Bulk Micromachining of Silicon

GREGORY T. A. KOVACS, MEMBER, IEEE, NADIM I. MALUF, MEMBER, IEEE,
AND KURT E. PETERSEN, FELLOW, IEEE

Invited Paper

Bulk silicon etching techniques, used to selectively remove silicon from substrates, have been broadly applied in the fabrication of micromachined sensors, actuators, and structures. Despite the more recent emergence of higher resolution, surface-micromachining approaches, the majority of currently shipping silicon sensors are made using bulk etching. Particularly in light of newly introduced dry etching methods compatible with complementary metal-oxide-semiconductors, it is unlikely that bulk micromachining will decrease in popularity in the near future. The available etching methods fall into three categories in terms of the state of the etchant: wet, vapor, and plasma. For each category, the available processes are reviewed and compared in terms ofetch results, cost, complexity, process compatibility, and a number of other factors. In addition, several example micromachined structures are presented.

Keywords—Bulk, etching, micromachining, silicon.

I. INTRODUCTION

The purpose of bulk micromachining is to selectively remove significant amounts of silicon from a substrate. This is sometimes done to “undercut” structures that are required to physically move; to form membranes on one side of a wafer; or to make a variety of trenches, holes, or other structures. A sampling of the wide variety of possible bulk-micromachined structures is illustrated in Fig. 1. The etching approaches used, while seemingly requiring quite aggressive chemistries, can be compatible with on-chip circuitry and even micrometer-scale mechanisms if the overall process flow is designed appropriately. This fact has allowed bulk micromachining to be combined with complementary metal–oxide–semiconductor (CMOS) circuitry to fabricate devices that take advantage of the unique properties of single-crystal silicon and the relatively large structures (both physically and in terms of mass) that can be fabricated from it.

The geometries of etched features lie along a continuum between fully isotropic (rounded, due to equal etch rates in all directions) to anisotropic (typically exhibiting perfectly flat surfaces and well-defined, sharp angles). This is illustrated in Fig. 1(a) and (b). These properties are defined by the nature of the chemical reactions, the diffusion of reactants and products, and a number of other factors, including the shapes of the masks used to define the etched regions.

Another key distinguishing feature of etchants is the phase of the reactants: liquid (or “wet” etchants, almost exclusively relying on aqueous chemistries), vapor, and plasma (the latter two being referred to as “dry” etchants). As might be expected, the reaction mechanisms, reaction rates, chemistries, and diffusion properties of these three modes are quite different, as are the associated equipment costs.

The etching reactions rely on the oxidation of silicon to form compounds that can be physically removed from the substrate. In aqueous chemistries, this tends to be accomplished using highly reactive species, such as acids and bases. Etching anisotropy exhibited by such reactions is due to differing chemical reactivities of certain crystal planes of the silicon. Most liquid-phase etches can be modulated by added dopants in the silicon as well as electrochemical biasing. In vapor phase, the reactions rely on the adsorption of halogen molecules or compounds, their subsequent dissociation into reactive halogen species, and the formation of volatile silicon compounds. These etches tend to be fully isotropic, diffusion driven, and with no preference for particular crystal planes. Both liquid- and vapor-phase etches can also be locally driven by the addition of external energy by optical means, such as a scanned laser beam. In the plasma phase, highly reactive halogen free radicals can be created, which react with exposed silicon to again form volatile silicon compounds, with the reactions often enhanced by directional bombardment by energetic ions arriving along paths perpendicular to the silicon surface. Secondary reactions can protect silicon surfaces perpendicular to the ion bombardment (i.e., the sidewalls of regions being etched). This results in very
high degrees of anisotropy for some plasma-based etching systems. A unique feature of these etches is the ability to control the degree of anisotropy during the etch process via the plasma chemistry.

As mentioned above, there are numerous other important etchant properties, such as etch rate, the ability to modulate etching with dopants or electrical bias, surface roughness, the availability of suitable masking films, health hazards, and the ease in which the etch chemistry can be reversed via a post-etch bake in oxygen. These factors and their interactions make the selection of the most appropriate etch chemical and process a crucial step in the success of bulk micromachining.

II. WET ETCHING

A. Isotropic Wet Etching

The most common isotropic wet silicon etch is “HNA,” a mixture of hydrofluoric acid (HF), nitric acid (HNO₃), and acetic acid (CH₃COOH) [1]–[5]. The HNO₃ drives the oxidation of the silicon, while fluoride ions from HF then form the soluble silicon compound H₂SiF₆. The acetic acid, which is much less polar than water (smaller dielectric constant in the liquid state), helps prevent the dissociation of HNO₃ into NO₂⁻ or NO₃⁻, thereby allowing the formation of the species directly responsible for the oxidation of silicon

\[ N₂O₄ \rightarrow 2NO₂⁻. \]

Despite this, the same mixture without the acetic acid is found to be nearly as effective for short etch times, until the NO₂ is depleted. The etching chemistry is complex (due to HF's instability), and the reaction order with respect to the components is unknown.
chemical mixture and silicon doping. The overall reaction is [6]

$$18\text{HF} + 4\text{HNO}_3 + 3\text{Si} \rightarrow 3\text{H}_2\text{SiF}_6 + 4\text{NO}_2(g) + 8\text{H}_2\text{O}.$$ 

A useful formulation for HNA is 250 ml HF, 500 ml HNO$_3$, and 800 ml CH$_3$COOH [7]. When used at room temperature, one obtains an etch rate of ≈4–20 μm/min, increasing with agitation (the agitation sensitivity of these etchants can sometimes make their repeatable use problematic). The etch can be masked with silicon nitride or silicon dioxide (however, this latter film is attacked fairly quickly, at 30–70 nm/min). It is also noteworthy that the HNA etch is slowed down in rate ≈150 times by regions of light doping (<10$^7$ cm$^{-3}$ n- or p-type) relative to more

As discussed by Petersen [7], the general mechanism of wet single-crystal silicon etchants is:

1) injection of holes into the Si to form Si$^{2+}$ or Si$^+$;
2) attachment of OH$^-$ groups to the Si$^{2+}$ to form Si(OH)$_2^{2+}$; 
3) reaction of the “hydrated” Si (silica) with a complexing agent in the solution; 
4) dissolution of the reaction products into the solution.

Thus, for such an etch, one needs a source of holes, OH$^-$, and a complexing agent. Since the etching is basically a charge-transfer-driven process, it makes sense that the dopant type/concentration and externally applied electrical
B. Anisotropic Wet Etching

As mentioned above, anisotropic (or "orientation dependent") etchants etch much faster in one direction than in another, exposing the slowest etching crystal planes over time. Those described here slow down markedly at the (111) planes of silicon, relative to their etch rates for other planes. Depending upon the crystal orientation chosen, the (111) planes are generally either at 54.74° to the wafer’s surface (for 100) silicon) or perpendicular to it (for 110) silicon), as illustrated in Fig. 1(b). Examples of classic papers describing these crystal plane dependent etches are Bean [8] and Bassous [9]. Most such etchants can be dopant or electrochemically modulated but slow down at the (111) planes regardless of the dopant(s). It is important to note that from the top view, etching at “concave” corners on (100) silicon stops at (111) intersections, but “convex” corners are undercut, allowing cantilevers rapidly undercut and released. This is illustrated in Fig. 1(c). A considerable amount of effort is often expended in designing mask patterns to take these effects into account [10]–[13].

While the reaction mechanisms have generally been elucidated, the mechanisms of dopant modulation and anisotropic etching along crystal planes have not been fully explained. The phenomena of propagating cracks and anisotropically etched features stopping on certain crystal planes are commonly thought to result from the plane with the “least surface density” of atoms, but this theory cannot account for all of the behavior seen. As an example, for cubic crystals (zincblende and diamond structures), the surface density of atoms does not vary by more than a few percent over all possible directions, and this cannot possibly account for the >100:1 anisotropies and cleavage preferences seen in practice. Another factor influencing this anisotropy is likely to be "screening" of the surface by attached H₂O molecules, which will be determined by crystal orientation. For potassium hydroxide (KOH) etchants (discussed below), values of the relative etch rates for the three planes of interest can be as high as (111) (reference) = 1, (100) = 400 [7], and (110) = 500 [8]. These values are extremely dependent on the chemical composition, concentration, and temperature of the etchant solutions used. In practice, relative values are lower (as much as an order of magnitude in some cases), but nonetheless very large, and account for the tremendous anisotropy seen.

Unfortunately, there are not yet any “master equations” that predict etching performance from user-controllable factors such as temperature, etchant concentration, etc. However, semiempirical methods do exist, and an example of such an etch-rate equation can be found in Seidel [14]. In practice, one typically needs to obtain etch rates and characteristics from the literature and by experiment. Also, many of the reported etch rates and other properties are time and usage dependent (i.e., only achieved for “fresh” etchants). Once the etch parameters and etch mixture lifetimes are well defined, most etchants give extremely reproducible results. Williams and Muller [6] is particularly useful since the entire study was carried out using a consistent set of materials.

1) Alkali Hydroxide Etchants: The hydroxides of alkali metals (i.e., KOH, NaOH, CsOH, RbOH, etc.) can be used as crystal orientation-dependent etchants of silicon. The chemistry is presently still under some debate. The reaction sequence appears to be the following [14], [15]. Silicon atoms at the surface react with hydroxyl ions. The silicon is oxidized, and four electrons are injected from each silicon atom into the conduction band

$$\text{Si} + 2\text{OH}^- \rightarrow \text{Si(OH)}_2^{2-} + 4e^-.$$  

Simultaneously, water is reduced, leading to the evolution of hydrogen

$$4\text{H}_2\text{O} + 4e^- \rightarrow 4\text{OH}^- + 2\text{H}_2.$$  

The complexed silicon, Si(OH)₂⁺, further reacts with hydroxyl ions to form a soluble silicon complex and water

$$\text{Si(OH)}_2^{2-} + 4\text{OH}^- \rightarrow \text{SiO}_2\text{(OH)}_2^- + 2\text{H}_2\text{O}.$$  

Thus, the overall reaction is

$$\text{Si} + 2\text{OH}^- + 2\text{H}_2\text{O} \rightarrow \text{Si(OH)}_2^{2-} + 2\text{H}_2.$$  

For KOH, Seidel [14] demonstrated that at 72°C, the etch rate was maximized at ≈ 0.9 μm/min for a 15 wt% KOH solution. In general, concentrations below 20 wt% are not used due to high surface roughness and the formation of potential insoluble precipitates. A more typical concentration of KOH is in the range of 40–50 wt% (for example, Williams and Muller [6] used 50 wt% KOH at 80°C for a reported (100) etch rate of 1.4 μm/min). A thorough overview of alkaline etchants, their properties, and their mechanisms can be found in Seidel et al. [15]. Price [16] showed that the addition of sufficient isopropyl alcohol (IPA), a less polar diluent, to saturate the solution greatly increases the selectivity for (111) versus (100) planes. Even without IPA, all of these alkali hydroxide etchants exhibit extremely high selectivity, etching the (111) plane up to 400 times more slowly than the (100) plane. In addition, they can all be dopant modulated [17]. The etch rate can be slowed down drastically in regions doped with boron to a concentration of ≥ 2 x 10¹⁸ cm⁻³. Apparently, the mechanism of this reduction is that in the heavily doped regions, the width of the space-charge-region layer at the silicon surface shrinks dramatically, leading to the rapid recombination of electrons generated by oxidation reactions (rather than their confinement to the surface) [17].

The reduction in the availability of these electrons limits the reduction of water to form OH⁻ ions that are necessary for the etch reaction to proceed.

A variety of thin-film materials can be used to mask alkali hydroxide etches, and silicon nitride and silicon dioxide are commonly used. Silicon nitride etch rates can be extremely low (for example, Seidel et al. [5] reported no measurable etch rate for chemical vapor deposited (CVD) nitride).
the order of 1–10 nm/min) and were shown to increase with increasing temperature and pH by Seidel et al. [5], who provided extensive tabulated data for thermally grown films. As for any etch mask films, the mode of their preparation (e.g., in situ growth, CVD, plasma-enhanced (PE) CVD, etc.) and their resultant chemical compositions and densities can have marked effects on their etch rates.

It is important to note that one can also use ammonium hydroxide or the so-called quaternary ammonium compounds, which contain no alkali ions (these ions, particularly sodium, can be extremely detrimental to MOS transistors, which may be present on fully integrated transducers). These etchants are discussed separately below.

For most micromachining and active circuit processing (e.g., MOS devices), (100) orientation material is used, for which hydroxide etches produce pyramidal pits with 54.7° (111) side-wall angles relative to the (100) surface. It is possible to obtain very low etched-surface roughness for a “mirror-like” finish (this is more difficult with the other anisotropic etchants such as ethylenediamine-pyrocatechol (EDP) and tetramethyl ammonium hydroxide (TMAH), which are described below). In general, KOH etching produces smoother surfaces at low and high molarities, with the maximum roughness occurring at 5–6 M (28–34 wt%), decreased dramatically with stirring (presumably displacing hydrogen bubbles), and increased temperature [18]. Hillock formation can also be suppressed by the addition of a suitable oxidizing agent (e.g., ferricyanide ions, Fe(CN)$_6^{3-}$), as described by Bressers et al. [19]. They reported a drastic reduction in hillock formation with the use of 18 mM K$_3$Fe(CN)$_6$ as an additive to 4-M KOH solutions, used at 70°C.

 Using (110) silicon, one can obtain “perfectly” rectangular trenches over considerable distances because the etch rate is so high in the (110) direction relative to the other two planes [8]. Unfortunately, however, secondary (111) planes appear across the ends of the channels, but in some designs, this is not a problem. Tuckerman and Pease [20] demonstrated the use of such trenches as liquid cooling fins for integrated circuits. One can bond such a microheat-sink directly to an ordinary silicon wafer (i.e., (100)) or other substrate on which active circuits could be fabricated.

Another useful reference on such devices is Kaminsky [21]. Further considerations of micromachining with (110) silicon can be found in Bean [8], Ammar and Rodgers [22], and Kendall and deGuehl [23].

General discussions of wet etching mechanisms for alkali etchants can be found in Seidel [14] and Seidel et al. [15], [17]. Descriptions of cesium hydroxide as an anisotropic silicon etchant can be found in Clark et al. [24] and Chambers and Wilkie [25], and rubidium hydroxide is discussed in Wang et al. [26].

2) Simple and Quaternary Ammonium Hydroxides: As mentioned above, there are hydroxide-based anisotropic etchants for silicon that do not incorporate alkali ions that can be detrimental to CMOS integrated circuits. Ammonium hydroxide (NH$_4$OH) is one such etchant that Kern [27] demonstrated the use of NH$_4$OH (0.7% in H$_2$O) to achieve 0.11 µm/min (6.6 µm/h) etch rates in (100) Si (temperature range 85–92°C). Little further work appears to have been done on this subject until Schnekenburger et al. presented their analysis of this etchant [28]. They explored a variety of concentrations from 1 to 18 wt% NH$_4$OH at a temperature of 75°C. They noted a maximum (100) silicon etch rate of 30 µm/h but extremely bad hillock formation (surface roughness). They reported that their best results were obtained at 3.7 wt% at 75°C for stirred etch baths. For this recipe, they demonstrated boron-dependent etch-rate modulation at 1.3 × 10$^{20}$ cm$^{-3}$ with a selectivity of 1 : 8000.

Ammonia-based etchants have not been popular for a number of reasons, including their relatively slow etch rate, hillock formation problems, and rapid evaporative losses of ammonia gas (noxious) when heated. Quaternary ammonium compounds in various formulations can be used in their place, with far superior performance and still without alkali metal contamination. TMAH ((CH$_3$)$_4$NOH) is a quaternary ammonium hydroxide compound and is one of the more useful wet etchants for silicon. It is already present in most cleanrooms in “MOS-clean” grade (low sodium) since it is used in most positive photolithographic developer processes (those that do not contain choline). It is safer than EDP (discussed below), can be modified with additives so that it does not etch aluminum, begins to slow down for boron doping levels above approximately 1 × 10$^{20}$ cm$^{-3}$, and is relatively low cost.

A potentially significant tradeoff with the use of TMAH is that, as for NH$_4$OH, the surface morphology tends to be rougher than that obtained with the other common etchants, although new formulations (discussed below) appear to control this effect. In addition, the (100):(111) plane selectivity ratio tends to be much lower than for alkali hydroxides, on the order of 10–35 for TMAH in the 10–40 wt% concentration range (etching at 90°C, with corresponding etch rates of 0.5–1.5 µm/min). For typical TMAH solutions, etch-rate and surface roughness decrease as the TMAH concentration is increased. Tabata et al. [29], [30] studied this and found that at 5 wt%, the surfaces are quite rough due to longer H$_3$ bubble residence times, becoming quite smooth at approximately 20% (note that the formula given above is 10 wt%, trading off some surface roughness for lack of aluminum etching).

TMAH exhibits useful selectivity for boron etch stops—the etch rate of TMAH falls off ten times at 10$^{20}$ cm$^{-3}$ boron concentration (conditions: 22 wt% TMAH, 90°C). The etch rate decreases up to 40:1 for 2 × 10$^{20}$ cm$^{-3}$ boron concentration (note that 2.5 × 10$^{20}$ cm$^{-3}$ is the solid solubility limit for boron in silicon) were reported by Steinsland et al. in 25 wt% TMAH at 80°C [31]. The etch-stop selectivity can also be improved somewhat by the addition of isopropyl alcohol (see Merlos et al. [32]).

A very useful property of TMAH is that typical masking layers show excellent resistance to etching. For example,
of 0.05–0.25 nm/min, and silicon nitride films also offer comparable performance. Additional TMAH etch selectivity data for various dielectrics versus (100) silicon were presented by Schnakenberg et al. [33], [34], Ristic et al. [35], and Merlos et al. [32]. An important caveat for such data is that it is very difficult to compare the etch rates of different films (particularly PECVD varieties) because the actual stoichiometry (and hence etch rate) can be quite process dependent.

Silicon can be dissolved in TMAH solutions to lower the pH, provide selectivity toward aluminum metallization (but increase surface roughness), and decrease (100) etch rate. A typical TMAH formulation (a modification of the formula described in Reay et al. [36]) that provides excellent etch characteristics with minimal aluminum etching is 250 ml of 10% TMAH (as obtained from Aldrich Chemical Co., Milwaukee, WI, 25 wt%), 375 ml deionized (DI) water, and 22 g silicon (dissolved in solution). The mechanism underlying minimal attack of aluminum with lowered pH is related to chemical passivation of the aluminum through the formation of a relatively insoluble aluminosilicate in the more acidic solution. Tabata demonstrated that lowering the pH with acids (NH_4)_2CO_3 or (NH_4)HPO_4 could offer the same protection of aluminum without the difficulty of having to dissolve silicon [37]. An alternative approach to lowering the pH (same effect as dissolving silicon) is to add silicic acid directly to the TMAH solution, as demonstrated by Hoffman et al. [38]. Using a solution of 80 ml of 25 wt% TMAH with 16 g silicic acid and sufficient DI water to bring the volume to 250 ml, they reported an etch rate of 35–70 nm/min at 70°C and relatively isotropic etch results, but with little attack of exposed aluminum.

5) EDP: EDP (sometimes referred to as EPW for ethylene diamine, pyrochatechol, and water) is a classic, but hazardous, anisotropic and dopant-modulated silicon etch, as described by Finne and Klein [39], Bassous [9], and Reisman et al. [40]. It should be noted that the (100):(111) selectivity of EDP formulations is on the order of 35, lower than those of alkali hydroxide etchants, but its selectivity for heavy p-type doping is much greater. A typical EDP formulation from Petersen [7] is 750 ml ethylene diamine, 120 g pyrochatechol, and 100 ml water. When used at 115°C, the reported etch rate was 0.75 μm/min, with an (100):(111) etch-rate ratio of 35:1.

The basic chemistry of EDP etching (after Finne and Klein [39]) includes the following steps: ionization of ethylenediamine

\[ \text{NH}_2\text{(CH}_2\text{)}_2\text{NH}_3 + \text{H}_2\text{O} \rightarrow \text{NH}_2\text{(CH}_2\text{)}_2\text{NH}_3^+ + \text{OH}^- \]

oxidation-reduction (oxidation of silicon)

\[ \text{Si} + 2\text{OH}^- + 4\text{H}_2\text{O} \rightarrow \text{Si(OH)}_4^- + 2\text{H}_2 \]

and helation of hydrosilica

\[ \text{Si(OH)}_4^- + 2\text{C}_6\text{H}_4\text{H}_2\text{O}_2 \rightarrow \text{[Si(C}_6\text{H}_4\text{O}_2\text{)}_3]^- + \text{2H}_4\text{O}. \]

Heavy (>7 x 10^{19} cm^{-3}) boron doping results in a non-isotropic etch rate, with (111) planes etching slower than (100). These etchants are more expensive and offer limited (100):(111) selectivity compared to TMAH. EDP etchants can be used for various purposes, including the removal of nitride and oxide films on silicon substrates.

4) Other Etchants: Other useful, but less popular, wet silicon etchants include hydrazine and amine gallate compounds. As described by Petersen [7] and Mehregany and Senturia [42], hydrazine/water mixtures provide useful etch rates (on the order of 2 μm/min) and can be used with similar masking layers as EDP. However, the (100):(111) etch-rate ratios are lower than those for KOH or EDP. The corrosive nature and potential carcinogenicity of hydrazine are comparable to those of EDP. Amine gallate etchants, as described by Linde and Austin [43], are composed of a mixture of ethanolamine (high-boiling-point solvent), gallic acid, water, pyrazine, hydrogen peroxide, and a surfactant. These etchants achieve high etch rates (up to 2.3 μm/min on (100) Si) and stop at high boron concentrations (>3 x 10^{19} atoms/cm^2), a lower concentration than it takes to stop EDP. Amine gallates are similar to EDP in terms of etching and masking layers (gentle on SiO_2), but apparently safer. Peroxide and pyrazine can be added to increase the etch rate, but surface roughness also increases.

5) Control of Surface Roughness: Hydrogen evolved from the etch reactions can form bubbles that can in turn cause local “micromasking,” resulting in hillock formation on the etched surface. If a sufficiently high hillock density forms on the surface of the silicon, the (100) etch rate can drop a great deal. The addition of oxidizers to consume the hydrogen as it is generated has been investigated by Petersen and others [40, 41].
and Campbell et al. [44]. Klaassen et al. described the use of peroxysulfate oxidizers in TMAH to eliminate hillock formation yet preserve aluminum etch protection via added silicon or silicic acid (adding peroxysulfates actually increases etch rates on the order of 25%, presumably through elimination of hydrogen masking) [45]. The optimum formulation reported was 5 g/l ammonium peroxysulfate (NH4)2S2O8 in 5 wt% TMAH solution with 16 g/l dissolved silicon at 80°C for an etch rate of 20.8 µm/min and a working life when mixed of 6–8 h (due to dropping pH over time).

Several nonchemical approaches to mitigating the hillock effect have been tried, including the use of surfactants, the use of ultrasonic agitation, and preemptive chemical elimination of the bubbles, as discussed above. Ohwada et al. noted that their use of ultrasonic agitation essentially eliminated surface roughness in KOH etching [46].

6) Etch-Rate Modulation: As discussed above for individual etchants, highly p-doped (p++) silicon regions greatly attenuate the etch rate. Selective p++ doping, typically done using a gaseous or solid boron diffusion source with a mask (such as silicon dioxide) to select doped regions, can be used to define specific regions of the silicon that remain, while the bulk is etched away. This is illustrated in Fig. 1(e). A classic example of this type of “lost wafer” process is that used by Najafi et al. to fabricate needle-like probes for recording the electrical activity of neural cells [47]. It should be noted that a limitation of diffusing the dopant is the maximum depth practically achievable (on the order of 15 µm). Examples of boron doped p++ etch-stop structures are shown in Fig. 2(a) and (b). In addition, surface or “buried” p++ etch-stop layers can be epitaxially grown and used very successfully with EDP, TMAH, and KOH-type etchants to form membranes or to limit etch depth beneath released structures such as cantilevers.

Electrochemical modulation of etch rates is also feasible [48]. For either η- or p-type doping, there is an open circuit potential (OCP) at which there is nearly zero current flow and the silicon etches just as if it were unbiased. As the potential applied between the silicon and the solution is made more positive (anodic current flow), more holes are supplied to the surface silicon atoms, speeding up the etching of silicon. As the applied potential is made further positive, eventually the passivation potential is reached where SiO2 is formed, effectively passivating the surface and stopping etching. For etchants such as KOH, TMAH, etc., this means that one can electrically modulate the etching up and then completely off if desired. If HF/H2O is used, the etch rate can be increased without a passivation limit since SiO2 dissolves readily in HF (it should be noted that if a highly concentrated HF solution, deficient in OH− ions, is used with an anodic bias, porous silicon can be formed [49], [50].

Another approach to electrochemical etch modulation is to form a p-n junction on the surface of a p-type wafer and set things up so that etching stops at the n-type silicon floats at its OCP and etches quickly. The p-n junction is held in reverse bias (positive potential applied to the n-type silicon) so that the etch is etched away, the positively biased n-type silicon (held well above its passivation potential) is directly exposed to the solution and prevents further etching. The bias voltage is applied to the n-type silicon, and the p-type silicon is either biased near the OCP or allowed to float. By stopping etching at a well-defined junction, very accurate thickness control of membranes can be achieved [48].

Using TMAH, the junction etch-stop technique can be used to postprocess prefabricated n-well CMOS chips, as demonstrated by Reay et al. [51]. Thermally and electrically isolated single-crystal silicon islands (including active regions) can be defined for each embedded transistor. The buried oxide and frontside silicon dioxide etch rate is modulated by a p-type etch-stop layer, forming a junction with the n-well CMOS chip. The OCP is used as the etch-stop potential. At the junction, the p-type silicon is etched away, and the n-type etches at the n-type silicon floats at its OCP and etches quickly. The p-n junction is held in reverse bias (positive potential applied to the n-type silicon) so that the etch is etched away, the positively biased n-type silicon (held well above its passivation potential) is directly exposed to the solution and prevents further etching. The bias voltage is applied to the n-type silicon, and the p-type silicon is either biased near the OCP or allowed to float. By stopping etching at a well-defined junction, very accurate thickness control of membranes can be achieved [48].

Using TMAH, the junction etch-stop technique can be used to postprocess prefabricated n-well CMOS chips, as demonstrated by Reay et al. [51]. Thermally and electrically isolated single-crystal silicon islands (including active regions) can be defined for each embedded transistor. The buried oxide and frontside silicon dioxide etch rate is modulated by a p-type etch-stop layer, forming a junction with the n-well CMOS chip. The OCP is used as the etch-stop potential. At the junction, the p-type silicon is etched away, and the n-type etches at the n-type silicon floats at its OCP and etches quickly. The p-n junction is held in reverse bias (positive potential applied to the n-type silicon) so that the etch is etched away, the positively biased n-type silicon (held well above its passivation potential) is directly exposed to the solution and prevents further etching. The bias voltage is applied to the n-type silicon, and the p-type silicon is either biased near the OCP or allowed to float. By stopping etching at a well-defined junction, very accurate thickness control of membranes can be achieved [48].
Another way to optically add external energy to modulate etch rates is the use of laser beams to drive chemical reactions, referred to as laser-assisted chemical etching (LACE). The etching is carried serially, by scanning a laser beam to remove a sequence of small volumes, making these processes slow but avoiding crystal-plane dependencies. A laser-driven wet-etching process for silicon has also been described by Ade et al. [58].

III. VAPOR-PHASE ETCHING

In addition to plasma-based methods (described below), dry etching can be achieved spontaneously with suitably reactive gases/vapors. In some cases, this approach yields the desired properties of plasma/reactive ion enhanced (RIE) etching without the need for complex and expensive equipment.

Although the fact that a family of fluorine-containing compounds (noble gas fluorides and interhalogens) will readily etch silicon (with nearly infinite selectivity to masking layers such as SiO₂) has been known for more than a decade, it is only recently that these techniques have been used for micromachining.

1) Xenon Difluoride Etching: A non-plasma, isotropic dry-etch process for silicon is possible using XeF₂ and provides very high selectivity for aluminum, silicon dioxide, silicon nitride, and photoresist [59], [60]. These properties make it an extremely useful etchant for postprocessing CMOS integrated circuits [61], although the etched surfaces produced are quite rough (an example XeF₂-etched CMOS integrated circuit is shown in Fig. 4).

Used originally for exposing the undersides of MOS transistors by etching away the underlying substrate silicon (Hecht et al. [62]), the applicability of XeF₂ to micromachined sensors and actuators was demonstrated by Hoffman et al. [63]. They showed that with a simple bell-jar setup run at 1 torr, XeF₂ could be sublimed from its solid form at room temperature and that this etch has excellent selectivity with respect to CMOS process layers.

The etch reaction, as discussed in Chang et al. [64], is (approximately)

$$2\text{XeF}_2 \rightarrow \text{Xe} + \text{SiF}_4$$

where only the silicon is in a solid phase. This reaction proceeds by nondissociative adsorption of XeF₂ at the silicon surface, dissociation of fluorine, reaction to form the adsorbed SiF₄ product, and desorption of the product and residual xenon. Etch rates are generally 1–3 μm/min (as high as 40 μm/min in some instances, as reported by Chang et al. [64]). It is noteworthy that the etched surfaces have a granular structure (10 μm and smaller feature size), making this etchant unsuitable for situations where smooth surfaces are required. In addition, unless the reaction rates are controlled or modulated (typically by using pulses of XeF₂), the heat generated by this exothermic reaction may adversely affect some microstructures. An important concern is that XeF₂ reacts with water (even moisture in air) to form Xe and HF (the latter can also unintentionally

in Fig. 3(a) and (b) [52]. Similar processes were also demonstrated by Olgun et al. [53] and Schneider et al. [54].

Optical energy can also be used to modulate wet etching of silicon, and Peeters et al. [55] described one such technique using photon pumping to generate carriers across a p-n junction while etching in KOH. Lehmann and Foll [56] and Lehmann [57] used photogenerated carriers to supply the necessary holes to the bases of prefabricated pits whose sharp bases served “high field points” to “focus” the holes and result in very highly anisotropic etch results. Extremely deep, high-aspect-ratio holes (>70:1) and trenches were
In terms of masking films, Pister reported that XeF$_2$ does not appear to etch several useful materials, including: photore sist, thermal silicon dioxide, phosphosilicate glass, boron phosphosilicate glass, Al, Au, TiNi alloy, silicon nitride, and acrylic [65]. More recently, Chu et al. confirmed that there was no measurable etch rate for Al, Cr, TiN, stoichiometric low-pressure (LP)CVD silicon nitride, thermal silicon dioxide, PECVD silicon carbide, and photore sist [66]. They measured some etching of Ti and W, with selectivities to Si etching given as Si: Ti = 85:1 and Si: Mo = 6:1. It should be noted that most metals form passivated and nonvolatile fluorides on their surfaces, preventing etching.

2) Interhalogen Etch Chemistries: Köhler et al. described a method to avoid the extremely rough silicon surfaces that are formed using XeF$_2$ etching yet still retain its advantages of being a simple-to-implement, dry-etch process [67]. To carry out the silicon etching, they used a thermal silicon dioxide mask and various interhalogen gases (BrF$_3$ and ClF$_3$) with a xenon diluent (note that the interhalogens were formed from single-element feed gases).

BrF$_3$ is formed from bromine and fluorine and reacts with exposed silicon, forming SiF$_4$ and elemental bromine, which can be reused. The mixture for which optimum results were obtained was 7 mbar (5.3 torr) of bromine, 21 mbar (15.8 torr) of fluorine, and 980 mbar (735 torr) of xenon [67]. Compared to silicon etched in pure fluorine, the surface roughness was reduced from $\approx 150$ nm to $<40$ nm (reportedly indistinguishable from the unetched, polished wafer surfaces). The etch results were nearly perfectly isotropic. Wang et al. [68] measured etch rates for thin films in BrF$_3$ and reported etch rates with respect to silicon for several materials:

- Si: LPCVD silicon dioxide = 3000:1;
- Si: silicon nitride (depends upon Si concentration) = 1000:1;
- Si: hard-baked AZ4400 and AZ1518 photore sist = 1000:1;
- Si: Al, Cu, Au, and Ni = $>1000:1$.

Such interhalogen etches could be quite useful in a number of silicon etching applications, and it is likely that they can be used to postprocess CMOS integrated circuits, for example. However, the complexities and dangers of working with pure halogen gases must be given serious consideration.

3) Laser-Driven Vapor-Phase Etching: As described by Osgood et al. [69] and Ehrlich and Taos [70], it is possible to use laser beams to selectively drive chemical reactions for dry etching of semiconductors, referred to as LACE. Direct heating of the silicon speeds the reaction locally, and if the appropriate wavelength is chosen (e.g., 500 nm for Cl$_2$), free radicals can be formed by photoly sis, as illustrated in Fig. 5(a). This process can be used to make via holes, channels, and very complex structures (undercut, overhanging structures cannot be etched, however). An example is shown in Fig. 5(b).

Etch resolutions approaching 1 $\mu$m$^3$ have been achieved in some materials (Bloomstein and Ehrlich [71]), and the Cl$_2$ process is extremely selective (1000:1) for silicon over SiO$_2$, so that buried (continuous) channels can be etched under an SiO$_2$ layer (potentially useful for fluidics, etc.). However, the Cl$_2$ and SiCl$_4$ must be able to diffuse to the opening of the channels at the edge of the die/wafer, limiting channel length.

LACE is not a parallel process and not fast enough for most manufacturing applications. Nonetheless, it may have utility in specialty micromachining or for making molds, etc. [72].

IV. Plasma-Phase Etching

The most common forms of “dry” etching of bulk silicon are plasma etching and RIE etching [73], [74]. In this class of dry-etching reactions, external energy in the form of radio-frequency (RF) power drives chemical reactions (i.e., takes the place of elevated temperatures or very reactive chemicals) in low-pressure reaction chambers. Energetic ions supply the necessary energy so that reactions can be achieved at relatively low temperatures (room temperature to a few hundred degrees Celsius). The full spectrum of isotropic through anisotropic etches is available using these methods.

RF power applied to a pair of plates accelerates stray electrons, increasing their kinetic energy to levels at which they can break chemical bonds in the reactant gases upon impact, forming ions and additional electrons. With ongoing input of RF energy into the chamber, electron/molecule collisions continue to yield ions and electrons, while exposed surfaces within the chamber absorb or neutralize these species. After a number of RF cycles, a steady-state discharge is reached in which the generation and loss processes are balanced. A wide variety of chlorofluorocarbon gases
Fig. 5. Example of the use of LACE for the bulk micromachining of silicon. (a) Illustration of the process by which optical energy leads to local heating, which in turn drives the etching reaction, after Bloomstein and Ehrlich [71]. (b) SEM of an example structure showing step sizes of 10 and 30 μm etched into a silicon substrate. The overall size of the structure is 500 × 500 μm × 50 μm deep, and the etching required 8 min. (SEM courtesy R. Aucoin, Revise, Inc., Burlington, MA.)
etched surfaces (less for small, deep pits, for example), and potentially diffusion of reaction products away from the etched features. These phenomena are sometimes taken into account by designing mask patterns appropriately, based on experimental etch results.

Many dry silicon etch chemistries do not attack the dielectrics and metals used in CMOS processing, and if isotropic can be used to undercut structures to form bridges and cantilevers. For example, Linder et al. [78] discussed various structures that could be fabricated by undercutting aluminum thin-film regions using plasma etching with an \( \text{SF}_6/\text{O}_2 \) chemistry. They reported a 1.3 \( \mu \text{m} \)/min silicon etch rate with \( >300:1 \) selectivity for aluminum and nearly full isotropy (0.8 : 1 undercut : depth ratio).

Contrary to popular belief, it is possible (in some cases) to modulate silicon plasma etch anisotropy via local dopant concentrations (Schwartz and Schaible [79]). Li et al. demonstrated that a Cl-based plasma etch can be used to etch lightly doped p- or n-type silicon anisotropically and heavily n-doped silicon isotropically [80]. By forming buried n- layers beneath a lightly doped epitaxial layer, they were able to selectively undercut structures above the buried n- regions.

1) High-Aspect-Ratio Dry-Etching Methods: The ability to etch deep, anisotropic structures in silicon is of considerable interest in the micromachining community for a variety of applications, including fabricating deep fluidic channels and single-crystal mechanical structures with extremely high aspect ratios and uniform, well-defined mechanical properties. There are several approaches to obtaining deep etching with high anisotropy, and at least three commercial etchers designed for this purpose are currently available.

Cryogenic cooling of the wafer can greatly enhance anisotropy of etching. Commercial machines have appeared on the market using this principle (e.g., Alcatel, San Jose, CA). By cooling the chuck to liquid nitrogen temperature (77 K) and using a helium gas flow under the wafer to transfer heat, the wafer’s temperature can be maintained at cryogenic temperatures during etching. Apparently, the mechanism is condensation of the reactant gases on the side-walls of the etched structures (condensing gas at the bottoms of the structures is removed by ion bombardment). A potentially important issue with cryogenic dry etching is that if microstructures become thermally isolated due to the etching, cryogenic temperatures (and hence high aspect ratios) may not be maintained locally.

Using pure \( \text{SF}_6 \), the Alcatel machine can yield aspect ratios on the order of 30:1 and is capable of etching all of the way through a full-thickness silicon wafer. For useful examples of cryogenic dry etching, see Murakami et al. [81], or Esashi et al. [82]. A very-high-aspect-ratio silicon etching method referred to as deep (D)RIE relies on a high-density (inductively coupled) plasma source and an alternating process of etching and protective polymer deposition [83] to achieve anisotropy on the order of 30:1 (side-wall angles 90° ± 2°).

selectivities of 120–200:1, and etch rates on the order of 2–3 \( \mu \text{m} \)/min (see Klaassen et al. [84] and Bhargwaj et al. [85]). The practical maximum etch-depth capability of this approach is on the order of 1 mm, and precise etch depths can readily be obtained using buried SiO\(_2\) etch-stop layers (e.g., formed by bonding an oxidized wafer to a second wafer).

The concept of alternating between etching and polymer deposition is described in the German patent of Lärm and Schilp [83]. The etching step uses \( \text{SF}_6/\text{Ar} \) with a substrate bias of \( \approx -5 \) to \( -30 \) V so that the cations generated in the plasma are accelerated nearly vertically into the substrate being etched. After etching for a short time, the polymerization process is started. A mixture of trifluoromethane (CHF\(_3\)) and argon is used (although other fluorocarbon gases can also be utilized), and all exposed surfaces (side walls and horizontal surfaces) are coated with a Teflon-like (polymerized CF\(_3\)) polymer layer approximately 50 nm thick. If ion bombardment, due to a small applied bias voltage, is used during the polymerization step, the formation of polymer on the horizontal surfaces can essentially be prevented. The etching step is then repeated, and the polymer deposited on the horizontal surfaces is rapidly moved due to the ion bombardment and the presence of reactive fluorine radicals. Commercial etchers of this type are available from Surface Technology Systems, Ltd., Redwood City, CA, and Plasma-Therm, Inc., St. Petersburg, FL.

It is possible to combine DRIE (or cryogenic dry etching) with fusion bonding to fabricate a wide variety of mechanical devices. As illustrated in Fig. 6(a), a “handle” wafer is pre-etched to form cavities. A second silicon wafer is fusion bonded to it, forming trapped spaces between the wafers. Etching down to the interface between the wafers—a thin SiO\(_2\) layer that stops the etching—features above the cavities are free to move, while those on the silicon dioxide are anchored. Thus, a number of dry-etched mechanisms can be fabricated, as illustrated in Fig. 6(b) and (c) [86].

2) Variable Anisotropy Etch Processes: Single-crystal silicon microstructures such as cantilevers, suspended beams, etc. can be realized using a combination of anisotropic and isotropic dry etches. By switching between them during the process, it is possible to form undercut structures. In the process described by Shaw et al. [87]–[89] and Zhang and MacDonald [90], [91], a silicon wafer was coated with 2.5 \( \mu \text{m} \) of PECVD silicon dioxide, and patterned resist was used to pattern this masking oxide using magnetron ion etching. This was followed by an anisotropic Cl\(_2)/\text{BCl}_3 \) RIE etch to form the trenches and the deposition of a thin (0.3 \( \mu \text{m} \) ) PECVD silicon dioxide layer to protect the side walls. The bottoms of the oxide-coated trenches were opened with a CF\(_3\) RIE etch, followed by a second anisotropic Cl\(_2)/\text{BCl}_3 \) RIE etch to deepen the trenches. The beam structures thus formed were then undercut using an isotropic SF\(_6\) etch, after which aluminum was sputtered to form electrostatic
Fig. 6. Examples of single-crystal silicon structures fabricated using DRIE. (a) Illustration of a process for fusion bonding two wafers together, the lower one with pre-etched pits, and fabricating movable and anchored single-crystal silicon structures by using DRIE (a hypothetical structure is shown for illustrative purposes in top and side view). (b) SEM of a single-crystal silicon leaf spring etched through a top wafer bonded to an underlying silicon substrate with pre-etched pits, allowing the released spring to move. (c) SEM of a DRIE-released accelerometer structure with on-chip CMOS signal-processing circuitry (shown at the upper left), with inset showing the individual silicon electrode fingers, which are 4 μm wide and 60 μm tall [86]. The mechanical sections of the accelerometer are 1 x 1.5 mm in size. (SEM’s courtesy of Lucas NovaSensor, Fremont, CA.)

There are many possible variations on such processes wherein combinations of anisotropic and isotropic etches are used to obtain the desired geometries. Naturally, there are geometric limitations on structures that can reasonably be undercut, generally restricting the structures obtained to assemblies of uniform cross-section beams. A variant of the process that can produce two levels of electrically isolated, suspended microstructures was reported by Hofmann and MacDonald [92]. Such single-crystal silicon structures can also be combined with active circuitry if they are formed on prefabricated, circuit-bearing wafers as a postprocessing step [93]. Fedder et al. [94], [95] demonstrated the use of a similar approach of combined anisotropic/isotropic dry etching to fabricate laminated aluminum/silicon dioxide mechanical structures using the top-level metal in a standard CMOS process as an etch mask for the silicon. They used an anisotropic silicon dioxide etch (CHF₃/O₂) to remove the silicon dioxide, followed by an anisotropic silicon etch (SF₆/O₂) and finally an isotropic silicon etch (SF₆/O₂ with lower O₂ flow and overall pressure) to release the mechanical structures. The process is illustrated in Fig. 7(a). An example structure fabricated using this type of approach is
V. CONCLUSION

There is a great variety of possible methods for bulk etching silicon. For a given application, the appropriate choice of etching method depends upon a number of factors, including the shapes of the desired structures, the resulting surface roughness, etchant cost, equipment cost, safety, process compatibility, and availability. In many cases, such as etching simple structures such as membranes, grooves, and reflective surfaces, beaker-based wet chemical etching of silicon will be entirely adequate. In others, such as the undercutting of delicate micrometer-scale mechanisms, that this paper will provide useful guidance in making such decisions.

Last, it should be pointed out that similar bulk-etching methods exist for other materials, beyond the scope of this review. Anisotropic wet etching is generally limited to crystalline materials such as semiconductors (see Runyan and Bean [96] and Hjort [97]). Isotropic wet and dry etchants exist for a wide variety of materials (see Vossen and Kern [98], Williams and Müller [6], and Runyan and Bean [96]). Anisotropic plasma-phase etches are also available for many substrates (see, for example, Zhang and MacDonald [99]), since the anisotropy is seldom greatly influenced by the crystal structure (or lack thereof) of a
A2O and HCOOH based etchants for silicon micromachining: In-HF-CH


References

[1] H. R. Robbins and B. Schwartz, “Chemical etching of silic-


[6] W. Kern, “Chemical etching of silicon, germanium, gallium ar-


[9] E. Basous, Fabrication of novel three-dimensional microstruc-


[22] G. Kamins and M. S. D. Vrentas, “Anisotropic etching of sil-


[24] S. Amiri and T. J. Rodgers, “V MOS transistors on (110) sil-

[25] S. Amiri and T. J. Rodgers, “V MOS transistors on (110) sil-

[26] S. Amiri and T. J. Rodgers, “V MOS transistors on (110) sil-
Gregory T. A. Kovacs (Member, IEEE) received the B.A.Sc. degree in electrical engineering from the University of British Columbia, Canada, in 1984, the M.S. degree in bioengineering from the University of California, Berkeley, in 1985, and the Ph.D. degree in electrical engineering and the M.D. degree from Stanford University, Stanford, CA, in 1990 and 1992, respectively. He is an Associate Professor of Electrical Engineering at Stanford University, where he has been a member of the Faculty since 1991. He teaches courses in electronic circuits and micromachined transducers. His present research areas include solid-state sensors and actuators, micromachining technologies, biological and medical applications of fluidic and microfluidic circuits, and micro/nanofabrication processes. He has authored more than 80 technical publications and has received several patents. He held the Robert N. Noyce Family Faculty Scholar Chair from 1992 to 1994. He has broad industry experience in the design of circuits and instruments, commercial product design, and intellectual property law consulting. In addition, he has been one of the founders of several technology companies, most recently Cepheid, Inc., in Sunnyvale, CA.

Dr. Kovacs received a National Science Foundation Young Investigator Award in 1993. He was appointed to the Defense Sciences Research Council in 1995. He became a Truman Fellow in 1994 and a University Fellow in 1996.

Nadin I. Maluf (Member, IEEE) received the B.E. degree from the American University of Beirut, Lebanon, in 1984, the M.S. degree from the California Institute of Technology, Pasadena, in 1985, and the Ph.D. degree from Stanford University, Stanford, CA, in 1991, all in electrical engineering.

From 1985 to 1988, he was with Siliconix Inc., Santa Clara, CA, where he was responsible for the development of high-voltage integrated circuits. His doctoral dissertation work at Stanford was on high-resolution optical, electron beam, and X-ray lithography. From 1991 to 1994, he was a Research Associate and Lecturer at Stanford University, where he worked on micromachined transducers for medical applications and taught graduate courses on solid-state physics and semiconductor devices. He presently is Manager of Advanced Technologies with Lucas NovaSensor, Fremont, CA, where his group is responsible for the development of advanced technologies and products for the automotive, medical, consumer, industrial, and aerospace markets. He also is a Consulting Assistant Professor of Electrical Engineering at Stanford University and serves on the advisory boards of a number of academic and industrial organizations. He has authored or coauthored more than 60 articles and patents in the areas of sensors and actuators, semiconductor technology, and microinstrumentation.

Kurt E. Petersen (Fellow, IEEE) received the B.S. degree (cum laude) from the University of California, Berkeley, in 1970 and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1975, both in electrical engineering.

He established a micromachining research group at IBM from 1975 to 1982. In 1982, he published the review paper "Silicon as a Mechanical Material" in the Proceedings of the IEEE (May 1982). This paper is still the most frequently referenced work in the field of micromachining and microelectromechanical systems. Since 1982, he has cofounded three successful companies in micromachining technology: Transovery Devices, Inc., in 1982, NovelSensor in 1985, and Cepheid in 1995. He has been founded with the purpose of applying advanced micromachining techniques to biomedical and bioanalytical microfluidic systems and instruments. He is the author of more than 80 papers and presentations and has received 19 patents in the field. Since 1994, he has been a Consulting Professor at Stanford University, Stanford, CA.