SURFACE STATES

Semiconductor electronic devices comprise one or more semiconductor variants, insulators, metals or other conductors, optical materials, sensors, among others. All these materials are bounded by exposed surfaces or by interfaces with each other. The physical and electrical properties of semiconductors (and other materials) at their boundaries are often quite unlike their bulk properties. The performance of a transistor, especially, can be heavily influenced or degraded by surface features and properties. The smaller the device, the more it can be dominated by surfaces. This article reviews the basic physical, chemical, and electrical aspects of the particular semiconductor surface features known as *surface states.* The emphasis is on the semiconductor material aspects, but, in addition, device consequences of surface states will be briefly described.

The abrupt truncation of a pure and perfect semiconductor at its surface creates a severe interruption of the periodic potential that governs electron disposition throughout the bulk of the material. The wave-function spatial distributions and the energy levels or states of electrons at or near the surface are much different from those of electrons in the bulk. In the bulk of a perfect sample, the wave functions are three-dimensional in character, and may often be considered as extending throughout the material. The bulk electron energies are grouped into distinct energetically continuous bands—conduction electrons and valence electrons, separated by forbidden zones or a bandgap. In contrast, surface electrons are in effect limited to a two-dimensional distribution; and are in many cases, "zero-dimensional," that is, confined to a single atom or small cluster of atoms. The energy levels of such surface electrons are not confined to the conduction or valence bands, and some fall within the bandgap. Further, the energy levels of surface electron wave functions are usually discrete, that is, they do not comprise a band or continuum. These surface electron wave functions, then, constitute the features known as surface states.

Surface states have been extensively studied in idealized semiconductor systems (1), and they have a rich physics. In nonideal systems such as composite structures and electronic devices, surface states can have important effects on electronic or optical properties. These effects are due to the random or systematic trapping and detrapping of electrons, and are usually, but not always, harmful in applications.The first two decades of research were largely concerned with the basic scientific aspects; the most recent four decades have seen the focus shift toward the manifestations and control of surface states in commercial devices. In particular, over the past 20 years, study of surface states in metal-oxide-semiconductor (MOS) transistors has almost totally dominated this technical sub-area.

The older and more inclusive term *surface states* has been supplanted by more accurate expressions over the years. *Surface states* is a term now more properly reserved for semiconductors bounded by a vacuum or inert gas. For boundaries with solids, the term *interface states* is now commonly used; in the MOS community, *interface traps* is recommended, because of a dominant device interest. In this article, the most appropriate term for the topic at hand will be favored; but the underlying physical situation is more important than the semantic distinction.

SURFACE STATES ON IDEAL SEMICONDUCTORS

In the earliest serious consideration of surface states, Tamm (2) understood that imaginary components of the wave vector in the Bloch functions describing electron disposition in a periodic lattice could well define the special character of states that would exist at the semiconductor boundary. Bloch wave functions are of the form

$$\Psi_{\boldsymbol{k}}(\boldsymbol{r}) = u_{\boldsymbol{k}}(\boldsymbol{r}) \exp(i\boldsymbol{k} \cdot \boldsymbol{r}) \tag{1}$$

where Ψ is the wave function, $\boldsymbol{k}$ is the electron wave vector, u is the potential energy, and $\boldsymbol{r}$ is the electron position vector. A boundary-truncated u of the form shown in Fig. 1 was considered. The potential at the left of the boundary represents the energy of escape from the semiconductor. For the bulk of the material, only real values of $\boldsymbol{k}$ have any meaning; and the resultant wave functions constitute virtual continuum bands of levels separated by forbidden zones, and extend throughout the bulk. At the surface, the imaginary components of the wave function in themselves lead to real solutions of the equation, which comprise a group of individual levels, one within each forbidden zone. These wave functions decay exponentially from surface to interior, and thus define surface states in the system. These states are commonly called *Tamm states.*

Tamm states are meaningful only with a fairly large (and perhaps unrealistic) atomic spacing in the semiconductor, and they are predicted on the basis of an asymmetrical truncation of the potential. Shockley (3) visualized that a nearly symmetrical truncation would result in surface states with perhaps more realistic features, and would be restricted to small atomic spacings. The Shockley potential is also shown in Fig. 1. The resultant energy level structure as a function of atomic

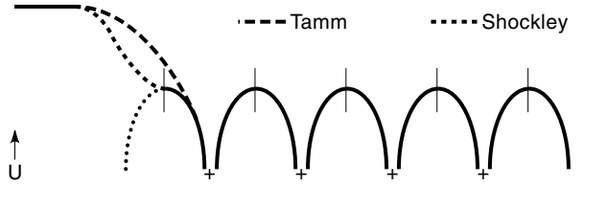

Figure 1. Electron potential energy U for a one-dimensional semiconductor lattice, with asymmetrical truncation used by Tamm and symmetrical truncation used by Shockley in their development of surface state models. Adapted from Ref. 3.

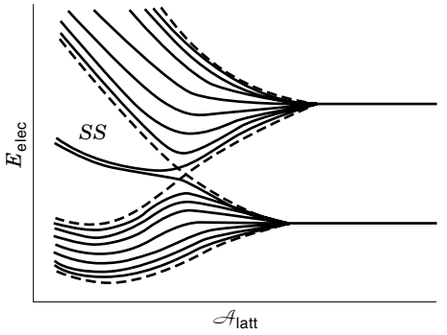

Figure 2. Electron energy states vs. lattice constant $\mathscr{A}_{latt}$ for one-dimensional lattice in the Shockley model for surface states. The paired midgap levels emerging at small atomic spacing are the Shockley states (SS). Adapted from Ref. 3.

spacing is shown in Fig. 2. At large spacing, where interatomic electron orbital overlap is weak, the model leads to two discrete levels, representing, say, widely separated impurity atoms. At a small lattice constant, there are two bands of levels separated by a gap, as in the bulk case, with two discrete levels in the gap, which are the surface states. These *Shockley states* offer a better correspondence with the attributes of real semiconductor lattices than do those of Tamm. Nonetheless, much work remains to be done in blending the earlier lattice-truncation approach with the more recent chemically based models emphasizing dangling orbitals, deviant structures, and bond energies of localized groups of atoms (4).

There have been numerous extensions, refinements, and more detailed physical models of surface states on ideal semiconductors. Nearly all lead to a density of states at the surface that is of order one localized level per surface atom, that is, about 10^{15} cm^{-2}. This number has been experimentally confirmed (5), but with difficulty; great care is needed to avoid inadvertent passivation.

MODELS OF SURFACE STATES IN NONIDEAL SYSTEMS

Clean, pure, exposed semiconductor surfaces are difficult to maintain in pristine condition, and are disadvantageous for device applications, partly because of the usually harmful effects of surface states. In application, semiconductor surfaces are almost always in intimate contact with another material, which may be electronically or optically active, or may be solely a surface passivant or insulating spacer. By far the dominant such composite is the Si–SiO_2 interface in the MOS transistor of contemporary integrated circuits (IC). The availability of well-controlled surfaces and the specific features and concepts emerging in research on surface states at the Si–SiO_2 interface have heavily influenced research, pulling it away from the classic physical theory and toward chemically oriented concepts and models.

ELECTRICAL ANALYSIS OF Si–SiO_2 INTERFACE TRAPS

The most common method for the study of interface traps in MOS structures is capacitance–voltage (C–V) analysis and variations thereof. In C–V, the capacitance of a test capacitor is measured as a function of dc bias with ac test signals of different frequency. (The dc bias also serves as the lowest test "frequency.") Important variations include measurement of conductance instead of capacitance; transient or pulsed application of test signals; time-variable thermal exposure; and optical stimulation. The analysis of the Si–SiO_2 interface is an enormous technical specialty (6), and cannot be treated beyond a simple but very important example here.

The capacitance as a function of bias voltage for high and low-frequency test signals applied to an ideal, trap-free MOS interface is shown in Fig. 3. The particular example is oxidized p-type Si. (The curves would be reversed left to right for n-type.) The high-frequency capacitance declines sharply near the crossover bias in the depletion region, between the majority-carrier-accumulation region on the left, and the inversion region on the right. The majority carriers (holes in p-type Si) can respond to both the low- and high-frequency signals with equal ease, yielding a large ac capacitance independent of frequency on the left side. On the other hand, the minority carriers are very sluggish, and cannot respond effectively to the high-frequency signal; thus their contribution to the MOS capacitance declines drastically on the right-hand side. The dip in the low-frequency curve near zero bias is due to the scarcity of surface carriers of either type in the depletion regime.

Interface traps introduce a source of capacitance in parallel with the carrier contribution. They can fill and empty at different biases which sweep their levels across the Fermi level, and can have various frequency responses. Their general effect is a distortion of the C–V curves, illustrated schematically in Fig. 3. From the distorted curves, the interface trap capacitance may be extracted, and the underlying trap density derived as a function of level position in the bandgap. (Trapping centers in the adjacent oxide near the interface may also exchange charge with the semiconductor at low frequencies, mimicking the behavior of the true interface traps. These oxide traps are not easily deconvolved from those sited in the interface, but they are tangential to the main topic here.)

Structure of Si–SiO_2 Interface Defects

It had long been assumed that a substantial part of interface traps in this system are due to dangling or nonbonded orbit-

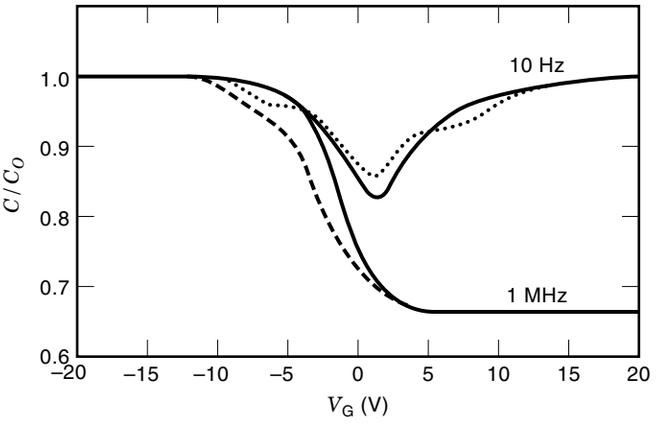

Figure 3. Schematic capacitance–voltage curves for an ideal metal-oxide-semiconductor (MOS) structure (solid) and with interface traps (dashed), showing high- and low-frequency response.

als on trivalent Si atoms, which had escaped oxidation. The existence of such trivalent Si centers at the interface has been demonstrated by electron paramagnetic resonance (EPR) (7,8). The interface on the (111) Si surface has given the most definitive picture, although the (100) surface is by far the most common in devices, due to a lower inherent number of traps. On the dry-oxidized (111) surface, there are typically between 10^{11} and 10^{13} traps per cm^2. About half of these traps are trivalent Si atoms with unbonded, dangling orbitals aligned perpendicular to the Si surface, as shown in Fig. 4. Dangling orbitals along the three tetrahedral bond directions inclined to the interface are not usually present in sufficient quantity to be detectable by EPR. The dangling-orbital defect on (111) Si is called the P_b center, and it is symbolized in the text as $\cdot Si{\equiv}Si_3$.

Only the (100) surface is used for MOS devices. Its features are empirically well characterized, and its manufacture is well controlled; however, it is more complex than (111), and not so well understood. It shows two types of P_b-like centers (9), Fig. 4. The first, termed P_{b0}, is very much like P_b on (111) in most respects, and occurs in two orientations, with a dangling orbital along one or the other of the two bond directions that exist on (100), both inclined to the surface. There is a second P_b-like center, termed P_{b1}, which also occurs in two symmetric orientations, but not precisely aligned with tetrahedral Si bond directions. The properties of P_{b1} are different in several important ways from those of P_{b0}. The structure of P_{b1} remains speculative at the time of writing; the structure shown is the earliest proposal. Modified models which are improvements in some respects have been proposed; but thus far, no model fits all the experimental evidence.

Correlation of P_b-Like Centers and Interface Traps

Verification of the role of P_b-like centers as a source of interface traps has been achieved by electrically controlled EPR and by close correlation with electrical determinations of the bandgap spectrum of interface trap density (D_{it}) (10). The D_{it} spectrum for dry-oxidized, unannealed (111) Si is shown in Fig. 5. Also shown in Fig. 5 is the bandgap occupancy change for P_b orbitals. The latter is derived from the EPR signal amplitude in a MOS structure with electrical bias applied to control the trapping or detrapping of electrons. The EPR peaks correspond quantitatively to the peaks in the D_{it} spectrum; if $d[P_b]/dE$ is subtracted from D_{it}, a smooth U-shaped continuum of traps remains. The same measurements on (100) correlate equally well (9), and in richer detail, because two types of P_b-like centers are usually present.

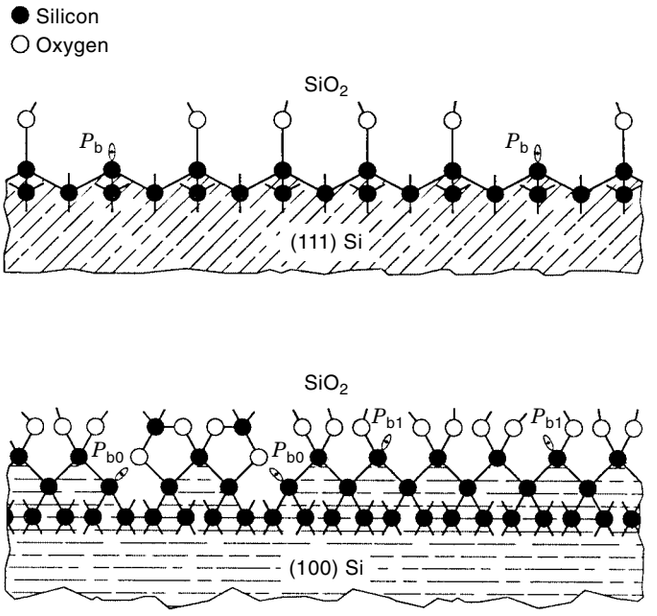

Figure 4. Dangling orbital defects on two principal crystal faces of Si, as derived from electron paramagnetic resonance studies. The structure shown for P_{b1} is the earliest model proposed, and it is in accord with some important features of the center; there is no model which fits all experimental features of the center.

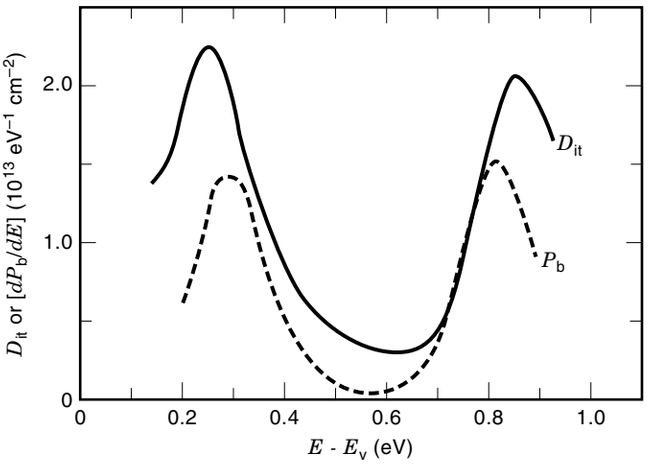

Figure 5. Bandgap spectrum of interface traps on oxidized (111) Si as determined by C–V analysis (D_{it}) and EPR (P_b).

The quantitative correlation of P_b with D_{it} has also been observed in samples with widely variable interface trap densities produced in systematic variation of oxidation conditions (9), in passivation by hydrogen and depassivation (7), and in generation of traps by particle or photon irradiation (11). Thus it has been well proven that P_b-like centers are major sources of interface traps. There are other trap centers, produced by radiation or hot-electron injection, with discrete levels (12); and the pervasive U-shaped continuum traps, which have not been observable by EPR (10). The latter are most probably due to the nonstandard character of bonds between the P_b central Si and its neighbors; they are present in quantity roughly comparable to the dangling-orbital traps. The nature of most damage-center traps is entirely speculative at this time. An additional large number of interface states may arise from Si–Si wrong-bond energies associated with oxidized Si sites. These will be very close to the band edges; they will escape detection by virtually all techniques, and will have no significant effect on devices.

EFFECTS ON DEVICE-RELATED PROPERTIES

Degradation of MOS Devices by Interface Traps

Traps at the Si–SiO_2 interface, between roughly $E_V + 0.2$ eV and $E_c - 0.2$ eV, can have several harmful effects on MOS transistors (6). First, the threshold or turn-on/turn-off voltage of the device can be reduced or increased, either condition being troublesome, especially if the fabrication processing is variable. In p-type Si (n-MOS transistors), traps become neg-

atively charged and thus oppose the influence of oxide positive charges; in n-type Si (p-MOS transistors), the traps become positive and thus enhance the effect of oxide charge. Second, transistor drain breakdown is affected—worsened for n-channel devices, and perhaps unexpectedly, improved for p-channels.

Third, the device transconductance or gain is reduced by interface traps. This occurs through two mechanisms. Applied signal voltage is forced to move the channel current carriers, and simultaneously to pump charge in or out of the traps. In addition, the channel conductance is reduced by carrier recombination via interface traps. Fourth, a further reduction in device sensitivity arises from the *generation* of carriers via traps near the drain junction of transistor (as compared with the *recombination* losses in the channel). The ensuing currents respond to the ac signal in a counterproductive way, thereby reducing sensitivity to small signals, and wasting power.

Fifth, and finally, flicker ("1/f") noise down to 10^4 Hz is generated by the randomly induced charging and discharging of interface traps. Decreasing channel volume increases the seriousness of this noise, an unfortunate circumstance for the reduction of device size.

Effects on Other Devices

Charge-coupled arrays find use when information packets must be transferred from one device to the next in a timed sequence. In one embodiment, they comprise a series of MOS diodes separated from the substrate Si by a common sheet of oxide. Interface traps charge and discharge as control pulses are applied, and some of them may seriously lag the transfer of the charge packets (13). The data packets are thus distorted or lost. Solar cells of the metal-insulator-semiconductor (MIS) type have a superficially similar structure, but the many conducting electrodes comprise tunneling-assisted Schottky diodes with a very thin SiO_2 layer. Excessive recombination by interface traps reduces light-to-electricity conversion efficiency (13). Solar cells made of thin films of amorphous Si suffer from P_b-like traps at grain boundaries (and others); the effects can be minimized by making the devices very thin, so that rapid carrier transit reduces opportunities for coupling with traps.

Devices made of other semiconductors can also suffer from interface traps. For example, CdS solar cells, which are structurally very similar to Si MIS cells, have gross amounts of interface states (13). Another compound semiconductor technology is that of GaAs microwave power amplifier transistors, where serious interelectrode leakage and breakdown are partly due to very high density of surface states. The analog of the complementary MOS (CMOS) transistor has not been possible with GaAs. There is no native oxide or simple thermal oxide comparable to SiO_2, and the surface and interface states are very hard to passivate satisfactorily. The traps are so numerous that they prevent inversion of the n-type surface, and they promote surface leakage and breakdown. In the GaAs/AlGaAs laser diodes, interface traps are a problem. In the solar cell and GaAs optical applications, the necessary sophistication of interface trap control is much less than in the IC MOS transistor.

A final example of device effects is the classic metal-semiconductor contact, Fig. 6. The effect of interface traps on this simple device (13) contributed to the early understanding of

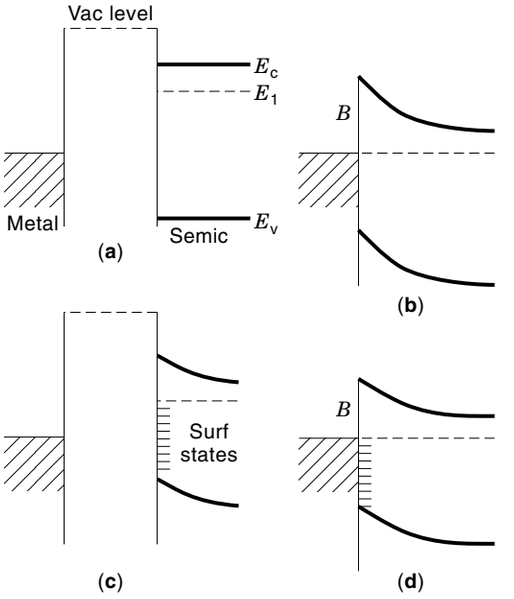

Figure 6. Buffering of energy barrier B at metal-semiconductor contact. (Left) Materials separated. (Right) Materials in contact. The charging of surface states shields the bulk of the semiconductor. The barrier is thus controlled by the surface, rather than the work function difference.

basic semiconductor and surface state physics. In the ideal trap-free contact, when the Fermi levels of the metal and semiconductor are coincident, the band edges are bent at the interface to create a barrier that is the algebraic difference of the electron work functions of the two materials. In the early days, many test junctions were made with various metals and semiconductors. Yet, there was only a small, unsystematic variation in barrier. The mystery was finally explained by the presence of surface states, which charge/discharge as the Fermi coincidence is being developed. If present in sufficient numbers, the traps can almost completely shield the semiconductor bulk, and thereby buffer the barrier against the effect of comparative work functions; the height is set mainly by the surface properties of the semiconductor (14).

PROCESSING-RELATED CHEMISTRY

The most serious generation of the very harmful midgap range interface traps in MOS structures occurs during thermal oxidation of the silicon. The original unoxidized surface, if subjected to a fresh etch, quickly reconstructs in vacuum to eliminate dangling orbitals. This precludes most of the traps in the middle region of the bandgap, roughly $E_V + 0.2$ eV to $E_c - 0.2$ eV. (The situation also prevents easy verification of the one-state-per-atom principle; most of the states are presumably packed tightly against the band edges, and not measurable by the usual methods.) If exposed to air, a thin "native" oxide forms, which breaks the reconstruction for most of the surface atoms, perhaps redistributing the "wrong-bond-energy" states near the band edges. Thermal oxidation extends this process. Although expected to be a trap-passivating agent, oxidation creates sources of seriously harmful traps in the middle region of the gap. As the oxide thickens, the oxide matrix gains strength; because Si and SiO_2 lattices do not

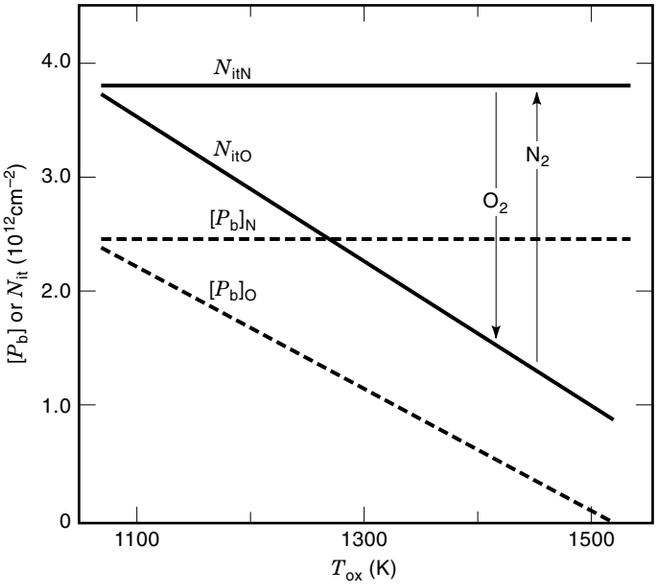

Figure 7. Interface trap densities (solid curves) after dry oxidation of Si, fast-cooled in O (N_{itO}) or slow-cooled in N (N_{itN}). The N_{it}may be recycled between upper and lower curves by repeated exposure to N or O, respectively. The correlated result for P_b is shown by the dashed curves.

quite match, P_b centers are produced where a Si atom is not oxidized.

The N_{it} and $[P_b]$ vary with the temperature of the oxidation, as shown in Fig. 7, the well-known Deal oxidation triangle (9,15), shown here for (111)Si. Thermal annealing of the sample in an inert (e.g., Ar or N_2) atmosphere produces a uniform higher density of traps; reexposure to oxygen restores the initial N_{it} or $[P_b]$ characteristic of the oxidation temperature. Although lattice mismatch is clearly a factor in setting the stage for trap existence at the Si–SiO_2 interface, it does not offer a convincing quantitative explanation; and it provides no good rationale for the effect of the inert anneal. The trap densities shown would disable an MOS device; the universally used (100)Si surface offers inherent N_{it} lower by a factor of three, but still much too high.

Interface trap densities are passivated in IC practice by anneal in hydrogen, usually in the form of H_2, but in former days, as atomic H produced by the reaction of Al electrode films with outer-oxide hydroxyl groups. Very low N_{it}, of order 10^9 cm^{-2}, is routinely achieved. The P_b centers are apparently altogether eliminated in the reaction

$$\cdot Si{\equiv}Si_3 + H_2 \rightarrow H{-}Si{\equiv}Si_3 + H\uparrow \qquad (2)$$

The associated U-shaped-continuum traps are reduced proportionally; thus hydrogen yields a very good passivation of traps throughout the significant middle range of the bandgap.

INTERFACE TRAPS GENERATED BY RADIATION OR ELECTRICAL STRESS

The Si–SiO_2 structure can be damaged by energetic radiation or by injection of hot electrons from Si into the oxide under excessive operating voltages (12). The greater part of the damage from radiation is not due to direct impingement on the interface, but rather, due to sequential physicochemical processes. Positive bias on the overlying electrode makes the damage much worse. Despite much research, the data accumulated have not allowed any unequivocal model mechanism to be developed. The N_{it} generated often arises from P_b centers, but more often, includes more or fewer of other centers which are not EPR-visible, and remain unidentified. Radiation resistance is much improved by holding post-oxidation temperatures to less than 900°C, which limits oxygen loss and chemical reduction of the near-interface oxide. In addition, hydrogen seems to be a factor in radiation susceptibility; radiation-hard processing tacitly includes control of H and H_2O in proprietary recipes.

Another important source of degradation is the negative-bias-temperature instability (NBTI) (15). The application of negative gate voltage at an elevated temperature produces this effect, a generation of P_b or N_{it} in H-passivated interfaces, along with oxide positive charges, Fig. 8. This effect requires H_2O near the interface, which can be present after steam oxidation or excessive exposure to atomic H. The problem has been much easier to eliminate in IC processing than has the radiation-damage susceptibility; much less research has been done (9). The NBTI is much worse in steam-grown oxides; and like radiation susceptibility, is controlled in practice by reduction of hydrogenous species in the oxide.

NATURE AND FEATURES OF SURFACE STATES ON OTHER SEMICONDUCTORS

Compared to Si, the underlying commercial applications for other semiconductors are so much smaller in scope that a detailed consideration of the limited surface-state research is not warranted in a review of this size (16). Possibly the most important semiconductor after Si is GaAs, which finds important specialty application in fast amplifiers and in electroopti-

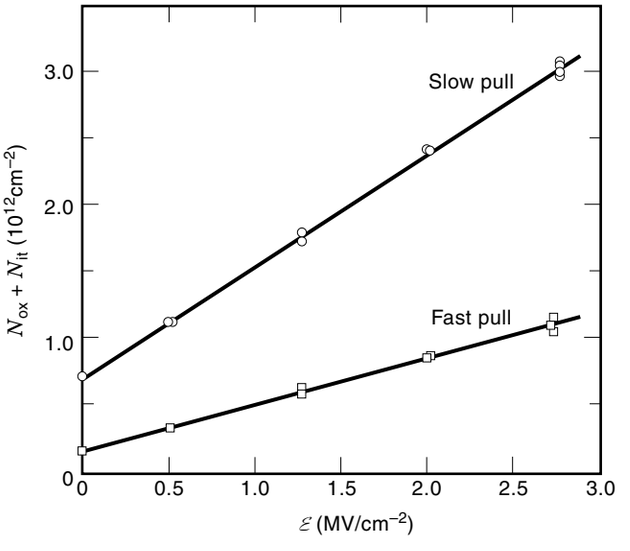

Figure 8. Negative-bias-temperature instability in MOS structures at 700 K. The gate bias produces an electric field of strength $\mathcal{E}$. Though not separated here, the resultant densities of interface traps N_{it} and oxide charges N_{ox} are equal.

cal devices. In addition to the very high density of surface states, the nature of the surface varies greatly with crystal orientation; and GaAs is variably reactive with metals and other materials with which it is placed in contact. Further, the surfaces of GaAs show no EPR signal; this precludes the straightforward identification of surface states that was possible on Si. The lack of EPR is a critical factor hindering surface state characterization, not only for GaAs, but for most other compound semiconductors.

BIBLIOGRAPHY

1. A. Many, Y. Goldstein, and N. B. Grover, *Semiconductor Surfaces,* Amsterdam: North Holland, 1965.
2. I. E. Tamm, Über eine mögliche art der elektronenbindung an kristalloberflächen, *Phys. Z. Sowjetunion,* **1**: 733, 1933.
3. W. Shockley, On the surface states associated with a periodic potential, *Phys. Rev.,* **56**: 317, 1939.
4. H. Flietner, Interface state distributions and the center of reconstruction model, In W. Eccleston and M. Uren (eds.), *Insulating Films on Semiconductors 1991,* Bristol: Adam Hilger, 1991.
5. F. G. Allen and G. W. Gobeli, Work function, photoelectric threshold, and surface states of atomically clean silicon, *Phys. Rev.,* **127**: 150, 1962.
6. E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology,* New York: Wiley, 1982.
7. P. J. Caplan et al., ESR centers, interface states, and oxide fixed charge in thermally oxidized silicon wafers, *J. Appl. Phys.,* **50**: 5847, 1979.
8. K. L. Brower, Electron paramagnetic resonance studies of Si–SiO_2 interface defects, *Semicond. Sci. Technol.,* **4**: 970, 1989.
9. E. H. Poindexter, MOS interface states: overview and physicochemical perspective, *Semicond. Sci. Technol.,* **4**: 961, 1989.
10. E. H. Poindexter et al., Electronic traps and P_b centers at the Si/SiO_2 interface: band-gap energy distribution, *J. Appl. Phys.,* **56**: 2844, 1984.
11. P. M. Lenahan et al., Radiation-induced trivalent silicon defect buildup at the Si–SiO_2 interface in MOS structures, *IEEE Trans. Nucl. Sci.,* **NS-28**: 4105, 1981.
12. T. P. Ma, Interface trap transformation in radiation or hot-electron damaged MOS structures, *Semicond. Sci. Technol.,* **4**, 1061, 1989.
13. S. M. Sze, *Physics of Semiconductor Devices,* 2nd ed., New York: Wiley, 1981.
14. J. Bardeen, Surface states and rectification at a metal semiconductor contact, *Phys. Rev.,* **71**: 717, 1947.
15. B. E. Deal et al., Characteristics of the surface state charge of thermally oxidized silicon, *J. Electrochem. Soc.,* **114**, 266, 1967.
16. C. Barret et al., Stoichiometry effects on Schottky barrier and interface states in GaAs(001)/Al system, in M. Averous (ed.), *Physics of Semiconductors,* Amsterdam: North Holland, 1983.

EDWARD H. POINDEXTER
Army Research Laboratory