Quantum two-dimensional calculation of time constants of random telegraph signals in metal-oxide–semiconductor structures

A. Palma,* A. Godoy, J. A. Jiménez-Tejada, J. E. Carceller, and J. A. López-Villanueva
Departamento de Electrónica y Tecnología de Computadores, Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain
(Received 21 April 1997)

The thermal and gate-voltage dependencies for the capture and emission times of random telegraph signals have been theoretically analyzed in a Si-SiO_2 interface. A quasi-two-dimensional treatment of the interaction between a neutral near-interface oxide trap and an electron in the subband of the inversion layer has been developed to obtain expressions for the capture and emission times where the influence of the trap parameters (energy depth, distance to the interface, and electron-phonon coupling factor) is clearly shown. This analysis combines multiphonon-emission theory, tunnel transition probability and the electrostatic Coulomb barrier effect, allowing us to reproduce experimental data for traps in different devices, temperatures, and bias conditions. As a result, trap distances to the interface, trap energy levels, and electron-phonon couplings have been calculated. The character of single electron transitions in this process let us show that the ground and first excited subbands, with similar capture and emission times, are the most important contributors to the phenomenon. [S0163-1829(97)07039-2]

I. INTRODUCTION

Random telegraph signals (RTS's) generated through the fluctuating occupancy of individual traps have been extensively studied in small area metal-oxide–semiconductor transistors (MOST's) for a great variety of gate bias and temperature conditions.[1–5] The importance of explaining this phenomenon is twofold. First, it provides a unique probe into the trapping dynamics of single defects. Second, $1/f$ noise in MOST's is supposed to arise from the RTS superposition.[1–3,6] In spite of being amply studied, the two main aspects of RTS analysis are not yet fully understood: its amplitude and the capture and emission times of the signal. In this paper, we have theoretically analyzed the second aspect: the capture and emission times and their dependencies on gate bias, temperature, and trap properties for a neutral oxide trap.

Experimental results about the thermal behavior of these times[1,6–9] clearly show that the trapping and release of a conduction electron at a near-interface oxide trap cannot be treated as a simple elastic tunnel process. This conclusion has also been suggested by capacitance-voltage data in metal-oxide–semiconductor (MOS) structures where evidence of trap relaxation processes has been observed.[10] In fact, the thermally activated behavior of the capture and emission times has been attributed to the multiphonon emission process.[6,9]

In previous reports, the main results include the use of an asymptotic expression of the multiphonon capture cross section, $\sigma_c=\sigma_0\exp(-E_B/kT)$, which is only valid at high temperatures.[11,12] In this expression, E_B is the capture activation energy (see Fig. 1), k the Boltzmann constant, and T the absolute temperature. σ_0 includes information about the tunnel interaction between the channel and the trap and gate-voltage dependence of the transition. In another report, at the limit of zero temperature, the multiphonon emission model has been applied to this process, showing the high degree of trap coupling near the Si-SiO_2 interface.[13] However, although advances made in this field have been quite significant, until now no clear relationship between the macroscopic measured time constants and the microscopic magnitudes associated to this transition has been established. In our opinion, a complete calculation of the transition times based on the relationship between trap properties (energy level, electron-phonon coupling factor, and distance to the interface) and environment conditions (temperature, gate bias, and bulk doping) is lacking. In that respect, a method taking quantum principles into account would be the most appropriate.

We therefore present a detailed quantum calculation of the transition probabilities between a neutral trap level inside the gate oxide and a channel state in the multiphonon emis-

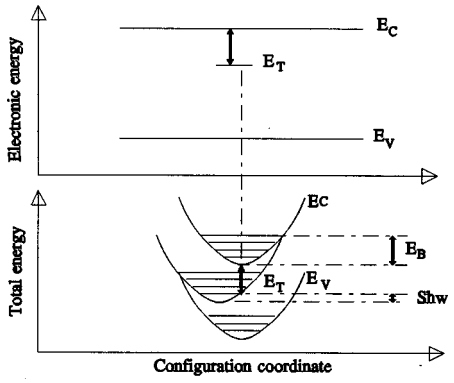

FIG. 1. Diagrams of the energy perturbation produced by a trap in the band diagram of a semiconductor. The upper diagram plots the usual vision of a trap with a single level E_T inside the semiconductor band gap. This represents only the electronic energy of the perturbation. The bottom diagram is the complete energy representation of the perturbation (electronic+vibrational energy) as a function of the configuration coordinate. E_B and $S\hbar\omega$ are the capture activation energy and semi-Franck-Condon energy shift, respectively.

sion approach. The contribution of the Coulomb energy[14,15] has also been included. This calculation has been made by assuming wave-function models for the trap and channel states in order to calculate the transition matrix element similarly as for trapping in bulk. However, given the two-dimensional behavior of electrons in the inversion layer,[16] capture has been taken into account from subbands in the semiconductor conduction band. With a numerical simulation of the MOS structure, the envelope wave function and energy minimum for each subband, Coulomb energy, and surface potential are available for different gate voltages, temperatures, and technological parameters (i.e., oxide thickness and bulk profile). All this information will be used to explain the capture and emission times observed in RTS's. We will clearly show that thermal behavior is mainly caused by the multiphonon process, and gate-voltage dependencies of these time constants are perfectly reproduced considering the microscopic parameters associated to the system. Therefore, both a quasi-two-dimensional treatment of the process and a complete calculation of the capture and emission times are the main contributions of this paper. In addition, comparing these results with experimental data provides information about the traps. Although the result obtained in this work is general, we focus on gate dependence and therefore, will assume very low drain voltages. A uniform channel has also been considered throughout this work.

The paper is organized as follows. The multiphonon theory basis for the calculation of the capture rate is in Sec. II. In Sec. III the transition matrix element is calculated according to the initial and final states, pointing out the differences with the traditional bulk capture calculation. In Sec. IV the Coulomb energy calculation is indicated from a quasi-two-dimensional point of view. Once the capture rate from one subband is obtained, the total capture and emission rates are calculated, the latter according to the balance principle. The next step, in Sec. VI, resides in comparing and discussing our results with experimental data, showing a very good fitting. Finally, the main conclusions are detailed in Sec. VII.

II. APPLICATION OF MULTIPHONON-EMISSION MODEL

In the capture and emission processes between a near-interface oxide trap and the inversion layer, we assume the following mechanisms: a tunnel transition assisted by a multiphonon emission or absorption given by the energy difference between the initial and final states, and the electrostatic Coulomb barrier effect, all resulting in an inelastic tunneling transition.

Therefore, total capture probability W_c can be expressed by the product of the multiphonon rate $W_{\rm mp}$ and the Coulomb barrier effect,

$$W_c = W_{\rm mp} e^{-\Delta E/kT}, \tag{1}$$

where ΔE is the Coulomb energy.

The first step is to apply the multiphonon emission model to the electron transition between a trap site and a channel state, in order to calculate $W_{\rm mp}$. Basically, this type of capture is assisted by the simultaneous emission of several phonons, in contrast with the single-phonon intraband scattering. In Fig. 1 the energy representation of an impurity level in the band gap of a semiconductor is shown. In the upper part of the figure the usual electronic energy diagram is represented, $E_C - E_T$ being the energy depth measured by emission transients. Moreover, including the vibrational energy caused by the electron-lattice interaction the total-energy representation is also plotted (bottom of Fig. 1). In this figure both the capture activation energy E_B and the relaxation energy (S phonons of energy $\hbar\omega$) involved in capture process[12] are clearly shown. Thus, more energy than E_T is involved in this phenomenon, and this fact plays an important role in the thermal dependence of the capture process according to the multiphonon model.

The main quantum-mechanic multiphonon-rate expressions[11,12,17,18] have the same thermal dependence (each within its valid temperature range) and almost the same quantitative results. In addition, it has been shown that adiabatic and static approaches for the problem give the same final result.[19] We have taken the approximate multiphonon probability calculated by Zheng, Tan, and Ng.[18] That expression, until now used only for the capture process by bulk traps, has been extended to this situation in which the interaction between a trap state and electrons in different subbands of the inversion layer is being analyzed. That is why, we have separated the electron energy into its potential and kinetic (parallel to the interface) components to calculate the number of phonons involved in the transition. Multiphonon probabilities for each subband have also been considered for consistency with the single-electron character of the transition. Therefore, in the single-energy phonon approximation, the nonradiative multiphonon transition rate between a state in the ith subband and a trap state $W^i_{\rm mp}$ can be approximately calculated from

$$W^i_{\rm mp} \simeq \frac{\pi}{\hbar} S|V|^2 \left(1 - \frac{p^i}{S}\right)^2 G(E^i_T),$$

$$G(E^i_T) \simeq \frac{1}{(2\pi)^{1/2}\hbar\omega}\, \chi^{-1/4} \left(\frac{\xi}{p^i + \chi^{1/2}}\right)^{p^i} \times \exp\left(\chi^{1/2} - (2n+1)S + \frac{p^i \hbar\omega}{2kT}\right),$$

$$p^i = \frac{E^i_T + E_\parallel}{\hbar\omega}, \quad n = \frac{1}{e^{\hbar\omega/kT} - 1}, \quad \xi = 2S[n(n+1)]^{1/2},$$

$$\chi = (p^i)^2 + \xi^2, \tag{2}$$

where p^i is the number of emitted phonons of energy $\hbar\omega$ in the transition from the ith subband, n is the phonon occupation factor, $E^i_T = E_i - E_T$ is the energy difference between the ith subband minimum E_i and the trap level E_T. Here we have only added the parallel energy corresponding to the thermal energy of a two-dimensional gas, $E_\parallel = kT$, without including any contribution caused by the longitudinal electric field. $|V|^2$ and S are the transition matrix element and the Huang-Rhys factor which characterize the interaction,

$$S \simeq \frac{1}{2(\hbar\omega)^2} |\langle b|U|b\rangle|^2,$$

$$|V|^2 = |\langle f|U|b\rangle|^2, \tag{3}$$

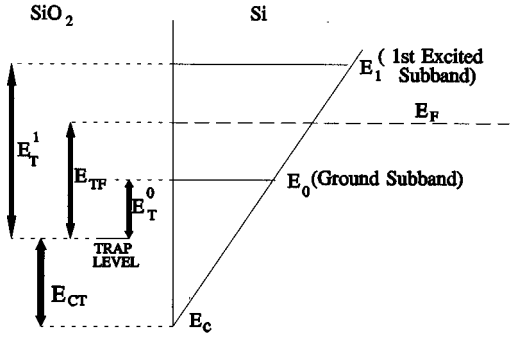

FIG. 2. Schematic representation of the energy levels involved in the interaction between a near-interface oxide trap and the inversion-layer electrons in the Si-SiO_2 interface. E_0 and E_1 are the ground and first excited subband minima, respectively, E_F is the Fermi level, E_c the conduction-band minimum at the interface. The trap level in the oxide is also shown.

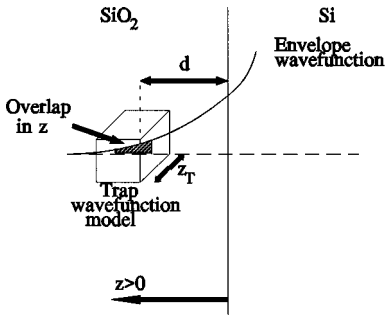

FIG. 3. Schematic representation of the trap solid-cube wave function inside the oxide, showing the distance to the interface, d, and the side of the cube, z_T. The overlap between trap wave function and envelope function is also shown as a shaded area.

where b and f refer to the bound (trap) and free (channel) states, respectively, and U is the potential for the electron-phonon coupling.

Equation (2) has to be evaluated for each subband; however, as we will show later, in most practical cases only the ground and the first excited subbands contribute significantly to the total capture rate. Only the energy E_T^i and the free-state wave function are different in each subband, so the number of emitted phonons p^i and the transition matrix element $|V|^2$ will both vary. In Fig. 2 a schematic representation of the band diagram near the Si-SiO_2 interface is plotted showing the important magnitudes for this study: subband minima E_0 and E_1, Fermi level E_F, and the trap level in the oxide E_T. The differences of the energies defined above are also shown.

Once the general equations for the calculation of the capture rate have been established, models to approximate the bound and free wave functions are required as well as the potential for the electron-phonon coupling. As we are considering a neutral trap before capturing an electron from the inversion layer, the trap model will represent a neutral bound state.

III. CAPTURE RATE MATRIX ELEMENTS AND PARAMETERS

Let x, y, and z be the rectangular coordinates, where z is the direction normal to the Si-SiO_2 interface. For our calculations, we have taken $z=0$ just at the interface and positive values for z in the oxide, as indicated in Fig. 3. Let us also consider d as the trap distance to the interface. Thus, z is the direction where the overlapping between the bound and free states is important. We have assumed the following models for the implied states.

A. Neutral trap wave-function model

Traditionally, two models are combined to approximate a trap wave function. One is defined by a solid sphere or box that very simply simulates the trap core effects. This model is considered valid for very short distances to the trap.[20] The other model, called the δ-function model,[21] is an approximation for distances far from the trap site. The latter model has usually been used to represent the trap wave function in bulk capture since, when trap atoms are far enough away from each other, most of the electrons are far from the trap site, and the core effects in the interaction can be neglected.

In the capture by a near-interface oxide trap, a different situation from bulk capture occurs. Taking into account the short distance d between the interface and the defect state in the oxide (typically less than 2 nm), the core effects in the interaction can become dominant in the interaction. Consequently, we have considered a solid-cube wave function as in Ref. 12 for the defect of a side length z_T, and with a volume V_T, calculated from the δ-function model radius, a_T.

Due to the lack of information on the trap properties, such as the trap potential, the estimation of this radius becomes a very difficult task. If the radius is calculated as trap is in the oxide bulk, experimental capture and emission times cannot be reproduced. In fact, it is reasonable to assume that this does happen, since Si can influence the effective trap volume due to its position. Moreover, it is also possible that the trap potential locally modifies the Si-SiO_2 barrier. For these reasons, we have introduced an empirical expression for this radius: the arithmetic mean between the radii obtained for trap in bulk oxide and in bulk silicon,

$$a_T=\frac{\hbar/[2m_{\rm ox}^*(E_{\rm Cox}-E_T)]^{1/2}+\hbar/[2m_{\rm Si}^*E_T^0]^{1/2}}{2}, \qquad (4)$$

where $m_{\rm ox}^*$ and $m_{\rm Si}^*$ are the carrier effective mass in the oxide and in Si, respectively, $E_{\rm Cox}$ is the SiO_2 conduction-band minimum, and E_T^0 the energy difference between the ground subband edge and the trap level. Thus, the wave-function model is

$$|b(x,y,z)\rangle=\left[\frac{1}{z_T^3}\right]^{1/2}=\left[\frac{1}{4\pi(a_T)^3/3}\right]^{1/2}=V_T^{-1/2},$$

$$-\frac{z_T}{2}\simeq-4^{1/3}\frac{a_T}{2}\leqslant x,y\leqslant\frac{z_T}{2}\simeq4^{1/3}\frac{a_T}{2},$$

$$d+4^{1/3}\frac{a_T}{2}\leqslant z\leqslant d-4^{1/3}\frac{a_T}{2}. \qquad (5)$$

B. Channel wave-function model

The free state in the ith subband can be represented by the normalized function:

$$|f^i(r)\rangle = \frac{1}{A^{1/2}}\,\zeta^i(z), \tag{6}$$

where $\zeta^i(z)$ is the normalized envelope function in the direction normal to the interface of the ith subband and A is the channel area.

C. Electron-phonon coupling potential

For the electron-phonon coupling, the potential form (polar or nonpolar) does not need to be detailed for this calculation since it will be included in the Huang-Rhys factor, S. However, in order to show the relationship of the matrix element with this factor, the simplest form of the potential has been used. Within the approximation given by Eq. (2) the S dependence in the final probability expression is the same whatever the potential.[12]

With these assumptions and the previous models, the matrix elements can be evaluated according to Eq. (3).

1. Huang-Rhys factor S

$$\begin{aligned}\langle b|U|b\rangle &= D_0\,\frac{1}{V_T}\int dr\; e^{iq\tau}\\ &= K_S\int_{-z_T/2}^{z_T/2} dx\; e^{iq_x x}\int_{-z_T/2}^{z_T/2} dy\; e^{iq_y y}\int_{d-z_T/2}^{d+z_T/2} dz\; e^{iq_z z}\\ &= K_S I_x I_y I_z,\end{aligned} \tag{7}$$

where we have separated the calculation in the directions parallel and normal to the interface. D_0 is the electron-phonon coupling constant and $K_S = D_0/V_T$. Taking into account that $I_x = I_y$ and that the maximum phonon module is $6^{1/2}/a_T$, to perform the sums over the phonon modes we approximate by means of[11]

$$S \approx \frac{V}{(2\pi)^3 2(\hbar\omega)^2}\,\frac{K_S^2}{}\left[\int_{-q_{xm}}^{q_{xm}} dq_x |I_x|^2\right]^2 \int_{-q_{zm}}^{q_{zm}} dq_z |I_z|^2,$$

$$q_{xm} = q_{zm} = \frac{\sqrt{2}}{2a_T}. \tag{8}$$

The evaluation of Eq. (8) relies on the factor K_S, which contains the unknown constant D_0. The value of the Huang-Rhys factor S will be fixed when comparing our results with experimental data. Very few theoretical estimations are found in the literature on the S factor for this type of transitions and those reported give high values.[13]

2. Transition matrix element $|V|^2$

Analogously, to the calculation of the S factor, we have for each subband:

$$\begin{aligned}\langle f|U|b\rangle &= D_0\left[\frac{1}{AV_T}\right]^{1/2} I_x I_y \int_{d-z_T/2}^{d+z_T/2} dz\; \zeta^i(z) e^{iq_z z}\\ &= K_V I_x I_y I_z'.\end{aligned} \tag{9}$$

Therefore, according to Eq. (3),

$$|V|^2 \approx \frac{V}{(2\pi)^3}\,K_V^2\left[\int_{-q_{xm}}^{q_{xm}} dq_x |I_x|^2\right]^2 \int_{-q_{zm}}^{q_{zm}} dq_z |I_z'|^2, \tag{10}$$

where $K_V = D_0/(AV_T)^{1/2}$. To eliminate some of the integrals in Eq. (10), the transition matrix element has to be expressed as a function of the S factor:

$$\begin{aligned}|V|^2 &\approx 2S(\hbar\omega)^2\,\frac{K_V^2\int_{-q_{zm}}^{q_{zm}} dq_z |I_z'|^2}{K_S^2\int_{-q_{zm}}^{q_{zm}} dq_z |I_z|^2}\\ &= 2S(\hbar\omega)^2\,\frac{V_T\int_{-q_{zm}}^{q_{zm}} dq_z |I_z'|^2}{A\int_{-q_{zm}}^{q_{zm}} dq_z |I_z|^2}.\end{aligned} \tag{11}$$

To study the contribution of the trap properties in the capture probability, a transformation of Eq. (11) is needed. To that end, an expression as a function of the real-space coordinates would be very useful. As a consequence, we have tried to relate the integrals of Eq. (11) with real-space functions. For usual values of d and z_T, the results of this analysis gave us the following approximations, where it is important to highlight that the overlap integral between the initial and the final states has been recovered:

$$\begin{aligned}\int_{-q_{zm}}^{q_{zm}} dq_z |I_z'|^2 &\simeq \pi\int_{d-z_T/2}^{d+z_T/2} dz |\zeta^i(z)|^2,\\ \int_{-q_{zm}}^{q_{zm}} dq_z |I_z|^2 &\simeq z_T.\end{aligned} \tag{12}$$

Therefore, Eq. (11) can be expressed for the ith subband by

$$|V|^2 \approx 2\pi S(\hbar\omega)^2\,\frac{z_T^2}{A}\int_{d-z_T/2}^{d+z_T/2} dz |\zeta^i(z)|^2. \tag{13}$$

As a result, the multiphonon emission probability of a carrier in the ith subband by the oxide trap, $W_{\rm mp}^i$ can be calculated from Eq. (2), using Eqs. (5) and (13):

$$\begin{aligned}W_{\rm mp}^i &\simeq 5\pi^2\hbar(S\omega)^2\,\frac{a_T^2}{A}\left[\int_{d-z_T/2}^{d+z_T/2} dz |\zeta^i(z)|^2\right]\\ &\times\left(1-\frac{p^i}{S}\right)^2 G(E_T^i).\end{aligned} \tag{14}$$

As can be observed in Eq. (14), the Huang-Rhys factor S and the energy of phonons involved in the transition $\hbar\omega$ are the parameters that arise due to the use of the multiphonon emission model. In order to quantify the influence of the values of these parameters on the multiphonon probability versus temperature, we have carried out calculations for several values of the phonon energy centered on a typical value of 20 meV.[13] In Fig. 4, temperature-dependent factors in Eq. (14) are represented for a typical situation of high-relaxation energy, $S\hbar\omega = 1$ eV, and an oxide-trap level near the Si

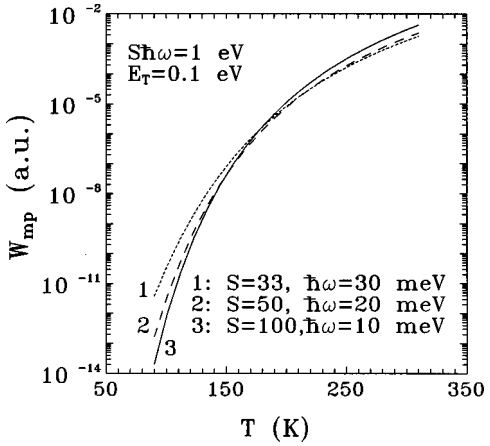

FIG. 4. Thermal dependence of multiphonon probability (in arbitrary units) for a trap with $E_T=0.1$ eV and high electron-phonon coupling, $S\hbar\omega=1$ eV, varying the Huang-Rhys factor S and the phonon energy $\hbar\omega$. Curves were calculated with (1) $S=33$ and $\hbar\omega=30$ meV, (2) $S=50$ and $\hbar\omega=20$ meV, and (3) $S=100$ and $\hbar\omega=10$ meV.

conduction-band minimum, $E_C-E_T=0.1$ eV. Keeping these conditions fixed, the phonon energy has been varied from 10 to 30 meV (with the proper value of S in each case). From this figure, it can be observed that the thermal behavior remains almost the same for the three cases, so within this interval it seems clear that this parameter is not critical in the multiphonon probability expression. Therefore, as in Ref. 13 we have hereafter taken the value of $\hbar\omega=20$ meV in all our calculations, obtaining S in the comparison with experimental data.

IV. COULOMB BARRIER

Once an expression for the multiphonon capture rate has been obtained, the Coulomb barrier effect must be included for a complete calculation of the capture rate.[14,15] Briefly, when an electron is trapped at the Si-SiO_2 interface, electrostatic charges in the substrate, channel and gate are modified. This charge variation expends an energy called Coulomb energy ΔE, which must be provided and which reduces the capture probability according to Eq. (1).

In this work, we have calculated the Coulomb energy considering the ground subband minimum in the inversion layer instead of the conduction-band edge. Thus,

$$\Delta E=Q_G(E_{FS}-E_{FM})+Q_S(E_{FS}-E_{v0})+Q_{\rm ch}(E_{FS}-E_0)-(E_{FS}-E_0), \tag{15}$$

where E_0 is the ground subband minimum, E_{FS} and E_{FM} are the Fermi levels in the semiconductor and in the gate metal, respectively, E_{V0} is the energy of the valence-band edge in the semiconductor substrate, and Q_G, Q_S, and $Q_{\rm ch}$ the normalized image charges on the gate, substrate, and channel, respectively. These charges have been numerically calculated in the quantum resolution of the MOS structure.[22] Numerical results coincide with those given in Ref. 14, although we have here considered that $Q_{\rm ch}$ must be raised to an energy of E_0 instead of the minimum of the conduction band to account for the subbands.

V. CALCULATION OF THE CAPTURE AND EMISSION TIMES

The next step in this calculation is to obtain the capture and emission times from the capture probability. Given the two-dimensional character of electrons in the inversion layer, we have defined (similarly as for the three-dimensional case) the capture coefficient per unit area or a surface capture rate for the ith subband, c', as

$$c'^i=W_c^iA. \tag{16}$$

In the three-dimensional calculation, the capture time can be calculated as the inverse of the capture coefficient times the carrier concentration. Therefore, in this case each subband contributes with its carrier concentration per unit area N_i:

$$N_i=\frac{kTn_{vi}m^*}{\pi\hbar^2}\ln(1+e^{(E_{FS}-E_i)/kT}), \tag{17}$$

where n_{vi} is the subband degeneration. Therefore, the mean capture time from the ith subband τ_c^i is given by

$$\frac{1}{\tau_c^i}=c'^iN_i\simeq 5\pi^2\hbar(S\omega)^2\frac{a_T^2}{A}\left[\int_{d-z_T/2}^{d+z_T/2}dz|\zeta^i(z)|^2\right]\times\left(1-\frac{p^i}{S}\right)^2G(E_T^i)e^{-\Delta E/kT}N_i. \tag{18}$$

For the calculation of the emission time τ_e^i, accounting for the Coulomb energy, the detailed balance principle has been used with the degeneration ratio $g=1$, giving us

$$\frac{1}{\tau_e^i}=\frac{1}{\tau_c^i}e^{(E_{TF}+\Delta E)/kT}, \tag{19}$$

where E_{TF} is the difference between the trap level and the Fermi level at the interface shown in Fig. 2.

Therefore, to calculate the capture and emission times of a system such as the one shown in Figs. 1 and 2, it is necessary to evaluate Eqs. (18) and (19). To do so, a complete quantum resolution of the MOS structure has to be carried out in order to obtain the subband minima, the electron concentration per unit area, the envelope wave function for each subband, and the surface potential. Image charges must also be calculated. Such numerical calculations have been carried out by the self-consistent solution of the Poisson and Schrödinger equations in an n-channel MOS with the substrate orientation according to the experimental data availability. More details of the numerical simulation are available in Refs. 22 and 23.

From Eq. (18) the different aspects involved in this process can readily be identified: (i) The multiphonon contribution is clear in factor S and in the $G(E_T^i)$ function; (ii) the integral between brackets that determines the overlapping between the subband and the trap state involving the tunnel transition contribution; (iii) the Coulomb barrier effect in the negative exponential; and (iv) the electron surface concentration, N_i.

The $G(E_T^i)$ function includes the main thermal dependence of the capture and emission times apart from the one contained in the Coulomb energy exponential. With regard to

the gate-voltage dependence, this can be observed in the fact that, as the gate voltage increases, the envelope wave function penetrates deeper into the oxide, making the overlap with the trap wave function greater. Furthermore, the electron concentration in the inversion layer also increases. In addition, the Coulomb energy decreases at different rates depending on whether the device is in the operation region of weak or strong inversion. In consequence, when the gate voltage increases, the capture process is favored, and therefore, the capture time decreases. On the other hand, it is important to note that the numerical capture time does not obey the Shockley-Read-Hall statistics, since we have verified that this time is not simply inversely proportional to the subband population, but clearly shows a steeper slope. This fact has been observed experimentally as well.[14] Below we will show the relative contribution of each factor of Eqs. (18) and (19) to the total time.

As mentioned, the numerical solution of the MOS structure for different temperatures and gate voltages provides different values for N_i, ΔE, E_{TF}, and E_T^i needed in these calculations. In the variation of the two latter magnitudes, the trap distance to the interface, d (Fig. 3), must be considered since the change of the potential in the oxide at a distance d induces a variation in the energy levels. These variations can be expressed as

$$\Delta E_{TF}(T,V_g)=\Delta E_T-\Delta E_{FS}=-q\,\frac{d}{T_{\rm ox}}\,[\Delta V_g-\Delta\phi_S(T,V_g)] - \Delta E_{FS}(T,V_g),$$

$$\Delta E_T^i(T,V_g)=\Delta E_i(T,V_g)+q\,\frac{d}{T_{\rm ox}}\,[\Delta V_g-\Delta\phi_S(T,V_g)], \tag{20}$$

where $T_{\rm ox}$ is the oxide thickness, V_g the gate-substrate voltage, E_{FS} the Fermi energy at the interface, and ϕ_S the surface potential. The first expression in Eq. (20) shows that the increase of the emission time with the gate voltage occurs as a consequence of the greater difference between the trap level and the surface Fermi level.

As a consequence of this thorough study, it can be observed that the thermal and gate-bias dependencies of the times are distributed among the multiphonon emission mechanism, Coulomb energy, subband electron concentration, wave-function overlapping, and the variations in the energy differences in Eq. (20).

VI. COMPARISON WITH EXPERIMENTAL DATA AND DISCUSSION

In our numerical procedure, we introduce the trap level, $E_{\rm Cox}-E_T$, and the trap distance to the interface, d, with the S factor remaining as the only free parameter. In the usual case of none of these three magnitudes being known, the double comparison with experimental capture and emission times versus temperature and gate voltage would provide enough information to fix the three parameters. The method of obtaining these parameters is the following: once the Coulomb energy values as a function of temperature and gate bias have been calculated, the distance to the interface is obtained from the comparison of the numerical and experimental slopes in the representation of the ratio time constants versus gate voltage. Then, the energy trap depth is derived by matching the experimental and numerical cross-points of the representations of capture and emission times against gate bias. Finally, the comparison of the numerical and experimental thermal dependencies fixes the Huang-Rhys factor. Therefore, when experimental thermal and gate-voltage dependencies are available, there is only one solution in the obtaining of the values of these parameters.

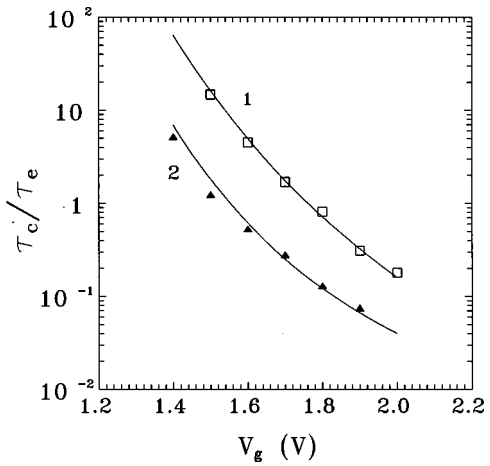

FIG. 5. Comparison between experimental ratio τ_c/τ_e (squares for trap 1 and triangles for trap 2) (Ref. 24) and our numerical results (lines) vs gate voltage for two observed traps at room temperature. In the numerical simulation, an oxide thickness of 8.6 nm and channel doping of $N_A=5\times10^{17}$ cm^{-3} were considered.

The results from our calculations have been compared with experimental data available in the literature for different samples and for different temperature and gate-bias conditions. Prior to the analysis of experimental data by our procedure, a detailed knowledge of the technological parameters of the MOS structure is necessary in order to introduce them in our simulator. That is why only cases where these parameters are available have been studied. Comparisons shown in the first part of this section only account for transition from and towards one subband in the inversion layer. At the end of this section, we will discuss the comparable contribution of the ground and first subband to the capture and emission times extracted from our numerical results and justify why only one subband is considered.

First, in Fig. 5, our numerical calculations (lines) have been compared with experimental time constant ratios τ_c/τ_e at different gate voltages,[24] for two traps (symbols) in the oxide at room temperature. For the calculation, oxide thickness and substrate concentration were included in the simulator to obtain the necessary magnitudes to evaluate Eqs. (18) and (19) as a function of gate voltage. We have obtained the following trap parameters: RTS labeled as 1, $d=0.6$ nm, $E_{\rm Cox}-E_T=3.15$ eV, $S\hbar\omega=2.2$ eV; RTS labeled as 2, $d=0.4$ nm, $E_{\rm Cox}-E_T=3.23$ eV, $S\hbar\omega=2.0$ eV.

In both cases, the thermal activation energy for the capture obtained by our results was approximately 0.52 eV, a value which is in good agreement with previously reported data.[6] The Coulomb energy calculated was between 168 meV at 1.4 V and 100 meV at 2 V in the gate. The introduction of the Coulomb energy and its dependence on gate voltage makes the distance d obtained different from other re-

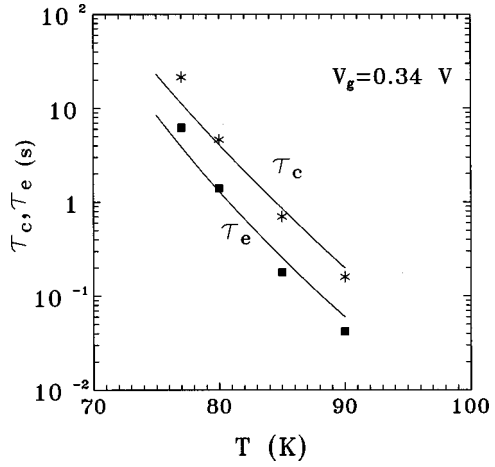

FIG. 6. Experimental thermal dependence of time constants for trap 2 in Ref. 9 (squares for emission times and asterisk for capture times) compared with our numerical results (lines) at a gate voltage of $V_g = 0.34$ V. An oxide thickness of 5.5 nm and an implanted channel doping with a concentration peak of 3×10^{17} cm^{-3} located at 0.1 μm to the interface were included in the simulation.

sults where this energy was not considered. As we will show throughout this work, the inclusion of the Coulomb energy is necessary to reproduce experimental data. Thus, distances can now be better estimated.

A more complete comparison has been made with RTS data from Shi, Miéville, and Dutoit.[9] In contrast with the above data, in this case, sample fabrication process details provided very useful information for including an accurate substrate profile in the simulations. Capture and emission time data from trap 2 in Ref. 9 were analyzed. In Fig. 6, comparisons for thermal dependencies are shown for both times, experimental with symbols (asterisks for capture times and squares for emission times) and our result with solid lines. Gate-voltage trends are reproduced in Fig. 7, giving the following parameter values as a result of the two comparisons: $d = 0.7$ nm, $E_{\mathrm{Cox}} - E_T = 3.11$ eV, $S\hbar\omega = 0.72$ eV, and Coulomb energy from 57 to 54 meV decreasing with gate voltage. It is important to highlight that both good

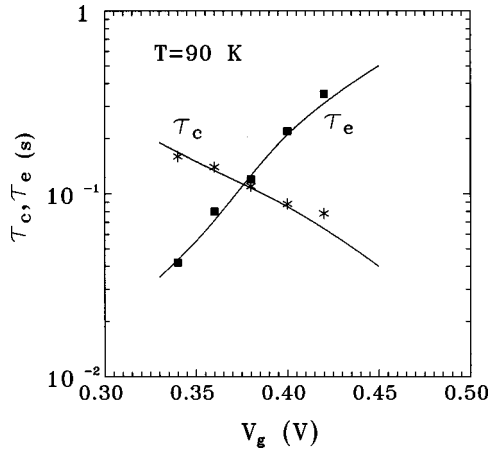

FIG. 7. Comparison between experimental time constants (squares for emission times and asterisk for capture times) and the numerical results (lines) as a function of gate voltage at $T = 90$ K. The trap analyzed is the same as in Fig. 6.

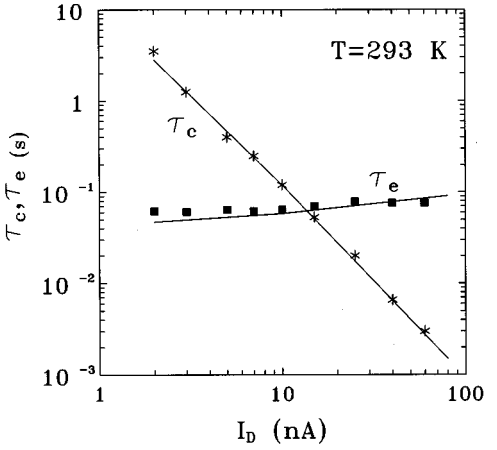

FIG. 8. Drain current dependence of time constants for trap analyzed in Ref. 14 (squares for emission times and asterisk for capture times) compared with our numerical results (lines) at $T = 293$ K. An oxide thickness of 17 nm and a channel doping of 3×10^{17} cm^{-3} were included in the simulation. Drain current calculation was performed with a short-channel metal-oxide-semiconductor field-effect transistor simulator developed by our research group (Refs. 22 and 23).

agreements have been obtained with the same trap parameters, which shows suitable modeling of the process. The high values of the relaxation energy, $S\hbar\omega$, obtained with these kind of traps are coherent with results for neutral traps in the oxide bulk.[25]

In addition, in Fig. 8, a comparison between experimental data from Shulz[14] and our calculations are shown as a function of drain current at 293 K. Now, trap parameters obtained from this fitting are $d = 0.9$ nm, $E_{\mathrm{Cox}} - E_T = 3.26$ eV, and $S\hbar\omega = 1.2$ eV, resulting in an activation energy of 0.29 eV according to experimental data. The Coulomb energy varies from 0.23 to 0.12 eV within the current drain interval from 2 to 80 nA. In this case, we have related gate voltages to drain current through the same simulator adapted to calculate current-voltage characteristics for short-channel MOS transistors.[23]

After comparisons with experimental data, our numerical procedure allows us to analyze in depth some interesting aspects of this process. First, we have studied the influence of the trap distance to the interface, d, in the time constants at different gate voltages. Obviously, the larger d is the lesser the overlap between trap state and channel envelope wave function. Thus, the integral of Eq. (18) is reduced and the capture and emission times increase. Moreover, this parameter influences the slopes of these times with gate voltage through the relations given in Eq. (20). Accordingly, if d increases these slopes also increase since the variation of the energy difference between states with gate voltages is more pronounced. Both considerations can be observed in Fig. 9, where we have represented the results of our calculation for different values of the d parameter, including the one plotted in Fig. 7. It is clear that for the highest values of d, 1.4 nm, the slopes are more pronounced than in the other cases. The change of time values with d agrees with this explanation. It is also important to point out how the Coulomb energy and transition matrix element $|V|^2$ dependencies on gate voltage decisively contribute to the reproduction of experimental

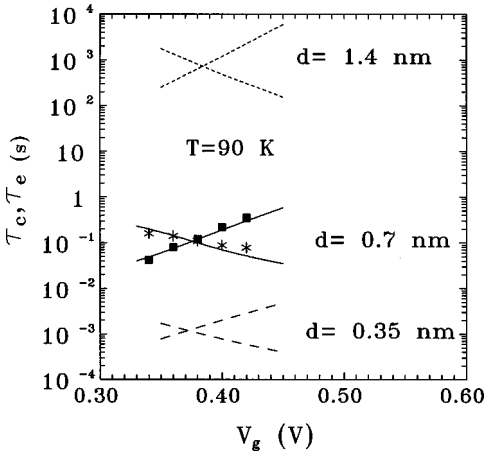

FIG. 9. Dependence of the capture and emission times on the trap distance to the interface, d. Numerical simulations have been done for $d=1.4$ nm (dotted lines), 0.7 nm (solid lines), and 0.35 nm (dashed lines) with the other parameters having the same values as in Fig. 7. Experimental data are also shown with symbols. Changes in slopes and actual values are quite evident.

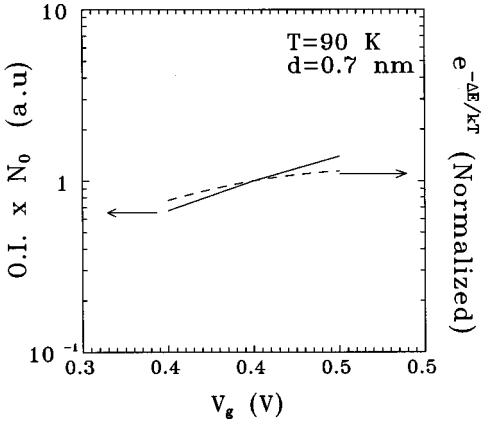

FIG. 11. Numerical normalized contributions of the overlap integral (OI) times ground subband population (solid line) and the exponential containing the Coulomb energy (dashed line) for the trap shown in Fig. 7. In this case MOSFET operation maintains almost the same state of inversion.

data. This fact can be clearly observed in Figs. 10 and 11, where the behavior of the overlap integral (OI) times the ground subband population N_0, and the exponential term including the Coulomb energy, both normalized, have been plotted as a function of gate voltage. In Fig. 10 the simulation data used correspond to trap 1 in Fig. 5. In this case, the increase of the capture rate is more pronounced for the Coulomb contribution because, as we have verified by simulation, the sample is in weak inversion, where the Coulomb energy varies quickly with gate voltage. On the other hand, in Fig. 11, where the data from Fig. 7 have been used, the contribution of the overlap integral and N_0 dominates due to the slow change of ΔE with gate voltage in this region of operation. Therefore, all these contributions influence the capture time behavior with gate voltage and explain the different d values obtained in our fitting and other works.

We will now treat the subject of the contribution of the different subbands to the process. In previous sections of this work, our procedure has allowed us to calculate capture and emission times for each channel subband. It is generally accepted that the experimentally observed RTS is a transition of a single electron between a specific channel subband and a trap site. Thus, it would be wrong to average the transition rates from each subband as if a multiple particle process occurred. We have, therefore, dealt with individual transitions from each subband. The contributions of the ground and first subband to the RTS times are plotted in Figs. 12 and 13. The samples studied correspond to those represented in Figs. 7 and 8, respectively, but now the capture and emission times obtained for each subband are represented with the experimental data. Moreover, it is important to note that these calculations are quite general because the two samples are in different thermal and operational conditions. It can be observed that for the range of gate voltages considered here, transitions are more likely from the first excited subband since the envelope wave function of this subband $\zeta^1(z)$ extends deeper into the oxide than the ground subband. As a consequence, the product of the overlap integral and N_1 is

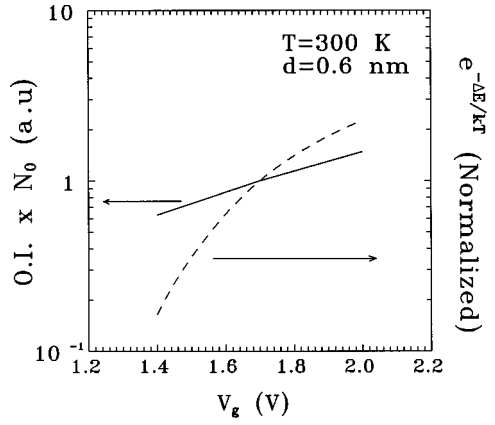

FIG. 10. Numerical normalized contributions of the overlap integral (OI) times ground subband population (solid line) and the exponential containing the Coulomb energy (dashed line) for trap 1 in Fig. 5. Gate voltages make the MOS transistor operate in the weak inversion region where ΔE undergoes considerable changes.

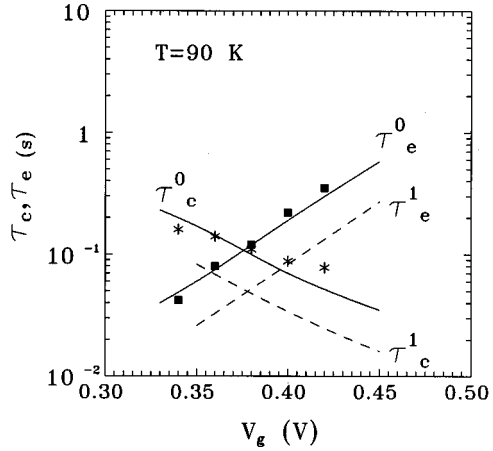

FIG. 12. Numerical capture and emission time for the ground subband (τ_c^0 and τ_e^0, respectively, in solid lines) and the first excited subband (τ_c^1 and τ_e^1, respectively, in dashed lines) and experimental data (symbols) corresponding to Fig. 7.

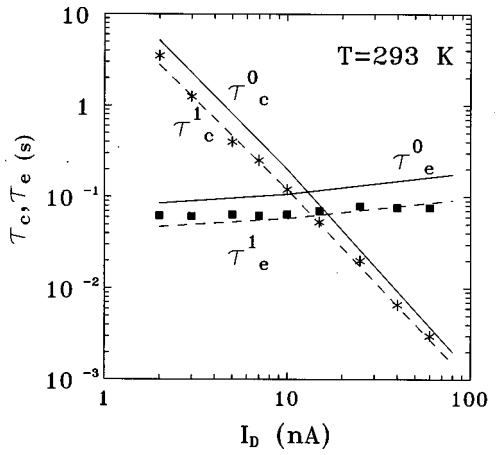

FIG. 13. Numerical capture and emission times for the ground subband (τ_c^0 and τ_e^0, respectively, in solid lines) and the first excited subband (τ_c^1 and τ_e^1, respectively, in dashed lines) and experimental data (symbols) corresponding to Fig. 8.

more important than the same magnitude calculated for the ground subband. As can be observed in Figs. 12 and 13, both time constants are quite similar, being separated by a factor of less than 2. It has been verified that time constants for higher subbands are much greater than those shown.

At this point, an interesting question can be raised. According to the calculation presented in this paper, each subband contributes with its own time constant, so if more than one subband participated, more than one time constant should be assigned to the up and down times of a RTS or, at least, and what is more likely to occur, a nonexponential distribution of times should appear. However, Figs. 12 and 13 show that these time constants are very close to each other. Thus, if both time constants were looked for, it would have been very difficult to separate them in the time constant histogram. Therefore, it would be quite possible for an almost perfect exponential distribution of times to be obtained, and it would appear no different from a process with only one averaged characteristic time, as has happened until now. We can also justify why the first comparisons were done considering only one subband. First, it is clear from Figs. 12 and 13 that the change in the trap parameters obtained by fitting from numerical results for ground and first excited subbands is unimportant. Of course, the two numerical results could be averaged, but that would not have given additional information, above all since there are not enough data about the experimental exponential distributions used to calculate the time constant. Moreover, depending on the average interval during the RTS measurements, the experimental results could eventually change.

VII. CONCLUSIONS

In this paper, a two-dimensional treatment of the capture and emission processes by neutral near-interface traps is presented. Multiphonon emission, tunnel transition, and Coulomb barrier effects have been combined to obtain expressions for the RTS time constants, where the contribution of each parameter can be easily explained. Numerically, solving Poisson and Schrödinger equations in a MOS structure has allowed us to obtain the information necessary to explain the different aspects involved in this phenomenon, which had not been done until now. The proposed numerical procedure has let us identify the contribution of the characteristics for each trap (energy depth, distance to the interface, and coupling factor) in the thermal and gate-voltage dependencies of the capture and emission times of RTS.

Our theoretical results have been successfully compared with different experimental data, showing good agreement for thermal and gate-voltage dependencies. When both experimental dependencies are available, comparisons with our results provide information about the trap, such as energy level and distance to the interface. In this case, the Huang-Rhys factor, which accounts for the electron-phonon coupling strength, is also fixed in the calculation. We have been able to determine that the trap distance to the interface is reduced from previous estimations since surface potential and Coulomb energy gate-voltage dependencies have been included. Therefore, it seems clear that trap distance estimations made with data in the weak and moderate inversion regimen are considerably overestimated.

Given the single transition character of this process, we have calculated time constants for each subband in the MOS inversion channel. According to our calculations, the ground and first excited subbands are the most important contributors to the RTS, with similar up and down times. This fact, apart from the similar values of the two time constants could explain why they have not been found.

To sum up, the quasi-two-dimensional behavior of electrons in the inversion layer lead us to believe that the treatment that has been developed in this paper could be a more realistic approach to explain this phenomenon.

ACKNOWLEDGMENTS

This work has been carried out within the framework of research Project No. TIC95-0511, supported by the Spanish Government (CICYT).

*Fax: 34-58-243230. Electronic address: apalma@gcd.ugr.es.

[1] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth, and D. M. Tennant, Phys. Rev. Lett. **52**, 228 (1984).

[2] R. E. Howard, W. J. Skocpol, L. D. Jackel, P. M. Mankiewich, L. A. Fetter, D. M. Tennat, R. Epworth, and K. S. Ralls, IEEE Trans. Electron Devices **ED-32**, 1669 (1985).

[3] M. J. Uren, D. J. Day, and M. J. Uren, Appl. Phys. Lett. **47**, 1195 (1985).

[4] K. R. Farmer, C. T. Rogers, and R. A. Buhrmann, Phys. Rev. Lett. **58**, 2255 (1987).

[5] E. Simoen and C. Claeys, Appl. Phys. Lett. **75**, 3647 (1994).

[6] M. J. Kirton and M. J. Uren, Adv. Phys. **38**, 367 (1989).

[7] A. Ohata, A. Toriumi, M. Iwase, and K. Natori, J. Appl. Phys. **68**, 200 (1990).

[8] B. J. Gross and C. G. Sodini, Tech. Dig. Int. Electron Devices Meet. **1992**, 881.

[9] Z. Shi, J.-P. Miéville, and M. Dutoit, IEEE Trans. Electron Devices **41**, 1161 (1994).

[10] M. E. Zvanut, F. J. Feigl, W. B. Fowler, J. K. Rudra, P. J. Caplan, E. H. Poindexter, and J. D. Zook, Appl. Phys. Lett. **54**, 2118 (1989).

[11] B. K. Ridley, J. Phys. C **13**, 2027 (1980).
[12] B. K. Ridley, *Quantum Processes in Semiconductor*, 3rd ed. (Clarendon, Oxford, 1993).
[13] W. B. Fowler, J. K. Rudra, M. E. Zvanut, and F. J. Feigl, Phys. Rev. B **41**, 8313 (1990).
[14] M. Shulz, J. Appl. Phys. **74**, 2649 (1993).
[15] H. H. Mueller, D. Wörle, and M. Schulz, J. Appl. Phys. **75**, 2970 (1994).
[16] F. Stern, Phys. Rev. B **5**, 4891 (1972).
[17] D. Goguenheim and M. Lannoo, J. Appl. Phys. **68**, 1059 (1990).
[18] J. H. Zheng, H. S. Tan, and S. C. Ng, J. Phys.: Condens. Matter **6**, 1695 (1994).
[19] K. Huang, Sci. Sin. **24**, 27 (1981).
[20] M. A. Amato and B. K. Ridley, J. Phys. C **13**, 2027 (1980).
[21] G. Lucovsky, Solid State Commun. **3**, 299 (1965).
[22] J. A. López-Villanueva, I. Melchor, F. Gámiz, J. Banqueri, and J. A. Jiménez-Tejada, Solid-State Electron. **38**, 203 (1995).
[23] J. B. Roldán, F. Gámiz, J. A. López-Villanueva, J. E. Carceller, and P. Cartujo, Semicond. Sci. Technol. **12**, 321 (1997).
[24] K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, IEEE Electron Device Lett. **11**, 90 (1990).
[25] A. Palma, J. A. López-Villanueva, and J. E. Carceller, J. Electrochem. Soc. **143**, 2687 (1996).