Solid-State Electronics Pergamon Press 1962. Vol. 5, pp. 285–299. Printed in Great Britain

AN INVESTIGATION OF SURFACE STATES AT A SILICON/SILICON OXIDE INTERFACE EMPLOYING METAL–OXIDE–SILICON DIODES*†

L. M. TERMAN

I. B. M. Yorktown Heights, New York

(*Received* 16 *October* 1961; *in revised form* 8 *February* 1962)

Abstract—A new solid-state device, the M–O–S diode, of which an oxidized silicon surface is an integral part, is introduced, and a theory for its operation in the absence of surface states is obtained. The capacitance of this device may be considerably more voltage sensitive than that of a p–n junction. The existence of surface states with non-zero relaxation times is introduced into the theoretical model. It is shown that the states may increase the capacitance of the device, as well as affect the proportion of applied voltage which appears across the silicon. A small-signal equivalent circuit is derived which includes the effect of the surface states. It is also shown that a comparison of the theoretical capacitance vs. voltage curve without states and a measured high-frequency capacitance vs. voltage curve may be used to obtain the distribution of all states, regardless of their time constants.

Results are given of measurements and calculations on two M–O–S diodes having different surface treatments before oxidation. Both surfaces have a total density of about 3×10^{12} states/cm^2. In both cases, the distribution of states is continuous and has its highest peak about 100 mV above $E_F(0)$, the position of the Fermi level at the silicon surface if there is no voltage drop across the silicon The time constants of the states extend from 10^{-8} sec to longer than 10^{-2} sec. There is a tendency for states located at deeper energy levels to have longer time constants, but some of the states in the high density of states above $E_F(0)$ have long time constants. The distribution of time constants with energy level is somewhat different for the two surfaces.

A comparison is made between the distribution of states obtained here with the distribution reported by others working in the field. The results are similar in density and location of the peaks of the distribution reported here, but differ in that some other sources report a discrete distribution.

Résumé—On introduit un nouvel élément à l'état solide, la diode M–O–S, dont la surface au silicium oxydée fait partie intégrale, et on obtient une théorie pour definir son mécanisme en l'absence d'états de surface. La capacité de cet élément peut être considérablement plus sensible au voltage qu'une jonction p–n. L'existence d'états de surface ayant des durées de lâchement définies est introduite dans le modèle théorique. On démontre que les états peuvent augmenter la capacité de l'élément aussi bien qu'affecter la proportion de la tension appliquée apparaissant à travers le silicium. On arrive à un circuit équivalent à petit signal qui inclut l'effet des états de surface. Il est aussi démontré qu'une comparaison de la capacité théorique avec la courbe de tension sans états et une capacité mesurée à haute fréquence avec la courbe de tension peuvent être utilisées pour obtenir la distribition de tous les états indépendemment des constantes de temps.

On donne aussi des résultats de mesures et calculs de deux diodes M–O–S ayant subi de différents traitements avant l'oxydation. Les deux surfaces ont une densité totale d'environ 3×10^{12} états/cm^2. Dans les deux cas la distribution des états est continue et a sa crête maximum à environ 100 mV au-dessus d' $E_F(0)$, la position du niveau Fermi à la surface du silicium s'il n'existe pas de chute de tension à travers le silicium. Les constantes de temps des états s'étendent de 10^{-8} à plus de 10^{-2} sec. Il existe une tendance pour les états situés à des niveaux d'énergie plus profonds d'avoir de plus longues constantes de temps mais quelques états dans la région de forte densité d'états au-dessus d'$E_F(0)$ one de longues constantes de temps. La distribution de ces constantes en fonction du niveau d'énergie est plutôt differente pour les deux surfaces.

Une comparaison est faite entre la distribution des états obtenus ici avec la distribution rapportée

* Prepared under Office of Naval Research Contract Nonr 225 (24), NR 373, 360, and jointly supported by the U.S. Army Signal Corps, the U.S. Air Force, and the U.S. Navy (Office of Naval Research).

† The work contained in this report was done at Stanford University, Stanford, California.

par d'autres travaillant dans cette branche. Les résultats sont similaires en ce qui concerne la densité et la location des crêtes de la distribution rapportée ici mais certaines autres sources rapportent un distribution distincte.

Zusammenfassung—Die M–O–S-Diode, ein neues Festkörpergerät, enthält als wesentlichen Bestandteil eine oxydierte Siliziumoberfläche. Für ihre Wirkungsweise in Abwesenheit von Oberflächenzuständen wird eine Theorie entwickelt. Die Kapazität dieses Geräts kann eine viel stärkere Spannungsempfindlichkeit besitzen als die eines p–n-Übergangs. In das theoretische Modell wird die Existenz von Oberflächenzuständen mit nicht verschwindender Relaxationszeit eingeführt. Es ergibt sich, dass diese Zustände die Kapazität des Geräts erhöhen können und ausserdem den am Silizium erscheinenden Bruchteil der angelegten Spannung beeinflussen. Ein Ersatzschaltbild für ein schwaches Signal wird gegeben, wobei der Effekt der Oberflächenzustände berücksichtigt wird. Ferner wird gezeigt, dass ein Vergleich der theoretischen Kapazität/Spannungs-Kurve in Abwesenheit der Zustände und einer durch Messung erhaltenen Kapazität/Spannungs-Kurve bei Hoch frequenz dazu dienen kann, die Verteilung der sämtlichen Zustände, unabhängig von ihrer Zeitkonstanten, zu bestimmen.

Die Ergebnisse der Messungen und Berechnungen beziehen sich auf zwei M–O–S-Dioden, deren Oberflächen vor der Oxydation verschiedener Behandlung unterworfen wurden. Beide Oberflächen haben eine Gesamtdichte von etwa 3×10^{12} Zuständen/cm^2. Bei beiden ist die Verteilung der Zustände kontinuierlich und hat einen Höchstwert etwa 100 mV oberhalb $E_F(0)$, der Position des Fermi-Niveaus an der Siliziumoberfläche, wenn am Silizium kein Spannungsabfall vorhanden ist. Die Zeitkonstanten der Zustände betragen von 10^{-8} bis über 10^{-2} sec. Die bei tieferen Energieniveaus liegenden Zustände haben meist längere Zeitkonstanten, doch haben auch einige der Zustände oberhalb von $E_F(0)$, wo die Dichte gross ist, lange Zeitkonstanten. Die Verteilung der Zeitkonstanten auf die Energieniveaus ist bei beiden Oberflächen etwas verschieden.

Die sich hier ergebende Verteilung der Zustände wird mit Ergebnissen anderer Arbeiten verglichen. Die Ergebnisse zeigen Übereinstimmung in Bezug auf die Dichte und die Lage der Höchstwerte der Verteilung, wie hier angegeben, doch ergibt sich eine Abweichung, da die anderen Quellen eine diskrete Verteilung nachwiesen.

SURFACE states due to the discontinuity at a clean silicon surface were first postulated by TAMM[1] and were analysed in a more general model by SHOCKLEY.[2] Similarily, at an oxidized semiconductor surface, the transition from ordered semiconductor crystal lattice to the oxide layer represents a major discontinuity which will result in dangling or unsaturated bonds. In an energy-band diagram, the effect of the unsaturated bonds as well as of impurities incorporated on the surface will appear as localized allowed energy levels in an otherwise forbidden band at the silicon/silicon oxide interface. These allowed energy levels are possible sources of charge at the semiconductor surface, and thus may strongly affect its properties.

In studying surface states, it is desirable to find the density of states vs. their location in the forbidden band, as well as the relaxation time of the charge in the states. The latter is a measure of how rapidly the charge redistributes itself if equilibrium conditions are disturbed.

This paper presents a new method of investigating surface states. It involves obtaining a theory for a new solid-state device, the M–O–S diode, of which an oxidized surface is an integral part, and measuring its characteristics. Results obtained by this method are presented. It should be emphasized, however, that owing to the pronounced sensitivity of surface states to surface preparation, oxidation process, and past history of the crystal, the results are not necessarily characteristic of oxidized silicon in general, but only of the specific treatments utilized.

NOTATION*

C_{ox}	oxide capacitance
ϵ_{ox}	dielectric constant of the oxide
W_{ox}	width of the oxide
C_D	silicon surface capacitance
Q_{net}	net excess charge in silicon surface
V_s	voltage drop across the silicon
V_{ox}	voltage drop across the oxide
E_{ox}	field in the oxide
ϵ_s	dielectric constant of silicon
Q_s	total charge in surface states
Q_{s0}	total charge in surface states at equilibrium
ω	angular frequency
q	electronic charge
k	Boltzmann's constant
T	absolute temperature
p_0	equilibrium hole concentration
p	hole concentration
C_s	capacitance due to surface states
R_s	loss resistance due to surface states

* Units are c.g.s.

R_{SL}	series loss resistance due to bulk silicon
R_{PL}	parallel leakage resistance through oxide and silicon surface
$E_F(S)$	position of the Fermi level at the surface of the silicon
$E_F(0)$	position of the Fermi level below the bottom of the conduction band deep inside the silicon
N_D	density of donors
n_i	intrinsic electron density
n_0	equilibrium electron density
τ	time constant of surface states
ϕ_0	Fermi level
ψ_s	electrostatic potential at the silicon surface
ψ_0	electrostatic potential deep inside the silicon
β	q/kT

I. THE M–O–S DIODE

A. *Surface states absent*

The M–O–S diode consists of layers of metal, silicon oxide and silicon in contact. Physically, it appears as in Fig. 1(a), and a diagram of possible energy levels between the metal, silicon oxide and silicon is illustrated in Fig. 1(b). Thermal oxide on ordinary silicon produces sufficient donors at the silicon/silicon oxide interface to make most

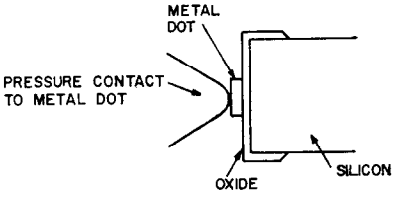

FIG. 1(a). Physical arrangement of the M–O–S diode.

surfaces n-type. In addition, for the purposes of illustration, it is assumed that the work function of the metal is less than the work function of the silicon. Constancy of the Fermi level, coupled with the assumed relative magnitudes of the work functions, results in the energy-level diagram of

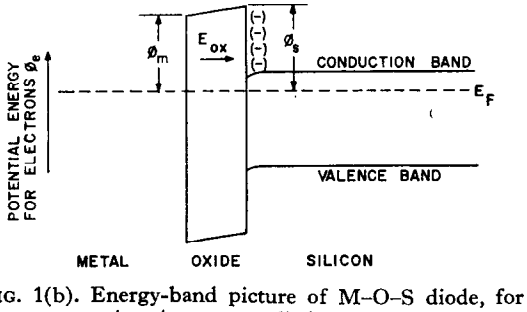

FIG. 1(b). Energy-band picture of M–O–S diode, for $\phi_s > \phi_m$, zero applied voltage.

Fig. 1(b) for zero applied voltage. The downward bend of the energy bands in the silicon at the oxide–silicon interface gives rise to an accumulation of negative charge in the silicon at the interface. As a result, a field must exist in the oxide which terminates on this negative charge. In the energy-level diagram it is assumed that the potential energy for electrons in the conduction band of the silicon oxide is high, i.e. silicon oxide is an insulator and presents a barrier to electrons.

Consideration of Fig. 1 will reveal the presence of the basic requirements for a capacitor, i.e. two conducting regions (the metal and silicon) separated by an insulating region (the oxide and any depletion region existing at the silicon surface). The total differential capacitance of the device will consist of two capacitances in series: that of the oxide layer, and that of the silicon surface. The oxide capacitance may be determined solely from geometry, and is given by*

$$C_{ox} = \frac{\epsilon_{ox}}{W_{ox}} \tag{1}$$

where ϵ_{ox} is the dielectric constant of silicon oxide and W_{ox} is the thickness of the silicon oxide.

The capacitance of the silicon surface cannot be expressed in as simple a manner. It depends upon the distribution of net charge at the silicon surface as a function of the voltage drop across the silicon, and must be calculated from the expression

$$C_D = \frac{dQ_{\text{net}}}{dV_s} \tag{2}$$

where Q_{net} is the net excess charge at the silicon surface and V_s is the voltage drop across the silicon.

Once the charge versus voltage characteristic of the silicon surface is known, it becomes a straightforward matter to calculate the capacitance versus voltage characteristics of the diode. The total capacitance will be the series combination of the oxide and silicon surface capacitance. The total voltage drop will be the sum of the drops across the silicon and the oxide. For a given voltage drop across the silicon there will be a known net surface

* Capacitance is expressed as capacitance per unit area in this section.

charge in the silicon. It has been assumed that no surface states are present, thus the D vector must be constant across the oxide–silicon interface. From this the field in the oxide may be determined. Since no charge is present in the oxide, the field will be uniform across it, and the voltage drop will be

$$V_{ox} = W_{ox}E_{ox} = W_{ox}\frac{\epsilon_s}{\epsilon_{ox}}Q_{\text{net}} \qquad (3)$$

where ϵ_s is the dielectric constant of the silicon, and the other symbols are as defined previously.

The asymptotic behavior of a capacitance vs. voltage curve of the M–O–S diode can be seen from simple considerations. If at zero bias the bands are as in Fig. 1(b), there will be an accumulation of electrons at the silicon surface. Consequently, a field will exist in the oxide that will terminate on these electrons. The capacitance will be very nearly the same as the capacitance of the oxide layer. If the silicon is made negative, there will be a further bending of the bands that will result in further accumulation of electrons at the silicon surface. This will continue until breakdown in the silicon oxide is reached.

If voltage is applied with such a polarity as to make the silicon positive, the effect will be to remove some of the charge from the surface of the silicon at the oxide–silicon boundary, thus weakening the field in the oxide. The energy bands in the silicon extend straight out to the surface at some voltage V_1. At this point the applied voltage

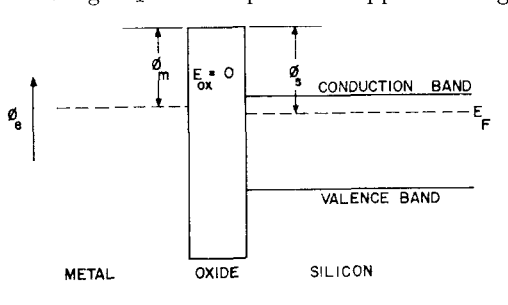

FIG. 1(c). Energy-band picture of M–O–S diode for $V_{\text{app}} = V_1$.

has acted to remove all the charge on the oxide–silicon interface which had been present at zero bias. The energy-band picture will appear as in Fig. 1(c), and the field in the oxide will have decreased to zero. V_1 is typically a volt or so, and depends upon the oxide thickness, metal work function, and semiconductor doping.

Increase of the bias voltage beyond V_1 will result in the growth of a depletion region in the silicon. A field will arise in the oxide which comes from the uncovering of the bound positive charges in the depletion region. This field will be in a direction opposite to the field intrinsic in the oxide at zero bias. This is shown in Fig. 1(d). If the applied field is sufficiently great so that a depletion region that

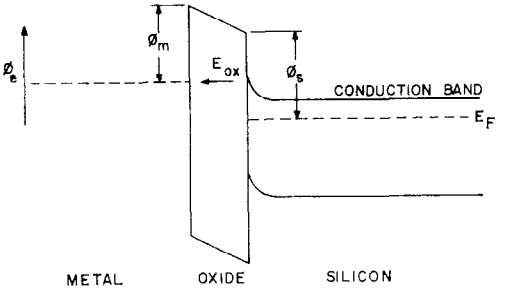

FIG. 1(d). Energy-band picture for $V_{\text{app}} > V_1$.

is several Debye lengths wide is formed, the differential capacitance becomes

$$C = \left(\frac{W_s}{\epsilon_s} + \frac{W_{ox}}{\epsilon_{ox}}\right)^{-1} \qquad (4)$$

This is the familiar form for two capacitances connected in series, one of magnitude ϵ_{ox}/W_{ox}, the other of magnitude ϵ_s/W_s. The first capacitance is the oxide capacitance discussed previously, while the second is the capacitance of the silicon surface due to the depletion region. With increasing voltage the depletion region widens, resulting in a decrease in the depletion capacitance, and an accompanying decrease in the total capacitance.

The decrease in capacitance of the M–O–S diode as the depletion region in the silicon widens is so analogous to the decrease in capacitance of a p–n junction as the depletion region widens as to require no further elaboration. It is interesting, however, to compare the capacitance–voltage characteristics of an M–O–S diode and a p–n junction. For large reverse biases (corresponding to wide depletion regions) the capacitance–voltage characteristics of the device will be almost entirely determined by the depletion capacitance, since it

will be a small capacitance in series with the much larger oxide capacitance. This is exactly the same case as exists in the p–n junction, where the capacitance–voltage characteristic is always determined by the depletion region. Hence, the characteristics of the M–O–S diode and a p–n junction of similar area and physical parameters should coincide for larger reverse voltages. The asymptotic behavior then is the approach to the p–n junction for large negative bias on the metal and to a constant value equal to the oxide capacitance at zero and positive bias on the metal.

The zero-bias capacitance of the M–O–S diode is determined by the thickness of the oxide, which can be made on the order of a hundred angstroms thick in practice, while the zero-bias capacitance of a p–n junction is determined by the inherent width of the depletion region which, for the resistivities of material used to obtain a reasonably rapid change of capacitance with voltage, may be on the order of several thousand angstroms. The M–O–S diode obtains its extended range of capacitance variation by going into higher capacitance at low voltage. A comparison of the capacitance–voltage characteristic of an M–O–S diode and a p–n junction are given in Fig. 2. V_1 was taken arbitrarily as equal to 0·5 V. A more complete analysis is necessary to compare the merits of the M–O–S diode and the p–n junction for specific circuit applications.

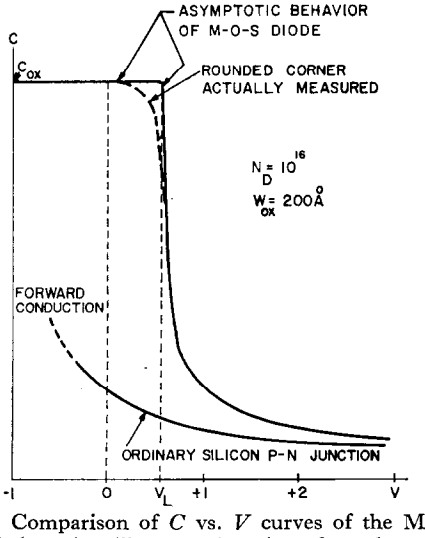

FIG 2. Comparison of C vs. V curves of the M–O–S diode and a silicon p–n junction of equal area.

The correct capacitance of the silicon surface is calculated, by means of the work of GARRETT and BRATTAIN,[3] in the Appendix, subject to the following three assumptions:

(1) All impurity atoms are ionized everywhere.

(2) The semiconductor is not degenerate.

(3) The semiconductor is at room temperature.*

The resulting expression for the capacitance of the silicon surface is

$$C_D = \sqrt{\frac{q\epsilon_s n_0 \beta}{2}} \frac{\exp(\beta\Delta\psi)-1}{[\exp(\beta\Delta\psi)-\beta\Delta\psi-1]^{1/2}} \qquad (5)$$

where $\beta = q/kT$, n_0 is the electron concentration in the silicon at room temperature and $\Delta\psi$ is given by $\psi_s = \psi_0 + \Delta\psi$, where ψ_s is the potential at the surface of the silicon, and ψ_0 is the potential at some point deep in the interior of the silicon. The quantity $\Delta\psi$ is thus the potential drop across the silicon.

From equations (3) and (A.4), the voltage drop in the oxide is

$$V_{ox} = \frac{W_{ox}}{\epsilon_{ox}} \sqrt{\frac{q\epsilon_s n_0}{2\beta}} [\exp(\beta\Delta\psi)-\beta\Delta\psi-1]^{1/2} \qquad (6)$$

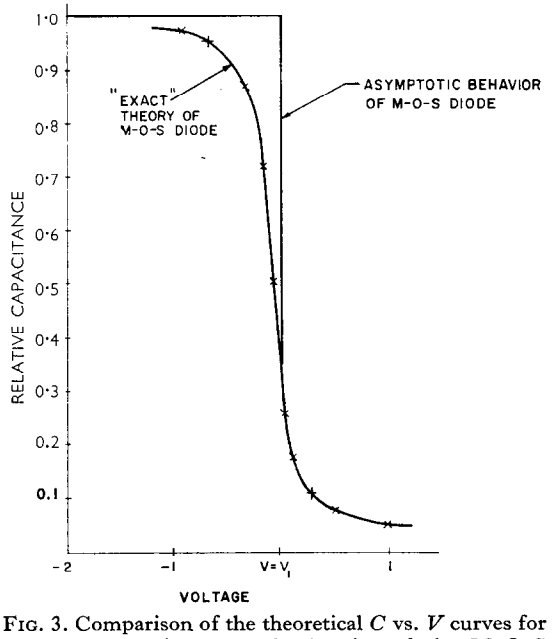

FIG. 3. Comparison of the theoretical C vs. V curves for the "exact" and asymptotic theories of the M–O–S diodes. Calculated for $N_D = 10^{15}$, $W_{ox} = 200$ Å.

* This assumption is needed to locate the Fermi level with zero voltage drop across the silicon.

A curve calculated from a doping level of 10^{15} donors/cm^3, and an oxide thickness of 200 Å is presented in Fig. 3. The disparity between the "exact" theory and the asymptotic behavior is quite evident. The "exact" theory has the noticeably rounded corner which is observed in practice.

B. *Effect of surface states*

The biggest complication to the theory for the M–O–S diode thus far presented arises from the existence of surface states. Since by their very nature the surface states exist only at the extreme proximity of the surface of the silicon, they will be represented by a delta function (in space) of charge at the silicon/silicon oxide interface. The space-charge densities and displacement current distributions will appear as in Fig. 4. It is to be expected that a time constant τ will be associated with the charging and discharging of the states, and that accompanying this time constant will be an energy loss mechanism.

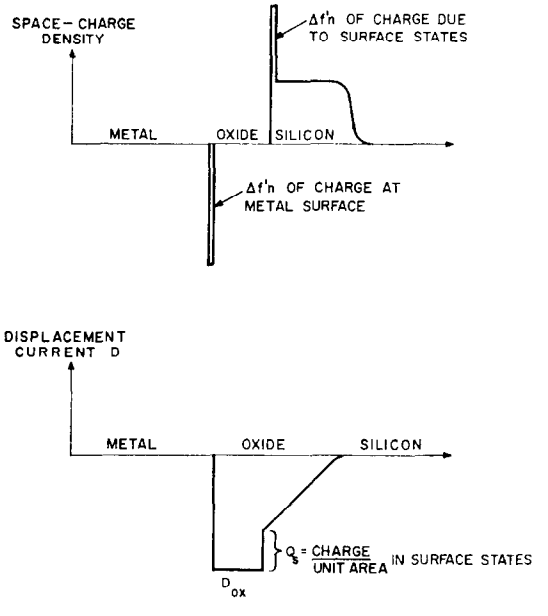

FIG. 4. Space-charge density and displacement current distributions for the M–O–S capacitor.

If $Q_s(t)$ is defined as the charge in the surface states at time t, and $Q_{s0}(t)$ as the charge that would be in the states at time t if equilibrium conditions were reached, the net rate of change of charge in the surface states is

$$\frac{dQ_s}{dt} = \frac{Q_s(t) - Q_{s0}(t)}{\tau} \tag{7}$$

(Note that Q_{s0} may be a function of time. It will be so if the surface potential of the silicon itself varies with time.) If the potential at the surface of the silicon is disturbed, the total charging current to the surface of the silicon will be the sum of two components, one which goes to charging the depletion region in the silicon, and one which goes to charging the surface states. Thus

$$C_{\text{total}} = \left.\frac{dQ}{dV_s}\right|_{\text{silicon}} + \left.\frac{dQ_s}{dV_s}\right|_{\text{states}} \tag{8}$$

Obviously, the admittance corresponding to the first term in the above expression will just be the capacitance associated with the depletion region which has been calculated previously. The admittance of the second term can be obtained by making the usual small-signal approximations for quantities varying sinusoidally in time. Then,

$$\begin{aligned} Q_{s0} &= Q_{00} + Q_{s0_1}\exp(j\omega t) & Q_{00} &\gg Q_{s0_1} \\ Q_s &= Q_0 + Q_{s_1}\exp(j\omega t) & Q_0 &\gg Q_{s_1} \\ V_s &= V_{s_0} + V_{s_1}\exp(j\omega t) & V_{s_0} &\gg V_{s_1} \\ I_s &= I_0 + I_1\exp(j\omega t) & I_0 &\gg I_1 \end{aligned} \tag{9}$$

where, for example, Q_{00} is the time invariant component of Q_{s0} and Q_{s0} is the magnitude of the time-varying component. If the quantities of (9) are substituted into (8), two independent equations result, one for the time-independent terms, and another for the time-varying terms. Considering only the latter, the second term in (8) may be written

$$I_1 = \frac{Q_{s_1} - Q_{s0_1}}{\tau} \tag{10}$$

The admittance may be found by noting that $I_1 = -dQ_s/dt$. Then,

$$I_1 = -Q_{s0_1}\left(\frac{j\omega}{1+j\omega\tau}\right) \tag{11}$$

But, for small applied signals,

$$Q_{s0_1} = \frac{dQ_s}{dV_s} V_{s_1}$$

so that the admittance due to the surface states becomes

$$I_1/V_{s_1} = j\omega\left(-\frac{dQ_s}{dV_s}\right)\frac{1}{1+j\omega\tau} \qquad (12)$$

This admittance has the same form as that of a series R–C circuit. The surface states may thus be represented by such a circuit, the capacitance of which is equal to $-dQ_s/dV_s$ and the time constant of which is τ, the time constant of the states themselves. It should be noted that since dQ_s/dV_s is a negative number, the capacitance due to the states will be positive; it is perhaps less confusing to write

$$\left|\frac{dQ_s}{dV_s}\right|$$

for the surface state capacitance.

Bearing in mind the presence of a capacitance due to the oxide which is in series with the silicon surface, the admittance per unit area of the whole M–O–S diode is

$$Y = j\omega \Big/ \left\{ \frac{W_{ox}}{\epsilon_{ox}} + \left[C_D + \left|\frac{dQ_s}{dV_s}\right| \frac{1}{1+j\omega\tau} \right]^{-1} \right\} \qquad (13)$$

The equivalent circuit corresponding to equation (13) is given in Fig. 5(a). The depletion capacitance is the only element of the equivalent circuit which varies directly with bias voltage. The resistance and capacitance due to surface states may vary indirectly with voltage to the following extent: Consider Fig. 6. If it is assumed that only donor states are present,* all the states above the Fermi level will be empty, all those below will be occupied.† Call $E_F(S)$ the potential at which the Fermi level intersects the surface of the silicon. If one calculates R_s and C_s in the equivalent circuit from a small-signal measurement, only those states located at or very near $E_F(S)$ will be affected by the applied signal and will hence contribute to the values of R_s and C_s measured. The effect of varying the bias is to shift the value of $E_F(S)$. Thus, by varying the bias, one may examine the properties of the surface states at different positions‡ in the energy gap. The resistance and capacitance due to surface states will thus be indirectly a function of bias voltage to the extent that the distribution of surface states

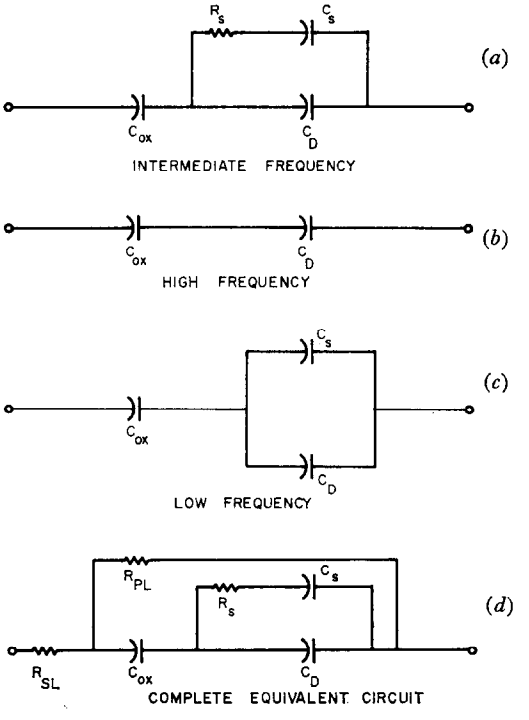

FIG. 5. a–d equivalent circuits for the M–O–S diode.

* Actually, both donor and acceptor states may be present. Both types of states will be in their most positive condition when above the Fermi level, and in their most negative condition when below. Since the effect of surface states appears as a result of the total change of charge present at the interface for a perturbation of the Fermi-level position, the effect of both donor and acceptor states will be identical. For simplicity, this report assumes only donor states are present.

† This statement is true only at absolute zero. At a non-zero temperature there will be a transition region from essentially full to essentially empty for energy levels within several kT of the Fermi level. For room temperature this represents only a small effect, and the above statement is essentially true for the purpose of this report.

‡ Position here does not mean a physical location but rather the depth in energy in the energy gap at which the states are located in a plot such as Fig. 6.

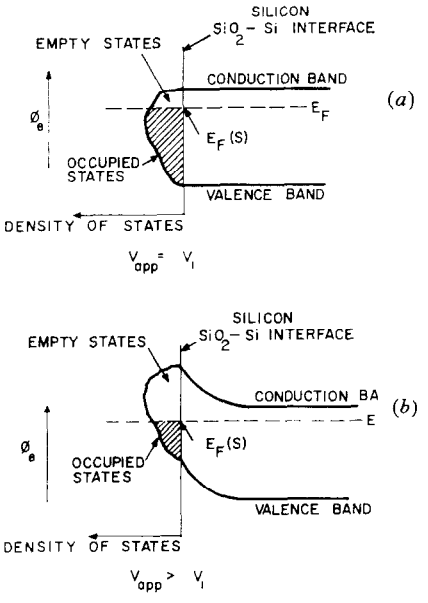

FIG. 6. Effect of bias voltage on surface-state occupancy.

and their time constants are not uniform with depth in the energy gap.

At high frequencies $\omega\tau \gg 1$, and the equivalent circuit of the device will be just the oxide and depletion capacitances in series, as in Fig. 5(b).* This will occur when the time constant is too long to permit charge to move in and out of the states in response to an applied signal. The small-signal equivalent circuit will be just that which would be observed if no states were present. At low frequencies where $\omega\tau \ll 1$, the equivalent circuit is that of Fig. 5(c). Here all the surface states affected by the applied signal are able to give up and accept charge in response to this signal. The surface-state capacitance appears directly in parallel with the depletion capacitance, resulting in higher total value of capacitance for the device than if no states were present. At intermediate frequencies where $\omega\tau \simeq 1$, some, but not all, of the surface state charge will participate in a small-signal measurement, and values of capacitance observed will be between the low and high frequency values.

In general, the capacitance of the device will decrease monotonically with increasing frequency at a given bias voltage. Physically, the increase in capacitance due to surface states arises from these states being located at the silicon surface, and thus representing a source of charge closer to the metal layer than the charge in the depletion region.

The presence of surface states will alter the shape of a capacitance–voltage curve even if the measurements are made at a frequency high enough that the states do not contribute their charge to the capacitance measurements. At a given bias, surface states lying above the Fermi level in energy will be empty. The empty states represent a density of positive charge at the silicon oxide/silicon interface, and will give rise to an additional field in the oxide. This will result in a larger voltage drop across the oxide than would be experienced if no states were present. The capacitance–voltage curve will consequently be shifted horizontally by the amount of the increased drop in the oxide. The amount of this shift will depend on the density and location of the states, and is 0·48 V/100 Å oxide thickness per 10^{12} states. For a typical oxide thickness of 200 Å and a density of states of 10^{12}, the horizontal shift will be 0·96 V. A comparison between the actual and theoretical high-frequency capacitance–voltage

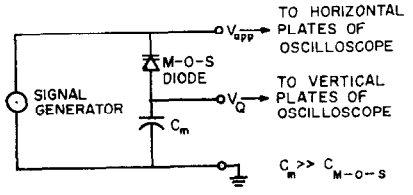

FIG. 7(a). Circuit to plot Q vs. V for M–O–S diode on oscilloscope.

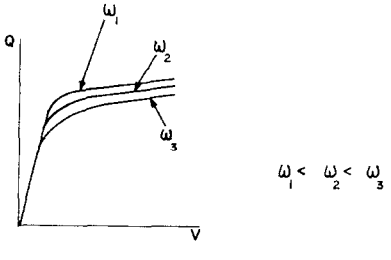

FIG. 7(b). Typical plot of Q vs. V for M–O–S diode on oscilloscope.

* A curve of capacitance vs. voltage made at such frequencies is termed a "high frequency" C–V curve.

curves may be used to determine the density and location of states.

The equivalent circuit of Fig. 5(a) corresponds to the ideal device. It does not include the effect of series losses in the bulk resistivity of the silicon, and the d.c. leakage through the oxide and depletion region. An equivalent circuit which does include these effects is given in Fig. 5(d). R_{SL} limits the quality factor Q at high frequencies, while R_{PL} limits Q at low frequencies.

C. *The induced inversion layer*

A final complication to the theory thus far presented arises from the creation of an induced inversion layer of charge in the silicon for higher voltages in the reverse direction. The electron current through any cross-section of the boundary must be zero, and electrons and holes obey the Boltzmann distribution law; thus,

$$n(x) = n_0 \exp(qV_s/kT) \tag{14}$$

and

$$p = p_0 \exp(-qV_s/kT) \tag{15}$$

Since V_s is a negative number for reverse bias, the density of holes will increase rapidly with reverse bias. For n-material with a doping density of 10^{15}, the unbiased hole density will be approximately 10^{10} less than the electron density. However, a reverse voltage drop across the silicon of 0·75 V will increase the density of holes, according to (15), to the point that it is of the same order of magnitude as that of the ionized donors in the depletion region. At this point any further positive charge needed to terminate an increased field in the device due to larger applied back voltages can be supplied by the holes, rather than the ionized impurity centers. The result is a large increase in charge for only a small increase in voltage, so that the presence of the induced inversion region will virtually eliminate any capacitance change with voltage, since the depletion region will change width only slightly with increasing voltage. Here the capacitance–voltage curve will have an almost flat region. Further increase in the reverse field results mainly in an increase in the oxide field, eventually resulting in breakdown in the oxide film.

An important ramification of the induced inversion region is that 0·75 V is close to the maximum d.c. drop attainable across the silicon. Since for doping levels of 10^{15} the Fermi level lies about 0·20 V below the bottom of the conduction band, it is very difficult to bring the Fermi level lower than about 1 eV below the conduction band with d.c. bias. Larger a.c. signals (tens of volts) can be applied to the silicon if the frequency is sufficiently high so as to prohibit minority-carrier generation. In silicon at room temperature, 1000 c/s signals are adequate.

II. MEASUREMENTS

A. *Validity of the theoretical model*

It would appear that the first step toward obtaining surface-state information by use of the M–O–S diode would be the experimental verification of the field and charge distributions for the basic device, operating in the absence of surface states. However, the existence of surface states (which are present in all real M–O–S diodes) prevents this simple comparison. As long as the magnitude of the slope of the measured high-frequency C–V curve is less than or equal to that of the theoretical curve (i.e. for the same value of capacitance on the two curves the increase in voltage of the actual curve over the theoretical curve always becomes monotonically greater with increasing bias across the device), any disagreement between the theoretical and the measured C–V curves may be explained by postulating the existence of states with the proper density and location.

Verification of the model for the M–O–S diode must come from determination of the surface-state distribution needed to bring the theoretical and actual C–V curves into agreement, and comparison of that distribution with the surface-state distribution obtained by another method of measuring the surface states.

B. *Surface-state measurement techniques*

Surface states manifest their presence in a number of ways. Thus, there are several different methods of demonstrating the presence of states and obtaining information on them. Perhaps the simplest and most direct of these is to plot a charge–voltage characteristic for the M–O–S diode, directly on an X–Y oscilloscope. This may be done with the circuit given in Fig. 7(a). A family of such characteristics, similar to that of

Fig. 7(b) will result from multiple exposures on the same negative of the scope trace for various applied frequencies. The vertical separation between traces of different frequencies is due to surface-state charge, and the difference in frequency of the various traces gives at least some idea of the time constants involved. Although this method is not suitable to making accurate quantitative measurements, it does directly indicate the charge present due to surface states, while other methods require obtaining this charge indirectly through calculation.

A second method of obtaining information is the comparison of the theoretical and actual high-frequency capacitance–voltage curves. Since the capacitance and voltage of the device can be measured quite accurately, this represents a sensitive measurement of the presence of states. As little as a few times 10^{10} states should be detectable, which is more than an order of magnitude less than the densities actually encountered. Unfortunately, this method reveals nothing about the time constants of the states; this information must be obtained by other measurements. Despite this, it is a useful tool for checking surface state distributions and total-charge measurements obtained by other methods. However, it should be noted that this method is essentially a d.c. technique, and will thus include the effect of all states, regardless of time constants. The results obtained may not be in agreement with those obtained from any technique requiring the application of an a.c. signal of non-zero frequency if some states have time constants too long to permit response to the a.c. signal. At worst, the d.c. measurement does provide an upper limit on the number of states present at any location in the energy gap.

One way in which surface states reveal their presence is by the addition of two elements in the equivalent circuit, R_s and C_s. The presence of these elements will be reflected in the two terminal impedance. Since the other elements of the equivalent circuit may be determined separately, it is possible to calculate numerical values for R_s and C_s from an impedance measurement, and from this to obtain the time constants and surface-state distribution. This represents the most effective way of obtaining the desired information.

If surface states at a given surface potential have more than one time constant, they should not be represented by a single R–C series circuit, but rather by a number of such circuits in parallel, each with an R–C product equal to a time-constant present, and a value of capacitance proportional to the number of states with that time constant. In this case, the R_s and C_s calculated from the impedance measurements do not represent any group of states with one time constant, but rather represent the average effect of all the states with their various time constants. The individual time constants may be obtained by calculating R_s and C_s at a number of frequencies, and then synthesizing the equivalent frequency-dependent impedance with parallel connected R–C circuits.

The position of the Fermi level at the silicon surface for a given total voltage drop across the device may be obtained from the theoretical capacitance–voltage curve, since, if the physical constants and dimensions are known, the ratio of the high-frequency measured capacitance to the oxide capacitance uniquely determines the potential at the silicon surface.

C. *Test-diode construction; measurements; data conversion*

Test diodes were constructed out of 1-Ω-cm n-type silicon from a grown crystal by lapping with no. 600 carborundum grit, etching in a 2:1 mixture of nitric and hydrofluoric acid, quenching the etching action in methanol, and either oxidizing directly* or cleaning the surface† and then oxidizing. Oxidation was effected in an oxygen atmosphere for 30 min at 800–850°C. After oxidation an alloyed gold contact was made to the silicon, which was then diced, and a 0·005 in. dia. aluminum dot was vacuum deposited on the oxidized surface of each chip. The chips were mounted in headers and a pressure contact made to the aluminum.

Small-signal two-terminal impedance bridge measurements were made at 1 decade intervals from 100 c/s to 1 Mc/s with additional measurements

* A surface thus produced is hereafter termed a quenched surface.

† A surface thus produced is hereafter termed a cleaned surface. The cleaning process was supplied by Texas Instruments Co., which requested that the exact procedure be kept private.

at 5 and 50 Mc/s. Some measurements were made at 100 Mc/s to insure that the 50 Mc/s measurements were truly above the effect of surface states. At each frequency, measurements were made at a dozen bias points.

From these measurements the average surface-state capacitance C_s and resistance R_s were calculated and the individual series R–C circuits synthesized. To do this, it was first necessary to determine the other elements in the equivalent circuit. R_{PL} was obtained by a simple V/I calculation of the d.c. leakage. R_{SL} was obtained as the resistance in the high-frequency equivalent circuit of the device, C_{ox} was obtained as the asymptotic value of the C–V curve, and C_D was obtained from the high-frequency C–V curve and the known value of oxide capacitance.

Finally, the static or d.c. curve of surface states was calculated from a comparison of the theoretical and actual high-frequency capacitance vs. voltage curves.

III. RESULTS

A. *Results and discussion*

Results obtained from two typical devices are presented, one having a surface prepared by the quenching method, and the other having a surface prepared by the special cleaning process. The data obtained from the small-signal measurements were converted to equivalent series resistance and capacitance, from which Figs. 8(a) and (b) were plotted. The static line in these figures represents a curve which would result if the measurements could be made at essentially d.c. It is calculated by adding the value of surface-state capacitance obtained from the static curve to the depletion capacitance and, knowing the oxide capacitance, calculating the resulting total capacitance.

Values of average C_s and R_s were then calculated and Figs. 9(a) and (b) plotted. The change in shape of these profiles from one frequency to another gives at least some indication of the time constants present. The static line indicates the profile of states obtained from the high-frequency C–V curve. Surface-state capacitance is converted to surface-state density by dividing by the electronic charge and the area of the device. $E_F(0)$, for $N_D = 5 \times 10^{15}$ is about 165 mV below the bottom of the conduction band.

Finally, the surface-state impedance as a

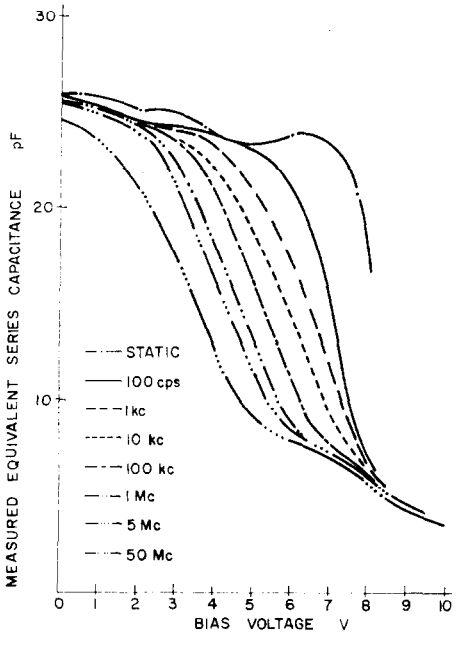

FIG. 8(a). C vs. V curves as a function of frequency for the "cleaned-surface" diode.

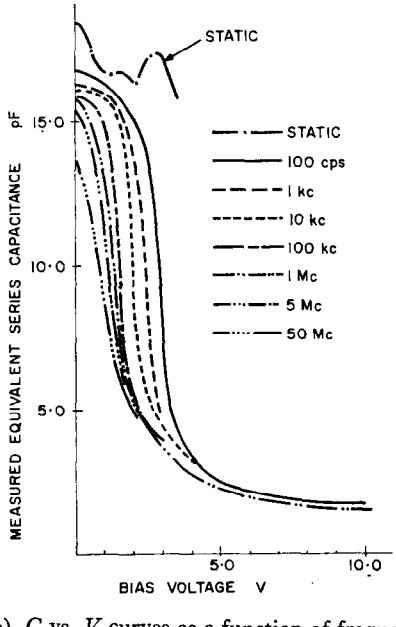

FIG. 8(b) C vs. V curves as a function of frequency for the "quenched-surface" diode.

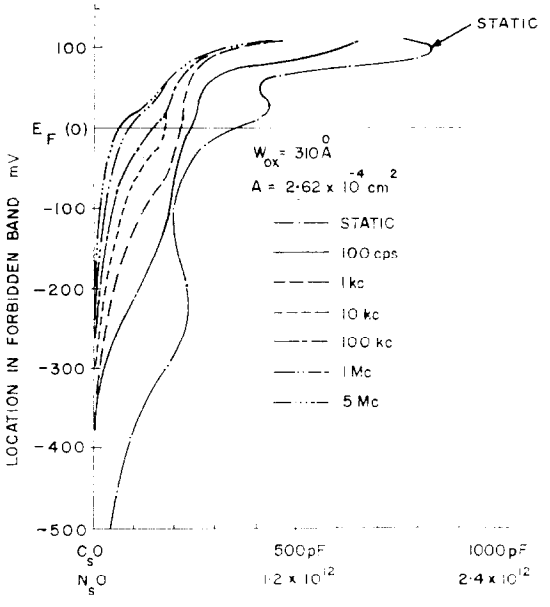

FIG. 9(a). Surface-state capacitance (C_s) and density of states (N_s) vs. location in the forbidden gap as a function of frequency, for the "cleaned" surface.

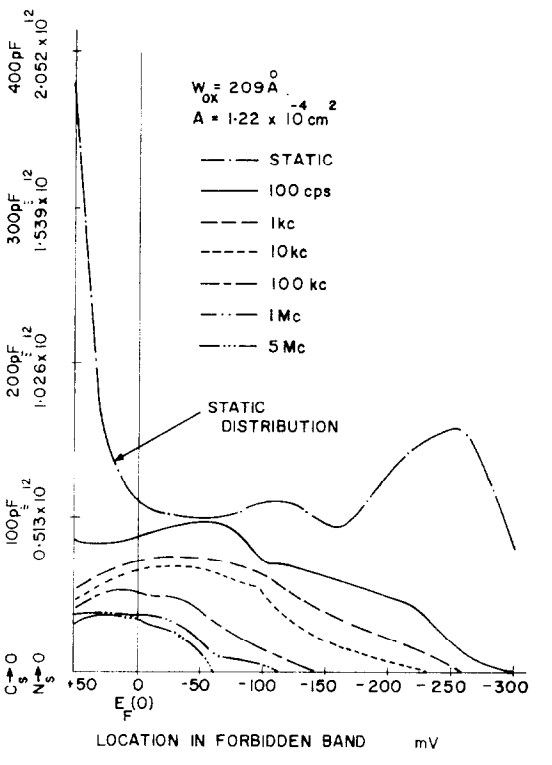

FIG. 9(b). Surface-state capacitance (C_s) and density of states (N_s) vs. location in the forbidden gap as a function of frequency for the "quenched" surface.

function of frequency was calculated, and this impedance synthesized. Figs. 10(a) and (b) are plots of the results. In these figures each point represents a calculated time constant, and the adjoining number is the value in picofarads (10^{-12} F) of the capacitance having that particular time constant. The values of capacitance given in brackets $\langle\ \rangle$ are the amounts of capacitance that must be added to the total capacitance obtained from the synthesis of the surface surface state impedance in order to obtain agreement with the static curve. They thus represent states with time constants too long to be observed at the lowest frequency of measurement.

Both surfaces have approximately the same profile of total number of states vs. depth in the energy gap. The total density of surface states is about the same for the two surfaces, being $3{\cdot}6 \times 10^{12}$ states/cm^2 for the cleaned surface, and $2{\cdot}8 \times 10^{12}$ states/cm^2 for the quenched surface. The spread of time constants is surprisingly large for both surfaces, encompassing at least six orders of magnitude. Both surfaces are similar, having a general trend towards a rapid decrease in the number of shorter time-constant states for deeper energy levels. The significant difference between the two surfaces lies in the short time-constant states located at shallow energy levels. The cleaned surface has a high density of such states, the quenched surface about an order of magnitude fewer.

It is evident that there is more than a single mechanism responsible for the states observed. One mechanism is needed to explain the shallow, long-time-constant states present on both surfaces, another to explain the large number of short time-constant shallow states on the cleaned surface, and perhaps a third for the scattering of other time constants observed.

B. *Comparison of results with those obtained by other workers in the field*

Although a fair amount of effort has been devoted to the study of germanium surface states,

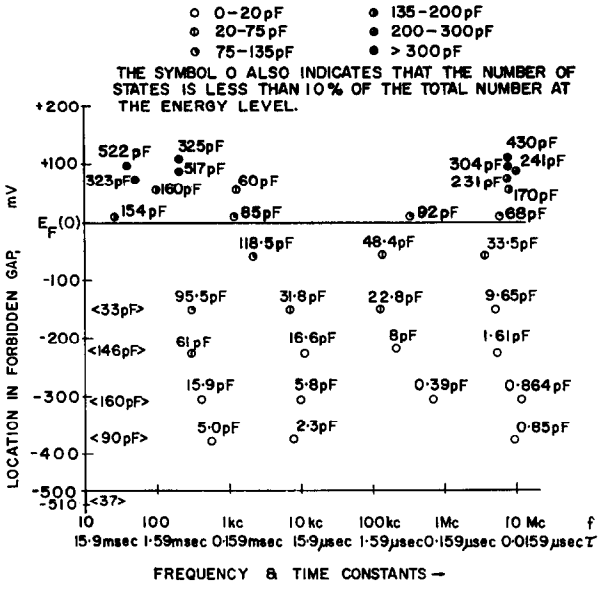

FIG. 10(a). Time constants and surface-state capacitance vs. location in the forbidden gap as calculated from synthesis of the surface-state impedance ("cleaned" surface).

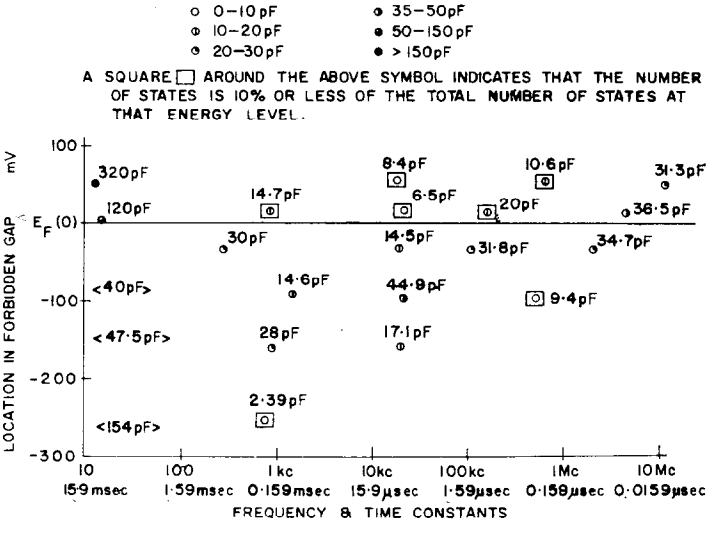

FIG. 10(b). Quenched surface.

and a number of articles presenting the results of this work have been published, the results of similar work on silicon are less numerous. Apparently, a major reason for this is the difficulty of inducing sufficient charge at the semiconductor surface in a standard field-effect measurement. Since the density of surface states is about an order of magnitude greater in silicon than in germanium, it is considerably more difficult to charge the states, and hence to measure their distribution, in silicon.

Of the references cited, none gave results in

direct contradiction with those obtained in the present investigation. Working with pulled silicon crystals, ATALLA *et al.* [4] reported that in many cases the oxidized surface was so strongly *n*-type that no conduction minimum was observed, even when a plate signal corresponding to about 2×10^{11} electronic charges/cm^2 was applied, indicating a density of donor-type states in excess of this figure. STATZ *et al.*[5] using *n*-type inversion layers on *p*-type silicon to investigate the upper half of the forbidden band, concludes that in the upper half of the forbidden gap a single group of states exists at 0·44–0·48 eV above the intrinsic Fermi level, with a density of about 10^{12} states/cm^2. An earlier paper by STATZ *et al.*[6] investigated the lower half of the forbidden band by use of *p*-type inversion layers on *n*-type silicon, and reported the existence of a similar density of states about an equal distance below the intrinsic Fermi level. Statz further concludes that there are very few states in the middle of the forbidden band. RUPPRECHT[7] and HARTEN[8] observe states located about 0·1 eV above the middle of the forbidden gap.

ATALLA's results are consistent with the total density of states observed for the surfaces measured in this report. He seems to indicate a continuous distribution of states on actual surfaces. Agreement with the work of STATZ is good to the extent that both indicate very few states in the middle of the band, and that the single group of states high in the forbidden band as reported by STATZ corresponds well, both in density and energy level, with the peak of states located above $E_F(0)$ on the surfaces investigated here. The states observed by RUPPRECHT and HARTEN agree, at least approximately, in location with the density of long time-constant states observed 250–300 mV below $E_F(0)$ here. STATZ, RUPPRECHT and HARTEN, however, report discrete distributions, while the present results indicate additional states which make up a continuous distribution. It should be noted, however, that the same sort of disagreement exists among workers in the field investigating germanium surface states, as indicated by MANY.[9]

There has been almost no previous work done on determining the time constants of the surface states, and virtually all the work which has been done has been centered about the slow states in oxidized germanium. These slow states are believed to exist on the outer surface of the oxide. ATALLA does not observe the existence of these states on oxidized silicon surfaces, and in any case the application of a metal layer to the oxide surface as in the M–O–S diode would eliminate their effect. ATALLA does observe a frequency effect in his field-effect measurements in the 25–300 c/s range when impurities are introduced during the oxidation process. MONTGOMERY and BROWN,[10,11] in high-frequency measurements on germanium, obtain results which indicate that the relaxation time of high-frequency states is in the range of 10^{-7}–10^{-8} sec.

It is, of course, impossible to make a direct comparison between the findings of ATALLA and MONTGOMERY on time constants and those of this report. However, the time constants observed here are certainly plausible on the basis of their findings.

There is a need for further work on silicon surfaces. In comparing the results obtained by various workers, however, it should be borne in mind that the state distribution and their time constants are sensitive to the surface treatment before oxidation, the oxidation process itself, and in any case are not highly repeatable at present.

APPENDIX

"Exact" Calculation of the Capacitance of the Silicon Surface

This calculation makes use of the work of GARRETT and BRATTAIN.[3] Three assumptions were made for the calculation:

(1) All the impurity atoms are ionized everywhere.
(2) The semiconductor is not degenerate.
(3) The semiconductor is at room temperature.

Then, following the development of Ref. 3, the net charge at the silicon surface will be

$$Q_{\text{net}} = \sqrt{\frac{4q\epsilon_s n_i}{\beta}}[\cosh\beta(\phi_0-\psi_s)-\cosh\beta(\phi_0-\psi_0) + \beta(\psi_s-\psi_0)\sinh(\phi_0-\psi_0)]^{1/2} \qquad (A.1)$$

where

$\beta = q/kT$
n_i = intrinsic hole concentration
ϕ_0 = Fermi level
ψ_s = electrostatic potential at the silicon surface
ψ_0 = electrostatic potential deep inside the silicon
ϵ_s = dielectric constant of silicon

The capacitance associated with the silicon surface will be given by the differentiation of equation (A.1)

with respect to ψ_s:

$$C = \sqrt{(q\epsilon_s n_i \beta)} \frac{\sinh\beta(\phi_0-\psi_0)-\sinh\beta(\phi_0-\psi_s)}{[\cosh\beta(\phi_0-\psi_s)-\cosh\beta(\phi_0-\psi_0)+\beta(\psi_s-\psi_0)\sinh\beta\ (\phi_0-\psi_0)]^{1/2}}$$

Let

$$\psi_s = \psi_0+\Delta\psi \text{ and } \lambda = \beta(\phi_0-\psi_0).$$

Then

$$\sinh\beta(\phi_0-\psi_s) = \sinh\lambda\cosh\beta\Delta\psi - \cosh\lambda\sinh\beta\Delta\psi$$

and

$$\cosh\beta(\phi_0-\psi_s) = \cosh\lambda\cosh\beta\Delta\psi - \sinh\lambda\sinh\beta\Delta\psi$$

For reasonably doped n-material, $N_D \gg n_i$, and

$$\cosh\lambda = \frac{1}{2}\left(\frac{p_0}{n_i}+\frac{n_0}{n_i}\right) \cong \frac{1}{2}\frac{n_0}{n_i}$$

$$\sinh\lambda = \frac{1}{2}\left(\frac{p_0}{n_i}-\frac{n_0}{n_i}\right) = -\frac{1}{2}\frac{n_0}{n_i}$$

so

$$\cosh\lambda = -\sinh\lambda$$

Then the capacitance of the silicon surface becomes

$$C = \sqrt{\left(\frac{q\epsilon_s n_0\beta}{2}\right)}\frac{\exp(\beta\Delta\psi)-1}{[\exp(\beta\Delta\psi)-\beta\Delta\psi-1]^{1/2}} \quad (A.3)$$

Since $C(0) = (q\epsilon_s n_0\beta)^{1/2}$ from a series expansion for the exponentials,

$$C = C(0)\frac{1}{\sqrt{2}}\frac{\exp(\beta\Delta\psi)-1}{[\exp(\beta\Delta\psi)-\beta\Delta\psi-1]^{1/2}} \quad (A.3')$$

Under the same approximations, the surface charge in the semiconductor is

$$Q_{\text{net}} = \sqrt{\left(\frac{q\epsilon_s n_0}{2\beta}\right)}(\exp(\beta\Delta\psi)-\beta\Delta\psi-1)^{1/2} \quad (A.4)$$

Acknowledgements—The author wishes to acknowledge his deep indebtedness to Dr. J. L. MOLL, who originally suggested the topic, and whose guidance and assistance was invaluable in the course of this study. Thanks are also due Dr. R. L. PRITCHARD of Texas Instruments Company, who made available the special cleaning process mentioned in Section II.

REFERENCES

1. I. TAMM, *Phys. Z. Sowjet.* **1,** 733 (1932).
2. W. SHOCKLEY, *Phys. Rev.* **56,** 317 (1939).
3. C. G. B. GARRETT and W. H. BRATTAIN, *Phys. Rev.* **99,** 376 (1955).
4. M. M. ATALLA, E. TANNENBAUM and E. J. SCHEIBNER, *Bell Syst. Tech. J.*, **38,** 749 (1959).
5. H. STATZ, G. DE MARS, L. DAVIS, JR., and A. ADAMS, JR., *Phys. Rev.* **106,** 455 (1957).
6. H. STATZ, G. DE MARS, L. DAVIS, JR. and A. ADAMS, JR., *Phys. Rev.* **101,** 1272 (1956).
7. G. RUPPRECHT, *J. Phys. Chem. Solids* **14,** 208 (1960).
8. H. U. HARTEN, *Z. Naturf.* **16a,** 459 (1961).
9. A. MANY, *J. Phys. Chem. Solids* **8,** 87 (1959).
10. H. G. MONTGOMERY and W. L. BROWN, *Phys. Rev.* **103,** 865 (1956).
11. H. G. MONTGOMERY, *Phys. Rev.* **106,** 441 (1957).