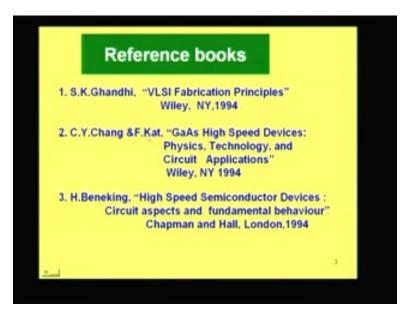
High Speed Devices and Circuits Prof. K. N. Bhat Department of Electrical Engineering Indian Institute of Technology, Madras

Lecture - 1

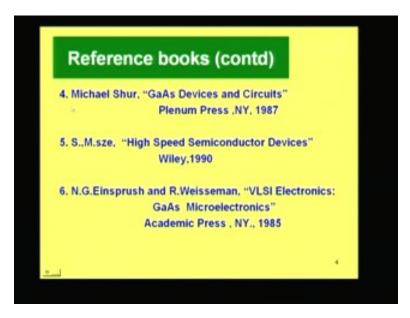
Introduction to Basic Concepts

Good morning. Today, we start our discussion on a series of lectures on high speed devices and circuits. My name is K. N. Bhat from the microelectronic section in the Electronic Engineering department, Indian Institute of Technology, Madras.

(Refer Slide Time: 01:36)



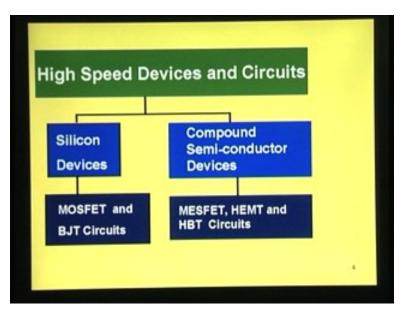
The reference books I think I can hand it over to you later. Just see, one of the books that I am referring constantly for material properties will be: S.K. Gandhi book on VLSI Fabrication Principles, second edition 1994, there is a book by Chang and Kat Gallium Arsenide high speed devices: Physics Technology and circuit applications. There is a book by Beneking high speed semiconductor devices circuit aspects and fundamental behavior. There are few things it is a very thin book it contains somethings which you do not usually see otherwise which are relevant to high speed circuits. That is another book we just can see, in fact this book is available in library. I have a copy of these two books I bought it especially for this course but, I do not know how much of it I follow but, still we will have material from various literature borrowed here.



(Refer Slide Time: 02:48)

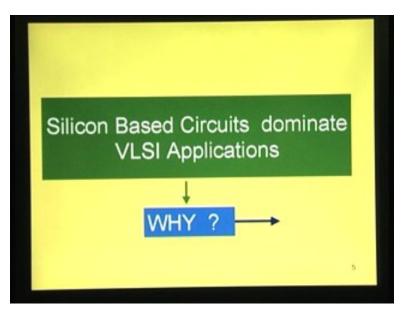
The other books which are available in library are: Michael Shur book on Gallium Arsenide Devices and Circuits, slightly old book, 1987, but, good enough for the basic studies. S.M sze, book 1990 which I think I am sure you have seen in the library for high speed semiconductor devices. Lastly, there is one the first time when I was seeing these areas I saw the book on VLSI electronics gallium arsenide microelectronics. This is also available in reference section. These are some of the books which we can use as reference book but generally, I will be giving you the summary or the list of various details related to the high speed devices as well as circuits. There is lot of emphasis on devices but, definitely circuit part also will be touched upon. I will not highlight what the syllabus is going to be, in fact, if you go to website you will see the syllabus.

(Refer Slide Time: 03:59)



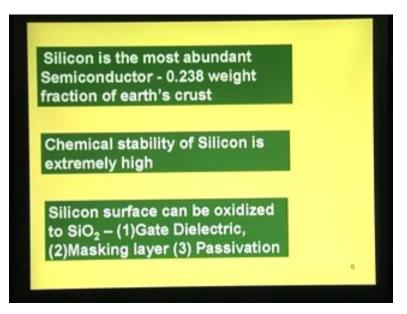
High speed devices and circuits when you say you can immediately see there are two classes of devices and circuits, one class based on the silicon devices and when you say silicon devices immediately you will see MOSFET and bipolar junction transistor circuit BJT that is very popular both in digital and analog circuits. Further classifications we will see later. Major device areas are these two. We will also see compound semiconductor devices; we will see why we should move from here to here for high speed during our course of discussion. When you deal with compound semiconductor devices, you will have MESFET that is metal semiconductor field effect transistor; **I** suggest MOSFET in semiconductor integrated circuits. Also you will have high electron mobility transistors as against MOSFET and JFET and other one will be HBT hetero junction bipolar transistors which actually is a replacement for the conventional bipolar transistors. We will definitely touch upon some of the circuit aspects. For example; some concepts of logic families involving these types of devices which are different from those MOSFET based logic circuits. Let me not name them right now because there is no point so, these are the general ideas.

(Refer Slide Time: 05:47)



Now, we will see silicon based circuits dominate the VLSI applications. First, we will have to see why it is so dominating, why it is all pervasive. Next, we would like to see why we should look for other materials. Those are the two aspects we will discuss now.

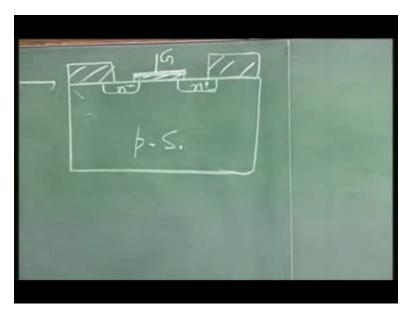
(Refer Slide Time: 06:09)



So, you notice I just put it deliberately on the left hand side so that, some space may be required for here if I shifted everything to the left slightly. Silicon is very omnipresent

because of various reasons; number one is you see silicon is the most abundant material on the earth's crust. It forms 0.238 weight fraction of earth's crust. In fact, there is a joke on this you can go to beach collect some sand that is silica and you can reduce it to silicon, so plenty of silica means plenty of silicon that is one of the major reasons why silicon is very popular. Now, apart from that, there are other reasons like the chemical stability of silicon is extremely high, when you say chemical stability is high it means, we can immerse silicon at room temperature particularly in all chemicals acids or alkalis. For example, we can dip it in nitric acid, HCL (hydro chloric acid) or even we can dip it in combination of HCL and HNO₃ that is agua regia. Even agua regia the most dreaded chemical which can eat or etch gold will not affect silicon. That is what we mean by saying it is chemically very stable it becomes very easy to process because, we can dip it in various chemicals. That is one of the major advantages that silicon has compared to most of the other semiconductor materials. Unless we require them very badly we will not go in for other semiconductor materials. Silicon surface can be oxidized to silicon dioxide this is another advantage you put silicon into or just subject silicon to high temperature like 1000 degree centigrade pass oxygen; you have immediately (Refer Slide Time: 08:43) silicon plus oxygen gives you silicon dioxide. Temperatures, we talk of are 900 degree centigrade to may be 1100 degree centigrade that is the range. So, you can easily convert it using oxygen or you can oxidize in water vapor H₂0, result is the same silicon dioxide. Silicon converted into silicon dioxide is called the native oxide. It has several uses why do we say oxidize the silicon so what, if you see these here a slide you can see this silicon dioxide is used as gate dielectric, in a MOSFET you have a gate and an oxide the gate dielectric is very effective and very powerful with silicon dioxide. Metal oxide silicon that is MOS structure you can make with the metal then, the silicon dioxide then, silicon the MOS structure. You also have other advantages like the masking and passivation let us take a look one by one. So, one is the gate dielectric Sio₂ can act as gate dielectric and second one is the masking layer.

(Refer Slide Time: 10:44)



This is like this: you have a substrate of silicon on that you can have a thick oxide and you have this that is the gate and then you have n plus do not know whether it is visible. This is actually the basic MOSFET structure source drain a schematic representation. So I am trying to say is you will have this portion which is colored that is actually the gate oxide which already you know I am just re iterating just filling whatever information you require. This is metal oxide semiconductor this is very effective as the gate dielectric. This is called gate dielectric you see the field oxide this portion that also a silicon dioxide. The second one is the masky layer for example when we do this n plus n plus diffusion you use the oxide as the mask which will prevent the diffusion of dopants through that layer. So, the silicon dioxide is a very effective masking agent against impurities that is what implies when we say masking layer. The third one is passivation layer notice here the junction is below oxide. The junction is protected by means of this oxide, so that is what we mean by saying junction is passivated that is otherwise it is very highly reactive. When we have the oxide on the junction its reactivity is reduced on surface, so contaminants do not go and reach the surface. It is a protection layer or a passivated layer. In layman's language we can say it is a protective layer in electronic term we can say it is passivate. Those are the advantages that we have for silicon dioxide. In fact, this is the major advantage of using silicon because it is converted to oxide very easily.

(Refer Slide Time: 13:31)

Energy band gap of Silicon is 1.1eV – enables device operation up to 150°C

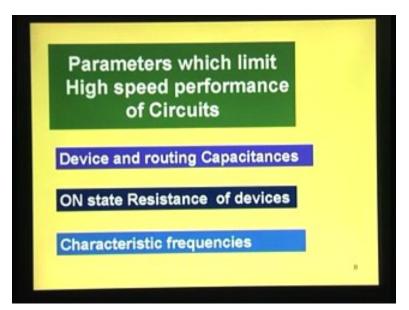
Single crystal silicon rods of 12 inch diameter easily grown by CZ method.

Other benefits of silicon in integrated circuits are the energy band gap. The gap between the conduction band and the valence band is 1.1 electron volts. In fact the energy gap 1.1 means it requires 1.1 electron volts energy to remove the electrons from the valence band and make it available for the conduction. That means sufficient amount of energy must be given to break the covalent bonds to remove the electrons from the bond. The big advantage of this is for example if we take germanium its band gap is 0.728 electron volts. That is one of the reasons, why germanium lost its market compared to silicon because this energy band gap is small. Whole electron pairs can be created very easily. So, if you make a junction, if the temperature goes up to 60 or 70 degree centigrade the junction becomes very leaky. If the junction becomes very leaky it can no longer effectively rectify junction. Result is the transistors etc., do not work properly the way you want them to behave. That is why it is one of the advantages. In fact, going back to the previous slide, silicon can be oxidized to get silicon dioxide for germanium we may ask, can you oxidize it? You can oxidize germanium where, we get germanium dioxide or germanium tri oxide but, that oxide is not chemically stable when you process it by dipping it into chemicals it just gets washed out very easily even the dilutest of dilute chemicals will etch out the germanium oxide. Whereas, silicon dioxide chemically very highly stable just like silicon the only chemical which will etch silicon dioxide is hydrochloric acid. In fact, you should be able to etch with some chemical otherwise we

cannot handle it at least there must be some boss to control that is hydrochloric acid.

Germanium oxide that is very weak, it can get etched very easily. So, on two counts germanium lost its market though the transistor initially was invented in germanium within a decade, silicon took over, because of: one, the advantage of high band gap 1.1 electron volts, two is the advantage of the oxide. The other advantage is that, we see now I am just putting it as an advantage, in the sense, since silicon was made use of for realizing integrated circuits; the vendors who sell the wafers started going bigger and bigger diameter wafers. In the sense grow a rod of silicon, slice them to get wafers, today you can grow twelve inch diameter rod that Intel etc., use only twelve inch diameter wafer by slicing them. In India, we have four inch diameter wafer BHEL semiconductor complex Chandigarh six inch diameter wafer, ITI Bangalore, Fithar, they use six inch wafers, so these are some of the companies. There are also other companies that deal with compound semiconductor. We will come back to these later because we will see whether we will need them at all. In fact there is one company in Hyderabad Getech, entirely deals with gallium arsenide base devices or high speed devices, high speed integrated circuits monolithic micro integrated circuits MMIC. That is the reality it is not a myth. So now, we know that, we can grow twelve inch diameter wafers by using Czochralski technique. That is melting silicon from poly crystal material dip inside crystal pull it to get that. This is I am sure you studied in your technology course initially. So, these are the things which have projected silicon into the forefront.

(Refer Slide Time: 18:10)

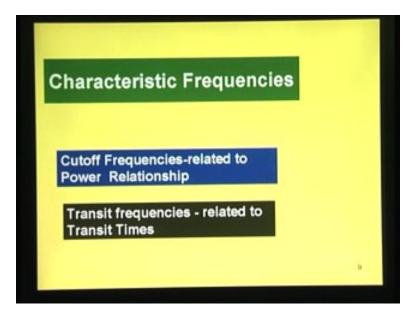


Now, let us see what are the parameters which limit the high frequency performance or high speed performance of the devices. I think without telling anything, we will say the parameters which kill the high speed operation of the devices are resistors and capacitors because, wherever resistors are there combined with the capacitors, we will have the RC time constant delays (18:46). It will affect the high frequency performance and also affect switching speeds because; wherever there are capacitors you have to charge those capacitors. If they are accompanied by the resistors they have to be charged through those resistors, so device capacitance and routing capacitance. You may have the integrated circuits where, a number of devices are there but those devices must be connected from one to another by means of inter connecting leads or wires. Those wires are not just connected like that it is routed on the surface of the oxide. Those wires will have capacitance to substrate and those capacitance are called routing capacitance or you can pronounce it like routing capacitors whichever way you like.

It depends which part of the world you are talking, if you are talking in America it is a routing capacitance. Other thing the device capacitance like junction capacitance and also when we talk about MOSFET the MOS gate itself has got the capacitance that has to be charged. Now, when it is getting charged from a preceding stage, preceding devices will be driving the capacitors and the preceding device if driving the capacitors, the ON

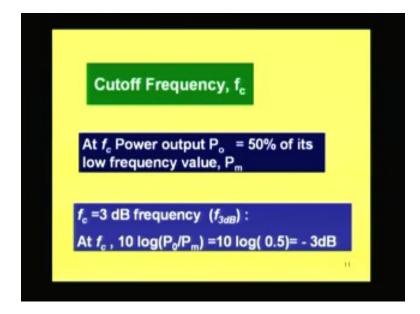
resistance of the preceding device will come into the picture that is the R of the previous device. For example, C MOS if we have the full of transistors that is the one we are charging this next stage. So that resistor is ON resistance of the device that ON resistance is the one that I am referring to here. That should be kept low; if it is high the RC time constant is lost that is when you are using the digital circuits. Now, there are some aspects which are known as the characteristics frequencies. These are of course the parameter which control those frequencies really, apart from that there are defined certain frequencies which will be the figure of merit of the device.

(Refer Slide Time: 21:01)



The characteristics frequencies of the devices particularly of a circuit, of that matter for example: The cutoff frequencies related to power relationship. Cutoff frequencies that are at low frequency get certain output power from device. Now, that power output will become 50% that is 3 db.10 log half, 10 log half is minus 3 db logarithm 2 base 10, 2 is actually 0.3 into 10, 3 db. So the 3 db frequencies is called the cutoff frequencies where the power output falls 50% 3 db point. Now (21:52), that one is evidently governed by the rc time constant of the devices and circuits. Then you have the transit frequencies. In fact, this may be slightly a different term which is related to transit time. Sometimes we see people referring to the cutoff frequencies and the transit frequencies one under the same thing they are different type. The transit frequencies are related to transit time.

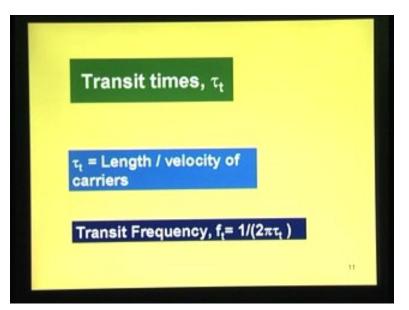
which is actually related entirely to the device and the cutoff frequencies are related to not only device capacitances plus the series resistance or shunt resistance which come across that device, we will see that.



(Refer Slide Time: 22:47)

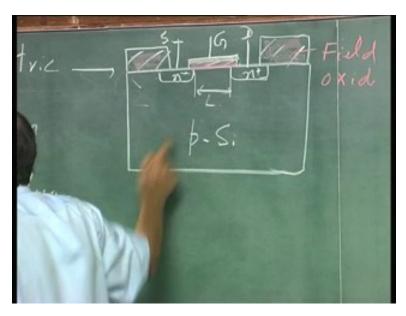
We have already defined what the cutoff frequency is. The cutoff frequency that is at f_c to define that f_c the power output P_0 at cutoff frequency is 50% of its low frequency value which is the maximum power output that you get. That is why; f_c is 3 db frequency because, at log P_0 by P_m into 10, 10 log P_0 by P_m is 10 log half that is finally minus 3 db at f_c 10 log P_0 by P_m is equal to 10 log 0.5 is equal to minus 3 db. So that is the cutoff frequency of the device. Now, what is the frequency, transit time, transit frequency?

(Refer Slide Time: 23:47)



The next thing that we will see is Transit time tau_t, this is I am sure very familiar for those of you who have taken at least one course on devices; this is property of the device. Transit time as the name itself indicates is actually the length divided by the velocity of the carriers.

(Refer Slide Time: 24:16)



If there is the length L for example if we have the channel length here this is actually the

channel length that is the channel length. The carriers get transported from the source to drain through these lengths with the velocity v. The transit time is the time required for the carriers to move from the source to the drain through the path length of l. So that is why that will be actually the length divided by the velocity and the velocity would depend upon the transport mechanism. For example, if we take the base region of the transistor there also you can talk of transit time. The mechanism by with the carriers transported which is not by drift it is by diffusion. Velocities are smaller compared to drift. So, length there is a base width divided by the velocity related to the diffusion that will give a transit time. Apparently or evidently it is related to the base width and some parameters which govern the velocity when it is drift the mechanism or the parameter which governs the velocity is when drift the velocity v is equal to mobility. Mobility of the carriers into electric field so that is called as the mu_n E where, E is the electric field mu_n is the mobility. Similarly, there is a velocity and length is L channel length. Similarly, the bipolar transistor the velocity is related to not mobility but diffusion coefficient because, the diffusion process is the driving force, driving force is the concentration gradient. The force is D_n into D_n by dfx where d is the diffusion coefficient, coefficients D_n by dfx is the concentration gradient here it is mu_n is the driving force in the electric field, there diffusion coefficients into driving force is the concentration gradient. So, it is the diffusion coefficient, diffusion coefficient is also related to mobility. What you see is now the velocity of carriers whether it is by drift or by diffusion it is proportional to, it depends upon mu_n or another parameter which is related to mu_n that is diffusion coefficient because, you know that D_n by mu_n is kT by q that is the Einstein's relationship D_n is the diffusion coefficients. Driving force in this case is diffusion coefficient in to D_n by dfx concentration gradient. These are of course the basic things which I am just reiterating.

Now, let us go back and see what the transit frequency is tau_t is the length and velocity. Now, let us just find out what happens here. Let me just go through a quick derivation of that. How much is transit time and how is it related to the term known as the transit frequency. (Refer Slide Time: 28:35)

Transit frequency is defined what it is, we will see afterwards physically meaning, if it is defined as f_t equals 1 by 2 pi tau_t where, tau_t is the transit time. Now, let us find out the transit time. This is the definition; we will see soon what it implies when you really see the equivalent circuit. Sometimes, we make a mistake of calling this as cutoff frequency itself it is the ideal cutoff frequency. So tau_t is length divided by velocity that is L divided by velocity. Now, the MOSFET, we will define this with respect to MOSFET. In a MOSFET metal oxide field effect transistor v is equal to mu_n into E, E along the channel electric field along the channel length source to drain, so that is actually equal to mu_n into what is the electric field along the channel when it is in linear region it is just V_{DS}. But once it goes to saturation the voltage drop is V_{GS} minus threshold voltage that is V_{GS} minus V_{Tn} divide by 1 that is the electric circuit voltage by voltage across the channel divide by length of the channel that is the average electric field.

(Refer Slide Time: 31:11)

Saturation

Now, in fact, I would write it one step mu_n into V_{DS} divide by L which is actually equal to mu_n into V_{GS} minus V_{Tn} divide by L when I_D is saturated (31:34). I_D in saturation gives you that what this actually is tau_t transit time is equal to L divided by V is actually equal to L square divided by L square divided by mu_n into V_{GS} minus V_{Tn} . Here, V_{Tn} is actually V_{GS} is gate to source voltage V_{Tn} threshold voltage those are the well-known definitions, we do not mention that initially but, for the completeness that is that. So, the transit time actually is this number. Now, let us take a look at the transit frequency by using this as the transit time what is the transit frequency.

(Refer Slide Time: 33:36)

Transit f

Transit frequency f_t 1 by 2 pi tau_t therefore omega_t which is actually 2 pi f_t is 1 by tau_t. Please note that is the definition that turns out to be then from here, from this equation we get this as the transit frequency. Now, let us look at manipulation. You have finally what you want to see is how this will be related to some of the circuit parameter which we talk of because, a circuit engineer may not like to say mobility threshold voltage or sort of things he may not like to analyse so let us see how. I remove this now. Notice that, what the transit time depends upon channel length and the mobility mainly, I can rewrite this as multiplied by C oxide. C oxide is the well-known term gate oxide capacitance per unit area so many farad per centimeter square in fact, the value is about 35 nanofarads per centimeter square if oxide thickness is 100 nanometers that is 1000 angstroms. If it is 1000 angstroms then that is 35 nanofarads per centimeter square.

(Refer Slide Time: 36:45)

I multiply that and then I multiply by W in to V_{GS} minus V_{Tn} I just multiplied by these two and then I divide by that. So, why did I do that, this I can write it as mu_n C oxide W into V_{GS} minus V_{Tn} divided by L just pulling out and then I will write it as L into W into C oxide. Here we should know what the various terminologies are. Let us just see that because, in case you we are using different symbols mobility of electrons oxide capacitance per unit area of the gate looking at the top, W is the channel width when I take the MOSFET. I just unfortunately rubbed of that, MOSFET has the depth and channel length, so channel length is L, W is the channel width MOSFET and gate source voltage, threshold voltage. What this quantity is L into W the area of the gate and C oxide is the capacitance for the unit area. L into W total gate capacitance, we can call it as input capacitance. This is actually equal to so L W in to C oxide is gate capacitance. I got C_i input capacitance we call C_g or C_i just called input capacitance. What this quantity is in terms of trans-conductance if you take. I_D in saturation for MOSFET is mu_n C oxide W divided by 2 L into standard square law simplified version of this MOSFET characteristic that is that square law. (Refer Slide Time: 38:58)

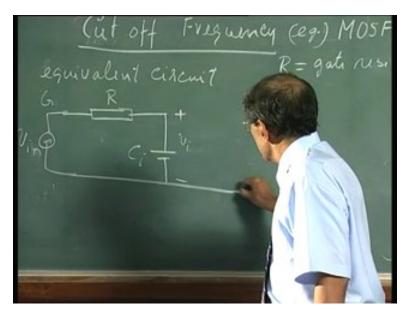
Therefore, g_m , which is trans conductance is equal to delta I_D by delta V_{GS} will be equal to differentiate that minus, so now you can see what is omega_t we defined it as 1 by tau_t and we finally come up to this and this portion we know it is equal to C_i gate capacitance or input capacitance and this portion is nothing but g_m . This actually is equal to g_m divided by C_i . In many places, we will see that, the figure of merit of this transistor is g_m divided by the input capacitance. In fact, we can show, I will leave it to you to show as exercise for a bipolar transistor also if we define the transit frequency as 1 by tau_t that omega_t will be equal to g_m by C_i that is why I put C_i to keep a general term. In a case of the transistors common emitter configuration it is the diffusion capacitance between the base and the emitter g_m divided by the capacitance that will be actually the 1 by tau_t. That is why I just kept it as C_i not C_g to keep it general. What is g_m in the case of bipolar transistors is V_t by Ic is R Ic by V_t is the trans-conductance (41:12) current divide by thermal voltage kT by q. If you have 1 milliampere current flowing V_t that is kT by q is 25 millivolts so 1 by 25 that is 40 milliamperes per volt that is the trans conductance for a bipolar.

Here, you can see it depends upon this quantity. I can have better and better value of that pi if a better value of this g_m and smaller value of the gate capacitance. These are some of the guidelines which are being used, g_m can be improved by increasing W but that are

affecting capacitance also. So, by increasing the area you will have the driving capability improved but, the capacitance also goes up that is not the way to improve the speed of the device. We cannot make it W but, we can make L smaller g_m goes up capacitance goes down. The way to increase the speed is reduce the channel length without affecting the capacitance, we are increasing the speed of the device. This is one of the key things more importantly you can improve the speed of the device if you can improve the mobility. How do you improve the mobility, for a given material we are stuck with that mobility, but mobility is better when the doping concentrations are low but, if we go to doping levels of ten to the power of seventeen of that level the mobility will start falling. You must be guarded in the conventional applications not to go to very high doping levels, we must remain at low doping levels it has its own repulsion, we will see as we go on. You need to go higher doping concentrations in certain cases, so one way is to change over the material which has better mobility or better velocity.

Now, let us take a look at the other aspects what we are telling, what the cutoff frequency is let us see that how is it different from this transit frequency. (43:52) of something is not very clear so I have the symbols are clear enough the g_m by C_i is the figure of (43:57) this is not only the device will talk circuit people will also talk. I am sure Prof. Radha Krishna should have mentioned g_m by C. That is the transit frequency which is omegat f with 1 by tau_t.

(Refer Slide Time: 44:20)



Now, let us see the cutoff frequencies. See the transit frequency is totally dependent on the transit time totally depending on device structure and the transport mechanism because that decides the velocity. Now, for this we will see the MOSFET example we will take MOSFET transit frequency of MOSFET equivalent circuit. Gate and then you have some resistor which may be the resistance of the gate region itself and also the contact resistance, so that comes up to the point and then we have a capacitor that is the gate capacitance C_i which were are talking of. C_i is the total gate capacitance which is equal to W into L into C oxide this is R is the gate resistance. It could be also when rest of the resistance call this as the g. So, this is the input point. Now, if this is V_i after all we are talking of small signals equivalent circuit now, which would mean that, when I apply ac signal the ac signal appearing across the gate capacitance contact resistance or the external applied resistance will be there from the supply. Then, what is the other equivalent circuit.

(Refer Slide Time: 46:48)

R= gate rusistan

This is actually the source point. From then on g_m times that voltage V_i that is the equivalent circuit because input you see from RC resistance and capacitance the output point is generating a voltage for giving a current source the drain current which is g_m times the gate voltage. So, that is currently by that is i_L put it as like that R_L we get the output there. Now, the question is what i_L is g_m times v_i which is actually equal to g_m times v_g divided by 1 plus jw R C_i. So, that is the i_L now what we are telling is the cutoff frequency is the frequency at which power output becomes half. Power output is I square L into R_L when the power output become half i_L becomes root two times at $f_c P_0$ is half of the P_m , P_m is low frequency maximum power that is therefore which is equal to half i_L square into so then i_L will be 1 by root 2 times low frequency value.

(Refer Slide Time: 49:33)

 i_L at f_c becomes 1 by root 2 times because square root of the low frequency value, so that becomes 1 by root 2 of low frequency value is g_m times v_g because this impedance will be large compared to that entire voltage across that. So, frequency value omega₀ value that is g_m times v_g . So, it becomes 1 by root 2 of that and then this becomes equal to 1. At omega_c omega R C_i equal to 1 without even growing through this we would have said cutoff frequencies related RC combination. We just went through that thing to keep the whole discussion open, now let us just go back here. So, you have got the f_c is 1 by 2 pi R C_i. You can see the whole thing depends upon the RC combination this cut of frequency. But, the transit time cutoff frequency depends upon g_m by C_i. Now, from this circuit point of view can you tell us what will be the condition for omegat here from this circuit you can see that, cutoff frequency depends upon R and C here. Now, if you go on reducing R it looks like as if cutoff frequency becomes infinite. It will not. This is not limited by this RC combination capacitance input and the resistance of the circuit. But as I keep on reducing the R what is the ultimate limit for the speed, transit time. Can you see what happens there what g_m by C_i will be that is what I try to illustrate. I want to bring it down to you to here that is if you call this as I input.

(Refer Slide Time: 52:14)

Last thing today is this aspect i_L divide by i_{in} how much is that is i_L is g_m times v_i . I can write v_i is i_{in} into, if I have a current coming in here what will be the v_i is i into impedance divided by j omega C_i. So i_L is equal to g_m into i_{in} divided by j omega C_i. I am getting at some thing which we usually do not see. Therefore, i_L divided by i_{in} is what magnitude, if it is the g_m divided by omega C_i. Now, you can see when omega is low that is something. Now, i_L divided by i_{in} equal to 1. When this current here is same as the current here, gives when i_L divided by i_{in} equal to 1, we get g_m divided by omega C_i equal to 1. Now, what is omega, then, omega equals g_m divided by C_i. That is the key thing I want to tell, so when i_L become by i_{in} is equal to 1 the frequency that happens at the frequency equal to g_m by v_i. What that actually is omega_t, so you have got now, what we have done now, is the transit frequency omegat is the frequency at which the current output is equal to the current input. That means it is no longer working as the way you want to work. We can put the resistors here or as well as here. You have the current that pumped in; we get certain voltage output I input into R_L. Now, if this current is same as that when we put it here, the output voltage that we get, i_L is equal to i_{in} that also into R_L , you would not get benefit of using that amplifier there and the circuit there.