The development of military and space electronics technology has traditionally been heavily influenced by the commercial semiconductor industry. The development of MOS technology, and particularly CMOS technology, as dominant commercial technologies has occurred entirely within the lifetime of the NSREC. For this reason, it is not surprising that the study of radiation interactions with MOS materials, devices and circuits has been a major theme of this conference for most of its history.

The basic radiation problem in a MOS transistor is illustrated in Fig. 1, where Fig. 1a shows the normal operation of a MOSFET. The application of an appropriate gate voltage causes a conducting channel to form between the source and drain, so that current flows when the device is turned on. In Fig. 1b, the effect of ionizing radiation is illustrated. Radiation-induced trapped charge has built up in the gate oxide, which causes a shift in the threshold voltage (that is, a change in the voltage which must be applied to turn the device on). If this shift is large enough, the device cannot be turned off, even at zero volts applied, and the device is said to have failed by going depletion mode.

In practice, the radiation-induced charging of the oxide involves several different physical mechanisms, which take place on very different time scales, with different field dependences and different temperature dependences. For this reason, the overall radiation response of a device or circuit can be extremely complex, sometimes to the point of bewilderment. However, the overall response can be separated into its components, and the components can be studied individually. In fact, this has happened. Many different individual investigators have studied different parts of the radiation response over a period of many years, and a reasonable degree of understanding has now been achieved. This understanding represents a major accomplishment of the NSREC and the NSREC community. Much of this understanding has been captured elsewhere already—Ma and Dressendorfer edited a major review volume. In addition, Oldham prepared another book, which was intended to update parts of the Ma and Dressendorfer book, to reflect later work. Both volumes discuss the same material to be presented here, and in far greater detail than is possible here.

II. Overview

We begin with an overview of the time-dependent radiation response of MOS systems, before discussing each of the major physical processes in greater detail. Then we will discuss the implications of the radiation response for testing, prediction, and hardness assurance. We will also discuss the implications of scaling (reducing the oxide thickness), and issues associated with oxide isolation structures and leakage currents.
Fig. 2 shows a schematic energy band diagram of a MOS structure, where positive bias is applied to the gate, so that electrons flow toward the gate and holes move to the Si substrate. Four major physical processes, which contribute to the radiation response of a MOS device, are also indicated. The most sensitive parts of a MOS system to radiation are the oxide insulators. When radiation passes through a gate oxide, electron/hole pairs are created by the deposited energy. In SiO$_2$, the electrons are much more mobile than the holes, and they are swept out of the oxide, typically in a picosecond or less. However, in that first picosecond, some fraction of the electrons and holes will recombine. That fraction will depend greatly on the energy and type of the incident particle. The holes, which escape initial recombination, are relatively immobile, and remain near their point of generation, where they cause a negative threshold voltage shift in a MOS transistor. These processes, electron/hole pair generation and recombination, together are the first process depicted in Fig. 2. In Fig. 3, this process determines the (maximum) initial threshold voltage shift.

![Figure 3](image)

**Figure 3.** Schematic time-dependent post-irradiation threshold voltage recovery of n-channel MOSFET, relating major features of the response to underlying physical processes.

The second process in Fig. 2 is the transport of the holes to the Si/SiO$_2$ interface, which causes the short-term recovery of the threshold voltage in Fig. 3. This process is dispersive, meaning that it takes place over many decades in time, and it is very sensitive to the applied field, temperature, oxide thickness, and (to a lesser extent) oxide processing history. This process is normally over in much less than a second at room temperature, but it can be many orders of magnitude slower at low temperature.

The third process in Fig. 2 and Fig. 3 is that when they reach the Si interface, some fraction of the transporting holes fall into relatively deep, long-lived trap states. These trapped holes cause a remnant negative voltage shift, which can persist for hours or even for years. But even these stable trapped holes undergo a gradual annealing, which is illustrated in Fig. 3.

The fourth major component of MOS radiation response is the radiation-induced buildup of interface traps right at the Si/SiO$_2$ interface. These traps are localized states with energy levels in the Si band-gap. Their occupancy is determined by the Fermi level (or by the applied voltage), giving rise to a voltage-dependent threshold shift. Interface traps are highly dependent on oxide processing, and other variables (applied field and temperature).

Fig. 3 is schematic in that it does not show real data, but it reasonably represents the main features of the radiation response of a hardened n-channel MOS transistor. The range of data, from $10^8$ s to $10^9$ s, is enormous, as it has to be to include qualitatively the four main processes we have discussed. For the oxide illustrated in Fig. 3, a relatively small fraction of the holes reaching the interface are trapped, which is why we say it is realistic for a hardened oxide. Many oxides would trap more charge than is shown here. In addition, the final threshold shift, including interface traps, is positive (the so-called rebound or superrecovery effect) here because the number of negatively charged interface traps finally exceeds the number of trapped holes. Not all oxides really have this behavior, but it is one of the results, which can be considered "typical."

### III. Description of Basic Physical Processes Underlying the Radiation Response of MOS Devices

Next, we consider these basic physical mechanisms in more detail, and provide critical references. But for a complete review, the readers should consult the references.

#### A. Electron-Hole Pair Generation Energy

The electron/hole pair creation energy, $E_p$, was determined to be $18 \pm 3$ eV by Ausman and
McLean, based on experimental data obtained by Curtis et al. This result has been confirmed independently by others, including a more accurate set of measurements and analysis by Benedetto and Boesch, which established $E_p = 17 \pm 1$ eV. From this value of $E_p$, one can calculate the charge pair volume density per rad, $g_0 = 8.1 \times 10^{-12}$ pairs/cm$^3$-rad. But this initial density is quickly reduced by the initial recombination process, which we discuss next.

### B. Initial Hole Yield

The electrons are swept out of the oxide very rapidly, in a time on the order of a picosecond, but in that time some fraction of them recombine with the holes. The fraction of holes escaping recombination, $f_e(E_{ox})$, is determined mainly by two factors: the magnitude of the electric field, which acts to separate the pairs; and the initial line density of charge pairs created by the incident radiation. The pair line density is determined by the linear energy transfer (LET), and is, therefore, a function of the incident particle type and energy. The line density is also inversely proportional to the average separation distance between electron/hole pairs—obviously, the closer the average spacing of the pairs, the more recombination will occur at a given field, and the less the final yield of holes will be.

![Schematic diagrams indicating pair separation distances for two recombination models](image)

**Figure 4.** Schematic diagrams indicating pair separation distances for two recombination models: (a) geminate (separate electron/hole pairs) and (b) columnar (overlapping electron/hole pairs).

The recombination problem cannot be solved analytically for arbitrary line density, but analytic solutions do exist for the limiting cases, where the pairs are either far apart, or very close together. These cases are illustrated in Fig. 4. Fig. 4a shows the so-called geminate recombination model, where the average separation between pairs is much greater than the thermalization distance, the distance between the hole and the electron. One can treat the interaction between the charges of an isolated pair, which have a mutual coulomb attraction, which undergo drift motion in opposite directions under the influence of the applied field, and which have a random diffusion motion driven by the thermal fluctuations of the system. But interactions with other pairs can be neglected. The geminate recombination model was first formulated by Smoluchowski, and later solved by Onsager, originally for the recombination of electrons and positive ions in gases. Experimental and theoretical results for the geminate process are shown in Fig. 5, where the theoretical curve was obtained by Ausman, assuming the average thermalization radius, $r_t$, to be 5 nm.

![Fractional yield as a function of applied field for Co$^{60}$ gamma rays, 12-MeV electrons, and geminate model calculations](image)

**Figure 5.** Fractional yield as a function of applied field for Co$^{60}$ gamma rays, 12-MeV electrons, and geminate model calculations.

The other case, called columnar recombination, is illustrated in Fig. 4b, where the separation between pairs is much less than $r_t$. There are several electrons closer to any given hole than the electron, which was its original partner, so the probability of recombination is obviously much greater than in the geminate case. The columnar model was originally solved analytically by Jaffe, extending earlier work by Langevin. More recently, Oldham has presented a more accurate numerical solution of the Jaffe equation, which extends the range of applicability of the model. Representative experimental data, for 2 MeV alpha particles, are presented in Fig. 6 along with theoretical curves. The parameter, $b$, is the half-diameter assumed for the initial Gaussian charge distribution. Recombination results for a variety of incident particles are summarized in Fig. 7. At a field of 1 MV/cm, there is a difference of more than an order of magnitude in the yields shown for different particles. Clearly,
recombination is an important effect, which must be accounted for, when comparing the effect of different radiation sources.

Figure 6. Fractional yield as a function of applied field for alpha particles incident on SiO₂. Solid lines indicate columnar model results for different initial column radii.¹⁴,¹⁶

As one might expect, there are also a number of intermediate cases of practical interest, where neither model is strictly applicable, as illustrated in Fig. 8. The original version of this figure was published by Oldham¹⁸ in 1984, based on experimental data by Stassinopoulos et al.¹⁹⁻²¹ and Brucker et al.²²,²³. But additional data points have been added from time-to-time, as other experiments were reported.²⁴⁻²⁶ These later experiments have generally reported lower yield (that is, more recombination) than the earlier experiments. The different experiments have all been done at different fields, so the geminate model limit is different in each case, as is now indicated in Fig. 8. The other key point is that the LET for a proton and an electron is not really the same until their energy is about 1000 MeV, well above the energy of any incident particle in any of these experiments. In the original version of Fig. 8, the geminate model was indicated schematically to apply from 1000 MeV down to about 150 MeV because it appeared to fit the data available at that time. But the more recent data indicates that the recombination process is not purely geminate until the proton energy is well above 100-200 MeV.

C. Hole Transport

The transport of holes through the oxide layer has been studied extensively by several groups,²⁷⁻³⁸ and it has the following properties: (1) transport is highly dispersive, taking place over many decades in time following a radiation pulse. (2) It is universal in nature, meaning that changes in temperature, field and thickness do not change the shape or dispersion of the recovery curves on a log-time plot. Changes in these variables affect only the time-scale of the recovery. (3) The transport is field activated. (4) At temperatures above about 140K, the transport is strongly temperature activated, but it is not temperature activated below about 140K. (5) The hole transport time, or recovery time, has a strong super-linear power law dependence on oxide thickness.

The best overall description of the experimental hole transport data seems to be provided by the CTRW (continuous-time-random-walk) hopping transport formalism, which was originally developed by Montroll, Weiss, Scher and others.³⁹⁻⁴². This formalism has been applied to hole transport in silicon dioxide by McLean,²⁷,³⁰,³¹,³₃⁻₃₅ and by Hughes.²⁸,²⁹,⁴₄ (However, the multiple trapping model²⁲,³⁷ also accounts for many of the features of the experimental data.) The specific transfer mechanism seems most
likely to be small polaron hopping of the holes between localized, shallow trap states having a random spatial distribution, but having an average separation of about 1 nm. The term polaron refers to the situation where the carrier interacts strongly with the surrounding medium, creating a lattice distortion in its immediate vicinity (also referred to as self-trapping). As the hole hops through the material, it carries the lattice distortion with it. The strongest evidence for the polaron hopping mechanism is the transition from thermally activated transport above about 140K, to non-activated transport at lower temperatures. This transition is a classic signature of polaron hopping. Many other features of the hole transport, such as dispersion, universality, and superlinear thickness dependence, can be attributed to a wide distribution of hopping times for the individual holes.

Representative experimental data are presented in Figs. 9-12. Fig. 9 shows the effect of temperature variation, and Fig. 10 shows the effect of varying the electric field. In both Figs., the results cover seven decades in log-time, following a short radiation pulse. In both Fig. 9 and 10, the flatband voltage shift is plotted as a function of log-time, normalized to the calculated shift before any transport occurs. In Fig. 9, the field is 1 MV/cm, and the strong temperature activation above 140K is apparent. In Fig. 10, all the curves are taken at T=79K. The universality and dispersion of the transport is better illustrated in Fig. 11, where all the curves from Fig. 9 are replotted for scaled time. The entire transport process covers 14 decades in time (!), and all the curves have the same "S" shape. The time, t_{1/2}, at which the flatband voltage reaches 50% recovery, has been used as
the scaling parameter. The solid line is an analytical fit of the CTRW model, where the shape parameter, $\alpha$, has the value 0.25. Finally, Fig. 12 shows how the hole transit time varies with oxide thickness, as $t_{tr} \propto \delta^{\alpha}$, or about $\delta^{0.4}$. This oxide thickness dependence arises because the farther the holes transport, the greater the probability that some of them will be in states where the next hop is a difficult one, one that takes a long time to happen. Then, the farther the holes go, the slower they move.

D. Deep Hole Trapping and Annealing

The most complete discussion of hole trapping and annealing is by Oldham.2 Deep hole traps near the Si/SiO$_2$ interface arise because there is a transition region where oxidation is not complete. This region contains excess Si, or oxygen vacancies, depending on how one looks at it. The oxidation process was described by Deal et al.46 in an early review article, which was based on original work done even earlier. Eventually, Feigl et al.47 presented a convincing model for the single oxygen vacancy. Basically, there is one oxygen atom missing from the usual lattice configuration, leaving a weak Si-Si bond, where each Si atom is back-bonded to three oxygen atoms. When a positive charge is trapped, the Si-Si bond is broken, and the lattice relaxes. The key point that Feigl added to earlier discussions was that the relaxation is asymmetric—one atom relaxes into a planar configuration, and the other remains in a tetrahedral configuration. The oxygen vacancy was also eventually connected to the E' center, originally detected by Weeks48 in $\alpha$-quartz, but also later detected in bulk glasses and thermally grown SiO$_2$.49 The correlation of E' centers, and oxygen vacancies, with radiation-induced trapped holes was first established by Lenahan and Dressendorfer.49

Oxide trapped holes are relatively stable, but they do undergo a long-term annealing process which can extend for hours or even years, with a complex dependence on time, temperature, and applied field. Generally, trapped hole annealing can proceed by either of two processes, tunneling or thermal excitation. At or near room temperature, tunneling is the dominant mechanism, but if the temperature is raised enough, the thermal process will eventually dominate. Tunneling has been analyzed by several authors,50-56 as has the thermal process.57

Both processes can give rise to the linear-with-log $t$ dependence that has been observed empirically,51-53 but one has to make different assumptions about the trap energy level distribution for the two processes.

![Figure 13. Trapped hole annealing; negative bias curve shows that “annealed” holes are not really removed.70](image)

The study of radiation-induced trapped hole annealing has led to new insights about the atomic structure of the oxide trap, which in turn has led to new insights into the structure of neutral electron traps, which play a critical role in breakdown studies and in other reliability problems. (For a full discussion, see reference 2.) One of the key results is illustrated in Fig. 13, originally reported by Schwank et al.70 An irradiated sample was annealed under positive bias at 100 C for about one week, and all the trapped positive charge appeared to be removed. But then they applied a negative bias, and about half the neutralized positive charge was restored within a day. This result led to the idea that annealing of radiation damage involved a compensation process. That is, defects were neutralized without being removed. Although other groups quickly confirmed the basic result,71 several years passed before Lelis et al. did a thorough study.72-74 One of their key results is shown in Fig. 14, where a hardened oxide is exposed to a short Linac radiation pulse, and then subject to a series of alternating positive and negative bias annealing steps. Under negative bias, a significant amount of neutralized positive charge reappears, but there is also a significant amount of “true” annealing, where the trapped charge really is removed. Lelis et al. proposed a model, illustrated in Fig. 15, to account for their results and results of others. Generally, it had been assumed that annealing proceeded by an electron tunneling
Figure 14. Alternate positive and negative bias annealing for capacitor exposed to 4-μs Linac pulse.

to the positively charged Si, neutralizing it, and reforming the Si-Si bond. Instead, Lelis et al. proposed the electron tunnels to the neutral Si, forming a dipole structure, where the extra electron can then tunnel back and forth to the substrate in response to bias changes. This model is consistent with the ESR work of Lenahan et al. and with the electrical results of Schwank and others, explaining a variety of complex results in terms of a single defect, which was already well-known. The transition from Fig. 15a to 15b was described by Feigl et al. The transition from Fig. 15b to 15c and back describes the switching reported by Schwank et al. and by Lelis et al. And the transition from Fig. 15c back to 15a indicates the true annealing, which is also observed.

The dipole hypothesis by Lelis et al. was attractive, because it explained many things very simply, but at first, it was also controversial. It was criticized by three different groups, for different reasons. The biggest problem was that putting an extra electron on a neutral Si atom instead of a positive Si, required overcoming an electron-electron repulsion. Lelis et al. pointed out that adding the extra electron to the positive Si would require changing a planar configuration of atoms into a tetrahedral configuration, moving around atoms in the lattice to change bond angles, which would require adding energy. The energy to rearrange the lattice might be greater than the electron-electron energy, which would mean stable dipoles would be energetically favored. But, initially, they lacked the means to quantify this argument, so debate continued for several years.

Several other, independent, experiments produced results, which seemed to support the dipole hypothesis. These included TSC (thermally stimulated current) measurements, first by Shanfield et al., and later by Fleetwood et al. In addition Walters et al. concluded that the dipole proposed by Lelis et al. acted as a neutral electron trap in their injection experiments. This work was an important independent confirmation of the dipole hypothesis, but it was also an important extension of it. Basically, they argued that the positive end of the dipole acted as an electron trap, so that a second electron could be trapped, making the whole complex net negative. The reason this result was important was that there is an enormous body of literature on neutral electron traps and the critical role they play in non-radiation-induced reliability problems. (This literature is too extensive to discuss here—see reference 2.) But suddenly, the Lelis et al. dipole hypothesis was critical for explaining all this other work. And finally, Conley et al.

Figure 15. Model of hole trapping, permanent annealing, and compensation processes.

The dipole hypothesis by Lelis et al. was attractive, because it explained many things very simply, but at first, it was also controversial. It was criticized by three different groups, for different reasons. The biggest problem was that putting an extra electron on a neutral Si atom instead of a positive Si, required overcoming an electron-electron repulsion. Lelis et al. pointed out that adding the extra electron to the positive Si would require changing a planar configuration of atoms into a tetrahedral configuration, moving around atoms in the lattice to change bond angles, which would require adding energy. The energy to rearrange the lattice might be greater than the electron-electron energy, which would mean stable dipoles would be energetically favored. But, initially, they lacked the means to quantify this argument, so debate continued for several years.

Several other, independent, experiments produced results, which seemed to support the dipole hypothesis. These included TSC (thermally stimulated current) measurements, first by Shanfield et al., and later by Fleetwood et al. In addition Walters et al. concluded that the dipole proposed by Lelis et al. acted as a neutral electron trap in their injection experiments. This work was an important independent confirmation of the dipole hypothesis, but it was also an important extension of it. Basically, they argued that the positive end of the dipole acted as an electron trap, so that a second electron could be trapped, making the whole complex net negative. The reason this result was important was that there is an enormous body of literature on neutral electron traps and the critical role they play in non-radiation-induced reliability problems. (This literature is too extensive to discuss here—see reference 2.) But suddenly, the Lelis et al. dipole hypothesis was critical for explaining all this other work. And finally, Conley et al.

Figure 16. E' density during alternate positive and negative bias annealing.

Conley et al. conducted ESR experiments, where they cycled charge back and forth by alternating bias, while monitoring the E' signal. Their main result is
shown in Fig. 16. The dipole model was the only one consistent with this result, so it settled the debate, at least on the experimental side. Even more recently, two theoretical groups have done quantum mechanical calculations, using different mathematical approaches, which have also indicated that dipoles should be energetically favored under certain conditions. In addition, Fleetwood et al. have recently extended this model to argue that it also accounts for 1/f noise results, which they reported.

We note that, in recent years, there has been much discussion of the role of border traps, oxide traps that exchange charge with the Si substrate. The proposal to call these traps border traps was made by Fleetwood in 1992. At that time, the dipole model by Lelis et al. had been in the literature for four years, and was already well known. Now, more than ten additional years have passed, and the defect described by Lelis et al. is still the only confirmed border trap, at least in the Si/SiO2 system. Other border trap structures have occasionally been proposed, but they have not done well in subsequent experimental tests.

E. Radiation-Induced Interface Traps

Radiation-induced interface states have been identified with the so-called P_{50} resonance in ESR studies, by Lenahan and Dressendorfer. This center is a tri-valent Si atom, back bonded to three other Si atoms, with a dangling bond extending into the oxide. This defect is amphoteric, negatively charged above mid-gap, neutral near mid-gap, and positively charged below mid-gap. Lenahan and Dressendorfer showed a very strong correlation between the build-up of the P_{50} resonance and the buildup of interface traps, as determined by electrical measurements. There is also an extensive literature suggesting that this same defect is also present as a process-induced interface trap. We cannot review all this literature here, but other reviews have already been published. (In particular, see the references in Chapter 3 of reference 2.) The basic picture, however, is that when the oxide is grown, there are still about 10^{13}/cm² unpassivated tri-valent Si centers. In subsequent processing, almost all these centers are passivated by reacting with hydrogen. However, they can also be depassivated, either by radiation interactions or by other environmental stresses.

There have been many conflicting models proposed to describe the processes by which radiation produces interface traps, and much controversy about them. However, a reasonable degree of consensus has finally emerged. All the models are consistent with the idea that the precursor of the radiation-induced interface trap is a Si atom bonded to three other Si atoms and a hydrogen atom. When the Si-H bond is broken, the Si is left with an unpassivated dangling bond, as an electrically active defect. The process by which the Si atom is depassivated is where the models differ. Now it has become clear that the dominant process is a two stage process involving hopping transport of protons, originally described by McLean, which was based on a series of experimental studies by his coworkers. This work was confirmed and extended in a series of additional studies by Saks and his coworkers. In the first stage of this process, radiation-induced holes transport through the oxide, and free hydrogen, in the form of protons. In the second stage, the protons reach the interface, they react, breaking the SiH bonds already there, forming H₂ and a tri-valent Si defect. One of the critical experimental results is shown in Fig. 17, which shows the results of bias switching experiments that support H⁺ transport model.

Figure 17. Experimental results from field switching experiments that support H⁺ transport model.
interface states for curves A and B is almost the same, however. For curve E, the bias is maintained negative throughout both stages, and interface trap production is suppressed completely. For curves C and D, the bias is negative during irradiation and hole transport, but switched positive later than for curve B. In all cases, bias polarity during the hole generation and transport made no difference, but positive bias during the proton transport was necessary to move the protons to the Si/SiO₂ interface. The time scale of the interface trap build-up was determined by the transport time of the protons. (For curves B, C, and D, the protons were initially pushed away from the interface, so it took them longer to get there after the proper bias was applied.) McLean also worked out the average hopping distance for protons to be .26 nm, which is the average distance between oxygen atoms. And he determined the activation energy for the interface trap buildup to be 0.82 eV, which is consistent with proton transport. Saks eventually succeeded in monitoring the motion of the protons directly. The two stage, proton transport model is a robust model at this point. It has been confirmed by different groups (McLean and coworkers, and Saks and coworkers), using different test structures (capacitors and transistors, respectively), different gate technologies (Al active metal and poly-Si, respectively), and different measurement techniques (C-V analysis and charge pumping respectively). Despite all these variations, this process has always been the main effect. However, it does not explain everything. Boesch have identified a second order effect, where a small part of the interface trap build-up seems to correlate with the arrival of transporting holes at the interface. Presumably, the holes break the Si-H bonds instead of protons. Also, Griscom and Brown had also proposed originally, that diffusion of neutral hydrogen, rather than drift transport of protons, was the main mechanism for interface trap production. But, Saks et al. were able to isolate the neutral hydrogen effect, and showed that it was also small. The key result is shown in Fig. 18, where the interface trap build-up between 120K and 150K is due to neutral hydrogen, and the build-up above 200K is due to proton transport. The vertical scale is a log scale here, so the neutral hydrogen process accounts for only a few percent of the total build-up. Griscom and Brown both eventually endorsed the McLean model.

There have also been several models proposed where trapped holes are somehow converted to interface traps—usually the details of the conversion process are not specified. These models are not well regarded today. For example, in Fig. 17, curves B and E have the same hole trapping. For this reason, one might expect these models to predict similar interface trap build-ups, contrary to what is shown in the figure. If one studies hole trap removal and interface trap build-up carefully, the two processes have different time dependences, different temperature dependences, and different bias dependences. They have the same dose dependence, but otherwise seem to be completely independent. Here, we cannot discuss these things in detail, but full discussions appear in Chapter 3 of reference 2, and in reference 88. However, Oldham et al. provided the most likely explanation for experimental results purporting to show trapped hole conversions. They pointed out that trapped holes that do not undergo a defect transformation can account for most of these results—that is, trapped holes look like interface traps in some experiments. (The model for these holes is discussed in the previous section.) The exchange of charge between trapped holes and the Si substrate has been extensively studied since then (usually under the name border traps), and the idea that such charge exchange takes place is no longer considered unusual.

Finally, some samples exhibit what has been called a latent interface trap build-up, which is illustrated in Fig. 19. This effect is thought to be due to hydrogen diffusing into the gate oxide region from another part of the structure, perhaps
the field oxide or an encapsulating layer. The latency period arises because the hydrogen is diffusing from (relatively) far away. From a

![Figure 19. Latent interface trap formation for OKI p-channel transistors.][1]

testing point of view, this is a difficult effect to account for. There is no trace of it on the time scale of most laboratory tests, yet it can eventually be a large, even dominant, effect, on a time scale of months. Saks et al. subsequently reported that the latent build-up is suppressed by a nitride encapsulating layer, which serves as a barrier to hydrogen diffusion. Unless one knows how the samples are encapsulated, it is not possible to predict ahead of time whether a latent build-up will occur, or not.

IV. Implications for Radiation Testing, Hardness Assurance, and Prediction

We have now completed our review of the basic physical mechanisms underlying the radiation response of CMOS devices. Next we consider these mechanisms in the context of device and circuit testing, hardness assurance, and prediction.

A. Rebound or Super-Recovery

The rebound effect is illustrated in Fig. 20, which shows threshold voltage shift ($\Delta V_T$) for a MOSFET, along with its components, oxide trapped charge ($\Delta V_{OT}$) and interface trapped charge ($\Delta V_{IT}$), during both irradiation and post-irradiation annealing. The annealing data is shown for two temperatures, 25 C and 125 C. After irradiation, the threshold shift is less than 1V, but this relatively small shift is obtained by compensating positive oxide trapped charge with negatively charged interface traps (in this n-channel device). When the hole traps anneal, however, the final threshold shift is positive, about 3.5 V, which is more than enough to fail the device. The effect of raising the temperature

![Figure 20. $V_T$, $V_{OT}$, and $V_{IT}$ annealing; long term response illustrates rebound effect.][2]
is to anneal the hole traps faster, but the same final state is reached at room temperature. One could imagine a device that failed due to trapped positive charge immediately after irradiation, that would later work properly for a time as some of the holes annealed, that would then fail again later from trapped negative charge as more of the holes annealed. There is very little change in interface trap density during the annealing process at either temperature, so the late time failure would be due to interface traps that were there at the end of the irradiation.

To test for rebound, there is now a standard test method, 1019.4, which calls for a 100 C anneal for 168 hours, which will detect the effect in oxides similar to the one used in Fig. 20. This method represents a compromise, because 100 C is not a high enough temperature to accelerate the trapped hole annealing in all oxides, but if one goes much higher in temperature, the interface traps may anneal, too.

The rebound effect is of great practical concern for space environments, because components are typically exposed to relatively low dose rates for very long mission lifetimes. There is now strong emphasis on using unhardened commercial technology as much as possible, which is reasonable to consider in space because oxide traps may anneal as fast as they are created, or nearly so. A component that fails at a low dose in a laboratory test from oxide trapped charge, as many things do, may work quite well in space, because of the low dose rate and annealing. But this discussion only applies to the positive
charge—negatively charged interface traps (in an n-channel device) will continue to build-up throughout the mission life. Therefore, it is necessary to check for rebound, too.

We note that rebound is only a problem on n-channel devices, because the hole traps and interface traps are both positive in p-channel devices. For this reason, their electrical effects add, instead of compensating. Of course, the hole traps do not change state with changes in applied bias in either n- or p-channel devices, while the interface trap state does depend on bias in general.

B. Apparent Dose Rate Effects

In CMOS devices the radiation response does not normally depend on the dose rate, except in the case of extremely short, intense nuclear-driven pulses. However, apparent dose rate effects are often observed if a given dose is delivered at two or more different rates in different tests, because the exposure times are different, and we have already discussed many different time dependent effects that will contribute to the overall response. Generally, if one allows the sample to anneal after the shorter exposure, so that the measurements are done at the same times, the response will be the same, within normal experimental error. The most definitive experiment showing the absence of a true rate dependence was reported by Fleetwood et al. where identical samples were exposed to the same dose, with the dose rate varied over 11 orders of magnitude. The results for $\Delta V_T$, $\Delta V_{OT}$, and $\Delta V_{IT}$ are shown in Fig. 21. In each case the sample with the highest rate exposure is allowed to anneal following irradiation. All the lower rate exposures fall almost perfectly on the annealing curve, indicating that if the dose and the time of the measurement are the same, the response will also be the same.

An approach, which has been used with some success, to predict the response of a CMOS device at dose rates other than those available in the laboratory is the use of linear systems theory. If one determines, by testing, the impulse response function of a device to a short radiation pulse, then one can determine the response to an arbitrary exposure by doing a convolution integral, as long as the response is linear with dose (meaning that the response to the different dose increments simply add up). The impulse function used in references 51-53

![Figure 21. $V_T$, $V_{OT}$, $V_{IT}$ annealing results showing absence of dose rate effect.](image)

Figure 21. $V_T$, $V_{OT}$, $V_{IT}$ annealing results showing absence of dose rate effect.

and 120 assumed that the annealing response was linear-with-log(t), which is reasonable for many unhardened commercial oxides. In general, the impulse response function may be more complicated, but linear systems theory can still be used, in principle.

![Figure 22. Dependence of circuit failure level on dose rate.](image)

Figure 22. Dependence of circuit failure level on dose rate.
A spectacular example of the mischief that can be caused by testing at different dose rates is shown in Fig. 22. In this case, a circuit was tested to failure at a wide range of dose rates. At high dose rates (the right hand side of the Figure), it failed at a dose of a few krad, because of the build-up of positive oxide trapped charge. At low dose rates (the left hand side), it failed at slightly higher doses due to negatively charged interface traps (rebound). But at one dose rate in the middle, where positive and negative charge generation were precisely balanced, it survived to very high doses.

C. Nonlinear Effects

In the previous section, we discussed the use of convolution integrals in linear systems theory to predict the radiation response of a component, which will work if the system response is linear with dose. Unfortunately, there are many cases of practical interest where the response is nonlinear. For example, hole trapping may saturate with dose, due to trap filling, space charge effects, recombination of trapped holes with radiation-induced electrons, or a balance between hole trapping and tunnel annealing. Space charge effects, in particular, play a critical role in at least two areas, SOI buried oxides and bipolar isolation oxides. Boesch et al. showed that SIMOX SOI buried oxides trap essentially all the radiation-induced charge, so that space charge fields are much larger than any applied field, and the response is dominated by space charge effects. And the enhanced low dose rate sensitivity (ELDRS) of some bipolar circuits has been shown to be related to space charge effects. We will not discuss these topics here, because they are covered in detail in other papers in this issue. But it is useful to keep in mind that space charge effects can happen in bulk CMOS devices, too.

D. Charge Separation Techniques

In order to do sensible testing and analysis, one naturally wants to be able to separate the overall radiation response of a device or test structure into its components. Therefore, after the hole transport is complete, it is common practice to write \( \Delta V_T = \Delta V_{OT} + \Delta V_{IT} \), where the right hand terms are the threshold voltage shifts due to oxide traps and interface traps, respectively. There are different methods for separating \( \Delta V_T \) into its components, but they all use the assumption that interface traps are net neutral at midgap, so that \( \Delta V_{MG} \) is a measure of oxide hole trapping (that is, \( \Delta V_{MG} = \Delta V_{OT} = -q \Delta N_{OT} / C_{ox} \)). Here, \( q \) is the electronic charge, \( C_{ox} \) is the oxide capacitance, and \( N_{OT} \) is the number of oxide traps. Then the shift due to interface traps is everything else, \( \Delta V_{IT} = \Delta V_T - \Delta V_{MG} \). For a capacitor, one can use the stretch-out between midgap and inversion, or the stretch-out between threshold and midgap on the I-V characteristic of a transistor (which usually requires extrapolating the subthreshold current to midgap). Rather than discuss the details of these procedures, we simply give a few key references. We note that the assumption of midgap neutrality for interface traps was first used by Lenahan and Dressendorfer, and reexamined later by McWhorter, and still later by Lenahan (again). McWhorter concluded that the point of neutrality for interface traps is close to midgap, perhaps 3kT below midgap. Lenahan concluded that neutrality for the \( P_{60} \) center is at midgap, but he also detected a second center, called \( P_{61} \), which is present in smaller numbers, and which is not neutral a little below midgap (consistent with McWhorter). So the assumption that \( \Delta V_{MG} \) is due entirely to oxide trapped charge is a useful approximation, which seems to introduce errors of only a few percent.

E. Dose Enhancement

Dose enhancement has been known and studied for many years, but it is a practical problem because of the widespread use of low energy (10 keV) X-ray sources. For a photon source (x-ray or gamma), most of the energy deposition is actually done by secondary electrons. The critical concept is called charged particle equilibrium (CPE). Normally, in a homogeneous slab of material, CPE is maintained because the number of secondary electrons scattering into any increment of volume is equal to the number of electrons scattering out. The problem in a MOS device is illustrated in Fig. 23, where there are several thin layers of different compositions, and therefore, different cross-sections. CPE is not maintained because more secondary electrons cross an interface from the high-Z side than from the low-Z side. In Fig. 23, the solid lines indicate the deposition profile that would be predicted for 10 keV X-rays, using the mass absorption coefficients alone, without any secondary electron transport. Fig. 23a indicates the situation when the oxide is thick compared to
Figure 23. Schematic diagram illustrating dose enhancement in thick and thin oxide layers; solid lines are bulk equilibrium doses and dashed lines represent actual dose profiles.

the range of the secondary electrons, where the broken line indicates the change in the depth-dose profile from the transport of secondary electrons into the oxide layer. The situation for a thin oxide is illustrated in Fig 23b, where the electrons go all the way through the oxide layer, and the dose-enhancement is indicated by the broken line. Dose enhancement as a function of oxide thickness is indicated in Fig. 24. For testing with an X-ray source, dose enhancement is an important effect, which means that the dose is different in different parts of the structure (e.g., gate oxides and field oxides). On the other hand, in a Cobalt source, the mass absorption coefficient for all the materials shown in Fig. 23 is essentially equal, and the dose is uniform because CPE is maintained. Dosimetry, in general, is covered in a separate paper in this issue, so we will not discuss it further here.

F. Implications of Scaling

In the history of the commercial semiconductor industry, few things have been more important than scaling, the regular shrinking of device feature sizes, so that larger and larger integrated circuits can be fabricated in a given chip area. Of course, the radiation effects community has been swept along, with a new generation of chips every few years, and similar hardening problems to be solved in each new generation. The fact that NSREC has SEE sessions is a consequence of scaling, but SEE is covered in other papers in this issue, so we won't say much about it here. However, there are impacts of scaling in the total dose response of CMOS, which are appropriate to cover here.

Figure 24. Measured and calculated dose enhancement as a function of oxide thickness for 10 keV x-rays.

The most obvious of these is the thinning of the gate oxide. Twenty years ago, oxides were typically about 100 nm thick, but now, commercially available oxides are less than 10 nm thick, and research samples are a lot less than that. McGarrity worked out the gains in gate oxide hardening that could be achieved merely by thinning the oxide, without special processing. The point of his analysis is that $\Delta V_T = Q_{ox}/C_{ox}$, which leads to the prediction that the threshold voltage shift is proportional to oxide thickness squared. The total charge in the oxide is proportional to thickness, and the capacitance is inversely proportional to thickness. (We note that other dependences have occasionally been reported, and oxide processing varies so much there is no reason not to believe data showing a

Figure 25. Threshold and flatband voltage shifts per unit dose as a function of oxide thickness at 80K. Dashed line is thickness squared, solid line and points are experimental results.

The most obvious of these is the thinning of the gate oxide. Twenty years ago, oxides were typically about 100 nm thick, but now, commercially available oxides are less than 10 nm thick, and research samples are a lot less than that. McGarrity worked out the gains in gate oxide hardening that could be achieved merely by thinning the oxide, without special processing. The point of his analysis is that $\Delta V_T = Q_{ox}/C_{ox}$, which leads to the prediction that the threshold voltage shift is proportional to oxide thickness squared. The total charge in the oxide is proportional to thickness, and the capacitance is inversely proportional to thickness. (We note that other dependences have occasionally been reported, and oxide processing varies so much there is no reason not to believe data showing a
different dependence for a particular oxide. But other dependences have not been shown to hold, in general.) The most important deviation from the \( t_{ox} \) dependence occurs in very thin oxides, where tunnel annealing eliminates, or at least neutralizes trapped charge near the interface. The point is that for thin oxides, this annealing process occurs at both interfaces, and accounts for all or nearly all of the trapped oxide charge. For thin enough oxides, the two tunneling fronts meet in the center of the oxide, leaving no net positive oxide charge. Data illustrating this effect is shown in Fig. 25.

Since mainstream commercial oxides are now thin enough that radiation-induced AVT has essentially vanished, the problem of hardening gate oxides is basically solved. This leaves field oxide isolation structures as the main remaining total dose problem, which we will discuss in the next section. However, there are three other gate oxide total dose effects, which we should mention here.

The first of these is the so-called stuck bit problem, which is caused by the total dose deposited by a single ion passing through the gate oxide of a transistor. Obviously, this only happens in very small transistors, but it has been commonly observed for some time, now. The effect was first reported by Koga et al.,

The basic effect is that the trapped charge deposited by a single ion is enough to cause a small threshold voltage shift, which causes a small increase in subthreshold leakage current. This is sometimes enough to cause the failure of an NMOS memory cell, in either a DRAM or in a four-transistor SRAM cell, because these cells are very sensitive to small leakage currents. Oldham et al. included oxide thinning in their analysis, and concluded that stuck bits would tend to go away in thinner, future oxides. But this has not happened as quickly as one might have predicted from that analysis. The likely reason was pointed out by Loquet et al.,

The second topic is radiation-induced leakage current (RILC), which has been addressed in several papers by Paccagnella and co-workers. For thin enough oxides, electrons can tunnel directly from the substrate to the gate contact, and the level of such current that can be tolerated is an important constraint on the design of a circuit. RILC is a variation on this idea, in that a radiation-induced defect increases the substrate-to-gate tunnel leakage current. The basic idea is that electrons tunnel from the substrate to a trap state in the oxide, which is induced by radiation, and then the electron also tunnels from the trap to the gate contact. Of course, the leading candidate for this defect is the hole trap described in section III D. above. This effect is a consequence of oxide thinning, or scaling, and it is a gate oxide total dose effect. The impact is that the effects of the excess leakage current may have circuit implications, even though the threshold voltage shift is insignificant. RILC may also be related to breakdown in the oxide, because it has been argued that leakage currents contribute to oxide wearout, which in turn leads to breakdown. But a discussion of breakdown is beyond the scope of this paper.

The third topic is mobility degradation—even though the total shift in threshold voltage is small, defects close enough to the interface can reduce carrier mobility by acting as scattering centers. For interface traps, this effect was originally identified and analyzed by Sun and Plummer. Their work was extended to also include the effect of oxide trapped holes by McLean and Boesch. Generally, interface traps will cause a larger effect at late times, but the effect of hole traps is detectable in some experiments.

V. Radiation-Induced Field Oxide Leakage Currents

For much of the recent history of the NSREC, the main total dose problem in MOS technology has been damage to field oxide isolation structures, which has often meant LOCOS (local oxidation of silicon) structures, as shown in Fig. 26. Charge buildup in the thick field oxide, or the bird's beak region, or both, turns on a parasitic leakage path. Current flows from source-to-drain, outside the active gate region, as indicated in the Figure. Experimentally, the
The initial I-V characteristic of a transistor is shown, along with the small shift it undergoes when irradiated—the shift is small because the oxide is relatively thin. There is also a parasitic field oxide device curve, which is not visible experimentally, initially, because it is far to the right of the gate characteristic. But the field oxide is much thicker than the gate oxide, so the shift per unit dose is much larger, and the curve eventually shifts past the gate characteristic. In the illustration, the post-radiation field oxide curve is on the left side of the Figure. The measured I-V curve, post-radiation, is indicated by the broken line, labeled “combined.” The leakage current at 0 V increases from the pre-rad value by several orders of magnitude, which is often enough to cause functional failure of a circuit. Oldham et al. and Terrell et al. have reported on the radiation response, including annealing, of several commercial field oxides, and there is wide variation in the results. It is basically impossible to predict the response of such an oxide, without studying the particular oxide in question. Some of them do anneal on a reasonable time-scale, which can be very useful in some applications (in space, for example).

As feature sizes were scaled below about .35 μm, the LOCOS approach became less effective, and the emphasis in the industry shifted to trench isolation schemes. But trench processes vary so widely, it is difficult to make general statements about them, either. Both LOCOS structures and trenches can be hardened, with sufficient investment, but the details cannot generally be discussed in the open literature. The companies consider the hardening recipes proprietary, and they are sometimes also classified.

VI. Conclusions

We have reviewed the total ionizing dose radiation response of MOS materials, devices and circuits. Generally the response is very complex, with many different physical processes contributing. Each of these processes has a different time dependence, a different field dependence, and a different temperature dependence. Even though the overall, combined response is extremely complex, a high level of understanding has now been achieved, by isolating the different mechanisms, and by studying them one at a time. The study of these mechanisms has been a major theme of the NSREC throughout most of its forty-year history. The understanding that has been achieved stands as a major success for the conference and the conference community.

References


92. P. S. Winokur and M. A. Sokoloski, Comparison of Interface State Build-up in MOS Capacitors Subject to Penetrating and Non-Penetrating


114. F. J. Grunthaner and P. J. Grunthaner, Chemical and Electronic Structure of the SiO₂/Si Interface, Materials Science


137. N. S. Saks, M. G. Ancona, and J. A. Modolo, Radiation Effects in MOS


