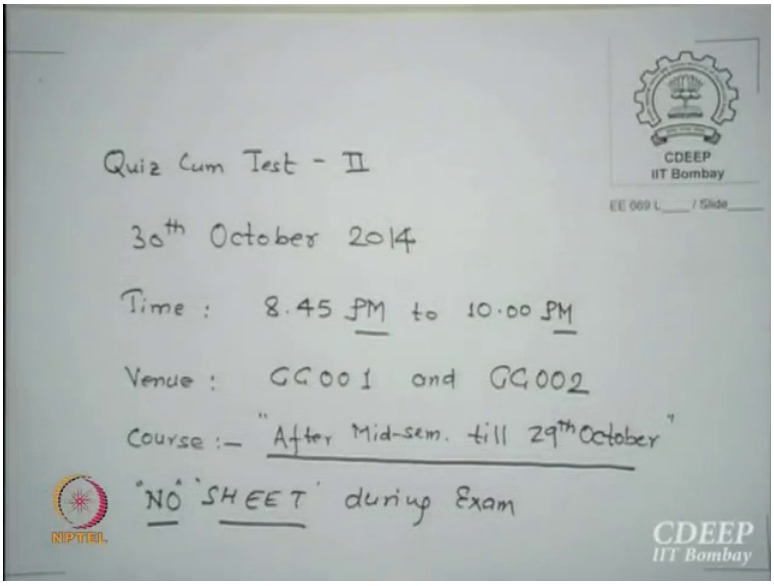
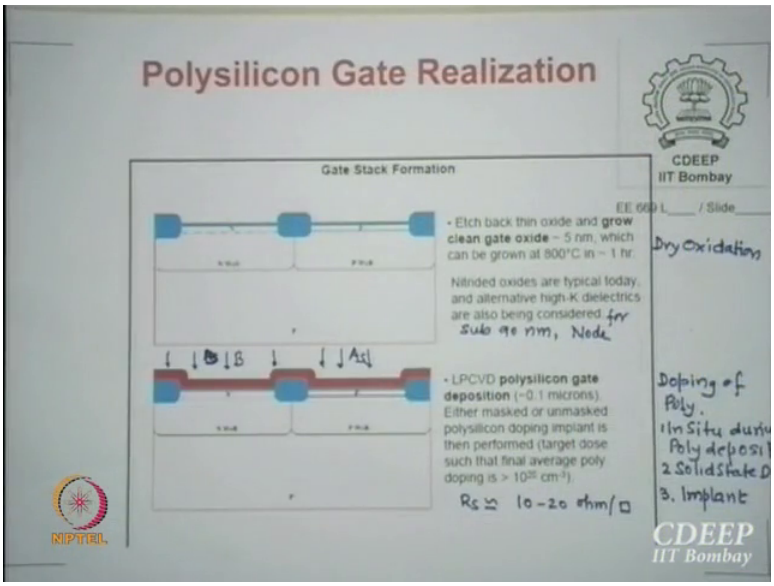
Fabrication of Silicon VLSI Circuits using the MOS technology
Professor A. N. Chandorkar
Department of Electrical Engineering
Indian Institute of Technology Bombay
Lecture no 23
Module no 01
Silicon IC Processing Flow for CMOS Technology

(Refer Slide Time: 0:19)



Okay so here we go I think this announcement has been already done so please note that my course I need not ask, whatever in last exam till the last day okay whatever has been taught is the part of the exam and since it is a quiz there will be some problem so to bring calculators.

(Refer Slide Time: 0:54)



We had last time gone little ahead but I will just go back a bit, we actually made the field oxide and then we made the Fox using low-cost process of this kind or we made SIT shallow trench isolation so either way we have now isolated the area for transistors, we clean everything on the wafer then accept the Fox and then the Silicon is also opened everything is cleaned, RC clean is very thorough at that time. Before any gate oxidation the wafer has to be ultraclean okay that is the major step for us for our MOS performance. So we etch back everything clean everything and the wafer is now ready for gate oxide and in our case right now I am talking of SiO2 but in other cases it can be high K in which case it will be deposition process but in this case it will be a thermally grown silicon dioxide typical it depends on the technology node.

It could be 20 Armstrong to 100 Armstrong or even earlier higher we used to have 300 Armstrong of gate oxide typically these days thin oxides are needed 800 degree centigrade which has the lowest growth rate probably steroid okay and maybe so that you have enough time for oxidation. So we actually have grown please remember that I forgot this before that we had done 2 things which is not shown here, one is channel implants with it channel stoppers and in one case we also did P and N implants for making threshold corrections. What I mean by threshold I do not believe that N-well concentration here and during the processing later will remain same and I want VT to be adjusted to the surface concentration so I actually implanted N and P to get my N-channel and P-channel device.

So here is after that implant has been done and everything cleaned up and all that, gate oxide has been grown, please remember that this process after this process gate oxide is never is going to be exposed again and therefore this as I keep saying this is the process step which is very crucial for all future successes. Immediately after the gate oxide has been grown we pushed the wafers into what we call as LPCVD reactor, LPCVD stands for Low-pressure CVD and typically around 1000 Armstrong of poly silicon is deposited. These days because of poly is not the best metallic system, we may try moly silicide, spluten silicide, tungsten, many silicides have been tried, moly is the most common one and for this right now I am talking of standard process so I say okay for poly silicon okay but in future processes this itself may get modified.

Now after the poly has been deposited which is 1000 Armstrongs we can either make an implant on this to make it by another masking step I can make one side N kind other kind, P kind or we may not do it right now at all okay this option I said when I do source-drain it will be exposed and that time I will get whatever poly depositions I want poly doping I want. Typically I am looking for less then 10 ohms centimetre 10 ohms per square as sheath resistance when I have a good poly okay so will P+ poly will have a lower sheath resistance or N+ poly? No, that is exactly a (())(4:33) and after so many days you will realize this. What is the maximum concentration boron can give and what is the maximum concentration arsenic or phosphorus can give? So it is always prefer to have N+ poly running everywhere because it will have least sheath resistance okay.

And therefore even if I for P-channel device I make P+ poly for the sake of gate but connection from there will not be P+ poly, it will run on only connecting with N+ poly ahead everywhere and that is most important that the poly runners or poly interconnects are always N+ doped preferably arsenic doped okay. But as I said these days we are not looking for poly runners too much maybe we will do silicides so it may not be that serious but essentially poly is very crucial for N+ poly. Then of course he said is not wrong, mobility has an issue but the major thing is one order higher concentration is possible okay... Interconnect from one transistor to other poly will run poly as a runner okay.

1st we will have to start with gate so there will be a poly so 1st layer may be a poly interconnect internally okay. If you do a design course and you must be doing it, you will say unless there is a

poly interconnect we will have difficulty in running metals okay, you need to many bridges then okay so the 1st runner is poly 1st runner is poly. Runner is a word used in design for interconnect so 1st runner is poly okay. So okay this was what we last time discussed, we have a poly which is either doped, some people dope it now but most people do not, and doping if at all has to be done is either by implant or by solid-state diffusion or during the growth of poly itself it can be doped by phosphine or (())(6:40) okay we can add gases along with that so process itself will dope it okay.

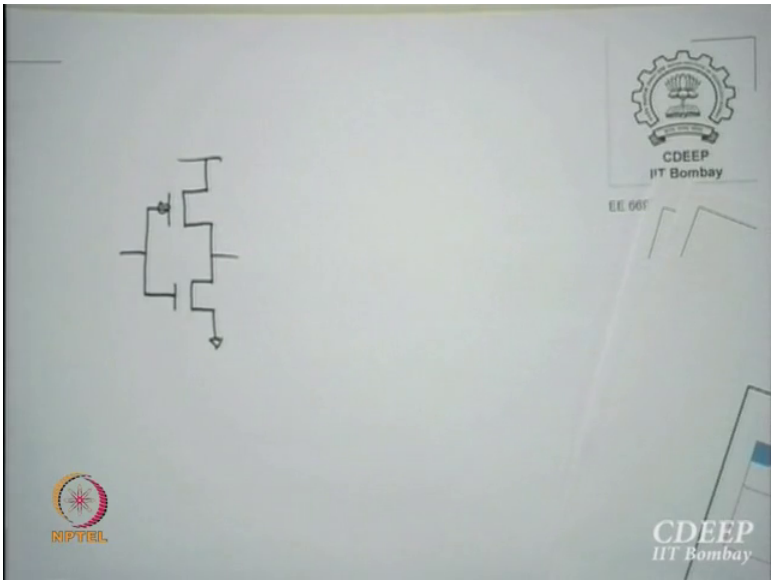
However, as I said rarely people do this but I just thought it is possible, why I wrote this specifically because poly is also used particularly for I f and analog blocks as resistors or even in memories, (())(6:59) so in statogram the load many times is these days because of variety of reasons can be restive, even in EPROMS you need now some what we called interconnect switches which also will require poly so in those cases we will require doping so this is one possibility how to dope a poly. So during the growth or during depositions or by implant or finally if you are bad enough everywhere then do solid-state diffusion.

Okay so till then we did this last time then I start now or maybe I also did this I am not sure, the 1st thing now I want to do is to create my gate because this is the most important lithography once again I will say because this will decide the W by L of the transistor okay so the etching here delineation here of the gate is the most crucial factor in deciding the I characteristics of transistor because here is where I am going to pick my width and length of the channel okay and therefore this is very crucial for me. So this step though it is like any other lithography step but it has to be handled very carefully because it is here where if I make 10 percent error I may make 100 percent and in my performance and therefore I am really worried how much error I may because then that will be shift everywhere some transistor will not have every channel stopper channels.

So if you are worried here of masking correctly and this is the step which takes care of your circuit performance requirement, we had deposited the resist as usual and etched that resist by developing, etched out the poly (())(8:42) I do not want gate okay. So you can see from here this and these gates are let us say to be connected in CMOS so where they will connect gate

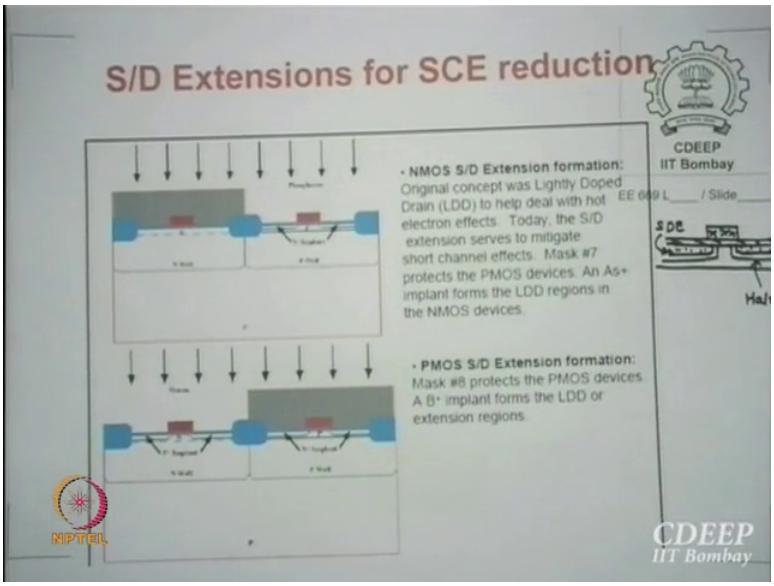
connected, inside the plane okay I mean if you go down on the plant then I can connect like this is that clear or what I should say.

(Refer Slide Time: 9:12)



If you see a normal CMOS inverter, this is what we this is a P channel so this gate connection is on the because this is circuit but on the cross-section I cannot show these 2 points because if I am taking something here I cannot show the connections but this connection is also possible during this masking itself so connection between 2 gates is also possible during this lithography itself okay, you should learn this plant elevation cross-section side elevation whatever it is. Okay so once this many times we have a rare possibility that some options people to refer reliability, some people actually retain the old oxide if they are already done good gate oxide before, some do not and they really etch out. So there were Ph.D. thesis from Stanford it proved that having fresh oxide is better okay for just this Ph.D. okay so think of it why so crucial okay.

(Refer Slide Time: 10:17)



After this the 1st thing we want to do is what we call as for the short channel devices particularly for the channel length less than 90 nanometre or 1.1 nanometre, 0.1 micron actually less than micron itself all short channel effects are seen but at least below 0.35, 0.25 you will certainly see huge number of problems. So one of the problems which we see is due to the scaling the voltages are not getting scaled whereas lengths and widths are so the electric field across the gate oxide as well as across the source-drain region are not scaled down by same proportion as widths and lengths and because of that electrical and until everywhere okay that is our major worry because as you reduce channel length and voltage is not scaled, V by L will increase as compared to normal scale device.

And this enhance field both across the gate and across the lateral channel has an issue which has to be sorted out because this additional electric field at the corner of drain particularly one finds that it may allow carriers actually to go into the gate oxide which is one reason how EPROMs are read, this is the process of EPROM writing which may happen without your actually wanting. If charges go into gate oxide what is the worst thing may happen? The threshold ratio, so everything is moving away from what you have designed for. So most crucially you must realise that this LDD called lightly doped drain structure so to deduce the electric field near the drain and also on the source said of course there is a matter in which asymmetric they say, (())

(12:00) we will not do but only on the drain said it is called asymmetric LDDs but right now standard process both sides.

So lightly drain doped have a depletion width, please remember depletion width is proportional to 1 upon root n okay, now lighter this the depletion width will be larger so it will not be concentrated near the drain but it will also be spreading below the lower regions of the drain and source region that means the electric field at the corner will be comparatively lower as compared to shorter depletion width okay. This issue is very important in more short channel effect device, the other problem of course is what we call channel length modulations, the 3rd problem is DIBL which is called drain induce (())(12:45) loading, I should ask much more on this before I started because otherwise I start teaching that course which I do not want.

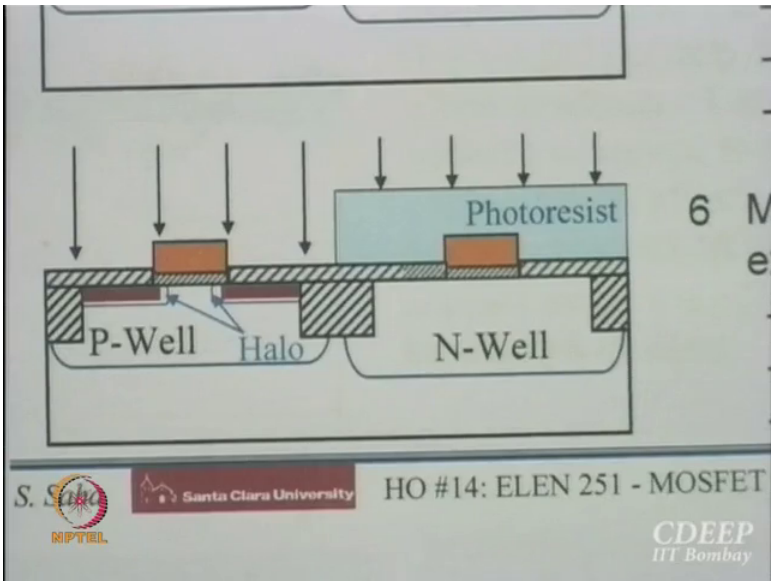
So if I really looking for my problem of short channel I must do something and one of the possibility is to do lightly doped drain structures. In another thing which I have shown here is also called Halo which also is for the safety when at least to reduce the depletion width going too far away so this is called limiters okay. But in this case there is no Halo shown the basic processes that please remember what we are trying in at the drain end we want N+, P+ to occur for N channel, P channel because that is where resistance is going to be okay so I want source and drain should have very highly doped regions. But very highly doped regions may create higher fields at the drain end because source will be grounded so drain will be a essentially huge freely.

So to spread out that field I reduce the doping near the drain and that process is what we call as lightly doped drain okay. So what we do is we actually mask let us say I am making first for n channel device so I mask this annual area and then do some phosphorus implants okay and since all implants will follow of course this you can see here I have not kept poly colour, since it is N and I am looking for N-channel so even if it goes into poly I am satisfied okay so I may mask that as well but I do not because I say it is N+ poly as well let it go. But in this case also it may happen, I may also have a P so I may have to compensate during N+ P+ that this P which I unnecessarily got there has to be 1st compensated. So there is another people who actually mask with this area get masking so they keep reduce keep there then implant so their method which means additional mask additional problems but otherwise can be done.

So essentially you can see it will follow etch and during (())(14:59) driving cycle it will get inside the gate little bit okay. Similarly, I mask with the P-well area for N channel device and then I opened for P-channel device and I do boron implant. Please remember this dotted line was what was that? This was the threshold corrections substrate equivalently saying substrate for the device, well is not substrate I created above which shallow implants so I have a P+, instead of P+ what? P implant I will do that is lightly doped P, lightly doped N and therefore arsenic was not put there, phosphorus was put, I was worried that it should not become N+ apriory so that is the trick okay and energy should be smaller because I do not want deeper implants okay.

Having done this then of course I may do arsenic implant for N+ or this some people also do arsenic instead of phosphorus but I am not the one who suggest that okay. Then you can see both mask are complimentary okay so I can use same mask resistive N resist mask complimentary either way I would can be done. So now I have both sides LDD structures for N channel device as well as P-channel device lightly doped regions have been created, please remember gate is protecting it below that is the channel area there is no doping which is only... What is the doping in the channel? Only the implant which you have done for threshold corrections okay that is the only implant there, is that okay.

(Refer Slide Time: 16:55)



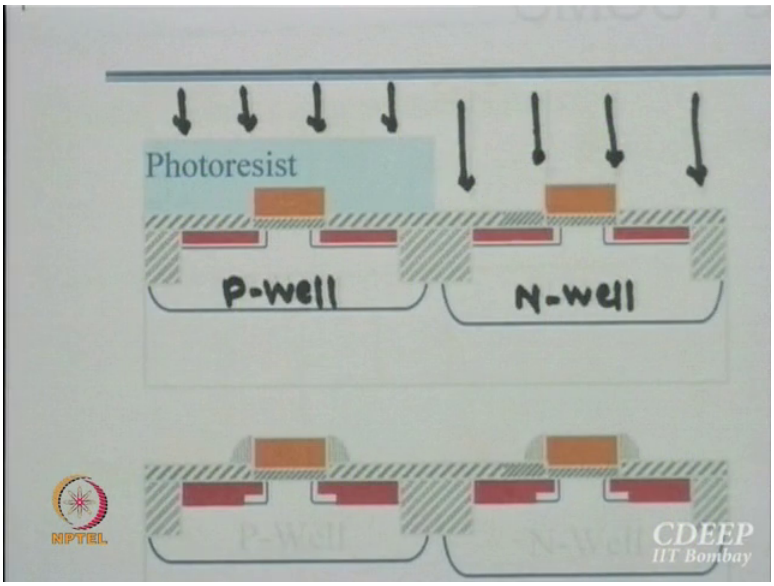
Just for the sake of this word Halo, Halo implants are very (())(16:50) just something another sheet which you can see C mask may be ahead should I do 6. This is a Halo implants but what is

Halo is essentially what we have done is that earlier P and N implant for threshold is there then we did LDD but before the LDD started we actually have another implant and that implant is the substrate, is that clear. Halo is made for substrate, Halo is not N+ P+ but it is for the substrate corners, this was your substrate and I made another implant there is that okay. And this is please remember for N+ source and drain extension of course we will create but we will have Halo of what kind? Boron okay boron is also a P well here so boron is the same as P well, there was already P threshold corrections but this implant is not in the channel, this is just at the edge of the channel.

So below the LDD implants you actually have Halo implant then LDD implants is that clear? 1st you have Halo implants which is same as well and then you have opposite of that to make LDD N or P implant, is that point clear? Halo is same as substrate implant I mean as the well implant so if you have P-well you have Halo of P type, if you have N-wells you have Halo of N-type okay this is called Halo this essentially is trying to create edges of the depletion region restricted only to that much place okay, a good figure may be coming and then see. So this is what Halo has done for so this is... what is HD implant, HD means LDD, HD implant is essential in this what he has written extension as is called, extension is for LDD.

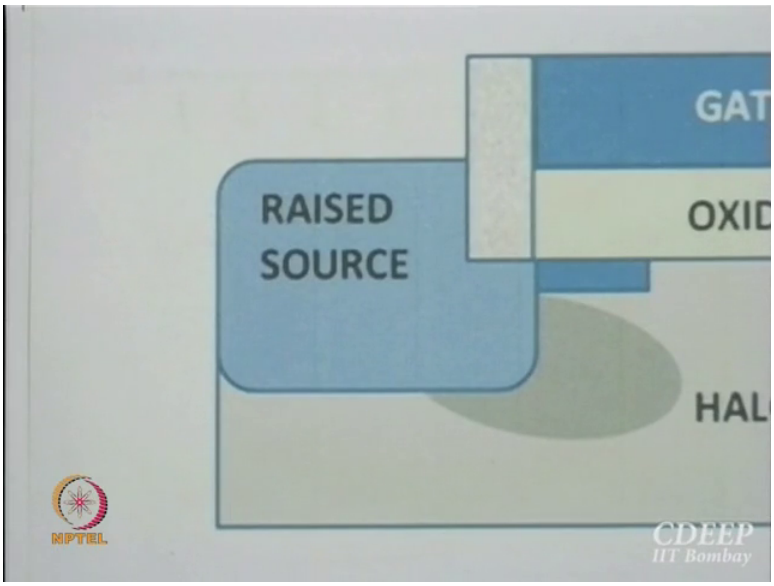
So HD is also... It is always N+, P+ for N-channel, P-channel, not + but N or P, whereas Halo will be the well implant please remember this part, Halo is essentially the substrate additional corner doping for the substrate okay or the wells at the area of so entry.

(Refer Slide Time: 19:11)



Well argument you can see from here, so I can also do now the other you can do P+ implant for source drain or P implant for source drain and N for Halo is that clear? N well means N for Halo and P implant for source and drain, same way for the other side so I have made Halo prior to my LDD this is that clear so it is a prior step and you do not need additional mask this is important you do not need additional mask during, whenever I was doing this I did write there both these processes. So I how do I do it? I just change the source and I will just change the energies is that clear? So one I want deeper the Halo has to be slightly deeper so I have little higher energy and in the case of N or P LDDs I will have lighter energy because I want at the surface is that clear.

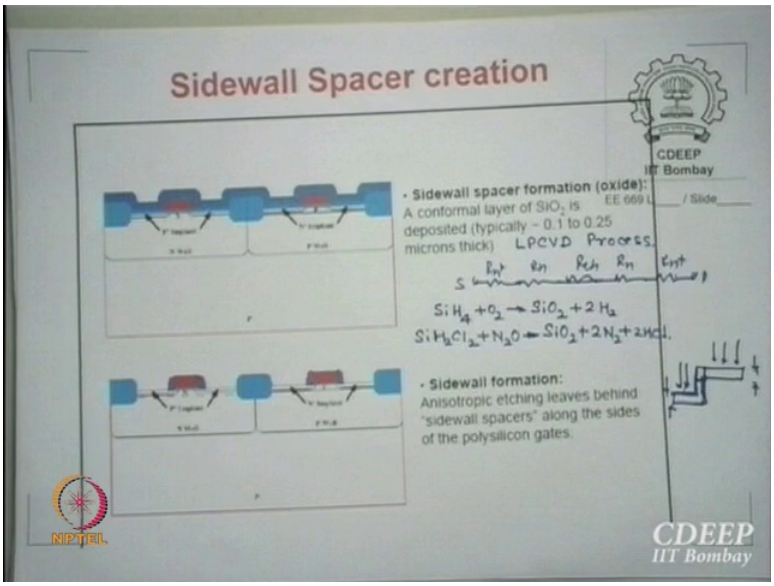
(Refer Slide Time: 20:48)



Red colour is N or P for N-channel, P-channel which is the source and drain LDD. White and blueish region, this is Halo this is well whatever well that same implant okay is that clear to you, this fact has registered many times people ask Halo, Halo world was used for something like this maybe I can show you why it was called Halo. It creates some peculiar show and it gives as if there is an image of something shown, so 1st time the ECM picture was taken the word was found Oh it looks like an Halo around okay so this word picked up from an ECM picture so do not go any other (())(20:54) there is no word essentially no expansion or HALO or something, it is essentially when the ECM was when the cut picture was seen, so someone said oh it looks Halo around substrate so okay so it is what it is so please get into it how word start but now it has got up okay now it has got up, this is called Halo implant.

Of course we will say there are many other like the circuit (())(21:21) technology which I choose from somewhere else, they are what is called gate, source and drain, this is another process step but right now forget about that I just wanted to use what is an Halo why the word was given Halo okay because when I see that I did not see uniform this, I say it is all around oh I said this hallowing all-around, Halo is a word which you say like most Indian Gods they show some Halo around the space you know backside same Halo okay.

(Refer Slide Time: 21:55)



Okay so having shown you then the next process step so far we did LDD so far we did Halo if needed then I do spacer, spacer as I said last time I discussed it but now I will show you where aside. Now after this I actually deposit on the wafer SiO2 deposit word please remember it can be even nitride, in most cases earlier ones we used to deposit silicon dioxide it can be even nitride okay. That process of deposition of SiO2 is through low-pressure CVD so this technique we have to learn what is how to do low-pressure CVD. And process that if you wish is silane is the gas used for silicon so silane + oxygen is SiO2 + 2H2 or if you are using dichlorosilane then SiH2Cl2 + the oxydent is N2O okay so the SiH2N this N2O is SiO2 + 2N2 +2HCl.

Normally I prefer not to have chlorine around, chlorine has some advantages and some great disadvantages, in 80s everyone thought chlorine will do wonders we have actually spent 3 years to put 3 years to put chlorine inside and finally learn that it has not done as good as what we thought all of us, many years many TC was tried, clear chlorine was tried, HCl oxidation was tried to improve the NIT or DIT of the MOS structure and we finally be realise that it reduces but it creates more draps TOH EK TARAH KA advantage DUSRA disadvantage so way which is better. So mostly silane will be used, only difficult with of course even dichlorosilane is poisonous not very deadly poisonous, silane is not really poisonous but it is extremely flammable. You are of this era so you will not learn you will not be knowing, earlier in English anything which is flammable used to be called inflammable okay inflammable.

Okay so now after this (())(24:09) this, I want to create sidewalls for the gate. So you can see from here oxide was... The wire deposition process is important because deposition is normally LP CVD in particular is conformal, what is conformal means? It follows the last contour, if something has to climb it will climb, if it goes down it will go down okay so it is a conformal process okay, since conformal oxide what does it mean? Thickness every point is roughly the same, every point thickness grown is roughly the same okay that has an advantage and disadvantage. Disadvantage is if I do it some normal etching in chemical not only this oxide will grow everything make go away so I want to restrict that etching process only 2 all other area except this sidewalls okay.

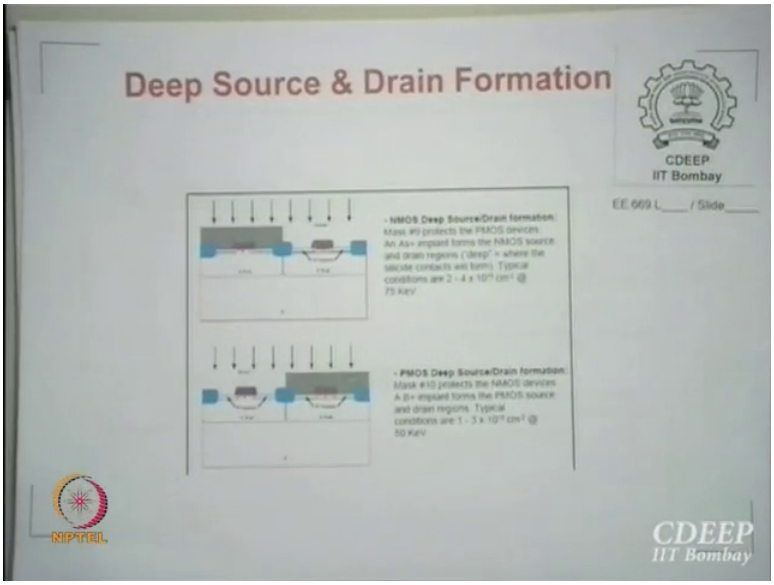
So I figure out that if I somehow do etching which is anisotropic vertically down and I said okay I have prepared I etched this oxide on the top let us say some T oxide that, I also can etch T oxide here but in this side is how much? 2T ox is that clear? Here is the T ox, here is the T ox, at the start there is 2 T ox, is my etchant is ionic etchant is such that it only can etch T ox then the right figure which you can see this will etch down and this will etch on this side but the sidewall it will not because it was twice there, is that clear. So the upper portion goes away, lower portion goes away, this goes away but space between the sidewalls of gate you always find oxide is retained that is called spacer, is that point clear? Why it is possible because I am etching down and I believe that the way I etch it, I will only etch T ox only okay. So upper T ox gone to, in between also oxide but sideway there is no way because ions were etching only vertically down okay and therefore I can create spacer.

This is what last time I discussed, spacer has its own advantage it does have some little bit of disadvantage as well there is something called boron depletion which is another worry but some other day okay. So once this sidewall formation has been done so do you think we have not used additional mask so far is that clear. I deposited oxide and I etched it so there is no mask so far used by me, I whatever ask 7, 6 or 7 whichever number I had how much it was the number when I just do not know which was the last number we have 5^{th} mask no not 5^{th}... 8^{th} mask we have done so we have still not gone for the 9^{th} mask, we have done now gate we have done white spacer, we have LDD, we have Halo but mask is only till 8 we have done almost everything what really device needs, so by 20 or 16 ITNI MEIN HOGAYA NA NAHI HOTA HAI ABHI.

Okay now 1st thing we have to do is, please remember the purpose of Halo. Let us say next step what is the source-drain implant I should have? Very heavy dose and it should fill up everything so that larger the depth of what is the better thing, the resistivity I mean the conductivity will be higher and therefore contact resistance will be lower, the sheathe resistance will be lower everything will go fair if I make deeper this but the technology may not allowing deeper source drain okay because there is sidewall capacitance issue so I want to reduce that. That means at least the minimum I should do is heavily dope that region, I cannot reduce now increased the junction depths because that I will be barred, I will be scaling it down for sake of capacitances.

So if I reduce that I will has increased... There is one problem which everyone is facing, smaller nodes the source-drain that is such smaller, thinner and because of that the resistance of source drain is not smaller now okay compared to earlier when I have 2 micron junction I have very low resistance, now I have 2000 Armstrong or less then I have much higher resistance at the source and drains, so the minimum I want if the surface concentration should be as high as possible okay.

(Refer Slide Time: 29:18)



Okay so I do now source-drain implant so 1st I for N channel so I mask with the N-well area for P- channel area and doing this I open all region wherever N transistor is going to come and deep enough is what because it should be at least filling up the complete source-drain area, if it is very shallow it will only go to the surface. There was a P implant you did, Halo implant you did, and I

want every area there be filled up with N+, is that clear? I have made an implant for threshold and then I made Halo and then I also have sidewall N or P there end. I want all that side on the left-hand side of the gate should be heavily doped so I must have deep enough and higher dose implant so that that depends N+ strong enough okay so that my resistance is smaller is that clear. Why my resistance is worrying me? This is the drain and source resistance in the transistor if you see even in analog for example, one of the major worries of this source resistance is what? Anyone?

Gain, it say common emitter with source what is called as degenerated source, the gain will be R d R e or R d by R s, if R s is pregnant that the gain can never increase maybe 4, 5, 8, I want 100 so this source resistance in analog is very crucial for us okay however, therefore we never go for 22 nanometre sorry I will work 0.25, 0.35, 90 BOHOT HOAGAYA TOH 65 okay that is it. So typical energies they have given for the process which is 0.25 micron process is 2 or 4 into 10 to the power 15 per centimetre square, please remember we talk about doses, is that clear? Depths are decided by what? The energy, so we only talk of dose and dose is decided by what? The current collected at the integral I dt whatever is our dose. So, dose is what we are deciding and not... Energy decides the depth, dose decides the amount of impurity I can push inside okay.

So I have just my dopes also there is a word which is important is this source drain also may get another contact on the top because it is silicon nitride so I will put the metal on that for making a contact and it should form a good alloy between the two is called silicidation. With poly what it is called if I make a contact with poly? Salicidation okay, silicon we call silicidation, poly silicon we call salicidation. So silicadation process is going to happen on source drain, some metal will actually react with source-drain so it will eat up some silicon is that clear? It will eat up some silicon which means your depth should be sufficient enough that this silicadation process does not take place. All are placed so there is no there is nothing left there so we have to be little careful, these are all design issues process design issues.

Typical energy he says is 75KV and this was what impurity will do? Arsenic why, N+ okay 10 to the power 21 I can go 4 into 10 to the power 21 is highest solidarity I can reach there. Of course yes it is named given polysalicide yeah but I am not saying this in this process itself + there is a separate salicidation process for ROMs so some other day some other technology. The word used

there if you read somewhere salicidation do not get worried it is essentially we are talking of polysilicides okay that is why I just say that number deem to you...

It will still not because the salicides (())(33:27) limit is coming at the temperature you arrive but that is the limit up to which no material we get what is I would say it still remain crystalline. So if you have too much of a base times the damage which it will create will not be restored okay so essentially there is a limit which is close to solid solidarity is that the divide goes to 1000 degree centigrade to anneal it so that it reaches maximum recovery as well as full dose. Okay, use the opposite mask opposite resist whichever you think of right now do for boron doping for P+ okay which is again the same technique, why boron has lower energy? The range of boron is longer for same energy so you think of it that we will discuss earlier we have shown you that number has been used 70 KV to 50KV typically dose is something 10 to the power 15 or 16 per centimeter square okay.

So now I made N+, P+ okay, I have made channel stoppers, I have made Halo is, I have made LDDs okay I have made everything what a transistor needs okay, I have almost made my both N channel and P-channel device so what should be the next thing I need? I need to make something contacts to all regions so that I can externally create a circuit out of that okay. After this implant which is normally high dose implant and particularly for arsenic the damage will be higher so what should be the anneal cycle? At least should be higher temperature till 30 minutes anneal are essential for recoveries, and what is the advantage of recovery during that? The impurities will also get substitutional so it is good to happen there okay.

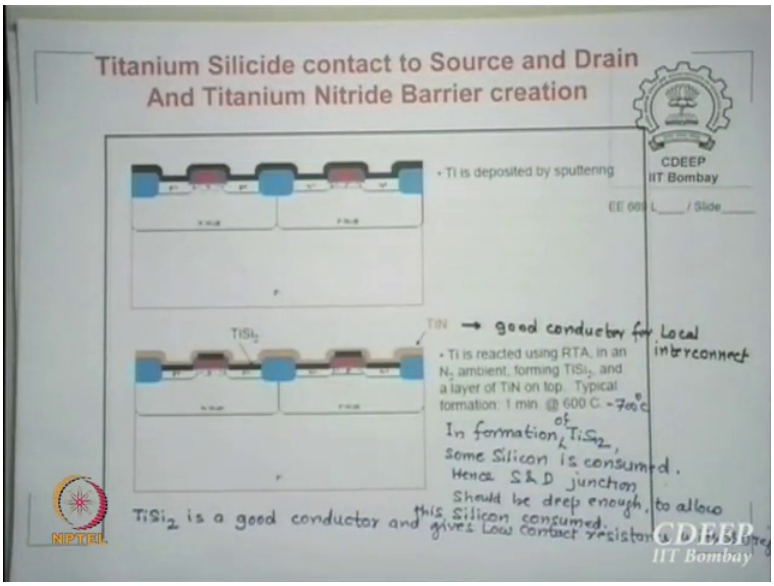
So there is a final anneal as the word says it is a high-temperature (())(35:27) it activates the implant at dopants, defuses junctions to their final depth typically 30 minutes, 900 or say 1 minute RTA, RTA is Rapid Thermal Annual at 1000 degree so that now I have a source drain, I have Halo of course not shown here, Halo + LDD + gate is also doped during those regions okay, so my transistors are ready okay. Now I need as I said contacts, I need contact from the source, contact from the drain, contact from the gates okay so I need contacts everywhere because unless I have takeout contact out of it I cannot actually connect externally anywhere okay on the wafer. So this next times is first thing is you whatever is the additional oxide here there you give, what

actually even not HF depth it is called to buffered HF depth say it is 10 to 1 or 100 to 1 HF diluted and you actually give 1 second (())(36:38) dope it and take it okay.

It removes all the circuit everything what is taken around, the idea is not to disturb any of the other oxides, sidewalls nothing should change okay so it actually will work is dipped into bath and taken away okay but not by hand by machine, it just goes in and goes after okay and then water is immediately poured on it okay then draw it and everything else. Okay so this is essentially we are now ready for this process which I finished now maybe little more 1st contact of course we will do is called Front end process, what is it called? Till the first contact we have not finished that but till the 1st contact is made for source-drain the process is called front end processing okay. That means for any circuit irrespective of whatever circuit you are looking for this process will be constant except W wired but that is decided.

So I have no transistors of (())(37:44) I have bought 1st contact to source-drain and everywhere and now what is the next thing you have to do is to create interconnect between number of such transistors the way you want circuit to perform is that clear? You want to make a flipflop you connect some other way, you want to make (())(38:04) you mark some other way, you use but these transistors are available to you with their basic connections out for any connection possibility, is that clear to you? So this is means why it was called front end because front end processes are constant, for a given technology node this will be a constant process okay no changes. But back end which is one which starts, back end means the interconnect parts ahead is the decision of a circuit man okay circuit people are telling what you want so that is called back end processing.

(Refer Slide Time: 38:47)



Okay so the 1st contact now I have to do is it is Titanium silicide contact to source and drain and titanium nitride are barrier creation, of course there are earlier we used to make even aluminium okay as the 1st contact itself, aluminium has all good things in impact I mean people always say why suddenly you went for titanium. Two reasons we went to titanium; aluminium has the problem and they are very thin they electromagnet very fast, some other day we will discuss this. So since say since very strong electron magnetic material, also it oxidises very fast. It is good because actually that is why here all grills everything are anodised means pacivated that means it becomes Alumina so it has stronger oxygen affinity for aluminium and aluminium oxide is an insulate that is correct so we do not want insulator to be contact okay so we want metallic contact, so aluminium is not so strong contender.

Earlier we had thicker aluminium so we had no problem, now when everything is reducing this aluminium has a penal problem, aluminium has oxidation problem then we said okay give it up so aluminium was declared person in non-greta and copper was replacing aluminium okay. Aluminium has good thermal conductivity and it has a good electrical conduct... 2 things we are looking for, good thermal conductivity and their electrical conductivity aluminium has all of it. Okay so we deposited titanium, deposition word please take it this coplanar or I mean it always forms the contours, I will say just now copper is replacing aluminium this is why we have copper interconnects.

Copper has disadvantage, copper has 2 levels in silicon okay say which are traps so it may clear mobility so you must protect copper from silicon so at least 1st interconnect cannot be copper is that clear? It has to have something else, 2nd, 3rd, 10 you can go for copper. Copper also has a problem it has a very strong diffusivity in oxide so one has to put some kind of a clad, glad means some kind of a jacket wherever copper goes. You have seen wire, it is always inside copper wire, it oxidises also very fast so it needs jacketing so additional process, this process was found by many companies and everyone claimed patents, it is called dancing process maybe we should discuss later on okay, copper cladding with copper interconnects okay that is where that vanadium oxide was worked as a clad.

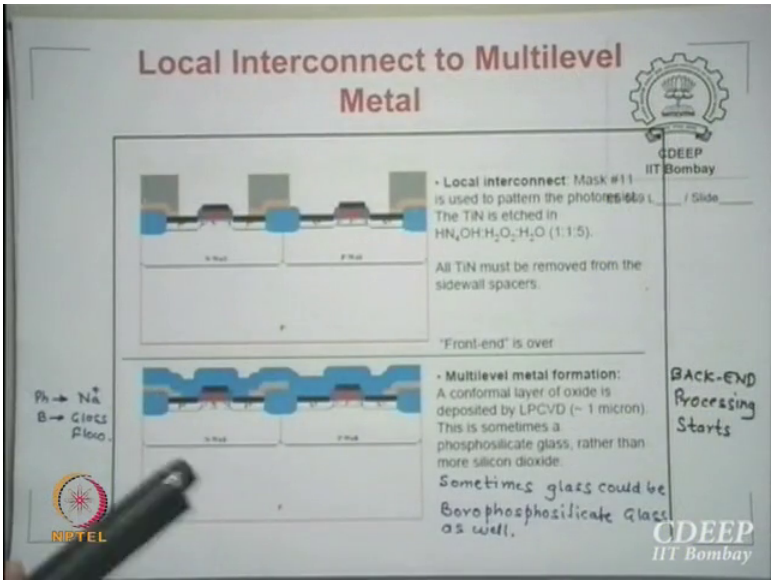
Okay so I deposited titanium everywhere then this process of deposition generally is by process called sputtering okay, the process which you will discuss is sputtering okay it is also one way of deposition metal frames, one of course is evaporation the other is sputtering. After this titanium has been deposited we actually allow it to react with nitrogen Ambien to form a titanium nitride layer on the top okay. Wherever this please remember formation there from silicon is being consumed, source-drain junction should be deep enough this word is I am keep saying why source-drain junctions are very important that deep how much deep? You want that some silicon will go away and that is why I rate again for this that source-drain depths are very crucial in present technology, for little less there is no source, little more it lost the conductivity issue okay.

So here is the issue which is very difficult to and that you people learn by experience, source-drain junction should be deep enough to allow titanium silicide as a good conductor which gives low contact resistance with source and drain that is why titanium are used in fact, it actually forms titanium silicide and it is an alloy okay it is an alloy and that can be done by using one... On top of this I actually passed nitrogen and create titanium nitride on that. Titanium nitride also a good conductor lucky, it is a great thing that titanium nitride is also actually is better than titanium silicide as a conductor running conductor okay, but creating titanium silicide everywhere uniformly is difficult okay allowing can be only restricted areas but nitride can run anywhere okay.

So the so-called whatever pink colour or orange colour is titanium nitride okay and this process normally is not (())(43:31) and it is generally this RTA because one-minute everything can be

achieved. Temperatures are 600 to 700, please remember in VLSI this thing has happening, all processes should be now lower and lower temperature because whatever earlier you have done that should not get disturbed with this additional temperatures. So next processing if implant has been annealed at 1000 or 900, no process next should happen for 900 because that DT product should not get added there, is that clear? So the next processing I had will always be lower temperature than the main implant process temperatures okay.

(Refer Slide Time: 44:23)



Having created nitride having silicide and everything done we use mask 11 to remove the nitride from the areas which you do not want, I only want nitride to be connected to the drain is source and maybe one of these and this contact may be internal to it which is not shown here, this so called orange one or nitride these are resist, you etch out the nitride from everywhere, people always need to know nitride etchant is ammonia NH4H + H2O 2H2O. Where do you use this etchant? In silicon cleaning, do you remember I say HCl + one is acidic cleaning other is basic, this is the same process okay.

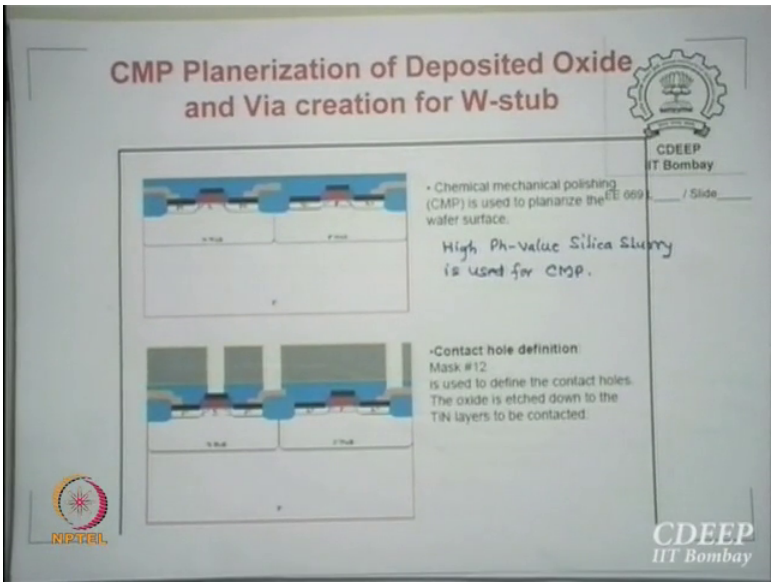
So we will remove titanium nitride in this normal one is to one is to 5 etchant and we will also remove... Please remember my worries are essentially that this spacer should not retain nitride so that is where the whole game was that that is why chemical agent was used because I want from everywhere nitride to go except the regions where I actually looking for interconnect is that point clear why the wet etching is done here and not ionic etching. Dry etching was not tried because I

want nitride to go from everywhere else is that clear to you because I said interconnect and it is not a contact to source-drain is that correct? It is an interconnect running.

We have titanium silicide as the contacts but outside to that I am running nitride layer titanium nitride, this has to be understood and therefore you need a mask so that titanium nitride is removed from all places except interconnect regions okay that is why this is an important step you can see from here everywhere we removed titanium nitride otherwise if it is a coplanar it will also go everywhere and if you protect it will also remain there so I want to remove from every other place. Then if I want a multilevel metal, now this is what I say front end, 1st metal as come front end and now what we start is back end, starts. We also do a conformal layer of oxide now again, I deposit silicon dioxide okay typically around a micron this blue one which goes everywhere, conformal means it sees the contours okay we will say LPCVD process.

Normally either the phosphosilicate glass or borosilicate glass we tried but off late people are trying what is called as Borophosphosilicate glass which has both boro glass as well as phospho glass, the advantage you will figure out is the following; phosphorus have the tendency to actually tag onto sodium, phosphosilicate glass will tag to sodium it will not allow sodium move. Where boron actually will allow because you know you have a glass which is there and you want to planarise latter so I want last to be little liquefied, I will not say liquified it is called reflow so it is slightly molten state. Borosilicate has a lower temperature of the flow as compared to phospho, so if I add borosilicate when actually want to planarise, I will flow at lower temperature so I will use borosilicate along with phosphosilicate it is a mixture okay so it is also given a name Borophosphosilicate glass.

(Refer Slide Time: 48:07)

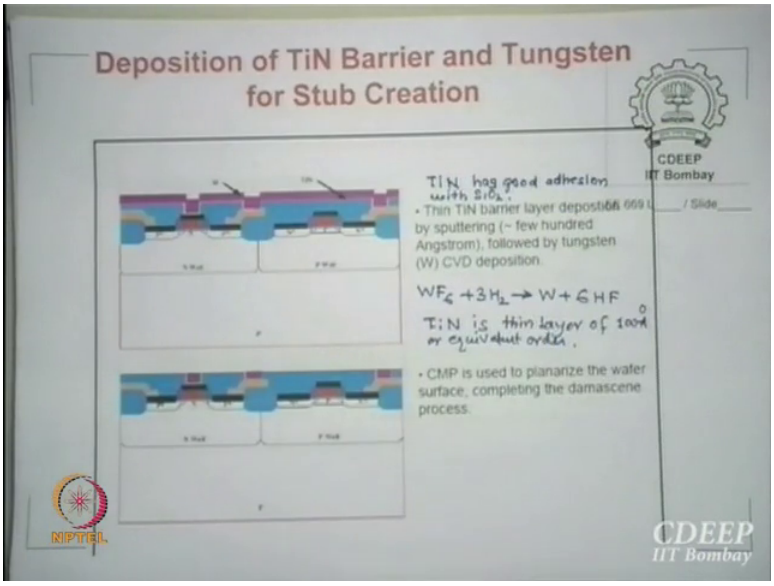


So once I do this which is the blue colour this, the 1st thing I do is to use a process which we discussed last time is called chemical mechanical polishing CMP, this was not in our time now this is a very common process, the wafers let us say if this is your wafer surface the polish is done like this on the Chuck which has a slurry silica slurry which normally has some higher pH means basic so slurry is slightly more basic than it should not be acidic, the acidic will attack everything so pH should be higher enough so you keep maintaining some ammonia ammonium hydroxide solutions around and there is a constant pH monitor is going on, you keep adding automatically. So using this the wafers are mounted on Chuck and actually polished mechanically not by hands, by machines and that essentially removes the top layer.

The earlier versions please just a minute we have a non-planar surface is that clear? Now please remember they have gone through so many process step, everywhere steps have different heights okay so we are now to make at least the top of the wafer one clay like this, this is called by using a CMP process I can planarise is that okay, this planarisation is a crucial process step. Now so far how many mask? 11, now here you have 12th mask I want to make contacts to this titanium nitride so I must protect all other even on titanium nitride I mean titanium silicide, I want contact gate, I want contact source, I want contact to drain so I must open all those regions where 2nd metal should actually come and make contact to that these are called either called Holes or called VIA V I A so I must create VIA.

In what material I should create VIA, only in resist it is good enough so 1st I actually create resist and I create VIA and I etch out all other oxides from region where contact is to be made is that clear, I remove from the gate, I remove from nitride, I remove of course here I did not show maybe backside and maybe here, wherever I want contact to the next layer I must open a window connection to there. What is being etched in mask? Oxide, please remember masking is always done oxide okay so this this will delineate what, the contact regions. Now this depth of VIA what we are creating is also crucial, how much thicker oxide should be? If it is too thin capacitance will be very high see Q ox is lower, if it is too thick the (())(51:12) depth is too high and the next process may create problem so depth of VIA is decided by how much CMP I do and what I retain as my actual oxide things, is that point clear.

(Refer Slide Time: 53:37)



To thickness of oxide I keep, the VIAs are too deep and the next metal may not go proper. If it is too thin between 2 layers if the oxide is too thin they will interconnect okay they will connect to itself by capacitance at high frequency, one of them is J Omega C. So my worry is how much so this circuit people will say this is the frequency they will operate so I will accordingly adjust my oxide thickness okay. Okay so the next process step I then after the contact is done I actually have another titanium nitride as the next layer of contact okay, please remember the... Pinkish layer is actually the tungsten, please please just listen, this pink layer is tungsten but below that

there is a thin layer of titanium nitride which I deposited which is not so much visible because it was below that.

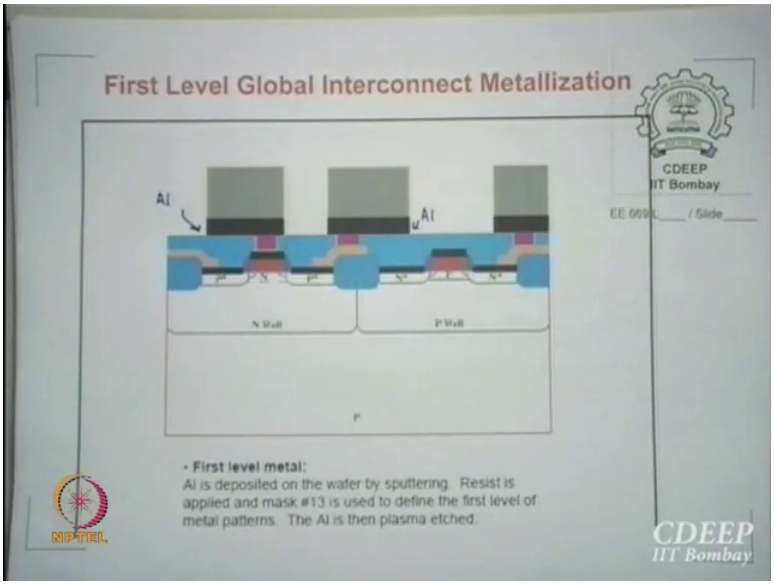
Of course in this region you can see but everywhere you will find one thin layer of titanium nitride over which tungsten is going to sit. Titanium nitride has some interesting features it is a layer typically 100 Armstrong or what we do and it acts like a barrier, firstly it has (())(52:53) with silicon dioxide, titanium nitride makes a good abrasion so that is what we are looking for. The other way it is a barrier between tungsten and the lower material lower region okay it should act between the 2 buffer. So titanium nitride is a buffer between tungsten and any other metal okay so everywhere wherever I run tungsten below that there will be a thin layer of titanium nitride and that will be typically 100 armstrong okay. How to defer the tungsten? tungsten hexafluoride treated with hard region at around 450 degree centigrade can reduce to tungsten and HF.

Now this tungsten, then I remove this tungsten please remember this is also conformal because it was deposited this is CVD so what do I do is I go through another CMP process, how much? As much as the thickness of tungsten was, so I if I remove this top portion of tungsten I get only tungsten just on the contact points when we want, these are called tungsten plugs what are they called? Plugs. So we create tungsten plugs, where do we create tungsten plugs? Wherever contacts you want above, for the 2nd metal you must create between the lower metal and the upper metal the conductivity is through tungsten and this W by L of this tungsten is decided by the contact resistance and the net resistance you can tolerate okay.

Okay so once I etch out the or rather I remove the tungsten from the top surface I have another planar surface there and why this is always possible? CMP allows me to grind things so that I should know up to where I have to grind so I finished that process so I get this Process which the people are talking. This was titanium silicide is not called 1st interconnect, the 1st interconnect is now going to come okay. Here they have shown aluminium but could be copper, I did not want to show copper because there is another process I have to show which I will show later, so this is essentially 1st metal is aluminium. How many metal layers you may expect? As large as 7 okay and maybe someday 10.

Why do we need so many interconnect? Because if there are 1 billion transistors on chip okay for 1 single circuit, how many particular if it is a DSP based some kind of a circuit there is only interconnect there are no devices, it is only here to there that are going from here to here here some filters, some adders, it keeps playing running around, LL, HH, HL, LH, filtering is going on and so interconnects are extremely heavy. Now because of such thing and we know 2 metals certainly will not like to go like this, if one metal is going here so it should be at least thick enough oxide and I run something on this. So I will say do not run this everywhere, do not run in between I run so if I decide it I will not run over everywhere as much as possible, I will require much more space to run the interconnects.

(Refer Slide Time: 57:24)



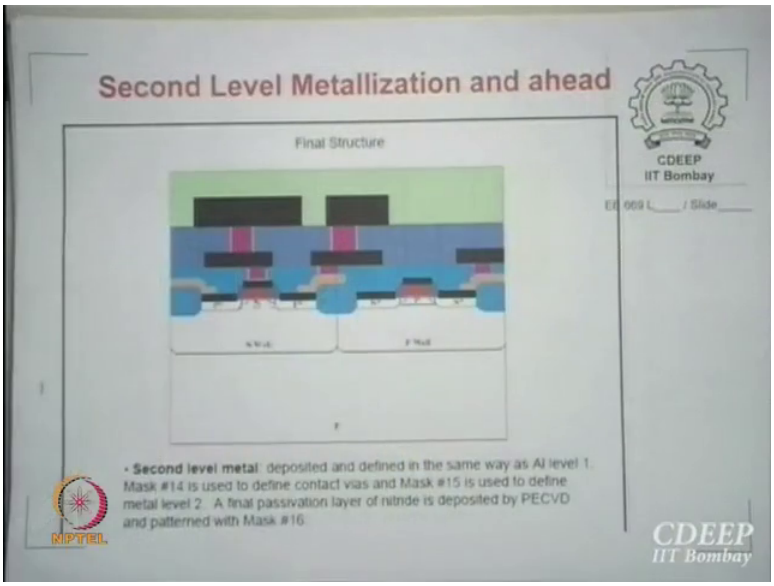
At least what I do is one should not interact with 3, 3 should not interact with 6 or 5 then next nearest layer it should not interact, below it goes it is okay to thickness of oxide and (())(56:51) but at least the nearest neighbor on the top or bottom should not get connected so run metal layer should not run 1 over the other as much as possible and this needs number of metal layers okay so that we avoid connection going from crossing one over okay. Okay so this is and since I did is aluminum not because I wanted it but this is what Plummer's book has so I thought I will show you but nowadays aluminum will be replaced by copper but the only problem then is as I say aluminum does not require any cladding, I have to create vanadium, some kind of a clad around

in which copper should sit okay so there is another gate going on okay maybe I will show you (())(57:40)

So I deposit metal and what do I do? Wherever that interconnect I want to retain okay I put photoresist, etch aluminum from elsewhere so which resist I should, let us say I using PPR, what kind of mask it should I repeat, I want to retain aluminum but I want be removed from all other regions, I am using PPR okay Clearfield mask. Why clear because whenever light will go photoresistance will (())(58:11) so aluminum will be easy to etch out there because there will not be resist there. The rest area that I am protecting people like should not go and therefore it should be... I repeat the mask is decided or resist is decided. Generally companies prefer one mask type or one resist type, they do not keep changing resist and mask generally not necessarily.

So which is this number mask for aluminum? 13 mask has been used to do 1st metal. If I want the 2nd metal what should I do again? I should deposit oxide again, if you need plugs you put plugs also then polish it and then again put the 2nd metal layer, create 14th mask on that, etch out the 14th layer okay. So this is 1, 2nd metal is I have another oxide and you can see another plug has been created, wherever I go from one metal to the lower metal that tungsten plug has to be created and tungsten plug can only be created by which it must have titanium nitride below it before so the same processing is done, 1st I deposit nitride and then I deposit in the same series actually I do it, it is called core CVDs so it is not actually CVDs, it is a molecular (())(59:43) from other word I can be one after the other, if I do LPCVD I have 2 reactors I will have to do twice TOH APNE YAHA HAI AISA.

(Refer Slide Time: 60:03)



This is 2nd metal, if I want 3rd what do I do? Etch out this, put them another oxide, put VIAs, put titanium nitride, tungsten, put new metal, put another keep doing as many metal layers as you wish to okay for example, we say another 2 of them may be final. The last after number of metals are over there is the last mask is used okay, this is called the generally the layer on the top which is shown here for example is silicon nitride okay. This is very important, 2 things silicon nitride is a relatively hard material what does that mean? During handling it does not gets crashed down in silicon area, it will actually protect the silicon wafer. So nitride the 2nd part of nitride is it is excellent passivating material, it does not allow any sodium potentia or any other specie or even water molecules to get inside it is not hydroscopic.

Because of that the last layer of the chip is nitride, but you need now another mask for what? Final contacts will go on the chip I mean this package so you need pad patterns to be opened. And opened where? Nitride is etched only at the pad patterns and nowhere else, pad is the last mask so your pad pattern which is now open in nitride and that is actually put it into the packaging and therefore either wire bonded or by bums bonding flip chip and I can make final packaging of the chip. So last mask irrespective whatever number of mask we did earlier it should be a passivation mask and normally it is and it should be very thin and very good nitride so its plasma enhanced CVD one of the CVD process which we use is plasma enhanced CVD, is that point clear to you?

So how many mask I did? 16 for the sake of 2 metals, you add 5 more metals so 5 minimum will go there okay. Why 5, we need more for the plugs so think of add anything additional mask for everything you do ahead of this, is that correct? Do not think that why not 20 mask this. Another thing you must remember, at no time the top most metal or the metal should go more than one layer down connection, is that clear this connectivity is clear, I cannot connect 7^{th} layer to 1 at any time directly so do not create a VIA of 7 layers, there is no possibility of anything going through okay. So what we will do is from 7 to 6, 6 to 5, 5 to 4, 4 to 6 so keep shifting okay so the area is constantly (())(62:58) because of where do you want which connection.

So venue layout a circuit you must plan very well that you do not need 7 to 1 connection anytime okay is that clear. So the designers when they create a layout also should think what is the technology available and how many minimum connectivity is to be given so it is not process person job, it is the designer's job to give me a mask which will allow me less number of interlayer connectivity that means top to layer I should not be asked to connect is that clear. I mean I am not saying I cannot do it (())(63:36) but that is not the best way of connectivity so try designs in a way, somewhere you use interconnect with your poly, somewhere you use only local titanium nitride, so do not run everywhere from every other place to every other place, this is something designer has to plan which layer I want next to that so any layout you may see colours there.

I think he has already come out with layout, has he? Have you shown some layouts to you? So you must have seen poly, red this, essentially base colours are mask for me, each colour is a mask for me okay so that time we should plan which connection when okay therefore that is the layout that is the mask I will receive, that is the mask I will print okay is that clear. So designer's job does not ends only on circuit simulation which many people believe, it is the layout which is the most limiting the success of chip design so layout should be very good, error free, good density with no issues in that all that and then you expect my technology to deliver what you are actually expecting which I may not promise at once, but I may go 2^{nd} time in one ram.

These are called turnarounds, once see I run through a mask 16 mask and my chip did not work so the first time I will go and verify there are test area there are this, I will test there I went... my is the mask problem or my problem okay. Most problem technology people will always say it is a

mask problem, designers say your process has the problem, these additional expenses are not mine, all statements are made but finally you have to play again the 2nd round and finally success has to come okay. There are no designs in the world which was formed in 1st turnout, there is always 1st turnout, one turn around it part of the gate this will be required.

All designers design with specifications and they keep saying you these are the... Technology people say these are the things which you must follow which are called design rules, you cannot separate diffusion less and this, put metal close to this, every route which I need I will inform him so its layout will take care and now there is an auto design checker so it will not allow even to draw many a times, you can foresee there is a bypass you can overrule it but normally DRC does not allow. Then why fail? Because this fail is something which humanly not possible. What designers tell the process people? Technology has set these design rules, what technology people receive from data from designers this you must give me, what is that he is asked? Technology has turned you cannot do this, you cannot do this, you cannot do this so I had a layout person I have to have taken care.

But what is something designer keeps telling this man, this you must promise what is that? He actually tells you the propagation time from source to drain he actually specifies, why propagation time? Because L by V is time but V is (())(67:02) okay so it essentially telling this much minimum mobility I am expecting at the surface from Mu because all my analyses are based on some Mu, is that clear? So designers tell only one thing, the propagation time for an electron to go from source to drain, he said I want propagation time of 10 picoseconds that is the only data he forces on you so the process should hold to that number any day in all transistors is that clear? So please do not think that designers do not tell, designers do not tell designers have very stringent force, they say if you even 1 percent or 5 percent error you go how much error I will have he will have to calculate.

So he has to always worry about speeds and then he will only tell what is called as propagation time, so do not think only one-sided is this, there is other side people keep telling where device processing is not good enough to get this much mobilities and I want for both P-Channel and N-channel this much okay and then only I can assure you it might it will perform, is that point clear to you. So do not anytime feel that things are very trivial, any small of course few of my Ph.D.

students work on reliability issues and we did figure out most processes will show you 10 percent variations okay.

Now either your designer should take care of 10 percent variations or at least you should say some of the processes may actually damage you more than the others so either there are techniques statistical techniques so you will have to do now statistical designs rather than what we called as fixed designs so you must take care of variations and these variations vary from technology node to technology node, to company to company. This (())(69:01) method is not very good that is what our (())(69:05) says that number of variables as much you take let us say you have 12 variables, so the... And pseudorandom generation requires 2 to the power 12 samples, so if you process it for a circuit with some 100 million transistors, and some next year its result come.

So we figured out that 2 to the power something instead of 2 into something if I do ay 2 to the power q and 2 into q and q is large, I will do much smaller simulations and will be only less than 0.01 percent away from (())(69:42) okay so that is other technique to show that we can do analysis much faster for any large circuit. Or you may even assign which are the critical areas where circuit need not be worried too much okay, slowest path so you anyway that is going to limit so why are you forcing VT everywhere? Only this area is limiting you, so first circuit person should talk with device and then to process and each should translate there what then you can make a better device so it is called statistical device so nowadays most of the designs are statistical okay.

So this finishes 16 mask standard CMOS, tomorrow we will start with 3 processes or rather to processes which we talked which we have not done, one we say deposition and in deposition also we said few different times, sometimes I deposited metals sometimes I deposited insulators okay so I need to have a CVD chemical vapor deposition which can either deposit metals or this. And then I will have etching, I will etch everywhere okay so I need to know etching, so 2 processes we will do so tomorrow we will start with CVD.