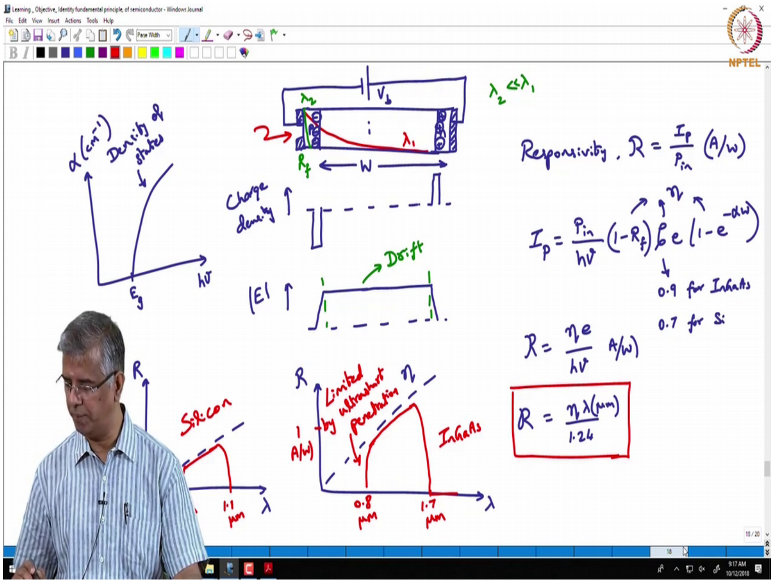
Introduction to Photonics
Professor Balaji Srinivasan
Department of Electrical Engineering
Indian Institute of Technology Madras
Semiconductor Detectors 2

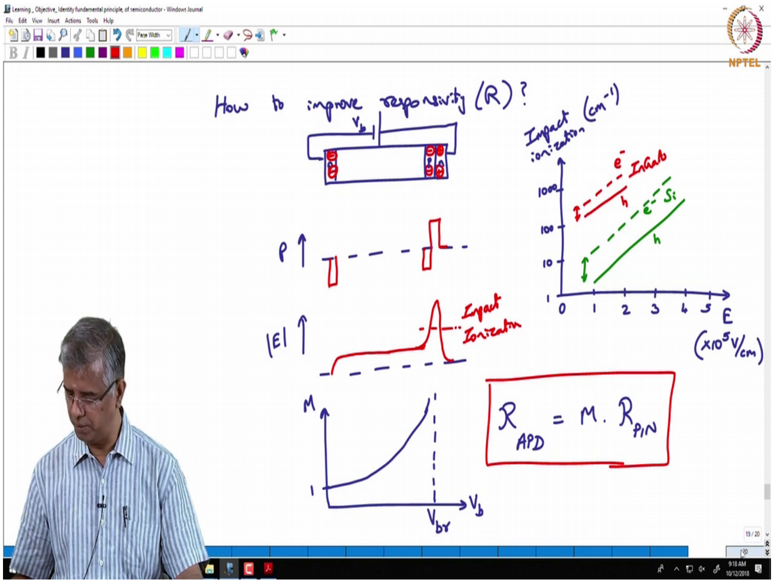
Okay welcome to introduction to photonics, we have been talking about Semiconductor light detectors and yesterday we went through some of the basic physics of semiconductors lights detectors, so essentially PN junction diode on which when we illuminate the junction diode with light you see a corresponding generation of electrons hole pairs as long as you like has energy greater than the band gap energy and then we saw that because of issues with respect to the carrier transport insight that medium, if you do not have an electric field you have carriers they are under the influence of diffusion wherein the velocity is much lower whereas if you are under the influence of an electric field you can actually push out these carriers much faster, so the response time of your photo diode is much better and that is the region where they would like to operate that is called the drift transport.

(Refer Slide Time: 1:40)



So to enable that we have been considering PIN diode and we looked at some of the characteristics of PIN photo diodes, we defined this responsivity as eta multiplied by lambda which is expressed in microns divided by 1.24. Eta is called the quantum efficiency but may be that is a misnomer because when we talk about quantum efficiency we talk about purely quantum effect. Eta has these 3 components 1 minus RF and then also 1 minus E power minus Alpha W which are based structure parameter, so the real quantum quantity is this zeta which is… what is zeta correspond to? That defines the fraction of the electron hole pairs generated for photons that are coming into the photodiode right that determines the conversion efficiency, so anyway we came up with this expression for R the responsibility and we looked at the physics of how the responsivity changes its function of lambda.

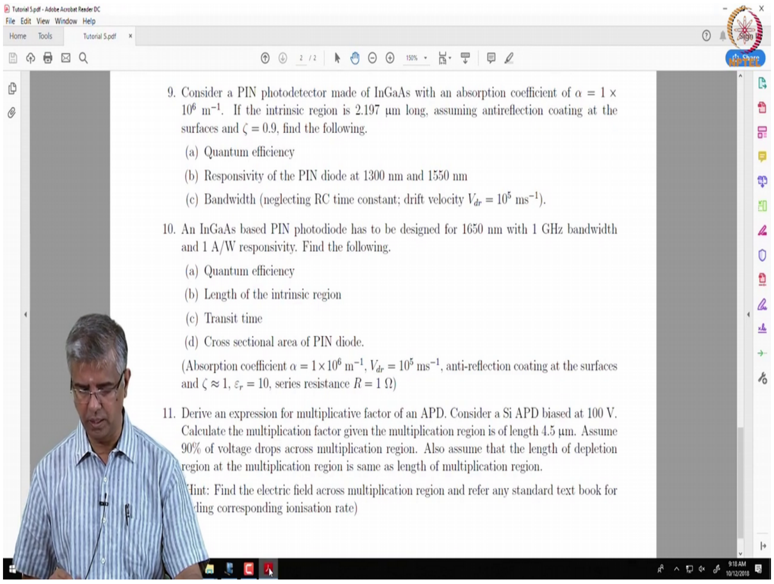
(Refer Slide Time: 3:13)



$$R_{APD} = M \cdot R_{PIN}$$

And then we went on to improving the responsivity by adding an extra layer in the PIN junction and that extra layer or such that it was creating a large potential in the charge density and correspondingly there is a large drop in the voltage across the junction, so when you apply an external bias most of the voltage drop happens in that junction in the PN junction on the right side and if the electric field is high enough we are talking about electric field magnitudes in the order of 10 power 5 volt per centimetre then you have impact ionisation happening which gives rise to avalanche multiplication of your carriers okay and then we start at the point where we said the responsivity of…APD is M times the responsivity of a PIN where M is determining by you know the electric field or in practical terms it corresponds to the voltage that we apply to this structure, so based on all of this let us just take up a couple of examples and try to solve those.

(Refer Slide Time: 4:45)



9. Consider a PIN photodetector made of InGaAs with an absorption coefficient of $\alpha = 1 \times 10^6\ \text{m}^{-1}$. If the intrinsic region is 2.197 μm long, assuming antireflection coating at the surfaces and $\zeta = 0.9$, find the following.
   (a) Quantum efficiency
   (b) Responsivity of the PIN diode at 1300 nm and 1550 nm
   (c) Bandwidth (neglecting RC time constant; drift velocity $V_{dr} = 10^5\ \text{ms}^{-1}$).

10. An InGaAs based PIN photodiode has to be designed for 1650 nm with 1 GHz bandwidth and 1 A/W responsivity. Find the following.
   (a) Quantum efficiency
   (b) Length of the intrinsic region
   (c) Transit time
   (d) Cross sectional area of PIN diode.

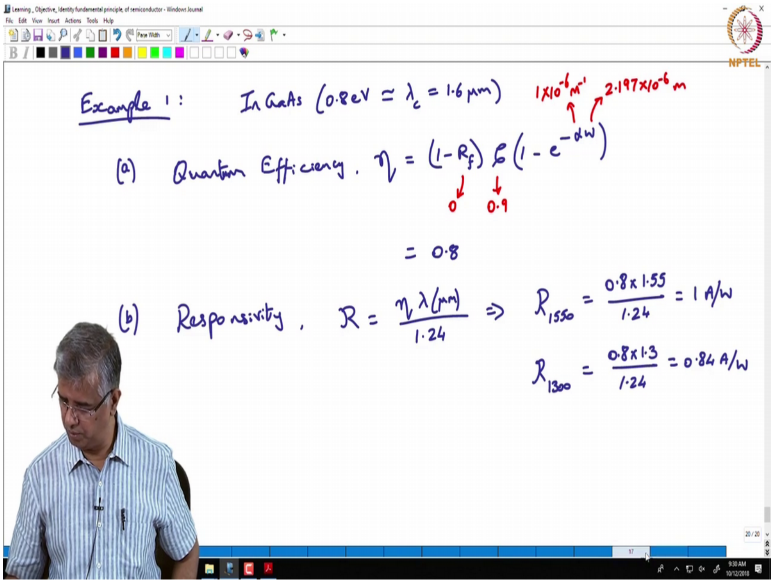
   (Absorption coefficient $\alpha = 1 \times 10^6\ \text{m}^{-1}$, $V_{dr} = 10^5\ \text{ms}^{-1}$, anti-reflection coating at the surfaces and $\zeta \approx 1$, $\varepsilon_r = 10$, series resistance $R = 1\ \Omega$)

11. Derive an expression for multiplicative factor of an APD. Consider a Si APD biased at 100 V. Calculate the multiplication factor given the multiplication region is of length 4.5 μm. Assume 90% of voltage drops across multiplication region. Also assume that the length of depletion region at the multiplication region is same as length of multiplication region.

   Hint: Find the electric field across multiplication region and refer any standard text book for ding corresponding ionisation rate)

I am going to pick examples from here question number 9 and 11 okay. Question number 9 we are talking about Indium gallium arsenide based PIN photo detector. The absorption coefficient is given that should be as a function of lambda but let just say that absorption coefficient correspond to what you have near the bandage. If the intrinsic region is about 2 microns long and assuming antireflection coating, the surface when you are assuming antireflection coating that means RF the factor that we define yesterday RF is equal to 0 and then zeta is given as 0.9 we are asked to find the quantum efficiency the responsivity and then the band width. Quantum efficiency responsivity we have been discussing yesterday, the bandwidth we did not talk about very much I will talk about it in little more detail today okay.

(Refer Slide Time: 6:09)



Example 1: InGaAs ($0.8\,eV \simeq \lambda_c = 1.6\,\mu m$)

(a) Quantum Efficiency. $\eta = \underset{\substack{\downarrow \\ 0}}{(1-R_f)}\,\underset{\substack{\downarrow \\ 0.9}}{\zeta}\,\left(1 - e^{-\overset{\substack{1\times10^{-6}\,m^{-1} \\ \uparrow}}{\alpha}\overset{\substack{2.197\times10^{-6}\,m \\ \nearrow}}{w}}\right)$

$= 0.8$

(b) Responsivity, $R = \frac{\eta\lambda(\mu m)}{1.24} \Rightarrow R_{1550} = \frac{0.8\times1.55}{1.24} = 1\,A/W$

$R_{1300} = \frac{0.8\times1.3}{1.24} = 0.84\,A/W$

So let us look at this let us see how we can determine these quantities. Example 1, so we have Indium gallium arsenide, Indium gallium arsenide actually has a band gap energy of about 0.8 electron volt which corresponds to lambda c that is a cut-off wavelength in the order of around 1.6 microns okay, so we said gallium arsenide has 1.42 ev right whereas Indium gallium arsenide has 0.8 ev almost half the band gap energy, so that is the beauty of working with semiconductors you add elements in this case Indium and you are able to change the band gap to such a large extent and so you can change the emission or detection characteristics of these semiconductors material over a wide region right.

So 1.42 ev if you say for gallium arsenide which we were using for LED's right that will correspond to a wavelength something in the order of 850 nanometres okay so gallium arsenide… LEDs would emit at 850 nanometres and then of course if use gallium arsenide itself as your detector then you can say that, that is going to be sensitive to 850 nanometres and below right the cut-off the bandage correspond to about 850 nanometres that is the cut-off wavelength.

So below that wavelength or energy greater than that energy you will have you know responsivity okay but we are talking about Indium gallium arsenide here because typically the reason why people are interested in Indium gallium arsenide is most of your optical communication, what is the band at which optical communication happens, wavelength band? 1550 nanometres, right 1.55 microns so you know we needed detectors for optical communication, so Indium gallium arsenide was actually proposed as one of these detectors. I should also mention that Germanium is another popular detector not as popular as Indium gallium arsenide these days but before Indium gallium arsenide was brought about you had germanium as your detector, germanium has a similar band gap energy.

So germanium is sensitive to near and (())(9:24) wavelengths okay but let us look at those problems where the given indium gallium arsenide and we are told that in the A part we are asked to figure out the quantum efficiency Eta which derive yesterday as 1 minus RF Zeta multiplied by 1 minus e power minus Alpha times w where w is the width of the intrinsic region which approximately can say is the width corresponds to the width of the PIN itself because the PN layers are the widths are much lower compared to the I layer width typically okay.

So in this problem we are told that this RF is 0 we have been asked to assume that RF is 0. This zeta is given as 0.9 and your alpha is given as 1 into 10 power minus 6 per meter and your W width is given as to 0.197 microns into 10 power minus 6 metres right, so it is just direct substitution once you recognize this, so E power minus 2.17 multiplied by 0.9 if you do the maths it will come out to be 0.8 okay, so does it have a unit? No, we are just talking about fractions, so eta is equal to 0.8, second part is the responsivity, we have been asked to figure out the responsivity at 2 different wavelengths 1550 and 1310, so responsivity is given by eta multiplied by lambda where lambda is expressed in terms of microns divided by 1.24, so this is once again direct substitution this implies the responsivity at 1550 nanometres is corresponding to eta is 0.8 multiplied by 1.55 in microns right 1550 is 1.55 microns divided by 1.24 and if you do the maths I believe you get 1 ampere per watt. Yes.

Student: (())(12:39).

Prof: Question is whether this Indium gallium arsenide is direct or indirect? It is actually direct band gap semiconductor, so you can use Indium gallium arsenide material for achieving lasers at 1550 also. It typically use Indium gallium arsenide phosphide, phosphide actually finely determines the band gap energy.

Student: (())(13:11)

Prof: Direct, yes.

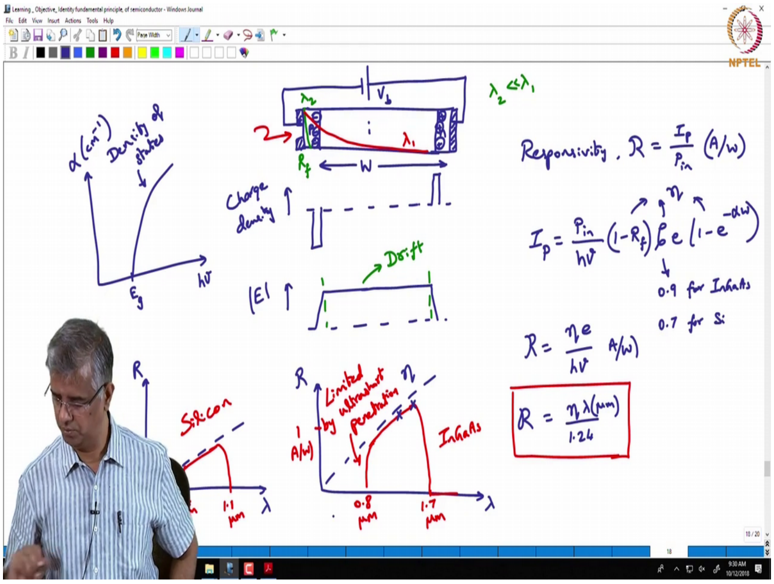
Student: (())(13:19)

Prof: So the question is, it is a good question, so LEDs we insist on having a direct band gap energy semiconductor…what about for detectors? So for sources yes for the emission to happen it has to be direct band gap semiconductors because you want to have this transition recombination happening without the need of generating this extra heat energy right because you want to maximise your conversion efficiency, radiative efficiency but in the case of detector do not have that sort of constraint because all your trying to do is just take the electron out of the valance band right.

If you just able to take the electron out of the valance band you do not care whether part of the energy is going into vibration energy and all that, so you do not have as much of constraint as far as detection is concerned compared to the emission, so indirect band gap semiconductors can be used but we have accounted that in zeta that conversion efficiency. Indirect band gap semiconductors have less conversion efficiency compared to direct band gap semiconductors. Right so then we have been asked to figure out the responsivity at 1300 nanometre 1.3 microns and if you do that you get a value of about 0.84 ampere per watt.

So lot of time people wonder whether you can have responsivities of greater than one amp per watt right, what do you think? Yes, no. Yes responsivity is just numerical quantity in terms of energy the quantum effects are all absorbed in zeta right so the responsivity can be anything, it can be even 2 amperes per watt if you have longer wavelength can get 2 amperes per watt, so there is once again people tend to think that responsivity should be a fractional value and does not have to be less than 1 it can be greater than 1 also but what you see here is it is roughly equal at 1550 and 1300.

(Refer Slide Time: 16:28)



Responsivity, $R = \frac{I_p}{P_{in}}$ (A/W)

$I_p = \frac{P_{in}}{h\nu}(1-R_f)\zeta e(1-e^{-\alpha W})$

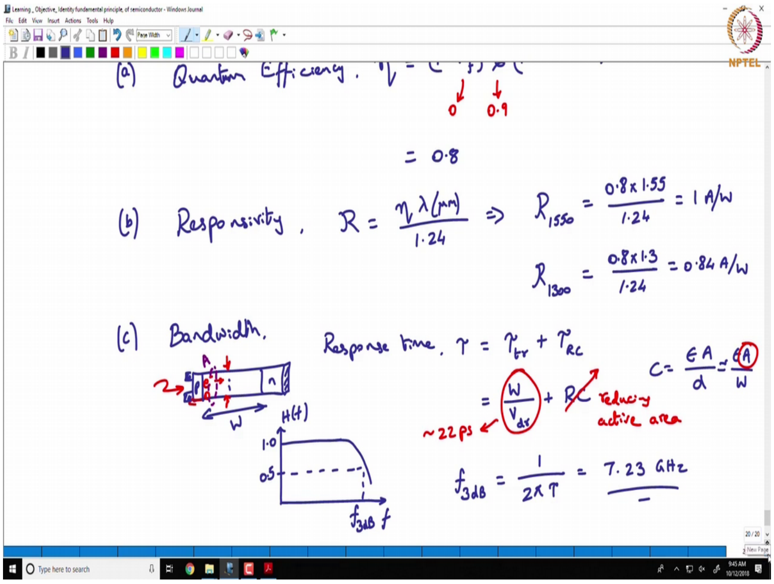
ζ → 0.9 for InGaAs, 0.7 for Si

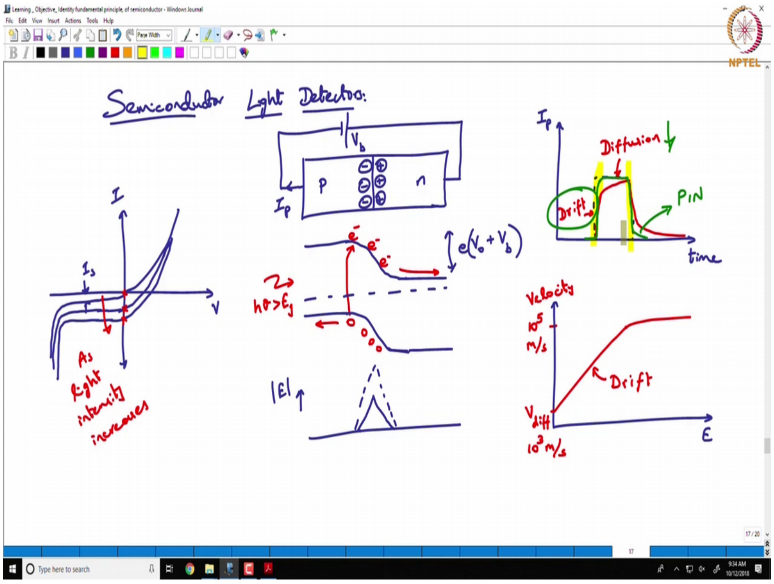
$R = \frac{\eta e}{h\nu}$ A/W)

$$\boxed{R = \frac{\eta\lambda(\mu m)}{1.24}}$$

Let us just go back what we were discussing her so 1550 is probably here, 1300 is probably here so we are just riding this curve over here okay. Any questions about this?

(Refer Slide Time: 16:44)





Then we will go to this other topic which is the bandwidth of the PIN junction diode it is not something that we talked about yesterday but if we consider a PIN structure you can tell that the response of the PIN structure is going to be limited by 2 different quantities right, one is that you may actually generate some carriers in the P region itself or initial part of the I region and as far as electrons is concern in that case the electron have to transfer across this entire structure, so we are talking about electron hole pairs that is generated. In electron hole pairs that is generated over here and we know that the holes are going to move this way.

The electrons are going to move this way but the electron would have to transfer that entire structure before it generates this external photo current, so in that sort of scenario you have a certain finite amount of time for transiting across that structure, so that is going to be the limitation as far as bandwidth is concerned. It will not be like an instantaneous response, it will be a response that is delayed by... the longest delay that you are going to see is corresponding to the case where electrons is going to have to transfer that entire structure right.

So that transit time will be one limitation, the other limitation is like we talked about over here when we have a pulse incident on this photo diode you want this to immediately rise up and when the pulses going off you want it to immediately come down, right and there is actually one…we already said that the transit time is going to be a limitation in terms of generating the photo current but the other limitations also that the PIN diode is going to have finite assistance provided because these are not infinitely conducting so they have a certain resistance and they have certain capacitance okay.

So what do we mean by the capacitance it is basically saying it is a structure that stores carriers okay, stores charges and there could be certain delay involved in sweeping out those carriers which we can actually attribute to a certain capacitance of the junction of a, so coming back here when we are talking about a bandwidth we will have to look at the response time of your PIN diode and that response time is going to have a transit time component and a component that is given by tau RC which is RC times constant, so we know that the transit time will correspond to transiting across this entire structure, so this is of with W.

So you can write this as W over the velocity right and the velocity typically that we are considering is the drift velocity which depends on the external electric field. The external electric field is 0 then VDR becomes…that will be their diffusion velocity, so you are essentially if there is no external bias applied then that velocity will be the diffusion velocity. Okay and so you know that your diffusion velocity is typical in the order of 10 power 3 metres per second whereas the drift velocity can be as high as 10 power 5 metres per second, so there is 2 orders of magnitude difference.

So depending upon the external bias applied at the response of your photo diode can be orders of magnitude different okay. So the response actually can be made quicker with the external bias you understand that, right that is one important thing to realize and in the other cases tau RC which depends on the series resistance of the junction multiplied by the capacitance, series resistance you know works out to be in the order of few ohms and then the capacitance is given by Epsilon a over d right. In this case if you look at the structure, the electrodes are over here it is between this electrodes and this electrodes that we are considering, so what is d here? That corresponds to the width of this entire furniture which you can approximate as W, so can write it as Epsilon a over W, what is a here?

Student: (())(23:37).

Prof: Right that corresponds to…if you were to put it in dotted lines that corresponds to the cross-sectional area over which the carriers are going across right so that corresponds to the area over here. So there is actually a very nice engineering aspect over here, you can see that W happens in the numerator the 1st term and W happens in the denominator in the 2nd term, so that is a contradiction right, so if you were trying to change W to change the response time you are typically you know when you go for smaller w right as far as your transit time is concerned you have a larger capacitance associated with that, so what do you do?

How do you get around this log jam. The free parameter is a right, a does not affect the transit time okay, so if I want to keep this RC component very low in the problem that you have given your been asked to assume that the RC time constant is negligible right. How do you make RC time constant negligible? You can do that by keeping the capacitance very low, how do you keep the capacitance very low? By playing around with the cross-sectional area.

So what you would find is if you look at commercial photo diodes you have these latively run off the mill commonly available photodiode which are like large area 3 millimetre, 5 millimetre active region will be there and when you look at the response time for that, that will be in the order of microseconds or somewhere around that. Whereas you have these specific photo diodes which are sold as high-speed photo diodes. When you look at the active area of high-speed photo diodes you will find that there are tens of microns in diameter, why? Because they are trying to play around with this RC time constant.

For low speed photo diodes you do not worry about the capacitance, it can be whatever and so you do not…you try to keep the active area larger but when you want to go to high-speed photo diode it is imperative that you keep your active diameter small, if you keep the active diameter small, what does that mean from the responsibility perspective? It is harder to focus all that lights on to that small diameter.

So you may take a hit in the responsivity, not all the photons are being absorbed by a photo diode, right if you cannot let us say 50 microns is your active diameter if you cannot focus all you light to 50 microns then you lose some photons without even…they are not even entering that photo diode structure okay so there may be a trade off in the responsivity, you understand this? Light is falling from here, right so if this is the active area that we are talking about, that facet area right, if that area is small then you are going to have to focus all your light into that.

Student: (())(27:51)

Prof: Yes, transit time depend upon this W.

Student: (())(27:57).

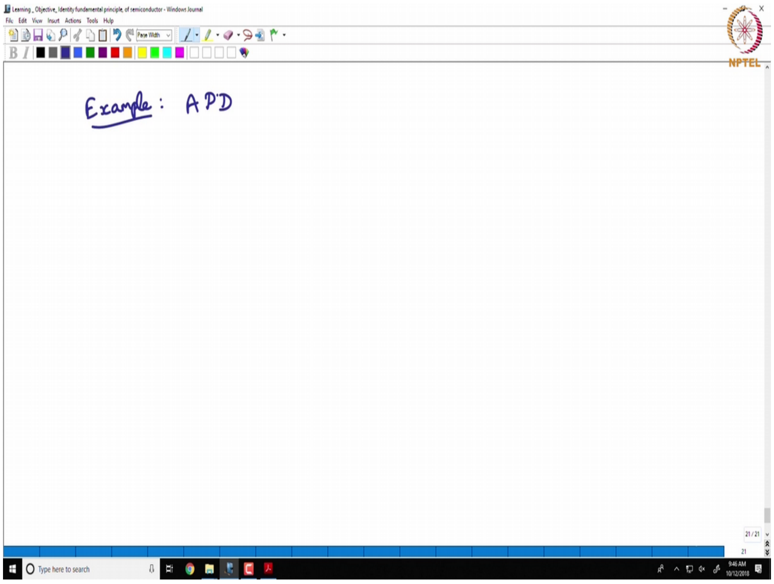
Prof: Response time, well so that is what you are saying response time, in the case when you have negligible capacitance as in the case that we have been asked to assume, you can basically neglect this by reducing active area, right by reducing the active area you can neglect that part.

Student: (())(28:36)

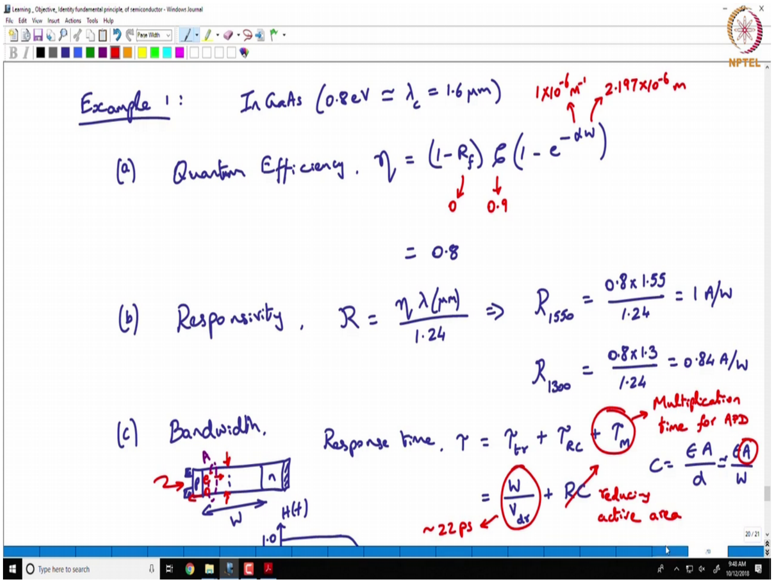
Prof: Area a is the area of cross-section cross that junction, so that is actually the area of the facet right this facet from which light is entering. Okay so response time is like that and if the response time is like that then if you plot the modulation transfer function h of f what you will find is it is flat at low frequencies and it drops down as you go to high-frequency because it is limited by the response time and what we are typically interested in is the 3DB bandwidth so f3db is going to be given by 1 over 2 pi multiplied by this tau which corresponds to the response time. In our case for this particular problem we have been ask to assume WS 2.17 microns and VDR to 10.5 metres per second.

So transit time component you will find is for this particular case that we are given it is roughly about 22 (())(30:35) seconds and if you substitute that back in this expression what you will find is this correspond to 7.23 gigahertz, so the bandwidth is 7.23 gigahertz for the particular example that we are given like I said to get to that bandwidth we are neglecting the RC time constant which means that your area if you go back and work out that area that you need to have so that RC time constant is negligible you will find that that area would correspond to something in the order of tens of microns square okay understand this? Now I do not have a lot of time but let me just quickly walk out this other example.

(Refer Slide Time: 31:47)



Example : APD



Example 1 : InGaAs $(0.8\,eV \simeq \lambda_c = 1.6\,\mu m)$ $\quad 1\times10^{-6}\,m^{-1} \quad 2.197\times10^{-6}\,m$

(a) Quantum Efficiency. $\eta = \underset{0}{(1-R_f)}\ \underset{0.9}{\zeta}\left(1-e^{-\alpha W}\right)$

$= 0.8$

(b) Responsivity, $R = \dfrac{\eta\lambda(\mu m)}{1.24} \Rightarrow R_{1550} = \dfrac{0.8\times1.55}{1.24} = 1\,A/W$

$R_{1300} = \dfrac{0.8\times1.3}{1.24} = 0.84\,A/W$

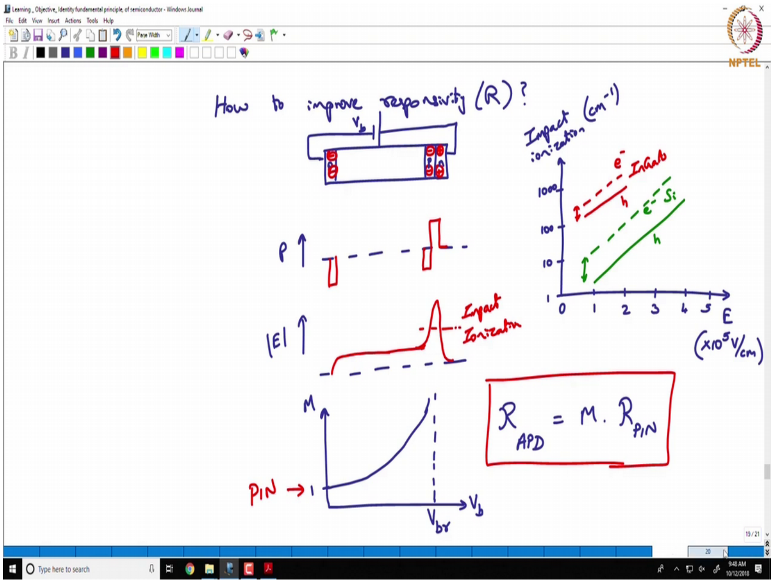
(c) Bandwidth.

Response time, $\tau = \tau_{tr} + \tau_{RC} + \boxed{\tau_M}$ → Multiplication time for APD

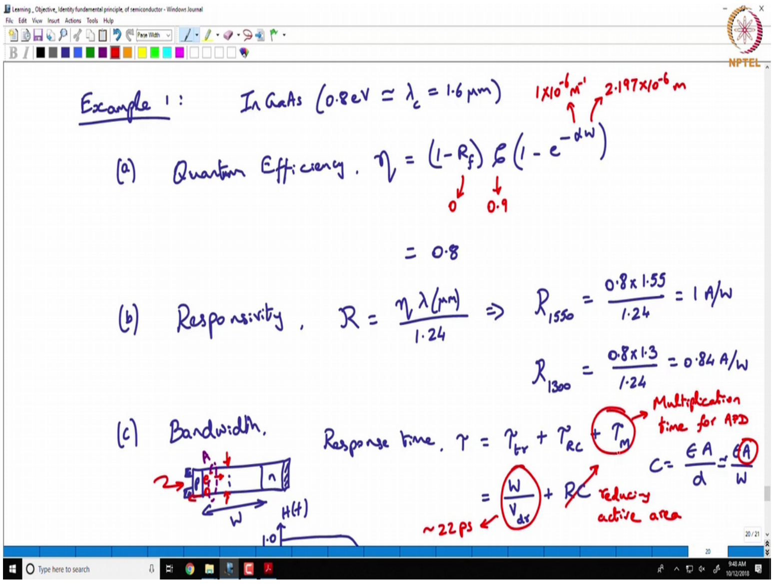
$= \boxed{\dfrac{W}{V_{dr}}} + \cancel{RC}$ reducing active area

$\sim 22\,ps$

$C = \dfrac{\epsilon A}{d} = \dfrac{\epsilon \boxed{A}}{W}$



$$\boxed{R_{APD} = M \cdot R_{PIN}}$$



Example 1 : InGaAs $(0.8\,eV \simeq \lambda_c = 1.6\,\mu m)$ $1\times10^{-6}\,m^{-1}$ $2.197\times10^{-6}\,m$

(a) Quantum Efficiency. $\eta = (1-R_f)\,\zeta\,(1-e^{-\alpha W})$ $R_f \to 0$, $\zeta \to 0.9$

$= 0.8$

(b) Responsivity, $R = \dfrac{\eta\,\lambda(\mu m)}{1.24} \Rightarrow R_{1550} = \dfrac{0.8\times1.55}{1.24} = 1\,A/W$

$R_{1300} = \dfrac{0.8\times1.3}{1.24} = 0.84\,A/W$

(c) Bandwidth. Response time, $\tau = \tau_{tr} + \tau_{RC} + \tau_M$ → Multiplication time for APD

$= \dfrac{W}{V_{dr}} + RC$ ~22 ps RC ↗ reducing active area

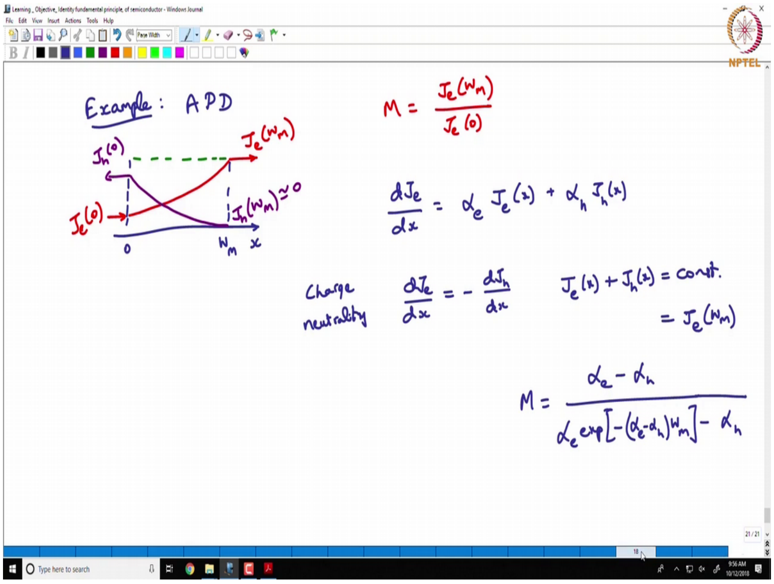
$C = \dfrac{\epsilon A}{d} = \dfrac{\epsilon A}{W}$

I will just give you a lead for that other example, example 2 let us consider design of an avalanche photo diode okay, so while we are at this topic I should also mention that if you are talking about avalanche photo diode you are just having one more layer added near the n region but in terms of the response time this response time will be there because this corresponds to the transit time of these carriers across the entire region in the RC time constant so what you will have is this extra term over here which corresponds to the multiplication, so since impact ionisation is sort of a serial process.

So you have one electron with high velocity coming and hitting an atom and ionising that atom and then whatever is scattered from the can go and hit another atom and ionise another atom and so on, so it is actually serially multiplying a right and so that will have certain multiplication time so this factor you need to add for APD based on this you can tell that APD response is going to be always slower compared to the PIN response right you understand that.

So because there is an additional multiplication time that happens in the APD and if you go back and look at what we discussed yesterday the APD without much bias actually has M of 1, so effectively at this point it behaves like a PIN, so take an (())(34:11) when bias voltage is 0 that effectively acts like a PIN (())(34:18) okay by concluding the external bias you are actually controlling the multiplication factor and that will incur a certain extra time right and that will limit the bandwidth of your (())(34:41).

(Refer Slide Time: 34:49)



Example : APD

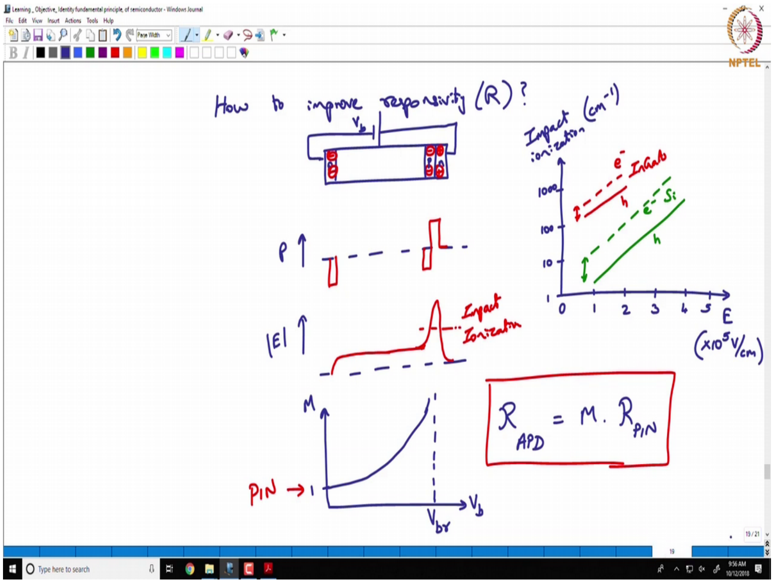
$$M = \frac{J_e(W_m)}{J_e(0)}$$

$$\frac{dJ_e}{dx} = \alpha_e J_e(x) + \alpha_h J_h(x)$$

Charge neutrality $\frac{dJ_e}{dx} = -\frac{dJ_h}{dx}$

$$J_e(x) + J_h(x) = \text{const.} = J_e(W_m)$$

$$M = \frac{\alpha_e - \alpha_h}{\alpha_e \exp[-(\alpha_e - \alpha_h)W_m] - \alpha_h}$$



How to improve responsivity (R)?

$$R_{APD} = M \cdot R_{PIN}$$

Let us just consider this region where the multiplication is happening and let us say this is a long (())(34:57) across that multiplication region this is 0 over here and this is Wm that is the width of the multiplication region okay. What do you know the left side of that location (()) (35:21) intrinsic region what is the right side of the Wm region? That is the M region okay, so just give you a perspective of where we are.

Now you can assume that we do not have any holes coming from this side right so because only absorption happened in the I region so let us say we do not have any more absorption happening beyond that, so there are no holes coming from this side but there is actually a finite current let us call it current density Je of 0 coming here and this is going to go through multiplication and what we want to figure out is the multiplication factor but the multiplication factor that we want to figure out is Je of Wm by Je of 0 okay you want to get an expression for that and we know that as far as holes are concern they are actually 0 here right so they go out as Jh of 0 here and here is Jh of Wn which can be approximated as 0 let us say we do not have any holes coming from the m region do not have any absorption happening there okay.

So in this scenario we want to find out the expression for m. To find that we will have to look at what is the rate at which things are changing across this junction, so let us say you have dJe over dx is what we want to find out okay, so what all contributing to this there could be impact ionisation due to electrons okay, so you have Alpha e corresponding to the impact ionisation coefficient for electrons multiplied by Je of x right, so that is one contribution.

Now the other contribution as you could have holes that are generated from the first impact ionisation, they could also participate if they have sufficient velocity, they could also participate in generating other electrons holes pairs, so you have Alpha h and Jh of x right that is in a general case. We could have electrons generated due to impact ionization by holes. Now we need to solve this and get that expression for the multiplication factor but solving that we can have some assumptions.

First of all you have charge neutrality, charge neutrality says that you are always generating electrons and holes as pairs, we have generated an electron there is corresponding holes also there okay so I can just say that dJe we are going to like differential of that current density corresponding to electron is going to be minus of dJh over dx which is saying that electrons are moving one way the holes are moving the opposite way right, that is one condition and you can also say that because they are always done in pairs Je of x plus Jh of x is a constant across this region right and it just call that total current density, so the total current density is constant across the entire region. If that is the case what is that value?

That value would have to be this because Jh of Wm is 0, so that constant and I can write as equal to Je of Wm because Jh of Wm is 0 right, so those are the things and of course we are assuming Jh of Wm is 0, so based on this you can get multiplication factor, I will just put down the final value, final expression and I will let you actually do this yourself final expression that you will get is alpha e exponential of minus Alpha h minus h, so that is the final expression we get and this expression if you plot that you will get something like this is m as a function of (())(42:02) it will get something like this which is shown exponential. Let me stop at this point.