# Chemistry of Si-$SiO_2$ interface trap annealing

Michael L. Reed[a] and James D. Plummer
*Center for Integrated Systems, Stanford University, Stanford, California 94305-4070*

(Received 9 November 1987; accepted for publication 17 February 1988)

The kinetics and chemistry of Si-$SiO_2$ interface trap annealing are examined in detail. Measurements of interface trap density $D_{it}$ as a function of anneal time were performed with several process variables as parameters: oxide thickness, anneal ambient, temperature, bulk carrier type, metallization damage, and orientation. Experiments were carried out using rapid thermal processing and capacitance-voltage measurements of aluminum gate metal-oxide-semiconductor capacitors. Anneal temperature and crystal orientation have the strongest effect on the kinetics. ⟨100⟩ interfaces can be described by a power-law temporal variation; ⟨111⟩ kinetics are slightly more complicated. In both cases the experimentally observed anneal behavior is in conflict with the commonly used second-order surface recombination model. We propose a two-reaction model involving atomic hydrogen dimerization and hydrogen/interface trap reactions. This model sucessfully predicts anneal kinetics over a temperature range of 170–500 °C, representing a $10^6$ dynamic range in anneal rates. The difference in anneal behavior between ⟨111⟩ and ⟨100⟩ interfaces is explained by postulating different trap anneal mechanisms for the $P_{b0}$ and $P_{b1}$ defect centers. This hypothesis is supported by trap production kinetics induced by extended anneals.

## I. INTRODUCTION

Defects at the Si-$SiO_2$ interface, such as fixed oxide charge, oxide trapped charge, and interface traps, continue to be limiting factors in the performance of many electronic devices, especially metal-oxide-semiconductor (MOS) transistors. Interfacial charges have serious deleterious effects on threshold voltage, subthreshold slope, and surface carrier mobility. These effects are more important in scaled devices as the defect densities do not scale with device length, and device operation often takes place closer to the surface. Control of interface defects is thus a critical step in modern integrated circuit (IC) manufacture.

One of these defects, the interface trap, has been studied with physical, electrical, optical, and theoretical methods. The generally accepted model of this defect, a trivalently bonded silicon atom with an unpaired electron, has evolved from these studies.[1] Interface traps are passivated by reacting them with hydrogen. This annealing is done at the end of the IC fabrication process by heating the devices in a hydrogen ambient.

This paper describes a kinetic study of Si-$SiO_2$ interface trap anneal chemistry. The goals of this study were to investigate the fundamental physical processes (i.e., transport, recombination, dissociation, etc.) involved in trap annealing, and to determine quantitative rates for these processes. The motivation for this work comes from several sources. For example, annealing behavior is strongly affected by the choice of metal or polysilicon as a gate electrode.[2,3] How the gate material affects reaction rates at the Si-$SiO_2$ interface is an interesting problem. Also, the mechanisms responsible for interface trap production when devices are made or operated in radiation environments are poorly understood. An understanding of anneal kinetics is basic to understanding the physical mechanisms underlying trap production.

This paper is divided into six sections. Section II is a brief introduction to interface trap structure on ⟨111⟩ and ⟨100⟩ interfaces. Section III describes the experimental portion of this work. Kinetic anneal models are introduced in Sec. IV and discussed in Sec. V. The main findings are summarized in Sec. VI.

## II. Si-$SiO_2$ INTERFACE TRAPS

Interface trapped charge, denoted by $Q_{it}$, is charge at the Si-$SiO_2$ interface that changes as the surface potential fluctuates.[4] The charge arises because defect centers, known as interface traps, may be occupied by electrons or holes, hence the net charge may be positive or negative. The trap density, in units of inverse square centimeters, is denoted by $N_{it}$.

Interface traps have energy levels distributed throughout the silicon band gap. The trap concentration at a specific energy is denoted $D_{it}$, and has units of $cm^{-2}\ eV^{-1}$. Interface trap concentrations are often specified as $D_{it}$ at midgap. Unless otherwise specified, we shall mean $D_{it}$ at midgap when we write $D_{it}$. The use of midgap $D_{it}$ as a measure of $N_{it}$ is discussed in Sec. II B.

Various models have been proposed for interface traps. These include interface disorder, misfit dislocations, oxygen atoms, oxygen deficiencies, three layer structures, and metallic impurities.[5] A review of these models may be found in Cheng.[6] The most widely accepted model is a trivalently bonded silicon atom at the Si-$SiO_2$ interface, denoted $Si_3{\equiv}Si\cdot$.[7–14] In this notation, the three horizontal lines represent the three bonds to bulk silicon atoms (the $Si_3$), and the $\cdot$ represents the fourth, unpaired valence electron of the silicon atom (the so called "dangling bond"). For brevity, we will shorten this to $Si\cdot$.

Typical values of $D_{it}$ are $\approx 10^{12}\ cm^{-2}\ eV^{-1}$ before annealing. After a low-temperature anneal, this drops to

[a] Present address: ECE Department, Carnegie Mellon University, Pittsburgh, PA 15213-3890.

 0021-8979/88/125776-18$02.40  

$\sim 10^{10}$ cm$^{-2}$ eV$^{-1}$. The atomic surface density is $\approx 10^{15}$ cm$^{-2}$, so even when the interface trap density is high, the individual traps are far apart. The spacing is sufficiently large to insure the electron wave functions of neighboring traps do not overlap. Therefore, the continuous energy distributions in the band gap are not bands, like impurity broadening, but are due to slight variations in structure and bond angle.[15] The large separation also means that only one trap is involved in a single anneal event.

Interface traps can be measured with a variety of techniques. Capacitance-voltage ($C$-$V$) techniques are the most widely used because they are fast and relatively sensitive, and a simple device structure can be used. In this work we used the high-low $C$-$V$ method[16] to determine $D_{it}$ on aluminum gate MOS capacitors.

### A. $P_{b0}$ and $P_{b1}$ defect centers

Considerable research on interface traps has been performed using electron-spin resonance (ESR). For historical reasons, the defect is denoted $P_b$ in the ESR literature. The identification of the $P_b$ defect as a trivalently bonded silicon atom was performed by Brower[11] by studying $^{29}$Si hyperfine interactions. Various studies have shown the correspondence of interface $P_b$ concentration (as detected with ESR) and the interface trap density $N_{it}$ from electrical measurements.[7,9,13,14] In particular, the correspondence of $Q_{it}$ and $P_b$ concentration under applied gate bias is highly suggestive of a common origin.

Further investigation has revealed at least two varieties of $P_b$. The first one observed, with the notation $P_{b0}$, is found on both $\langle 111\rangle$ and $\langle 100\rangle$ interfaces. On $\langle 111\rangle$ interfaces, the unpaired spin is in the $[111]$ direction. With extreme oxidation conditions (i.e., ones that produce high $D_{it}$), spins oriented in the $[1\bar{1}1]$, $[\bar{1}11]$, and $[11\bar{1}]$ directions can also be found, but these are considerably less numerous than the $[111]$-oriented spins.[17] The intuitive picture of a silicon atom at the $\langle 111\rangle$ interface, backbonded to three substrate silicon atoms with an unsaturated bond sticking straight up into the oxide, is more or less reasonable. The $P_{b0}$ structure on $\langle 100\rangle$ is nearly the same, except the unpaired spin is oriented in the $[111]$ or $[1\bar{1}\bar{1}]$ direction, and is thus at an angle to the interface.

The other $P_b$ center, denoted $P_{b1}$, is found only on $\langle 100\rangle$ interfaces. Like the $P_{b0}$, it is an amphoteric defect, but the correlation energy is smaller, 0.35 eV versus the 0.55 eV of $P_{b0}$.[8] The structure of $P_{b1}$ is not known. Originally, it was thought to be a silicon atom backbonded to two substrate silicon atoms, with the third saturated bond attached to an oxygen atom ($Si_2O{\equiv}Si\cdot$),[7] but this identification has been found to be incorrect, as the calculated energy levels and correlation energy for this defect do not agree with experiment.[18] (They do agree for the $P_{b0}$ defect.) Some suggestions for the $P_{b1}$ structure have been put forth, but these are considered speculative at the present time.

Gerardi *et al.*[8] found approximately equal numbers of $P_{b0}$ and $P_{b1}$ defects on $\langle 100\rangle$ interfaces. The ratio of $P_{b0}$ to $P_{b1}$ concentrations is thought to be a strong function of the oxidation conditions[7,19]; however, a systematic study has yet to be performed.

Little work has been done with ESR on the anneal behavior of interface traps. This is partly because the absolute spin concentrations are very low, particularly with $\langle 100\rangle$; trap annealing reduces the ESR signal-to-noise ratio to unusable levels. Nishi[20] noted the $P_b$ centers can be bleached in the atmosphere at 80–280 °C. He found a 0.66-eV activation energy for the annealing time constant. We will compare this result with our findings in Sec. V. A detailed review of ESR applied to the Si-$SiO_2$ interface can be found in Poindexter.[1]

### B. Midgap $D_{it}$ as a measure of $N_{it}$

Much work has been done in correlating the concentration of unpaired spins of $P_b$ centers, found from ESR, to the interface trap density found from electrical measurements. Poindexter *et al.*[7] have found a 1:1 correlation between $[P_b]$ and midgap $D_{it}$ in wafers pulled from the oxidation furnace in dry oxygen. (The square brackets denote concentration.) Interestingly, there is a numerical correlation between the two measurements, even though the units are different: a spin concentration of $10^{11}$ cm$^{-2}$ corresponds to a midgap $D_{it}$ of $10^{11}$ cm$^{-2}$ eV$^{-1}$. For oxides that undergo a high-temperature anneal, the situation is more complex. The relationship between $[P_b]$ and $D_{it}$ is still linear, but the constant depends on the anneal ambient, and in some cases, the bulk conductivity type. In general, there are more traps measured by electrical measurements than can be accounted for with ESR.[8]

This point is in dispute. Chang, Wu, and Lyon[21] have shown that *all* interface traps in the central 0.7-eV portion of the band gap can be ascribed to amphoteric $P_b$ centers. This applies to traps created with $x$ irradiation and Fowler–Nordheim injection. It is not clear if this also applies to oxidation produced traps. The uncertainties of many measurements, both ESR and electrical, make conclusions about the absolute correlation beyond a factor of 2 unreliable.

Ideally, the interface trap concentration $N_{it}$ should be found by integrating $D_{it}$ over the band gap. This is often impractical and other estimates of $N_{it}$ are used. The integration of $D_{it}$ is not used because measurements of $D_{it}$ near the band edges are difficult, and interface trap response time decreases exponentially away from midgap. The net result is $D_{it}$ measurement accuracy degrades as one moves further away from midgap; integrating these inaccurate $D_{it}$ values will give erroneous values of $N_{it}$.

As discussed above, Poindexter *et al.*[7] have shown a linear correlation between $[P_b]$ and midgap $D_{it}$. We will use this result throughout this study by making the approximation $[P_b] \approx [Si\cdot] \approx D_{it,midgap}$, with the understanding this is a *numerical* correlation, despite the difference in units. Additional justification for this approximation is provided by the following evidence:

(1) Integrated values of $D_{it}$ over the accessible portion of the band gap correlated well with midgap $D_{it}$ in our experiments.

(2) No evidence of different anneal rates for traps of different energies was found.

(3) The finding that all of the measurable traps are $P_b$ centers[21] supports the contention that the midgap value is

formed from the sum of the tails of the broad, amphoteric $P_b$ center peaks.

(4) With proper processing, $D_{it}$ values below $10^9$ $cm^{-2}$ $eV^{-1}$ have been measured with deep level transient spectroscopy.[22] This represents the small number of interface traps present when nearly all of the $P_b$ centers have been annealed. This quantity of traps is nearly an order of magnitude smaller than the minimum densities observed in our experiments. Thus, even if the midgap $D_{it}$ has a component not related to $[P_b]$, its magnitude is probably insignificant.

Although widely used, the approximation of midgap $D_{it}$ for $N_{it}$ is not without its detractors. Counterevidence is provided by nonamphoteric interface traps, which are produced with positive bias temperature aging.[23,24] Traps are produced only in the upper half of the band gap, which is uncharacteristic of amphoteric $P_b$ centers. ESR measurements of traps produced by this method have not been reported, so the nature of these traps remains a mystery.

The use of midgap $D_{it}$ as a measure of $[P_b]$ has been nicely summarized by Poindexter and Caplan[1]:

> "Despite all these caveats, however, (a) the striking, nearly 1:1 quantitative, general correlation of $P_b$ and $D_{it}$, and (b) the detailed tracking of $P_b$ and $D_{it}$ as a function of processing variables are strong evidence of a meaningful physicochemical connection between the two."

It is beyond the scope of this study to conclusively prove the correlation of $D_{it}$ and $N_{it}$. In the interest of progress, we will use $D_{it}$ as an approximation to study the anneal kinetics of interface traps.

### C. Hydrogen annealing of $Q_{it}$

Interface traps are annealed during the last furnace step of integrated circuit manufacture during the "forming gas anneal." In addition to trap annealing, this step also sinters the aluminum/silicon contacts, lowers the contact and metal sheet resistances, and alloys the aluminum lines to the $SiO_2$ dielectric layer. Forming gas is a mixture of hydrogen and nitrogen, generally from 10% to 25% hydrogen. Annealing is typically performed between 400 and 450 °C for 10–45 min.

Many experiments have clearly delineated the importance of hydrogen in the anneal process.[2,3,5,25–37] These experiments can be summarized as follows:

(1) Polysilicon gate devices anneal much slower than devices with an active metal gate, such as aluminum or magnesium. Polysilicon gate devices will partially anneal without hydrogen (i.e., in a furnace ambient of pure nitrogen or argon), presumably because hydrogen is trapped in the poly-grain boundaries. However, lower trap densities will be obtained if annealing is done in an ambient containing hydrogen.

(2) The speed of annealing is dependent on the lateral geometry of polysilicon gate devices, but not aluminum gate devices. This demonstrates that lateral diffusion of hydrogen under polysilicon gates is a rate limiting step.

(3) Devices with active metal gates anneal without ambient hydrogen. Inclusion of a silicon nitride layer between the oxide and metal gate severely impedes the annealing. The interpretation is that the nitride, in which hydrogen diffuses more slowly than in $SiO_2$, blocks the hydrogen generated at the metal-dielectric interface. Similar effects are seen in polysilicon gate devices, where the nitride covers the entire surface. Again, annealing is slower because the nitride impedes the transport of hydrogen from the ambient.

(4) The final trap density after isochronal annealing of a metal gate device decreases as the electronegativity of the metal increases. For example, devices made with magnesium gates, which is more reactive than gold, show trap densities after annealing less than half that of those made with gold electrodes.

(5) Final trap densities of isochronal annealed aluminum gate devices are independent of oxide thickness over a range of 1500–5400 Å. Our results extend this range downward to 200 Å. This indicates hydrogen transport is *not* the rate limiting step in annealing aluminum gate devices, in contrast to polysilicon gate devices.

(6) Annealing of bare oxide interfaces (with no gate) proceeds more rapidly in hydrogen than nitrogen. No ambient effect is seen if annealing is performed after an aluminum gate is deposited.

(7) Reverse annealing (i.e., an increase in trap density) is seen in aluminum gate devices with long anneals at high temperatures. This is not observed with polysilicon gate devices.

(8) Traps produced by x-ray, $\gamma$-ray, or ultraviolet irradiation will anneal without externally supplied hydrogen, even with polysilicon gates.

(9) Implantation of hydrogen near the interface results in lower interface trap density than that obtained solely by annealing.

The picture that emerges from these experiments is one of hydrogen making its way to the interface, where it reacts with the traps and renders them ineffective at exchanging charge with the silicon. The presence of aluminum or other active metals acts to produce hydrogen internally so none need be supplied from the ambient.

The fact that annealing under metal gates is both faster and more effective than with polysilicon gates, or no gates, strongly suggests that two species of hydrogen are involved in the two experiments. In polysilicon gate devices, molecular hydrogen from the ambient or the polysilicon grain boundaries diffuses through the oxide. Atomic hydrogen is responsible for annealing traps when an aluminum gate is present. This has been demonstrated by Johnson and coworkers,[34,35] who annealed interface traps in plasma generated atomic deuterium. (Deuterium was used to facilitate subsequent tracing by secondary ion mass spectroscopy, SIMS.) Annealing by an atomic species mimics the effect of a post-metal anneal.

A major contrast between anneal mechanisms with polysilicon and aluminum gates involves the rate limiting step. Hydrogen transport limits the annealing with polysilicon gates (molecular hydrogen), but not with aluminum gates. This fact has been exploited by Schols and Maes[3] and Fishbein, Watt, and Plummer[37] to measure the diffusion coefficient of $H_2$ in $SiO_2$. Aluminum gate devices were used exclusively in our experiments, even though they are rela-

tively obsolete for most integrated circuit applications. A kinetic study of trap annealing in polysilicon gate capacitors will only yield information on the transport properties of $H_2$ in $SiO_2$; however, annealing of aluminum gate devices is limited by reactions at the Si-$SiO_2$ interface.

Balk[2] was the first to propose that atomic hydrogen is produced at the Al-$SiO_2$ interface, which then diffuses to the Si-$SiO_2$ interface and reacts with the interface trap. Hydrogen is produced by aluminum reacting with trace amounts of water at the interface. This reaction will produce a thin layer of aluminum oxide as well as atomic hydrogen. In most cases, the water is in the form of hydroxyl groups strongly bound to surface silicon atoms. (Adsorbed molecular water can be removed by heating to about 100 °C, but removal of the hydroxyl requires a vacuum anneal above 600 °C.[38,39] Hence, there is always a source of hydrogen at the Al-$SiO_2$ interface.)

An estimate of the initial hydrogen concentration can be obtained from the oxide atomic surface concentration. Assuming the hydrogen is produced by the reaction

$$\mathrm{Al} + \mathrm{OH} \rightarrow \mathrm{AlO} + \mathrm{H}, \tag{1}$$

the surface concentration of hydrogen in $cm^{-2}$ is equal to the OH surface concentration. (The oxide is denoted AlO, rather than $Al_2O_3$, because the composition is not known; in fact, it probably changes during annealing.) This in turn is approximately one-third of the total atomic surface concentration, since the hydroxyl groups bond only to surface silicon atoms, and not oxygen atoms. Finally, the average starting hydrogen concentration in $cm^{-3}$ is related to the hydroxyl surface concentration by

$$[\mathrm{H}]_0 = [\mathrm{OH}]/h, \tag{2}$$

where $h$ is the distance over which the hydrogen penetrates. (Note that [H] has units of $cm^{-3}$, while [OH], like [Si•], is expressed in units of $cm^{-2}$.) SIMS results of Johnson *et al.*[35] show the hydrogen is mostly confined to the oxide for anneal conditions used in this study, so we will use the oxide thickness, $T_{ox}$, as an approximation to $h$. Compensation of shallow acceptors, which is seen when hydrogen penetrates the interface,[34,40,41] was not observed on the $P$-type samples, providing further evidence of hydrogen confinement. These considerations place an upper limit on $[\mathrm{H}]_0$ of about $10^{20}$ $cm^{-3}$. This will be less if reaction (1) does not go to completion, $h$ is larger than $T_{ox}$, or if [OH] is reduced somehow. $[\mathrm{H}]_0$ can be higher if the sample is irradiated, because of oxide damage which releases hydrogen.

We will use a single value of $[\mathrm{H}]_0$ (at a given temperature) to characterize the initial hydrogen concentration through the oxide. This is valid, except near the Al-$SiO_2$ and Si-$SiO_2$ interfaces where some pileup occurs.[35] The hydrogen concentration, as used here, represents an average value over the oxide thickness. The distinction is not important since macroscopic transport of hydrogen though the oxide is not the rate limiting step in annealing metal gate devices.

## III. EXPERIMENTS

Interface trap anneal kinetics were studied experimentally by partial anneals of aluminum gate MOS capacitors. In this section, we describe the sample preparation, anneal and test procedures, and the experimental kinetics as a function of several process parameters.

The general idea in these experiments was to test and retest the device electrical characteristics after repeated partial anneals. Annealing was performed in a commercial rapid thermal annealer (RTA).[42]

### A. Sample preparation

This study used aluminum gate MOS capacitors exclusively. Depending on the experiment, some changes in the fabrication would be made; but for the most part, the sample preparations were similar. The standard procedure was a backside implant, oxidation, backside oxide strip, metallization, and pattern definition.

Most samples were made from 3-in., $N$-type, ⟨100⟩-oriented, device grade silicon. The wafer resistivity was $\approx 8$ Ω cm, which corresponds to a bulk doping density of about $6\times10^{14}$ $cm^{-3}$. The backsides were implanted with $10^{15}$-$cm^{-2}$ phosphorus at 50 keV to improve the backside contact properties. Preoxidation cleans in $H_2O$:$H_2O_2$:HCl, $H_2O$:$NH_4OH$:$H_2O_2$, and 50:1 $H_2O$:HF followed after implantation. Oxidations were performed in dry oxygen in a resistance heated quartz furnace tube. All oxidations in this study were performed at 1000 °C. After the desired oxidation time, the ambient was switched to argon (at 1000 °C) and ramped down to 800 °C. The samples were slowly ($\approx 0.5$ cm/s) pulled in argon to insure a small and reproducible value of $Q_f$ from run to run. Except for the experiment which investigated the effect of different oxide thicknesses, the oxide thickness was generally 200–300 Å.

After oxidation, the wafers were coated with photoresist to mask the oxide during the next step, a backside oxide etch. After stripping and cleaning, pure aluminum was deposited on the frontside, usually by electron beam evaporation. (Flash-evaporated and sputtered aluminum gates were also investigated.) The aluminum was patterned into capacitors using standard negative resist photolithography. If needed, a backside layer of aluminum was applied to improve the resistance of the bulk contact. Square capacitors with an area of $6.08\times10^{-3}$ $cm^2$, surrounded by guard rings, were used as test vehicles. Previous work with aluminum gate capacitors has demonstrated the anneal mechanism is not sensitive to gate area.[2,25,26,43]

### B. Test and anneal procedures

For each sample, five to ten devices were selected. Variation in $D_{it}$ was at most 10%, but typically much less. All of the $D_{it}$ values described in this work are the average of three or more individual devices. High-frequency and quasistatic capacitance-voltage measurements were taken with an HP 4061A measurement system. $D_{it}$ was calculated using the high-low method.[16] We estimate the absolute measurement error in interface trap density is less than $1\times10^{10}$ $cm^{-2}$ $eV^{-1}$.

In the early stages of these experiments, variations in the starting $D_{it}$ values were noted. Values from $9.4\times10^{11}$ to $1.9\times10^{12}$ $cm^{-2}$ $eV^{-1}$ were found on wafers that had presumably identical processing. The problem was the interface

traps annealed at a much lower temperature than expected: 175 °C, the temperature of the photolithography singe step. Because the singe time was not monitored closely, different samples had variations in the starting $D_{it}$ because of various singe durations. This problem was solved by eliminating the singe step (it is not necessary for photoresist adhesion to aluminum) and by limiting the wafer temperature after aluminum deposition to 120 °C, which was sufficiently low to prevent partial annealing of the interface traps. When these precautions were taken, the starting $D_{it}$ values were all close to $4.2\times10^{12}$ cm$^{-2}$ eV$^{-1}$ (for electron beam evaporated gates).

Initial anneals were 5 s at the desired temperature. The instant the RTA lamps extinguished, the sample tray was quickly pulled out of the anneal chamber to provide the fastest possible temperature quench. The sample was immediately removed from the tray with metal tweezers to further increase the cooling rate. Retesting of the samples would commence within 60 s after annealing.

Further anneals of 5 s were performed until the cumulative anneal time reached 20 s (four anneals). Because each subsequent anneal produced a smaller change in the *C-V* characteristics, progressively longer anneal times were used to produce measurable changes in $D_{it}$. In most cases, cumulative anneal times of 0, 5, 10, 15, 20, 30, 60, 120, 180, 300, and 600 s constituted a single anneal run.

Generally, no sample cleaning was done between anneals. Periodic testing for mobile ions was performed using biased temperature stress of annealed devices. $N_m$ values were less than $1\times10^{10}$ cm$^{-2}$, confirming the cleanliness of our anneal/test procedure.

### C. Results

The next seven sections detail the results of our experiments. In each section, we will show graphs of midgap $D_{it}$ versus cumulative anneal time, with some third variable (oxide thickness, temperature, etc.) as a parameter. The kinetic data is plotted on log-log scales throughout. For ease in comparing different data sets, all graphs are scaled identically: the time axis starts at 1 s and covers four decades to $10^4$ s; the trap density limits are $10^9$–$10^{13}$ cm$^{-2}$ eV$^{-1}$, also four decades. Initial values, not plotted, were $\approx4\times10^{12}$ cm$^{-2}$ eV$^{-1}$. Unless otherwise specified, the samples were prepared using *N*-type, ⟨100⟩ wafers and electron beam evaporated aluminum.

#### *1. Validity of the anneal procedure*

The purpose of this experiment was to determine if the anneal/test procedure described in the previous section was sufficient to produce accurate kinetic data. A potential problem is a systematic error involving the RTA ramp-up and ramp-down periods.

Sample annealing can be done two ways: (1) the "cumulative" method, described in the previous section, and (2) the "direct method." In the cumulative method, a single sample is annealed over and over and tested between anneals. The effect of cumulative annealing is assumed to be the same as a single, longer anneal. For example, when we plot data for an anneal time of 15 s, the sample has actually undergone three separate 5-s anneals. In contrast, the direct method uses as many samples as anneal times. Each sample is tested, annealed for the desired time in one step, retested, and not used again.

Ideally, these two procedures should give identical results. However, a problem could arise from the fact that the temperature-time profiles are not perfectly square, but have finite slewing periods and some variation in temperature during the (supposedly) constant temperature portion of the cycle. Some annealing is certain to occur during the ramp-up and rampdown periods. Taking again the example of a 15-s anneal, both the cumulative and the direct methods will have the initial ramp-up period (at 0 s) and the final ramp-down periods (at 15 s). Samples annealed using the cumulative method, however, will have additional ramp-up and ramp-down cycles at 5 and 10 s. The cumulative method will thus *overanneal* the devices, resulting in a distortion of the time axis when the data are analyzed.

This error can be minimized by using the direct anneal procedure, but this has the disadvantage of requiring many more samples. We investigated the error in the cumulative method that results from ignoring the temperature ramping periods by comparing trap densities obtained after annealing using both methods. The results are shown in Fig. 1, which plots midgap $D_{it}$ from the cumulative anneal procedure against that obtained using the direct method. The anneal temperature is 320 °C. The rightmost point represents the preannealed values, around $10^{12}$ cm$^{-2}$ eV$^{-1}$. The next point down and to the left represents $D_{it}$ after 10 s of annealing. The minimum value of $\approx5\times10^9$ cm$^{-2}$ eV$^{-1}$ occurs after 10 min of annealing.

The data fall very close to a line with slope 1 going through the origin. This demonstrates that the error introduced by ignoring the ramping periods in the cumulative method is minimal. Evidently, much more annealing occurs during the 5-s constant temperature portion of the cycle than during the transient portions. This result would definitely change if one were to shorten the minimum anneal time; we did not attempt anneals shorter than 5 s to avoid this problem. Its validity thus shown, the cumulative method was used in the remainder of experiments.

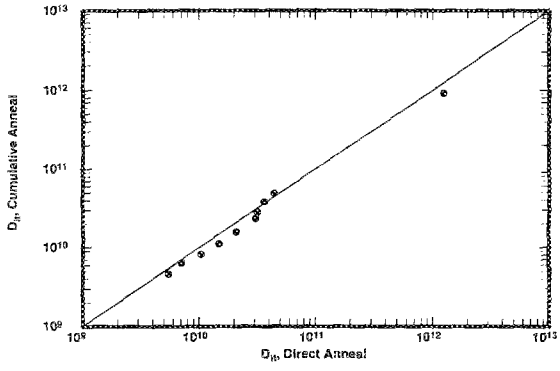


FIG. 1. Midgap interface trap density $D_{it}$ measured by cumulative and direct anneal methods.

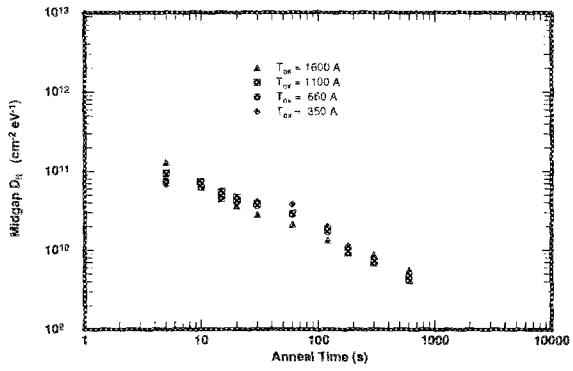


FIG. 2. ⟨100⟩ anneal kinetics at 300 °C for various oxide thicknesses.

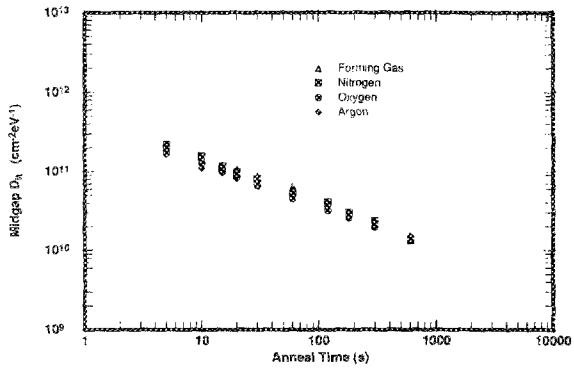


FIG. 3. ⟨100⟩ anneal kinetics at 280 °C in various ambients.

### 2. Oxide thickness

How is interface trap annealing affected by varying the oxide thickness? The answer to this question will tell us if the rate limiting step in annealing is hydrogen transport through the $SiO_2$ bulk. For example, annealing of polysilicon gate devices depends on the lateral distance molecular hydrogen travels in the oxide, under the gate. We wish to see if a similar mass transport limit applies with aluminum gate devices.

Samples were prepared with oxide thicknesses of approximately 350, 660, 1100, and 1600 Å. All of the wafers were in the furnace tube during the 350-Å oxidation to minimize run-to-run differences at the interface. During each anneal, one sample of each thickness was loaded in the RTA chamber to minimize errors from run-to-run variations in anneal cycles. Anneals were performed at 300, 340, and 380 °C in forming gas for this experiment. The 300 °C results are shown in Fig. 2; results at other temperatures are similar.

There is no systematic dependence of anneal behavior on oxide thickness for 350 to 1600 Å. The scatter in the data can be ascribed to experimental uncertainty. This result extends the earlier observation by Yeow, Lamb, and Brotherton[32] that no variation in midgap $D_{it}$ exists after isochronal anneals in the thickness range from 1500 to 5400 Å. From other experiments we have been able to infer no dependence on oxide thickness down to 200 Å. This strongly supports the assertion that hydrogen diffusion through the oxide, from the Al-$SiO_2$ interface to the Si-$SiO_2$ interface, is not a valid model of interface trap annealing in aluminum gate devices.[32] Instead, a model based on a reaction-rate-limited process is more likely.

### 3. Anneal ambient

Previous work with isochronal furnace anneals of aluminum gate capacitors showed similar final values of $D_{it}$, regardless of the ambient.[25] This contrasts with polysilicon gate devices, which show different kinetics in ambients that do not contain hydrogen.[37] The reason for this has already been discussed: an aluminum gate reacts with OH at the Al-$SiO_2$ interface to generate atomic hydrogen, so no hydrogen is needed from the ambient. The goal of this experiment was to determine what effect, if any, the ambient would have on the annealing kinetics, instead of simply looking at the trap densities after isochronal anneals.

Four experimental runs were performed at 280 °C. Anneals were performed in forming gas, nitrogen, oxygen, and argon. Water content of the gases was not measured but was expected to be minimal as they came from liquid sources. Figure 3 shows the results. There is clearly no effect on the anneal kinetics from changing the ambient. These results may be compared to some data of Fishbein *et al.*[37] with polysilicon gate capacitors, replotted in Fig. 4. The poly gate anneal kinetics at 500 °C for nitrogen and forming gas ambients are distinct. This supports the conclusion that aluminum gate capacitors do not depend on hydrogen from the ambient to anneal.

Note that the aluminum gate capacitors were annealed at 280 °C, while the polysilicon gate devices were annealed at 500 °C. Even at the lower temperature, the atomic hydrogen mechanism for annealing aluminum gate devices is faster than the molecular mechanism which anneals capacitors with polysilicon gates. Because of this large rate difference, there is no effect on the aluminum gate kinetics when annealing in hydrogen. We will use this result later as justification for ignoring the effect of molecular hydrogen in modeling the anneal kinetics of aluminum gate devices.

### 4. Temperature

Determination of the anneal kinetics as a function of temperature yields the energetics of the anneal chemistry. Figure 5 shows the kinetics at a variety of temperatures from

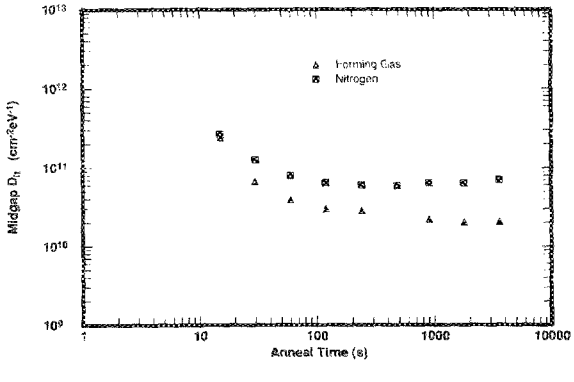


FIG. 4. Polysilicon gate capacitor anneal kinetics at 500 °C in nitrogen and forming gas. After Fishbein *et al.*[37]

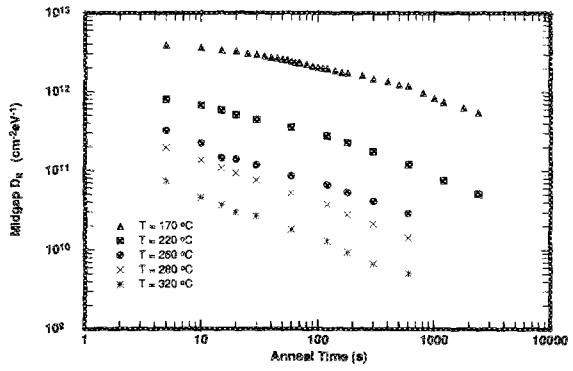


FIG. 5. ⟨100⟩ anneal kinetics from 170 to 320 °C.

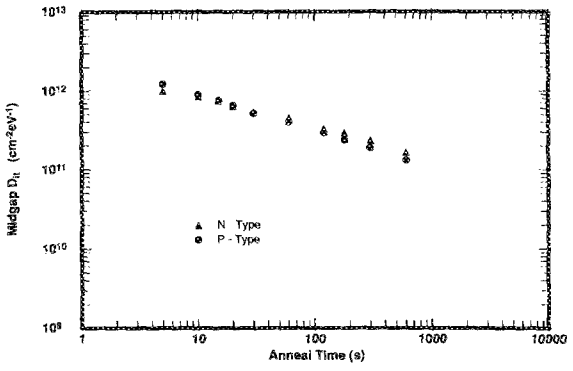


FIG. 7. ⟨100⟩ anneal kinetics at 220 °C for *N*- and *P*-type samples.

170 to 320 °C. Results obtained using temperatures from 340 to 460 °C are shown in Fig. 6. (All of these results are for ⟨100⟩ oriented interfaces.)

A surprising result is just how fast the anneal process actually is. Standard manufacturing anneals are 15–45 min at 400–450 °C. This is necessary to get an acceptable trap density for polysilicon gate devices (Fig. 4), but is clearly overkill for aluminum gates. In fact, anything over 10 min above 380 °C will cause the trap density to increase above a minimum value of approximately $3\times10^{9}\,\mathrm{cm}^{-2}\,\mathrm{eV}^{-1}$.

The upper curve on Fig. 5 (kinetics at 170 °C) demonstrates the difficulty encountered with photoresist processing described earlier. After 20 min at this temperature, midgap $D_{it}$ dropped from $4.2\times10^{12}$ to $7.5\times10^{11}\,\mathrm{cm}^{-2}\,\mathrm{eV}^{-1}$, an 82% decrease.

Figure 6 shows the interface trap production which occurs above 380 °C. The kinetics of this process are discussed elsewere.[44]

### 5. Bulk carrier type

Production of interface traps by ionizing radiation is more efficient with a positive gate bias.[33,45] The reason for this is not clear. The increase in efficiency is thought to be related to the shift in surface potential rather than the oxide field itself.[46] It is interesting to ask if a similar effect occurs during interface trap annealing. Instead of applying a bias, we shifted the surface Fermi level by changing the bulk conductivity type from *N* to *P*. The reasons for this were twofold: (1) A trap production mechanism known as the negative bias temperature instability[47] can obscure the anneal kinetics under bias, and (2) the annealer was not configured for biased anneals.

Sample preparation was identical to the *N*-type samples, except that a lower resistivity ($\approx 3\ \Omega$ cm) starting material was used. This was done to insure the sample remained extrinsic during annealing. A temperature of 220 °C was chosen so that the intrinsic carrier concentration of $\approx 3\times10^{13}$ $\mathrm{cm}^{-3}$ would be comfortably below the extrinsic concentration of $\approx 4\times10^{15}\,\mathrm{cm}^{-3}$. The results are shown in Fig. 7. The *N*- and *P*-type kinetics are the same within experimental error.[48]

This result is somewhat surprising. The difference in surface potential for these two cases is sufficient to change the dominant trap occupancy from $+2$ to 0 (net charge $-q$ to $+q$). The annealing process causes some sort of bonding change of the defect; this result says this bonding change occurs regardless of the number of electrons occupying the trap. This strongly suggests the annealing agent is *neutral*. If, for example, the hydrogen were in the form of a bare proton, one would expect very different behaviors as it approached centers with net negative, neutral, or positive charges.

Griscom[46] explains the bias effect in trap production by requiring an electron from the silicon bulk in one of the reactions. Positive gate bias shifts the surface Fermi level towards the conduction band, increasing the electron density and facilitating the production mechanism. If the reverse mechanism occurred during anneal (that is, an electron were released from the defect), the annealing process would not necessarily be rate controlled by the electron density at the surface. The electron mobility in $SiO_2$ is high; a released electron can move around and easily leave the oxide either at the interface or at the gate. As a result, we cannot say anything definitive about this aspect of Griscom's model based on this experiment.

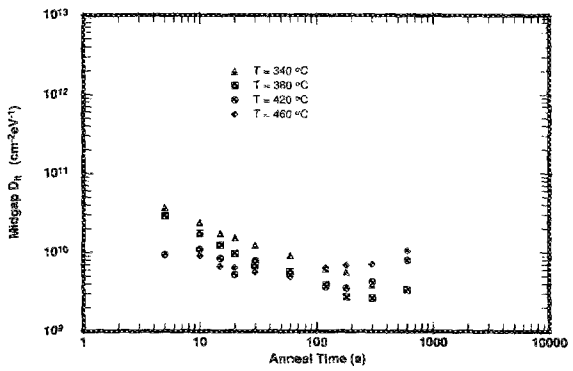


FIG. 6. ⟨100⟩ anneal kinetics from 340 to 460 °C.

### 6. Deposition method

This section describes two experiments. The first portion compares the anneal kinetics of devices with aluminum deposited by flash and electron beam evaporation. This is of interest because flash evaporation does not induce radiation damage in the oxide or interface. The second section de-

scribes results obtained with sputtered metal. Sputtering affords some control over the Al-$SiO_2$ interface and, as we will see, the annealing rate.

The results presented to this point have been on samples with electron beam (*e* beam) evaporated gates. *E* beam gates were used in this study to keep ionic contamination low. However, *e* beam evaporation is known to produce oxide and interface damage from soft x rays. Since this strongly affects interface trap production, its effect on annealing kinetics was investigated.

Control samples were prepared using a wire-feed, flash evaporation system. To minimize thermal cycling after aluminization, these samples were shadow masked to form the capacitor dots, instead of the usual photolithography. Starting midgap $D_{it}$ for the flash evaporated capacitors was approximately $7\times10^{11}$ cm$^{-2}$ eV$^{-1}$. This value is in agreement with that found by Razouk and Deal[49] for $N\langle100\rangle$ silicon oxidized at 1000 °C in dry oxygen. Recall that the initial midgap $D_{it}$ for *e* beam aluminum gate capacitors was approximately $4\times10^{12}$ cm$^{-2}$ eV$^{-1}$, about 600% greater. This is a result of the soft x rays produced by the 6-keV electron beam (continuum flux) and the 1.5-keV $K\alpha$ line radiation characteristic of aluminum.

Annealing kinetics at 260 °C are shown in Fig. 8. We see that the form of the kinetics (i.e., the shape of the curve) is the same for both types of deposition methods. However, the flash evaporated samples, at any given anneal time, have a slightly higher trap density than the *e* beam samples, even though their starting trap density was much smaller than that of the damaged interfaces. Thus, there is a difference in the absolute anneal *rate*, which is evidenced by this vertical translation of the log-log curves.

For modeling purposes, the form of the kinetics is more important than the rate, but it is interesting to understand why there is a rate difference for damaged and undamaged interfaces. A possible explanation follows: ESR measurements[10] show an increase in atomic hydrogen in thermal $SiO_2$ following irradiation. This hydrogen comes from broken Si-OH and Si-H groups which are present in large numbers in thermal oxides. We will show later that the anneal rate is directly proportional to the amount of atomic hydrogen in the oxide. The annealing rate is higher in damaged oxides because of this additional source of hydrogen. Interestingly, the actual $D_{it}$ values are about the same in both cases because the amounts of radiolytic hydrogen and interface damage are both proportional to the radiation dose. In other words, if one increases with radiation the starting trap density by $X$, the anneal rate also increases by $X$ from the increase in radiolytic hydrogen. Since the initial trap density and anneal rate both increase by roughly the same factor, we see comparable values of $D_{it}$ after annealing.

This also explains the sometimes reported difference in anneal behavior between oxidation- and radiation-induced interface traps. Supposedly, traps remaining after oxidation require hydrogen to anneal, while those produced by radiation (such as *e* beam evaporation) will anneal without hydrogen.[50] The problem here is the radiation also generates hydrogen in the $SiO_2$, so observations of anneal behavior in different furnace ambients are not conclusive. Our results show there is no difference in the anneal kinetics between the traps created by different processes; by implication, there are no structural differences either.

In fact, ESR spectra of interface traps created by a variety of techniques are remarkably similar. Defects induced by $\gamma$ and x irradiation, corona discharges, mechanical stress, and hot-carrier injection are the same $P_b$ centers observed after thermal oxidation. The fact that identical anneal kinetics were observed in this experiment provides additional evidence of a common origin of interface traps produced by different processes.

A third method of depositing aluminum is sputtering. Like *e* beam evaporation, sputtering also damages the oxide and interface. An advantage of sputtering is the substrate can be cleaned *in situ* by exposing it directly to the plasma, while the target is shuttered (backsputtering). This experiment was conducted to compare the anneal kinetics of interface traps produced by the sputter process to those produced by *e* beam evaporation, and to see if the anneal rate could be changed by backsputtering the oxide surface.

Samples were prepared from $N\langle100\rangle$ wafers with 240-Å oxides. Backsputtered samples were etched to a 200 Å thickness. 5000 Å of pure aluminum was deposited. Initial midgap $D_{it}$ for both samples was $1.8\times10^{12}$ cm$^{-2}$ eV$^{-1}$, about 2.6 times greater than the flash evaporated aluminum, but less than half that of the *e* beam metal devices.

Anneals were performed at 260 °C. Figure 9 shows the kinetics of the sputtered samples, along with the *e* beam ca-

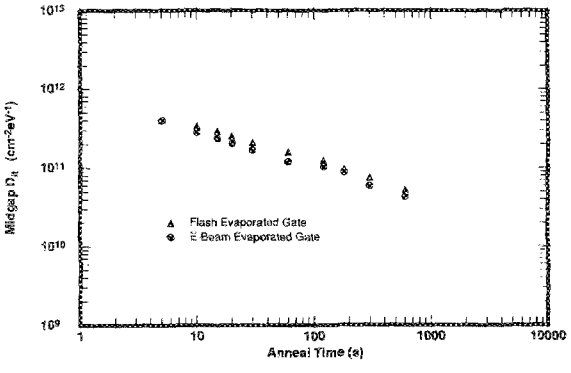


FIG. 8. ⟨100⟩ anneal kinetics at 260 °C for flash and electron beam evaporated gates.

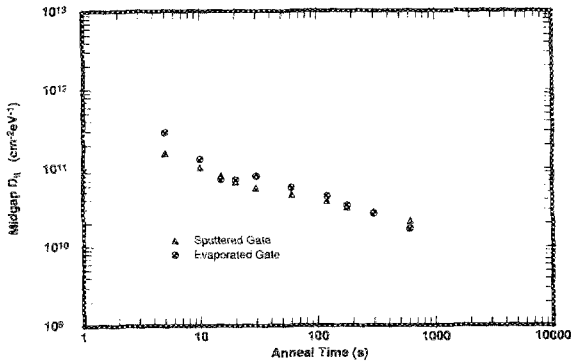


FIG. 9. ⟨100⟩ anneal kinetics at 260 °C for evaporated and sputtered gates.

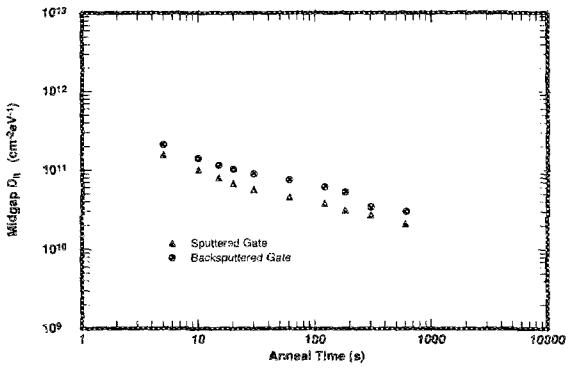


FIG. 10. ⟨100⟩ anneal kinetics at 260 °C for sputtered and backsputtered gates.

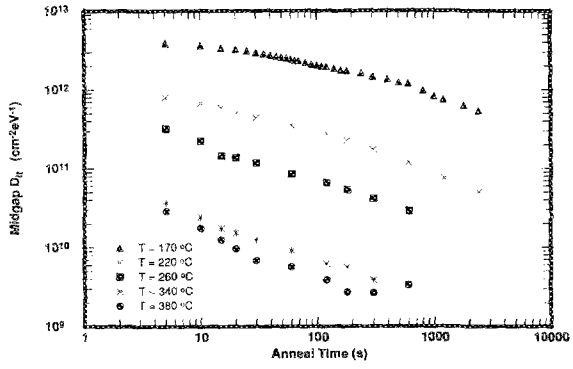


FIG. 12. ⟨100⟩ anneal kinetics at various temperatures. Symbols correspond to those in Fig. 11.

pacitors. As expected, the anneal kinetics are the same.

Figure 10 shows the anneal kinetics of the sputtered and backsputtered devices. The shape of the kinetics are the same, but we see a slightly slower rate with the backsputtered aluminum. Recall that hydroxyl groups bonded to the oxide surface are the source of atomic hydrogen for annealing in aluminum gate capacitors. By backsputtering the oxide, we have removed the hydroxyl groups and thus decreased the amount of hydrogen available to anneal the traps. This reduction in hydrogen is responsible for the slower anneal rate. However, the anneal rate does not go to zero, because the radiation which creates the interface damage also releases hydrogen in the oxide bulk. This hydrogen is still available for annealing. Another source of hydrogen is the bulk Si-OH and Si-H groups which are exposed during the backsputter etch. Those that are now on the surface can react with the aluminum, releasing hydrogen. Thus, annealing still takes place on the backsputtered devices, although at a reduced rate.

### 7. Orientation

Interface trap structure of ⟨100⟩ and ⟨111⟩ interfaces is different. How this difference in structure affects the anneal kinetics was the subject of this experiment.

$N$ ⟨111⟩ samples were prepared using the same procedure as ⟨100⟩. Although unannealed ⟨111⟩ interfaces have higher oxidation-induced trap densities than ⟨100⟩,[49] this difference was obscured by the large number of traps created during $e$ beam evaporation. Initial $D_{it}$ values for both orientations were comparable.

The ⟨111⟩ anneal kinetics at 170, 220, 260, 340, and 380 °C are shown in Fig. 11. For comparison, ⟨100⟩ results at the same temperatures are shown with corresponding symbols in Fig. 12. Results for both orientations at 420 and 460 °C are shown in Fig. 13.

Unlike most of the other variables studied, orientation has a strong effect on the anneal kinetics. At any given temperature, ⟨111⟩ samples anneal slower than ⟨100⟩. However, the difference in rate is not the only contrast. Rather, there is a fundamental difference in the shape of the $D_{it}$ versus anneal time curves, which reflects a difference in the mathematical form, and hence the mechanism, of the anneal process.

Another difference between the two orientations is the minimum value of $D_{it}$ obtained. The lowest trap density for ⟨100⟩ was $\approx 2\times 10^{9}$ cm$^{-2}$ eV$^{-1}$, approaching the resolution limit of the capacitance technique. The ⟨111⟩ samples never dropped below $10^{10}$ cm$^{-2}$ eV$^{-1}$.

The difference in kinetics between ⟨100⟩ and ⟨111⟩ interfaces is a central problem. Later, we will consider a variety of models that attempt to explain this difference. We will find that although there are several differences between the two orientations (such as different $Q_f$ and trap structure), explaining the difference in anneal kinetics is not an easy problem.

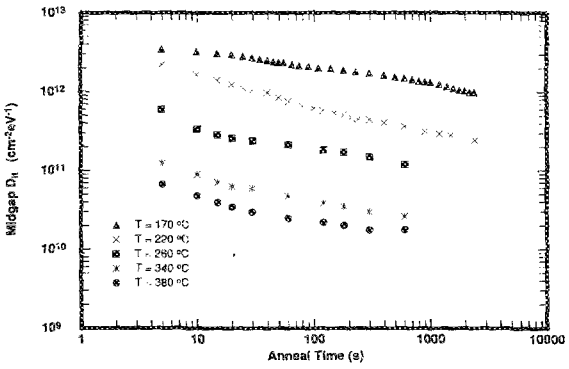


FIG. 11. ⟨111⟩ anneal kinetics at various temperatures.

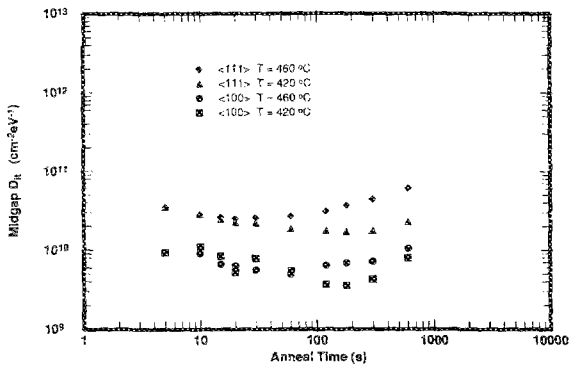


FIG. 13. ⟨111⟩ and ⟨100⟩ anneal kinetics at 420 and 460 °C.

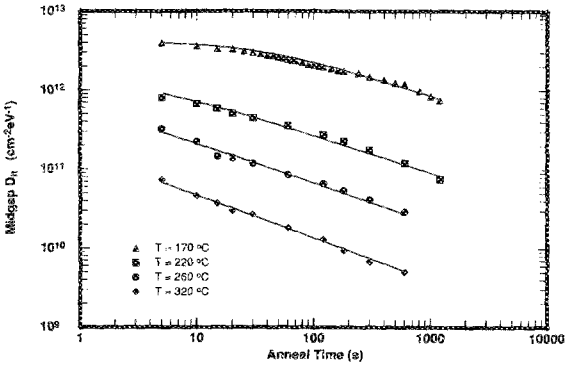


FIG. 14. Empirical power-law fits of ⟨100⟩ anneal kinetics at various temperatures.

### D. Empirical description

The ⟨100⟩ data can be fitted to an empirical equation of the form

$$D_{it}(t) = [D_{it}(0)]/(1+Kt)^{\eta}. \tag{3}$$

In this equation $D_{it}(0)$ is the initial interface trap density, $t$ denotes cumulative anneal time, and $K$ and $\eta$ are constants. The general behavior is a power law.

Weighted least-square fits of Eq. (3) to the ⟨100⟩ data were performed. The results are shown in Fig. 14 for various temperatures. There is good agreement below 380 °C. Above this temperature, trap production occurs and Eq. (3) no longer applies.

The temperature dependence is contained in the constant $K$, which is the anneal rate in inverse seconds. $K$ is thermally activated with an energy of 1.2 eV. The value of the exponent $\eta$ is approximately 0.55 with no systematic dependence on temperature.

Except for orientation, none of the parameters studied have any effect on the mathematical form of the anneal kinetics. Only the speed of annealing is affected by temperature and deposition method. The difference in rate is reflected in the plots of $\log(D_{it})$ vs $\log(t)$ by a vertical translation of the power-law curve, which is a straight line on this type of plot. (Strictly speaking, this is true only when $Kt \gg 1$; this condition is satisfied for $t > 5$ s for all but the 170 °C data.) The difference in ⟨111⟩ kinetics can be seen most dramatically in Fig. 15, which plots ⟨111⟩ and ⟨100⟩ kinetics at 340 °C together. Clearly, there is no way that one set of data can be vertically translated to match the other. Hence, there is more than just a difference in anneal rate between the ⟨111⟩ and ⟨100⟩ interfaces.

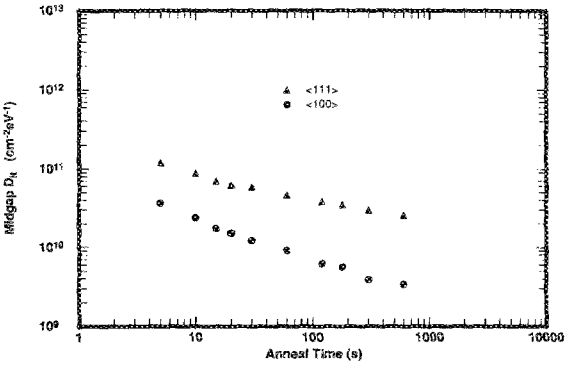


FIG. 15. ⟨111⟩ and ⟨100⟩ anneal kinetics at 340 °C in nitrogen.

If one assumes a power-law dependence for ⟨111⟩ anyway and performs least-square fits of Eq. (3), the mean value of $\eta$ is somewhat lower, about 0.35, and is not as tightly distributed as the ⟨100⟩. However, careful examination of Figs. 11 and 15 suggests that ⟨111⟩ kinetics are not adequately described by a power-law formulation. Although a reasonable fit can be obtained during the early stages, the trap density curve departs from the power law by bending upwards. Because this upward bending is seen in every case, we conclude it is not experimental scatter from a power law (with $\eta \approx 0.35$), but a distinct kinetic behavior.

The next section describes a two reaction model for interface trap annealing. We have shown, from the results on oxide thickness, that hydrogen transport through the oxide is not the rate limiting process. Modification of the Al-$SiO_2$ interface (by backsputtering, for example) affects only the anneal rate. This, coupled with the strong dependence of kinetic form on orientation, indicates reactions at the Si-$SiO_2$ interface are key to understanding the anneal chemistry.

## IV. MODELING

### A. One reaction model

The traditional model of interface trap annealing is a bimolecular recombination of hydrogen and Si•:

$$\mathrm{Si\bullet} + \mathrm{H} \xrightarrow{k_1} \mathrm{SiH}. \tag{4}$$

(SiH is shorthand notation for $Si_3 \equiv Si\text{–}H$.) The rate constant $k_1$ can be found from bimolecular reaction rate thoery[51–53]:

$$k_1 = 2\pi\rho^* D_H, \tag{5}$$

where $\rho^*$ is the equivalent reaction radius and $D_H$ is the diffusion coefficient of the mobile species, in this case atomic hydrogen. $k_1$ has units of cm$^3$ s$^{-1}$. The differential equation corresponding to the second-order reaction (4) is

$$\frac{d}{dt}[\mathrm{Si\bullet}] = -k_1[\mathrm{Si\bullet}][\mathrm{H}]. \tag{6}$$

(Note that [Si•] has units of cm$^{-2}$, while [H] has units of cm$^{-3}$.) According to reaction (4), for every interface trap passivated, one hydrogen atom is consumed. Therefore, we can write [H] as

$$[\mathrm{H}] = (1/T_{ox})(T_{ox}[\mathrm{H}]_0 - [\mathrm{Si\bullet}]_0 + [\mathrm{Si}]), \tag{7}$$

where the subscript 0 denotes initial values. Substituting this expression into Eq. (6) allows us to solve for [Si•]. The final result is

$$[\mathrm{Si\bullet}] = \left[T_{ox}[\mathrm{H}]_0 - [\mathrm{Si\bullet}]_0\right] \Big/ \left\{ \frac{T_{ox}[\mathrm{H}]_0}{[\mathrm{Si\bullet}]_0} \times \exp\left[\left(\frac{T_{ox}[\mathrm{H}]_0 - [\mathrm{Si\bullet}]_0}{T_{ox}}\right) k_1 t\right] - 1 \right\}. \tag{8}$$

An expression equivalent to Eq. (8) was used by Brown *et al.*[54] to model anneal kinetics.

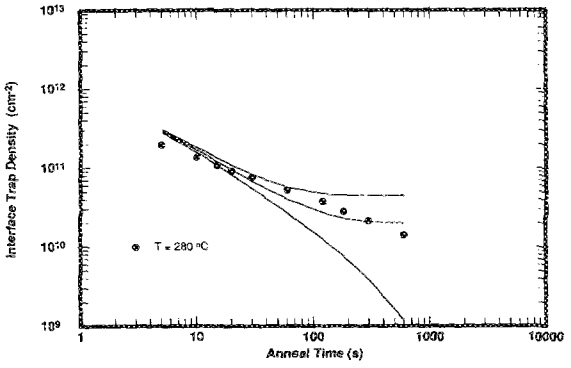


FIG. 16. Second-order solutions of single reaction model, shown with ⟨100⟩ anneal kinetics at 280 °C.

Figure 16 is a plot of Eq. (8), the second-order solution, for various values of $[H]_0$. Because we have only an upper bound, but not an independent experimental estimate of $[H]_0$, we will treat it as a parameter in our modeling. Also shown are the experimental annealing kinetics at 280 °C. Clearly, the second-order solution does not fit the data. The single bimolecular interface reaction (4) is not an accurate model of interface trap annealing.

### B. Two reaction model

In this section we will show that the experimental kinetics can be modeled if, in addition to the interfacial recombination, there is a second reaction[55]:

$$2H \xrightarrow{k_2} H_2. \tag{9}$$

Reaction (9) describes dimerization of atomic hydrogen into $H_2$, which occurs in the $SiO_2$ bulk.

Although it is generally recognized that hydrogen is essential for annealing interface traps, the details of the passivation mechanism are unclear. Specifically, the formation of interfacial SiH has never been demonstrated directly. Measurements of [SiH] before and after annealing have been attempted with infrared spectroscopy, but have yielded inconclusive results.[56] It is possible that hydrogen is involved in the conversion of Si•, without actually being consumed at the interface. Formally,

$$P_b + H \xrightarrow{k_1} \oplus + H, \tag{10}$$

where we have written the annealed $P_b$ center as $\oplus$. Later, we will suggest some definite mechanisms that are represented only schematically in reaction (10).

The calculation of [Si•] as a function of time can thus take two paths: one can assume either a nonconsumptive or a consumptive reaction at the interface. We will take both cases in turn.

#### 1. Nonconsumptive mechanism

The reactions are

$$P_b + H \xrightarrow{k_1} \oplus + H, \tag{11}$$

$$2H \xrightarrow{k_2} H_2, \tag{12}$$

and the corresponding differential equations are

$$\frac{d}{dt}[Si\bullet] = -k_1[Si\bullet][H], \tag{13}$$

$$\frac{d}{dt}[H] = -2k_2[H]^2. \tag{14}$$

Equation (14) can be solved for [H]:

$$[H] = \frac{[H]_0}{1 + 2k_2[H]_0 t}. \tag{15}$$

This solution for [H] is substituted into Eq. (13). The final solution for [Si•] is

$$[Si\bullet] = \frac{[Si\bullet]_0}{(1 + 2k_2[H]_0 t)^{\eta}}, \tag{16}$$

where

$$\eta \equiv k_1/2k_2. \tag{17}$$

Equation (16) has the same form as Eq. (3), the empirical description of the ⟨100⟩ anneal kinetics.

The exponent, $\eta$, is given by the ratio of the two reaction rates. We can estimate its value by substituting in expressions for $k_1$ and $k_2$:

$$\eta = \frac{k_1}{2k_2} = \frac{2\pi\rho_1 D_H}{(2)4\pi\rho_2 D_H} = \frac{\rho_1}{4\rho_2}. \tag{18}$$

The reaction radii are

$$\rho_1 = r_H + r_{Si\bullet}, \tag{19}$$

$$\rho_2 = r_H + r_H. \tag{20}$$

The radius of a hydrogen atom is $\approx 0.5$ Å. The capture radius of an interface trap is not known, but is at least as large as the silicon tetrahedral radius, 1.2 Å, and perhaps up to 50% larger. Taking these two values as limits, we find

$$0.43 < \eta < 0.58. \tag{21}$$

This range is in excellent agreement with the empirical values of $\eta$ found from experiment.

More important, though, is the prediction from the two reaction model that the exponent $\eta$ is independent of temperature (assuming, of course, that the atomic radii do not change, a good assumption in the temperature range of interest). Note the diffusion coefficient $D_H$, which appears in both $k_1$ and $k_2$, cancels out since $\eta$ depends on the ratio of the reaction constants.

The annealing rate, in $s^{-1}$ is given by the product of $k_2$ and $[H]_0$. We need the dependence of $[H]_0$ on temperature before we can calculate the activation energy of $k_2$. This is discussed in Sec. V.

#### 2. Consumptive mechanism

If we assume that the hydrogen bonds to the interface trap, the reactions become

$$P_b + H \xrightarrow{k_1} SiH, \tag{22}$$

$$2H \xrightarrow{k_2} H_2, \tag{23}$$

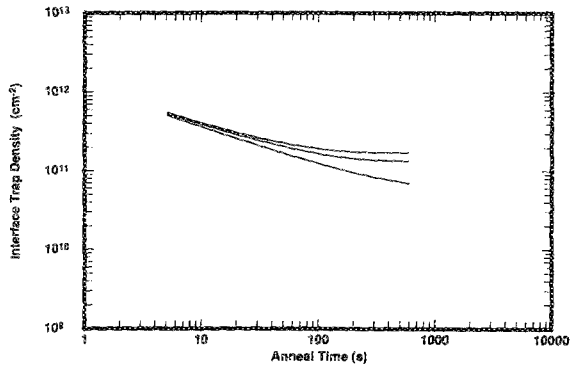


FIG. 17. Effect of changing $[H]_0$ while holding $k_2[H]_0$ constant. Increasing $[H]_0$ causes the solution to more closely approximate a power law.

and the corresponding differential equations are

$$\frac{d}{dt}[\mathrm{Si}\cdot] = -k_1[\mathrm{Si}\cdot][\mathrm{H}], \tag{24}$$

$$\frac{d}{dt}[\mathrm{H}] = -\frac{k_1}{T_{ox}}[\mathrm{Si}\cdot][\mathrm{H}] - 2k_2[\mathrm{H}]^2. \tag{25}$$

Equations (24) and (25) are coupled and cannot be solved analytically. To find [Si·] for the consumptive case, we must resort to numerical techniques.

A numerical solution of these equations is shown in Fig. 17. The various curves correspond to different starting concentrations of hydrogen, $[H]_0$. At short anneal times, the consumptive solution looks like a power law, which is the solution to the nonconsumptive model. The two models also converge in the limit of high $[H]_0$. The reason for this can be seen by comparing Eqs. (14) and (25). They differ only by a term describing hydrogen consumption at the interface. When [H] is relatively high (i.e., at the beginning of the anneal) this term is negligible compared to the dimerization term, since the interfacial recombination is a two-dimensional process, while the dimerization is a three-dimensional bulk process. However, as the hydrogen is used up, consumption at the interface becomes comparable to the bulk dimerization, and the solution for [Si·] departs from a power law.

The consumptive model cannot be used to describe the characteristic power-law kinetics of the ⟨100⟩ results. The value of $[H]_0$ needed to fit the experimental results is over $10^{21}$ cm$^{-3}$, an order of magnitude higher than the estimated upper bound on $[H]_0$.

However, the consumptive model *does* fit the ⟨111⟩ results. This is illustrated in Fig. 18, which shows numerical solutions of the consumptive model, along with ⟨111⟩ kinetics at 220 and 340 °C. The agreement at these and other temperatures is excellent. Because the shape of the kinetics depends strongly on $[H]_0$, the dependence of $[H]_0$ on temperature can be extracted from the ⟨111⟩ data. In this way, we can separate the kinetics of hydrogen production at the Al-$SiO_2$ interface from the Si-$SiO_2$ interface trap anneal kinetics.

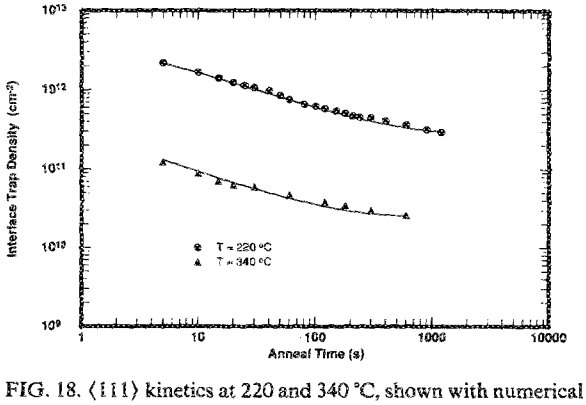


FIG. 18. ⟨111⟩ kinetics at 220 and 340 °C, shown with numerical solutions of the consumptive two reaction model.

## V. DISCUSSION

In the previous section, we showed that nonconsumptive and consumptive models of interface trap annealing are consistent with experimental kinetics on ⟨100⟩ and ⟨111⟩ interfaces, respectively. As a working hypothesis, we will assume the $P_{b0}$ defect anneals by bonding with hydrogen, but the $P_{b1}$ anneal is characterized by the nonconsumptive reaction. In this section, we will examine additional evidence that supports this hypothesis, and also review other explanations of the orientation effect. We conclude by discussing other aspects of the two reaction model.

### A. Numerical results

Quantitative values for the reaction constants and other model parameters were extracted by comparing the experimental data to numerical solutions of the two-reaction model. An interactive differential equation solver, employing a fourth-order Runge–Kutta algorithm, was developed to solve for [Si·] and [H] simultaneously. Parameter values were optimized by minimizing the squared error between simulated and experimental values of [Si·].

Interface trap production induced by high-temperature anneals (see Fig. 13) was modeled by including a production term in the differential equations. It has been shown that the elementary reaction responsible for the increase in [Si·] is[44]

$$\mathrm{SiH} \xrightarrow{k_1} \mathrm{Si}\cdot + \mathrm{H} \tag{26}$$

for the consumptive case; the corresponding reaction for the nonconsumptive mechanism is

$$\oplus \xrightarrow{k_{-1}} \mathrm{Si}\cdot . \tag{27}$$

Because there is a stoichiometric relationship between [SiH] (or [⊕]) and [Si·], these mechanisms can be incorporated into the kinetic equations without difficulty.

For a given initial trap density, four parameters are needed: the three rate constants $k_1$, $k_2$, and $k_{-1}$, and the initial hydrogen concentration $[H]_0$. This is reduced to three because $k_1$ and $k_2$ are not independent. This is illustrated in Fig. 19, which plots optimum values of $\eta (\equiv k_1/2k_2)$ as a function of temperature. The horizontal line represents a constant value of 0.55 which was used in subsequent simulations to reduce the number of parameters. At low temperatures, where thermal trap production is not observed, there are only two parameters: $k_2$ and $[H]_0$.

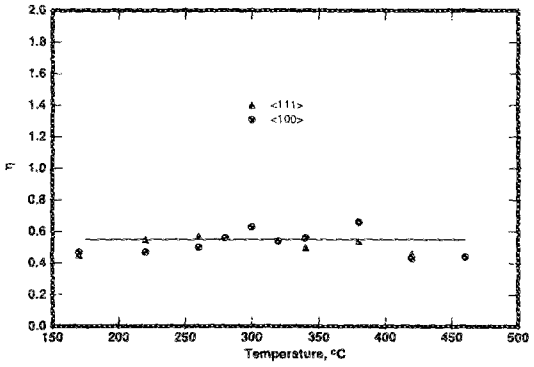


FIG. 19. Optimum values of $\eta$ for both $\langle 111\rangle$ and $\langle 100\rangle$ simulations, vs temperature. The horizontal line represents the fixed value $\eta = 0.55$ used in the simulations.

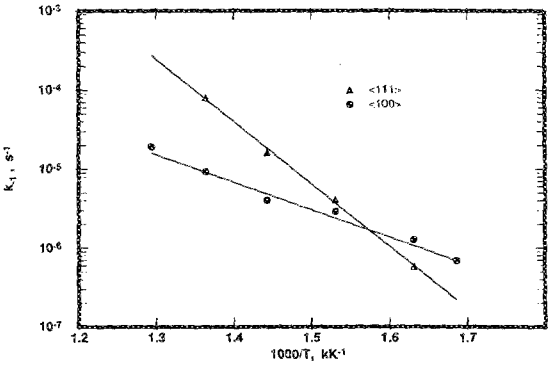


FIG. 21. Reverse anneal rate $k_{-1}$ as a function of inverse temperature. $E_A$ = 1.56 eV for $\langle 111\rangle$, 0.69 eV for $\langle 100\rangle$.

These Arrhenius expressions for the model parameters were derived:

$$k_2 = 1.3\times 10^{-12}\exp -(0.75\ \text{eV}/kT)\ \text{cm}^3\ \text{s}^{-1}, \quad (28)$$

$$[\text{H}]_0 = 3.3\times 10^{23}\exp(-0.46\ \text{eV}/kT)\ \text{cm}^{-3}, \quad (29)$$

$$k_{-1} = 4.3\times 10^{6}\exp(-1.56\ \text{eV}/kT)\ \text{s}^{-1}\langle 111\rangle, \quad (30)$$

$$k_{-1} = 5.0\times 10^{-1}\exp(-0.69\ \text{eV}/kT)\ \text{s}^{-1}\langle 100\rangle. \quad (31)$$

Equations (28) and (29) can be combined into

$$k_2[\text{H}]_0 = 4.2\times 10^{11}\exp(-1.21\ \text{eV}/kT)\ \text{s}^{-1}, \quad (32)$$

which describes the overall temperature dependence of the anneal process. Figure 20 plots $k_2[\text{H}]_0$ versus inverse temperature for both $\langle 111\rangle$ and $\langle 100\rangle$. The Arrhenius behavior described by Eq. (32) is clearly seen.

Figure 21 plots the reverse anneal rate $k_{-1}$ versus inverse temperature. Unlike the other parameters, $k_{-1}$ has a strong dependence on orientation.

The interfacial reaction constant $k_1$ is related to $k_2$ by

$$k_1 = 2k_2\eta = 1.4\times 10^{-12}\exp(-0.75\ \text{eV}/kT)\ \text{cm}^3\ \text{s}^{-1}. \quad (33)$$

This result be compared to Nishi's result[20] of 0.66 eV for the activation energy of $P_b$ center bleaching. (Because Nishi's samples did not have metal gates, we are not including the $[\text{H}]_0$ activation energy in this comparison.) These values are in reasonable agreement.

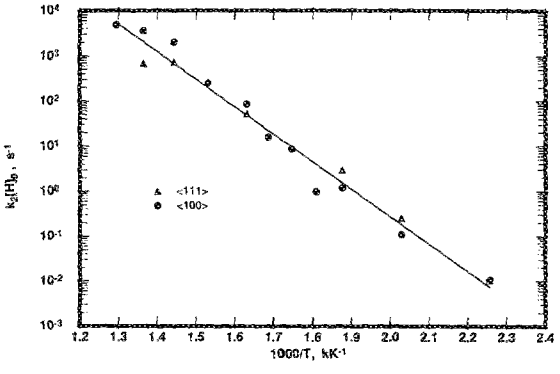


FIG. 20. Forward anneal rate $k_2[\text{H}]_0$ as a function of inverse temperature. $E_A$ = 1.21 eV.

## B. Orientation effect

To model the distinct kinetics observed on $\langle 100\rangle$ and $\langle 111\rangle$ interfaces, we have postulated different reactions involving hydrogen and the $P_{b1}$ and $P_{b0}$ defects. In a sense, this assignment is formal; there is no direct experimental evidence a consumptive reaction occurs for $P_{b0}$ and a nonconsumptive for $P_{b1}$. We have been driven to this conclusion because other explanations for the orientation effect are all lacking in some regard. In this section, we will examine these other possibilities, and will also look at evidence which supports the dual mechanism explanation.

### 1. Initial hydrogen concentration

Mathematically, the nonconsumptive model is a special case of the consumptive model, in the limit of high hydrogen concentration (see Fig. 17). Perhaps, then, the consumptive model holds for both interfaces, but there is some difference in the rate constants or $[\text{H}]_0$ which is reflected in the kinetics. The limiting cases are found in the two components of Eq. (25), the rate equation for hydrogen removal. The limit of high hydrogen concentration can be expressed as

$$[\text{Si}\cdot] \ll T_{\text{ox}}[\text{H}]/\eta. \quad (34)$$

For this explanation to work, we must therefore have one of the following:

(1) $[\text{Si}\cdot]_0$ on $\langle 100\rangle$ is much smaller than $[\text{Si}\cdot]_0$ on $\langle 111\rangle$, or

(2) $[\text{H}]_0$ on $\langle 100\rangle$ is much larger than $[\text{H}]_0$ on $\langle 111\rangle$, or

(3) $\eta$ on $\langle 100\rangle$ is much smaller than $\eta$ on $\langle 111\rangle$.

At first glance, the first possibility looks reasonable, since it is well known that interface traps on $\langle 111\rangle$ outnumber those on $\langle 100\rangle$, by a 3:1 ratio. However, this ratio applies only to oxidation-induced traps. The orientation effect persists with radiation-induced traps, where the starting concentrations are approximately equal for both orientations.

A difference in $[\text{H}]_0$ for the two interfaces could come about because of an orientation dependent surface hydroxyl concentration [see Eq. (2)]. There could conceivably be a difference in [OH] for oxides grown on $\langle 100\rangle$ and $\langle 111\rangle$ samples, because the oxide structure is in some respects controlled by the orientation of the substrate (the so-called

"template effect").[57] Since OH groups bond only to Si atoms on the $SiO_2$ surface, variations in the oxide structure could cause changes in [OH], and hence in [H]. The problem with this scenario is that at least a tenfold difference in $[H]_0$ is required for the kinetic form of the annealing to change enough to account for the orientation difference. This is far beyond the change in oxide surface structure that can likely be attributed to the template effect. Furthermore, the oxides in this study were all grown at 1000 °C, above the glass transition threshold of $\approx$960 °C.[58] This serves to lessen any differences in oxide structure originating from interface orientation.

Could $\eta$ be orientation dependent? This is certainly a possibility, since $\eta$ is partially determined by the interface trap reaction radius, which is almost assuredly orientation dependent. However, a difference in $\eta$ only would cause the ⟨111⟩ kinetics to look like a power law, like the ⟨100⟩, but with a different slope, which is not observed experimentally.

The orientation effect is therefore not related to differences in any of the parameters in inequality (34).

#### 2. Reverse reaction

In Sec. V A we listed Arrhenius expressions for the reverse rate constant $k_{-1}$, which controls the rate of trap production for extended thermal treatments. The production mechanism is also orientation dependent, which is not surprising considering the difference in interface trap structure on the two interfaces.

A question arises: can the data be described by using only one of these models but with suitable values of $k_{-1}$ for each orientation? The answer is no. To see this, we write the differential equation for the increase in [Si•], for the case when the production term exceeds the annealing term:

$$\frac{d}{dt}[\mathrm{Si}\bullet] \approx k_{-1}([\mathrm{Si}\bullet]_0 + [\mathrm{SiH}]_0 - [\mathrm{Si}\bullet]). \tag{35}$$

Trap production becomes observable only when [Si•] is negligible compared to $[Si•]_0$; therefore, we can further approximate this by

$$\frac{d}{dt}[\mathrm{Si}\bullet] \approx k_{-1}([\mathrm{Si}\bullet]_0 + [\mathrm{SiH}]_0). \tag{36}$$

The solution is

$$[\mathrm{Si}\bullet] = k_{-1}([\mathrm{Si}\bullet]_0 + [\mathrm{SiH}]_0)(t - t_r), \tag{37}$$

where $t_r$ is a characteristic time. The increase in interface traps is linear in time. On a log-log plot, the data should asymptotically approach a line with slope 1 until [Si•] approaches $([Si•]_0 + [SiH]_0)$, where the approximation that $[Si•]_0 \gg [Si•]$ breaks down.

Inspection of the data will confirm that this description is not adequate. Referring to Fig. 15, it is apparent that one data set cannot be translated into the other simply by adding a linear production term. Until the trap density reaches a minimum ($\leqslant 10^{10}\ \mathrm{cm}^{-2}\ \mathrm{eV}^{-1}$), the production term can be ignored.

#### 3. Charged hydrogen transport

Along with interface trap structure, the most definitive difference between ⟨111⟩ and ⟨100⟩ interfaces is the higher fixed charge density on ⟨111⟩. Razouk and Deal[49] have demonstrated a 3:1 ratio; this was confirmed in our experiments where the ⟨100⟩ $N_f$ was $\approx 8.0\times10^{10}\ \mathrm{cm}^{-2}$, while the ⟨111⟩ was $\approx 2.4\times10^{11}\ \mathrm{cm}^{-2}$. If the hydrogen is ionized, then the fixed charge will repel it from the interface. The higher $N_f$ on ⟨111⟩ will be more effective in repelling the hydrogen, so a slower anneal rate would be expected, which is observed experimentally.

To verify this hypothesis, we need to demonstrate the atomic hydrogen in the oxide is charged. In Sec. III C 5 we showed there is no effect on the anneal kinetics when the surface potential is varied by changing the bulk conductivity type. As an additional test of the hydrogen neutrality, we also performed anneals with the device under bias. Sufficient voltage was applied across the electrodes to double and negate the oxide field from the fixed charge. No difference in the annealed $D_{it}$ was noted with polarity reversal; thus the electrostatic differences arising from fixed charge variation on the two interfaces is not the origin of the orientation effect. (In Sec. V E, below, we will discuss *chemical* influences of the fixed charge defects on the anneal process.)

#### 4. Negative bias temperature instability

An increase in fixed charge and interface trapped charge is observed when MOS devices are heated under bias. This effect is known as the negative bias temperature instability, or NBTI.[59,60] $N_f$ and $N_{it}$ increase in roughly equal amounts, and the increase in charge is linear in applied field, although the increase is nonzero with no applied bias.[43] More charge is generated on ⟨111⟩ interfaces than on ⟨100⟩; also, the increase is related to the as-grown fixed charge density by $N_{f,\mathrm{final}} \propto (N_{f,\mathrm{initial}})^{0.72}$.[61]

Although NBTI is strongly orientation dependent, it cannot be used to explain the difference in anneal kinetics between ⟨111⟩ and ⟨100⟩. For the oxides used in this study, the increase in $D_{it}$ is given by[24,62]

$$\Delta D_{it} \propto t^{\xi}, \tag{38}$$

where $\xi \approx 0.25$. The combination of an NBTI production mechanism and either the consumptive or nonconsumptive anneal process does not describe the experimental data.

In fact, the increase in [Si•] observed for anneals above 380 °C (Fig. 13) is not due to NBTI. We have sucessfully modeled the trap production by dissociative processes which predict linear increases rather than the NBTI power-law behavior. The characteristic $t^{0.25}$ increase was seen only for anneals performed at 500 °C; NBTI trap production for unbiased anneals is insignificant at normal annealing temperatures.

None of these possibilities provide a quantitative explanation of the distinct kinetic behavior of ⟨111⟩ and ⟨100⟩ interface trap annealing. The dual mechanism hypothesis (consumptive for $P_{b0}$, nonconsumptive for $P_{b1}$) provides a framework for understanding the orientation effect; other clues which support this theory are described below.

Figure 22 shows the 340 °C kinetics for both orientations, along with the model results (solid lines). The theoretical predictions were obtained using the same parameters for both orientations; only the reaction product was

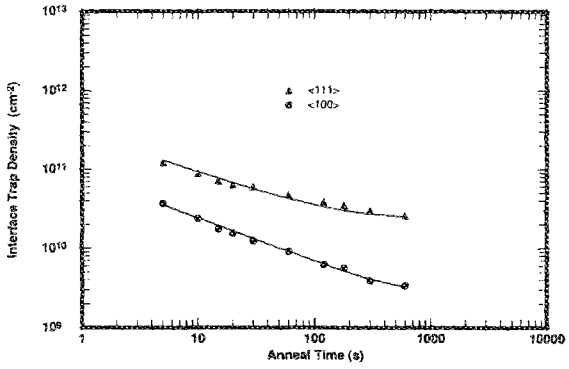


FIG. 22. ⟨111⟩ and ⟨100⟩ kinetics at 340 °C, along with consumptive and nonconsumptive simulations.

changed. The agreement between theory and experiment is striking.

Second, we observe that the temperature dependence of the anneal mechanism is independent of orientation, as shown in Fig. 20. Both the activation energy and the absolute anneal rate do not depend on orientation. This agrees with the dual mechanism hypothesis as the anneal rate, $k_2[\mathrm{H}]_0$, does not depend on any properties of the interface, but only on $SiO_2$ properties.

Finally, the rate constant $k_{-1}$, which quantifies the increase in trap density for long anneals, does depend on orientation (Fig. 21). This is to be expected, since we have postulated different structures for the annealed $P_b$ centers on ⟨111⟩ and ⟨100⟩; the difference in structure is reflected in different reverse anneal mechanisms, with different temperature dependencies (1.56 eV for ⟨111⟩, 0.69 eV for ⟨100⟩).

We have implicitly assumed that $[\mathrm{Si}\cdot] = [P_{b1}]$ on ⟨100⟩, i.e., that the $P_{b0}$ centers on the ⟨100⟩ interface can be neglected. This cannot be justified *a priori*. As mentioned previously, the ratio of $P_{b0}$ to $P_{b1}$ depends upon process conditions, but the dependence has never been studied systematically. If there are substantial numbers of $P_{b0}$ centers, the kinetics will depart from the strict power-law dependence. However, the overall kinetics cannot be modeled as the linear superposition of the individual $P_{b0}$ and $P_{b1}$ kinetics since the differential equation for [H] is nonlinear. If the individual defect concentrations at $t = 0$ were known, one could easily obtain the total $P_b$ concentration as a function of time. One could, in principle, separate the individual defect concentrations by performing ESR measurements on the unannealed samples, or by the shape of the $D_{it}$ curve across the band gap. However, we did not attempt this refinement because the departure from a strict power law by even a substantial $P_{b0}$ fraction would probably not be observable in the experimental data, given the inherent scatter, and would only make minor numerical changes to the model parameters. Therefore, the approximation of $[\mathrm{Si}\cdot] = [P_{b1}]$, while not strictly true, does not materially affect the reasoning of this section.

### C. Nonconsumptive mechanisms

In explaining the orientation effect, we have challenged the usual assumption that interface trap annealing must necessarily consume hydrogen at the interface. It is well established, though, that hydrogen is necessary for trap annealing. How can these two viewpoints be reconciled?

One possibility is that hydrogen catalyzes some sort of bond rearrangement at the interface, whereby the unattached bond of the Si• defect becomes unavailable for carrier capture. We will not speculate on the details of this process; at the present time, our understanding of $P_{b1}$ center does not even extend to the structure of the defect itself.

Another scenario: atomic hydrogen breaking Si-O-Si linkages in the $SiO_2$, releasing oxygen, which then reacts with the interface traps.[56] Nonbridging oxygen can also be generated by this process, which would be equally effective at neutralizing interface defects. Interstitial oxygen released by diffusing hydrogen is another source of oxygen. Regardless of the source, hydrogen combining with the released oxygen would form hydroxyl through

$$\mathrm{O} + \mathrm{H} \xrightarrow{k_3} \mathrm{OH}. \tag{39}$$

The familiar dimerization reaction also occurs in the oxide bulk:

$$2\mathrm{H} \xrightarrow{k_2} \mathrm{H}_2. \tag{40}$$

At the surface the hydroxyl reacts with the $P_{b1}$ center, denoted here by Si•:

$$\mathrm{Si}\cdot + \mathrm{OH} \xrightarrow{k_4} \mathrm{SiOH}. \tag{41}$$

Finally, the hydrogen dissociates,

$$\mathrm{SiOH} \xrightarrow{k_5} \mathrm{SiO} + \mathrm{H}, \tag{42}$$

leaving a trap passivated by oxygen, and the original hydrogen atom.

Ignoring for simplicity the reverse reactions, the differential equations for this four reaction model are

$$\frac{d}{dt}[\mathrm{Si}\cdot] = -k_4[\mathrm{Si}\cdot][\mathrm{OH}], \tag{43}$$

$$\frac{d}{dt}[\mathrm{H}] = -k_3[\mathrm{O}][\mathrm{H}] + \frac{k_5}{T_{ox}}[\mathrm{SiOH}] - 2k_2[\mathrm{H}]^2, \tag{44}$$

$$\frac{d}{dt}[\mathrm{O}] = -k_3[\mathrm{O}][\mathrm{H}], \tag{45}$$

$$\frac{d}{dt}[\mathrm{OH}] = k_3[\mathrm{O}][\mathrm{H}] - \frac{k_4}{T_{ox}}[\mathrm{OH}][\mathrm{Si}\cdot], \tag{46}$$

$$\frac{d}{dt}[\mathrm{SiOH}] = k_4[\mathrm{OH}][\mathrm{Si}\cdot] - k_5[\mathrm{SiOH}]. \tag{47}$$

Under the steady-state hypothesis, the reaction intermediates SiOH and OH are near equilibrium, so [SiOH]′ and [OH]′ are zero. Equations (46) and (47) can be solved to give

$$[\mathrm{OH}] = (k_3/k_4)T_{ox}([\mathrm{O}][\mathrm{H}]/[\mathrm{Si}\cdot]), \tag{48}$$

$$[\mathrm{SiOH}] = (k_3/k_5)T_{ox}[\mathrm{O}][\mathrm{H}]. \tag{49}$$

Substituting this result for [SiOH] in Eq. (44) yields

$$\frac{d}{dt}[\mathrm{H}] = -2k_2[\mathrm{H}]^2. \tag{50}$$

This equation was solved earlier; the solution is

$$[H] = [H]_0/(1 + 2k_2[H]_0 t). \tag{51}$$

This result is used in Eq. (45) to solve for [O]:

$$[O] = [O]_0/(1 + 2k_2[H]_0 t)^{\eta'}, \tag{52}$$

where $\eta' = (k_3/2k_2)$. Noting that, stoichiometrically, $\Delta[Si\cdot] = T_{ox}\Delta[O]$, the final solution for [Si·] is

$$[Si\cdot] = [Si\cdot]_0 - T_{ox}[O]_0 + [T_{ox}[O]_0/(1 + 2k_2[H]_0 t)^{\eta'}]. \tag{53}$$

The four reaction model, with hydrogen liberated oxygen in the form of OH as the annealing agent, predicts the observed power-law kinetics.

Another possibility is that hydrogen releases oxygen from the network, but does not form OH. The reactions would then be

$$(O_3{\equiv}Si{-}O) + H \xrightarrow{k_6} (O_3{\equiv}Si) + O + H, \tag{54}$$

$$2H \xrightarrow{k_7} H_2, \tag{55}$$

$$Si\cdot + O \xrightarrow{k_7} SiO, \tag{56}$$

where we have written the oxygen source as nonbridging oxygen, but could just as well be network forming oxygen. It is a straightforward exercise to show

$$[Si\cdot] = [Si\cdot]_0 - T_{ox}[O_3{\equiv}Si{-}O]_0 + \frac{T_{ox}[O_3{\equiv}Si{-}O]_0}{(1 + 2k_2[H]_0 t)^{\eta''}}, \tag{57}$$

where $\eta'' = (k_6/2k_2)$. Again we have a power-law kinetic form predicted by a specific nonconsumptive mechanism.

From the kinetic data there is no way to decide among these or other competing hypotheses for $P_{b1}$ anneal mechanisms. Interestingly, diffusion of OH is the rate limiting step in the production of interface traps by the negative bias temperature instability, according to one model.[62] A complete model of trap annealing would have to account for this phenomenon as well. Without more detailed knowledge of the $P_{b1}$ structure, definitive annealing mechanisms cannot be evaluated adequately. However, it is clear that the hydrogen can be a necessary factor in trap annealing, even without the formation of interfacial SiH.[63]

### D. Hydrogen diffusion

The diffusion coefficient of atomic hydrogen in $SiO_2$ can be found from the dimerization rate constant $k_2$. We find an activation energy of 0.75 eV, higher than the 0.3 eV which is often quoted,[64] but certainly within the range of 0.07–0.92 eV found in the literature.[45,65]

ESR studies of thermal oxide[10] and fused quartz[65] irradiated at cryogenic temperatures show a large concentration of atomic hydrogen, which disappears upon warming at approximately 130 K. By observing the ESR signals of other defects, Griscom[46] has concluded that most of the atomic hydrogen is dimerizing into $H_2$. From the isochronal anneal data of Brower *et al.*,[10] the diffusion coefficient of H was calculated to have an energy of 0.18 eV. Based on these results, the dominant mobile hydrogenic species in $SiO_2$ is assumed to be molecular, rather than atomic, hydrogen at room temperature and above. The dimerization is so rapid above 130 K that molecular and atomic hydrogen are in equilibrium, which favors the molecular over the atomic species. This conflicts with the central hypothesis of this work, that hydrogen dimerization occurs at processing temperatures at a measurable rate, so that molecular and atomic species are not in equilibrium.

There appears to be no straightforward way to reconcile these two viewpoints. A large body of experimental evidence indicates that atomic, not molecular, hydrogen is responsible for annealing traps in metal gate MOS structures. Additionally, we have been unable to model the power-law temporal dependence (or the slightly more complicated ⟨111⟩ kinetics) of [Si·] with mechanisms involving molecular hydrogen.

As a possible resolution of this dilemma, we should consider the nature of the $SiO_2$ irradiation process, which might influence the hydrogen transport. The disappearance of the hydrogen ESR signal is observed in samples irradiated by 100-keV x rays[66] and $^{60}Co$ $\gamma$ rays ($\approx$1 MeV).[10] In both these cases the $SiO_2$ was subjected to energies much higher than those used in our experiments—only 6 keV from the electron beam evaporator. It is not unreasonable to suppose the higher energies, which produce more extensive oxide damage, are creating low-energy diffusion pathways which will greatly accelerate the hydrogen dimerization. The 0.18 eV energy seen in highly damaged oxides may be characteristic of an extrinsic diffusion mechanism created by correlated x-ray and $\gamma$-ray damage tracks. In contrast, the less severe damage inflicted by the evaporation process allows hydrogen to diffuse at more of an intrinsic rate.

In a study of various radiation-induced defect annealing experiments, Brown *et al.*[54] showed that a large number of isochronal anneal investigations in the literature fall into two broad groups, where the difference between the two is related to the $SiO_2$ OH content. Our experimental results do not fall into either of these groups. The diffusion energy seen in this and other studies may thus be related to the oxide damage, or lack thereof, and comparisons to other results should take this into account.

### E. Influence of fixed charge defects

We have explained the orientation effect by suggesting different anneal mechanisms for $P_{b0}$ and $P_{b1}$ defects. Earlier, we described an experiment to determine if the electrostatic influence of fixed oxide charge on an ionized $H^+$ specie could account for the different kinetics observed on ⟨111⟩ and ⟨100⟩. Although we found no evidence for ionized hydrogen transport, it is not clear that $Q_f$ has no role in explaining the orientation effect. Perhaps the defect responsible for the fixed charge exerts a *chemical*, rather than an electrical influence on the annealing mechanism. This is plausible given the general correlation of $N_{it}$ and $N_f$ (Ref. 49) which probably represents a spatial correspondence of the two defects. The 3:1 ⟨111⟩/⟨100⟩ ratio of $Q_f$ might then explain the difference in anneal kinetics.

One way to check this is to prepare ⟨100⟩ samples with

large amounts of fixed charge, and see if the anneal kinetics follow ⟨100⟩ or ⟨111⟩ behavior. This experiment was performed by heating ⟨100⟩ interfaces in dry oxygen for several hours at 700 °C. This increased $N_f$ to $\approx 2.8\times10^{11}$ cm$^{-2}$, an increase of 350% (and about 15% higher than the ⟨111⟩). The anneal kinetics, measured at 260, 300, and 380 °C resembled ⟨111⟩ kinetics, suggesting the $Q_f$ defect has a strong effect on the anneal kinetics.

Or does it? In fact, this experiment is inconclusive, which is why it was not described earlier. The problem is that the process used to increase $Q_f$ also changes the $P_{b1}/P_{b0}$ ratio in an undetermined manner, and also changes the oxide structure and hydrogen concentration due to the excess oxygen present in the $SiO_2$. If the high-temperature oxygen anneal increases $[P_{b0}]$, one would expect to see ⟨111⟩-like kinetics, but the additional fixed charge would not necessarily be involved. Unfortunately, there is no known way to decrease the ⟨111⟩ $N_f$ to ⟨100⟩ levels, so this point remains unresolved.

## VI. SUMMARY

We have performed an extensive study of Si-$SiO_2$ interface trap anneal kinetics by rapid thermal processing of MOS capacitors. The reaction kinetics are accurately described by a two reaction model, consisting of an interface reaction between the traps and atomic hydrogen, and hydrogen dimerization in the $SiO_2$. Quantitative estimates of the reaction coefficients have been obtained through numerical simulations of the two reaction model.

The anneal chemistry of defects on ⟨111⟩ and ⟨100⟩ interfaces is significantly different. We have considered a variety of explanations for this orientation effect; the most successful involves a consumptive mechanism for the $P_{b0}$ defect, forming SiH, and a nonconsumptive mechanism for the $P_{b1}$ center. This hypothesis is supported by rate measurements of thermally induced trap production.

The two reaction model is the simplest scheme that reproduces the experimentally observed kinetics. We have firmly established that this model is to be preferred over the single bimolecular surface reaction model. We consider our explanation of the orientation differences to be plausible and consistent, but it is somewhat tentative since there is no direct experimental evidence for either the consumptive or nonconsumptive reactions. Progress in instrumentation to measure interfacial SiH concentration will provide a meaningful test of this hypothesis.

## ACKNOWLEDGMENTS

The authors are pleased to acknowledge many helpful suggestions from P. Griffin, B. Fishbein, J. Watt, D. Brown, E. Nicollian, and E. Poindexter. We also thank the staff and students of the Stanford Integrated Circuits Laboratory for their help. This work was supported by the Defense Advanced Research Projects Agency and the Semiconductor Research Corporation. M. Reed was supported by a generous fellowship from the Fannie and John Hertz Foundation, which is gratefully acknowledged.

[1]E. H. Poindexter and P. J. Caplan, Prog. Surf. Sci. **14**, 201 (1983).
[2]P. Balk, in *Proceedings of the Electrochemical Society Fall Meeting* (Electrochemical Society, Buffalo, NY, 1965), p. 29.
[3]G. Schols and H. E. Maes, in *Symposium on Silicon Nitride Thin Insulating Films* (Electrochemical Society, Princeton, NJ, 1983), Vol. 83-8, p. 94.
[4]E. H. Nicollian, in *Semiconductor Silicon*, edited by H. R. Huff, T. Abe, and B. Kolbesen (Electrochemical Society, Princeton, NJ, 1986), p. 437.
[5]S. Kar and W. E. Dahlke, Solid-State Electron **15**, 221 (1972).
[6]Y. C. Cheng, Prog. Surf. Sci. **8**, 131 (1977).
[7]E. H. Poindexter, P. J. Caplan, B. E. Deal, and R. R. Razouk, J. Appl. Phys. **52**, 879 (1981).
[8]G. J. Gerardi, E. H. Poindexter, P. Caplan, and N. M. Johnson, Appl. Phys. Lett. **49**, 348 (1986).
[9]P. M. Lenahan and P. V. Dressendorfer, J. Appl. Phys. **54**, 1457 (1983).
[10]K. L. Brower, P. M. Lenahan, and P. V. Dressendorfer, Appl. Phys. Lett. **41**, 251 (1982).
[11]K. L. Brower, Appl. Phys. Lett. **43**, 1111 (1983).
[12]N. M. Johnson, D. K. Biegelsen, M. D. Moyer, S. T. Chang, E. H. Poindexter, and P. J. Caplan, Appl. Phys. Lett. **43**, 563 (1983).
[13]P. M. Lenahan and P. V. Dressendorfer, J. Appl. Phys. **55**, 3495 (1984).
[14]E. H. Poindexter, G. J. Gerardi, M.-E. Rueckel, P. J. Caplan, N. M. Johnson, and D. K. Biegelsen, J. Appl. Phys. **56**, 2844 (1984).
[15]F. Herman and R. V. Kasowski, J. Vac. Sci. Technol. **19**, 395 (1981).
[16]R. Castagne and A. Vapaille, Surf. Sci. **28**, 157 (1971).
[17]A. Stesmans, Appl. Phys. Lett. **48**, 972 (1986).
[18]A. Edwards (unpublished results).
[19]E. H. Poindexter (unpublished results).
[20]Y. Nishi, Jpn. J. Appl. Phys. **10**, 52 (1971).
[21]S. T. Chang, J. K. Wu, and S. A. Lyon, Appl. Phys. Lett. **48**, 662 (1986).
[22]W. Eades, Ph. D. thesis (Stanford University, 1985).
[23]N. Shiono, M. Shimaya, and O. Nakajima, Appl. Phys. Lett. **48**, 1129 (1986).
[24]N. Shiono, O. Nakajima, and C. Hashimoto, J. Electrochem. Soc. **130**, 138 (1983).
[25]P. L. Castro and B. E. Deal, J. Electrochem. Soc. **118**, 280 (1971).
[26]B. E. Deal, E. L. MacKenna, and P. L. Castro, J. Electrochem. Soc. **116**, 997 (1969).
[27]S. Kar, Solid-State Electron **18**, 723 (1975).
[28]G. Schols, H. E. Maes, G. Declerck, and R. Van Overstraeten, Rev. Phys. Appl. **13**, 825 (1978).
[29]G. A. Schols, H. E. Maes, and R. J. Van Overstraeten, J. Appl. Phys. **51**, 3194 (1980).
[30]T. W. Hickmott, J. Appl. Phys. **48**, 723 (1977).
[31]T. W. Hickmott, J. Appl. Phys. **51**, 4269 (1980).
[32]Y. T. Yeow, D. R. Lamb, and S. D. Brotherton, J. Phys. D **8**, 1495 (1975).
[33]P. Balk and N. Klein, Thin Solid Films **89**, 329 (1982).
[34]N. M. Johnson, D. K. Biegelsen, and M. D. Moyer, J. Vac. Sci. Technol. **19**, 390 (1981).
[35]N. M. Johnson, D. K. Biegelsen, M. D. Moyer, V. R. Deline, and C. A. Evans, Jr., Appl. Phys. Lett. **38**, 995 (1981).
[36]N. S. Saks, IEDM Tech. Dig. **710** (1981).
[37]B. J. Fishbein, J. T. Watt, and J. D. Plummer, J. Electrochem. Soc. **134**, 674 (1987).
[38]R. O. Lussow, J. Electrochem. Soc. **115**, 660 (1968).
[39]J. A. Cunningham, L. E. Sharif, and S. S. Baird, J. Electrochem. Soc. **1**, 242 (1963).
[40]N. M. Johnson, Appl. Phys. Lett. **47**, 874 (1985).
[41]S. Dadgar, Technical Report T86012, Semiconductor Research Corporation, P. O. Box 12053, Research Triangle Park, NC 27709 (1986).
[42]AG Associates, 1325 Borregas Ave., Sunnyvale, CA.
[43]B. E. Deal, M. Sklar, A. S. Grove, and E. H. Snow, J. Electrochem. Soc. **114**, 266 (1967).
[44]M. L. Reed and J. D. Plummer, Appl. Phys. Lett. **51**, 514 (1987).
[45]F. B. McLean, IEEE Trans. Nucl. Sci. **NS-27**, 1651 (1980).
[46]D. L. Griscom, J. Appl. Phys. **58**, 2524 (1985).
[47]E. S. Schlegel and G. L. Schnable, J. Electrochem. Soc. **119**, 165 (1972).
[48]For the doping levels in this experiment, a significant change in the kinetics would arise from the change in surface Fermi level at trap densities below approximately $2\times10^{11}$ cm$^{-2}$ eV$^{-1}$. Above this level, surface pinning by the interface traps will dominate. Since there is no abrupt change in the kinetics near this value, electrochemical effects do not appear to be important in the anneal process. In Sec. V B 3, we consider another approach to this question.
[49]R. R. Razouk and B. E. Deal, J. Electrochem. Soc. **126**, 1573 (1979).

[50]B. E. Deal (private communication).
[51]T. R. Waite, Phys. Rev. **107**, 463 (1957).
[52]R. M. Noyes, Prog. React. Kin. **1**, 129 (1961).
[53]M. L. Reed, Ph. D. thesis (Stanford University, 1987).
[54]D. B. Brown, D. I. Ma, C. M. Dozier, and M. C. Peckerar, IEEE Trans. Nucl. Sci. **NS-30**, 4059 (1983).
[55]M. L. Reed and J. D. Plummer, IEEE Trans. Nucl. Sci. **NS-33**, 1198 (1986).
[56]G. A. Ruggles, R. Koba, and R. E. Tressler, J. Electrochem. Soc. **133**, 2549 (1986).
[57]E. A. Taft, J. Electrochem. Soc. **125**, 968 (1978).
[58]D. Chin, S-Y. Oh, and R. W. Dutton, IEEE Trans. Electron Devices **ED-30**, 993 (1983).
[59]D. J. Breed, Appl. Phys. Lett. **26**, 116 (1975).
[60]D. J. Breed and R. P. Kramer, Solid-State Electron **19**, 897 (1976).
[61]A. Goetzberger, A. D. Lopez, and R. J. Strain, J. Electrochem. Soc. **120**, 90 (1972).
[62]K. O. Jeppson and C. M. Svensson, J. Appl. Phys. **48**, 2004 (1977).
[63]One might ask if interface traps could be passivated by a hydroxyl group, leaving a structure like $Si \equiv Si–OH$ at the interface. To our knowledge, this possibility has not been considered in the literature. The much higher anneal rate of metal gate devices (compared to polysilicon gates) suggests hydrogen, which diffuses faster than hydroxyl in $SiO_2$, is a more likely candidate for defect passivation. We cannot, however, rule out the possibility of OH bonding at the interface with these experiments.
[64]A. G. Revesz, J. Electrochem. Soc. **126**, 122 (1979).
[65]D. L. Griscom, J. Non-Cryst. Solids **68**, 301 (1984).
[66]D. L. Griscom, M. Stapelbroek, and E. J. Friebele, J. Chem. Phys. **78**, 1638 (1983).