Physical Chemistry of Hydrogenous Species in the Si–SiO_2 System

Edward H. Poindexter
Army Research Laboratory, Fort Monmouth, NJ 07703, U.S.A.

Z. Naturforsch. **50a,** 653–665 (1995); received February 6, 1995

Dedicated to Prof. W. Müller-Warmuth on the occasion of his 65th birthday

Research on hydrogenous species in the Si–SiO_2 system is reviewed and examined. Some aspects of thermal silica on silicon are explained by comparison with crystalline quartz or bulk amorphous fused silica. Hydrogen behavior in the Si–SiO_2 system is complicated by the unique features of device processing technology, an electrified interface, and high electric fields. An electrochemical model of the negative-bias-field-induced degradation of the system is used as a starting point for discussion of diffusion and solubility, atomic H disposition, thermochemical phenomena, and radiation damage. It is thereby hoped to provide new approaches for complete modeling of hydrogen physical chemistry in the Si–SiO_2 system.

Introduction

Hydrogen is very important in integrated circuit (IC) technology, which is now about 30 years old; it is critical in the pervasive metal-oxide-silicon field-effect transistors (MOS FET). It is used for passivation of harmful electrical traps at the Si–SiO_2 interface. Despite this beneficial role, various problems are related to hydrogen. Perhaps strangely, there has been no systematic research on the physical chemistry of hydrogen and its compounds in the Si–SiO_2 system. Much knowledge has been accumulated, but it is fragmentary. All these disconnected findings, however, now comprise a substantial lore. In this paper, selected observations and interpretations will be further examined. Our own recent studies, which correct one important deficiency in the knowledge base, will be presented. Building upon this latter work and other insights, we will attempt to coordinate important but discordant findings into tentative hypotheses for modeling of hydrogen in the Si–SiO_2 system.

Two of the bulk forms of silica, quartz and amorphous fused SiO_2, have been extensively studied and provide a large amount of fundamental physical and chemical knowledge which can be applied to understanding of the thin-film, thermally-grown silica of IC technology. Thin-film silica, however, offers the important distinctions of novel device processing methods, an electrified interface, and the easy application of a high electric field (>1 MV cm^{-1}). Unique chemical and electrochemical behavior thus emerges, which can in turn shed new light on the bulk forms of SiO_2.

We will first present observations on several different hydrogen-related phenomena, without interpretation, since it is necessary to consider interrelated aspects of the whole in order to develop hypotheses for the individual parts. We will then discuss and interpret the phenomena in the same sequence.

Experimental Techniques

Since the main thrust of this paper is a scrutiny of published works, we will not present details of experimental techniques. Our own research on thin-film silica was dominated by electron spin resonance (ESR), with some use of simple MOS capacitance-voltage (C-V) techniques. More sophisticated electrical techniques were used by our colleagues elsewhere. We say nothing about the methods used by researchers for bulk silica. The interested reader should consult cited references for thorough discussion of techniques used for both thin-film and bulk material.

Observed Hydrogen-Related Phenomena in the Si–SiO_2 System

Negative-Bias-Temperature Instability

The negative-bias-temperature instability (NBTI), a seemingly specialized topic, is presented first, because it introduces a new kind of hydrogen-based physicochemical approach to the MOS area. The usual IC tradition is to do research with "dry" or "wet" oxides, as produced by the common IC processing oxidation regimens. It is rare to find an Si–SiO_2 study in which

---

Reprint requests to E. H. Poindexter.

0932-0784 / 95 / 0700-0653 \$ 06.00 © – Verlag der Zeitschrift für Naturforschung, D-72027 Tübingen

$[H_2O]$ was systematically varied throughout a wide range for any purpose, and there are none other that we know of where point defects were correlated with finely graded water content. This may seem surprising; it may be partially due to a reluctance to upset well-tuned production furnaces with non-standard thermal procedures.

The NBTI also serves to illustrate an important difference between studies of thin-film thermal silica and bulk amorphous silica. Hydrogen chemistry of thin-film silica is very often probed via the behavior of atomic-scale or point defects, and especially charged defects, rather than by the macroscopic approaches used for bulk SiO_2. In thin-film silica, the main defects of interest are positive oxide charge Q_{ox}; interface trapped charge Q_{it}; dangling Si bonds on the Si side of the interface P_b centers ($\cdot Si \equiv Si_3$, source of much of Q_{it}); and dangling Si bonds in the oxide, E' centers ($O_3 \equiv Si \cdot .... {}^{+}Si \equiv O_3$, source of some kinds of Q_{ox}).

The NBTI has been a long-standing mystery in MOS integrated circuit devices [1, 2]. The NBTI is manifest as a temporal increase in both interface traps and oxide fixed charge under the action of negative bias stress in the MV/cm range, and/or thermal stress up to 700 K, Figure 1. Each agent individually can produce the effect; but a much stronger or at least faster effect is produced by their combined action. The effect is not generally reversible. Until recently, there was no extant chemical model of the NBTI; nor indeed, has any serious attempt been made to conceive such a model, despite quite a number of studies, cited in our own recent works [3–5].

The NBTI effect is stronger in wet oxides than dry ones, but no systematic study of NBTI as a function of $[H_2O]$ had ever been undertaken. In our recent study it was at first qualitatively observed that samples into which H_2O or atomic H had been introduced above 700 K showed an NBTI, but samples with molecular H_2 did not. For quantitative study, samples were then prepared with graded hydrogenous species content. They were stressed with field (to 10 MV cm^{-1}) or heat (to 700 K), applied independently. The surprising results are shown in Figure 2. The final interface trap Q_{it} does not vary monotonically with hydrogen content; it is maximized at a value of 2×10^{13} cm^{-2} (referred to interface area). The activation energy E_a for depassivation is very low at this value of hydrogen concentration.

It was further observed that induced $Q_{it} = Q_{ox}$, and that both were linearly dependent on field. These

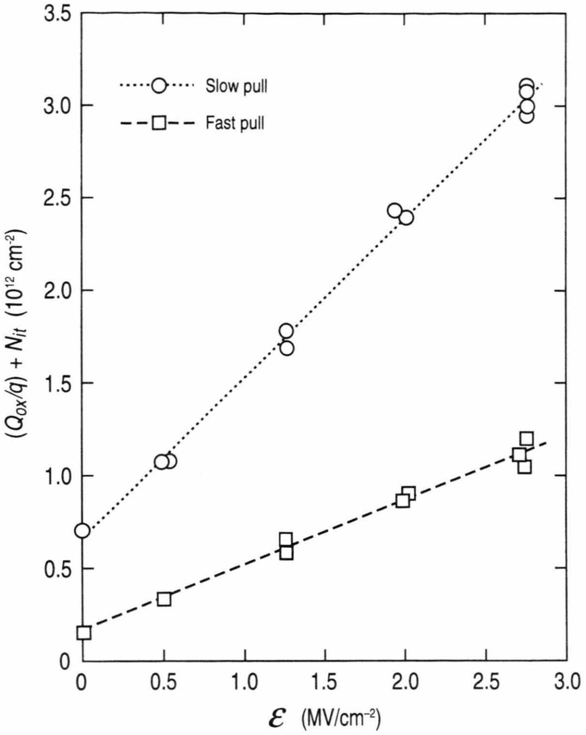

Fig. 1. The negative-bias-temperature instability as a function of electric field. Samples were oxidized in nominally dry O_2, and cooled to 300 K by either a slow pull (>15 s) or fast pull (<3 s) in the oxidizing gas. Adapted from Deal et al. [1].

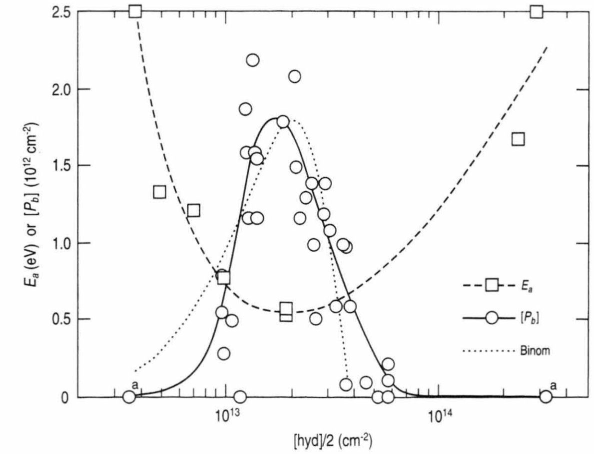

Fig. 2. Generation of $[P_b]$ by electric field at 300 K; E_a for thermal generation without field; $[P_b]$ predicted for binomial distribution of H_2O molecules in SiO_2 pores. The total concentration of hydrogen atoms in all moieties present is [hyd].

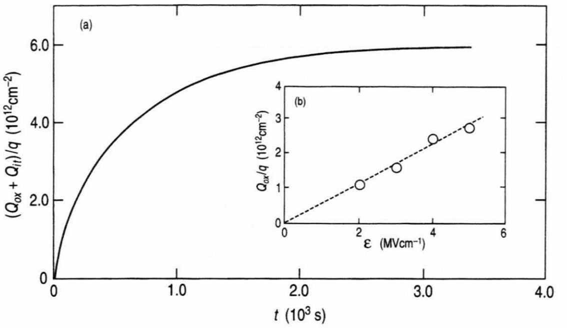

Fig. 3. (a) Logarithmic growth of total charge near the Si–SiO_2 interface by stressing with negative bias field (5 MV cm^{-1}) at 620 K. (b) Linear dependence of deconvolved limiting oxide charge as a function of stressing field.

products developed as a simple logarithmic function of time. The observed time-dependent growth of total charge $(Q_{ox} + Q_{it})/q$ is shown in Fig. 3 for one value of electric field $\mathscr{E}$. The inset shows that the limiting value for Q_{ox}/q is linear in applied field. Inspection shows that the time constants are essentially equal and independent of $\mathscr{E}$. The limiting values of Q_{ox} were defined by

$$Q_{ox} = Q_{obs} - Q_0, \tag{1}$$

where Q_{obs} is the total observed charge, and Q_0 is the initial charge. Both include contributions from charge located at the interface, which had to be deconvolved. It was found that the final Q_{ox} introduced by the NBTI was directly proportional to $\mathscr{E}$.

$$Q_{ox} \sim \mathscr{E}\,. \tag{2}$$

From MOS electrostatics, we have the interface hole (h^+) concentration:

$$[h^+] \sim \mathscr{E}\,. \tag{3}$$

The final value of the Q_{ox} introduced by bias stress was thus determined to be linearly proportional to the initial $[h^+]$, which is the only controlling reactant concentration.

Diffusion and Solubility of H_2O

The low-temperature penetration of H_2O into thermal silica is not consistent with diffusion at high temperatures, where an activation energy of about 0.8 to 1 eV is derived from steam oxidation kinetics [6]. Even below 400 K, water penetrates thin thermal oxides very easily; e.g., in a 90 nm oxide, water vapor at 360 K produces a hydrogen-rich (3×10^{13} cm^{-2}) region 10 nm thick adjacent to the Si–SiO_2 interface in only 10 min [7]. A study of H_2O diffusion at these low temperatures found $E_a = 0.3$ eV [8], a very low value, and clearly indicating a different mechanism than at high temperatures. Water infused below 400 K, however, does not create the NBTI or electron trapping sites in the oxide [7, 9], indicating a different speciation than the higher-temperature infusions.

The overall solubility of H_2O in thin-film SiO_2 is of great importance. Water is much more soluble than oxygen, and hence it accumulates in the oxide layer at substantial concentration if the oxidizing ambient is only very slightly damp. The solubility is even higher near the Si–SiO_2 interface. As a results, it has been impossible to produce truly dry oxides by even the most careful technique. The concentration of hydrogenous species achievable near the interface spans only a very limited range, barely two orders of magnitude, 10^{19} cm^{-3} hydrogen moieties in the driest, 10^{21} cm^{-3} in the (oversaturated) wettest [5].

The form and disposition of all water-related species in the thermal thin film have not been clearly deduced from the various experiments in which water or other hydrogenous species take part. Water does not long remain exclusively as H_2O. Isotope studies show the exchange of oxygen between mobile species and the lattice [10]; this indicates complications in the diffusion process.

Atomic Hydrogen

The behavior of purported atomic hydrogen in Si–SiO_2 is ambiguous and paradoxical. Atomic H is readily detected in radiation-damaged thermal films at low temperature by ESR [11]. The ESR disappears above 100 K, indicating a great decrease in [H]. Nevertheless, the existence of atomic H in some guise in thermal SiO_2 at 300 K and above is indicated by the much easier passivation of P_b centers or interface traps Q_{it} as compared to the effect of molecular H_2. The atomic H was produced either by rf plasma [12, 13] or by reaction of SiO_2 surface silanols (hydroxyls) with an aluminum metallization above 600 K [14]. Simple experiments comparing atomic H and molecular H_2 are shown in Figure 4 [13]. A great difference in anneal temperature between H and H_2 is seen. Measurements from the latter and from some better studies show activation energy $E_a = 0.3$ to 0.6 eV for passivation by atomic H [14], compared to 2.6 eV for

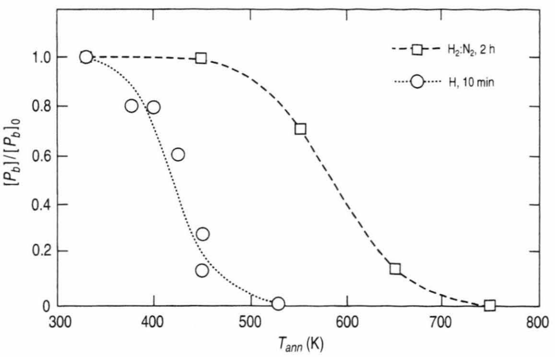

Fig. 4. Passivation of P_b centers by molecular H_2, and rf-plasma-generated atomic H. Adapted from Johnson et al. [11] and Caplan and Johnson [12].

molecular H_2 [15]. The contradiction regarding atomic H above 100 K may require other H-species for explanation.

Radiation-Generated Interface Traps

Generation of Q_{it} by attack of radiolytic hydrogen has been proposed several times [16–18], in accord with a schematic reaction

$$H{-}Si{\equiv}Si_3 + H \rightarrow \cdot Si{\equiv}Si_3 + H_2 \,. \qquad (4)$$

Some appealing experiments have supported this idea or variants thereof. Among them, the temporal growth and decline of positive oxide charges Q_{ox} upon infusion of molecular H_2, correlated with the subsequent growth of Q_{it} [18], Fig. 5, supports a two-stage process. The first is the cracking of H_2 by E' damage centers near the Si/SiO$_2$ interface, again, schematically only:

$$H_2 + E' \rightarrow H{-}E + H^+ \,. \qquad (5)$$

Molecular orbital calculations predict that hydrogen passivation of E' would be an easy reaction [19]; widely unheralded experiments have shown that molecular H_2 passivates E' centers at room temperature, Figure 6 [20]. Later experiments confirmed the ease of the reaction [21, 22] and showed E_a to be very low, only about 0.3 eV [21], also Figure 6. The protons are driven to the interface by the applied positive bias (Radiation damage is generally greater with positive bias.), where they are neutralized upon picking up an electron from the Si surface. Interface traps Q_{it} are then proposed to be generated by the reaction (4).

Unfortunately, these interesting experiments were not performed with ESR; the postulated presence and

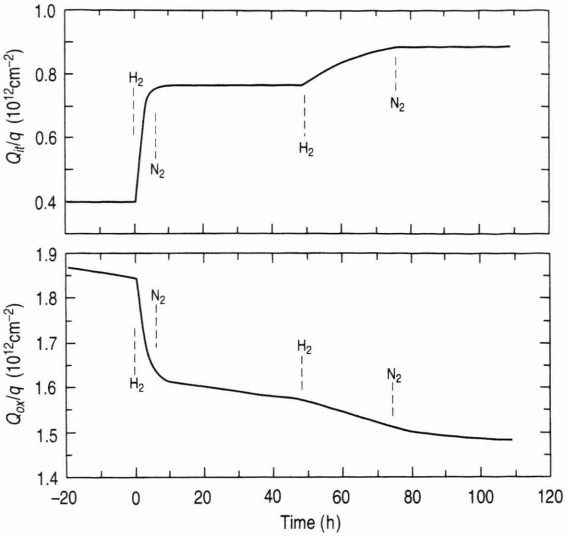

Fig. 5. Concerted evolution of oxide charges Q_{ox}/q and interface traps Q_{it}/q. Radiation at −65 h; gas introduced as marked. After Stahlbush et al. [17].

role of E' centers could only be inferred. Moreover, as will be discussed further, the depassivation of P_b sites by atomic H is in strong disagreement with careful kinetic studies of thermochemical passivation.

Passivation and Depassivation of P_b Centers

Unlimited passivant. Passivation with unlimited supply of H or H_2 shows the inconsistency of rad-damage and annealing results. In Si/SiO$_2$ samples annealed with a constant supply of passivant, either H_2 or (plasma-generated) H ultimately approached the same Q_{it}, characteristic of the anneal temperature [23], Figure 7. The only difference was that atomic H acted much faster. Higher temperature anneal effected a greater reduction in Q_{it}, and a sample previously annealed at a higher temperature would depassivate at a lower temperature, asymptotically reaching the Q_{it} characteristic of the same temperature. These results counterindicate any forward and reverse reaction equilibrium incorporating (4), since the same equilibrium condition (revealed in Q_{it}) could not occur for the widely differing [H] and [H_2] imposed by the contrasting ambients.

The above findings have recently been extended with direct ESR detection of P_b centers produced by exposure to atomic H at 300 K [24]. Oxides were grown on (111) Si in nominally dry O_2 at 1200 K. As

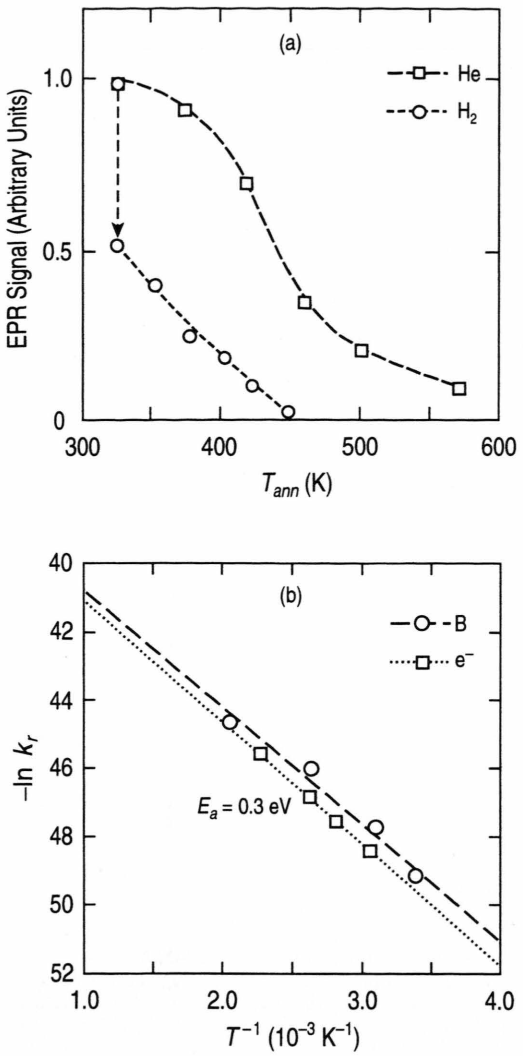

Fig. 6. (a) Anneal of E' centers in thermal thin film SiO_2, generated by α-particle irradiation, by either He or molecular H_2 for 10 min at each point. (b) Anneal of E' centers, generated by boron implantation or electron irradiation, in molecular H_2. After Vikhrev et al. [19] and Li et al. [20].

grown, they showed no detectable P_b. Upon vacuum anneal, a substantial P_b signal developed, $[P_b] > 2 \times 10^{12}\ cm^{-2}$. Both wafers were then subjected to extended (2 h) anneal in atomic H at 300 K. Interestingly, while in the vacuum annealed wafer $[P_b]$ declined markedly, in the as-grown wafer a P_b signal developed, with both wafers tending toward the same $[P_b]$, about $0.4 \times 10^{12}\ cm^{-2}$. These results are shown in Figure 8.

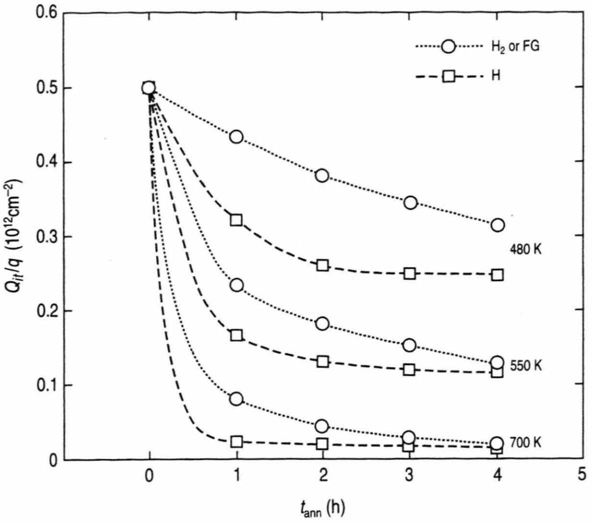

Fig. 7. Interface-trap passivation by molecular H_2 or rf-plasma atomic H. Adapted from DoThanh and Balk [22].

Limited-passivant ambients. Passivation with a limited supply of passivant offers different evidence on depassivation mechanisms. There have been two common kinds of such studies. In one kind, a previously H_2-passivated sample is exposed to nitrogen, argon, or air at 800 K or above [25]. Any unbound hydrogenous species are free to diffuse and evaporate from the outer surface. The result is depassivation of P_b centers and increase in Q_{it}. In the other kind, an aluminum film is deposited atop an unpassivated oxide and then heated [14]. The supply of passivant is the finite amount of atomic H released by reaction of the Al with outer surface silanols. In this case, Q_{it} at first declines, but then ultimately increases again as time is prolonged, shown in Figure 9. In these two kinds of experiments, the possibility of the eventual depassivation by atomic H in the reaction (4) was deemed plausible, but found to be very inconsistent with the observed kinetics. Like the unlimited passivant results, they are in strong disagreement with the atomic-H radiation damage hypothesis.

Oxidation Triangle

An old and well-established experimental principle for MOS oxidation is the so-called oxidation (or Deal) triangle [1, 26], Figure 10. It has been found that if the oxide is grown in well-dried oxygen, and the sample is rapidly cooled in O_2, then the resultant concentra-

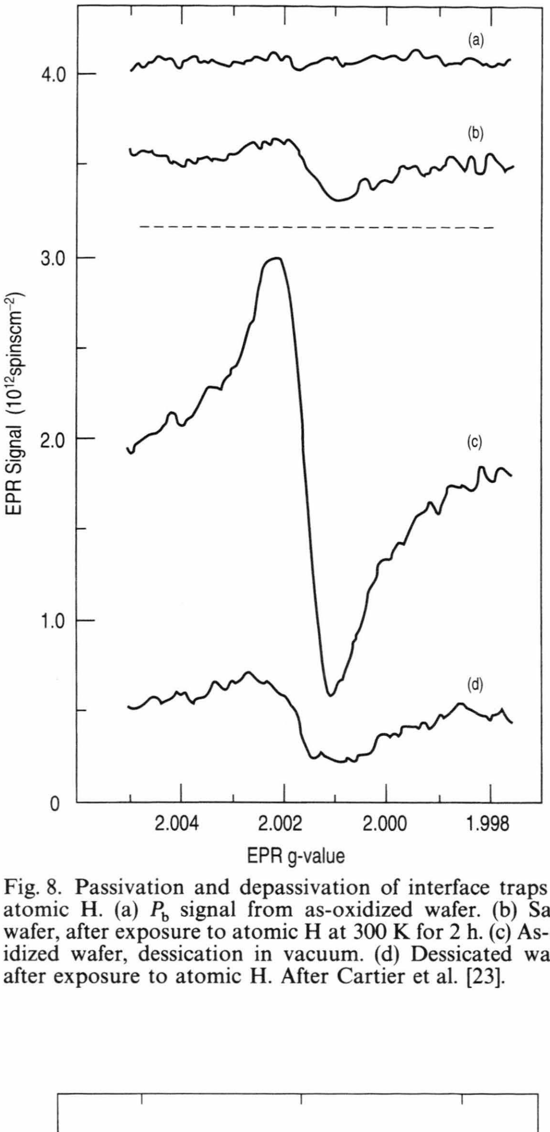

Fig. 8. Passivation and depassivation of interface traps by atomic H. (a) P_b signal from as-oxidized wafer. (b) Same wafer, after exposure to atomic H at 300 K for 2 h. (c) As-oxidized wafer, dessication in vacuum. (d) Dessicated wafer, after exposure to atomic H. After Cartier et al. [23].

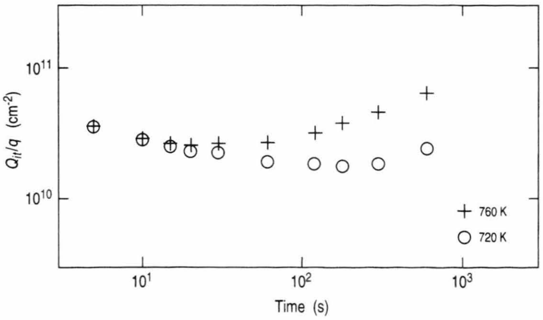

Fig. 9. Passivation/depassivation of interface traps in (111) Si wafer with atomic H from Al/silanol reaction on outer oxide surface. After Reed and Plummer [13].

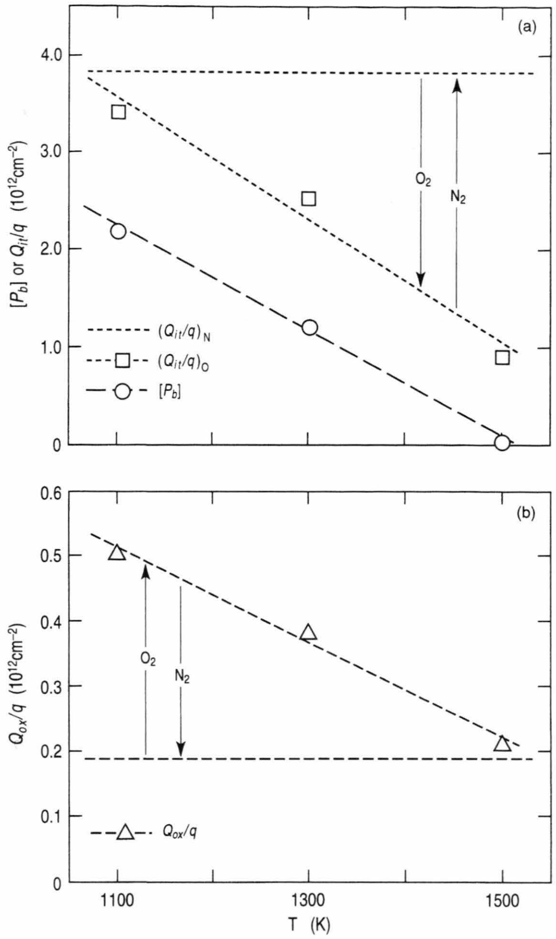

Fig. 10. Deal triangle phenomena for thermal oxidation and inert anneal of the (111) Si surface. (a) Interface trap and P_b center concentrations as a function of thermal treatment in O_2, fast pull, (sloping plot) or N_2, slow or fast pull, (horizontal plot). The Q_{it}/q value for oxidation is characteristic of the temperature used; wafers may be cycled repeatedly between the N_2 and O_2 states. (b) Analogous behavior for oxide charge.

tions of interface dangling bond centers $[P_b]$ or interface traps (Q_{it}), and near-interface oxide positive charges (Q_{ox}) decline linearly with the oxidation temperature. Unhappily, $[P_b]$ and Q_{it} are measured on the as-oxidized wafer, whereas Q_{ox} is usually determined after hydrogen anneal, a rather drastic treatment. The annealed Q_{ox} has long been tacitly assumed to be proportional to the as-oxidized one; recent measurements confirm that the initial Q_{ox} in a dry-oxidized wafer is equal to Q_{it} [4]. Initial $[P_b]$ can be as high as $Q_{it}/2q$ after low-T oxidation, where q is the unit electronic charge;

but additional species contribute to Q_{it}, especially after radiation damage or other stress. The low-T region has been tested to only about 900 K, since the oxide growth is very slow in that region, and other effects obscure the simple relationship.

The designation *triangle* comes from the effect of nitrogen anneals on Q_{ox}; they bring Q_{ox} to a low horizontal baseline value. Nitrogen anneal raises $[P_b]$ and Q_{it} to a uniform high value. Re-exposure to O_2 restores all quantities to the value characteristic of the temperature. The triangle has sometimes been interpreted to arise from a variable degree of lattice mismatch, depending on the different thermal expansions of Si and SiO_2. However, this idea has not been rigorously assessed; and it lacks the flexibility to explain other thermal effects, such as the nitrogen anneal, and even more pertinently, the imprecision and variation in triangle parameters from different batches or laboratories. Another triangle mechanism would be welcome.

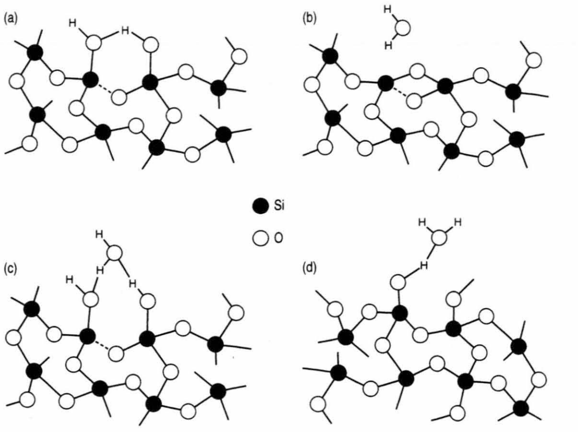

Fig. 11. Disposition of H_2O in amorphous SiO_2. (a) H-bonded vicinal silanols, comprising chemisorbed H_2O. (b) H_2O molecule thermally ejected from silanol pair, with condensation of dehydrated site to form a siloxane linkage, $\equiv Si-O-Si\equiv$. (c) H_2O H-bonded to vicinal silanols, comprising first physisorbed H_2O. (d) H_2O H-bonded to isolated silanol. The highly strained dual siloxane structure $=Si=O_2=Si=$ has not been demonstrated; the vicinal silanols may not be located on nearest-neighbor Si.

Review of Hydrogen Disposition in Bulk Silica

Some features of the well-researched [27–30] situation of hydrogenous species in quartz and bulk amorphous silica are broadly useful in understanding thermal silica films. At equilibrium, hydrogen in bulk silica eventually exists only in a molecular or other bound form, mostly as silanol (OH) or adsorbed H_2O. Free molecular H_2O exists in amorphous bulk SiO_2 if present in quite large concentration, or if there are pore spaces larger than the usual interatomic voids, with suitable pretreatment. Most of the water is normally bound as silanol or OH groups on Si atoms which are not 4-fold coordinated to oxygens. These silanols may be either isolated and standing alone, or situated adjacent to one or more other silanols, Figure 11. If two silanols are immediately adjacent, they are termed vicinal. A pair of such vicinal silanols represents the situation of a chemisorbed H_2O molecule in silica. Heating between 400 K and 800 K eventually liberates the H_2O from the two OH, and the remanent oxygen reforms a bonding siloxane ($\equiv Si-O-Si\equiv$) bridge involving the former host silicons. In contrast, isolated silanols are very difficult to dislodge, requiring temperatures of 1200 K; and then they may liberate protons or H atoms instead of the entire OH group.

If a wet silica is well dessicated above 800 K, it is quite difficult to re-instill silanols by infusion of H_2O at lower temperatures. Some ill-researched higher temperature or growth process is required to incorporate silanols, even though they are a persistent feature of silica prepared in several different ways. One might expect re-entering H_2O to attack preferentially any strained siloxane linkages; about 4 percent strain makes them vulnerable [31]. A completely siloxanated surface, however, is quite hydrophobic; and the regeneration of silanols proceeds very slowly. There is some evidence that the process is initiated at the few isolated silanols remaining on the surface. (Isolated silanols are much more difficult to dislodge than vicinal ones [27]).

Excess molecular H_2O has been found to attach to isolated or vicinal silanols, forming physisorbed H_2O, also shown in Figure 11. Such water may be easily removed at temperatures of about 400 K. Interestingly, a second tier of H_2O molecules adsorbed onto the first physisorbed tier has a higher binding energy than the first tier [27, 32]. This reaffirms that clusters of non-adsorbed H_2O would be viable entities in SiO_2, perhaps in a variable equilibrium with the physi- or chemisorbed H_2O. Nonetheless, the disposition of water in silica remains somewhat ambiguous, despite the numerous studies.

The in- and out-diffusion of H_2O in the temperature range above 800 K indicates that molecular dissociation is part of the processes, since the ultimate concentration is proportional to the square root of the overlying vapor pressure, Figure 12. Ionization into H^+ and OH^- seems unlikely, and the following reaction has been proposed [33]:

$$H_2O + \equiv Si{-}O{-}Si\equiv \;\rightarrow\; \equiv Si{-}OH....H{-}Si\equiv . \tag{6}$$

The diffusional process, however, appears to be carried on by molecular H_2O, with the silanol state providing only a temporary harbor. A number of studies is only in fair agreement on the process and the diffusion parameter D, but a consensus is

$$D = 4\times 10^{-7}\exp(-0.76/kT), \tag{7}$$

where the energies are in eV.

Additional details on the bulk forms of silica will be introduced when pertinent in the discussion to follow. The $Si{-}SiO_2$ system, however, reveals some phenomena which are absent or undetectable in bulk silica, and difficult to explain by bulk silica models. New moieties and mechanisms are required.

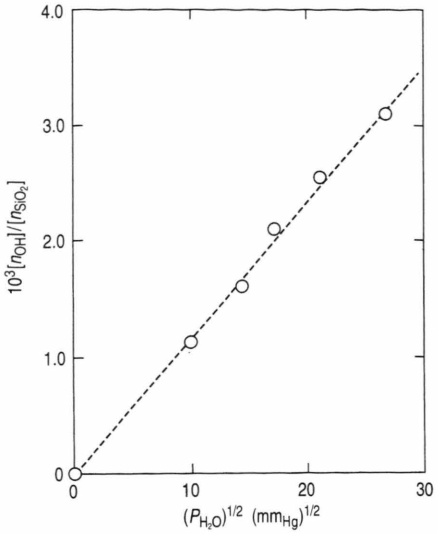

Fig. 12. Solubility of H_2O in vitreous silica as a function of ambient H_2O pressure at 1300 K. The H_2O solubility is expressed as the equivalent number of OH groups per SiO_2 unit, in accord with postulated infusion mechanism. After Moulson and Roberts [32].

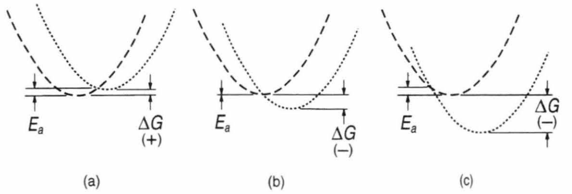

Fig. 13. Reaction coordinates for cases where free energy ΔG may be controlled. In case (a), the reaction is endothermic, with a finite energy of activation E_a. In (b), the reaction is slightly exothermic, and the E_a is zero. In (c), the reaction is still more exothermic, and E_a is again finite. The latter case, where the reaction is retarded as it becomes more exothermic, is called the Marcus inversion regime [34].

Interpretations and Proposed Working Hypotheses for $Si{-}SiO_2$

Negative Bias-Temperature Instability

The NBTI has no parallel in bulk silica; similar phenomena there are simply unknown. Thus we must rely on physico-chemical knowledge from other areas. Reactions which show a minimum in E_a at some intermediate situation are familiar in electrochemistry, and one particular explanation for a general case of minimum E_a is termed the Marcus effect [34] represented in the reaction coordinate diagrams of Figure 13. The left-hand parabola of each pair represents the envelope of bound hydrogen energy levels in the reactant, and the right-hand, in the product. In the situation at (a), a transfer of H from left to right is endothermic with a finite activation energy E_a. In (b), conditions have changed so that the reaction is slightly exothermic, and E_a has vanished. Finally, in (c), the reaction has become still more exothermic, and E_a has again become finite, thus retarding the reaction despite the greater exothermicity.

Consideration of the NBTI reaction kinetics described earlier, the Marcus mechanism, and many electrochemical studies over the years, led to the proposal of the NBTI reaction

$$H{-}Si\equiv Si_3 + H_2O + h^+ \rightarrow \cdot\, Si\equiv Si_3 + H_3O^+ .$$
$$(h^+ = \text{interface hole}) \tag{8}$$

The experiments were analyzed via the kinetic equation for the reaction (8). We have for the rate of creation of product $\cdot\,Si\equiv$ or H_3O^+:

$$\frac{d[H_3O^+]}{dt} = \frac{[\cdot\,Si\equiv Si_3]}{dt} = k_f[H{-}Si\equiv Si_3][H_2O][h^+] - k_r[\cdot\,Si\equiv Si_3][H_3O^+], \tag{9}$$

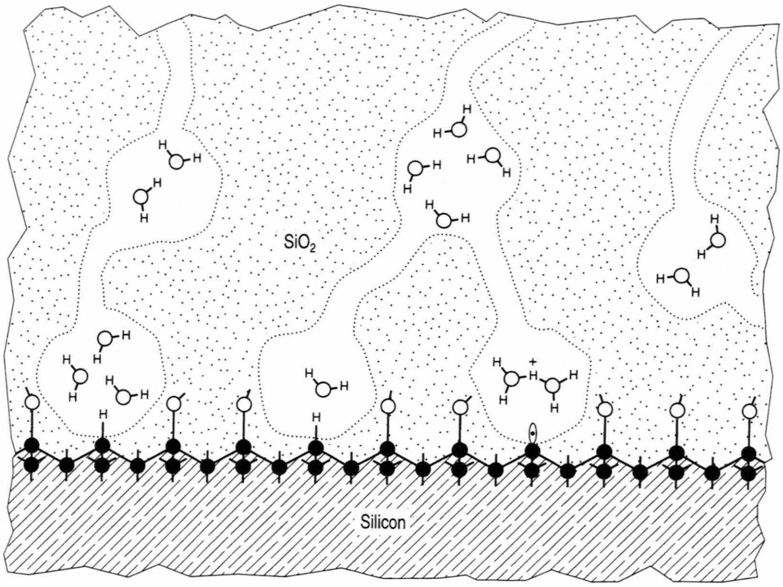

Fig. 14. Occupancy of pores in SiO_2 by H_2O molecules and depassivation of P_b centers by optimum number of H_2O.

where the forward and reverse reaction rate constants are k_f and k_r, respectively. They are assumed to be of Arrhenius form, e.g.

$$k_f = k_{f0} \exp(-E_a/kT). \tag{10}$$

Because only one reactant species (h^+) controls the single-logarithmic progress of the reaction, reactants H_2O and $H-Si \equiv Si_3$ must be present in concentration significantly greater than the limiting product concentration, $\sim 6 \times 10^{12}$ cm^{-2} at 5 MV/cm. Finally, the single-logarithmic dependence also means that there is no back reaction, which would also introduce a quadratic term, i.e., $k_r = 0$. The NBTI reaction is first order in [h^+], and the observed phenomena are thus well-modeled by (8). It is of further interest that holes attracted to the interface do not weaken the H–Si bond, in accord with the invariant reaction rate and molecular orbital calculations [5].

The binding energy of the proton to H_2O has been calculated by molecular orbitals and measured by electrochemical cycles, and is found to range from 8 eV for H_3O^+ to 13 eV for $H_3O^+ \cdot 4H_2O$ [35]. In accord with Marcus, it was suggested for the NBTI that E_a goes through its minimum with two H_2O molecules. If one assumes that SiO_2 is a matrix with pores that can each hold four H_2O molecules, and starts to fill them randomly, then the number of pores with 2 molecules is maximized when their aggregate capacity is half full. The distribution of product concentration predicted by binomial or Poisson statistics is included in Fig. 2, and pore-filling argument for the NBTI is shown in Figure 14.

The argument offered as an explanation for the NBTI is derived from long-standing models developed for the electrode chemistry of electrical cells and batteries. For the moment, it is the most detailed model of a particular MOS degradation process that is extant. With that in the background, we now turn our attention to other Si–SiO_2 problem areas, for some of which no model yet exists.

Diffusion and Solubility of H_2O

The diffusional ingress of H_2O into thin-film thermal silica and its state in the lattice seem to be in general accord with aspects of quartz and bulk vitreous silica, despite lack of detailed models. The diffusion parameters derived for thin-film silica above 700 K [6] agree fairly well with the directly-measured diffusion constants of bulk amorphous silica [33]. Further, the diffusional anneal of radiation-induced defects in thin-film silica above 400 K [36] is in good accord with bulk silica radiation/diffusion studies [37, 38]. Last, the demonstration of oxygen exchange with the lattice confirms a (reversible) transformation of H_2O into silanols proposed from bulk silica.

The rapid ingress of water into thermal silica at 350 K seems rationally explained by an interconnected porous structure in the SiO_2. However, the failure of this low-T water to generate either H_2O-related bulk-oxide electron traps [7] or the NBTI implies complications. There may be regions of dense oxide, between pores or channels, which cannot be effectively permeated at 350 K; or the water may exist only in multi-molecular liquid droplets which disfavor the NBTI or an analogous formation of hydrated electron complexes [39]. Inclusions of liquid water are well known in synthetic quartz, and diffusion via "pipes" is also known [30]. The slower diffusion at high temperature may well be due the drag imposed by the lattice exchange of oxygen, a process which would probably require heat to overcome an activation barrier.

Atomic Hydrogen

The EPR spectrum of atomic H disappears similarly at about 100 K in thin-film thermal silica, bulk amorphous silica [37, 40], and quartz [41, 43]. One might then expect that the bulk materials would help to explain the thin-film silica. Unfortunately, that is not the situation. Unique behavior of atomic hydrogen in thermal SiO_2 is clearly seen in the chemical reaction with dangling Si orbitals at the $Si-SiO_2$ interface. If the same speciation of H takes place in bulk forms of SiO_2, it has not been detected, much less interpreted. Not even the lesser problem of the disappearance of H upon warming has been satisfactorily explained, despite the detailed EPR studies. Altogether, bulk silica yields only a few tangentially pertinent findings to aid in development of a chemical model for $Si-SiO_2$.

Exhaustive infrared studies of quartz have found a defect moiety in which a proton is bonded to a three-coordinated oxygen, itself sited between an Si atom and an Al atom, $O_3{\equiv}Al-(OH)-Si{\equiv}O_3$ [30, 44]. In another study of a particular type of E' or oxygen-vacancy center in quartz, H is found situated between two Si atoms, $O_3{\equiv}Si{=}H\ldots{\cdot}\,Si{\equiv}O_3$; the structure is bistable, with H transferring from one Si to the other [45]. Finally, the diffusion of O_2 in quartz is found to be retarded by the presence of H_2, but much aided by H_2O. An obscure proton effect is deduced [46]. None of the above observations is of direct help in understanding the H– vs. H_2-dilemma in $Si-SiO_2$.

The problem in $Si-SiO_2$ may be related to the purported bahavior of atomic H in semiconductor Si. There, its disposition and transport have been found to be explainable by a dual-moiety model [47–49]. Atomic H seems to exist in (at least) two states in silicon: as a proton, H^+, and as a negative ion, H^{-2}. The proton finds an energetically stable site in the so-called bond-centered location, symbolized as $\equiv Si-H^+-Si\equiv$. The negative ion is found to be stable in the tetrahedrally-coordinated interstitial position. There is currently some debate about whether the H^{-2} site is further stabilized, (The ion is weakly stable in isolation.) by negative electron–electron correlation involving the surrounding lattice, or by some immediate or delayed interaction with the proton. We cannot consider this in detail here; it is important only that a similar dual-ionic hydrogen situation may exist in SiO_2. We cannot conjecture on the diffusional transport of such a dual moiety, or its reaction scheme with interface dangling bonds.

Radiation Damage

The thermal passivation/depassivation findings strongly counterindicate any generation of P_b centers by radiolytic atomic H, no matter what unknown H speciation may exist above 100 K. Objections to atomic-H-induced depassivation models might be partially answered by a model in which the proton, bias-attracted electron e^-, and P_b site are incorporated into one reaction, instead of the two proposed earlier:

$$H^+ + H-Si \equiv Si_3 + e^- \rightarrow \cdot\, Si \equiv Si_3 + H_2\,. \quad (11)$$

A difficulty with any model proposing proton transport lies in the extreme attraction by H_2O (binding energy a minimum of 8 eV); it is likely that the traveling proton will abstract water from vicinal silanols. Thus, it might be wise to consider that an H_3O^+ reaction may occur, but proceed very slowly at 300 K, as in Figure 5.

$$\begin{aligned} H_3O^+ + H-Si \equiv Si_3 + e^- \rightarrow &\cdot\, Si \\ &\equiv Si_3 + H_2O + H_2\,. \end{aligned} \quad (12)$$

The NBTI reaction coordinates, Fig. 5, and tests with high positive bias indicate that this reaction would be energetically possible only in a "dry" oxide, if it can occur at all. High positive bias fields did not produce any P_b in oxides when $[H_2O]$ was set for maximum P_b generation under negative bias [3]. Overall, radiation damage mechanisms must be regarded as unsettled; essential reactant concentrations are undetermined at this time.

Passivation and Depassivation

Unlimited passivant. Since the reaction of (4) is counterindicated by kinetic studies, there must be other hydrogenous species which establish the ultimate Q_{it} equilibrium with either H or H_2 ambient and either high or low initial Q_{it}. The 300 K ESR results are (probably unintentionally) informative and suggestive. The as-oxidized P_b is not detected by ESR, meaning that $[P_b]$ is smaller than about 1×10^{11} cm^{-2}. This is not consistent with the Deal triangle, which would predict about 1×10^{12} cm^{-2} at 1200 K. The most common cause of a lower-than-predicted P_b is a trace of water in the oxidant; a few parts per million is sufficient. This probable inclusion of H_2O suggests a model for the buffered $[P_b]$ in the H- or H_2-saturated oxide.

From the NBTI model, there is a range of $[H_2O]$ which minimizes the activation energy for the depassivation. The reaction rate is strongly affected by two exponential factors: that due to the activation energy, which can vary by a factor of 10^5 with the observed change in $[H_2O]$; and the concentration of holes (or electrons), which can vary by a factor of 10^{10} with internal or external fields. This very sensitive reaction sets the equilibrium point and buffers the system against seemingly important, but relatively much weaker influences, such as other reactant concentrations. We consider several reactions which are reasonable proposed for passivation by H at the interface:

$$H + \cdot Si \equiv Si_3 \rightarrow H{-}Si \equiv Si_3, \qquad (13)$$

$$H_2O + H{-}Si \equiv Si_3 + h^+ \rightarrow \cdot Si \equiv Si_3 + H_3O^+; \qquad (8)$$

and in the oxide:

$$2H + O_2 \rightarrow H_2O. \qquad (14)$$

The latter reaction assumes residual dissolved oxygen from the oxide growth process.

In the N_2- or vacuum-dried oxide, initial $[P_b]$ is high (Processing lore suggests about 3×10^{12} cm^{-2}), but $[H_2O]$ is low. The infusion of atomic H drives reactions (13) and (14) to the right, creating passivated P_b centers H_2O. Increase in H_2O eventually reaches the point where E_a for (8) is at its minimum, which sets up a competing reaction to offset (13). The buildup of positive charge then snuffs reaction (8), and the system would be buffered at an intermediate value of $[P_b]$.

In the damp-oxygen-grown sample, initial $[P_b]$ is zero and $[H_2O]$ is higher than in the dried sample but lower than the critical value for $E_a = 0$, about 5×10^{12} cm^{-2}. Reaction (14) occurs, but there is no immediate reaction (8) or (13). After a while, $[H_2O]$ approaches the minimum E_a value, and reaction (8) commences. Reaction (8) is snuffed by accumulation of positive charge, while reaction (13) has started to passivate the newly-created P_b centers. Eventually, the same buffered condition is obtained as in the dried oxide case.

The effect of temperature, Fig. 7, may also be rationalized by the above reaction scheme. At low temperature, the solubility of H and H_2O will be high, and a high [H] will be maintained by the constant ambient pressure. There will be, however, a loss of H_2O by diffusion and evaporation. If the temperature is raised, all three reactions will be slowed, (13) and (14) due to reduced solubility of reactant [H], and (8) by reduced product $[H_2O]$ from reaction (14). The change in E_a with $[H_2O]$ gives reaction (8) much greater leverage; and once into the region where declining $[H_2O]$ raises E_a, the depassivation reaction is very much retarded. As a result, $[P_b]$ or Q_{it} declines with increasing temperature.

The same ultimate Q_{it} is attained for either H or H_2, though attainment of steady state requires a much longer time with H_2. This may mean that the same general types of reaction take place, but the additional step of cracking the H_2 molecule retards the entire process.

Limited passivant. The aluminized-sample experiment was found by careful kinetic study to be explained by simple dissociation of $H{-}Si \equiv Si_3$, but the NBTI experiments clearly show that H_2O is formed by atomic H in the oxide at the temperatures concerned, about 700 K. As above, competition between reactions occurs; but since the applied [H] is finite, the ultimate condition develops differently. Observations were not extended to reveal the limiting situation. As to the depassivation by heating in N_2 or inert ambient, the molecular H_2, previously applied as passivant, may have produced water; but the reaction would have been much slower. So a role for H_2O in this situation might be less important, but possibly still significant. Careful reconsideration of the kinetics and sample preparation is necessary.

The Deal Oxidation Triangle

A modeling of the Deal triangle on the basis of hydrogenous species is now perhaps obvious from the preceding discussions. It is proposed that the triangle, supposedly applicable to perfectly dry oxides, is in effect a result of an invevitable small amount of water. The same three schematic reactions (13, 8, 14) apply here. At high temperature, reduced solubility of H_2O favors a lower initial $[P_b]$ or Q_{it}, and Q_{ox}. A nitrogen anneal at any moderately high temperature further dries the oxide, reducing Q_{ox} by reduction in $[H_3O^+]$; and it also allows interface $H{-}Si{\equiv}Si_3$ centers to dissociate, increasing $[P_b]$. There are, of course, additional reaction possibilities to consider, e.g.

$$Si + O_2 \rightarrow SiO_2 \qquad (15)$$

$$Si + 2H_2O \rightarrow SiO_2 = 2H_2 \text{ (or 4H)}, \qquad (16)$$

$$H_2 + \cdot Si \equiv Si_3 \rightarrow H{-}Si \equiv Si_3 + H\,. \qquad (17)$$

We do not attempt to include these in the argument here.

The significance of H_3O^+ is reinforced by the oxide charge condition after steam oxidation. The oxide charge is very nearly independent of temperature, Q_{ox}/q being about $0.3 \times 10^{12}\ cm^{-2}$ over the range of Figure 10 [50]. The constant value in the face of a large $[H_2O]$ shows the importance of water in establishing a buffered interface condition. Of further interest is the variation of Q_{ox} as a function of silicon dopant character. Higher Q_{ox} is consistently observed for p-doped Si, consistent with the idea that oxide H_2O could more easily abstract a hole from the Si surface to form H_3O^+.

This discussion of the triangle tacitly offers H_3O^+ as the source of Q_{ox} in the as-oxidized wafer. There has never been an accepted model for the oxidation-induced positive charge; EPR has shown that it is definitely not due to positively-charged E' (oxygen vacancy) centers. The reaction (8) provides a plausible explanation for the initial, as-oxidized equality of Q_{it} and Q_{ox}. Thus, H_3O^+, strongly implicated in the NBTI, is a reasonable source for the Q_{ox} resulting from the thermal oxidation.

Concluding Remarks

Throughout the history of $Si{-}SiO_2$ chemical studies, two crucial factors have been largely overlooked: reactant concentrations and the influence of the electrified interface. As a result, many significant electrical and chemical effects have been but vaguely modeled. Among hydrogen-related phenomena in the $Si{-}SiO_2$ system, only the NBTI has been kinetically modeled with attention to these aspects. It is perhaps the only chemical reaction for which the key reactants, including the governing hydrogenous species, have been systematically considered as a function of their concentration, and for which the importance of the electrified interface has been fully acknowledged.

Several other important $Si{-}SiO_2$ phenomena still lack similar physico-chemical exploration and modeling: diffusion and solubility of hydrogenous species, the evasive behavior of atomic H, radiation damage, passivation/depassivation of interface traps, and the oxidation triangle. In this paper we have suggested tentative working hypotheses for some of these unmodeled phenomena, based on a greater importance of H_2O and a postulated involvement of H_3O^+. To develop complete and quantitative models for all these other hydrogen-related features of the $Si{-}SiO_2$ system would require extensive and systematic exploration. Whether or not the ideas here are valid, it is hoped that they will provide a framework which is useful in tracking the ultimate explanations.

[1] B. E. Deal, M. Sklar, A. S. Grove, and E. H. Snow, J. Electrochem. Soc. **114**, 266 (1967).
[2] E. H. Nicollian and J. R. Brews, MOS (Metal-Oxide-Semiconductor) Physics and Technology, (Wiley-Intersci. New York 1982, p. 794.
[3] G. J. Gerardi, E. H. Poindexter, P. J. Caplan, M. Harmatz, W. R. Buchwald, and N. M. Johnson, J. Electrochem. Soc. **136**, 2609 (1989).
[4] C. E. Blat, E. H. Nicollian, and E. H. Poindexter, J. Appl. Phys. **69**, 1712 (1991).
[5] G. J. Gerardi, E. H. Poindexter, M. Harmatz, W. L. Warren, E. H. Nicollian, and A. H. Edwards, J. Electrochem. Soc. **138**, 3765 (1991).
[6] B. E. Deal and A. S. Grove, J. Appl. Phys. **36**, 3770 (1965).
[7] R. Gale, H. Chew, F. J. Feigl, and C. W. Magee, in The Physics and Chemistry of SiO_2 and the $Si{-}SiO_2$ Interface, C. R. Helms and B. E. Deal, Eds., Plenum, New York 1988, p. 177.
[8] G. L. Holmberg, A. B. Kuper, and F. D. Miraldi, J. Electrochem. Soc. **117**, 677 (1970).
[9] E. H. Poindexter and G. J. Gerardi, Colloids and Surfaces **45**, 213 (1990).
[10] R. Pfeffer and M. Ohring, J. Appl. Phys. **52**, 777 (1981).
[11] K. L. Brower, P. M. Lenahan, and P. V. Dressendorfer, Appl. Phys. Lett. **41**, 251 (1982).
[12] N. M. Johnson, D. K. Biegelsen, and M. D. Moyer, J. Vac. Sci. Technol. **19**, 390 (1981).
[13] P. J. Caplan and N. M. Johnson, unpublished results.

[14] M. L. Reed and J. D. Plummer, J. Appl. Phys. **63,** 5776 (1988).
[15] K. L. Brower, Phys. Rev. B **38,** 9657 (1988).
[16] D. L. Griscom, J. Appl. Phys. **58,** 2524 (1985).
[17] N. S. Saks and D. B. Brown, IEEE Trans. Nucl. Sci. **36,** 1848 (1989).
[18] R. E. Stahlbush, A. H. Edwards, D. L. Griscom, and B. J. Mrstik, J. Appl. Phys. **73,** 658 (1993).
[19] A. H. Edwards, unpublished results.
[20] B. I. Vikhrev, N. N. Gerasimenko, and G. P. Lebedev, Mikroelektronika **6,** 71 (1977).
[21] Z. Li, S. J. Fonash, E. H. Poindexter, M. Harmatz, F. Rong, and W. R. Buchwald, J. Non-Cryst. Solids **126,** 173 (1990).
[22] J. F. Conley and P. M. Lenahan, Appl. Phys. Lett. **62,** 40 (1993).
[23] L. DoThanh and P. Balk, J. Electrochem. Soc. **135,** 1797 (1988).
[24] E. Cartier, J. H. Stathis, and D. A. Buchanan, Appl. Phys. Lett. **63,** 1510 (1993).
[25] K. L. Brower, Phys. Rev. B. **42,** 3444 (1990).
[26] E. H. Nicollian and J. R. Brews, ibid., 790.
[27] R. K. Iler, The Chemistry of Silica, Wiley, New York 1979, p. 622–648.
[28] H. P. Boehm and H. Knözinger, in Catalysis Science and Technology, J. R. Anderson and M. Boudart, Eds. Springer, Berlin 1983, pp. 40–207.
[29] A. V. Kiselev and V. I. Lygin, Infrared Spectra of Surface Compounds, Wiley, New York 1975, pp. 75–103, 123–157, 162.
[30] A. K. Kronenberg, in Silica: Physical Behavior, Geochemistry, and Materials Applications, P. J. Heaney, C. T. Prewitt, and G. V. Gibbs, Eds., Mineralogical Society of America, Washington 1994, p. 123.
[31] M. Heggie and R. Jones, Phil. Mag. Lett. **55,** 47 (1987).
[32] K. Klier and A. C. Zettlemoyer, J. Colloid Interface Sci. **58,** 216 (1977).
[33] A. J. Moulson and J. P. Roberts, Trans. Faraday Soc. **57,** 1208 (1961).
[34] R. A. Marcus, Discuss. Faraday Soc. **29,** 21 (1960).
[35] B. E. Conway, Ionic Hydration in Chemistry and Biophysics, Elsevier, Amsterdam 1981, p. 394, 404.
[36] E. H. Poindexter, P. J. Caplan, and G. J. Gerardi, in The Physics and Chemistry of SiO_2 and the $Si-SiO_2$ Interface, C. R. Helms and B. E. Deal, Eds., Plenum, New York 1988, p. 299.
[37] D. L. Griscom, Nucl. Instrum. Meth. Phys. Res. **B-1,** 481 (1984).
[38] D. L. Griscom, in Structure and Bonding in Noncrystalline Solids, G. E. Walrafen and A. G. Revesz, Eds., Plenum, New York 1986, p. 369.
[39] D.-F. Feng and L. Kevan, Chem. Rev. **80,** 1 (1980).
[40] T. E. Tsai, D. L. Griscom, and E. J. Friebele, Phys. Rev. B **40,** 6374 (1989-II).
[41] R. A. Weeks and M. Abraham, J. Chem. Phys. **42,** 68 (1965).
[42] M. E. Markes and L. E. Halliburton, J. Appl. Phys. **50,** 8172 (1979).
[43] J. Isoya, J. A. Weil, and P. H. Davis, J. Phys. Chem. Solids **44,** 335 (1983).
[44] A. Kats, Philips Research Repts. **17,** 201 (1962).
[45] J. Isoya, J. A. Weil, and L. E. Halliburton, J. Chem. Phys. **74,** 5436 (1981).
[46] S. C. Elphick and C. M. Graham, Nature London **335,** 243 (1988).
[47] S. K. Estreicher and D. M. Maric, Phys. Rev. Lett. **70,** 3963 (1993).
[48] C. G. Van de Walle, P. J. H. Denteener, Y. Bar-Yam, and S. T. Pantelides, Phys. Rev. B **39,** 10791 (1989-II).
[49] N. M. Johnson, C. Herring, and C. G. Van de Walle, Phys. Rev. Lett. **73,** 130 (1994).
[50] R. R. Razouk, L. N. Lie, and B. E. Deal, J. Electrochem. Soc. **128,** 2214 (1981).