Effects of Post-Stress Hydrogen Annealing on MOS Oxides after $^{60}$Co Irradiation or Fowler-Nordheim Injection

Code 6810, Naval Research Laboratory
Washington, D.C. 20375
(* SFA, Inc., Landover, MD)

Abstract

Changes in interface trap density $D_{it}$ have been determined in MOSFETs as a function of time during hydrogen annealing at 295K. Large increases in $D_{it}$ are observed during $H_2$ annealing in MOSFETs previously stressed by either $^{60}$Co irradiation or Fowler-Nordheim electron injection. The annealing behavior is very similar for both types of stress, which suggests that the $D_{it}$ creation mechanism involves similar chemistry for hydrogen reactions. Studies of the time dependence of $D_{it}$ creation as a function of MOSFET gate length show that the time dependence is limited primarily by lateral diffusion of molecular hydrogen ($H_2$) through the gate oxide. An activation energy of 0.57 eV, which is consistent with $H_2$ diffusion, is obtained from the temperature dependence.

INTRODUCTION

When MOS oxides are exposed to various types of stress (irradiation, hot electron or hot hole injection, Fowler-Nordheim injection, etc.), the oxide is damaged by the creation of positive oxide charge $Q_o$ and interface traps $D_{it}$. It is widely believed that the creation mechanisms are similar or identical for many types of stress, although this has yet to be shown in a rigorous way. Recent research on irradiated oxides has shown that hydrogen is clearly involved in $D_{it}$ creation, and the chemistry of this process is believed to be reasonably well understood [1-5]. However, the mechanism of $D_{it}$ creation by other types of stress is less clear and may be somewhat more complex. For example, studies of $D_{it}$ creation during electron injection suggest that $D_{it}$ creation by recombination of electrons with trapped holes [6,7] is important in addition to hydrogen related mechanisms.

In an effort to elucidate the role that hydrogen may play in $D_{it}$ creation, experiments have been performed wherein MOS oxides are first irradiated and then exposed to hydrogen ($H_2$) gas at room temperature [8-14]. In these experiments, large increases in $D_{it}$ are observed during the $H_2$ exposure; for example, $\Delta D_{it}$ during $H_2$ anneal is typically as large as, or larger than, $\Delta D_{it}$ due to irradiation alone. Stahlbush et al. [9] have proposed the following mechanism for this effect: (1) Irradiation creates positive charge in the oxide; (2) During the subsequent $H_2$ anneal, the positive charge reacts with $H_2$ to create $H^+$ ions; (3) the $H^+$ ions drift to the Si-SiO$_2$ interface where they react to break Si-H bonds and produce interface traps. In this model, the hydrogen chemistry of trap creation is identical to that which occurs during irradiation itself [1-5]. $D_{it}$ creation during the $H_2$ anneal occurs over long times from 1 to $>10^6$ s at 295K. Stahlbush and co-workers have proposed that this slow time dependence arises from the slow cracking reaction of $H_2$ molecules at positively charged defects [10,11]. Alternatively, Mrstik and co-workers have proposed that the rate-limiting step arises from the diffusion of $H_2$ through the oxide, prior to the $H_2$ cracking reaction [12,13].

In this work, we have measured $D_{it}$ creation during $H_2$ annealing. This work has three goals: (1) $H_2$ annealing measurements have been performed following different types of oxide stress: irradiation, Fowler-Nordheim (FN) electron injection, and channel hot carrier (CHC) degradation. Whether the results obtained for different types of stress are the same or significantly different could help determine if the $D_{it}$ creation mechanisms for the different stress types are the same, or different. Here we will report results for irradiation and FN stress, both of which create uniform damage in the oxide. Results for CHC damage, which creates non-uniform damage in a region highly localized near the MOSFET drain, are reported elsewhere [15]. We find many similarities in these experiments among the different stresses, but there are also a few instructive differences. (2) This research was undertaken in part to obtain some insight into the chemistry of hydrogen reactions, and to try to resolve the different interpretations of the cause of the time-dependent $D_{it}$ build-up during $H_2$ anneal (discussed above). Our studies of the time-dependence, gate-length dependence, and temperature dependence show that the time dependence is determined primarily (but not exclusively) by the diffusion of $H_2$ molecules to reaction sites in the gate oxide. Thus, unfortunately, these data do not help much in understanding the hydrogen chemistry involved in $D_{it}$ creation. (On the other hand, these data may yield useful information about the nature of the $H_2$ diffusion process in gate oxides. Indeed, Mrstik et al. [13] have used similar data to show that the $H_2$ diffusion rate is dependent on MOS oxide thickness and is a factor of 10-100 slower than obtained for bulk silica glass. This result implies that thin gate oxides are slightly densified compared to bulk SiO$_2$.) We also identify two cases where the time dependence is not due to $H_2$ diffusion and should be studied further. (3) There are several recent reports of unexpected increases in $D_{it}$ after irradiation [16] or hot carrier stress [17].

U.S. Government work not protected by U.S. copyright
Both reports propose that hydrogen-related mechanisms could be responsible for these effects. In this work, hydrogen is deliberately introduced into the oxides to see if effects similar to refs. [16] and [17] are observed.

SAMPLE FABRICATION AND EXPERIMENTAL TECHNIQUES

N-channel MOS transistors were fabricated at NRL with LOCOS field isolation and phosphorus-doped polysilicon gates. Gate oxides were grown in dry O₂ at 900°C to a thickness of 25.4 nm. A few similar results were obtained on N-FETs with a 19 nm dry oxide, and p-FETs with 35 nm dry oxides. All samples were fabricated without inter-metal-level oxides or a final passivation layer; these layers would probably interfere with diffusion of hydrogen into the gate oxide during the hydrogen annealing experiments. Irradiations were performed with a 60Co source with +2.0 MV/cm applied to the MOSFET gates. FN electron injection stresses were performed with about +8 MV/cm applied to the MOSFET gate for 5-10 s to obtain an integrated injected charge of 3-5 x 10¹⁶ e/cm². Interface trap densities Dₜ were measured using charge pumping (CP) on (typically) 10 µm gate length MOSFETs. The CP frequency was 1.5 x 10⁷ Hz with 1 μs rise/fall times and 3-4V ac pulse heights. Using these CP parameters, the Dₜ values given here are midgap values averaged over approximately the central 0.6 eV of the silicon band gap. Initial Dₜ values for unstressed FETs are typically ~1.0 x 10¹⁰ traps/cm²-eV.

RESULTS AND DISCUSSION

A. Anneal time and gate bias dependence

Charge pumping results for a typical measurement sequence (virgin, post-irradiation, and during H₂ anneal) are shown in Fig. 1. The gate oxide was biased at +2 MV/cm during irradiation and H₂ anneal. The magnitude of the CP current Iₛ is proportional to Dₜ in the MOSFET, while the voltages at which Iₛ changes from zero to the maximum and vice versa are related to the MOSFET flatband and threshold voltages. After irradiation, the magnitude of Iₛ increases, indicating more than an order of magnitude increase in Dₜ, while the negative shift of the curves indicates a buildup of positive oxide charge Qₒ. After irradiation, the samples were annealed in 100% H₂ at 295K for up to several weeks. At least 12 hours was allowed to lapse between irradiation and the H₂ anneal to allow any time-dependent changes after irradiation to occur. The increase in Dₜ during the H₂ anneal was measured with CP as a function of anneal time. In general, CP measurements were taken at a single fixed value of Vₒ, where Iₛ is flat and independent of Vₒ, rather than obtaining the entire Iₛ curve. This approach is adequate because changes in Qₒ are small during the H₂ anneal.

Typical data showing the time dependent increase in Iₛ during exposure to H₂ at 295K following 1 Mrad(Si) irradiation are shown in Fig. 1. Irradiation causes a large increase in Iₛ, corresponding to an increase in Dₜ from 0.11 to 1.7 x 10¹⁶ traps/cm²-eV. A subsequent exposure to H₂ then causes a dramatic further increase in Dₜ. Note that this is not a small effect; ADₜ due to a 24 hr anneal in H₂ is slightly larger than ADₜ due to the irradiation alone. Other annealing experiments in N₂ demonstrate that the time dependent increase in Dₜ is negligible unless H₂ is present. These results make it clear that, at least in our samples, the Dit increase in Fig. 1 is not just due to post-irradiation time dependence, as described by Bhat and Vasi [18]. There is a slight movement of the CP curves towards positive Vₒ as the H₂ anneal proceeds, which is due to a small reduction in the radiation-induced positive Qₒ.

Dₜ as a function of time during H₂ annealing is shown in Fig. 2. These results were obtained after FN annealing, rather than irradiation, and are qualitatively similar for both types of stress (as discussed below). Results are shown for two different MOSFETs from the same chip which were FN-stressed identically (5 x 10¹⁶ e/cm² at +19V) but biased differently during the anneal. The initial increase in Dₜ after the FN-stress is about the same in both samples. At long H₂ anneal times, the increase in Dₜ due the anneal is large and about equal to the FN-stress induced increase. In general, ADₜ during H₂ anneal is proportional to ADₜ during the stress; this holds for all types of stress (irradiation, FN-stress, and CHC stress). Results of H₂ annealing after FN stress are generally quite similar to H₂ annealing after irradiation. Different gate oxide fields (+2 MV/cm) were applied to the two FETs during the H₂ anneal. As shown in Fig. 2, we find that: (1) ΔDₜ is smaller for negative gate bias, and (2) ΔDₜ tends to saturate at very long times with negative bias, but not with positive bias. At positive gate bias, Dₜ at very long
Figure 2. Increase in interface trap density \( D_i \) as a function of \( H_2 \) anneal time for two samples stressed by Fowler-Nordheim electron injection of \( -5 \times 10^{16} \text{e}/\text{cm}^2 \). For positive \( V_g \) during anneal (+5.1 V), \( \Delta D_i \) is larger, and a clear saturation is not obtained at long anneal times.

B. Gate length dependence

In an attempt to determine the rate-limiting step of the time dependence shown in Figs. 1 and 2, we have measured the time dependence as a function of the MOSFET gate length \( L_g \). If the rate-limiting step is determined by \( H_2 \) diffusion into the gate oxide as proposed by Mrstik [12,13], then the time dependencies should be a strong function of the gate geometry. Experiments were performed on a FET array of devices with the same gate width \( W_g \) (29 \( \mu \)m) but different gate lengths \( L_g \) (2 \( \leq L_g \leq 24 \mu \)m). (Values for \( L_g \) were measured on a neighboring chip using a scanning electron microscope with an accuracy of \( \pm 0.1 \mu \)m.) The FETs were irradiated to 1.0 Mrad(Si) and then annealed in \( H_2 \) at \( V_g = -5.1 \text{V} \). A negative gate bias was chosen to obtain the best saturation characteristic -- see Fig. 2. Fig. 4 shows a plot of the increase in the number of interface traps \( N_t \) as a function of \( H_2 \) anneal time on a log-log scale (note that \( N_t \) is not scaled to device area).

Results in Fig. 4 show that the increase in \( N_t \) is independent of gate length at small times (less than about 3000 s), but is strongly dependent on gate length at long times. For short anneal times, \( \Delta N_t \) increases at about the same rate in all FETs. As time increases, saturation is achieved, with the shortest length \( L_g = 2 \mu \text{m} \) FET saturating at the shortest time. Similar (less extensive) behavior was reported by Mrstik et al. [13]. This behavior is readily understood if one considers an (abrupt) \( H_2 \) diffusion front moving into the gate oxide from the edges of the gate at the MOSFET source and drain (shown schematically in Fig. 5). At first, the amount of diffusion is identical in all devices, independent of gate length, until the two diffusion fronts meet at the center of the gate when saturation is achieved. FETs with longer \( L_g \)'s saturate at longer times because of their longer diffusion paths. For this model to be correct, "lateral" \( H_2 \) diffusion through the gate oxide must be very much faster than "vertical" diffusion through either the relatively thin poly-crystalline silicon (p-Si) gate on top, or through the underlying crystalline silicon (c-Si) substrate (see Fig. 5). Diffusion of \( H_2 \) through the c-Si...
Figure 4. Number of traps created during H₂ anneal in FETs with different gate lengths (2-24 μm) and the same gate width (29 μm). Samples were irradiated to 1 Mrad(Si) dose before the H₂ anneal. The N₄ growth rate is initially the same in all FETs until saturation is achieved in the shortest gate length FET. Each longer FET then saturates at a higher N₄ level. This behavior is consistent with lateral H₂ diffusion from the edges of the gates towards the center.

Figure 5. Schematic diagram (not to scale) of proposed "lateral" H₂ diffusion through the gate oxide from the edges of the poly gate towards the center.

Substrate can clearly be ignored because (a) it is much thicker (~300 μm) than any gate oxide dimension, and (b) it has a much smaller H₂ diffusion coefficient (~10⁻¹⁵ cm²/s at 295K [19]) compared to the diffusion coefficient in bulk silica glass (~10⁻¹¹ cm²/s) [20]. However, it is not clear that H₂ diffusion through the p-Si gate can be so easily ignored because (a) the p-Si gate is relatively thin (~0.5 μm), and (b) H₂ diffusion in p-Si may be enhanced by diffusion along the poly-crystalline grain boundaries. However, a recent study of H₂ diffusion in poly thin film transistors (TFTs) [21] concludes that H₂ diffusion is actually retarded at the grain boundaries, and that H₂ diffusion in TFTs occurs primarily by "lateral" diffusion through the gate oxide as shown in Fig. 4.

The dependence on gate length is summarized in Fig. 6, which shows the Dₙ formation time τₗ₁ (≈ time to create 1/2 the final ΔDₙ) as a function of gate length. Results are essentially identical for all devices whether irradiated (data from Fig. 4) or FN-stressed. τₗ₁ increases rapidly with gate length approximately as Lₙ₁.5. Lines show calculated dependencies for 1-dimensional and 2-dimensional models based on H₂ diffusion (see text).

A simple model for the time dependence can be developed assuming a one-dimensional H₂ diffusion model. We assume the H₂ diffusion can be characterized by a sharply defined diffusion front [13,22]. Throughout the diffused region, the H₂ concentration is sufficiently high that the H reactions proceed immediately to completion. In simple diffusion processes, the motion of the diffusion front is
Figure 7. Schematic diagram of 1-dimensional ("1-D") and 2-dimensional ("2-D") \( \text{H}_2 \) diffusion models. The 2-D model accounts for \( \text{H}_2 \) diffusion along all four sides of the MOSFET.

proportional to \( \sqrt{(\text{time})} \), which implies \( \tau_{\text{rel}} \propto L^2 \). Mrstik et al. [13] have reported just such a \( \sqrt{(\text{time})} \) dependence. We also find straight lines if the data are plotted with a \( \sqrt{(\text{time})} \) abscissa and \( \Delta \text{Dit} \) on a linear scale (not shown). However, Fig. 4 does not show \( \sqrt{(\text{time})} \) dependence because this log-log plot accentuates the region where \( \Delta \text{Dit} \) is very small and the diffusion front is probably not sufficiently abrupt compared to the short diffusion distance. In Fig. 6, the predicted \( L^2 \) dependence (solid line) is compared with experimental data. \( L^2 \) increases much more rapidly than is observed. This discrepancy is somewhat resolved by using a 2-dimensional model which accounts for diffusion from all 4 sides of the poly gate, which must be considered when the gate length is approximately the same size as the gate width (see Fig. 7). A 2-dimensional model is derived in Appendix A, which gives the relative change in \( \Delta \text{Dit} \) based solely on FET geometry (i.e., \( L \) and \( W \)). This model is fit to the data at \( L < < W \) where the 2-D model reduces to a dependence on \( L^2 \) as expected. For \( L < W = 29 \mu \text{m} \), the 2-D model shows that \( \tau_{\text{rel}} \) is smaller than expected from the 1-D model because additional diffusion in the 2-D model causes saturation to be achieved more rapidly. Agreement between the 2-D model and the data in Fig. 6 is much improved. However, agreement is still less than satisfactory, especially at \( L > 6 \mu \text{m} \) where the slope should approach \( L^2 \) in both models, but does not. The reason for the remaining discrepancy is unclear. One source of error is that the FET source and drain are overlapped by \(-0.2 \mu \text{m} \) at the edges of the poly gate. Because of the heavy source/drain doping, \( \Delta \text{Dit} \) cannot be measured in the overlap region which introduces an error into the value for \( \tau_{\text{rel}} \) (whose real value will be smaller than measured). The assumption of a sharp diffusion edge may not be adequate in the shortest \( L \) devices. Finally, the time dependence of other sequential steps (such as diffusion of \( \text{H}_2 \) through the 0.5 \( \mu \text{m} \) thick field oxide, or perhaps some hydrogen reactions, etc.), may not be negligibly small in the smallest \( L \) FETs as assumed.

C. Temperature dependence

To further investigate the time dependence of \( \Delta \text{Dit} \) creation during \( \text{H}_2 \) annealing, the temperature dependence has been determined. Different FETs (\( L = 1.5 \mu \text{m} \)) were FN-injected to \(-5 \times 10^{16} \text{e/cm}^2\). \( \text{H}_2 \) anneals were performed with \( V_\text{g} = -5.1 \text{V} \) at temperatures in the range 278-360K. The resulting normalized increase in \( \Delta \text{Dit} \) is shown in Fig. 8. The process is clearly strongly temperature activated. An Arrhenius plot of the creation time \( \tau_{\text{rel}} \) vs. \( 1/T \) (K) obtained from the data in Fig. 8 is shown in Fig. 9. Each data point represents data from 2-3 FETs whose \( \tau \) values typically agreed within \( \pm 10\% \). An activation energy \( E_\text{a} \) of 0.57 eV is obtained from a least-squares fit to these data. Activation energies of 0.45 eV [20] and 0.47 eV [23] have been measured for \( \text{H}_2 \) diffusivity in bulk silica glass near room temperature. A 0.45 eV activation energy (matched at 295K) is shown in Fig. 9 for comparison. The difference in these activation energies represents a factor of 3.0 difference in \( \tau \) over the measured temperature range. Although the quality of the data in Fig. 8 is clearly somewhat imperfect (note the abrupt small changes at normalized \( \Delta \text{Dit} = 0.9 \)), it seems unlikely that errors in determining \( \tau \) could lead to a (relative) factor of 3 error.

Next, experimental values of \( \tau_{\text{rel}} \) are compared with calculations from a simple \( \text{H}_2 \) diffusion model. An approximate diffusion length \( x \) may be calculated from \( x = \sqrt{(D \Delta t)} \), where \( D \) is the \( \text{H}_2 \) diffusivity \((=1.4 \times 10^{-11} \text{cm}^2/\text{s})[20]\). For a 2 \( \mu \text{m} \) FET, \( x = 0.5 \mu \text{m} \) when the diffusion front has penetrated halfway from the gate edge to the center (note that diffusion occurs from both sides of the gate). Thus, time =
Figure 9. Arrhenius plot of $D_n$ formation time $t_{1/2}$ versus $1000/T$. The experimental activation energy is 0.57 eV. Also shown for comparison (dashed line) is the activation energy for $H_2$ diffusion in silica glass (0.45 eV).

Experimental $t_{1/2} = 1.9 \times 10^3$ s (2.0 µm FET, Fig. 6) which is a factor of about 10 larger than predicted by the diffusion model. Mrstik et al. [13] have compared experimental $D_n$ buildup times with $H_2$ diffusion times calculated using a tarnishing model. They find that experimental buildup times are a factor of 10-100 larger than predicted from $H_2$ diffusion in bulk silica. They attribute the discrepancy to a 3% densification of thin SiO$_2$ films grown on crystalline silicon, compared to bulk silica glass. They also observe that compaction should lead to an increase in the $H_2$ diffusion activation energy. Fishbein et al. [22] have measured the annealing of interface traps in hydrogen at 400 - 500°C in poly gate MOS structures. They find that $D_n$ is about 30 times smaller compared to bulk silica, and the activation energy for $H_2$ diffusion is larger (0.58 eV). Both observations are in excellent agreement with results here. These results demonstrate convincingly that the time dependence of the $D_n$ creation measured here is related to the $H_2$ diffusion rate.

D. Time-dependent effects

Stahlbush et al. [9] have proposed the following model to explain how $H_2$ annealing after irradiation causes a large $D_n$ increase (this model is based on the $H^+$ model for $D_n$ creation after irradiation [11]): (1) Irradiation creates positively-charged defects $Q_{ox^+}$ in the oxide. (2) During exposure to an $H_2$ ambient, $H_2$ diffuses into the oxide. (3) $H_2$ reacts at $Q_{ox^+}$ defects to produce positively charged hydrogen ions ($H^+$) and neutral hydrogen ($H^0$). (4) $H^+$ and $H^0$ drift/diffuse to the Si-SiO$_2$ interface, where they react to create interface traps. (Once close enough to the interface, the $H^+$ ions can pick up an electron, which tunnels from the substrate, to produce $H^0$. Thus the actual reaction chemistry at the interface is the same for both $H^+$ and $H^0$). If $V_g < 0$ during the $H_2$ anneal, the $H^+$ ions will drift away from the Si-SiO$_2$ interface, and cannot contribute to the increase in $D_n$. On the other hand, $H^0$ motion is independent of the applied bias. Therefore, $\Delta D_n$ is larger when $V_g > 0$, as is observed in previous $H_2$ annealing experiments [9-13]. In this work, $\Delta D_n$ for $V_g < 0$ is about 50% of $\Delta D_n$ for $V_g > 0$. This ratio is larger than in previous experiments (ratio about ~25%) [9-13]. The reason for this is not understood, but may be associated with (unknown) differences in the oxides studied.

In the Stahlbush model, $H_2$ is cracked by positively-charged defects $Q_{ox^+}$ remaining after irradiation or FN-stress. Using electron spin resonance, Conley and Lenahan [14] have recently shown that the density of positive $E'$ centers decreases in proportion to the increase in $D_n$ ($E'$ defects are radiation or stress-induced hole which are trapped at an oxygen vacancy). This result is strong evidence that the $H_2$ molecule is cracked at $E'$ centers. In addition, both Stahlbush et al. [9] and Mrstik et al. [12] have observed that the density of positive oxide charge $Q_{ox^+}$ decreases roughly in a 1:1 ratio with the $D_n$ increase during $H_2$ annealing experiments. This result strengthens the identification of the $E'$ center as the $H_2$ cracking site, because $E'$ defects are probably the majority of the radiation-induced positive charge after irradiation. In this work, we also observe a small reduction in $Q_{ox^+}$ during $H_2$ annealing after FN-stressing or irradiation (see Fig. 1).
However, the data are not accurate enough to determine whether the ratio is 1:1. The following experiment has been performed to investigate the correlation between post-stress $Q_{ox}$ and $D_{it}$ growth during a subsequent $H_2$ anneal. After irradiation, several FETs were uniformly injected with electrons at low oxide field ($1 \text{ MV/cm}$) using substrate hot electron (SHE) injection [24]. These samples and an un-injected control were then annealed in $H_2$ with $V_g > 0$ as shown in Fig. 10. Electron injection clearly causes a smaller increase in $D_{it}$ during the subsequent $H_2$ anneal. Significant electron trapping at low electron injection levels indicates a large trap cross section, which also implies that the trapping defects are probably positively charged. Similarly, a reduced increase in $D_{it}$ was observed in e-injected samples with $V_g < 0$ during the $H_2$ anneal. Similar results were also obtained for e-injected FN-stressed oxides. It is not surprising that FN-injection results are so similar to irradiation because oxide damage for both types of stress involves holes. In FN-stress, although only electrons are probably injected, holes are created in the oxide by avalanche multiplication [25]. Using a MOSFET carrier separation technique, we find that the ratio of oxide hole current to electron current during FN-injection is about $3 \times 10^4$.

One aspect of the experiment in Fig. 10 which remains unclear is why $\Delta D_{it}$ continues to fall at the highest ($\approx 10^{18} \text{ e/cm}^2$) e-injection level. Careful measurements of the threshold voltage shifts as a function of $N_{hi}$ show that essentially all positive charge is removed at $N_{hi} = 10^{15} \text{ e/cm}^2$, as would be expected for removal of positive charge with a large capture cross section for electrons ($\approx 10^{-15} - 10^{-14} \text{ cm}^2$) [25]. Why then does $\Delta D_{it}$ continue to fall at higher values of $N_{hi}$? It is possible that some of the positive radiation-induced charge is compensated with nearby trapped electrons (similar to the compensation of near-interfacial traps [26]), which would cause the trapped positive charge to appear to have a small cross section.

The cracking reaction of $H_2$ molecules at stress-induced defects is perhaps the most interesting feature of these experiments, since other processes ($H_2$ diffusion, H reaction to create $D_{ox}$, etc.) have been studied in some detail previously. This research was undertaken partly in the hope that detailed measurements of the time-dependence could be used to explore this reaction, as attempted by Stahlbush et al. [10]. Unfortunately, it now seems clear that the time dependence is determined primarily by $H_2$ diffusion, and not $H_2$ cracking reaction. There are, however, two instances of time-dependent behavior which are not associated with the $H_2$ diffusion: (1) In Fig. 2, for positive $V_g$ during anneal, $\Delta D_{it}$ continues to increase without saturation even at very long times. (2) Studies of $\Delta D_{it}$ during $H_2$ annealing after channel hot carrier (CHC) stress have also been performed on the same FETs used here [15]. In this work, $\Delta D_{it}$ increases without saturating to at least $10^9 \text{ s}$. For CHC stress, defects are created in a non-uniform region very close (=0.2 $\mu$m) to the edge of the drain. From Fig. 6, we estimate the $H_2$ diffusion time to be relatively small, $\approx 10^{-100} \text{ s}$, which appears to be too small to account for the continued $D_{it}$ increase at long times. Qualitatively, $\Delta D_{it}$ increases approximately as $\log(\text{time})$ at long times in both the CHC data and here in Fig. 2. From these results, we tentatively conclude that this behavior is due to a slow hydrogen-related reaction. Note that this reaction must be different from the $H_2$ cracking reaction which mostly dominates in this work, and which must be much faster than the $H_2$ diffusion times. Further research is necessary to elucidate the nature of this second hydrogen reaction.

De Schrijver et al. have recently observed long-term $D_{it}$ instabilities (increases) in MOSFETs after CHC stress [17]. $D_{it}$ was found to increase approximately as the log(time) up to $10^6 \text{ s}$ after stress, and $\Delta D_{it}$ is larger for positive compared to negative $V_g$ during anneal. Both of these observations are consistent with CHC results reported in [15]. In both CHC stress papers [15,17], and in this work (Fig. 10) most of the long term $D_{it}$ increase is eliminated with a short electron injection after the stress. This result suggests that the hydrogen reactions involve positively-charged oxide defects. The main difference in these works is that FETs were deliberately exposed to $H_2$ after stress in [15] and here, so these results must be hydrogen related (essentially no $D_{it}$ increases were observed in unexposed control oxides). On the other hand, results in [17] were not obtained by direct exposure to an $H_2$ ambient, but rather were found to be associated with the presence of Si$_3$N$_4$ passivation layers. The similarity in results suggests that the presence of the passivation layer introduces hydrogen in some form into the gate oxide. Kohler and co-workers [27], among many others, have demonstrated that the radiation and CHC sensitivity of gate oxides can be affected by passivation layers, and that this effect probably arises from hydrogen introduced from the passivation layer. However, the exact source of the log(time) dependence in the CHC experiments is unclear, as discussed above, it cannot be due to simple $H_2$ diffusion.

In similar experiments, Schwank et al. [16] have reported "latent" $D_{it}$ increases at very long times ($10^6 \text{ s}$) after irradiation. Hydrogen was not deliberately introduced into these structures, but it could arise from the passivation layers, perhaps due to irradiation. An activation energy of 0.45 eV was obtained for the "latent" $D_{it}$ increase which, as discussed above, is in good agreement with the activation energy of $H_2$ diffusion in bulk silica glass [20]. Based on this diffusivity, the authors calculate a diffusion distance of 130 $\mu$m, which is much larger than the gate length of the MOSFETs (3 $\mu$m). This disagreement is not surprising since, in this work, we measure a time constant for $H_2$ diffusion of $\approx 10^8 \text{ s}$ at 295K, whereas the time constant of the latent buildup is much longer, $\approx 10^{10}-10^{12}$. Thus, it appears that the rate-limiting step of the latent buildup cannot be due to $H_2$ diffusion through the gate oxide of the FET. We propose that the rate limiting step could be diffusion of radiation-released $H_2$ through the
passivation layer. H₂ diffusion through the passivation layer must be slower than in bulk silica glass to account for the results.

SUMMARY AND CONCLUSIONS

D₂ growth has been measured during H₂ annealing in previously FN-stressed or irradiated oxides. A large D₂ growth is observed during H₂ exposure, in agreement with several previous reports. This result demonstrates that hydrogen chemistry is involved in the D₂ creation process. Results from FN-stressed and irradiated samples are almost identical, suggesting the same defects are created by FN stress and irradiation, defects which then react with H₂ during the D₂ creation process. Experiments using MOSFETs with different gate lengths show that the time dependence of D₂ growth during H₂ annealing increases rapidly with increasing gate length. This result strongly suggests that the time dependence is determined primarily by "lateral" H₂ diffusion through the gate oxide. This conclusion is supported by the measured activation energy of the time dependence, 0.57 eV, which is consistent with the value for H₂ diffusion in bulk silica glass, 0.45 eV. We argue above that the differences between H₂ diffusion in silica glass and results obtained here are the result of compaction of the thermal gate oxide as proposed by Mrstik [13]. This argument is supported by the excellent agreement between data here and data on H₂ diffusion distance and activation energy (0.57 eV) measured in MOS oxides during a study of D₂ annealing [22]. Electron injection experiments have been performed after the stress but before the H₂ anneal, which confirms that the H₂ molecules are cracked primarily at positively-charged stress-induced defects. This result is consistent with the spin resonance data of Conley and Lenahan [14] who identify the defect as the well-known E' center. In the case of CHC stress [15,17], and for positive bias at long times, the time dependence cannot be explained by H₂ diffusion, and may be associated instead with slow hydrogen reactions. This case needs further investigation.

It is clear from this work and a number of similar experiments described above that exposure of gate oxides to hydrogen can cause serious reliability problems. In this work, the FETs were deliberately exposed to hydrogen, but in other cases the hydrogen may arise from the passivation layers [16], or even may be liberated from the packaging during irradiation [27]. Our understanding of these effects is still clearly imperfect and bears further investigation. The development and use of improved passivation layers which do not contain hydrogen would also clearly be desirable.

Acknowledgements

This work was sponsored by the TREE/RM office of the Defense Nuclear Agency under DNA MIPR #93-522 and by the Office of Naval Research.

Figure 11. Schematic diagram of H₂ diffusion from all four sides of the poly gate of a MOSFET.

APPENDIX I. Calculation of 2-dimensional model for H₂ diffusion.

In this appendix, we calculate the characteristic diffusion time τ₁/₂ using a geometric 2-dimensional model as shown in Fig. 11. In this model, diffusion is assumed to occur from all four sides of the MOSFET gate, as opposed to the simpler 1-dimensional model where diffusion occurs only along the larger dimension. The linear diffusion distance d(t) is assumed at time t which defines an (assumed) abrupt edge to the diffusion region [13]. Using Fig. 11, it is readily shown from geometrical considerations that the normalized rate of Nₓ increase is given by:

\[ \frac{dN_x}{dN_{x,\text{max}}} = \frac{2dW_g + 2d y}{L_g W_g}, \]

where \( L_g \) and \( W_g \) are the MOSFET gate length and width, respectively, \( N_{x,\text{max}} \) is the maximum saturated value of \( N_x \), and \( y \) is defined in Fig. 11. Since, by definition, \( \tau_{1/2} \) is the time when \( \frac{dN_x}{dN_{x,\text{max}}} = 1/2 \), by setting Eq. (2) = 1/2, we obtain a value for \( d (\approx d_{1/2}) \):

\[ d_{1/2} = \frac{W_g}{4} [1 + L_g/W_g \cdot (1 + L_g^2/W_g^2)^{1/2}]. \]

As a check on Eq. (3), it can be shown that \( d_{1/2} = L_g/4 \) in the limit \( L_g < W_g \) (the range of \( d \) is from 0 to \( L_g/2 \)). Also, \( d_{1/2} = W_g/4 \) in the limit as \( L_g > W_g \). Thus, the 2-dimensional model reduces to the 1-dimensional model in both limits. Finally, since \( d \) is a 1-dimensional diffusion distance, \( d \) may be expressed as \( d = C_1 t^{1/2} \), where \( C_1 \) is a constant incorporating the H₂ diffusivity. Consequently, \( d_{1/2} = C_1 \tau_{1/2}^{1/2} \). Solving for \( \tau_{1/2} \), we obtain:
The relative increase in $\tau_{1/2}$ will be proportional to $d_{1/2}^2$, which can be calculated from Eq. 3 using solely geometric factors.

REFERENCES

24. R.B. Klein, N.S. Saks, and R.R. Siergiej, "Charge trapping in an ONO gate dielectric", in Physics and Chemistry of $SiO_2$ and the Si-$SiO_2$ Interface, edited by C.R. Helms and B.E. Deal, Plenum Press, to be published (1993). The pure $SiO_2$ control oxide in Fig. 3 is the same oxide as in Fig. 10 here.
36. R.B. Klein, N.S. Saks, and R.R. Siergiej, "Charge trapping in an ONO gate dielectric", in Physics and Chemistry of $SiO_2$ and the Si-$SiO_2$ Interface, edited by C.R. Helms and B.E. Deal, Plenum Press, to be published (1993). The pure $SiO_2$ control oxide in Fig. 3 is the same oxide as in Fig. 10 here.