The Effects of Oxide Traps on the MOS Capacitance

F. P. Heiman, Member, IEEE, and G. Warfield

Abstract—The trapping of electrons and holes at a semiconductor surface by traps located in the oxide adjacent to the semiconductor has been considered. It is shown that the effective capture cross section of an oxide trap viewed by a carrier at the semiconductor surface is reduced by a factor which increases exponentially with the distance the trap is located from the interface. A pseudo-Fermi function in this position variable is developed which gives the probability that a trap will be filled (or emptied) in a measurement time, $T_m$. The trapping kinetics developed in the first part of the paper are applied to yield the full frequency and bias dependence of an MOS capacitor for an arbitrary spatial and energy trap distribution. Specific examples are given and the problem of voltage hysteresis is dealt with quantitatively. The conclusion is that very little information about the energy distribution and capture cross sections of the oxide traps is obtained from the analysis of MOS-capacitance curves.

I. INTRODUCTION

EVER SINCE the pioneering experiment of Shockley and Pearson [1] confirmed the existence of surface states at a semiconductor surface, much work has been done toward obtaining a greater understanding of their behavior through field-effect measurements [2], [3]-[5], surface recombination measurements [6], [7], surface photovoltage measurements [8], [9], and recently by capacitance measurements [10]-[14]. McWhorter [15] has shown that surface states are responsible for the $1/f$ noise commonly associated with semiconductor surface devices.

From this work, primarily with germanium, it became customary to divide the surface states into two groups: fast states that reside on the semiconductor surface, and slow states that reside on the oxide which normally covers the surface of a semiconductor. Since this oxide on germanium is generally between 10 and 40 Å thick, transfer of charge can take place between the slow states and the semiconductor by tunneling through the barrier or by thermal excitation over the barrier, producing long response times. The large density of slow surface states, controlled largely by the ambient conditions and surface preparation, generally determines the surface potential of the semiconductor.

The use of silicon for many commercial semiconductor devices has produced a wealth of technology concerned with the preparation of clean, passivated silicon surfaces. By properly producing a thermally grown silicon-dioxide layer over the silicon surface [16], the density of traps and recombination centers at the interface (fast states) can be greatly reduced. Since the SiO$_2$ films are generally

5) Use low voltage circuits with moderate values of resistance.
6) Use balanced circuits such as differential amplifiers in which gain is dependent upon resistor ratios.
7) Place resistors and transistors which must track with temperature at equal temperature points in the chip.
8) Connect substrate to the most negative supply in the circuit to reduce substrate capacitance.
9) Use resistors as underpasses.
10) Place all conductors over heavy oxide layers to reduce capacitance and increase voltage breakdown on conductor runs.
11) Allow adequate spacing between conductor runs.
12) Use short distance conductor runs for all high impedance connections.
13) Isolate all conductor pads with $p$-$n$ junction and place over heavy oxide layers to reduce capacitance.
14) Cross isolation regions with minimum conductor lengths.
15) Avoid saturating transistors in nongold doped circuits.
16) Avoid use of substrate as a conductor in low impedance circuits.

BIBLIOGRAPHY


The Effects of Oxide Traps on the MOS Capacitance
thicker than 200 Å, there is little charge transfer directly between the semiconductor and the “slow states.” The electronic properties of the silicon surface are now sensitive to the trap structure in the oxide and the terminology, “slow and fast states” no longer applies.

It is the purpose of this paper to analyze the effects of these traps in the oxide on the behavior of MOS (Metal-Oxide-Semiconductor) structures. In particular, we shall show that the presence of these oxide traps can lead to erroneous interpretations of MOS capacitance measurements used to find the distribution in energy of surface states. In addition, we shall discuss the effect of these traps on the frequency response of MOS capacitors and the voltage hysteresis associated with capacitance measurements.

The structure considered in this paper is a metal-silicon dioxide-silicon sandwich and all numerical calculations will assume the parameters associated with these materials. However, the equations derived will be applicable to all materials which can be prepared to yield a low density of interface states. The oxide will be represented by a perfect insulator containing a spatial distribution of traps, whose energy distribution is arbitrary. Space-charge effects, tunnel-current effects, and the effects of polarized molecules are assumed to be negligible [17]. The influence of mobile charges in the oxide is omitted.

II. DETAILS OF OXIDE TRAPPING

A clear understanding of the effects of traps and recombination centers on the photoconductive gain-bandwidth product and response time of insulators is now available [18]. These traps are located in the bulk of the photoconductor and have energy levels which place them in the forbidden gap. We extend the concepts presented by Rose [18] to traps and recombination centers located in the oxide adjacent to a semiconductor. It is clear that the electrons reach these trapping sites by quantum mechanical tunneling, but a detailed analysis of this mechanism [19], [20] would involve knowledge about the trap structure. Since this information is unavailable, a semiclassical approach will be taken in which the details of the trapping sites are included in an average capture cross section, \( \sigma \).

Consider the problem of an electron impinging upon a barrier of height \( W \) as shown in Fig. 1. The solution to Schrodinger’s equation for this simple one-dimensional configuration is [21],

\[
\varphi(x) = \begin{cases} 
A e^{ikx} + B e^{-ikx}, & x < 0 \\
C e^{-ikx}, & x > 0 
\end{cases}
\]

where

\[
k^2 = \frac{2m* \xi}{h^2} \quad \kappa^2 = 2m*(W - \xi)/h^2
\]

From the usual matching conditions at the boundary, the constants are found to be

\[
B/A = e^{i\alpha} \quad C/A = 1 + e^{i\alpha}
\]

where

\[
e^{i\alpha} = (ik + \kappa)/(ik - \kappa)
\]

An electron with energy \( \xi \) approaching the barrier from the left may be described by the wave packet

\[
\varphi_{in}(x, t) = \int_{-\infty}^{\infty} dk A(k) e^{i(kx - \omega t)}
\]

where \( A(k) \) is peaked about \( k_0 \) and \( \omega = \varepsilon = \hbar k^2/2m* \). The stationary phase argument places the peak of the incoming packet at

\[
x_{i,n} = \frac{\partial \omega}{\partial k} \frac{\kappa}{k} t = \bar{v}t,
\]

where \( \bar{v} \) is the group velocity of the packet.

The reflected packet is simply,

\[
\varphi_{ref}(x, t) = \int_{-\infty}^{\infty} dk A(k) e^{-i(kx - \omega t + \alpha)}
\]

and its peak is located at

\[
x_{ref} = -\bar{v}t + \frac{2}{\kappa_0}
\]

Equation (8) may be interpreted as follows: the packet enters the barrier and emerges after a delay of \( T \) seconds, where

\[
T = \frac{2}{\kappa_0}
\]

The penetration depth is simply \( 1/\kappa_0 \) since the packet travels twice this distance in \( T \) seconds; typically \( T \approx 10^{-14} \) s for an electron traveling at thermal velocity impinging upon a 1 eV barrier.

For the following calculation, we will assume that an electron spends \( T \) seconds in the classically forbidden region and during that time interval, the wave function will be taken as,

\[
\varphi(x) = \begin{cases} 
\sqrt{2\kappa_0} e^{-\kappa x}, & x \geq 0 \\
0, & x < 0
\end{cases}
\]

where the coefficient in (10) insures normalization; i.e., the entire electron is localized in the oxide during this interval. Actually, the amplitude of the wave function gradually increases and then decreases as the packet is
reflected from the barrier, but the step approximation assumed here should only alter the numerical coefficient of the subsequent results. Also, the use of one decay constant \( \kappa \) for all components of the wave packet is not restrictive, since the packet is assumed to be peaked about \( k = k_0 \) for the stationary phase argument to be valid.

With this assumed model, the probability that an electron impinging upon the oxide will get caught in a trap located in the region \( \Delta x \) about \( x \) is,

\[
\Delta P = \left( \frac{T}{\tau_n} \right)^n \phi(x)^n \Delta x
\]

where \( \tau_n \) is the lifetime of an electron in the oxide and \( T/\tau_n \) is the probability that an electron will be trapped in time \( T \) (for \( \tau_n \gg T \)). Actually, the wave function chosen does not represent a free electron wave packet and the mean lifetime of an electron in a classically forbidden region is not the same as that of a free electron. However, this can be incorporated into the numerical value of the capture cross section which will be assumed to be energy-independent in this first-order trapping theory.

The total number of electrons trapped per square centimeter per second (by traps located at \( x \)) is simply,

\[
\Delta S_x = \frac{1}{4} n_r \left( \frac{T}{\tau_n} \right)^n \phi(x)^n \Delta x
\]

where \( n_r \) is the concentration at the surface and \( \bar{v} \) is the average thermal speed of the nondegenerate free electron gas in the semiconductor.

There are several assumptions implicit in (12). By using the random current density for electrons \((n_r \bar{v} / 4)\) we are assuming that every electron approaching the interface, irrespective of its direction of approach or its total energy, may be described by the same "penetration function" given in (10).

The justification for this assumption is that the decay constant \( \kappa \) in (2) is insensitive to the energy associated with motion in the \( x \) direction. Since the barrier potential is a function of \( x \) only,

\[
\kappa^2 = 2m^* (W - \varepsilon_x)/\hbar^2
\]

where \( \varepsilon_x \) is the "\( x \) directed energy" for an electron and \( \varepsilon_x \leq \varepsilon \), the total energy of an electron. Practically all the electrons are near the bottom of the conduction band in the semiconductor and if the barrier is sufficiently high so that \( \varepsilon \ll W - \varepsilon_x \), to a good approximation, we can neglect \( \varepsilon \) in comparison with \( W \) and to first order the decay constant is the same for all electrons.

In this paper we shall assume the barrier is sufficiently high to satisfy this condition. For a barrier of approximately one eV, the decay constant is such that

\[
1/(2\kappa_0) \approx 1 \text{Å}
\]

A second assumption we shall make is that the mean thermal speed of the free electron gas is equal to the group velocity of the wave packet as defined implicitly in (6). Since \( \Delta S_x \) is the number of electrons captured per square centimeter per second in the region \( \Delta x \) about \( x \), \( \Delta S_x/\Delta x \) is the rate at which electrons are captured into traps per unit volume at \( x \). Thus,

\[
\frac{dn_r}{dt} = - \frac{dp_r}{dt} = \Delta S_x
\]

where \( n_r \) and \( p_r \) are the concentrations of filled and empty traps, respectively.

Substitution for \( T \) and \( \phi(x) \) and the use of the relation

\[
\tau_n = (\bar{v} \sigma_x)^{-1}
\]

where \( \sigma_x \) is the capture cross section of a trap for an electron in the oxide, yields

\[
\frac{dp_r}{dt} = -n_r p_r \bar{v} \sigma_x e^{-2\kappa x} = -n_r p_r \sigma_x
\]

From (17) it is apparent that the time constant of the traps \( \tau_r \) is

\[
\tau_r = (n_r \bar{v} \sigma_x)^{-1}
\]

and, most importantly, the capture cross section of a trap located a distance \( x \) from the interface into the oxide is reduced by the factor \( e^{-2\kappa x} \) when viewed by an electron at the semiconductor surface.

\[
\sigma_x = \sigma e^{-2\kappa x}
\]

It is important to distinguish between the electron life time and the trap time constant as used in this paper. The life time of an electron \( \tau_n \) is the mean time before an electron is captured. It depends upon the number of available (empty) traps \( p_r \) as indicated in (16). The time constant of a trap \( \tau_r \) is the mean time before the trap captures an electron. It depends on the concentration of electrons available for trapping as indicated in (18), and is independent of the number of traps.

It will be shown subsequently that \( \tau_r \) (not \( \tau_n \)) is the meaningful quantity to use when discussing the frequency response of the traps; hence, the time constant of a trap is not an inherent property of the trap but depends on the surface density of free carriers.

III. SMALL SIGNAL TRAP RESPONSE

The relationships derived previously will be extended to both hole and electron trapping, and (19) will be used to describe the effective capture cross section. The method used below is similar to McWhorter's analysis for \( 1/f \) noise [15].

As shown by Sah, et al., [22], the rate of change of the electron density in traps having capture cross sections \( \sigma_x \) and \( \sigma_p \) (for electrons and holes, respectively), and an energy level \( \varepsilon_x \) is (see Fig. 2)

\[
\frac{dn_r}{dt} = \left( \frac{n_r}{\tau_n} - \frac{p_r}{\tau_p} \right) - \left( \frac{p_r}{\tau_p} - \frac{n_r}{\tau_n} \right)
\]
where

\[ \tau_e = (\sigma_e \Xi_p)^{-1} \quad \tau_h = (\sigma_h \Xi_n)^{-1} \]  

(21)

and \( N_e \) and \( N_h \) are the effective densities of states in the conduction and valence bands of the semiconductor, respectively.

The first two terms on the right-hand side of (20) represent electron capture from and electron emission to the conduction band of the semiconductor, respectively; the last two terms describe hole capture from and hole emission to the valence band, respectively.

Throughout the rest of this paper, we will assume \( \sigma_{ne} = \sigma_{nh} = \sigma_n \) and invoke thermal equilibrium statistics in the semiconductor\(^1\) to give

\[ p_n n_e = n_e^2, \quad \frac{\delta p_n}{p_n} + \frac{\delta n_e}{n_e} = 0 \]  

(25)

From (20),

\[ \frac{\delta n_e}{\delta t} = \sigma_n [p_n - n_n - p_n + p_p] \]  

(26)

If we consider small deviations from an average value i.e., \( n_e = n_{e0} + \delta n_e \), etc., and replace \( d/dt \) by \( j_\omega \), (26) becomes

\[ j_\omega \delta n_e = \sigma_n \left[ - (n_e + n_n + p_n + p_p) \delta n_e \right] \]  

(27)

Thus,

\[ \delta n_e = \frac{1}{2 \omega} \ln \left[ \frac{1}{1 + e^{-\omega \tau_e}} \right] \]  

(28)

Since no dc current flows, thermal equilibrium statistics is valid provided that the frequency of the ac signal is low enough so that excess charge distributions can follow.

Hence, the characteristic small signal trap time constant \( \tau_e \) depends on both the hole and electron densities at the surface, as well as the cross section and energy level of the trap. The temperature dependence of the time constant is implicit in (29).

IV. SPATIAL DISTRIBUTION OF OCCUPIED OXIDE TRAPS

Consider (26) under the conditions: \( p_i = N_i, n_i = 0 \); i.e., all traps are empty at \( t = 0 \). The initial rate at which traps are filled is

\[ \frac{\delta n_i}{\delta t} \bigg|_{t=0} = \sigma_n N_i (n_e + p_n) = \frac{N_i}{\tau_e} \]  

(30)

which represents electron capture (the \( n_e \) term) and hole emission (the \( p_n \) term). As the traps become filled, the rate of filling will decrease. To a first approximation we assume an exponential rise in the trap population with the time constant \( \tau_e \) as defined in (30).

\[ n_i(x, \ell) = n_i(x, \infty) [1 - e^{-\ell/\tau_e}] \]  

(31)

That is, we assume that eventually the traps will reach their thermal equilibrium occupancy as described by the Fermi function.

We ask now for the occupancy of the traps at energy \( E_i \) after a time \( T_\ell \), which is defined to be the measurement time in an experiment. Thus at \( T_\ell \),

\[ n_i(x, E_i) = N_i \left[ \frac{1}{1 + e^{(E_i - \phi) / kT}} \right] \]  

(32)

From the definitions of \( \tau \) given in (30) and \( \sigma_e = \sigma_{ne} \) given in (19),

\[ T_\ell / \tau = T_e e^{-2kT \sigma_e (n_e + p_n)} \]  

(34)

It is convenient to define a distance \( x_\ell \) associated with \( T_\ell \) as

\[ x_\ell = \frac{1}{2\sigma_e} \ln \left[ T_e \sigma_e (n_e + p_n) \right] \]  

(35)

leading to

\[ n_i(x, E_i) = N_i \left[ \frac{1}{1 + e^{(E_i - \phi) / kT}} \right] \left[ 1 - \exp \left( - e^{2kT \sigma_e (n_e + p_n)} \right) \right] \]  

(36)

In general, then, we may write

\[ n_i(x, E_i) = N_i f(\phi_i) g(x, E_i) \]  

(37)

where \( f(\phi_i) \) is the Fermi function and

\[ g(x, E_i) = 1 - \exp \left[ - e^{2kT \sigma_e (n_e + p_n)} \right] \]  

(38)
The function \( g(x, \varepsilon_i) \) is plotted in Fig. 3 for \( x_m = 15 \AA \). The transition from \( g \approx 1 \) to \( g \approx 0 \) occurs at \( x = x_m \) over a width of \( \Delta x \approx 5(1/2K_0) \). As we saw in (14) this width is of the order of 5 Å for a barrier approximately 1 eV high.

We may interpret the function \( g(x, \varepsilon_i) \) as a pseudo-Fermi function in the variable \( x \), or as the probability that a trap of energy \( \varepsilon_i \) at \( x \) will have its thermal equilibrium occupancy. Because of the relatively sharp functional dependence of \( g \) on \( x \), we can see that essentially those traps at \( x > x_m \) will not be occupied in the measurement time \( T_m \), while those at \( x < x_m \) will have their equilibrium occupancy. Note that \( x_m \) depends upon the electron concentration at the surface and hence on the temperature.

For a general distribution of oxide traps \( K_i(x, \varepsilon_i) \) in units of \( \text{cm}^{-2} \text{eV}^{-1} \), the total number of filled traps \( N_{ss} \) per unit surface area is

\[
N_{ss} = \frac{1}{w_o} \int_{0}^{x_m} \int_{\varepsilon_o}^{\varepsilon_i} K_i(x, \varepsilon_i)f(\varepsilon_i)g(x, \varepsilon_i) \, d\varepsilon_i \, dx
\]  

where \( w_o \) is the oxide thickness. The integral will be a function of the position of the Fermi level \( \varepsilon_F \).

It is interesting to note that both the small signal trap time constant \( \tau \), and the depth \( x_m \), to which the traps can be filled in the given measurement time depend upon the temperature through \( n_s \), \( p_s \), \( n_i \), and \( p_i \), even though a tunneling process has been invoked to account for the capture process.

Just as there is a maximum depth \( x_m \) to which traps can be filled during a measurement time \( T_m \) after the position of the Fermi level is raised by the application of a voltage, there is a maximum depth \( x_{\text{tr}} \), to which they can be emptied in the same measurement time after the Fermi level is lowered. By using an argument for hole capture analogous to that for electron capture, we find

\[
x_{\text{tr}} = \frac{1}{2K_0} \ln \left( \frac{T_m \sigma_i(p_s + n_i)}{\sigma_e} \right)
\]

Essentially all traps above the Fermi level will be emptied to a depth \( x_{\text{tr}} \) in the time \( T_m \).

The fact that \( x_s < x_m \) under certain bias conditions leads to hysteresis effects which will be considered in the examples of surface state calculations in Section VII.

V. FIELD DEPENDENCE OF EFFECTIVE SURFACE STATE DENSITY

Since a trapping site resides in the oxide, its energy level (with respect to the bottom of the semiconductor conduction band) will be a function of the electric field in the oxide as well as of the distance \( x \). This is indicated in Fig. 4(a), where a monoenergetic trap level is assumed to exist in the oxide when no field is present. A consequence of this effect is that the single trap level shown in Fig. 4(a) gives rise to a uniform density of surface states (as seen from the semiconductor) which depends on the electric field \( E_o \) in the oxide. Explicitly, the density of surface states \( \mathcal{N}_{ss} \) is

\[
\mathcal{N}_{ss} = \frac{N_i \Delta y}{e E_o \Delta y} = \frac{N_i}{e E_o}
\]

\( N_i \) is the volume density of traps, and \( \mathcal{N}_{ss} \), is the effective surface state density in units of \( \text{cm}^{-2} \text{eV}^{-1} \). Numerically, a trap density of \( 10^{18} \text{cm}^{-2} \) under an applied electric field of \( 10^6 \text{V/cm} \) gives rise to an effective density of \( 10^{12} \text{cm}^{-2} \text{eV}^{-1} \), which is a typical surface state density.

The energy range over which this density exists may be estimated as follows. Since the traps can be filled to a depth \( x_m \) which is typically about 20 Å, then,

\[
\Delta \varepsilon = e E_o x_m = e \times 10^6 \times 2 \times 10^{-7} = 0.2 \text{eV}
\]

which is about one-fifth of the forbidden gap in silicon. Experimentally, a trap level of this kind will give rise to an effective surface state density as shown in Fig. 4(b).

The fact that the trapped charge resides in the oxide,
and not at the interface, will have a negligibly small effect on the voltage drop across the oxide. To see this, note that the particular solution to Poisson’s equation for a uniform density of bound charge $eN$ can be written as

$$\psi = \frac{eN}{2e_{ox}} x^2$$

(43)
or

$$\frac{e_{ox}}{kT} \psi = \frac{e^2N}{2e_{ox}kT} x^2 = \left(\frac{x}{\lambda_{ox}}\right)^2$$

(44)

for a judicious choice of origin. Here $\psi$ is the electrical potential, $\varepsilon_{ox}$ is the permittivity of the oxide, and $\lambda_{ox}$ is a Debye length based on the charge density $eN$. For $N = 10^{16}$ cm$^{-2}$, $\lambda_{ox} = 10 \text{ Å}$. Thus if these traps are filled to a depth of 20 Å into the oxide, $e+/kT \sim 4$.

In practice this is a negligibly small correction. Thus for computing the total voltage drop across the oxide, all the charge in surface states will be assumed to exist at the silicon-silicon dioxide interface.

VI. CAPACITANCE OF AN MOS STRUCTURE

Now that the details of carrier trapping have been investigated, the full frequency and voltage dependence of the capacitance of an MOS structure will be determined. The capacitance and potential profiles associated with a trap-free semiconductor surface have been studied by Garrett and Brattain [23], and others [24]-[26], but the fundamental equations will be briefly rederived here to establish the notation. We will assume that the zero-bias thermal equilibrium situation leaves the semiconductor bands flat (no net charge in surface states) and that the contact potential between the metal and semiconductor is zero. Inclusion of a finite surface state charge and a non-zero contact potential merely offsets the total voltage but adds no physical significance. The bulk of the p-type silicon will be maintained at ground potential; its surface potential will be $\psi_s$ and the potential of the metal electrode will be $V$.

Before proceeding with the details of the calculations, consider the following expression for the capacitance. By Gauss’s law,

$$\varepsilon_{ox} E_{ox} = \varepsilon_{ox} \frac{V - \psi_s}{w_{ox}} = C_{ox} (V - \psi_s) = -Q_{tot}$$

(45)

where $C_{ox}$ is the oxide capacitance per unit area and $Q_{tot}$ is the total charge per unit surface area in the semiconductor surface and the oxide traps. Since

$$C = -\frac{dQ_{tot}}{dV} = C_{ox} \left(1 - \frac{d\psi_s}{dV}\right)$$

(46)

one sees that the capacitance is completely determined by the ability of the surface potential $\psi_s$ to follow the applied voltage. If a heavy surface state density or space charge accumulation causes $\psi_s$ to change little when $V$ is changed, then $d\psi_s/dV \approx 0$ and the oxide capacitance is measured. If a large depletion region is present and a change in $\psi_s$ results in a small change in the total surface charge in the semiconductor, then $d\psi_s \approx dV$, $d\psi_s/dV \lesssim 1$, and the capacitance measured is very small.

If $Q_{tot}$ in (45) is decomposed into

$$Q_{tot} = Q_{ss} + Q_{ox}$$

(47)

where $Q_{ss}$ and $Q_{ox}$ are the charges per unit area residing in surface states and in the semiconductor, respectively, from (45) we can write

$$V = \psi_s - \frac{Q_{ss}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} = \psi_s + V_{ss} + V_{ox}$$

(48)

where $V_{ss}$ and $V_{ox}$, defined by this equation, are the voltage equivalents of $Q_{ss}$ and $Q_{ox}$, respectively.

As we shall see shortly, $V_{ss}$ (or $Q_{ss}$) is completely determined by the surface potential $\psi_s$. $V_{ox}$ is a function of both $\psi_s$ and $V$ because of the field dependence discussed in Section V. Therefore, a determination of these functional forms permits a solution of the form $V = V(\psi_s)$, and the capacitance may be written as

$$C = C_{ox} \left[1 - \frac{1}{\left(\frac{dV}{d\psi_s}\right)}\right]$$

(49)

where

$$C_{ox} = \left[\frac{dV_{ss}}{d\psi_s} + \frac{dV_{ox}}{d\psi_s}\right]$$

(50)

and

$$C_{ox} = \left[\frac{dV_{ss}}{d\psi_s} + \frac{dV_{ox}}{d\psi_s}\right]$$

(51)

First we turn to the calculation of $V_s(\psi_s)$. Poisson’s equation for a p-type semiconductor having an acceptor concentration $N_A$ under the assumption of completely ionized acceptors is,

$$\frac{d^2\psi}{dx^2} = -\frac{e}{\varepsilon_s} \left[N_A(e^{-\frac{\psi}{kT}} - 1) - n_0(e^{\frac{\psi}{kT}} - 1)\right]$$

(52)

where $n_0$ is the thermal equilibrium electron concentration in the bulk and $\varepsilon_s$ is the permittivity of silicon. One integration yields,

$$\left(\frac{d\psi}{dx}\right)^2 = \frac{2e}{\varepsilon_s} \left[\psi(N_A - n_0) + N_A \left(\frac{kT}{e}\right)(e^{-\frac{\psi}{kT}} - 1) + n_0 \left(\frac{kT}{e}\right)(e^{\frac{\psi}{kT}} - 1)\right]$$

(53)

Relating

$$E_s = -\frac{d\psi}{dx}_{\psi=\psi_s}$$

to $V_{ss}$ by Gauss’s law yields,

$$V_{ss} = \pm 2 \left(\frac{kT}{e}\right) \left(\frac{\psi_s}{\varepsilon_{ox}}\right) \left(\frac{w_{ox}}{\lambda_{ox}}\right) G(\psi_s)$$

(54)

The terminology “surface states” will be synonymous with oxide traps in this paper.
\[ G(\psi_0) = \left[ e^\beta \psi_0 - 1 + e^{-\alpha \psi_0} \right. + \left. \frac{n_0}{N_A} (e^{\beta \psi_0} - 1 - e^\beta \psi_0) \right]^{1/2} \]  

(55)

where

\[ \lambda_s = \left[ \frac{2kT_e}{eN_A} \right]^{1/2} \]  

(56)

is the Debye length for the bulk semiconductor. The plus sign is for \( \psi > 0 \); the minus sign is for \( \psi < 0 \). Fig. 5 is a plot of \( V_s(\psi) \) for \( w_{ox} = 200 \, \text{Å} \) and \( N_A = 10^{14} \, \text{cm}^{-2} \) for a silicon silicon-dioxide structure.

Differentiation of (54) gives,

\[ \frac{dV_s}{d\psi} = \pm \left( \frac{\varepsilon_r}{\varepsilon_{ox}} \right) \left[ \frac{1 - e^{-\alpha \psi} + \frac{n_0}{N_A} (e^{\beta \psi} - 1)}{G(\psi)} \right] \]  

(57)

which is shown in Fig. 6.

The specific functional forms of \( V_s \) and \( dV_s/d\psi \) will depend on the energy distribution assumed for the oxide traps. These equations will be worked out in detail for the examples chosen in Section VII. However, the frequency response of both the semiconductor capacitance and the surface state capacitance can be discussed without reference to a specific distribution of surface states.

When the surface potential \( \psi_s \) is negative, an accumulation of holes exists at the surface. Since an ohmic contact is presumed to exist at the back of the semiconductor wafer, any change in the hole concentration called for by a change in applied voltage can be supplied in a time of the order of the dielectric relaxation time. For p-type silicon \( (N_A = 10^{14} \, \text{cm}^{-2}) \), this time is about \( 10^{-10} \) seconds. Hence, the accumulation layer can easily follow the applied signal for commonly used test frequencies, up to \( 10^6 \) radians per second.

When \( \psi_s \) is made positive, an accumulation of electrons (minority carriers) exists and for sufficient bias, most of the semiconductor charge is carried by electrons. The frequency response of this inversion layer has been discussed by Pfann and Garrett [26], Lehovec et al. [10], [27], and most recently by Hofstein [28]. Hofstein has shown that this response is typically as low as 1 to 100 c/s for 10 ohm-cm silicon, but under certain conditions may be much higher. For example, in p-type silicon thermally oxidized in wet oxygen to produce an n-type inversion layer the response may be as high as 100 kc/s.

Since the purpose of this paper is to investigate the effects of oxide traps on the capacitance of the MOS structure, we will assume that the space charge in the semiconductor, whether in an accumulation or inversion layer, can follow the test signal and stay in thermal equilibrium with the bulk of the semiconductor. Thus we shall assume that \( (dV_s/d\psi_0) \) given in (57) is independent of frequency.

3 The notation 2.00 \(-1\) is synonymous with 2.00 \times 10^{-1}, etc.
However, the variation in occupancy $\delta n_t$ of that group of states $K_t(x, \varepsilon_t)\Delta x\Delta E$, will vary with frequency as we have already discussed in (28), which can be rewritten as

$$\delta n_t(x, \varepsilon_t, \omega) = \delta n_t(x, \varepsilon_t, 0)\hat{h}(x, \varepsilon_t, \omega) \quad (60)$$

where

$$\hat{h}(x, \varepsilon_t, \omega) = [1 + j\omega\tau(x, \varepsilon_t)]^{-1} \quad (61)$$

and $\tau$, is given by (29). Hence,

$$\frac{dV_{ss}}{d\psi_s} = \frac{e}{C_0} \int_{\varepsilon_s}^{\varepsilon_f} \int_{\psi_s}^{\psi_f} K_t(x, \varepsilon_t)g(x, \varepsilon_t)\left[1 + j\omega\tau(x, \varepsilon_t)\right]^{-1} d\varepsilon_t dx \quad (62)$$

where $\varepsilon_f$ is the position of the Fermi level for the given applied bias voltage. Notice that only the quantity in the first bracket is differentiated.

Of the three factors in this bracket, $K_t$ is independent of $\psi$, but both $f$ and $g$ are dependent on $\psi$, through their dependence on the instantaneous Fermi level $\varepsilon_{Ft}$. Thus we can make the transformation

$$\frac{d}{d\psi_s} = \frac{d}{d\varepsilon_{Ft}} \cdot \frac{d\varepsilon_{Ft}}{d\psi_s} = \frac{e}{C_0} \frac{d}{d\varepsilon_{Ft}} \quad (63)$$

From (35) and (38) we notice that $g$ is a function of the Fermi energy only through $\varepsilon_{Ft}$, which depends on the surface concentration of electrons $n_s$. Because $\varepsilon_{Ft}$ is essentially the maximum depth to which traps can be filled in the measurement time $T_m$ (typically 120 seconds), $\varepsilon_{Ft}$ can not follow the applied signal, which may have a typical period of less than $10^{-1}$ second. Therefore,

$$\frac{dV_{ss}}{d\psi_s} = \frac{e}{C_0} \int_{\varepsilon_s}^{\varepsilon_f} \int_{\psi_s}^{\psi_f} K_t(x, \varepsilon_t)g(x, \varepsilon_t)\left[1 + j\omega\tau(x, \varepsilon_t)\right]^{-1} \frac{d\varepsilon_t}{d\varepsilon_{Ft}} dx \quad (64)$$

The factor $\partial f / \partial \varepsilon_{Ft}$ is peaked around $\varepsilon_{Ft} = \varepsilon_{F0}$ with a spread of several $kT$.

Physically, this result says that only those trap states which are within a few $kT$ of the Fermi position of the energy can contribute to the surface state capacitance. Those states several $kT$ above $\varepsilon_{Ft}$ will never be occupied by electrons, while those several $kT$ below will always be fully occupied during the period of the test signal provided they are within a distance $x_m$ of the interface.

Because $dV_{ss}/d\psi_s$ is frequency dependent, the MOS capacitance will be frequency dependent. Moreover, since the small signal time constant $\tau$, depends upon the applied bias through $\psi_s$, the frequency dependence will itself be a function of the bias. Therefore, no real information about the traps can be obtained by analyzing the experimental capacitance versus applied voltage curves measured at an arbitrary frequency.

As pointed out by Lehovec et al. [10], one should measure the $C$ vs. $V$ curve at a sufficiently high frequency that none of the trap states can follow the applied signal. (However, this signal frequency must be sufficiently low so that the charge in the space charge layer can follow the test signal.) If these conditions can be met, the MOS capacitance will be

$$C(\psi, \omega) = C_0 \left[1 + \frac{dV_{ss}}{d\psi_s} \right] \quad (65)$$

where $C(\psi, \omega)$ represents the capacitance measured at this appropriate frequency.

This is the same capacitance that would be measured at a lower frequency in an identical structure with no surface states.

A plot of this capacitance, normalized to $C_0$, as a function of the applied voltage $V$ for a $p$-type silicon silicon-dioxide structure with no surface states and with $w_{ox} = 200 \, \AA$, $N_s = 10^{14} \, \text{cm}^{-3}$ is shown in Fig. 7.

![Fig. 7. Normalized capacitance of a Si-SiO$_2$ structure with no surface states as a function of applied voltage.](image-url)
actual voltage with the theoretical voltage for each value of capacitance (i.e., for each $\psi_e$), one can determine from (48) the total number of electrons per unit area $N_{\text{ox}}$ that can be trapped in a time $T_m$ as a function of surface potential. The derivative, $dN_{\text{ox}}/d(\psi_e)$, is the effective surface state density $\sigma_{\text{ox}}$. Generally, $T_m$ is several minutes and the high frequency measurements yield the density of surface states that can respond within several minutes or less. This procedure for determining the effective surface state density is employed in one of the examples in Section VII.

The slight asymmetry in the theoretical $C$ vs. $V$ curve shown in Fig. 7 points out an interesting practical consideration. When measuring the $C$ vs. $V$ curve of a high resistivity sample, it is difficult to differentiate between an $n$-type sample, and a $p$-type sample containing an $n$-type inversion layer at zero bias due to the presence of positively charged surface states. However, it is shown by Heiman et al. [29] that the slope $dC/dV$ of the high frequency $C$ vs. $V$ curve evaluated in the valley, is negative for $p$-type material and positive for $n$-type material.

VII. EXAMPLES OF SURFACE STATES CALCULATIONS

In this section several specific forms for the energy distribution of traps with a uniform spatial distribution through the oxide will be assumed. For each distribution we shall calculate the MOS capacitance as a function of the applied potential.

Example 1 — Monoenergetic Trap Level

We assume that in the absence of an electric field in the oxide there is a monoenergetic trap level at an energy $E_{\text{ox}}$ with a uniform spatial concentration $N_t$. When there is a field $E_{\text{ox}}$ in the oxide, the energy $E_t$ of a trap depends upon its distance from the interface. From Fig. 4(a), it is apparent that

$$E_t(x) = E_{\text{ox}} - eE_{\text{ox}}x$$  \hspace{1cm} (66)

provided the electric field is uniform within the oxide. At a given $x$ there are traps of only one energy. Conversely, traps of a given energy will be found only at one value of $x$.

The trap distribution function for this case may be expressed most conveniently in terms of a delta function in terms of either $E_t$ or $x$. Because of the order of integration used to evaluate the integrals, it is more convenient to use $E_t$ as the variable. Thus,

$$K_1(E_t, x) = N_t \delta(E_t - (E_{\text{ox}} - eE_{\text{ox}}x)) = N_t \delta(u - u_0)$$  \hspace{1cm} (67)

with the usual conditions

$$\int_{-\infty}^{\infty} \delta(u - u_0) \, du = 1$$  \hspace{1cm} (68)

$$\int_{-\infty}^{\infty} f(u) \delta(u - u_0) \, du = f(u_0)$$  \hspace{1cm} (69)

Hysteresis effects in the capacitance versus voltage curves, which are considered in Example 2, will be avoided by performing the calculations only for increasing values of the surface potential.

If we use (67) in the integral of (39), the integral over $\mathcal{E}_t$ may be evaluated through the use of (69) to yield

$$N_{\text{ox}} = N_t \int_{-\infty}^{\infty} f(E_t - eE_{\text{ox}}x) g(x, E_t - eE_{\text{ox}}x) \, dx$$  \hspace{1cm} (70)

This integral was evaluated on a digital computer\(^5\) and the result is shown in Fig. 8 for the specific parameters $N_t = 10^{18} \text{ cm}^{-2}$, $T_m = 120 \text{ s}$, $\sigma = 3 \times 10^{-15} \text{ cm}^2$, $N_A = 10^{14} \text{ cm}^{-3}$, $w_{\text{ox}} = 200 \text{ Å}$ and $E_{\text{ox}} - E_{\text{ox}} = 0.327 \text{ eV}$.

It is significant that the total number of trapped electrons continues to increase even when the Fermi energy moves well above $E_{\text{ox}}$. This happens because $x_m$ increases when the electron concentration at the surface increases as the Fermi level continues to rise.

In performing the integration over $x$, the dependence of the electric field in the oxide on the total number of electrons trapped in the oxide was taken into account by an iterative procedure in the computer program. That is, the initial calculation of $N_{\text{ox}}$ was carried out using a field based on no trapped charge. From this initial value of $N_{\text{ox}}$, a new magnitude of field was calculated and this new field was then used to calculate a second value for $N_{\text{ox}}$, etc. This iteration was terminated when self-consistent values of $N_{\text{ox}}$ and $E_{\text{ox}}$ were obtained.

To compute the MOS capacitance we evaluate (64) using (67) for $K_1$. After integrating over $\mathcal{E}_t$ using (69), we used a digital computer to evaluate the remaining integral over $x$, with the surface potential as a parameter. This result was then used in (50) to compute the normalized MOS capacitance $C/C_{\text{ox}}$. The result, for a signal frequency $\omega = 10^4 \text{ r/s}$ is shown in Fig. 9.

Example 2 — Hysteresis Effects

Consider an MOS capacitor fabricated on $p$-type silicon. We start the calculation of the $C$ vs. $V$ curve with a large negative voltage on the metal electrode so that a heavy accumulation layer of holes exists at the silicon surface. If this voltage is applied for a sufficiently long time, all the oxide traps will be empty because of hole capture from this accumulation layer. If we assume the oxide traps are of the acceptor type, so that each center is neutral when empty, there will be no net charge in the oxide at the start of the "measurements."

As the electrode voltage is made more positive, the Fermi level will move toward the conduction band and the traps will fill. For any specific value of $\mathcal{E}_t$, the total number of electrons in traps, designated $N_{\text{ox}}(\mathcal{E}_t)$ to indicate the number for increasing bias, will be given by (39). With the maximum applied positive bias, the traps below the maximum Fermi level $\mathcal{E}_{Fm}$ are filled to a maximum

\^1This work made use of computer facilities supported in part by National Science Foundation Grant NSF-GP579.
Fig. 8. Total number of electrons trapped in the oxide vs. surface potential for a monoenergetic trap distribution at $8 = \varepsilon_t$.

Fig. 9. Normalized capacitance of a Si-SiO$_2$ structure with a monoenergetic oxide trap distribution. Parameters are given in Fig. 8.

depth $x_{mn}$ determined by the electron concentration at the surface and

$$N_{ss} = \int_0^{x_{mn}} \int_{\varepsilon_{so}}^{\varepsilon_{to}} K_i(x, \varepsilon_s) f''(\varepsilon_s) g''(x, \varepsilon_s) \, d\varepsilon_s \, dx$$

where $f''$ is the Fermi function for holes and $g''$ is the pseudo-Fermi function for holes; i.e.,

$$f''(\varepsilon_s) = 1 - f(\varepsilon_s) = [1 + e^{\varepsilon_s(x-x_s)}]^{-1}$$

$$g''(x, \varepsilon_s) = 1 - \exp \left[ - e^{-2\varepsilon_s(x-x_s)} \right]$$

where $x_s$ was defined by (40).

If $x_s < x_{mn}$ for a given value of surface potential, the magnitude of $N_{ss}(\varepsilon)$ as found from (39) will be smaller than $N_{ss}(\varepsilon)$ found from (72). Therefore, the magnitude of $V_{ss}$, when $\psi_s$ was established by increasing the applied voltage from a large negative value, will be smaller than its magnitude when $\psi_s$ was established by decreasing the applied voltage from the maximum positive bias. The spread between these two magnitudes depends upon the maximum applied bias.

The magnitudes of $N_{ss}$ given by (39) and (72) were calculated for a system in which the oxide traps were distributed uniformly in energy throughout the forbidden gap of silicon and had a uniform spatial distribution with a total spatial concentration of $10^{16}$ cm$^{-3}$. In the computer solutions it was assumed that $T_s = 120$ seconds, $\sigma = 10^{-18}$ cm$^2$, $N_s = N_v = 2.5 \times 10^{19}$ cm$^{-3}$ and $(1/2K_o) = 1$ Å. The applied voltage was varied over a range sufficient to sweep the Fermi level at the surface of the silicon from the valence to the conduction bands. From the computed values of $N_{ss}$, the surface state potential $V_{ss}$ was calculated and the results are plotted in Fig. 10.

The dependence of the trap distribution function on the electric field in the oxide was omitted in these calculations. Since the zero-field distribution is assumed to be uniform in both position and energy, the electric field cannot change this distribution provided the field is uniform throughout the oxide.

To find the MOS capacitance as a function of bias for increasing bias we carried out the calculations called for by (64) (at $\omega = 10^4$ rad/s) and then used this calculated value for $dV_{ss}/d\psi_s$ in (50) to give the normalized capacitance curves. This result is plotted in Fig. 11 as the curve for increasing $V$.

The analogous calculation for decreasing bias starts with the calculation of

$$\frac{dV_{ss}(\varepsilon_s)}{d\psi_s} = \left( \frac{\varepsilon^2}{C_{ox}} \right) \frac{dN_{ss}(\varepsilon_s)}{d\psi_s} + \left( \frac{\varepsilon^2}{C_{ox}} \right) \frac{dN_{ss}(\varepsilon_s)}{d\varepsilon_s}$$

where the response time of the traps must be included in the expression for $dN_{ss}(\varepsilon_s)$. Thus,

$$\frac{dV_{ss}(\varepsilon_s)}{d\psi_s} = -\frac{\varepsilon^2}{C_{ox}} K_i \int_0^{\varepsilon_{ss}} \int_{\varepsilon_{so}}^{\varepsilon_{to}} g(x, \varepsilon_s) \left[ 1 + 2\sigma \tau_s(x, \varepsilon_s) \right]^{-1} \frac{\partial f''}{\partial \varepsilon_s} \, d\varepsilon_s \, dx$$

(76)

In writing (76), $K_i$ was removed from the integral because
it is a constant in both $\xi$ and $x$, and $g'(x, \xi)$ was not differentiated for the same reasons given in the argument leading to (64). Using the computer solution for (76) in (50) leads to the curve for decreasing $V$ shown in Fig. 11; the hysteresis in voltage is evident.

**Example 3—Experimental Determination of Surface State Density**

In this example we will use the computer to simulate the analysis of MOS capacitance vs. voltage measurements in order to extract the effective surface state density in accordance with the procedure described at the end of Section VI. We will assume a uniform spatial distribution of acceptor-type traps whose energy distribution is peaked about the midgap energy. For convenience, let

$$K_i(\xi_i, x) = \frac{N_i}{\tau_0} \left( \frac{x}{\tau_0^2} \right) \sin \left( \frac{\pi (\xi_i - \xi_{oc})}{\epsilon g} \right) \text{cm}^{-3} \text{ev}^{-1}$$

where the normalization insures that

$$\int_{\xi_{oc}}^{\xi_i} K_i(\xi_i) d\xi_i = N_i \tag{78}$$

We will avoid the hysteresis effect by performing the "measurements" for increasing $\psi$, only. We will also neglect the field-dependence of $K_i$. As yet we have not succeeded in carrying out the calculations taking this dependence into account. While the neglect of this field-dependence will result in a curve for $N_s$, which differs in detail from the correct curve, it will not change the gross features of the curve which are of physical interest. A more accurate computation is now being attempted.

Direct substitution of this distribution into (39), (64), and (50) yields the normalized capacitance vs. voltage curves shown in Fig. 12 for $\omega = 10^7, 10^8, 10^9, 10^{10}$ r/s. The values $N_i = 10^{19} \text{cm}^{-3}, N_0 = 10^{14} \text{cm}^{-3}, \tau_{ox} = 200 \text{Å}$, $T_m = 120 \text{ s}$ and $\sigma = 10^{-18} \text{ cm}^2$ were chosen. It is seen that a frequency of $10^8$ r/s is sufficiently high so that the capacitance is no longer frequency dependent.

The procedure calls for one to compare the total voltage of the experimental curve with the theoretical voltage if no surface states were present at each high frequency capacitance value. It is clear from (48) that this will just yield $V_{ss}$, the voltage equivalent of the total surface state charge. In Fig. 13, we have plotted the total number of trapped electrons as a function of surface potential, which was determined by substituting the sinusoidal distribution given above into (39). If one assumes that all of this charge resides in interface states and ignores the possibility of oxide traps, then the derivative of this curve, $dN_{ss}/d(\Phi_s)$ is the surface state density. This differentiation was carried out and the result appears in Fig. 14. There is little resemblance between this curve and the sinusoidal energy distribution originally assumed. Actually, the assumed distribution has its maximum density at midgap and is symmetrical about this energy, whereas the effective surface state density is asymmetrical about the midgap energy and exhibits a minimum density there! The physical explanation for this type of result is clear. Since the depth $x_m$ depends on the hole and electron concentrations at the surface, it is a minimum when the Fermi level is at midgap and only those traps located near the interface can communicate with the semiconductor in two minutes. There are insufficient carriers available for either hole or electron capture into traps away from the interface and these traps are located too far from either band edge for hole or electron emission. In fact, it is possible for $x_m$ to be zero over a range of surface potential for a wide band-gap material; i.e., it is impossible to alter even the interface state occupations in several minutes when the Fermi level is near the center of the forbidden gap.
Fig. 12. Normalized capacitance of a Si-SiO₂ structure for a sinusoidal oxide trap distribution.

Fig. 13. Total number of electrons trapped in the oxide for the sinusoidal oxide trap distribution. Parameters given in Fig. 12.

Fig. 14. Effective surface state density that would be measured with sinusoidal oxide trap distribution.

References


