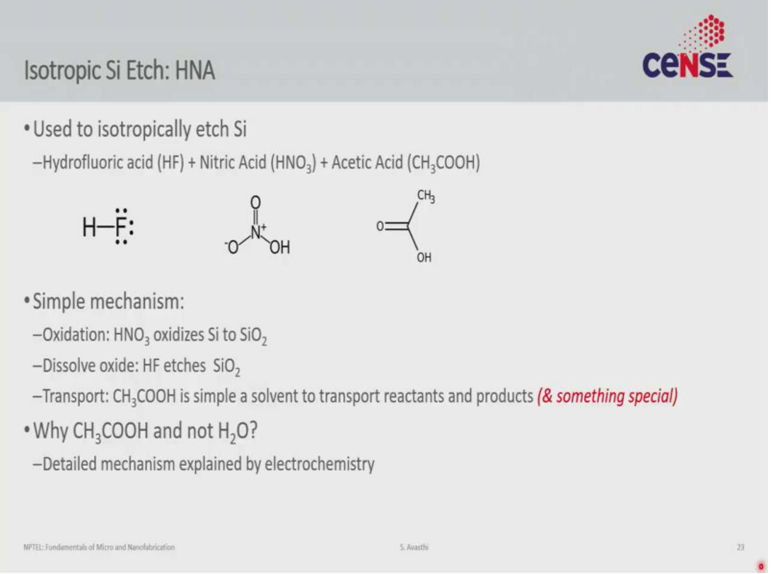
Fundamentals of Micro and Nanofabrication
Prof. Sushobhan Avasthi
Centre for Nano Science and Engineering
Indian Institute of Science, Bengaluru

Lecture – 44
Wet Etching Recipes

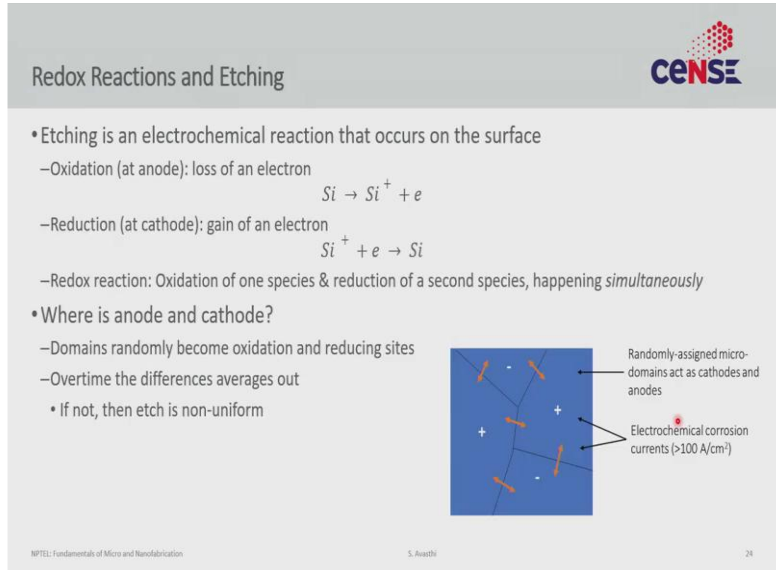
We shall continue the module on subtractive processing. In this lecture, we shall go into a little more detail about some specific wet etch recipes, specifically that do anisotropic etching. We will also look at the detailed mechanism of the HNA etching and try to make some generalizations.

(Refer Slide Time: 00:55)



In the first lesson, we shall talk about the isotropic silicon etch, colloquially called HNA etch. The H here strands for hydrofluoric acid, N for nitric acid, and A for acetic acid, the diluent. The primary mechanism is easy to understand. Silicon is a reduced elemental semiconductor. You first oxidize it using HNO_3 to SiO_2 that gets etched with the HF to form a complex that we saw in the last lecture. Acetic acid is simply a solvent to transport the reactants and products. The HNA is an interesting etch. The acetic acid also does something special that we will discuss. You only understand an etch's details if you consider it an electrochemical process, not just simple dissolution chemistry.

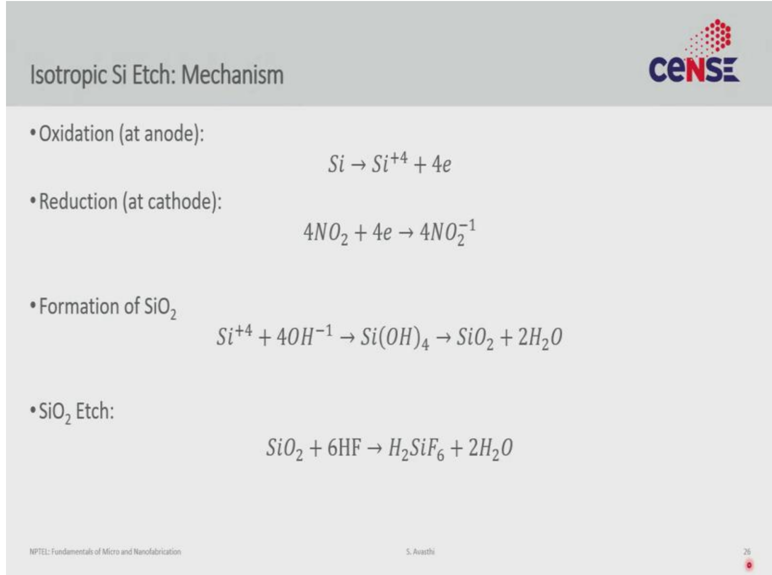
(Refer Slide Time: 02:05)



On any etch surface, you randomly form positive and negative domains (anodes and cathodes), where the electrochemistry happens. You get an electric current between these positive and negative domains because of the potential difference. Since they are so close to each other, the electrochemical corrosion currents that go through the surface can be enormous; as large as 100 A/cm^2. This course is not a device course, but those with that background may appreciate that this current is relatively high. These domains don't remain static. They come up because of perturbations, not because of the applied voltage. They may change polarity, size, and direction randomly over the etching process.

During the HNA etch, at the anode, Si loses an electron to the anode to form Si^+. At the cathode, Si^+ accepts an electron and reduces to form elemental Si. A redox reaction is one in which one species oxidizes and one reduces simultaneously. These reactions happen randomly over the whole substrate; At some point, an oxidizing surface can start reducing. If the etch is uniform, you expect these two mechanisms to even out over the etch-time, so you don't see any differences and get a smooth film. Otherwise, you start seeing non-uniformities.

(Refer Slide Time: 04:33)



Silicon is a group IV element and has a valency of 4. It oxidizes to Si^{+4} and gives $4e^-$ to the anode. At the cathode, 4 NO_2 molecules react with the electrons to form 4 NO_2^- ions. From where does this NO_2 come? We start with HNO_3. The NO_2 comes from HNO_3, and we will discuss the mechanism in the next slide. For now, assume there is NO_2 in the system that reduces at the cathode. These two reactions are equal and opposite electrochemical reactions. Si^{+4} ions react with the hydroxides to form silicon hydroxide $Si(OH)_4$, which dehydrates to form SiO_2 and water. The HF reacts with SiO_2 to form this soluble complex, thereby etching SiO_2 and exposing new silicon that can once again oxidize at the anode, and the process continues.

So, the intuitive understanding that we started with in the first slide was not too off. The exposed silicon gets oxidized with NO_2, not HNO_3. We have still not answered where this NO_2 comes from and the unique role of acetic acid in this combination.

(Refer Slide Time: 06:47)

Isotropic Si Etch: Role of HNO_3

- HNO_3 is little complex
  - But ultimately, it provides NO_2 that oxidizes Si

$$2HNO_3 \rightarrow 2NO_3^{-1} + 2H^{+1} \quad \leftarrow \text{Ionization}$$
$$HNO_2 + HNO_3 \rightarrow 2NO_2 + H_2O \quad \leftarrow \text{Formation of } NO_2$$
$$2NO_2 + 2e \rightarrow 2NO_2^{-1} \quad \leftarrow \text{Reduction}$$
$$2NO_2^{-1} + 2H^{+1} \rightarrow 2HNO_2 \quad \leftarrow \text{Regeneration of } HNO_2$$

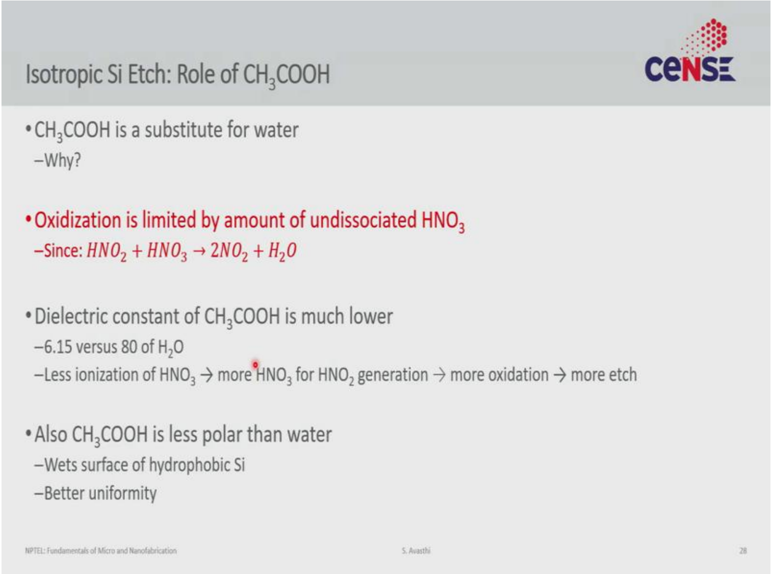
- HNO_2 is required
  - Thankfully some of it spontaneously generates in HNO_3
  - But we need to wait for some starting HNO_2 → needs an "induction" period.
  - Else add NH_4NO_2
- Subsequently, HNO_2 is auto-catalytically regenerated

Let's talk about the presence of NO_2. HNO_3 is oxidizing acid, but the exact mechanism is a little complex. It provides NO_2. One of the products that form during this process is HNO_2. The first reaction should be obvious; HNO_3 is a strong acid and likes to ionize. It creates NO_3^- and H^+ ions. Competing with this ionization process is the oxidative ability of HNO_3. What is the oxidation reaction? The next three are the set of oxidation reactions. So, let us assume there is a little bit of HNO_2 in the system. HNO_2 reacts with HNO_3 to form 2 NO_2 molecules and water. I have marked it in red because this is what leads to NO_2 formation. It is also the slowest step in this chain, and hence, rate limiting. The amount of NO_2 decides your ability to etch. NO_2 reduces to NO_2^- ions, as we have seen in the previous slide. Finally, NO_2^- ion reacts with the H^+ to form HNO_2.

Notice that we start with 1 HNO_2 molecule but end up with 2 HNO_2 molecules. In this mechanism, HNO_2 participates as a catalyst. A self-regenerating catalyst. Even if you have one molecule of HNO_2, if you give the solution enough time and something to oxidize, it will automatically generate more HNO_2, which will enhance HNO_3's oxidation ability. From where does the HNO_2 come? Maybe because of impurity or random perturbations, but once even one molecule forms, the mechanism generates its HNO_2 to catalyze itself. HNO_2, thankfully, generates spontaneously; it is not something we have to add explicitly. Sometimes the HNO_3 solution is not active from the get-go. In that case, for some initial time, the etching does not work.

Hence, there is an induction period. It takes a certain amount of time for these reactions to happen and for enough HNO_2 to generate. Till that time, the reactant seems to be doing nothing, but if you wait, you will see them becoming active. For very shallow etches, where you cannot tolerate the induction period or the variability introduced by this induction period, you can always add some NH_4NO_2 into the solution. It provides you with the NO_2^- immediately, thereby avoiding the induction period altogether. We have discussed the source of NO_2 and HNO_2; however, we have still not answered where, in this whole scheme of things, does acetic acid plays a role?

(Refer Slide Time: 12:05)



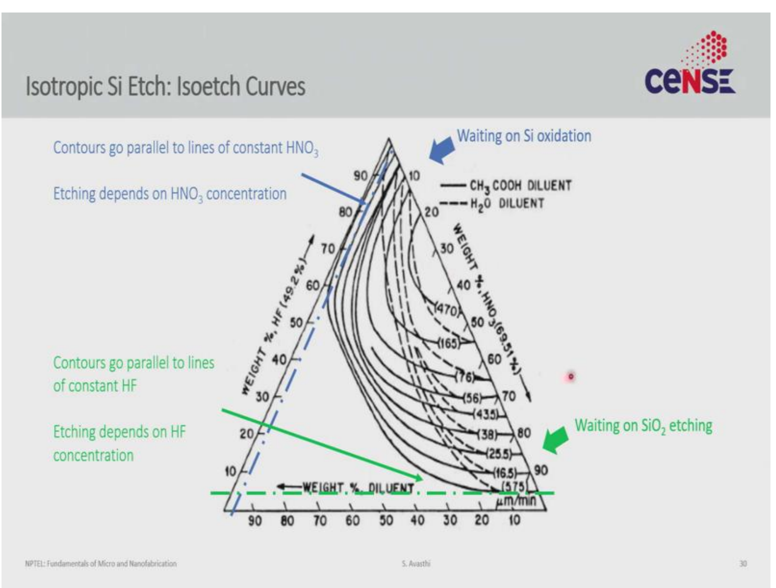
What is acetic acid substituting? I said acetic acid is the diluent or a solvent. So, it is a substituent for water. We could have done this HNA as HN water instead of acetic acid, but we deliberately choose acetic acid. So, how is acetic acid different from water? Think about it; another hint is the red reaction (dissociation of HNO_3 and HNO_2 into NO_2 and H_2O) limits oxidation. What do you think is the role of acetic acid?

The dielectric constant of acetic acid is only 6.15, as it is an organic molecule. It's relatively low if you compare this to water, a polar molecule with a high dielectric constant of 80. The dielectric constant reduces the energy of ionization. In a more polar or a high dielectric constant liquid like water, you tend to ionize more than a solvent with a low dielectric constant. Does that get you closer to the answer? In the previous slide, the first equation described ionization. If you do this HNA etch in water instead of acidic

acid, a higher dielectric constant increases screening, and ionization. If HNO_3 doesn't ionize, there is more available to react with HNO_2 to form more NO_2. These are the competing mechanisms; HNO_3 can either ionize as a strong acid or dissociate with HNO_2 to act as a potent oxidizing agent. Now that we have reduced the possibility of ionization, it helps oxidation, and since that is the rate-limiting step, it enhances the etching rate. In general, with acetic acid as the diluent, you get more etching than with water.

Also, acetic acid is less polar than water and has lower surface energy. It wets the surface of hydrophobic silicon. In cleaning, we discussed that after stripping away its oxide with HF, water does not stick to the wafer surface very well; it is one of the oxide etch-completion indicators. After you etch SiO_2, the new silicon you expose is hydrophobic. If you do this in a medium that wets the hydrophobic Si surface better, you get better performance, uniformity, and etching. Acetic acid is one such solvent that wets the hydrophobic silicon better than water. Those two are the reason you use acidic acid.

(Refer Slide Time: 16:23)



Look at a specific composition; 50 % HNO_3, 20 % acetic acid, and 30 % hydrofluoric acid. Where is this composition on this iso-etch curve? Go to the 50 % point and draw a line parallel to the tick mark on the HNO_3 axis. Then draw a line for the 20 % acetic acid and 30 % HF. These lines intersect at a point. Each point on this graph represents a unique combination of these three components and a corresponding etching rate. For example, at this composition, the known etch rate is around 50 µm/min. If you notice,

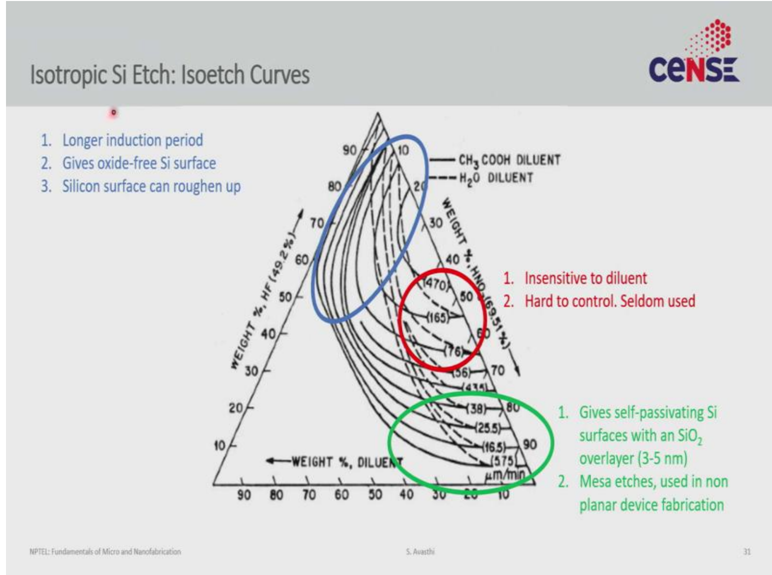
this composition is close to this contour with a value of 56 µm/ min. It is no surprise that the etching rate is nearly 50; All the points on this contour have the same etch rate – 56 µm/min. The next one has an etching rate of 43.5 µm/min. The highest etching is where you both have a lot of HNO_3 and a lot of HF, which is not surprising. So, very little diluent, a lot of HNO_3, and a lot of HF lead to a lot of etching. The peak contour represents all the compositions, where the etching rate is 470 µm/min. You can go through a whole wafer in a minute!

The dashed lines represent the etching rates with water as a diluent. For example, this composition is very close to the contour with an etching rate of 56 µm/min, but if the diluent was water, the nearest dashed line has an etching rate of only 38 µm/min. It is around 25 – 30 % lower. It is true of all compositions; with water, you get lower etching rates than acetic acid. So, that is the great advantage of using acetic acid as the diluent.

Let's discuss the corners of the isoetch curve. The blue line (dot-dash) represents all those compositions where the nitric acid is minimal. It is parallel to the tick mark on the HNO_3 axis, corresponding to a low concentration of HNO_3, about 2 %? The rest of the 98 % is either HF or acetic acid. There are contours (solid lines) parallel to this. All these compositions have the same etching rate, denoted by this contour; 16.5 µm/min. For the combinations with very little HNO_3, it does not matter how much HF you have; whether you have 60 % or 90 %, you get the same etching rate. If you shift this line from 2 % HNO_3 to 10 % HNO_3, the line becomes parallel to a contour with a higher etching rate. Again, independent of the HF concentration. Remember that oxidation and etching need to happen in a series. With very little HNO_3, you have to wait for the silicon surface to oxidize. Oxidation, being the slowest step, decides the rate. Adding more HF will make the oxide etching faster, but it is not the bottleneck.

Similarly, you can look at the other edge. This green line represents all those compositions with minimal HF, say 4 %. The balance 96 %, some part is HNO_3, some acetic acid. There is a little bit of diluent at the bottom right corner, but most of it is HNO_3. Say 70 to 90 %. For all those cases, the etching rate is 5.75 µm/min, independent of HNO_3 concentration. If you change the HF concentration from 4 % to 10 %, this line shifts up. Now, irrespective of how much HNO_3 you put in, you get an etching rate of 16.5 µm/min. Here, you are waiting on the SiO_2 to get etched. Oxidation of silicon happens quickly; there is a lot of HNO_3 in the system.

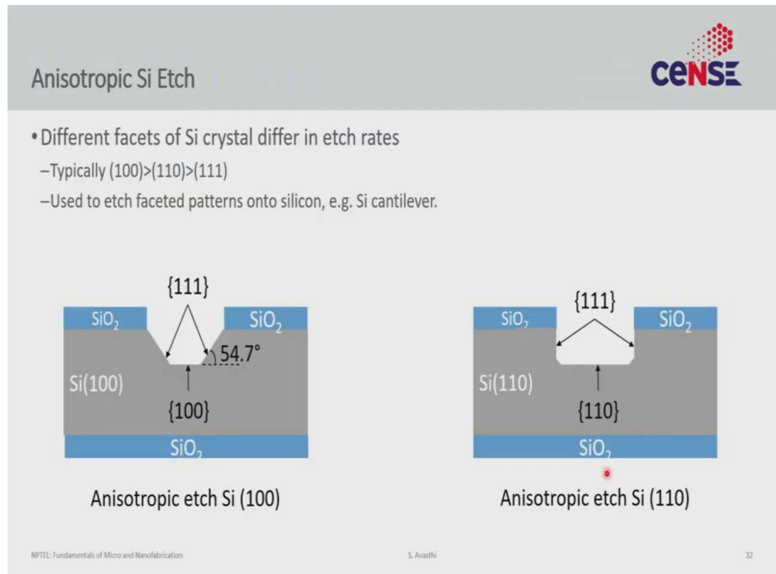
(Refer Slide Time: 24:15)



The corners marked by the blue ellipse with a little HNO_3 and large HF proportion, and the green ellipse with a large HNO_3 and a small HF fraction have the same rate - 16.5 µm/min. What you choose depends on the application. In the blue corner, the HNO_3's ability to oxidize silicon limits the etch rate. So, all the peculiarities of HNO_3 as an etchant will affect the etching. It might be sensitive to the induction period for HNO_2 generation. As the surface is always waiting on the silicon to get oxidized, it is invariably oxide-free after the etch. That can be an advantage or disadvantage depending upon the application. As there is no surface oxide and a lot of HF, you get surface roughening. As we have discussed in cleaning and HF etching, HF does not etch silicon, but it can micro roughen it. So, all the compositions in this corner suffer from that issue.

In the green corner, you wait on HF to show up and etch SiO_2. What characteristics can you expect? The silicon surface you get from these compositions will have a SiO_2 layer on top after etching because SiO_2 etching is the bottleneck. At any moment, there is always some 3-5 nm thick SiO_2. Because of this passivating SiO_2 and low concentrations of HF, you do not roughen up the surface. You get much smoother, planar etch. It is useful if you want a polished finish. The red region in the middle has both a lot of HNO_3 and HF. Small changes in concentration cause considerable changes in etching rates, which are very hard to control. That is why you rarely use this red zone. You want to either be in the blue zone or the green zone. If you need bare silicon, choose blue; for a smoother/polished surface, choose green.

(Refer Slide Time: 27:19)

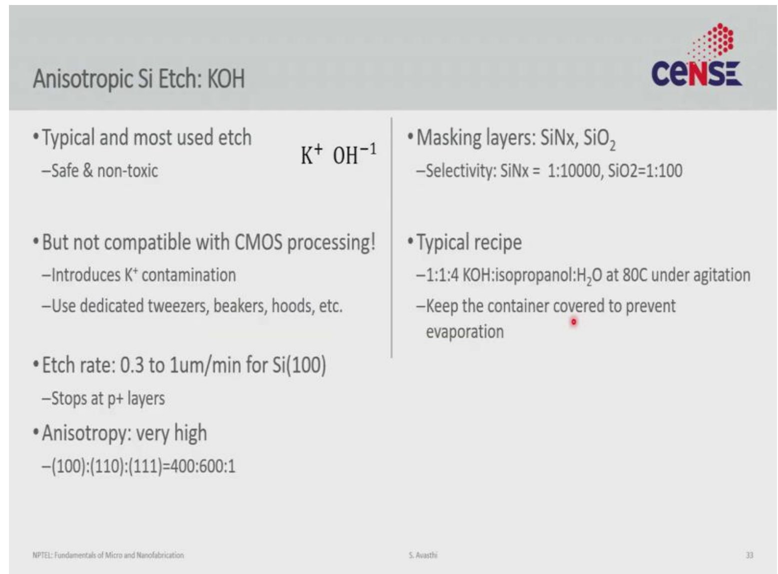


We have discussed the HNA etch. Let us examine another prominent class of silicon etching - anisotropic etch. It is an etch dependent on silicon's facets or orientations and is not uniform in all directions. Anisotropic etching of silicon (100) gives you these slanting (54.7^{o}) profiles. It etches (100) and (110) surfaces very quickly, but (111) very slowly. So, the (111) faces at 54.7^{o} to (100) remain, while the other faces etch away, and you get these sloped cross-sections. Why is there an anisotropy in the etch rate?

We talked about how (111) orientation is the easiest to make during crystal growth because it is more stable. The same order applies in reverse. The etching of (100) is faster than (110) and much quicker than (111). It is advantageous in MEMS, where you want to make a mechanical structure like a silicon cantilever. Towards the end of this course, we have a lab module cantilever fabrication. It requires anisotropic etch.

The etch profile strongly depends upon the properties of the material as well as the orientation. In an anisotropic etch of silicon (110), you don't get this triangular or a trapezoidal profile, but precisely vertical as the exposed (111) faces are perpendicular to the (110) surface. Some of these etches are very useful because it is tough to get precisely vertical profiles in an etch. This anisotropic process automatically gives you an utterly vertical sidewall, which is very useful for microfluidic applications.

(Refer Slide Time: 29:57)



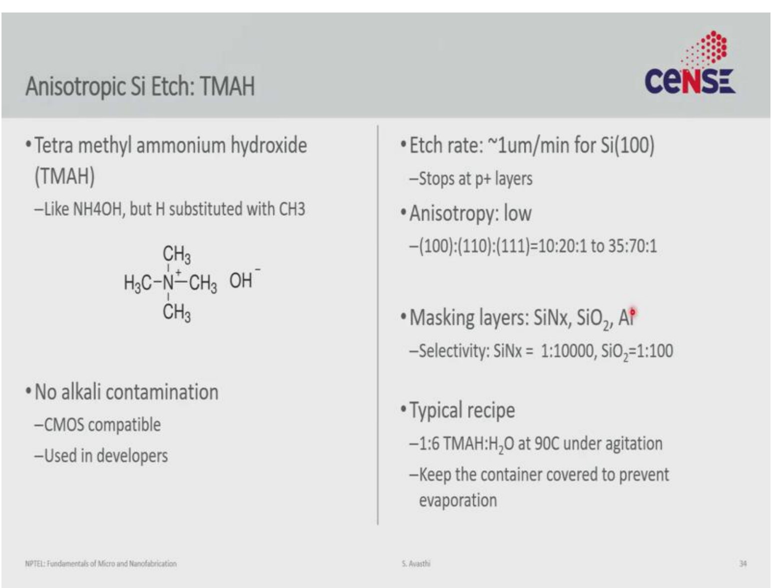
What types of anisotropic etches can we access? Most of the anisotropic etchants are bases. The simplest one is potassium hydroxide (KOH). It is safe, non-toxic, cheap, and a strong base, so it readily ionizes into K+ and OH- ions. The problem is it is not CMOS processing compatible. During cleaning, we talked about preventing alkali metal contamination like Na^+, K^+, how sweat is terrible, and you should never touch the wafer, etcetera. The ITRS requirement allows a minimal amount of Na^+ and K^+ ions. And here, we are proposing the use of a chemical that explicitly has potassium contamination. It is incompatible with CMOS. CMOS is what Intel does; they would never allow this process inside their fab.

MEMS fabs are not interested in electronic devices, at least of the caliber that Intel makes. They are okay with potassium hydroxide. Another example is silicon solar cells. For researchers working in an academic fab, we can't afford different fabs for CMOS processing and MEMS work. We can afford separate tweezers, beakers, hoods; you must try to segregate CMOS and non-CMOS processes as much as possible; otherwise, you have a soup where nothing works. The typical etch rate is around 0.3 - 1 µm/min for silicon (100); Interestingly, this etchant does not etch the highly doped p-type layers very fast. You would think this is an issue, but it is an advantage; we will talk about this as an etch stop a little bit later. In this process, the anisotropy is very high; the etching rate of the (100) facet is ~ 400× that of the (111) face. The etching rate of the (110) is ~ 600 times (111) etch rate.

What do you use as a mask? KOH is a base. During the photolithography lecture, you must have discussed developers. Chemically, developers are also bases, which means the photoresist dissolves in bases. KOH may dissolve photoresist, and hence, a photoresist is not a good masking layer for anisotropic etching. You have to use hard masks. Si_3N_4 or SiO_2 are common choices. You first pattern the SiO_2 using HF and then use it to pattern silicon. You can get very high etch selectivity; 10000:1 for nitride and 100:1 for oxide mask. A mere 100 nm Si_3N_4 is enough to etch 100 µm silicon.

A typical recipe has some KOH, IPA, and water. We will not discuss what IPA does. We do this at a little higher temperature and under agitation. You may look up the conditions where anisotropic etching becomes more probable. These etch's can be reasonably long if you want to etch deep structures. During those hours, you don't want the concentration of the etchants to change too much. You do it in a closed container to reduce evaporation. At 80°C, if you keep it open for a few hours, most of the water will just evaporate.

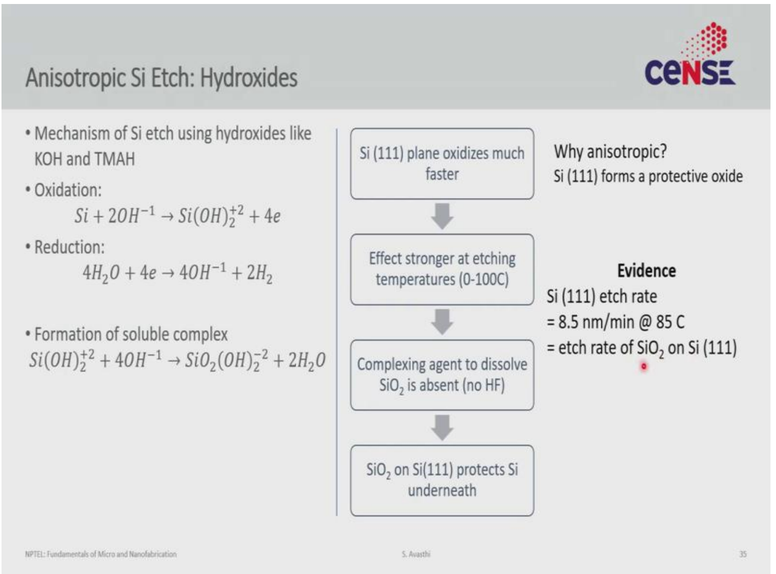
(Refer Slide Time: 34:29)



For those of you who want a CMOS compatible anisotropic etch, there is good news. KOH is not the only etchant in town; it is just the cheapest one. For CMOS compatibility, you can use tetramethylammonium hydroxide (TMAH). It is similar to ammonium hydroxide, where methyl groups (CH_3) substitute the hydrogen. TMAH is also a strong base and does the same job as KOH. It provides OH^- that do the etching. The advantage is there is no alkali like Na or K. In fact, CMOS compatible metal ion free

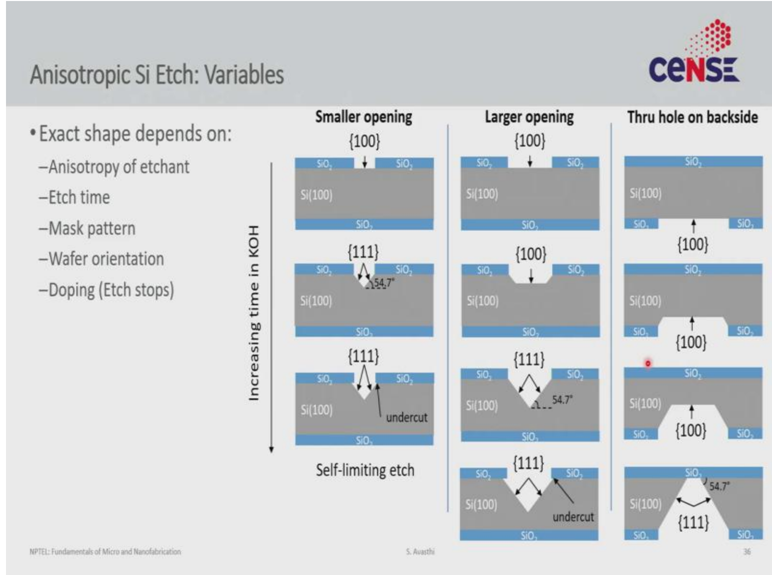
developers use TMAH as a base, not KOH. However, you still require a hard mask as the photoresist dissolves in TMAH. The etch rate is 1 μm/min for silicon (100). Incidentally, this etch also stops at p+ layers. Anisotropy, as compared to the KOH, is a little lower. Between (100) and (111), you only have anisotropy of 10 – 35, not 100s as earlier. The typical recipe has a dilute TMAH at 90°C under agitation. Again, these etch's are long. So, you must cover them, otherwise the etching concentration and performance changes.

(Refer Slide Time: 36:29)



What is the underlying mechanism? Both the hydroxides KOH or TMAH functionally do the same thing. They provide the OH^- which forms a soluble complex with silicon. Why is it anisotropic? Fundamentally, this is an oxidation reaction. Silicon compounds with the OH^- and gives e^- that reduce water at the counter domain. The rate at which the silicon (111) oxidizes is much faster. It forms a silicon dioxide layer that protects the underlying silicon (111), reducing the etching rate. Silicon (100) does not have that protective oxide. So, it continues to dissolve, and you get all etching only in that direction. What is the evidence for this mechanism? If you look at the etch rate of the silicon (111) in TMAH, it is 8.5 nm/min at 85°C, a low number. It's the same as the etching rate of SiO_2 that forms on (111), which strongly suggests that the slow step preventing Si (111) etching is removing the passivating oxide.

(Refer Slide Time: 38:25)



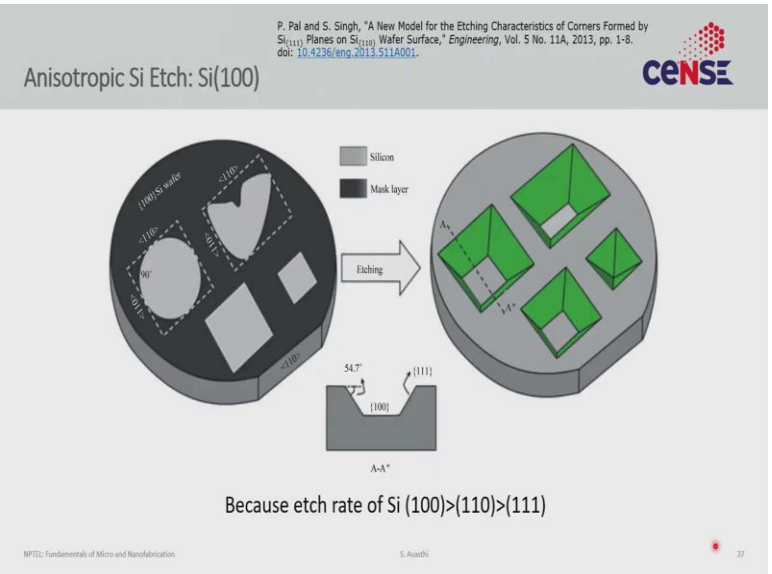
We shall now discuss practical detail on implementing this in the MEMS system. What shape you get depends on the agent; TMAH versus KOH versus HNA, but also on the etch time, mask pattern, wafer orientation, and etch-stops. The last four are not obvious, and through examples, I will highlight each of these.

You pattern an opening in the SiO_2 hard mask on Si (100) wafer while protecting the back. Put this system with an exposed (100) area in anisotropic etch like KOH. As the etch progresses, you reach a point where the (111) planes touch. Then, you have no exposed (100). The etch slows down. If you keep it in the solution longer, you will get some etching because (111) also dissolves at 8.5 nm/min, but very slowly. All you will get is an undercut. The longer you keep it, you get a little more undercut. The shape will not change because you expose only (111), the slowest etching facet. This self-limiting etching happens for small openings or long etching times. If you have a larger window, it will take longer to get to a self-limiting behavior. The etch-time is on the y-axis. For the same etch-time, the mask pattern with a small opening has a triangular etching cross-section, while in the large one, it looks like a trapezium. All of this discussion is for silicon (100). A couple of slides ago, I showed that the cross-section of (110) was vertical and looks very different.

The final example is etching a large hole at the back or a through-hole across the wafer's thickness. Those interested in geometry may have noticed that this self-limiting etch-

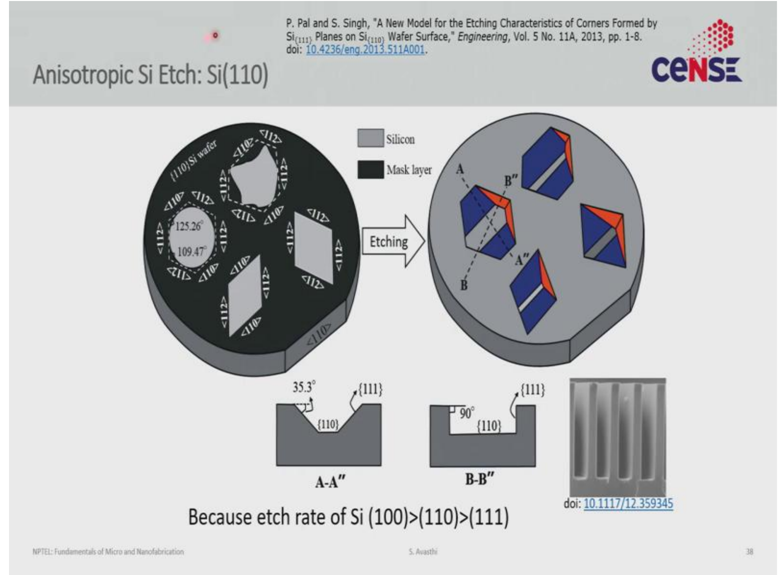
depth is trigonometrically related to the opening you need. You can calculate the window for etching 400 µm. For that large an opening, you can carve a through-hole, where the etching doesn't self-limit for the whole thickness. If you keep etching for a very long time, you will hit the SiO_2 mask on the other side. Now, you have exposed (111) facets that go very slowly and SiO_2, which etches even slower. If you remove the hard mask, you have a hole through the silicon wafer, useful for various applications.

(Refer Slide Time: 42:55)



These lovely are figures from this open-access paper. You have a Si (100) wafer with <110> flat and a circular window in the mask. If you etch it, you get these slopes on a transcribing square parallel to the <110>. In the A-A" cross-section, you have (111) faces at a 54.7^o angle, represented in green. The grey flat at the bottom is (100). It doesn't matter if you have a circle, a square, or some arbitrary shape. In all the cases, you etch these rectangular features on the (100) wafer. If the size is tiny, you get self-limiting etching, where you don't have any exposed (100).

(Refer Slide Time: 44:33)



Let us talk Si (110) anisotropic etching. The primary flat is in the <110> direction. You can draw a hexagon tangential to it, along <110> and <112> direction family. Anisotropic etching of this wafer is a little more complicated. Let us take two cross-sections; one is A-A", similar to what we had in the (100) case. The (110) flat base in grey and (111) slopes in blue are at a different angle than (100). It is because the angle between the (100) and (111) is different from the angle between (110) and (111), which is 35.3^{o}. There is another cross-section B-B" along which the exposed (111) facets (in orange) are precisely vertical. This anisotropic-etch gives you completely flat and 90° channels, which is very hard to do without dry etching. By leveraging anisotropy and selecting the right facets, you get these vertical channels, very useful for microfluidic applications.

In the next lecture, we shall continue anisotropic etching of other materials, specifically gallium arsenide. We shall look at a general class of anisotropic etchant used to delineate defects. We will also look at process integration and put all of this together to form a silicon cantilever. The lab visit for that will come later. We will discuss it theoretically now, and you can then correlate it with what you see in the lab later.