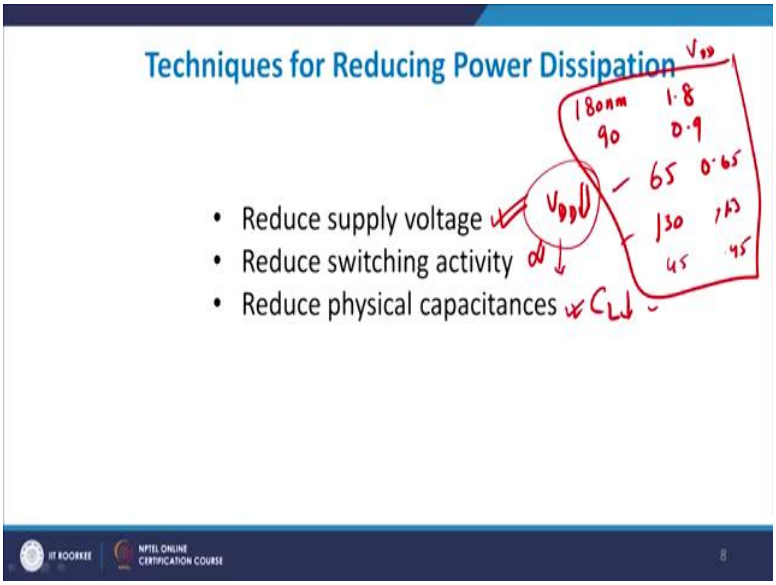
Now, we define a new term which is basically power delay product PDP and it is given by $C_L V_{DD}$ square by 2 at this stage we need not worry about too much about it and we define energy delay product which is PDP multiplied by tp and therefore $C_L V_{DD}$ square by 2 into tp gives you the value of your EDP which is energy delay product,

So we have two types of products which is there with me one is an energy delay product and we have a PDP and we try to optimize the energy sorry we try to optimize the we try to optimize the delay and therefore, if you want to up to reduce the reduce the delay dynamic power dissipation what is what is there in your hand a very good idea to reduce the dynamic power dissipation is to reduce $V_{DD}$, because if you reduce $V_{DD}$ you have almost a parabolic decrease in the value of your power because it is square you are talking about and that is the reason there is a sudden drop which you will see here in this case.

You can also operate at lower frequency of operation, so in most of the cases when you do not want the power where you do not want a very high frequency of operation you can actually afford to keep your $f_{max}$ low, once you do that your PDP and as well as your total power actually reduces drastically. So these are the few things which you should keep in mind as far as designing is concerned.

Generally PDP is a constant for any system PDP and EDP are a constant, so energy delay product and power delay product are constant which means that if the power dissipation rises if power dissipation becomes high then that can be only accommodated provided your system becomes slow, so tp has to go down, tp going down basically mean that your system is getting faster in this case.

(Refer Slide Time: 25:03)



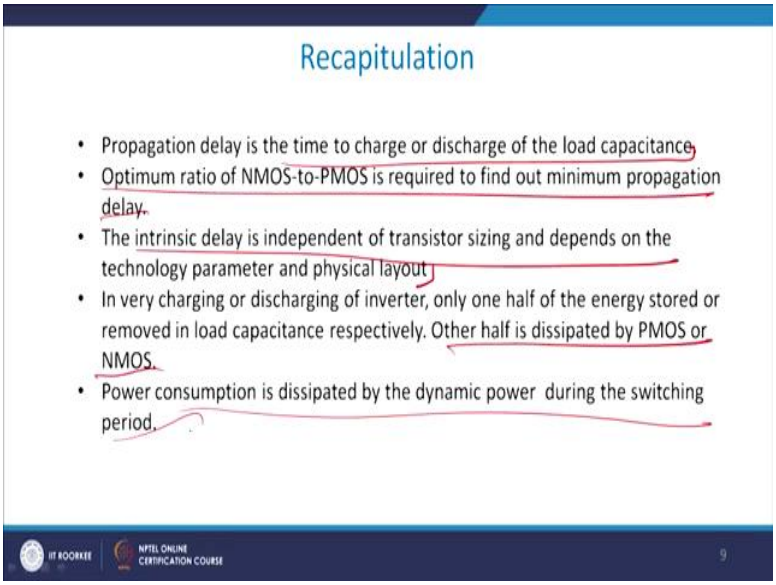

Now this is what you get from the power delay product and energy delay product concept, techniques for reducing power dissipation let me see, we have, we can do it by as I discussed with your reducing power supply, reduce your switching activity $\alpha$ and reduce physical capacitance, so reduce $C_L$, reduced $\alpha$, reduce $C_L$ and reduced $V_{DD}$, once you take care of all these three you will have a reduced power dissipation which is there in this module.

But then the problem is that if you reduce $V_{DD}$ beyond a particular point your current reduces and the time taken to charge or discharge the capacitance becomes large and therefore, the $\tau$ becomes a large factor, so you have to optimize the value of $V_{DD}$ to a larger extent.

So the rule of thumb is that typically the rule of thumb is that if you are working with 180 nanometer technology your $V_{DD}$ should be 1.8, you have only 90 nanometer is 0.9, if you want 65, 0.65, if you are working with 130 then 1.3 volts and so on and so forth, if you are working at 45, 0.45 and typically if not in a very layman sort of a rough back-of-the-envelope you do it and you get these are the $V_{DD}$ values.

Best way to do is to reduce $V_{DD}$, but then if you reduce $V_{DD}$ the price you pay for it is of course that you are ending up having a higher lower current and therefore, a larger $\tau$ and therefore the time taken will be also larger in this case.

(Refer Slide Time: 26:30)




So let me therefore recapitulate what we did in this idea, we tried to find out the minimum what is what was the time taken to charge or discharge the load capacitance, we also tried to find out the optimum ratio of NMOS to PMOS for minimum propagation delay.

We saw that the intrinsic delay is independent of the transistor size and depends only on the physical layout and the technology parameter. Now in every charging and discharging as we saw that the inverter actually loses half only half of the energy stored is removed to the load capacitor, other half is dissipated by PMOS and NMOS I think this clear.

So what we do is that for any single cycle take half $C_L V_{DD}$ square is the total energy I take, half I dissipate to the ground and half of it is dissipates through PMOS and NMOS sink. Power consumption is dissipated by dynamic power during switching period. So we have a switching period, how can you reduce power? Three techniques are available reduce your $V_{DD}$, reduce your $\alpha$ switching activity, or reduce your $C_L$. Reducing your $V_{DD}$ not a very good idea beyond a particular point because the current will be there for reducing and your delay will be increasing.

So these are the few important takeaways from this from this module, from this lecture of inverter, why was this important? Because now since you know how a inverter, you can optimize its power and delay, you can now therefore do small modules of NAND gate, NOR gate, XOR gate using CMOS technology, right, maybe we can do it in the next time and show it to you how it works out in those domains.

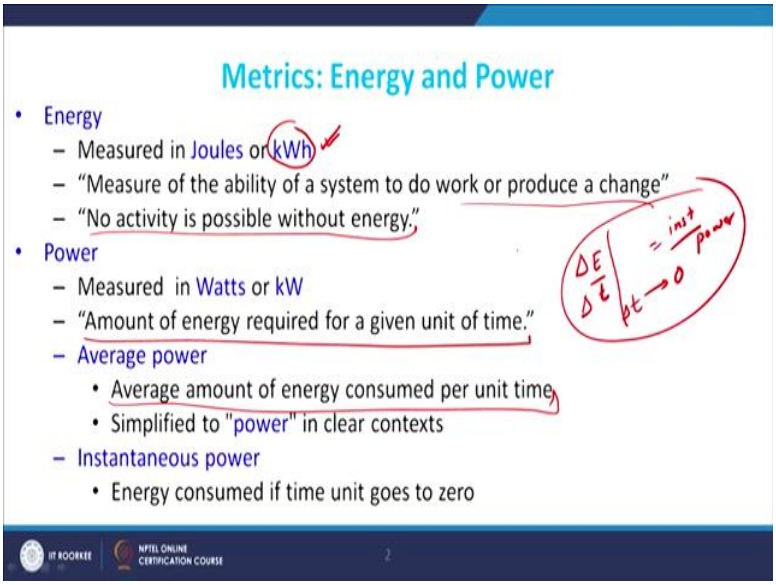
And then optimize using sizing you optimize to get the minimum delay and the highest speed or we do an optimizing of a PDP and EDP of the delay, fine? With this, let me thank you for your patience hearing, okay.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture – 21**
**Power Analysis - I**

Hello everybody and welcome to the NPTEL online certification course of Microelectronics: Devices to Circuits. In our earlier modules and interactions we have understood what is the meaning of propagation delay of an inverter, we have also seen how an inverter using an N-channel MOSFET and a P-channel MOSFET can be formed and what is the basic electrical functionality of a inverter, we have also seen what is the meaning of noise margin as far as inverter characteristics is concerned, what is the meaning of voltage transfer characteristics, how we are able to calculate the high noise margin and low noise margin, and its importance in terms of digital as well as analog logic.

We have also seen the various functionalities of basic inverter and the basic concept of power dissipation. In this slide or in this module we will be actually dealing with the concept of power analysis in a much more detailed manner, so that is what the outline of this talk is that and therefore, it is named as power analysis part 1. Let me give you an let me give you an idea about what the power analysis is all about.

(Refer Slide Time: 01:44)




Generally you must be aware of the fact that energies are always represented in form of Joule or also referred to as kilowatt hour, right and it is also defined as the measure of the ability of a system to do work or produce a change so therefore, no activity is possible without energy,

right? So that is the reason why we always talk about energy in terms of whenever we have an electronic circuit coming into picture because without energy you will not be able to do any computation you will not be able to do any logical operation.

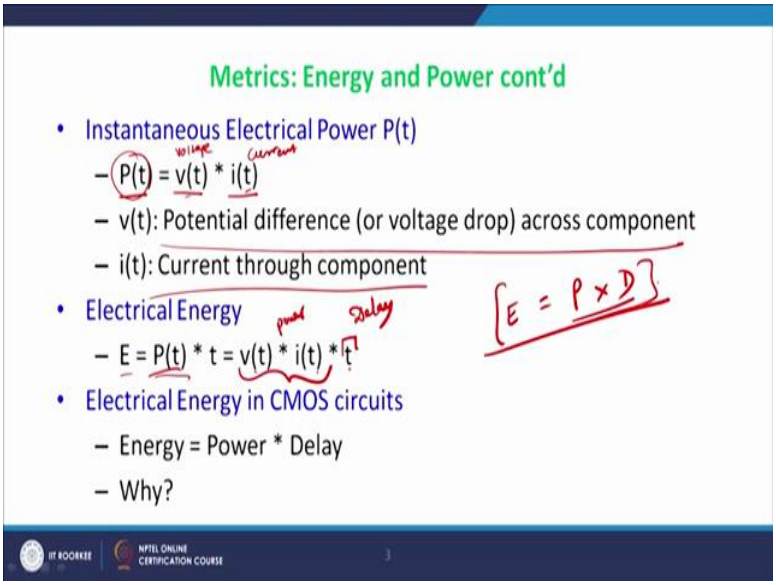
If that energy per unit time is basically power, so energy by time is basically my power and that is what is written. An amount of energy required for a given unit of time is basically referred to as power. So it might be true that I might be doing a very small amount of spending small amount of energy over a period of time, but if my time domain is very small I could have I could spend large power in a small interval of time, right so that is pretty important.

When we say average power it primarily means that the average amount of energy consumed per unit time. So let us suppose you were consuming 5 kilowatt during the first hour, 3 kilowatt in the second hour, then 5 plus 3, 8 by 2 will be for a 4 kilowatt will be the average power. So average power is therefore, defined as a power where you take the average in the time domain, right?

So 2 things take away here from the slide, energy is primarily the ability to do work or if you want to do some work you have to have energy with you and what is power energy per unit time is defined as power, you might have a high power and low energy as well, right, because power is energy by time so if your time domain is pretty small and your time domain is relatively small by energy spending is large you might get a large power.

We have also defined the average power concept here, we now therefore, define also an important term known as instantaneous power. Instantaneous power primarily means that when this t tends to zero, right, so $\partial$E $\partial$t is my power and when $\partial$t tends to zero we define this to be as an instantaneous power which means that when we are calculating power and this power we are just transient at a particular time t, we define that to be as an instantaneous power, right, okay. So we have 2 types of power instantaneous power and average power.

(Refer Slide Time: 04:12)




As I discussed with you therefore instantaneous power when you want to find out P of t is exactly equals to v of t and i of t, so well voltage v of t is the voltage and i of t is the current which you see this is the current so voltage into current which is a function of time t will give you the power here so v(t) is basically the potential difference across the component, function of time t and i (t) is the current component.

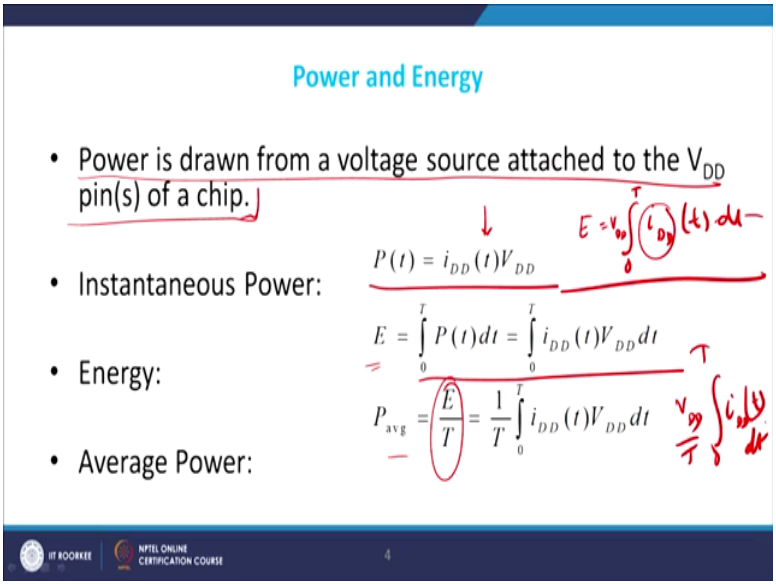
So electrical energy will therefore will be electrical energy will therefore will be power per unit time power into time, sorry, so power into time will be there so v into i into t will be the value of electrical energy, right. I am not going into the details of this one I think these are pretty simple basic concepts even taught in class 12$^{th}$ sort of a mechanism or even in your school days you have taught all these things.

All right, we define a new term also and therefore, if you see very clearly you have a component this of power, right, and this is primarily your time t across which the power is being spent also referred to as a delay. So energy will also refer to as power into delay, right and that is why it is known as a power delay product, energy is also referred to as a power delay product, right.

And this has been used for quite a long time to find out the power delay product in terms of energy, so if you multiply the power multiplied by the delay in the signal going from primary input to the primary output and if the power dissipation is P then we can find out the energy is to be equals to P into D, right and that is what standard mechanism of power and energy is all about.

Therefore in all practical examples you will see that you have $V_{DD}$ rail and you have a ground rail passing through the chip. Now, whenever you accept power, you accept the power from the $V_{DD}$ rail, $V_{DD}$ rail is the rail where you are giving the voltage a DC voltage or an AC voltage you are giving and from this DC voltage you are dragging power and feeding it into various active components of the device, right?

(Refer Slide Time: 06:17)



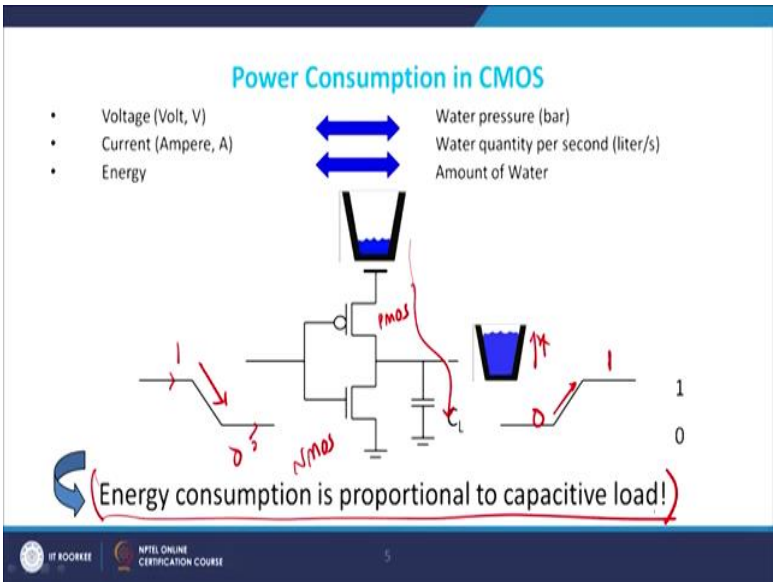

So what I am trying to tell you therefore, is that power is drawn from a voltage source attached to $V_{DD}$ pin of the chip, so I have a $V_{DD}$ pin, it connects to the voltage source and I am able to therefore, have a power being drawn from the voltage source at this point, we have just now discussed that the instantaneous power P (t) is equals to i (t) into $V_{DD}$, $V_{DD}$ is a DC bias so it will not be a function of time t, right.

We define energy therefore, is equal to 0 to t, integral 0 to t P(t) dt and therefore, I can safely write down P (t) dt as $i_{DD}$ (t) $V_{DD}$ dt, so again you can also say that since $V_{DD}$ is constant you can bring it outside and therefore, I can say write down energy to be equals to integral 0 to t $V_{DD}$, right into $i_{DD}$ (t) dt, right. Now since current is a function of your time t you will always have energy integrated from zero to time capital time T where capital time T is the time period across which we are measuring the energy.

Average power as I discussed with you is energy per unit time and therefore, if we divide this expression by t, I get this thing so I get $V_{DD}$ by t $V_{DD}$ by t integral of 0 to t $i_{DD}$ (t) dt. So we will see later on that this drain current or the $i_{DD}$ current is not a constant current, but it is a function of many other factors. Like for example whether the transistors are in saturation,

whether they are in active mode or they are in the non-saturated state, are they working as a voltage variable resistor, depending upon that the value of $i_d$ will be changing, so $i_d$ will be either linear, it will be either nonlinear, or it will be saturated and we should be very careful about that particular point that where we are talking about in terms of power.

(Refer Slide Time: 08:09)

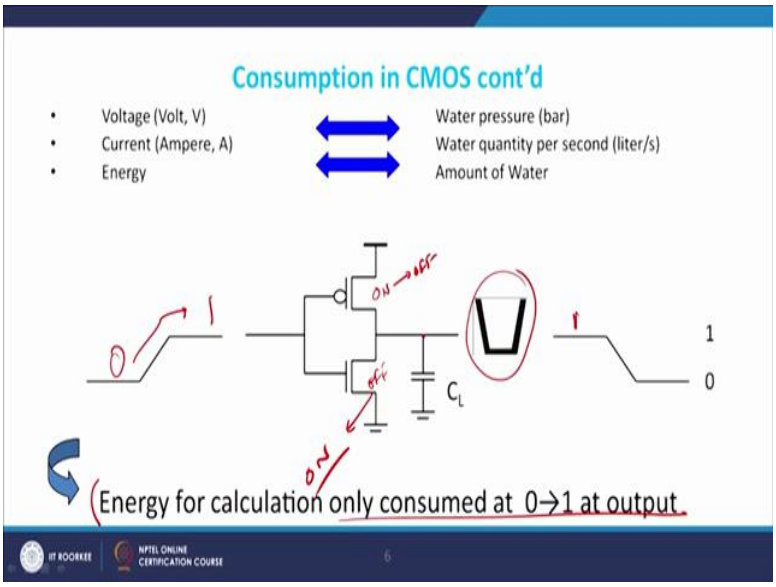



Now let me see what is the power consumption let us suppose that this is a MOS device this is PMOS, right and this is your NMOS here right and your input goes from 1 to 0, right, so when it was initially 1 this was on and therefore, $C_L$ had discharged and now what you are doing you are giving a 1 to 0 transition. So once you give a 1 to 0 transition, right, then what you see a very interestingly that the voltage here falls across this path, this path and charges my capacitor $C_L$ and therefore, the water here rises and therefore, you have a 0 to 1 transition taking place here, right. So I have a 1 to 0 transition and it goes from 0 to 1, fine and this gives you an idea about how fast or slow your design is, right.

So when you are doing it you are actually drawing power or you are consuming power from the $V_{DD}$ rail right because charges has been drawn from the $V_{DD}$ rail and you are able to therefore, have this power available to you. As you can see therefore energy consumption is proportional to the capacitive load, why? Because higher the value of your capacitive load, more charge it can store and therefore, more charge can be drawn from the $V_{DD}$ rail right and therefore, you will be consuming larger amount of power from the $V_{DD}$ rail right and that is the reason that energy consumption is proportional to capacitive load.

So if you want to reduce your energy reduce your capacitive load, once you do that you require less amount of charge and less amount of charge will be therefore discharge from the $V_{DD}$ rail and you will have a lower energy profile available with you, but otherwise it will remain so in a larger state.

(Refer Slide Time: 09:42)




Then let us see what happens when your input goes from zero to 1, so I have initially zero and then it goes to 1, so initially zero means this is ON state this is ON right, and this is OFF, right and therefore, this whole $C_L$ is filled with water which means that this $C_L$ is actually 1, voltage across the $C_L$ is actually equal to $V_{DD}$ and now I go from zero to 1 which primarily means that this goes from ON to OFF state and this OFF it goes to OFF to ON state.

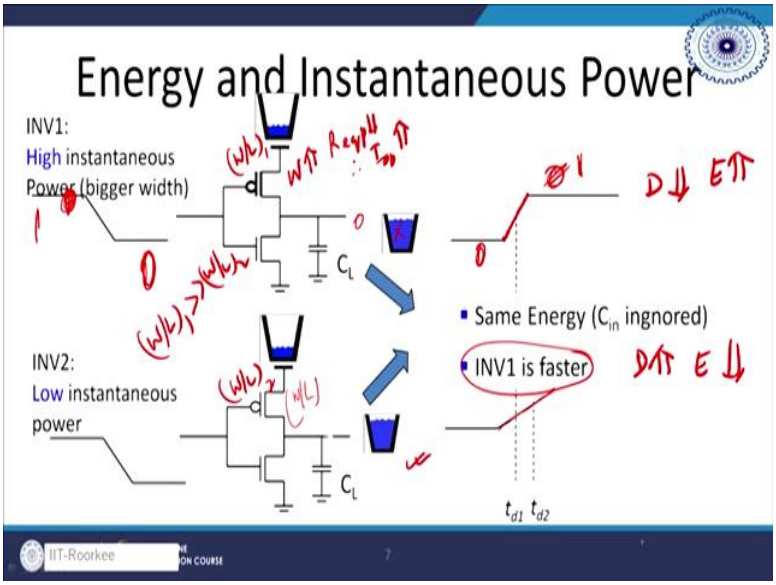
As it goes to ON state I will explain to you, now you see now this discharging takes place to the ground, right, why this discharging? Because now you have a low impedance path with you in the input side, right and therefore, you will always have this output available to you, right. And therefore, you see the CMOS inverter your CMOS design, one very critical aspect is that your output node is typically not a floating node, it is either connected to $V_{DD}$ or it either connected to ground, right.

And that is quite an important observation at this point of time we are not going into further details of it and we see that it is there.

So now energy for calculation only consumed at 0 to 1 transition when the output was 0 to 1 transition that was the only point when you consumed energy, when the output is going from

1 to 0 you are actually dissipating energy and you are able to dissipate energy to the environment or to the heat sink whatever you name it.

(Refer Slide Time: 11:17)




Now the idea therefore is let us see what the idea is, the concept is that, let us suppose your PMOS width was PMOS width was very large, right, so if the width is very large I will expect to get a smaller resistance, so R on or R equivalent P will be small and therefore $i_{dd}$ will be relatively large and as a result what will happen is that if you look very carefully that is what is happening, when input is going from 0 to 1, the output is going from 1 to 0, right, I will just show it to you, or it go to 1 to 0.

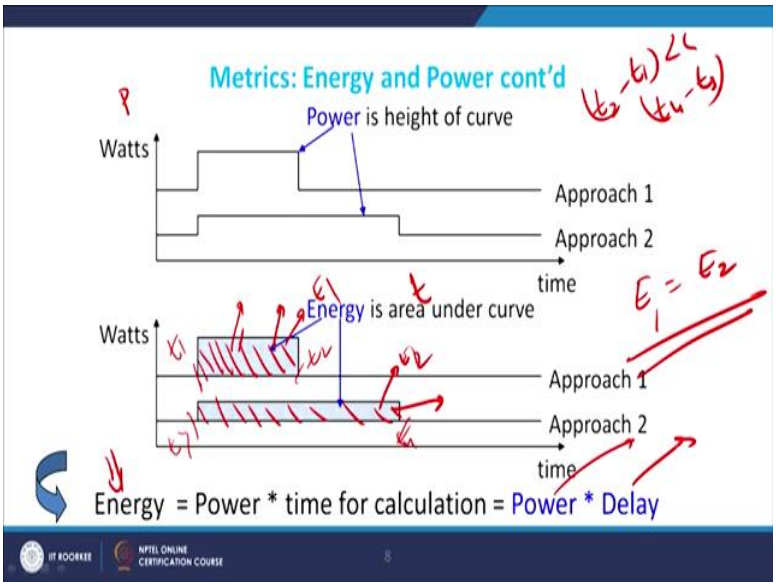
But the rate at which it goes from 1 to 0 is very fast, sorry 0 to 1 I am sorry 0 to 1 to 0 to 1, so this is going from 1 to 0, this is going from 1 to 0 and therefore, output will be 0 to 1, right. Now, 0 to 1 means basically that if you look very carefully look at the first figure here, look at the first figure here, so it was initially unfilled which means it was initially 0, now you start giving your voltage, this starts to fill up means the voltage is rising and the voltage is rising like this, after that it remains constant here.

Same concept applies here also, but only thing is that here W by L ratio for the PMOS is smaller, so this has got a bigger width W by L is larger and this is W by L is smaller, right, so W by L 1, let us suppose 2, so W by L 1 is much larger as compared to W by L 2. Primarily meaning that you allow larger currents to flow and charge your $C_L$ therefore, the rate of charging is smaller and you are able to achieve 1 very fast, right, but when your W by L ratio is small you allow a larger amount of time to move and therefore, inverter 1 is faster so the

inverter 1 is faster, so why it is faster? Because I am able to do a large amount of transfer within a short period of time, which means that, that since the delay is lower now I would expect to see the energy to be high and that is true also because for a smaller duration of time you are sending a large amount of charge, so the energy is large.

So what happens is that the delay though is reducing, but the energy is basically typically very large in this case, whereas in this case delay is large right and therefore, energy is reduced in the second case, second case W by L ratios are very small for the pull-up device in this case PMOS devices.

(Refer Slide Time: 13:46)



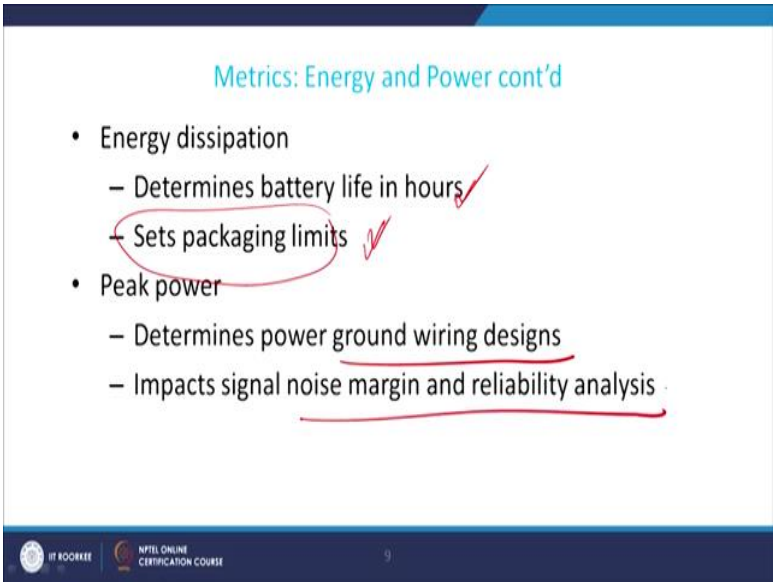

Now therefore, there are 2 approaches which is available with us, now if you look at the approach this is power versus time, right, the area under the graph is basically your energy P because P into t is basically your energy, right so this is $E_1$ and it is $E_2$, both see if we look at in that way both $E_1$ equals to $E_2$, it primarily meaning that the energy consumed by both the transistors please understand, both the first transistor and the second transistor is large, but the first transistor does this energy at a shorter duration of time, whereas the second one it does it larger duration of time t and this is a larger duration of time it is doing it and this is doing it a shorter duration of time, right.

As a result so energy is the area under the curve, right, so energy is same, energy is same because the area is same, but then you are dissipating more amount of power here in the first case the reason being you are doing it in a smaller time domain, so this is $t_1$, $t_2$, right and this

is $t_3$ and $t_4$ . Now if I subtract $t_2$ - $t_1$ it is much smaller as compared to $t_4$ - $t_3$, as a result this energy $E_1$ so the power is much larger as compared to the second power.

Okay, so we write down therefore, power into delay equals to energy, so energy is equals to power into delay, right, so this is this is what a standard mechanism which we follow.

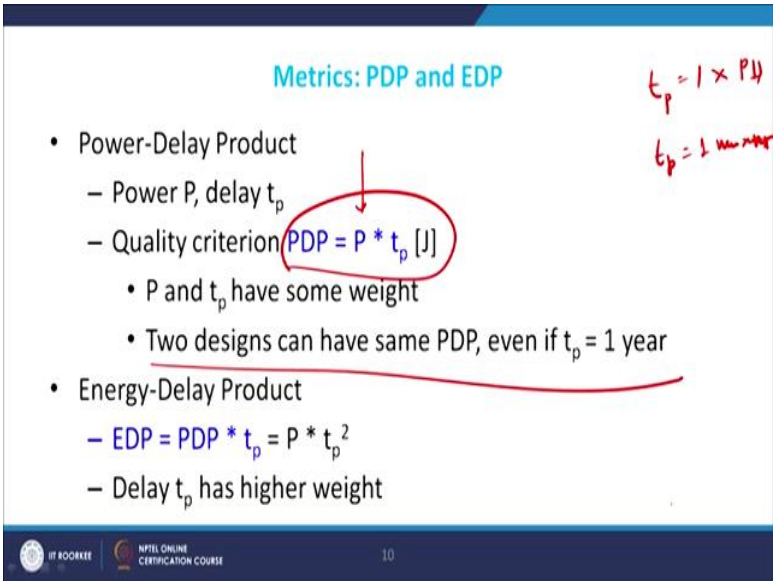
(Refer Slide Time: 15:19)




Now, let me therefore come to the matrix the problem of energy dissipation is that the battery life will be in hours rather than minutes if I reduce the power if I reduce the power dissipation and it sets the packaging limit, why packaging limit? The reason being if it is highly dissipative in nature you should have proper sinks, heat sinks in the packaging so that you are not able to let the heat or the power dissipate within the chip, but outside the chip, right, so your packaging limits are set in that, right.

Peak power is defined as the power which is peaking at any particular time t and therefore, depends upon the wiring design, as well as a noise margin and reliability analysis so we will not go into details of this 1, but primarily therefore, we if you look very carefully I can have therefore, 2 energy which are equal, but in 1 case the power will be higher and in other case the power will be lower, right that is what we know as basic concept which you see here.

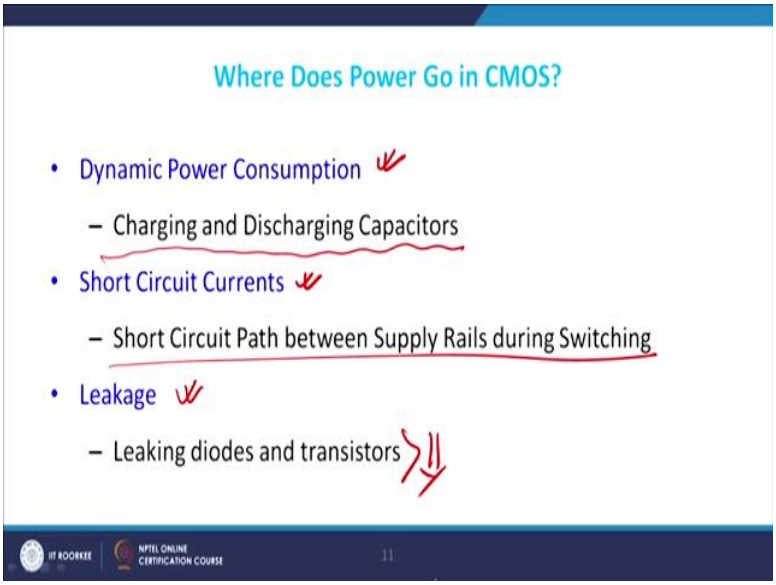
(Refer Slide Time: 16:15)




So therefore, the quality criteria power delay product as we discussed was an energy PDP was equal to $P^* t_p$, p is the power dissipated multiplied by $t_p$ which is the propagation delay, right. So you see 2 designs can have the same PDP even if $t_p$ is 1 year, I hope you understand. For example in 1 case the $t_p$ is 1 year and another case $t_p$ is basically say 1 month or whatever small value, then for the same PDP this will have a lower power, whereas this will have a much higher power, right.

So if you if you are able to achieve a functionality within a small duration of time, you end up having a larger energy or larger power, whereas if you do the same computation for a larger period of time you are spending less amount of power. So power depends not only on the total energy, but also on the time domain analysis of each 1 of them right, and that is quite an interesting one.

(Refer Slide Time: 17:25)

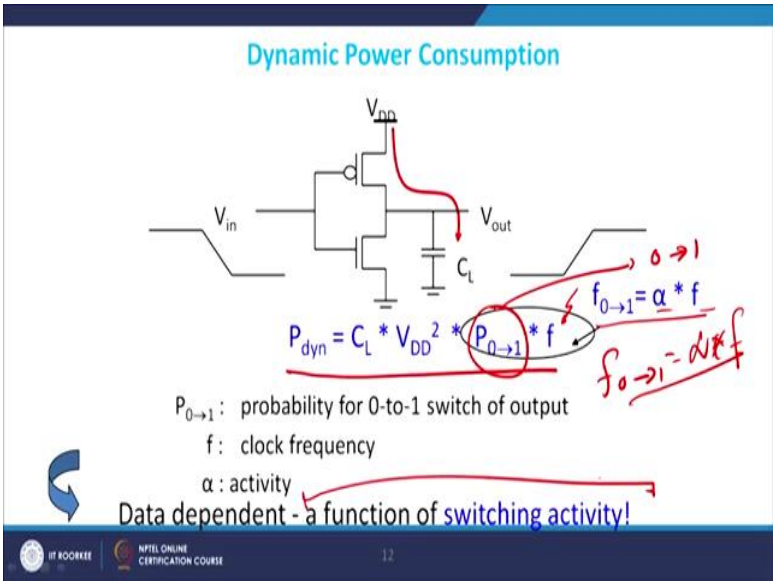



So where does the power go in a CMOS? That is pretty intriguing question and people have been asking this question for a long time and there is a straight answer available here also. We have 3 types of power consumptions, we have got dynamic power consumption we have short-circuit power consumption sorry short-circuit and leakage, so you have dynamic power, we have short-circuit and we have leakage.

So dynamic as the name suggests it is basically not fixed, but goes on changing with respect to time which means that charging and discharging of the capacitor of the design is defined as the dynamic power consumption. So faster the charging discharging more will be the power available with you or more will the power dissipated for you. Then of course as I discussed with you in a previous term, you will have a short-circuit power between supply rail and ground during switching and there will be also some leakage current which is primarily your reverse leakage current, also due to Zener leakage current and so on and so forth and the transistors will show you a drop here.

So these are some of the problem areas which you will face, the last 2 and the major contribution which the this will give you is basically I will have a larger dynamic participation or a smaller dynamic power dissipation, I will have a larger or a smaller short-circuit power dissipation and we also have a smaller leakage which is basically a sub-threshold leakage with us and this is what we get from here.

(Refer Slide Time: 18:49)




Now let me come to the again the basic dynamics or the basic equations of dynamic power dissipation and if you remember correctly half $C_LV_{DD}$ square you are taking from the $V_{DD}$ rail and then you are pushing the half $C_LV_{DD}$ square into the ground, right. And therefore, for a 1 cycle you are actually taking half plus half which is actually exactly equals to $C_LV_{DD}$ square and that is the reason you get $C_LV_{DD}$ square as the dynamic power, initial dynamic power.

Then what you do is you multiply this with P 0 to 1, P 0 to 1 is the probability that you get 0 to 1 switching done in the circuitry, easier said than done because you do not have any control over the circuitry and the second thing is that if you do a 0 to 1 transition you will automatically have almost half the maximum frequency of operation and that is the reason the probability for 0 to 1 switch of output is kept at the last point, f is the clock frequency and α is the activity, we had already discussed the activity concept in the previous turn.

So for example, if you have a clock, the clock will have a surely an activity and that therefore, will have our activity equals to 1, if you do not have a clock, but you have some complex logic gate maybe your α will be 0.75 and so on and so forth. So on 2 factors primarily a dynamic power depends and that is known as the frequency of operation as well as on the probability that the output will go from 0 to 1, fine, that is what is dynamic power dissipation is all about.

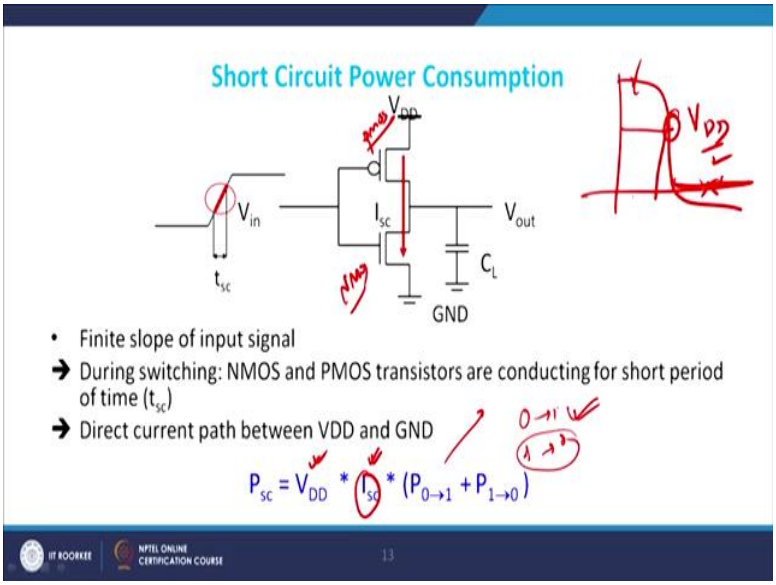
Now let me come to therefore, as I was discussing with you, you get 0 to 1 and this f of 0 to 1 is defined as α into f, please understand that f of 0 to 1 is basically the meaning of

functionality from zero to 1 is defined as α multiplied by f, α is the activity vector and f is the frequency of operation of the device and f is the clock frequency.

Let us look at the dynamic power dissipation as I discuss with you ($C_L V_{DD}$ square)* the probability that you are going from zero to 1, since so zero to 1 multiplied by f, where f is the frequency of operation of the device. As I discussed with you data depended, why is it data dependent? And it is a function of switching activity, why? The reason is it is data… but data dependent in the sense that if your both inputs are 1 1 and you have a NAND gate also, logic is also there, then your pull down network will be fully activated and your output will be very strong 1 to 0 transition.

So which means that it is very difficult for you to go from 0 to 1 transition and the reason being that if your data depends on the functionality of the switching circuit activity which means that how the input, how the output goes high or low will depend upon whether the device is switched ON or OFF, right and that is the problem area of the data being inserted here.

(Refer Slide Time: 21:41)




Let me come to the next part that is the short circuit power dissipation or consumption, we have just now discussed with you the dynamic power which primarily depends upon the frequency of operation and the capacitive charge, we not depend short circuit, we have already discussed this point earlier also, but I will dealt in slightly more detail.

If you remember we were discussing that the during the ON state when your device was ON, we define the device to be in the ON state and we also say that the output will be switched

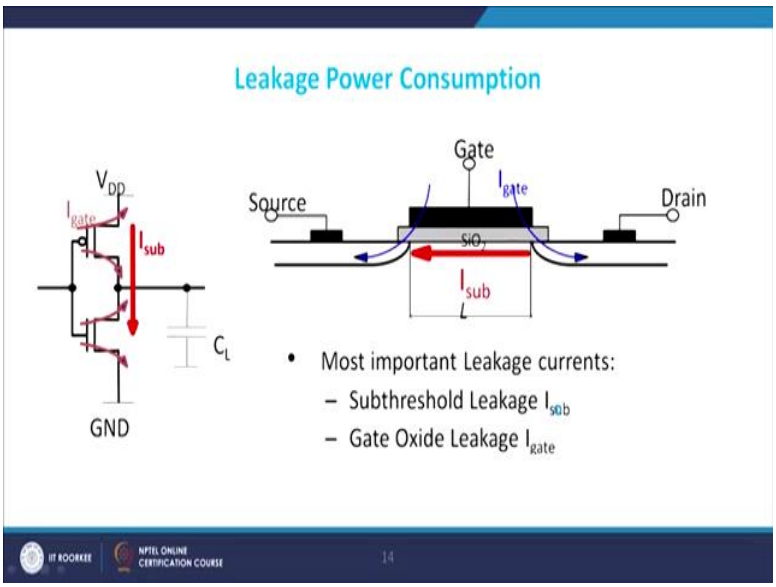
OFF because typical inverter characteristics we discussed yesterday, right something like this. So when your input is low output is high, if input is high output is low, the input is high output is low.

Now if you are biasing a device here or here you will have a problem and the problem is that you will… though you will have no current flow through it, but there will be a loss in the voltage value. Somewhere here in the middle where you have switching thresholds available approximately equals to $V_{DD}$ by 2, then your NMOS and PMOS, this is your PMOS and this is NMOS, these 2 are equally ON, getting my point? Which means that both are equally ON which primarily means that both the transistors will be able to pull the charge carriers from the peripheral onto the capacity load, fine, is it okay?

So that is the reason the short-circuit power dissipation is very very low in case of in case of a simple CMOS inverter right, and this is what is an interesting idea. Now as I discussed with you there is a direct path between $V_{DD}$ and ground, so I write $V_{DD}$ into $I_{SC}$, $I_{SC}$ is the short-circuit current multiplied by P 0 1, P 0 1 is the probability that the output goes from 0 to 1 and then you have a probability the output goes from 1 to 0.

So we are actually not interested in this domain, we are either interested in this domain which means to say is that given a complex logic gate can you find the probability that the output goes from 0 to 1, so there is a, we will do it as we move along, but that is what it is all about here, so short circuit depends upon all these values.

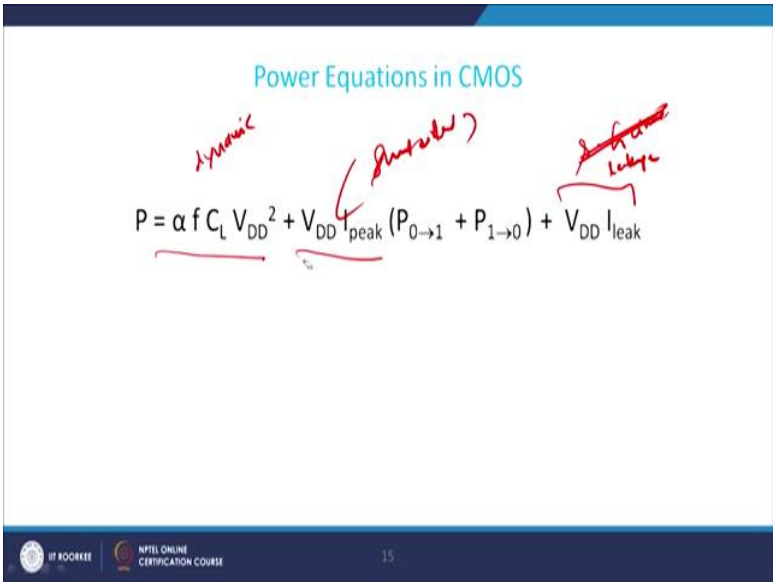
(Refer Slide Time: 24:30)

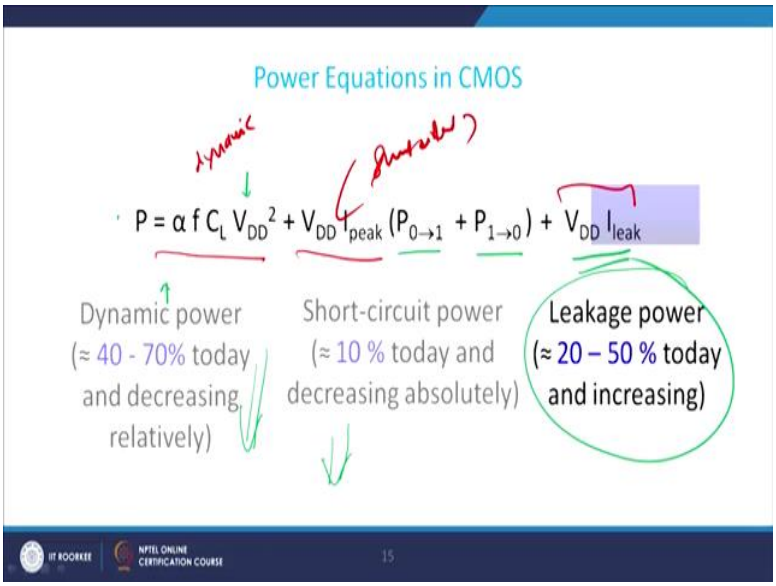


Let me come to therefore, the leakage, so I had saturation in which the gate voltage was very large or very small and therefore, the output will be either switched ON or switched OFF. The second issue which came up was that of the static power dissipation or yes the short-circuit power dissipation, when both the devices are equally ON you will have a direct path between the 2 end up with last leakage power which is of sub threshold leakage.

So when your gate voltage is just smaller than the threshold voltage device there will be large carrier movement which will be available which will result in the sub threshold leakage, we will not go further than this, but P peak equals to I peak into $V_{DD}$, now I peak can be written as $I_{sub}$ plus I get multiplied by $V_{DD}$, so this is what an important this thing we get.

(Refer Slide Time: 25:00)



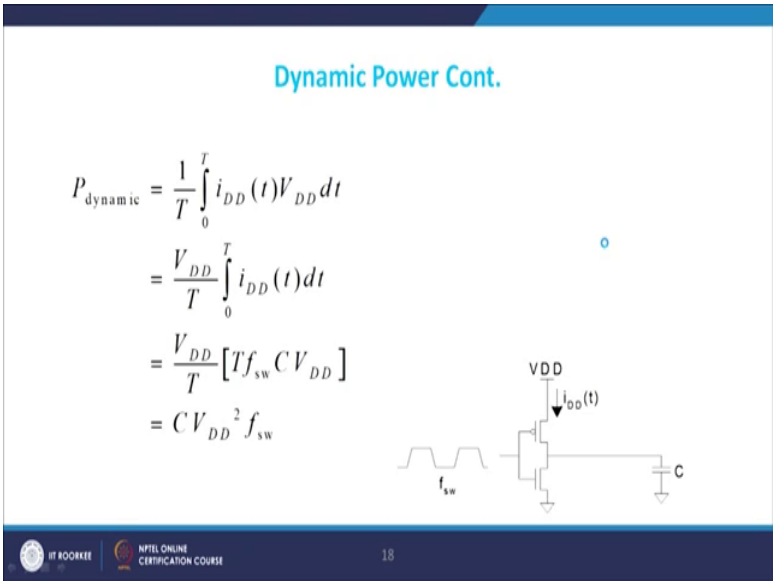
Next we get the power equation, the power equation as I discussed with you $C_L$ f times $V_{DD}$ square is your dynamic power dissipation, we have $V_{DD}$ times I peak as the so this is basically your, so this is a dynamic, this is your short-circuit, short-circuit power dissipation, this is your not short circuit, this is a short circuit, short circuit, circuit power dissipation and this is basically your power dissipation by virtue of leaking so I have a leakage available with me.

Now if you see carefully in with all these discussions or all these ideas you will you will appreciate that if you look carefully it tells me that the total power dissipation can be broken down into 3 parts and those 3 parts are mentioned in this region that some of them will be in, so what I tell you is that if you have some voltage variations your $V_{DD}$ will change and your P will change.

Similarly, if your probability of 0 to 1 and 1 to 0 if it happens you will obviously get added up with the total probability and that is the reason you will get a higher energy. Similarly, $V_{DD}$ into I leak will also give you a higher power. Now a dynamic power is approximately 40 to 70 percent of the total power and it is decreasing yearly and that is sort of a good thing for us.

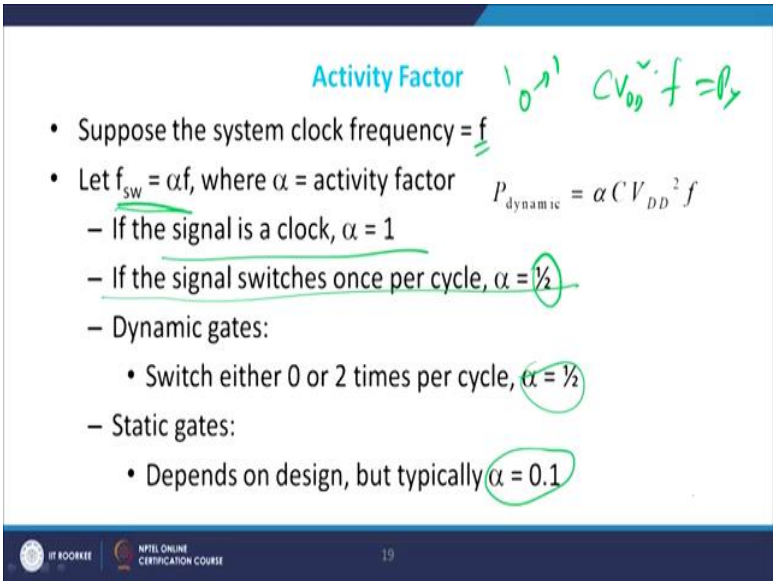
The second is short-circuit power is relatively small is approximately 10 percent of the total - ,,,,,,,, to total charge available and this is decreasing, so this is decreasing, this is also decreasing, what is increasing is basically my leakage power because of the reduced dimension you do have a larger chance of the charge carriers being leaking and that is the reason it is leaking by 20 to 50 percent, then here where people are actually concentrating large for optimizing power.

(Refer Slide Time: 26:47)




As I discussed with you dynamic powering, I will not go into further details of it, but I will give you a brief idea about dynamic power dissipation, it is actually $V_{DD}$ by T f SW t times f SW is basically the frequency operation or number of times the output change before time t and this is what is known as this thing. We came to dynamic power dissipation given by $CV_{DD}$ square times f of SW, f of SW is the frequency of your short way, the frequency of your wave.

(Refer Slide Time 27:21)



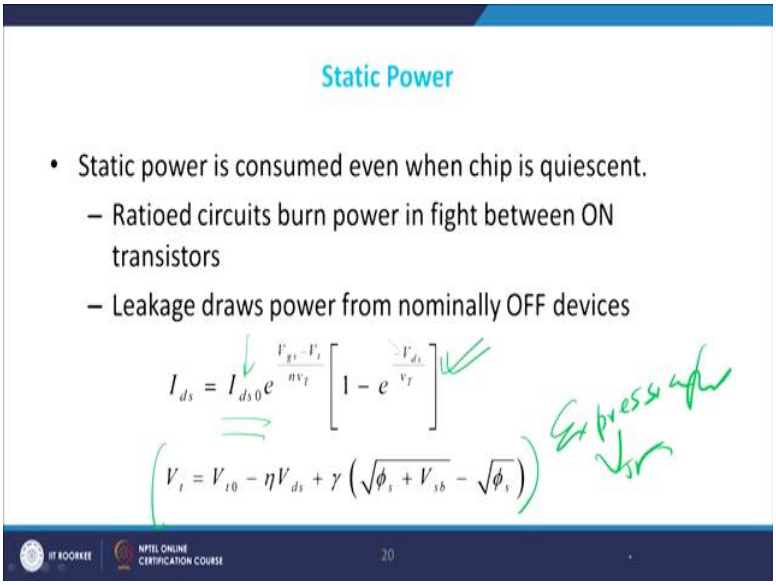

Now suppose a system has system clock of f, then $f_{SW}$ is let us suppose α into f, but α is the activity factor. Now if signal is a clock as I discussed with you, α will be equals to 1 and therefore, I will get $CV_{DD}$ square multiplied by f as equals to P dynamic. Similarly, but if the

signal switches once per cycle, so there is a one 0 to 1 transition between the 1 full cycle of this input, we can give you an α equals to 0.5 or 1 by 2.

For static gates typically α will depend upon the value of the static gates it is equals to 0.1 and if you are using a dynamic gate at this point 0.5 or 1 by 2. So we have seen one important point of activity factor is that activity factor depends upon the type of logic level you are deciding, as well as on the structure which you are designing.

Now let me come to the static power, we have already discussed static power, the static power is primarily the power which is dissipated across the board or across the device when you do not give any dynamic power, so which means that when you do not give any $V_{DD}$ at that point of time whenever the device is in the OFF state maybe or just simply OFF you have leakage currents available or you have a sub threshold current available which basically gives you a larger power.

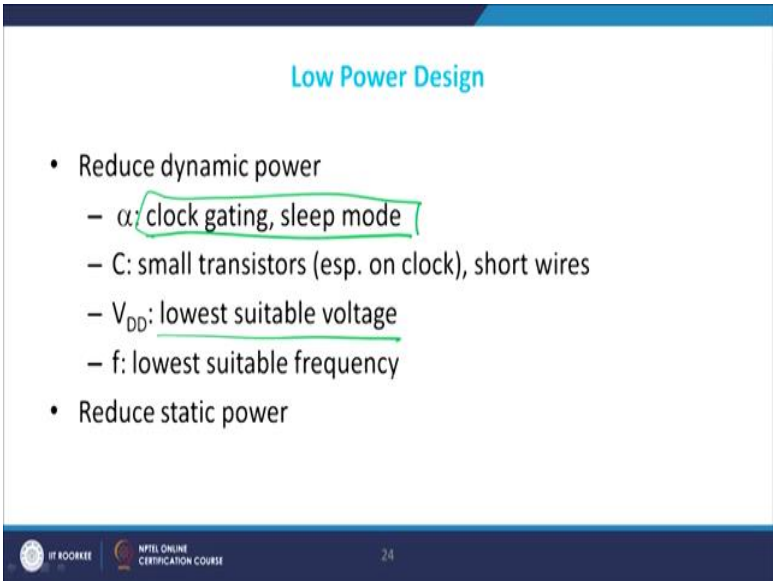
(Refer Slide Time: 28:45)




So this is a larger power, $I_{ds}$ equals to $I_{ds}$ 0, $I_{ds}$ 0 at V equals to 0, the amount of current flowing is given as $I_{ds}$ 0 and we get this big equation available here, and I got the equation, but the threshold voltage.

So this is the expression for, expression for threshold voltage and that is quite an interesting result which we get, but if you forget about this 1 the last 1 you will see that with increasing $V_{DS}$ $e^{-VDS}$ will go on decreasing and therefore, 1 minus that will go on increasing and as a result you will get larger ideas value. So mathematically it seems that this should work fine.

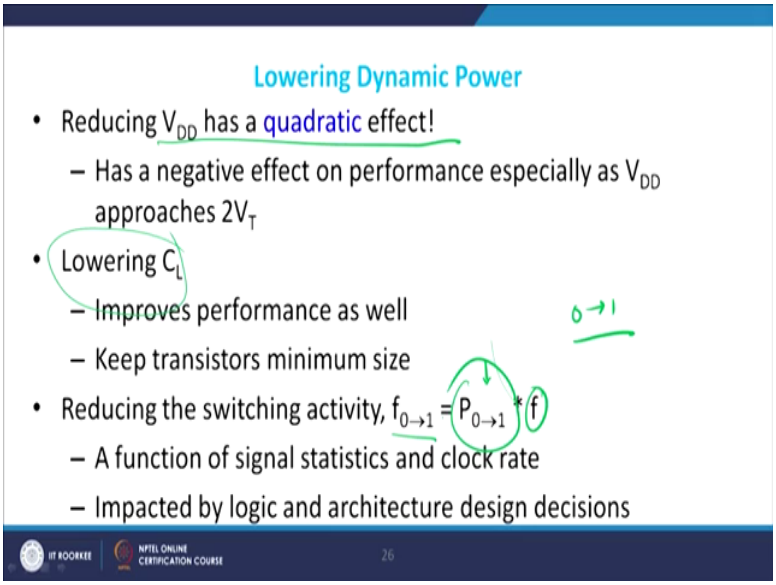
(Refer Slide Time: 29:29)




Now, let us see how can you reduce power, that is the most important part which we learned from all these things. The first thing is that reduce your clock gating or do you put it into sleep mode? So when the system is not working either you put it into sleep mode or you put your the devices into clock gating, so once you put through clock gating, it primarily tells me that you can reduce the value of α, C should be small and therefore, you do not have to work with large wires, it should be small in length and that will give you a small transistor based design.

You should also work with this lowest suitable voltage in all sense and the reason being if it is not the power dissipation levels will be will be typically very large and you require the lowest possible frequency. So possible frequency, voltage power and you should actually therefore, rely on small transistors and on short wires and this takes care of our low design concept.

(Refer Slide Time: 31:21)



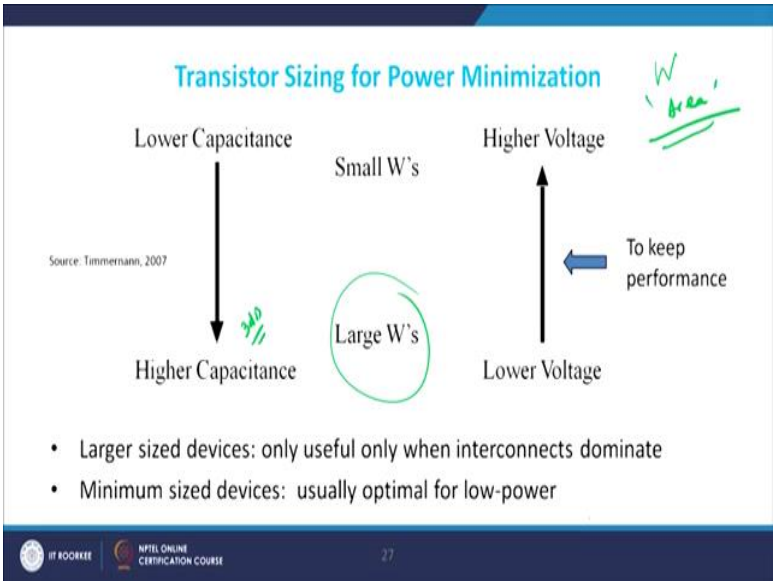

So how do you decide, how do you reduce your dynamic power as I discussed with you α, C capacitance, $V_{DD}$ which is the lowest suitable voltage available to you and α f is the lowest suitable frequency, how do you, so this is that, how do you reduce static power? Selectively ratioed logic selectively use low $V_T$ devices you will have leakage reduction provided you arrive at stack devices, body bias and temperature.

So what you see is that I can reduce static power by using certain changes in the circuits itself making it ratioed logic, whereas if you want to reduce dynamic power or want to reduce the branding power to a bare minimum I need to concentrate primarily on for the purpose of reduction on the dynamic part of it. Now if you look carefully how do I lower my dynamic effect?

So since $V_{DD}$ has a quadratic influence so therefore, even if you double your $V_{SS}$ which is doubling a power supply that will lower your power dissipation by 1/4th and that is quite an interesting phenomena people have observed. Okay, so lowering a $C_L$, lowering $C_L$ of course will be meaning that you will be able to improve your performance because when a capacitive load reduces, you achieve a much better performance with you.

Now reducing the switching activity, switching activity is again defined as the probability of 0 to 1 multiplied by frequency of operation, right. So this is the probability that you have got 0 to 1, the 0 to 1 transitions are power consuming transitions and that multiplied by f will give you the frequency of 0 to 1 transition in this case.

(Refer Slide Time: 32:11)

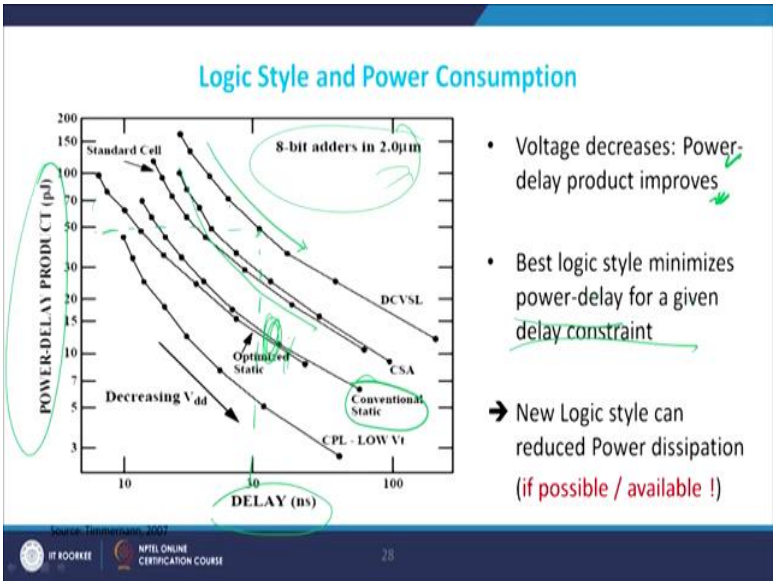



Now, let me come to the transistor sizing concepts, so if you have a small W which means the width is small, I would expect to see, I would expect to see of course the lower capacitance because small W primarily means that the area under the this thing gate is very small and therefore, you require a lower capacitance. Whereas if you make your capacitance higher and higher I obviously at larger W, I get a higher capacitance and therefore, this restricts my 3 dB bandwidth and so on and so forth.

What happens if you want to work at higher voltages? And the reason is that sometimes my requirement is not high speed by high performance which means that I cannot sacrifice on the performance and therefore, what I do? I increase the value of $V_{DD}$, once I increase the value of $V_{DD}$, I make it larger the performance increases at the same instant of time the power dissipation also becomes large.

So this is a methodology which people use that they try to do what is known as an adaptive scaling in terms of this thing in terms of both $V_{SS}$ and in terms of other devices as well.

(Refer Slide Time: 33:22)




So I will not go into details of this 1, but just give you an idea if you look at this graph it is between the PDP power delay product versus the delay, right. So if you see that the delay increases the PDP also goes on increasing for all that design, which means that you did not contribute to the design beyond a particular time limit and therefore, you see it is actually getting reduced drastically.

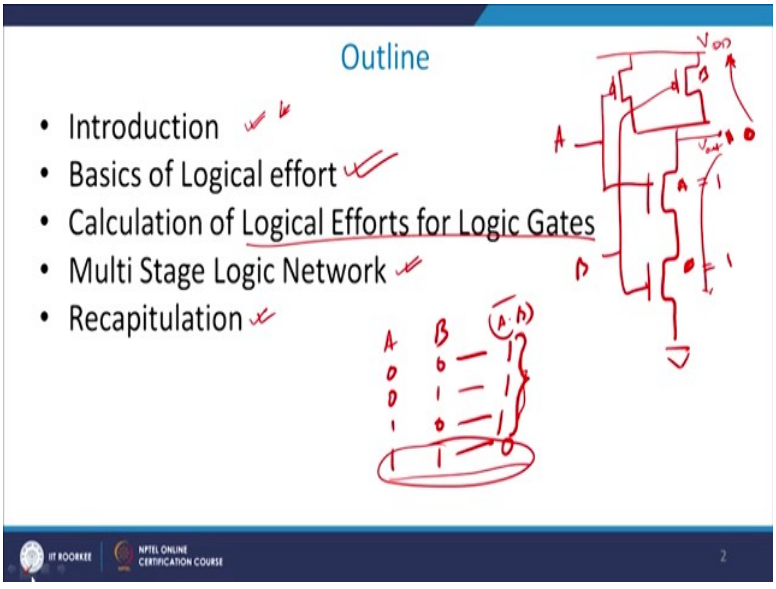
For the conventional static CMOS this is the equation of the graph of power design failure and for non-standard 1s these are the case of the design available with you. So this gives you a best logic style for input-output combination and gives you an idea about how this works for, so this is this example is for an 8-bit adder, so I had used an 8-bit adder in a 2 micron CMOS technology and I got this these are the outputs so these dots are basically your output available to you for the PDP versus delay so when your delay is 30, or whatever 30 microsecond or nanosecond, I will see this to be equals to approximately 50 PDP of pico joules, as you make your delay large, this PDP will fall down and that is 1 of the major areas or advantages of this system. As you can see therefore, as the voltage decreases the power delay product improves. Now the best logic style minimizes power delay product for a given delay constraints, this is the example, okay. You can achieve a new logic style and reduce the power, but that is the way beyond the scope of this work. So we have learned 2 things, we have learned in this chapter that we had learned power dissipation, as well as the concept of delay and what is the meaning of PDP and EDP. With these word let me finish off today's lecture and when we meet next time we will do something else regarding the logic design, okay thank you very much!!!

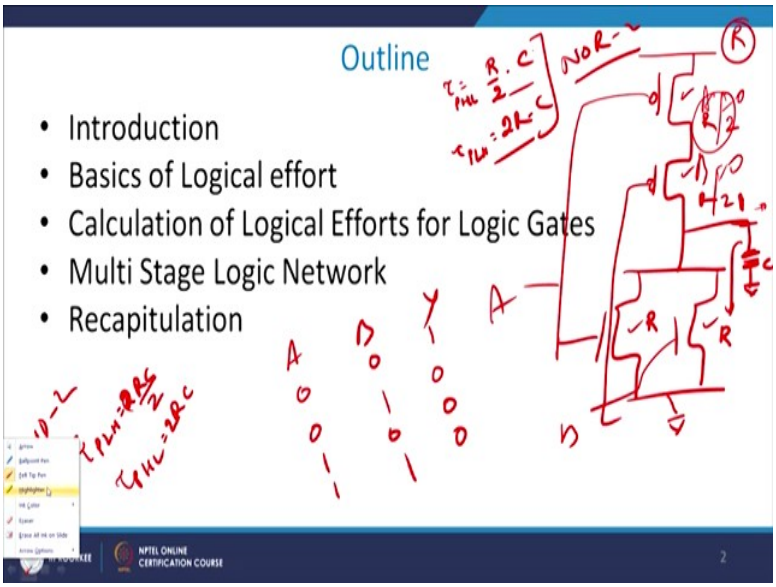
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture – 22**
**Logical Efforts - I**

Hello and welcome to the NPTEL online certification course on Microelectronics: Devices to Circuits. In our previous interactions we have seen that we have understood the basic concept of power dissipation on what issues these power dissipations can be lowered, how can they be lowered, we also looked into the concept of delay which means high to low and low to high propagation delay, we saw that the delay was obviously dependent on the width of the transistor beat pull-up or pull-down and therefore, higher the width lower was the value of a $t_{PHL}$ or $t_{PLH}$, but the price you paid for it was that higher the width means more will be the capacitance and therefore, it will be difficult for the signal to drive the gate, right because your if your width increases the area under the gate increases and therefore, that is the problem area. So you need to optimize the design.

And in one of the optimization problems which we saw in our previous case we saw that, if you are able to keep a fan-out of approximately the stage ratio to approximately 3 we can get the minimum delay available to us. Keeping that in background let us start with one module which is a critical model for estimating delay between point A and point B of a circuit, right and that is what is the methodology which is adopted to do that is basically known as logical effort.

(Refer Slide Time: 2:12)






So today's topic will be referred to as a logical effort. So we will look at logical effort concept. So what will be doing is we will be introducing to you the logical effort what is the basic, sorry, I will give you the basic idea of logical effort and basic idea of logical effort, then we will be coming to, so we will introduce to you the logical effort what is the motivation for studying logical effort, and then what is the meaning of logical effort and where it is effectively used so basics of logical effort will be known to you.

Then how do you calculate logical effort for certain logic gates, right and then look at the multi stage network and then we will finally recapitulate, right. Before we move forward let me give you an idea about the concept or let me give you an idea about how you design it. For example if you have for example a gate which is something like this which is basically a two input NAND gate, right, right and you sometimes do like this and then you do like this then this is A and this is B, then if you put A and B as your inputs, right, then you will see for 0 0 since both A and B is, so this is A, this is B, so when both A and B are 0 output will be 1 when either one of them this is still output will be 1 and when both are 1 output will be 0. So this is your NAND gate approximation, so A dot B bar.

Now the idea here is that you can only have one transition from, so if you are storing 1 here you can only go to 0 provided both A and B are equals to 1, which is this one, but if this was initially 0, then you can go to 1 either of these 3, so if you fall back to your previous understanding you will know that as I discussed with you that 0 to 1 transitions are the transitions which are power consuming cycles. So therefore, if the… initially you have a data of 0 available with you, the probability that it will go to $V_{DD}$ is much higher as compared to it

going to down assuming that all the inputs are statistically possible and independent with respect to each other.

Then if initially you had 0 in your $V_{out}$ output, then it can go to $V_{DD}$ only if under all these three conditions, whereas it can stay to 0 only and 1 the condition which is basically this condition.

So you see that a power dissipation is therefore, also depends upon the statistics of your input and the type of gate whereas, if you look at this point and then if we let us say you design something like this that now your pull-up is basically your pull-up is basically series, right and you pull down is basically in parallel so if you have A here and you have B here so if you look here then if you plot A, B and Y, Y is the output here then if for this condition, right when both are 0 then only you will have 1, for all other cases you will have 0 which means that you initially if you have 1 here then the probability it to go to 0 is just that only one of the inputs or both of them should go to 1, so the probability is just to go to 0 is much higher whereas, if it wants to go to 0 then both A and B has to go to 0.

So again I am assuming that the, if it is statistically independent and equal probability of getting all the signals are available with you then I would not expect to see, I would expect to see power dissipating cycle more in a and NOR 2 logic as compared to a NAND 2 logic, this is the first observation. The second observation is that you see when you are pulling up you have two transistors in series, right and when you are pulling down in this case you have only one… even one transistor if switches ON you will be obviously in parallel.

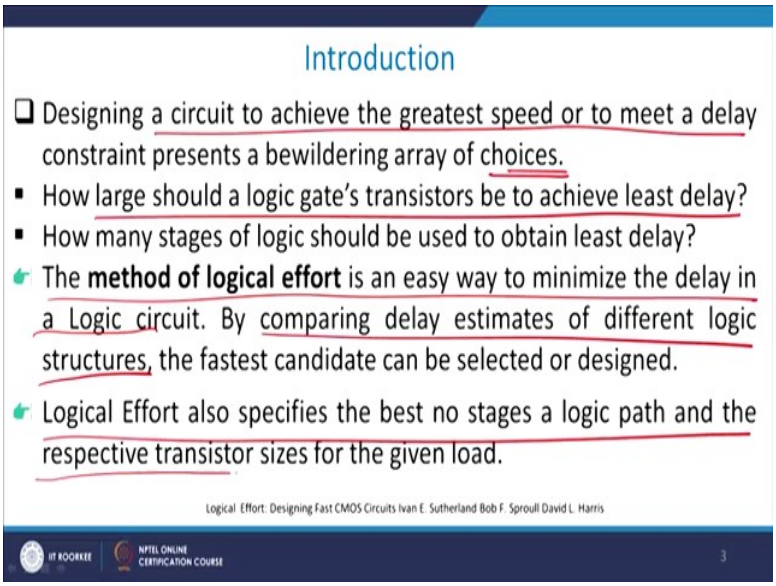
So, if we assume that each transistor is the resistance of R, then and if output is basically let us suppose is C, then I can safely say that if both the transistors are ON, both will be in parallel I will get $\tau$ equals to R by 2 right, into C whereas, if I assume this to be also R this to be also R, then $\tau$ this is $\tau_{PHL,}$ right, and this is $\tau_{PLH}$ will be equals to 2R times C. So means if I assume that NMOS's and PMOS's have exactly the same value of resistance as being offered during the ON state, then $\tau_{PHL}$ is RC by 2 whereas, $\tau_{PLH}$ is 2 RC, right and that is quite critical.

Therefore, low-to-high propagation delays 4 times larger as compared to high to low propagation delay. The same thing will reverse when you do a NAND-2 logic, this is for NOR-2. If you do a NAND-2 logic, for NAND-2 logic if you want to find out, then what will happen is that you will get an automatic reductions so $\tau_{PLH}$ will be equals to 2 RC and sorry

$\tau_{PLH}$ will be equals to RC by 2, and $\tau_{PHL}$ will be equals to RC twice RC I hope you understand the reason.

So therefore, if I for in this case if I want to make that my high to low and low to high are equal I need to size my PMOS in such a manner that by high to low and low to high are equal, so they will be in series so if I am able to make this one R by 2 and this one R by 2 then even in series these will add up, you will get R only and they will make me equal. So R by 2 primarily means that you make the width larger, but when you make the width larger then this capacitive loading also increases drastically, right. So this logical effort gives me an idea that what changes should I do in my circuitry in order to achieve the best possible results in terms of delay, right, with this basic introduction let me introduce to you the topic which we which we were supposed to do.

(Refer Slide Time: 8:44)




So, as I discussed with you that designing a circuit that is what I was saying that designing a circuit to achieve the greatest speed or minimum delay will give you large number of choices, right? So you need to choose which is the best choice in front of you. Now the idea is therefore, how large should a logic gates transistor be to achieve the least delay? Please understand that just making it large does not make it… make that achieve the least delay because we just now saw that making it large also increase the capacitive loading and therefore, your capacitive loading will be larger.

Now this is true for a single stage amplifier, single stage design, but generally in today's world you have multi stage circuits, right, so you have one stage driving another, driving

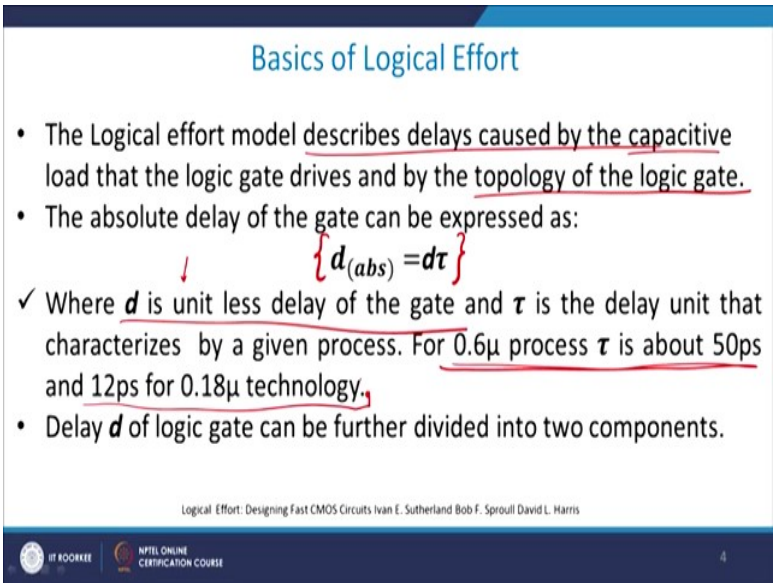
another and so on and so forth. In such a scenario if you look very carefully the first stage is loaded by excessively by the subsequent stages, so the subsequent stage capacitive load will be so high that it will try to reduce the delay, try to enhance the delay of the first stage.

So can we therefore, find out the total number of stages which gives you the minimum delay? So what we say the method of logical effort if you look very carefully what is written here then the method of logical effort is an easy way to minimize the delay in a logic circuit, right, so what we do? We compare the delay estimates of different logic structures. For example we compare for a NAND-2, NOR-2, XOR-2 and then we say that the fastest candidate can be given by this particular design.

Logical effort also specifies the best number of stages in a logical path and the respective transistor sizes for a given load, so which means that if I know for sure that my output load is $C_L$ or load is given to me which is exactly equals to some load then can we do a back calculation and say that okay my transistor size should be this much for this much load to be driven by an input to get the minimum delay, right?

So the overall justification or overall idea is that given a set of choice or given a set of paths or data path can we in some way or other find out a methodology by which we can optimize the delay, right? So this is what we get, right.

(Refer Slide Time: 11:22)



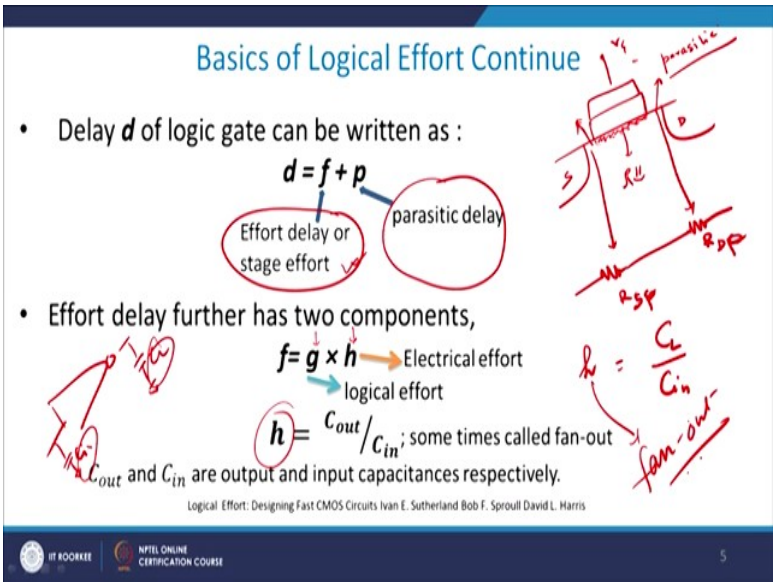

Okay, so two things logical effect takes care of is what is the capacitive load right?, so we have a so you will have a delay because of capacitive load right, and you also have a delay by virtue of the topology of the logic gate, this we were discussing in the starting slide that there

will be two delays associated with any design. The first delay will be the intrinsic delay which is basically because of the device, because of the capacitances and so on and so forth, right. Whereas, the second delay is basically depending upon the logic structure which you are using, are you using a NAND-2 logic, are you are using a NOR-2 logic, are you using what type of universal gate are you using in a digital logic in order to express it?

Now, the absolute delay of a gate can be expressed as, as you can see is given by this basic formula, where d absolute is equals to d multiplied by τ. Now d is basically unit less delay of the gate and it depends upon the type of process through which you are designing that particular gate, so this basically intrinsically you cannot do too much, you cannot do too much a manipulation in that because it is basically depending on the type of process the fabrication process through which the structure has gone through. So typically for 0.6 micron technology τ is approximately 50 microseconds, picoseconds and it is 12 picoseconds for 0.18 micron technology.

And that is quite interesting which means that as you go lower in the technology node, your intrinsic delay starts to lower down, so you see a lowering down of intrinsic delay of 50 picosecond to 12 picosecond, right. Now, so why do you multiply both of them? The reason being that, the reason you multiply is that τ is the delay unit, so if τ is the delay unit right, in picosecond, since τ is 1 picosecond and d is the delay unit by virtue of the fabrication, we can actually find out the total delay to be equals to d into τ, right.

(Refer Slide Time: 13:22)

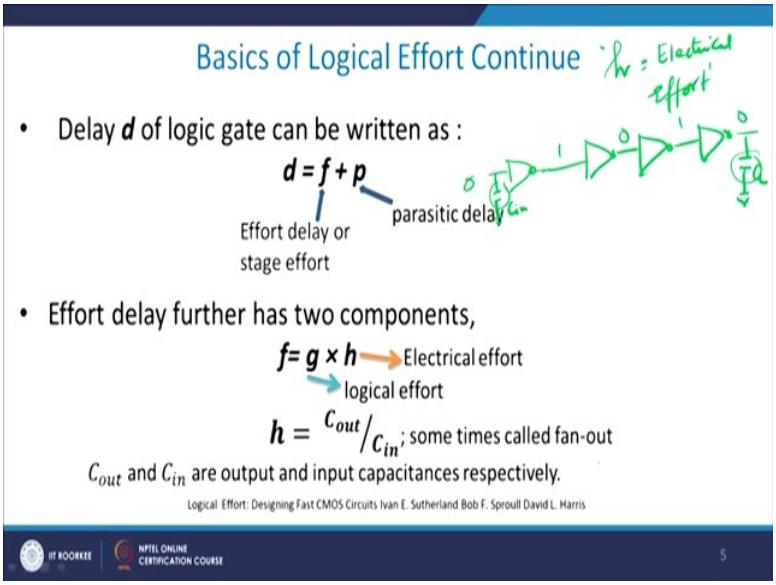


Now, the delay of a logical gate d can be therefore, broken up into two parts, one is known as a stage effort or effort delay and another is the parasitic delay, what is a parasitic delay? Well parasitic delay is very simple straightforward we already discussed this point earlier also just to refresh your memories, I will just do it, let us suppose you have a gate here and you have a gate here and you have a source and drain here, right?

Now, what is happening is that the inversion layer is formed only below this region, so this is where the inversion layer is form and you have large carrier concentration here the resistivity here actually falls down, so resistance is lowered much here, but you see if you look at this region under also as under lap region these two regions, this, as well as this, then typically in these two regions you will automatically have a much much higher delay or the intrinsic delay will be larger because there are no charge carriers here and therefore, this will act as a parasitic, which means that it is basically adding so this sort will happen is something like this it will be one R here corresponds to this, and then you have a straight wire and then one R will be here, so this is $R_1$ and this is $R_2$ or $R_S$ and $R_D$, so $R_{SC}$, $R_{DC}$, parasitic so $R_{SP}$ and $R_{DP}$, so $D_P$ is the drain side parasitic and S is the source side parasitic.

So the delay is by, one delay will be by virtue of this and you will also have a stage effort which is there this stage effort has got two component one is known as logical effort another is known as electrical effort, electrical effort is the most easiest one h and it is given as $C_L$ or $C_{out}$ by $C_{in}$. So, let us suppose I have a buffer or let me say a simple inverter and if my output load capacitance is $C_L$ and my input load capacitance is $C_{in}$, then $C_L$ this by this is effectively my h, right, this h is also sometimes referred to as fan out, right, so this is also referred to as sometime as fan out, $C_{out}$ and $C_{in}$.

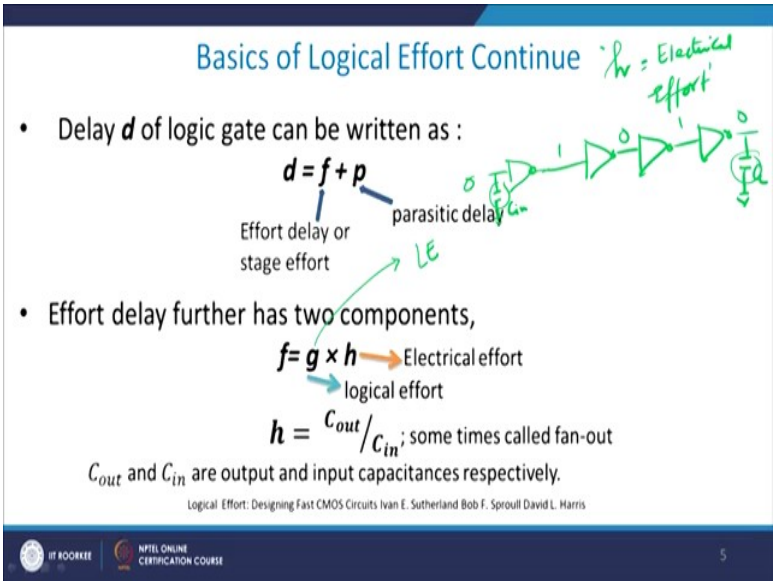
(Refer Slide Time: 15:42)

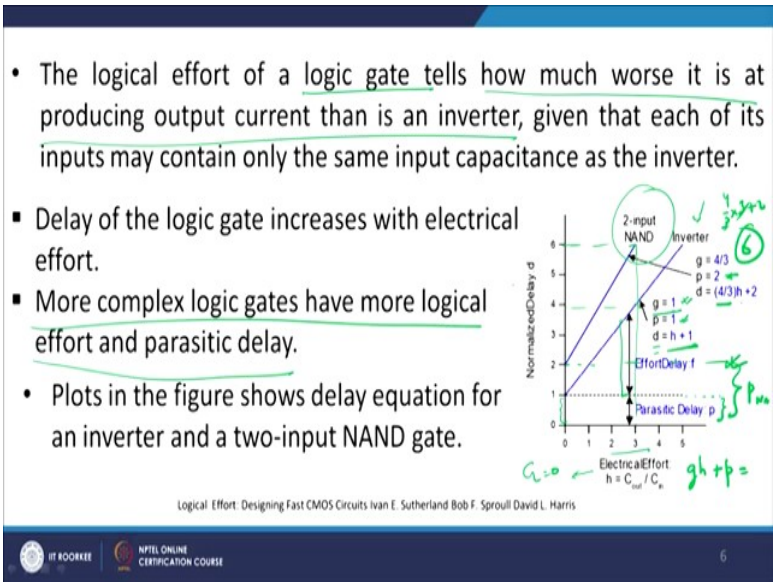



Now, you can also have something like this that you do have let us suppose you have something like this that let me say, let me say you have so you have let us suppose a buffer here which means that you have let us say this, so I will get 1 0 1 0 and I will let us put a 0, I will get a 0, so this is a buffer. Now, in the buffer this is the load capacitance which you see $C_L$ are external load and this is my $C_{in}$, so this, by this is referred to as the electrical effort, electrical effort French electrical effort is there.

So higher the electrical effort of course as you can see more will be the load capacitance as compared to $C_{in}$ and therefore, higher delay will be effectively large so f will be larger in that case if you get a higher electrical effort and that is the problem area which people face as far as electrical effort is concerned.

(Refer Slide Time: 16:46)




So what is a logical effort? Logical effort tells you that of a logical gate, logical effort for a logical gate tells you how much worse it is at producing output current than an inverter, right. So assuming that I have an inverter, I give an input to an inverter and I check out how much amount of current is flowing in the output side for a pair of inputs, then what I do I take the same pair of input and put it into a logical gate any logical NOR-2, NAND-2 whichever you want to calculate and see how much current is flowing, right.

If it is better good equal to an inverter this logical effort will let me know, right. As we have already discussed therefore, that logical gate, the delay in the logical gate will depend upon or will increase with the electrical effort, obviously it will do so, what is the electrical effort?

Because h, higher the value of h you will have higher will be the delay and therefore, why? Because your loading is typically very large, right.

And if you therefore, see a plot between electrical effort and normalize delay this graph here you see you will have a parasitic delay P is a constant one, this is independent of number of inputs, number of gates so on and so forth it is basically process and device centric and therefore, this parasitic delay is always constant. So even when your electrical effort is 0, you still have some amount of normalized delay. For example a two input NAND gate, we will talk about inverter later on. For example a two input NAND gate even with some electrical effort, electrical effort equals to 0, 0 means basically $C_L$ equals to 0, I will automatically have this two, this two is because of primarily because of the because of the parasitic delay which you see.

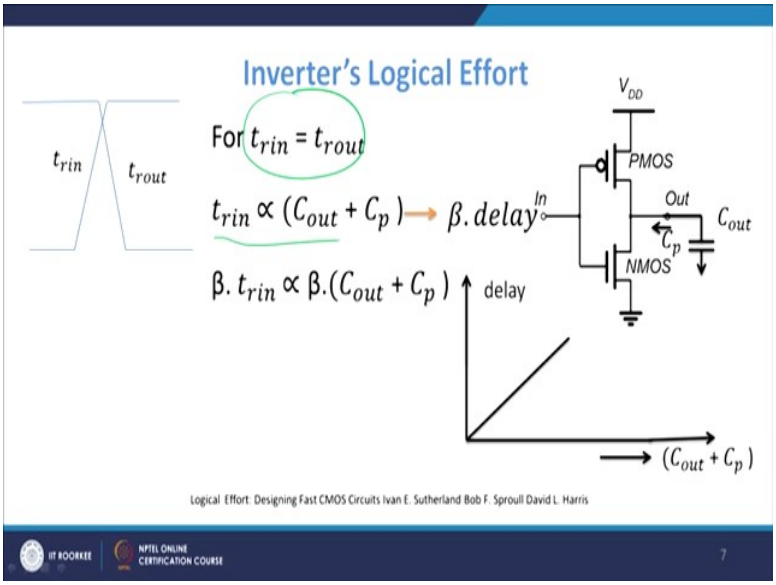
Now, if you go on increasing the electrical effort you automatically get a linear increase in the value of normalized delay and this is true also because then your load you in the external world is increasing and therefore, your delay will go on increasing, right. We define therefore, in this case, so let us suppose, let us suppose we take an inverter which is a skewed inverter, then we define this to be as a parasitic delay this p and this f is my effort delay, this f is my effort delay.

So at about approximately 3 the value of your normalized delay is 4 and we define g as equals to 1 and p equals to 1, what is g and what is g? If we look back to your previous slide, g is this, which is basically my logical effort, so g is my logical effort and if you go back to your here, what is p? The p value is the parasitic delay, parasitic delay is fixed always fixed, so I get f plus p equals to d, so what I am doing here is the p is 1, $c_p$ is 1, what is g? Electrical effort is also let us assume it to be 1 because for our invert rate is 1, delay which you get capital D is equals to h plus 1, what is h? h is basically your electrical effort and you get d equals to h plus 1, right, that is how you work out.

And therefore, more complex logical gates have more logical effort, so if we have got rather than a two input NAND gate, if you have a three input NAND gate life becomes more difficult and therefore, your logical effort is larger in that case, right and you also have a larger parasitic delay if for example in that case. For example if you look back here a two input NAND gate you see your parasitics are actually 2, this is parasitic for a NAND gate, so you see p equals to 2, we will see how why is it g equals to 4 by 3, but p equals to 2, right.

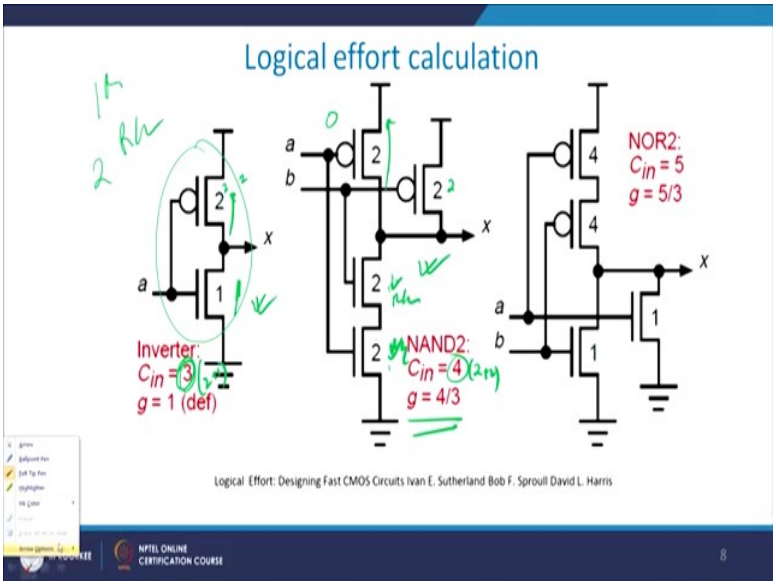
The formula is g times h plus the parasitic delay is your total delay, so you see 4 into 3 into h plus 2, right, where h is the your electrical effort, right, so if you are plotting here somewhere here let us suppose you are plotting somewhere here, then here h equals to 3, 4 by 3 into 3 plus 2 is equal to 6, so normalized delay is approximately equals to 6, fine. How we got 4 by 3 for g, we will see that just in the next subsequent slide. So I get for 2 input NAND gate, now if it would have more complex and let me say it would have been 2 inputs XOR gate then the graph will change drastically in this case as compared to an inverter.

(Refer Slide Time: 21:18)




Let me come to the next slide and let us see how a logical effort works or in virtual logical effort assuming that my input rise time and output fall time exactly the same I can write down input rise time is proportional to $C_{out}$ which is this 1 plus $C_P$, $C_P$ is the loading capacity which you see which is given as β times delay, β times delay or delay therefore, can be written as β times $C_{out}$ plus $C_P$ and if you plot $C_{out}$ plus $C_P$ versus delay you will always get a linear curve whose slope will be equals to beta. So this is a standard way of looking at it.

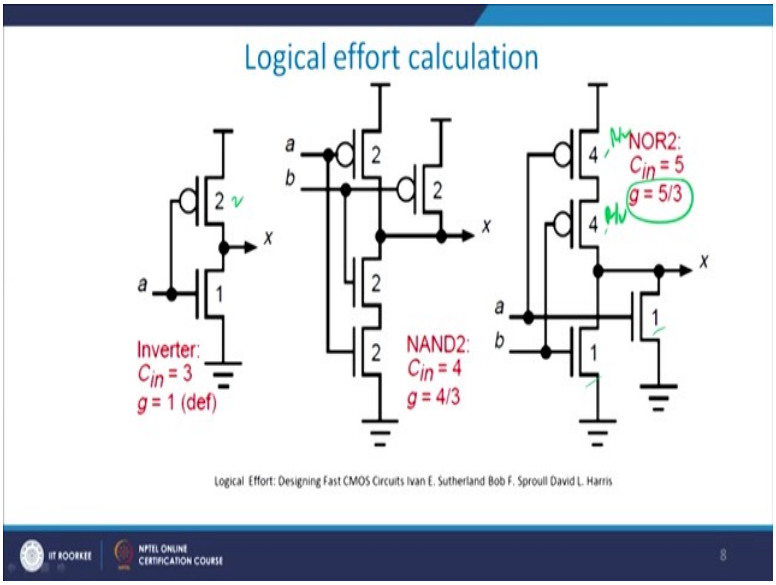
(Refer Slide Time: 21:46)




Now, how do you do a logical effort? You do it in this manner, that if you have a inverter which is basically as a single input inverter we make that PMOS is 2 times larger as compared to 1, so that $t_{PHL}$ equal to $t_{PLH}$, now we try to make a NAND-2 logic and we have two PMOS's in parallel and 2 NMOS's is in series. Now, if you want this delay to be exactly equal to this delay, you have to make the lower transistor size double, then only I will get R by 2, R by 2 available to me, for pull up there is no issue because pull up this is 2, for any 0 I get 2 here and therefore, my pull up will be equal.

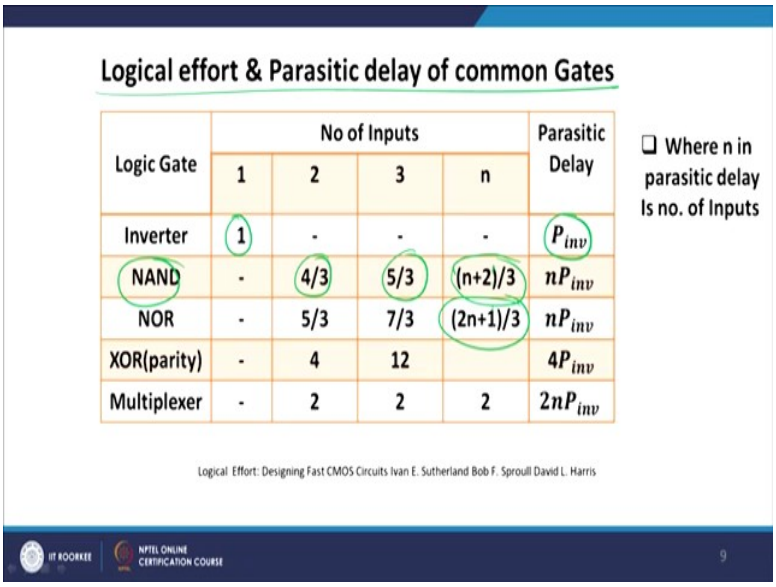
But for pull-down this is 1 and since this is 2, right, so this is 1, so if I have to double it because it is in series so this will be R, if 1 corresponds to R then 2 will corresponds to R by 2, so R by 2 plus R by 2 is always equals to 2 and the pull-down will be equal in that case. So pull up and pull down will be again 2 is to 1 available to you. So you see in this case $C_{in}$ is equal to 3, why 3? Because 2 for PMOS 2 for PMOS and 1 plus NMOS, why is it 4 here? Because for single gate it is 2 for this thing and 2 for NMOS, right, so we define g to be 4 by 3, 4 by 3 means 4 is the effective capacitance seen from the NAND-2 logic, right divided by equally sized inverter which will give you the same delay, so 4 by 3, right.

(Refer Slide Time: 23:30)



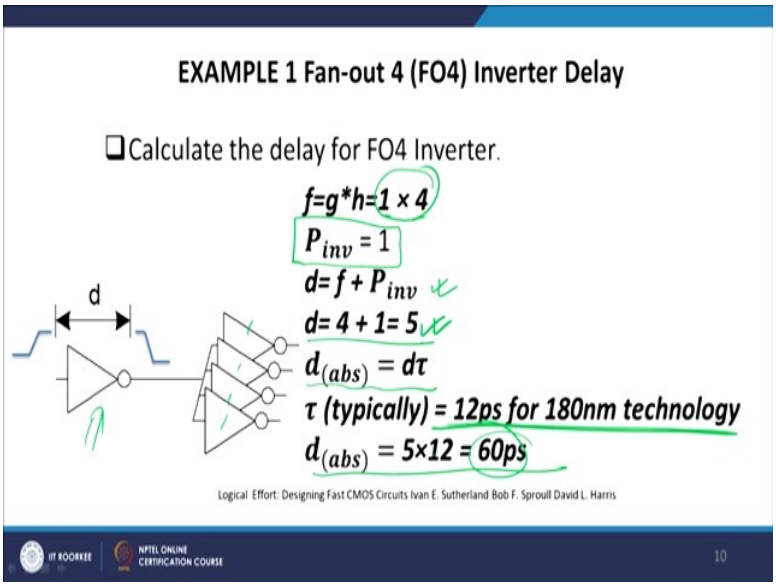

For a NOR-2 logic things change slightly and the reason is that they changed slightly, why? Because for NOR-2 logic understand if you want to compare this with this, then for pull-down I can afford to be 1 because both are in parallel so each one of them is 1 no problem, but in the pull-up I require this to be as 4 because then only it will become R by 2 because this is already 2, so this has to be R by, to make it R by 2, I have to double the size, so I am doubling the size so 4. So 4 means each block is looking at 4 plus 1, 5, so 5 by 3 is basically my g, which is which you see, right. So this is how you calculate the value of g, you compare that with a standalone inverter, right and that is how you get it.

(Refer Slide Time: 24:12)



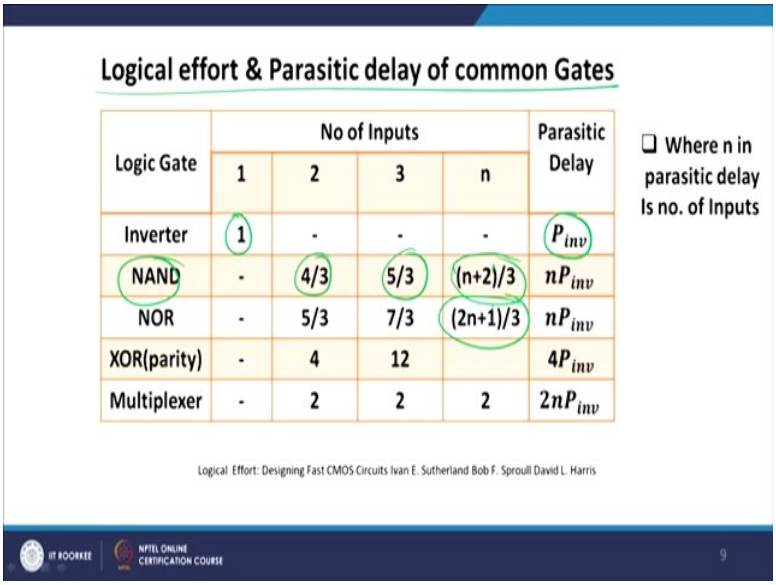
So if you look at the some of the logical efforts and common gates you see that inverter which is basically the 1 input obviously input has got only 1 for an inverter, parasitic delay is P inverter, for a NAND logic for 2 input 4 by 3, 3 input 5, so you see as the number of gates inputs increases the logical effort also increases, the general formula is n plus 2 by 3 for NAND gate and NOR gate it is 2n plus 1 by 3, where n is the number of inputs which you see right and that is a standard way of looking at it.

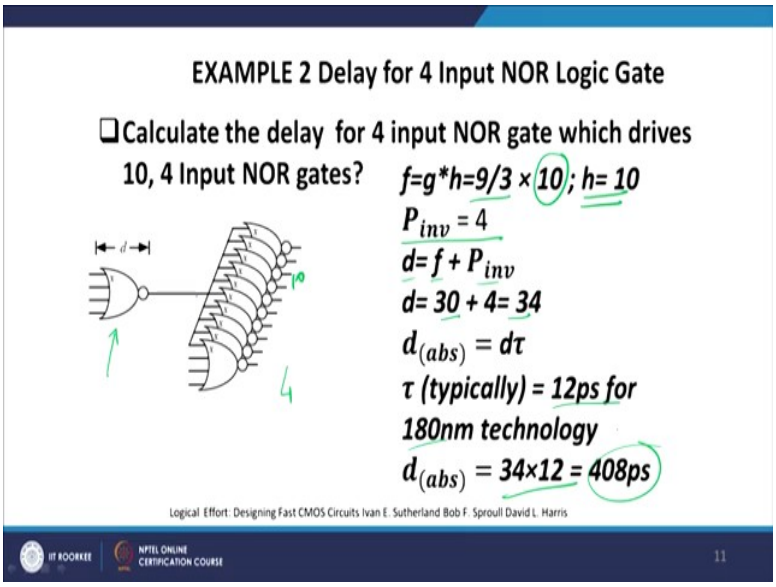
(Refer Slide Time: 24:44)




I will give an example of FO4 inverter delay and if I have a inverter something like this then f is equal to g into h, g if you see single inverter it is 1, so 1 into assuming that you are driving 4 such inverters, so 1 into 4 is the output 1, P inverter is the parasitic delay assuming it to be equal to 1 I get d equals to f plus P inverter, right f is nothing but 4 so 4 plus 1 is 5, right so d absolute is equal to d into τ typically as I discussed with you 12 picoseconds for 180 nanometer so I get d absolute is equal to 5 into 12 which is 60 picoseconds so I got d from here multiply that with 12 I get 60, which means that if I have a single transistor driving 4 similar transistors with fan-out 4, I will get approximately 60 picosecond delay between the input and the output.

(Refer Slide Time: 25:34)



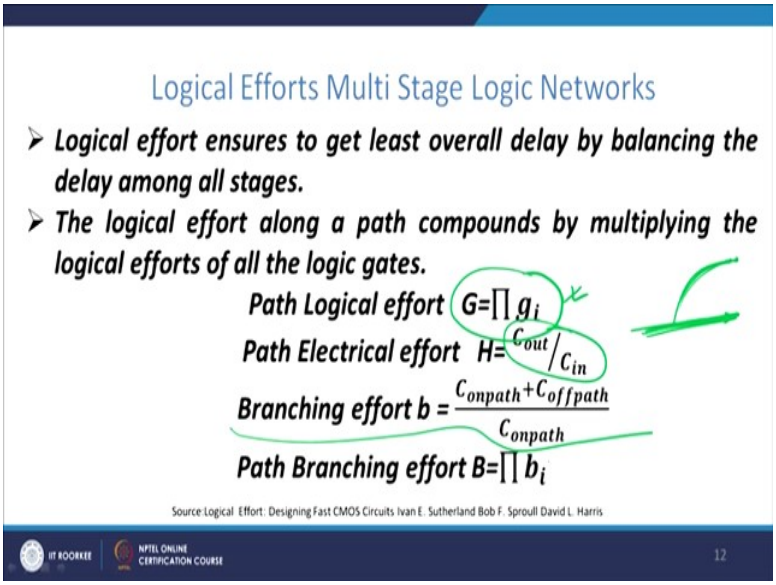




Let us say an example of a NOR-2 logic, let us suppose I have a 4 input NOR logic driving 10 such gates, so as I discuss with you f will be g into h, g in this case will be 9 by 3, right, it will be 9 by 3 you can check it for a 4 input NOR gate for example just let me just show it to you for a 4 input NAND gate 4 2s are 8 plus 1, 9; 9 by 3 which you get, right, and that is what you get 9 by 3 multiplied by 10 because equally sized so h will be equals to 10, right so I get and let us suppose the parasitic for inverter is 4, so I get d equals to f plus inverter so I get 30 plus 4; 34 and for again 180 nanometer 34 into 12 is 408 picosecond.

So therefore, a 4 input NOR driving 10 such 4 input NOR gate the picosecond 408 picosecond is the gate, so first you have to find out the f value, then you find out the if P parasitic is given to you, d equals to f plus p inverter when you once you find out the value of

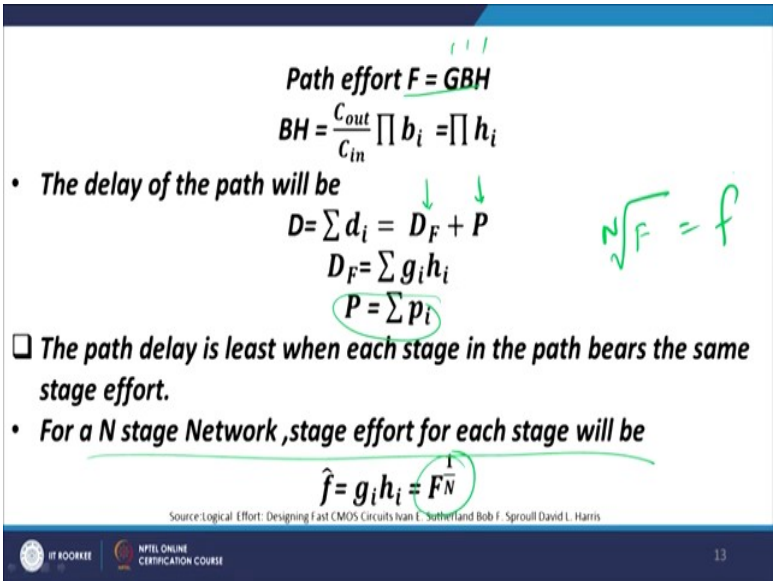
d you then find out multiply it by τ, τ is always fixed for a particular technology and from there you can calculate the total delay between the gates.

(Refer Slide Time: 26:55)



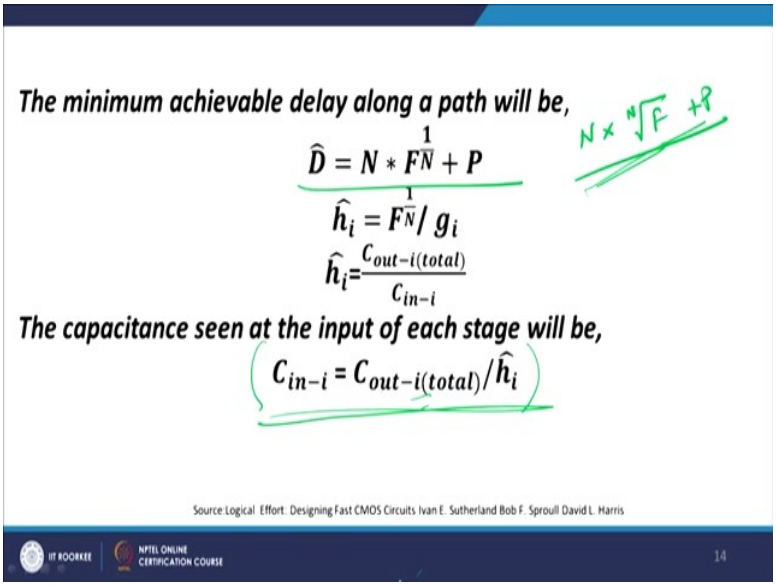

So what we will, what we will try to do is give you a brief idea about the logical effort for a multi, so path stage a path so now if you have a large path then we what we do is that we define a path logical effort, so in a large path we go on adding the logical effort for each one of them, we define path electrical effort as $C_{out}$ by $C_{in}$, we have already discussed this path, we define a new term which is known as branching effort which is $C_{on}$ path versus plus $C_{off}$ path by $C_{on}$ path, which means that let us suppose I have this and my signal goes via this path, then this is $C_{off}$ path and this is $C_{on}$ path, so we see the $C_{on}$ path plus $C_{off}$ path thereby $C_{on}$ path gives you this value and the total branching effort is given by this formula.

(Refer Slide Time: 27:34)




From these explanations I get that total logical path will be $P_i$, $P_i$ will be the individual stage so if you add all the individual stages you get the total branch effort, so what we do is, we different path effort to be equal to g b into h, right and from there we find the delay as this plus p which is p is the parasitic delay for each stage, right? We therefore, say for N stage network, stage effort for each stage will be F to the power 1 by n so the square root or $n^{th}$ root of F is my stage effort which we see, right?

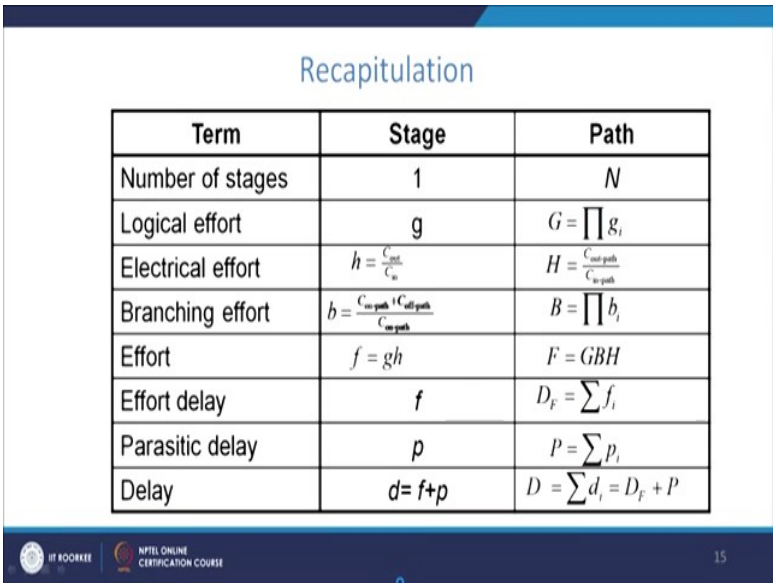
(Refer Slide Time: 28:10)




So what we do is we delay we define as N into F, so N into $n^{th}$ root of F plus P is basically by delay, right and that is what you get the total delay minimum achievable delay along a path, right. Now, so the minimum capacity $C_{in}$ where capacitance $C_{in}$ at each stage will be given

by this formula, where $C_{in}$ is basically the input capacitance of $i^{th}$ stage and I get total in the output stage for the $i^{th}$ stage I get $C_{out}$ divided by $h_i$ cap, $h_i$ cap is the basically the logical effort this F root N divided by logical effort.

(Refer Slide Time: 28:45)




So, let me recapitulate therefore, if we have logical effort g, electrical effort as h, branching effort as b, then we define the total effort f is equal to g into h and effort delay to be equals to f and p and therefore, the total delay is always equals to f plus p, how do you find out f? g into h you find out, so in a single stage, so in multiple stage you can do it, so first of all find this and then you multiply with this and then you then you add to p and you get the total delay, right.

So we have understood two important points, I will come into details of this one at later course also, but at this stage we actually understood that given that means if a particular logic gate is driving some other logic gate, right maybe it is a N input logic gate we can find out the delay between point A and point B for the design, right that will help you to gain in this area, with this let me thank you for your hearing, thanks a lot, thank you!

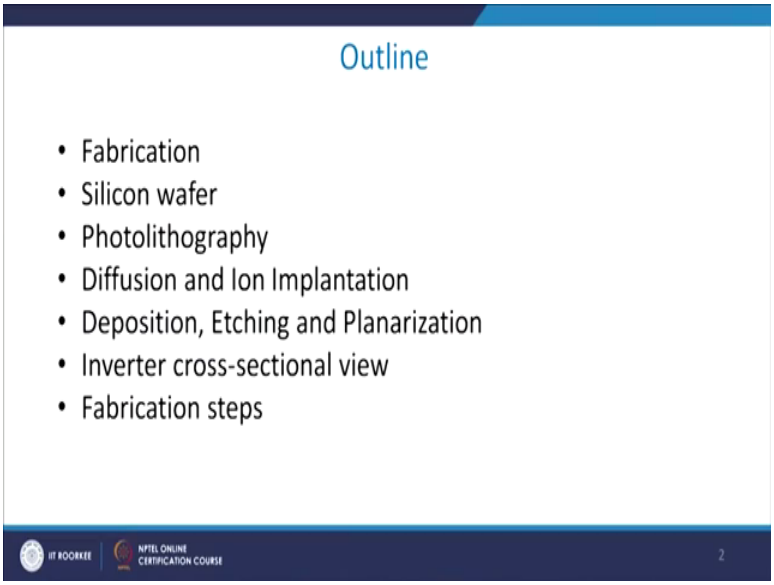
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−23**
**Fabrication Process - I**

Hello everybody and welcome to the NPTEL online course on Microelectronics: Devices to Circuits. This module is dedicated to the fabrication procedure for an ordinary silicon based design. So, what we will be looking into this module is if you want to manufacture IC or integrated circuits on silicon then what are the steps you should do.

In our previous discussion we have already discussed about in detail about bipolar junction transistors and metal oxide semiconductor field effect transistors and we have also seen what are its possible usage in both digital and analog domain. We will be taking up MOS as an amplifier in our subsequent modules. So, typically as the course is progressing you might be seeing that I am giving you a flavour of both digital as well as analog electronics, right?

Subsequent few modules will be devoted for analog and during the later courses or later sections of this module we will be actually moving to, when we go to week 10, 11 and 12 we will be actually looking into digital applications of these MOS devices. So, this gives you a full breadth of a design flow as far as silicon MOS device is concerned. Now, we have already looked into those basic facts. Now, let me give you a basic idea about the basic fabrication steps which is available with us typically in the design.

(Refer Slide Time: 2:07)

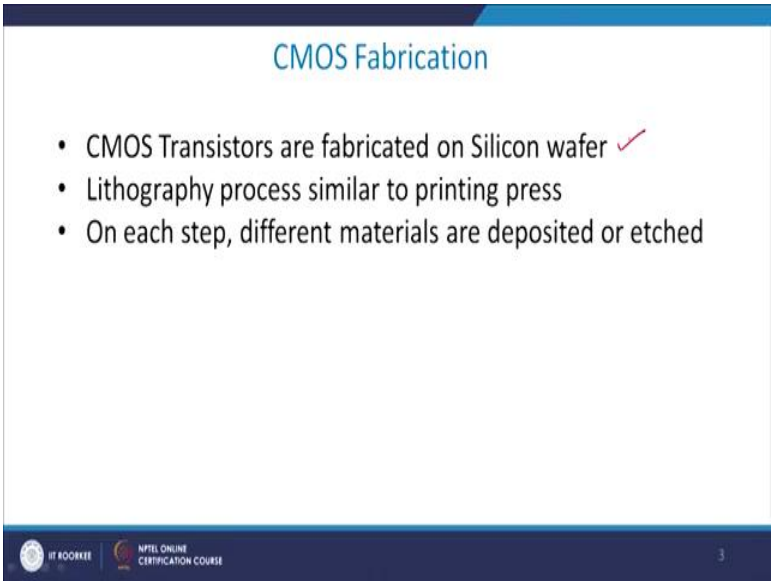



Now, the outline of the talk therefore will be we will be going to the basics of fabrication, right. We are looking into the basics of fabrication. We will actually look into the various concepts or various meaning. What is the meaning of silicon wafer? What are the limitations of silicon wafer? We will also be looking into a very important term known as photolithography which is actually determines the channel length or the dimension of the devices. So, that is photolithography we will be looking into.

We will be looking into diffusion and ion implantation. These are primarily meant for doping doping and then this doping if we want to change the doping or we want to alter the doping then we use diffusion or ion implantation system. Then, for example, if you want to do a metallisation, for example you want to grow a metal then you need to do a deposition or even maybe a simple silicon dioxide growth over silicon will require a deposition.

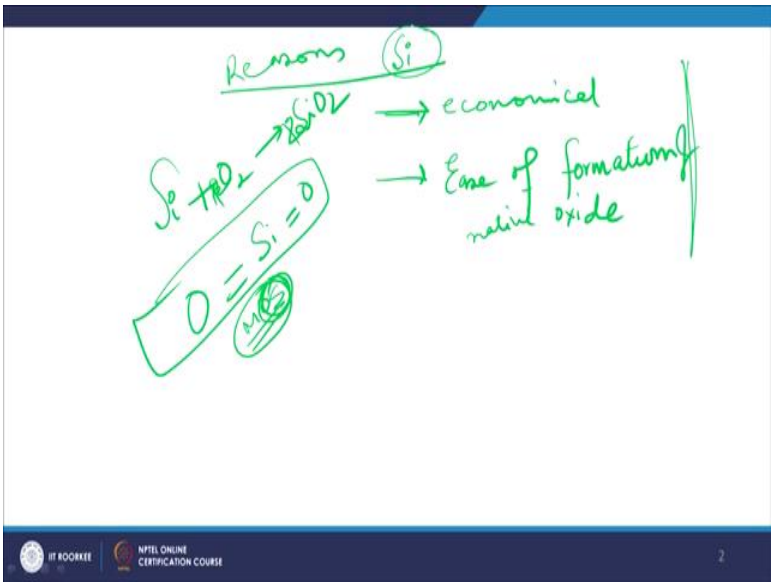
We require etching because etching will help us to remove the areas where you want to grow metallization. And then planarization means making it plane by doing a by doing a polishing. We will therefore, then going into the inverter cross sectional view from the fabrication point of view. And then finally we will look into the various fabrication steps in this design, right? So, that is a typical outflow for this for this outline for this design for this module. Let me see what is the silicon fabrication?

(Refer Slide Time: 3:43)




Most of the time as you can see CMOS transistors are fabricated on silicon wafer, right? There are a few reasons for that silicon for 2 reasons. Silicon is first of all, if you go back very clearly, silicon if you look very carefully then you will understand that silicon is very economical in nature, right?

(Refer Slide Time: 3:56)



It is very economical. So, that is the reason we use silicon, so reasons for using silicon. If you want to use reasons for silicon, it is pretty economical right, and the second is ease of formation of native oxide, of native oxide. So, if you leave even a simple silicon wafer here it mixes with oxygen to form silicon dioxide, right? So, this is what you get silicon plus silicon dioxide which means the silicon automatically does a double bonding with O and forms silicon dioxide.

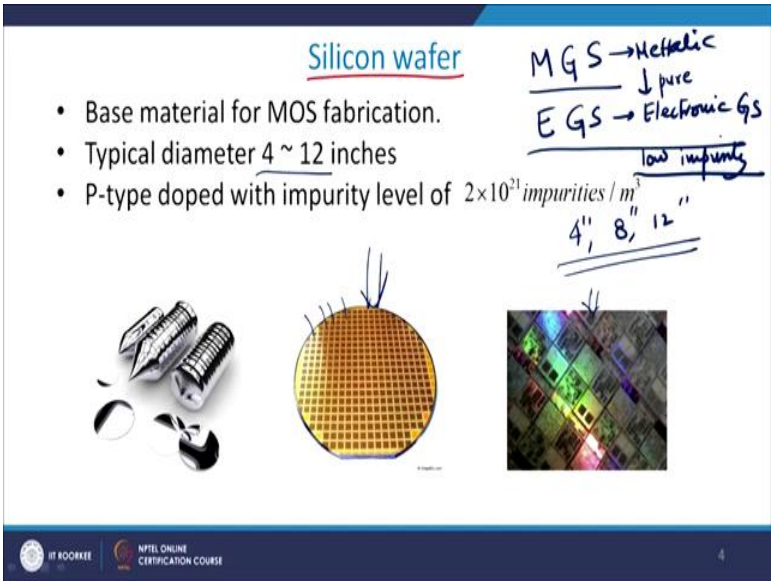
So, formation of native oxide is very easy. So, if you want to grow a MOS, this MO metal, sorry MOS, OS which is oxide and silicon is a very natural growth, right? So, that is the reason we use silicon for all practical purposes, right? The cost we pay for it is, for example, we cannot use silicon for optical domain, right? And that is the problem area of silicon then we have to go for III-V, III-V semiconductors. For the time being, we will be concentrating on silicon itself.

Now, the lithography as I discussed with you is just like a printing press. So, what you do is, you have a sort of a silicon wafer over which you want to write something by using by using certain techniques that is known as lithography. Now, every lithographical steps will have different materials. For example, if I want to grow interconnects then I will be using copper in the lithographic techniques. If I want to grow, for example, just the contact maybe I will be using something else, maybe I will be using aluminium to do that and so on and so forth.

So, depending upon the usage of those photolithographic (tech) lithographic techniques, you generally have photolithography an electron (E beam lithography) or direct E beam writing also is there. But these two are the most important ones. Photolithography as the name suggests is very simple. If you if you go back to your basic, this photography right; when you do a negative, what do you is that you do have a negative and you let the aperture open for a small duration of time, light flows in, right?

Wherever there is a human being or a structure outside the reflectivity is low. As a result, the chemical within that negative gets reacted less slowly and other parts gets more reacted and therefore, when you make a positive one you automatically get the image. Exactly the same thing happens in silicon as well. So, photolithography is an important step on each step we use different devices.

(Refer Slide Time: 6:39)



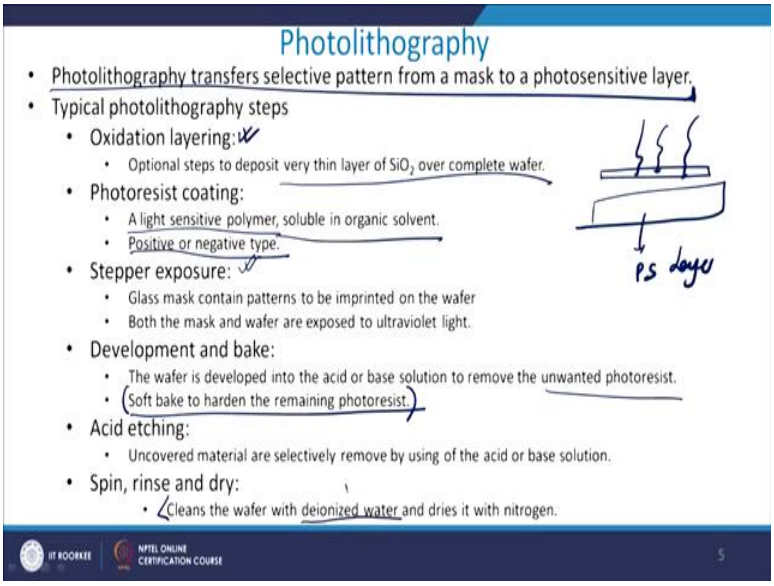

Now, the starting materials is what is basically what is known as a silicon wafer, right? There are 2 types of silicon available to us one is known as MSG which is known as, sorry MGS, sorry. Let me rub it for you. It is basically, it is basically your right, it is basically metal metallic grades silicon. So, it is metallic grade silicon. You also have an electronic grade silicon. So, this is basically your electronic electronic right grade silicon and this we have metallic grade silicon. So, as you can appreciate this is much more purer, so this is much more pure as compared to metallic grade.

So, typically when you fabricate silicon wafer, you first fabricate MGS and then you do more refining to get EGS. So, in EGS you will have very-very low impurity, very-very low impurity both in terms of doped external agent in terms of external agent as well as the crystal itself. The crystal itself will be so even that you would not be able to distinguish between a bad crystal and a good crystal, right? Typical diameter is from generally 4, 12 so we typical have diameters of 4, 8 and 12 inches, right?

So, these are fixed. You cannot have you cannot have more than these diameters of silicon because there is certain fundamental problem area if you go beyond this particular diameter. You will not be able to stabilize the system and moreover there will be large amount of defects inserted within silicon if the diameter of the system silicon grows too large, right? Okay! So, this is the photograph which you see is basically that wafer has been cut into smaller parts. So, each

one of them is basically a silicon chip here and you can actually start doing wafering there. This is the final form of wafer which you see where you are actually doing the layout and the designing.

(Refer Slide Time: 8:37)




With this knowledge, let me come to the photolithography. Photolithography transfers selective pattern from mask to a photosensitive layer, right? So, I have a mask, for example, if you mask right and if your mask is an open mask here like this then light can filter through this mask and fall onto a maybe a photosensitive layer. So, this is my photosensitive layer, photosensitive layer, right? When it falls there is some chemical reaction. The first reaction is basically your oxidation layering, right?

Typical photography step is that, it is an optional step why because you do not want the wafer to go bad. So, what people do is that let a small width of silicon dioxide grows over silicon. So, a small layer of silicon dioxide is grown over the silicon, right? Over the complete wafer. When you do a resistive coating then what you do is that a light sensitive polymer, photosensitive polymer which is soluble in an organic solvent or an agent and you put a layer of that over this silicon dioxide, right?

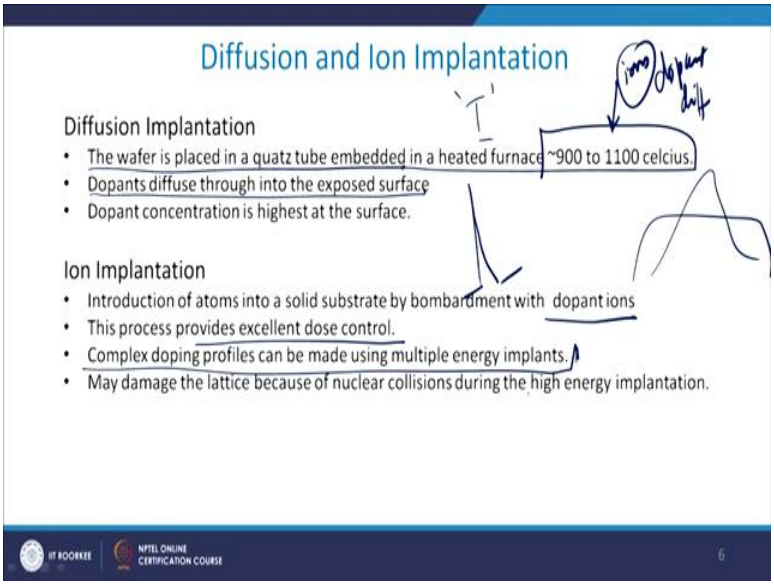
It can be positive; it can be negative as well. Positive basically means you can drive it out and let the material remain there and negative means let the material come out the other parts be there.

Now, my third point is basically my stepper exposure. So, you do not have a glass mask over which you have imprinted the pattern by which you want to photolithography. You put it over the wafer and then both mask can wafer are exposed to ultraviolet light. So, when the UV light falls wherever there is a mask there is a transparency the UV light will fall on the silicon and will expose it whereas those areas where you do not want to expose it will remain fixed like this.

Then the wafer is developed into acid or base solution to remove unwanted photoresist. So, whenever unwanted photoresist will be there, you just remove it by via chemical etching and then soft bake to harden the remaining photoresist. The remaining photoresist will be what will be useful to you. So, what you do is soft bake, soft bake means that you do it at very low temperatures and automatically only those photoresists remain there, right?

Acid etching basically means that once you have actually done all these things and removed unwanted photoresists and you have hard baked your photoresist which you want or the required photoresist then you do acid etching. What is acid etching? You uncover by using acid or base you do the partial removal of the uncovered material, right? So, that you do not end up having spots in your wafers. Then you do what? You clean the wafer with a deionized water distilled water also referred to as deionized water and dry it with nitrogen, right? So, we get the wafer with us in a detailed fashion.

(Refer Slide Time: 11:27)



Now, let me do a diffusion and ion implantation. Well diffusion as the name suggests is basically for (purp) generally for the purpose of dope inserting dopant species and as the name suggests what does doping do? It carries species from region of high concentration to a region of very low concentration and as a result what happens is that there will be a diffusion species between point A and point B.

Depending on the concentration gradient as well as on the diffusion constant of that particular material, the particular material will either sink deep inside the original species or if it is not much ahead of it, right? So, as I discussed with you dopant, so typically the wafer is placed in a quartz tube because it is quite pure and you do not have any reaction on quartz and it is heated in a furnace with 900 to 1000 degree centigrade. So, when you heat it becomes more reactive in nature, right?

And therefore, you allow dopant species to diffuse through it, through the exposed surface. Of course, as you can understand the diffused species dopant concentration will be highest at the surface and as you go inside the surface, the percentage of the dopant species starts to fall down, right? And these are standard problems area which people face while doing doping is that the doping is always highest on the surface and goes on reducing as you move within the surface, right? And therefore, the doping concentration is highest on the surface.

Please understand diffusion is again temperature dependent phenomena, right? So, on chip temperature you do which is basically by heated furnace will determine the most optimal value or the procedure for fast movement of ions, right? Fast movement of dopant ions in this case through using diffusion. So, it is very critical to set your temperature of the quartz tube in such a manner that it is the most optimized one.

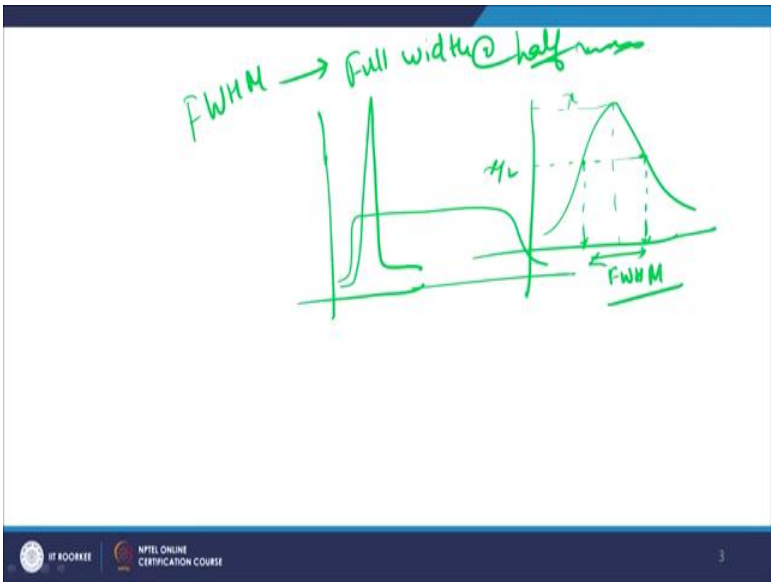
So, for example, if you keep it very largely heated up the chances are that the dopants will move so fast that it might not sit even with the silicon because energy is very high. Similarly, if the gradient is very-very small then it would not be moving from point A to point B by virtue of diffusion and that is you should be very careful about as far as diffusing an ion implantation is concerned. Let us look at ion implantation.

What is ion implantation? Ion implantation is bombardment of dopant ion using an accelerator. So, you should have a good accelerator. This process provides an excellent dopant control, right?

And complex doping profiles can be made using multiple energy implants. Say you want to first of all make something like this, right? And then you went on to make something like this. This is the dopant profiles. You can easily do using ion implantation, right?

For example, you want to make a profile something like this up so peaked, right? Then use a Gaussian profile and make your FWHM - Full Width at Half Maxima as possible and keep your doped D as large as possible, dopant species as large as possible. So, within a short domain, since, we have developed FWHM, FWHM is at Full Width at Half Maxima, right?

(Refer Slide Time: 14:55)

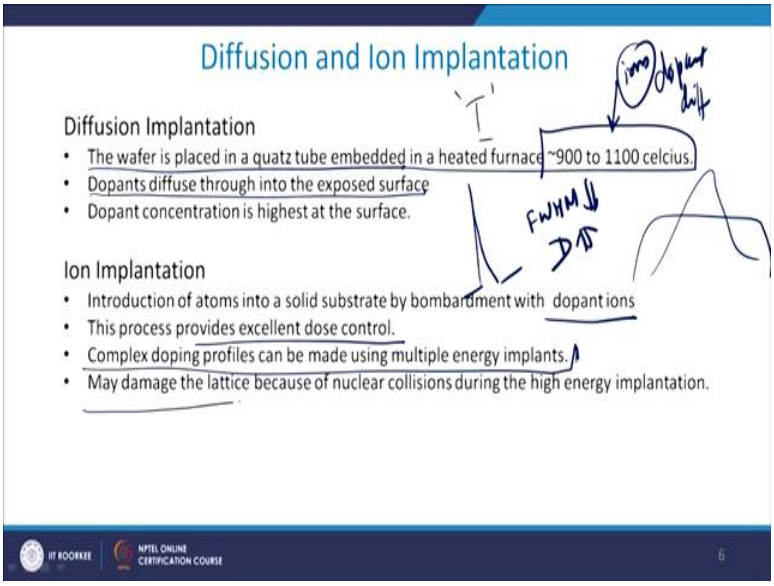



FWHM if we want to point out FWHM. This is basically Full Width at Half Maxima which means that if I plot using a may be a Gaussian surface, then we try to find the maximum value which is this one, X. And we draw a graph which is from X by 2. This is X by 2 and then we check the values of dopant species here and then we say that this is my FWHM, fine? So, this is Full Width at Half Maxima, right? So, when you get Full Width at Half Maxima it means.

So, now if you want that the profiling should be peaked or you should have a peaked profile, keep your FWHM small and you can actually get a peaked profile. Similarly, if you do not want a peaked profile you want a wide profile then it is also advisable to keep wide profile and you get something like this. So, depending upon the type of dopant species you are using and the type of

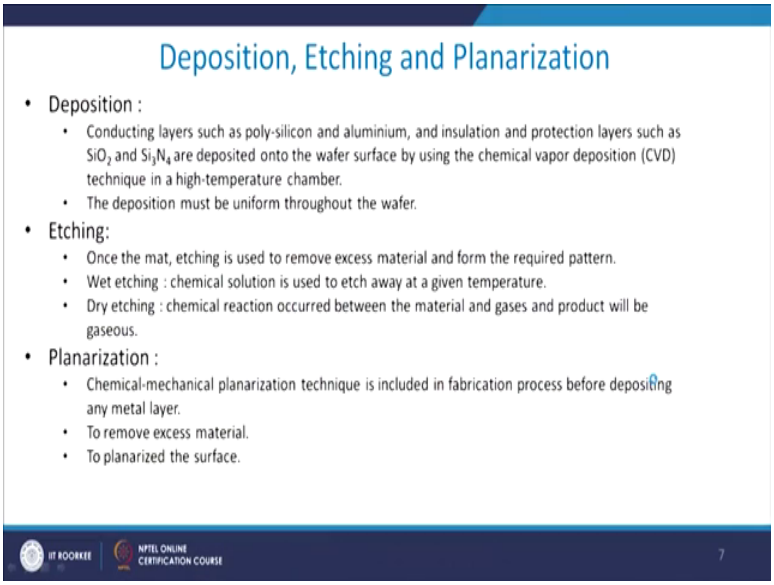
temperature which you are using you can actually have either a either a like you can either have a proper control over the channel and ion implantation and other channel wise.

(Refer Slide Time: 16:06)




So, but the problem with this ion implantation is that the energy is so high the energy is so high that you end up having maybe dislocation of the atom itself through which these charged particles are passing, right? And that is a problem area which people face.

(Refer Slide Time: 16:25)



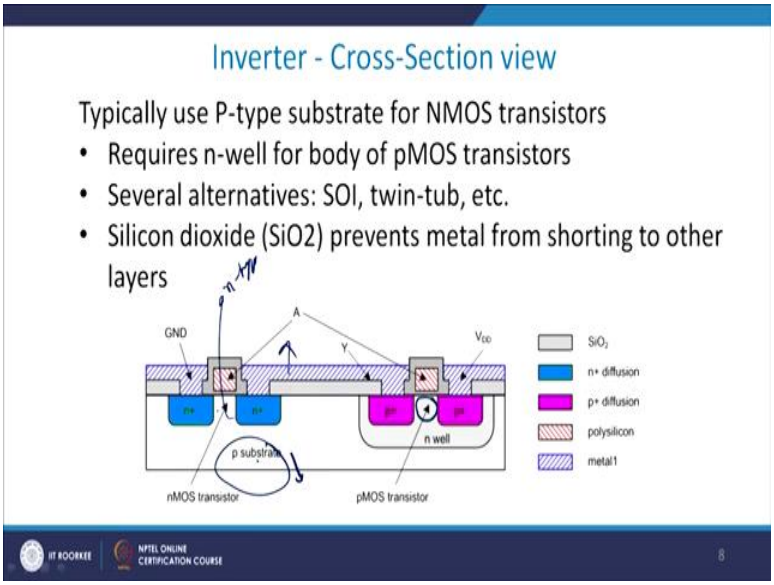
Let me come to deposition. When deposition is conducting layers such as poly-silicon and aluminium, and insulation and protection layers such as silicon dioxide and silicon nitride and they are done by CVD process at high temperature high temperature chambers. So, I have a high temperature chamber, right? So, once I do a CVD which is a chemical deposition and I deposit layers of silicon dioxide and nitride which is both by the way directive in nature over this area.

Now, so, once you have actually the material and then you have been covering it with the material across the board then what you do is that depending upon the pattern where you want to photolithography to work, you start to remove; if you are doing a positive photolithography start to remove those areas or etch those areas through which you want the light to go inside, right? And you do that easily using any of the dry etching or the wet etching, right?

So, chemical, in case of wet etching we use a chemical solution to etch away at given temperature. What is dry etching? Chemical reaction occurred between the material and gases and the product will be the gaseous itself, right? So, seldom people use dry etching most of the people use wet etching and that is the reason why, the reason for wet etching is simple but the cost you pay for it is that you are not very accurate in terms of appropriate etching, right?

What do you mean by planarization, it basically meaning chemical mechanical planer mechanical planarization and is used to techniques fabricate any metal layer to remove excess material and to planarize the surface, right? So, what will we do is that they have a planarization technique by which excess layer is removed first of all. And after the removal of excess layer what you try to do is that you try to have the model layer perfectly plane in dimension.

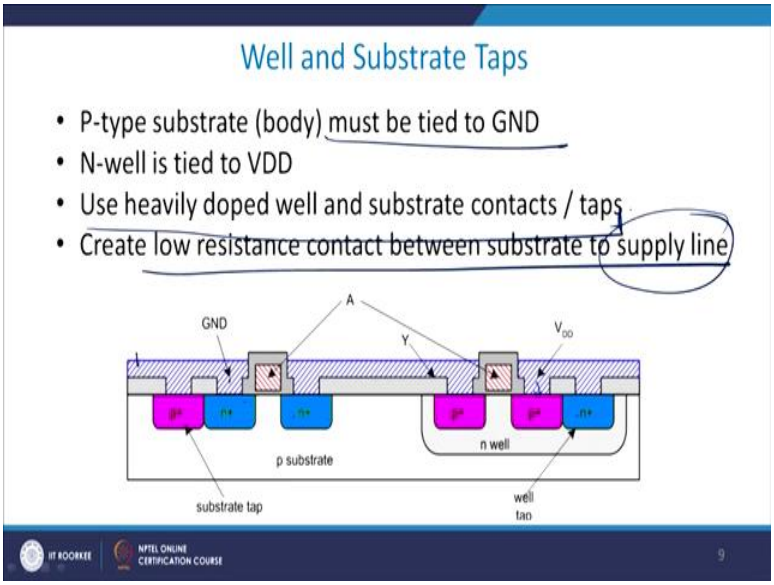
(Refer Slide Time: 18:24)




So, let me see inverter cross section view. In this you can basically see it is a twin-tup process and in twin-tup process I have p type substrate. This will act as the substrate for NMOS. However, you need to grow an n-well within the p substrate, for a p type substrate to be formed here, right? So, for a p type substrate to be formed here you require a, you want to have a pMOS transistor, try to make this one. So, this is basically your n-well and you try to make within the well another well which is basically n-well and from there you can get the value of your p type material.

Similarly, the p substrate itself adds a substrate to the MOS transistor which is basically your n type. And therefore, this is your n type substrate which you get. So, I want n type and p type, they are separated by this metal 1, right? So, this metal 1 is separating the 2. And from here and here and here you take the output value of voltages or currents, right? Okay!

(Refer Slide Time: 19:18)

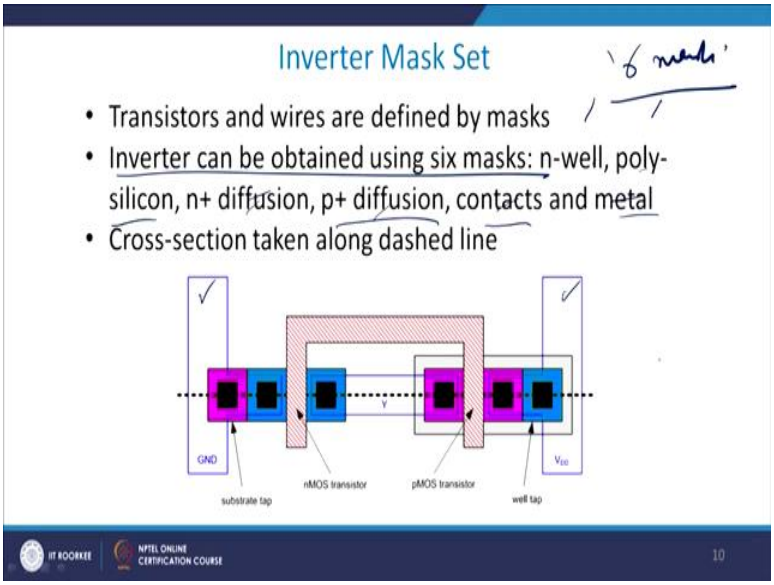



Let me look at the well and substrate taps. So, I have a well tap which is basically my, the well tap and substrate tap if you look carefully. This is my ground tap, right? This is my ground tap. These 2 are the taps where you put an input, right because this is gate to, this is gate to source. This is also gate to source and therefore, I would expect to see these to be latched together and this is the value of $V_{DD}$ which you use. And here is another well available with you, right? So, these are a few n-well processes which you use which is typically there with us.

Now, any transistor or mask, sorry, any substrate must be tied to the ground. So, if a p type substrate that must be tied to the ground, right? So, I have a p type substrate always tied to the ground and n-well is always tied to the $V_{DD}$. N type devices are always tied to higher value of voltage and p type devices are always tied tied to a lower value of voltage, right? And use heavily doped wells and substrate contacts taps. So, if you use if you use very heavily doped wells and also substrate contacts, your effect of n will be reduced drastically in a device, right? Now, you have moreover you have to create a low resistance contact between substrate and the supply line.

Otherwise, if you do not do that there will be excessive voltage loss across the across the across the supply line. And therefore, effective voltage available to you for running the MOS device would be relatively small, right? And that is the problem over here.

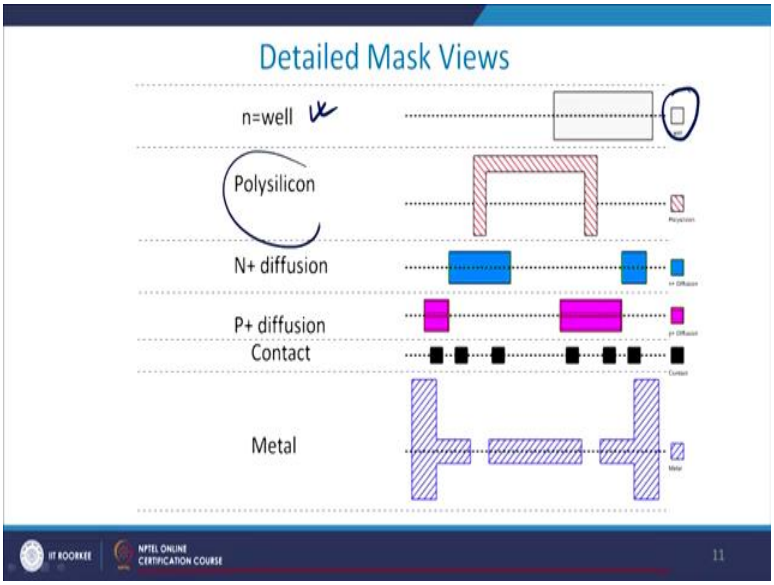
(Refer Slide Time: 20:44)




Now, let me come to the inverter mask set, right? We have got transistors and wires are defined by masks. So, these are basically the masks which are available with you. And inverter can be obtained by using 6 masks; n well, poly silicon, $N^+$ $P^+$ diffusion contact and metal, contact and metal. So, it is 1, 2, 3, 4, 5 and 6. So, you require 6 masks in order to do it right and 6 masks can easily be purchased and can be used.

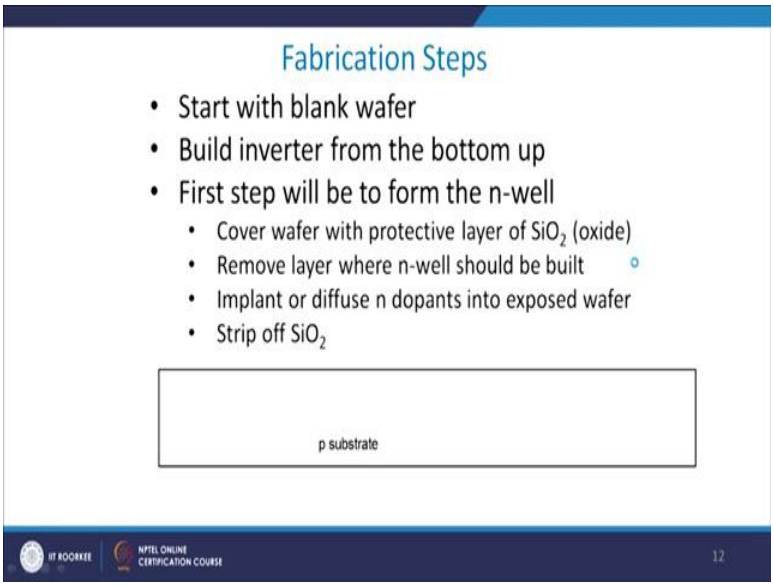
But, when you use these masks for fabrication purposes the fabricator will come to you and ask you to change this mask according to his fab lab right according to his fab lab. So, please be careful that just by saying that you have been able to generate the fabrication mask means that you have done everything, no. Fabrication mask as well as its final utilisation is quite important as far as inverter mask is concerned.

(Refer Slide Time: 21:38)




Let me give you a brief idea nothing to worry about you have to just learn. For n-well this is the mask which you see in front of you for p poly-silicon it is this thing, right? For $N^+$ diffusion we have we have $N^+$ diffusion shown by red. This is for $P^+$ diffusion, right? $P^+$ and this is the contact which is given by the black one, right? So, so if you look at this point metal is this and this are metal, right? And this is also a metal which is available with us and it is connected in series like this, right?

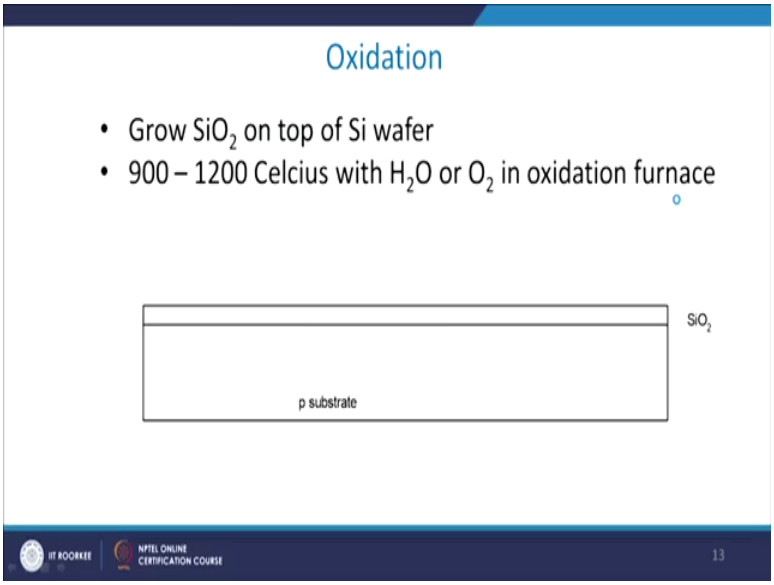
(Refer Slide Time: 22:06)



Let me start now with a blank wafer and let me innumerate the fabrication steps which are there. We start with the blank wafer build inverter from the bottom up. First step is to form an n-well cover wafer with protective silicon dioxide we discussed this point, right? Sorry, with silicon dioxide remove layer where n-well should be built. I think that is clear why n-well should be built. So, remove the layer where you want to remove where you have where you have plan to have n-well and implant or diffuse dopant into exposed wafer, right?

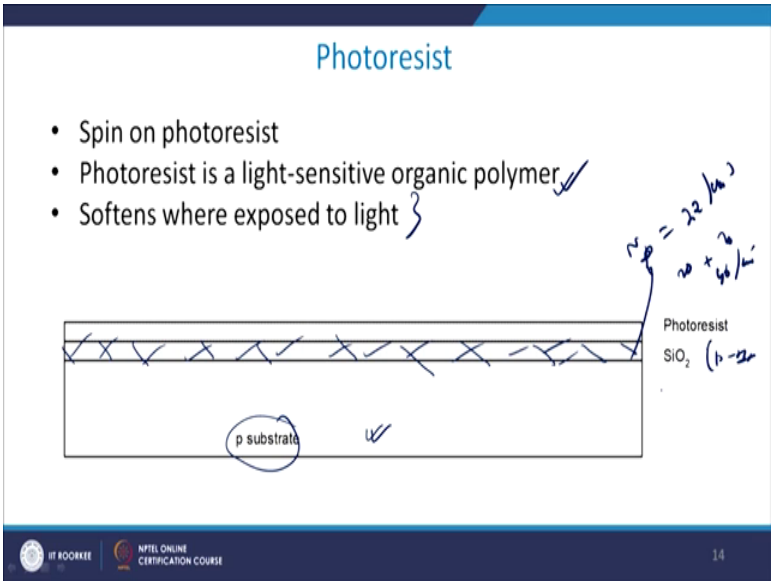
So, we appropriately change into n type or p type and then strip off silicon dioxide. So, you have a silicon dioxide initially present over that you form the n-well. And after the formation of n-well you try to do a protective covering of silicon dioxide. Remove it and remove it, and try to do your implantation or diffusion using dopant species which are basically your non silicon dioxide based, right? And then if you are using silicon dioxide based, then strip off the silicon dioxide and throw it outside.

(Refer Slide Time: 23:12)




Now, oxidation is you have to grow silicon dioxide on top of silicon and the temperature range is basically 600 to 1200 degree centigrade, right? With $H_2$ Oand $O_2$ is basically the reason you are putting this acts as catalyst, right? They act as catalyst means they themselves do not part in the reaction but they hasten the speed of the reaction in a much more detailed manner. So, we use oxidation for doing all these things and it is done at 900 to 1200 degree centigrade.

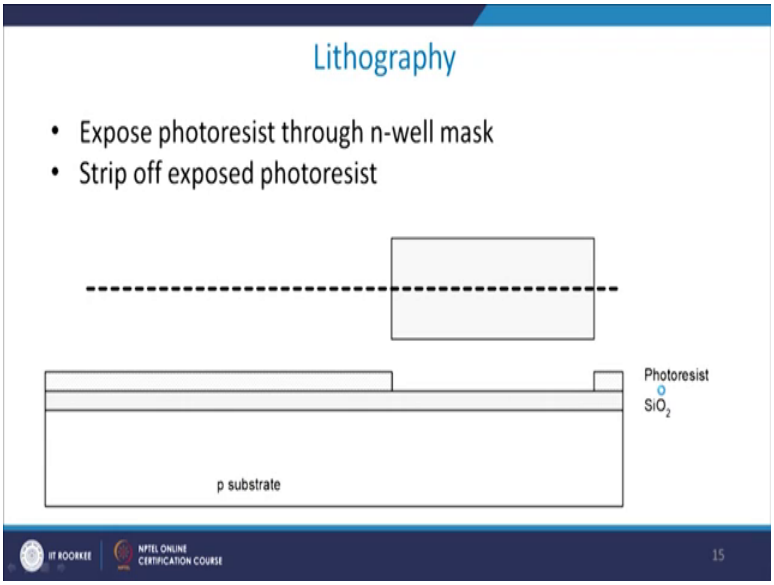
(Refer Slide Time: 23:40)




So, what is a photoresist? Photoresist as I discussed with you is a polymer photosensitive material in which if the photons fall, it becomes sensitive to input voltages and there is a problem. Now, as I discussed with you photoresist is a high-sensitive organic polymer. For example, there are many organic polymers available and what are the examples that happen to be photoresist? Now, so when you expose to light it softens that the photoresist itself softens when light is exposed to it. So as you can see here I have the native p type substrate I what do you mean I do something like this which means that I try to make this internal one as to be as silicon dioxide maybe p type.

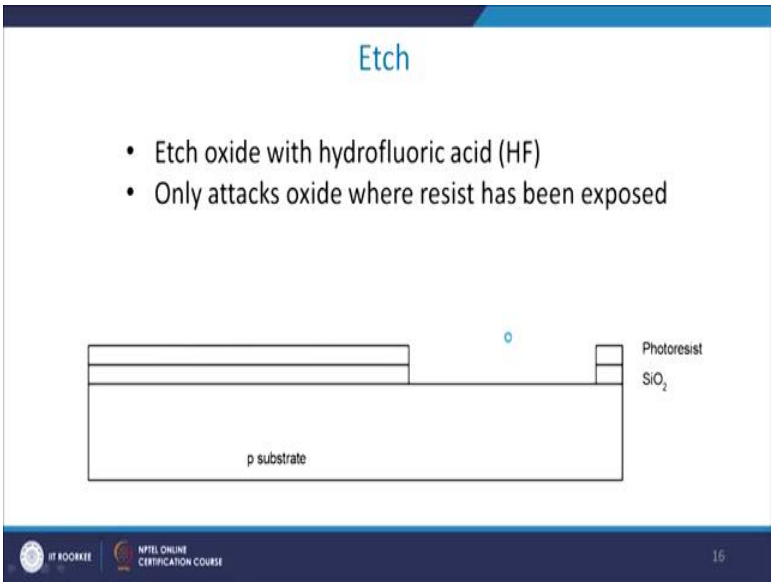
So, what we get from here is that that you get np to be equals to say you had 16 plus and suppose you add 16 and now you have 6, 16 plus 6 will be 22 per centimetre cube, right? Whereas if you have metal layer you will have already 20, so 20 plus large number 26 per centimetre cube. So you can see same technology and you get a larger doping profile as compared to the previous case.

(Refer Slide Time: 25:02)




So, let me come to lithography, you need to expose the photoresist through an n-well mask, right? And then a sign of strip off the photoresist. Once you do that you are stripping off your all other things. So, what you do is that this is basically you're the photoresist that has fallen on to this bank this bank. And as a result, you strip off the photoresist from this directly, right?

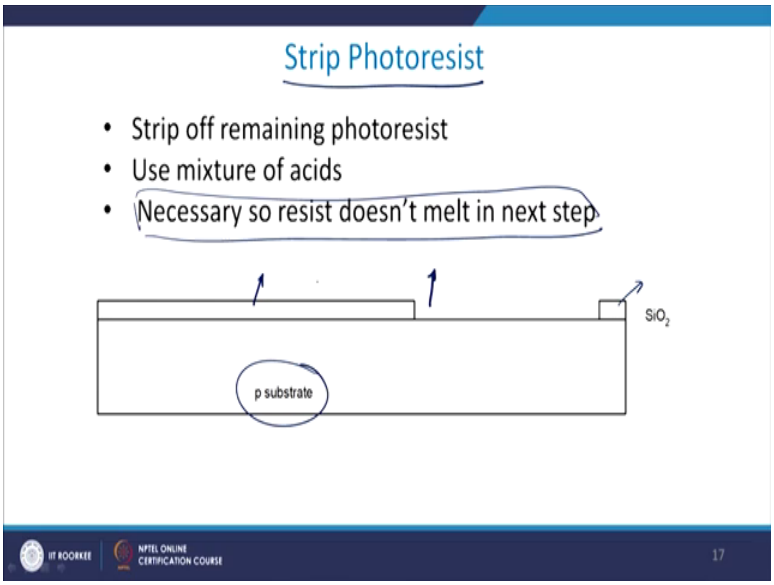
(Refer Slide Time: 25:23)




What is etching? Etching means that, You etch the oxide with Hf. So, I have an oxide formation means I want to remove it for whatever reasons. Then what I do is I put Hf or hydrofluoric acid

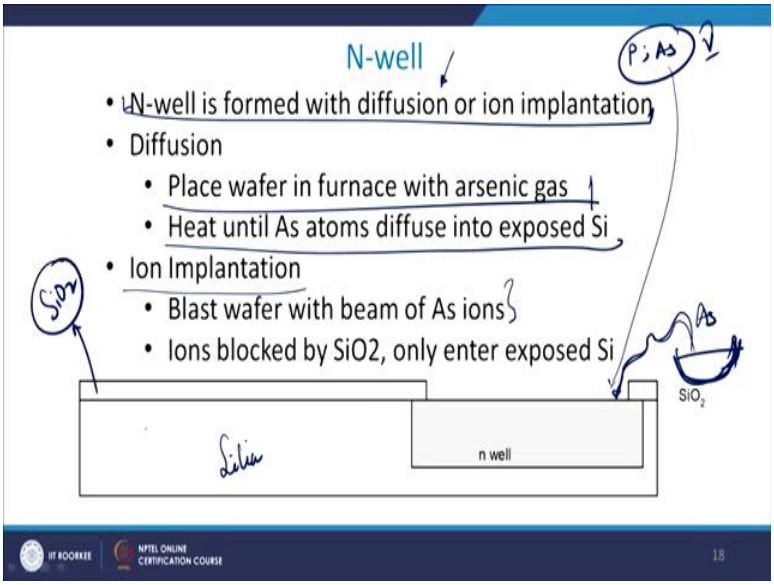
and that strikes off that metal layer and there is and makes it almost non-existent. HF actually attacks the oxide layer where there is no, where the resist has been exposed, right? So, wherever the resist has been exposed this HF will try to attack those exposed parts in order to reduce the value of your or to reduce the attacks, right, parasitic attacks or others.

(Refer Slide Time: 26:26)




Let me come to the strip photoresist. As we are discussing we strip off the excess photoresist, we generally use in this cases. So, we have a strip photoresist which is basically what you do is remove the photoresist from here and the remaining part is what is there, right? So, the reason we strip off is that we do not want the other part to melt away, right? So, in the next step when we do we do not have to melt the photoresist. So, we strip off the remaining photoresists. This is what we have stripped off. So, I have a silicon dioxide layer here and I have a p type substrate over which I am doing the photoresist, right?

(Refer Slide Time: 26:37)




We come to the n-well formation. Now, n-well basically meaning that you are doping with either n type dopant concentration which is basically either phosphorus or arsenic, right. This is group V element which you are doping with and what you do is that this n well is formed with diffusion or ion implantation, right? So, typically in most of the organisation or fabrication techniques we use diffusion, right?

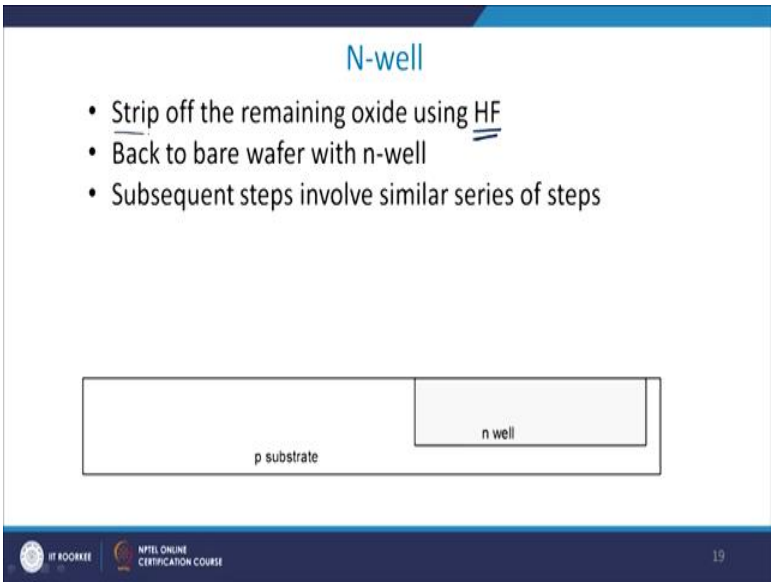
What we try to do is that the phosphorus and arsenic are kept in a chamber at elevated temperatures and then and this material is kept at a lower temperature respectively. So, there is a temperature gradient and there is also a dopant gradient which results in the atoms of these phosphorus and arsenic coming towards and sitting at these positions, right? And you do have do have diffusion.

You can control the speed of the diffusion by changing the temperature also by changing the dopant species. You can also change the amount of dopant species by increasing or decreasing the dose in the ion implantation machine. So, the diffusion if you look we use arsenic gas in a furnace, right? What we do is that arsenic gas is there and we heat it as arsenic atoms diffuse into exposed silicon. So, the arsenic atoms will diffuse and sit, these arsenic atoms will from maybe a this thing a glass and from there the arsenic atom will be coming and sitting on the silicon wafer which is this.

The silicon wafer which you see, right? And silicon dioxide, right? In ion implantation we try to do blast with arsenic atoms. So, arsenic ions are there since, you are accelerating you require the particle to be charged and therefore, you have to first ionize your arsenic or phosphorus atom. So, when you ionize your phosphorus or arsenic atom, then you through an electric and magnetic field you try to push it and a large dose of that charged particle falls onto the material.

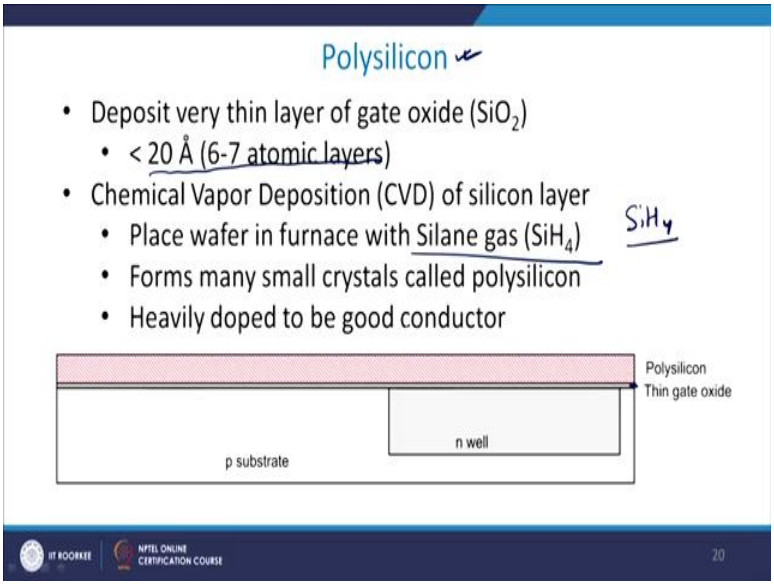
As I discussed with you earlier, I can change the dose as well as the doping concentration directly. Now, so now you can understand why this silicon dioxide is so important. The silicon dioxide is important because then this silicon dioxide does not allow for these arsenic or phosphorus atoms to enter in these regions. So, this region is devoid of those group V materials whereas this region is having large amount of group V materials and therefore, it is basically your n type well is formed, right?

(Refer Slide Time: 29:14)




Now, we come to the n-well. We are using HF or hydrofluoric acid. We strip off the remaining oxide, right? So, whatever oxide layer was here at this point and at this point we just use HF to remove it. Then therefore, now what we have is the bare silicon wafer which is this one and we also have n-well which is kept here, right? And we will be using similar steps later on.

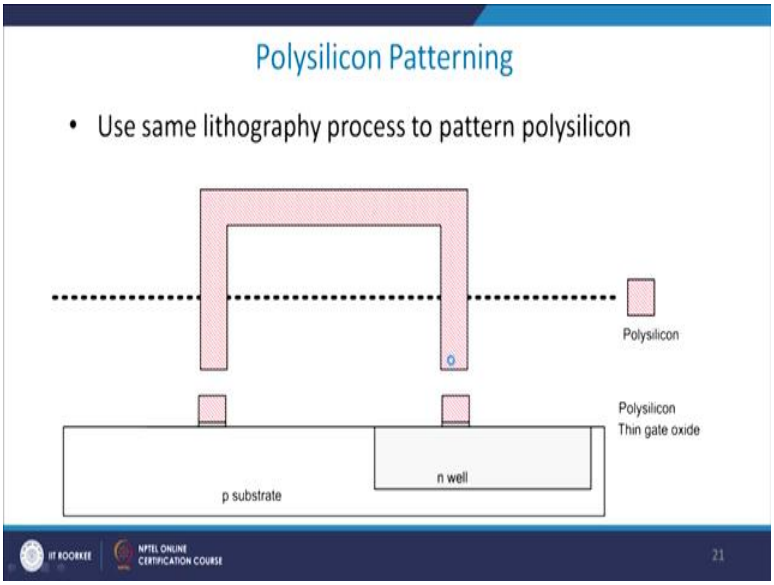
(Refer Slide Time: 29:55)




After this we grow a layer of polysilicon. This layer is polysilicon, right? We deposit a very thin layer of gate oxide. So, this is basically a thin layer of gate oxide which you see here. This is a thin layer of gate oxide, right? There is approximately 20 $A^0$ which means approximately 6 to 7 atomic thickness available to me. Then you have then after this after we have grown the gate oxide which is basically native growth you allow for CVD and you use a silane gas $SiH_4$ to do that, right? Silane gas will be forming and form small crystals and they form polysilicon.

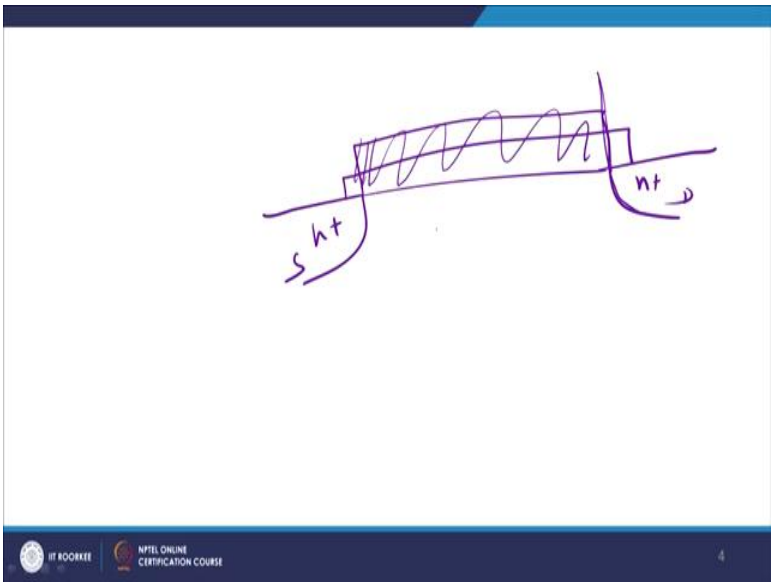
So, polysilicon is basically sort of a silicon arranged in a special manner. They are crystal in small domain but in larger domain they are still amorphous in nature. So, they are heavily doped. You can heavily dope it to be a good conductor, right? So, polysilicon is used in that manner. So, now you have polysilicon, you have a thin gate oxide then you have n well and then you have p type substrate.

(Refer Slide Time: 30:38)



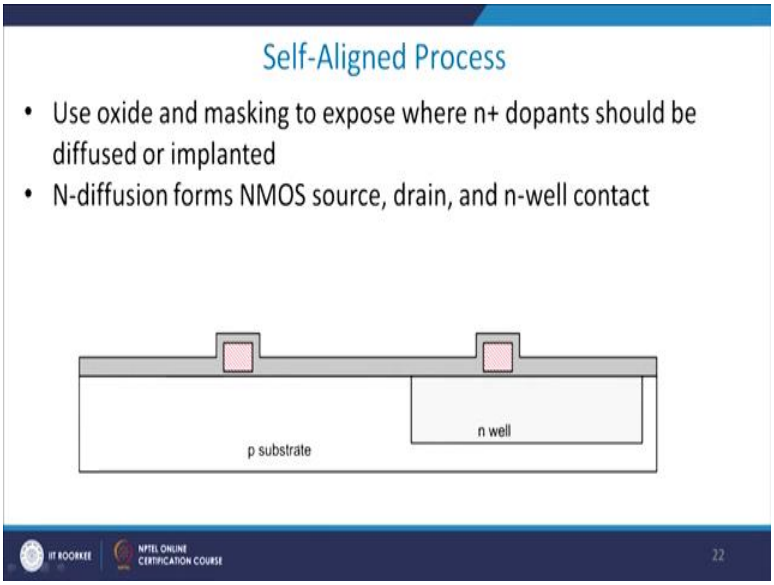

Now, you use the same lithographic process to pattern the polysilicon. So, what you do? You use a polysilicon you have a n well here, and then you pattern it say, for example, you want to grow here and grow here; the same procedure which I discussed earlier you can actually etch and remove the oxide layer and the polysilicon layer at these places, where you have marked it which is discussed in our previous lecture. An important point which is current technology is that you require self-align processes. Self-align primarily means that if you have a.

(Refer Slide Time: 31:20)



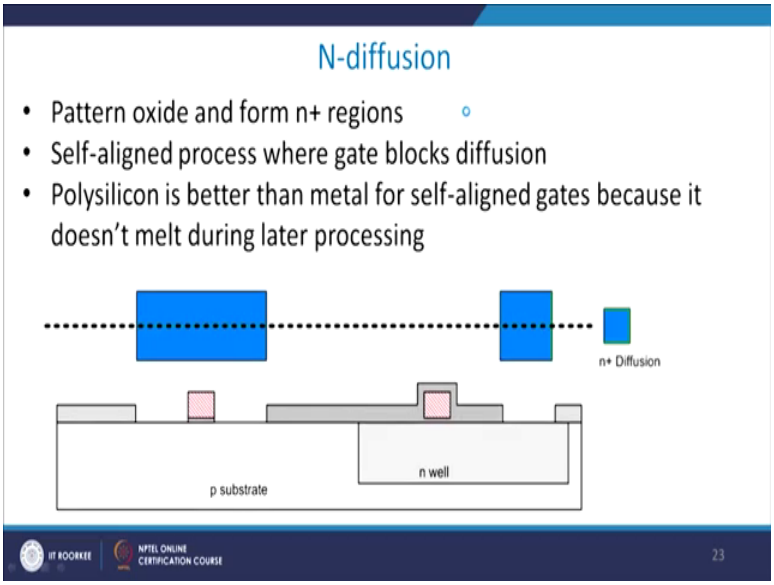
Let us suppose you have a device you have a device which is something like that this is your n and this is your p. $N^+$ $N^+$ and this is your source and drain, right? If this is overlap so you would be overlapping with it. Self-align primarily means that you should have your oxide layer like this. This is the oxide layer. This is the oxide layer and this self-align means, when the alignment of the oxide and the metal is exactly that way as that of the source and drain material source and drain material.

(Refer Slide Time: 31:43)




And that is how you do it here. So, what do you do is that you grow it like this and then use oxide and masking to expose $N^+$ dopants should be diffused or implant means now, what we do is we expose the polysilica and we try to those places where we want the $N^+$ dopant to go we expose it and for other reasons we stop it. Now, n diffusion forms NMOS of course, and if you want to have p type device you will have p type diffusion and therefore, you will have n-well contact, you will have source and drain available with you.

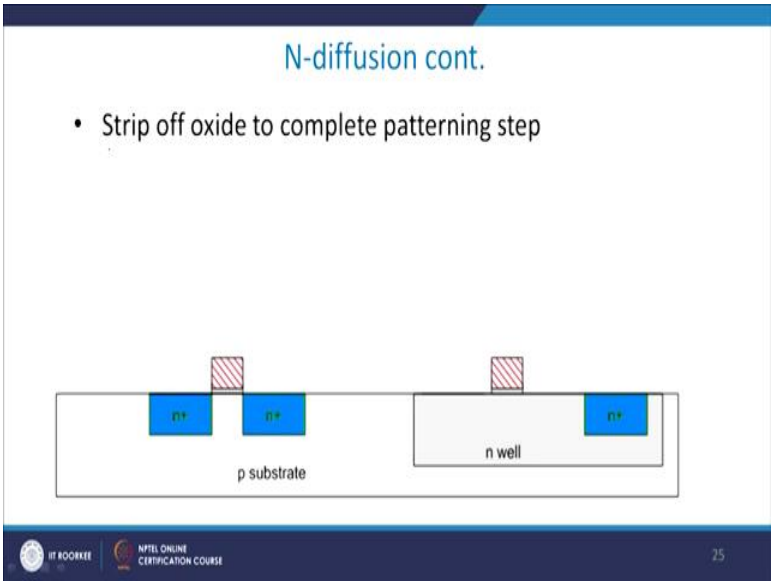
(Refer Slide Time: 32:14)




We also use the n diffusion as I discussed with you these are actually self-aligned processes where gate blocks the diffusion. So, gate will be typically polysilicon. You try to put your diffusion from outside and the gate will block it from going inside just below the gate. So, polysilicon is better than metal for self-aligned gates because it does not melt during later processes, right. So, its melting point is relatively high for polysilicon as compared to even metal, right. And that is the reason you make it. So this is the diffusion, right? So, this will be $N^+$ diffusion here and there will be $N^+$ diffusion here as well as at here.

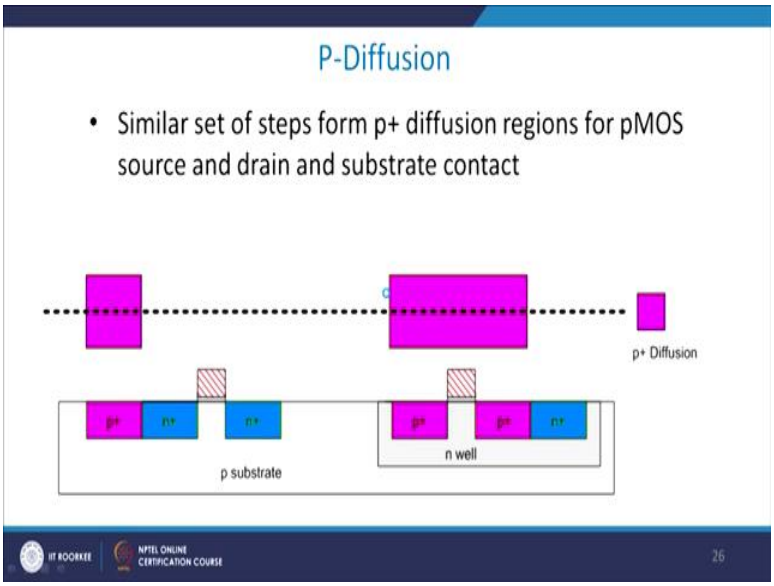
So, this gate will not allow this diffusion to go inside this material, right? Okay, So, after you have actually diffused into the n type material now you have $N^+$ region here you have $N^+$ region here and you have $N^+$ region already from, right? Now, usually nowadays initially we used to have diffusion but now we used ion implantation to do it but still historically it is done by diffusion. So, I have an n-well here and I have diffusion. So, this is a diffusion, this is a diffusion and this is a diffusion generated region. Now, what you have done is let me see what we have done. Therefore, now, what you can do is you can strip off this layer of polysilicon, right?

(Refer Slide Time: 33:35)



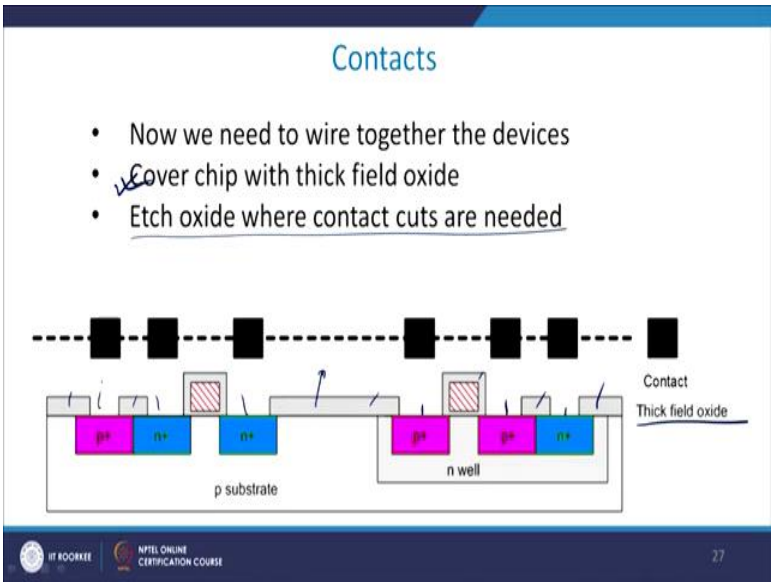

So, next slide as you see I have removed this polysilicon by doing oxide layer. You remove the oxide layer and therefore, what we have done is all the oxide layer here has been removed. If you go back to the previous slide this was the oxide layer which was there. Now, by doing HF etching or using some other technique I can remove this oxide layer directly and I get almost devoid of any oxide layer in this case, right?

(Refer Slide Time: 33:58)



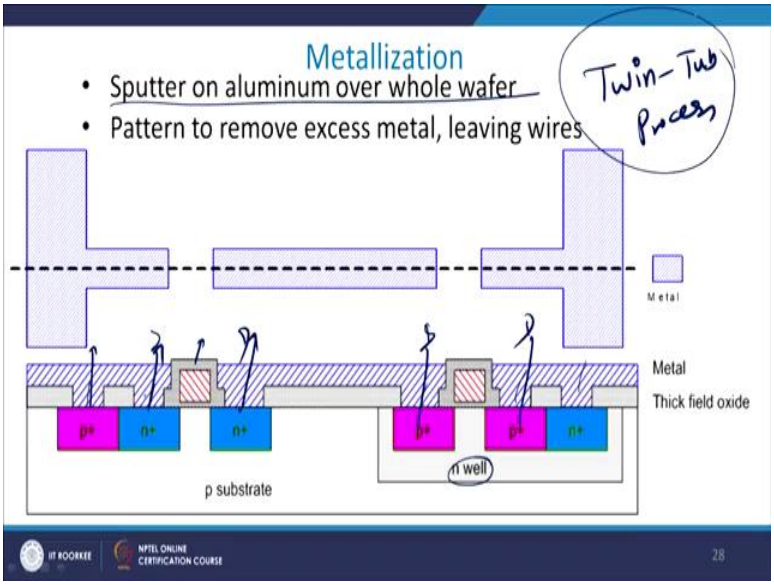
If you want to do p type dopant and n type dopant is finished, I have an NMOSFET formation here but I also hold a p formation over the n-well. So, what I do is that, we exactly as we started with diffusion with n type MOSFET we do diffusion using a p type material which is group III boron material. And we dope it with source and drain and substrate contact. So, what we do source, this is the source, this is the drain right, and the substrate contact which is this one. All these 3 can be done using the p well diffusion process.

(Refer Slide Time: 34:29)




After we have done the p well we need to have the final layer which is the metal layer and to do that you require the grower contact. So, what we do, we cover everything with a field oxide. Field oxide is a very thick oxide. So, this is the field oxide which you see, thick field oxide. This are very thick field oxide here, right here, here,,, When you grow it you do not allow any external species to enter anywhere else except the region which you have opened. Now, what do you do? You cover it with thick field oxide and then etch oxide where contact cuts are needed. So, I etch at this point at this point at this this this and this, right? And then you grow the contacts here.

(Refer Slide Time: 35:10)




Once, you have grown the contacts which is this one; we have grown the contacts, we then sputter after we have grown the thick oxide layer. Now, my now my region where you want to grow the oxide is opened. Then what I do? I sputter with aluminium over the whole wafer. So, I do a sputtering using a sputtering technique; I sputter over a whole aluminium. The aluminium actually sits at those places which was opened, right?

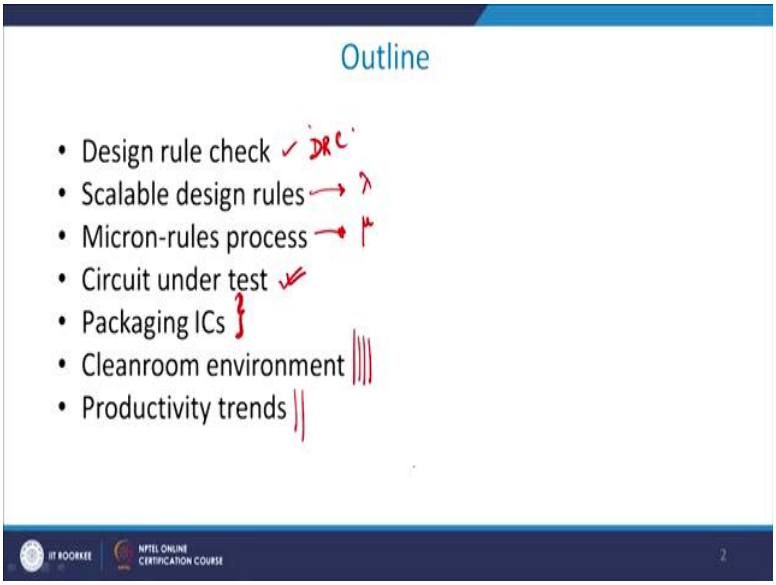
These were open spaces and the metal fills in all these open spaces, right? And as a result what happens is that sorry, as a result what happens in this case is that now you do have now what you can do is now you can now you can remove this etched oxide here, right. And you will have a contact here, a contact here and a contact here. So, these two will be source drain contact for NMOSFET. This will be source drain contact for source drain contact for PMOSFET.

This is also known this whole procedure is also known as twin tub process, right? Because you grow a p substrate, you grow an n-well within that and then this is known as a twin tub process. With this I have given you a brief insight into the flow of fabrication in a silicon and we will take up further issues in this as we move to the next slide. Thank you very much!

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−24**
**Fabrication Process - II**

Hello everybody and welcome to the NPTEL online course on Microelectronic: Devices to Circuits and we take up fabrication processes part 2. So this is basically VLSI technology part which we are dealing with, so there will be two modules which we will be devoting to fabrication. Actually it requires a whole set of lectures to go in details of it but at least for a design course you should know the basic idea of fabrication and hence this lecture and the previous one.

(Refer Slide Time: 1:07)

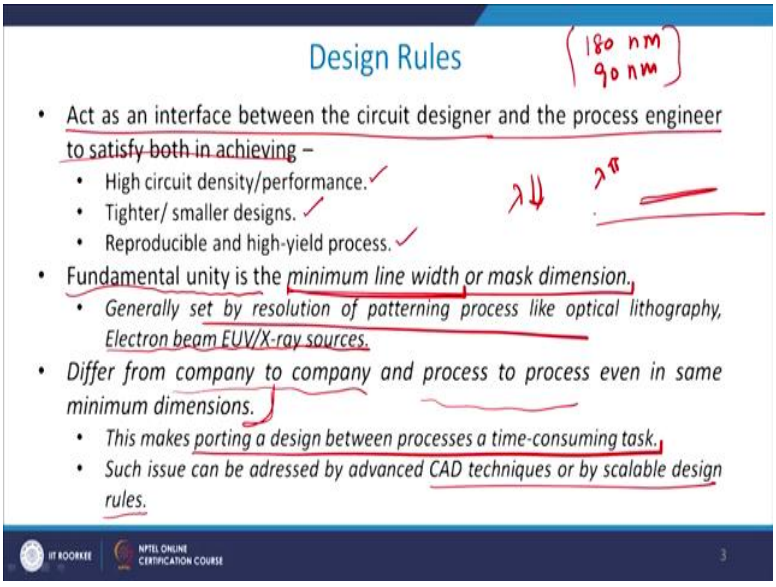



What we will be doing in this lecture series is the outline of the talk will be, so I will give you certain design rule checks, also known as DRC. So means that once your design has been done on a simulation platform, your layout has been done, you need to check whether this,, for example, whether the width of the contact which you have drawn is as per the requirement of the fabrication facility, right? So it might be true that the fabrication house can only make a contact which is say in a very layman's language 1 millimetre by 1 millimetre, and you have made a 2 millimetre by 2 millimetre so it will be rejected once your design actually goes to the Fab house. So the work is that you need to therefore do a checking of your design and therefore this is known as design rule check, so whatever design you have done on silicon using simulation platform tools, you need to do a thorough checking of that.

There are two types of rules such as Lambda-based design rules which is called Scalable design rules as you can see, right and we have a Micron-based design rules and this is the second design rule which we see, so this is basically **λ**, this is basically Micron right. And have been device from quite a long time but as the name suggests, Scalable design rules are design rules which are scalable in the sense that if you have a design which has higher dimensions right, say it is 4**λ** then you can scale it down to 2**λ** and everything will be scaled down by half in a sense so that is known as the scalable design rule. Whereas what is the micron-based design rule? Micron base design rule primarily means that the dimensions of the structure are actually fixed and given to you beforehand so you need to fix up those processes.

We will look into the fourth one which is basically circuit under test. We will look at some issues of packaging which itself is again a very big topic but I will be dealing in a detailed manner later on but at this stage we will touch and go for the packaging. We will have a look at typical cleanroom environment which is with us, and then finally we will look finally at productivity trends which is the current VLSI technology of fabrication gives me in a real sense so what is the design rule first of all let us see that.

(Refer Slide Time: 3:22)

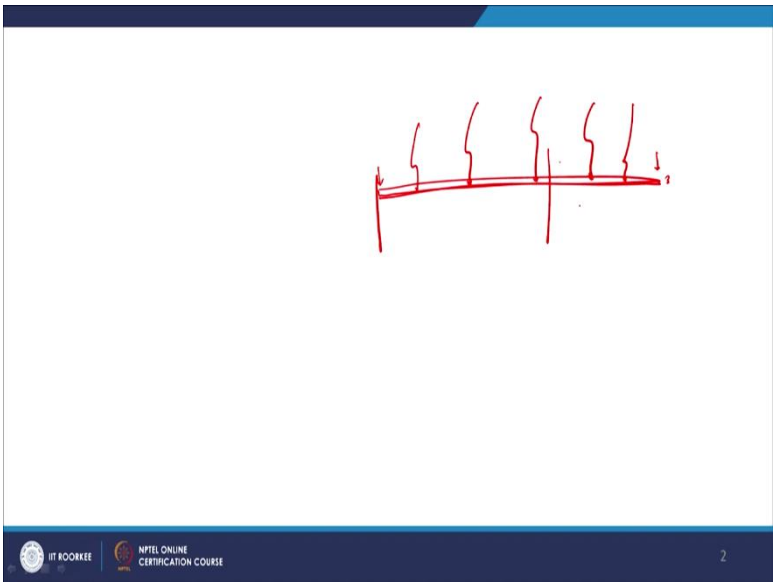



This acts as an interface between the circuit designer and the process engineer to satisfy both in achieving high density or performances, smaller design or smaller area and high yield processes. So as I was saying in the first slide that basically design rule checks are such that in is a common set of rules which a process engineer and circuit designer both have to follow,

so this is sort of a talking point between the circuit designer and the process engineer, right. And therefore if a circuit designer follows these rules very thoroughly then chances are that his design will be accepted by the process engineer and he will be able to send it for proper fabrication. If not, then he has to again iterate back to the circuit designer to change his design rules.

Why they are required? They are required because we require a very high performance which basically means that for example, I will give you, for example, if your contact is very large in dimension, you eat up into area but again if it is very small you then have larger resistance offered by the contact, so you need to optimize the contact area. And the process people have already optimised the contact area for you, say the metal contact should be same 0.5 micron by 0.5 micron, so they will let you know so then you design that much exactly onto your profile and you get the best results. Similarly, since we already optimise designs so the area of silicon utilised will be typically very small so there will be much tighter design that is the second point and the third point is since the circuit has been based on perfect design rules and they are satisfying the perfect design rules, the yield will be also very high because the production will be also very high, there will be less chances that the fab will fail in delivering the circuit.

(Refer Slide Time: 5:35)




Now the fundamental limit or the fundamental unity is the minimum line width or mask dimension, so this is basically very important. See, as I discussed with you in the previous slide that if you have for example mask and the mask has got something like this opening then you can ensure that means you can understand that this much amount of mask with this

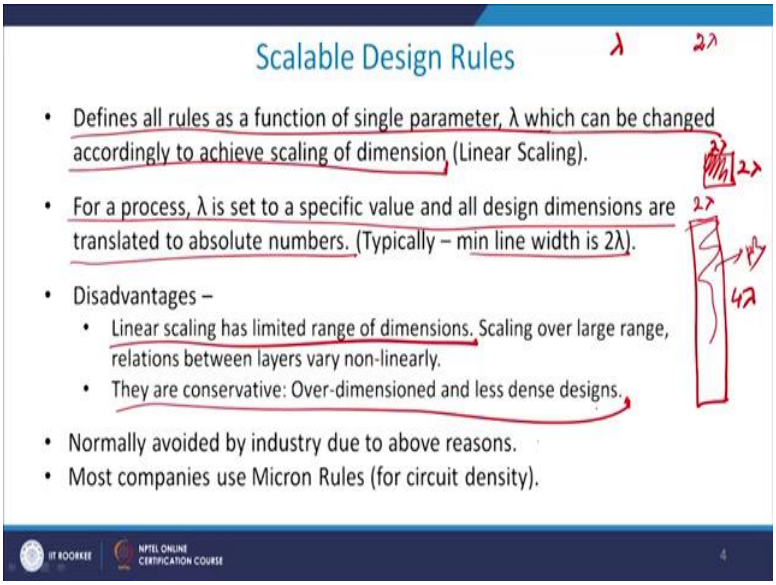
much amount of opening will be actually responsible for all the photolithographic raised to go through it and as a result you will have a dashing table typically available with you. So the line width which you see here, minimum line width which can be drawn and you can draw it safely with best results is basically the fundamental unit also referred to as minimum line width, so minimum line width is available dimensions. Who sets it? Of course it is set by the resolution of the patterning process.

Say for example, if you are using optical lithography where you are shining using an optical light typically an optical light you are doing it, then the **λ** values are typically large right. So if you are using **λ** large, you very well know that you will have heavy scattering right and therefore, if you wanted to grow let us suppose this much amount of Channel length, you might end up growing this much amount of channel length, right. So what you do, to make the design tighter you lower the **λ** value, and to lower the **λ** value you ensure that **λ** value is lower you have less scattering available to you, you go to E beam as well as X-ray sources.

Once you do that then you are not only able to reduce the dimensions which is an obvious advantage, but you will be also able to make the design much more stable because for example, when you have optical lithography there is chances of diffraction at the edges of the mask will be very high, so you wanted to grow 1 micron you end up growing 1.5 micron. Whereas, when you are using silicon or when you are using X-ray or direct E-beam that is very-very small, the diffraction are very-very small and therefore the exact length will be replicated on silicon, right. So therefore, it depends upon the type of source which you are using to a larger extent. Now this also differs from company to company and process to process, and with the same dimensions the process will change and it differs from company to company what type of process flow the company takes.

Now therefore that is what is important, porting a design between processes is a time-consuming task, which means that let us suppose today I have a process flow at 180 nanometre right, I want to port it at 90 nanometre right this is my job. This porting is quite difficult because it is not that you just reduce the dimensions by half and everything and everything will be fine, no it is quite complicated and therefore it is very difficult to port it from either lower to higher or higher to lower. So therefore there are certain CAD tools or CAD techniques which scale the design so we use scalable design rules, so I can either use CAD tool to do that or I use scalable design rule to take care of that this thing. Let me therefore come to what is scalable design rule.

(Refer Slide Time: 8:45)



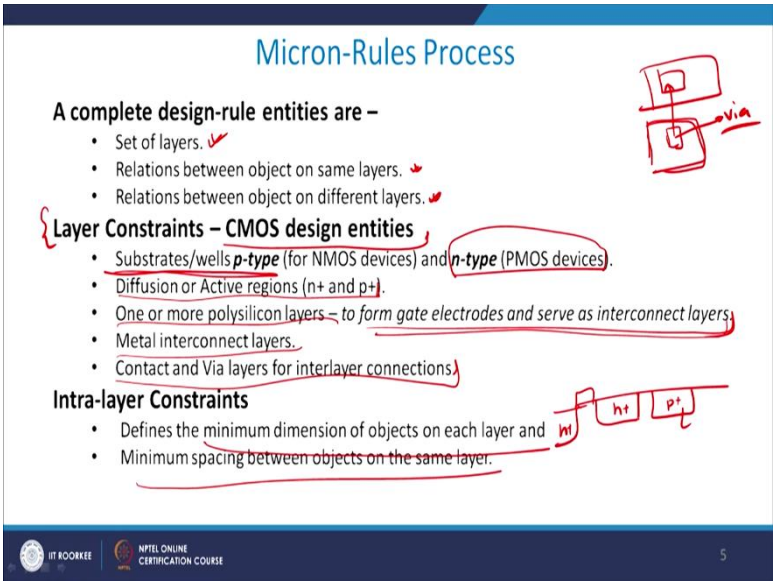

Now what is happening is that the scalable design rule is that defines all the rules as a function of single parameter **λ** and which can be changed according to achieving the scaling of the dimensions, so linear scaling which is here. Secondly for a process, **λ** is set to a specific value and all the design dimensions are translated to absolute numbers typically minimum line width is 2**λ**. So let us suppose I have a **λ** right process width **λ**, so we define minimum line length to be equal to 2**λ** there are certain reasons for that. Therefore, let us suppose I have a contact whose dimensions are 2**λ** by 2**λ**, so there is a minimum contact which you can get.

Similarly, you have a poly silicon which is basically say 2**λ** by 4 **λ**, this is poly this is poly let us suppose right so it will be always dimension of so it may be 2 **λ**, 3 **λ**, 4 **λ**, 5 **λ**, 6 **λ**, so that is the reason it is scalable. So once your **λ** is defined you just have to put the integer value to change it from the high to low value, so on and so forth. Now and therefore the advantage is of course that it is scalable and therefore you can use it but disadvantage is that it is not scalable over a large range of design or over a long large range right because you can see here the linear scaling has limited range in dimensions right and they are conservative in nature and therefore they are dimensioned and less dense design is available to you, which means that they tend to give you though they are scalable but has a limited range in dimensions and they also set a nonlinearity at higher values and therefore they are less dense design with us.

Now generally avoided by industry to a larger extent because of the above problems as most of the industries use micron rules for their purposes and they use micron quite conveniently in this case. Micron rule is a set of layers as I discussed with you earlier, so **λ** based design

rules are based on the fact that you have minimum length of **λ** and generally typical line widths of 2 **λ** right. And if you take 2 **λ** as a dimension and then you take either even or odd multiples of 2 **λ** and you automatically start getting the scale designs so you can have 2 **λ**, 3 **λ**, 4 **λ**, 5 **λ** as a scale design and so on and so forth.

(Refer Slide Time: 11:13)




In the micron-based design rule or the micron rules, a complete design rule entities are there are set of layers right, and relationship between object on the same layer and on different layers. So what we do is, in this case we have layers for example, if you have an interconnect layer which is a metal layer, then you have poly silicon layer then you have field oxide layer then you have silicon layer, so on and so forth. At each layer we start defining the distance between 2 active points or 2 passive points right, we can also define the distance between 2 layers across the particular board, right.

So what are the constraints the layer constraints? The layer constraints are you have CMOS, let us look at CMOS design styles. So substrates are p well or p-type right and is n-type, for PMOS device of course if you remember substrate has to be n-type and for NMOS device the substrate should be p-type right. Second layer is diffusion or active layer $N^+$ or $P^+$ by which you make actually the source and drain regions. One or more poly silicon layers to form the gate electrodes and serve as interconnect layer, and then this is for gate electrode whereas, this is for metal interconnect layer.

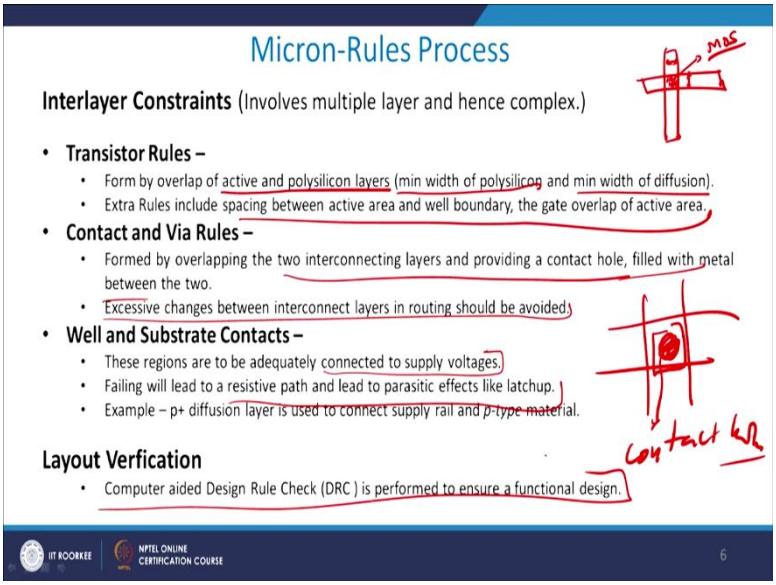
So interconnect basically means that so there are 2 layers of metal, one is the metal layer itself which connects the contact with the external world and you have an interconnect layer

which contacts the external world with the peripheral devices. Now contact and via layers for interlayer connections, so I have a contact like this right and there is a via available here which starts to make contact with the upper layer right, makes a contact with the upper layer so this is known as via. So via is a sort of a through silicon metallisation right and that is known as via right.

Now so you already have an intra-layer constraint, now within the inter-layer constraints are, defines the minimum dimensions of objects on each layer, and the minimum spacing between the objects on the same layer. So what does it tell me is two things micron is what should be the minimum dimension for example, an $N^+$ diffusion layer or a $P^+$ diffusion layer, right that tells me that, and then secondly it also tells me what should be the minimum distance between $N^+$ layer and $P^+$ layer.

Suppose you are able to get pictures for example, you are designing a system which is something like this and you have an $N^+$ layer and you have a $P^+$ layer then what does it tell me is this is substrate and this is source drain and this is source and then you have something like this, so what do they tell me is what should be the minimum dimensions of this $N^+$ layer and what should be the minimum distance between these 2 layers which is this one right. If it is closer than this, the design will fail so two things gives you the intra-layer constraints in a certain sense. Now to continue with the micron process, for inter-layer constraints are obviously more complicated because between the two layers you need to talk.

(Refer Slide Time: 14:22)




Now what are the transistor rules? The first rule is that when an active layer and a poly silicon layer cross each other we get a transistor that is a very simple basic rule number 1 right. And therefore the minimum width of poly silicon and the minimum width of diffusion is primarily, so if you have something like this as a poly and you have something like this as a diffusion then typically this is your active MOS device right transistor. What are the extra rules? Include the spacing between active area, well boundary and gate overlap of active area, so you need to set out what is this, what is this distance, how much should be this distance, so on and so forth.

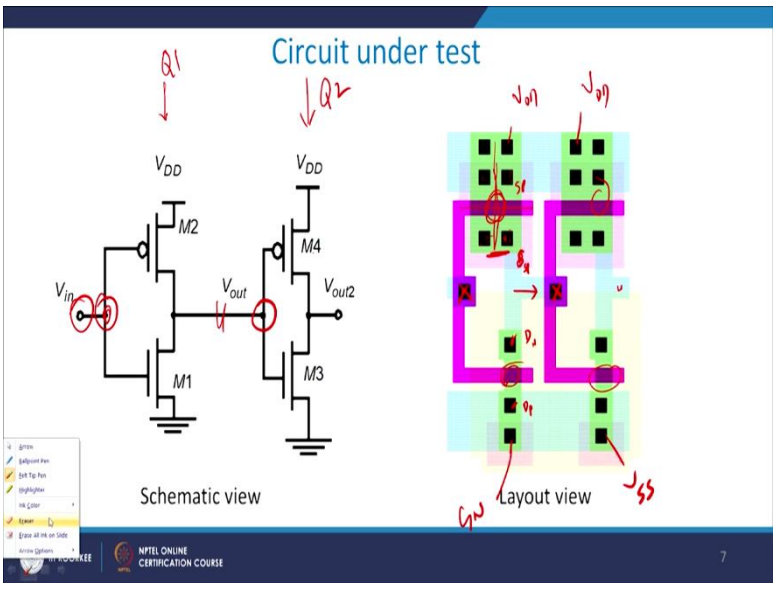
Now contact and via rules, so we have understood transistor rules, transistor rules are those in which the active and poly silicon layer cross each other and dimensions are set by the crossing principle. Contact and via rules, well they are formed by overlapping of two interconnecting layers and provide a contact hole filled with metal between the two. So let us I have one layer here and another layer here, we have this empty here, here I fill up with metal right and therefore this we refer to as a contact hole, right. Excessive changes between interconnect layers in routing should be avoided, I think this is very interesting. So you might have, so I was discussing with you only two metal layers right, but you might have multiple metal layers looking into the fact that you might have multiple interconnect layers right. So you have 1 interconnect layer 2, 3, 4, the last one becomes your $V_{DD}$ layer which is the final power supply layer.

Typically while designing we should not transverse between layers too much, within the layer we make a design and then we move to the next layer and we keep these micron-based design

rules effectively okay with this. Now “Well and substrate contacts”, they should be adequately contact connected to the supply voltage as we have discussed already. Failing this will lead to resistive path and lead to parasitic effects such as latch up right. This is pretty important, you seldom found problem areas specially the micron-based design rules is that, if you do not connect the substrate and your wells, substrates and your n-well, p-well procedures either to the supply or to the ground and you keep it floating, chances are that there is a problem of latch up.

Latch-up basically means that maybe it latches to a particular state and stays there permanently which you do not want it to happen right, so that is one of the pretty drastic problem which happens due to latch up, so that is what I wanted to say for the well and substrate layer right. Generally your CAD tool for example is performing a functional design check using DRC. And you can use one of the standard Cadence for example is one of them which uses the typical layer , there is a DRC for you. If there is a problem, it gives you the problem at which point you have a problem in layout and then you can correct those again on the DRC. When this DRC clean then you freeze your design and send it to the fabricator.

(Refer Slide Time: 17:30)

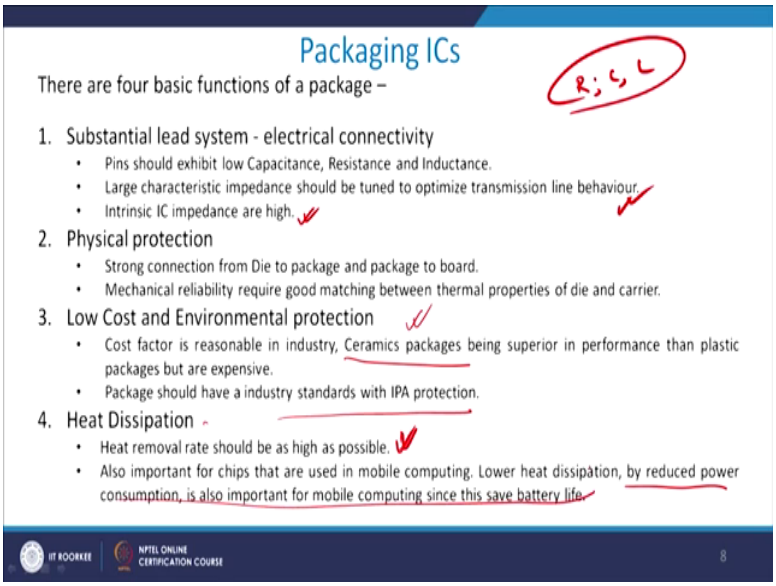



So look at the circuit under test, as you can see we have got two CMS inverters driving each other, so this first one is let us suppose Q1, next one is basically Q2. So Q1 is driving Q2 and therefore as you can see here very interestingly that you have a contact here right so this contact is this one right. And when the poly and the metal layer, this metal layer and the poly layer cuts, this is basically your MOS device similarly this is your MOS device here, similarly this is your MOS device which is M4 and this is your MOS device M3. Since this is

PMOS you will have metal layer twice the width because you do not want it to be skewed, you want the $t_{PHL}$ to be $t_{PLH}$ and therefore the width is almost double that of NMOS here that is we can see.

So this is your source drain and this is your $V_{DD}$ and ground, so this is your $V_{DD}$, $V_{DD}$ right and this is your GND and this is your $V_{SS}$ or GND, and this is your this contact is this contact right and this contact is this contact or this contact in fact, and this is drain and source. So you have a drain here so this is drain of NMOS, this is drain of PMOS, this is again the source of PMOS and this is source of PMOS and these 2 are shorted and I get this, so these 2 are interconnect, similarly this is interconnect which is visible to you. So this is a typical circuit under test, typical layout which you will see here as you can see here right and these are the metal layers which you see, so metal layers are responsible for carrying the $V_{DD}$ ok.

(Refer Slide Time: 19:20)




Let us look at the packaging of IC, well there are 4 basic functions of the package. Generally, you should have pins which will be IO pins means input output pins, they should exhibit low resistances, capacitances and inductances so that the delays are minimise across the pin. It should also have large characteristics impedance to reduce the noise at those particular pins which is the second part. And intrinsic IC impedance should be high which we have already discussed in the second part.

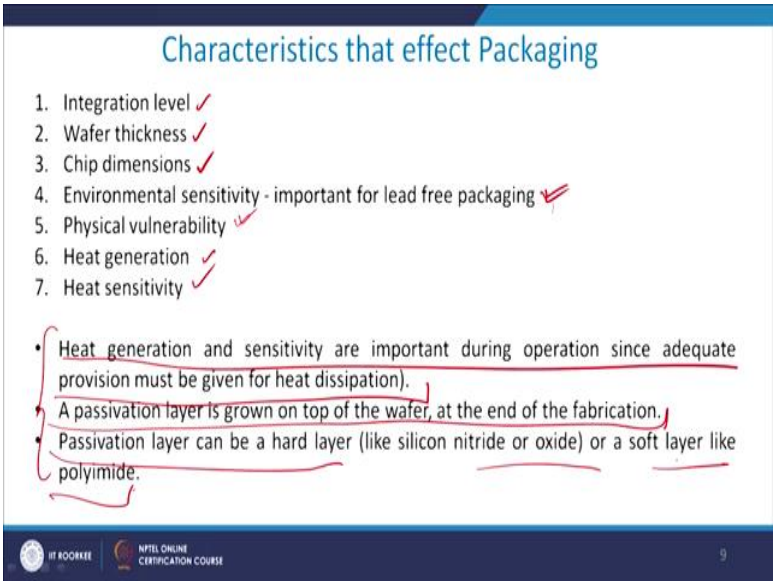
The second point is the physical protection, physical protection means strong connection between dye and package, and package and the board. So if you have those dye, even if you remove it physically you can remove it easily that should not be the case and it should be

properly bound with the board. Now mechanical reliability require good matching between thermal properties of dye and the carrier right so that is very important that mechanical reliability is important for example, if there is shaking of the table or something like that, your dye should not be extracted right from the board itself.

It should be low-cost, the cost factor is pretty from industry point of view, ceramic packages are most superior in performance then plastic packages but then they are very expensive in nature. So whenever you go to the Fab and you tell them I want the ceramic package they will provide you a set of ceramic packages, if that matches with IO of your design input output map of your design, you can take that ceramic package and do it. So typically what people do over the years, if you want a package design you first look at the package and then you because if you want to do a package design then you first of all look at the package, look at the number of pins and IO pins and then decide how to change your design so that it fits to the package right.

At the last stage you cannot do it so therefore it is always advisable to look at the package first and have a look into it. And it should have industry standards as I discussed with you right with IPA protection. Typically, you should be protected from surges electrical surges and so on and so forth. What is heat dissipation? Heat removal should be as good as possible so you should have good heat sinks, but you can do it by reducing the power consumption and generally we use mobile computing for saving the power nowadays right and most of the time the packages help you to reduce the power drastically right.

(Refer Slide Time: 21:39)

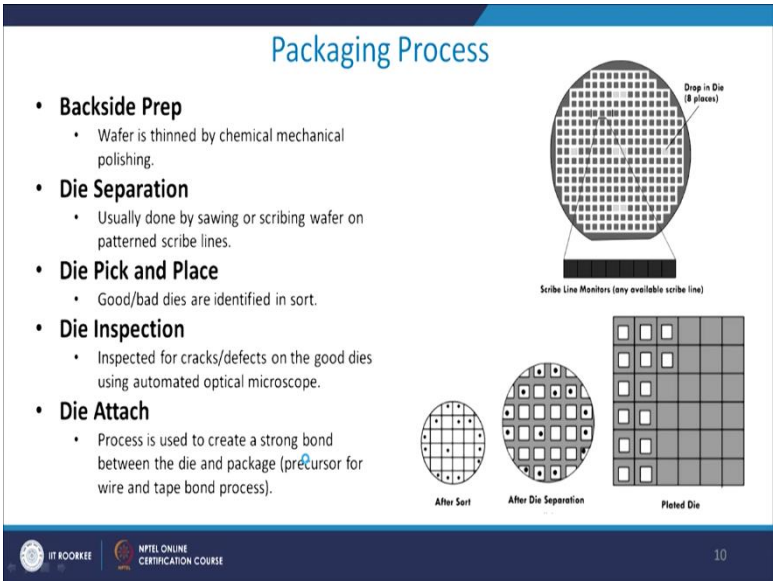



What are the characteristic that affect packaging? Well of course, what type of packaging you will be using it depends upon the type of integration level which you are using. Similarly, how much wafer thickness are you using, what are the dimensions of the chip which you are using right, and then you look into an environmental sensitivity for example, people generally do not like obviously for obvious reasons arsenic, lead as part of the packaging technique and therefore it is not a very good idea to have these materials as part of the package.

And then it is physical vulnerability because how much it is prone to mechanical instability as such. Heat sink and heat sensitivity are two important points, heat generation and heat sensitivity in the sense that even if the heat is generated for example, all mixed signal block generates very heavy heat and analog will generate very heavy heat but then we should ensure that it behaves as a very good heat sink and removes all the heat even before everything comes to the surface. Now therefore as I discussed with you that heat generation and sensitivity are important during operation since adequate provision must be provided for heat dissipation, right.

A passivation layer is grown on top of the wafer at the end of the fabrication, and we know the reason why because the passivation layer will be there, it will help you that the lower devices active devices are not in touch with the active environment and therefore we do not expect it to get back. Passivation layer can be a hard layer like silicon dioxide or a soft layer like polyimide right. And that is quite an interesting one which we needs to look into that they have a soft layer like polyimide or even a hard layer passivation can be done right, so this is the effects of packaging which you see.

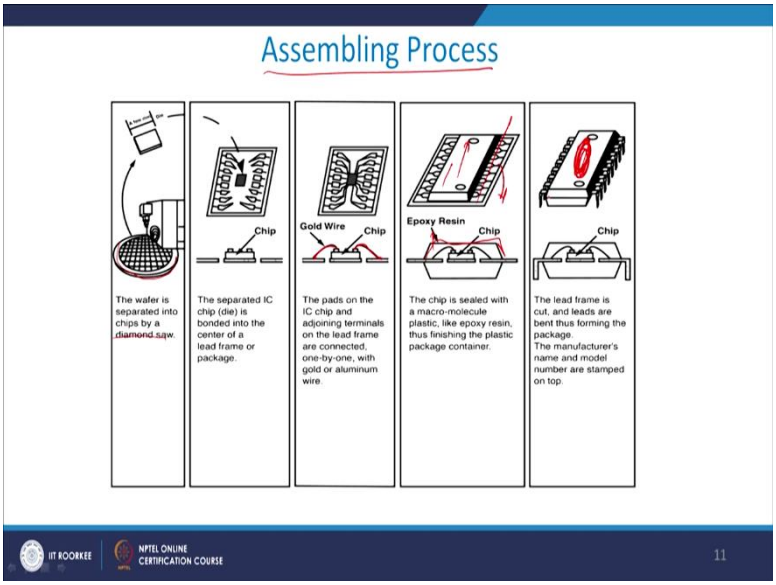
(Refer Slide Time: 23:23)




Packaging process if you look very carefully, we generally do wafer thinning by chemical CMP which is chemical or mechanical polishing. Generally these are known as dyes and these dyes are separated by diamond patterning, by diamond patterning you can actually remove these scribes, you can describe you can remove it, so these are actually known as Scribe line monitor as you can see this is the scribe line monitor. Over this you will have a dye on this side, you will have a dye on this side. Good dyes and bad dyes are sorted and they are identified and they are identified even when they are part of the silicon wafer itself.

You do dye inspection for physical cracks and defects, and generally we use automated optical microscope and that is enough to see whether a dye has cracked, or if it has not cracked then is there any physical discontinuity within the package itself, and if there are it is automatically removed from the system. You have to also have a good process to attach your dye with the package right, and therefore there are techniques of doing packaging and dying to a larger extent. And this is quite an interesting or important steps by which you can process these dyes ok.

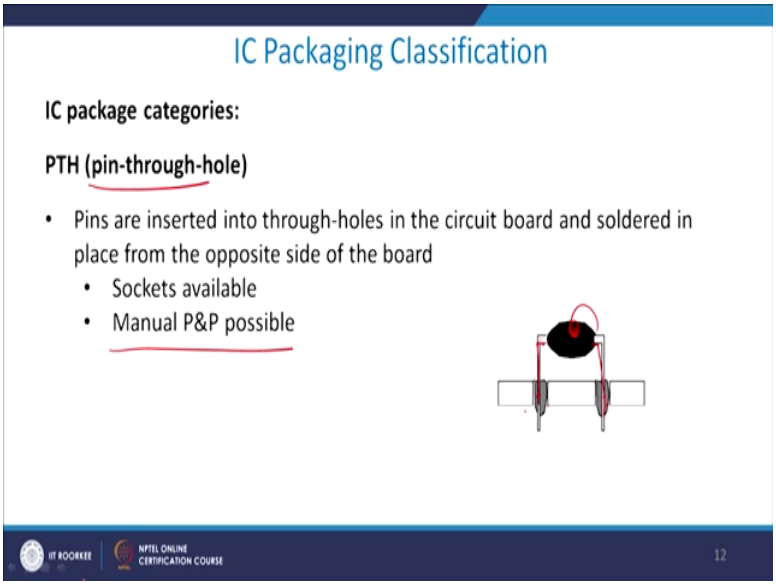
(Refer Slide Time: 24:47)




As I discussed with you, let me just give you a brief of assembling process. As I discussed with you the wafer, if you look at this is the wafer right. This wafer is separated using a diamond saw as I discussed, the separated IC chip which is this one is bonded into the centre of a lead frame or package. So I have a chip here and I place it here, these are your IO pins which you see in the pack, you do not able to see it but these are the IO pins. And these are the wires which connect the IO pins to the external world, IO pins and these are gold wires which connect to the external world. So the pads on the IC chip and the adjoining terminals are connected one by one with gold or aluminium wires, right.

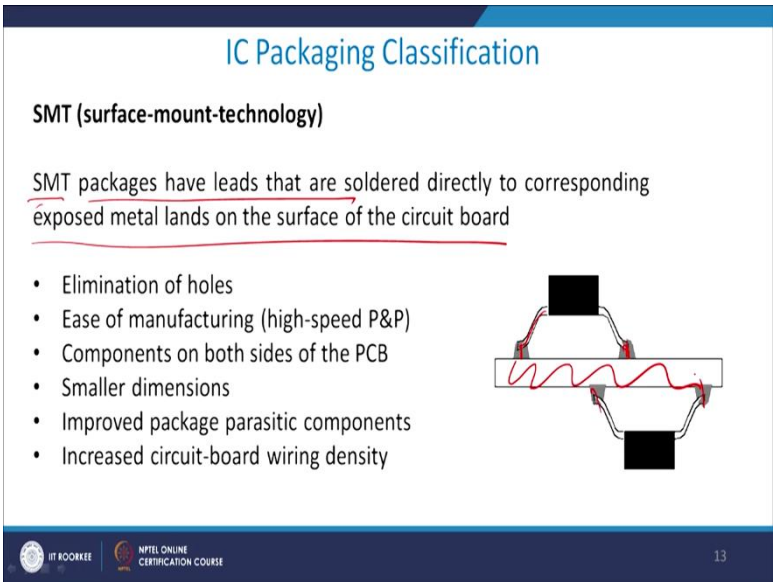
So after you have connected them, the chip is basically sealed, this is the proxy seal, is sealed with a micro-molecular plastic like epoxy resin and this is the plastic container which you can do it right. So this chip and this is the epoxy resin, so epoxy resin is over the whole chip so that it does not get disturbed right. The net frame is then cut and bent, so what you can do is you can cut it and then bend it around this direction and then we can put the name of the firm which has made this process and so on and so forth, what type of IC is this so on and so forth, its identification mark can be placed here in this case right.

(Refer Slide Time: 26:14)




IC packaging, well there are 2-3 important packaging techniques, the first is a pin-through-hole. Pins are inserted through holes as you can see here in the circuit board and soldered in the place in the opposite side of the board. So this is soldered from where? So this is through holes and then you soldered it here right, so manual PNP is possible manual pin through holes can be possible and you can actually solder it manually to a larger extent.

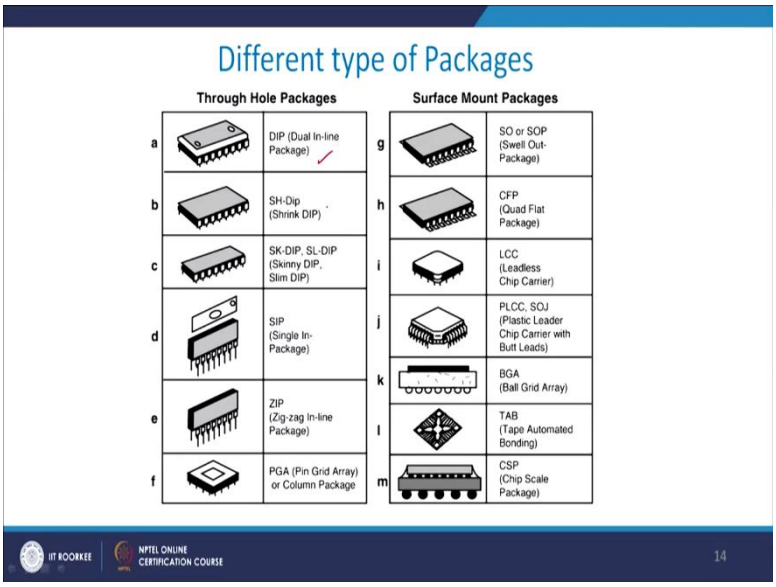
(Refer Slide Time: 26:44)




And then you have surface mount technology in which what you do is that so you do have SMT packages or surface mount technology packages that have leads right that has soldered to the surface of the exposed metal lands. So this is the circuit board, you just have to solder it here right you solder it here. So you do not have holes, elimination of holes is the first

advantage, it is easily manufactured and you can use both side of PCBs to do that, they have much smaller dimensions and therefore they have reduced parasitics and larger wiring density and therefore you do have a tighter control over the wire package right and that is what is the advantage of IC packaging in this case.

(Refer Slide Time: 27:28)

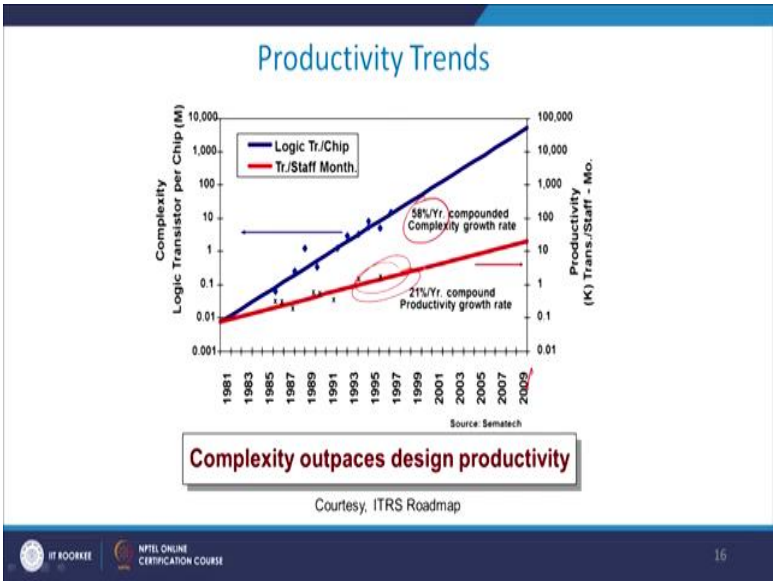



Different types of packages you can get it on various books. We have a DIP which is dual in line, we have shrink DIP, we do have SK-DIP these are all through hole packages right. So which means that you draw the whole through the system through the silicon and then you wire bound it here or you can do bonding here or you can do moulding here right and these are all types of packages available with you. So the left one is basically Through Whole Package and the right one is basically the Surface Mount Packages right. And so surface mount packages and through hole packages are available with you, but SMTs give you much much lower parasitics as compared to through hole and they are more reliable as compared to through holes right that is what you get from this idea.

(Refer Slide Time: 28:29)

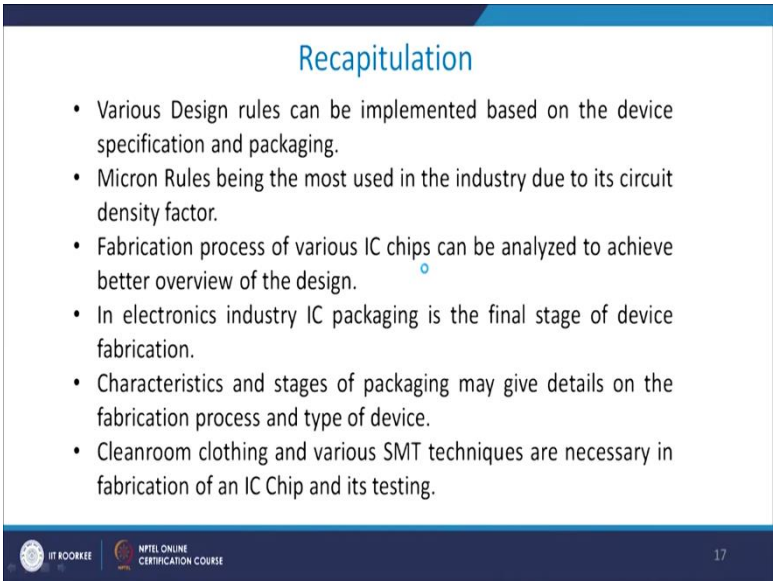

Let me give you a small maybe a figure or a photograph of the cleanroom. This is from TSMC which is basically Taiwan semiconductor manufacturing company, courtesy their photograph is there. And basically you can see the cleanroom, it is filtered such that the particles are 35 particles in the per cubic feet per unit volume which means we use cleanroom clothing to remove all the dust particles, and there is an air flow maintained and generally you can have it. So you generally have cleanrooms marked as 10, 100, 1000, 10,000 so on and so forth. 100 basically is the best one because there are 100 numbers of particles per unit volume of that area. Similarly 10,000 will also have some particles, so lower the value, higher is the cost for running as well as fabricating the cleanroom environment right, so this is typical environments which you get in cleanroom.

(Refer Slide Time: 29:06)



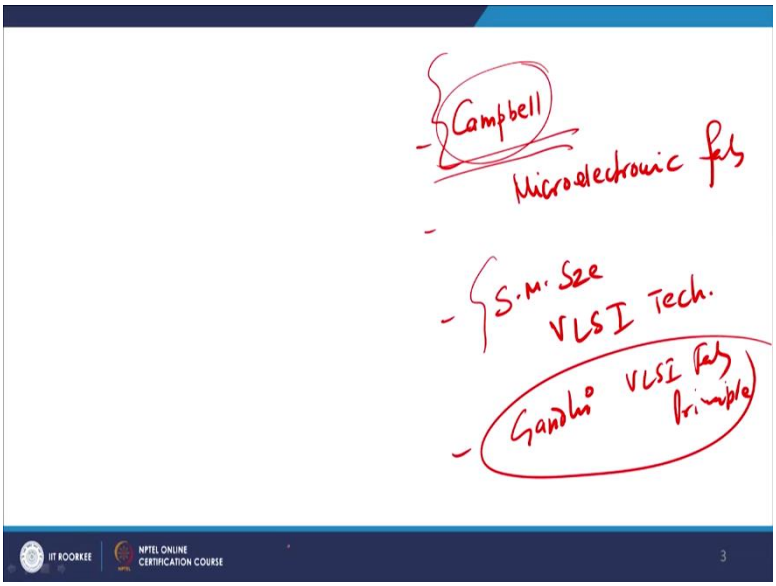

Now the Typical productivity trends which you see over the period of year. Though this is a bit old data 2009 data, holding good still is that the logic is actually growing much faster as compared to the transistor growth rate. It is approximately 58 percent growth is there in the logic right whereas, transistor per unit star is relatively small  relatively smaller 21 percent growth is there right and therefore higher the complexity, better is the trend nowadays happening so yield is getting higher and higher in this case. So this gives you a basic trend, so let me recapitulate what we have done till now.

(Refer Slide Time: 29:46)



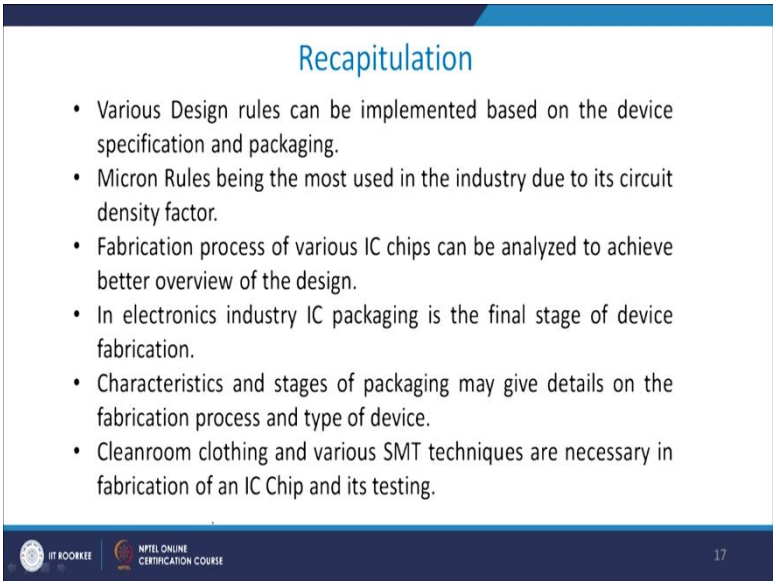
We have understood what are the various design rules which one is the lambda based design rule and another is the micron based design rule. Micron based design tool is the most used in companies due to its circuit density factor. Fabrication process of various IC chips can be analysed to achieve better review, so what you do? You just have to go through the fabrication processes, there are certain good books at which if you refer quite interested to go through a book, maybe I can give you some idea about the books which you can use.

(Refer Slide Time: 30:20)




There is a good book by Campbell on Microelectronics Fabrication. You also have typically SM Sze a good book on VLSI technology, Tata McGraw-Hill I think and but this is what there is a book by Gandhi, VLSI Fabrication principles right Fab principles but I would recommend that this is quite a heavy book, only meant when you are actually wanting to study deep inside the fabrication. Most of you what you will study in Campbell as well as in Sze, Sze is also slightly higher, I would recommend that at this stage if you are doing graduation and if you are actually in the first or second year of the study, have a look at Campbell. If you are slightly ahead and you have already done your bachelors and you are doing Masters, try to read Sze. And if you are faculty somewhere and you are a Ph.D. or post-doctorate maybe look at Gandhi first of all right. So these are the few books which you should be capable of or you should be able to handle them

(Refer Slide Time: 31:30)



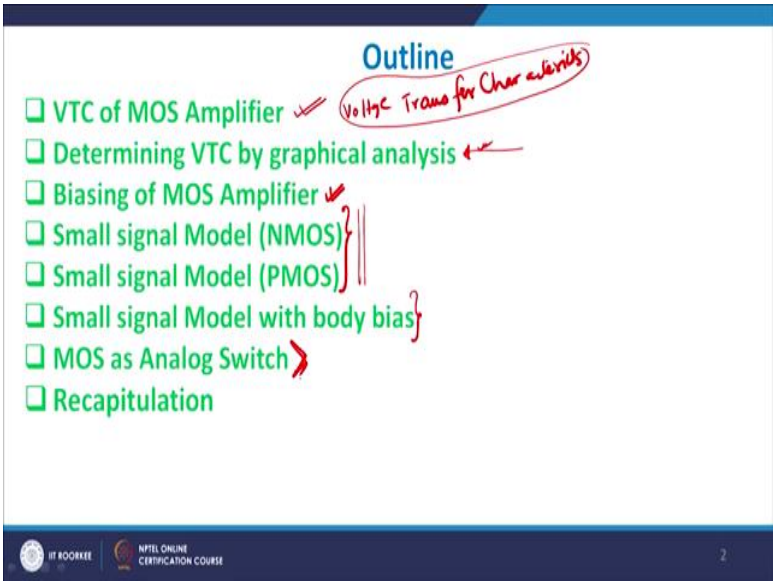

We have also understood that the packaging may be given in detail, we have discussed the packaging type. We have understood 2 types of packaging; through holes and SMT surface mount technology, and we have seen how a cleanroom facility typically looks like as well as the trend in the industry towards logic as well as on Semiconductor in general right. With this I thank you for your patient hearing, till we meet next time goodbye!!!!!

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−25**
**Biasing of Amplifier and its Behaviour as an Analog Switch-I**

Hello everybody and welcome to NPTEL online course on Microelectronics: Devices to Circuits. Till now we have actually done more of devices in perspective which means we have learned, understood the concept of BJT, MOSFET, various operation modes of the MOSFET. We have also understood how a BJT works and how does a MOSFET works, what are the various domains of its working principle. We have also seen in our previous discussions, how fabrication can be done in terms of MOS devices which means that I do have MOS device with me, I need to fabricate it on silicon so what are the procedures one needs to follow.

Our last Module dealt with basically the principle of fabrication as well as packaging, so that we know that what is the final goal of the whole course. So finally we need to have a device which is basically integrable onto silicon and then can be packaged properly and delivered as an entity. With this previous about 4 to 4 and half weeks of lecture we were able to finish therefore the basic device concepts. Now let me come to the circuit aspect of it, and the next about 8 to 9 weeks or say 7 to 8 weeks we will be devoting much on Circuits and understanding the circuit aspect. To do that we will be starting with today's lecture and that is basically the biasing of MOS amplifier and its behaviour as an analog switch so that is the topic of today's lecture "Biasing of MOS amplifier and then behaviour as Analog switch.

(Refer Slide Time: 2:15)




So what we will be doing in the first module of this course is we will be looking into the VTC of the MOS device, right. So I have a MOS amplifier, let me look at the basic simple MOSFET and then in a simple MOSFET let me draw for you the Voltage Transfer Characteristics, so VTC is basically your Voltage Transfer Characteristics. Using this we will be understanding certain things and certain areas of operation of the device. After this when we know VTC, we should actually graphically know can do the graphical analysis of this VTC right. So you have VTC available with you and you want to do a graphical analysis, then we actually how to go about doing it is basically my second point.

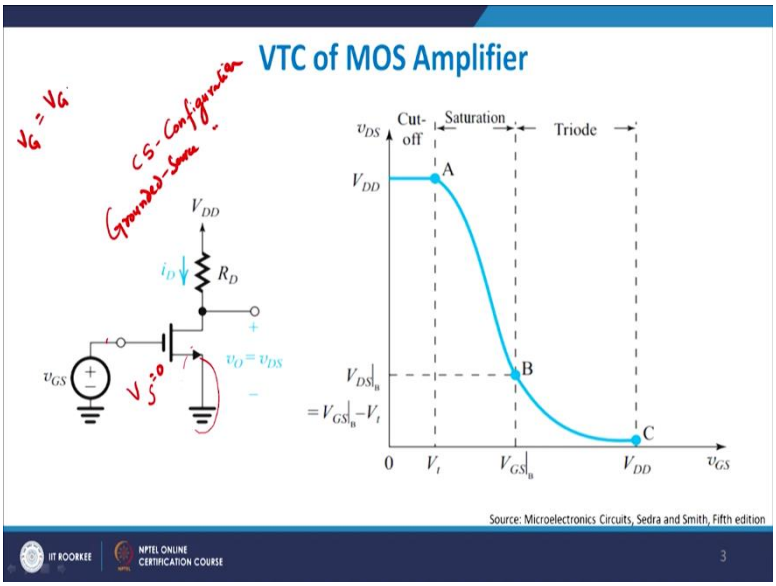
The third point is, we will learn about MOS amplifier how to bias it, now that is pretty interesting and important and the reason is if you remember very carefully that MOS can operate in triode region, in saturation region or in the cut-off region. In the triode region it works as a resistor, in the deep triode region it can work as a voltage variable resistor we have already discussed with you, in the saturation region it works as a current source and in Cut-off it is basically switched off. So whenever you want to use MOS device as a switch, you just have to shift it from saturation to the cut-off and vice versa, you just have to go from saturation to cut-off and Cut-off to saturation right that is known as switching right, so you know how to bias it.

But then if you want to use MOS device as an amplifier, you should know where the initial biasing of the device should be so that I get linear amplification with me, which means that the amplification $A_V$ or voltage amplification voltage by voltage is not a function of the input voltage. So even if you are varying the input voltage drastically, my gain should remain

constant and should be very high right. So you should know therefore how you bias your amplifier that is the reason this third one is important one right. We will be therefore looking at the small signal circuit models for NMOS and PMOS, obviously reasons asked is why will you require it?

Well, we require it because finally we have to use those MOS devices in a circuit environment and therefore if you are able to predict how does a circuit behaves or if you can predict how does a MOS device behaves, we can just take its equivalent model and place it in those places where MOSFETs have been there. This makes our calculation and understanding very easy from circuit perspective that is the reason we generally do small signal modelling here right. We again therefore look at small signal model with body biased and this I will come to this later on as we have discussed that body bias are important point and then we finally end our module by explaining how MOS works as an analog switch right, and we will see how it works as an analog Switch overall.

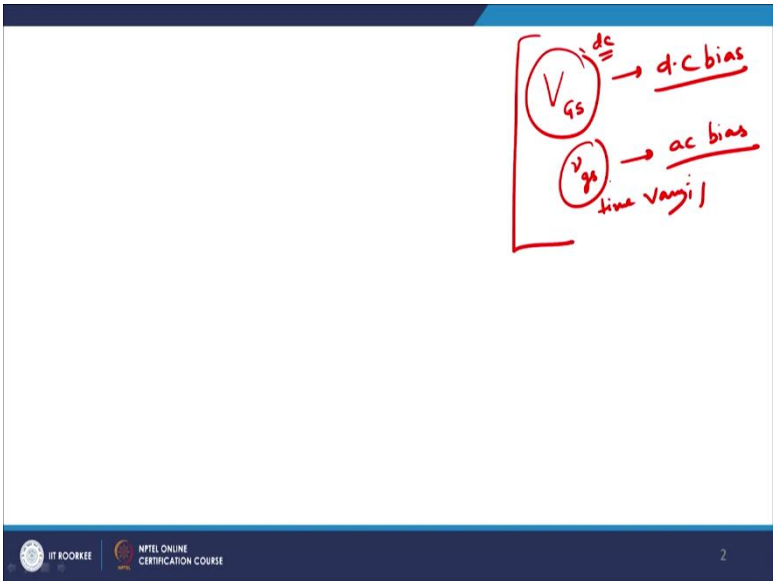
(Refer Slide Time: 5:07)




Now, I will just give you a brief insight into the voltage transfer characteristic of MOS amplifier. Now if you look very carefully, this is also known as common source configuration, if you look in front of you is basically a Common Source Configuration also known as Grounded Source Configuration right Grounded Source Configuration, and the reason being your source is basically grounded right. So source voltage is always held at equals to 0, so whatever gate voltage you give here is actually equals to gate to source voltage so your gate voltage $V_G$ in this case is always equal to $V_{GS}$ fine that this you should be very

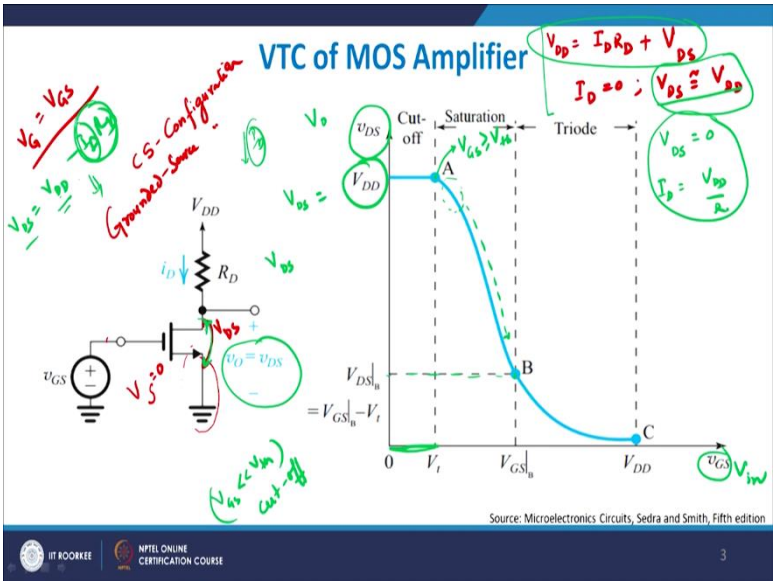
careful about this point. I again want to re-stress some important point which is you should be always careful to be handle about notation.

(Refer Slide Time: 5:56)




Notation is that if it is capital $V_{GS}$ we have right, you can have small v gs right, so capital $V_{GS}$ is basically the meaning of DC bias, DC biased primarily means that you have given a fixed value of gate voltage with respect to source voltage and it is basically a DC and therefore it is defined like this, v gs is basically your AC bias right or AC signal which means that it is a time varying signal so this is basically your time varying signal which you see time varying which you see. So you should be able to distinguish between capital $V_{GS}$ and small $v_{gs}$ and what does it stands for right at this stage.

(Refer Slide Time: 6:40)

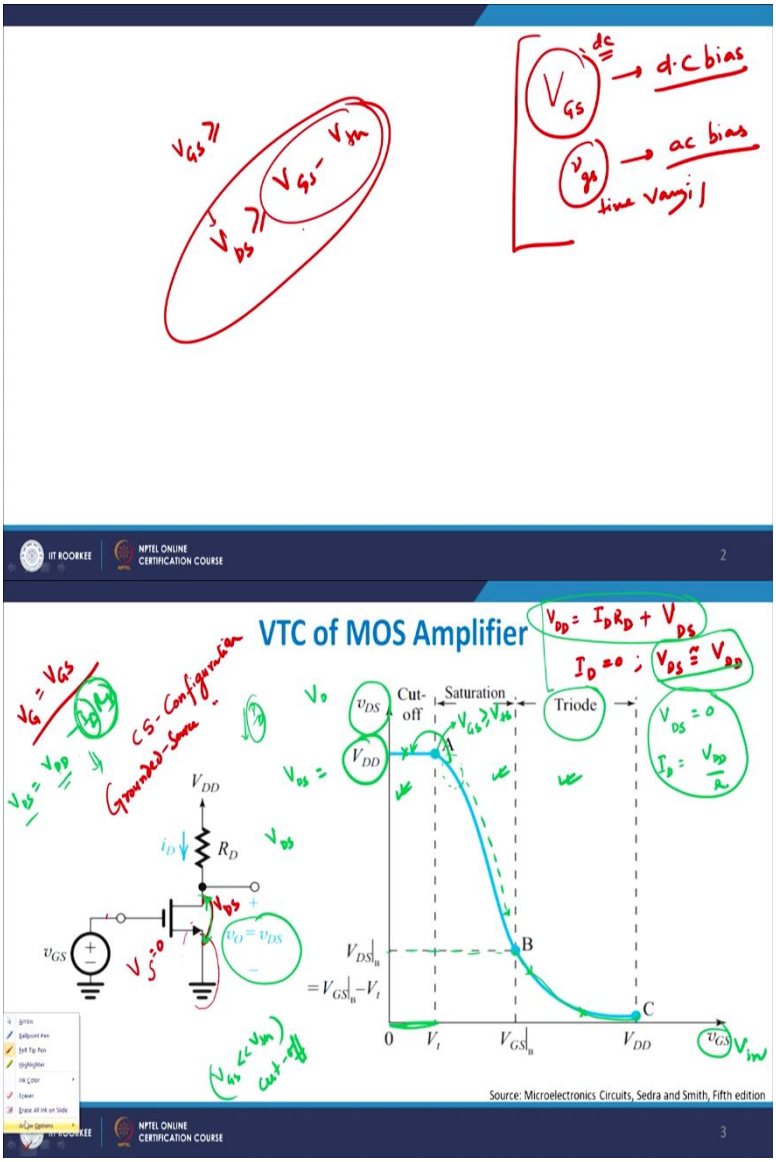


Now if you look carefully at this graph this profile, you will see I can simply write down for this case that $V_{DD}$ equals to $I_D R_D$ plus $V_{DS}$ drain to source voltage right fine. Now which means that if your $I_D$ is equal to 0, right, $I_D$ is equal to 0, then $V_{DS}$ is approximately equal to $V_{DD}$ right. And when your $I_D$ is not 0, $I_D$ is typically a large value or $V_{DS}$ is equal to 0 let me just tell you that you have $V_{DS}$ is equal to 0 which means that so your $V_{DS}$ is equal to 0 implies that your $I_D$ will be equal to $V_{DD}$ by $R_D$ fine. If you therefore see very carefully that under the condition that you drain current is 0 and your device is cut-off, the output voltage will be dragged to $V_{DS}$. So as you can see here, $V_O$ is equal to $V_{DS}$, right. And similarly, when your drain to source voltage is equal to $V_{DS,}$ your $I_D$ is equal to 0 cut-off. And whenever my drain to source voltage is 0, I have a finite value of $I_D$ flowing into it and that is what this with me.

Now understand that if I plot $V_{DS}$ which is drain to source voltage versus $V_{GS}$ right which is the input voltage, so this is my output voltage and this is my input voltage right. So I have an input voltage which is $V_{GS}$ and output voltage which is $V_{DS}$. When my $V_{GS}$ is below threshold voltage of the device, my device as I discussed earlier is in Cut-off mode right. Cut-off means this is cut-off so there is no current flowing through you, $I_D$ is equal to 0 and as you can see $V_{DS}$ is equal to $V_{DD}$ see so this $V_{DS}$ is equal to $V_{DD}$ fine till a point you reach A, at A your MOS device turns ON, because your, at this point at node A, I get that $V_{GS}$ is just greater than equal to $V_{TH}$ threshold voltage of the device and as a result the device switches ON and the current starts to flow, $I_D$ starts to flow.

$I_D$ is initially low and then it increases, when it is initially low then if you can find out from this equation, from this equation if you find out $V_{DS}$ therefore will be equal to $V_{DD} - I_D R_D$. So even $I_D$ is low, this quantity is small and therefore $V_{DS}$ is almost close to $V_{DS}$ which is this region. As $I_D$ starts to increase, this quantity becomes larger and larger and therefore $V_D$ minus that quantity starts to fall down and this is what is happening, your falling down, so the output voltage is falling down till you reach point B right. Point B is the point where you reach to a point where your $V_{DS}$ drain to source voltage $V_{out}$ has actually fallen so low that you have driven this almost into the edge of the triode region fine.

(Refer Slide Time: 10:26)

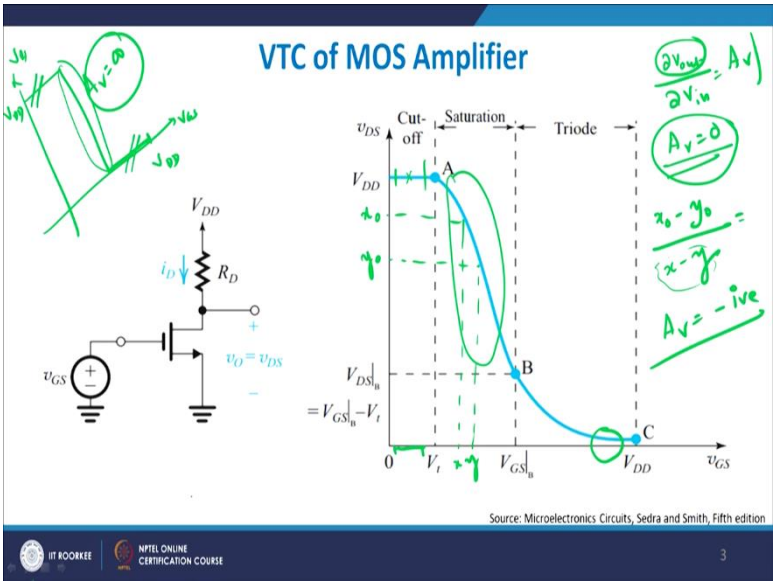



So you remember that if $V_{DS}$ was high right, you were into saturation if $V_{GS}$ was larger than, so the condition for saturation was that $V_{DS}$ should be great than or equal to $V_{GS}$ minus $V_{TH}$ this is one condition and that is the reason $V_{DS}$ should be quite large because this is quite large and therefore $V_{DS}$ should be quite large. So if $V_{DS}$ falls below a particularly limit which is happening here right and output voltage is falling below a particular limit you enter into triode region and that is the reason you see this is triode region.

And in the triode region, the current is the nonlinear function of the input voltage, and you see as the current increases the voltage further rises till you reach point C, and at point C the $V_{DS}$ is so small that there is no current flow and therefore your voltage transfer characteristic stops at this particular point. So we have therefore understood the VTC of MOS amplifier is divided into 3 parts; one is the cut-off, another is saturation, another is triode region right. So

if you want to use it as a switch, you need to move from cut-off to saturation and from saturation to cut-off very fast, right.

(Refer Slide Time: 11:35)

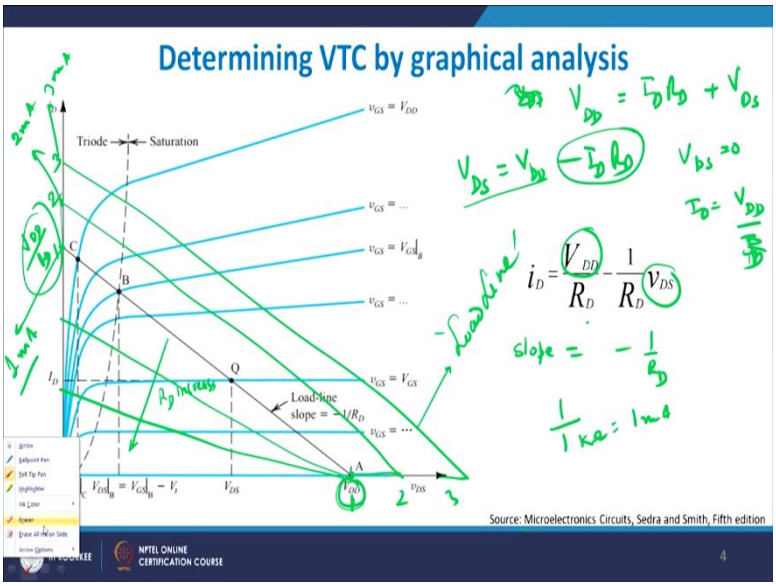



If you want to use it for example if you are biasing here or here which means at the cut-off triode region then you can safely see or you can actually see that if you are biasing somewhere here let us suppose right. And we try to find out output $\partial V_{out}$ right and $\partial V_{in}$ which is basically my voltage gain right peak to peak. Then you will see that even if I shift my input voltage by some amount, my output voltage still remains fixed which means that $A_v$ is equal to 0 because $\partial V_{out}$ is 0, which means that as you go on shifting your input voltage, your output voltage does not change and your voltage gain fix to 0, same thing will happen somewhere here where it is almost flat right.

So where the gain will be there? Gain will be there in the saturation region, where $V_{out}$ is drastically varying even for a small change of $V_{in}$. So even if your $V_{in}$ has changed from this point say to this point because of heavy nonlinearity, the output has actually changed from so if this is your input X and output Y, then this will be $X_O$ $Y_O$ and therefore we can define the gain to be equal to $X_O$ minus $Y_O$ divided by Y minus X or X minus Y. So it will be X minus Y actually but what we do is we write something like this, we write this to be as X minus Y so therefore this will be negative in dimensions because X is smaller than Y and therefore this gain $A_v$ will be actually a negative quantity but will be a large quantity. So more steep is the slope, more will be the drain current, more will be the source.

So ideally VTC looks like this ideal, right, this is actually equals to $V_{DD}$ here and this is also equal to $V_{DD}$ here, this is $V_{GS}$, this is $V_{DS}$, right. And if you bias your device somewhere here I should ideally get $A_v$ to be equals to infinity voltage gain to be equal to infinite. At this point the output latches to 1, at this point is latched to 0 so it is almost working as a digital switch, right, so by application of external bias I am able to switch it off from this side to this side. So with this let me therefore come to you how do I do it by graphical method.

(Refer Slide Time: 13:58)



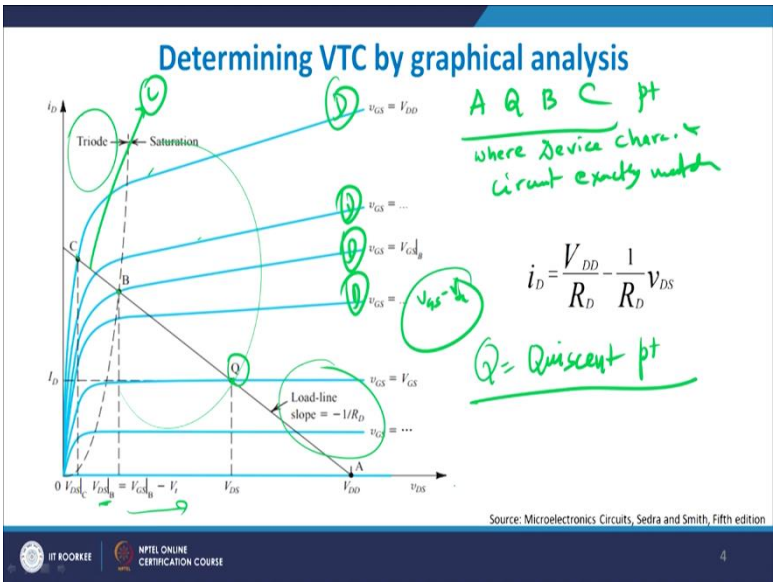

See as I was discussing with you that if you plot $I_D$ versus $V_{DS}$ right, I will get one point in $I_D$ axis which is equal to $V_{DS}$ so as I was discussing with you $V_{DD}$ if you go back to the previous slide, $V_{DD}$ equals to $I_D R_D$ plus $V_{DS}$, right. So what I can write down safely is that $V_{DS}$ equals to $V_{DD}$ minus $I_D R_D$. Now when in this case, so if you look at the X axis, it is $I_D$ which means that $V_{DS}$ equals to 0. $V_{DS}$ 0 means $I_D$ will be equal to $V_{DD}$ by $I_D$ so this point is basically your $V_{DD}$ sorry, $R_D$ that will be this point. And what about this point? This is the point where your $I_D$ is equal to 0. So when $I_D$ is equal to 0, this quantity equals to 0 and therefore $V_{DS}$ is equal to $V_{DD}$ that is the reason this is $V_{DD}$ fine. So I have a line which connects one point on the Y axis as $V_{DD}$ by $R_D$ to another point on Y axis which is basically equals to $V_{DD}$ fine.

And therefore if you drop this line, the slope of this line is given so if you actually plot this line from this equation, I get $I_D$ equal to $V_{DD}$ by $R_D$ minus 1 $V_{DS}$ by $R_D$, $V_{DD}$ is a DC bias and $V_{DS}$ is the applied variable bias. So as you can see, the slope of this curve is equal to minus 1 by $R_D$ so higher the value of $R_D$ less is the slope available to you. So if you make $R_D$ large then what happens is that this quantity will drop down and this will not change, this

will be $V_{DD}$ and therefore if you increase your $R_D$ it shifts to this, this is the increase in $R_D$, $R_D$ increases right. Whereas, if you would have increased $V_{DD}$ then this would have shifted to this point and this could have also shifted by the same amount to this point and therefore I will get a set of parallel lines when I increase $V_{DD}$.

So let us suppose your $V_{DD}$ was say 1 volt here then 2 volts, 3 volts, so on and so forth and so this will be 1, this will be 2, this will be 3. And this divided by $R_D$ will say your… so 1 volt divided by 1 kilo ohms will give you the value of 1 milliamp right, so this will be 1 milliamp, 1 milliamp, similarly this will be so this is 1, this is 2, 2 by 1 will be 2 milliamp and this will be 3 milliamp right and so on and so forth. So these are known as the Load lines so to understand this issue this is actually known as the Load line, right and the reason of the Load line is that they exactly follow the issue that inverse of that is basically the slope, in fact inverse of the load is basically the slope right and higher that value lower is the slope which is available with you.

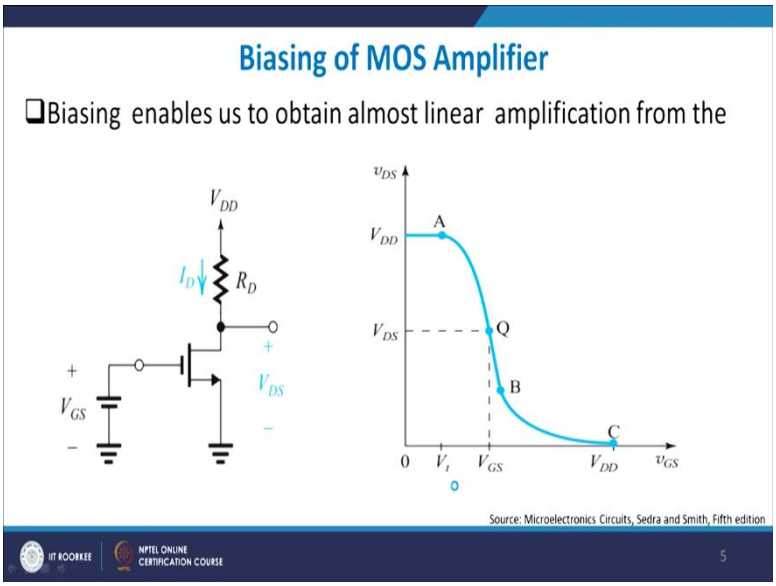
(Refer Slide Time: 17:15)




Now with this knowledge what we can therefore write down is we learned two things out of it, the first thing is basically that you do have always current which is flowing through the device, will depends upon many factors but primarily upon external circuit factors. The second idea was that always there will be a load line whose slope will be equal to minus 1 by $R_D$, and whenever it cuts your IV characteristics of the device which is this, this, this are the points AQBC so AQBC are the points where device characteristics right and circuit they exactly match, right why because this is the device but this is device, this is device, this is device, this is device whereas this is basically a circuit.

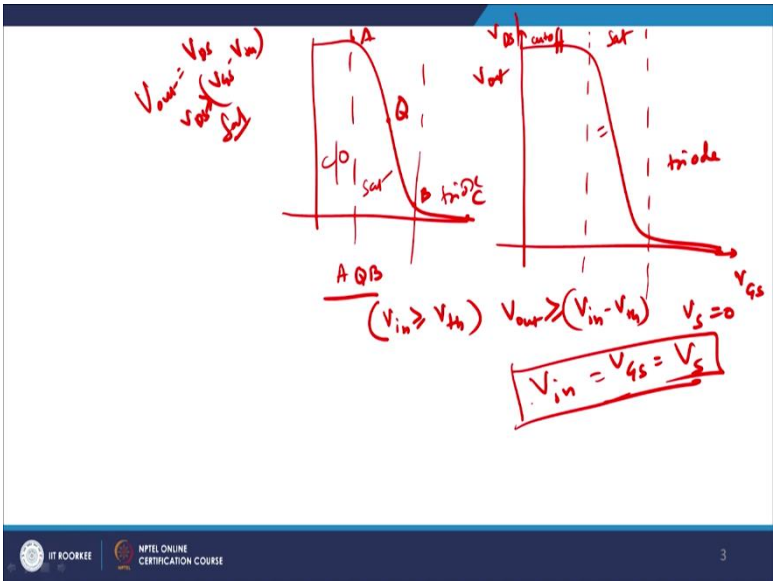
So when it cuts both of them are the points where both devices and circuits properly give you a match current available with you right and therefore these are effectively known as the points of operation of the device out of which we take one point which is also referred to as Q point quiescent point, Q point is known as quiescent QUISCENT primarily meaning that this point is very-very insensitive. So even if your voltage is varying slightly, the Q point will be varying but output will be very stable in output side right and this is where it cuts almost in the middle of the saturation region. So this is the saturation region right, and this is the middle of saturation region you will find the Q point, this is your triode region when your $V_{DS}$ is very-very small alright so that is what we were discussing here that in the bulk side you will have a low value and as you increase it, so this is basically your $V_{GS}$ which is equal to $V_{GS}$ -$V_{TH}$ also known as over drive right.

(Refer Slide Time: 19:14)




With this knowledge we come to important point that I always want a linear amplification primarily meaning that my gain should be almost independent of the input voltage and should be fixed for a given value of or a range of input voltage right. So you can see here from this basic operation principle that you have this region available with you so I have to $V_{DS}$ and $V_{GS}$ and threshold voltage $V_{TH}$ and so on and so forth. Let me now explain to you that given a VTC right, how can we therefore draw the output characteristics of the device and what will be the fundamental characteristics of the device.

(Refer Slide Time: 20:01)

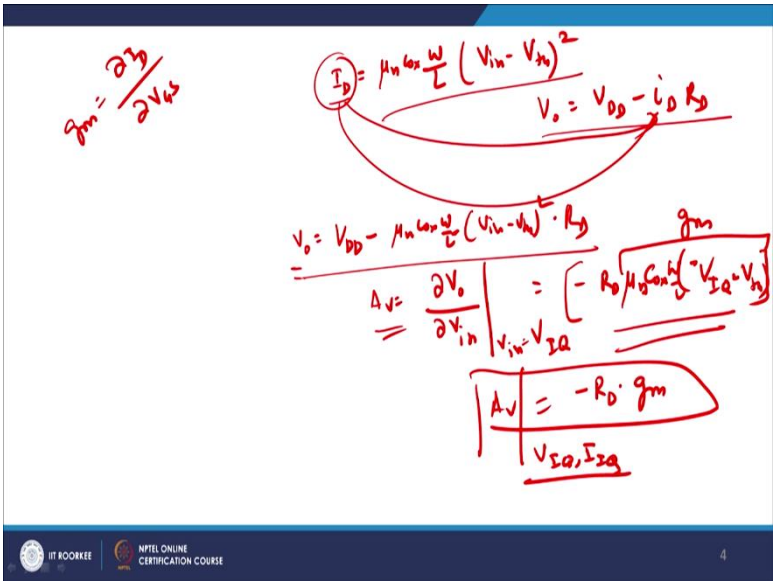



So let me just draw for you right and this is your point, right, this is your cut-off, this is your saturation and this is the triode region and this is $V_{GS}$ and this is your $V_{DS}$ also referred to as $V_{out}$ fine. So let us at each point at each section of our voltage transfer characteristic let us see if we can evaluate the $I_D$ -$V_D$ characteristics right and then try to find out voltage gain available here. So with this knowledge let me first start with the saturation region of operation which is this one and let us see how current is formed there, current is formed in saturation region of operation right.

So what we do here is, we take up this case and then we draw a profile here right and we say that it is like A B and then how many like this, so we referred to this as point B, this is point C, this is point A and let us suppose this is point Q, this is the quiescent point right and we mark it as this and then this. Of course we discuss with you saturation, cut-off and this is your triode region. And the behaviour of current is totally different in all these 3 regions; in cut-off region current is 0, in triode region it is a strong function of $V_{GS}$, in saturation it is almost linear function of $V_{GS}$. Now, in this region, in region AQB let us suppose, I will get $V_{in}$ input voltage should be greater than equal to threshold voltage of the device, we have already discussed this point and $V_{out}$ is greater than equals to $V_{in}$ - $V_{TH}$.

Remember why? Because $V_{out}$ is equal to $V_{DS}$ right and $V_{DS}$ should be a greater than equal to $V_{GS}$ - $V_{TH}$ for saturation right and therefore I get $V_{out}$ to be greater equals to $V_{in}$ - $V_{GS}$. Now since $V_S$ is equal to 0 right therefore, $V_{GS}$ is equal to $V_S$ and therefore this equal to $V_{in}$ input voltage which you see is almost equal to the source voltage in general, right. Now the second is the output, so output is given as $V_{in}$ - $V_{TH}$.

(Refer Slide Time: 22:33)



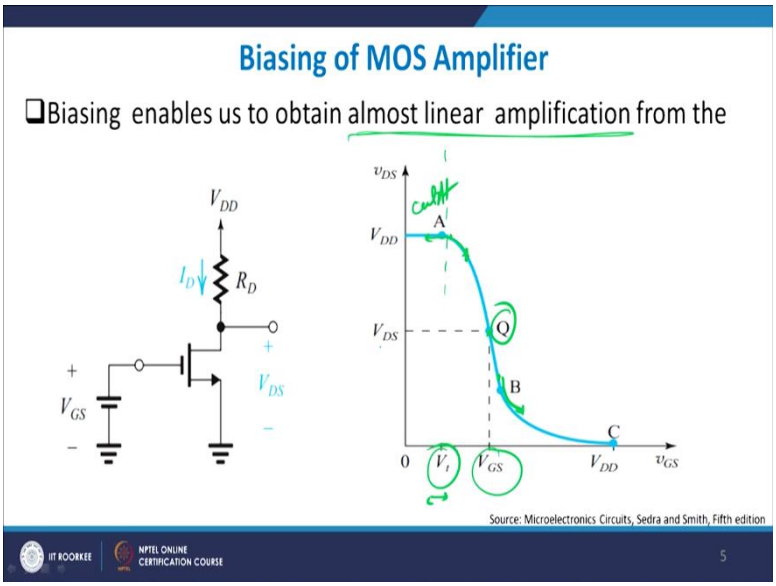

With this knowledge we can safely write down $I_D$ which is the drain current will be equal to $\mu_n$ C oxide right, W by L into $V_{in}$ minus $V_{TH}$ whole square which therefore again means that so therefore $V_O$ will be equal to $V_{DD}$ minus $I_D R_D$ and therefore if you place this value of $I_D$ back into this equation and get $V_0$ equals to $V_{DD}$ minus $I_D$ is $\mu_n$ C oxide W by L right multiplied by $V_{in}$ minus $V_{TH}$ whole square into $R_D$ right and this is what you get. Now if you therefore see this is your $V_0$ right so I need to find out therefore $A_v$ as Del of $V_0$ to $\partial$ of $V_{in}$ right and $V_{in}$ is equal to V IQ, this comes out to be minus $R_D$ $\mu_n$ C oxide right W by L into $V_{IQ}$ minus $V_{TH}$, this is the gain in the saturation region fine.

So let me again recapitulate what we did, I am assuming that device is fully into saturation right and there is no CLM and therefore I can write down $I_D$ to be equal to $\mu_n$ C oxide W by L into $V_{in}$ -$V_{TH}$ whole square, and $V_{out}$ will be equal to $V_{DD}$ minus $I_D R_D$. We also saw that in saturation that $V_{DS}$ or $V_{out}$ should be greater than equals to $V_{GS}$ minus $V_{TH}$. Therefore putting the value of $I_D$ into this equation here I get this into consideration, then if you find the differential of $V_{out}$ to $V_{in}$ which is basically the voltage gain, I get this consideration and therefore I will just give you an idea that $R_D$ into this is known as trance conductance gm this whole thing is known as gm also referred to as trance conductance of the device.

Tranceconductance primarily means $g_m$ basically means $\partial I_D$ and $\partial V_{GS}$, which means that how fast is your drain current changing with respect to similar change in the value of gate to source voltage right which means that how much your device is sensitive to input gate

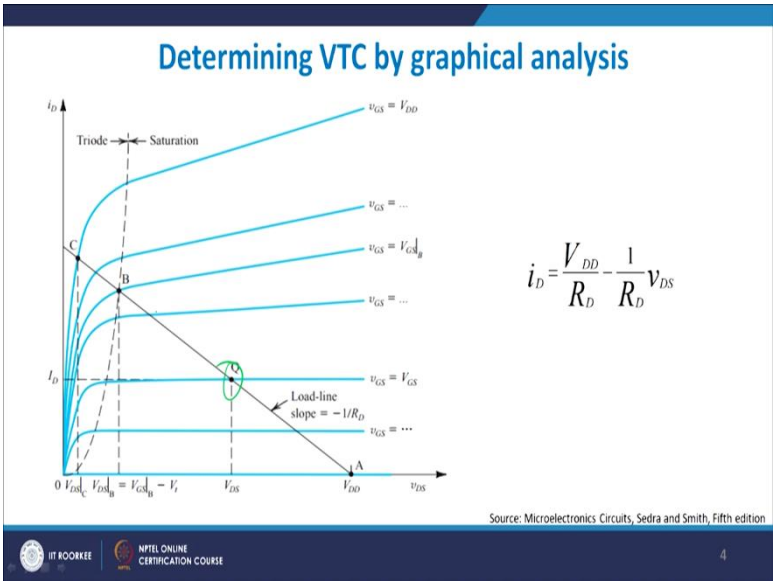
voltages right. If $g_m$ is high transconductance of the device is high, I would expect to see a large change in the output current for the same change in the value of $V_{GS}$. Whereas when the transconductance is low, I will get a drain current which is actually smaller with same change in the value of $V_{GS}$ right. So please understand therefore the gain as this point but one point to be noted at this stage is that gain at which point only at $V_{IQ}$ or $I_{IQ}$ which means that if you go back to your previous slide which means that.

(Refer Slide Time: 25:44)



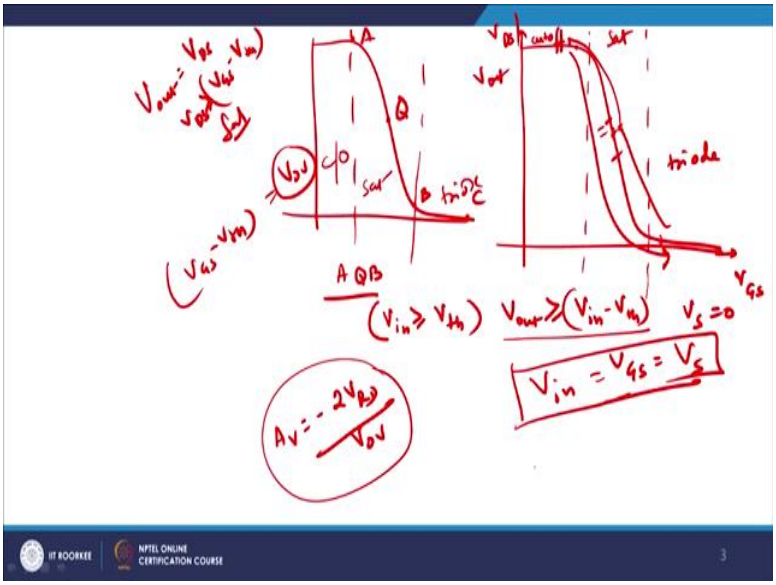

I have set my Q point to be Q. If you shift your Q point somewhere here or here very edge to the triode region or very edge to the saturation region right then even if there is a small change, so you see you have placed it here right. Now you do a small change in $V_A$ then this will either shift back into A or it will come this side, if it will come this side you will have great nonlinearity as I discussed with you. And if it goes this side then it will go to cut off fine and both ways I do not want it and then similarly if I place it here and I allow it to subtract then most of it will go to triode region, so people thought that let us put a Q value that is somewhere in the middle of the two and help both of them right and that was the reason why this all sorts of variations came into picture here.

(Refer Slide Time: 26:52)




But I wanted to stress again here one important point is that just as you saw in the previous case, try to keep your Q point in the middle of the saturation region rather than in the edge here at the edge here that makes sense also. So we have understood the saturation region and we have understood the voltage gain

(Refer Slide Time: 27:09)




Another popular formula for voltage gain is basically given by this formula that $A_v$ equals to minus times 2 V $R_D$ divided by V overlap, V $R_D$ is the voltage drop across the resistor R which is there on the drain side so V $R_D$, divided by V overlap is basically the voltage $V_{GS}$ minus $V_{TH}$ is basically the voltage overlap right and that is quite interesting phenomena which we get. So I am not deriving that part, you can do it yourself but that is typically what

we are trying to do. With this idea understood and we have understood therefore in this lecture the basic concept of amplifier, how does an amplifier works, why do you want to bias the amplifier so that it works in a proper manner and does not give you a distortion. We also looked into the fact that given a MOS device and cut-off region, saturation region and triode region, exactly at which point should you bias your device so that I get the best available linear amplification available to me.

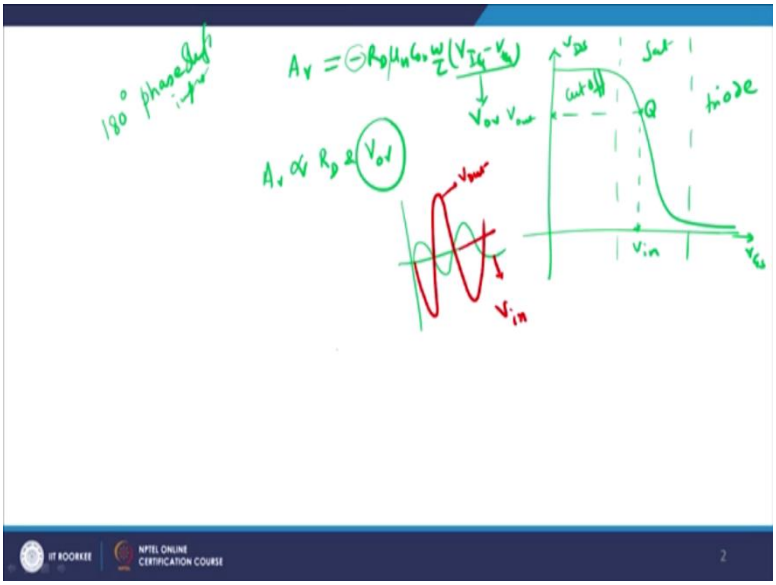
As I discussed with you, if you keep your Q point very close to either the cut-off region or the triode region, chances are that by application of input voltage the Q point may actually shift into cut-off or into saturation region or triode region, and as a result you will get non-linear output so it is always advisable to keep your input signal in a small signal mode and therefore the peak to peak input signal should be as small as possible then I will be able to maintain linearity into the system which I just showed you that which means if you bias it here and then if you move from this place to this place it is almost linear, but if you bias it here and then apply a very strong input voltage then it go somewhere here and then it again comes back here to here and so on and so forth right.

So if you see very carefully, you will find that these are the areas where you will have excessive nonlinearity coming into picture but there will be surely large gain available with you at this point and this point. We have also understood the basic functionality, what are the regions of operation for a MOS device. In our next class in our next module we will be actually following the triode region as well as the various biasing schemes in a MOSFET fine with this let me thank you for your patient hearing, thanks a lot!

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−26**
**Biasing of Amplifier and its Behaviour as an analog switch-II**

Hello everybody and welcome to the NPTEL online certification course on Microelectronics: Devices to Circuits.

(Refer Slide Time: 00:52)




We will start from where we left in the previous turn, we had actually seen that the gain depends upon the overdrive and we saw from the previous turn that when the device is in saturation region, right when the device is actually in the saturation region then you do have a profile here and therefore if you look very carefully that this is the saturation region which we are talking about this one, this is the cut-off and this is your triode and this is the saturation. So if you bias your device somewhere here right, somewhere here with this to be as $V_{in}$ and this is to be $V_{out}$, right this is your $V_{in}$ right.

Then I will be able to achieve a gain which is given as $A_V$ was equals to, if you remember correctly in your previous turn to be equals to minus $R_D$ times $\mu_n$ C oxide W by L into $V_{IQ}$ minus $V_{TH}$. $V_{IQ}$ is nothing but the input at this point Q, right and this is therefore the nothing but the overdrive, V overdrive, right.

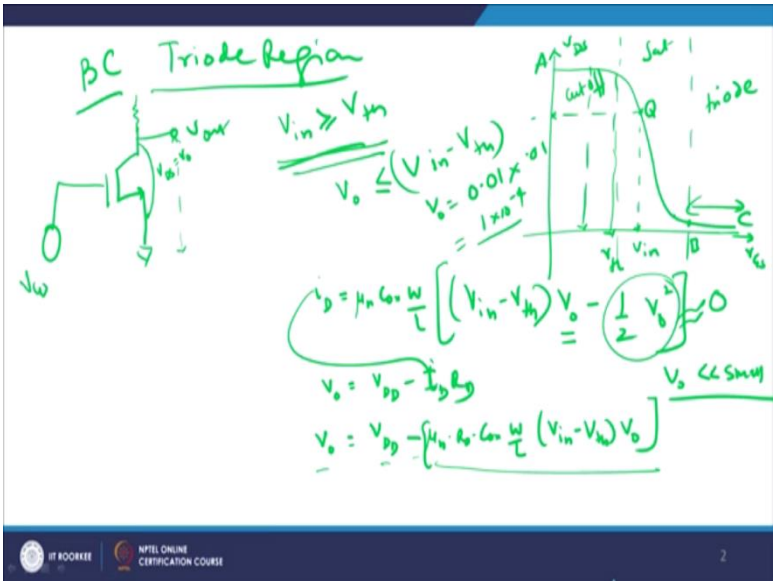
So $A_V$ voltage gain is therefore proportional to $R_D$ and also onto the overdrive V overdrive. So higher the overdrive more will be the gain and of course higher the $R_D$ you will also get the gain, but you see you have minus sign here negative sign which tells me that a MOS

device, when you have a source grounded condition will always have the output 180 degree phase shifted right, phase shifted as compared to input, right.

So you will always have a 180 degree phase shift and therefore it will always be negative in nature. So if you have input something like this, right then the output will be the output if you want to find the output, the output will be something like this that it will be like this, it will go like this and then like this and this, so it is 180 phase shifted. This is the $V_{out}$ right, and this the $V_{in}$, right and if you want to find out the gain it is basically $V_{out}$ by $V_{in}$ at any particular point.

So it is 180 degree phase shifted. Now let us come back to our original discussion and see how does this works in the triode region, right so what we will be doing is we will be now discussing what is happening in the triode region. So let me erase some part of it and just give you an idea about the basic functionality here that if you did have a triode region, we saw the gain in a saturation region is independent of the applied input voltage, it only depends on the value of $R_D$ and it does not depend upon anything else, it depends upon the overdrive voltage as well.

(Refer Slide Time: 03:36)




So with this discussion let me come to the region, let us suppose this is point B and this is C, this is A, right. Let me discuss about the point region BC which is this point, right. This is actually your triode region of operation, triode region. In the triode region if you see very carefully then input voltage is obviously larger than threshold voltage because already threshold voltage is somewhere here or somewhere here is the threshold voltage $V_{TH}$.

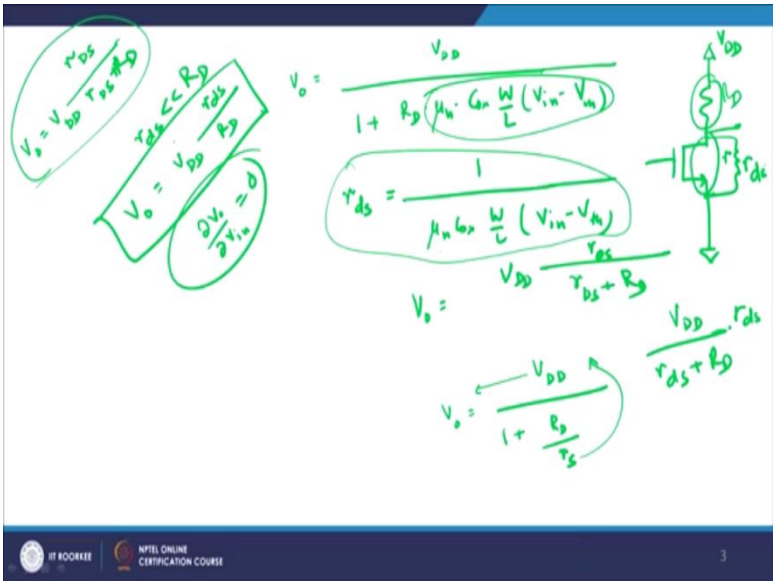
So your input voltage is obviously larger than threshold voltage but what has happened is that your $V_{out}$ has actually fallen below $V_{in}$ minus $V_{TH}$ that is quite critical. That, your $V_{out}$ has fallen below, so your idea was that $V_{DS}$ should be greater equals to $V_{GS}$ minus $V_{TH}$ for saturation, but once your $V_{out}$ falls below $V_{in}$ minus $V_{TH}$, right then you are in a fix that now the condition for saturation will not be applicable.

Now the device will start behaving like a resistive network or resistive device and therefore I can write down in this case $I_D$ to be equals to $\mu_n$ C oxide W by L, right and then we right down $V_{in}$ minus $V_{TH}$ into $V_o$, I will tell you why, into half $V_o$ square. Why? Because if you remember the previous discussion which we were doing MOS device was something like this and you did have $R_D$ here and this was your $V_{GS}$ and this was grounded and you are taking the $V_{out}$ from here. So if you take $V_{out}$ from this to ground, it is exactly the same as drain to source because source is grounded. So therefore $V_{DS}$ is equals to $V_{out}$ and that is the reason we change from $V_{DS}$ to $V_0$, right and $V_0$ is the $V_{out}$ which you see from here.

Where $V_0$ is exactly equals to $V_{DD}$ minus $I_D$ $R_D$. So what I do, I just put the value of $V_0$ equals to $V_{DD}$ minus $I_D$ I just plug it from here and place it here and we write down to be equals to $\mu_n$ into $R_D$ into C oxide into W by L into $V_{in}$ minus $V_{TH}$ into $V_0$ assuming that my $V_0$ is very very small.

So if $V_0$ is very very small, $V_0$ square will be negligibly small and this can be approximated to be equals to 0. For example $V_0$ is say 0.01 or so whatever then if you multiply it 0.01, I get 1 into 10 to the power minus 4 and so it is very very small as compared to this quantities here and therefore you can safely say that $V_0$ equals to $V_{DD}$ minus this whole quantity into $V_0$.

(Refer Slide Time: 06:32)



So we can write down therefore with this explanation, we can write down $V_0$ to be equals to $V_{DD}$, $V_{DD}$ divided by 1 plus $R_D$ times $\mu_n$ times C oxide into W by L into $V_{in}$ minus $V_{TH}$ and you get this. The effect was $R_{DS}$ as the drain to source resistance as 1 by $\mu_n$ C oxide W by L $V_{in}$ - $V_{TH}$ and therefore I can write down $V_0$ to be equals to $V_{DD}$ into $R_{DS}$ into $R_{DS}$ plus $V_{DS}$ $R_D$.

How did I do that? See, simple if you look very carefully here if you take $R_{DS}$ to be equal to this quantity because $R_{DS}$ is 1 by this quantity, $R_{DS}$ is the drain to source resistance of the MOS device in the on condition.

So once this is true if you place $R_{DS}$ here, so this quantity can be written as 1 by $R_{DS}$ remember, so I can write down $V_0$ to be equals to $V_{DD}$ upon 1 plus $R_D$ by $R_S$. So if you take LCM and do also some manipulation, $V_{DD}$ comes outside, this denominator will come in the numerator and I will get $R_S$ into so on and so forth. Finally, I get this thing:

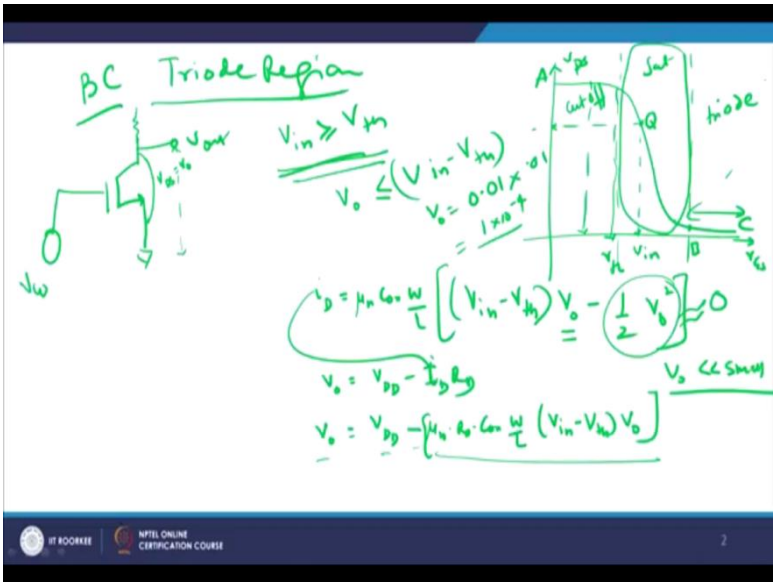
So I will get $V_0$ equals to $V_{DD}$ into $R_{DS}$ upon $R_{DS}$ plus $R_D$, $R_D$ is the resistance on the drain side and if you look very carefully how it should like also, because when we were discussing just know if you look here carefully then this is $V_{DD}$ and this $V_{DD}$ is being broken into two parts as a voltage divider network and these two parts will be inversely proportional to the resistance offered by this as well as this. This is $R_{DS}$, so this is basically your $R_{DS}$ and therefore $R_D$ and since they are in series to each other, so the total current flowing through circuit will be nothing but $V_{DD}$ by $R_{DS}$ plus $R_D$.

This you multiply with $R_{DS}$ to get the voltage at this particular point and that is what you are getting, $V_{out}$ is equals to $R_{DS}$ upon $R_D$ plus, $R_{DS}$ plus $R_D$. Typically what you get is that $R_{DS}$ is very very small as compared to capital $R_D$ and therefore we write down $V_0$ to be equals to $V_{DD}$ divided or $R_{DS}$ right by $R_D$. This is your output voltage which depends upon $V_{DD}$ into $R_{DS}$.

Now as you can see very carefully or interestingly that $V_0$ here is actually independent of $V_{in}$. So if you want to find out $\partial V_0 \partial V_{in}$, this will be actually equal to 0 because you do not get, you do not have any $V_{in}$ on the right hand side but it depends upon the value of $V_{DD}$, $R_{DS}$ and $R_D$, clear?

So as a result, I want to stress one important point here that in the triode region your gain will be fixed, will be almost 0 not only that your output voltage will also fixed and will not depend upon the input voltage and that is the reason as I explained to you. You will always get a constant value of the output voltage independent of the input voltage. This is one thing which I wanted to propose at this stage. So looking at the voltage transfer characteristics, I can only therefore safely say that I can only safely say that whenever my gate to drain whenever...

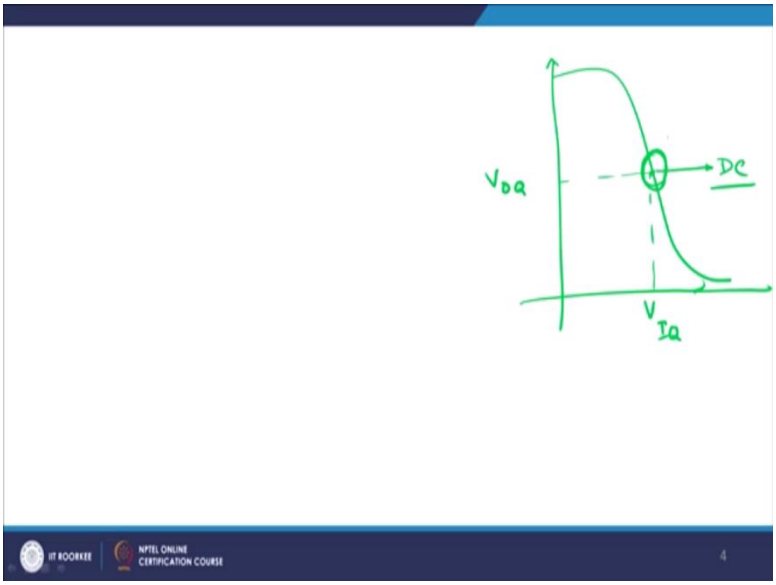
(Refer Slide Time: 10:28)




So if I want to bias the device as an amplifier it is always advisable to keep the device in the saturated region. This is the one take-away which you should do from this discussion. You place it here or here you are in a problem, you place it in a saturation region which is this, this region. You automatically get a very high gain which is available with you and this gain is

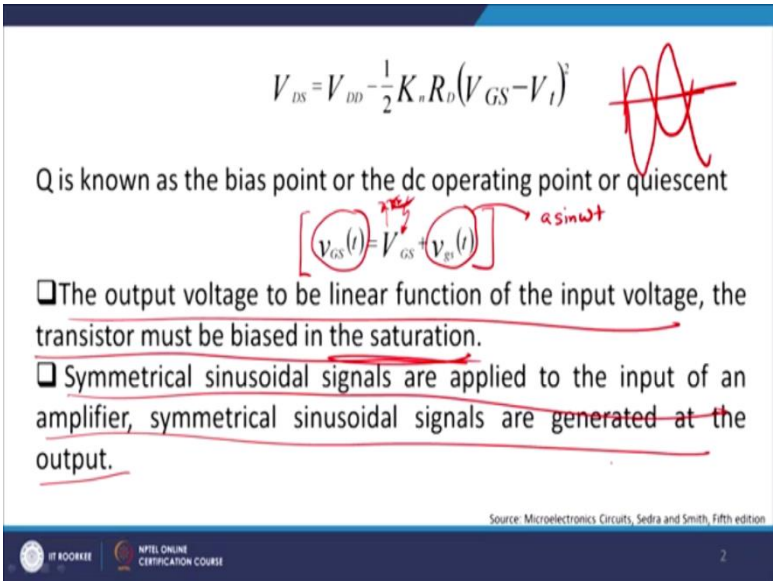
typically very high, right, of the order of few 10 to the, even $10^2$, $10^3$ voltage by voltage by doing certain manipulation of course, and you can do it. Okay, with this let me explain it to you what do I get from here.

(Refer Slide Time: 11:07)




Now what I want to tell you is that since we have understood where to bias the device, say you have biased the device somewhere as expected, you have biased the device somewhere here. So almost in the middle of the region and this your $V_{IQ}$ and this your $V_{OQ}$, so output Q input Q. So which means that you have to bias it by an external DC source, right so that initially even without an input signal it is placed somewhere in this region. Please be very careful, this is very very important that you need to first of all bias your device to sit in the saturation region before application of any input signal. That is what I was trying to tell you.

(Refer Slide Time: 11:48)



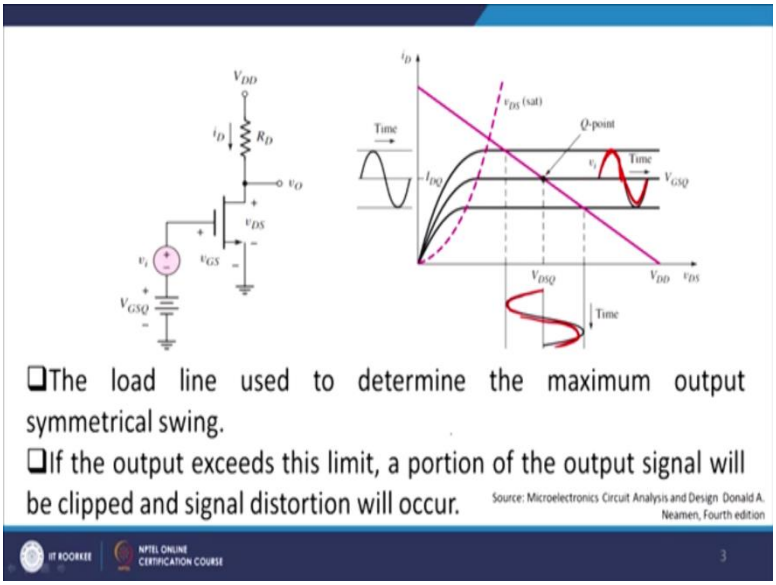

And therefore if you look very carefully this expression here, that my overall gate to source voltage which is a function of time will be the DC biased which you put $V_{GS}$, this is the DC operating point and the input voltage. So this is my time varying input voltage, maybe an Asinw, sort of input voltage and you do have $V_{GS}$ which is basically the DC bias which is fixed bias which is available.

Then you add simply to that and then you get this. So this basically a small signal output. Now the output voltage to be a linear function of the input voltage the transistor must be biased in saturation, this is the one single point which you should be very careful. Now what you do is that you apply a symmetrical sinusoidal signal to the input of an amplifier and we get a symmetrical sinusoidal output in the output side.

So you apply maybe a sine wave like this and you automatically get in the output side an amplified side wave in something like this, 180 degree phase shifted, fine. So this is one important point which you should be very carefully take care of in all due respects as far as this device is concerned, right.

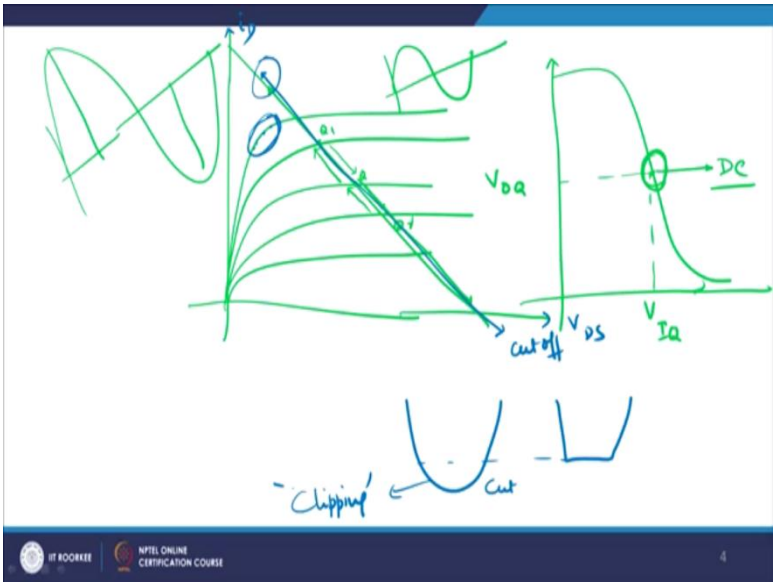
Now let us see, as I discussed with you in our earlier discussion that the load line plays a very important role, why? Because the load line tells you those points within the IV characteristics of the BJT or the MOSFET where you should bias your device for proper application of amplification, right. So you should be very careful of the load line analysis or load line as such.

(Refer Slide Time; 13:23)




As I discussed with you this is the Q point, so this is your input and input goes from like this. And your output goes something like this in the time domain and you can get sorry, this is the input which you get and this is the output which you get from the this thing.

(Refer Slide Time: 13:42)




Now the idea here is that let us suppose I was discussing with you earlier point, let us suppose you had biased the device, this is my IV characteristics of my BJT, right. This is my IV characteristics, you had put the load line something like this and you biased it here, right. Once you biased it here, well, it is pretty simple then you apply an input signal, this Q point goes to this, suppose there are large number of curves here as well.

It goes to this point, so this your Q this your Q1 and this your Q2, right. So it goes to from here from here it goes to let me say it goes to Q1, right. Then from Q1 it goes to Q again and

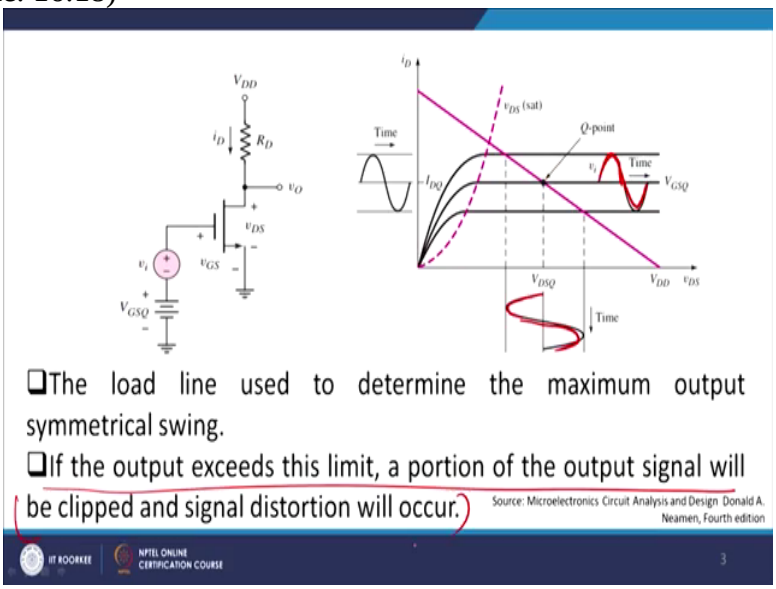
the Q to Q2 and then again Q2 to Q. So one single transition of input wave cycle will make my Q point traverse 1, 2, 3, 4, agreed? But if this signal itself is a very large signal which is this much big peak to peak right, then rather than going to Q1 it might actually go to this place or even go to this place, fine, are you getting my point? And therefore it is quite interesting or important to add that your input signal should be as small as possible or really small in nature, otherwise you force the Q point to go to the in this case into the triode region and this case if you go down into the cut-off region, in cut-off.

Which means that once you go to triode or cut-off, triode will ensure that your gain is a nonlinear amplification is there which means gain becomes a function of your voltage, please remember you can see that. See, this is current and this is $I_D$ versus $V_G$ $V_{DS}$.

If you see here, the resistance is varying if the resistance is varying obviously your gain will vary and as result if you force your device to be triode region you end up having a nonlinear amplification. Whereas if you force the device to go into cut-off, your output voltage will be clipped, so your output should have gone something like this but what happens is that because of cut-off this goes like this.

You have cut-off here, this is cut-off. So this is also known as clipping, so you have clipping taking place. So you need to ensure that your device biasing is very proper, right and that is what is the next topic which I want to come to?

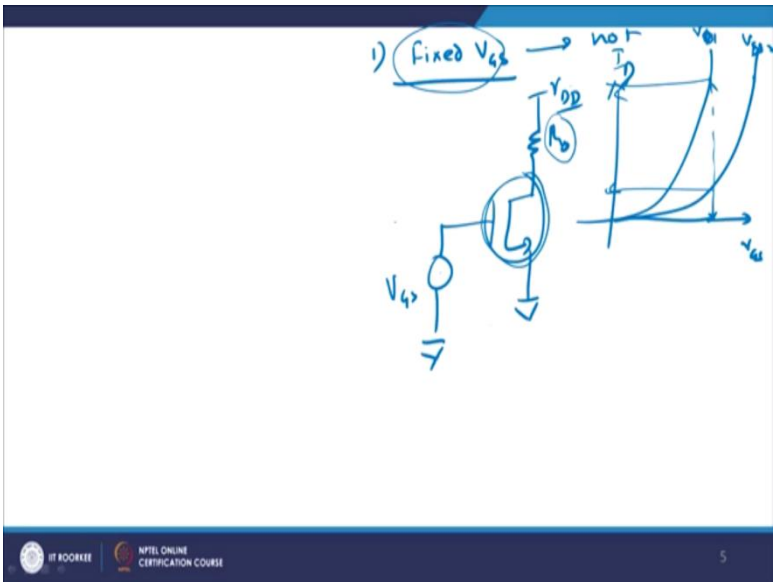
(Refer Slide Time: 16:18)




And therefore I was saying that if the input, output exceeds this limit a portion of the output signal will be clipped and the signal distortion will take place. Now this explanation of this

idea which I was discussing with you, let me explain to you the various biasing techniques which are available to us as far as MOS device is concerned.

(Refer Slide Time: 16:45)

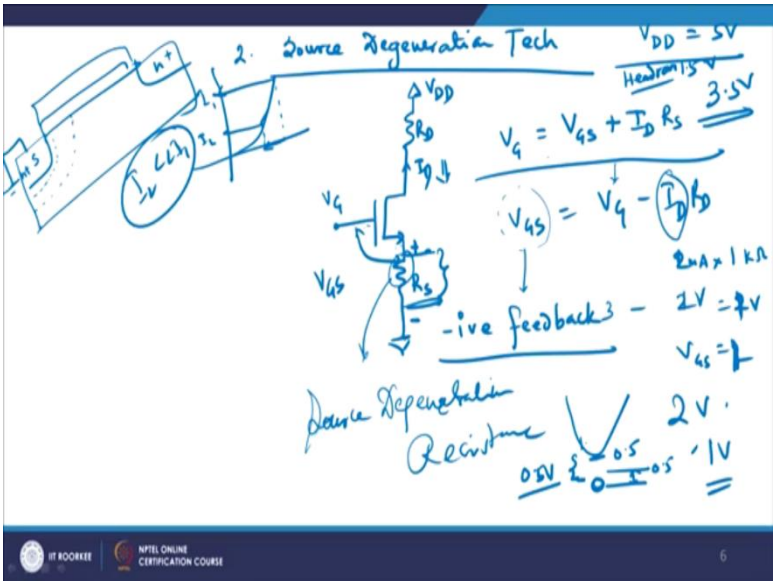



The first biasing technique which people used to follow much earlier, obviously not a very good biasing technique is what is known as a fixed, first is the fixed $V_{GS}$ technique or a fixed gate to source voltage. It is not very of course, not good which means that in fixed $V_{GS}$ what you do that you have got a drain source voltage, you do like this and then you fix up a gate voltage here and you fix up $V_{GS}$ and I was discussing.

So this is basically a grounded source application with $R_D$ here and $V_{DD}$ here, this you have been learning it till now. But this way but the way it has been done is that you fix the value of $V_{GS}$ in saturation region but the problem is that if you suppose replace this BJT by other BJT, if you replace this $R_D$ or $V_{DD}$ then chances are that the Q point will shift, so what might happen is something like this. So you have one device, so if you plot $I_D$ versus $V_{GS}$ let us suppose and then if you plot it then this will be this and maybe this.

So this is for $V_{GS\ 1}$ let us suppose and this for $V_{GS}$, $V_{DS\ 1}$ and $V_{DS2}$. So even if your drain to source voltage varies for the same values of $V_{GS}$, two devices may be so different from each other that the current obtained will be so large, so different from each other. It might also happen that for one calculation you have done properly for other one when you fix at the MOS device it forces it go in saturation region or in the nonlinear region. So fixed $V_{GS}$ technique is not a very good technique and has been often like we do not assume it to be a very good technique.

(Refer Slide Time: 18:11)




The second one which people follow is defined as source degeneration technique. This is source degeneration technique of biasing, so what I am trying do, I am trying to bias it from an external source using a DC bias, I am doing it on external source and trying to find out. Let me draw for you how it is different.

It is exactly the same as the first case, the only thing is that here we are adding an extra source resistance in series to your $R_D$ and $V_{DD}$. So I have a $V_{GS}$ of course available here. If you write it, so this is the value of $V_G$, so if you write it $V_G$, this will be nothing but $V_{GS}$ plus $I_D$ times $R_S$, this is the value of $V_{GS}$ here. This is $V_{GS}$ and this is $V_{GS}$ plus $I_D$ is the current which is flowing here, so $I_D$ into $R_S$ is the voltage drop across these two points. You add this voltage drop with this $V_{GS}$ and you get the total gate current.

So the total gate current is therefore made up of $V_G$ minus $V_{GS}$, therefore since $V_{GS}$ is almost fixed I can safely write down that $V_{GS}$ will be equal to $V_G$ minus $I_D R_D$. Now what happens is that say if you assume that my $V_{GS}$ has to vary for a fixed value of $V_{GS}$ even if $I_D$ varies even if $I_D$ varies because of some reason or other then my $V_{GS}$ will also vary in the opposite direction in order to compensate for the variation.

I will just show you how it works out. Say $I_D$ was say 1 milliamp, so 1 milliamp multiplied by $R_D$ say 1 kilo ohms, so if you multiply these two I get 1 volt, say $V_G$ was equals to 3 volt, so 3 minus 1 is 2 volts, so $V_{GS}$ was equals to 2 volts, right.

Now, let us suppose this becomes 2 milliamp, so 2 into 1 is 2 and I get this to be as 1 millivolt and therefore $V_{GS}$ from 2 volts it becomes 1 volt. $V_{GS}$ becomes 1 volt. Which means

that with increase in current, the gate to source voltage reduces and therefore it helps you to reduce the current. So, this is basically a negative feedback control, I will explain it to you just now, how?

Say current is flowing, so this is basically at higher potential as compared to this, of course. As a result this will reverse bias my source gate junction or the source, so it is something like this, on the source side it is source side and this is $N^+$ let us suppose and this is $N^+$ drain side, I have a gate here and then I have a metal function here and then I have got something like this.

Now what I do, on the source side, I apply a positive bias now, so what happens? The depletion thickness will be much larger here, if we did not apply any bias, it will be something like this. If you now apply the bias which is basically $R_D$ into, $I_D$ into $R_S$, the depletion thickness will become more stronger and therefore your current will drop down. So what happened? You increase the current right and decrease in $V_{GS}$ force the current to go down, so something like this. If you plot $I_D$ plus $V_{DS}$, your current increased, so what happens was your $V_{GS}$ was large you get a large $I_D$. You force the $V_{GS}$ to go low and it forces the current to go low here and therefore the current becomes something like this.

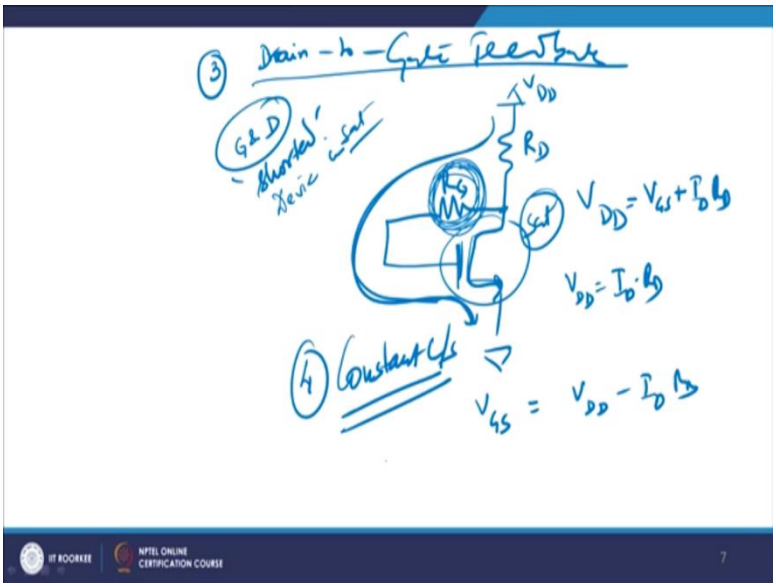
So you had I1, I2, I2 is much smaller as compared to I1 and automatically you get a much more stabilized system. This resistance is known as a source degeneration resistance. This is known as source degeneration resistance. Higher the value, better the control is, right. Let us also understand the cost you pay for it, the cost you pay for it is that as the voltage drops here, and this voltage drop can never be recovered, right and therefore the overall, so you have a headroom and a legroom. Headroom is basically, say you are you have applied $V_{DD}$ say 5 volts, right?

Now whatever you do, the output voltage cannot go beyond 5 volts, then otherwise you will be violating Kirchhoff's law and there will be clipping, right? Let us suppose the due to amplification the output voltage peak voltage goes to say 3.5 volts, then the headroom available to you is defined as 1.5 volt, this is known as headroom, fine.

Also we define the legroom, legroom is when you go to the bottom side, so if you have a current flowing through the device and your bottom should be actually grounded but you are going to a value say equals to (1 point) say, your maximum value you are going to say, the value you are going like this, this value is say 0.5, then the legroom available to you is 0.5 volts and you can play with it.

Now if you already have $I_D$ into $R_S$ drop with you then this rather than 0, this will look something like 0.5 already, so you have actually eaten away into the legroom of the device. So the price you pay for it is that you peak to peak swings are therefore restricted now for a source degeneration technique based biasing, right. So you will be very careful about this technique that though you gain something you also lose something. Two things before we move forward, okay.

(Refer Slide Time: 23:50)



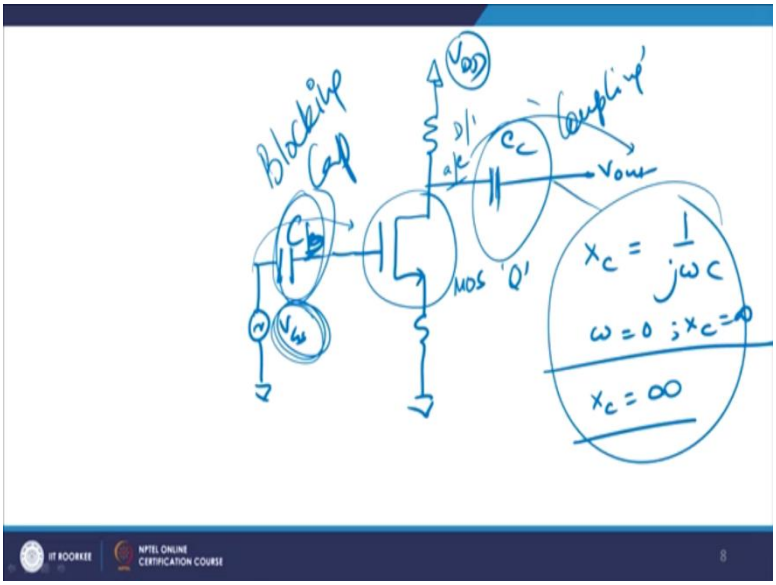

So let me go to third one, third technique for biasing is basically your drain to gate feedback technique, drain to gate feedback. It is basically drain to gate feedback. So let me draw for you how it looks like, it is something like this and you have a this, so $R_D$, $R_G$, this is $V_{DD}$ and so on and so forth. So I get $V_{DD}$ equals to $V_{GS}$ plus $I_D$ $R_D$ of course. But if you go by this technique or from this path, I guess $V_{DD}$ equals to $I_D$ times $R_D$, is it okay? And you will also get $I_G$ times $R_G$ but $I_G$ is very very small, you do not have any gate current available with you. So I get therefore, even if you follow this path the voltage drop across this will be negligibly small, right and you will get a more sort of idealistic situation.

I will get therefore $V_{GS}$ to be equals to $V_{DD}$ minus $I_D$ $R_D$, what is the advantage of having a feedback loop here which connects the gate to the drain ends? That typically now your gate and drain are bond by this $R_G$. Since $I_G$ equals to 0, the voltage drop across this thing will be negligibly small, so I could expect the gate voltage to be approximately equals to drain voltage and therefore this device will always be in saturation whatever be the input.

This is a quite interesting phenomena which takes place that since your gate and drain are shorted, then what happens is that you force the device to be in saturation for all values of

$V_{GS}$. And this is what you get. The fourth technique, I will not discuss it here is known as constant current source technique and it has been used quite often across the board but this is the constant current source technique.

(Refer Slide Time: 25: 50)




Please understand one more important point before we move forward is that whenever suppose we draw may be a source degeneration resistance based design, right, then if this is the gate voltage which you see, this is the output here, so we have a gate voltage we apply here that the input is given here and we have a gate voltage here $V_{GS}$, CC, CB we write down and then CC and this is $V_{DD}$. We have got two types of capacitors here, and these capacitors are also known as blocking capacitors CB, this is known as blocking cap or blocking capacitor and this one is known as a coupling capacitor, So, please understand that once we have this BJT biased at the saturation region, I do not want it to move away from the saturation region by external DC source.

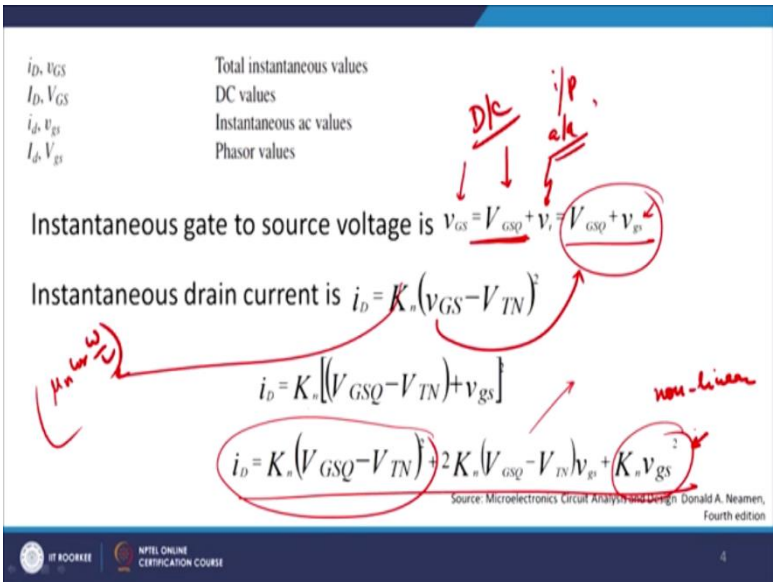
Now any power supply in this case $V_{GS}$ is a power supply, any $V_{GS}$ which is power supply will always have some DC component superimposed on the AC component, it will not be pure AC which you will get. Now that DC component should not change the Q point of this MOS device. This MOS device Q point should not be changed by that DC input of this $V_{GS}$. So what people do, people put a capacitor here. Once you out a capacitor here remember the reactive impedance of the capacitor is 1 upon jwC.

So your w when it is 0 exceeds infinity and therefore for a DC bias which is omega equals to 0 the reactive impedance is 0 and the capacitive impendence is infinity sorry and therefore it does not allow any DC source to go here and it looks very good. Similarly, when you have a

coupling capacitor then I would like to see that all the AC signal which is available here should go to this next point and all this DC bias coming from $V_{DD}$ should be stopped.

So you put a coupling capacitor, same principle you apply here as well and your $V_{out}$ therefore will pure AC and there will be no DC bias available with you. So these are the blocking and coupling capacitors which we seldom use while DC while biasing of circuitry. So please be careful when we discuss the basing of the circuitry for all practical purposes.

(Refer Slide Time: 28:27)

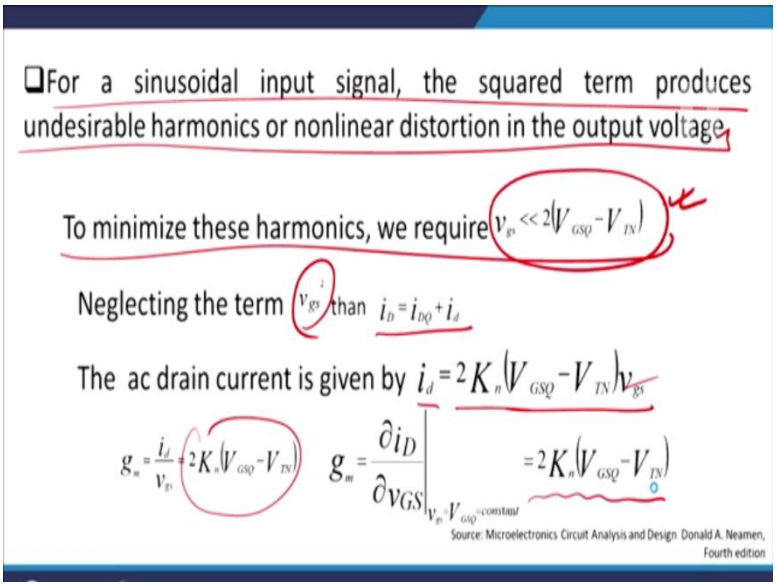



With this let me discuss with you that okay, so what happened was, so let me see therefore what happens, that now I have a $V_{GS}$, I have a $V_{GSQ}$, I have a DC bias which is somewhere in the middle. I have a DC bias here and DC bias I am overwriting the input voltage VI, this is my AC bias which I am giving and this is my total current or total voltage is sum of the DC voltage here plus the input AC.

So I get $V_{GSQ}$ because it is at Q point plus $V_{GS}$ gate to source voltage. So instantaneous drain current $K_n$ is nothing but $\mu_n$ C oxide W by L. So this is the $K_n$ value, this is $K_n$. So $I_D$ equals to $K_n$ $V_{GS}$ minus threshold voltage whole square in the saturation region and therefore $I_D$ is equal to $K_n$, in place of $V_{GS}$ minus $V_{TN}$ you just have to put this value now. So I get $V_{GSQ}$ minus $V_{TN}$ plus $V_{GS}$ whole square, you got the point? Because just in place of this I place this and I get. Now if you expand it I get something like this, this is no problem because this is we have been talking about quite a lot time. So $V_{GSQ}$ minus $V_{TN}$ whole square the current in the saturation region, no problem. This is also a no problem because current is a linear function of $V_{GS}$. Where you have a problem is the last part where the current is parabolic function of $V_{GS}$ which means that if $V_{GS}$ increases by $\partial V_{GS}$ amount that current will be parabolically

increasing. So there will be heavy nonlinearity provided you do not, you have this component of the current available with you and that makes the life difficult for overall picture.

(Refer Slide Time: 30:04)




So for a sinusoidal, input signal what I am saying….for a sinusoidal signal input signal, the square term produces undesirable harmonics or nonlinear distortion in the output side. So to minimize this distortion we require that $V_{GS}$ should be as small as possible, how small? Much smaller than 2 $V_{GSQ}$ minus $V_{TN}$. So you are into deep triode region now. So therefore if you neglect therefore in such a scenario if you left at $V_{GS}$ square I get $I_D$ equals to $I_{DQ}$ plus $I_D$. The AC drain current is given by this quantity we have already discussed this point and therefore I get $G_M$ is defined as rate of change of drain current with gate voltage which is nothing but this, because if you differentiate with respect to gate voltage this vanishes off and you get this. So $G_M$ we get two times K times $V_{GSQ}$ minus $V_{TN}$ and this is what the value of transconductance which you get. So this is what you get from the basic fundamental principle, so let me stop here because of the want of time. What we have understood today is given a MOS device where can I place the MOS device to obtain the most linear and the best saturation, first thing and the second thing is how to bias it using a DC source from an external world. From the next turn onwards we will look at the small signal model of MOS device and then we will look into the various modes of operation of MOS device, common source, common gate and common drain.

That will take care of our almost the analog part of the MOS device, amplification part at least, right? Thank you for your patient hearing. Thank you!!!

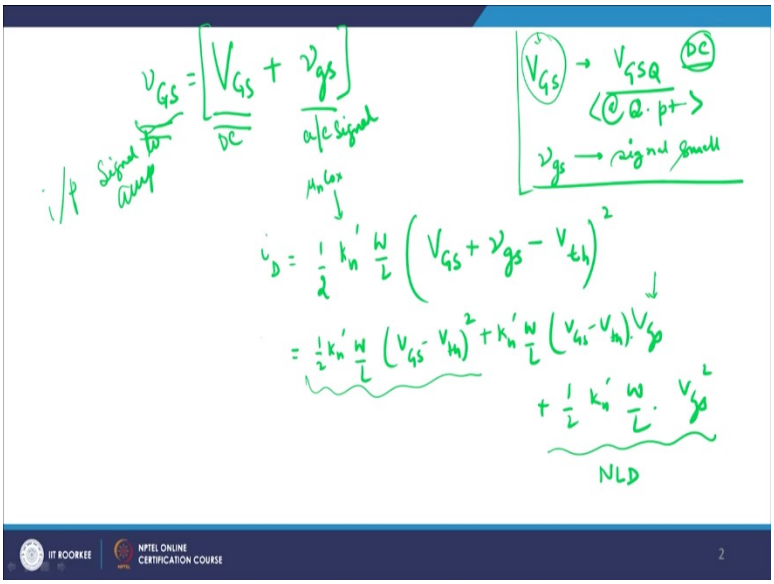
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture - 27**
**Biasing of MOS Amplifier and its Behavior as Analog Switch – III**

Hello and welcome to the NPTEL online course on Microelectronic: Devices to Circuits. In our previous interactions, we have seen that in order to amplify an input analog signal, we required to bias the MOS transistor in the saturation region of operation, which primarily means, that if you want that the amplification should be linear, which again means, that your amplification is independent of the input voltage, then we need to bias the MOS device, in the saturation region. And it should be well in the middle of saturation region, so that, even with the application of an input voltage, your output voltage does not make your Q-point, either go into the triode region or in the cut-off region, right? Otherwise, there will be a distortion in the output side.

We also saw one important point that your input signal should be as small as possible. So, that linearity is sustained for the I/O characteristics. We also kept this into mind, that while doing so, we require an external DC biasing and superimposed on the external DC biasing, we will have the input's AC signal, right? That is what we have learnt. And there are many types of biasing available to us, which we have already yesterday learned. For example, direct gate biasing. We have a feedback, drain gate feedback biasing, we have a source degeneration biasing and so on and so forth.

The best one option available to us was the source degeneration biasing, which if you, we, if we discussed in the previous turn, we saw that though your gain was reducing, but I had stabilized operation. There was a negative feedback, and therefore, whatever increase in the current was there will be reduced by virtue of a negative feedback, by virtue of the larger $I_D$ available to you. So, this was what we have learned in our previous section.

(Refer Slide Time: 2:48)




What we will learn today, is again the same continuation of my previous lecture and we will take up one important point. That now, let us suppose, we do have DC bias, which is fixing my MOS device into the saturation region, right. And let us suppose, that my saturation region, in the saturation region, let us suppose, that you do have, in the saturation region, the gate voltage is suppose $V_{GS}$, right. This is the, also defined as $V_{GSQ}$, let us suppose, which means that gate to source voltage at Q-point. So, this is at Q-point, right, at Q-point. So this is that Q-point.

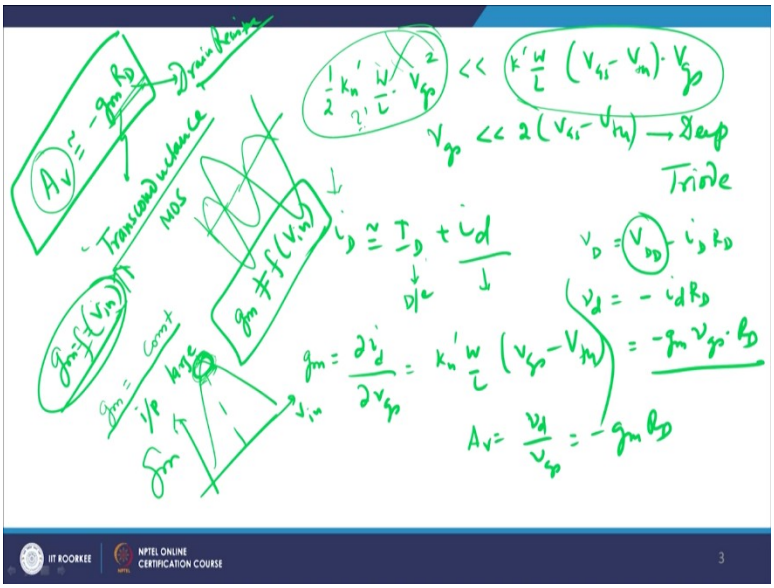
Now, superimposed on that, so this is basically a DC bias, given by an external world, right. You already have the blocking capacitors or bypass capacitor, but this gives you a DC bias, which is basically $V_{GS}$ right, gate to source voltage. Superimposed on that we will have, a small $v_{gs}$, which basically is the signal. So, this is your small signal, small signal voltage, which superimposes on the DC bias of $v_{gs}$, right. So, this is what, what we try to do. And therefore, we can safely write down $V_{GS}$ is equals to $V_{GS}$ plus $v_{gs}$, which means that, this is actually applied signal to the amplifier, is composed of one DC bias right, this is the DC bias component and this is the AC signal which you are giving. This is the actual signal, signal to the amplifier, to amplifier right, amplifier input signal, this amplifier.

Now, what I should do is, I just need to feed this value, into our drain current equation assuming that it is already in the saturation. So, I can write down Id to be equals to, 1 by 2, $K'_n$, which is $\mu_n$ C oxide, into W by. So, this is basically your $\mu_n$ C oxide, W by L, right. And then, we write down in place of $v_{gs}$ minus $v_{gs}$, $V_{th}$, we write down, $V_{GS}$ plus $v_{gs}$ minus $V_{th}$ whole square. If, you break it open, because it is a, it is basically, if you want to break it open, I will get something this to be equals to half, $K'_n$ W by L, $V_{GS}$ minus $V_{th}$ whole square, right, plus $K'_n$ W by L $V_{GS}$ minus $V_{th}$, into $V_{gs}$ plus and then half $K'_n$ W by L into $V_{gs}$ square.

Now, if you look at the first term, well, there is no problem because in any case, it is the term of the current which you will get in the saturation, saturate case, which means if the device is in saturated case, this amount of the current will surely flow, you cannot remove this. The second term, current is a linear function of $v_{gs}$, so, also, we do not have any problem. But, third term is where the problem is.

We have a heavy non-linear distortion. And the reason being, that now your input (volt) current is a square function of your gate voltage, which means that, a small change in the input gate voltage will give rise to a large change in the value of output drain current. Now, therefore, your define this to be as, a one of the major non-linear distortions, distortions available to you, right.

(Refer Slide Time: 5:48)




So, with this knowledge or with this idea, if I assume that half K'$_n$ right, W by L $V_{gs}$ whole square is much smaller than K'$_n$, W by L, $V_{GS}$ minus $V_{th}$ into $V_{gs}$, which is the second term, and then, we can write down therefore, that $v_{gs}$ is much smaller than 2 $V_{GS}$ minus $V_{th}$. So, this is known as deep triode region, deep triode.

Then, I get $i_d$, is approximately equals to $I_D$ plus $i_d$. So, this $i_D$ is the actual output current which you see, from the amplifier. This $I_D$ is the DC bias, DC current and this $i_d$ is the current, due to the signal, input signal. So, if you find out $g_m$, which is the transconductance, then it is basically $\partial i_d$, right, $\partial$ $v_{gs}$. If, you solve it, I get K'$_n$, W by L, $v_{gs}$ minus $V_{th}$, right. So, this is nothing, no problem because you need to just differentiate this problem.

This is gone because this is very, very small. See, see where are you operating? You are operating at very, very low value of $v_{gs}$, right. And therefore, when you squared it, this term can be negligible, can be neglected, can be removed. And therefore, this term comes out to be K'$_n$, W by L, $v_{gs}$ minus $V_{th}$, right. Similarly, you all, we have already seen that $V_D$ is equal to $V_{DD}$ minus $i_d$ $R_d$. And therefore, $v_d$ can be written as minus $i_d$ times $R_d$. Why? Very simple because what are you doing is that, you are replacing $V_{DD}$ by output voltage, shorting it.

And therefore, I can right down this to be as minus $g_m$ times $v_{gs}$ times $R_d$, right. This equals to $v_d$. So, what is our gain? Gain is equals to $v_d$ by $v_{gs}$, source to drain, is equal to minus $g_m$ times $R_d$. So, this is what we get from all this discussion. That the overall gain for any amplifier is approximately equals to $g_m$ times $R_d$, where gm is basically known as the transconductance, right, transconductance of the device, which device? The MOS device. And $R_d$ is the drain resistance.

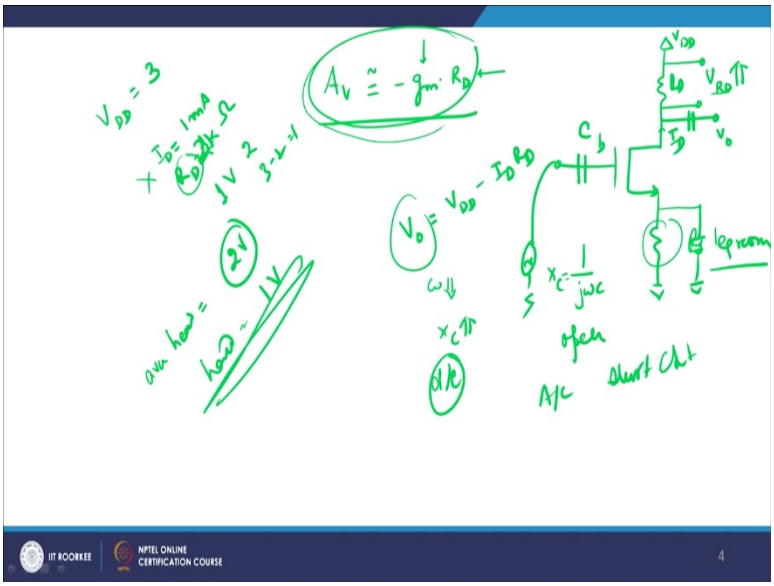
So, $A_v$, the minus sign, shows you that it is basically a, it is basically a 180 degrees phase shifted. So, if your input is something like this, your output will be something like this, right. And therefore, this is known as 180 degree phase shifted, as long as this is a small signal model, anyway.

Now, this assumption that, $A_v$ equals to minus $g_m$ times $R_d$, holds good when your $g_m$ is typically, is considered to be independent of the value of $V_{in}$. In reality, you will see that, the transconductance $g_m$ is again, in second order, we are not dealing it here. Second order, $g_m$ is actually a strong function of $V_{in}$. We will learn that later on.

So, whenever you do a large signal modeling, so whenever you do a large signal modeling, your $g_m$ actually increases or decreases, depending on the value of $V_{in}$ which you are choosing, which means that, this transconductance value, which we have assumed to be constant, and therefore, $g_m$ into $R_d$ which is assumed to be constant, equals to $A_v$, will not hold good, when, when, when your input signal is having a large signal, right. Your $g_m$ then starts to be a function of $V_{in}$.

Till now, we were assuming that $g_m$, is not a function of $V_{in}$, right. But in reality, we will see later on that, if you are having, if your $V_{in}$ variation is very large, then gm becomes a function of $V_{in}$. Typically, if you plot, $g_m$ versus $V_{in}$, it is something like this, right, something like this. If you plot, $g_m$ versus $V_{in}$, right, very small region it is peaking and then it is dropping down drastically. We try to keep our input voltage in this range and therefore, we get the maximum gm available with us, right.

(Refer Slide Time: 10:19)




So, with this knowledge, we therefore, know that, for any amplification to occur, or to any amplification to be available to me, I will have, the voltage gain $A_v$ to be approximately equals to $g_m$ times $R_d$, where $R_d$ is applied drain voltage, resistance. So, there are two ways of increasing the amplification of CA stage or any stage amplifier using MOS device. You either increase the value of gm or increase the value of $R_d$, right. But, let us see the cost; you will pay for, increasing the value of $R_d$.

Let us suppose, so your $R_d$ was this one, right, and this is your MOS device, which is there and let us suppose you have source resistance at this point and this was $V_{DD}$, right. So, as I discussed with you, V zero, which is this one will be nothing but $V_{DD}$ - $I_D$ $R_D$, where $R_D$ is this one and drain current $I_D$ is flowing in this manner.

Now, the point is, therefore, the V zero is given. Now you see, if you, if you would increase the value of $R_D$, because you want a higher gain, you then, ensure that the drop across $R_D$, which is let us suppose $V_{RD}$ is therefore, will increase. And therefore, the available headroom to you will reduce, right. So, let us suppose, your $V_{DD}$ is say, 3 volts, right. And your $I_D$, $I_D$ $R_D$ is, $I_D$ is 1 milliapm and $R_D$ is 1 kilo ohms, right. Then, I get this; if you multiply these two, I will get 1, so you will get 1 volt, and therefore, 3 minus 1 volt is 2 volt.

So, available headroom to you is, available headroom is, headroom is 2 volt. If you increase $R_d$ to say 3 ohms or maybe 2.5 ohms, 2 ohms, then, you will get, 2 into 1 is 2 and therefore, 3 minus 2 equals to 1, and I will get only 1 volt as a headroom. So, I can do only manipulations up to 1 volt so, my headroom gets reduced.

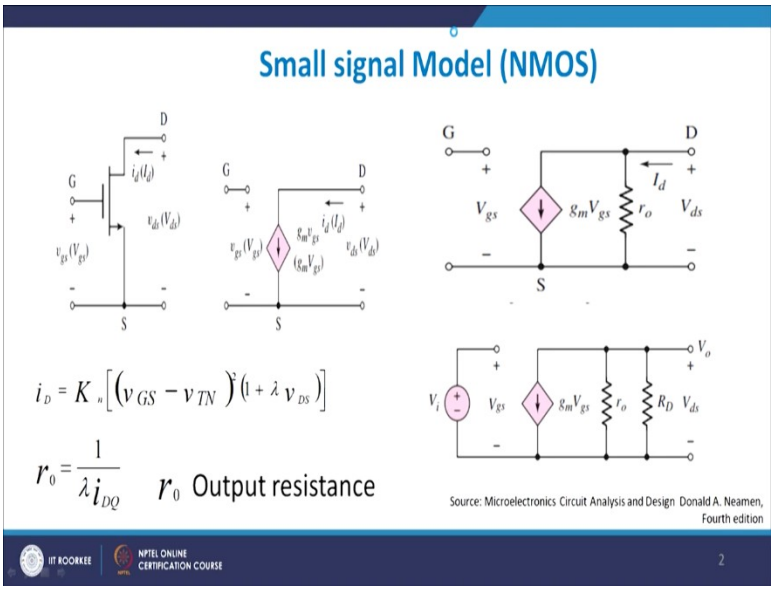
Similarly, if you go on increasing $R_s$, your leg rooms, leg rooms gets reduced. So, the cost you pay, for a higher gain is offset by a reduction in your head rooms, for an analog design. So, please understand, this very interesting and important property that do not think, by simply increasing the value of $R_D$, you will be able to achieve a larger gain, always. You will be able to achieve, but the cost you will pay, will be too much. So, therefore, it is always advisable to restrict yourself to very small signals of inputs, right, very small signals of input, right, this one.

Now, second thing, as I discussed with you yesterday, that this, I have a blocking, I have a blocking capacitor here, right, also I had a coupling capacitor. I had blocking capacitor here as well, right. Coupling or blocking, whatever you want to name it, you can name it. But, you can remember yesterday's talk, last part of the talk was that $X_c$ or the trans, $X_c$ means, reactive capacitance, of this one will be 1 by jwc.

So, when, w is very, very low, $X_c$ is very, very large, which means that for a DC bias, this acts as a open circuit and, for AC bias, right, this act as a short circuit, capacitances. So, whenever, we are doing AC analysis, short, that is the reason, we say that, when we do AC analysis, please short all your capacitances, right. When you do a DC analysis, please open all your capacitances, so that it acts as an open circuit.

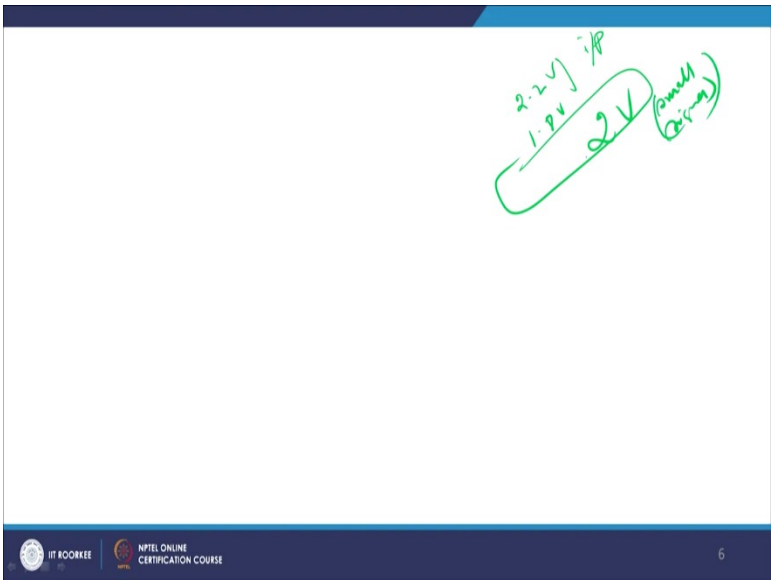
Now, with this idea or with this, we have therefore, recapitulated one point, that gain is this one. But, then gain depends also on $g_m$ as well as $R_D$. Higher the value of $R_D$, more is the gain, but less is the headroom available to you to work with,

(Refer Slide Time: 13:59)




With this knowledge, let me come to, very important topic that is known as a small signal model of a device. So, we have understood NMOS, NMOSFET and let us, therefore, take the NMOSFET and see, if we can do small signal model, means by the input voltage is varying by very small.

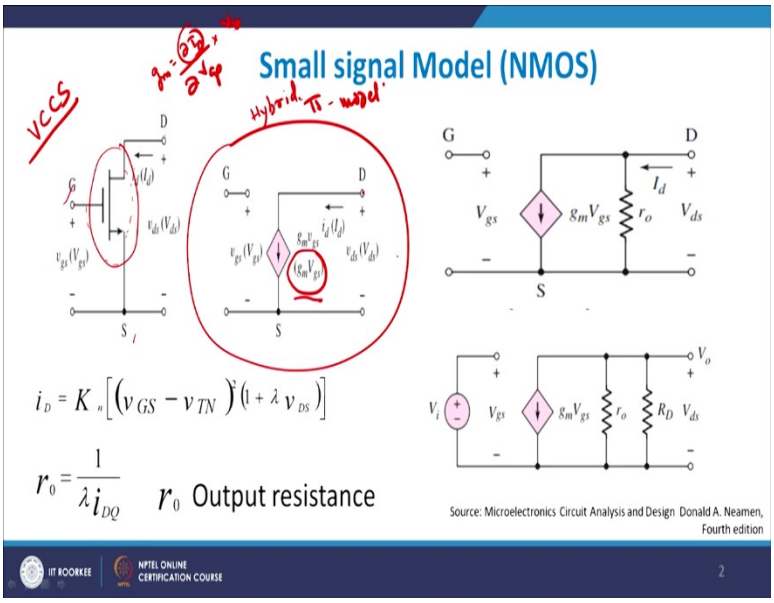
(Refer Slide Time: 14:20)




Typically, when we say small signal, there is no quantitative analysis as such, but typically what people expect is that, a plus minus, 10 percent, from Q-point, from Q-

point. That is will be defined as a small signal. So, if your Q-point is somewhere around 2 volts, suppose it is 2 volts, 2 volts, then it will be 2.2 to 1.8 volt will be the input swing. Then, we will define this to be as a, 2.2 will be defined as small signal. So, this will be defined as a small signal, right. So, I will have small signal, available with me, provided, this is true or this functionality is true.

(Refer Slide Time: 14:57)




Now, let me come to you, the small signal model of the FET device. If you look here carefully, this is the N-MOSFET, which is in front of you, right. I have a current flowing which is a drain current. $V_{ds}$, is the drain to source voltage, which you see here. And you have a gate to source voltage, which is available at this particular point. So, it is basically a three terminal device gate, drain and source.

And therefore, as you can see, it is basically a voltage controlled current source. So, it is basically a $V_{CCS}$, right, it is basically a voltage controlled current source. Who is controlling, controlling this current, $I_d$, is $v_{gs}$. So, therefore, this is also referred to as a voltage controlled current source, right.

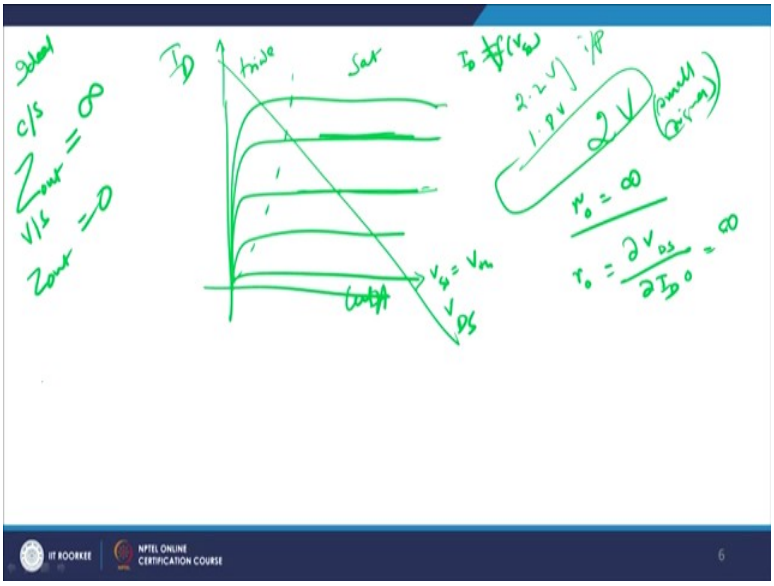
Now, so what we can do, simply is, I can, I can, make this transistor or this MOS device, as a ideal current source. And therefore, I can simply write down this to be as, a current source with $g_m V_{gs}$ as the current value, because, that is what we know. $G_m$ was referred to

as $\partial I_d$ $\partial V_{gs}$, remember, right. So, If, you multiply this with $V_{gs}$, I get simply, $I_d$, and therefore, this gives you the drain current.

And, therefore, on the drain side, I refer to this as $g_m$ times $V_{gs}$ and the gate is open. Why the gate is open, shown as open? Because please understand, it is basically a voltage source. A voltage source, in its equivalent form will always be open circuit because, a voltage source by definition, means that, its output impedance is zero, right. Its output impedance is zero, and will give you the same voltage, even if you draw any, any current, right.

So, therefore, a voltage source is always shown by an open circuited and a current source is, its equivalent form, if you want to find out, will be always shorted right; whenever you want to find out its equivalent form. So, I get this, so, this was my, this was my model. This is also referred to as a hybrid, hybrid π model, right. This reason, for the hybrid π, why, it is known as hybrid π model. But it is also referred to as a hybrid π model. So, we have understood why.

(Refer Slide Time: 17:12)



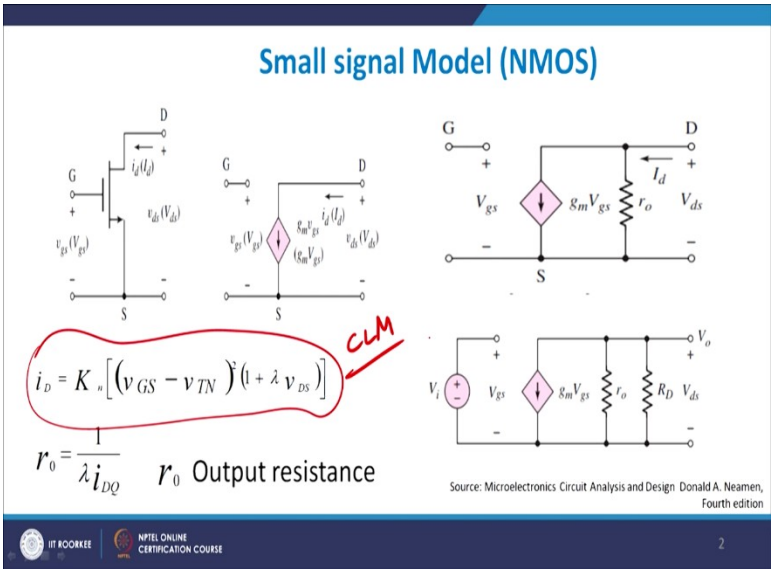

Now, let us understand, this was with the assumption, that my I-V characteristic of the MOS device is perfectly straight in the saturation region, which means that, it is absolutely, a perfect current source, in the saturation region. So, this was the saturation,

by definition this is saturation, this is triode and then if you go, like this, this is $V_{gs}$ equals to $V_{th}$; this is cut-off.

So, you are biasing your device somewhere in the middle, by drawing the load lime, you are biasing somewhere here. But, then I am assuming that, this saturation current source is basically, a fully straight line. So, this is $I_d$ versus $V_{ds}$. So, when this is a perfectly a straight line, I assume that, therefore, Id is not a function of $V_{gs}$, right. It is not a function of $V_{gs}$; it is a constant current source. If, that is so, what happens to your output impedance, will be infinitely large. Because, $r_o$ or output impedance, will be referred to as $\partial$ V, $V_{ds}$ by $\partial$ $I_d$.

Now, if you travel from this point to this point, even if you are varying in $V_{ds}$, your $I_d$ is still, $\partial$ $I_d$ is still zero. So, this is zero, R zero will be infinity large, right. So, therefore, a current source, therefore, let me write down a current source, an ideal current source will have, its $Z_{out}$, which is output impedance, to be infinitely large. A voltage source, a ideal voltage source, $Z_{out}$, will be equals to zero.

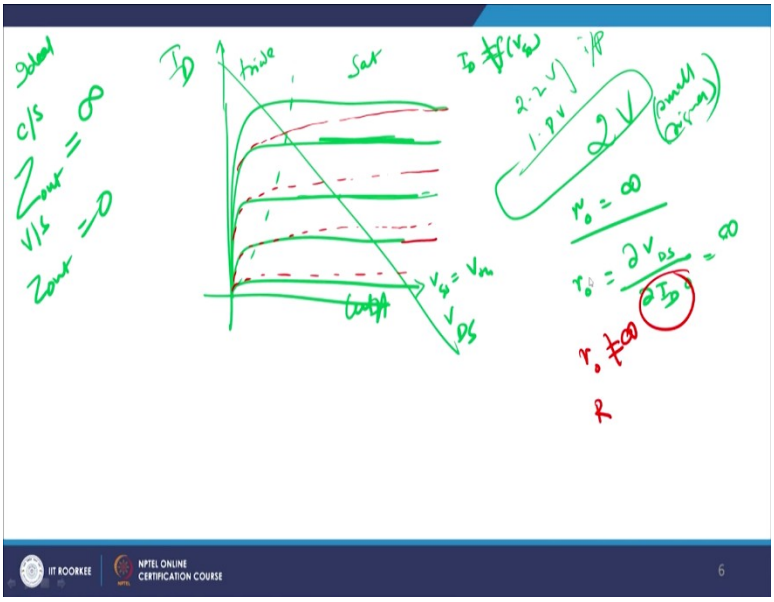
(Refer Slide Time: 18:52)




But, in reality as we have discussed in our previous turn or in our previous understanding, that $I_d$ is actually depending on the value of $V_{ds}$, in the saturation, also, known as Channel

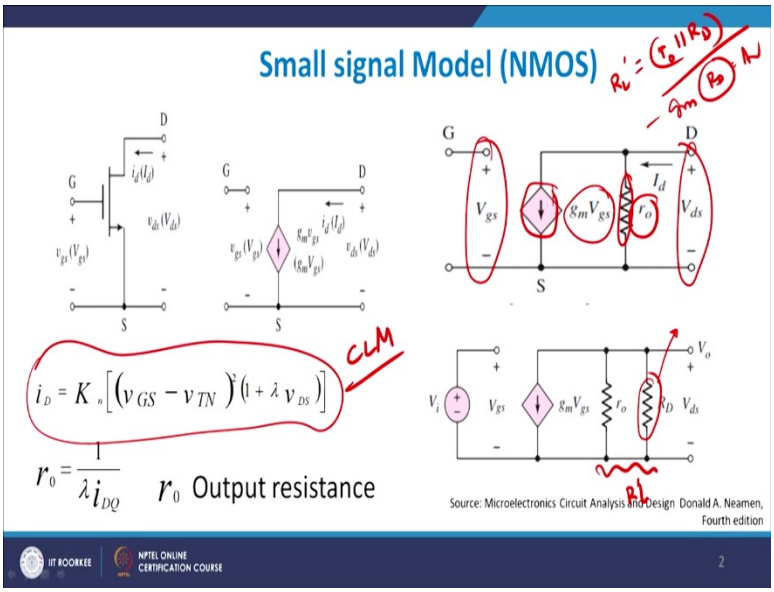
Length Modulation phenomena, right. We have discussed this point, when we were discussing MOS device physics as such.

(Refer Slide Time: 19:21)




And therefore, due to Channel Length Modulation phenomena, what happens therefore is, that rather than this remaining, like perfectly, perfectly straight or perfectly constant, I actually start getting a small rise in the current, with increase. So, with increasing the value of $V_d$, $V_{gs}$, right, $V_{ds}$, I start getting a more and more current with me, right. So, this is higher values of $V_{gs}$, which means that now; since this is not a straight line and you do have a finite $\partial$, you do have a finite $\partial I_d$, $I_d$, therefore, $r_0$ will be, not, will not be equals to infinity, but will have a value, right, but will have a value; which in this case, we have considered to be $r_0$.

(Refer Slide Time: 19:56)



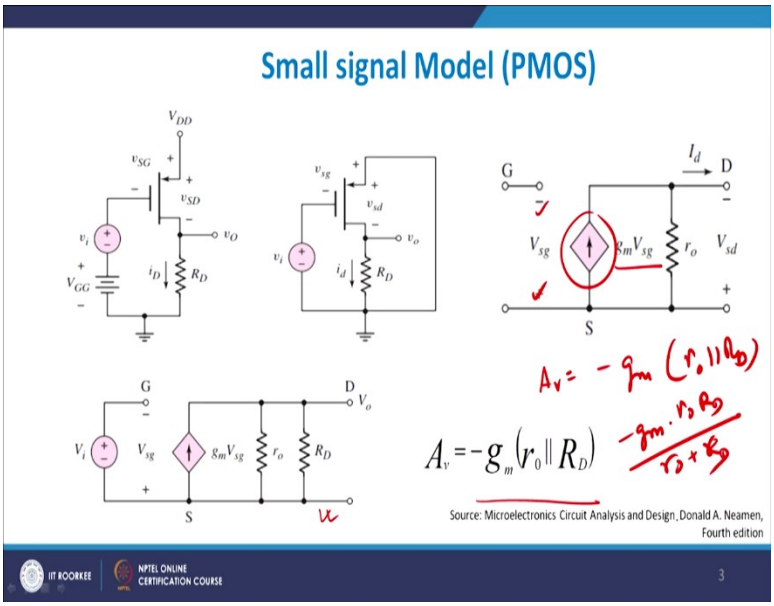

So, I just take $r_0$, in parallel to the current source here, where, $r_0$ is basically, the resistance offered by the device, when the device is operating in the saturation region, fine. And therefore, there will be a current $I_d$ flowing here, there will be $V_{ds}$ flowing here, and this $g_m V_{gs}$ is current.

So, if you, if you treat, gate, drain and source as three terminal gate, will be obviously open circuited, as we discussed. Drain, will be open circuited, but the drain will be terminating on to a current source, whose value is $g_m (V_h) V_{gs}$, followed by a resistance, which is in parallel to the current source, to prove that Channel Length Modulation phenomena is taking place, right.

Apart from this, you will also, automatically, have the drain current resistance, drain resistance, right. Why drain resistance? Any external resistance which you will be putting will be defined as drain resistance and you will always have the drain resistance, in the output path, right. And, this will give rise to a, change in the value of your, if your, output impedances. And this is what we learned from; we will discuss this in detail.

We will see that, since therefore, I can safely say, $R'_L$, let us suppose $R'_L$ is basically this, $R'_L$, right, $R'_L$. Then $R'_L$ is $R_O$ parallel to $R_D$, right. And therefore, only less than the least resistance will, will appear with us, right. And therefore, only $R_D$ will appear, because $R_0$ is typically very large, $R_D$ will appear. And again, $g_m$ times $R_D$, will be your $A_v$ value, which you get for a small signal model, right.

(Refer Slide Time: 21:31)

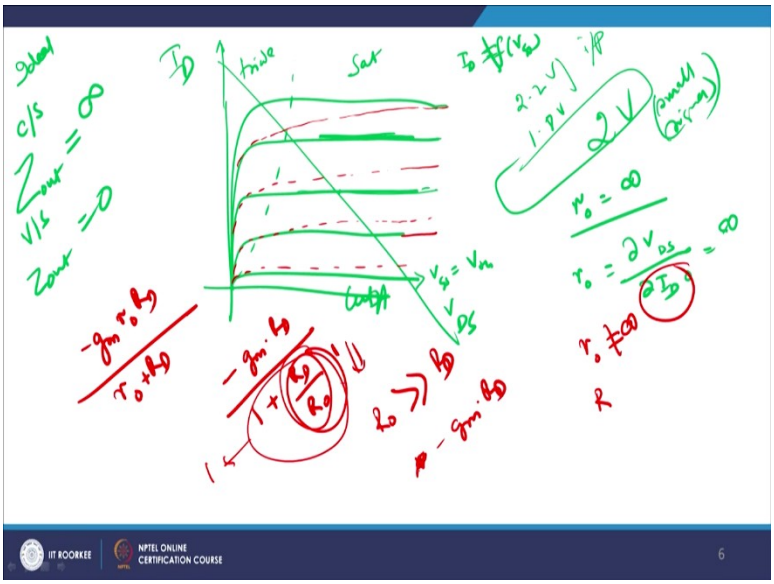



Let me explain to you for the PMOS model. The same thing, the same thing; though, just let me remove this, is exactly the same thing, but you have to see, that the current direction is just switched. So, you had a, sorry.

So, the current direction was initially this side, down and now if you look back, the current direction, is up. And, the gate to source voltage, also has changed between, positive and negative to, to negative and positive, right and therefore, we get $V_{sd}$ to be equals to $g_m$ times $V_{sg}$ into $r_0$, you will get.

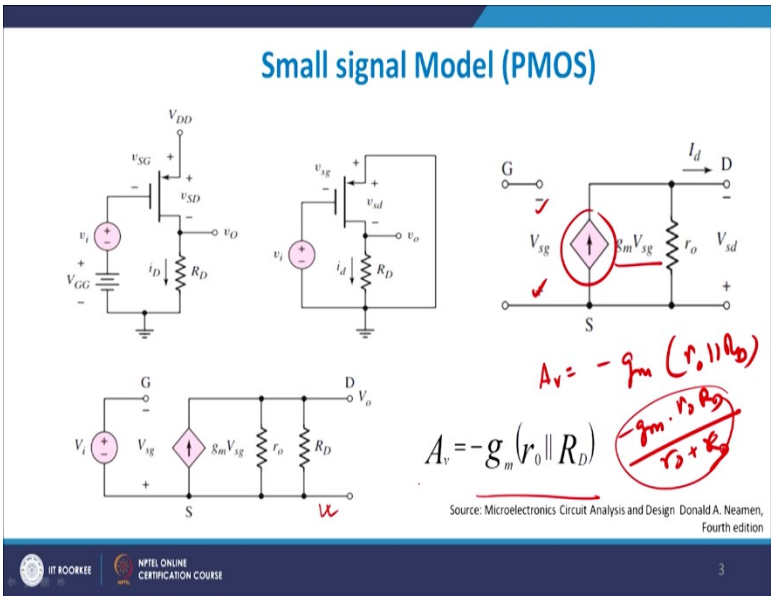
Similarly, if you put, a drain voltage outside, I get minus $g_m$ times $r_0$ parallel to $R_D$. So, I get $A_v$ to be equals to, minus $g_m$ times $r_0$ parallel to $R_D$. So, if you solve it, I get minus $g_m$ times $r_0$, $R_d$ upon $r_0$ plus $r_d$, fine, you get this as your final value. So, I get, minus; so, I get what?

(Refer Slide Time: 22:29)



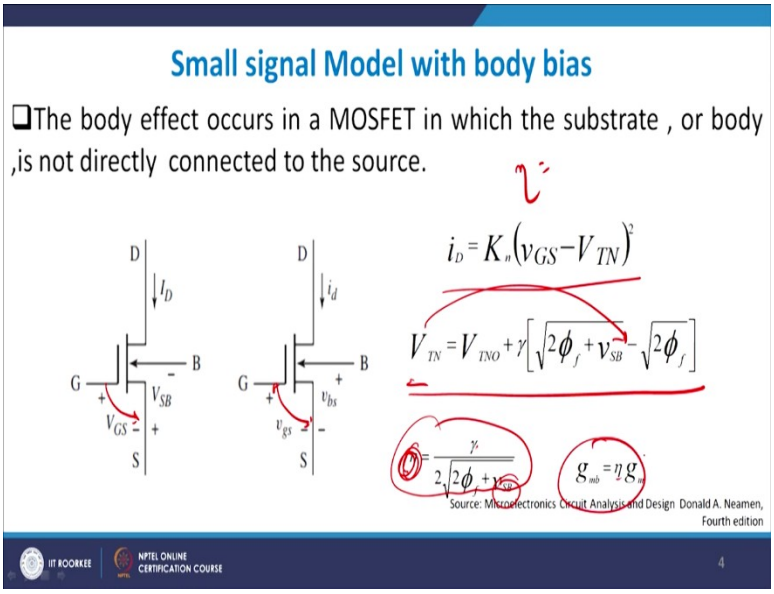

So, I get minus, minus $g_m$ r $_0$, $R_D$ divided by r $_0$ plus $R_D$. So, if you divide by say, $R_D$, numerator and denominator or r $_0$ by denominator and numerator, I get $g_m$ $R_D$, divided by 1 plus $R_D$ by $R_O$. Now, if your $R_O$ is much, much (large), smaller as compared to $R_D$, then this, this whole quantity will be equal to 1, right, If, $R_O$ is much, much larger as compared to $R_D$, then, this whole quantity will be much smaller, very low. And therefore, 1 plus very low quantity, will be approximately equals to 1. And therefore, I can safely write down this to be, equals to $g_m$ times $R_D$ with a negative sign, right.

(Refer Slide Time: 23:16)




And, this is what we get here, that, $A_v$ will be minus $g_m$ times $r_0$ parallel $R_D$, where $R_D$ is the applied drain resistance which you see.

(Refer Slide Time: 23:21)

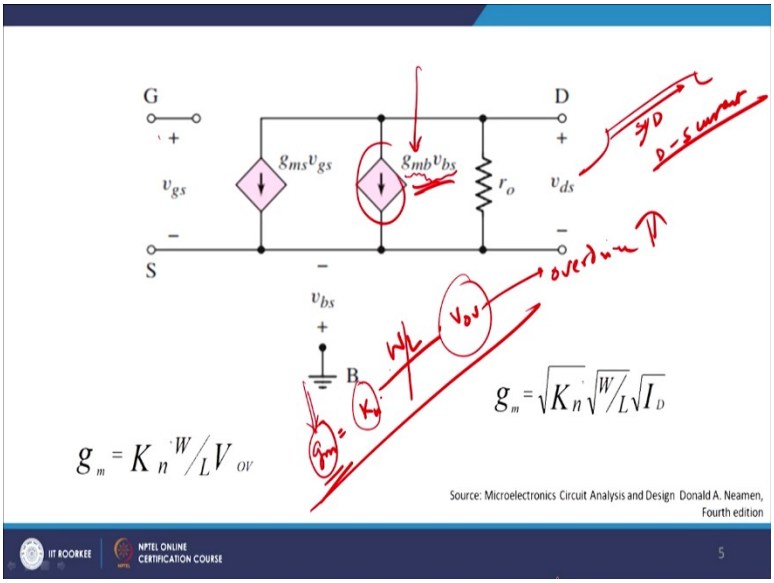



Now, let me also discuss with you, that if you remember, that a MOS transistor is not always, three terminal device. It is actually, a four terminal device, with substrate being the fourth terminal, right. This, we have already discussed earlier, that is known as body effect, right.

So, I had discussed with you, that $i_D$ will be equal to $K_n$ times $v_{GSN}$ minus $V_{TN}$, $v_{GS}$ gate to source of N device, gate to source of N device, this, minus $V_{TN}$, threshold voltage of N device. Now, threshold voltage of n-channel MOSFET is itself written in this manner, which means that it is a strong function of $v_{sb}$. And therefore, I can write a new term $\eta$ to be equals to this; so there is a derivation, I am not doing it, in this interaction. But, I am doing a derivation, which tells me that $\eta$ will be equals to $\gamma$ upon 2 $\varphi f + v_{sb}$.

So, if you vary $v_{sb}$, my $\eta$ will change and as a result, you will expect to see $g_{mb}$ to be n times, $\eta$ times $g_m$. And therefore, on depending on the value of $\eta$, the $g_{mb}$ means, transconductance with bulk, with bulk potential and $g_m$ is transconductance only with the gate potential. So, I will see a change, in this case.

(Refer Slide Time: 24:36)




So, what I am done, therefore is, if you remember correctly in your precious discussions, my current is flowing in this direction, for source to drain, right, source to drain current is flowing, So this, so this is, the gate, drain and source. So, if you remember, this $g_{mb}$ times $v_{bs}$, is nothing but the, again drain to source current, right, drain to source current. Because, there is no other way, the current can move.

Because, you are applying bias from the drain side, your source is grounded. So, all the charge particles generated, assuming, there is no bulk as such or the bulk potential is

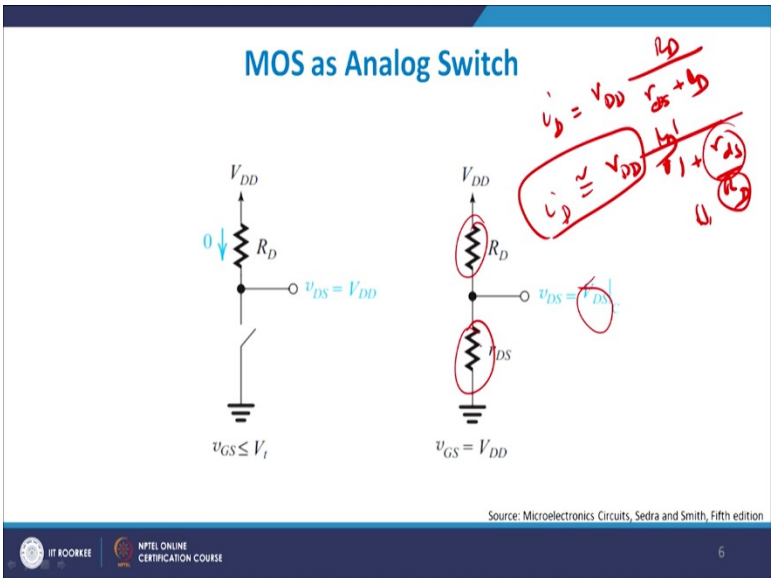
relatively smaller or no change, then I will expect to see it, moving across source to bulk. That is what is happening here, I get $g_{mb}$.

So, $g_{mb}$ is defined as a transconductance of the device, with substrate bias, right, substrate bias. Now, $g_m$, as I discussed with you will be referred to as $K_n$ times WL, right, W by L into V overlap. So, $K_n$ times W by L into V overlap, is actually equals to $g_m$, right. So, if you want to increase the value of $g_m$ one is that straightforward, W by L ration, you make it higher. As you make my, W by L ratio higher, I get a larger current, and therefore, I get a larger gm for the same change in value of $v_{sb}$.

Second thing is, if you try to increase the overlap voltage, V overlap, right, or V, V over or even $K_n$, I should also get a larger value of $g_m$, right. So, your overdrive voltage is a large, I will expect to see $g_m$ to be also very large, right. And, as a result, higher the overdrive, higher will be the transconductance of the devices itself.

So, let me take up MOS as analog switch, where we have already discussed, this point in detail, in our previous discussion. But, let me just; still reframe the whole network for you. As, I discussed with you, whenever my input is basically more than threshold voltage of the device, the device turns on, and, as a result, there is a direct path between $V_{DD}$ and the capacitance. And, the capacitance gets to, starts to get charge and the charge, and the output voltage, goes from 0 to 1, right. Similarly, in the reverse bias, when your input voltage is negative, PMOS switches on, and output voltage, will therefore, go from 1, 0 to 1, because it is moving towards $V_{DD}$.

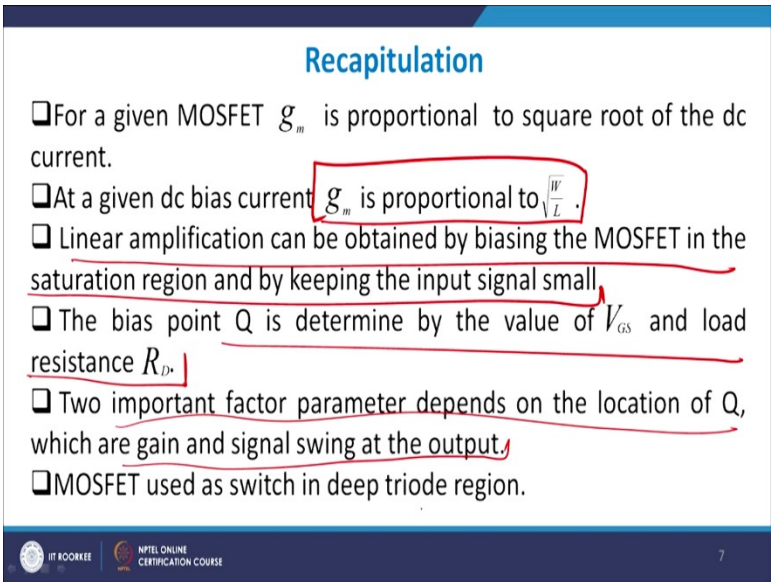
(Refer Slide Time: 27:03)




Now, assuming that, you are doing; assuming that you are having something like this here, then I can safely say that $i_D$, right, is basically equals to $V_{DD}$ into $R_D$ upon $r_{ds}$ plus $r_D$, fine, voltage divider. So, $V_{DD}$ divided by 1, if we divide by 1, 1 plus $r_{ds}$ by $R_D$, right. If, your, $R_D$ is very very small as compared to $r_{ds}$, right, this whole quantity can be neglected, and $V_{DD}$, $i_D$ will be approximately equal to $V_{DD}$, approximately equals to $V_{DD}$, multiplied by $R_D$, multiplied by $R_D$.

So, if you see very carefully here, MOS is an analog switch, if you look very carefully, I have an $R_D$ value and $r_s$ value, coming into picture and therefore, $v_{DS}$ drain to source voltage, is equals to $V_{DS}$, in the center frequency. And, $r_{DS}$ is the resistance. So, it is basically known as two resistance model of a MOS device, right, and it depends upon the value of current flowing through this terminal, and value of voltage, at this particular point, because of this current, right.

(Refer Slide Time: 28:06)




So, let me therefore, recapitulate what we have leaned till now, that transconductance is always proportional to square root of dc the current. This we get, in another means, but do not worry about it. But, transconductance primarily depends upon value of mobility W by L ratios and the square of the, square root of the dc current. Now, at any fixed dc value of current, $g_m$ is always proportional to W by L root over, right, and this is an expected or an accepted fact across the world.

Linear amplification can be obtain by biasing the MOSFET in the saturation region and by keeping the input signal small. This, we have already discussed time and again, to remove the nonlinear distortions. Now, how you determined the value of Q-point? By using $V_{GS}$ and $R_D$. $R_D$ is the external parameter and $V_{GS}$ is also the external parameter. But, how it switches on the MOS transistor, is basically the internal parameter. And, therefore the bias point Q, is determined by the value of $V_{GS}$ and load resistance $R_D$, fine. And, it is the standard method of finding out the Q-point, technically.

Two important factors, decide the location of Q-point- what is the gain, what is the signal swing at the output, you got the point. Because, if your Q-point, is very close to the, either the saturation or cut-off, then even in the small change in the input voltage, I would expect to see, somewhere cut-off in your output voltage, And, as a result, it is not a good idea to have cut-off voltage available to you, right.

If, you want to use MOSFET as a switch, please bias it in the triode region. So, that is very, very important. You bias it, in deep triode, right, so, this will make you off, right. Is it alright? So, it will make you off and that gives you very good idea of the MOSFET can be actually used as an analog switch, right. And $\tau$, as you all know, 0.69 R times C, is my typical, this thing, delay element for a switching diode, or a switching MOS device.

With this, let me finish, today's lecture and thank you, for your patient hearing. The next turn, we will look into the other facts, as far as this course is concerned. Thank you very much!!!!!!

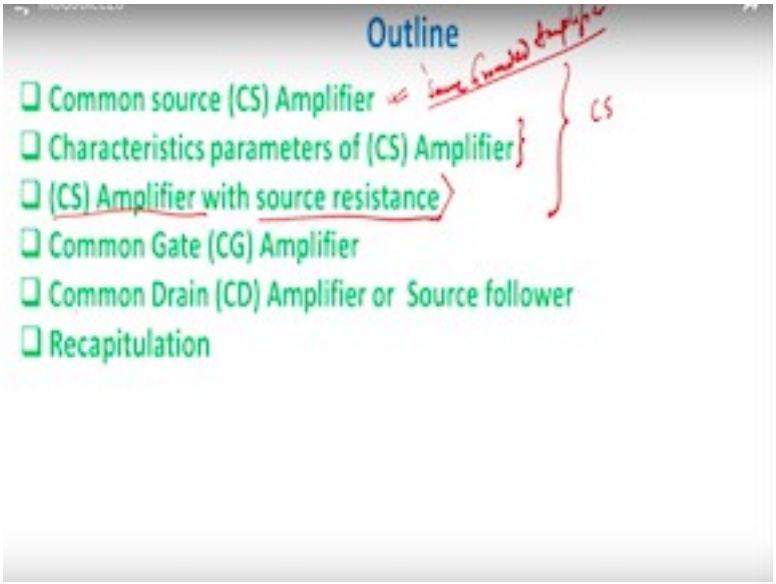
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture - 28**
**CMOS CS/CG/CD Amplifier Configuration**

Hello and welcome to the NPTEL online course on Microelectronics: Devices to Circuits. In our previous file or previous have understood what is known as the small signal model of the MOS device; wherein he also understood as π; hybrid π model and what is a T model.

So for both MOS device why was it important? Reason was that whenever you encounter a MOS device in a circuit you can replace it by the equivalent circuit model and then calculation of various gains and output gain, current, impedances becomes much easier. We have also seen that in a previous example of the previous slide that your MOS device is basically a voltage control current source, right. And therefore by using an external voltage source, which is gate to source my drain current flowing through the device can be changed, right.

What we will see today is basically the various configurations of amplifier and therefore the name of the topic of today's discussion is basically CS CG CD amplifier configuration, CS primarily means common source, CG means common gate and CD means common drain. They are also known as source grounded, gate grounded and drain grounded amplifier configuration. So these are all actually amplifier, which is, which is there with us.

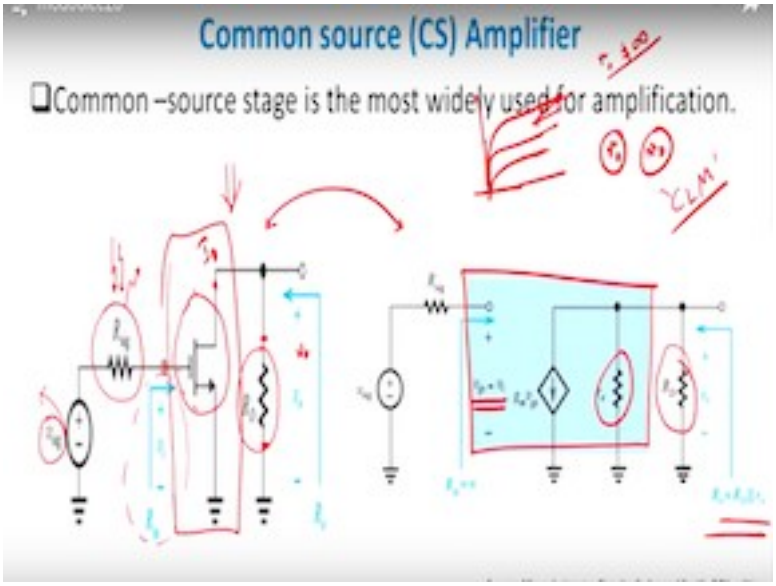
(Refer Slide Time: 2:05)




So let us see the first what will be the outline of my talk. The first thing, which we will be starting basically my common source amplifier, which is basically also known as source grounded. This is the most commonly used amplifier design across the world and for reasons we will discuss later on. We will see the various characteristics of this amplifier, which means we will be looking into the various output characteristics of this amplifier. As we have seen in our previous discussion if we apply a source resistance, right, source resistance, how does my CS amplifier change or how does its output characteristics change.

After we have understood these three important points regarding CS we go to common gate and then source follower or common drain, right. Common drain also referred as source follower and there are certain reasons, why we study CG and CD, but the most commonly used amplifier is basically the common source amplifier, right.

(Refer Slide Time: 3:02)




Common source amplifier is this, this is the diagram of a common source amplifier, right. This is the diagram of a common source. You see here you have a NMOS device which is this one active device, this is your output resistance of the input signals source, so this is the signal source which is there with me, right.

So assuming that I have already DC biased it, which means that I have already fed these devices with an external $V_{DD}$ supply and source supply and biased it at saturation region so that its works as amplifier. This V signal is a small signal input voltage, which I am giving to the MOS device and R signal is basically my resistance offered by the voltage source. So a voltage sources please understand output impedance is very low, but in reality you might get some amount of resistance offered by it and that is given by R signal, right. So R signal therefore comes in series to V signal and V signal is the input voltage, which you give.

Then this $r_D$ is the load resistance on external voltage resistance, which you see on the external world. So you see the effective $V_{in}$ showed in this blue diagram in this blue curve is basically the $V_{in}$ is the input voltage. So you see depending on the value of your $R_{sig}$, right, and this resistance $r_G$ or $r_D$ some part of $V_{sig}$ will appear as $V_i$ so not all of $V_{sig}$ will appear as $V_i$ or it might appear, but you will have a voltage divider network and therefore, we will see later on a part of $V_{sig}$ will appear as $V_i$, so that is nothing but $V_{gs}$ because source is grounded and whatever gate voltage you give will be basically your gate to source voltage. That will generator a current $I_D$, that current

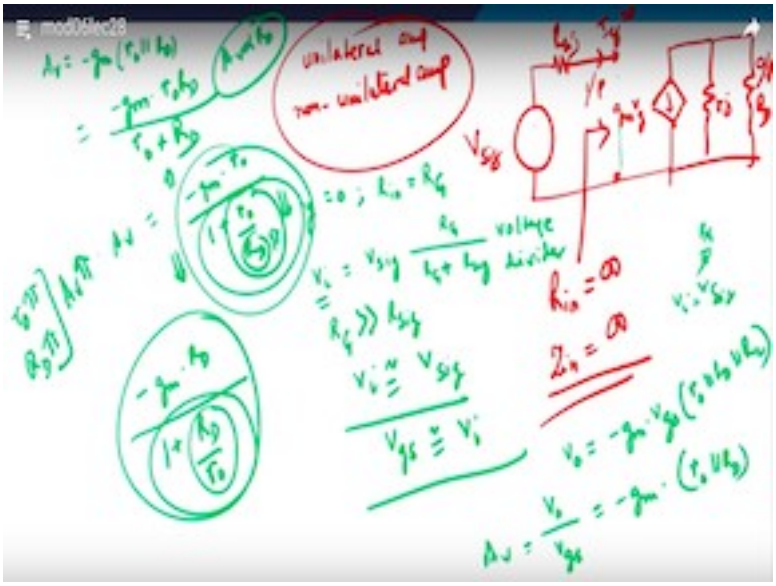
$I_D$ multiplied by this $r_D$ is basically the voltage drop $V_0$, which you get and therefore that will be the output voltage, which you see, fine.

With this knowledge you have gained till now let me therefore do a matching of one to one matching between the actual circuit and the equivalent circuit model. So, what I do? I replace this MOS device here, this device here by its equivalent circuit which you see is this one. Please understand this is very very important to how you are replacing it. You are replacing it, why? Because please understand that $G_m V_{gs}$, this one is nothing but the current source, its the drain current of this MOS device. So I have the MOS device whose applied input voltage is approximately equals to $V_{gs}$, why? Because input voltage I am giving here, gate to source voltage input voltage and therefore I get $g_m V_{gs}$ is the drain current.

We have also understood in the previous, when we were discussing the small signal was that I will always have a $r_o$ which is the output impedance of the device by virtue of the fact that your output will be slightly CLM. You will have always a channel in modulation phenomena and therefore at this point, at this point if you want to find out $\partial I_D$ you will always have $r_0$ which is not equal to infinity, but typically a relatively a large value will be there, right, and that is what is $r_0$ is here.

So this $r_0$ primarily comes out because of CLM, which means that assuming that my device, which is a MOS device here is behaving like a non-ideal current source and therefore I will have this $r_0$. This $r_D$ is the applied drain resistance in the output side or the load resistance whatever you see. So the effective resistance seen by the output world is nothing but $r_D$ parallel to $r_0$, fine. This is $r_D$ parallel to $r_0$, which you see and therefore this $r_D$ parallel to $r_0$.

(Refer Slide Time: 7:04)




So let me just discuss with you that how will I look into the same facts, so let me draw for you this diagram, this diagram if you look very carefully then I will have signal here, right. I will have R signal here, right, and then I will have $G_m V_{gs}$, right, as a current source and then I will have $r_0$ and then I will obviously have $r_D$, with us right. This is what we have learnt till now and this is what you get, so this is your $V_{sig}$, this is your $R_{sig}$.

Please understand therefore your input impedance is actually infinity, why? Because your gate current will be 0, why gate current will be 0? Because see, where you are inserting your current or you are inserting it on the gate side of your MOS device, the gate is separated from the channel by a oxide layer or a thick oxide layer, right. This oxide layer is primarily if you look very carefully is a dielectric material of course and therefore I would not expect to see any current flowing through the MOS device as a gate current.

And therefore, looking from the gate side which is the input side your impedance will be very large so you are $Z_{in}$ will be typically infinity, ideally it will be infinity so that, therefore, I keep it open input side is kept open, right. The input side is kept open this is the input side and this is my output side. Input side is kept open.

With this knowledge let me define what is, there is a term what is, known as unilateral and then you have another type of amplifier which is non-unilateral. The meaning of unilateral is that if there is no feedback between output and input, which means that

the effect of output impedance, right, does not load your input impedance and vice versa then we define that to be as a unilateral.

If there is a feedback loop which tells me that, no if I start loading the output, loading means I increase the value or decrease the value of your load resistance then your impedance; input impedance also changes, we define that to be as a non-unilateral amplifier, right.

So we have two types of amplifier typically, amplifier and this is one is unilateral amplifier and we have a non-unilateral amplifier. So if have a non-unilateral amplifier it means that if I change the output I would expect to see a change in the input by some closed loop control, right, whereas unilateral amplifier; both are totally different.

So as you can see since my input impedance is infinitely large actually my input does not talk with the output direct and as a result you will actually see that there is no change in the value of your input impedance because of the change in the output impedance.

Now with this knowledge let me therefore start doing the derivation for the gain for the, for the CS stage design and let us see how it works out. We discuss just now that of course your gate current will be 0 and therefore your input resistance will be approximately equal to the gate resistance, which means that whatever your input resistance is there will be equal to gate resistance therefore I can safely write down $V_i$ equals to $V_{sig}$, right. $R_G$ upon $R_G$ plus $R_{sig}$, why?

Because $R_G$ is a resistance which is the gate resistance which you see here, I am assuming it to be infinity but let us assume that it is not infinity at this stage it is having a finite value then we define $V_i$ to be equals to, so you see if $R_G$ is infinite value, right, if it is infinite value then this denominator will be at actually equals to just, if you look the denominator will be just $R_G$ by $R_G$. So $R_G$ $R_G$ will cancel and $V_i$ will be equals to $V_{sig}$, right and that is true also.

That means if your ideally, if your device, MOS device would have shown you an infinite resistance my input signal and my actual signal given by the voltage source will be exactly equal to each other, but since they are not I will have a voltage divider,

voltage divider network here and therefore gives me a value $V_i$ here, right, but since my $R_G$ is much larger as compared to $R_{sig}$, I will automatically get that $V_i$ is approximately equals to $V_{sig}$, right, and therefore we can safely write down $V_{gs}$ is approximately equals to $V_i$, right. So whatever you give the $V_i$ that will be equals to $V_{gs}$ as such, right.

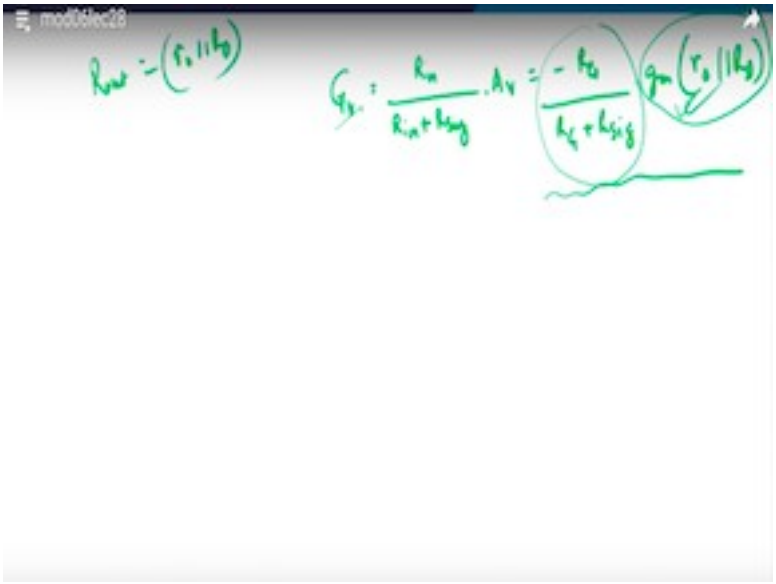
With this knowledge we have already if you remember from my previous discussion that $A_V$ voltage gain was equal to $G_M$ times $r_D$, with that knowledge let me see how it works out, let me suppose I want to find out $V_o$, output voltage, it will be minus gm times, right. $V_{gs}$ times $R_o$ parallel to $R_D$ parallel to $R_L$, $R_L$ is the load resistance, there must be some load resistance in the output side and let me suppose this is there.

So, if you want to find out gain so it will be $V_0$ by $V_{gs}$ will be nothing but $g_m$ $r_D$, so $g_m$ times $r_0$ parallel to $r_D$ assuming that $R_L$ is very very large, load resistance is very very large as compared to both of them. This into consideration, this is nothing but $A_V$, voltage gain, right. So I get voltage gain therefore $A_V$ for common source amplifier is $g_m$ times, right $r_o$ parallel to $r_D$ and therefore if you solve it, I get minus $g_m$ times $r_o$ $R_D$ divided by $r_o$ plus $R_D$, is it clear?

So if you want to find out the effect of $R_D$ for example then just simply divide the numerator and denominator by $R_D$ and you get what, you get minus $g_m$ times $r_o$ divided by 1 plus $r_o$ by $R_D$. So if you increase the value of $R_D$, right, this quantity will decrease and therefore this quantity will decrease and therefore this whole quantity will do what? It will increase and therefore your gain will become larger, so therefore, gain is directly proportional to $R_D$ as we have already learnt. Similarly gain is also proportional to the value of $g_m$.

Let us see what happens to $r_o$, right, again the same concept will come, if you divide $r_o$, I get $g_m$ times $R_D$ divided by 1 plus $R_D$ by $r_o$. So if $R_D$ increases, this decreases, this decreases, this increases. So even if $r_o$ increases or $R_D$ increases I will automatically get a higher gain with me, fine. So this is what we get overall gain, which you see.

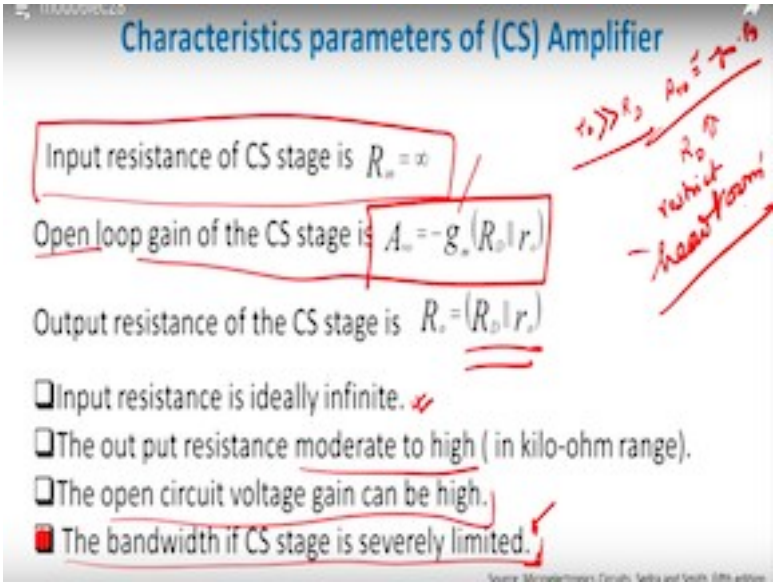
(Refer Slide Time: 13:50)



Now, if you want to find out the overall gain, total gain of the system, then we write down $G_v$, which is the total gain to be equals to $R_{in}$ upon $R_{in}$ plus R signal into $A_v$ which is equals to minus $R_G$ times $R_G$ plus $R_{sig}$ into $g_m$ times $r_o$ parallel to $R_D$, right and therefore this is this gives you the value of your overall gain system. So this is the gain due to the device itself and this is the gain due to the circuit itself and therefore I get this, but as I discussed with you overall, since $R_G$ is very very large this will cancel off with each other and I will get over all gain to be equal to gm times, $R_D$ with a negative sign, negative sign because you will get 180 degree phase shift between the drain and the gain of your device, right.

What is your $R_{out}$ in this in this case, $R_{out}$ will be therefore equals to $r_o$ parallel to $R_D$, right, and that is what we get. So with this information let me see what we have understood until now.

(Refer Slide Time: 14:56)




We have understood that the input resistance, right, in the CS stage is always equals to infinity right very important point, but it is infinity. Why is it infinity? And we have discussed point just now, why it is infinity? It is infinity because that you are applying to the gate side, right, you are applying to the gate side, once you have applied to the gate side, gate is open circuited; when gate is open circuited you automatically get input impedance to be infinitely large.

Now, how do you define the open loop gain? So what I am trying to tell you is that the gain, forget about open loop let us just discuss the gain here, the gain of a CS stage is given as minus $g_m$ times $R_D$ parallel to $R_o$ and output impedance is given as $R_D$ parallel to $R_o$, right. So as I discussed with you is ideally infinite, input impedance is infinite or the output impedance is moderate to high and the open circuit gain can be also be high depending on the value of $g_m$ and the bandwidth of the CS stage is severely limited. We will not this discuss this at this stage but this for information so you can write down a term known as bandwidth of the CS stage is severally limited capacity.

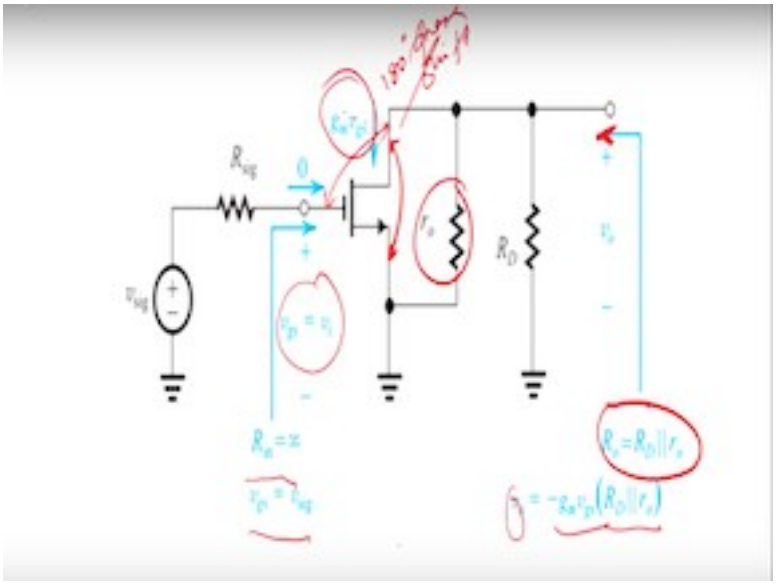
We will discuss this when we discuss the switching behaviour, means we will go to frequency modeling of MOS devices. At this stage you can just keep this in mind that it is restricted. This will be, this last point which is here, which is this one the last point will dealt, when we discuss with you the modeling of the device for frequency

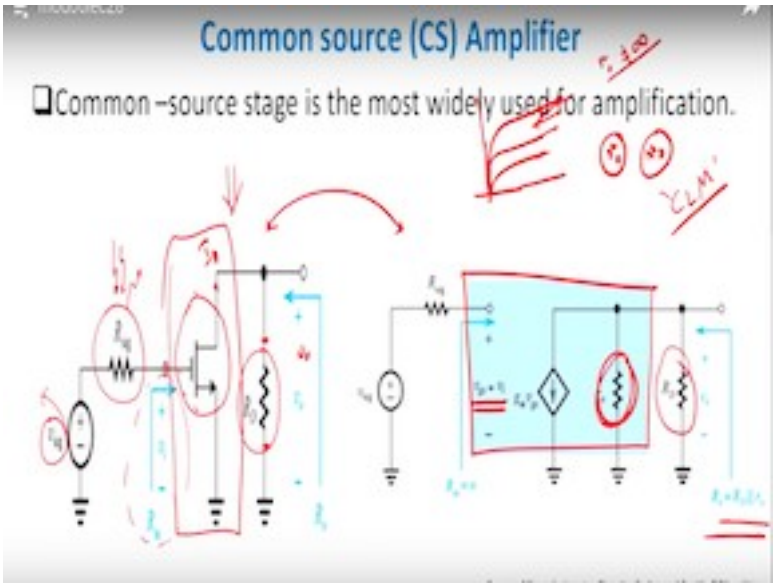
modeling; high frequency, moderate frequency and low frequency and there we try to find out the bandwidth of the MOS device.

But three things take away from all these discussion; infinite input impedance, moderately high output impedance, very large gain $A_V$ and so on and so forth.

Now, as you can see therefore the price again the same thing discussed earlier also that let us suppose your $R_D$ is suppose your $r_o$, right is very very large as compared to $R_D$ then I can safely write down $A_{V0}$ to be equals to approximately equals to minus $g_m$ times $R_D$ because they are in parallel and therefore again by the previous discussion if I increase $R_D$ I restrict my head room, right, and therefore I am playing with there will be a chance of large order of discontinuity it the output side, ok.

(Refer Slide Time: 17:22)

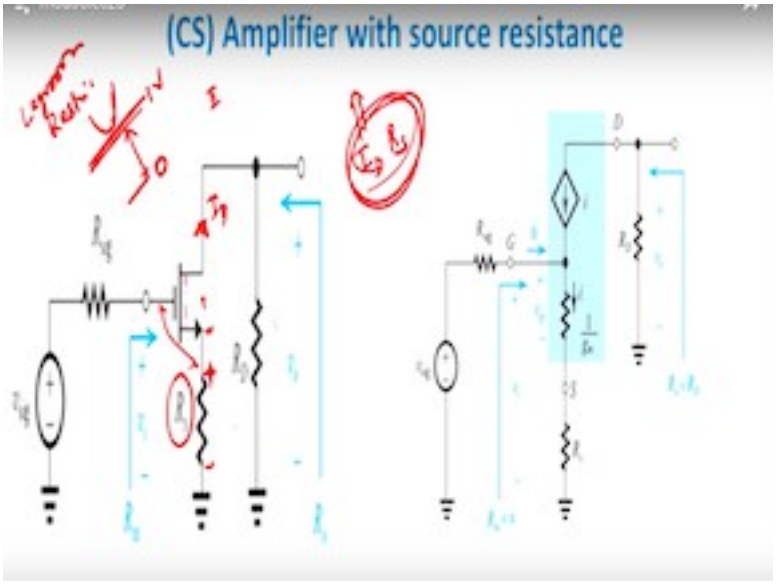





So this is how the CS amplifier looks like and let me come to the next stage also referred to as, so this is what I was talking about so you have, so you have your, just let me just come to the point, let me come, discuss the previous point as well as I was discussing with you, so you see $V_{gs}$ is equals to $V_i$, so $R_{in}$ equals to infinity and therefore since $R_{in}$ equals to infinity and $V_{gs}$ equals to V signal, drain current is nothing but $g_m$ times $V_{gs}$, this is the output voltage which you see, which is equal to nothing but $g_m$ $V_{gs}$ multiplied by $R_o$, right, that is what is because current multiplied by output impedance will be the output voltage and therefore you get $g_m$ times $V_{gs}$ multiplied by $R_D$ parallel to $R_o$ is equal to $V_o$ negative sign because gate and drain of a MOS device between these two, you will always have a 180 degree phase shift, right. Phase shift you will always have that is perfectly possible.

Now this also circuit takes into account the $r_0$ turn right it was not there in the previous discussion as you can see this was there in the previous one if you look at the previous one discussion this was grounded whereas this was also grounded but we have connected to the source showing that it is between these two nodes exactly which is happening, right. Now, so this is what we get a common source configuration.

(Refer Slide Time: 19:00)



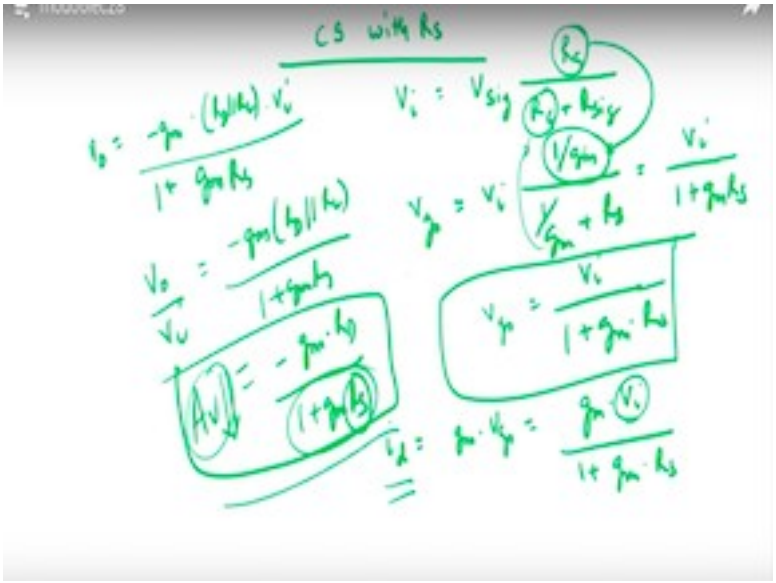

Now let me come to the common source amplifier with source degeneration resistance. So member what change we are doing, what change we are doing is we are replacing the source side which was initially grounded by a source resistance $R_S$. So this is source resistance $R_S$ which is in series to the MOS device, right, so I have a MOS device here and I will put a series resistance here. I have discussed already with you why this is important in terms of application.

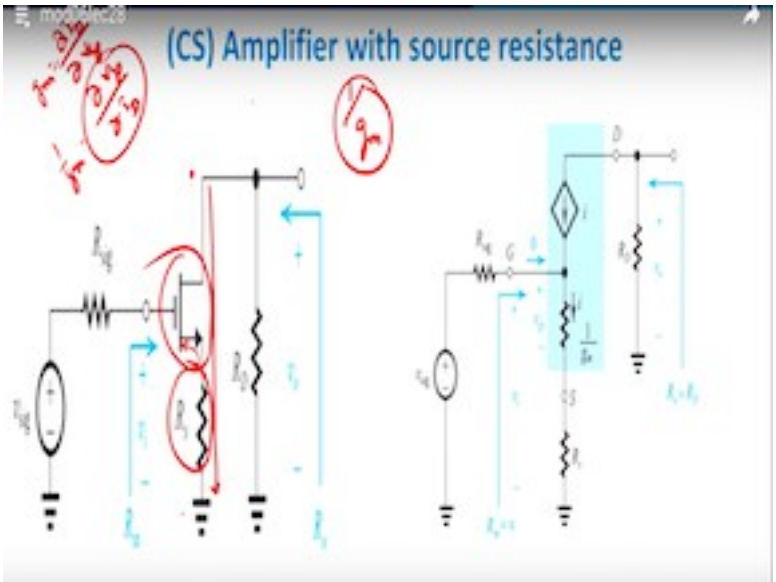
See it is important because as you make your $R_S$ large or you make your source degeneration resistance large you end up having the voltage drop across this larger and larger, fine, because let us suppose current $I_D$ is flowing, sorry $I_D$ is flowing, when once the $I_D$ is flowing through the circuit $I_D$ multiplied by $R_S$ is basically the voltage drop across $R_S$. So if your $I_D$ increases, right, $I_D$ $R_S$ increases and therefore this voltage increases which in fact means that you are able to reverse bias this gate to source junction, fine.

Why? Because it is MOS device, n channel MOS device, so this is basically N and this is basically P, you have an N region here and since on the N side you are giving a positive bias you are reverse biasing it. Once you are reverse biasing it your current is falling down, so whatever current increase you got by virtue of some problem you will be able to reduce it by giving a negative feedback, right. So this is also a negative feedback.

Now, now so therefore, but therefore the problem what is the cost you pay for it is something like this. The cost you pay for it is now your leg room is restricted, leg room is restricted, why? Because out of the negative say you want to go to 0, you actually went to 0 at one point of time. $I_D$ $R_S$ is 1 volt, so out of 0 your 1 volt is already taken care of by $I_D$ $R_S$, so your 1 volt here, so you can only go up to this much point. You cannot go below this much point. If you go you will have, you will have clipping, you will have wave-shipping problem so on and so forth.

(Refer Slide Time: 21:11)




So this is a source degeneration resistance, which you see and I told you what is that advantage of source degeneration resistance, right. Let me therefore show you source degeneration, so CS stage with $R_S$; let me just do some small derivation here, CS

mean a common source with source degeneration resistance. Now let me write down $V_i$ therefore again same thing, $V_{sig}$ signal into $R_G$ upon $R_G$ plus $R_{sig}$, right.

As I discussed with you therefore $V_{gs}$ is equals to $V_i$ and this is quite interesting 1 by $g_m$ upon 1 by $g_m$ plus $R_S$. Now, this is quite interesting I will just explain to you from this discussion. See now the idea is that if you go from this point to this point. From this point to ground, right, you are encountering two resistances. One is $R_S$, right, another is the resistance offered by this device, right, and that offer is that looking from the source side the resistance offered is basically 1 by gm. Remember $g_m$ was equals to $g_m$ was $\partial I_D \partial V_{gs}$, right, so if you do 1 by $g_m$, it is $\partial V_{gs}$ by $\partial I_D$, right.

So $\partial V_{gs}$ $\partial I_D$ primarily means that for the same change in the value of $V_{gs}$, how much $I_D$ is changing? That can only be done provided you know what is the resistance offered so this is basically the resistance. So 1 by $g_m$ is the resistance offered by the MOS device.

So therefore this is very very important you can solve it yourself and get the principle get this clear that obviously therefore I can replace this $R_G$ by 1 by $g_m$, $R_G$ by 1 by $g_m$ so I am replacing this $R_G$ by 1 by $g_m$ which I was assuming to be infinite in the previous case I am assuming it to be 1 by $g_m$; yes and 1 by $g_m$ will be relatively small quantity and therefore $R_S$ will be there here and I will get something like this.

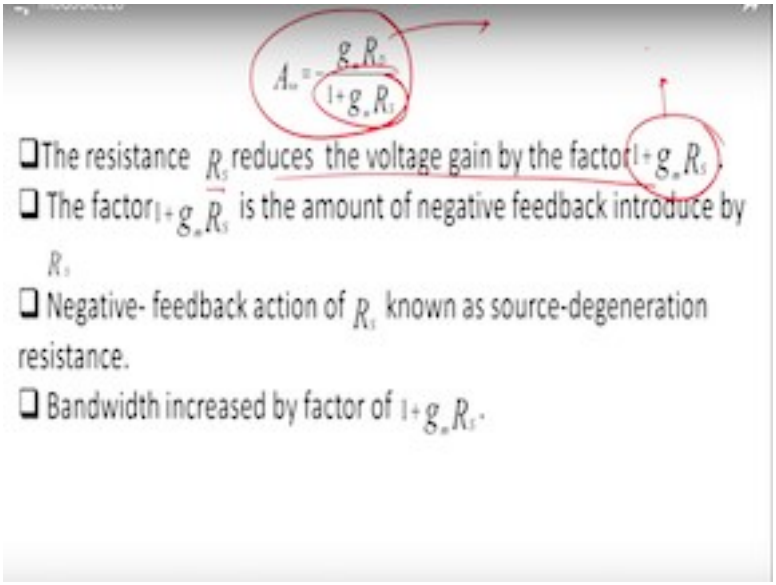
Now, from here this could be written as or this could be written as this could be written as $V_i$ upon 1 plus $g_m$ times $R_S$ because if you cross multiply this will be 1 plus $g_m$ $R_S$, this $g_m$ will go in the numerator, right, and this $g_m$ will cancel with this $g_m$ and you will remain with this. So I get what? $V_{gs}$ equals to $V_i$ upon 1 plus $g_m$ times $R_S$, right, this is the gate to source voltage.

Therefore, I can write down $i_d$, which is the drain current flowing to the device to be $g_m$ times $V_{gs}$, $V_{gs}$ is nothing but for this quantity $g_m$ times $V_i$ upon 1 plus $g_m$ into $R_S$, right. This is what I get and therefore if I want to find out the value of $V_0$ output voltage I just have to multiply $I_D$ with $R_D$; $R_D$ parallel to $R_L$. So what I get I get minus $g_m$, right into $R_D$ parallel to $R_L$ output resistance multiplied by $V_i$, this $V_i$ divided by 1 plus $g_m$ times $R_S$, right.

And therefore, I can write down this $V_0$ to be equal to minus $g_m$ times, right $R_D$ parallel to $R_L$, so $V_0$ by $V_i$, 1 plus $g_m$ $R_S$, so if you solve it I get minus $g_m$ times $R_D$ assuming that $R_L$ is very very large quantity as compared to $R_D$ I get 1 plus $g_m$ $R_S$. This is your voltage gain, its minus $g_m$ $R_D$ upon 1 plus $g_m$ $R_S$.

So you see as compared to CS stage without source degeneration resistance, with source degeneration resistance your gain is actually fallen by a factor 1 plus $g_m$ $R_S$, right, and that is quite an interesting observation or quite an interesting idea, that what has happened therefore is that once your, once you are increasing the value of a source degeneration resistance because you wanted to be more stable and you do not wanted to be going beyond a particular limit or you do not want to clip out output device what happens is that this increasing $R_S$ will result in a reduced value of $A_v$ or the gain will reduce, right and that is the price you pay for a reduced gain with higher $R_S$. So what we have learned therefore is that $A_V$ will be given as minus $g_m$ $R_D$ upon 1 plus $g_m$ $R_S$, right.

(Refer Slide Time: 25:28)

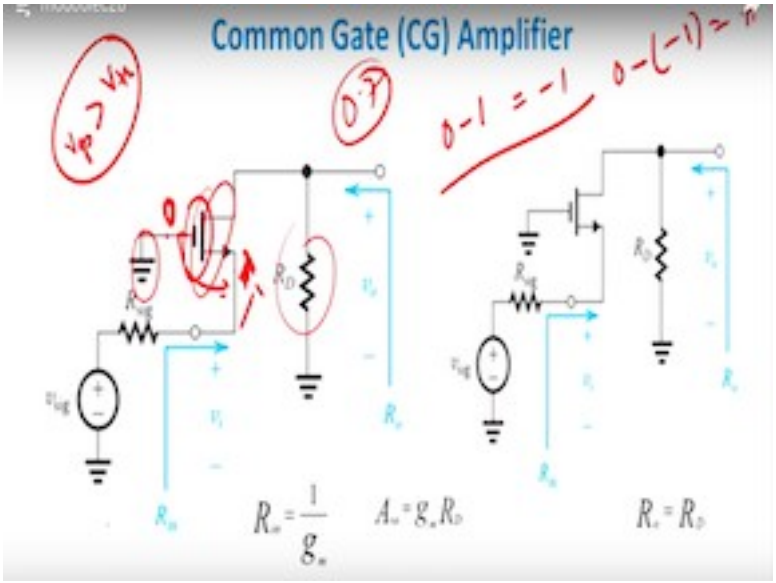



So let me come to this point and this is what I was talking about. So $A_{V0}$ is equal to minus $g_m$ $R_D$ upon 1 plus $g_m$ $R_S$ and therefore, the resistance $R_S$ reduces a voltage gain by a factor of 1 plus $g_m$ $R_S$, this is the reduction. Initially we had $g_m$ $R_D$ now reduces by factor 1 plus $g_m$ $R_S$.

So therefore, we define 1 plus $g_m$ $R_S$ is the amount of negative feedback which is introduced by $R_S$, right, so you have $R_S$ resistance, the voltage drop across it will be

negative bias reverse biasing your source to channel contact, or source to gate contact and as a result you will have negative feedback, this negative feedback will result in what? Will result in a reduced gain. So what happens is that if the negative feedback is also known as, this $R_S$ is known as source degeneration resistance, right, and the bandwidth increases by the factor of 1 plus $g_m R_S$. We will come to this later on in our discussion of frequency response. So this is what we get from common emitter or sorry common gate, sorry common source with source degeneration resistance.

(Refer Slide Time: 26:36)



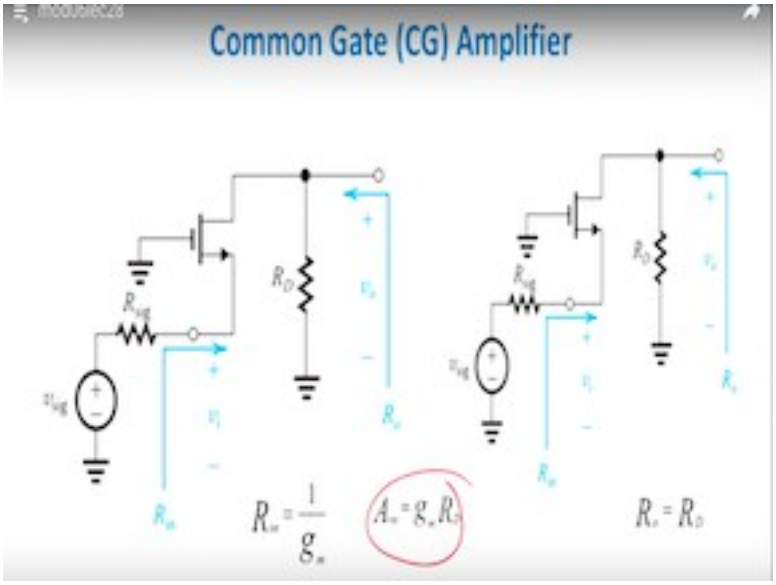

Now, let me come to common gate configuration, in common gate configuration again as I discussed with you the gate will be actually grounded so it is also known as gate grounded configuration, right. So I have a gate grounded configuration what is it means, that my gate is basically grounded which means that the gate is grounded here as you can see and therefore when get is grounded we automatically have a device and if it is an NMOS device it will be basically in a cut-off mode as such.

And therefore if you apply any signal here, right, what was the idea that $V_{gs}$ should be great than $V_{th}$ for the device to be your, but you see your gate is giving you 0 voltage here, so if your source voltage is positive, right, if its positive gate voltage which you are giving then gate to source voltage will always be negative, right.

Because let us suppose this is 2 volt here, and maybe 1 volt here, so 0 minus 1 will be equal to minus 1 volt, right, and therefore you need to switch it on to give, but in the

negative cycle suppose this is 1 volt here and sorry this is minus 1 volt here, 0 minus of minus 1 will be equal to +1 and this is larger than the given threshold voltage. Let us supposed threshold to be 0.7 in that case it will switch on and there will be current flow, right, and so therefore I have $R_D$ and so and so forth.

(Refer Slide Time: 28:06)




So let me therefore delete this option here or remove all the, remove all the inks here and you see for a common gate, if I have a common gate option or we, I will prove it maybe next time that in this case also I get $g_m$ $R_D$ as my output resistance, but in this case quite interestingly you do not have a phase change of 180 degree. So the output will be exactly in phase with input side unlike in the CS stage or in the CD stage here you will not get the same profiling or the same this thing in this case, right.

So with this let me stop here today and explain what we have done, let me just recapitulate; we have done common source, we have done common source with source degeneration, we will take up common gate in the next turn and common drain in the next turn and that will finish the amplification action of a MOS device.

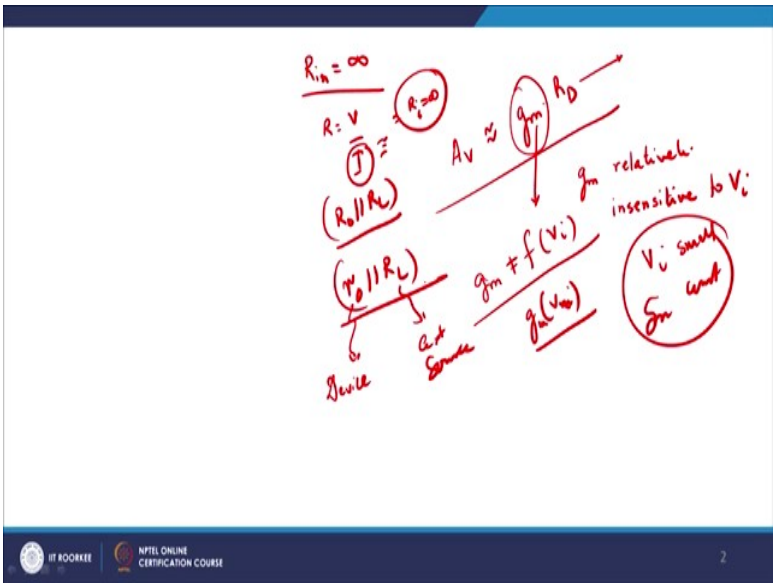
Since MOS devices is actually a device which is, which is, VCCS or a voltage controlled current source, we, output impedance of the device will be typically low but the input impedance will be typically very high, right, and therefore I can use it in the MOS device provided I am able to fix it in the saturation region of operational device. So with this let me thank you for your patient hearing until we meet next time. Goodbye!!!!

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture – 29**
**CMOS CG/CD Amplifier Configuration**

Hello everybody and welcome to the NPTEL Online Certification Course on Microelectronics: Devices to Circuits. Today's lecture will be primarily concerned with common Gate and common drain amplifier configuration. In our previous schedule, we had looked into common source amplifier with source degeneration resistance as well as without source degeneration resistance. We also saw the advantages of having an source degeneration resistance.

And the advantage primarily was to do with enhanced bandwidth but a reduced gain and more stability. What we will do today is have a look at common gate configuration and common drain. But, please take into consideration that most of the time common source configuration is used as an amplifier because of its relatively high gain which depends upon the value of transconductance of the MOS device under study.

(Refer Slide Time: 01:31)



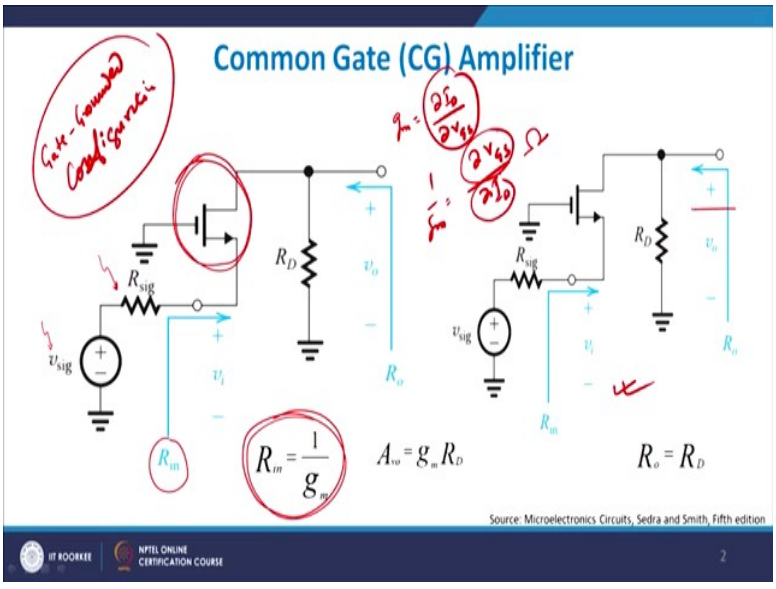

So, higher the transconductance, we remember yesterday in our previous discussion, if you look very carefully, what we did, what we saw was the gain in general is considered to be equals to $g_m$ times $R_D$. This we have already learnt where $g_m$ is the transconductance of the device and $R_D$ is the drain resistance. Now, this $g_m$ we have already discussed this point that is you need to be relatively insensitive to $V_i$. Right.

Which means that to input voltage. Which means that gm should not be a function of $V_i$. In reality, it is not true and therefore, $g_m$ happens to be a function of input voltage. But if your $g_m$ or $V_i$ is relatively a small signal then my $g_m$ is considered to be constant right. Under that condition, we say that my linearity is sustained and I am in a saturation region. We have also seen in the previous interactions and discussions that the input impedance of MOS device is always infinity.

$R_{in}$ is typically infinitely large because you are actually feeding your signal to the gate side. And since the gate side has got oxide layer inbuilt into it you will not have any current and therefore, your resistance which is voltage by current, since your current is approximately equals to 0 you get R equals to infinity. And that is what typical input resistance of the amplifier is all about.

Output impedance will depend upon many factors but typically it is $R_o$ parallel to $R_L$, where $R_o$ is the resistance offered by the device itself. $r_o$ should be small parallel to $R_L$ or $R_D$, where $R_o$ is basically the resistance offered by the device itself and $R_L$ is the resistance offered by the external source. So, with this basic logic or with this basic understanding, let us now undergo what is known as a Common Gate Amplifier.

(Refer Slide Time: 03:27)




In the Common Gate Amplifier you need to therefore ground your Gate. So, if you look at this diagram here, this is your NPN transistor which you see or sorry it is an N channel MOSFET which you see and its gate is basically grounded here which means this is also referred to as Gate Grounded Configuration. Right. So, this is Gate Grounded Configuration and this is also referred to as Gate Grounded Configuration and the Gate is grounded.

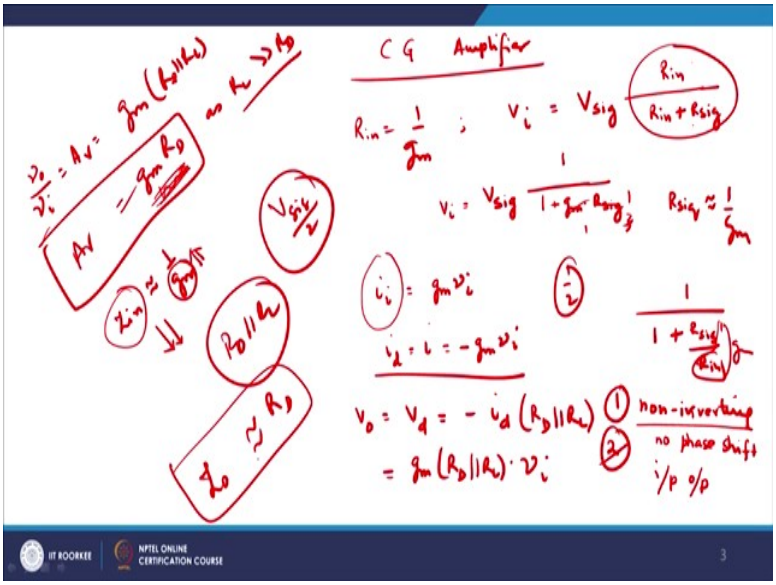
Now, please understand, therefore you are inserting a signal to the source end of my gate and you are extracting the signal from the drain side. So, this is your $V_{sig}$ is my input signal and $R_{sig}$ is my resistance offered by this voltage source. Resistance has a series to it, typically relatively very small and $R_{in}$ is basically the resistance offered by the device looking from the source side.

So understand, looking from the gate side it was infinitely large because there was an oxide layer. But, looking from the source side it is not infinitely large but typical value is 1 by $g_m$. Right. So, I will not derive it here but a source, from the source side if you want to look in a MOS device the resistance offered by it is 1 by $g_m$, right. So, higher the trans conductance of the device, lower will be your $R_{in}$.

And intuitively you can understand why is it like that. If you remember, $g_m$ was equals to $\partial I_D$ by $\partial V_{GS}$, right. So, if you take 1 by $g_m$ it is just the reciprocal of this one and you get $\partial V_{GS}$ by $\partial I_D$, which means that for how much change in $I_D$ I should get for a typical change in $V_{GS}$. And if you look at this dimensions it is basically in ohms.

So, this is the resistance offered by the device when you are looking from the source end of its operation. On the right hand side which you see here, which is this one, if you look at this one, we are assuming that the diagram is exactly the same which you see on your left and right. But, here I am assuming that my impedance looking from the drain side is actually equals to $R_o$.

(Refer Slide Time: 05:48)



With this knowledge let me just do a small derivation, a brief interaction of CG amplifier. So, this is basically your common gate amplifier, right. If you solve it, I get $R_{in}$ equals to 1 by $g_m$, this we have already seen just now. I can say $V_i$ equals to $V_{sig}$ into $R_{in}$ upon $R_{in}$ plus $R_{sig}$. $R_{in}$ is the input voltage, input resistance which you see and $V_{sig}$ is the voltage of the signal which you see. Therefore, I can right down $V_i$ equals to $V_{sig}$ signal divided by 1 upon 1 plus $g_m$ times $R_{sig}$. How do I get it?

So, how do I get it, if you divide numerator and denominator here by $R_{in}$. So, if you divide this whole by $R_{in}$, numerator and denominator, I get 1 in the numerator, I get 1 in the denominator plus $R_{sig}$ by $R_{in}$. But 1 by $R_{in}$ is nothing but $g_m$. Therefore, I get $g_m$ times $R_{sig}$. Therefore, I get input current ii equals to gm times $V_i$. Please understand here, why are you getting an input current because you are feeding the voltage through the source terminal of the MOS device. That is the reason you are having an input current, right.

This would not have been there provided you are actually doing on the gate side, right, and that is the reason it is $g_m$ times $V_i$. So, therefore the drain current which you get is also therefore equals to $g_m$ times $V_i$, but it will be with a negative sign. So, this will be minus gm times $V_i$, negative, why, because there is a phase shift between there. Sorry, because the current direction will be just reversed as that of the source side.

Therefore, I get $V_o$ which is output voltage also referred to as $V_d$ equals to minus $i_d$ times $R_D$ parallel to $R_L$. If you solve it, I get $g_m$ times $R_D$ parallel to $R_L$ multiplied by $V_i$. So, $V_o$ by $V_i$ is nothing but, therefore $V_o$ by $V_i$ is nothing but $A_v$ and that is equals to $g_m$ times $R_D$ parallel $R_L$, right. Therefore, I can safely right down this to be equals to $g_m$ times $R_D$ as $R_L$ is much larger as compared to $R_D$. Therefore, I get $A_v$ to be equals to $g_m$ times $R_D$. You see here, right.

So, your Common Gate Amplifier is basically $g_m$ times $R_D$ you see here $A_v$ gain, right it is $g_m$ times $R_D$. So, what I get from here is that it can act as a unity drain device provided $R_{sig}$ is approximately equals to 1 by $g_m$. So, if it is 1 by $g_m$ that cancels out I get 1, so 1 by 2 half. So, what I get half here, I can have $V_i$ to be equals to $V_{sig}$ by half, right. So, half the $V_{sig}$ is available to you in the input side of your device and therefore your gain will also reduce in that sense.

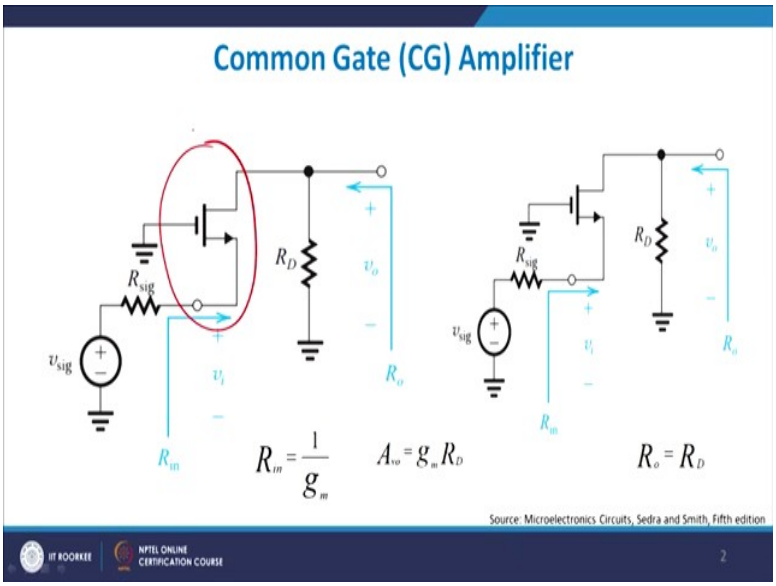
One thing is there, on the second part of it is that, it is one characteristics is that it is noninverting in nature. So, it is noninverting, which means that, the phase, there is no phase change between input and output, right. Because you are feeding from the source side and not

from drain side. There would have been a phase change provided you had taken it from Gate side and taken from the drain side, output would have been from the drain side.

But, in this case you are feeding from the source side and you are taking it from this drain side. Therefore, it is also refer to as non-inverting design with no phase shift between input and output, first thing. Second thing is, if you look very carefully the input impedance of this device is approximately equals to 1 by $g_m$ which we already understood.

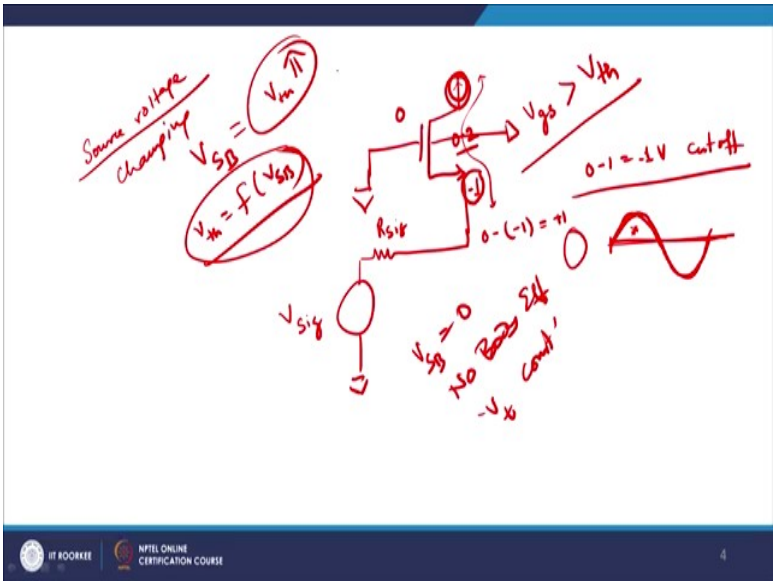
Now, this is relatively very low. Because $g_m$ is relatively very high, right. And therefore, $Z_{in}$ is typically low, so the input impedance of the device in the Common Gate Amplifier is relatively low, right. So, that we have understood that it should be low and output impedance can be found out as $R_D$ parallel to $R_L$ and if $R_L$ is typically large it is approximately equals to $R_D$ which you get $Z_o$ output impedance, right, for a Common Gate Amplifier.

(Refer Slide Time: 10:37)




One small thing, which I just wanted to point out to you here is, that that if you look, for example, if you look from this side, this diagram here and just concentrate on this particular point then let us see what is the problem.

(Refer Slide Time: 10:42)




One small problem area of Common Gate. So, please understand, you are feeding the signal from this $R_{sig}$ right, $V_{sig}$ right and this is $R_{sig}$ and you are feeding on to the source terminal of the battery. And the gate is grounded, right. If you look very carefully, when this, so for the device to be on, gate to source voltage should be larger than the threshold voltage of the device. That is the basic concept we have been learning for, we know very well.

Now, if gate is 0 volt and let us suppose my source voltage is positive and let us say it is 1 volt then 0 minus 1 will be equals to minus 1 volt and therefore my device will be in cut off, right. And that is the problem area that whenever my device goes to negative or the positive. Let us suppose, I have a positive and a negative cycle, then for the positive cycle you will automatically have this problem.

In the negative cycle, let us suppose this is minus 1, then what is get is 0 minus of minus 1 which is equals to plus 1. And let us suppose the threshold voltage is point 2 then obviously my device is in the on state, right. So, the negative half cycle it works fine but the positive half cycle you do have a problem that your device might be in the cut off state. And that is the one of the problem areas of Common Gate as such but we will discuss no further than this in this case.

Secondly, why there is no phase difference between Drain and source because you will see, if you look very carefully, whatever potential you are giving here the same potential will appear at this edge, right. Because, the same current has to flow between these two points and therefore it is not out of phase it is perfectly in phase. And therefore, basically it is a non-

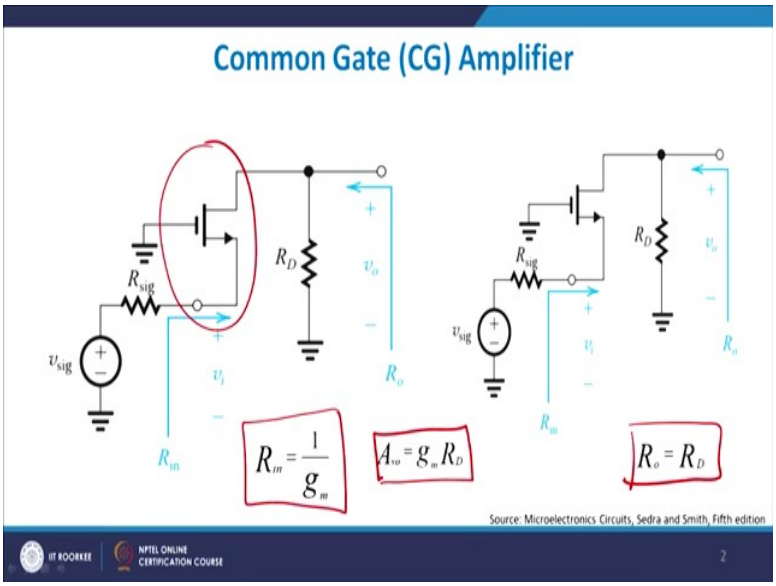
inverting. Common Gate is basically non-inverting amplifier and the second thing is $Z_{in}$ is very low and $Z_{out}$ is relatively high.

With this, Okay, One more thing which maybe I can discuss with you is, one more problem area is that, now you see very effectively your source voltage is changing, right. Source voltage is always changing, right. Why it is changing? Because we applied a $V_{sig}$ here. The signal voltage will change and therefore this voltage will go on changing, right. Once this change, obviously $V_{SB}$, let us suppose our base is grounded, then $V_{SB}$ will also change source to bulk and therefore my threshold voltage, which you remember will be function of $V_{SB}$.

In the previous, all my examples, since my source was grounded and even if it is not grounded provided in a MOS device, I am able to plug my bulk to the most negative part of the potential. Then my $V_{SB}$ will be approximately equals to 0 and therefore, I will get automatically no body effect. So, under the condition that your $V_{SB}$ is equals to 0, you will never have and there will be no body effect and your $V_t$ will be effectively constant, right.

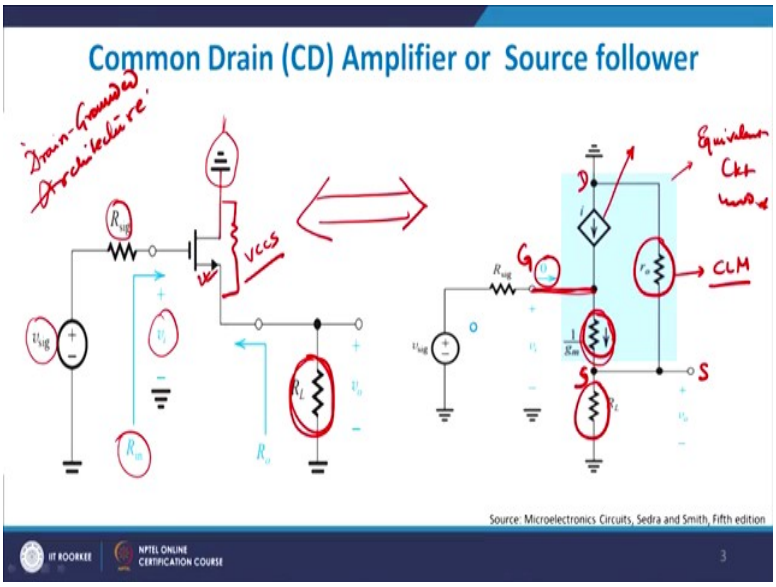
But, in this case specifically when you are doing a Common Gate structure, you end up a variation in the voltage of the source side and therefore source to bulk voltage goes on varying and therefore threshold voltage becomes the strong function of the source to bulk voltage. And therefore, threshold voltage will vary. And therefore, we are not sure whether the device will be on or off at particular point of time. So, this is one of the problem areas which people face when you are doing a Common Gate translation.

(Refer Slide Time: 14:23)



We come to the next part and that is basically, just a minute, so Common Gate we have finished, we have seen that $A_{vo}$ is $g_m$ times $R_D$ and output impedance is approximately equals to $R_D$ and input impedance is 1 by $g_m$, right. So, it is just Common Gate Configuration which you see.

(Refer Slide Time: 14:35)




Let me come to Common Drain architecture, Common drain as I told you there are termed as a Common Drain, your drain will be grounded and therefore it is also referred to as Drain Grounded Architecture. It is referred to as a Drain Grounded Architecture. This Drain Grounded Architecture which you see, is primarily means that the drain has been grounded, obviously to the 0 potential here and applying at signal back to the gate side, right.

So, only in one condition which is CG which is Common Gate is the point where you apply the signal through the source side. For all other conditions, for all other amplifier configuration you apply the signal voltage on to the gate side. So, I am applying a $V_{sig}$ here through an $R_{sig}$ here applying to this. $V_i$ is the available input voltage and $R_{in}$ is the input impedance.

So, $V_i$ is the available input voltage out of, so $V_{sig}$ was your total voltage given to the device out of which $V_i$ appears as the input voltage to the Gate side of this device and you get something like this. And then you have $R_L$ here which is the load impedance and there will be a output voltage will be seen from this side. So, if you do a short channel. Sorry, small signal analysis and maybe replace it by a T model then let us see how it works out. So, if you see very carefully, this blue box is my Equivalent Circuit Model. I will explain to you how we got it.

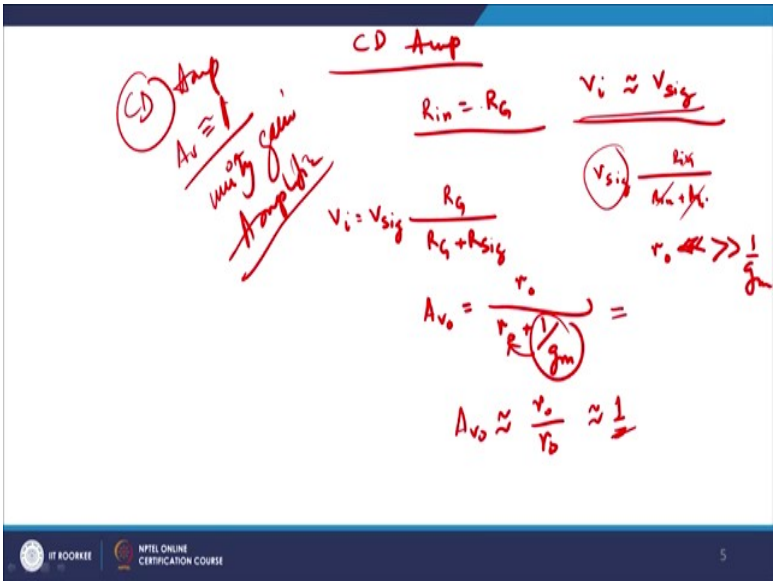
First of all see, this ro is between this point and this point and this primarily comes out because of CLM. So, Channel Length Modulation remember, I am assuming that my output impedance will be infinitely large, not true, there will be a finite impedance available between the drain. So, this is your drain end and this is your source end and this is your drain end. So, you will automatically have a impedance or a resistance between Drain and Source, which you see in front of you, this one. And therefore, this is your source actually and therefore you get this thing.

What is 1 by $g_m$? 1 by $g_m$ is nothing but this device itself. Because, remember this is basically a VCCS (Voltage Control Current Source). So, I can safely write down the resistance offered by the device is 1 by $g_m$ looking from the source end. We discussed this point earlier also, right. This is the point. And this $R_L$ is basically resistance offered by external load. So, $R_L$ is external load which you see in front of you. And This is the MOSFET current source.

So, I am assuming this to be an ideal current source, right. So, its output impedance is infinitely large and then what I am trying to do is, its corresponding resistance looking through the source side is given by this. Now, why I am applying, why I am adding this here, the reason is very simple, that since no current is flowing through this point because it is a gate terminal.

No current, I can safely join the Gate terminal to this particular point, without violating any Kirchhoff's law. Because, there is no current flowing through these two arms. So, what is the net equivalent circuit, I have $r_0$ here 1 by $g_m$ $R_L$ and a current here which is basically an ideal current source and I get $V_{sig}$ at this particular point, right.

(Refer Slide Time: 17:56)

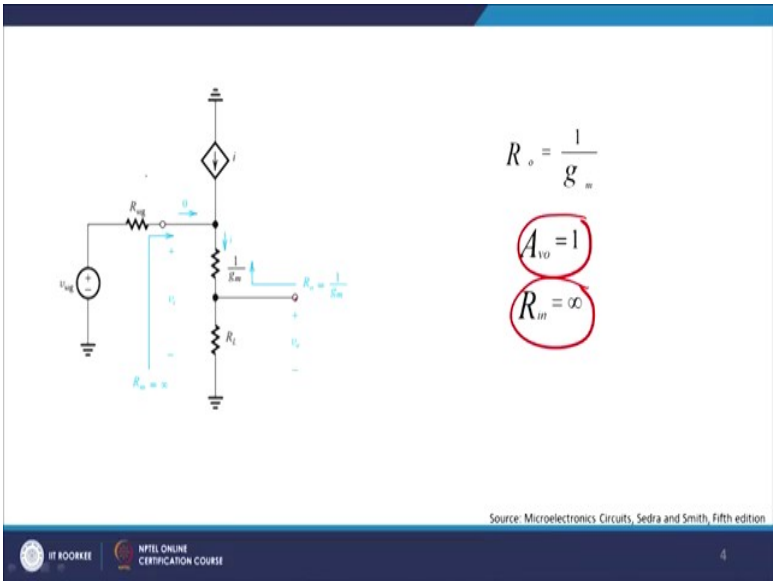



Now, let me explain to you the CD (Common Drain Amplifier), what you get is $R_{in}$ equals to $R_G$ the gate voltage, of course, and Vi therefore will be approximately equals to $V_{sig}$, and the reason is that, if you remember the concept was that your, if you remember in previous term or previous discussion, we saw that $V_{sig}$ was actually getting divided by $R_{in}$ and $R_{in}$ plus $R_G$. Now, if $R_{in}$ is relatively very large this cancels off and therefore I get $V_i$ equals to $V_{sig}$.

So, under the condition that your repeat impedance is relatively very high, I automatically get my $V_i$ equals to $V_{sig}$. So, whatever $V_{sig}$ you are giving, say you are giving a 2 volt supply, 2 volt will apply at the Gate side of MOS device. With this knowledge, I can write down $V_i$ equals to $V_{sig}$, right, $R_G$ upon $R_G$ plus $R_{sig}$, right. So, if you solve it I get $A_{v0}$ to be equals to $r_0$ divided by $r_0$ plus 1 by $g_m$. And therefore, if you see, sorry it is very-very large as compared to 1 by $g_m$, right.

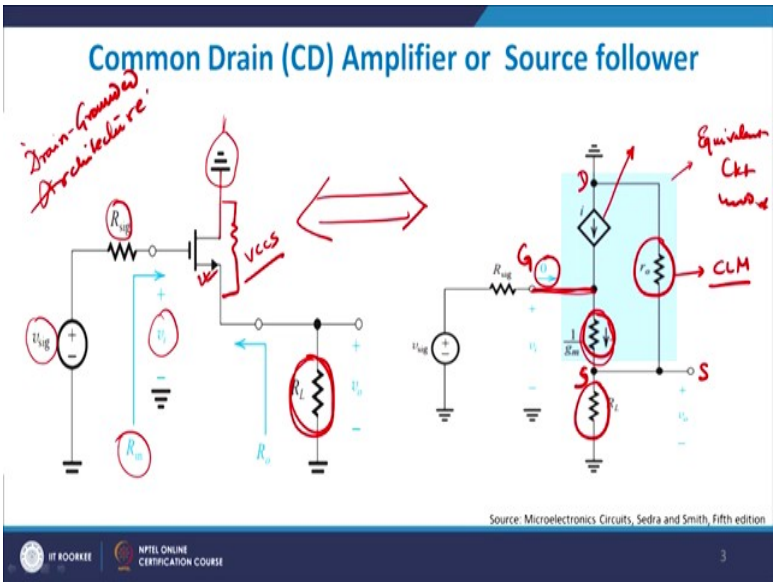
This r0 which is primarily because of the CLM phenomena is very-very large compared to 1 by gm. So, I can safely write down that Av0 will be approximately equals to r0 by r0 because this will be very small as compared to r0. And therefore, I can safely write down this to be equals to 1. Therefore, I can safely say that the CD amplifier (Common Drain Amplifier) is having a Gain of unity which is basically a Unity Gain Amplifier. This is basically a Unity Gain Amplifier, right. This is Unity Gain Amplifier which you see. So, it is basically uses a buffer, right.

(Refer Slide Time: 19:56)



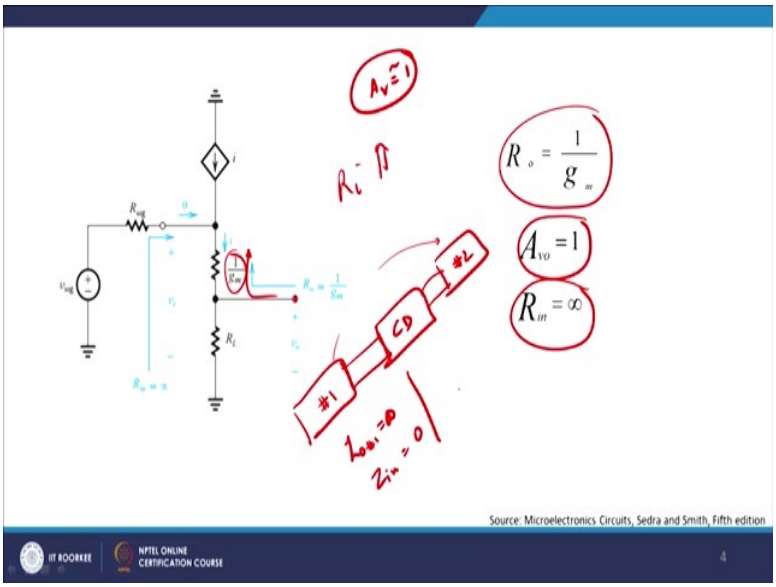

Now, if you come here therefore, you see $A_{v0}$ is 1, $R_{in}$ is infinity for reasons which you have already understood by now because, $R_{in}$ is infinity why? Because you are actually seeing from the Gate side. So, as I discussed with you infinity. $R_o$ is basically the impedance looking from the drain side onto the device.

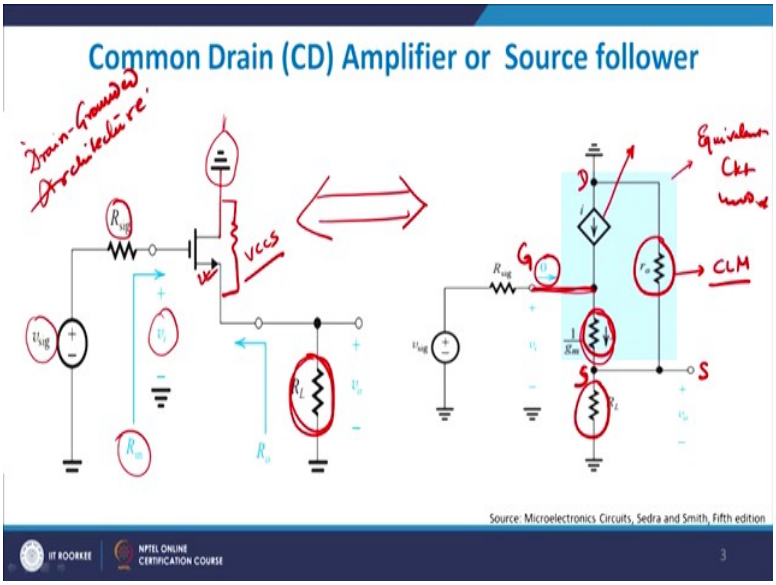
(Refer Slide Time: 20:17)




So, if you come back here, you are looking from the Drain side onto the device, means you are looking from this side onto this side which is nothing but 1 by $g_m$ as we discussed earlier.

(Refer Slide Time: 20:26)



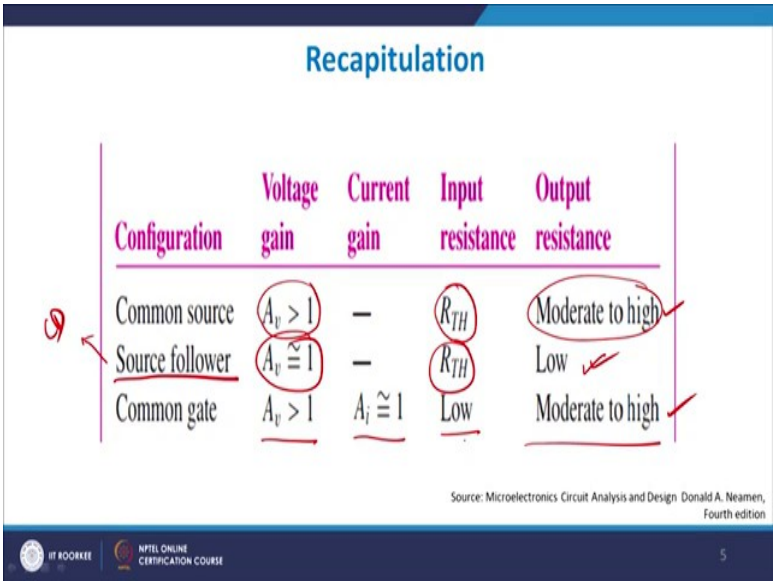

If you are looking at the output here the equivalent Circuit model, If you are looking from this, then this is my output and then looking into this region, right, from the Gain side, not from the drain side, sorry. I will just explain to you, this is my output right, so you are looking from this side. Looking from this side is basically looking from the source end of the device. Source end of the device means 1 by $g_m$. So, I get $R_0$ equals to 1 by $g_m$, right. So, if you want to use it as a impedance matching purposes, you can do that, provided so this will be somewhere sitting in the buffer of IU design, right. It can be used as a buffer.

Why? Because since its $A_v$ is approximately equals to 1, it does not give you any Gain. But, what it does is, its input impedance is relatively very high, $R_i$ is relatively very high, so if I have one block here and another block, this is my, I have another block here. So, this is my CD right, since it is input impedance is infinite, the output impedance of previous stage, so

this is stage number 1 and this is stage number 2 and you want to have the maximum signal to be transferred from stage 1 to stage 2.

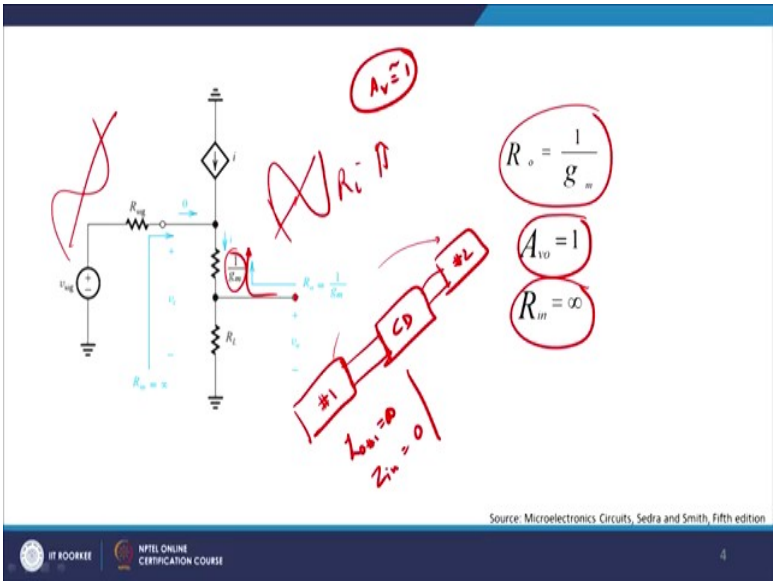
Then, if the output impedance of stage 1, $Z_{out}$ of stage 1 is infinitely large, then you connect this to this very easily because, input impedance of CD is infinitely large. And if I can connect to the, like this, then CD drives stage 2 whose input impedance is 0. And therefore, I can have a maximum power being transferred from 1 to 2. And therefore, reliability is not a major question provided you have a CD gap in between them, right.

(Refer Slide Time: 22:43)



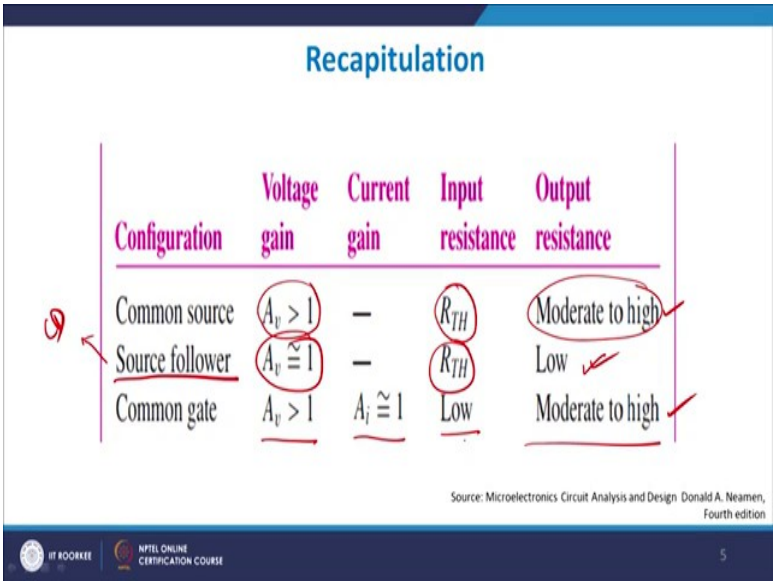

With this knowledge, we have almost done the most part of the design functionality of the Common Gate, Common Drain. Let me just recapitulate therefore all the 3 configuration to you. The first is Common Source, as I discussed with you, obviously its $A_v$ is much larger than 1, right. Input impedance is the Thevenin equivalent of the total resistance offered by the device. Output resistance is moderate to high. We just now saw source, Common gate $A_v$ is greater than 1, current gain is approximately 1, input resistance is low and moderate to high. Ok, Common Drain this is also referred to as CD, Common Drain is basically a source follower.

(Refer Slide Time: 22:48)




I will explain to you, why it is source follower. Just a minute! Source follower primarily mean that, whatever voltage you give on the source side, right. Exactly same appears across the output. So, source follower means whatever input you are giving in the source side, right. Exactly same thing appears in the output side without any gain or loss, because your Av equals to 1. Therefore, this is also referred to as a source follower.

(Refer Slide Time: 23:14)




How do you define a source follower? Source follower is that the signal on the source side of a device will be exactly equals to the input signal and therefore it is also referred to as a source follower with $A_v$ approximately equals to one, right. Gain is approximately equals to.

So, Common source, Common gate, both are very high Gain out of which Common source has got much higher Gain.

Output impedance is relatively high for both of them. But, for Source follower the output impedance is low because it is 1 by gm and input impedance is also relatively small which you see in front of you. So, what we have done till now therefore is that, we have finished with configurations of Common gate, Common source and we have also seen into the fact of $Z_{in}$ and $Z_{out}$. Therefore, we are now in a position to use these MOS devices under various configurations.

So, if I want a high gain fix it to Common source. I don't want a high gain, I want only the impedance matching, do Common drain or Source follower. If I want to go for moderate, low input impedance, go for Common gate structure, right. And therefore, you will be able to handle all of them independently in order to design it. With these words, I thank you for your patience hearing. Thank you!!!

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−30**
**Internal Cap Models and High Frequency Modeling-I**

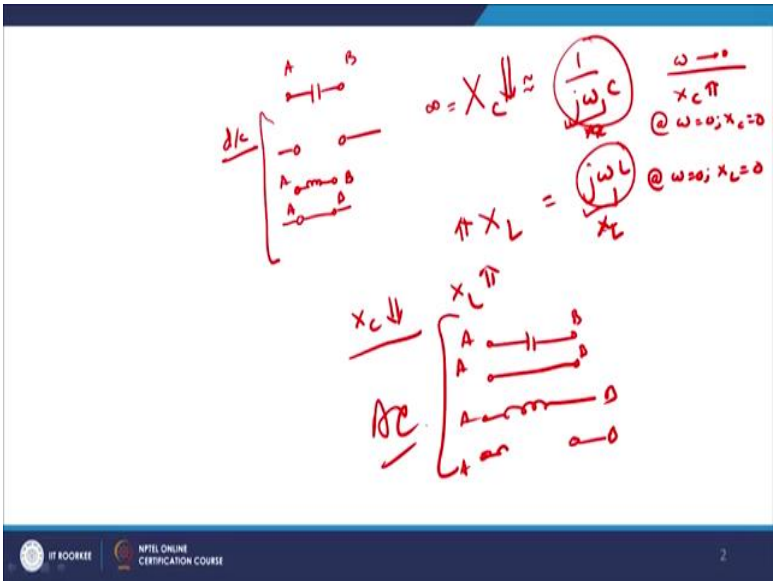
Hello everybody and welcome to the in NPTEL online course on Microelectronics: Devices to Circuits. We start with the new module today, which is basically on internal capacitance modules and high frequency modelling, right. So, what we have done till now? We have looked into various amplifier configurations using MOS device and we saw that various configurations have their own advantages and disadvantages in terms of working principle and in terms of output parameters being extracted.

However please understand, that these signals, the input signals were restricted to very small signal to maintain linearity and we were also assuming that the MOS devices actually biased in the saturation region of operation, it is neither in the cut-off nor in the triode region or nor at the edge of these regions right, so they are somewhere in the middle of the saturation region.

While this is good for working principles but generally the MOS device for example has to work at relatively high frequencies of operation because finally you will have to assimilate these MOS amplifiers in an analog scenario or in a mixed signal scenario where your frequencies of operation may be of the few MHz to few GHz.

Under such a scenario, your device should work properly at such high frequencies and that is the motivation behind this lecture, that this lecture will give you an idea about what are the device problems at very high frequencies of operation, input frequency and to understand the appreciate that we need to therefore understand the various capacitances associated with the device which become prevalent at very, very high frequencies. We will show you, I will give a reason why frequency is an important concept in a capacitance or inductance.

(Refer Slide Time: 02:27)



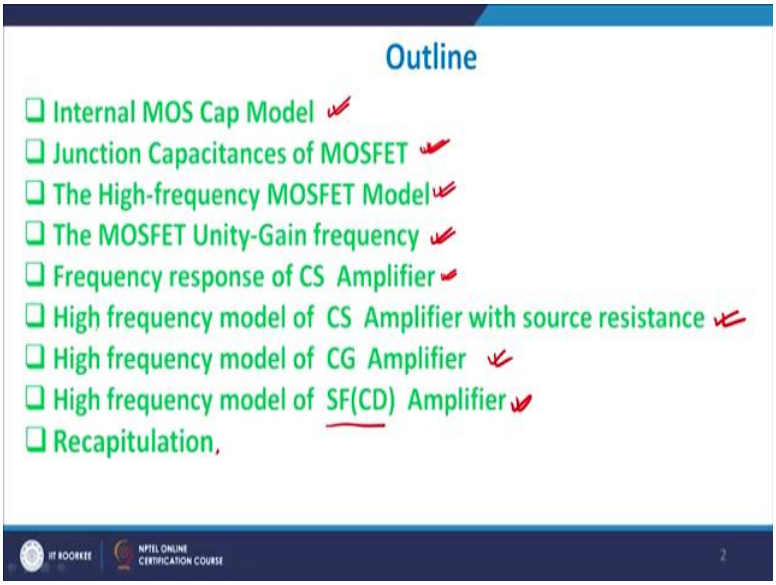

All of you must be aware of the fact that if you want to find out the capacitive reactance it comes out to be equals to 1 by jwC whereas for inductance it is equals to jwL, right this jw is referred to as $X_C$ and this is referred to as $X_L$ right, sorry yes I am sorry this is not jw, this is referred to as $X_C$ and this is referred to as $X_L$. Now if you look very carefully in case of capacitances if w tends to 0 my $X_C$ becomes very large and therefore I can therefore say that whenever you are doing a DC analysis right whenever you are doing a DC analysis, you can just keep your capacitances you need to.

So, if you are doing DC analysis this is point A and point B and you are doing a DC analysis then the equivalent circuit will have something like this, this will be acting as an open circuit why because at w equals to 0, $X_C$ equals to infinity, right. So, at w equals to 0, $X_C$ equals to infinity. Similarly, at w equals to 0, $X_L$ is equals to 0, so what you do is? When you do a DC manipulation or DC analysis you keep your capacitances open but you if you have an inductor between point A and point B right, if you have an inductor right what you do is that you then short it, right when you do a DC analysis.

But (little) so, but let us look what happens in AC analysis. So, when the frequency increases $X_L$ also increases but $X_C$ starts to drop down, so at relatively large values of frequencies where w is relatively large $X_C$ can be actually reduced to very, very small quantities and $X_L$ can be reduced to very high quantities, so $X_L$ will be relatively very large. So, if I want to find out therefore the cap modelling at high w or a high frequencies this then will be shortened right and what happens to my inductor? Inductor is let us suppose A, B then under AC bias I can I can make assume it to be an open circuitry, so this is under AC.

So, that is the reason that when you go for high frequencies, the points in the circuit which would initially made it made to be opened, now starts to become closed right and therefore access current or movement of current changes and therefore it is totally a different ball game altogether. So, we will be looking into this aspect as we move ahead and to understand therefore the first part of the lecture is that, we will be understanding the internal capacitance model. So, as you can see here the topic is internal capacitance models and it is high frequency modelling, right. So, let us see what the first, what are the outlines of the talk?

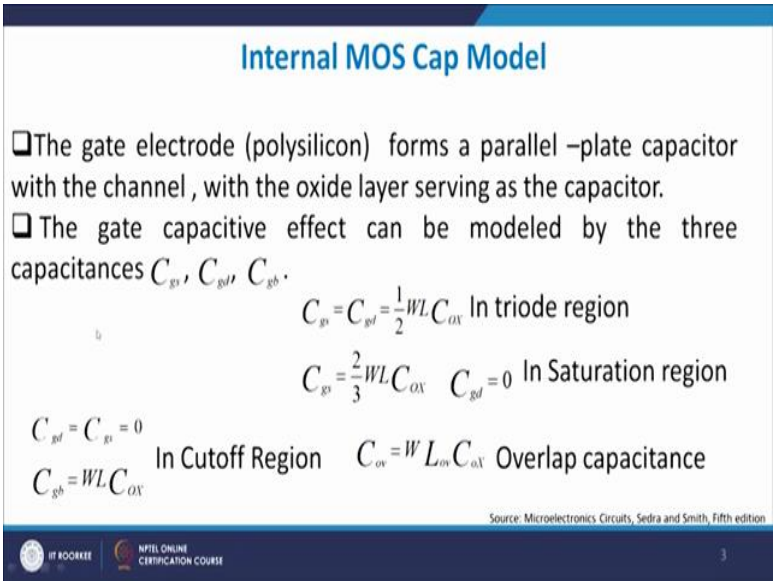
(Refer Slide Time: 05:11)




We, will first talk about internal MOS capacitance model, we will look at the junction capacitance of model of the MOSFET, right. We will look at the high frequency modelling of MOSFET, after we understood high frequency we will be coming into MOSFET unity gain amplifier or a unity gain frequency, what is the meaning of unity gain frequency? Then, frequency response of CS, we have already studied CS in our previous interactions, where CS with source resistance, CG and then SD also known as source follower right and then we will recapitulate.

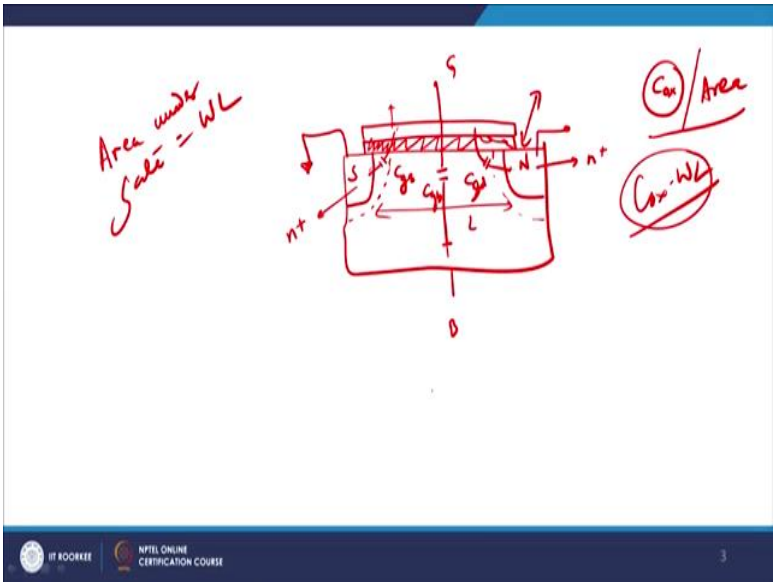
So, we will do high frequency modelling for all the amplifier design which you have already (let) left. And to do that, certain important characteristics of amplifier should be clear to you as we move forward. Let us look at the internal MOS capacitance model, right.

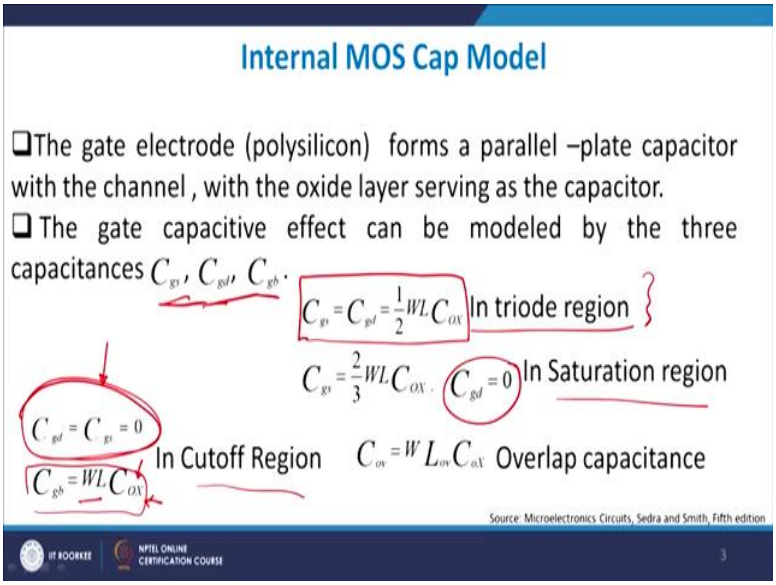
(Refer Slide Time: 05:52)




As you can see primarily, if you look at the MOS device, right.

(Refer Slide Time: 05:58)






It is basically and it is divided into majorly into three or four important capacitance models. So, let us suppose this is my source and drain right and I have an overlap here, right I have a diffusion capacitance here and I also have a metal capacitance, so this is my gate and then I have my bulk, let us suppose bulk so forget the timing, i have bulk. So, I will have a depletion region here, I will have a depletion region here this I give it positive bias, this I am grounding it for all practical purposes.

This is a typical structure of a MOS device, right this is the typical structure of a MOS device. Now, if you look very carefully there will be always a gate to source capacitance, gate to source and this is $C_{gs}$, there will be also source to drain capacitance $C_{gd}$, right. So, this is $C_g$ sorry, so this is $C_{gs}$, $C_{gd}$ and you also have $C_{gb}$ why $C_{gb}$? Gate to bulk, so you have an gate to bulk, no capacitor $C_{gb}$.

What is the origin of these capacitances see? Gate to sources primarily because what is the basic definition of capacitance that if you have got two charges separated by a distance we refer to that as a capacitance. So, if you look at this figure here, the gate is always heavily charged because it is a metal gate and this is highly doped, so $N^+$ region. Similarly, this is also $N^+$ region, so two $N^+$ regions separated by a dielectric here is primarily a capacitance, right.

So, I have as $C_{gs}$ here and I have a $C_{gd}$ here, I have $C_{gb}$ here. So, there are three capacitances which is available to us, let us see how it works out when you are in the triode region, in the saturation region and in the cut-off region, right. Cut-off is the most easiest one so we can we can deal with it much earlier or much faster, please assume or please understand that you also

have a $C_{ox}$ or the oxide capacitance per unit area, what does it mean? It primarily is basically the area under the so what is the area under the? Because if this is your length, if this is your length of the channel then W is in the inside the inside the board.

So, the area under the gate is actually goes to W into L, if $C_{ox}$ is the oxide capacitance per unit area then the total capacitance will be equal to $C_{ox}$ into W into L, fine. So, this is what I was explaining $C_{ox}$ therefore is the oxide capacitance per unit area right and that is what we were trying to see. If let us look at what happens in the triode, saturation and cut-off region.

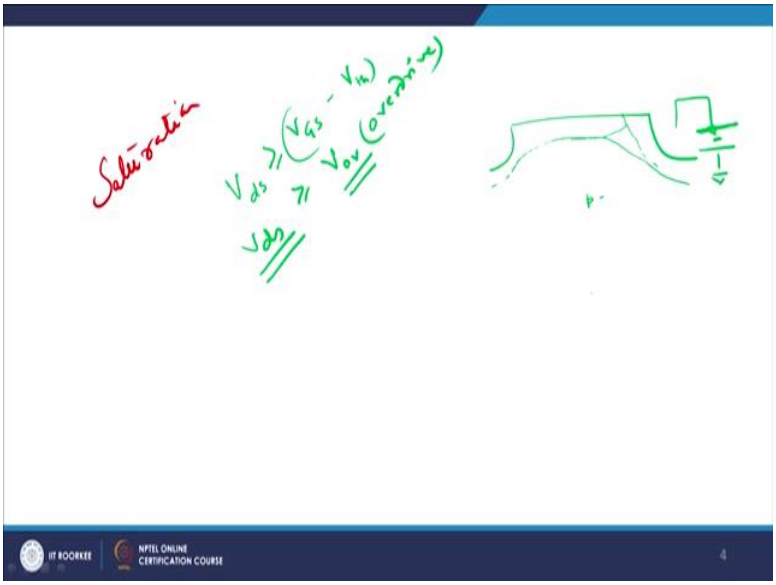
In the cut-off region your $C_{gd}$ and $C_{gs}$ are both equals to zero, right and the reason is very simple cut-off means you do not have any inversion charge, right when there are no inversion charges available your gate voltages are relatively very, very low and therefore gate to source and gate to drain almost equals to zero, almost equals to zero they are not exactly goes to zero they are very close to zero.

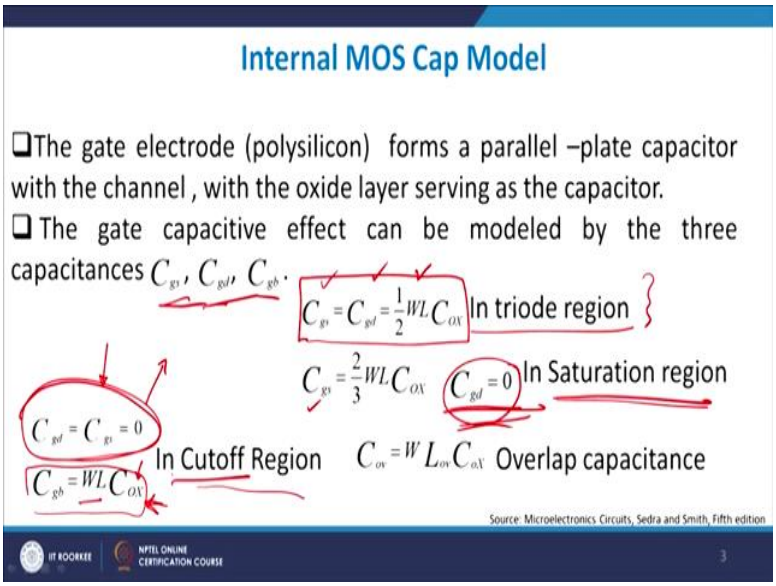
Whereas gate to bulk, you will always have a charge, depletion charge is always there and therefore that will result into a $C_{gb}$ of approximately W into $C_{ox}$, right this is what you get when you get $C_{GB}$ right and this is what you get when you have cut-off region. So, when you have cut off region, you do have this $C_{gd}$ and $C_{gs}$ is equals to 0 and $C_{gb}$ is equals to W into L width into length of the transistor multiplied by oxide capacitance per unit area.

What happens the next subsequent leaves the triode region? You have that $C_{gs}$ and $C_{gd}$ is half of WL into $C_{ox}$, which means that this is quite interesting and there is a debatable topic but in reality it is a safe assumption that whatever oxide capacitance you are having the can be divided into two parts source and drain.

In the being the triode region please understand your drain voltage is very, very low, source is already grounded, so the amount of charge deposited at source and drain will be almost equal to each other and therefore what we say it is half WL into $C_{ox}$. In the saturation region whereas $C_{gd}$ equals to 0, let us see why is it equal to 0. In the saturation region remember (what is the) what was the concept of saturation region?

(Refer Slide Time: 10:26)




The saturation region, if you remember saturation region if you point out, right in the saturation region always $V_{ds}$, right $V_{ds}$ sorry I will just come back to you $V_{ds}$ is greater than equals to $V_{gs}$ minus $V_{th}$, this is the condition for saturation which means that the $V_{ds}$ should be at least greater up this is also referred to as V overdrive, right. So, my drain to source voltage should always be larger than the overdrive voltage, fine.

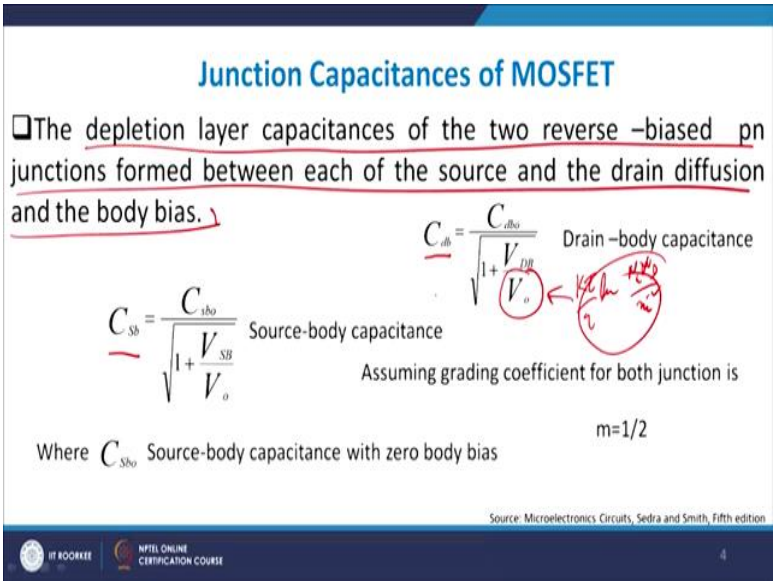
Which means that, the my $V_{ds}$ should be typically very large but understand unlike in the previous case in this case drain in always sorry, it not grounded but I give a positive bias which means that if this is an n channel MOSFET then then this is this is p type and then if this is source, I will have a depletion region is something like this initially, right you go on

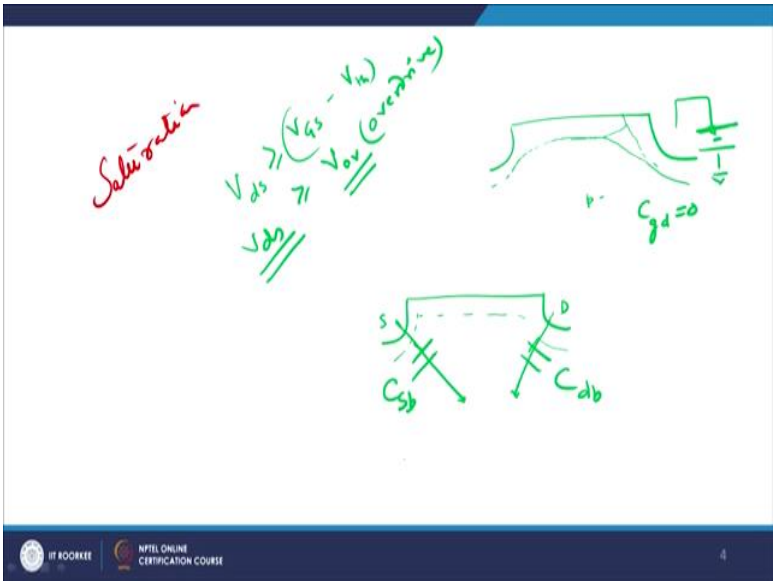
increasing the depletion region and what will happen is that depletion region will eat away into so this depletion region it away into the channel.

So, there will be no channel formation here, so there will be no charge, no charge primarily means you that $C_{gd}$ is exactly equals to 0, that is what I am trying to say you here but $C_{gd}$ is equals to 0 in saturation, right. And $C_{gs}$ is 2 by 3 W L C oxides, you will ask you whether the one third C oxide goes? Well, that is a debatable topic but typically bulk takes the maximum value of the of the one third of your W by L, C oxide.

Which means that under saturation triode and cut-off region you do have the change in your overall capacitance? So, even if your bias you and you and you let the bias point move from saturation to cut off and vice versa then you will see the capacitors will go on bearing, right. In the triode region, what we see? We get $C_{gs}$ equals to $C_{gd}$ equals to 1 by 2 half W L oxide and then $C_{gs}$ equals to 2 by 3 W L C oxide into C oxide where $C_{gd}$ equals to 0 for this thing. In cut-off region, both my $C_{gd}$ and $C_{gs}$ are equals to 0 and all of the oxide is going to gate to bulk region for operation, fine.

(Refer Slide Time: 12:38)





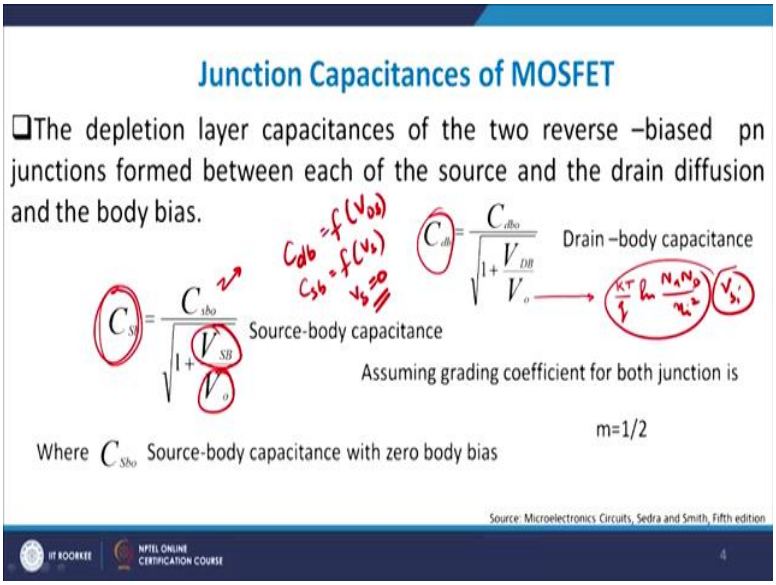

This is the basic model of a MOS capacitor. The third or the fourth type of junction capacitance which you feel is basically, please remember as I was discussing with you just now that your source and drain region, right remember is always sort of a reversed biased at least the drain region is reversed bias which means that there will be always a depletion capacitance at this particular point, right.

So, the depletion layer capacitance is of the two reversed bias pn Junction formed between each of the source and the drain diffusion of the body bias is defined as my $C_{db}$ or drain to bulk. So, I have $C_{sb}$ and $C_{db}$, $C_{db}$ is drain to bulk and this is source to bulk, right. If you go back it is drain to bulk and then source to bulk, so we will have one capacitance here 1 capacitance, so this is $C_{db}$, this is $C_{sb}$ and this will be primarily a junction capacitance which will depend upon the drain diffusion and the source diffusion. So, $C_{db}$ is given as $C_{dbo}$, $C_{dbo}$ upon 1 plus $V_{db}$ by $V_0$, $V_0$ is basically your applied.

So, let me see $C_{dbo}$ is the 0 bias drain to bulk, so (when he did not) when you do not apply any bias of the drain side you still will have a depletion capacitance between this drain and the bulk, right even when you do not apply any external bias there, right it will be still 0 when it is still 0 you will see that $C_{db}$ will be equal to $C_{dbo}$ upon $V_{db}$.

So, $C_{dbo}$ is the 0 bias drain to bulk capacitance, $V_{db}$ is the (drain to bulk capa) drain to bulk voltage which you are giving which you see and $V_0$ is basically the built in voltage which you see, this is a bit built-in voltage depending upon the this thing. So, it is basically K T by q is K T by q by ln of $N_a$, $N_i$, $N_d$ upon $N_i$ square this is what is equals to $V_0$.

(Refer Slide Time: 14:50)

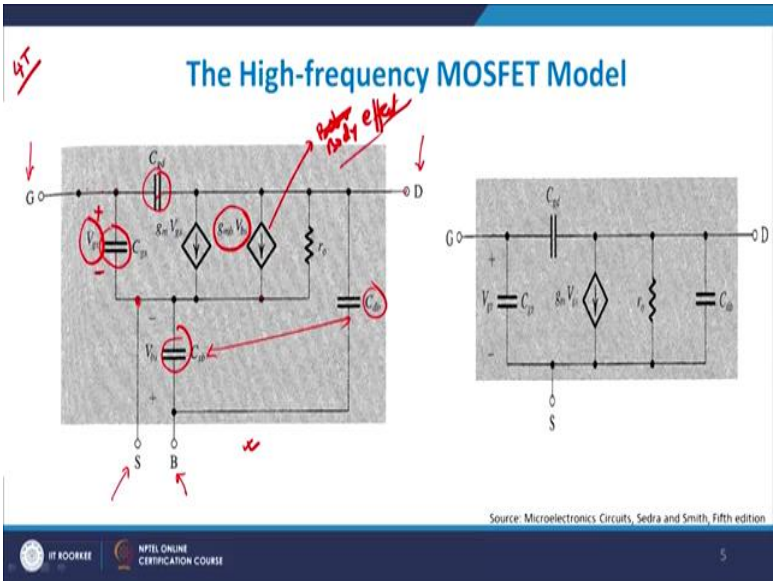



So, $V_0$ is basically as I discussed with you just now, $V_0$ is the built in voltage it is K T by q ln of $N_a$, $N_d$ upon $N_a$ square and this is basically a built in voltage and this is the minimum voltage which you need to overcome. Similarly, when you do source to bulk it is sb zero, it is a zero bias source to bulk capacitance $V_{SB}$ is the applied source to bulk voltage right and $V_0$ is the built in voltage which is there with us.

So, your $C_{db}$ and $C_{sb}$, please understand the values of $C_{sb}$ and $C_{db}$ depend upon the value of course the reverse bias which we do. Typically, source will be grounded, so this $C_{sb}$ is almost fixed, you know they do not vary with bias but $C_{db}$ varies with the bias because your drain to bulk will always vary right unlike, so $C_{db}$ is a function of your drain voltage, $V_{DS}$ whereas $C_{sb}$ is also a function of $V_s$.

So, this $V_D$ and $V_S$ but since my $V_S$ is equals to 0 the value of $V_{sb}$ is almost fixed when you talk about junction capacitances, right. So, we are understood overlap capacitances and diffusion capacitances, we also learned about depletion capacitance is also known as junction capacitances and we saw that as you vary the biases the capacitance value also varies in in these particular cases. Let me now, therefore come to the high frequency model of a MOSFET and that is pretty interesting or important.

(Refer Slide Time: 16:25)




So, now if you look very carefully look at the left hand figure here, again as I discussed with you, you have got four terminals gate, you have drain right, you have got source and you have got bulk. So, it is basically behaviour it is basically a four terminal device behaviour right out of which if you look very carefully as I discussed with you gate to sources, so when you apply your gate to source voltage you apply between gate and source, so this is your $V_{gs}$ where this is positive and this is not.

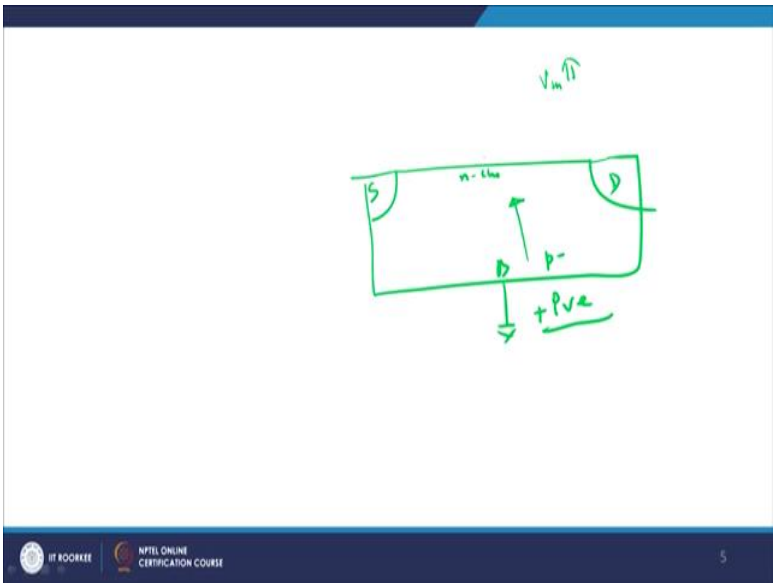
So, this is my source point, so the difference between this and this is by $V_{gs}$ and therefore the capacitance $C_{gs}$ appears between gate and source. Similarly, $C_{gd}$ appears between gate and drain and $C_{sb}$ is between source and bulk and $C_{db}$ is between drain and bulk, so these two are actually your depletion capacitances and these two are actually your diffusion capacitance which is visible to you.

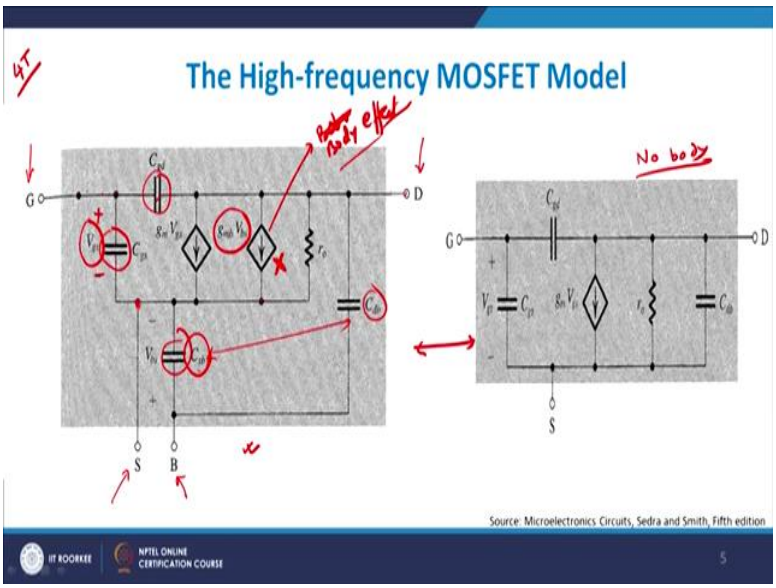
So, therefore if this this $g_m$ times $V_{gs}$ is nothing but the device itself, the channel current because $g_m$ is transconductance is $\partial i \, \partial V_{gs}$, so when you multiply with $V_{gs}$ I get the current. So, therefore I am showing it by the ideal current source here and it is basically $g_m V_{gs}$ and as you can see it is between drain and source, this is my source and this is your drain, so this is a drain and source.

This is also a current which is virtually between drain and source known as the bulk $g_{mb} V_{bs}$, this is because of the body effect, sorry body effect, right. At the body effect why? Because you always have a $g_m$. So, apart from $g_m$, so assume that you do not have any body effect then this you will not have this this current source. So, under so but if you have a body effect

by virtue of the fact that threshold voltage and therefore there is an excess current flowing. You have g $_{mb}$ times V $_{m}$ why this is true? I will just give you a small brief idea why this is true it is something like this.

(Refer Slide Time: 18:23)



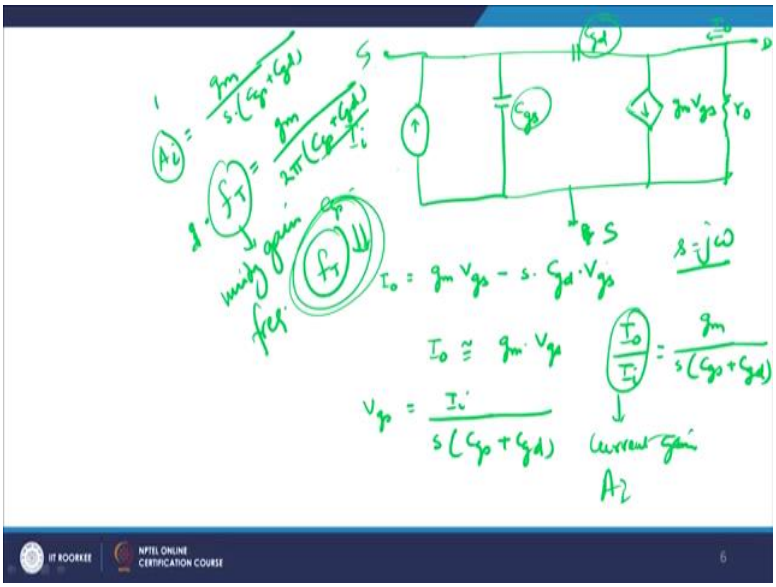

So, if you remember in our previous discussion when we are doing if this is my bulk and this is my source, right and this is my bulk and drain, if I vary the bulk potential and this was initially p type you want it to be n channel, n channel this is in channel, so this is p type and then you apply suppose let us suppose a positive bias on the bulk. It primarily means that it will push the holes towards the side and therefore you do not aid your channel and therefore your threshold voltage will actually rise because you require therefore a larger gate voltage to pull electrons near the surface, right.

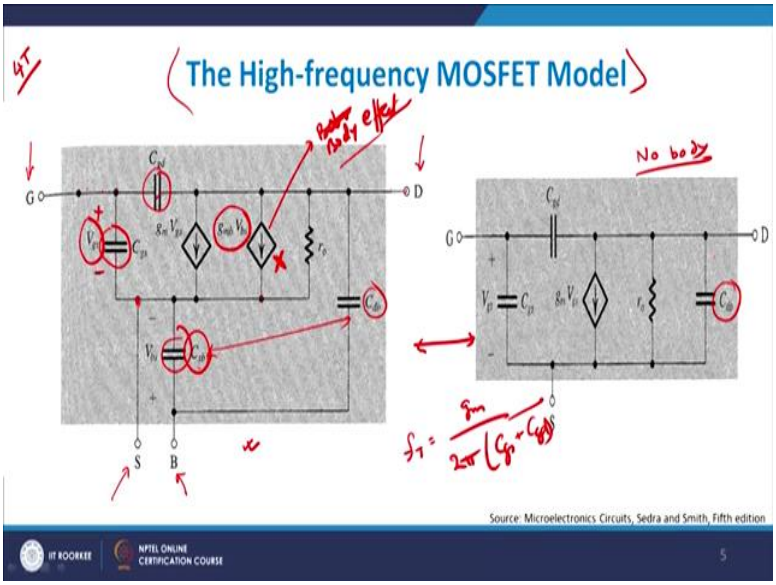
As a result, what will happen is for the same amount of gate voltages earlier your current will fall down, right that is what is known as $g_{mb}$ a $V_{BS}$, right. So, that current which is there between source and drain by virtue of the fourth terminal base current fourth terminal is basically referred to as $g_{mb} V_{BS}$, right. So, this is the current on the fourth because of the fourth terminal.

So, this is what you get as a fourth terminal $r_o$, as I discussed with your previous term $r_o$ is but the due to channel length modulation effect the resistance offered is known as $r_o$, right. So, this is basically the high frequency model of a MOSFET, this has been if you look at the right hand side, this one has been done with everything else remaining the same but I have removed $C_{sb}$ because as I discussed with you $C_{sb}$ does not vary with bias, so let me remove it makes my life easier and I have $C_{db}$ only here, i have also removed $g_{mb} V_{BS}$ here.

So, there is no body, so no body effect so there is no body effect on the right hand side and this is what you get finally from your from your this thing from the from the understanding purposes. So, with this let me give you a brief idea about, let me see how I can work out that I do have.

(Refer Slide Time: 20:15)






Let me say, I have a high frequency model and I have a current source here $I_i$ and then I have $V_{gs}$ and $C_{gs}$ being available to us, so this is $C_{gs}$, right I have $C_{gd}$ here, $C_{gd}$ right and then you have the current source which is basically your $g_m V_{gs}$ here and then you have $r_o$, right and then this goes to bulk and this is your grounded, let us suppose bulk is grounded and this is your drain and this is your, so this is the $C_{gs}$.

So, this is a gate and source available here, right or let us suppose this is not bulk, it is supposed this is a source and therefore this is gate here and then $C_{gd}$ gate to source, to gate to source, gate to drain you have got $r_0$ and this is your current source here. So, if you current is $I_0$ is flowing a typical two port Network, current is assumed to be flowing inside the device, so what I say? Whether I can right down is $I_0$ equals to $g_m$ times $V_{gs}$ minus s times $C_{gd}$ multiplied by $V_{gs}$, right.

If we look from this side it is $g_m$ times $V_{gs}$ current and then $C_{gd}$ multiplied by $V_{gs}$ into s is this current, $I_0$ is therefore relatively equals to $g_m$ times $V_{gs}$, this is a very simple straight forward so I can write down $V_{gs}$ to be equals to $I_i$ upon s times $C_{gs}$ Plus $C_{gd}$. So, I can write down $I_0$ by $I_i$ the ratio of output current which is basically the current gain to be equals to $g_m$ upon s times $C_{gs}$ Plus $C_{gd}$, right where s equals two therefore s is basically equals to jw.

So, I get this is basically my current gain, current gain which is also referred to as A I, so I can refer to as A I therefore A I is equal to $g_m$ upon s times $C_{gs} + C_{gd}$. Now, A I if so let us suppose I want a unity current gain, right so A I will be equals to 1 right and therefore I can safely write down to, so s if I take this if (s equals to) A I equals to 1 I take s equals to this

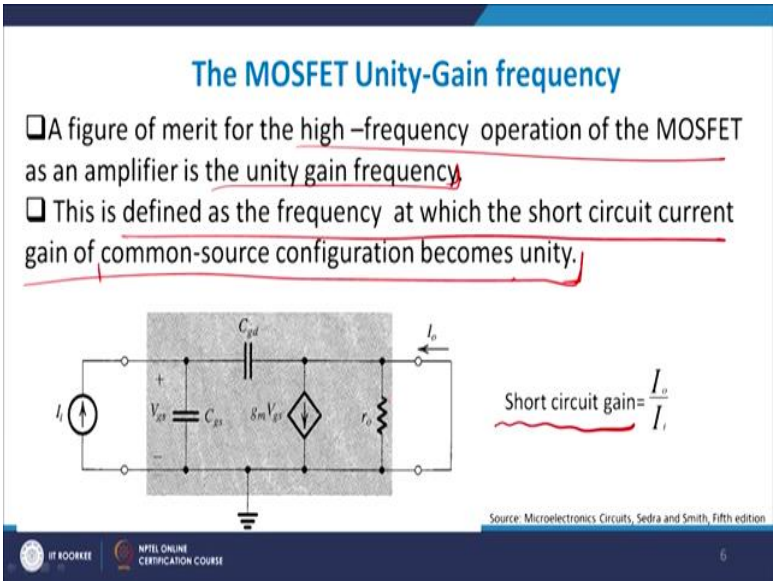
side and now then replace s by (2) w by 2 π F then I get $f_T$ is equals to $g_m$ by 2 π $C_{gs}$ plus $C_{gd}$, this is referred to as unity gain frequency.

What is the meaning of unity gain frequency? It tells me that, that frequency at which the ratio of output current to input current will be approximately equals to 1 will be referred to as a unity gain frequency and it depends upon the transconductance and it also depends upon the value of $C_{gs}$ in $C_{gd}$. So, you see if you increase the value of $C_{gs}$ and $C_{gd}$ your $f_T$ actually starts to drop down.

So, you cannot operate at very high frequencies, right you if you want the unit again device for whatever reasons for example for matching your impedances and you want a unity gain device then if your gate to source of gate to drain capacitances becomes large then you end up having a lower $f_T$ which means that you operate at a lower $f_T$ or a lower frequency domain, right.

So, that is quite it and that is what I was saying that capacitance model tends to change your frequency of operation drastically that is what I wanted to just remind you in from this observations or from this basic idea, right. So, with this we come to the high frequency model once again and we saw this thing and we just now saw that my $f_T$ will be equal to $g_m$ upon $2\pi C_{gs}$ plus $C_{gd}$, right and you see the depletion capacitance which is $C_{db}$ does not play a role as far as unity gain is concerned, right. It is only the gate to source and gate to drain capacitances which play a play important role that is what I was coming to therefore.

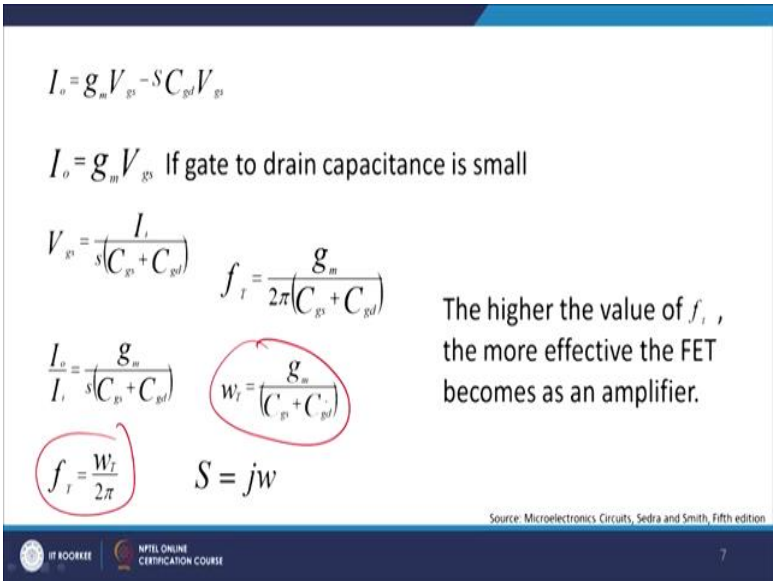
(Refer Slide Time: 24:25)




Therefore, one of the figure of merit for the high frequency operation of the MOSFET is the unity gain frequency. So, this is defined as the frequency at which the short circuit current gain of a common source configuration becomes unity right. So, whenever you short circuit gain means? We the output is basically shorted, so and therefore the current $I_o$ is flowing by virtue of this MOS device.

So, by short circuit current is grounded and therefore in a common source configuration $I_0$ by $I_i$, $I_0$ is the output current, $I_i$ is the input current if you divide then that we define that to be the short circuit gain and we just now saw this to be equals to this whatever depending on the transconductance of the device as well.

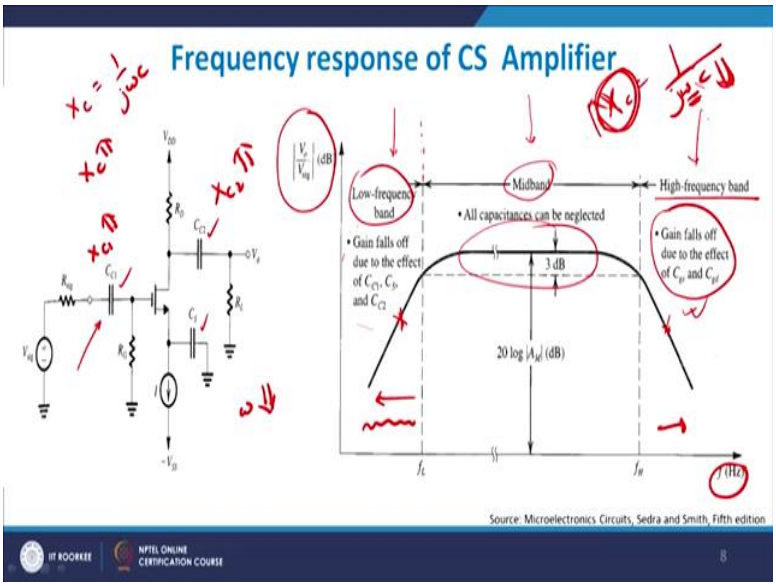
(Refer Slide Time: 25:06)



So, this is basically your MOSFET unity gain amplifier as such. As you see therefore, I have already derived this where this value and as I discussed with you $f_T$ if you look and therefore w T if you find out is $g_m$ by this thing, so f will because 2 $g_m$ by 2 $\pi$ $C_{gs}$ plus $C_{gd}$. So, I will not go into details we have already discussed this point in details for this case. So, it becomes a much better amplifier right it becomes a much better amplifier as such in this case, ok.

Let me come to the frequency response of the common source amplifier, right. It quite an interesting study, frequency response primarily meaning means that if you are able to plot the variation of gain on the y axis.

(Refer Slide Time: 25:49)




So, you have gain on the y-axis, so $V_0$ by $V_{sig}$ is my gain, $V_{out}$ by $V_{sig}$ is my gain and then the x axis, you have do have the frequency then we define this to be as a frequency response, right. We generally divide the frequency response into three major component, one is the low frequency band which is therefore below a particular set of frequencies we define this to be the low frequency band, anything above a particular frequency is defined as high frequency band and somewhere in the middle we define this to be as the mid band.

So, typically we have three components or three frequency components of a common source amplifier, the first is basically the low frequency band, the second is high frequency band, third one is a mid-band, right. It has been seen for almost all the amplifiers that the mid band frequencies is almost constant independent of the frequency applied right, so you have a large range of frequencies across which you would expect to see almost constant values of a gain

right and typically if you operate under such a criteria or certain conditions here then your gain will be independent of frequency.

So, you vary frequency as much as you want and your gain is stable and gives you a stabilized gain but if you some or other fall in this region or in this region then you again becomes a becomes a function of your frequency, right. So you vary a frequency or gain varies very highly nonlinear approach to your design, right. Let us see what happens and why this happens? You see I had discussed with you that $X_c$ is equals to 1 by j w c right, fine.

So, you see once you go to the low frequency domain somewhere here your w are relatively small, right, when your w are relatively small your $X_c$ are relatively very, very large so $X_{c1}$ is very high, similarly, $X_{c2}$ is also very high and so on and so forth. So, when these are very high this start behaving as a open circuit right, because the impedances are very high in this case.

So, if you concentrate only forget about everything else even and concentrate only on a $C_{c1}$ which is basically the coupling capacitor you know why, we had referred to this as coupling capacitor or a blocking capacitor? The reason being that I do not want any DC bias of $V_{sig}$ to appear on to the gate side of my MOS device, right that was my main aim and that is the reason you are putting a coupling capacitance here.

Once you put that a coupling capacitor then at lower frequencies of w the resistance offered by the $C_{c1}$ is typically very high and therefore most of the $V_{sig}$ does not appear on the gate side of the MOS device and therefore again starts to fall down, right. So, it is $X_{c1}$ starts to behave like a more and more like an open circuit and as a result the gain starts to fall down as you lower your frequency from a particular value and that is the that is what is written here that the gain falls out due to the effect of $C_{c1}$, $C_{c2}$ and $C_s$.

So, I have this, this and this all starts to behave like an open circuit whereas at very high frequencies your depletion capacitances they become almost very, very large at very high frequencies, so and therefore as the frequency increases those values become large and as a result the gain starts to fall down, right. As you can see therefore that at very high frequencies at relatively high frequencies $X_c$ again as I discussed with you will be 1 upon j w c.

Now, at very high frequencies this is very high agreed therefore $X_c$ has to be low but your C becomes so small your $C_{gs}$ and $C_{gd}$, gate to source and gate to drain it becomes so small that this $X_c$ capacitance is almost negligibly, negligibly large right or it becomes very large. So,

w is very small C drops down $X_c$ becomes very, very large and as a results as the gain starts to fall down at high frequency regions, right.

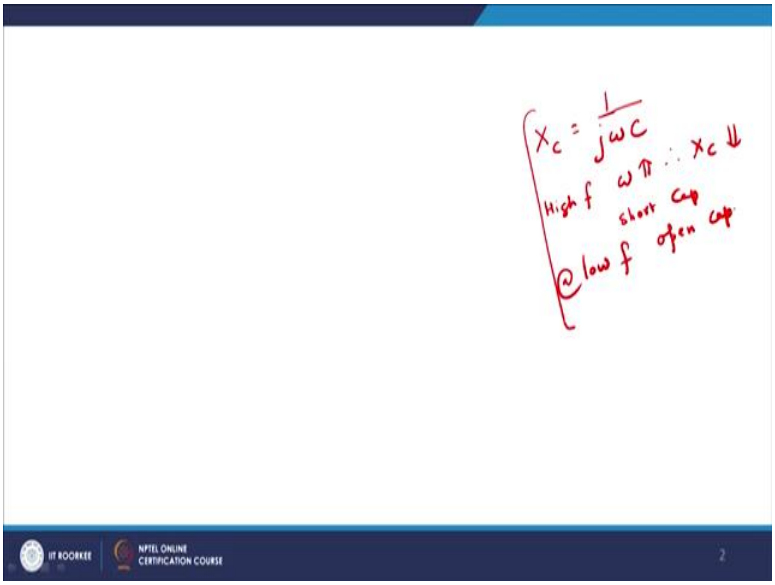
So, I just started with this thing, maybe in the next interactions we will carry forward with where we left today and give you a basic idea about the frequency response of amplifier what. We learnt today? we learned about the high frequency modelling of a MOS device, of a MOS amplifier, we also learnt how to design or how to predict a unity gain, unity gain current gain of amplifier on what factors does it depend.

We finally ended up, looking into the frequency response of a common source amplifier and we saw that mid band is the place where you should bias your device, so that it is independent frequency, at higher and lower frequencies the gain will fall down with respect to frequency. So, when the frequency increases or decreases your gain will fall down typically in high band, high frequency band and low frequency band respectively, right. So, with this let me stop here today and then when we meet next time we will start forward we will move forward ahead of this, is it ok, thank you very much!!

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−31**
**Internal Cap Models and High Frequency Modeling-II**

Hello everybody and welcome to the NPTEL online certification course on Microelectronics: Devices to Circuits. What we will be doing today is basically look into the high frequency modelling of the bipolar Junction transistor of the MOS devices. In our previous section, we have seen what are the various capacitive components which are there within the MOS device.

(Refer Slide Time: 00:52)

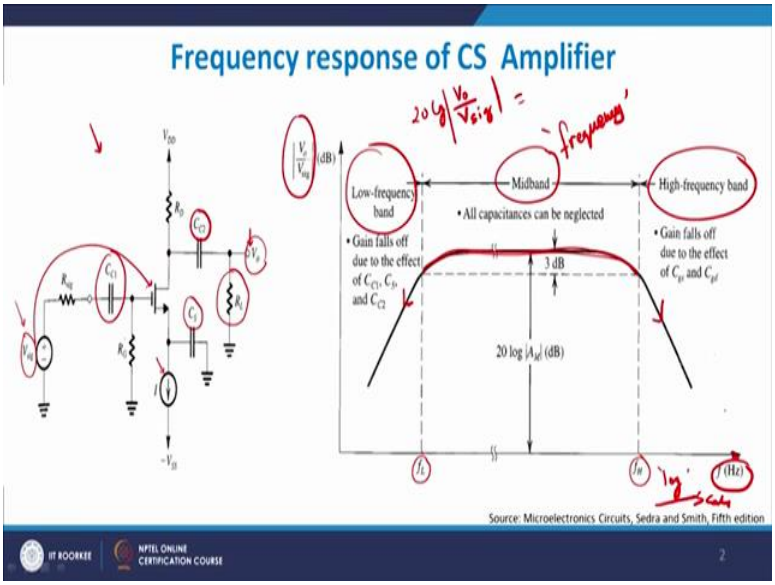



We also saw that, if you have a MOS device and if you have a capacitive model it can be predicted as a capacitance for example the overlap capacitance then it can be written as like this 1/jwc, so at high frequencies your ω, at high frequencies your ω is obviously very large which therefore your Xc are very, very low and therefore at high frequencies I can actually short all my capacitors, right short capacitors.

Where at low frequencies just as the name suggests, we will open circuit the capacitance, right. But this is only true for certain range of capacitances right, which of a certain range of frequencies. So, at very high frequencies you need to short it and certain open frequencies you need to close it. So, what we will be looking today is, that if I take a common source amplifier.

Let us, for example, MOS based common source amplifier and we have all those capacitances which is parasitic capacitances as well as bypass capacitances as well as blocking as capacitances then how does my amplifier behave is the main course of study at this module itself. So, let me come back to the come back to the slide itself and show you how it looks like overall our system looks like.

(Refer Slide Time: 02:14)




So, to your left if you look at this point is basically your $C_s$ based, common source based design of a common source amplifier and I have attached current source on to the source end and there is a $C_s$ which is the emitter sort of a or a source bypass capacitance, you also have a coupling capacitor in $C_{C2}$ here which couples the drain output to the output of the (pins) drain of the MOS device to the output $V_0$ and you also have a $C_{C1}$ which is basically the input coupling capacitances which connects your V signal input signal to the gate of my MOS device, right.

$R_g$ is the gate (capacitance) resistance, $R_d$ is the drain resistance and of course there will be also a $R_s$ but since it is in series with the current source $R_s$ will be negligibly small as compared to an output impedance of the current source and therefore that $R_s$ is missing at this stage and $R_L$ is the load resistance which you see here. Now, if you plot on the y axis V0 which is the output voltage divided by $V_{sig}$, $V_{sig}$ is basically the signal input voltage and if you make it in dB. So, 20 log ($V_{out}$ / $V_{sig}$ ) if you try to find out this will give you in dB, right and then you try to plot it with frequency on the x axis right and this is basically in log scale typically, this is in log scale, so you will have 1 Hertz, 10 Hertz, 100 Hertz, 10^3, 10^4, 10^5,

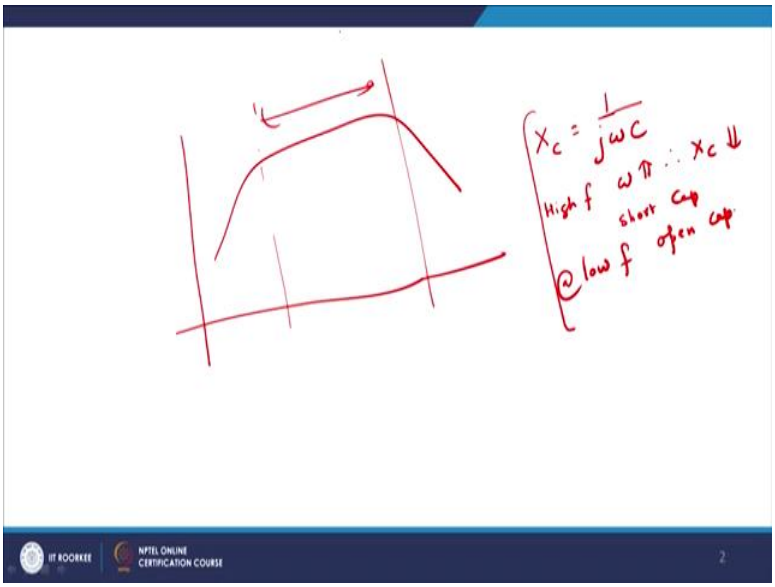
6, 7, 8 so on and so forth. So, this is, so you please understand your y axis will not be a linear scale it will be a basically a log scale right.

And your x axis will be a component, sort of a transfer function, which gives you output voltage by input voltage expressed in dB and mod of that basically. So, mod ($V_0$ / $V_{sig}$ ), in dB of that because at this stage we are not interested whether you have a phase difference or not, so we are just interested in looking to the magnitude a component available to you.

Now, if you look very carefully in this slide or in this overall approach, you will see that typically somewhere between $f_L$ which you see here, we will discuss this later on and $f_H$ your gain is almost constant see, so you are plotting gain v/s frequency and this gain is almost constant, right it is independent of frequency, it is independent of anything else, right this band is known as mid band, so mid band frequency right.

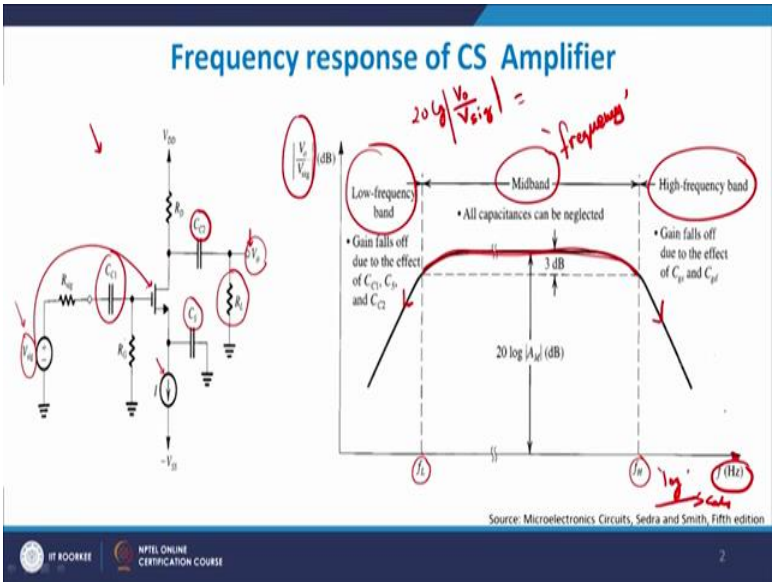
So, what is the mid band frequency? Mid band frequency is that frequency across which the output or the gain is independent of the frequency which you see, right. Now, beyond that particular point, beyond this $f_H$ and beyond $f_L$ as you go on increasing frequency or decreasing frequency your gain starts to fall down, right this is what is happening, right.

(Refer Slide Time: 5:17)



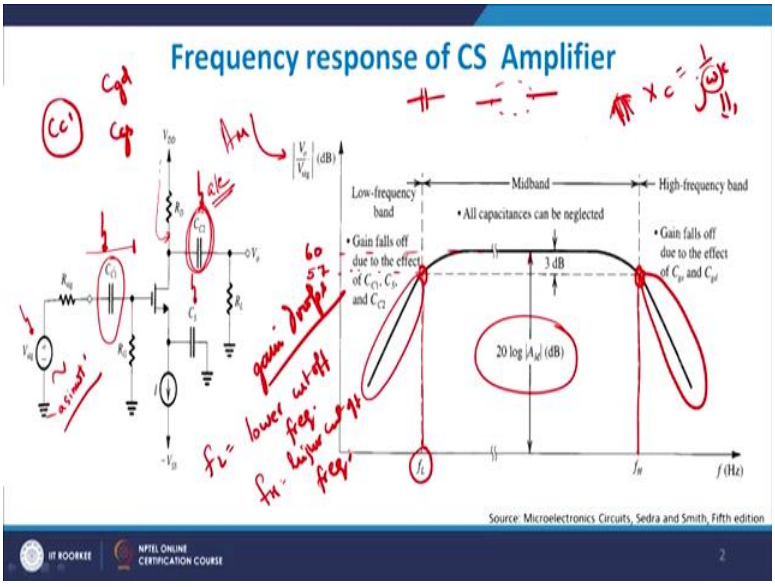

So, what we see therefore is, if you properly plot it is that you get a very steep profile here, you get a very steep profile here but somewhere in the middle you will have automatically a gain which is almost independent of frequency this is known as mid band, as I discussed with you this is mid band.

(Refer Slide Time: 5:32)




This is low frequency band and this is high frequency band right and a time might come that is it is large frequency or gain will be absolutely equals to 0. So, let us see why this happens or why you actually see a drop in the gain at very low frequencies and at very high frequencies?

(Refer Slide Time: 05:47)




If you remember, if you look very closely and if you remember your previous discussions you will appreciate that you will always have a coupling capacitor here and a coupling capacitor here and you have a $C_s$ which is basically a source capacitance. Now, why these capacitances were used? Initially they were used primarily so that for example, why $C_{C1}$ was

used? $C_{C1}$ was used primarily, so that any DC component of this $V_{sig}$, so $V_{sig}$ is basically an AC signal.

So, it is basically a sin(wt) but since it is derived from a DCs you do have a source available within itself, there will be a DC component always associated with V sig, right. You do not want that DC component to sit on the gate side of this MOS device otherwise it will change the bias point and that is the reason you put a $C_{c1}$ here, (a coupler) a coupling capacitor, in fact known as a blocking capacitor and as a result you only allow out the AC signals to go through and you block the DC.

Similarly, if you look at $C_{C2}$, this also allows only the AC signals to go through and block the DC right, say only allow AC signals to flow through and block DC, which DC? DC capping for example from $V_{DD}$, right coming from $V_{DD}$ this will be blocked by $C_{C2}$, so that is the reason but then let us see what happens therefore at low frequencies. Now, at low frequencies as I discussed with you $X_c$ was equals to 1/jwc, which primarily means that at low values of w, $X_c$ will be typically large right.

So, as well as this, right, its effective value is typically very large, the resistance offered by this is very large and therefore as you lower you start to lower your frequencies this start would behave more like an open circuit rather than short circuiting and therefore gain starts to fall down, the gain so your gain drops right, your gain drops why does your gain drops? The gain drops because of a simple reason that as you lower your frequency the 1/jwc goes on increasing $X_c$ goes on increasing and then the assumption that you are shorting it now you have to open it to a larger section.

And as you make it open obviously you can understand there will be no signal flowing and therefore your output will be equals to be 0. That is primarily the reason why the gain falls at the low frequency band, right. Let us see, let us look at the fact what happens at the high frequency band which is somewhere here more than $f_H$, right. If you look at more than $f_H$ part then remember there were two components parasitic component known as $C_{gd}$ and $C_{gs}$, $C_{gd}$ and $C_{gs}$ gate to source and gate to drain capacitances, right.

Now, at such high frequencies at this, what we were assuming? Was that these capacitances will not playing a role because overlap is small or they are minuscule, they are very small in dimensions but at such high frequencies of operation right the resistance offered by $C_{gd}$ and

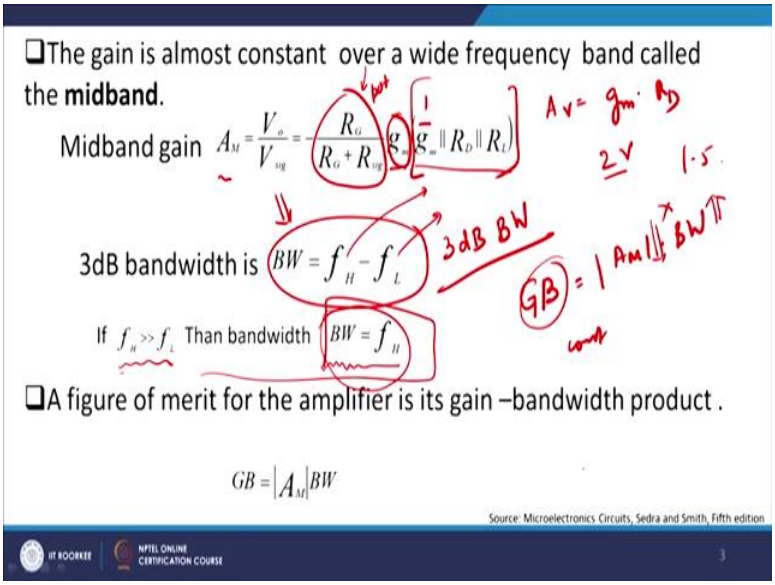
$C_{gs}$ though is relatively small but the gate to drain capacitors which is available to me has a extra effect on to the circuitry and gain starts to fall down.

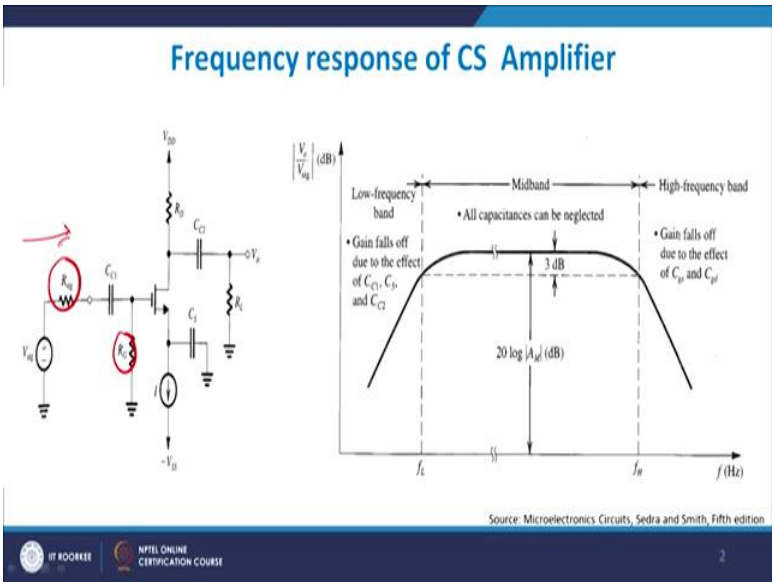
Which means that at such a high frequency though Xc is low for Cgd and Cgs right, which means that your gate and drain are not coupling properly in a sense and therefore gain starts to drop. So, gate starts to lose control over the channel, you know in a sense and as a result the gain starts to fall. So, the high frequency the gain reduction in is primarily by virtue of reduction in $C_{gd}$ and $C_{gs}$, right and your gain losing control over the channel and the source and drain side.

And on the low frequency part if you look very carefully it is primarily because of the CC1, $C_{C2}$ getting effectively shorted, so sorry effectively open and then when they open the gain starts to fall down, right. So, this is the two reasons why primarily you will see this so most of the characteristics of the amplifier circuits will look like this, this is 20 log $A_m$, $A_m$ is nothing but this is $A_m$, this is $A_m$, right 20 log ( $V_0$ / $V_{sig}$) in dB is defined as this point, right. Now, let me come to two important cross points in this characteristics and these cross points are this one and this one, right.

So, what you try to do is you just look at the top, suppose this is equals to say your 60 dB then what you do is? You try to go from here to 57 dB, so draw a wire which is 57 dB. So, 3 dB drop is there and then check out two points where you get the 3 dB drop then we drop the point across the frequency axis and we define fL as the lower cut-off frequency and $f_H$ to be equals to higher cut-off frequency, right. So, I have a lower cut-off frequency and I also have a higher cut-off frequency, the lower cut-off frequency is the point where the gain has dropped by 3 dB and higher cut-off is also again point where it is of 3 dB.

(Refer Slide Time: 11:25)




Now, the therefore the distance between the two in frequency domain is defined as my bandwidth, right also referred to as a 3 dB bandwidth. So, bandwidth is defined as difference between higher cut-off to lower cut-off, right and this is known as bandwidth. If $f_H$ is obviously much larger than $f_L$, so $f_H$ is 10 to the power 7, 10 to the power 8 and I feel may be how the hardly on order of 10.

So, if we subtract the 2 it is approximately goes to $f_H$. so, whatever we do the value of $f_H$ that will be the bandwidth of the input signal, right. Now, if you look at the mid frequency gain right, mid frequency gain which is basically your Am then it is $V_0 / V_{sig}$ and $V_0 / V_{sig}$ how did I find out or how did I come to know was that if you look at, so if you go back to your previous slide, sorry.

Let me just drop this point for clarity sake, right and let us look from this point of view, right. if we look this point $V_{sig}$ actually is the $V_{sig}$ is the voltage given by the signal but some portion of that goes $R_{sig}$, $C_{C1}$ and so on and so forth. The remaining falls across $R_G$ and that is the gate voltage given to this MOS device, right. So, if you look at, sorry so if you look at the next slide here you see $R_G$ / $R_G$ + $R_{sig}$ is the division which you do because it is a basically a voltage divider network.

So it is RG, so this is $R_G$ / $R_G$ + $R_{sig}$ is the is the voltage divider network which you see, this one, right multiplied by because this is the voltage divider network multiplied by $G_M$ x $G_M$ parallel $R_{Dx}$ $R_L$, right to why? So, it is it should be actually 1 / $G_M$ $R_D$ $R_L$, right. So, it is basically $G_M$ transconductance, so if you remember gain from $G_M$ times $R_D$ remember, so this is GM and this is your effective $R_D$ which you see but this factor comes from the fact that not all the signal voltage appears across the gate side.

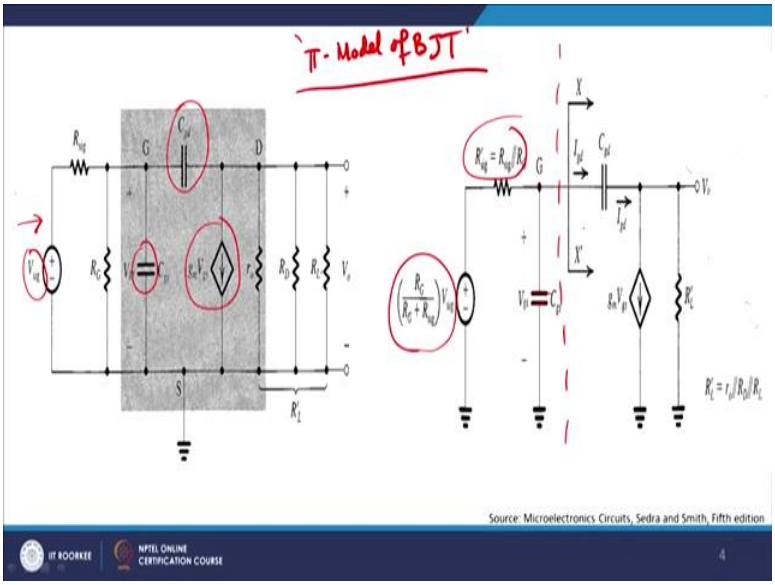
A part of it appears because of the potential divider technique because of potential divider technique, so if you are giving V 6 a is equals to 2 volts not all the two volt will appear at the gate side may be 1.5 peak to peak will appear though, what is the reason for that? The reason for that is that you do not want large signal distortion to play into picture.

So, if your input is very, very large peak to peak then you might end up having a nonlinear distortion in your system, right and as a result what will happen is that it output will be heavily distorted that is compared to your (())(13:59) and as a result you will never get a linear profile which is there with you. So, I discussed with you just now that a bandwidth is defined as $f_H$ minus $f_L$, since $f_H$ is many orders larger than $f_L$ we define that bandwidth can also be written as $f_H$ just the value of $f_H$ itself.

A very important figure of merit which we typically use an amplifier design is defined as the gain bandwidth product, we also refer to as gain bandwidth GB and it is referred to as $A_{m\ x}$ bandwidth and this is generally constant. So, what you do is? That if you want to therefore make your bandwidth high, you have to reduce your gain, so that the product of these two remains constant, right.

And that is the problem for any amplifier that if I want my gain to be high my bandwidth will be restricted and if I want my bandwidth to be high my gain will be restricted, you cannot take both of them together and make both of them high, right, and that is the sort of a price you pay for designing this basic cons, this problem area, right.

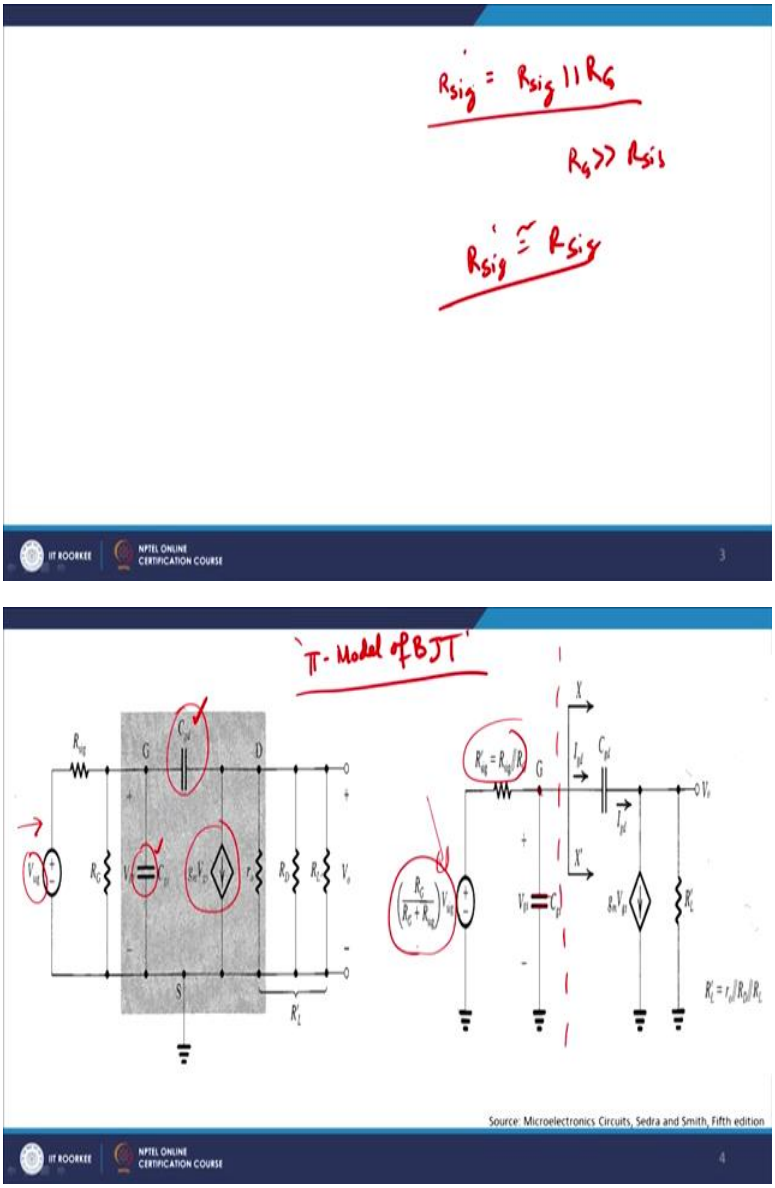
(Refer Slide Time: 14:59)




With this knowledge let me do a mathematical treatment for high signal frequency and low frequency response. So, if you look here (click) clearly for the high frequency component if you remember in the previous, I have just replaced the MOS device with it is equivalent circuit model pi. So, this is basically a pi model of BJT is used right, of BJT is used and if you look carefully it is C gd here and here what $C_{gs}$ here, we have got current source $g_m * V_{gs}$ and $r_0$, $R_D$, $R_L$, right.

So, if you look from this side from the left hand side which is the input side then a part of basic appears on the gate side given by this formula, so $V_{sig}$ x ($R_G$ / $R_G$ + Rsig ) is that effective value then what we do Rsig' equals to Rsig because these two will be in parallel, so $R_{sig}$' will be equal to $R_{sig}$ || $R_G$ and then the same voltage is fell into the gate side which is internally having a $C_{gs}$ gate to source voltage which is available with us and this is the input side so, fine.

(Refer Slide Time: 16:12)

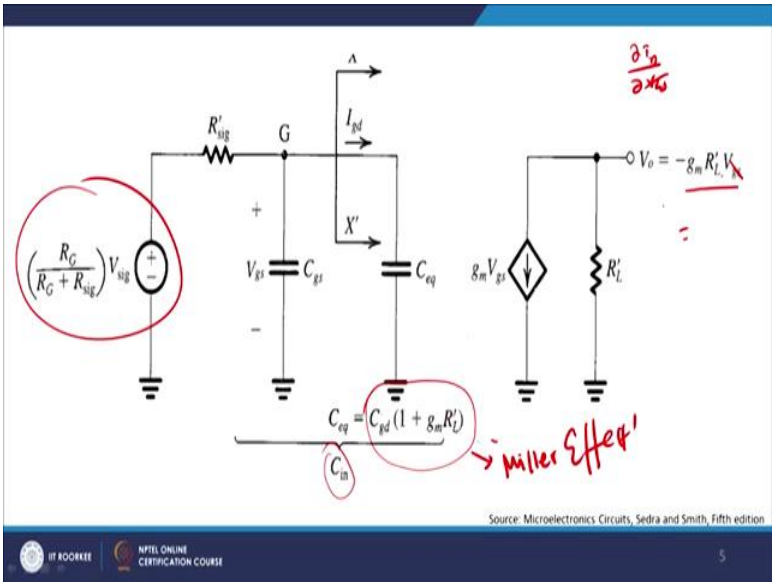



And therefore, what people have found out over the years is that $R_{sig}$', right $R_{sig}$' will be approximately equal to Rsig || $R_G$ but since $R_G$ is much larger as compared to $R_{sig}$, what we safely can write down is $R_{sig}$' is approximately equal to $R_{sig}$ which means that if you go back to a previous, this slide then $R_{sig}$' which is the potential or the resistance offer to the gate side is exactly almost of the offered by this structure right and that is what we get from here in this case, right, ok so, this is what we get.

So, if you look at the pi model therefore from therefore looking from this side you what you can do it is quite an interesting summation that, so when you are doing a high frequency modelling or we needing a sorry, a low frequency modelling $C_{gd}$ will come into picture $C_{gs}$

will come into picture and therefore they will change the output characteristics of the device, right.

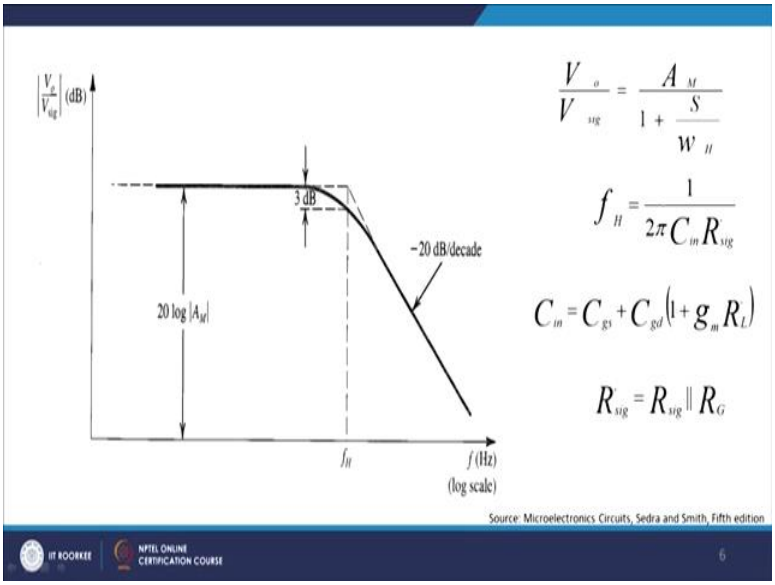
(Refer Slide Time: 17:07)




Now, if you look very carefully then what happens is this is, the input voltage which you see in the output side quite interestingly I get $C_{gd}$ * ( 1 + $g_m$*$R_L$'), right. I am a deriving it in the class but $C_{gd}$ or C equivalent or input capacitance will be $C_{gd}$ * ( 1 + gm*$R_L$), which means that initially the device would be looking only at $C_{gd}$ but because of a coupling between gate and drain, right, you have a multiplication term 1 + g $_m$ * $R_L$', this phenomena is known as Miller effect, this is known as Miller effect, right.

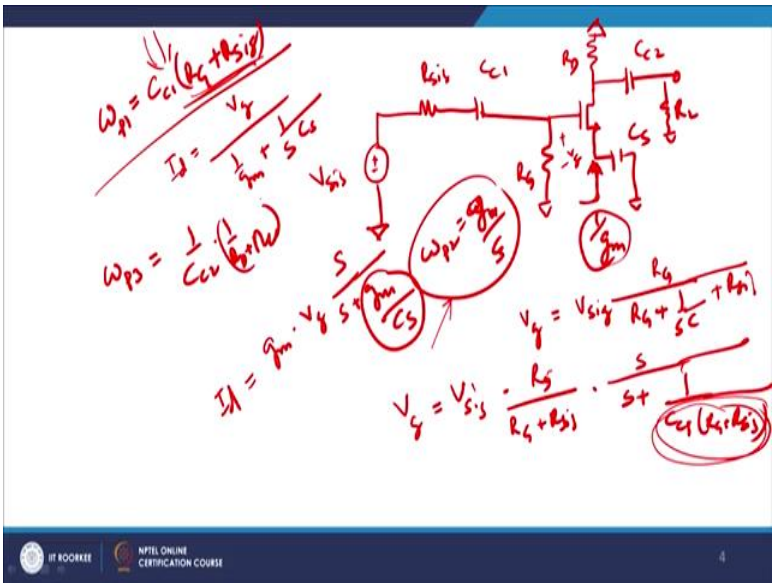
And therefore, I get the overall voltage will be $g_m$*$R_D$, so $g_m$* $R_D$ is nothing but gm times $R_L$*$V_{gs}$, why? Because $g_m$ is del ($I_D$ ) / del ($V_{gs}$ ), so this $V_{gs}$ cancels with this $V_{gs}$, $I_D$ * R $_L$ will give you the output voltage V0, right and from here you can get the value output voltage in a much easier manner, right. If we are now therefore come to the low frequency domain, low frequency domain of this device maybe we can explain to you, so, right.

(Refer Slide Time: 18:19)




So, let me come to the, sorry, so if you see very closely or if you have a look at the and we also do a low signal, low slow frequency modelling.

(Refer Slide Time: 18:34)



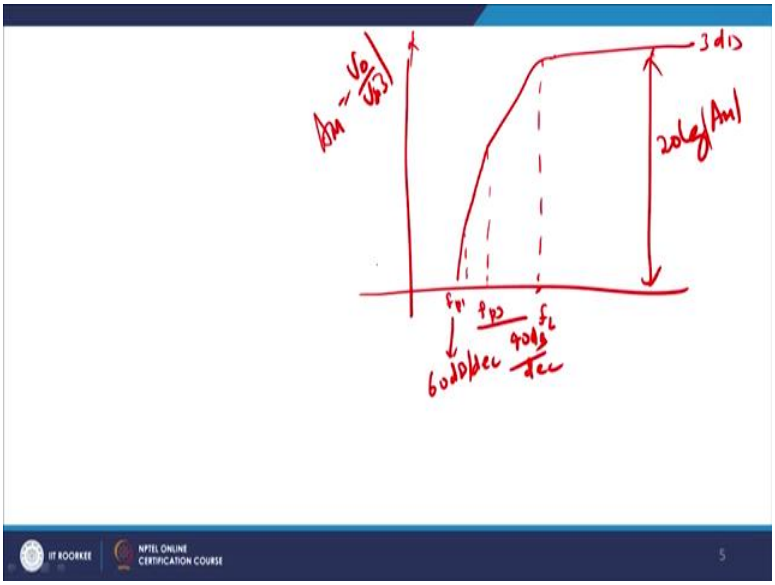

Then in the low frequency, as I discussed with you, right. All the capacitances will remain because they will not be closed as such and therefore I can write down something like this as the overall gate to source variation, right. So, this is $R_D$, $R_G$ $C_{C1}$, $R_{sig}$ and this is $V_{sig}$, this is grounded, and then this is also grounded because DC biases are missing and you have $C_{C2}$, right and then you have got $R_D$ or $R_L$ as part of it the $C_s$, right and this is your plus minus $V_g$ which you get.

So, looking from the source side the impedance offered is 1 / gm, right so keep this is in mind a standard methodology. So, I can write down Vg to be equals to $V_{sig}$ * $R_G$ / ( $R_G$ + 1/sc + $R_{sig}$), right and therefore, you can safely we write down Vg to be equals to $V_{sig}$ * ( $R_G$ / $R_G$ + $R_{sig}$) * ( S / ( S + 1 / $C_{C1}$ ( $R_G$ + $R_{sig}$))) where this value is referred to as $w_{P1}$ or the first point cross point frequency, first cost point frequency is $C_{C1}$($R_G$ + $R_{sig}$). Which means, the first cross point frequency or the frequency beyond which you will see a drop in the gain is governed by the $C_{c1}$, the coupling capacitor attached here as well as the resistance offered by the RG and Rsig value. Now, therefore $I_d$, $I_d$ as I discussed with you is $V_g$* $g_m$ will be $V_g$ / (1 / $g_m$ + 1 / $SC_s$ ), fine why? Because looking from this side if you look very carefully 1 / $g_m$ is the source impedance as well as 1 / $SC_s$ is the impedance offered by this capacitance.

So, since they are in sort of a parallel connection therefore what you do is? That you add them up and therefore Vg upon that will give you the value of drain current, right. So, if you solve it and if you want to get the picture I get $g_m$ * $V_g$ (S / S + $g_m$ / $C_S$), this is your Id, right and therefore I get a second crossover point $w_{P2}$ as gm/s, so have a first crossover point given by this and a second across over point given by this, right.

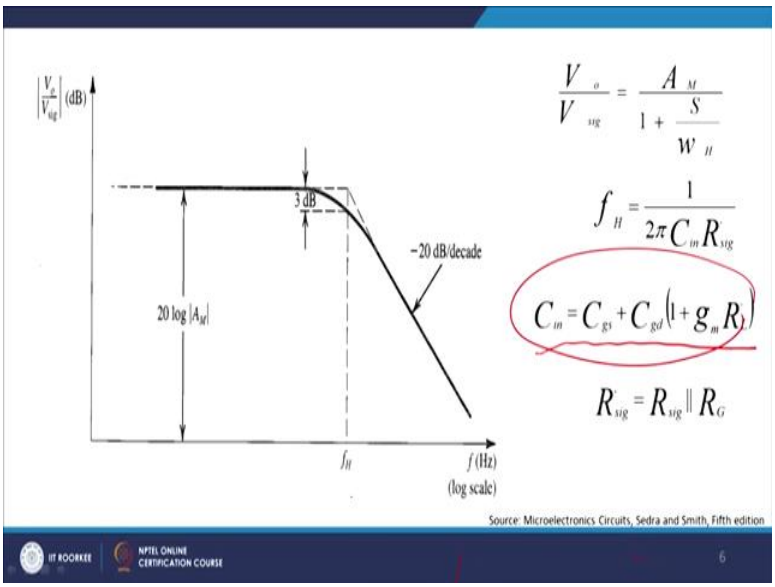
You also get a I am not deriving in the class but you also get a third crossover point which is given as $w_{P3}$ is ( 1 / $C_{C2}$ * 1 / $R_D$ + $R_L$) which means that each of the capacitor $C_{C1}$, $C_{C2}$ as well as $C_S$, right starts to give you extra dip in the in the output characteristics. So, as you lower your frequency domain is they start to fall drastically. So, if you plot it is frequency.

(Refer Slide Time: 21:56)




If you plot its frequency versus gain graph then you get something like over 3 dB bandwidth obviously available with you, this shows our drop here and then this shows a drop here and then this shows a drop here. So, if you look very carefully this, is your $f_H$, right this is your $f_L$, $f_H$ and this is your $f_{P3}$ and this is your $f_{P1}$, this gives you approximately 60 dB per decade drop, this gives you 40 dB per decade drop, right and then $f_L$ is there which is the low frequency gain which you see, right. This is nothing but 20 log of $A_m$, right where $A_m$ is the gain which you see, $A_m$ is therefore equals to $V_0$ by $V_{sig}$ for a fixed value of this value, right.

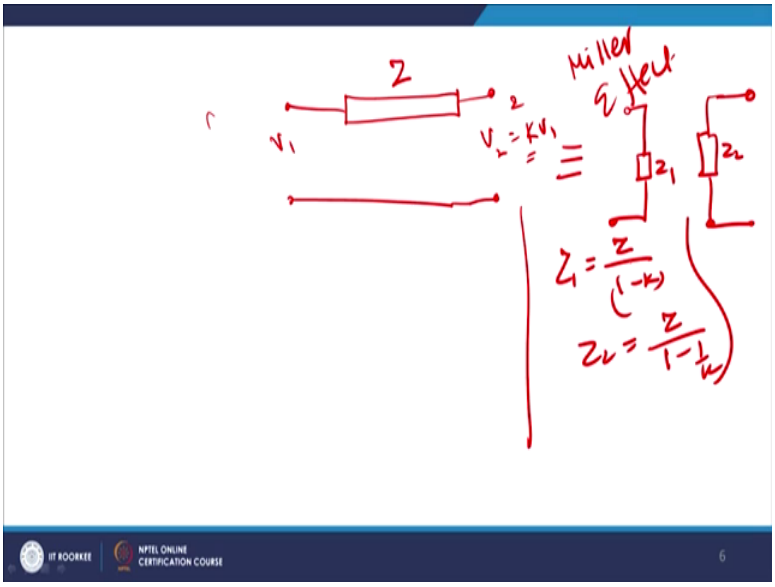
(Refer Slide Time: 23:13)




So, this is what you get from the gain and therefore if you look carefully then we get what you get is basically $C_{in} = C_{gs} + C_{gd} *( 1 + g_m * R_L)$, right. So, let me do the, this is this for the

low frequency design and therefore let me do for a high frequency design to make that to make it much more better understood in the sense that we can make it high frequency design.

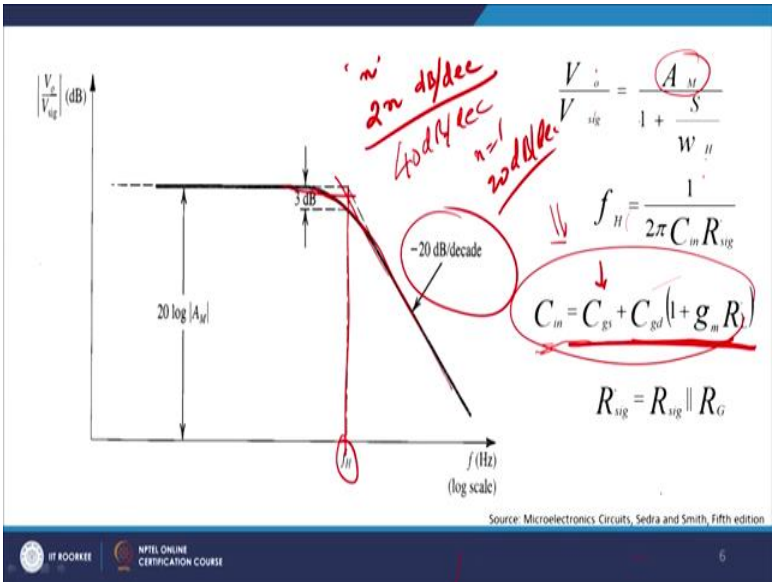
(Refer Slide Time: 23:25)




Within the high frequency design, we also try to find out or try to see, what is known as a Miller multiplication, so I have got two ports here 1 and 2, 1 and 2, let us suppose input voltage is $V_2$, $V_1$ and output voltage is K * $V_1$ which means that the input and output are variated through a costing multiplication factor K.

Then this and this is Z is the impedance then this is equivalent to saying as this Z 1, right and then this is Z 2, right and Z 1 is referred to as Z / 1 - K and Z 2 is referred to as Z /( 1 - 1 / K), right I am not deriving it in the class, in this lecture but this is how it looks like and it and this is known as Miller effect. So, Miller effect tries to enhance the overall gain by making coupling between gate and drain in a much prettier fashion or easy fashion right and that is what we get from the Miller capacitance as such.

(Refer Slide Time: 24:44)

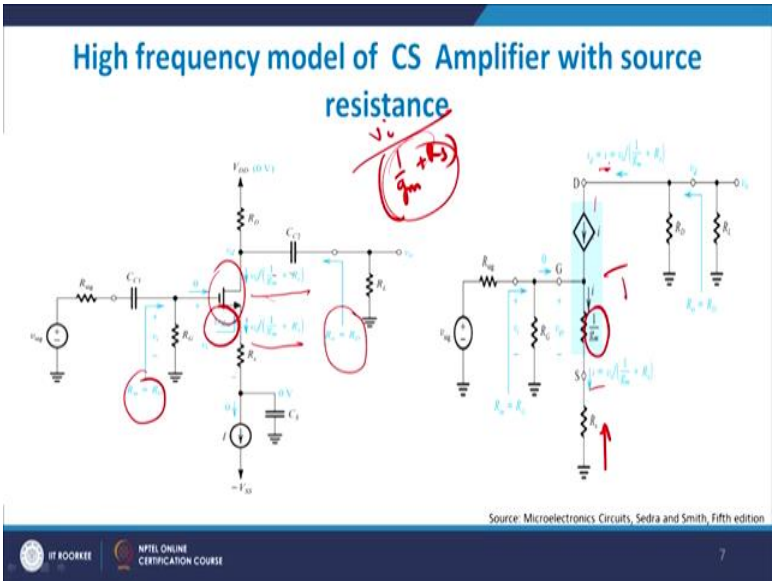



So, with this what we have done therefore is, that we have actually seen therefore that as I discussed with you for every one pole added to the right you get, so for an nth order filter for example this is known as the decade or the role off rate is always given as 2 to the 2n dB per decade. Which means that, if a second order system is there right then it will be 40 dB per decade because n is equal to 2, if n equals to 1 then a single order will be have 20 dB per decade, fine and you can see therefore that this is the this is how it forms.

So, you backtrack this one, the back side in your backtrack this one, the cut is basically your half your f H which is basically your high band design. As I discussed with you earlier therefore V0 / Vsig will be given as $A_m$ / 1 + S / $w_H$ and $w_H$ is given by this formula $f_H$ equals to 1 / 2*pi*$C_{in}$, $C_{in}$ * $R_{sig}$, $R_{sig}$ is the approach value of voltage but $C_{in}$ is written $C_{gd}$ + $C_{gd}$ * 1 + $g_m$ * $R_L$.

So, as you can see that adding an extra potential or a capacitance in the feedback loop in sort of it an inner loop and assuming that C gs will always come into picture irrespective of any other thing we see that my input capacitance can be raised from a value $C_{gs}$ to $C_{gd}$, $C_{gs}$ plus this whole quantity. And that is quite an interesting and impressive one that by doing so, you will be able to raise the overall gain of the system, right.

(Refer Slide Time: 26:16)

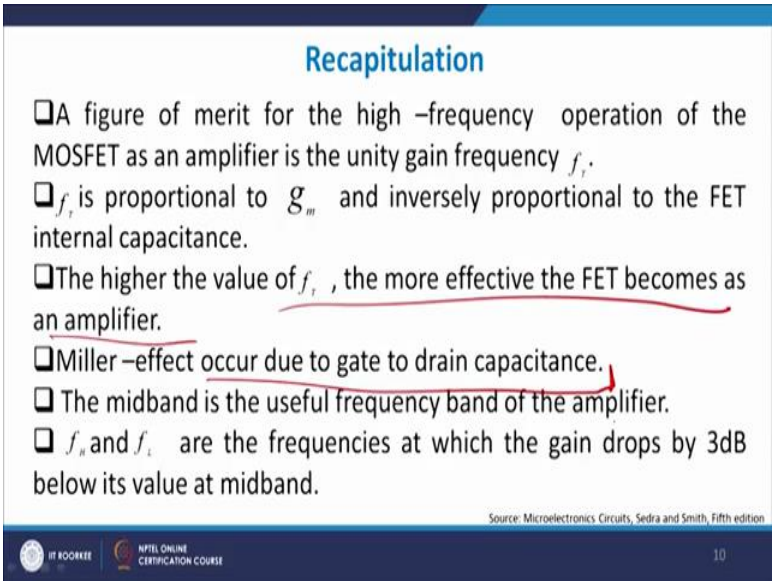



Let me come to the high frequency model of the CS, so that was the low frequency model and let me come to the high frequency model as I discussed with you. If you look very carefully, this is again the same thing $R_n = R_g$ and from looking from this side it is only $g_m$, right it is only $g_m + R_S$. So, the resistance offered at the source end is $1 / (g_m + R_S)$ right and then if you divide $V_i$ input voltage by this quantity is basically the current which is flowing through this resistor R right and that is what is the current is.

Now, since the current has to be the same, I see that $V_I$ * this quantity and this quantity will also be remaining the same. So, on the right hand side $R_0$ will be equals to $R_d$ which is basically the input impedance which you see from here. Now, if you replace this MOS device by it is corresponding T model then we say that I will be equals to $V_I / ( 1 / g_m + R_S)$, ( $1 / g_m$ + total current flowing ).

Since of the gate side, from the gate side you do not have any exiting an entry of current almost the same current will flow through the resistor and the evidence is that there will be 1/ gm drop there and therefore looking from the source end it is $V_I / 1 +( g_m + R_S)$, right. So, looking for the source side current and both the currents are equal and therefore, there is no problem and as far as I understand in the configuration is concerned. So, we are finished with high frequency model and we have seen how frequency model works. Let me therefore recapitulate the whole discussion on that on this topic.

(Refer Slide Time: 27:58)




Well with the figure of merit for high frequency operation of MOSFET is an amplifier is the unity gain frequency, how did you define unit gain frequency? Where a gain falls to 1, we define that to be as the unity gain frequency. $f_T$ or the transit frequency is or is the proportional to $g_m$ and inversely proportional to FET internal capacitances, which means that higher the value of g m more will be the bandwidth and more will be the your $f_T$ but if it in general capacitance is also higher than your reduction will be there in your f $_T$ as well as $g_m$.

The higher value of $f_T$, the more effective the FET becomes as an amplifier, right. So, one thing which we learn is that your $g_m$ should be high and your $f_T$ should also be high, once you ensure these two to be very high then automatically everything falls into place in a detailed manner. Miller effect occurs due to gate to drain capacitances right and this is quite an interesting problem which people face that the Miller effect which will do may be in the next turn and you show it to how you Miller effect works out.

But Miller effect does what primarily is that it adds to the overall capacitance of a system and that addition is also not linear it is effective addition. So, some there is some constant, you multiply that with $C_{gd}$ to get the new value of $C_{gd}$, so gate to drain coupling capacitors is there. Now, Miller effect therefore works due to gate to drain capacitances that is what we have just now understood.

The most useful sub band of amplifier is basically your mid gap band, why? Because the gain is almost constant and independent of the frequency and therefore, I can safely say if you are operating your amplifier within that range my output will be almost constant depending only

in the value of $V_0$, right. Then you have two important frequencies, we have seen $f_H$ which is also referred to as low cut-off frequency, high cut-off frequency and $f_L$ which is basically refer to as a low cut-off frequency.

They are typically 3 dB points below the maximum gain position and beyond these points the gain starts to fall down with increasing or decreasing frequencies, right. So, the most reliable one is basically your the mid band frequency. Last thing is, that there is a figure of merit which is basically known as gain into bandwidth, which is also known as gain bandwidth product that should be always constant for a system.

So, if you are increasing the gain pay the price of a lower bandwidth if you are increasing the bandwidth you pay the price of a lower gain, right. So, we with this we have almost learnt most part of the amplifier design and look into the very critical aspect of amplifier design, next time we will take up another issue as far as this course is concerned, all right thank you very much for your patience hearing.

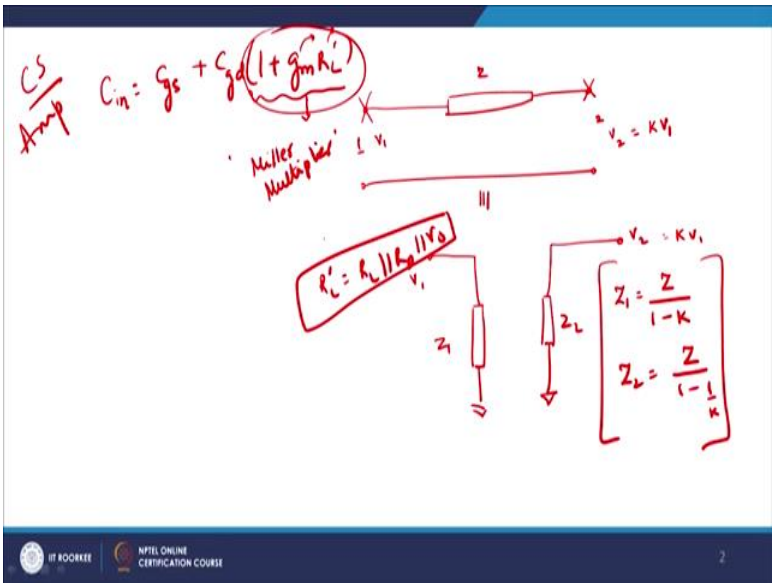
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−32**
**JFET, Structure and Operation**

Hello everybody and welcome again to the NPTEL online certification course on Microelectronics: Devices to Circuits. In our previous lecture schedule, we had seen how a common source amplifier works and what do you mean by mid band frequency? What do you mean by lower cutoff frequency? And what do you mean by higher cutoff frequency or upper cutoff frequency? How do you define bandwidth? What is the meaning of gain bandwidth product? And this is one of the figures of merits which we have discussed.

We also saw, why does the gain false of at relatively high frequencies and low frequencies and remain stable, independent of the applied frequency somewhere in the mid band. Then we saw the mathematics of it and we saw that, if we do have common source design at low frequencies at the device the capacitance are open circuited and high frequency is there shorted.

We are left with two important parts which we did not carry forwarded in the previous lecture series and that is what we have actually seen is the Miller capacitance is, we have already discussed with them together but I will just refurbish you memories here.

(Refer Slide Time: 01:35)



So, let us suppose I have I want to just refurbish the whole thing. So, I will let us have a two port network, this is port one and this is port two and I have V1 is the voltage source at port 1 and I say at port 2 the voltage V2 is equals to k * V1, which means that the voltage at V2 is a function of V1, right. So, there is sort of feedback loop between nodes 2 and 1, right.

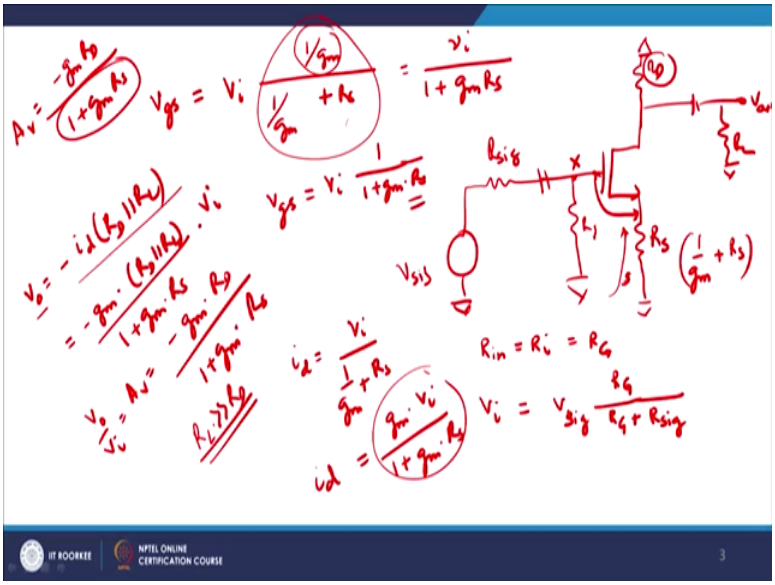
Now, this can be replicated as that is sort of string is, it can be replicated as like this, right where this is basically is z1, this is v1 once again and this is z2 and this is again replicated where this is v2 and where we supposed to be is k * V1 where we can write down Z1 = z / 1-k and Z2 = Z1 /( 1 – 1/k ).

Now, formal proof at this stage is beyond the scope of this course but this I how it works out which means that, if I ground those two capacitances I get it. Now, in my previous discussion we saw that when we doing high frequency response we that saw that C in will be equal to $C_{gs}$ for a common source amplifier, for a common source amplifier $C_{gs} + C_{gd}(1+ g_m * R_L')$ right this is high frequency response which you got, where $R_L'$ is nothing but the resistance $R_L \| R_D \| r_o$, right.

So, this is basically your $R_L'$, right and this quantity is referred to as a Miller multiplier, Miller multiplier which means that? The gate to drain capacitors which was initially a low value in reality gets multiplied with $1 + g_m * R_L'$ and gives effectively gives you a larger value which means the reason been gm is typically large, right $R_L$ is also large, so this quantity is a relatively large quantity.

So, that means even if the gate to drain capacitance is small it is effect on the output is relatively large, right that is what I wanted to give you an idea about place an idea about. And the reason being that which means that therefore when two ports are connected by an impedance just as gate and drain current at through a gate impedance then we can refurbish or re-clarify it and we get this structure into consideration, right and this is what we get as the output capacitances.

(Refer Slide Time: 04:38)




Now, let me come to, we have also seen that, sorry we have also seen that there was a source degeneration resistance concept which we have seen, right. And we saw that if that is true which means that if I have got a common gate that means I have got a this, this and now capacitance here and I am driving by $V_{sig}$, right and I also have a resistance here, right this is your $R_g$ and you have $R_D$ here, right and then this is a cap here and then you have $R_L$ here, right and this is your $V_{out}$ and this is your $V_{dd}$ , let me save ground and gate.

Now, under these criteria this is also referred to as source resistance Rs, right and what happens is that? The effective resistance seen from the source side is basically $1 / g_m + R_s$, fine. So, if you look from the source side which is a source end of this MOS device, the effective resistance seen by the devices $1 / g_m + R_s$, with this knowledge let me write down for you Rin to be equals to $R_i$ that is equals to almost equals to $R_G$, right because the gate resistance is the largest one and you get this.

Therefore I get Vi input voltage which mean input voltage is the voltage at this particular point at point x will be equal to $V_{sig}$ signal, right signal * $R_G + R_{sig}$, so this is Rsig, right. So, $V_i$ equals to $V_{sig} * R_{sig}$, so this is basically potential divider network which you see. Therefore I can safely write down $V_{gs}$, right equals to $V_i * ((1 / g_m) / (1 / g_m) + R_s)$, why? Very simple, if you look $V_{gs}$ is gate to source voltage which is applied here, right $V_{gs}$ is the gate to source voltage, gate to source voltage, this one.

Will be nothing but, so not all of $V_i$ will appear as gate to source voltage apart of it will appear across gate to source voltage depending on this value of $((1 / g_m) / (1 / g_m)$, why?

Because this is the resistance of the device itself and this is the resistance of the device shrink from the source side and therefore if we divide it I get this into consideration, this will be approximately equals to $V_i$ / ( 1 + $g_m$ * $R_s$).

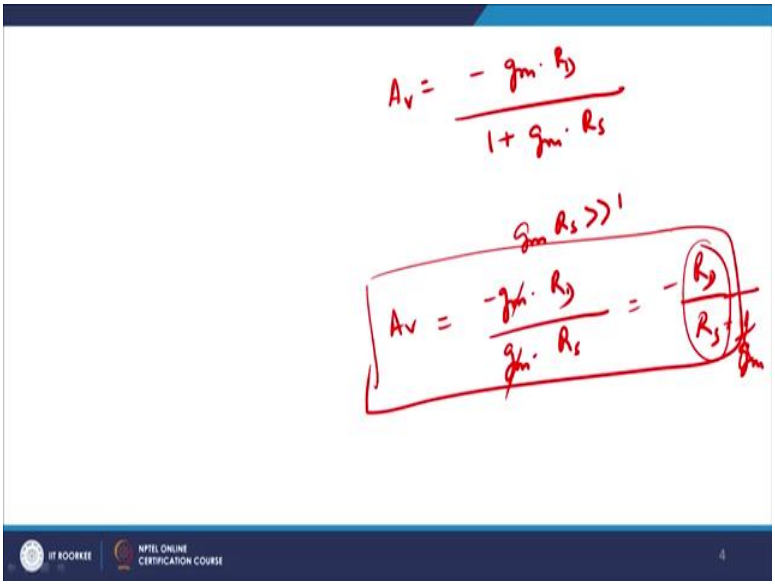
So, if you if therefore, if you find out $V_{GS}$ therefore, I am sure to be Vi *( 1 / 1 + $g_m$ * $R_s$ ), right and this is the source resistance which you see, right. Therefore, if I want to find out the drain current $i_d$, will be given as Vi / ( 1 / $g_m$ + $R_s$ ) and this will be approximately equals to ( $g_m$ * $V_i$ ) / ( 1 + $g_m$ * $R_s$ ) is equals to $i_d$, right. Now, we very well know that Vo output voltage will be - $i_d$ , right $R_D$ || $R_L$, where $R_D$ and $R_L$ what?

Suppose this is $R_L$, this is $R_D$ then we define V0 equals to -id phase is change is there therefore there is negative sign. So, I can safely write down this to be as - $g_m$ * $R_D$ || $R_L$, right divided by (1 + $g_m$*$R_s$) * $V_i$, this is what we are doing, we are placing it here and plugging it here and therefore $V_0$ / $V_i$ which is nothing but Av gain is equals to – ($g_m$ * $R_D$)/ 1 + $g_m$*$R_s$ where $R_D$ is the drain resist.

If you actually see the overall resistance assuming that your $R_L$ is much, much larger as compare to $R_D$ and this is true also, right. So, I get overall gain $A_v$ to be equals to - $g_m$*$R_D$/ (1 + $g_m$*$R_s$). So, if you look very carefully or if you want to have look into the careful consideration we will see that your presence of Rs actually lowers the gain, right and it makes the gain low because as you increase the value of Rs at the denominator increases and as a result it becomes slow and as a result what happens is ?

That your gain starts to fall down but the interesting part is that, as I discussed today earlier also that this presence of Rs actually helps you to do a negative feedback and stabilizes it. So, though your gain falls down your bandwidth increases because the gain bandwidth product is always fixed for a system, right. So, that is the advantage of having this resistance as available to you.

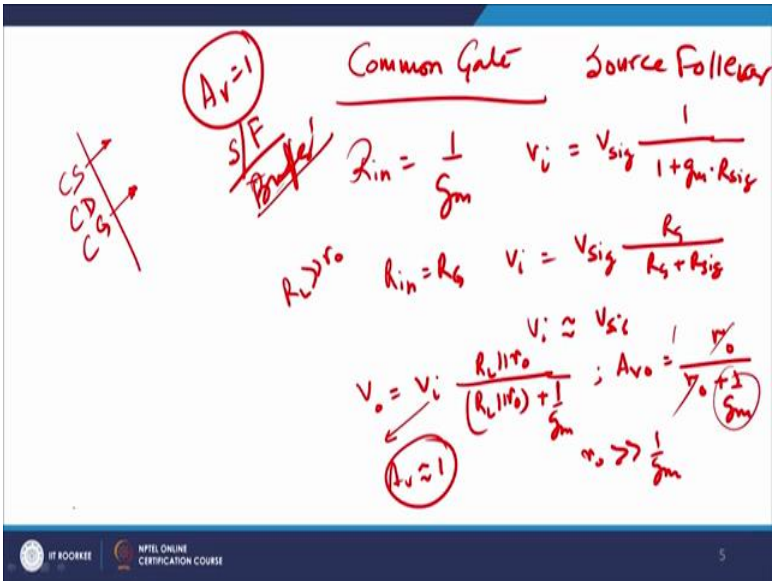
(Refer Slide Time: 09:25)




Now, if you look at this equation here, so let me just write down in a clearer fashion that Av equals to - $g_m$ * $R_D$ / ( 1 + $g_m$*$R_s$ ). Now, obviously $g_m$ * $R_s$ is much larger as compare to 1 therefore I can safely write down Av to be equals to - $g_m$ * $R_D$ / $g_m$ * $R_s$, right. So, this gm cancels out and you are left with minus $R_D$ / Rs equals to gain, right which means that if you want to find out the gain just by inspection, look at the drain resistance, look at the source resistance.

But there is an issue here that you will see the source resistance extra term 1 / $g_m$ will also come into picture, right because when you sweep in the source side that is the extra gm which you which as come already if you remember correctly it is already come. So, you need not put this one but typically $R_D$ by $R_s$ will be the value of $A_v$ which you need get in such a scenario, right. So, this is what we get from a previous discussion and our knowledge as far as this design is concerned. We have also understood the basic concept, let me come to the next one, which is basically a common gate design, right.

(Refer Slide Time: 10:41)




Common gate, gate is grounded and therefore, I so I can do common gate design not very seldom used to it, it is not basically seldom used it is also known as a source follower design which means that it is follower which means that it is gain is almost equals to 1. So, Rin equals to $1 / g_m$ as I discussed with you, Vi equals to $V_{sig}$ $1 + g_m * R_{sig}$, right and therefore I get $R_{in}$ equals to $R_G$ approximately and therefore I get $V_i$ equals to $V_{sig}$ into $R_G / R_G + R_{sig}$.

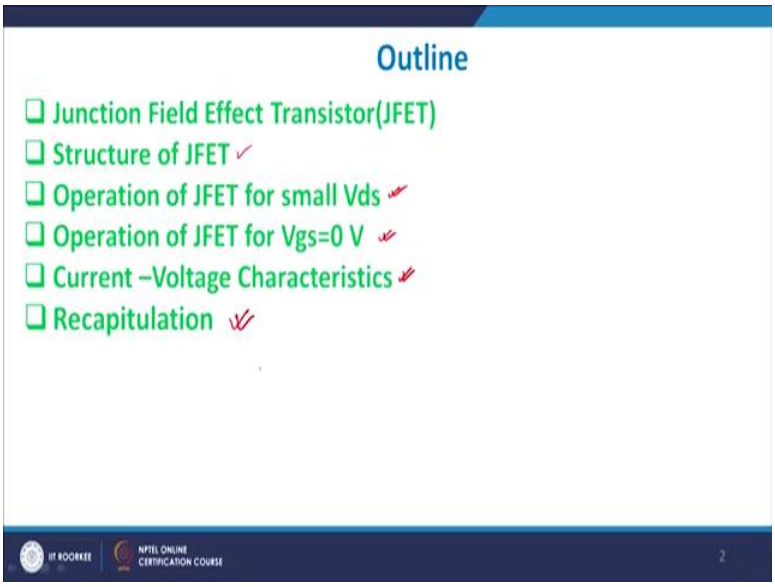
So, therefore Vi will be approximately equals to $V_{sig}$ because $R_G$ or $R_{sig}$ because $R_{sig}$ is very small as compared to RG. So, I get $V_0$ to be equals to $V_i$ into $R_L$ parallel to $r_0$, right into $R_L \parallel r0$, right + $1 / g_m$, fine and therefore I will get open loop gain $A_{v0}$, so this $V_i$ will come into denominator since $R_L$ is much, much larger as compare to $r_0$, I can really write down this to be as $r_0 / r_0 + 1 / j$, right.

Since, in reality $r_0$ is much, much larger as compare to $1 / g_m$, we can write down $A_v$ to be approximately equals to 1 because this will be very small, this cancels out and I get 1, right. So, in a common gate configuration, the gain is always equals to 1 and therefore it is also referred to as a source follower and generally there used as buffers, right they used as buffer. So, with this discussion we are finished with the three configurations available to us and these key configurations are common source, right.

We have got, you have got common drain and you also have the third type which is common gate. So, there are three configurations, the most seldom used as this one because of it is high gain, this is source follower which you used for impedance matching purposes, right. With this we have finished the basic concept of single stage amplifier, how to design a single stage

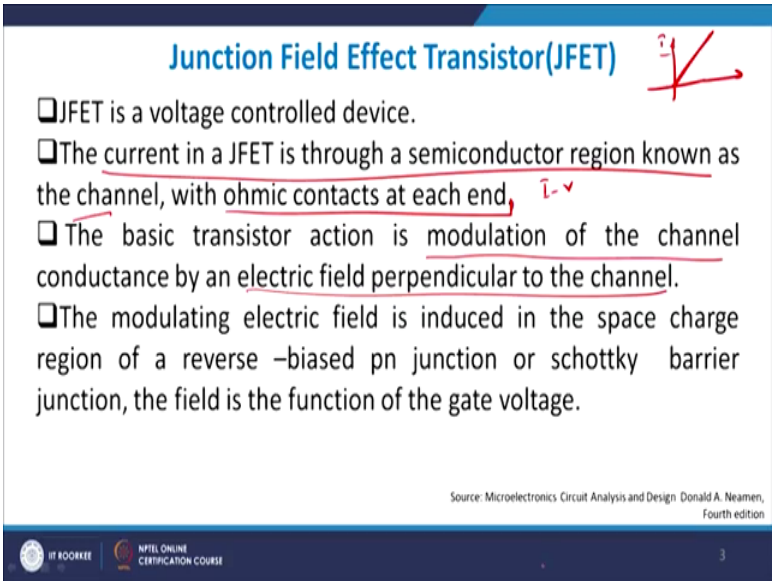
amplifier from MOS devices. We now come to a new device, which is the topic of this module and that is known as JFET or junction field effect transistor. We will look into it is structure and it is operation in this module and in the subsequent modules we will discuss other parts of the structure, ok.

(Refer Slide Time: 13:32)




Let me come to the, what will be the outline, the outline will be look into JFET or junction field effect transistor. We will looking it is structure, how it is looks like then how does it operate when you are apply a small drain to source voltage and when you apply $V_{GS}$ is equals to 0, how does it operate? We will look at the I V characteristics and then recapitulate our idea is about JFET, right.

(Refer Slide Time: 13:38)

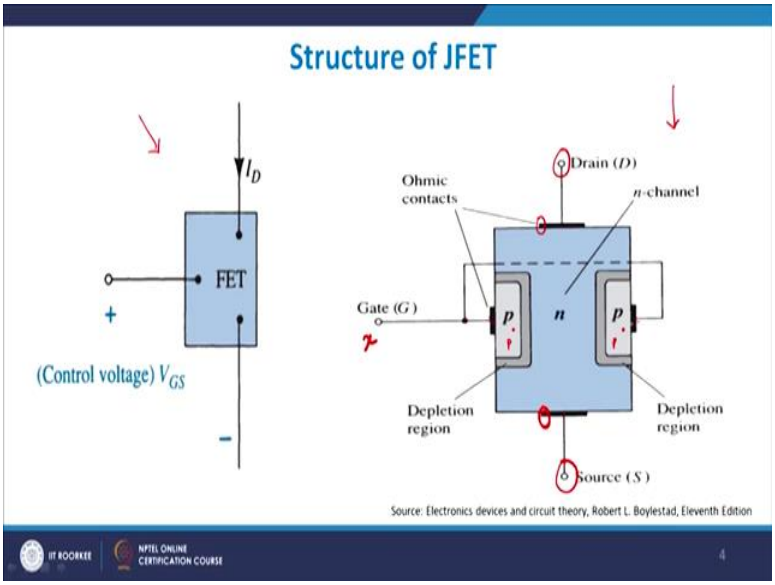



So, it is what we will doing here? So, again as have, as a MOS device, this also voltage control device which means that this basically again a voltage control current source. So, you will be applying a voltage and the current will be changing on the output side. So, the current in the JFET is therefore through a (resi) semi-conductor region known as channel and the with the ohmic contacts at each end, ohmic contacts means? Which follow I V current versus Voltage linear graph.

So, this is basically, this is the ohms law, right because so I versus V, so any contact in which I versus V is a linear graph is basically my ohmic contact, what we will be looking into his how it operates and that operation will be explain to you in subsequent slides. But, the basic transistor action is the modulation of the channel, right by electric field perpendicular to the channel, so which means that it is the something like this.

So, if I have a channel, in suppose a water is flowing through the channel, right then what I do? As I close the channel from outside by pressing it from both the sides, so channel is getting narrower and narrower, then what will happen? The volume of water will reduce, right and as a result, so I increase the volume of water by opening of the channel or decrease it by closing the channel. It is conceptually the same thing exactly happens in a junction field effect transistor and let the see how it is works out, how what it is structure looks like.

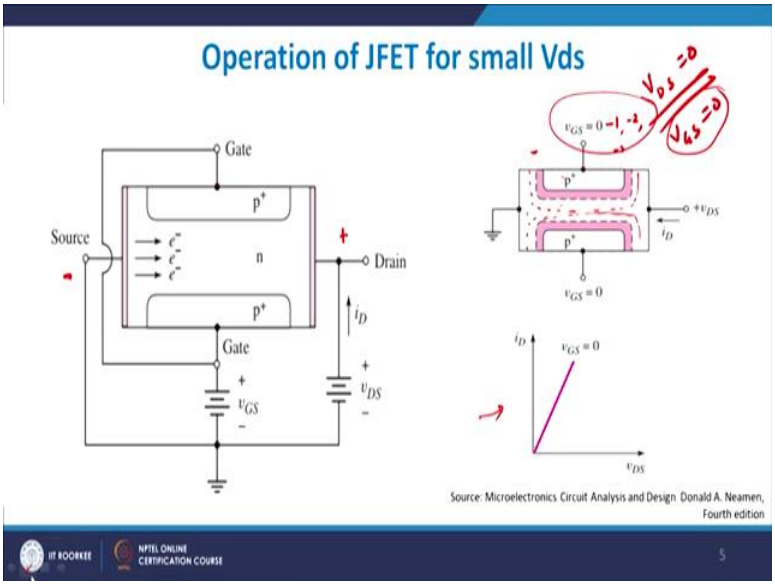
(Refer Slide Time: 15:05)

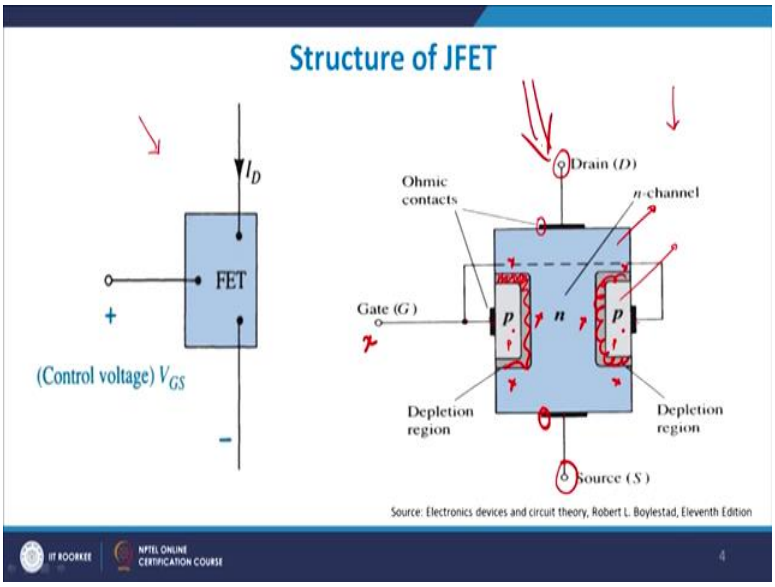



So, if you look at this on the left hand side is the schematic of JFET, on the right hand side would you see here is the actual diagram cross section of a JFET. So, if you look closely I have this ohmic contact here, I have drain region, I have a source region exactly like a MOS device, this is ohmic contact, so it is a basically a metal contact, metal semiconductor contact here, I have a gate region also in ohmic contacts are 1 ohmic, 2 ohmic, 3 ohmic and 4 ohmic contacts then I have a P region, this is basically P type, this is P type and this whole substrate is basically your n type, right this whole substrate is basically n type.

Now, if you look very closely from your previous discussion and these two gates are this gate and this gate are shorted with respect to each other. Which means that, if I give a voltage x here the same voltage appears here as well as here, right so it is known as the (shorted gain) shorted gate or the two gates are shorted with respect to each other. So, the same voltage is applied to both the gates.

(Refer Slide Time: 16:13)




On the such a criteria, we will see (how those) how it works out or how it works in a general sense. So, before we move, so this is a structure which you see, as I discussed with you, the doping concentration here is pretty large, this is also doping concentration is pretty large and drain and source you have metal contact here. Now, as I discussed with you even a without application of any external source or external bias there will be obviously a depletion region is will be here, right.

There will be always a obviously depletion region which will be here, right and this depletion region thickness will depend upon the relative doping concentrations of the P side and n side but the interesting point will be the depletion region will be equally thick here, here as well as here, fine I hope you agree with the statement that the depletion since I am not giving any

bias anywhere all of them look equally reverse biased or forward biased and therefore, the depletion region will be formed equally in all the directions, right because the doping concentration is same across the all the directions.

Now, let us see how it works out. Let me see that you did not you first of all, now let us suppose we explain in this manner that I apply on the gate side, this is the gate voltage I am applying, a positive bias means I have forward biasing the gate with respect to the source and applying a positive bias here and a negative bias here, right, let us see how it works out. Let us suppose I this is like a let me see, will be how we will understand.

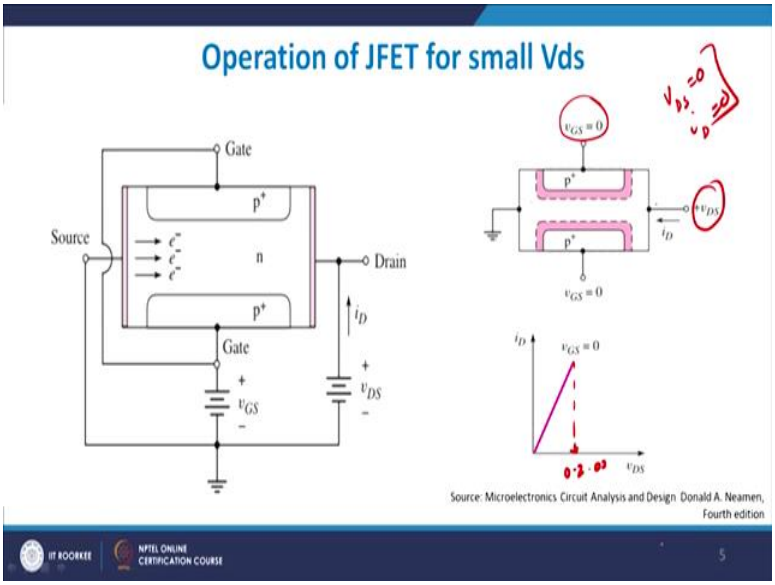
Let me see let me put VG, let me have $V_{DS}$ equals to 0, right also my VGS equals to 0 which is exactly the this graph which you see or this profile which you see. Now, what I do is I keep my $V_{DS}$ equals to 0, right and I go on increasing the value of VGS, right and how do I increase it? I make it more and more negative, so -1, -2, -3 and so on and do forth.

So, once I put to the positive sign and negative voltage obviously this will reverse bias with respect to n channel, right and as a result what will happen is? This will grow and this will grow what? Equal in all directions, can we got the point, right. So, will VGS equals to 0 with channel is fully open, you go and increasing the value of VGS keep your $V_{DS}$ constant at this stage, go on increasing the value of VGS in a negative fashion for n channel.

If you have P+ as your source and drain and n as channel will go on, giving more and more negative value to the gate and you will see that the depletion region in the both edges will start drowning together and a time will come when they will approximately touch with each other though that is a debatable issue but they will touch each other and therefore the channel will be lost and therefore even now if you apply a drain voltage nothing will happen because you are the because those areas divide of any free charge carriers, right.

So, this is what I want to do say that by simply application of negative bias on the gates side I was able to control the width of my channel and therefore the net flow of charge carriers from source to drain. Now, with this idea or with this basic concept let me therefore explain to you this graph here which is given at this point. Now, let me do one thing since you understood how a channel will be closed, let me do one thing.

(Refer Slide Time: 19:49)

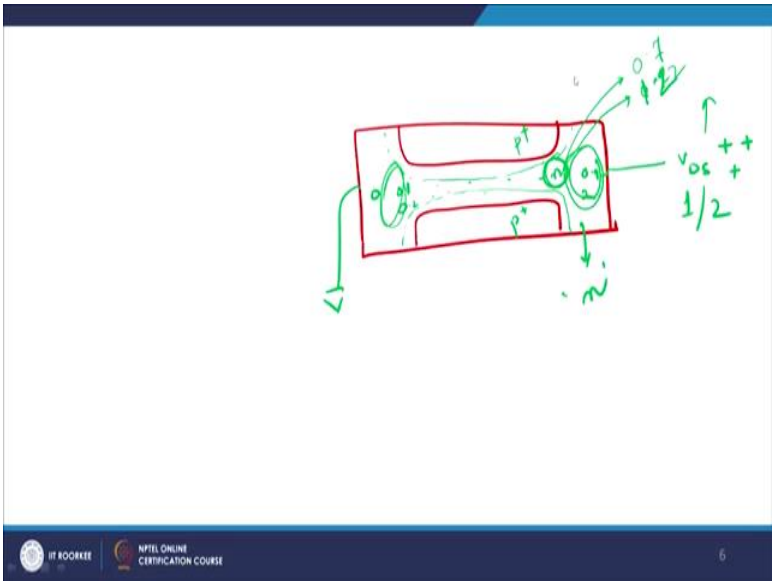



Now, what I do is? I make the VGS equals to 0 and then go on increasing the value of $V_{DS}$ which I am doing it here. So, you will be obviously certain that when $V_{DS}$ equals to 0 when $V_{DS}$ equals to 0, right your $i_D$ will also be equal to 0 because it is no potential, there is no chance of any electric field and therefore when there is no electric field no electrons will be moving from source to drain and as a result, source is grounded and as a result there will be no current but as you go and increasing the value of $V_{DS}$ in a very, very small domain this $V_{DS}$ is relatively very small say 0.2 volt or 0.3 volts, right.

The electrons will be moving and it will move fast from this end to this end and a current will be fall but and quite interestingly and therefore we just compare it with flow of water, I have water is flowing but tap is there I open the tap fully or I open tap and then let I do I start closing it but when I am closing it, I am closing it very slowly at the starting.

So you will not feel the difference in the water flow as such and therefore it will be almost linear increase in the current because you go on increasing the value of $V_{DS}$ more and more electron will be flowing and therefore you will get a larger amount of current. But, understand one important point that and that is quite interesting. Let when you make it $V_{DS}$ larger or larger, right you see say I will give it example to give you an idea about.

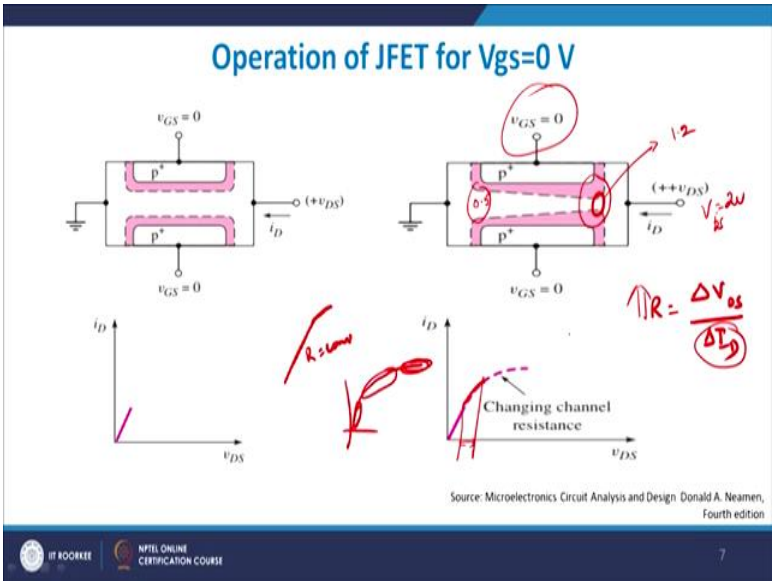
(Refer Slide Time: 21:22)




So, let us suppose this is your JFET, right and this is your P+ region, this is your P+ region and then what you do is? You apply a bias and you had a this and this, so assuming it to be equal. Now, you apply a $V_{DS}$ here and we go on and making it more and more positive as you get, so if you apply 1 volt here it means that it is 0, so it is 0.1, 0.2, 0.3, 0.9 and then 1, if it is 2, it is 0.2 here, 3, 4, 5 and then it will come 2 volts here which means that, as you go on and increase the value of $V_{DS}$, this side will become more and more positive as compare to this side means the all the channel will become more positive but this side will be more positive.

So, here it will be 2 volt, here will be, so if you look at very concentrated this point when you apply $V_{DS}$ equals to 1 volt this will be my 0.7, when you apply 2 volts this exactly the same point will be 0.9 maybe or 1.2 maybe, right but understand this is n channel MOSFET, this is n channel JFET. So, within n channel and you would apply a positive bias of the n side your effectively reverse biasing the P plus n region and therefore the depletion region on this side will be larger, right.

(Refer Slide Time: 22:46)

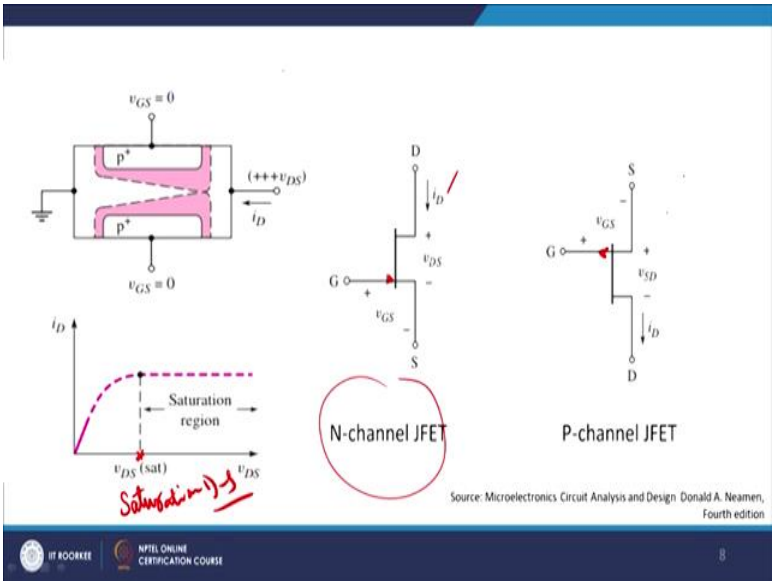



I will just show you what I was trying to say, you see this side depletion region is thicker as compare to this side because what this been see is just it will see 0.5 volts whereas this been see maybe 1.2 volts if you apply $V_{DS}$ equals to 2 volts. So, this side it becomes more and more (thick) more and more thick and therefore your channel starts to shallower and shallower, right it goes and becoming thinner and thinner.

As a result, now what happen is that the channel resistance starts to rise obviously if you are closing something the resistance offered by the channel will become more and more therefore the current which was following linearly straight line starts to show deep here, why? Because if you remember resistance is equals to $\partial (V_{DS}) / \partial (I_D)$. So, when I was in the linear region resistance is constant somewhere here for the same change in the (value), so somewhere here for the same change in the $V_{DS}$ is a smaller change in $I_D$ as therefore this is small and therefore R is large.

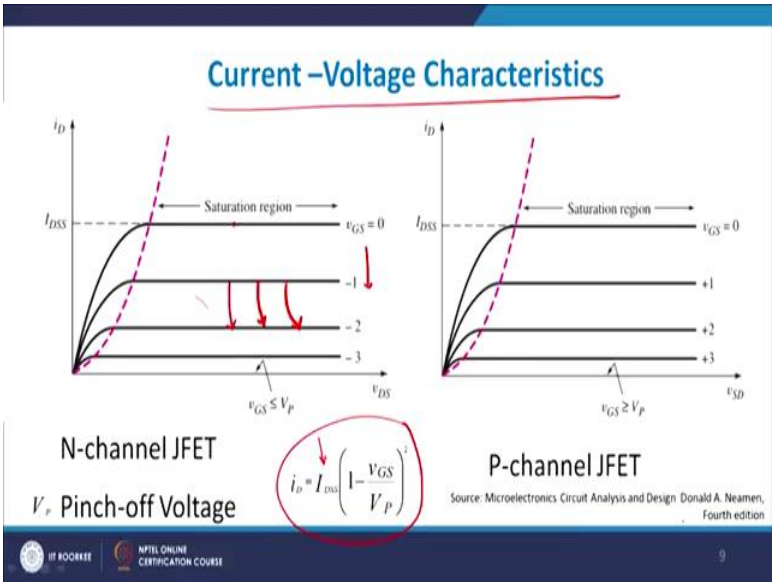
So, more the slope is less will be the resistance offered. So, the here is heavy resistance, here is almost very small resistance and here is infinitely large resistance available to you, fine. So, when you take $V_{GS}$ equals to 0, this is can take admission on $V_{GS}$ equals to 0, right and I get this current into profiling. I hope I made myself clear that this is basically with $V_{GS}$ equals to 0, I have increase the value of $V_{DS}$ and I get this into consideration.

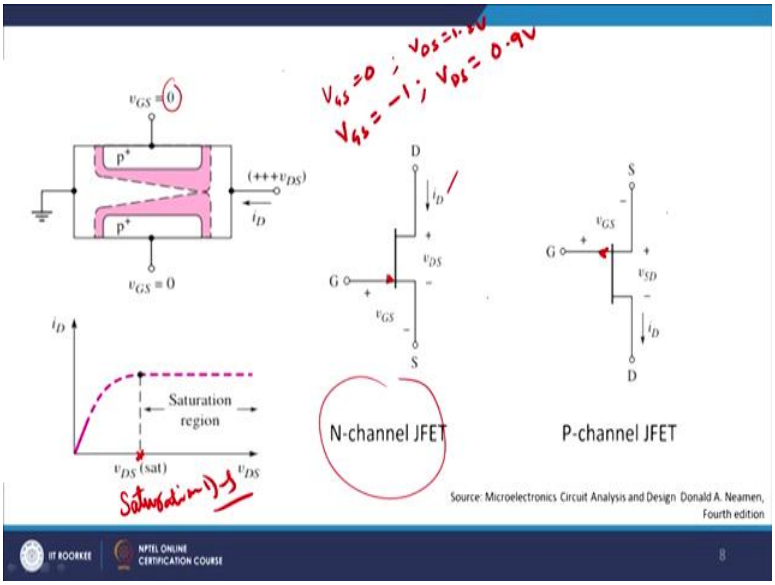
(Refer Slide Time: 24:24)




Now, what I do is? Therefore I define a term known as $V_{DS}$ sat or saturation drain to source voltage, saturated saturation drain to source voltage at which? The current gives independent of $V_{DS}$ it is almost too straight line and the resistance offered is infinitely large. This the schematic of n channel JFET and if you look it is pointing gate is pointing inwards and similarly gate is pointing outwards in a P channel JFET and the direction of current is already shown to you, right and this is what JFET look like.

(Refer Slide Time: 24:58)





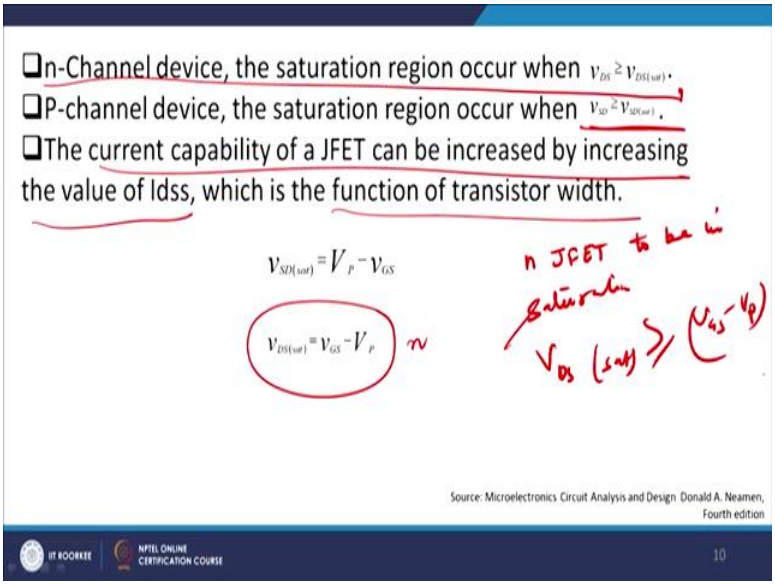

As I discussed with you therefore current versus voltage characteristics of a JFET if you want to look it looks iD versus VGS and it goes like this and as you can see, now you see as you make VGS more and more negative for the same amount of $V_{DS}$, the current starts to fall down, right and the reason being what? Reason being that now when your VGS is more negative, it is helping to formed a depletion layer more and more.

So, when VGS equals to 0, this was forming at say when VGS equals to 0 then it was the depletion region of the depletion region was touching each other as say $V_{DS}$ equals to 1.2 volts, clear. But, now what is happening is that let us suppose my VGS equals to minus 1 volt then my $V_{DS}$ at which the pinch-off will occur will be just 0.9 volts maybe. So, at a lower drain voltage I am able to achieve higher current or almost the same current, so that is what I wanted to tell you here, show you here.

In the saturation region as the gate to source voltage increases becomes more and more negative, the current starts to drop down or current starts to fall, right and we define a voltage pinch-off bad gate voltage at which the current goes to 0 is defined as my pinch-off voltage exactly the same definition which we saw for the MOSFET threshold voltage, right. And this is the equation for the drain current for case of a JFET, $I_{DSS}$ is the value of the current when VGS equals to 0.

So, this is your IDSS, right for VGS equals to 0 and 1 - $V_{GS}$ / $V_P$ where VP is the value which is you see, right. So, I have given you the discussion for n channel JFET, you please try to extend it to a P channel JFET as well, right and the answer remains the same as such occurring to you.

(Refer Slide Time: 26:49)

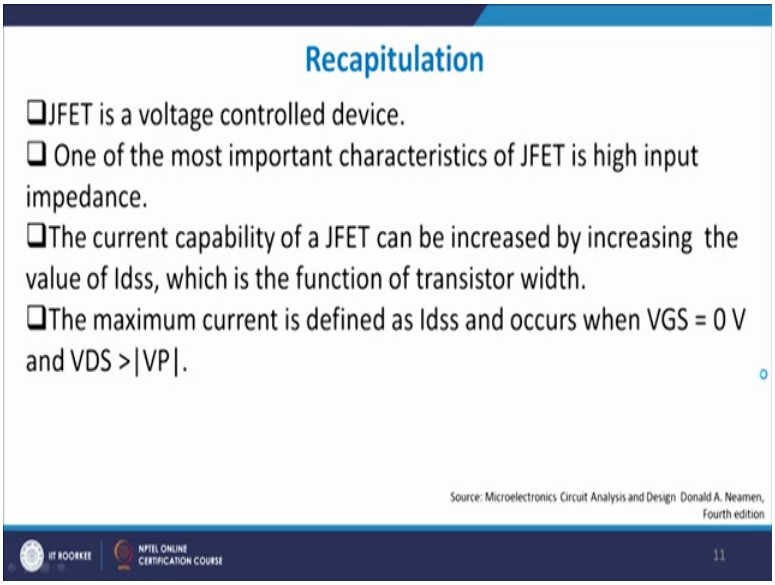



As I discussed with you in n channel JFET, therefore the saturation region occurs when $V_{DS}$ is greater $V_{DS}$ sat, exactly the same principle for a MOSFET, so I am going bit fast in this case because we are already deal with MOSFET in a detailed manner, conceptually, mathematically same as MOS device but cross section wise slightly difference in each operation value.

For, P channel this is what we sat that VSD is greater than VSD sat. Now, therefore the current capability of JFET can be increased by increasing the value of $I_{DSS}$ which is the function of transistor width. I hope you understand more is the width more number of electrons will be there and therefore even with VGS equals to 0 there will be substantially larger amount of current which will be there and as a result you will get a much larger profile in this case.

As i discussed with you for n channel therefore $V_{DS}$ sat should be equals to VGS - $V_P$ which means that for a device n JFET, right to be in saturation, to be in saturation $V_{DS}$ sat should be greater equal to $V_{GS}$ - $V_P$, please keep this in mind when $V_{GS}$ gate to source voltage and $V_P$ is the pinch-off voltage.

(Refer Slide Time: 28:00)




So, let me recapitulate to you, what we did in this small module of JFET because we will not be using too often because now JFETs are not too much used across the board. So, I should give you an idea about how it is working principle and how does it look like, right if time permits we will return back to this later on and explain to you some second applications but primarily it is voltage control device.

It is input impedance is relatively very high because on the gate side as I discussed with you, you have a depletion region working in the depletion region and therefore your forward biased characteristics are typically very low and reverse bias characteristics $i_D$ will be approximately equals to 0 and therefore impedance of a (dristrib) relatively very high. The current capability of JFET can be increased by increasing the value of W, right and typically the current flow is restricted by two important point VGS and $V_{DS}$.

So, for n channel JFET if you increase the value of VGS more negative I will get smaller current for the same amount of $V_{DS}$, right and so on and so forth. The maximum current is defined as I DSS on occurs when VGS equals to 0 and $V_{DS}$ greater than mode of V P which is primarily for n channel JFET, right. So, this gives you brief idea about my working principle for the JFET and basic structural parameter of JFET, right.

Two important parameters, I DSS, the maximum current flow through device in n channel JFET when VGS equals to 0 and VP the pinch-off voltage the value of (that drain to source) that gate to source voltage at which the current is exactly equals to 0, right and for the device

to be in saturation $V_{DS}$ sat should be greater than equals to VGS minus $V_P$, right. With this let me close the module here and thank you for your patient hearing, thank you.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−33**
**Multistage and Differential Amplifiers-I**

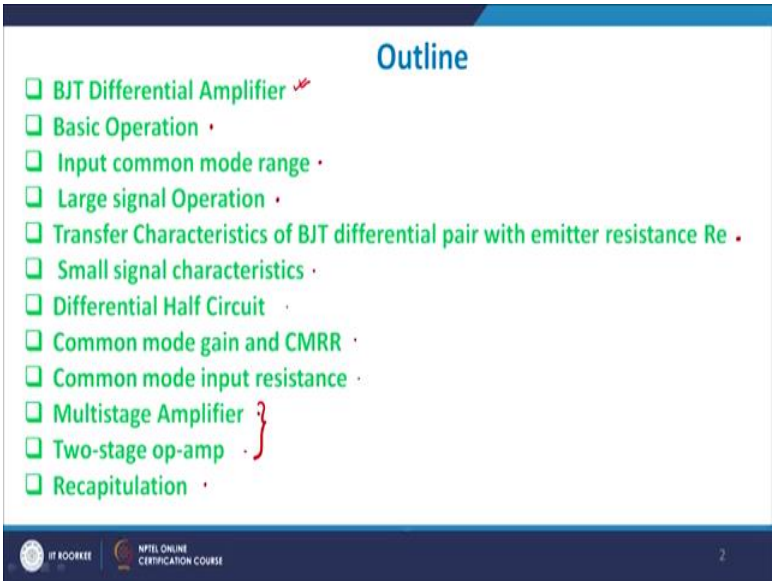
Hello everybody and welcome once again to the next module of NPTEL online certification course from Microelectronics: Devices to Circuits. Today's lecture will be on multi stage and differential amplifiers. We have studied in our previous few modules, about a single stage amplifier which primarily means that, if I have a simple common source amplifier bit a MOS or a BJT and I give a input signal which is basically a small signal ensuring that my devices are in the linear region of operation.

Then my output voltage will be undistorted without any nonlinear distortion available to it and therefore the gain will be almost independent of the input voltage. But a single stage amplifier, as you can always understand will give you a gain just quite small right, it is the order of few tens a few hundred or maybe something like that, when you want a very large gain out of some particular for a particular operation you either require a differential amplifier or even a multi stage amplifier.

So, what we try to do is? That in the first half of the lecture, we will be looking into the concept of differential amplifier and we will also look into the fact that using a single stage amplifier and by cascading it properly or adding stages to it properly how can I make my gain high. One more important property of differential amplifier which is not there in a single stage amplifier is.

It is capability of rejection of noise which means that unlike single stage amplifier which amplifies both signal as well as noise differential amplifier has got a capability of blocking the noise and enhancing the signal, right. So, signal to noise ratio, sort of if you want to take that as a parameter will be quite high in front of a differential amplifier, so with this basic knowledge or with this basic background.

(Refer Slide Time: 02:28)

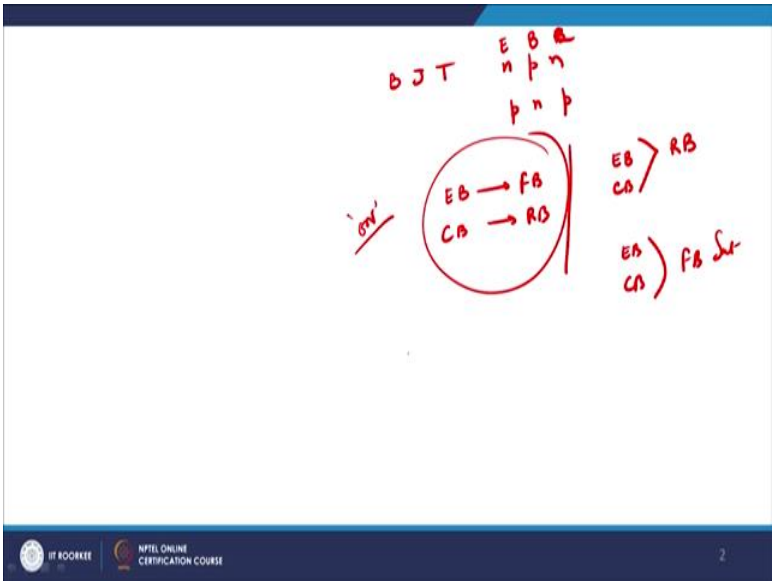



So, the outline of the present course module is, we will first of all look into a BJT differential amplifier in under which, we will look into its basic operation of the device, we will also understand what is the input common mode range with us then last signal operation of the device and then we will see that if we have an emitter resistance how does the output transfer characteristics change.

And then, we will look into small signal characteristics, we will also look into differential half circuits, CMRR we will which is basically common mode rejection ratio and then we will look at common mode input resistance and then we will finally have a look into multistage amplifiers and two stage op-amp as an example and then we will recapitulate, right.

So, these two the last two we will be dealing in a much more detailed manner in subsequent slides or subsequent lectures but today's topic will be more concentrated on differential amplifiers, right. So, let us first look at a differential amplifier and to go ahead let me give you an brief insight if you remember from your very basic device physics courses.

(Refer Slide Time: 03:37)

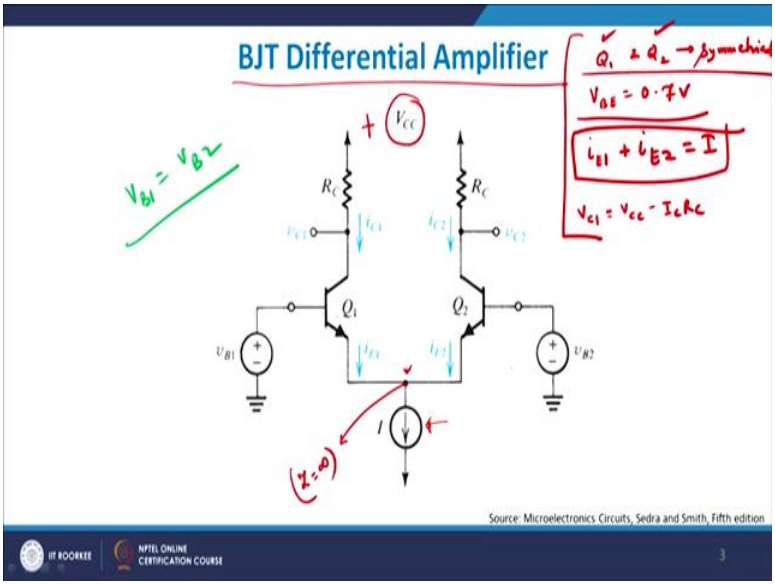



That if you have a BJT which is basically an npn or you will have pnp, right this is emitter base and Junction for the base and collector and therefore, if you want the device to be operating in the saturation region or in the triode region you require the emitter base Junction to be forward biased, right and collector base Junction should be reversed biased. If you wanted to be in cut-off then emitter base as well as collector base both should be reversed biased, right I hope you understood these points.

If you want to work it in the in the active region of operation then maybe emitter base and both collector base should be forward bias, this is your saturation region of operation, right. So, typically we want to keep this is where we define the been on state and we get better idea, right. So, please understand when you want to move the device BJT from a saturated stage to non-saturated stage, you spend large amount of time in switching because the charges in the base has to be removed or flushed out before the next cycle starts.

So, the time taken is very important, we will see in a differential amplifier that we do not like the devices to go into saturation region and that is the beauty of the whole thing that it gives you a much better gain and not only again it gives you a much faster switching speeds. We will discuss one by one as you move along but we will just look at the first of all at a basic differential BJT differential amplifier.

(Refer Slide Time: 05:07)




Which is something in front of you, which you can see here, right. This is a basic BJT differential amplifier in which we have got two transistors Q 1 and Q 2, we are assuming that Q 1 and Q 2 are perfectly symmetrical, symmetrical in the sense that they have exactly the same profile, they will give you the same amount of collector current for the same amount of base emitter bias and so on and so forth, that is the meaning of Q1 and Q2 being symmetrical, right and their $V_{BE}$ is since the silicon we assume it to be 0.7 volts because it is the silicon junction which we are assuming.

We also obviously a load resistance $R_C$ of the order of few ohms or kilo ohms, we have a plus $V_{CC}$ here which is basically the power supply and these two Q 1 and Q 2 are fed by a current source which is I which as you can see here and this current source I is basically a current source which is basically and is considered to be an ideal current source and therefore, if you look at the impedance at this point Z will be obviously infinity in infinity, right for ideal current source.

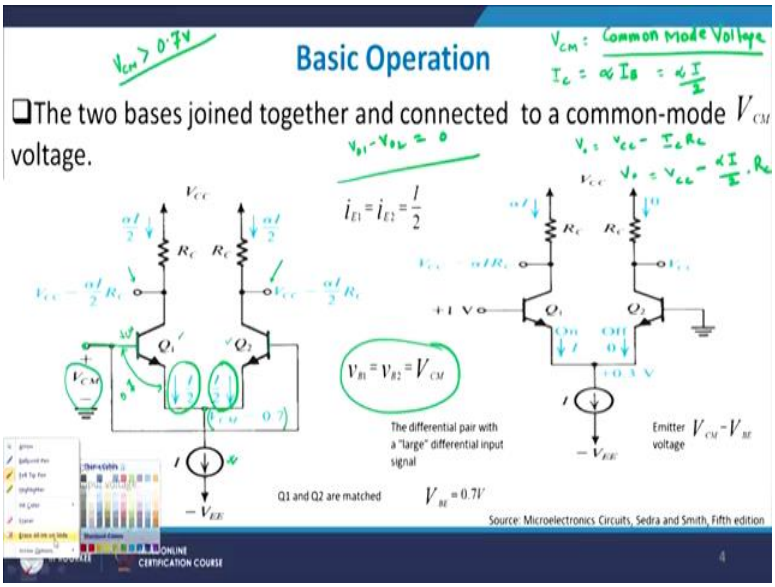
Which means that, at this point it is the impedance looking at this point is particularly infinitely large and whatever current flows through Q 1 and Q 2, if you add the currents you will always get the total current to be equals to I. So, what I get is? The $i_{E1}$ plus $i_{E2}$ will always be equals to I, capital I. so, this constraint always holds good, that the total current flowing through the left arm, which is this one and the right arm, which is this one will always be charge.

If you want to find out the value of $V_{C2}$ and $V_{C2}$, then it is very simple $V_{C1}$ will be equal to $V_{CC} - I_C * R_C$, right because this is the voltage drop which you see, right this is the $-I_C*R_C$ plus there will be also $V_{ce\ sat}$ which will be coming because of this one for the device to be saturation if provided I allow Q1 and Q2 to go into saturation, will see that we will not allow it to do that but these are the few constraints which one has to work upon, right.

That the total current will be always fixed to I and we are having $V_B$ equals to 0.7 which is the cutting voltage for an npn. We also assume that Q 1 and Q 2 are perfectly symmetrical in nature, with this knowledge we also say that now we apply a voltage $V_{B1}$ on the base side of Q 1 and $V_{B2}$ on the base side of Q 2, right.

If we assume that these two are correct then what we can safely say is something like this that that we can say that which that the collector current $I_{C1}$ and $I_{C2}$ which is the current flowing through the two collector arms of these networks are also equal provided $V_{B1}$ is equals to $V_{B2}$, right. So, if I assume that $V_{B1}$ equals to $V_{B2}$, $V_{B1}$ equals to $V_{B2}$ we will see that mathematically later on then if the same bias is there everything else is same $I_{C1}$ will because $I_{C2}$, right.

(Refer Slide Time: 08:12)




So, this is the basic differential amplifier this thing diagram. Let me see, what happens if we apply the same potential to the base of Q 1 and Q 2, so what I have done? I have simply applied a positive bias here and the same potential is applied to the base terminal of Q 2 and this is basically known as a common mode $V_{CM}$, $V_{CM}$ is referred to as a common mode voltage, right.

Let me or explain all these things later on but at this stage just assume it to be common mode voltage as the name suggests, this is the voltage which is common to both the arms, right. So, you will ask me why do you require it? Well see, you initially as I discussed with you earlier also that you want this Q 1 and Q 2 to be in the region where you can actually achieve an amplification right.

So, you require a DC bias from an external world, right a DC bias which will fix this device into certain regions of operation where it can do an amplification then superimposed on that you will have an AC signal being available to you. So, these basically or a DC signal which are giving, assuming at this stage what I am saying is correct and under such a criteria let us see what happens.

If I assume, that I am giving $V_{B1}$ equals to $V_{B2}$ equals to $V_{CM}$, the common mode signal, right so I gave $V_{CM}$ here. Obviously, this will give you a potential drop of 0.7, fine and therefore the potential at this particular point will be $V_{CM}$ minus 0.7. so, you have applied a potential DC bias here which is $V_{CM}$ a 0.7 voltage drops takes place here, so at this point the value of the voltage will be $V_{CM}$ minus 0.7, right.

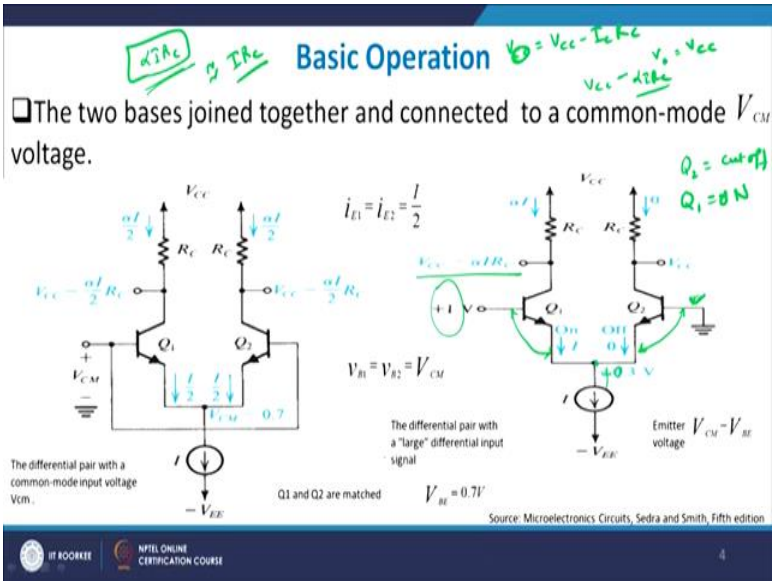
So, which means that both the devices Q 1 and Q 2 will be in all state provided $V_{CM}$ is greater than 0.7, fine. So, if my $V_{CM}$ is greater than 0.7 volts I would expect to see that both the devices will be non-state, right. So, they will be on, when they are (equ) they are on they will be carrying since both of them are on equally they will be carrying equal currents. So, therefore if total current is basically equals to I, Q 1 will carry I by 2, right and Q 2 will also carry I by 2.

So, half the current will be flowing through the left arm and half the current will be flowing through right arm and therefore if you want to find out the current through this collector remember, if you are from a basic semiconductor the code says is that IC will be equals to alpha times $i_E$, so if your $i_E$ is equal to I by 2, I get $I_C$ equals to alpha times I by 2. So, $I_C$ is equal to alpha I by 2.

So, therefore as I discussed with you that $V_{CC}$ or output $V_{out}$ will be equals to $V_{CC}$ minus $I_C$ times $R_C$ right, so I get $V_{CC}$ minus $I_C$ is nothing but alpha I by 2 into $R_C$ this is equals to $V_0$, right. Now you see, since it is perfectly symmetrical this output voltage and this output voltage are exactly goes to each other. So, if you if you want to find out the differential of that means if you subtract $V_{out1}$ - $V_{out\,2}$, $V_{out1}$ - $V_{out2}$ automatically I will get a 0 volt.

Which means that, it is very, very important stage that any bias equal bias given or DC bias equal bias given to the two base of a symmetric differential amplifier the output will always be equal to 0, right. So, DC bias and equal biases to both the arms does not give you a differential output voltage. Now, with this knowledge or with this idea let me move forward and maybe show it to something else.

(Refer Slide Time: 12:07)




That now, what I do is? I make a small change that at Q 2, I ground it base of Q 2 is grounded and my base of Q 1 is given 0.7 volts, so what will happen is? So, this Q 1 will have this again 0.7 volt here, so 1 minus 0.7 you will achieve a point 3 volt here, right. Now, if you give a base which is grounded, so base to emitter junction will be just 0.3 now because this is 0 and this is plus point 3 please understand this an npn transistor, so you reverse biasing a your emitter base Junction, right.

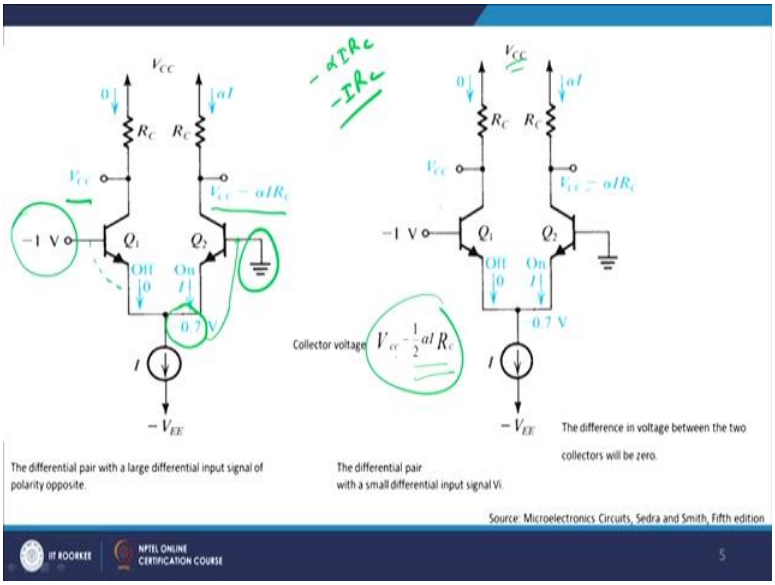
And therefore Q 2 will be basically cut-off and Q 1 will be on, I hope you agree with this point that Q 1 will be on and Q 2 will be off. Which means that, Q 1 will be responsible for biasing all the currents through it on to IC, fine and therefore as you can see here if you if by again by the previous statement which I gave you, we sorry $V_{out}$ will be equals to $V_{CC}$ - $I_C$*$R_C$.

Now, if you if you see $I_C$ a for the right arm Q 2 $I_C$ equals to 0, so $V_{out}$ will be equals to $V_{CC}$ and for the left arm it is $V_{CC}$ minus alpha by 2, it only alpha by 2 divided alpha I $R_C$, see VCC minus alpha I $R_C$, fine. Now, if you find the difference between this and this, what the answer you will get is alpha I$R_C$, right. Which means that if I give a difference voltage of + 1 volt

between the left arm and right arm I get output which is basically a DC output and is given by alpha I times $R_C$, alpha is very close to 1.

So I Can approximately this to be as it equals to I times RC where I is the current flowing through the device. This is with the assumption that Q 2 is perfectly off and Q 1 is perfectly on and all the current is routed through Q 1 all the current of this current source is routed through Q 1, so, this is with the basic assumption is there. Now, so Q 1 and Q 2 are matched I am assuming this to be true right.

(Refer Slide Time: 14:29)



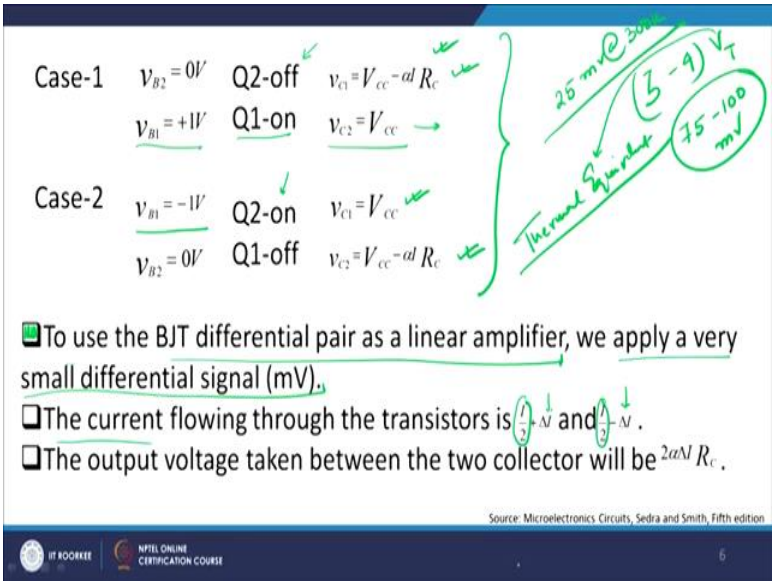

Now, let me do one small thing and let me make a small change once again. With this let me say, that now let me say my Q 2 was grounded, right but my Q 1 I have given a minus 1 volt here which means that if this is minus 1 volt here and if this is grounded, right assuming this there is a potential drop here, I will get 0.7 minus 0.7 volts here, minus 0.7 primarily means that this will be 0 bias between this and this it will be 0 bias.

So, what it will happen is? That Q 2 will be now on and Q 1 will be off, right. So, understand the relationship I am going for both the positive swing and the negative swing in my input when I did a positive swing my Q 1 was on and Q 2 was off with Q 2 grounded with my negative swing of course Q 1 will go off and Q 2 will on, why? Because my Q 1 is going to have because you are reverse biasing the emitter base Junction and therefore Q 1 goes to off state but then this ensures that Q 2 goes to a non-state because minimum point 7 volt is already there between the emitter base Junction.

And therefore it will be an on state and therefore the reverse will happen that my $V_{out}$ 1 will be equal to $V_{CC}$ and $V_{out2}$ will be equal to $V_{CC}$ minus alpha times $I_C$ and therefore the output will be actually equals to minus alpha IRC assuming alpha to be approximately equals to 1, I get output 2 because 2 $IR_C$. let us see, so with this basic knowledge what things we have been able to found out? We have been able to recapitulate or found out or find out that if same voltage is given to the two base of the inverter BJT the output will be always equals to 0, right.

What therefore, what does the differential amplifier do? It tries to sense the difference between the two signals and then tries to amplify it, right so that is the reason it is known as a differential amplifier. So, if you look at the collector voltage, it is $V_{CC}$ right minus $I_C$ so it is alpha I by 2 into $R_C$ provided both the devices are on, I get minus alpha by 2 alpha I by 2 into $R_C$ which is $V_{CC}$ minus alpha $R_C$.

(Refer Slide Time: 16:43)




With this knowledge or with this basic idea, let me therefore move on. So, as I discussed with you in the first case Q 2 will be off and therefore V C1 will be equals to $V_{CC}$ minus alpha I times R $_C$ and $V_{B1}$ when it is plus 1 volt Q 1 will be on and then V $_{C2}$ will be equals to $V_{CC}$. So, whichever the transistor is off the output will be latched to $V_{CC}$ and whichever transistor is on the output will be latched to $V_{CC}$ minus alpha $R_C$.

In case two when you get $V_{B1}$ is equals to minus 1 volt in that case Q 2 will be on as I discussed with you and therefore Q 1 will be off, Q 1 off implies that VC1 will be equal to $V_{CC}$ and Q 2 on implies that $V_{C2}$ will be equals to $V_{CC}$ minus alpha $IR_C$. so, all the current is

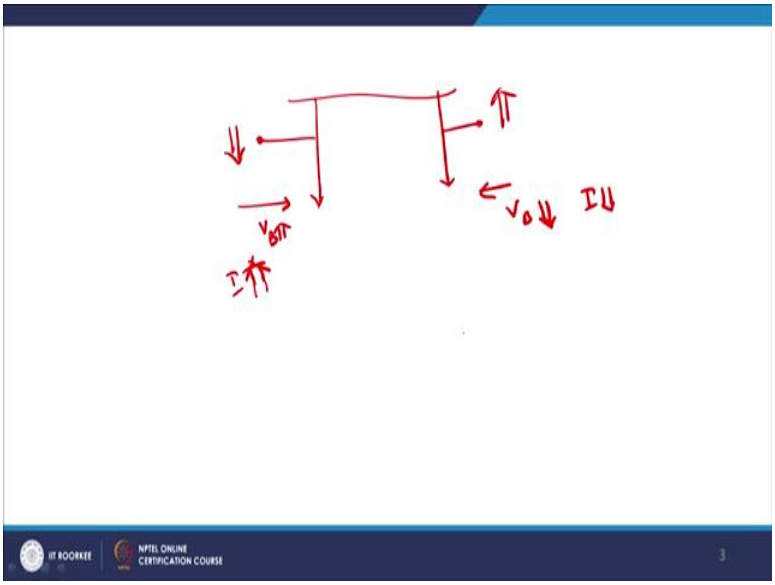
routed through Q 2 now. So, there are two cases across which the differential amplifier to work.

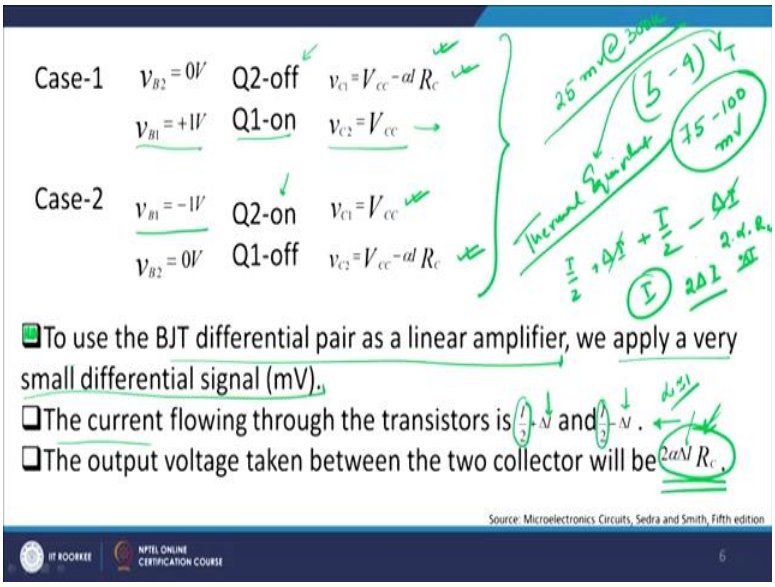
Therefore just see this point which I am shading here, that to use the differential amplifier BJT differential amplifier, as a linear amplifier linear amplifier means? That the gain is independent of the input voltage, a fixed gain is available to me, we have to apply a very small differential signal of the order of few millivolts, right. Typically generally it is approximately 3 to 4 $V_T$ we say, if the difference is this much we will expect to see a change 3 to 4 $V_T$ is approximately 200 millivolt difference would be there, $V_T$ is thermally equivalent voltage is approximately equals to at 300 Kelvin.

This is, we have already discussed this point is known as thermal equivalent voltage which means that minimum difference of 3 to 4 $V_T$ should be there for the switching or cut-off or switching to be proper. Now, therefore the current through the transistors, the second point will be I by 2 plus delta I and then I by 2 minus delta I, you understand reason why it is, in one case it is minus 1, once case it is plus?

And the reason is something like this, that assuming that as I discuss with you in the previous slides, let us suppose let me just explain to you here, right and I explained here, ok. Let me explain to you, therefore if I give up in one case if I give a base voltage which is higher in another case I reduce the base voltage then the current here will be larger and on the right hand side it will be smaller. So, I will just explain to you what I am trying to say.

(Refer Slide Time: 19:22)






Let us suppose, that I have got two arms here right and these two arms are going on here, in this arm I increase the base voltage $V_B$ and in this arm I reduce the base voltage $V_B$, so if I increase it the current becomes larger, right the current increases, as the current increases this voltage will drop down here. Similarly, when the base voltage reduces the current reduces and therefore the voltage at this point actually increases.

So, if I therefore make my base voltage make it a differential base voltage, the current will be also differential in both arms and therefore the voltage will also be differentiating in both the arms, that is what I am try to say in the second statement here. Secondly that delta I and plus delta I minus delta I. So one, in one case with the base voltage is increased there will be

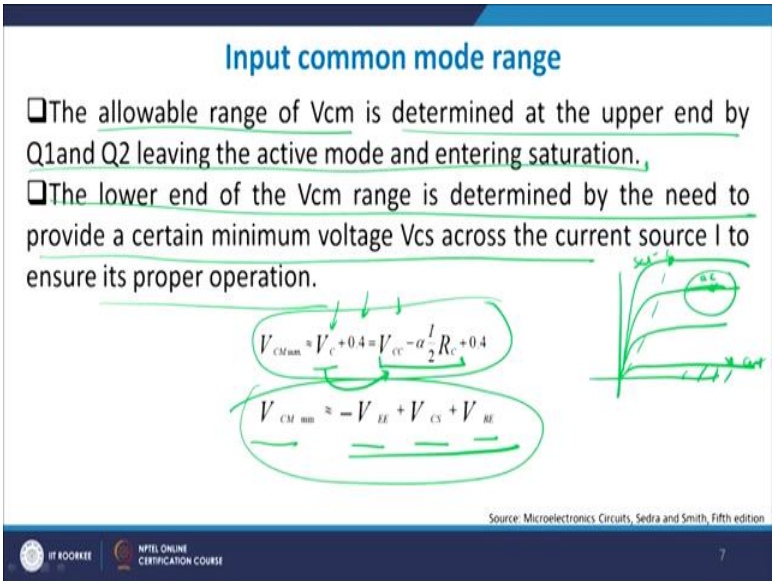
increase in the current plus delta I and in another case where there is a decrease there will be a change in the value of current.

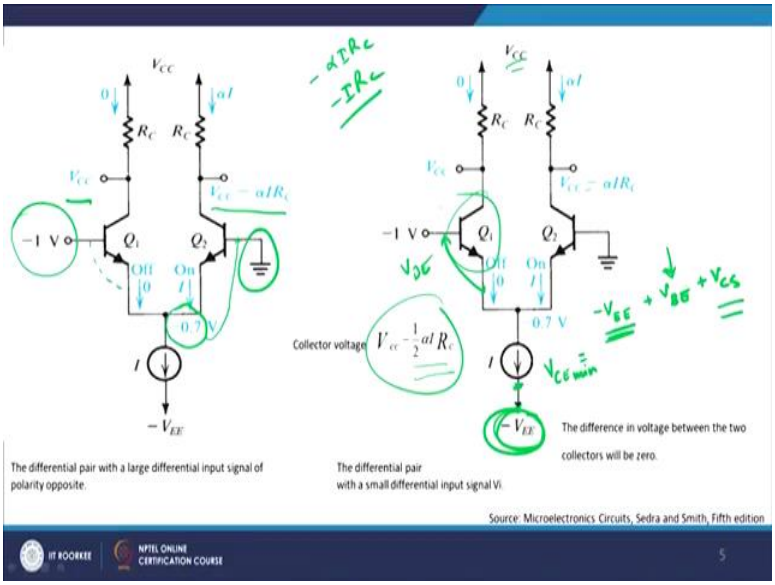
But please understand this delta I will always be same for both arms because if you add these two together I should get I, so I by 2 right plus delta I plus I by 2 minus delta I if you see this cancels off and if you add these two I get I, which is the basic emitter current from the current source, submit violating the basic principle, right and that is the reason this is this thing. And therefore, if you want to take the output you will get 2 times alpha delta I by $R_C$, right.

It will be 2 times alpha times delta into $R_C$, how why he will get two times alpha ? the reason being you get you when you, so when you subtract the two together I get, so if I if you subtract say for example the current is I by 2 plus delta I, so if you add those to subtract the 2 I get I my 2 cancels out and I get 2 delta IComing into picture, so I get 2 delta I by primal means that 2 times alpha times $R_C$ times, right I get delta I which is basically I here, so, I get this much delta I to be this much, so I get this much as the current.

So what, which means that output voltage difference will be higher provided the current difference is also higher, right alpha is typically very close to 1, so that does not play a major role, so delta I higher the delta I value larger will be the differential voltage output difference between the two.

(Refer Slide Time: 21:42)






Now, let us see, what is the allowable common mode voltage? So, the initial voltage is the DC voltage which I am giving to a Q 1 and Q 2 before actually inserting AC bias over superimposing over it. Let us see till what point $V_{CM}$ is allowed, right. Now, the allowable as you can see the allowable range of $V_{CM}$ common mode voltage is determined at the upper end by Q 1 and Q 2 leaving the active mode and entering into saturation mode.

So, what I ensure is that both Q 1 and Q 2 should be in the active mode and from active, it should go back to cut-off and my higher limit of $V_{CM}$ should be such that, that it does not allow you to go into saturation, so I am sitting somewhere in the edge of my active region, please understand this very carefully that if you remember from your basic when we are discussing the BJT and npn transistors, we saw this to be as the saturation region, so this was my active, right this is a saturation region and this was my cut-off, this is cut off below this is cut-off.

So, if you are biasing somewhere here, then you do not bias it very close to saturation region, keep it very much in the active region because even amplification to be there but the highest value of common mode voltage which you can put is basically somewhere here, right and the lowest value is somewhere here. So, the highest value is very near to saturation and the very lowest value is near cut-off, but you do not keep it there because then small change in $V_{CM}$ will result in the device going into saturation and thereby reducing your speed or going to cut off and thereby giving you as erroneous output result.

So, keep it in the middle to find out. The, that is what I will say, the lower end of the $V_{CM}$ range is determined by the need to provide a certain minimum voltage across the current

source to ensure the operation. So, let us see what is $V_{cmax}$? $V_{cmax}$ typically is the collector voltage which you give plus 0.4 which is nothing but $V_{CC}$ minus alpha into I by 2 into $R_C$ plus $R_C$.

So, this $V_c$ is nothing but this whole thing, right and a $V_{CC}$ minus alpha into I by 2 $R_C$ plus 0.4 approximately and $V_{cmin}$ is equals to minus $V_{EE}$ plus $V_{CS}$ plus $V_{BE}$, why? Please understand, why is it like that. If you go back to a previous slide, you see you have bias this with minus $V_{EE}$, right and this is your base to emitter voltage $V_{BE}$, so what was written here? It was written as minus $V_{EE}$, right.

So, minus, so I will write down sorry, minus so it is minus $V_{EE}$ plus sorry, plus $V_{BE}$, right plus $V_{BE}$ plus $V_{CS}$, right. So, this minus $V_{EE}$ is and I am saying this to be as approximately the value of your $V_{CE\ min}$, minimum value why? Because any volt, see so for the device to be on state I require minimal $V_{BE}$ to be potential given to the base side otherwise it will go into off state.

So, assuming that your $V_{BE}$, Q 1 is in the on state, I have to give a minimum voltage of 0.7 $V_{BE}$ onto the Q point, so that takes care of this one, right is it ok, what is $V_{EE}$? $V_{EE}$ is this potential, so minimum this potential has to be there in order, so because this potential is appearing so this is basically a $V_{EE}$ and this potential is appearing somewhere at the edge of your current source here, right.

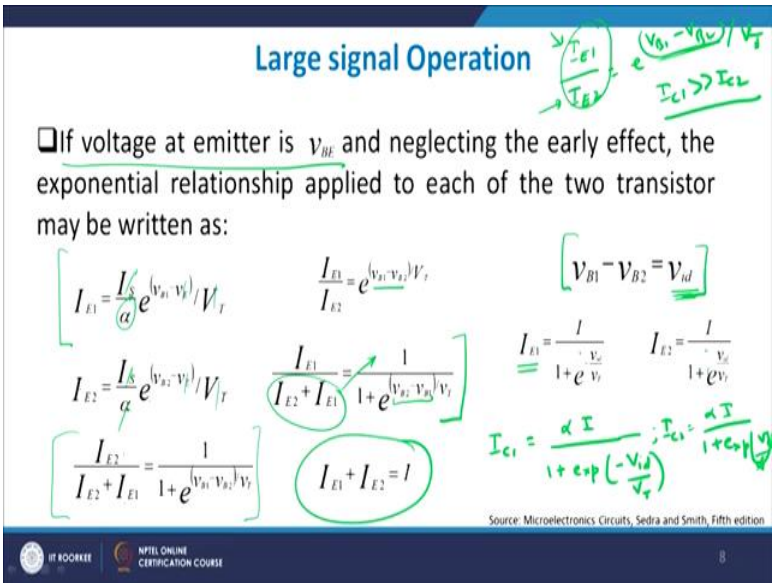
So, the minimum amount of DC bias, so because if you do not give this much amount this will actually reverse bias your sorry, it will forward bias it will give one every time and therefore you require minus $V_{EE}$ as one of the point and we $V_{CS}$ is basically the collector to source potential, which you need to add to $V_{BE}$ in order to ensure that the device is non-state.

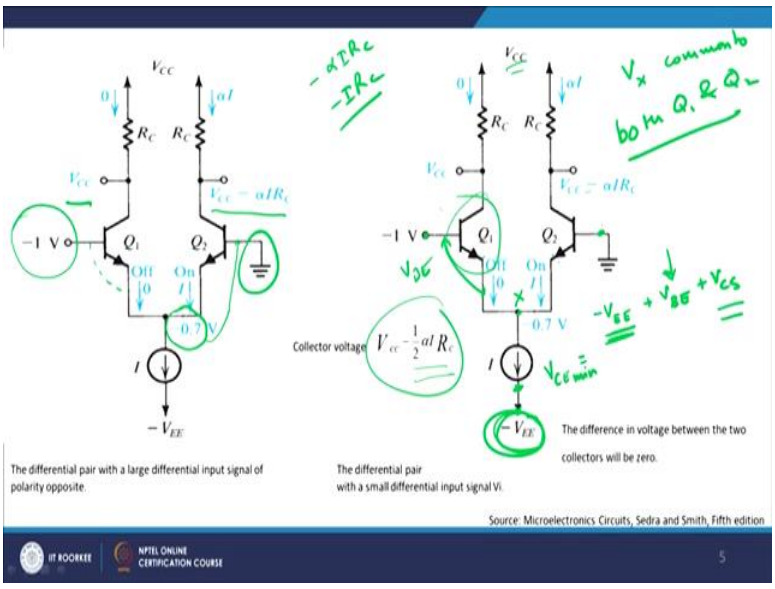
So, the requirement therefore is that for the $V_{cmax}$ common mode max is given by this and common mode min is given by this, right. So, if your $V_{CC}$ is typically around 3 volts or 4 volts or 5 volts then you can see what is the value of $V_{CM\ max}$. so, higher the value of $V_{CC}$ more you can go for $V_{cmax}$ without entering into the saturation region of operation and the critical part is that we do not allow it to enter in the saturation region because then it compromises in the speed of the device, right.

Let us look at the therefore the, we have understood therefore the small signal amount not the small signal amount it, the basic operation. So, what I will do? Is I will give a differential voltage here, if therefore I will get a 180 degree phase shift at signal between Q 1 and Q 2, I

will get an output which is 180 degree phase shifted if you therefore take the difference of the two collector voltages, the output voltage will be twice and therefore I will get a differential of twice approximately.

(Refer Slide Time: 26:52)




Now, if obviously if the voltage at the emitter is given by $V_{BE}$ and therefore assuming that there is no early effect then I Can safely write down the emitter voltage $I_E$ to be equals to $I_S$ by alpha, IS by alpha into e to the power $V_{B1}$, $V_{B1}$ minus $V_E$ by $V_T$, my basic semiconductor equation and $V_{E2}$ equals to $I_S$ by alpha into e to the power $V_{B2}$ - $V_E$. Now, if you add and therefore this equation if you do like this manipulation, I get this into consideration and if I do some $I_E$, so the $I_{E2}$ by $I_{E1}$ plus $I_{E2}$, I get this, I will also get $I_{E1}$ by $I_{E1}$ plus $I_{E2}$ equals to this, right.

Now, you see so $I_{E1}$ by $I_{E2}$, $I_{E1}$ by $I_{E2}$ this will cancel off, right this will cancel off, this will cancel off sorry, this will remain there and therefore we $I_{E1}$ by $I_{E2}$ is equals to e to the power $V_{B1}$ minus $V_{B2}$ by $V_T$ and there even without knowing the physics of the device you can appreciate that higher this difference between the base (voltage) voltages, right more will be the ratio this one.

So, which means that if $V_{B1}$ is much, much larger as compared to $V_{B2}$, right everything else remaining the same $I_{E1}$ will obviously be much larger as compared to $I_{E2}$ and therefore $I_{C1}$ will also be much, much larger as compared to $I_{C2}$. So, that is the bottom like a or that is the sort of a the whole crux of the whole issue, that I need to make one side base voltage much larger as compared to another side.

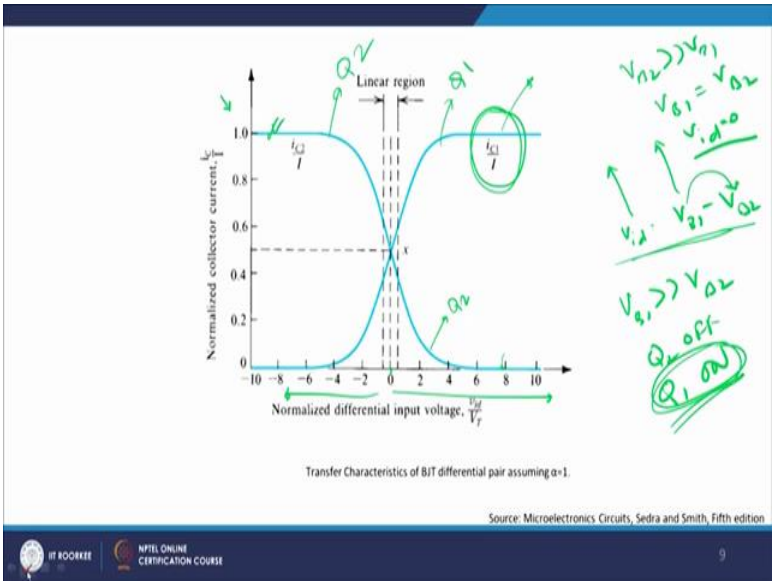
Now, I we also know that $I_{E1}$ plus $I_{E2}$ will be always equals to I, you see this one and we refer to $V_{BI}$ minus $V_{B2}$ as $V_{id}$, $i_d$ means input difference, input difference between the two, why we are saying input difference between the two? Because you see, this potential is always constant, right I will just come back to the previous slide and you will explain, this potential is always common to the both the emitters of Q 1 and Q 2.

So, let me name it as X then $V_X$ is common to both Q 1 and Q 2 therefore since this is common to both I just need to find out the difference between this voltage and this voltage that is all, so that is the reason I am referring to this as $V_{B1}$ minus $V_{B2}$, right I am forgetting the common node because unnecessarily complicate the matters and therefore I refer VB1 minus $V_{B2}$ equals to $v_{id}$.

Therefore I Can write down, $I_E$, right $I_{E1}$ to b equals to $I_{E1}$ to be, so this is basically your I, so this is I this I, I am take it in the numerator and I get $I_{E1}$ plus e to the power $V_{B1}$ minus $V_{B2}$ is $V_{id}$. So, $V_{id}$ minus $V_{id}$ by $V_2$ and $I_{E2}$ equals to I upon 1 plus $V_{id}$ by e to exponential plus $V_{id\,2}$, right. So, 1 is a got a negative sign and other is good as a positive sign.

So, these are the emitter currents which are available to us right and if you want to just find out the collector current you multiply this with alpha, so $I_{C\,1}$ will be equals to alpha times I upon 1 plus exponential minus $V_{id}$ by $V_T$ right and $I_{C2}$ will be equals to alpha I upon 1 plus exponential $V_{id}$ by $V_T$, fine, ok.

(Refer Slide Time: 30:16)




Now, with this knowledge which you have gained till now, right let me just therefore give you a small idea about the transfer characteristics. Now, on the y axis you will see it is basically the normalized collector current which means that collector current normalized with respect to the total current which is available to us and therefore the maximum I Can go is 1 and on the x axis we have $V_{id}$ by $V_T$, $V_{id}$ is basically the difference voltage of the base

So, $V_{id}$ is as I discussed with you is basically $V_{B1}$ minus $V_{B2}$, right so if we divide that by $V_T$ thermal equivalent voltage this is what I get, this is the point 0 where $V_{BI1}$ equals to V $V_{I2}$, right and as you move to the left or right let us see how it works out. As you move to the let us suppose left, right as we move to the left or let us say we are moving to the right that is more easy to understand.

It primarily means that my moving to the right means this is becoming more and more positive which means that VB1 is becoming more and more higher as compared to, so $V_{B1}$ is getting higher and higher as compared to $V_{B2}$ which therefore means that the Q 2 is in the off state and Q 1 is in the on state, when Q 2 is in the off state obviously so sorry, when Q 1 in the on state I will get on current available here at this particular point, right.

And therefore, if you look very carefully that as I and therefore as you can see the collector current which is flowing, now when you make it $V_{id}$ very, very large right this is very, very large as compared to $V_{B2}$, right so Q 1 is in the on state, right Q 1 is my on state, Q 1 is in the on state primarily meaning that you have a large collector current available to it and therefore you see you have large collector current $I_{C1}$ by I is the large collector current whereas since

already Q 2 is a switched off mode the current goes to 0 and you can see the current goes to 0.

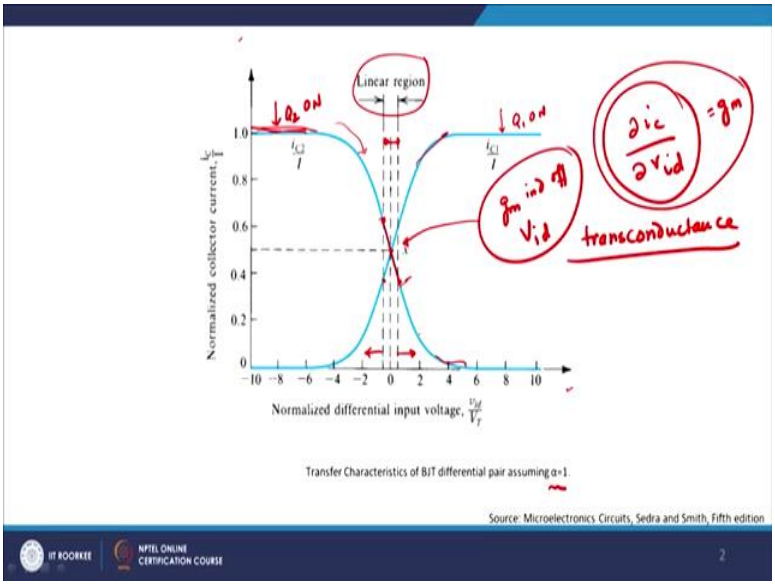
So, this is the Q 2 current and this is Q 1 current sorry, this one is Q 1 current and this is Q 2 current. Similarly, if we go to extreme left somewhere here then $V_{B2}$ will be much, much larger as compared to $V_{B1}$ and therefore your second Q 2 will be on having a large current and Q 1 will be off. So, this is C is a perfectly symmetrical sort of a butterfly available with, it is a perfectly symmetrical, somewhere in the middle where you will have $V_{B1}$ equals to $V_{B2}$ which means that $V_{id}$ equals to 0 you automatically get a place where the current changes drastically with respect to $V_{id}$.

So, we will start from here in the next slide in the in our next discussion, we will see how it works out and we will discuss from the linear region of operation, how we can move ahead and give you the value of CMRR and give you the value of the differential mode common signal, right thank you, thanks a lot.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−34**
**Multistage and Differential Amplifiers-II**

Hello everybody and welcome to the NPTEL online certification course on Microelectronics: Devices to Circuits. In our previous model we had looked into differential amplifiers and we will start of from where we left and let us look at the therefore the transfer characteristics.

(Refer Slide Time: 00:45)




As we discussed in the previous term that, if you go to extreme this is basically the plot between collector current, normalize collector current with respect to difference voltage, normalize with respect to the thermal equivalent voltage which is $V_T$, 25 millivolts. If you go to extreme right as you can see, you will see that the Q 2 will be basically off and therefore the current will be 0 whereas Q 1 will be on and therefore you will get a large current, so this is Q 1 is on, right.

On the extreme left, Q 2 will be on and therefore you will get Q 2 a high current whereas Q 2 will be off and you will get a Q 1 current. Somewhere in the middle where linear region has been maintained, is the region where the differential of current with respect to input differential voltage, so this is basically if you want to find out the slope of this curve from this point to this point, this point to this point it is basically $\partial (i_c) / \partial (V_{id})$, right.

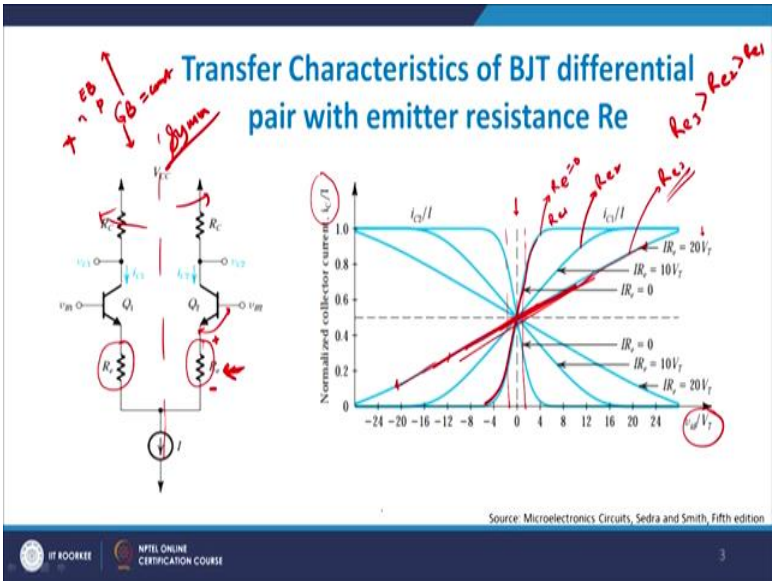
So, if that is the slope you are referring to $\partial ( i_c) / \partial ( V_{id})$ is maximum somewhere in the middle of this region, right somewhere in the middle of this linear region and if you look very carefully this is nothing but the transconductance of the transistor, so this is transconductance, why it is a transconductance? Because you are giving on to the base the voltage, the difference of the base and $i_c$ is the difference of the collector current.

So, it is basically $g_m$ for the whole differential pair, for the whole differential pair this is $g_m$ which you see transconductance. And as you can see, this is the region where gm is almost linear and therefore $g_m$ is almost independent of $V_{id}$, right $g_m$ is almost independent of $V_{id}$ at this particular point. Which means that, take any value of $g_m$ any point, at any point you bias your device and give you peak to peak swing between this point and this point and you would expect to see your $g_m$ to be almost constant and therefore gain to be almost constant and the highest gain also you will get here.

As you move towards this region, this effect sample if you look at this point, the gain will be almost equals to 0, right because the current is not changing with respect to variation in the differential voltage. Similarly, the here the gain will be slightly less but as you shift towards the middle where you have a linear region of operation you get the highest gain, right and therefore, I am also assuming that approximately $\alpha$ is approximately equals to 1 which I am which basic assumption I am taking here.

To improve these linear regions the problem is, if you look very carefully the problem is at linear region is only restricted by this much amount of $V_{id}$, hardly of the other of few millivolts, right few millivolts. So, typically if you want to make it more realistic differential amplifiers you have to increase the linear region of operation which means that you want to increase this 1 unit to increase it, right and make it shift go to both right hand side.

(Refer Slide Time: 03:41)




How do you do it? Well the methodology which people adopted was that they inserted a resistance $R_e$ in series to the (emitted region) emitter region and if you plot there for the graph between $i_c$ by I which is basically collector current normalize with respect to total current versus $V_{id}$ by $V_T$, the same graph as the above, you see as the $R_i$ value becomes larger, right.

So, this is for various values of $R_i$, this is the value with $R_i$ is equals to $R_e$ equals to 0, right this is with certain value of $R_e$ such that $I_e$ into $R_e$ 10 times $V_T$ and this is with the value i into $R_e$ equals to 20 times $V_T$. So, obviously this is suppose $R_1$, $R_{e1}$, $R_{e2}$ and $R_{e3}$ then $R_{e3}$ is greater than $R_{e2}$ is greater than $R_{e1}$, in fact, $R_{e1}$ is approximately equals to 0 we have sorted it, the previous one which means that, if you look very carefully you are actually increase the linearity, look at the point that when you increase, when your $R_{e3}$ is largest, your linear region is extending from maybe from this point to this point to larger extend, right.

So, your linear region is extending, so you can play with this for a large linear region whereas when you make a $R_e$ equals to 0 your linear region is very, very small maybe this just small around as I discussed with previous slide. And therefore putting $R_e$ we will discuss why is it like that but increasing $R_e$ mixed the it is more linear but cost to pay for it is that the transconductance in the gain therefore reduces when your $R_e$ is large, we will also explain this point one by one.

But as you can see the slope here around this particular point around mid when $V_{id}$ equals to 0 is small slope as compare to a slope of this line. So, you see that it putting $R_e$ makes my linear region of operation broaden but the cost I pay for it is that I reduce my overall gain, so

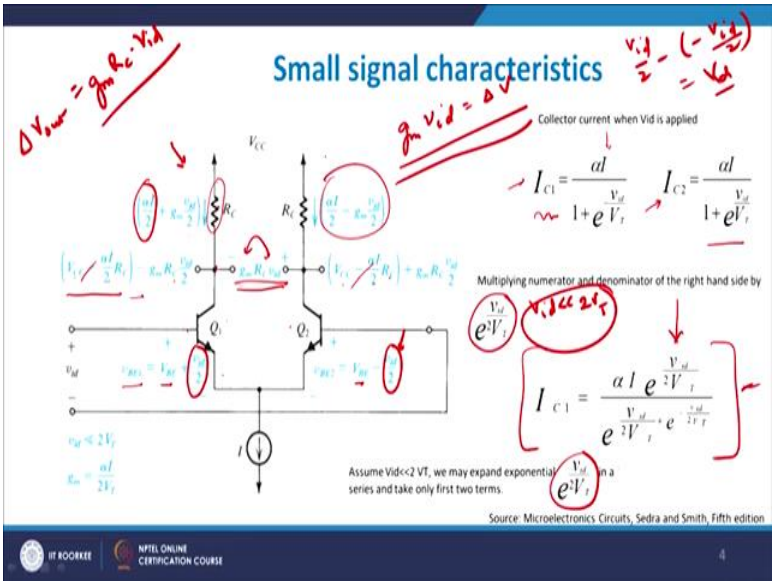
g m reduces transconductance reduces, why does it happen? We will just discuss that point. Say, you have inserted Re here, right you have inserted Re, first let us see why transconductance and therefore your gm or gm reduces in this case.

See, once you reduce when as you put $R_e$, your $R_e$ actually has a bias which is something like this, right and therefore in a sense it is reverse biasing your npn (transfer) emitter base junction. So, I have an emitted base junction, if you given it is npn, so if you give a positive potential on the n side your reverse biasing emitter base junction and it is reverse biasing it, what is happen? The gains fall down because you are not letting down large amount of emitted current to come across and therefore your collector current will be also small.

That is the reason your gain starts to fall down when you insert $R_e$, right and that is the reason your gain and therefore you see gm values are reducing, gain starts to fall down, why does it increase the linear region? Because if you remember gain bandwidth product is always constant, so therefore when the gains fall down you can work with the larger bandwidth, your bandwidth increase and therefore your linear region of operation becomes larger and larger in this case, right.

So, that is the reason why we use, so we can use these two generally make it this thing. Please understand all these discussions are with the classist assumption that my both the arms here, left arm and right arm are exactly symmetrical in nature. If they are not, then this is not a whole good, there will be some changes which you need to do.

(Refer Slide Time: 07:14)




Ok, let me come to the small signal characteristics of the of a BJT bipolar transistor as differential amplifier. So, I have Q 1, Q 2 here, I also have $V_{id}$, $V_{id}$ is basically the difference in the potential of base to emitter of Q 1 and Q 2, so if see $V_{B1}$ is nothing but $V_B$, $V_B$ is the basically the DC bias which we applying + $V_{id}$ / 2, so the difference is $V_{id}$ then if I write this to be as plus $V_{id}$ by 2 then they should be - $V_{id}$ / 2 why? Because $V_{id}$ / 2 minus of minus $V_{id}$ / 2 will give you plus $V_{id}$, right and that is a difference which you see here, that is the reason we write plus $V_{id}$ / 2 or - $V_{id}$ / 2.

Similarly, we can write $I_{C1}$ to be equals to α times $I_{e1}$, now $I_{e1}$ is nothing but I upon 1 plus e ^ - $V_{id}$ / $V_T$ and $I_{C2}$ will similarly will be α I upon 1 + e^($V_{id}$ / $V_T$ ), right. Now, if you multiply the numerator and denominator for this transistor and this one by e to the power $V_{id}$ by e ^ $2V_T$, we get $I_{c1}$ to be equals to this much, right and $I_{c1}$ we get is and therefore I mean if assume that $V_{id}$, right is much smaller as compare to $2V_T$ which is typically is.

Let us assume, we may expand the series in exponential this one, this one is in the numerator and in the series and take only two first two terms. So, what the first two terms will be? I will just discuss that in the next slide but assume this to be as the $I_c$ which is available to us again α $I_e$ to the power so multiply $V_{id}$ by 2 times $V_T$ and then I multiply this to the power e^($V_{id}$ / $2V_T$) plus if I by $2V_T$, I will get minus $V_d$ / $2V_T$, right and that will therefore that is what I am doing here.

If you look, therefore on this side from the on the base side, I get $V_{cc}$ - α I / 2 * jab Rc, remember why? Why is it minus α Ic / 2? The reason been the total current flowing through

Rc on the left hand side will be nothing but α I / 2 which is the dc bias which you see, we have already discussed this point earlier plus g m multiplied by Vid / 2 because Vid/2 is the approximate voltage which is visible between base emitter that you multiply with gm you get the currents, so this is what you get $I_c$.

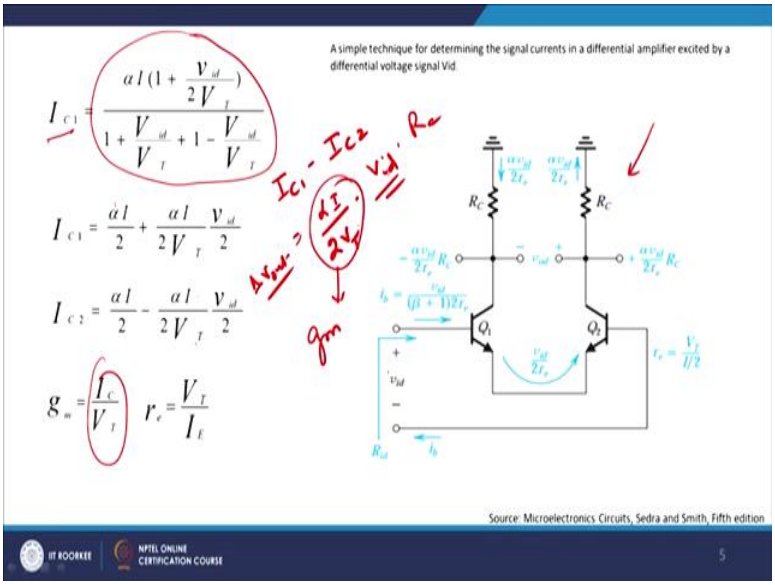
But on the right hand side, you will get a minus V id because you see this is minus $V_{id}$ / 2, so I get minus $g_m$ * $V_{id}$ / 2, right we get minus $g_m$ * $V_{id}$ / 2. Now, quite interestingly if you simply add subtract these two α I by 2 at α a by 2 cancels off and you what you get is? Basically $g_m$ times $V_{id}$. So, even without knowing anything if you just simply find the difference between this and this, the current difference this is ∂ voltage difference, sorry ∂ voltage difference which you see, right and you get ∂V * $I_c$ which you see.

Now, you see, so what I get if we closely therefore come into the point then I can simply this, so if you look from the right hand side I can simply multiplied $V_{cc}$, right minus (alp) so if the potential drop across $R_c$ I need to find out will be nothing but $V_{cc}$ minus α I by 2 into $R_c$, right that is the potential which is appearing across $R_c$ - $g_m$ * Rc into V id by 2, right because you have to subtract these voltage to achieve the voltage at this particular point.

Similarly, if get this I have to multiply this with R $_c$ and therefore I get $V_{cc}$ - α I by 2 into $R_c$ α I / 2 * $R_c$ plus, why plus? Because there will be a negative sign here, so minus into minus is plus gm * $R_c$ / $V_{id}$ / 2. But if you subtract this from this, this whole thing get cancelled with this whole thing, this remains so this minus of minus this will give me $g_m$ * $R_c$ * $V_{id}$, fine.

So, the output voltage difference between the two is nothing but gm $R_c$ * $V_{id}$. So, this is ∂ $V_{out}$ is equals to gm $R_c$ * $V_{id}$. So, if your $V_{id}$ is large you get a very large value of ∂ta $V_{out}$, if your $g_m$ is large you also get a very large value of $V_{out}$ and this is true also, we have understood from our basic understanding previously.

(Refer Slide Time: 11:42)

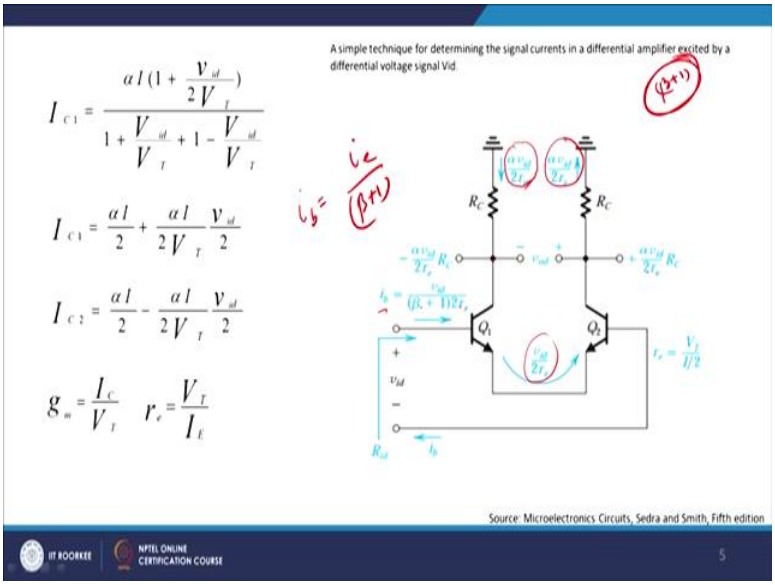



Now, let me find out the, this what I was discussing with you an earlier sense, that IC1 therefore can be written as this by expanding it by expanding the previous discussion and similarly $I_c$ and therefore if I break it down I get α I by 2 plus α I V is by 2 times $2V_T$ and $I_{c2}$ is equals to α I by 2 minus α I $2V_T$ / $V_{id}$ / 2. So, if you add these tw o, two, so if get the $I_{c1}$ - $I_{c2}$ as I discussed with you, you will get the this will gets cancelled out and there will get you, this will get added up and you get α I by $2V_T$ into $V_{id}$, we will get, fine.

That is basically difference in current, this multiplied by $R_c$ will give you with the difference in voltage. Now, you see very well that this is nothing but the Charles conductors of the device gm and therefore we can safely, so that is what I am saying $I_c$ / $V_T$ is transconductance of the device and therefore I can safely write down this to be as $g_m$ * $V_{id}$ * $R_c$ which we have already derived in a our previous our case, right, ok.

So, this gives me an idea about the basic understanding and therefore as we have already discussed the transconductance of a device higher the transconductance of the device more will be the output voltage, higher the value of $R_c$ more will be the value of it. So, this is basically your ∂ ($V_{out}$) which you see, right, If you look at this graph here and what I am trying to tell you here is, if you look at the figure here, you see the Q 1 and Q 2 are both in active region of operation.

(Refer Slide Time: 13:12)




And if we now try to find out a how much amount of input resistance is there? See, when you are looking at MOS as devices, MOS based devices, right in MOS base devices you were actually looking into the gate side, with then signal was inserted to the gate side of a MOS device which is obviously giving you infinitely large input impedance because gate was having a dielectric constant of oxide and therefore this was problem whereas in this case which is in this case you are actually looking into the base side.

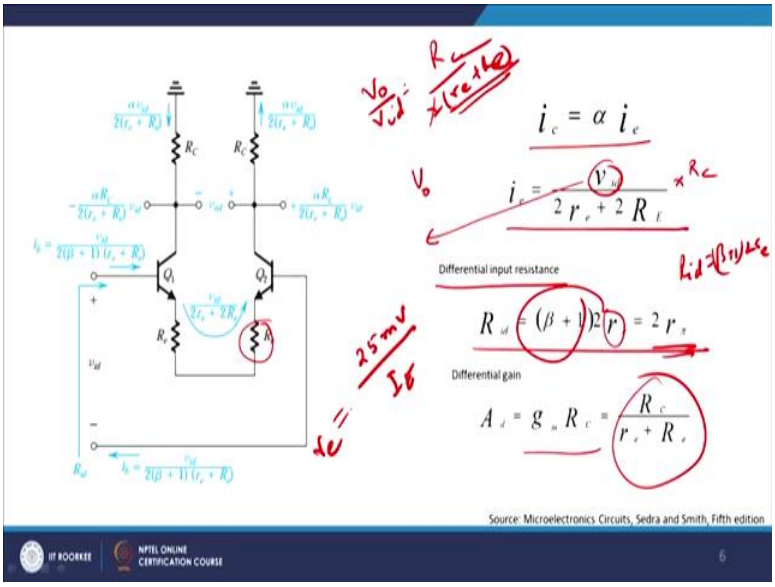
So, base side will have a finite resistance and that is what you are trying to find out, how do you find out? You find out the base current, right. So, base current if you want to find out if it was basically a basic idea that if the resistance looking from the source side, is let a suppose twice R e, right then on the base side it will be B plus 1 times $R_e$, so that is what I am trying to, so we it will you have to multiply by β plus 1 when you looking from the base side.

So, the base current will be Vid by $2R_e$ * β + 1. Similarly, on this side, so where Vid by 2Re is the difference voltage which you see in front of you therefore if the same current flows through, so α times Vid, so current is basically $V_{id}$ / 2 $R_e$ that is emitter current if you multiply by α that is a collector current, so this is the collector current here and this the collector current here.

If you multiply this with Rc I get the voltage at this particular point with a negative sign and since there 180 degree phase shifted I will get a positive voltage here which is a α $V_{id}$ by 2 $R_e$ * $R_c$, right and the base current will be $V_{id}$ / 2 $R_e$; $V_{id}$ / 2 $R_e$ is nothing but emitter current

divided by β plus 1. So, $i_b$ equals to emitter ie upon β + 1, remember and from there I get this as the emitter current which you see, right.

(Refer Slide Time: 15:05)

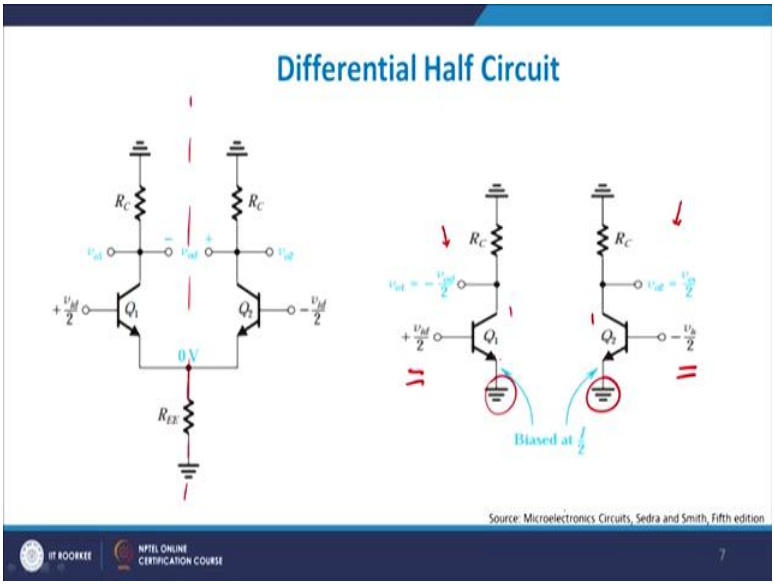



Now, with this region I said $i_c$ equals to α i and i therefore is equals to $V_{id}$ / 2 $r_e$ + 2 $r_e$ we get. So, the input differential resistance $R_{id}$ is equals to β plus 1 * 2 $r_e$, right and this β plus 1 into 2 $r_e$ is also referred this is $r_{pi}$ and therefore we referred to this as $R_{id}$ is equals to β + 1 * 2 re, re is the resistance looking from the emitter side, right.

And the differential gain is $g_m$ times $R_c$, $g_m$ we have already find out to be some value which is $R_c$ by $r_e$ + $R_e$, right $R_c$, how did you find out differential gain if you want to this will be this you need to multiply with $R_c$, right you multiply with Rc you get the Vout. Now, this Vid if it goes down I get $V_o$ / $V_{id}$, right I will get $R_c$ / 2 $r_e$ + $R_c$, so this 2 will be very small as compare to the anything else I get $R_c$ / $r_e$ + $R_e$, sorry re, right $R_e$ is the resistance offered from the emitter side and $r_e$ is the dc resistance which you have actual resistance which you have put.

So, if you see a differential gain depends upon the ration of your collector resistance to that of the emitter resistance plus the $r_e$ value, $r_e$ value is the small signal $r_e$ which you see to it is basically 25 mv millivolt by $I_E$, this is equals to $r_e$, right and this is what we get as a differential gain.

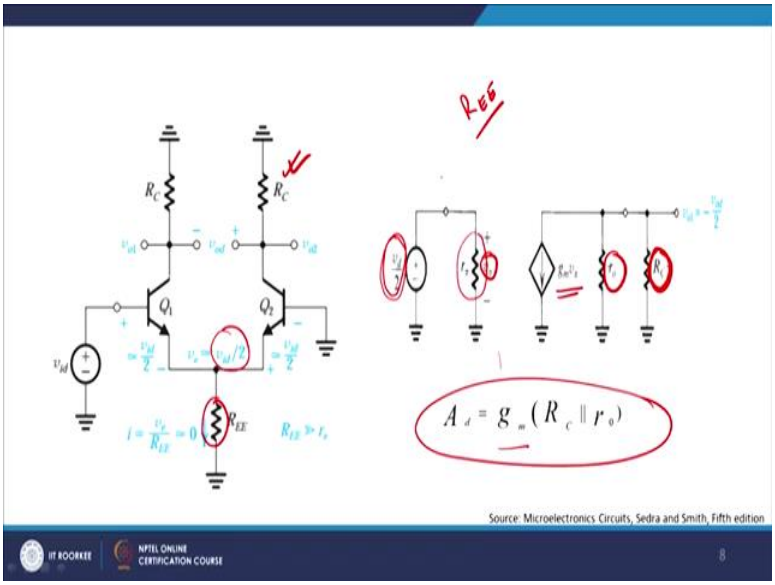
(Refer Slide Time: 16:36)




Now, the concept here is that if everything is symmetrical both the left arm and right arm are symmetrical, we can divide the whole differential circuit in a two half circuits. So, what we do typically is that we break it down this middle and then we say that on the left hand side $V_{id}$ / 2 is the input voltage, on the right hand side - $V_{id}$ / 2 voltage and they are biased such that I / 2 current flows through this arm and this arm and therefore $V_{od}$ / 2 at this point and $V_{od}$ / 2 at, so minus plus is there, right and therefore the output difference, did you find the difference between the two? It will be still V equals to $V_{od}$.

So, this is the concept of differential half circuits that overall transconductance will be divided by 2, you will also have the bias current divided by 2 and you what you do? Is that you emitted part of the BJT is actually grounded here, right they grounded here and once they grounded they act as a differential half circuit, right.

(Refer Slide Time: 17:34)

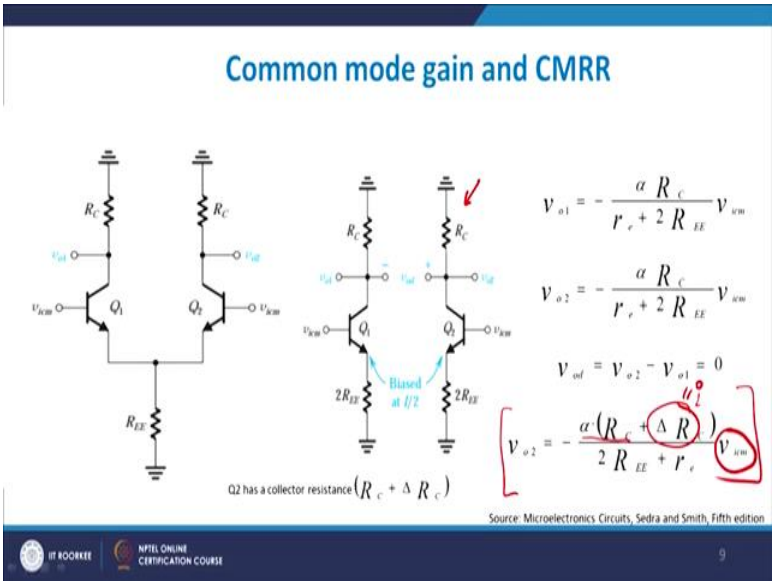



The advantages it is much easier to calculate various formulas from differential half circuits. With this knowledge, let me therefore come to a small signal characteristic of the differential amplifiers. As you can see my input voltage, I have already told you is differential $V_d$ / 2, right so I have giving $V_d$ / 2, this $r_{pi}$ is nothing but the input side resistance offered by the differential pair. So, it is basically the pairs $r_{pi}$, right.

If you look at the right hand side since it is $V_{cc}$ voltage control current source, any $V_{pi}$ which you give in the input side you will have $g_m$ * $V_{pi}$ as the current flowing in the output side, right. So, this is your this, this is your collector resistance which you have already found out and $R_0$ is the resistance offered by the device itself, right, so, these two are in parallel. So, if you even close your eyes do any problem, I simply can get the differential gain to be equals to $g_m$ * $R_c$ || $R_o$, right.

So, transconductance multiplied by output impedance will be actually equals to the voltage gain and that is what we have shown here, right. This is with the assumption that my $R_{EE}$ which is the, so what I have done here? Is that I replace the current source by your resistance and this resistance value is very, very large, it is typically very large, right and it does not let any current to flow through a typically and therefore I can safely write down V to be equals to $V_{id}$ / 2 which you see here, $V_{id}$ / 2. So, the emitter resistance, so emitter voltage is exactly equals to $V_{id}$ been $V_{id}$ / 2 in both the cases.

(Refer Slide Time: 19:04)

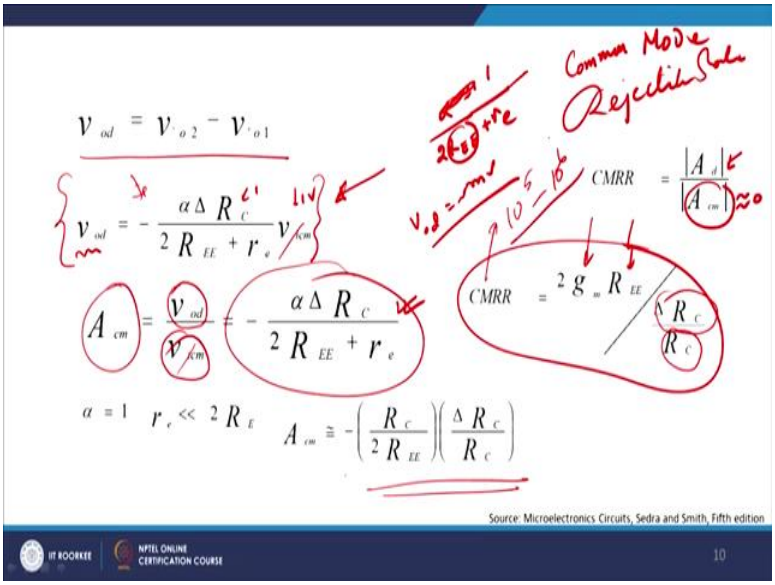



Let us look at the common mode gain and CMRR. To what we will be doing and what we will be understanding is basically that, let us suppose that in reality, of course, your left hand, right hand side will never be symmetrical with respect to each other, that we already know because there will be some mismatch in the resistances, some mismatch in the emitter resistances, some mismatch in the character resistances, the transistor itself will be mismatch in terms of thermal equivalent voltage is through and so on and so forth.

Now, under a criteria that you do have a change, let us suppose that my one of the arms $R_c$ changes by $\partial R_c$ then I can safely write down $V_{o2}$, you can do it yourself is given by minus $\alpha$ times $R_c$ plus $\partial R_c$ / 2 $R_{EE}$ + $r_e$ * $V_{icm}$, $V_{icm}$ is nothing but the input common mode voltage, right, I am not giving any differential voltage now. So, under the presence of a input common mode voltage my output voltage $V_{oc}$ comes out to be this.

So let us suppose there would not have been any $\partial$ R, so this would have been 0, this would have been 0, right this would have been 0, and I would have got a $\alpha$ $R_c$ / 2 $R_{EE}$ + $r_e$. Now, because of this $\partial R_c$ I have an extra voltage term which comes into picture.

(Refer Slide Time: 20:12)




Now, therefore $V_{od}$, which is a output difference voltage is $V_{o2}$ - $V_{o1}$ will be given by this quantity. So, which means that even in presence of a equal voltage at the base of the two transistors because of a $\partial$ R change in one of the resistances I always will get a $V_{od}$ equals to this value which is given by this value. So, $\partial$ $R_c$ say 1 ohm then I will always get off, so V in equals to 1 ohm, this is 1 volt and this is also equals to 1 ohm, then there will be an α upon 2 $R_{EE}$ + $r_e$, right α is very close to 1, so I will get almost equals to 1 upon this whole quantity, $R_{EE}$ is very large quantity.

So, typically what I will get is, $V_{od}$ few millivolt there will be an always offset available to me, the order of few millivolts. We define therefore $A_{cm}$ which is basically my common mode we defined this to be as common mode gain, common mode gain is basically my output voltage under the condition that I am given a common mode voltage in the input side, I get this to be true, you see this $V_{od}$.

So, if you divide by $V_{icm}$, this $V_{icm}$ cancels with this $V_{icm}$ and you are left with $A_{cm}$ to be equals to this much. Now, you can write down $A_{cm}$ to be also breaking up into this value, you just simply can do a small manipulation and CMRR which is basically known as common mode rejection ratio is defined as the difference of the or ratio of the differential gain to common mode gain. Which means that, if I give a common mode signal to both the arms of my bipolar differential amplifier I record a gain, right and that gain will be by virtue of $\partial$ R whatever change is there and I also record the gain by giving a differential signal to both the arms, I divide 1 upon another and should get the CMRR. Ideally CMRR should be infinitely

high, the reason you can understand, why? Because your $A_{cm}$ should be (approxi) will should be ideally equals to 0, ideally your $A_{cm}$ should be 0.

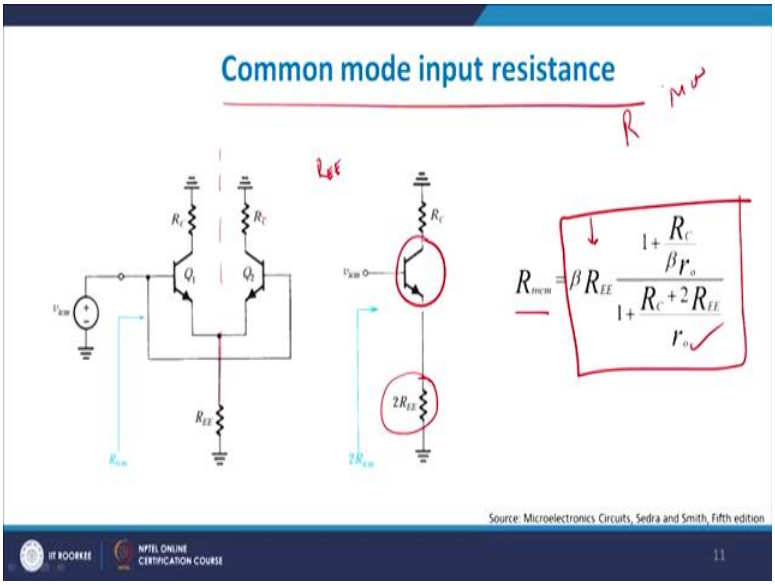
We, have seen in the first few slides of my previous picture that $A_{cm}$ should be approximately equals to 0, the reason been for a exactly the same bias apply to do transistors, the output was actually equals to 0, so my $A_{cm}$ should ideally equals to 0, it is not by virtue the fact that both the arms are not symmetrical in natures, there is some difference, so there is some finite value of $A_{cm}$.

Similarly, Ad will be some particular value. So, if you divide $A_d$ /$A_c$, I get this as my CMRR, so it is 2 $g_m$ * $R_{EE}$ upon $\partial$ $R_c$ / $R_c$, right and this gives you a value of your CMRR, Re obviously is very large value, $g_m$ is also very large, $\partial$ $R_c$ is very small, $R_c$ is also relatively large of the order of kilo ohms. So, you get CMRR very large of the order of 10 to the power 5 to 10 to the power 6, ideally you should get infinity but this is a order which you get for CMRR.

Higher the value of therefore the CMRR, a common mode rejection ratio better the differential amplifier in rejecting common mode signals and accepting differential mode signals or amplifying differential mode signals, right. Typically you will see as you may be we can stick up in the next slide or maybe in the next lecture, that noise is generally common to both the arms, noise is not differential.

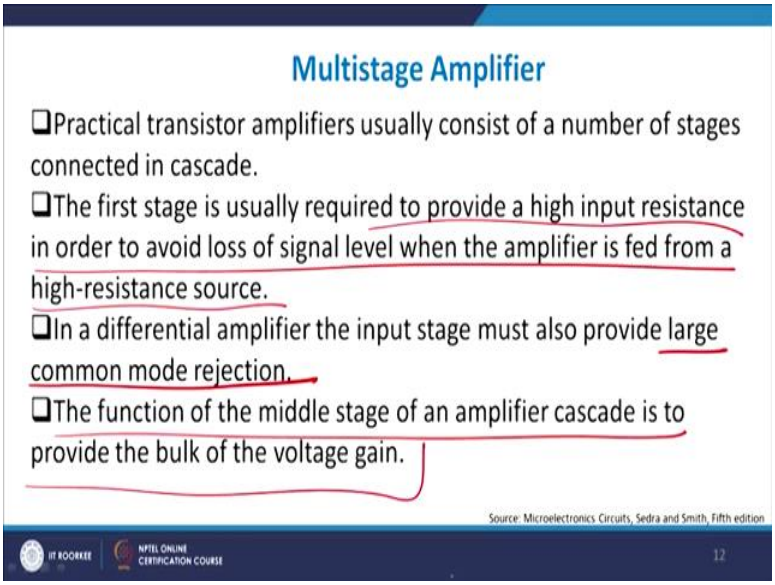
So, therefore a differential amplifier will be a very good rejecter of noise and very good amplifier of signal and therefore higher the CMRR value of differential amplifier better the design is as far as it is input resistance is concerned, ok.

(Refer Slide Time: 23:56)




Let me therefore explain to you, what is known as common mode input resistance? We have seen it already and it is resistance looking from the input side to the base side of (my) of a design, right. So, remember half mode signal I / 2, so if I take (so I will get R $_E$) so I get $R_{EE}$, so I get $2R_{EE}$ available to me because this time just overlapping with respect to each other Rc, $R_c$ and therefore I get $V_{icm}$, so $R_c$ in comes out to be just definition I am not deriving it in the class but it is β times $R_{EE}$ upon 1 + $R_c$ / (β) $r_o$ and this is the value of your $R_{incm}$. So, this is the input resistance, it is actually infinite in case of MOS differential pairs but infinite value when you have this into consideration, right. Let me come to the multi stage amplifiers, well in practical stages you will have maybe each differential amplifier driving another differential amplifier so on and so forth.

(Refer Slide Time: 24:56)

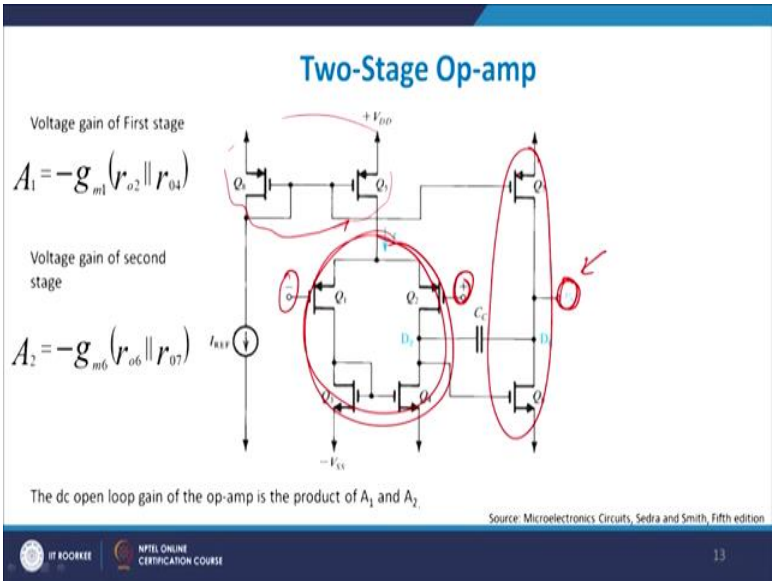



The first it is usually, so if you have a multistage amplifier typically, the first stage is require to provide a high input resistance and order to avoid loss of signal level. So, if I, so if you have put two stage amplifier, the first stage is generally used to match the impedance, so the first stage is got a high input impedance because typically my source has got a high output impedance to match the two, I generally the first stage is basically having high input impedance.

In the, in differential amplifier the input stage must have very large CMRR, right and therefore that is what I written here, right. So, the function of the middle stage of an amplifier cascade is to provide the bulk of the voltage gain. So, if I gave got two stages, three stages let us suppose, the first stage is just to match the impedance level, so that the maximum power is transferred.

The differential part will be responsible for high gain by virtue of by noise reduction, by virtue of high value of CMRR and maybe the third stage or middle stage will be responsible for high value of voltage gain, right. So, we will come to that in a much more detail manner.

(Refer Slide Time: 26:04)

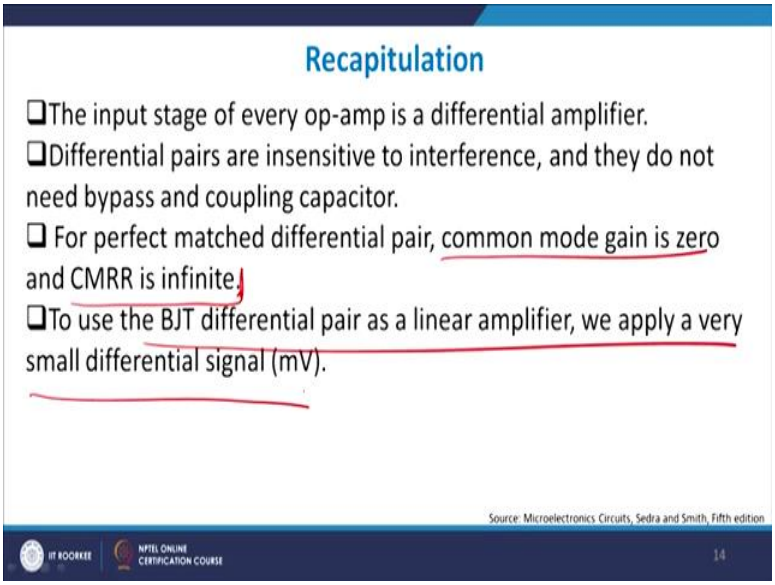



At this stage, this is just a diagram of a two stage op-amp. We have just give you an idea about in reality because when we because after this lecture we will do the MOS differential pair then it will become clear, maybe we will revisit once more when we come to op-amp later on but if you look carefully this is nothing but an replication of a BJT differential pair. If you look carefully, this is the differential pair which you get here, right and this is the differential pair which you get here where I am giving this is the input differential R which I am giving here and this is the single ended output which I am taking from the sub, right.

So, this pair is responsible for doing all impedance matching, this one is responsible for giving you a large gain and this one is responsible for again impedance matching and giving a proper output, right. So, there are three stages, one, two, three stages available here. Let us now look into this basic fact here, these are simple facts which you should be aware of. At this stage, we will discuss op-amp maybe in details, we will come to revisit this formula once again.

So, just to recapitulate or what we have done till now. So, the first stage of any amplifier design is basically a differential amplifier, right gives you a very high gain. As I discussed with you differential amplifiers are insensitive to noise, they are very good rejecters of noise but they are very good enhancers of input signal.

(Refer Slide Time: 27:35)




For perfect CMRR common mode gain is zero and CMRR is infinite as I discussed with you. So, if you want to use BJT as a linear amplifier, we apply a very small differential signal, if you apply a large differential signal you might end up having a highly nonlinear profile available with you, right because then you move over from the linear region. So, you generally apply a small input signal, right.

So, this may we have actually taking care of the very basic understanding of a differential signal. We, in the next lecture series, we will take a MOS differential amplifiers, right and then explain to you how is a difference from BJT difference amplifier then we will see that MOS is relatively slower in switching as compare to BJT because in BJT if you remember one basic idea was we are moving from active to cut-off and vice versa, we are not atoll moving the saturation.

So, once you do not like let the device move into saturation, the time take into come out from active to cut-off is much faster as compare from saturation to the to active, right because you are in saturation you are storing large amount of charge in the base side which you are have to first remove, right and therefore it is much faster. So, BJT based amplifiers are much, much faster as compare to MOS based, right.

So, but we will look into others factor as well for example power dissipation is higher much higher in case of the MOS devices as compare to in case of BJT as compare to MOS device, fine. With this I think, we have finished with single stage amplifier, basic understanding

using BJT. In the next we will look into the single stage amplifier, differential amplifier using MOS devices, thank you.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−35**
**MOS Differential Amplifier-I**

Hello everybody and welcome to the NPTEL online course on Microelectronics Devices to Circuits. In our previous module we have understood a differential amplifier implementation using a bipolar transistor and we saw that that differential amplifier has gotten an added advantage that it is a very good rejecter of common mode signals which means the signals which are common to both the inputs they will reject it and the differential signals are getting amplified.

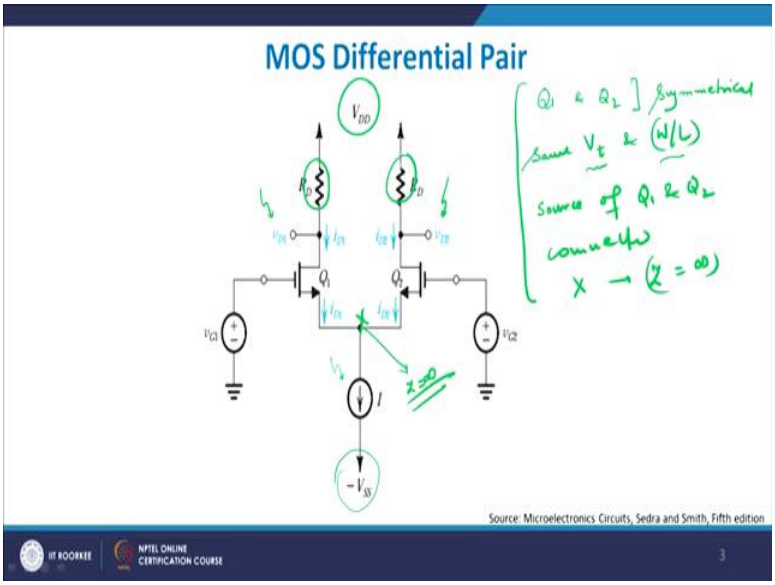
So as long as the signals are phase shifted with respect to each other by 180 degree. I would expect to see a large amplification taking place in the output whereas a common mode signal will relatively have almost 0 amplification and we also therefore, saw that common mode rejection ratio which is basically the differential gain upon the common mode gain $A_{DM}/A_{CM}$ is actually infinitely large for a ideal differential amplifier.

So we also saw that it is commonly used for operational amplifier, the differential amplifier is the first stage in an operational amplifier where gain is where gain has to be very high. Of course, the price you have to pay for is, of course, higher power dissipations because you are working with current source, you are also working with three to two input devices and to load devices and so on and so forth. So, switching activity will be higher and therefore your power will be also very large.

We also saw in the previous discussion that you do not allow the device to go into the saturation mode, you let it go from active to cutoff and cut off to active, and the reason being when you put it into saturation mode and you want to move away from saturation mode you have to remove large amount of charge from the base side, and therefore, the time taken for you to switch from cutoff to active and vice versa will be relatively large.

So, therefore BJT switching speeds are relatively very high right but the loss which you get is a larger power dissipation, right. So what we will do today is have a look at most differential amplifiers so what we did in the previous turn was a BJT or a bipolar transistor based MOS amplifier, this time you will be looking into a CMOS or a MOSFET based amplifier.

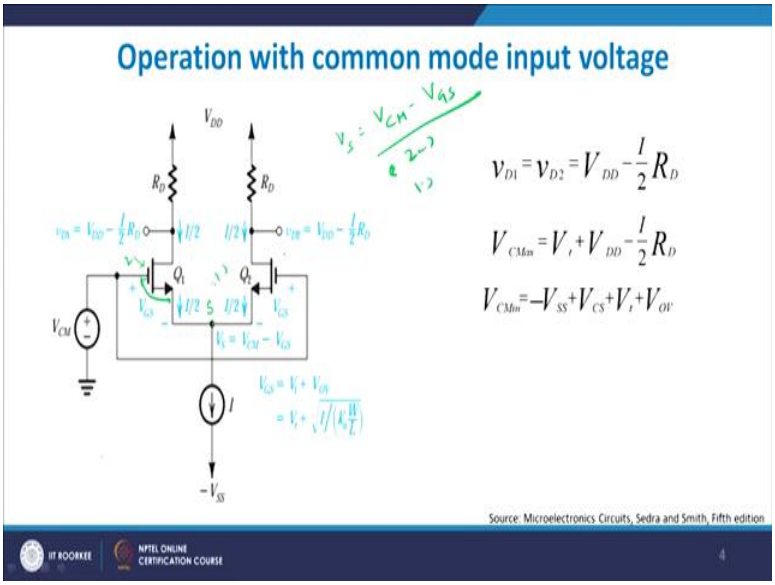
(Refer Slide Time: 3:06)




So the topic of this discussion is the outline of this slides of this module is we will first look at a MOS differential amplifier, just as we did for BJT we will first look at the common mode and then with differential input signals. We will look at the last signal operation and then we look into differential to single end conversion. We will look at active mode MOS differential amplifier we will also look into the active mode differential gain, and then look into input resistance, output resistance of a differential gain and then look at the CMRR and then finally we capitulate the whole thing right.

So, this will be the flow of today's lecture and so we will be first looking into a MOS as a differential pair. It is exactly the same like a BJT design, only thing is here only thing which you should remember at this stage is that when the gate voltage crosses a threshold voltage of the device, the device is in the on stage and it draws current from the power supply, so this is the only thing which is which is not there in a BJT. In a BJT you need to have we have your base voltage at least 0.7 larger than your source voltage for the base emitter voltage to be forward biased. Here you need the gate voltage to be at least larger than the threshold voltage of the device and therefore, that will ensure that you are into the active region of operation.

(Refer Slide Time: 4:29)




Let us look at the differential most differential pair so this is basically a small differential pair which you see this most differential pair is primarily made up of two transistors Q 1 right and Q 2 right and we are assuming that Q 1 and Q 2 are both symmetrical which means that they are exactly same having the same $V_T$ and W / L right. So they have exactly the same threshold voltage and same aspect ratio and we applied two gate voltages $V_{G1}$ plus $V_{G2}$ are the two input signals.

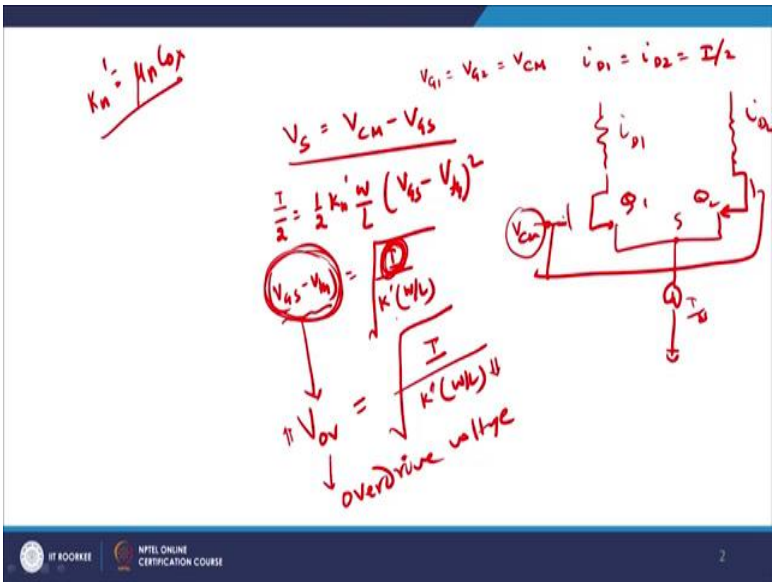
We also have a current source here I which is connected to a negative power supply minus VSS and this eye is basically current source which primarily meaning that this impedance here is typically very high. So, here if you want to find out the value of Z it will be approximately equals to infinity because it is terminating on a current source. We also have a load here Rd so there are two loads here $R_d$ and you take two differential outputs $V_D$ one is one differential output and $V_D$ two is another differential output, right.

As I only discuss with you this potential let us suppose this we name it as X right X will obviously have the same potential and therefore so if you look very carefully the source of q1 and q2 are connected together so what we are doing is that the source of q1 and q2 are connected right the source of q1 and q2 are connected. And we can always say that that this part X right since its Z is equals to infinity high impedance node I can safely say that whatever current is flowing is basically by virtue of the charge carriers available in the MOS device right and that makes our life or calculation easy, relatively easy.

It is not connected to $V_{DD}$ which is the basically the power supply and I have a here you have got that here I have got the source which is available here. Now, let me explain to you therefore with the knowledge with the basic concept or knowledge here let me give to you an idea about two things. Firstly, is that let us assume that you have a common mode operation so if you see here this is the common mode operation right so what we what we will be looking into is that up with common mode input voltage so if I have a common mode input voltage how does it work out.

So if you see here what I have given here is I have given an the $V_{cm}$ is the common-mode voltage, a DC bias which is given here and I am giving a the same DC bias exactly at this particular point $V_{CM}$. So if you look it (trans) clearly $V_{gs}$ will be nothing but $V_{CM}$ – or this this protects oppose this this point is then vs will be equals to $V_{CM}$ - $V_{gs}$ right because if this is 2 volt $V_{gs}$ is say 0.7 then 2 minus 0.7 is, 0.2 - 0.7 is 1.3 so as will be equals to 1.3 volts. So this is what we what we get from here.

(Refer Slide Time: 7:46)




With this knowledge we can write down therefore right we can write down that let us suppose $V_{G1}$ equals to $V_{G2}$ equals to $V_{cm}$ right and therefore $I_{D1}$ equals to $I_{D2}$ equals to I / 2, $i_{d1}$ and $i_{d2}$ are basically the current flowing through. Let me draw for you the diagram here once again so this is the current which is there. This is Q1 Q2 and this is $I_{D1}$ and $I_{D2}$ is the current flowing through it and this is basically your Vcm which ever applied + $V_{cm}$ here and same + $V_{CM}$ is applied here and then you have a current source here which is basically $i_{ss}$ and this is what you get.

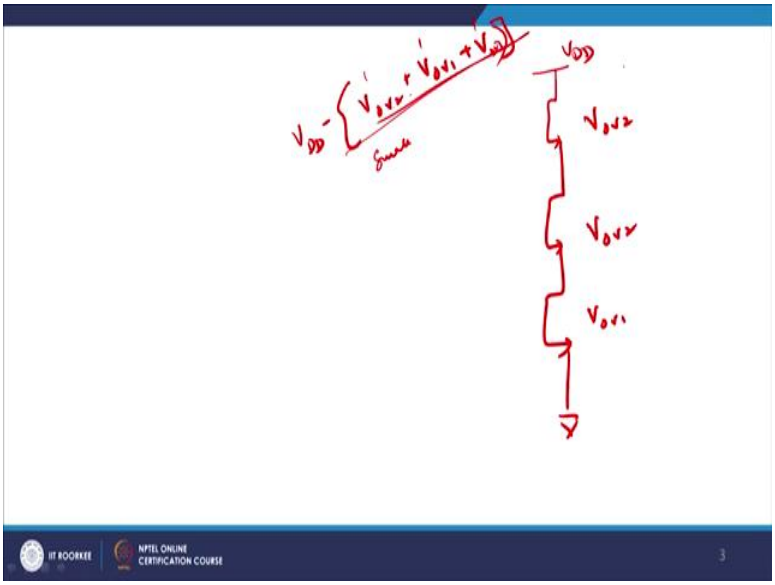
So if you if you this is point S, so I what I get is that $V_s$ will be equals to $V_{CM}$ - $V_{GS}$ right so I can safely write down if I assume that Q 1 and Q 2 are perfectly symmetrical and I apply the same value of voltage $V_{cm}$ to both both the base of the transistor here other gate of the transistor here then I can safely write down that I / 2 is equals to 1 / 2Kn' W / L $(vgs - vth)^2$ where $K_N$' is equal to mu * C oxide.

Fine, I get this into consideration and therefore I can write down vgs - vth to be approx. equal to by K' W/L which means and if you look very carefully this is nothing but $V_{overlap}$ so $V_{overlap}$ is overdrive sorry $V_{overdrive}$ is nothing but 1 / K ( W / L ). This is known as overdrive voltage in a MOS device. So what drive voltage is difference between the gate to source voltage and the threshold voltage of the device right and we define that to be as equals to I / K' * ( W / L ).

Which means that if you make it W / L small right please understand you make it overdrive large. Overdrive large means here $V_{gs}$ - $V_{th}$ is large implying that your current is large right which we so sorry your $V_{gs}$ - $V_{th}$ is large primarily meaning is a current is large but when your current is large you also therefore end up having a larger gain.

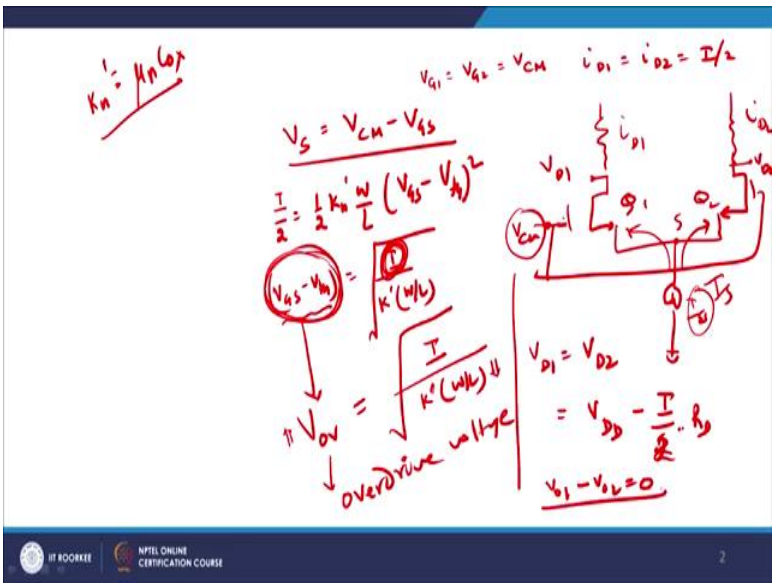
So that is how so what drive is related to gain in such a manner right we will see later on that this what is the drawback of such a scenario is that higher the overdrive is do again is high but your head rooms as I discussed in the previous turns is reduced right. So, if you have a stacked transistor each with head rooms of each with over drives of X Y then your overall head rooms will reduce by X plus y right. I will just try to explain to you what over drive mean to say by that.

(Refer Slide Time: 10:41)




Let us suppose I have got three transistors available here right and the three transistor this is grounded and this is $V_{overdrive1}$, $V_{overdrive}$ to and this $V_{overdrive3}$, then the available headroom at the top will be $V_{overdrive2}$ plus $V_{overdrive1}$ plus $V_{overdrive}$, sorry $V_{overdrive3}$. Which means that higher these values are because you want again to be high more these values will be and then 1 minus $V_{DD}$ minus that quantity will be typically very small. So you will be not left with much room to play with either the near $V_{DD}$ or near $V_{SS}$ right and that's the price you pay typically for the purpose of design itself.

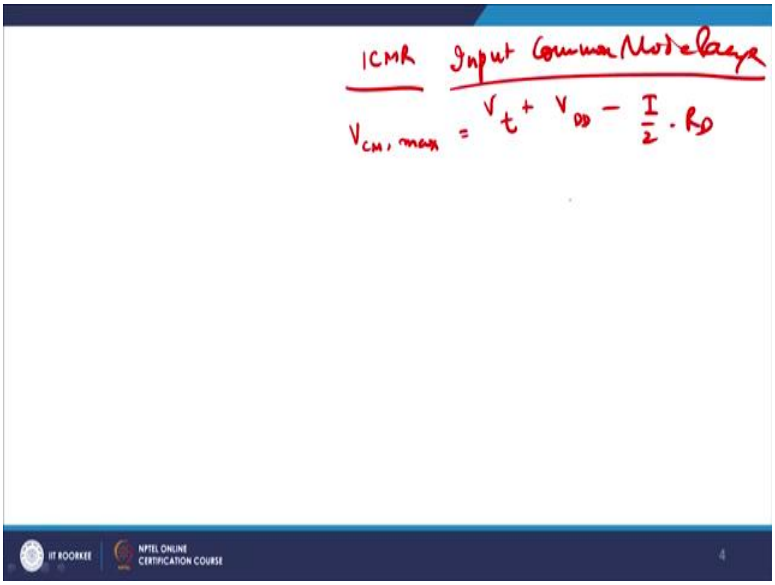
(Refer Slide Time: 11:31)



Therefore, what we can write down all these discussion is that I can write down so $V_{D1}$ equals to $V_{D2}$, $V_{D1}$ $V_{D2}$ are where, this is $V_{D1}$ right and this is $V_{D2}$ right, so we do 1 equals to $V_{D2}$ equals to $V_{DD}$ minus I guess by 2 into $R_D$, understood why $I_S$ by 2 because this is this is the total current is basically, Is so this will move to half on this side and half on this side, so I / 2 right, I / 2 , $V_{DD}$ - I/2 * $R_d$ this much is your $V_{D1}$ and $V_{D2}$

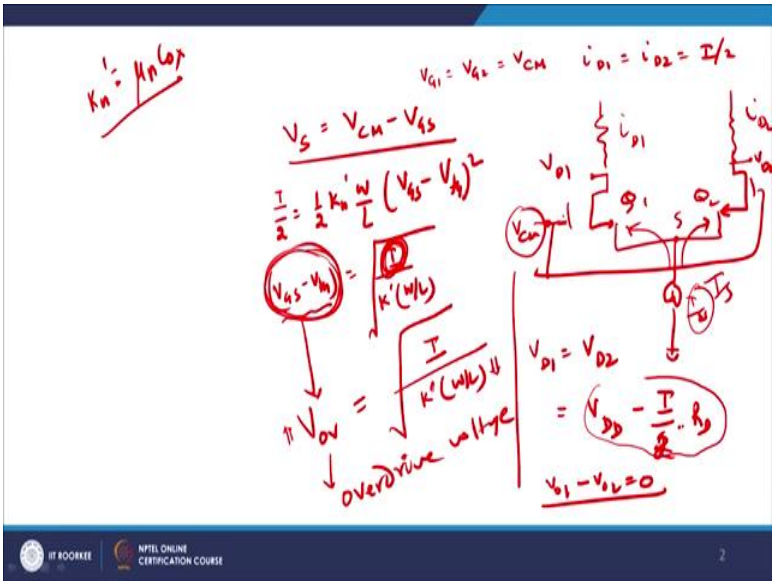
Now, if you since you've applied the same potential on the gate side of the MOS device which is basically $V_{CM}$ therefore subtract $V_{01}$ minus $V_{02}$, i will get 0 which means that if you have a perfectly equal signal being applied to the gate side of the input transistors you do not get any output available right and that is quite an interesting one.

(Refer Slide Time: 12:28)



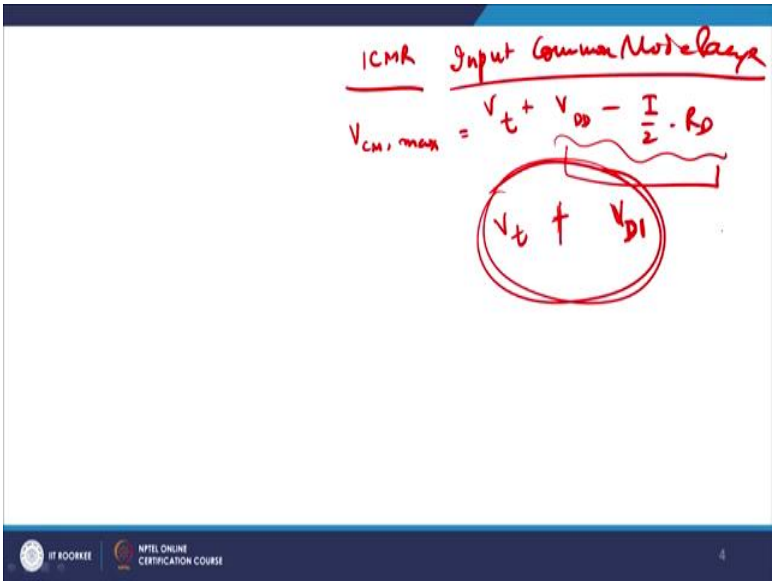

Let us look at therefore the ICMR which is also defined as input common mode range. The right and what do I mean by that I mean to say that what is therefore the minimum and the maximum value, so therefore, so I have applied, we see this, this is what I am trying to say that once you apply input, positive input voltage I need to figure out what is the maximum value of that input bias and what is the minimum value of that input bias right. So, that is my next aim as far as designing is concerned. The next time therefore if so we write down $V_{cm}$ common mode max, which is the maximum value which you still get is $V_T$ plus $V_{DD}$ minus (I / 2) * Rd.

(Refer Slide Time: 13:24)




So if you go back to your previous discussion you saw that your the drain voltage which we are giving was $V_{DD}$ - I / 2 * $R_d$ right.

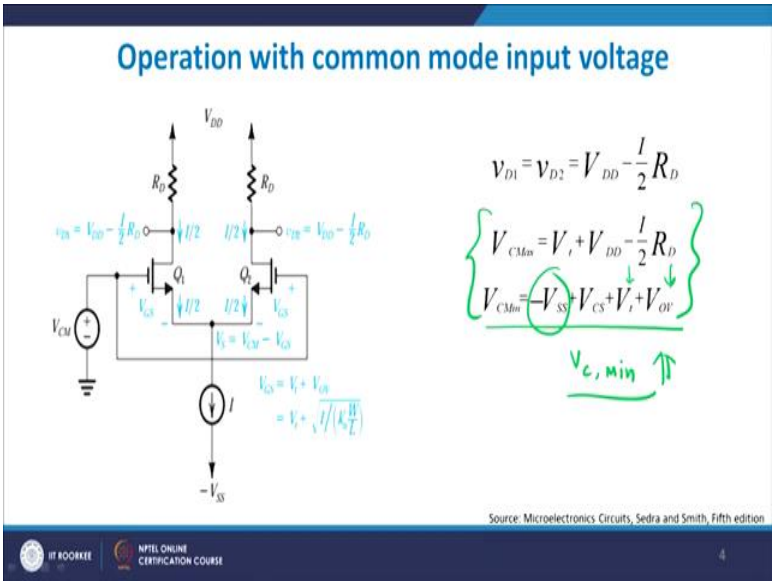
(Refer Slide Time: 13:35)




So it is $V_{DD}$ minus so this is what you get so you get this this is what's the value of $V_{D1}$ is right and we say $V_{Tmax}$ - because this much why because you see when you apply a DC bias it should at least switch on the devices because if it doesn't switch on the device we will never be able to achieve the switching or we will never be able to steer the current to south to left arm or writer arm.

So you want the device to should get switched on. Now to get it switched on the minimum voltage required on the gate side is basically equal to the threshold voltage which is $V_T$ and that is the reason I write $V_T$ here. Plus the same DC bias should able to give a particular value at the output side, of the drain side which is basically $V_{D1}$. So, until and unless you maintain a common mode DC bias of $V_T + V_{D1}$ right, $V_T + V_{D1}$ minimum value I will not be, $V_{D1}$ is the output in the drain side, right, you will not be able to get this much amount of signal in the output nor you will be able to switch on the transistor fully right.

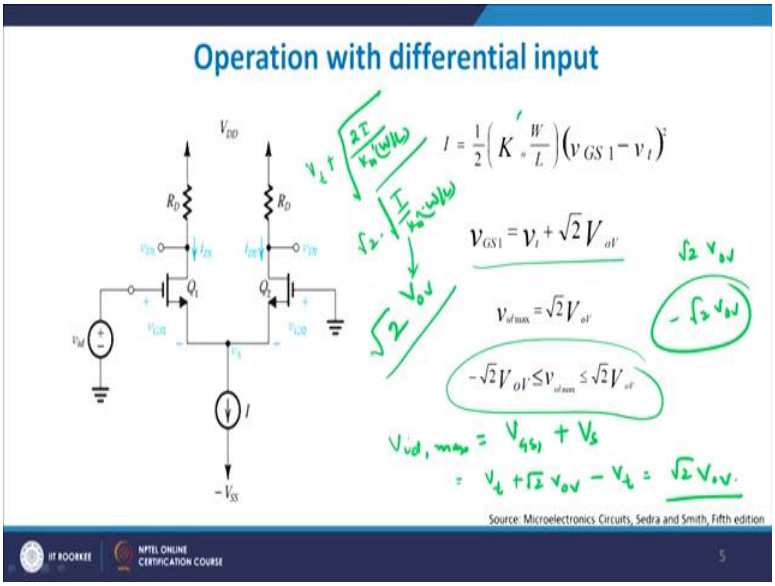
(Refer Slide Time: 14:41)




What is the minimum value which you see, the minimum value is given by this and therefore, you have to ensure that this minimum value which is equals to minus $V_{ss}$ + $V_{cs}$ + $V_T$ + $V_{overlap}$, if anything, if the carbon mode signal goes below any one of them your it might be even true that the current source which we are using might be switched off right and therefore, I will not be able to get any amplification out of it and that is the reason you always apply a minus $V_{SS}$ in the input side plus threshold voltage plus overlap voltage here for $V_{CM}$ right.

So you see if I increase the value of $V_{overlap}$ going on increasing it the price I pay is that my $V_{cmin}$ is actually going high right and that is where pretty difficulty should cover because I want my overall up to be high for my gain to be high but once my gain is high, my overlap is high my common mode signal minimum also starts to become higher. So, I cannot play with large amount of DC biases $I_C$ bias is the input side for a longer duration of time right. So this is what we have done or what we have understood from this basic implication.

(Refer Slide Time: 16:00)

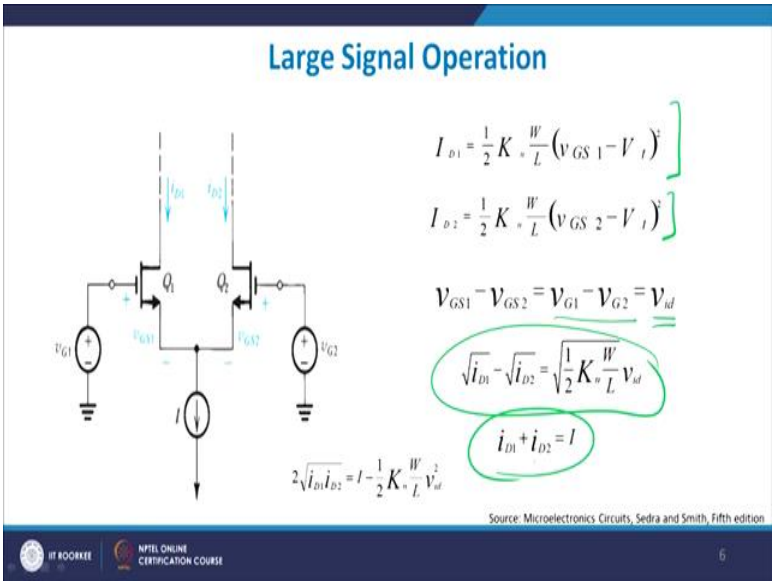


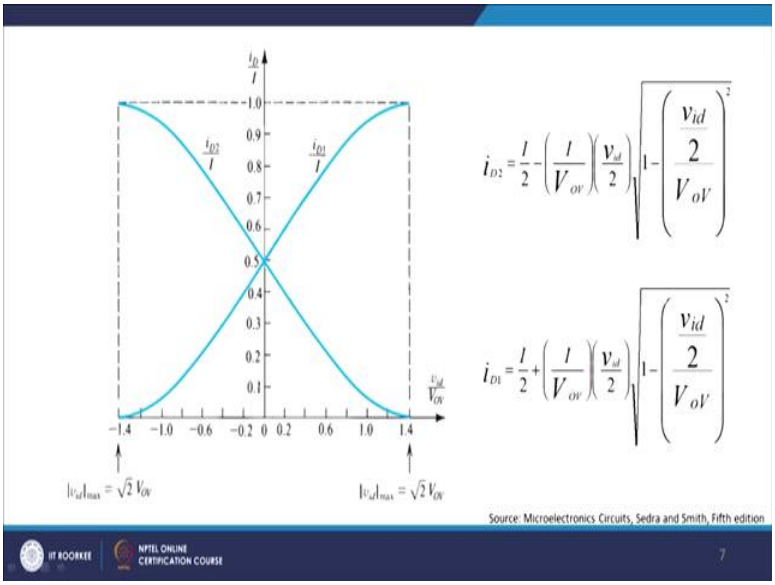

Let me come to the differential mode of operation. The device so as I discussed with you current will be equals to 1 by 2 $K_N$' , this is( $K_N$' * W / L) * $(Vgs1 – Vth)^2$ right. So, if you find out Vgs1 it will be Vt + root(2) Voverlap, why Voverlap because shall tell you. If you if you solve it you solve it I get $V_T$ + root (2i / $K_N$' * ( W / L )) and then if you look very closely it is nothing but root (2 )* root I / ( $K_N$' * W / L) and this is nothing but Voverlap.

So it is root( 2 ) * $V_{overlap}$ right. So, what I get from here is that my Vgs1 should be $V_T$ + root $2V_{overlap}$ right and if I remember $V_{idmax}$ differential voltage max should be equals to Vgs1 + $V_s$ why this is true because gate to source is nothing but the applied common mode signal minus Vs so if you take $V_s$ of the right hand side it becomes plus $V_s$ so they get $V_{Gs1}$ plus Vs. So if you plot it I get $V_T$ + root 2 $V_{overlap}$ - $V_T$ and this comes out to be root 2 $V_{overlap}$ right, and therefore the minimum and maximum value of $V_{ID}$ is just root (2 ) $V_{overlap}$ in the positive side and minus root (2) Voverlap at the negative side.

Which means that if my $V_{IDmax}$ goes beyond root 2 $V_{overlap}$, it goes beyond VD root (2) $V_{overlap}$ then I might enter into a nonlinear region of operation of the device right and as a result I might even intend to try out digit of operation and there will be heavy non-linearity associated with the device. So, I restrict myself to root (2) $V_{overlap}$ both sides negative and positive right and that is a quite an interesting formulation which we see.

(Refer Slide Time: 18:14)



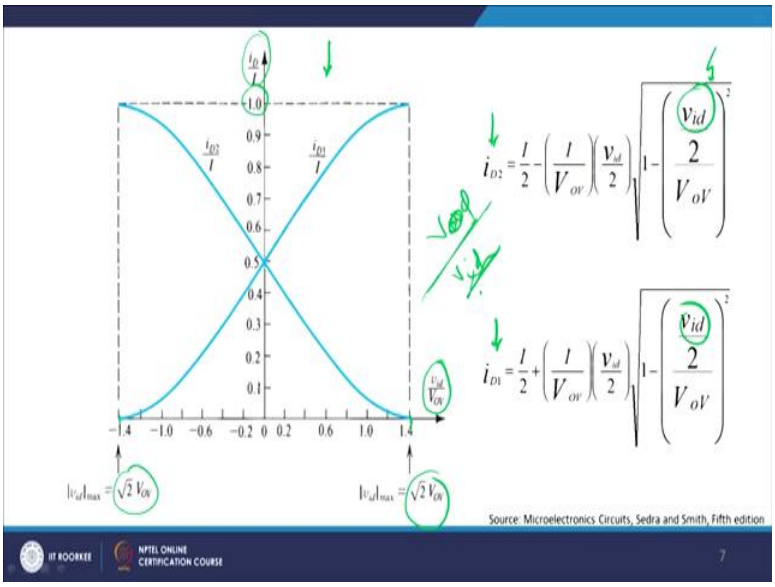




Now with this knowledge which you have gained till now we will do a small not signal analysis here and last signal analysis is required that we have both the transistors here q1 and q2 they are charged with an external voltage source of $V_{G1}$ and $V_{G2}$ and there is a current source here which is basically behaving like an ideal current source, which is giving a current I, right.

And, therefore, I can safely write down that $I_{D1}$, I can write down so I get yes this $I_{D1}$ equals to 1 / 2 $k_N$ W / L ($V_{GS1}$ - $V_T$ ) and $I_{D2}$ is nothing but ($V_{GS2}$ – $V_T$ ), right, so if you subtract $V_{GS1}$ $V_{GS2}$, I get $V_{G1}$ - $V_{G2}$ and that is equals to $V_{ID}$, because source voltage is common to

both of them. So, if I take square root of that I get this into consideration and I with the constraints that $I_{D1}$ plus $I_{D2}$ equals to I. We can safely write down $I_{D1}$ $I_{D2}$ as these quantities that I / 2 - I / $V_{overlap}$ $V_{I\,D}$ by 2 into this quantity and you have got another this quantity here.

(Refer Slide Time: 19:21)




Now you see quite interestingly that $I_D$ both the currents $I_{D\,1}$ I will come to this graph later on but look at both the currents with the currents as such now if you look if we look at both the currents then $I_{D\,2}$ and $I_{D\,1}$ both are basically a function of $V_{ID}^2$ right and that makes my gain which gain will be obviously $V_{I\,D}$ by sorry $V_{I\,D}$ which is the sorry $V_{OD}$ / $V_{ID}$ $V_{OD}$ is the output difference voltage divided by input difference voltage.

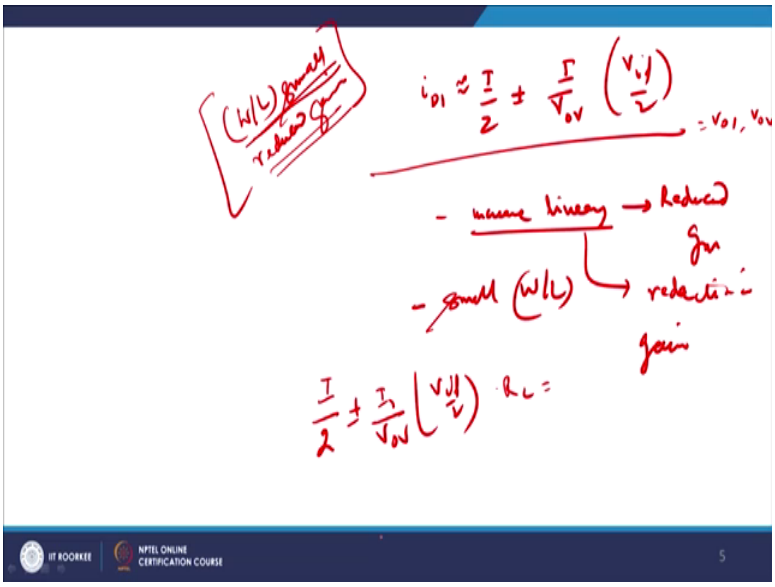
But you see the currents $I_{D\,1}$ and $I_{D\,2}$ are by itself nonlinear function of $V_{ID}$ because $V_{ID}^2$ is here right, so it will be a parabolic term which will be there and therefore, $I_{D\,1}$ and $I_{D\,2}$ will not be a linear function of $V_{ID}$ but will have a nonlinear term because it's square tower a parabolic term available here.

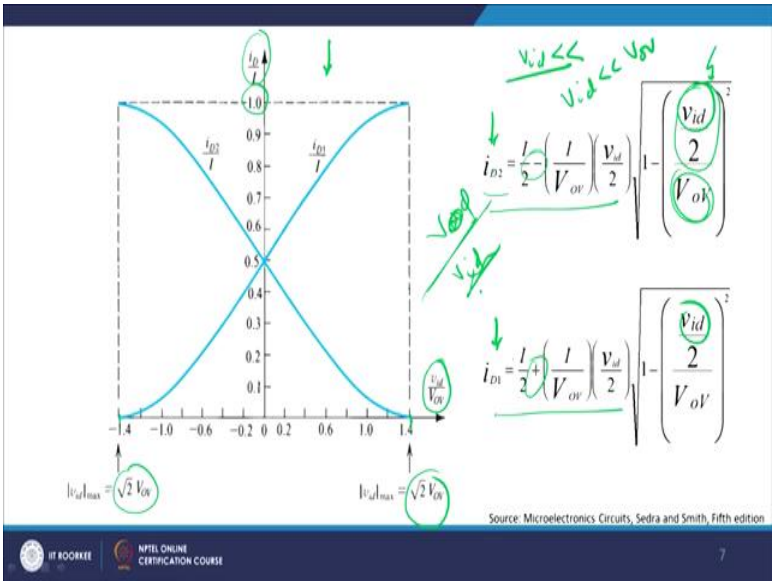
So with this knowledge if you just look at this graph which you see in front of you so you see I have plotted exactly like the previous turn we have plotted $I_D$ / I on the y axis and we have plotted $V_{ID}$ / $V_{overlap}$ on the x axis right this $V_{ID}$ by $V_{overlap}$ and that is $I_D$ / I and on the y axis and of course as I discussed with you $I_d$ even maximum value of $I_D$ will be equals to I and therefore $I_D$ by I will always be equals to 1 and that's the reason a maximum I am going is plus 1, right.

And as I discussed with you that that my $V_{IDmax}$ can go up to my maximum root (2) $V_{overlap}$ and minimum minus root(2) $V_{overlap}$ so I am restricting myself to this this 1.4 here and 1.4 here just a numerical value, do not worry about it, that is a numerical value but this ensures to me that my $V_{idmax}$ which is the difference voltage difference right, the voltage between the two difference is exactly this root (2 ) $V_{overlap}$ and therefore, you get the negative side due to $V_{overlap}$ out so at the negative side.

Now, so the whole idea is therefore to make this current independent of $V_{ID}^2$ term because square term introduces large amount of non-linearity right and you do not want to do that and that is the reason you somehow ever have to remove this $V_{ID}$ we will see how to do that later on. But with this knowledge you have gained, with this knowledge which you have just seen here.

(Refer Slide Time: 21:52)





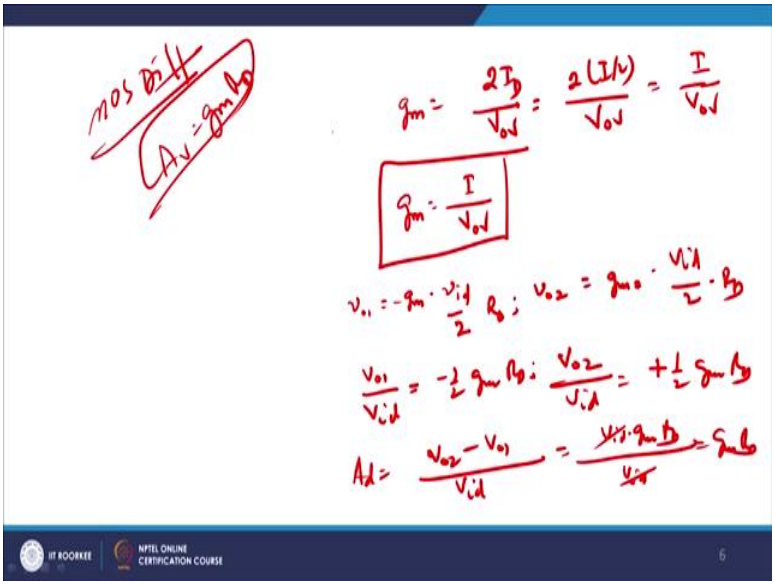

Let me write down for you certain statements and certainly issues that that I could write down that $I_{D\,1}$ therefore can be written as I / 2 +/- I / $V_{overlap}$ into $V_{ID}$ / 2, $V_{ID}$ / 2. Now, how it is $V_{ID}$ by two, provided what I do is, I provided I make, if I make my $V_{ID}$ much smaller as compared to other values then this square term will be very small as compared to $V_{overlap}$, so if I make my $V_{ID}$ much smaller as compared to $V_{overlap}$ then this quantity will be very small and therefore 1 upon this will be typically very large and therefore this will be approximately goes to 0 and what I get will be $I_{D\,2}$ will be I / 2 - 1- I / Voverlap into $V_{ID}$/2 right, this is what we get, if I do $I_{D\,1}$ $I_{D\,1}$ will be therefore this was minus this will be plus, fine so I get $I_{D\,1}$ and $I_{D\,2}$ values with us here.

Now let me see what happens if we increase the linearity of this region right that is my job because as a designer I want that my linearity should be of larger range so that for a larger input voltage range my gain is almost independent of the input voltage right that is the major functionality of this this equation. So what we intend to do therefore is we take two or three important steps here to understand the basic features here and the first thing is that we if you want to increase linearity right you end up having little reduced transconductance.

And second thing is that you require a small W / L right so increase in linearity comes from a reduction in gain right, so we need to find out methodologies by which we can actually enhance the gain even with this this stage differential stage amplifier okay. So we required therefore, as I discussed with you since this this has come because $V_{I\,D}$ square is very small and therefore that can be neglected with respect to 1 and therefore I will be left with $I_{D\,1}$ equals to I / 2 +/- I / $V_{overlap}$ * $V_i$ / 2 as the value of your $V_{o1}$ right.

Now if you see $V_{o1}$ + $V_{o2}$ because $V_{o2}$ will be just minus of that, so when you want to add those 2 so I get I / 2 right plus minus (I / $V_{overlap}$) * ($V_{ID}$ / 2) right, this is the $I_D$ value which you get, this if you multiply with $R_L$ load resistance you get the overall voltage in the output side, so this will be $V_o$ right. If I want to as I discussed with you if I want to improve the linearity of the system make your W / L small cost you pay is reduced in reduced gain right and not only it is give me will have other problems as well but W / L small prime facie will require a reduced gain right so even for a differential amplifier.

(Refer Slide Time: 25:13)




Let me therefore explain to you how a GM looks like, GM is defined as the transconductance is defined as to $I_D$ by V $_{overlap}$. So, it will be to 2(I/2) / $V_{overlap}$ right, so this is nothing but to 2 2 will cancels I / $V_{overlap}$ right, so I get GM transconductance to be equals to I / $V_{overlap}$. Now, therefore I can write down $V_{01}$ to be equal to $g_m$ times minus $g_m$ times $V_{I\,D}$ by 2 into $R_d$ and $V_{01}$ equals to GM2 * ($V_{ID}$ / 2) * $R_d$ but since $g_m$ 1 equals to GM equals to GM 2 can just simply replace by this formulation, so therefore I get $V_{01}$ / $V_{ID2}$ equals to half GM times Rd and $V_{o2}$ / $V_{ID}$ is equals to plus 1/2 $g_m$ $I_D$.

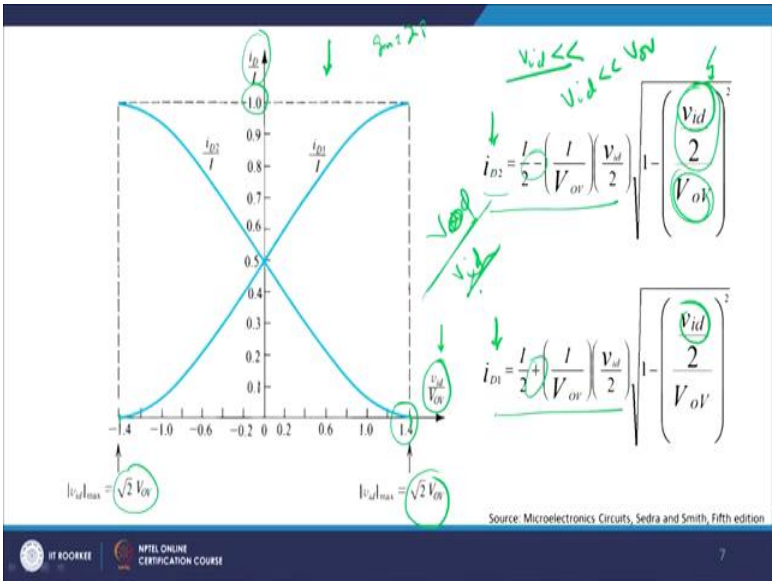
Therefore, if I want to find the differential gain will be $V_{O2}$ - $V_{O1}$ / $V_{ID}$. Now $V_{O2}$ , $V_{out\ 1}$ will be nothing but you will have $V_{ID}$ multiplied by $g_m$ $R_d$ this if you divide by $V_{ID}$, this cancels off and I get $g_m$ times $R_d$ which we have already derived earlier also that means for a MOS transistor, for a MOS based, for a MOS based differential pair where $g_m$ is the acting load I can safely write down my $A_V$ gain to be equal to $g_m$ times $r_d$ fine and this is what you get $g_m$ times $R_d$ is the best value of a voltage which you get.

Now if you if you therefore have this $g_m$ $R_d$ available with you or has got a $g_m$ $R_d$ which is digitally smaller therefore if you want to improve the gain you need to make your gm's larger right. But if you want to make a $g_m$ larger as I discussed with you the previous term, even this term also you end up having a larger non-linearity also coming into picture right. So you have to be very careful while designing operational amplifiers or for the differential amplifiers to a larger extent.

With this knowledge let me come to the next section of our talk and that is basically finding out the CMRR value and to skate the CMRR value maybe let me switch it over and let me explain to you here that as discussed yesterday as well this left hand side is basically $I_D$ / I, the current normalized with respect to the overall current and this is difference current divided by $V_{overlap}$.

As you can see when my $V_{ID}$ is very-very high almost near to root (2) $V_{overlap}$ the maximum value I get all the current, all the charge carriers are flowing through $I_{D\,1}$ through Q 1 and Q 2 is giving you 0 current whereas when your input voltage is minus times root (2) $V_{overlap}$ then you all the current is flowing through Q 2 and you are not getting any current to Q 1 right. So this is the extreme value anything larger than that no problem, but then you will explain extend larger non-linearity.

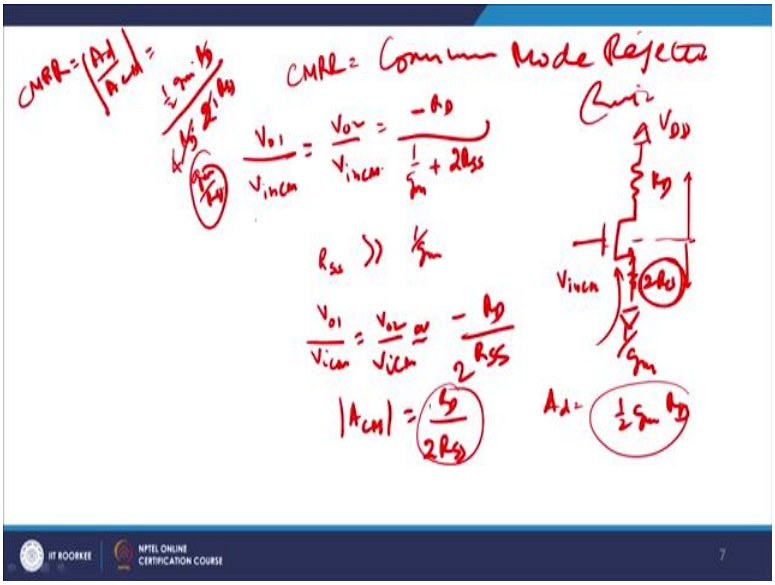
(Refer Slide Time: 28:35)




Similarly if you look at the trans conductance profile which is basically $g_m$ which is $\partial (I_D)/ \partial (V_G)$ then somewhere here if you see around mid-point you will see the maximum value of

$g_m$ is available here which means that if you bias your device somewhere in the middle and try to give your $V_{ID}$ just this difference between these two right, I would expect to see a very large gain anything larger than that no problem, but it will to go into this nonlinear region and in your nonlinear gain with you and therefore, that is not a very good idea to stop with.

(Refer Slide Time: 29:11)



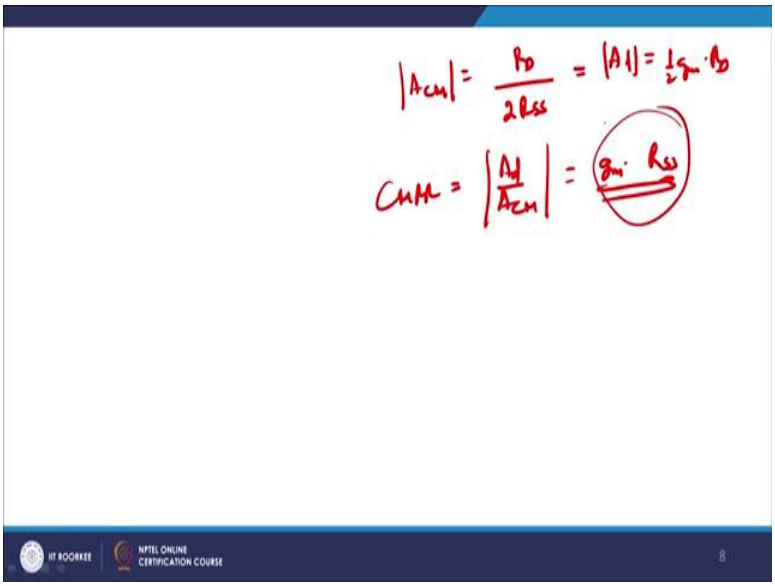

With this knowledge of it this idea let me just also explain to you an important term which is basically your CMRR also known as Common Mode Rejection Ration right. If you find out Common Mode Rejection Ratio, if you look at that graph it is basically $R_d$ and then you have got right so this is your this and this is a $V_{incm}$, this is your $R_d$ and let us suppose 2 $R_{SS}$ and this is my $V_d$ fine then $V_{01}$ output 1 / $V_{incm}$ is equals to $V_{02}$ by $V_{input\_cm}$ right equals to minus $R_d$ / (1/ $g_m$ + 2 $R_{SS}$).

See this is one of the techniques which people used in later days that but just by inspection only you can tell me the sort of a volume sort of a gain of a circuitry. So how do you do take the active device right go from numerated from this active device to up and from activity device to down then you will see they divide the total impedance seen from active device to $V_{DD}$ you divide that by total impedance seen from the device to the ground, and that will be your overall gain with a negative sign because you will have a 180 degree phase shift so with the minus $R_D$ as you move to the top, if you go below you have $2R_{ss}$ as your load plus 1 by $g_m$ remember 1 / $g_m$, 1 / $g_m$ is very important term.

Which means that looking from the source end of my active device, MOS device, the resistance offered is 1 / GM right, and since RS, twice $R_{SS}$ is a series to that we automatically get this much amount of this was amount of this your $V_{01}$ / $V_{incm}$. Now $R_{SS}$ is generally greater than 1 / K small $g_m$ and therefore $V_{01}$ / $V_{icm}$ equals to $V_{02}$ / $V_I$ cm right is approximately equal to $R_d$ / $R_{SS}$ or 2 $R_{SS}$ right so which tells me that mod ( $A_{cm}$) is nothing but $R_d$ / 2 * $R_{SS}$.

Why, because you still have not inserted any differential signal you are only inserting $V_{incm}$ and therefore the overall gain is $R_d$ / 2* $R_{SS}$ right $R_d$ / 2 $R_{SS}$ with a negative sign and mod of that will a positive value. Now your differential gain $A_D$ will be equals to 1 / 2 $g_m$ * $R_d$ this we have already seen earlier and therefore your CMRR can be written as $A_D$ / $A_C$, $A_{cm}$ and $A_{cm}$ is how much, $A_{cm}$ is $R_d$ / 2 $R_{SS}$, $A_D$ is this much so I just need to solve the value is $A_D$ is 1 / 2 $g_m$ * $R_d$ this divided by $R_d$ times 2 $R_{SS}$. So this took so I have got 4 here, this $R_D$ gets cancelled out and I get $g_m$ by $R_{SS}$ fine, and that is the differential gain which you see here.

(Refer Slide Time: 32:24)




So if you want to find out CMRR at which you let me again give you an idea that $A_{CM}$ will be equal to $R_d$ / 2 $R_{SS}$ which is equal to and your mod ($A_D$) will be equals to 1/2 $g_m$*Rd so therefore CMRR is given as $A_D$ / $A_{CM}$ with the mod sign right and this will come out to be approximately equal to $g_m$ * $R_{SS}$. So you see unlike in the previous case it was $g_m$ * $R_d$ when you want to find out CMRR it is $g_m$* $R_{SS}$. So gain is $g_m$ * Rd and CMRR is $g_m$ * $R_S$.

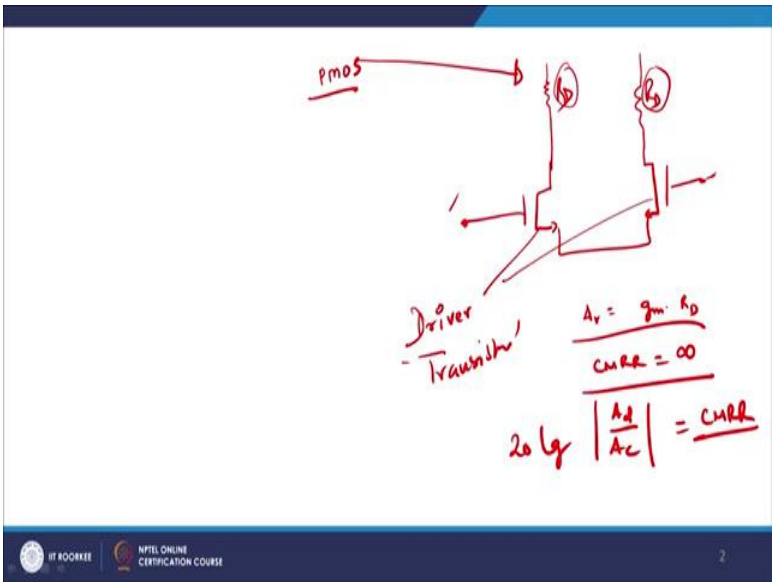
So, therefore if you want to improve your CMRR the best option available to you increase the value of $R_{SS}$. So make it to the order of Giga ohms or something very large value as a result it will never be able to get the output swings but the price then you have to pay is that there will be large power dissipation across these resistors, right and voltage swings and the leg room will also be restricted (33:21) directly because there will be some voltage drop across $R_{SS}$ as such.

If you therefore so if you look very carefully I get $g_m$ * $R_{SS}$ so if you want to improve it a CMRR make it higher make your $R_{SS}$ large and $g_m$ large and this takes care of approximately our understanding of CMRR in the basic principle. In our next discussion, our next slide we will be actually looking into the various other aspects of the MOS device based differential amplifier and its applications right. Thank you very much.

**Microelectronics: Devices to circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−36**
**MOS Differential Amplifier-II**

Hello everybody and welcome to the NPTEL online certification course on Microelectronics Devices to Circuits. We start from where we left in the previous module. We had initially done the work on MOS differential amplifier and we saw that the overall gain of an amplifier depends on the value of your transconductance of the driver transistor and the drain resistance.

(Refer Slide Time: 00:52)



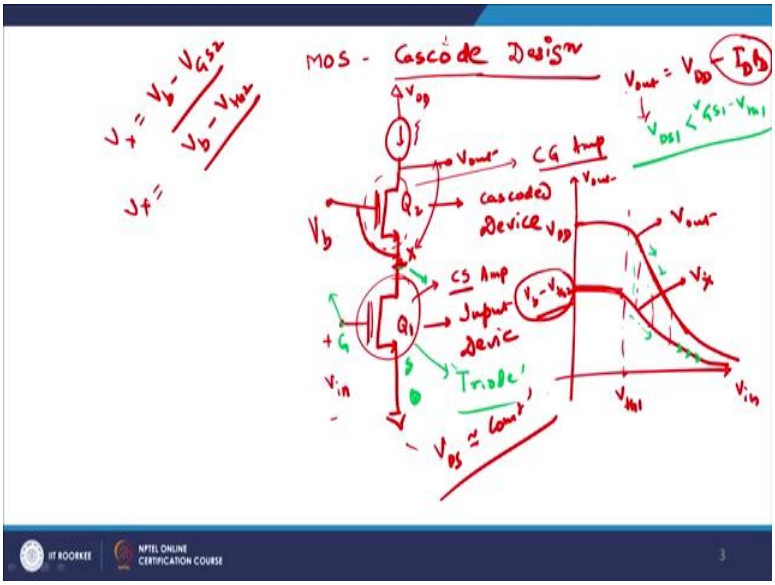

So if we, if we discuss the previous term what we saw was that if you, wherever you are giving the input, please understand, wherever you are giving the input voltages this is the input voltage right this one, this one, these transistors are known as driver transistors, right. Further if you remember we had $R_D$ here which is the drain resistance right $R_D$, this is the $R_D$ value.

But in reality when you do it on IC or on a chip when you want to fabricate it this resistance is will be performed by MOS devices. Typically if we use a PMOS device here you have a larger resistance being offered at this particular point. So we will see as we move along a lot of how it works out, but typically you have learnt 2 things as the gain is given as $g_m$ times $R_D$

by CMRR is infinitely high because of the fact that you are $A_d$ it is basically defined as the $A_d$ by $A_c$ right.

If you take in db then it will be 20 log of $A_d$ by $A_c$ and this is equals to your CMRR, right and this is typically very high quantity which you get, approximately 10 to the power 5 or 6, ideally this should be infinitely large. What we will do today is we will actually look into MOS differential amplifier part 2 where we will be actually replacing those $R_D$ the drain resistances by the typical PMOS.

(Refer Slide Time: 02:20)




Before we go there let me just do one more amplifier design. That is known as MOS CASCODE design, please understand it is not cascade it is CASCODE, right, so you have a O here, CASCODE design, right, you have a MOS CASCODE design. So, let me draw for you how, this is but please understand this is we are slightly moving away from differential amplifiers. Differential amplifiers, you are feeding signals in a differential fashion to both the inputs of the gate.

Here, it is single stage amplifier it is not a differential stage, so, but, we will see what the advantage of it is as far as designing is concerned. So we apply at voltage source or current source here and then there is (03:03) $V_{DD}$, right and then, what we do is, this is quite interesting, that this is Q1 this is Q2 and this is plus minus $V_{in}$ we apply the voltage here $V_{in}$ and we apply a fixed bias here $V_b$, this Q2 is known as the CASCODED device and this is

your input device, you are feeding the input to this and you are taking the output from here, right.

Now, you see this is if you look very carefully this Q1 is basically a common source a design, so this is a Common Source (CS) amplifier, whereas this one if you look very carefully is basically a Common Gate (CG) amplifier, so when common source amplifier and common gate amplifier are stuck together or they are made together, then we defined that to be as a CASCODED amplifier, right.

So what is a CASCODED amplifier you have a common source amplifier you have a common gate amplifier and when they are, when they are, when they are attached like this such a manner that the source of one is connected to the drain of the common source and then we define that as to be a CASCODED, CASCODED case. Now, if I, if I try to plot the $V_{out}$ versus the $V_{in}$ to get the transfer characteristics curve here. Right let us suppose this point is x, right, then if I this is $V_{out}$, then I get something like this.

This is your $V_{th1}$, this is $V_{dd}$, right, and this is $V_b$ minus I will explain to you, $V_{th2}$. If you see as my $V_{in}$ happens to be very low, right, when my $V_{in}$ is very, very low I can safely say that my Q1 is cut off, right. So, I am just finding out the voltage at this particular point. Then if this is cut off, which is in the off state, if this is the $V_{in}$ is very, very low in the off state then you automatically get what that, this is for, this is for, sorry this is for, this is for $V_{out}$ here and this is $V_x$, so $V_x$ is plotted in this manner and $V_{out}$.

So let us look at $V_{out}$ first, when Vin is very very low, right, as I discussed with you, you can safely assume that Q1 will be cut off and therefore even whatever be the case $V_{out}$ will be lashed to $V_{dd}$, because you do not have any current flowing through the system and therefore $V_{out}$ will be latched to $V_{dd}$. As you start to increase your $V_{in}$, as you start to increase your $V_{in}$ you draw current from Q1, because this is the threshold voltage of the device.

Till $V_{th1}$ the output will be latched to $V_b$ minus $V_{th2}$ what is $V_b$, $V_b$ is this much, $V_{th2}$ is the threshold voltage of this device. So remember your $V_{GS}$ minus $V_{th}$ was the overdrive, right. So, so this if you look very carefully, this potential here will be nothing but $V_b$ minus $V_{th2}$, right. So this is your, this is your $V_{GS2}$, right. So if you look very carefully $V_x$ will be nothing but $V_b$ minus $V_{GS2}$.

But, for the device to be on $V_{GS2}$ will be typically equals to $V_{th}$, so $V_b$ minus $V_{th2}$ will be the value of $V_x$, and that is what you are getting here. Till a point, till what point, till the point when the input voltage is just crosses the threshold voltage of the device, which device the common source device which is the Q1 value. As it just crosses the device which is on and both the currents, and the currents starts to flow through Q1 and Q2 because they are in series. And as it starts to flow the voltage at output starts to fall because $V_{out}$ will be referred to as $V_{DD}$ minus IDRD, say it is the resistance is also here available with you.

So when $I_D$, so initially what was happening $I_D$ was 0, so you had what this was 0 and therefore $V_{out}$ was equals to $V_{DD}$. Now, when the $I_D$ starts to flow the $V_{out}$ starts to fall down and that is what is falling down here, right. Similarly, your $V_x$ will also since your $V_{out}$ is falling, your $V_x$ has to also fall, so that $V_{DS}$ between this and this of Q2, $V_{DS}$ of Q2 is almost constant. And therefore you see it is almost the same drop is there between the two initially, initially when the $V_{in}$ is relatively just larger then $V_{th1}$, right.

I will just explain to you once again what I just not talked about it, what I talked about it is initially when your $V_{in}$ is very low less than threshold voltage device, Q1 is in the cut off mode, when Q1 is cut off mode no current is flowing through this CASCODE structure as a result what will happen is your $I_D$ will be equals to 0, therefore your $V_{out}$ will be latched out to $V_{DD}$, standard case and your $V_x$ will be equals to $V_b$ minus $V_{th2}$. And we explained why it should be like that, it should be $V_b$ minus $V_{GS2}$, but $V_{GS2}$ equals to $V_{th2}$ because that is the on state of the device and therefore that is the overall voltage which you will get here.

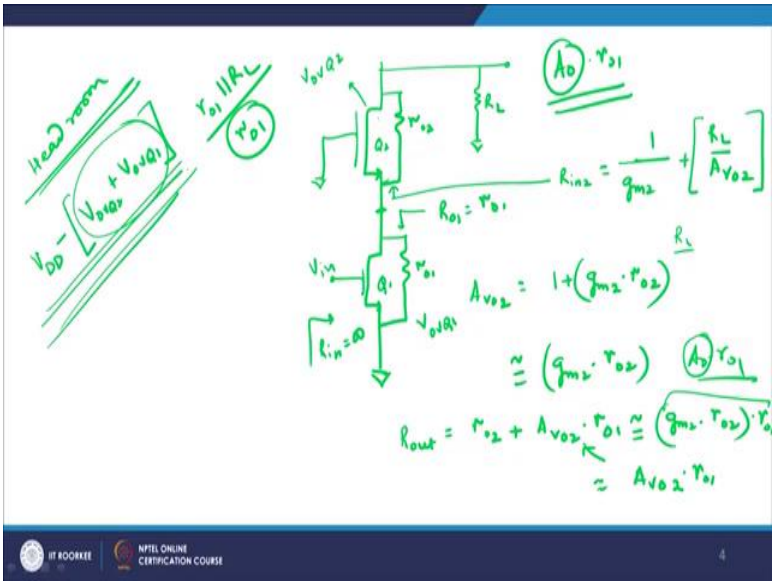
Now, as the input voltage crosses threshold voltage of the first device, the device gets on current starts to flow and when the current starts to flow through Q1 it the same current flows through Q2, and therefore the $V_{out}$ starts to fall down, right, and therefore the $V_{out}$ starts to fall down, as I discussed with you just now, the $V_{out}$ starts to fall down, so $V_{out}$ will fall down as you can see.

But as the $V_{out}$ falls down you, you have to ensure that Q2 is in saturation and therefore $V_x$ also starts to fall down and the difference almost remains the same till few values of $Q_h$. As $V_{in}$ further goes on increasing, Vin further goes on increasing this value goes on still further increasing, right, as it goes on increasing, this is, this is basically a gate and source, right. And this is always grounded.

So, $V_{GS}$ is becoming very, very large which primarily means its over drive is increasing, current is increasing, as a current increases then it forces this voltage to further go down, so that is the reason it is going down and down, right. Ensure it at one point of time it might also happen that $V_{DS}$ of 1 might be even smaller then $V_{GS}$ of 1 minus $V_{th}$ of 1 and forcing the Q1 to enter into triode region, fine.

So, so what I wanted to tell you from this whole, whole, whole idea is that by doing so there is a problem here that common source if you remember was very good amplifier and common gate was a very good a impedance matching purposes. So CASCODE structure helps you to do both of them together that it helps you to match the impedances as well as gives you a improved gain with the with much better gain available to you.

(Refer Slide Time: 10:30)




So if I draw the impedance diagram here or if we draw the small signal diagram here and external load is $R_L$ and let us suppose Q2 is the input current and Q1 is the driver CASCODED current, then $R_{01}$ and $R_{02}$ are basically the, the resistances offered by those devices, if I assume this is Q2, if I assume this is Q1 and then I am giving a voltage here $V_{in}$. So if you look from this side, Rin will obviously be equal to infinity, fine. And, and if you from therefore if you look from this side affectively value of R01 will be equals to $R_{01}$. So from this point if you want to look inside affective resistance offered will be equals to $R_{01}$. I think I need not to explain it to you why is it like that.

Similarly, if you look from this side, right, and that is quite interesting then $R_{in2}$ resistance, input resistance of this one will be from source side, is 1 by $g_{m2}$, right plus $R_L$ by $A_{V02}$ I will explain to you what do I mean by that. If you remember the whole discussion was that even by inspection also you can predict the value of input impedance, how did you do that just look from the source side and from the source side if a the input impedance will be 1 by transconductance of the device, so that is what we have written here 1 by $g_{m2}$ and if you why do you add this $R_L$ plus $A_{V0}$ because $A_{V0}$ is basically the voltage gain right.

Now, so if you divide $R_L$ by the effective voltage gain I get the input impedance at particular point because $R_L$ is the external load resistance which you see, which you see here, right. $A_{V02}$, $A_{V02}$ which is the open circuit voltage gain of the second transistor the common gate one is given as the 1 plus $g_{m2}$ multiplied by $r_{02}$, right. Now, of course this could be approximately return as $g_{m2}$ into $r_{02}$, the reason being that a $g_{m2}$ into $r_{02}$ is much larger then 1 and therefore $A_{V0}$ is equal to $g_{m2}$ into $r_{02}$. Why $g_{m2}$ into $r_{02}$ very straightforward $g_m r_d$, so $r_d$ is the resistance offered by the device $g_m$ is the transconductance of the device.

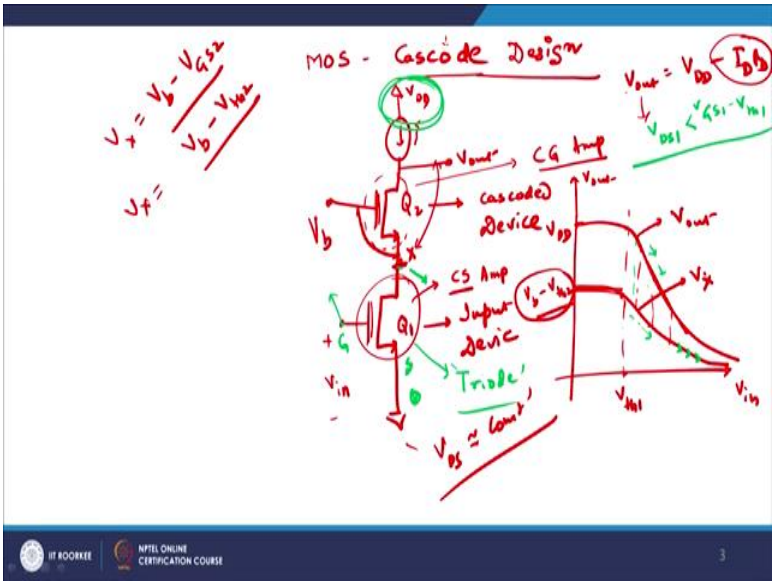
Now, if you find out $R_{out}$ and I am not deriving it, it here lot of scope it will be given as $r_{02}$ plus $A_{V02}$ into $r_{01}$ and if you find out this will come out to be $g_{m2}$ into $r_{02}$, right, into $r_{01}$ and this is nothing but $A_{V02}$ into $r_{01}$ which we just found out here, right, $A_0$ into $r_{01}$. So if you look very carefully on this profile here what does it tell me that the output impedance is raised by a factor of A0, so suppose, suppose you did not have any CASCODE, you did not have any CASCODE device Q2 was not there, then I would have seen $r_{01}$ parallel to $R_L$ now since $R_L$ is very very large as compared to $r_{01}$ the output impedance would have been $r_{01}$.

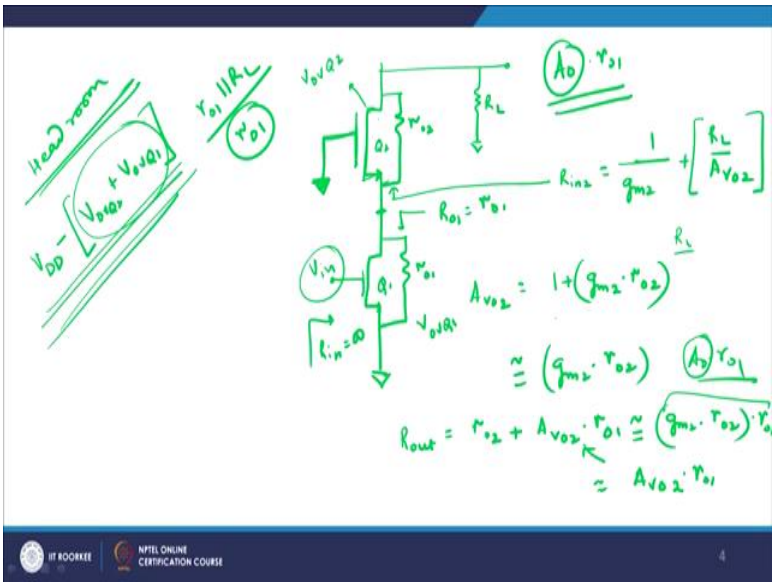
Now, when you place a transistor Q2 CASCODED over Q1 you actually increase the impedance by a factor of $A_0$, where $A_0$ is the gain of my Q2, so you multiply this with $r_{01}$ to get the overall gain $R_{out}$. So CASCODE structures, so CASCODE structures improves the gain by a factor of $A_0$ into $r_{01}$ and, and gives you a very large value of your output gain, right. So this is a, a brief introduction or a structure of, of CASCODED structures, right, and we will revisit this maybe at a later stage, when we come back. Two things to take away from this discussion on CASCODED important for impedance matching as well as for high gain.

The cost we pay for it is that now you have 2 transistors in CASCODE therefore if you look at this very carefully this will be $V_{overlap}$, overlap, sorry $V_{overlap}$ Q2, and this will be $V_{overlap}$ Q1. So the available headroom for me will be, will be affectively equals to if you look very

carefully will be $V_{DD}$ minus right, $V_{overlap}$ Q2 plus $V_{overlap}$ Q1, right. So this is the available headroom to you in the, in the output side. Which means that if you want the gain to be high for the CASCODED structure you end up having a reduced headroom, so this is basically the headroom, headroom associated with this, right.

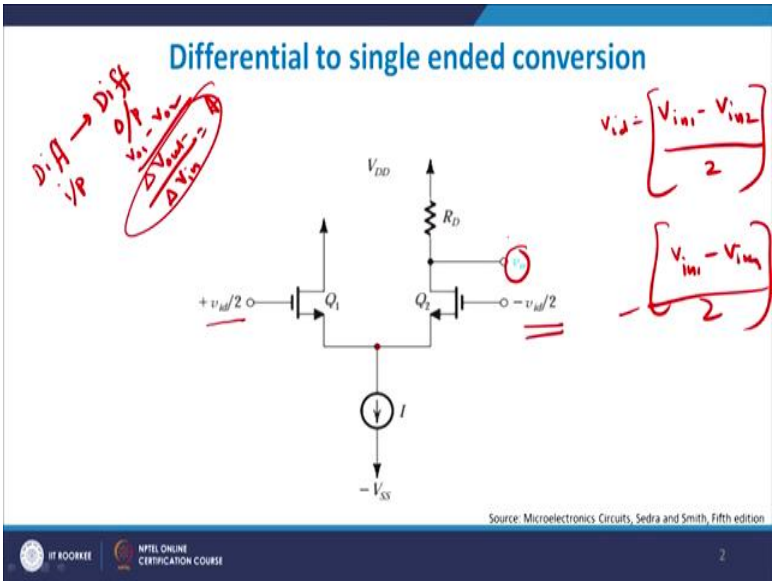
(Refer Slide Time: 15:14)




Which means that, so this is connected to so if you look very carefully, if you look at the, this was connected to the $V_{DD}$ remember, connected to $V_{DD}$, which means that it is $V_{DD}$ minus the sum of the over drives which you see. While looking at this point you please see that the while, while deriving this all quantity you please see that the base of the gate of Q2 has been grounded, right, whereas, I am giving an input signal to Q1, right, please find out why is it like that, why have we grounded Q2 and why we have taken out the signal through Q1 here in this case, right. This, this is to rather extent explains the overall features of CASCODED structures using MOS devices.

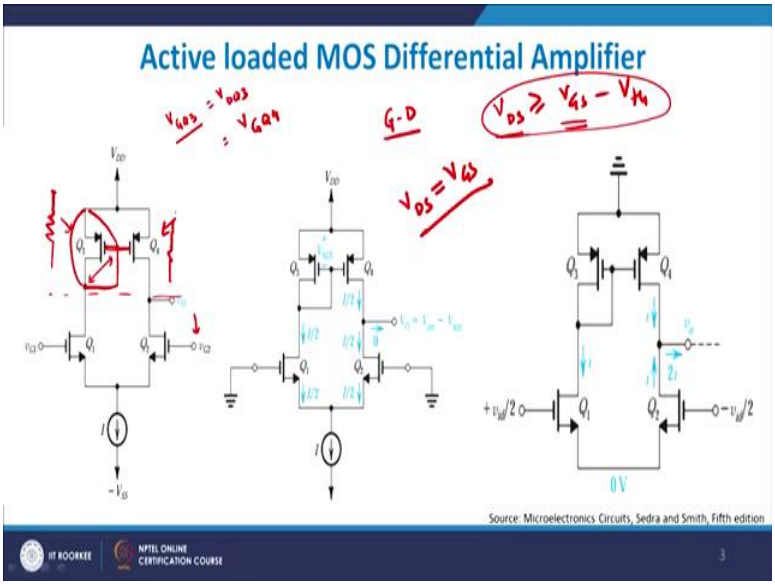
(Refer Slide Time: 15:50)




Let me come to the differential amplifier and let me see how can we change because till now we were what we are looking we were looking in the fact that we had a differential we had a differential 2 differential output. So, I have a differential input and I have a differential output and then what I was doing was I was subtracting $V_{01}$ from $V_{02}$ or vice-versa. And this is this was del $V_{out}$ right and already I had del $V_{in}$ and this was your overall gain which I solved. This is what we are doing till now, we also require to sometimes have a differential to single ended conversion.

So this is the diagram for that, that you apply signals here $v_{id}$ by 2 so the difference divided by 2 plus and then the difference divide by 2 minus we already know what is $v_{id}$. So vid is basically v input 1 ($v_{in1}$) minus v input 2 ($v_{in2}$), let us suppose, so it is basically this by 2 on this side and then you apply $v_{in1}$ minus $v_{in2}$ by 2 with the negative sign on this side perfectly differential signals. As you may, as you mean infinitely high impedance. Then we find put the output impedance at this particular point $V_0$, right, and a straightforward way of saying that conversion of a high from a, from differential to single ended input.

(Refer Slide Time: 17:40)




As I discussed with you in the previous turn, what we were doing was that we were, we need not to replace the $R_D$, the drain resistance by load, actual PMOS load. So what people thought was that let us apply PMOS load in place of $R_D$, the reason was PMOS is effectively having giving you a larger resistance because of lower mobility of holes, for the same dimensions of W by L. For the same aspect ratio the resistance offered PMOS will be 3 to 4 times larger as compared to NMOS of the same dimensions. So, if you look, this is exactly the same this below side is exactly the same if you look at the top, $R_D$ is replaced by Q3 and Q4, right, and these are PMOSs.

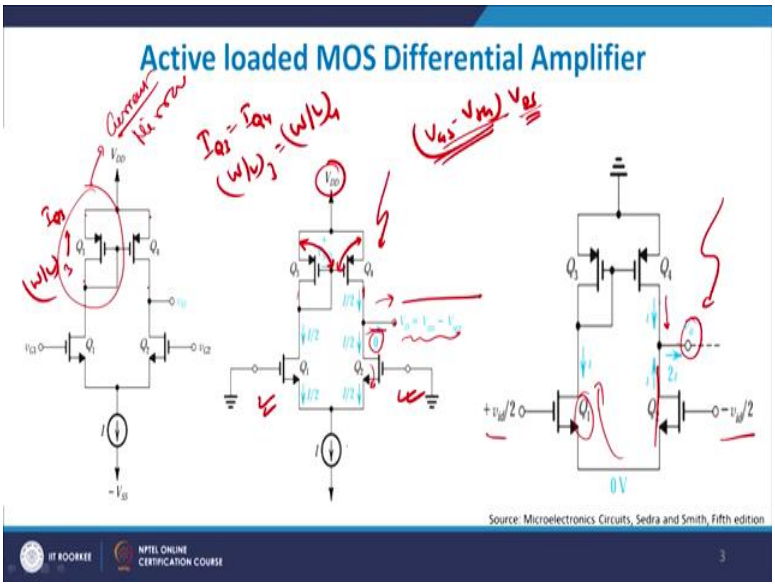
And quite interesting what we have done one thing is that we have connected the gate of Q3 to the drain of Q3. So, now if you say $V_{GQ3}$ right is exactly equals to $V_{DQ3}$, right. So gate of, gate of Q3 and drain of Q3 are exactly the same which is also equals to $V_{gate}$ of Q4 ($V_{GQ4}$), which means that now, gate to source of Q3 and gate to source of Q4 is exactly the same. If I join these 2 then it ensures that gate to source of Q3 and gate to source of Q4 are exactly same which means that, which means that if I assume that Q3 and Q4 are perfectly symmetrical their over drives and resistance will also be equal to each other.

What happens when we join the drain to the gate side of it, right? When you join drain to gate you ensure that the device is basically moving in to a saturation mode and behaving like a resistor, right. If you, if you short gate and drain this is the reason the reason being, remember what was the reason $V_{DS}$ should be greater than equal to $V_{GS}$ minus $V_{th}$, right. Now, $V_{DS}$ is

this much, $V_{DS}$ is this difference and $V_{GS}$ is this difference. Now, if these 2 are equal then $V_{DS}$ is always equal to $V_{GS}$.

$V_{DS}$ drain to source will always be equal to $V_{GS}$ get to source, agree, so drain to source will always be equal to get to source because your gate and drain have been shorted with respect to each other. Which ensures that which, which is quite interesting that which means that this condition can never be applicable which means that my device can never go into saturated region at all, right, but in triode region behaving like a resistor. So this can be replace by a simple resistors. If you look back since this is a symmetrical Q4 will also look like a resistor.

(Refer Slide Time: 19:59)




This action is basically known as current mirroring, this, this is known as current mirroring, because this, I will explain it to you later on maybe current mirror. The reason being that the current flowing through Q3 will be exact if the W by L ratios of Q3 and Q4 are same the current flowing through Q3, suppose this is $I_{Q3}$ will be exactly equals to $I_{Q4}$ provided W by L of Q3 is exactly equals to W by L of Q3 is exactly equals to W by L of Q4.

The same current will flow, why because remember current is depending on $V_{GS}$ minus $V_{th}$ whole square, right, it also depends upon the value of $V_{DS}$, if you, if you find out, in the triode region, if you are trying to find out, it depends on the value of $V_{DS}$ is $V_{GS}$ minus $V_{th}$ multiplied by $V_{DS}$.

So if I am able to ensure over drive to be equal and $V_{DS}$ to be equal and everything else taken care of my current will also equal, right. So if you look very closely this difference between

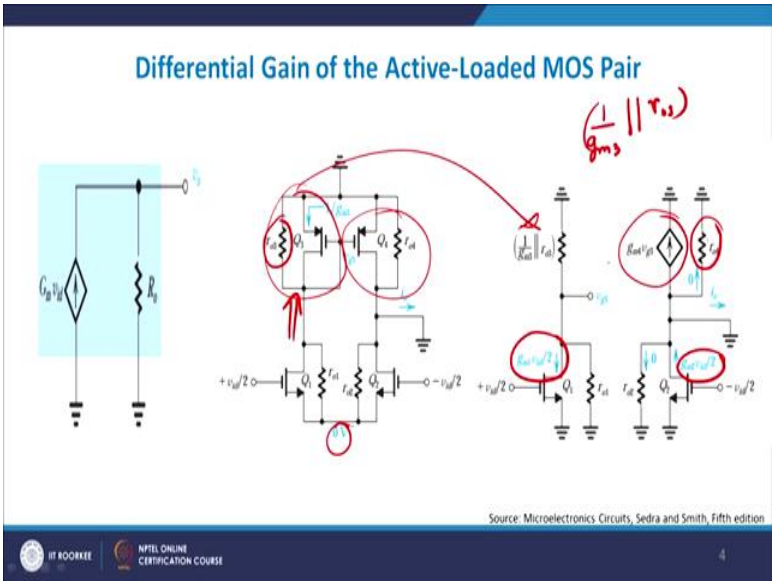
this point and this point will also be equal to the difference between this point and this point. And therefore $V_{DS}$ will be also equal my over drive will be equal and therefore they will behave like exactly symmetrical structures.

With this knowledge let me say we have this structure available here, so I have, so I, so as I discussed with I have a current source I, it divides into I by 2 here, I by 2 here, so this is I by 2 here, since this is I by 2 as I discussed with you, you will obviously have I by 2 flowing here because it is a current mirroring action here. As a result the current flowing through these will approximately equals to 0 because all I by 2 have has to be routed through Q2, so the net current flowing here will be 0, so that is what is written here.

And therefore $V_{out}$ will be equals to $V_{dd}$ minus $V_{GS3}$, why, $V_{dd}$ this voltage minus gate to source, source to gate this to this. So if you subtract this voltage or even if subtract this voltage I get the voltage available at this particular point (())(21:55) Kirchhoff's law. And as a result you will see this. This is for the case when both your Q1 and Q2 are either shorted or having equal dc bias.

Let us come to this point you will have differential voltage given, so I will have I, so this is minus $V_d$, minus $V_d$ by 2 and this is plus $V_d$ by 2 all current will be routed here. So this is the current flow here, so the other current flowing is here, so net current flow was I plus I 2I and therefore $V_{out}$ will be having 2I available to you, right. So this is what you get from the, this active load.

(Refer Slide Time: 22:29)

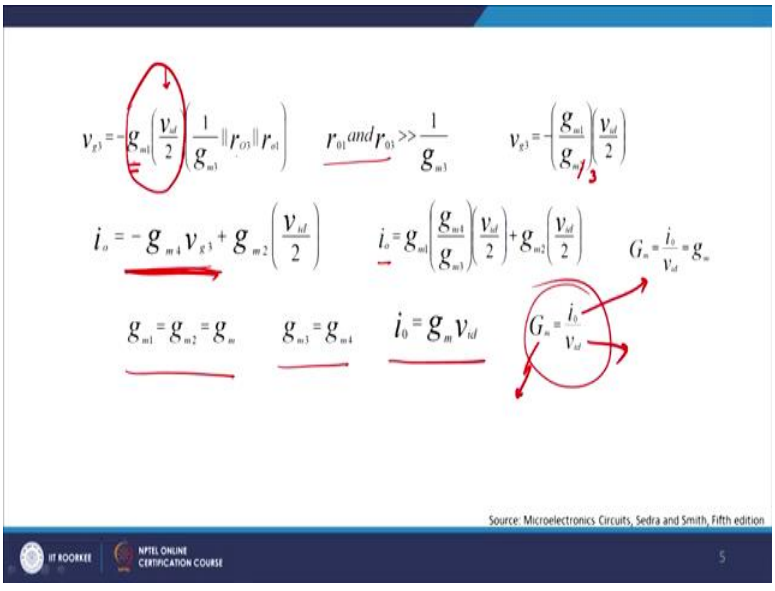


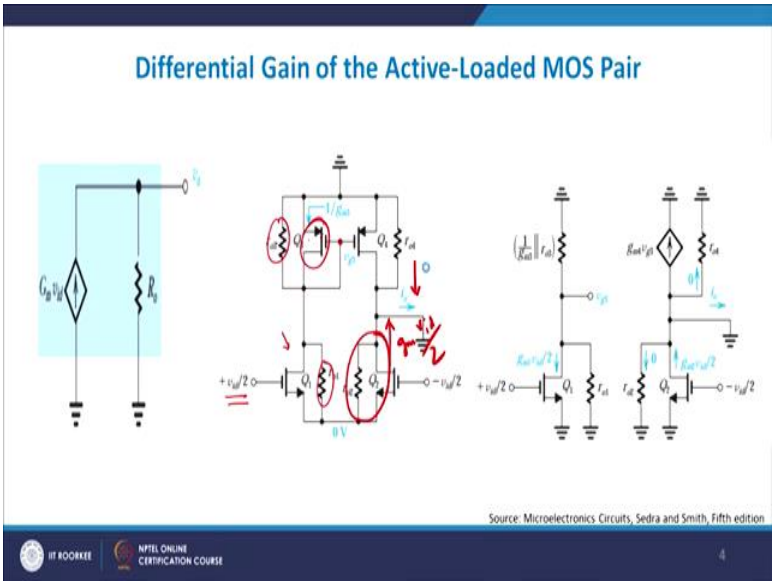
Now, therefore if you look here which is what I have done here is I have broken the Q3 Q4 Q1 & Q2 with the effective transient resistance. Where $r_{01}$ is the drain to source resistance of Q1, $r_{02}$ is the drain to source resistance of Q2, so are and so forth. And I have sort of grounded so 0 voltage is given to the input side to the source side, right. So I will assume that since it is 0 grounded any voltage above them with the threshold voltage of the device give to you gate side will switch on the device, right, switch on the device.

Now this is $r_{03}$ and this is $r_{04}$ and so on and so forth. So, if you look from if you look from this side towards this side the effective resistance offered will be $r_{03}$ is parallel to Q3. And Q3 as I discussed with you if you look from the drain and source side the amount of is given by this, this is the resistance offered. So this resistance is in parallel to $r_{03}$ and that is what I have written here. So this is replaced by this quantity, right. Whereas if you look where carefully at this one this is basically again a Q4 is basically a $V_{CCS}$ a voltage control current source.

And therefore gm4 multiplied by $V_{G3}$ is the value and R4 is therefore the, the resistance offered between the two, between the two. Assuming that this is the perfect current source I am assuming that there will be no current flowing through $r_{04}$ as such, right. So what I see is, if $v_{id}$ by 2 is the voltage given to the left arm and right arm and I can safely write down $g_{m2}$ into $v_{id}$ by 2 as the current flowing here, and gm by $v_{id2}$ opposite direction current will be flowing here, right, and this is what we get.

(Refer Slide Time: 24:16)






Now, let me come to the mathematics of it, so if you look very carefully $V_{G3}$ and this will require some amount of flipping of the power point presentation, $V_{G3}$ is minus gm times $v_{id2}$ into the sole quantity. What is $V_{G3}$, $V_{G3}$ is, let me rub this to make it clear, $V_{G3}$ is nothing but the voltage at the gate of third transistor, right it is nothing but $g_{m1}$ into $v_{id}$ by 2, why it is $g_{m1}$ into $v_{id}$ by 2 see $g_{m1}$ is the transconductance of Q1 into vid2 is the current which the transistor will allow it to flow, so the same current will flow to Q3, so that is the, sorry, so that is the Q3 current here right, I am sorry.

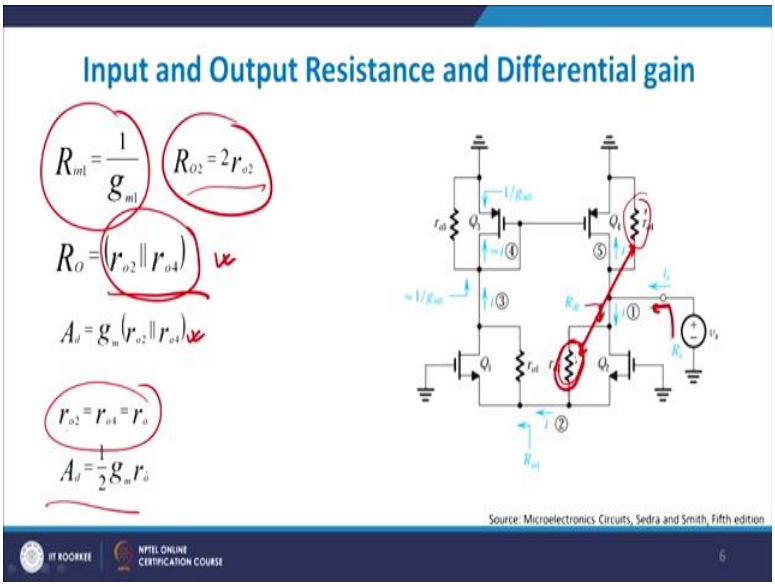
So I get $g_{m1}$ into $v_{id}$ by 2 is the current this if you multiply with the impedance that will give you the voltage, what is the impedance, impedance is nothing but 1 by $g_{m3}$ because of this one and ro3 this one as well as r01 because you see this, this and this these 3 will be in parallel to each other. As well as you will get $g_{m1}$ into $v_{id2}$ into 1 by $g_{m3}$ in parallel to $r_{03}$. Now $r_{01}$ and $r_{03}$ is much, much larger as compared to 1 by $g_{m3}$. So, what I can safely write now gm1 by $g_m$, $g_{m1}$ by much larger then $g_{m3}$, sorry this will be gm3 I think, this will be $g_{m3}$ so I will make a correction.

So $g_{m1}$ by $g_{m3}$ into vid by 2, right, now $i_0$, $i_0$ is the output current right, output current in $g_{m4}$, if you look very closely $g_{m4}$ into $V_{G3}$, why, $V_{G3}$ is the same voltage you are applying on to the fourth transistor Q4 right. So $V_{G3}$ multiplied by $g_{m4}$ is one component of the current and another component of the current is coming from this transistor $g_{m2}$ multiplied by $v_{id}$ by 2. So one current component is coming from here another is this one.

This one is $g_m$ multiplied by $v_{id}$ by 2 and this one is basically yours, sorry, this one is basically yours $g_{m4}$ into $V_{GS}$ by 3 with a negative sign, right. So if you do a small manipulation I get $i_0$ equals to $g_{m1}$ into $g_{m4}$ by $g_{m3}$ into $v_{id}$ by 2 plus $g_{m2}$ $v_{id}$ by 2, as you mean that $g_{m1}$ and $g_{m2}$ is equal to $g_m$ and $g_{m3}$ and $g_{m4}$ are equal I can say $i_0$ is equal to $g_m$ times iv.

And therefore capital $G_m$ is equal to $i_o$ by $V_{id}$, what is $i_o$, it is the output current, $V_{id}$ is the input differential voltage and therefore I use a capital $G_m$, capital $G_m$ is therefore the effective transconductance for the differential pair, fine, so capital $G_m$ is the effective transconductance of the differential (())(27:04), whereas small gm is the transistor level transconductance, fine, so this is the difference between the two.

(Refer Slide Time: 27:17)

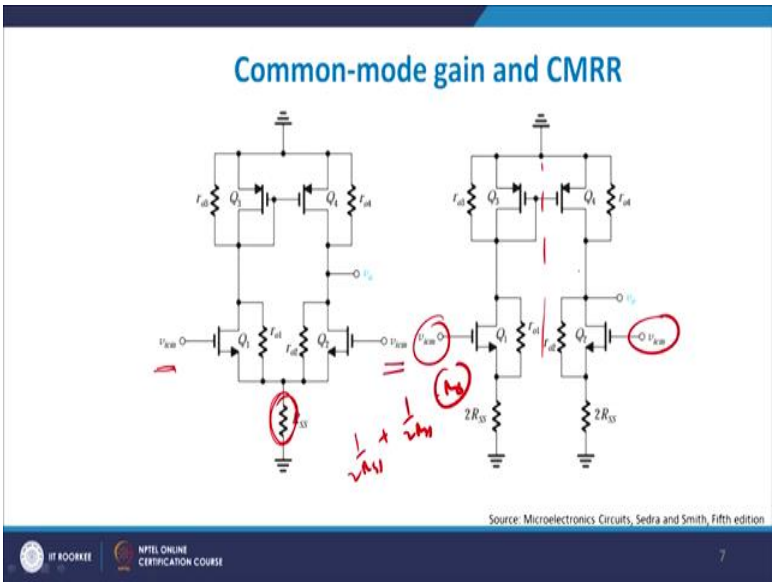



Let us look at the input and output resistance of the differential pair if you look input resistance will be given as 1 by $g_{m1}$ we have already looked into the fact that, that input resistance given by here is $g_{m1}$, $R_{02}$ will be 2 times $r_{02}$ the reason being that if you look from this side, if you look from this side you have $r_{01}$ coming from here, one resistance is coming from here and another one is basically coming from here, so they are equal and therefore assuming to be true.

And therefore, R0 though effective resistance seen at this particular point these two are in parallel, these two are in parallel and therefore I get $r_{02}$ parallel to $r_{03}$, right. We will explain to you this why $R_{02}$ is actually equals to $2r_{o2}$, $r_{02}$ which is looking from this side you will have 2 times $r_{02}$ right, and looking from this side you will have this these 2 in parallel that is what

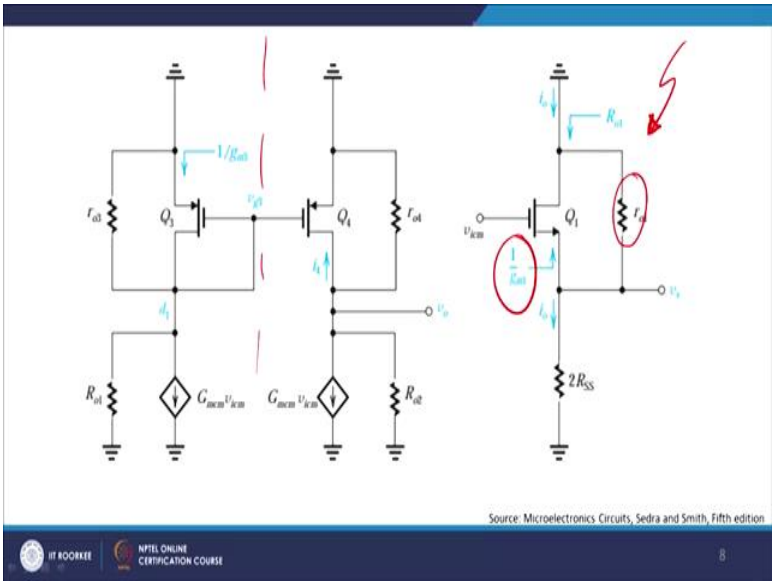
you get here, and therefore the differential gain will be nothing but gm times $r_{02}$ multiplied by r04 assuming $r_{02}$ equals to $r_{04}$ equals to equals to $r_0$. I can safely write down Ad to be equals to 1 by 2 $g_m$ $r_0$, fine, because they are in parallel, write so this is what you get parallel.

(Refer Slide Time: 28:20)



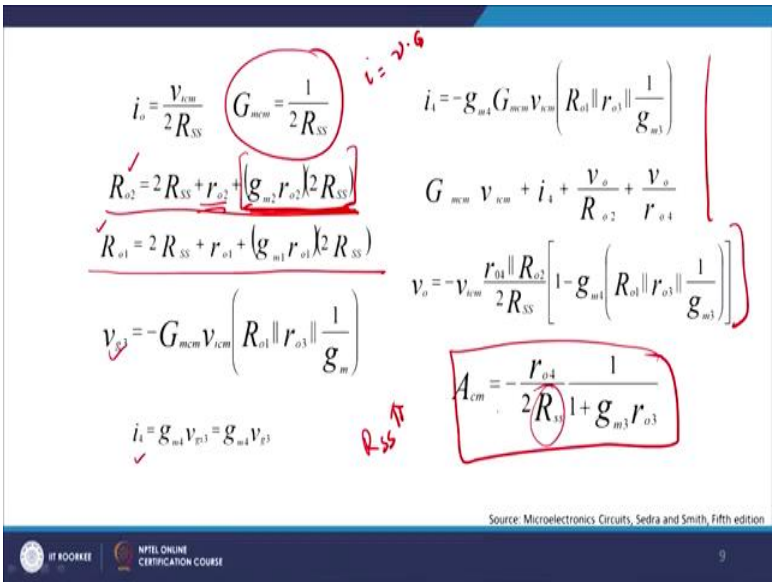

Now, we will look into common mode and differential gain so common mode gain if you remember what we did we inserted $V_{icm}$ as the input mode voltage and I replace my current source by $R_{ss}$, $R_{ss}$ is of the very large value of your resistances being offered here. It is a quite a large value of resistance because output impedance of the, of the current source will be typically very high. And then what we do, we do a half circuit realization and we divided into 2 parts and we give $V_{cm}$ here, $V_{cm}$ here and then this is parallel so 1 by 2 $R_{ss}$ plus 1 by 2 $R_{ss}$ will give you $R_{ss}$ and that is the reason we divided into 2 parts and we just separate it out, right, for calculation purposes.

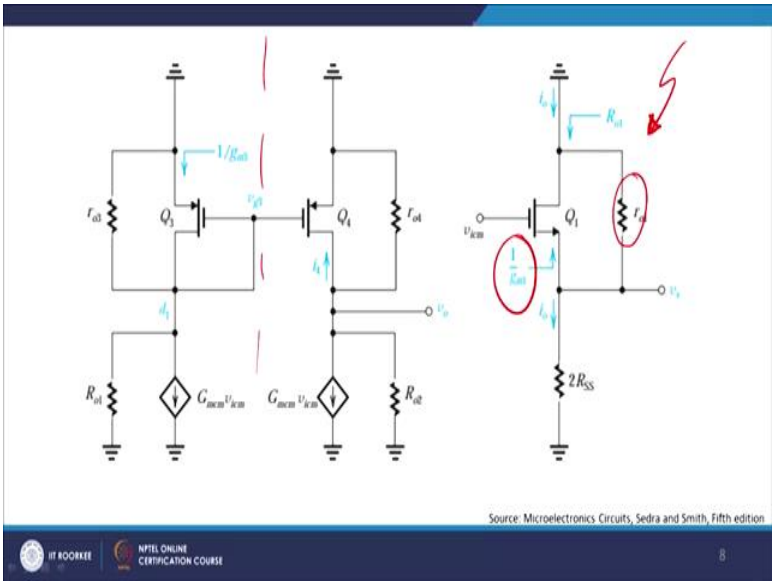
(Refer Slide Time: 28:59)




Similarly, when we look in to the, the overall transconductance features here so, so this basically one of the half circuits. This is one of the basically the half circuits. So if you look at the half circuit it is Q1 with 2 $R_{ss}$ here and $r_{01}$ as the drain to source resistance offered by this thing and therefore looking from the source side is 1 by $G_{m1}$ is the effective resistance offered, offered by the device, right. So what we have done is that we have just broken up in to half circuit and explain it to you this basic idea.

(Refer Slide Time: 29:29)





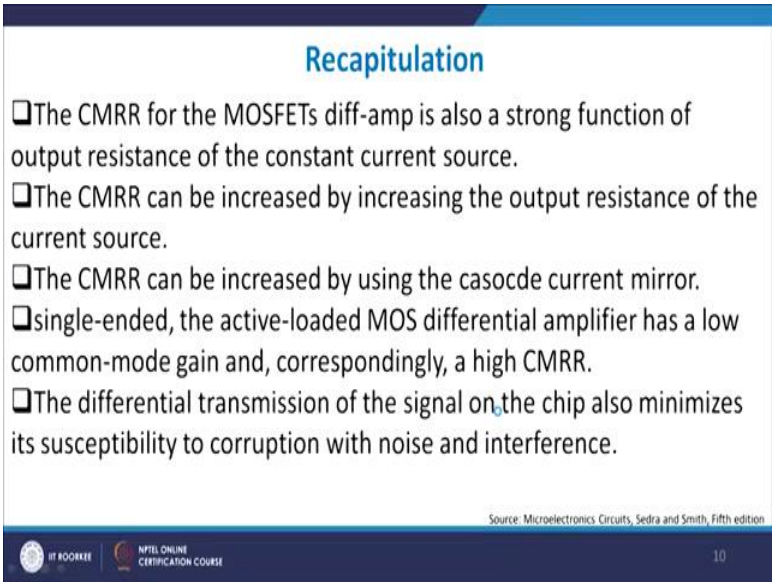

So, let me come to you and explain to you each one of them individually, so i0 will be equals to v, so i0 will be the output current is $V_{icm}$, why it is $V_{icm}$, $V_{icm}$ is, $V_{icm}$ is basically this $V_{icm}$, this divided by 2 at twice Rss because twice $R_{ss}$ is the resistance offered by it, whereas, $G_m$, therefore $G_m$ cm will be 1 by 2 $R_{ss}$, right because current is equals to G into so if you remember current will be equals to, therefore voltage into G, G is the transconductance.

And therefore, I can safely write down $G_{2b}$ equals to 1 by 2 $R_{ss}$. $R_{02}$ is given by this formula, $R_{02}$ is the, $R_{02}$ is basically the resistance at this particular point at Q2, this is $r_{02}$. Looking from this side this is $R_{02}$, right, $R_{02}$ if you look at this is $2R_{ss}$ why $2R_{ss}$ because you will always have $2R_{ss}$ here, right you will have $r_{02}$, the resistance plus $g_{m2}$ $r_{02}$ into $2R_{ss}$, what is this quantity $g_{m2}$ into $r_{02}$ into $2R_{ss}$ into $2R_{ss}$ will give you the overall resistance offered by the device, right.

Similarly, $R_{01}$ by symmetricity you will get this, similarly if you (multi) therefore, $R_{01}$ and $R_{02}$ are known to you, you can predict the value of $i_4$ and $V_{g3}$ and from here, I will get common mode signal value to be equal to $A_{cm}$ to be equals to this much. Please do it yourself these internal I am not doing it in the lecture module, but, please solve these yourselves to get these values and you will get $r_{04}$ by by $2R_{ss}$ by 1.

So if you want your $A_{cm}$ to be really, really small, right, keep your $R_{ss}$ infinitely large and that is the reason why we require a almost an ideal current source here to make my $R_{ss}$ as large as possible, right. That makes my $A_{cm}$ very small and therefore by CMRR will also be very large.

(Refer Slide Time: 31:28)




So, let me recapitulate what we did today in this, in this thing, the CMRR MOSFET is a strong function of output resistance we have seen that, and therefore, we have to ensure that it is a constant source or an ideal current source. How can you increase a CMRR by increasing the $R_0$ value that we have also seen in the previous discussion. If you use a CASCODED current mirror your CMRR will increase because your $A_{cm}$ will decrease and $A_d$ will increase.

We discussed that point today also that the differential voltage gain will be much higher in a CASCODED current mirror design. Then, the single ended active loaded MOS differential amplifier as a low common mode gain, and correspondingly high CMRR and the differential transmission of single on the chip also minimises the susceptible corruption with noise and interference. This is quite interesting the last one which take me that if you do have a chip in which you have got large impedances $R_{out}$ is typically very large, it is actually a very good rejecter of noise also.

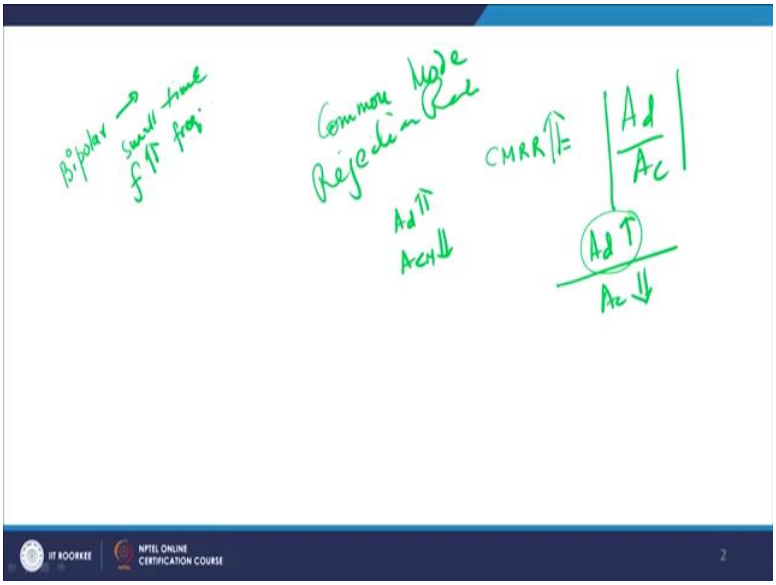
So noise resistance capabilities of differential amplifier is typically very high, we have already discussed this point, but not only because differential operation but because also of high impedance nodes in those devices, right. So with this we finish of this module and from next module we will take up another subject, right. Thank you very much.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−37**
**Small Signal Operations and Differential Amplifiers-I**

Hello everybody and welcome to the next addition of NPTEL online certification course on Microelectronics Devices and Circuits. In our previous modules we have actually seen differential amplifier using bipolar technology and we came to a major important issue that all common signals are being rejected and all differential signals are getting amplified. Common signals primarily mean that noise signals are getting rejected and all your differential signals which are the input signals given will always be amplified.

This results in one important parameter known as Common Mode Rejection Ratio - CMRR which is defined as a ratio of the differential mode gain to that of the common mode gain. So it is basically $A_D$ by $A_C$. So to just give you an idea about what I was talking in our previous discussion is to remind you because we will be starting from where we left in our previous term and we will see how it works out in this case.

(Refer Slide Time: 1:28)



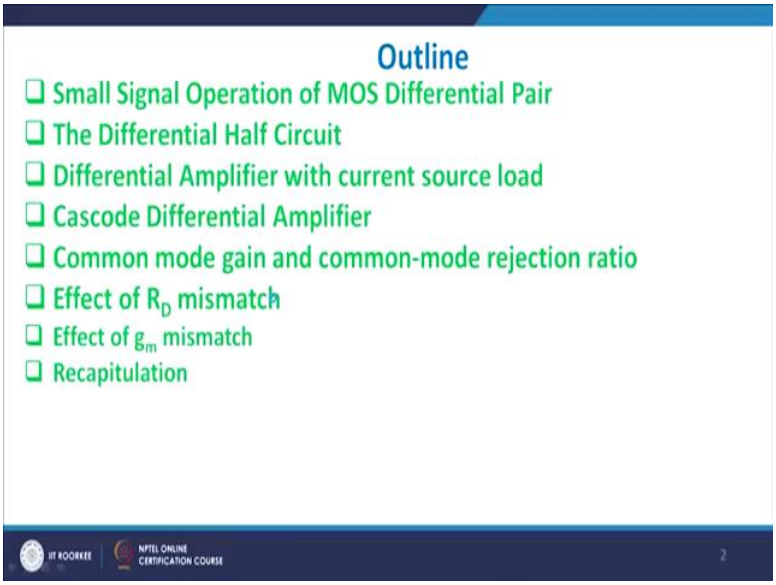

So we had defined another term that is known as CMRR known as Common Mode Rejection Ratio and we saw that this to be equals to $A_d$ by $A_c$, we do not take the sign of it but it is differential gain upon the common mode gain. Now, typically since $A_d$ is large and $A_c$ is very-very small, right? Because the idea is that the common mode signal needs to be

suppressed and therefore amplified less as compared to differential mode design. And that is the reason your CMRR is relatively very high, right?

So, this is one of the design criteria's we generally follow in an amplifier that especially in differential amplifiers whose CMRR should be high. So we understood to things that $A_d$ should be high and $A_c$ the common mode signal CM should be low, right? To do that we first buyers the device in a BJT in the active region and we do not let it go into the saturation region but we go from active to cut off and cut off to active because we want the switching speeds to be higher.

So if we take a bipolar technology, right? Bipolar technologies have typically very small switching times or its frequency of operation is very-very high, right? Its frequency of operation means the frequency by which the switching can takes place is very-very high and is one of the major advantages of bipolar technology. What we will do in today's lecture is we will replicate the same thing which we did in our earlier design or in our earlier bipolar technology and we will try to have it through MOS technology because CMOS is the standard sort of industrial standards.

(Refer Slide Time: 3:19)



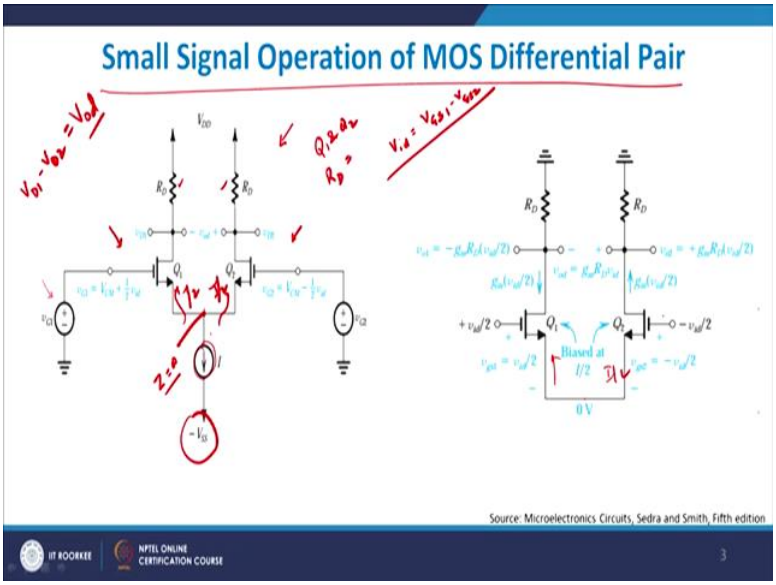

And therefore what we will do is we will just look into the following outline of the talk. We will look into the MOS differential pair, how a MOS differential looks like? What you understand by differential of circuit? We have already seen it but we will see it in details. Then we will put a current source load at the output and see how it works out. Then we will

look at the Cascoded differential amplifier CMRR the meaning of CMRR with respect to CMOS technology.

And then this comes with the fact that till now we were all assuming that the resistances in the left and right arm of a bipolar technology based differential amplifier are exactly symmetrical which means that both the transistors are both the bipolar will have the same value of beta, Alpha, will have also the same value of the drain resistance RD, in such a scenario you are able to reject the common mode noise very easily.

But in reality when you actually fabricate you will always have certain mismatches in the transistors, in terms of its transconductance as well as the drain resistance $R_D$. So that we will be studying in the next 2 slides or the next 2 sections that is $R_D$ mismatch and gm mismatch. So we will look into these mismatches and see how does it affect our overall picture as far as the small single, small signal operation of MOS device is concerned?

(Refer Slide Time: 4:33)




Let me come to the small signal operation of a MOS device and as you can see here what we have done here is that, please remember from our previous discussion that whenever you did have a common mode signal, right? We discussed the fact that a common mode signal will always give you a 0 differential output, right? Another reason was very well known to us because the same current was drawn through both the arms and therefore, the voltage drops at the output were exactly the same, if you subtract 2 equal quantities you always get 0 voltage.

Now, if you look at this left hand side slide, you will see that what we have done is, that we have given a gate voltage onto the gate side of Q1. We are assuming that Q1 and Q2 are 2

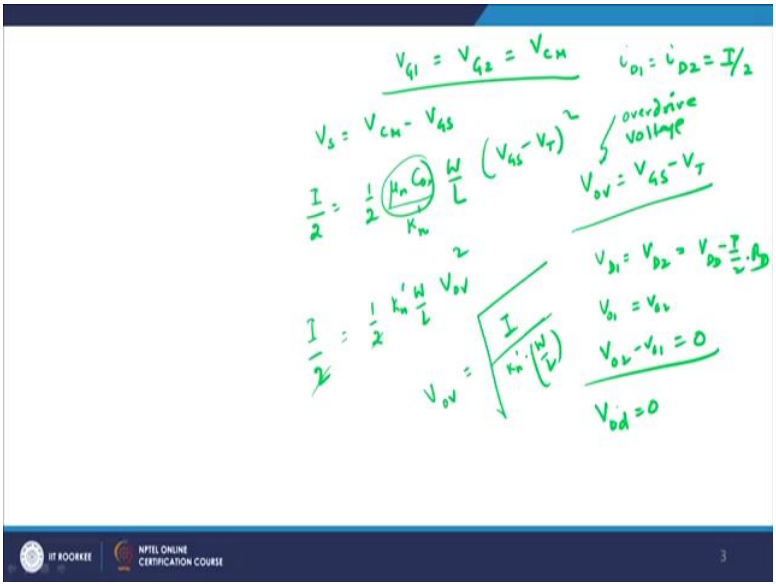
symmetrical transistors and $R_D$ is also the same value which is in the left and right arm under such a scenario if I give a differential input then $V_{id}$ is defined as the difference between the $V_{GS1}$ minus $V_{GS2}$, right?

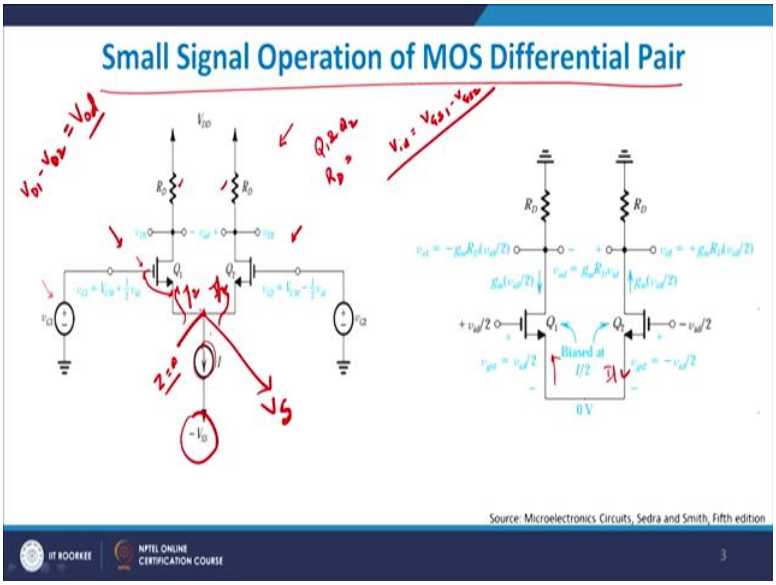
So different signal is given. So what I do is that, so I divided by 2, why? So the difference signal divided by 2 and we give first one half to this part and the negative half to this part, so that they are 180 degree phase shifted with respect to one another and therefore, I would expect to see maximum gain out of it, right? And we try to find out the value of $V_{D1}$ minus $V_{D2}$ or $V_{D2}$ minus $V_{D1}$ and I get what is known as $V_{od}$ output difference.

So, $V_{id}$ means input differential voltage, $V_{od}$ means output differential voltage, right? Assumption again that Q1 and Q2 are perfect transistors and you have a current source which is basically your $i_{ss}$ or the current source here which is given by i. It is connected to minus $V_{SS}$ on the side and therefore this is basically your high impedance note is the Z equals to infinity at this particular point and therefore whatever current is flowing will be equally distributed across this arm and this arm as I by 2 and I by 2.

With this knowledge let me shown here that since this is basically, the current shows it is terminating here and therefore, what I have shown here is, it is not grounded, right? So it is basically a floating note sort of and therefore the overall current will always remain the same in these 2 arms. So, I by 2 I by 2 provided the amount of voltage given to Q1 and Q2 are exactly the same, right? Then only we will see that the overdrives are equal and therefore the same current will be drawn in a respect.

(Refer Slide Time: 7:40)






So with this knowledge let me show you how does it work out in terms of working principles. So I can safely say that $V_{G1}$ is equals to $V_{G2}$ equals to let us suppose $V_{CM}$, right? Let us suppose I have a common mode signal been given whose value is equal to $V_{CM}$ then I get $i_{D1}$ equals to $i_{D2}$ equals to I by 2, right? Because you are drawing i current and then therefore half the current will go to left and half the current will go to right arm.

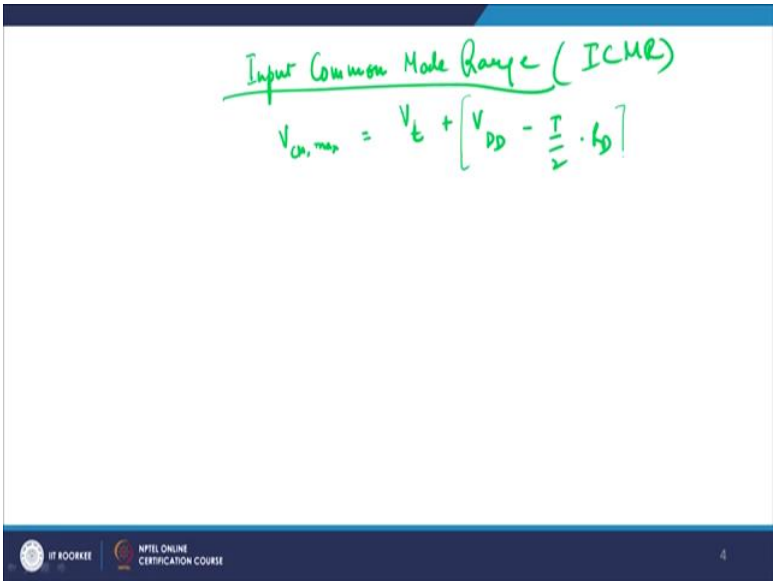
Therefore, $V_S$ is equal to $V_{CM}$ minus $V_{GS}$. What is $V_S$? This point is S, so the voltage at this particular point is nothing but the applied voltage at this particular point minus this difference gate to source will be this voltage. And that is what I have shown here so $V_{CM}$ minus $V_{GS}$ that I get I by 2 therefore current equals to 1 by 2 mu , right? C oxide W by L which is the process trans conductance parameter we have already seen this into $V_{GS}$ minus $V_T$ whole square this can also be written as $K_n$' me to a larger extent. Where we can write down $V_{overdrive}$ to be equals to $V_{GS}$ minus $V_T$ this is known as overdrive voltage, right?

It is overdrive voltage which is equal to $V_{GS}$ minus $V_T$ therefore I can just replace this and write down half Kn', right. * W by L into V overlap Whole Square this is equals to I by 2. Which means that if you therefore find out I overlap this 2 will cancel out and therefore my V overlap comes out to be equal to how much, you just check it out it will be given as I divided by Kn prime into W by L this thing in square root of this one that I is the current drawn to say. W by L is the aspect ratio and Kn prime is given by $mu_n$ into C oxide, right? And I get $V_{overlap}$ to be equals to root of this whole quantity.

Therefore I can write down in such a scenario when VG 1 equals to VG 2. I can write down therefore $V_{D1}$ equals to $V_D$ to equals to $V_{DD}$ minus $I_D$ is I by 2 into $R_D$ and therefore I get that

which basically means that $V_{O1}$ and $V_{O2}$ are exactly the same and therefore if we subtract $V_{O2}$ minus $V_{O1}$ I will get output voltage to be 0 or I can write down $V_{OD}$ to be equals to 0. So therefore any common mode signal as I discussed in your BJT also. In a MOS device also the common mode signals that is vanished off because its output value is illustrated in this manner.

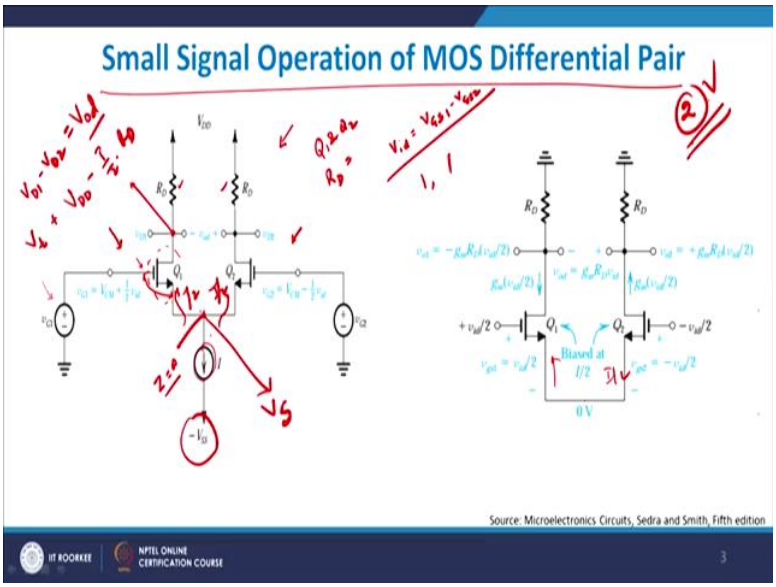
(Refer Slide Time: 10:08)




Now, let me therefore explain to you what is known as Input Common Mode Range. So what the idea is, we also refer it to as ICMR. Now ICMR basically tells me therefore what should be the maximum input range to ensure that the transistor which you are working is actually working in active region and going to cut off.

So you have to be very careful that you do not let the transistor go into the saturation directly but right from the active region it goes to the cut-off region and vice versa. Now, let me just write down for you and we will explain this later on that $V_{CM\text{-}Max}$, $V_{CM}$ means the common mode maximum voltage which you can give is basically $V_t$ plus $V_{DD}$ minus I by 2 $R_D$, you know the reason why, because, you see $V_{DD}$ minus I by 2 $R_D$.

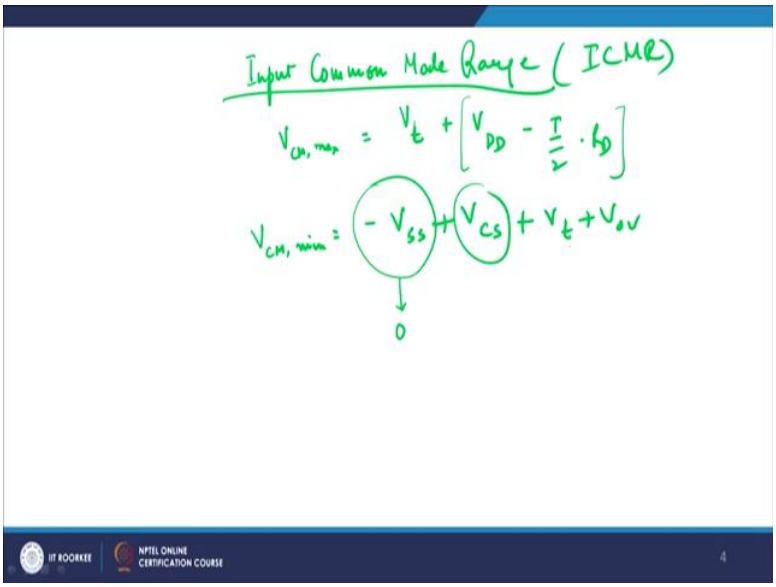
(Refer Slide Time: 11:12)




If you look very carefully here is the voltage at this particular point, this is $V_{DD}$ minus I by 2 into $R_D$ is the voltage at this particular point. You are added $V_t$ to it threshold voltage of this device Q1 this has to be the minimum voltage. See, if my voltage falls below this, right? Say my threshold voltage is, my $V_t$ is say 1 volt and $V_{DD}$ minus 1 by 2 $R_D$ is equals to say 1 volt again.

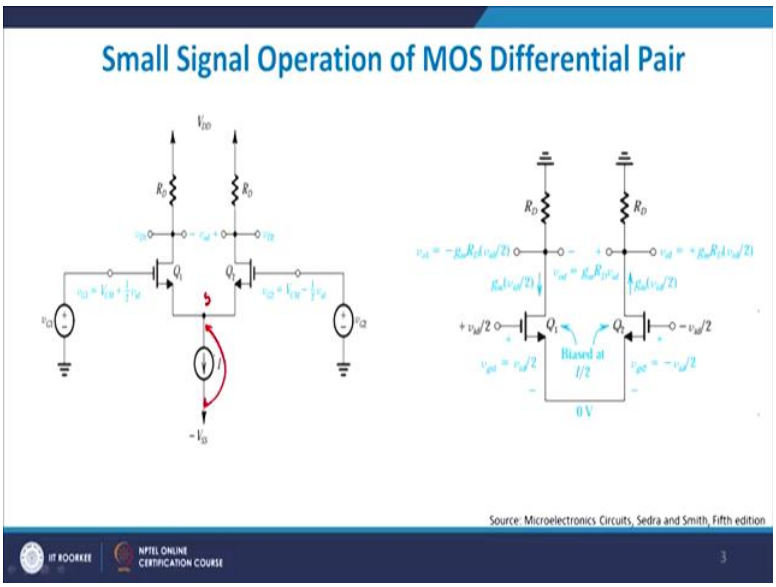
So 1 plus 1 is 2 volts, so minimum common mode voltage which you should give to the input of Q1 or Q2 assuming database symmetric minimum is 2 volts. If you keep the voltage less than this 2 volts then either Q1 will be turned off permanently or you will never get this much amount of voltage drop at the output of your $V_{01}$, right? And therefore, this is the reason why we define this to be as the maximum value which you see.

(Refer Slide Time: 12:12)



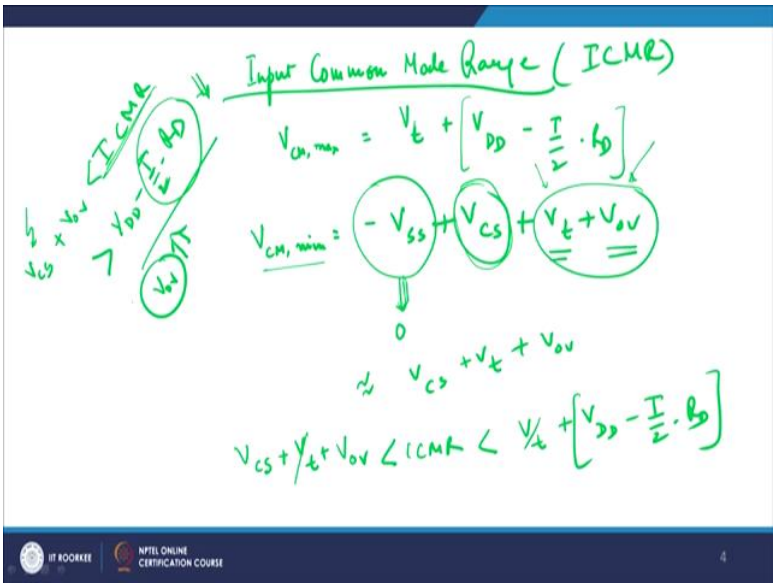

Now, what is the meaning of $V_{CM\ min}$, minimum CM voltage is minus $V_{SS}$ plus $V_{CS}$ plus $V_t$ plus $V_{overlap}$. Now, obviously this could be grounded also, so this can be put to 0 otherwise it is connected to a negative terminal. $V_{CS}$ is this voltage which I will just show it to you.

(Refer Slide Time: 12:40)




This is the voltage which I discussed with you just now. $V_{CS}$ means this is the S value, right? From this point to this point this is the $V_{CS}$ value which is the voltage across the current source that is the minimum value plus what do you require?

(Refer Slide Time: 12:50)




Plus you require $V_t$ threshold voltage of the Q1 transistor and $V_{overdrive}$. You understand why these 2 are kept because otherwise the system will not get on. So your minimum common mode voltage should be this much that it should be able to switch on your device, draw the device into the active region by $V_{overlap}$ and should sustain a minimum value of voltage between the source and through the current source otherwise it will not work fine.

And VSS for all practical purposes always equals to 0 and therefore I can write down this to be typically equals to $V_{CS}$ plus $V_t$ plus V overlap. So the ICMR values which you see, right? The minimum value is this much, assuming $V_{SS}$ equals to be equals to 0 should be greater than $V_{CS}$ plus $V_t$ plus $V_{overlap}$, right? And should be less than $V_t$ plus $V_{DD}$, right, minus I by 2 into $R_D$, fine. So this is what we get maximum.

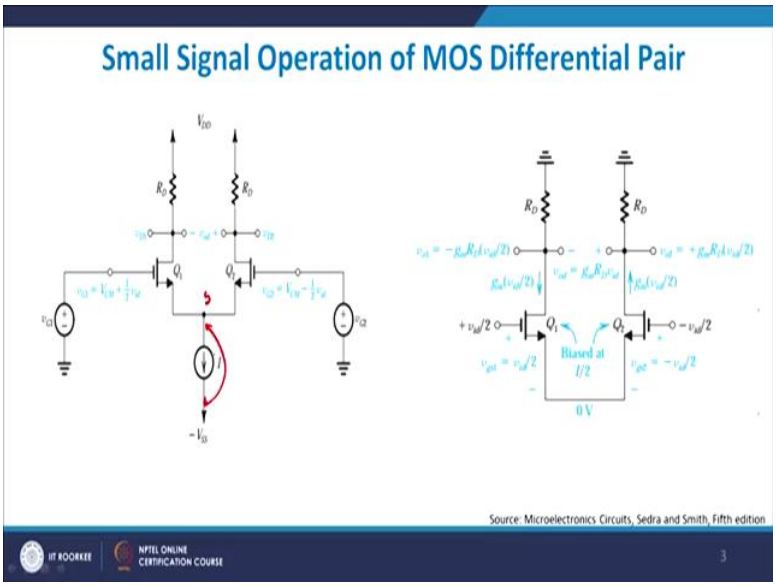
So if we relate these 2 from both the sides because they do not play a role as such I can safely write down that ICMR that $V_{CS}$ plus $V_{overdrive}$ ICMR, right? And it should be greater than $V_{DD}$ minus I by 2 times $R_D$, right? So you see therefore, if my $V_{DD}$ is high, right? My input voltage should be also going higher and higher. My input Common mode range will also go higher. Similarly if my overlap is high my ICMR will also go high and the reason being in front of you that, it should be at least greater than $V_{CS}$ plus $V_{overlap}$.

VCS is almost fixed because assuming that it's an ideal current source obviously the voltage across the current source will always remain fixed and therefore you add overlap you get ICMR. So ICMR is always greater than $V_{overlap}$ in all respects. If you go on increasing

$V_{overdrive}$ you will be increasing the current and therefore what will happen is, this quantity will increase and therefore this quantity will actually decrease, right?

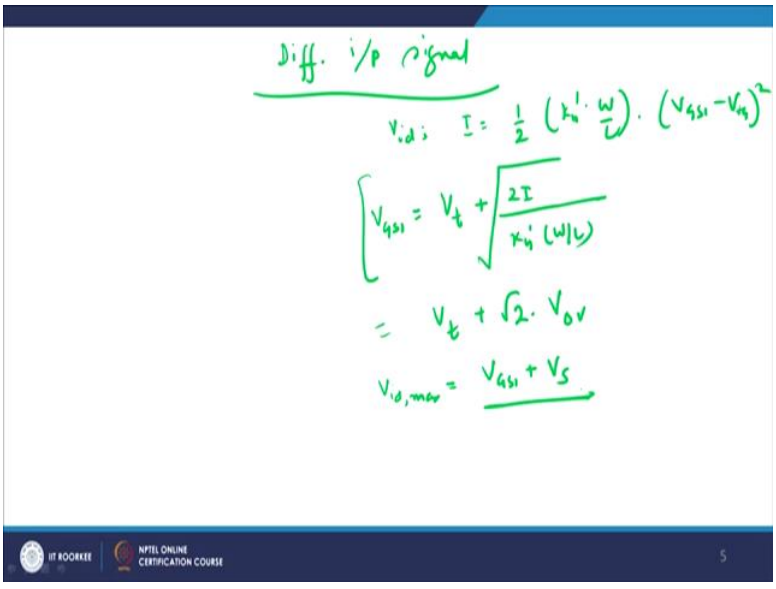
And therefore this will be obviously greater than the decrease quantity, right? So what people generally do is that, they try to make the overdrive slightly larger in order to sustain this relationship in real sense, right? And this is what is ICMR looks like in a real sense of operation.

(Refer Slide Time: 15:23)




This is again as I discussed with you the small signal model, right? And let us now put a input differential signal and see how it works out.

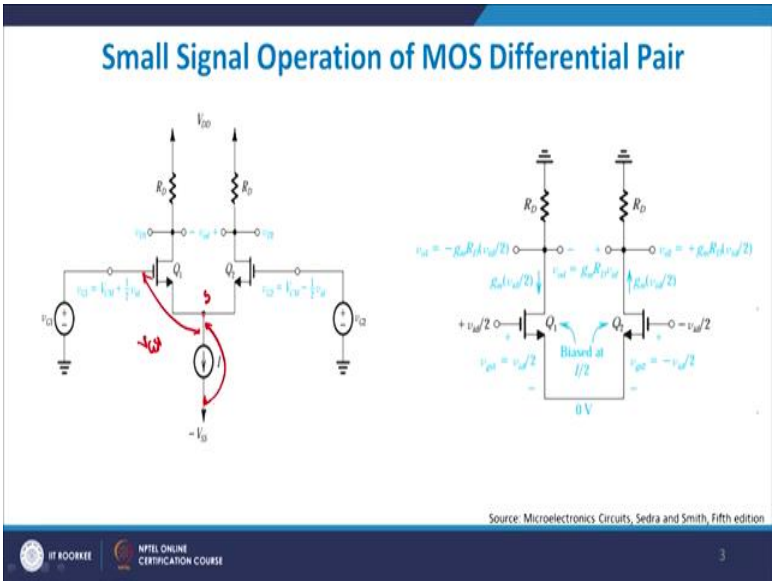
(Refer Slide Time: 15:31)



So, what I do I now put a Differential Input Signal and then what I write down is I refer to as $V_{id}$ and therefore, my current is equals to 1 by 2 $K_n$ prime into W by L into $V_{GS1}$ minus $V_{th}$ whole square assuming it to be in the active region. Therefore, $V_{GS1}$ is equal to $V_t$ plus 2 I by $K_n$ prime into W by L root over. This is equal to $V_{GS1}$, right? So that is nothing but $V_t$ plus root 2 $V_{overlap}$ because if we take to outside than its root 2 into root of I divided by $K_n$ prime W by L that is nothing but $V_{overlap}$, right?

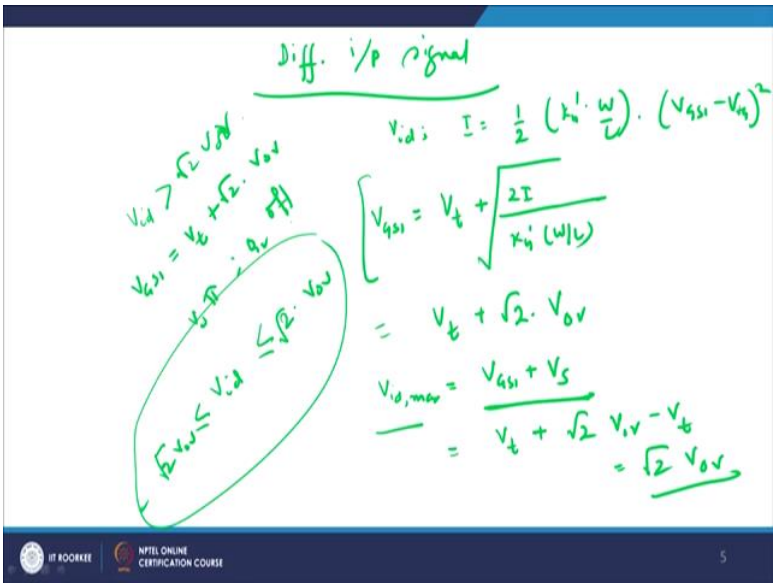
So I get current to be equals to this much. Assuming that have the current is routed through both the arms. Similarly, I get $V_{iD\ Max}$, input voltage difference should be equals to $V_{GS1}$ plus $V_S$ and the reason being that, why it should be greater than $V_{GS1}$ plus $V_S$ is that, $V_{GS1}$ is nothing but gate 2 source of…

(Refer Slide Time: 16:46)




This is $V_{GS1}$, right? And this is $V_S$ value, so this potential plus this will give you the potential at this particular point.

(Refer Slide Time: 16:57)

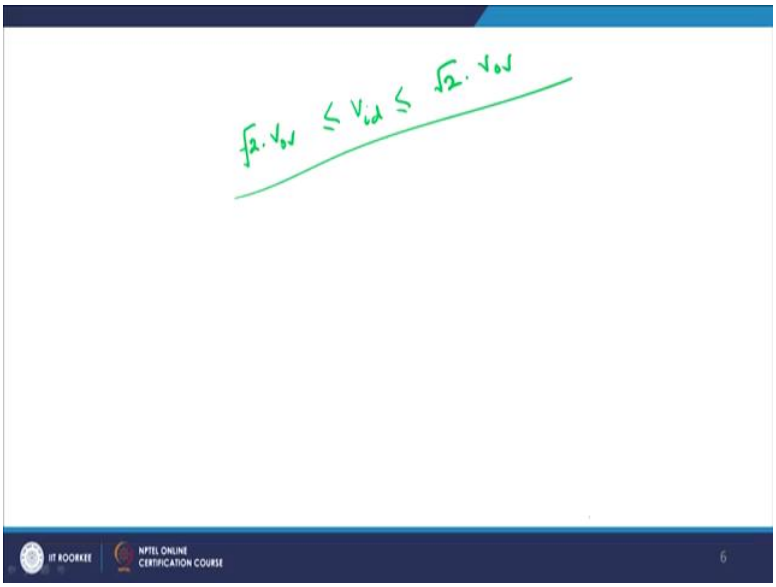



That is what I am trying to say here that $V_{GS1}$ plus $V_S$ should be here. If you solve this and do some small manifestation I get $V_t$ plus root 2 $V_{overlap}$, right? Minus $V_t$ and therefore this will come into root 2 $V_{overlap}$. So $V_{iD\ Max}$ is this much which you get. Now which means that if $V_{id}$, right? If it exceeds root 2 $V_{overlap}$ then what will happen is that all the current from one arm will be routed through other arm, the other arm will remain almost in the cut of position, right?

And that is the reason you do not like to draw it. And $V_{GS1}$ will be written as $V_t$ plus root 2 $V_{overlap}$ which means that as $V_S$ will go on increasing, right? Q2 will be switched off, why? Because let us suppose your overdrive is increasing, right? And your source voltage is also increasing as you move along. If the source voltage is increasing and you are trained voltage is fixed, right? Then $V_{DS}$ will go on decreasing.

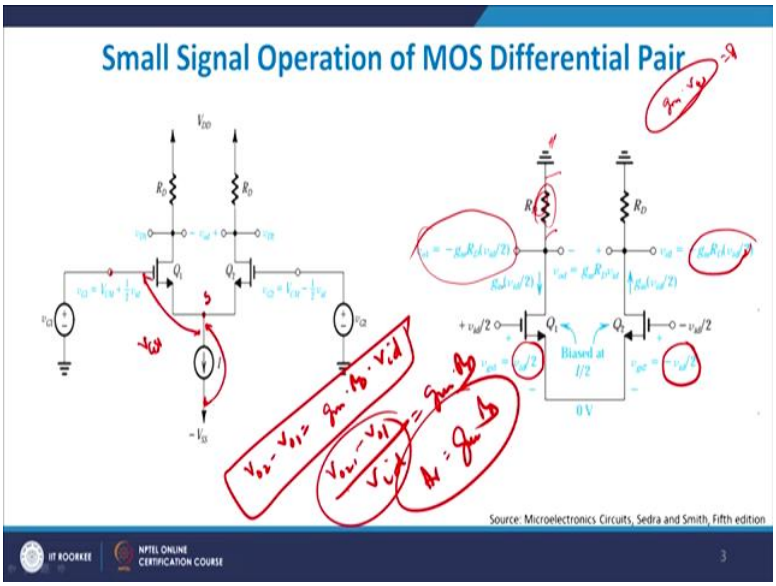
Decreasing primarily means that you are forcing the device to go into the deep triode region but that is not my aim, right? And therefore, I cannot let my $V_S$ fall down drastically below a particular point. And therefore, I do not want the Q2 to be fully off, right? And therefore, the maximum value which can be given to a system differential signal is that $V_{id}$ should be greater than less than equals to root 2 $V_{overlap}$ and should be root 2 $V_{overlap}$, this is the final explanation of $V_{id}$.

(Refer Slide Time: 18:45)




So, the differential signal maximum which you can draw is basically minus of root 2 into $V_{overdrive}$ and this is maximum of root 2 into $V_{overdrive}$, this is the range of input signals which you can give. Now, the idea therefore is that by giving such a signal I am able to sustain it. Anything larger lesser than this, larger then this will force further the transistors to go into cut off and other will not go into cut off in a general sense. So this is the maximum value of signal which one gets, right?

(Refer Slide Time: 19:22)

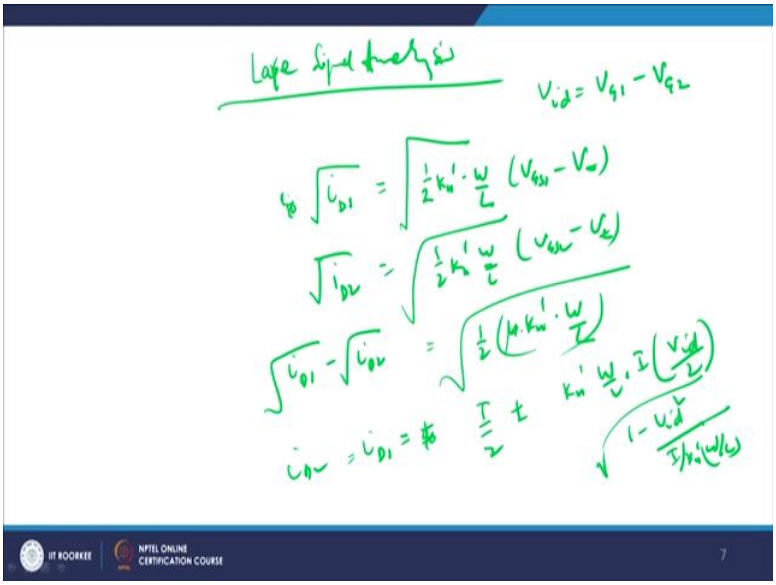



So, let me show to you therefore what is happening here. As you can see a bias state, I have $V_{GS1}$ equals to $V_{id}$ by 2 and $V_{GS2}$ is therefore minus $V_{id}$ by 2, right? And therefore this much

amount of current is flowing through system gm times VGS is basically my ID, so VGS here is nothing but $V_{id}$ by 2. So $g_m$ multiplied by $V_{id}$ by 2 is this current, this multiplied by $R_D$, right? Is the voltage across this device, since this is already grounded I get $V_{01}$ to be equals to minus $g_m$ $R_D$ upon into $V_{id}$ by 2.

Similarly, if you look at this side you will get gm $R_D$ into $V_{id}$ by 2 which you get, right? And therefore if you subtract this minus this which is $V_{02}$ minus $V_{01}$ I get this plus this and therefore I get gm times $R_D$ times $V_{id}$, right? Therefore, $V_{02}$ minus $V_{01}$ by $V_{id}$ is nothing but gm times $R_D$ and this is only for the voltage gain because this is the output voltage is difference divided by input voltage difference turns out to be therefore $g_m$ times $R_D$, right, which is exactly what we have already derived in our previous discussion and slides.

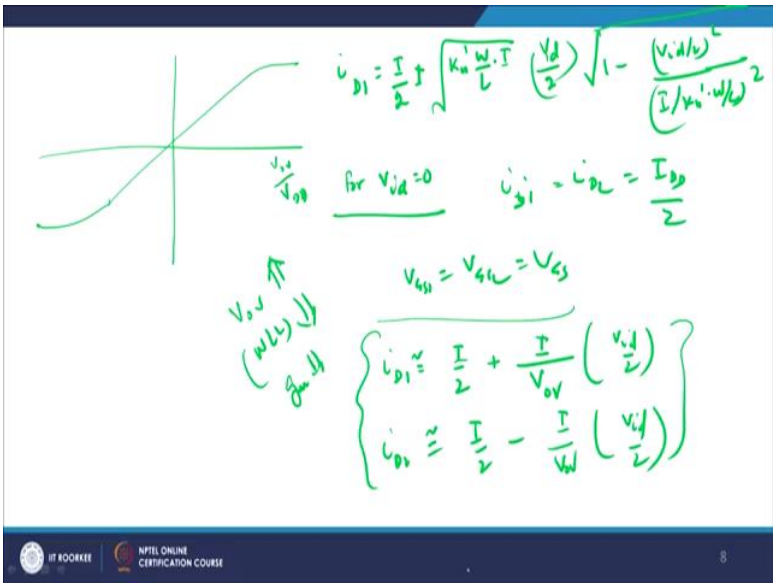
(Refer Slide Time: 20:45)




So, with this knowledge let me show to you the large signal analysis here. In the large signal analysis what I can write down is, I am assuming that $V_{id}$ equals to $V_{G1}$ minus $V_{G2}$. So, if I solve it I get root of $i_{D1}$ is equals to 1 by 2 Kn' W by L root over into $V_{GS1}$ minus $V_t$. similarly root $i_{D2}$ to be equals to 1 by 2 $K_n$ prime W by L into $V_{GS2}$ minus $V_t$, fine. So if you subtract $i_{D1}$ minus $i_{D2}$ square root of that, right?

Something will cancel out and I get root of $i_{D1}$ minus root of $i_{D2}$ equals to half Kn' into W by L, right? This is what you get as the overall $i_{D1}$ minus $i_{D2}$ and if you do a small manipulation I get $i_{D1}$ to be equals to $i_{D2}$ equals to, right, I by 2, right? Plus minus, so this is plus minus Kn prime W by L into I into $V_{id}$ by 2, right? And then square root of 1 minus $V_{id}$ square divided by I upon Kn' W by L.

(Refer Slide Time: 22:22)




So, I will just write down for your information said what I got. I got $i_{D1}$, right? Equals to I by 2 plus I by 2 plus minus root over (Kn prime into W by L into I , right? This is into Vid by 2, right? And this is root over 1 minus Vid whole square divided by I divided by Kn' into W by L whole square. This is your $i_{D1}$, similarly $i_{D2}$ will be with the negative sign of course here.

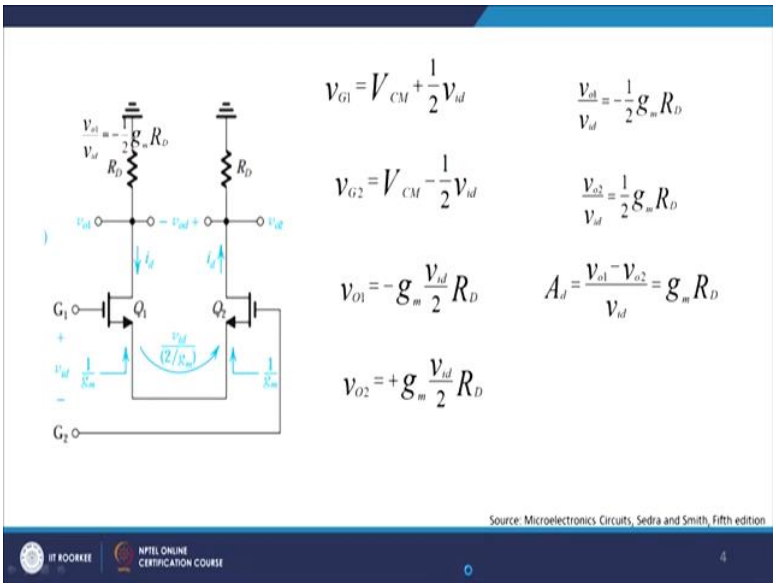
Now, for Vid equals to 0 which is then there is no differential signal implies that your $i_{D1}$ will be exactly equals to $i_{D2}$ equals to $I_{DD}$ by 2, half the current will flow through this arm. Now in that case $V_{GS1}$ equals to $V_{GS2}$ equals to $V_{GS}$ gate to source of first transistor, right? And that gives me a proper idea about where you will be biasing your device in order to get such a figure.

So, what I get from this figure, all this discussion is finally I get I by 2, right? I by 2 plus I by $V_{overlap}$ into $V_{id}$ by 2 right? And then I get $i_{D2}$ to be approximately equals to I by 2 minus I by $V_{overlap}$, right? $V_{id}$ by 2 I get which means that the currents which I see they are separated by each other by 180 degree phase shifts, right? So this is another issue which we have also dealt in BJT when we are plotting the current versus $V_{overlap}$ by $V_{DD}$ and you will see that all of them crosses through 0 and then they get something like this.

So the gain actually starts to roll off at higher values of overdrive. So if you want to sustain a large amount of linearity that is always advisable to keep your $V_{overdrive}$ to be as large as possible, right? Which implies the W by L should be as small as possible, right? And therefore the trans conductance will be a small as possible.

The reason being that $g_m$ when you keep overdrive voltages very large, right, which means that $V_{GS}$ minus $V_T$ is typically very large which also sustains that your $g_m$ will be also large and therefore gm times $R_D$ becomes large which means that the overall gain of the system actually increases drastically. The cost you pay for it is heavy nonlinearity in a real sense. So that is very-very heavily non-linear and gives you a difficult time in terms of understanding the theory behind it and it is really difficult to get the whole of it.

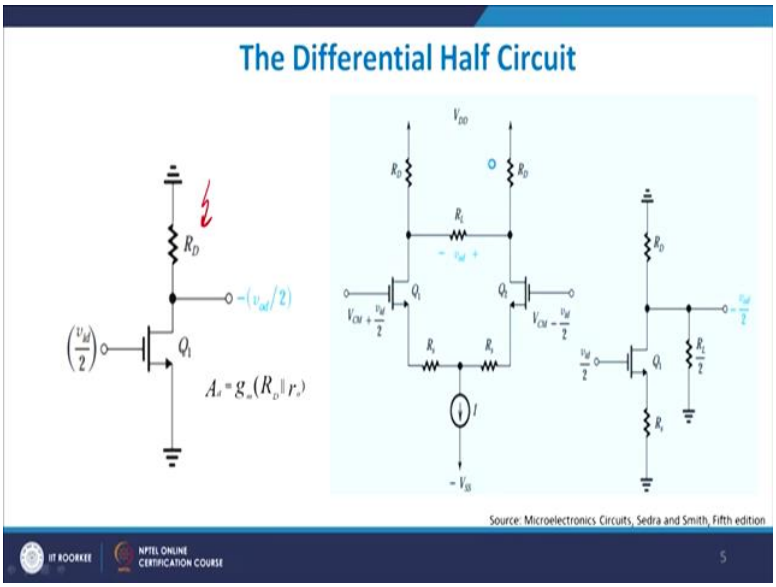
(Refer Slide Time: 25:24)




So what I do, as I discussed with you just now that if I do a simple design here which is basically even your half circuit design, this is basically a half circuit design, I can show to you that using this half circuit design we can actually do it. What is half circuit? Well, half circuits are very standard methodology used in analog circuit design or mixed signal design where we assume that since the ground and power supply are the same for both the arms.

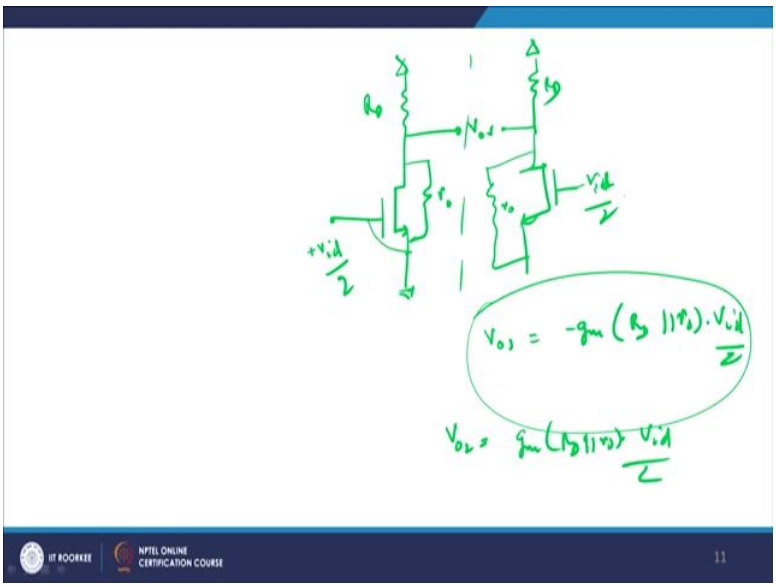
I can divide the 2 arms into 2 parts and then just fine for one something, some important term and do a complimentary equation for the second one assuming that they are perfectly symmetrical with respect to each other and these are known as differential half circuits.

(Refer Slide Time: 26:07)




Maybe, I can show you, yes. As you can see it is a differential of circuit, right? So this is the differential half circuit which you see here, right? And we will discuss this in detail as we move along. But this is what your differential amplifier based design, right?

(Refer Slide Time: 26:34)

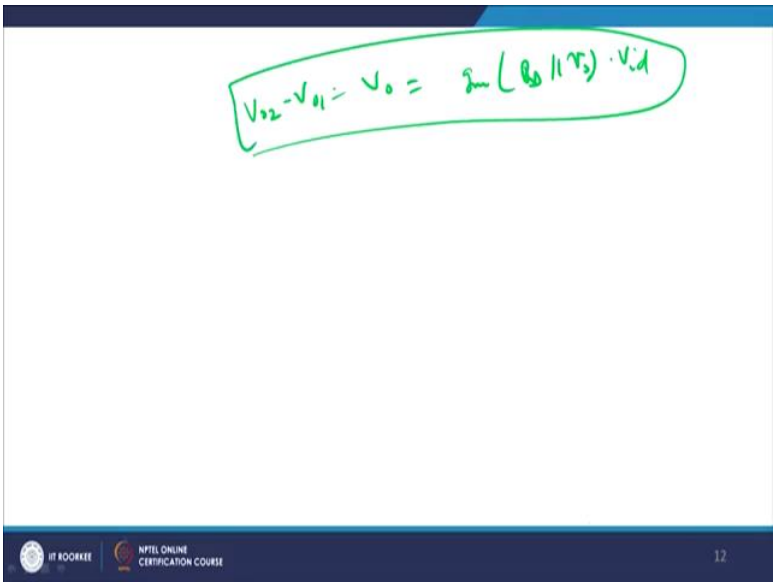



So with this knowledge or with this idea let me come to differential half circuit and explaining to you how it works out, right? So the differential half circuit, does what? It divides the whole phenomena or the whole devices or the whole material or the circuit into 2 parts. The first part is basically your $R_D$, right? And of course this is the plus $V_{id}$ by 2 half the value of the difference then what you do is you grounded first of all.

And then assume that across the gate to source you have r0 here and you have RD here and then you have V here and then you have $R_D$ here, right? And this is grounded, this is also grounded then as you come down to enter into this thing, so this is my minus $V_{id}$ by 2 because they are perfectly anti-Semitic with respect to each other, right? And therefore this will also have resistance which is basically given as r0 also in this case.

So this is the half circuit which you see, right? Half under this side, half under the side therefore I can write down V01 to be equals to minus gm times $R_D$, right? Parallel to r0 multiplied by $V_{id}$ by 2 because this is grounded, so $V_{GS}$ is nothing but $V_{id}$ by 2. So, $V_{id}$ by 2 minus that threshold voltage of the device is the overdrive and from there I get $V_{01}$ to be equals to this value which you see. And similarly $V_{o2}$ hour output voltage 2 will be given as $g_m$ times $R_D$ parallel to $r_o$ into $V_{id}$ by 2 of course just with a negative sign changes there.

(Refer Slide Time: 28:18)




With this knowledge I can safely write down $V_{02}$ minus $V_{01}$ which is $V_0$ to be equals to $g_m$ times $R_D$ parallel to $r_o$, right? Into $V_{id}$, right? So therefore if I do a pure differential analysis or pure differential analysis as such then what we get is that we get the output voltage difference to be equals to gm times $R_D$ parallel to $r_o$ into $V_{id}$, right? And therefore higher the value of $g_m$ more is the gain which you get at the cost of linearity.

So, linearity is always adversely getting affected by high gain because when you do a high gain you primarily also increase the non-linear distortion as compared to your signal like both of them increase and therefore it's sort of losing out the whole thing as far as design is concerned. In our next segment of slide we will do some mismatching of $R_D$ and $g_m$ which we

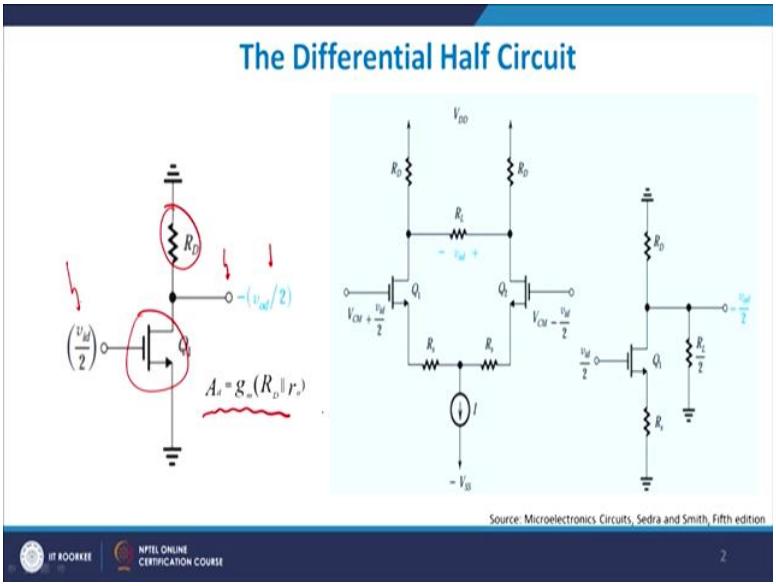
were supposed to do in this part and then we will discuss about differential amplifier cascoded and then common mode games we will discuss, right?

So, I think we will stop here at this stage and in the next module we will discuss the $R_D$ and gm mismatch as we will also discuss about CMRR and the Valley of $A_{CM}$ for CMRR equals to infinity, thank you. Okay, thanks a lot, thank you very much.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−38**
**Small Signal Operation and Differential Amplifiers-II**

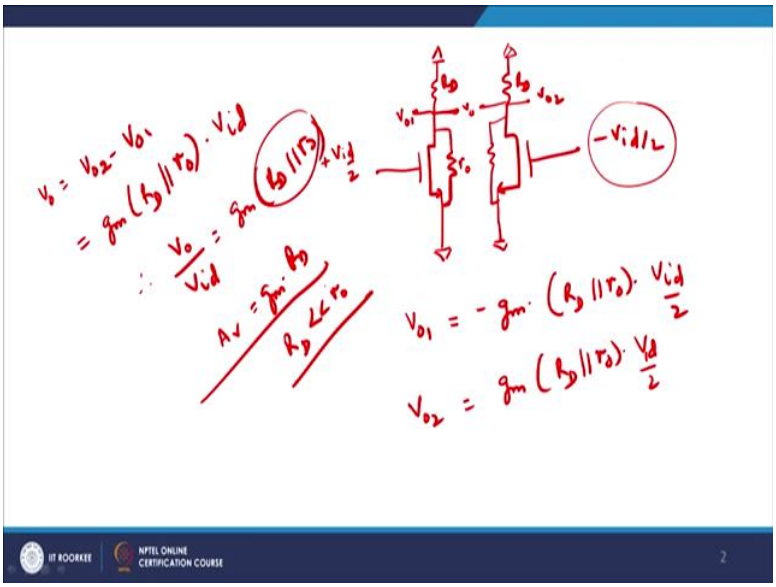
Hello and welcome to the NPTEL online certification course on Microelectronics Devices and Circuits. In our previous module we had seen the basic concept of MOS based differential amplifier. We have also seen the concept of half circuit as applied to MOS-based differential amplifiers and we saw the advantage that a MOS-based differential circuit will have mathematically advantage because once we are able to find out the voltage at a particular point, for a particular bridge the opposite of that will be applied to another bridge and therefore we can easily calculate the overall gain. So, what we will do now is, we will just continue from where we left in our previous discussion.

(Refer Slide Time: 1:25)




And we saw that this is the differential, so you see this is basically your half circuit which is in front of you, differential half circuit which I was discussing with you that you do have a Q1 which is basically your MOS transistor here and I apply a voltage $V_{id}$ by 2 where $V_{id}$ is basically the input voltage difference and $R_D$ is resistance offered. And we get minus $V_{od}$ by 2 is the output voltage with the negative sign because it is 180 degree phase shift. In such a scenario the overall gain is given as gm times $R_D$ parallel to ro, right?

(Refer Slide Time: 1:55)

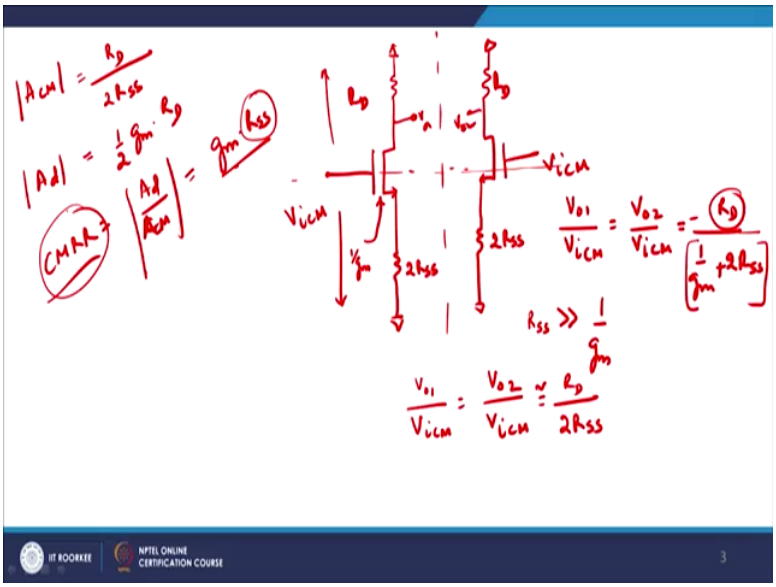



And let me show you how this works out. If I have got to circuits here, right? And they are primarily given like this and this is plus $V_{id}$ by 2, this is $R_D$ and then this is again the same here. So there are 2 half circuits which is here and then this is the series resistance of the device itself, so this is Ro, this is $R_D$ and this is again V0 output voltage which we are trying to find out and this is minus $V_{id}$ by 2.

So, you see $V_{01}$ which is this one, this is the output $V_{01}$, right? And this is $V_{02}$. So $V_{01}$ if you want to find out will be minus gm times $R_D$ parallel to ro multiplied by $V_{id}$ by 2. Similarly, $V_{o2}$ will be equals to $g_m$ times $R_D$ parallel to $r_o$ into $V_{id}$ by 2, so it would have been minus but since there is a minus sign already available into the input signal, minus and minus becomes plus and I get this into consideration.

Therefore, output voltage $V_o$ will be equal to $V_{o2}$ minus $V_{o1}$ and this comes out to be equals to gm times $R_D$ parallel to $r_o$ multiplied by $V_{id}$ and therefore, Vo output by $V_{id}$ is nothing but gm times $R_D$ parallel to $r_o$. Now, of course, as you very well know that my $R_D$ will be, $r_o$ is much much larger as compared to $R_D$ and therefore, for all practical purposes the gain will be equal to gm times $R_D$. In reality this gm $R_D$ parallel ro which is this parallel resistance which you see because we have assumed all we know that $R_D$ is much smaller as compared to $r_o$, right? And we get that into consideration for all practical purposes.

(Refer Slide Time: 4:13)




Now with this knowledge and we therefore see that, now let me put the same concept. It will do the same concept here and explain to you that for a common mode signal for example. For a common more signal let say what happens. For a common mode signal I need to find out CMRR now and therefore what I will do is, I will just take 2 $R_{SS}$.

This is with the current source, tail current source, remember. So I have divided into $R_{SS}$ by 2 because when you take in parallel this 1 by 2 $R_{SS}$ will give you just $R_{SS}$ as the output voltage because they are in parallel with respect to each other. So assume that I have given $V_{iCM}$ here, I have also given $V_{iCM}$ here. This is $R_D$ and then I have put across this concept here and we will try to find out the output voltage here.

So this is $V_{01}$ and this is $V_{02}$. If you want to find out the output voltage, so this is basically my half circuit here, so why is it $R_{SS}$ by 2? Because we have divided it into 2 parts and we got $R_{SS}$ by 2 as the overall signal. So, if I find out $V_{01}$ by $V_{iCM}$ this will be equal to $V_{02}$ by $V_{iCM}$ will be equal to minus $R_D$ upon 1 by $g_m$ times 2* $R_{SS}$, why?

Because you remember in one of our previous discussions we had seen that if you want to find out the gain or the output voltage gain just by inspection then you just have to do one thing that you divide the whole thing into 2 such that somewhere in the middle, the device is in the middle and this is the output profile and this is your grounded profile.

So, when you go from the side you just check out how much amount of resistance is being offered in the line below. Below this dotted line it is 1 by gm because when you look from

the source side into the MOS device the resistance offered is 1 by $g_m$ and series side is $2R_{SS}$. So, the denominator is therefore 1 by $g_m$ plus 2 $R_{SS}$ and from top if you come, from $V_{DD}$ which is grounded here I get $R_D$ as the output.

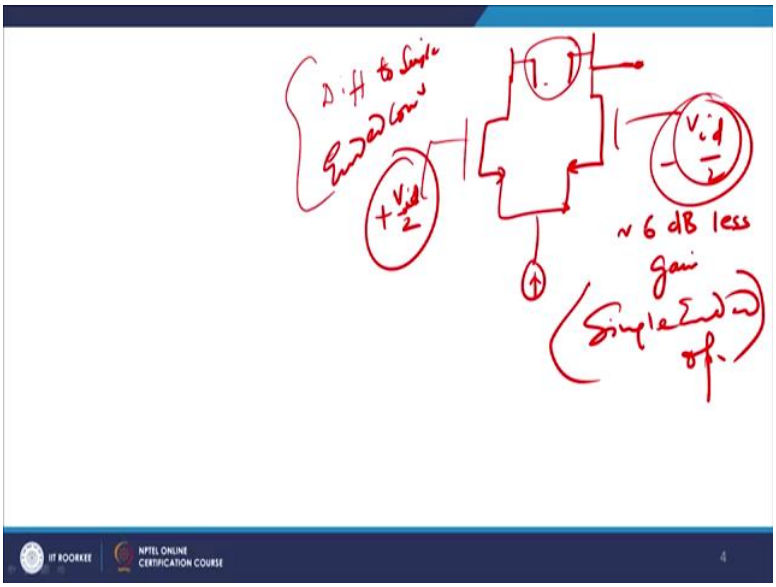
So I get minus $R_D$ upon 1 by gm plus 2 $R_{SS}$ as my value. We already know that $R_{SS}$ is much-much larger as compared to 1 by $g_m$ and therefore I can safely write down $V_{01}$ by $V_{iCM}$ equals to $V_{02}$ by $V_{iCM}$ this is the common mode signal to be equal to $R_D$ upon, right? 2 times $R_{SS}$ because gm is much smaller and I can neglect that value here and I get $R_D$ by 2 $R_{SS}$ as my overall picture.

So, let me find out common mode gain therefore and make it mod. I get $R_D$ upon 2 $R_{SS}$ and I get $A_D$ which is the differential mode gain to be equals to 1 by 2 $g_m$ times $R_D$, why 1 by 2? Because now you see I am trying to find out for a single circuit. So that the reason this is not $g_mR_D$ it is $g_mR_D$ by 2. If you want a differential you multiply it by 2 and I get $g_m$ $R_D$, right?

And therefore I get CMRR which is common mode rejection ratio to be equals to mod ( $A_D$ by $A_{CM}$ ) of that and if you try to find out this overall picture I get $g_m$ times $R_{SS}$. And since RSS is relatively very-very high because you are using tail current source which is assumed to be an ideal current source $R_{SS}$ value is very-very high and therefore CMRR happens to be very high 10 to the power 5, 10 to the power 6 orders of magnitude.

So, with this knowledge what we have understood therefore is that if we do have a differential or a common mode signal. The common mode signal will give you a large value of CMRR. Now, let me just give you a brief inside into one important point which needs to be looked into. Let me see maybe I can explain to you in this case, right?

(Refer Slide Time: 8:18)

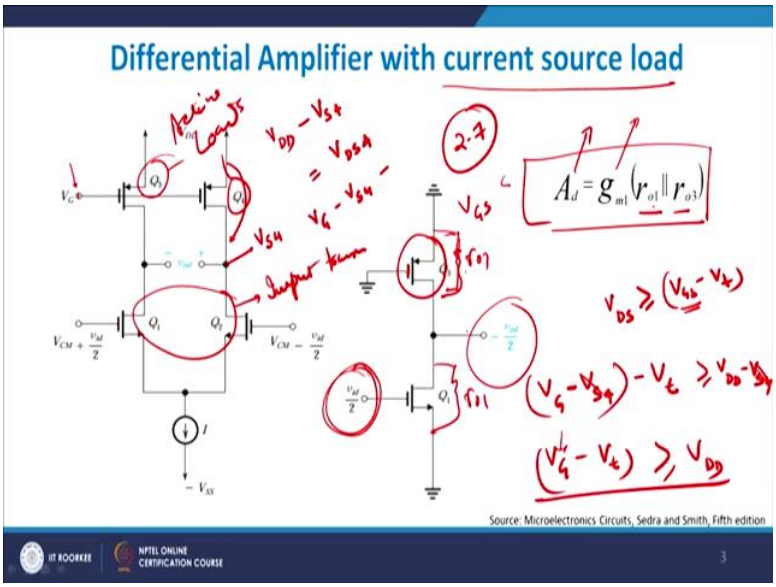


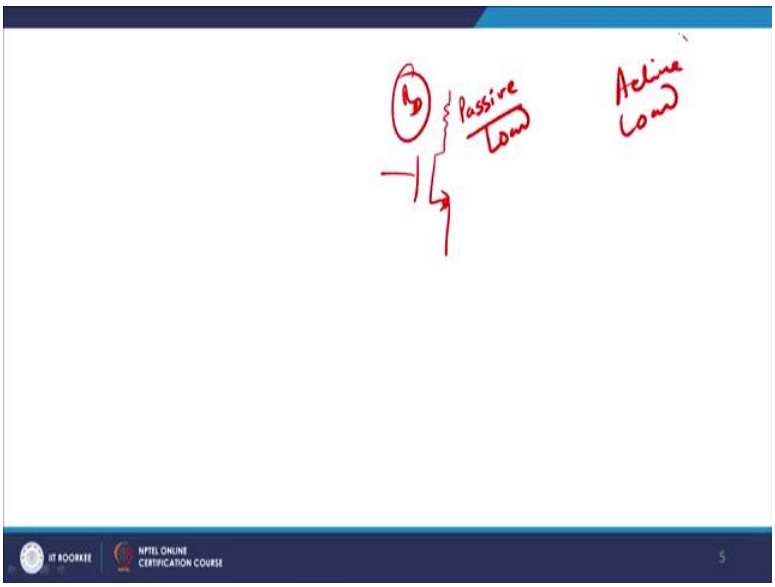

What we were doing is, that they were actually looking into 2 important fact that whenever you do have, let us suppose I have a device and we have another device here and then we have current source which is here. So rather than taking differential output, we can also take a single ended output then your gain falls by 2 but you get a differential, you get a single ended, this is known as a differential to single ended output converter.

So from a differential input and you will get differential input here minus $V_{id}$ by 2 plus $V_{id}$ by 2, right? And I am giving this and this, right? And am trying to find out if it would have been purely differential I would have found out here but no, I am only interested in single ended output, I get this. So you will see that and you can do it yourself by issuing small manipulation and small derivation that this primarily will be giving you twice lesser gain approximately 6 dB less gain will be there, if you do a single ended operation, right?

If you do a single ended operation you will get 6 dB less gain in a more so in a different manner. And second cost you pay for it is, of course, your common mode rejection is not very good, right? So, this is one of the problem areas which people faced when we were doing a differential amplifier, the single ended amplifier.

(Refer Slide Time: 10:12) 10:44




With this knowledge let me therefore explain to you the various other factors that, when we do a differential amplifier with the current source, so what happened was, with this large gain, it also rejects noise as we discussed common mode signals. But if you go back to your previous discussions and understandings you will see that we did have a $R_D$ here, remember. We had a $R_D$ here so, I have a $R_D$ here.

Now this $R_D$ is generally realized, so these are known as Passive loads because you're putting a resistance here, right? We also have what is known as an active load. Active load means a MOS device acting as a register that is known as an active load. So most of the time but we

try to do is, we try to insert in place of $R_D$ which is a passive load because it has high power dissipation we try to insert an active load, right?

And this is one of the examples of the active load which you see. So also known as a current source load an active load. So what I do is, I give a gate voltage here and these are basically your P MOS device, so these are N MOS and these are P MOS devices, right? And if you give a gate voltage here such that it is lower of threshold voltage of the device, switches it on keeps in the triode region and starts working as a diode.

So what is the $V_{DS}$ value here? The difference between $V_{DS}$ value is, threshold is connected to $V_{DD}$, if we talk of Q4 then $V_{DD}$ minus let us suppose this is, let me say source, $V_{source}$ of 4. So, I get $V_{DD}$ minus $V_S$ of 4 is nothing but $V_{DS}$ of 4, drain to source of 4, right? And therefore, gate is $V_G$ and minus $V_{S4}$, this is gate to source voltage because, if you subtract this because you remember, what was the condition $V_{DS}$ should be greater than equal to $V_{GS}$ for an NMOS, right? And should be less than equals to $V_{GS}$ minus $V_t$ for PMOS.

So if you find out $V_{GS}$? $V_{GS}$ is gate to source is $V_G$ minus $V_{S4}$ is gate to source, right? So this is your gate to source minus threshold voltage of the MOS device which is Q4, right? And this should be greater than equal to $V_{DS}$ which is nothing but $V_{DD}$ minus $V_{S4}$, right? If, you look very carefully these 2 will get cancelled out $V_{S4}$ $V_{S4}$. I get $V_G$ minus $V_t$ should be greater than $V_{DD}$ for this device to be staying in saturation, right?

So once you are able to sustain a large value of $V_G$, so if your $V_{DD}$ say 2 volts and threshold voltage of these devices are 0.7 then you have to at least gave 2.7 voltage at the gate side to make them behave as ON state devices and they will be acting as resistors here, right? So these are active load devices also which are there with us. So this is what we get.

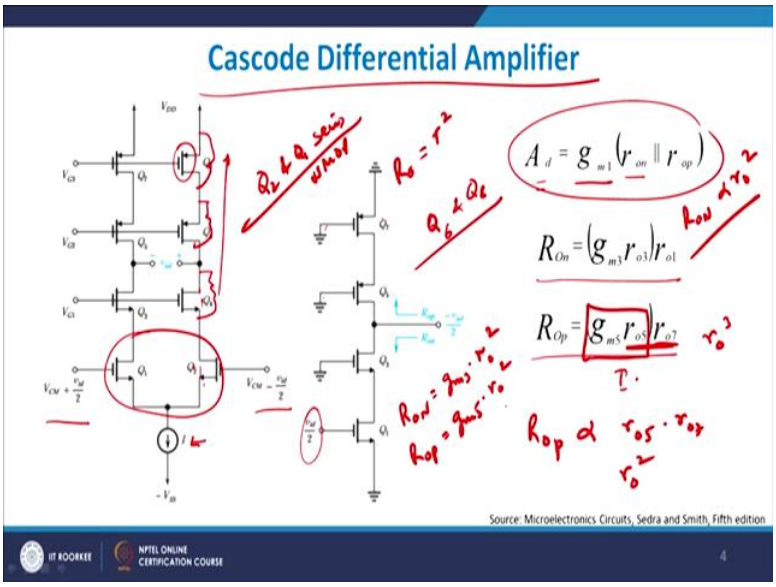
Similarly, in P MOS I can do one more thing that this Q3 can be actually grounded, so if you ground your MOS device and if your drain voltage is already grounded here and your gate is also grounded then gate to source is this one, right? Will always be negative value and therefore this will always make your Q3 on ON state and therefore they will all be like pure resistive domain architecture and they will give you $V_{OD}$ by 2.

If I give, half circuit concept, so if I give $V_{id}$ by 2 here I get $V_{OD}$ by 2 in the output side. And therefore, I get $A_d$ equals to $g_{m1}$ into $r_{01}$ into parallel to $r_{03}$ which means that the overall differential mode gain is depending on the transconductance of the input transistor. So these

Q1 and Q2 are referred to as input transistors, right, whereas Q3 and Q4 are referred to as loads.

In this case active loads, right? So they are referred to as active loads. And therefore when you do $r_{01}$ parallel to $r_{03}$ because if you look carefully $r_{03}$ is here and you have $r_{01}$ here. Since these 2 are in parallel the effective load seen by the Q1 will be $r_{01}$ parallel to $r_{03}$ and you can automatically therefore gates the value of drain voltage in the output side.

(Refer Slide Time: 14:42)




Another method which people adopted and it is still being adopted with certain disadvantages of course is, what is known as a Cascode differential amplifier. So, what we do is, your Q1 and Q2 which is basically your driver transistors they are N MOS devices driven by $V_{CM}$ plus $V_{id}$ by 2 and $V_{CM}$ minus $V_{id}$ by 2 as we discussed earlier and we have a current source here.

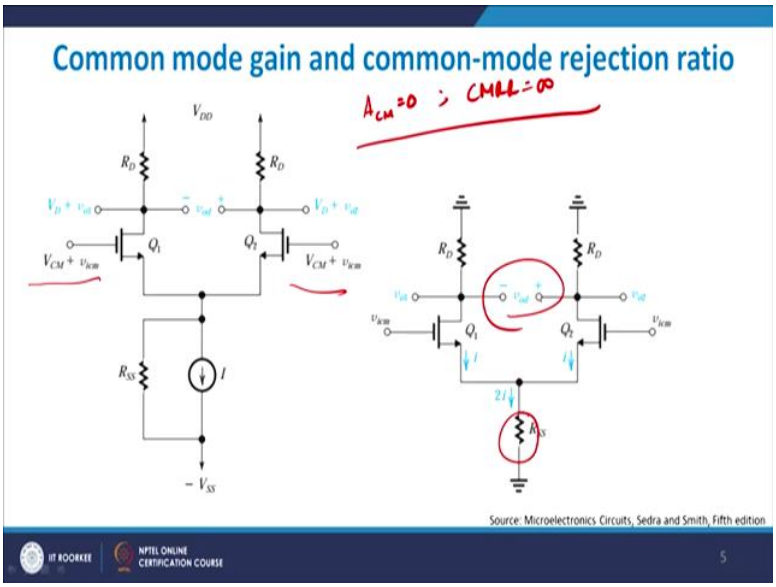
But what we tried to do is in the pull up state, in the higher state we try to have large number of devices and stacked, so in a Cascoded fashion. So that the overall gain can be increased higher and higher because you will have so much larger of devices in parallel with respect to each other. Therefore, I get in this case to be equals to $g_m$ 1 which is again this one into $r_{on}$. $r_{on}$ is basically $r_o$ output resistance of N MOS which is this one.

And then parallel resistance of, gives you $r_{op}$, so $g_{m1}$ into $r_{op}$ into $r_{on}$ is effectively a value of Ad, so if you look at the half circuit concept once again I ground these 3 small signal analysis you have $V_{id}$ by 2 here and I get $r_{op}$ equals to $r_{on}$ which is $r_{on}$ equals to this much. $r_{on}$ and $r_{op}$ is what? $r_{op}$ if you look very carefully is basically looking from the side if you go up.

So you will always have $r_{o7}$ and $r_{o5}$ in series. You see $g_{m5}$ into $r_{o5}$ is nothing but the current flowing through the resistance into the MOS device Q5 current that if you multiply with $r_{o7}$ which is basically the parallel impedance then we get automatically the value of this thing. So if you remember when we try to find out the resistance, it is basically the small signal resistance square which we are trying to find out, right?

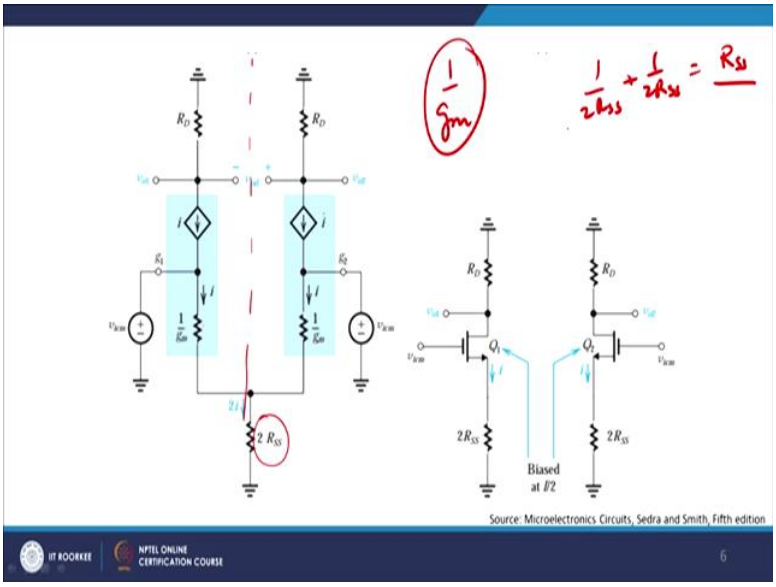
So, Ro is nothing but the square. So if you see $r_{op}$ is proportional to $r_{o5}$ into $r_{o7}$ which is basically if I assume them to be equal I get $r_o$ square, where you do therefore triple Cascode I will get $r_o$ cube. Similarly $R_{ON}$ will be proportional to $r_o$ square because there are 2 transistors in series. So, Q2 and Q4 are in series and they are N MOS and therefore, I get this much. Similarly, Q 6 and Q 8 are again in series and I get $g_{m3}$ multiplied by $r_0$ square. So what I get is, $R_{ON}$ is equals to $g_{m3}$ multiplied by $r_0$ square. I get $r_{op}$ to be equals to $g_{m5}$ multiplied by $r_0$ square. And therefore, simply by choosing higher value of $g_{m3}$ and $g_{m5}$ I actually manipulate the overall resistance or overall gain of the system.

(Refer Slide Time: 17:37)




If you look very carefully therefore that we have discussed in the earlier phases that common mode gain will be always 0 and your CMRR will be always infinitely high. So $A_{CM}$ will be approximately equals to 0 and your CMRR, we have already discussed this point I really should be equal to infinity, right? And that is what I have shown here, that if you have $V_{iCM}$ $V_{iCM}$ I get this and if you do a design than what I do is, I make it $R_{SS}$ assuming that both the currents or i1, i is current flowing I get 2i current flowing here and therefore, $V_{od}$ will always give you a 0 value at the value of common mode signal.

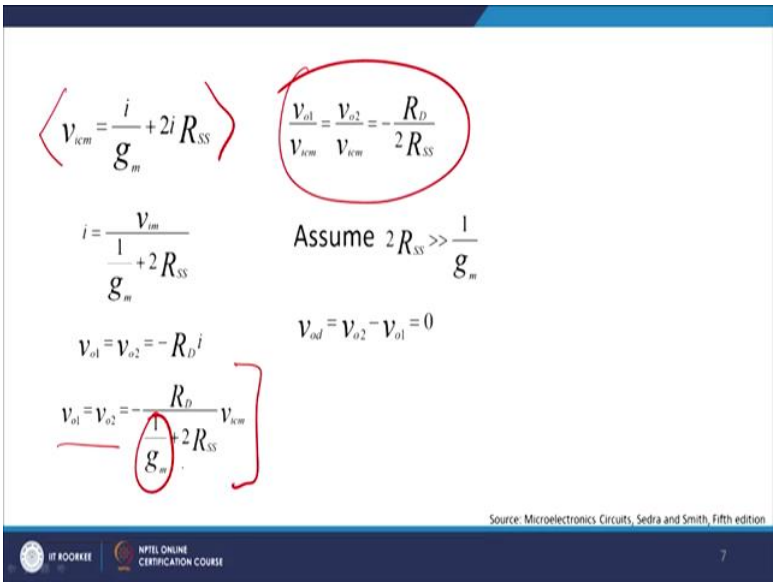
(Refer Slide Time: 18:19)




Therefore, as I discussed with you just now and I was deriving also with you. Do a half circuit here, remember? I remove the MOS device here and show it as a current source in series with 1 by $g_m$, remember because looking at the source side the amount of resistance offered is actually equals to 1 by $g_m$.

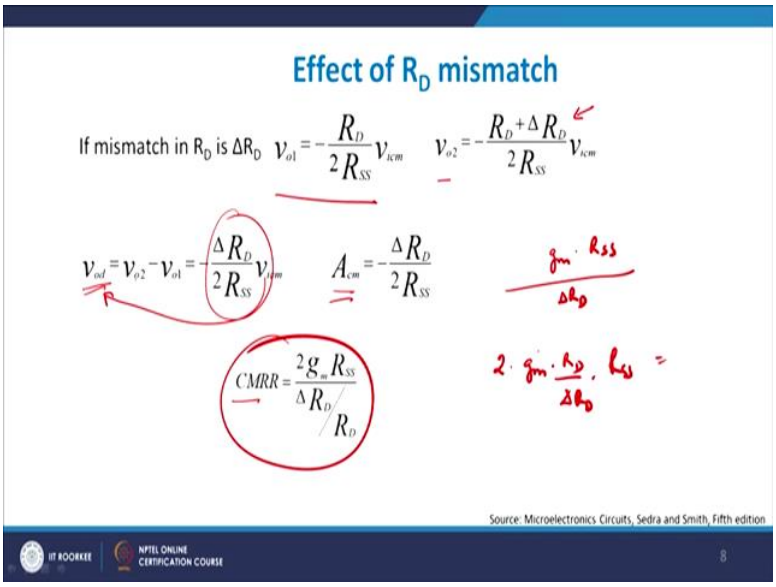
So, I simply do a series connection between the current source and 1 by $g_m$, current source and 1 by $g_m$. And then to $R_{SS}$ because one by 2 $R_{SS}$ plus 1 by 2 $R_{SS}$ in parallel will always give you $R_{SS}$ value in more respect and that is the reason when you do a half circuit I get 2 $R_{SS}$ $V_{o1}$ are ready and then I get I buy 2 current is flowing within 2 arms, right? And this is what the general scheme of things are?

(Refer Slide Time: 19:04)




Now, if you see Victim as I discussed with you will be I by gm plus 2i $R_{SS}$, right? And similarly as I discussed with you just now that $V_{01}$ by $V_{CM}$ plus $V_{01}$ by $V_{ic}$ is our ready by 2 $R_{SS}$. If I assume them to be equal I get $V_{01}$ equals to $V_{02}$ equals to minus $R_D$ by $g_m$. What we did was, that 1 by $g_m$ is much smaller and therefore is $R_D$ by 2 $R_{SS}$ the value of your output voltage.

(Refer Slide Time: 19:33)




Let me, therefore, come to the next stage in $R_D$ mismatch, let me explain to you what is $R_D$ mismatch? See what was happening till now, all this discussion was with the assumption that

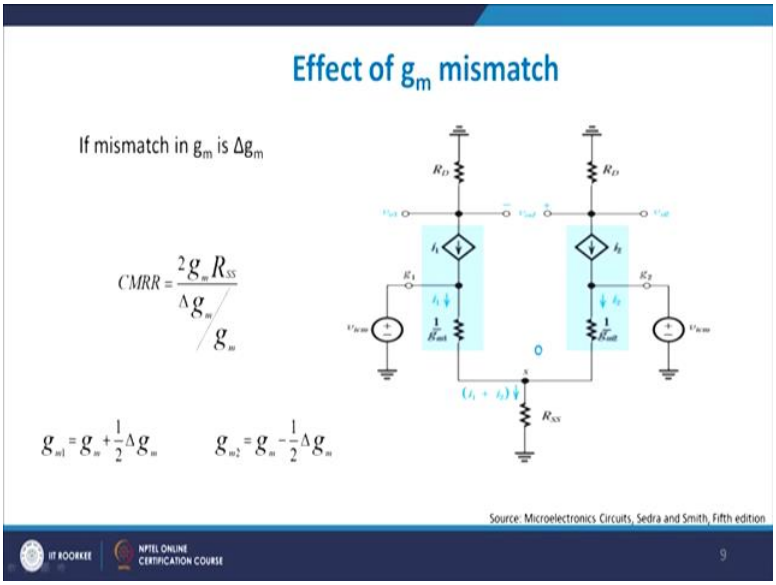
there is no mismatch of the resistances and the transconductances of the device used in the left arm and right of the differential amplifier. They are perfectly symmetrical and therefore all the common mode signals were lost and all the differential signals were magnified.

In reality not true. When you actually fabricate a circuit you will always have a mismatch of the resistors which means that you cannot have a fixed value of register always available with you, they will always vary with each other. So, let us also come to the point that if there is a mismatch in the drain resistance, for example, then how does it influence my common mode rejection ratio?

If you look very carefully if there is a mismatch, as I have discussed with you it is basically our $R_D$ by 2 $R_{SS}$, remember? We just now saw, $V_{01}$ is equal to $R_D$ by 2 $R_{SS}$ and $V_{02}$ will be equal to $R_D$ plus let us suppose is Del $R_D$ by 2 $R_{SS}$ into $V_{icm}$. So we have seen the effect of Del $R_D$ and $V_{icm}$. So, $V_{od}$ output difference is $V_{01}$ minus $V_{02}$ turns out to be this value, right?

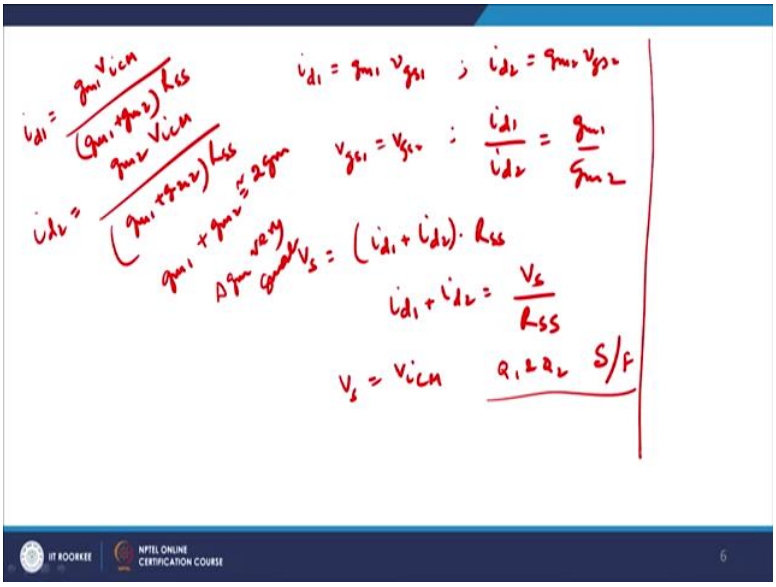
And therefore $A_{CM}$ is basically, so what I do is, I put it into denominator and therefore I get $A_{CM}$ equal to Del $R_D$ by 2 $R_{SS}$. So CMRR is how much? Will be equal to 2 because if you remember, so you will get Del $R_D$ by 2 $R_{SS}$, so you will get gm times $R_{SS}$ divided by Del $R_D$ by 2 $R_{SS}$. This 2 $R_{SS}$ will go to the top and you will get this value of CMRR, right? How much you get? You get 2 times $g_m$ times, right? $R_D$ by Del $R_D$ into $R_{SS}$ as your CMRR.

(Refer Slide Time: 21:34)




Now, if you do a $g_m$ mismatch which is transconductance mismatch. Let us see how it works out in that sense?

(Refer Slide Time: 21:47)

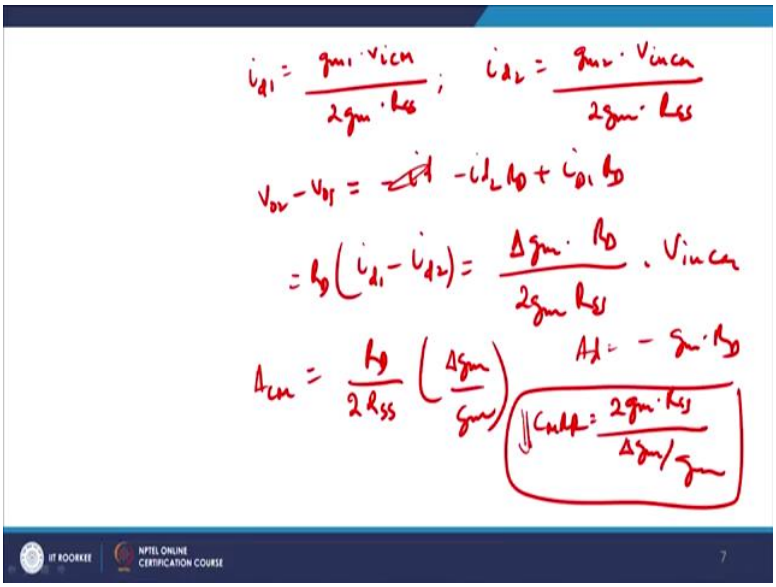



So, if we do a trance conductance mismatch, let us suppose I have got id1 to be equals to gm1 $V_{gs1}$ and $i_{d2}$ to be equals to $g_{m2}$ $V_{gs2}$, right? And if I assume that $V_{gs1}$ is equal to $V_{gs2}$ and then I get $i_{d1}$ by $i_{d2}$ to be equals to $g_{m1}$ by $g_{m2}$ and therefore I get $V_s$ which is the source resistance to be equals to $i_{d1}$ plus $i_{d2}$ multiplied by RSS and therefore, I get $i_{d1}$ plus $i_{d2}$ to be equals to $V_s$ by $R_{SS}$, right?

Now, Vs is nothing but ViCM, because that is a voltage difference between the 2, why? Because Q1 and Q2 are basically source followers, so whatever value of voltage you are giving on the gate side appears across the source side. So the difference between them is almost equal to 0 and therefore I get as this thing. So, I get idi equals to $g_m$ times $V_{icm}$ upon $_{gm1}$ plus $g_{m2}$ divided by $R_{SS}$.

And I get $_{id2}$ to be equals to $g_{m2}$, for this is gm1 into ViCM upon $g_{m1}$ plus $g_{m2}$ divided into RSS, right? Now, if I assume that $g_{m1}$ and $g_{m2}$, the difference is very small then I can safely write down gm plus $g_{m2}$ is approximately equals to twice $g_m$. So, I can write down twice gm here and assuming that Del $g_m$ is very small, it is not 0 but it is very small. So it is relatively very small, right? So if you get this into consideration or you get this into idea. I can write down, right?

(Refer Slide Time: 23:40)

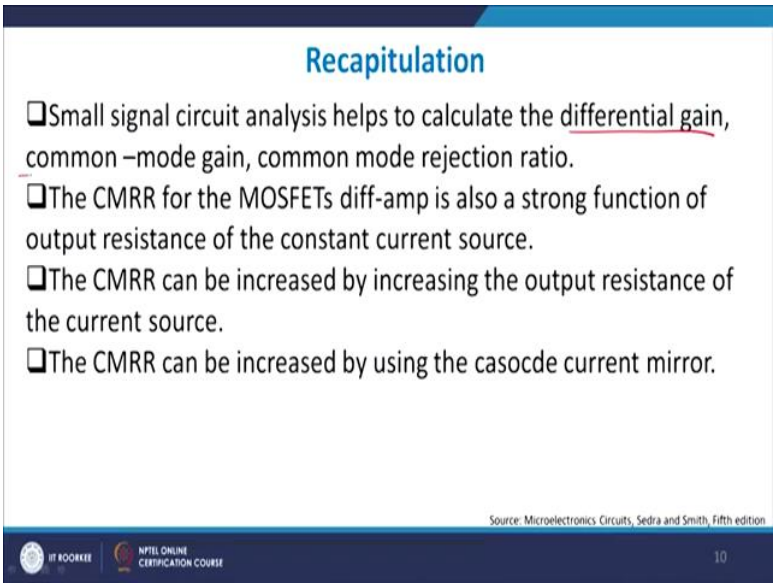



I can write down id1 to be equals to $g_{m1}$ times $V_{iCM}$ divided by $2g_m$ times $R_{SS}$ and $i_{d2}$ to be equals to $g_{m2}$ multiplied by $V_{in}C_m$ divided by 2gm times $R_{SS}$. So, I get V02 minus V01 is equal to minus minus $i_{d2}R_D$ plus $i_{d1}R_D$ region. So if you solve it, I get $R_D$ ( $i_{d1}$ minus $i_{d2}$) which is nothing but Delta $g_m$ times $R_D$ upon $2g_m$ $R_{SS}$ multiplied by $V_{inCM}$. So, if you take $V_{inCM}$ in the denominator I get $A_{CM}$ equals to $R_D$ by 2 $R_{SS}$ into Delta $g_m$ by $g_m$, right?

And since already we know that Ad equals to minus gm times $R_D$, I can write down CMRR to be equal to $A_d$ by $A_c$ which is nothing but $2g_m$ times $R_{SS}$ divided by Del g_b by $g_m$, right? So this is for the overall CMRR which we see when we have a $g_m$ mismatch, right? At this gm therefore depends upon the value of Del $g_m$. So when your Del $g_m$ is typically very large, right?

It tends to lower your CMRR, right? So if your mismatch is low both in terms of $R_D$ or $g_m$ it lowers your CMRR and the reason is very simple, why does it lower it? Physically also you can understand because then the common mode signal itself gets differentiated and gets amplified. Therefore the differential amplifier is not able to distinguish between differential mode signal and a common mode signal because both of them look the same to the differential amplifier and therefore it equally amplifies both the signals together, right? And that's a problem area which people face as far as mismatch is concerned.

(Refer Slide Time: 26:14)




So let me recapitulate what we have done till now and let me see what we will learn in the further. That any small signal circuit analysis where we assume that my transconductance is always there for the device which converts a voltage into a current source. We will be able to calculate the differential gain, common mode gain and the common mode rejection ratio this 3 very-very important.

In fact these 2 are very important automatically this is the third one which comes out of the difference between the differential mode and common mode gain. As we saw that CMRR will obviously be a strong function of the output impedance. The reason being higher the output impedance larger will be the differential mode gain and as a result will get a larger CMRR in that sense.

How can I increase it? I can increase the CMRR by increasing the output resistance of the current source which is very true also. More ideal your current source is better your CMRR is because you ensure that your source resistance of the MOS device is infinitely high, which is typically very high and therefore it sort of acts like a virtual ground, right? Virtual ground means though it is grounded in a sense but current is not flowing out of that particular arm.

And we also saw that CMRR can be increased if I do a Cascoded amplifier which means that I Cascade devises top of each other, stacking the devices as a result my overall output impedance increases and therefore my gain also increases or overall gain increases which

means my differential gain increases not at the cost of the common mode gain and as a result my CMRR goes to a very high value, right?

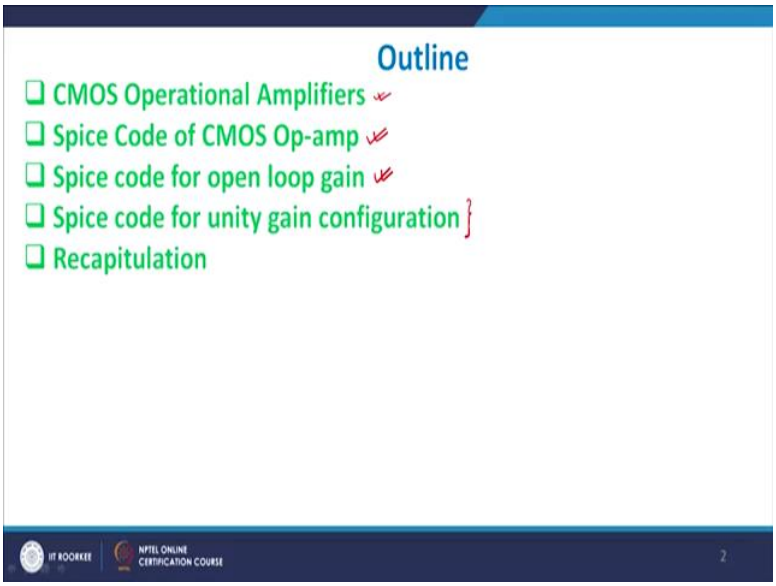
This we have learned through large number of simulation tools, through large number of understanding. We have also understood that this $R_D$ mismatch and $g_m$ mismatches will lower your CMRR. And it will not improve your CMRR and therefore from fabrication point of view one has to always entertain or one has to always see that you always have the perfectly symmetrical arms available to you.

Most of the time not available then my CMRR falls down from a large value to a relatively small value, right? So, lower your mismatches and try to make the gm of the input devices and the output impedances to be too large in order to increase the CMRR and therefore reject the noise, right? With this let me thank you for your patient hearing and we will discuss other topics in the upcoming tutorials and slides, thank you very much.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−39**
**Multistage Amplifier with SPICE Simulation**

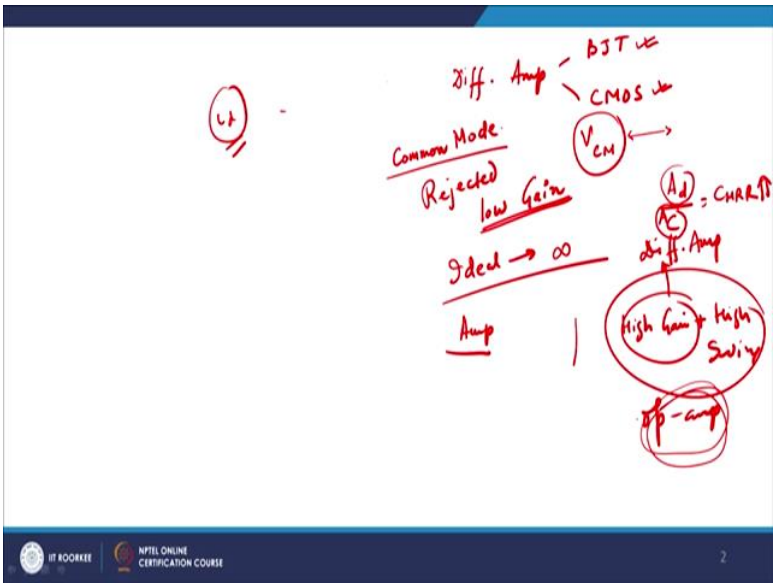
Hello everybody and welcome to the NPTEL online certification course on microelectronics devices to circuits. We start today's module the name of today's module is multistage amplifier with spice simulation. So what we will do in this module of approximately half an hour is, giving you an idea about the structure of a multistage amplifier and then we will look into how to design a spice deck file in order to solve a Multistage Amplifier. And what parameters need to be extracted from the amplifier design. So, let me just recapitulate what we need to do.

(Refer Slide Time: 1:15)




So we will first start with CMOS operational amplifier in the outline of the talk as you can see. And then we will do the spice coding of the CMOS op amp and then we will look into the open loop gain and closed loop gain for operational amplifier and then we will also show you a spice code for unity gain amplifier or unity gain configuration of operational amplifier. Before we go forward therefore let me recapitulate what we did in a previous turn.

(Refer Slide Time: 1:47)




In our previous turn be actually looked into what are known as differential amplifiers, right, both the BJT design as well as CMOS-based design. We also saw that you need to give a common mode signal to both of them initially, so that they are in the active region. And we don't want them to go into saturation region drastically and therefore from active to cut-off and cut off to active should be decision parameter which will be available for BJP as well as for CMOS.

We also saw that common mode signals are, that is a major prime importance of concern, is that all your common mode signals, right? Which is basically noise, for example, common mode signals to board the arms of BJT they get rejected. So all your common mode signals have got very-very low gain, right? And as a result it gets rejected very easily, which means that if you free the noise to both the arms of symmetrically designed differential amplifier the noise will be easily rejected.

Whereas the different signal which we had defined as id or idv as a different signal between the 2 arms of the amplifier they gets amplified drastically. And the ratio which we have already defined earlier differential mode gain divided by common mode game here was referred to as CMRR and this should be as high as possible ideally value is as I discussed with you it should be equal to infinity.

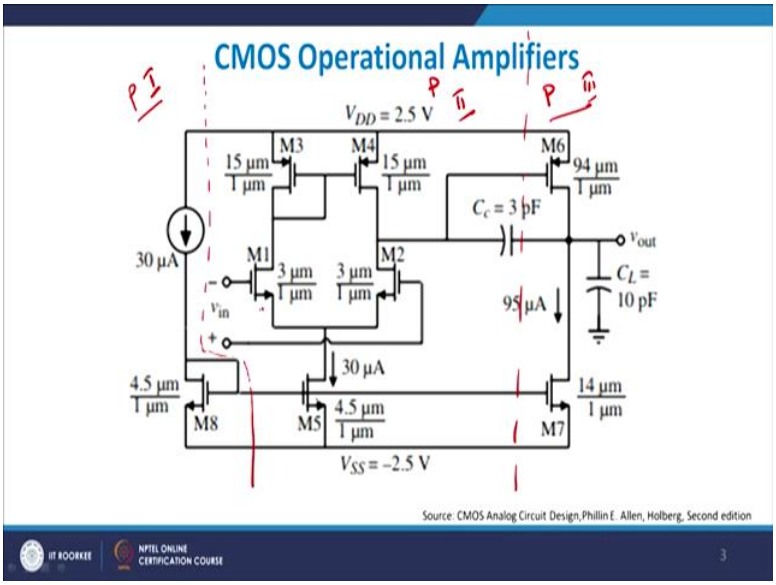
So if you design and operational amplifier or a differential amplifier. Differential amplifier is no doubt the first days of the operational amplifier. So today I'll show you the second stage

of the differential amplifier and I will show how it works out properly as a signal stage op amp. So we will show you the circuit diagram of an operational amplifier when we cascade sort of differential ended operation with a single stage amplifier, right?

We also saw and we also came to the conclusion that whenever you are designing an amplifier, it is always advisable to first of all design an amplifier with let us say high gain, right? And then cascade it with high swing, right? And that is what is basically an op amp, right?

So this will be a differential amplifier dif Amp, so the first age of any operational amplifier is basically a differential amplifier. And the second stage is basically a high swing stage or a high gain stage and that results in basic understanding of the operational amplifier. So this is the basic flow which we have been doing it for quite a long time. And let me therefore come and show it to you the basic diagram of a CMOS operational amplifier.

(Refer Slide Time: 4:46)

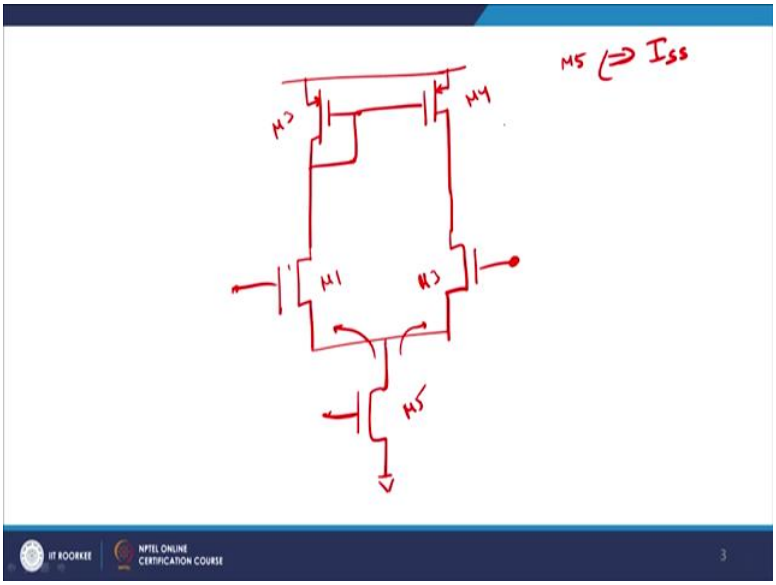



This is the typical diagram of the operational amplifier out of which if you see very carefully, I will just divide the whole thing into few parts, so it is easier for you to gather information for each of them and I am doing it like this. So this is part number 1, right? So this is part number 1, right? And then we have part number 2 here, this is part number 3 and this is part number 2, right?

So you have part 1 here, right? P1 and then you have P2 and you have got P3 here it all respects, you do have your design here. I will first concentrate on P2 and then we will come to P3 and before as we move along. If we look at P2, right? The part 2 which you see here, it

is nothing but differential amplifier with M1 and M2 as input devices and M3 and M4 our basically the cascaded device or the low device and if you look at M5 it is nothing but the current source which is available to you.

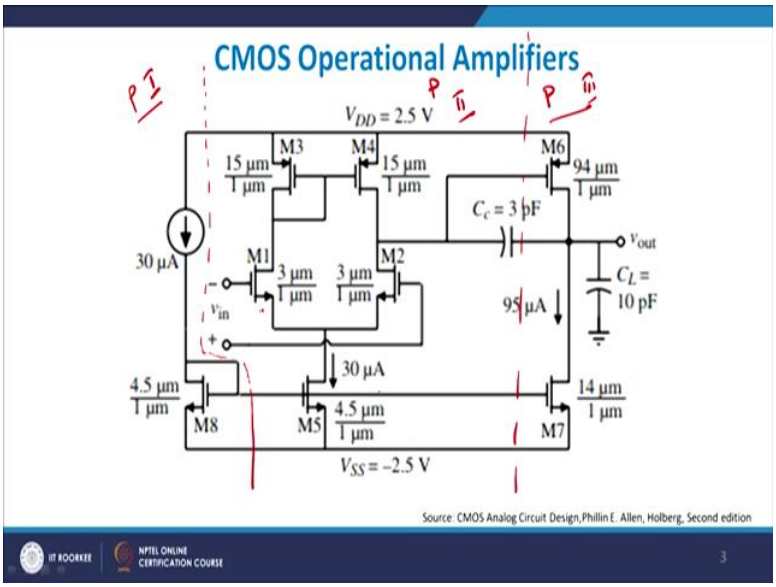
(Refer Slide Time: 5:56)




Now, let me draw you this diagram only in a schematic mode. Any will see how it works out. As you can see therefore it will be something like this. This is again CMOS and we are using PMOS load and this PMOS load is, this is mirrored here, so this is current mirrored and then we have got here 1 NMOS driver another and NMOS device here and then we have got this coming out like this, this, this and this.

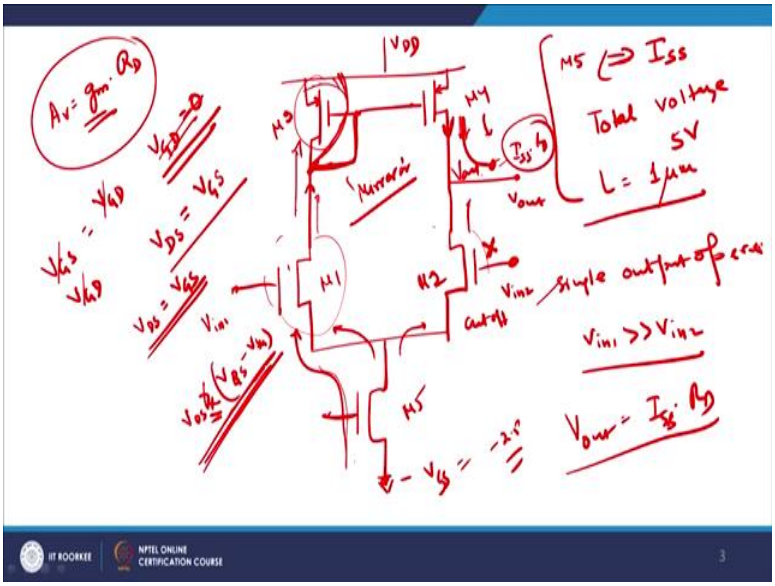
So, this is M1, M2, M3, M4 and M5, right? So this M5 is primarily the ISS. Which means that this is basically the tail current source, right? And this is responsible for giving the current to both this arm as well as this, right? And you have to make this M5 is analogous to your ISS or the current source.

(Refer Slide Time: 07:02)




And therefore if you see its aspect ratio is very large, you see 4.5 by 1 micron. So this is design is based on 1 micron technology which means that channel length of all the transistors used here is approximately 1 micron. And the ratio if you look at W by L ratios of M5 is typically very high, right? It is 4.5 micrometer by one micrometer, right? And your VSS is latched to minus 2.5 volts, so it is not grounded here, if you go back.

(Refer Slide Time: 7:31)




It is not grounded here but this is VDD which is plus 2.5 and we have got minus VSS which is basically minus 2.5 and therefore the total swing, the total voltage available, swing available to me is basically equals to 5 volts, right? And this is L equals to 1 micron

technology which people have been using in this case. I will give you the operation and then we can go step-by-step, one step ahead.

See, if you remember that the gain was given as equals to gm times are RD, right? Where gm was a trans conductance of this M1 and M3. Assuming M1 and M3 are equivalent in nature and also assuming that they have been properly biased in the active region I can safely write down the overall gain to be equals to gm times RD where gm is the trans conductance of this device as well as this device.

And RD is basically the resistance offered by M3 and M4 low devices, right? So that is the reason you typically like to have these PMOS devices because PMOS have got a whole as the charge carriers and therefore, its mobility is low and therefore its current is small which results in a higher resistance. Higher resistance means obviously a higher gain in the output side.

So the output has been taken from this place, if you see very carefully, the output is taken as a single ended output, right? So this is output which you see. So if you look very carefully this is the output here V out. So this is also known as a single output operation. Why single output? Because you're only taking output from the single end they are not taking from the double end, right? So its gain will be approximately half.

Let us see how it works out? Your V in 1 is given here and this is say M1 M2 and this is V in 2 and let us suppose V in 1 is much-much larger sum but to V in 2 and therefore V in 2 is cut off, right? At this stage and all the current flows through this arm and it goes to M3, right? So what am trying to say is that V in 1 is typically very large much larger than a threshold voltage of M1.

Therefore, what it does is, it carries all the current through M5, M1 on to M3. Now in M3 your gate and drain our short head. So your $V_{GD}$ is equals to 0. Gate to drain is always equals to 0. Which means that $V_{GS}$ gate to source this to this, right? Is this one and gate to drain are both equal to each other, so $V_{GS}$ equals to $V_{CD}$, right? $V_{GS}$ gate to source is equals to gate to drain, right?

Because Gate to drain is actually equals to 0, gate to drain equals to 0. I will just make some small corrections here. See here is that your drain voltage and your gate voltage are equal, so $V_G$ and $V_D$ are equal. So if you put source here $V_{DS}$, right? And you put $V_{GS}$, so they are

always equal. So your drain to source voltage is always equal to gate to source voltage implying that the device is always into saturation region of operation.

Because if you remember the region of operation was VDA should be greater than equals to $V_{GS}$ minus $V_{TH}$, right? That was what was told to you, it should be less than or equal to. For PMOS it should be less than equal to but since it is always equal to therefore this will always be in saturated state.

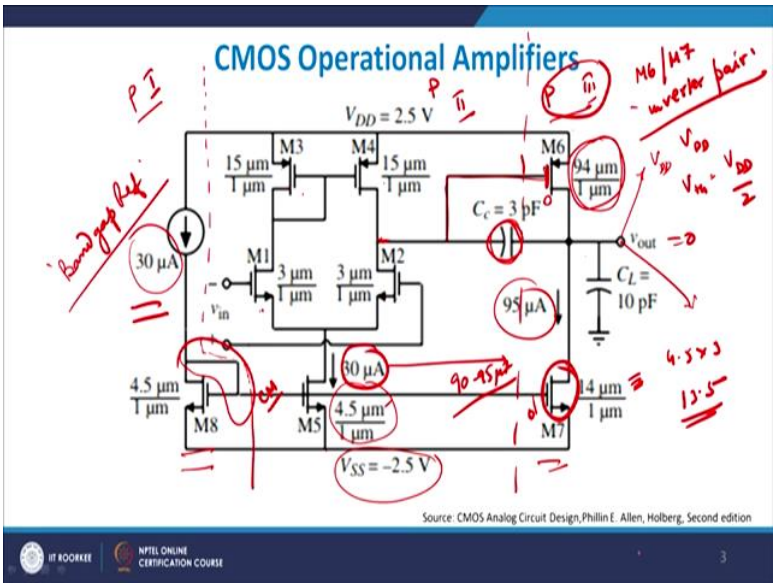
So they will always behave like a resistor if you shot this thing gate to drain, if you short the gate to drain. Further you also ensure that this potential, M4 gate potential is exactly equal to the drain potential of M3, right? And therefore this is sort of replica here, so all the current which is flowing from here the exactly same current will be flowing through this arm, right? This is known as the mirroring action in analog design.

This is known as mirroring action at the result, if you remember for the current to be saved in 2 arms of a 2 identical MOSFET gate to source voltage should be equal first of all threshold voltage should also be equal. And you also have to ensure that if it is in the priorities of operation the drain to source voltage should also be equal, right? So all these 3 are ensure to be equal for both M3 and M4, if you simply shot gate and source of M3, right?

And as a result the same current is flowing from this are exactly the same current flows through this arm and goes to V out because this is cut off, so the same current flows through $V_{out}$, right? And therefore whatever current was flowing through this ISS is coming out here this multiplied by $R_D$ is $V_{out}$ so therefore you get $V_{out}$ to be equals to $I_{SS}$ times $R_D$ in a general sense.

And as a result that is the voltage drop which you see as equals to V out, right? And that is what is $V_{out}$ is all about. So now if you go on increasing the value of your M2 gate voltage and lowering your V in 1 then part of $I_{SS}$ will be transferred to M2 and therefore you will automatically, so this will be less than $I_{SS}$ and this will be slightly greater than 0 and if you add, sum of them will always be equal to $I_{SS}$, total current source.

(Refer Slide Time: 13:12)




Now that is the general scheme of things which you see now. You do have a capacitor here which is basically a role of a blocking capacitor because it does not let any DC component to pass through it and reach to this particular point it only reads the A/C component to pass through it because you don't want any DC component to load your subsequent stage. The subsequent stage now is basically your P3, right?

We are finished with P2, this is P2 which we have done, so this M3 was mirroring with M4 and the current was going through this point as a result it is going through the same current multiplied by RD is the voltage here, voltage appears across the PMOS here, right? Now if you look very carefully this is basically a pseudo NMOS inverter and therefore part 3 is basically a high swing stage.

I will give you an example why? Say for example, if you look very carefully M6 & M7 forms an inverter, right? Forms an inverter pair which results in what? Which results in the fact that if you're using a $V_{DD}$ here then switching threshold will be approximately equals to $V_{DD}$ by 2, so if the voltage here just crosses $V_{DD}$ by 2 this will ensure $V_{out}$ to go to latch to $V_{DD}$ and then it is just below $V_{DD}$ this will ensure it goes to some other voltage.

You see you are feeding to the gate and of my device, right? So when this voltage is 0, this is on and therefore this output voltage is latched to $V_{DD}$ and the $V_{out}$ goes to 0. When this goes to one and this goes to, say this is one that this goes to 0 and therefore this output goes to 0

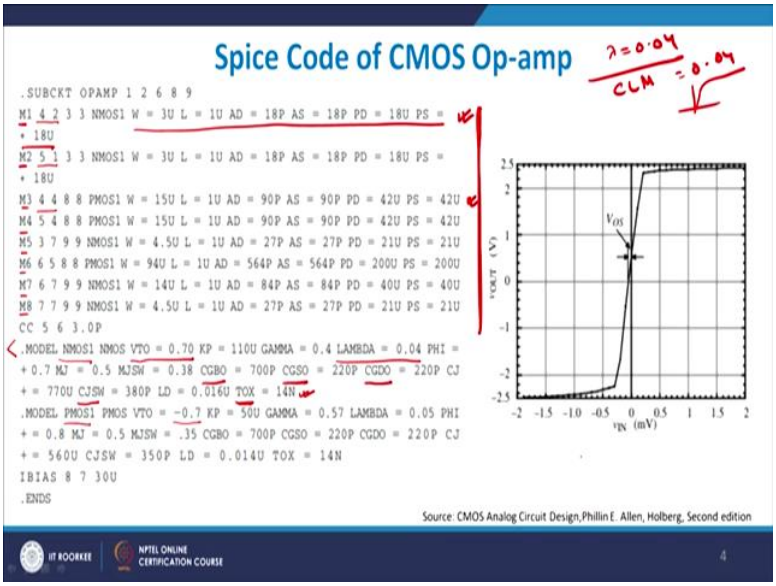
when this M7 is 1, right? And as a result you will have a large amount of current flowing through this particular point.

You see M6 has been made very-very large and I mentioned, right? This is to ensure that your typical resistances are small and voltage across this is very-very small and it really get high-end $V_{DD}$ full swing available to you in the output state. Who drives M7 ? It is being driven by this M8. Look at the fact that this M8 it is M7 and you have a current source 30 micron ampere this is also referred to as band gap reference we will discuss that particular point.

Assume that band gap Reference is basically current source which is basically temperature independent and it is a very fixed current source which is given to operation amplifier the same current source is getting replicated at every other point, so if this is 30, this is also 30 because you have a current mirroring action here, this is your CM current mirroring action here and as a result you have 30 microamperes here and approximately the same here provided but it is approximately 2.5 times, this is 4.5 and this is 14, right?

And therefore, what you will see, you expect to see a large increase in the current here and that the reason you see this to be as equals to 90 microamperes the reason being this approximately 3 times which you see, right? Now, for example, 4.5 into 3 if you do I get 13.5 which is approximately 14 micron which means 30 microamperes will appear as 90 to 95 microamperes at this particular point and that the reason you see a large amount of current flowing through here. This is the general structure of a CMOS operational amplifier,

(Refer Slide Time: 16:49)



Let me come to the spice code of a CMOS operational amplifier. Spice code is basically a code written in spice language. We will come to spice we will have one module on spice later on but at this stage the circuit in front of you gives you an basic idea about the sub circuit of the operational amplifier and you can see here basically M1, M2, M3, M4 are the entities which you are naming here and these are basically the transistors which are being named here.

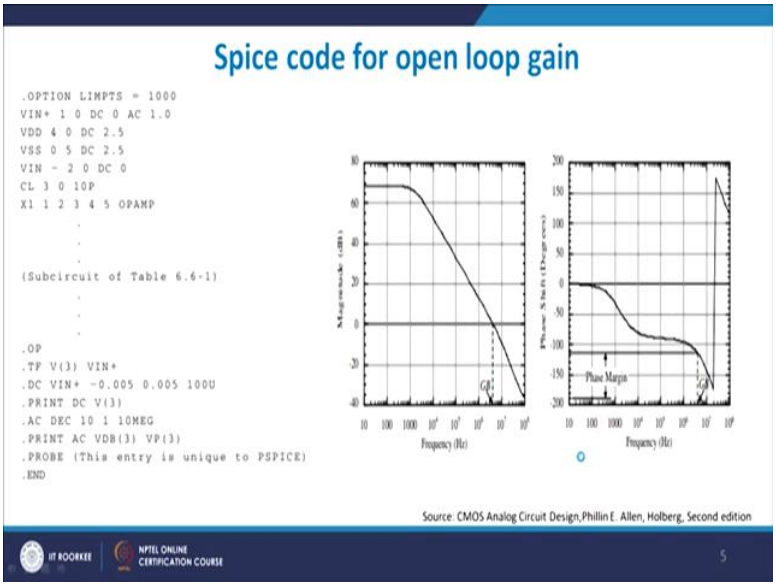
And these are the nodes across which they are connected to each other. So 4, 2, 5, 1 and so on and so forth. This is width, this is length, this is the area of drain, area of the source then you have got the parameter of the drain and parameter of the source. These are the 5 quantities which are written for each one of them and these are basically process dependent parameters which are inserted onto the system by the spice module developer.

So you do have certain sub circuits for all of them and you have large number of, for all the devices we first of all defined its structure and how it looks like and across which to nodes they are connected then we define a dot model file which gave you an idea about the type of device it is. For example VT0 is basically the threshold voltage of NMOS is been taken to be equals to 0.70 voltage, right?

0.7 volt is the threshold voltage of NMOS into consideration. We are also assuming this is not an ideal current source but a non-ideal current source with lambda equals to 0.04, so this is 0.04 primarily meaning that CLM parameter is basically 0.04 and therefore when you got a graph it will be something like this. It will not be exactly the straight line across Y axis. Similarly this is the capacitance of gate to bulk, gate to source and gate to drain in Pico farads, right?

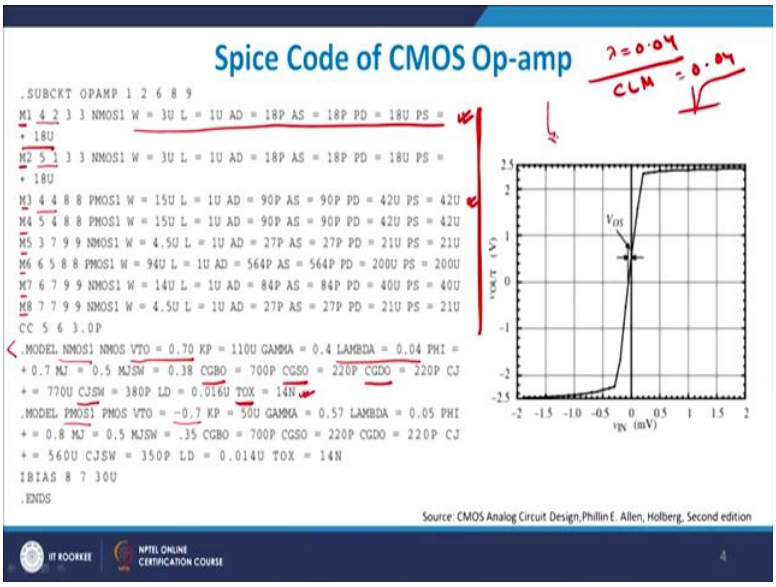
We have junction capacitances here and we have got oxide thickness as 14 nanometers and so on and so forth. So these are the model files for NMOS 1 and these are the model files for PMOS 1. So PMOS one is exactly minus 0.07, so therefore fairly symmetrical with respect to each other but everything else almost remains the same, right? And this is how you write your spice code file for the CMOS operational amplifier.

(Refer Slide Time: 19:21)



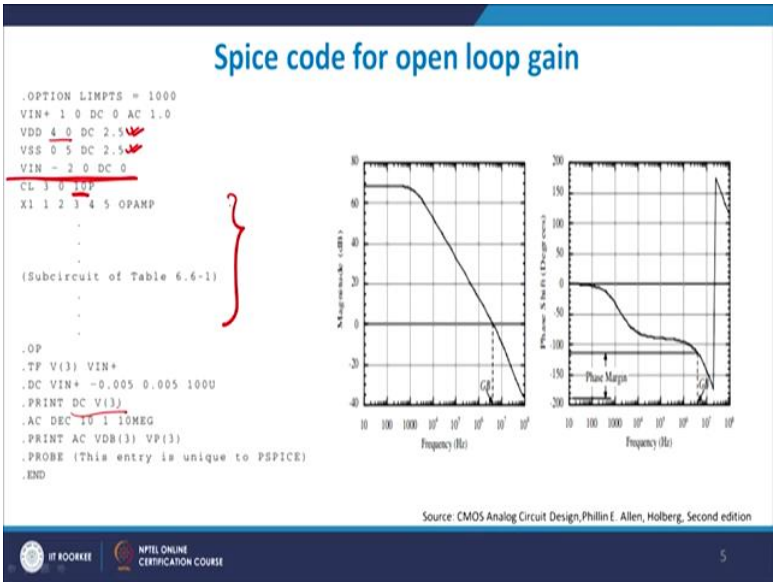

Now when you want to find out the open loop gain, right?

(Refer Slide Time: 19:30)



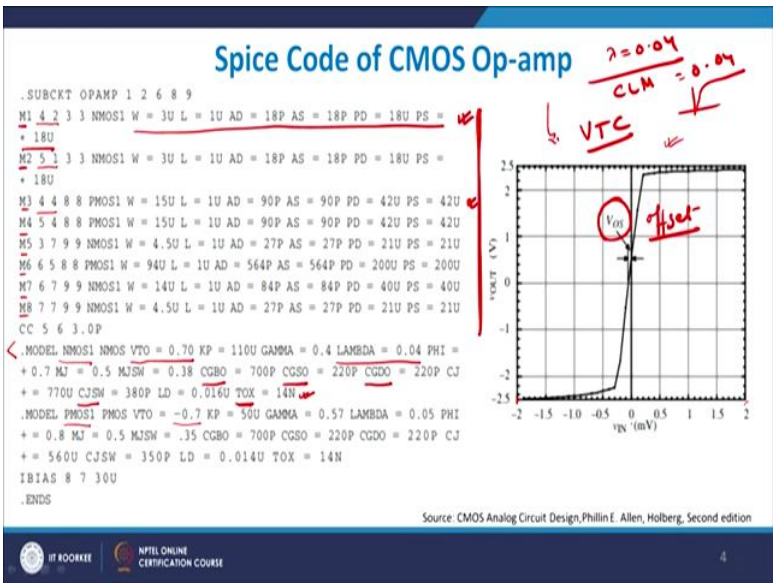

So, let me come back to this point basically if you look at graph here this is V out versus V in which is basically also referred to as VTC, so this is basically the VTC often operational amplifier. So you give an input swing from minus 0.2 to plus 0.2, right?

(Refer Slide Time: 19:43)




And that's what is given here. So if you see at $V_{in}$, if you look at $V_{in}$ then you see that I am giving VSS to be equals to 2.5, right? It cannot be between node 4 and 0 and DC bias and $V_{DD}$ is 2.5 which am giving here $V_{in}$ is going from minus 2 volts, right? Minus 2 volts to 0 volts, right? And I'm able to get the picture here then I define load capacitance as 10 Pico farad between note 3 and 0 and so on and so forth. So what happens is that, you just read the circuit here and then you try to find out the value of the DC bias or the DC output bias at this particular point.

(Refer Slide Time: 20:35)



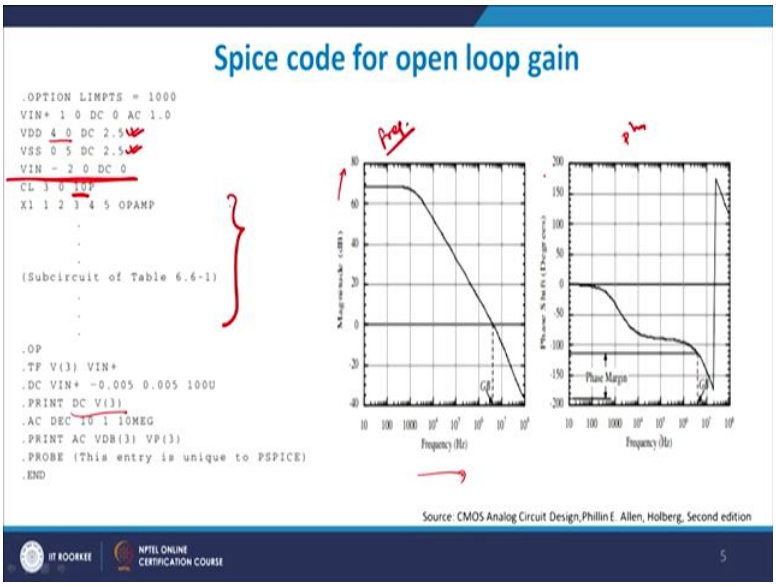

Then what you get from this discussion is that you try to find out $V_{out}$ versus Vin which means that you give a sweep of $V_{in}$ for every value of $V_{in}$ you try to find out the $V_{out}$ and as

expected the VTC looks something like this which is shown in front of you, so when $V_{in}$ is very-very low the output is 0, right? And it is very-very low when $V_{in}$ is very-very high one of the $V_{in}$ is very-very high you get output to be very high, right?

And this is we have already discussed this point in our earlier discussions of differential amplifier. But remember that when $V_{in}$ equals to 0 the difference between them should be equal to 0 then output also should be equal to 0 but in reality not true. So therefore you always have an offset with you, that the reason I write here $V_{OS}$.

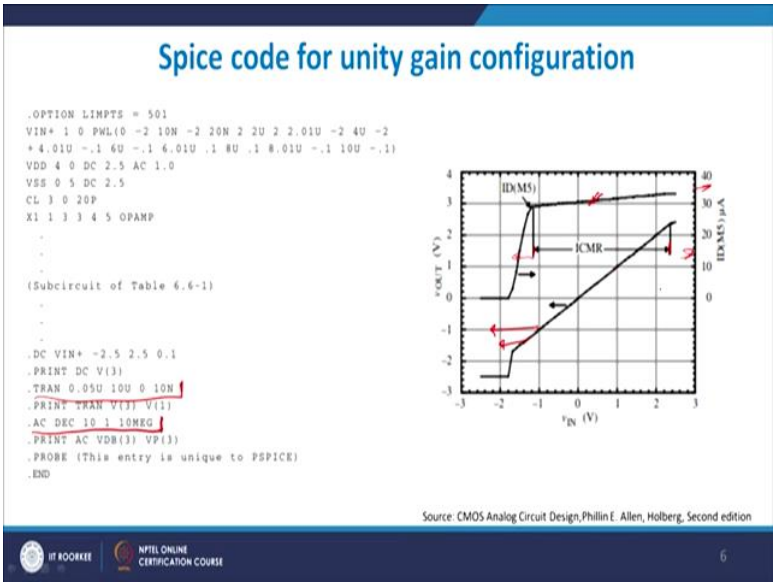
$V_{OS}$ basically the offset. This is basically your offset. Voltage offset at corresponding to Vin equals to 0, right? This offset does what? You have to subtract this offset later on into your system. If it's a positive offset you need to subtract it from the output and if it is a negative offset you need to add it to the output to get the achievable signal.

(Refer Slide Time: 21:46)




So we plot a magnitude which is basically gain versus frequency. So this is your frequency spectrum which you do, right? And then you also plot phase versus frequency also using spice code.

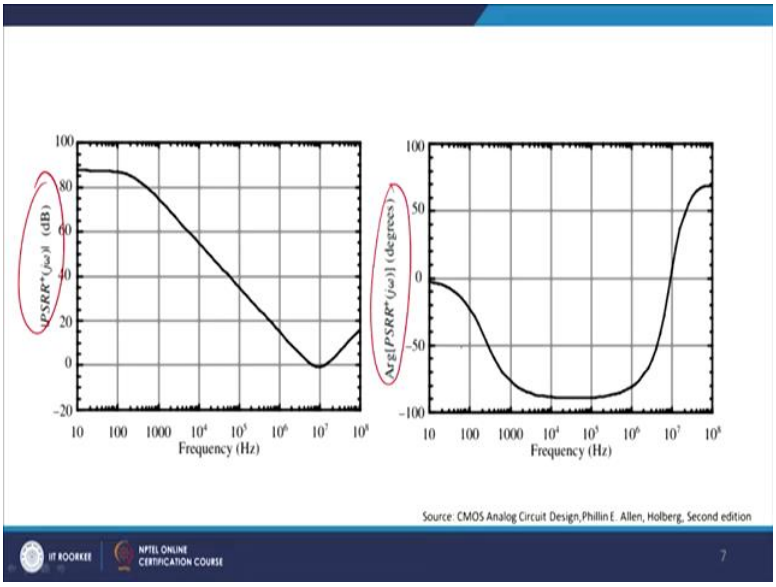
(Refer Slide Time: 21:59)




If you want to do it for unity gain configuration. Unity gain basically means that you just short the input and output, right? So you have a unity gain feedback loop and therefore that gives you a unity gain output. You get a $V_{DD}$ as I have discussed with you again $V_{DD}$ but here we use dot AC because you are giving an AC cycle and across the 10 mega ohm resistance you are trying to do it and then you are trying to find out the transient value of the voltages at every particular point.

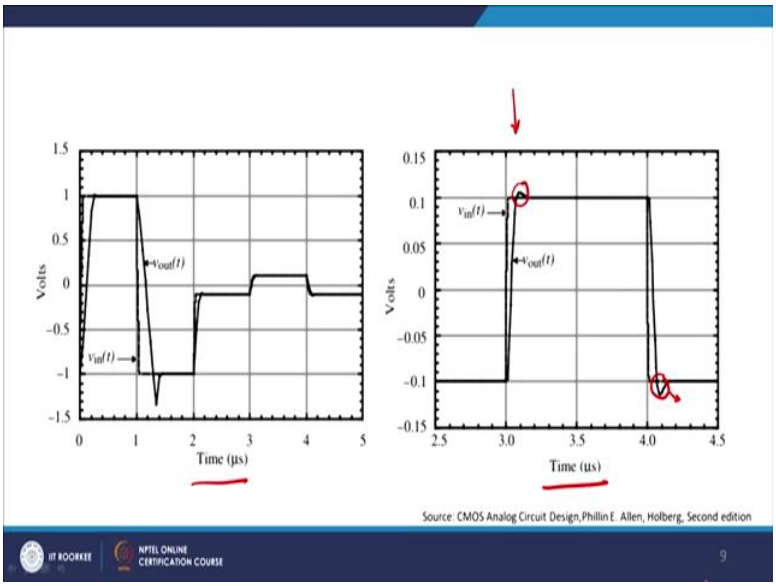
So, when I find out $V_{out}$ versus $V_{in}$ I see that, this is basically your $I_D$ which is basically on this side and this one is on this side, right? So, I see that this is basically my ICMR from this part to this part is my ICMR input common mode range, right? And anything lower than this you force the device to be into the deep linear region and if anything greater than this you allow it to be going to saturation region, right? And this is the spice code for the unity gain buffer which you see.

(Refer Slide Time: 23:11)




Then we plot the same plot for PSRR and for your phase of PSRR this is power supply rejection ratio, we will come to this later on but very important part is that this power supply rejection ratio is basically the capability of the operation amplifier to remove noise from the power supply. And as we can see at lower frequencies the gain is typically very high it can remove very well but as the frequency goes on increasing there is a drop in the output in this case. Similarly, for low frequency we get the same thing.

(Refer Slide Time: 23:50)



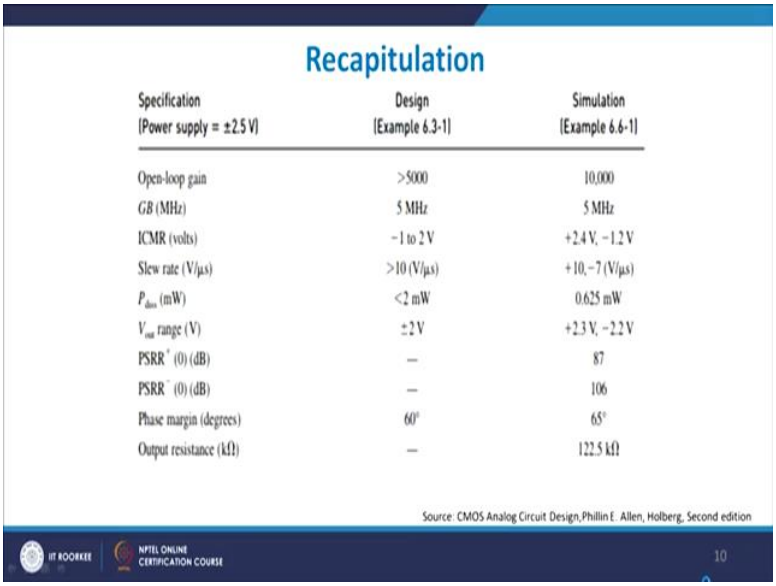

We come to the last part and that is basically transient analysis and we see that for various different time analysis, how a timing analysis? How we get the output voltage with us and the methodology adopted here is that again we do a transient analysis. So rather than doing a DC

analysis and we do a transient analysis. In the transient analysis for a very short duration of time we try to find out the output voltage changes because of the change in input voltage.

And you will see that when my $V_{in}$ rises my $V_{out}$ also rises, this diagram. But it arises with certain extra delay which is expected also because that is basically the delay of the inverter itself. Not only that but one important point is that you see a small hump here and a small hump here. This hump is basically to do with the clock feed through which means that because of very fast changing output there is a capacitive coupling between the output and the $V_{DD}$ and as a result you will certainly see a small hump in the output.

Same thing happens in the lower transition when the output goes below particular point which is minus 0.1. So even if you are $V_{DD}$ is 0.1 to minus 0.1 for a very short duration of time the output voltage can actually exceed 0.1 and go below minus 0.1, right? And that is because of the fact that you do have a capacity of coupling between the input and output which results in a sort of clock feed through mechanism in this case, right? And this results in a larger change.

(Refer Slide Time: 25:28)




Let me therefore recapitulate to you and show you overall picture of the operational amplifier which we have designed till now. This example, we have taken it from a standard operating manual. And we have found out that two-stage op amp your typical design we have taken out to be 5000. The same design principles if we put it on a spice profile we get approximately 10,000 as my simulation output.

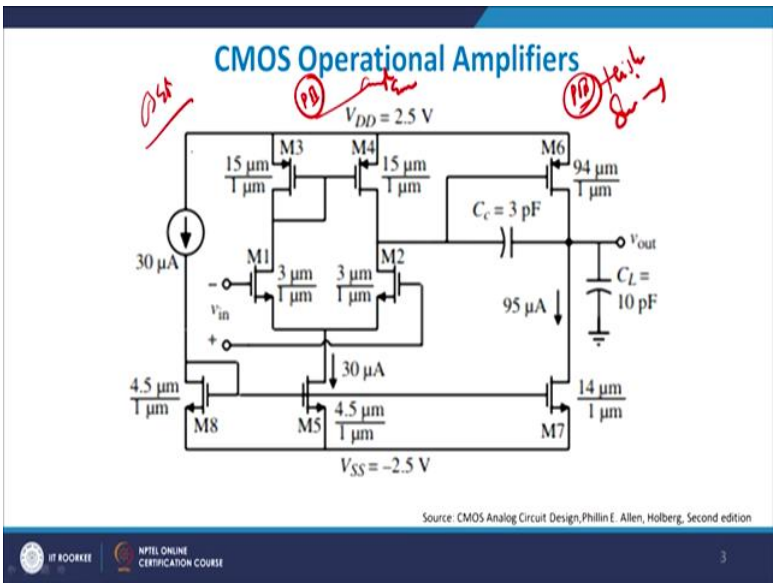
This is the gain bandwidth product of 5 megahertz and we also get 5 megahertz here. ICMR is minus 1 to 2 worlds and it is plus 2.4 to minus 1.2 volts, so it is almost the same thing. The slew rate are also equivalent. The power dissipation is typically very small. My output range is plus minus 2 volts for design. I was planning for plus 1 volt to minus 1 world variation in output, right?

PSRR and phase noise margins and output resistances are also within the domain which is available to us. Now, the idea here was that to just give you a brief inside into spice, right, and how you can use the spice for multistage operational amplifier calculation. They will also come back to this spice module in details in maybe about 6 to 7 modules later when we discuss spice.

Then again I would recommend that you come back to this module once again, right? So that once you study the spice properly you can come back once again to this spice module for multistage amplifier, right? How to write a net least? How to write a schematic entry in a spice?

All will be told to you in subsequent modules but at this stage I just showed to you that it is possible using simple spice tool to actually extract the values of gain, bandwidth, CMMR, Slew rate, PSSR, open loop gain bandwidth product for a simple two-stage operational amplifier. So what we discussed here was basically your two-stage operational amplifier only.

(Refer Slide Time: 27:48)



If you look very carefully here what we discussed was basically a two-stage operational amplifier. And as you can see here this two-stage operational amplifier was working because of 2 main reasons that you had part 2 which is basically, primarily the differential amplifier part which gives you a very high gain part and you have got part 3 which helps you to give you high swing part, right?

So you have an inverter which actually therefore pulls your output voltage directly to $V_{DD}$ up or it goes down to $V_{SS}$ at a bottom low. So part 3 which is basically this one helps you to go for high swings, right? High swings and this is high gain. This is your band gap reference which you see BGR and this is your $V_{SS}$ which is minus 2.5 volts supply which you give at this stage, right?

So what we did in this half an hour module was to give you an idea about the working principles of a two-stage operational amplifier where the first stage is differential and the second one is high swing rate. We also saw the spice coding how to do it, a methodology adopted.

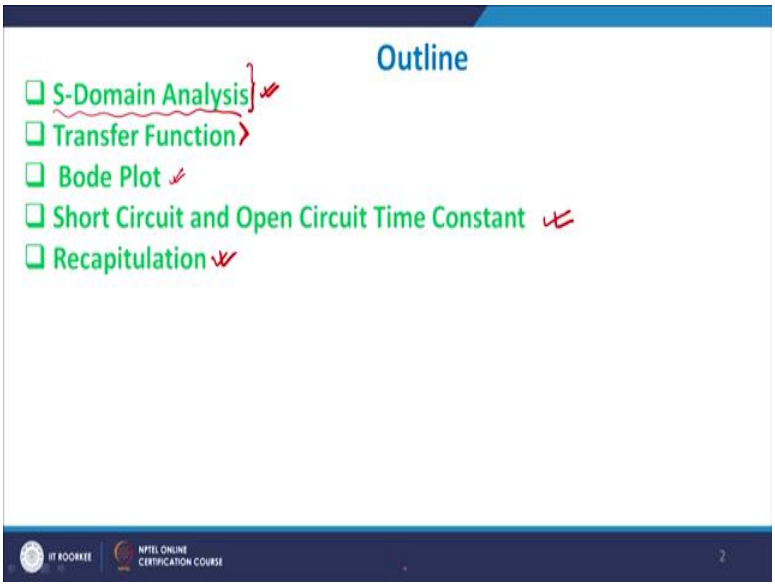
We will revert back to this module once again when once you finish spice module later on and then we recapitulate it by doing comparison between a simulation results and the spice results, right? Simulation spice results and handwork results together. And we saw they are very close to each other at certain point of time. I thank you for your patient hearing, thanks once again.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−40**
**S-Domain Analysis, Transfer Function, Poles and Zeros-I**

Hello everybody and welcome to the next edition of NPTEL Online Certificate Course on Microelectronics: Devices to Circuits. In our previous module, we had looked into simple multistage amplifier, differential amplifier based operational amplifier and we also saw the methodology for extracting various parameters of the OP-AMP.

In today's lecture we slightly change gears and we try to do what is known as S-Domain Analysis. We will start with S-Domain Analysis today and we will show to you that how do you actually do a frequency response which means that with varying frequency how does the gain of an amplifier changes. We will also look into the concepts of poles and zeros and what is a transfer function.

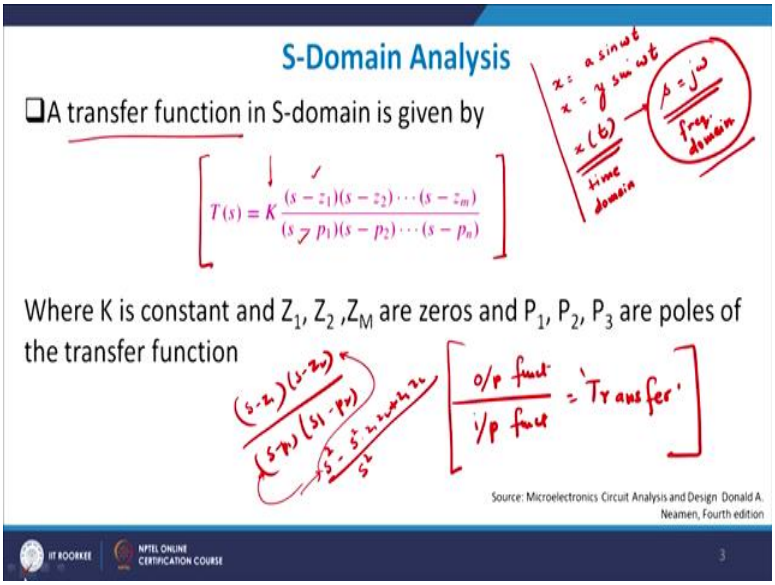
(Refer Slide Time: 01:48)




So, the lecture module is named as S-Domain Analysis, Transfer function and Poles and Zeros. So, these 3 we will be covering maybe by this module and then we will move along and we will see as we go ahead. So, the outline of the stock is something like this that we will start with first of all, we will explain to you later on maybe what is S-Domain, though I have written it here at the very first instance but we will be using this later on.

So, we will be using this later on, right. So, we will be using the concept of S-Domain Analysis later on and we will finally come to, we will first initiate our discussion with Transfer Function, explain to you what is the Bode Plot and then look into what is a Short Circuit and Open circuit Time Constant and we recapitulate our work. So, this is the general scheme of things which will be following as far as this course is concerned.

(Refer Slide Time: 02:29)




Now, when we say in S-Domain, typically most of the input signals which you will be encountering will be in time domain for example let us say x equals to a sin of omega t or let us say x equals to y sin omega t. Which means that x is basically a function of time t. So, this is in time domain. Now, there was a requirement, we will not underline at this stage not why but to convert this time domain into frequency domain, right.

So, we define s to be equals to j omega and this is what we define as frequency domain. So, this is known as the time domain analysis and we have got here a frequency domain analysis. So, in most of the cases we will be actually looking into frequency domain analysis and then when time permits we will go to time domain analysis. But, at this stage we are not looking into the methodology by which you can convert a time domain into frequency domain and vice versa.

We are assuming that we do already have in S-Domain which is in the frequency domain we do have the certain characteristics available and from there I can find out the values of this thing, various characteristic of the device of circuit. We define a new term here which is

known as Transfer Function. Transfer Function as the name suggests is a very simple and basic pack. It is basically output function divided by input.

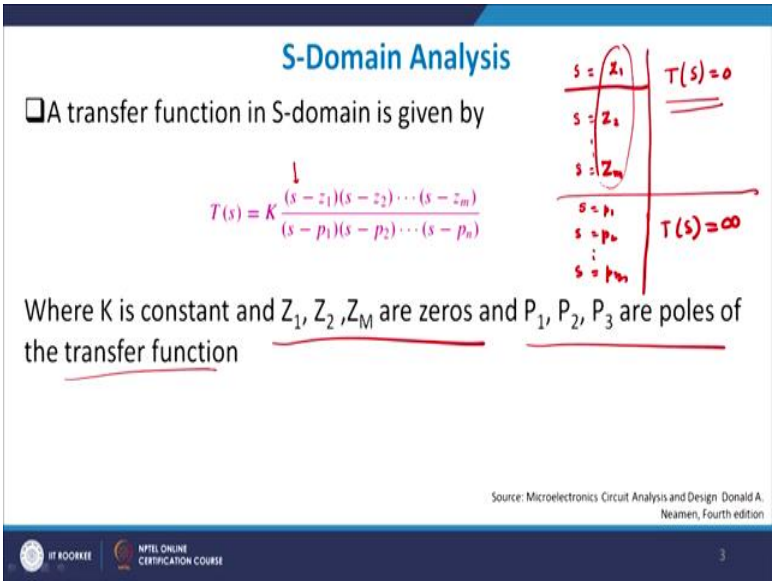
So, this is known as Transfer Function. As the name suggests, it is transferring from output to input. So, we define this to be an output function divided by input function. We define output function, output can be current voltage, input can also be current voltage. So, therefore, I can have sort of 4 types Transfer Function available to me. Because, there are 2 inputs and there are 2 outputs, so total about to 4.

I can have 4 combinations of transfer function available (to us) to me in a much more detailed manner. But, before we move forward, let me show to you that what is the meaning of poles and zeros before we move forward. I will be keeping to bare minimum understanding so that we can concentrate more on the circuit aspects of it. For understanding more of poles and zeros, I would refer that you go for a basic course of the network theory and synthesis part of the network theory and there this will be discussed in a detailed manner.

Look at this Transfer Function, if you look Ts is given as K times, in the numerator you have s minus $Z_1$, s minus $Z_2$, s minus zm divide by s minus $p_1$, s minus $p_2$, s minus pn. Now, we all know that the output current or voltage or input current or voltage can be expressed in a linear relationships and this can be there for factorized as given as this and this, right. This K is basically constant, which is basically an s, it does not depend upon the frequency, it's basically an integral constant and it is kept outside here, right.

It is also referred to as gain for most practical purposes. I guess, therefore s minus $Z_1$, s minus $Z_2$, so if you for e.g. if this would have been s minus $Z_1$ and then s minus $Z_2$ divided by s minus $p_1$ into s minus $p_2$, then you open it up, I will get s square minus square $Z_1Z_2$ plus $Z_1Z_2$ divided by s square, so on and so forth. So, this is basically a quadrative equation here and I can remove the quadrative equation in form of partial factors and these are the partial products of each one of them, right.

(Refer Slide Time: 06:36)



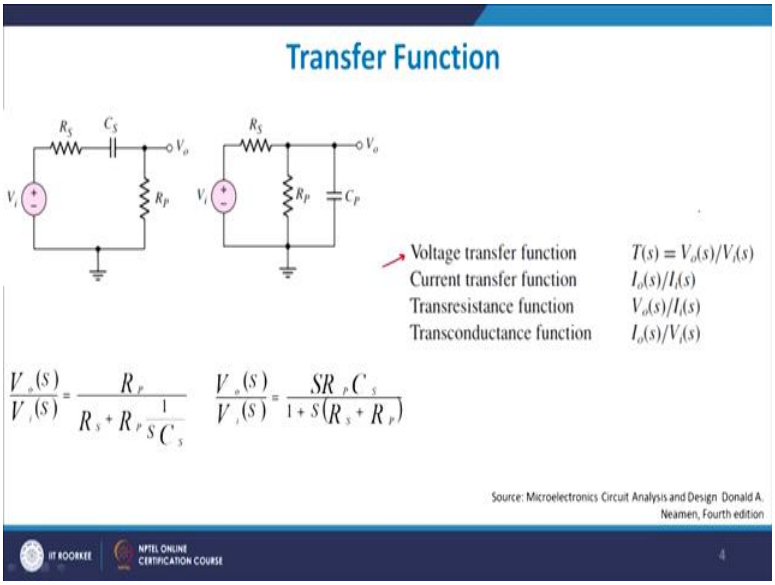

Therefore, it gives me an idea that, I can explain to you from this basis equation itself that if, let us suppose my s happens to be equal to $Z_1$ let us suppose. Then what happens is that, (this z minus) s is equal to $Z_1$, so z minus $Z_1$ is 0 and all of T(s) is equals to 0. Similarly, when s equals to $Z_2$ or till s equals to zm for all the cases I get T(s) equals to 0, which means that in the numerator if the solution of one of the points of the numerator happens to be either $Z_1$ to zm then the whole function actually goes to 0.

Similarly, in the denominator if each of the function s goes to $p_1$, s goes to $p_2$ and so on s goes to pm then the output T(s) goes to infinity, right. So, these are the points, $Z_1$,$Z_2$ till zm which are referred to as 0 are the points where if you are biasing your device in frequency domain at those particular points, the gain will be exactly equals to 0 whereas, if you are biasing your points at $p_1$, $p_2$, pm, then your gains will be infinitely large.

So, if you biased your amplifier whose transfer function is given by this equation which is shown in front of you, then, and if you biased it at $p_1$, $p_2$, $p_3$ to pm, I would expect to see a maximum gain and if you are able to biased it at $Z_1$, $Z_2$, $Z_3$, $Z_4$ then we define that the gain will be equals to 0. So, therefore we define this to be as poles and zeros, right.

Poles are the points in frequency domain where your gain becomes infinitely large and zeros are the points in the frequency response where the gain is approximately equals to 0 or exactly equals to 0. Therefore, $Z_1$ and $Z_2$ to zm are defined as zeros and $p_1$,$p_2$ and $p_3$ are referred to as poles of the transfer function. K is the constant which is independent of frequency. So, this is the basic understanding of this thing.

(Refer Slide Time: 09:05)

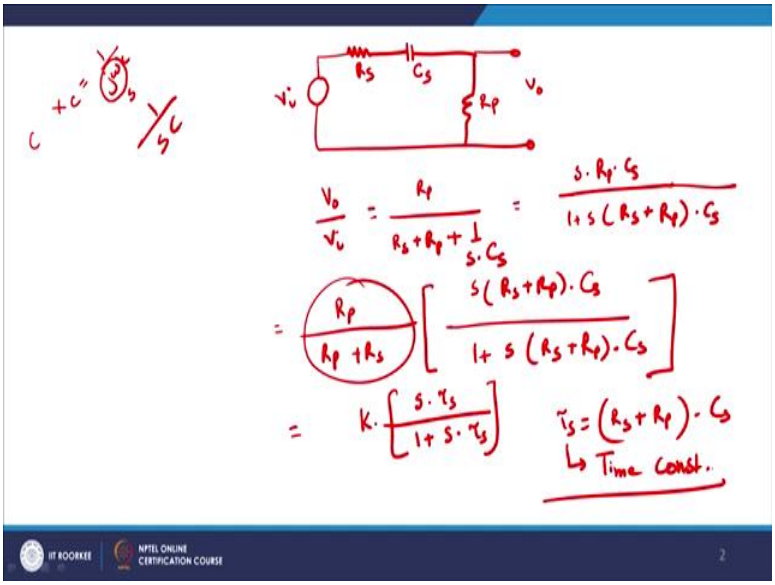



Therefore, as I discussed with you just now, since there are 2 quantities (current and voltages) and their 2 ports (input and output) I can have therefore about 4 types of Transfer Function. The first Transfer Function is very simple, Voltage transfer which refers to as output voltage by input voltage. So, $V_0(s)$ or Vo(s) divided by Vi(s) is basically giving meaning that output voltage by input voltage.

If you look at the Current transfer function here then it is basically output current by input current, right. If you are doing a Transresistance function then output voltage by input current and transconductance if you see then output current by input voltage, right. So, the first two are dimensionless quantities and (the next two are) these two are basically your dimensionless and this is of the order of Ohms and this is 1 upon Ohms.

So, we have got semens here. So, we have got therefore 1, 2, 3, 4 types of Transfer Function available to us in a distinct sense of it. If we give you a brief idea about. Let us say for e.g. How to derive a Transfer Function? So given a circuit, given any circuit how can you derive the Transfer Function. I will just show you one example for that using basic this thing, we can actually go ahead and show it to you.

(Refer Slide Time: 10:37)

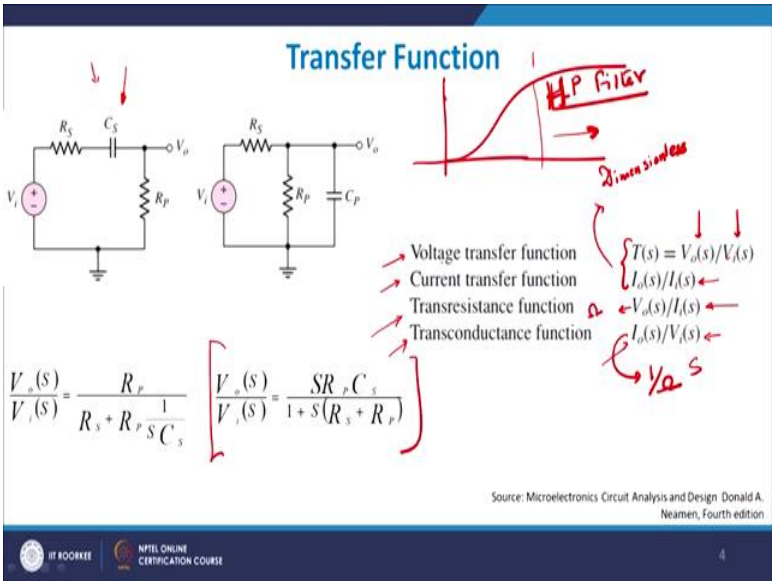


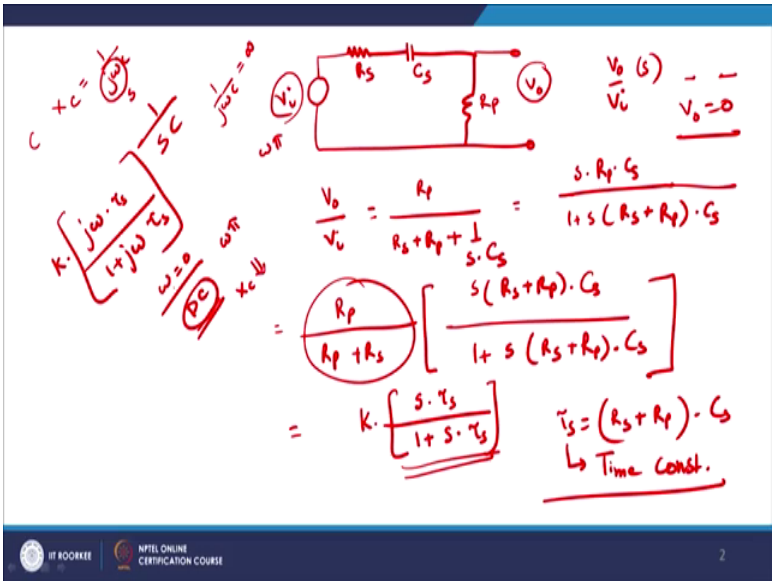

For example we will take the same circuit as this one and let us see how it works out. Let me just draw for you. I have got this and then we have got Rs here, then we have $C_s$ here and then we have $R_p$ here and then we have $V_{out}$ coming like this. So, this is $R_p$, we have $V_i$ here. So, if you plot $V_0$ by $V_i$, I get $R_p$ upon, because it is a core potential divide sort of technique plus 1 by SCs. SCs basically meaning that, if you remember (you will) the capacitance, any capacitance see will have its capacitive reactance to be given as 1 upon jwc, right.

So, this jw is replaced by s, so I get 1 upon SC, so that is the reason I get 1 upon SC here in this manner, right. And therefore, you can write down this to be as equal to S into $R_p$ into $C_s$ divided by 1 plus s times $R_s$ plus $R_p$ into $C_s$, right. And then if you break it down, I get $R_p$ upon $R_p$ plus $R_s$, right. And then if you write it down I get S upon Rs plus $R_p$, right, into $C_s$ divided by 1 plus S times $R_s$ plus $R_p$ times $C_s$, right.

I get this into consideration into all these things and therefore I can write down this. So, this basically a K which is basically a constant independent of frequency. I get S times Tow S into 1 plus S times Tow S, right. Where Tow S is given as $R_s$ plus $R_p$ multiplied by $C_s$ also referred to as time constant of the circuit. So, I get K times S Tow S upon 1 plus S Tow S and circuit.

(Refer Slide Time: 16:35)




Now, if you come back to this, from this explanation as you can see, this is what I got also in the previous discussion just now. And you will see that, sorry just a minute, so I get yes, this is what I got, (I got) so if you see if I replace this S by j omega I get K times j omega Tow S upon 1 plus j omega Tow S. You might be asking the reason, why not eliminate 1 because 1 will be very small compared to j omega Tow S.

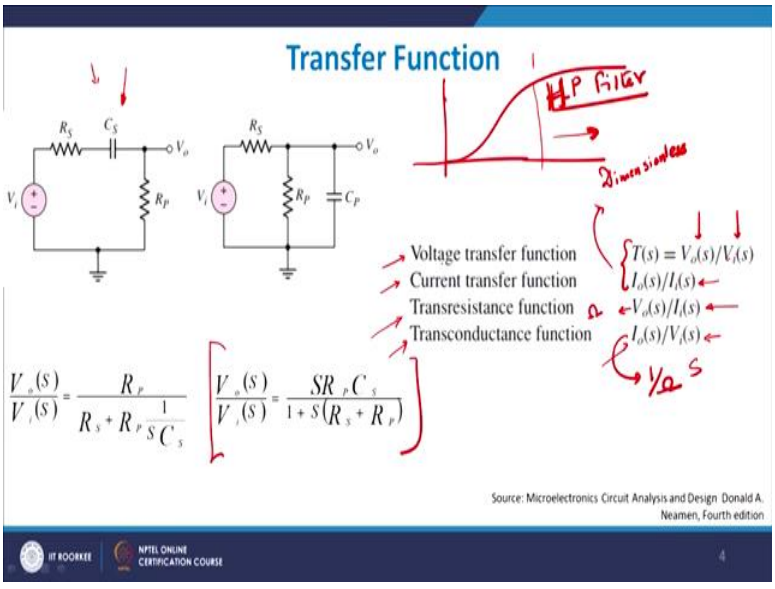
The reason we are not doing it is because for very low values of omega this condition will not hold good that j omega Tow S will always be greater than 1, right. So, for not all values of omega you will have that condition. For certain value of omega of course it will be greater than one then it becomes independent of K.

Now, let me give you a physical insight and then we will mathematically connect that with this equation which we just now derived, right. See, if you look at this 2 port network where this is my $V_{in}$ and this is my $V_{out}$ and then maybe fix the value of $V_i$, that is ok, so we find out $V_0$ by $V_i$ which is function of (s). Now, if you start varying my omega frequency from a very low value to a very high value, let us see how it works out when your omega is very very small.

Let us say omega is 0, which means you are doing a DC bias, then what will happen, the output will be 0 and the reason is when omega equals to 0, 1 upon j omega c will be infinitely large and therefore the capacitance will behave like an open circuit and we already know that DC bias condition capacitance starts to behave like an open circuit and therefore, $V_0$ equals to 0, right. At omega equals to 0, $V_0$ will be equals to 0 and therefore your Transfer Function will give you a gain equals to 0.

Now, if you go on increasing the value of omega, right, then this $X_c$ value starts to fall down from the infinity value it comes down. And therefore, it allows more and more signal to be transferred in the output side and the gain therefore starts to increase as we move ahead, right, the gain starts to increase.

(Refer Slide Time: 15:50)

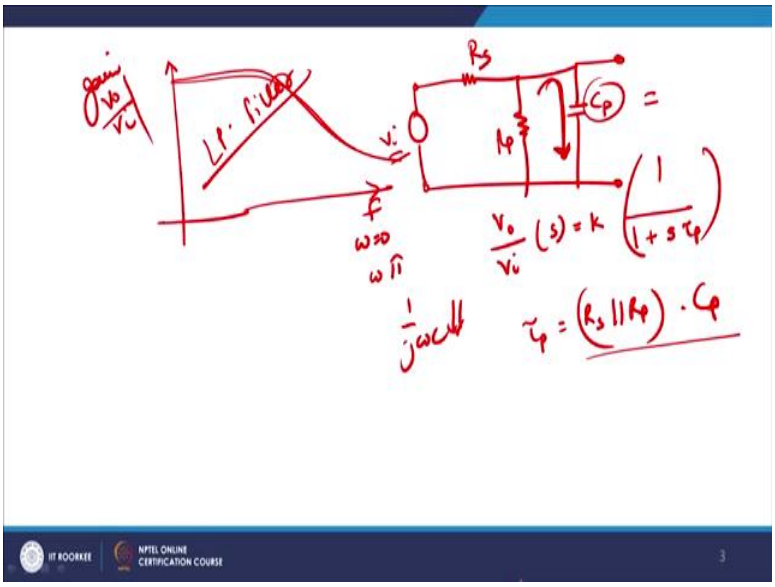



So, if you look back here, therefore for e.g. if you look at this value, then as you therefore make your omega larger and larger, the xc, the resistance offered by the $C_s$, right, that starts to come down and therefore (it gives you) the gain starts to increase. So, what happens is that at omega equals to 0 the gain is almost equals to 0. But, as the omega value goes on

increasing the gain output voltage goes on increasing and therefore the gain goes on increasing.

So, I can safely say that the output function is something like this, right. Therefore, it starts to behave as a low-pass filter. Why low pass? Sorry, it is a high-pass filter. I am sorry, high-pass filter because at high frequencies it is allowing you to the system to pass and at low frequencies it is blocking the system in a detailed manner, right.

(Refer Slide Time 16:46)



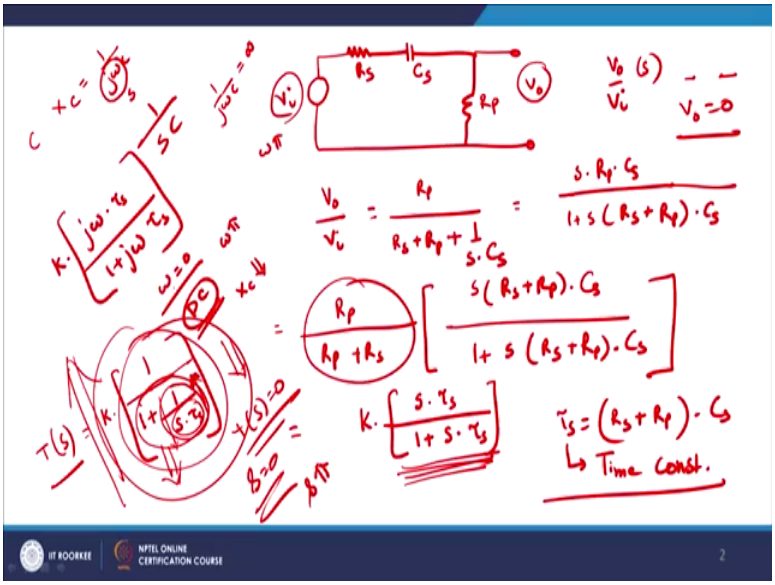

So, in reality if you look very carefully therefore, that it is acting as a filter design, basic filter and it is basically high pass filter which you are actually finding out here. Now, if you do a small change in this whole thing and then try to do that you have $R_s$ here and then you have got $R_p$ here and then parallel to $R_p$ you have $C_p$. So, rather than a series resistance capacitance, you have now a parallel capacitance here, right, and this is $R_s$, this is $R_p$ and this is $C_p$ which you get, $V_i$ which you get here, then I get $V_0$ by $V_i$ as a function of (S) to be equals to K into 1 upon 1 plus s times Tow p.

This is you will get, where Tow p is incidentally $R_s$ parallel to $R_p$ multiplied by $C_p$, right. So, it is $R_s$ parallel to $R_p$ multiplied by $C_p$. When parallel means, you will always, the output will always be less than the least. So, out of $R_s$ and $R_p$ whichever is the least value, you will get for that and multiplied by $C_p$, where $C_p$ is this value which you get. Now, you see incidentally here, if you look very carefully, then at omega equals to 0, this will behave as an open circuit, right, this will behave as an open circuit and as a result all $V_i$ will appear at the output side, right, or part of it will appear at the output side.

Whereas, as you make your omega go on increasing 1 by j omega c starts to decrease and more and more frequency curves, high frequency domain circuits are inserted on to the system, right. And this does not appear in the output side, the output starts to fall. So, if you look very carefully, it will be something like this, the output profile will look something like this. If you plot frequency versus (gain) output gain, output gain is $V_0$ by $V_i$, right.

If you plot it you get something like this and therefore it starts to work as a low-pass filter, right. So, I can therefore by simply changing the placement of my capacitor from a series connection to a parallel connection, I can actually convert the high pass to low pass and the ociated with that also gives you the same principle, right, and that you can find out from here, you can find it incorporate that, let me see how it works out.

(Refer Slide Time: 19:03)

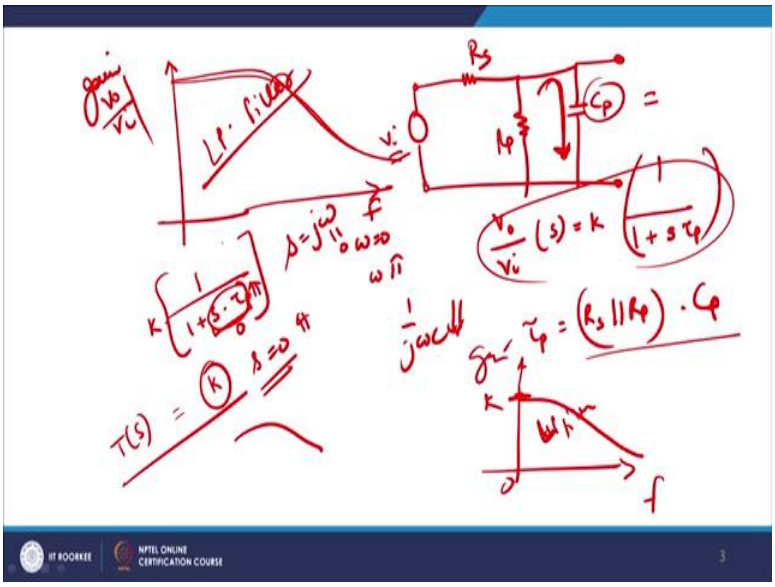



Let us suppose I found out this one, right. So, if you look at this expression of the transfer function then if I divide the numerator and the denominator by s Tow s, I get K times 1 over 1 plus 1 by s times Tow s and this will be equals to your T(s) or the transfer function of s. So, you see as (s) was equals to 0, right. If (s) equals to 0 which is a DC bias, when this is 0 this quantity actually shoots very very high, becomes very very large, right. It becomes very very large, primarily therefore it means that this quantity comes down to a very low value and therefore T(s) is almost equals to 0 at s equals to 0, fine.

So, s equals to 0 at T(s) equals to 0 and s equals to 0. Now, as S starts to increase or as S becomes more and more higher, this quantity, right, this quantity starts to reduce and therefore 1 plus this quantity starts to reduce and therefore this quantity starts to increase,

fine. And as a result you will see that the gain will start to become higher. So, the plot which I made just now, this plot exactly summarizes the plot in the earlier cases, right. So, I just wanted to make it sure that you understand whatever we are doing is in consonance with the output characteristics.

(Refer Slide Time: 20:40)



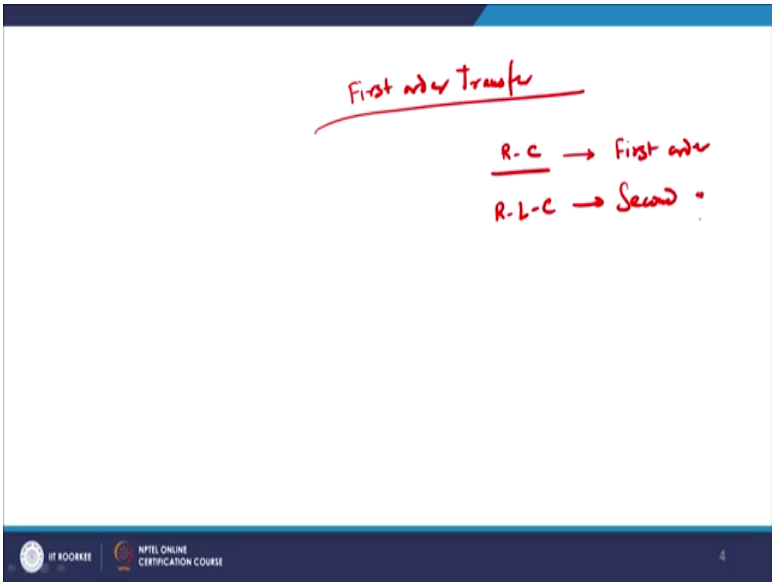

I will show you here also, say for e.g. if you take this into consideration, I get K times 1 over 1 plus s times Tow p, right. Now, if your S is 0 for example then I get output equals to K. If S equals to 0, S equals to basically equals to j omega, that omega equals to 0 I get S equals to 0. S equals to 0 implies that this quantity goes to 0 implies that the output T(s) will be always equals to K and therefore at S equals to 0, if I plot frequency versus gain here, I will get K here at T equals to 0.

As my S goes on increasing, this quantity goes on increasing and therefore 1 upon this quantity goes on decreasing and therefore my gain starts to fall down and (())(21:28) starts falling down like this at higher value of F. So, it starts to behave as a low-pass filter. So, physically also and numerically also, I can show it to you by using transfer function I can predict output characteristics of the device or the circuit in a sense that I will be able to predict its behaviour in terms of output gain with respect to change in the frequency in the input side, right.

That is what is written in this, that is what it is shown in this module or in this worksheet and it gives me an idea that simply by changing the position of the capacitor in this case, for

example, from series to para combination, I am able to change the filter from a high pass to a low pass one. So, this is the basic understanding or operation of the design which we get.

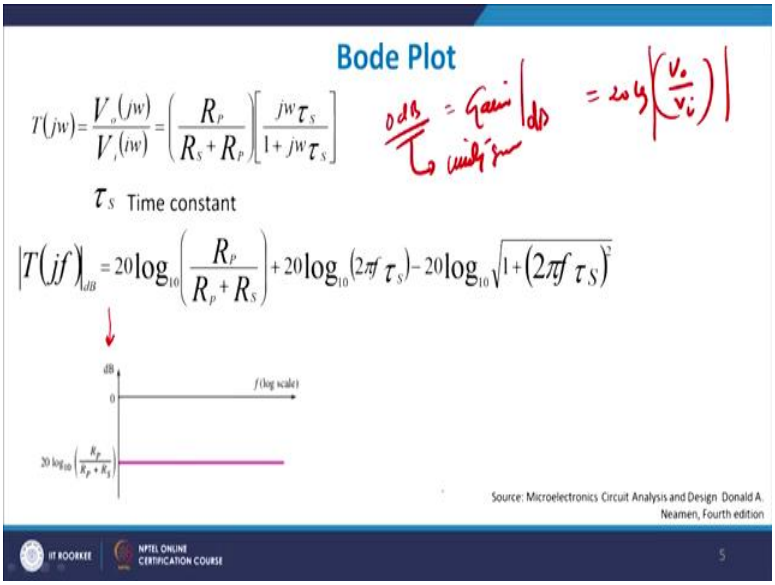
(Refer Slide Time: 22:37)




These are all defined as first order transfer functions. First order basically means that where, apart from register you have only one capacitor into consideration, right. Now, if you have a capacitor as well as the inductor RLC then we would define that as a second order transfer function, right. So, if you have a simple $R_L$, suppose you have got simple R and C network then the equation which defines it is basically or the transfer function defines basically as first order transfer function.

But, if it is an R, L and C, then we define this to be second order transfer function. We will stick ourselves to first order and explain to you from first order itself how you can derive various quantities here. So, we see that therefore you have got 4 types of functions. Transfer function can be given, we can evaluate and we can get the values.

(Refer Slide Time: 23:29)

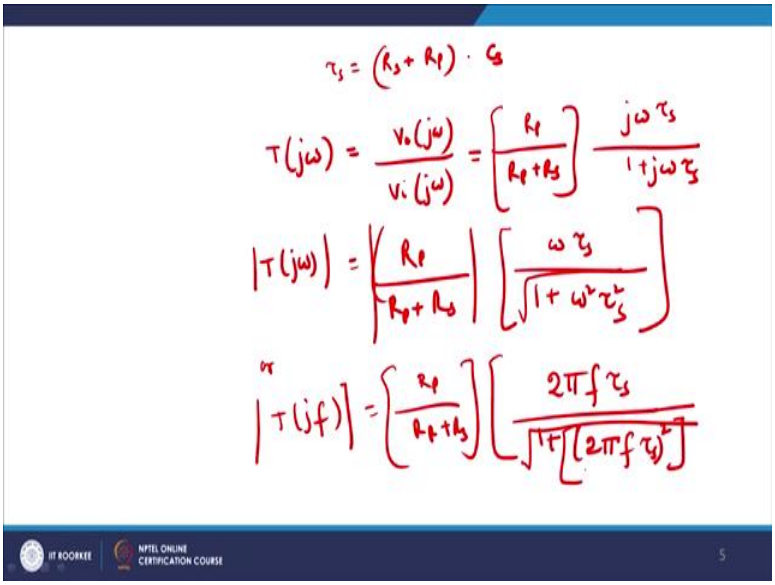



Now, let me come to a very important term or a very important aspect of designing is basically the Bode Plot. The Bode Plot is basically this one, the variations of the (gain) voltage gain in dB with respect to frequency in log scale because, if you remember frequency can vary from 0 to maybe gigahertz value or maybe megahertz or gigahertz value.

Therefore, if I keep it in a linear scale I will never be able to get the actual profile because, linear scale will give you a very large linear scale with us. Therefore, what people have proposed is that the X axis or the X scale should be basically a log scale, right. And in the log scale we have the frequency which is there and on the Y axis we have the gain and gain is in dB.

So, it is basically 20 log $V_0$ by $V_i$, this is your gain in, voltage gain in dB. It is 20 log $V_0$ by $V_i$, right. So, when $V_0$ equals to $V_i$, log 1 will be equals to 0 and therefore our gain will be equals to 0 dB. So, 0 dB primarily means that you do not have any gain. It is unity gain. So, it is basically unity gain available to me, right and so on and so forth. We should also therefore, we will see how you can generate a Bode Plot or a Bode plot, we will see each one of them individually.

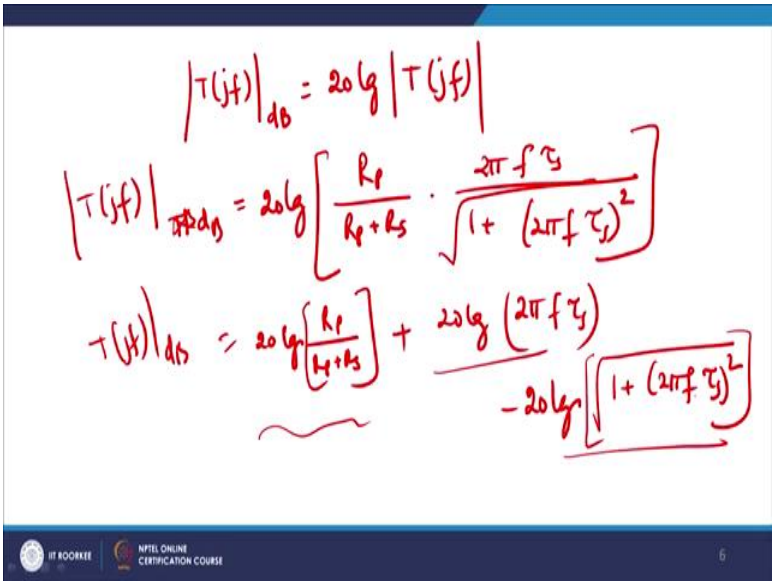
(Refer Slide Time: 25:24)

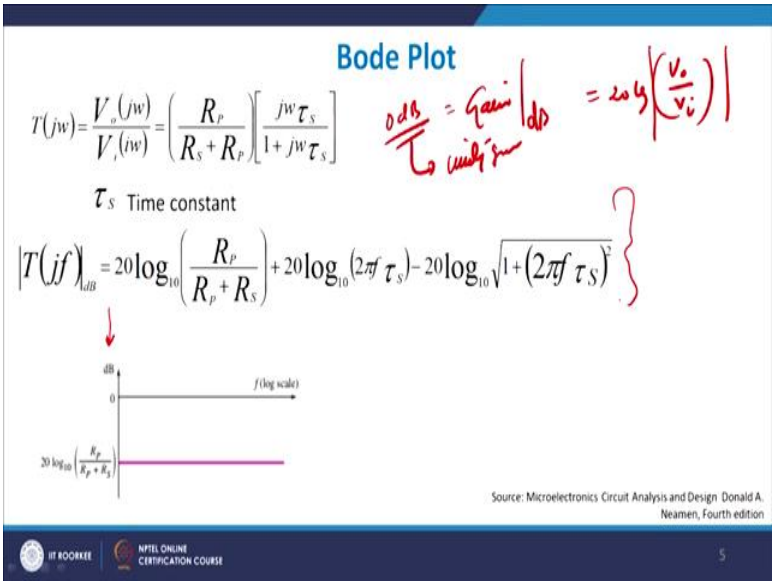



Now, we just now saw that, we just now saw that, let me just again come back to this same concept here and let me explain to you that, we just now saw that Tow S was equals to Rs plus $R_p$ into $C_s$ and T of j omega because T of S was equals to $V_o$ of j omega upon $V_i$ of j omega, right, and is given as $R_p$ upon $R_p$ plus $R_s$ into j omega Tow S divided by 1 plus j omega Tow S.

So, if you take the mode of T of j omega then I get square of this whole quantity and I get $R_p$ upon $R_p$ plus $R_s$, magnitude of this one ofcourse remains the same, into omega Tow S upon 1 plus omega square Tow S square or if I do it in terms of T (j f ) because it is 2 pi f equals to omega, I get $R_p$ upon $R_p$ plus $R_s$, right, into 2 pi f of Tow S upon 1 plus 2 pi f of Tow S whole square and then this will be square root, so this will be square root here. So, I will get this consideration in this particular form.

(Refer Slide Time: 26:55)




Now, if I therefore write T(jf), right, in dB, I just need to find out 20 log of T(jf), right. And therefore, T(jf) in tB, sorry in dB, will be given as 20 log of $R_p$ upon $R_p$ plus $R_s$ multiplied by 2 pi f Tow S upon 1 plus 2 pi f of Tow S whole square square root, right. If you look at this, it is exactly the same thing as this quantity. Because, why this quantity? Because if you just break it down I get, I can break it down into 2 parts. I get 20 log of $R_p$ upon $R_p$ plus $R_s$, right. I can break it into this, plus I get 20 log of 2 pi f of Tow S, right.

And then I can get a minus sign 20 log of 1 plus 2 pi f Tow S whole square square root of this one. That is what we defined now as T(jf) in dB. So, the whole quantity which we found out just now, can be therefore broken up into 3 parts, 1, 2, 3 and we will see in the next turn, in the next module that when we start with this and we add these 3 together how the profile

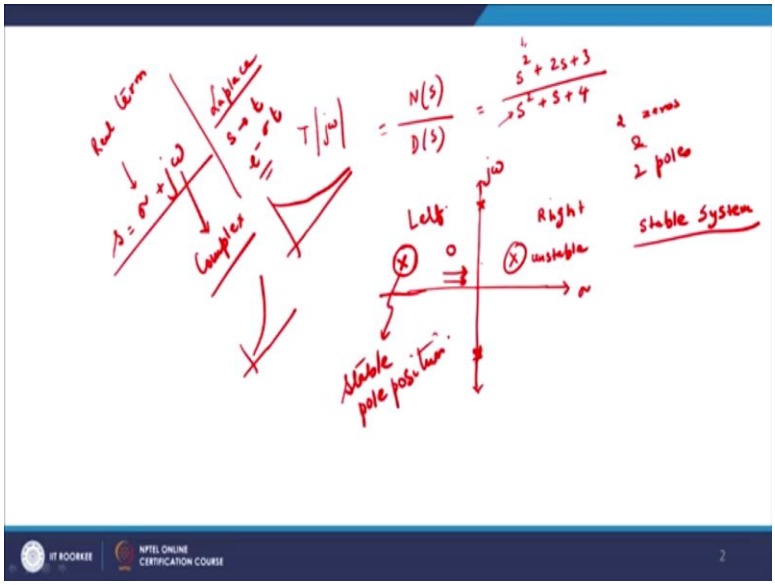
comes out in the output side, right. So, we will just do a principle of super position for all these 3 and check out the values of the output Bode Plot, fine. In the next turn when we take up, we will discuss this point. Thank you.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−41**
**S-Domain Analysis, Transfer Function, Poles and Zero- II**

Hello everybody and welcome to the NPTEL online certification course on Microelectronics Device Circuits. We start from where we left in the previous module and this time we will be studying again the S-Domain Analysis, Transfer Function, Poles and Zeros, we will do the part II.

Yesterday, we have seen, in the previous turn we have seen that how do you define a pole, a pole is basically a value of frequency at which the transfer function which is basically the gain function goes to infinity and 0 is that frequency at which the transfer function goes to 0.

(Refer Slide Time: 01:23)




So I can have multiple zeroes and multiple poles with me. We also, we did not learn previous turn but we will be learning it later in our course structure, is basically that if you have a say numerator by denominator in both S term, this is the transfer function. Let us suppose we are talking about say j omega. Then the maximum order of your S will determine the number of zeroes, so if you have got S square plus 2S plus 3 divided by S square plus S plus 4 let us suppose, then since the maximum order is 2 here and the maximum order is 2 here, I will get two 0s, and I get two poles.

So whenever we plot these poles and zeroes, generally we plot it in what is known as an S curve or S circle. So generally we plot it in this manner that this is my sigma and this is my j omega. So S is written as a mixture of sigma plus j omega where sigma is basically the real term, real term and j omega is basically the complex quantity available with me.

Whenever we therefore try to find out poles for example, it is been seen, we will not derive it in this classroom or in this module that if the poles are on the left half plane which so this is your right half plane and this is the left half plain of j omega plot. If your poles are generally shown by this crosses and zeroes are shown by such elements, so if the poles are on the left half plain of the j omega axis, then we define the system to be a stable system, which means that you will never the system to actually be oscillatory or even if it is oscillatory it will be infinitely oscillatory.

And stable means that the transfer function will go on will not go on increasing in differently onwards. So therefore your left half plain is one of the reasons why we require to do it and there are certain reasons for that if you we will learn it later on if the time permits when we do a Laplace transform and we covert from frequency space to time domain, we get e to the power minus sigma t. So if it is minus sigma t, it will always be exponentially decaying function and therefore that is a stable function which you see.

If it is an exponentially increasing function like this it will always be a non-stable function with me. So typically we what is the importance of getting poles and zeroes is that the poles helps you to find out whether the system is stable or not. So as long as they are on the left half plane of the system, you will automatically get a stable system.

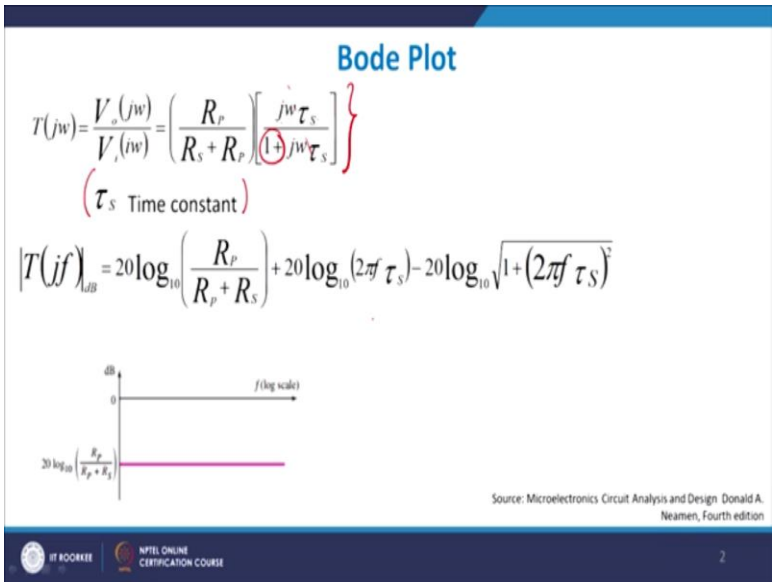
This is basically an unstable pole position, this is basically a stable pole position and this gives you a stable pole and this gives you an unstable pole. Where you place your poles? The poles can also be placed on j omega axis here as well as here, no problem, it can be placed on these j omega as well and as a result you might have a oscillatory behaviour of the circuit. For example pure oscillation, RC oscillations which are there in which the poles falls on the j omega axis.

The second point which you should remember at this stage only and we might not be revisiting it again once again is that more the pole is near the j omega axis more better the more stable the system is. So more pole closer to the j omega axis on the left half plane better it is and these are the few just small nuances of this sigma j omega plot and therefore just by

looking at these positions of poles and zeros we can predict whether the system is stable, whether the system is stability is in the system or not.

We come back to, with this basic knowledge which we left I think yesterday we did not do this, we come back to yesterday's talk again and let me show to you where we left in the yesterday.

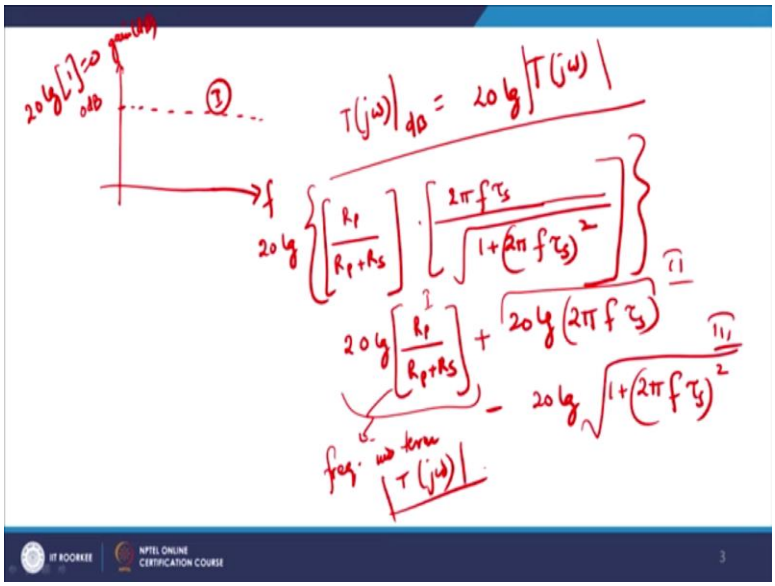
(Refer Slide Time: 05:45)




So if you remember yesterday we tried to find out this TJ of J omega and we found out to be $R_P$ upon $R_S$ plus $R_P$ into j of omega tau $_S$ upon 1 plus j omega tau $_S$ where tau $_S$ is defined basically the time constant of the circuitry and omega is the frequency which is you get and which is equals to 2 Pi f.

Therefore if you look very carefully, even if omega is very large quantity and if it is very large as compared to 1 then this will get cancelled out and you are left with the $R_P$ upon $R_P$ plus $R_S$. Assuming that $R_S$ is very very small I will get overall gain to be equals to unity and that is quite an interesting phenomenon.

(Refer Slide Time: 06:35)




But as we discussed in our previous turn, we told you that we generally find out the we define we gain in terms of dB, in terms of dB as 20 log of T of J omega, this was the definition which we did. So when we find out this definition and we placed the 20 log of this quantity, so what we found out in the previous or yesterday what we did was something like this that we got 20 log of $R_P$ upon $R_P$ plus $R_S$ and multiplied it by 2 pi f tau $_S$ upon 1 plus 2 pi f tau $_S$ whole square into square root. This is your log of this whole quantity is basically your transfer question.

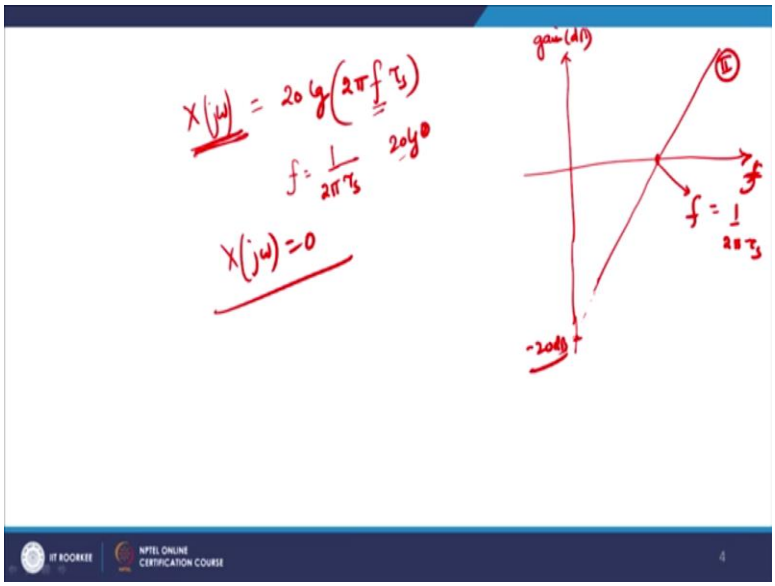
So if you break it open, you get three terms which is with you. One is 20 log of $R_P$ upon $R_P$ plus $R_S$, this is the first term, the second term is of course 20 log of 2 pi f of tau $_S$ and the third term will be the negative sign because this is the denominator, will be given by this thing, 20 log of 1 plus 2 pi f of tau $_S$ whole square. So this is the term number one, this is term number two and this is term number three and the overall therefore the plot which is basically the gain versus frequency in dB we can do it by principle of superposition of this, number 1 plus number 2 plus number 3.

So let us look at one first, it is very straight forward and simple. If you look very carefully this quantity $R_S$ is very small as compared to $R_P$ and therefore log this becomes log of 1 and therefore 20 log of 1 will be equals to 0. So if you plot a dB plot you get something like this, so it is your this is your 0 dB and this is your gain in dB and this is your frequency in log scale and this is what you get, that you will get always a straight line because of term number 1 and it will be independent of frequency. As you can see here it is almost independent of frequency and therefore I can safely write down this to be as a straight line moving from 0 dB

onwards to higher and higher dB, sorry higher and higher frequency and the dB is 0 dB remains constant.

Now this definition which means that this is basically a frequency independent term in your T of j omega, let me come to the second term now and explain to you how a second term will look like.

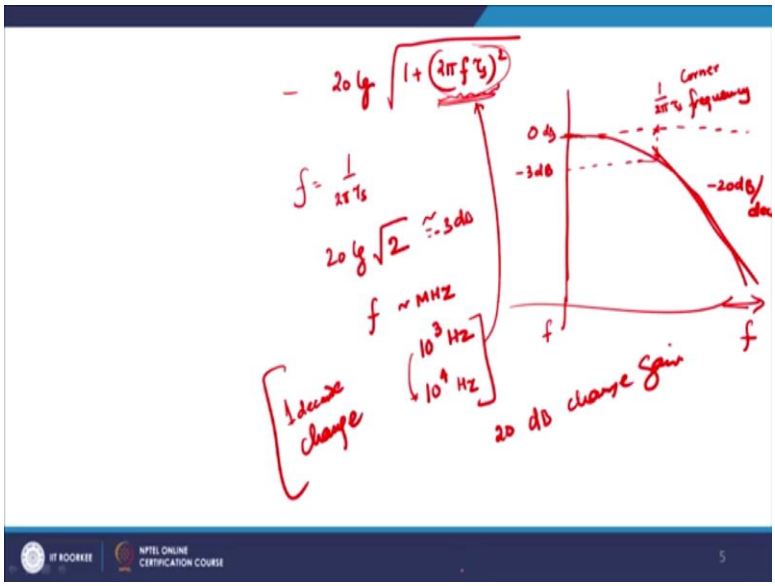
(Refer Slide Time: 09:34)




If you look at the second term, it is basically 20 log of 2 pi f of tau $_S$. So if you plot again a gain in dB versus frequency then I can be amply sure that at f equals to 1 by 2 pi tau $_S$. If I put f equals to 1 by 2 pi tau $_S$ I get this to be, suppose this is X of j omega then X of j omega will be equals to 0. And then therefore this is the, sort of a corner frequency or cut-off frequency, and at this value the f value is defined as 1 by 2 pi tau $_S$. Now you see X of j omega is actually a linear function of f. So as you increase f, X of j omega will go on increasing. As you decrease f, X of j omega will go on decreasing, right. X of j omega will go on decreasing.

So this is what you get, similarly if you extend it backward, and this is where it cuts, this is the place where f equals to 0, so when f equals to 0 I get 20 log of 0. Log of 0 equals to 1 and therefore I get minus 20 dB here, fine. So, this will be a straight line, so this is part two of your network. It is part two of the network and therefore that is what you see.

(Refer Slide Time: 11:16)




Let us go to part number three and that is basically a negative sign, so it will be 20 log of 1 plus 2 pi f of tau $_S$ whole square and then square root of this whole quantity. If you plot again the same thing, the same concept here that is the 0 degree here and this is your frequency in log scale. Then when f equals to 0, I get again, f equals to 0 means this will be whole term will be 0, implying that log of 1 will be equals to 0 and therefore at f equals to 0, I will get 0 dB, so I will get 0 dB here.

Now, as f goes on increasing, for low values of S f, this quantity will be very very small because square of that quantity and as a result it will still remain 0, right. But as f goes on increasing, this quantity starts to become higher and higher and therefore 1 plus that quantity goes on increasing because it is a negative sign attached to it, this will show a something like this, something like this drop will be shown to you and this will be the overall shape of the curve which you will see.

We define a new term here what is known as a corner frequency and it is given by 1 by 2 pi tau S and again you can understand the reason why, when f equals to again, as you can see, when f equals to 1 upon 2 pi tau$_S$, then I get log of root 2, right and this comes out to be approximately equals to 3 dB. So this is your 3 dB point, minus 3 dB. So this is therefore referred to as a corner frequency. So how do you define a corner frequency?

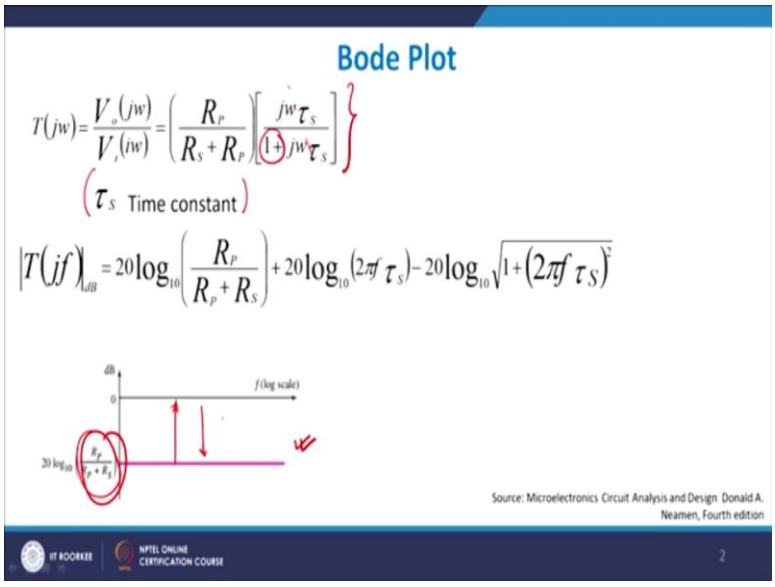
Corner frequency is the point where the frequency is equals to 1 by 2 pi tau S and where the gain has fallen to 3 dB, minus 3 dB from its actual value, so the original value was 0 degree, now it has fallen to minus 3 dB, so you have minus 3 dB, this thing is there, a gain is there. And therefore as you can, so this is what you get, now as f goes on increasing drastically

higher and higher somewhere here, see if you give f a decade increase, for example if f was equal of the order of few megahertz and then you make it say it was 10 to the power 3 hertz, let us suppose and then you make it 10 to power of 4.

Then how does it influence your overall picture, right? That you must be aware of. Generally it is seen that if you feed these two for example figure into this formula, then for every so this is one decade change in the frequency, because it was 10 to the power of 3, it goes tend to power of 4, so this is basically a 1 decade change. So a 1 decade change in frequency should result in a 20 dB change in the gain. If you put it you will find these values here and therefore if you try to find out the slope of this graph, this graph here, it will basically be minus 20 dB per decade, which means that the gain actually falls down at the rate of 20 dB per decade, fine and that is quite interesting that for every 1 decade rise in the frequency, I would expect to see a 20 dB fall in the gain in the voltage gain in the output side.

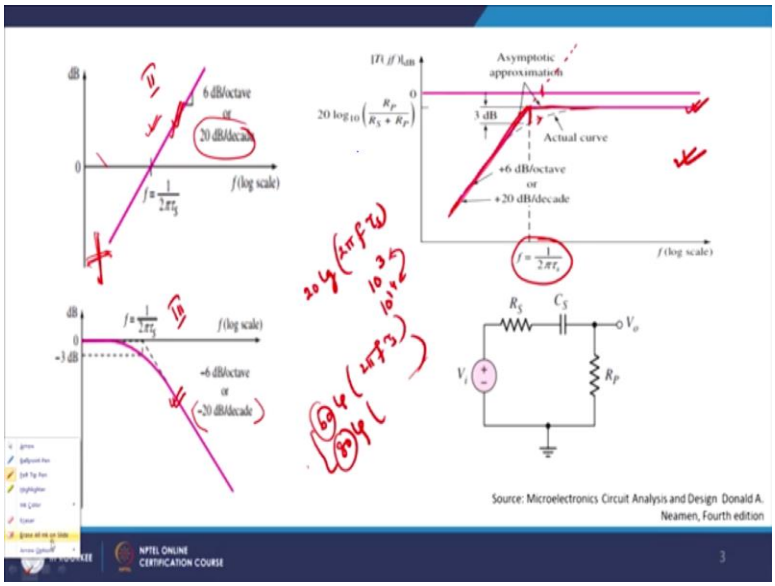
So, this is what you get when you do a 20 dB, when 20 dB drop is available to me, this is for 20 dB per decade. So decade means you are increasing it by a factor of 10, 10 times increase or 10 times decrease is there in terms of frequency. So now I have got 3 mechanisms, so let me switch back to the basic concept here.

(Refer Slide Time: 15:28)




As you can see here the same graph which I was plotting for you, I got this as my frequency scale, this is 20 log $R_P$, assuming that $R_P$ and $R_S$ is very small, I can, this could be safely assumed to be almost equals to 0, if not it will be lower than this because obviously this quantity will always be smaller than 1 and therefore this will be slightly less than zero.

(Refer Slide Time: 15:50)




We come to the next case and as I discussed with you, this will be a straight line which will be a linear curve with a 20 dB per decade, you will have an increase and here you will have a 20 dB per decade decrease. So there is a negative sign at this particular point. Now you can see also why is it like this, this one, this is 20 log of 2 pi f of tau $_S$. So this was a definition which people saw, if I got 10 to the power 3 then this comes out, if I do 10 to the power 3, then this 3 will come out and come out to be 60 log of something in terms of 2 pi f of tau $_S$.

Now if it is 4, I will get what? 4 will come outside I will get 80 log of something, so you see there is a 20 dB change which you see here, 60 and 80, so for every decade change in the frequency from 10 to the power of 3 to 10 the power of 4, I expect to see a 20 dB per decade drop. So I get 20 dB per decade drop here, now if you add number 1 drop, number 2 and number 3, this is number 2 and if you add number 3 here and then put it all together at one graph then this is what you see here.

So what you see here is basically, since assuming that 20 log of 10 $R_P$ upon $R_S$ plus $R_P$ is a figure which is slightly smaller than 1 by virtue of the fact that $R_S$ is not negligibly small, then its constant value will be less than 0 dB and it will be a constant value independent of frequency, which is this curve which you see. And then if you superimpose on that this curve, the second one, then this is the second curve which you see in front of you. This crosses the 0 dB point at 1 by 2 pi tau S and then the third part is from this point.

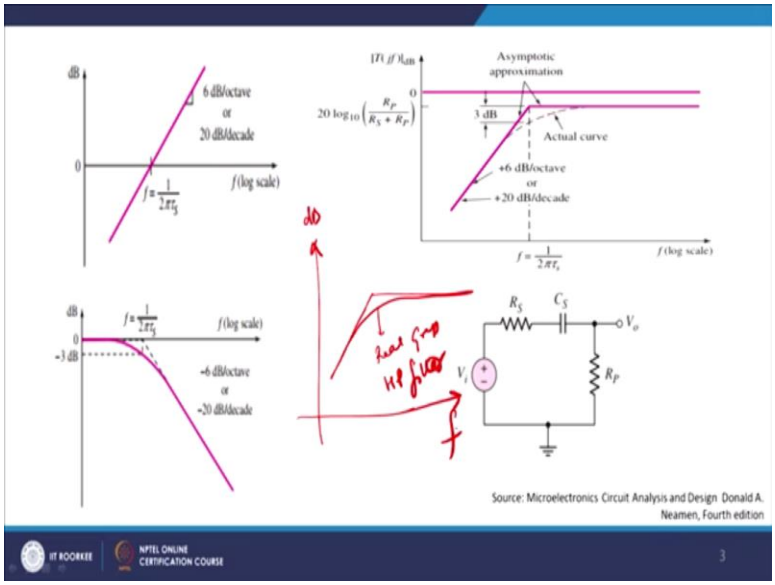
So if you look very carefully, if you add this to this then you can see because of this linear fall here, right, you actually see a linear fall in the output voltage, output gain here. And as you move further in the frequency domain, it goes on increasing somewhere around this

corner frequency here, you actually start to see constant value because the rate of rise here and the rate of fall here are exactly equal. And therefore if you add graph number two and graph number three, I will almost get a constant profile independent frequency, right.

Where this 1 by 2 pi tau S is basically my corner frequency, also referred to as frequency. Now if you look very carefully, this is basically a graph of low pass filter, sorry, high pass filter, why? Because at high frequencies you are allowing the gain to be very high, at low frequency the gain is very very small in dimensions.

So this is what we have learned or we know how to therefore deal with Bode's plot, how to deal with Bode's plot, how to make the Bode's plot work for me. And this works because of the fact that you do have a frequency dependent term. I have to ensure that at every decade rise or fall in frequency, I have a 20 dB change in the value of your gain. If you consider that as a basic one you can carry forward. Similarly, you will always get at the corner frequency a 3 dB drop, right. This is the 3 dB drop which you will get at corner frequency. So this the actual curve, this is the idealized curve and this is the actual curve which you see.

(Refer Slide Time: 19:31)

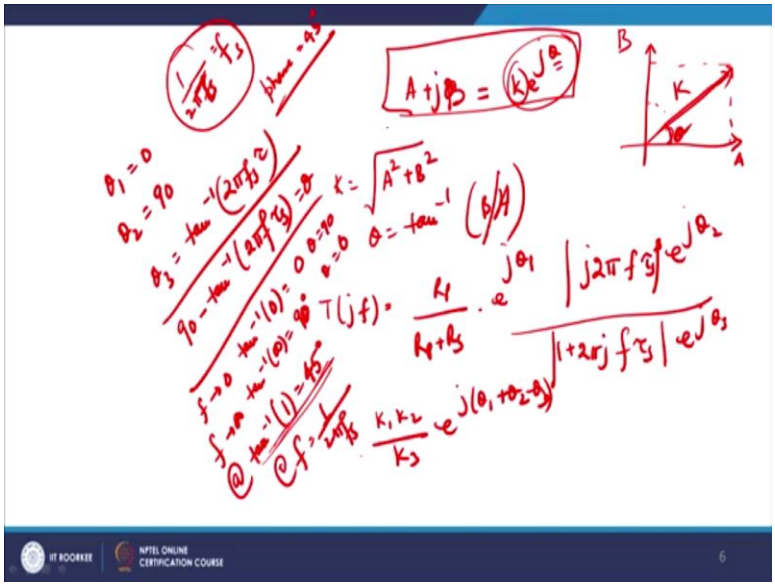



So if I remove all these things, then this is the actual curve, the dotted one is the actual curve, this one is the actual curve. So you see it is something like this, so it is like this, this and then the ideal one should be like this and this and the real one will be asymptotic here and then it will become asymptotic across this.

So this is my real graph and this is basically a real graph of HP high pass filter and this is your gain in dB and this is your frequency in decade in log scale basically and that gives you

quite an interesting phenomena or at least gives me a first-hand implication of how you can design a basic filter also.

(Refer Slide Time: 20:30)




Let me go forward and explain to you but before I move forward let me explain to you about something about polar coordinate system, you can also refer to as A plus j beta right j B as equals to K e to the power j theta. So many a times we represent a Cartesian coordinate system to actually a polar coordinate system. This is A and this is your K theta, this is theta, this is K and therefore I get K e to the power j theta. This is B and therefore K happens to be equals to A square plus B square root over.

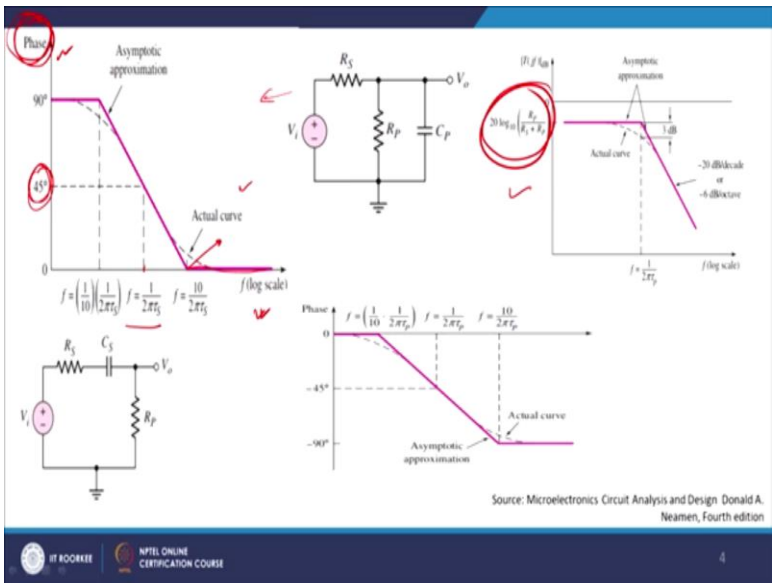
Whereas theta is equals to tan inverse B by A. So if you solve it I get T to the power j f was equals to $R_P$ by $R_P$ plus $R_S$, right e to the power j theta 1 mod of j 2 pi f of tau S, e to the power j theta 2, divided by 1 plus 2 pi j f of tau S, e to the power minus j theta 3. So if you solve it I get K 1, K 2 by K 3 e to the power j theta 1 plus theta 2 minus theta 3, because theta 3 is in the denominator. So I get when theta 1 therefore is equals to 0 and theta 2 equals to 90 degree, I get theta 3 to be defined to be as tan inverse 2 pi f S into tau. This is what we get it.

Similarly and therefore so the theta is basically 90 minus this quantity, this quantity right, so I get 90 minus tan inverse 2 pi f of tau S. This happens to be your theta. Similarly, therefore so I am trying to find the phase margin, when f tends to 0 tan inverse 0 is equals to 0. So theta equals to 90 degree and f equals to infinity I get tan inverse infinity is equals to 1, 90 degree in degree terms it is 90 degree and theta is equals to 0 degree and therefore and at tan inverse let us suppose 1 is given us 45 degree and this will happen at f equals to 1 by 2 pi f S.

So let me just recapitulate what we did by simple mathematical derivation that at corner frequency your phase margin will be approximately equals to 45 degree. The phase margin of the input, you got the phase margin from where I am getting it? See, if you look, see any complex conjugate quantity or complex quantity can be broken down into K e to the power J theta sort of term where theta is basically the phase margin, K is the gain which you see.

So this will be K and if this is theta which you see then I get that 45 degree angle. At 45 angle we will, what is the value of your corner frequency? Corner frequency is place where you get a 3 dB drop in your gain. At that point I get f is equals to 1 by 2 pi f S right and this is tau S sorry, tau S is equals to f S and at this frequency I get my phase margin to be equals to 45 degree and if you go on increasing the value of phase margin if you go on increasing the value of your this thing input, it goes to high value otherwise it goes to a low value.

(Refer Slide Time: 24:33)

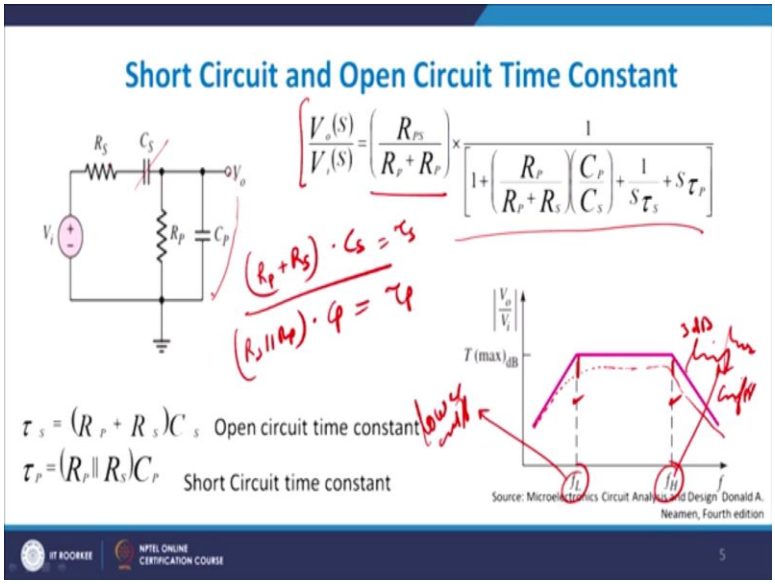



Now with this knowledge let me come to the short circuit and let me come to the previous slide. So this is what I was talking about, if you look very carefully therefore if you plot phase versus log scale for this filter design then we see that around, when it is corner frequency 1 by 2 pi tau S, your phase is basically 45 degree.

So 45 and then as you move forward the phase falls to 0 approximately and at very high values it is then, so when the frequency is very small you theta is equals to 90 degree. And when the frequency is very large, just now I discussed with you, the frequency theta value, the phase is almost equals to 0 degree and this is what I am getting here.

So it is almost 0 degree at very high frequencies and almost 90 degree at very low frequencies. At corner it is exactly equals to 45 degree which you see, so this is sort of a phase scale which you see and how the phase behaves with respect to frequency. We have just now seen how does a gain behaves with respective frequency. So these both are taken care of in a detailed manner as far as this one is concerned.

(Refer Slide Time: 25:45)




If we plot the graph of output voltage to input voltage and we try to find out the various principles here, we see that we can write down in this manner and given by this function, then we define $R_P$ times $R_P$ plus $R_S$ times $C_S$ to be as the open circuit gain. So if you look very carefully at this place, if it an open circuit means you open up $C_P$, $C_P$ is opened up, so $C_P$ does not come into picture. These two are in series, so I get $R_S$ plus $R_P$ multiplied by $C_S$ and this happens to be your time constant $tau_S$.

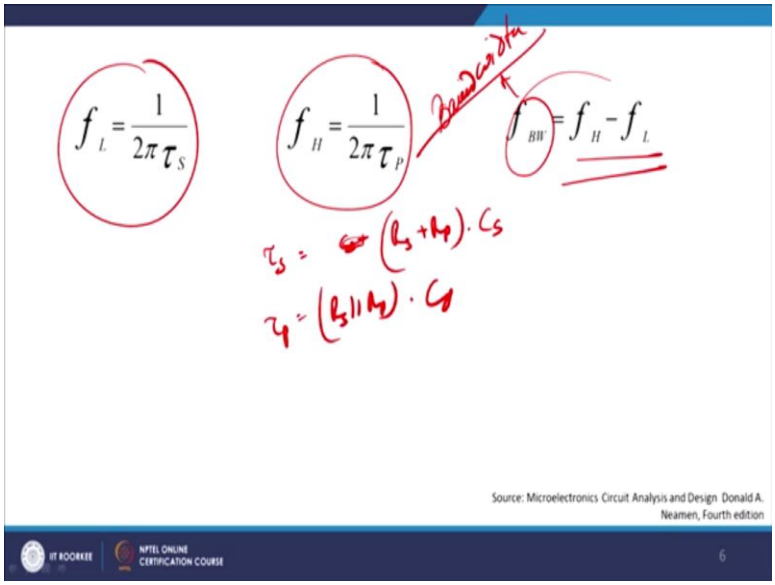
Whereas if you short it when you are shorting it then $C_S$ vanishes off $C_S$ vanishes off, your $C_P$ stays with you but $R_S$ and $R_P$ are parallel to each other. When you short this output voltage with respect to ground, then $R_S$ is parallel to $R_P$ and then this is in sense dot of $C_P$ will give you value of tau P. So $tau_S$ is basically the short circuit time constant when you shot the output and you do not have any $C_P$ coming into picture and you only have $C_S$ into picture and both $R_S$ and $R_P$ are basically series combination resistances.

Whereas when we talk of $tau_P$ and make sure it is to be as a short circuit. So we short it and we try to find out the value of $R_P$ parallel to $R_S$. So we define two sort of boundary elements here, one is known as $F_L$ and $F_H$, $F_L$ and $F_H$ are the points where your 3 dB bandwidths have been maintained. So if I actually plot the graph it will be somewhere like this something like

it will come out and it will follow like this asymptotic and it will go down something like this.

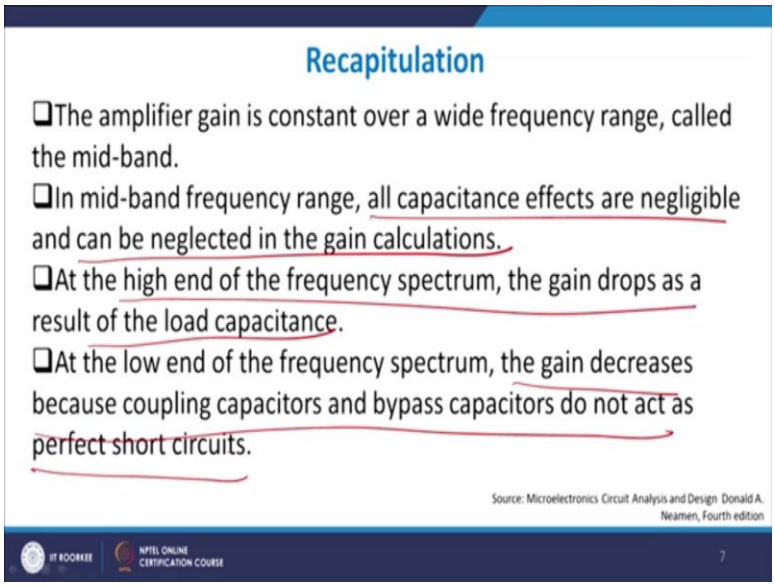
So this drop is basically your 3 dB and we define $F_L$ to be as the lower cut-off frequency, this is known as the lower cut-off and this is known as the higher cut-off. So I have a higher cut-off, I have a lower cut-off and this tends to make the difference between the two bandwidth available to us.

(Refer Slide Time: 27:57)




So if you look very carefully therefore what we finally get is that the lower cut-off will be given by this quantity and the higher cut-off is given by this quantity where $tau_S$ is the short circuit gain given by $C_S$, $R_S$ plus $R_P$ multiplied by $C_S$ .Whereas tau P is given as $R_S$ parallel to $R_P$ multiplied by $C_P$ and this is what you get. And $f_H$ minus $f_L$, high cut-off frequency minus low cut-off frequency is defined as my bandwidth. So this is my bandwidth which I get, this is my bandwidth, so that is the difference between high and low cut-off frequency.

(Refer Slide Time: 28:39)




We come to the, let me recapitulate therefore, generally amplifier gain is constant somewhere in the mid frequency band. At very high and low frequencies the gains starts to fall down for high frequencies fall down with increase in frequency. At low frequency it falls down with decrease in frequency. The mid-band frequencies are all the constant and in that case all the capacitive effects are negligible and can be neglected in gain calculation.

So in gain calculation we do not have any capacitive effects coming into picture. At a high end of the frequency spectrum, the gain drops because of the load capacitance, we have discussed this point already. Because of a heavy load capacitance your system has to actually charge or discharge the capacitor at in the output side and it therefore takes time to do it and that is the reason your frequency, the gain drops down drastically. At the low end of the frequency, the gain also deceases because of the coupling capacitor and bypass capacitor do not act as perfect short circuits.

So there are two reasons why a gain is getting lowered with increasing (freq) with decreasing frequency in the input side when the frequency is low. So, if you remember $X_C$ is which is the capacitive reactance is 1 by j omega C. So if your omega is very very small, $X_C$ is typically very large and as you lower your value of omega, $X_C$ still goes on increasing.

As a result what happens is that the resistance offered is going on increasing and it does not let the signal to pass through from point A to point B. As a result, you always have a loss of gain whereas at very very high frequencies typically high frequencies what happens is that as the frequency becomes very large, the 1 by j omega C as usual drops down and therefore it starts to short your output and therefore the gain starts to fall down in that case.

So for both the reasons you do have a gain drop which is there with us for all practical purposes. So in this way we have finished the concept of frequency spectrum, high frequency mid frequency and low frequency spectrum and how we are able to extract that from a system, what is a Bode's plot and how a Bode plot is generally plotted in a log scale. So we have done all these things and the next time we will take up common emitter and common base configurations for high frequency modelling. Thank you very much.

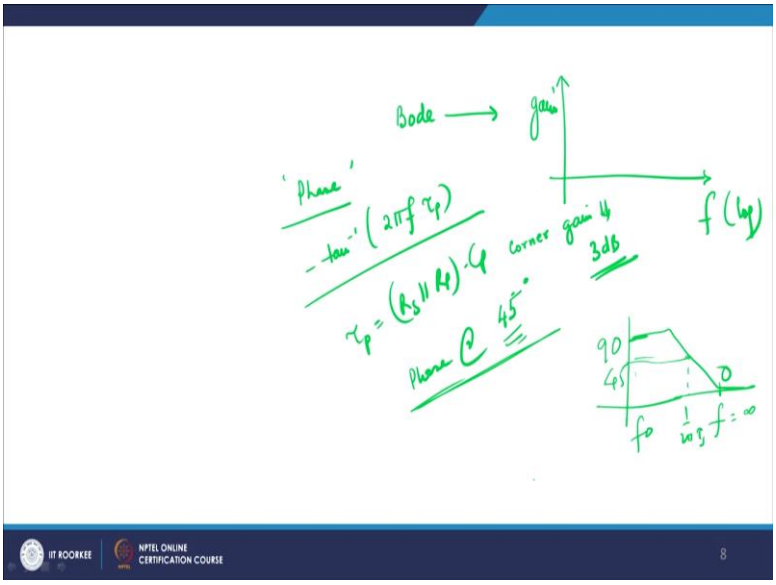
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−42**
**High Frequency Response of CS and CE Amplifier**

Hello and welcome to the next module of the NPTEL online certification course on Microelectronics: Devices and Circuits. We take up today's module as high frequency response of common source and common emitter CMOS-based amplifier. And we will be concentrating on the frequency spectrum or the frequency, how does the gain of the amplifier change with respect to frequency. What is the motivation for such a module? The motivation is that since these are basically used as audio amplifiers, therefore they need to be giving me a constant gain irrespective of large change in the frequency of the input.

So, my voice goes from 20 hertz to 20 kilo hertz, right, the typical audio frequency range. Now, if the amplifier frequency, if the amplifier high-frequency gain suddenly drops down, say at about 18 kilo hertz, 18.5 kilo hertz, the gain drops down. Then those frequencies or those voices, the voices with those frequencies, the audio range in those frequencies will not be properly amplified and I will not be able to hear those frequencies in a better manner.

So, ideally the gain should be almost independent of frequency for CS and CE mode design. And that is the reason we need to know, what are the factors which influence the behaviour of the spectrum of the frequency or the band of the frequency. And we will be concentrating currently on the high-frequency part. Just before we move forward, let me recapitulate what we did in our previous module. We saw that, in our previous module we saw that we were actually doing what is known as the Bode plot, right, we were doing Bode plot.

(Refer Slide Time: 2:51)

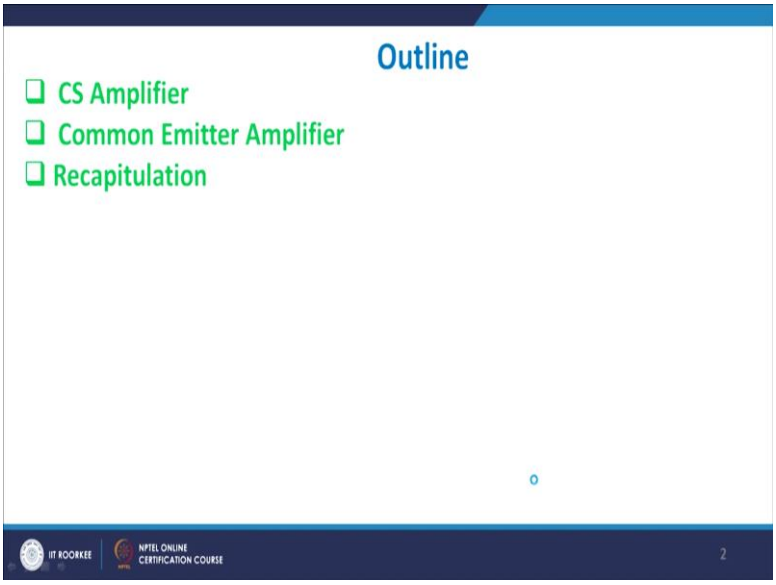



And in the Bode plot we were actually looking into the, what is a Bode plot, that we saw that Bode plot was basically gain with respect to frequency we saw that and frequency should be in log scale. We also saw how to make this bode plot and we expected what is known as the coordinate frequency where the gain drops to 3 dB of its original value and at that particular point, what is the frequency which should be able to gather.

We also understood what is known as a phase or phase margin or phase and is defined as minus of tan inverse 2 pi f of $\tau_P$ where $\tau_P$ is given as $R_S$ parallel to $R_P$ multiplied by $C_P$. Right and typically the phase at unity gain, which is basically meaning whenever you have unity gain, is approximately equal to 45 degrees. Right, the phase margin, the phase is basically 45 degree. So, we should be able to find out at 90 degrees when you have, you will have added a given frequency is very 0, frequency 0 you will have 90 degree and then it goes on and becomes, phase becomes at very large value it becomes 0.

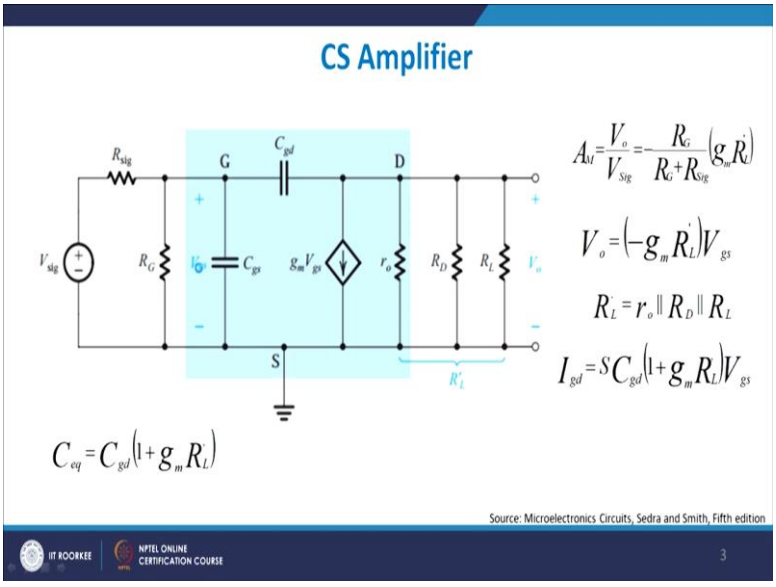
Somewhere in the middle where the frequency drops to 1 by 2 pi $\tau_S$, this is 45 degree. So this is how the phase varies for an amplifier, right. And the gain how it varies, we have already seen working as a low pass, and the high pass, as a bandpass filter we can I do it. We now come and we now discuss about the frequency response.

(Refer Slide Time: 4:24)



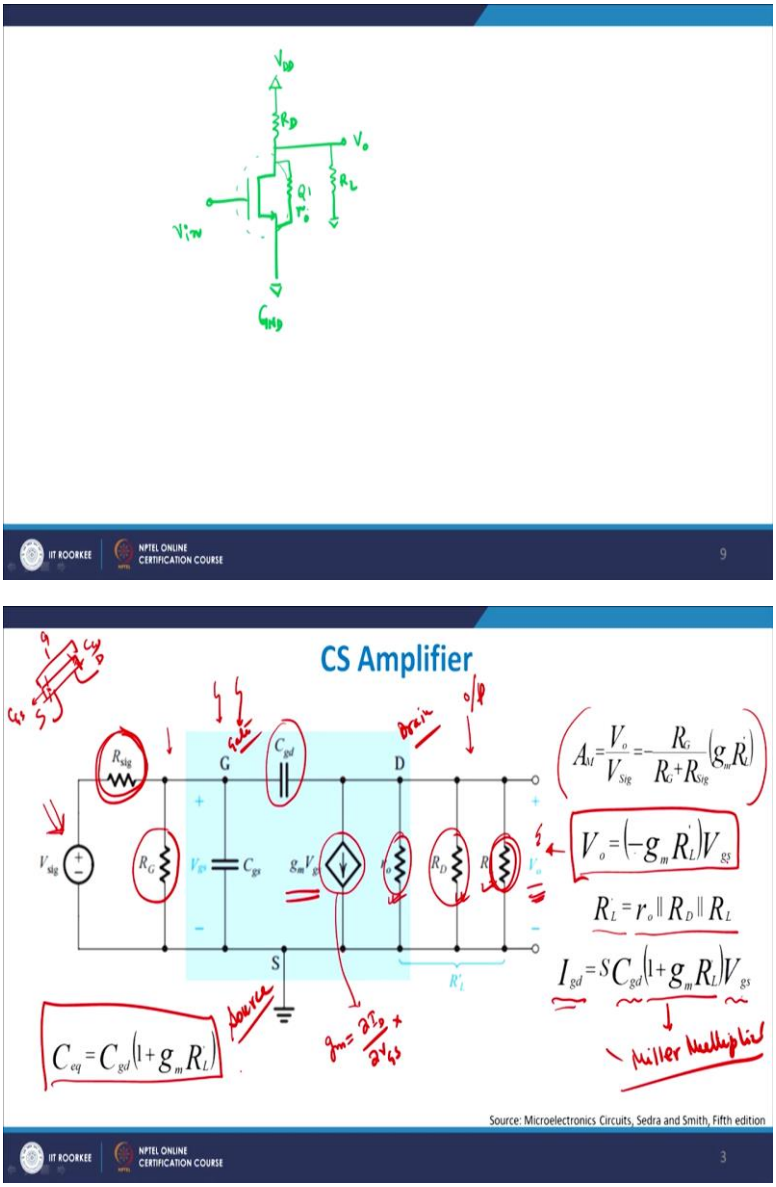

So, what is the outline of this current topic? The current topic is that we will take up another CS amplifier which we have already been doing previously also. This is basically common source amplifier, right and then we will look into the common emitter amplifier in the BJ T configuration and then we will be recapitulating the high-frequency response for CS and this thing, for common emitter.

(Refer Slide Time: 4:54)




This is the small signal diagram for common source amplifier.

(Refer Slide Time: 4:57)




Let me just draw for you against the common source amplifier itself. The common source amplifier looks something like this, we have already explained it earlier also, that it looks something like this. Right, I am not giving any source degeneration resistance here. And this is the output I am taking and this is my $V_{DD}$, this is ground, and we have input here, then we do have gate here, we do have a MOS device here, which is basically the MOS device here and this is one output. Right, so if you do a corresponding small signal analysis here, I can replace this Q1 by GM $V_{GS}$.

Because GM is basically the trance conductance, right and GM is defined equal to $\partial I_D / \partial V_{GS}$. So, if you multiply this with $V_{GS}$, $V_{GS}$ gets cancelled out and you get, we are left to the current source. So, therefore we have already discussed this point that GM multiplied by $V_{GS}$

will give me this current source here. Look at the input part 1st of all, so this, we will look at the blue part just now. This blue part is actually the device part. And this is certain the circuit which we insert in input side and this is the circuit which you insert in the output side.

So, if you go back, we generally have an external load here, $R_L$, right, and this is my $R_D$ and I have also a resistance across source and drain given by $R_0$. So, this $R_0$, $R_D$ and $R_L$ are actually parallel to each other. We will come to that later on. This is the signal source which we are giving, right and this is the ideal voltage source. So, it will have almost 0, very low output impedance. But in reality there will be always a signal impedance available to me and therefore that is given by this $R_{sig}$ which you see.

So R sig is in series to the voltage source. You also have $R_G$ which is basically the resistance which is at the gate side of my design. It is typically very large value because the gate resistances are very large. Right, and they are held parallel to the gate terminal or the gate node. So, it is very large primarily means that for all the V sig will appear across the gate. So, what happens primarily is that your signal voltage, which is V sig here goes through a potential divider network of $R_{Sig}$ and $R_G$ and depending on the relative values of this $R_{Sig}$ and $R_G$, a part of $V_{Sig}$ appears on the gate side of it.

You will ask me why we do all these things, the reason being that if you do not do it, there might be a chance that the voltage, that suppose the voltage, the voltage source suddenly gets shorted and the voltage source suddenly the rises to a very large value. Then it might destroy the MOS device permanently and that is the reason we generally give resistance biasing or potential divider biasing at the input side, so that not all of the voltage of the gate appears, of the input signal appears on the Gate side.

So this is all about your left-hand side. Now, on the right-hand side, which is basically your output side, if you look very carefully, you have $R_D$ which is basically the drain resistance, you have a load resistance $R_L$ here and you have output voltage $V_0$. And this $R_0$ is the applied voltage. Now, if you look very carefully, if we look at this diagram, for example, very simple, it is output by input, V 0 is output voltage by V Sig is the input voltage. It is given as GM multiplied by $R_L$', where $R_L$' is basically the parallel combination of $R_0$, $R_D$ and RL.

So, this is the effective resistance seen by the output voltage source from this side. So, GM multiplied by $R_L$' multiplied by $R_G$ divided by $R_G$ into $R_G + R_{Sig}$. Therefore, I can safely write down $V_0$ to be equals to minus GM times $R_L$ into $V_{GS}$ because $V_{GS}$ is the input voltage and

your minus GM times $R_L$ is basically the voltage gain which you see and $R_L$' is this one. Now, the gate to drain voltage, now let me to this blue part and explain to you the other principles here.

See, you do have, this is the, MOSFET is basically as you remember from your basic days, that is basically 3 terminal device. It is 4 terminal but for all practical purposes in this case, we will take as 3 terminal. So, this is your gate, this is your drain, right and this is your source. So, source, gate and drain, so these 3 terminals are there with me. Now, a gate to drain, you will always have a capacitance which is the depletion capacitance by virtue of overlap between gate and drain.

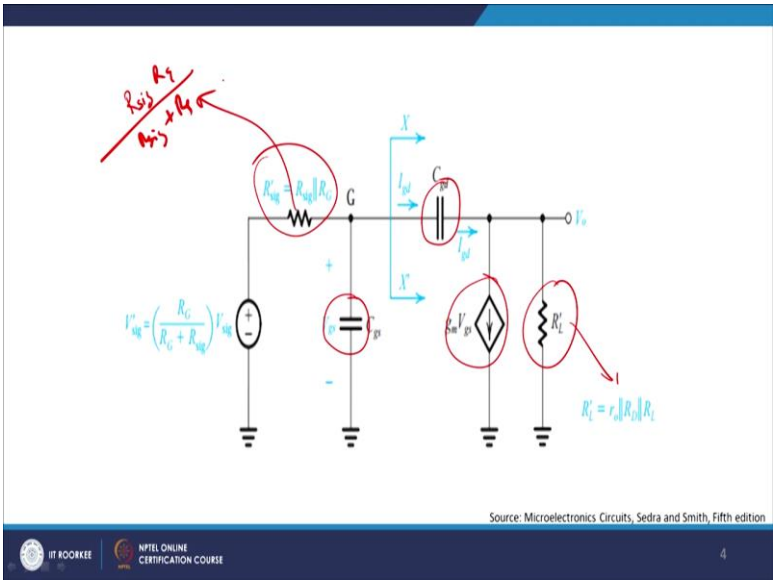
Remember this is $C_{GD}$, right, you also have $C_{GS}$, gate to force. So, if you remember your basic MOS device structure and then it looks something like this, then you have something like gate and this is your gateway, this is your gate, this is your source and drain. So, you will always have a capacitance here and the capacitance here. This is $C_{GS}$, right and this is your $C_{GD}$, right. So, I have $C_{GS}$ and I have a $C_{GD}$ capacitance across the 2 ends which is with me.

Now what happens is that once you multiply, so if you want to find out the total current flowing through $I_{GD}$ or the gate to drain current, then because of the overlap capacitances, then it is basically $C_{GD}$ multiplied by $V_{GS}$ that is the current. But you also have a 1 + GM multiplied by RL Prime. This is known as what is known as the Miller multiplier factor. This is the Miller multiplier. It means that your C equivalent, if you look at C equivalent, it is nothing but $C_{GD}$ multiplied by 1+ GM times $R_L$'.

So, which means that quite an interesting phenomena, that gate to drain capacitance appears in the output side as a increased value of capacitance by how much, 1+ GM times R'. So, higher the value of your transconductance of your MOS device, more will be the current of course but then more will also be the gay to drain capacitances. From where I am getting this, we will not discuss at this stage but at this stage we will just assume it to be Miller indices or Miller multiplier.

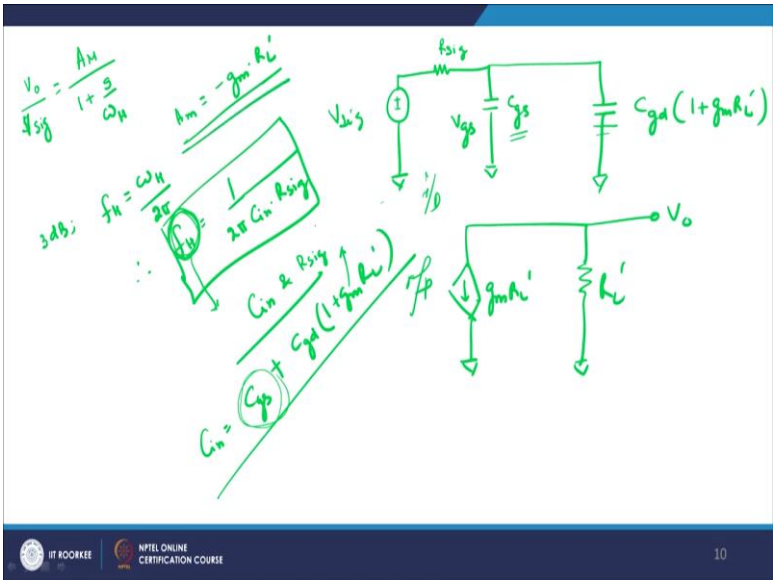
If time permits we will come to this later on which means that because of some overlap, you generally have a Miller capacitances which comes into picture and these Miller capacitances tend to show an increased value of your capacitances. Right and therefore I get $C_{GD}$ 1+ GM times $R_L$ which you see here.

(Refer Slide Time: 12:07)




And from here if I go back to my previous, and therefore if we just drew the small signal model of it, I get $C_{GS}$ here, I get $C_{GD}$ and this is GM $V_{GS}$ is a current source, this $R_L$ Prime takes care of $R_0$ + parallel to $R_D$ parallel to $R_L$. And this $R_{Sig}$ is basically $R_{Sig}$ is parallel to $R_D$, it is basically $R_{Sig}$ $R_G$ divided by $R_{Sig}$ + $R_G$. Right, this value, this is a typical value which you get. And this is the effective value which you get for all practical purposes.

(Refer Slide Time: 12:44)




Now, let me come to explanation here and explain to you maybe a simple derivation here. Let me show to you that I have a $V_{Sig}$ here, I have $R_{Sig}$ here, there is $C_{GS}$, gate to source across which you will have $V_{GS}$ and this is your $R_{Sig}$, this is your $V_{signal}$ and then you have your $C_{GD}$ into 1+ GM times $R_L$'. So, this is the effective value which you see in front of you. Now, if you therefore plot the output side, I can write it safely as GM times $R_L$', right, this will be in parallel to $R_L$' and there will be output here.

So, this will be the input side and this is your output side, right. And I get $R_L$' multiplied by $V_0$. Now, $R_L$' consists of 3 resistances and these 3 resistances will be responsible for giving you the overall picture. Now, therefore, if you plot $V_0$ by V, sorry $V_{signal}$, output voltage by $V_{signal}$ voltage, I will typically get a general form will be 1 + S by $omega_H$ in terms of S. Where $A_M$ is the mid-frequency gain, which is independent of frequency, remember, is nothing but minus GM times $R_L$.

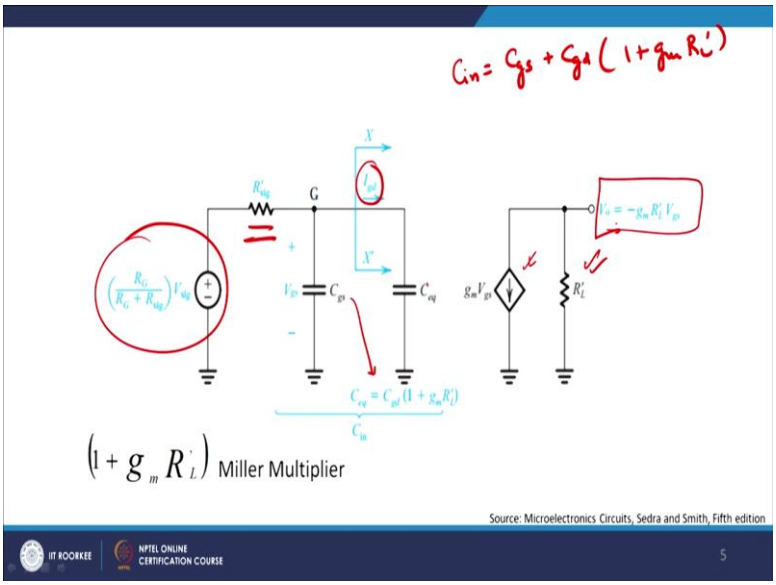
And this is the mid-frequency that you get here. At 3dB if you want to find out, then I get F of H equals to omega H by 2 pi and therefore, I get F of H to be equal to 1 by 2 pi $C_{in}$ into S into, sorry, let me just again write down this whole thing, so this is easy here and then, so I get 2 pi, right, $C_{in}$ into $R_{Sig}$. This is your sort of 3 dB sort of corner frequency. So the corner frequency which is again a 3 dB corner frequency is being primarily determined by the input capacitance here and the $R_{Sig}$.

What is input capacitance? Input capacitance is nothing but $C_{GS}$, gate to source + $C_{GD}$ into 1 + GM times $R_L$'. Why, because this and this are not parallel to each other. And therefore,

they will be adding directly with respect to each other. Right, and you can see quite an interesting phenomena which has come out from here. That, therefore depending on the value of $C_{GS}$, the frequency of the, cut-off frequency or the 3 dB frequency will go on changing. Similarly, if you make your GM higher, because you want the current to be high because you want the gain to be high, this came to be high for example, you end up having a smaller value of $F_H$, right.

So, your smaller value of $F_H$ primarily means that your cut-off frequencies are lowered now. It is not a good idea if you want your bandwidth to be very large and that is the reason that the designing is a bit critical, that just simply by increasing the transconductance of the device will not always make your life easier and this is what the problem area which you face in general you get.

(Refer Slide Time: 16:21)

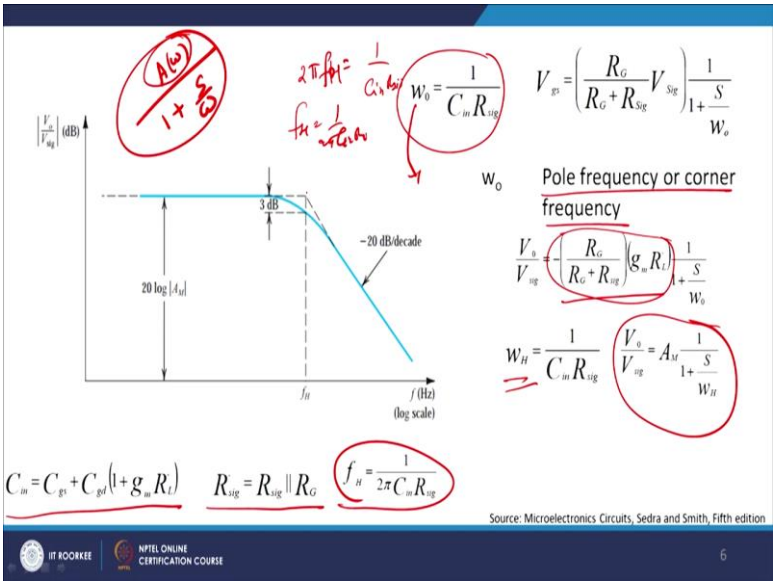



So, this is the structure which you get from here, let me come therefore to the effective as I was discussing with you just now. That I have got, this is the effective resistance which is seen from the device, this is the signal resistance, and current flowing through is $I_{GD}$, gate to drain and C equivalent is given by $C_{GD}$ 1+ GM times $R_L$, whereas $C_N$ is therefore equal to CGS + this quantity. So, $C_{in}$ will be the sum of these 2 quantities and that is the reason I was telling you that effective C in value will be given as $C_{GS}$ + $C_{GD}$ into 1 + GM times $R_L$'.

And that is what you will get here. This is $R_L$' with you get and this is the MOS device equivalent circuit and this is my output voltage, final output voltage given by minus GM times $R_L$' minus $V_{GS}$. Negative sign because it is phase shifted by 180 degree and therefore

you will always have a negative sign attached to it. With this knowledge, let me therefore plot for you the function which we have just now derived.

(Refer Slide Time: 17:17)




And if you look, in this omega as I discussed with you was 1 by $2C_{in}$, so this omega 0 is basically 2 pi F of H is equal to 1 by $C_{in}$ $R_{Sig}$. So, F of H will be equal to 1 by 2 pi $C_{in}$ $R_{Sig}$ and that is what is written here also. So, this omega is referred to as the pole frequency or the corner frequency and it is given by this. So, if I write down in terms of $V_0$ by $V_{Sig}$, it is minus $R_G$ by $R_G$ + $R_{Sig}$ into GM $R_L$ into 1 upon 1+ S by omega 0.

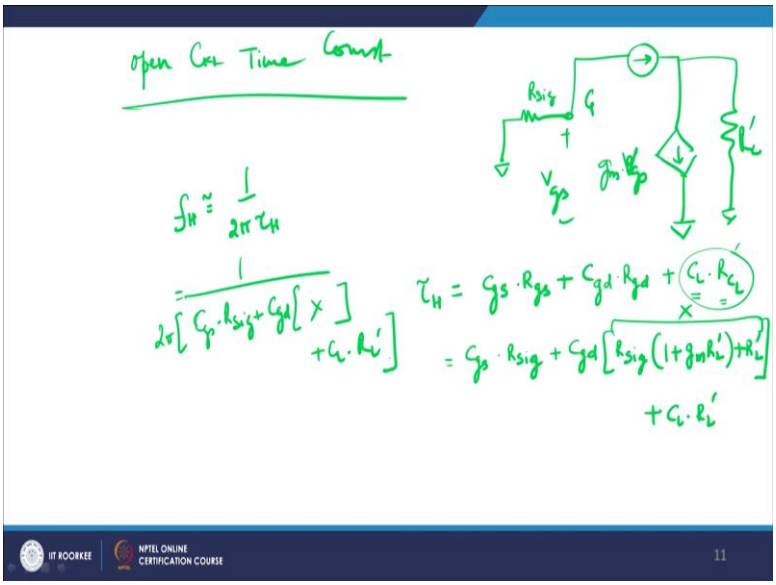
Because if you, as I discussed with you, general formula is A of omega, which is the mid frequency gain upon 1+ S by omega that is the general formula for any gain structure of any amplifier. This is the mid-frequency gain, mid-frequency gain is given by this quantity. So, this divided by 1+ S by j omega. Omega H is given by 1+ $C_{in}$ into $R_{Sigma}$ and therefore this is the formula which I just now discuss with you here. Where therefore, as I discussed with you, $C_{in}$ is equal to $C_{GS}$ + $C_{GD}$, research, where $R_{Sig}$' is equal to this much and $F_H$ equal to1 by 2 pi $C_{in}$ into $R_{Sig}$.

Therefore, at $F_H$, which is this one, I would expect to see a 3 dB drop in my gain. Right, so the corner frequency is that frequency at which I get a 3 dB drop in my gain. And the point at which this happens is determined by the value of C and an $R_{Sigma}$ which is the output impedance of my voltage source which is driving the MOS device and the input capacitance is seen by the MOS device. Please understand that is no output capacitance still playing a role here, apart from $C_{GD}$, gate to drain.

So, gate to drain is again multiplied by 1+ GM times $R_L$'. So, the input capacitance comes from $C_{GS}$, gate to source and the output side comes from gate to drain and then both get added up together and that gives you the value of your this thing. So, this is your 20 log A of M, this is your mid-frequency gain which you get, at higher frequencies therefore they start to drop. Therefore a common source or a common emitter mode configuration at higher frequencies also shows a drop of 20 dB per decade, which we have already discussed in our previous modules.

And we have seen that this works pretty fine for all the practical purposes. We will come to common emitter, before we come to common emitter, therefore let me explain to you what is known as an open circuit timing, open circuit time constant and explain to you what is known as an open circuit time constant and I will show to you how it works out.

(Refer Slide Time: 20:15)



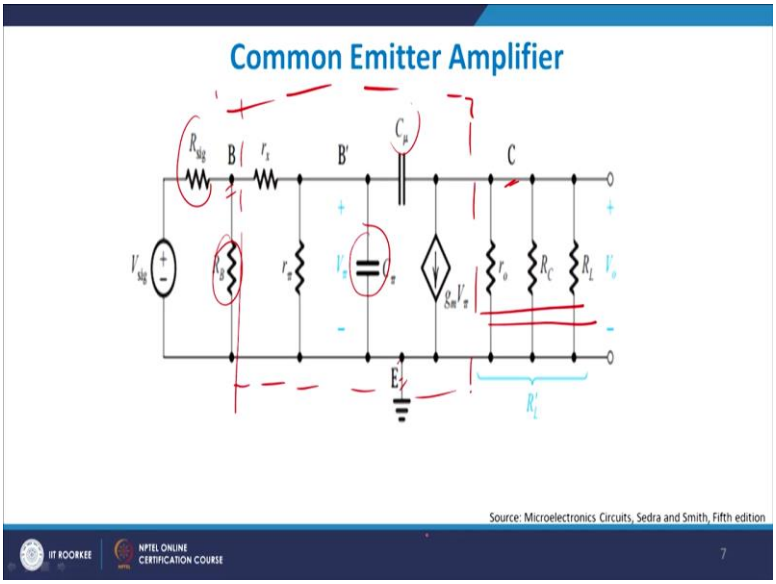

So, let me give you, so I have a signal source here and I have grounded and this goes to get, this is R Sig which you see and then the same thing can be continued and I will have the current source here, current source here, and then this is GM times $V_{GS}$ and then you have got $R_L$'. This is your current. So, this is your plus minus $V_{GS}$. Right, so I get from here, if you do small analysis, I am not going to dig it the whole because it is out of scope, I will get the total this thing will be given as $C_{GS}$ multiplied by $R_{GS}$ + $C_{GD}$, right, let me write down for you, let me just $C_{GD}$ multiplied by $R_{GD}$ + $C_L$ multiplied by $RC_L$.

So, $C_L$ is the load capacitance multiplied by the resistance offered by the load capacitance in terms of loading. So, if I do a small rearrangement, I get $C_{GS}$ multiplied by $R_{Sig}$ are assuming

that RG is approximately equal to R Sig. And if I do CGD, near, I will get R Sig into 1+ GM times $R_L$', right + $R_L$', why because this RL Prime is coming here, right + $C_L$ times RL Prime. So, I will get $C_L$ times $R_L$' here.

So, if you solve it I get F of H becomes approximately equal to 1 by 2 pi $tau_H$. So, if you put $tau_H$ tau, I will get this as 1 upon 2 pi $C_{GS}$ multiplied by $R_{Sig}$ + $C_{GD}$ multiplied by this whole quantity, suppose this is X + $C_L$ times $R_L$'. So, you see that $F_H$ value depends upon the value of your gate to source capacitance, it also depends upon the value of gate to drain but multiplied by 1+ GM $R_L$ because of Miller multiplication factor and it also depends upon the value of an external load voltage which is seen to us. And that gives you a typically nice idea about the open time constant for this case.

(Refer Slide Time: 22:59)

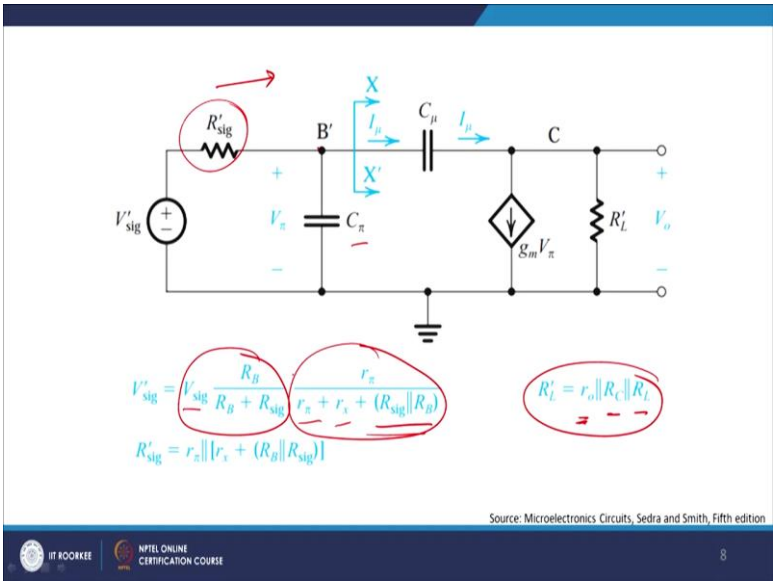



For the case of common emitter mode configuration let me come to common emitter amplifier, we have a BJ T based common emitter amplifier. So, now you have, this is the base, right, this is the collector and this is the emitter. Now, this is your signal, so this is your signal and this is your base resistance which you see, right, $R_B$. So $R_G$ there is analogous to $R_B$ here and $R_0$, $R_C$ and $R_L$ are effective load in the output side.

So, this is your actual device, this is the actual device, where this is the base, this is the base terminal, emitter terminal and collected terminal and we have got 2 capacitance here, $C_{pi}$ and C mu, these are basically the capacitances, depletion capacitances between emitter base, collector and base emitter. So, emitter base is $C_{pi}$ and base collector is $C_\mu$. And GM times $V_{pi}$

is basically the amount of voltage or the current flowing through the bipolar junction transistor, through the BJT, right.

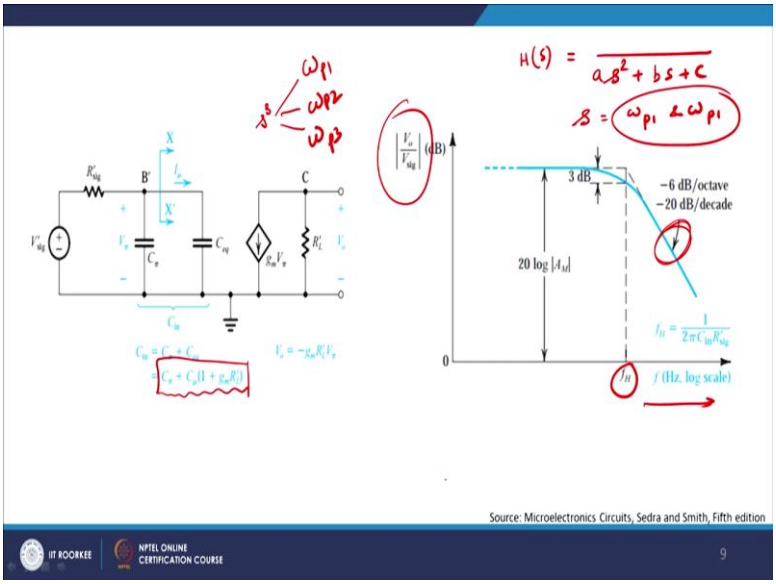
(Refer Slide Time: 23:48)




Now, if you solve it, it is exactly the same as the previous case, the only thing here is that you now have, if you look from the left-hand side, which is this one for common emitter base configuration. I get $V_{Sig}$ multiplied by $R_B$ upon $R_B$+ $R_{Sig}$. This gives you sort of voltage divider network and gives you effective value of voltage seen from the base side of the common emitter configuration divided into $R_{pi}$ Upon $R_{pi}$ + $R_X$.

So, this is again a voltage divider network which you see here. $R_{Sig}$ + $R_B$, which is $R_{Sig}$ is basically is this part, right parallel to $R_B$, $R_B$ is the base resistance offered by the device and $R_{pi}$ and $R_X$ are nothing but the pi and X values of your emitter, base, collector junction. And $R_L$ is basically $R_0$ $R_C$ $R_L$. So, collector resistance, load resistance and the resistance offered by the device itself, right. And that gives you the value of $R_{Sig}$' and VSig Prime.

(Refer Slide Time: 24:49)




And with this if you plot again, exactly the same thing as your previous case, so common source and common source MOS amplifier and common emitter bipolar transistor, high-frequency looks almost the same, qualitatively as well as most quantitatively. So, if you plot again, therefore gain versus frequency and you see that your cut-off frequency $F_H$ comes out of the again somewhere here, which depends upon the value of $C_{in}$ an $R_{Sig}$' , right.

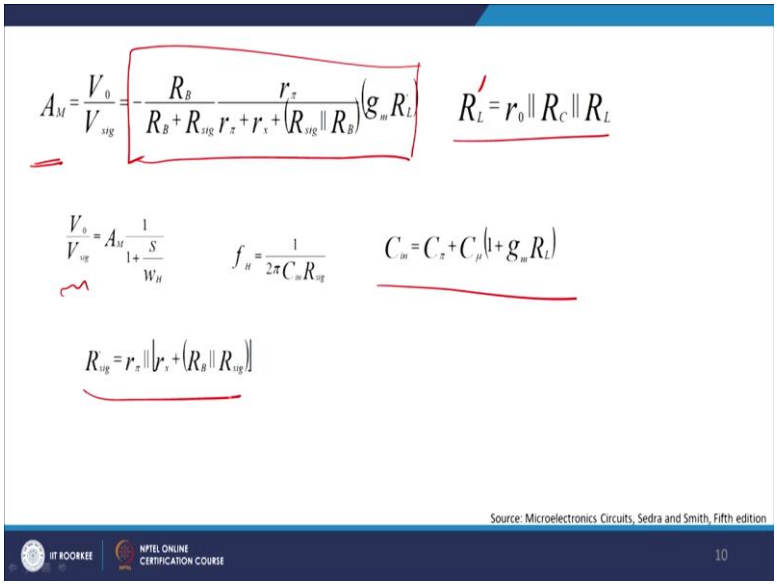
And this is again a Miller capacitance which you see, Miller indices, maybe the next module I will just give you a small insight into the Miller capacitances because this is quite an important term which we will be encountering time and again, especially in amplifier circuits. And I might go into the details of this one at a later time, maybe in the next module.

So, I get 20 dB per decade drop here which you see, this gives me a drop which is quite substantial drop in terms of voltage gain. If you therefore see and try to find out one important point which you should know or you should be able to find out is that as I discussed in the starting of the lecture, that typically if you take up a transfer function H(S) and you do have second-order transfer function say $AS^2 + BS + C$ then you typically have 2 poles, omega $_{P1}$ and omega $_{P2}$.

Why, because this is a square term here, so there will be 2 poles are associated with it. Similarly if there is a cubic term in the denominator, you will have 3 poles associated with it omega $_{P1}$, omega $_{P2}$ and omega $_{P3}$, right. So, these poles will depend upon, the placement of

these poles will depend upon the frequency at which they are operating in Sigma G omega plot.

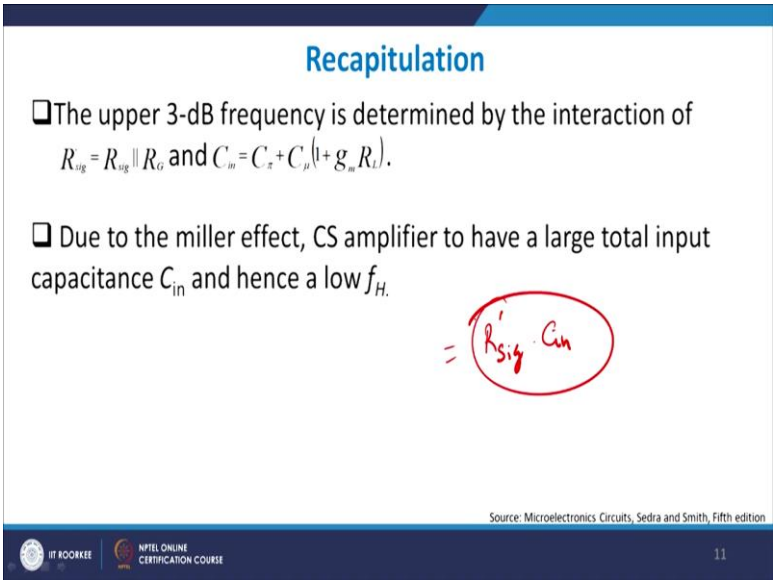
(Refer Slide Time: 26:37)




So, same again, I am just repeating the same thing which I did earlier. That I get $A_M$, the mid-frequency gain is given by this whole quantity multiplied by GM times RL Prime, right, with the negative sign because again here a 180 degrees phase shift, RL, this basically RL Prime, so $R_L$' will be equal to $R_0$ in parallel to $R_C$ parallel to $R_L$. And I get this, again $V_0$ by $V_{Sig}$ I get $C_{in}$ equal to this much and I get $R_{Sig}$.

So, I wanted to just give you an idea that if you take MOS-based amplifier in CS configuration or BJT in CE configuration, the behaviour is almost the same qualitatively as well as quantitatively, right. And the primary pole, the first poll or the major poll comes because of $C_{in}$ times $R_{Sig}$, right. Because of this you will get frequency which is a 3 dB frequency available to me, right.

(Refer Slide Time: 27:35)




So, let me recapitulate the upper 3 dB frequency is determined by the interaction of as I discussed with you $R_{Sig}$ and $C_{in}$. As I discussed with you, therefore the typically your 3 dB bandwidth will be, 3 dB cut-off point will be determined by the product of $R_{Sig}$ and $C_{in}$. So, this will be the product which will determine the $F_H$ value.

Now, one important point which we have also studied this time around is because of Miller effect, we will take up this Miller effect in detail, maybe next module. If we take the Miller effect, then CS amplifier or even BJT shows you an exaggerated value of your output capacitances, which is multiplied by a typically factor of GM times $R_L$.

So, your effective value increases and that is the problem area. That though you want to increase again by increasing the value of transconductance of the device, you end up having also a larger value of your input, effective value of your input capacitance and therefore, your cut-off frequencies also get reduced drastically and you will not get a large band of constant gain feature.

And that is the problem area which people face as far as this specially with Miller capacitance is concerned. With this we have almost finished with the concept of this thing. Next time when we come back, we will discuss source follower and a part of Miller effect to give you a feeling about what Miller effect is and how does it influence our high-frequency response of CS or CE or source follower amplifier. Okay. Thank you very much.

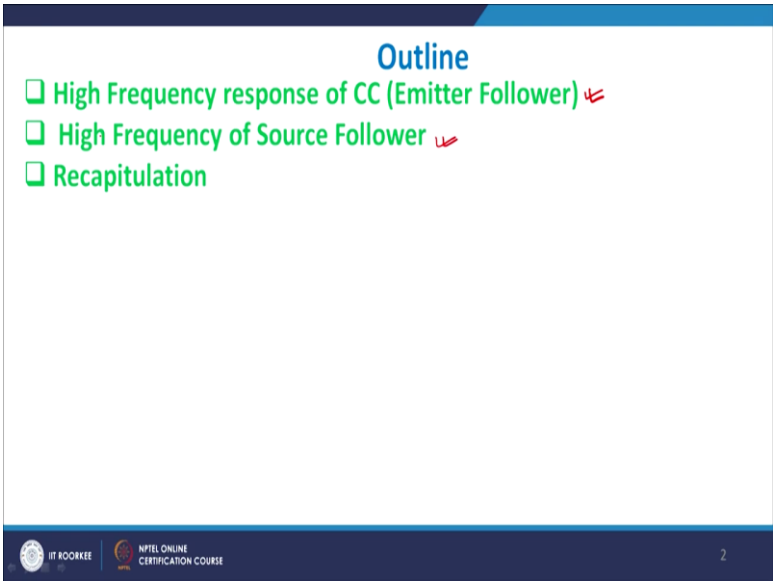
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−43**
**Frequency Response of CC and SF Configuration**

Hello everybody and welcome to the NPTEL online course on Microelectronics: Devices and Circuits. We start today's module, which is named as frequency response of common collector and source follower. So, 2 configurations we will be taking up today, as far as this module is concerned. In our previous model we have looked into the common emitter configuration design as well as common source design when we do CMOS configuration. And we try to find out the high-frequency plots from that using Bode plot.

And we also saw that at high-frequency there is a degradation in the characteristics and that is because of the coupling capacitance. We also saw that there is a limit or there is a sort of a marker and that is basically a 3 dB crossover point where the gain falls by 3 dB. And at that particular point you will have a crossover between mid-frequency and high-frequency range.

After that we do have a 20 dB per decade drop in your gain and that is what we have learnt through our understanding. In our previous section we also have seen the concept of Miller capacitances being used quite often. So, today I will take up a bit of understanding of what is Miller capacitances and then we will follow the frequency response of common collector and source follower technique.

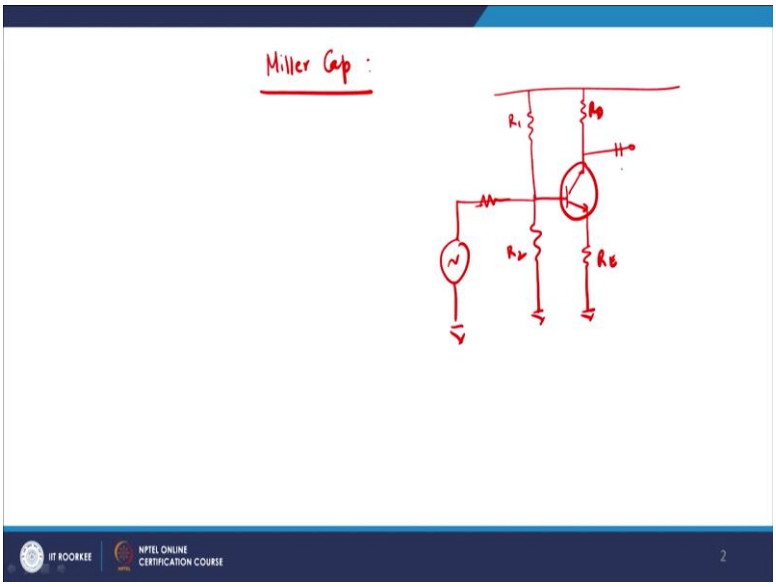
(Refer Slide Time: 2:04)



So, the outline of this module is that we will do an emitter follower and then we will do a source follower, both in high-frequency region. So, we already know what these amplifiers are, for example source follower we know has got a gain of 1 and typically you take the output from the source of the amplifier and you give input to the Gate side.

And typically, the source follows the source output follows the Gate output. And therefore, it is also referred to as source follower, that is the reason is also referred to as a source follower and then we will recapitulate. But before we move to this portion of the talk where we talk of high-frequency response of common collector.

(Refer Slide Time: 2:48)



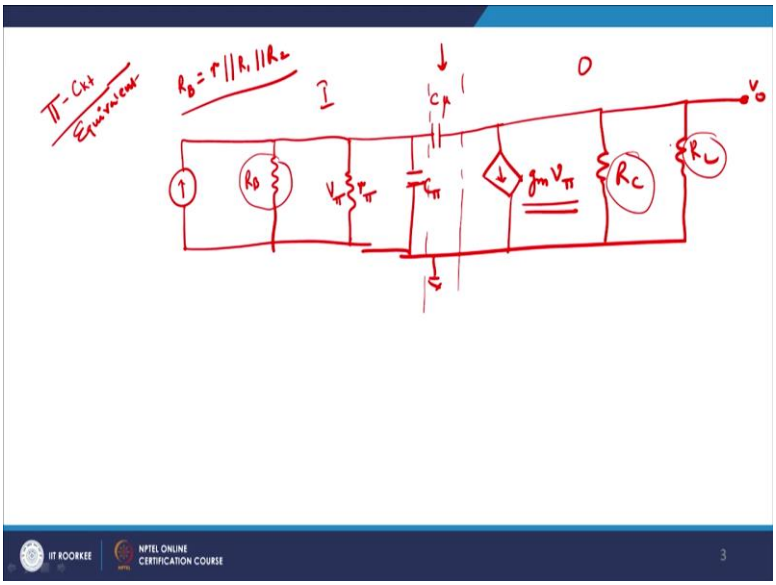

Let me explain to you as I had promised earlier the concept of Miller capacitances, because it is quite an important issue which we will be facing time and again when we will be doing analog design and Miller capacitances is one of the most important capacitances which comes into picture because of open loop gain of the system. Now, the problem is that the Miller capacitance has got a problem in the sense that if you want to increase the gain, open loop gain of the system, you end up having also a larger Miller capacitances. So, optimisation is a very important fact. So, we will take up Miller capacitances from the point of view of, let us suppose common emitter based BJT in voltage divider technique or voltage divider networks.

So, if I plot the graph of voltage divider, it looks something like this and then, common emitter if we take up, then we come to this and then we have this, right and this is $R_1$, $R_2$ and let us suppose you have got $R_S$ here and you are taking an output from here and you have a

blocking capacitor here and this is $R_1$, $R_2$ and this is $R_E$ and this is your $R_D$, this is your common emitter based circuit and you will have obviously resistance here as well as an input signal.

So, you have biased it properly in the active region and then give an input signal which is basically sinusoidal in nature to this base signal and then we will see, we will see its small signal analysis and we will see how its small signal analysis can be done.

(Refer Slide Time: 4:31)




To do that, so we say that we do have equivalent circuit model. It is also known as pi circuit equivalent. So, we will do a pi circuit equivalent and therefore we will see this to be as $R_B$, we will refer to this as $R_B$, base resistance, followed by $R_{Pi}$ and across it you will find $V_{pi}$, which is actually the input resistance and then you have got $C_\mu$ and you have got pi*μ*$C_{pi}$, I will explain to you these terms just now.

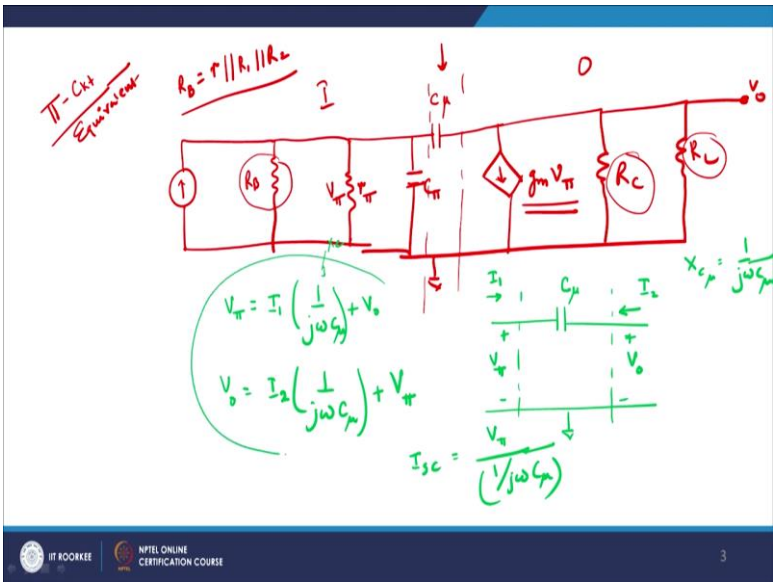
This is equal to $g_m$ times $V_{pi}$, then you have got RC here, which is the collector resistance followed by the external load resistance which is $R_L$ and this is $V_{out}$. So, we are at this stage looking into the fact that, just for practical purpose I will do like this. And this is your $R_B$, $R_B$ is basically your base resistance, which is nothing but parallel combination of $R_{in}$ parallel to $R_2$ and small r is the effective resistance between collector and the emitter side of the circuit.

Now, you see $V_{pi}$ is basically the voltage which appears across the base terminal of the BJT, the common emitter configuration. And who is responsible for giving that, $R_B$ is responsible for finding the value of voltage across the 2 ends and it also depends on the value of $R_{pi}$ which you see. $C_{pi}$ is basically the capacitance seen at this stage between base and emitter,

right because there will be a depletion capacitance as well as diffusion capacitance. And those if we add up happens to be $C_{pi}$.

What is more important at this stage is we should know what is $C_\mu$, right. $C_\mu$ is basically the capacitance which you see, which is basically a feedback capacitance, which connects the input port to the output port of a 2 port network, especially a BJT here. $g_m$ Times $V_{pi}$ is nothing but, because BJT is a current controlled device and therefore $g_m$ Times $V_{pi}$ gives you the current flowing through the BJT. $R_C$ is the collector resistance and $R_L$ is the load resistance for all practical purposes.

(Refer Slide Time: 7:16)

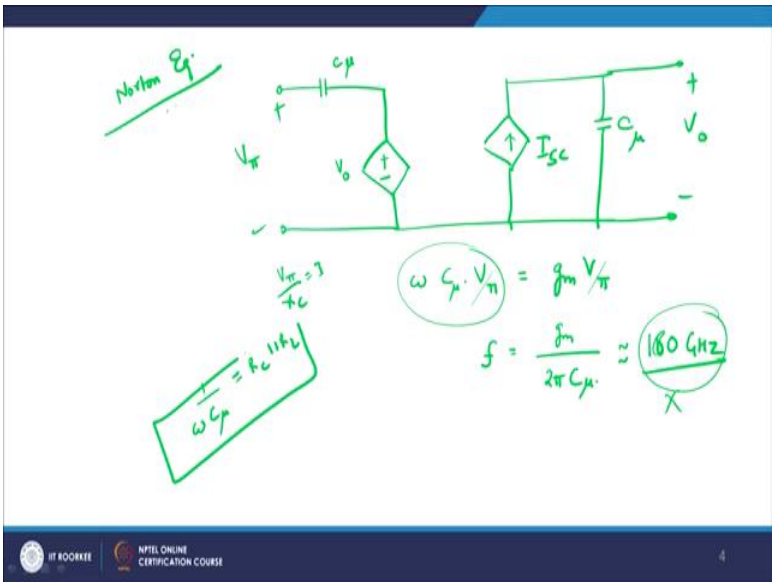



Now, we will like to model this $C_\mu$ with, this is where the Miller capacitance portion will be coming. We will like to model the $C_\mu$ here and let us see how we can actually do a $C_\mu$ modelling here in this case. Now, in this case we can model $C_\mu$ as a 2 port network and we will see how a 2 port network can actually work out and give me a result in this case. So, I have got $C_\mu$ here and I have got current $I_1$ and Current $I_2$, both going into the black box and we have got $V_{pi}$ as the input voltages and $V_0$ is the output voltage.

So, basically a 2 port network as we have already seen, you must have already done in your network theory classes and you have got $C_\mu$. Now, I can do I-V relationship, knowing that the effective impedance available to you is basically 1 by j $\omega$ $C_\mu$. So, $XC_\mu$ will be equal to 1 by j $\omega$ $C_\mu$, right. And that is what we will be writing now as far as this is concerned. So, we can write down V of pi to be equals to $I_1$ into 1 by j $\omega$ $C_\mu$ + $V_0$. And we can also write down $V_0$ to be equal to $I_2$, right 1 by j $\omega$ $C_\mu$ + $V_{pi}$.

So, you can see if you look very carefully, 1 by j ω $C_\mu$ is nothing but $X_C$. So, this is nothing but $X_C$, so $X_C$ multiplied by I1 is the voltage drop across the objective capacitance which is $X_C$ here and we define $I_{SC}$ which is basically the current flowing through $C_\mu$ to be equal to $V_{pi}$ divided by 1 by j ω $C_\mu$. Right, where $V_{pi}$ is the input voltage which is visible to the device, right. So, with these 2 networks or this 2 port network. we can draw the Norton equivalent circuit and we get something like this.

(Refer Slide Time: 9:08)



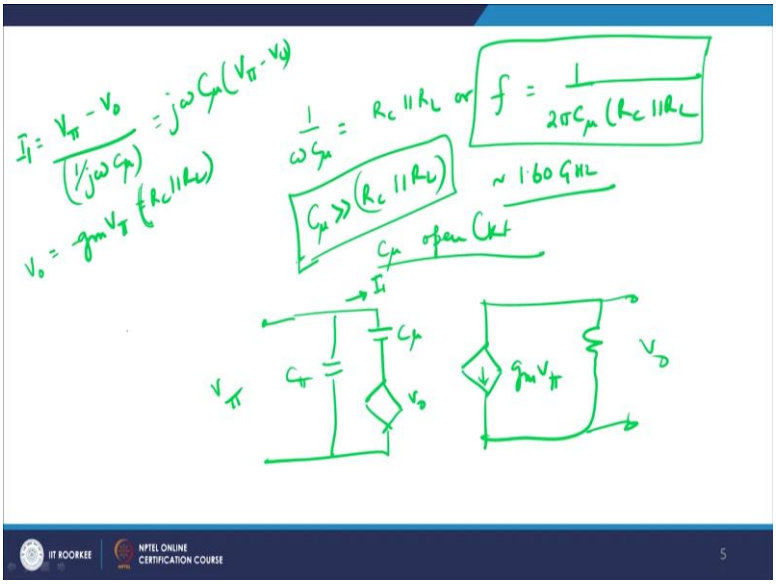

We have got $C_\mu$ here, this will be a voltage source, right, voltage source $V_0$ which is given by a voltage source here, this is equal to $C_\mu$ and then on this side we have got $I_{SC}$ as the current source because of the voltage source $V_0$. And parallel to this you will have, of course, $C_\mu$, right and then you will have $V_0$ here, this is your $V_{pi}$. And then you have got $I_{SC}$ which is basically the current source. So, this is basically my Norton equivalent circuit for the common emitter based configuration.

And as a result you see this overall profile which you see in front of you. If you look at this overall profile here, then if I assume that, if this is true which you see, then I can safely write down that ω times $C_\mu$ multiplied by $V_{pi}$ is nothing but equals to $g_m$ times $V_{pi}$ because both of the currents, right ω times $C_\mu$ into $V_{pi}$ is nothing but $V_{pi}$ by $X_C$. So, this whole thing is basically your $V_{pi}$ by $X_C$, $X_C$ is the reactive capacitance, which gives you the current. And that must also be equal to $g_m$ times $V_{pi}$, where $V_{pi}$ is the input voltage with you. So, therefore, if you look very carefully, and this will be 2 pi F, so $V_{Pi}$, $V_{pi}$ will get cancelled out and F will

be equal to $g_m$ upon 2 pi times $C_\mu$, right. Now, if you put a typical value of $g_m$ and $C_\mu$, this comes out to be approximately 160 gigahertz, right, 160 gigahertz. Now, here comes the problem, therefore BJT, obviously cannot work at 160 gigahertz because it is a very large frequency which is seen to you, therefore it cannot work in such a large region of operation or range of operation.

So, what we should do? What we should do is we need to therefore find out what is the condition under which 1 by $\omega$ $C_\mu$ is equal to $R_C$ parallel $R_L$ which means that the effective capacitive reactance and 1 upon that, of course, 1 upon j $\omega$ C and the resistive, pure resistive network, when they exactly become equal to each other, we say that that is the point where you will have achieved the stability.

(Refer Slide Time: 12:00)



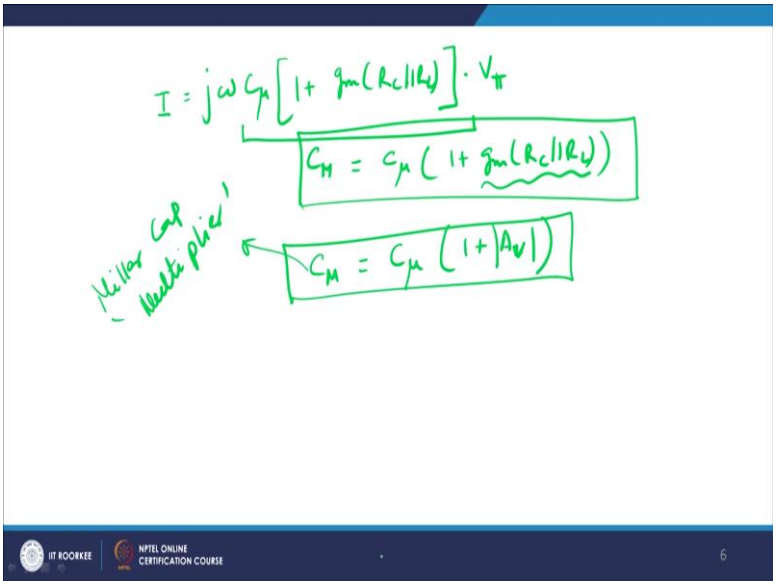

Under such a criteria, we see that 1 by $\omega$ Times $C_\mu$ equals to $R_C$ parallel to $R_L$ or f equals to 1 by 2 pi $C_\mu$ into $R_C$ parallel to $R_L$. Right, this is the frequency of operation of the common emitter based BJT. And if you put approximate values of $R_C$, $R_L$ and $C_\mu$, I get this to be approximately equal to 1.60 gigahertz. So, we have drastically reduced the frequency of operation from 160 to 1.60, about 100 times decrease has been there in terms of the operating frequency.

And this can happen, provided $C_\mu$ is much larger as compared to $R_C$ parallel $R_L$, right. So, you have to ensure that the capacitive reactance of the device is much much larger as compared to $R_C$ parallel $R_L$. Right, and therefore I can safely assume that if $C_\mu$ is basically an open circuit therefore, right. Why, because it is much much larger compared to $R_C$ parallel $R_L$

and therefore, it is open circuit and therefore if I draw the effective circuit diagram, I get something like this, this is $V_0$, this is $C_\mu$ and then you have got $C_{pi}$, right and then this goes to the input side.

And then in the output side you have got $g_m$ Times $V_{pi}$ and then you have got resistance RL and this is your output voltage $V_0$ and this is $V_{pi}$. And from here, this is a current I which is flowing through the device, $I_1$. Then I can safely write down $I_1$ to be equals to $V_{pi}$ minus $V_0$ divided by 1 over j ω C mu because that is what we get, it must be equal to j ω C mu into $V_{pi}$ - $V_0$. This will go up and you will get j ω $C_\mu$ into $V_{pi}$ minus $V_0$. We also know therefore that $V_0$ is equal to - $g_m$ Times $V_{pi}$, $g_m$ Times $V_{pi}$ multiplied by $R_C$ parallel $R_L$, right.

(Refer Slide Time: 14:21)



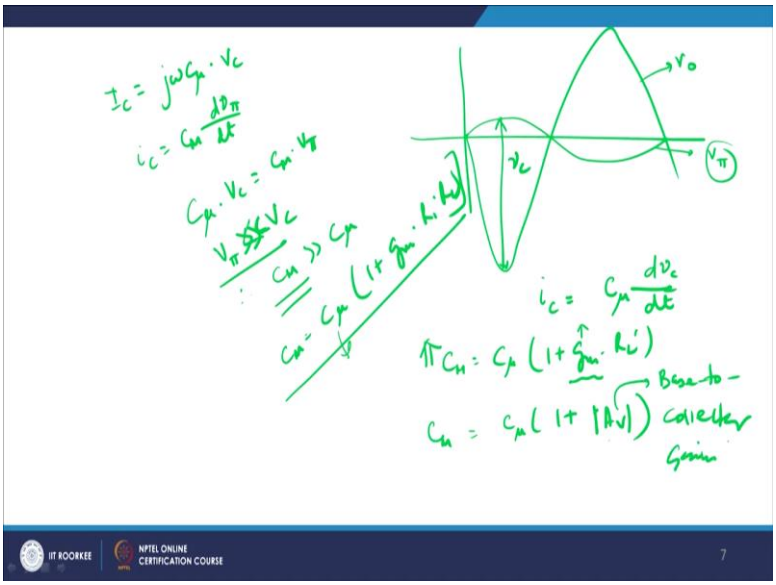

And therefore, I can write down the current I to be equals to j ω $C_\mu$, right into 1+ $g_m$ Times $R_C$ parallel $R_L$, right into $V_{pi}$. And therefore, you see $C_\mu$ which was initially the effective capacitance seen is actually getting translated into a slightly larger value of capacitance and this is basically my CM or the Miller capacitance and is given as $C_\mu$ into 1+ $g_m$ times $R_C$ parallel $R_L$. Fine.

So, this is your CM which you get, this CM is equal to $C_\mu$ into 1+ $g_m$ times $R_C$ parallel $R_L$. And if you look very carefully, this is nothing but the voltage gain and therefore I can write down Miller capacitance as it CM, $C_\mu$ is equal to 1+ $A_V$ mod of that. And this is what is known as the Miller Cap multiplier. Right, so this is the Miller cap multiplier, which occurs because you do have some gain, open loop gain of the system, which gets multiplied with $C_\mu$

and that appears on the output side as CM. This component is basically therefore known as the Miller component.

So, what is the Miller component in BJT or CMOS? It basically is the component of capacitances which connects between the input and output and raises the value of voltage in the output side beyond a particular limit. That is what is known as Miller capacitances. Let me therefore give you physical reasoning behind this one.

(Refer Slide Time: 16:12)




So, let us suppose I have got an input and output and my input is given by the small $V_{pi}$, this is $V_{pi}$, this is my input $V_{pi}$. My output will be obviously 180 degree phase shifted but highly amplified and therefore I get this as my $V_{out}$, this is my $V_{in}$ and this difference suppose we refer to this as $V_C$ or the voltage across the capacitance C. I can safely write down IC to be equal to C of μ $dV_c$ by dt, right. We can also write down $I_C$ to be equal to j ω $C_\mu$ into $V_c$ as we have discussed just now.

$I_c$ is also equal to CM times $dV_{pi}$ / dt, so if you equate these 2 together, I get $C_\mu$ Times $V_C$ equal to CM times $V_{pi}$. Right. Now, we already know that V pi is much much larger as compared to $V_C$. And therefore CM will be obviously, sorry, V, this is true that $V_M$ is much much smaller as compared to $V_C$ and therefore my CM will be much larger as compared to $C_{pi}$. And that is true also because CM is nothing but $C_\mu$ into 1+ $g_m$ times $R_1$ into $R_2$ or $R_1$ parallel to $R_2$, right.

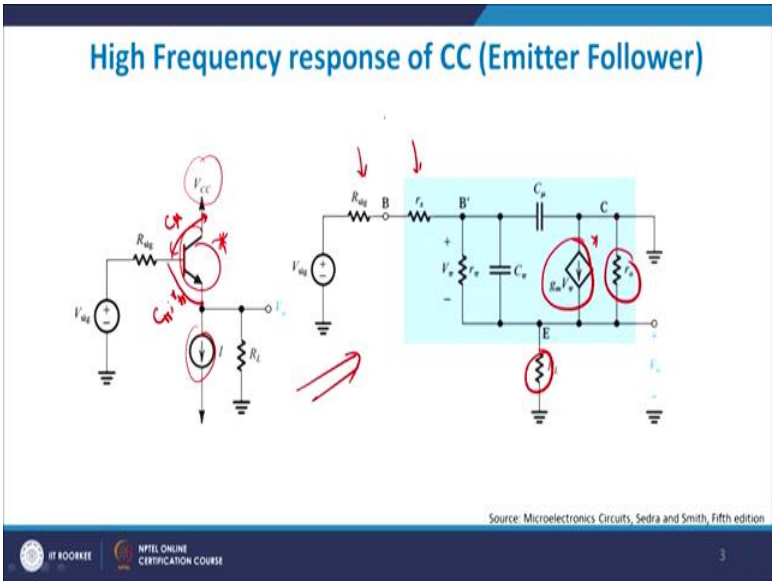
And you get overall Miller capacitances available to me, depending on the value of your feedback capacitance. This is the basic criteria or the basic concept of a Miller capacitance

and as we have seen just now that Miller capacitance gives you an enhanced value of your this thing, of the output. But then you see, as I discussed with you, CM will be equals to $C_\mu$ into 1+ $g_m$ times $R_L$'. So, if you want to increase the gain or bandwidth or you want $g_m$ to be large, you also end up having a larger CM.

A larger CM implies that your speeds will be restricted drastically and therefore, you have to do a sort of manipulation, a sort of optimisation between input and output in this case. And they have to make them look such that they are almost equal to each other. So, this is basically gain and it is also written as $C_\mu$ into 1+ mod of $A_V$, right. And this equals to CM, this is also referred to as base to collector gain.

So, this base to collector gain which you see as $A_V$ and this gain is typically of the order of 10 to the power 5, 10 to the power 6, depending upon the type Op-Amp which has been chosen in this case. With this knowledge we have understood what is the Miller capacitance, what is the origin of Miller capacitance and how are Miller capacitances related to the overall capacitances in the system. Let me switch back to today's course and explain to you how it works out.

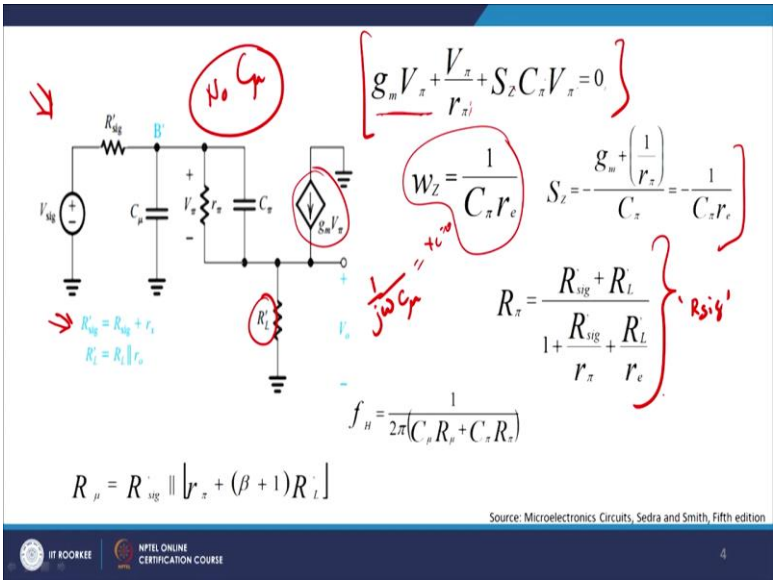
(Refer Slide Time: 19:42)




For example if you are doing common collector or emitter follower design, then this is what it looks like come on the emitter side you have the current source and the collector is connected to $V_{CC}$ and the signal is inserted to the base side of the BJT. If you plot its equivalent circuit diagram, it looks something like this. So, $g_m$ times $V_{pi}$ is nothing but the device itself, so this is the device itself, right. Who is surrounding the devices, the collector is connected to $R_0$, it is basically an open circuit resistance value.

You will have $C_\mu$ and $C_{pi}$ as the resistance between base and collector. So, between base and collector we will have $C_\mu$ and between base and emitter you will always have what is known as $C_{pi}$ and $R_{pi}$. Fine. And this makes our life difficult in the sense that these capacitances become higher and higher at enhanced values of operation or resistance of the device. So, $R_0$ is basically my output resistance of the device, $R_L$ is the load resistance, $R_X$ and $R_{Sig}$ are the input resistance seen by the device when the device is operating in saturation region.

(Refer Slide Time: 21:07)




So if you go back to your previous discussion, what I was doing was $C_\mu$ if you see carefully has been removed. So, there is no $C_\mu$ and the reason is that at such a high, so when you are, so you remember 1 upon j ω. So, when your frequencies are relatively very large, then these are actually almost equal to 0 and therefore they can be shorted and therefore the capacitance have been shorted, the $C_\mu$ capacitances have been shorted.

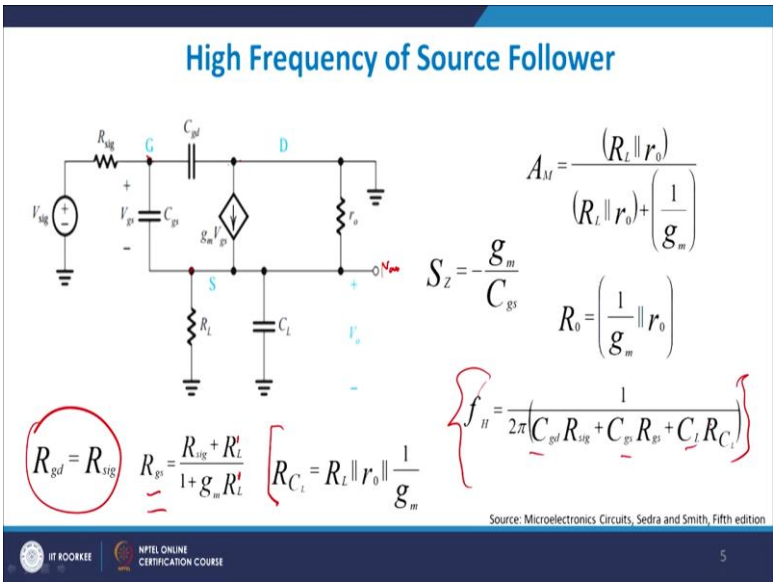
And you get this picture here which you see in front of you, where $V_{Sig}$ is applied signal voltage and $C_\mu$ is the applied capacitance and C of pi is basically the pi value or the input capacitance of the system. $g_m$ times $V_{pi}$ is nothing but the current which is flowing through the devise because this is basically a current control device.

So, I get $R_{Sig}$' to be equal to $R_{Sig}$ + $R_X$, and $R_L$' with equal to $R_L$ parallel to $R_0$, right. $R_L$ prime equal to $R_L$, $R_L$ is the load capacitance seen from the outside world and $R_0$ is the output impedance of the device. So, if you do a small Kirchoff's law solution then I get $g_m$ times $V_{pi}$ which is nothing but the current must be equal to $V_{pi}$, again the current, must be equal to S Z $C_{pi}$ $V_{pi}$ and this almost be equal to 0.

So, the net current should be equal to 0 because before also the current was equal to 0. From there I get ω Z equals to 1 by $C_{pi}$ $r_e$, right. And therefore I can write down f to be equal to 1 by 2 pi $C_{pi}$ $r_e$, right. Similarly, we can write down SZ to be GM +1 by $R_{pi}$ upon $C_{pi}$ is equal to 1 minus 1 by $C_{pi}$ times re, this is the value of SZ. And similarly, $R_{pi}$, which is basically the resistance offered by the other resistances is given by this formula where $R_{Sig}$' + $R_L$' upon 1+ $R_{Sig}$' by $R_{pi}$ + $R_L$' by $r_e$.

So, if you look at $R_{Sig}$ is nothing but the signal resistance offered by the signal to the device itself. So, if you plot this graph or if you try to find it out, I will get $F_H$ equal to 1 by 2 pi $C_\mu$ $R_\mu$ + $C_{pi}$ $R_{pi}$. And this is what very important result which we get that the holding frequency or the frequency where you are holding the mid-frequency gain is basically determined by the capacitances and resistances of the device itself. And that gives me quite an interesting result as far as designing is concerned.

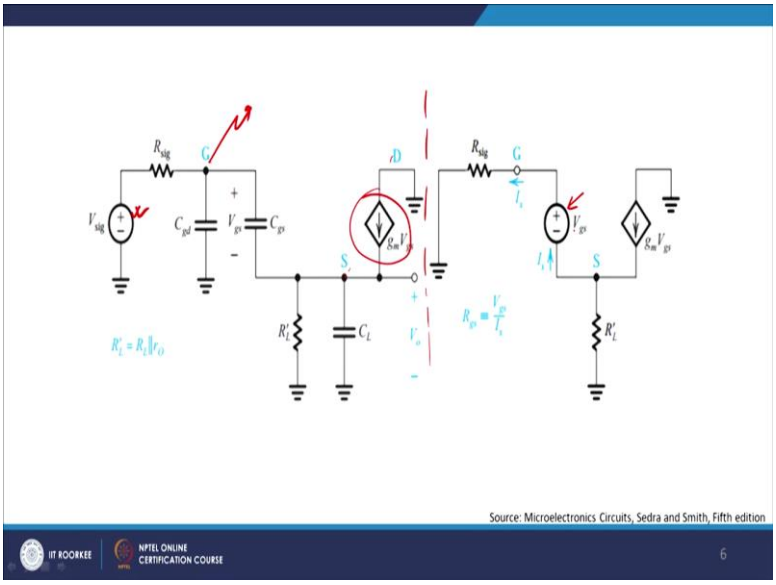
(Refer Slide Time: 23:48)




Now, if I do, so we are finished with common collector, let me do a source follower. As I discussed with you, source follower if you look very carefully, you are actually extracting the voltage from the source itself. So, this is $V_{out}$ here, right and you are extracting it from a source. Similarly, if you look very carefully, $R_{GD}$ equal to $R_{Sig}$, $R_{GD}$ is gate to drain, gate to drain is basically your $R_{Sig}$ and RGS equal to $R_{Sig}$ + $R_L$' upon 1+ $g_m$ times $R_L$', right, these are all primes here which you see.

And from there I get $RC_L$ to be equal to $RC_L$ to be equal to $R_L$ parallel to $R_0$ parallel to 1 by GM, right. And therefore, your $F_H$ happens to be all the combinations of $R_{GD}$, R $C_{GS}$ and CL, load capacitances. So, in a source follower technique or in a high-frequency source follower technique, the resistances offered by the gate to drain overlap and gate to source overlap makes quite a lot difficult calculating the value of the frequency $F_H$. And that is quite an interesting task.

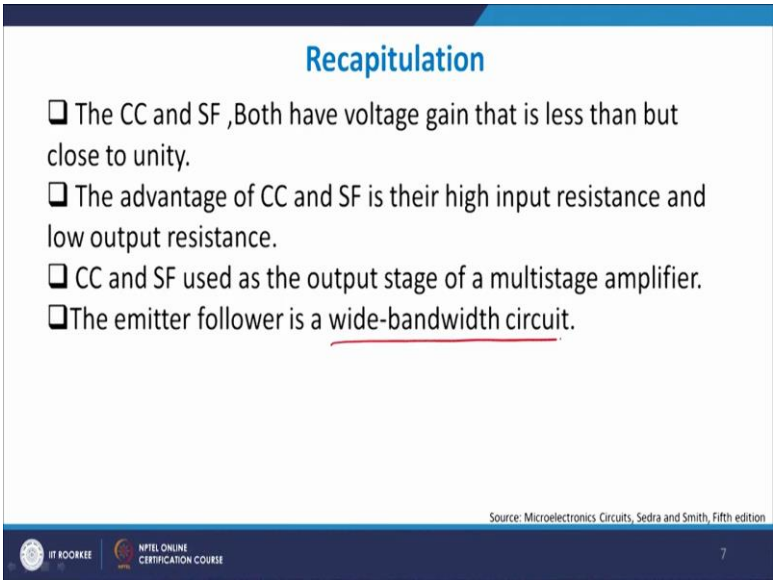
(Refer Slide Time: 25:00)




So, if I just break it down into a smaller piece here and show it to you, then I get $R_{Sig}$ into $V_{Sig}$ into $C_{GD}$. So, this is my gate terminal here and I have got $C_{GD}$ and this goes via the source site to the drain side. So, this is the source and drain and this is the collector, drain current which is flowing through the device given by GM times $V_{GS}$. Right, so if you see and if you see the plot between these 2, they are exactly the same, the only thing is that this $V_{sig}$ has been replaced by $V_{GS}$ and this splits in the input side of the design.

And GM times $V_{GS}$ is the amount of current which is flowing in the output side. So, if you make it $V_{GS}$ large, this also becomes large, so that the same current is flowing through both the arms of the device. And $R_L$ is the tail current source resistance which is being offered by the device itself.

(Refer Slide Time: 26:08)




So, let me recapitulate what we did till now and let me show to you how it works out. If you look very carefully, both have voltage gain less than unity, both common collector and source follower have lesser gain which is quite less than unity. Obviously the advantage of this common collector and source follower is that it has got typically very high input impedance and low output impedance and therefore, matching of signals can be done very well using this technique, using this either CC or SF. Similarly CC or SF are also used as the final stage of multistage amplifier. And emitter follower has got a very wideband gap or bandwidth circuit in reality.

So, it has quite a large bandwidth by which it works fine and gives me a very good result as far as designing is concerned, in terms of both the system design as well as for the design available to us through the overall network of things. So, this takes care of approximately most of the understanding part of our common collector and source follower technique. So, we have seen that common collector and source follower, both have voltage gain less than one because the gain is very small but they are very good impedance matchers because their impedance is dependent on the input and output profile.

The advantage is that both have the high input impedance and very low output impedance and these impedances can be changed by varying certain input parameters. Generally, these CC and SF are used as the final stage of a multistage amplifier. So, typically if you are using a multistage amplifier, say 2 stage or 3 stage, then the last stage will be your CC or SF. Where you get the unity gain but your impedances are properly matched between input and

output. Emitter follower is basically a very wide band gap circuit, not only this which gives me typically a very large bandwidth to a larger extent.

The cost I pay for it is basically that though its bandwidth is large, your gain is relatively small, it is almost unity gain which you see. And therefore the price we pay for the larger bandwidth is the lower gain, right. And that is a n important drawback of this CC and SF. So, let me recapitulate what we did today. I explained to you what is a Miller capacitances and how Miller capacitances play in important role as far as determining the mid-frequency gain is concerned.

And we also saw that if you have Miller multiplier factor, the output capacitances will actually be multiplied by a factor of $1+ g_m$ times $R_d$ which is basically my $A_V$ and will appear in the output side, that makes my life difficult. And thirdly is that if I want to therefore make my capacitance small, I also end up having reducing my gain of the mid-frequency gain of the amplifier, right.

Subsequently we looked into the common collector and source follower technique and we saw the various principles and formulae used for doing that. Where these will be used, we also saw that and what will be the voltage gain for common collector and source follower technique. This we have followed in this module. I thank you for your patient hearing.

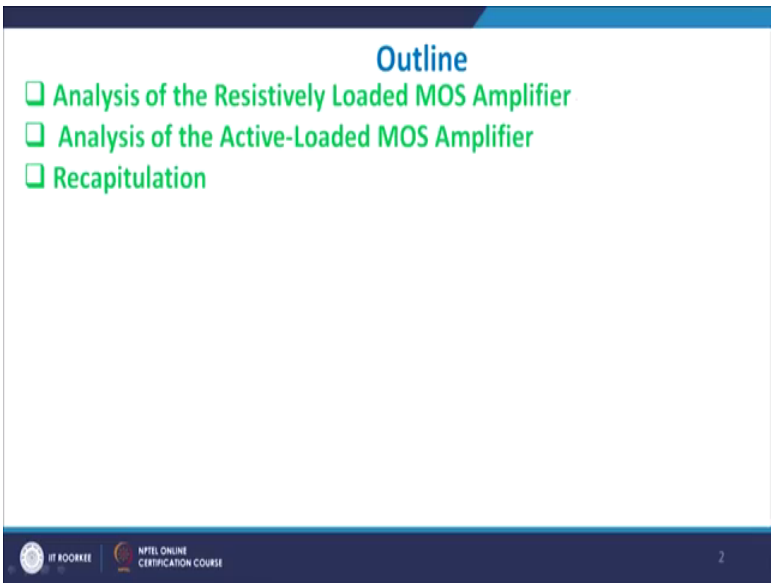
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−44**
**Frequency Response of the Differential Amplifier**

Hello everybody and welcome to the next edition of NPTEL Online certification course on Microelectronics Devices to Circuits. Today we will be taking up frequency response of the differential amplifiers. We have already seen in our previous discussions and modules that a differential amplifier gives you a relatively larger gain as compared to a single stage amplifier. The second advantage is that it is a very good rejecter of common signals to both the inputs of the differential amplifier which means that any noise will be heavily rejected.

And therefore, signal to noise ratios for differential amplifiers are relatively very high as compared to single stage amplifier. However, the price we pay for it is it has got a much larger power dissipation because now you have to drive 2 transistors in 2 loads. Not only that there is also a problem of mismatch between the 2 arms of the differential pair, right? So, if this is small difference in terms of let us say load resistance or transconductance that reflects your common mode gain also.

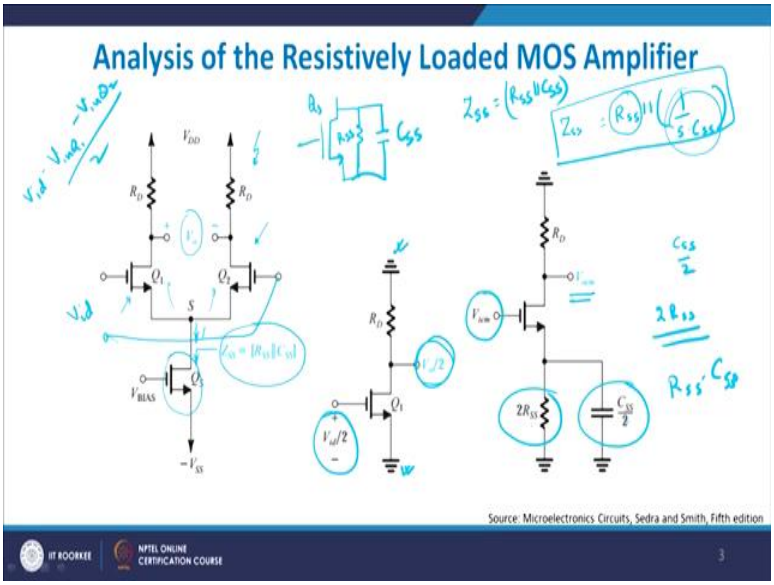
So, in reality you common made common mode gain should be close to 0, ideally it should be 0 which means that any common mode signal will not be at all amplified. And all differential mode signals would be highly amplified. So, $A_{DM}$ upon $A_{CM}$ should be very large quantity but due to these problem areas your $A_{CM}$ also becomes large and therefore, CMRR which is adm by $A_{CM}$ also does not go to a very high value.

(Refer Slide Time: 2:15)




So, we will look into this frequency response of a differential amplifier and the outline of this talk will be analysis of resistively loaded MOS amplifiers. We will take up that first of all and then actively loaded MOS amplifier we will take. So, one is the resistive loading which basically means that you are loading at the pull up stages basically by $r_d$. This was a resistance and your actively loaded MOS amplifier will be that you do have a current mirror based MOS device or loaded device and, of course, then we will recapitulate, right? And then we will go for cascoded stage amplifier designed in our subsequent talk.

(Refer Slide Time: 2:47)




Let me show you how a resistively loaded MOS amplifier looks like. You see the left hand side if you look here is basically you have got 2 input devices q1 and q2, right? And these are all active devices and these are basically MOS devices. You also have a qs which is basically the current source which is giving current to both the arms q1 and q2 here, right? Then we have got RD the resistance here and the output is taken between the 2 nodes here which is referred to as v not. As you can see therefore, that there effective value of impedance seen from node S towards S from the ground is basically $R_{SS}$ is parallel to $C_{SS}$.

What is $R_{SS}$? If you look very carefully $R_{SS}$ is nothing but if you if you plot qs here, right; this is qs. The $R_{SS}$ is nothing but this, right? And $C_{SS}$ is nothing but the capacitances in parallel to it. So, this is your $C_{SS}$ and this is your $R_{SS}$, right? And the parallel combination of that gives you ZSS are also referred to as the impedance seen from the from the from the source end of the input transistor. Now, if you remember $C_{SS}$ can also be written as 1 over S of $C_{SS}$, right; j omega $C_{SS}$. This is $Z_{SS}$, right?

So, whenever we talk about so basically it is a parallel combination and therefore, overall resistance and impedance will be less than the least or will be least of course. Now, what happens is that very high frequencies of operation, relatively high frequencies of operation. This 1 by S of $C_{SS}$ goes low and therefore, it is capacitively dominated and relatively low values of

capacitances or frequencies $R_{SS}$ which is the DC resistance offered by this transistor qs comes into picture, right?

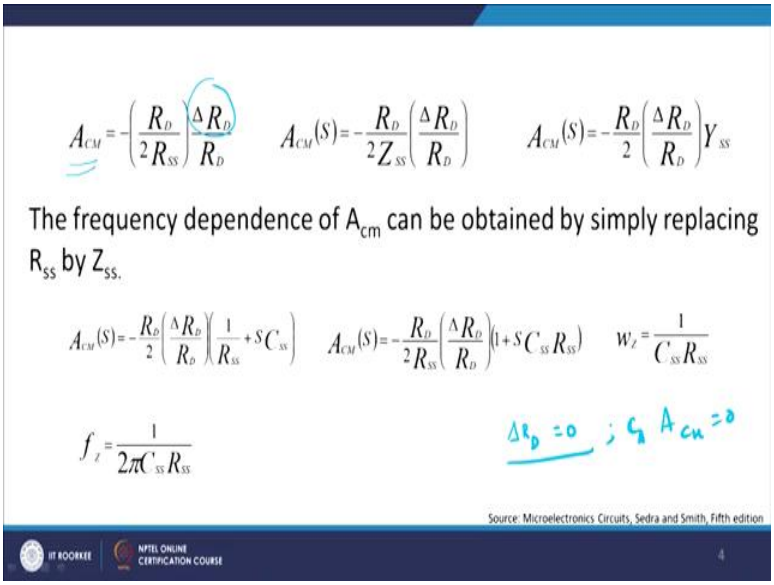
But, nonetheless you will always have a $Z_{SS}$ value which is basically a parallel combination of both comes into picture. Now, to do a single stage, so to understand this we have already seen our previous discussion that we all short our dc biases so $V_{DD}$ is shorted and my ground is also grounded here. And we apply a signal which is $V_{ID}$ by 2, right; half the difference signal between the2. So, remember basically what w do is that if this my input here then I apply my $V_{ID}$. $V_{ID}$ is nothing but v in q1 minus $V_{in}$ q2.

So, $V_{ID}$ by 2 will be this much. So, this much amount of voltage is applied here and we try to get the value of $V_{02}$, $V_0$ by 2 here the output voltage here; which is nothing but the voltage visible at this particular point. Now, as I discussed with you earlier that q1 and q2 need to be initially biased by a dc supply which is basically a common mode supply which will result in these two q1 and q2 working in the saturation region and therefore, this is the $V_{icm}$ which is the common mode signal.

And this is my $V_{OCM}$ which is coming out of it. If you see very carefully the capacitance is half because, the capacitance is half the resistance is double because 2 $R_{SS}$ comes into picture because there are if you look very carefully if this is single device it can be broken into two compound devices and whose capacitances will be in parallel will be in series. And therefore, I get $C_{SS}$ by 2 as the effective capacitance.

However, since they are in they are in series to each other we get 2 $R_{SS}$ as the overall resistance seen from the particular arm. So, if you multiply 2 into see if you see the time constant which you get as $R_{SS}$ into $C_{SS}$ because if you multiply this 2 gets cancelled off and you get $R_{SS}$ into $C_{SS}$, right? So, this is the typical structure of any MOS device which you see.

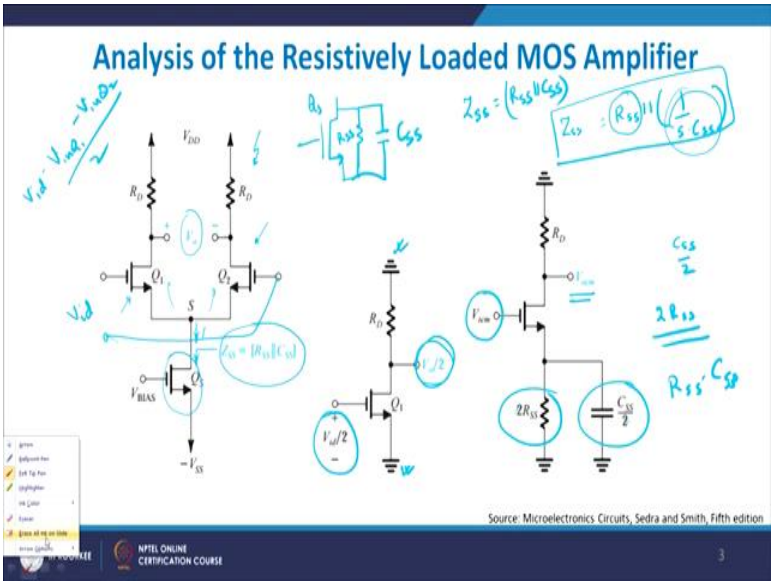
(Refer Slide Time: 6:37)




Now, let us suppose you have mismatch of the drain resistance, right? If there is a mismatch of drain resistance whose value is equal to del $R_D$ then I get $A_{CM}$ the common mode signal equals to $R_D$ upon 2 $R_{SS}$ into del $R_D$ by $r_d$. I am not deriving it here but this is what the value is which you will get.

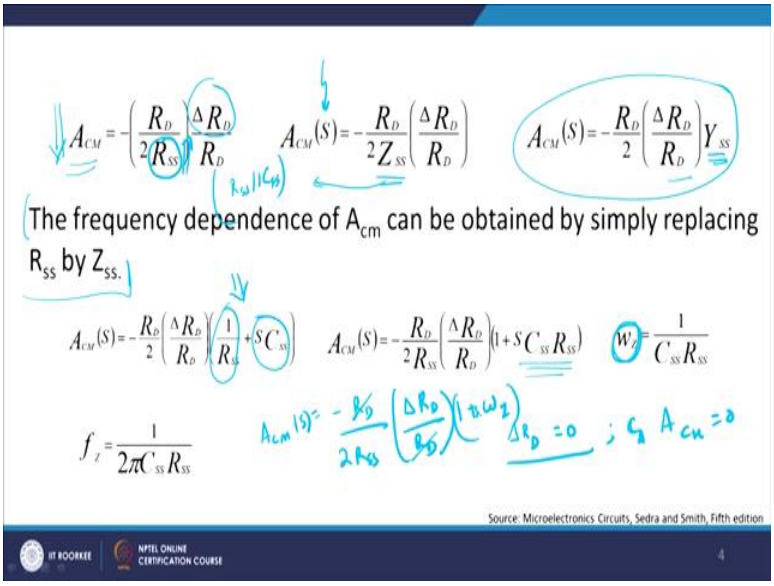
The reason ideally if you look very carefully if you have no mismatches then this del $R_D$ will be equals to 0 and therefore, my $A_{CM}$ shall be equals to, $A_{CM}$ should be equals to 0, right? But if there is some mismatch between the 2 I get this into consideration and so best methodology is you can see is to improve your $A_{CM}$ almost to 0 value a low value is to keep your $R_{SS}$ also very high which effectively means that.

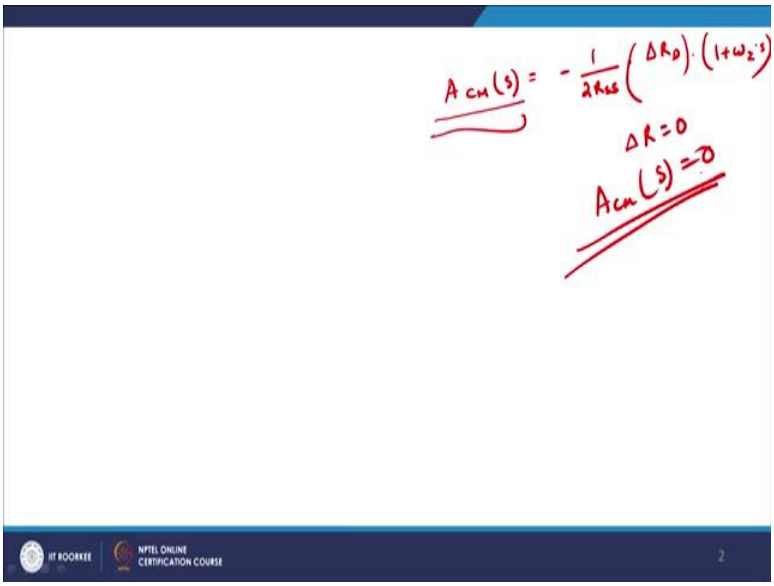
(Refer Slide Time: 7:15)




If you go back to this slide again and then let me rub all these things, then this transistor qs has actually to behave like an ideal current source. If it behaves like an ideal current source then resistance offered output impedance offered by it is relatively very high.

(Refer Slide Time: 7:36)






So, RSS is relatively very high and as a result what you will see is $A_{CM}$ is very low in that case. So, when I go in frequency domain I replace $R_{SS}$ by $Z_{SS}$ and $Z_{SS}$ is given by RSS parallel to $C_{SS}$, right? And we get $R_D$ upon this thing into consideration. Now, 1 upon $Z_{SS}$ is nothing but $Y_{SS}$ and therefore, I get $A_{CM}$ to be equals to minus $R_D$ by 2 into del $R_D$ by $R_D$ into $Y_{SS}$, right?

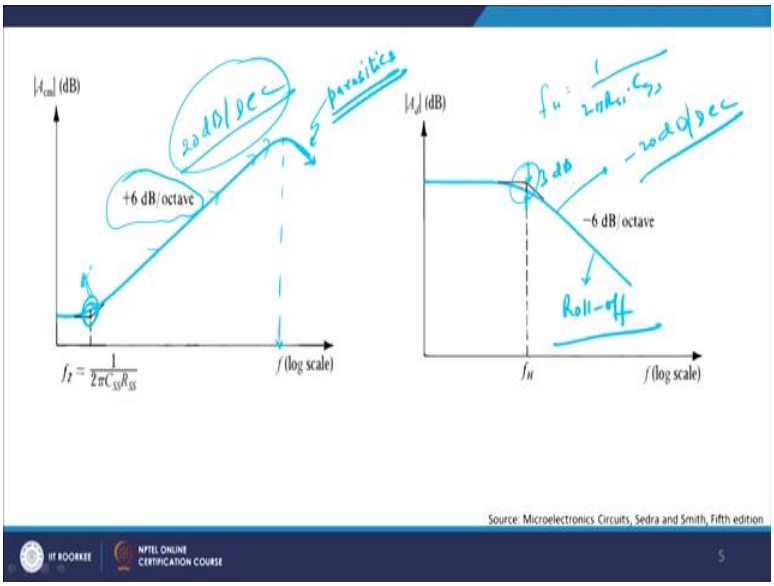
As I discussed with you therefore the frequency dependence of ACM can be obtained by simply replacing RSS by ZSS. That is what we have done strictly in this case. Therefore, in S domain I can just simply simply write down this this quantity $R_D$ basically it is $R_D$ by $R_{SS}$ $R_D$ by 2 into del $R_D$ by $R_D$ into $Y_{SS}$. So, $Y_{SS}$ I can simply write down as equals to 1 by $R_{SS}$ plus S of Scc, right? Because Y is nothing but 1 by Z so, if you remember Z is equal to v by I I by v, right?

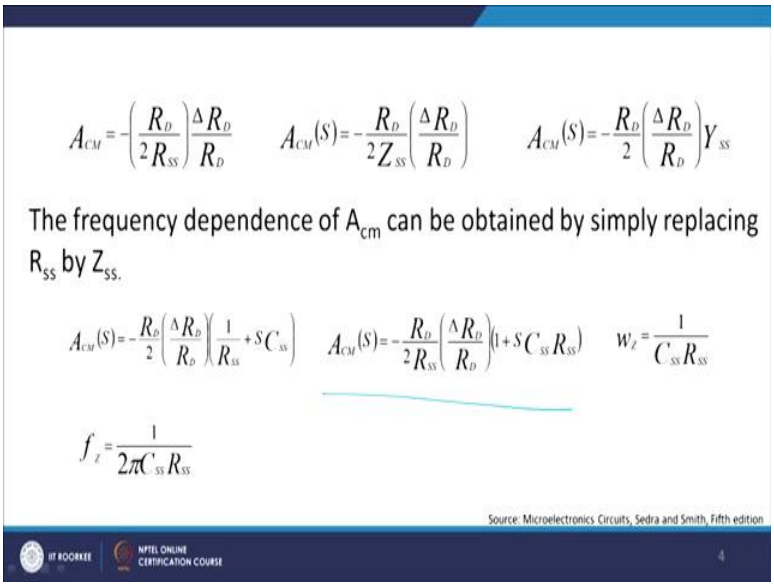
So, I by v is nothing but 1 by r 1 by r. so, this is what you get from here. And since they are parallel to each other we simply add those 2 together. Similarly, if I do a small manipulation here it will be S times RSS here, right? And this RSS will come in the output side and I will get $R_D$ by 2 RSS into del $R_D$ by $R_D$ into $R_{SS}$. Now, you refer to it as omega Z omega Z and therefore, I can write down $A_{CM}$ in S domain to be equals to minus $R_D$ by 2 $R_{SS}$, right? Into del $R_D$ by $R_D$ into 1 plus swz, right?

So, now you see that $R_D$ also gets cancelled out and you are left with the fact that your your $A_{CM}$ of S is equals to minus 1 by 1 by 2 RSS right multiplied by del $R_D$ del $R_D$ into 1 plus omega into omega Z into S. so, this is your common mode differential signal which you see. So, in this case

as you can see del $R_D$ equals to 0 I get $A_{CMs}$ equals also equals to 0, right? And that is quite an interesting idea which you see. This is basically omega Z is basically my 1 by $C_{SS}$ or $f_z$ is equals to because omega Z will be 1 by 2 pi r. So, 1 by 2 pi comes here and I get the frequency to be 1 by 2 pi $C_{SS}$ $R_{SS}$, right? Right!

(Refer Slide Time: 10:29)




Now, with this knowledge let me plot for you the differential gain versus frequency here. And as you can see at very-very low frequencies at typically very low frequencies if you if you just see the sorry, if you look at the point then the gain is differential gain is almost constant independent

of frequency. But as the frequency goes on increasing somewhere near $f_h$ is nothing but approximately 1 by 2 pi $R_{SS}$ into $C_{SS}$.

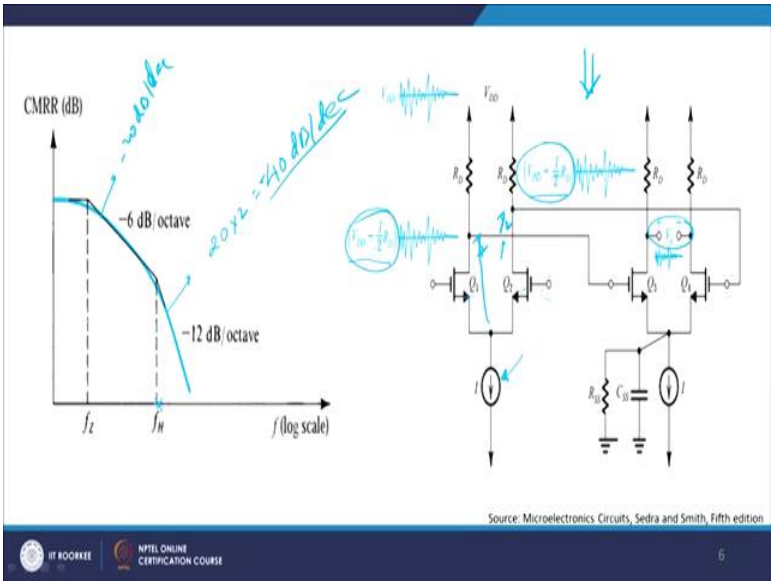
Somewhere, at this point you get a 3 db drop here. This is minus 3 db gain or a drop here. And after this you get minus 20 db per decade drop in terms of differential voltage gain. So, this is this is defined as a roll of which you see. This is known as a roll of, right? So, I have got a roll of available at this particular point. This is your low frequency gain behaving like a low pass amplifier low pass filter actually where your low frequency gain is relatively high and at high frequency the gain starts to fall off.

Whereas if you look at the common mode signal, right; and go back to the previous slide then then you will see that the common mode signal is given by this formula. The gain at least is given by this formula, right; in S domain. So, when S equals to 0 or it is very-very small, your value is very small, right? And that is what you get here. So, what do you get when S is very small this is also very small. Beyond a particular point given by 1 by 2 pi $C_{SS}$ $R_{SS}$ as the S value goes on increasing, $A_{CM}$ value also goes on increasing, right?

And the increase is basically 20 db per decade or plus 6 db per decade octave increase. We will stick to this single formula 20 db per decade enhancement is there. It has been seen that beyond a particular point your actually the $A_{CM}$ value the common mode gain starts to fall down or again there is a roll of point which you see beyond which the frequency it falls off. And this is primarily because of parasitics. So, this is primarily because of parasitics which you get which makes your output capacitance getting loaded.

So, apart from $C_{SS}$ and $R_{SS}$ you will also have an output capacitance which will be in series to overall picture. And as a result overall capacitance will increase and your gain will start to fall down beyond a particular limit. But this is sort of 3 db drop here this is 3 db gain here and the 3 db drop here for the differential amplifier design.

(Refer Slide Time: 13:02)



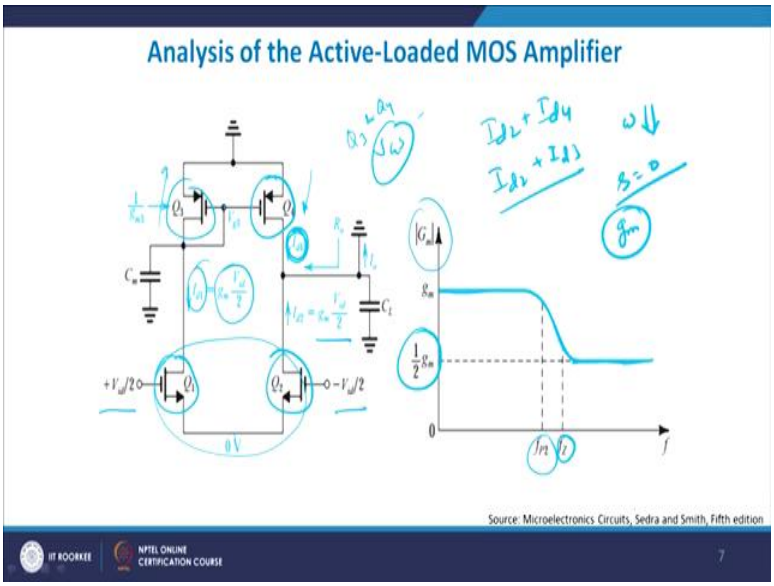

So, if you look at CMRR as I discussed with you we define $f_h$ be equals to position where you will have 1 by 2 pi $R_{SS}$ $C_{SS}$ beyond which you will get actually a 20 into 2 which is basically 40 db per decade drop, right with a negative sign. So, I will get initially 0 and then I will get a minus 20 db per decade, right? And then I get 40 db per decade drop here and this roll off is very high at higher frequency.

Now, if you look at this slide this slide last part here q1 and q2 are basically the transistors which you have given and I have a current source which is the tail current source you give. Depending upon the input value of voltage I get $V_{DD}$ by 2 into $R_D$ which means that half the current flows here and half the current flows here I get $V_{DD}$ minus 2 $R_D$ as the output voltage here and $V_{DD}$ minus I by 2 $R_D$ as output voltage here.

The same voltage are fed into q3 and q4 the second stage of the differential amplifier and as a result you will get a v out value depending on these 2 values here, right? So, as the current goes on increasing I goes on increasing this quantity goes on decreasing this goes on decreasing right? So, how does current goes on increasing by simply make both these transistors switch on at higher values of threshold voltages. And more current will be drawn from I. Its maximum value is I by 2.

You cannot have currents greater than I by 2 but it will be if current is less than I by 2 then you will automatically have voltages here slightly larger, right? And, therefore, forcing q3 and q4 in saturation. As a result we will get an output which is depending on the value of q3 and q4 here which state the q3 and q4 devices are run. Similarly, I have just blocked the current source through $R_{SS}$ and $C_{SS}$ in this case.

(Refer Slide Time: 15:17)




So, now what I do is I replace so till now what we were doing is we were concentrating on the fact that you will have a resistance in the pull up transistor which is rd. But now what I do is I replace RD by q3 and q4 which is basically pMOS loaded actively loaded pMOS and which it is basically current mirror source and the gate. Sorry the drain and the gate of q3 are shorted with respect to each other. Once you short it, it starts to behave like a resistive element and therefore, the current flowing through q3 is replicated in q4, right? Because this gate voltage here and the gate voltage at this particular point is exactly the same. So, gate to source voltage of both the transistors q3 and q4 VGS is exactly the same, right?

If I assume that q3 and q4 are the same transistors with same threshold voltage the current flowing through m which is mu nc oxide w by l $V_{GS}$ minus $V_{TH}$ whole square will also be same and as a result same current will be flowing through q4, right? So, I will have obviously a current flowing through q3 and there will be a same current flowing through q4. So, if I assume that

there is no change here but the current source is ideal current source. So, I have removed it because its output impedance infinitely high.

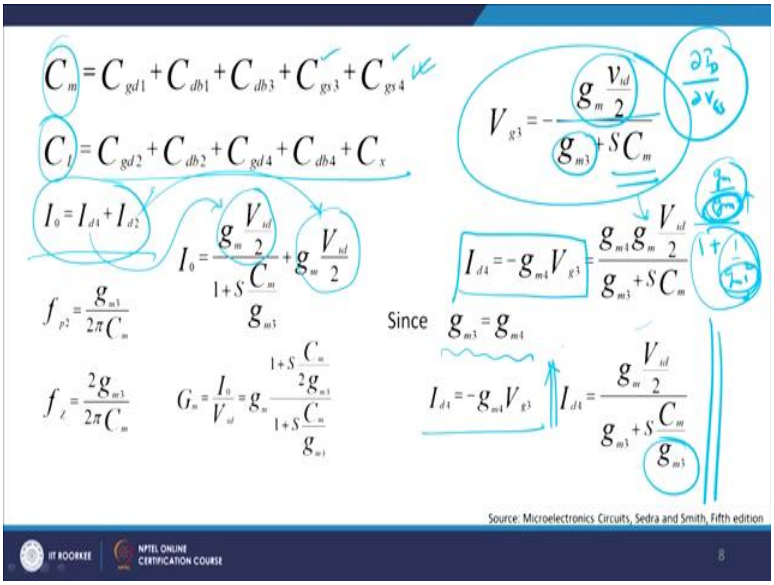
I have given $V_{ID}$ by 2 $V_{ID}$ by 2 here. Once you have given $V_{ID}$ by 2 I get a current which is basically gm times $v_g$. So, $g_m$ into $V_{ID}$ by 2 is 1 which you see. This id1 will come here and $i_{d2}$ will be flowing which is $g_{m2}$ into $V_{ID}$ by 2. I will get id4 which is primarily the current flowing through q4 because of q3 and current mirroring action here. And therefore, the total current will be equals to $i_{d2}$ plus $i_{d4}$. Now, $i_{d4}$ is also equals to $i_{d3}$. So, $i_{d2}$ plus $i_{d3}$ I can also write down as the total current flowing through the device, right?

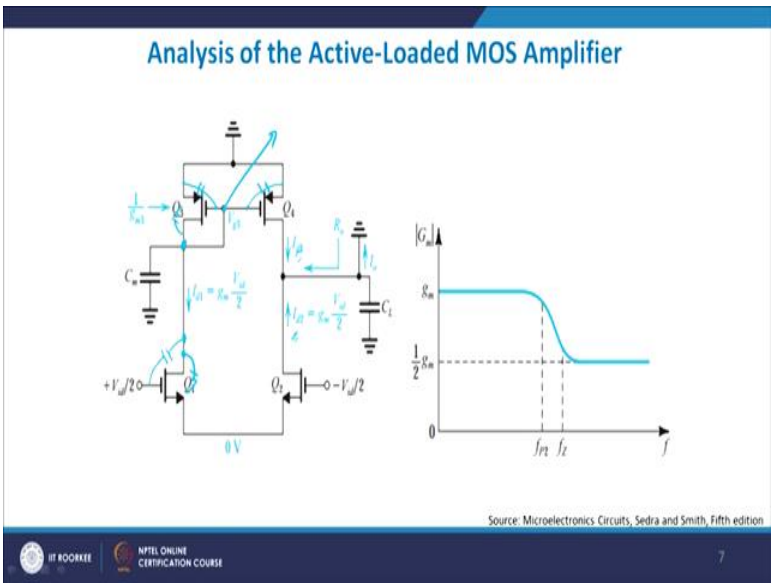
And that much amount of is across the load capacitance, right? So, if you plot capital $g_m$ which is basically the transconductance of q1 and q2 taken together then you will see that till a particular frequency the trans conductance is almost independent of frequency. It is almost a straight line beyond which you will see a certain drop here and then it again remains constant approximately given by $g_m$.

So, I will see initially first gm and then it goes to half $g_m$ in essence, and the reason can be found out from this picture only because very low frequencies of operation when the frequency is relatively low and you do not have any other frequency issues at this particular point, then what happens is that when frequency is relatively low omega is low your S is also approximately equals to 0 in dc case.

Then this capacitance transistors q1 and q2 work in pure active region, right? And they there biased points do not move about. There biased points are fixed. So, they are overall trans conductance is nothing but $g_m$. Once, I increase my frequency and I goes to a value of $f_z$ 1 by 2 pi $R_{SS}$ into $C_{SS}$ then at a particular point the gm starts to drop down, the gain starts to drop down. And the reason being now because you have larger parasitic into picture and the overall capacitance becomes half $g_m$ and therefore gain also drops down at a particular point. That is the reason a differential amplifier your gain at a high frequency starts to drop down, right?

(Refer Slide Time: 18:33)




Let me show to you therefore id. So, total current which is flowing through the device is nothing but $i_{d4}$ plus $i_{d2}$. What is $i_{d4}$ and $i_{d2}$? $I_{d4}$ and $i_{d2}$ is nothing but this is $i_{d4}$ this is $i_{d2}$ and this is $i_{d4}$. So, if you look at $i_{d4}$, this is the total current which flows through a device. This can also be written as gm into $V_{ID}$ by 2, right because that is the current which is flowing through the arm.

This divided by 1 plus S c by $g_{m3}$ sc by $g_{m3}$ because capacitance by transconductance will give you a value which is fundamentally resistance. As a result what you will get is voltage by

resistance will give you a current flow here. And that is must be equal to $g_m i_d$ by 2 which is which is coming from $i_{d2}$. So, this is your $i_{d4}$ and this is your $i_{d2}$, right? And you simply add those 2 together in in reality.

And so what I get from here is I get $v_{g3}$ right? $V_{g3}$ is the gate voltage of 3 transistors is equal to gate voltage of number 3 transistors is this is if you remember del I is, del id del VGS right? So, that is what you get. Del id is nothing but gm times $V_{ID}$ by 2, right? Because half of the half of the voltages applied here divided by you get gm3 transconductance of the device plus S times cm which is basically the total capacitance seen by at node 3 which will include of drain to drain to bulk drain to bulk of 3 gate to source of 3 and gate to source of 4.

So, gate to source of 4, gate to source of 3 and gate to drain of 1. I hope you understand this point, right? As you can see here this is gate to source of 4, right? It is gate to so I have a capacitance here I have a capacitance here, right? You also have gate to source of 3, 4 gate to drain of 1. Gate to drain of 1 basically means that gate to drain of 1. So, this one, right because this is connected to this particular point. So, this particular point is connected to this point. So, this point this point and this point are in parallel to each other.

And that is the reason you add it, right? And then you have got cdb1 and $c_{db3}$. What is cdb1? Drain to bulk of 1, drain to bulk of 1, right? And then drain to bulk of 3. Now, as you can see here all connected through this node, right? So, this node is a parallel node which is to all of them. So, you just have to simply add those capacitances and what do you get is basically my cm. similarly, I get cl to be equals to this whole quantity here, right?

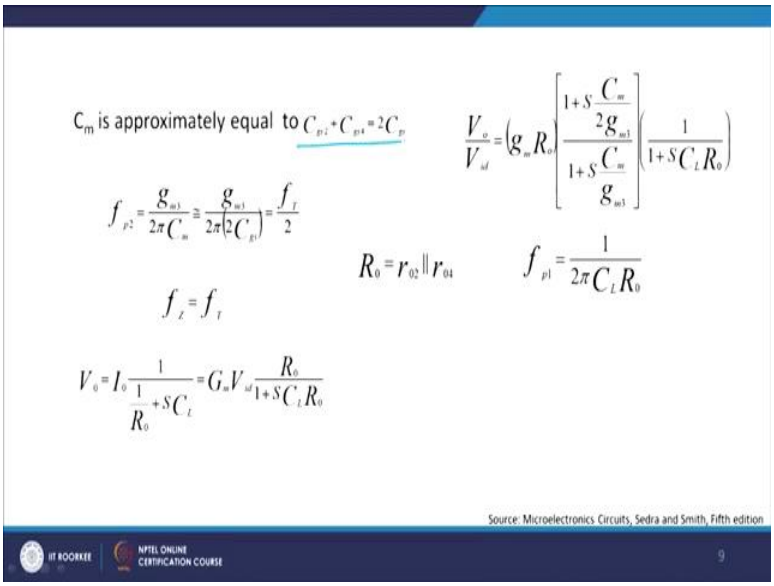
And you need to simply place these capacitances here to get the value total value. So, I get $i_{d4}$ to be equals to $g_{m4}$ into $V_{g3}$. Now, why $V_{g3}$? Because $V_{g3}$ is equals to $V_{g4}$ so, simply you have to multiply the trans conductance of 4 th transistor multiplied by $V_{g3}$ and you get this. So, I can write down gm4 into $V_{g3}$ can be just this 1 can be replaced here and I get this into consideration. As you mean that $g_{m3}$ is equal to gm4 because that is obviously because same w by l ratios same voltages which you see.

I get $i_{d4}$ equals to $g_{m4}$ into $V_{g3}$ which is equals to $g_m$ $V_{ID}$ by 2 $g_m$ $V_{ID}$ by 2 divided by $g_{m3}$ plus S of gm by $g_{m3}$. So, I get $g_{m3}$ into consideration here as $i_{d4}$, right? So, I get $i_{d4}$ to be this value. So, as you can see here as you make your $g_{m3}$ if you if you divide by $g_{m3}$ in the numerator and

denominator what you get is basically $g_m$ upon $g_{m3}$ here, right, divided by 1 plus if you divide it again by 1 by $g_{m3}$ square 3 square. So, if you make your $g_m$ go on increasing this quantity will increase and therefore this quantity will decrease.

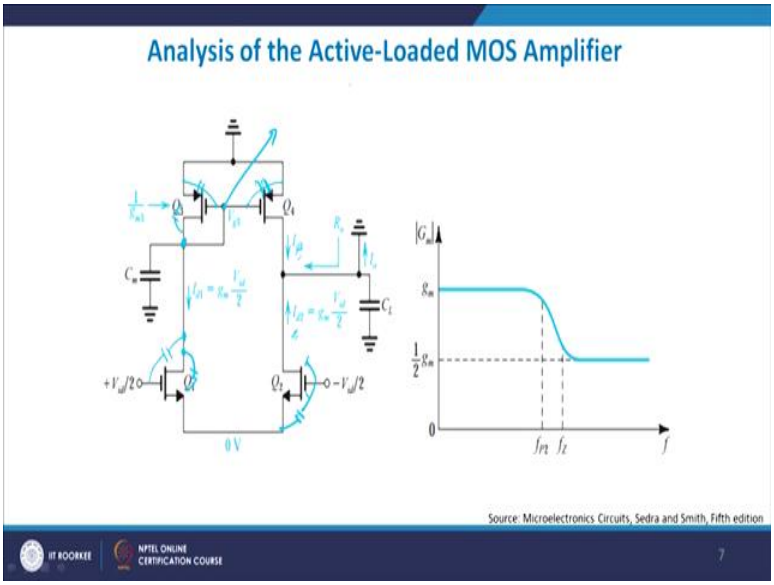
This will also increase and this will decrease and therefore overall decrease will be there. And therefore, this $i_{d4}$ will go on increasing as $g_{m3}$ goes on increasing, right? And that is what we have seen also. The $g_{m3}$ goes on increasing I would expect to see more and more of drain current flowing through the device structure.

(Refer Slide Time: 22:36)



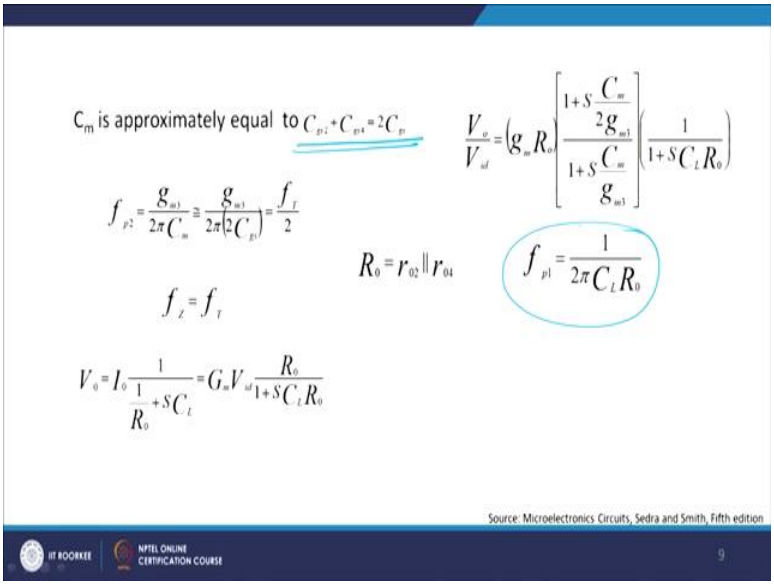

Therefore, as I discussed with you that cm will be approximately equals to the $C_{GS2}$ and gate to source 4. So, 2 and 4 would be added, right? Let me show you why.

(Refer Slide Time: 22:46)




2 and 4 is basically $C_{GS2}$ and $C_{GS4}$. $C_{gs2}$ is gate to source of gate to source of 2 plus gate to source of 4, $C_{GS\,4}$. These 2 will be again parallel.

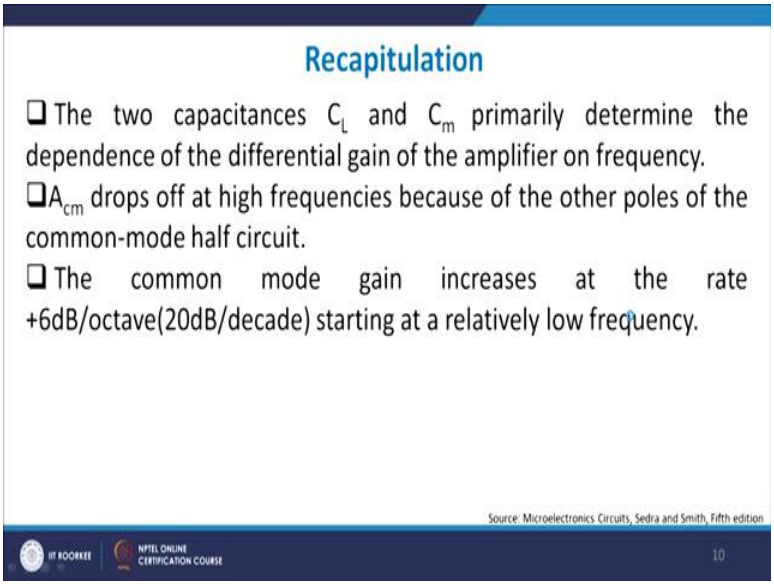
(Refer Slide Time: 23:01)




They will almost be equal to each other and as a result I will get $C_{GS2}$ equal to 2 $C_{GS}$. If you solve again the whole transistor I get finally the value of pole first 4 frequencies 1 by 2 pi $C_L$ times $r_0$

and that is effective value of frequency which is visible to you as far as finding out the values is available.

(Refer Slide Time: 23:18)



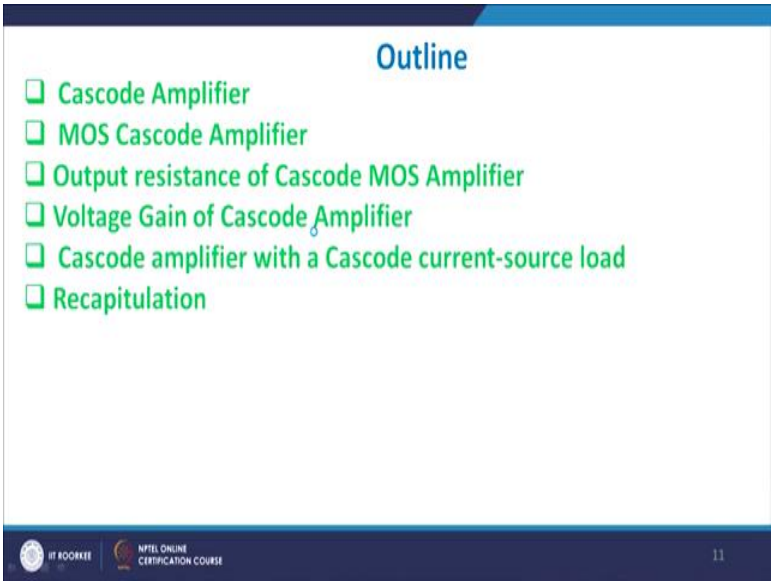

So, therefore, let me recapitulate what we have learnt till now for the for the high frequency domain analysis of this structure. The 2 capacitances are cl and cm. What is $C_L$? $C_L$ and cm are basically the sum of the gate to source and the drain to bulk capacitances of all the 4 transistors which determine the cut off frequency of the differential amplifier, right? So, if you make them high, of course, your bandwidths will be limited and your cut off frequencies will be reduced. $A_{CM}$ very-very important point drops off at a very high it increases at as the frequency increases but then at very high frequencies it starts to drop down, right?

Because there are other issues apart from that you have large amount of parasitic capacitances as well as due to formation of poles for the for the for the common mode half circuit here. The common mode gain increases at the rate of approximately 20 db per decades starting at relatively low frequencies. This we have already seen. The differential gain the differential amplifier gain starts to fall beyond the cut off frequency by approximately 20 db per decade whereas the CMRR, sorry $A_V$, $A_{CM}$ actually is very low at low values of frequencies.

As the frequency increases the common mode gain increases but beyond a particular frequency the gain again starts to fall down and that makes my life difficult as far as this is concerned or

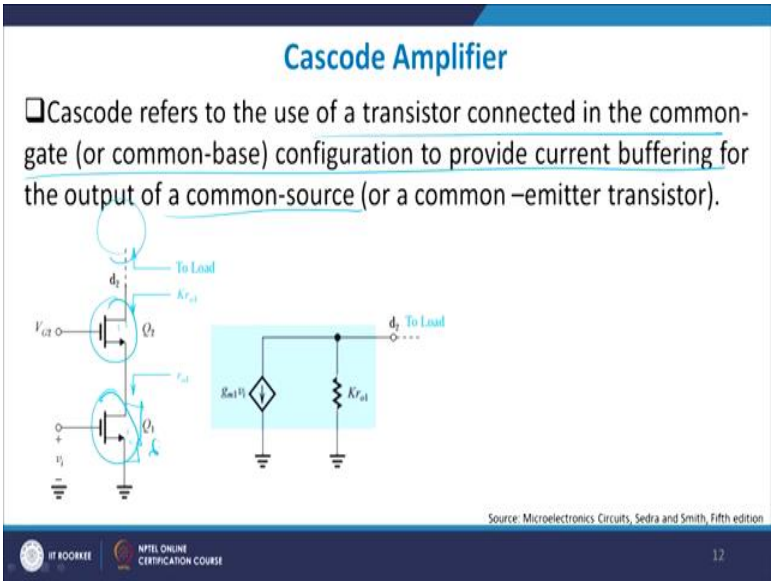
this idea is concerned. Okay! So, that takes care of our understanding of the basic concepts which you see here or the basic idea which you see here.

(Refer Slide Time: 25:07)




Let me now come to the last part of our amplifier design portion or the major portion of the amplifier and we start with first of all cascode and we take up MOS, cascode amplifier and cascode amplifier with cascode current source, right? So, we will be doing that.

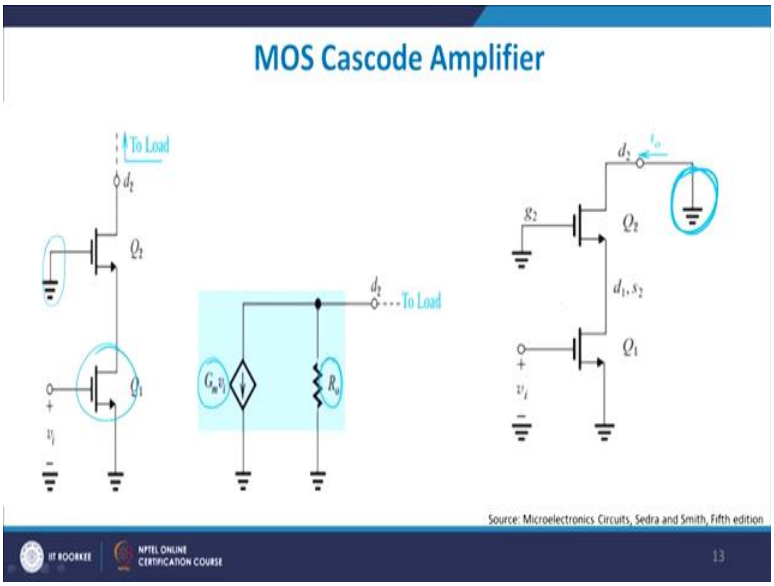
(Refer Slide Time: 25:13)



So cascoded amplifier is refers to the transistor in which a common gate is connected to a to a common source. So, common gate is connected to common source. So, you see this 1 is cascoded. This is my common source, right? And q2 is cascoded with respect to q1, right? And if you look very carefully this is so you have a load here and you are driving $V_{g2}$ and we are giving an input voltage here and you are looking at the output somewhere here at this particular point.

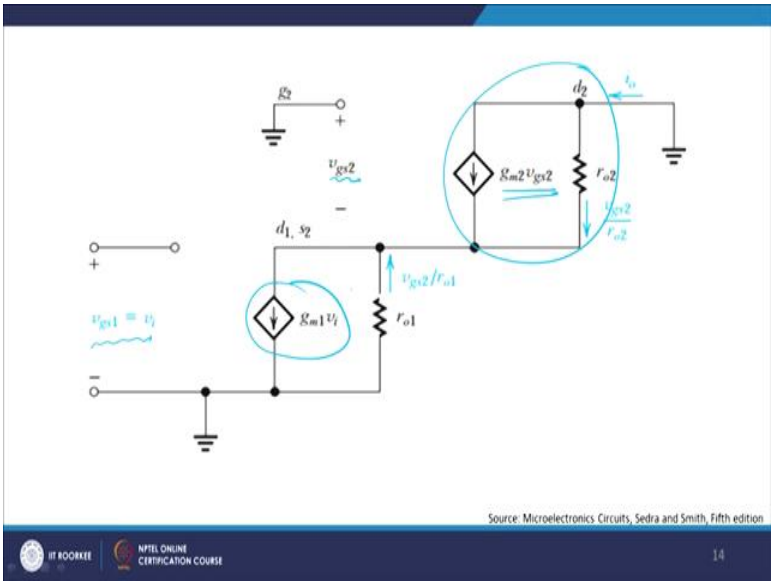
And therefore, I can replace these transistors q1 by $g_{m1}$ $V_{i1}$ where $V_{i1}$ is the input voltage. It is behaving like a current source. Parallel to this you will have a resistances which is effectively the resistances at this 1. So, this is $r_{01}$ which you see, right? And this is the cascode amplifier general design which you see.

(Refer Slide Time: 26:02)



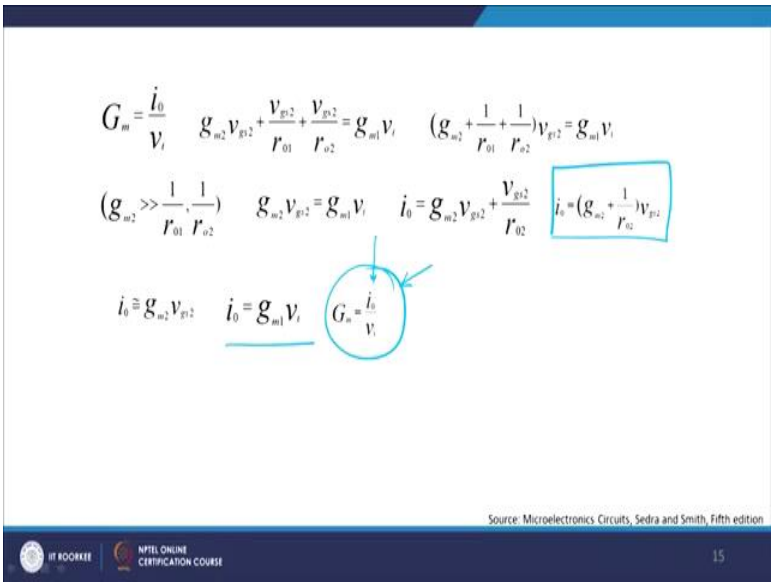

If you look at the MOS cascode amplifier therefore, as I discussed with you can actually you can actually short your gate and to the ground and then in that case it is capital $g_m$ $V_i$ which is basically the trans conductance of q1 into $r_0$ which is the output impedance as you can see from the load. So, as I discussed with you earlier also therefore, that if you short your modelling signal analysis if you short your output dc by here and then try to find the total current flowing through q1 and q 2 which will depend upon the value of $V_i$ which you have inserted into the cascode amplifier.

(Refer Slide Time: 26:39)




If you look very carefully therefore, this is your cascode amplifier which is given by $g_{m2}$ $V_{gs2}$ is the current source and $g_{m1}$ $V_i$ is the q1 which you see and $r_{01}$ and $r_{02}$ are respectively the drain to source resistances offered by these devices. And as you can see here, $V_{GS1}$ is nothing but $V_i$ and $V_{GS2}$ is nothing but shorting it. So, basically $V_{GS2}$ is that which you see in front of you. I am shorting it, right?

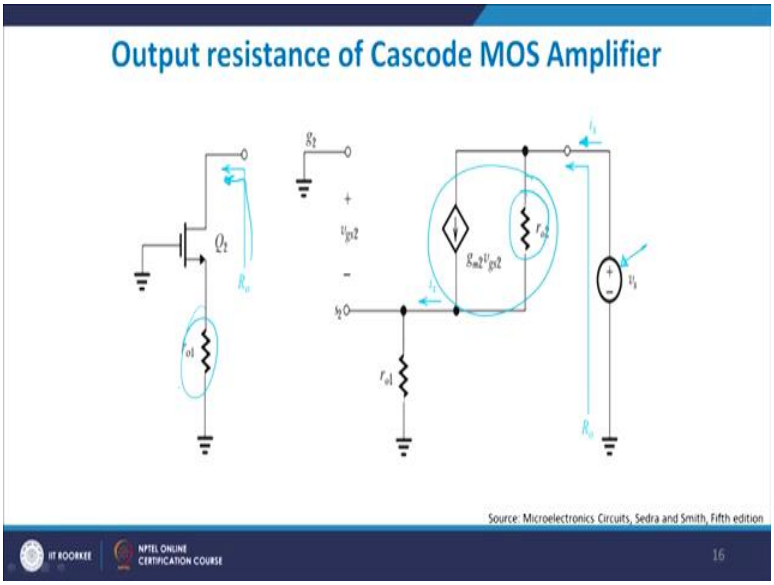
(Refer Slide Time: 27:08)



So, if you solve it and do small amount of manipulation which I did here in this in this slide you can see that the total current is basically given by $g_{m2}$ plus 1 by $r_{02}$ into $V_{gs2}$, right? And that gives you the value of output current and therefore, your $i_0$ is equals to gm1 into vi and your capital gm is equals to $i_0$ by $V_i$, right? And this is what you get finally at the end of the day that the capital gm which is the net trans conductance of both the devices in cascoded is given by the total current flowing through the device divided by the input voltage given to the common source stage MOS amplifier.

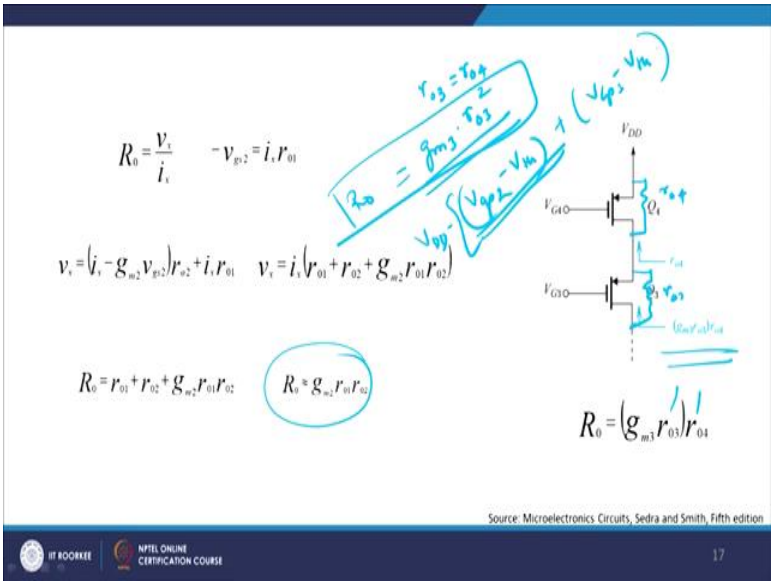
Now, if you want to find out the output impedance or the output resistance of a cascoded MOS amplifier the methodology adopted is that we generally short the input and we apply a unit voltage at the output and then see how much current is flowing through them. So, the voltage by current in the output side will give you the output impedance.

(Refer Slide Time: 28:00)




And therefore, you see this is basically your $g_{m2}$ $v_{gs2}$ which you see with $r_{02}$ as the output impedance offered by the device and you have a current source here $i_x$ here and you have applied a $V_x$ voltage here, right? And therefore, looking from this side you have $r_0$ which is defined. Now, as you can see you do not have q1 coming into picture. Q1 is actually been going because you have already moved q1 and replaced it by $r_{01}$ in this case.

(Refer Slide Time: 28:31)

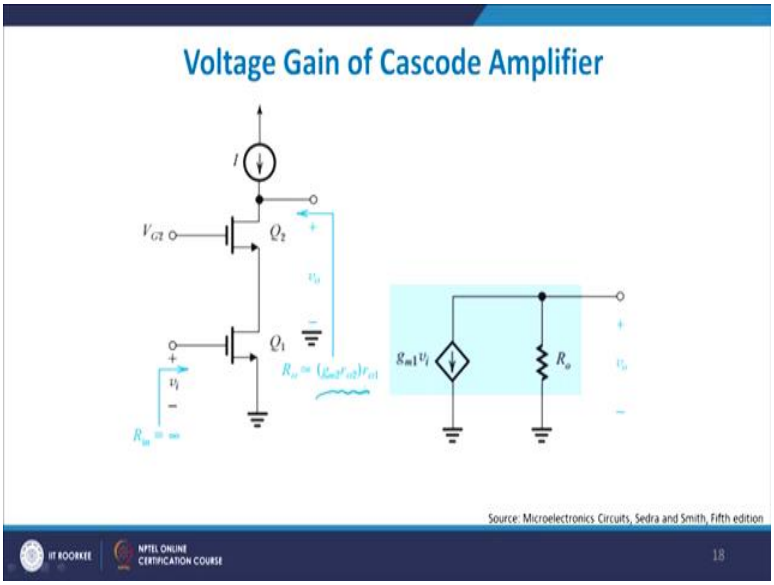



So, if you solve it the effective impedance comes out to be equal to $g_{m2}$ into $r_{01}$ into $r_{02}$ and therefore, that is quite interesting that that the overall impedance seen from the output side is basically the product of $r_{02}$ and $r_{03}$ which is basically the resistances at these 2 particular points. So, I have got $r_{04}$ and I have got here $r_{03}$. And therefore, they are sum they are product. So, basically multiplication of the 2. So, if $r_0$ if I assume that $r_{03}$ is equal to $r_{04}$ then I can safely write down as $g_{m3}$ $r_{03}$ whole square, right?

And that happens to be your capital $r_0$. So, you see if I do a single cascode I get this much if I do a double cascode my output impedance will be further high and therefore I will be able to achieve a larger gain. But the cost I pay for it is very simple that more I cascode the device more head rooms are meeting in the output because for every device to be in saturation, this much amount of overdrive is lost, right?

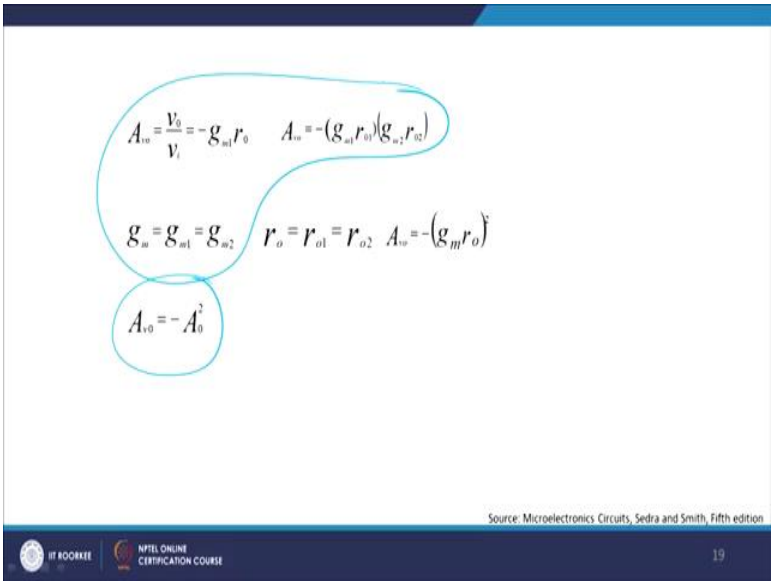
And as a result from $V_{DD}$ you have to subtract this much amount of overdrives for the cascoded device. This is $V_{gs2}$ this is $V_{gs3}$ even if it is threshold you need to subtract it. So, you see if these are relatively large then this quantity will be relatively small and head rooms you will see and your signal integrity will be a big issue for a cascoded device.

(Refer Slide Time: 29:55)



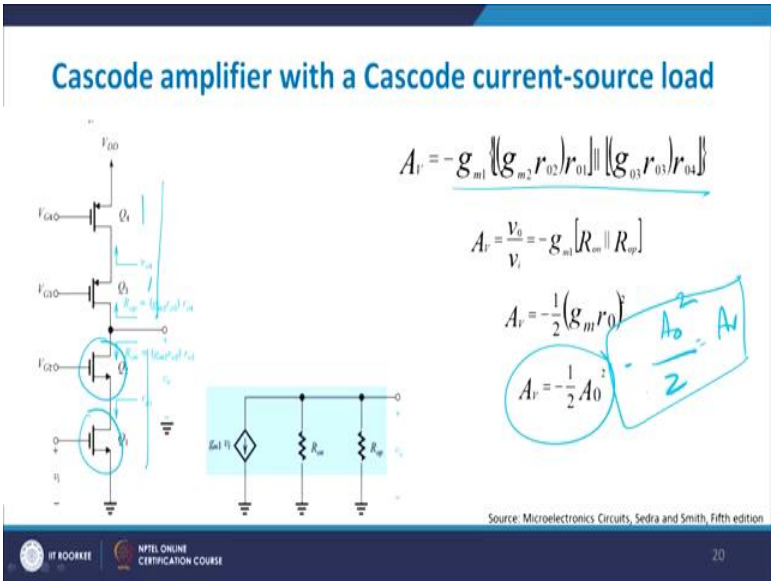

If you look at the voltage gain is very simple you just have to the input impedance will obviously be infinity because you are looking from the gate side. From the gate side obviously your resistances will be infinitely large but from the drain side if you look gm2 by r02 into r01 as I discussed with you.

(Refer Slide Time: 30:14)



And therefore, I can safely write down the voltage gain to be equals to $A_{v0}$ equals to minus $a_0$ square. So, what we have seen is that our overall gain, right; will be nothing but the square of each stage gain which means that you have 2 transistors $i_b$ $c_s$ and $c_g$ then multiplication of the each stage gain will give you the value of $A_{v0}$ with a phase change of 180 degree and therefore, you have a minus sign here. I am not doing derivation well documented in all the books and you can find it in all the text. But primarily this is what the important result and therefore, as I cascode device here, I end up having a higher gain not only that I also end up having larger $r_0$ which we have already discussed now.

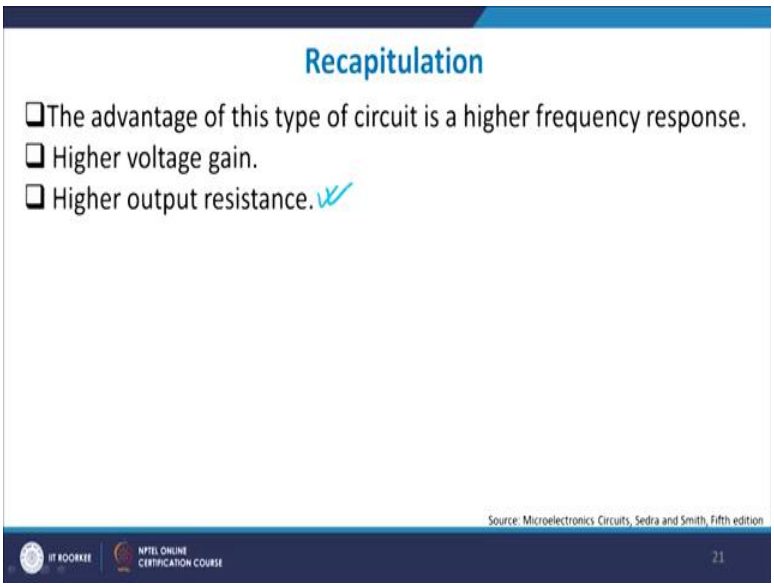
(Refer Slide Time: 30:54)




If you look at the cascoded current source which is being basically a current source which you see. So, this this is my input device q1 this is my cascoded device and these q3 and q4 are basically my load, right? And these loads will give you a value of voltage gain which is equals to $a_0$ square by 2. This will be with a minus sign and you will get $A_v$, right?

The derivation again is very simple and straightforward because looking from this side this will be in parallel to this. This is what I have done here, right? So, these 2 are in parallel. Do some small manipulations and I get overall gain to be equals to a square. So, so when you have a cascoded current source load this is a cascoded current source load q3 and q4 then your overall gain falls to half square root of 2. So, that is what you get from here 1 by 2 what you get.

(Refer Slide Time: 32:00)




So, let me recapitulate what we did in this in this case? We saw that if I do a cascoded amplifier in place of a differential amplifier, the gain will be relatively high at the cost of voltage head rooms; my gain will be also very high, right? That is what we have seen here. So, that is what you see here that they have got higher output impedances and they have got a higher voltage gain. So, cascoded transistor means CS and CG in series with respect to each other we define them.

Cascade will be when the output of the first transistor is to the gate of the second transistor. Then we define it to be as a cascade. Cascode means when CS and CG connected to source and drain valuable to the cascade, right? So, higher the stacking or higher the number of cascode structures more will be the voltage gain and more will be the output impedance, fine? So, this is 1 of the techniques by which we can simply improve the voltage gain of a single stage amplifier. So, this is not differential single stage amplifier and you can improve it drastically, right? So, with this we finish off this module and thank you for your patient hearing. Okay!

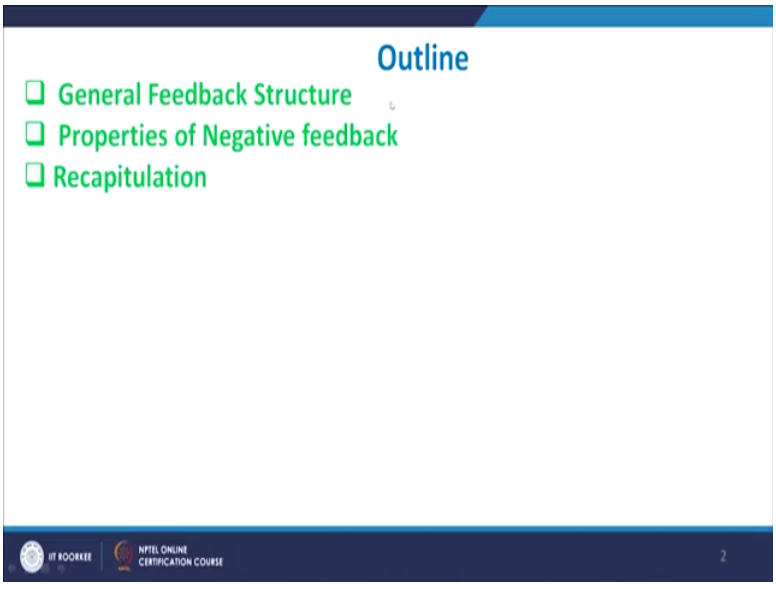
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−45**
**General Feedback Structure and Properties of Negative Feedback**

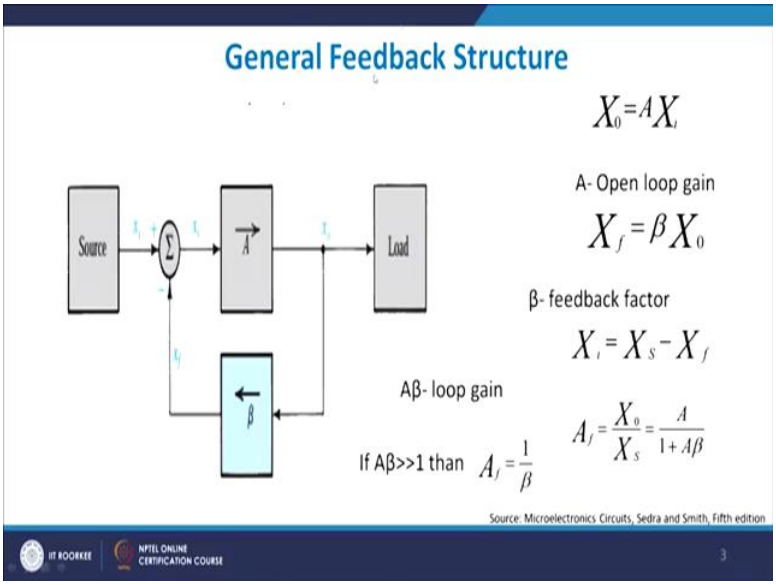
Hello everybody and welcome to the NPTEL certification online course on Microelectronics Devices Circuits. In our previous modules we have discussed about amplifiers, what are the various types of topologies of amplifier? And we have also discussed single stage and differential amplifiers. We have seen the effect of various components, passive elements to the concept of gain and also cutoff of frequency rule of.

We have also looked into the Bode plot and then compared the Bode plot of each and every configuration of amplifier. What we will do now, we will just change our tracks a bit and start off, if not entirely a different but slightly different module and that module is basically negative feedback. So we will start with the concept of feedback. What is the meaning of negative feedback?

And the general properties of negative feedback. After we have learned that we will learn into various topologies of negative feedback and then take up a valid example. Since we have already dealt with amplifiers we will take up each one of the amplifiers and explain to you how negative feedback helps me to achieve certain important output parameters of the device. So with this in mind this module was developed or discussed.

(Refer Slide Time: 2:10)

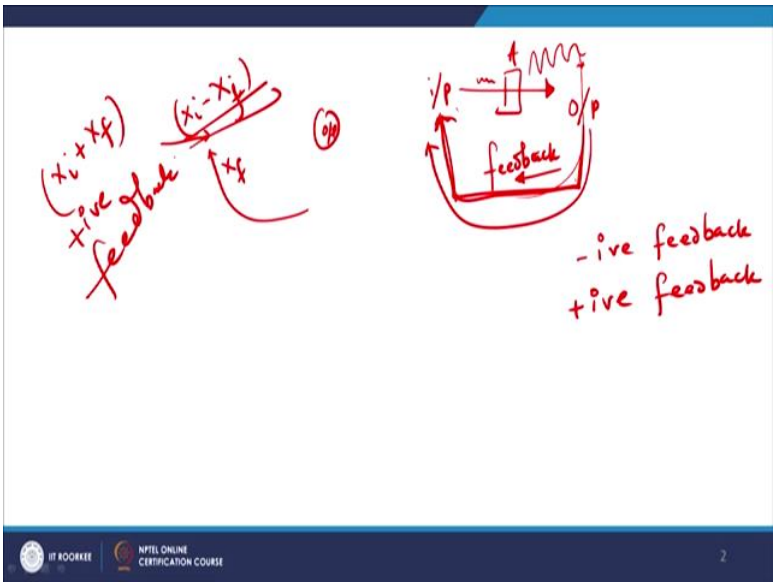





So the outline of at least this module there will be 4 or 5 modules related to negative feedback. For this module we will be restricting ourselves to what the general structure of the feedback, right?

So we will be looking into our general feedback concept which is basically the feedback structure which you will be seeing, right? And then we will look into the properties of negative feedbacks and we will be restricting at this stage only to negative feedback. They will not do anybody feedback. And of course we recapitulate what we have done till now, right?

(Refer Slide Time: 2:54)




Now let me explain to you why feedback was an important essential part as such in the structure for any amplifier or for that matter any structure. See feedback primarily means that

you take a portion of the output, so if I have got some output here I take a portion of that and then feed it back into the input once again, right? And as a result have feedback. I will give you basic examples.

For example say you are going through these NPTEL lecture courses of various professors across India. And you have already seen or understood one lecture. But after about 5 to 10 days you seem to have forgotten it and therefore you want to re- consolidated it. So what you do? You can go back to the same set of modules of the lecture, go through the lecture once again and then sort of give a positive feedback to your mind which means that you are again adding to your memory. Once again the same whole thing. As a result now your memory becomes much more stronger and it is able to retain the data to a larger extent. So this is a very basic sort of a layman approach feedback.

So feedback primarily means that you in a sense are able to get or add on a particular data from the output once and for all. So this path is basically known as a feedback path this mechanism is known as a feedback and this path is known as a feedback path which was from output to the input, right? And of course there will be always an input to output to transition because you need an amplifier.

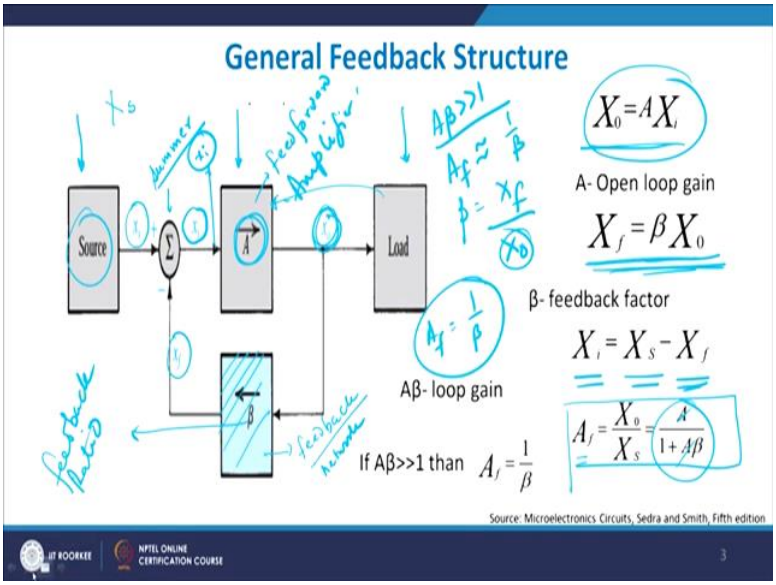
For example a small input you give you get a very large output here and from the output you take part of the output and you feedback into the input side, right? And we will see mathematically why it is important or why is it interesting to see, why it will be important at a later duration of time? But the general scheme is something like this.

Similarly, so let us, so there are 2 types of feedback which is basically one is known as a negative feedback, right? Another one is known as a positive feedback, right? As the name suggests negative feedback primarily means that you take voltage or current from the output and then feedback into the input such that you abstract it from the main input source. So I have an input current source here I will discuss with you just now.

So I have an input current source suppose $X_i$ and I feedback certain from a path from the output, right? And such that suppose this is $X_f$ the output will be $X_i$ minus $X_f$ then it is referred to as a negative feedback. So if this is 180 degree phase shift between your input and your feedback signal and then if you add or apply superposition principle you will automatically get a lower value of subsequent signals.

Whereas, if they are 0 degree or 360 degree out of phase and you add those 2 signals together what you will get is basically $X_i$ plus $X_f$, so this is known as a positive feedback, right? So a negative feedback will lower the voltage or current and the positive feedback will increase or improve the gain, right? So with this basic knowledge of feedback or the concept of feedback let me explain to you the general structure of a feedback loop.

(Refer Slide Time: 6:07)




If you look very carefully it comprises of, of course source is there, right? Which will be responsible for giving a voltage or current and there will be of course a load. So for any electronic system you will always have a source and a load. The load might be a resistive load, a capacitive load but primary a load will be there across which you will be able to find out the voltage or current.

Then you will have the amplifier in the middle, so this A is referred to as a feedforward amplifier, right? And its job is to basically amplify signal which is available at this particular point, right? So let me start and go through across the loop, across this. So source gives me, see input signal $X_s$, right? That is X of s means the source signal, you will have a summer here which is basically summing 2 signals.

And then the output is given as $X_i$ which is fed into A. So output of A will be, feedforward amplifier will be A times $X_i$. So $X_0$ if you see is A times $X_i$. $X_i$ is the input signal, so $X_i$ is nothing but the signal at this particular point or the input of an amplifier that if you multiply with A you automatically get $X_0$ as A into $X_i$, right? Now what you do is, your $X_0$ is here. Now you also have a loop here.

So this blue colored box which you see is basically my feedback gain or a feedback module. What it does? It takes a portion of this $X_0$ and it feedback's $X_f$ is nothing but the portion of this $X_0$. So my $X_f$ which is also referred to as the feedback voltage or current is given as $\beta$ times $X_0$, so $\beta$ is basically what? Is given as $X_f$ by $X_0$, what does it mean? It means that $\beta$ is a fraction, say $\beta$ is 0.5, 0.5 primarily means now that my feedback or a feedback voltage will be approximately 0.5 times that of the output voltage, right?

So half of the output voltage is being fed back. So this $\beta$ element is primarily referred to as a feedback network. So you have got a feedforward amplifier, I have got a feedback network here and we refer to $X_f$ as $\beta$ times $X_0$. This $X_f$ gets added up with $X_i$, so now since $X_f$ we are doing 180 degree phase shift here and adding it to the source signal. I get $X_i$ to be equals to X of s minus X of f, right?

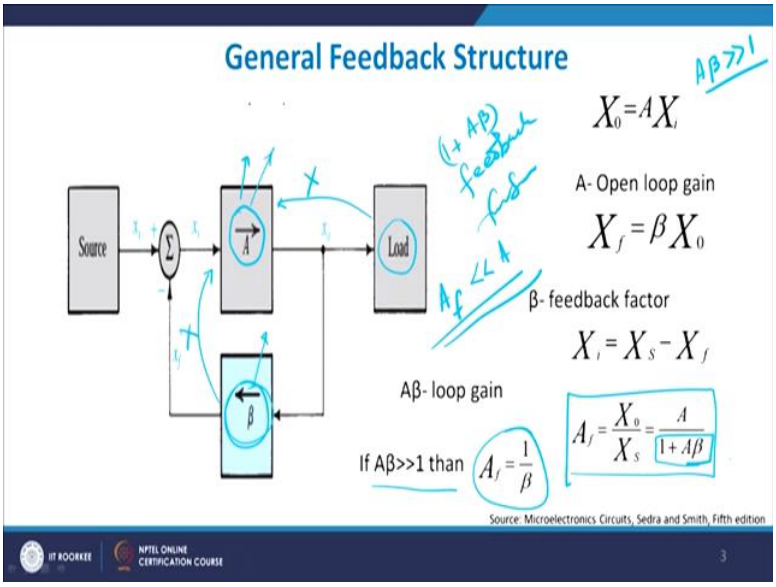
So X of f is $\beta$ times $X_0$ and X of s is already available with me. So I can write down $A_f$ to be equals to, $A_f$ means the gain with feedback to be equals to $X_0$ by $X_f$. $X_0$ is what? This by $X_s$ which is basically this. So gain is basically output by input. Output here is $X_0$, input here is $X_s$ and therefore $X_0$ by $X_s$ happens to be A upon 1 plus A $\beta$. And therefore, if A $\beta$ is much larger than 1, right?

Then I get $A_f$ to be approximately equals to 1 by $\beta$, so why? Because A $\beta$ plus 1 will be very large as compared to 1. AA gets cancelled out from here and I am left with 1 by $\beta$. So the feedback factor or the gain with feedback does not depend upon the gain of the feed forward amplifier it does not depend upon this gain here or the feed forward amplifier it only depends upon this feedback factor, so $\beta$ is referred to as a feedback ratio or a feedback factor, right?

And as a result we always get that $A_f$ will not depend upon the value of A, it only depends upon $\beta$ and $\beta$ is basically passive element which can be very accurately modelled and therefore my $A_f$ which is the voltage gain or current gain with feedback is very stable and it does not depend upon any factor which is temperature dependent or something like this. It is almost fixed depending on the user value.

While designing these we have tacitly assumed to important points that this load actually does not load our amplifier. Quite interesting I will expand to you what does that mean?

(Refer Slide Time: 10:55)

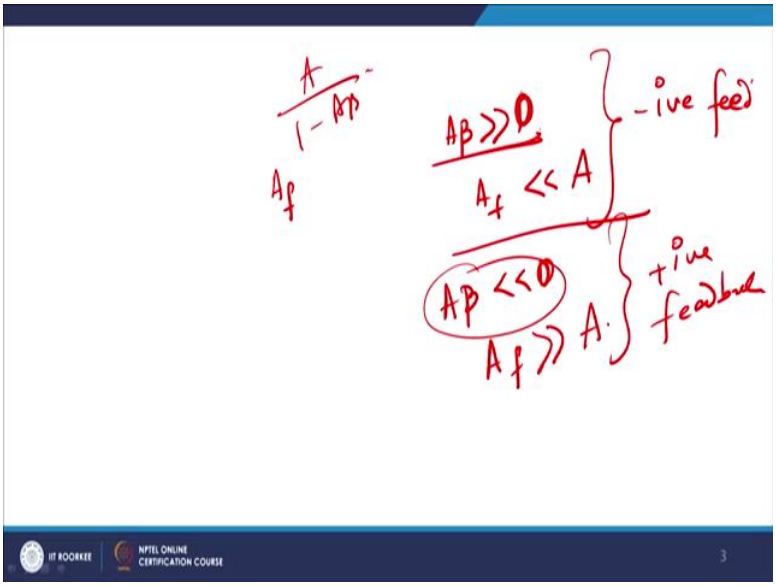



It means that, you have this load, right? We have this load, so this might be resistive load. Now I am assuming that resistive loading or for that matter this β does not load my amplifier in this manner or this does not load in this manner. There is no loading effect, right? So there is no loading, so I am assuming that. Then, second assumption is that my signal through my feedforward is always moving from input to output, so there is no reflection of signal within the amplifier itself.

Similarly, there is no reflection of signal within the feedback network also. So it is always going from input to output in the direction of the arrow here. So it is going in the direction of arrow here, this is going in the direction of the arrow here and that is what I have assumed here for all likely purposes. 1 plus A β is referred to as a feedback factor. Now you see that quite an interesting part here that $A_f$ is given as A upon 1 plus A β and under this condition I get $A_f$ to be equals to 1 by β, right? And β is basically as I discussed with you very small quantity.

Now, since A β let us assume it to be greater than 1, it primarily means that $A_f$ will always be less than A, right? Which means that whenever you have A β greater than 1 you will have automatically this to be denominator will be large and therefore you will always have a negative feedback and as a result you will automatically get $A_f$ less than A.

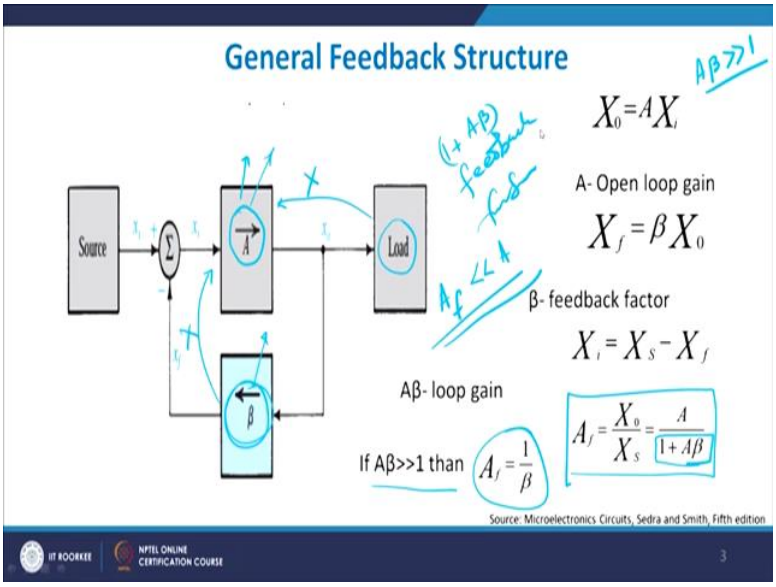
(Refer Slide Time: 12:38)




Whereas if A β, so let me write down for you explicitly that if A β is greater than 1 which is in most cases it is then I will get $A_f$ to be equals to much smaller as compared to A, right? And therefore we refer to this as negative feedback. Whereas if A β is less than 1 then therefore A β is negative and therefore A upon 1 minus A β, $A_f$ will be always larger than A. As a result this will be actually your positive feedback.

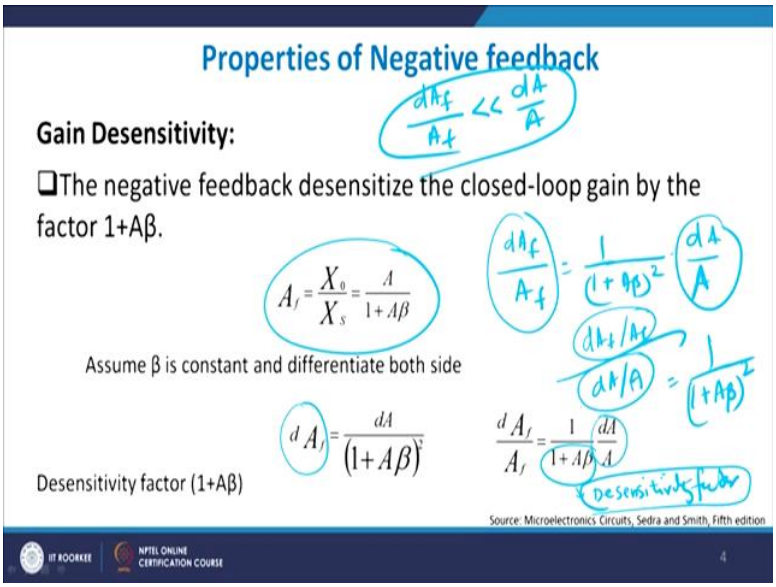
So whenever you want a positive feedback you have to ensure A β is to be less than 0, so it will be negative quality. Whereas if it is greater than 0 A β that we refer to this as a negative feedback. So any A β product of A β if it is positive, you have negative feedback. If product of A into β is negative you will always have a positive feedback, right? And so this is how it works out in a real sense.

(Refer Slide Time: 13:50)




So, now what you see here is that your $X_s$ is almost equals to, so what I see is $X_i$ is basically referred to as the error signal. Now, what will typically happen is that $X_s$ and $X_f$ will be very close to each other, right? This is the source signal and a feedback signal they will be very close to each other and therefore your error signal will be relatively small. And the error signal will go on reducing with more and more feedback loop available and $X_s$ will be very close to $X_f$, right? And therefore we say that the input signal will track each other. So $X_s$ and $X_f$ they track each other. So there is a tracking phenomenon which takes place between the 2, fine and there is a tracking which takes place between these.

(Refer Slide Time: 14:43)



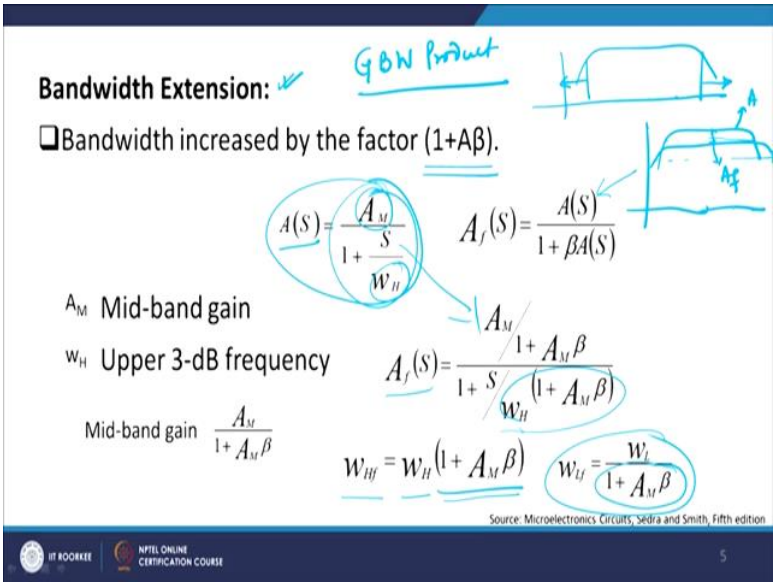
With this general property let me come to the properties of negative feedback. The first property of negative feedback is, the gain is desensitized which means that though our gain reduces in case of negative feedback, but the gain itself is desensitized which means that the gain is now more stable independent of frequency, independent of any of the systems available to us, any of the device parameters or circuit parameters independent of that.

Now, how can you do that? How can you show it? You can just simply $A_f$ is equal to $X_0$ by $X_s$ which is A by A β. And if β is constant and then if you differentiate both side with respect to $A_f$, so I get $dA_f$ equals to dA upon 1 plus A β whole square. If you divide this by $A_f$ then this becomes divided by A and I get $dA_f$ therefore by $A_f$ equals to 1 by 1 plus A β whole square into dA by A.

So you see dA by A is nothing but change of the gain of the amplifier, right? With respect to the amplification itself, whereas $dA_f$ by $A_f$ is the change when you have got feedback into consideration. Now therefore, if you find out $dA_f$ by $A_f$ divided by dA by A if you do it. It comes out to be 1 upon 1 plus A β whole square. Now, since 1 plus A β whole square is a very large quantity I get obviously this to be smaller as compared to this.

Which means that, it effectively means that $dA_f$ by $A_f$ will be much smaller as compared to dA by A which means again that you are able to control the gain with feedback in a much stricter sense as compared to without feedback, right? So I refer 1 plus A β, this quantity, in this case to be as desensitivity factor, so this is basically your desensitivity factor and this comes out from 1 upon 1 plus A β and A is referred to as dA by A, right?

(Refer Slide Time: 17:03)



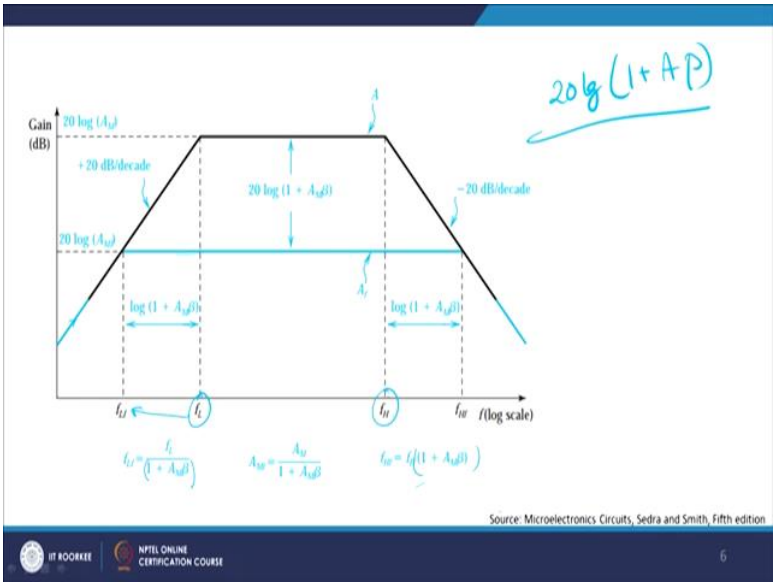
So, just let me explain to you therefore the next issue which is prevalent and that is known as bandwidth extension, right? What happens is that, what we have seen is that the bandwidth increases with A β. So your gain decreases by 1 plus A β but your bandwidth increases by 1 plus A β that I am just doing it to you here. So I already know that gain is basically given by mid-frequency gain upon 1 plus S upon omega H.

If you have with feedback, so this is given by AS upon 1 plus β AS, as we have discussed, right? And therefore $A_f S$ will be nothing but I will just place this whole thing here, right? And I get 1 plus S upon this whole thing here and therefore I get omega of Hf is equals to omega of H 1 plus A β times f and omega L of f is equals to omega L upon 1 $A_{ML}$ which means that my high frequency cut-off point increases by a factor of 1 plus A β and my low frequency decreases by factor of 1 plus A β.

Which means that if you are actually plotting again, so this is my low frequency, this is my high frequency cut-off, this shifts to right by 1 plus A β and this shifts to the left by 1 plus A β. So your effective bandwidth becomes, so what happens is that your, so, say this is your without feedback, right? And then with feedback the gain will drop down, right? But your bandwidth will become larger. So this is $A_f$ with feedback and this is A.

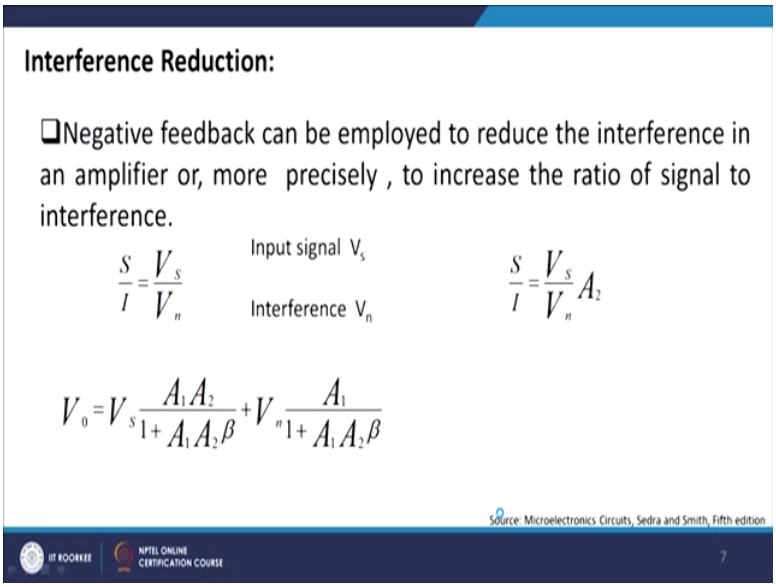
So gain has come down but your bandwidth has become larger as compared to your with of course with the (())(18:37), right? So this is one thing such that and therefore your Gain Bandwidth Product is always constant. So the gain falls down but the bandwidth increases by certain factor such that the product of them is always constant with respect to each other.

(Refer Slide Time: 18:55)



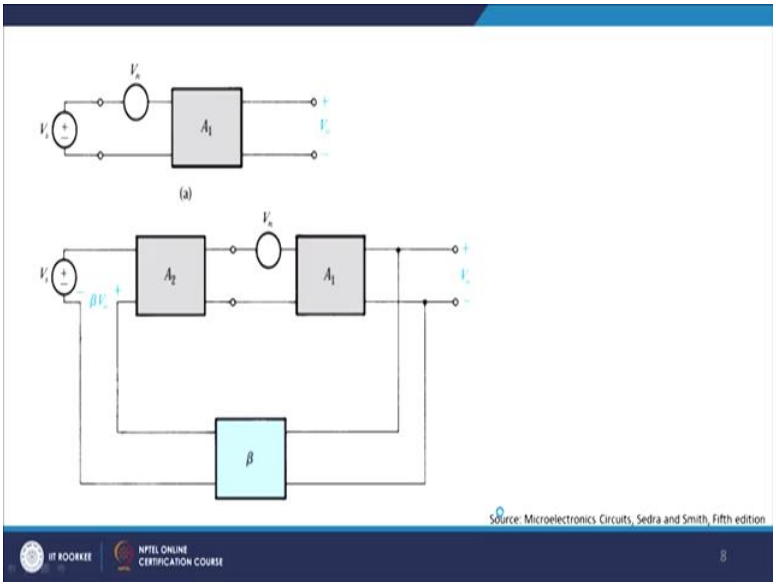
And that is what you see here, this is without feedback, right? And as you apply feedback this shifts to the left by virtue of the fact that you divide it by 1 plus $A_M$ times β whereas this case shifts to the right by a factor of 1 plus $A_M$ times β. So you see that with feedback the blue color with feedback is actually having a much lower drop. So your gain drops by about 20 log of 1 plus A times β, right? And you have what? 20 db gain here and 20db drop here. So that makes my life difficult in the sense that it makes the game fall down drastically.

(Refer Slide Time: 19:39)



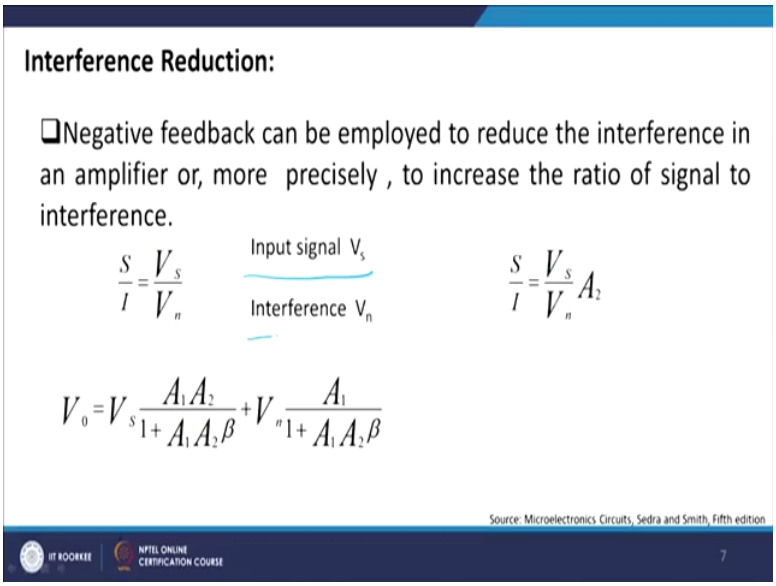

We come to the negative feedback concept and show to you that this also helps you to reduce noise reduction. So noise reduction is also there whenever you do have a negative feedback. How do you do that? Let me just explain to you how do you do that. It will not be visible here but let me show it to you.

(Refer Slide Time: 20:03)



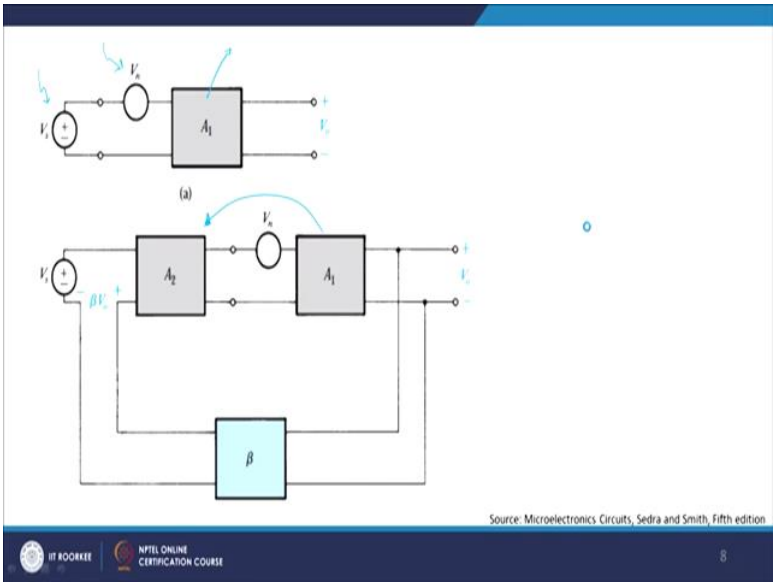

So I have got amplifier $A_1$ which is basically a noisy amplifier and I have a source signal here and at the input referred noise of this $A_1$ is basically $V_n$. Now what I claim is, that if I had this $A_1$, right? It should be preloaded by an amplifier $A_2$ whose noise source is almost 0, right? If this is the case then I automatically get a much better profile as far as designing is concerned.

(Refer Slide Time: 20:39)



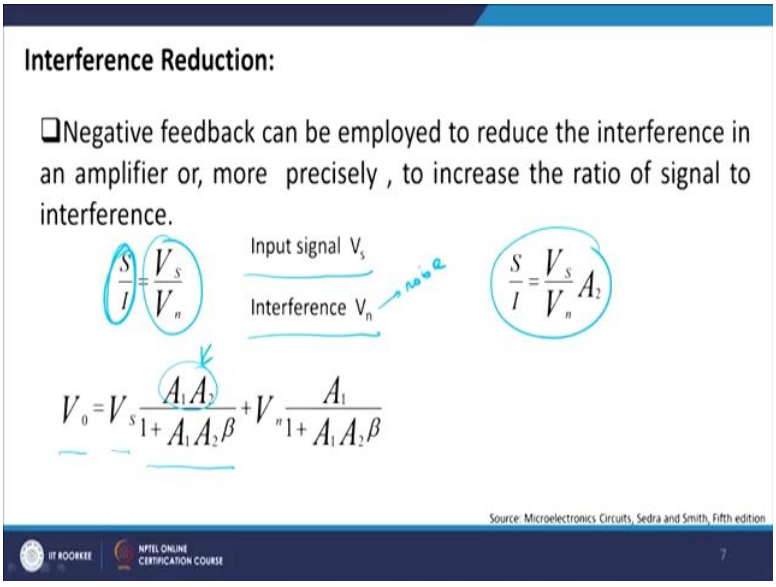

So, I will just show you what I am trying to tell you here that I have an input signal which is Vs, I have an interference which is $V_n$, so this is my noise voltage, input referred noise voltage. So I have Vs by $V_N$ which I get and therefore I signal by interference S by I is Vs by $V_n$ into $A_2$ because $A_2$ is nothing but this signal where am getting it.

(Refer Slide Time: 20:54)




So I am a negative feedback concept here. I have a β here, I have got 2 transistors $A_1$ and $A_2$ and therefore you see V0 will be nothing but Vs times $A_1$ $A_2$ upon 1 plus $A_1$ $A_2$ into β, why? Because now you see your overall gain in the feed forward part will be $A_1$ into $A_2$. So wherever you had A you will be replacing it by $A_1$ and $A_2$, right? So that is what you are doing here.

(Refer Slide Time: 21:24)

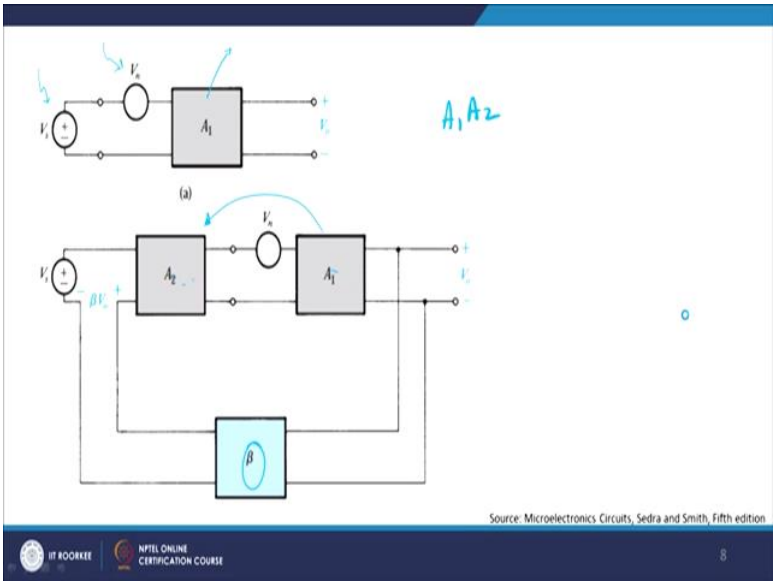



You are replacing A by $A_1$ $A_2$, here also replacing A by $A_1$ $A_2$. So what is happening is, your source signal is getting multiplied by this thing where is your noisy signal is only getting multiplied by $A_1$ upon 1 plus $A_1$ $A_2$, why? Because noisy signal is coming between $A_1$ and $A_2$. So only $A_1$ is responsible for increasing the noisy signal and $A_2$ is responsible for

increasing the source signal and as a result your signal to noise ratios will be improved provided you are able to sustain this into consideration.

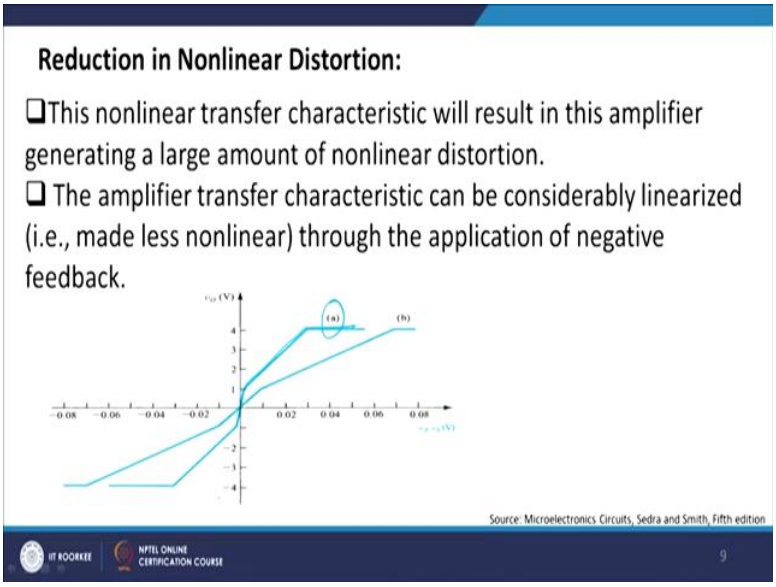
Now, if you look carefully here, so this if you take Vs into $A_1$ $A_2$ upon, if you take this plus $V_n$ into $A_1$, if you take this as common factor and then 1 upon 1 plus $A_1$ $A_2$ by β, right? So you see if your noise voltage increases this factor will become larger, right? But if you are able to make your $A_1$ small than the noise voltage effect on the overall output voltage will be bear minimum provided you are able to reduce the value of $A_1$, right?

(Refer Slide Time: 22:37)




So what you do is that you make this amplifier a very low gain amplifier almost near to unity if possible and make this which is noise free amplifier a large amplification. Once you are able to do that you are able to sustain a much larger voltage in the output side. So also this helps you to do a non-linear reduction in non-linear distortion properties what happens is that, for example if you look at A it has got (()) (23:04) non-linear at this particular regions, right?

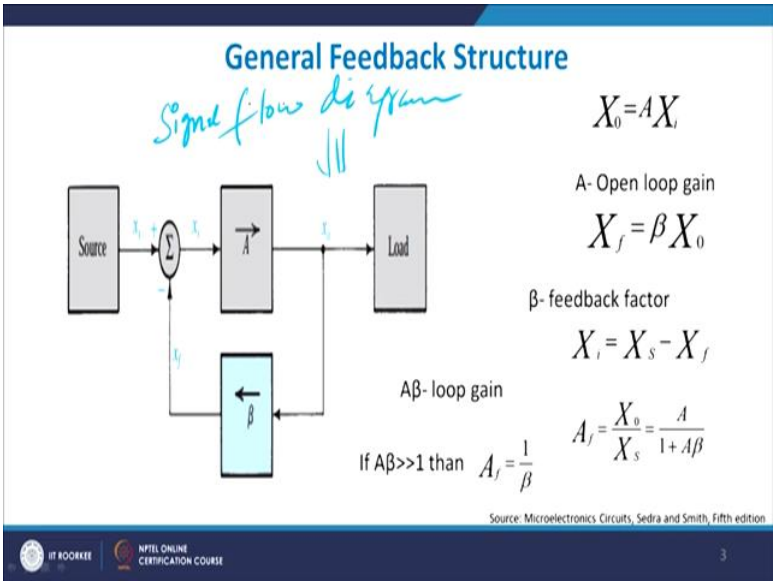
(Refer Slide Time: 23:09)




So when you do a negative feedback your non-linear distortions get reduced drastically, right? This is feedback and this is without feedback. With feedback it is almost linear in dimensions and nature whereas here you will have peace wise linear approximation because of, so when you apply a negative feedback distortion also reduces, right? The non-linear distortion also reduces drastically.

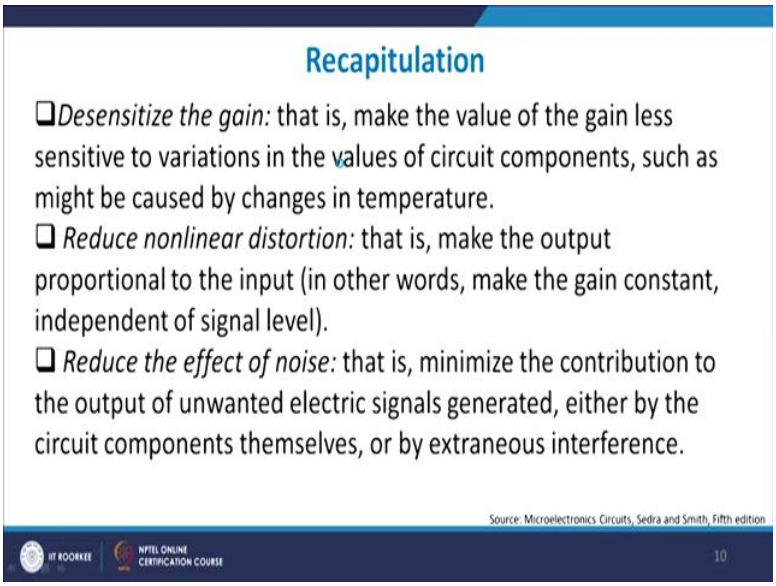
So this is one of the advantages of negative feedback as far as this is concerned. So with this let me wrap up this module and explain to you the other factors that while we have understood what is negative feedback, we have understood the topology of a basic signal flow, just to give you an idea this is also referred to as a signal flow diagram.

(Refer Slide Time: 24:02)



This is referred to as a signal flow diagram. So from signal flow diagram we were able to look at the general feedback structure.

(Refer Slide Time: 24:16)



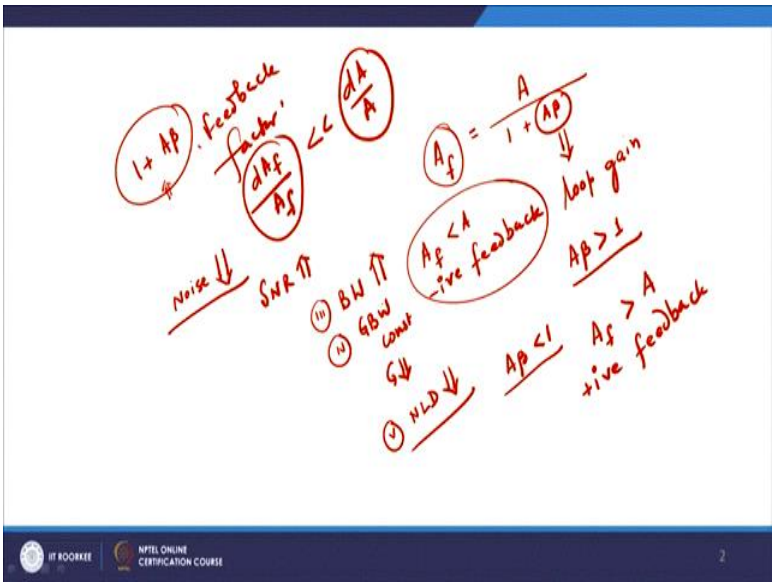

We were also able to see what are the advantages for example my again reduces but my again becomes less sensitive. So desensitization of gain is one of the important point and it becomes less sensitive to variations in temperature and so on and so forth. There is also a reduction in the non-linear distortion, but it also makes my bandwidth increase at the cost of a lower gains such that game into bandwidth is always constant.

And also if we reduce the effect of noise, but that is not so straightforward in that you have to reduce the gain of the noisy amplifier and improve the amplification of the preamplifier. So $A_2$ is referred to as a preamplifier, please understand this point also, that $A_2$ is basically referred to as a preamplifier. So preamplifier should be actually less noisy, but its gain should be very high and $A_1$ should have a much smaller gain as compared to 1 and your noise signals will be suppressed if you do a this thing.

So we have been able to see the advantages of negative feedback when we come next time, when we join next time we will be able to handle the other products of the system also and see the advantage in real life in real amplifier design, right? Thank you for your patients hearing.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−46**
**Basic Feedback Topologies**

(Refer Slide Time: 1:16)




Hello everybody and welcome to an edition of the NPTEL online course on microelectronics devices and circuits, what we will do today is look into the basic feedback topology, if you remember in our previous module we had discussed about the concept of negative feedback, what is the meaning of negative feedback? We have seen that in order to have a negative feedback the general scheme is basically with the amplification, with feedback is equals to amplification without feedback A/(1+Aβ).

Where Aβ was referred to as the loop gain right and as long as Aβ > 1 right, you will automatically have $A_f$ < A and we refer to this as a negative feedback, that means, with feedback the gain is reduced, similarly, if Aβ < 1, then you get your $A_f$ > A because a denominator actually becomes less than 1 and this is refer to as a positive feedback right, so in most of the cases which we will be encountering, we will be concentrating on negative feedback, per say, right and we have also seen that with this loop gain, 1 + loop gain if you see and that is known as the feedback factor.

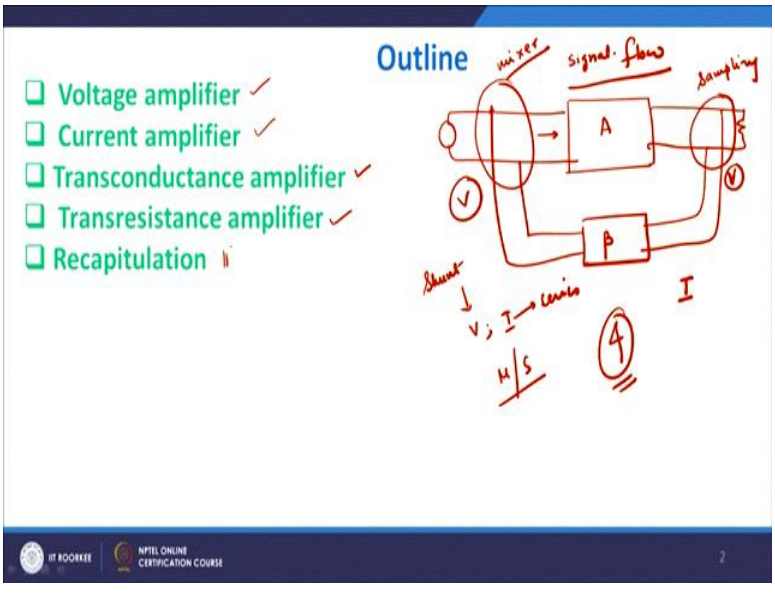
So this (1+Aβ) is refer to as a feedback factor right and if you look at the feedback factor higher the value lower is your $A_f$ and therefore in most practical cases we try to keep this one

as high as possible, but the advantage of all this things is that, that your gain desensitisation takes place, so $dA_f/A_f$ is reduced drastically as compared to dA/A, so this is gain without feedback, this is relative gain with respect to feedback right, so your gain becomes most stabilised, what you lose is basically the gain, so gains falls down, but your gain becomes much more stabilised with respect to variation in temperature and output electrical characteristics.

Similarly your noise also is reduced, noise is also reduced and your signal to noise ratios becomes larger, provided you are able to have a preamplifier which is less noisy in series to the noisy amplifier right, if you do not do that, then it is not possible. Otherwise, we generally have a lower SNR, the third thing which is, this is second part, the third part is basically your bandwidth increases right, so bandwidth increases because lower cut-off frequency goes smaller and the higher cut-off frequency goes higher and therefore the bandwidth increases.

But your gain bandwidth product, which is basically GBW is constant and therefore what happens to the gain is that the gain starts to fall down, so your bandwidth increases, gain falls down more stability, less of noise and there will be non-linear distortion will be also reduced right and these are the very important points of negative feedback.

(Refer Slide Time: 3:39)



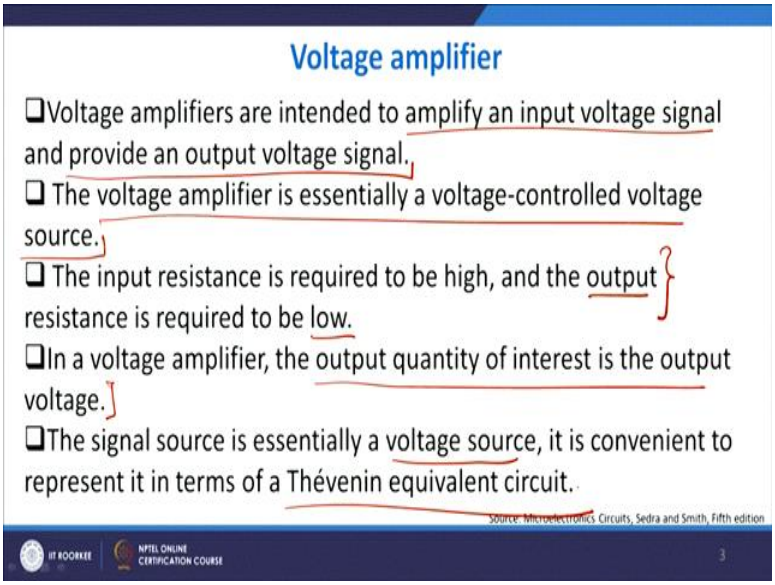

Today, what will be looking into is, we will be looking into the various topologies of feedback amplifier and if you remember, in our previous talk, in our previous discussion, the general scheme of things was the signal flow diagram if you look was you had A and then you had β here, so β was basically my feedback factor and then you have a source of voltage

source or a current source here, which was feeding the amplifier, then you have a load here right and this load was talking to my β here and a part of β was fed into this amplifier, and as a result, you will have a feedback factor.

So this was your basically your signal flow diagram, if you remember from my previous discussion, now, depending on how you manage this side and this side, this is known as sampling portion and this is known as a mixture portion, so depending how you sample and mix, you get a four types of configurations available to me, whenever you sample in a parallel form, you are actually sampling out voltages and whenever you are sampling in series you are, so this is basically voltage, this is also voltage right, because this is parallel, this is parallel, so parallel, parallel basically, means that your voltage, voltage feedback network is there.

Similarly, if you have series extraction, then you have current extraction taking place, so we have got two notations, current and voltage right, so current voltages shunt because it is parallel and this is basically your series right and then you also have, you have two voltage on current as a two-part and you have therefore two mixers, mixer and samplers, so in effective here are kind of 4 configurations available to me right and these 4 configurations are written here, one is a voltage amplifier, another is a current amplifier, you will have transconductance amplifier and trance resistance amplifier and then we will recapitulate the whole discussion once again right, so we start with voltage amplifier and then you go to current amplifier, transconductance amplifier and so on so forth, now if you go at, if you go to the voltage amplifier right.

(Refer Slide Time: 5:45)




Let us discuss the voltage amplifier first, so voltage amplifier's main job is to amplify a voltage, as the name suggests, so you will amplify an input voltage which is very, very small in dimensions, peak to peak and will provide output voltage signal, so at this stage we are not talking about, at this stage at least we are not talking about frequency change or for that matter, the change in the frequency of the inputs source, so on and so forth, we are keeping the input stable at, say, omega and we are trying to find out the variation in the peak to peak signal of the output with respect to A and β, A is the feed forward amplifier, configuration factor and β is the feedback network exponent.

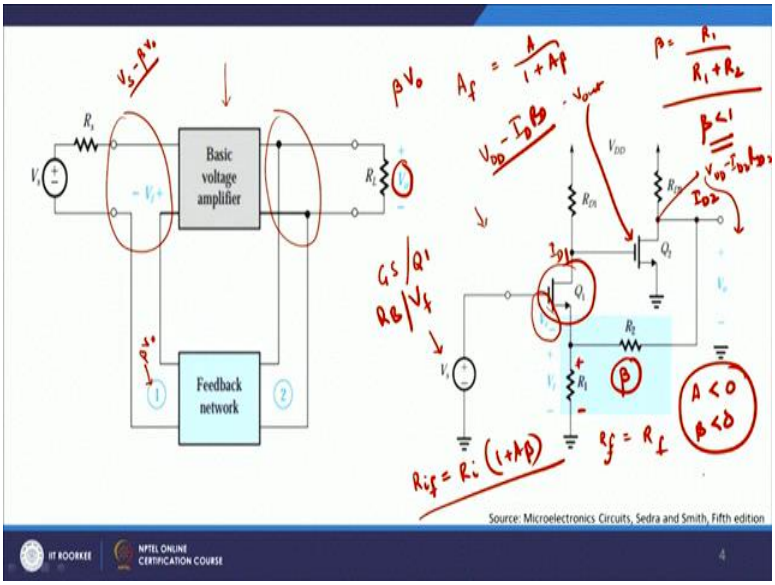
So therefore a voltage amplifier is actually a voltage-controlled voltage source, I suppose you can understand this why, because if you look very carefully you are actually extracting voltage from the sampler and you are feeding voltage into the mixers, so both are shunt, shunt configuration or both are basically parallel configuration, now why it is known as a VCVS or voltage control voltage source, the reason is that you are, this is basically voltage amplifier, so the output is basically a voltage which you are getting right.

So therefore output is the voltage source, so that is the reason it is said to be voltage source, why is it voltage control because the amplifier which you will be using typically here, will be controlled by a voltage right and therefore this basically a voltage controlled voltage source, in this case, as we will see the input resistance is quite high and the output resistance is quite low right, just like an op-amp , so op-amp is an example of a voltage amplifier, we already know that and therefore if you look at op-amp, its input resistance is infinitely large or very

large and output resistance is very, very small, same thing happens to a voltage amplifier here.

What we will be looking into is that the output quantity of the interest is the output voltage, so we will looking at the output voltage and since the signal source is essentially a voltage source, it is always advisable that you use Thevenin equivalent right, so what will do is that we will use and Thevenin equivalent when we do a voltage amplifier action or a voltage amplifier division.

(Refer Slide Time: 7:59)




So this is the basic architecture of your voltage amplifier here, basic amplifier so as you can see this is your sampler right and this is your mixer here right, you do have, this is the load and what you do is that this feedback network tries to take a portion of this $V_0$ and therefore you will have a output β times $V_0$ coming at this point, so this will be β times $V_0$, which will be coming here, then you will have a feedback network and therefore you will have ($V_S$ – β)$V_0$ will be coming at this basic network here.

And then what you get from here or what you get from this basic configuration is that in this case also you will have $A_f$ the feedback with will be equals to A/(1 + A β) right and that is what happens to a larger extent right, now if you go ahead and do it, for example I will give you an example, for example, the actual voltage gain series network which you will see is, if you see here, this is basically an example of a voltage amplifier, the circuit here.

This $R_1$ and $R_2$ is basically your β right, so β is basically the parallel combination of $R_1$ and R2 which you will see here and therefore if you are taking out the voltage out of this particular form, I get β to be equals to $R_1/(R_1 + R_2)$ right, and as a result β will be always less than 1, so when β =1, it primarily means that your all of the voltage in the output side is fed back in input side, if $\beta < 1$ that the percentage of the output voltage is fed back.

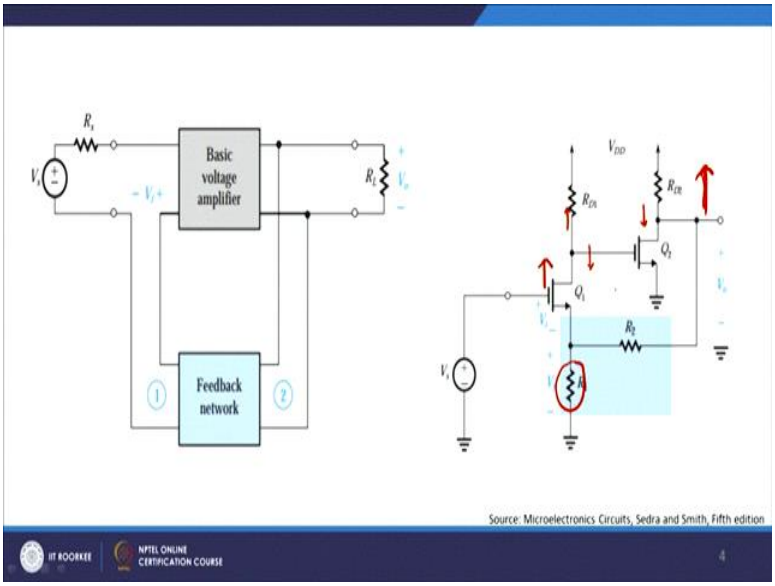
Now if you see very carefully the input voltage, this is the source voltage which you are giving right, the input voltage which appears between the gate and source of $Q_1$ is referred to as $V_I$, so this is my $V_I$ right, so this my $V_I$ right, input voltage and this gives you a current here $I_D$ right and $I_D$ current flows through this point, this generates a voltage, how much voltage will be generated, it will be $(V_{DD} - I_D R_D)$ right, that will be $V_{out}$, this $V_{out}$ will be fed where, will be fed into the gate of $Q_2$, this $Q_2$ gate will generate a current here which is $I_{D2}$ right.

Therefore this is $I_{D1}$, this is $I_{D2}$, then $I_{D2}$ will generate a voltage here which is given as $(V_{DD} - I_{D2} R_{D2})$ right, let us suppose, this is the voltage here, now that voltage is your $V_{out}$, so a fraction of that depending on the value of β is again fed back negative, why negative? Because if you see very carefully that this voltage is, this is positive and this is negative and this will be reversed biasing my gate to source of $Q_1$ right, so my gate to source of, gate to source of $Q_1$ will be reversed biased with respect to $V_F$, because $V_F$ this is positive, just like a source degeneration resistance remember, exactly the same happens here as well and therefore he tries to reverse bias this transistor $Q_1$ right and therefore you have negative feedback network here.

The obvious reason why generally we have two transistors is that one is a driver transistor another is a transistor, which is being driven by this $Q_1$ and therefore that is the voltage which you see from the outside world right and you are able to sustain this formulation here, depending on how you look into it, we will see automatically that the by application of a reverse bias here your input impedance actually rises with feedback.

So with feedback $R_f$ if I want to find out input, so if I want to find out $R_{if}$ which is input impedance with respect to feedback I get without feedback $[R_i (1+ A \beta)]$ as the increase in input impedance, in this case you will see that this is negative feedback primarily meaning therefore that $A < 0$, $\beta < 0$, so when you multiply both of them you get a positive sign and therefore you will get a negative feedback.

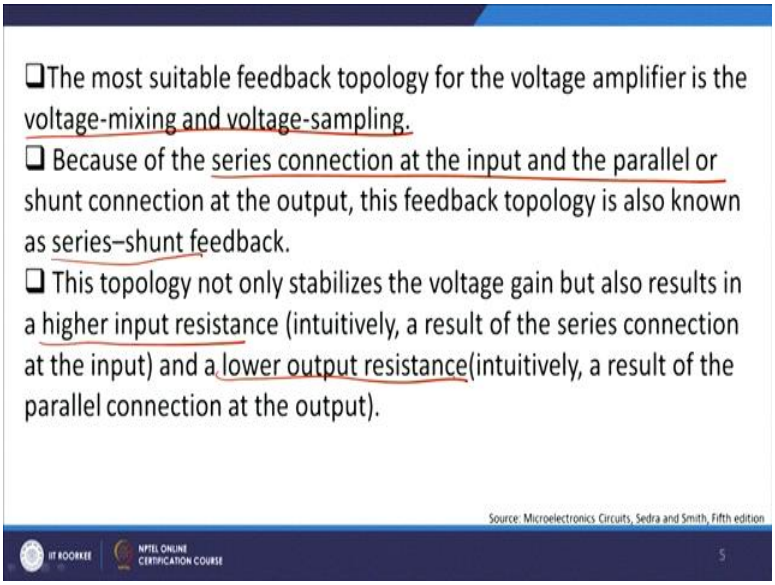
(Refer Slide Time: 12:04)

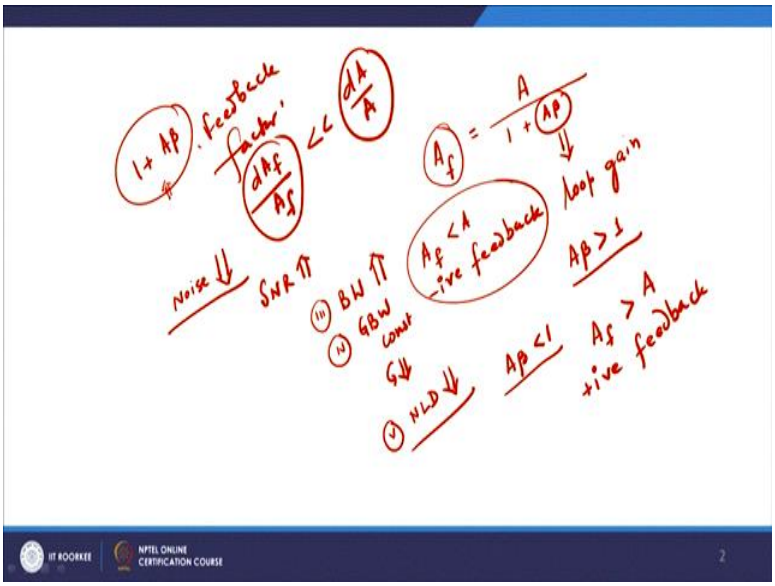



From this physically, let me explain to you how I get negative feedback, so as I discussed with you as this VS increases, this voltage increases, this current increases, so your this voltage drops down, which means that this current here will drop down, which means that this voltage will increase, so with increase in $V_s$ here I get an increase in voltage here right, but this increase in voltage is primarily means that my $V_f$ we also go on increasing.

And as a result, it will negative feedback my $Q_f$ in a much more larger term and that is the reason, that is the basic origin of negative feedback here and as you can see here, therefore, that depending on the effective values of the gain of $Q_1$ and $Q_2$, this factor will be typically very large, you can get a very large gain here at the output of $Q_2$ and that will be fed through $R_1$, $R_2$ network back as a negative feedback on to $Q_2$.

(Refer Slide Time: 12:55)



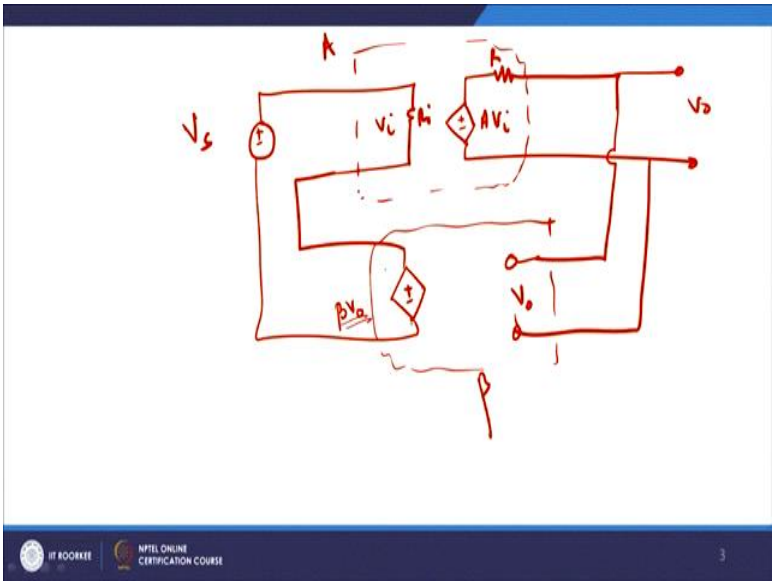




So this is your voltage feedback network, the most, as I discussed with you, the most convenient form is basically the voltage mixing and voltage sampling, no need to say we have discussed this point time and again, because of the serious connection at the input and the parallel connection at the output, this topology is also known as series shunt feedback, because we referred to this, mixer side as the series first and the sampler side as a shunt and therefore we referred this to as a series shunt mechanism or a series shunt mechanism which we see.

Now if you look at the series shunt feedback amplifier what we can and what we can write down is that as we discussed just now that we get a higher input impedance and a lower output impedance right and that we will just show it to you through a derivation maybe and

this derivation will remain the same for almost all the cases, the methodology by which we have adapt.

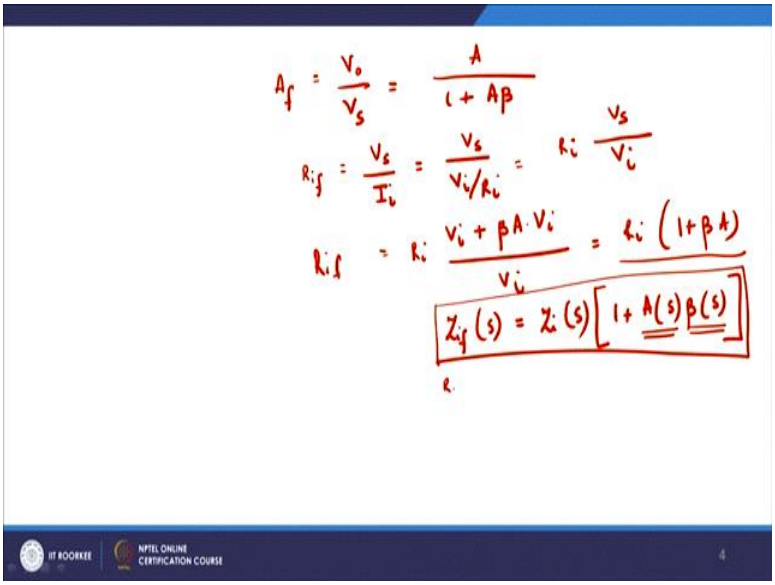
(Refer Slide Time: 13:52)




So if we, let me draw you the effective diagram of a series shunt feedback network and explain to you how it works out right and I will explain to you how this is being done, so it is $AV_i$, I will explain these one of them individually, then you have R here and then this is your $V_0$, this is output voltage and then you have, you take a parallel combination out of it, assuming $V_0$ is the output here, we take a factor of it given as $\beta V_0$ and I get this factor and then this is connected to this side and then we get something like this right.

So this is your amplifier A, for feed forward amplifier A, where this is your $V_i$ and this is your $R_i$ and this is your $R_o$ or R output resistance and this is your voltage, so this is your feedback network, this is your β, feedback network and this is your feed forward amplifier and this is a voltage source $V_S$, so if you see very carefully $AV_i$ is nothing but the output voltage which you see in the output side of the amplifier and that is the reason I refer to this as $AV_i$ and out of $V_0$, $\beta V_0$ is the output voltage of the feedback network, so this is the effective circuit diagram which is seen and can be explained to you.

(Refer Slide Time: 15:15)




And let me write down therefore that with feedback I get $V_0/V_S$ and that is equals to A/( 1+ A β) and that is a standard form of you see but let me write down, let me show a derivation of good impedance with feedback I get $V_S/I_i$ output, source by current, I get $V_S/(V_i /R_i)$ because $I_i$ is the input current, which is nothing but the voltage, input voltage by input current and this can be written as resistance times $V_S/V_i$ right.

So this can be written as $[R_i(V_i + \beta AV_i)/V_i]$, so I get this to be equals to $[R_i(1+ \beta A)]$ as the value of $R_{if}$, so your $R_{if}$ shows increase by a factor of (1+ βA), so in general, if you want to generalise at I can write down $Z_{if}$ which is impedance with respect to frequency, now I am writing now is equals to $Z_i(s)[1+A(s) \beta(s)]$, see what we are assuming till now was that, we are assuming that the feedback network is not loading your amplifier, neither your load which is there, RL is loading your amplifier, so just they are working on a standalone basis right, in reality this is not true, there is always a loading and then second thing is that it is also frequency dependent.

For example, the gain of the amplifier is again a frequency dependent quantity, you must have known by this time, that in the mid frequency only your gain is stabilised and gives you a constant value, at very low frequencies and high frequencies the gain starts to fall down, so therefore it has to be a frequency dependent quantity and therefore I right down to be as A(s), similarly β if it is a purely passive network it will be independent of frequency but if it is an active network it will be also having a frequency dependent term coming into picture and that

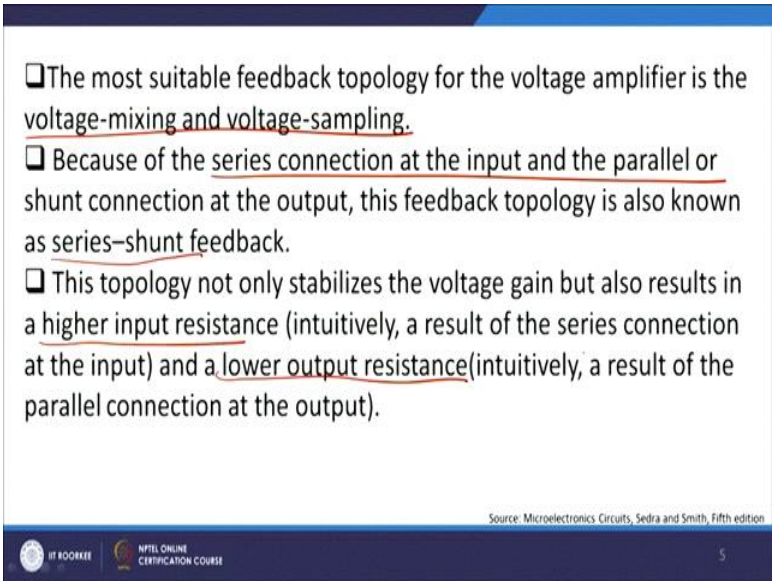
is the reason I write a generalised formula for any voltage, voltage amplifier in this form, in the form given here.

(Refer Slide Time: 17:21)

Let me show you for $R_{of}$ which is the output impedance here, let me do it in the next page, maybe, $R_{of}$ which is the output resistance, is given as $V_i$ / I, where I is given as $(V_i - AV_i)/ R_o$ because you see its - because you are doing a negative feedback, now what we do is that, for finding out the output impedance we shot the input and therefore $V_s = 0$, I get $V_i = -V_f = -\beta V_0$, which is nothing but $(-\beta V_i)$ because $V_i$ and $V_0$ are constant in this case.

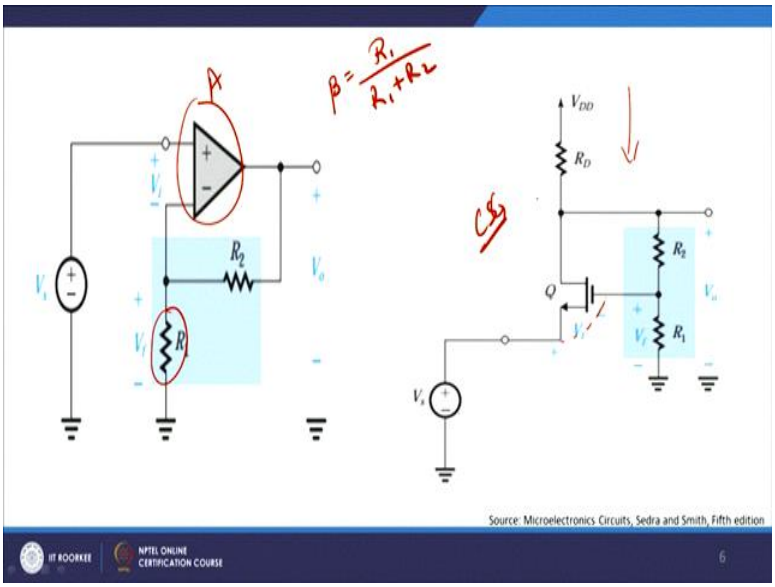
So I get current to be equals to $(V_i + A\beta V_i)/R_0$, so I get therefore $R_{of}$ right, $R_{of} = R_0 /(1+A\beta)$ because you see we take $V_i$ common right, you take $V_i$ common and so this will be $I = V_i(1 + A\beta)/R_o$ right, so if you take this to this side and just vice versa, I will get $R_o/(1 + A\beta)$ right is equals to $V_i$ / I, now $V_i/I$ is nothing but $R_{of}$, so $R_{of} = R_o /(1+A\beta)$ right and that is what you get and in general, we can write down therefore $Z_{of}$ with feedback output impedance can be given as $Z_o(s)/(1+ A(s)\ \beta(s))$ and this is what we get from the overall discussion, this value which is available with me, so we have seen that the input impedance rises and the output impedance falls down.

(Refer Slide Time: 18:58)




And that is the intuitive result of which we have already found out because it is a parallel, output is a parallel connections, output parallel connection primarily means that output impedance will fall down right and that is the reason you get a reduction in the output impedance.

(Refer Slide Time: 19:11)

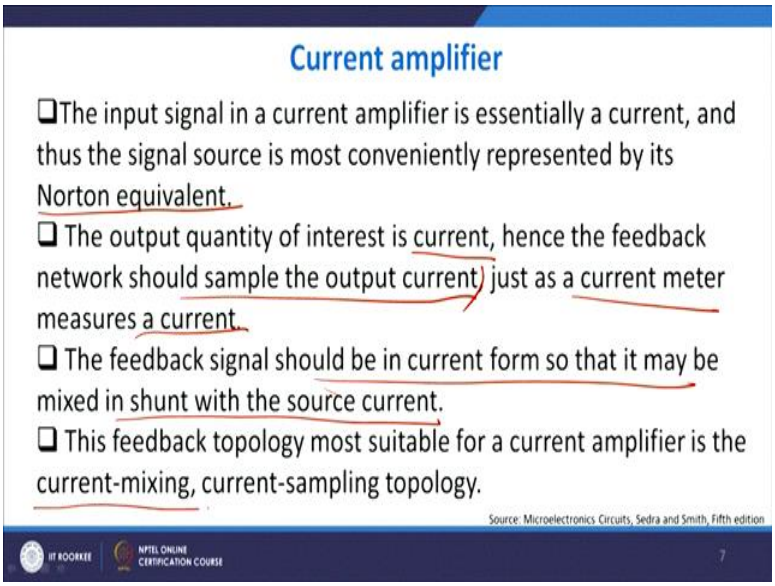



Let me show to you now an example, wherein you have another example here, which is basically, just let me, this is your current amplifier, another example is basically this one, wherein you have got your amplifier here, this is your A, so this basically an op-amp and this gives you a feedback region and this $R_1$ is responsible for giving you a feedback to the

amplifier and β therefore will be equals to $R_1 /(R_1 + R_2)$ right, because there are in series with respect to each other.

Similarly, if you want to make it much more simpler, you will see that it is exactly the same as this one here right and this is the feedback network which you see here right, this is a common source stage because, sorry common gate because you are giving the source input signal to the source side of it and so source, so $V_i$ will be between $V_{GS}$ which is between this two points and therefore you will automatically get a variation in the output frequency, output voltage here.

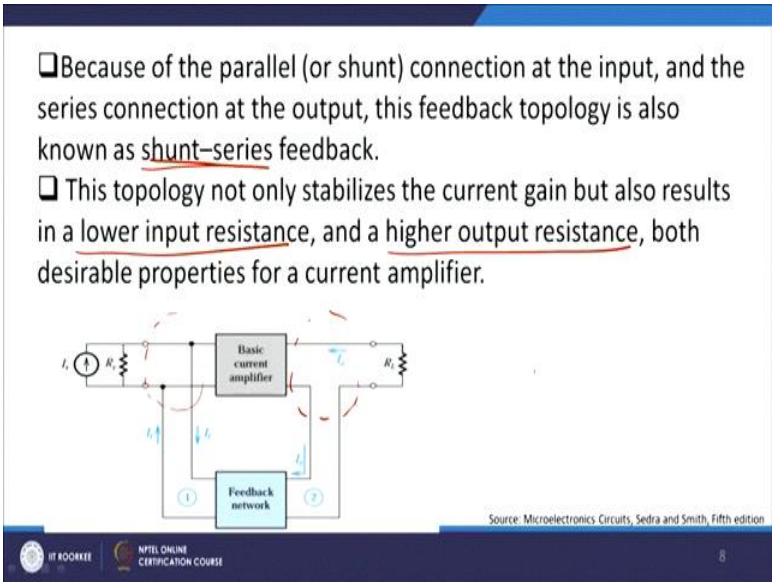
(Refer Slide Time: 20:09)




Let me come to the current amplifier here now, first was a voltage, this is the current, so input is a current output is also a current, so and therefore since it is a current amplifier we represent it by a Norton equivalent right, so if it is a voltage, you must be knowing from your basic network theory courses or you electrical technology courses that voltage sources or voltage related can be done very well with Thevenin and current can be done very well with a Norton equipments.

The output interest as I discussed with you is current hence the feedback network should sample output current and therefore output current will be sampled in the output side, the feedback signal should also be a current, so it may be mixed in a shunt with the current source and the feedback, most feedback topology is basically current mixing and current sampling topology and I will see to you how it works out.

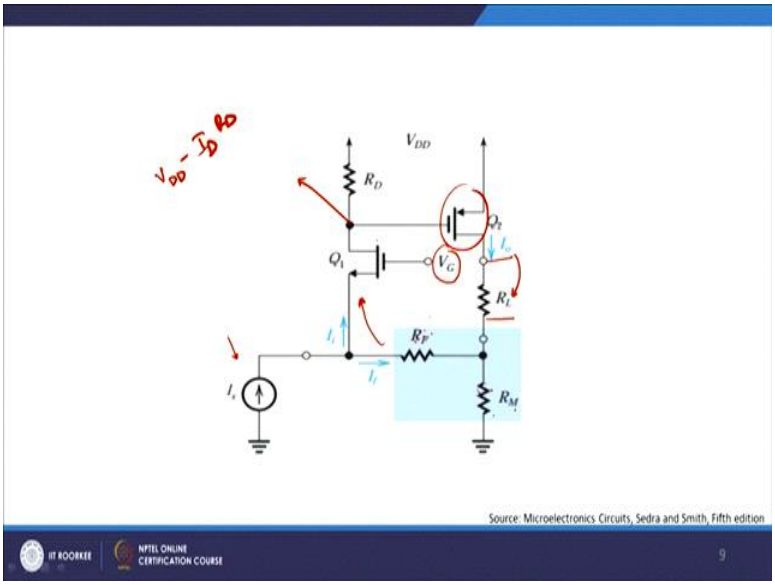
(Refer Slide Time: 20:58)




Because if you have understood voltage amplifier, this is just a complimentary form of voltage amplifier, this is basically, the previous one was series shunt, this is basically a shunt series amplifier, so we take series in the output side because we are trying to take out current, so the output side, you will have a series connection and the input side your feeding a voltage, so you will have automatically a voltage consideration right, as I discussed with you, you will have a lower input resistance and a higher output resistance, just the reverse of the previous cases and please understand that when you are discussing a shunt series amplifier for example, you have to be very cautious that it is basically a current source right, it is basically a current source, which you are referring to.

Now if you remember a current source output impedance is infinitely high right and as a result what will happen is, when the current source infinite is relatively high, you want your output impedance to be high with feedback because if your output impedance is high with respect to feedback, you automatically get much more stabilized current source with you and that is the reason currencies feedback or a current, current series helps you to achieve that phenomenon.

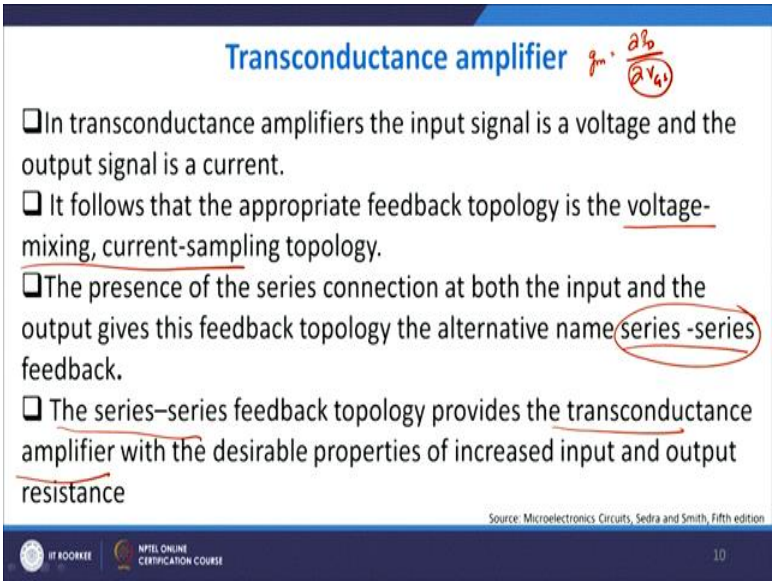
(Refer Slide Time: 22:18)




This is an example of your current series feedback network and as you can see here, I am feeding the through a current source here $I_s$ and this is my gate voltage which is fixed, of course, this you as a discussed with you, you will be taking output in the parallel side, so the parallel is basically $R_F$ and $R_M$ are in parallel with respect to each other and therefore depending upon the voltage appearing across $R_L$, part of it will be divided between $R_F$ and $R_M$ and there will be a negative feedback to you $Q_1$ right.

As a result this $Q_1$ will be reversed biased with respect to each other, as I discussed with you earlier also the voltage here will be nothing but $V_{DD}$ - $I_DR_D$ and therefore that voltage will be switching on my $Q_2$ and there will be current $I_0$ flowing through this point, $I_0R_L$ will be the voltage, that voltage will be divided between $R_F$ and $R_M$ and a part of it will be feed into negative feedback of $Q_1$ and you will get a negative feedback amplifier available in this case.

(Refer Slide Time: 23:10)

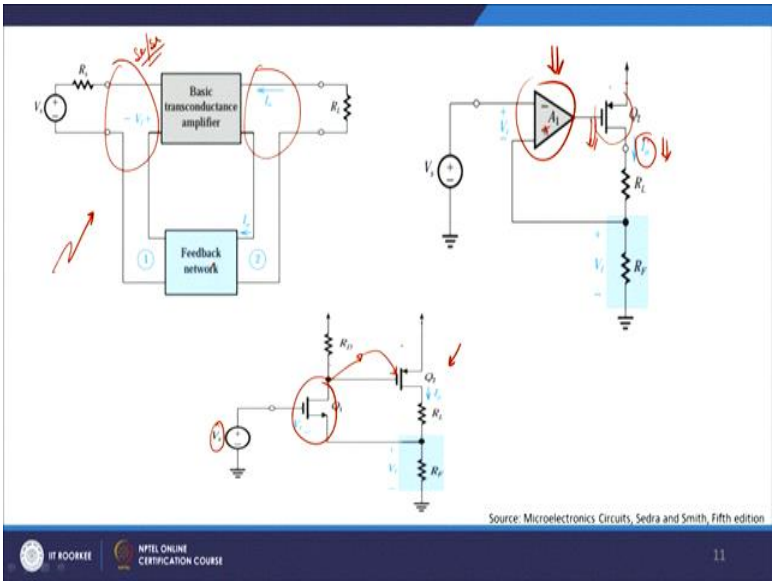


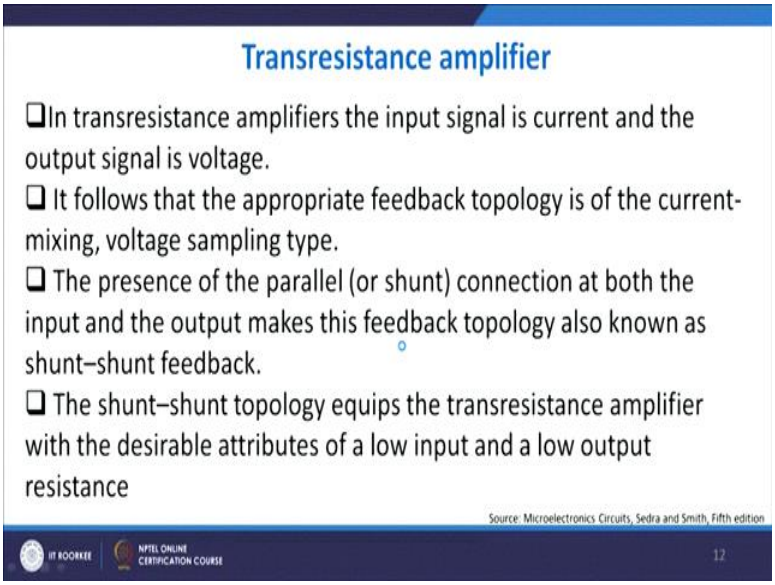

Let me come to the transconductance amplifier, a transconductance amplifier is that when you are, so the first one was voltage, so input was, voltage output was voltage, I used VCCS, which is voltage control current source, then we have a current control, current source, so it is a current amplifier second one, now we will have the first time where the input is basically a voltage and the output is basically a current, so my input, so I will have a voltage controlled current source right, so I have voltage controlled current source, so VCCS right, so the first one was voltage controlled voltage source, the second was current controlled current source, the third one is basically voltage controlled current source.

So as I discussed with you, you will be sampling, you will be mixing voltage and sampling current in this topology, this is also referred to as a series-series, because both sides sampling and input side will be having both side will have series-series combinations with you and that is the reason you can have a, it is also known as series-series feedback topology, provides transconductance amplifier, which means that it takes voltages as the input, remember where transconductance term is coming is basically $g_m$, so it is $\partial I_D / \partial V_{GS}$ right.

So if you look at transconductance is refer to as $g_m$ and refer to as $\partial I_D / \partial V_{GS}$, so which means that I give an input voltage a gate to source voltage which results in output current right, so voltage to current amplification is been shown in this case for all practical purposes.

(Refer Slide Time: 24:36)



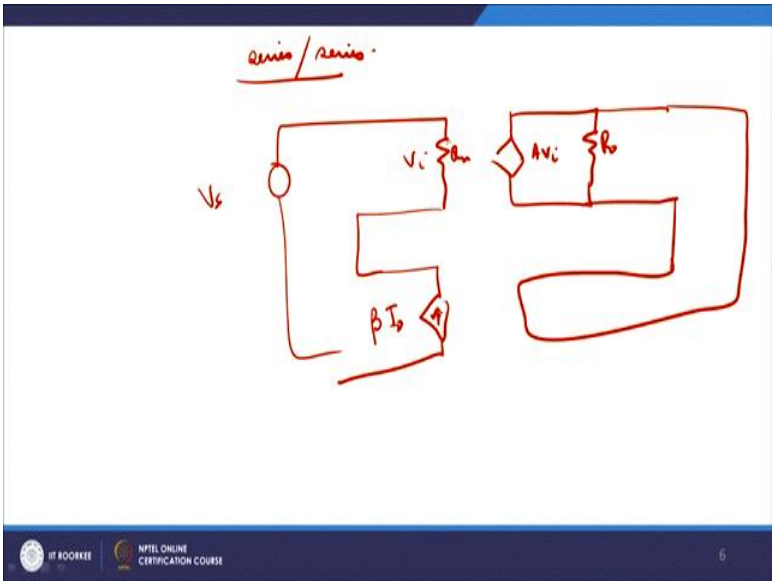




This is an example of a basic topology of your series-series network right, because both sides are series, so you have an sampling side as well as your in mix side, both side are basically a series-series combination here and you have a feedback network coming here right, if you want to do a analogy with respect to real circuits, well, you have an op-amp here and you do have a, this drives your $Q_2$ right, and as a result, you will see this is basically a P-MOS right.

So and you giving to the negative side here $V_I$ and a positive side is connected to the feedback network, so when this voltage, when this current becomes larger and larger, the voltage drop across $R_F$ increases, this feedbacks in a negative sense to $A_I$ and as a result, the voltages starts to fall, the current starts to fall here right and therefore this is basically an example of a negative feedback network for all practical purposes.

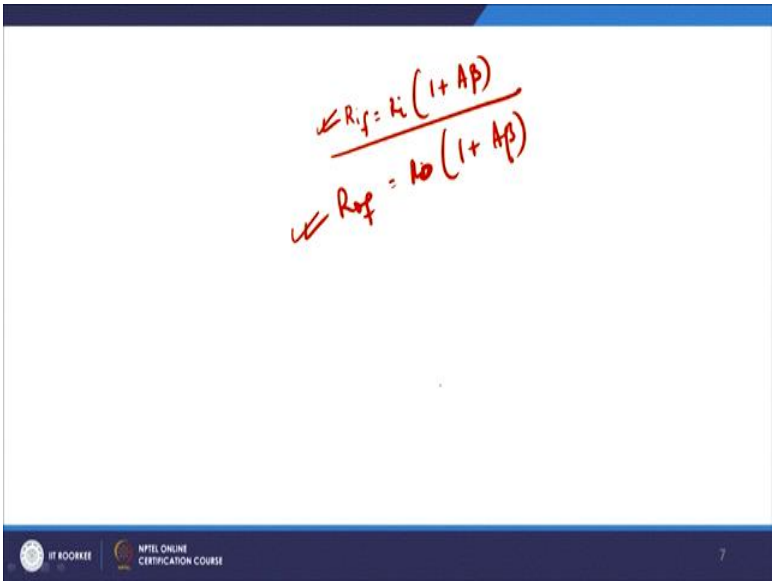
The same example can be seen here as well, where we have replace this op-amp by this $Q_1$ network here, so this is basically a transconductance because we are giving a voltage source, $V_S$ here, which converts this into a current and that converts into a voltage here which feeds this gate of $Q_2$ and the same conceptually this remains the same, so you have series-series network here right and then you do have a series-series network which comes into picture.

(Refer Slide Time: 26:02)




Now if you look at the series-series network, let me just right down for you, the, for the series-series network, if you write down series-series, you will see that I can write down in this format that you have got a currents, your voltage here and then of course you will have a current here and then you will have $R_0$ here right, this is $R_i$, this is $V_i$ and the same thing goes on this side and then, alright and then this is shorted and then you have got $\beta$ times $I_0$, because this will be a current, this will be a current and then this will be going like this and this is the voltage source $V_s$ right.

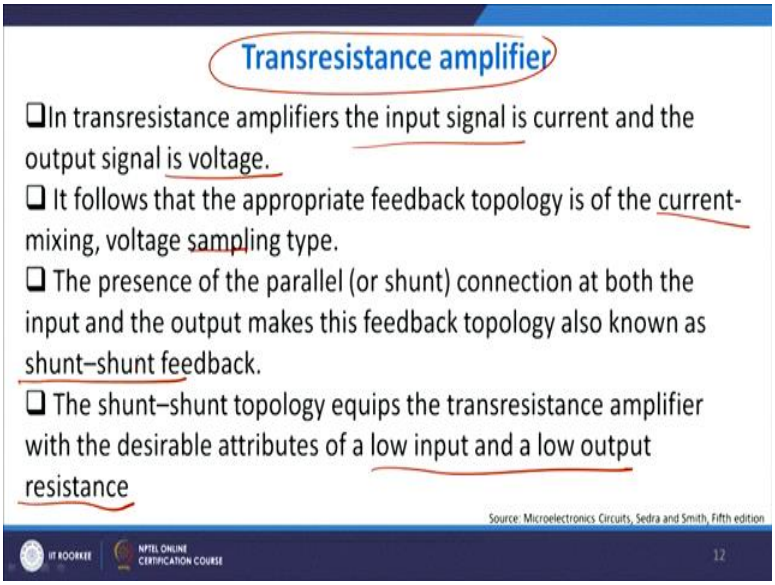
(Refer Slide Time: 26:43)




If you take into consideration and do the same formulation as we did earlier, I get $R_{if}$ to be equals to $R_I(1 + A \beta)$, so your input impedance also rises and $R_{of}$ is also equals to $R_o(1 + A \beta)$ in this case, so your input impedance as well as your output impedance both rise with respect to negative feedback in a series-series combination, so whenever, just you can, even if you do not understand the mathematics of it, whenever you have a series combination, the impedance rises, obviously.

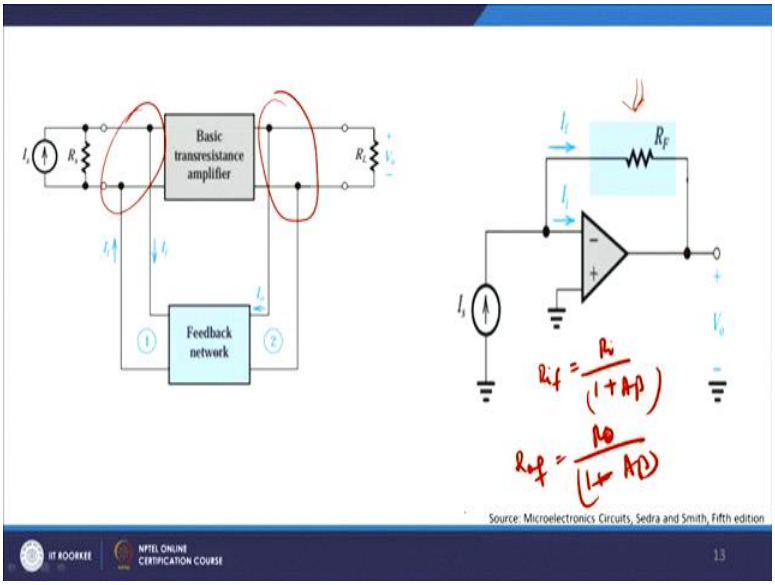
And you have a parallel combination, the impedance falls, now when you have a both side, if you have both side shunt, then both the impedances will fall, both sides series, both the impedances will rise, one side series, one side shunt, the series side the impedance will rise and the shunt side the impedance will fall, with respect to feedback, so that general rule of thumb is most important for you to remember as far as feedback theory is concerned.

(Refer Slide Time: 27:40)




We come to the last part of the amplifier which is basically my transresistance amplifier, in this case, just it is just the reverse of the previous case, where my input is basically an input signal and the output is basically a voltage right, so I will have current mixing and voltage sampling here and it is also referred to as a shunt-shut because both sampling and mixture will have shunt-shunt regulation and so as I discussed with you, it will give you a low input as well as a low output resistance because it is a shunt-shunt feedback resistance.

(Refer Slide Time: 28:11)




And you see that I am taking a shunt here, I am taking a shunt here and therefore this is my $R_S$ source resistance, load resistance in the feedback network and one of the best examples is basically your operational amplifier design, wherein you will have $R_F$ here which is basically,

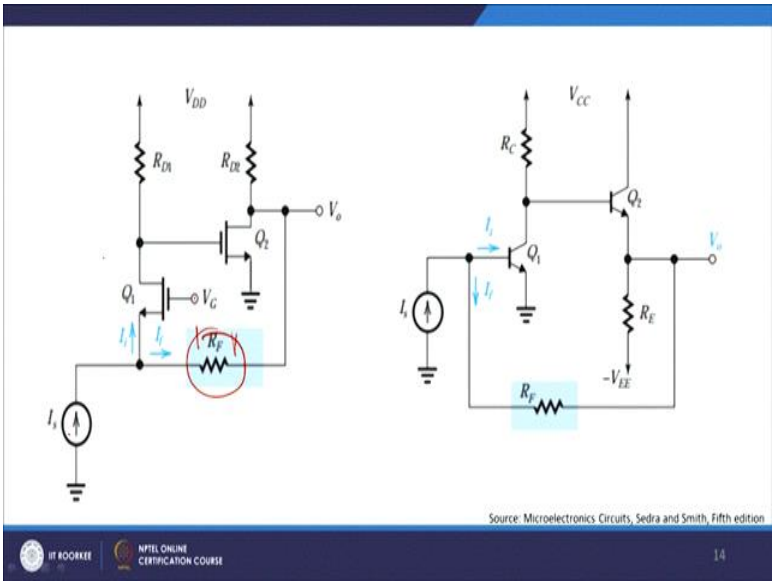
so you take a shunt here, because you are taking parallel, from here you are taking parallel output from here and therefore you are sampling in it parallel and you are mixing it in parallel in this case, so basically this is a shunt-shunt network.

Here in this case, you get $R_{if}$ to be equals to $R_i / (1 + A \beta)$ and therefore there is a drop and $R_{of}$ is also equals to $R_o/(1 + A \beta)$, so in both the cases $R_i$ as well as $R_f$, your output impedance with feedback actually falls down drastically and therefore there is a decrease in the value of your shunt-shunt resistances in this case, so one thing you should always remember is that in presence of shunt-shunt the impedances fall down, series-series both the impedances rise up, shunt series and series shunt, series will rise up, shunt will fall down and in all the four cases your gain will always drop down, if you have a negative feedback.

So this is a common thread across all the amplifier which you see, you will ask me why this is important? Well, this is important because now you have an amplifier because you remember your maximum power transfer theorem from your basic network theory courses, wherein you need to make your output impedance of the driver equal to that of the load, input impedance of the load for a maximum power to be transferred, so when you have these type of amplifier sitting in between your load and your source, then you can change your output impedances in such a manner that it becomes almost equals to the input impedance of the external load and therefore the maximum power will be transferred from the chip or on the circuit to the load, external load right.

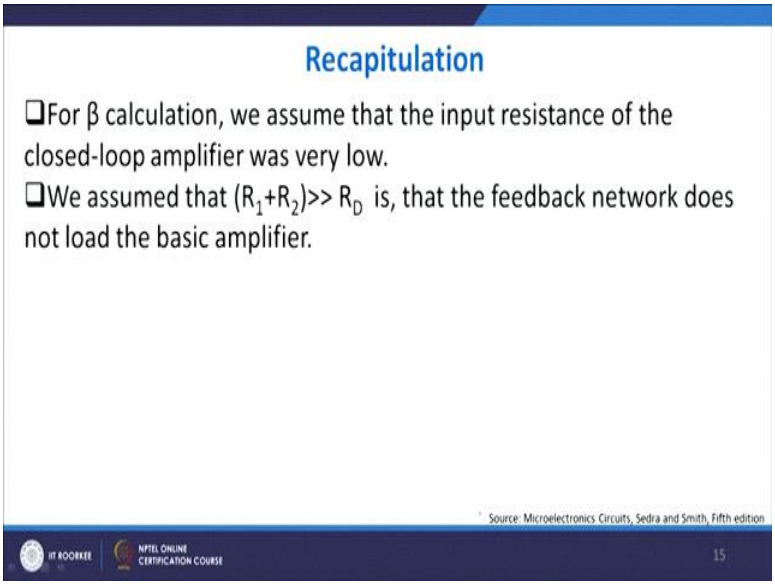
So that is one of the major advantages of having feedback that you can control the impedances in the input and output form by simply choosing the type of configurations which you want to do right and that is the important part which you will see here.

(Refer Slide Time: 30:18)




There is another example, I will not talk about the BJT maybe, maybe MOS device you can see here, this is basically shunt-shunt again, so you see here, I am giving you $V_G$, I am giving $I_S$ as in the previous case, this $R_F$ is basically in series to this one, as a result this voltage across this will be feeding back this in reverse direction and you are automatically get a negative feedback concept in this case right, for a shunt-shunt case right.

(Refer Slide Time: 30:39)




So let me recapitulate what we did. What we did was that two important assumptions, we have taken that there is no loading of the amplifier or the feedback network, in reality not true, you will always have a loading, howsoever the loading is low is still there, the second thing is we have taken them to be almost independent of frequencies, but therefore our

definition of understanding here stands very well within the domain of mid-frequency band, where your gain is almost stable with respect to frequency right and at higher and lower frequencies, if you remember, the gain starts to fall down and therefore there these things might not be very valid in the sense.

We have also assumed that the feedback network does not load your basic amplifier, well, that is very important that it does not load and this gives me a proper feedback, advantage of such a network, also studying such topologies is, that you have a methodology with you by which you can vary the impedance levels and input and output of an amplifier and therefore you can have a good driver available with you while driving an input signal to an external load, right, so with this, let me thank you for your patient hearing and we will talk about other topologies in the next module. Thank you very much.

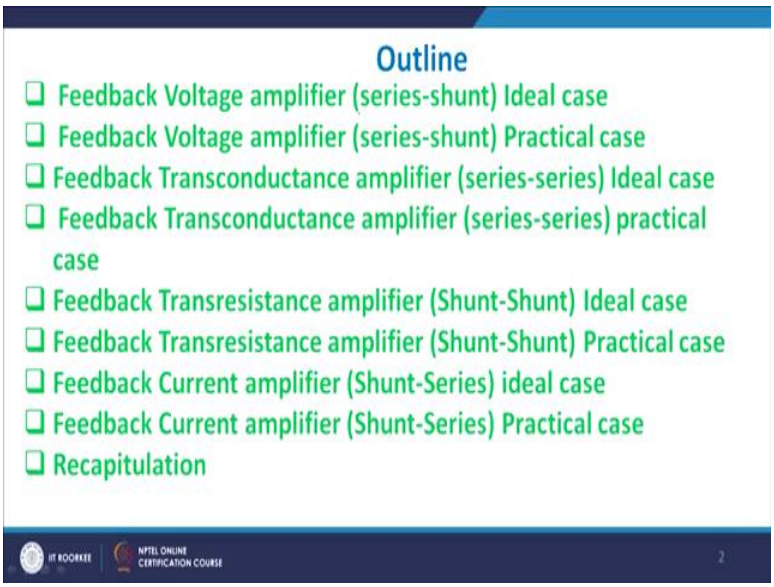
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−47**
**Design of Feedback Amplifier for all Configuration**

Hello everybody and welcome to the next edition of NPTEL online course on Microelectronics Devices to Circuits. Today we will look into the design of feedback amplifiers against configuration various configurations of the feedback amplifier. In our previous module we had looked into the fact that what is a negative feedback? How does a negative feedback work? What are the criteria for the multiplication of A and β which is basically the forward feed forward amplifier and β which is basically my feedback ratio so that I am able to achieve a value of negative feedback.

And we saw that when Aβ is greater than 1 we will automatically get a lower gain. The advantage of a negative feedback we get a higher bandwidth, we get a less distortion nonlinear distortion. We also have a much more stabilized gain in the output side. And these were the few advantages which we saw. We also saw in our previous term that when you do have those configurations of series shunt or shunt series network we generally are able to modify our input impedance and output impedance of an amplifier.

And the advantage of such therefore negative feedback is that I am able to therefore have my output impedance change with various configurations. Now, since the amplifier output drives the extensive load and the load will be varying load, I can therefore match the output impedance of the amplifier with the load itself for the maximum power to be transferred. And I can use, how can I do that, by applying various configurations of negative feedback. So, we have learnt 4 series shunt, series series, shunt series and shunt shunt, right? And we have seen how the input impedance and the output impedance change.

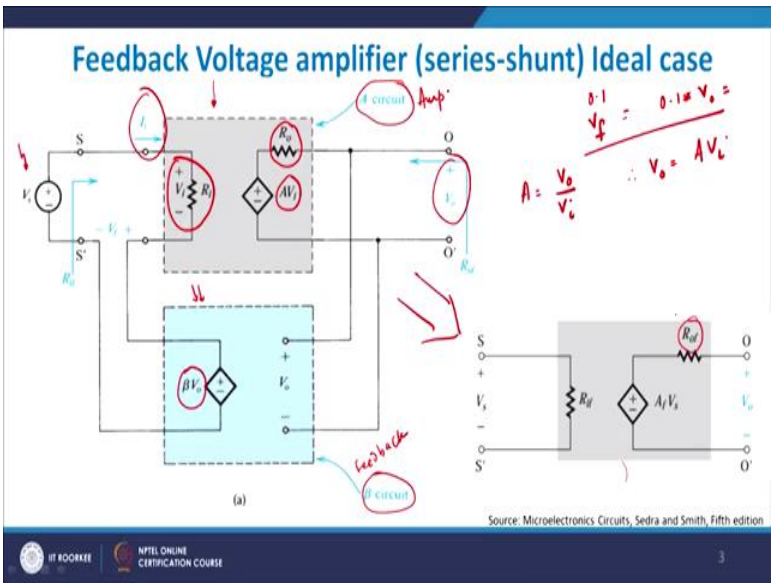
(Refer Slide Time: 2:31)




Today we will the topic today's topic is basically by feedback voltage amplifier. So, we look at ideal case of a series shunt amplifier, right? And then we will look into a practical case of ideal of the series series shunt. Then, we will look into an ideal and real case of series series feedback network, right? So, we will look into series series feedback. We will also look into shunt shunt case. So, for all the cases we will look into ideal case ideal case when you do not have any loading, right?

Which means that your feedback network does not load your amplifier and your amplifier does not load your feedback network and there are 2 independent entities that we have assumed. In the second case we will take a practical amplifier; we assume that there is a loading effect coming into picture. So, we have all the 4 amplifier voltage amplifier, current amplifier, transconductance amplifier and transresistance amplifier for ideal case as well as the real cases.

(Refer Slide Time: 3:32)




So, let us look at the feedback voltage amplifier. This basically is series shunt and let me let me show you the series shunt feedback network here. As you can see here this part which is basically my dotted part here is basically the amplifier. So, this is the A circuit when I say it is basically your amplifier circuit and the blue one is basically my β circuit, which is basically my feedback circuit. So, this is my feedback circuit, right? And this is my amplifier circuit, right? So, amplifier is a feed forward and this is basically my feedback.

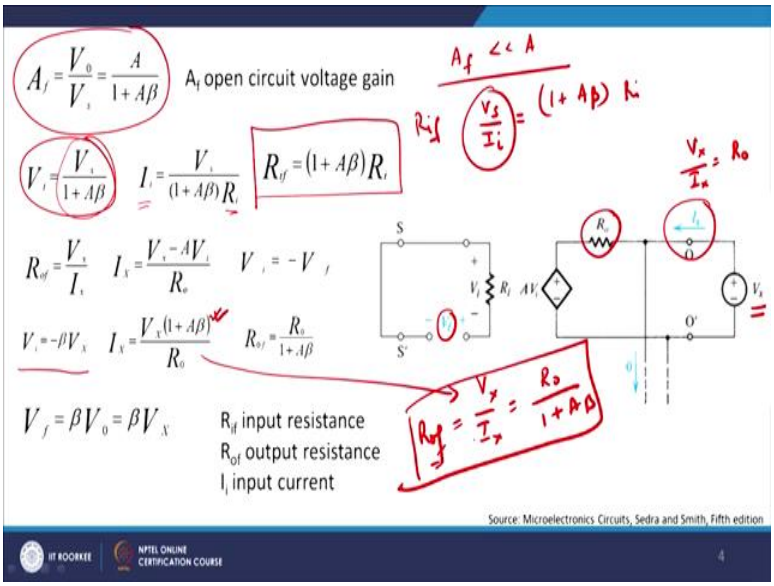
Now, this is series shunt which means that I will be extracting voltage from the output and feeding it in the form of current in input side. So, the input side will be input current here, right? And you have an output voltage which is $V_0$. At this $V_0$ therefore appears here, right? And a part of $V_0$ appears as a voltage source in the output feedback network. So, say your β is equal to 0.1. It primarily means that $0.1*V_0$ will be your output voltage here of this network, right? of this network. So, your $V_f$ will be equal to $0.1* V_0$.

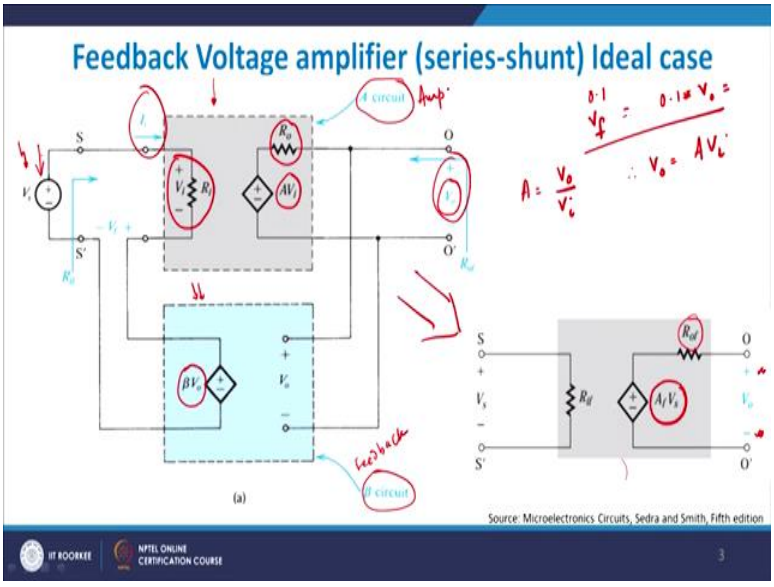
And as a result this is how we will generate the feedback ratio. $V_i$ is basically $V_s$ is the source voltage which you see out of which a particular voltage $V_i$ appears here depending upon the value of $R_i$ which is input resistance of the amplifier. Now, typically if it is voltage gain $A=V_o/V_i$, right? And therefore, $V_o = A V_i$. That is what is written here $A* V_i$ which you see. You

also have a output impedance of the voltage source here and therefore, that maintains and the output side, right?

And depending on the value of β and $A_0$ we will see how it works out. So, therefore we can converge into this factor in order to get this form that I have a source voltage which terminates onto $R_{if}$ which is basically my input resistance with feedback. I have a output resistance with feedback $R_{of}$ f suffix means feedback. And my voltage gain is given as $A_f$ * $V_s$, right? And my input voltage will be given by $V_0$ here; it is given by $V_0$ here.

(Refer Slide Time: 6:05)



So, with knowledge or this basic idea let me show to you how I am calculating the $R_{of}$ and $R_{if}$. Now, we know that $A_f$ will be equal to $V_0/ V_s$, right? $A_f$ means that amplifier gain with feedback, right? Amplifier gain with feedback primarily basically it effectively means that that I am trying to look at this voltage divided by this voltage, right? And that is what we are trying to find out when we say $A_f$. And if you if you solve it you get $A/(1 + A\beta)$ which we have already earlier also and we saw that $A_f$ therefore is always less than A, right? Therefore, there is a drop in the gain.

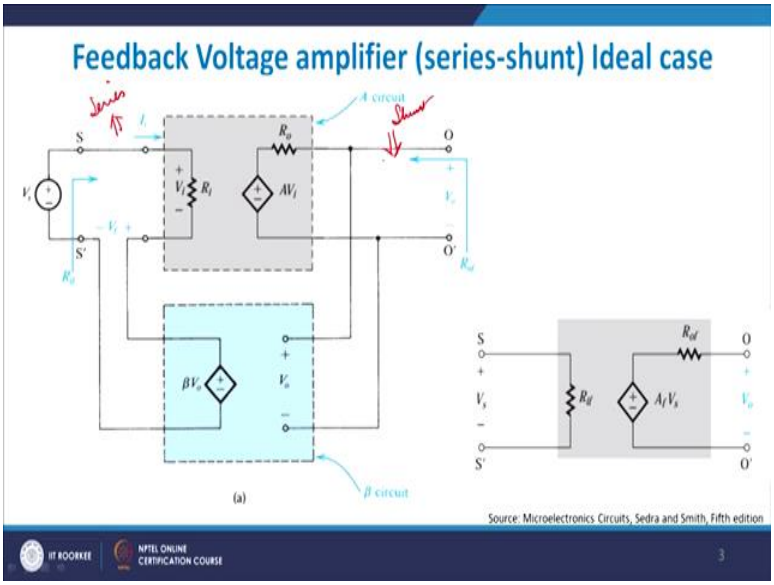
Now you see $V_i$ will be equals to $V_s / (1+A\beta)$. From where did you get this $V_i$, $V_s / (1+A\beta)$ is basically that so what is $V_i$ is the input voltage. Where is the input voltage actually referring to? Input voltage is basically this voltage which you see, right? So, I am looking at $V_i / V_s$ this by this will nothing but be a voltage divided network and I get $V_s / (1+A\beta)$. Therefore, if you want to find the input current, you have to divide this whole quantity by $R_i$. That is what I am doing here.

Now, so therefore, if you look very carefully if you transfer this to right hand side I get $V_s$ / Ii is basically equals to $(1+A\beta)*R_i$. This is nothing but $R_{if}$. So, $R_{if}$ equals to $(1+A\beta) * R_i$ which means that input impedance rises with feedback. Look at the output impedance. How do you find output impedance is very simple. You try to short the input and give a feedback voltage here and then apply a source voltage $V_x$ at the output side and try to find out the current. $V_x/I_x$ will be equals to $R_o$, right?

And with respect to feedback I will get the value of ro. So, I write $V_i$ is equal to $- \beta * V_x$. Why this why there is a negative sign? Because you are doing a feedback. So, a part of $V_x$ appears as $V_i$ input side $- \beta *V_x$. So, if I want to find $I_x$ which is the current flowing through this R. This this current I want to find out. Then this will be nothing and $V_x$ is the voltage divided by $R_o$ which is this 1, right into $(1+A\beta)$, right? Into $(1+A\beta)$. And therefore, if you want to find out you will get so this equation you will get basically $V_x / I_x$ will be equal to $R_o / (1+A\beta)$.

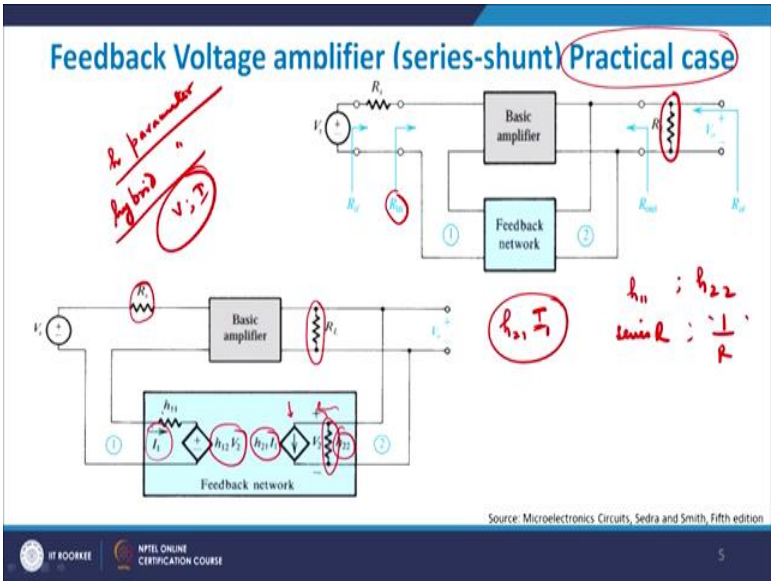
And this is nothing but $R_{of}$. So, I get $R_{of}$ is equal to $R_o / (1+A\beta)$ which means the output impedance which we feedback falls down by falls back to the $(1+A\beta)$. But my input impedance rises by a factor of $(1+A\beta)$. So, my input impedance rises and output impedance falls down. You can see it physically also.

(Refer Slide Time: 8:52)




Why because if you look very carefully then you will see that that this is basically a series addition here. And therefore, the impedance levels will go high. This is a shunt or a parallel addition here and therefore, impedance is lower down. That takes cares of approximately all the understanding here.

(Refer Slide Time: 9:09)

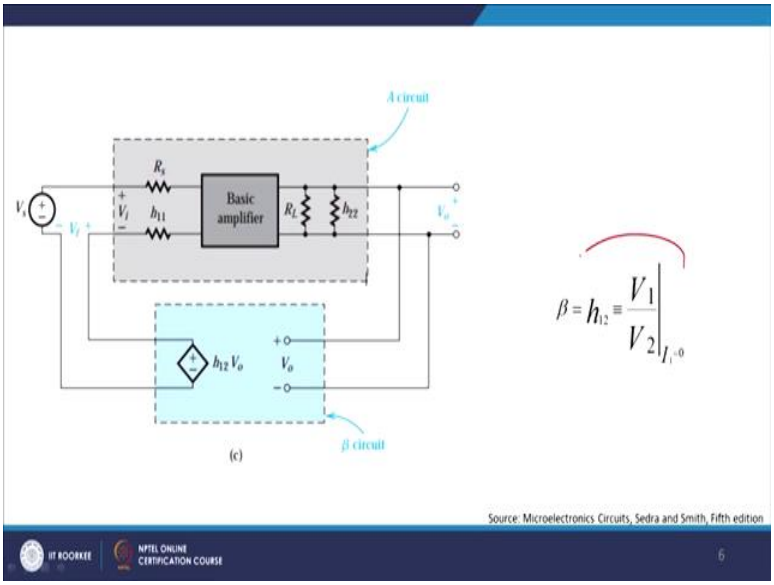


Now, we do the same thing for all the 4 networks and see how it works out. This is basically a series shunt the previous case but let us look at practical case in in reality which means that you do have a loading here. I do have a load here and I have $R_{if}$ is input which you see, $R_{in}$ is input resistance without any feedback and $R_{if}$ is input resistance with feedback. Now, $R_S$ is basically the voltage source here. What we try to do is that we try to we try to convert into a 2 port network, right?

Which is basically my h parameter h parameter analysis which we do h parameter analysis. h parameter is also referred to a hybrid parameters and it correlates voltages and currents together, right? So, so what we do is basically is that we design the feedback network properly and we say that there is a load resistance which affects your basic amplifier. I have a source resistance here. Out of the voltage available at this point which is $V_0$ I say that I can model is $h_{21}$. $h_{21}I1$ and this will be basically $h_{21}$ will be basically a current amplifier and into $I_1$ if you do I will get the overall current flow.

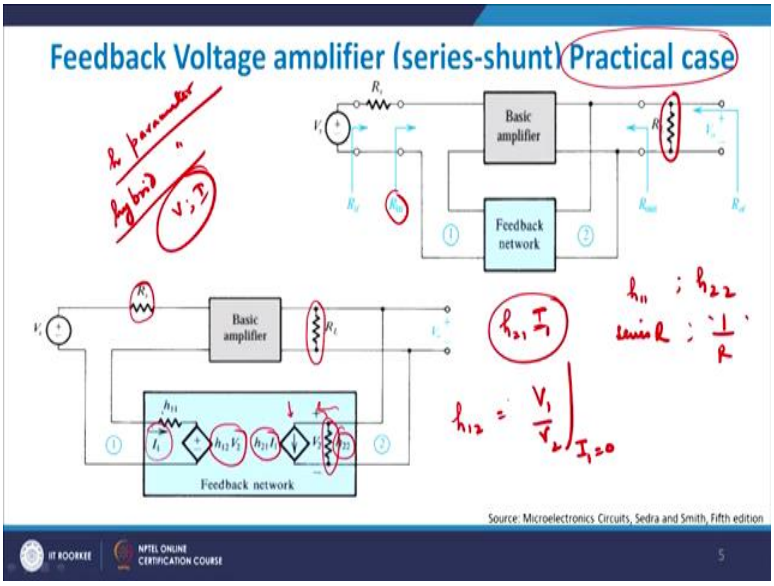
Therefore, this is represented by current source and similarly $h_{12}*V_2$. So, there are 2 terms here $h_{11}$ and $h_{22}$. $h_{11}$ is nothing but basically a series resistance series resistance. $h_{22}$ is basically 1 over resistance term. You see it is coming parallel to the current source, right? As you can appreciate as well because current source is ideal current source is we automatically have a large output impedance and that is taken care by $h_{22}$ in reality. $h_{12}$ $V_2$ is nothing but the voltage source you see here. So, it is basically a voltage source here and current I1 is flowing through this R. And current $I_2$ is flowing through this R and this is $I_2$ in a 2 port network. This is $h_{11}$ and $h_{22}$ which you see.

(Refer Slide Time: 11:19)




With this knowledge you can safely write down $h_{21}$ is nothing but $V_1 / V_2$ with $I_1$ equals to 0.

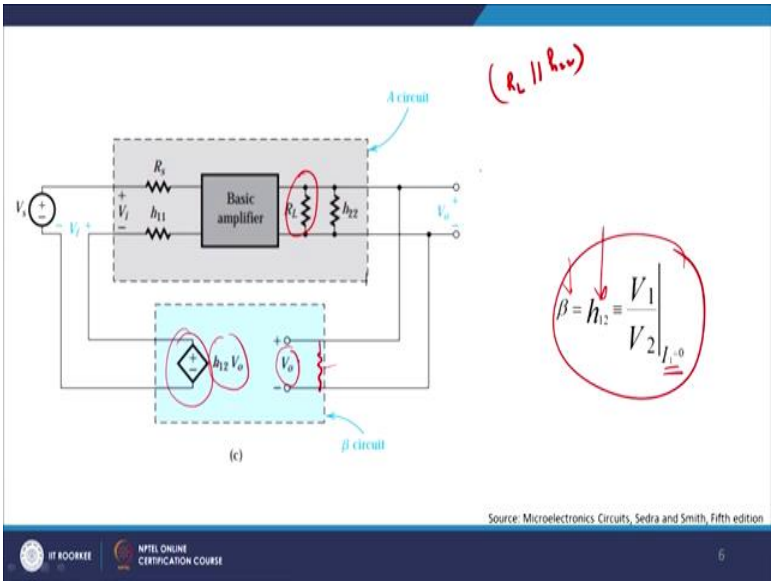
(Refer Slide Time: 11:27)




What is $h_{12}$ therefore? $h_{12}$ is nothing but if you look very carefully $h_{12}$ I am referring to as $V_1 / V_2$ with $I_1$ equals to 0. $I_1$ is open $I_1$ is equal to 0. $I_1$ 0 primarily means that you are opening the input site and that is pretty important that you are opening the input site. So, the input site is open. In

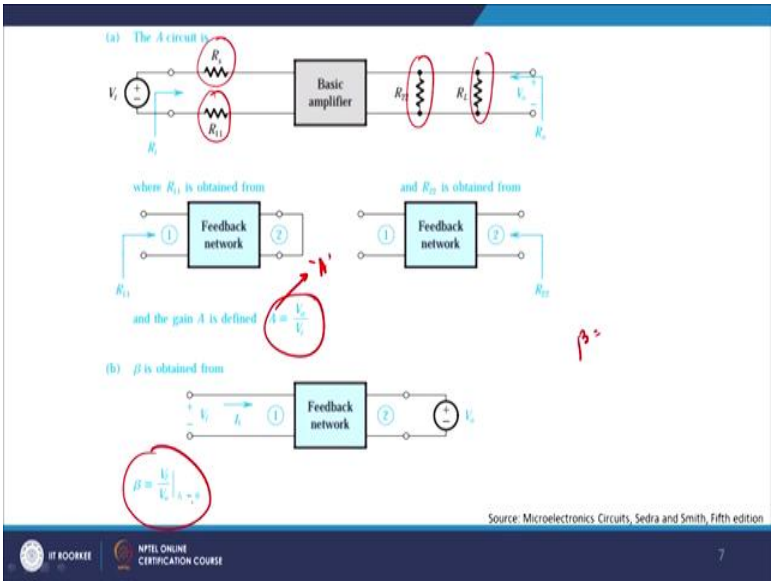
this case this is the input side and this is considered to be open. In that case I get $h_{12}$ equals to ratio of the 2 voltages $V_1$ / $V_2$.

(Refer Slide Time: 11:59)




And therefore you write here $V_0$ which is the voltage source and then $h_{12}$ $V_0$ because this is the input voltage. This input voltage multiplied by $h_{12}$ will give you the value of the output voltage and there is also referred to as β and also referred to as $h_{12}$, right? And I get therefore I get so what do I do is that I shift this $h_{22}$ which was initially here back to this place. So, now you have got $R_L$ parallel to $h_{22}$ as overall loading on the output side, right? Similarly, the $h_{11}$ which was initially here comes in series to $R_s$ we place it in series to $R_s$. And then, that is what we are trying to do in this case.

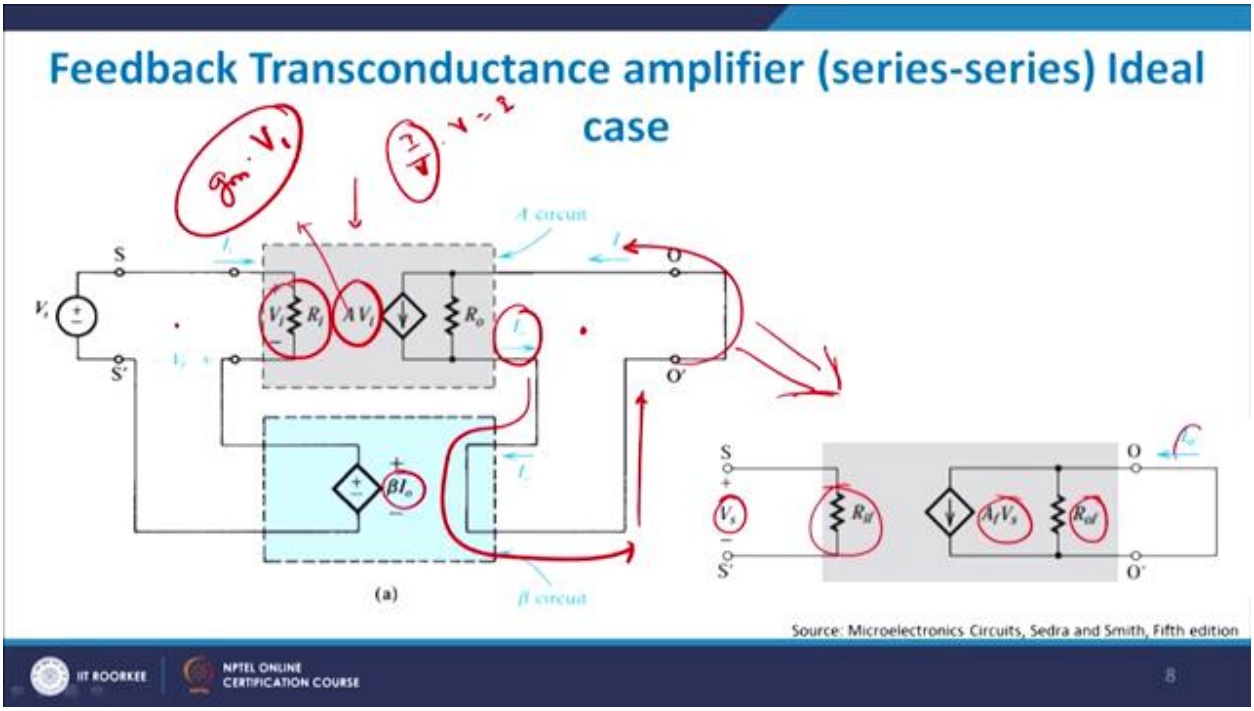
(Refer Slide Time: 12:39)




So, if you therefore fall back and let us see what or how it works out we see that the amplifier circuit will be modelled as in $R_{11}$ in series to basic amplifier. And $R_{22}$ and $R_L$ parallel to the basic amplifier in terms of loading, right? So, as I defined to you the gain a is $V_0/V_i$ which is where $V_0$ is the output voltage and vi is the input voltage and $V_0/V_i$ will give you the amplifier gain which is basically a here, right? How do you obtain the how do you obtain the value of β which is the feed? So, β is referred to as $V_f/V_0$ when I1 equals to 0 and β is obtained from this network that what you try to do.

You try to open circuit the input side. When you open circuit the input side, $I_1$ equals to 0. And then what you do? You do voltage $V_0$ and then try to find voltage at these 2 points by $V_0$ using the value of feedback factor β, right? And that is what an interesting part of this whole network part is there.

(Refer Slide Time: 13:30)



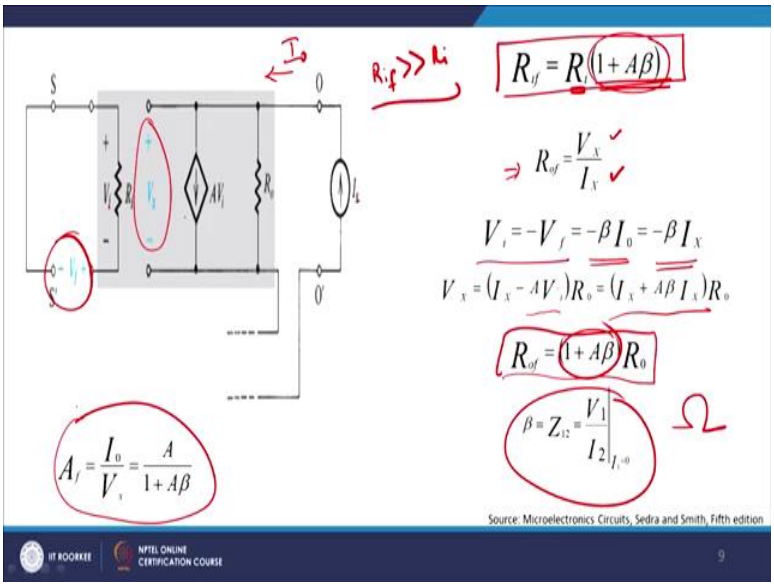

We come to now the Feedback Transconductance Amplifier and this basically is a series series case. Series series as you can see that I am adding the voltage and current here in series and extracting current here in series as well, right? So, I have a current source here which is giving me $I_0$ output volt current. This flows through this path, right? And you get this into consideration and this current flows back into the system and therefore I can replace it by giving $R_i$ here and $V_i$ here as in the previous case.

And then I say that $A*V_i$ is the voltage which you see here and parallel to that. But this will be a basically current source here. So, $A_i$ is basically I/V which you say, right I/V which you say. Multiplied by V gives you current and therefore, that is a current source here $A*V_i$. Here, $A_i$ is basically my $\partial I/\partial V_G$ which is basically transconductance. And that is the reason a will be referred as transconductance.

Many people refer to it as $g_m * V_1$ here also. And they get a reason for that, right? And the current source $I_0$ and then $\beta * V_0$ is the voltage which you see here. And that gives me the value of output voltage which is the feedback voltage at this particular point. You also therefore, can

come back to this equation this diagram and show that therefore I have therefore my input resistance feedback resistance here. I have my output with feedback $R_{of}$ here. And then $A_f * V_s$ is nothing but a current source which appears to me by virtue of my feedback current or feedback values. So, this is my source voltage and this is my output voltage in practical sense.

(Refer Slide Time: 15:11)



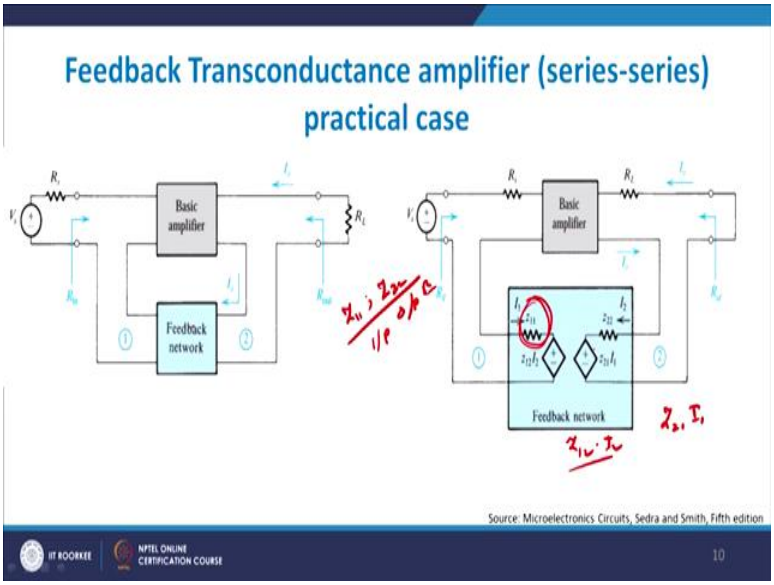

Now, the same thing we bring it back here. I will show to you what is what is the implication here. And in the both the cases you will see $R_{if}$ I am not doing the whole derivations given in all the standard books. I get $R_{if}$ basically equals to $R_i$ (1+Aβ) which means that the input impedance of the transistor in this case an amplifier depends on the factor of Ri which is basically without feedback the resistance available to me on the side multiplied by (1+Aβ). So, what happens is that $R_{if}$ is obviously greater than $R_i$ by a factor of (1+Aβ). Since, Aβ is relatively larger as compared to 1 I can safely say that $R_{if}$ is equal to $R_i$* Aβ.

Now, $R_{of}$ if you want to find out which is basically the output impedance then I can say simply it is equal to $V_x$ / $I_x$. Where is $V_x$? $V_x$ is this voltage which you see in front of you and this is the voltage which is appearing here upon $I_x$ which is the total flowing through the device. And that is your $R_{of}$. So, if I get $V_i$ equals to - $V_f$ because of negative feedback $V_i$ equals to - $V_f$ and therefore I can replace $V_f$ by β * I0 by my previous definition which can be also referred as - β * $I_x$.

So, I0 is the output resistance output current which you see here which is exactly the same as value of $I_x$ here. So, now what I try to do is I get $V_x$ equals to ($I_x$ - $AV_i$ )* $R_0$ because $AV_i$ will give you the current here multiplied by $R_0$ and then if you solve it I get A times. So, $V_i$ will be broken up into A*β *$I_x$ into $R_0$ and therefore, $R_{of}$ is equal to (1+Aβ) * $R_0$. So you see now as I was predicting earlier also the input resistance with feedback and the output resistance with feedback both increases by a factor of (1+Aβ), right?

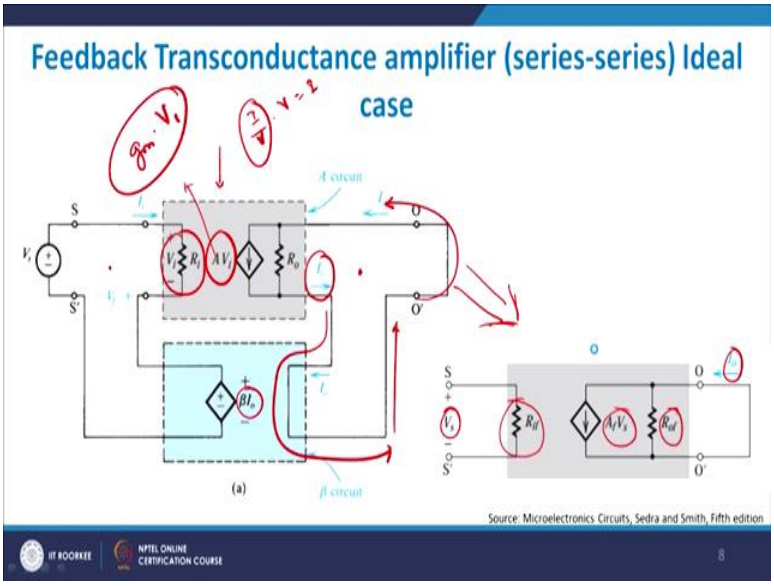
Now, that is quite a heavy increase here where A is the feed forward open loop up to the amplifier and β is the feedback fraction. β in this case is again referred to as Z12 and is given by $V_1/I_2$ and its dimensions in the order of ohm with $I_1$ equals to 0 primarily means that your input is basically open. So, you see that the input has been kept open here, right? I am trying to find the current available at. So, $A_f$ will be defined as A/(1+Aβ) as per my previous definition as we have already discussed this point in details.

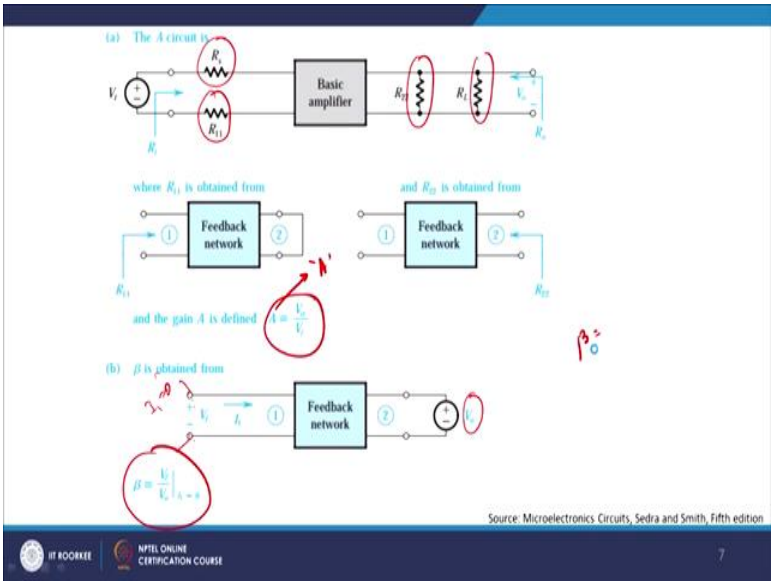
(Refer Slide Time: 17:35)

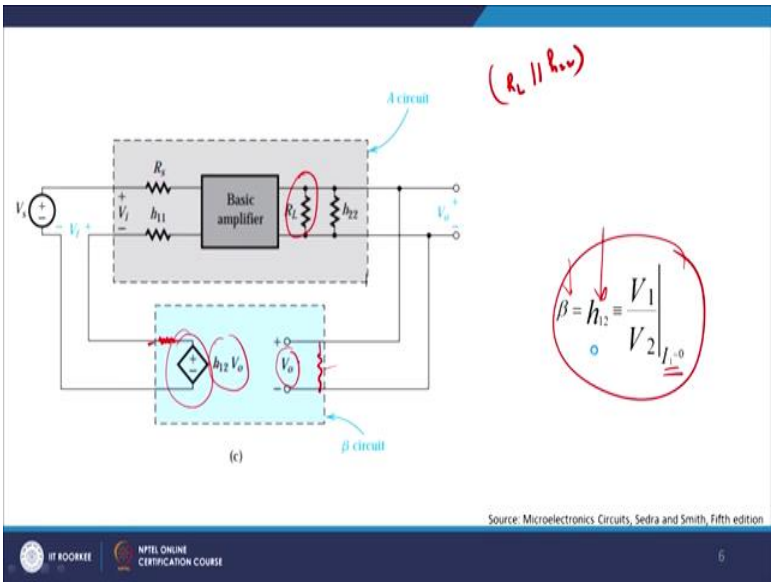



Now, let me come to a practical case of series series amplifier and you see here that I have broken down the feedback network into $Z_{11}$. So, it is basically a Z parameter which is available to me. So, $Z_{11}$ and $Z_{22}$ are basically the resistance input resistance and output resistance offered to the device and I get $Z_{21}$ $I_1$and $Z_{12}$ * $I_2$. So, you see your feedback network is changing.

(Refer Slide Time: 18:02)

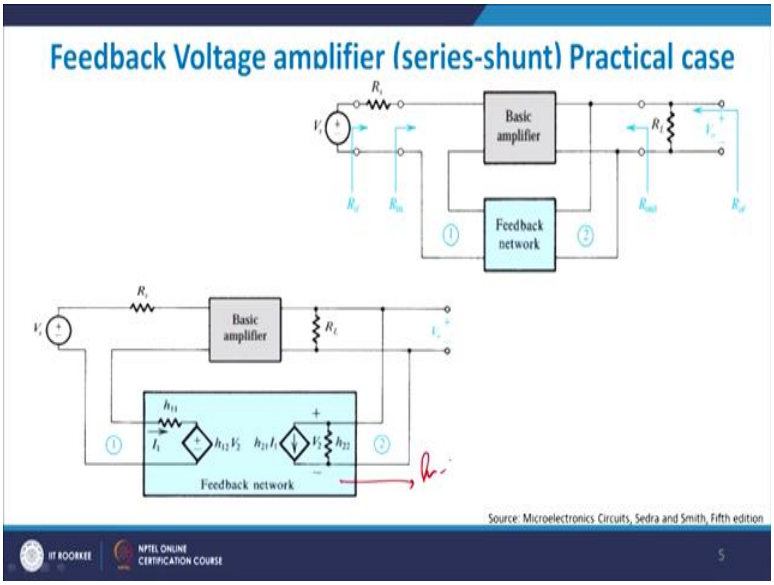




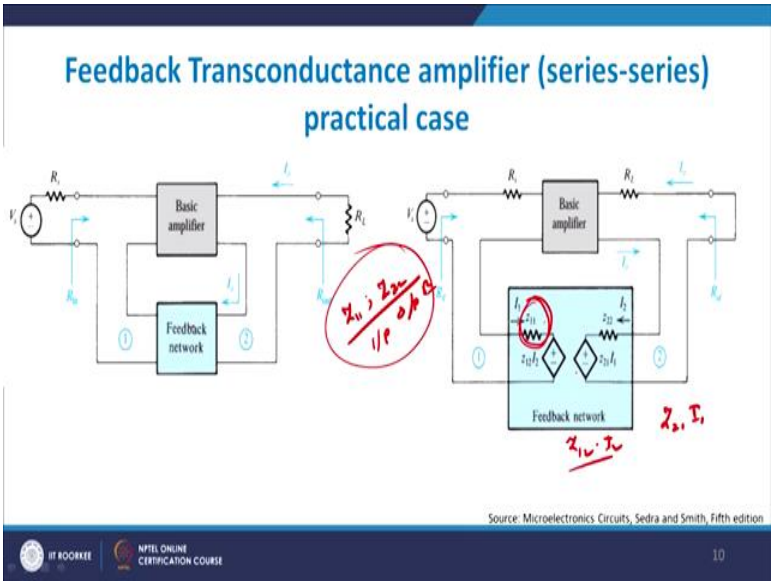

If you look back here previous discussion here, here or maybe this previous case that if you look here and you try to find out … if in this case you see we were talking in terms of.

(Refer Slide Time: 18:22)



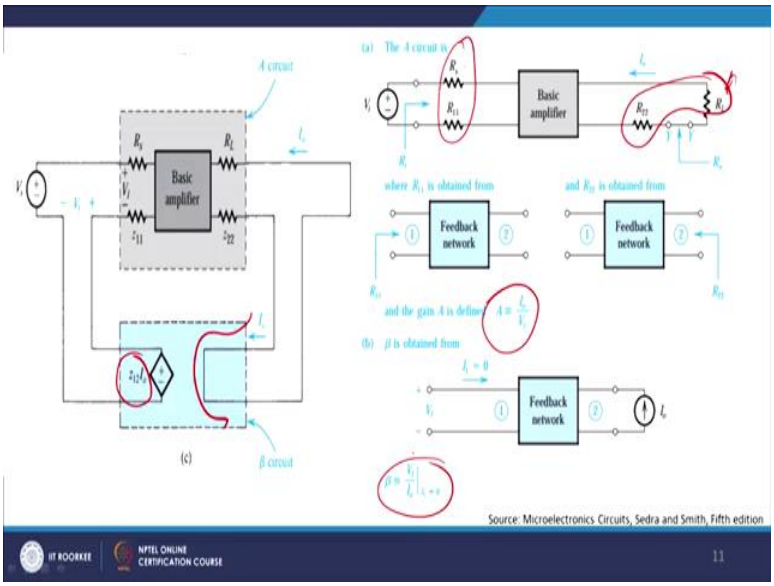

If you look very carefully here what we were talking was that your this was your h parameter network, right? h parameter h parameter because you are dealing with both current and voltages here.

(Refer Slide Time: 18:38)




Whereas now when you are dealing with both sides current or both side voltages or both side current maybe then you need to deal with the fact that you should have a current amplifier with you and that is what the reason is current amplifier is with us and therefore, $Z_{11}$ and $Z_{22}$ are nothing but voltage sources or the impedances seen to it. $Z_{12}$ $I_2$ is the current flowing through the device and $Z_{21}$ $I_1$ is the current input into the device in the feedback network, right, and I get $I_0$ as the total current flow into the device.

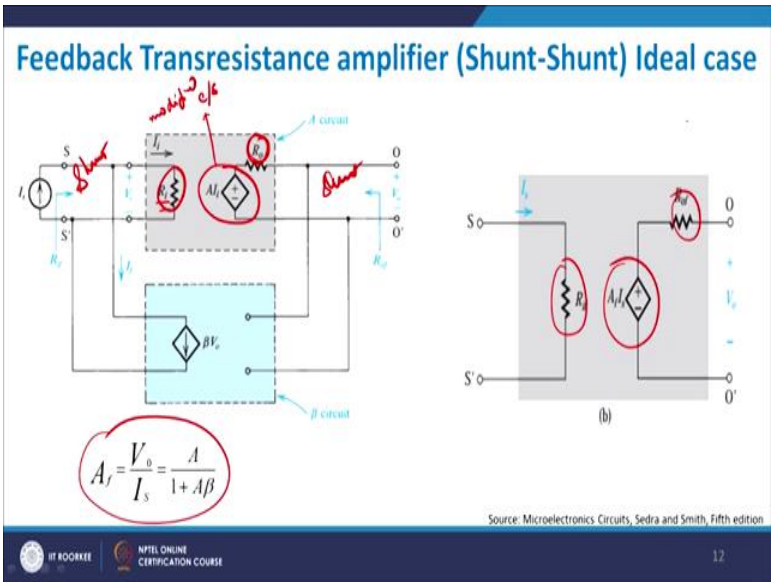
(Refer Slide Time: 18:59)




Now, we will look at the basic amplifier and then we will just take $Z_{11}$ from its initial position and make it into the $R_{SS}$ series value. And $Z_{22}$ in parallel position to in parallel position to $R_L$ and then I get $Z_{12}$ $I_0$ as the input current source here, right? So, I have a current source here which is shorted and I get $Z_{12}$ $I_0$ as the output current source which is fed back into the basic amplifier which converts the current into a voltage corresponding voltage and the output value is created. As I discussed with you in a typical example which we follow we see that these 2 are in input side and these 2 are in output side, right?

And we say that $R_{22}$ and $R_L$ which is the load external load are related to each other just like $R_s$ and $R_{11}$ is related to each other. We also define gain to be equals to $I_0$ by $V_i$ which is basically output current by input voltage. Now, β is obtained from the fact that if $I_1$ equals to 0 , I get β equals to $V_f/I_0$ with $I_1$ is equal to 0. This is how you calculate the value of β, right? So, its feedback voltage by input current that is how you define your β value in all respects as far as this feedback this feedback is concerned.

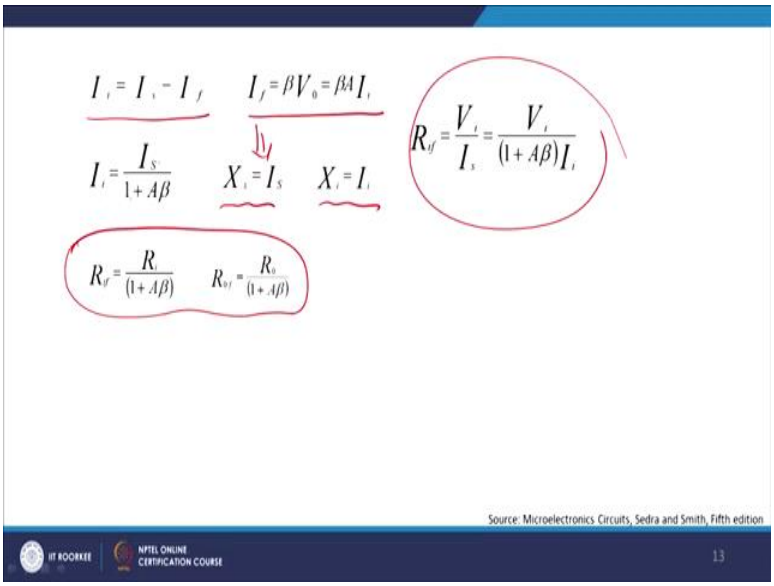
(Refer Slide Time: 20:19)




Now, let me come to the shunt shunt ideal case. The shunt shunt if you remember will be both sides parallel and therefore, you see this is $R_i$ which is the input resistance which you see and then you will have output resistance $R_0$ and this A into $A_i$ or A into $I_i$ is nothing but a current source. What do you say? A current source which is basically your modified current source as compared to an ordinary one and this is $R_i$ which you see and therefore, this $R_o$ is the output impedance you see here.

They are parallel with respect to each other because both sides you have got shunt and shunt, right? And therefore, I can write down $A_f$ equals to $A/(1+A\beta)$, right? And if you look if you go back to your ideal shunt cases I get $R_{if}$ I get $R_{of}$ and you have got $A_fI_s$ which is basically a controlled current source which gives me a value of output current for various values of $A_f$ chosen, right?

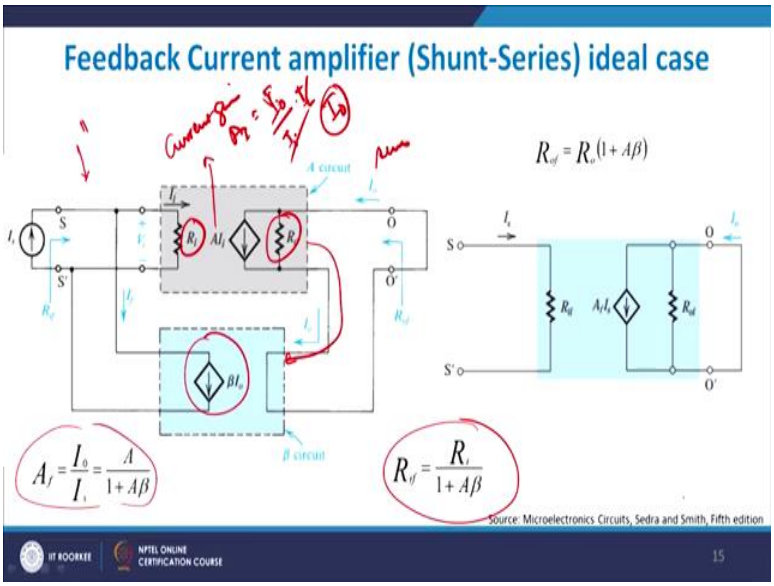
(Refer Slide Time: 21:14)




Now, therefore I get Ii the input current equals to $I_s$-$I_f$ because of negative feedback. We already if to be equals to $\beta * V_0$ and this is $V_0$ is nothing but $A*I_i$. I can safely write down if to be equals to $\beta * A*I_i$, right? So, I get $I_i$ equal to therefore equals to $I_s / (1+A\beta)$ and therefore I can safely write down input impedance or input source is basically $X_s$.

Output input current source is Xi and therefore, I can safely write down $R_{if}$ equal to $V_i/[(1+A\beta)*I_i]$ and $R_{if}$ equals to $R_i/(1+A\beta)$ and $R_{of}$ is equal to $R_o/(1+A\beta)$. Which primarily means that whenever I try to find out an impedance $(1+A\beta)$ term will surely come in the denominator. And that will make my impedance level go down with feedback. We can change the value of A and β and get the new values of $R_{if}$ and $R_{of}$, right? So, this is what we have learnt in this module. In general, we have learnt this basic concept here.

(Refer Slide Time: 22:31)

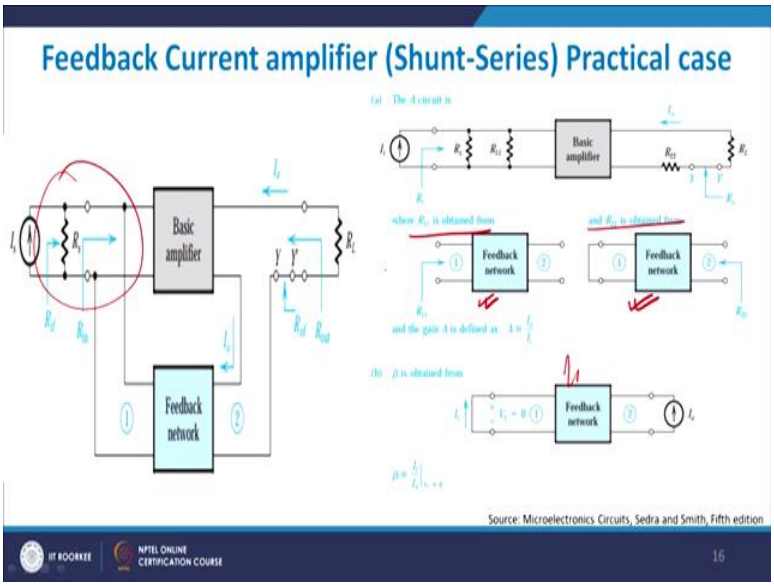



Now, let me come to the case of feedback sorry feedback resistances over here, feedback shunt resistance feedback shunt series resistance. So, I get a shunt series basically means that I will take a parallel output here. And I will do a series input at this particular point, right? And therefore, you see I have got $R_i$ here and then $A*I_i$. A is basically my current gain, right? And is given by $A_i$ and this will be current gain which is which is visible to me.

So, the total current would be nothing but $I_0 / I_i$ multiplied by $I_i$. $I_i$ gets cancelled and we get $A_i$ is equals to $I_0$, right? So, when you short and the resistance is low your current becomes large. And as a result, we also get a large current on the output side which is beyond our control on an extent. All the current is now being routed from the upper arm.

So what we do is we put an extra resistance here which is basically in parallel and then try to route this current through this arm onto a onto a device which is basically takes care of the increased value of voltages and therefore, increases the current flow through it. So, as I discussed with you as this is a shunt series feedback I will get my input impedance lowered by $R_i$ is equal to $R_o/(1+A\beta)$. However, my output impedance will also get lowered by $(1+A\beta)$. So both the input impedance and the output impedance falls by a factor of $(1+A\beta)$.

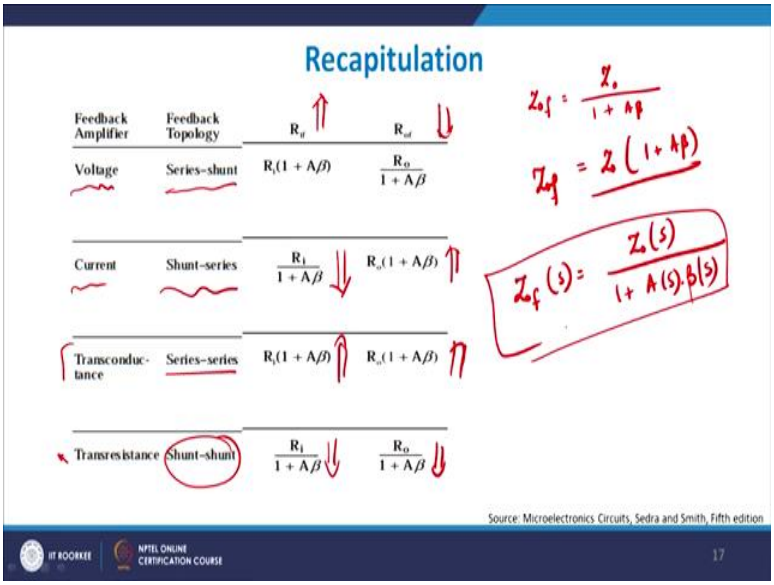
(Refer Slide Time: 24:11)




Let us look at the feedback current amplifier. It is basically a shunt series application so I have got a shunt here, right? And I have got the series application here. So, I get $R_s$ is parallel to $R_{11}$ and I get $R_{12}$ here in this place multiplied by $R_L$, right? And this is output impedance I have shorted it because any output impedance to confuse at this point of time.

Where $R_{11}$ is defined from this network from this network and $R_{12}$ is obtained from this feedback network, right? And from these two feedback networks they can actually find the value of the resistances or the impedances which is there in the system. β can be obtained from this feedback network in a general in a general manner. Okay!

(Refer Slide Time: 24:45)




To recapitulate, therefore let me show to you few recapitulation points that when you have a series series topology or a series shunt topology your feedback amplifier is generally a voltage. Whenever you have shunt series it is basically your current, whenever you have series series trans conductors and whenever you have got a shunt shunt you do have a you do have a 1 of the 1 of the scholars take care of your problem, right? That is basically the transresistance shunt shunt shunt shunt resis transresistance shunt shunt. So, in transresistance shunt shunt both the resistance fall down, right? In case of series both the resistance goes up input output.

In case of shunt series input falls down, output increases and in case of series series this rises and this falls off. So, depending upon whether you are having series or shunt the resistance offered will be always increasing or decreasing depending upon the virtual position of the human being, right? And that is what we learnt from this whole idea; that wherever you are placed in circuit tree depending upon whether it is series or shunt, you will be able to predict first principle from first principles whether you get a low resistance path there or high resistance path in that case.

Okay! So, I have got 4 types of network available with me series shunt, shunt series, series series and series shunt. Whenever you get series you get an increase in impedance, shunt you get a loss in the impedance. If you want to generalize the whole term I get $Z_{of}$. Z is the impedance which you see with feedback equals to Z just Z you can write down. $Z_o$ which is basically your output

impedance into 1 / (1+Aβ) when you want to show a decrease and this is how when you want to show an increase.

We have already discussed this point earlier. We also can show you that overall if you want to make it as a function of frequency, then $Z_{of}(s)$ equals to $Z_o(s)/[(1 + A(s)\beta(s)]$, right? And that is quite critical if because we have till now believed that the amplifier is basically frequency independent and basically we are talking about the mid frequency gain which your as will be independent of the frequency, right? And that is the reason this is a safe way to play out values of $Z_{of}$ and $Z_{if}$, right?

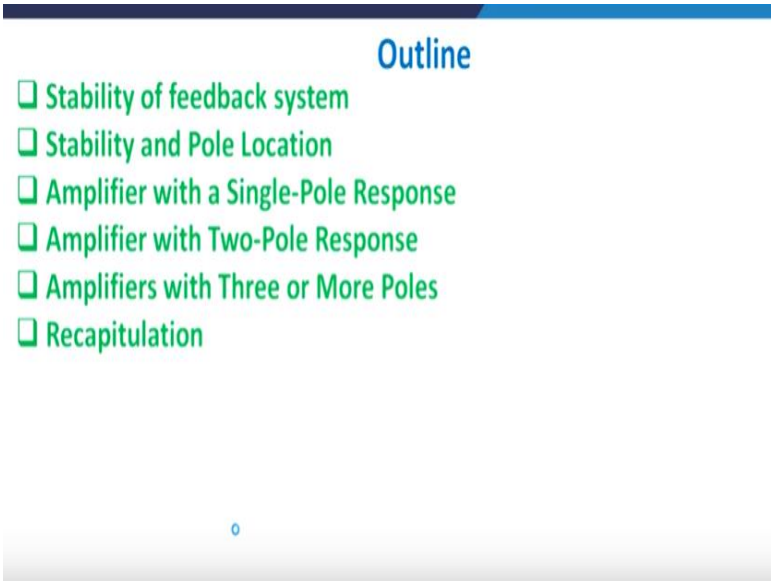
And you can actually get those values in a detailed manner here. And if you remember what we did therefore, was applying Kirchhoff's law first order Kirchhoff's law to get the values of the impedances here, right? With these words, let me thank you for your patient hearing and we will take up the next module of stability and pulls and zeroes applied to amplifier in our next module. Thank you very much!

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−48**
**Stability and Amplifier Poles**

Hello everybody and welcome to the this edition of NPTEL online certification course on Microelectronics Devices to Circuits. We will be now starting an important part of amplifier design and that is basically understanding the stability and the stability will be understood from the position of the poles in the S-plane right. So, we will first understand how do you will design S-plane and then there how do you, what is a pole and which we have already in our earlier section and then using that pole, or the location of that pole we will try to analyze whether this amplifier is stable or not.

What do I mean by stability? Stability primarily means that if you have a sustained oscillation in the output and the peak to peak of the output voltage is constant with respect to time then we define that to be sustained oscillation. If the voltage starts to fall with respect to time or rise with respect to time, we define that to be as an unstable element specially if the voltage rise with respect to time you enter into a loop which is never-ending and therefore, the overall gain, for example, if you have a positive feedback right, you will always go on adding voltage to the input and therefore the output voltage will always be going to be greater in time domain. So, with this basic background let me start today's topic. Let us see how it works?

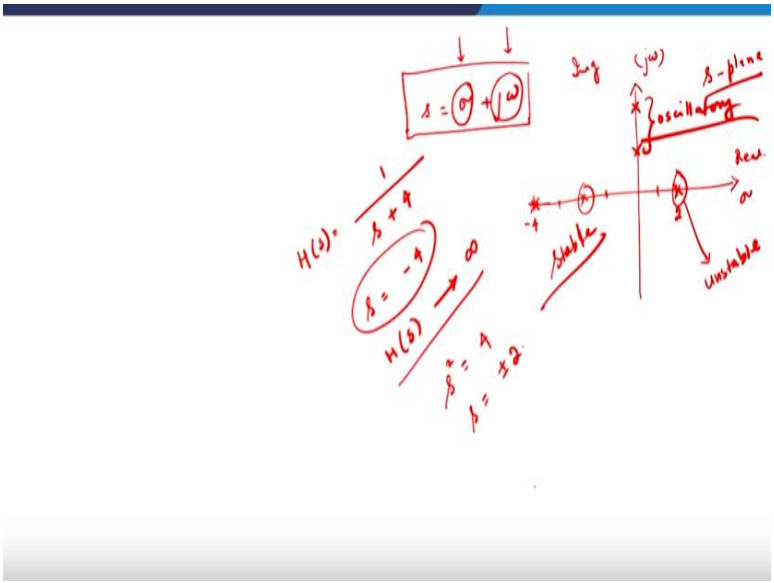
(Refer Slide Time: 2:03)




So, with this basic background let me start today's topic. Let us see how it works? So, we will be looking at what is known as the stability of the feedback system. So what are the criterias? We will study the criterias for stability in an amplifier design, right. We will also look into the position of stability and pole location. So, how do I relate the fact that where the poles are located and how stability can be asserted from the position of those poles in the S-plane?

Now, amplifier will look at the single-pole response, two-pole response and multiple pole response right. So, this single pole response, multiple or three-pole, single-pole primarily means that you have only single pole to the left half-plane. Two poles means there are two poles on the left half-plane, three or more poles is the number of poles three or more poles again on the in the left half-plane. So, I am assuming that all the poles are on the left half-plane of the axis. Now, let me explain to you where this concept comes into picture?

(Refer Slide Time: 3:03)




Generally, S is equal to right (σ + jω) right. So, this was general overall any complex quantity will consist of a real quantity plus imaginary quantity right. So, whenever you plot, this plot where this is a real part on the X-axis and this is the jω part and this basically the imaginary part, imaginary part and this is a real part here. Then we define this to be as a S-plane. We will see that if I remember in your previous discussion we were defining the transfer function as for example 1/(S+4) right.

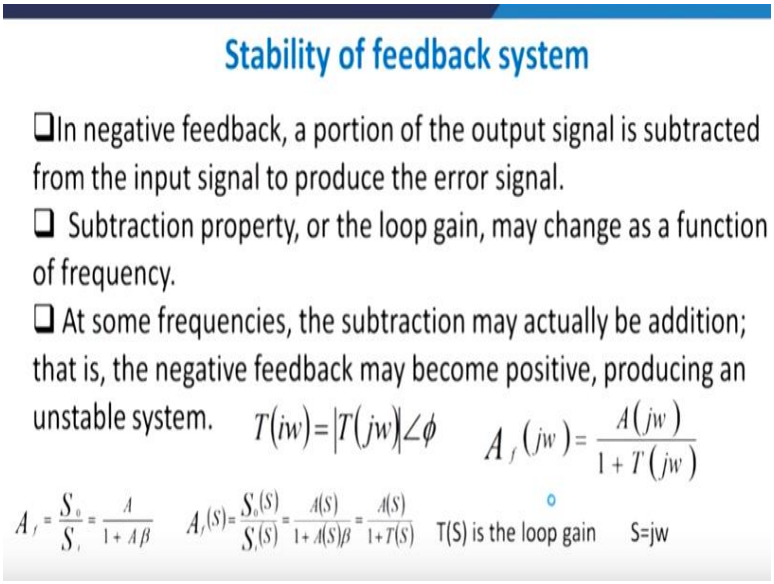
Now, if you see very carefully at s is equal to -4. The value of H(S) will actually go to infinity right and therefore I can one of the poles of this session is S is equal to minus 4. So, where is -4? This is -1, -2, -3, and -4. So, if you place it here I will get a very large output amplitude available to me.

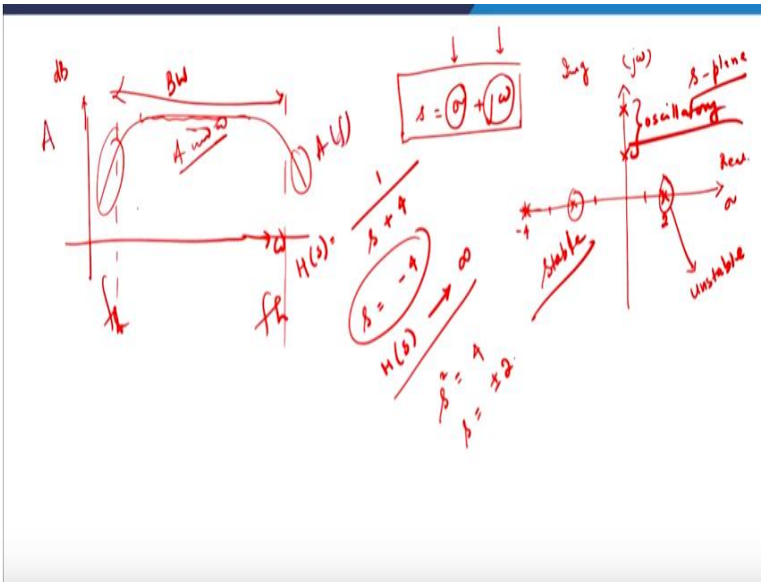
Similarly, let us suppose that S square is equal to some say again say some value say 4, then S will be equal to ±2 right. But, for +2 this is +2 here one pole here and one pole on this side and -2 one pole on this side. This will be stable but this will be unstable. So, on the right-hand side, this will be unstable on right side and everything on left-hand side will be stable in nature.

We will come on this in detail later but just for information sake at this stage whenever you refer to S-plane it is basically σ versus jω, σ is the real part of S and jω is basically the imaginary part and when you place it in this manner, you get this. If the poles are available onto your jω axis

this will let you what is known as the oscillatory wave. So, I will have an oscillatory wavefront here sustainable oscillations are available here.

(Refer Slide Time: 5:08)




So, with this basic knowledge of S-plane, let me explain you the feedback like feedback. Now, with respect to the feedback system. We have already seen in our previous discussions that the negative feedback primarily means that the output signal is subtracted from the input signal right to produce an error signal. So, not all output signal is subtracted. A part of the output signal is subtracted from the input signal and an error is taken care of. This error is then added up with input or this error signal which is available to you is fed into the amplifier. The amplifier again

amplifier is a signal and this process is iterative process through the feedback loop. We have already studied that portion earlier.

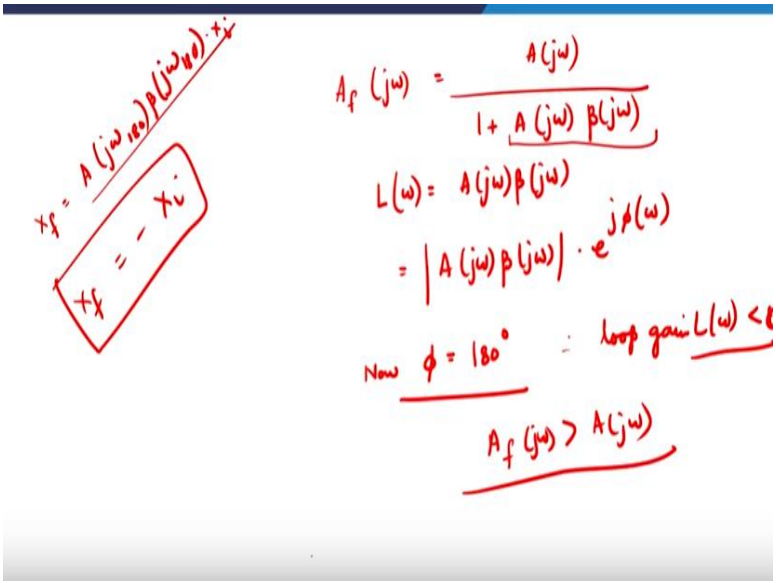
Now, as I discussed with you that if you remember how I define my loop gain. Loop gain was basically equal to Aβ and this Aβ will be the function of frequency, right. Assuming initially that they are not, but in reality they might be a function of frequency right.

As we have discussed in our earlier terms that if you plot A verses gain normalized gain in dB versus frequency let us say ω then typical values which you get is something like this. So, this is a mid-frequency gain is almost constant from this point to this point and this is a bandwidth actually and this is $f_L$ and this is your $f_H$ right. So, this part is the part where A is independent of ω and these are the portions where the gain this is your gain the gain is a function of frequency right.

So, with this knowledge, I can safely say that for a large change in the frequency you might have the change in the value of your amplification factor. At some frequencies the subtraction may actually be the addition, the negative feedback may actually be look as positive feedback producing an unstable system. I will explain it to you how I am talking to it.

See, as I discussed with you T is transfer function will consist of its mode value and its phase right and if you have a generic diagram of $A_f$ with feedback is equal to A, without feedback is equal to 1 plus transfer function of feedback then in our case as in our previous discussion $A_f$ is equal to A/(1+Aβ) then you see that af with respect to β comes out to be in S terms comes out to be this where ts is the loop gain so A β is the loop gain which you see right.

(Refer Slide Time: 7:52)

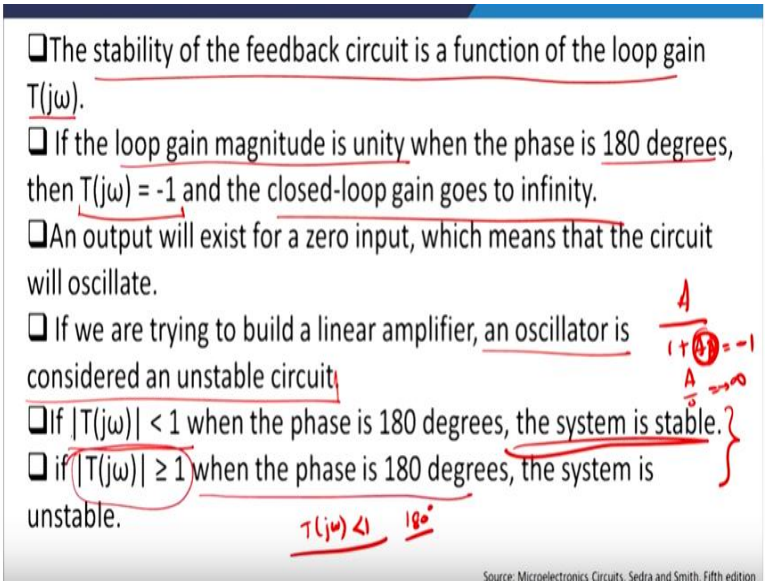



Now, if you look carefully again get back to the whole discussion. What we can write down from all this discussion is that, that $A_f(j\omega)$will be equal to $A(j\omega)$ right divided by 1 plus $A(j\omega)$ into $\beta(j\omega)$ right. So the loop gain $L(\omega)$ is equal to $A(j\omega)$ into $\beta(j\omega)$ right. Now, this can also be written as if you want to find out its in the form which is most understandable to most of us that we find out its mode value so the magnitude multiplied by $e^{j\phi(\omega)}$ right where $\phi(\omega)$ is the phase which you get.

Now, let us suppose $\phi$ happens to be equal to 180 degree then my loop gain loop gain which is basically my $L(\omega)$ will actually be negative and therefore, if it is negative so this whole quantity becomes negative this whole quantity and therefore $A_f(j\omega)$ is greater than $A(j\omega)$ right and that is the problem area that generally you do have your with feedback gains are increases as compared to without feedback. So $X_f$ in general sense. I can write $A(j\omega_{180})$ right into $\beta$ $(j\omega_{180})$ into $X_i$. This will be equal to $-X_i$. So, $X_f$ is equal to $-X_i$. So where A and $\beta$ are the basically the feed-forward feedback factor at $\omega$ equals to 180 degrees.

So, what I am trying to tell you is that when the phase angle is equal to 180 degrees. Your negative feedback actually converts into a positive feedback. So, till 180 you will have a negative feedback and your system will be relatively stable as you cross 180 degrees, your system becomes unstable. That is what I wanted to prove from the statement as far as this is concerned.

(Refer Slide Time: 9:58)



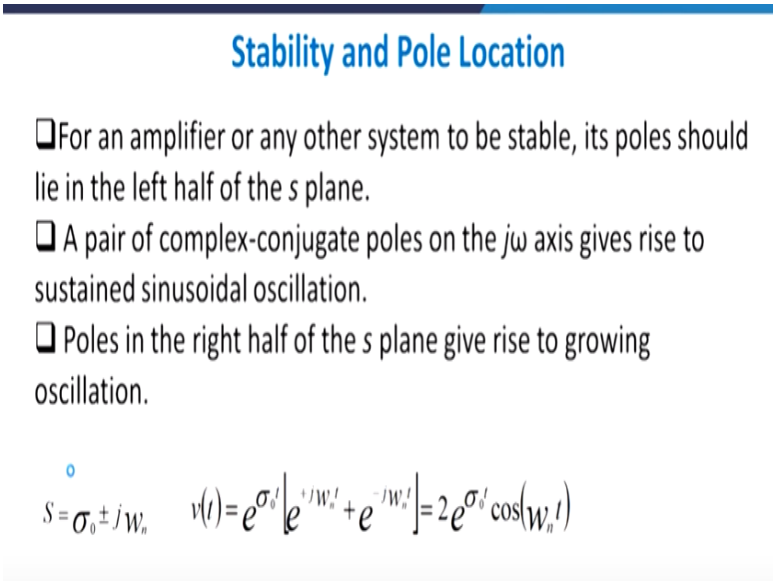

So, therefore as I discussed with you the stability of the feedback system in the circuit is a function of the loop gain T(jω) right. Now, if the loop gain magnitude is unity when the phase is equal to 180 degrees then as I discussed with you T(jω) will be equal to -1 and the closed-loop gain goes to infinity, why? Because if you remember it was A/(1+Aβ). So, now if A β is happens exactly to be equal to -1 then A upon 0 is basically tending to be equal to infinity to a rather extent.

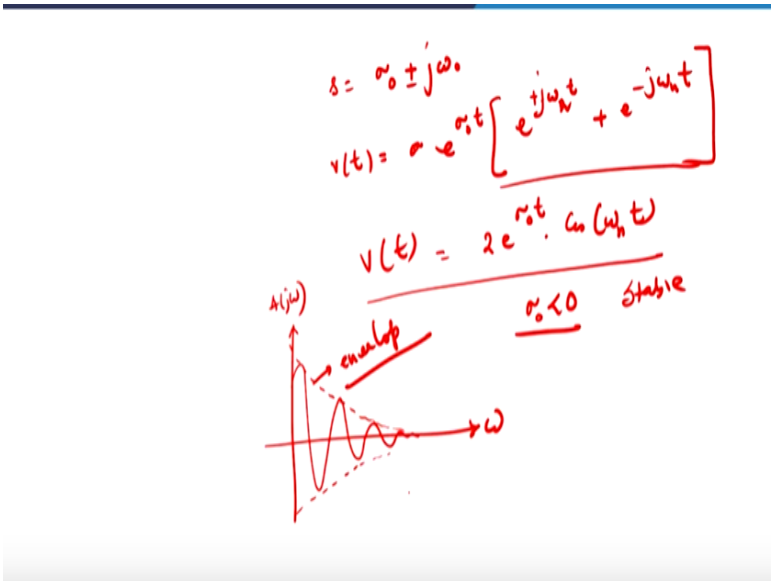
Idea is how do you define therefore oscillation? Oscillation is defined therefore as even with the small input on to the input side of the system or the oscillator I should be able to get an external oscillations available to be right that is what an output will exist as 0 input right. Why this 0 input? 0 input means you do not have any input right given as a signal state, but you might have some small noise signal, some thermal signal is available to you which tries to make it available it to circuit state right.

Now, if you try to have a linear amplifier obviously an oscillator is considered to be an unstable circuit. I hope you understand why is it like that. Linear circuit primarily means that with respect to frequency there will be a linear rise in the gain let us suppose. But that is not true when you have oscillation gain into consideration right. Your oscillator, sustained oscillation will give you a gain independent of frequency, mid-frequency gain and therefore linear amplifier designed with an oscillator is a difficult task almost impossible task.

Now, therefore to sum up all these discussions these two points are very very important that whenever my T(jω) is less than 1 so I get a negative quantity and when the phase is 180 degrees right, then the system is stable right so whenever it is less than 1, A/(1+Aβ) will be A will be therefore less than A and the system will be stable as when the phase is 180 degree at that point of time if your T(jω) is transfer function is greater or equals to 1 then in the denominator you get T(jω) to be negative and as a result automatically you will get 1 minus small quantity will give you a small output in the denominator and therefore you will get the system to be unstable in nature right because it will be rising wave function as such for your case right. So, this is what you get as far as switching is concerned.

(Refer Slide Time: 12:22)






Now, let me come to the concepts of poles and zeroes. As I discussed with you just in the starting of this module that for an amplifier to be stable or any system to be stable its poles should lie on the left half of the s plane. So, as a rule of thumb across the board and we will see just now a numerical reasoning for that that if you have s plane within the S plane you want to generate a stable system then try to ensure that the poles are at actually situated on the left half side of the plane right. That is pretty important where your σ is less than 0. That is very very important where your σ is less than 0.

As I discussed with you earlier also a pair of complex conjugate poles on the jω axis gives you a sustained oscillations. It will rather extend right. Now, poles on the right-hand side will always give you an oscillations which are growing with respect to time. So, it will be a primarily unstable design for for any oscillator or for the amplifier. The reason being on the right-hand side you will always get the output will be always gaining with respect to time and therefore it will be a never-ending process as far as this is concerned.

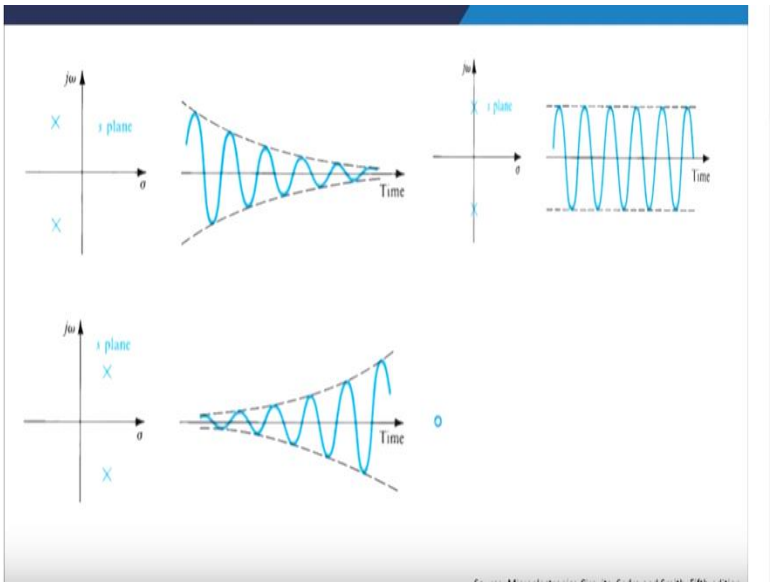
Now, let me let me come to the explanation here that if you see this signal s is equal to $\sigma_0 \pm j\omega_n$ right and therefore if this is the value of s which you see then we can write down voltage.

I just write down for you that let me say let me say is given S equal to $\sigma_0 \pm j\omega_0$ and therefore V(t) voltage is given as $e^{\sigma_0 t}[e^{+j\omega_n t} + e^{-j\omega_n t}]$. So this is the sign of $\omega_n t$ multiplied by $e^{\phi(t)}$ gives you these value. Now if you solve this I get $2*e^{\sigma 0 t} *\cos(\omega_n t)$. This will be your V(t).

Now, if the poles are situated to the left half-plane which means the $\sigma_0$ is less than 0 right then we get a stable system otherwise we will get an unstable system. Now, if it is less than 0 you will actually see something like this that the output will be like we will go on decreasing as you move with respect to time. So, if you plot a graph here which is the envelope of the whole design you will see that the envelope is actually a falling envelop. So, this is your gain in $j\omega$ versus $\omega$ right and this gives you the decay right.

So, this basically is your envelop, envelop of the cell and you can see where $\sigma_0$ is less than 0 so you will get e to the power $\sigma_0$ coming into picture which means $\sigma_0$ is less than 0 means $\sigma$ will be negative e to the power negative will actually be a falling function and therefore you will automatically get a smaller and smaller function in general right. So, this is the basic concept which you see here. As I discussed with you $2*e^{\sigma 0t} *\cos(\omega_n t)$.

(Refer slide time: 15:50)

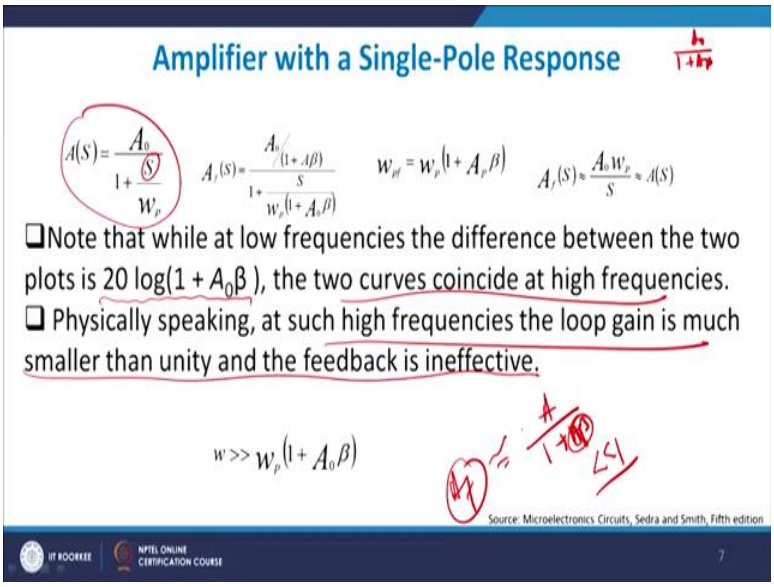


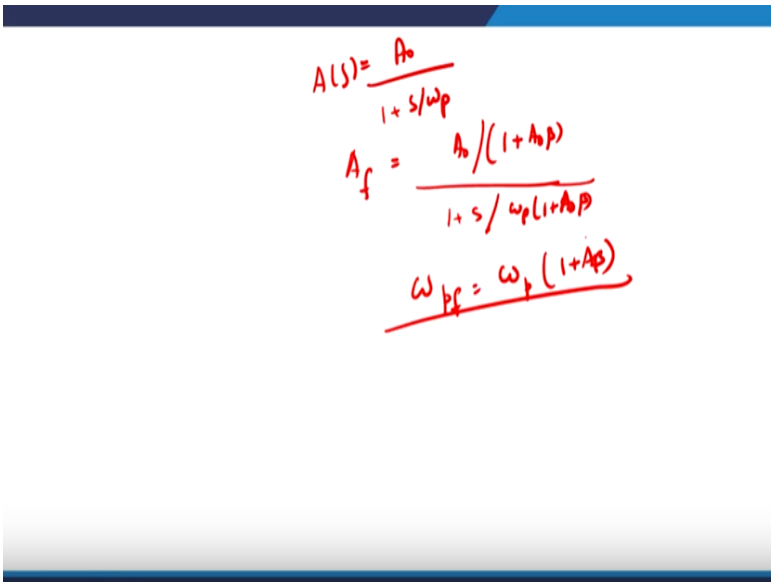

Now therefore, if the two therefore let us suppose you have got two poles and they are on the left half-plane and they are complex conjugate. Conjugate means one is a real part and another is an imaginary part and therefore they are referred to as complex conjugate. So, if you have such types of planes you will automatically get a function which is e to the power $(j\omega_n t)$ and then there will be term $\beta$ here. If $\beta$ is negative this function will always be a falling function as such and this is falling now right and therefore the envelope is also falling down which means that you are restricting the output to a very very low peak to peak voltage.

Similarly if is in the S plane the design is such that that the S falls to the right for example 1 upon S mimus 4 then I get equals to hs then I get s equal to plus 4 for the poles so which means that it falls on the right-hand side of the plane and therefore it becomes sort of a rising function which you see the envelope is rising envelop this is a falling envelop of the output side right. If they are exactly on the plane as you can see on this graph then it will always give you a sustained oscillations to a larger extent.

So, the job of a designer when designing an oscillator primarily is oscillator or amplifier is that you try to keep your poles exactly on the jω axis that makes a sustainable sustainable growth in terms of peak to peak voltage. So, that is a basic idea that is all about and then we come to the fact that let us suppose that we do have a single-pole response which means that I do have a single design or a single-pole and the single pole will have on the left-hand side.

(Refer slide time: 17:54)



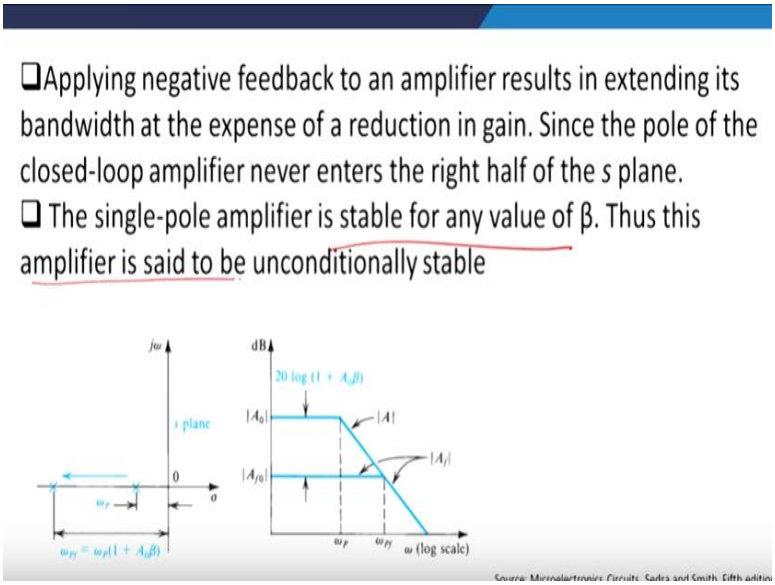



Then we can safely write down $A_0$ so we can safely write down for a single-pole system that A(S) equals to $A_0$ upon 1 plus S upon $\omega_p$ right single pole you will also get only single S function here. A of s will be feedback $A_0$ upon 1 plus Aβ so you remember it is basically $A_0$ upon 1 plus A β so I am replacing this A by A plus 1 β so I get A plus 1 by A β and I get $A_0$ here so what I get from here is that my with gain the feedback factor increases.

So, I get As is equal to $A_0$ divided by 1 plus s by ω p right and therefore I get A with feedback to be equals to $A_0$ divided by 1 plus $A_0$ into β right divided by 1 plus S divided by $\omega_p$ 1 plus $A_0$ times β fine so I get therefore from here if you solve it I get $\omega_{pf}$ with feedback $\omega_p$ 1 plus A β. So, you see your gain frequency starts to rise with feedback and that is what you see here. Note that while at a low frequency is a difference between the two plots is 20 log(1+ $A_0$ β), the two curves coincide at high frequencies right.

Physically speaking at such high frequencies the loop gain is much smaller than unity and the feedback factor is ineffective right. So, if you remember it was A/(1+Aβ) at such high frequencies this value of loop gain which is A β is much smaller to as compare to 1. Therefore, your af which is a feedback factor almost equals to A and it remains A for quite duration of time and therefore feedback is quite ineffective right that is what we have learned through this discussion.

(Refer slide time: 20:18)

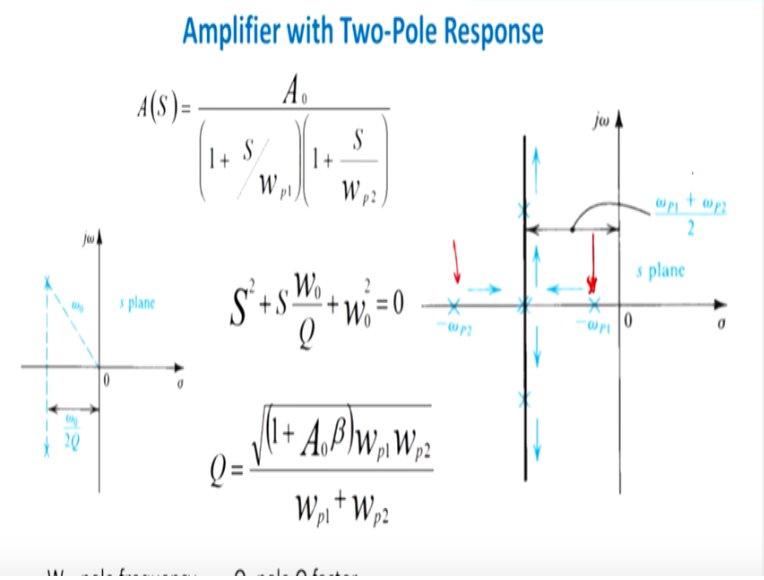



Now, therefore applying a negative feedback as I discussed in your previous turn you are able to increase the bandwidth of the amplifier at the cost of its gain right. Since the pole of the closed-loop of the amplifier never enters the right half of the s plane, you automatically get a dominant much more stable system with you right.

As I discussed with you a single-pole amplifier is stable for any value of β. Thus, this amplifier is said to be unconditionally stable right pretty important statement that this is unconditionally stable which means that if you do not have to put in extra efforts to get its stable unconditionally stable. As you can see in my diagram here on the left half-plane you have got $\omega_{pf}$ right and this is $\omega_p$ which is basically your left half pole right and this pole is basically on the left half-plane therefore it is stable but by some mechanism or other I am able to shift this by doing negative feedback actually I am able to shift this pole from this point to this point so the access value that you get is basically this much and that is the reason I add $\omega_p$ plus $\omega_p$ into $A_0\beta$. so, this extra zone which you get is not coming here and you get unstable.

On the right-hand side if you look very carefully it is actually a plot of gain versus frequency and as you can see as the gain falls down its 3dB bandwidth it is basically the higher cut of the frequency which is this one is actually falling down right. So, gain in bandwidth product again is constant. This is the standard rule of thumb which we apply in reality.

(Refer Slide Time: 21:26)

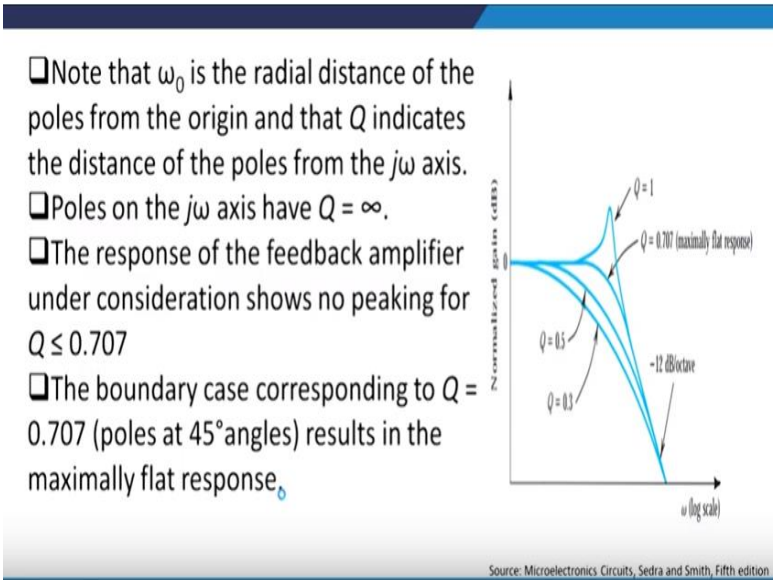


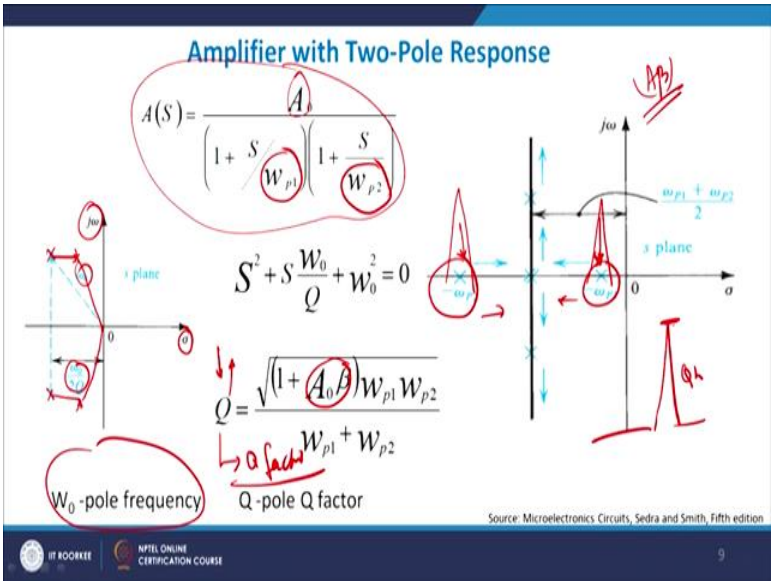

Now, let us look at a two-pole response or a two-pole design. I will explain to you how a two-pole design works typically in this scenario. But before you, ok so as I discussed with you that phase should never go beyond 90 degrees right. It should be between 0 and 180, but if you want to find out a positive feedback it should not be above 90 degrees right.

Let us look at the amplifier with the two-pole response. So two-pole response primarily means that I have one pole here $\omega_{p1}$ and I have got another pole here $\omega_{p2}$ right and then I start applying frequency. So I do a characteristics equation. This is very simple and straight forward. It is a low-frequency gain or low-frequency parameter which I am fixing up. My my now what is happening is that I am varying my $\omega_{p1}$ and $\omega_{p2}$ right and we check out how these work out.

So, if you want to if you want to solve it and gets its pole the denominator has to be standardized in the form of $a^2+2ab+b^2$ and they have done that using this formula and from here I can get the value of s to be some value as well as two values of plus and minus sign. Similarly, Q will have Q is basically known as pole a Q-factor or the Q factor here right Q factor. It is typically height by the full-width half maxima that is what you define as so if you have a sharp profile then you have a high Q because the selectivity is very very high. Not only that its height is also very large. So this height by this height happens to be relatively small quantity, but still because of the fact that this gives you an idea about how large is your function which you are dealing with. It is quite an important formula which you see in front of you here. So, I got Q-factor available with me.

(Refer slide time: 23:24)




Now, note that $\omega_0$ is the radial distance of poles from the origin and Q indicates distance from the poles from the jω axis. So, it is quite interesting that in the previous slide if you look very carefully then you will see that for two-pole response as I discussed with you you have one at $\omega_{p1}$ with a negative sign, you have another at minus $\omega_{p2}$ on to the negative side here right on the negative side.

Now, so if as you increase the value of the A β gain into loop gain goes on increasing these two poles start to come closer to each other right. How is it possible? You see it is possible because of this reason. $A_0$ β will go on increasing and therefore this quantity will go on increasing, as a

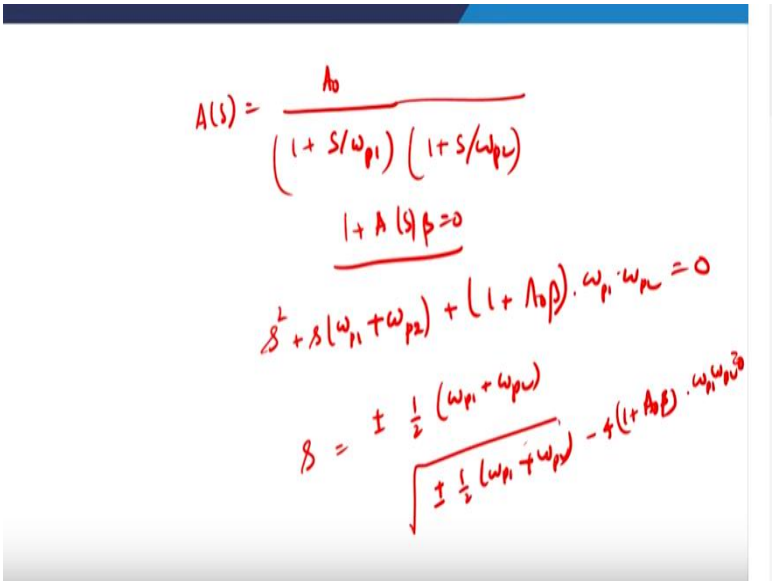
result, your peaks will be sharp. So, corresponding to this you will get one peak here, corresponding to this you will get one peak here. Bring it closer and you will start getting single peak of a large dimension within the typical value of σ right and that is very important issue which people are looking into it.

$\omega_0$ is referred to as the pole frequency of the design right $\omega_0$ is defined as the pole frequency right and this pole frequency gives you the value of your this thing. So in a jω versus σ plane if you want to find out then we define this to be as the σ plane and this to be jω axis, then if my pole is complex conjugate and lying on the left half-plane we try to first plot them and then check the value of $\omega_0$ and $\omega_0/2Q$. Which is this is distance right.

Now, if you bring closer here if you bring closer here then you will see that let us suppose you want here then this and you will have this right and as a result what will happen is that you are bringing the poles closer to the jω axis so you are allowing it to go for a sustained oscillations right but that will be only possible provided you are able to enhance the values of your or reduce the values of your Q. Then only this will be possible.

Now, let me come to a two-pole system as I discussed with you Q indicates the distance of the poles from the jω axis. The poles on the jω axis have Q equal to infinity and that makes sense also. So at, if you are on jω axis you put a pole, right put a pole here on the jω axis obviously Q has to be infinity because in this case your S is equal to 0 or σ equals to 0 right. So, the real function is gone. Only what you will have is an oscillatory function that is because of the complex conjugate pair of pair of output signal.

(Refer slide time: 26:11)

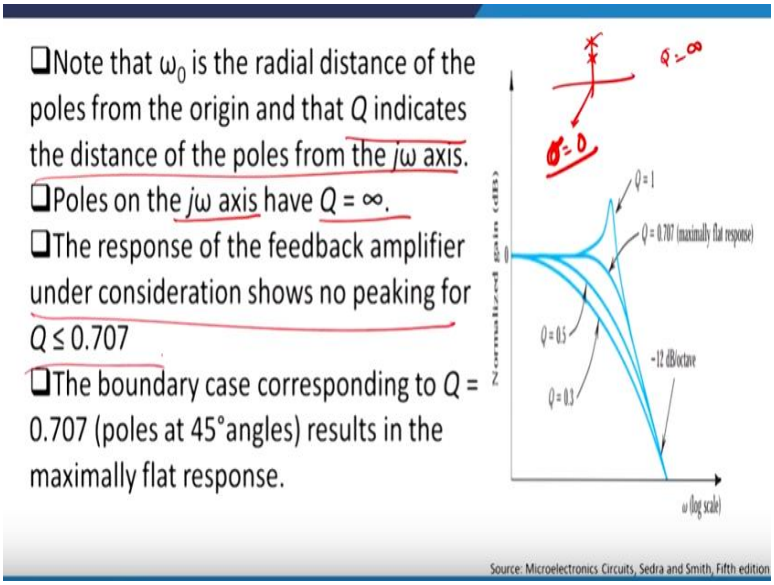



So, what you do is I will give you just a thought and we can move forward that if you have a multiple let us suppose that we have a multiple system then the two-pole response again let me show $A_s$ is given as $A_0$ upon 1 plus S by $\omega_{p1}$ right into 1 plus S upon $\omega_{p2}$. A generalized statement of a two-pole network. Then if I assume that this is the time when I will get the overall picture then again I write down as ssquare plus S into $\omega_{p1}$ plus $\omega_{p2}$ plus 1 plus $A_0\beta$ into $\omega_{p1}$ into $\omega_{p2}$. This must be equal to 0. If it is 0 then you automatically you get this function to be high.

So I get S to be equal to plus minus 1 by 2 right $\omega_{p1}$ plus $\omega_{p2}$ right. This multiplied by multiplied by obviously plus minus 1 by 2 $\omega_{p1}$ plus $\omega_{p2}$ right $\omega_{p2}$ and then minus 4 times 1 plus $A_0$ times $\beta$ multiplied by $\omega_{p1}$ and $\omega_{p2}$.

So the same formula which I have been using time and again minus B by A and so on so forth and then you get s to be equals to this quantity here right. So as and as as as therefore as $A_0$ into $\beta$ goes on increasing right so what happens is that this quantity goes on decreasing sorry, yes this quantity goes on increasing and therefore this quantity goes on decreasing when this quantity goes on decreasing this quantity goes on increasing. So, I will the poles will start moving to the left of the plane right-left of the plane here right.

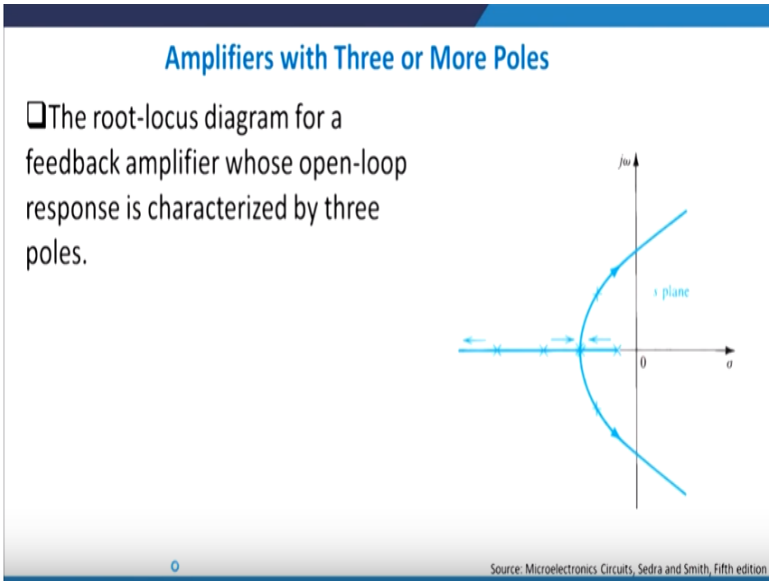
(Refer slide time: 28:05)




So the response of the feedback amplifier shows no peaking at Q equal to less than 0.707 right and that is the reason you will get a stable system at Q equals to 0.707. That is also referred to as maximally flat response right. If Q is greater than that you might get a peak here which is not a very good sign. If Q is less than this value the 0.707 and 0.03 my mid-frequency gain will be very very low and I will not be able to have a stabilized values of frequencies where they will fit into each other in a proper fashion.

Now, boundary case value is a value is equal to 0.707 which corresponds to a phase margin of 45 degrees. So, 45 degree is the phase margin which you see as there the poles location will be at 45 degrees also referred to as pole position right.

(Refer slide time: 29:00)

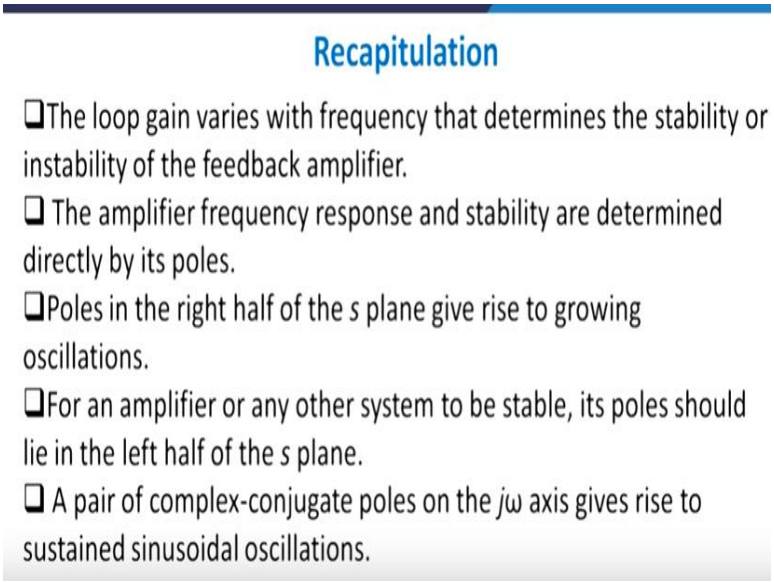



Now, amplifier with three poles is that just explained to you this thing and then we can move forward. What we say is that the last there are three or last one as we move away from most away from the origin axis starts moving to the left away from the pole as the value of the frequency goes on increasing. However, however, the other two poles starts to move closer to each other right. A time will come when they will actually override with respect to each other and then start moving to the left and to the right up and down.

So, till this movement, you will only have a real real frame movement. So σ movement as available to you, but be on this particular point you do have complex conjugate movement of these two poles as you move towards the jω axis. So, that gives you a proper idea about the three-loop system or the three-loop system in a sense is there. It is much more complex conjugate and makes it unstable as such.

So with this let me show to you also as I discussed with you in the previous this thing that if you go back so if you back to your previous discussion then you will see that you require such high end maximally flat amplifiers where your gain is almost independent of frequency.

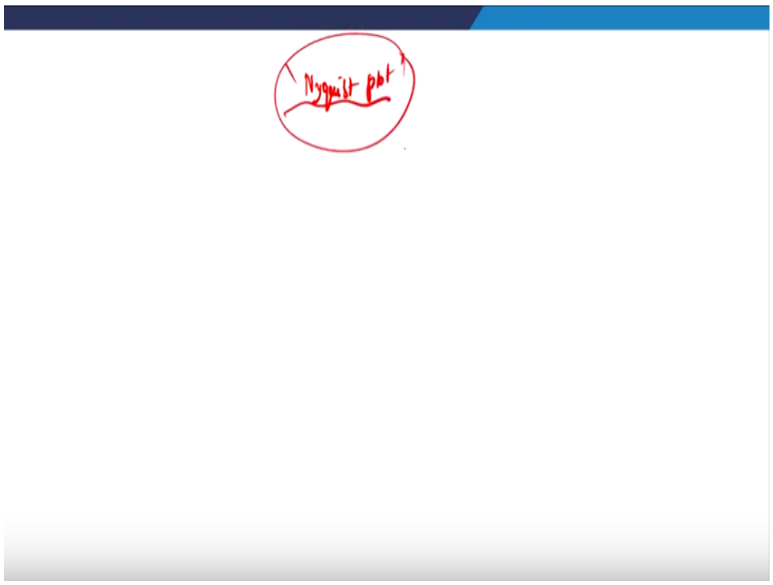
(Refer slide time: 30:31)




So, let me recapitulate what we studied in this module for stability the loop gain has to be less than greater than 1 and if it is equal to 1 or the phase margin equals to 180 degrees I enter into the region of unstability of (oscillation) amplifier. The amplifier frequency responds and the stability is determined by the poles. So, typically the poles of left half-plane it is much more stable. Poles in the right half-plane gives rise to growing oscillations right and that is quite difficult to achieve. If not difficult it is not desirable to achieve in common physical systems.

For an amplifier or any other system to be stable, its poles should be in the left half of the s plane, we have seen that. A pair of complex conjugate poles on the jω axis gives rise to sinusoidal oscillations in reality right. So, you have will always have sinusoidal oscillations if the two poles are located on the jω axis in a sense right and that gives you an idea about why we are requiring such a such a concept here.

(Refer Slide Time: 31:37)

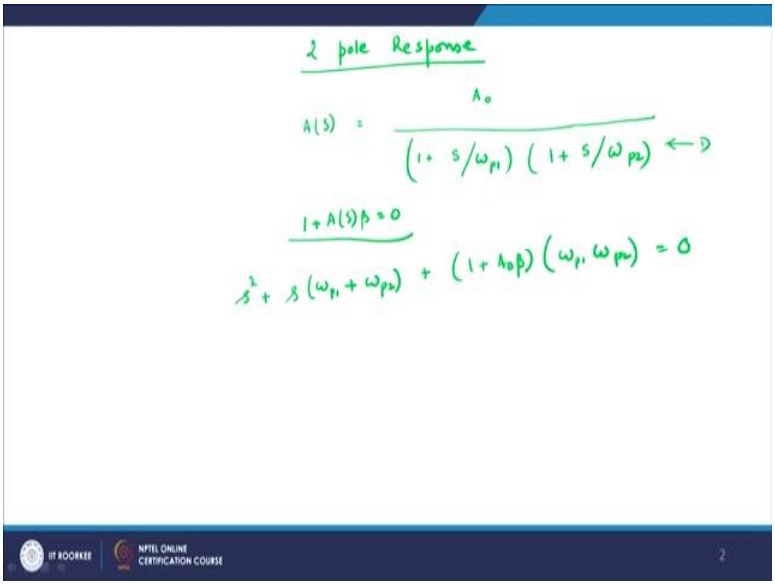



Now, I will just stop this discussion with one small enhancement and that is known as Nyquist plot and I will discuss this Nyquist plot may be in the next turn giving you an idea what is Nyquist plot right. We will discuss this Nyguist plot in the next module. Thank you for your patience hearing.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−49**
**Bode Plots and Frequency Plot**

Hello everybody and welcome again to this NPTEL knowledge certification course and on Microelectronic Devices to Circuits. In our previous module we had looked into the stability criteria from the point of view of poles and zeroes, poles specially, and we saw that the poles should be on the left half plane of the system and the phase margin as I discussed with you in the previous turn, should not exceed 180 degree for the proper stability. And we will discuss that in detail as we move along. So we will actually look into the today's discussion is on bode plots and frequency responses.

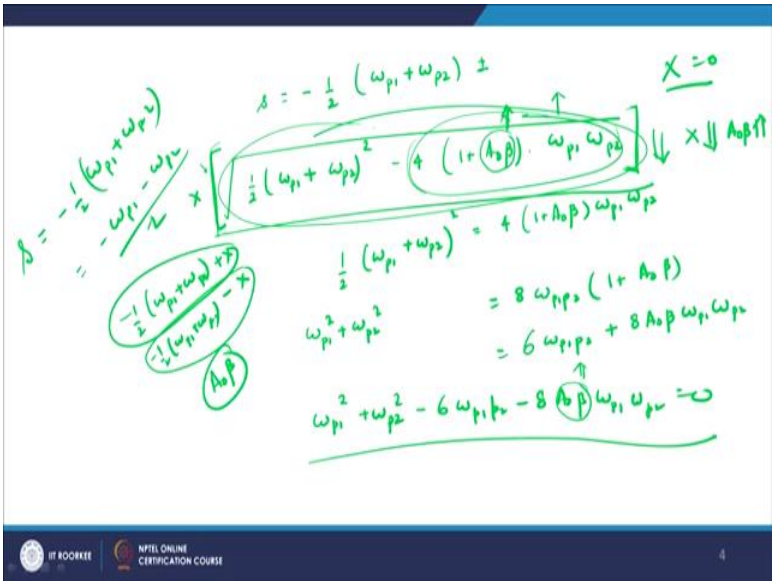
(Refer Slide Time: 01:14)




So before I move forward, let me just discuss in detail slightly about the two pole networks. Two pole network primarily means that the 2 pole response actually which we did in our previous turn also, but I just wanted to just to refurbish again the whole thing, in 2 pole response the general your response transfer function looks like this S upon $\omega_{p1}$. Right? And this is 1 plus S upon $\omega_{p2}$.

With this if you solve it, I get 1 plus A(S) times β equals to 0 is the condition when you get a sustained oscillations under such a criteria if you solve if you breakdown this denominator. Right? So the denominator broken up into S square plus S times $\omega_{p1}$ plus $\omega_{p2}$ plus 1 plus $A_0$ time β. Right? Into $\omega_{p1}$ into $\omega_{p2}$ this must be equals to 0.

(Refer Slide Time: 02:21)




Which effectively means that s will be equal to plus minus. Right? Or sorry s will be equal to minus half $\omega_{p1}$ plus $\omega_{p2}$. Right? And then plus minus taking sign of root over 1 by two $\omega_{p1}$ plus $\omega_{p2}$ whole square minus four times 1 plus $A_0 \beta$ into $\omega_{p1}$ into $\omega_{p2}$. So this are the two roots of the equations, one root is with the positive sign another root will be with respect to the negative sign.

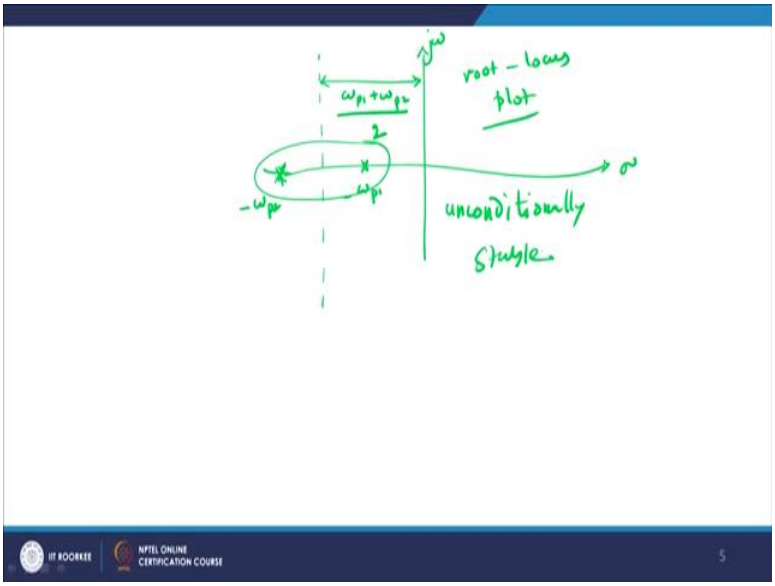
If you look carefully the whole quantity within this root this under root sign can be made to 0 then you only you will naughts have $A_0$ or $\beta$ coming in the picture, in that scenario I get 1 by 2 $\omega_{p1}$ plus $\omega_{p2}$. Right? Whole Square if that equals to 4 times 1 plus $A_0$ times $\beta$ times $\omega_{p1}$ $\omega_{p2}$ that makes these two sides equal and therefore thus the sustain this to be whole this let us suppose this is X then X will be equal to 0.

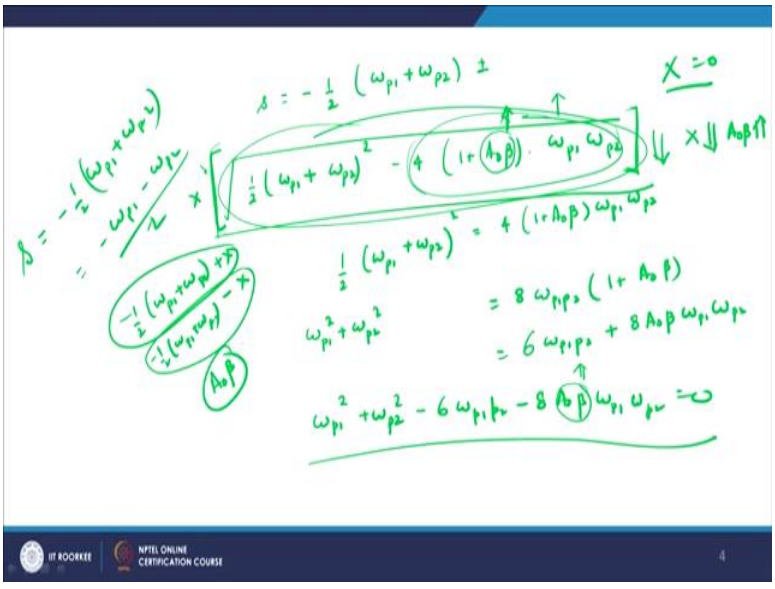
If you solve it I get $\omega_{p1}$ square plus $\omega_{p2}$ square plus 2 $\omega_{p1}$ $\omega_{p2}$. Right? Will be equals to 8 times $\omega_{p1}$ $\omega_{p2}$ into 1 plus $A_0$ time $\beta$. So if you solve it I get 8 times $\omega_{p1}\omega_{p2}$ plus 8 times $A_0$ times $\beta$ times $\omega_{p1}$ $\omega_{p2}$. Right? So with this so if you look if you see carefully the two sides of the equation then this will cancel with this and make it 6 so this will this will go and this will become 6 times so this will go and therefore what I will get from here is this will becomes and I can vanish this. Right? And I can get I can get 6 $\omega_{p1}$ $\omega_{p2}$ plus 8 times this thing into $\omega_{p1}$ $\omega_{p2}$. Right? And this what I get overall picture for this whole design.

So or in other sense I can wright down this to be as $\omega_{p1}$ square plus $\omega_{p2}$ square minus 6 $\omega_{p1}$ $\omega_{p2}$ minus 8 times $A_0$ $\beta$ $\omega_{p1}$ $\omega_{p2}$ must be equal to 0. When you get this into the consideration I

get S equal to minus half of $\omega_{p1}$ plus $\omega_{p2}$ which is equal to minus $\omega_{p1}$ minus $\omega_{p2}$ by 2. So which is therefore the left half can plane so it is stable.

(Refer Slide Time: 05:25)




So with this condition let me plot draw the plot for you and show to you, suppose this is my this is my jω axis and this is my σ axis and I have got here this is this is $\omega_{p1}$ and this is your minus $\omega_{p2}$ and this is your point which is the half line point between the two, and this is actually your $\omega_{p1}$ plus $\omega_{p2}$ by 2. Right? So the distance of the half line from this is what you get it.

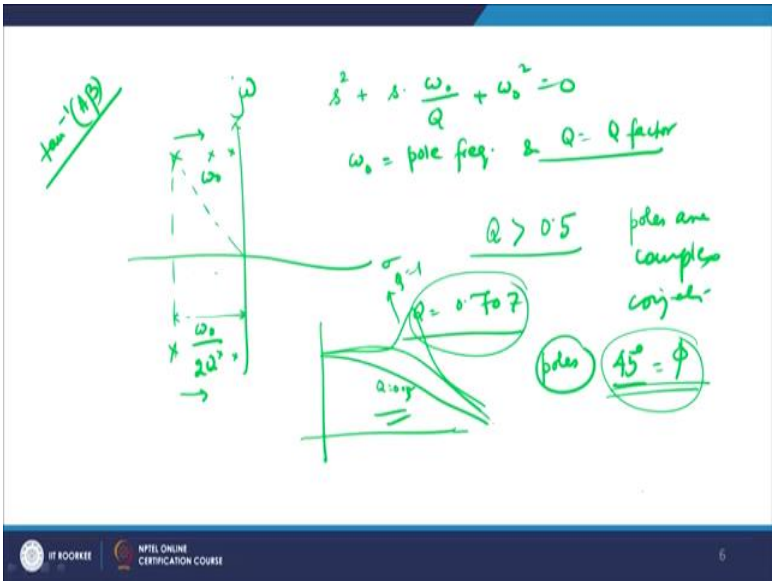
Now an interesting idea if you look very carefully from all this discussion is as you start increasing your frequencies which means that or if you go on increasing gain $A_0$ times β if you go on increasing it, if you let us look at this characteristics and if you go on increasing $A_0$

times β then this quantity goes on increasing and therefore this quantity goes on decreasing. Right? So I get suppose so I get 1 by 2 $\omega_{p1}$ plus $\omega_{p2}$. Right? I get plus $X_1$ once and then I get minus $X_1$ once. Right?

Now, when this $A_0$ times β goes on increasing then you will see that this X quantity because this increases means this whole quantity will increase. Right? Which means the whole quantity will actually go on decreasing. So X quantity so X will decrease when $A_0$ times β goes on increasing. Right? So when $A_0$ times β increases X increases so this quantity will go on increasing whereas this quantity will go on decreasing because this is a minus sign here.

And as a result what will happen is that with increase in $A_0$ times β. Right? These two location of these two poles on this two occasions they will start shifting away from each other and therefore that is what happens when they shift away from each other. This is also known as a root locus plot in the X axis and this is the condition so this whenever the two on left half plane we defines this to be an unconditionally stable condition. Stable condition. Right? So this is for 2 pole response which we get in reality. Right?

(Refer Slide Time: 07:59)




Now, let me explain you the Q factor because that was left out in the previous turn. So let us suppose the characteristic equation in again a 2 pole network is given as $\omega_0$ by Q plus $\omega_0$ square is equal to 0. So $\omega_0$ is defines as the pole frequency and Q is defined as the Q factor. Q factor or quality factor. Right? Typical rule of thumb is that if Q is greater than 0.5. Right? The poles are a complex conjugate. Right?

And they will be lying it will be something like this so it will be lying suppose this is jω, this is σ they will be lying somewhere complex conjugates. So this will be the path which they will take, this is your $\omega_0$ the distance is $\omega_0$ and this is $\omega_0$ and this if we join these two curve then this distance is referred to as $\omega_0$ divide by 2Q.

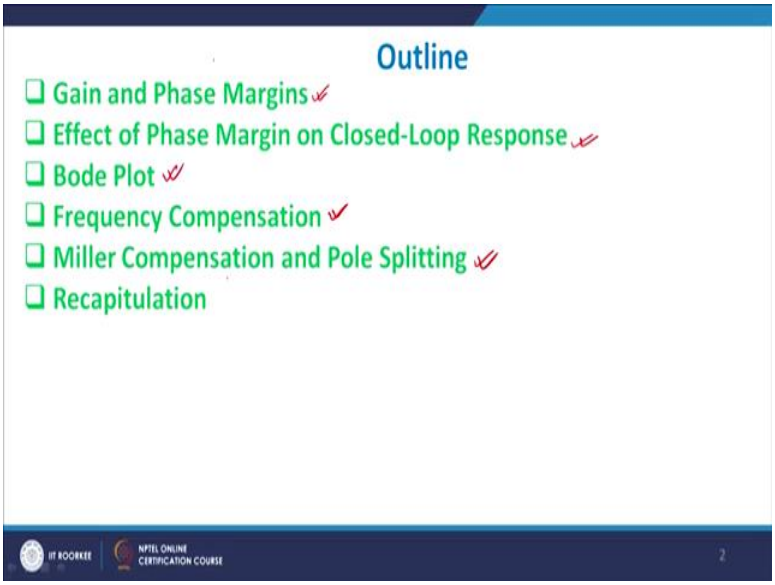
So if you make your Q higher and higher this distance goes on reducing this distance and therefore these two poles start to shift towards the jω axis. Right? And Therefore so at a higher Q this will be the new position of pole at further higher Q this will be further closer to jω axis and therefore you when you go on making Q larger you are actually adding a sinusoidal component to it and therefore you will have sustain oscillation as these two fall on the jω axis in reality.

Further if Q is equals to 0.707 I am not proving it here then you will get maximally flat response which means you will get something like this and maximally flat. If Q is equal to 1 then I get something like this and then it increases like this so this is for Q is equal to 1. For Q less than 0.707 I get something like this, so this equal to 0.5 or let us suppose something like this. Right.

And therefore, 0.707 is maximally flat at this position at this point the poles should be around 45 degree. The phase margin phi should be at 45 degree. Right? What is the phase which you see? Phase is basically the tan inverse of Aβ. Right? If it is a function of this thing and therefore whenever phase margin is equal to 45 degree I will get the most stable option available with me in reality.

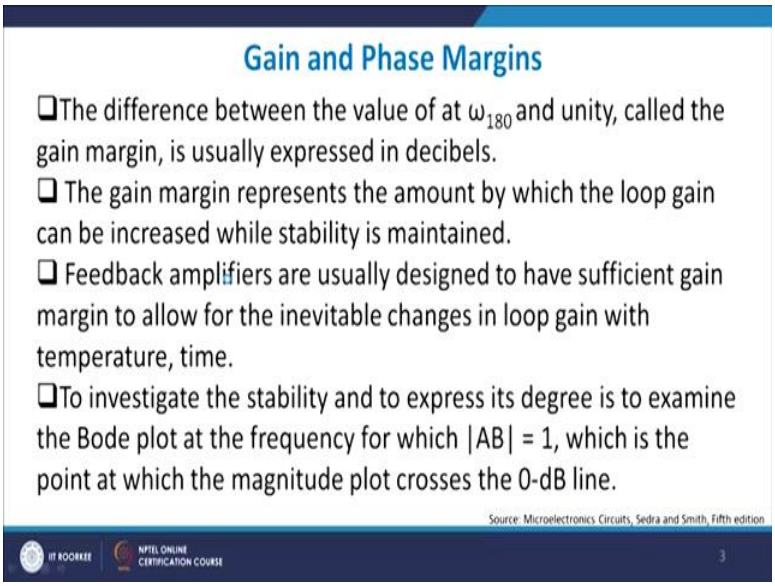
So that takes care of our understanding in reality and let me therefore come back to the this module's major portion and let us see what we will be studying here.

(Refer Slide Time: 10:42)



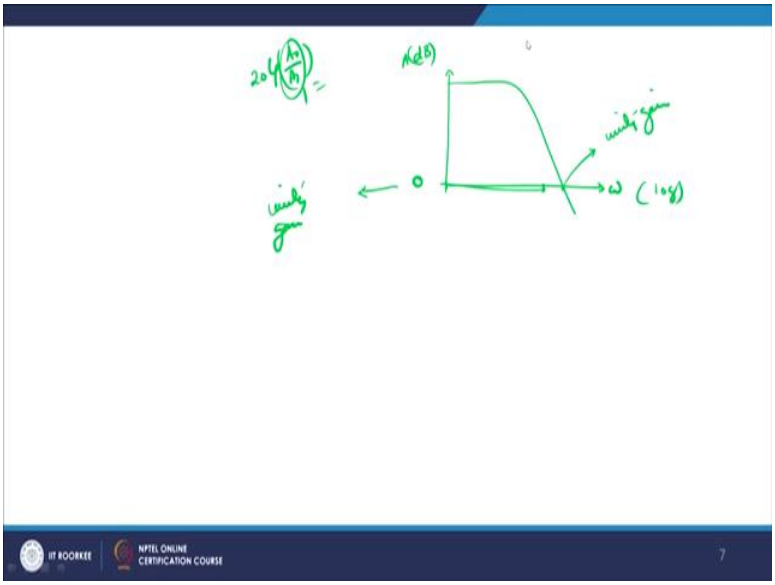

We will look at gain and phase margins. Right? Phase margin, then effect of phase margin on closed loop response, we have already look into the bode plot earlier but with respect to amplifier we will look into the bode plot. Then we will go for frequency compensation and miller and poles splitting we will see. Right? We will go in details of this first 4, but next last two we will keep in a bit low key of air.

(Refer Slide Time: 11:04)



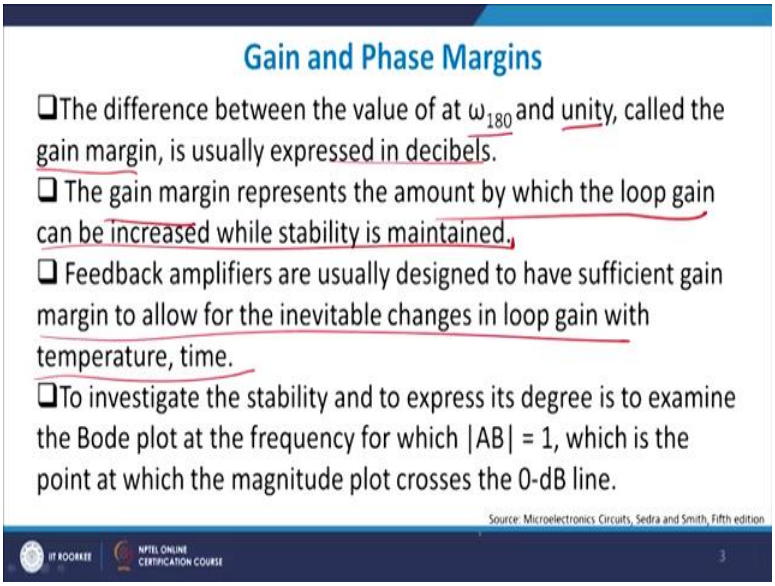

Let us look at the gain and phase margin the difference see as I discussed with you in a bode plot whenever you have whenever you are plotting it.

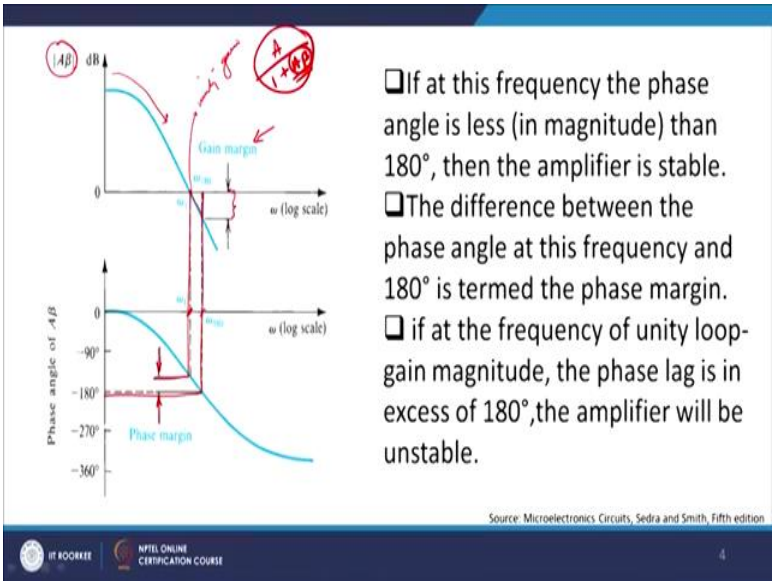
(Refer Slide Time: 11:18)




Suppose you are plotting from plotting a gain. Right? So gain is dB, gain is in dB and then you are plotting here ω in basically log scale. Right? So if you look very carefully if you take 0 dB this is where you get a unity gain this is unity gain. Because if you remember it is 20 log of output by input. Right? So if output by input is equal to 1 log of 1 equal to 0 therefore in dB scale you will get 0 here. So whenever my logs scale comes and crosses here let us suppose then this corresponds to unity gain. Right? Unity gain which you get.

(Refer Slide Time: 12:06)






Now, so with this difference between the values of the gain at $\omega_{180}$ degree unity is called the gain margin and usually express in decibels. I will explain to you from the diagram I do not know I do have a diagram may be let me check.

Yes I have a diagram here. And if you look very carefully this one on the y axis you do have the Aβ term basically A actually because β is very small let us suppose, then we get loop gain verses ω then as I increase the frequency loop gain starts to fall down be on the particular region, it crosses this point ω1 where gain equals to 0 and therefore this point basically your unity gain.

Now, we define and then we start let it move on the negative side so the dB goes to negative side, as it goes to the negative side it crosses so this $\omega_1$ let us suppose some values of ω, say 130, and then as it crosses this particular point and comes to this point where you have $\omega_{180}$. Right? At 180 degree then we define this difference between this point and this point as a gain margin. Right? So how you define the gain margin?

The difference between the gain from which point to which point? From the point of unity gain to a point where you shifted to ω equals to 180 degree. Right? So if you do its corresponding region so if you see this curve goes to 180 degree. Right? Phase margin and this is the point where you have gone to some 130, 140 whatever, then this is defined as my phase margin. Right? This is been defined as my phase margin. The difference between when unity gain to ω equals to 180 is defines as my phase or the gain.

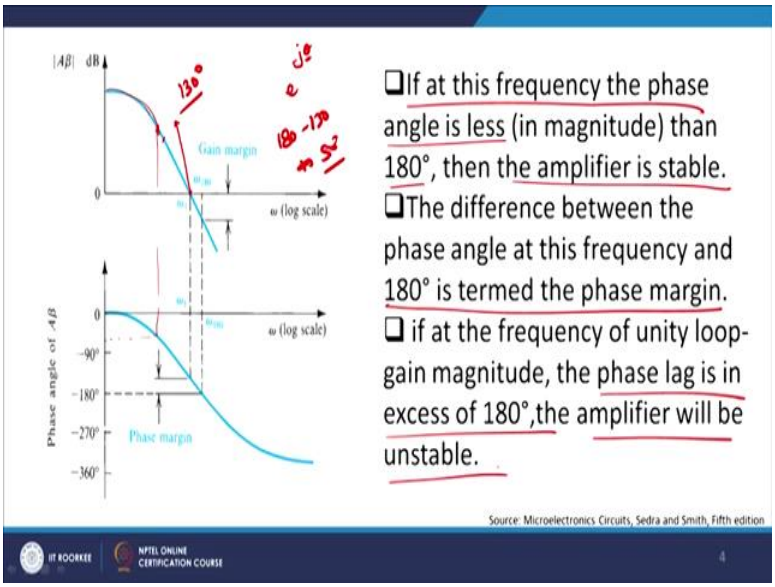
So I have got two margins here one is known as a gain margin and one is known as a phase margin. If you will go back therefore the gain margin therefore represents the amount by

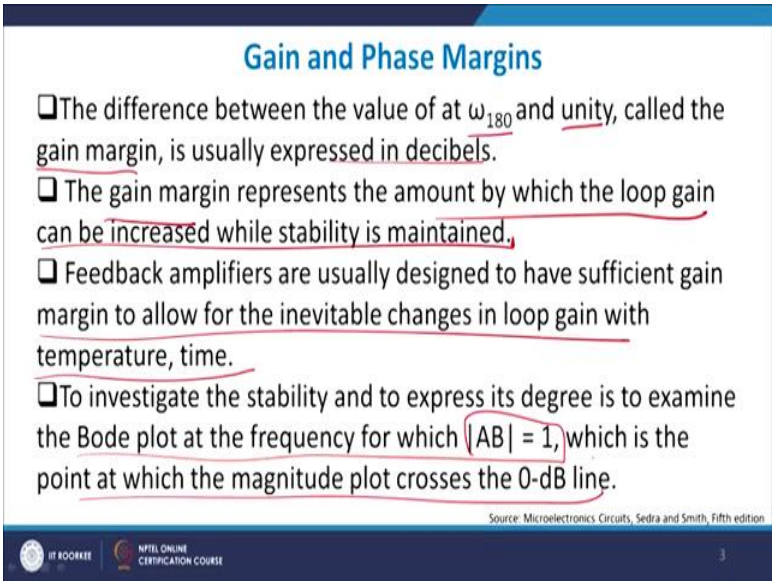
which the loop gain can be increased while the stability is maintained. What does it mean? It means that if you look very carefully that this difference is if the gain margin is very high then I can increase the loop gain because loop gain is Aβ, so I can still have loop gain go higher and I do not have the stability problem issue coming into picture, remember? What was the stability problem? Stability problem was that A upon 1 plus Aβ was there.

Now, if Aβ has to be a negative quantity then this quantity will shoot up. Right? And this Aβ will become negative quantity at ω is equal to 180 degree. We have already discussed that point, so the I do have a time of making my Aβ even go to 0 value or very close to 0 value without going to negative value and I can still have a large amount of negative feedback coning into the picture. Right? That is what is defined as my so what is gain margin. Right? So it represents the amount by which the loop gain can be increased without affecting my stability.

Now, feedback amplifiers are usually designed to have a sufficient gain margin to allow for the inevitable changes in loop gain with temperature and time. I will keep this point very clear.

(Refer Slide Time: 15:26)






See what happens is that if you suppose let us suppose you take this graph into the consideration. Right? And you have inadvertently done what thing? That you have placed your device somewhere here or you have placed somewhere here. Generally what happens around so this is your 45 degree point. Right? And this is you have placed here, so if you place somewhere here or somewhere here, somewhere you are stable and it is working fine.

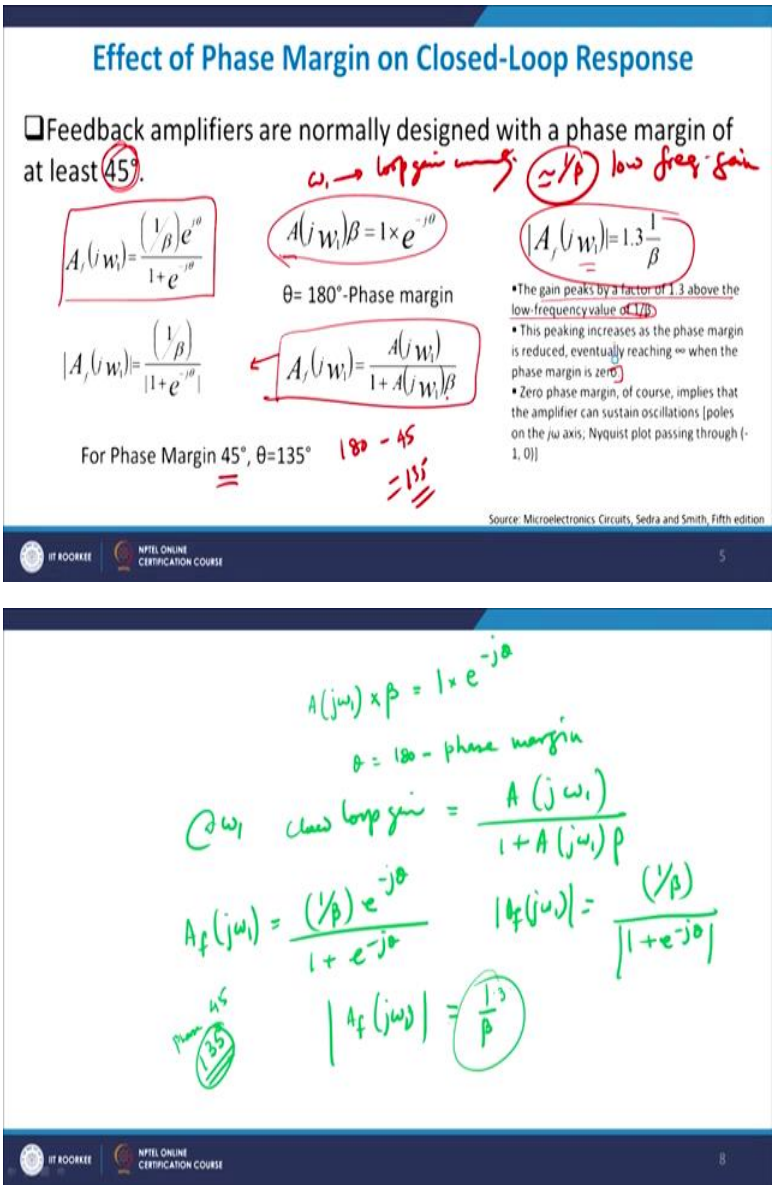
But especially in analog design an designs where the currents are high you do have the chip temperature goes high. So the temperature goes high your β your A which is the gain those also change and when they change the Aβ value also changes so I should have a large margin between the point where your unity gain till equals to ω equal to 180. So that will allow me to give a large headroom for the Aβ value so that I am not going away from the stability criteria.

Now, so to investigate stability and to express its degree so we example the bode plot at frequency at which Aβ is equal to 1 which is the point at which magnitude plot crosses the zero dB line. We have just discusses this last point here. Now, if at this frequency which frequency? At this frequency the phase angle is less than 180 degree then the amplifier is stable. And that is what I was saying, that if at any frequency your phase angle is less than 180 degree you are stable if it is more than 180 it is unstable and the reason is you remember e to the power jθ. Right? And therefore at 180 degree this will becomes to minus 1. And as a result you will automatically have a positive feedback.

The difference between the phase angle and at this frequency and 180 degree is term as the phase margin, which means that let us suppose I will give you an example, let us suppose at this angle my phase is 130, then we define phase margin as 180 minus 130 that is 50 degree.

Will be a phase margin. Which means that this much of amount of phase I am able to fore go even if without compromising on the stability criteria of the amplifier. Right? If then again the phase lag is in excess of 180 degree the amplifier will be unstable. We have already discussed this point just now.

(Refer Slide Time: 17:48)




So let us look at the effect of phase margin the close look response, I have the close loop response, close loop means basically I do have feed forward amplifier connected to Aβ which is basically my feedback network and then it is mixed to a mixture or to a sampler and then the error signals goes again to the amplifier. So this is a basic close loop feedback.

Now, generally in most of the cases we defined feedback amplifiers are designed such that your phase margins are approximately equal to 45 degree. Right? At 45 degree you have

maximum stability criteria coming into picture. Right? And that gives you a pretty descent values of this thing. Why? Because tan45$^o$ if you see it will give you 1. Right? And tan45$^o$ 1 primarily means that the most stable criteria will be coming into picture.

Now, if you see that for phase 45$^o$ I told you the phase margin will be this minus 45 and that is equal to 135 degree. Right? So I get A($j\omega_1$). Right? $\omega_1$ is the point where I am assuming that loop gain is unity. So $\omega 1$ is the point where my loop gain is unity. Right? And my low frequency gain will be how much? 1 by $\beta$ approximately. This will be my low frequency gain. So if you look at this particular issue then I get if you solve it I get low frequency gain is as I said 1 by $\beta$ so 1 by $\beta$ e to the power $j\theta$ upon 1 plus e to the power minus $j\theta$. Right? Converting into polar of coordinate systems.

So I can wright down therefore that if we take mod value of that I get mod of this equal to 1 by 1 upon because this will go to 1 and I get 1 by $\beta$ coming into picture, 1 plus e to the power minus $j\theta$ comes into picture and I can safely wright down this to be is 1 into the power minus $j\theta$. Right? Therefore I get with feedback to be equals to this consideration is there with me it is always equal to the value.

Now so the closed loop gain, this is my closed loop gain with feedback. Right? Is a quantity which is given by this, now let me do one thing let me show to you that how I am getting it, so if you do a small derivation here I can safely wright down that A of j $\omega 1$ which is unity gain into $\beta$ will be equals to 1 into e to the power minus $j\theta$ where $\theta$ is phase margin, where $\theta$ is equal to 180 minus the phase margin. Right? At $\omega_1$, $\omega_1$ where unity gain this thing is there, the close loop gain. Right? Will be given A j of $\omega_1$ divided by 1 plus A ($j\omega_1$) into $\beta$.

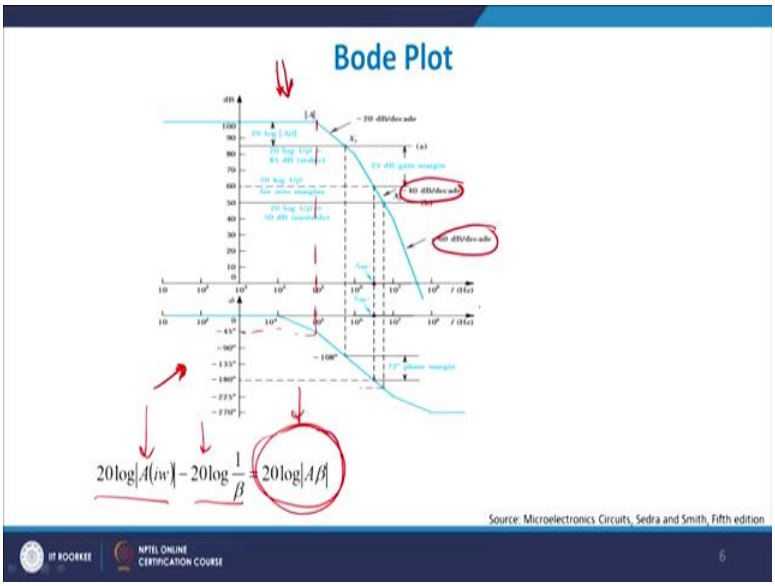
So I get $A_f$( $j\omega_1$) with feedback equals to 1 by $\beta$ as I showed I the earlier slides e to the power minus J$\theta$ upon 1 plus e to the power minus $j\theta$. So if I take the mode value of $A_f(j\omega_1)$ with feedback then this comes around 1 by $\beta$ because mod value of this will come out to 1 and I can remove it and from here I can get 1 plus e to the power minus $j\theta$ mod of that. Right? Mod of that.

Now, mod of that if you solve it this comes out to be equals to 1 and therefore I get $A_f$( $j\omega_1$) mod of this happens to be equal to 1 by $\beta$ times. Right? And I get this to be approximately equals to 1.3. Right? Assuming that phase angle is 45 I get phase margin to be equal to be 135$^o$. Right? So tan inverse of 135$^o$ becomes out to be 1.3 by $\beta$. Right?

So this is what I want to do here that A of f means that with feedback at unity gain bandwidth I get 1.3 upon β, so the low the gain peaks by a factors of 1.3 above low frequency value of 1 by β. This peaking increases of the phase margin is reduced eventually reaching to infinity when the phase margin is 0. Right? So what happens is with feedback your gain became accessibly high and it becomes infinity at phase margin equals to 0. Phase margin primarily means that your θ is 180 degree. So at 180 degree is the point beyond which if you increase your phase margin you are in a problem that you will enter into a non-zero part. Okay.

So, 0 phase margin of course implies that the amplifier can sustain oscillations. Right? So poles are in jω axis. Right? And that gives you 0 phase margin, so the 0 phase margin is the best one, but if you want the stable oscillation keep it at 45 degree that is what we have learned across this whole thing.

(Refer Slide Time: 23:13)




Lets come to the bode plot and we already and from bode plot we can also calculate the stability criteria for the design. Let us see how we can do that. Remember 20 log of jω minus 20 log of 1 upon β will be 20 log Aβ. So, if I subtract this and this I get this, and I will be able to plot this with respect to ω. Right? Therefore if you look at this particular issue and this is my phase margin plot and this is my frequency gain operation plot. Then you see that it is maximally stable at up till this much frequency. Right?

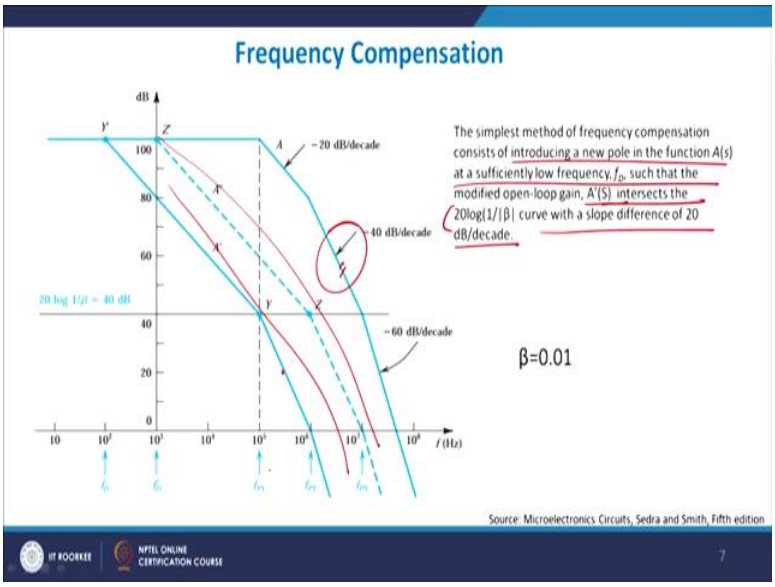
If you see this will be around 45 degree so I get maximum stability at 45, till 45 get a maximum stable, then I get 20 dB per decade drop because this first order system, I get 20

DB per decade drop. It reaches till $X_1$ which is basically a point where you have got to reach the value of 135. Right? And 135 you have reached.

Further if you go down I reached to a point equals to 180 and beyond this particular point your become unstable. Right? So your gain falls down drastically lower it reduces at the rate of minus 40 dB per decade and here minus 60 dB per decade. So you see even with the one decade change in the frequency you have 60 dB drop taking place here, which means the gain is not at all stable here and it falls very fast beyond a phase margin of 180 degree.

So this plot gives you an idea about the fact that 180 is basically therefore the phase margin where the stability criteria can be violated if you go beyond that particular point and therefore the you can also get the stability criteria understanding from this bode plot as well.

(Refer Slide Time: 25:15)



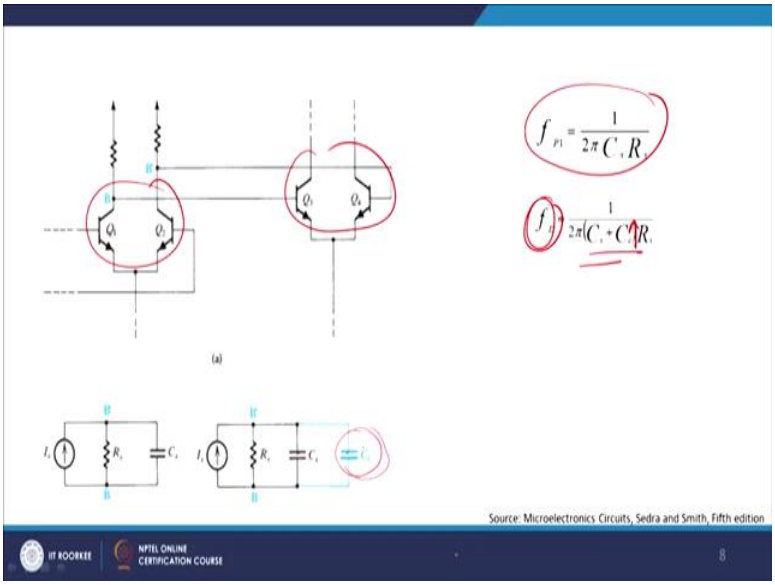

What is the meaning of frequency compensation is that the simplest method of frequency compensation is introducing what is the meaning of frequency compensation? See the problem which we have faced earlier in our earlier discussion that till 180 degree you will have stability and beyond 180 you will your gain will be falling very fast and you will have unstable system with you.

We required to compensate for an increased frequency here, and to do that what we do is that we introduce a new pole at very very low frequencies at sufficiently low frequencies such that this A'(S) intersects 20 log 1 by β curve with the slope difference of 20 dB only. Which means that I allow so I just told you previous slide that it will be 20, 40, 60 dB per decade drop, but if you are able to sustain a 20 drop even beyond 180 degree phase margin then it

will be relatively less or it will be relatively more stable. Right? And that gives me better frequency compensation in reality.

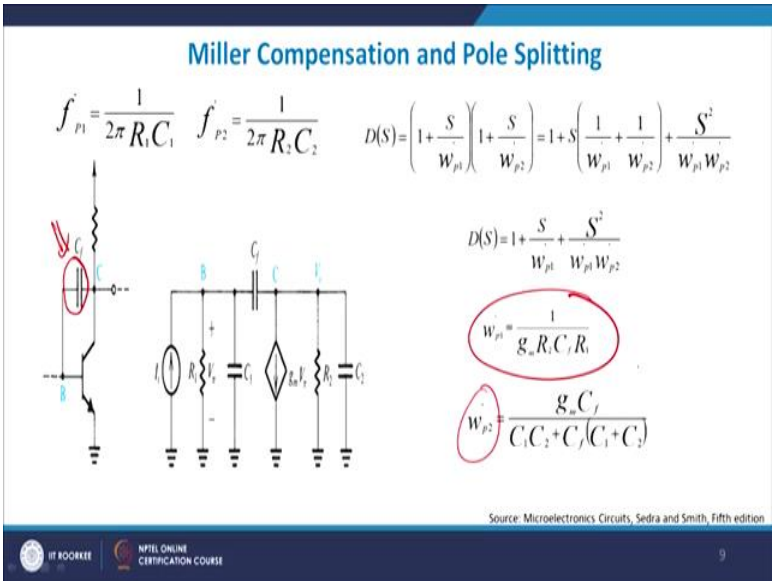
As you can see here I only for example I have just remove this 40 dB decade drop here and if you go ahead and do this I get this compensated profile and I get this to be a compensate profile. Which means I have left alone 40 dB here and I only get a 40 dB drop at this particular point beyond a particular point $f_{p1}$, so this is permanently more stable as compared to previous case.

(Refer Slide Time: 26:47)




This is another method of showing a frequency compensation, how do you do that you put BJT here, you have BJT complex here and you have a BJT complex here and this is represented by a current source and a resistance current source and a RC resistance then you apply a compensation capacitor here. As you apply compensation capacitor here the frequency is inversely proportional to the capacitance and since these two capacitors this is are in parallel they add up and as a result the $f_d$ prime value actually shifts to the left, because this reduces because of the increased value of the capacitance here. And as a result what will happen is that the point where it cuts the 180 degree curve will shift to the left and you will automatically get a better compensation in this case.

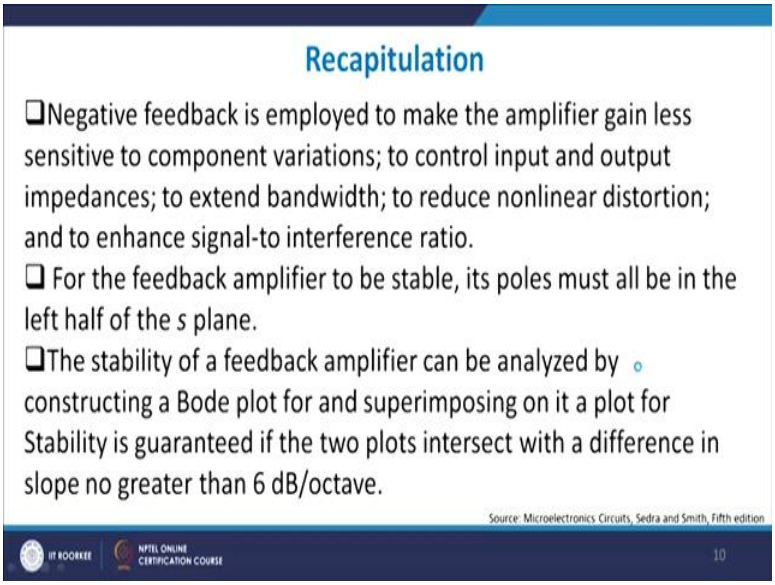
(Refer Slide Time: 28:04)




We will not discuss in detail this slightly ahead of its course but just to give regularity and to give you an idea about where you are looking to discuss this point and we will close here. That we also try to do one more important point and that is we try to split the poles into two parts, so rather than having a single from a single order system can be have multiple poles and those multiple poles can add to the stability of the system. Right?

And this is one of the methods of doing it and I will not go in to the details of whole method but typically applied to a BJT, here you apply $C_f$ which is basically my feedback capacitance here and also referred to as a Miller capacitance and if you put an feedback capacitance here my ωp1 actually becomes a function of $C_f$, similarly $\omega_{p2}$ also becomes the function of $C_f$ and simply by making $C_f$ higher and higher I can shift relatively the $\omega_{p1}$ and $\omega_{p2}$ from its initial values and this is known as pole splitting in analogue design. We will not talk any further than this at this stage because it is slightly ahead of this course. Right?

(Refer Slide Time: 29:05)




Therefore let me give you recapitulation of what we did in this module and the previous one, we have taken up the fact that negative feedback is employed. Right? And to make the amplifier again less sensitive we have already learned that we have less reduced non-linear distortions enhance signals to noise ratios, the poles as I discussed with you for stability the poles all will be in the left half of the plane S plane. Right?

The stability of the feedback amplifier can be analyse by constructing a bode plot and superimposing a plot for the stability, and we ensure that difference between the two plot is not more than 20 dB per decade or 6 dB per octave difference is not there. So we sustain a 20 dB per decade drop even after 180 degree phase margin and then the stability will be maintain.

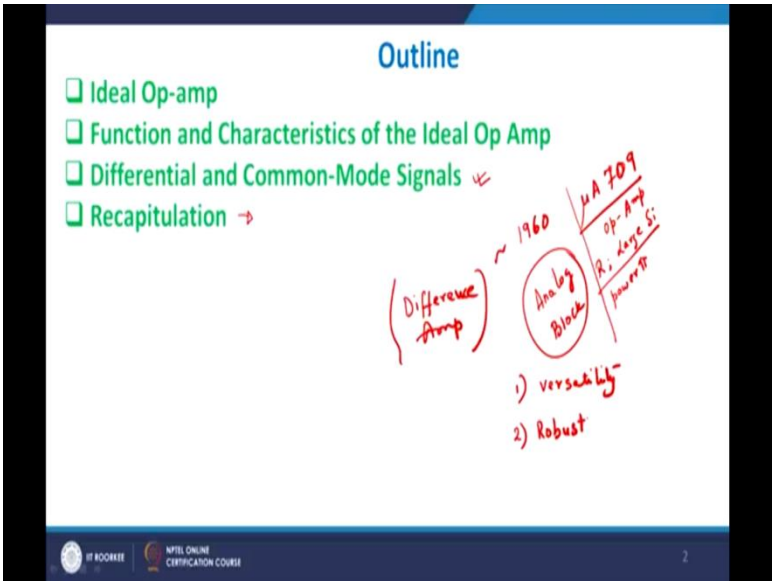
Similarly, we can also therefore use frequency compensation for the stability of the amplifier, by choosing an appropriate value of the compensation capacitor and making the poles. We have also learned finally about what is known as the poles splitting but that you need not worry about because it requires further basic knowledge to understand this, if time permits later on we will revisit this part latter on. Right? And that keeps my this thing ready.

From the next module onwards we will actually looking into operational amplifiers and design of operational amplifiers and then utility of operational amplifiers for analog circuits. Right? So till the next week the lectures will be over as far as analog design is concerned and then we will shift to digital combinational block design as per the syllabus of the NPTEL course. Right? Thank you very much for your patient hearing.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−50**
**Ideal Operational Amplifier and its Terminal**

Hello and welcome to the NPTEL online certification course on Microelectronics Devices to Circuits. Today we will start with a new chapter, or a new section as per the syllabus and the chapter is on operational amplifiers and its terminals. So the name of the module is ideal operational amplifiers and its terminals. What we will be doing in this module is the following.

(Refer Slide Time 1:04)




We will look into what is known as an ideal Op-Amp, when we say Op-Amp, we mean to say operational amplifiers and we will see why is it named so. We will look into the function and characteristics of an ideal operational amplifier. Then we will be looking into the differential and common mode signal here, right? And then we will come to non-inverting inverting mode and then we will finally recapitulate our discussion.

And now way back in 1960's or around 50's and 60's, you had the first operational amplifier been designed which was in IC form, in so, it approximately in 1960's or 65's. You had this your operational amplifier which was named as μA709. So this was a first sort of an operational amplifier which was designed. It consists of large

number of resistors. It consists large silicon area, right. And the silicon area was typically very large in this case.

As a result, it also consumed very high power. So the power was also very high, right. And this was in the analog block available tome. So it was basically an analog block and as I discussed with you, this analog block did consume large amount of power, but this was the first time we are able to look into the operational amplifier. What is the operational amplifier basically means? It primarily means that it senses the difference between two voltages and then amplifies it to have a output voltage, which is just a difference of the two voltages.

So I, if I have got two voltages maybe sine wave, 2 sine waves and then we subtract the 2 and then multiply with the fixed gain value 'A' and that is what output is all about. So that is the reason it is also referred to as a difference amplifier, right. It is also referred to as a difference amplifier. Why difference? Because it is actually trying to look into the difference of 2 voltage sources and trying to find out the output voltage depending on the difference of the 2 voltage sources.

Now, the idea was why it became famous? Or operational amplifier was used quite often in large amount of instrumentation design, was first of all it is versatility, right. Versatility so, the first reason why it became very famous was versatility, and the reason why it was so versatile is that you put this operational amplifier anywhere within the structure, whenever you have an analog signal or for that matter even if digital signal is available to you, but primarily 90 percent of the cases, analog signals. You can do large amount of computation in analog domain using operational amplifier, right.

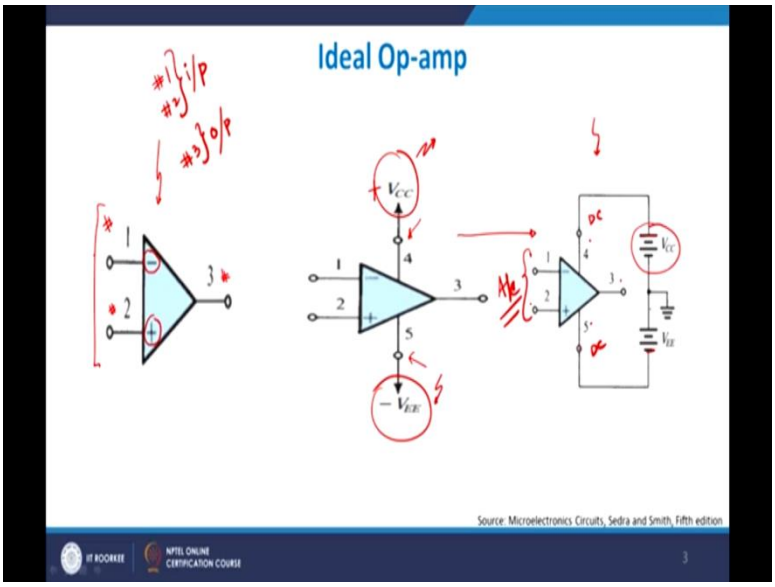
So you can add 2 signals. You can subtract 2 signals. You can do multiplication of 2 signals. You can do cascading of 2 signals so on and so forth. So this was not possible earlier, remember. It was possible through a bipolar technology, but operational amplifier through bipolar technology is also available to us and the gains are relatively very high. So you get a large gain and you also get a versatility and you can use this across many signals, signal domains.

The second thing is its robustness. It is quite robust, in the sense that the offsets or the variabilities are very very low in this case. So if you have designed a Op-Amp which is say working with an amplification of 100 then you will be 100 percent sure or you

will be pretty sure that the gain is fixed at 100, irrespective of the ambient, irrespective of the variations in the parameters of the devices used in the operation amplifier. It is almost fixed at approximately 100, right.

So it is two advantages, it's versatility and it is very robust in design. That was the reason Op-Amp is still very very important in IC whenever you are, you are using it in specially analog and mixed signal blocks of your VLSR design.

(Refer Slide Time 5:17)




Let me show you how an ideal operational amplifier looks like. Ideally for a user, the ideal operational amplifiers look like this, which is this one and it has got 3 terminals here. Terminal number 1, so this is terminal number 1, right. This is terminal number 2 here and this is a terminal number 3. So there are 3 terminals and there is a specific sign to the first terminal minus. Specific sign to the second terminal plus, and you have an output here which is $V_{out}$. So, this is your output terminal.

So, terminal number 1 and terminal number 2 are basically your input terminals, right. And terminal number 3 is your output terminal. So, it is physically a 3 terminal structure where in the first two terminals 1 and 2 are basically your input terminals and structure 3 is output terminal here.

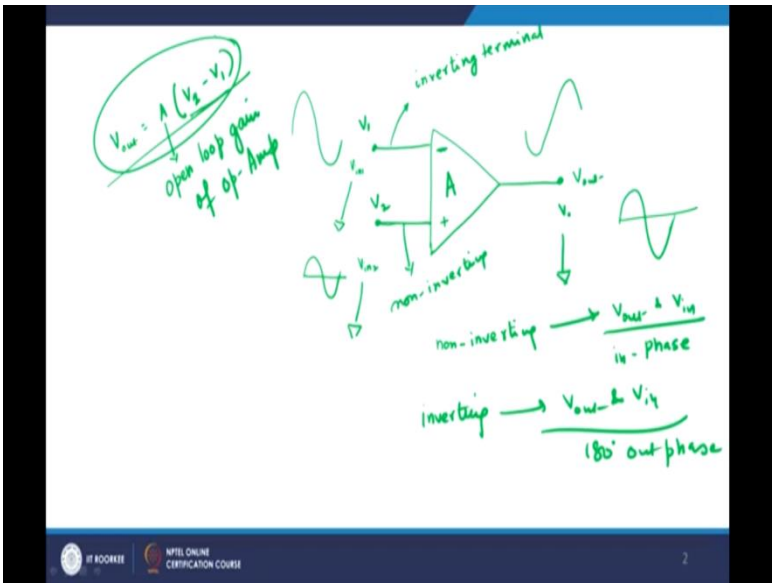
Now, since this is active device obviously it is operational amplifiers are made up of differential amplifiers that you possibly be knowing also or maybe I can give you a extra talk on that, but these are internally they are made up of difference amplifiers, differential amplifiers which we have already studied on our earlier modules.

So, you require an external power source to drive these operational amplifiers. And this is where this fourth and fifth terminal comes into picture. We generally do not show these terminals or we do not show whenever, whenever drawing an operational amplifier, we do not show it but typically these remain there and this is given as plus $V_{CC}$ here and a minus $V_{EE}$ here. So, I have a plus $V_{CC}$ here and I have a minus $V_{EE}$ at this particular point. Obviously, this is third terminal which is available to you.

Now therefore, I can therefore correlate this to this terminal, that terminal 4 is biased in this in this manner such that the positive terminal of $V_{CC}$ is connected here and the negative terminal of $V_{EE}$ is connected to terminal number 5. So I have got therefore typically at this stage understanding purposes, 1, 2, 3, 4 and 5. 4 and 5 are primarily DC biases. 1, 2 are input and 3 is the output.

Now, if you look very carefully across this network which you see here, these are the DC biases which you apply and here you apply the AC signal, right. So sort of a mixed signal approaches here, but DC biases is done in order to or in order to make those transistors or a bipolar transistor or even a MOSFET in the active region of operation so that they start behaving like a current source and you do have a large gain available with you. So this is the basic structure of an operational amplifier and basic structure looks like this in a real sense.

(Refer Slide Time 8:09)




So, let me show to you or let me give you an idea of what the idea is. The idea is something like this. So if I have got a as I discussed with you that I have a operational amplifier like this, right. And I have got negative and positive, so this is your this and

this your this. This terminal, negative terminal is referred to as an inverting terminal, inverting terminal. And this is referred to as a non-inverting terminal.

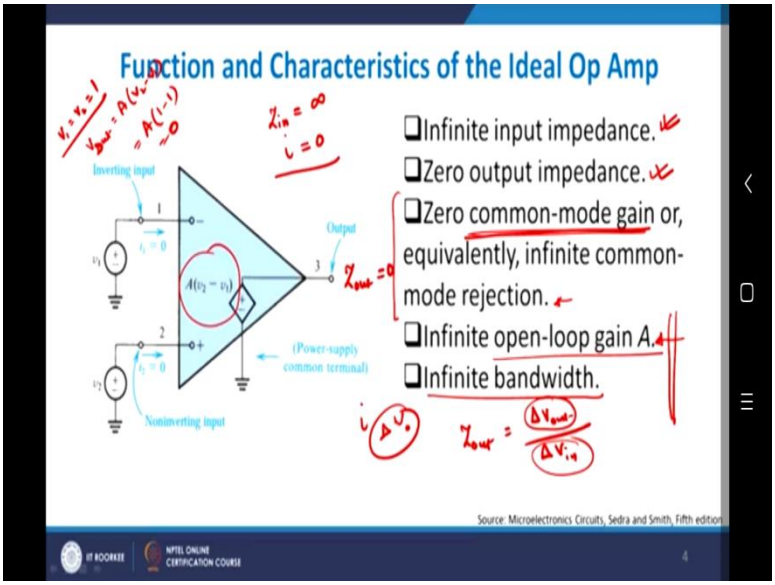
Inverting terminal primarily means that any signal you give here right, will be inverted or will have a 180 degree phase shift between input and output, output where? Terminal number 3. So this is your output, $V_{out}$ and this is always measured with respective ground. So all of these measurements, so this is a $V_{out}$. So all these measurements which you do here, this is with respect so so this is $V_{in1}$ let us suppose again with respect to ground. And then you have $V_{in2}$ here, again with respect to ground. So all are with respect to ground here.

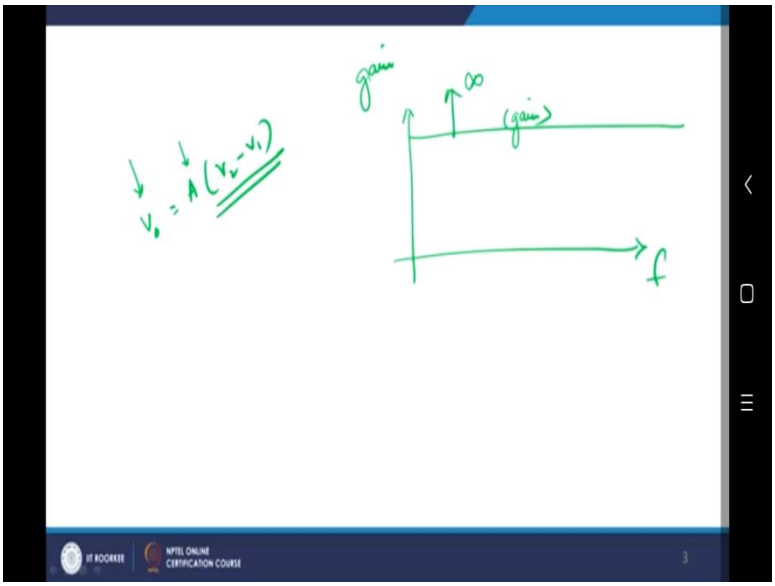
And whatever signal you give on a inverting terminal here will appear as 180 degree phase shift. So if you have got something like this it it will appear something like this in the output side. Whereas in the non-inverting terminal if you give like this it output will be something like this, right and you have to ensure that, so they they will be in phase. So if I talk of non-inverting terminal, non-inverting inverting then I will have $V_{out}$ and $V_{in}$ are in phase, right, in phase. And if you have got inverting terminal then $V_{out}$ and $V_{in}$ will be 180 degree out phase this is quite important.

And therefore, and therefore, let us suppose this is suppose, this is $V_1$ and this is $V_2$, then my output looks like $V_{out}$ or say $V_{out}$ in this case will look like A multiplied by $V_2$ minus $V_1$. This will be the general equation of an ideal operational amplifier. Now you see, it is $V_2$ minus $V_1$ why? Because it is connected to the inverting terminal. So whatever input you give, it will be always phase shifted by 180 degree whereas $V_2$ is basically my inverting terminal.

Now, when I so, so whatever signal I am giving, is the difference between the 2 signals, is basically my output multiplied by A. A is referred to as open loop gain of operational amplifier. So, A is defined as the open loop gain of the operational amplifier which means that A, basically this is the gain of the differential amplifier which is there within the operational amplifier, right. And therefore, you get $V_{out}$ to be equals to A times $V_2$ minus $V_1$, right and this is what general scheme of things appears here.

(Refer Slide Time 11:28)




Now, what you see in this case is that if you look, now let us draw it's appropriate circuit diagram its equivalent circuit diagram. So if you see here, I have an inverting terminal. I have a non-inverting terminal. It has been seen that the inverting, that the input of my input of my Op Amp is having $Z_{in}$ which is input impedance to be ideally infinite which mean that the current is actually equals to 0. So, no current is entering into the terminal of Op Amp, right. And you can understand from your basic theory also.

See if you remember from your differential amplifier that 2 inputs are towards the gate, remember. The gate of the MOSFET, right. The gate of the MOSFET, automatically, if you remember has got a very very high input impedance because

there is an oxide layer and therefore, it is a dielectric. And that is the reason you do not have any current flowing through the gate side, right. Almost 0 current, which means that the input impedance is relatively very high. In ideal Op Amp the input impedance should be infinitely high which means that your voltage remains fixed, right and your current is almost 0.

Whereas in the output side your $Z_{out}$, will be equal to 0 in an ideal case. Which means that you can draw nay current you want to do, right. You can draw any current and and typically your voltage will remain fixed. Or otherwise or otherwise the current will remain fixed and you can vary any voltage which you want $\Delta V_{out}$. So how I will define my $Z_{out}$? $Z_{out}$ is defined as $\Delta V_{out}$ by $\Delta V_{in}$, right $\Delta V_{in}$, which means that if which means that $\Delta V_{out}$ should be equals to 0, which means that I should not see a change in the voltage with change in current.

So I can do, I can do a large variation in the current in the output domain, but my voltage across the terminal 3 and ground should remain fixed and therefore, that is my 0 output impedance which you see here. I will come to this point just now but let us concentrate here, that open loop gain for ideal operational amplifier is infinitely high

and it's infinite bandwidth, both of this treatments are for ideal operational amplifier.

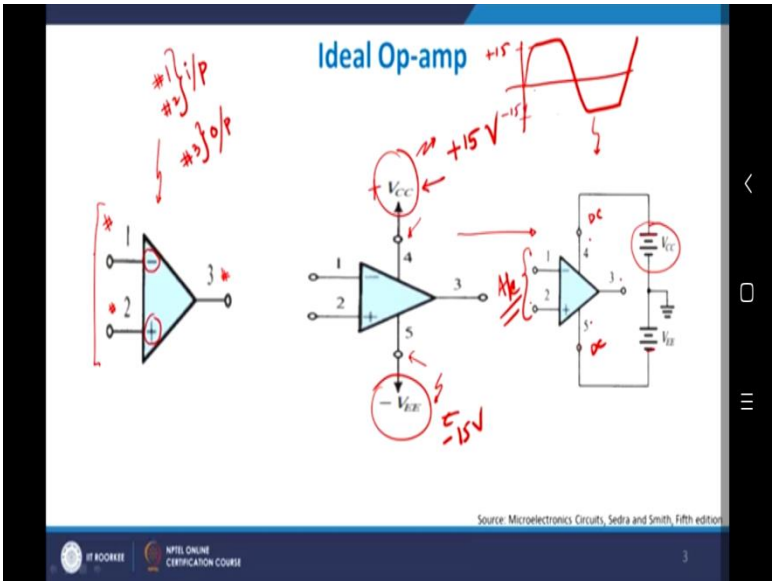
Which means that if you are able to plot, if you plot gain versus gain versus omega or frequency for operational amplifier, right, it will look something like this, for ideal case it will be infinitely bandwidth and the gain is always fixed and the gain is always fixed and very high gain, infinite gain. So this will be around infinite, right and that is what is happening here.

Now you see here, within the blue triangle which is a schematic of the Op Amp this is the voltage which you will see, A times $V_2$ minus $V_1$. So it is a voltage source here and therefore, the output is given as A $V_2$ minus $V_1$ for all practical purposes. We come to this one now. See, it tells me that common mode gain is 0 which means that if any signal which is common to terminal 1 and 2 will always be rejected. So if I have $V_1$ equals to say equals to $V_2$ equals to 1 volt then of course ideally my $V_3$ or $V_{out}$ will be equals to A times $V_2$ minus $V_1$ and therefore, A times 1 minus 1 will give you 0.

So my $V_{out}$ will be equals to 0, which means that any signal which is common to both the terminals will be automatically rejected or will not be accepted or gain will be almost equals to 0. This is known as a common mode gain. Common mode, why? Because you are giving a common signal to both the inputs of your operational amplifier and therefore they are gain, right. And therefore, and therefore so, common mode gain is high.

Now, but my differential gain is relatively high. Why differential gain? Because differential gain by definition if you see is $V_{out}$ is equals to A times $V_2$ minus $V_1$. This is your differential gain. The difference between the 2 signal gets amplified by a large value which is infinitely large in ideal case and therefore this will be infinity large. But we will see that it does not, it will never be infinity, it will have some values associated with it, can you tell me why it will not be infinity? Can you think about it why it will not be infinity even if your A is infinity?

(Refer Slide Time 15:20)

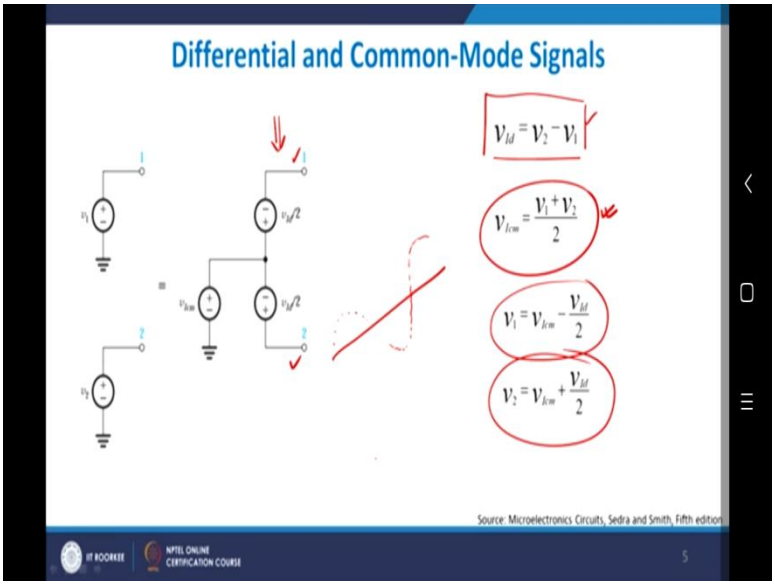



The reason is see I discussed with you just now that in the previous slide that this is limited by plus $V_{CC}$ and minus $V_{CC}$. So if your plus $V_{CC}$ is plus 15 and let us suppose this is minus 15 volts right. You do anything you want to do, the $V_{out}$ value cannot exceed plus 15 in the positive cycle and cannot exceed minus 15 in the negative cycle because these are limited by the power supply here. And if it crosses there will be clipping.

And therefore, what you can see is if the, if typically the A value is large you will expect to see something like this. Clips. So it will clipping at plus 15 here and it will

clip at minus 15 here. And so your distortion will take place. And therefore, an ideal Op Amp has got a 0 common mode common mode gain and has got a large infinite gain with with you, right.

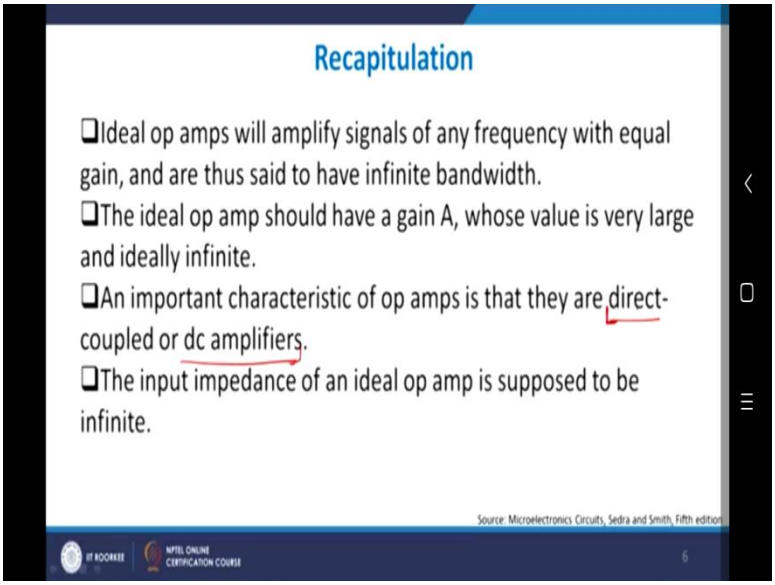
(Refer Slide Time 16:08)




So, let me come to the next case and explain to you the differential and common mode signals so I said differential signal $V_{Id}$ will be given as the difference of the 2 signals $V_2$ minus $V_1$ whereas common mode signal is given as the average of the 2 signal, $V_1$ plus $V_2$ by 2, right. So, what is common mode signal? Common mode signal as you can see here is basically a DC bias which is given to or even a AC bias which is given to both, common to both the both the terminals, inverting and non-inverting terminal of my design. You will ask me why is it given?

Well, this is given very simply because that you want to bias it in the active region, both the MOS devices, remember. So, you need to give an external bias which will try to do it, that plus we will see later on that has to do with the noise of the source also. So I get $V_{Icm}$ is equal to $V_1$ plus $V_2$ by 2. Therefore, if you just place it and do some manipulation here, I get I get this V1 equals to VIcm minus $V_{Id}$ by 2 and I get $V_2$ to be equal to $V_{Icm}$ plus $V_{Id}$ by 2. So, I have $V_{Icm}$ and then the voltage goes up $V_{Id}$ by 2 and it goes down by $V_{Id}$ by 2 so on and so forth, right. So this is the output voltage which you see or the input voltage what you see right.

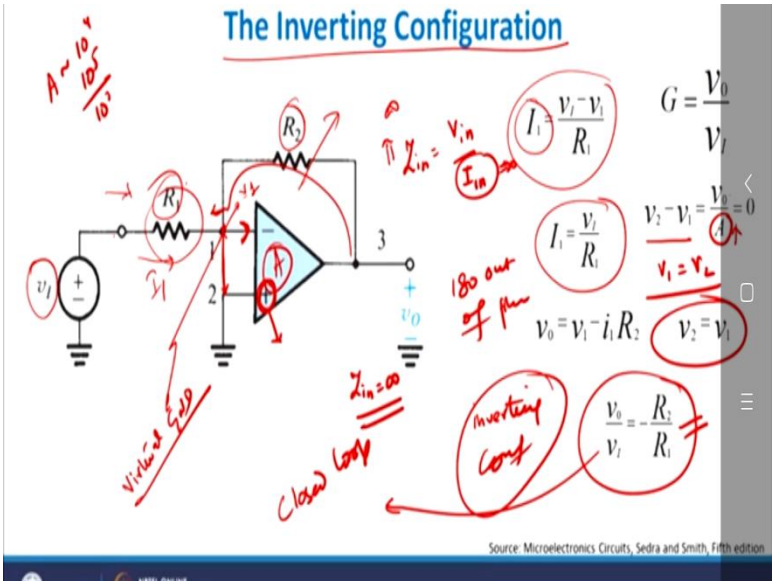
(Refer Slide Time 17:22)




So, let me recapitulate what we learn and then we will go for negative and inverting terminal, inverting terminal, non-inverting terminal in the general sense. So ideal Op Amps will have infinite bandwidth, large gains, A will be also infinitely large. These these are basically input impedances are relatively very high. $Z_{in}$ equals to infinity and output impedances is 0, in reality. These are again a direct coupled or DC amplifiers, we will discuss this later on when time permits, but these are known as Direct coupled or DC amplifiers, right.

Because they are able to able to amplify signals which have very very low frequency signals. So even at a very low frequency good good amplification is available when you do an operational amplifier design, right.

(Refer Slide Time 18:21)




With this knowledge, let me come to very important configurations now which means that Op Amp by itself has got no use until and unless you do have a supporting configuration to to help it. For example, this is the first configuration which we will be studying and known as inverting configuration. As a result, what we will see later on is that this configuration helps you to invert an signal, analog signal. So if I have a sine wave, I will actually get automatically I will get a sine wave, 180 degree sine wave and I can change the gain of the output wave form also.

So how it is done? It is done in this manner. My this positive is grounded. My inverting, non-inverting terminal is grounded here and the inverting terminal is connected across R1 and R2 and through this thing, they connected to the Vout value here. Now the current flowing through this R I1 will be equals to V1 minus V2, Vi minus V1 which is the voltage, this is V1. So, Vi minus V1 by R1 will be the voltage which is the current flowing through this R, I1, right.

And similarly I1 can be written as Vi by R1. Why? Because you see, it is quite interesting that this concept of virtual ground that in a operational amplifier since my Z in equals to infinity, right. Voltage at arm number 2 arm number 1 sorry, here, will try to actually equalize itself to arm number 2 because remember Zin was defined as Vin by I-in, right. Now I told you input impedance is infinitely high so I-in equals to 0. 0 means this will go to infinity and Z in equals to 0.

Similarly, I can also define it in this manner that my V so that is the reason infinity large bandwidth, but this point, point number 1 is also referred to as a virtual ground.

Why is it referred to as a virtual ground is that, if you remember a ground is a place where you will have high currents and low voltages. Because low voltage, because you are grounding it to the ground. High current because since it is grounded, it is a 0 potential, all the current will flow through it.

Here interestingly, 1 will follow 2 so its voltage will be around ground so its very very low voltage, but there will be no current flowing through into this. Because its a high impedance node. So therefore, they are referred to as a virtual ground, right. So, though they are grounded, though their voltage is 0, but you do not get any current into it or out of it and therefore they are known as they are referred to as this thing.

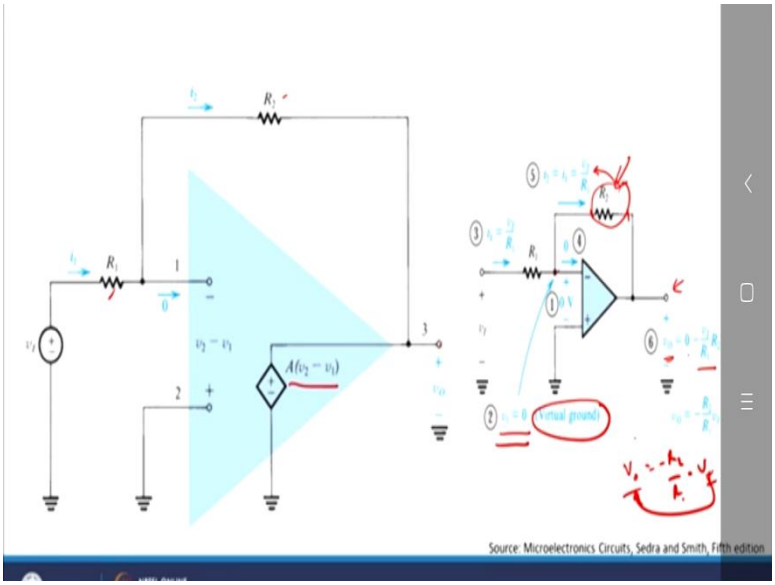
Similarly, so therefore if I have infinitely a large value I get $V_2$ minus $V_1$ equals to $V_0$ by A right, because $V_0$ was equals to A into $V_2$ minus $V_1$. I took it denominator, A is infinitely large so I get $V_2$ equals to $V_1$ which means that $V_1$ equals $V_2$. That was what was talking about. That whatever voltage is there at terminal 2, terminal 1 also starts to go towards that voltage right. This is because of an infinitely large value of amplification open loop gain of operational amplifier.

With this knowledge I do a small derivation and get $V_0$ by $V_i$ equals to minus $R_2$ by $R_1$ right, minus $R_2$ by $R_1$. Now this is known as an inverting configuration, inverting configuration. Why inverting? Because you have a negative sign here, which means that the input and output will be 180 degree out of phase, out of phase right. And its now this is referred to, this is referred to as a closed loop gain. Closed loop why? Because now your loop has been closed by $R_2$ and $R_1$. Open loop when you open it up, we define it open loop gain. That will be just the amplifier gain which you see here. Which means that the closed loop gain will be relatively much smaller as compared to an open loop gain.

Typically A is of the order of 10 to the power 4, 10 to power 5 maybe or 10 to the power 3000 volt by volt. Whereas this might be very very low, maybe 100 or maybe 10, 20 whatever.

So, what we have done is that using this negative feedback concept, we have actually reduced the gain, right but my gain is more stable and it does not depend upon the open loop gain of the operational amplifier which I am using here. That is the major advantage. So therefore, by simply changing the value of $R_2$ and $R_1$ you will have various value of the output voltages in the case of the inverting configuration.

(Refer Slide Time 22:44)

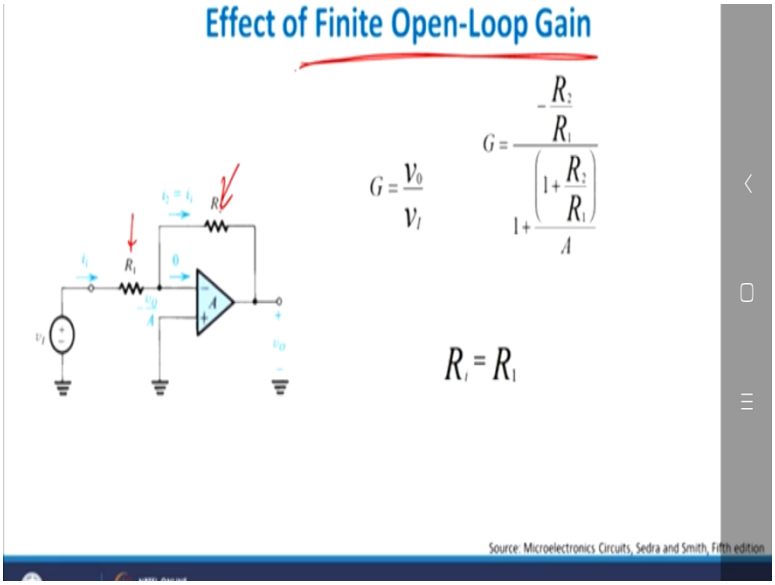



In inverting configuration let me show you the corresponding your corresponding circuit diagram. This is the equivalent circuit diagram which is there with us and therefore I can show it you that in this case for example I can write down that say this is your $R_1$, $R_2$ and therefore this is the voltage difference $V_2$ minus $V_1$ because they are inverting and non-inverting terminal. You multiply A into $V_2$ minus $V_1$ and you get 3 here.

Now therefore, as I discussed with you $V_1$ is approximately equals to 0 because it is a virtual ground and therefore, the current flowing here is $V_i$ by $R_1$ and the same current will trough $R_2$ because there is no other path for the current to flow down. And therefore, at this output terminal I get, output will be equals to 0 minus, why 0? Because this is grounded and therefore 0 minus the voltage drop here will be nothing but minus times $V_i$ by $R_1$ into $R_2$. Because this is the current, current multiplied by this resistance will give you the voltage drop here. And that must be equals to $V_{out}$.

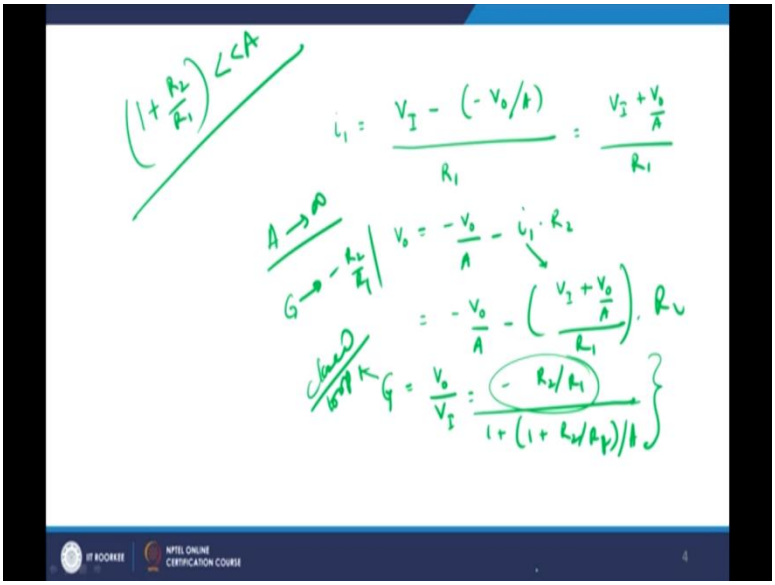
So $V_{out}$ I get equals to minus $R_2$ by $R_1$ into, so $V_{out}$ equals to minus $R_2$ by $R_1$ into $V_1$. So if I take $V_i$ down here as a denominator, I get minus $R_2$ by $R_1$ as my explanations. So that is the reason why we get virtual ground in in in a detailed manner right. And we get typical virtual ground here.

(Refer Slide Time: 23:56)




Now the second aspect is that, let me see, effect of a finite open loop gain, which means that till now when we when we are dealing with your operational amplifier per say, we were looking into the fact that, we were assuming that the gain is infinitely large when in reality not true. Whenever we have a closed loop gain, for example a feedback loop here, $R_2$ is feeding back into the input side, you do have a finite open loop gain. So your open loop gain is not infinitely large, right. And I get finite open loop gain. So let me see, let me explain to you how do I how do I get finite open loop gain.

(Refer Slide Time 24:30)

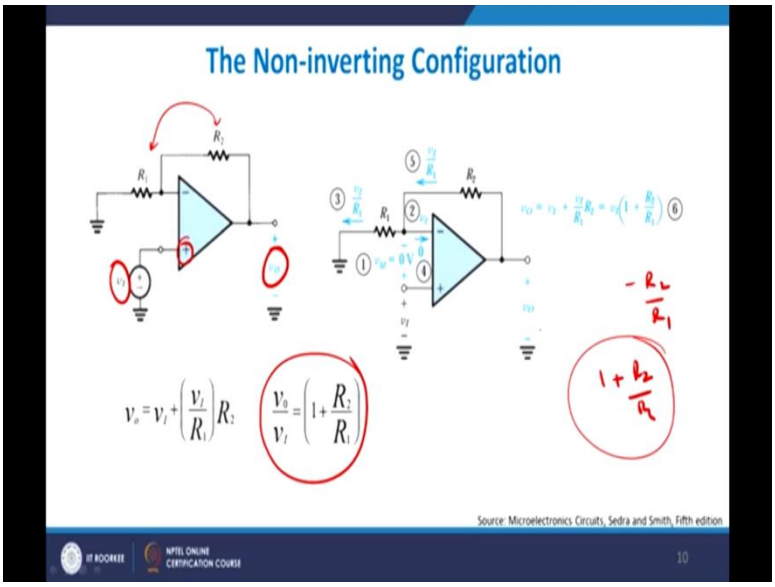


So I get current $i_1$ equals to $V_I$ right minus minus $V_0$ by A. Why $V_0$ by A? By $R_1$. $V_0$ by A is output voltage by A will give you the difference in the input voltage. So, $V_I$ minus that by $R_1$ is the current. So, I get $V_I$ plus $V_o$ by A by $R_1$ equals to $i_1$. This $i_1$ flows through $R_2$, $R_1$ and $R_2$, the same current is flowing. So I get $V_0$ equals to minus $V_0$ by A minus $i_1$ into $R_2$, right.

And therefore, I get minus $V_0$ by A minus of $V_I$ plus $V_0$ by A divided by $R_1$ into $R_2$, just I have replaced $i_1$ by this one. If we do a small so gain which is the closed loop gain, I get $V_0$ by $V_I$ equals minus $R_2$ by $R_1$ divided by 1 plus 1 plus $R_2$ by $R_1$, right, $R_2$ by $R_1$ divided by A. This gives you the value of G. Now as A tends to infinity or very large value, this quantity goes to 0 and G tens to minus $R_2$ by $R_1$.

So you see, from this formula or from this definition I can achieve a infinite open loop gain if A is infinitely large, I get my closed loop, this is my closed loop gain. G is my closed loop gain. Closed loop gain, will be equal to minus $R_2$ by $R_1$, right. So the condition, therefore the condition is 1 plus $R_2$ by $R_1$, this should be as small as compared to A. When this is as, so A if it is very large you automatically get this into consideration, right. So this is the effect of finite loop gain, so I get $R_i$ equals to $R_1$ right.

(Refer Slide Time 26:30)



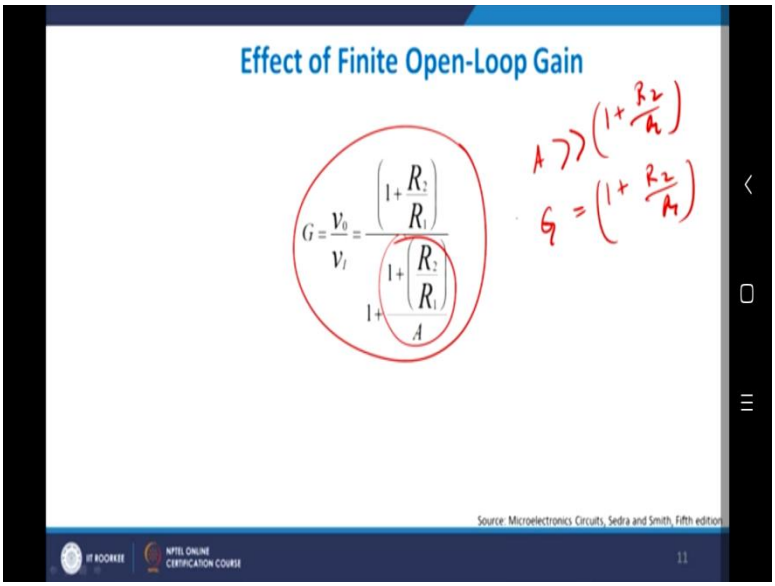

Let me come to the, if you understood inverting and non-inverting is just the just the just the complimentary of that, so non-inverting configuration will come. In non-inverting what we do, we apply the signal to the non-inverting mode, but we have the same configuration of the closed loop gain and $R_1$ and $R_2$ are here. So, if you

solve it, am not doing it on the class, am not doing in the module. If you solve it, I get $V_0$ by $V_i$ equal to 1 plus $R_2$ by $R_1$.

So what are you getting in the inverting mode was minus $R_2$ by $R_1$, here you are getting 1 plus $R_2$ by $R_1$, and with the positive sign. With the positive sign, primarily means that they are in phase. So this $V_i$ and this $V_o$, this $V_i$ and this $V_o$ will be in phase in all respects, right.

We can do small derivation to get these value of $V_o$ and $V_i$. They are very simple straight-forward. I would expect you to do it if you know what is a Kirchoff's law and how you know, you solve a Kirchoff's law. So I get 1 plus $R_2$ by $R_1$ as the gain for this case, right and that is what you, what the gain which you get in this case, right and that is what we get for all practical purposes.

(Refer Slide Time 27:25)

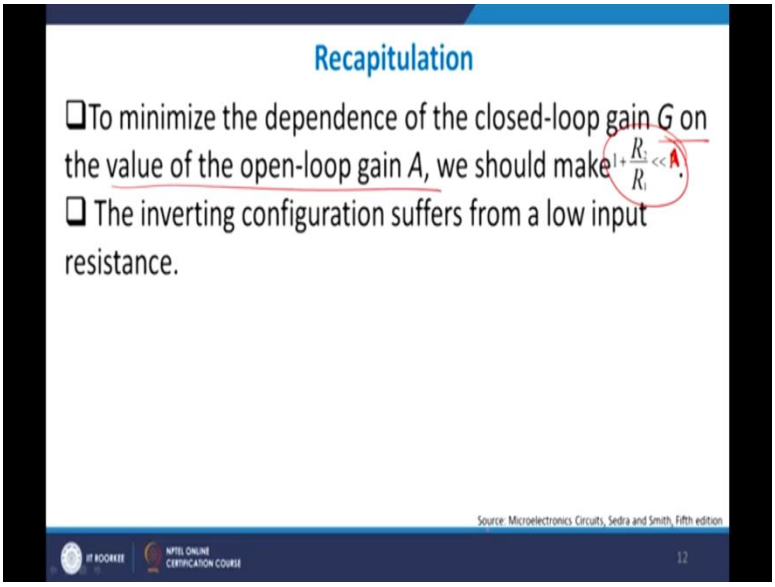



Again, if you take a finite open open loop gain in case of a non-inverting terminal, I get this into consideration that I get gain equals to $V_0$ by $V_I$, 1 plus $R_2$ by $R_1$ upon 1 plus $R_2$ by $R_1$ by A, 1 plus of this one. So in this case also if A is very large, sorry if A is very large as compared to 1 plus $R_2$ by $R_1$, then what you get is that that this automatically vanishes and I get open loop gain, closed loop gain to be equals to 1 plus $R_2$ by $R_1$.

So in both the cases, you do have, if your amplification factor is very high you can afford to have almost 0 gain in the closed loop configuration, right, and it only depends upon the resistors. The ratio of the resistors. So you change the value of $R_2$

and $R_1$ and you can get variable resistance, a variable gain with respect to your requirements, right. And that is what an important part is.

(Refer Slide Time 28:21)




So, let me recapitulate what we learnt from this small module which we discussed now, is that to minimize the dependence of closed loop gain on the value of open loop gain we should make 1 plus $R_2$ by $R_1$ much smaller than A actually, it should be smaller than A here. Much smaller than A right. And the inverting configuration suffers from low input impedance so inverting configuration has a problem, we did not discuss this but we will come to this in the next module. That they suffer from a low input resistances. So input resistances are pretty small in the case of inverting configurations, right.

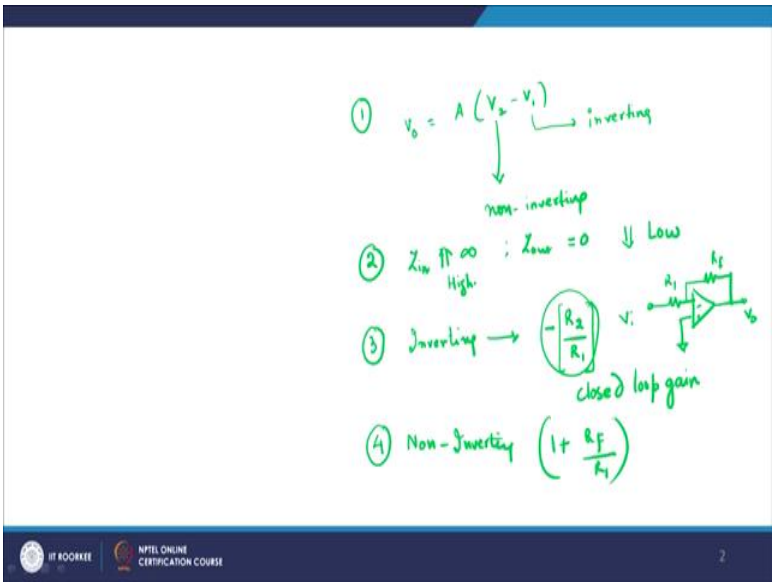
When we meet time, we will take care of voltage follower networks and then we will look into Op Amp as an integrator, as an differentiator, and we will look into as a waited summer. So all these 4 or 5 configurations we will finish in the next turn so that we are able to have Op Amp as a circuit element for the purpose of amplification, right and that is an important phase.

So this module takes care most of, this and the subsequent module will take care of the analog design and then we will revisit the combinational logical design at the last part of our module, right. Thanks a lot for your patient hearing and thanks a lot.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−51**
**Op-amp as an Integrator and Differentiator**

Hello and welcome again to the NPTEL Online certification Course on Microelectronics: Devices to CiRCuits. This module will be devoted to an application of an operational amplifier which is primarily an integrator and differentiator. In our previous lecture, we had seen that an operational amplifier is a difference amplifier which actually finds the difference between the 2 voltages applied to 2 nodes, inverting and noninverting node of an op-amp, multiplies that with the gain of the operational amplifier and put it in the output side. This is known as an open-loop gain. So, there is no loop, or is no feedback loop available to me and therefore I typically have my output voltage almost equals to A times $V_2$ minus $V_1$, right.

(Refer Slide Time: 01:31)



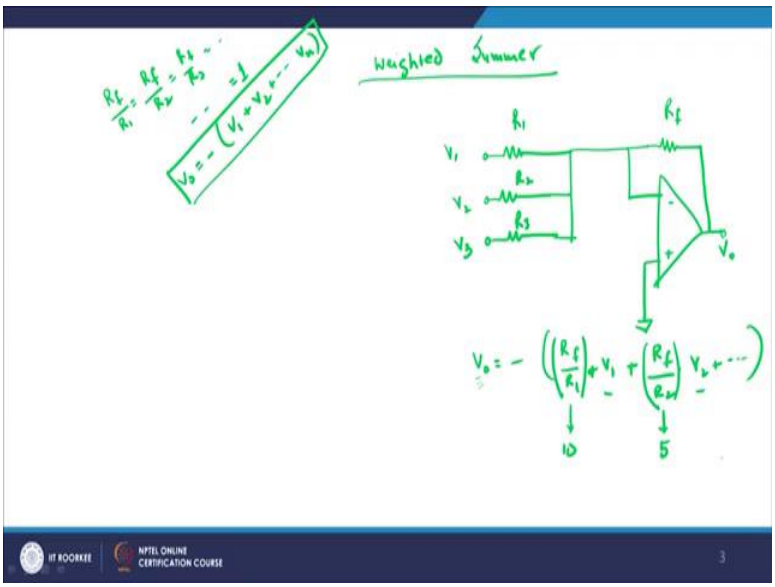

So, in our previous discussions we have seen that the output voltage $V_0$ right, output voltage $V_0$ is given as A times $V_2$ minus $V_1$, where $V_1$ is fed into the inverting terminal inverting terminal and $V_2$ is fed into a noninverting terminal, inverting terminal right. We also, So, this is the first property, the second property was that input impedance of an operational amplifier ideally is very high. Ideally it should be infinity but really it is very high and my output impedance is ideally equals to zero but in reality it is very low. So, I have got low output impedance and high input impedance.

This is the second important property of the operational amplifier which you saw in the previous discussion. We also saw that I can I can have an inverter or I can have an operational amplifier which can act as an inverting amplifier, right and we saw that the inverting amplifier can work with the gain of $R_2$ by $R_1$. Which is $R_2$ is the feedback resistance and $R_1$ is the input resistance. So, the operational amplifier looks something like this and you had $R_f$ here and you had $R_1$ here.

Now This was referred to as the closed loop, so this is basically your closed loop gain, closed loop gain, right. I have a closed-loop gain in which I have inverting mode, because it is minus sign. So, it is basically a phase change between input and output by 180 degree and it is minus sign showing it is a phase change and it is a ratio between 2 registers $R_2$ and $R_1$, right. And therefore, this is the inverting. We also studied, I think noninverting consideration non inverting in which we studied that it is basically 1 plus $R_f$ by $R_1$ and therefore this was the basically the gain which you which you got in the case of a noninverting case.

Now, we will take care of this one and we will go ahead and see what are the implications of having for example, a noninverting. So, can we have operational amplifier work as a voltage summer. So, if I want to sum n number of analog voltages, what is the best from of doing it and one of the best form is actually using operational amplifier as a summer.

(Refer Slide Time: 04:04)




So, let us just show to you what is known as a weighted summer So, it will be a summing of 3 or 4 analog voltages and it looks something like this. That is your $R_1$ here, right, and you have got $R_2$ and your $R_3$, all these 3, if you feed here then you get to negative side and then

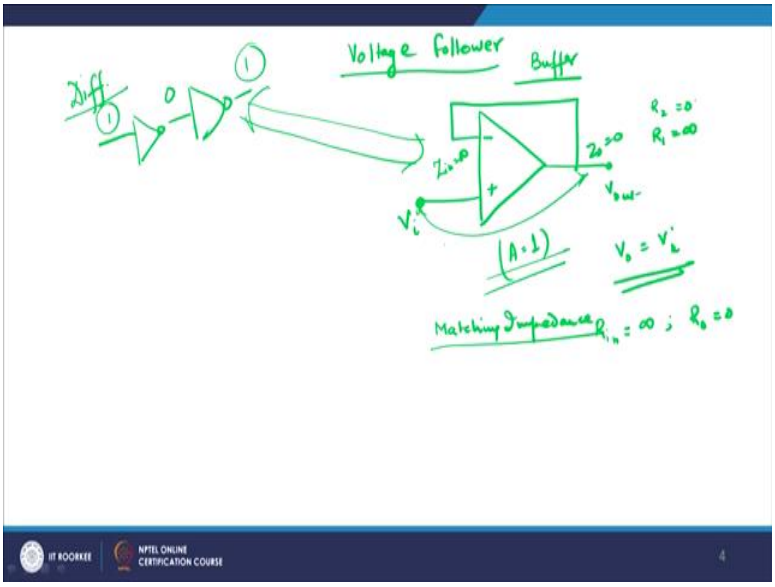
this becomes equals to $R_f$. This is your $R_f$, this you ground and this is $V_1$, $V_2$, $V_3$ and you got $R_1$, $R_2$, $R_3$.

With this knowledge, since this is in inverting mode, I get $V_0$ to be equals to minus $R_f$ by $R_1$ plus into $V_1$ right, plus $R_f$ by $R_2$ into $V_2$ plus so on and so forth, which means that the output voltage which is available at this point, right, is basically a sum of input voltage but weight weighted with respect to $R_f$ by $R_1$, $R_f$ by $R_2$, $R_f$ by $R_3$ and so on and so forth. Which, an interesting thing therefore is that if you want to just sum the voltages, you just have to keep $R_f$ by $R_1$ equals to $R_f$ by $R_2$ equals to $R_f$ by $R_3$ so on and so forth, all equals to 1.

If you sustain that I get $V_0$ equals to minus of $V_1$ plus $V_2$ so on and so forth till $V_n$. So, this is the simple summing of voltages across the board, right. This is a very straight forward way of looking at it. Another methodology or another interesting part of it is that, that is if you want to have this, suppose you want that the output voltage of voltage $V_1$ should be ten times.

So just keep $R_f$ by $R_1$ 10, you keep this one whatever you want to keep it, 5 and so on and so forth, then voltage $V_1$ will have a larger influence on the output voltage because you are applying principle of superposition here and all the voltage will be additive in nature. And therefore, you automatically get output voltage which is sum of all the voltages here. This is known as a weighted summer. So, you can you can sum analog voltages in this manner, right. So, if n number of analog voltages you can sum all the voltages in this manner.

(Refer Slide Time: 06:42)



Now, let me come to another one and that is basically a very simple one, that is known as a Voltage follower network and if you look very simplistically it is something like this, that you have got and Op-amp and this is given here, this is applying voltage $V_i$ and what I do is I simply short my input and output. So, the noninverting mode is shorted with respect to output $V_{out}$.

Under such a scenario, my $R_2$ obviously is zero and $R_1$ if you see is infinity because it is input impedance of the operational amplifier as infinity. This is basically, so so what I am trying to tell you is since $R_2$ equals to zero the closed loop gain A equals to 1. So therefore, this type of configuration gives you a voltage gain of approximate equals to 1. However, in this $V_{out}$ therefore follows $V_i$, because gain equals to 1, you are not doing any change in the gain therefore $V_o$ equals to $V_i$ and therefore your $R_{in}$ is equals to infinity and we have discussed this point and $R_o$ equals to zero.

$R_{in}$ equals to infinity means input impedance is relatively high or very high, which again means that no current will be flowing through the operational amplifier, only a large change in the voltage will visible to you. But that will not entertain any current whereas since $R_0$ equals to zero ideally, it primarily means that I can have any current flowing through this but the voltage at output point will be relatively small in dimensions. So, this is a voltage follower.

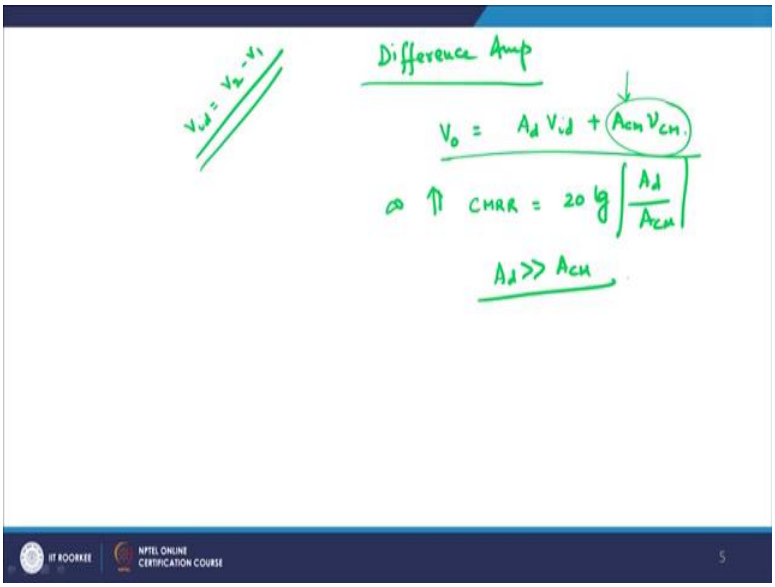
So, when do you use a voltage follower? Because its gain is 1 and therefore you generally do not use it for high gain applications. But, you use it for, you use it as a buffer ciRCuit, right. So, it is basically sort of an analog buffer which you see. If you remember we had studied digital buffer in our other modules. When we go for digital buffers, we use for example, a static inverter series. So, it is 1 zero 1.

So, whatever input you are giving here, output is also 1. So, so and you remember in a CMOS inverter, static inverter, your input $Z_{in}$ is very high and $Z_{out}$ is very low. And exactly the same thing happens in an Op-amp that here $Z_{in}$ is infinitely large as I discussed with you. And $Z_{out}$ here is very very low. It is almost analog. So CMOS inverter is analogs to a voltage follower in a sense.

And you can use this voltage follower for all your other issues for example, used for the purpose of, for matching the impedances. So, primarily voltage followers are used for matching impedances. And for transfer, so what it implies is that it allows you to transfer

large amount of power from $V_{in}$ to $V_{out}$, right. Because your impedances are perfectly matched and you have a large power pin transferred from input to output. So, this was important aims of a differential amplifier.

(Refer Slide Time: 09:59)



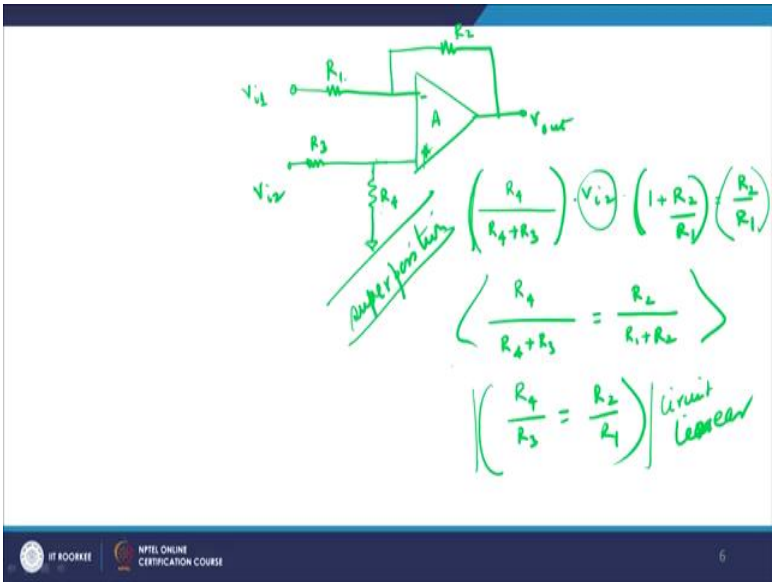

We come to the last part here and the last part of the ciRCuit configuration and that is known as Difference Amplifier. Difference Amplifier, right. What does difference amplifier do? That if you I can write an expression, that output voltage is equals to $A_d$ right into $V_{id}$ plus $A_{cm}$ into $V_{cm}$. I will explain to you what do I mean by these terms. Well, $V_{id}$ is nothing but the difference of $V_2$ minus $V_1$. This is $V_{id}$, right. So that if you multiply with Ad, Ad is the differential gain.

Differential gain means that part of the gain which actually tries to amplify the differential signals across the 2 terminals of the operational amplifier. The next part is $A_{cm}$ into $V_{cm}$, where $A_{cm}$ is basically the voltage gain, right, for the common mode signal. Common mode signals are those signals which are common to both the inputs of your operational amplifier. Now, typically we define, as I discussed with you earlier also, CMRR is equals to 20 log of $A_d$ by $A_{cm}$, right. Mode value of that, right. That is how you define your CMRR.

And the CMRR ideally is infinitely high or very high, very large value, which primarily means that $A_d$ has to be much much larger as compared to $A_{cm}$. And makes sense also, because see the differential gain if it is large, then any differential input will be amplified by the operational amplifier. Whereas, any common mode input will be suppressed by the

operational amplifier. And therefore, any Op-amp is a very good selector or a rejecter of noise, right.

(Refer Slide Time: 11:47)




Now, let me come and show to you what is a how does a difference amplifier looks like or what is the configuration of difference amplifier. So, I have got a negative terminal here and this is my noninverting terminal and this is my $R_3$, right. So, I have a voltage divided network at this particular point and this is $V_{i2}$ and there I got a $V_{i1}$ and we have got $R_1$ here, right, and then exactly the same as an inverting terminal $R_2$ and this is $R_1$ and this is $V_{out}$, right.

So, so if you if you if you see here, this is $R_1$, $R_2$ and this is $R_3$ and $R_4$ and this is operational amplifier with voltage gain of A. So, I can write down in this case, I can write down $R_4$ upon $R_4$ plus $R_3$, right. Why $R_4$ upon $R_4$ plus $R_3$? Because, if you see very carefully between these 2 points, these 2 are in series and so it is $R_4$ plus $R_3$ is the total resistance and the voltage across $R_4$ appears into the noninverting terminal of A and therefore $R_4$ upon $R_4$ plus $R_3$ into $V_{i2}$ is the voltage at this particular point, right, multiplied by multiplied by 1 plus $R_2$ by $R_1$ right, and this this must be equals to $R_2$ by $R_1$.

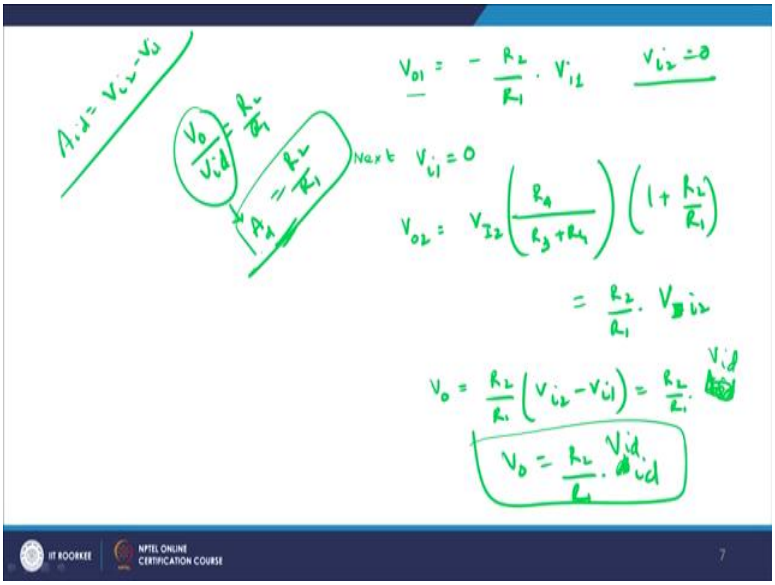
Why it should be equals to $R_2$ by $R_1$? Because, you see primarily if I assume that $R_4$ upon, because you are taking the voltage across $R_4$ and feeding it into the noninverting terminal of the amplifier. It primarily means that out of some value of voltage $V_{i2}$ only $R_4$ upon $R_4$ plus $R_3$ is available to you at A, right. Similarly, why do you multiply that with 1 plus $R_2$ by $R_1$? Because, that is nothing but the gain of this stage. So, I multiply these two together and I get

$R_2$ by $R_1$ coming into picture here, must be equal to $R_2$ by $R_1$ which is the closed loop gain of this system.

So, what I get is that $R_4$ upon $R_4$ plus $R_3$ must be equals to $R_2$ upon $R_1$ plus $R_2$. You can solve it and get it for all possible reason. Now, this 2 will be equal provided I make $R_4$ by $R_3$ equals to $R_2$ by $R_1$, right. $R_4$ by $R_3$ is equals to $R_2$ by $R_1$. Once you sustain these 2 inequalities, then the ciRCuit is linear and you automatically get, so this will imply that the ciRCuit is linear and as a result you can apply principle of superposition, right. So, you can apply principle of superposition.

Now, please understand this is the ratio of the registers, right. so it is basically $R_4$ by $R_3$ and $R_2$ by $R_1$, which basically means that if this is true, your Op-amp is balanced and your ciRCuit is considered to be linear for all practical reasons. now with this knowledge, let me show to you that what will be the value of a differential gain.

(Refer Slide Time: 15:13)




So, $V_{o1}$ equals to minus $R_2$ by $R_1$ into $V_{i1}$,right,because $V_{o1}$ is the output voltage because of input voltage. Assuming that $V_{i2}$ equals to zero. So, $V_{i2}$ equals to zero that is how you do how do you apply principle of superposition, right. How do you do that? You take If there are n number of signals you take one signal and try to give an input, all other signals are made zero. Similarly, now you take the second signal first and all the other signals are made to zero and then output you go on calculating for each of the signals.

Once the output is available to you, you simply have to add those outputs and for those outputs you will get the input, for these inputs you will get those outputs. Which means that if

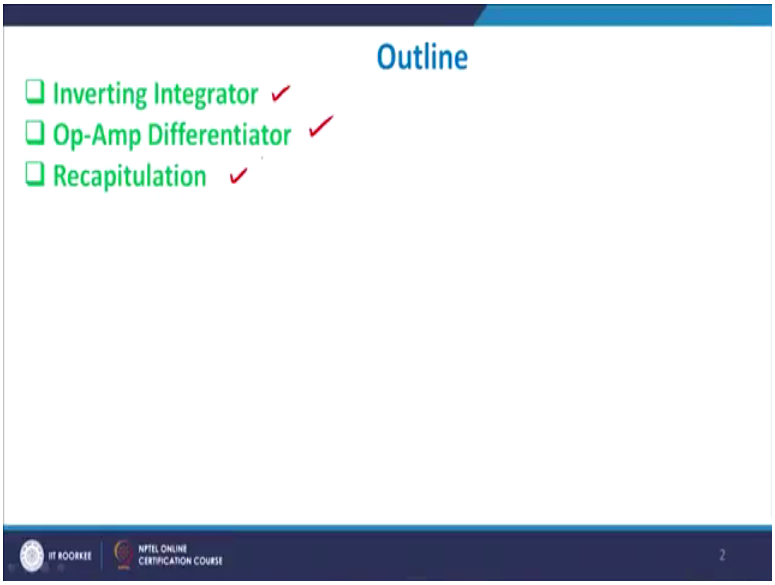
all these n number of inputs work together the output will be the sum of the outputs and this can only happen provided superposition principle is not violated under any ciRCumstances in the ciRCuit which you are using. So, I get $V_{01}$ equals to $R_2$ by $R_1$ into $V_{i1}$.

Next if I take $V_{i1}$ equals to zero, I get $V_{o2}$ to be equals to $V_{i2}$ into $R_4$ by $R_3$ plus $R_4$ remember. Because $V_{i2}$ is the input voltage multiplied by the voltage divided network into 1 plus $R_2$ by $R_1$ which is the gain of the system. This must be equals to $R_2$ by $R_1$ multiplied by $V_{i2}$. $V_{i2}$. $V_{i2}$ is the input voltage multiplied by $R_2$ by $R_1$ because that is in the noninverting mode and therefore you get this. Sorry, inverting mode and therefore you get $R_2$ by $R_1$.

So, if you solve it I get $V_0$ equals to $R_2$ by $R_1$, right, and $V_{i2}$ minus $V_{i1}$ which is also equals to $R_2$ by $R_1$ into $A_{id}$, right. So, $V_0$ equals to $R_2$ by $R_1$ into $A_{id}$, right. Where $A_{id}$, What is $A_{id}$? Aid is nothing but $V_{i2}$ minus $V_{i1}$, right. And therefore I get, therefore if you look very carefully, I get sorry this is $V_{id}$, $V_0$ equals to $R_2$ by $R_1$ into $V_{id}$. This will be $R_2$ by $R_1$ into $V_{id}$. So, I get $V_{id}$. So I get $V_0$ by $V_{id}$ equals to $R_2$ by $R_1$.

Now, $V_0$ by $V_{id}$ is nothing but your $V_0$ by $V_{id}$ output voltage with input voltage is a gain, closed loop gain and that is equals to $R_2$ by $R_1$, right. Also referred to as differential gain, fine. So, your differential gains are just the ratio of $R_2$ by $R_1$ in case of a difference amplifier, right. And that is what you get out of difference amplifier design or issues available to you.

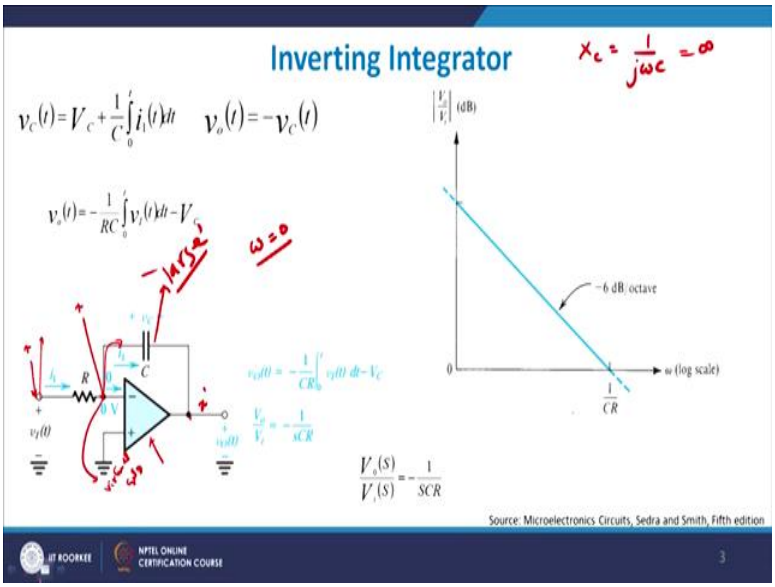
(Refer Slide Time: 18:19)




Now, let me come to, let me explain to you the, the so, let me come to now the topic itself. Inverting Integrator, Differentiator and then we will recapitulate what we have, what has been thought to you. So, we will take up first of all inverting integrator, op-amp differentiator and

so on and so forth. Let me explain it to you, how it works out, if you see the diagram looks something like this as shown in the presentation, that you have an operational amplifier here with again inverting and noninverting terminal.

(Refer Slide Time: 18: 43)



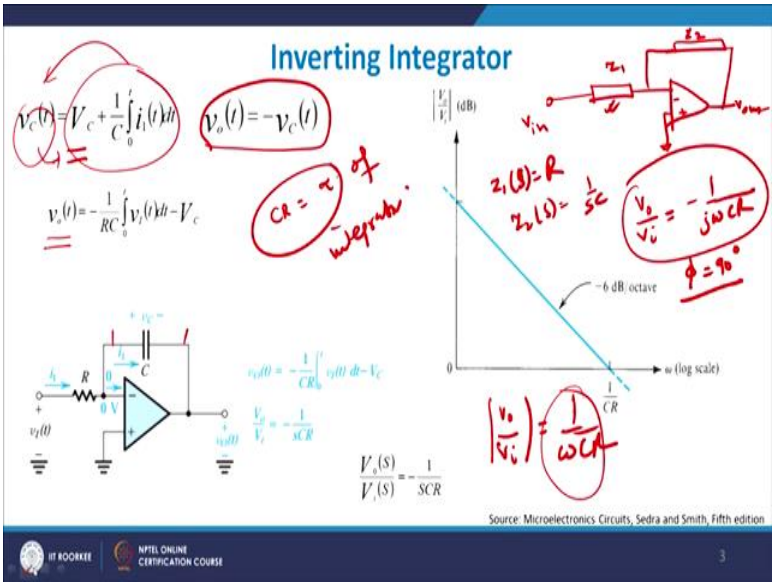

And you feed your input to the inverting terminal and you have a capacitance in the feedback loop, right. Now, quite interestingly at ω equals to zero, this capacitor will be open ciRCuited and there will be no feedback. As the ω increases, the feedback quantity also starts to increase, right. And therefore, with rising gain or rising frequency your gain will start to fall down. This is without even doing any basic mathematics or understanding it.

We can do it by simple methodology that when ω equals to zero, $X_C$ which is basically the capacitive reactance which is given as jωC, that comes to be infinitely large, right. So, this is infinitely large. Now, when you apply a voltage here, now you apply a voltage here this voltage will appear at this point. Suppose you apply voltage x here and the same voltage will appear across here approximately as x. And the same voltage in reality will also appear here as some value, say x'.

Now, as I discussed with you in our previous module, that this is basically your virtual ground. Virtual ground primarily means that ground where though the voltage is zero but the currents are not zero, right. So, where the current is flowing? Current is coming from here to here. A large amount of current is flowing. So, whatever voltage you are giving at the input here, here write all the, drop that voltage by R is all the current is flowing through this

capacitor C. With this knowledge let me just show to you, how does one how does one formulate a policy for inverting differentiator.

(Refer Slide Time: 20: 36)




Now, you see $V_c(t)$ is nothing but the voltage across the capacitor t, with respect to time t. So, the time time dependent voltage across capacitor C must be equals to $V_c$ which is the initial value of the voltage across the capacitor plus 1 by C integral zero to i, zero to t $i_1$ dt where $i_1$ is the current flowing in the capacitor. The standard differential equation, the standard equation which we have used based on the physics of the device and we also know that $V_0(t)$ will be minus of $V_c(t)$. Right. And the reason being here connecting a capacitor to the inverting terminal of your op-amp.

And as the result, the output will be a negative part of the input, right. So, if you want to find out the $V_0$ value, right $V_0$ value, so what you do is you will have a negative sign minus RC here integral zero to t $V_1(t)$ dt minus $V_c$, right. So, minus $V_c$ will come on this side and minus RC by, minus 1 by RC will come here. How did I do that? Well simple, you just take this to the left hand side right, and take this to the right hand side, so I get $V_c$ is equals to minus $V_c$ so I get minus , so I get $V_0(t)$ is equal to 1 by RC $V_1(t)$, right.

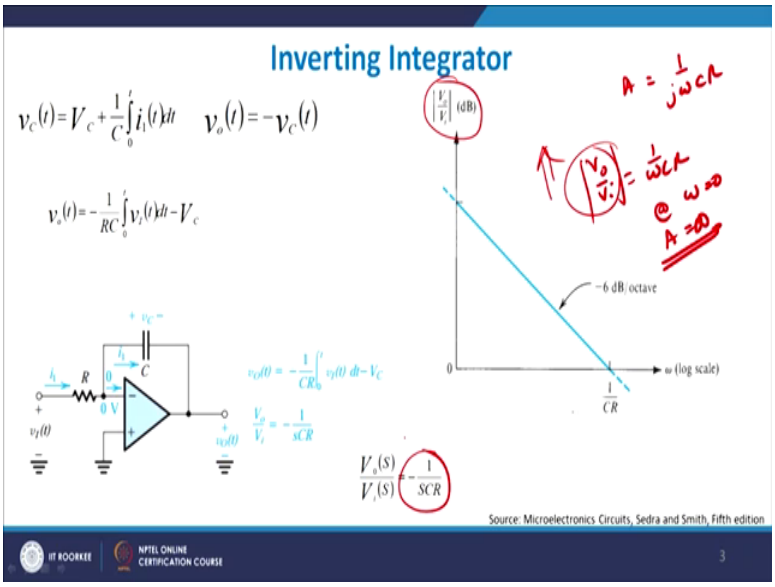
Now, this CR is referred to as τ of integrator, right. $V_0$ is defined as the CR time constant of the integrator. Now, if I do a more generalized formulation then I can safely use it like this. I have got, say this is your $Z_2$ and this is your $Z_1$, right. And this is your output $V_{out}$. Let us suppose your positive terminal which is the noninverting terminal is grounded. Under such a

direct scenario I can write down my my $Z_1$ $Z_1S$ will be equals to R and $Z_2S$ will be equals to 1 by SC.

So, if you place it in our initial equations, I will get $V_0$ by $V_i$ to be equals to minus 1 by jωRC. So, I get the open loop gain to be equals to minus 1 by jωRC. Now, with the phase ofcourse of 90 degree. I am not interested in phase at this stage. So, if you want to find out the mod value of $V_0$ by $V_i$, right, I get this, I will get to be equals to 1 by ωCR. So, j square will be equals to minus 1 and that minus will cancel with minus and I get 1 upon ω square C square R square root over that comes out to be 1 by ωCR, fine. And that is the value of your output the gain that you see in front of you.

Therefore, if you just unable to see this, let me just see how it works out in terms of plotting a Bode Plot, right.

(Refer Slide Time: 23:52)

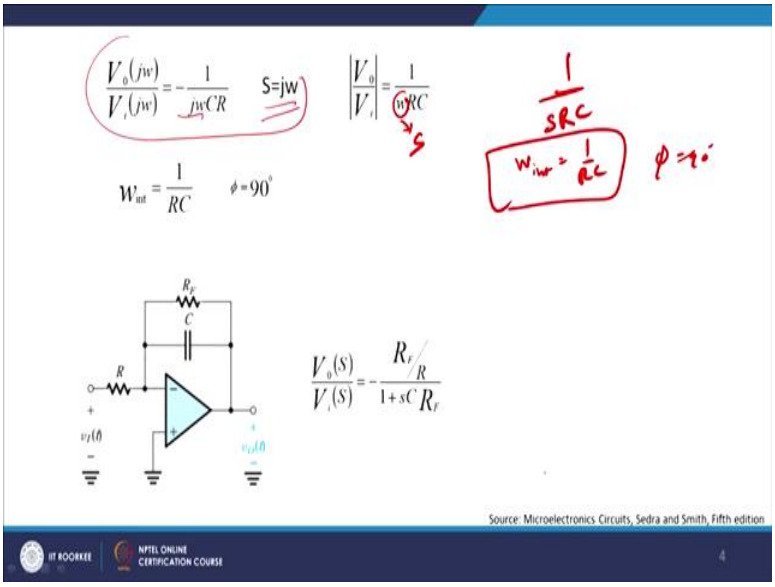



So, if you look here, this is basically $V_0$ by $V_i$ in dB with respect to ω which is in log scale. Now, at by my previous discussion or definition we saw that my ω, ω is, so my gain is is basically 1 by jωCR and if you do all sorts of manipulation I get 1 by ωCR to be equals to $V_0$ by $V_i$. So, at ω equals to zero, right, your your gain should be equals to infinity, closed loop gain should be equals to infinity, right, because your ω equals to zero. But, that means the denominator is equals to zero and therefore this shoots upto infinity.

This is not true and there are certain reasons for that because operational amplifier works under DC bias which is given to the operational amplifiers and any voltage larger than that will be clipped right away, right. And therefore you will not get such type of variations in

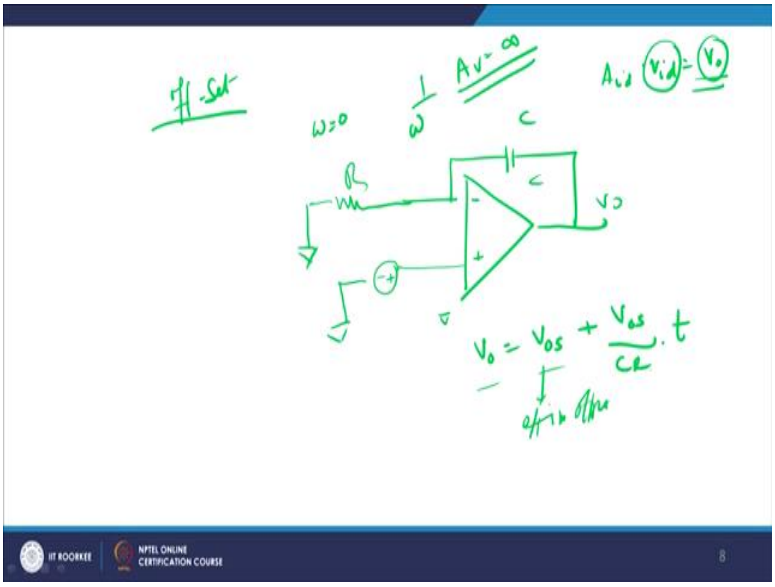
inverting amplifier. So, you see as I discussed with you I get 1 by SCR with a negative sign as the value of the output voltage with respect to the input voltage.

(Refer Slide Time: 25:00)




If you see, I told you 1 by jω SCR, S can be written as jω, so I get 1 by ωRC or you can write down this to be as equals to SRC. So, 1 by small SRC as your value $V_o$ by $V_i$ and similarly $\omega_{int}$ is equals to 1 by RC. 1 by RC is basically the time constant for the circuitry with phase equals to 90 degree ofcourse.

(Refer Slide Time: 25:36)




Let us see what happens otherwise, what is the problem otherwise, let me say that you do have a problem that there is a off-set voltage, which means that, so what is the meaning that,

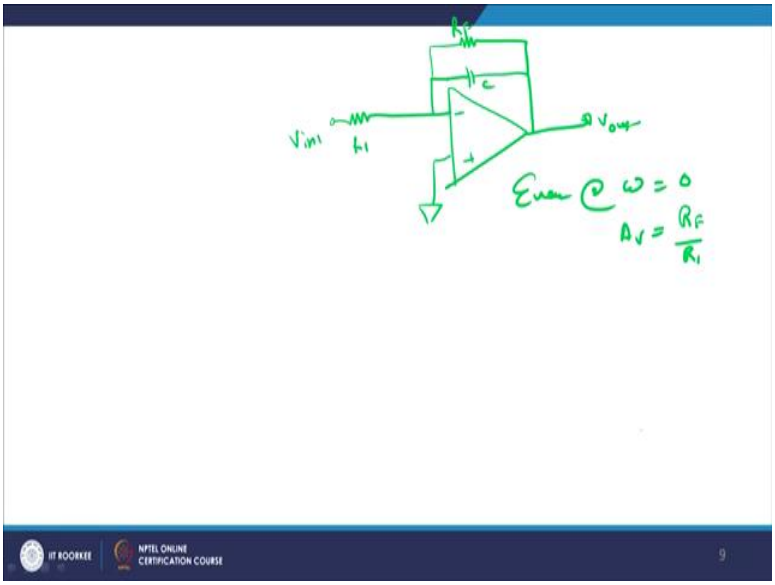
it primarily means that if the differential voltage Vid between the 2 terminals of op-amp is zero the output should be exactly equals to zero. Because, A into $V_{id}$ is equals to output is equals to $V_{out}$. Ofcourse the common mode signal is there but this is the output voltage.

Which means that, when $V_{id}$ is equals to zero, $V_o$ should be equal to zero. In reality, not true, and that is what is known as a op-amp off-set. You have an off-set in op-amp, meaning that howsoever best your op-amp is, there is some finite DC off-set and that results in this chain, first thing. The second thing is, remember my previous discussion I had ω in the denominator, which means that at ω equals to zero my gain should be equals to infinity, right.

But this is not true, we very well know that any gain cannot be equals to infinity. So, how to go about doing it. So, ok, so, we will do something like this, that means if ω equals to zero, $X_c$ opens and your feedback loop closes, there is no feedback available to you. And therefore when there is no feedback your gain will not be stabilized. That was the major reason why we went for a stability principle.

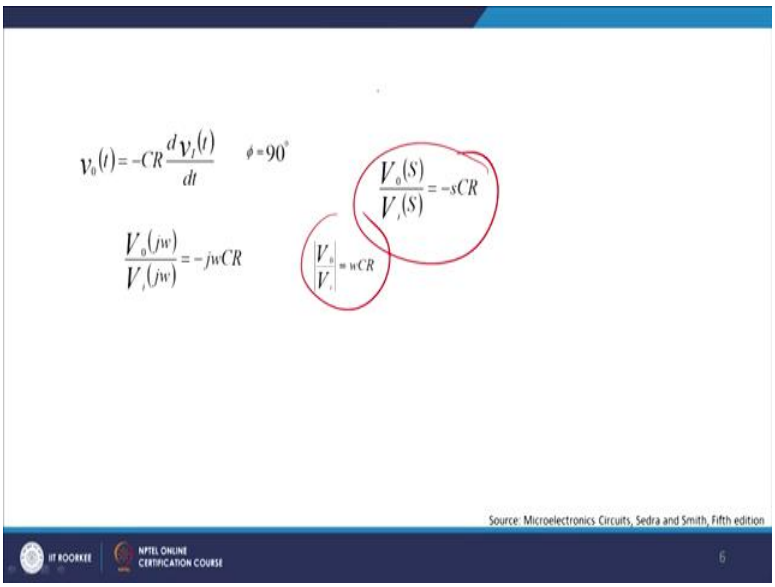
To remove that, I will show you what is the DC off-set therefore, so if I have got a negative and a positive one, right, and I have got a C capacitance here, which is $V_0$ and this is C and this is R, and this is grounded again, right, and this is grounded and this is C. Now, if you go back to this $V_{out}$ will be equals to $V_{os}$ plus $V_{os}$ by C into R into t, right, which means that $V_{os}$ is nothing but the output off-set voltage. $V_{os}$ is the input off-set voltage. So, so it is something like this that you do not you do not you do not, I will just show it to you, so you what you do is you do not let it apply here and then you have voltage source and then goes to zero and this is what you what you get. So, I get $V_0$ equals to $V_{os}$ plus $V_{os}$ by CR into t, right. Into t.

(Refer Slide Time: 28:04)




Now, what people did over the years was that, this was quite an interesting addition to the basic amplifier configuration, that if you are able to have capacitance here, right, which is $V_{out}$ and this is positive $R_1$ and this $V_{in1}$, then if you are able to sustain or have a feedback resistance here equals to $R_f$, right, then even at ω equals to zero you will have a finite gain given by $R_f$ by $R_1$. So, what people did was that even at ω equals to zero my $A_v$ will be equals to $R_f$ by $R_1$, right. And That is quite critical, which means that there is no concept of an infinite gain as such, right, and I have this $R_f$ into consideration.

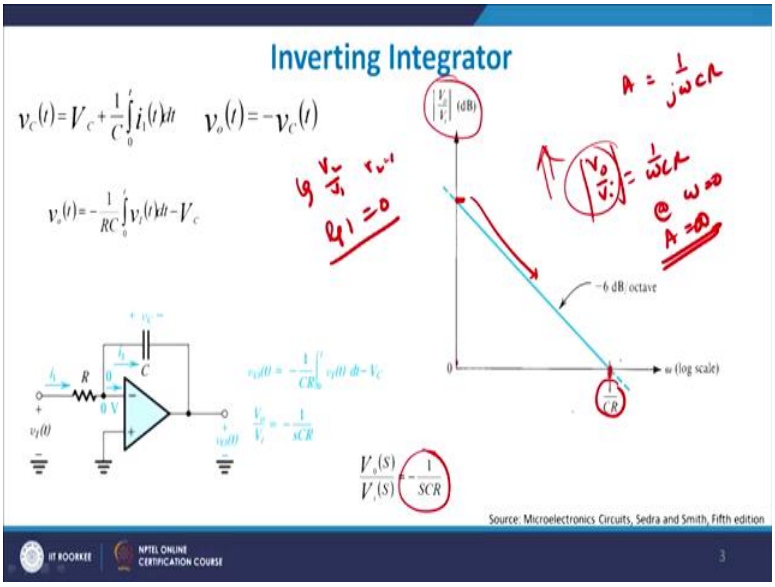
(Refer Slide Time: 29: 04)




Now, with this knowledge or with this idea, I can I can next come to the, in here that is given by minus SCR and therefore it is ωCR as your output impedance in this case. So, the

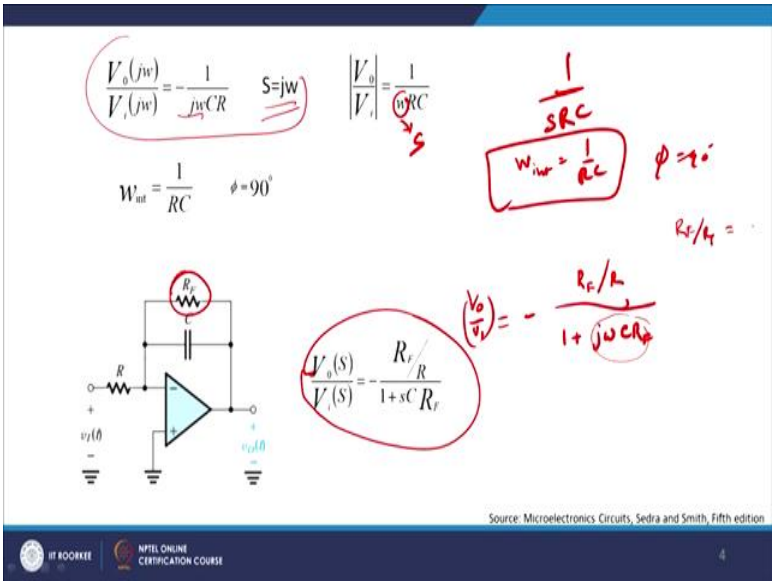
integrator starts to behave like a low-pass filter with a corner frequency equals to zero. Why? Because, you see, if you look at. Sorry.

(Refer Slide Time: 29:27)



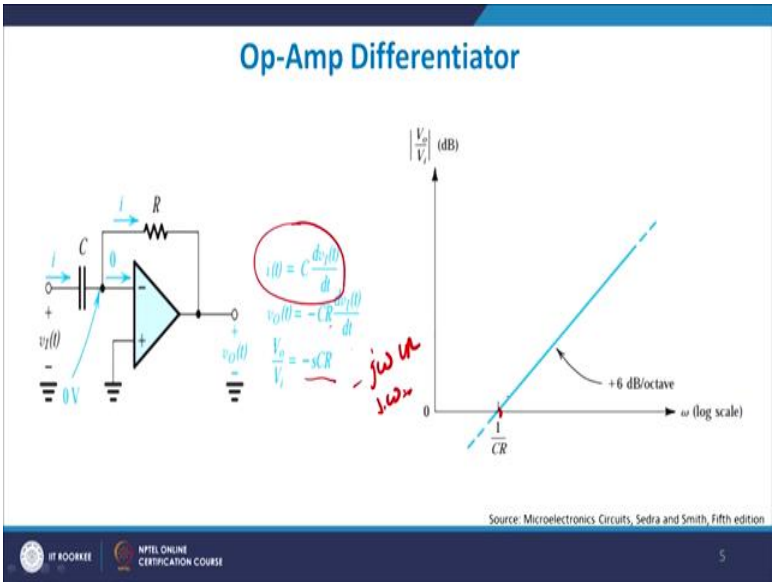

If you go back, yes, if you look at ω equals to zero, I get this part as the ω starts to increase there is a linear drop with approximately 20 dB per decade or 6dB per octave drop in the gain, right, and where it cuts the ω axis we define that to be as a unity gain point. Because, log $V_2$ by $V_1$, when $V_2$ equals to $V_1$ I get log 1 which is equals to zero. And therefore, this is the point where you get unity gain or the gain is unity, right. If you go below this or the negative side of the dB what will have is that the phase margin will change and all these things will happen.

(Refer Slide Time: 30:10)




Okay, so that is the reason, with with this feedback resistance, this is the new value of your new value of your of your transfer function or the gain function which is minus $R_f$ by R 1 plus SC, so I get $R_f$ by by R into 1 plus S can be written as $j\omega CR_f$, this must be with a negative sign given to $V_0$ by $V_1$, right. And therefore, at ω, now if you see at ω equals to zero only this quantity goes to zero and you still have $R_f$ by $R_1$ as the output voltage, variation of the output voltage.

(Refer Slide Time: 30:49)

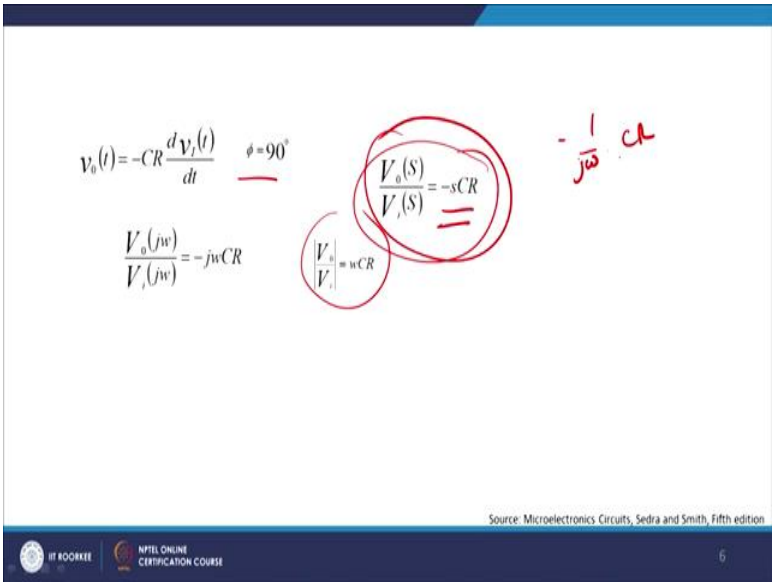



Let me come to the differentiator part, Op-Amp differentiator and if you see differentiator we have just switched to the place of capacitor and register here, so capacitor is here and register is here, so I get i equals to $C(dV_i/dt)$ differential equation. So, if I take differential signal and

we try to evaluate, I get $V_0(t)$ equals to minus $CR(dV_i/dt)$. Now, now therefore $V_0$ by $V_i$ will be equals to minus SRC, right. Minus SRC will be the value of $V_0$ by $V_i$. Now, this is the gain, so if you see at S it is jωRC with a negative sign.

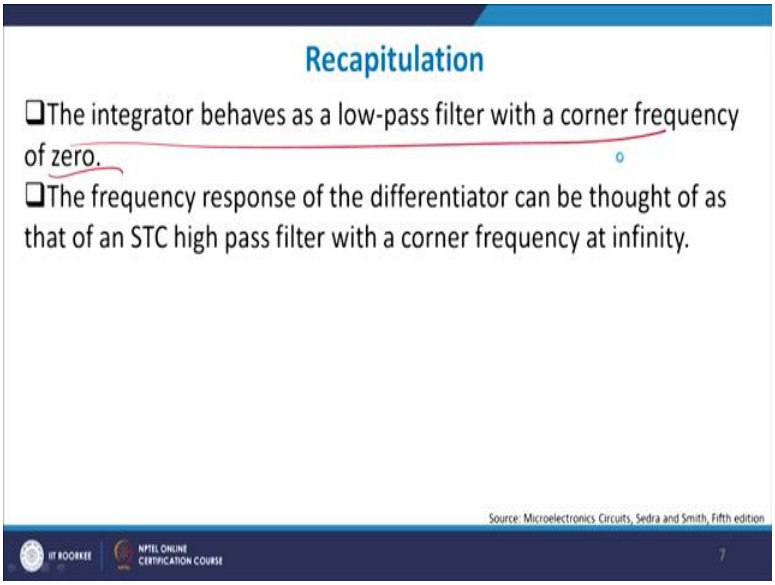
So, with this case ω equals to zero, I get this thing, so in a case of op-amp differentiator even at ω, at gain unity gain you will have your ω equals to 1 by C into R, right. So, you not gaining anything but you are actually trying to imply that at ω at certain ω equals to ω maybe t or ω b which is basically the frequency, 3dB bandwidth frequency or unity gain frequency, sorry. I get this much as amount of my output gain, right, and that is what one needs to find out or why.

(Refer Slide Time: 32:07)




Exactly, the same thing as I did it in the previous turn, ϕ equals to 90 degree, I get output voltage $V_0$ by $V_R$ equals to minus s times CR, right. So, minus S is minus 1 by jω into C into R, right and that what gives you the value of your output voltage.

(Refer Slide Time: 32: 18)



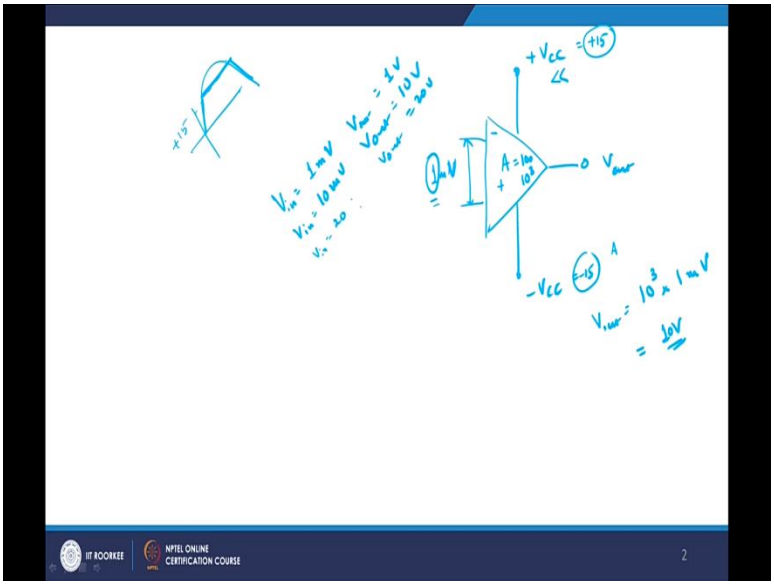

So, what we have what we have done till now, we have seen that an integrator can actually behave like a like a  like a low pass filter with a corner frequency equals to zero because that is what it was leading to. And the frequency of the differentiator can be thought of as a that of an STC means high-pass filter, right. Your time constants, your unity gain time constants, high-pass filters, with a corner frequency at infinity, right. So, differentiator will have corner frequency equals to infinity whereas integrator will have a corner frequency of zero in ideal conditions.

However, this is not true and we will discuss all these things as we move along. But, this is the primarily 2 major portions of operational amplifier, differentiator and integrator. In the next, I will start with the numerical example and see how it works out in case of differentiator amplifier, right. Thank you very much for your patience hearing.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communications Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−52**
**Large Signal Operation of Op-amp and Second Order Effects**

Hello and welcome to the next edition of NPTEL online course on Microelectronics, devices to circuits. We will start today's module on last signal operation of operational amplifiers and second order effects. The reason for doing this module was that typically as of now we are assuming that the input to the signal or input signal to the operational amplifier's peak to peak values are relatively small and therefore even when you multiplied that with the open loop gain of the operational amplifier, it never crosses the values of plus $V_{cc}$ and minus $V_{ee}$.

(Refer Slide Time: 01:59)




Remember in my previous discussion which we had during the starting of the module on Operational amplifiers, you will understand or you will notice that whenever you are dealing with Op Amps, your op amps are primarily having a negative and a positive terminal and you also had plus $V_{cc}$ and minus $V_{cc}$ or minus $V_{ee}$ and you had one output terminal here. Now if you multiply typically A, say suppose this is equals to plus 15 and this is equals to minus 15 and if the difference between the two signals is say 1 milli amp, 1 milli volt let us suppose and A is equals to 100 or maybe even 10 to the power 3, so I get output $V_{out}$ to be equals to 10 to the

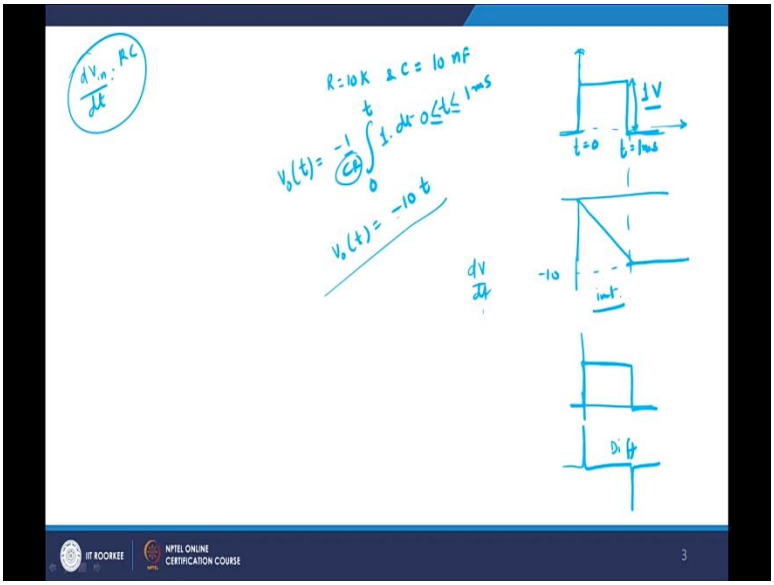
power 3 into 1 milli volt. So that becomes equals to 1 volt, now 1 volt is much smaller as compared to $V_{cc}$, right.

Similarly, if this now comes out to be 10 milli volt then I get 10 volt as the output, so if this is, so $V_{in}$ is 1 milli volt, I get $V_{out}$ to be equals to 1 volt. If $V_{in}$ equals to 10 milli volt, I get $V_{out}$ to be equals to 10 volt and so on and so forth. Now therefore if this Vin is let us suppose say 20 then my $V_{out}$ should be equal to 20 volts but that exceeds this value here and here which means that my output waveform which I will get will be much larger than plus 15 which is being said by the power supply.

So I am basically allowing the device to trip and therefore there will be the output waveform will be clipped at plus 15. Now what we will be studying in this module is that thing only that given a possibility that you do not, you let the whole system to saturate and you allow the voltage to go beyond the permissible values then how does the operational amplifier behaves. So that will be the major goal of this study.

Before that let me give you a small problem and you might try to solve it, the methodology I can give it to you. Say I have an input, this is regarding an integrator right. So I have an integrator in which the input waveform, this is t is equals to 0 and this is t equals to 1 millisecond and this is equals to 1 volt. So in an integrator I am giving a square wave, just like as input waveform whose peak value is 1 volt and whose width, pulse width is 1 millisecond and you need to calculate how it looks like, so how it looks like we need to calculate.

(Refer Slide Time: 03:55)




Let me draw for you, so before that if the resistance is given to be as 10K and C is given to be as 10 nano farads. So I get V0(t), right is equal to 1 by CR integral 0 to t 1 dt where t is between 1 millisecond and 0. So if I solve it I get V0(t) to e equals to minus 10 t by putting the values of C and R. So if you look very carefully this is basically a negative sign with a 10, so it will be a ramp so and the voltage goes like this, this will be typically a ramp here, ramp till how much, minus 10 volts and after this it will be a straight line once again right, there will be 0, sorry it will go to, it will go to 0 here.

Why it will go to 0 here because see here you are integrating, you are integrating it right? You are integrating it, so when the, so this is what you get as far as integration is concerned. This is how it looks like right, so integration of a constant value will give you a constant output and that is the reason we will get it, whereas if you look at the differentiator for example, differentiator is basically dVin/dt into RC, well that is how you look into the differentiator which means that in a differentiator if you get a constant. For example, the same profile if I ever got, this is the input then the rising edge right, rising edge my dV/dt would have been very large, so I would have got a spike here, but when my voltage goes to constant, then constant differential is always 0 and therefore this will, I will just see a spike here and then I will since it is coming down I will see a spike here.

So a differentiator looks something like this and integrator looks something like this. So this is a pretty wonderful method, differentiators are pretty wonderful methods for generating spikes out of given waveforms but differentiators are very seldom used because they influence large amount, they insert large amount of noise because of electromagnetic interference and so on and so forth whereas integrators do not do that. So just for information's sake this is in consonance or in continuation of my earlier discussion as far as this course was, as far as this course was concerned.

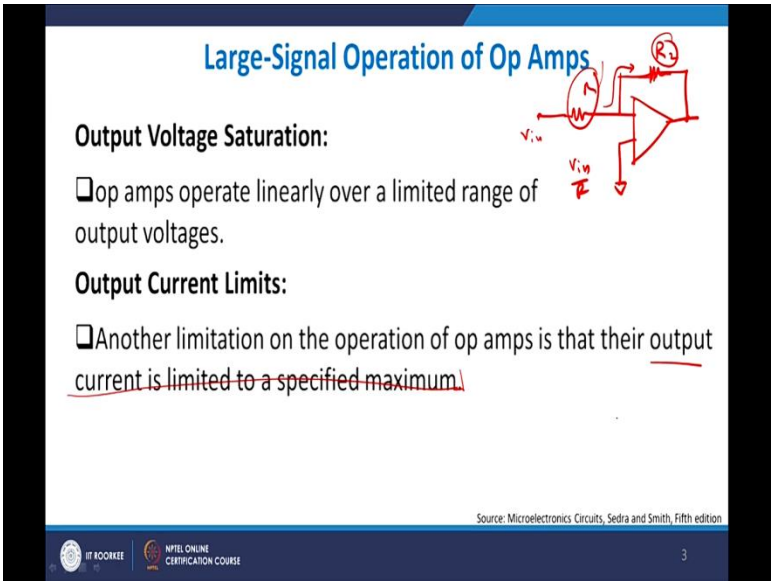
(Refer Slide Time: 6:37)


Let me come to therefore the next part here and let me come to the large signal operation and let us go step by step what we will be doing here. So we will be looking into large signal operation of op-amp, we understand what is a slew rate, why is it important as far as designing is concerned, in operational amplifier design is concerned. We will also look into one important point that we are assuming till now that your loop, your gain is infinitely large and your bandwidths are also large but in reality your bandwidths are always restricted bandwidths right, and gains are also restricted. And therefore under such a scenario how will your open loop gain change and how will your input impedance and your output impedance change.

We will also look into Frequency Response, Gain Bandwidth Product, the concept of Offset Voltage and the concept of Offset Compensation Voltage and then a Bias Current, right. We will

take one of them individually and understand, try to understand the basic features of these individual voltages or understand the basic concept, right.

(Refer Slide Time: 07:20)




You see we had just now discussed that an operational amplifier if you give an input voltage, my output voltage will be as I discussed with you, difference of the two voltage multiplied by a which means that if the difference remains the same right, that my output voltage will be independent on the both the voltages. I hope you understand this point, because A into $V_2$ minus $V_1$. As long as $V_2$ minus $V_1$ is held constant and I do not mind and my $V_0$ is constant, output voltage is constant.

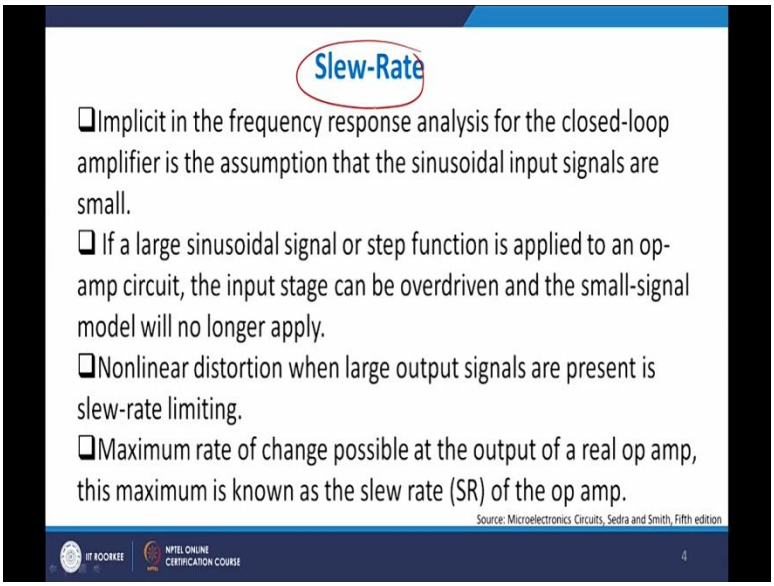
This is basically sort of a linear property of the device which primarily means that in such a scenario your amplification is independent of the input voltages but then operational amplifiers remain linear for a limited range of output voltages right and what happens is that as you reach towards plus $V_{cc}$ and minus $V_{cc}$ in the output side, you try to drive or or the signal tries to drive the operational amplifier into saturation, right. As it enters into saturation, there is a problem and the problem is non-linearity comes into picture to a larger extent.

Now another limitation of the operation of operational amplifier is that the output current is limited by the maximum specified value which means what, an operational amplifier remember, output current is given by even if you take an integrator or a differentiator. So if you see it will be resistance here and the capacitance here right, or maybe the inverting one will be much easier

to understand so I have got $R_2$ and $R_1$ here, as I discussed with you suppose this is $V_{in}$ then $V_{in}$ by R is the current flowing through this resistor but since this is a virtual ground I would expect to see all the current to move to this R and therefore $R_2$ will also have the same current here, the same current will reach at the output of the current.

But please understand the total output current is limited by power dissipation of operational amplifier right, so operational amplifiers are designed to operate at optimal power dissipations. Now if the current is very high right, it might trouble the operation of an operational amplifier and therefore you always have a limit on the total amount of current flowing through the operational amplifier in the output loop factor. So this is one problem area which people face.

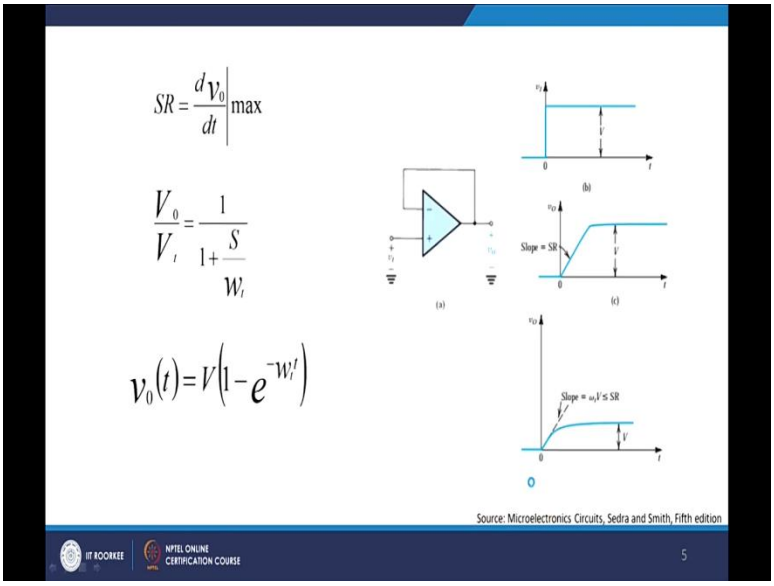
(Refer Slide Time: 09:44)

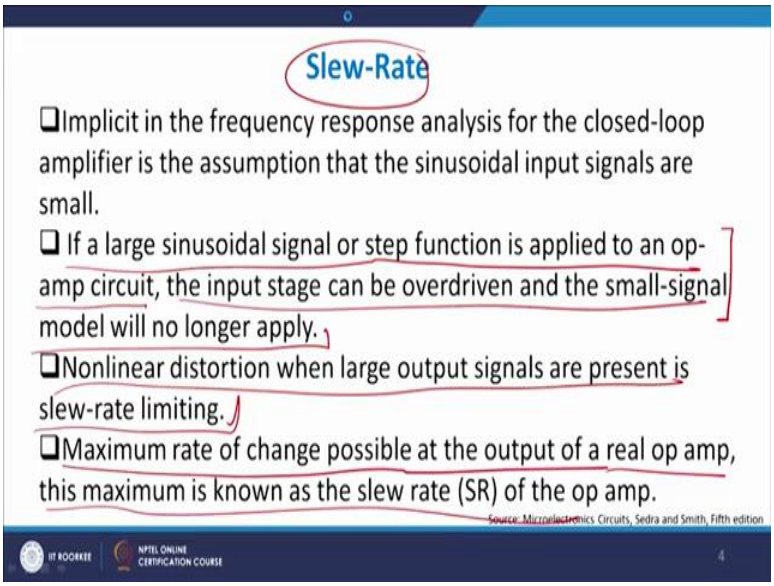



The third problem area is of the slew-rate, you would not be as far as designing is considered in most of the cases if it is a small signal you would not be having these problems coming in the picture but whenever you have a large-signal model for operational amplifier you might have these problems. Slew-rate is primarily if you look very closely, look at the second point will give you a give you an idea. If a large sinusoidal signal or a step function is applied to an operational amplifier circuit the input stage can be overdriven and the small-signal model will no longer apply, as I was saying to you.

Now if there is a heavy non-linear distortion when there is a large distortion signal, then we define it to be the slew-rate limited signal. So how do you define slew-rate, the maximum rate of change at the output.

(Refer Slide Time: 10:49)

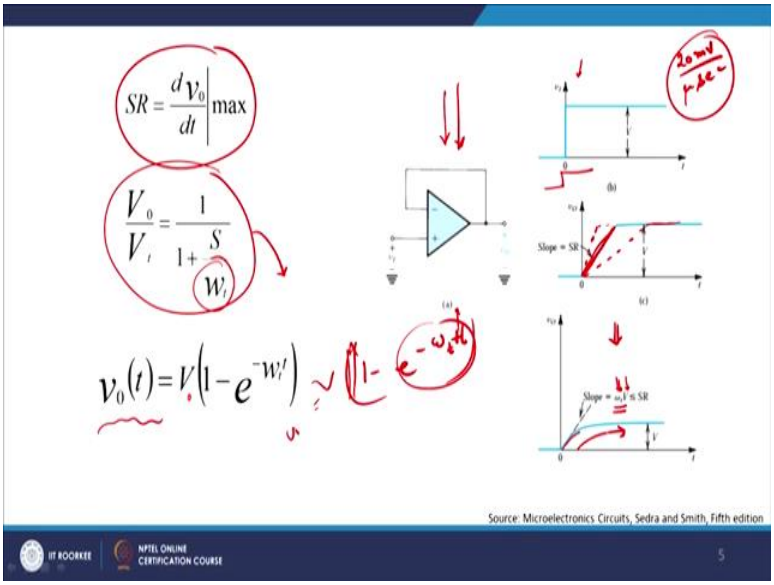



The maximum rate of change possible at the output of a real op-amp is known as the slew-rate which means that let me just show you then we will come back here. This is my definition $dV_0/dt$ means the rate of change of output voltage with respect to time, the maximum it can sustain and why I am saying the maximum it can sustain is there is certain specific reasons for that.

The maximum, the reason for that is that the variation of the voltage with respect to time may may, let it be at input or output is always governed by the RC time constant of that particular point or that particular node, so you cannot expect that if I give an input which is suddenly rising like a step voltage, you would not expect that the same performance will be at the output of your operational amplifier, it will be always delayed and there will be always a RC time constant available to you at the output side right, so we define slew-rate as the maximum rate of change of output voltage with respect to time which the operational amplifier can safely sustain. You make it slightly larger than that output variation and you would not get any change in the output cycle right or it would not be or will be actually slew-rate limited output supply.

(Refer Slide Time: 11:57)




So as I was discussing with you let us suppose I have a voltage follower right, I discussed this point or I discussed this circuit in your previous turn as well and let us suppose I have a voltage follower. In this voltage follower I give a step input which is given as $V_I$, what I am seeing is that because of the step input ideally I would have seen the output to follow the input because it is given to the positive terminal, non-inverting terminal and therefore output should ideally follow like this but because op-amp has got a problem, therefore output will always be a like this which means that it will flow or it will rise at the maximum pace.

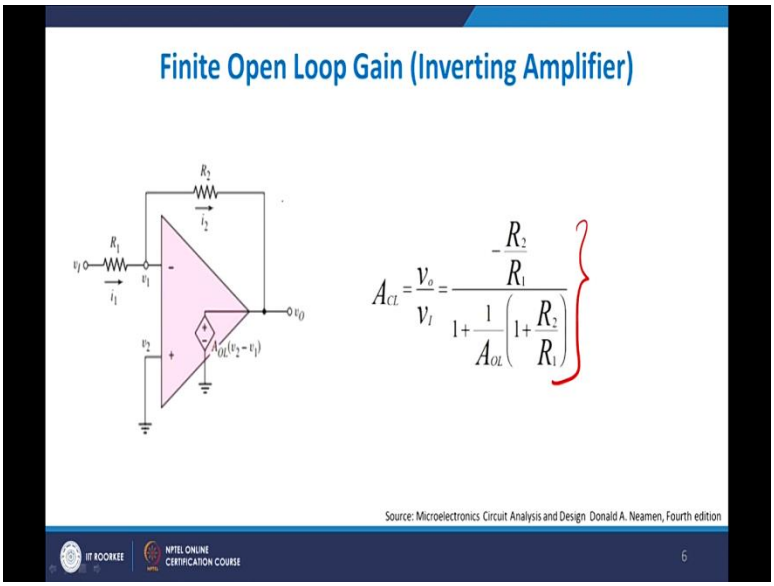
So if the slew-rates are large it will be like this, if the slew-rates are small, it will be something like this, you got my point. Therefore, these are known as slew-rate limited operational amplifiers right. So they will tell you that it is basically 20 milli volts per second or per microsecond which means that for every one microsecond change in the output I would expect to see a 20 milli volt change in the output voltage that is the maximum that the op-amp can sustain. Anything larger than that it cannot be sustained with operational amplifiers, right? So this is the basic concept here which you see.

Now as I discussed with you, therefore, if you look at the curve which is the last one here then this slope which you see that means it will be slew-rate limited then this slope gives you the value which is the $\omega_T$ into, $\omega_T$ is the unity gain bandwidth frequency multiplied by voltage must be less than equals to slew-rate. So so this we will discuss later on if time permits. But $\omega_T$ into,

$\omega_T$ into V is the input voltage given is basically the rate of change of my output voltage and that should be always less than the slew-rate, then you will get such type curve in reality, right.

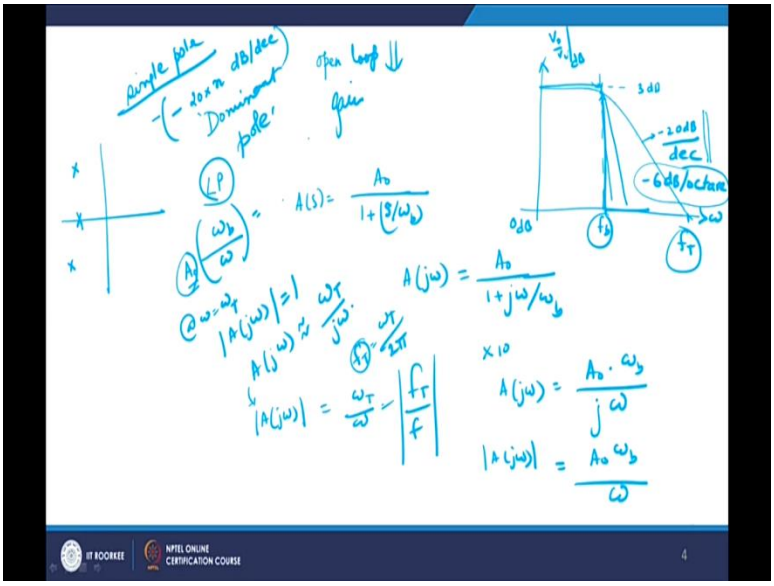
So I get $V_0$ by $V_t$, $V_0$ is the output voltage by $V_t$ which is the input voltage is 1 by 1 plus s by $\omega_T$ where $\omega_T$ is basically the 3 db bandwidth frequency, so if you do transform, Laplace transform of this one you get $V_0$, inverse Laplace transform sorry, you get $V_0(t)$ equals to $V_0$ into 1 minus e to the power minus $\omega_T$. So 1 minus e to the power minus $\omega_T$ into t into $V_0$ is equals to $V_0$, so you see as the time increases this quantity will drop down this quantity will increase and that is what is happening, fine. So this is Slew-rate dependent phenomena which you see.

(Refer Slide Time: 14:35)




Now if you, what we were discussing in the previous turn was that if you have an inverting amplifier, I mean inverting amplifier, then the output will be 180 degree phase shifted with respect to input and therefore you put a negative sign in your closed loop gain. So I get $A_{CL}$ equals to minus $R_2$ by $R_1$, remember? Now what has happened is that my loop gain or my gain was considered to be very high right and therefore my, but if it is not high and it is restricted by a certain value, for example $A_{OL}$ which is the open loop gain, then the closed loop gain is related to the open loop gain by this this factor, right? And this is what I get; this is what I get as the difference in the open loop gain and the closed loop gain.

(Refer Slide Time: 15:35)




So just to show you this basic concept as far as effect of finite loop gain is considered. Let me just show it you what I am talking about and let me show you how it works out that, let us suppose I have a filter whose output characteristics look something like this so I have, this is basically your, this is your 3 dB bandwidth or 3 dB point and the voltage is defined as $f_b$ let us suppose where it cuts is defined as $f_T$ unity gain because it is 0 dB, this is my gain $V_0$ by $V_i$ in dB and this is my ω in terms of frequency we will get some value.

Now so what I am trying to tell you is that open loop gain, open loop gain, let us suppose it drops down or it falls then how does it influence my in this thing. So I have a low pass filter let us suppose and transfer function is given by A(S) equals to A0 upon 1 plus s by $\omega_b$, where $\omega_b$ is basically my this frequency here. So in terms of jω I get, I just replace S by jω I get $A_0$ upon 1 plus jω by $\omega_b$, right?

So, so what I am trying to tell you is that let us suppose now you assume that your frequency is increased by 10 times as compared to $\omega_b$, so I am somewhere here right, then I get A of jω equals to $A_0$ into $\omega_b$ by ω, got the point, jω, why because if ω is very very large as compared to $\omega_b$ this quantity will be very large as compared to 1 and I get $A_0$ into $\omega_b$ by ω s equals to A of jω right? So if I take the mod value of a of jω then I get this to be equals to $A_0$ $\omega_b$ by ω.

So I get the gain to be equals to $A_0$ multiplied by $\omega_b$ by $\omega$. $A_0$ is fixed because that is open loop gain multiplied by $\omega_b$ by $\omega$. This is my sort of a closed loop gain and a finite loop gain which you see, right. So this gives you a value, now at $\omega$ equals to $\omega_T$ where you have got a unity gain bandwidth product I will automatically have this equals to 1 and therefore a of $j\omega$ will be approximately given as $\omega_T$ by $j\omega$ because this is 1 $\omega_b$ will convert to $\omega_T$ and $\omega_T$ by $j\omega$ I get where $f_T$ will be given as $\omega_T$ by $2\pi$, right?

So this is your unity gain unity gain bandwidth frequency right? And therefore if you want to find out this, if you want to find out the this the magnitude of A of $j\omega$ from this case right, I get this to be equals to $\omega_T$ by $\omega$, this is nothing but $f_T$ by f. So it is the ratio of the unity gain frequency divided by the frequency at which you are trying to find out the gain, the ratio of that will be equals to your gain right at that particular frequency. Now that is with the consideration that your finite loop gain is already available to you.

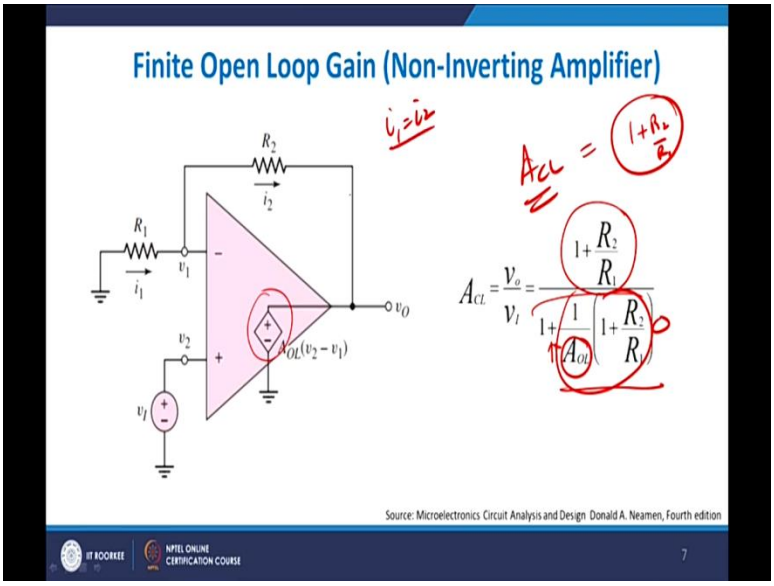
Now generally if this is a single pole dominated transfer function, single, if there is a single pole on the left half plain, then you will automatically this will be around minus 20 dB per decade or minus 6 dB per octave. So please try to find out why it is 6 dB per octave, I leave it as an exercise to you but you can ask questions later on through the discussion forum but try to find out why it can be also referred to as minus 6 dB per octave or minus 20 dB per decade provided you have a single pole on the S plain.

So if you have a single pole here then we define that to be as a this thing, if there are multiple poles, for example this complex conjugate poles here so there are two poles, so it will be 20 times n dB per decade. I will not derive it here, so this is minus 20 times n dB. So you multiply 20 with the number of poles, so if there are two poles, 40 dB per decade. So you would expect to see a steeper fall. So larger the number of poles right, more steeper will be your fall in the low-pass filter and better will be your characteristics, you will be reaching to the much.

So the ideal characteristic is something like this right, for a low pass filter. So now you will get much more better because this fall will be very fast here, so which means higher number of poles means higher order of your filter, you automatically get a larger and larger value of your input profile and that is what is known as a dominant pole approximation. This is known as a dominant

dominant pole, right? So this is as far as understanding the dominant pole approximation is concerned.

(Refer Slide Time: 20:39)



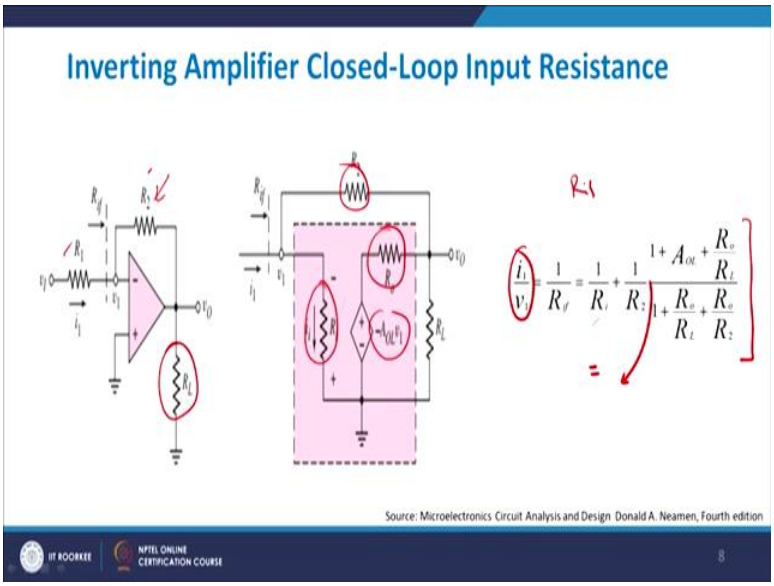

Now if I take a closed loop control or a closed loop consideration I get $R_2$ by $R_1$ into 1 plus 1 by $A_{OL}$ 1 plus $R_2$ by $R_1$. I am not deriving it in the class but this is what you get out of it. Now you see if your open loop gain $A_{OL}$ is extremely high or infinitely high then this quantity, this quantity goes to 0 and I get again $R_2$ by $R_1$ which is basically my initial inverting mode, remember, inverting mode operational amplifier. Got the point?

So when your open loop gain is very very high 10 to the power 7 in the ideal cases, you get a closed loop gain almost near to $R_2$ by $R_1$ but if your $A_{OL}$ is not very high it is 10 or 20 or 30 or 40 whatever, then these also lower your overall closed loop gain right and that is the inverting amplifier case.

Let us look at the non-inverting amplifier case, exactly the same only thing the numerator is replaced by 1 plus $R_2$ by $R_1$ and you have denominator given as this similarly if $A_{OL}$ goes to a very high value, this quantity this quantity goes to 0 and therefore I get 1 plus $R_2$ by $R_1$ as the value of your $A_{CL}$ closed loop gain, right? And as you can see that for the small signal model this has been replaced by a voltage source of a value $A_{OL}$ into $V_2$ minus $V_1$, fine.

And this is the current source so I am assuming that so I will be assuming that $I_1$ equals to $I_2$, well assuming or it is true in a sense true also because this is basically high impedance node available here, right? So I have a finite open-loop gain that gives me a non-inverting amplifier here.

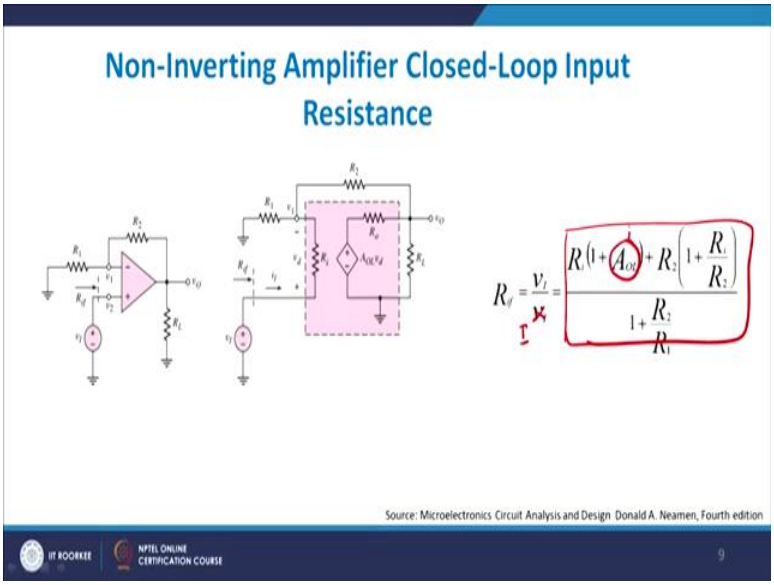
(Refer Slide Time: 22:23)




Now if I want to find out the inverting amplifier closed loop input resistance then I have to find out $R_{if}$ which is $I_1$ by $V_1$. Now if you look at this this thing here, so I have got $R_1$, $R_2$ and let us suppose I have a load resistance $R_L$. This can be made to equivalent small-signal models, so this is my input impedance $R_i$, this is my $R_2$ which is the feedback resistance here and I have got output resistance $R_0$ and minus $A_{OL}$ into $V_1$ because you are inserting your signal into the inverting terminal and therefore minus $A_{OL}$ into $V_1$, why $V_1$, because $V_2$ is grounded. So I get minus $A_{OL}$ into $V_1$ into $R_L$ is the external node.

If you solve it by doing simple simple derivations I get this the overall picture as the input resistances, so sorry so I get this plus sorry this will be equal to, so this will be here equal to right and I get, 1 over that will give you the value of your $R_{if}$ input impedance for the closed loop gain.

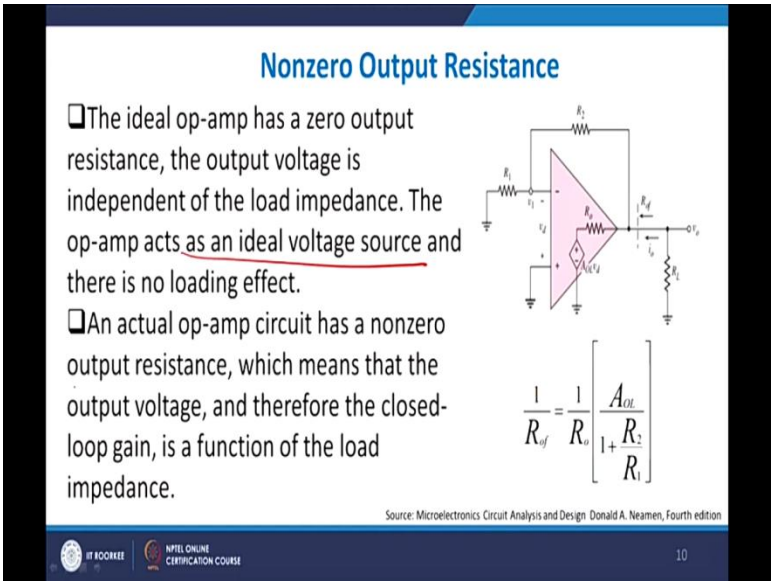
(Refer Slide Time: 23:16)




If I do a non-inverting exactly the same happens, non-inverting case $R_{if}$ case comes out to be equal to this one right and this is sorry this is actually $V_I$ by $I_I$, right, input current and this is the input resistance 1 plus $A_{OL}$ plus $R_2$.

So as you can see if this goes to very high value right, your $R_{if}$ also goes to a very high value right and this is true also if a open loop gain is very very high of an operational amplifier your input impedances will also be very high right and you can get this value here. So is it a $R_1$, $R_2$, $R_L$ is exactly the same as the pervious case and we get this into consideration here with $V_d$ is the voltage difference between $R_1$ and $R_2$ between this point and this point, right.

(Refer Slide Time: 24:09)



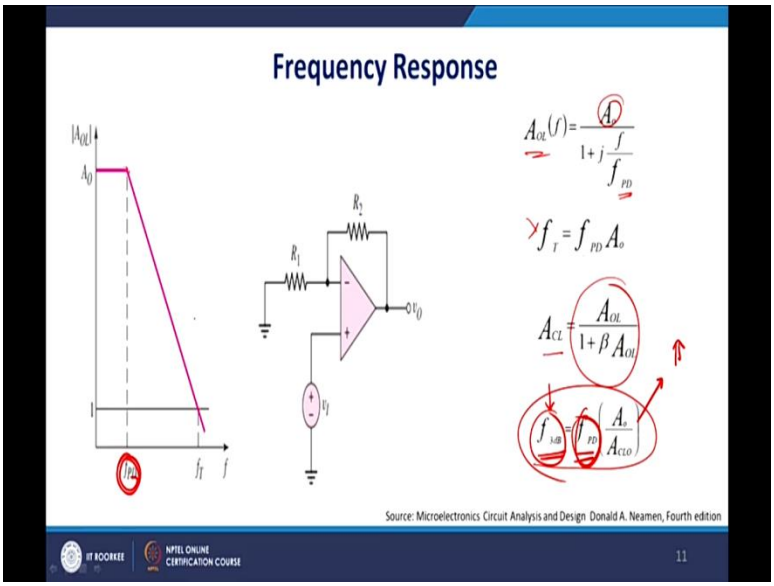

We come to the non-zero output resistance; typically the ideal op-amp has a zero output resistance. We have already seen that but and the output voltage therefore is independent of the resistances seen at the output load and that is therefore op-amp starts to act like an ideal, what is an ideal voltage source? Ideal voltage source is a source which gives you, wherein you can draw any large current but the voltage across the that source will always remain constant right and that was the ideal voltage source which means that there is no loading, in reality not true that the output the ideal op-amps will have some non zero in resistance in the output side.

As I discussed with you therefore the actual op-amp circuit will have a non-zero output resistance which means that that output voltage and therefore the closed loop gain is as function of the load impedance, right. Why? Because see if let us suppose, let us suppose that output voltage is independent of load impedances right, so whatever your load impedance is 10 kilo ohm, 5 kilo ohm, 100 ohm or 40 ohm or shortage, the voltage level at the particular point will remain fixed and so when your impedance levels are say 100 kilo ohm, you will be drawing a smaller current but if voltage remains fixed right.

So when you want to make your voltage independent of load impedances then you assume that op-amp is behaving like an ideal voltage source, in reality not true as I discussed with you and therefore your output voltage will start to become a function of your output impedance, right? And therefore when you output impedance changes, your voltage will also change.

It is given by this formula here1 by $R_{of}$ to be equals to 1 by $R_o$, $R_o$ is the, 1 by $R_o$ is basically my output impedance under ideal conditions right, but with this you get $A_{OL}$ coming into picture. So if you just make it, if you just do the reciprocal of this one I get $R_{of}$ to be equals to $R_o$ right and then you reciprocate this 1 plus $R_2$ by $R_1$ right, divide it by $A_{OL}$. As you can see make open-loop gain infinitely large and $R_{of}$ goes to 0. So ideal output output voltage always be close to 0.

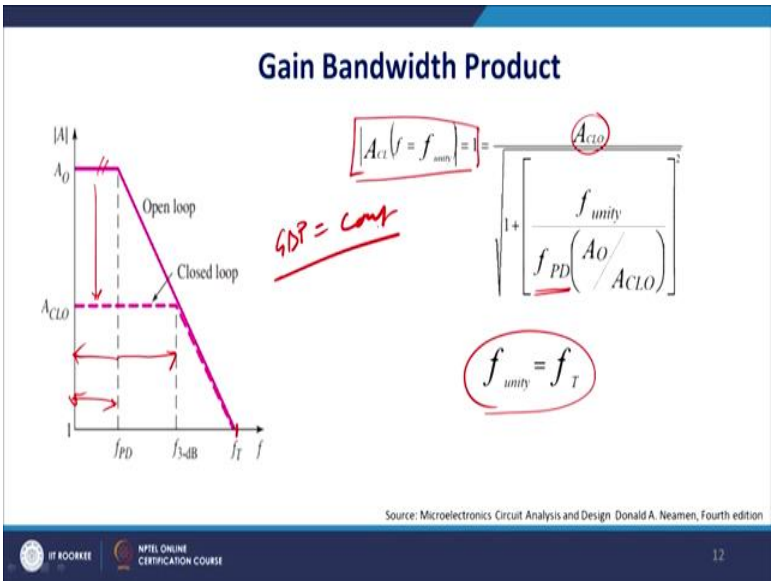
(Refer Slide Time: 26:10)




Look at the frequency response, as I discussed with you frequency response for any other cases will be just like a low-pass filter and $A_{OL}$, open loop gain is given as $A_o$ upon 1 plus j by $f_{PD}$, PD is basically your 3 dB point, so $f_T$ which is the unity gain frequency is given by $f_{PD}$ into $A_o$ and therefore the closed loop gain is given by $A_{OL}$ upon 1 plus β times $A_{OL}$, right? And therefore you get f 3dB equals to f $_{Pd}$ into $A_{OL}$ upon $A_{CLO}$, which means that 3 dB frequency depends upon your $f_{PD}$ and also depends upon the ratio of your open loop gain on the closed loop gain, right.

So if you know, if this quantity is high which generally it is then your 3dB bandwidth is higher than $f_{PD}$, right? And that is what you get here $f_{PD}$ is basically my 3dB, 3dB bandwidth and f $_{PD}$ is my, $f_{PD}$ is this point where it starts to go down and this $f_{3dB}$ is the bandwidth at which you have a 3dB gain here. So we will again come back to this frequency response later, all in a detailed manner. At this stage you can understand it is almost behave like an integrator and low-pass integrator.

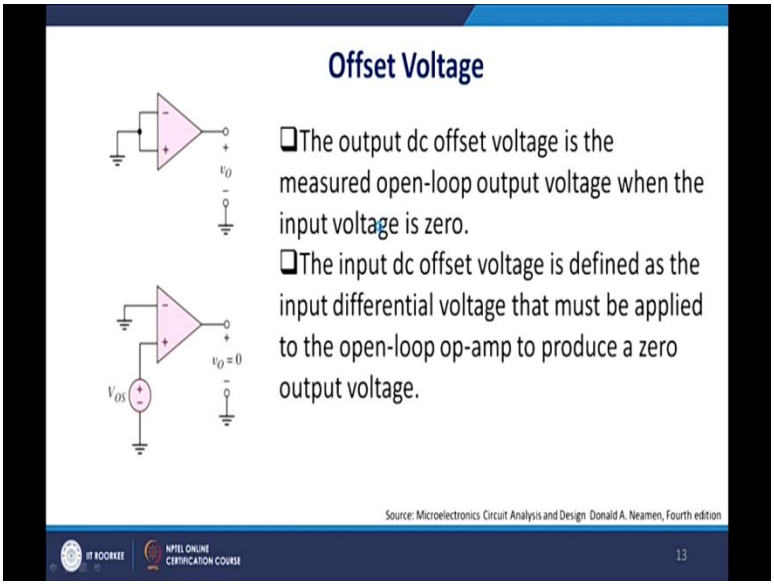
(Refer Slide Time: 27:28)




Now gain bandwidth product is given by closed loop gain if you look very carefully, unity $f_T$, $f_T$ is unity gain bandwidth is given by CLO which means that the closed loop circuitry 1 plus f unity; f unity is $f_T$ divided by $f_{PD}$, $f_{PD}$ is basically my 3dB into $A_{OL}$ by $A_{CLO}$. And therefore I get $f_T$ equals to, so f unity is equals to $f_T$, provided I get $A_0$ by $A_{CLO}$ very high, right.

So what happens is that when you open-loop your gain is higher, when your closed loop your gain falls down but your 3dB bandwidth increases so your initial 3dB was this much right and now your 3dB has improved by this much, so your 3dB bandwidth is an and that is the reason your GBP, Gain Bandwidth Product is always constant. So the first case without when you are in an open loop conditions, you automatically have your gain very very high, in the closed loop conditions your gains falls off drastically, right. Okay

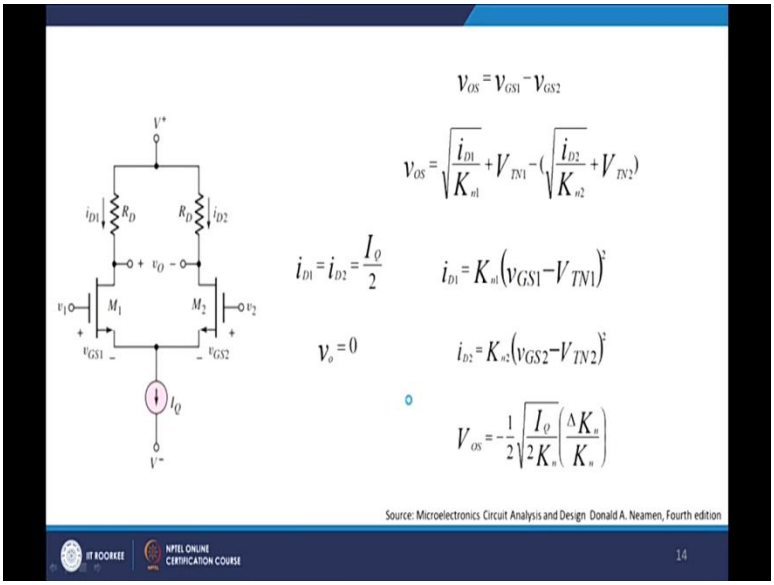
(Refer Slide Time: 28:20)




We come to another important issue of an operational amplifier and that is known as an Offset Voltage. Now ideally if you remember if I short the input of my operational amplifier, the inverting and the non-inverting input, then for this, then I will get 0 output available to me because $V_2$ minus $V_1$ is 0. In reality not true, there is some amount of DC voltage available to me in the output side. So I can see here the output DC offset voltage is measured open-loop output voltage when the input voltage is zero, right. The input DC offset voltage is defined as the input differential voltage that must be applied to the open-loop op-amp to produce a zero output voltage. You got the point?

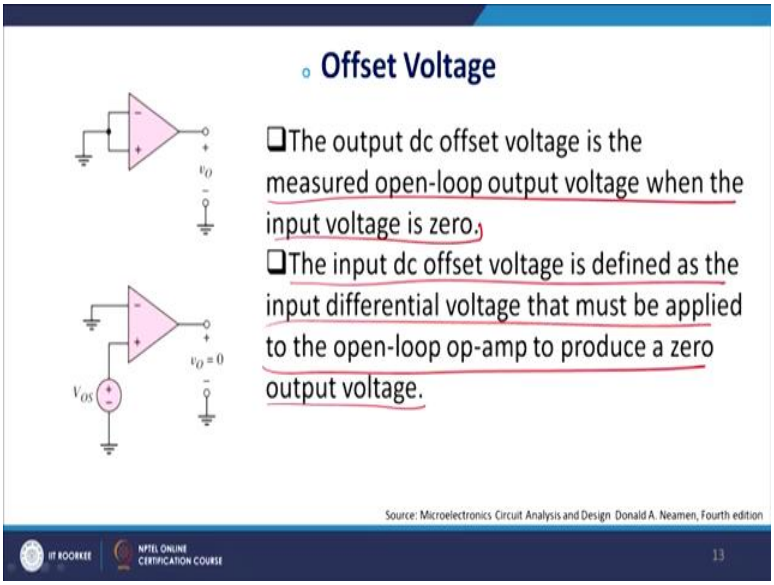
So you see when you do not apply anything you get some voltage right? Now if we apply a voltage in the reverse direction, then this high value will go down to zero. So this amount of voltage is, input side voltage is defined as my input differential offset voltage right or input offset voltage. I can have an output offset voltage also I can have an input offset voltage. What is an input offset voltage, at the input side the amount of voltage you need to give differential between two voltage sources so that the output goes to zero is my defined as my input differential voltage.

(Refer Slide Time: 29:40)



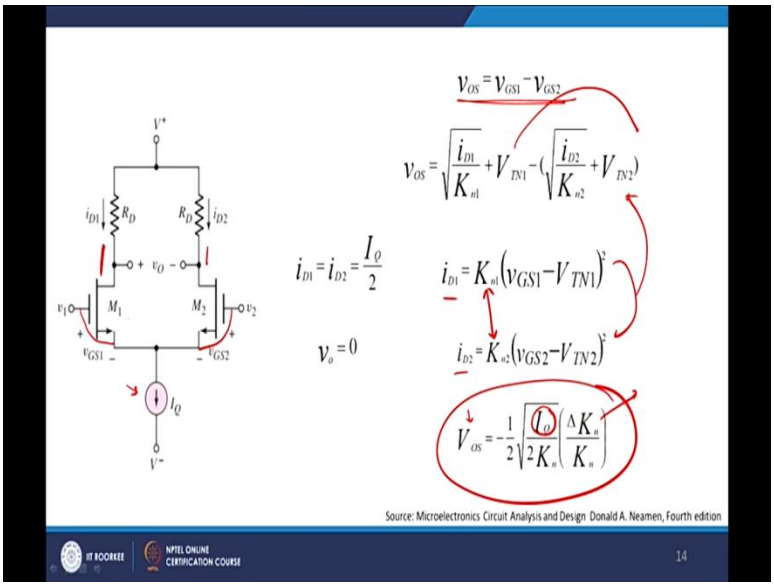

I will not go into the details of the derivations here but typically if you look very carefully assuming that, I have a differential pair here right and it is driven by this, by this ideal current source here. And so this is the $V_{GS1}$ is the drive voltage, $V_{GS2}$ and I am trying to do the voltage difference between $V_0$ and $V_1$. So offset voltage will be given as $V_{GS1}$ minus $V_{GS2}$ right, remember because these are the two inputs I am giving to the operational amplifier. Now I am assuming that these voltages are not equal, so I am trying to find out the difference between the two voltages, right.

(Refer Slide Time: 30:23)




So in reality if you look very carefully, in the previous slide if you, if you go back to the previous slide see this is the $V_{OS}$ I am giving here and I have grounded this I am giving $V_{OS}$, so that this goes to 0.

(Refer Slide Time: 30:32)

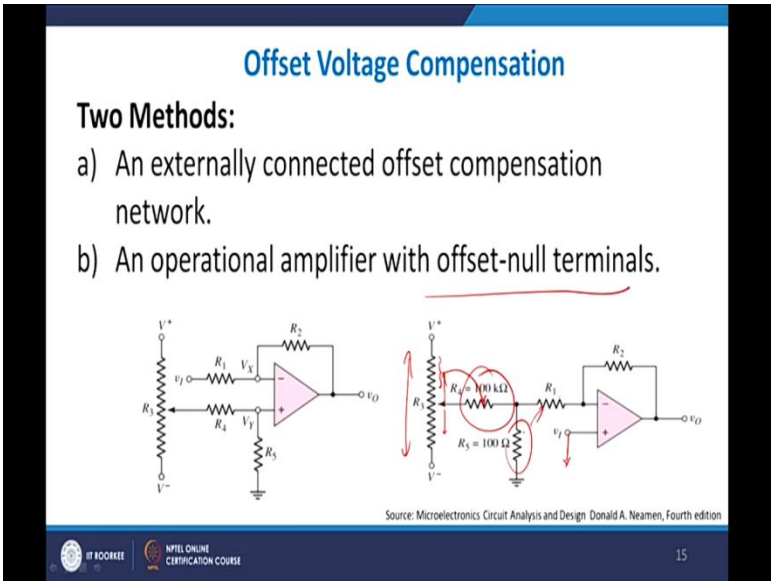



So with this concept I have applied $V_{OS}$ as $V_{GS1}$ minus $V_{GS2}$ right. This is from assuming that there is saturation I get root of $I_{D1}$ upon $K_{n1}$ where $K_{n1}$ is $\mu_{n1} C_{ox}$ (W/L) by l plus $V_{TN1}$ threshold voltage of N-MOS minus this whole quantity where $I_{D1}$ is given by this $I_{D2}$ is given by this. $I_{D1}$ is the current flowing through this arm and $I_{D2}$ is the current flowing through these two arm. If you

put all these values back into this equation, do a small solution assuming that $V_{TN1}$ equals to $V_{TN2}$, so they cancel out. I get this output voltage the value here, right, where, just let me see, yes.

So I will get this output voltage where $I_Q$ is the sum of the two currents which you see and delta $K_n$ is basically the difference in the value of your, these two values, right. Difference of these two values and $K_n$ is basically the one of the, $K_n$ is the process transconductance parameter of the differential pair. As you can see here higher the current more will be offset in a negative sense right, because it is the inverting terminal which I am giving to you. Okay.

(Refer Slide Time: 31:54)



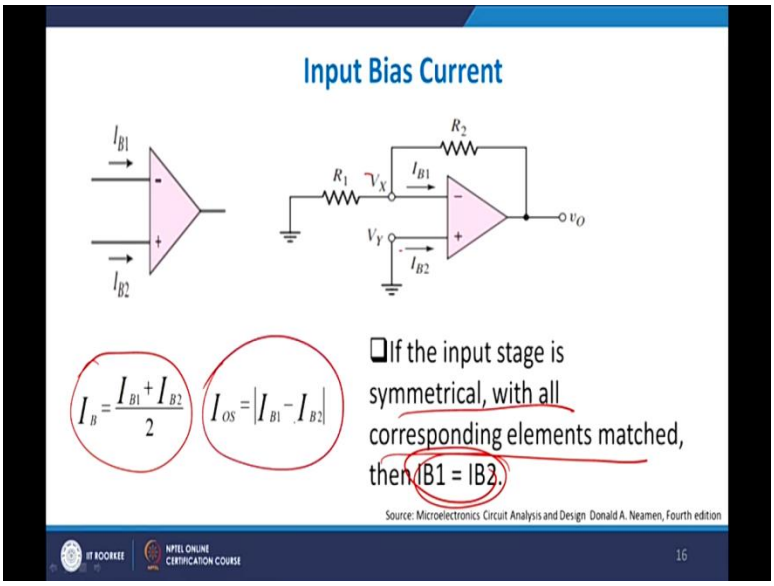

We come to the offset compensation voltage compensation, so there are two methods by which you can compensate the offset voltages and these two compensations are externally compensation, externally connected of set compensation network and with the null offset terminals. So we would not go into the details but I will give you an idea how it works out in a very simple manner. So you see I have $V_i$ here which I am giving. I also have a 100 ohm $R_1$ and 100 kilo ohm here.

I have a potential divider at $R_3$, so If I move this up or down right, some amount of resistance will be in series 200 kilo ohm and that will be in parallel to 100 ohm here and this will be feeding the voltage to $R_1$, you getting my point? So by simply choosing the potential divider

network up or down I can actually make the resistance value in a potential divider network change right and make it exactly equal to the offset voltage.

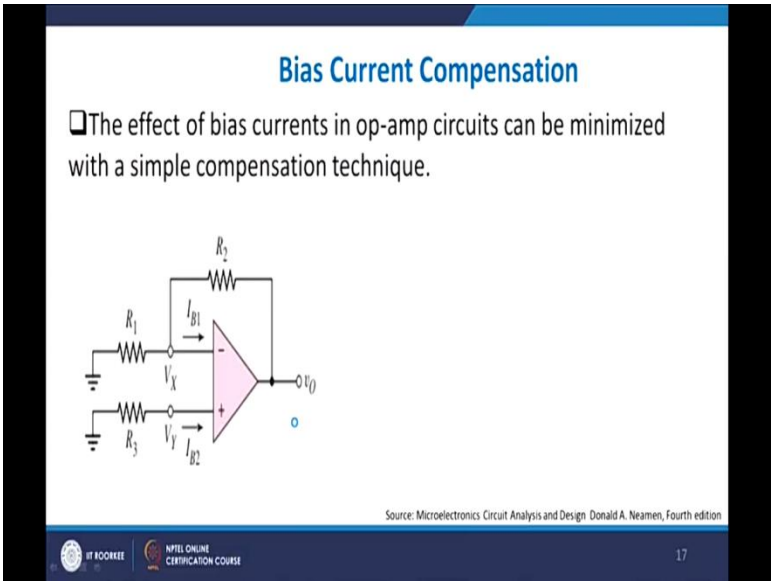
So what I do, what I initially experimentally how how I do it, I go on shifting this potential divider network up or down and check out that when $V_0$ equals to 0. So when $V_0$ is equals to 0, I stop and that is basically my input offset voltage, right. So this is one of the methodologies people adopt across the current.

(Refer Slide Time: 33:01)



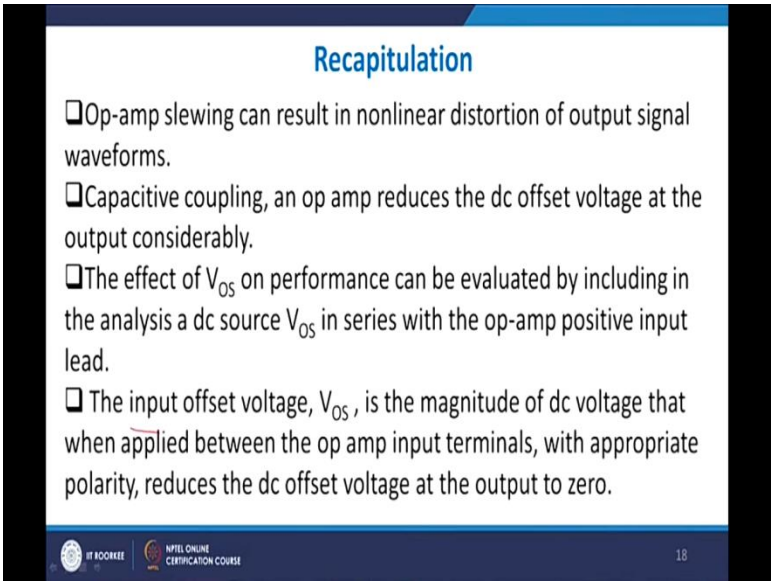

Then you also have an Input Bias Current, well this is very simple. They will be generally symmetrical if everything is matched for the top and bottom. If the input stage is symmetrical with all the corresponding elements match, $I_{B1}$ will has to be equal to $I_{B2}$. So I define $I_B$ equals to $I_{B1}$ plus $I_{B2}$ the average current and offset current is defined as difference between $I_{B1}$ and $I_{B2}$. In reality, $I_{B1}$ and $I_{B2}$ must be equal but they are not and therefore there will be difference. Okay.

(Refer Slide Time: 33:35)




I come to Bias Current Compensation, so I was doing voltage compensation there. Now I do a bias compensation and this bias compensation is given by this manner by that which we have got $V_X$ and $V_Y$ here and simply by choosing appropriate value of $R_3$, I can choose $I_{B1}$ equal to $I_{B2}$, right. I will not go into further details of this one, because this is slightly higher than what you are supposed to know at this stage and therefore you need not worry about from where I am getting this bias compensation.

(Refer Slide Time: 34:26)



So let me recapitulate what we did, we were actually looking into an op-amp slewing action. What is the meaning of slew right and what is slew rate, slew rate limited operation? What is a DC offset voltage right and how do you define DC offset voltages? What is the meaning of saturation, operational amplifier saturation? What is the meaning of input offset voltage $V_{OS}$?

And how can we determine the input offset voltage, similarly what is the input offset current and how can you remove it. So this we have learned in this module and this takes care of approximately our whole understanding of the analog part of the syllabus. So from next turn onwards, we will start with the combinational logic which is a digital part, right? And for the next ten five to six hours of our lecture which is left we will be actually doing a lot of differential a lot of digital logic design.

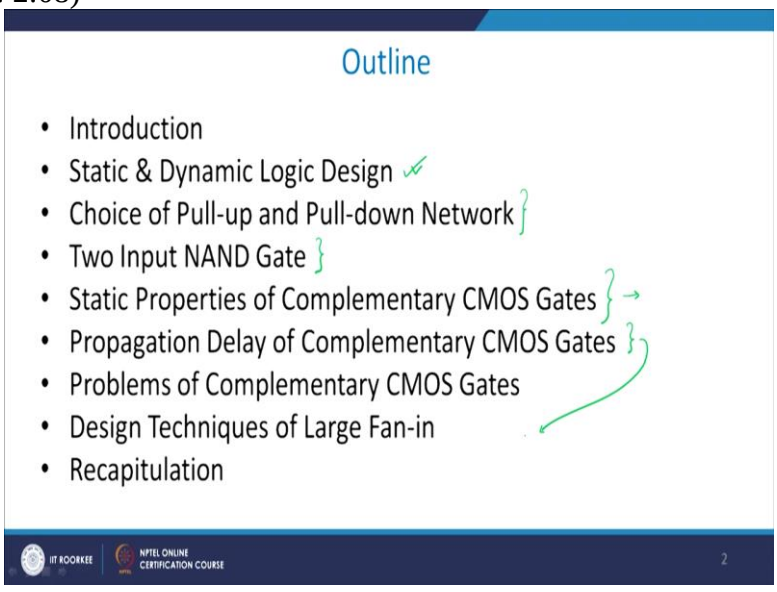
So this is where we stop our mixed signal or analog block out of the circuits, so if you look at the whole structure of the course it is first devices which is the first few weeks, one or two weeks. Then subsequently from three to lecture week number seven it is primarily a analog flow and then we have a digital flow at the last places. So next time onwards we will start the digital flow. Thank you very much.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−53**
**Combinational Logic Design -I**

Hello everybody and welcome once again to the NPTEL online certification course on microelectronics: devices to circuits. Till the previous module, we had covered the basic of devices and the usage of devices for analog circuits. So primarily, till the previous module which is about week 6 or week 7, we have actually dealt with the whole of analog design. The most, at least the basics of analog design and so that given a MOSFET, either an N channel or P channel, now you will be in a position to actually design for example basic filters, amplifiers, differential amplifiers, operational amplifiers, you would also be in a position to understand the device physics as we have understood in our previous module.

From this module onwards, we start our journey for digital logic design. This we will keep it short, so about say about 4 hours will be devoted for digital logic design because the same has already been done in your previous modules. So we will be repeating the slightly varied amount, keeping in view your requirements for this course. So we will start with today's module or talk on combinational logic design and this is part 1 module of the combinational logic design.

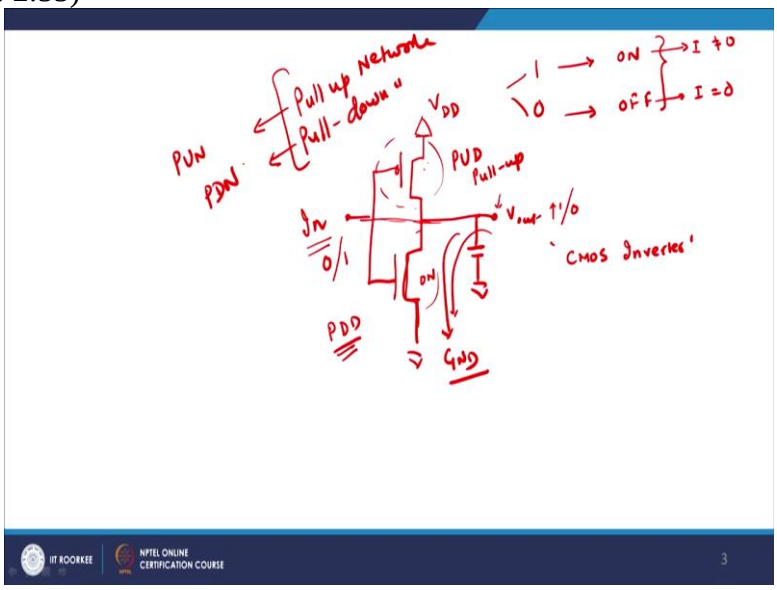
(Refer Slide Time: 2:08)



So what we will be doing in the next half an hour or so is, give you an idea about what is combinational logic design, we will have a look into what is known as a static and dynamic logic design. So this will be, we will be looking into the static and dynamic logic design, right, what is the meaning of that and then what makes your device go ahead and choose pull up and pull down network.

So which device will you use for pull up networks and which device should you use for pull down networks? We will explain each one of them. Then we will take an example of a two input NAND gate and then we will go into static properties like for example, complementary gate and then we will look into the propagation delay and then basically designing of large fan-in circuits and then we will recapitulate the whole thing. So this is the basic flow which will happen.

(Refer Slide Time: 2:55)




Now just to give you an insight of what we are trying to do here is, that till now we were assuming that you have an input which is basically a sine wave and you are supposed to know how does the MOSFET or the active device behave at each and every point of the input cycle which means that if the device is triode region, saturation region, active region and so on and so forth. You were also told that the current equations for each region should be known to you right.

In digital logic design whereas, you actually work with only 2 phases, which is switching on and off. You must be knowing that in digital logic, you only work with 1s and 0s. So this corresponds to on state and this corresponds to generally you know off state right? And therefore

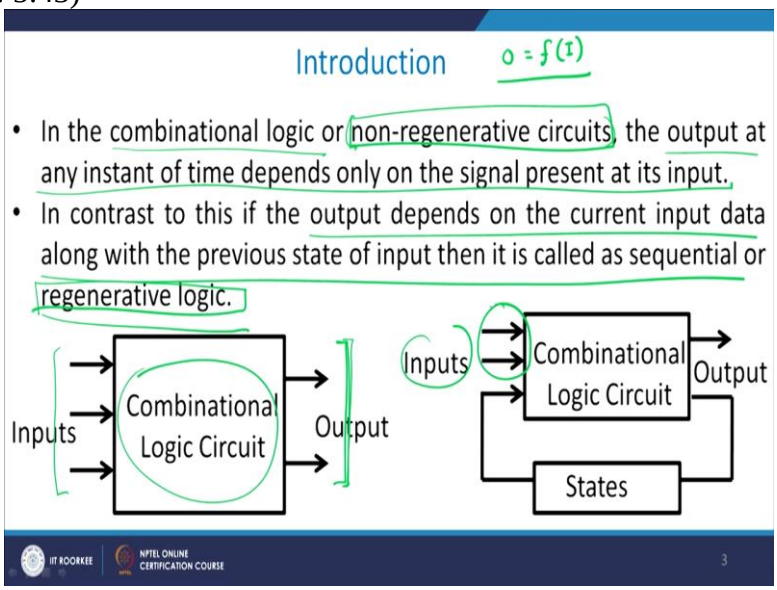
there will be only 2 states and therefore in this case there will be some current flowing, in this case the current will be equals to 0 and this case the current will be equals to nonzero, right.

So that makes my life relatively easier when we are dealing with digital logic. We have already seen in our previous turn that whenever we are discussing a CMOS transistor, if you have an NMOS and PMOS then when you give, say this is your input right and this is your output, $V_{out}$ and you have a load capacitance here. And this is your $V_{DD}$ then when you given input here which is equals to say 0 then this switches on and $V_{out}$ goes to 1, right?

When this becomes 1, the NMOS switches on right and this voltage here drops down to 0 and therefore this goes to 0. So and this is known as a basic CMOS inverter, right? This is inverter. Now in this case, as you can see, we can divide the whole thing into 2 parts. The top part and the bottom part. So this is known as pull up device, PUD and this is known as a pull down device, PDD. So this is pull up device. Why pull up? Because this device helps you to pull up the voltage at this particular point to $V_{DD}$.

And why pull down device? Because this device tries to pull down the voltage at this point down to ground. This is your ground, right. Now if you have combinations of PMOS in the pull up, then we define that to be as the pull up network. If there are combinations of $PMOS_S$. And similarly, if you have combination of $NMOS_S$ design at the bottom, then we define this to be as a pull down network. So we refer to this as PUN and this as PDN right. This is the basic fundamental principles based on which we will be starting to look into our whole issue of combinational logic.

(Refer Slide Time: 5:43)




Let me come to the combinational logic, first of all give you an introduction of combinational logic. Combinational logics are the most simplest circuits of digital logic design and as I discussed with you or I will be discussing just now, is that in a combinational logical circuit if you look at this particular diagram here, you see I have got a set of inputs here on the left-hand side and I have a set of outputs here on the right-hand side right?

Now if you look at this 1st bullet which you see, in the combinational logic right, we will not discuss at this stage non-regenerative but that is in the combinational logic, the output at any instance of time depends only on the signal present at that input which means that depending on the inputs available here at a particular instance of time, the output is determined, right. So if you can write down, output is actually only function of inputs at that particular time. It does not depend upon anything else, right?

Whereas, if the output which is the 2nd bullet statement, if the output depends on the current input data which is this one, along with the previous state of the input, we define this to be as a sequential or regenerative logic which means that if the output from the previous sets of input helps you to determine the current state, we define that to be the sequential logic. So what is the primary difference between sequential logic and combinational logic?

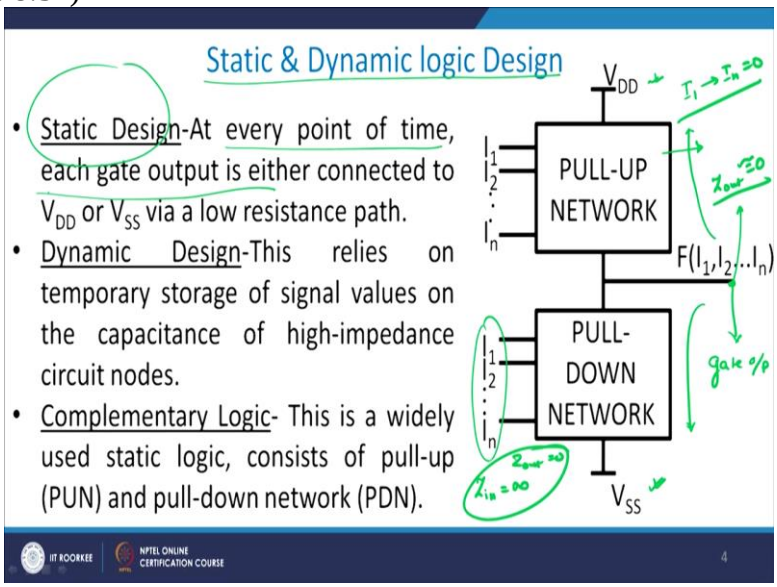
Combinational logics are responsible for giving you an output based on current inputs. For example NAND gate, NOR gate, XOR gate, Ex- OR, any of the standard gates available to you

whereas, a sequential logic output not only depends upon the present state of input, it also depends upon the previous states right? And therefore you have a memory sort of memory here because you are storing at least some amount of data for a finite amount of time.

Therefore, these are also known as sequential logic, also known as regenerative because you are regenerating the output based on the previous inputs whereas, the combinational logic is basically non-regenerative. And therefore combinational logics are also referred to as non-regenerative circuits whereas sequential logics are referred to as regenerative logics and circuits, fine. So, combinational logic therefore let us understand once again, it depends only on the inputs in the present states.

Sequential logic depends on the present state inputs as well as remember, previous state inputs right, on the previous outputs. Now, what we will be looking into this about 2 modules or so or 3 modules or so, we will be concentrating on the combinational logic itself right. Within the combinational logic, we have 2 types of logic once again. One is known as static and another is known as dynamic logic right.

(Refer Slide Time: 8:37)




Now static logic primarily means that at every point of time, each of the gate output which is this one, this is the gate output, right is either connected to $V_{DD}$ or $V_{SS}$ via a low resistance path.

I suppose you can understand. So let us suppose your all your inputs here, $I_1$ to $I_n$ are all 0, right and suppose all the $PMOS_S$ are in series here right. Then all the $PMOS_S$ will be switched on, so therefore I will have a low resistance path between output and $V_{DD}$ and this $V_{DD}$ will appear on the gate side, right. Similarly, if all your inputs are 1 here and all your NMOS devices in the pull down network are in series, then all your devices will be switched on together and you will have a low resistance path from your output to the $V_{SS}$ right.

So you will have a low resistance path at this particular point right. Then we define such a design to be a static design, right. So what is a static design? Whenever the output is connected to a low impedance load or is connected to a $V_{DD}$ or $V_{SS}$, which is output and it is not floating therefore, please understand, it is not floating. It is either connected to $V_{DD}$ or $V_{SS}$. So this is therefore, the impedance at this particular point, $Z_{out}$ will be approximately equals to 0, right.

So output impedance is 0 and input impedance is infinite because, why input impedance is infinite in static design or for that matter, any CMOS structure is because you are sending the signal onto the gate side. The gate is always separated from the channel by oxide layer and therefore there will be no current flowing in this direction. So V by I will be infinitely large and therefore your input impedance will be infinitely large. So in ideal cases, $Z_{in}$ equals to infinity and $Z_{out}$ equals to 0, right. So this is your static design.

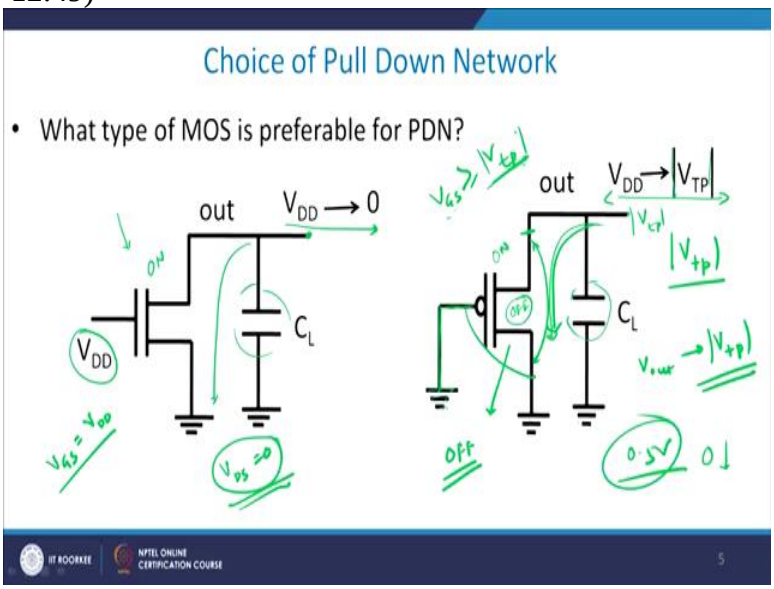
What is a dynamic design? We will look into dynamic design later on but just to give you a definition in thoroughness. This relies on the temporary storage of signal values on the capacitance of high impedance nodes, right? Now this is quite interesting that in a dynamic design, you hold the value of the voltage at the output node through $C_L$, right till the time when you are not changing, means you need to change it externally, right. So you will be storing some amount of the time the output side, right.

When you are storing the data, it has to be minimum stored till a point till your new sets of inputs do not arrive, right. So typically, it is a floating node sort of node in a dynamic operation. Now the design which you see in front of you in this slide is basically known as a complementary logic, right. This is widely used static; the most widely used static logic design is the complementary design, as you can see here. Complementary means, you will have $PMOS_S$ in the pull up and $NMOS_S$ in the pull down and complementary why?

Because if you give input high, output will be low; if you give input low, output will be high. So they are complementary of each other, a basic CMOS static. So what we have learned across this, static where output node is connected to either $V_{DD}$ or to $V_{SS}$, having your output impedance almost equal to 0, dynamic is a floating node and where the charges and the voltage is stored in the output side till a new set of inputs are not given to it. Most of the time, we will be discussing about the static design only, right. Okay.

Let me see why should we therefore choose a pull down or, what is the reason for choosing pull up and pull down network? Now, I have been telling you that all my pull up networks should always be made up of PMOSs and all my pull down networks should be always made of NMOSs and there is a specific reason why we do like that, right. The reason is something like this and I will tell you the reason in front of you. Say for example, your, this is your NMOS right.

(Refer Slide Time: 12:49)




This is your NMOS, let us suppose and you give a high-voltage $V_{DD}$ from the gate side. So what will happen is, this will switch on and the output voltage here will be dragged to 0 and therefore $V_{DD}$ goes to 0, right. There is no problem at all. $V_{DD}$ will go to 0 and what happens to your, finally what happens? Your $V_{DS}$ becomes equals to 0 because this is grounded. So $V_{DS}$ is equals to 0. Why? Because source is grounded, drain is also grounded, finally, you get $V_{DS}$ equals to 0 and $V_{GS}$ equals to $V_{DD}$.

So when $V_{DS}$ equals to 0, finally I will get no current and I will get output equals to 0 and this is stable condition. Let us put a PMOS here and let us see how it works out. See, what happens in this case is that suppose you, because to switch it on, you have to ground it, right because $PMOS_S$ require grounded to be switched on, $NMOS_S$ require a positive voltage to be switched on because of threshold voltage.

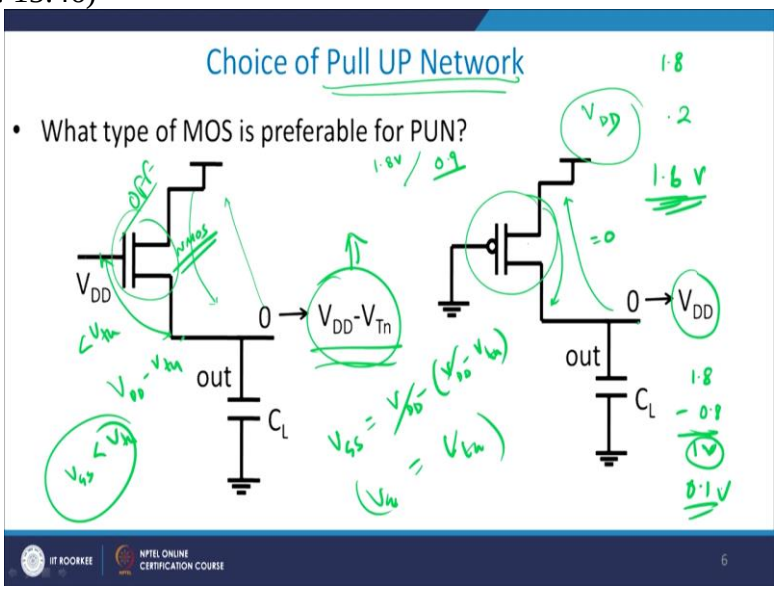
If you apply a bias here which is grounded, this is switched on, agreed with you and therefore the voltage at this point will start to fall down, right? Who is storing the voltage? This capacitance is storing the voltage in both the cases. So this voltage will start to fall down but here comes the big issue that just as the voltage here, right becomes equals to mod of $V_{TP}$, $V_{TP}$ is the threshold voltage of the device, this device goes into cut-off. Right?

And that is the basic problem area of a PMOS that, whenever my V out reaches approximately equals to mod of $V_{TP}$ right, this ensures that this is switched off. And as it switches off, you will automatically, you can understand the reason why. Because if this is mod of $V_{TP}$ the difference between these 2 is mod of $V_{TP}$ but understand, for device to be in the on state, $V_{GS}$ should be greater than equals to the threshold voltage, right? That was what my basic definition is all about.

So gate to source voltage is 0 but I have a mod $V_{TP}$ here which is basically a fixed value. It cannot be, never be greater than mod of $V_{TP}$ and at mod of $V_{TP}$ this switches off and as it switches off, my output voltage latches to a value equals to, just equals to mod of $V_{TP}$ which means that my output voltage is not able to go directly to 0 but it only latches to a value equals to mod of $V_{TP}$ which means that let us suppose the PMOS has got the threshold voltage of 0.5 volts, then rather than the output going to 0, my output will actually fix to 0.5 volts.

So I will not be able to get the whole swing from $V_{DD}$ to 0. That is the reason we use a pull up, we use a PMOS for all practical purposes. Similarly, let us come to pull down network.

(Refer Slide Time: 15:46)




Sorry, in a pull up network. We were doing pull down, we will do pull up now. In pull up, let us suppose, I have a PMOS right, I have a PMOS grounded again, this is connected to $V_{DD}$ right? There it was 0, here it is connected to $V_{DD}$. Now my, this is initially 0, this is switched on. This voltage rises to $V_{DD}$, right and goes to $V_{DD}$. Now as it goes to $V_{DD}$ right, the idea is that this will actually, can eventually go up to $V_{DD}$. Why? Because then the $V_{DS}$ will be equals to 0.

And that will ensure my device will be switched off and therefore it will be switched on only at $V_{DD}$. So which ensures that the PMOS will actually raise my voltage at the output side to $V_{DD}$. So there is a full swing available to me. Whereas in this case, if we take NMOS, then when this is switched on and I have 0 here, if it is switched on, again a low resistance path is there and therefore this voltage will start rising. It will rise, agreed but it will rise to a value equals to $V_{DD}$ minus $V_{TN}$. Why?

Because if this is $V_{DD}$ minus $V_{TN}$, right and then $V_{GS}$ if you find out, this is equals to $V_{DD}$ minus of $V_{DD}$ minus $V_{TN}$ which is nothing but $V_{TN}$ because this will get cancelled out. So I get $V_{GS}$ equals to $V_{TN}$. Now if the voltage rises above this, then $V_{GS}$ becomes less than threshold voltage of the device and this device is switched off. Exactly the same as I discussed in the previous slide. Are you able to get the picture?

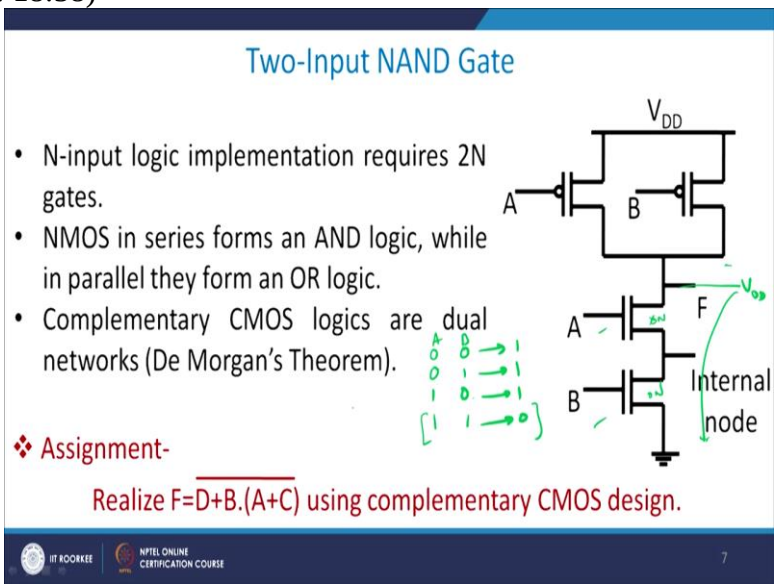
See, so the idea therefore is the output has to go from 0 to $V_{DD}$ but as it goes to $V_{DD}$ minus $V_{TN}$, this device which is NMOS here, will be cut off, will go to cut off because your $V_{GS}$ becomes

less than $V_{TH}$ as the output voltage goes above this and therefore the output voltage, this voltage latches to $V_{DD}$ minus $V_{TN}$ and which means that if I use a NMOS in the pull up network, my output will be only latched to $V_{DD}$ minus $V_{TN}$.

So if you say, $V_{DD}$ is 1.8 volts and $V_{TN}$ is 0.2 volts, then I will be only latched to 1.6 volts and I will not be able to go to the whole swing attached to me, right? We will see later on that if you are not using a whole swing, there is some problem, right and you are very close to $V_{DD}$ by 2. I will give you an example. Say 1.8 volts and your threshold voltage of the device is 0.8. So 0.80 if you subtract, I get 1 volt right. Now typically if we use a 1.8 volt supply, right your switching thresholds are 0.9, which means that when the voltage, input voltage crosses 0.9 above, it is read as input 1.

As it goes below 0.9, it is read as 0 but if your voltage itself is 1 volt, right then just you have a difference of 0.1 volt, you see. A small noise voltage, some small change in the power supply will switch on from on to off state and that is the reason, it is very critical that you use the whole swing from $V_{DD}$ to ground right. And that is the reason we use a PMOS at the pull up stage and we use NMOS in the pull down state.

(Refer Slide Time: 18:58)

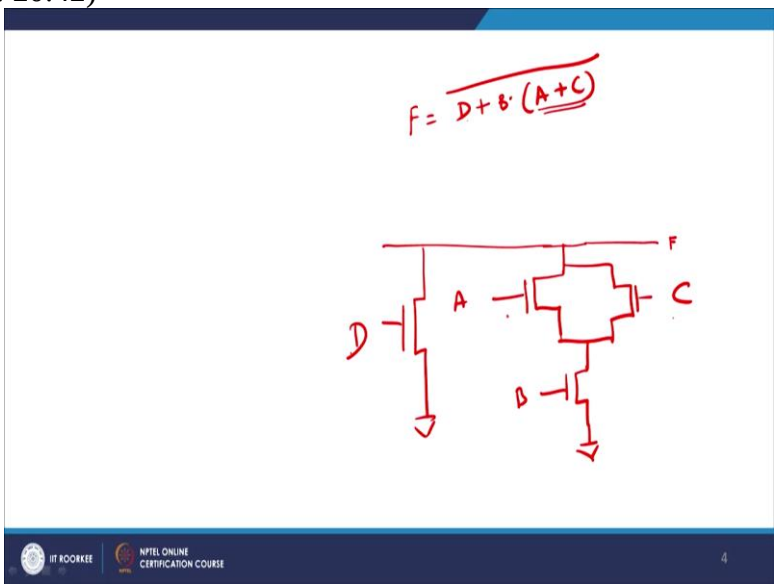



Let us look at a two input NAND gate and I suppose you all are aware of it. It is pretty simple. If you look at the logic of NAND, it is something like this. If I have got two inputs, right, let me put it like this right, then this will be, sorry 0 1. Let me do it once again. Let me just rub this off.

So I get 0 0 1 1, I get 0 1 0 1, I get 0 0 will give me 1, 1 0 will give me 1 right and this will give me 0 which means that whenever you get one, these 2 will be on right and therefore this output which is F will go to 0 and therefore output will be 0.

For all other cases whenever you get 1 0, the PMOS is on and even if your one NMOS is on, since they are in series, you have to make them both on together. So either A or B is 0, I will get pull up network to be on and my output will be latched to $V_{DD}$. Only in the case, when your 1 1 case, I will get output equals to 0 because both your $NMOS_S$, A and B are switched on simultaneously, fine? And that is what is written all these are written here in this nature.

(Refer Slide Time: 20:42)



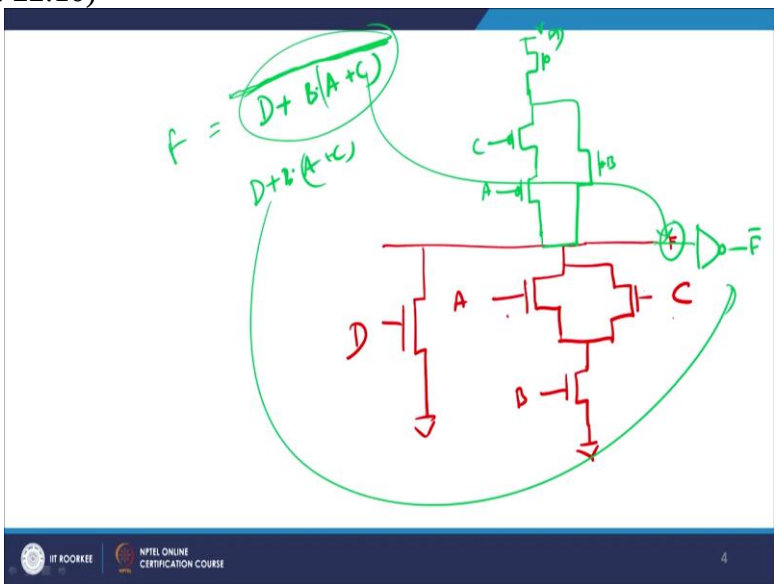

Now let us see, how we can actually make for example, realize, say for example this. The methodology how you realize our combinational logical block in a CMOS circuitry right? Let me say it is D. So it is basically your D. So F equals to D, right? It is equal to D plus B dot A plus C complementary, right? So I have to design this one. So what you do is that, you first of all design, let us say pull down. So pull down, so to do the pull down, I make F here which is the output and then I go pull down.

See, A plus C basically means that you should have 2 gates in parallel to each other, because this is a OR combination. So I should have 2 gates in OR combination and I get something like this, right. So they are OR condition here. So this is your A and this is your C because either of them

is 1, the circuit works right? If this is 1, this is 1, there will be a path available to me. Now D is in series to because this is an ANDing here.

So I will have a B here, right? And then D is ORing with this whole thing. So D is ORing primarily means that, I stop here and then D is ORing means I will have D something like this. D is OR here. Fine. So I have A plus C dot B OR D, right. Now if I have to use it say for example, now this is a pull down network right. This a pull down. Let me design for you the pull up network. Please understand, since this is a complementary logic right, since this is a complementary logic, I would expect to see that the PMOS should be complementary of NMOS which means that wherever you get OR gate, you do AND gate.

(Refer Slide Time: 22:10)



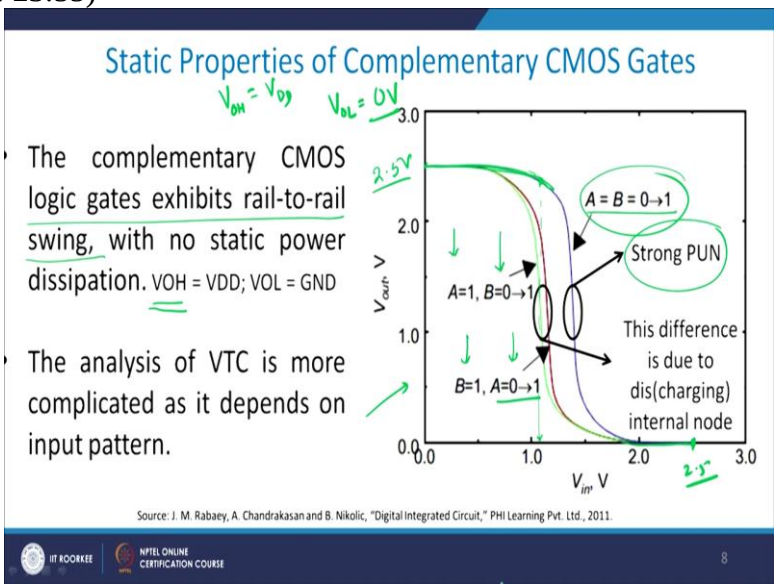

So you just start from here and then make one NMOS here which is A and then make C also in series to this point, right and then since B was in series to A and C, you make something like this. You will make it B like this. B and D was in parallel, so D should be in series again here. So I will do like this. I am sorry. And all will be $PMOS_S$, so all will be $PMOS_S$. So I can have something like this. So this is $V_{DD}$ and this is ground.

So this is how you realize this D plus B dot A plus C. And why do you get a complementary automatically? Because this is a CMOS structure, right? So if you want to get back the value of F, let us suppose F would not have, did not have this complementary signal, then you need to put a static inverter here and this will give you F bar and therefore, since this is F, you will get F bar

and therefore D plus B dot A plus C will be available to you at this particular point and this will be available at this particular point.

So what we have discussed is, we therefore know that by using certain techniques, I would be able to generate a combinational logical block, right. And this combinational logical block will have these networks available to me in which I can place these networks right and I can have these, such type of designs. So I will recommend that you take any of the standard Boolean expression and try to realize using CMOS switch logic here. For example, this one which we have already done in the class.

(Refer Slide Time: 23:55)



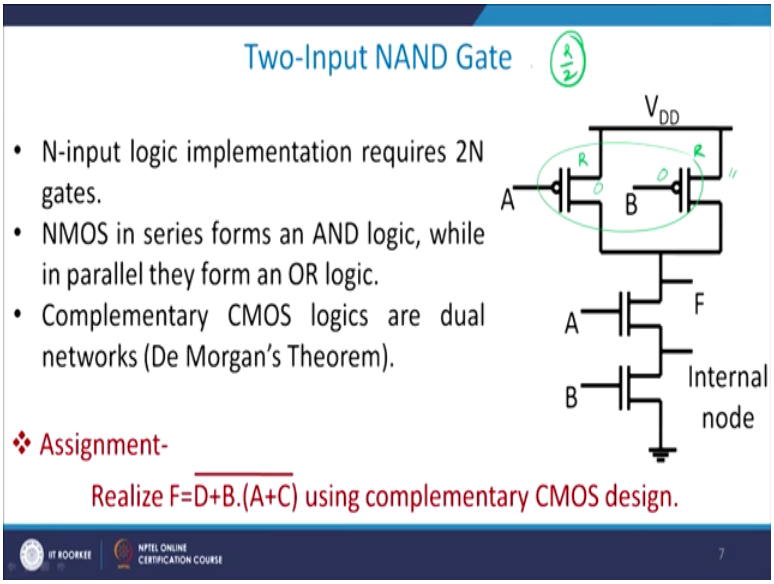

Now look at the static property of CMOS inverter. Let us suppose, as I discussed with you, CMOS inverter logic gives you a rail to rail swing. Rail to rail means $V_{DD}$ to 0. So the 1st, the power rail is $V_{DD}$ and you have a ground rail. When I say rail to rail swing, I mean to say the output goes from $V_{DD}$ to ground, right? And this is one of the important merits of static CMOS logic design that you are able to put your output swing up to $V_{DD}$ and down to ground, the whole limits you can do.

So therefore, my output high, $V_{OH}$ means $V_{OH}$ is output high and $V_{OL}$ is output low. OH is output high and this is equals to $V_{DD}$. The output low is basically equals to 0 volts in this case. Now let us see, that you did have but there is a problem here and the problem is, I will just point out from

this figure which you see in front of you, this one. Let us suppose, A and B for example you had a maybe a 2 input NAND gate right and I had A and B both equals to 0 and both goes to 1.

So A and B were both equals to 0 and let us suppose A, sorry A and B were both equals to 0 and both goes to 1. So A goes to also 1, and B also goes to 1. Now when both were equals to 0, please understand, in a 2 input NAND gate, if we look back, the two input NAND gate which was this one, your two PMOS$_S$ were in parallel, right? Let me erase it.

(Refer time slide 25:34)




These two PMOS$_S$ are in parallel which means that if both are 0, both are on and therefore if each carries a resistance of R, then the actual resistance is actually equals to R/2. So the resistance between F and $V_{DD}$ rather than being R is now R/2 which means that I can pull up the voltage here to $V_{DD}$ in a much easier manner. Agreed? So with this statement or with this knowledge, see what happens.

It means that when both A and B are equals to 0, my output was equals to 1, right. But you see, the output remains 1 for a larger value of your V in. You see. From here, it starts to fall down, somewhere here, around 1 volt. You are using a $V_{DD}$ of let us suppose, say 2 volt, so around beyond even 1 volt you are able to sustain the input voltage to be equals to, the output voltage to be equals to $V_{DD}$ and therefore this is known as a strong pull up network. Why?

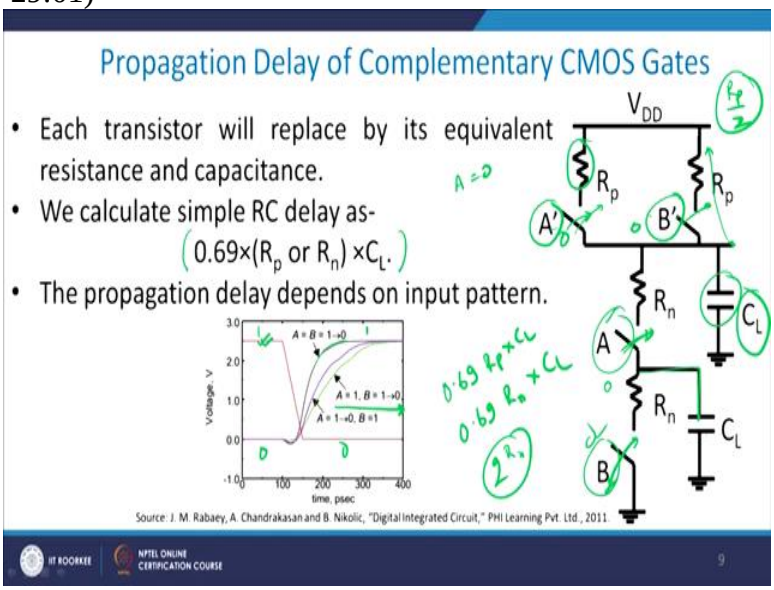
Because both your PMOS$_S$ are in parallel and therefore the effective resistance goes down and you are able to pull the output voltage at F to $V_{DD}$ in a much, much better manner. And therefore you see that output remains at $V_{DD}$ for a larger duration of time. So I am using a $V_{DD}$ of 2.5 volts. So this is approximately, it goes up to 2.5 here. So both sides, 2.5, right, so your pull down networks are relatively weak in because they are in series.

So I get R and R. So I get 2R as the effective resistance seen downwards and therefore it is very difficult to pull it down and therefore you see, the 0 remains only till 0.5 volts, where as 1 remains till more than even 1 volt, fine and that is the reason you say it is a very strong pull up network. Similarly but you see here very interestingly that for the condition when A equals to 1 and B goes to 0 to 1 and B equals to 1, A goes to 0 to 1, in both the cases, the structure looks approximately the same right? A 1, B goes to 0 to 1 implies that your pull down network gets activated whenever B goes from 0 to 1 and A goes to 0 to 1 implies that it gets activated whenever A goes from 0 to 1, right.

Please understand a two input NAND gate, either of the inputs should be 0 for pull up network to be on, right. Whereas for the pull down network to be on, both the inputs should be equals to 1. So you see, when both the inputs are 0, your output resistance offered is R by 2 whereas if it is only one 0, then you still have a pull up path but the resistance is R only. So it is difficult to pull you up, pull the voltage up from F towards $V_{DD}$ and that is the reason I wanted to show here.

So it is a bit complicated in the sense that you are not, your output voltages, though it gives you a rail to rail swing but it is not input dependent right and how you, how the input transition takes place, the output depends upon that value as such, right. Let me come to the concept of propagation delay in a CMOS inverter and then we can switch our grounds.

(Refer Slide Time: 29:01)




Now what we do is that, we replace each transistor by a switch and a resistance. So whenever my device is in the on state, this will, whenever my A equals to 0, this will go down. So therefore I write A complement and B complement. When I say A complement and B complement, I mean to say 0 means the switch is on and 1 means the switch is off. Whereas for A and B, when A is equals to 1, it switches on. A equals to 0, it switches off.

So I have a load capacitance here, I have a load capacitance here. Now this is the intermediate capacitance here and is known as the load capacitance. Now we can calculate a simple RC delay from your 1st order because these are all first-order circuits. So from your network theory basic course, you can calculate the simple RC delay to be equals to 0.69 $R_P$ into $C_L$. why? Because this is a $C_L$ and this is a net $R_P$ when you design your external network.

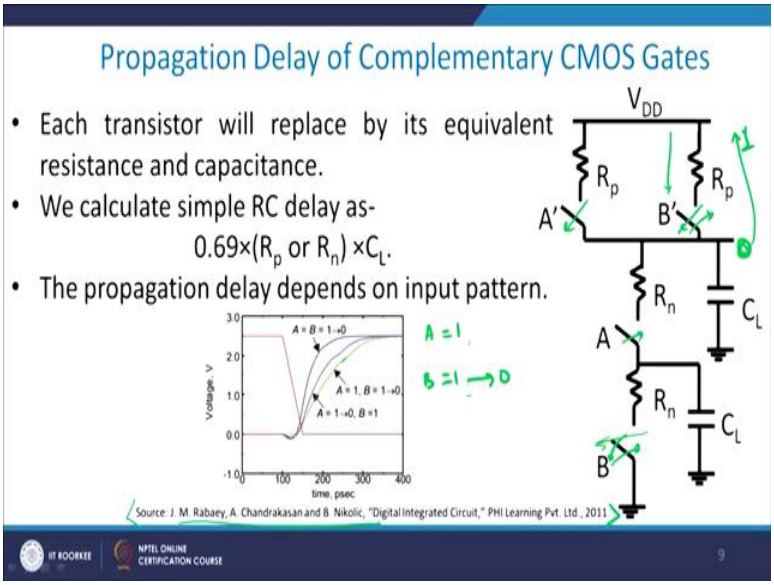
Similarly in the pull down network, it will be $R_n$ into $C_L$. $R_n$ is the resistance looking downwards, right. Now of course, I will let you know why the propagation delay depends upon the input pattern right. So this is my input, right and this is my output. So when input goes from 1 to 0, the output goes from 0 to 1. So this is 1 to 0 here and this is from 0 to 1, right. So I mean complementary in nature.

Now when both A and B were equals to 1 and both goes to 0, you see the time taken to go to the higher node is much faster and you can understand why. See when both were equals to 1, both these transistors here and here were on, right. And you had these two resistances is coming into

series. So I get $2R_n$ effectively value. Now what has happened is suddenly, you have made this go to 0, sorry, this 0 and this is also becoming 0. When these two become 0, this switches on and therefore both the transistors help you to pull up the voltage at this particular point to $V_{DD}$.

And since the effective resistance $R_P$ will be $R_P$ by 2, I will, so this at this point the delay will be half almost and therefore you are able to switch on the value of voltage very fast to one value. This is quite interesting that therefore, your output for a static logic depends upon the input transition also, right. Similarly, if you have, I will not go into details of the 2nd one. I leave it as an exercise to you.

(Refer time slide 31:48)



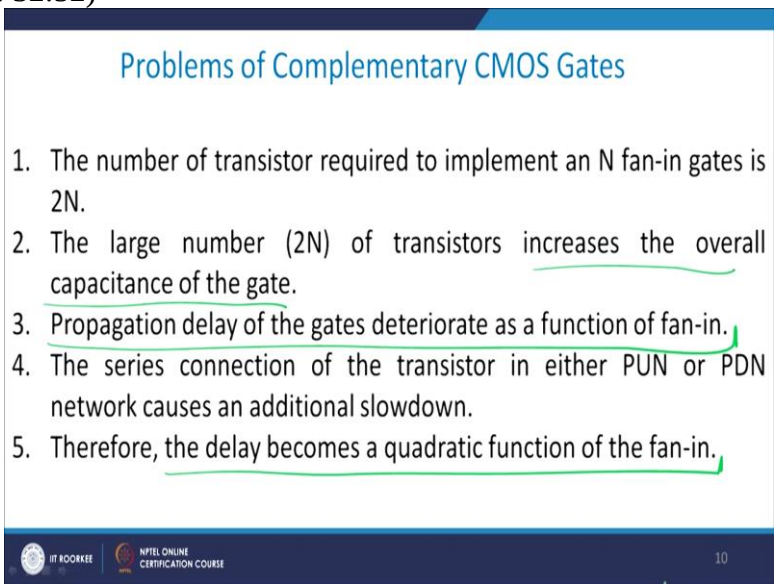

For example, when A was equals to 1 and B goes from 1 to 0 right, when A and B both were equals to 1, your pull down was on, if this was 0. Now what has happened, I am keeping A equals to 1 but I am moving B equals to 0 which means that A equals to 1 means this is on, this is off now and when this is, so let us look at this particular point that let us look at A equals to 1 and B equals to 1 to 0. So when A and B both were equals to 1 which is this one in the on state, so this was on, this was on, this was off, this was off and you had output equals to 0.

Now what has happened? B is equals to 1. B is equal to 1 means B goes from 1 to 0. 1 to 0 means this opens and this closes, right? And therefore, only one path is available you to pull down. There is no pull down path because you have switched off this. So the circuit is broken

down in the pull down. But your pull up path, only one resistance is switched on and therefore this goes from 0 to 1 right. This, the violet line which you see is basically, sorry the green line which you see is basically 0 to 1.

Similarly, when 1 to 0, you see the time taken is relatively smaller as compared to the previous one. I will leave it as an exercise for you to find out. You can get all these discussions in this book by Nikolic, Chandrakasan and Rabaey, one of the standard books which we use across networks, across the whole digital, CMOS digital logic design. You can get the results from this one.

(Refer Slide Time: 32:52)




Let us look at the problems of complementary CMOS. Therefore the problem with complementary CMOS logic is that for N input, you require at least 2N Gates. That we have already seen. For example, two input NAND gate, I require two $PMOS_S$ and two $NMOS_S$. For 3 input NAND gate, I will require three $PMOS_S$ and three $NMOS_S$. So for an N input NAND gate, NOR gate, I require 2N number of gates and therefore the area is relatively large.

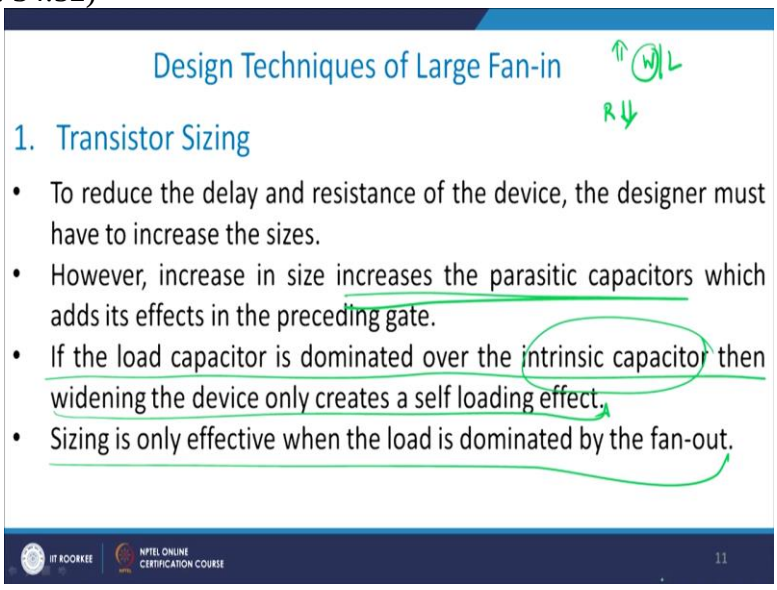
Larger number of gate also implies that the overall capacitance is also large because when you have larger number of gates connected to the output, the output sees a larger capacitance which is there with you and therefore, time taken to charge or discharge the capacitance also rises and therefore delay becomes larger. And therefore if you go from a NAND2 to NAND3 logic, your delays will be typically larger in that case, right.

Propagation delay of the gate deteriorates, that is what I was saying. The propagation delay of the gates deteriorates as a function of fan-in. So was the fan-in starts to rise, means you, as large number of gates becomes is available to you, you end up having a larger propagation delay and you can understand why? Because your loading, load capacitance which is in the output side starts to become higher and higher.

As I discussed with you, since pull up and pull down networks are complementary with respect to each other, we have to ensure that so if let us suppose your pull up network is in parallel, your pull down network will be in series. If the pull down network is in series, then the overall resistance will be the sum of individual resistances, right? $R_1$ plus $R_2$ plus $R_3$. As a result, the overall propagation delay will be much, much higher because the resistance is higher.

So you are paying both in terms of higher resistance pull down network as well as higher load capacitive load. And therefore as you can see, the delay becomes a quadratic function of fan-in. And you can understand why assuming quadratic? Because it is once you are paying for resistance, another one you are paying for capacitance and that is the reason, these are the problems of CMOS logic, right.

(Refer Slide Time: 34:52)




Let me therefore, as I discussed with you, so let us see what are the design techniques available to you for large fan-in circuitry, right? What are the design techniques generally available to you for large fan-in. One is that simply increase the size of the transistor. So if you increase the size

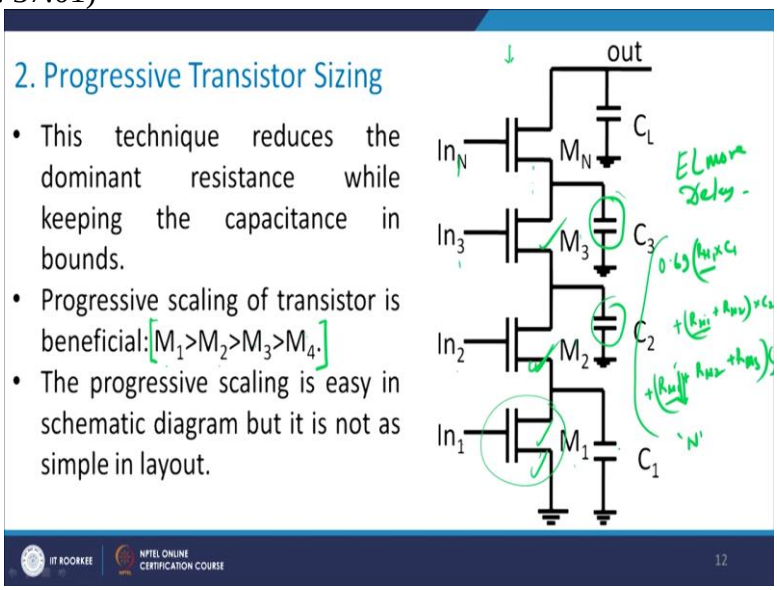
of the transistor, W by L increases. You increase the value of W, right the area goes on increasing and therefore the resistance starts to fall down and you automatically have a lower delay.

However, please understand when you increase the value of your W, you also end up paying the price of increase in the parasitic capacitances, right. So though you reduce the resistance of the gate, single gate but you add up to the parasitic capacitances of the previous gate and that makes it a bit, slightly difficult to design. That is what I was saying that if the load capacitor is dominated over intrinsic capacitor then the widening of the device only creates a self loading effect.

See, generally a device has got a self capacitance. For example, a $C_{oxide}$, oxide capacitance of the device right. It depends only on the area of the device. So epsilon A by $t_{oxide}$. So if we increase the W, area increases, oxide capacitance increases. That is known as intrinsic capacitance but you also have load capacitances, right. Load capacitances by virtue of a $C_{GD}$, $C_{GS}$ which is basically the overlap capacitances and so on and so forth.

Now if your those capacitances are higher, load capacitances are higher, then increasing the size, if you want to increase the size, you are actually increasing the value of $C_L$ rather than increasing the value of intrinsic capacitances. And that makes the life difficult as such. And therefore sizing is only effective when the load is dominated by fan-out. So which means that if you are driving a larger number of such devices which is higher fan-out, then only sizing is an important issue or you need to look into sizing issue but if your fan-in is very large right, the sizing will not help you too much as far as designing is concerned for lower propagation delay.

(Refer Slide Time: 37:01)




Now, let me give you a concept of progressive transistor sizing. See, if you look carefully at the pull down network which is shown here in this case, so when both in 1, 2, 3 and 4, let me say this is In N, there are N number of logics. So there are $M_1$, $M_2$, $M_3$, $M_N$. So there are N number of gates available here. Now suppose all the gates are having 1 1 1 1 1.

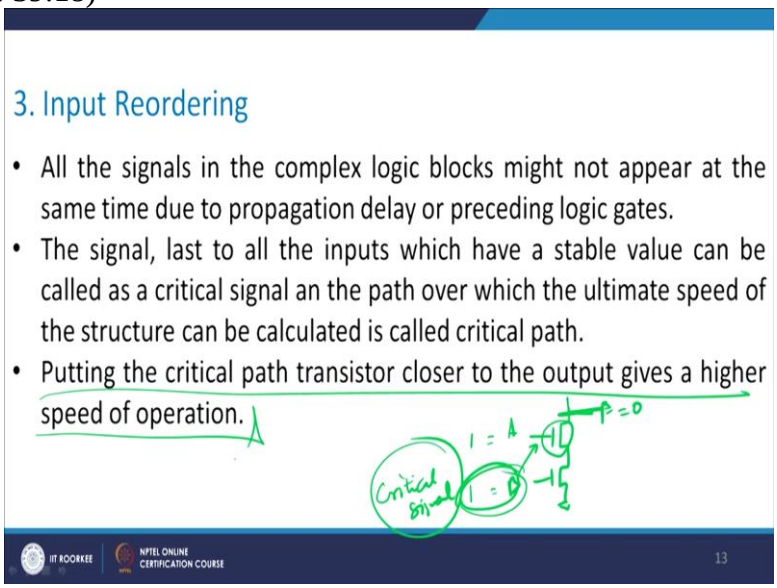
All the $NMOS_S$ are on, then you can see that if you look at the delay, I get a 0.69 right multiplied by $M_1$ which is the this one multiplied by $C_1$, resistance of $M_1$ multiplied by $C_1$ plus resistance of $M_1$ plus resistance of $M_2$ into $C_2$, right plus resistance of $M_1$ plus resistance of $M_2$ plus resistance of $M_3$ into $C_3$. You are getting my point? This plus this plus this, this plus this plus this multiplied by this.

This plus this multiplied by this and this multiplied by this, this is also known as Elmore delay. So you see, this $RM_1$ is coming thrice and if there are N number of gates, you will get $RM_1$ N number of times. So, it is always advisable to make the lowest gate right? The lowest gate means this, you got it in series, the largest gate. You make it largest, this resistance will be the smallest and therefore you go on adding it, you get the smallest delay with respect to the overall delay.

So it is always advisable that try to keep the aspect ratio of the gate which is farthest away from the input largest. Once you ensure that, your overall delay starts to reduce and this is what is known as progressive scaling in transistor. So $M_1$ is greater than $M_2$ is greater than $M_3$ is greater

than $M_4$, right. Though it looks very simple on a pen and paper, in actual layout it becomes very difficult to achieve it.

(Refer Slide Time: 39:18)



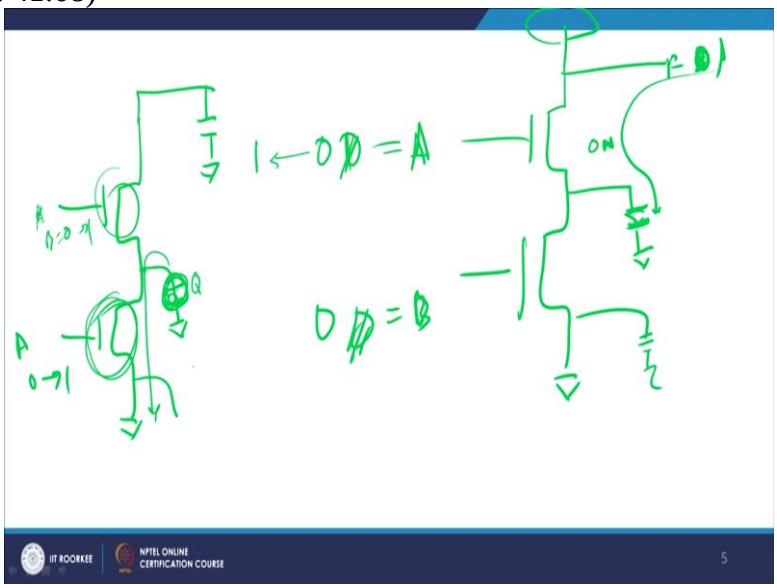

What is known as input reordering? I will just use a single term here. You can go into it later on. Now try to make, see whenever you have a combinational logic, you will have N number of signals coming together, right? For example, I will just you a small example, let us suppose I have got these 2 signals here right and I have got this F here and this is my pull down network. So there are 2. I have got A and B here. Now let us suppose A arrives earlier.

So even if A arrives earlier, this switches on but still, suppose A equals to 1 and this switches on, my B is equal to still 0, right and it will go to 1 but it will go to 1 at a later stage. But till it goes to 1, my F cannot go to 0 because there is no direct path between F and ground. Are you able to get the picture? Which means that, you have to wait till the B signal comes into picture, right. So we define B signal to be as the critical signal.

So what is the critical signal? Critical signals are those signals which are the last to arrive and which are, when they come last to arrive, they evaluate the output also last. So they help you to, so you have to wait till the critical signal arrival for the output to go to 0 or 1, right? The rule of thumb is and there is a reason for that that you try to keep this, if B is the critical signal, try to keep it closest to the output. So putting the critical path transistor closer to the output, gives a higher speed of operation.

This is a rule and there is a reason for that. You please find out yourself what the reason is, but this is typically the reason that the critical path signal should be kept closer to the output. It has to do with the discharging, output discharging, right. I will tell you how, I can give you a reason for that.

(Refer Slide Time: 41:08)



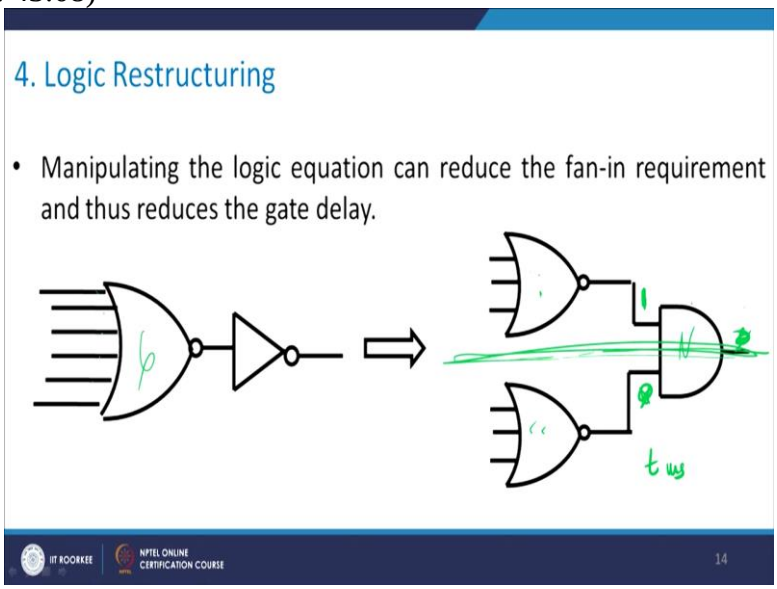

Let us suppose I have this as the consideration and this is A and this is B, right? This is B and this is, so this is your F and this is your pull up network, something is here. Let us suppose B is the critical signal, right? A is the non-critical signal. So when A becomes equal to, both were 0 and output was equals to, so both sorry, I am sorry. Both were equals to let us suppose 0, output was equals to 1 because both were off. Now A goes from 0 to 1 but B still remains 0 and it has to wait till then.

So when it goes to 1, this becomes on. This becomes on means that, so you have a capacitance here and a capacitance here also. Some amount of this 1 will appear across this capacitance here but it has to, some amount of charge at F will appear will charge this $C_L$. But then, it has to wait till B arrives right? So what you try to do is quite interesting is, that you try to make your B arrive here and A arrive here. So A has already gone from 0 to 1 right and therefore, so this is grounded, this is there and this is with me and I have a capacitance loading here.

So A goes from 0 to 1 implies that this has switched on, all my charge, extra charge available at this point has gone to the ground and therefore, simply if this goes from, B goes from 0 to 1 now,

this becomes on, then this extra charge which was already available here and need not be discharged. It has already been discharged within the time. So you do some amount of time sharing. So the time till which your critical signal was not appearing to you, you discharge the extra capacitance or discharge the extra charge at the drain end of my transistor A. And therefore, you need not therefore discharge large charge. You have to discharge a lower charge and therefore your speed becomes larger, right.

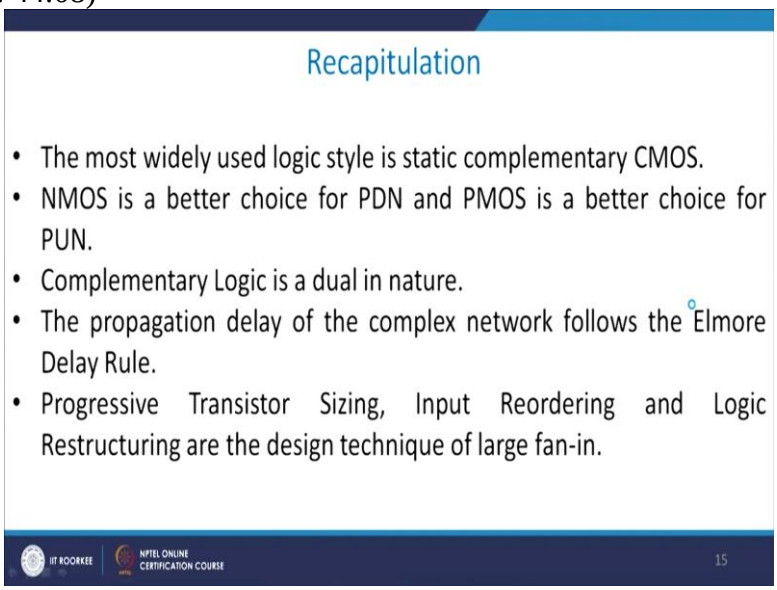
(Refer Slide Time: 43:08)




Again one methodology is that logic restructuring methodology which is logic number 4. You have to manipulate it in such a manner that you try to make it as symmetrical as possible. So when you make it symmetrical, the delay, so typically let us suppose it was not symmetrical and let us suppose, this reaches 1 and this reaches 0 right, so output will be 0 right because it is an AND gate. But if this is 1 and if this is let us suppose, reaches some delay by say $t_{ms}$, then this output has to wait till $t_{ms}$ for the output to appear.

So what you try to do is, you try to make it symmetrical in nature. So, for example, 1, 2, 3, 4, 5, 6. 6 input NOR gate can be broken up into 2-3 input NOR gate and 1 NAND gate. So it is a very high fan-in, not a very good idea to manipulate, break it into 2 smaller fan-in transistors and you try to make it more symmetrical in design. And that is known as logic restructuring, right?

(Refer Slide Time: 44:08)




So let me recapitulate what we done till this module. The most widely used static module design is CMOS logic. As we discussed just now, the NMOS transistors are better pull down networks, PMOS devices are very good pull up networks, they are part of pull up networks. Complementary logic is very, is dual in nature. Please understand, this is very, very important when you are designing combinational logic. But whenever you are doing a complementary logic, your pull up network and pull down network are dual of each other which means that if upper 2 transistors are in series, the lower 2 will be in parallel and vice versa.

So you should keep in mind and should do a large amount of practice of, so if I give you a Boolean expression, you should be able to design its logical function or the complementary logic. The propagation delay of the complex logic follows NMOS delay rule which we have discussed and therefore, it is always advisable to keep your transistor which is most away from the output, the largest in size, therefore resistance falls down and that is the one methodology.

The 2nd is that when you do, so this is known as a progressive sizing of the transistor. You also need to understand that the critical signal should be kept very close to the output for your lower delay and then you also have to do logic restructuring. Try to keep your fan-in low and try to make it a symmetrical path so that you do not have glitches or there is no problem in this case right. So these are the few statements or the few areas in which people have worked on. And I

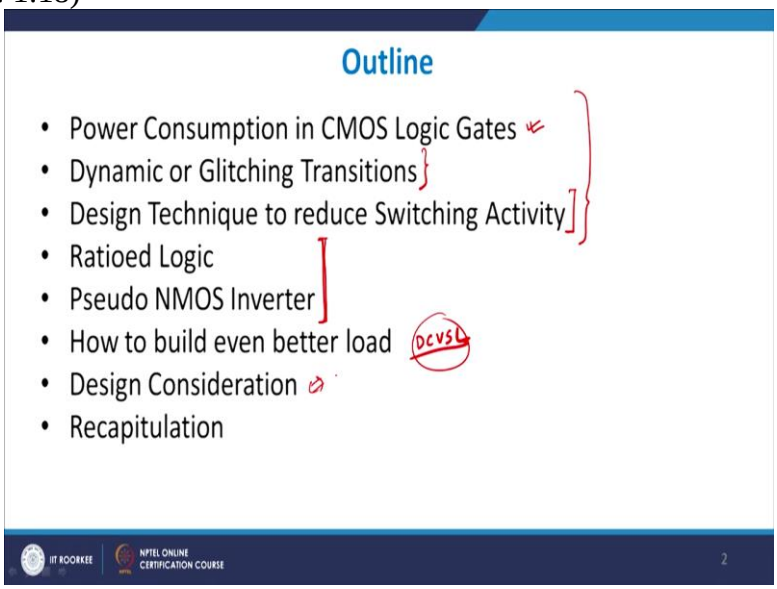
hope you have understood this module in a better manner. Okay. Thanks a lot. Thank you for your patient hearing.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−54**
**Combinational Logic Design -II**

Hello, welcome to the next edition of the NPTEL online certification course on microelectronics devices to circuits. We start with the 2nd module of combinational logical design. In our previous module, we have seen what do you mean by propagation delay and how can you reduce your propagation delay in a complementary logic. We also saw what is the meaning of combinational logic and how can you design in a combinational logic, given a Boolean expression.

So if there is a Boolean expression available to you, how can you design it on a complementary logic? What we will be doing now is look into another important parameter and that is known as power dissipation. Let me give you the outline of what we are going to do.

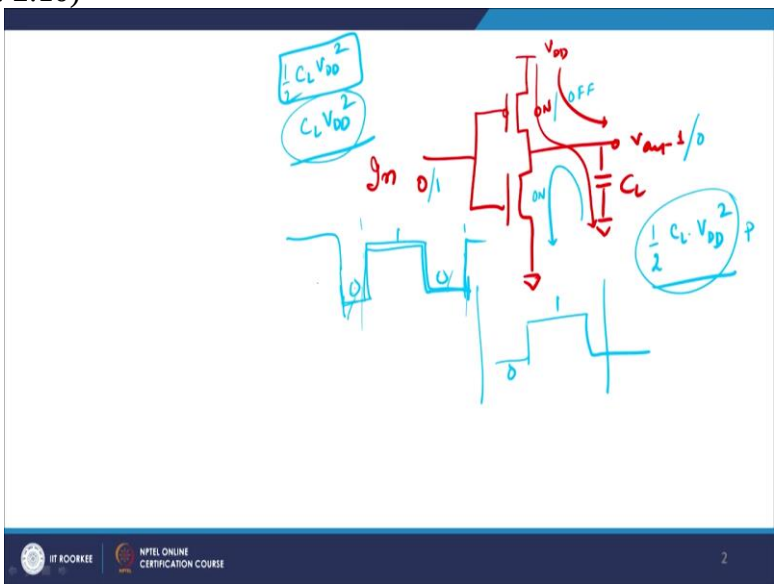
(Refer Slide Time: 1:18)




So we will look into the concept of power consumption and dissipation in CMOS logic gate. We will look into dynamic or glitching transitions and then how we can do, reduce the switching activity. And after we have understood all these basic concepts, we will take up what is known as a ratioed logic and within which we will study pseudo NMOS inverter.

And then we will look into what is known as a DCVSL which is also a better form of a ratioed logic. We will see how it works out. And then therefore, using DCVSL what are the design considerations for a DCVSL logic and then we will finally recapitulate for this particular module.

(Refer Slide Time: 2:10)



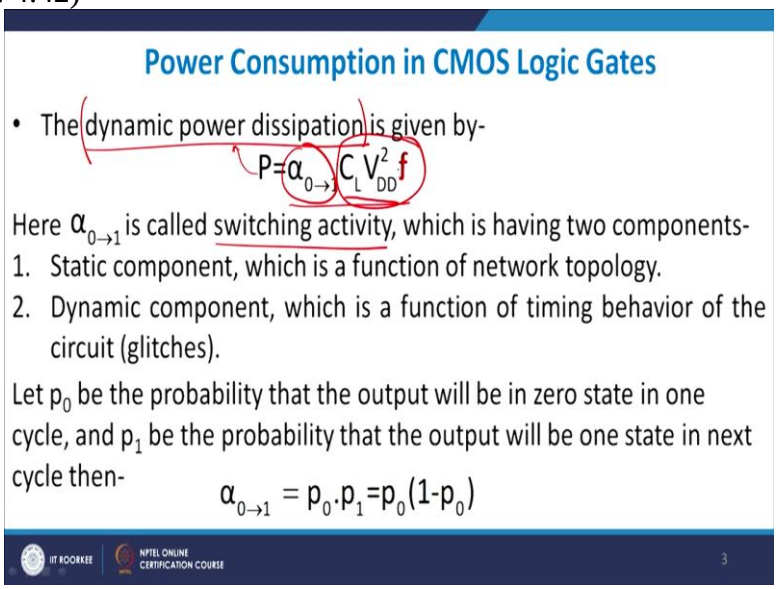

Let me explain to you before I move any forward, let me give you an idea about what is power consumption because we need to find out from basic this thing, inverter. You remember, if in an inverter, you have an input here right and then you have an output $V_{out}$, there is a load capacitance here. So whenever you have 0, this switches on as I discussed. Therefore this goes to 1 because your $V_{DD}$ tries to charge. So you have this charging path in this manner. So $C_L$ gets charged to 1.

Whenever now, whenever your input becomes equals to 1, this becomes off and this becomes on right? And when this becomes on, this charge which was there across $C_L$, this charges and this becomes 0. So this is the basic, fundamental principle of this inverter. Now you see, whenever your input is 0, you are drawing power from the $V_{DD}$ rail and therefore, you have half $C_L$ into $V_{DD}$ square as the total amount of power which you are actually extracting from the $V_{DD}$ rail over a single cycle.

So I am doing say, a 0 to 1 transition and that is all. I am doing a 0 to 1 transition, right. Now in the next half cycle, when 1 comes into picture, this switches on and the same charge or the energy or the power actually goes down to the ground. So over a cycle of, over a period of cycle or over one period of cycle, 0 to 1 transition, you actually utilized $C_L$ times $V_{DD}$ square. Half plus half because half you have taken and half you have done but then total will be, total power dissipation will be still equals to half $C_L$ $V_{DD}$ square, right.

Now with this knowledge, with this basic idea, we can say that the total power consumption over say one cycle of operation, so 0 1 0, so this is one cycle of operation. We have 2 times 0 coming into picture. So half plus half is equal to $C_L$ into $V_{DD}$ square, right. So over 1 period of cycle from this to this, this is one period of cycle right. So you are starting from here, ending here, you have got 0 0 and therefore we get $C_L$ $V_{DD}$ square as the power consumed from the $V_{DD}$ rail.

(Refer Slide Time: 4:42)

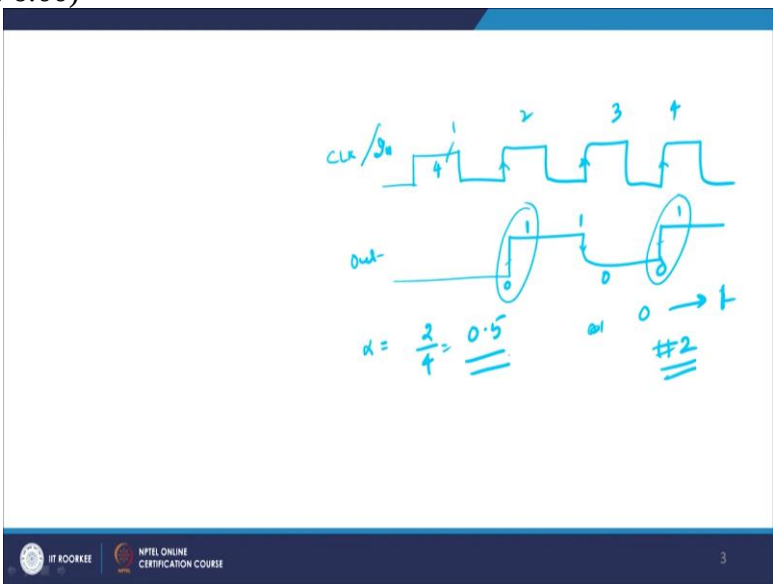



With this knowledge, what we do is we define another term known as dynamic power dissipation. It primarily means that, whenever the inverter is in operation mode, operation in the sense that whenever your input is varying from 0 to 1 and 1 to 0 and there is a large input bit train coming into picture and the output is also therefore varying from 0 to 1 and 1 to 0 we define this to be as a dynamic operation. What is a static operation? Static means that when it is given 1

or 0 and output is latched to either 0 or 1 and you have fixed that operation, we define that to be a static operation.

Then we define dynamic power dissipation as P equals, P is the dynamic power dissipation. It is given as this quantity, where alpha is defined as a switching activity or switching parameter. And $C_L$ $V_{DD}$ square we have already discussed is the power multiplied by frequency, you can understand why. Because in one cycle, you are dissipating $C_L$ $V_{DD}$ square. So if there are f such cycles available to you, this will be the total power which you will be dissipating in f such cycles. So that is the reason, you multiply with f. An interesting term comes here which is 0 to 1, alpha 0 to 1. I will explain this to you in a detailed fashion.

(Refer Slide Time: 6:00)

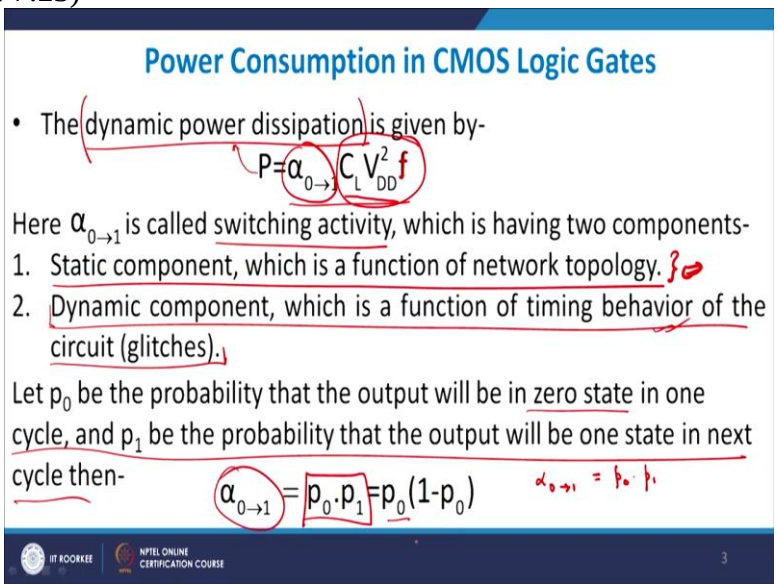



Let us see, let us say you have a clock in this manner right? You have a clock and there are 4 clock cycles which are there. Then let us suppose my output goes something like this. This is my clock or input, let us suppose and this is my output. So output I have got, so I have got something and then at this rising edge of the clock, I have a 0 to 1 transition. Similarly, at this rising edge of the clock, I have a 0 to 1 transition. So please understand, at this edge, at this rising edge of the clock I have 1 to 0 transition, right?

So please understand 0 to 1 transitions are power consuming and power dissipating sort of cycles, right. So you see, of the 4 clock cycles here, 1, 2, 3, 4, there are two 0 to 1 transitions. How many? 0 to 1. How many? There are 2 number of 0 to 1 transitions which means that we

define activity factor as equals to 2 by 4 which equals to 0.5. If there are 3 such cycles, it will be 3 by 5. If there are 4 such cycles, it will be 1 right? And then 4 by 4 is equals to 1. So activity factor is that.

(Refer Slide Time: 7:23)



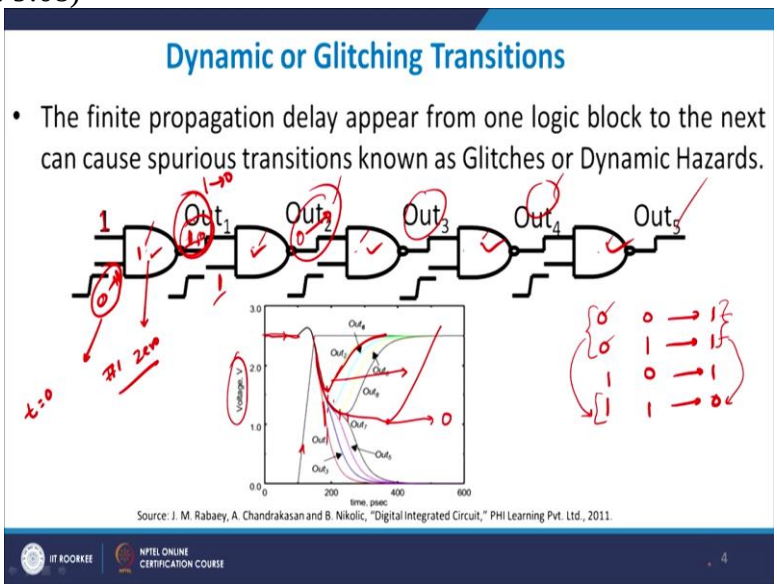

Now activity factor as I discussed with you, we will have two components, static component which is a function of the network topology. This is what I was talking about. So depending upon is it NAND to, NOR to logical or whatever logic you are using, you will have fixed number of 0 to 1 transitions for certain number of input clock period. That will determine your alpha 0 to 1 transitions. Now dynamic component which is the function of timing behaviour of the circuits.

So dynamic behaviour, dynamic component, it depends upon the timing behaviour, which means that say your clock is there but your, the input data train is there but the input data train is not same for both the inputs. There is some delay between them. Then there will be some glitches which we will discuss later on about which also alpha remains. But primarily, we will be looking into this 1st thing which is basically your static component, right?

We also define one important point that see, as I discussed with you, how do you calculate the probability? See, the issue is that the probability or alpha going from 0 to 1 is defined as the probability that one state is 0 and another state is 1. Of course, you can understand why. Because 0 to 1 means the initial state should be 0 and the final state should be equals to 1.

So if $P_0$ is the probability then the output will be in 0 state in one cycle and $P_1$ is the probability that the output will be in one state in the next cycle. Then the alpha 0 to 1 is also defined as $P_0$ into $P_1$, product of the 2 multiplications and therefore $P_1$ can be written as 1 minus $P_0$ because either you can have 0 or 1. Therefore 1 minus $P_0$ is the effective value of alpha 0 to 1.

(Refer Slide Time: 9:08)




Now let me come to dynamic or glitches, we will come back to this concept of alpha 0 to 1 later on. At this stage, you just have to keep in mind that this is basically the activity factor whose value is between 0 and 1 right? 1 is the highest value and which is basically a clock and it will go on decreasing depending on the topology which you are using it. $C_L$ is the effective load which is visible to you at the end of the, at the combinational logic block, $V_{DD}$ is the applied source voltage, voltage power supply and f is the frequency of operation.

Let us look at this concept of glitching and I will explain to you how does glitching affect your overall delay. Now, these are NAND gates here. So I have got 1, 2, 3, 4, 5 NAND gates and I am giving an input, 0 to 1 at this particular point, 0 to 1. So if you look at the NAND gate, its typical truth table will look like this, right? So suppose it is 1 right and here input goes from 0 to 1 in the input side. So initially when it is 1, 0, output will be equals to 1 and then when it goes to 1, this will glitch, this will go back to 0, right.

But this glitching takes place equally across the path. So you see, I have plotted here, output voltage here, the voltage at these particular points, out 1, out 2, out 3, out 4 versus the time and

this is my, so this is my input here, right and this is my output here. So whenever my input was low, input was low means one of them is, so whenever the input is low, means either this or this, the output is 11 and therefore output is latched to a high-value.

Input goes high and your, one of the other input was already high which means that I am going from this state to this state. So I am going from output 1 to output equals to 0. But the whole issue is this would have been true that all of the output will go to 0 together provided these did not have any intrinsic delay, right? If these did not have any delay, then this going from 1 to 0 will imply that this will happen instantaneously as 0 to 1 transition is taking place in input side. So out 1 transition of 1 to 0 will take place instantaneously as 0 to 1 transition in input side provided NAND 1, number 1 has got 0 delay. But this is not true.

And therefore, this 1 to 0 or 0 will appear in the output side of 1 only after a finite delay which is equal to the delay of this NAND gate. So let us see how does it influence my output. So you see, the output 1 goes to 0. You see. The output 1 goes to 0. This curve is basically output equals to 1 goes to 0. Why it goes to 0? Because when my input goes from 0 to 1, the output goes from 1 to 0. Now what has happened is, whenever my out, so you see, this has already gone to 1 right? Till the time when this output was equal to 0, till the time when it was, it came to 0 because initially you had 1 and then 0. 1 0 will give you 1 here.

So initially, you had 1 1 right? 1 1 means out 2 was start to falling down. So look at this, out 2 is the green one. Out 2 started to fall down. Why? Because at some value, suppose this is you did this at T equals to say, 0 milliseconds right and this has got a delay of 1 millisecond. So till 1 millisecond, this output will be equals to 1. So output will be 1 means this is 1, this is 1, this is 1, this is 1 implying that this will be always 0 and that is the reason, you see the output starts to fall to this value. It comes towards, it tries to go towards 0 but after 1 millisecond, this 0 appears.

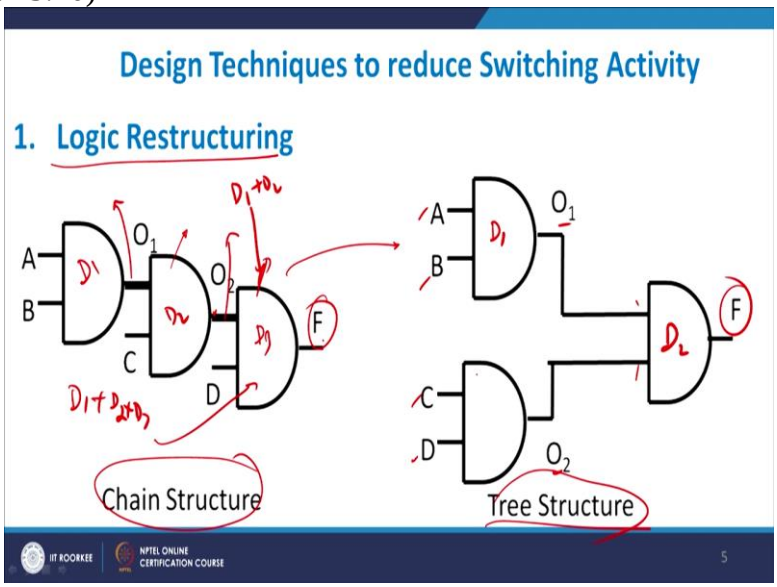
Why? Because you know how, you have a transition taking place in the output side. So out1 goes from 1 to 0. 1 to 0 primarily means that out 2 will go from 0 to 1. So now you see, the out 2 goes to 0 to 1. So you see, so similarly and so on and so forth, if you go in this direction, out 2, out 6, out 4, out 6 and out 8, all will first show a drop and then will go on increasing and as you can see in time domain, they are shifting this wards because this delay plus this delay plus this delay plus this delay plus this delay is all adding up.

The delays get added up and therefore the time till which it goes to 0 is actually shifting towards the right, right? And this is known as the dynamic hazard or a glitch. But ideally, so ideally if this chain would have been very long, a time would have come that I would have actually seen this going down like this and then increasing. So even if it is less than $V_{DD}$ by 2, I would have seen it as 0 in reality whereas it should have actually gone to 1, right.

This is known as the dynamic transition or a glitching transition in a CMOS logic and that is one of the major important problem areas of glitching. So what the people do is something like this, that they try to do what is known as logic restructuring. So, how to reduce therefore the switching activity, right? Switching activity, remember was actually sorry, switching activity is actually equals to 0 to 1 transition. Now if you go back to 0 to 1 transitions, you see 0 to so what you are doing here, you are doing a 1 to 0 here, you are doing a 0 to 1 here, you are doing 1 to 0 here and 0 to 1, right.

So you are allowing larger number of 0 to 1 transitions and therefore power dissipation goes up, right. So this glitch helps you to increase the power dissipation, not a good idea right.

(Refer Slide Time: 15:20)



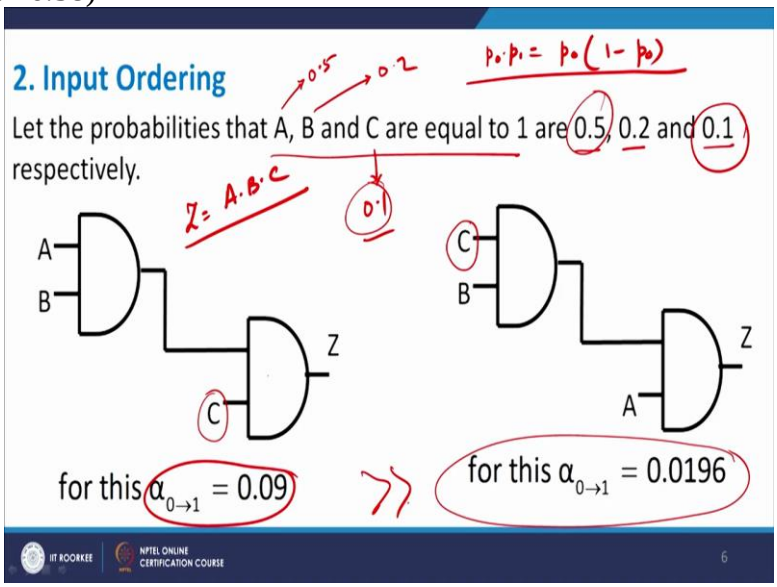

And therefore what we try to do is that we try to normalize the glitches or we reduce the glitches. How do we do it? By simply making logic restructuring and try to make it as symmetrical as possible, right? And rather than a chain structure, you do a tree structure. I will explain to you

how does it help you. In a chain as I discussed with you just now, you have to, so $O_1$ gate which is a AND gate has to wait till the delay of the 1st gate for this to evaluate from value of $O_2$.

Similarly, this has to wait, so $O_2$ has to wait till $D_1$ plus $D_2$ till this comes out and F has to wait till $D_1$ plus $D_2$ plus $D_3$. Okay, this D3 is the delay. You got the point? And therefore, these internal blocks will be glitching. And therefore, they will add to the power, total power, glitching power, right? So what you do? You convert this chain structure, this is known as the chain structure to what is known as the tree structure. The tree structure is the most symmetrical in nature which means that whenever AB is available to you and CD is available to you, $O_0$ and $O_1$ appear together and therefore this appear together and therefore F appears at exactly the same instant of time, right.

Overall delay also reduces because now you will have a overall delay of $D_1$ plus $D_2$. Initially, you had $D_1$ plus $D_2$ plus $D_3$. So the delay was much larger, frequency of operation will reduce, right. And you have a logic restructuring here which is basically a tree structure here and gives you a much lower logic, lower sort of activity here.

(Refer Slide Time: 16:58)



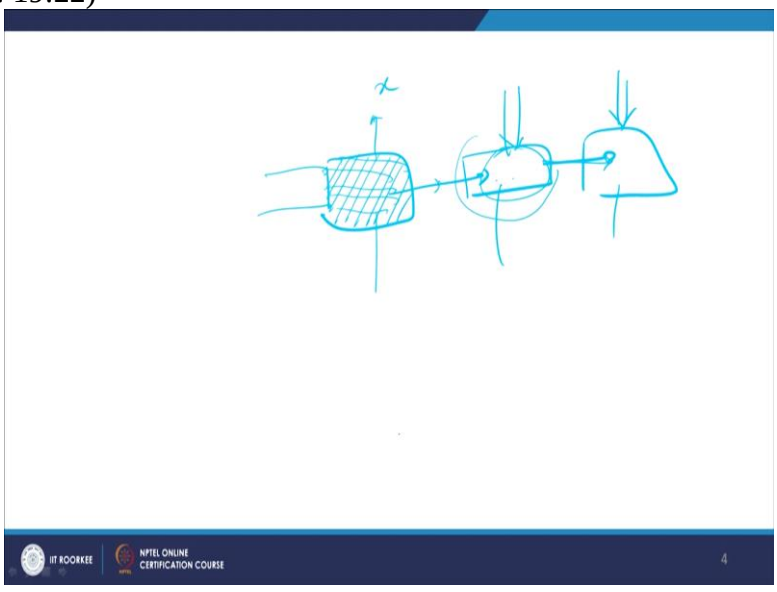

Another methodology which people adopted was what is known as the input reordering. Input reordering means that try to, so let us suppose that the probability of A, B, C equal to 1 are 0.5, 0.2 and 0.1. So A is basically 0.5, B is 0.2 and C is 0.1, the probability that they are equal to 1. Now if this C is inserted, so I have 2 logics. So if you look very carefully, Z is equals to A dot B

dot C, right? So it is a NAND logic which I am trying to realize. Sorry, AND logic, for 3 input AND logic. So that is why, if you are using a 3 input, we still can have, you can also have a 3 input AND gate but I am using a 2 input and gate, two 2 input AND Gates.

So what I do? I put A, B here and then C here. But C has got a probability of, the least probability of going to 1 right? In that case, if you try to find out alpha 0 to 1 right, from that initial P0 multiplied by P1 comes out to be equals to P0 into 1 minus P0, right? Similarly, you know what the values are. You can put the values here and you can get the value of alpha 0 to 1. But there is another method that the least value which you get, the least probability if you shift it to the earliest one and the most probability, if you shift to the later stage, you get alpha to be equals to 0.196 which is much lower than this value which means that, those inputs which have the highest probability of switching, try to put them closest to the output.

We have already discussed this point when we were discussing with you the propagation delay. Now in case of input reordering also, or input ordering also, you have to ensure that highest transition inputs which gives you highest probability of transition, should be delayed and should be placed at the last most gate for the consideration. As you can see, you have approximately 10 times decrease in the value of your alpha. And therefore, your dynamic power dissipation also goes down. Another methodology is time multiplexing the sources.

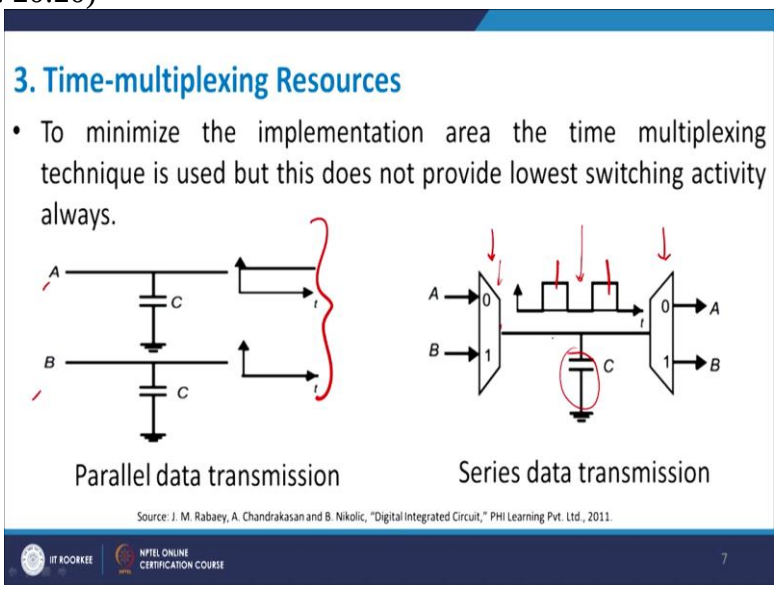
(Refer Slide Time: 19:22)



This is quite interesting that, what we try to do is that let us suppose I have 2 modules working together but, let us suppose I have got let me show it to you in this manner that I have got 2 modules working together and I have a block here and therefore they are talking with respect to each other. Now these are all driven by let us suppose the clock or even the data. Now this is behaving in such a manner that they are getting a certain intrinsic delay, say X, right. Now when the, so now you have fed an input here.

This is processing right? This is processing. Similarly, you do have some data, so this whole process give you a processed output, goes to this combinational logical block and then you forward it to this block. So what, what the concept is that till the time this is being processed, this 1st block, right, you can do some processing of the internal blocks as well. So that is what is known as the time multiplexed resource allocation. So you need to allocate resources here and try to make it lowest switching activity in this case.

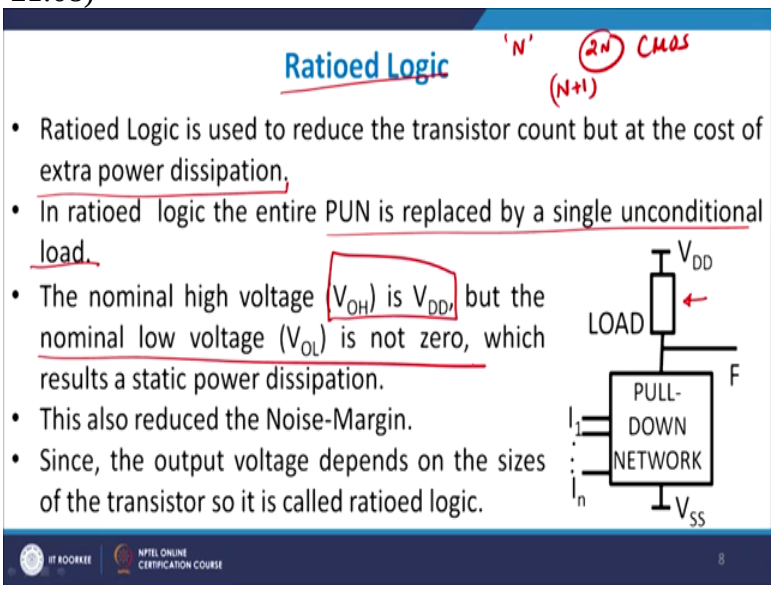
(Refer Slide Time: 20:20)




I will give you an example. For example, you have a parallel data coming into picture from A and B right and this C is the data. So I have a Mux here and I have a Demux here and this Mux serializes the data and this deserializes the data. So this is basically a serial deserialiser concept and you do have these output coming into picture here. So if we use such type of mechanism which is this one, then you allow for the capacitance to get charged and discharged and there will be sort of a talking between A and B.

As a result, there will be heavy glitches in this particular point. Whereas, in this case, you will have minimum glitch because there is only one line which is carrying the data from this Mux to this Demux. And therefore, this will give you a much lower power dissipation as compared to the previous case.

(Refer Slide Time: 21:08)




This takes care of approximately our understanding of complementary logic to larger extent and how the complementary logics can be used for a least delay and least power in a very sort of first-hand application area. We come to another one where it is known as the ratioed logic and ratioed logic primarily means that this logic has got lower transistor count as the previous one as a consideration. So if you remember, in a CMOS, P or CMOS architecture, you require for an N gate CMOS, for a N input CMOS, you require 2N Gates for a CMOS right?

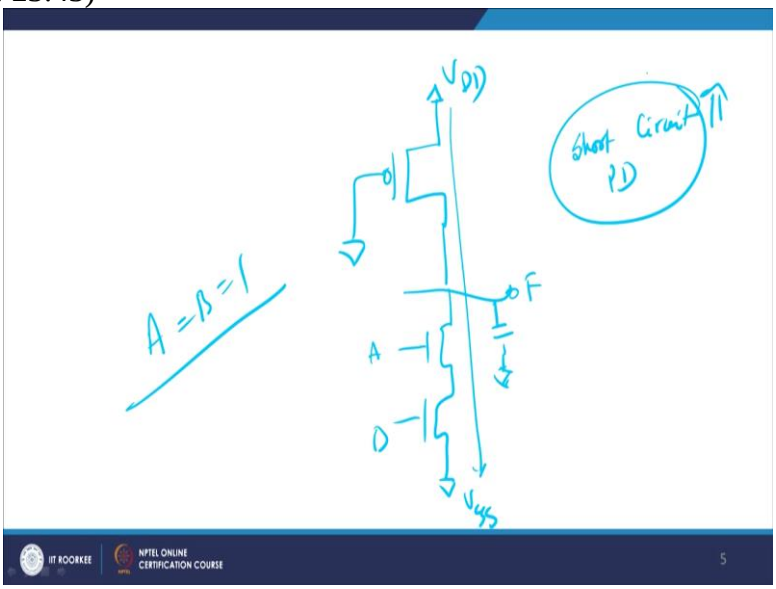
For a ratioed logic, it is just N plus 1. So if you are using a, for example a 4 input will be an AND gate or whatever gate, then for a CMOS logic input, you require 8 transistors. Whereas in this case, you just require 5. So the transistor count reduces, the area reduces. The cost you pay for it is higher power dissipation. So they have got extra power dissipation. I will explain to you why is it that. So what we do is, in a ratioed logic, what we do is we always keep the pull up network as a single PMOS device which is always on, let us suppose.

So what I do? I have a load here, the load which is given here in this figure. This is basically a PMOS and this PMOS is always on means its gate is always grounded. So that is what I am

saying that PUN is replaced by a single unconditional load right by a single PMOS and having exactly always on state which means that obviously, your output voltage will always be latched to $V_{DD}$ because by that definition, till even a complementary CMOS. So therefore, my $V_{OH}$ which is output high will be equals to $V_{DD}$, no problem.

But the problem is that your low voltage will not be equals to 0. Low voltage will not be equals to 0. It will depend upon the type of inputs you have given here right. Let us suppose you have given two transistors in series right and you have given 0 1, then 1 will be cut off and therefore you will never get to 0 right. Is it okay? But there is a problem here and therefore, you will have I will explain to you why do you will have static noise margin or you will have static power dissipation, I will explain to you.

(Refer Slide Time: 23:43)




See, for example, if you are, this is your PMOS right and a PMOS has been switched on. So it is grounded. So it is always on. And now I have got let us suppose a AND gate. So I have got A, B right and this is my F and this is $C_L$ here. This is $V_{DD}$ and this is your $V_{SS}$. Now if A and B are both equals to 1, then this will be switched on. But when this is switched on, you automatically have what? You have a direct path between $V_{DD}$ and ground. So your short-circuit power dissipation, short-circuit is quite high, PD.

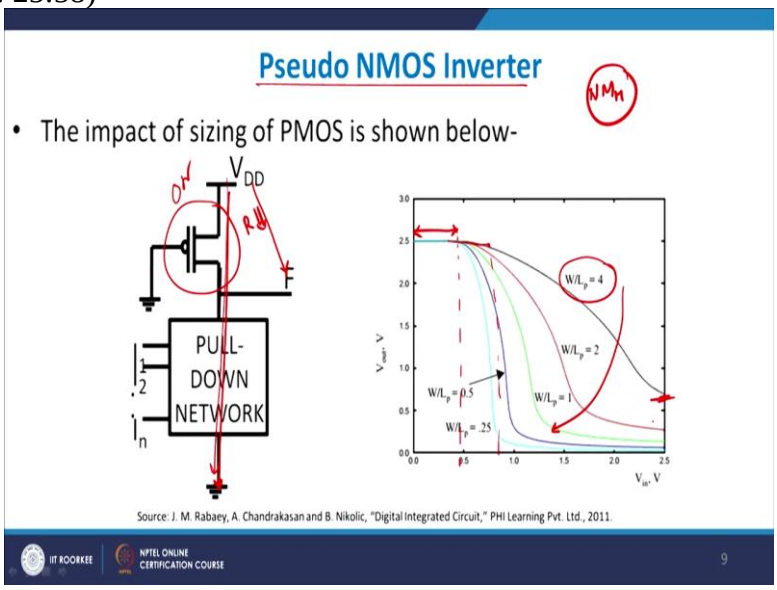
Your short-circuit powered situation is quite high. Why? Because now you have a direct path between $V_{DD}$ and $V_{SS}$. So though the logic will give you a good value but otherwise, it will not

give you a good value in real sense. And that is the problem with static noise margins, with sorry, with the static power dissipation. 2nd thing is, we will not discuss in detail but it will reduce the noise margin as well. So pseudo-NMOS logic or a pseudo-logic it is also referred to as a pseudo-NMOS logic, also known as the ratioed logic will have a lower noise margin, right.

And the reason is that your pull up network and pull down networks are not equally strong. For example, your pull up network might be very weak because you are using a low value of W/L because you want the RL to be very high, load to be very high, you want W/L to be small but once you make W/L small, you are making a pull up weak. When pull up is weak, your high noise margin reduces, right.

And that is the major area of concern for ratioed logic, right and that is the problem area this thing. Now, therefore since the output voltage depends on the size of the transistors, so it is called a ratioed logic. Why is known as the ratioed logic? Because the output value of voltage will depend upon the size of the transistor. Higher the size better the transistor will be able to make the output voltage go to ground and lower the size, difficult it will be go to the ground. So therefore, this is known as the ratioed logic as I discussed with you. Also, therefore also this is also referred to as a pseudo-NMOS inverter as I discussed with you.

(Refer Slide Time: 25:58)




So what I do? I PMOS, I make it grounded. So this is always on state right and then I have a pull up network with me. Now as you can see here, my pull up network is not strong because I have

only one PMOS here and therefore, your noise margin which is high noise margin, is very small. You see, it is just this much, almost, approximately this much for W by L $P_5$, is this much which means that, if the noise flow or the noise is typically low, even then this will not be rejected by the inverter at high values.

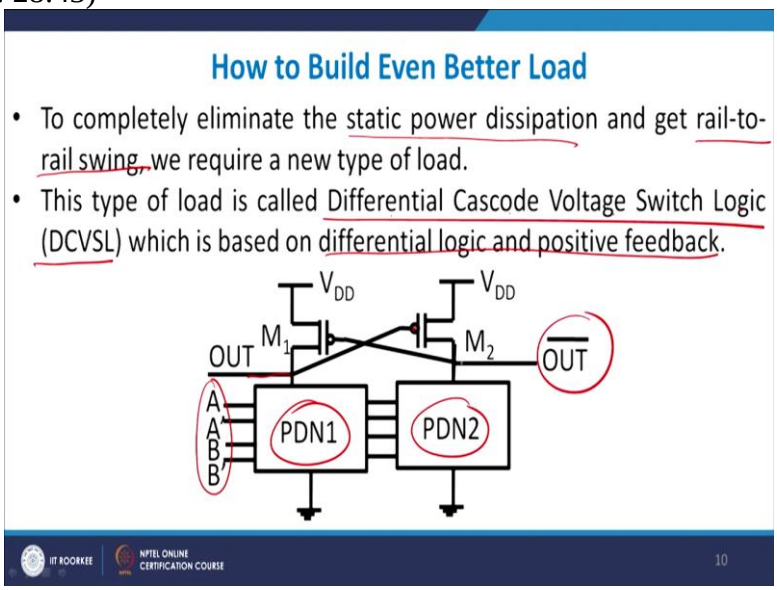
Now as you can see here, so this is W by L of PMOS. So as the W by L of PMOS is increased, increased means R value goes on reducing and you can easily have $V_{DD}$ appear as F and therefore you see your noise margin goes on increasing. So the noise margin for W by L $P_4$ is approximately this much, whereas it goes on reducing as you go on reducing the value of W by L of P. So once the aspect ratio of your pull up transistor goes reducing, your noise margin reduces.

Not only noise margin reduces, you also have a problem that, for example if you are saying a W by L P equals to 4, your output when you are input is high should so when you are input is high, output should go to ideally go to 0 because let us suppose, it is a two input AND gate, then 11 will give you 0. So but it does not go to 0. Why it does not go to 0? The reason is, you always have a direct path between $V_{DD}$ and this $V_{SS}$. And higher the value of W by L of PMOS, lower is resistance and that larger current will be flowing through $V_{DD}$ and $V_{SS}$.

As a result, it is never latched to 0 in this case. So at the cost of higher noise margin, my output does not go to the 0 value where it should go. And therefore, when you cascade a ratioed NMOS inverter to another one where you expect that the, so if you are cascading this inverter with another inverter and that inverter is expecting a 0 from this inverter, but in reality, it is not giving me 0, it is giving me 0.5, 0.6, 0.7 voltage, there will be problem as far as its operation is concerned. And that is the major concern of a pseudo-NMOS logic.

There are 2 concerns. One concern is that your output does not even go to 0. It goes to 1 very well, it does not go to 0 properly because of the always on, this PMOS is always on right. And the 2$^{nd}$ property is its static power dissipations are very high. So its overall power dissipations are very, very high.

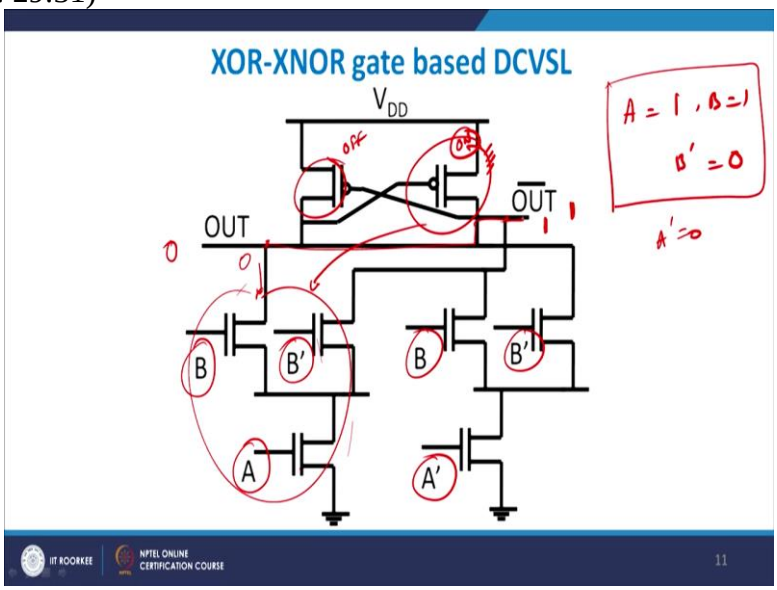
(Refer Slide Time: 28:43)




So let me see how can you build a better load or maybe a good one. And this is known as what is known as a Differential Cascoded Voltage Switch Logic, DCVSL. Differential Cascoded Voltage Switch Logic, also known as DCVSL, based on differential logic and positive feedback. So what you do is very simple. You have a PDN and you took a PDN2 which is just complementary of PDN1 and you apply both A and its, so if you have two signals, you also apply its complementary signal and what it does?

It totally removes your static power dissipation and you automatically get a rail to rail swing and I will show you how it works out. So you see this PMOS is connected, the output here is connected to the gate of PMOS $M_2$ and out bar is connected to $M_1$ gate here.

(Refer Slide Time: 29:31)



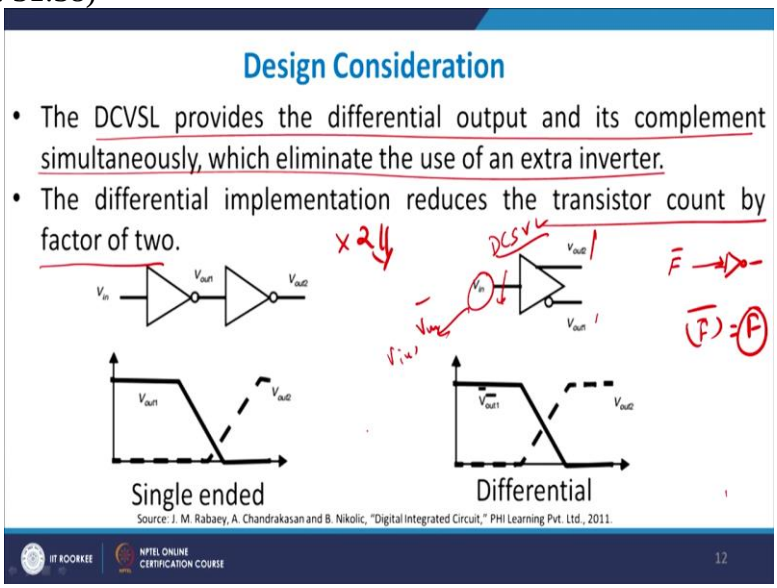

Let us look at the functionality of a DCVSL XOR-XNOR gate. Now you see, in this case if A equals to 1 right, B equals to 1 and therefore B bar will be equals to 0. In such a scenario, what I get is that I have, so what I did was, I told you PDN should be complementary in the sense that if you apply A signal here, you should apply A bar signal to PDN2. If you apply B signal here, you should apply B bar here. If it is B bar, you apply B here, right. So let us suppose A equals to 1, B equals to 1, B bar equals to 0.

When A equals to 1 and B equals to 1, this out goes to 0, so out of goes to 0. In this case, A equals to 1 primarily meaning A bar equals to 0. 0 means this is cut-off. When this is cut-off, means this goes to high. So this goes to 1. This goes to 1 primarily means that this PMOS is goes to off and this is on. Once this is on, so this transistor works with this module to give you an output here. So out becomes equal to 0 and out bar equals to 1. So in a DCVSL logic, I insert differential signals, I also get differential signals back which is 0 and 1 here, right.

So I input differential signals, A and A bar, A and B bar and I also get differential signals out from out and out bar. This is what we get in DCVSL logic but the idea there is that there is no direct path between $V_{DD}$ and ground at any point of time which was there in a pseudo-NMOS logic. In a pseudo-NMOS logic, you did have a direct path from $V_{DD}$ to ground. In this case, you do not have any direct path. So you do have a simple, so you see, when this is on and when this is on, you do not have a direct path.

So you see this is goes like this and this is the output which you see. So this is on right and this is switched on here. This is switched on, this is switched on and therefore this output goes to 0 and therefore it almost out bar equals to 1. And therefore, this gives you a very, very low static power dissipation right.

(Refer Slide Time: 31:38)

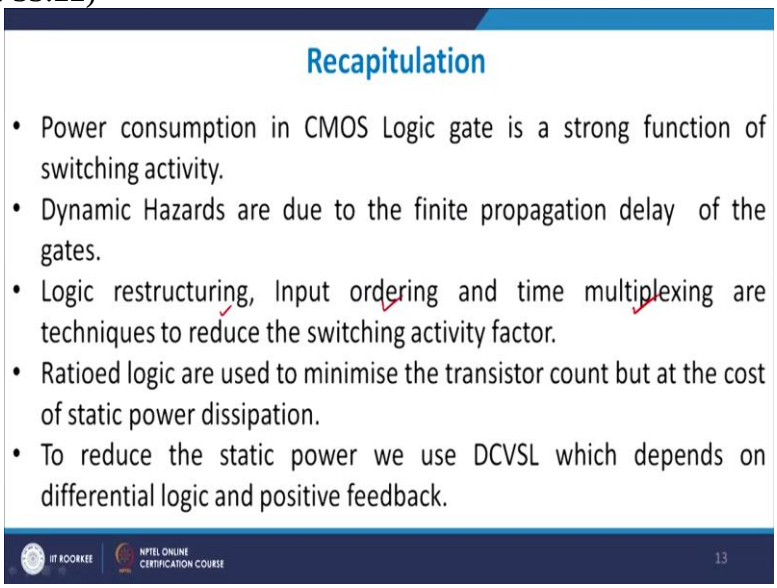



So therefore, DCVSL provides a differential output and its complement and therefore it eliminates the use of external resistor. So you remember, when we were using a ratioed logic or when we were using a standard CMOS complementary logic, my output will always be complementary. Remember, by virtue of the fact that it is an inverter. So you have to put an extra static inverter to complement it back to its original form. So I generally get F bar and therefore when you pass it through a static inverter, I get F bar of bar and therefore I get F.

So I am using one more static inverter to achieve the signal. In this case, we do not do that. In this case, we already get both the signals and its complementary together. The differential implementation reduces transistor count by a factor of two. So you get a 2 times reduction in the transistor count in this case. Now, so I get $V_{out}$. So this is a $V_{out1}$, $V_{out2}$, $V_{out1}$, $V_{out2}$. This is single ended and this is a differential operation as you can see $V_{in}$ I do and I get $V_{out2}$. So this is basically your DCVSL but the problem is that, $V_{in}$ should also be Vin and you should also give Vin bar, the complementary forms should also be given right and that gives you quite a good idea about this whole thing.

So let me recapitulate up this module what we have done till now. We have learned that the power consumption is a very, very strong influence on the switching activity. Switching activity plays an important role in determining the power consumption.

(Refer Slide Time: 33:22)




The power consumption depends upon $C_L$ the load, $V_{DD}$ the applied voltage, and f the frequency of operation. It also depends upon the topology which you have with you and also depends upon dynamic glitches. So the methodology we adopted for reducing dynamic glitches are logic restructuring, right, logic restructuring, input ordering and time multiplexing some of the techniques because this will reduce your dynamic hazards.

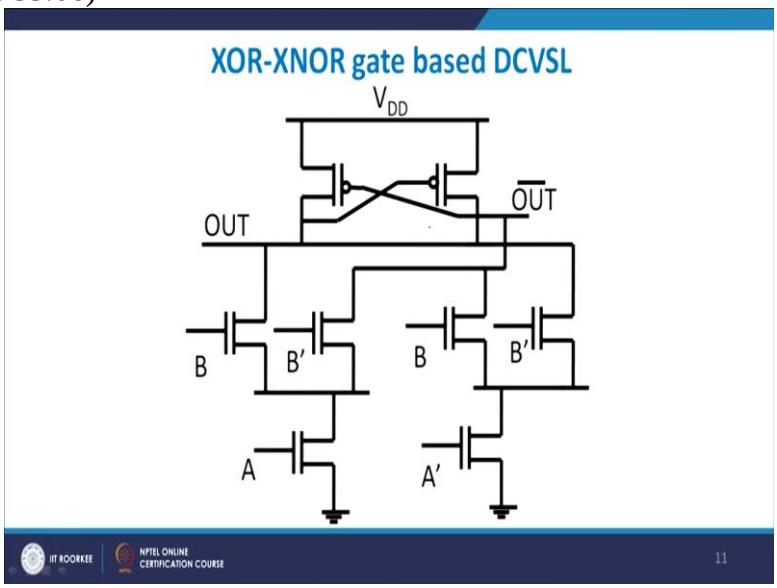
And dynamic hazard appear because of finite propagation delay of the Gates, right and that gives you an important one. To remove it people went for ratioed logic, in which ratioed logic was what? That you replace the pull up network by a single PMOS device which is normally on and you make it always on because you ground the gate of the PMOS and then you try to get the output. But the problem there was that you have two problems.

One was that you did not allow the output to go to fully to ground if your load was having a large W by L ratio because now the load will try to pull up the voltage towards itself and therefore you will not allow it to go to ground and that is the major problem area and therefore you and therefore, the 2nd problem which came into picture was that the noise margins were reduced

drastically, specially the low noise margins were reduced drastically. High noise margins will depend upon the W by L ratios of the PMOS which you take.

So higher the W by L ratio, larger will be the static noise margin in this case. Now to reduce static power and to get a differential output, we did DCVSL which is Differential Cascoded Voltage Logic and in which case, you use a differential input, you do a positive feedback and then you do this thing. Just to give you a brief insight what positive feedback helps you to do what I will tell you.

(Refer Slide Time: 35:06)




It tries to reduce the delay drastically. See when you have a positive feedback and suppose this is 0 and this is basically your 1 here, this 1, so this 0 will try to switch on this device very fast, right and therefore the overall delay of a DCVSL is reduced drastically. So since this is a positive feedback, you do not have to wait till a particular time for the output to appear and therefore suddenly when it goes to 0, because by application of A equal to B equals to 1 and B bar equals to 0, I automatically get a much lower profile here, right?

So we have understood this whole logic here. We have therefore taken care of ratioed logic, DCVSL, understood what is power dissipation and we also know what is propagation delay. So by this, end of this module, you as a deliverable, you should be able to deliver the logic, you

should be able to design a logic in a much better manner using CMOS technology, using pseudo NMOS technology and using a DCVSL.

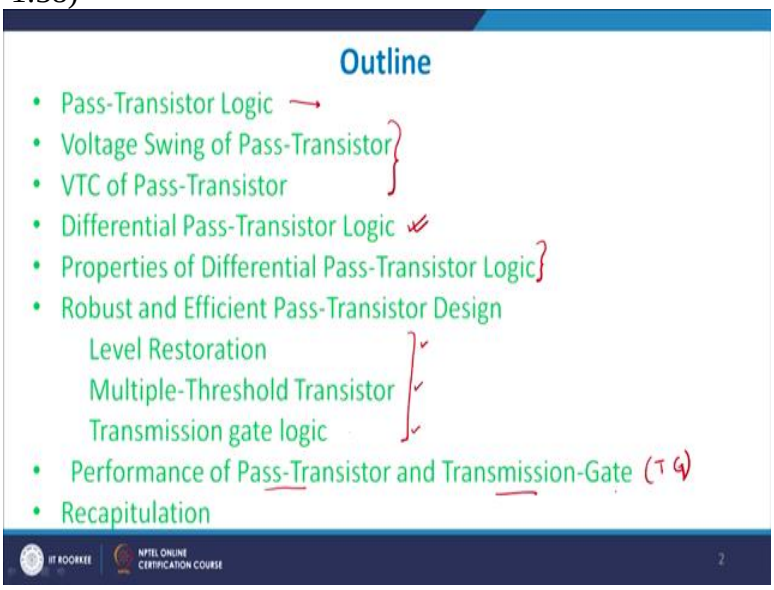
And please try to do it as a sort of a home assignment once you go back. Good book to look into is basically, there is a good book by Digital integrated circuits by a designer's perspective by Chandrakasan, Rabaey and Nikolic. I will recommend that you please go through that book for at least the digital part. Analog part you will not get there but the digital part of this whole course structure is very well defined in that case. Thank you. Thank you for your kind, patient hearing.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−55**
**Combinational Logic Design -III**

Hello everybody and welcome to the next edition of NPTEL online certification course on microelectronics devices to circuits. We start today's module from the combinational Logic design III. So we have actually covered two modules of combinational logic design in which we have seen what is the meaning of combinational logic, how can you design a combinational logic from a Boolean expression, we have also seen how does a combinational logic, what are the various parameters for example power, delay, how they are related to the aspect ratio of the transistors and so on and so forth.

So we have studied two types of styles. One is the complementary style, one is the ratioed logic we have studied and we have seen that, complementary style is the most robust style but the area count is typically high. Whereas ratioed logic gives you more amount of static power dissipation and noise margins are limited. Today we will take up what is known as a pass transistor logic in applied to digital logic design.

(Refer Slide Time: 1:38)

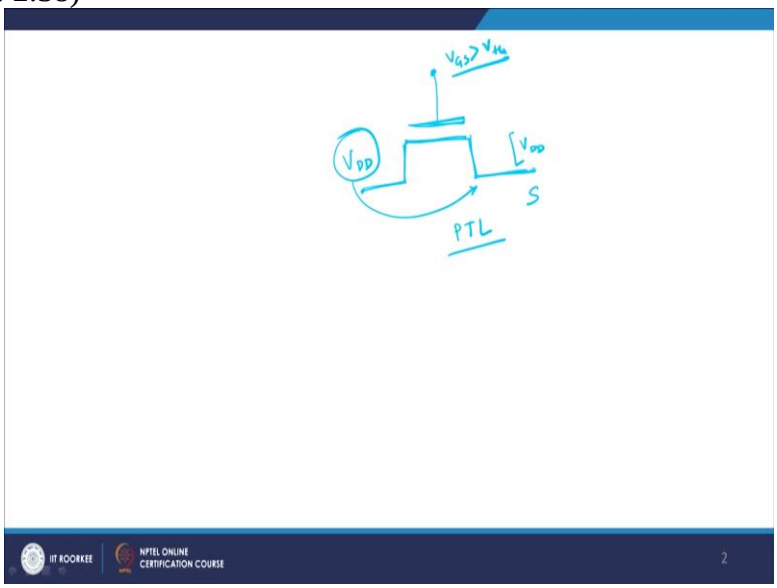



So the outline of this module is that we will first look into pass transistor logic and then look into the VTC and therefore the voltage swing of the pass transistor logic. So we will look into the fact

that what is a pass transistor and therefore what is the voltage transfer characteristics. Then maybe I have a look at differential pass transistor logic. We will look into properties of differential pass transistor logic and then we will understand these three quantities here which is level restoration, multiple threshold transistor design and transmission gate logic, right?

And then we will compare pass transistor logic with TG logic which is transmission gate logic. This is also referred to as TG, right. Transmission gate. And then we will recapitulate. So this is a general outline of the module structure for this particular lecture. And we will go step-by-step and see how it works.

(Refer Slide Time: 2:38)

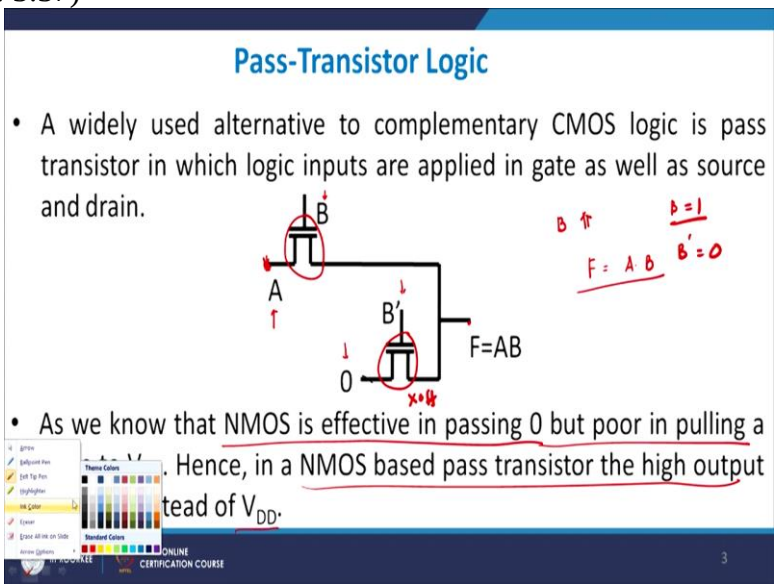



If you remember now, till now we will first now look into the transistor logic. What is a pass transistor? Let me give you a very simple idea what is a pass transistor. A pass transistor is something like this that if you have a gate, right you have a gate here and you apply a $V_{DD}$ here, it is a drain side and this is your source side. Now, what you do is that you do not apply and signal on the gate side right? You apply the signal actually on the drain side, right.

And let us suppose you apply a gate voltage which is greater than threshold voltage, then since this is on right, this will be on, this $V_{DD}$ will be transferred to this particular point and this will be $V_{DD}$. Obviously, of course there will be a threshold voltage drop. I will not discuss that point here at this stage but the name, it is pass transistor logic, also referred to as PTL, pass transistor

logic, it allows you for this $V_{DD}$ to move forward and appear in the source side, right. So this is the basic pass transistor logic, basic idea.

(Refer Slide Time: 3:37)



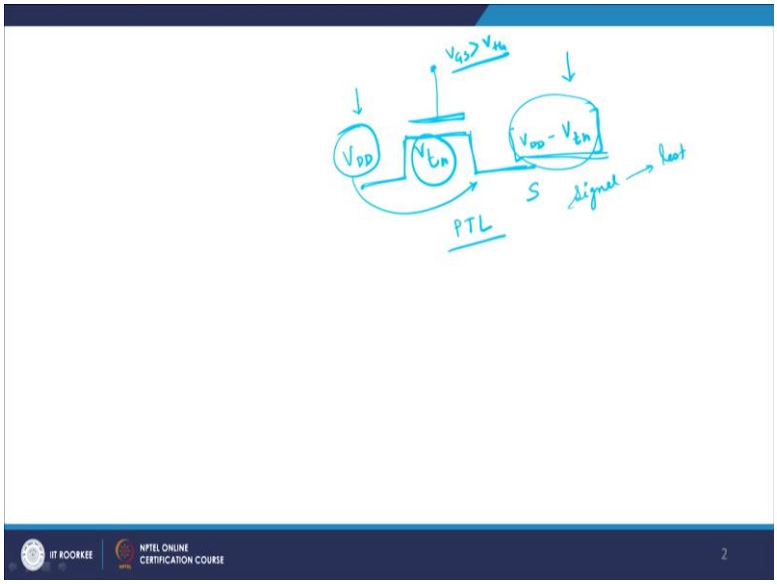

Now with this knowledge, let me give you an idea about basic AND gate using pass transistor logic. So you see here, I have got one transistor here, another transistor is here, and we apply the signal B here and we apply a complementary signal B here, we apply a 0 here and we apply the A here. So whenever my B is high right, whenever my B is high, whatever be the value of A appears outside and therefore safely I can write down F equals to A dot B. Why? Because whatever the value of A will appear at this point.

So if A is equals to 0, F will be equals to 0. If A equals to 1, F will be equals to 1 provided B equals to 1. Now whenever B is higher, B equals to 1, you ensure that your B complement is basically zero which meaning then this will be cut off. And therefore, this your output will be just latched or bar to this input A here. We have already dealt in your previous turn that if you remember the question which we asked was, why NMOS is a very good pull down network candidate whereas PMOS is a very good pull up candidate?

And we told you that or we discussed that NMOS is a very good, it is a very good, it allows you to do a very good passage of zero whereas PMOS allows you to do a very good passage of 1. So NMOS is a very good passing of but a poor in $V_{DD}$. Hence in NMOS-based pass transistor, the

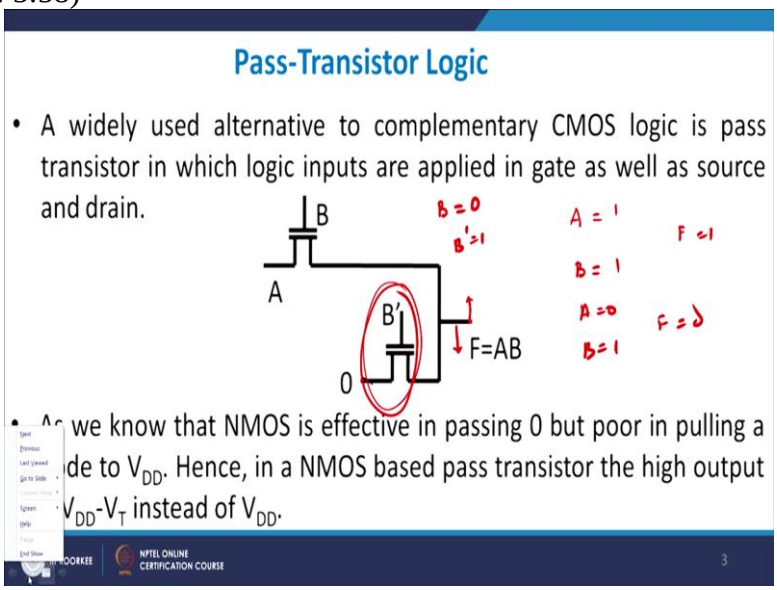
high output is $V_{DD}$ minus $V_T$ instead of $V_{DD}$. That is what I was saying that if you have got a high input transistor in a pass transistor logic, the output will appear as $V_{DD}$ minus $V_T$.

(Refer time slide 5:24)




So the discussion which I was having here, you will have $V_{DD}$ minus $V_{TN}$ here, right where $V_{TN}$ is basically the voltage drop here, right. So whatever the voltage here, you get a drop in this case, right. And therefore, and please understand this drop cannot be restored at any point of time, right. And therefore there will be a loss of the signal. So the signal will be lost, right. So a pass transistor logic has this problem that the signals will be lost.

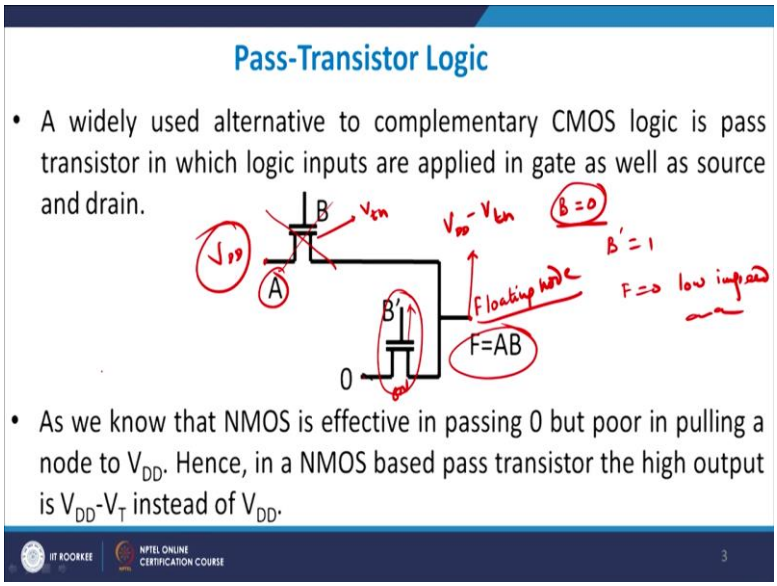
(Refer Slide Time: 5:58)




Now if you look at this, again this structure here, we have discussed this point that now you see, in this whole structure, even if you do not give this transistor this transistor, you still will have an AND application but the problem is that there will be a floating node available to you whenever my A is equals to 0 or 1. So now suppose my B equals to 1 and my A equals to 1, then output of course will be 1 and this will be at high-voltage, high. Similarly, if A equals to 0 and B equals to 1, output F will be equals to 0 and this will be at low voltage.

But then, this will not be connected to any input and therefore there will be a problem of floating node, right. And that is the reason you always have this transistor with you. I will give you an example. For example let us suppose B was 0, right. So B bar will be equals to 1. When B is zero, suppose this transistor would not have been there when B equals to 0 is there. So when B equals to 0, this will be acting as a floating node. You are getting my point?

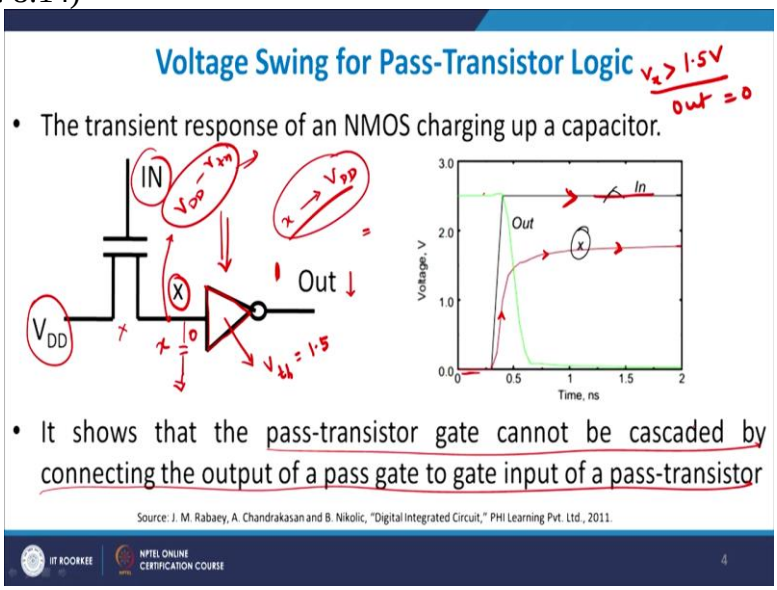
(Refer time slide 7:05)




See, suppose we assume that just to give you a brief insight, say B equals to 0.B equals to 0 implies that this transistor is cut off. And suppose this transistor would not have been there in the design, then this node F would have been a floating node, right. Floating node and all sorts of voltage spikes, and everything would have been visible here. So what people thought or what researchers thought was that if you give B goes to 0, then B bar equals to 1 right, which implies that this will be on and when this is on, this is zero will, so your output node will be plugged to 0 which is a low impedance node.

So your output F will be equals to 0 which is a low impedance node. So at no point of time, this F will be a floating node and therefore you will have no problem related to high distortions. So first thing is this one, the second thing is as we discussed with you, the voltage here will be equals to $V_{DD}$ minus $V_{TN}$ where $V_{TN}$ is the threshold voltage of this one and this voltage is equals to $V_{DD}$, fine. So this is the pass transistor, basic pass transistor logic application.

(Refer Slide Time: 8:14)



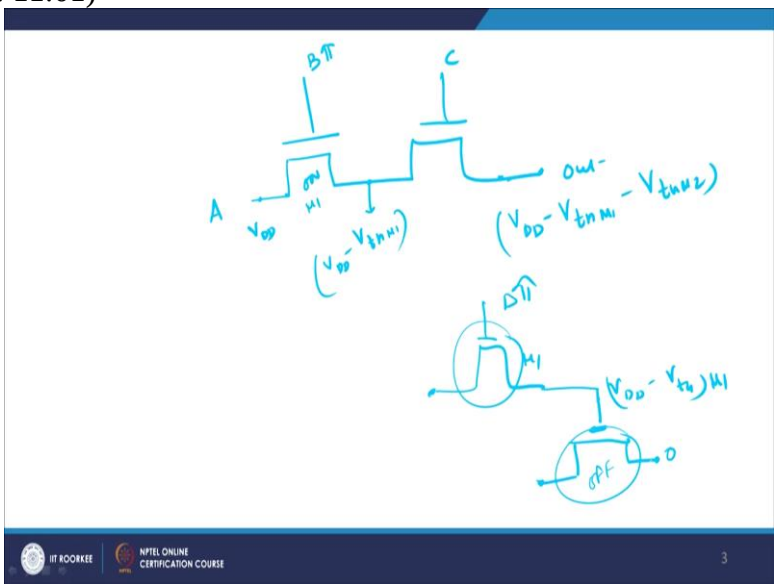

Now let us look at the property of voltage swing in a pass transistor logic, right. So we need to also look into the fact that you do have a voltage available with you and that is swinging right. It is swinging from low to high. So I have an input, so my $V_{DD}$ is given on the drain side, let me define this point source to be point X and then of course, I have a static inverter here and I am giving an input here right and there will be a capacitance, of course there will be a capacitance here, right.

And this capacitance will be charged by virtue of the current here. Now this is the input, this black colour is my input. This black colour which you see here, is my input. Now when the input was low, right, when the input was low right, let us suppose this was low, it means this is cut off. Once this is cut off, the output goes to high. So output is high now right and now the input starts to rise. As the input starts to rise, look at this red one. Input starts to rise which means that X now starts to reach towards $V_{DD}$.

And that is the reason you see the red coloured graph here, it starts to $V_{DD}$ but what happens is that, beyond a particular time limit, this X latches onto only $V_{DD}$ minus $V_{TN}$. So your response is $V_{DD}$ minus $V_{TN}$. So if your $V_{TN}$ is approximately say 0.7, 0.8, you automatically have a lower value of X at this particular point. So when X is high, out will be low at this point. But it can be only low, understand, it can be only low provided the value of voltage at point X exceeds the switching threshold of the CMOS, static CMOS.

Say the switching threshold for this one, $V_{TH}$, switching threshold, is say 1.5 volts. Then the value of voltage at X should be larger than, so V of X should be larger than 1.5 volts for out to be equals to 0. But what has happened is that X rather than going towards $V_{DD}$ and out therefore going to 0, your X only went up to $V_{DD}$ minus $V_{TN}$. If your $V_{TN}$ is relatively large for certain reasons or other, this output voltage will not cross the switching threshold and as result, it will be lesser than that and this inverter, static inverter will read it as zero only. And therefore output will be still latched to 1, right. And this is the problem area which you are facing. And therefore, the pass transistor gate cannot be cascaded by connecting the output of the pass gate to gate input of the pass transistor. I will explain to you what do you mean by that.

(Refer Slide Time: 11:01)



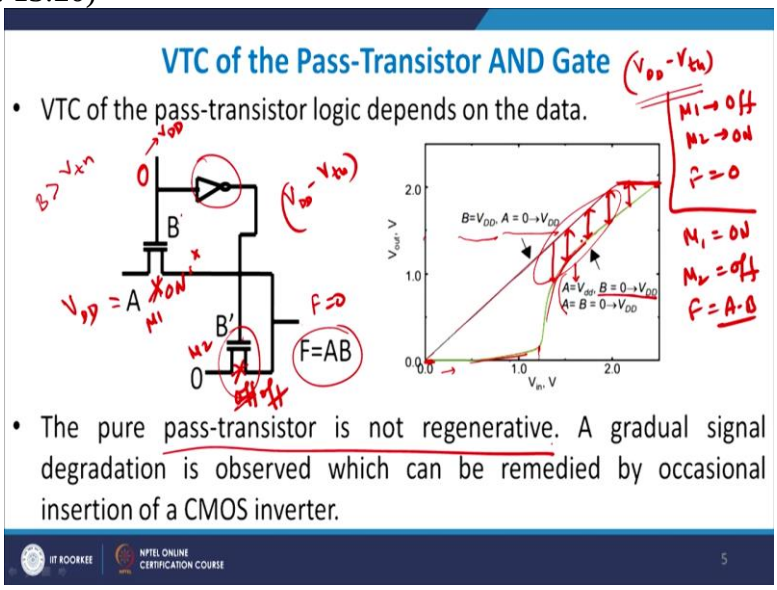

Let us suppose now let us suppose, now I have got two pass transistors connected in this manner right. Is it okay? So this is A, this is say B. This is B. This is C and this is your out, fine. So what will happen is that if this is $V_{DD}$ here, you are given this one high, so this is on. This $V_{DD}$ appears here as $V_{DD}$ minus $V_{TN}$ of say this is $M_1$, of $M_1$, right and at this point, it will appear as $V_{DD}$ minus $V_{TN}$ of $M_1$ minus $V_{TN}$ of $M_2$ which means that in a pass transistor logic, if you cascade in this manner, then the signal as you move from primary input towards the output, the signal level starts to fall down by one threshold voltage as you cross one transistor.

And in no way you are able to restore it back to its original state until and unless you put a static inverter, right? And that too, I should ensure that the voltage level does not go above or below

the switching threshold, right. And therefore, this type of cascading does not work in a pass transistor logic for large logic designs. This is one problem area which people face. Another problem area is, let me show you what will happen if this is connected as the gate, something like this.

This we have already done, right? And therefore when B is going high, this $V_{DD}$ will appear as $V_{DD}$ minus $V_{TN}$. $V_{TN}$ of what? $V_{TN}$ of this transistor say $M_1$, of $M_1$. And that is the voltage here which is given here. Now if this voltage does not exceed the threshold voltage of this second transistor, this will be switched off and the output will be held to 0 voltage or whatever the original voltage was right. So what I wanted to point you out was that it is very difficult to cascade two pass transistor logics in a similar manner and you need to restore the signal level by simply using a static inverter in between the two transistor levels.

(Refer Slide Time: 13:20)




Now let me come to the voltage transfer characteristics of the pass transistor AND gate, right. So AND gate I have already discussed with you. I have got A, B Prime, B and what I am now I am adding is, I am adding this gate output or the obviously, this gate is inverted by this static inverter and I am feeding it at B prime and F equals to A into B. Now you see, suppose B equals to $V_{DD}$ and A goes from 0 to $V_{DD}$, so therefore when input equals to 0 and as my input starts to increase, this goes on increasing, output goes on increasing, this output goes on increasing right.

And this is a linear, obviously rise will be to you but beyond a particular $V_{in}$ right, how much it will be? $V_{DD}$ minus $V_{TN}$. You will see that it will be almost saturated and therefore this will be a saturated case. Why is it like that? Because when B equals to $V_{DD}$, this is switched on. This is switched off of course, this is off right. So A will appear at point X by degrading it by one threshold of the device right, one threshold of the device. But whenever $V_{DD}$ is much smaller as compared to $V_{TN}$, my $V_{out}$ will follow $V_{in}$ and therefore this will be a straight line path available to you till a point where you reach $V_{DD}$ minus $V_{TN}$.

At such a point, this device will enter into deep saturation and the voltage will be fixed at $V_{DD}$ minus $V_{TN}$ right. And therefore the voltage is fixed at $V_{DD}$ minus $V_{TN}$. Similarly, let us look at the fact that, suppose A was equals to $V_{DD}$ and B was equals to going from 0 to $V_{DD}$, B is this one, right? So it was initially zero and now it is going from 0 to $V_{DD}$ and A has been fixed to $V_{DD}$. So when initially this was zero right, this will be cut off.

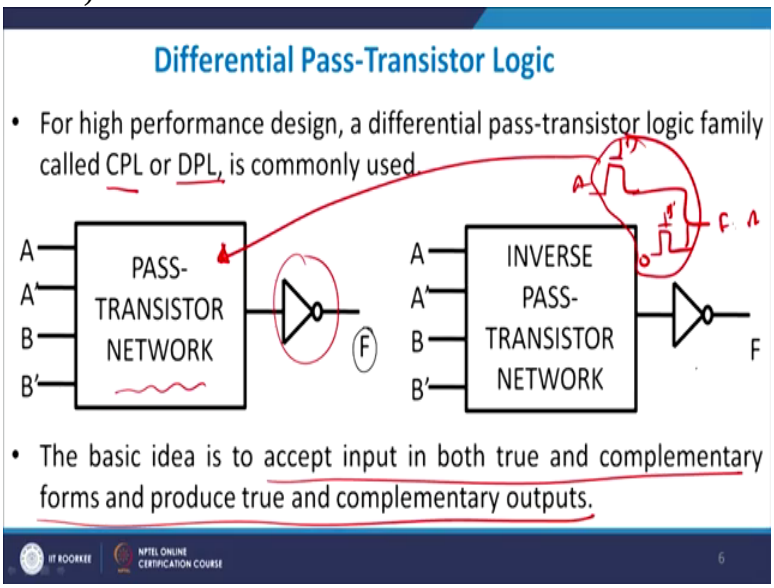
So this transistor will be cut off. As this transistor is cut off, this will be fully on and the output F will be equals to 0. And that is what you are getting here. So whenever your $V_{in}$ is very, very small, output is also small. It goes like this. Just at the point where your threshold voltage of the device is there, when B goes above the threshold voltage of the device, this switches on and then this switches off. Sorry, this switches on and this switches off, yes, off.

So it was initially on and this was, so let me assume this to be as $M_1$ and this to be as $M_2$. So initially $M_1$ was off and $M_2$ was on and F was equals to 0. Now what has happened? $M_1$ is going to on state, $M_2$ is off state and F is equals to A dot B which you see. And therefore you start following this curve in reality. So this is the curve you follow. So as your $V_{in}$ increases, so since $V_{DD}$ is already fixed, I get $V_{DD}$ minus $V_{TH}$ curve till the point I get up to this much right.

B goes from 0 to $V_{DD}$. So when the voltage starts, so your…this is fixed, B is varying from 0 to $V_{DD}$. When this B is varying from 0 to $V_{DD}$, this goes like this right and the reason is that this switches on, this switches off and if A dot B comes into picture here right and that is quite critical or understanding it. So you can see your $V_{out}$ in this case is already down by $V_{TN}$ as compared to the previous state. It is already down by quite a large quantity.

Below $V_{TN}$, it is already so difference is there, above $V_{TN}$ also there is no difference but around this part, you will have $V_{TN}$ difference between the two, threshold voltage difference is there. So this is the, so as I discussed with you therefore the pure pass transistor logic is not regenerative, means the voltage once lost is gone, you cannot regenerate it and as gradual signal degradation is observed, which can be remedied by the occasional insertion of a static CMOS inverter because inverter will either latch it to $V_{DD}$ and zero and therefore you have to insert $V_{DD}$ resulting in a larger power dissipation.

(Refer Slide Time: 17:40)



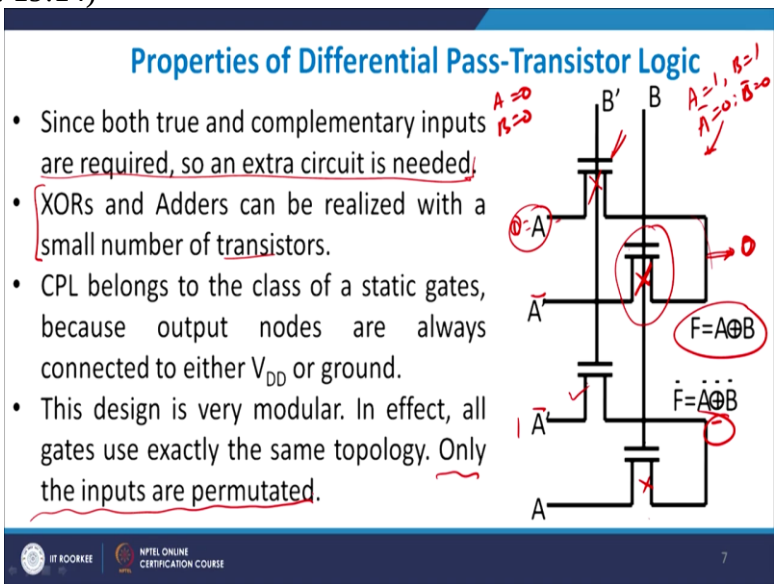

Now, typically when you do a high-performance design, we use a differential pass transistor logic in which we assume that my input and its complements are available to me. So A, A bar, B, B bar, C, C bar, are all available to me and I do a pass transistor network here and then I put a static inverter to get a full swing in the output side.

Now the basic ideas as I discussed with you accept the input in both true and complementary forms and produce a true and complementary outputs. So the output will also be a true and complementary one. Well, that is simple because you do have a static inverter here, right? This is also referred to as a differential pass logic or even a complementary pass logic here right.

Differential, most common is the differential pass transistor logic or DPL or DPTL right and it gives you quite an interesting result as far as this is concerned. So what is this pass transistor network? The same thing which we just now studied that there is you have got A, B right and

then it comes like this and then the same thing which you have done. So this is B prime, this is B, this is A and this is zero. So F equals to A into B. So this whole thing which you see is inserted in this black box, right. And I will term this as a differential pass transistor logic.

(Refer Slide Time: 19:14)




With this knowledge or with this idea, let me come to the next one that let me take up the properties of differential pass transistor logic and let us see how it works out. So since both true and complementary inputs are required, so extra circuitry is needed just to complement the signal. So I require an extra static inverter to complement the signals. Of course, XORs and Adders can be realized using small transistors, one example is given here where XOR and XNOR are designed.

So you can see here, when A, I have got A, A bar, A bar, A right? So and this is, so if your A equals to 1 and B equals to 1, then what will happen is that when A equals to 1, this is 1. So your A bar will be equals to 0 and B bar will be equals to 0. So this will be zero means this will be cut off right and B equals to 1 meaning this will be on and therefore A bar will appear in the output side here, A bar. And therefore that is nothing but zero, right.

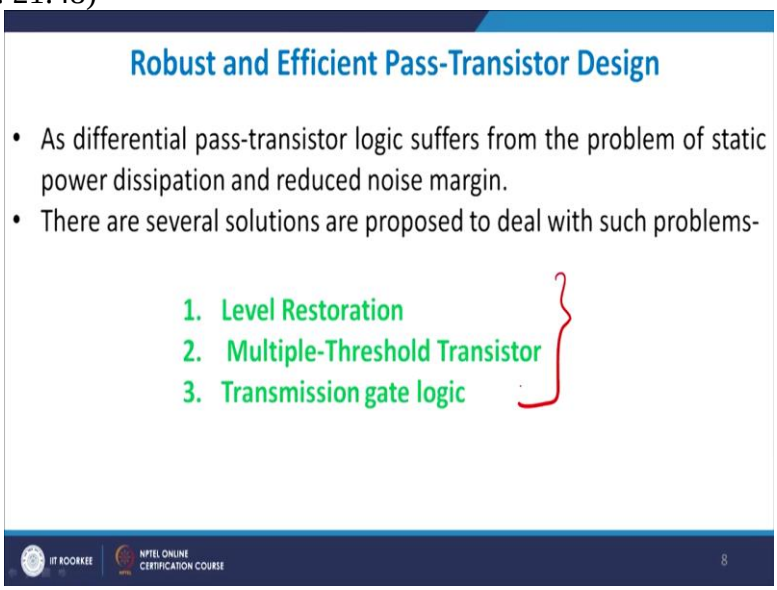
Which means that when A and B are both equals to 1, I get a zero. When both are 0 0, sorry when equals to 0 and B equals to 0, so A equals to 0 and B equals to 0 implies that this is cut off now, this one is cut off. So with B equals to 0. With A equals to 0, what will happen is that, with

A equals to 0 sorry. So with A equals to 0, what will happen is that this will be 0, of course 0 and B prime will be 1. So this will be switched on. So I will have 0 coming into picture here and when B prime equals to 0, this will be cut off.

So B prime will be 1, this will be switched on. And B will be equals to 0, this will be cut off and A bar will appear here, so 1 will be appearing here, right? So if I want to find out XNOR, XOR I will get this. If I want to find XNOR, I will get this into consideration. So I can have an XOR bar here, right. And so we just need to permutate the inputs in the proper manner so that I get the proper output here, right.

However, as I discussed with you, there will be always a problem of signal degradation between the pass transistor logic input and output and therefore cascading this across a large number of logic is a futile exercise because everywhere you will get one $V_{TN}$ drop and if that exceeds the switching threshold of the static inverter, I will actually read 0 as 1 or 1 as 0 and that is quite a important problem area which you will see.

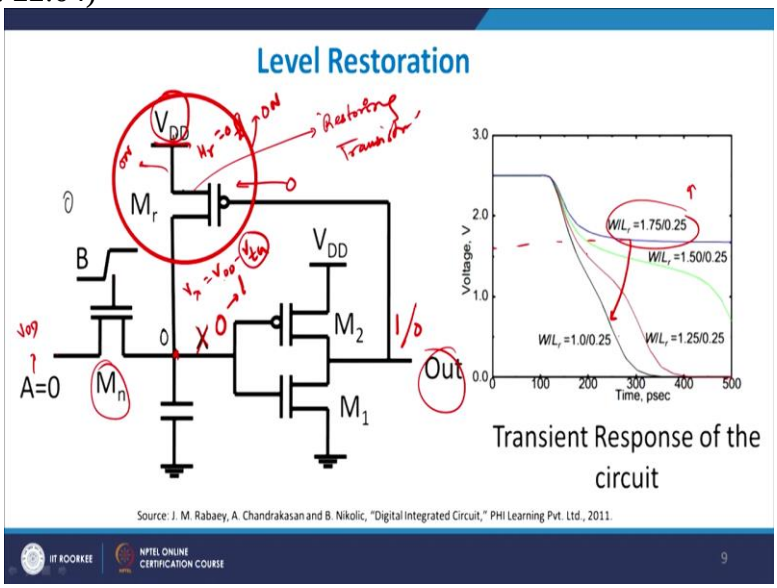
(Refer Slide Time: 21:48)




Now as I discussed with you, differential pass transistor logic suffers from the problem of static power dissipation. Why power dissipation? Because now you are directly allowing the output to see the input right and therefore there will be a large current and therefore this will result in a larger power dissipation. So there are certain solutions which is level restoration, multiple

threshold and transmission gate. We will take each one of them and explain to you how it works out.

(Refer Slide Time: 22:04)




Say for example, I have a level restorer. Means I want to restore the level at the inputs. So this is my static inverter here. Just like an AND gate, I have A and I have B here and I have this is $M_n$ and X is the point, this particular point and this is the output here. Now you see, when this is equals to 0 right, when this is 0, this will be 1 and therefore $M_r$ will be equals to off. When this is off, this will be again 0 because you are connected to the ground through this capacitance.

Now when X equals to 1, sorry X equals to yes, 1, this will be 0. This will be 0 means this will be switched on. When this will be switched on, this $V_{DD}$ will appear at this particular point and try to pull the value of voltage here above $V_{TN}$ because if you see, this is A equals to 1 means A equals to $V_{DD}$. Then my $V_X$ will be equals to $V_{DD}$ minus $V_{TN}$. $V_{TN}$ is the threshold voltage of $M_n$. Now this will switch on my transistor and as a result, this will be zero. When this is zero, this will switch on my $M_R$ and this $V_{DD}$ will be connected to ground and the loss which you get through $V_{TN}$ will be made up by this $V_{DD}$.

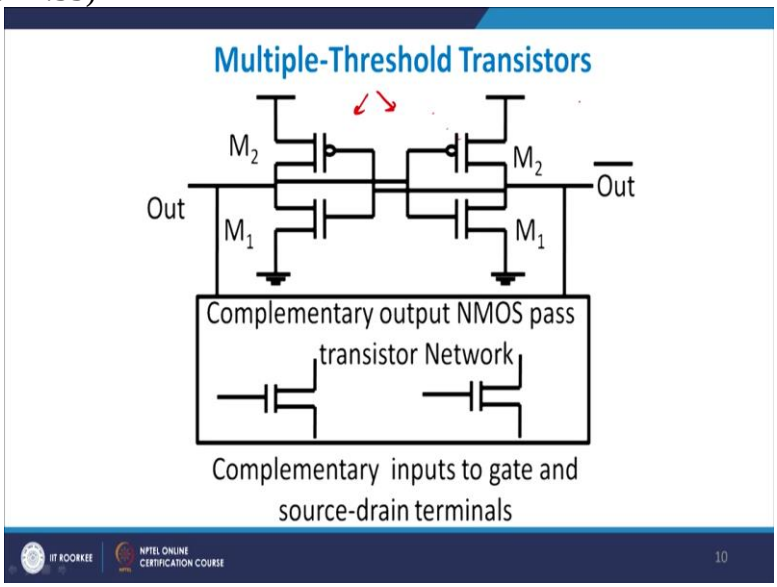
But please understand the loss V will give you is that at that particular instant of time, when your MR is on, you have a direct connection between $V_{DD}$ and ground. And therefore your short-circuit power dissipation will be larger in such a case, right. And that is, this is known as level

restoration at the cost of a slightly higher power dissipation. You see, therefore as the W by L ratio of the, this is known as restoring transistor. This is referred to as a restoring transistor.

So as the aspect ratio is increased, which means the W of the restoring transistor is increased, the voltage level at this particular point is not allowed to fall drastically. So when you reduce your W, it goes down to almost 0 but as you increase the value of W from 1 to 1.75 with L fixed as 0.25, the voltage is actually latched to a value approximately 1.75 or approximately 1.60 or something like this. So you got the principle correct that, so why is it like that?

Go back to my previous module to appreciate this point that when your aspect ratio is large, you have made the device stronger, higher current is flowing through the device and therefore the output voltage which this was trying to go to the ground is now pulled back, is pulled up by the $M_r$, restoration transistor and therefore you see this is what do you get, the profile which you get right. And therefore, you are not able to get the peak to peak swing so at the cost of that you are able to restore the internal level at node X.

(Refer Slide Time: 24:53)

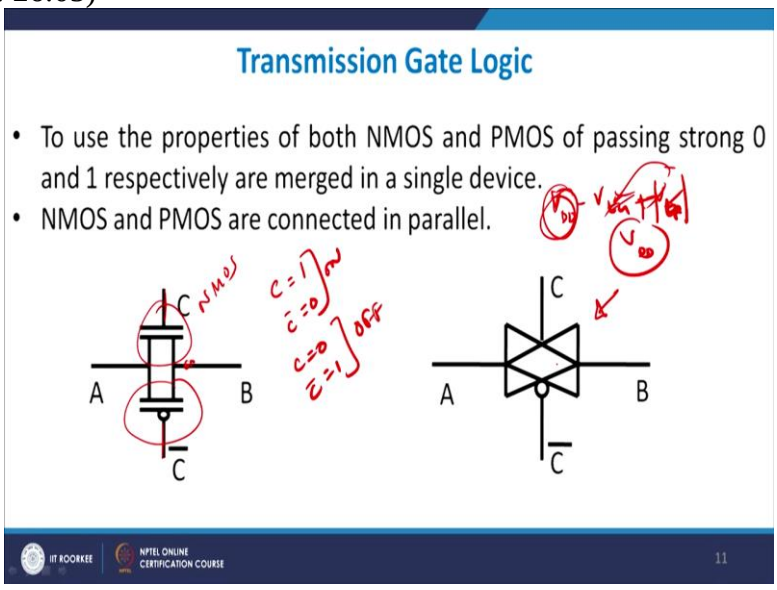



Let me come to multi-threshold voltages. What we do here is that we have a, so this is basically your two cross coupled inverters connected back-to-back right and I will get out and out bar and then this is complementary output NMOS pass transistor logic, right. So what we do is that we try to make the threshold voltages of this one slightly larger as compared to the threshold voltage of the devices kept inside the complementary NMOS pass transistor logic.

Once you initiate that or once you have that into consideration, you automatically are able to correlate or able to do some amount of correlation between the output at this stage and output at this stage, right; so basically what we do is that we try to keep the complementary output module threshold voltages much lower and we try to keep the threshold voltages of the cross coupled inverter much higher right.

And what it does therefore is that report if it is much higher, they are more strong and therefore the voltage of out and out bar are actually latched to its value and these cross coupled inverters try to stick to these values in a proper manner, right. So this is the advantage of having a multiple threshold transistors.

(Refer Slide Time: 26:05)



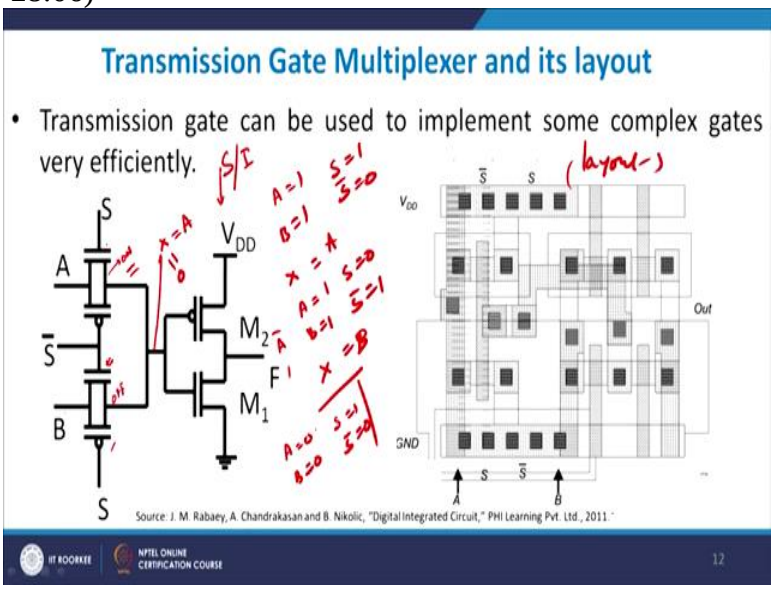

Now let me come to the last part of the or sort of, this is transmission gate and transmission gate is very simple straightforward. You do have an NMOS here, right? You have an NMOS here. For AND, a PMOS in connected back-to-back. This is driven by C and this is driven by C bar. So whenever my C equals to 1 and my C bar therefore equals to 0, I ensure that this is on right, on. And whenever C equals to 0 and C bar equals to 1, I will assume it to be off or it will be off also in a sense because you see when C equals to 1, this NMOS on and C bar equals to 0 means this PMOS is on.

So both the transistors are on and you have a direct connection between A and B. When C equals to 0, C bar equals to 1, this is cut off and therefore there is a high resistance between source and drain of this transistor and therefore A and B are disconnected with respect to each other right. And therefore they are connected in parallel and they are driven by one clock or one data input, C and its C bar, it's complementary.

This is the schematic which you see here. Remember why do we do a TG is very simple. You remember that NMOS was a good puller of 0 and PMOS was a good puller of 1. We have already seen that point. Now therefore if you combine the two together, two transistors together, then both of them cancel each other's domain assuming the threshold voltages are exactly equal and the whole transistor is able to pull the input node right till $V_{DD}$, right.

So what you will get in the output side is something like this. You will get $V_{DD}$ minus $V_{TN}$ right and then you will subtract $V_{TN}$. So you will add $V_{TN}$ here and this cancels up and you get $V_{DD}$ in the output side. So we add mod of $V_{TP}$. Now assuming that mod of $V_{TP}$ equals to mod of $V_{TN}$, these two gets cancelled off and I get again $V_{DD}$ in the output side. Right and that is a standard way of looking at a transmission gate.

(Refer Slide Time: 28:06)




I will show you how I can implement this logic gate and let us see how it works out. I have this, this is basically a static inverter. This part is static inverter. Let us suppose A equals to 1 and B equals to 1 and let us suppose S is also equals to 1, then S bar equals to 0. So when S equals to 1,

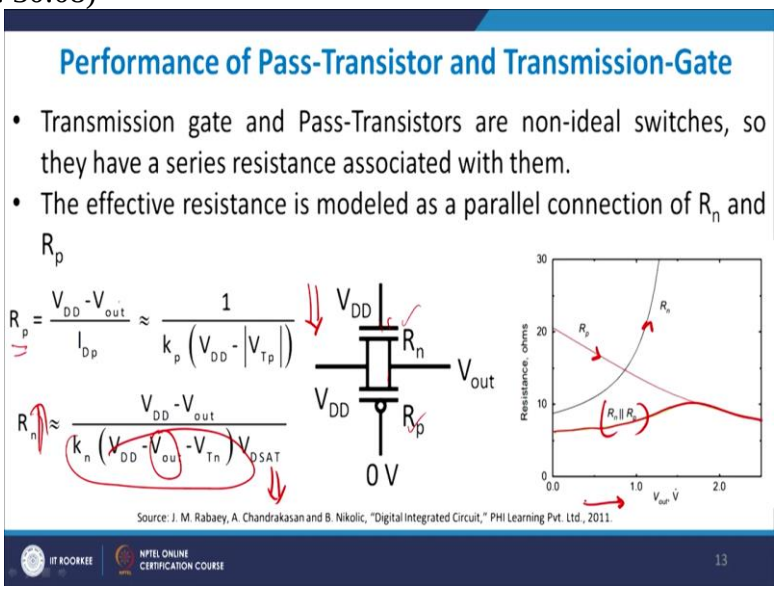
S bar equals to 0, right, my this point let us suppose this is X, X will be equals to, so when S equals to 1 and S bar equals to 0, this is on right and this is off. If this is on, then A will appear in the X. Fine.

Similarly, if A equals to 1 and B equals to 1 and S equals to 0 and S bar equals to 1, what will happen? S equals to 0 and S bar equals to 1 primarily again means that this will be on, right S bar equals to 1 and S equals to 0 means B will be on. And X will be equals to B. Fine. And X will be equals to B. Is this conceptually clear? Right and there will be no drop in the voltage. In fact, there will be no switching of the voltage from $V_{DD}$ to $V_{DD}$ minus $V_{TN}$. It was still remain the same.

Similarly, if A equals to 0, B equals to 0 let us suppose and S equals to 1 and S bar equals to 0, in that case what will happen? If S equals to 1 and S bar equals to 0, means this condition will be there and A and B are both equals to 0 implies that this will be switched on and therefore X will be equal to 0, right. And therefore I can get all sorts of logic principles here. Now 0 will be obviously will be made equals to 1. When you have A, you have got A bar consider here. So I can implement certain logic gates in this case.

Just tell me what this logic gate will be all about right. You need to find out. You yourself do it and need to find out. This is the layout. On the right-hand side is basically the layout of this logic gate. It is quite complex in nature because you are using multiple gates with multiple clock loading. And that makes life difficult for us.

(Refer Slide Time: 30:08)




Let me come to the performance of the transmission gate. So remember, transmission gate is made up of NMOS and PMOS. This is NMOS and PMOS. So you see, $R_P$ which is the resistance of PMOS will be $V_{DD}$ minus V out by $I_{DP}$, where $I_{DP}$ is the current flowing through the device which is approximately equals to 1 by $K_P$ $V_{DD}$ minus mod of $V_{TP}$. And similarly, $R_n$ is $V_{DD}$ minus $V_{out}$ upon $K_n$ minus $V_{DD}$ minus $V_{out}$ minus $V_{TN}$ into $V_D$ sat, assuming that both are in the saturation region.

So you see, therefore as the input or the output voltage goes on increasing right, the NMOS resistance goes on increasing. The NMOS resistance goes on increasing. And the reason is that this value of voltage goes on increasing means this whole thing starts to decrease. As a result, RN goes on increasing, whereas RP actually starts to decrease with $V_{out}$ going high and high. And that is see here, red curve.

So the overall curvature of $R_n$ parallel to $R_P$ because these two are parallel in dimension, is something like this in the green curve almost independent of the value of $V_{out}$ and that is one of the major advantages of a transmission gate that the transmission gate output or the effective resistance offered by a transmission gate is almost independent of the value of the signal which you give, right.

Whereas for all other cases, it happens to be a strong function of the signal, right and that makes my life easier as far as performance of the pass transistor logic is concerned. So we have

understood the pass transistor logic, we have also understood the complementary logic, we have also understood the PTC which is pass transistor logic as well as transmission gate logic. So for reduction in the transistor account, we use pass transistor.

For example, an AND gate would require 4 transistors assuming the complement is also to be taken out for one of the signals. Whereas when you use a AND gate for your CMOS transistor, it takes six transistors into consideration. In a differential pass transistor, I will assume both true and its complementary inputs and both types of output, true and its complementary output is formed. Typically as I discussed with you, when you do a pass transistor, there will be a fall in the voltage.

To remove that, we use a level restoration circuit. We can also use a multiple threshold circuit or we can use a transmission gate, either of the three. The problems with these logic designs are that they might replace CMOS but they are not as robust as your CMOS is. So they are not too much preferred in your logic domain or in design domain, right. In some cases, we do use transmission gate based but then you are actually increasing the number of transistors in that case, right and your clock load also starts to increase.

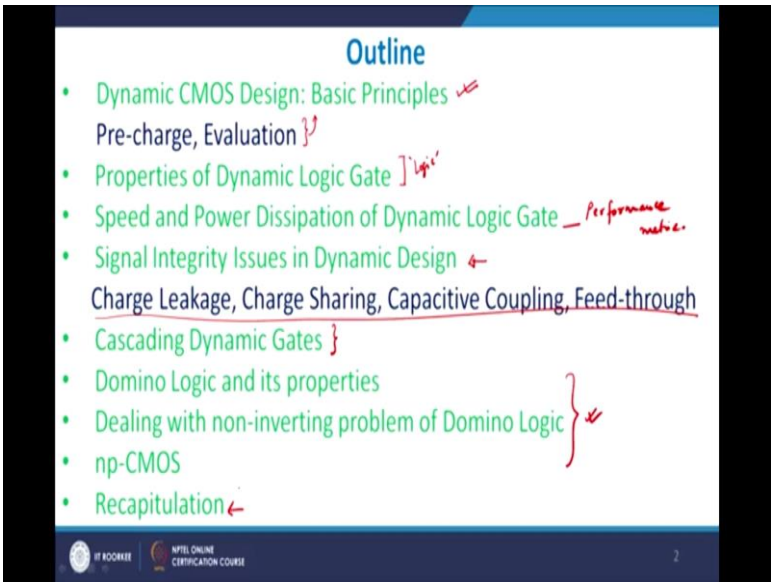
With these words, we have therefore understood the basic difference between PTL, pass transistor logic and transmission gate, we have also understood what is a pass transistor logic and how is it depending on the value of threshold voltage, what is the meaning of cascading of logics and how can I remove the problem of a loss of signal using level restoration circuits, right. With this, we finish this module and we will come back in the next module with other explanations. Thank you.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−56**
**Combinational Logic Design-IV**

Hello everybody and welcome to the Online Certification NPTEL Online Certification Course on microelectronics device circuits. This module is combinational logic design module number four. In our previous modules we have looked into past transistor logic, we have also looked into transmission gate logic and we have seen how these logics behave under various signals given to it. What are the relative advantages and disadvantages of each of these logic design.

Today, in this module we will take up dynamic logic and see how is it different from static logic. As I discussed with you in the initial slides the very first slides of my combinational logical blog that static logics are those logics which are in which the output is either connected to $V_{DD}$ or to ground or $V_{SS}$ and therefore it is basically a low impedance node whereas when you talk of dynamic logic you can have a floating node picture available to you.

(Refer Slide Time: 1:33)

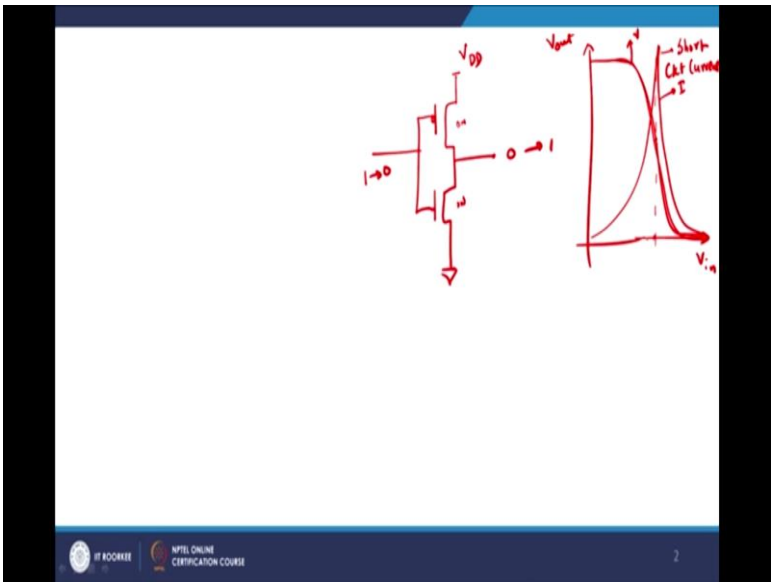



So with this let me come to the first outline of the module and the outlines of the module are: we look into dynamic CMOS circuit design, we understand what is the meaning of pre-charge and evaluation with respect to dynamic CMOS, we look into the properties of dynamic CMOS logic

design, dynamic CMOS logic, then speed and power of dynamic logic, so these are performance matrix here right and we then look into the signal integrity issues of a dynamic design. Typically dynamic design has got a major problem of signal integrity that means there is heavy leakage and so on and so forth so that is what I was in blue it is marked here that within this we will be looking into feed-through, capacitive coupling, charge sharing and charge leakage.

We will try to understand how we can cascade dynamic gates for a performance enhancement and then looking into domino logic and then NP-CMOS we will look into. This is the final one we will do and then we will recapitulate finally the whole discussion. So this will be the outline of the whole talk, as far as this is concerned.

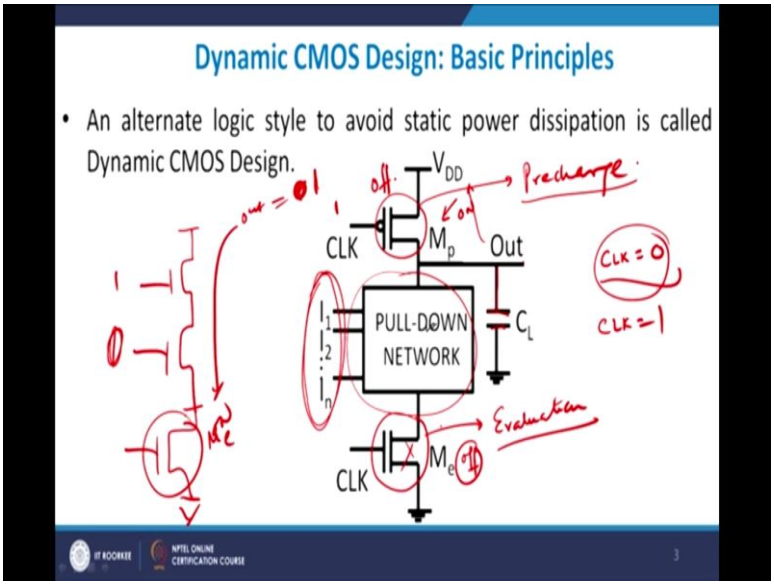
(Refer Slide Time: 3:06)




Let us look at the dynamic logic here. An alternate style to remove static power dissipation is called a dynamic CMOS logic design. If we remember at one point of time when we were using CMOS architecture, simple CMOS architecture and you had NMOS and CMOS as your PUN and PDN. Then remember when you are going from 1 to 0 and the output was going from 0 to 1. If you plot the VTC it comes something like this right? So if plot $V_{in}$ versus $V_{out}$, I get the profiling something like this, right? The profile was something like this, but somewhere in the middle right here effectively both your devices are in the on state for a very short duration of time and as a result if you plot the current flowing through the both the arms you will see that this current suddenly peaks here and then falls down.

So this is the current and this is the voltage. This peak this current is referred to as a short circuit current right so this is the short circuit current which you will see. So this is the short circuit current which is there with you and this short circuit current implies that you will always have a large current at certain point of time.

(Refer Slide Time: 4:10)



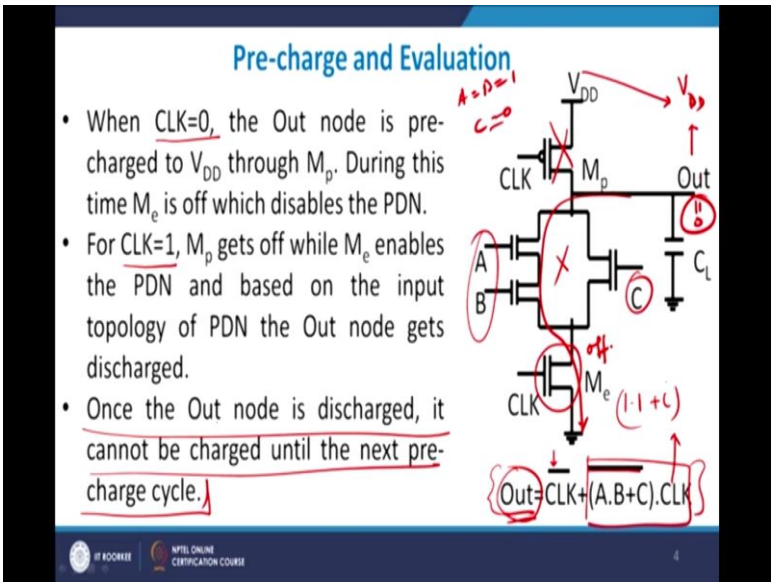

Now let us come back to dynamic logic and see how it is removed. Dynamic logic you have one single pull-up just like your CD NMOS inverter PMOS and you also have an evaluation transistor known as $M_e$ and so this is an NMOS also referred to as an evaluation transistor $M_e$ and this is $M_p$ also known as a pre-charge transistor. We will explain these terms individually but we will come to that later on.

So, and this is the pull-down network which you see, this black box is basically pull-down network with n inputs $I_1$ to $I_n$ here and the output is basically here with this is the load capacitance, which you see. Now it is very simple and straightforward. Whenever your CLK is equals is equal to 0 right your $M_e$ is switched off right and this is switched on. So when this is switched off and this is switched on, output goes to $V_{DD}$ because this is switched on but your, this transistor is not coming to picture at all because here this is been cut off. So when your CLK equals to 0 right and your this transistor switches on, your output goes to 1.

Now when your CLK goes to 1 right, then this CLK this also goes to 1, right and therefore this switches off but then depending on the value of your input vectors I1 to I n and the types of logic design which is there in the pull-down network, either the output will go to 0 or it will not go to 0. For example, if your pull-down network is something like this and then this is your $M_e$ this is already on right? If this both inputs are 1 1 then out which was initially held to 1 will now go to 0 across these transistors, but if your input is 0 here these will hold the output equals to 1, you are getting my point?

So two things happens, one is the type of network which you are choosing in the pull down network that, and the type of input you are choosing here will determine whether the output. So you see at no point of time, $V_{DD}$ is directly connected to the ground and as a result because you have CLK, because you have this driving these CLK transistors. As a result you will automatically have much short circuit power dissipation.

(Refer Slide Time: 6:37)

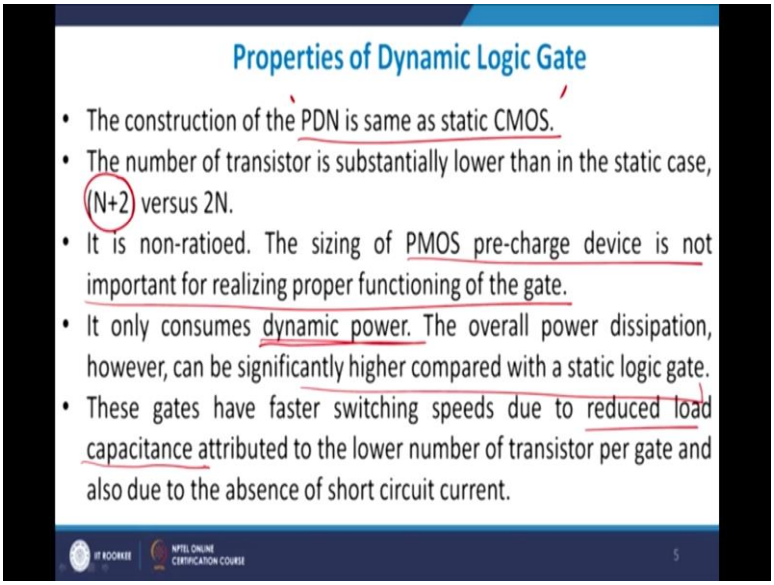



That is what is all written here and so that evaluation, for example, when CLK equals to 0 out will go to, will be pre-charged to $V_{DD}$. This will go to $V_{DD}$, the out value will go to VDD and this will be disabled or this will be switched off right? A pull down network will therefore be disabled so this will not work and therefore out will go to $V_{DD}$, now when CLK goes to 1, $M_p$ switches off right and based on the values of A B and C out will either go to 0 or remain stable.

Now, once you have A B C suppose both, say A equals to B equals to 1 and C equals to 0 then out will discharge through this path and out will go to 0, clear? When out goes to 0 we get an important issue here that you will never be able to recapture or have the value equals to $V_{DD}$ until and unless your CLK goes to 0 once again. So till your CLK goes to 0, your output will never go to back to $V_{DD}$. So in the sense that once out node is discharged it cannot be charged until the next pre-charge cycle.

You understand this point, and therefore out can be written as CLK bar because CLK bar when it is high then CLK will be low so this will be 0 right and therefore out will be equals to $V_{DD}$ out is equal to 1. Similarly when CLK equals to high and CLK bar will be low so this will go to 0 this will go high and then depending upon A B and C out will be determine right and this is basically your, the complement value will be shown to you. Why? Because, for example, if I take 1 dot 1 plus 1 I should get output should be equals to 0 and that can only happen if we have a NOR function or a NAND gate.

(Refer Slide Time: 8:44)



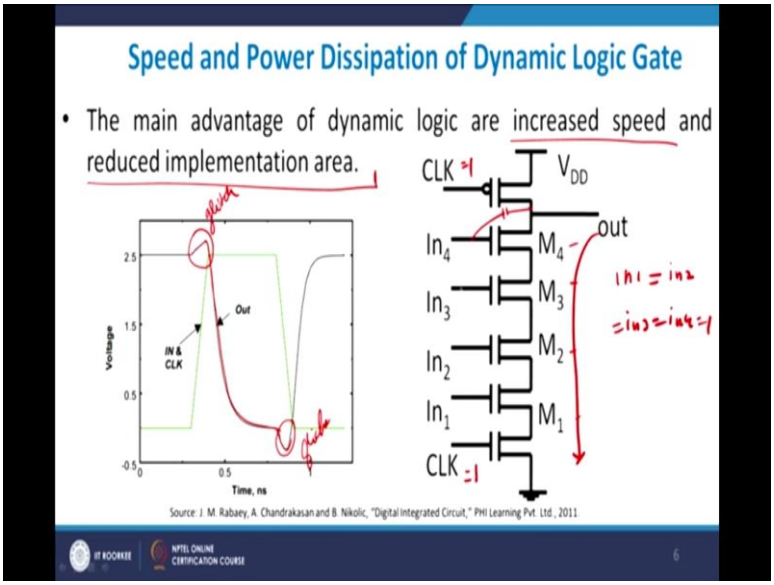

What is the property of dynamic logic? The construction of PDN is therefore exactly the same as a, the construction of the PDN is the same as your static CMOS logic. We have already done that. So PDN network is exactly the same. However, it is much lower than static the CMOS logic. You actually require N plus 2. N plus 2 why? Because N is the number of pull-down networks and plus 2 because 1 pre-charge transistor and 1 evaluation transistor. Of course, it is a

non-ratioed logic right which means the sizing of PMOS pre-charge device is not important for realizing proper functioning of the gate.

Please understand. Proper functioning means logical function of the gate is not determined by the W by L ratio of the PMOS transistor. PMOS transistor therefore W by L ratio will only help you to pre-charge your output node to $V_{DD}$. That is all. It does not play a role in determining the VTC, voltage transfer characteristics in anyway and therefore we define this as a non-ratioed. As I discussed with you, it only consumes dynamic power and therefore overall power dissipation can be highly compared to static logic gate because dynamic power is very high and since we are using CLK, its power can be very high.

However, these gates have much faster switching speeds because if you are reduced load capacitances because why because now the number, see in a CMOS inverter you had almost double the number of gates. Here the number of gates have drastically reduced and therefore your loading of the capacitance actually reduces and the speed of the circuitry improves so dynamic circuits are relatively faster in nature.

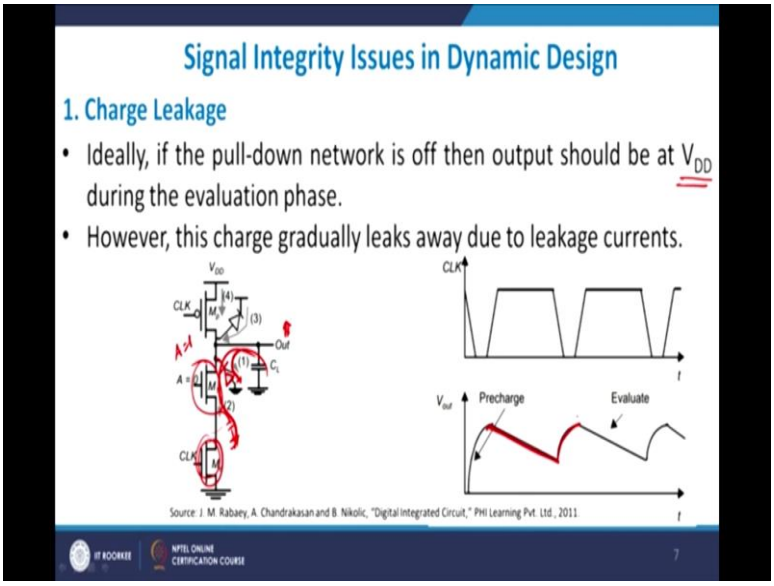
(Refer Slide Time: 10:29)




Let me get into speed and power dissipation in dynamic logic. The main advantages was, as I discussed with you, increased speed and reduced implementation area because of less number of transistors with you. However there is a small problem here that when the input suppose your input goes high and the CLK goes high, then the output should go low and it goes low also but the problem is that output going low across the $M_4$ $M_3$ so whenever my CLK goes to 1 and this goes to 1 then depending upon the value of in 1 equals to in 2 equals to in 3 equals to in 4 equals to 1 then only out will actually discharge through this path which means that the discharging path has got a large resistance being offered right?

So your delays will be typically large and therefore you see your out is basically delayed by a large quantity here right and not only that, your out is also showing a small glitch here and a glitch here so you have a glitch here and a glitch here. This is known as the glitch. This is because of a capacitive coupling we will see later on as we move forward but this is because of capacitive coupling and therefore you will see suddenly, the reason this happens is, that as the input suddenly increases there is a capacitive coupling between input here and the output here right and as a result the output shows a certain glitch in the output side and therefore that increases the capacitance.

(Refer Slide Time: 11:46)



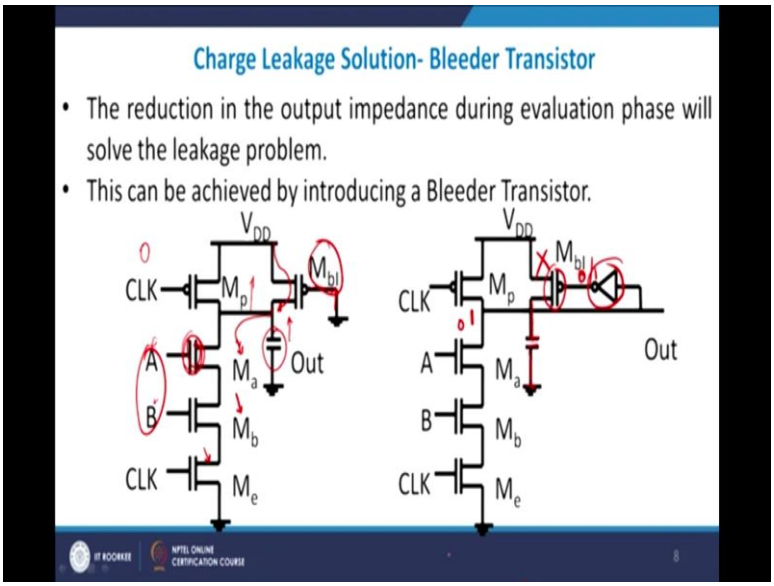

In a dynamic logic, you do have a problem of charge leakage and that is very important or that is very difficult to do that. Ideally, as I discus with you, your pull down methods is off when the output should is at $V_{DD}$. So when your output is at $V_{DD}$ right since your CLK is at low voltage and so this is cut off right? However, there is a problem the problem is that there is a, so when you pre-charge it, your output goes to high value right?

It goes to high value but though your $M_e$ is switched off you still have let us suppose A equals to 1 now and this transistor was switched on, M1 was switched on so any cap here or any reverse biased connected diode here which is basically the depletion capacitances, the charge will actually leak through these domains and therefore the pre-charge value will start to fall down until and unless the next CLK pulse comes and tries to raise the value to a high value.

So you are losing the charge because of leakage, charge leakage specially in a case when you have single integrity issues in a dynamic design. So whenever you have a dynamic design you do have a problem that your pull down network though ideally is supposedly switched off but if let us suppose one of the gates connected to the output node is switched on then the output node charge might actually go through this particular device which is a on state device and reside in its capacitances, say a reverse biased PN junction diode capacitance or the $C_L$ load capacitances or maybe parasitic capacitances.

So what will happen is that you will lose it. Once you lose it you do not have any other option but to wait for the next pre-charge cycle to come right? And so that loss is pretty steep so you see it is actually gone down to a larger extent. So now what happens is that when you go now to the evaluation phase which is when your CLK equals to 1 and CLK equals to 1, then you start from here rather than from here so you are not able to utilize the whole swing of the design in this case. So, this is one of the problem areas which you will see.

(Refer Slide Time: 14: 24)

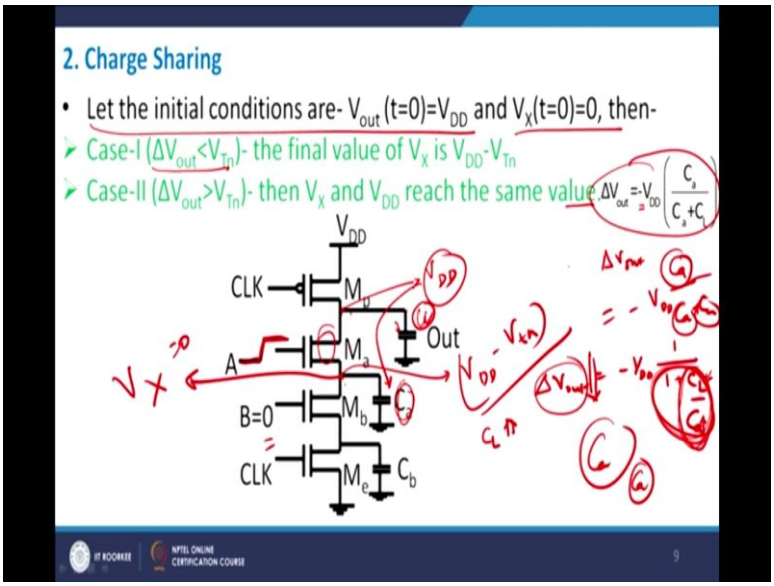



So, what people thought and people did also, they introduced what is known as a bleeder transistor. Now, let us see how it works out very simple and straight forward. What I do is I have $M_a$ $M_b$ so it is an AND gate. This is my evaluation transistor here and this is $M_a$ input 1 input 2 A and B are inputs here CLK and I have put a bleeder transistor $M_{b1}$ there right? So let us see how it works out. So let me say that CLK is basically equals to 0.

This charges up to a very high value, so this charges up to very high value this is a normally on device because PMOS is connected to ground and since this typically high value sorry this becomes a high value even if these transistors are switched on if suppose, this transistor is switched on even if some discharge takes place I still have $V_{DD}$ connectivity here and the $V_{DD}$ will appear and charge this node once again so this node will be again charged to a high value. So you do not lose charge because you are applying a charge from a external world.

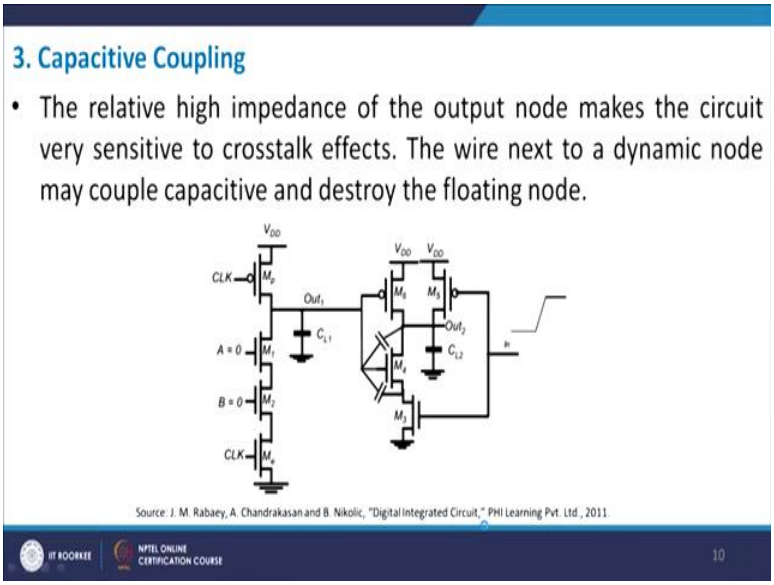
Similarly if I take a NOT gate here and if I put suppose this is for example if I take this to be as equals to 1 right, because it has charged to this value this will become 0, this will become 0 implies that $M_{b1}$ will be on $M_{b1}$ will be on means that this $V_{DD}$ will be always connected to this particular capacitance and the capacitance charge will never be lost. You will have always a high capacitance value there and when my this goes to 0, when this goes to 0 this becomes equals to 1, this cuts off and you enter into the evaluation phase. So this $M_{b1}$ is defined as my bleeder transistor, bleeder means it is basically bleeding the charge from $V_{DD}$ on to the capacitance which you see in front of you.

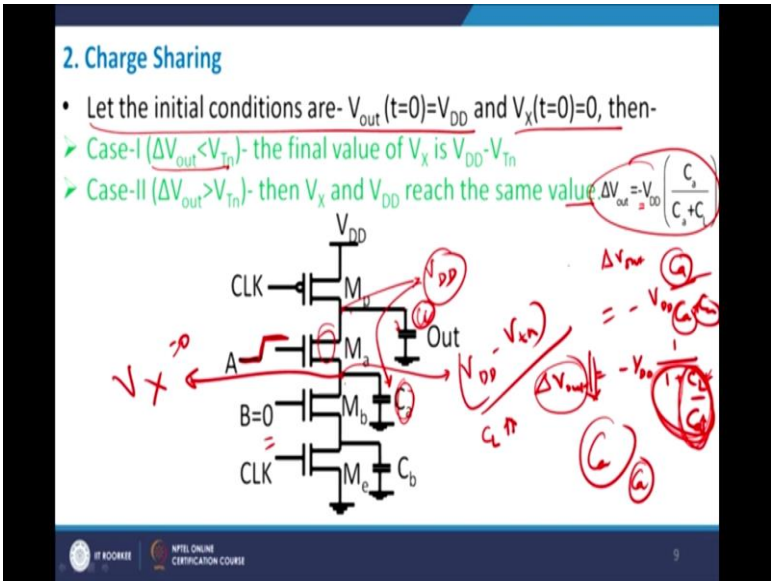
(Refer Slide Time: 15:50)




One classic example which you see, one of the classic example is basically to do with, I will just discuss with you the classic problem.

(Refer Slide Time: 16:07)




The classic problem of this thing is that when you have a capacitive coupling between the two we also have a charge sharing. What does a charge sharing mean? Charge sharing means that whenever your device switches off, the charge which is there within the device has to sit somewhere right? It has to go somewhere right? And therefore, it will, that charge will be shared between two or three devices and its voltage will go up.

I will explain to you this is worked out. Let us suppose initial conditions are my $V_{out}$ is equals to $V_{DD}$ and $V_x$ equals to 0 so $V_{out}$ is $V_{DD}$ so I have already this is $V_{DD}$ right? And I am already into

evaluation phase. $V_x$ which is the voltage at the intermediate point. This is intermediate point x $V_x$ equals to 0 because obviously you are switched on and now what I am doing is I am giving a signal A equals to 0 to 1 and B is stabilizing at 0 right?

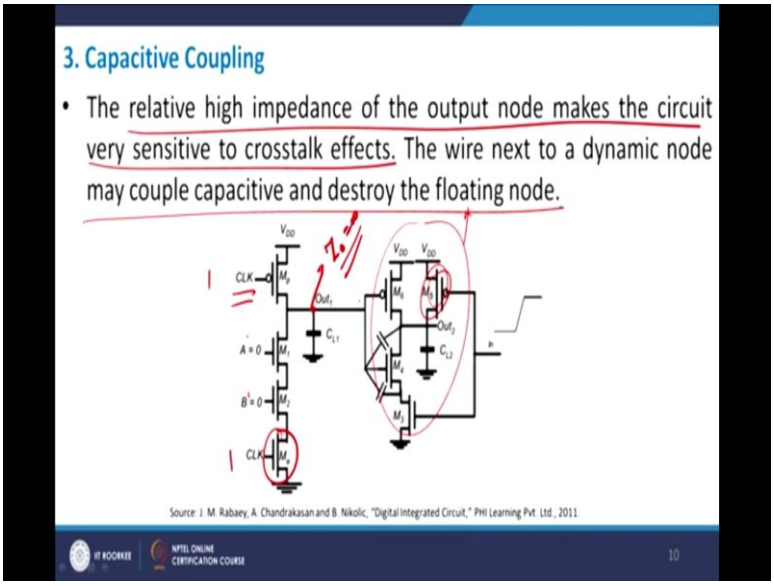
Now, if $\Delta V_{out}$ is not smaller than threshold voltage of this device right? $\Delta V_{out}$ being the change in gate voltage gate voltage is smaller, then if this is switched on this voltage will appear here. It is behaving like a past transistor logic as $V_{DD}$ minus $V_{tn}$ right? Now but if $\Delta V_{out}$ change in the $\Delta V_{out}$ value is very large and is much larger than the threshold voltage of this $M_a$ then $V_x$ and $V_{DD}$, these two will reach to the same value and your $\Delta V_{out}$ will be, what is the condition for $\Delta V_{out}$, is $V_{DD}$ upon $C_a$ by $C_a$ plus $C_L$. $C_a$ is this one and this is $C_L$ right?

Which means that it is basically a capacitive coupling and depending on the effective values of $C_a$ and $C_l$, my $V_{out}$ will be there. So if you look at $\Delta V$ out it is equals to minus $V_{DD}$. If we divide numerator and denominator by $C_a$ I get 1 by $C_L$ upon $C_a$ so you see if your $C_L$ is very very large as compared to $C_a$. This quantity will be very large right and therefore this quantity will be very small right and sorry yes so it will be very small and therefore $\Delta V_{out}$ will be very large.

It will be so also because if $C_L$ is very large the most of the charges will be accumulated at $C_L$ and the voltage change at output will be very large. Similarly if your $C_a$ is larger then $C_a$ is larger means then this is large which means that that this falls down and therefore, sorry I am sorry. I will just reframe my question, the understanding here, if $C_L$ becomes large then this quantity becomes large right and therefore $\Delta V_{out}$ goes down, the reason is that with increasing value of $C_L$, more and more charge will be accumulated here whereas when you increase the value of $C_a$ the more distribution will take place and your $\Delta V_{out}$ will actually become larger and larger right?

This is known as charge sharing. So whenever you switch on $M_a$ right because a goes from 0 to 1 some charge from this $V_{DD}$ node goes on to $M_a$ node and therefore $\Delta V_{out}$ shows a small drop right and this drop is you see negative, you see it is negative actually and as a result you will always see a drop right and this is charge sharing phenomenon in a domino logic in a dynamic logic.

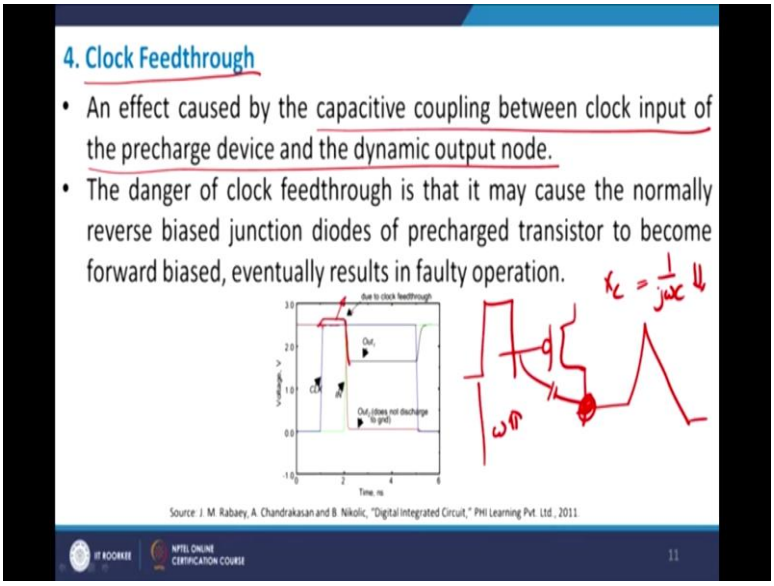
(Refer Slide Time: 19:18)




Let me discuss with you the capacitive coupling therefore. The capacitive coupling, so my output impedance is relatively high and therefore very sensitive to cross talk effects. That is the only problem area of a dynamic logic because this is quiet high $Z_{out}$ is equals to infinity ideally right? The wire next to the dynamic node may couple capacitive and destroy the floating node right and that is your problem area which you face for a dynamic logic design.

So what people did so this is a bleeder circuit which you see in front of you and this is a bleeder transistor here which tries to make the level again go to back to $V_{DD}$. So even if there is a problem with your output because of charge sharing or because of leakage. This helps you to raise the value to a high value, right but the problem here is that why is it Z equals to infinity is it is behaving like a floating node because why because your CLK has actually gone to 1 CLK is equals to 1. This is switched on but let us suppose A and B are both zero then this will be acting as floating node with high output impedance right and therefore they will all be affected by crosstalk noise and electromagnetic interferences will happen here to a larger extent right? And this is therefore known as, this will result in externally signal being degraded to a larger extent.

(Refer Slide Time: 20:34)

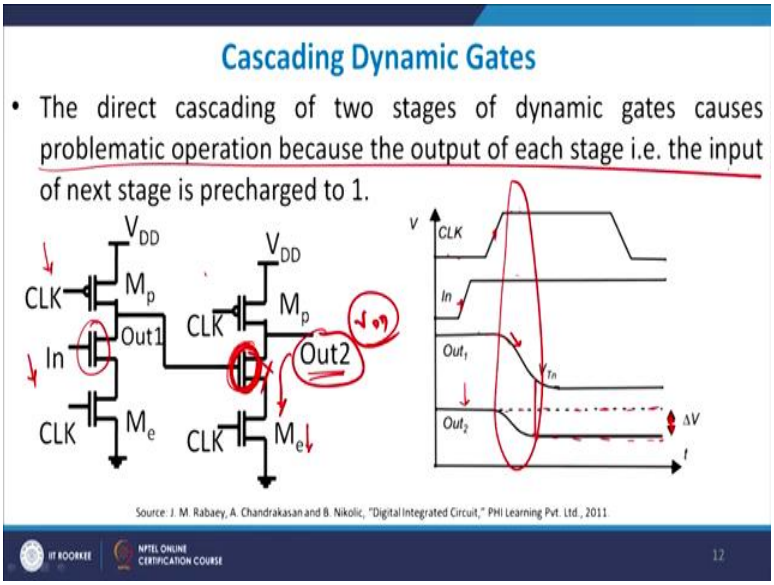



Last, let me come to the last problem area of this thing and this is what is known as a clock feed through. Clock feedthrough is basically a capacitive coupling between the clock input of the pre-charge device and the dynamic output node. So if you look back here this was clock here and this was the output here. So your clock goes high like this right? Now because of a coupling between this and this, the node voltage also goes like this.

So you see here when the clock goes right high the output should remain stable like this to $V_{DD}$ but actually it goes beyond $V_{DD}$ for some duration of time and then starts to fall back. This is because of the fact that for a finite duration of time because you see, clock is rising very fast which means that this ω is very very large right? So ω is large means ω is 1 by jωC is very very small Xc.

So Xc is small means it is looking almost like a transparent medium. So any small change in the clock here, high frequency change will be easily detected at the output node of my pre-charge and that is the reason you will have a voltage and as a result what happens is that the voltage rises beyond $V_{DD}$ right and that results in what is known as a clock feedthrough and this results in a larger power dissipation and signal integrity issues so power dissipation is a problem.

(Refer Slide Time: 21:54)




Now let us look at cascading of two dynamic gates. So I have cascaded how so output is cascaded back into the input of the second gate. Now you see if my CLK goes high and my input also goes high my Out1 will start to fall down. Obviously the input goes high means this is on and the output starts to go down. It starts to go down means, this so it was initially on this was high right it was high, it was high means I should get Out2 be equals to 1 to 0 why it is 0? See initially when your CLK is low and input is low CLK is low input is also low out 1 will be high which means that this is on but then your CLK is low and input is low therefore this will be off and Out2 will be so the CLK is low means Out2 will be also latched to $V_{DD}$. That is what is happening here.

Now as you input those high it take some finite amount of time to output to fall down output starts to fall down as it crosses the threshold voltage of the next transistor. At this stage my this goes to off state. Which goes to off state? The next device goes to off state. As it goes to off state my out then fixes to that same value where it goes to off state. Ideally it should have been here only you understand?

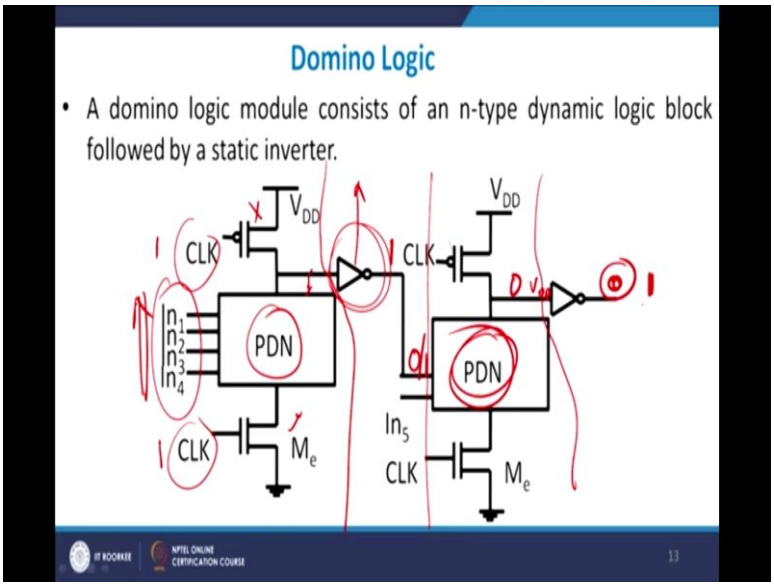
Ideally this should be here because Out1 was initially high and then starts to fall down but because of a finite delay between the two this transistor will get switched off at certain amount of time which is threshold voltage of the device and as a result what will happen is that the Out2 will be latched to a value which is not exactly equal to its initial value and therefore there will be

a ΔV change here and I cannot recover it until and unless I come back to the next pre-charge side.

So you see, Out1 has fallen down out1 is falling down so Out1 when it was initially high CLK is high and so CLK is high and input is also, so if you look at this, concentrate on this part so when CLK is high and input is high then $M_e$ is on, right? And out should actually fall down and it is falling also but it should ideally go to a low value, it should go down if you allow this transistor to go to large value but then Out2 when it crosses the threshold voltage of the second transistor it cuts off this second transistor and this latches to a value which is a not equal to its initial value right?

And therefore you lose so when you cascade dynamic logic you end up losing that is the problematic operation because of the output of each stage okay? And that is the problem. Direct cascading is therefore not a good idea.

(Refer Slide Time: 24:27)



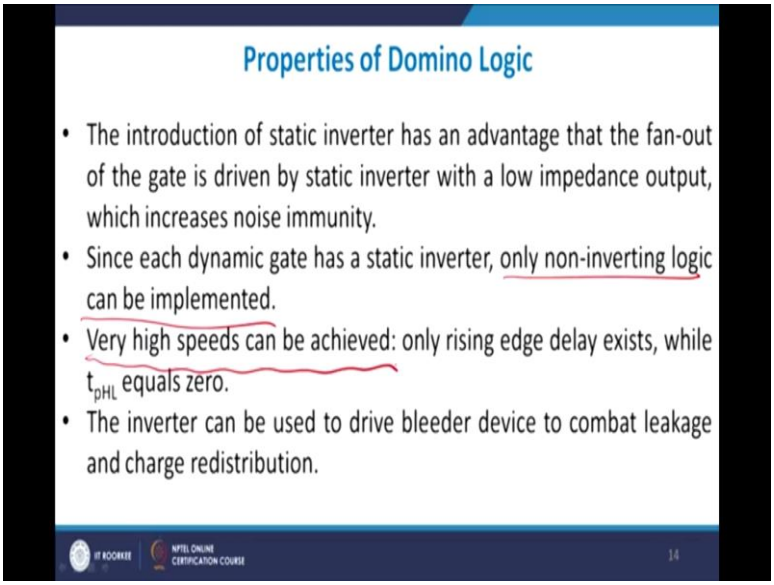

So what people thought was, that since dynamic cascade dynamic you cannot directly cascade two logics because the logic falls down and not only that you will not be able to recover that logic until the next pre-charge cycle come. We enter into what is known as a domino logic. So I have inputs here of course and I have a PDN here and a PDN here. I also have a CLK and a CLK

bar. So this, if you look at closely it is exactly the dynamic logic. Similarly this is also a dynamic logic here right? But they are cascaded by a static inverter right? And they are cascaded by a static inverter.

So you see, this static inverter will always ensure that if suppose CLK is equals to 0 then this will $V_{DD}$ so this will be 0. This will be 0 implies that your PDN will be switched off and therefore if it is switched off and CLK equals to 0 this will be actually $V_{DD}$ and output will be equals to 0 right? Now when this input goes all high let us suppose and it is a AND gate, CLK goes to 1 here. This switches off and this switches on and this voltage starts to fall but nothing will happen here until and unless this goes below the switching threshold of this transistor which will make this goes to 1 and this will go 1 and therefore then it will switch on my PDN.

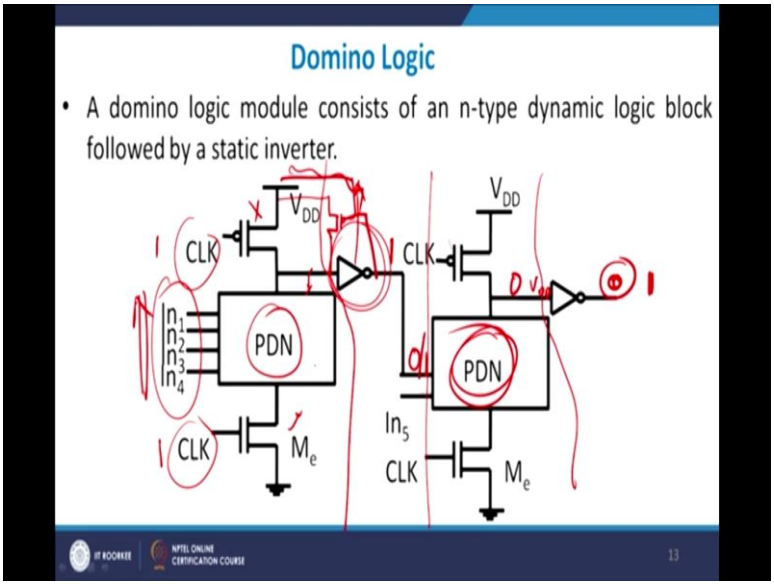
When the PDN goes high this goes to 0 and then output becomes equals to 1 right? So by just inserting a static logic here a static CMOS logic here I am able to restore the signal integrity issues in a domino logic right? So the introduction of a static inverter has an advantage where the fan out of the gate is driven by static inverter which is obviously a low impedance node and therefore it increases the noise immunity. So noise immunity problem is gone.

(Refer Slide Time: 26:11)



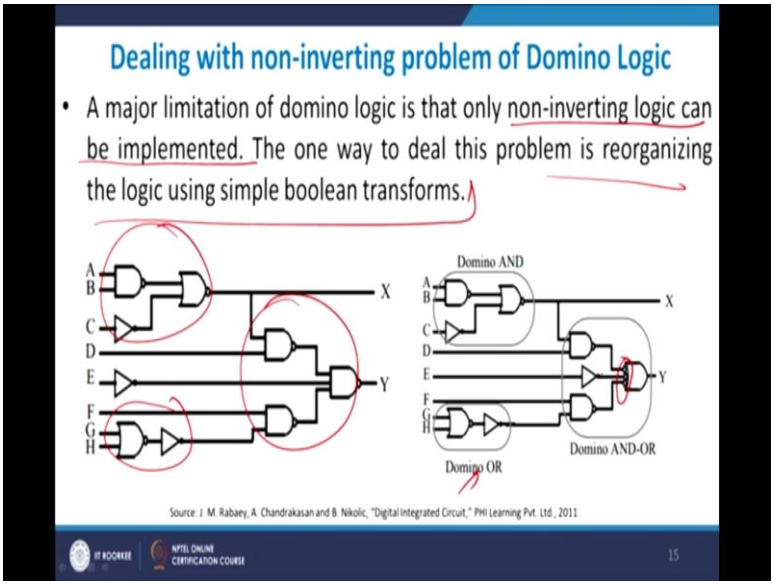
Since each dynamic gate has a static inverter only non-inverting logic can be implemented but obvious very high speeds can be achieved in this case and the reason being that $t_{PHL}$ being approximately equals to 0. High to low will be approximately equals to 0 right? Now the inverter can be used to drive bleeder device to combat leakage and trans distribution.

(Refer Slide Time: 26:40)




How you can do that sorry how you can do that, you can add a bleeder transistor here right and this bleeder transistor will be added here and this will be $V_{DD}$ it will be taken care of in this manner and therefore you can have a bleeder transistor directly inserted here. So charge loss will be taken care of. The charge sharing will be taken care of and the domino logic gives you a very good idea about this whole charge description.

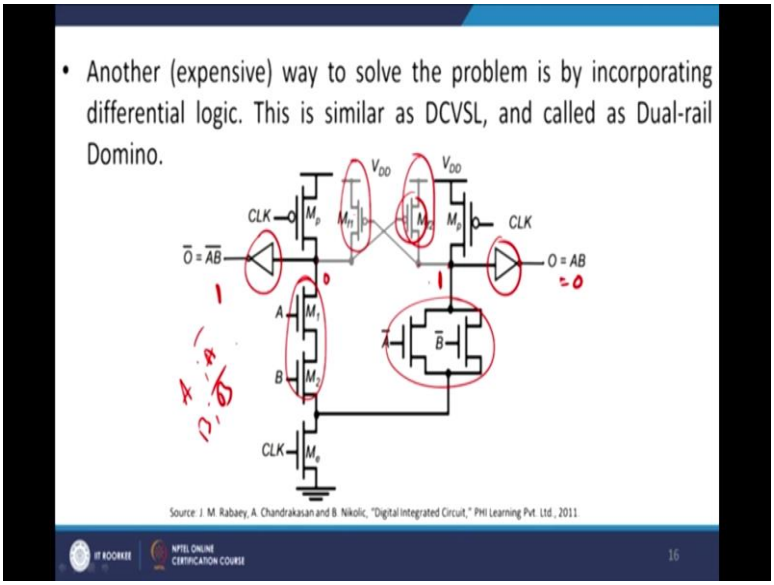
(Refer Slide Time: 27:01)




Now so as I discussed with you therefore a major limitation of domino logics is that only non - inverting logics can be implemented like non-inverting logics means those who do not invert the output. So you just need to re-organize your logic function using simple Boolean transform. For example, I can do a domino NAND, this is to be a domino NAND and this to be as a domino NOR which you can do and this is your domino NAND NOR operation by putting a bubbled OR gate right? I will not go details, will not go into the details of this one too much a large extent but simply by choosing appropriate Boolean functions we can actually remove the problem of non-inverting in logic in a domino function.

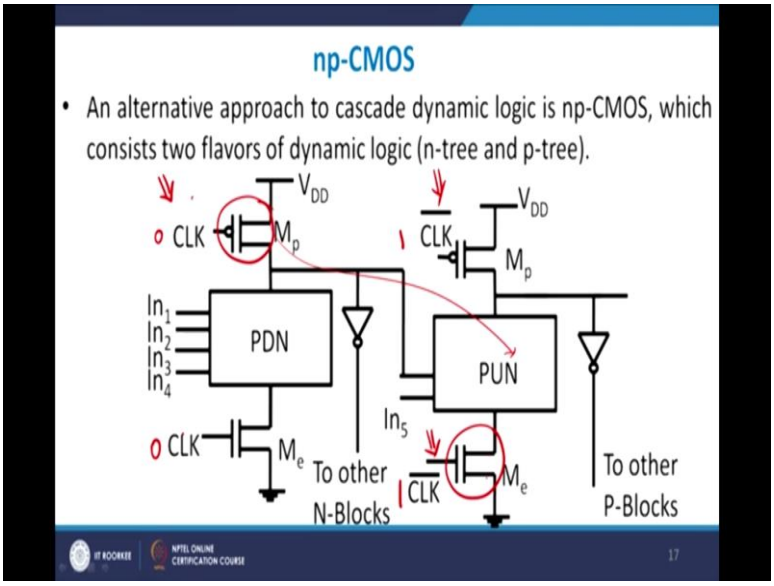
Another way to solve this problem is differential logic. It is exactly the same as DCVSL which we have already learned in our initial modules and this is also known as dual rail domino logic in which we do I have got a static inverter applied here and the static inverter applied here. I have got a feedback or a feedback resistance applied here so you have a positive feedback latch here so once this goes to 0 this switches on this device and this goes to 1 here. When this goes to 1 this goes to 0 here and this goes to 1 here so you have got 0 and 1 coming into picture.

(Refer Slide Time: 28:20)




Now this is therefore you see this is AND gate this is basically your NAND gate so you have series connections here see, so you have a parallel connection here in this place right? So once you ensure that this is 0 and 1 these PMOS transistors $M_{f1}$ and $M_{f2}$ help you to speed up the logic right and helps you to make this go to 0 very fast go to 1 0 in a fast manner. The advantage here is you get both inverted as well as non-inverted logic but the disadvantage here is that you require to generate both your input signal and its inverted signal. So both A A bar has to be generated B and B bar has to be generated in this case right?

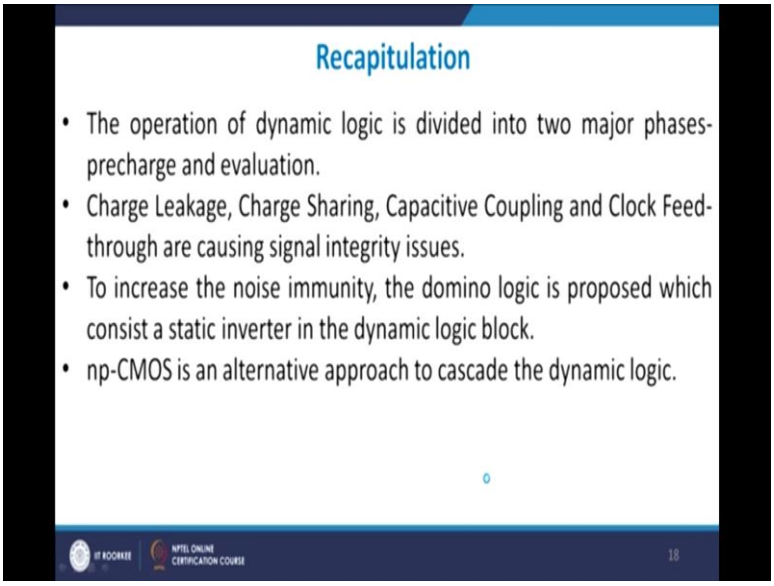
(Refer Slide Time: 29:03)




And that is the problem area for this logic taken into consideration. Another alternative approach is using NP logic. NP logic what we do is that, it is exactly the same as domino logic only thing is that we have a do is that we have a dynamic logic with CLK and CLK and then we drive it with a dynamic logic with CLK bar. So whenever your CLK is high so when CLK is suppose high or suppose CLK is low so this is low, this is low but then this is high and this is high.

So out of all these four designs right if it is 0 CLK is 0 CLK bar equals to 1 means this evaluation transistor will be on and this will be on. This both will be on together and as a result what will happen is that you will have a direct path between these two points provided Me is on provided your input five is a high value which you see right but the NP I will like you to go through it in a much more detailed manner from this book actually and it gives you a very good idea about NP-logic design.

(Refer Slide Time: 30:06)



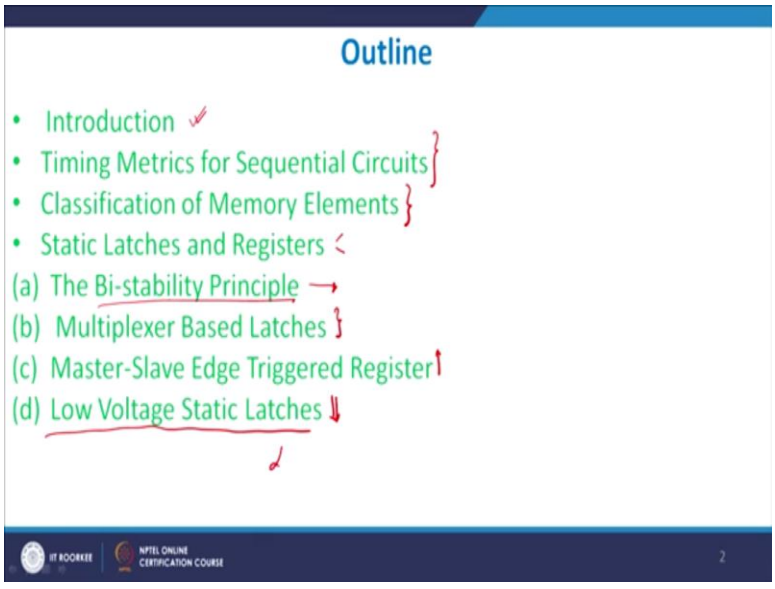

So let me recapitulate what we did till now. We looked into the operation of dynamic logic and we saw that it can be divided into pre-charge and evaluation. We also looked into charge leakage charge sharing phenomenon of dynamic logic and we also saw that there is signal integrity issue in dynamic logic. How to remove it there are methods of doing it, using, for removing the by using static CMOS logic we have seen that and NP-CMOS is an alternative side to cascade the dynamic logic. So cascading ordinary dynamic logic is difficult task. NP logic is one of the solutions to this particular design right? With this we finish off the dynamic logic case for a combination logical blog. Thank you very much.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics & Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−57**
**Sequential Logic Design- I**

Hello everybody and welcome again to the NPTEL online certification course on Microelectronics Devices to Circuits. We will start today with a new topic within the domain of microelectronics and that is design also and that is Sequential Logic Design and this is part one of the sequential logic design.

In our previous modules we had studied about combinational logic, what is the meaning combinational logic and then realization of various gates using combinational logic. We also understood what is the meaning of propagation delay, power area and on which factors this depends. So by this time given therefore a logic Boolean expression, you will be able to design a simple logic diagram, means a simple logic using gates at transistor level. That is at least relevant for the combinational logical block and you should also be therefore able to calculate how much amount of power is dissipated and amount of when it is in dynamic operation and when it is in the static operation. Today, we will start with a totally new section of this design and that is sequential logic design. So let us just have a look at the outline of the course.

(Refer Slide Time: 01:56)



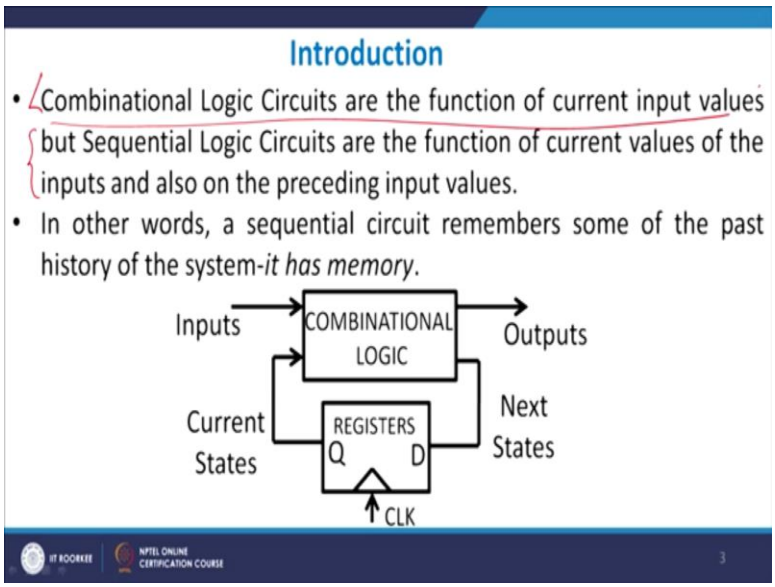

The outline is something like this, we will introduce to you the logic design, we will introduce to you the logic design first of all, what is the meaning of sequential logic, then

very very important term, we will look into the timing metrics of sequential circuits. And we will see that when we are discussing combinational logical blocks, the timing of the clock or the timing of the data was not very sacrosanct or not very important but in a sequential logic the timing at which the data is coming with respect to the clock is a very important perimeter and we would be looking into that in a detailed manner.

Of course, therefore we will explain to you why sequential logic is a memory and therefore the classification of memory elements. We will look into static latches and registers, 2 types of memory. Then we will come to Bi-stable principle, which means that what are the stability points in a sequential design using say for example a simple latch and we will see what is the meaning of bi-stability principle in a latch.

We will also look into mux based which is multiplexer based latch, then Master-Slave edge triggered latch and then finally end our discussion with low voltage static latches. These low voltage static latches is primarily meant for extremely low power technology. So this is for low power, this is Master-Slave is generally meant for those cases where you want to cascade, so this will be a cascaded mode and these are all latches. So mux based latches, and each one us we will, each one of it will be discussed in details regarding its functionality as well as timing requirements.

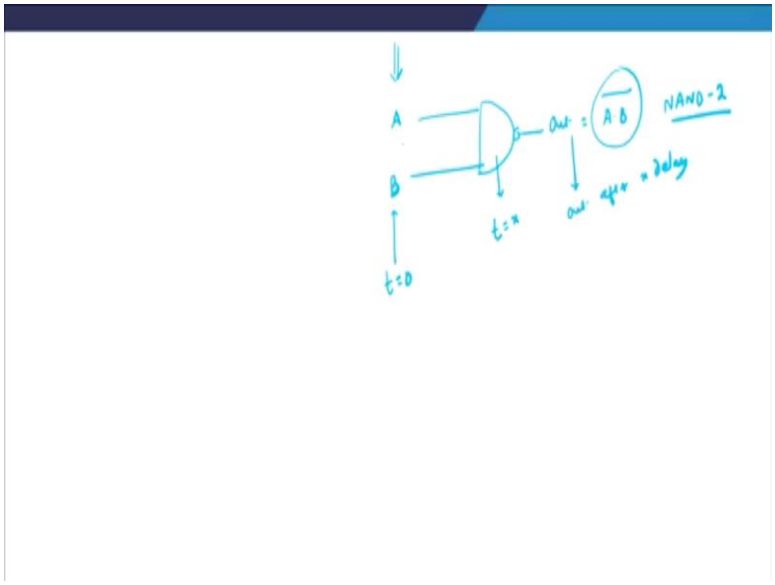
(Refer Slide Time: 03:46)




We have also seen or we will be seeing, we have also seen that if you look at a combinational logical block for example, the as you can see the first bullet here, the combinational logic circuits are the functions of current input values. Let us not, you just do not have a look at

these last 2 lines but just have a look at the first line of the introduction part and you see the combinational logic circuits are function of current input values.

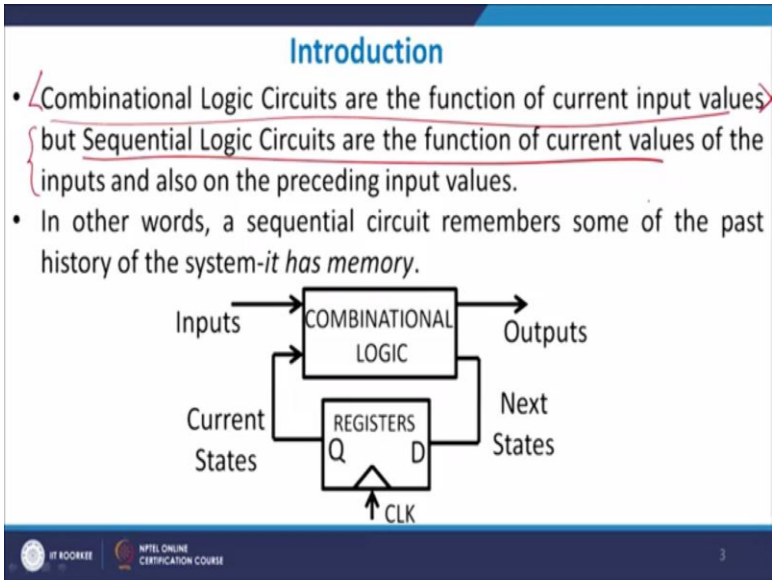
(Refer Slide Time: 04:20)




I will give you an example right? We have already seen this earlier but to give you, just to refresh your memories as far as this is concerned, you can see, understand what I am trying to say. Say for example, you did have a 2 input NAND gate. Now you have 2 inputs suppose and input A and input B and this was output and this was equals to A dot B compliment. So this is a NAND 2 operation which means that output at any point of time which is A dot B compliment will only depend upon A and B which is the input at that particular instance of time.

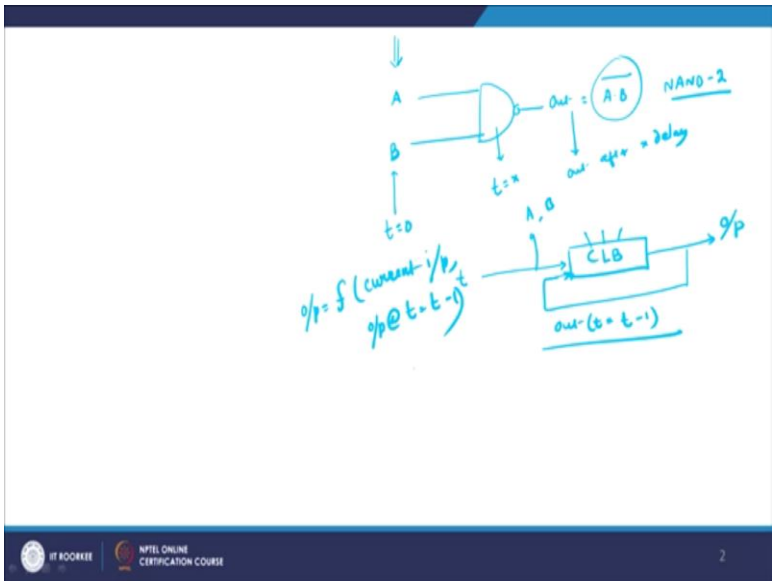
So let us suppose this is happening at t equals to 0 assuming that this is say you have a delay of X then after X millisecond or X nanosecond the output will appear here, that will appear after X delay and this will only depend upon A and B which is true also, it does not depend anything else if you very carefully look. Be it NAND 2 logic, NOR 2 logic, XOR anything it does not depend upon anything else except the primary inputs which are fed into the input side of the NAND gate.

(Refer Slide Time: 05:28)




Same thing cannot be said about sequential logic and that is where it is different but sequential logics are a function of the current values, no doubt just like combination logic but they also depend upon the preceding input values, which means that the output of a sequential logic will depend upon the primary input coming now plus the input available to it at t minus 1th precedence. So everything has happened let us suppose at T, then let us suppose everything is happening at t then in the sequential logical file at t let us suppose I have got a combinational block.

(Refer Slide Time; 06:12)

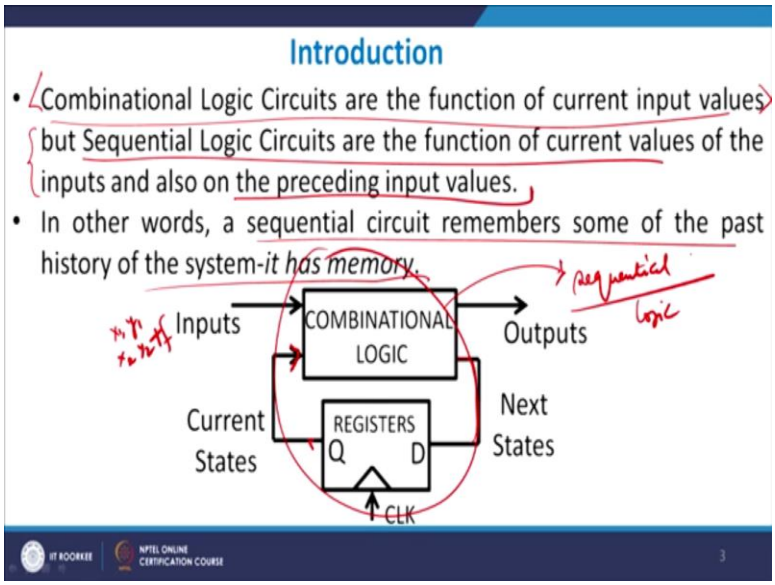



Let us suppose I have got a combinational block C combinational logical, logical block we refined to the CLB and then we get output right but now if this input another input I am

adding here this input is basically let us suppose A and B there are two inputs here and I am adding in the third input which depends upon out of t equals to t minus 1.

Which means the previous state output is not feedback in to the input side and therefore the output now is not only a function of the current inputs but it also depends upon the output at t equals to t minus 1 and that is the basic difference between the sequential logic and a combinational logical block.

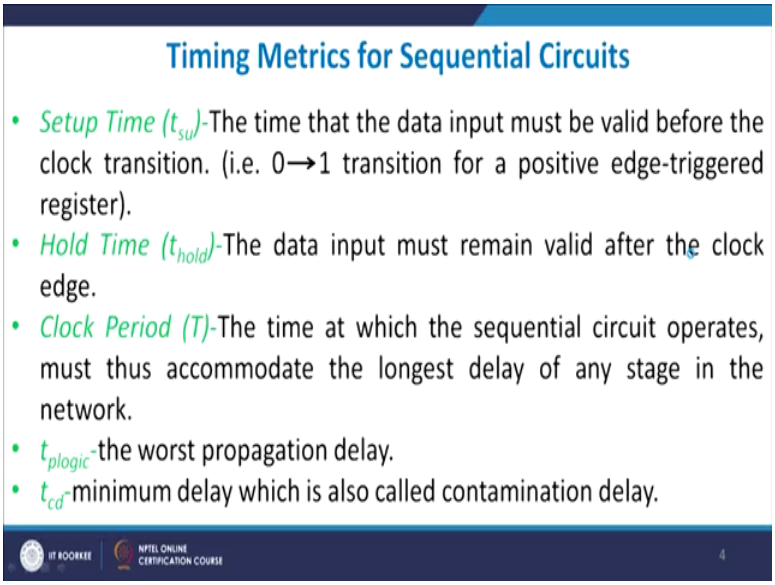
(Refer Slide Time: 06:59)




If you see the same diagram, the sequential circuit remembers some of the past and therefore it has got the memory, as you can see why past? Because you are at least storing some data till the next set of inputs are coming to you. Say you have set some input x and y right, it gets evaluated by combinational logic bar, gets an output and then output is stored in this register here, this is basically a memory element, do not worry about it. And then this are fed into a current value, what current value? The next x.

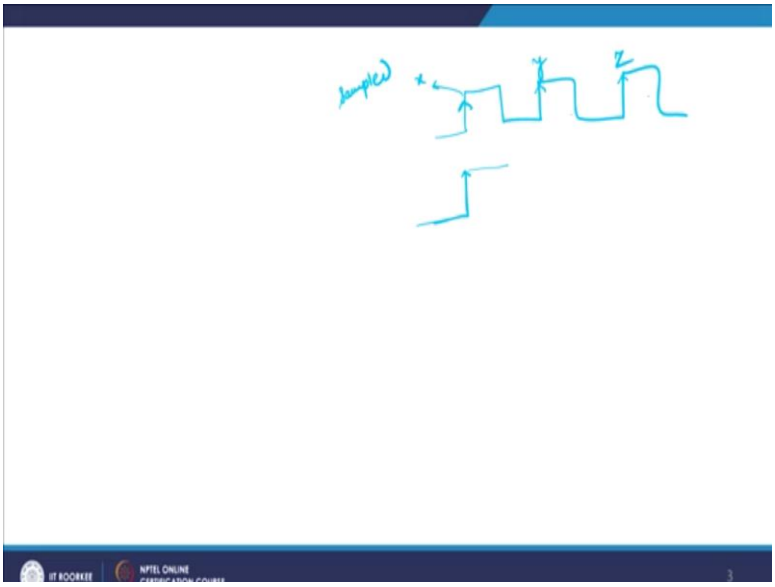
Suppose it is $x_1$ $y_1$ and the next is $x_2$ $y_2$ but $x_2$ $y_2$ plus the feedback from the register will also come which means that this whole logic. So this total is known as sequential logic. This is actually your sequential logic, this whole block of sequential logic will actually store certain bits of information onto its register till a time the next pulse or next input wave train comes. Therefore, output not only depends upon the primary input at that instance of time, it also depends upon the input from the previous state and therefore we also refer to is that sequential logical design also has a memory in it.

(Refer Slide Time: 8:10)




Important points to be remembered, because these are very very important timing diagram metrics for sequential logic and we will go one by one in this case. I will explain to you what do I mean by that, see typically when you have a sequential logic it is clock driven. So what means that…?

(Refer Slide Time: 08:36)

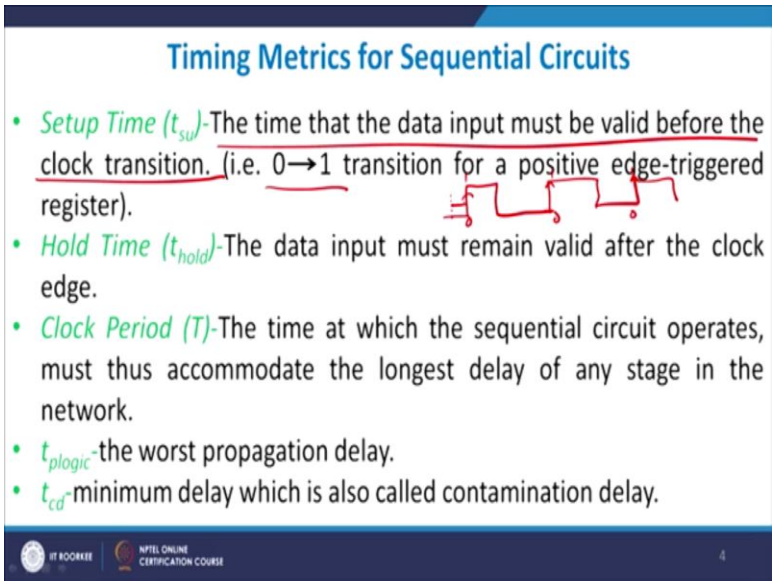



Suppose let us suppose I have a clock here and whenever my clock goes high, my data which is coming from outside gets inserted into the system. So, I will get here data will be inserted, here data will be inserted and here data will be inserted or in technical terms these are referred to as the data will be sampled at these particular points, at x, at y and at z. And so

this is known as and this is therefore known as edge triggered design which means that the data, external data it can only be fed into the system provided the clock is rising.

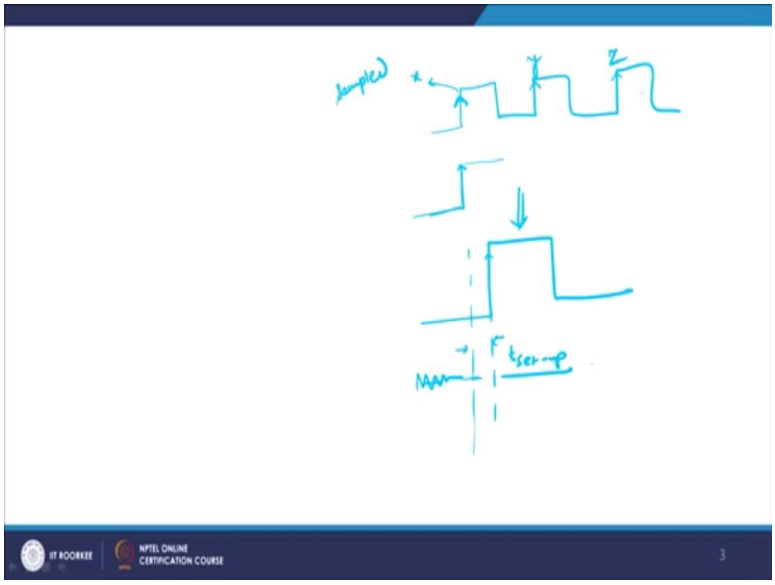
So when the clock is rising and that instant of time you will just catch hold of the data from the outside world. So therefore there are 4 clock cycles, I can catch four sets of data with me. Now the idea, the setup, the concept of setup time is and I will just discuss with you what is the setup time.

(Refer Slid Time: 09:25)




That the time the data input must be valid before the clock transition. So clock will go to 0 to 1 transition as I discussed with you. So when you have edge triggered, positive edge triggered, this is 0 to 1, this is 0 to 1, similarly this is 0 to 1. So these are the all 0 to 1 transitions which is taking place, also referred to as the positive edge-triggered register or positive edge-triggered design. Now you see the setup time is defined as the minimum time as the which the data should remain stable before the rising edge of the clock.

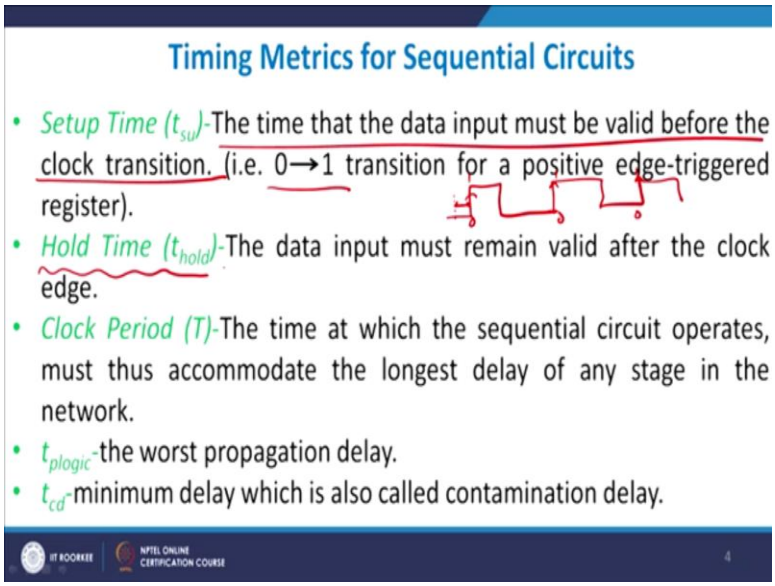
(Refer Slide Time: 10:04)




So, what I am trying to say is that, let us suppose my clock is rising here at this instance of time, right. So your data must be held stable, so your data is moving like this let us suppose, up and down, but minimum it should become stable at this particular point.

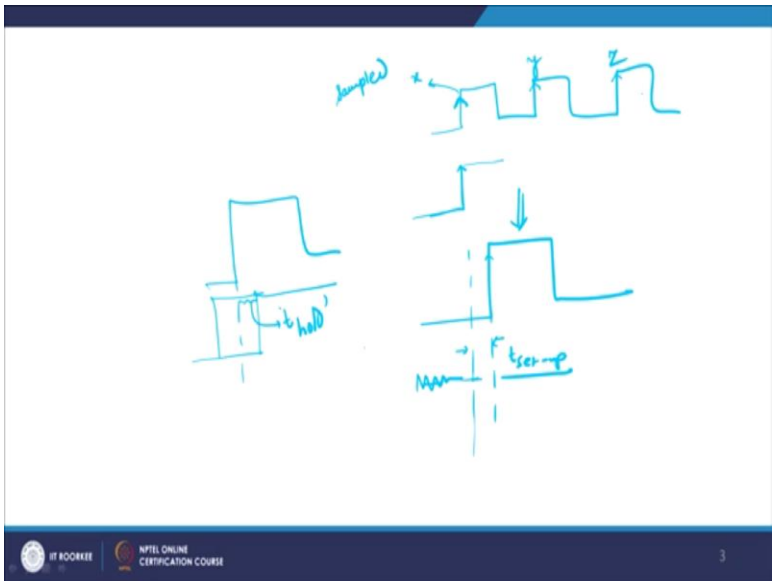
After this, if there is some oscillation in the data from 0 to 1 and 1 to 0, your register or your latch will not be able to evaluate that signal, it will not be able to understand the signal, it will take it as arbitrary noise signal only, you are getting my point? So we define this to be as $t_{su}$, so what is $t_{su}$? $t_{su}$ is the minimum time before the rising edge of the clock in which the data should be held stable and that will allow you to evaluate the signal properly or give you it properly.

(Refer Slide Time: 10:48)



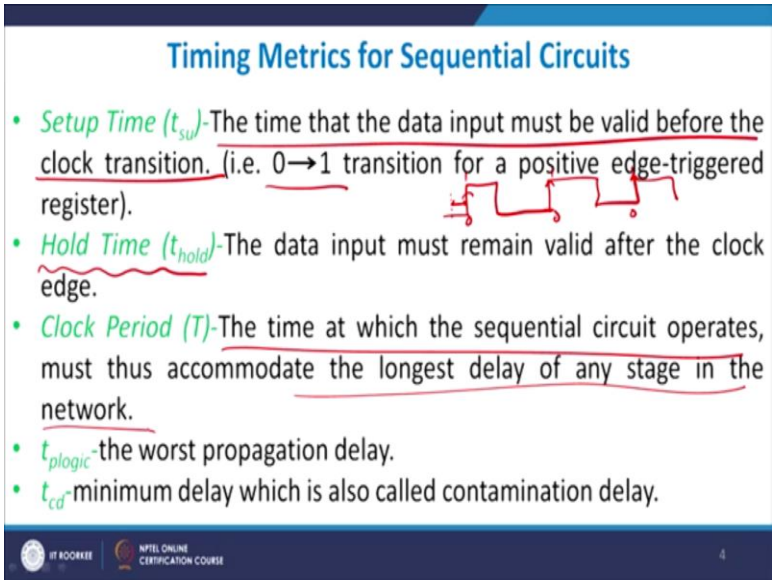
Let me come to the second definition which is known as hold time. Hold time is again very simple, the data input must remain in the same state after the pulse has passed, so setup time is before the pulse has come, arrived and hold time is after the pulse has arrived, you should actually require to have this much.

(Refer Slide Time: 11:09)



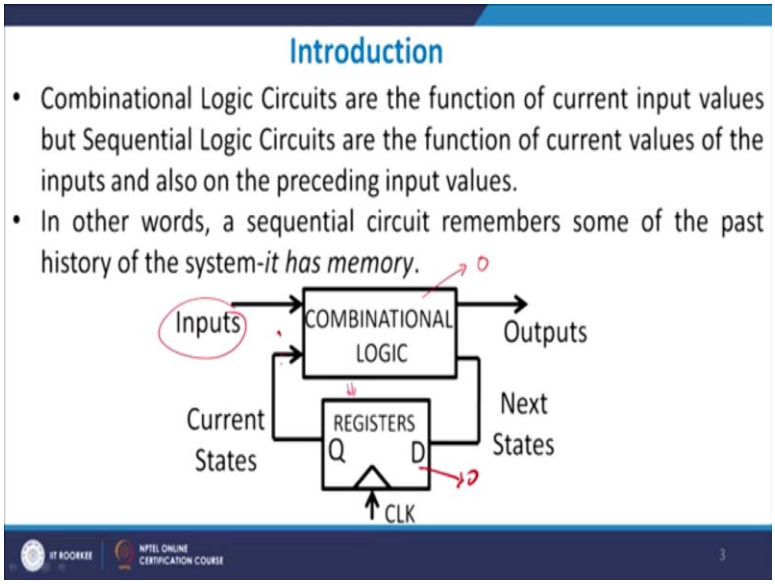

So what is a hold time? If you look diagrammatically, it looks like, so this is the rising edge of the clock and then the data should hold at least till this much point, right. It goes high here and then holds up to this much point, this is known as $t_{hold}$ time. Am I correct? So suppose I am legible to you, in the sense that hold time is the minimum time after the passage of the clock when you are able to have the data stable, after the passage of the clock. And the second time is before the passage of the time and that is what known as hold time.

(Refer Slide Time: 11:45)




Of course, the time at which the sequential circuit operates must thus accommodate the longest delay of any stage in the network. I hope you can understand this point, from the previous diagram you understand.

(Refer Slide Time: 12:05)

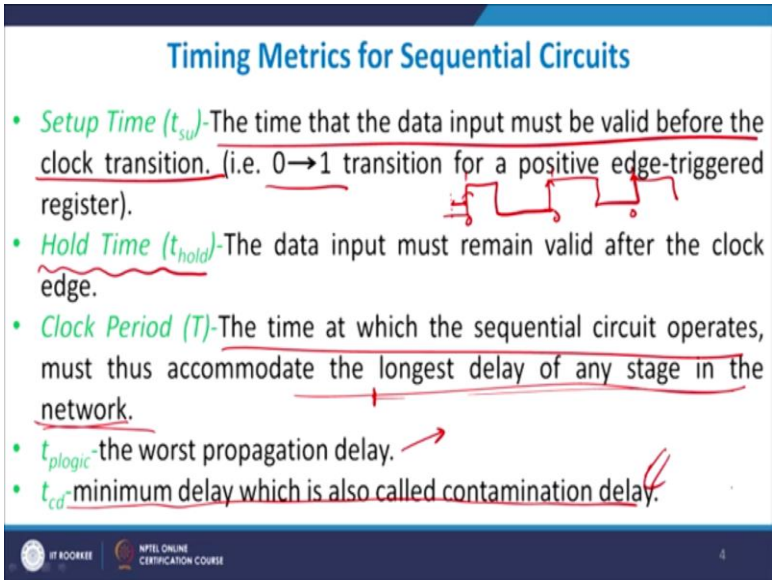



See, if you look here in this slide, you will see that this register has to be minimum delay. Suppose this register would have 0 delay, then what will happen? That, you send an input at this point, at this point you send an input and exactly that say, assuming that this has got 0 as the delay, this has also got 0 as delay, that instantaneously as you insert input here, you will get an output at this particular point.

But understand my basic definition was that the combinational logical box should be able to evaluate the current inputs and the inputs coming from the previous cycle, but, in this case you will find it becoming the same cycle assuming this to be 0 not true therefore. And therefore we define the clock period as the minimum, longest delay, so it should accommodate the longest delay of any stage in the network, this of course you can understand.

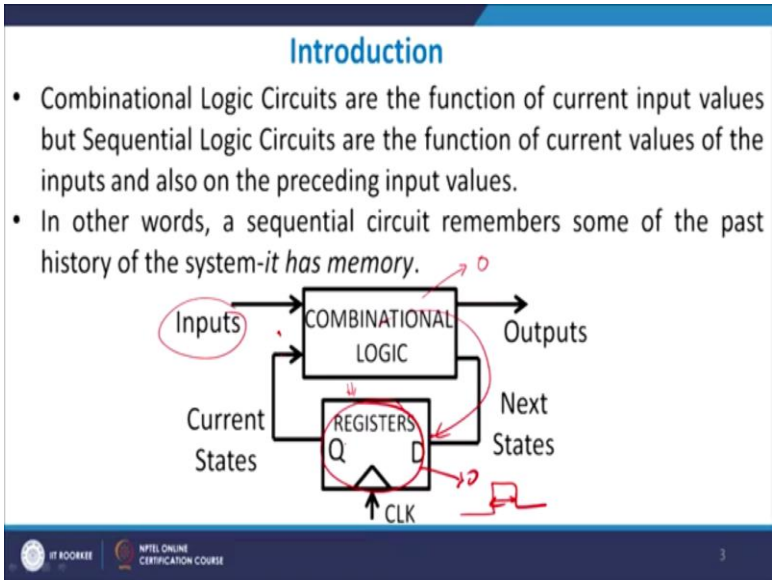
So see, you are driving this clock right, you are driving this clock, this clock is driving this register. So the clock width should be at least that much that it should be able to hold the charge or the data till the whole output is not stabilized, right and that is the basic combinational logical block design is there.

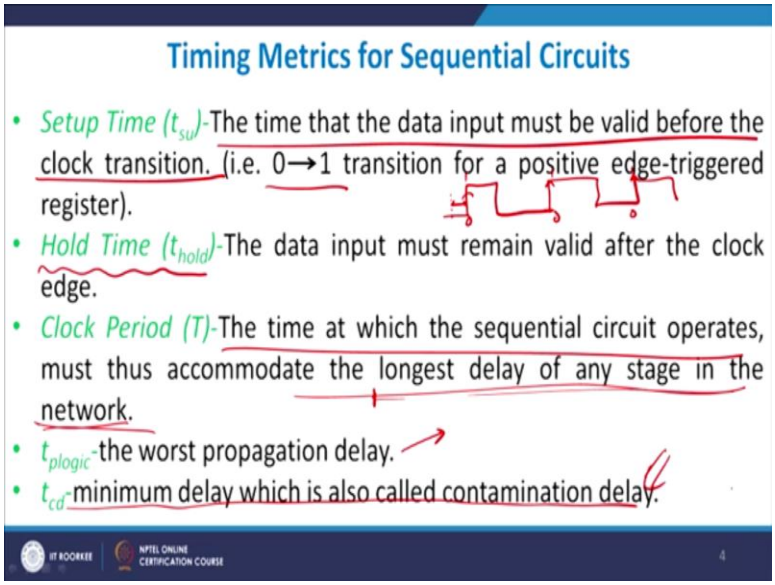
(Refer Slide Time: 13:25)




The worst delay of the logic block is basically referred to as $t_{plogic}$ and the minimum delay is also referred to as $t_{cd}$ and also known as contamination delay. So we have four types of delay: set up time, hold time, we also have clock period which takes care of the generic clock. We have $t_{plogic}$ which is basically the worst case propagation delay across the network, combinational logical block, and $t_{cd}$ is the minimum delay which is also known as contamination delay. $t_{cd}$ delay is also referred to as contamination.

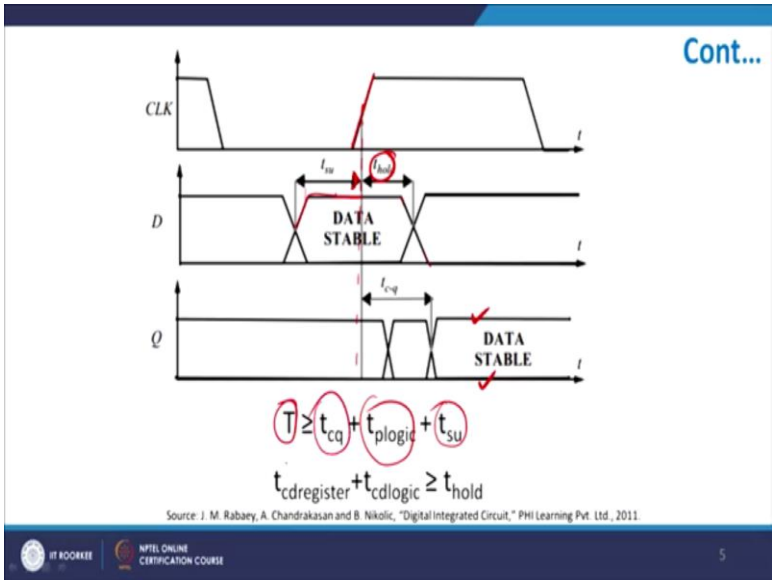
(Refer Slide Time: 13:58)




C to D is basically referred to, if you back the previous slide, C to D is basically this to this delay, right, overall delay. This has to be actually very small or it should be very small. Contamination delay should be 1. So if you want the sequential logic to move fast, you need to reduce these delays drastically.

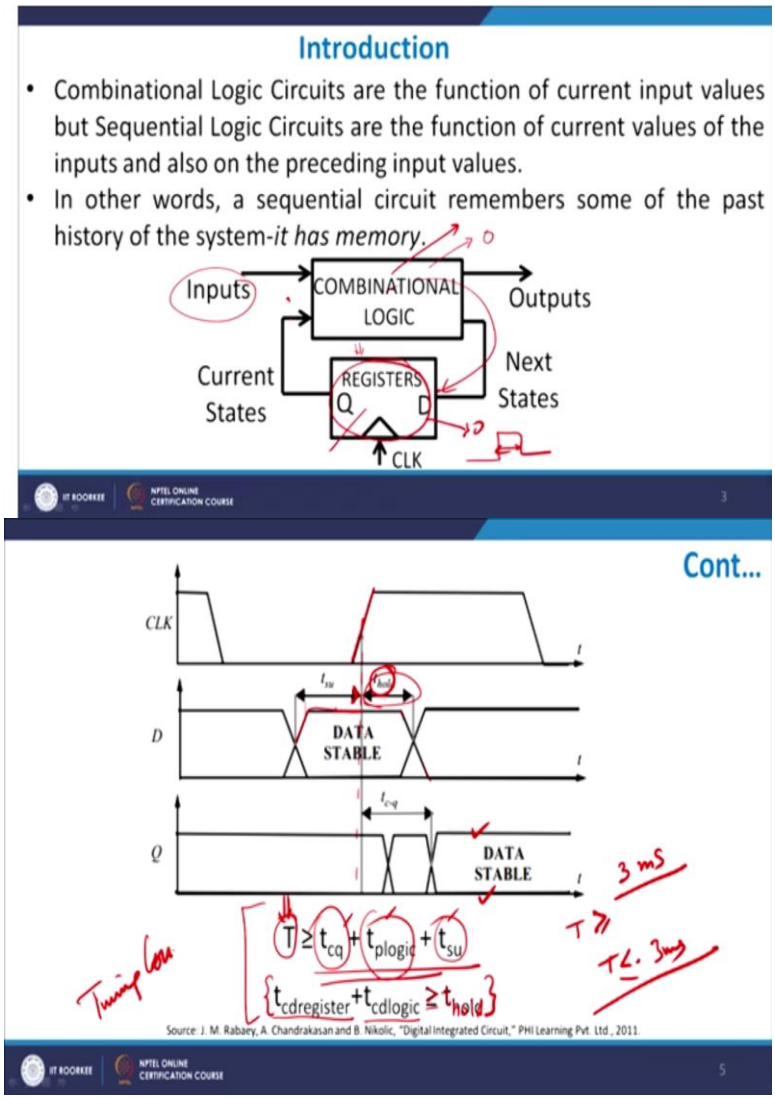
(Refer Slide Time: 14:14)




The same which I discussed with you just now has been shown diagrammatically here. So you see setup time is data, so the data is stable here. So this is the rising edge of the clock, let us suppose. This is the clock, right, so clock is rising edge assuming that it has got finite rise and fall time. So around 50 percent if I take, then we define this to be as a setup time that the time taken for the data to remain stable and we define this to be as $t_{hold}$ because that is the minimum time at which the data has to hold for it can be properly evaluated. What is $t_{cq}$? $t_{cq}$, as I have discussed with you is basically the logic block delay and therefore you see the output has actually fallen down here and you are not able to hold it and that is basically my $t_{cq}$.

After this the data has again held to be stable, why? Because no combinational block, nothing is happening, the data is remaining stable. So we say that the capital t which is the clock of the whole system should be at least greater than $t_{setup}$, plus $t_p$ logic, plus $t_{cq}$, right and why is it important, what is $t_{cq}$? What is $t_p$ logic?

(Refer Slide Time: 15:24)

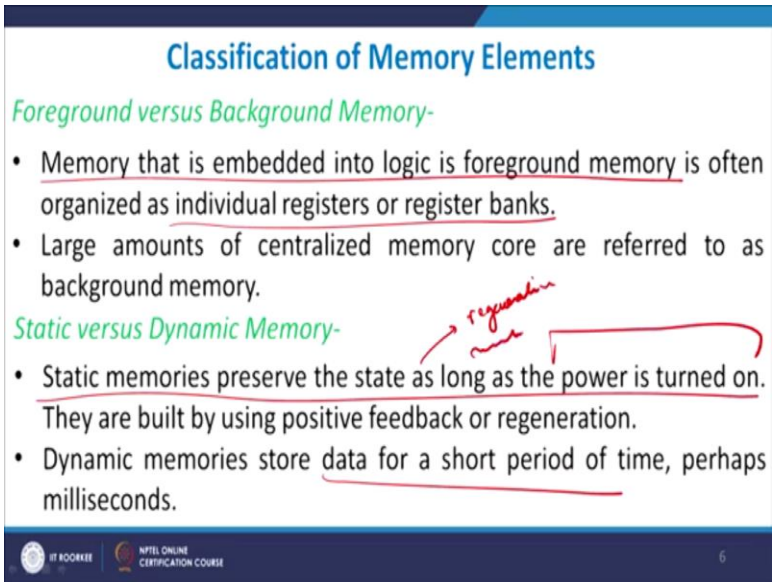



The combinational logical block here, this delay plus this delay, right plus the setup time here is defined as my combinational logic as the total, as means….it is basically your stability criteria and the minimum time should be actually the sum of all these 3. That means if you say, say this is 1, 1, 1, so let us suppose 3 millisecond. If t is greater than equal to 3 millisecond, no issues, if t is less than or equal to 3 millisecond then there is a problem.

The problem is that means you are not allowing at least one function to take care of your sequential logic and therefore it will not remain as a sequential logic, right. That is the one thing, then we define the another set of timing constrains here and that is $t_{cd}$ registered plus $t_{cd}$ logic should be greater than equals to $t_{hold}$ here, what is $t_{hold}$? $t_{hold}$ is this, remember? So your data must hold, till how much time it should hold? The $t_{cd}$ logic, the combinational logical block delay plus the register.

The register itself should be so, should at least hold that much time so that it is greater than the $t_{hold}$. Say it is less than $t_{hold}$ then what will happen? Your data is held stable but then you are already ready to give the next set of data. So, you combinational logical block will not be able to understand which data you are talking about in reality and that is the reason this are the two timing constrains which are very very important in the sequential logical design, timing constrains, right and there are two timing constrains here.

(Refer Slide Time: 17:09)




Let us look at the classification of memory elements, the classifications are that you have a foreground and background memory. The memory that is embedded into the logic is defined as the foreground memory, right and is often individual registers or register bank. So in a register you will have CLBs, so these combinational logical blocks acts as a memory array in reality.

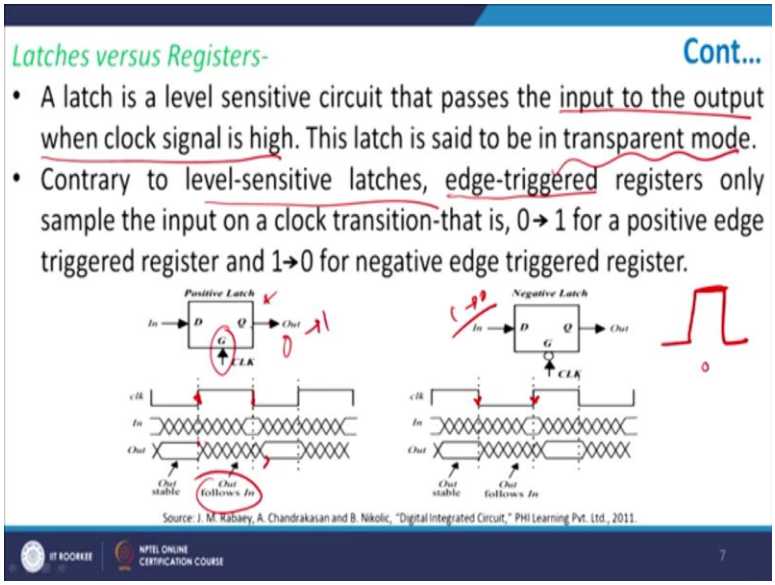
Whereas the background memory is your SRAM, your RAMs which you generally have and these store memory for a longer duration of time. Whereas the memory which is embedded onto the logic, the logic being embedded into the memory in the foreground memory and it is relatively a very low power, right.

Most of the time, whenever we do all sorts of design problems and sequential we use the foreground memory but whenever you want to store a data for a longer duration of time, we use the background memory, right and that is the general rule of thumb which we follow. Now static versus dynamic memory, as I discussed with you static memory are those memories which hold data for as long as the power is on, right, static memory is preserve the

state as long as the power is on and dynamic memories only do for a short period of time perhaps milliseconds. For example, the example might be dynamic RAM, DRAMs, so it is stored in its capacitance small charge but for a few millisecond and then again it will refresh it to let it store for a longer duration of time.

Whereas in a static RAM they preserve as long as the power is on, right and so as long as the power is on, then I can do a positive feedback too. So even if the voltage goes down because of charge getting reduced, I can do a positive feedback network and add charge to it to maintain the level, right. Therefore this is also known as a regenerative network memory. This is also referred to as regenerative memory.

(Refer Slide Time: 19:02)



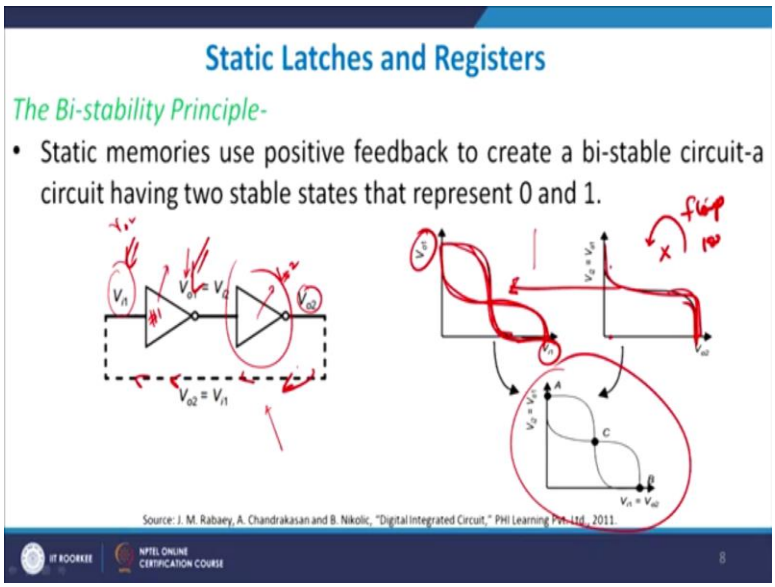

So with this knowledge let me explain to you the difference between latch and register, you must be knowing also, latch is a level sensitive circuit and passes the inputs to the output when the clock signal is high. So this is my latch, so whenever at this state when the it is level, when it is level sensitive and high level it is reached, then only the latch becomes transparent and is able to perform any operation and let the input go to the output with the clock is high.

This latch is transparent latch as I discussed with you. Contrary to level sensitive latches, how I define edge sensitive or edge-triggered registers? With sample input only when during the rising edge of the clock, 0 to 1 positive edge-triggered. When you sample it 1 to 0 then we define it to be as a negative edge-triggered register.

So if you look at this block diagram here, this is basically your positive edge-triggered and on the D and Q, D is the input and Q is the output and this is my clock here. So you see when the clock is going from 0 to 1 right, output goes 1 and it remains stable. When it goes from 1 to 0, the voltage also falls down to follow and therefore out follows in this case by certain delay. Whereas in this case where you have a negative latch, negative latch basically means during the negative 0 to 1 transition.

So this is 0 to 1 transition of the clock, this is 1 to 0 transition of the clock, you will have negative edge-triggered design and that is how it follows. It is almost the same as you look here, only thing is that during the negative edge of the trigger you will have following edges with you and this is positive edges rising here and you will have the output being latched. So when the positive goes high the data is being replicated here, when it goes low the data is left and that is what is a difference between a latch and a register.

(Refer Slide Time: 21:00)




Let me come to a concept of bi-stability principle. Well, what we do is that we have got two static invertors in series to each other as you can see on the left hand side. So if I define $V_{i1}$ as the input voltage to static inverter, number 1 and $V_{i2}$ to number 2 then what will be the output? It will be $V_{O1}$ will be the output of the first invertor which will be exactly equal to $V_{i2}$ for the second invertor and output will be $V_{O2}$. So $V_{O2}$ now goes inside if it is feedback look and now $V_{O2}$ and therefore if it is $V_{O2}$ here which is just, what is $V_{O2}$? $V_{O2}$ is just the output of this one, you will feedback into the first invertor. I will show to you that such a pair will only give you two sets of stable points.

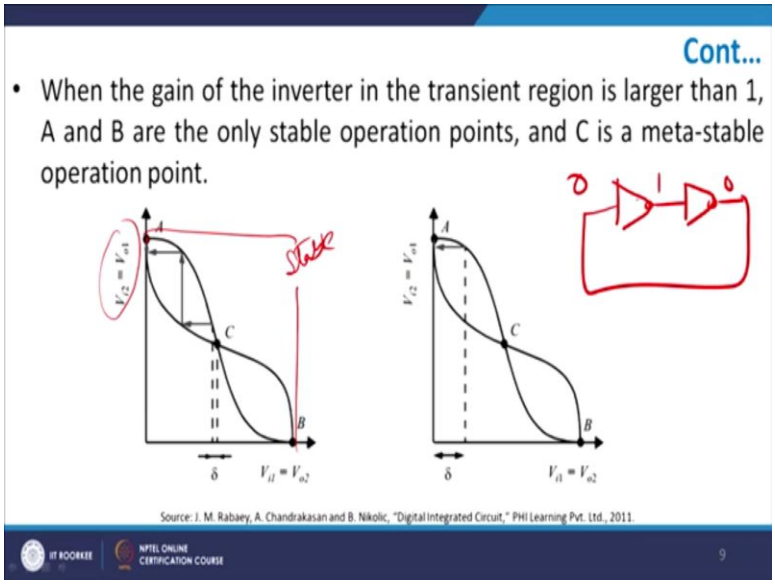
So if you look carefully you forget about this now and you draw the VTC for the first one, you get $V_{out}$ versus $V_{in}$ to be this. So this is $V_{i1}$ here versus $V_{O1}$ which is this one. Therefore, since it is a static invertor it will behave exactly like the VTC of the static invertor which we have already studied when input is low and output is high, when input is high then output is low, fine, we have already done this point.

Now if we go for the second one therefore, you understand the second one's input is the first one's output. So if I want to draw the VTC, I just have to make x to y plane and y to x plane and then the job is done. Once you do that, I get this into consideration right, I get $V_{O2}$ here and it is $V_{I2}$. What is $V_{I2}$? $V_{I2}$ is the input to the second transistor right and when it is low, the output is high and when the input is high the output is low, it is very straight forward and simple.

Now what you do is you just flip this flip this by 180 degree right, flip it by 180 degree and place each axis over other, then you will get a figure something like this. You are getting my point? So this is your one, you have another one which is something just like this, so just you just place one over the other. You will get something like this. So this is I am just shifting it on this side right? And this is what you will get.

This is a bi-stable circuit for a memory, I will explain to you where the bi-stability comes into picture but I suppose now you appreciate the point that based on the two stable states and there is one unstable state, using positive feedback I can actually achieve. So this is the feedback loop which I am drawing to, giving, again I will be able to achieve certain important things.

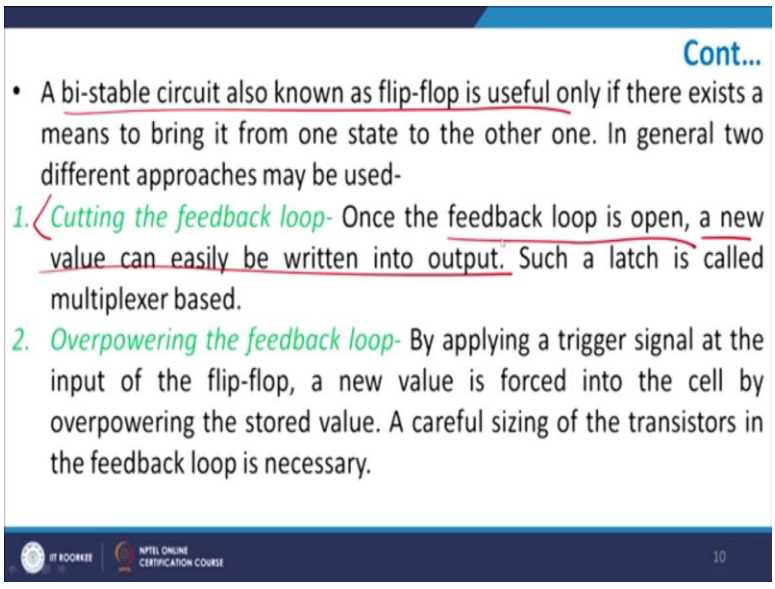
(Refer Slide Time: 23:52)




Now if you look at this VTC, then you see this point A and point B are the most stable points, these are the most stable points and the reason being for example, look at the point B, point B is basically Vinput equals to Voutput2 because you feedback directly from the system and this corresponds to $V_{i2}$ equals to $V_{O1}$.

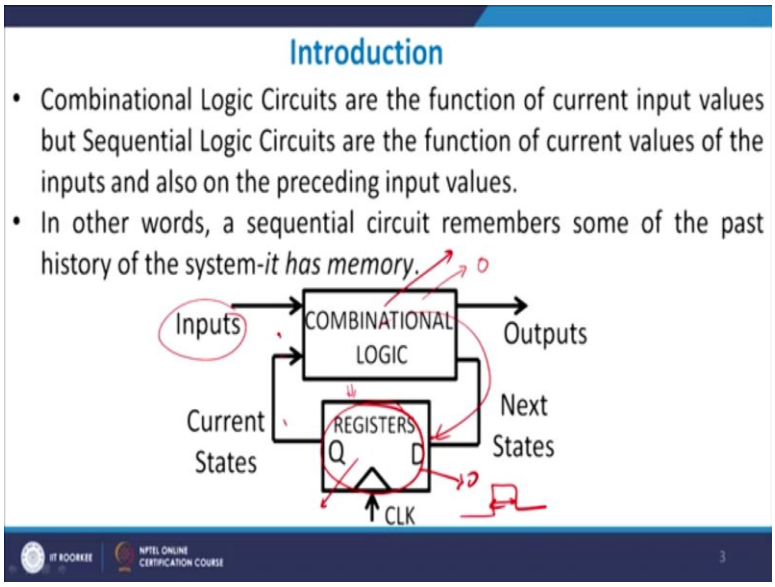
So now if you just place it in A or B, it will always remain stable because you can see very clearly that if you have something like this then if it is 0 1 0, then 0 is feedback into the input side and this is the most stable configuration of two CMOS inverter connected back to back and that is what I am trying to show you in front of you.

(Refer Slide Time: 24:39)



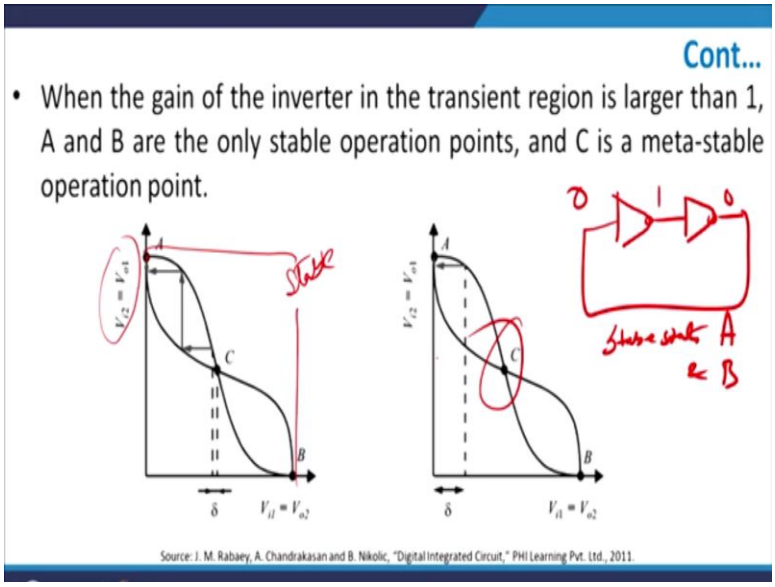
Now with this knowledge a bi-stable circuit can also be known as flip-flop, why flip-flop? Because they are two stable states and by doing some manipulation I can move the stable state from A to B and I can also bring it back from B to A right. I will show to you how this is done. There are two basic approaches by which you can achieve a bi-stable circuitry, obviously you should cut your feedback loop and therefore now the feedback loop is open and now you can write very well onto the output.

(Refer Slide Time: 25:27)



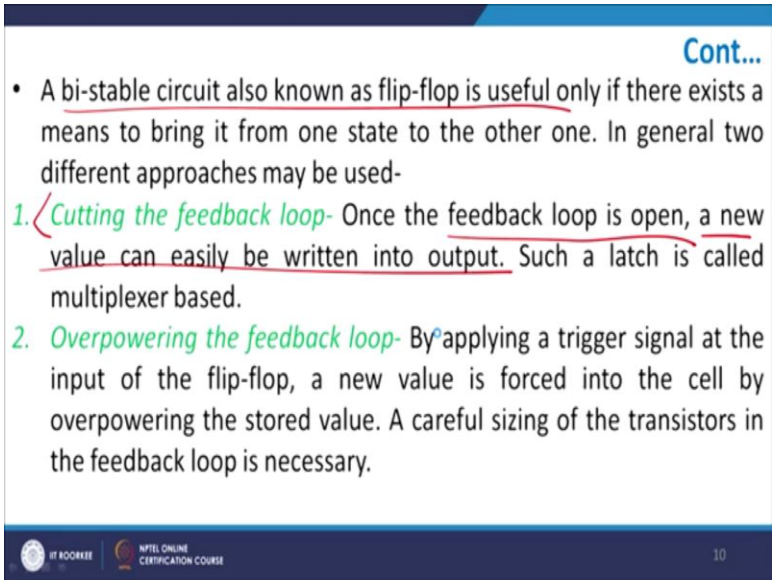

Which means that it is something like, that means, sorry, if you look in this case then if you, for example if you open the feedback loop here, I will just show you from the figure which is there, right. So if you open the feedback loop which is this one, then there is no problem, then it becomes a combinational logical block and as I discussed with you, when you are discussing having a combinational logical block, there is no concept of delay as such and that is the reason you will get a stable state here.

(Refer Slide Time: 25:40)




You will have a stable state at A, so stable states are what? Stable states are A and B, any other state is sort of an unstable state, the most unstable state is basically the C where you have got two VTCs crossing each other, why crossing? Because input of two is exactly goes to the output of one. So I am just overwriting each one over other and I get this into consideration, A and B are the most….

(Refer Slide Time: 26:13)

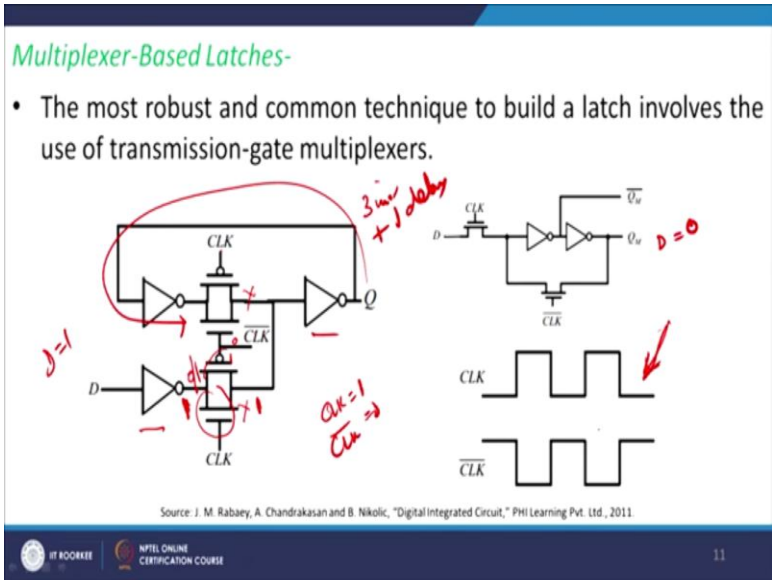



So how do you do it? You when you keep the feedback loop open, no value can be easily written and such a latch is called multiplexer latch, we will discuss this point. We can also overpower the feedback loop by applying a trigger signal at the input of the flip-flop and a new value will be forced on the cell by overpowering the stored value. So if your stored value

is there but I am able to overpower that value by an external source voltage then we can say that your flip-flop or your this thing, the sequential logic is prone to external voltages.

Now you see if you want to make it, another way is that you want to overwrite over the cell within the combinational logical block, then you have to ensure that your feedback loop is relatively very weak. If it is weak it will not be able to push around its value onto the basic combinational logical block. So if you want a combinational logical block to sustain, this is a particularly good idea people use across the research community.

(Refer Slide Time: 27:21)



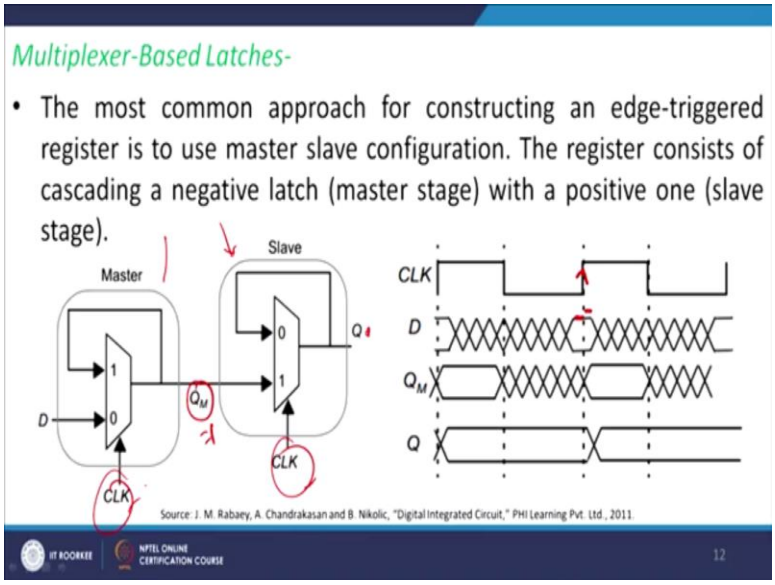

Let me come to a multiplexer based latches, the most popular one is this one and as you can see here, this is basically a Q and so when D equals to 1, suppose invertor this is 0, now let us suppose clock equals to 1 and therefore clock bar equals to 0. So this is 0 and this is 1. So when as I discussed with you in TG when this is 1 this is cut-off, this is cut-off and nothing happens. Now what happens is that suppose D equals to 0 and therefore output of this equals to 1, this makes slightly difficult easier for us that this will therefore be, if it is 1 then this clock will be switched on and similarly this clock bar so if it is 1 the output will be 0, sorry output will be, so if it is 1 if clock equals to high, I we would expect to see a 1 at this particular point.

But please understand in this case the total delay between the D and Q is the sum of this inverter, this inverter and this inverter, why? Once you set the data it passes through this, this but this is not making a difference and then this, so it goes via this, this and then backs to this point. So, I have got three inverter delays, plus 1 this delay because of the clock because of

the two transmission gates or maybe one transmission gate. Now this gives you an idea that having transmission gate reduces your overall on current but the variation of its resistance with respect to input voltage is almost constant and we know that, that is the reason we use a TG design in most of the cases.

This is the non-overlapping clock I was talking about because this is clock and this is clock bar and you will be able to sustain a non-overlapping clock for a longer duration time once you are able to design these muxes.

(Refer Slide Time: 29:16)



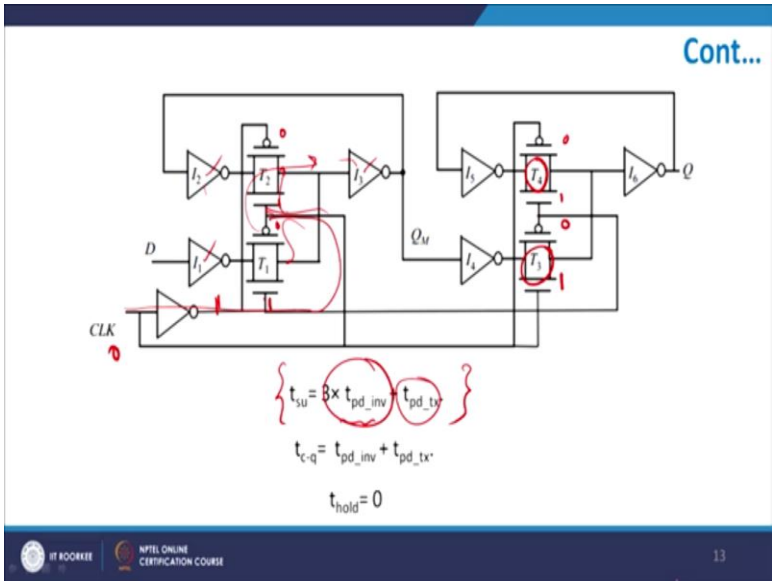

This is another mux based latch, the most common consideration using an edge-triggered, so I have a master here and I have a slave here right .So when the master comes Q1 is equals to 1, it gives 1 here. Whenever the clock is high the slave accepts a data from Q, compares with 0 and 1 and then processing starts place.

So you see this is my clock here. When the clock goes, so this is basically positive high which means that if the clock goes high let us suppose, the data has to be held stable before the rising edge of the clock and it should be minimum held stable till this much point after the passing edge of the clock. And therefore we should be very careful about how we are clocking the whole thing. The second thing is, as I have discussed with, you are applying clock at this stage and at this stage.

Please ensure that these clocks if you are doing it in a lab or somewhere please ensure that this clock and this clock are exactly of the same dimensions giving you exactly same peak to

peak voltage, otherwise you will not be able to achieve such a good master slave configuration of laches.

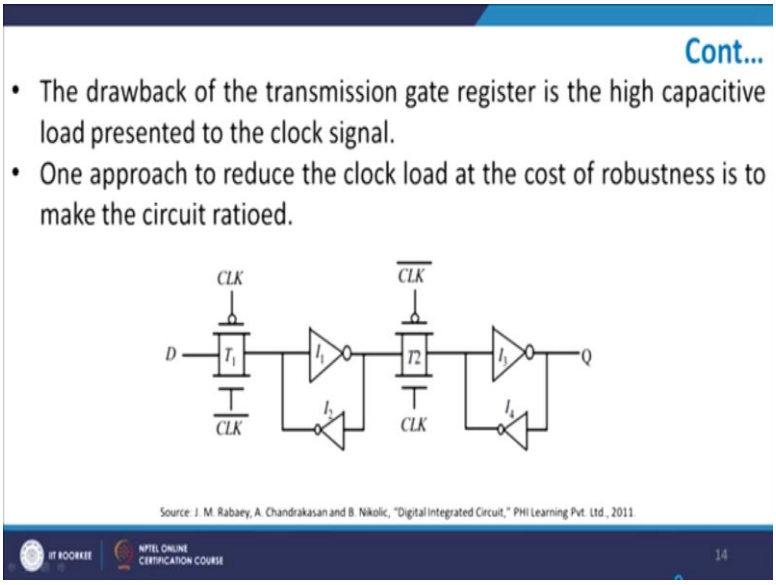
(Refer Slide Time: 30:30)




This is another example which you see and that if you look very carefully here when clock goes say 0, this is 1, this is 1 means this is 1. When this is 1 means this is 1, this is 0 and therefore this is 0. So what happens is that if this is 1 this goes to suppose here right and let me just try the same thing here 1 0 1 0.

So whenever the clock one goes and tries to overwrite here, since it is fed to the PMOS terminal, $T_3$ does not react at all but $T_4$ shows a change drastically. So the overall delay will be actually governed by the $T_4$ delay in a combinational logical block. The total setup time therefore is three times inverter delay, why three times? Because there is a 1, 2 and 3 because the data has to travel from here, it has to go here and then either it should come to this point or come to this point and then finally to the output side. And therefore your total delay is basically 3 pd (propagation delay) inverse, plus 1 transmission gate, so either of the three transmission gates will be required finding out the total delay. We define $t_{cq}$ as equals to $t_{pd}$ inverse plus $t_{pd}$ transmission, if you add those two together I get $t_{cq}$ as one of the important points. Hold time is 0 in this case for all practical purposes.

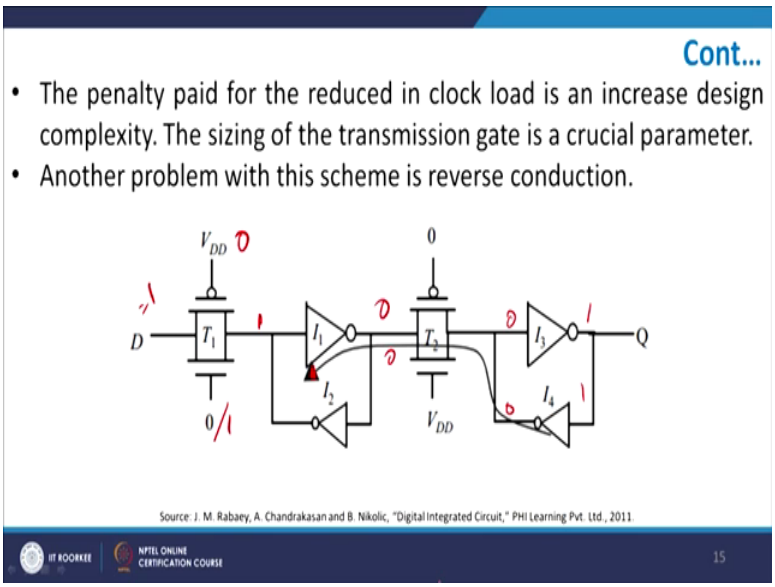
(Refer Slide Time: 31:50)




Now the drawback of transmission gate register is the high capacitive load, we all know this point, because you remember this transmission gate obviously terminates with both PMOS and NMOS on the same substrate and therefore you will have excessive loading of the substrate onto the device itself and as a result you will have large amount of capacitive delay.

Another methodology which people have approached is to make it ratioed logic, remember till now we were discussing that these are all non-ratioed logic and we will not be requiring any ratioed logic in this case but we do require ratioed logic because of the fact that we want to reduce the clock load.

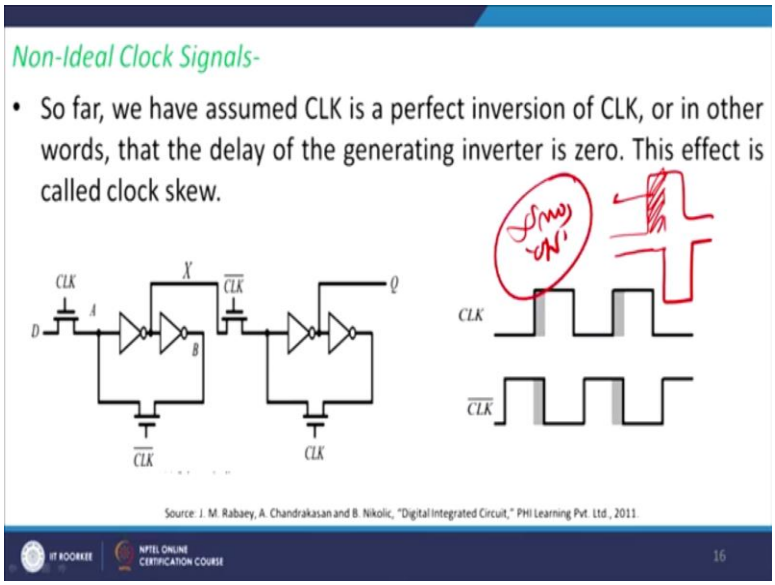
(Refer Slide Time: 32:35)



However, if you want to reduce the clock load, the penalty you want to pay is basically design complexity right and therefore the size of the TG is quite important because I will give you an example. If the TG size is so small that the resistive elements goes very high on chip then you will end up having a very high delay in the output side. Similarly, if these are shorted, you will never have large amount of current to drive your inputs and therefore proper sizing of transmission gate is a crucial parameter.

Now if you see D equals to 1 and $V_{DD}$ equals to 0, so this is 0, this becomes 1 let us suppose. It transmits 1 here, goes to 0 here and then comes to 0 and this is 1, this is 1 and this is 0. And so we get back 0. So I gave 0 and I got back as 0 but the only thing is that now I have a transmission your gate based design which takes care of it. The reverse conduction we need to stop it, so make the area large and we can stop the reverse conduction.

(Refer Slide Time: 33:32)

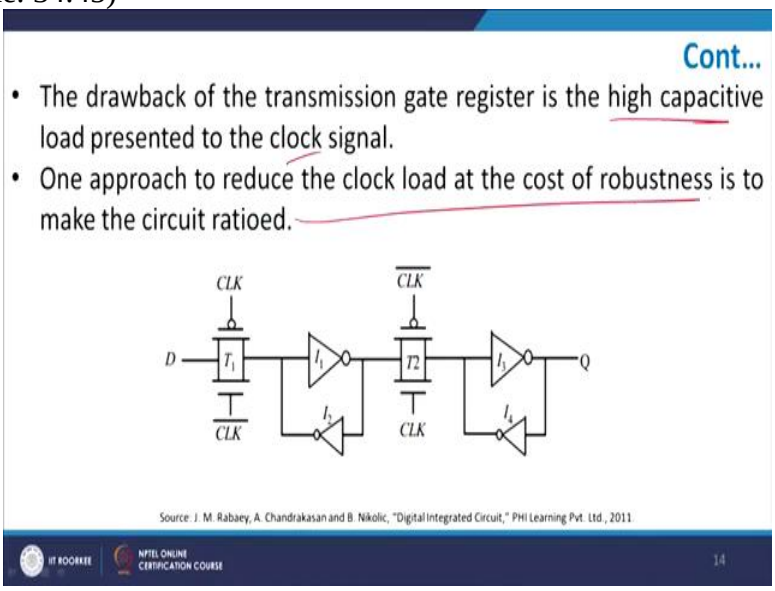



This is how you generate a non-ideal clock signal right. However there will be some delay between two clocks right and that is known as Q delay or clocks skew and this is the diagram of the schematic of the non-overlap clock generation, you can use this as a plug and play sort of option.

What is the problem? This option means this delay means if there if let us suppose your clock and clock bar are not having perfect overlap, then what will happen is then when the clock is high, your clock bar is also clock bar which was supposed to be 180 degree out of phase will be shifted by this much amount and it will go like this.

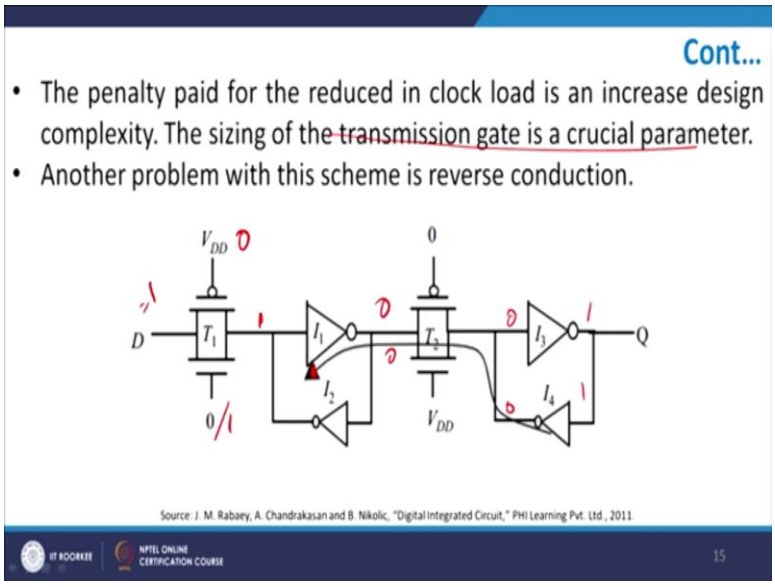
So till this time your NMOS will be still on and you will be actually sampling data even. You do not want the sample to take place but actually it is happening because of the overlap, this is known as a 1 1 overlap in a combinational logical block, and therefore it is also known as a reverse transmission.

(Refer Slide Time: 34:45)



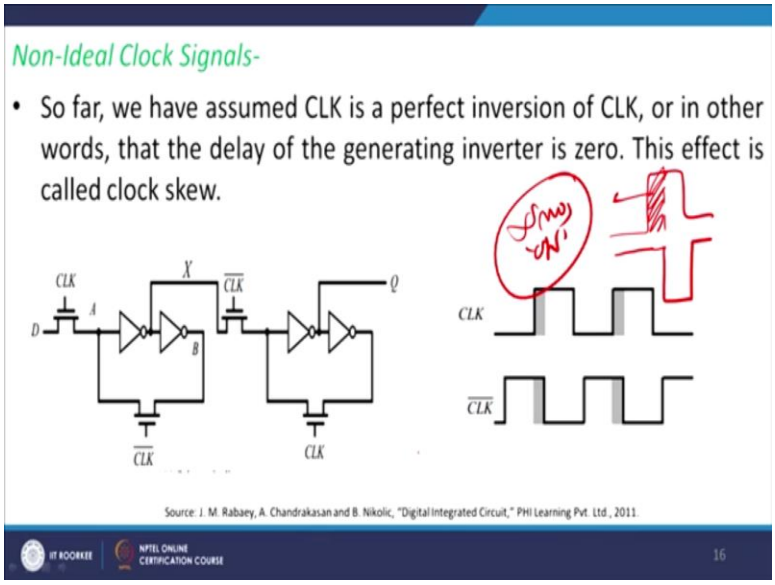

The drawback of, as I discussed with you is the high capacitive load and robustness of the circuit is an important issue.

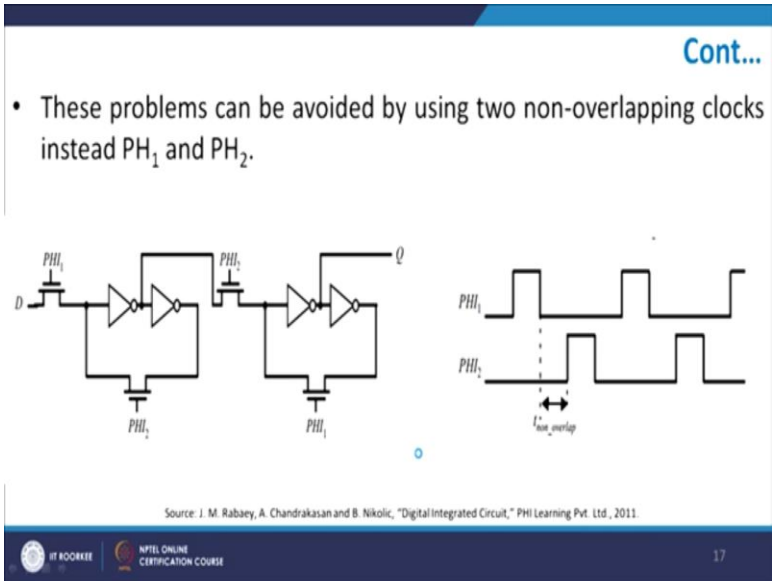
(Refer Slide Time: 34:50)




We have also looked to the penalty of paying, if you want reduce the clock gate and the sizing of the transmission gate is quite important. Larger the size lower will be the loading effect and that is generally taken care of across the board.

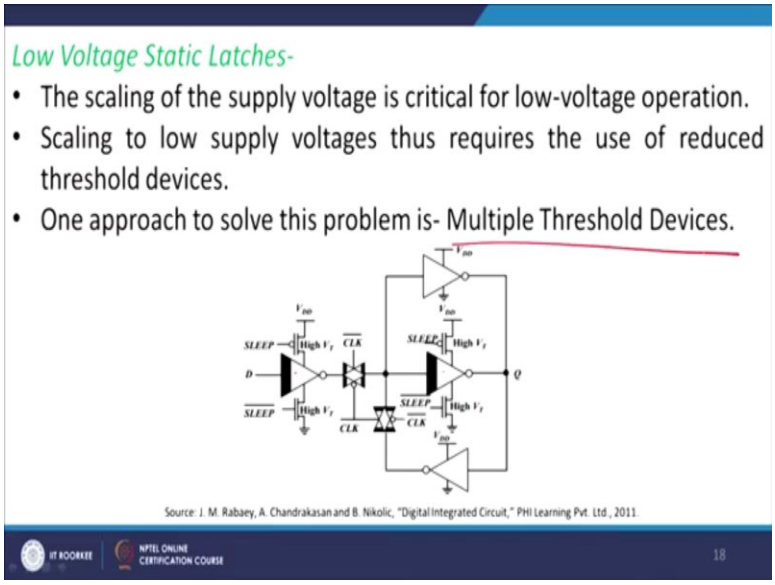
(Refer Slide Time: 35:04)



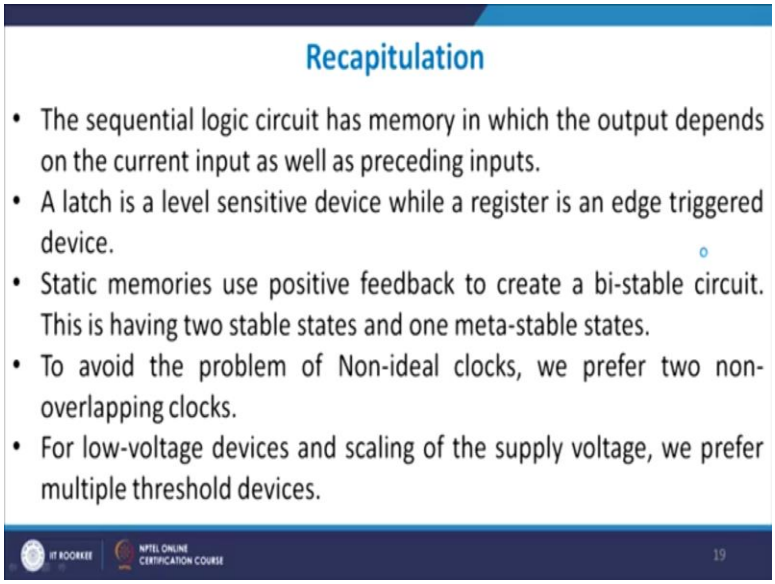

So as I discussed with you non-ideal signal will be those signals which the clock is a non-perfecting clock and this can be done by using two overlapping clocks of just pure non-overlap of each other and we should be able to generate the output file.

(Refer Slide Time: 35:22)




Okay, then we come to the last part of our this module and that is basically low voltage static latches. This is critical for low voltage operations. What we do is that reduced threshold voltages are used MOS devices in the critical path so that even when the dynamic participation is high that can be reduced drastically by applying this sleep mode transistors with a reduced threshold voltage of the device. This is also referred to the multiple threshold device we can use. This is a quite interesting one.

(Refer Slide Time: 35:54)




So let me recapitulate what we have done in this small module, we looked into sequential logic design, how it's made, what are the diagrams we have already seen it. A latch is a level

sensitive device whereas edge-triggered, resistor is an edge-triggered device. Static memory used as a feedback can create bi-stable element, DRAM is one of the examples.

To avoid the problem of non-ideal clocks, we refer two non-planner overlapping clocks, this is what I was trying to say till yesterday. For low voltage devices, the scaling of supply voltage we prefer multiple threshold voltage which means that when we go for low voltage domain, we require multiple threshold devices.

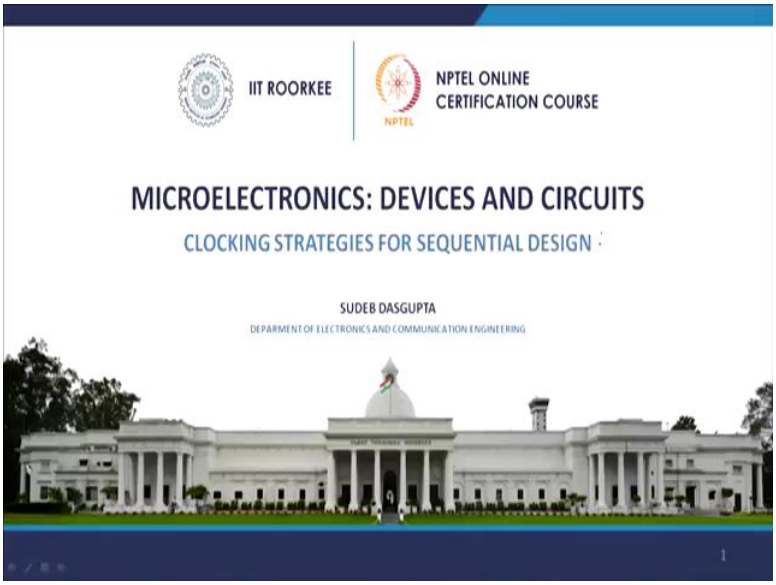
In the critical path try to keep a low voltage devices so that they are switched on and off very easily and the signal can be easily passed from primary input to primary output. So this is what is all about static memory and dynamic memory, we have taken care of the basic ideas of sequential logic. In the next module we will start off again from the sequential logic flow only. Okay. Thank you for your patient hearing.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−58**
**Clocking Strategies for Sequential Design-I**

Hello everybody and welcome once again to the next addition of NPTEL online certification course on Microelectronics Devices to Circuits. In a previous module we have looked into the concept of combinational sequential logic and we have also understood that is equation logic is basically logic where memory is inbuilt into it and we also saw that unlike combinational, in sequential logic timing is a very important issue.

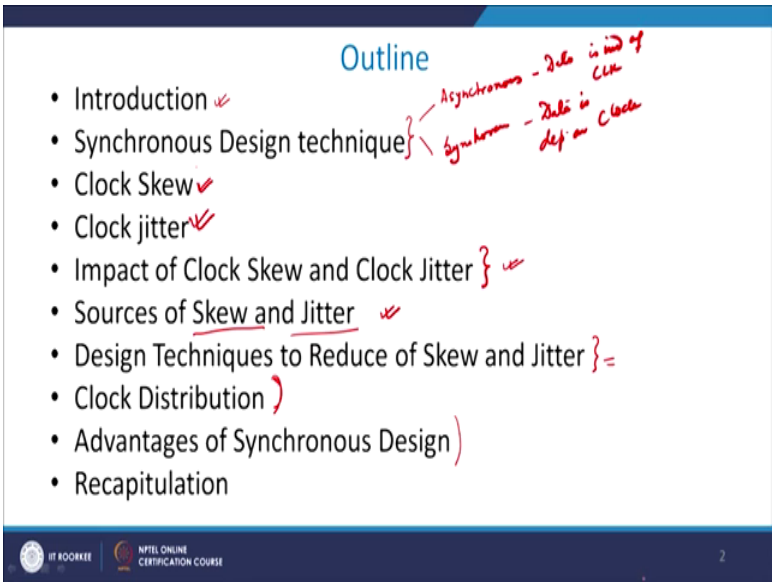
So we should be able to therefore perceive that at a particular rising edge of the clock if my data has to be inserted into the system then how will the sequential logical block react if the data would have inserted much earlier than the rising edge or much later than the rising edge and what would be the effect on the overall performance of the sequential logic. Keeping in mind this module is basically, if you look is termed as clocking strategies for sequential design.

(Refer Slide Time: 1:33)




So this is a topic of this module and outline of this module is as follows.

(Refer Slide Time: 1:40)



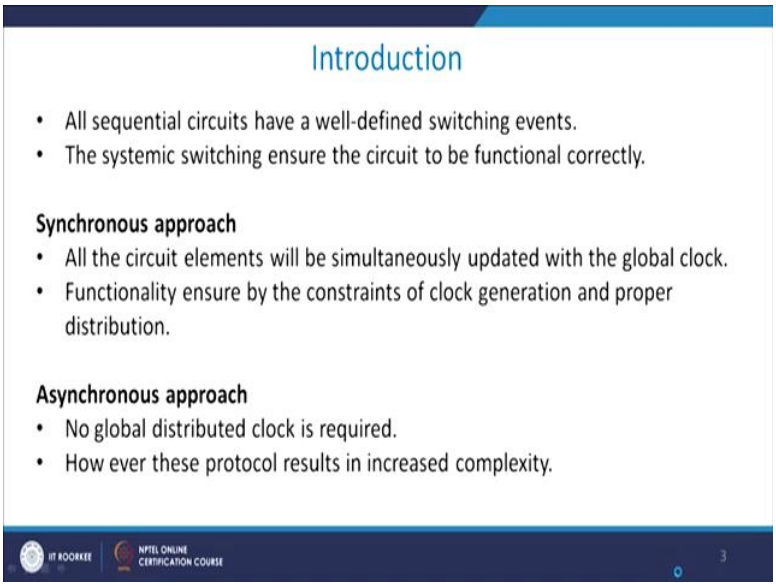

We will first introduce to you the basic concept of timing and then we will look into the synchronous design techniques. As they have discussed earlier there are 2 types of design, one is A synchronous, right? And we have another is synchronous. What is A synchronous? Well, Asynchronous is that when the data is independent of clock, right? External clock and when data is dependent on clock we define that to be as Asynchronous design.

So for Asynchronous design timing issues are not important but for synchronous design timing issues with respect to clocks are very important. In these respects we look into 2 major problem areas of clock and one is the clocks skew, another known is clock jitter and we will see how this influence the overall frequency of operation of a sequential logical block. And that is what is the fifth point here is all about that its impacts on the overall delay of the system.

We will also look into the fact from where skew and jitter do come into picture because if you know the sources from where it is coming we should be able to therefore reduce the skew and jitter here, right? So we would look into what is the meaning of skew and jitter. The origin of skew and jitter and then ways and means of reducing the values of skew and jitter. Then we will have a look at the clock distribution problem and then advantage of synchronous design as compared to Asynchronous design then finally we will recapitulate the whole thing.

So as far as this module is concerned, it gives you typically a brief idea about the locking strategy.

(Refer Slide Time: 3:26)



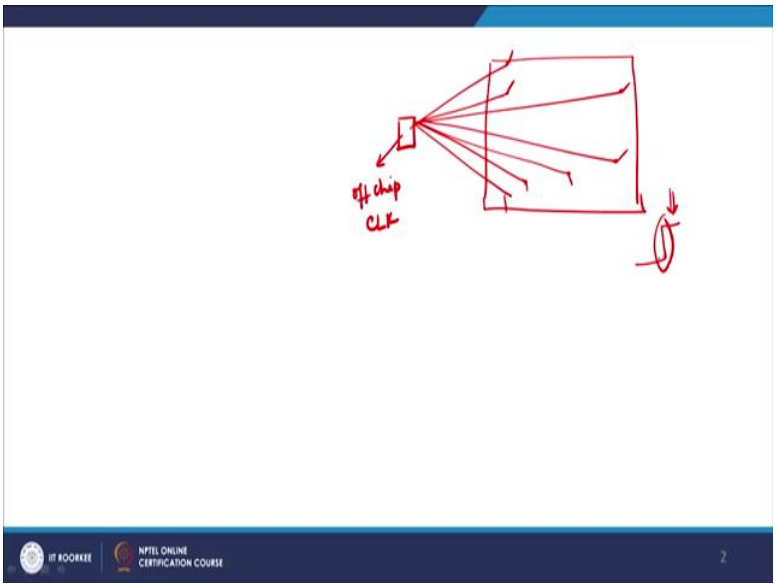

Now let me come to the basic approach and we will go step-by-step explaining to you each one of them. So as I discussed with you all your sequential circuits, right? Have well-defined switching events. As I have discussed with you which means that these switching systems ensure that the functionality of the circuit is not compromised. As I discussed with you earlier also that let us suppose it is a positive edge triggered register maybe which is accepting a data and then it is giving to a combinational logical block where some computation is being carried out maybe a logical operation and after that it is fed into another register where it is storing for a few amount and then giving it to the output side.

Now the way the data will travel from input to output through these 2 registers as well as combinational logical block, we will determine the speed of the clock. Maximum speed of the clock. Now if for example I have discussed with you that if the input is much much earlier than the setup time of the register then that data will not be at all inserted into the system and you will get a miss.

Similarly if the data does not hold its value beyond a particular limit after the clock is passed which is known as $t_{hold}$ you will not be also able to evaluate the system. So that is the reason why we define that sequential logical are well-defined switching events. Based on switching events. Now there are 2 approaches as I discussed with you in the starting slide we have got synchronous approach and we have got Asynchronous approach.

In this case as I told you all the circuit elements will be simultaneously updated with the global clock which means that there will be a global clock which will be driving all the blocks digital block within the system and this global clock assuming that the global clock is being run in such a manner that the time at which it goes from 0 to 1, rising edge of the clock is exactly the same in all the blocks then we will accept that and all the blocks the data will be inserted at the rising edge.

(Refer Slide Time: 5:41)

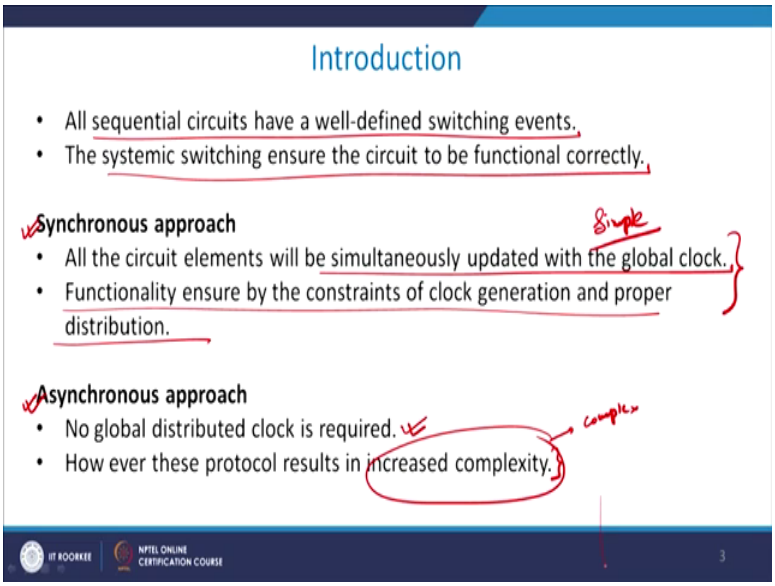



I will give you an idea let us suppose what we are assuming, we will see that this assumption is wrong later on but for understanding purposes, let us suppose you are driving and off chip clock. So this is my of chip clock here, right? And it is driving through this interconnects to the various parts of the sequential block here, right? So these 2 are almost equal in diamonds.

This is the farthest, these 2 are farthest and this is somewhere in the middle which you see. So somewhere here also you can get. Now I am assuming that when there is a rising edge of the clock this event is replicated at this point, does point, this, this, this, this and this, which means that irrespective of the routing interconnect between the clock and the area where you are giving the clock irrespective of the distance mode.

I assume that the clocks always at 0 to 1 transition at t equals to 0 let us suppose, right? And that is a major assumption we will break off this later on.

(Refer Slide Time: 6:44)

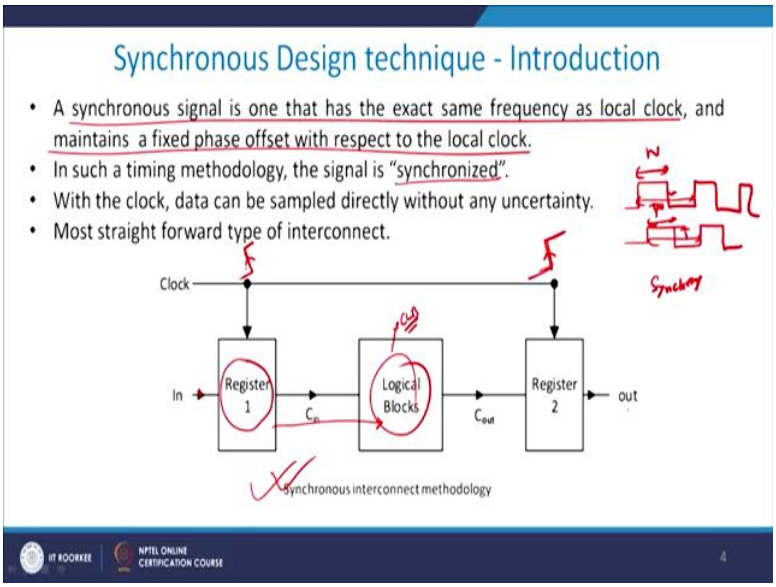



So that is one important clock that there will be a global clock. And therefore the functionality will ensure that by the constraints of clock generation and proper distribution. As I discussed with you therefore that clock going into the system and generating an output logic will depend therefore on the proper synchronization of the clock with respect to individual blocks of the sequential data. This is for the synchronous approach.

You can also have an Asynchronous approach which is there. Asynchronous approach means that there is no global distribution of the clock. So there is no clock is required, however, so it depends upon what? It depends only upon the rate of flow of data directly coming into the system, however in these cases there is an increase complexity which is with us.

The complexity is typically very large, they are complex circuits, whereas these are simple circuits where you need just to rout the signal at a particular point. So this is the basic introduction of a clock.

(Refer Slide Time: 7:47)




Now let me therefore define certain definitions I will explain to you. Asynchronous signal is one that has exactly the same frequency as the local clock and maintains a fixed phase offset with respect to the local clock. Which means that if I have a local clock which is coming like this, this is a local clock than if I were clock which is something like this, so this is T, this is also T.

Say this is width this is width, this is T of the main clock this is also T of the main clock but it suffers from a basic phase difference here then we define this to be as Asynchronous signal. Signal is synchronous. The signal is said to be synchronized. As I discussed with you, why synchronized? For example if you look back into this design here in front of you. There I have a register here, clock here.

I am assuming that the clock transition, these are all edge triggered register suppose. So whenever my edge goes high, it also goes high. And this register is able to insert one data from the data train here and transfer it to the logical block. So this is your combinational logical block which does all sorts of computation, a logical block and gives you $C_{out}$ goes to the register 2 and that is output.

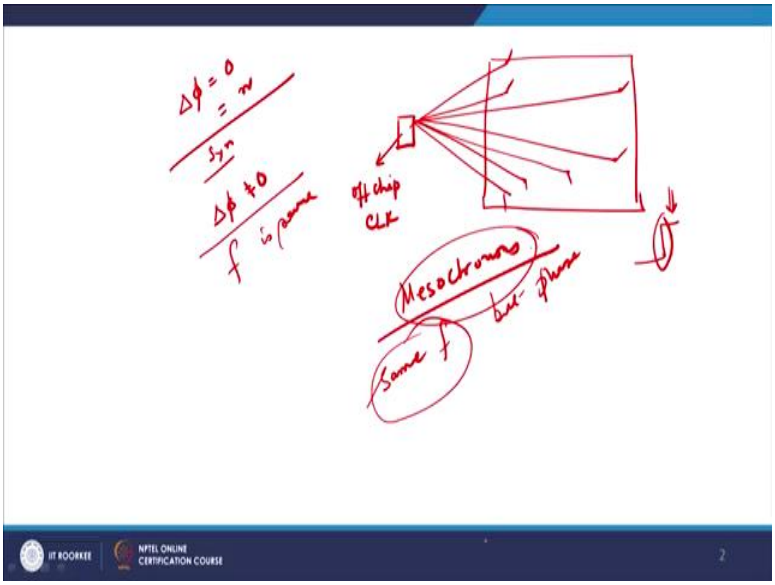
So which means that, if I assume that the logical block has got 0 delay then I would expect that the insertion of data at register 1 and output at register 2 are exactly the same time domain provided a delay is 0 for the combinational logical block. In reality not true and

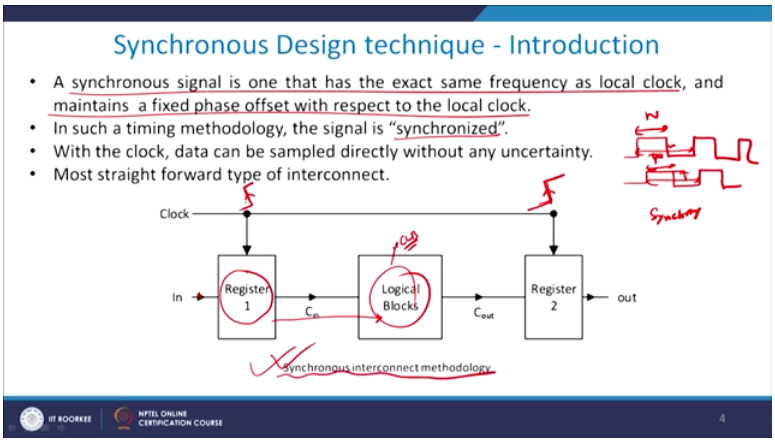
therefore that gives you a constraint on the total time period of the external clock for sequential logic.

Now with the clock as I discussed with you data can be sampled directly without any uncertainty and so this is the most straightforward simplest interconnect design for using your synchronous digital technique and it is pretty easy to implement also.

There is also another 2 types of important designs which is there with us. One is known as Mesochronous us and other is known as plesiochronous.

(Refer Slide Time: 10:10)



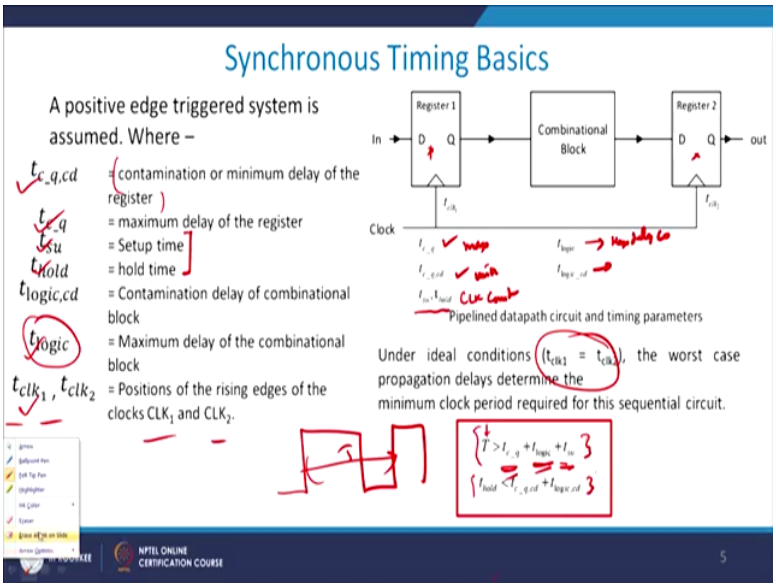




So I have got 2 type of designs one is known as Mesochronous. In a Mesochronous, synchronous we have already discussed. Exactly in phase or equal phase and they are same frequency. Mesochronous have the same frequency, so 2 clocks, so I have a master clock and I have a clock, I have a data train then the data train is said to be Mesochronous provided they have the same frequency but we cannot say it about the phase. So phase will be varying.

In the first synchronous case, you will have the same frequency of course but the phase will be fixed. Phase difference will be fixed, so the phase difference $\Delta\phi$ will be either 0 or some value, integral value N and it will remain fixed for all the edges or all the designs and therefore this is known as synchronous, right? Whereas let us suppose $\Delta\phi$ is not equal to 0 and it remains like this for a period of time we will define that to be as a Mesochronous though your frequency is same, fine.

Plesiochronous basically means that neither this frequency nor the phase is same in both the cases, so there are 3 types of interconnects models available we will not go into details of that one at this stage but I will give you a brief idea later on if time permits. So this is basically synchronous interconnect methodology which is in front of you and gives you quite a good definition of synchronous timing.

(Refer Slide Time: 11:38)




Now let me come to the basic timing analysis, we have already done this in our previous class but I will still like to give you an inside window once again, so the first delay which you see is the setup and whole time, very simple. Setup time is defined as the time before the rising edge of the clock when the data should be held stable, so that the input register is able to properly accept the data.

What is hold time? Hold time is the minimum time after the rising edge of the clock, same clock till which the data must be held stable for the clock to understand the fix data has come to our system. Now maximum delay of the combinational logical block is known as $t_{plogic}$ or

even $t_{logic}$, $t_{clk1}$ and $t_{clk2}$ are the positions of the rising edges with respect to clock 1 and clock 2, so we have done with this, we have done with this understood setup and hold time.

Now the maximum delay of these registers this or this is referred to as $t_{c\text{-}q}$. Now $t_{c\text{-}q}$ is defined as the contamination or minimum delay of the register. So $t_{c\text{-}q}$ is the maximum delay of the register whereas $t_{c\text{-}q,cd}$ is basically defined as the minimum delay of the register and $t_{logic,cd}$ is the combinational logical block error in terms of the delay, higher delay higher error will be there.

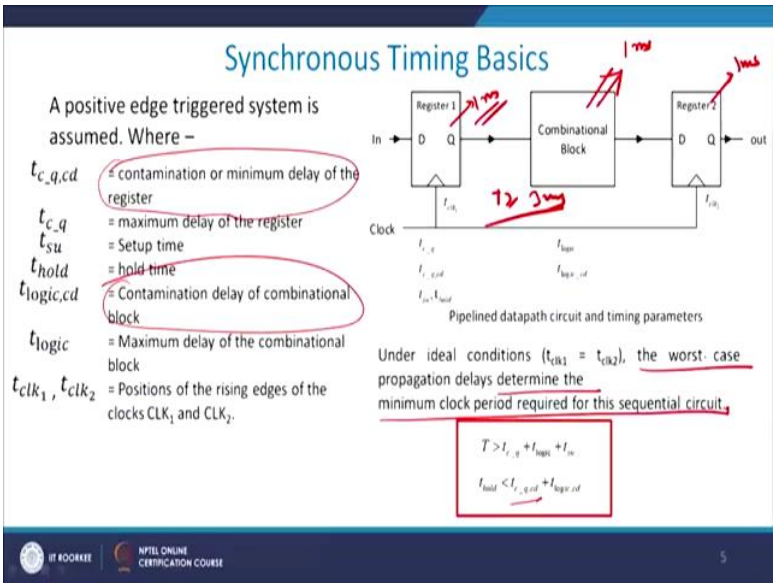
So we have got 5 quantities here $t_{cq}$, which is your maximum delay of the register $t_{cq,cd}$ is the contamination of the minimum delay of the register. So this is max delay, this is min delay, this t setup and hold is clock constraint and then $t_{logic}$ is basically your maximum delay of combinational logical block and $t_{logic,cd}$ is basically your delay because of the logic itself.

So under ideal conditions as I told to you in our earlier discussion that $t_{clk1}$ will be exactly equals to $t_{clk2}$ and the worst-case propagation determine the minimum clock period require for the sequential circuit is this much. I think you will understand, see $t_{cq}$ plus $t_{logic}$ plus $t_{setup}$, why it is like that? Because your minimum clock period of the input clock must take into account the setup time violations or setup time logic $t_{logic}$ is basically the contamination delay of the combinational logical block or at this stage is the total delay of the combinational logical block and $t_{cq}$ is the maximum delay of the register.

Because you see the signal passing through registers through combination logical blocks and then finally going to the output. So the minimum capital T, capital T means the clock period, right? This is the clock period should be at least greater than $t_{su}$, right? Because otherwise you will have a problem tlogic also the same thing and $t_{cq}$ also the same thing, so capital T should be greater than $t_{cq}$ plus $t_{logic}$ plus $t_{su}$. What is $t_{hold}$? $t_{hold}$ is that I discussed with you.

After the rising edge of the clock the time till which the data should remain stable is defined as $t_{hold}$. And it is given as $t_{logic,cd}$ minus…

(Refer Slide Time: 15:04)



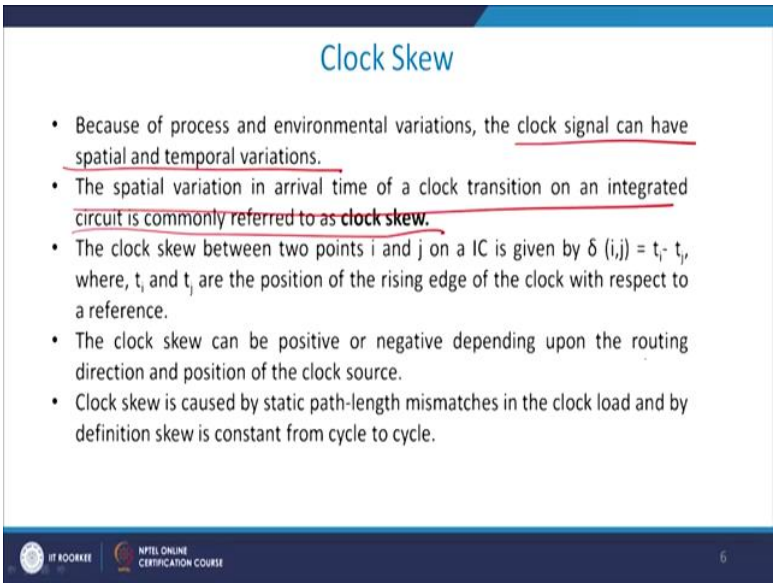

So $t_{hold}$ is equals to $t_{c\text{-}q,cd}$, what is $t_{c\text{-}q,cd}$? It is the contamination delay of the register of the minimum delay of the register plus $t_{logic,cd}$ which is contamination delay or the minimum combination and you can understand why? So this logic will have say one millisecond delay, right? And then this is having 1 millisecond delay, right? And let us also assume that this is having 1 millisecond delay.

So capital T should be at least larger than equal to 3 milliseconds. If it is not then any of the 3 operations will have to be violated which is not possible in a sequential logic, right? And the worst-case propagation delay requires that the delay determines the minimum clock period required for the sequential logic. I will give an example. For example if you are in a situation where you are using a sequential delay for a very stress condition.

What does it mean? That your voltage levels and current levels are very high then what does it do? Is that it gives you the worst-case propagation delay for a sequential logical block, right? So this gives you the worst-case preparation delay in a sequential logical block. As you can see capital T is greater than $t_{cq}$ plus $t_{pd\ logic}$ plus $t_{setup}$ plus $t_{logic}$ and $t_{hold}$ should be less than $t_{c\text{-}q,cd}$ and $t_{logic,cd}$.

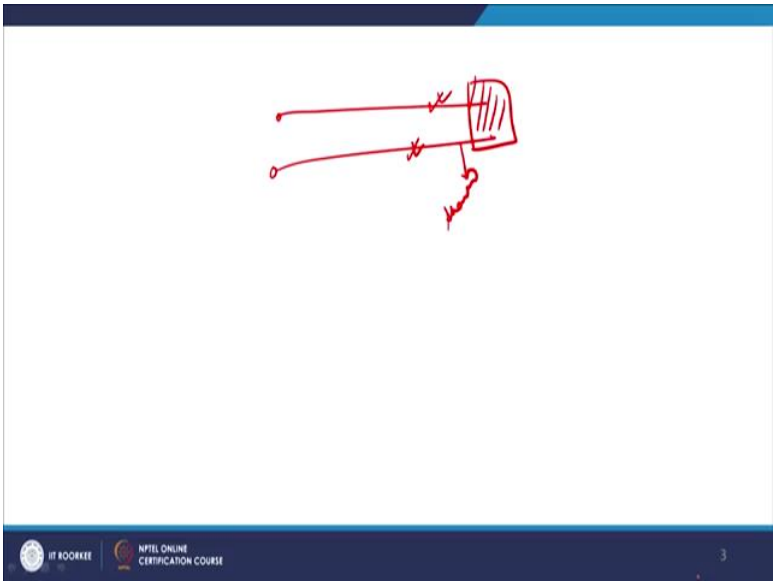
Now why $t_{hold}$ should be less than that? Let us suppose $t_{hold}$ is greater than that then what will happen is that either of these 2 problem areas will be inserted into the logic principle and there will be an issue as far as output is concerned.

(Refer Slide Time: 16:42)




So let us look at now at clock skew. We have understood the basic idea about the delay concepts, let us come to clock skew because of process variations and environmental variations the clock signals can have spatial and temporal variations, so how did you define clock skew? The spatial variation in the arrival time of the clock on an IC is referred to as clock skew.

(Refer Slide Time: 17:10)



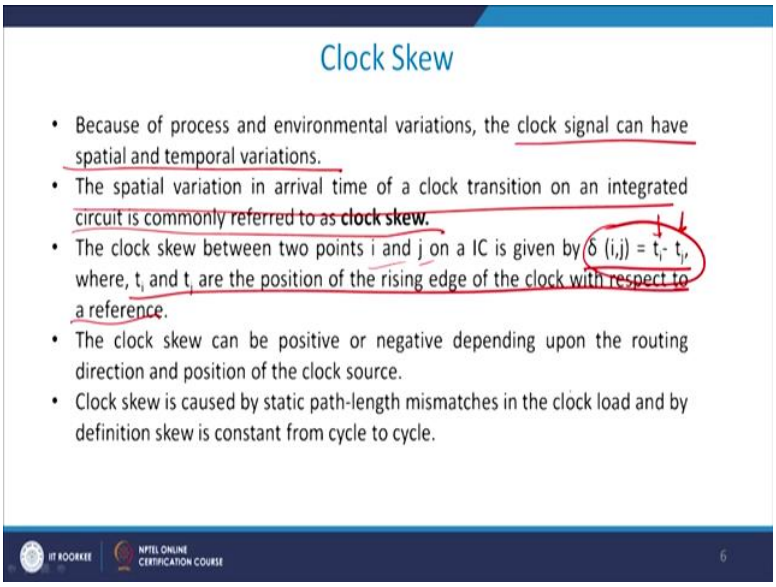

Which means that let me give you an example, let us say I have a IC here and this is driven by 2 external clocks. If this clock take higher time as compared to this for whatever reasons we define that this clock is basically having a skew design and you need to formulate it, why

these are required? Because see until and unless you send all the signals in transparent mode as well as properly you will not be able to evaluate the output signal.

And that is what is happening here, so what is the clock skew? The spatial variation in the arrival time of a clock transition or an IC is referred to as clock skew. Now the clock skew between the 2 point i and j is given by $\delta$ (i, j) but $t_i$ minus $t_j$ and this is quite important also. So for example if you have clock here.

Maybe I can, right? Maybe I can discuss here, right? But nonetheless, let me see how it works out.

(Refer Slide Time: 18:20)



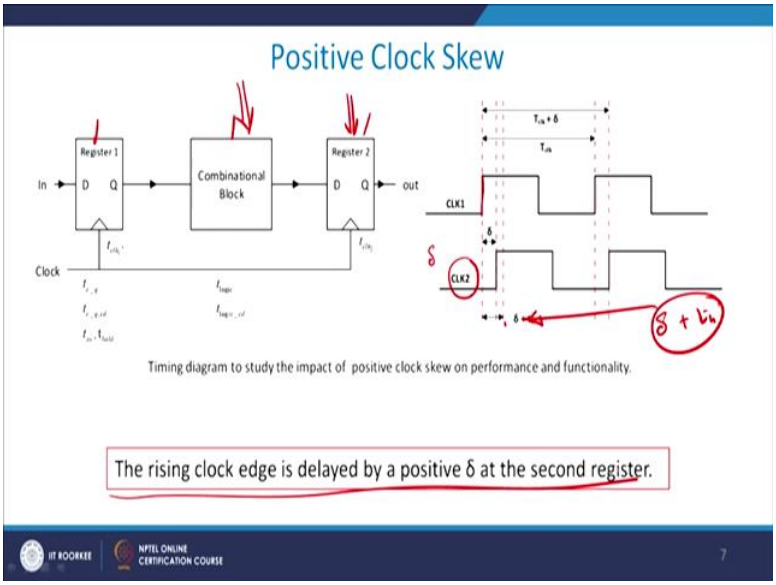

Now you see that the spatial variation in time of arrival of the clock. Transition on an IC is referred to as a clock skew. In the arrival time, please understand it is not on the terms of delay, it is not in the terms of frequency it is arrival time it is important. So if it has arrived later than the clock skew it is known as positive skew. If it arrives ahead of the rising edge of the clock we define that to be as the premature case.

So one has to be very careful what is clock skew. As I discussed with you the clock skew between 2 points in an IC with having i and j as the subscript, we can write $\delta$( i, j) is equal to $t_i$ minus $t_j$ where $t_i$ is the time taken for the data to reach at the particular point and $t_j$ is basically the output point with respect to the rising edge. So $t_i$ and $t_j$ are the position of the rising edge of the clock with respect to a reference.

Now the clock skew can be positive or negative depending upon the routing of direction. So I will like you to find it out. If the direction of the clock and the data is same you will get positive skew, if the direction of the clock and the direction of the data moment is opposite to each other we will get a negative skew and I would like you to find out why so? Now clock skew will remain same from cycle to cycle, right? It does not change.

Because it does not depend upon high mismatches and so on and so forth. It only depends upon clock load but unlike jitter which we will see later on clock skew is very stable in terms of the output which is there with me, right? The clock skew as I discussed with you positive or negative depending upon the routing and the direction of the positive nature of the clock source.

(Refer Slide Time: 20:15)

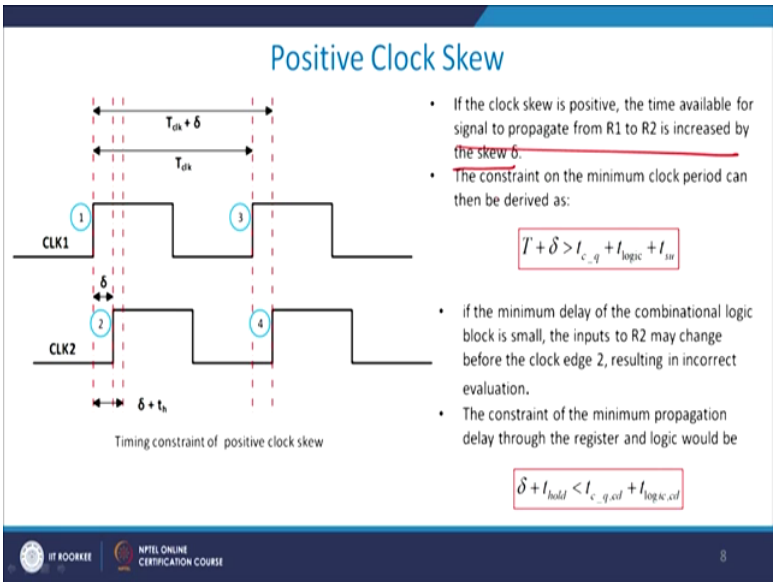



So let us look at what is known as a positive clock skew. What is a positive clock skew is, I have register 1, register the combinational logic and register 2 is there with me. Now I have a tclk1 tclk2. So at the rising edge of the clock 1 nothing is happening because in clock 2 you already have a skew which means that the rising edge will shift by right to accept this data and therefore you get δ is equal to this much.

Now in reality it has to also hold a data equals to $t_h$, the total delay is basically δ plus $t_h$ and that is what we have seen at this particular point, fine. Is it conceptually cleared? Right. So combination logical blocks job is to accept the data and then do some logical operation on it and then send the data to the registered number 2. So till a time this happens register number 2 is sitting ideal.

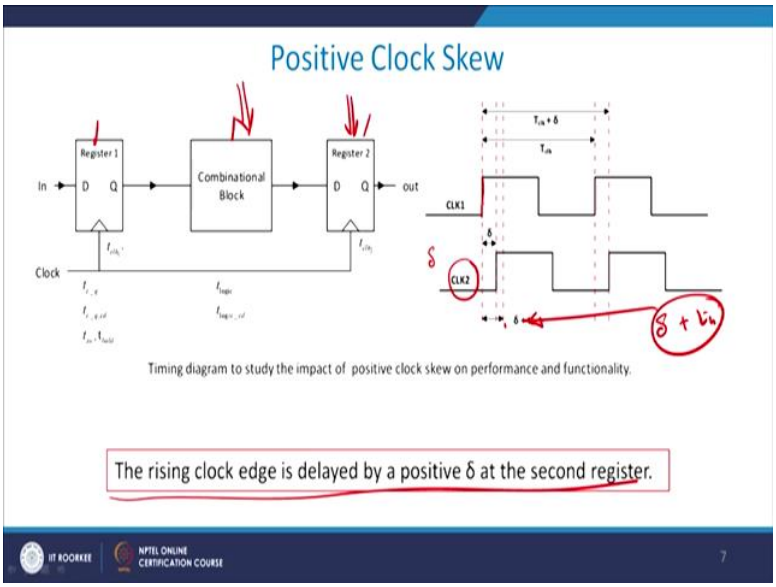
And what does combinational logical block try to do? It tries to accept the data from register 1 and processes it and give it to register 2. Register 2 within a few period of time manipulates it and send it to the output, right? So the rising edge is delayed by a positive δ the second register, you got the point. Why? Because if you see carefully by the time this reaches the second register the clock one might have actually gone behind as compared to clock 2 or maybe ahead also and therefore you can get a rising edge delay in terms of Δ in second register.

(Refer Slide Time: 22:02)



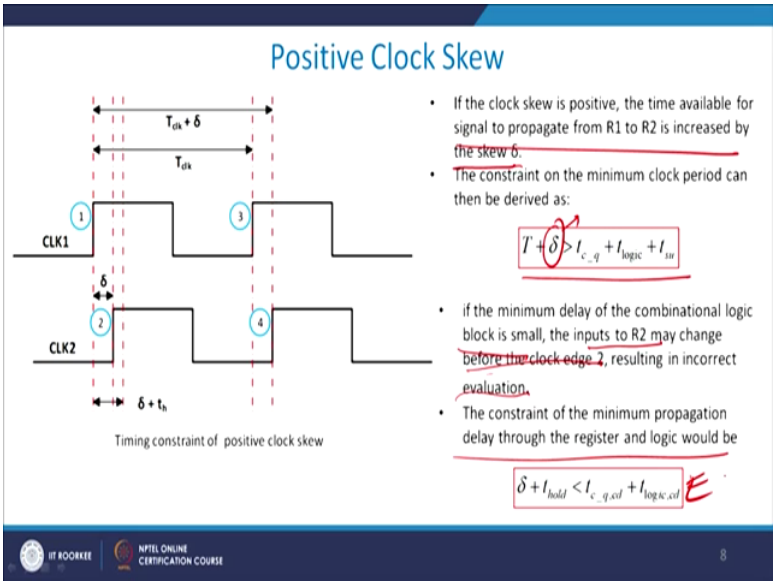

Okay, as I discussed if the clock skew is positive the time available for signal to propagate from $R_1$ to $R_2$ and increased by δ, I suppose this is clear no need to explain.

(Refer Slide Time: 22:13)




For example if the clock skew is positive then I get a larger time to integrate my design, right? And therefore it is much easier to do that.

(Refer Slide Time: 22:27)

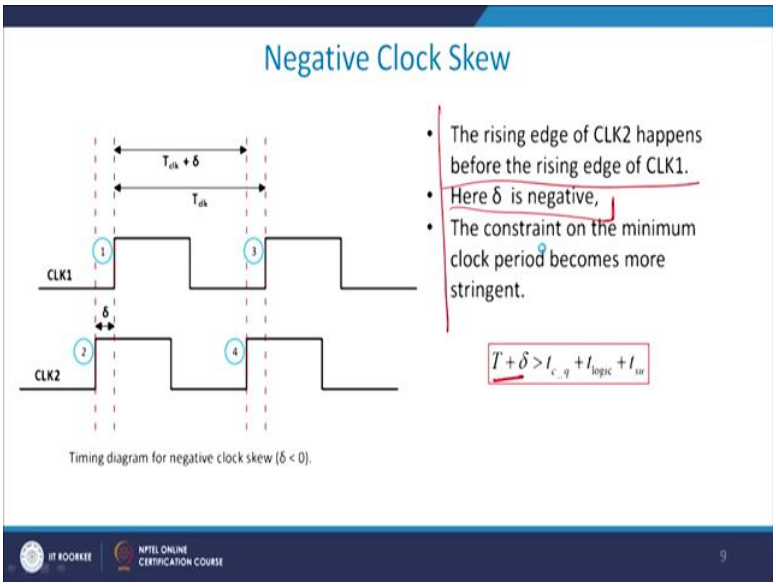


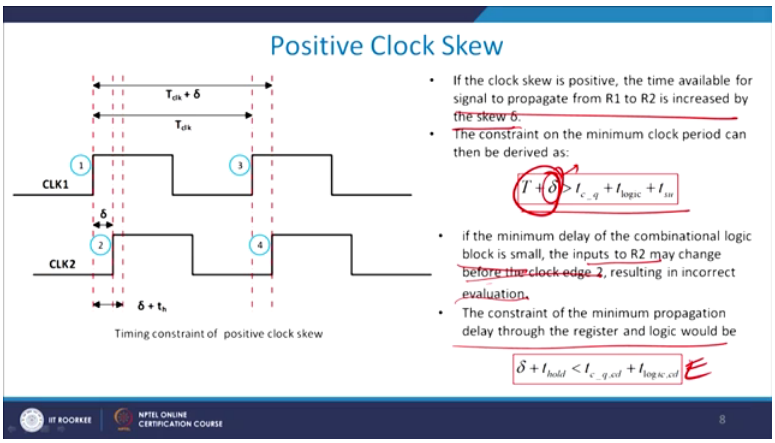

And therefore if you look very carefully positive clock skew is always beneficial in nature. Now let us look at the constraint of the total timing delay and the total timing delay is given by T plus δ, why T plus δ? Because that is the capital T is the total time period when the system is going from 0 to 1 and 1 to 0 and the value of δ can be found out from this relationship.

But the value of δ will not be varying, will be fixed one for the fixed cases of interest. So if the minimum delay of the combination logical block is small the input register 2 may change before the clock edge of 2 resulting in incorrect evaluation, you got a point. That means if the combination logical block has very low delay then the output of gate number 3 exactly matches with the gate number 1 rising pulse.

But that has to be available to gate number 2 in order to evaluate that person, right? But if it is not available, it is lost. So that is the problem area of T logic or transistor logic. The constraint of minimum propagation delay through the register in the logic would be as I discussed with you, it is not there would be this much, right? So the constraint of minimum propagation delay, what is the minimum propagation delay? It is the maximum error will be given by this particular formula here, right? That is what we get from the positive clock skew.

(Refer Slide Time: 23:57)

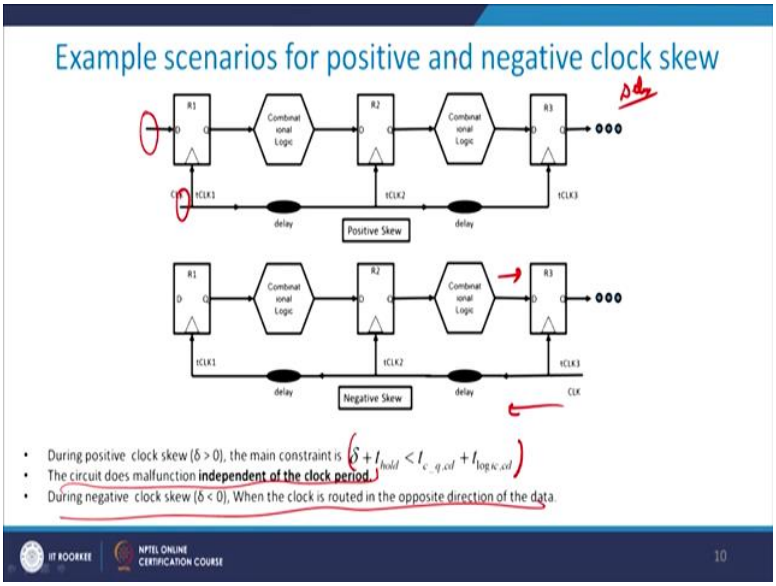


Let me come to negative clock skew. Negative clock skew is just the reverse of positive. In the positive clock skew the data part and the person who is looking at the skew they were in the same direction. So if the data path is going from X to Y the clock was also moving from X to Y. Here the reverse is happening when the clock is moving from left to right the device is operating from right to left and therefore we define negative is to be negative in nature.

The rising edge of the clock 2 happens before the rising edge of clock 1 and therefore δ is negative and therefore here δ is kept to be negative. The constraint on the minimum time period becomes more stringent that T plus δ should be greater than or equal to this much. Why it is δ? Because this δ is a minimum difference between the set up time and the rising edge of the pulse for a MOS device in the study and that the reason it is converted into this simple form where you can understand the T plus δ which is basically your this quantity. It should be always greater than this thing $t_{su}$, $t_{cq}$ and $t_{logic}$.

(Refer Slide Time: 25:18)




With this we come to an example scenario of a positive and negative clock skew, right? And I can say here that for positive clock skew this input coming from $R_1$ is sort of a delay element which goes to $R_3$ directly so $R_1$ to $R_3$, so there are 3 registers $R_1$, $R_2$ and $R_3$ and there are 2 combinational logical blocks, combinational logic 1, combinational logic 2 but please understand to move from $R_1$ to $R_3$ it has to be properly tuned in terms of potential, so that it can move easily from point A to point B This is the case of positive skew.

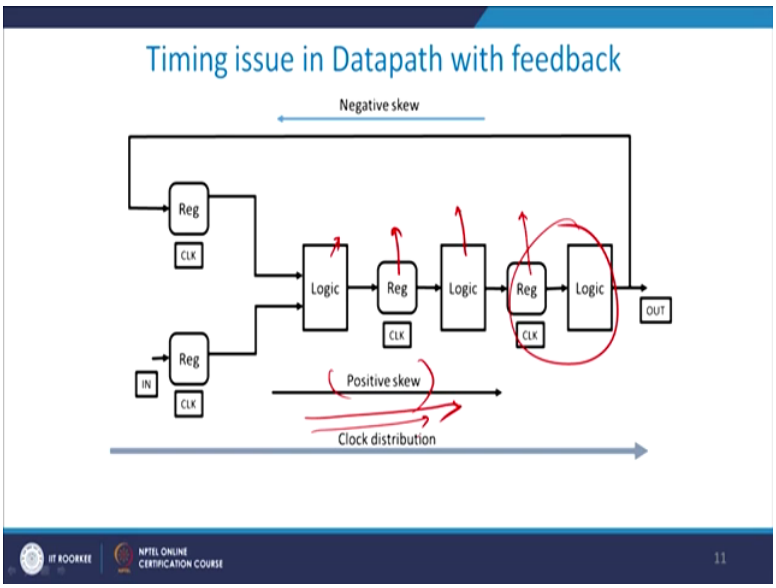
Positive skew means the data train which is this one and the clock which is this one they move exactly in the same direction, they move exactly in the same direction. So the delay of

the combinational logical block is basically your internal delay. Similarly if the clock is in the opposite direction as compared to data you will have a negative delay at this point and this point.

So during the positive edge of the clock δ is greater than 0 and therefore I should maintain δ plus $t_{hold}$ should be less than equal to this quantity. And the circuit does malfunction independent of the clock period. Sometimes there is a malfunctioning in the clock and therefore it stops working. During negative clock skew when the data is routed to the opposite direction of the clock we get the negative skew in this case.

So we understood what is the positive skew, what is a negative skew and it works.

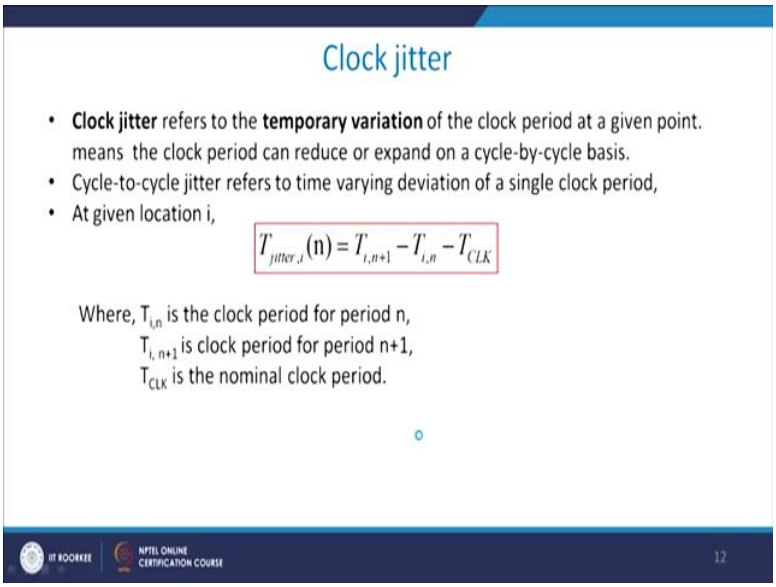
(Refer Slide Time: 26:54)




Now timing issue in a data path with feedback. Let us suppose I have a negative skew and this is a feedback which is there with my… have a logic here. I have logic which takes care of the logical experts then we have got a register in our house then this is basically a logic analyzer. Same is the case with of this one, these 3 are basically simple players in this game. So this is basically a positive skew.

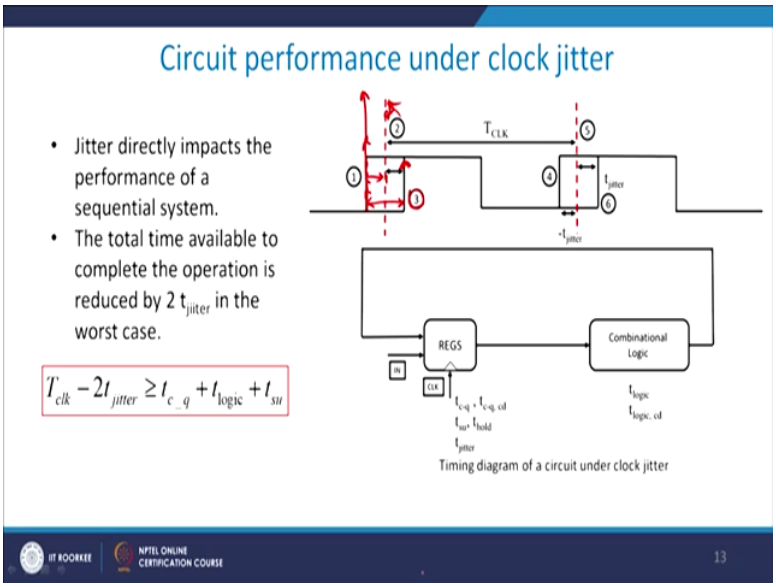
Because this is by data train path and this is my moment of the various blocks from left to right and by using a simple rather than a problem area.

(Refer Slide Time: 27:36)




Let me come to one more important point and that is clock jitter and we will take up this one in the detail manner in the next turn. Clock jitter refers to temporary variation of the clock period. So the clock skew was spatial variation whether there is a physical movement of the clock from point A to point B.

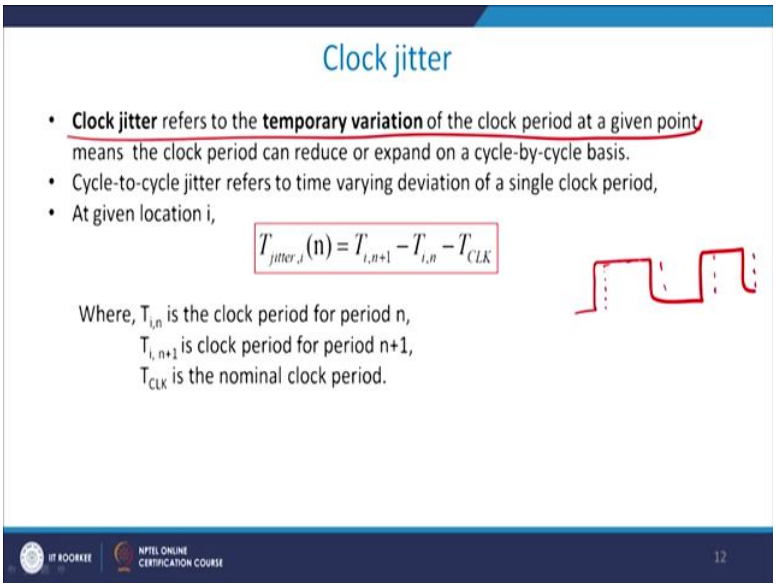
(Refer Slide Time: 27:59)




Temporary variation primarily means that you do have this much amount of variations so you initially had 2. At 2 you had this thing and then what happened is after part 2, at one suppose you are giving a positive edge to get design here. But what has happened is that you have to wait till this edge number 3 to come for which you need to wait till this much amount of time domain when the next clock higher edge comes.

So initially it was triggered at 1, it was supposed to be triggered at 1, what has happened is, it is not triggered to either at 3 or 2, so there will be a small this $t_{jitter}$ which will be there. This is known as a temporal variation, in time domain you will see a shift and it can vary from 1 block to another block, the skew remains constant from block to block but the jitter may vary from 1 block to another.

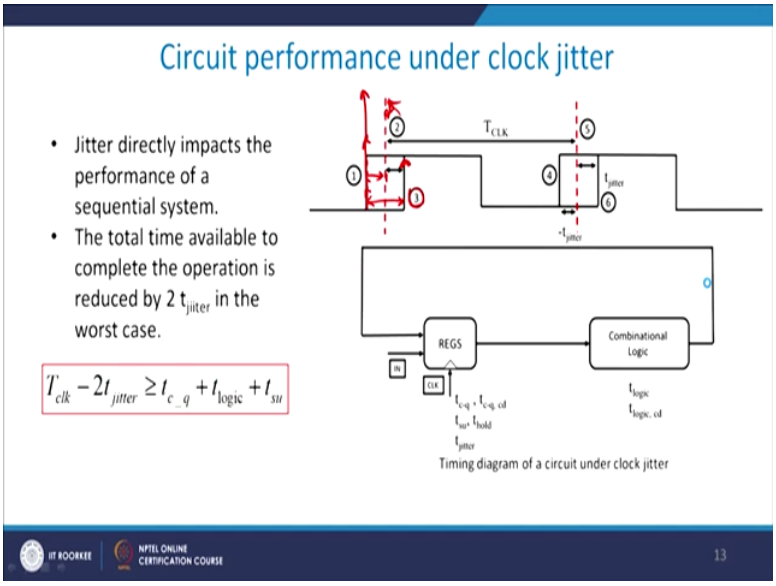
(Refer Slide Time: 29:05)




We will work out the details of it in the later stages but primarily if you look very carefully this case you can see that therefore clock jitter refers to the temporary variation of the clock period at the given period of time. So initially you had this into the consideration. Now what has happened is because of some problem we will discuss that later on that it has not shifted from this side to this side.

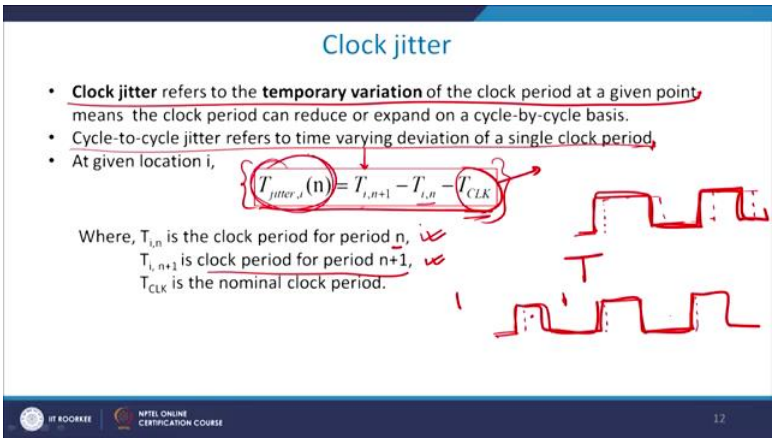
Everything has shifted by a particular phase from this side to this side, this shifting is there.

(Refer Slide Time: 29:31)




As you can see this shift is there. So this was initially here this shifted here.

(Refer Slide Time: 29:35)




And as you can see therefore there is a cycle to cycle refers to time varying deviation of a single clock period. So you had a single clock period of capital T. Now because of cycle to cycle variation in the jitter your new value at location I will be given as $T_{i\ jitter}$ to be equals to $T_{i(n+1)}$ minus $T_{in}$ minus $T_{clk}$, where $T_{in}$ is basically the clock period for period n and $T_{in}$ plus 1 is for $n^{th}$ plus 1.

So when you vary from first-period to the second period, so this is one and then when you vary for the second period there is a lateral shifting. So what has happened is that we defined $t_{jitter}$ to be this minus this the difference because that is extra 1 minus clock, why? Because clock is common to both, so when you subtract that the difference will come out to you between the various values of skew and jitter right.

So we will take-up the next part of the strategy in the next clocking strategy in the next morning. Thank you very much.

---

Okay so clock jitter refers to temporary variation of the clock period at a given point of time which means that the clock can expand or reduce from one cycle to another. So this is one cycle, let us suppose this is one cycle and I have another cycle, so this is one cycle, this is another one cycle and then this is another cycle. So between this cycle and this cycle you might have a difference in the rising edge of the clock this is known as jitter.

So this is a temporary domain whereas skew is special domain. Skew does not change from cycle to cycle whereas jitter changes from cycle to cycle. We define jitter at any particular point in given location i to be equals to $T_{i,\ n\ +\ 1}$ which is basically the clock period for $(n + 1)^{th}$

minus Ti n minus at nth clock. So when you go from $T_{(n+1)th}$ clock to $n^{th}$ clock the difference between the 2 is basically your jitter.

But you understand within that you already have the clock period common to both of them. So if use subtract, when you subtract that you actually get the overall jitter which is there, right? So this is basically from cycle to cycle. Once you have one cycle then another cycle the jitter varies from cycle to cycle, right? With this we have finished with this first module of your sequential logical clocking strategies.

In the next block we will take care of the combined effect of skew and jitter on the performance of a sequential logical block, thank you very much.

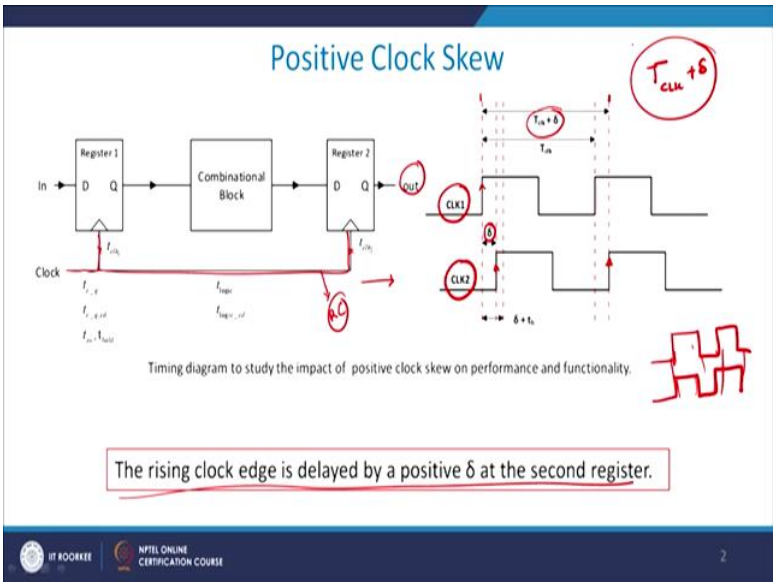
.

**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−59**
**Clocking Strategies for Sequential Design-II**

Hello everybody and welcome once again to the next edition of NPTEL online course on Microelectronics Devices to Circuits. We start off with the second module of the clocking strategies for sequential logic and we will see what we have learned earlier and then we will go forward from the place where you have left especially the Jitter. We learn that the concept of Skew and we said that this Skew happens because the interconnect lengths at this particular point and at this particular point are different.

Which means that if a clock is routed through this much path and the clock is routed through a larger part than this will give you some RC delay and as a result this clock will be slightly delayed as compared with the first clock. So if your first clock rising edge is somewhere here and your second clock rising edge is somewhere here then we define this to be as my clock Skew, right?

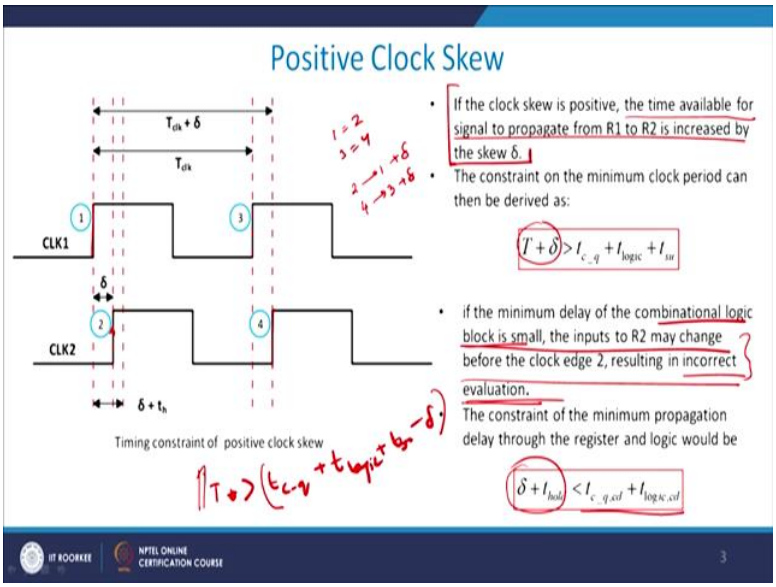
(Refer Slide Time: 1:20)




And therefore what happens is that, so this clock1 and this is your clock2 and then the distance between this point and this point is of course $t_{clk}$ plus δ, why? Because this whole thing has shifted, so you see after clock2 only your output will be in the output state and therefore what has happened is the rising edge, clock edge is delayed by a positive δ.

So what is happening is that, you are actually evaluating at this particular point during the rising edge of the clock 2. Which means that the total time spent is $T_{clk}$ plus δ which means that you initially had only $T_{clk}$ to do if they were not Skew because if they were not Skewed than clock 1 and clock 2 will exactly fall each over them exactly same and therefore there will be no difference between the Skew between the 2. This is Skew you get an extra δ coming into picture and that reduces the performance of the system.

(Refer Slide Time: 2:21)




Now if the clock Skew is positive, if the clock Skew is positive which I have shown already then the time available for the signal to propagate from $R_1$ to $R_2$ is increased by the Skew factor of δ.

As you can understand since I am evaluating at edge and number 1 and evaluating at edge number 4 if there would not have been any Skew 1 will be equals to 2 and 3 will be equals to 4. In place of Skew 2 is shifted from one by a factor of δ. Similarly 4 is shifted from 3 by a factor of δ and therefore the total time is basically T. So that the reason when we write T plus δ, so if you remember our previous discussion that T should be greater than $t_{cq}$ plus $t_{logic}$ plus $t_{setup}$.

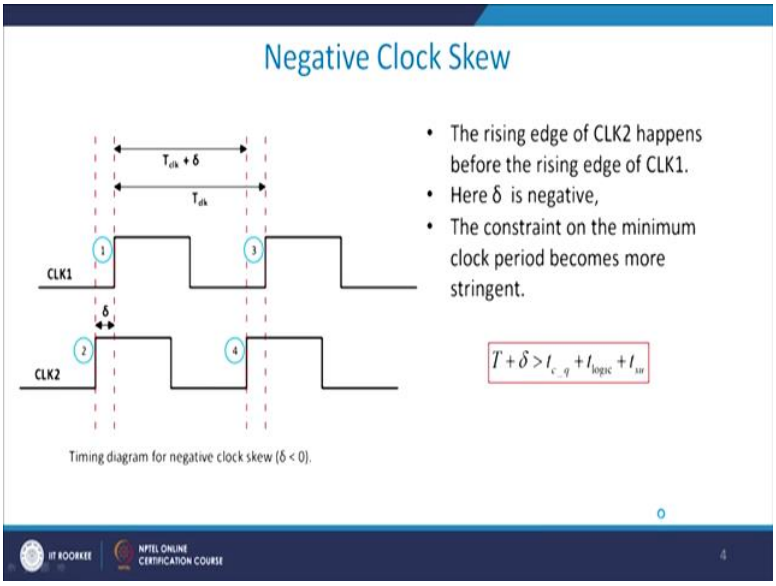
But now I say that T plus δ should be this much, so if you take δ on the opposite side. So T should be greater than $t_{cq}$ $t_{logic}$ and plus $t_{setup}$ minus δ, so if the δ goes higher and higher this Skew increases then this reduces and as a result your performance becomes better and better. In the sense that you are now able to achieve a larger, faster clock.

So the frequency improves but the cost you pay for it is that that there might be a problem, that is what I am coming to the second point. If the minimum delay of the combinational logical block is small the inputs to $R_2$ may change before the clock edge 2 resulting in incorrect evaluation. So though you have gained in terms of higher clocking speeds in terms of because T has increased but the price you are paying for it is basically that let us suppose your combination logical block.

Which was there between the 2 registers, if the delay is very-very small, so what has happened is, that if the register are 2, if the minimum delay of the combinational logical block is small the input to $R_2$, right may change even before the rising edge of the clock this second, are you getting my point? So your output was synchronize with respect to clock 2. Now what has happened is, since the delay of the CLB is very small it has actually come before the rising edge and you are not able to evaluate it and therefore resulting in incorrect evaluation and that is the constraints which you get.

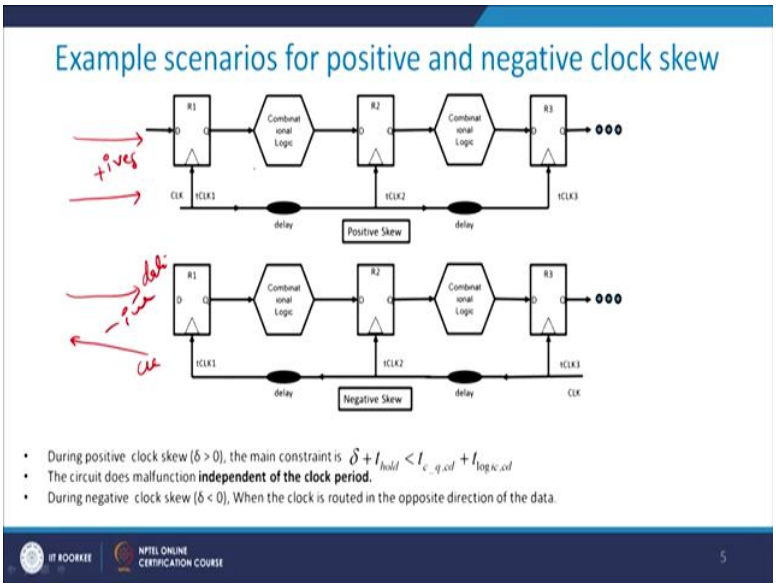
The constraints of the minimum propagation delay through the register is $\delta$ plus $t_{hold}$ should be less than $t_{cq,\ cd}$ plus $t_{logic,\ cd}$. If it is not then there will be a problem of evaluation also for a positive edge triggered.

(Refer Slide Time: 5:05)



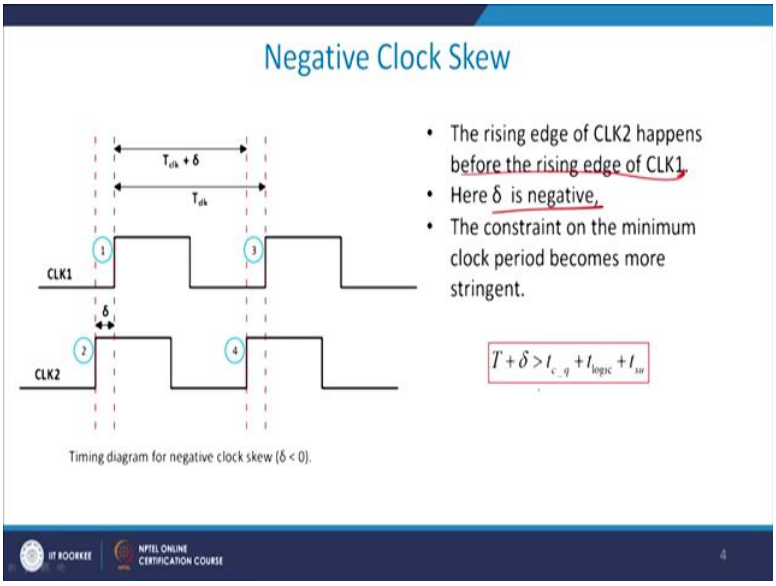

How do you define negative edge triggered? Negative edge triggered is when clock2 is moving ahead of clock1. So this is clock2 and clock1, this will happen when?

(Refer Slide Time: 5:14)




As I discussed in the previous turn that when the data, this is the data path movement and this is my clock movement, I will always get positive Skew. If my data path is opposite to my clock, this is my data and this is my clock and then we get a negative Skew, so you need to point out this particular point.

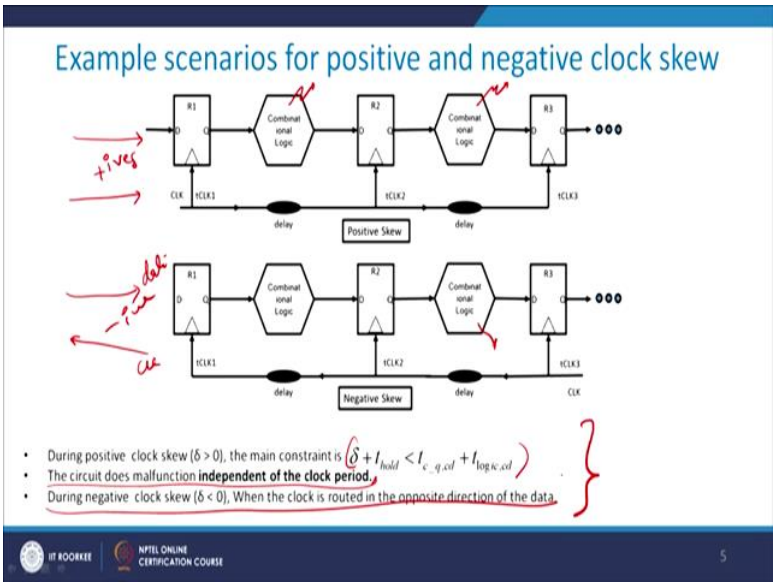
(Refer Slide Time: 5:33)




And therefore the rising edge of the clock 2 happens before the rising edge of clock 1 and therefore the δ is negative or we define that the clock Skew is negative but the constraint is therefore again as I discussed with you T plus δ should be greater than or equal to this. So the positive or the negative Skew does not influence by overall performance. It only tries to make

the point that the CLB should have a minimum, the combination logical block should have a minimum delay so that the edge 2 is properly being refer to.

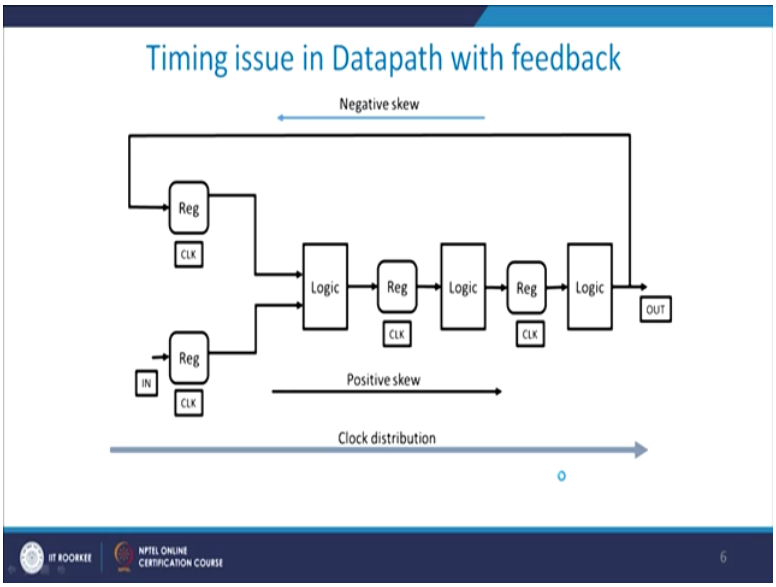
(Refer Slide Time: 6:06)




This is what I was talking about in my previous turn also that δ plus $t_{hold}$ should be less than equal to $t_{cq,\,cd}$ plus $t_{logic,\,cd}$. Now therefore the logic does malfunction independent to the clock period, you got the point. If the combination logical block for example here this case, right? Or this case, right? Or even this case, if this combination logical block relatively have a very very low delay then the clock has already arrived but your data is not available to be evaluated, so get a wrong data output in the case.

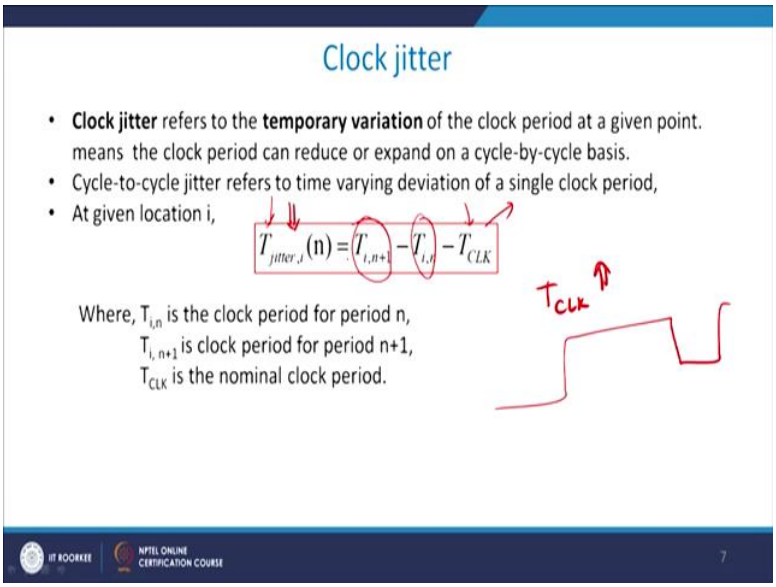
So as I have discussed with you in the negative Skew the clock is always routed in the opposite direction to the data. Whereas in the positive Skew they are always routed through the same data.

(Refer Slide Time: 6:51)




Now clock distribution if you look very carefully, as I have discussed with you. As you move from input to output if you move in the direction of the data flow you will always start getting positive Skew and if you move in the opposite direction to dataflow you will get a negative Skew, right? But both will help you in terms of performance and reduce delay at the cost of wrong evaluation. So functionality might be under question if this is there.

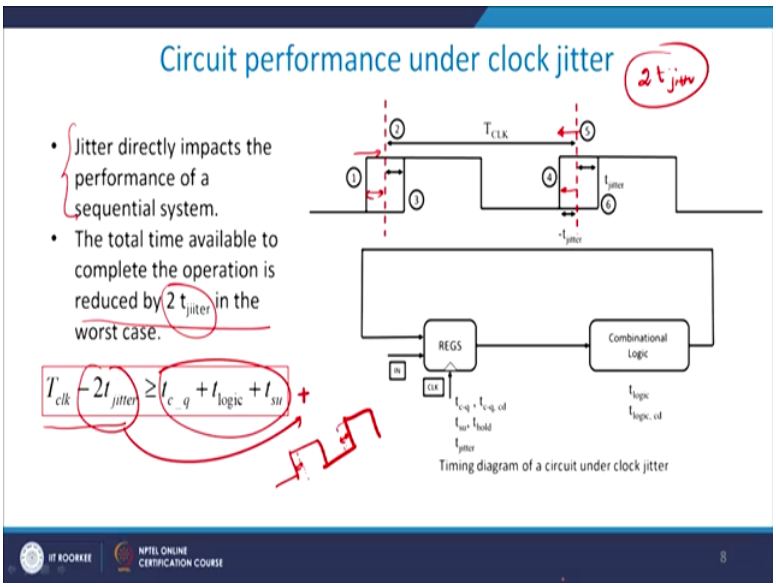
(Refer Slide Time 7:16)




We also discussed clock Jitter that is a temporary variation and we define that to be as $T_{i, n+1}$ minus $T_{i, n}$ minus $T_{clk}$ and so on and so forth that is defined to be as $T_{Jitter}$, so if the clock period is very-very high, large then I will get a lower value of $T_{Jitter}$. So if my $T_{clk}$ is large which means that I have a large $T_{clk}$ I will get a lower Jitter but then you can understand large T means a lower frequency.

So a lower frequency will allot you lower Jitter, so a lower Jitter can only achieve through a lower frequency operation of the system.

(Refer Slide Time: 7:59)

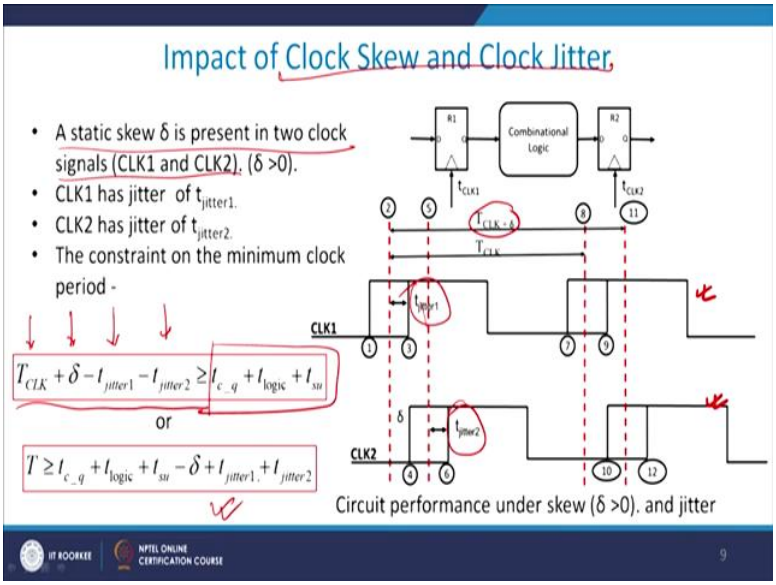


So as I was discussing with you the Jitter directly impact the performance of a sequential circuit. The total time available to complete the operation is reduced by 2 Jitter in the worst case. As you can understand see what will happen is, if by far this 1 moves to let us say 2 and this 5 moves to 4, let us suppose. So this moves by $1t_{Jitter}$ and this moves by $1t_{Jitter}$, so the maximum movement one can handle is $2t_{Jitter}$.

And that is what is written that the worst-case Jitter is basically $2t_{Jitter}$. Why $2t_{Jitter}$? Because if this is your clock, right? And this is your rising edge and falling edge of the clock then let us suppose this is moved by this direction and this is moved by this direction temporarily. So I lost 1 this much and I lost this much, so this is $2_{tJitter}$ which you get. So I get $T_{clk}$, $T_{clock}$ minus $2t_{Jitter}$ should be always greater than this value.

Which means that if you take $2t_{Jitter}$ on the right-hand side it is always adding up. So if you take this to the right-hand side it will always add up which means that Jitter will always lower the frequency of the output signal, right? Whereas Skew in fact improves it to an extent but then at the cost of functionality of the chip in question. Whereas Skew will actually increase the overall delay means lower the frequency actually.

(Refer Slide Time: 9:32)

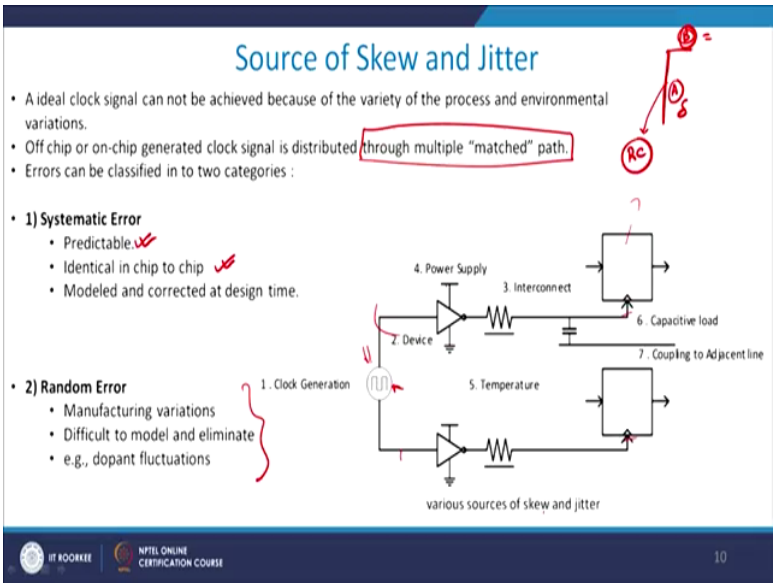



Now if I take both clock Skew and clock Jitter together, let us see how it impacts the overall circuit. As I discussed with you a static Skew δ is present between 2 clock signals clock1 and clock2, so let us suppose this is clock1 and this is your clock2, this is $T_{clk}$ for the first case and this is the $T_{clk}$ for the second case, right? And then you have $T_{clk}$ plus δ coming into picture because of the clock Skew which is there with us.

And there is also a Jitter which $t_{jitter1}$ here, right? And you have $t_{jitter\ 2}$ here, so there are 2 different clocks and 2 different Jitters. So the constraint on the minimum clock period is the $T_{clk}$ plus $\delta$ minus t Jitter 1 minus t Jitter 2 because the worst-case scenario was twice to t Jitter. But now let us suppose that the Jitter of first clock and second clock are unequal then we write $T_{clk}$ plus $\delta$ minus $t_{jitter\ 1}$ minus $t_{jitter\ 2}$ must be greater than equal to this value.

Or I can shift everything on this side and we take minus $\delta$ plus $t_{jitter\ 1}$ plus $t_{jitter\ 2,}$ so this is a generic form of having Skew and delay. Skew will always try to improve the frequency, Jitter will always try to reduce the frequency in a sense.

(Refer Slide Time: 10:50)




Now let us come to this source of Skew and Jitter and we will see how it works in that sense. An ideal clock signal cannot be achieved as I discussed with you because let us suppose you have a problem, the problem is something like this. If the clocks are generated generally in the off chip because on chip the power dissipations are very high, it is getting heated up and therefore you generate clock off chip.

Once you generate clock off chip and then you route the signal from outside the clock to inside the chip and driving an active device or maybe a combination logical block or a register then you will see that the routing distances is not equal. So if am moving from this point to this point and I am moving from this point of this point, suppose this is path A and path B.

Path A will obviously have a larger RC delay as compared to path B and therefore I would expect to see a larger Skew here as compared to point B that is what is written and therefore

though through paths if they are matched paths which means the RC δ are equal then I will not get the Skew. If they are unmatched you will always get a Skew. There are 2 types of error therefore.

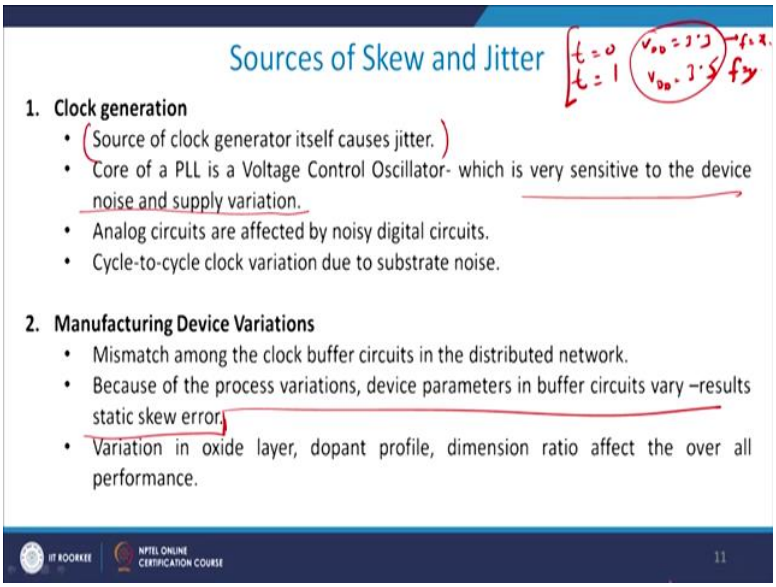
One is known as the predictable error which you can predict by virtue of the mathematical treatment in terms of how many, what is the interconnected length and so on and so forth. And then there identical from chip to chip, so if you take the same length in any of the IC which you are fabricating the delay will almost be the same in this case. And therefore it can be modelled and corrected at the design time itself.

So at the design time you know what is the length of the interconnect and therefore you can predict what the value of your Skew will be and I can therefore give leveraged to my $t_{clk}$, so that the Skew is taken into consideration. However, what is a random error? When you have manufacturing defects or manufacturing variations and they are difficult to model because these are random in nature for example dopant fluctuation is one of them.

Now as I discussed with you therefore if you look at this, I am generating an example of various Skews and Jitter. So if am generating a clock here and then I have a power supply through which and I have an inverter let us suppose. And then there is an interconnect with drives register here then what will happen is, even if these 2 lines are exactly equal, right? But the clock loads might vary which means for example that a single clock might load 6 devices together.

And as a result the clock loads will vary and therefore the capacity of loading will vary and therefore Skew will vary and that there is reason you will have various Skews and Jitters in reality.

(Refer Slide Time: 13:43)



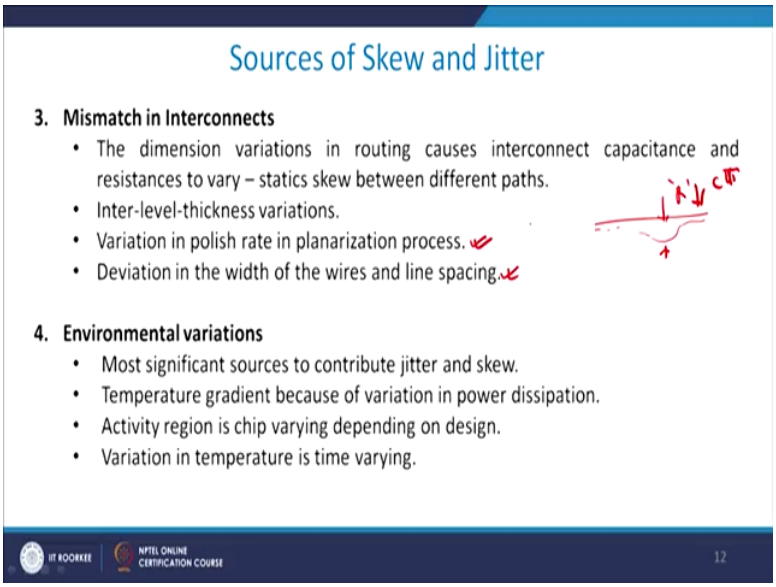

Now the source of clock generator itself causes the Jitter for example generally we use VCO which is voltage control oscillator or even PLL which is responsible for generating the clock but the problem is these are very very sensitive devices and they are prone to noise and supply variation. So even if your $V_{DD}$ which you have supplied is changing by even a fraction of amount that frequency will change slightly and therefore that gives you a Jitter.

So suppose you are at t equals to 0 your $V_{DD}$ was equals to 3 volts or 3.3 volts suddenly at t equals to 1 this $V_{DD}$ goes to 3.5 volts then this will give a frequency correspondence to x. Suppose this goes to y which means that in time domain or temporal domain you have shifted from X to Y because of just because of the variation in the Valley of $V_{DD}$ and that is what is written in the second point, bulleted point here.

Similarly cycle to cycle clock variation due to substrate noise, so you will have substrate noise you will have thermal noise especially in analog circuits and they will give rise to change in the frequency. As I discussed with you random variables or random functionalities your manufacturing device variations because of process variation device parameters in buffer circuits vary and results in Skew error.

For example if your oxide thickness varies, your dopant concentration varies, your W by L ratio varies in all these cases in Manufacturing defects you will have a variational threshold voltage for example which results in output frequency variation and therefore they will be changed in the variation of the value of a Skew and Jitter.

(Refer Slide Time: 15:21)



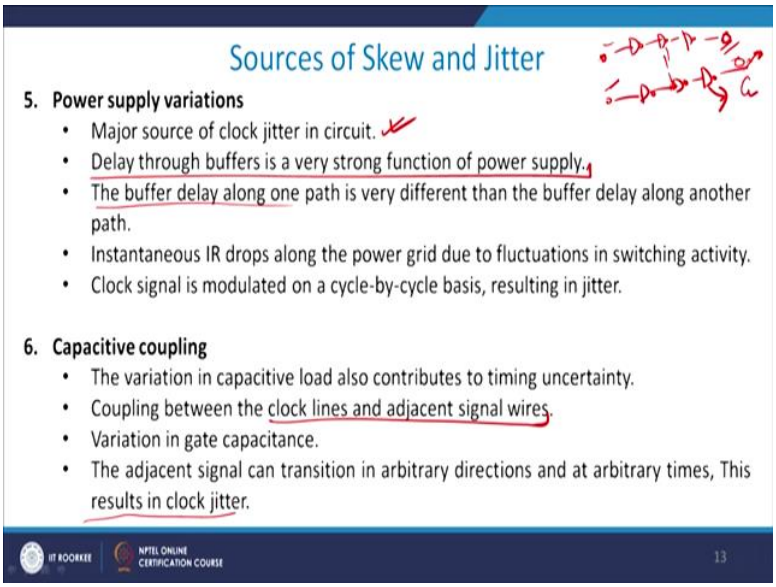

Similarly the third is basically the mismatch in interconnects this is the most prominent one and the reason is that the paths even if they are matched there might be some variation when you actually fabricate it on chip and that might result in the mismatch in this Skew and Jitter in reality. Similarly one example which people have been facing is the variation in planarization process.

The deviation, the widths of the wires and line spacing so though you think your line is like this internally the line maybe something like this. So here the width is large and therefore the resistance value here falls down and your capacitance value rises up. So you never know what is happening across the interconnect here in reality and as a result your Skew will be or Jitter will be there in that case.

Environmental variations are the most important contributors to the Jitters and Skew variations for example if there is a temperature gradient or if there is a variation in the power you automatically get a higher Jitter because at larger temperature variation implies at that particular point you will have large changes in the resistances of the wire which you are using as a result your Jitter will change and that is an important part which people play.

The fifth part is basically as I discussed with you earlier, power supply rejection of power supply variation. This is a major source of clock Jitter in circuits.

(Refer Slide Time: 16:51)

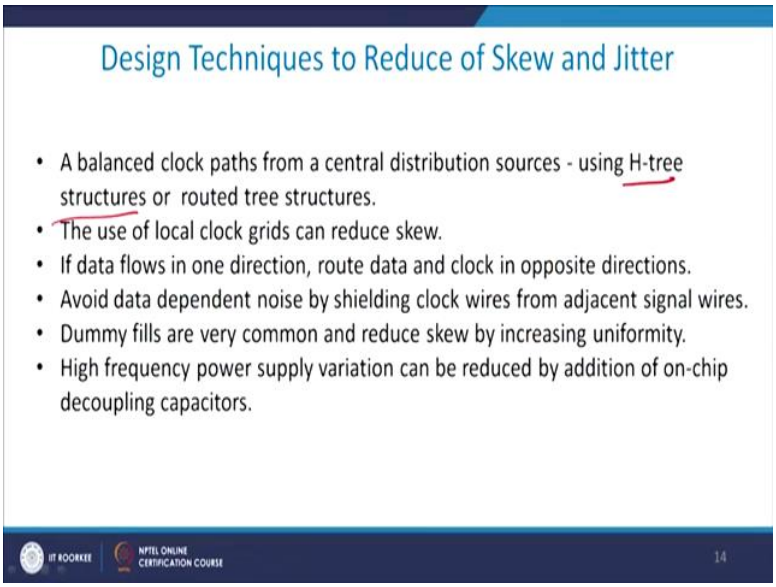



Similarly the delay through buffer is a strong function of power supply we have studied this when we were discussing FO4. Similarly the buffer delay between any 2 path, so I would delay here and I am putting buffers here.

Even if the buffer lengths are equal between 2 paths we will see that, even if the buffer lengths are equal it might be true that this path has got a higher delay as compared to this one because here it has to drive an external load which is of high value as compared to a load here. So your CL value or logical effort is typically very high in this case as compared to this case and the delay therefore is larger.

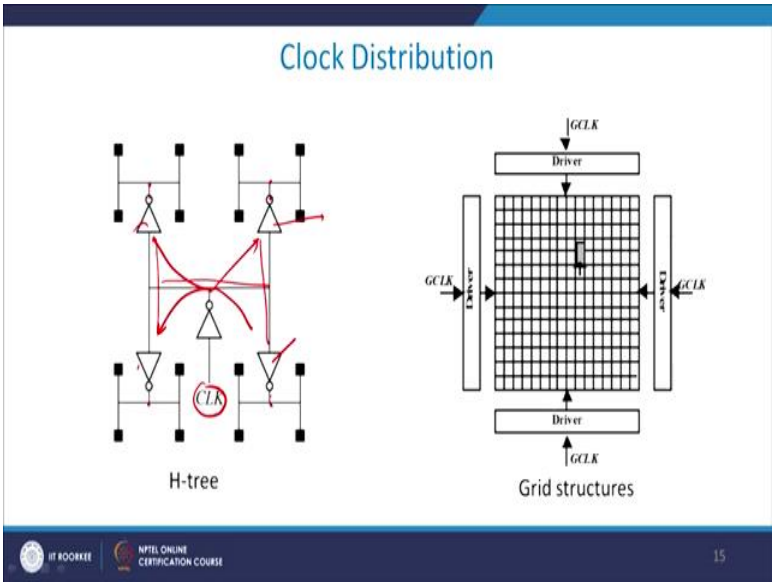
Capacitive coupling. Capacitive coupling basically means that you have 2 adjacent lines carrying the data, right? Then there will be coupling between the clock lines and the adjacent signal wires and this will vary the overall gate capacitances and this results in clock Jitter in a sense. Instantaneous loading, capacitive loading goes on changing because you do have a talking between the 2 wires very close to each other, so you have a problem of electromagnetic interference in the sense.

(Refer Slide Time: 18:03)




Okay how to reduce the Skew and Jitter. The best option available to you is try to make the clock as symmetric as possible and one of the example is basically H-tree.

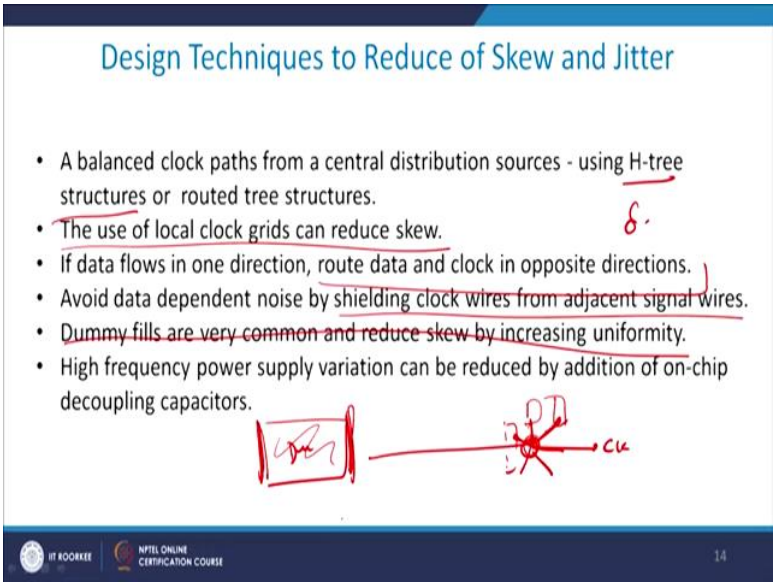
(Refer Slide Time: 18:16)




H-tree, I may have a diagram, yes. This is the H-tree which you see, so I have generated a clock here and using an inverter I generate the clock and then this is H. so this is your H and therefore this distance, from this point to this point, this point to this point, this to this and this to this are exactly same. Assuming that the delay of these inverters are exactly the same.

The clock will reach at these points exactly the same instant of time and this is known as a clock distribution for H tree clock distribution.

(Refer Slide Time: 18:58)




Now another methodology which people have adopted is that rather than routing clock from external source down to the whole register modules, what we can do also is that, let me route a clock from here to here and then locally if you have large number of blocks here, here, here then you route the clocks from here, so you make a hub and a spoke. So this hub will be the basic clock and from this clock you will have…

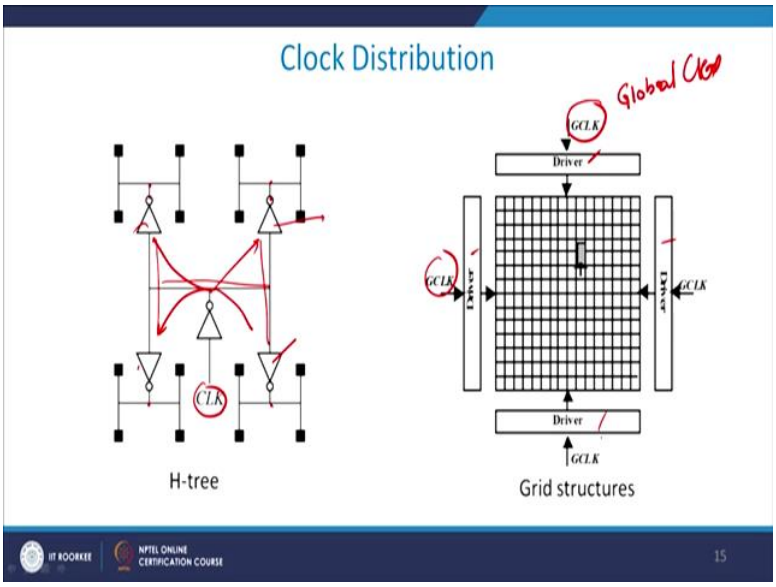
Now since this distance were very-very low or very-very small they are effect on Skew will be relatively small as compared to having 6 different paths from external clock to this path it is always advisable that you should have a single path terminating near the blocks and from there you route smaller wires to these clocks in that way you can almost eliminate 80 to 90 percent of this Skew.

As I discussed with you if the data flows in one direction route the data and clock in opposite direction you automatically get a negative Skew. Once you get a negative Skew you are an advantage in the sense. Similarly you can shield the adjacent wires as I discussed with you, you can shield the clock wires from the adjacent signal wires this will result in noninterference and therefore you will automatically get better reduced Jitter.

One methodology which people adopted is at layout level was Dummy fills. Dummy fills are primarily, the areas where you have blank spaces, you have a silicon what you do is, that you fill there with some dummy structures, so that between 2 edges there is less amount of interference. So let us suppose I have got one structure here, another structure here this is empty space I can have easily interaction.

What I do? I feel this up with a dummy fill in that act as a very good logic breaker between the 2. You can add decoupling capacitors we need not worry about it and try to make your $V_{DD}$ as stable as possible in order to reduce the Jitter problem, so I discussed with you this is the clock distribution of an H tree.

(Refer Slide Time: 21:15)

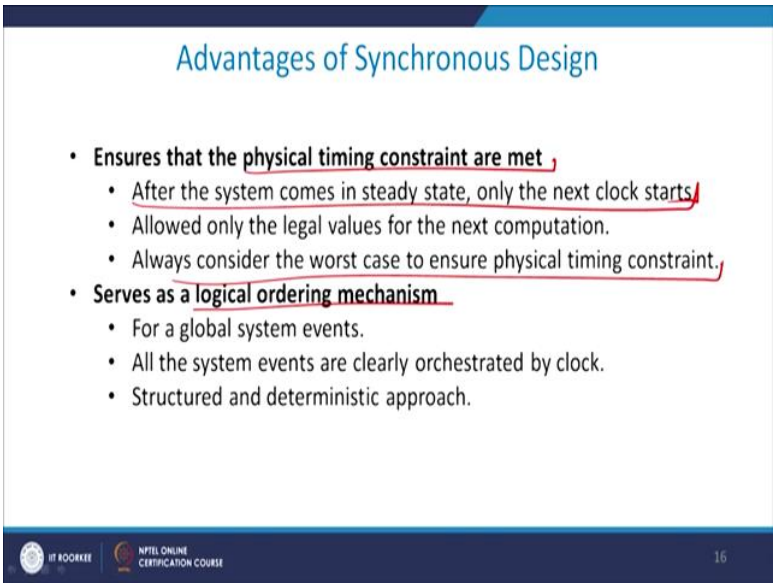



So what I do is, so GCLK is basically my global clock and this global clock actually drives the drivers here which drives the local clock. And this local clock distance is from each of the blocks are so small that you effectively do not get any delay as such.

So let me come to the last part of our talk and that is what is the advantage of synchronous design. This ensures that the physical timing constraints are met. Of course this is very very important. And second part serves as a logical ordering mechanism which means that whenever a data is being processed, then the data should be processed in a serial fashion.

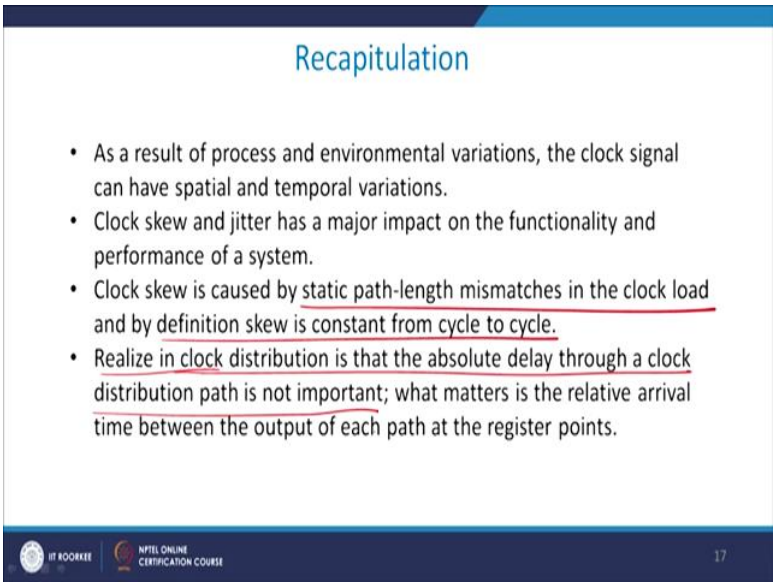
Because the first data gets processed and then second grade of data, third grade of data that is can be only done if you do a synchronous design because the clocks will be responsible for feeding the data to the register and the combinational logical blocks.

(Refer Slide Time: 22:12)




And the first one ensures to me that whenever you have clock you can store information till a point when the next clock arrives, so after the system comes in steady-state only the next clock starts and that is pretty important. Third thing is that you always consider the worst-case to ensure timing constraints. So you add $t_{cq}$, you add $t_{logic}$ you add $t_{setup}$ and the total delay your global clock should be at least larger than that. The time period for that and therefore the frequency should be less than that in any case.

(Refer Slide Time: 22:47)




So let me recapitulate the whole lecture which we did for the 2 parts of clocking sequences. We saw that what is the clock Skew, we also saw what is the clock Jitter, they effect the

functionality and performance of a system. Clock Skew as I discussed with you is caused by the static path-length mismatched in the clock load and by definition clock Skew is constant from cycle to cycle.

Whereas in case of Jitter it is not it varies from cycle to cycle. Clock Skew is positive if the data and the clock move in the same direction whereas it is negative if the data and clock moves in the opposite direction. Realize in the clock distribution is that the absolute delay through a clock distribution path is not important. It is more important that what is the relative separation in time domain between 2 clocks edges energizing my output and that is very very important in each case.

So this is what we have done and that takes care of the whole module as far as sequential logical block is concerned. I thank you for your patient hearing. Thank you.

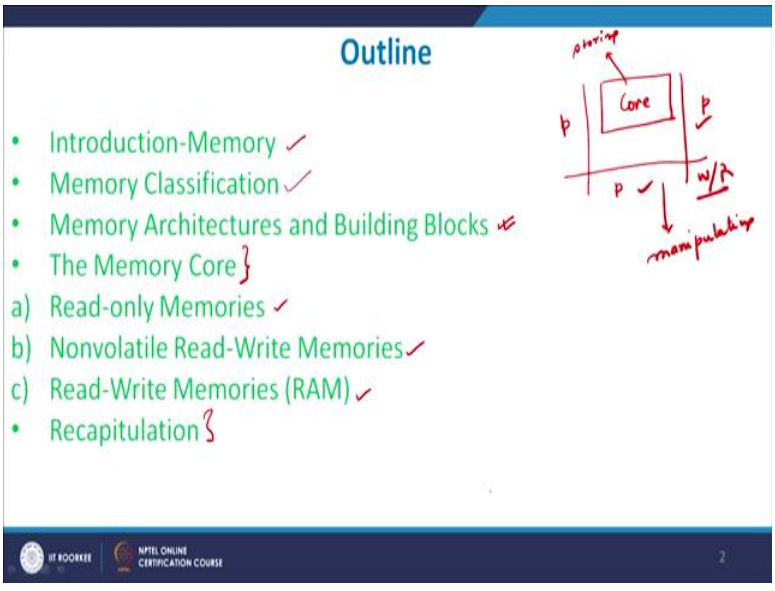
**Microelectronics: Devices to Circuits**
**Professor Sudeb Dasgupta**
**Department of Electronics and Communication Engineering**
**Indian Institute of Technology Roorkee**
**Lecture−60**
**Memory Design**

Hello everybody and welcome to the NPTEL online certification course on Microelectronics Devices to Circuits. This is the last model of this whole course, by this we will be finishing this module or this lecture series on microelectronics. The module name is Memory Design and therefore we will be looking into the last section of the digital integrated circuits and one of the applications is usage of Digital Integrated Circuits from Memory.

So let me show you what is the outline of the course, the outline of the course is that we will be first introducing to you the memory itself and then looking into memory classifications, a basic building block and its architecture and then we will look at the memory core. Core means basically whenever you actually look into a memory typically we have a core here.

This core is the area where we are storing information and you will have peripheral here, you will have peripheral here which will be responsible for extracting or writing information within the core itself.

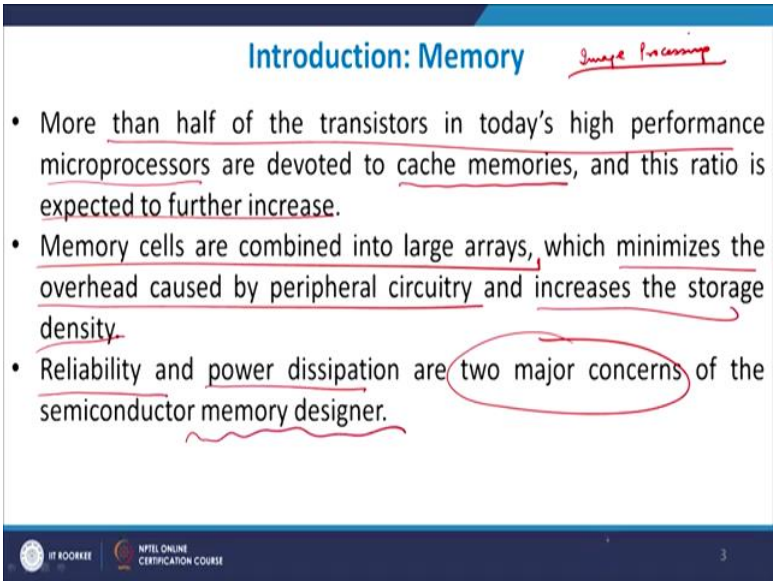
(Refer Slide Time: 1:31)




So the core will be actually storing the data and the peripherals will be responsible for manipulating the data which means that should be able to read or write the data. So primarily read and write will be responsible for the peripherals, so we are looking into that. Then we

will look into ROM read-only memory and nonvolatile read-write memories and then read-write memory as well and then recapitulate the whole thing and that is what the whole structure is all about.

(Refer Slide Time: 1:57)




So let me introduce memory here. So more than half of today's transistors when we use in a microprocessor the memory part of it are devoted to cache memories and this is the ratio expected to increase which means that the cache memory is going to increase. The reason is that the computation is also increasing and you need to store large amount of data. I will give you an example.

Whenever you are discussing for example let us say and image processing algorithm, then you are actually taking an image and then breaking it down into pixels and those pixels are getting processed at every stage. So if your image is basically high-density image, the number of pixels will be very large and therefore you require a memory to do that and therefore it is very-very important that the performance of the memory is kept at the optimal level.

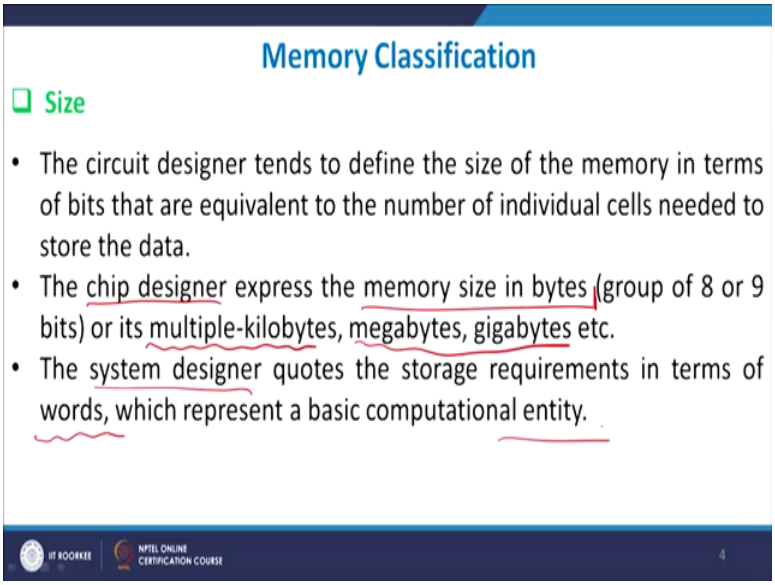
What we try to do is that memory cells are combined into large arrays, so it is an arrayed architecture and why is it so? So that it minimizes the overhead cost by peripheral circuit. So which means that if I have one core and one peripheral then it becomes very non-robust architecture because if there are N numbers of memory core storing N number of bits you require N number of peripheral circuitry.

And that makes life very difficult therefore what people have done is that your core memory and the peripherals are therefore multiplexed across the whole memory. And this increases

the storage density as I discussed with you. 2 important problems which are there in a semiconductor design is primarily, first of all is the reliability and the second is the power dissipation.

And these are the 2 major concerns of memory designer, reliability when I say, I mean to say are we able to reliably retrieve the value of the data from the memory court to the external port or you are able to reliably write down the value of the external data onto the memory, What is your excess time? What is your time taken to retrieve the data? So on and so forth. Let me come to the memory classification where we define that one core of memory will store maybe one bit of data, so if you have a 1 memory core we will see that later on we will store one bit of data.

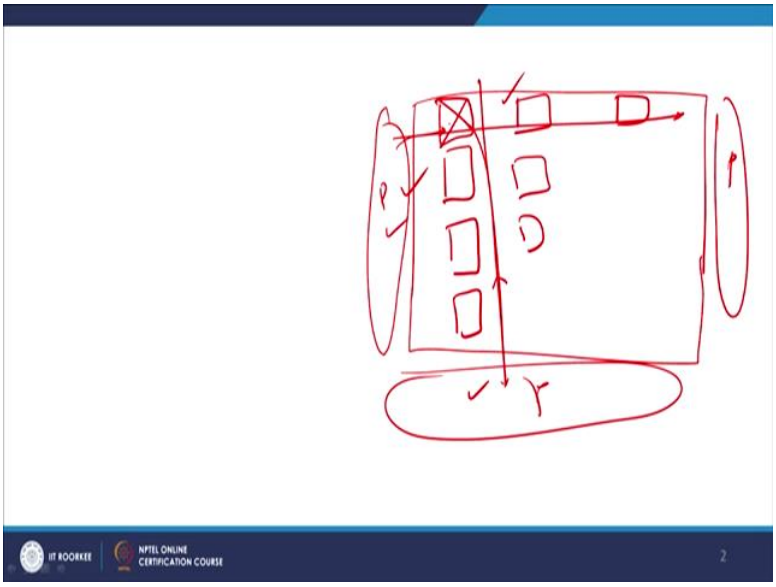
(Refer Slide Time: 4:22)




Typically chip designers express memory size in bites so group of 8 we represent we can also represent in terms of multiple kilobytes, megabytes and gigabytes right the cache memory. Where a system engineer will, system designer will be storing in terms of words, how many words you are storing which is a basic competition entity for the memory in a sense. So these are the few areas in which we but for our purposes as an electronic engineer or designer we generally referred to as bits, so I will have 64 kb memory can have up to 56 memory kb, I can have 512 kb be memory and so on and so forth.

And therefore the array of the architecture which you require should be quite large in dimensions. Typically if you look at the array, it is typically square array, so if you want to generate a memory core.

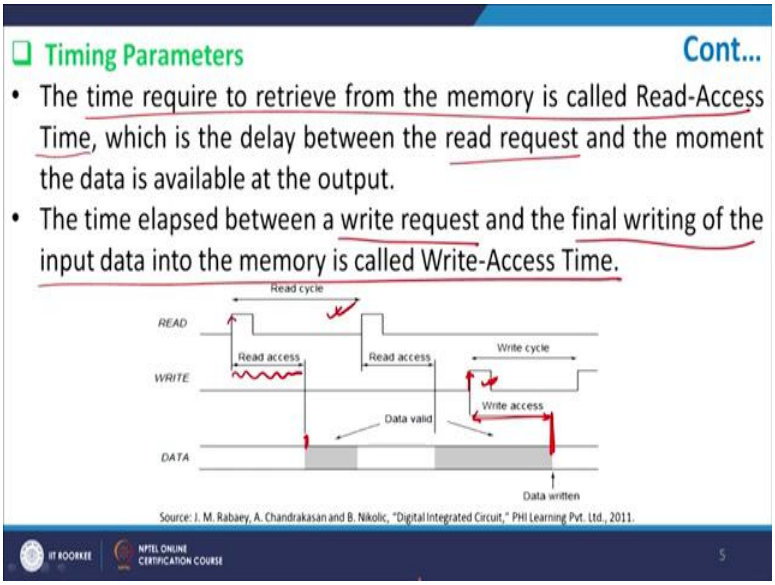
(Refer Slide Time: 5:11)

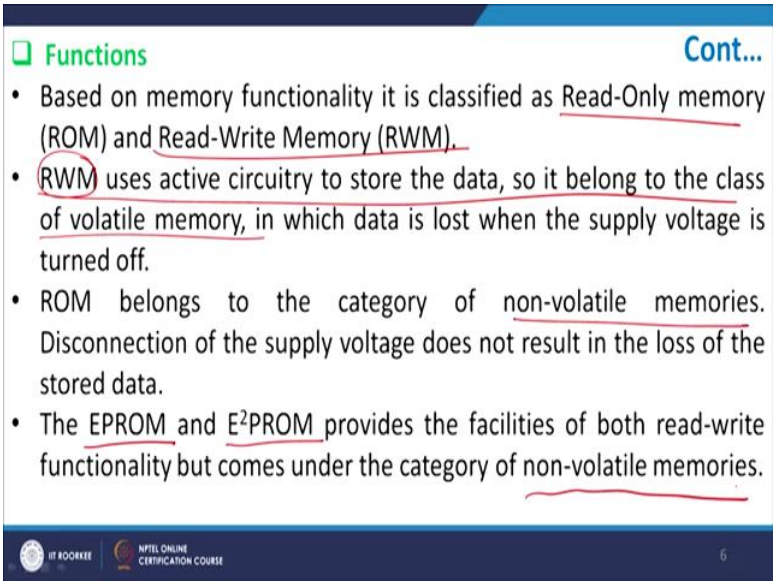


The memory core maybe something like this big and this should be as square as possible. The aspect ratio should be as close to 1 as possible and the memory elements are all arranged in this fashion that is the general trend which you follow And this is the fashion in which you and then you will have peripherals here and you will have peripherals here and you will have peripherals here.

Then you will be recalling the data and you will be writing the data, you will be recalling this data and this data. So by this peripheral you will be switching on this say cell and by this you will be switching on this. So by using these 2 lines I can select any particular cell and generate right a data or read a data.

(Refer Slide Time: 5:52)






Now typically the time required, so there are certain timing parameters. The time required to retrieve from a memory cell is defined as read access time. What is read access time? Already data is there in the memory core you send a read request and there is a certain time interval between the read request and the data coming to the output that difference in time is referred to as read access time or we say RAT.

So if you see, these are the read signals which you see. So the read signal goes high means you have sent a logic or you have sent to the signal to read the data from the memory core and the data has been valid only at this particular point. So the data has actually appeared in the output. So in time domain this is basically my read access time. Now the time elapsed similarly between right request and the final writing of the input data into the memory is defined as write-access time.

And therefore you see when I give write access this is my write going high and the data is written at this particular point till this much point then we define that to be as the write excess time. So this is the right cycle and this is the read cycle, so the idea is to make the read and write cycle closer to each other which means that the frequency of operation will increase and you should be able to do that.

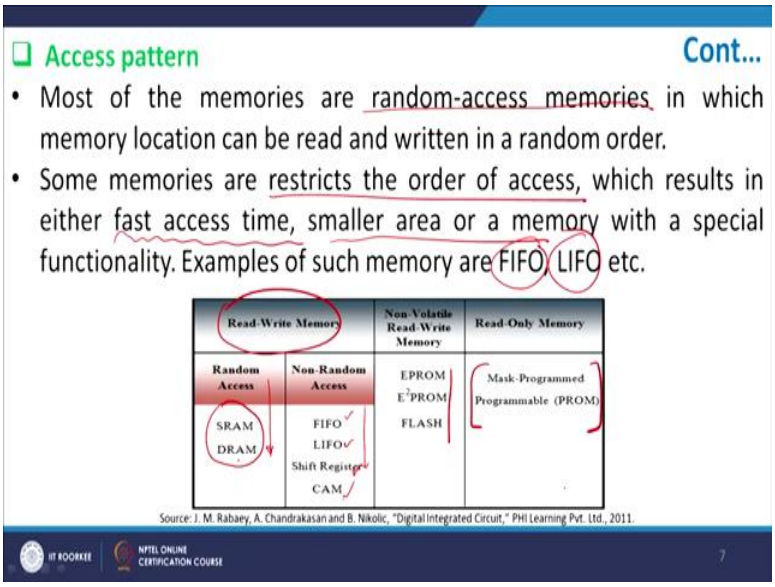
But the problem is you cannot reduce the read and write cycle beyond particular limit and the reason being they will be always interconnect delays and so on and so forth which restricts the high frequency operation of the memory. So basic timing we have understood, we have also understood the basic concept of the size of a memory. Let me excellent you based on memory functionality what are the issues available to us.

So our read-only memory and we have read-write memory. Read-write memory uses active circuitry to store data. So it belongs to the class of volatile memory. Volatile memory is what? When you switch off the supply voltage the data is lost and we refer to that as a read-write memory as such. Whereas ROM belongs to the category of nonvolatile memory which means that the disconnection of the supply voltage does not result in the loss of the store data.

So for example whenever you switch on a computer or any of your devices for example even your mobile phone, tablets and laptops the operating system, the OS is actually loaded on the ROM Because you do not want the OS to be installed every time you switch on the system, it should be already pre-installed in ROM. So even if you switch off the power, when the power is gone it is stored there and when you switch on the power the OS gets activated from ROM and you are able to do it.

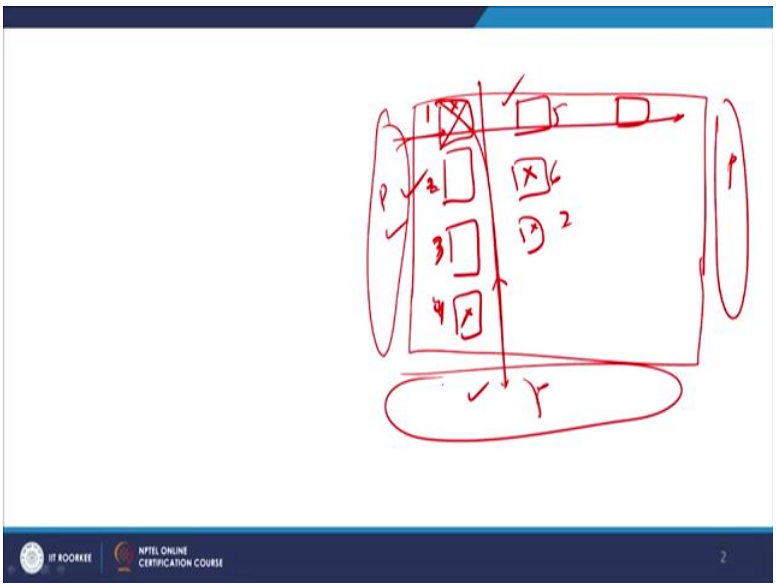
There are 2 types of other one and this is EPROM electrically programmable ROM and electrically erasable PROM programmable read-only memory and these are actually examples of nonvolatile memory and these are the few examples which are there for the memory.

(Refer Slide Time: 8:51)



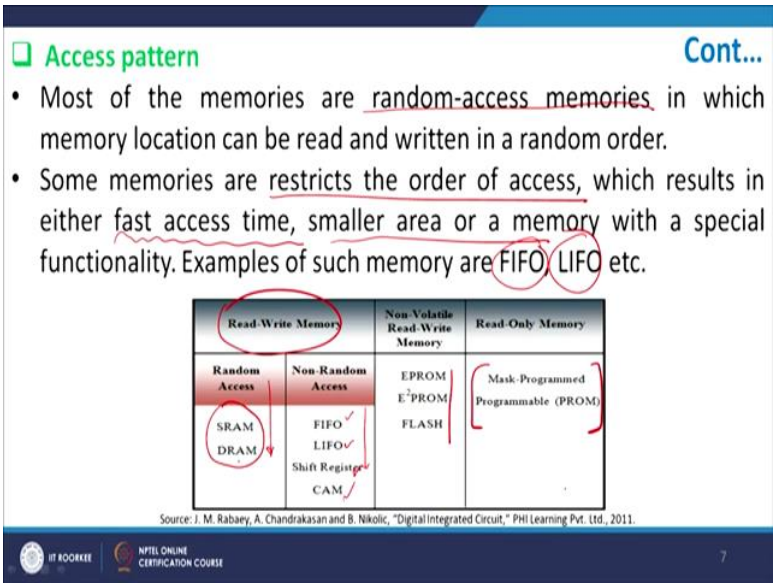

Now how do you access the memory and that the most important part as far as timing and frequencies are concerned. Most of the memories which you will study is the random access memory means which is basically that I can read or write any cell in a random order.

(Refer Slide Time: 9:14)




It is not that for example the diagram which are have drawn suppose this is cell number 1, 2, 3, 4, 5, 6, 7 it is not that I will only be able to read 1 then 2 then 3 then 4, no. it is all random, so I can even read 6 first and then I come to 1 and then I come to 2 and then I come to number 4 and so on and so forth. So this is the random access operation of the memory.

(Refer Slide Time: 9:30)

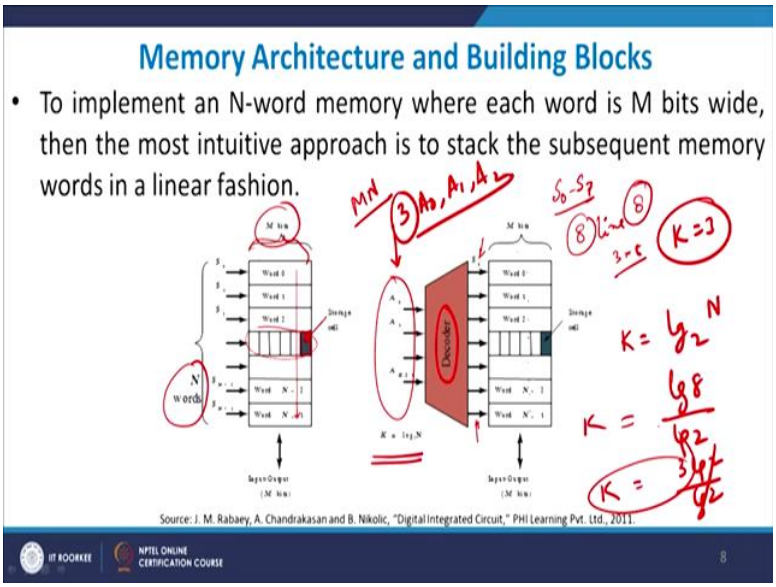



However there are certain memories which actually restricts the order of excess but then the game which you get out of it is, you get a faster access time and smaller area. For example FIFO and LIFO are some of the examples. So FIFO is first in first out and LIFO is last in first out. So if you look at read/write memory which is basically your volatile memory.

We have 2 we have got SRAM static random access memory and DRAM which is dynamic random access memory. So this is a random access and in the nonrandom access we have FIFO, LIFO, shift register and CAM. So content addressable memory, now you see for example shift register. You can have 4 type of shift register. You can have parallel in, parallel out, PIPO, you can first in first out, you can have first in last out, so on and so forth.

So there are 4 types of shift register which are all giving you a nonrandom access to you whereas in nonvolatile memory which is there, we will have EPROM, flash and EEPROM then read-only memory will have ROM and PROM. This is PROM which is programmable read-only memory and these are all read-only memory are obviously as I discussed with you non-volatile memory.

(Refer Slide Time: 10:59)




Let see what are the basic memory architecture because people have been doing it for quite a long time. And what people have done is, that I can just show you here is that you store information in terms of bits and these bits are stored in this word. So there are storage, so you see this mart here. Your brown colored mart is basically a cell which stores one bit of information.

And similarly for 8 bits of information we define this to be as one word here. And similarly you will have large number of words between this point to this point. So there are m bits, let us suppose there are m bits and there are n number of words. So there are n number of words with each word having m-bit and you can have therefore M into N is the total number of bits which you can store in this case.

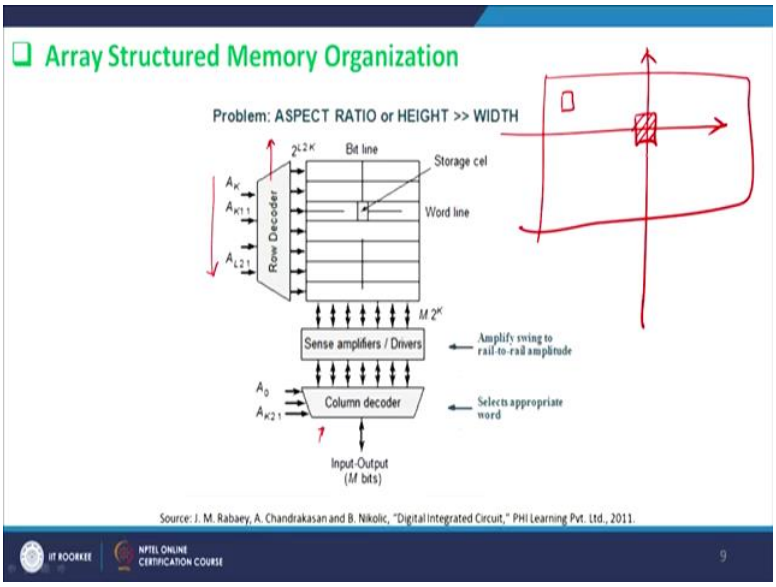
So what people do is that, they put a decoder here in the input side. What does decoder do is, that it basically tries, so if there are say 8 number of select lines $S_0$ to $S_7$ Say there are $S_0$ to $S_7$ select lines you are choosing which word you want to read or write then there are 8 lines which is available to me then I require a 3 is to 8 decoder. So there will be 3 address lines which you will be inserting.

So you see K is basically your log N to the base 2. So K is equals to log N to the base 2 Where N is the number of word which you see. So if the number of word is say 8, so log 8 you will get, so K will be equals to log 8 by log 2, so this you can break down in 2 to the

power 3. So this will be 3 log 2 by log 2 and therefore log 2 gets cancelled out, K equals to 3. So I get K equals to 3 means you require 3 words here.

3, A0, A1 and A2 as the decoder input bits, so you require this into consideration.

(Refer Slide Time: 13:06)

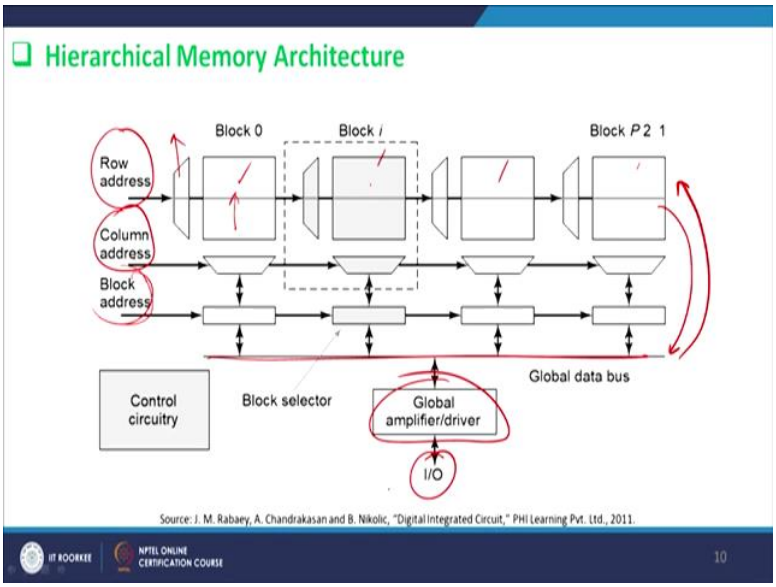



Similarly, typically you can have Row decoder as I have discussed with you with values of A here and you can also have column decoders which is shown here and this column decoders will be attached with sense amplifiers, you do not have to worry what is this but column decoders and Row decoders allow you to choose a particular bit within the array structure of the memory code.

So you have an array structure of memory code. So what I am trying to tell you is that I have any array structure of memory port and let us suppose I want to choose this then I will switch on this column, this row and this column. When I choose using the address decoder I am able to choose this one now I can do reading and writing. So decoders helps you to choose a particular cell and then I can read and write over the particular cells.

Cell amplifiers are drivers which are primarily responsible for making the circuitry fast because you want to read a cell, so I will require that the voltage should come to the output side at the very fast manner and that is required. It also allows you to give you a rail to rail swing. So $V_{DD}$ to 0 swing will be available through this sales amplifier design. This is the hierarchical memory architecture, if we go step-by-step, hierarchy.

(Refer Slide Time: 14:18)




Then we will see that we have Row addresses, we will have column addresses and therefore this row will have Row decoders here and therefore they will be Block 0, Block 1, Block 2, Block 3 so on and so forth and these column addresses will be responsible for taking the column of the block, particular block I will insert the block address here and these will be responsible for talking with these blocks for extracting it.

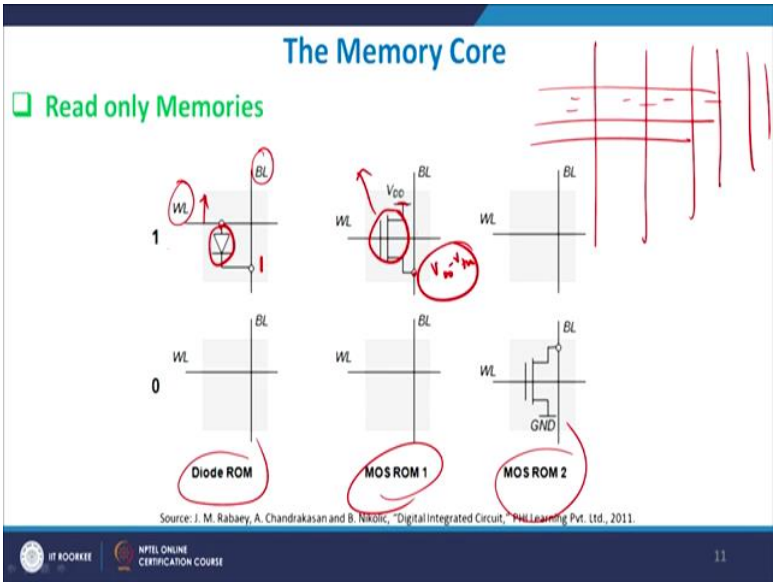
Then the global bus will be responsible for extracting the data or writing the data onto the system which will be driven by this global driver or amplifier and of course there will be control circuitry, there will be a clock which will be driving it and so on and so forth and it will be finally governed by the IO. So if you look very carefully to the whole hierarchical architecture, it is governed by multiple IOs here.

Then you will have amplifier here and then you will have a global data bus over which the data is being fed from external world into the memory or from the memory into the external world and then you have got Row addresses, column addresses and block addresses to ascertain from where the data is to be written or extracted. Say you want to go for the seventh block Ward number 3 bit number 7, let us suppose.

Then these 3 information will be fed by a block, column and row addresses and then you can fetch the data from the memory architecture.

Let us look at the memory code this is only the read-only memory which is basically a nonvolatile memory which I have discussed with you.

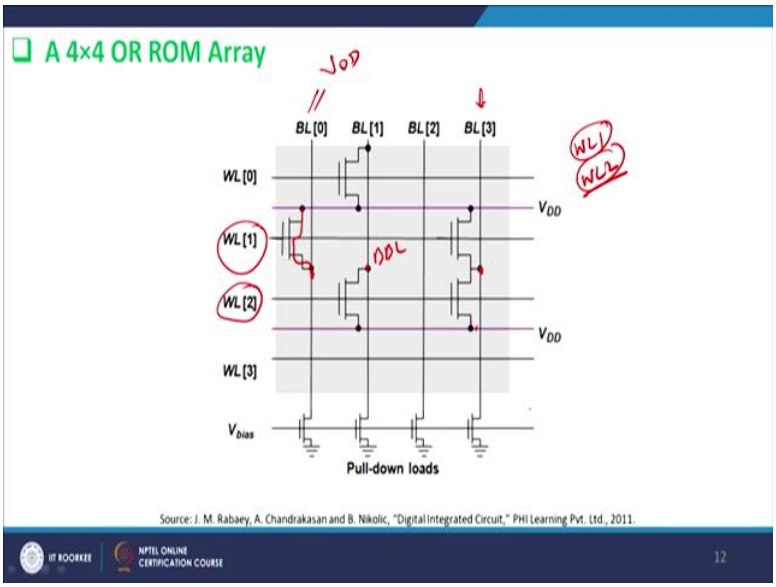
(Refer Slide Time: 15:46)




And therefore for example if the word line is high these diode switches on and one appears on the bit line a very straightforward and simple. Similarly when the word line height this MOSFET switches on, this $V_{DD}$ appears here at a particular point of course there will be one threshold voltage drop of the transistor but that is immaterial at the stage. Similarly, so this is a diode based ROM.

So I have a bit line here, I have a word line here. So word line runs like this for selecting a particular row and the bit lines are responsible for selecting the particular column. Similarly you will have MOS ROM 1, so this is the MOS ROM 1 and this is again the MOS ROM 2 where you have ground and bit line and so on and so forth you can have various architectures with the memory core.

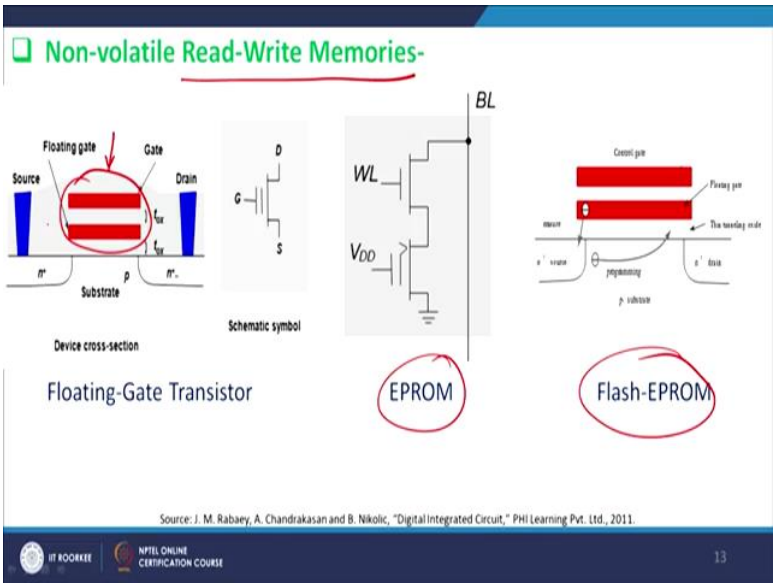
(Refer Slide Time: 16:31)




Let us look at the 4 by 4 ROM array. In this ROM array it is all MOS based ROM array and therefore let us suppose your WL[1] is high then what it does is, this $V_{DD}$ through these appears on bit line, so your bit line equals to $V_{DD}$. Similarly if your WL[2] is high then I ensure that $V_{DD}$ is written onto BL[1], similarly if WL[2] is high then I will ensure that this $V_{DD2}$ is written on this bit line

Similarly if WL [1] is also high and WL [2] is also high, 2 word lines really they become high, if either of these 2 lines are high then I will get a high-end BL [3] and there what I do therefore is, I can therefore extract the information at particular cross-section for the output world and that is what we have been doing in this case.

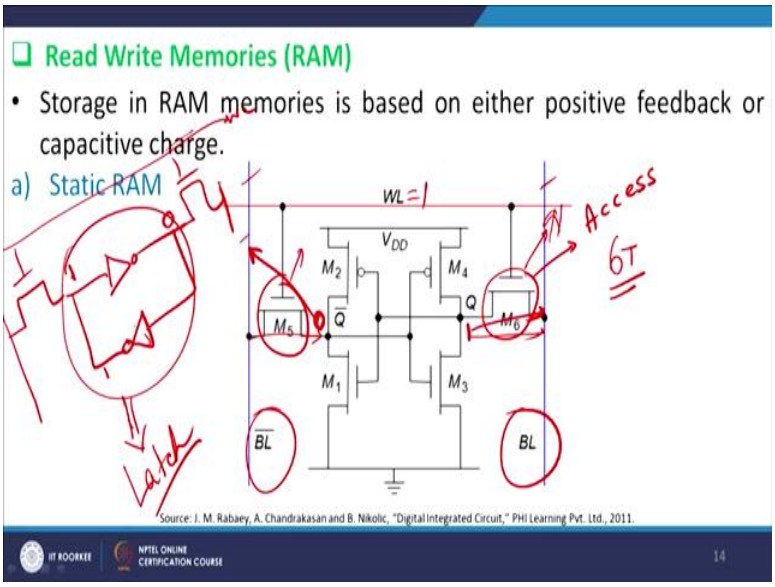
(Refer Slide Time: 17:23)




Let us look at the nonvolatile read/write memories. Nonvolatile is switch off your memory will be lost. You'll be losing its data across it. One of the example is the floating gate and then you have an EPROM and you do have a Flash-EPROM. I will request you to study all these if you are interested in this area study, good book on memory as far as these read/write memories are concerned.

We do not have time to go into details of each memory architecture to give you an idea but these are actually your nonvolatile memory so if you switch it off even, so for example this floating gate. If you look at the floating it whenever you switch it off the data is actually stored in this floating gate as a capacitor, so ideally it stores data for infinitely long duration of time and that is the example of a nonvolatile read/write memory.

(Refer Slide Time: 18:18)

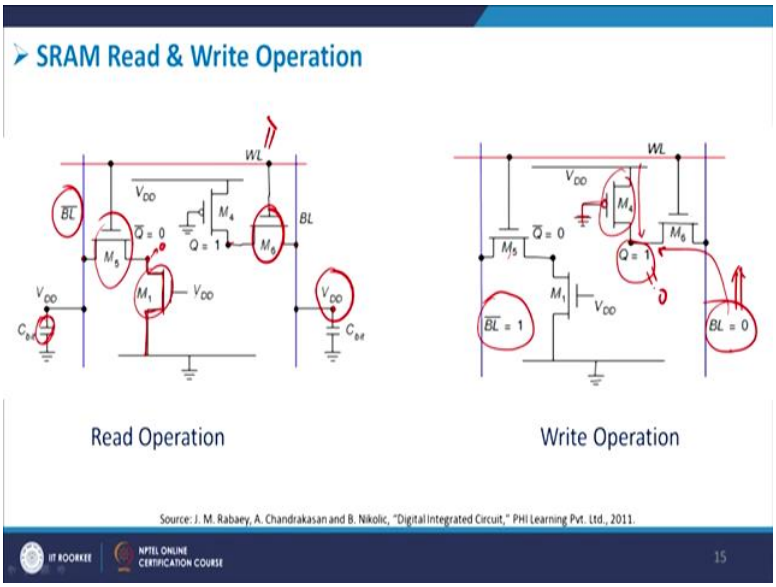



Let us look at the static RAM and one of the first example of a static RAM is basically your 6T cell or 6 transistor cell. So if you look very carefully this $M_5$ and $M_6$ are referred to as access transistors and this is basically cross coupled inverter so if you just do one small job here it looks something like this, it looks something like this. So this is your word line and this is your bit line and bit line bar. So this is your bit line bar and this is your bit line.

If I store 1 here automatically 0 will be stored here. 0 will feed into this inverter make 1 here. So this behaves as a latch which means that I am showing one bit of data here in this latch right? Now if I want to write or read a data I just have to switch on my WL. When I switch on WL equals to 1 then $M_5$ and $M_6$ switches on and then this bit line and bit line bar or bit line and bit line bar gets access to the value of Q and Q bar here.

So if it is 1 here, your hundred percent sure 0 will be here. The 0 will appear here and 1 will appear here. So if you want to read a cell, you have written a cell, suppose Q equals to 1 and Q bar equals to 0. Simply make the word line high, once you make the word line high the excess transistor switches on because it is NMOS transistor. As it switches on the Valley of voltages go onto bit and bit bar line and you are able to therefore read the data from that particular cell.

(Refer Slide Time: 19:50)

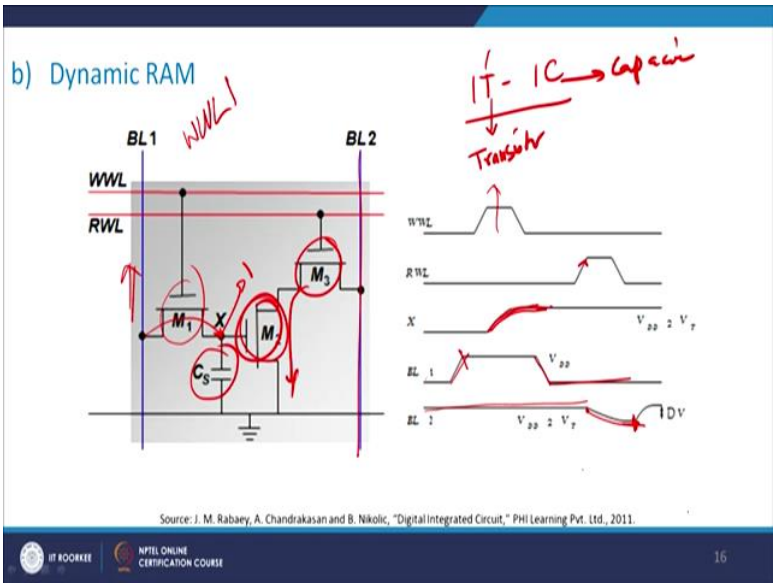



As I have discussed with you the read operation is something like this that you have Q here word line goes high, switches on this $M_6$ and you are able to transfer this Q onto bit line bar with the one voltage, threshold voltage drop, this appears across this therefore you have a $V_{DD}$ appearing here. Whereas when you have bit line bar getting activated this goes on, $M_5$ goes on.

Since Q bar equals to 0, so I have pulled on getting activated because $V_{DD}$ was attached to it. This is on and therefore this goes to 0. When this goes to 0, 0 also appears as bit bar line and it is stored in this C bit here, fine. So you have a C bit stored here and that is the read operation. If I come to the right operation what we try to do is at the same exactly the same as the previous case the only thing is that you are already bit line bar here 1 and bit line equals to 0.

You again switch on WL because you want the access transistors to be on and then you allow the 0 to be written through this arm. So you had initially Q 1 0 here, now if this goes to 0, if it goes to 0, this is switched on and this $V_{DD}$ appears here, so you have to make it bit line voltage slightly higher to overwrite those value and therefore this will be converting into a 0. So this is the right operation for SRAM memory.

(Refer Slide Time: 21:01)




Now another example is basically my DRAM which is dynamic RAM and it is consisting of also referred to as 1T-1C design which is one transistor, so this is one transistor and one capacitor and therefore if you see very closely even WWL is high, which you see, this is high. This high means $M_1$ is switched on, when $M_1$ switches on The bit line is suppose 1.

The bit line is 1 and bit line to, so I have got bit line 1 is high at this stage and bit line 2 is remaining constant, let us suppose and the RWL which is the read line has gone now high so it has high here, so when WWL is high it switches on $M_1$ and therefore the value of voltage here starts to rise, why does it start to rise is that, initially if your BL 1 is high and your $M_1$ is switched on this voltage will be written on this X and CS will gets charged and that the reason you see an exponentially rising function at X And that gives you arising function.

Now let us suppose you want to read a cell, so read this particular cell then you make it read high, once you make it read high then let us suppose the bit line has gone down then you will see that at this particular point bit line 2 which is this one Bit line 2, so when your reading goes high this $M_3$ switches on you had stored data here one, this will switch on your $M_2$ and what will happen is, this voltage will fall to 0 and therefore you see that your bit line 2 is going down to 0 And that way you can actually have the operation of dynamic RAM.

So dynamic RAMs are therefore responsible for 1T 1C cell, so this capacitance CS is able to store the data till the next read or write cycle comes into picture So this is what the basic timing diagram of this is.

Let me recapitulate for what we have done for this memory design. You can actually look into there are certain very good books on memory, you can have a look into it. We can discuss it on discussion forum also regarding the books and open sources from where you can get this data. What we have discussed is that, we have discussed the memory time, what is the meaning of various timing definitions, look at the access pattern and the application.

We also looked into the fact that we have looked into the basic read-write memory, volatile memory and ROM which is basically a nonvolatile memory. Finally we went for RAM and we looked into 6T cell, 6 transistor SRAM cell and we looked also into 1T 1C DRAM cell which is basically a volatile memory and example of that and how to read and write an information from a cell

So these basic concepts I have tried to make clear in this particular module, I hope you have enjoyed this module and we look forward to meeting you during the NPTEL course structure when this is released. Thank you very much.