Chemical Etching of Silicon¹

IV. Etching Technology

B. Schwartz*

Bell Laboratories, Murray Hill, New Jersey 07974

and H. Robbins*

Galamar Industries, Palo Alto, California 94304

ABSTRACT

The etching of silicon in HNO₃-HF based systems proceeds by a sequential oxidation-followed-by-dissolution process. In those composition regions where the solution is very low in HNO₃ and rich in HF, the rate-limiting process is the oxidation step. Consequently, electron concentration, surface orientation, crystal defects, and catalysis by lower oxides of nitrogen play an important role. In those compositions where HF is in limited supply, dissolution of the formed oxide is the rate-controlling step and diffusion of the complexing fluoride species is the important factor. Therefore, crystal orientation and conductivity type independence as well as hydrodynamic control are the consequences. In order to meaningfully select an etching composition to solve a specific processing problem, it is necessary to understand this composition-mechanism interaction. Corollary with the mechanism understanding, sample geometry effects have been followed as a function of solution composition. The HF-HNO₃-H₂O solution composition plane has been characterized into various regions where the two basic mechanisms interact and specific processing utilization is shown. Similar results are shown for the system HF-HNO₃-HC₂H₃O₂. In addition, a number of particular etching problems are posed, and solutions offered, that make use of these composition characterizations, and show how one can use their information to solve other practical processing problems.

Most of the data on silicon etching that one can find in the literature involve studies of the mechanisms and kinetics of the dissolution process (1-4). There are a few examples of practical applications, but they deal primarily with crystallographic aspects of the etching, e.g., defect delineation (5) or anisotropic crystal plane etching (6). Because of the sophistication of present-day silicon technology, it appears to be desirable to be able to control the geometrical aspects of a silicon slice, e.g., from sharp, possibly peaked corners and edges to smooth and rounded edges and corners, merely by controlling the chemical etching environments. It would, therefore, seem appropriate to present some data on the geometrical effects observed on the etching of silicon rectangular parallelepipeds in solutions of HNO3, HF, and H2O with and without HC2H3O2. These observations were made at the same time as the previously published kinetics data were being obtained (1,2), but it is only recently that many requests for geometric information have made us aware of the usefulness of disseminating these configurational data.

In order to make the picture more complete, some of the original iso-etch-rate figures will also be included, for comparison purposes.

Experimental Procedure

The technique used for the etching and thickness monitoring has been adequately described in previous publications (1-3). However, in order to make this paper useful, we will have to repeat some of the experimental detail, but will stress only the salient features of the sample and solution preparation methods and how the geometric aspects were followed.

The sample dice were rectangular parallelepipeds $(0.125 \times 0.125 \times 0.025$ in. on side) cut from n-type silicon slices with resistivities about 2 Ω -cm. The dice were cut so that the large-area surfaces were along the (111) plane while the remaining surfaces were

* Electrochemical Society Active Member.

¹ All of the data used in this paper were obtained while the authors were employed at Hughes Aircraft Company.

Key words: silicon, etching, geometry control.

not controlled with respect to any particular crystal plane. Also, the dice were used as received, which means that the exposed surfaces contained saw damage.

All dice were etched three times, and each etch was performed in a fresh portion of the same solution. Since each etching period had been chosen to remove 4 to 6 mils from the specimen, it may be assumed that the work damage had been removed after the first etch. The third etch was performed with a few milligrams of NaNO2 added to the solution for possible catalytic purposes (1). After each etching step, the sample was rinsed in distilled water, dried, and measured for dimensional changes with a micrometer to \pm 0.05 mils, and then the geometry of the specimen was examined under a microscope.

Concentrated hydrofluoric acid, as normally supplied. may vary in concentration from 48 to 52 weight percent (w/o). Similarly, nitric acid may vary between the limits of 69 and 71%. In order to establish control over the composition of the etching solutions, it was necessary to standardize the stock acids from which the solutions were to be made. This was done by titrating samples of the acids against a standard sodium hydroxide solution. Potassium acid phthalate was used as the primary standard for the sodium hydroxide. The first bottle of hydrofluoric acid analyzed at 49.25%, and this concentration was accordingly selected as the standard HF reagent concentration. All subsequent bottles of hydrofluoric acid were adjusted to $49.25 \pm 0.05\%$ by the addition of 60% HF or water as dictated by the chemical analysis. In a similar manner, the value 69.51% HNO3 was established as the standard nitric acid concentration and all subsequent stock bottles of nitric acid were adjusted to this value ± 0.05% by the appropriate addition of water or 90% HNO3. In a second series of experiments designed to cover a larger area of the composition range, 90.45 \pm 0.05% HNO3 and 59.75 \pm 0.05% HF were the standard reagents used. In a third series of experiments, glacial acetic acid was used as a diluent, and the only water present was that found in the original HNO3 or HF reagents.

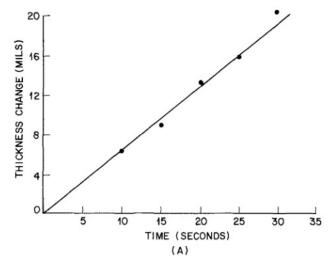
The etching was done in a Teflon beaker 1 in. in diameter and 21/2 in. deep. Agitation was provided by an electric stirrer equipped with a polyethylene paddle. The dice were etched one at a time in 10 cm3 of solution, and duplicate results were obtained by repeating the entire experiment. To quench the reaction at the proper time, a large volume of water was poured into the beaker. The etching technique was later modified when the higher concentration reagents were used. Here the dice were etched two at a time, in 20 cm3 of solution, while encased in a small perforated polyethylene basket. The basket and dice were agitated for the required time in the solution, then immersed in a beaker of cold water to quench the reaction. The agitating action consisted of approximately 150 vertical strokes per minute, with the etching jig rotated between the thumb and middle finger as rapidly as possible during the up and down strokes.

All etching was performed in a constant temperature bath regulated at 25.00° ± 0.02°C. Unfortunately, the heat transmission of polyethylene and Teflon is so poor that the temperature of the bath could not be realized in the etch solution, and even after long periods of equilibration the maximum initial etch temperature was only 24.70°C. As a further consequence of the poor heat transmission of the etching container, the more rapid etches were run under essentially adiabatic conditions.

It was realized early in the study that in some etch compositions the rate of attack on the die would be so high that if the specimen were left in the solution for one minute, either the sample would be etched away completely or it would be so thin that it would be very difficult to handle. For this reason an attempt was made to choose an etch time for each composition such that only 4 to 6 mils of material would be removed. The etch rate was then linearly extrapolated to obtain the 1 min value. The justification for the extrapolation is illustrated by Fig. 1 where the change in sample thickness is plotted as a function of time for two different etch compositions. It is quite possible that for compositions where the etch rate is very high, the linear relation between the amount of material removed and the time is not valid for the entire minute, since the reaction is known to be strongly exothermic and to proceed under adiabatic conditions. Nevertheless, it is believed that by limiting the etch time to minimize the effect of changes in experimental conditions, one could obtain numbers by linear extrapolation which are reasonably valid. The change in concentration of the etch system during the experiment was also considered an important factor. However, if one allows only about 0.005 in. of silicon to be etched away during the operation, this would correspond to reacting only about 0.4 mequiv. of silicon, and the change of concentration of the reagents due to reaction would be at most 1%, which is a negligible consideration.

Since the etching systems under study here were composed of three components (i.e., HF, HNO3, and H2O) with a fourth component present only when HC2H3O2 was added, it was decided to represent the solution concentration field with a triangular coordinate system. The decision was also made to set two of the vertexes at 100 w/o of the concentrated HF or HNO3 acids and to make the third vertex 100% added water or glacial acetic acid. This normalization of the axes to the concentrated acids results in data good only for the particular concentrations chosen as the basis for the study. The weight percent system, though, allows for ready conversion to other concentration units such as mole fraction and molality, and knowledge of the specific gravity of the solution2 also enables one to calculate the molarity of the solution. By converting the concentrations to mole fractions, one can compare any concentration system with any other.

²The specific gravity of each solution was measured, but for brevity we will not discuss this point further.



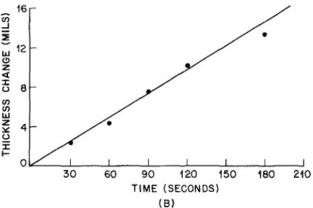


Fig. 1. Change of die thickness plotted as a function of die exposure time: (A) solution composed of 55% HF, 45% HNO₃; (B) solution composed of 30% HF, 70% HNO₃.

The solution composition field was subdivided into a number of evenly spaced points representing different compositions. The amounts of each reagent corresponding to each composition were then calculated, and the solutions were prepared by putting together the required amounts of reagents. The hydrofluoric acid was weighed out directly to \pm 0.01g, but the weights of nitric acid and water were converted to volumes and measured out in a burette to \pm 0.01 cm³.

Experimental Results

Figures 2 and 3 represent the iso-etch-rate contours determined for the normal and high concentration

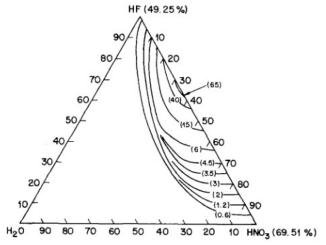


Fig. 2. Curves of constant rate of change of die thickness (mils per minute) as a function of etchant composition, in the 49% HF-70% HNO₃ system.

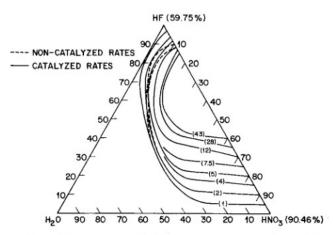


Fig. 3. Curves of constant rate of change of die thickness (mils per minute) as a function of etchant composition in the 60% HF-90% HNO3 system; the effect of added catalyst (NaNO2) is shown as the dashed lines.

reagents [to obtain the etching rate on either of the (111) faces, divide the listed value by 2], with added water as the third vertex. Figures 4 and 5 are the same data, illustrating the shape of the etch-rate surface as a function of solution composition. The various regions and the mechanisms involved in the dissolution process were adequately covered previously (1-3). We shall just note that in the region around the HF vertex the reaction is primarily oxidation-reduction controlled, and in the region around the HNO3 vertex, the reaction is primarily diffusion controlled. A consequence of the diffusion-governed kinetics is that different crystallographically oriented surfaces should all etch at the same rate. For two widely different etch compositions, the rate of attack on the (111) plane has been found to be identical to the rate of attack on the (110) plane. In preferential etch systems (e.g., NaOH), the (111) is the slowest etching plane and the (110) plane is one of the fastest etching planes.

Figures 6 and 7 are plots of the same solution composition fields, respectively, but in the latter two figures are shown the observed geometry effects on the initially rectangular parallelepipeds.3 In these two figures it is seen that the geometry of the etched

³ Although solid lines are drawn from which sharp discontinuities could be inferred, this is not the case. There are graded regions in going from one area to another, but this would make illustration well-nigh impossible.

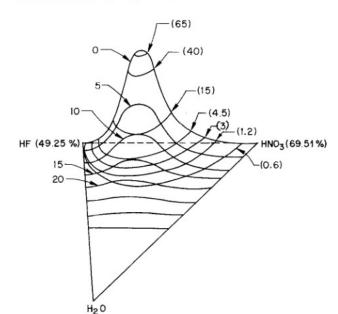


Fig. 4. Etch-rate surface as a function of etchant composition in

specimen undergoes a gradual transition as the etching formula is changed along the HNO3-HF axis. In the high HF region the specimens have rough, pitted surfaces and sharply peaked corners and edges. At about

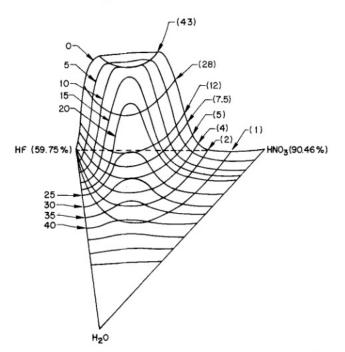


Fig. 5. Etch-rate surface as a function of etchant composition in the 60% HF-90% HNO3 system.

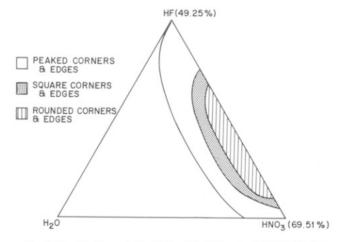


Fig. 6. Resultant geometry of the etched die as a function of the etchant composition in the 48% HF-70% HNO₃ system.

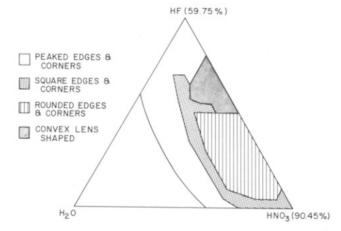


Fig. 7. Resultant geometry of the etched die as a function of the etchant composition in the 60% HF-90% HNO3 system.

the 48% HF-70% HNO₃ system. 158 130 63 39 address Redistribution license or convright: see ecso 25% HNO₃ on this axis (Fig. 6), the geometry becomes square and the surface has a matte appearance. Further increase of HNO₃ above 35% is accompanied by progressive rounding of the corners and the development of a highly specular surface. The latter geometry and appearance prevail over the remainder of the composition range except that in the very high HNO₃ region the geometry again becomes peaked. A tendency toward peaking of corners is also observed as the composition is varied toward increasing water.

The natural tendency of a diffusion-controlled reaction is the development of rounded corners and edges, since the availability of reagent is greater at the corners and edges of the specimen than in the center. A factor must therefore exist that enhances the rate of attack of the reagent in the central portion of the specimen. This factor is the effect of catalysis on the reaction rate. We know that in some etching solutions, a work-damaged surface will etch readily while a preetched surface will be very difficult to etch. In the latter case, it has been observed that if the etching action can be started by any one of a number of means (e.g., seeding the solution with a die having a damaged surface), there is then no difficulty in sustaining the reaction. This phenomenon can be explained on the basis that one of the reaction products catalyzes the reaction. To verify this hypothesis, certain solutions were seeded with NaNO2 in order to generate nitrogen oxides (1). The results are shown in Fig. 3; the enhancement of the etching rate by the addition of NaNO2 in the high HF region and in the high H₂O region is evidence of catalysis. The reaction is thought to be autocatalytic throughout the entire composition range, but the effect of added catalyst is negligible in those regions where the rate is sufficiently high so that any generated catalyst would far exceed the added amount. Since the reaction is autocatalytic, swirling dissipates the catalyst more rapidly from the corners and edges than from the central portion of the specimen. This might explain the high degree of attack in the central portion of the die relative to the attack on the edges and corners.

The rough surfaces of specimens etched in the high HF region can also be correlated with the autocatalytic mechanism. When the HNO3 concentration is low, the reaction can begin only at low activation energy sites. such as crystal imperfections. From these sites, the reaction spreads to adjacent areas as the catalyst diffuses. If the reaction is slow, the catalyst is liberated slowly and there is a high probability that it will be dissipated before it can act to propagate the reaction. When the HNO3 concentration is very low, the surface can be seen to be covered by isolated pits and craters. As the HNO3 concentration increases, the density of pits and craters increases until they finally merge into a rough "orange peel" pattern. When the concentration of HNO3 becomes sufficiently large, the entire surface becomes susceptible to attack, and as HF becomes the rate-limiting reagent, the surface becomes smooth and

The ratio of the attack rate on the two (111) planes to that on the four unoriented surfaces is greater than unity in the high HF region, and much less than unity in the high HNO3 region. This can also be correlated in the high HF region with the greater dependence of the rate on catalyst, and in the high HNO₃ region with the diffusion mechanism. Because of the greater availability of reagent at the edges, this region should show an enhanced etch rate with respect to the center. The fact that the edges etch more slowly than the center in the high HF region indicates that the dissipation of the catalyst is the predominating factor governing the behavior of etchants rich in HF. In the region of etch composition where the oxidation step ceases to be rate determining, the etch rate on the edges of the specimen is in accord with the expectation that these areas will etch faster.

Another observation lending itself to interpretation in terms of the autocatalytic nature of the reaction is the peculiar dependence of the etching rates on the water concentration. In the lower region of Fig. 2 and 3, the contour lines run parallel with lines of constant HF for a considerable distance and then suddenly break upward. The rates at constant HF are independent of increasing water or decreasing HNO3 up to a critical value. Around this critical point, the rates vary sharply with composition. It is reasonable to assume that as the concentration of nitric acid decreases, a point is reached where the rates become dependent on the nitric acid concentration, and thus on the coupling of the generated catalyst with the reacting system. The decrease in rate resulting from a decrease in the nitric acid concentration is amplified by the decreased coupling of the catalyst with the surface, so that as a result of this "positive feedback" effect the rates are strongly affected by relatively small changes in etch composition.

In the 60% HF-90% HNO₃ system there is a composition region (see Fig. 5 and 7) where the rate reaches a plateau value and the geometry is extremely rounded. This is a region of extremely rapid reaction rates, and the mechanism in this region is not very well understood; however, it is thought to involve the formation of a vapor barrier at the silicon surface owing to the high generation rate of SiF₄.

Figures 8 and 9 are, respectively, the equivalent kinetic and geometry data for the acetic acid diluent system. A comparison of Fig. 2 and 8 clearly brings out the relationship between the two etching systems. The rates in the two systems obviously must converge to identical values along the HF-HNO₃ axis. In the

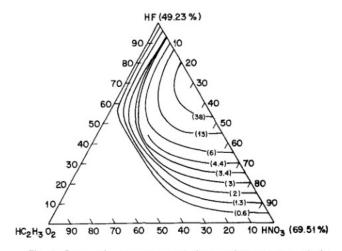


Fig. 8. Curves of constant rate of change of die thickness (mils per minute) as a function of etchant composition in the 48% HF-70% HNO₃-HC₂H₃O₂ system.

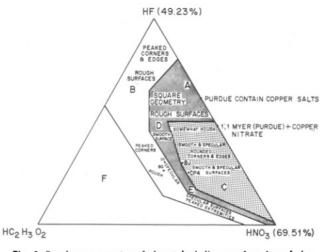


Fig. 9. Resultant geometry of the etched die as a function of the etchant composition in the 48% HF-70% HC₂H₃O₂ system.

lower portion of each plot, the constant etch-rate contours of both systems run identical courses over a considerable range of compositions. In the upper portion of each plot, however, the contours diverge immediately upon the addition of diluent. The noteworthy feature here is that the contours for the acetic acid system run parallel with lines of constant nitric acid concentration and the plot in general bears a striking resemblance to the high concentration acid system shown in Fig. 5. The equivalent etching systems HF + HNO $_3$ + diluent are seen to show a much greater tolerance for acetic acid than for water as the diluent.

Earlier in this section it was shown that in the high HF region the etch rates are affected by the oxidizing ability of the nitric acid and the reaction is critical with respect to the coupling of the generated catalyst. The tolerance of the system for acetic acid is thus in accord with the known enhanced oxidizing power of nitric acid in acetic acid over that of nitric acid in water (7,8). The addition of acetic acid does not reduce the oxidizing power of the nitric acid until a fairly large amount of the diluent has been added. Therefore, the rate contours remain parallel with lines of constant nitric acid over a considerable range of added diluent.

The same explanation is valid in the high HNO₃ region. The sudden dependence of the etch rates on the amount of added diluent has been associated with a transition from an oxidation-independent mechanism to an oxidation-dependent mechanism. The effect of the acetic acid is primarily to defer the onset of the oxidation-dependent mechanism until a very large amount of this diluent has been added.

The striking similarity of the water diluted etching system and the acetic acid diluted system, coupled with the consistency of the etch rates irrespective of the nature of the diluent, provides an additional argument in favor of the diffusion-governed mechanism previously proposed. The fact that the etch rates are not affected by the enhanced reactivity of the nitric acid in acetic acid can only be explained on the basis that the rates are diffusion limited, not reactivity governed.

The acetic acid diluent etch composition plane may also be subdivided into general areas having different geometric etching characteristics (see Fig. 9). The principal difference between the two systems is that the areas in the acetic acid system are more extensive.

In Fig. 9, we have indicated several commonly used etches. The BJ etch and the CP4 etch are seen to be located in area C, and they should have similar properties. On the basis of our interpretation of the reaction mechanism in this area, we would venture that the bromine in the CP4 formula serves no useful function in the etching of silicon to obtain a specific geometry. The same comment may be made with respect to the copper in the Purdue etches. There may be factors, though, which we have not evaluated, such as resulting electrical properties of devices etched with these solutions, that might show a need for the additive. From our point of view, however, the use of an additive in such compositions is pointless.

The 111 etch is interesting because it lies in a region where the resultant surfaces are specular, but where the corners tend to peak. It should produce better geometry retention than the BJ etch and it should be more reliable in the absence of added catalyst than its neighbor, the BQ etch. The latter is very critical, with respect to slight changes in composition, and with respect to the silicon surface. Infinitesimal composition changes such as the addition of one drop of acetic acid to 20 cm3 of etchant, or the use of 48.5% HF instead of 48.9% HF to prepare the solution, have been observed to cause the second etch rate (i.e., etching of a work-damage-free surface) to vanish. It has been observed that the etch rate in the BQ composition may have practically any value between zero and the diffusion-limited value for that composition. However, these disadvantages may possibly be overcome by the introduction of a catalyst, and the BQ etch might prove to be quite useful in device work.

In summary, the behavior of the water diluted and acetic acid diluted etches based on HF and HNO3 is qualitatively very similar and is governed by the same considerations. The differences are primarily quantitative and reflect the greater tolerance of the system for acetic acid than for water. This does not mean that the acetic acid diluted system is less critical than the water diluted system, with respect to composition, but merely that a greater amount of diluent may be added before the system becomes critical with respect to diluent or composition. Except in the high HF region, both systems are initially indifferent to the addition of diluent, and the etching behavior is unaltered until the composition is reached where the direction of the contours begins to change. Then feedback mechanisms come into play and the behavior of the solution becomes erratic. It is significant that the etching characteristics reflect the critical dependence of the etch rate on the local catalyst concentration in this region, so that a favorable geometry is obtained at the expense of erratic or unpredictable behavior of the etchant. Any attempt to make the etchant noncritical, such as an increase in temperature or the excessive addition of catalyst, will simultaneously produce a change in the geometry of the specimen to that characteristic of a diffusion-limited mechanism.

In the high HF region, the solution is critical with respect to the addition of water, because the critical amount of diluent is already contained in the concentrated stock HF. The use of acetic acid in this region may serve some purpose in making the area less critical, and some useful sizing etches may be found here. It is surprising that the common etches containing acetic acid, such as CP4, were not formulated in this region, but in the region where the addition of acetic acid seems to serve no useful purpose. (The amount of acetic acid in CP4 is a little less than critical.)

It is often said that acetic acid moderates the reaction and slows the rate. This is not true in the high nitric acid region and in the high hydrofluoric acid region. However, in the vicinity of the maximum etch rate, the acetic acid does have the alleged effect.

It seems on the basis of the small amount of information we have presented that any etching behavior attainable with an acetic acid diluted etch can be duplicated with a water diluted etch. The choice of diluent is thus a matter of individual preference.

Applications

A process specification for the fabrication of a semiconductor device usually includes one or more chemical etching steps. The etching operation is employed for the attainment of certain objectives such as the reduction of reverse current, machining to size, or the removal of unwanted material. The diversity of circumstances under which the etching is performed, coupled with the infinite variability of the etching formulas, has created a fertile field for the sorcerer and alchemist to practice their arts. It was inevitable that some excellent etching procedures should have been discovered in the course of time. However, these procedures were applicable only to the attainment of limited objectives with specific devices. Every situation became a separate problem which had to be coped with on a hit-or-miss basis, without the guidance of scientific principles. In such a state of affairs, a method for the selection of the optimum etch to accomplish a given objective appeared to be extremely valuable.

A key to the selection of an etch is an understanding of the etching characteristics of a solution as a function of the composition. The physical factors that influence the etching behavior have aready been discussed. It may now be desirable to recapitulate some of the concepts developed above and show how they can be applied to the formulation of an etching solution having certain specific properties.

For convenience, the solution composition plane in Fig. 6 will be divided into six regions (see Fig. 10). The characteristics of each region will be discussed in relation to the basic physical processes that play important roles in these regions. When this foundation has been established, a few basic rules of thumb will be given for formulating etching solutions that can be applied to specific practical problems.

Let us first focus attention on area A in Fig. 10. As Fig. 6 shows, the rates in this region are very rapid. Control of the amount of etching in this area is very difficult, because of the extreme sensitivity of the rate to small fluctuations in composition. There are also very few desirable characteristics in this region to make etching in it worthwhile. For these reasons, no etching compositions in area A will be considered further.

Area B covers the largest range of compositions. The boundary between B and F is located where the etch rates become too low to be generally practical for device fabrication. The compositions in area F have not been investigated, and no further reference to this area will be made.

The behavior of region B is best understood if one considers the compositions in the upper portion of the field. It is also convenient here to refer to the contours represented in Fig. 3 because the greater range of compositions enables observation of the behavior of the system which can only be estimated by extrapolation from Fig. 2. On the other hand, the latter system is the most useful from the practical stand-point of application because of the ready availability of the materials, the decreased hazard to personnel, and the greater ease of handling, all of which are obtained at no sacrifice in versatility whatsoever.

In the high HF region, nitric acid and catalyst are both necessary for the reaction to proceed at a perceptible rate. In the absence of added catalyst, therefore, it is necessary to find a "spark" to kindle the reaction. Such a spark exists in the presence of crystal imperfections and high energy sites. Once these undergo attack, catalyst is generated and the reaction spreads to adjacent sites. The propagation of such a reaction is clearly dependent upon the experimental conditions. Vigorous stirring may dissipate the catalyst so rapidly that the reaction is quenched. Further reaction is then dependent upon the existence of additional sites at which it can be kindled anew. For this reason, a previously etched surface, from which all work damage can be presumed to have been removed, will not be attacked by certain etch compositions in this region. On the other hand, in the absence of stirring the reaction will proceed rapidly, as the

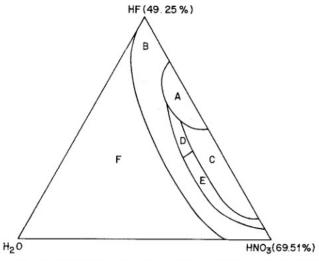


Fig. 10. Subdivision of the composition plane into general areas on the basis of the effect of the changing composition on the etching characteristics in the 48% HF-70% HNO₃-HC₂H₃O₂ system

catalyst will be able to "couple" with the reacting system, thereby increasing the rate to the point where the limiting process of diffusion takes over. Once a reasonable etching rate has been established, the presence of the required catalyst becomes assured and the reaction will, from then on, proceed even under considerable agitation. Also, if the nitric acid concentration is increased, a greater number of reactive sites will be simultaneously attacked, so that the probability of coupling of catalyst with the reacting system is enhanced. When sufficient nitric acid is present to sustain a reasonable rate, the coupling of catalyst with the reaction becomes assured and the etchant will reetch a specimen even in the absence of added catalyst.

The reaction of nitric acid with silicon is, as described previously, an oxidation-reduction process. The essence of such a process is the transfer of electrons from the reducing agent to the oxidizing agent. It is now apparent that factors influencing the availability of electrons will affect the rate. The rates will therefore be subject to modification by the effect of galvanic couples or pn junctions. Etching compositions have been found that will not attack either p- or n-material alone but will, if a junction is present, etch the p-region preferentially.

In region C of Fig. 10, the rates are independent of the nitric acid concentration, and electron transfer processes have a minor effect on the geometry. One might say that the entire surface is covered with a sheath of SiO₂. The more reactive centers may begin to etch first, but they soon become covered with an oxide layer. The oxide protects the reactive center from excessive oxidation while the remainder of the surface "catches up." The net effect is that the hydrofluoric acid, which is the rate-limiting factor, acts on a uniform layer of oxide. The result is a smooth, specular surface with rounded corners and edges. The rounding of the geometry is also affected by the rate of the reaction. When the reaction is fast, the diffusiongoverned mechanism produces considerable rounding. By decreasing the hydrofluoric acid concentration, it is possible to decrease the rate to such an extent that catalyst coupling again becomes important in spite of the high concentration of HNO3. As a result, the geometry is observed to become square or slightly peaked in the extremely high HNO₃ region.

The etching characteristics of region C in Fig. 10 may be summarized as nonselective, producing specular surfaces and rounded geometry. The surface quality is mediocre at the upper end of the region, but it improves considerably a short distance from region A. Any surface detail such as wavy lines can probably be attributed to a nonhomogeneous surface environment such as concentration or thermal differences, or perhaps major crystal defects.

Regions D and E of Fig. 10 correspond to compositions in Fig. 2 where the direction of the contour lines is changing. This indicates a transition of mechanism from an HF-limited process to an HNO₃-limited process. The geometry of specimens etched in regions D and E is square as the result of the transition to a catalyst coupling mechanism. The difference in the etching characteristics between regions D and E lies in the surface quality of the etched die. In area D the reaction is relatively more HNO₃ dependent, and vestigial evidence of selective etching is apparent. The result is a somewhat rough surface. In area E, the reaction is more HF dependent, and highly lustrous, specular surfaces result.

Let us now consider a few practical etching problems. As the first example, let us look at the problem of a suitable sizing etch for thinning a silicon chip. It is desired to remove a rather large amount of material, control the resultant thickness within narrow tolerances, retain square edges and corners, and preferably end up with a smooth surface for the subsequent operation, which could be an evaporation step. Area B is not ideal for such an etch because of the catalyst coupling factor; the rates in this region are too erratic, and they are also much too sensitive to small composition variations. The surfaces are also a bit too rough, although the geometry is favorable. In region C, the rates vary slowly with composition and the resulting surfaces are good, but the geometry is unfavorable. Possible solutions to the problem lie in the lower portion of area C, where the geometry turns square, or preferably in area E. If a large amount of material must be removed, then a two-stage process is possible where the etch may first be performed in a region of B where the rates are moderate but catalyst coupling is assured, and then finished up with a composition from regions C or E.

For the second example, let it be desired to remove the p-region from a large-area pn junction and to have the n-region relatively unaffected. A solution lies in choosing a "selective" etch, preferably near region F. The etch rate should be so low that the etching is essentially quenched by dissipation of the catalyst. Under these conditions, the attack will stop at the junction provided sufficient agitation is used, and provided that the volume of etch is sufficiently great that a negligible buildup of catalyst concentration occurs.

For the third example, consider the configuration shown in Fig. 11A. The problem is to remove a p-layer adjacent to an aluminum layer in order to contact the base n-type layer. One of the etches normally used gives a configuration such as is shown in Fig. 11B. Severe local etching in the vicinity of the aluminum plateau caused severance of the base area before any substantial portion of the p-region had been etched. The problem here was that a very selective etch was being influenced by the galvanic effects or catalytic effects of the aluminum layer to produce undesirable results. An answer to this problem lay in using a non-selective etch from area E of Fig. 10. An etch composed of 1HF:8HNO₃:1H₂O (by weight) was tried, and it gave the configuration illustrated in Fig. 11C.

There is a striking similarity between the problems of the second and third examples. In both instances it was desired to remove a p-layer and stop at a junction. In the former instance a selective etch did the job best because it had the built-in feature of stopping at the right time without danger of overetching. An alternative solution would be to use a nonselective etch, but this would require careful control of the time, temperature, and other factors governing the amount of etch, and also knowledge of the thickness of the p-layer. The presence of the aluminum and a second junction complicated the picture. The selective etching in the vicinity of the aluminum may have been caused by the trapping of generated catalyst first in the corner and later in the cavity. Under these circumstances it became necessary to sacrifice the automatic cutoff feature and use the alternative solution.

In the next example, consider the problem of bringing out the structural detail of a crystal by etching. The solution here also lies in a selective etch from region B or region F. A very slow etch must be chosen in order that catalyst coupling may be held to a minimum and mass action rather than diffusion may determine the mechanism, otherwise smudging of the structural detail might occur. The etch composition should therefore be in region F. Another solution might lie in selecting a nonautocatalytic etching system, in which diffusion is relatively unimportant, such as NaOH-H₂O.

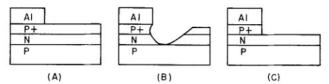


Fig. 11. The influence of solution composition on preferential etching: (A) initial configuration, (B) configuration after etching with 1HF:1HNO₃:2HAc, (C) configuration after etching with 1HF:8HNO₃:1H₂O.

For a final example, examine the problem of junction etching for the improvement of the electrical properties and stability of a device. In the high HF region, etching will be localized at the junction, although all areas will be attacked to some extent. The tendency to attack attached leads made of copper, nickel, silver, etc. would be minimized because of the weakly oxidizing nature of the solution. The problem of lead attack is not merely one of weakening the structures, but also one of contamination of the etch by the reaction products. The localized etching in the desired portion of the device may allow for shorter time of exposure of the device to the etch. However, the roughness of the surface may provide a greater surface area for leakage and sharp points from which arcing may take place. Deep erosion of the junction also provides an overhanging roof to the device that is subject to mechanical damage and that constitutes a potential site for arcing. The possibility of severe local etching in the presence of metallic leads, aluminum layers, or crevices is also to be guarded against.

In the high HNO₃ region, the etched surfaces will be smooth. The junction will be visible as a fine line, but no major discontinuity in the surface will be noted. The tendency of the device to etch all over will be partly offset by the factor that the edges will etch faster than the (111) surfaces. The attack on metallic attachments will be somewhat greater than in the high HF region, but the chances of undercutting and local etching will be obviated. The minimized surface area may somewhat reduce surface leakage and the surface smoothness may possibly be conducive to better high voltage stability.

It is unfortunate that our knowledge of the surface, and of the factors contributing to high surface leakage and low breakdown voltage, is too meager to permit a statement of the problem in terms of etching objectives. In this instance, the method was to examine the etching characteristics of each region and make an educated guess as to which approach is the most promising. The answer will then probably be as valid as the premise that etching, and not some other unknown factor, holds the key to better electrical properties.

Acknowledgments

The authors wish to express their gratitude to the many former colleagues at the Hughes Aircraft Company who contributed their helpful comments, suggestions and in some cases direct assistance. S. Pryor, P. Walker, F. Ludwig, and H. Winston are only a few of those who helped. Finally, we would like to thank the management of Bell Laboratories for allowing the time to prepare this paper for public dissemination.

Manuscript submitted May 27, 1976; revised manuscript received Aug. 18, 1976.

Any discussion of this paper will appear in a Discussion Section to be published in the June 1977 JOURNAL. All discussions for the June 1977 Discussion Section should be submitted by Feb. 1, 1977.

Publication costs of this article were assisted by Bell Laboratories.

REFERENCES

- H. Robbins and B. Schwartz, This Journal, 106, 505 (1959).
- H. Robbins and B. Schwartz, *ibid.*, **107**, 108 (1960).
 B. Schwartz and H. Robbins, *ibid.*, **108**, 365 (1961).
- 4. B. A. Irving, in "The Electrochemistry of Semiconductors," P. J. Holmes, Editor, p. 256, Academic Press New York (1962)
- Press, New York (1962).

 5. P. J. Holmes, in "The Electrochemistry of Semiconductors," P. J. Holmes, Editor, p. 329, Academic Press, New York (1962).

 6. R. M. Finne and D. L. Klein, This Journal, 114,
- 965 (1967).
- E. Briner, Helv. Chim. Acta, 18, 363 (1935).
 E. Briner and P. Bolle, ibid., 18, 368 (1935).