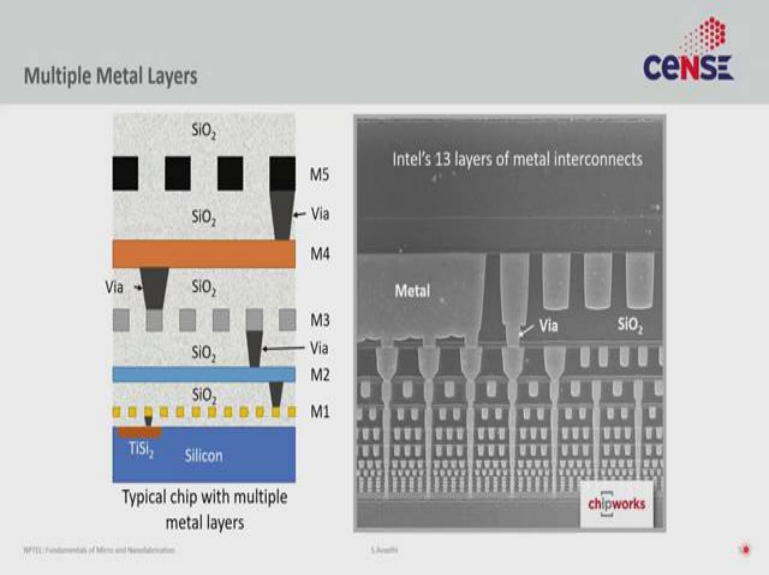
Fundamentals of Micro and Nanofabrication
Prof. Sushobhan Avasthi
Centre for Nano Science and Engineering
Indian Institute of Science, Bengaluru

Lecture – 26
Metallization: Contact resistance

In this lecture, we discuss metallization in the continuing series on additive processing.

(Refer Slide Time: 00:45)



We have looked at various ways of doing physical vapor deposition. Specifically, evaporation, sputtering, and in much lesser detail, pulsed laser deposition. We will now discuss how the metal deposition affects the device properties, electromigration, and multiple metal deposition layers. In a modern device, you often need multiple layers of metals. If you are connecting a city like Bangalore with roads, having all of them at one plane level creates a bottleneck. You need to add traffic lights at every intersection. There would be a jam at some of them. One way to avoid that is to have over-bridges that allow the vehicles to go over each other without stopping.

Something very similar happens in a device. What you are looking at is a cross-section of a real intel chip. The scale bar is missing here, but at the bottom, we have a silicon substrate on which all the devices exist. At this scale, they are too small to be seen. All these devices have to be interconnected. If you look closely at the image on the right, you can see small metal connections at the bottom edge. There are many layers of metal

connections. As you go from bottom to top, the metal layers become larger. The higher-level metal connections accumulate the current from the lower level metals. To avoid delay, they must have low resistance, and the way to get low resistance is to increase the geometry. So, top-level metal layers are often very thick and very big; the lower ones tend to be thinner and smaller.

You must insulate these metal layers from each other, or the signals will cross. The insulation must have as low capacitance as possible to minimize the delay. Therefore, the dielectric constant ϵ needs to be as small as possible.

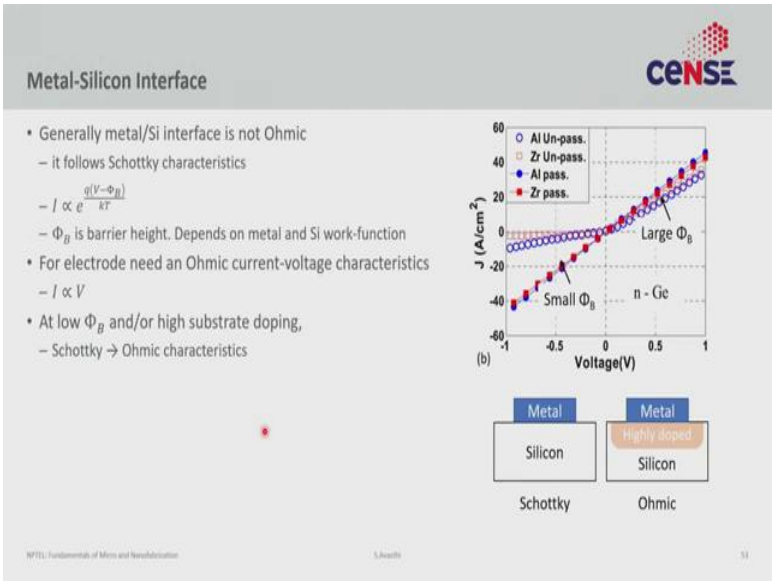
The best insulator is air, but in practical devices, that is hard. There would be mechanical problems of having an overhanging metal. We fill it with SiO_2 instead. The slide shows an SEM (scanning electron microscope) image on the right and an illustration on the left. You have some layer, typically a silicide, that makes the metal contact with lower contact resistance. This black thing touching titanium silicide is also a metal, a deep fill called a Via. This Via connects you to the first layer of metal, M1. To join the first layer of metal and the second, you again need a Via. As you go up, the dimensions increase and the thickness between the metal layers also increases.

We have discussed several ways of depositing SiO_2. The native oxide growth gives you very high-quality silicon oxide, but it requires silicon. Because it needs silicon, the top layers cannot be native oxides as they do not have access to the silicon. Maybe the bottom-most layer can be, but no more. There is an additional consideration. You deposit SiO_2 on top of these metal layers. We have learned in the diffusion lecture that metals diffuse very fast. Once you have deposited a metal, you can never take the wafer to higher temperatures. Hence, you must deposit all these SiO_2 at lower temperatures, low enough so that the metal does not diffuse. Low-temperature oxides usually have poor qualities. We have to deal with it.

We discussed the term back end of the line (BEOL) processing but did not go into the details. The top layer depositions are BEOL processing. All these SiO_2 depositions are also BEOL because they have metal on it. You cannot do the RCA cleans; you have to do gentler cleans that are metal compatible. You cannot use clean furnaces for these steps and must dedicate a separate set of metal-contaminated equipment. The process

temperature should be low enough such that metals do not diffuse from these contacts, vias, through SiO_2 into the silicon. Metal should not move from top to bottom.

(Refer Slide Time: 06:36)



Let us discuss the metal-silicon interface. Usually, the metal-silicon interfaces show Schottky characteristics. Metals are good conductors, semiconductors are reasonable conductors, but the interface can be a worse conductor than metal or semiconductor alone. At the interface, you can have a built-in field that actively counters any flow of carriers. The value of ϕ_B determines how bad that field is. The larger the barrier, the smaller the current. To use metal as a contact, you want the Schottky-barrier as low as possible to push as much current as possible because the interface resistance losses would lead to heating and would also require you to put more voltage than you need.

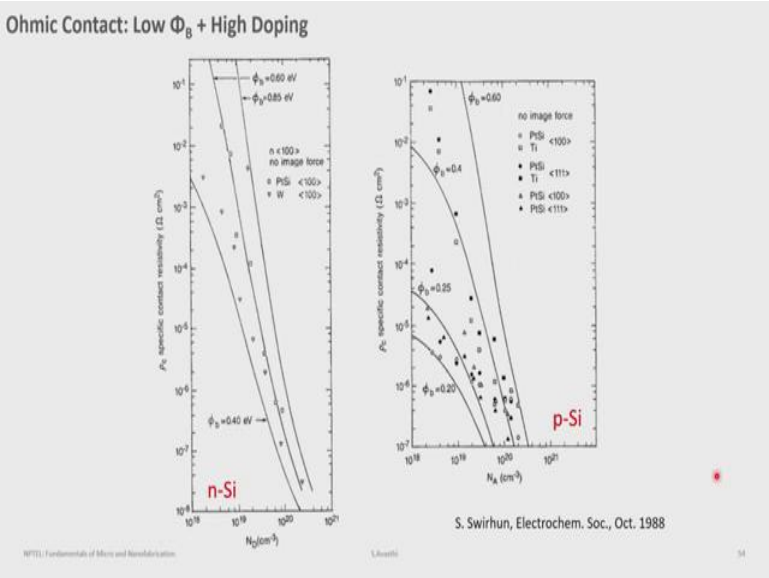
Now, for devices like diodes, we go for larger barriers; but for contacts, which is the application we are discussing today, you want this ϕ_B to be as small as possible. Now, this ϕ_B depends upon the metal work function and the silicon work function. In general, the less difference between the two work functions, the smaller the barrier. One way that happens in practical devices is that the barriers become very small. For all the current limits that we need to use, you can approximate this expression (shown in the slide) to be linear. In that case, we say that the barrier is so small that the device is not functioning like a Schottky barrier but like an Ohmic barrier. So, that is what we seek.

On the right is an example of how such a barrier would look. You are looking at current versus voltage. The current scales with the area (A), so we plot the current density J(A/cm^2). This plot is for metal contact on n-type germanium. You see similar behavior for Si or any other semiconductor. The open circles/squares represent metal-semiconductor junctions with large ϕ_B, and the filled circles/squares represent small ϕ_B.

With a large barrier, you get a diode characteristic. You can push the current in the forward-bias, but you do not in the reverse-bias. Now, this is a problem. Supposing you want to flow an AC signal with both positive and negative voltages (forward and reverse biases). If you have a diode at the interface, then it will rectify the AC signal. You want an Ohmic behavior, where the resistance is the same for both the negative and the positive bias. So, you get a linear I-V curve. It is possible for small barriers. One easy way to achieve this is to change the metal work-function. Zirconium and aluminum have a slight difference in work function. If you look at the open circles, the blue one is aluminum; the red one is zirconium. Both of them are Schottky and do not allow current to pass through in the negative bias. Aluminum passes a little bit more current. You get a considerable difference if you do passivation. It is interface engineering that reduces the barrier despite these materials having very different work functions.

The most common way of making good contact is to dope the silicon at the interface. You take silicon and grow a very highly doped layer. If the substrate is n-type, the highly doped layer must also be n-type. So, we have n++ silicon under the metal layer. It allows you to tune the work function of silicon and reduces the barrier width. If the barrier between metal and semiconductor becomes very narrow, the carriers can effectively tunnel through it. These are common tricks to get ohmic contacts even when the work functions between the silicon and the metal do not agree.

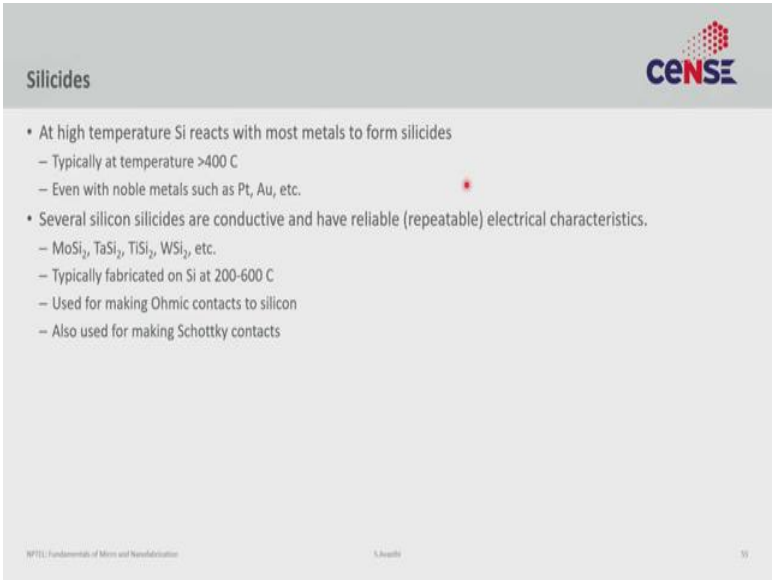
(Refer Slide Time: 11:24)



The figure shows published data about n-type and p-type Si specific contact resistivity as a dopant concentration function. Notice the unit on the y-axis is $\Omega.cm^2$. The larger the contact area (A), the smaller the resistance. It inversely scales with A. So, whenever you talk of contact resistance, you have to talk about a normalized quantity, specific contact resistivity. The specific contact resistance is high for lowly doped films and low for highly doped films, similar to what we discussed in the previous slide. If you can grow a highly doped layer at the interface, your specific contact resistance will be lower.

Also, the specific contact resistance changes with the φ_B. For any φ_B, high doping is better. If you cannot create a highly doped region, you can play with φ_B. The contact resistance is low for lower φ_B. This specific case (shown in the slide) is for two contact metals on n-type Si; tungsten and platinum silicide. We shall discuss how to make platinum silicide later. You see similar effects in p-type silicon. In general, you get a lower specific contact resistance for the higher doping and lower barrier.

(Refer Slide Time: 13:24)



What is silicide? Several metals alloy with silicon. If you put platinum on top of silicon and heat it at some given temperature, it would form an alloy with silicon. That is called (platinum) silicide. Typically it requires a high temperature and depends upon the metal. The range is typically 400-900^{o}C. It can form even with noble metals because this is not a chemical process, but an alloying process like copper and gold, or copper and silver alloy. Similarly, silicon can form alloys with gold or platinum. Whether a metal and silicon will form an alloy is given by the phase diagram.

The solid solubility determines the percentage of gold and silicon in an alloy, which you can see in the phase diagrams. The reason silicides are interesting is that they are conducting, very conducting. For example, molly silicide, tantalum silicide, a titanium silicide, tungsten silicide are all conductive silicides.

This alloy formation is very beneficial because it allows you to overcome some of the challenges you have with barrier height. Suppose you have an issue with contacts because the metal and silicon's work functions do not match very well. A common trick is to anneal it at a high enough temperature that you form the silicide. Then, you have a metal silicide – silicon contact instead of a direct metal – silicon contact. And that allows you to reduce the contact resistance.

There is another trick people often use to get ohmic contacts on silicon. If this silicide has a very different work function, then it can also be used to form Schottky contacts.

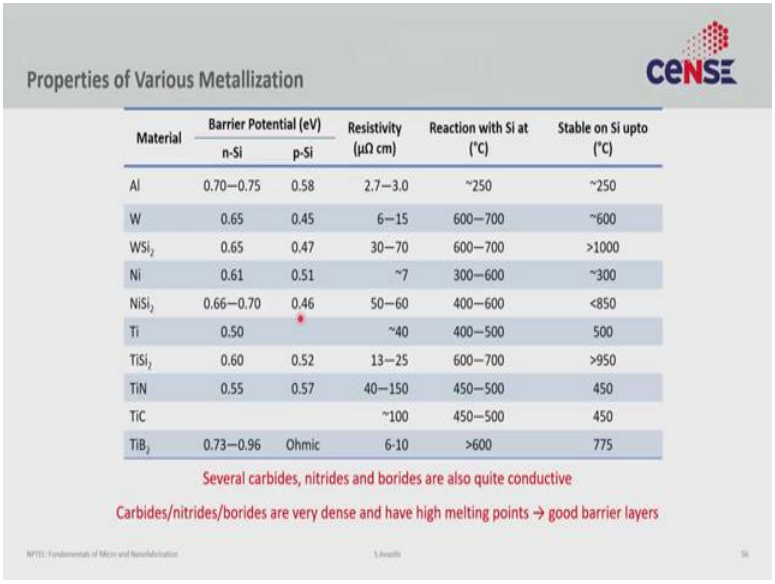
Those are the cases where you do not want to make contact but a diode. In those cases, silicide is a more reliable way to form a diode or a Schottky – contact. Why it is more reliable requires a little bit of discussion in the practicalities of metal deposition.

We have discussed why cleaning the silicon surface is so essential, especially before critical depositions. Many of those cleaning steps that contain acids work for native silicon but become chemically incompatible with metal. Even with hydrogen passivation or hydrofluoric acid dips, there is always a little bit of oxygen on a silicon surface.

Similarly, when you deposit metal, we have discussed maintaining the purity of physical vapor deposition by making the base pressure very low, keeping immaculate materials, keeping the mean free path very long, and the arrival ratios high, etcetera. But all of those are still limited in their effectiveness; for any given deposition, there always be some impurity in the metal. Any impurity, whether it is at the metal – silicon interface or within the metal, would slightly change the barrier. Since the contact resistance exponentially depends upon the barrier height, tiny changes in it cause considerable changes in contact resistance.

Getting a consistent contact resistance from the metal-silicon interface is often challenging. Silicides are a little more dependable as they are thermodynamically stable. The silicide would form due to solid solubility, despite the small number of impurities in the metal, silicon, or at the interface. It is a bulk material, not an interface. Controlling interface property is very hard; maintaining the bulk property is a little easier. As the bulk properties are more repeatable, the silicide contact tends to be more repeatable.

(Refer Slide Time: 17:52)



Here is an example of the various types of metals you can deposit on silicon. Aluminum is one of the easiest and most often used. The table shows the measured values of the barrier potential for n-type and p-type Si with various contacts, including silicides.

Ideally, if you add the barrier potential between n-type and p-type silicon, you should get the bandgap. If that rule breaks, it is usually because of the interface effects. So the theoretical and experimental barrier potentials are often not the same. Those of you who have done a device course may remember image charges and a few other non-idealities that creep. We shall not discuss those non-idealities; we will simply say that theoretical and practical band barriers are often different, but theoretically, the sum of these two should be the bandgap. Therefore, for a particular metal, if the Schottky-barrier with p-type is shallow, it should be high with n-type.

The contact resistance depends on whether you are making contact with p-type silicon or n-type silicon. There is one exception to this rule and this exception is when you create a very highly doped layer at the interface, the contact becomes insensitive to the barrier. Then it doesn't matter whether it is a p-type or n-type semiconductor. As long as it is highly doped, any metal would give an ohmic contact. Otherwise, you have to decide which metal to use, guided by whether you are making contact with n-type material or p-type.

Aluminum gives a higher barrier for n-type than p-type, which means you will have a higher contact resistance on n-type than p-type. The aluminum itself is very conductive. Often the resistance of the whole system is not decided by the resistivity of the metal but the specific contact resistance.

Aluminum and silicon start alloying with each other at around 250^{o}C. Aluminum melts at around 630^{o}C, so 250^{o}C for aluminum is a relatively high temperature. The flip side is that Al-Si interfaces are only stable up to 250^{o}C. If you are depositing aluminum on silicon, do not heat your wafer after the deposition to more than 250^{o}C, or you will see a change at the interface; if you can live with that change, fair enough. So, the temperature at which the Si and the metal alloy also tells you till what temperature the interface is stable.

Another prevalent material is tungsten. It is a refractive (high-temperature) metal that we used as a boat or crucible in evaporation. Interestingly, it gives you a barrier for both n-type and p-type silicon, but it is not very large. 0.45 eV is relatively small. For p-Si and W, resistivity is higher than that of aluminum. So, tungsten is not that conductive. If you are using tungsten only in Vias, then the actual length of the contact is so small that the higher resistivity is something we can live with. The reaction between tungsten and silicon to form a silicide happens at 600-700^{o}C. Partly because tungsten is a high melting point material, it sort of diffuses less. So, it means that you can make stable tungsten silicon junctions up to 600^{o}C. Because of its high-temperature stability, we often use it for military or space hardened devices.

Once you form tungsten silicide, the silicide itself would have specific properties. The tungsten silicide has very similar properties to tungsten, and that is one of its significant advantages. So, the barrier heights are essentially the same. Resistivity is a little higher but still on the same order of magnitude, and it is stable up to 1000^{o}C. Once you form the tungsten silicide, its interface would be stable up to 1000^{o}C, a reasonably high temperature. The disadvantage here is the formation temperature is also high. If you cannot tolerate these high temperatures because you have other metals that may start defusing into the system, you cannot use tungsten.

You may look at other metals that form silicide at a slightly lower temperature, maybe not as low as aluminum or as high as tungsten. A common one that people use is nickel.

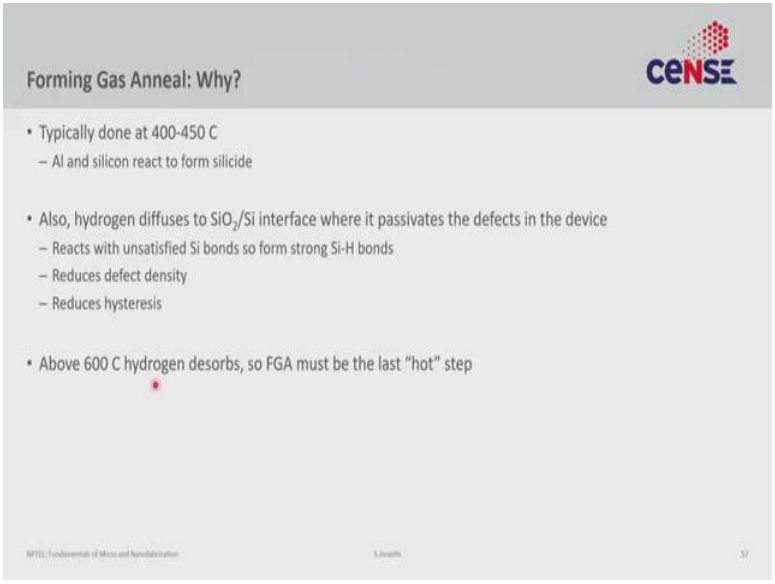
Ideally, you start seeing ohmic characteristics for ~ 1 A current at barriers of 0.3 eV or so. 0.3 eV, for all practical purposes, is a pretty low barrier. Nickel is reasonably conductive, so that is another advantage. A reaction with silicon starts at much lower temperatures 300-600^{o}C; it depends on what type of silicide you are forming. Since nickel can create multiple silicides, it is only stable up to 300^{o}C. The silicide is more resistive. However, it is acceptable. Nickel silicide, once it forms, is stable up to 850 C.

Titanium is another very commonly used metal, I will not go through the details of everyone, but you see the general trends; the materials that I will talk about are titanium nitride, titanium carbide, and titanium boride. Titanium nitride is a useful material, also reasonably conductive. It is stable up to 450-500^{o}C. More interestingly, it is a remarkably dense material. Generally, nitrides and carbides are very hard. So, TiN, TiC, and TiB_2 do not allow diffusion. You can use TiN as diffusion barriers. For example, if you have a metal and are worried that metal at a specific temperature will start diffusing out into the silicon, one way to prevent that is to encapsulate it by a diffusion barrier. Titanium carbide and borides are useful for similar reasons.

To make good contact, we typically need a low barrier; if that is not possible, form a silicide or create a highly doped region under the metal. The temperature at which silicide form is an upper limit for the metal-semiconductor interface stability. Once you make that interface, the wafer cannot go to higher temperatures. It is a number that you always have to keep in mind when you are creating process flow.

Hence, metallization tends to be the last step in a typical fabrication process. You do all your high temperature-sensitive steps first. You grow your gate oxides, where you require a very high amount of cleaning, followed by a silicon epitaxy, where you need a very high amount of cleaning. From those very pure processes, you then transition to moderately dirty techniques such as RIE, etching, etcetera. Finally, you do physical vapor deposition of metal, silicides, nitrides such that you do not create contamination.

(Refer Slide Time: 26:30)



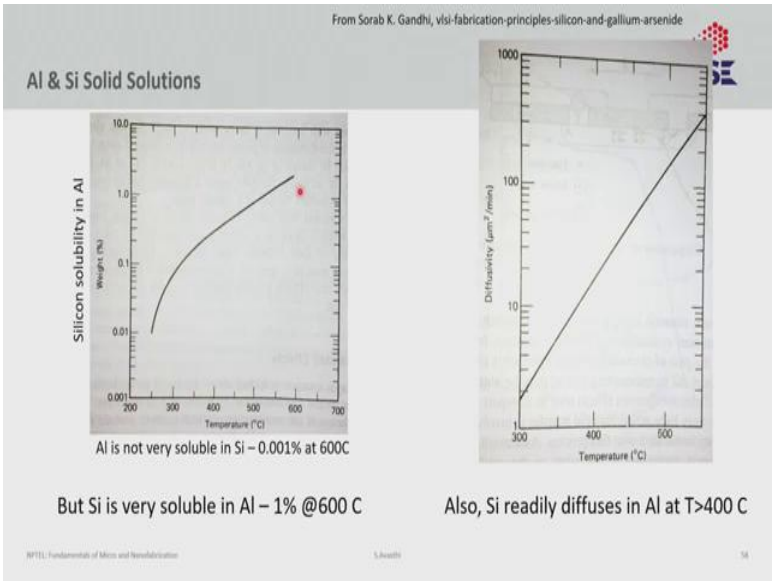
We looked at Forming Gas Anneal (FGA) when we were discussing silicon oxide interfaces. FGA is essentially a mixture of hydrogen, nitrogen. You keep hydrogen low enough such that the mixture becomes safer to use than hydrogen alone. The hydrogen is the active component and diffuses through materials very fast. It passivates the defects at the silicon-oxide interface. FGA can also improve the qualities of the metal-silicon interface.

You typically do the FGA at 400-450^{o}C, not higher. At any higher temperature, the hydrogen tends to desorb not adsorb. Also, at high temperature, the metal will start diffusing, forming silicides, and the interface starts changing. In BEOL processing, you want the temperatures to be kept low. Ensure that the temperature of the FGA is low.

The hydrogen diffuses, and as a reducing agent, removes any amount of oxide (which prevents you from making good contact between your metal and silicon) at the interface. The Si-H bonds are strong. Hydrogen from the FGA passivates any defect present at the interface by making Si-H bonds.

When you have aluminum in the system, and you flow forming gas, it reacts with the aluminum to form nascent hydrogen, not molecular hydrogen (H_2), but nascent atomic hydrogen. That nascent atomic hydrogen is much more potent at removing defects. If you have aluminum as a metal, the effectiveness of the forming gas is much better. Let us discuss on aluminum-silicon interface, simply because so many people use these.

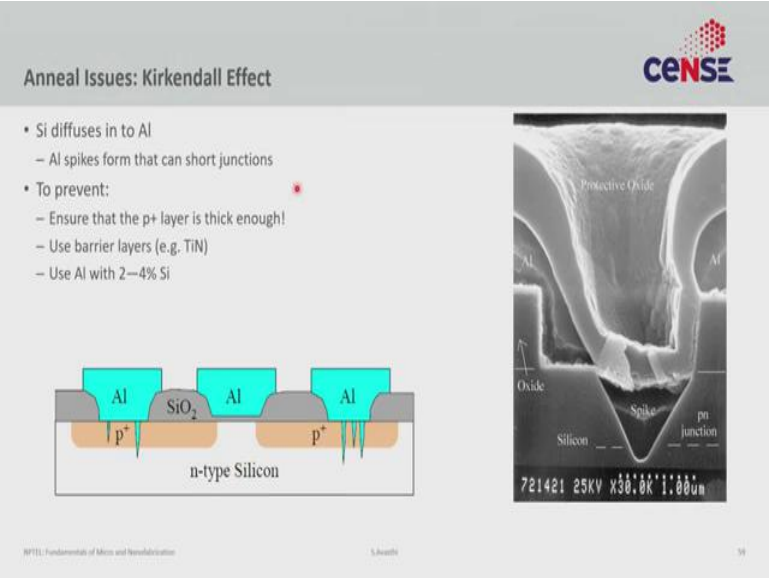
(Refer Slide Time: 29:02)



Till now, we looked at the solubility of the metal or a dopant or an impurity into silicon. If you want to analyze how much aluminum diffuses into silicon, you look at the solubility of aluminum in silicon. However, diffusion is a two-way street. Aluminum diffuses into silicon, but silicon might also diffuse into aluminum. After all, there is a gradient there too. So, you must also look at the solubility of silicon in aluminum.

Looking at this plot, you realize that aluminum is not very soluble in silicon. It is only around 0.001 % at 600^{o}C. So, aluminum does not dissolve in silicon a lot. If you look at the phase diagram, you will see nothing. However, silicon is hugely soluble in aluminum, up to 1 %. It is a relatively large number that increases with temperature. Also, silicon diffuses remarkably fast in aluminum. Any silicon that dissolves in aluminum because of solid solubility would again move very fast in the aluminum layer.

(Refer Slide Time: 30:32)



When you deposit aluminum on silicon and heat the system to say 400-500^{o}C or 600^{o}C, not only do you form a silicide, but you also get spikes.

In the slide, if you refer to the SEM image shown on the right, you can see that the oxide is etched, and the Al is in contact with Si. If you take this configuration to high temperatures (say 600^{o}C) during back end processing, the Si starts diffusing (moving up in the figure) into the Al because the solid solubility and diffusivity of Si in Al is high. This movement of Si into the Al creates a void that is filled by Al, resulting in spikes of Al through the layer. In a p-n junction, the spikes protrude through the p+ regions resulting in a short circuit, as Al comes in direct contact with n-type silicon. You do not have a p-n junction then, only a Schottky-junction with a low enough barrier that you form an Ohmic-contact. You get an Ohmic-contact where you expected a diode. That is a routine problem when you take this system after deposition of metal to higher temperatures and when your p+ regions are thin.

Be careful while doing rapid thermal annealing of very shallow implants. Be very careful in using aluminum; even at low temperatures, you might have a short-circuit. To prevent that, either use a passivation layer, say a titanium nitride at the interface or grow a silicide or do something to make sure the aluminum and silicon are not in intimate contact with each other.

In the next lecture, we shall discuss another exciting issue with metals called electromigration, with some discussion on the damascene process and the use of barrier layers.