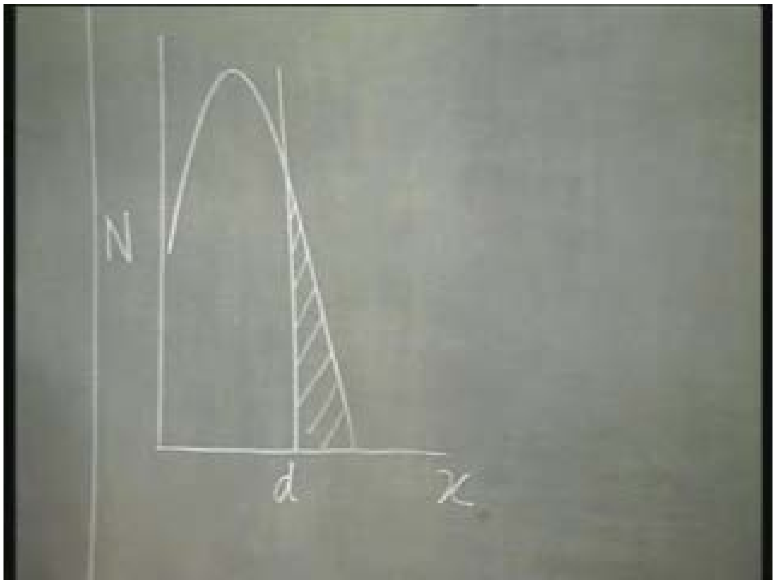
# VLSI Technology
**Dr. Nandita Dasgupta**
**Department of Electrical Engineering**
**Indian Institute of Technology, Madras**

## Lecture - 22
## Masking during Implantation + characterization of doped layers

We have covered most of the aspects of ion implantation. You know how ion implantation is to be done, how the doping peak concentration, junction depth, etc., can be varied by varying the dose and the energy; you know about the basic ion implantation system. But remember, whenever you are doping in a semiconductor, doping is done selectively. That is some portions have to be protected from the doping by use of masking. At the beginning of our discussion about ion implantation, I mentioned that ion implantation is a room temperature process and therefore you have a larger choice of mask material. You do not have to use silicon dioxide always, like in case of diffusion. Diffusion by contrast, is a high temperature process; therefore you must use a masking material that can withstand this high temperature, right.

But, in case of ion implantation you can even use photoresist. Photoresist will not withstand high temperature, but since ion implantation is a room temperature process, you can use photoresist as the mask. But the point is, even though some portions of the semiconductor is covered with this mask material, when the ion beam is incident on the semiconductor it is incident over the mask also and you see, these ions are energetic particles, so they will penetrate the mask material, agreed.

(Refer Slide Time: 3:15)

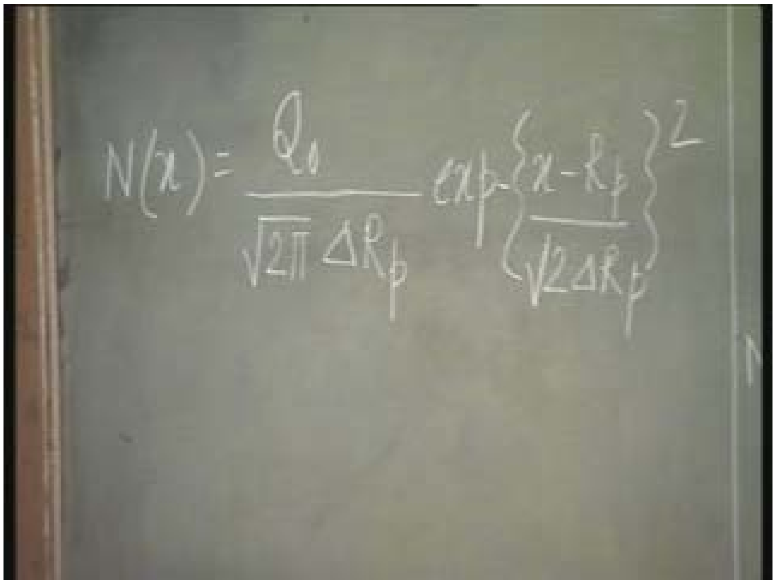


In other words, if I look at the mask material alone, then just as in the silicon inside the mask also, the energetic ions will have a Gaussian profile. No matter what the material is, the energetic ions will anyway have a Gaussian profile; I can approximate it to be a Gaussian profile. Of course, the range and the straggle will depend on these two particular material combinations, which may be different from silicon to the dopant combination. But it will have a particular R p and delta R p and it will follow a Gaussian profile. So, inside the mask also will have a profile like this; a Gaussian profile, with a peak and a standard deviation, agreed.

Now, this masking material has a particular thickness. Whether you have used silicon dioxide or photoresist, you have put a particular, a layer of some thickness on top of semiconductor, isn’t it. So, if let us say, the thickness of the masking material is something called d and if this is the value of d that is this is where the mask material ends, below this is semiconductor; then this much of the impurity is going to be put in the semiconductor, agreed. In other words, this is quite obvious that the masking efficiency will depend on the thickness of the masking material you have put. If d is smaller, then more impurity will be put in silicon, if d is larger less impurity will be put in silicon and since we want those regions not to have any impurity, we will say that the masking efficiency is large, when d is large.

Now, how exactly does the masking efficiency depend on the thickness of the mask layer? In other words, if I want to protect certain regions from getting an implantation dose, what thickness of the mask layer should I choose? I know what is going to be the dose and energy of the implantation. According to that what should be my masking layer thickness?

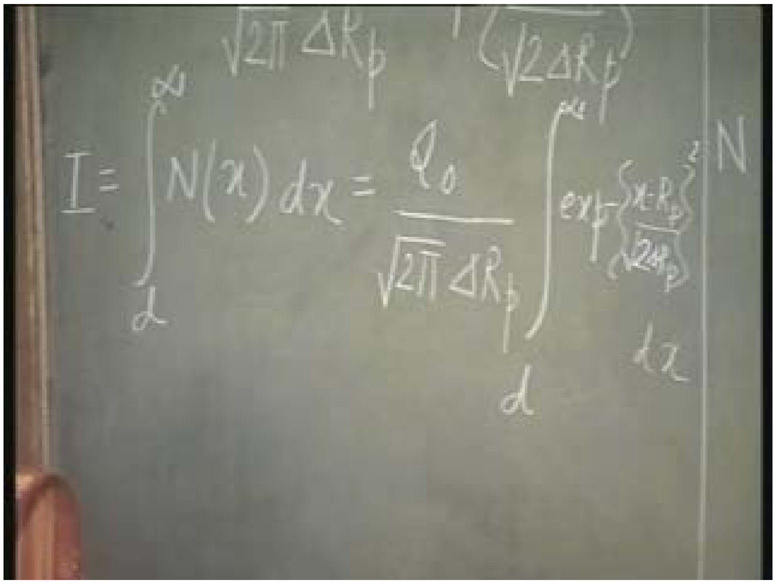
(Refer Slide Time: 6:52)



$$N(x) = \frac{Q_0}{\sqrt{2\pi}\,\Delta R_p} \exp - \left\{ \frac{x - R_p}{\sqrt{2}\,\Delta R_p} \right\}^2$$

Now, you know that the implantation doping profile is given simply by …… isn’t it. This Gaussian distribution is for the implantation doping profile that is this curve. Then, what is the value of this hatched region? What is this hatched region?

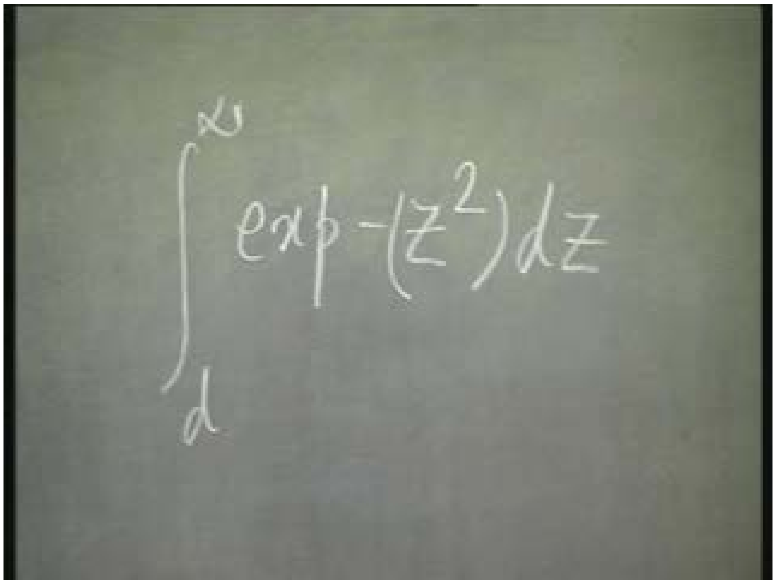
(Refer Slide Time: 7:50)



$$I = \int_{d}^{\infty} N(x)\,dx = \frac{Q_0}{\sqrt{2\pi}\,\Delta R_p}\int_{d}^{\infty} \exp -\left\{\frac{x-R_p}{\sqrt{2}\,\Delta R_p}\right\}^2 dx$$

This hatched region is nothing but, see, if I integrate from zero to infinity that will give me the area under this total curve. So, if I want to get only this portion, then I must integrate from d to infinity. That is what I have done. If I integrate N x from d to infinity, then I get the area, the hatched region. This is the total impurity which is put in the silicon, which is no longer protected by the mask, since the mask ends here at the point d; so, anything beyond d is going to be inside the semiconductor, right. So, this can be written as Q 0 by root 2 pi delta R p integral d to infinity exponential …..

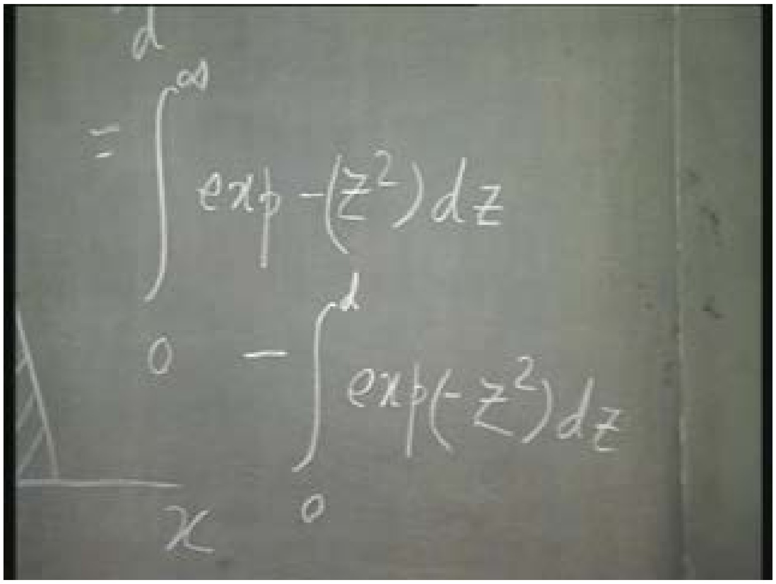
(Refer Slide Time: 9:46)



$$\int_d^\infty \exp -(z^2)\,dz$$

Now, let us see. What I have is of this form, integral something to infinity, d to infinity exponential minus z square dz.

(Refer Slide Time: 10:18)

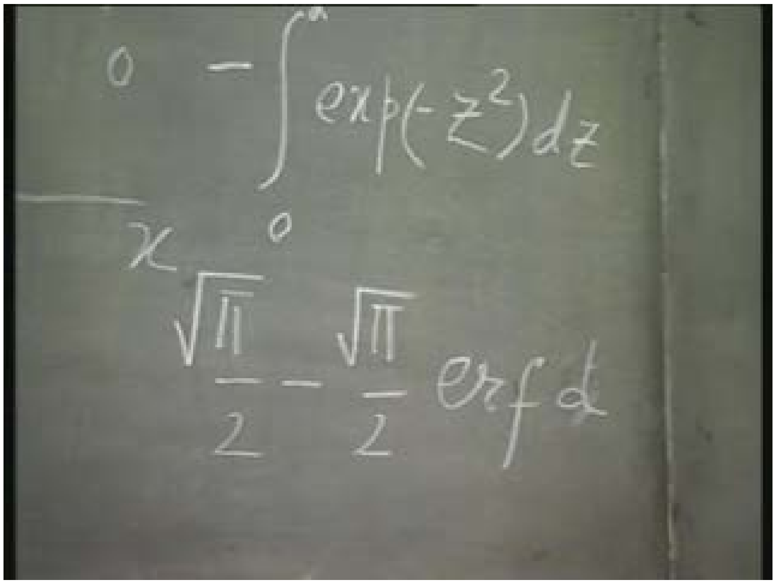


d

$$= \int_0^\infty \exp -(z^2)\,dz - \int_0^d \exp(-z^2)\,dz$$

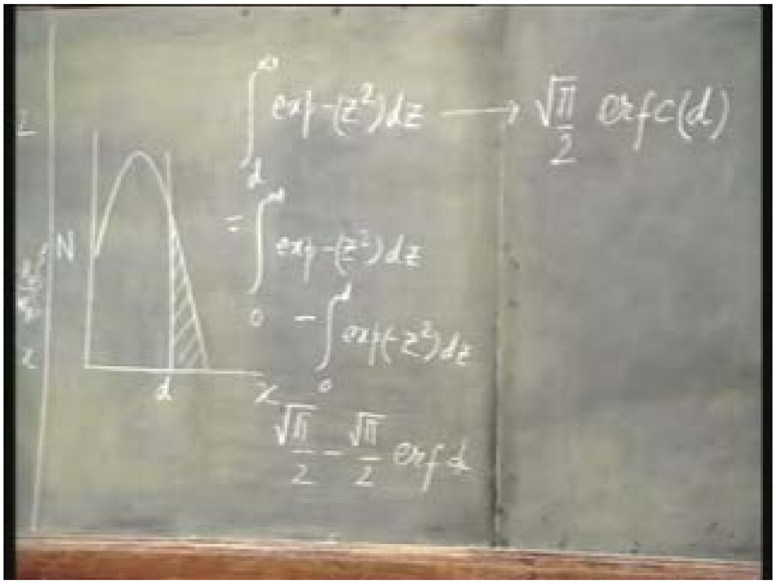
I can write it as ……, can I not? In other words, this hatched region is this total minus this portion, agreed. This is a definite integral with a value of root pi by 2.

(Refer Slide Time: 11:06)



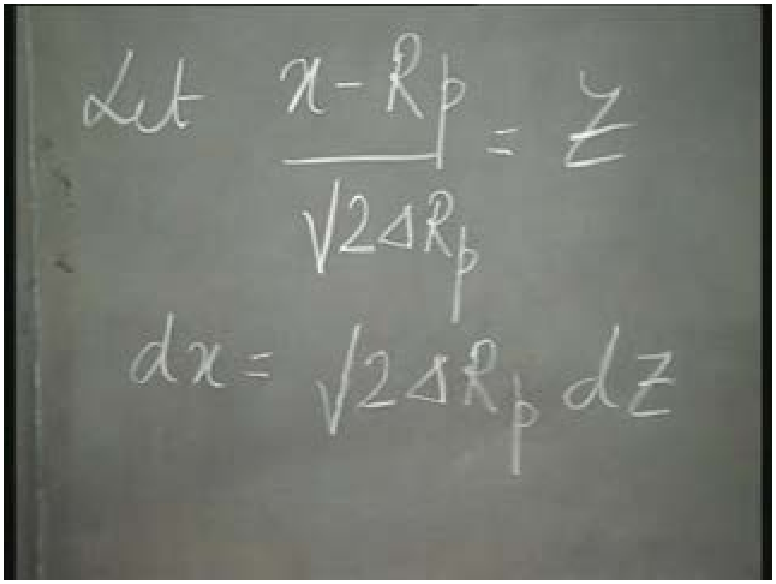
So, this can be written as …. and what is this? Integral zero to d exponential minus z square dz, isn't it? Root pi by 2 error function of z, sorry error function of d, right.

(Refer Slide Time: 11:56)



In other words, this can be written as root pi by 2 times 1 minus error function of d that is error function complement of d, right. So, if I have an expression like this, integral d to infinity exponential minus z square dz, this can be given as root pi by 2 error function complement of d, agreed. Now, let us look back at this. If I have to express the integral in this form, then I have to change the variable.

(Refer Slide Time: 12:55)

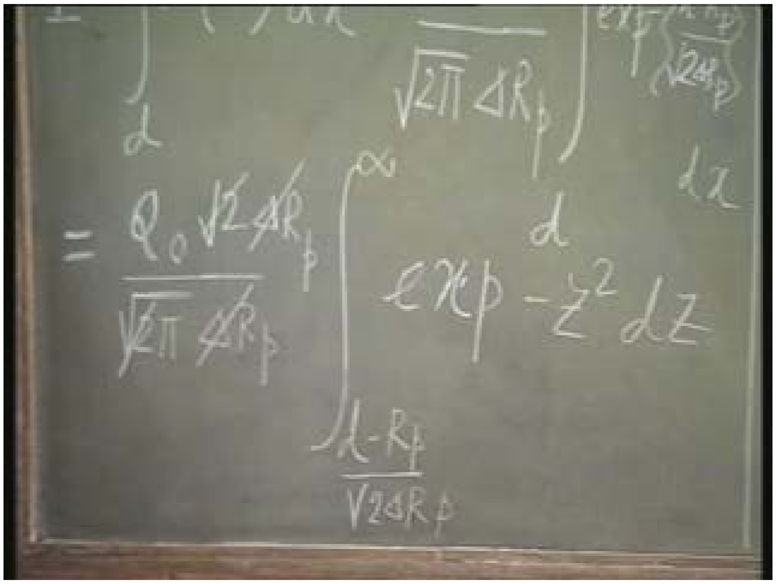


Let $\frac{x-R_p}{\sqrt{2}\Delta R_p} = Z$

$dx = \sqrt{2}\Delta R_p \, dZ$

If I have to get it in z square fashion, then I have to write let x minus R p by root 2 delta R p be equal to z. If that is so, then ….., right.

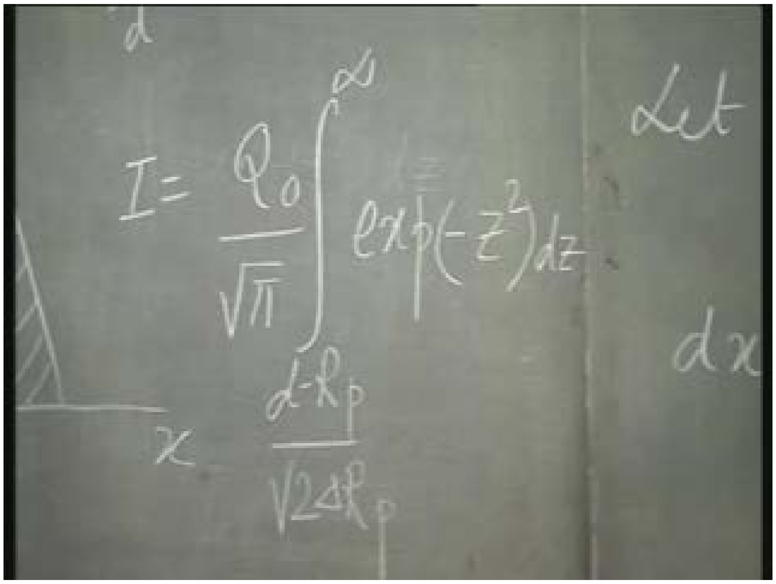
(Refer Slide Time: 13:35)



$$= \frac{Q_0 \cancel{\sqrt{2}} \cancel{\Delta R_p}}{\sqrt{\cancel{2}\pi} \cancel{\Delta R_p}} \int_{\frac{d-R_p}{\sqrt{2}\Delta R_p}}^{\infty} \exp -z^2 \, dZ$$

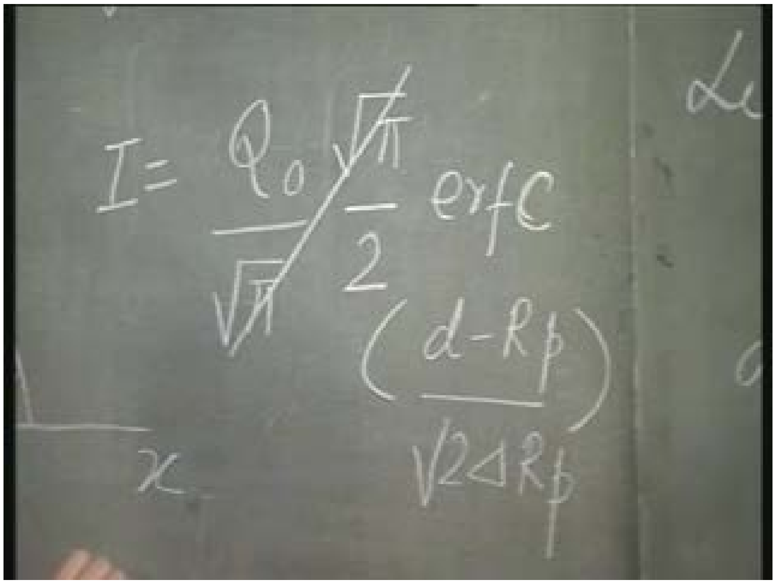
Then, this integral becomes ….., right. But, do not forget, it is a definite integral. When I change the variable, I must also change the limits. So, x was to be integrated from d to infinity, now z that is x minus R p by root 2 delta R p must be integrated from d minus R p by root 2 delta R p to infinity, isn't it. When x becomes equal to d, in that limit z becomes equal to d minus R p by root 2 delta R p.

(Refer Slide Time: 15:21)



So, I can cancel out delta R p, can cancel out root 2, so that I is Q 0 by root pi integral …… This is my I. Now you know, integral d to infinity exponential minus z square dz is going to be root pi by 2 error function complement of d. So, this can simply be written as…..

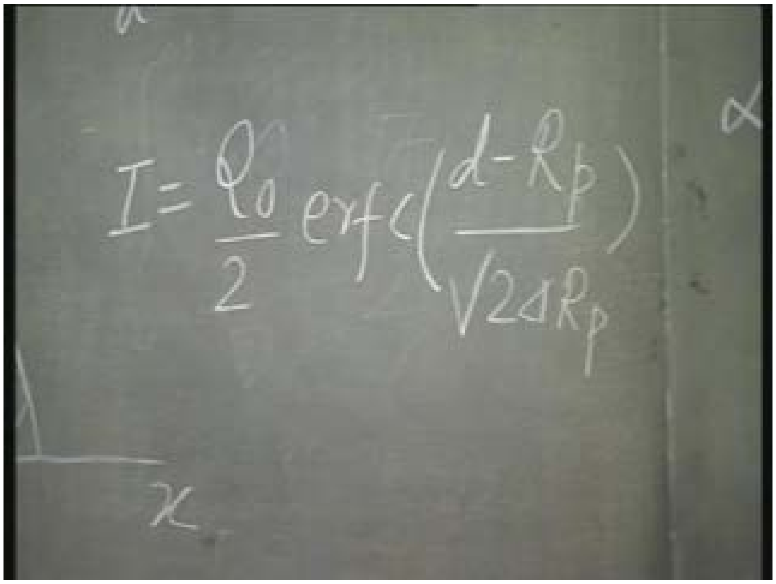
(Refer Slide Time: 16:12)



$$I = \frac{Q_0}{\cancel{\sqrt{\pi}}} \frac{\cancel{\sqrt{\pi}}}{2} \operatorname{erfc} \left( \frac{d - R_p}{\sqrt{2} \Delta R_p} \right)$$

Or in other words, I is given as …..

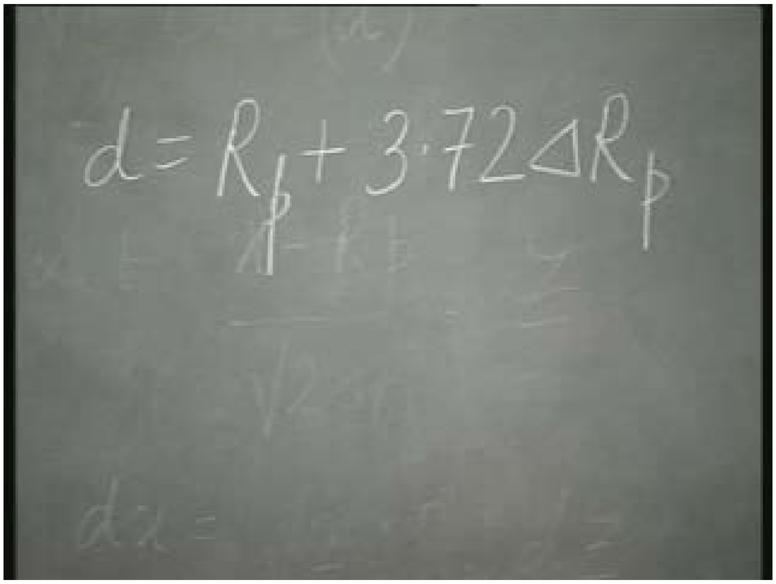
(Refer Slide Time: 16:38)



$$I = \frac{Q_0}{2} \operatorname{erfc} \left( \frac{d - R_p}{\sqrt{2} \Delta R_p} \right)$$

This is the amount of the impurity which is not protected by the mask, which is going inside the semiconductor. This is the amount Q 0 by 2 complementary error function of d minus R p by root 2 delta R p. So, this is all mathematics. Let us look at it physically, now. What do I mean really? What I mean is very simple.

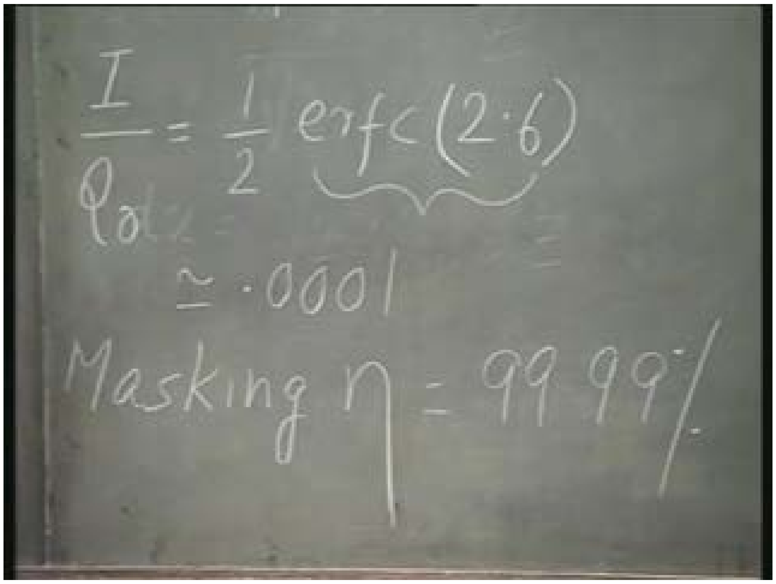
(Refer Slide Time: 17:58)



$$d = R_p + 3.72\Delta R_p$$

That means if I have a mask layer thickness, d greater than R p by a certain extent, let us say if I put, I have just pulled out a number from somewhere. All I am trying to say is the masking layer thickness is greater than the projected range, by a multiple of delta R p.  It is nearly, it is greater than R p by nearly 4 times delta R p.

(Refer Slide Time: 18:46)



$$\frac{I}{Q_0} = \frac{1}{2}\underbrace{\text{erfc}(2.6)}$$

$$\simeq .0001$$

$$\text{Masking } \eta = 99.99\%$$

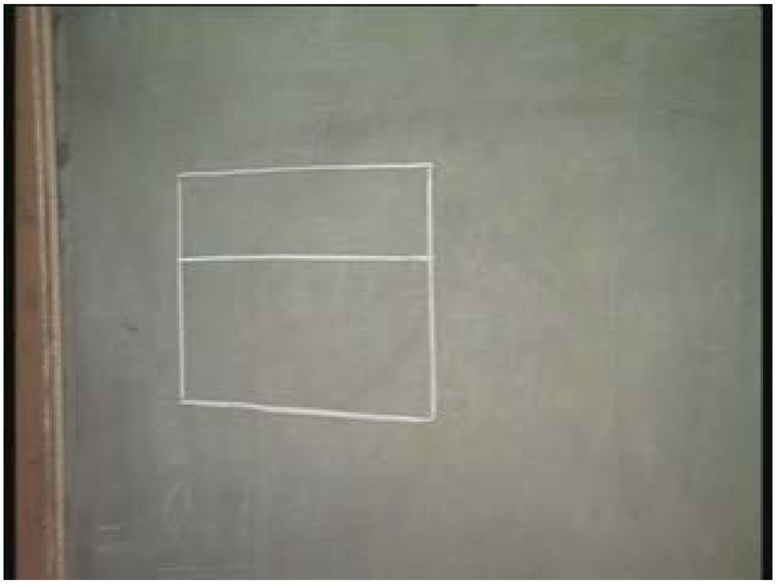
If that is so and if you put this value here, then you can say that this I that is the total amount of impurity which is going into the semiconductor by the total dose of impurity, remember Q 0 is the total dose that you are putting, I by Q 0 is going to be half error function complement of 3.72 divided by root 2, which works out to be about 2.6 and you can verify from the error function table that this multiplied by half will work out to be about 0.0001. That means I by Q 0 is only 0.0001 that is point zero zero zero oneth part of the total dose is going into the semiconductor. In other words, we will say that the masking efficiency of such a layer is 99.99% that is the masking efficiency. So, this corresponds to the masking efficiency; this is the accepted symbol for efficiency, masking efficiency is 99.99%, right. So, you see, if you know R p and delta R p, you can always choose a masking layer thickness, so that most of the impurity is masked by the layer. Obviously, R p is going to be dependent on the energy of implantation. So, those are the factors you know. So, accordingly you can select the thickness of the masking layer, right.

So, now you have your doped layer whether by ion implantation or by diffusion, you have been able to create to a doped region in semiconductor. Now, you must evaluate this doped layer. That is, so far we have discussed about the theory; theory of diffusion, theory of ion implantation, what the profile should be, where the junction should be situated, right; what should be the surface concentration, what should be the peak concentration. These are all the theory. You must check whether after the implantation or after the diffusion it follows the expected value that is you must have an experimental verification. In other words, you must be able to evaluate the doped layer.

So, if I want to evaluate the doped layer, what are the factors I want to know? First and most important is to know about the junction depth, right. Junction depth is a very important parameter. I must know after the implantation or after the diffusion, where the junction is situated. So, number one is junction depth. Number two what do I want to know? I want to know what about the doping profile, what about the surface concentration? How does the concentration change from surface to bulk that is the doping profile and what is the surface concentration, right? So, you would want to know about the junction depth and you would want to know about the surface concentration as well as about the doping profile.

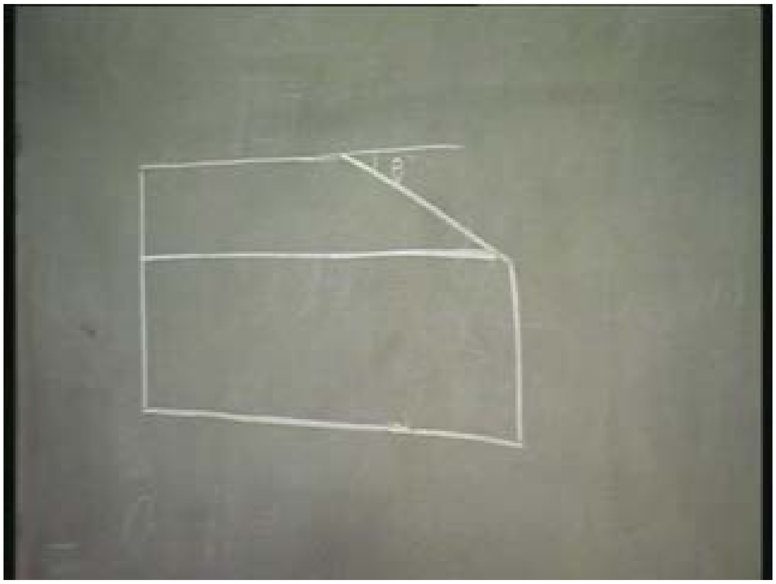
How does one evaluate these things? It is going to be our next topic of discussion - evaluation of the doped layer. Let us first talk about the junction depth. How does one measure the junction depth?

(Refer Slide Time: 23:56)



Well you have, this is the cross sectional diagram of your junction. This is p, this n or this in n, this is p. You want to measure the junction depth. One very come way of measuring the junction depth is by lapping and staining. What is done is the junction is angle lapped. What do I mean by angle lapping?

(Refer Slide Time: 24:37)

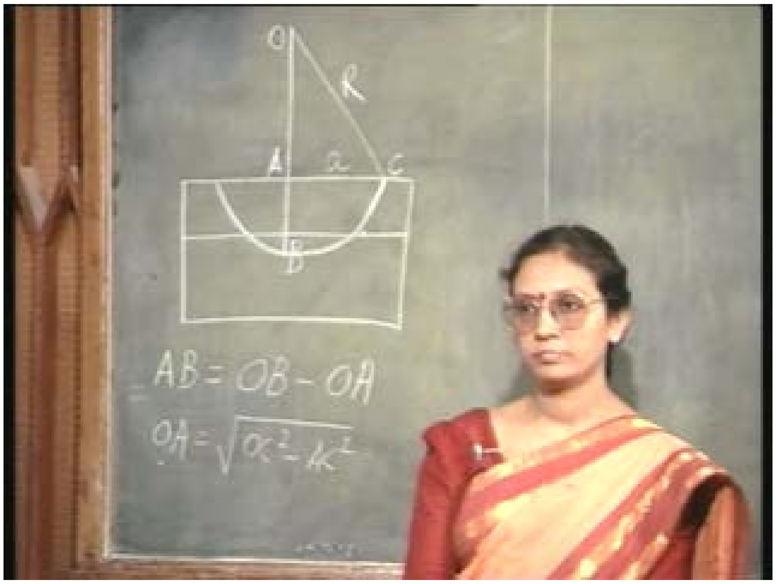


This is angle lapping. Actually this is done at a very small angle, so that it looks somewhat like this. Why is it done? This angle lapping is done to magnify the junction. You see, if you do not do any angle lapping, the junction that you are seeing is this. But, if you angle lap it, you are extending this portion and if you know this angle, this is your theta, which is usually half degree or so and if you can measure this, then by trigonometry you will be able to find what this is, right. So, it provides an ease of measurement. You are visually magnifying the junction area.

So, first you do the angle lapping. Then, you stain it using a copper sulphate and dilute HF solution. Dilute HF will etch the surface oxide and copper sulphate is going to preferentially stain, it is going to preferentially stain p region over n region. So, now if you view it under the microscope, then you can measure this. This is the p region. This will be stained in preference, so you can measure the stained region and you know the theta, the angle. So you can find out what the junction depth is going to be. It is a very common way of measuring the junction depth, by angle lapping and polishing.

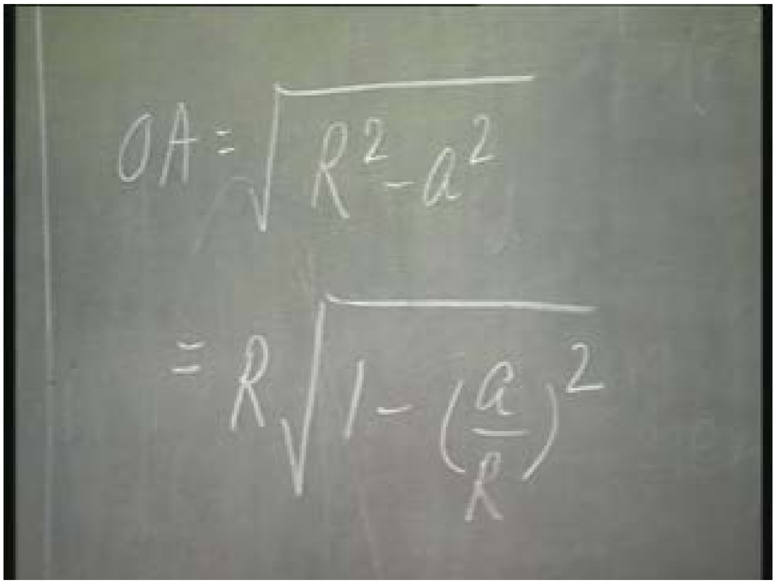
There is an even better, even more accurate method of doing that and that is called cylindrical groove technique. So, in cylindrical groove technique, what happens is you have cylinder with a radius, R and you use this cylinder to grind a groove in the semiconductor.

(Refer Slide Time: 27:43)



That is you have a cylinder with particular a radius R and you use it to grind a groove like this. This is R and this is the groove you have ground in the semiconductor. Remember this is a destructive technique. You are going to grind a groove, therefore you are going to destroy the surface. But, what you want to measure is the distance A B. A B is actually the junction depth, right. I can write depth A B as O B minus O A, right. Now, let me see. What is O A? If I look at this triangle O A C, then O A is given by square root of O C square minus A C square, Pythagoras theorem; I hope you member, right. You already know O C; O C is actually R, the radius of the cylinder.

(Refer Slide Time: 30:10)

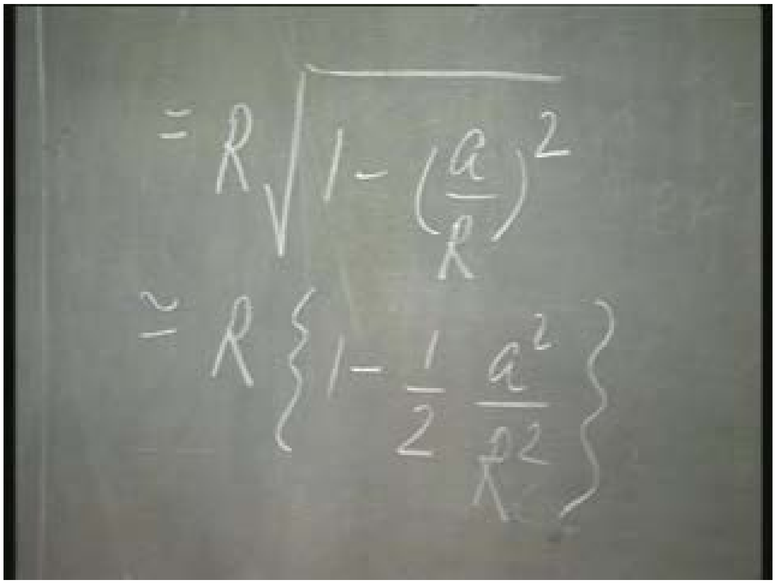


$$OA = \sqrt{R^2 - a^2}$$

$$= R\sqrt{1 - \left(\frac{a}{R}\right)^2}$$

Let me call this A C as small a, so that I can write that O A is equal to square root of R square minus a square, which can be written as ……

(Refer Slide Time: 30:38)

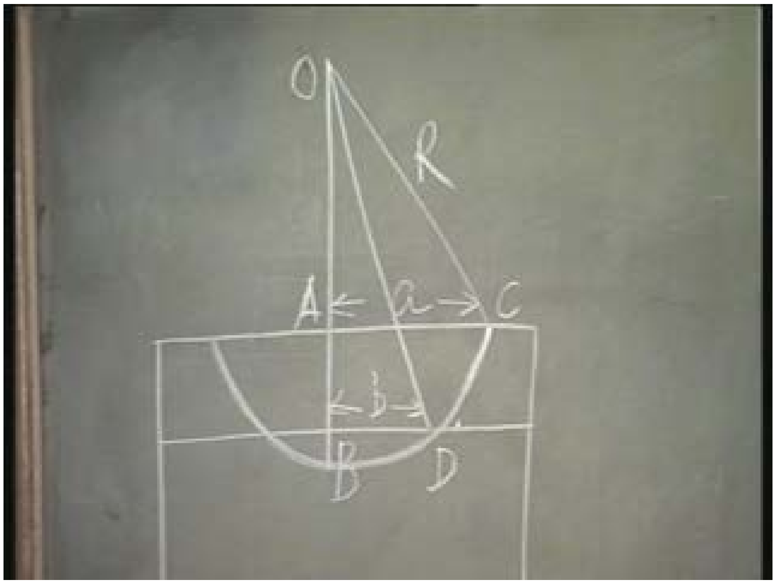


$$= R\sqrt{1 - \left(\frac{a}{R}\right)^2}$$

$$\simeq R\left\{1 - \frac{1}{2}\frac{a^2}{R^2}\right\}$$

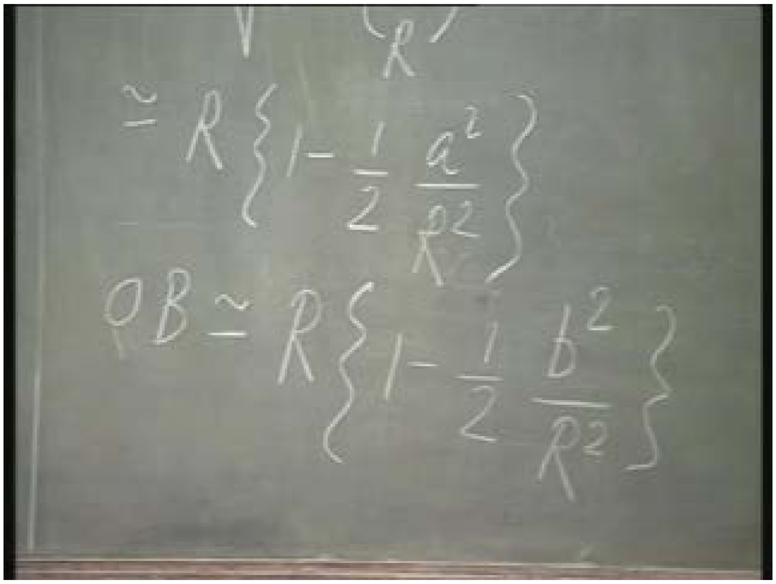
Now, if the radius, capital R is much larger than this small a, then I can approximate it to be ….., can I not? 1 minus x is, 1 minus x, if x is very much smaller than 1, that is a is much smaller than R, then I can approximate it to be 1 minus half x. That is what I have done.

(Refer Slide Time: 31:10)



Similarly O B can be given as square root of R square minus B D whole square in the triangle O B D, isn't it.

(Refer Slide Time: 31:53)

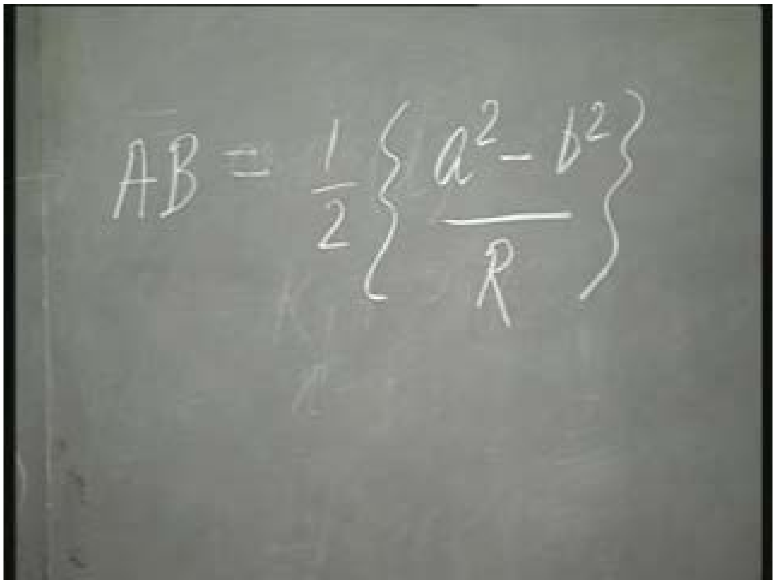


$$\simeq R\left\{1-\frac{1}{2}\frac{a^2}{R^2}\right\}$$

$$OB \simeq R\left\{1-\frac{1}{2}\frac{b^2}{R^2}\right\}$$

Then I could say, O A is given by this; O B is given by this, if I call this B D to be small b.

(Refer Slide Time: 32:25)



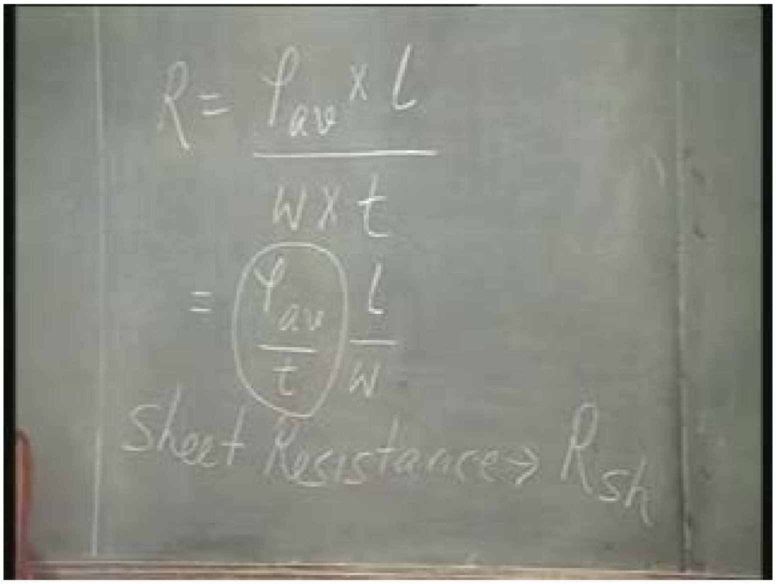
Then, my junction depth A B is given simply as the difference of O B and O A. So, you see, R gets cancelled out, what I have is a square minus b square divided by twice R, right. So, this is the cylindrical groove technique of measuring the junction depth.

There is another technique of measuring the junction depth and that is the interference fringe method. In the interference fringe method also what is done is you have a lapped sample; on top of that you provide an optical flat and you subject it to a monochromatic radiation, usually a sodium vapour lamp. So, you will see fringes. There will be dull fringes in the p region. You can count the number of fringes and that will be related to the junction depth. So, it is actually given as a function of the number of the dull fringes multiplied by the wavelength, wavelength of the monochromatic light. So, from that since you know the wavelength of the monochromatic light, you can find out what the junction depth is going to be. So, these three are the three most common techniques of measuring the junction depth.

Once you know the junction depth, next what you want to know is how about the doping distribution, the impurity distribution. Now you see, you can measure the total impurity distribution by doing some spectroscopy analysis like SIMS, secondary ion mass spectroscopy, which will tell you exactly how much impurity is put inside the material. But, it will not tell you whether this impurity is electronically active or not, whether it is sitting in the substitutional site or it is just sitting anyhow, anywhere inside the semiconductor. From the point of view of VLSI of course, you would want to know how much of the impurity is electronically active and you can measure it only electrically. By looking at the electronic activity of the semiconductor you can say how much of the impurity is electronically active. So, you have no go, but to look at the current-voltage characteristics.

Now, let us say you have a doped region, you have a doped region of a particular length, a particular width and you also know the junction depth, therefore it is of a particular thickness. I have already finished measuring the junction depth, you see, by using one of these three techniques.

(Refer Slide Time: 36:12)



$$R = \frac{\rho_{av} \times l}{w \times t}$$

$$= \boxed{\frac{\rho_{av}}{t}} \frac{l}{w}$$

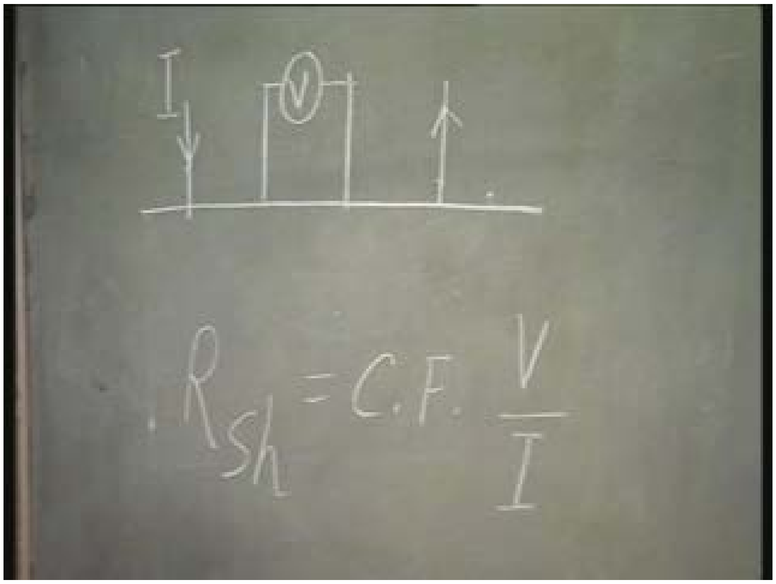
Sheet Resistance → $R_{sh}$

So, now if I look at this doped region which has a particular length, width and thickness, then the resistance of this doped layer, R will be given by its average resistivity into l that is the length divided by this is the cross sectional area, right, width into thickness; rho average into l by a that is going to be the resistance of this doped region. I can write this as ….. This rho average divided by thickness, this is called the sheet resistance, sheet resistance, R sh. Sheet resistance, what is the unit of sheet resistance? It is Ohms per square. That is if I take both l and w to be equal that is I am measuring in a square, then you know, resistance the unit is ohms, so it is ohms per square. That is the unit of the sheet resistance ohms per square and sheet resistance is a very important parameter.

In fact, in most of the processing steps, at the end of the, it is specified. Suppose you are doing a base diffusion, a base implantation, so it is specified that at the end of this process, the sheet resistance must be so much say, 200 ohms per square. So, what you have to do is to measure the sheet resistance, see whether it is conforming to your process specification which brings me to the question, how do I measure the sheet resistance? Of course from this expression you know it is very simple. How to measure the sheet resistance? You measure the current and voltage and then measure the number of squares and then you can find out what the sheet resistance is. But, more usually what is done is we use a four point probe technique in order to measure the sheet resistance.

What is a four point probe technique? It is exactly what the name suggests.

(Refer Slide Time: 39:04)



It is a probe with four points; you have four points placed on the surface. The outer two points are for current, current going in and current going out and the inner two points are for measuring the voltage. So, what actually happens is you have an instrument, a four point probe, which has four small tips like this. Take your sample and then these four tips are placed like this, touching the sample surface. The outer two, you connect it to the current source; you are sending the current through this and the inner two, you connect the voltmeter across it, so that you can measure the voltage across that and under this condition, the sheet resistance is directly given by….. You measure the voltage and the current. The sheet resistance is directly given as V by I times a correction factor.

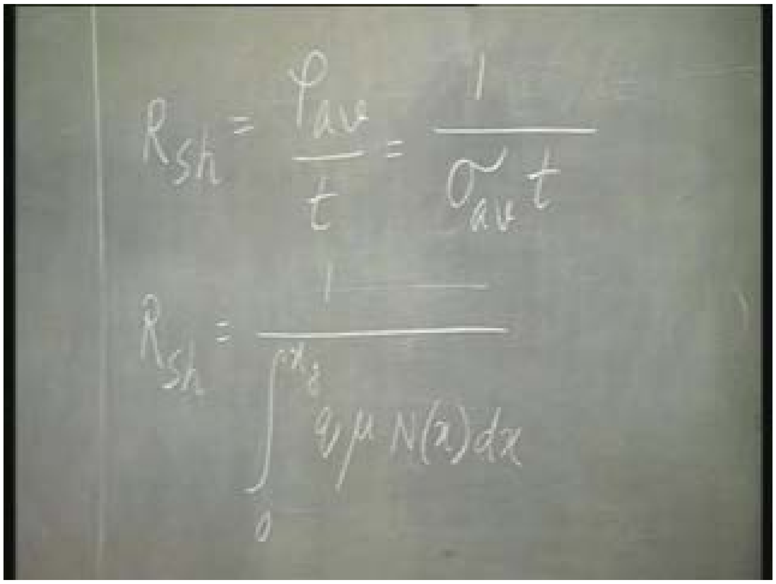
Now, this correction factor value tends to 4.53 to 4.5, provided this sample that you are using, it is a long sample. That is the sample length is much larger compared to the probe spacing. These four probes are spaced close together.

(Refer Slide Time: 41:02)

$$R_{sh} = C.F.\ \frac{V}{I}$$
$$\simeq 4.5\ \frac{V}{I}$$

If your sample length is much larger than this probe spacing, then this can be approximated as 4.5 times V by I, right. So, you have found the sheet resistance. Four point probe technique, it is a very common technique; you have found the sheet resistance. Now, as I said that in most cases the sheet resistance will be a parameter, so it has, if your process, after, at the end of process if you find that the sheet resistance conforms to your expected value, fine, you can go ahead. But, this sheet resistance can also be used directly to give the surface concentration. You can also get the surface concentration from the sheet resistance value. How can we get that? Well, we have already said that sheet resistance is given as rho average by t. Now let me try to write it in a slightly different manner.

(Refer Slide Time: 42:12)

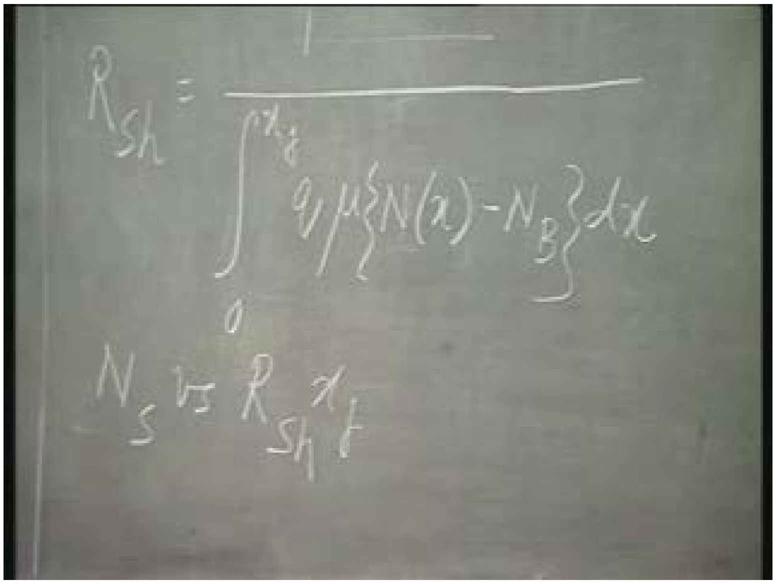


$$R_{sh} = \frac{\rho_{av}}{t} = \frac{1}{\sigma_{av} t}$$

$$R_{sh} = \frac{1}{\int_0^{x_j} q \mu N(x)\, dx}$$

I can write it as ….. or I can write it as 1 upon sigma average into t. That is the sheet resistance. Now, this sigma average into t can be written as ….. agreed. What is sigma average into t? It is actually the integral of N dx and of course, multiplied by q mu. You know, q mu N is actually the conductivity. So, over the entire doped region from zero to x j that is the thickness of the doped layer, over the entire doped region, if I integrate q mu N x, then I get sigma average into t, agreed, yes? I have of course here neglected the background doping concentration. What you should have here is the actual doping concentration that is the difference, N x minus N B. I have neglected N B thinking that N B will be much smaller.

(Refer Slide Time: 44:25)

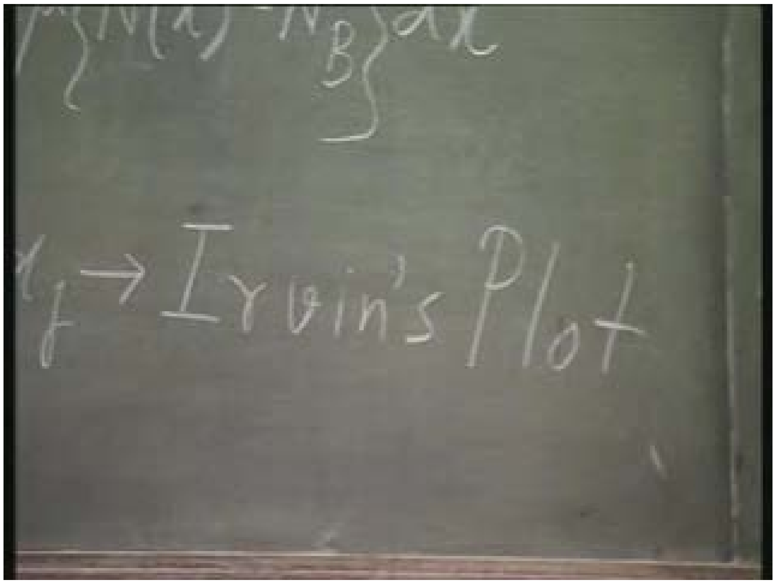


$$R_{sh} = \frac{1}{\int_0^{x_j} q\mu\{N(x) - N_B\}dx}$$

$N_s$ vs $R_{sh} x_j$

If I want to make it more accurate, then I should write it like this. So you see, I have measured the sheet resistance, I know the junction depth; I have already measured the junction depth. So, can I solve this equation and get the value of the surface concentration? I cannot do that simply because mobility is not really a constant, but mobility is also of function of the doping concentration. You know that as the doping concentration is increased, mobility is going to decrease, right. So, this equation, solving this equation is not a very straight forward thing, but it can be done. Computer derived solutions do exist for various profiles. That is if it is a complimentary error function profile or if it is a Gaussian profile and also for various background concentration, it is possible to obtain computer derived solution and we have a set of plots available, which plots N s as a function of R s times x j, for error, complementary error function profile as we well as for Gaussian profile and for various background concentration.

(Refer Slide Time: 46:22)



These plots are called Irvin's plot. So you see, since this Irvin's plot do exist and you have already measured the sheet resistance and the junction depth, so you know what R sh times x j is going to be. Now, depending on whether the profile is, you expect it to be Gaussian or complementary error function you can find out what the surface concentration is going to be, depending on the background doping concentration, right, which brings me to another key problem. How do I know what the doping profile is going to be like, how do I know whether it is going to follow complementary error function or Gaussian? That is from theory I can say. If it is infinite source diffusion, it is going to be complementary error function; if it is going to be a limited source diffusion, it is going to be Gaussian, if it is an ion implantation profile, it is going to be Gaussian.

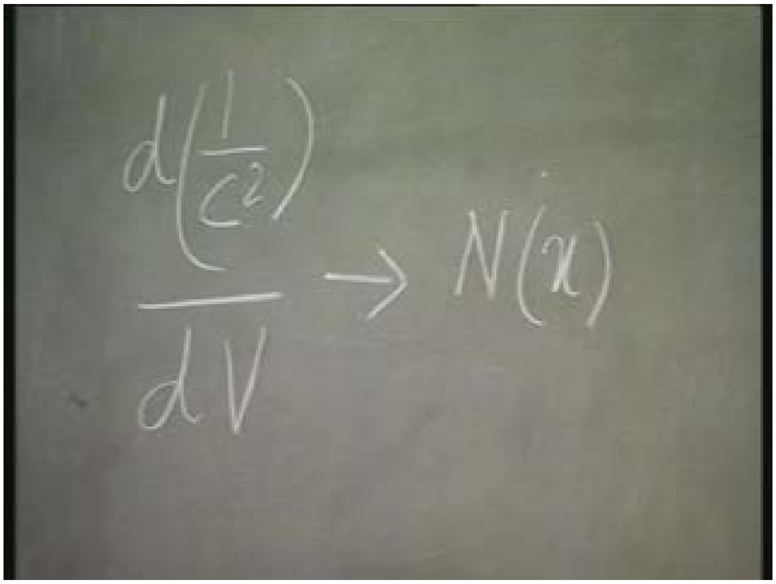
But, how do I check, how do I verify and you already know like in case of phosphorus diffused in silicon, the doping profile does not really match Gaussian. For boron implantation in silicon, you know it does not really match Gaussian, it is a Pearson ….. type distribution. So, how do I know these things, how did people find out these things?
So, which brings me to the question of how does one find the doping profile? There is one very simple technique, by anodic oxidation. What you do?

You just anodically oxidize the semiconductor. Since in the anodic oxidation, the oxide thickness is going to be proportional to the applied voltage, you know exactly how much oxide you have grown, that is exactly how much silicon you have consumed. Now, if you measure the sheet resistance before and after that is first you measure the sheet resistance on the surface. Then you grow a thin layer of oxide, etch that thin layer of oxide; again measure the sheet resistance on the new fresh exposed surface. So, the difference in these two sheet resistance values will be related to mu times N, right. It will be actually given by 1 upon q mu N into dx, the difference. You know dx; dx is the amount of silicon you have consumed. Since this dx is very, very small, you can approximately consider that the doping concentration is going to be nearly constant in this very thin region. So, what you know is the product mu into N.

Now, these plots are also available, mu versus N plots are also available and you know mu times N. So, you can find out what N is going to be. You keep doing that. This is called the successive etch technique. You keep on doing the anodic oxidation, etching and again you measure the sheet resistance, till you reach the junction. So, you have a number of points going from surface to bulk. You can find the entire doping profile. This is one way of measuring the doping profile.

The other way, which should be of more interest to you people that is by C-V technique, capacitance-voltage technique. You have a junction, a pn junction; apply a reverse bias to it, measure the capacitance. Change the reverse bias, increase the reverse bias, what is going to happen? The depletion region width is going to be increased. So, the change in the capacitance can be related to the change in the depletion layer width. From this, you can find out what the change in the charge is.

(Refer Slide Time: 51:13)



$$\frac{d\left(\frac{1}{C^2}\right)}{dV} \rightarrow N(x)$$

So, from C-V technique also you can find out the doping profile and the most common practice is to plot 1 by C square versus V, which is going to be related to the doping concentration. So, very simply stating, if you plot 1 by C square versus V and it follows a straight line, the slop of that curve actually gives you the doping concentration. So, you plot 1 by C square versus V, if it is a straight line that means the slope is constant; that means the doping concentration is uniform. If it is not a straight line, then you know it is not uniform. So, you can do, you can measure the slope at points all over the depletion layer width and you can find out how the doping profile is going to change. So, this is the C-V technique of measuring the doping profile. We will stop here today.