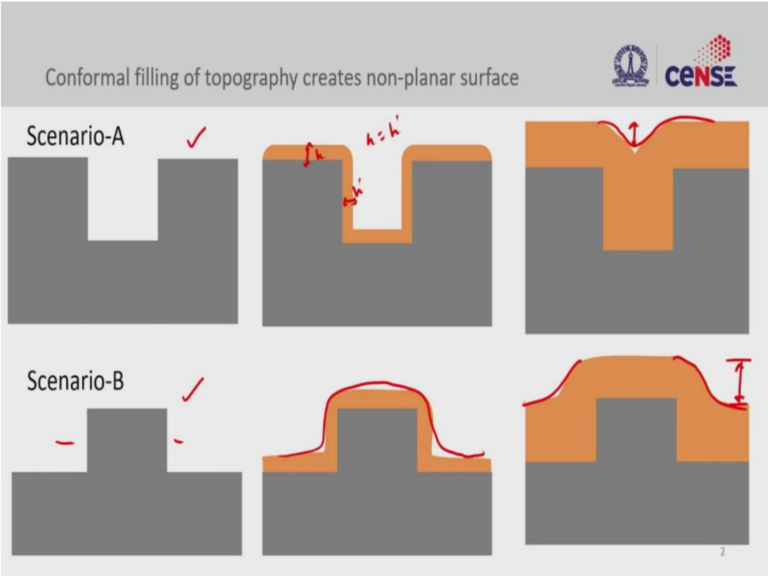
Fundamentals of Micro and Nanofabrication
Prof. Shankar Kumar Selvaraja
Center for Nano Science and Engineering
Indian Institute of Science, Bengaluru

Lecture – 51
Chemical Mechanical Polishing (CMP): Basics

In this lecture, we are going to look at Chemical Mechanical Polishing. So, chemical mechanical polishing is a process that is used to make wafers flat. So, let us look at why we want to make it flat and why the wafers are not flat to start with.

(Refer Slide Time: 00:46)



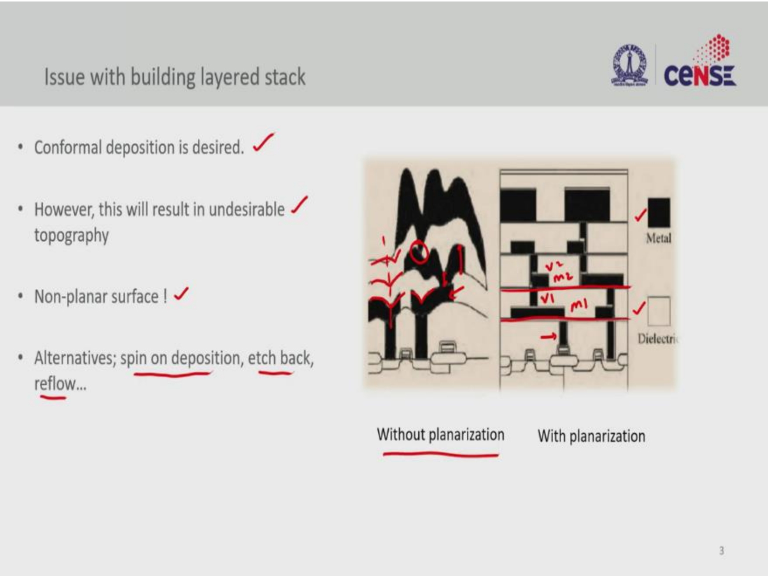
So, here we see two scenarios, there is scenario A and scenario B. So, in scenario A, we have a trench, and we are trying to fill up the trench using conformal deposition, so this is the starting cross-section. So, the slide shows the three different stages; the cross-section on the wafer with a trench, and then we start filling the trench with a material, which is finally conformally filled.

Conformal means the vertical thickness, h, and horizontal thickness, h', should be equal. So, if we have a uniform thickness coating at the top, the sidewall, and the bottom, we call it a conformal deposition. It is good to have conformal deposition for many applications, but then if we fill a trench, we end up with a topography that is not a flat surface.

Now let us look at an isolated line with spaces at the sides. Here again, we start with a conformal deposition process where we deposit the material around and then we increase the thickness.

We end up with the sloped sidewall, and the surface is not going to be flat. Both cases are problematic if we want to build circuits on top of this. Not necessarily many layers, with even one extra layer, the topography here could cause serious issues.

(Refer Slide Time: 02:45)



Let us look at some general expectations from the wafer fabrication process and the issues we face when building a multi-layer stack. So, we tend to use conformal deposition, which gives a very good fill but results typically in topography on the surface; that means it will not be flat anymore, so we call non-planar surface.

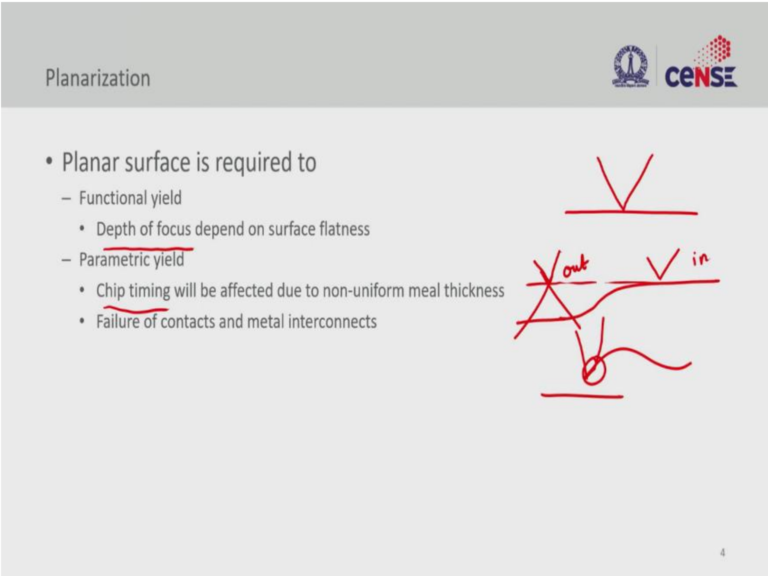
In the figure, we see the result of such a conformal deposition, so we always have sort of neck that is formed. Then the subsequent layers are also going to follow the same topography leading to layers of unequal thickness. For instance, the black region is a metal region. But, if we look at the metal thickness in two different regions, they are different. Also, the region shown by the red circle could result in device failure because the resistance in this particular contact will be very high.

Ideally, we want flat geometries. Once the metal deposition is done on the front end, the surface can be made flat. Then we deposit metal layer-1 and via-1 followed by flattening the surface for subsequent metal-2 and via-2 and so on.

In order to achieve this, two types of materials are used; metal and dielectric. So, we want to make sure that we will be able to create a flat surface while depositing metal and dielectric. There are some alternate methods to address this topography issue: spin on deposition, etch back method, and also reflow.

But some disadvantages come along with these methods. For instance, every material cannot be coated by a spin on deposition. We can deposit dielectrics using spin on, but we can not do the same for metal deposition. In reflow, the solid material is nearly melted, and once we melt the material, it tries to reflow and becomes flat. Metal and dielectrics require very high temperatures to melt them. The thermal budget of the process would not allow heating them, so we would not be able to do this reflow also. Therefore, we need better ways to make this surface flat.

(Refer Slide Time: 05:44)



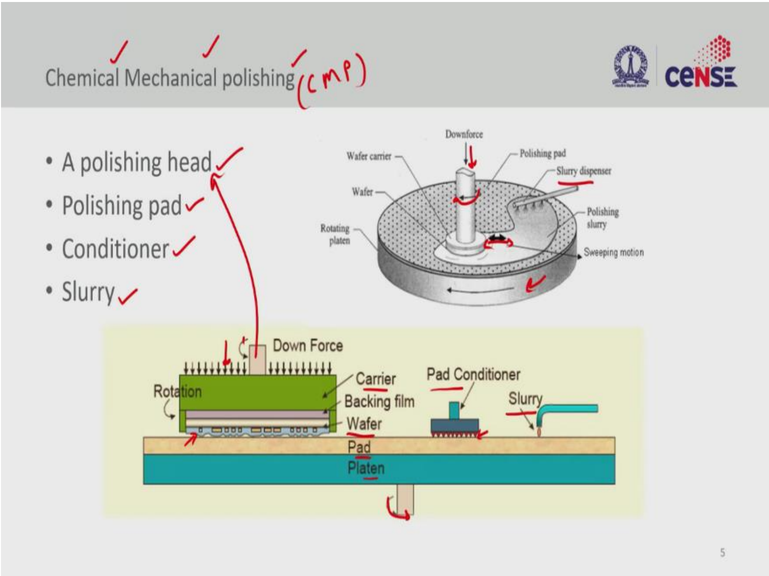
Planarization is needed because there are functional and parametric yield issues are related to this flatness. So, one important thing is the depth of focus, which we saw in optical lithography where the focus onto the surface is very important; this is where the image will be very sharp. But if the surface is non-planar, then the focal point is going to be very different.

For instance, if the surface has topography and we focus on the point which is relatively at greater height than the rest of the regions on the surface, then the rest of the regions will be out of focus. This could result in lithography challenges. The next is the parametric yield. So, we saw in the earlier slide, the metal line thickness can vary, and in some cases, the whole cross-section can be very different. This would result in a change in resistance and will also bring in delays. So, let us say we have multiple metal lines connecting different transistor blocks or different functional blocks, and we want to transmit and receive data that is synchronized. If we do not have uniformity, we will have timing errors which will be challenging to adjust.

So, we want the metals to be uniform in cross-section, and for that, we need a flat surface. Earlier, we saw thin metal contact on the topographic surface, which will act as a device failure point. So, we want to make sure that there are no failure points when connecting two metal layers.

It could result in an open circuit at some point or result in reliability issues where we could have very high current crowding. And then, we will start seeing electron migration and other effects. So, we want to ensure that we will avoid these scenarios that could result in failure or suboptimal operation of the chip. But at the same time, the subsequent processes should not be affected. In this case, we do not want lithography to be suboptimal because we do not have a flat surface.

(Refer Slide Time: 08:26)



So we use chemical mechanical polishing. As the name suggests, we use chemicals and mechanical force to do polishing. The setup has these four critical elements; a polishing head where the wafer is loaded, a polishing pad, a conditioner, and a slurry.

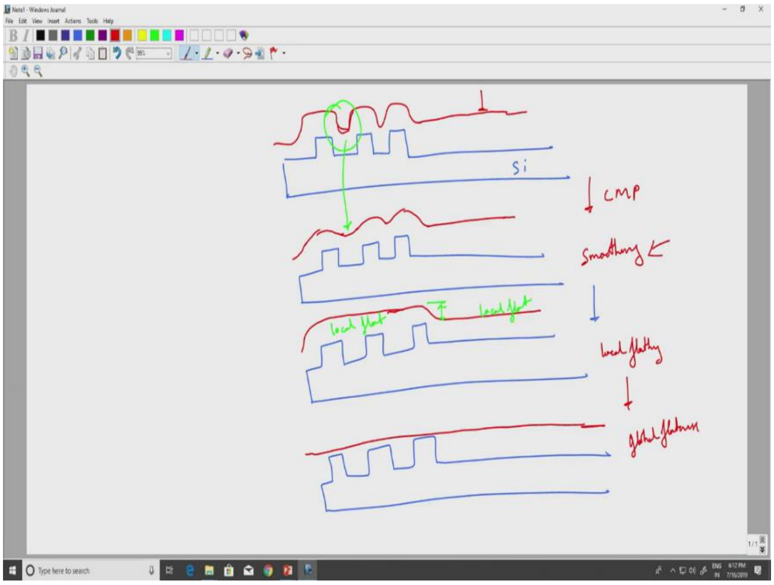
So, the bottom image in the slide depicts how we do chemical mechanical polishing. So, there is the polishing head, and the wafer is put upside down. We can see the topography on the wafer due to the non-uniform coating of silicon dioxide on silicon. The wafer is held by a carrier attached to the polishing head. It is then pressed against a soft pad with force. The soft pad is situated on a very hard plate.

Both the hard plate and the polishing head can rotate, and the soft pad will flatten or polish this topography. The pad can become smooth over time. This is similar to sandpaper. If we want to remove the rust from the metal and make it smooth, we rub it against sandpaper. The sandpaper, after a while, becomes smooth; because it will collect all the material from the rusty metal. In the case of the pad, a conditioner is used to roughen it. So, the conditioner will have this diamond tip that is going to make the pad rough. Similar to abrasive particles in sandpaper, we also need an abrasive here, and a slurry in a liquid form supplies it. So, these are the four elements required in polishing.

The picture on the top has the configuration showing how the polishing is done. So, we have a very large plate that has the wafer head. And then it has a pad at the bottom which will remove the material in the presence of slurry. So, the slurry is dispensed on the pad and then the polishing head is pressed against the pad. And then, we can rotate both the polishing head and the plate. So, there are different motions possible; both can move in the same direction or opposite directions.

The rotation direction and also the configurations vary from method to method. But in principle, we want at least one of these elements moving so that we have friction between the wafer surface and pad to remove material.

(Refer Slide Time: 12:35)



Let us look at the working of a CMP process. So, let us start with a very simple structure. We have a silicon substrate with a certain topography, and we deposit a material, let us say, silicon dioxide, on the top of it. Now the silicon dioxide is going to follow the topography. The goal is to flatten the topography and level the surface.

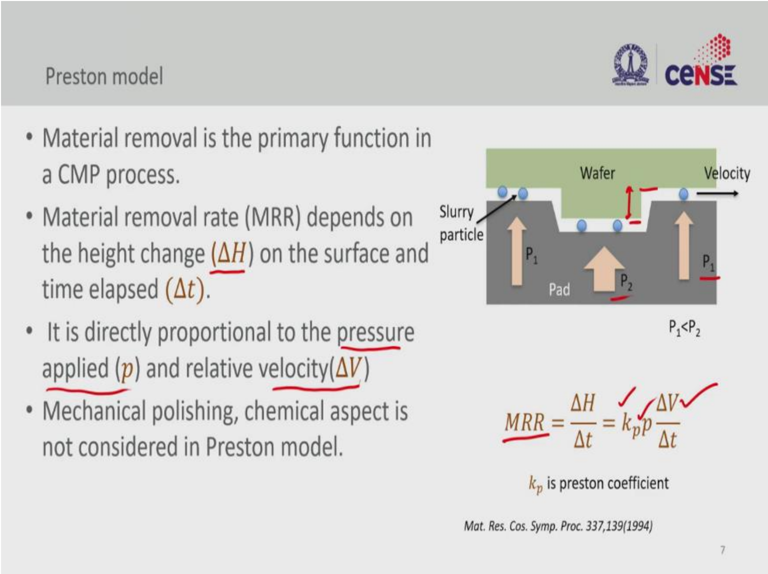
The first step is the smoothing process. So, we will put the wafer through the CMP process; that means we are putting this wafer against a pad and then start rubbing in the presence of slurry. Then, the topography smoothens out a bit, and we get a layer indicated in the second diagram from the top. So, in the smoothening step, we still have topography. But we have converted large topography to a very small topography.

The next process that happens while we are doing a CMP is local planarization. It is shown in the third diagram. So, now we have a flat surface in the local region, but globally the surface is still uneven. We get a flat wafer surface in the subsequent step called global flattening or global polishing.

So, this is how we do the chemical mechanical polishing process. When we first put the wafer into the pads and start polishing, the smoothening process is the first thing that would happen. Once we get reasonable smoothness or reduction in the topography, we achieve local flatness in the next step. Finally, we get global flatness; the whole wafer is now flat.

If we do not do global fattening, we get local flattening alone. But still, we do get some planarization range that is undesirable in many of the applications. So, this is how the CMP process progresses.

(Refer Slide Time: 17:26)

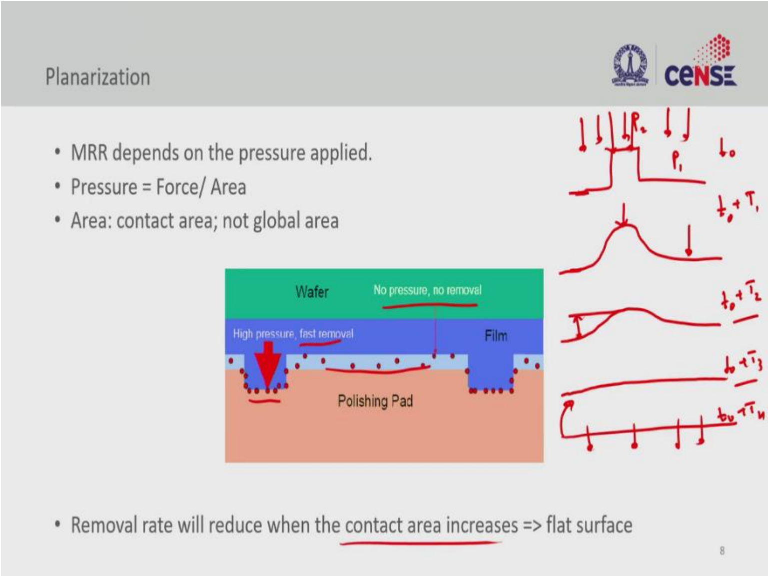


So, let us look at the mechanism used in removing the material. We shall see how the pad, the slurry, and the pressure between the wafer and the pad affect the material removal.

The material removal rate depends on the height difference between the topography and the flat surface and the time spent on polishing. That is something straightforward. But it also depends on the pressure we apply, which depends on the height difference. Two pressures depicted here are P_1 and P_2, where P_1 is less than P_2 because of the topography here. That is the reason in the earlier statement we said it is a height difference that matters. In other words, it is the pressure difference that we have between the flat layer and the topography and also the relative velocity between the wafer and the pad that matters.

Just applying pressure will not help in stationary function. Either wafer or pad should move or both of them could move. They should move at a certain velocity. The faster they move, the removal rate is higher. The pressure and the velocity play a huge role in dictating the material removal rate. So, the material removal rate depends on the velocity gradient that we have, the pressure that we apply, and the Preston coefficient or constant. The Preston coefficient depends on the particle that we use, the type of pad we use, and other configurations.

(Refer Slide Time: 20:34)



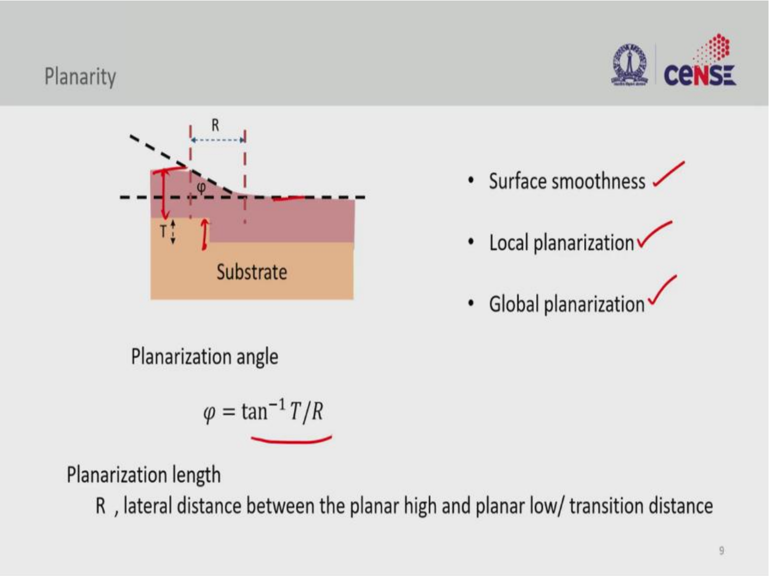
This is how the pressure distribution will look like. On the top is the wafer and the bottom is the polishing pad. This red dot is showing the pressure distribution. When the wafer is flat, we do not have much pressure that is being applied. So, there is no removal at all because it is a physical process; we are pressing against the surface and rubbing it.

If the pressure is low, the removal rate is very low. But then, if the pressure is very high because we have this topography, we will have faster removal of the material.

Let us say we have a topography, as drawn in the slide. When we apply plate pressure, in this case, from the top, pressure P_2 on the topography is high. A smaller force is applied on the sides. So, there will be a higher removal rate at the center and a lower removal rate at the outside. In the first step, i.e., the smoothening process, more material is removed from the center than the sides. The first diagram is the wafer at a time, say t_0. And after smoothening at $t_0 + T_1$, we see the topography is reduced a bit. If we keep on progressing at $t_0 + T_2$, because the topography is reducing, the pressure will also reduce. After a while, at $t_0 + T_3$, we achieve flatness. Beyond this point, the flatness is maintained. However, there will be loss of material.

So, when the topography is reduced, we have reduced force and reduced removal. It is because the contact area increases, and the pressure reduces and eventually reduces the removal rate. So, we need to understand at what point we should stop to maintain the flatness and not lose material.

(Refer Slide Time: 24:05)



Here is how we characterize the planarity. Planarity is nothing but the angle between the global flatness or the local flatness that we have. It strongly depends on the topography that we start with and the material we have, and the flatness we are looking for.

So, based on this, we could characterize the flatness, which is the topography over the length scale, and how far these two flat surfaces are to the topography that we have. So, as we keep on polishing, we would have a smooth surface initially, and we try to get local planarity done. This is the intermediate stage. Then, as we keep on progressing, we will get global planarity. So, the planarization angle we mentioned here comes typically when we have local planarization; the local structures are flat, but the global flatness is not yet achieved.

If we keep on progressing in the polishing process, we will achieve global planarization and typically, the planarization angle will be nearly zero.

So, this brings us to the end of the understanding CMP process. We have seen what happens on the surface and how the CMP process progresses, starting from very sharp features smoothened to get local flatness and finally achieve global flatness. In order to achieve this, we need four components. The wafer that is in a moving or stationery holder, and pad that comes into contact with the wafer holder, and then we need a slurry that supplies abrasives and, also, we need a conditioner. So, slurry in a fresh pad can work for some time. But after a while, the slurry is going to occupy the pad completely. Then conditioner is the one that is required to remove all the settled particles and make the pad more abrasive.

So, once we understand the importance of these processes, we can design and understand each step's implication, which we will see in the next lecture.