Fundamentals of Semiconductor Devices
Prof. Digbijoy N. Nath
Center for Nano Science and Engineering
Indian Institute of Science Bangalore

Lecture - 42
Heterojunction transistors

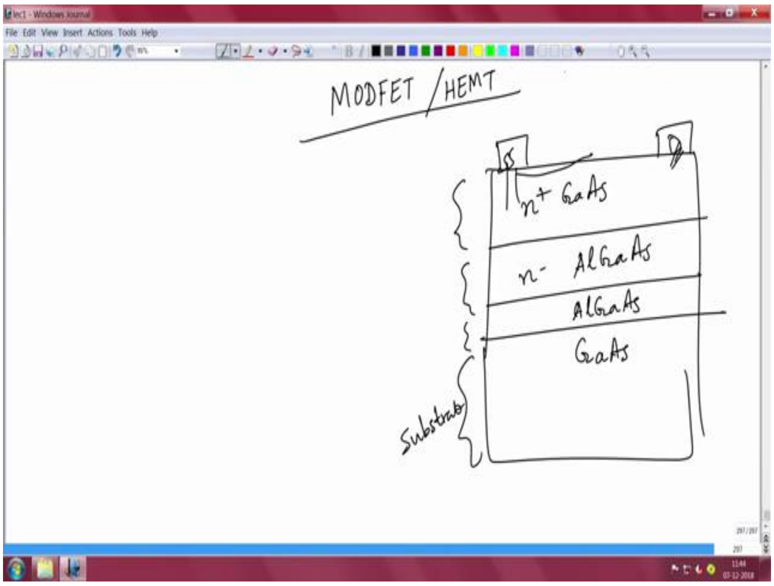
Welcome back, so we were at the in the discussion mode for 3-5 or compound semiconductor devices, today in this lecture and may be perhaps the next lecture we shall wrap up the part for a compound semiconductor. So, if you recall whatever we have studied till now we introduced different kinds of compound semiconductor, I told you that you can change the band gap, you can tune your band gap and other properties. And that also depends on the material science because in practically you have to grow the material with different lattice constant, different band gap by changing the aluminium mole fraction for example in aluminium gallium arsenide aluminium nitride.

We have discussed about grading of the band gap Heterostructure band diagrams, how to draw p n junctions of band of different materials they have that have different band gaps. We also discussed about the different kinds of band alignment staggered, straddling and broken, I have taken some examples of p n junctions. I also told you that different devices need this you know heterojunction because they have far more flexibility and freedom in designing different devices optical or electronic devices compared to silicon. Because, you can change the band gap and hence the properties and also the heterostructure is a very rich area of you know research where you can do lot of this new devices.

I told you that one of the good examples of heterostructure devices is a modulation doped FET with which we concluded last lecture if you remember; modulation doped FET or HEMT High Electron Mobility Transistor is a classic example of a new type of transistor it is not new now of course, it is been many decades old now. It is a type of transistor which is different from silicon MOSFET in that your channel charge is physically separated from the doping dopants and so you have very high mobility because the scattering is low and also you can confine them in something called a quantum well.

So, we will quickly again recap that transistor called MODFET and from there I shall go to how wide band gap devices can help in or different band gap devices can help in HBT, where are those devices used in real life. And then I will also touch base on how this can be used for heterojunction devices like optical devices LEDs and so on and from there we shall start a few lectures on optical devices that is the plan.

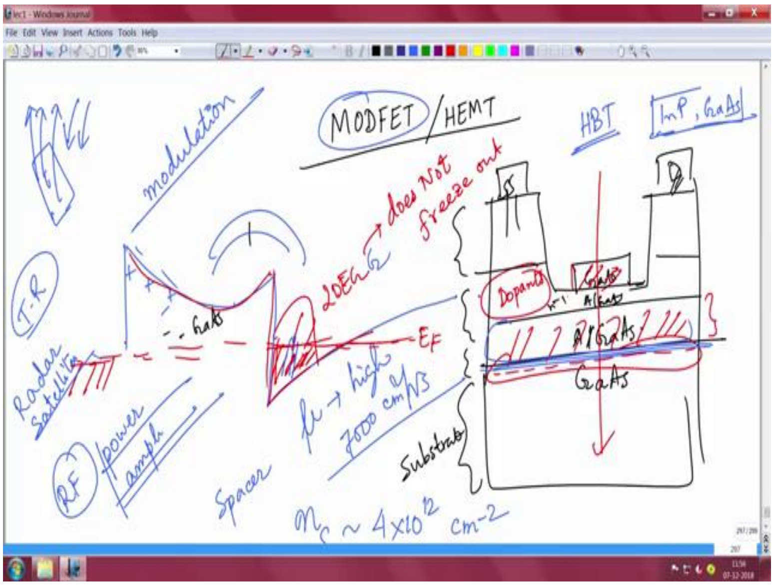
(Refer Slide Time: 02:43)



So, let us come back to where we ended in the last lecture, so if I if you recall I told you that there is this device called MODFET modulation doped FET or you can call it high electron mobility transistor HEMT. Essentially what happens is that you have you have a layer of very highly doped gallium arsenide and then you have a dope layer of aluminium gallium arsenide and then you have a layer of undoped aluminium gallium arsenide undoped and then you have a layer of undoped gallium arsenide.

These layers are very thin epitaxial layers this is like a substrate this is like a substrate so this is very thick ok, essentially what happens is that you want to make an source drain contact here I told you the source drain contact will be here very good. And, then you will essentially what you will do is that you will you will actuate the middle region, so that you only have a region like this right you will have a region like this.

(Refer Slide Time: 03:35)



So, your gate will come here and the gate will be on top of your AlGaAs that is doped right and then there is AlGaAs which is also undoped. The composition of aluminium and gallium can be adjusted depending on how much you need and of course there will be a very high density of electron gas that forms here it is called the 2 dimension electron gas that is the channel. So, the dopants are here the n plus dopants are here this layer is very thin and it is undoped it is called spacer layer, this physically separates the dopant layer from the channel layer ok, it can be few nano meter to 10 nano meter for example thick.

So, how the band diagram looks like along this direction you know I told you that there is a Fermi level there is a Fermi level in equilibrium I am drawing it right and then this is your gate that is there right. So, the gate has a Fermi level metal Fermi level there it looks like this right. So, essentially what happens is that this is your doped n type AlGaAs layer and this is your a thin spacer layer that essentially physically separates there is a electron gas here very high density of electron gas I call it 2D electron gas.

The electrons have come from this donor atoms here you have doped this layer no you have doped this layer. So, all these donors have donated the electrons they actually come here they come here and they form this very high density of charge and this layer at a separate few nanometer layer AlGaAs this basically separates this donor this donor ions from that a charge here. So, the scattering is low and hence mobility is very high mobility is high I told you know you can get 7000 or 8000 cm^2/V-s room temperature mobility very high the mobility can be achieved.

In this kind of structure that is why it is called high electron mobility transistor and the reason it is calling MODFET it is because it is modulation doped you are doping it here you are doping it here right you are doping it here but your channel is forming here. So, this is called modulation doping it is your called modulating, you are modulating the doping essentially you are changing the doping here and you are getting a charge here ok, so that is why it is called modulation doping.

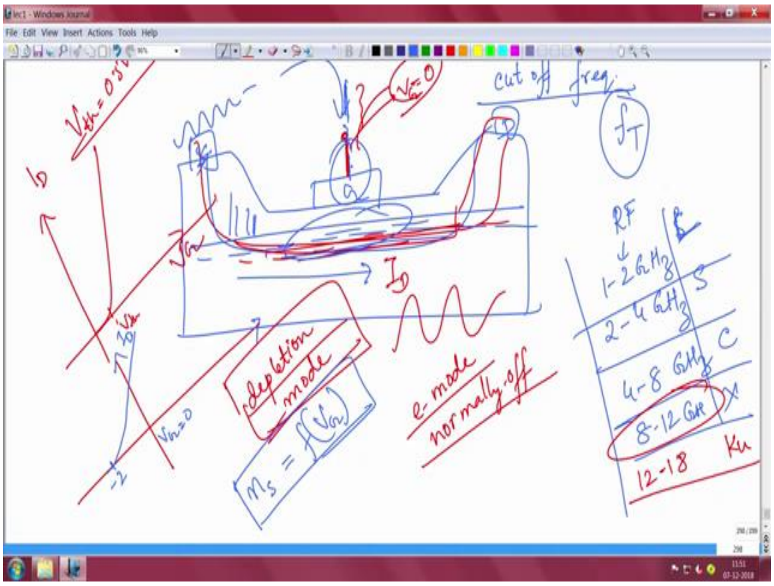
Of course you cannot dope this region very high otherwise the Fermi level will come very close and you will get a parasitic channel like an extra channel in this doped layer you do not want this doped layer to conduct also right. So, you have a very good charge density the typical charge density the 2 D charge density is called n_s typically and it is in the range of $4 * 10^{12} / cm^2$. So, this is an epitaxial interface here. So, the mobility is very high their interface scattering is low and you can get good currents good linearity and stuff like that. So, this is the high electron mobility transistor gallium arsenide that is widely used in RF.

In RF transistors especially in RF power amplifiers you know RF power amplifiers are those devices essentially they have devices plus RF passives which essentially amplify your RF signal. For example, your cell phone has to you know if you have a cell phone it has to the antenna has to receive the signal from the nearest base station the signals becomes very weak when it receives. So, you have to amplify the signal of course there will be low noise amplifier power amplifier it has to process and then it has to send back again by you know the another antenna you have to it has to send back to the nearest base station you need to boost up the power and send stuff like that.

So, there is a power amplifier power amplifiers RF power amplifiers, so it is in the TR module transceiver receiver module. So, those transceiver receiver modules in your you know smartphones could be either made of this kind of high electron mobility transistor made of gallium arsenide or it can be made of HBT heterojunction bipolar transistor, even their you might be using indium phosphide or gallium arsenide sort of HBT. These are essentially compound semiconductor devices that are enabling this efficient performance in your cell phones and other kind of gadgets where you need RF signals.

And even of course, places like radar wherever you need very high RF power and amplification you in takes places like radars or satellites for satellite communications you use this kind of gallium arsenide MOSFET, because you know it can compare to silicone it is mobility is high it is mobility is high it is velocity also is very high it can have much higher cut off frequency there is something called Cut off frequency.

(Refer Slide Time: 07:33)



Cut off frequency that is the frequency up at up to the maximum frequency up to which it can deliver gain one is called f_T which is the current gain cut off frequency. So, this kind of gallium arsenide indium phosphide sort of devices can actually give you much higher f_T and hence much higher gain than silicon devices.

So, silicon devices of course cheaper relatively but you know if you want very good performance for example, then this kind of devices can give you good RF output power essentially your RF signal will be say few giga Hertz you know they are different bands you can have between 1 to 2 giga Hertz, these are different bands that are split up 1 to 2 giga Hertz this is I guess l S band this is L band this is S band then 4 to 8 giga Hertz this is called C band these are actually different classifications and then 8 to 12 giga Hertz this is called X band and so on. So, these are different bands in RF frequencies and what I mean is that the signal has to come at that is high frequency and in your transistor this signal will be fed into your gate of the transistor right.

So, essentially it is not a simple thing you have this source here your drain here right and then your gated it is here essentially your electron gas is here right. So, this is your transistor your gate that is there know in the gate you feed the small signal and if you bias the gate of course at a DC voltage and you feed the small signal at the gates. So, this is the drain current right this is the drain current that goes this drain current can be modulated with the small signal you can get the large signal gain here.

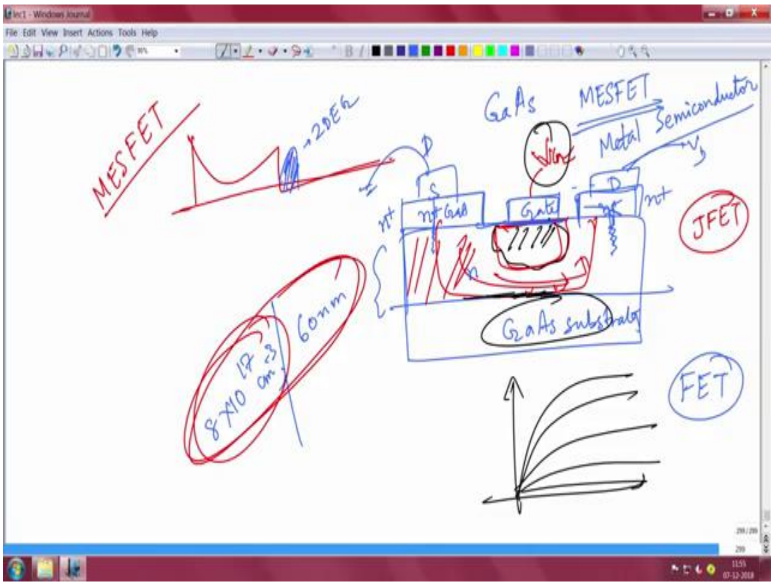
So, you use that for RF signal amplification and this kind of gallium arsenide devices would perform superior to silicon MOSFET in terms of RF signal amplification at this very high frequencies of X band then even higher frequencies for example, you know 12 to 18 giga Hertz is called Ku band. For example, all this area in all this high frequencies gallium arsenide devices really perform well and that is why they have been dominating this landscape for a long time.

Also please remember that this kind of devices of gallium arsenide the 2D electron gas this electron gas is always there even if you apply 0 voltage on the gate I told you this before in the last class also, that is why these are called depletion mode devices. Devices in which the charge exist at 0 gate voltage or you know if you keep the gate floating and you apply a source drain voltage you will get very high current, because the channel is on, these are called depletion mode devices and if the channel is not on in many of the silicon MOSFET. For example, at 0 voltage on the gate your channel inversion does not from those are called enhancement more devices or e more devices or called normally off devices same thing actually.

So, silicon MOSFET n MOSFET could be normally of most of the time, because your if you recall your silicon MOSFET classes at 0 gate voltage you will not invert the channel you have to apply some positive threshold voltage like say 0.5 volt to invert the channel. So, at 0 voltage you know if you draw a silicon MOSFET ID the channel current versus VG you know at 0 voltage on the gate you will have nothing and then you will have this is the threshold voltage at which the current start. Unlike that in this kind of gallium arsenide devices you have very high charge density at 0 gate voltage. So, you have to apply some minus 2 or minus 3 volt actually to turn it off and then at 0 voltages it has very high current. So, this kind of devices are called depletion mode devices.

And this is one of the many examples of heterojunction devices that are used widely used in modern day electronic specially RF electronics and in devices that require amplifier in a amplification also there is another that is heterostructure based devises it is called MOSFET. There is a lot of physics of course behind it on this electron gas that forms and also of course a lot of mathematical expressions as to how we can this is a charge that is formed below the gate it is called n s the 2D electron gas how is the charge controlled by the gate voltage right it is a function of the gate voltage. So, there mathematical expressions behind is that depends on the doping density that you are using in the barrier layer that is the spacer thickness and so on.

(Refer Slide Time: 11:29)



The those if you are interested you know you can ask me email me and I can send you texts and text books and other things that you can read, but this is not we will not go to the mathematics right. Now here another very important devices is compound semiconductor device that is widely used is Gallium Arsenide MESFET, gallium arsenide MESFET MESFET actually stands for metal semiconductor field effect transistor metal semiconductor field effect transistor and this is not modulation doped. It is basically what happens is that you have a lightly doped n n minus a lightly doped gallium arsenide region which could be say you know around 60 nanometer or maybe 100 nanometer.

When I say lightly doped you might be using something like say 8 into 10 to the power 17 per centimetre cube maybe that is the doping maybe and this is the thickness this is of course, in undoped gallium arsenide substrate on top of which you grow this mildly doped thing and then of course you grow very thick a very highly doped and plus gallium arsenide to make contact. So, that you can make a source contact n plus essentially this is very highly doped you make drain contact.

You know your contact resistance has to be low otherwise when you apply this is ground. Of course, when you apply the voltage and you have a current that flows lot of voltage will unnecessarily drop across this contacts, so you do not want the contact resistance to be bad and hence why you and that is the reason why you put an n plus layer here and here so that your contact is good you contact resistance is low.

And of course, it is grown everywhere it is grown everywhere you just etch over here over here you etch away and you put it a gate here a gate you put you etch away you etch away this region this region, see because the gate you do not want to put on n plus otherwise the gate will leak the whole premise of a field effect transistor is that the gate should not leak it should be an insulating gate ok. So, the gate will deplete the this area no so this area will be depleted by the gate as you apply more negative voltage, it will basically deplete this and this depletion will what will happen because of depletion the charge that is flowing from the source and the drain this channel length will basically this thickness of the channel length will reduce right.

So, it will deplete some area, so it will deplete some area below here as you increase the gate voltage. So, your the path that the current flows this channel this thickness reduces right, so your current can be modulated by the gate voltage. As you keep applying more and more gate voltage you will actually deplete more and more, so your what happens is that the channel becomes even tinnier at one point you will be such that you will apply such a high gate voltage that the entire will be depleted it can be depleted and then your channel will be pinched off and then your essentially channel is turned off ok.

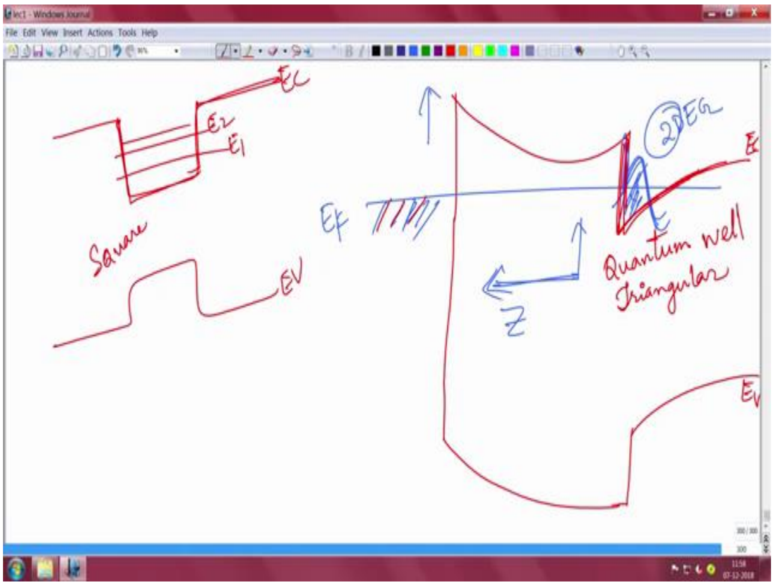
It is pinched off as in the sense your channel has at a very high negative voltage your device has stopped working, I mean the current has become off because the current now cannot flow here it is gone and this is insulating so it will not go there. So, this is how you basically get the ID Vd characteristics you know ID Vd characteristics that a negative voltage you will eventually be not able to conduct at a negative voltage you will not be able to conduct. So, because the channel thickness is gone again if you want to conduct you will apply you will apply a lesser negative voltage, so that depletion is like that and your current can go away in this way that is called like a MESFET.

Of course, there is another variation called JFET in which instead of a this is short key barrier you are using here to actually deplete this is short key barrier depletion, in JFET you can use a p type junction here to deplete that is also similar do not worry about that so much. But this MESFET you know this MESFET you are using this is also used very widely for RF linear applications that require linearity, you know this MESFET is basically having a n minus doped layer and this sort of a device can actually freeze out at very low temperature like 4 Kelvin or so.

This carrier concentration will reduce because career freeze out will happen so MESFET are not preferred to be operated at low temperature. But a MODFET you remember the MODFET band diagram I told you it is like this right the Fermi level is here you see the Fermi level is going above the conduction band. So, you have an electron gas here, whenever your Fermi level crosses the conduction band and you have an electron gas like 2 D electron gas here if you recall the band diagram from here right this is the this is this is your this is your band diagram right.

If you recall and this is your Fermi level, your Fermi level is crossing above the conduction band here this is 2 d electron gas right this this does not freeze out this does not freeze out it stays constant because your Fermi level is above the conduction band here the conduction band is gone below this is the degenerate electron gas you can say and this does not freeze out in the low temperature. So, MODFET can be used at very low temperature also even low it as low as one Kelvin or 4 Kelvin at. In fact, at low temperature your mobility will become better in this kind of devices, so you get higher current in this MODFET. So, this MODFET and MESFET are 2 different kinds of devices in a way please remember also that.

(Refer Slide Time: 16:09)



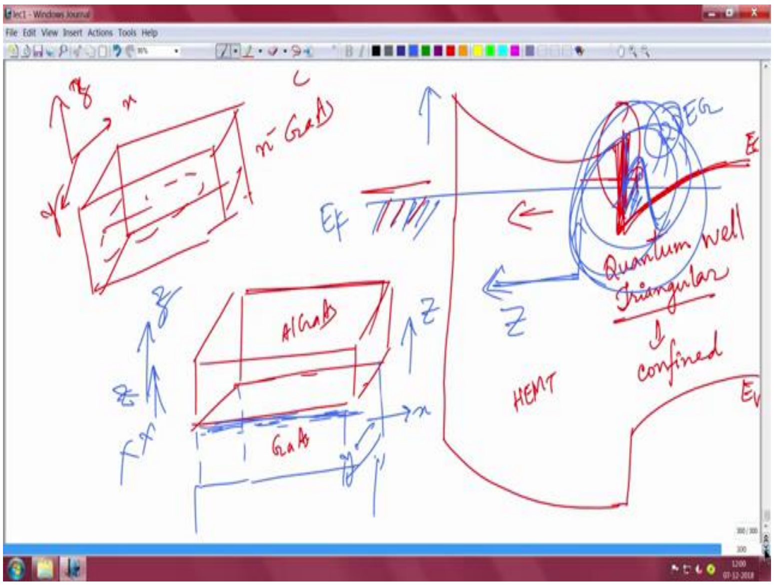
When you have MODFET like that you know let me draw the Fermi level here this is Fermi level EF metal here and then I draw I have drawn the band diagram like this, of course your valence band also will look like this ok. This is your conduction band this is a valence band of course, and this is a Fermi level right this is a Fermi level.

So, the reason you know there is a Fermi level is going out of the conduction band and there is this electron gas I keep this is a wave function actually of the electron gas I keep drawing like this is electron gas is called 2 dimensional electron gas the reason is called 2 dimension is because, this electron can only move in 2 direction which is x and y it cannot move in this z direction this is the z direction this is energy.

Of course, so this is energy you cannot visualize x here but essentially what happens is that this electron gas can only move in 2 direction which is x y the plane of the wafer, it cannot move in the z direction that is in the direction of the growth of the vapour the thickness of the vapour because there is the barrier here. This barrier see this barrier will essentially prevent it and there is a barrier here this is a Quantum well essentially, quantum wells can be of different shapes and sizes this is a quantum well formed at heterostructure interface and is called a triangular quantum well it is called a triangular quantum well because it is like a triangle.

Remember I told you we can have this kind of square quantum wells this is E_C this is E_V right, I told you this can be as you can have sub band levels like E_1 E2 which you studied in high school about potential in a periodic you know sorry particle in a box it can be physically realised in Heterostructure. I told you this is like a square quantum well right this is like a square quantum well that can be formed also from heterostructure only different kinds of heterostructure. But in this gallium arsenide modulation doped or HEMT this is the HEMT right.

(Refer Slide Time: 17:49)



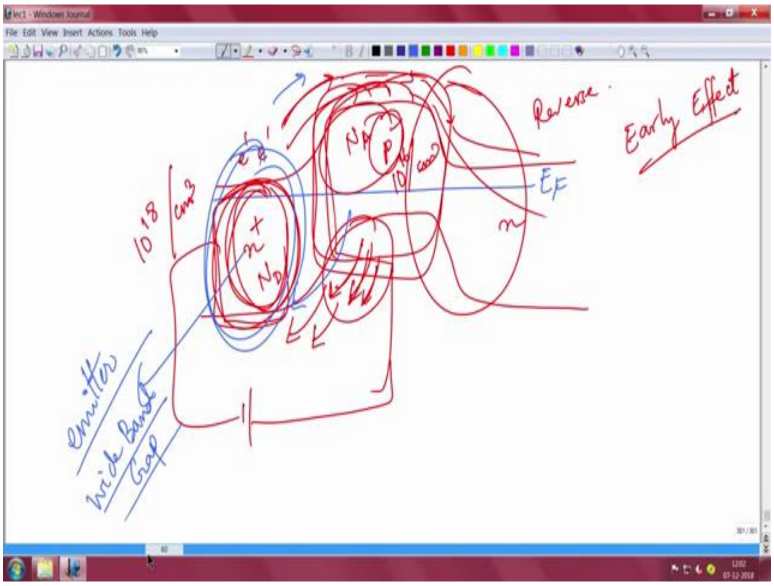
Here the quantum well that is formed is a triangular quantum well because it is look like a triangle of course and it is a quantum well in quantum well carriers are confined carriers are confined in only in 2 dimension. So, in x and y in the plane of the device it can move but vertically in this direction it cannot move in the z direction because, there is the confinement here this confinement this confinement will not allow the electrons to move.

So, unlike in a bulk semiconductor so if you take you know a bulk semiconductor for example I am taking a bulk semiconductor here right, I am taking a bulk semiconductor here which is n minus doped gallium arsenide or something carriers the carriers that are here can move in all direction x this is z you x y in all direction carriers can move if we apply a voltage in a way.

But here if you even if you apply voltage in this direction you know carriers cannot move. So, what I am trying to say is that if I take an example of gallium arsenide wafer I am drawing a square (Refer Time: 18:44) wafer and a device can be anything. So, this is the top suppose this is the top suppose AlGaAs layer right this is the top AlGaAs layer and then below it you will have the gallium arsenide layer here again right. If this is the top AlGaAs layer aluminium gallium arsenide layer suppose and the bottom is the gallium arsenide layer then essentially you will have a 2D electron gas thus below here no 2D electron gas below at the interface ok.

This is the 2D electron gas that I am talking about this is the z- direction of growth this is the direction in which you are depositing the material layer by layer because, this is the vapour and this is the direction in which you are growing the material. So, it cannot the electrons cannot move in this direction in this z- direction the electrons cannot move because, there is a barrier here this is interface is actually barrier here electrons can only move in x - direction and y- direction ok. That is why it is a 2 dimension electron gas unlike in MESFET this happens in MODFET or a HEMT. So, that is why carrier this is called carrier confinement ok. So, that is one of the classic examples and then if you recall our BJT discussion if you recall I had enough many lectures back.

(Refer Slide Time: 19:49)



We have discussed BJT suppose this is a Fermi level and you have I was talking about silicon npn BJT right, this is a npn BJT n p n the same band gap of course if you recall the doping here was N_D and a doping here was N_A and in forward bias the way you operate BJT is that you forward bias the base emitter junction here right you forward bias the base emitter junction here.

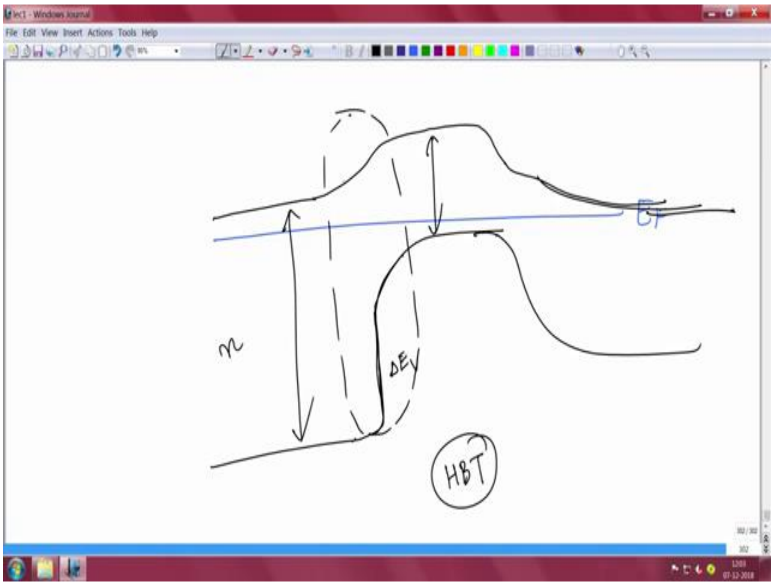
So, the electrons are injected here holes are back injected and the idea is that electrons have to diffuse and essentially go here and get collected this part is reverse biased ok, so that will be swept away by the field here and they will be collected. The problem here of course was that you do not want this back injected component to be more because, this back injected component does not contribute to gain or do anything you only want that injected electrons should be very high. Whereas, the back injected electrons holes should be very low and that is why you need to dope with very high n- plus and you need to dope with low.

So, that if this doping is very high say $10^{18}/cm^3$ and if this doping is very low say $10^{16}/cm^3$, then the current that you are injecting this side due to electrons will be at least 100 times more than the electrons that holes that you are inject in this side. So, that is why you have a n plus region p region here but if you dope this high you know you cannot keep doping it high because one to (Refer Time: 21:03) your base resistance increases because if you dope with low, that base resistance increasing you know your gain will suffer eventually not your gain. But your power gain power performance will suffer and a base resistance increase so it will lose the transistor action eventually.

If the base resistance also is low then you will have early effect if you recall and that early effect will also essentially deteriorate the device, so you do not want the base to be doped light. But if the moment you dope the base high your emitter you know if it comes equal to almost the emitter doping your gain will not be there because, the number of electrons you are injecting this side will be equal to number of holes coming this side it will be very terrible.

So, and if you cannot dope the emitter very high because after sometime the doping is so high that the band gap will narrow the band gap will shrink, that is why the solution I told you was that in the emitter side here this emitter and this was also proposed by (Shockley and Herbert kromer if you if you use the emitter as wide band gap. If you use if you use a wide band gap material on this emitter side, then lot of this problems will be solved. What I mean by that is that you have suppose the Fermi level here.

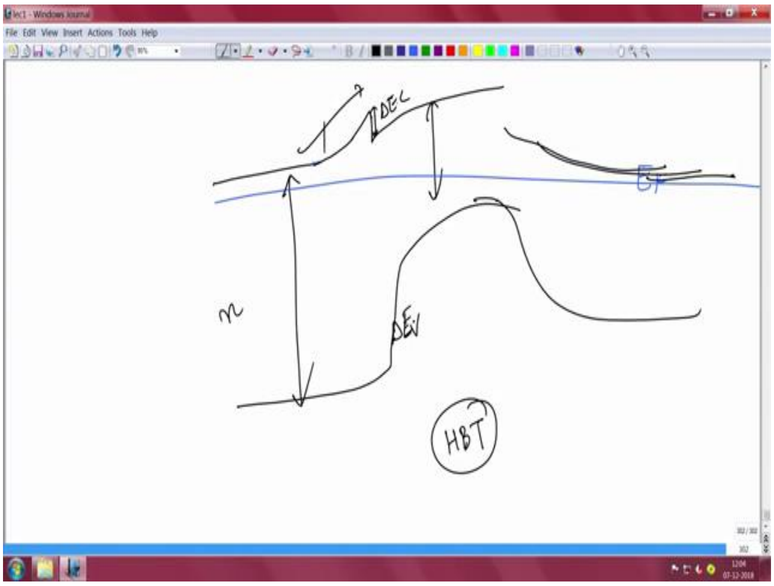
(Refer Slide Time: 22:05)



You use a wide band gap material here for an n region and then on and you can adjust the doping and the grading between emitter and base I told you grading can be done so that thing are smooth ok. So, you can actually have a narrow band gap material on the base side here like this, this is a delta E_V by the way this delta E_V the discontinuity under valence band and then you can have something like that and then you can again have sorry the collector here.

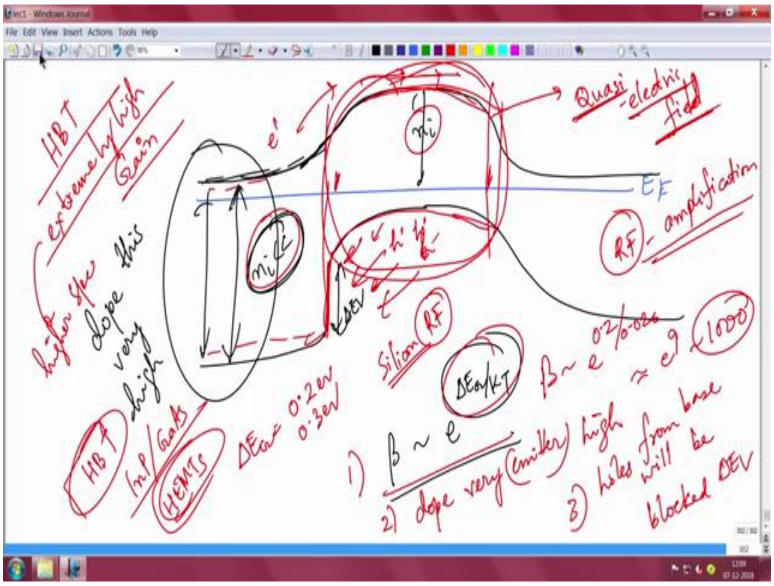
So, essentially what happens is that this band gap is lower than this band gap this is called HBT Heterojunction Bipolar Transistor because, you was I had probably told you in the BJT classes also this part that you are grading this part that you are having this heterojunction you can actually do this tricks in heterojunction engineering such that the most of the discontinuity falls across the valence band and very little discontinuity falls across the conduction band. Because if you do not do this engineering trick then what will happen is that right you will end up getting a discontinuity in both valence and conduction band, so it will look like this it will look like this.

(Refer Slide Time: 23:11)



So, it will look like this and you dont want that kind of notch here because this is delta EC this is delta EV right conduction and valence band discontinuity it is electrons will have this problem you dont want this kind of a structure. So, what you do is that you make this grading smooth and you can decide the way you grade such that you have a suppose you have a Fermi level here.

(Refer Slide Time: 23:35)



You know you want that the conduction band should have smoothness like this, but the valence band should have discontinuity like this ok. The valence band should have discontinuity this discontinuity delta E_V and the conduction band discontinuity should be smooth you can do this tricks actually, otherwise there will be also there will be notch a discontinuity in the conduction band which is not good.

Now the benefit of this is that because this is a wide band gap device your n_i is much smaller than much much smaller than the n_i here and because of that what will happen is that your gain your gain will increase by a factor of e to the power delta EG/ kT because of this n_i difference because n_i depends on E_G.

Because the n_i there is a difference in the band gap between here and here your n_i will be dramatically smaller here and that is why that will manifest itself as a extra in the gain. Also you can dope this very high now you can dope this very high to have higher gain, you can dope this very high even if you dope it very high even if the band gap shrinks that band gap shrinking that band gap shrinking little bit of band gap shrinking will not matter because your delta E_V is very large.

So, that is benefit number 2. Benefit number 1 is that your gain will automatically increase because of this n i difference, gain number 2 is that you can dope very high you can dope the emitter very high, I mean to say dope emitter very high you can dope the emitter very high for a higher injection, that means electrons will go better and holes will not come still your band gap narrowing will not affect anything because delta E V will take care of that and your gain will automatically be high.

And number 3 is that this hole from the base will not be back injected because there is a barrier here this barrier will prevent the holes from back injecting. So, holes from base will be blocked will be blocked will be blocked by the delta E_V barrier right, this barrier will block the holes that the holes that are trying to come. So, that will even decrease the hole component and your emitter injection efficiency will increase. So, this HBT and this delta E_V by the delta E_C total even if this band gap difference is only say 0.2 eV or 0.3 eV it is exponentially dependent. So, your gain will increase by e to the power 0.2 by 0.026 which is almost e^9 , e^9 is like 1000.

So, your gain will probably increase by thousand times can you believe that a gain will increase by 1000 times. So, this HBT can provide extremely high gain exceptionally high gain extremely high gain this can provide extremely high gain, also they can provide you very much higher speed higher speed they can provide the reason is you see this base region you can actually when the electrons drift across the base on.

I mean transit across the base drift- diffusion you can make the electrons move faster across the base by also band engineering this base layer, what you do is that you can grade this base layer you can grade this base layer. So, as to develop a quasi electric field remember I told you in the last class or probably last to last class. When you have a grading in the layer your conduction valence band can be slope can we adjusted accordingly, because of the gradual change in the band gap it is called grading

And that can lead to an additional electric field called the quasi- electric field in the base layer which is not possible if the semiconductor has only one band gap, then people try to dope it little bit differently from one point you dope it differently to another point. So, that there is a field but compared to that this is much better you have a quasi electric field you are changing the band gap in a way gradually such that there is an inbuilt electric field that is developed here that will make the transit of the electrons even faster and you will get a better in a faster device with much higher gain.

So, this devices are typically made up of indium phosphide gallium arsenide sort of structure, along with this indium phosphide gallium arsenide HBT you also have these HEMT that I have talked about together with HEMT and HBT they have dominated the RF landscape for a long time, the RF amplifier RF amplification longs long time and silicon RF devices are still there please do not get me wrong.

But device in areas applications where you need really high performance and these are not that expensive of course, I means relatively expensive than silicon. But all your cell phones for example, have this gallium arsenide indium phosphide HEMT or HBT of the radars and the satellite communication many of the applications where you have RF device you know signal processing you need this is HBT and HEMT and the next incoming 5 G technology.

For example will also depend as lot on HBT HEMT of course silicon RF devices silicon RF devices are also competing because, they have the cost advantage and integration at CMOS platform you know CMOS platform also has their advantage that because gallium arsenide gallium phosphide platforms are not as high as dense you know as silicon. So, those are things you should be aware of, but anyways these are devices that are very you know widely used.

So, what we will do here is that we will wrap up the class here today this lecture here ok, we will end the lecture here today and in the next lecture we shall start or not we shall start we shall continue a little bit on that will be the last lecture in compound semiconductor. I shall introduce very briefly to another class of compound semiconductor called 3 nitrides that you if you recall it is gallium nitride based technology compound semiconductor, a reason I will touch base very briefly on gallium nitride technologies because you should be just aware or just basically be familiar with the very basics of 3 nitride technology.

They are pervading our life very much all your white LEDs that you have seen in the shops and street lights and all are made of gallium nitride. So, we should be slightly aware to the very basic minimum if you know someone asks you like why how is this LED working you should be able to tell something.

So, we will touch base quickly on gallium nitride materials and devices, I will tell you briefly about the applications that are going on and then that will be the end of compound semiconductor portion and from there and then from the next to next class we will start a few lectures maybe 5- 6 lectures on optical devices, solar cells LEDs and photo detectors that will be the agenda. So, we will end up class here I will hope to see you in the next class.

Thank you.