Model for the generation of positive charge at the Si-SiO_2 interface based on hot-hole injection from the anode

Massimo V. Fischetti
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 17 May 1984; revised manuscript received 1 November 1984)

A comprehensive model for the generation of positive charge and fast interface states in metal-oxide-semiconductor structures during electron injection is quantitatively analyzed. According to this model, the injected electrons are accelerated in the SiO_2 conduction band by the external electric field. Once they reach the anode-SiO_2 interface, a significant fraction of them lose their kinetic energy by exciting surface plasma oscillations. The decay of these collective excitations into hot-electron-hole pairs results in the injection of holes into the oxide and their trapping at the Si-SiO_2 interface. The theoretical predictions agree with the experimental dependence of the phenomenon on anode field, temperature, gate material, and oxide thickness.

I. INTRODUCTION

The degradation of the Si-SiO_2 interface affects the reliability of metal-oxide-semiconductor (MOS) devices. This degradation is observed mainly after electron injection into SiO_2,[1-8] and is marked by the appearance of positive charge and fast states at the Si-SiO_2 interface. Considerable progress has been made in understanding the microscopic nature of these defects,[9,10] but the mechanism responsible for their generation remains poorly understood.

Several mechanisms have been proposed in the past, but none of them is completely satisfactory. Models based on the generation of electron-hole pairs in the SiO_2 via interband impact ionization[3,11,12] do not account for the positive charge which appears even when the applied gate voltage is less than 13 V,[13] which is the threshold for impact ionization. Models based on the release of hydrogen from water-related electron traps in SiO_2,[14] and its migration to the interface, fail to account for the large amount of positive charge generated in "ultradry" samples, provided the applied field is sufficiently high.[15] Furthermore, hydrogen diffusion[8,16] cannot account for the high rate at which positive charge (stable[17] or unstable[18]) is formed, even at temperatures as low as 77 K. Other models assume direct bond breaking by hot electrons,[8] but they cannot account for the presence of positive charge at the Si-SiO_2 interface during injection at positive bias,[1] because the electrons are not sufficiently hot at that interface. Other possible mechanisms involving diffusion of neutral species such as excitons have also been proposed.[19]

The purpose of this paper is to propose and evaluate quantitatively a model for the generation of the positive charge at the Si-SiO_2 interface during electron injection in MOS structures. This model is based on electron heating in the conduction band of the SiO_2 under the applied electric field. As the hot electrons reach the anode-SiO_2 interface, a significant fraction of them lose their kinetic energy by emitting surface plasmons. These are collective motions of the valence electrons of the anode localized at the anode-insulator interface. For a typical case of interest, the energy needed to excite a surface plasmon at the Al-SiO_2 interface is 8.5 eV. At oxide fields in the range of $(6-12)\times 10^6$ V/cm, a fraction of $10^{-3}-10^{-1}$ of the injected electrons will achieve kinetic energies higher than this value as they enter the anode.[20] These hot electrons emit surface plasmons with probability as high as 0.3. Once emitted, the surface plasmons will quickly decay mainly via the excitation of one electron-hole pair.[21] Some of the hot holes so generated (typically a fraction $10^{-4}-10^{-2}$, the higher fractions being obtained at higher values of the field in the insulator close to the anode-SiO_2 interface) will be injected into the SiO_2 valence band and subsequently could be trapped in the stressed region near the Si-SiO_2 interface,[9] giving rise to the positive charge observed experimentally. It should be noticed that mechanisms involving hole injection from the anode have been considered in the past.[22,23]

Support for this model comes from the experimental observation that the electric field at the anode-SiO_2 interface is the key parameter controlling the rate at which the positive charge is generated during electron injection.[15] Additional experimental work has provided further evidence that hole injection from the anode is likely to take place during electron injection into SiO_2.[13] The quantitative analysis presented in this paper is intended to provide a sound basis for this model and to show its quantitative consistency with the experimental data.

The paper is organized as follows. In Sec. II we analyze the interaction between the electrons and the surface plasmons in the simplified case of electrons which tunnel elastically from the cathode and are transported across the insulator without energy losses. The rate of emission of surface plasmons is obtained by using the effective-mass approximation.

The restriction of elastic transport in the SiO_2 is dropped in Sec. III, by introducing the electron-phonon interaction in the insulator. Then the efficiency for emission of surface plasmons is obtained by integrating the result of Sec. II over the energy distribution of the hot electrons at the anode-SiO_2 interface.

In Sec. IV, a further step of the mechanism is considered. Ignoring the details of the band structure of the

anode, each generated plasmon is assumed to decay into one electron-hole pair, the total energy being equally divided between the two carriers. The probability for a hole to be injected into the SiO_2 is evaluated. Finally, the hole trapping at the Si-SiO_2 interface is qualitatively evaluated, and in Sec. V the theoretical results are compared to the available experimental data.

II. INTERACTION BETWEEN ELECTRONS AND SURFACE PLASMONS

The mediation of the surface plasmons in the process of exciting electron-hole pairs near the anode-SiO_2 interface is important for two reasons (see Fig. 1). First, the efficiency for generating hot-electron–hole pairs via the plasmon-mediated process is much higher (as high as 0.1 or more) than the efficiency for the direct process, as explained below. Second, the holes are generated close to the Si-SiO_2 interface and have a high probability of being injected into the SiO_2. The first reason is explained by the picture suggested by Pines.[21] Low-energy electrons having energies (measured from the Fermi level of the metal or the bottom of the Si conduction band) of 1—5 eV will lose their energy mainly to single-particle excitations. If, on the contrary, the electrons have energies higher than what is needed to excite a collective motion (typically 10—30 eV in the bulk, specifically 8.5 and 9.5 eV at the Al-SiO_2 and Si-SiO_2 interfaces), the larger density of final states provided by these collective excitations favors them with respect to the excitation of a high-energy electron-hole pair. Indeed, all valence electrons take part in a collective motion, while only a few electrons are involved in intraband or interband transitions of a given energy. Moreover, the Coulomb screening in the solid has been shown by Nozières and Pines[24] to favor a single large energy loss to a collective mode over a loss of a lower energy to a few electron-hole pairs. Instances in which the direct electron-electron interaction might have an importance higher than (or comparable to) the plasmon-mediated process are given by low-field situations (very few carriers are sufficiently energetic to excite surface plasma modes), and when considering anode materials with a very low plasma frequency at the anode-SiO_2 interface. Gold is a good example of this situation, since its surface plasma energy, $\hbar\omega_{SP}$, is about 2.5 eV and the plasmon-generated holes will not be sufficiently energetic to efficiently tunnel into SiO_2. In these cases, neglecting the direct excitation of single-particle states results in underestimating the quantum efficiency of the process.

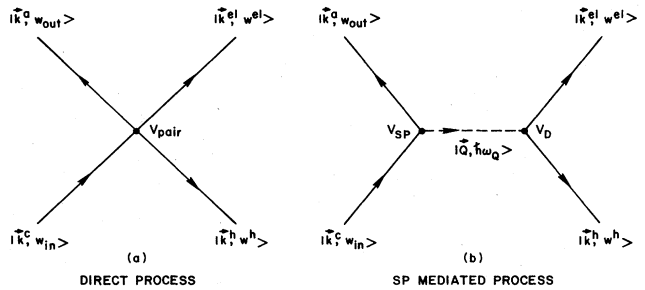

FIG. 1. Diagram representing the (first-order) generation of one electron-hole pair by one electron entering the anode with energy w_{in}. (a) represents the direct excitation of a single-particle state. In (b) this excitation is mediated by a surface plasmon (SP). Process (b) is more efficient than process (a) when $w_{in} \simeq \hbar\omega_{SP}$. If Δw is the energy of the electron-hole pair, for $\Delta w \ll \hbar\omega_{SP}$ one has $V_{pair}/V_{SP} \simeq (\Delta w/\hbar\omega_{SP})^2 \ll 1$. If $\Delta w \simeq \hbar\omega_{SP}$, then $V_{pair}/V_{SP} \simeq n'/n_v \ll 1$, where n_v is the electron density in the valence band and n' is the density of the electrons which can take part to the intraband or interband transition. In the diagram, the vertex V_D represents the SP damping and it is assumed to yield a rate much higher than the SP emission process.

In order to evaluate the interaction between the electrons and the surface plasmons, it is convenient to start by considering a monoenergetic beam of electrons injected into SiO_2, and compute the efficiency for producing a surface plasmon localized at the anode-SiO_2 interface, in much the same way as this process is evaluated in metal-insulator-metal structures.[25–28]

The interaction between the electron and the electromagnetic field associated with the surface plasmon is given by the Hamiltonian

$$H_{int} = e \int d\mathbf{r}\, \phi(\mathbf{r}) \psi(\mathbf{r})^{\dagger} \psi(\mathbf{r}) \tag{1}$$

in the electrostatic limit. Here, $\phi(\mathbf{r})$ is the surface-plasmon electrostatic potential and $\psi(\mathbf{r})$ is the electron field.

In order to obtain the matrix element corresponding to the emission of one surface plasmon by one injected electron, the interaction term given in Eq. (1) must be written in a second-quantized form. This is obtained by promoting the fields ϕ and ψ to operators, expressing them in terms of creation and annihilation operators (that is, their Fourier components), and extracting the term representing the annihilation of one electron with the simultaneous creation of a surface plasmon and a new outgoing electron.

A. Quantization of the surface-plasmon field ϕ

The quantization of the electromagnetic field associated with the surface plasmon in a MOS structure is, in general, a difficult task. Ngai and Economou[25] have provided a general scheme allowing the reduction of this problem to the simpler case of the quantization of a scalar field (the electrostatic potential). This is done by considering only surface plasmons of short wavelength and neglecting the imaginary part of their dispersion relation, that is, their damping. The first approximation is certainly justified in the present work, since only surface plasmons of short wavelength are localized at the anode-SiO_2 interface, where the electrons have sufficient kinetic energy to excite the plasmon modes. The finite lifetime of the surface plasmons corresponds to a broadening of the emission line and will be considered in Sec. III. Thus, following the calculation performed in Ref. 25 step-by-step, and treating the insulator as an ideal dielectric characterized by a dielectric constant κ_{ox}, one can write (in the notation of Ref. 25)

$$\phi(\mathbf{r})=\sum_{Q}\frac{1}{4\pi\epsilon_0}\left[\frac{\epsilon_0\hbar\omega_Q}{QG_Q}\right]^{1/2}\times\exp(i\mathbf{Q}\cdot\mathbf{R})\Lambda_Q(x)(c^{\dagger}_{-Q}+c_Q)\,, \tag{2}$$

where ω_Q is the frequency of the surface plasmon, x is the coordinate normal to the anode-SiO_2 interface, $\mathbf{R}$ is the spatial coordinate in the interfacial plane, $\mathbf{Q}$ is the plasmon wave vector, and $c^{\dagger}_Q$ and c_Q are the surface-plasmon creation and annihilation operators, respectively. In the electrostatic ($Q\gg Q_0=\omega\kappa_{ox}^{1/2}/c$) and high-momentum ($Q^{-1}\ll$oxide thickness) limits considered in Ref. 25, one has

$$\omega_Q\simeq\omega_{ba}/(1+\kappa_{ox})^{1/2}\,, \tag{3}$$

where κ_{ox} is the SiO_2 high-frequency dielectric constant [$\simeq$2.2 (Ref. 29)] and ω_{ba} is the bulk plasma frequency of the anode. In deriving Eq. (3), the Drude expression for the dielectric constant of the anode has been assumed. This is a good approximation for Si and Al (definitely not for Au) in the range of energies of interest. Figure 2 illustrates the surface-plasmon dispersion relation for an Al-SiO_2-Si system with oxide thickness $d_{ox}=200$ Å under these approximations. It should be stressed that deviations from ideality of the SiO_2-anode interface, such as interface roughness, could further enhance the reduction of the plasma frequency at the interface[30] from the values of 8.5 eV (Al) and 9.45 eV (Si) as computed from Eq. (3). In the same limit, the function G_Q appearing in Eq. (2) is

$$G_Q\simeq\frac{\exp(-Qd_{ox})}{2(1+\kappa_{ox})}\,, \tag{4}$$

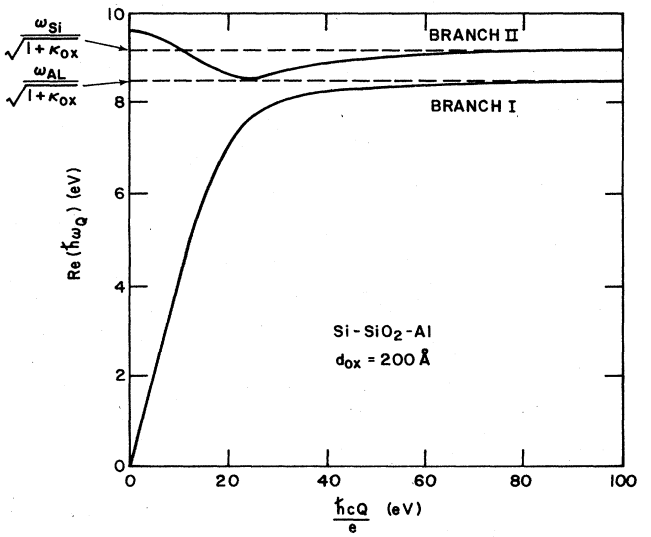

FIG. 2. Dispersion relation for the surface plasmons in a Si-SiO_2-Al system with an oxide thickness of 200 Å. Branch I extends only down to values of Q such that oscillation damping is negligible (Ref. 25). For high values of Q, it represents a mode localized at the Al-SiO_2 interface. In the same limit, branch II corresponds to a mode localized at the Si-SiO_2 interface. In the low-Q region, ω_Q^{II} depends on the oxide thickness, approaching the value $[(\omega_{Al}^2+\omega_{Si}^2)/2]^{1/2}$ for $d_{ox}\to 0$.

and, to the leading order in $\exp(-Qd_{ox})$, the function Λ_Q, expressing the spatial variation of the plasmon field, is given by

$$\Lambda_Q(x)\simeq\begin{cases}-\dfrac{\exp(-Qx)}{1+\kappa_{ox}} & \text{if } x\geq d_{ox}/2 \quad \text{(anode)}\,,\\ -\dfrac{\exp(-Qd_{ox})}{1+\kappa_{ox}}\exp(Qx) & \text{if } -d_{ox}/2\leq x\leq d_{ox}/2 \quad \text{(oxide)}\,,\\ O[\exp(-\frac{3}{2}Qd_{ox})] & \text{if } x\leq -d_{ox}/2 \quad \text{(cathode)}\,.\end{cases} \tag{5}$$

A second branch exists in the surface plasmon-dispersion relation, as shown in Fig. 2. In the high-Q limit, this branch corresponds to a mode with frequency $\omega_{bc}/(1+\kappa_{ox})^{1/2}$ (where ω_{bc} is the bulk plasma frequency of the cathode) and with energy localized at the cathode-SiO_2 interface. At low Q, branch II corresponds to oscillations of small amplitude, while branch I has low energy. Thus, only the high-Q limit of the "anodic" branch (I or II) is of interest, as stated above.

B. Quantization of the electron field ψ

The electron field can be quantized in the effective-mass approximation with parabolic bands with the resulting semiclassical external potential illustrated in Fig. 3. In the following a single- (isotropic) electron mass m will be considered. This will greatly simplify the notation, and the introduction of different effective masses will not

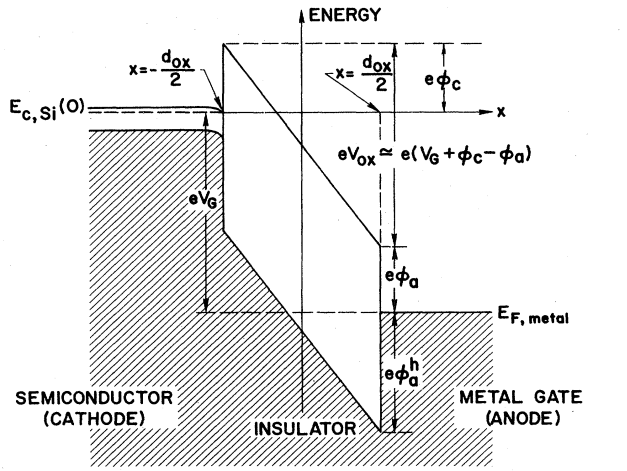

FIG. 3. Schematic diagram of the effective potential vs distance for a MOS structure for the case of positive bias applied to the metal.

change the final results in a significant way. By using these approximations, the field can be expanded over two (unitarily equivalent) convenient single-particle basis eigenfunctions:

$$\psi(\mathbf{r})=\sum_{\mathbf{k}}[a_{\mathbf{k}}^{(c)}\chi_{k_\perp}^{(c)}(x)\exp(i\mathbf{k}_{||}\cdot\mathbf{R})+b_{\mathbf{k}}^{(c)}\xi_{k_\perp}^{(c)}(x)\exp(i\mathbf{k}_{||}\cdot\mathbf{R})]$$
$$=\sum_{\mathbf{k}}[a_{\mathbf{k}}^{(a)}\chi_{k_\perp}^{(a)}(x)\exp(i\mathbf{k}_{||}\cdot\mathbf{R})+b_{\mathbf{k}}^{(a)}\xi_{k_\perp}^{(a)}(x)\exp(i\mathbf{k}_{||}\cdot\mathbf{R})]\ . \tag{6}$$

where $\chi_{k_\perp}^{(c)}$ and $\chi_{k_\perp}^{(a)}$ are normalized solutions of the effective-mass one-dimensional equation representing one-electron tunneling from the cathode and from the anode, respectively. Their expressions are given in the Appendix. The ξ's are eigenfunctions orthogonal to the χ's, and $a_{\mathbf{k}}$ and $b_{\mathbf{k}}$ are the corresponding particle operators. The notation $|\mathbf{k}\rangle_a=a_{\mathbf{k}}^{(a)\dagger}|0\rangle$ and $|\mathbf{k}\rangle_c=a_{\mathbf{k}}^{(c)\dagger}|0\rangle$ will be used to distinguish the one-electron states corresponding to the two representations appearing in Eq. (6). Finally, $\mathbf{k}_{||}$ and $k_\perp$ are the components of the wave vector parallel and normal to the plane of the cathode-SiO_2 interface.

C. Surface-plasmon emission rate (monoenergetic electrons)

The interaction term given by Eq. (1) can be rewritten in terms of the second-quantized fields given by Eqs. (2) and (6). A simplification can be obtained by neglecting the residual small interaction near the cathode. Furthermore, the electron—surface-plasmon interaction in the oxide, although *a priori* strong since the electrons are completely unscreened, is effective only when the electrons have energy higher than $\hbar\omega_Q$ before entering the anode. In the range of electric field values of interest here, very few electrons are expected to have such a high energy in the oxide. Therefore, the integration in Eq. (1) can be performed only in the anode (i.e., for $x\geq d_{ox}/2$). Then, the term expressing the emission of one surface plasmon by one electron coming from the cathode is given by

$$H'_{\rm int}=e\pi\left[\frac{\hbar}{\epsilon_0}\right]^{1/2}\sum_{\mathbf{k},\mathbf{k}',\mathbf{Q}}V_{\mathbf{Q},\mathbf{k},\mathbf{k}'}c_{\mathbf{Q}}^{\dagger}a_{\mathbf{k}}^{(a)\dagger}a_{\mathbf{k}'}^{(c)}\ , \tag{7}$$

where

$$V_{\mathbf{Q},\mathbf{k},\mathbf{k}'}=\delta_{\mathbf{Q},\mathbf{k}'_{||}-\mathbf{k}_{||}}\left[\frac{\omega_Q}{QG_Q}\right]^{1/2}I_{Q,k_\perp,k'_\perp}\ ,$$
$$I_{Q,k_\perp,k'_\perp}=\int_{ox/2}^{\infty}dx\ \Lambda_Q(x)\chi_{k_\perp}^{(a)*}(x)\chi_{k'_\perp}^{(c)}(x)\ . \tag{8}$$

Defining a matrix element

$$M(Q,k_\perp^{(c)},k_\perp^{(a)})=e\pi\left[\frac{\hbar\omega_Q}{\epsilon_0 QG_Q}\right]^{1/2}I_{Q,k_\perp^{(c)},k_\perp^{(a)}}\ ,$$

the rate of surface-plasmon emission, $R_{\rm SP}$, can be obtained by the Fermi golden rule. To leading order in M, one has

$$R_{\rm SP}=\frac{2\pi}{\hbar}\sum_{\mathbf{Q},\mathbf{k}^{(a)},\mathbf{k}^{(c)}}|\,{}_a\langle\mathbf{k}^{(a)},\mathbf{Q}\,|\,H'_{\rm int}\,|\,\mathbf{k}^{(c)}\rangle_c\,|^2\delta_{w^{(a)}+\hbar\omega_Q,w^{(c)}}$$
$$=\frac{2\pi}{\hbar}\sum_{\mathbf{Q},\mathbf{k}^{(a)},\mathbf{k}^{(c)}}|M(Q,k_\perp^{(c)},k_\perp^{(a)})|^2 n_{k^{(c)}}(1-n_{k^{(a)}})(1+N_Q)\delta_{\mathbf{Q},\mathbf{k}_{||}^{(c)}-\mathbf{k}_{||}^{(a)}}\delta_{w^{(a)}+\hbar\omega_Q,w^{(c)}}\ , \tag{9}$$

where n_k and N_Q are the Fermi and Bose thermally averaged occupation numbers for the electrons and plasmons, respectively.

Transforming, now, the sums over momenta to integrals over energies, ignoring tunneling in the reverse direction (only large applied bias will be considered), and carrying out the integrations over the δ functions, one has

$$R_{\rm SP}=\frac{Am^{5/2}}{2^{1/2}(2\pi)^6\hbar^6}\int_0^{2\pi}d\phi\int_{Q_0}^{\infty}dQ\int_0^{\infty}dw_{||}^{(c)}\int_0^{\infty}dw_\perp^{(c)}\frac{Q}{(w_\perp^{(c)})^{1/2}K_a}|M(Q,k_\perp^{(c)},k_\perp^{(a)})|^2$$
$$\times[N(\hbar\omega_Q)+1]f(w_{||}^{(c)}+w_\perp^{(c)})[1-f(w_{||}^{(c)}+w_\perp^{(c)}+eV_G-\hbar\omega_Q)]\ , \tag{10}$$

where $w_\perp$ and $w_{||}$ are the transverse and parallel energy components defined by $\hbar^2k_\perp^2/2m$ and $\hbar k_{||}^2/2m$, f is the cathode Fermi function, ϕ is the angle between the ingoing electron wave vector $\mathbf{k}_{||}^{(c)}$ and the plasmon wave vector $\mathbf{Q}$, V_G is the external bias, and A is the area of the device. The normal components $k_\perp$ of the electron wave vectors are functions of the energy, as follows from the integration over the δ functions expressing energy and momentum conservation:

$$k_\perp^{(c)}=(2mw_\perp^{(c)})^{1/2}/\hbar\ ,$$
$$k_\perp^{(a)}=\left[\frac{2mw_\perp^{(c)}}{\hbar^2}-Q^2-\frac{2m\omega_Q}{\hbar}+\frac{2Q(2mw_{||}^{(c)})^{1/2}}{\hbar}\cos\phi\right]^{1/2}\ ,$$

and the anode wave vector K_a has been defined as

$[(k_\perp^{(a)})^2+2meV_G/\hbar^2]^{1/2}$.

The final result of this section, Eq. (10), represents the rate at which electrons of energy $w^{(c)}$ are injected into the insulator and emit surface plasmons of momentum Q in the anode, integrated over all allowed initial and final states.

III. EMISSION OF SURFACE PLASMONS BY THE HOT ELECTRONS

Equation (10) has been obtained by assuming that the electron tunneling and transport in the SiO_2 occur elastically. The inelastic tunneling channels are the primary interest in studying the conductance of thin-insulator tunnel junctions, as in Refs. 25–27. In the present work, however, much thicker MOS structures are of interest. In this case the electrons travel across tens or hundreds of angstroms in the SiO_2 conduction band before reaching the anode-insulator interface, undergoing many collisions with the lattice vibrations. Therefore inelastic tunneling can be ignored at first order, but we must account for the fact that the electron-energy distribution in the anodic region will be significantly different from the monoenergetic situation considered in Sec. II.

To account for this, Eq. (9) can be modified by introducing intermediate states $|\mathbf{k}^{(f)}\rangle_a$ of energy $w^{(f)}$, weighted by a density matrix ρ whose elements $\rho(\mathbf{k},\mathbf{k}')$ represent the probability that an electron having entered the insulator near the cathode in the state $|\mathbf{k}\rangle_a$ will reach the anodic region in the state $|\mathbf{k}'\rangle_a$. If the electron transport in the SiO_2 occurs in a steady-state regime, the matrix ρ is actually independent of the initial state $|\mathbf{k}\rangle_a$. Then, the surface-plasmon emission rate can be expressed as

$$R_{\rm SP}=\frac{2\pi}{\hbar}\sum_{\mathbf{Q},\mathbf{k}^{(a)},\mathbf{k}^{(c)}}\sum_{\mathbf{k}'^{(f)},\mathbf{k}^{(f)}}|\,{}_a\langle\mathbf{k}^{(a)},\mathbf{Q}\,|\,H'_{\rm int}\,|\,\mathbf{k}^{(f)}\rangle_a\,|^2\rho(\mathbf{k}^{(f)})\,|\,{}_a\langle\mathbf{k}'^{(f)}\,|\,\mathbf{k}^{(c)}\rangle_c\,|^2\delta_{w^{(a)}+\hbar\omega_Q,w^{(f)}}$$
$$=\frac{2\pi}{\hbar}\sum_{\mathbf{k}^{(c)},\mathbf{k}'^{(f)}}|\,{}_a\langle\mathbf{k}'^{(f)}\,|\,\mathbf{k}^{(c)}\rangle_c\,|^2\sum_{\mathbf{k}^{(a)},\mathbf{k}^{(f)},\mathbf{Q}}\rho(\mathbf{k}^{(f)})\,|\,{}_a\langle\mathbf{k}^{(a)},\mathbf{Q}\,|\,H'_{\rm int}\,|\,\mathbf{k}^{(f)}\rangle_a\,|^2\delta_{w^{(a)}+\hbar\omega_Q,w^{(f)}}\,. \tag{11}$$

In order to obtain the quantum efficiency for emission of surface plasmons, the rate at which electrons are injected elastically into the SiO_2 conduction band, R_T, and the density matrix ρ, will be considered next. The quantum efficiency $\eta_{\rm SP}$ will be obtained in Sec. III C.

A. Rate of elastic tunneling of electrons into SiO_2

The elastic tunneling rate R_T can be obtained by considering the matrix element

$${}_a\langle\mathbf{k}^{(a)}\,|\,\mathbf{k}^{(c)}\rangle_c=\langle 0\,|\,a^{(a)}_{\mathbf{k}^{(a)}}a^{(c)\dagger}_{\mathbf{k}^{(c)}}\,|\,0\rangle\,,$$

which can be evaluated by finding the unitary transformation connecting the two bases employed in Eq. (6). Multiplying the transmission amplitude $|\,{}_a\langle\mathbf{k}^{(a)}\,|\,\mathbf{k}^{(c)}\rangle_c\,|^2$ by the flux of incoming particles,

$$(\hbar/2im)(\chi^{(c)}_{k_\perp^{(c)}}\partial_x\chi^{(c)*}_{k_\perp^{(c)}}-\chi^{(c)*}_{k_\perp^{(c)}}\partial_x\chi^{(c)}_{k_\perp^{(c)}})\,,$$

ignoring the effect of the anode Fermi function, and summing over the density of states, one obtains

$$R_T\simeq\frac{Am}{4\pi^2\hbar^3}\int_0^\infty dw_\parallel^{(c)}\int_0^\infty dw_\perp^{(c)}F(w_\perp^{(c)})f(w_\perp^{(c)}+w_\parallel^{(c)})\,, \tag{12}$$

where $F(w_\perp^{(c)})$ is an algebraically complicated function which reduces to the usual expression for Fowler-Nordheim tunneling[31] when employing the WKB approximation described in the Appendix:

$$F(w_\perp^{(c)})\simeq 8\left[\frac{w_\perp^{(c)}}{e\phi_c}\right]^{1/2}\exp\left[-\frac{\beta}{E_{\rm ox}}\left[1-\frac{w_\perp^{(c)}}{e\phi_c}\right]^{3/2}\right]. \tag{13}$$

In deriving Eq. (12), it has been assumed that $eV_G\gg e\phi_c$ (the cathode barrier height) and that $eV_G\gg w_\perp^{(c)}$, conditions which are certainly met in the case of oxide thickness larger than 60 Å. Also, a factor 2 has been introduced to account for the electron spin. $E_{\rm ox}$ is the electric field in SiO_2 ($\simeq V_G/d_{\rm ox}$, neglecting the work-function difference). The coefficient β above is given by the usual expression,

$$\beta=\frac{4}{3}\frac{(2m_{\rm ox})^{1/2}}{e\hbar}(e\phi_c)^{3/2}\,.$$

where the electron effective mass $m_{\rm ox}$ in the SiO_2 band gap has been restored.

B. Electron-energy distribution in SiO_2

The calculation of the density matrix ρ is an open problem. In the past years scattering with longitudinal-optical (LO) phonons has been considered to be the dominant mechanism for electron-energy loss in SiO_2.[32–34] With this assumption, the theoretical calculation predicted the existence of a critical field [usually varying from as low as 1.5 MV/cm (Ref. 34) to 10 MV/cm or higher (Refs. 32 and 33)], above which the LO phonons could not prevent an unlimited acceleration of the carriers and the distribution would become unstable. At sufficiently high energies, impact ionization would become effective.[32]

Recent experimental results obtained from direct measurements of the energy of the electrons injected into SiO_2 at high fields,[20,35] and recent Monte Carlo simulations for electron transport in SiO_2,[36] actually suggest a different picture. At electron energies of about 3 eV and starting from 3 to 4 MV/cm, the LO-phonon, uncontrolled runaway is limited by scattering with the acoustic phonons. This occurs much below the threshold for interband impact ionization and it results in the existence of a steady-state electron distribution at energies higher than that predicted by the LO-phonon-based calculation. These recent results, both experimental and theoretical, provide

good estimates of the average electron energy. However, additional problems arise when an accurate determination of the high-energy tail of the electron-energy distribution is needed. Referring to the work by Theis *et al.*,[20] DiMaria *et al.*,[35] and Fischetti[36] for a discussion of this issue, in the present context one is left with the problem of evaluating $\rho(w^{(f)})$ in the high-energy region. Since accurate data in the high-energy tail are not yet available, scattering with the LO phonons only will be considered here, although it should be stressed again that this procedure is employed only because of our lack of knowledge of the electron transport in amorphous SiO_2, when the electron energy is higher than the width of the SiO_2 conduction band. The nonparabolicity of the bands, the "collision broadening" due to the uncertainty of the electron energy between successive collisions (as required by the uncertainty principle at high collision rates), and the breakdown of the semiclassical approach as the electron mean free path becomes comparable to the interatomic distance are all effects which, at present, prevent our gaining an accurate knowledge of the high-energy tail of the electron distribution.

When solving the Boltzmann equation, attention must be paid to the large-angle ($>\pi/2$) scattering with the LO phonons. As pointed out by Dumke in a different situation,[37] this scattering event is needed to prevent the electron distribution from running away at low fields. Thus a steady-state distribution can be obtained up to fields of 7.5 MV/cm (Ref. 32). For fields above this threshold an energy cutoff will be introduced. The only justification for doing so lies in the fact that the final results will be almost unaffected by the choice of this cutoff in the range 30—60 eV for fields up to 11 MV/cm. An approximated steady-state solution of the Boltzmann equation has been given by Dumke[37] by considering only electrons with a large component of longitudinal momentum $k_\perp$. With this approximation and considering a parabolic conduction band in SiO_2, one has

$$\rho(w^{(c)},w^{(f)})\simeq\rho_{\rm LO}(\bar{w}_\perp^{(f)}) \simeq \begin{cases} \rho_0\exp\left[-\int_{\hbar\omega_{\rm LO}}^{\bar{w}_\perp^{(f)}} d\bar{w}_\perp\sigma(\bar{w}_\perp)\right]=\rho_0\exp[-\Sigma(\bar{w}_\perp^{(f)}] & \text{for } \bar{w}_\perp^{(f)}>\hbar\omega_{\rm LO}\ , \\ \rho_0 \quad \text{for } \bar{w}_\perp^{(f)}\le\hbar\omega_{\rm LO}\ . \end{cases} \tag{14}$$

Here the energies $\bar{w}_\perp$ are measured from the bottom of the SiO_2 conduction band at the anode, and the function $\sigma(\bar{w}_\perp)$, neglecting small terms of order $1/\bar{w}_\perp$, is given by

$$\sigma(\bar{w}_\perp)=\left[\frac{1}{\hbar\omega_{\rm LO}}-\frac{e\lambda_S E_{\rm ox}\bar{w}_\perp}{(\hbar\omega_{\rm LO})^2\ln(4\bar{w}_\perp/\hbar\omega_{\rm LO})}\right]+\left[\left[\frac{1}{\hbar\omega_{\rm LO}}-\frac{e\lambda_S E_{\rm ox}\bar{w}_\perp}{(\hbar\omega_{\rm LO})^2\ln(4\bar{w}_\perp/\hbar\omega_{\rm LO})}\right]^2+\frac{2\ln(2)+e\lambda_S E_{\rm ox}}{(\hbar\omega_{\rm LO})^2\ln(4\bar{w}_\perp/\hbar\omega_{\rm LO}}\right]^{1/2}. \tag{15}$$

Using the Fröhlich Hamiltonian, the scattering length per unit energy, λ_s, is given by[28–30]

$$\lambda_s=\frac{4\pi\hbar\varepsilon_0}{e^2 m^*_{\rm ox}\omega_{\rm LO}}\left[\frac{1}{\kappa_\infty}-\frac{1}{\kappa_0}\right]^{-1}. \tag{16}$$

To give some feeling of the mean free paths involved in these interactions, by assuming a value of $m^*_{\rm ox}\simeq 1.3m$ for the polaron-corrected electron mass in SiO_2,[38] setting $\hbar\omega_{\rm LO}=0.153$ eV and $\kappa_\infty^{-1}-\kappa_0^{-1}\simeq 0.143$ (Ref. 33), one has

$$\frac{\lambda_s}{\ln(4w_\perp/\hbar\omega_{\rm LO})}\simeq\frac{18.5}{\ln(4w_\perp/\hbar\omega_{\rm LO})}\ \text{Å/eV}\ \text{(small-angle scattering)}\ ,$$

$$\lambda_s/\ln(2)\simeq 27\ \text{Å/eV}\ \text{(large-angle scattering)}\ .$$

The low-energy LO phonons (0.063 eV) can be included trivially in this picture. The corresponding parameter λ_s is about 108 Å/eV. The effects of the lattice temperature can be included by modifying the scattering parameter λ_s:

$$\lambda_s(T)=\frac{\lambda_s}{1+2N_{\rm ph}}=\lambda_s\tanh\left[\frac{\hbar\omega_{\rm LO}}{2k_B T}\right], \tag{17}$$

where the thermally averaged phonon number $N_{\rm ph}$ is given by $[\exp(\hbar\omega_{\rm LO}/k_B T)-1]^{-1}$ in the absence of dispersion.

C. Quantum efficiency for surface-plasmon emission

The quantum efficiency for surface-plasmon emission, $\eta_{\rm SP}$, can be now computed by simply taking the ratio $R_{\rm SP}/R_T$ from Eqs. (11) and (12). In the following, the zero-temperature expression for the Fermi functions is considered. Then, evaluating the sums in Eq. (11) with the same procedure employed in Sec. II [Eq. (9)], the quantum efficiency $\eta_{\rm SP}$ becomes

$$\eta_{SP}=\frac{m^{3/2}e^2\pi^2}{2^{1/2}(2\pi)^4\epsilon_0\hbar^2}\int_0^{2\pi}d\phi\int_{Q_0}^{\infty}dQ\int_{-e(V_G-\phi_a)}^{\infty}dw_\perp^{(f)}\int_{-e(V_G-\phi_a)}^{\infty}dw_\parallel^{(f)}\rho(w^{(f)})\frac{\omega_Q}{G_Q K_a[w_\perp^{(f)}+e(V_G-\phi_a)]^{1/2}}$$

$$\times|I(\mathbf{Q},\mathbf{k}^{(f)},\mathbf{k}^{(a)})|^2[N(\hbar\omega_Q)+1]\,, \tag{18}$$

where the integration over the energy $w^{(f)}$ is extended above the bottom of the SiO_2 conduction band at the anode, the matrix elements $I(\mathbf{Q},\mathbf{k},\mathbf{k}')$ are now taken between the anodic wave functions $\chi_{k_\perp}^{(a)}$, as shown in Eq. (11), and the wave vector $\mathbf{k}^{(a)}$ can be expressed in terms of the integration variable $w^{(f)}$, as was done in Eq. (10).

Inserting, now, the density matrix for the intermediate states given by Eq. (14) into Eq. (18), one finally obtains the quantum efficiency for emission of surface plasmons. In order to perform the integration in Eq. (18), the following approximations are used: (1) $w_\parallel^{(f)}$ is neglected (i.e., it is assumed that $w_\perp^{(f)}\gg w_\parallel^{(f)}$, the electron wave vectors being mostly in the forward direction), consistently with the derivation of Eq. (13), so that the integral over ϕ is trivial. (2) The dispersion relation of the surface-plasmon energy is ignored and, accordingly, (3) the limit $Qd_{ox}\to\infty$ is considered, as discussed in Sec. II A. Then, with the notation given in the Appendix,

$$I(Q,k_\perp^{(f)},k_\perp^{(a)})\simeq-\frac{1}{1+\kappa_{ox}}\frac{1}{|g^{(a)}(k_\perp^{(f)})|^{1/2}|g^{(a)}(k_\perp^{(a)})|^{1/2}}$$

$$\times\int_{d_{ox}/2}^{\infty}dx\,e^{-Qx}[\mu_1^{(a)*}(k_\perp^{(a)})e^{-iK_a x}+\mu_1^{\prime(a)*}(k_\perp^{(a)})e^{iK_a x}][\mu_1^{(a)}(k_\perp^{(f)})e^{iK_f x}+\mu_1^{\prime(a)}(k_\perp^{(f)})e^{-iK_f x}]\,,$$

with $K_f=[2m(w_\perp^{(f)}+eV_G)]^{1/2}/\hbar$. (4) The WKB approximation for the coefficients μ with large external bias $[eV_G\gg e\phi_a\gg w_\perp^{(a)}-E_{c,a}(0)]$ is employed. (5) The limit $Q^2\ll(K_f\pm K_a)^2\approx K_f^2$ is considered, since even in the electrostatic limit the electron wave vector is larger than the wave vector of the surface plasmons. (6) Finally, the surface-plasmon occupation number $N(\hbar\omega_Q)$ is neglected, since the high value of $\hbar\omega_Q/k_BT$ causes a negligible number of surface plasmons to be present at thermal equilibrium, and the strong plasmon damping largely inhibits the possibility of stimulated emission. Then, one obtains for the efficiency η_{SP} for the generation of one surface plasmon per injected electron,

$$\eta_{SP}\simeq\frac{\sqrt{2}}{8}\frac{e^2m^{1/2}(e\phi_a)^{1/2}\omega_{SP}}{\pi\epsilon_0(1+\kappa_{ox})}\Phi(E_{ox})\,. \tag{19}$$

The function $\Phi(E_{ox})$, modified to include a Lorentzian plasmon line shape (i.e., the finite plasmon lifetime, so far neglected), and accounting for the zero-temperature Fermi function in the anode, is

$$\Phi(E_{ox})=\frac{\int_{\hbar\omega_{LO}+e\phi_a}^{\infty}dw_\perp(\exp\{[-\Sigma(w_\perp-e\phi_a)]L(w_\perp)\}/(w_\perp-e\phi_a)w_\perp)}{\hbar\omega_{LO}+\int_{\hbar\omega_{LO}+e\phi_a}^{\infty}dw_\perp\exp[-\Sigma(w_\perp-e\phi_a)]}\,. \tag{20}$$

The integrated line-shape factor is assumed to be of the form

$$L(w_\perp)=\frac{1}{2}+\frac{1}{\pi}\tan^{-1}\left[\frac{w_\perp-\hbar\omega_{SP}}{\Gamma}\right]. \tag{21}$$

Here, $\hbar\omega_{SP}$ is the surface-plasmon energy in the high-Q limit, $\Gamma=\mathrm{Im}(\hbar\omega_{SP})/2$ is the plasmon linewidth ($\simeq$2 eV for Si, 0.4 eV for Al), and ϕ_a is the anode-SiO_2 barrier height. In deriving Eqs. (19) and (20), the narrow-line limit ($\Gamma\ll\hbar\omega_{SP}$) has been considered. To give an idea of the strength of the electron–surface-plasmon interaction, it can be noticed that an electron entering the Al or Si anode with "resonant" energy $\hbar\omega_{SP}$ has a probability of about 0.3 of exciting a surface plasma oscillation.

IV. HOLE INJECTION FROM THE ANODE

The overall efficiency for emission of real ($w_\perp\geq\hbar\omega_{SP}$) or virtual ($w_\perp<\hbar\omega_{SP}$) surface plasmons, after averaging over the electron distribution, does not exceed 10^{-1} in the range of fields we are concerned with. Also, the surface plasma oscillations are strongly damped both in Al and Si (in 10^{-16}–10^{-15} sec). This implies that the average number of surface plasmons in a steady-state regime will be negligible as long as less than 10^{16}–10^{17} electrons/cm^2 sec will be injected (i.e., 10^{-3}–10^{-2} A/cm^2). Therefore the products of the decay of the plasma modes are generated at the same rate at which the plasmons themselves are excited.

A. Hole tunneling probability

The decay of the surface plasmons will be assumed to occur mainly via excitation of electron-hole pairs, thus neglecting possible competitive channels.[28] This process can be viewed as the "time reversal" of the emission process. However, it occurs at a much higher rate. As pointed out by Ngai and Economou,[25] a surface plasma oscillation has very high components of momentum perpendicu-

lar to the interface, because of its exponential decay along that direction. This induces a strong coupling between the surface-plasmon field and the single-particle excitations in the anode, which results in strong damping. The detailed transitions involved in the plasmon decay are determined by the band structure and the oscillator strength of interband and intraband transitions in the anode. In order to compute the rate at which the generated holes will be injected into the SiO_2, the crude approximation will be made that the plasmon energy is equally divided between the electrons and the holes.[39] If $S(w^{(h)},w_\perp^{(h)})$ is the probability that a hole of total energy $w^{(h)}$ has transverse momentum $k_\perp^{(h)}=(2mw_\perp^{(h)})^{1/2}/\hbar$ pointing toward the anode-insulator interface, and $\tau_\perp(w_\perp^{(h)})^{-1}$ is the rate at which such a hole tunnels into the SiO_2, the rate $\tau(w^{(h)})^{-1}$ at which holes of total energy $w^{(h)}$ enter the oxide is given by

$$\frac{1}{\tau(w^{(h)})}=\int_0^{w^{(h)}} dw_\perp^{(h)} \frac{S(w^{(h)},w_\perp^{(h)})}{\tau_\perp(w_\perp^{(h)})} .$$

Assuming an isotropic distribution of holes and employing the WKB approximation for $\tau_\perp(w_\perp^{(h)})^{-1}$, one obtains

$$\frac{1}{\tau(w^{(h)})}\simeq\frac{8\hbar}{mw^{(h)}}\int_0^{w^{(h)}} dw_\perp^{(h)}\exp(-2\bar{\gamma})(e\phi_a^{(h)}-w_\perp^{(h)})^{1/2}w_\perp^{(h)}\left[\frac{w^{(h)}}{w_\perp^{(h)}}-1\right]^{1/2},$$

where $\bar{\gamma}=\int_{x_t}^{d_{ox}/2} dx\{2m_h[e\phi_a^{(h)}(x)-w_\perp^{(h)}]\}^{1/2}/\hbar$, x_t is the tunneling distance $[e\phi_a^{(h)}(x_t)=w_\perp^{(h)}]$, and $e\phi_a^{(h)}(x)$ is the potential barrier for holes. If the holes are generated with energy larger than the barrier $e\phi_a^{(h)}$, it will be assumed that $\bar{\gamma}=0$. The hole mass m_h appearing in $\bar{\gamma}$ is assumed to be $0.5m$, i.e., the electron effective mass in the band gap of the SiO_2.

The tunneling rate given above must be compared with the rate at which hot holes lose their energy via electron-electron and electron-phonon interactions. This rate, τ_r^{-1}, depends strongly on the band structure of the anode and the energy of the holes. If no strong interband transitions occur with threshold energies lower than the hole energy, only the electron-phonon interaction has to be considered. This is the case for Si, since the surface-plasmon-generated holes will have energy lower than the threshold for impact ionization [$\simeq$5 eV (Ref. 40)]. It will be assumed that the hole lifetime is controlled by the electron-phonon interaction also in Al, because of the strength of this mechanism and the fact that only one interband (vertical) transition occurs at 1.5 eV (Ref. 41). Then, the probability that a hole of energy $w^{(h)}$ will tunnel into the SiO_2 before suffering one scattering event is $[1+\tau(w^{(h)})/\tau_{e\text{-ph}}]^{-1}$, $1/\tau_{e\text{-ph}}$ being the electron-phonon scattering rate.

B. Quantum efficiency for hole injection

Assuming again that the plasmon emission is the rate-limiting step of the entire process, the overall quantum efficiency η_{hole} for one injected electron to produce the tunneling of one hole into the SiO_2 can be obtained by coupling the results of this discussion to Eq. (19):

$$\eta_{\text{hole}}(E_{ox})\simeq\frac{\sqrt{2}}{8}\frac{e^2m^{1/2}(e\phi_a)^{1/2}\omega_{SP}}{\pi\varepsilon_0(1+\kappa_{ox})} , \tag{22}$$

with

$$\Psi(E_{ox})=\left[\int_{\hbar\omega_{LO}+e\phi_a}^{\infty} dw_\perp \exp\left[\frac{-\Sigma(w_\perp-e\phi_a)}{w_\perp(w_\perp-e\phi_a)}\right]\int_0^{w_\perp}\frac{d\varepsilon}{\pi}\frac{\Gamma}{(\hbar\omega_{SP}-\varepsilon)^2+\Gamma^2}\left[1+\frac{\tau(w^{(h)})}{\tau_{e\text{-ph}}}\right]^{-1}\right]$$

$$\times\left[\hbar\omega_{LO}+\int_{\hbar\omega_{LO}+e\phi_a}^{\infty} dw\exp[-\Sigma(w-e\phi_a)]\right]^{-1} . \tag{23}$$

The second integral in Eq. (23) represents the sum over all possible processes involving the emission of an intermediate surface plasmon (real and virtual) of energy ε, and the hole energy $w^{(h)}$ is $\varepsilon/2$ (Al) or $(\varepsilon-E_{gap})/2$ (Si).

C. Generation of positive charge

The overall efficiency for the generation of positive charge can be obtained by introducing an empirical macroscopic capture cross section σ_h for the hole trapping at the Si-SiO_2 interface. It is believed that stressed Si—O and Si—Si bonds in the interfacial region can capture holes, giving rise to positive charge and fast states.[9,42] Usually, during an electron-injection experiment, the areal concentration N^+ of positive charge is observed to increase with the electron fluence f according to a first-order kinetics of the form[14,15]

$$N^+(f)=N_\infty^+[1-\exp(\sigma^+f)] . \tag{24}$$

The generation rate is given by the generation cross section σ^+, which, in the present approach, is given by

$$\sigma^+=\eta_{\text{hole}}(E_{ox})\sigma_h . \tag{25}$$

The saturated density N_∞^+ (i.e., the maximum number of interface defects which can be generated) is strongly dependent on the structure of the oxide and on its preparation conditions, but it should be independent of the surface-plasmon emission mechanism. Therefore, this issue is outside the scope of the present work.

Table I lists the values of the parameters entering Eqs. (22) and (25) which have been employed to evaluate numerically η_{hole} and σ^+. Before comparing these estimates with the experimental data, it should be stressed that errors might affect these theoretical predictions in many respects, as is obvious from the long sequence of processes which have been considered. First, it has been already noticed that the nature of the electron transport in SiO_2 is a source of major concern. Second, some uncertainty is present in the numerical values listed in Table I. Possible errors could arise in choosing the value of the hole tunneling barrier (possibly lowered by quantum image-force effects), the anode surface plasma lifetime and frequency (possibly lowered by the polycrystalline nature of the material, particularly with poly-Si gates), and the hole lifetime in the anode, which might be affected by the electron-electron interaction. Finally, the approximations employed to obtain Eq. (22) are probably satisfied for electric fields between 5 and 10—12 MV/cm: At low fields, particularly in thin oxides, the condition $eV_G \gg w_1^{(f)}$ is not satisfied, and the direct process of Fig. 1(a) may be significant. More important, the procedure employed to obtain the electron distribution given by Eq. (14) underestimates the fraction of electrons able to emit surface plasmons at low fields, as can be seen from experimental data of Ref. 35. At very high fields (i.e., above 1.2×10^7 V/cm), the electrons may reach the anode with an energy much larger than the surface-plasmon energy, and the condition $w_1^{(a)} - E_{c,a}(0) \ll e\phi_a$ [employed to simplify the integration leading to Eq. (19)] may not be satisfied.

V. COMPARISON WITH EXPERIMENTS

The most direct experimental support to the model presented here could come from the detection of both the excitation of surface plasma modes at the SiO_2-anode interface and of a hole current in the oxide during electron injection. Attempts to detect the surface plasma modes can rely only on indirect observations, such as electron-energy loss or the radiative decay of the plasma modes.[35] Measurements of this kind have been performed in the past and actually show that the emission of surface plasmons is a strong energy-loss mechanism, particularly in thin metal films.[35,43]

A. Hole current in SiO_2

As far as the existence of a hole current during electron injection is concerned, it has been already observed that,

TABLE I. Parameters employed in the numerical computation.

Parameter	Symbol	Numerical value
Electron effective mass in the SiO_2 band gap	m_{ox}	$0.5m_{\text{free}}$[a]
Electron effective mass in the SiO_2 conduction band (polaron corrected)	m_{ox}^*	$1.3m_{\text{free}}$[b]
Surface-plasmon energy (eV)	$\hbar\omega_{SP}$	9.46 (Si)[c] 8.51 (Al)[c]
Surface-plasmon full linewidth (eV)	2Γ	4.1 (Si)[d] 0.8 (Al)[e]
Electron barrier height (eV)	$e\phi_{a,c}$	3.2 (Si)[f] 3.15 (Al)[f]
Hole barrier height (eV)	$e\phi_a^h$	4.7 (Si)[g] 5.85 (Al)[g]
Hole lifetime (e-e or e-ph collision time) (sec)	$\tau_r = \tau_{e\text{-ph}}$	4.5×10^{-14} (Si)[h] 7.5×10^{-15} (Al)[i]

[a]Z. A. Weinberg, J. Appl. Phys. **53**, 5052 (1981).
[b]R. C. Hughes, Phys. Rev. Lett. **30**, 1333 (1973).
[c]D. Pines, Rev. Mod. Phys. **28**, 184 (1956).
[d]H. R. Philipp and H. Ehrenreich, Phys. Rev. **129**, 1550 (1963).
[e]H. Ehrenreich, H. R. Philipp, and B. Segall, Phys. Rev. **132**, 1918 (1963).
[f]S. M. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981).
[g]From the electron barriers and assuming the band gap of SiO_2 to be 9 eV.
[h]From the hole mobility (f).
[i]From dc conductivity (e) and assuming a temperature dependence T^{-1} for $T \lesssim T_{\text{Debye}} \simeq 430$ K.

during positive-bias electron injection in n-channel field-effect transistors (FET's), the gate current is due to two distinct components: the usual tunneling electron current (observed as channel current) and an additional current, observed as substrate current, whose origin is uncertain.[44,45] Recent investigations by Weinberg and Fischetti[18] could not prove unquestionably that this substrate current is due to holes transported in the insulator. Nevertheless, they seem to rule out the possibility that this current is due to electrons tunneling from the silicon valence band, because of the very high ratio observed between the substrate and the channel currents. This ratio, typically ranging from 10^{-3} at 9 MV/cm to 10^{-2} at 12 MV/cm (although in some samples even higher values were observed), has the field dependence and order of magnitude predicted by the present model for the hole current in SiO_2. This can be seen in Fig. 4, where the quantum efficiency $\eta_{\text{hole}}(E_{\text{ox}})$, calculated from Eq. (22), is plotted against the experimental values of the ratio $I_{\text{sub}}/I_{\text{chan}}$ measured by Eitan and Kolodny[44] and Weinberg and Fischetti.[18] Considering that the theoretical curve is calculated without any adjustable parameters, and noticing the scatter of the experimental data, the agreement appears to be satisfactory. However, it should be stressed again that no conclusive proof has been given that the substrate current is (totally or partially) due to holes coming from the oxide.

B. Generation rate of the positive charge

A more interesting and unambiguous result can be obtained by calculating the generation cross section σ^+ of the positive charge, as given by Eq. (25), and comparing it with the available experimental data. In this comparison the anode field must be used, since both the electron-energy distribution at the anode as well as the rate of hole injection are controlled by this field. Some uncertainty exists in the choice of the value for the hole capture cross section σ_h. Values within 1 order of magnitude (from 10^{-14} to 10^{-13} cm^2) have been reported in the literature. It is also known that σ_h depends on device-processing variables.[40] Figure 5(a) shows the results obtained by evaluating Eq. (25), having chosen $\sigma_h = 6\times10^{-14}$ cm^2 in order to fit the room-temperature data for the case of injection from Si into SiO_2 with Al as the anode.[15] This particular value of σ_h is in agreement with the experimental value reported by Tzou *et al.*[42] (ignoring the field

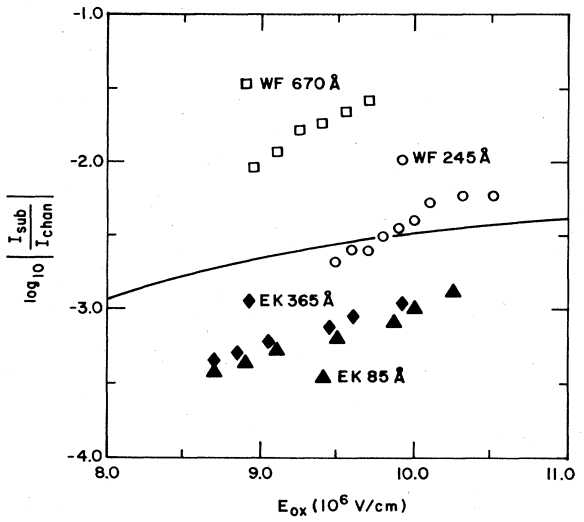

FIG. 4. Quantum efficiency for hole injection from the anode during a high-field electron injection calculated from Eq. (22) of the text (solid line). The experimental points represent the ratio $I_{\text{sub}}/I_{\text{chan}}$ measured during high-field injection at positive gate bias in n-channel MOSFET's measured by Eitan and Kolodny (Ref. 44) on samples with 85- and 365-Å-thick SiO_2, and by Weinberg and Fischetti (Ref. 18) on samples with 245- and 670-Å-thick insulators.

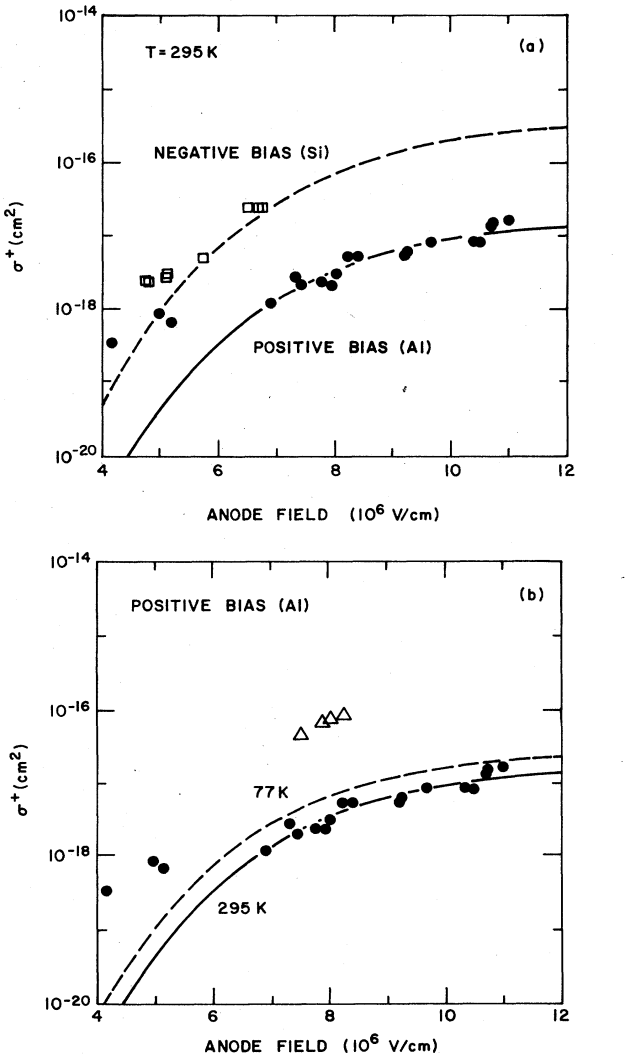

FIG. 5. Generation cross section for the interfacial positive charge calculated from Eq. (25) of the text with $\sigma_h = 6\times10^{-14}$ cm^2. The structure is a Si-SiO_2 capacitor. In (a) the theoretical calculation and experimental data are compared for the case of injection to Al (solid line and circles) and Si (dashed line and squares) at 295 K. In (b) the comparison is done for the case of injection to Al at 295 K (solid line and circles) and 77 K (dashed line and triangles).

dependence observed at low fields). It can be seen that, except for the expected disagreement at low electric fields, Eq. (25) provides the correct behavior for the rate of generation of the interfacial positive charge in Al-SiO_2-Si structures under injection at either polarity. Aluminum appears to be less efficient than silicon because of both a shorter hole lifetime and a higher barrier for holes. In Fig. 5(b) the temperature dependence is shown. The theory predicts a higher efficiency at 77 K than at 295 K because of the longer electron mean free path in SiO_2 (significant for the 0.063-eV LO-phonon scattering) and, again, because of the longer hole lifetime in the anode. As shown in Fig. 5(b), the temperature dependence observed experimentally in stronger than the theoretical calculation. Again, the assumptions made to obtain Eq. (15) may be responsible for this disagreement. The effect of the temperature is much stronger when the effect of the acoustic phonons (having energies not larger than 30 meV) is included in the theoretical evaluation of the electron transport in SiO_2.

C. Threshold in gate voltage for the generation of positive charge

Apart from these encouraging agreements, two additional experimental observations confirm the existence of two of the steps involved in the long sequence of processes considered here.[13] Figure 6 shows the generation efficiency of the positive charge as a function of the total voltage applied to the insulator V_{ox}. To avoid complications due to the field dependence of σ^+, the generation efficiency measured on a sample of a given thickness has been normalized to the efficiency measured on a much thicker sample at the same field. This procedure emphasizes the dependence on V_{ox} itself. A threshold for the generation of the positive charge can be seen to occur around 7—8 V, as V_{ox} is reduced by reducing the oxide thickness while keeping the field constant. Such a threshold is just what one would expect from the data listed in Table I for silicon: As the maximum energy the electrons gain drops below the energy required to generate a surface plasmon at the Si-SiO_2 interface, no more positive charge is observed. For a surface plasma energy of 9.5 eV with a half-width of 2 eV, such a threshold should be expected at 7.5 eV. The theoretical curves shown in Fig. 6 are obtained from Eq. (22) by assuming a steady-state energy distribution for the electrons arriving at the anode, but truncated at energy $eV_{ox}-e\phi_c+e\phi_a$ at two different oxide fields. This procedure amounts to assuming that the electron distribution has already reached steady state, even in the thinner samples. This is certainly true for thicknesses larger than 70 Å.[35,36]

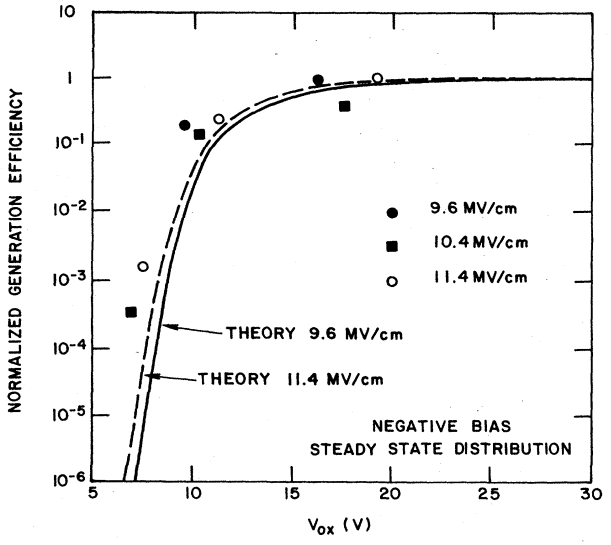

FIG. 6. Generation efficiency of the positive charge (normalized to its value for large oxide thickness) as a function of V_{ox} for samples of various oxide thicknesses (69, 99, 169, and 484 Å) during high-field electron injection in Si-SiO_2-Al structures. The cutoff at the Si-SiO_2 surface plasma energy is clearly visible in both the theoretical curves (calculated at two fields) and the experimental data (from Ref. 13).

D. Effect of gate metal on the generation of positive charge

Figure 7 shows the generation of positive charge during electron avalanche injection in MOS capacitors with different metal gates.[13] As the barrier for holes $e\phi_a^{(h)}$ is increased (from Au to Al), the amount of charge which is formed decreases, and eventually no charge is detected for $e\phi_a^{(h)} \geq 6.3$ eV, as in Mg-gate samples. Quantitative estimates of the generation cross section σ^+ for Mg indeed show an unmeasurably small rate ($\hbar\omega_{SP} \simeq 6.5$ eV at the Mg-SiO_2 interface.[21]). Calculations for Au are complicated by strong d-band effects in both the surface plasma dispersion relation ($\hbar\omega_{SP} \simeq 2.5$ eV with about 0.1 eV of linewidth[46]) as well as in the hole lifetime. Furthermore, non-plasmon-mediated processes are undoubtedly important, as remarked previously. Nevertheless, the strong dependence of the phenomenon on the hole barrier at the anode-SiO_2 interface indicates that hole injection from the anode takes place.

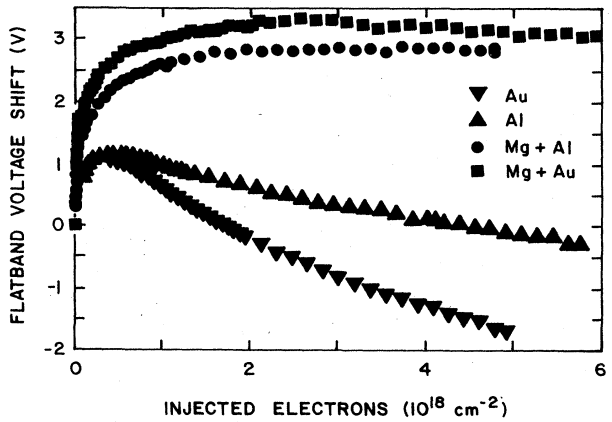

FIG. 7. The "turn-around" effect (proportional to the interfacial positive charge) observed during electron avalanche injection in samples having gate metals with different hole barriers at the anode-SiO_2 interface. Mg, having the highest barrier, shows no sign of positive charge. Au, having the lowest barrier, shows the largest effect.

VI. SUMMARY AND CONCLUSIONS

A model has been presented to explain the generation of positive charge at the Si-SiO_2 interface during high-field electron injection in MOS structures. According to this model, the injected electrons gain energy during their transport in the insulator. A significant fraction of the accelerated electrons lose energy at the anode-SiO_2 interface by emitting surface plasmons. Once excited, the surface plasmons decay quickly into electron-hole pairs. A small fraction of these holes can be injected into the SiO_2 (as previously suggested by Weinberg *et al.*[22] and Grunthaner *et al.*[23]) and can be trapped at the stressed Si-SiO_2 interface.[9] Finally, slow and fast interface states can be produced after the recombination of the trapped holes with the electrons flowing in the oxide, as proposed previously by Lai.[42]

The generation rate of the positive charge has been calculated and it has been found to be reasonably consistent with the available experimental data. Moreover, the dependence of the process of positive-charge generation on the oxide thickness (below 100 Å), on the field at the anode-SiO_2 interface, on the metal gate, and on temperature, as well as the existence of a substrate current during positive-bias electron injection in *n*-channel FET structures, are well explained by the model.

It should be noted that two major issues are insufficiently treated in the present work. First, the high-energy tail of the electron distribution in SiO_2 has been treated by considering only scattering with LO phonons, in the absence of better knowledge of the band structure of amorphous SiO_2. Second, a phenomenological approach has been taken in considering the mechanism of hole trapping at the Si-SiO_2 interface. This process is largely influenced by the structure of the interface, and varies with different oxidation and annealing processes. These processing effects are clearly neglected in the present "electronic" approach, but they can play a significant role in real devices. MOS structures with a polycrystalline-silicon gate probably constitute a good example of this point, since the higher temperatures obtained during the gate deposition and doping are known to cause a significant modification of the underlying SiO_2 layer.

As a final observation, we note that the generation of positive charge and destructive breakdown in SiO_2 have been often associated in the past.[11] For this reason, the model presented here could represent a good starting point for understanding the mechanisms leading to dielectric breakdown in silicon dioxide.

ACKNOWLEDGMENTS

The author is pleased to thank D. R. Young and Z. A. Weinberg for many interesting suggestions and discussions, D. J. DiMaria and S. D. Brorson for information regarding the electron heating in SiO_2, and T. N. Theis for useful criticisms concerning the role of surface plasma modes in MOS systems. Thanks are also due Z. A. Weinberg, J. R. Kirtley, and F. Stern for having critically read the manuscript.

APPENDIX: EXPRESSIONS FOR THE SINGLE-PARTICLE ELECTRON WAVE FUNCTIONS $\chi_{k_\perp}^{(a)}$ AND $\chi_{k_\perp}^{(c)}$

In the effective-mass approximation the electron wave functions are expanded in terms of Wannier functions whose envelope $\psi(\mathbf{r},t)$ obeys approximately the equation

$$\left[-\frac{\hbar^2}{2m_\perp^{(i)}}\frac{\partial^2}{\partial x^2}-\frac{\hbar^2}{2m_{||}^{(i)}}\nabla_{\mathbf{R}}^2+eV_{\text{eff}}(x)\right]\psi(x,\mathbf{R},t)=i\hbar\frac{\partial}{\partial t}\psi(x,\mathbf{R},t)\ , \tag{A1}$$

where $\mathbf{R}$ is the coordinate in the plane of the electrode-SiO_2 interfaces and x is the coordinate along the perpendicular direction. The masses $m_\perp^{(i)}$ and $m_{||}^{(i)}$ are the components of the effective-mass tensor along these directions in the ith material (anode, insulator, cathode). As stated in the text, a unique (isotropic) electron effective mass m will be considered.

The effective potential $V_{\text{eff}}(x)$ has the form

$$eV_{\text{eff}}(x)=eV_{\text{ext}}(x)+\begin{cases}E_{c,c}(0)=E_{F,c}-\zeta_c & \text{in the cathode}\ ,\\ E_{c,\text{SiO}_2}(0) & \text{in the oxide}\ ,\\ E_{c,a}(0)=E_{F,a}-\zeta_c & \text{in the anode}\ .\end{cases}$$

It has been assumed that the minima of the cathode and anode conduction bands $E_{c,c}(0)$ and $E_{c,a}(0)$ occur at $\mathbf{k}=\mathbf{0}$. $E_{F,c}$ and $E_{F,a}$ (ζ_c and ζ_a) are the Fermi levels of the electrodes measured from zero energy (from the cathode and anode conduction-band minima) and V_{ext} is the external potential applied. The zero of the energy will be taken at $E_{c,c}(0)$.

If a bias V_G is applied to the electrodes, calling $e\phi_a$ and $e\phi_c$ the anode-SiO_2 and cathode-SiO_2 barrier heights, and assuming $\zeta_a,\zeta_c \ll eV_G$ (thus ignoring the band structure of the electrodes), one can write

$$eV_{\text{eff}}(x)=\begin{cases}0 & (x\le -d_{\text{ox}}/2)\ ,\\ \dfrac{e}{2}(\phi_c+\phi_a-V_G)-\dfrac{ex}{d_{\text{ox}}}(V_G+\phi_c-\phi_a) & (-d_{\text{ox}}/2\le x\le d_{\text{ox}}/2)\ ,\\ -eV_G & (x\ge d_{\text{ox}}/2)\ .\end{cases}$$

Looking for solutions of the form

$$\psi(\mathbf{r},t)=\chi_{k_\perp}(x)\exp[i(\mathbf{R}\cdot\mathbf{k}_{||}-w_k t/\hbar)]\ ,$$

the functions χ can be shown to obey the equation

$$\left[\frac{d^2}{dx^2}-\frac{2m}{\hbar^2}[eV_{\rm eff}(x)-w_\perp]\right]\chi_{k_\perp}(x)=0\ , \tag{A2}$$

where $w_\perp=w_k-\hbar^2k_{||}^2/2m=\hbar^2k_\perp^2/2m$. The continuity of the solution with its derivative at the interfaces renders (A2) a self-adjoint Sturm-Liouville problem, for which one (for $-eV_G\le w_\perp\le 0$) or two (for $w_\perp>0$) independent solutions can be found in terms of the Bessel functions $J_{\pm1/3}$ or $I_{\pm1/3}$. For $w_\perp>0$ the first solution corresponds to one electron propagating from the anode and being partially transmitted and reflected:

$$\chi_{k_\perp}^{(a)}(x)=N^{(a)}\times\begin{cases}\exp(iM_cx) & (x\le -d_{\rm ox}/2)\ ,\\ \mu_2^{(a)}(k_\perp)y^{1/2}J_{1/3}(\frac{2}{3}y^{3/2})+\mu_3^{(a)}(k_\perp)y^{1/2}J_{-1/3}(\frac{2}{3}y^{3/2}) & (-d_{\rm ox}/2\le x\le d_{\rm ox}/2)\ ,\\ \mu_1^{(a)}(k_\perp)\exp(iM_ax)+\mu_1'^{(a)}(k_\perp)\exp(-iM_ax) & (x\ge d_{\rm ox}/2)\ .\end{cases}$$

Considering only the conduction band of the cathode, for $-eV_G\le w_\perp\le 0$ this is the only physically acceptable solution, representing an electron tunneling from the anode and exponentially decaying in the cathode. The second solution is nonvanishing only if $w_\perp>0$, and represents a symmetrical situation in which one electron tunnels from the cathode:

$$\chi_{k_\perp}^{(c)}(x)=N^{(c)}\times\begin{cases}\exp(iM_cx)+\mu_4^{(c)}(k_\perp)\exp(-iM_c) & (x\le -d_{\rm ox}/2)\ ,\\ \mu_2^{(c)}(k_\perp)y^{1/2}J_{1/3}(\frac{2}{3}y^{3/2})+\mu_3^{(c)}(k_\perp)y^{1/2}J_{-1/3}(\frac{2}{3}y^{3/2}) & (-d_{\rm ox}/2\le x\le d_{\rm ox}/2)\ ,\\ \mu_1^{(c)}(k_\perp)\exp(iM_ax) & (x\ge d_{\rm ox}/2)\ ,\end{cases}$$

where $M_c=k_\perp$, $M_a=[2m(w_\perp+eV_G)]^{1/2}/\hbar$, $E_{\rm ox}$ is the oxide field $(V_G+\phi_c-\phi_a)/d_{\rm ox}$, and the variable y is defined as

$$y(x)=\frac{(2emE_{\rm ox})^{1/3}}{\hbar^{2/3}}x+\frac{(2m)^{1/3}}{(eE_{\rm ox}\hbar)^{2/3}}\left[w_\perp-\frac{e}{2}(\phi_c+\phi_a-V_G)\right]\ .$$

It must be understood that if $y<0$, then

$$J_\alpha(\tfrac{2}{3}y^{3/2})=i^\alpha I_\alpha(-\tfrac{2}{3}|y|^{3/2})\ .$$

The factors μ are obtained from the continuity of χ and its derivative at the interfaces, while the factors N are normalization factors which must satisfy the condition

$$\int_{-\infty}^{\infty}dx\,\chi_{k_\perp}^{(a)}(x)\chi_{k_\perp'}^{(a)}(x)^*=\int_{-\infty}^{\infty}dx\chi_{k_\perp}^{(c)}(x)\chi_{k_\perp'}^{(c)}(x)^*=\delta(k_\perp-k_\perp')$$

allowed by the orthogonality guaranteed by the self-adjointness of Eq. (A2). However, notice that the system $\{\chi_{k_\perp}^{(a)},\chi_{k_\perp'}^{(c)}\}$ is not orthonormal, since $\chi_{k_\perp}^{(a)}$ and $\chi_{k_\perp'}^{(c)}$ are not orthogonal if $k_\perp=k_\perp'$. The functions ξ must be introduced to orthogonalize these basis functions also in each $k_\perp$ eigenspace in order to diagonalize the free-electron Hamiltonian.

In the text the WKB approximation has been extensively employed to minimize the amount of numerical work required by the exact solutions given above. Within this approximation, the wave functions can be written as follows:

$$\chi_{k_\perp}^{(a)}(x)\simeq g^{(a)}(k_\perp)^{-1/2}\times\begin{cases}\exp(iM_cx) & (x\le -d_{\rm ox}/2)\ ,\\ \dfrac{\mu_2^{(a)}(k_\perp)}{\kappa(x)^{1/2}}\exp\left[\displaystyle\int_{-d_{\rm ox}/2}^{x}dx'\,\kappa(x')\right] & (-d_{\rm ox}/2\le x\le d_{\rm ox}/2)\ ,\\ \mu_1^{(a)}(k_\perp)\exp(iM_ax)+\mu_1'^{(a)}(k_\perp)\exp(-iM_ax) & (x\ge d_{\rm ox}/2)\ ,\end{cases}$$

for the anodic solution, and

$$\chi_{k_\perp}^{(c)}(x)\simeq g^{(c)}(k_\perp)^{-1/2}\times\begin{cases}\exp(iM_cx)+\mu_4^{(c)}(k_\perp)\exp(-iM_cx) & (x\le -d_{\rm ox}/2)\ ,\\ \dfrac{\mu_3^{(c)}(k_\perp)}{\kappa(x)^{1/2}}\exp\left[-\displaystyle\int_{-d_{\rm ox}/2}^{x}dx'\,\kappa(x')\right] & (-d_{\rm ox}/2\le x\le d_{\rm ox}/2)\ ,\\ \mu_1^{(c)}(k_\perp)\exp(iM_ax) & (x\ge d_{\rm ox}/2)\ ,\end{cases}$$

for the solution representing an electron tunneling from the cathode. In the above expression, $\kappa(x)$ (real or imaginary) is defined as $\kappa(x)=\{2m[eV_{\rm eff}(x)-w_\perp]\}^{1/2}/\hbar$. Defining $\kappa_\pm=\kappa(\pm d_{\rm ox}/2)$ and $\gamma=\int_{-d_{\rm ox}/2}^{x_t}dx\,\kappa(x)$, where x_t is the tunneling

distance [i.e., $eV_{\rm eff}(x_t)=w_\perp$], the coefficients μ are given by

$$\mu_1^{(a)}(k_\perp)=\frac{\exp[\gamma-i(M_a+M_c)d_{\rm ox}/2]iM_c}{4(\kappa_+\kappa_-)^{1/2}}(1-\kappa_+/M_a)(1-i\kappa_-/M_c)\ ,$$

$$\mu_1'^{(a)}(k_\perp)=\frac{\exp[\gamma+i(M_a+M_c)d_{\rm ox}/2]iM_c}{4(\kappa_+\kappa_-)^{1/2}}(1-i\kappa_-/M_c)(1+i\kappa_+/M_a)\ ,$$

$$\mu_2^{(a)}(k_\perp)=\frac{\exp(-iM_cd_{\rm ox}/2)iM_c}{2\kappa_+^{1/2}}(1-i\kappa_-/M_c)\ ,$$

and, for $w_\perp>0$, by

$$\mu_1^{(c)}(k_\perp)=\frac{4i\exp[-\gamma-i(M_c+M_a)d_{\rm ox}/2](\kappa_+\kappa_-)^{1/2}}{M_c(1+i\kappa_+/M_a)(1+i\kappa_-/M_c)}\ ,$$

$$\mu_3^{(c)}(k_\perp)=\frac{2\exp(-iM_cd_{\rm ox}/2)\kappa_-^{1/2}}{1+i\kappa_-/M_c}\ ,$$

$$\mu_4^{(c)}(k_\perp)=\exp(-iM_cd_{\rm ox})(1-i\kappa_-/M_c)(1+i\kappa_-/M_c)^{-1}\ .$$

In the limit $\gamma\gg1$, the normalization factors are

$$g^{(a)}(k_\perp)\simeq\pi[\mu_1^{(a)}(k_\perp)\mu_1^{(a)}(k_\perp)^*+\mu_1^{(a)}(k_\perp)\mu_1'^{(a)}(k_\perp)^*]\ ,\quad g^{(c)}(k_\perp)\simeq\pi[1+\mu_4^{(c)}(k_\perp)\mu_4^{(c)}(k_\perp)^*]\ .$$

In this limit, $\xi^{(a)}\to\chi^{(c)}$ and $\xi^{(c)}\to\chi^{(a)}$, so that the two bases chosen in the expansions given by Eq. (6) of the text coincide. This is the usual approximation done within the "effective tunneling Hamiltonian" formalism.[28]

[1]K. R. Hofmann and G. Dorda, in *Proceedings of the Conference on Insulating Films on Semiconductors,* edited by M. Schultz (Springer, Berlin, 1981), p. 122.

[2]Y Nissan-Cohen, J. Shappir, and D. Frohman-Bentchkowsky, J Appl. Phys. **54**, 5793 (1983).

[3]M. Knoll, D. Braunig, and W. R. Fahrner, J. Appl. Phys. **53**, 6946 (1982).

[4]J. Maserjian and N. Zamani, J. Appl. Phys. **53**, 559 (1982).

[5]M. V. Fischetti, R. Gastaldi, F. Maggioni, and A. Modelli, J. Appl. Phys. **53**, 3129 (1982); **53**, 3136 (1982).

[6]D. R. Young, E. A. Irene, D. J. DiMaria, R. F. DeKeersmeacker, and H. Z. Massoud, J. Appl. Phys. **50**, 6466 (1979).

[7]S. K. Lai and D. R. Young, J. Appl. Phys. **52**, 6321 (1981).

[8]C. T. Sah, J. J. Sun, and J. J. Tzou, J. Appl. Phys. **54**, 944 (1983); **54**, 2547 (1983).

[9]F. J. Grunthaner, P. J. Grunthaner, and J. Maserjian, IEEE Trans. Nucl. Sci. **NS-29**, 1462 (1982).

[10]P. H. Lenahan and P. V. Dressendorfer, J. Appl. Phys. **54**, 1457 (1983).

[11]T. H. DiStefano and M. Shatzkes, Appl. Phys. Lett. **25**, 685 (1974).

[12]P. Solomon and N. Klein, Solid State Commun. **17**, 1937 (1975).

[13]M. V. Fischetti, Z. A. Weinberg, and J. Calise, J. Appl. Phys. **57**, 418 (1985).

[14]F. J. Feigl, D. R. Young, D. J. DiMaria, S. K. Lai, and J. A. Calise, J. Appl. Phys. **52**, 5665 (1981).

[15]M. V. Fischetti, J. Appl. Phys. **56**, 575 (1984).

[16]R. Gale, F. J. Feigl, C. W. Magee, and D. R. Young, J. Appl. Phys. **54**, 6938 (1983).

[17]M. V. Fischetti and B. Ricco, J. Appl. Phys. (to be published).

[18]Z. A. Weinberg and M. V. Fischetti, J. Appl. Phys. **57**, 443 (1985).

[19]Z. A. Weinberg and G. Rubloff, Appl. Phys. Lett. **32**, 184 (1978).

[20]T. N. Theis, D. J. DiMaria, J. R. Kirtley, and D. W. Dong, Phys. Rev. Lett. **52**, 1445 (1984).

[21]D. Pines, Rev. Mod. Phys. **28**, 184 (1956).

[22]Z. A. Weinberg, Appl. Phys. Lett. **27**, 437 (1975); Z. A. Weinberg, W. C. Johnson, and M. A. Lampert, J. Appl. Phys. **47**, 248 (1976).

[23]F. J. Grunthaner, B. F. Lewis, N. Zamani, J. Maserjian, and A. Madhukar, IEEE Trans. Nucl. Sci. **NS-27**, 164 (1980).

[24]P. Nozières and D. Pines, Phys. Rev. **109**, 741 (1958), and references quoted therein.

[25]K. L. Ngai and E. N. Economou, Phys. Rev. B **4**, 2132 (1971).

[26]L. C. Davis, Phys. Rev. B **16**, 2482 (1977).

[27]A. J. Bennett, C. B. Duke, and S. D. Silverstein, Phys. Rev. **176**, 969 (1968).

[28]J. Bardeen, Phys. Rev. Lett. **6**, 57 (1961).

[29]*Handbook of Chemistry and Physics,* 61st ed. (Chemical Rubber Co., Cleveland, Ohio, 1980–81), p. E-392.

[30]R. H. Ritchie, Phys. Rev. **106**, 874 (1957); Surf. Sci. **34**, 1 (1973).

[31]G. Krieger and R. M. Swanson, J. Appl. Phys. **52**, 5710 (1981).

[32]D. K. Ferry, Solid State Commun. **18**, 1051 (1976).

[33]W. T. Lynch, J. Appl. Phys. **43**, 3274 (1972).

[34]H. J. Fitting and J. U. Freiman, Phys. Status Solidi A **69**, 349 (1982).

[35]The vacuum emission results are discussed by D. J. DiMaria, T. N. Theis, J. R. Kirtley, F. L. Pesavento, D. W. Dong, and S. D. Brorson [J. Appl. Phys. **57**, 1214 (1985)]. More evidence for the "anomalous" heating of electrons of SiO_2 can be found in T. N. Theis, J. R. Kirtley, D. J. DiMaria, and D. W. Dong, in *Insulating Films on Semiconductors,* edited by J. F. Verwey and D. R. Wolters (North-Holland, Amsterdam, 1983), p. 134.

[36]Massimo V. Fischetti, Phys. Rev. Lett. **53**, 1755 (1984).

[37]W. P. Dumke, Phys. Rev. **167**, 783 (1968).

[38]R. C. Hughes, Solid-State Electron. **21**, 251 (1978); Phys. Rev. Lett. **30**, 1333 (1973).

[39]Note that the surface-plasmon decay may or may not conserve momentum, so that the assumption that the energy $(\hbar\omega_{SP}-E_{gap})/2$ is transferred to the hole is not obvious, even for a direct-gap semiconductor at $\mathbf{k}=\mathbf{0}$. For a metal, the detailed band structure should be considered. Momentum conservation during the decay of bulk plasmons in semiconductors has been assumed by A. Rothwarf [J. Appl. Phys. **44**, 752 (1973)], and it has been discussed by R. C. Alig, S. Bloom, and C. W. Struck [Phys. Rev. B **22**, 5565 (1980)].

[40]See, for instance, S. M. Sze, *Physics of Semiconductor Devices (Wiley, New York, 1981)*.

[41]H. Ehrenreich, H. R. Philipp, and B. Segall, Phys. Rev. **132**, 1918 (1963).

[42]The processing dependence of the hole capture cross section has been studied by A. R. Stivers and C. T. Sah [J. Appl. Phys. **51**, 6292 (1980)]. More recent results about its field dependence can be found in J. J. Tzou, J. Y. Sun, and C. T. Sah [Appl. Phys. Lett. **43**, 861 (1983)]. It should be noted that holes can be trapped at the Si-SiO_2 interface with or without the creation of the permanent damage consisting of fast and slow interface states, also called "anomalous positive charge." During electron-injection experiments, the large electron concentration may favor the transformation of most of the trapped holes into "anomalous positive charge," following the electron-hole recombination process. This idea has been discussed by S. K. Lai [J. Appl. Phys. **54**, 2540 (1983)].

[43]For a comphrehensive, although old, review of experimental data on electron-energy loss in solids, see L. Marton, Rev. Mod. Phys. **28**, 172 (1956). The radiative decay of surface plasmons during electron injection in metal-insulator-metal and metal-insulator-semiconductor structures has been recently observed by J. Kirtley, T. N. Theis, and J. C. Tsang [Phys. Rev. B **24**, 5650 (1981)] and J. R. Kirtley, T. N. Theis, J. C. Tsang, and D. J. DiMaria [Phys. Rev. B **27**, 4601 (1983)]. There, the observed luminescence was attributed to the decay of surface plasmons localized at the outer metal-vacuum interface (fast modes) whose radiative efficiency was shown to dominate over the efficiency of the (slow) junction modes.

[44]B. Eitan and A. Kolodny, Appl. Phys. Lett. **43**, 106 (1983).

[45]C. Chang, M. S. Liang, C. Hu, and R. W. Brodersen, *Carrier-Tunneling Related Phenomena in Thin Oxide MOSFET's* (IEDM Technical Digest, Washington D.C., 1983), p. 194.

[46]P. B. Johnson and R. W. Christy, Phys. Rev. B **6**, 4370 (1972).