

## Monte Carlo study of electron transport in silicon inversion layers

M. V. Fischetti and S. E. Laux

*IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598*

(Received 21 September 1992)

Electron transport in Si inversion layers at 300 K is studied using a self-consistent Monte Carlo solution of the Boltzmann transport equation coupled to the two-dimensional Poisson equation and the one-dimensional Schrödinger equation. Physical elements included in the model are (1) nonparabolicity effects to treat quantization in the inversion layer; (2) static screening of the Coulomb interactions accounting for the population of many subbands; (3) anisotropy of the deformation-potential interaction, shown to be quite important in the case of a two-dimensional electron gas (2DEG); (4) a careful analysis of the dynamic screening of the deformation-potential interaction, showing that the interaction between electrons and acoustic phonons can be approximated by the unscreened interaction in the nondegenerate limit of a 2DEG; and (5) the inclusion of interface SiO<sub>2</sub> optical phonons. Up to ten subbands have been included to study the 2DEG together with a bulk-transport model employed to handle high-energy electrons. We have obtained mixed results: In the Ohmic regime, we have found a phonon-limited mobility that exhibits the correct dependence on carrier density, but which is about 20% larger than the experimental data. This still represents an improvement upon previous nonempirical theories, and even better quantitative agreement is obtained at the very low and very high carrier densities at which Coulomb scattering and scattering with surface roughness, respectively, control the mobility. At high longitudinal fields we find a bulklike saturated velocity, in agreement with some experimental results, but not with many others that we consider more reliable.

### I. INTRODUCTION AND OVERVIEW

In their 1982 review of the electronic properties of two-dimensional systems, Ando, Fowler, and Stern<sup>1</sup> stated that the “theory of phonon scattering is [ · · · ] at an unsatisfactory stage,” being unable to account for the measured field-effect electron mobility in Si inversion layers. An equally negative picture had been already painted by Basu in 1978 regarding the high-field drift velocity of electrons in *n* channels.<sup>2</sup> From a Monte Carlo study of high-field electron transport, he concluded that his “model appears inadequate to account for the variation of the drift velocity with electric field in (100)-oriented inversion layer.” After more than a decade since these statements were made (and more than three decades since field-effect transistors have permeated our lives), the situation has not improved much. In the Ohmic regime, experimental results are consistent and complete,<sup>3–8</sup> but the phonon-limited electron mobility is overestimated by many theoretical models,<sup>9–12</sup> good agreement being obtained only using simplified<sup>13,14</sup> or empirical<sup>15,16</sup> physical pictures. The high-field properties of electron transport appear to be confusing even at the experimental stage, saturated drift velocities ranging from the bulk value<sup>5,17</sup> ( $\approx 10^7$  cm/s) down to about half this value.<sup>18–21</sup>

The purpose of this paper is to describe our attempts to improve the situation, focusing our attention on the room-temperature behavior, more complicated than the low-temperature behavior, but certainly more interesting from a practical point of view. These attempts have been made in the same spirit of our previous simulations of electronic transport in bulk semiconductors.<sup>22,23</sup> We try to embrace a physical model as complete as today’s com-

puting facilities allow us. As a rule, we relax many approximations employed in the past and try to understand the effect of these approximations. Occasionally, the large computing power at our disposal lets us improve the physical model beyond the accepted “state of the art.” Monte Carlo simulations are chosen for their flexibility as far as implementation of the physics is concerned.<sup>24,25</sup> Moreover, the solution of the transport equation provided by the Monte Carlo technique needs not to be limited by additional approximations dictated by mathematical difficulties. This is quite a positive feature, since approximations dictated by our ignorance of the physics are already many and significant. Finally, notice that we have relied on Monte Carlo solutions even in the Ohmic regime, in which direct approximate solutions of the Boltzmann transport equation are usually obtained. Indeed, given the numerical complexity already needed to implement our self-consistent approach and to evaluate the scattering processes we have considered, the additional computational burden of the Monte Carlo technique is minor.

Considering the quite large amount of formulas and the level of details we shall go into in the bulk of this paper, it is useful to give here an overview of the work, emphasizing the key physical elements of the model, some of them new, and the main results we obtained. The physical model we have employed includes scattering with all bulk phonons, scattering with surface optical modes, scattering with surface roughness, and Coulomb scattering with ionized impurities, and interface/oxide charges. What differentiates our approach from previous work are the following elements.

(1) We have considered the effect of the band nonpara-

bolicity on the subband structure of Si inversion layers. This leads to a 20% reduction of the phonon-limited mobility at 300 K.

(2) The anisotropy of the deformation-potential interaction between electrons and acoustic phonons is found to be quite important in two-dimensional systems. This leads to a 10–20 % enhancement of the mobility, almost cancelling the effect of nonparabolicity.

(3) The scattering rates have been computed using Fermi's Golden Rule, but otherwise dropping any additional approximations. Thus all electronic form factors are computed using self-consistent wave functions and realistic scattering potentials. Intracuband and intersubband, intravalley and intervalley processes have all been included, given the small incremental computational cost of accounting for often negligible scattering processes.

(4) Static screening (applicable to Coulomb and surface roughness scattering) has been included by accounting for the population of many subbands, by extending Stern's two-subband calculations,<sup>26</sup> similar to what was done by Yokoyama and Hess<sup>27</sup> in the case of the  $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$  system. On the other hand, the deformation-potential electron-phonon interaction has been left unscreened: We have investigated in detail the role of screening in this case and found that dynamic screening must be used to treat this interaction in two dimensions, in contrast to the bulk case. In the range of low-to-medium densities ( $\approx 10^{11}$ – $10^{12}$   $\text{cm}^{-2}$ ) and high temperatures at which the electron-phonon interaction controls the mobility, a partial cancellation of dynamic screening and antiscreening at short and long wavelengths, respectively, is found to reduce the role of screening to such an extent that leaving the interaction unscreened amounts to a very small error.

(5) The role of the surface optical modes (coupled modes between the  $\text{SiO}_2$  longitudinal optical phonons, and interface electromagnetic waves) is investigated in detail, paying attention to the coupling between the surface phonons and the two-dimensional plasmons.

In addition, we have improved upon numerical issues, such as the large number of subbands employed, the inclusion of both two-dimensional and bulklike carriers, the full self-consistency among the Poisson, Schrödinger, and transport equations, the inclusion of degeneracy<sup>11</sup> in the scattering processes, and the generality of situations we are able to simulate. For example, full self-consistency in the broadest sense has been obtained in the case of Si channels by Shihirata, Taniguchi, and Hamaguchi,<sup>28</sup> and for the  $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$  system by Yokoyama and Hess,<sup>29</sup> but only in the case of a uniform field. Instead, we allow for a nonuniform channel, with an arbitrary field configuration, as dictated by the Poisson, Schrödinger, and transport equations.<sup>30</sup> While most of these issues are decoupled from the physical model, we believe that in many cases such numerical issues could have influenced results obtained in the past.

After all these efforts, our main message is not one of total success. The calculated phonon-limited mobility appears to be about 20–30 % larger than the experimental data, while the high-field saturated velocity is con-

sistently higher than the experimental data we consider most reliable. Surprisingly, surface roughness and Coulomb scattering appear to be treated quite well in our model, their effects being noticeable at large and small electron sheet density  $n_s$ , respectively. While our results are not fully satisfactory, they do constitute an improvement over previous theories which, as we mentioned above, either overestimate the phonon-limited mobility<sup>9–12</sup> or can explain the experimental data in the Ohmic regime or at high fields only empirically<sup>15,16</sup> or with major simplifying approximations (such as the use of variational<sup>15,16</sup> or approximated Airy-type wave functions<sup>2</sup> of equilibrium subband structure even at large fields,<sup>2,13–15,27,31,32</sup> approximated form factors in the evaluation of the scattering rates,<sup>2,13,14,31,32</sup> and by using adjustable parameters<sup>15,16,32</sup> or neglecting screening<sup>13–16</sup>).

The paper is organized as follows: In Sec. II, we shall present the physical model we have employed, discussing in detail (1) the role of a realistic conduction-band structure on the dispersion of the two-dimensional electron gas (2DEG); (2) the electron-phonon interaction and the anisotropy of the scattering with acoustic phonons; (3) the role of interface optical modes; (4) Coulomb scattering with impurity and oxide charges, and scattering with interface roughness; and (5) screening in a situation when many subbands are populated, paying particular attention to the problem of deciding whether static screening is appropriate in the case of the deformation potential interaction. In Sec. III, we shall discuss numerical and device issues and details concerning the implementation of the physical model into a Monte Carlo (MC) simulation program. Finally, in Sec. IV, we shall conclude by presenting our results related to Ohmic mobility versus electron sheet density at 300 K, and drift velocity versus field.

## II. PHYSICAL MODEL

### A. Subband structure

The effect of conduction-band nonparabolicity on quantized electron states in low-dimensionality systems has been investigated quite extensively in the context of III-V compound semiconductors.<sup>33–37</sup> However, in the case of Si inversion layers, a parabolic-band approximation has almost always been employed, the only exception being a short investigation by Falicov,<sup>38</sup> who studied the effect of nonparabolicity on the transverse effective mass. However, considering the complicated structure of the conduction band near the  $X$  symmetry point, and the fairly large nonparabolicity of the six valleys with minima along the  $\Delta$  lines, it is sensible to ask what effect a realistic band structure might have on the subband structure of Si  $n$  channels.

In principle, the role of the band structure on the electron states in Si inversion layers can be obtained by considering the Schrödinger-like “effective-mass” equation<sup>30</sup>

$$[\epsilon(-i\nabla) + V(z)]\phi(\mathbf{r}, z) = E\phi(\mathbf{r}, z). \quad (1)$$

Here  $z$  is the coordinate along the quantization direction (that is, normal to the Si- $\text{SiO}_2$  interface in our case), while

upper and lower case vectors, here and in the following, denote three-dimensional vectors and vectors in the plane of the Si-SiO<sub>2</sub> interface, respectively. So  $\mathbf{r}$  is the spatial coordinate in the plane of the interface,  $\varepsilon(\mathbf{K})$  is the energy dispersion [so that  $\varepsilon(-i\nabla) = -\hbar^2\nabla^2/(2m_c)$  for a parabolic spherical band with effective mass  $m_c$ ],  $\hbar$  is the reduced Planck's constant, and  $V(z)$  is the external confining potential. The unknown  $\phi$  is the "envelope" wave function, i.e., a modulation of the periodic Bloch factor of the full wave function.

Equation (1) was originally derived by Slater,<sup>39</sup> by Luttinger,<sup>40</sup> and by Luttinger and Kohn,<sup>41</sup> in the context of shallow electronic states at donor centers.<sup>42</sup> Here we think it is appropriate to obtain a solution of Eq. (1) in a "constructive" way, by outlining briefly a procedure<sup>43</sup> which parallels the original derivation but is applied to our situation of a Si channel. In so doing, the approximations which constitute the "effective-mass" framework will be clearly pointed out. This derivation is workable for any type of dispersion  $\varepsilon(\mathbf{K})$ , but assumes full translational symmetry of the unperturbed problem [that is,  $V(z)=0$ ], so that we must deal with only one type of crystal, and no heterojunctions or wave-function penetration into the SiO<sub>2</sub> can be handled. As an example relevant to our model, we shall present solutions for the case where the dispersion is obtained using empirical pseudopotentials. The Schrödinger equation we must solve can be written as

$$\left[ -\frac{\hbar^2\nabla^2}{2m} + V_L(\mathbf{R}) + V(z) \right] \psi(\mathbf{R}) = E\psi(\mathbf{R}), \quad (2)$$

where  $V_L(\mathbf{R})$  is the periodic potential of the lattice and  $m$  the free-electron mass. For the case of Si inversion layers, we shall assume that the wave function  $\psi(\mathbf{R})$  vanishes at the Si-SiO<sub>2</sub> interface ( $z=0$ ) and at a distance from the interface,  $z_0$ , sufficiently large so as to avoid errors. As we stated above, the necessity of avoiding wave-function penetration in the oxide is required by the need of full translational symmetry of the unperturbed problem, and it amounts to an approximation which by now constitutes common practice, although it may be quite unsatisfactory particularly for thin insulating layers. We shall further solve the physically equivalent problem in  $(-z_0, z_0)$  with  $V(-z)=V(z)$ . To fix the notation, let us recall that in the bulk [i.e.,  $V(z)=0$ ], and approximating the lattice potential with a local pseudopotential  $V_p$ , the solution of Eq. (2) is obtained by expanding  $\psi$  as

$$\psi(\mathbf{R}) = \sum_{\mathbf{K}} \psi_{\mathbf{K}}(\mathbf{R}) \equiv \sum_{\mathbf{K}, \mathbf{G}} f_{\mathbf{G}}(\mathbf{K}) e^{i(\mathbf{K}+\mathbf{G})\cdot\mathbf{R}}, \quad (3)$$

$$\sum_n \left[ \frac{1}{2} \left[ \varepsilon \left( \mathbf{k}_0 + \mathbf{k}, k_{0,z} + \frac{m\pi}{z_0} \right) + \varepsilon \left( \mathbf{k}_0 + \mathbf{k}, k_{0,z} - \frac{m\pi}{z_0} \right) \right] \delta_{nm} + V_{nm} \right] a_n^{(\mu)}(\mathbf{k}) = E_{\mu}(\mathbf{k}) a_m^{(\mu)}(\mathbf{k}), \quad (6)$$

where

$$V_{nm} = \frac{2}{z_0} \int_0^{z_0} dz \sin \left[ \frac{n\pi z}{z_0} \right] V(z) \sin \left[ \frac{m\pi z}{z_0} \right],$$

where the  $\mathbf{G}$ 's are vectors of the reciprocal lattice. Then Eq. (2) implies

$$\frac{(\mathbf{k}+\mathbf{G})^2}{2m} f_{\mathbf{G}}(\mathbf{K}) + \sum_{\mathbf{G}'} V_p(\mathbf{G}-\mathbf{G}') f_{\mathbf{G}'}(\mathbf{K}) = \varepsilon(\mathbf{K}) f_{\mathbf{G}}(\mathbf{K}). \quad (4)$$

The eigenvalues  $\varepsilon_{\lambda}(\mathbf{K})$  and eigenvectors  $f_{\mathbf{G}}^{(\lambda)}(\mathbf{K})$  of the linear problem above yield the dispersion (band structure) and pseudo-wave-functions for band  $\lambda$  at wave vector  $\mathbf{K}$ . When the confining potential is turned on, we can expand  $\psi$  around a local conduction-band minimum at  $(\mathbf{K}_0 = \mathbf{k}_0, k_{0,z})$  as

$$\psi(\mathbf{R}) = e^{i\mathbf{K}_0\cdot\mathbf{R}} \sum_{\mathbf{k}} \phi_{\mathbf{k}}(\mathbf{r}, z), \quad (5a)$$

where

$$\phi_{\mathbf{k}}(\mathbf{r}, z) = e^{i\mathbf{k}\cdot\mathbf{r}} \sum_n a_n(\mathbf{k}) \sin \left[ \frac{n\pi z}{z_0} \right] \times \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{R}} f_{\mathbf{G}} \left[ \mathbf{k}_0 + \mathbf{k}, k_{0,z} + \frac{n\pi}{z_0} \right]. \quad (5b)$$

The goal now is to substitute the expansion (5) into Eq. (2), eliminating the lattice (pseudo)potential thanks to Eq. (4). In so doing, we obtain an equation for the coefficients  $a_n$ . The form of this equation is significantly simplified provided we assume (1) that the confining potential is "slowly varying" over a unit cell, and (2) that its matrix elements between Bloch functions in different bands are negligible. The first approximation poses no concerns, while the second approximation may be questionable and will prevent us from including valley splitting beyond the effective-mass approximation.<sup>1</sup> Another crucial approximation is required so that the solution can be factored into a periodic Bloch term modulated by a slowly varying "envelope," which is the essence of the effective-mass approach: (3) Since we have already assumed a slowly varying potential  $V(z)$  compared to the lattice potential, it is reasonable to ignore the  $n$  dependence of  $f_{\mathbf{G}}$  compared to  $\sin(n\pi/z_0)$ , so that

$$f_{\mathbf{G}} \left[ \mathbf{k}_0 + \mathbf{k}, k_{0,z} + \frac{n\pi}{z_0} \right] \simeq f_{\mathbf{G}}(\mathbf{k}_0 + \mathbf{k}, k_{0,z}).$$

Thus from Eqs. (2) and (5) we obtain

and the index  $\mu$  labels the independent eigenvectors and eigenvalues of the linear problem (6), corresponding to the different subbands. Note that for valleys which are symmetric under reflections on the  $x, y$  plane, the diago-

nal terms on the left-hand side of Eq. (6) simplify further. For some of the  $X$  valleys of Si, however, this is not so and the eigenvalues result from an average of the “left-going” and “right-going” waves of wave numbers  $\pm n\pi/z_0$ . The wave functions, as desired, are now expressed as products of a Bloch factor and a term,  $e^{i\mathbf{k}\cdot\mathbf{r}}\xi(\mathbf{k},z)$ , which is a solution of Eq. (1) with the required boundary conditions and which acts as a modulation of the periodic, high-frequency component of the wave function:

$$\phi_{\mathbf{k}}(\mathbf{r},z) \simeq e^{i\mathbf{k}\cdot\mathbf{r}}\xi(\mathbf{k},z) \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{R}} f_{\mathbf{G}}(\mathbf{k}_0+\mathbf{k},k_{0,z}), \quad (7)$$

where

$$\xi_{\mu}(\mathbf{k},z) = \sum_n a_n^{(\mu)}(\mathbf{k}) \sin \left[ \frac{m\pi z}{z_0} \right]. \quad (8)$$

The factorization (7) was the original intent of the effective-mass approximation. Unfortunately, when abandoning the parabolic approximation in favor of a more realistic band structure, we lose some simplicity. Most notably, the “envelope” wave functions  $\xi$  depend on  $\mathbf{k}$  since we cannot separate the dispersion  $\varepsilon(\mathbf{k},k_z)$  into a “normal” and a “parallel” component, as in the parabolic-band approximation. This translates directly into a computational stumbling block from a numerical perspective: An ensemble Monte Carlo simulation would require a solution of Eq. (6) for every particle at every time step. We can consider two approximations which remove this difficulty: The first one consists in ignoring the  $\mathbf{k}$  dependence of  $\xi_{\mu}(\mathbf{k},z)$ . In this case the eigenvalues  $E_{\mu}(\mathbf{k})$  are approximated by

$$E_{\mu}(\mathbf{k}) \simeq E_{\mu,0} + \sum_n |a_n^{(\mu)}|^2 \Delta E_n(\mathbf{k}), \quad (9)$$

where the subscripts 0 indicate that the quantities are evaluated at  $\mathbf{k}=\mathbf{k}_0$  (i.e., at the valley minimum), and

$$\Delta E_n(\mathbf{k}) = \frac{1}{2} \left[ \varepsilon \left[ \mathbf{k}, k_{z,0} + \frac{n\pi}{z_0} \right] + \varepsilon \left[ \mathbf{k}, k_{z,0} - \frac{n\pi}{z_0} \right] - \varepsilon \left[ \mathbf{k}_0, k_{z,0} + \frac{n\pi}{z_0} \right] - \varepsilon \left[ \mathbf{k}_0, k_{z,0} - \frac{n\pi}{z_0} \right] \right]. \quad (10)$$

This approximation is exact when  $\Delta E_n(\mathbf{k})$  does not depend on  $n$  (as for parabolic bands), it preserves many features of the exact dispersion (such as the anisotropy of the valley) remarkably well in the case of Si, and it has the main advantage of allowing us to factor out the explicit dependence of wave functions on  $\mathbf{k}$  and to separate the normal and parallel components of the dispersion. Yet, as for the case of the exact solution, the numerical difficulty remains of evaluating density of states (DOS) and scattering rates in a transport model. Therefore we shall employ the even cruder approximation of assuming a bulk dispersion of the form:<sup>44</sup>

$$\varepsilon(\mathbf{K}) \simeq \gamma_b(\mathbf{K}) [1 + \alpha \gamma_b(\mathbf{K})], \quad (11)$$

where  $\alpha$  is the nonparabolicity parameter, and

$$\gamma_b(\mathbf{K}) = \frac{\hbar^2}{2} \left[ \frac{k_1^2}{m_1} + \frac{k_2^2}{m_2} + \frac{k_z^2}{m_z} \right],$$

$m_1$  and  $m_2$  being the effective masses in the two Cartesian directions on the plane of the interface, and  $m_z$  the effective mass along the quantization direction. Using (11), the Schrödinger-like Eq. (1) can be recast as

$$\left\{ -\frac{\hbar^2}{2m_z} \frac{d^2}{dz^2} + \alpha \left[ \frac{\hbar^4}{4m_z^2} \frac{d^4}{dz^4} - \frac{\hbar^2 \gamma_{2D}(\mathbf{k})}{m_z} \frac{d^2}{dz^2} \right] + V(z) \right\} \xi_v(\mathbf{k},z) = E_v(\mathbf{k}) \xi_v(\mathbf{k},z). \quad (12)$$

Here the parabolic “parallel energy”

$$\gamma_{2D}(\mathbf{k}) = \frac{\hbar^2}{2} \left[ \frac{k_1^2}{m_1} + \frac{k_2^2}{m_2} \right],$$

takes different values for each of the two “ladders” of subbands in Si. We can solve Eq. (12) using first-order perturbation theory, in the spirit of Eq. (11), treating the nonparabolic corrections proportional to the parameter  $\alpha$  as a perturbation to the parabolic problem:

$$\left\{ -\frac{\hbar^2}{2m_z} \frac{d^2}{dz^2} + V(z) \right\} \xi_v^{(0)}(z) = E_v^{(0)} \xi_v^{(0)}(z). \quad (13)$$

Therefore the perturbed dispersion in subband  $v$  is given by

$$E_v(\mathbf{k}) \simeq E_v^{(0)} + \alpha \langle (E_v^{(0)} - V)^2 \rangle_v + \gamma_{2D}(\mathbf{k}) [1 + \alpha \gamma_{2D}(\mathbf{k}) + 2\alpha \langle E_v^{(0)} - V \rangle_v], \quad (14)$$

where

$$\langle g \rangle_v \equiv \int_0^{z_0} dz g(z) |\xi_v^{(0)}(z)|^2$$

denotes the expectation value of the function  $g$  in subband  $v$ .

In Fig. 1, we show the dispersion obtained from the pseudopotential method (using about 50 sine waves and the local pseudopotential form factors given by Chelikowsky and Cohen<sup>45</sup>) compared to the conventional parabolic approximation in the case of an inversion layer of a 0.1- $\mu\text{m}$ -long channel Si metal-oxide-semiconductor field-effect transistor (MOSFET) previously studied.<sup>22,46</sup> Note that the solutions in the figure are *not* self-consistently obtained from a simultaneous solution of Eq. (1) and the Poisson equation. This is due to the almost prohibitive numerical task of obtaining the DOS when using the pseudopotential dispersion.<sup>47</sup> This self-consistency will be recovered later in our simulations, using the first-order nonparabolic approximation. The confining potential at the (100)-oriented Si-SiO<sub>2</sub> interface at a location about midway along the channel from the source to the drain contact has been chosen as a typical practical situation encountered in many Si MOSFET's. Figure 1(a) shows the situation in a valley having the heavy, longitudinal mass  $m_z = m_l = 0.91m$ , along the direction normal to the interface. Such valleys are usually referred to as “unprimed” valleys. The opposite situation (transverse

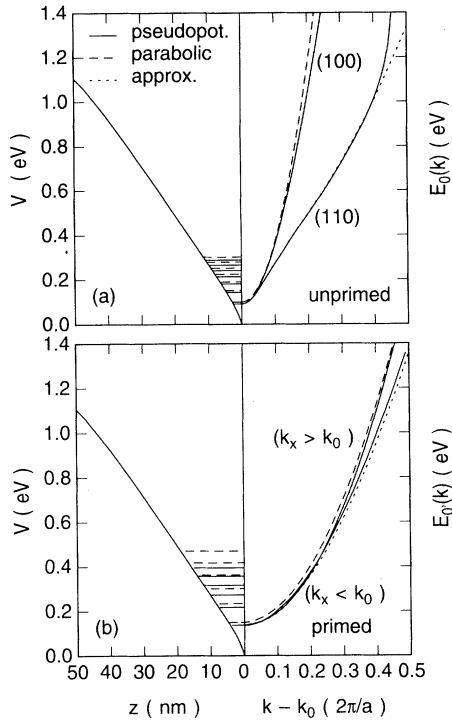


FIG. 1. The confining potential and bottom of the first subbands (left) and the  $E_0(\mathbf{k})$  dispersion of the ground state (right) obtained at midchannel of a  $0.1\text{-}\mu\text{m}$ -long channel of a Si-MOSFET with the pseudopotential, approximated pseudopotential (labeled “approx.” in the figure), and the parabolic model for the conduction bands of Si, as given by Eqs. (6), (9), and (14) of the text, respectively. In (a), the quantization is done along the longitudinal-mass direction (unprimed valleys), in (b) along the transverse-mass direction (primed valleys). In (a), we show the dispersion of the ground state in the (100) and in the (110) directions away from the conduction-band minimum at  $\mathbf{K}_0 = (0, 0, 0.85) (2\pi/a)$ , where  $a$  is the lattice constant. By definition, the parabolic model yields the same dispersion in every direction. Note that the “approximated-pseudopotential” dispersion coincides with the “exact” pseudopotential dispersion along the (100) direction. In (b), we show the dispersion when moving along the (100) direction from the band minimum  $\mathbf{K}_0 = (0.85, 0, 0) (2\pi/a)$  toward the zone center ( $k_x < k_0$ ) and toward and across the symmetry point  $X$  into the second conduction band of the next Brillouin zone ( $k_x > k_0$ ).

mass,  $m_z = m_t = 0.19m$ , along the quantization direction, “primed” valleys) is shown in Fig. 1(b). Note how the parabolic and pseudopotential minima of the subbands differ, particularly in Fig. 1(a). The dispersion in the ground-state subband deviates significantly from the parabolic dispersion at high energies [Fig. 1(a)] and in the transverse (110) direction away from the minimum at  $\mathbf{k}_0$  [Fig. 1(b)]. The dotted line shows the validity of the approximation given by Eq. (9): only for the extremely high energies shown in Fig. 1(a) and along the ( $k_x < k_0$ ) direction [Fig. 1(b)] does the approximation fail visibly; otherwise it is identical to the full pseudopotential treatment.

Note that every correction to the parabolic approximation results in energy levels more closely spaced and dispersions which have a lower slope. This is bound to have strong consequences on electron transport in Si inversion layers resulting in a lower electron velocity.

In Fig. 2, we present a comparison between the “pseudopotential,” the parabolic, and the perturbative nonparabolic approximation for the situation of Fig. 1 restricted to the “unprimed” (low-energy) ground-state subband. A nonparabolicity parameter  $\alpha = -0.5 \text{ eV}^{-1}$  has been used in the latter model. We see that the (spherical) nonparabolic approximation yields satisfactory values for the “subband bottom”  $E_0(\mathbf{k}=0)$  and averages the “exact” behavior in the (100) and (110) directions. Above about  $0.5 \text{ eV}$ , the first-order expansion of Eq. (11) fails and no agreement can be expected.

The first-order nonparabolic approximation (simply called “nonparabolic” in the following) presents two major computational advantages: First, the full translation symmetry is not required any longer, so that interfaces and heterojunctions can in principle be treated by allowing for a space-dependent effective mass and nonparabolicity parameter. Here we will not make use of these extensions, but they are trivially implemented. Second, it allows us to express analytically many quantities needed in the study of transport: For example, the DOS,  $\mathcal{D}_\nu(E)$ , at energy  $E$  in subband  $\nu$  is given by

$$\begin{aligned} \mathcal{D}_\nu(E) &= \theta(E - E_\nu) \frac{g_\nu m_{d,\nu}}{\pi \hbar^2} [1 - 2\alpha(E - E_\nu) \\ &\quad - 2\alpha \langle E_\nu^{(0)} - V \rangle_\nu] \\ &\equiv \mathcal{D}_\nu^{(p)}(E) \kappa_\nu(E - E_\nu), \end{aligned} \quad (15)$$

where  $g_\nu$  is the degeneracy of the subband (2 for the unprimed subbands, 4 for the primed subbands),  $E_\nu \equiv E_\nu(\mathbf{k}=0)$ ,  $m_{d,\nu}$  is the DOS effective mass in subband  $\nu$ ,  $(m_{1,\nu} m_{2,\nu})^{1/2}$ , and  $\theta(x)$  is the step function. We have expressed the nonparabolic DOS as the product of the

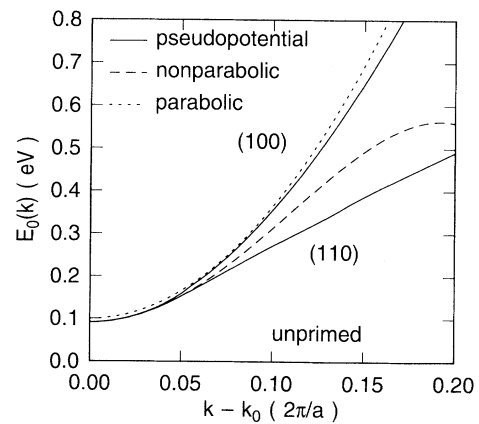


FIG. 2. The dispersion in the ground-state subband for the situation of Fig. 1(a), as obtained from the pseudopotential, the parabolic, and the nonparabolic approximations. The latter two are spherically symmetric and therefore apply to both the (100) and (110) directions.

parabolic expression  $\mathcal{D}^{(p)}$  and a correction factor for the subband  $\nu$ ,

$$\kappa_\nu(E) = 1 - 2\alpha E - 2\alpha \langle E_\nu^{(0)} - V \rangle_\nu,$$

which we will employ in the following. Similarly, the velocity of an electron in subband  $\nu$  with wave vector  $\mathbf{k}$  is

$$\mathbf{v}(\mathbf{k}) = \mathbf{v}^{(0)}(\mathbf{k}) [1 + 2\alpha \gamma_{2D}(\mathbf{k}) + 2\alpha \langle E^{(0)} - V \rangle_\nu], \quad (16)$$

where  $\mathbf{v}^{(0)}(\mathbf{k})$  is the parabolic velocity  $\nabla \gamma_{2D}(\mathbf{k})/\hbar$ . Finally, the equilibrium electron sheet density at temperature  $T$  as a function of Fermi level  $E_F$  is

$$\begin{aligned} n_s(E_F, T) &\equiv \sum_\mu n_\mu(E_F, T) \\ &= \frac{k_B T}{\pi \hbar^2} \sum_\mu g_\mu m_{d,\mu} \left[ \kappa_\mu(E_F - E_\mu) \ln(1 + e^{-\eta_\mu}) \right. \\ &\quad \left. - 2\alpha k_B T \int_{\eta_\mu}^\infty dx \frac{x dx}{(1 + e^x)} \right], \end{aligned} \quad (17)$$

where  $k_B$  is the Boltzmann constant, and  $\eta_\mu = (E_\mu - E_F)/(k_B T)$ . This relationship between  $n_s$  and  $E_F$  is needed to account for degeneracy effects in the Monte Carlo simulations, which we have implemented as described in Ref. 23.

Having finally settled for the nonparabolic model (12), the self-consistent solution is obtained in a “conventional” way: The external potential energy is given by the sum of various contributions:<sup>1</sup>

$$V(z) = V_d(z) + V_s(z) + V_i(z) + V_{xc}(z), \quad (18)$$

expressing the potential due to the ionized impurities in the depletion layer, to the induced charges in the space-charge layer, to the image charges at the semiconductor-insulator interface, and to the exchange and correlation effects.<sup>48–55</sup> We have written this last term as a simple local function of  $z$ , although in general this is not the case,  $V_{xc}$  being usually expressed as a functional of the electron density. In the present work, we have considered only the first two terms, which are obtained from the Poisson equation

$$\frac{d^2}{dz^2} [V_d(z) + V_s(z)] = \frac{e}{\epsilon_{sc}} \left[ \rho_d(z) - e \sum_\mu n_\mu |\xi_\mu(z)|^2 \right], \quad (19)$$

where  $e$  is the magnitude of the electron charge, and  $n_\mu$  is the occupation of subband  $\mu$ , given either by the corresponding term in Eq. (17) at equilibrium, or by a direct “particle counting” in a high-field Monte Carlo simulation. The eigenfunctions  $\xi_\mu$  are the solutions of Eq. (13), which we will use in the following, having dropped the superscript (0) for simplicity. Also,  $\epsilon_{sc}$  is the semiconductor permittivity, and  $\rho_d(z)$  the charge of the ionized impurities, given by the difference between the ionized acceptor and the donor charge densities,  $e(N_D^+ - N_A^-)$ , in the depletion region of thickness  $z_d$ , vanishing otherwise.

As shown by Stern,<sup>56</sup> in the extreme quantum limit (low temperature and only the bottom subband populated), the image potential approximately cancels the many-body corrections given by the exchange and correlation term above. At high temperature, many-body effects have been considered by Das Sarma and Vinter<sup>54,55</sup> both in a perturbative approach as well as with local-density-functional calculation. In the general situation in which we are interested (high temperature, high source-drain fields resulting in off-equilibrium electron population in many subbands), we found these rigorous approaches computationally too expensive and have decided to ignore both exchange-correlation and image potential effects. Details about the numerical procedure followed to solve Eqs. (12) and (19) together with the Boltzmann transport equation in a Monte Carlo context will be presented below, in Sec. III.

### B. Bulk phonon-electron interaction

We have considered the interaction between electrons and bulk phonons in a rather conventional way, as described, for instance, in the reviews by Price<sup>57</sup> and Ridley,<sup>58</sup> considering both intravalley and intervalley nonpolar processes. In this subsection we shall give the relevant equations and the parameters we have employed. Yet we must stress one major exception we took from the conventional approach, concerning the anisotropy of the deformation-potential interaction between electrons and acoustic phonons.

Herring and Vogt<sup>59</sup> have analyzed the anisotropy of the intravalley deformation potential in the ellipsoidal valleys in Si. Expanding the electron-phonon matrix elements over spherical harmonics and retaining only the leading terms, they have expressed the anisotropy of the interaction in terms of the angle  $\theta_Q$  between the wave vector  $\mathbf{Q}$  of the emitted/absorbed phonon and the longitudinal axis of the valley. They have shown that the matrix element is proportional to  $Q$  via the deformation potential  $\Delta_i(\theta_Q)$  ( $i = \text{LA}$  or  $\text{TA}$ ) given by

$$\Delta_{\text{LA}}(\theta_Q) \simeq \Xi_d + \Xi_u \cos^2(\theta_Q), \quad (20a)$$

$$\Delta_{\text{TA}}(\theta_Q) \simeq \Xi_u \cos(\theta_Q) \sin(\theta_Q), \quad (20b)$$

where the labels TA and LA denote transverse and longitudinal acoustic phonons, respectively. Note that Eq. (20b) accounts for the contributions of both TA branches.  $\Xi_u$  and  $\Xi_d$  are the uniaxial-shear and dilation deformation potentials, respectively. In bulk Si, this anisotropy is usually ignored by using an effective deformation potential  $\Xi_{\text{LA}}^{\text{eff}}$ , for the interaction with longitudinal modes and ignoring the role of the lower-energy TA modes. This approximation is partially justified for the following reason: Acoustic modes are most effective at low energy. In this regime and in the usual elastic and equipartition approximations, thanks to the linear dependence on  $Q$ , scattering of electrons at energy  $E$  samples almost uniformly the equienergy ellipsoid  $\epsilon(\mathbf{K}) = E$ . Therefore, one can take the average values of  $\Delta_i$  over the ellipsoid. Since there is nothing to fix an energy scale in the problem, this averaging procedure is independent of the electron energy. To

make it explicit, using Eq. (20) and ignoring (for now) nonparabolicity, the intravalley electron-TA/LA photon-scattering rate  $1/\tau_i(E)$  in bulk Si can be written as

$$\frac{1}{\tau_i(E, \beta)} = \frac{\sqrt{2}k_B T m_d^{3/2} E^{1/2}}{8\pi^2 \hbar^4 \rho c_i^2} \int_0^\pi d\beta' \sin\beta' \int_0^{2\pi} d\gamma' \Delta_i^2(\theta_Q), \quad (21)$$

for either emission or absorption, having aligned the longitudinal axis of the valley with the  $z$  axis and having expressed the initial and final electron wave vectors  $\mathbf{K}$  and  $\mathbf{K}'$  in terms of the polar angles  $\beta$  and  $\beta'$ , and of the azimuthal angles  $\gamma$  and  $\gamma'$ . In Eq. (21),  $c_i$  is the transverse/longitudinal sound velocity,  $\rho$  the Si volume density,  $m_d$  the DOS effective mass in the given valley, and the phonon wave vector  $\mathbf{Q}$  is simply  $\mathbf{K}' - \mathbf{K}$ . Using energy and momentum conservation, the angle  $\theta_Q$  can be obtained in terms of the quite awkward expression (setting  $\gamma = 0$  without loss of generality)

$$\begin{aligned} \cos(\theta_Q) = & [K(\beta)\cos\beta - K(\beta')\cos\beta'] \\ & \times \{ [K(\beta)\sin\beta - K(\beta')\sin\beta'\cos\gamma']^2 \\ & + K(\beta')^2 \sin^2\beta' \sin^2\gamma' \\ & + [K(\beta)\cos\beta - K(\beta')\cos\beta']^2 \}^{-1/2}, \end{aligned}$$

where

$$\frac{1}{\tau_{i;\mu}(\mathbf{k})} = \frac{k_B T}{4\pi^2 \hbar^3 \rho c_i^2} \sum_{\nu} \theta[E_{\mu}(\mathbf{k}) - E_{\nu}] \kappa_{\nu}[E_{\mu}(\mathbf{k}) - E_{\nu}] m_{d,\nu} \int_0^{2\pi} d\beta' \int_{-\infty}^{\infty} dq_z |\mathcal{F}_{\mu\nu}(q_z)|^2 \Delta_i(\theta_Q)^2, \quad (23)$$

where once again  $\beta$  and  $\beta'$  are the polar angles of the initial and final wave vectors, respectively, on the plane of the interface, and  $\mathcal{F}_{\mu\nu}$  is the electronic form factor

$$\mathcal{F}_{\mu\nu}(q_z) = \int_0^{\infty} dz \xi_{\mu}(z) e^{iq_z z} \xi_{\nu}(z). \quad (24)$$

The angle  $\theta_Q$  is given by another awkward but trivial set of expressions which we quote here, since they will be used repeatedly in the following:

$$\cos(\theta_Q) = \frac{Q_l}{(q^2 + q_z^2)^{1/2}}, \quad (25)$$

where  $Q_l$  is the component of the phonon wave vector along the longitudinal axis of each valley. The magnitude  $q$  of the component of  $\mathbf{Q}$  on the plane of the semiconductor-insulator interface can be expressed in terms of the electron energy using energy and momentum conservation:

$$q^2 = k^2 - 2k'k \cos(\beta - \beta') + k'^2, \quad (26)$$

where  $k'$  is the magnitude of the final electron wave vector which can be written as

$$k' = \frac{[2(E - E_{\nu})\bar{\kappa}_{\nu}(E - E_{\nu})]^{1/2}}{\hbar} \left[ \frac{\cos^2\beta'}{m_{1,\nu}} + \frac{\sin^2\beta'}{m_{2,\nu}} \right]^{-1/2}, \quad (27)$$

$$K(\beta) = \left[ \frac{\cos^2\beta}{m_l} + \frac{\sin^2\beta}{m_t} \right]^{-1/2}.$$

Comparing Eq. (21) with the usual expression obtained from an isotropic model, we can define the effective deformation potential by averaging over the initial polar angle  $\beta$ :

$$(\Xi_i^{\text{eff}})^2 \equiv \frac{1}{8\pi} \int_0^\pi d\beta \sin\beta \int_0^\pi d\beta' \sin\beta' \int_0^{2\pi} d\gamma' \Delta_i^2(\theta_Q). \quad (22)$$

The equation above shows explicitly that the electron energy  $E$  does not enter the definition of the effective deformation potential. Using the values  $\Xi_d = -11.7$  eV and  $\Xi_u = 9.0$  eV, as we shall discuss below, we find that the averaged effective values are 9.9 (LA) and 5.3 eV (TA). The first value is close to that typically employed in the Monte Carlo literature.<sup>60</sup>

Moving to the two-dimensional situation, we find that we cannot follow a parallel path to arrive at an isotropic, energy-independent effective deformation potential. This is bound to have a profound effect on the strength of the electron-TA/LA-phonon interaction and affect negatively our mobility results. Using once more the elastic, equipartition approximation suitable to the high-temperature simulations of interest here, and reintroducing nonparabolicity, we can write the rate for an electron with parallel wave vector  $\mathbf{k}$  in subband  $\mu$  to scatter into any other subband  $\nu$  by emission or absorption of a TA/LA phonon as

where

$$\bar{\kappa}_{\nu}(E) = 1 - \alpha E - 2\alpha \langle E_{\nu}^{(0)} - V \rangle_{\nu}.$$

We can recover the “usual” isotropic expression by ignoring the dependence on  $\theta_Q$  in Eq. (23), replacing  $\Delta_i(\theta_Q)$  with  $\Xi_i^{\text{eff}}$ , obtaining

$$\begin{aligned} \frac{1}{\tau_{i;\mu}(\mathbf{k})} \simeq & \frac{k_B T}{2\hbar^3 \rho c_i^2} \sum_{\nu} m_{d,\nu} (\Xi_i^{\text{eff}})^2 \theta[E_{\mu}(\mathbf{k}) - E_{\nu}] \\ & \times \kappa_{\nu}[E_{\mu}(\mathbf{k}) - E_{\nu}] F_{\mu\nu}, \end{aligned} \quad (28)$$

where  $F_{\mu\nu}$  is the “usual” form factor

$$F_{\mu\nu} = \int_0^{\infty} dz \xi_{\mu}(z)^2 \xi_{\nu}(z)^2 \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} dq_z |\mathcal{F}_{\mu\nu}(q_z)|^2. \quad (29)$$

Comparing Eq. (28) with Eq. (23), we see that if we want to define the equivalent of the effective averaged deformation potential given by Eq. (22) for the bulk case, this quantity must now depend explicitly on the pair of initial and final subbands  $\mu$  and  $\nu$ . Averaging as before over the initial polar angle  $\beta$ , it is expressed as

$$(\Xi_{i;\mu\nu}^{\text{eff}})^2 \equiv \frac{1}{(2\pi)^3 F_{\mu\nu}} \int_0^{2\pi} d\beta \int_0^{2\pi} d\beta' \int_{-\infty}^{\infty} dq_z \Delta_i^2(\theta_Q) |\mathcal{F}_{\mu\nu}(q_z)|^2. \quad (30)$$

The angle  $\theta_Q$  is again given by Eq. (24). Explicitly, for the unprimed valleys we have

$$\cos(\theta_Q) = \left[ \frac{q_z}{k} \right] [(\cos\beta - \cos\beta')^2 + (\sin\beta - \sin\beta')^2 + (q_z/k)^2]^{-1/2},$$

which depends on  $\beta - \beta'$ , by symmetry, so that the integration over either  $\beta$  or  $\beta'$  can be made trivial. For the primed valleys we have instead

$$\begin{aligned} \cos(\theta_Q) = & [k(\beta)\cos\beta - k(\beta')\cos\beta'] \\ & \times \{ [k(\beta)\cos\beta - k(\beta')\cos\beta']^2 \\ & + [k(\beta)\sin\beta - k(\beta')\sin\beta']^2 + q_z^2 \}^{-1/2}, \end{aligned}$$

where the function  $k(\beta')$  is given by Eq. (27). The main observation is that, due to the presence of  $q_z$  in the equations above,  $\theta_Q$ , and, therefore,  $\Xi^{\text{eff}}$ , depend on the electron energy. The reason is that the form factor  $\mathcal{F}_{\mu\nu}(q_z)$  introduces an energy scale in the problem by fixing the “fuzzy” component, the phonon wave vector  $q_z$ . For instance, in the unprimed valleys, at low electron energy the average phonon wave vector will be mostly aligned

with the longitudinal axis of the ellipsoid, so that  $\Xi_{\text{LA}}^{\text{eff}} \approx |\Xi_d + \Xi_\mu|$  and  $\Xi_{\text{TA}}^{\text{eff}} \approx 0$ . As the electron energy increases, the phonon wave vector rotates toward the plane normal to the longitudinal axis, so that in this case  $\Xi_{\text{LA}}^{\text{eff}} \approx \Xi_d$  while, as before,  $\Xi_{\text{TA}}^{\text{eff}} \approx 0$ . The coupling with TA modes will peak at some intermediate energy, maximizing the angular average of Eq. (20b). This is illustrated in Fig. 3. We have plotted the effective deformation potentials in the lowest-lying unprimed [Fig. 3(a)] and primed [Fig. 3(b)] subbands. Note the strong energy dependence introduced by the anisotropy of the interaction. Its effect on the scattering rates can be seen in Fig. 4, which shows the total electron-phonon-scattering rates in the device described below, for electrons in the lowest unprimed subband, at low density ( $\approx 2.6 \times 10^{11} \text{ cm}^{-2}$ ) for parabolic bands. The anisotropic correction depresses remarkably the scattering rate at low electron energy. This is bound to have a profound effect on the electron mobility in the medium-density range, at high temperature, where scattering with acoustic phonons dominates the picture. In Fig. 5, we show the effect of the nonparabolic corrections to the total electron-phonon-scattering rates in the same situation.

The last issue related to scattering with acoustic phonons is the choice of the values for the deformation potentials  $\Xi_u$  and  $\Xi_d$ . The available experimental data are,

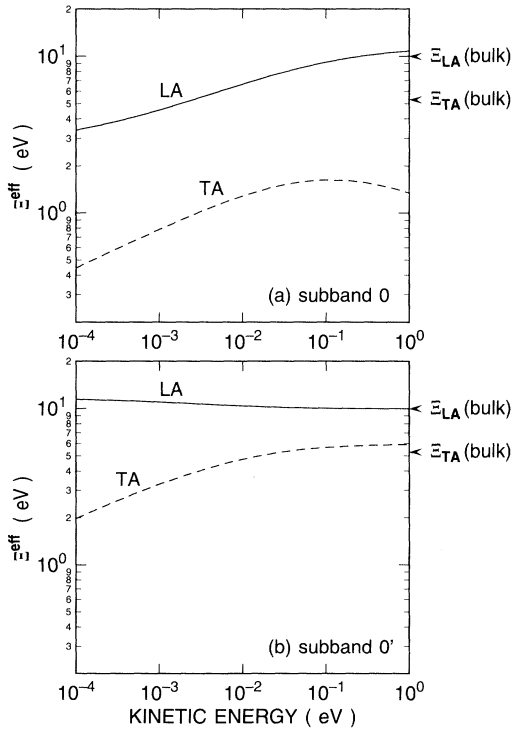


FIG. 3. Effective intravalley deformation potentials for intrasubband scattering with acoustic phonons in the lowest-lying unprimed (a) and primed (b) subbands. A parabolic model is employed in this figure. The “effective” average deformation potentials in bulk Si are also shown.

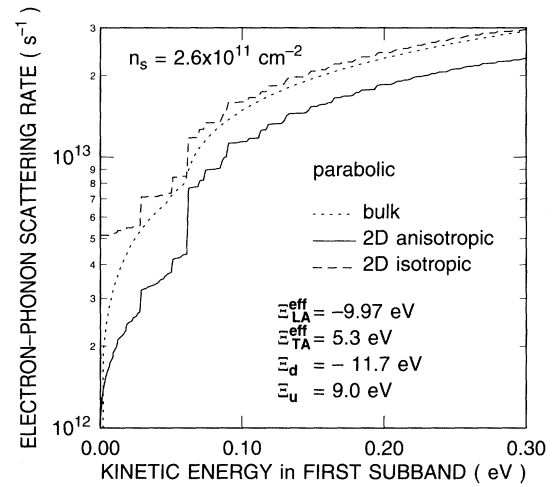


FIG. 4. The total electron-bulk-phonon-scattering rate at 300 K for electrons in the bottom unprimed subband using a parabolic-band structure in the channel of the device described in Sec. III of the text, at an electron density of about  $2.6 \times 10^{11} \text{ cm}^{-2}$ . The effect of the anisotropy of the intravalley scattering with the acoustic phonon is very pronounced at low energy. The “steps” are due to the thresholds of emission and absorption processes, as well as to the appearance of additional subbands. 40 subbands have been included in the calculation. The scattering rates in bulk Si are also shown for comparison. Note that the anisotropic scattering rate does not go to zero as the kinetic energy goes to zero, but approaches the value of about  $6 \times 10^{11} \text{ s}^{-1}$ .



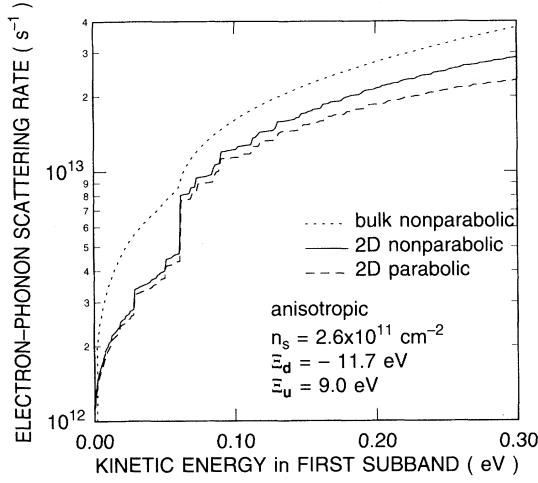


FIG. 5. The calculated total electron—bulk-phonon-scattering rate at 300 K in the situation of Fig. 4, using parabolic and nonparabolic models for the band structure.

unfortunately, not fully consistent among themselves. The uniaxial-shear deformation potential is probably the most well known, with values of  $\Xi_u \approx 8.6$ – $9.2$  eV having been reported.<sup>61–64</sup> The value of  $\Xi_d$  is less certain from an experimental point of view. However, we can rely on the several constraints which  $\Xi_d$  must satisfy. Among them,  $\Xi_d$  must be consistent with the value for the deformation potential  $E_1 = \Xi_d + \Xi_u/3$  at the  $X$  point of the first conduction band. From the point of view of transport, for a fixed value of  $\Xi_u$ , the ratio  $D_\Xi = \Xi_d/\Xi_u$  determines the phonon-limited bulk electron mobility.<sup>59</sup> Rescaling the Herring-Vogt results for the mobility at 100 K to account for the larger value of  $\Xi_u = 9$  eV we adopt, we find that a ratio<sup>65</sup>  $D_\Xi$  of about either  $-1.3$  or  $0.8$  yields the correct phonon-limited electron mobility of about  $12\,000\text{ cm}^2/\text{Vs}$  at 100 K (Ref. 60). This implies that  $\Xi_d$  must take a value of either  $-11.7$  or  $7.2$  eV. Only the former is consistent with measurements of  $E_1$  yielding  $|E_1| = 8.7$  eV (Ref. 66). In addition, theoretical estimates, based on the rigid pseudoion model,<sup>67</sup> provide negative values for the ratio  $D_\Xi$  and values for the uniaxial-shear and dilation potentials not too far from those we select here. Therefore, we shall choose the values of  $-11.7$  and  $9.0$  eV for  $\Xi_d$  and  $\Xi_u$ , respectively. However, we should keep in mind that uncertainty still remains. Since the electron mobility depends on the square of these parameters, it is clear that small errors on the deformation potentials can result in large errors in the final results. Note that previous workers<sup>13–16,28,68–72</sup> have almost always ignored the TA phonons and selected large, isotropic values for the effective potential (typically 12 eV) for scattering with LA modes.<sup>9,10,28</sup>

Finally, nonpolar intervalley scattering can be treated in a more conventional fashion. (Note that intravalley scattering with optical phonons vanishes in Si for electrons sufficiently close to the bottom of the  $s$ -type valley minima along the  $\Delta$  lines, and we shall ignore this process.) These processes are treated assuming bulk phonon dispersions and deformation potentials.<sup>24</sup> Table I lists,

among others, the  $f$ - and  $g$ -scattering parameters we have chosen. Labeling these processes by the index  $r$ , the rate at which an electron with wave vector  $\mathbf{k}$  in subband  $\mu$  can scatter into subband  $\nu$  by emitting or absorbing a phonon of energy  $\hbar\omega_r$  can be expressed as

$$\frac{1}{\tau_{r,\mu\nu}^{(iv)}(\mathbf{k})} = \frac{(\Delta_r^{(iv)})^2 m_{d,\nu} g_{\mu\nu}^{(r)}}{2\hbar^2 \rho \omega_r} \kappa_\nu (E_f - E_\nu) \theta(E_f - E_\nu) \times \left(\frac{1}{2} \mp \frac{1}{2} + n_r\right) F_{\mu\nu}, \quad (31)$$

where  $E_f$  is the final electron energy  $E_f = E_\mu(\mathbf{k}) \pm \hbar\omega_r$  (the upper/lower sign is for absorption/emission of a phonon),  $\Delta_r^{(iv)}$  is the deformation potential for the  $r$ th transition,  $n_r$  is the thermal (Bose) population of phonons of type  $r$ , and the form factor  $F_{\mu\nu}$  is given by Eq. (29). The degeneracy factor  $g_{\mu\nu}^{(r)}$  expresses the multiplicity of the final state for the  $r$ th process and can vanish for particular transitions, as dictated by symmetry: Only  $g$  processes ( $g_{\mu\nu}^{(r)} = 1$ ) are available for transitions between unprimed subbands, while both  $g$  ( $g_{\mu\nu}^{(r)} = 1$ ) and  $f$  processes ( $g_{\mu\nu}^{(r)} = 2$ ) are available for transitions between primed subbands. Transitions from an unprimed to a primed subband are assisted by  $f$  processes only ( $g_{\mu\nu}^{(r)} = 4$ ), and, finally, transitions from primed to unprimed subbands are also assisted by  $f$  processes only, but with  $g_{\mu\nu}^{(r)} = 2$ .

Note that in Eq. (23), as well as in (31), it is not particularly computationally expensive to consider intravalley, intersubband transitions, as it amounts only to additional bookkeeping in an already sizable computer program. Therefore, we have retained all possible transitions from any subband  $\mu$  to any other subband  $\nu$ , as in Eqs. (23) and (31). As a rule, however, intravalley intersubband processes are strongly depressed by the small form factors and do not represent a significant contribution.

### C. Scattering with interface modes

The presence of the Si-SiO<sub>2</sub> interface undoubtedly alters the dispersion of the phonons, their nature, and their coupling to the electrons. The nature of surface acoustic modes has been studied by Ezawa and co-workers.<sup>9,10</sup> As clearly outlined in Ref. 1, they have idealized the insulator as an infinitely soft medium, so that the eigenmodes could be obtained assuming a stress-free boundary. Several independent modes, called “surfons,” have been found, one of them corresponding to the well-known Rayleigh wave. The results of the monumental efforts by Ezawa and co-workers<sup>9,10</sup> are somewhat disappointing. Compared to bulk acoustic modes, the surfon-limited electron mobility is lowered by 20% or less in the extreme quantum limit. We should keep in mind that SiO<sub>2</sub> is not as soft as assumed in Refs. 9 and 10, so that the difference between bulk and surface acoustic modes probably has been overestimated. Moreover, at room temperature, intervalley processes play a significant role and intravalley acoustic-phonon scattering matters less, particularly at low densities. Therefore the difference between bulk-phonon-limited and surfon-limited mobility should be even smaller than 20%. For these reasons—and for the difficulty of accounting correctly for the nonzero stress of the interface, beyond what was done by

TABLE I. Semiconductor and insulator parameters.

Symbol	Quantity	Value	Unit
<b>Bulk Si:</b>			
$m_t$	transverse effective mass	0.19	m
$m_l$	longitudinal effective mass	0.91	m
$\alpha$	nonparabolicity parameter	-0.5	eV <sup>-1</sup>
$\Xi_u$	uniaxial-shear def. pot.	9.0	eV
$\Xi_d$	uniform dilation def. pot.	-11.7	eV
$\rho$	density	2.33	g/cm <sup>3</sup>
$c_L$	longitudinal sound velocity	9.2	10 <sup>5</sup> cm/s
$c_T$	transverse sound velocity	4.7	10 <sup>5</sup> cm/s
$\epsilon_{sc}$	static permittivity	11.7	$\epsilon_0$
<b>Intervalley electron-phonon scattering parameters:</b>			
$\Delta_1^{(iv)}$	g-scattering, TA def. pot.	0.5	10 <sup>8</sup> eV/cm
$\Delta_2^{(iv)}$	g-scattering, LA def. pot.	0.8	10 <sup>8</sup> eV/cm
$\Delta_3^{(iv)}$	g-scattering, LO def. pot.	11.0	10 <sup>8</sup> eV/cm
$\Delta_4^{(iv)}$	f-scattering, TA def. pot.	0.3	10 <sup>8</sup> eV/cm
$\Delta_5^{(iv)}$	f-scattering, LA def. pot.	2.0	10 <sup>8</sup> eV/cm
$\Delta_6^{(iv)}$	f-scattering, TO def. pot.	2.0	10 <sup>8</sup> eV/cm
$\hbar\omega_1$	g-scattering, TA phonon energy	12	meV
$\hbar\omega_2$	g-scattering, LA phonon energy	18.5	meV
$\hbar\omega_3$	g-scattering, LO phonon energy	61.2	meV
$\hbar\omega_4$	f-scattering, TA phonon energy	19.0	meV
$\hbar\omega_5$	f-scattering, LA phonon energy	47.4	meV
$\hbar\omega_6$	f-scattering, TO phonon energy	59.0	meV
<b>Bulk SiO<sub>2</sub>:</b>			
$\epsilon_0^{(ox)}$	static permittivity	3.9	$\epsilon_0$
$\epsilon_i^{(ox)}$	intermediate permittivity	3.05	$\epsilon_0$
$\epsilon_\infty^{(ox)}$	high-frequency permittivity	2.5	$\epsilon_0$
$\hbar\omega_{TO,1}$	bulk TO-phonon, low energy	55.6	meV
$\hbar\omega_{TO,2}$	bulk TO-phonon, high energy	138.1	meV
<b>Interface roughness</b>			
$\Lambda$	step correlation length	1.3	nm
$\Delta$	step rms height	0.48	nm

Ezawa—we have ignored the presence of the interface in dealing with acoustic modes. Intervalley processes are assisted by short-wavelength phonons, so that a bulk picture seems appropriate. We are left with the possibility of electrons scattering with the fringing fields of polar modes in SiO<sub>2</sub>. This possibility was first considered by Hess and Vogl,<sup>73</sup> later studied by Moore and Ferry.<sup>74,75</sup> Once more, the effect of the surface optical (SO) modes was found to be small: On the one hand, the large energy of the modes prevents a strong coupling with low-energy electrons, so that the mobility is largely unaffected by them. On the other hand, the scattering rate with SO modes decreases with electron energy, although not as fast as bulk polar scattering. Therefore, their effect on the high-field properties of transport is relatively small. Despite these considerations, in our attempts to find ways to improve the agreement between experimental data and results of our simulations, we have investigated in some detail the optical interface modes and their coupling to interface electromagnetic waves (polaritons) and two-dimensional plasma excitations of the 2DEG. The com-

plicated nature of these coupled modes has the effect of reducing the importance of the SO modes even beyond the already small corrections found by Moore and Ferry,<sup>74,75</sup> although, eventually, we will end up treating these processes in a way very similar to Refs. 73–75.

The main motivation for reconsidering the results of Wang and Mahan,<sup>76</sup> of Hess and Vogl,<sup>73</sup> and of Moore and Ferry,<sup>74,75</sup> is that surface polar modes may couple strongly to plasmons at the high densities usually present in Si inversion layers. This coupling is well known in the context of III-V compound semiconductors,<sup>77–81</sup> and it is reason for concern. Scattering with plasmons—or modes mixed to plasmons to some extent—is difficult to treat in a transport model. Plasmons are not excitations “external” to the electron system, as phonons are. The momentum and energy transferred from single particles to collective electronic excitations are not lost, but mainly “redistributed” among the electrons. Only when plasmons are quickly dissipated via scattering with “external” excitations (e.g., phonons or impurities) can they be treated as effective in removing energy and

momentum from the carriers. In heavily doped bulk Si, this is probably the case, as discussed by one of us before.<sup>82</sup> On the contrary, this is not generally true for low-effective-mass, polar materials, as shown by Sirko and Mills.<sup>83</sup> Care must be taken to understand whether plasmons feed back their energy and momentum to the electron system before claiming that plasmons affect the momentum and energy relaxation rates.<sup>84,85</sup> In the context of Si inversion layers and SO modes, we have to face the same problem and we want to understand over which range of wavelengths of SO modes we can ignore this difficult issue.

In order to study the optical interface modes, we draw from the large literature on interface modes (plasmons,<sup>86–90</sup> phonons,<sup>91–95</sup> and mixed modes such as plasmarons<sup>96</sup> and other excitations<sup>97</sup>) subject to electromagnetic boundary conditions,<sup>98</sup> as done originally by Fuchs and Kliever.<sup>91–93</sup> We consider two ideal systems of media characterized by a frequency-dependent dielectric constant (possibly nonlocal, as is the case for the 2DEG). We take Fourier components of Maxwell's equations and solve them, subject to the boundary conditions of continuity across interfaces of the component of the electric fields parallel to the interfaces,  $\mathbf{E}_{\parallel}(\omega)$ , and of the electric displacement fields normal to them,  $\mathbf{D}_{\perp}(\omega)$ . With this procedure we obtain a secular equation yielding, for a given radian frequency  $\omega$ , the eigenmodes of the system; that is, the dispersion of the interface excitations. The first system consists simply of two semi-infinite media, an insulator with permittivity  $\epsilon_{\text{ox}}(\omega)$  for  $z < 0$ , and a semiconductor with a nonlocal permittivity corresponding to the 2DEG dielectric response given by Dahl and Sham,<sup>99</sup>  $\epsilon_{2\text{D}}(\omega; z, z')$ , for  $z \geq 0$ . This geometry approximates the situation we expect for very thick oxides. We shall assume that two transverse optical modes exist in the oxide, with energies  $\hbar\omega_{\text{TO},i}$  ( $i=1,2$ ), with  $\omega_{\text{TO},2} \gg \omega_{\text{TO},1}$  so that<sup>100</sup>

$$\epsilon_{\text{ox}}(\omega) = \epsilon_{\infty}^{(\text{ox})} + [\epsilon_i^{(\text{ox})} - \epsilon_{\infty}^{(\text{ox})}] \frac{\omega_{\text{TO},2}^2}{\omega_{\text{TO},2}^2 - \omega^2} + [\epsilon_0^{(\text{ox})} - \epsilon_i^{(\text{ox})}] \frac{\omega_{\text{TO},1}^2}{\omega_{\text{TO},1}^2 - \omega^2}, \quad (32)$$

where  $\epsilon_{\infty}^{(\text{ox})}$  and  $\epsilon_0^{(\text{ox})}$  are the high-frequency and static permittivities, respectively, while  $\epsilon_i^{(\text{ox})}$  is the permittivity at “intermediate” frequencies. For the dielectric function of the 2DEG, we shall follow Dahl and Sham and consider the long-wavelength response in the extreme quantum limit. These approximations will be relaxed below considering screening of the Coulomb scattering processes. For now, in view of the illustrative nature of the present analysis, the qualitative behavior of the results should not depend on these restrictive assumptions. Therefore, we have

$$\epsilon_{2\text{D}}(\omega; z, z') = \epsilon_{\text{sc}} \delta(z - z') - \frac{e^2 n_s}{m_{\parallel}} |\xi_0(z)|^2 |\xi_0(z')|^2 \frac{1}{\omega^2}, \quad (33)$$

where  $m_{\parallel}$  is the (conductivity) mass on the plane of the

interface ( $=m_i$  when only the lowest-lying subband is considered). In the absence of the insulating layer, the collective excitations of the 2DEG, as obtained from Eq. (33), have the dispersion  $\omega_p(q)$  given implicitly by

$$\omega_p^2 = \frac{e^2 n_s}{\epsilon_{\text{sc}} m_{\parallel}} \left[ q^2 - \frac{\omega_p^2}{c^2} \frac{\epsilon_{\text{sc}}}{\epsilon_0} \right] G_{q,00}, \quad (34)$$

where  $c$  is the speed of light,  $\epsilon_0$  is the permittivity of free space, and

$$G_{q,00} = \int_0^{\infty} dz \int_0^{\infty} dz' |\xi_0(z)|^2 G_q(z, z') |\xi_0(z')|^2. \quad (35)$$

The function  $G_q$  is the Green's function satisfying

$$\left[ \frac{d^2}{dz^2} - \left[ q^2 - \frac{\omega^2}{c^2} \frac{\epsilon_{\text{sc}}}{\epsilon_0} \right] \right] G_q(z, z') = \delta(z - z') \quad \text{in the Si channel}, \quad (36a)$$

$$\left[ \frac{d^2}{dz^2} - \left[ q^2 - \frac{\omega^2}{c^2} \frac{\epsilon(\omega)}{\epsilon_0} \right] \right] G_q(z, z') = 0 \quad \text{otherwise}, \quad (36b)$$

with the additional boundary conditions fixed by the continuity of fields across the interface(s), as mentioned above. The solution of Eq. (36) for the geometry under consideration (and for other cases) has been given by Dahl and Sham. In the no-retardation approximation (i.e., ignoring the thickness of the inversion charge layer compared to the wavelength of the excitations), the dispersions of the interface excitations for this geometry are given by the solutions of the secular equation:<sup>101</sup>

$$K_{\text{ox}} \epsilon_{\text{sc}} + K_{\text{sc}} \epsilon^{(\text{ox})}(\omega) = \epsilon_0 \frac{\Omega_{\parallel}^2 K_{\text{ox}} K_{\text{sc}}}{\omega^2}, \quad (37)$$

where

$$K_{\text{ox}} = \left[ q^2 - \frac{\omega^2}{c^2} \frac{\epsilon^{(\text{ox})}(\omega)}{\epsilon_0} \right]^{1/2},$$

$$K_{\text{sc}} = \left[ q^2 - \frac{\omega^2}{c^2} \frac{\epsilon_{\text{sc}}}{\epsilon_0} \right]^{1/2},$$

and  $\Omega_{\parallel}^2 = e^2 n_s / (\epsilon_0 m_{\parallel})$ . Figure 6 illustrates the dispersions of the eigenmodes given by Eq. (37) with the parameters given in Table I for the case of the lowest-lying unprimed subband populated at two values of the electron density  $n_s$ . At low  $q$  the three upper branches coincide with the dispersion given by Wang and Mahan in the case of only one TO mode in the insulator: As  $q \rightarrow 0$ , two branches,  $\hbar\omega_{+,1}$  and  $\hbar\omega_{+,2}$ , approach the energy of the LO phonons in bulk  $\text{SiO}_2$ , a high-frequency mode of energy<sup>100</sup>  $\hbar\omega_{\text{LO},2} = \sqrt{\epsilon_i^{(\text{ox})}/\epsilon_{\infty}^{(\text{ox})}} \hbar\omega_{\text{TO},2}$ , and a lower-frequency mode of energy  $\hbar\omega_{\text{LO},1} = \sqrt{\epsilon_0^{(\text{ox})}/\epsilon_i^{(\text{ox})}} \hbar\omega_{\text{TO},1}$ . The oxide

and Si lightlines,  $\hbar\omega = \hbar c q / \sqrt{\epsilon_{\infty}^{(\text{ox})}/\epsilon_0}$  and  $\hbar c q / \sqrt{\epsilon_{\text{sc}}/\epsilon_0}$ , respectively, are indicated by dash lines, the Si lightline being the lower curve. Dotted lines indicate the “uncoupled” dispersions obtained by ignoring the phonon-plasmon coupling. Branches above the lightlines are radiatively damped. The lower branch  $\hbar\omega_-$  at low  $q$  is

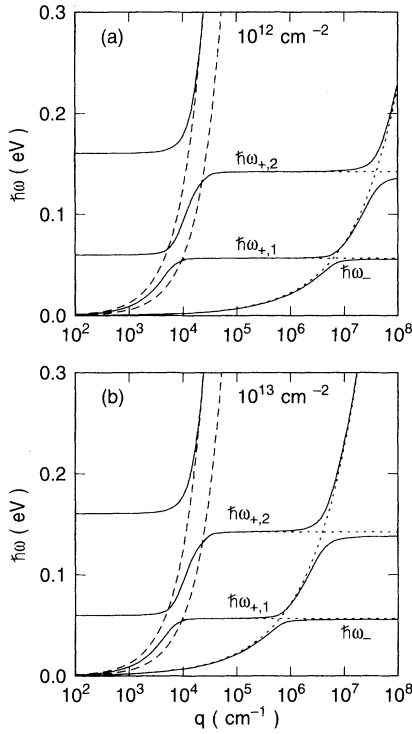


FIG. 6. The dispersion of the coupled optical-phonon-plasmon modes at the interface between Si and SiO<sub>2</sub> for an infinite thickness of both the oxide and the space-charge region of the semiconductor for two values of electron sheet densities. The dashed lines indicate the SiO<sub>2</sub> (upper) and Si lightlines (lower), the dotted lines the dispersion of the interface excitations in absence of plasma-phonon coupling. Branches—or portion of branches—above the oxide lightline are radiatively damped. The parameters used for the calculation are presented in Table I.

mainly the 2D plasmon. Note how at the highest density of  $10^{13} \text{ cm}^{-2}$  the SO modes couple with the plasma modes at a relatively large wavelength. Therefore, there appears to be only a small window of wavelengths over which the SO modes are “good” excitations: At small  $q$  they are radiatively damped, while at large  $q$  they become coupled surface optical phonon/plasmons modes. In this intermediate range the energy of the modes is given approximately<sup>100</sup> by

$$\hbar\omega_{+,1} \simeq \hbar\omega_{\text{SO},1} \equiv \left[ \frac{\epsilon_0^{(\text{ox})} + \epsilon_{\text{sc}}}{\epsilon_i^{(\text{ox})} + \epsilon_{\text{sc}}} \right]^{1/2} \hbar\omega_{\text{TO},1}, \quad (38a)$$

$$\hbar\omega_{+,2} \simeq \hbar\omega_{\text{SO},2} \equiv \left[ \frac{\epsilon_i^{(\text{ox})} + \epsilon_{\text{sc}}}{\epsilon_\infty^{(\text{ox})} + \epsilon_{\text{sc}}} \right]^{1/2} \hbar\omega_{\text{TO},2}, \quad (38b)$$

while the phonon content of these excitations,<sup>102,103</sup>  $S(\omega_{+,i}) \simeq (\omega_{+,i}^2 - \omega_p^2)/(\omega_{\text{SO},i}^2 - \omega_-^2)$  is approximately unity. Note that we are being cavalier about intersubband plasmons,<sup>99</sup> ignoring them altogether. These are excitations of higher energy, which in principle could couple to the interface optical modes. Their fields are somewhat

smaller than the field associated with the intrasubband plasmon we have discussed, but the possibility of complications due to these modes should be kept in mind. On the other side, two arguments work in favor of the simpler picture used in the works by Hess and Vogl and by Moore and Ferry. In first place, the phase space of the region in which the SO modes are radiatively damped is very small. Second, what Fig. 6 does not show is the region in which 2D plasmons are Landau damped, decaying into single-particle excitations. This process may prevent the SO-mode-plasmon coupling we are worried about. Thus, to summarize, we can consider the SO modes sufficiently decoupled and undamped over an intermediate region of  $q$  space large enough to justify the approaches employed previously.<sup>73–76</sup>

The effect of a finite oxide thickness and of the thickness of the space-charge region in the semiconductor may be investigated by considering a second geometry in which a metal of permittivity  $\epsilon_M = \epsilon_0$  occupies the half-space  $z \leq -d_{\text{ox}}$ , where the oxide is the layer of thickness  $d_{\text{ox}}$  in  $-d_{\text{ox}} < z \leq 0$ , the space-charge layer occupies the region  $0 < z \leq d_{\text{sc}}$ , while, finally, the half-space  $z > d_{\text{sc}}$  is bulk Si with permittivity  $\epsilon_{\text{sc}}$ . For this geometry the equivalent of the secular equation (37) becomes<sup>101</sup>

$$\coth(K_{\text{sc}}d_{\text{sc}}) + \frac{K_{\text{sc}}\epsilon^{(\text{ox})}(\omega)}{K_{\text{ox}}\epsilon_{\text{sc}}} \coth(K_{\text{ox}}d_{\text{ox}}) = \epsilon_0 \frac{\Omega_{\text{sc}}^2 K_{\text{sc}}}{\epsilon_{\text{sc}}\omega^2}. \quad (39)$$

The corresponding solutions are shown in Fig. 7. Compared to Fig. 6, note how the plasma energy is depressed at low  $q$  by the finite oxide thickness. The higher-energy SO mode exhibits a “hump” above the asymptotic value  $\hbar\omega_{\text{SO},2}$  at intermediate values of  $q$ , but both modes are otherwise unaffected by the finite thickness of the oxide and space-charge layers.

Having cleared from our minds some concerns about the complicated nature of the optical interface modes, we shall proceed in a rather conventional way in order to compute the scattering rates between the interface modes and the electrons: we still need the scattering potential associated with these modes, and their matrix elements associated with the electronic wave functions. We assume a flat dispersion for the two SO modes, as given by Eq. (38). Following the derivation by Kittel<sup>104</sup> as rephrased by Wang and Mahan, the electrostatic field  $\varphi_q^{(\text{SO},i)}$  associated with the  $q$  component of the two SO modes ( $i=1,2$ ) in the inversion layer is approximately given by

$$\varphi_q^{(\text{SO},i)}(z) = \left[ \frac{\hbar\omega_{\text{SO},i}}{2\pi\epsilon_i q} \right]^{1/2} e^{-qz}, \quad (40)$$

where

$$\frac{1}{\epsilon_1} = \frac{1}{\epsilon_i^{(\text{ox})} + \epsilon_{\text{sc}}} - \frac{1}{\epsilon_0^{(\text{ox})} + \epsilon_{\text{sc}}}$$

and

$$\frac{1}{\epsilon_2} = \frac{1}{\epsilon_\infty^{(\text{ox})} + \epsilon_{\text{sc}}} - \frac{1}{\epsilon_i^{(\text{ox})} + \epsilon_{\text{sc}}}.$$

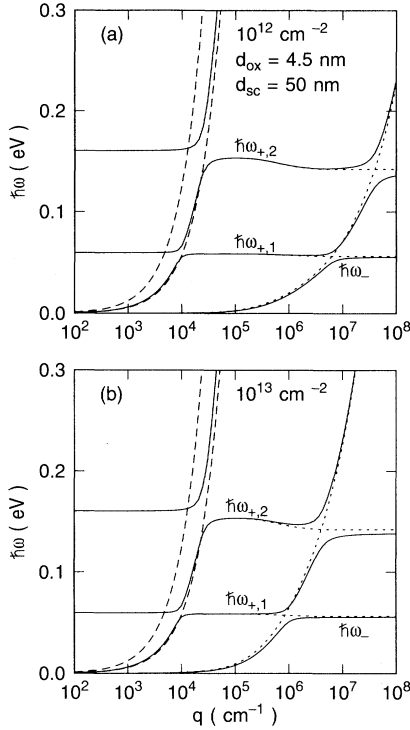


FIG. 7. The dispersion of the coupled optical-phonon-plasmon modes as in Fig. 6, but for finite thickness of both the oxide and the space-charge layer. The finite oxide thickness (4.5 nm in the figure) depresses the plasma dispersion at long wavelength and causes two humps in the dispersion of the surface optical modes.

Ngai and Economou<sup>90</sup> have described a procedure to obtain the field associated with surface plasmons. Following their canonical quantization scheme we find that the potential  $\varphi_q^{(P)}$  associated to the two-dimensional plasmons is

$$\varphi_q^{(P)}(z) = \left[ \frac{\hbar\omega_P(q)}{\pi(\epsilon_0^{(\text{ox})} + \epsilon_{\text{sc}})q} \right]^{1/2} e^{-qz}. \quad (41)$$

The matrix element for a transition from subband  $\mu$  to subband  $\nu$  due to these scattering potentials is simply

$$M_{q,\mu\nu} = 2\pi e \int_0^\infty dz \zeta_\nu(z) \varphi_q(z) \zeta_\mu(z). \quad (42)$$

In principle, electron-plasmon scattering is a legitimate scattering process and should be accounted for. However, there are two reasons to ignore this process. First, we must face the problem related to the decay of the plasmons and their “feeding back” energy and momentum to the 2DEG. Second and more important, in our simulations the Poisson equation is solved self-consistently with the transport equation. In complete analogy to the situation we have already discussed at length for the case of bulk semiconductors,<sup>22,105</sup> the long-wavelength Coulomb interactions among carriers are already accounted for classically by the self-consistent procedure. The introduction of an explicit electron-plasmon scattering mechanism would result in double counting these processes. Therefore, they should be ignored. Although the results given in this section

represent a better approximation to the physics of the collective excitations of the 2DEG than the classical solution of the Poisson equation could provide, the latter accounts automatically for the plasmon decaying and feeding back to the electron ensemble.

We are left with the simple task of obtaining the rate required for an electron with wave vector  $\mathbf{k}$  in subband  $\mu$  to make a transition to subband  $\nu$  with the absorption or emission of a SO mode. From Eqs. (40) and (42), we get

$$\begin{aligned} \frac{1}{\tau_{\mu\nu}^{(\text{SO},i)}(\mathbf{k})} &= \frac{e^2 \omega_{\text{SO},i}}{\hbar^2 \epsilon_i} \kappa_\nu [E_\mu(\mathbf{k}) - E_\nu \pm \hbar\omega_{\text{SO},i}] \\ &\times \theta[E_\mu(\mathbf{k}) - E_\nu \pm \hbar\omega_{\text{SO},i}] \\ &\times \int_0^{2\pi} d\beta' \frac{|I_{\mu\nu}[q(\beta')]|^2}{q(\beta')} \\ &\times K(\beta')^2 \left( \frac{1}{2} \mp \frac{1}{2} + n_{\text{SO},i} \right), \end{aligned} \quad (43)$$

where  $K(\beta')$  is defined before Eq. (22), the upper and lower signs are for absorption and emission, respectively,  $n_{\text{SO},i}$  is the thermal population of the SO mode of type  $i$ ,

$$I_{\mu\nu}(q) = \int_0^\infty dz \zeta_\nu(z) e^{-qz} \zeta_\mu(z),$$

and  $q(\beta')$  is given by Eqs. (26) and (27), by modifying trivially Eq. (27) to account for the different final energy entering this inelastic process.

#### D. Coulomb and surface-roughness scattering

In this section, we consider Coulomb scattering with charges located in the semiconductor space-charge regions (the ionized impurities), in the oxide (fixed oxide charges), and at the Si-SiO<sub>2</sub> interface (interface fixed charges). We shall also consider scattering with the roughness of the Si-SiO<sub>2</sub> interface. Electrons are assumed to interact with the *unscreened* scattering potentials, deferring the discussion of screening effects to the next section.

Stern and Howard<sup>106</sup> have given the expression for the potential due to charges located in the regions of interest. Under their expressions, also summarized in Ref. 1, and employing the Born approximation, we can express the Coulomb scattering rate (with obvious notation) as follows:

$$\begin{aligned} \frac{1}{\tau_{\mu\nu}^{(r)}(\mathbf{k})} &= \frac{e^4 N_r}{8\pi \hbar^3 \epsilon_r^2} \kappa_\nu [E_\mu(\mathbf{k}) - E_\nu] \theta[E_\mu(\mathbf{k}) - E_\nu] \\ &\times \int_0^{2\pi} d\beta' \frac{H_{\mu\nu}^{(r)}[q(\beta')]}{q(\beta')^2} K(\beta')^2, \end{aligned} \quad (44)$$

where, again,  $q(\beta')$  is given by Eq. (27), and the form factor  $H_{\mu\nu}^{(r)}$  is defined as

$$H_{\mu\nu}^{(r)}(q) = \int_0^\infty dz \int_0^\infty dz' \zeta_\nu(z) \zeta_\nu(z') I_q^{(r)}(z, z') \zeta_\mu(z) \zeta_\mu(z'). \quad (45)$$

The index  $r$  runs over impurity, oxide, and interface charges. For impurity scattering, after integration over the semiconductor space-charge layer of thickness  $d_{\text{sc}}$ , we have

$$\epsilon_{\text{imp}} = \epsilon_{\text{sc}}, \quad (46a)$$

$$N_{\text{imp}} = (N_A + N_D)d_{\text{sc}}, \quad (46b)$$

$$I_q^{(\text{imp})}(z, z') = ((|z - z'| + q^{-1})e^{-q|z - z'|} + 1/(2q))(e^{-q(z + z')} + e^{q(z + z' - 2d_{\text{sc}})}) \\ + \bar{\epsilon} \{ (z + z' + q^{-1})e^{-q(z + z')} - 1/(2q)[e^{-q(z - z' + 2d_{\text{sc}})} + e^{-q(z' - z + 2d_{\text{sc}})}] \} \\ + \bar{\epsilon}^2 e^{-q(z + z')}(1 - e^{-2d_{\text{sc}}})/(2q)/d_{\text{sc}}, \quad (46c)$$

where  $\bar{\epsilon} = (\epsilon_{\text{sc}} - \epsilon_0^{(\text{ox})})/(\epsilon_{\text{sc}} + \epsilon_0^{(\text{ox})})$ . For scattering with charges uniformly distributed in  $(-d_2, -d_1)$  in the oxide, after a similar integration over the distribution of the charges, we find

$$\epsilon_{\text{ox}} = (\epsilon_{\text{sc}} + \epsilon_0^{(\text{ox})})/2, \quad (47a)$$

$$N_{\text{ox}} = N_{\text{ox}}^{(3D)}(d_2 - d_1), \quad (47b)$$

$$I_q^{(\text{ox})}(z, z') \\ = e^{-q(z + z')}(e^{-2qd_1} - e^{-2qd_2})/[2q(d_2 - d_1)], \quad (47c)$$

where  $N_{\text{ox}}^{(3D)}$  is the volume density of charges in the oxide. Finally, for scattering with interface charges, we have

$$\epsilon_{\text{it}} = (\epsilon_{\text{sc}} + \epsilon_0^{(\text{ox})})/2, \quad (48a)$$

$$N_{\text{it}} = \text{areal density of interface charges}, \quad (48b)$$

$$I_q^{(\text{it})}(z, z') = e^{-q(z + z')}. \quad (48c)$$

Intervalley processes can be considered by accounting for the degeneracy of the final valleys and modifying the wave vector  $q$  to account for the valley separation in the Brillouin zone. In general, the larger wave vector in the denominator of Eq. (44) will reduce the strength of these collisions.

The last scattering process we consider is scattering with the roughness of the Si-SiO<sub>2</sub> interface. We have followed Ando's approach,<sup>107</sup> modified only in order to account for intersubband transitions, consistent with the generality of the scattering rates we have obtained so far. We have

$$\frac{1}{\tau_{\mu\nu}^{(\text{SR})}(\mathbf{k})} = \frac{\Delta^2 \Lambda^2}{2\hbar^3} \kappa_{\nu}[E_{\mu}(\mathbf{k}) - E_{\nu}]\theta[E_{\mu}(\mathbf{k}) - E_{\nu}] \\ \times \int_0^{2\pi} d\beta' K(\beta')^2 |S[q(\beta')]|^2 |\Gamma_{\mu\nu}[q(\beta')]|^2. \quad (49)$$

Here  $S(q)$  is the term of the matrix element which depends on the spectral distribution of the "steps" at the interface,  $\Delta$  being the rms step height and  $\Lambda$  the autocorrelation of the step distance. For the original Gaussian model,<sup>107</sup> we have

$$|S(q)|^2 = e^{-q^2 \Lambda^2/4}, \quad (50)$$

while for the exponential-decaying model proposed by Goodnick and co-workers<sup>108,109</sup>—which we have adopted—we have

$$|S(q)|^2 = (1 + q^2 \Lambda^2/2)^{-3/2}. \quad (51)$$

In Eq. (49) we have defined

$$\Gamma_{\mu\nu}(q) = \Gamma_{\mu\nu}^{(0)} + \frac{e^2}{\epsilon_{\text{sc}}} \bar{\epsilon} H_{\mu\nu}^{(\text{SR})}(q), \quad (52)$$

where

$$\Gamma_{\mu\nu}^{(0)} = \int_0^\infty dz \left\{ \zeta_{\mu}(z) \frac{\partial V(z)}{\partial z} \zeta_{\nu}(z) - E_{\nu} \frac{d\zeta_{\mu}(z)}{dz} \zeta_{\nu}(z) \right. \\ \left. + E_{\mu} \frac{d\zeta_{\nu}(z)}{dz} \zeta_{\mu}(z) \right\} \quad (53a)$$

and

$$H_{\mu\nu}^{(\text{SR})}(q) = \int_0^\infty dz \zeta_{\nu}(z) \varphi_q^{(\text{SR})}(z) \zeta_{\mu}(z), \quad (53b)$$

with

$$\varphi_q^{(\text{SR})}(z) = (n_s + N_{\text{imp}})e^{-qz} + \frac{q^2}{16\pi} \left[ \frac{K_1(qz)}{qz} - \frac{\bar{\epsilon}}{2} K_0(qz) \right]. \quad (53c)$$

$K_0$  and  $K_1$  are the modified Bessel functions. Note that Eq. (53a) is occasionally expressed in terms of the derivatives of the envelope wave functions at the interface.<sup>110</sup> We prefer the form given above, since it is more accurate when numerical solutions are available for  $\zeta_{\mu}$  (Ref. 111). Finally, note that we ignore additional complications due to the finite oxide thickness, which may further reduce the mobility via scattering with remote roughness.<sup>112</sup>

### E. Screening

Coulomb and roughness scattering have quite a strong effect on the electron mobility, particularly at low and high electron densities, respectively. Obviously, the scattering potentials are strongly affected by the screening of the mobile charges in the inversion layer, and every theory hoping to explain the density and temperature dependence of the electron mobility must account for these screening corrections. However, because of the difficulty of treating it correctly, screening has either been ignored, as done originally by Sah, Ning, and Tschoop,<sup>113</sup> or has been approximated rather crudely,<sup>70,71</sup> since the approximations introduced by Ning and Sah<sup>114</sup> have been embraced by many authors with additional simplifications. For example, the  $q$  component of the scattering potential  $\varphi_q(z)$  is sometimes screened by approximating the low-frequency dielectric function  $\epsilon_{2D}(\mathbf{q}, \omega \rightarrow 0)$ , with  $\epsilon_{\text{sc}}[1 + q_s(q)/q]$ , where the two-dimensional screening parameter  $q_s(q)$  is quite often taken to be independent of  $q$  (e.g., Ref. 71, which ignores the form factor accounting for the finite extent of the wave

functions) and to be given by the equivalent of the Debye-Hückel expression in two dimensions:

$$q_s^{(\text{DH})} \simeq \frac{e^2}{2\epsilon_{\text{sc}}} \frac{\partial n_s}{\partial E_F} \rightarrow \frac{e^2 n_s}{2\epsilon_{\text{sc}} k_B T}, \quad (54)$$

the last step being taken in the high-temperature limit. When working in the extreme quantum limit, Stern<sup>115</sup> has employed the random-phase approximation (RPA) to show that the dependence of  $q_s$  on  $q$  is actually quite strong, the screening parameter decreasing fast at short wavelengths, so that the effects of screening are reduced compared to what is predicted by Eq. (54) above. Even accounting for these form-factor corrections, Ning and Sah<sup>114</sup> have noticed that screening is overestimated in the Debye-Hückel approximation, except at low electron densities. Another issue, usually ignored, is related to the population of many subbands in a realistic room-temperature situation. Obviously, the different spatial and energetic distribution of the carriers in higher-lying subbands alters the picture. To our knowledge, the only attempts to estimate the effect of screening resulting from carriers populating more than one subband are due to two-subband ( $\mu=0$  and  $0'$ ) calculations by Stern,<sup>26</sup> and to the more recent and sophisticated approach by Yokoyama and Hess<sup>27</sup> for the  $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$  system, while Stern and Howard<sup>106</sup> and Siggia and Kwok<sup>116</sup> have studied the case of the multiple-subband population and its effect on impurity scattering only at a formal level. We can easily apply these results to the case of an arbitrary number of subbands within the RPA, following Stern and Howard,<sup>106</sup> Stern,<sup>26</sup> and Yokoyama and Hess.<sup>27</sup>

The screened potential  $\varphi^{(s)}(\mathbf{r}, z)$  due to the external charge distribution  $\rho^{(\text{ext})}(\mathbf{r}, z)$  satisfies the Poisson equation

$$\nabla \cdot (\epsilon \nabla \varphi^{(s)}) = -(\rho^{(\text{ext})} + \rho^{(\text{ind})}), \quad (55)$$

where  $\rho^{(\text{ind})}$  is the induced charge density. Let us consider the Fourier-Bessel components  $\varphi_q^{(s)}(z)$ :

$$\varphi_q^{(s)}(\mathbf{r}, z) = \int dq J_0(qr) \varphi_q^{(s)}(z), \quad (56)$$

where  $J_0$  is the Bessel function, and assume linear screening, so that the induced charge density is proportional to the expectation value of the potential in each subband,  $\langle \varphi_q^{(s)} \rangle_v \equiv \varphi_{q, \nu\nu}^{(s)}$ :

$$\rho_q^{(\text{ind})}(z) = -2\epsilon_{\text{sc}} \sum_v q_{s, \nu}(q) |\xi_\nu(z)|^2 \varphi_{q, \nu\nu}^{(s)}. \quad (57)$$

The matrix elements of the scattering potential are

$$\varphi_{q, \mu\nu}^{(s)} = \int_0^\infty dz \xi_\nu(z) \varphi_q^{(s)}(z) \xi_\mu(z), \quad (58)$$

and the screening parameter  $q_{s, \nu}(q)$  can be evaluated using the high-temperature limit given by Fetter:<sup>117</sup>

$$q_{s, \nu}(q) = \frac{e^2 n_\nu}{2\epsilon_{\text{sc}} k_B T} g_1(q \lambda_\nu) \equiv q_{s, \nu}^{(\text{DH})} g_1(q \lambda_\nu), \quad (59)$$

where the “Debye-Hückel” parameter  $q_{s, \nu}^{(\text{DH})}$  is simply given by Eq. (54), replacing  $n_s$  with the population  $n_\nu$  in

the  $\nu$ th subband, and  $\lambda_\nu$  is the thermal wavelength of electrons in subband  $\nu$ ,  $[2\pi\hbar^2/(m_\nu k_B T)]^{1/2}$ , where  $m_\nu$  is the conductivity mass in the  $\nu$ th subband, and the function  $g_1$  is defined as

$$g_1(x) = \frac{2\pi^{1/2}}{x} \Phi \left[ \frac{x}{4\pi^{1/2}} \right],$$

where

$$\Phi(y) = 2e^{-y^2} \int_0^y dt e^{t^2}$$

is the plasma dispersion function.<sup>118,119</sup> Using the Green’s function (36), the Fourier-Bessel components of the screened potential can be written as<sup>120</sup>

$$\begin{aligned} \varphi_q^{(s)}(z) &= \varphi_q^{(\text{ext})}(z) \\ &\quad - 2 \sum_\nu q_{s, \nu}(q) \int_0^\infty dz' G_q(z, z') |\xi_\nu(z')|^2 \varphi_{q, \nu\nu}^{(s)}, \end{aligned} \quad (60)$$

where  $\varphi_q^{(\text{ext})}(z)$  can be viewed as the unscreened scattering potential. Multiplying Eq. (60) by  $|\xi_\mu(z)|^2$  and integrating, we find that the intrasubband matrix elements of the screened scattering potential,  $\varphi_{q, \mu\mu}^{(s)}$  can be obtained from the unscreened matrix elements  $\varphi_{q, \mu\mu}^{(\text{ext})}$  by inverting the linear problem:

$$\varphi_{q, \mu\mu}^{(s)} = \varphi_{q, \mu\mu}^{(\text{ext})} - \sum_\nu \frac{q_{s, \nu}(q)}{q} \mathcal{G}_{\mu\nu, \nu\nu}(q) \varphi_{q, \nu\nu}^{(s)}, \quad (61)$$

where the form factor  $\mathcal{G}_{\mu\nu, \lambda\lambda}(q)$  is given by

$$\begin{aligned} \mathcal{G}_{\mu\nu, \lambda\lambda}(q) &= \int_0^\infty dz \int_0^\infty dz' \xi_\mu(z) \xi_\nu(z) \\ &\quad \times \tilde{G}_q(z, z') \xi_\lambda(z') \xi_\lambda(z'). \end{aligned} \quad (62)$$

Here  $\tilde{G}_q$  is simply the Green’s function given by Eq. (36) multiplied by  $2q$ , in order to exhibit explicitly the “usual”  $q$  dependence in the denominator of Eq. (61). For the simple geometry of two semi-infinite media:

$$\tilde{G}_q(z, z') = e^{-q|z-z'|} + \tilde{\epsilon} e^{-q(z+z')}, \quad (63)$$

$\tilde{\epsilon}$  having been defined after Eq. (46c). Having obtained the “diagonal” intrasubband terms, the intersubband matrix elements are similarly obtained from Eq. (60):

$$\varphi_{q, \mu\nu}^{(s)} = \varphi_{q, \mu\nu}^{(\text{ext})} - \sum_\lambda \frac{q_{s, \lambda}(q)}{q} \mathcal{G}_{\mu\nu, \lambda\lambda}(q) \varphi_{q, \lambda\lambda}^{(s)}. \quad (64)$$

Were we wishing to simplify the picture somewhat and employ the Hartree approximation instead of the RPA, we would replace the screened matrix elements  $\varphi_{q, \mu\mu}^{(s)}$  in the right-hand sides of Eqs. (61) and (64) with the unscreened values  $\varphi_{q, \mu\mu}^{(\text{ext})}$ , so that the inversion of the dielectric matrix required by Eq. (61) would not be needed. However, from a computational point of view, we would not gain much, since the numerically intensive step is setting up the dielectric matrix, considering the double integrals in Eq. (62). Therefore, we adopted Eqs. (61) and (64) as they stand.

Static screening, as derived above, is certainly satisfactory in the case of Coulomb and roughness scattering.

The unscreened potential for scattering with interface charges playing the role of  $\varphi_q^{(\text{ext})}(z)$  in Eq. (60) is simply

$$\varphi_q^{(\text{it})}(z) = \frac{e}{(\epsilon_0^{(\text{ox})} + \epsilon_{\text{sc}})q} e^{-q|z|}. \quad (65)$$

For scattering with surface roughness, the matrix elements of the unscreened potential are given by Eq. (52), up to a multiplicative constant which is immaterial here. In the case of scattering with oxide charges and impurities, there is a slight complication due to the integration over the distribution of charges which we have performed to arrive at Eqs. (44) and (46c) and Eqs. (44) and (47c), respectively. Thus we must consider the potential due to an impurity or oxide charge at  $z_e$ , consider its screened matrix elements, square them, and integrate them over the charge distribution in order to replace the factor  $H_{\mu\nu}^{(r)}(q)/(\epsilon_r q)^2$  in Eq. (44) with its screened expression. The potential in the semiconductor due to an impurity at  $z_e$  is

$$\varphi_q^{(\text{imp})}(z, z_e) = \frac{e}{2\epsilon_{\text{sc}}q} [e^{-q|z-z_e|} + \tilde{\epsilon} e^{-q|z+z_e|}], \quad (66)$$

while for a charge at  $z_e$  in the insulator, the unscreened potential for  $z > 0$  is

$$\varphi_q^{(\text{ox})}(z, z_e) = \frac{e}{(\epsilon_0^{(\text{ox})} + \epsilon_{\text{sc}})q} e^{-q|z-z_e|}. \quad (67)$$

We shall describe later in Sec. III C how we have handled the scattering rates with screened potentials from a numerical point of view.

The effect of screening on electron-phonon scattering is not as clear as the effect of static screening on Coulomb or roughness scattering. Scattering with the SO modes involves excitations at large frequency, much larger than the plasma frequency of the 2DEG. Therefore, short of using a formulation of dynamic screening much more sophisticated than we are capable of handling, it is probably "less wrong" to leave those scattering potentials unscreened. Intervalley scattering with short-wavelength phonons will be largely unaffected by screening, in view of the large wave vectors entering the transitions. Moreover, for most of those transitions the frequencies involved are once again much larger than the frequencies at which the two-dimensional plasma can respond.

Finally, we must discuss the complicated role played by screening in intravalley scattering with acoustic phonons. The conventional static dielectric function has been used in many instances in the bulk.<sup>121-123</sup> More recently, after some debate on whether screening of the deformation-potential interaction is appropriate in order to analyze transport in modulation-doped heterostructures,<sup>124,125</sup> the RPA formalism used by Stern<sup>115</sup>—extended to finite temperatures following Maldague<sup>126</sup>—has been employed to screen statically the deformation-potential interaction in two dimensions.<sup>127</sup> However, the issue is not trivial. At first, the use of static screening seems intuitively correct, since at the low frequencies of long-wavelength acoustic phonons, free carriers should screen the ionic deformation potential in a trivial way, as it happens in the case of the long-range Coulomb poten-

tial. On second thought, from the microscopic perspective of the electron-phonon interaction, such as presented in the excellent review by Vogl,<sup>128</sup> one may wonder why the potential of the ions responsible for the short-range electron-phonon coupling, already screened by the valence electrons, should be modified to such a large extent by a few extra free carriers in the conduction band. Indeed, as shown in Ref. 128, this short-range component of the electron-phonon interaction is left unaffected by the free carriers. Boguslawski and Mycielski,<sup>129</sup> on the other side, have shown that in bulk semiconductors, long-range components of the electron-phonon interaction, when screened by free carriers, are responsible for a partial cancellation of the short-range component, resulting in an effective screening of the deformation-potential interaction which can be expressed as the unscreened potential divided by the static (Thomas-Fermi type) dielectric function. The net result is that the naive picture we started from (i.e., conventional static screening) does indeed hold. Does this hold also in two dimensions? A qualitative idea of whether static screening is appropriate or whether dynamic screening should be employed in this case can be obtained by assuming that a simple (local and scalar) dielectric function can be defined for the 2DEG and by expressing the frequency dependence of the dielectric function in terms of the static function,  $\epsilon_{2D}(\mathbf{q}, 0)$ , using the Lundqvist-Overhauser plasmon-pole approximation.<sup>130</sup> Following Vinter,<sup>51</sup> we have

$$\epsilon_{2D}(\mathbf{q}, \omega) \simeq \epsilon_{\text{sc}} \left[ 1 + \frac{\omega_p^2(q)}{\omega^2 - \omega_q^2} \right]^{-1}, \quad (68)$$

where

$$\omega_p(q) \simeq \left[ \frac{e^2 n_s}{(\epsilon_{\text{sc}} + \epsilon_0^{(\text{ox})}) m_{\parallel}} \right]^{1/2} q^{1/2} \quad (69)$$

is the plasma frequency, and

$$\omega_q^2 = \frac{\omega_p^2(q)}{\frac{\epsilon_{\text{sc}}}{\epsilon_{2D}(\mathbf{q}, 0)} - 1}. \quad (70)$$

If we account for the finite oxide thickness, and going to the limit of a very thin oxide, the low- $q$  plasma frequency above should be replaced by the dispersion relation obtained from Eq. (39):

$$\omega_p(q) \simeq \left[ \frac{e^2 n_s d_{\text{ox}}}{\epsilon_0^{(\text{ox})} m_{\parallel}} \right]^{1/2} q. \quad (71)$$

We see that static screening represents a good approximation when dealing with the scattering potential associated with excitations for which  $\omega^{(\text{ex})}(q) \ll \omega_p(q)$ . On the contrary, when  $\omega^{(\text{ex})}(q) \gg \omega_p(q)$  screening becomes ineffective, since the free carriers are unable to respond to the fast perturbation. In the bulk semiconductor, the picture is qualitatively similar. However, in this case, at sufficiently high carrier densities the plasma frequency takes a value almost independent of  $Q$ , while the frequency of the acoustic phonons,  $\omega^{(i)}(Q)$ , goes to zero approximately as  $c_i Q$  ( $c_i$ , with  $i$ =longitudinal or transverse, is



the sound velocity) at small  $Q$ . It follows that static screening is always appropriate. If we move to the two-dimensional case, we notice a dramatic difference: While the frequency of the plasmon approaches zero at low  $q$ , according to Eqs. (69) or (71), the wave vectors of phonons with which the electrons interact always have a nonzero component normal to the interface,  $q_z$ , which, for intrasubband transitions in subband  $\mu$ , is of the order of

$$\bar{q}_\mu \sim \int_{-\infty}^{\infty} dq_z |\mathcal{F}_{\mu\mu}(q_z)|^2 = 2\pi F_{\mu\mu} \sim \frac{2\pi}{\bar{z}_\mu}, \quad (72)$$

where  $\bar{z}_\mu$  is the average width of the charge density associated with the  $\mu$ th subband. Therefore, as  $q \rightarrow 0$ , the frequency of the phonons involved in intrasubband transitions within the  $\mu$ th subband,  $\omega^{(i)}(q) = c_i(\bar{q}_\mu^2 + q^2)^{1/2}$ , approaches a nonzero value, so that  $\omega^{(ex)}(q) \equiv \omega^{(i)}(q) \gg \omega_p(q)$  and screening becomes ineffective. In physical terms, if we are willing to ignore intersubband plasmons (i.e., the “secondary screening” effects of Ref. 131, shown there to be in general small), perturbations propagating along the direction normal to the interface cannot be screened by the electrons, as they are “frozen” into their wave functions and cannot oscillate in the quantized direction. We can reach the same conclusion by going to the extreme quantum limit and considering the dielectric function given by Stern<sup>115</sup> in the RPA. We find that as  $q \rightarrow 0$  and  $\omega \sim q$  (as we expect if the phonons propagated in the plane of the interface), then  $\epsilon \rightarrow \epsilon_{sc}(1 + q_s/q)$ , where  $q_s = e^2 m_i / (\epsilon_{sc} \hbar^2)$  in the 0 subband. This is the expected result, i.e., conventional static screening. However, if we let  $q \rightarrow 0$ , but  $\omega \sim c_i \bar{q}_\mu = 0$ , as it happens in reality, then the real part of the dielectric function approaches  $\epsilon_{sc} - e^2 n_s q / (2m_i \omega^2)$ , i.e., it approaches the unscreened value with a small amount of antiscreening, as noted by Ridley.<sup>58</sup> We see that while we could neglect the energy of the acoustic phonons compared to the average electron energy in order to obtain the electron-phonon-scattering rates at large enough temperatures, on the contrary we are not allowed to ignore it when dealing with the dielectric response, since  $\omega^{(i)}(q)$  is comparable to (or larger than) the plasma energy.

These arguments are only qualitative and indicate that static screening may be inadequate at long wavelengths. We need more general arguments to tell us under which conditions the deformation-potential interaction should be screened or not. We can obtain such a quantitative picture in a couple of meaningful cases, by selecting the extreme examples of intrasubband LA phonons—for which we expect significant deviations from static screening, due to their relatively large frequencies—and the smaller corrections relative to TA phonons. A quantitative measure of the role played by dynamic screening in modifying the phonon-limited mobility can be obtained by considering only electrons in the lowest-lying subband (assumed here to be parabolic for simplicity), and by going to the limit of high temperatures and low densities, i.e., nondegenerate situations. In this regime we can replace<sup>132</sup> the static screening parameter  $q_{s,v}(q)$  given by Eq. (59) with its dynamic value  $q_{s,v}(q, \omega)$ , as given by

Fetter and Walecka:<sup>118,133</sup>

$$q_{s,v}(q, \omega) = q_{s,v}^{(1)}(q, \omega) + i q_{s,v}^{(2)}(q, \omega), \quad (73)$$

where

$$q_{s,v}^{(1)}(q, \omega) = q_{s,v}^{(DH)} \frac{\pi^{1/2}}{q \lambda_v} \times \left\{ \Phi \left[ \left[ \frac{m_v}{2k_B T} \right]^{1/2} \left[ \frac{\omega}{q} + \frac{\hbar q}{2m_v} \right] \right] - \Phi \left[ \left[ \frac{m_v}{2k_B T} \right]^{1/2} \left[ \frac{\omega}{q} - \frac{\hbar q}{2m_v} \right] \right] \right\} \quad (74a)$$

and

$$q_{s,v}^{(2)}(q, \omega) = q_{s,v}^{(DH)} \frac{\pi \hbar \omega}{q \lambda_v k_B T} \exp \left[ -\frac{m_v \omega^2}{2k_B T q^2} - \frac{\hbar^2 q^2}{8m_v k_B T} \right] \times \frac{\sinh \left[ \frac{\hbar \omega}{2k_B T} \right]}{\frac{\hbar \omega}{2k_B T}}, \quad (74b)$$

evaluated along the LA-phonon dispersion  $\omega = c_L(\bar{q}_v^2 + q^2)^{1/2}$ . We can now obtain the screened electron-phonon matrix elements by solving Eqs. (61) and (64) employing the unscreened acoustic-phonon potential  $\sim \exp(-iq_z z)$ . As a significant example, we can evaluate the screened intrasubband matrix elements employing the phonon potential at the particular value of  $q_z$  given by Eq. (72) above, just to get a quantitative idea in a particular, “average” case. In Fig. 8(a), we show the real part of the dynamic screening parameter. Notice the large amount of antiscreening at long wavelengths as compared to the large amount of screening obtained in the static approximation. The resulting correction to the phonon potential has been expressed in terms of the magnitude of an “effective” inverse dielectric function  $\epsilon_{sc} \epsilon_{00}^{(LA)}(q)^{-1}$ —defined as the ratio between the magnitude of the matrix elements of the screened phonon potential and the magnitude of those of the unscreened potential, which is shown in Fig. 8(b). It appears that screening has no effect, except for the two large structures at small values of  $q$  resulting from the quasisingular behavior of the dielectric matrix as electrons in the two most populated subbands exhibit complete antiscreening: Given their slow response, as indicated by the low plasma frequency at small  $q$ , carriers cannot keep up with the faster-oscillating deformation potential: At very long wavelength they do not screen at all, while for particular but

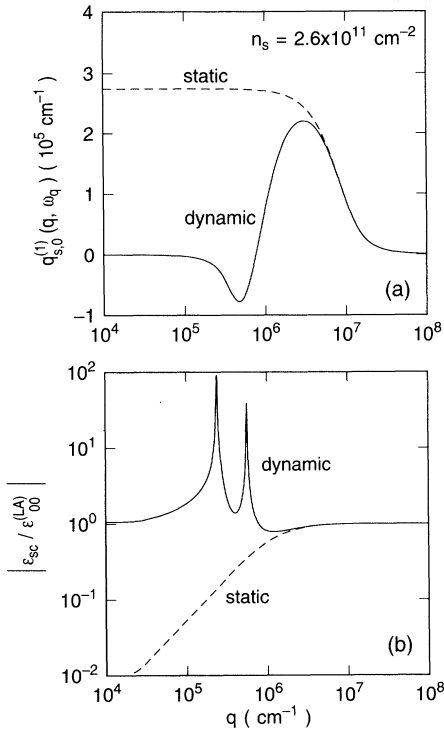


FIG. 8. The dynamic screening parameter in the lowest-lying subband (a) and screening corrections to the intrasubband electron-LA-phonon-scattering potential (b) evaluated in the high-temperature limit using screening due to the first five subbands of each ladder. In (a), the dynamic screening parameter is evaluated along the linear dispersion of the LA phonons  $\omega = \omega^{(LA)}(q) \approx c_L(\bar{q}_0^2 + q^2)^{1/2}$ , accounting for the “fuzzy” component of the phonon wave vector along the quantization direction  $z$ . In (b), the correction to the scattering potential is shown for the case of intrasubband scattering in the lowest-lying subband, and it is compared to the much stronger screening correction obtained in the static approximation.

still long “resonant” wavelengths they end up being in complete opposition to phases with external perturbations, thus actually enhancing, rather than weakening, the external potential. The small value of the imaginary component of the screening parameter, Eq. (74b), is not effective in smearing these structures, while collisional broadening will probably reduce them.<sup>134</sup> We can now use this result to estimate its effect on the LA-phonon-limited mobility for intrasubband scattering for electrons in the bottom subband. The momentum relaxation rate  $1/\tau_p^{(LA)}(E)$  is proportional to the angular average of the screening corrections, i.e.,

$$\frac{1}{\tau_p^{(LA)}(E)} \propto \int_0^{2\pi} d\beta (1 - \cos\beta) |\epsilon_{sc} \epsilon_{00}^{(LA)}(q)^{-1}|^2 \Delta_{LA}(\theta_Q)^2, \quad (75)$$

where  $q = 2k \sin(\beta/2)$ ,  $k = (2m_e E)^{1/2}/\hbar$ , and  $\beta$  is the

scattering angle. The LA-phonon-limited electron mobility is then proportional to the integral over the equilibrium electron distribution, i.e.,

$$\mu_{LA} \propto \int_0^\infty dE \tau_p^{(LA)}(E) f_0(E) [1 - f_0(E)], \quad (76)$$

where  $f_0(E)$  is the Fermi function in subband 0. Dividing Eq. (76) by the similar quantity obtained by ignoring screening, we find that dynamic (anti)screening actually *reduces* the electron mobility by about 14%. Admittedly, this is an extreme example. A similar calculation for the 0' subband yields a mobility only 5% lower than the unscreened value, due to the almost complete cancellation between long-wavelength antiscreening and short-wavelength screening. The role of TA phonons is quite different. Their lower frequency allows proper screening to occur at longer wavelengths, so that the corrected TA-phonon-limited mobility in the lowest-lying subbands is now larger than the unscreened value, by amounts varying from 11% in subband 0', to almost 40% in subband 0. Combining such a set of results for the most populated subbands (0, 0', and 1), we estimate that dynamic screening would result in a less than 20% enhancement of the intrasubband, intravalley phonon-limited mobility at a density of about  $2.6 \times 10^{11} \text{ cm}^{-2}$ . This is only part of the whole picture, since we should keep in mind many additional factors which also affect the mobility, although not to a dramatically large extent: The effect of a finite oxide thickness will slow down the plasma response at long wavelengths [see Figs. 7 and Eq. (71)], thus reducing even more these already moderate corrections. The role of intersubband plasmons, able to screen via a conventional static model, will be that of increasing moderately the mobility,<sup>131</sup> while intravalley phonons would be largely unaffected by screening, as mentioned above. Moreover, at higher densities and/or lower temperatures, the corrections probably will be in the direction of conventional screening (i.e., yielding a higher mobility), since degenerate electrons will sample mainly the relaxation rate at the Fermi energy and, therefore, the role of the screened large- $q$  phonons will be dominant. On the other hand, these are precisely the conditions for other scattering mechanisms (Coulomb and surface-roughness scattering most notably) to control the mobility. Therefore we conclude that *at room temperature and in nondegenerate situations, treating screening within the static approximation may result in an excessive reduction of the deformation-potential interaction and that ignoring screening altogether is not likely to introduce major errors.*

As a result of this long discussion, we can identify two limiting cases: At low temperature and/or high densities, phonon scattering does not play any major role in fixing the electron mobility, while dynamic screening of the deformation-potential interaction matters the most. In this situation, screening is quite intractable, due to the necessity of accounting for degeneracy corrections and for the screening effects of intersubband plasmons, among other complications. On the contrary, in a nondegenerate situation (i.e., low density and/or high temperature), electron-phonon collisions play a major role, but dynamic screening is not dramatically important, as we

have seen above. Therefore, we have decided to leave this interaction unscreened.

### III. NUMERICAL ISSUES

The task of implementing the machinery developed in the previous section into a computer program is not trivial. In this section we describe the approach we have taken. Before going into details, we feel it is necessary to stress a basic condition that our implementation must satisfy. Our goal is not a “one-shot” simulation of electron transport in Si inversion layers under simplifying assumptions, such as uniform field, absence of injecting contacts, rigid specifications of geometrical configuration, etc. Rather, the results presented below represent only the necessary “calibration” step before transport in realistic devices can be studied with some—perhaps only moderate—degree of confidence. Therefore, the numerical implementation of the physical model described above must be performed while keeping in mind a quite general situation of real, nonuniform Si channels, the presence of highly doped, field-free contacts, and a very general geometry. These requirements open a series of problems, some of them involving not only crude numerical issues, but also some physical questions of a nontrivial nature. As an example, given our obvious inability to “quantize the whole device,” we must face the problem of letting bulklike, three-dimensional electrons coexist with two-dimensional electrons in the quantized channel and of being able to transform one population into the other, as transport dictates, without unduly mistreating the basic physics. Also, in order to account for the nonuniformity of the electric field along the channel, a scheme must be devised to maintain the full self-consistency of the Boltzmann, Poisson, and Schrödinger equations in such a situation, along the entire channel. This problem has never been faced before. Our task is rendered somewhat easier by the availability of self-consistent Poisson–Monte Carlo computer program we have described at length before.<sup>22,23,105</sup> We refer the reader to those references for a detailed description of the self-consistent procedure in two spatial dimensions which we have followed.

Our discussion of the numerical issues will proceed as follows: In Sec. III A, we shall consider the spatial and “energy” regions in which electrons are treated according to the two-dimensional transport model of Sec. II. We shall also describe how we have attempted to satisfy energy and momentum conservation when particles leave and enter these regions. In Sec. III B, we describe how we obtain full self-consistency during a Monte Carlo run, and in Sec. III C we shall dwell upon some specific issues related to the efficient computation of the scattering rates. Finally, in Sec. III D we shall describe the device we have simulated.

#### A. The “quantum region”

Starting from a simulation dealing with bulk electrons (which we shall refer to as 3D electrons in the following, as opposed to the 2D electrons we are interested in here), the initial step toward the inclusion of quantization in the

inversion layer is the definition of a *spatial quantum region*  $\mathcal{R}_S$ , defined as the region in the  $x,z$  plane of the device (the  $y$  direction being absent in the simulation), such that two-dimensional transport can be used to deal only with electrons in  $\mathcal{R}_S$ . In addition, we must define an *energy quantum region*  $\mathcal{R}_E$ , in which 2D transport is used to treat electrons in  $\mathcal{R}_S$  with energies lower than some critical value  $E_{th}$ . The necessity of defining these regions should be quite obvious: As stated above, we cannot use 2D transport in the whole device. Similarly, the simulation of high-energy carriers would require too many subbands to be practical. Moreover, even if it were computationally feasible, it would be very well approximated by the “classical limit” of 3D transport. Therefore, we must handle carrier transport within a bulk-transport model above some suitable energy.

The actual definition of the boundaries of  $\mathcal{R}_S$  and  $\mathcal{R}_E$  is nontrivial. Let us consider first the relatively easier task of determining the value of the threshold energy  $E_{th}$  which defines  $\mathcal{R}_E$ . In order to choose the maximum threshold energy above which electrons are treated as 3D carriers, we have taken care to represent the dynamics of the carriers (i.e., their scattering rates) as accurately as possible. To illustrate better what we mean, consider Fig. 9, in which we show the electron-phonon-scattering rates for electrons in the 0 subband, computed using a variable number of subbands. Considering the result for 40 subbands as “exact” (as it is indeed the case in the energy range shown in the figure), we see that the scattering rates are incorrectly underestimated already at very low kinetic energies when using the widely used set of the three subbands ( $\nu=0, 0'$ , and 1), since too many final states are dropped out of the picture even at low energies. As pointed out by Imanaga and Hayafuji,<sup>71</sup> calculations performed using three subbands and 2D electrons *only* should be regarded with suspicion, since the errors are seen to be quite large in Fig. 9: The almost flat scattering

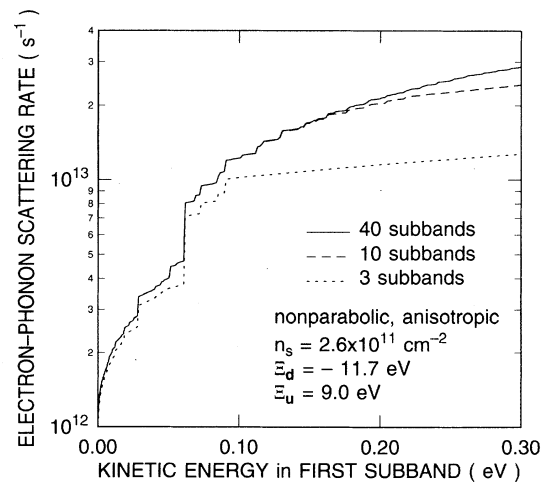


FIG. 9. The calculated electron-phonon-scattering rate at 300 K in the anisotropic and nonparabolic approximations, as in Fig. 5, evaluated using a variable number of subbands. The error made when employing only the three lowest-lying subbands is clearly seen.

rates above 0.1 eV (the slight slope being due only to non-parabolicity corrections), results in very incorrect carrier dynamics even at moderate source-to-drain (longitudinal) electric fields, unless we were willing to consider a switch-over to bulk 3D transport at very small kinetic energies,<sup>71</sup> thus underestimating by definition the importance of quantization in the channel. By contrast, the use of ten subbands is satisfactory up to energies of about 0.2 eV. Looking at Fig. 5, we see that switching to a 3D model at that energy would result in an error of about 15–20 % for the scattering rate when using a bulk, first-order non-parabolic model. In our case, the bulk model at those energies uses anisotropic rates obtained from the pseudopotential band structure of Si, and the error—unavoidable, unless Eq. (6) is employed—is of comparable magnitude. After some “empirical” trial-and-error experimentation, we found that over the range of densities we are interested in, we represent correctly the scattering rates by switching from 2D to 3D transport whenever a carrier has *total* (kinetic + potential) energy larger than the bottom of the fifth unprimed subband. Thus  $E_{\text{th}} = E_5$  (which is about 120 meV in Fig. 5). Typically, seven subbands are populated in the Ohmic regime (the four lowest-lying unprimed and the three lowest-lying primed subbands). However, additional subbands are retained in the simulations for reasons to be discussed below.

A major problem we had to face concerns energy and momentum conservation whenever an electron leaves or enters the region  $\mathcal{R}_E$ . The 3D→2D transition may happen in two cases: (1) after a phonon emission process, or (2) whenever a 3D carrier moves toward a region of high potential energy and drops below the kinetic-energy threshold  $E_{\text{th}}$ . For the opposite of both cases, a 2D→3D transition is in order. If the electron is at position  $(x, z)$ , conservation of kinetic energy and conservation of “parallel” momentum  $\mathbf{k}$  upon a 2D→3D transition involving subband  $\nu$  would demand, classically,

$$E_\nu(\mathbf{k}) + V(x, z) = \varepsilon(\mathbf{k}, k_z) + V(x, z), \quad (77)$$

where  $E_\nu(\mathbf{k})$  is given by Eq. (14) and  $\varepsilon(\mathbf{k}, k_z)$  is the bulk electron dispersion. Since the position  $z$  of a 2D electron is an ill-defined concept, Eq. (77) is more correctly recast by requiring conservation of the expectation value of the kinetic energy on subband  $\nu$ :

$$\langle E_\nu(\mathbf{k}) - V(x) \rangle_\nu = \varepsilon(\mathbf{k}, k_z), \quad (78)$$

which, for the case of a bulk dispersion given by Eq. (11), is equivalent to

$$k_z^2 = \frac{2m_z[E_\nu^{(0)} - \langle V(x) \rangle_\nu]}{\hbar^2}. \quad (79)$$

Note that nonparabolicity is accounted for in Eq. (79), since, to first order in  $\alpha$ , the “parabolic” value  $E_\nu^{(0)}$  is related to the nonparabolic value  $E_\nu(\mathbf{k}=0)$  by

$$E_\nu^{(0)} = E_\nu(\mathbf{k}=0) - \alpha \langle [E_\nu(\mathbf{k}=0) - V(x)]^2 \rangle_\nu,$$

as is seen from Eq. (14). Equation (79) exhibits a real problem, ultimately due to the nonphysical and artificial nature of what we are doing. In a 3D→2D transition,

Eq. (79) cannot be satisfied: the 3D electron cannot be converted to a 2D carrier in subband  $\nu$  conserving energy and momentum exactly. In order to minimize the error, we scan the subband index  $\nu$  over the subbands at our disposal, conserve  $\mathbf{k}$ , and choose the subband which yields the lowest energy mismatch. There are no problems in the case of a 2D→3D transition, since Eq. (78) can be solved easily without additional concerns, by assigning the (now) 3D electron the component  $k_z$  required by Eq. (79), and a  $z$  coordinate selected with probability distribution  $|\xi_\nu(z)|^2$ .

This discussion brings us to the main criterion which guides us in the selection of the boundaries of the spatial region  $\mathcal{R}_S$ . Most of the troublesome 3D→2D transitions occur at the boundary between contacts (i.e., the heavily doped source and drain regions) and the channel. Particles seeking to enter  $\mathcal{R}_S$  must undergo the procedure outlined above. By moving the left and right boundaries  $(x_{\min}, x_{\max})$  of  $\mathcal{R}_S$  as far as possible into the contacts, the wider confining potential in those regions causes a very small subband spacing, thus allowing us to make only “forgivable” violations of energy conservation, on the order of a few meV. Of course, as we already mentioned, the problem stems from the unnatural attempt of forcing a 2D→3D separation in the electron ensemble. Better approaches have been investigated, but only for the case of sharp and well-defined “channels,” such as the study of Ref. 135. In the present investigation, these problems are irrelevant, since, once launched in the channel, memory of the injecting procedure is soon lost. But our concerns, as always, are related to possible artifacts in device simulation. Finally, the boundaries of the region  $\mathcal{R}_S$  along the  $z$  axis,  $(z_{\min}, z_{\max})$ , are quite naturally selected: The Si-SiO<sub>2</sub> interface is the lower boundary, so  $z_{\min} = 0$ . The upper boundary is given by the intersection of the threshold energy  $E_{\text{th}}$ , with the potential along a “vertical slice” at  $x$ , that is  $V[x, z_{\max}(z)] = E_{\text{th}}$ , since electrons at  $z > z_{\max}$  obviously cannot have total energy lower than  $E_{\text{th}}$ .

So far we have fully specified the regions in phase space in which electrons are treated as 2D carriers, and we have given rules to handle electrons leaving and entering the quantum region during ballistic transport. One question which remains open is how to account for scattering-assisted 2D→3D transitions. In order to circumvent the awkward problem of defining “mixed” scattering process between 2D and 3D electron states, we have included in the simulation a few “buffer subbands.” Scattering rates for 2D electrons are computed for a number of subbands larger than specified above. Thus a 2D electron can absorb a phonon and scatter into subband  $\mu > 5$ , being temporarily treated as a 2D electron with total energy above the threshold  $E_{\text{th}}$ . At the beginning of the next time step, the electron will be reclassified as 3D and a 2D→3D transition will be handled as described above. With this procedure, the dynamics of the scattering process is treated correctly. On the other hand, a 3D electron losing energy after phonon emission will be immediately converted to 2D whenever its energy is lower than  $E_{\text{th}}$ , without further complication. As stated above, typically ten subbands are included (three of

them of the “buffer” type). Occasionally, up to 20 subbands have been handled, with  $E_{\text{th}} = E_{10}$ , without detecting noticeable differences in the results.

### B. Self-consistency

We now describe how full self-consistency between the Schrödinger, Poisson, and Boltzmann equations is obtained numerically. Let us start by considering a simulation of a Si MOSFET with the Si-SiO<sub>2</sub> interface in the  $x, y$  plane, at  $z = 0$ . The device is assumed to be uniform in the  $y$  direction, so that the Poisson equation is solved in the two dimensions  $x, z$ . Electron transport in the device is first simulated employing only a bulk electron-transport model, solving the Poisson equation self-consistently with a particle Monte Carlo transport technique, as described in Refs. 22 and 23. In order to do this, a mesh is set up in the  $x, z$  plane (the *Poisson mesh*), the pointlike charge of the simulated particles is mapped onto the nodes of this mesh as described in detail in Ref. 105, and the Poisson equation is solved on this mesh.

In order to account for quantization in the channel, starting from the potential obtained from the bulklike simulation, the Schrödinger equation (13) is solved along each of the “vertical” mesh lines  $x = x_\sigma$ , with  $(\sigma = 1, N_x)$ , where  $x_1 = x_{\min}$  and  $x_{N_x} = x_{\max}$  are the left and right boundary of the spatial quantum region  $\mathcal{R}_S$ . To render the computation more efficient, a one-dimensional uniform mesh (the *Schrödinger mesh*) is introduced along each line, so that Eq. (13) can be solved by a fast direct matrix inversion. The wave functions are assumed to vanish at the interface and at a distance  $z_0$  which is empirically determined in order to avoid errors in the spectrum of the subbands and the form of the wave functions. Typically,  $z_0 \simeq 50$  nm for the device described below. Thus we obtain eigenvalues and eigenfunctions  $E_\nu^{(\sigma)}$  and  $\xi_\nu^{(\sigma)}(z)$  along the channel at the mesh lines. Typically, the number of points employed in the Poisson mesh,  $N_z$ , is about 100–200, resulting in a spacing of 0.25–0.5 nm. The level of coarseness of this mesh affects not only the accuracy of the Poisson-Schrödinger self-consistent procedure, but also the accuracy of the form factors entering the expressions for the scattering rates. These results are now needed for two purposes: we must evaluate the charge density to be fed back to the Poisson solver, and we must evaluate transport parameters.

The charge density is evaluated by assigning each simulated electron in subband  $\nu$  its own envelope wave function, by interpolating along the  $x$  direction the wave functions  $\xi_\nu^{(\sigma)}(z)$ . The square wave function is then mapped back from the uniform Schrödinger mesh to the (in principle nonuniform) Poisson mesh. The charge density—obtained by multiplying the squared wave function by the particle statistical charge weight<sup>105</sup> and the electron charge and by summing over all 2D particles—is then integrated over each Poisson-mesh element and mapped onto the Poisson nodes in a way similar to the “cloud-in-cell” method employed in a bulk situation.<sup>22, 105</sup> The Poisson equation is then solved by accounting for both the bulk charge density resulting from 3D particles and the 2D charge density obtained in this way. Particles

are then moved in the new field configuration, allowed to scatter, and the procedure is repeated again at the successive time step. We use very small time steps ( $\sim 10^{-16}$ – $10^{-15}$  s) to evolve the system self-consistently in time. A similar interpolation between “vertical cuts” is employed also for the subband spectrum, needed for evaluating scattering rates. It should also be noted that the effective longitudinal component of the electric field,  $F_{\nu, x}$ , driving 2D particles in subband  $\nu$  in their ballistic flights between collisions is obtained as

$$F_{\nu, x} = -\frac{1}{e} \frac{\partial E_\nu(\mathbf{k}=0)}{\partial x}, \quad (80)$$

when the energies  $E_\nu$  are all measured from a fixed reference throughout the entire device. In the case of a nonuniform channel, this accounts for the driving force resulting from the shift of the subband bottoms which move up or down as the channel gets narrower or wider.

### C. Numerical evaluation of the scattering rates

The idea of evaluating the scattering rates of Sec. II in a straightforward way is soon ruled out if we want to be able to handle nonuniform channels, as implied above. Every 2D particle carries its own wave function, depending on its position along the channel, and we must perform a different rate evaluation for every particle at every time step. Necessarily, this evaluation must proceed as efficiently as possible. On the contrary, the multidimensional integrals appearing in the expressions for many scattering rates are quite some consuming. Screening effects would render the situation simply hopeless. Therefore we had to invoke many embodiments of the procedure, well known in the Monte Carlo community, of “internal self-scattering.” Let us consider first the electron-TA/LA-phonon deformation-potential interaction, Eq. (23). In that form, a triple integral should be performed, and we cannot afford this luxury from a computing-time perspective. Therefore, we abandon temporarily the anisotropic model of the interaction, by replacing  $\Delta_i(\theta_Q)$  in Eq. (23) with its upper bound  $\Xi_{i, \max}$ , given by  $|\Xi_d| + |\Xi_u|$  for scattering with LA phonons,  $\Xi_u/2$  for scattering with TA phonons. Thus we are left with an expression similar to Eq. (28), with  $\Xi_i^{\text{eff}}$  replaced by  $\Xi_{i, \max}$ , which requires the numerical evaluation of one integral only. The expression thus obtained can be viewed as a self-scattering rate. Using this upper bound to the real scattering rate, electrons which undergo this “fake” collision must also undergo a further check: After each scattering event, a value of  $q_z$  is selected at random with probability distribution  $|\mathcal{F}_{\mu\nu}(q_z)|^2$ , a random angle  $\beta'$  is selected uniformly in  $(0, 2\pi)$ , so that the angle  $\theta_Q$  is uniquely determined via Eq. (25). The ratio  $r = \Delta_i^2(\theta_Q)/\Xi_{i, \max}^2$  is compared with a random number  $\xi$  in  $(0, 1)$ : If  $\xi > r$ , the interaction is considered to be of the self-scattering type, and the electron proceeds as if it never scattered. Otherwise, the collision is real, and the electron is scattered normally. Note how this procedure also allows the determination of the final state after collision.

Intervalley processes given by Eq. (31) require only the evaluation of a one-dimensional integral and pose no

problems. The rates for scattering with SO modes, Eq. (43), involve a double integral, which is tolerable. Not so for screened Coulomb or surface-roughness interactions: Eqs. (44) and (49) in themselves already involve triple and double integrals, respectively. To render the task even harder, the squared matrix elements given by Eq. (45) must be screened and numerically integrated over the charge distribution. Invoking again the self-scattering technique, we proceed in two steps: First, for all 2D particles, these scattering rates are evaluated by replacing the form factor  $H_{\mu\nu}^{(r)}$  in Eq. (44) with unity, and an upper bound to screening is introduced by multiplying the integrand in Eq. (44) by  $[1 + \hat{q}/q(\beta')]^{-2}$ . The parameter  $\hat{q}$  is introduced only to provide a lower cutoff for the magnitude of the  $q$  vector and avoid the divergence of the scattering rates as the energy of the electrons approaches the bottom of the subband. Typically we use  $\hat{q} \approx 10^{-4} q_s^{(\text{DH})}$ . In the case of scattering with surface roughness, the integrand in Eq. (49) is similarly replaced by  $[1 + q(\beta')^2 \Lambda^2/4]^{-3/2}$ . As a second step, for those particles which undergo this type of fake collisions, we compute less approximated scattering rates, reintroducing the dependence of the form factors on the wave functions, but otherwise still considering the interactions essentially unscreened. Finally, for those particles which undergo this second type of interaction, we select a final state by obtaining a value for  $q$  with the rejection technique,<sup>24</sup> using the probability distribution  $H_{\mu\nu}^{(r)}(q)/(q + \hat{q})^2$  as in Eqs. (45) or  $|\Gamma_{\mu\nu}(q)|^2$  as in (52). For this value, we screen the corresponding potential inverting the matrix implicitly defined in Eq. (61) [or also using Eq. (64) for intersubband processes]. In the case of Coulomb scattering with impurities or oxide charges, we select a random location  $z_e$  of the scattering charge with probability distribution  $|\varphi_{q,\mu\nu}^{(\text{imp})}(z_e)|^2 N_{\text{imp}}(z_e)$  or  $|\varphi_{q,\mu\nu}^{(\text{ox})}(z_e)|^2 N_{\text{ox}}(z_e)$  [see Eqs. (66) and (67) above] to perform this last step. Note that in the simulations we have actually assumed a uniform distribution of impurities in the depletion layer. Finally, we consider the ratio  $r = (1 + \hat{q}/q)^2 |\varphi_{q,\mu\nu}^{(s)}(z_e)|^2 / |\varphi_{q,\mu\nu}^{(\text{ext})}(z_e)|^2$  (where ext labels the external impurity, oxide, or interface charges unscreened potential), and accept or reject the collision, depending on whether  $\xi \leq r$  or  $\xi > r$ , as done above in the case of anisotropic scattering with acoustic phonons.

#### D. The device

Before presenting the results of our simulations, we must describe the device we have employed. From the discussion of the previous subsections, it is clear that our machinery is well equipped to handle small, nonuniform devices in which off-equilibrium transport dominates; however, the extraction of Ohmic mobility and drift-velocity characteristics requires large devices and a homogeneous situation. Therefore, we must stretch our simulation capabilities and face problems of device design similar to those experimentalists also face. We have considered a device we have studied successfully before, described in detail in Refs. 22 and 136. It consists of highly doped source and drain regions with a peak doping density of about  $1.5 \times 10^{20} \text{ cm}^{-3}$  obtained from a double im-

plant of Sb and As. A deep channel implant is used to prevent punch-through in short-channel devices and to minimize the degradation of the mobility at the surface of the semiconductor. This implant is modeled by a Gaussian profile peaking 70 nm below the Si-SiO<sub>2</sub> interface, with a width (standard deviation) of 30 nm and a peak concentration of  $7 \times 10^{17} \text{ cm}^{-3}$ . Thus the average doping charge in the depletion layer is about  $5.5 \times 10^{11} \text{ cm}^{-2}$  in strong inversion. The gate oxide is 4.5 nm thick. The major modification of the design with respect to previous simulations is that now the channel length we are forced to choose is at the upper limit of what our CPU-time-consuming program can handle: The metallurgical junctions source/substrate and drain/substrate are separated by 1  $\mu\text{m}$ , yielding an effective channel length of about 0.95  $\mu\text{m}$ . This distance is sufficiently long to ensure that transport reaches steady state toward the drain-end of the channel even at the lowest electron sheet densities at which phonon scattering yields the longest momentum relaxation times. To ensure a good channel uniformity with the thin oxide we have chosen, we have employed the equivalent of the resistive gate<sup>20,21</sup> often employed experimentally to achieve the same effect. The gate contact actually consists of about 20 small contacts, the voltage dropping in a stepwise linear fashion from source to drain, so that the gate-to-source and the gate-to-drain potential drops are equal. The channel density and the longitudinal field were found to be remarkably constant along the channel in this configuration. Typically 10 000 particles were employed in the simulation, about 3000 to 7000 (depending on gate bias) of them treated as 2D carriers. Time steps ranging from  $2 \times 10^{-16}$  to  $10^{-15}$  s were chosen, as dictated by the value of the source-drain field (smaller time steps at higher fields). Very long simulations had to be performed initially in a pure 3D-transport mode, in order to achieve a good equilibrium solution of the steady-state problem. Indeed, at longitudinal (i.e., a source-to-drain) fields of 1.0 kV/cm, transit times of the order of  $10^{-10}$  s are expected. Even after a state solution has been obtained, moving to another steady-state solution with a different gate and/or drain bias can require a comparably long-time evolution of the model.

### IV. RESULTS

#### A. Ohmic mobility

In Fig. 10(a), we show our results for the field-effect electron mobility  $\mu_{\text{eff}}$  versus the electron sheet density, obtained from the electron velocity in the  $x$  direction along the channel at a longitudinal (i.e., source-to-drain) field of 1 kV/cm. The electron velocity has been calculated by mimicking what is done experimentally, i.e., by integrating along the  $z$  direction the  $x$ -directed current density and dividing it by the electron density similarly integrated along the  $z$  direction. A direct evaluation of the  $x$ -directed electron velocity was also performed, obtaining, in general, results which were identical within the statistical errors. Only at the lowest density and for simulations including only phonon scattering, some differences were observed. Since in this case the momen-

tum relaxation length is long enough to amount to a significant fraction of the channel, steady state was observed to occur only toward the drain end of the channel. Therefore, the direct evaluation of the  $x$ -directed velocity in the drain-end section of the channel was preferred. Figures 10(b) and 10(c) show the energies of the bottom of the subbands—measured from the surface potential—and the subband population at various densities. The topmost solid line in Fig. 10(b)—corresponding to the bottom of the fifth unprimed subband—marks the critical energy value  $E_{th}$  above which electrons are treated according to a bulk-transport model. The fraction of

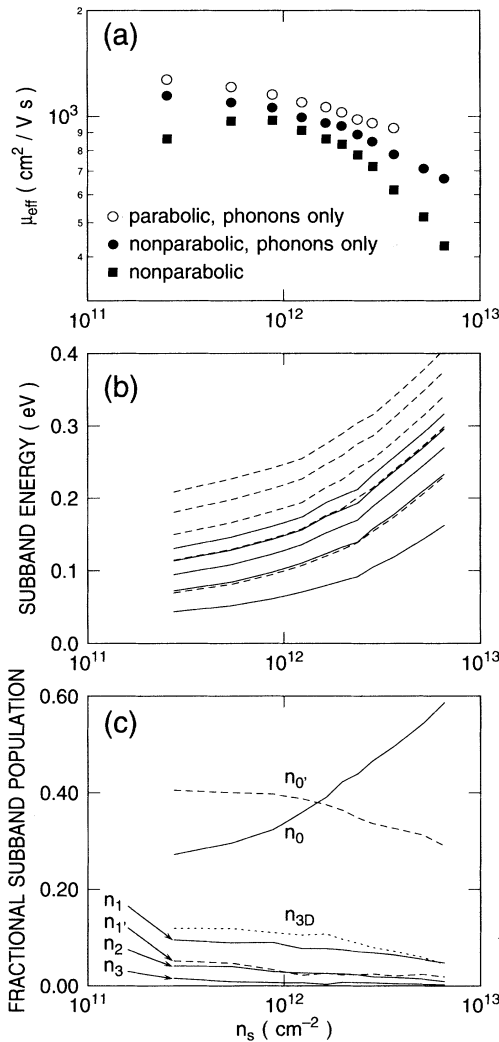


FIG. 10. The calculated field-effect electron mobility (a), subband energies measured from the surface potential (b), and the electron population in each subband (c) vs the electron channel density. Solid lines refer to the unprimed subbands, dashed lines to the primed subbands. In (a), results obtained with the inclusion of phonon scattering only in a parabolic- and nonparabolic-band model are compared to the results obtained by using a nonparabolic model and including scattering with ionized impurities and surface roughness.

electrons treated in this fashion is seen in Fig. 10(c) never to exceed 15% at the lowest densities, dropping to less than 10% for  $n_s > 10^{12} \text{ cm}^{-2}$ . In Fig. 10(a), we can see the effect due to nonparabolicity on the phonon-limited mobility (as obtained from simulations performed including phonon scattering only), and the role played by Coulomb and surface-roughness scattering at low and high densities, respectively. In Fig. 11, we have decomposed the electron mobility into its separate contributions, the phonon-limited mobility  $\mu_{ph}$ , the Coulomb mobility due to scattering with the ionized impurities in the space-charge region  $\mu_{imp}$ , and the mobility limited by scattering with surface roughness  $\mu_{SR}$ . Assuming that Matthiessen's rule holds, for the sake of illustration, we see the following trends (indicated by the dashed straight lines in the figure): The expected dependence of the phonon-limited mobility on electron density,  $\mu_{ph} \propto n_s^{-1/3}$ , is obtained only for  $n_s \gtrsim 2 \times 10^{12} \text{ cm}^{-2}$ . At this density the Fermi level is less than  $k_B T$  below the bottom of the lowest subband—so that the 2DEG approaches degeneracy—and the population of the 0 subband exceeds the population of the 0' subband. The  $n_s^{-1/3}$  behavior is expected from the fact that the form factor  $F_{\mu\nu}$  increases with density roughly as  $n_s^{1/3}$ , as originally obtained by Kawaji<sup>72</sup> for the case of a single occupied subband, and later demonstrated for the case of two subbands by Hamaguchi and co-workers.<sup>15,16</sup> At small densities the small subband energy spacing complicates the picture, since intersubband and intervalley processes are very effective. At larger densities degeneracy effects may also depress the mobility, as suggested by Kim *et al.*<sup>137</sup> (but see Ref. 138). We should also keep in mind that at very large densities ( $n_s \gtrsim 5 \times 10^{12} \text{ cm}^{-2}$ ) our treatment of screening may become inappropriate, since it is based on the nondegenerate, high-temperature approximation. Scattering with surface roughness reduces the mobility at high densities as expected,  $\mu_{SR} \propto (n_s + N_{imp})^{-2}$ , from Eqs.

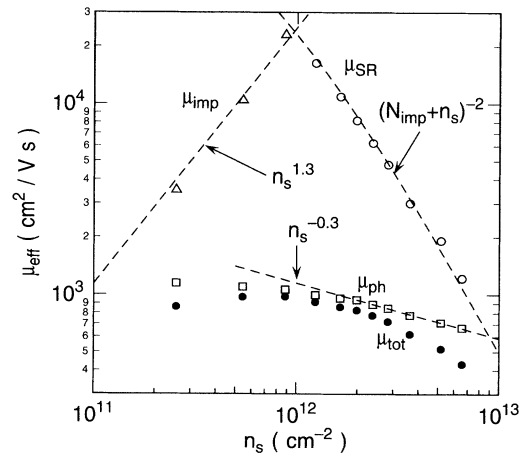


FIG. 11. The calculated field-effect electron mobility and the various contributions due to scattering with phonons, ionized impurities, and surface roughness obtained assuming that Matthiessen's rule is valid. The dashed lines are for the purpose of illustration only.



(49) and (53c). Coulomb scattering loses its strength roughly linearly with increasing electron density, as screening gains importance. However, note that the quasilinear dependence indicated in Fig. 11 is no more than suggestive because of the very few points we have calculated.

The results of Figs. 10 and 11 are encouraging, as they reproduce qualitatively all features we expected. But when we compare our results with experimental data, we see that our phonon-limited mobility is larger than the measured values. In Fig. 12, we have plotted experimental data from Refs. 3–7 in addition to unpublished Hall mobility data by Fowler reproduced in Ref. 1. Looking at the results of our simulations performed including only scattering with phonons, we see that our model predicts values consistently about 20% higher than the experimental data at the low densities at which the mobility is limited by phonon scattering. Coulomb scattering with ionized impurities and scattering with surface roughness improve the situation. Indeed, we are able to obtain very good agreement with the experimental data at high densities, while at low densities the Coulomb mobility shown in Fig. 11 compares favorably with the measured one of Ref. 7. A similarly satisfactory result related to Coulomb scattering is shown in Fig. 13, which illustrates in (a) the effect of interface charges on the effective electron mobility, and in (b) the  $N_{it}^{-1}$  dependence of the interface-charge-limited mobility  $\mu_{it}$ . These results are somewhat surprising, since we thought at first that our theoretical understanding of Coulomb and roughness scattering was in a much worse situation compared to phonon scattering, which is so well understood in bulk Si. Corrections to the impurity scattering rate obtained by going beyond the first Born approximation, the complications of treating Coulomb screening correctly, and the still largely empirical nature of the conventional treatment of surface roughness are all elements which made us uneasy about

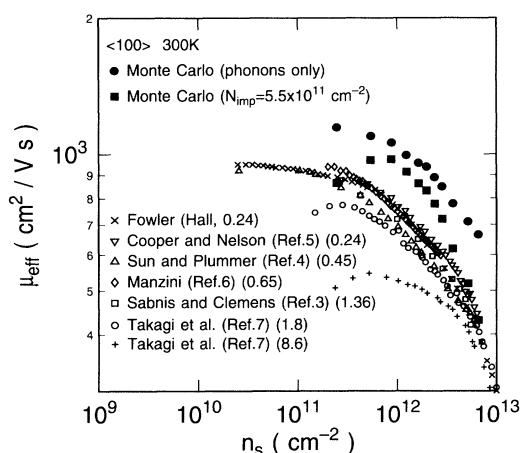


FIG. 12. The calculated field-effect electron mobility vs electron channel density compared to experimental data. The numbers in parentheses indicate the total density of ionized impurities in the depletion layer in units of  $10^{11} \text{ cm}^{-2}$  in strong inversion. Note that Fowler's data (unpublished, reproduced in Ref. 1) are Hall mobilities.

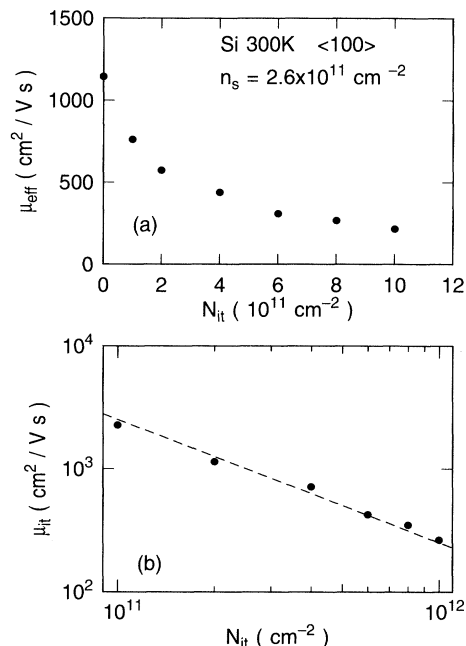


FIG. 13. The calculated field-effect electron mobility at low density in the presence of scattering with phonons and interface charges as a function of density of interface charges (a). In (b), the effective mobility due to scattering with interface charges exhibits the expected dependence  $N_{it}^{-1}$ , shown by the dashed line. Impurity scattering has not been included in this case.

the theory. Yet, considering that we employ no adjustable parameters (the values of the surface-roughness parameters  $\Delta$  and  $\Lambda$  have been taken from Ref. 108 without any modification, despite their dependence on processing), we are surprised by the excellent agreement. Instead, phonon scattering remains the real cause of concern. Figure 14 shows that the situation is even worse when plotting our results versus experimental data following the engineering custom of employing the “effective” normal field  $F_{\text{eff}} = (e/\epsilon_{\text{sc}})(n_s/2 + N_{\text{imp}})$  to “scale away” the difference of substrate doping (and the corresponding difference in subband structure) and compare a wider set of experiments. Despite the lack of a physical justification for this “universal curve,” its widespread use in the engineering community forces us to pay attention to this way of plotting results. The disagreement is worse because of the quite unusual channel doping profile we employ in our simulation.

Finally, we would like to briefly mention results we have obtained from using simulations employing the isotropic approximation to treat the deformation-potential interaction and using various forms of the static approximation to screen the interaction between electrons and acoustic phonons. The former approximation would allow us to obtain lower values for the mobility ( $\approx 850 \text{ cm}^2/\text{Vs}$  at a density of about  $3 \times 10^{11} \text{ cm}^{-2}$ ), in almost perfect agreement with data in the lower portion of the experimental range. This is expected, considering the larger intrasubband electron-phonon scattering rates



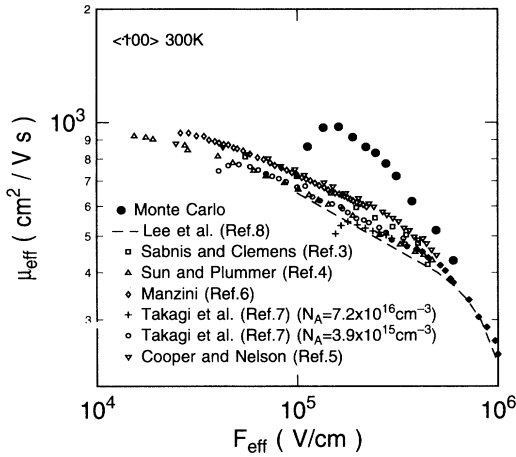


FIG. 14. The calculated field-effect electron mobility compared to experimental data as in Fig. 12, but plotted as a function of the “effective” normal field,  $F_{\text{eff}} = (e/\epsilon_{\text{sc}})(n_s/2 + N_{\text{imp}})$ .

shown in Fig. 4. The use of static screening, on the other hand, yields disastrous results: The simplified form  $\epsilon(q) = \epsilon_{\text{sc}}(1 + q_s^{(\text{DH})}/q)$  yields phonon-limited mobilities about a factor of 2 larger than those found when using unscreened phonons, while multisubband static screening, as discussed above, yields values of about 1600  $\text{cm}^2/\text{V s}$  at the lowest densities we have investigated, i.e.,

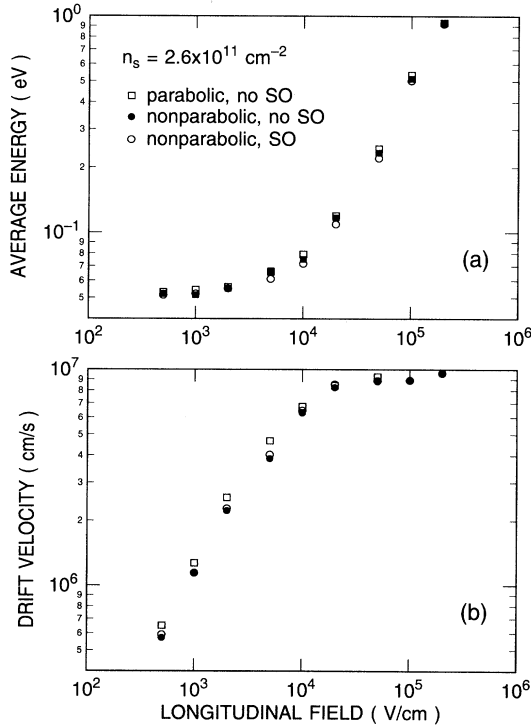


FIG. 15. The calculated expectation value of the electron kinetic energy (a), and electron drift velocity (b) vs longitudinal field at low electron density. Results obtained with and without scattering with SO modes and within the parabolic-band approximation are also shown.

about 50% larger than the unscreened values and almost a factor 2 above the experimental data.

### B. Velocity-field characteristics

In this subsection, we discuss results related to hot-electron transport at a large longitudinal field, i.e., at large source-to-drain bias. Figure 15 illustrates the average kinetic energy of the electrons (a), and their drift velocity along the  $x$  directions (b) versus the longitudinal electric field. In Fig. 16, we also show the subband energies (a) and the population in each subband (b). These data refer to simulations performed at a small value of the electron sheet density ( $n_s = 2.6 \times 10^{11} \text{ cm}^{-2}$ ), corresponding to a normal (i.e., along the direction  $z$  normal to the Si-SiO<sub>2</sub> interface) field of about 1.3 kV/cm. Only phonon scattering has been considered in these simulations. Note that the average kinetic energy of an electron in subband  $v$  has been defined in a rather unconventional way as the expectation value of its kinetic energy,  $\langle E_v(\mathbf{k}) - V(x) \rangle_v$ , rather than the average kinetic energy itself. This definition has the advantage of allowing a meaningful averaging over 2D and 3D particles which coexist in the simulation. The disadvantage is, obviously, the dependence of the equilibrium values (i.e., the “thermal” energy) on the subband structure. In Figs. 15(a) and 15(b), we

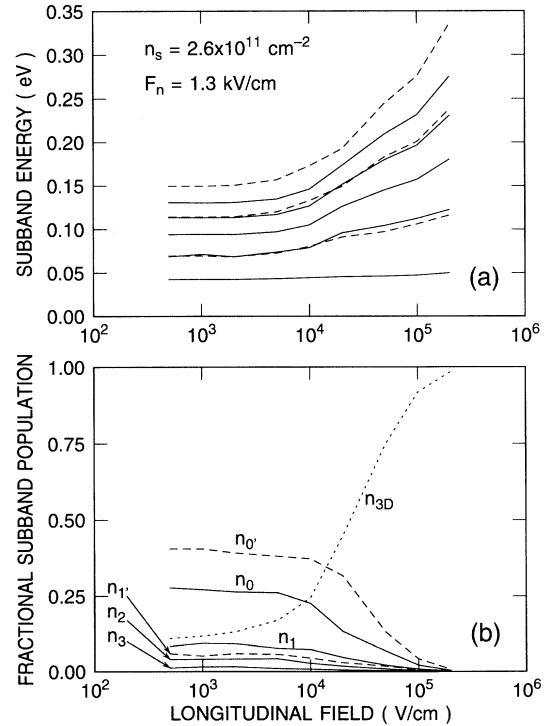


FIG. 16. Calculated subband energies (a), and electron population in each subband (b) vs longitudinal field for the conditions of Fig. 15 in a nonparabolic model including scattering with SO modes. Solid lines refer to the unprimed subbands, dashed lines to the primed subbands. Note in (a) the change of subband energies at a fixed electron density as the carriers become hot and the charge density in the inversion layer changes from its equilibrium spatial distribution.

have also shown the results of simulations performed without the inclusion of scattering with SO modes. Note that this process has a very minor effect on either average energy or velocity. The latter is actually slightly *increased* at fields around  $10^4$  V/cm, as SO scattering is introduced, as carriers are removed from the high-energy, low-velocity region of phase space and put in the low-energy, high-velocity regime by these small-angle collisions. This is exactly the opposite of what was reported by Moore and Ferry.<sup>74</sup> The discrepancy is most likely due to the anisotropic model we have employed for the deformation-potential interaction: The scattering rates at low energies are strongly reduced in this approximation when compared to the isotropic model. Thus electrons which lose a large amount of energy in collisions with SO modes enter low-energy regions of much longer momentum relaxation rates (which implies much larger velocity) than is the case in the Moore-Ferry model. In Fig. 16(b), note that at longitudinal fields larger than about  $10^4$  V/cm transport becomes almost completely bulklike, as more than half of the total electron population has a kinetic energy exceeding the threshold value we have set. This worries us to some extent. On the one hand, our approach is certainly more accurate than models which employ only a small number of subbands and 2D electrons only. As remarked above in discussing Fig. 9, and as also noted by Imanaga and Hayafuji,<sup>71</sup> the incorrect way the scattering rates are treated in this case renders the results quite suspect. On the other hand, we are unable to deal with a smooth transition from 2D to bulk transport and we must switch transport models rather abruptly at the threshold energy  $E_{th}$  [the topmost solid line in Fig. 16(a)] in order to retain a dynamics which is approximately correct and avoid the artifact of a “flat” scattering rate as we “run out of subbands.” We continue to wonder whether transport can really be described correctly by a bulk picture even at these high effective temperatures and whether the bulklike saturated velocity seen in Fig. 15(b) is caused by this approximation. We should remark here that our 3D electrons were subject to Coulomb scattering in a bulklike model, accounting for the phase-shift corrections described in Ref. 82, whenever Coulomb scattering was turned on for 2D electrons. Note in Fig. 16(b) that, as pointed out by Shihirata, Taniguchi, and Hamaguchi,<sup>28</sup> the full self-consistency between the transport model, the Poisson solver, and the Schrödinger equation is an essential ingredient for simulating high-field transport, as the subband energies change—for a fixed sheet density—due to the different spatial distribution of the hot carriers at high longitudinal fields. From Fig. 16, it is clear that any model using a purely 2D transport model and a subband structure calculated at equilibrium (i.e., at zero longitudinal field) will severely mistreat both the carrier dynamics and the subband and field configurations.

In Fig. 17, we compare our low-density results with experimental data (a) and results of previous simulations (b) obtained in samples with channel doping ranging from negligible ( $N_{imp} \approx 2 \times 10^{10} \text{ cm}^{-2}$  in Cooper and Nelson’s experiments<sup>5</sup> and  $N_{imp} \approx 8 \times 10^{10} \text{ cm}^{-2}$  in Cohen and Muller’s samples<sup>20</sup>) to moderately large ( $N_{imp} \approx 2 \times 10^{11}$

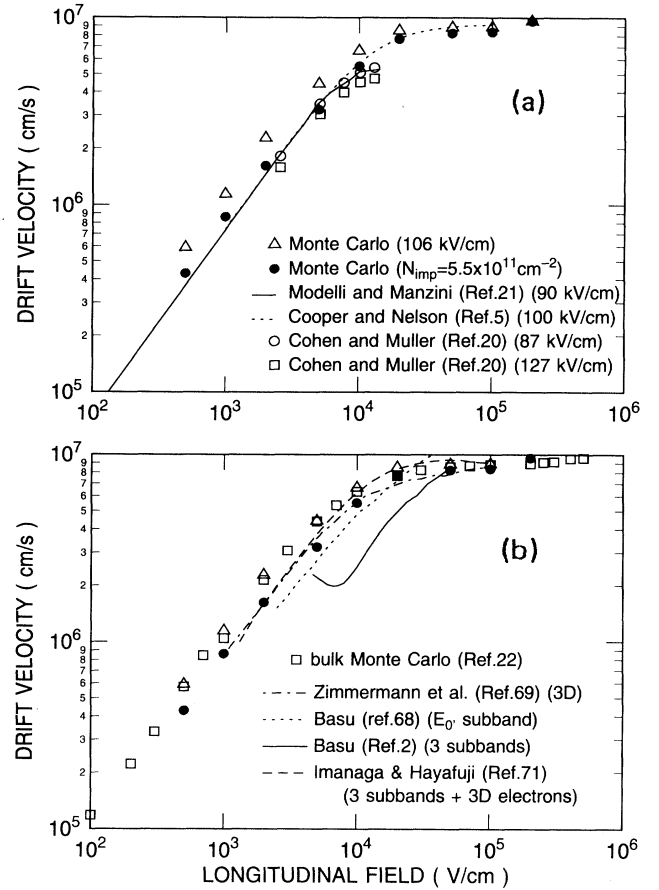


FIG. 17. The calculated electron drift velocity vs longitudinal field with and without the inclusion of scattering with ionized impurities compared to experimental data at similar values of the normal field (indicated in parentheses) (a), and compared to previous simulations and to the results of bulk simulations (b).

$\text{cm}^{-2}$  in Modelli and Manzini’s experiments<sup>21</sup>), at values of the “effective” normal field  $F_{eff}$  indicated in parentheses in the figure. Given the non-negligible amount of impurity and interface charge scattering affecting some of the experimental data, we think it is not too unfair to compare them with results of simulations performed by introducing scattering with ionized impurities ( $N_{imp} \approx 5.5 \times 10^{11} \text{ cm}^{-2}$  in our device). At low fields, we note the same disagreement between experimental data and our “phonon-scattering-only” simulations as we have already seen in the phonon-only data of Fig. 12. The inclusion of Coulomb scattering obviously improves the situation. In the velocity-saturation regime we find, as already discussed, bulklike velocity-field characteristics. These results agree with the measurements by Cooper and Nelson,<sup>5</sup> but we do not believe that this agreement settles the issue: Results by Fang and Fowler<sup>18</sup> shown in Fig. 18, by Cohen and Muller,<sup>20</sup> and by Modelli and Manzini<sup>21</sup> all give much lower values. These last measurements have been obtained using either very thick oxides<sup>18</sup>—so that the criticisms by Müller and Eisele<sup>17</sup> do not apply—or resistive gates.<sup>20,21</sup> On the con-

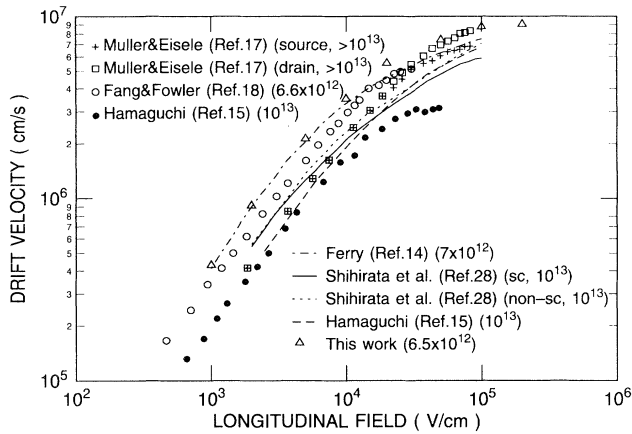


FIG. 18. The calculated electron drift velocity vs longitudinal field for a large value of the electron sheet density. Also shown are experimental data (dots) and results of previous calculations (curves) at the comparably large densities indicated in parentheses.

trary, the results of Ref. 5, obtained using a time-of-flight technique, require a quite complicated analysis to extract the data and are related to a situation of quasidepletion, rather than inversion, of the channel, as remarked in Ref. 21. We are willing to state that this issue is still unsettled and refrain from drawing strong conclusions. However, we are frankly inclined to consider the lower values as more credible, despite our failure to reproduce them. Concerning Fig. 17(b), we see that the bulk simulation of Ref. 69 predicts a bulklike saturation velocity, ruling out possible explanations of a lower channel saturation velocity based on the different energetics of the carriers,<sup>1</sup> while Basu's results, obtained using only one<sup>2</sup> or three<sup>68</sup> subbands, predict even larger velocities, as expected from the lack of 3D carriers at higher energies. Imanaga and Hayafuji's results<sup>71</sup> are similar to ours, except in the Ohmic regime, where Coulomb scattering and the different model for the deformation-potential interaction which we have employed explain the difference.

Finally, velocity-field characteristics at a larger density ( $n_s \simeq 6 \times 10^{12} \text{ cm}^{-2}$ ) are considered in Fig. 18. Comparison is made with both experimental data (indicated by symbols) and previous theoretical results (indicated by curves). The experimental data by Müller and Eisele<sup>17</sup> are split into two groups, drift velocity at the source end of the channel and at the drain end, according to their procedure. The former show a smaller saturated velocity, which they claim is due to the pinning of the longitudinal electric field at the source end of the channel in their thin-oxide (50-nm) samples. Note that these data, as well as those by Hamaguchi,<sup>15</sup> are taken at a larger electron density. The small value of the saturated velocity of Ref. 15 is surprising, since it is lower than the data obtained by Fang and Fowler<sup>18</sup> at a density not much smaller, and it is also in contrast with the small dependence of the saturated velocity on sheet density reported by Modelli and Manzini.<sup>21</sup> The theoretical calculations range from Ferry's three-subband, drifted-Maxwellian model (includ-

ing scattering with surface roughness and Coulomb scattering, both with adjustable parameters), to the full self-consistent Monte Carlo model by Shihirata, Taniguchi, and Hamaguchi.<sup>28</sup> Their non-self-consistent results are also shown, together with the results by Hamaguchi.<sup>15</sup> Our results behave quite well in the low-field Ohmic regime, thanks to the good behavior of our model for interface scattering. At high fields the situation parallels the low-density results discussed above.

### C. Density dependence

We have already discussed our results in the high-field regime and our moderate unhappiness with the inability to obtain a saturated velocity lower than the bulk value. Here we wish to discuss our results concerning the electron field-effect mobility as a function of carrier density (Figs. 10–12). These can be better appreciated considering Fig. 19. Our results, with and without the inclusion of impurity and surface-roughness scattering, are compared once more to experimental data and to results of previous investigations. We have not included in the figure the qualitative exploratory attempt by Kawaji,<sup>72</sup> and the results by Hamaguchi<sup>15</sup> and by Shihirata and Hamaguchi,<sup>16</sup> which only show that the overlap integral  $F_{\mu\nu}$  yields the correct dependence of  $\mu_{\text{eff}}$  on  $n_s$ , but require an empirical fit to obtain good quantitative agreements with the experimental data. The first complete quantitative calculations by Ezawa, Kawaji, and Nakamura<sup>10</sup> gave quite large values for the mobility, ranging from  $3500 \text{ cm}^2/\text{Vs}$  when considering only surfon scattering within the ground subband, to  $3150 \text{ cm}^2/\text{Vs}$  with the inclusion of intrasubband scattering within higher subbands, and dropping finally to about  $1900 \text{ cm}^2/\text{Vs}$  with the additional inclusion of intersubband, intravalley scattering. In Ref. 10, it was noted that the inclusion of intervalley scattering depresses the bulk Si mobility by a factor of almost 3, and that a correction of a factor of 2 could have been expected for the electron mobility in Si inversion layers. This was found to be the case

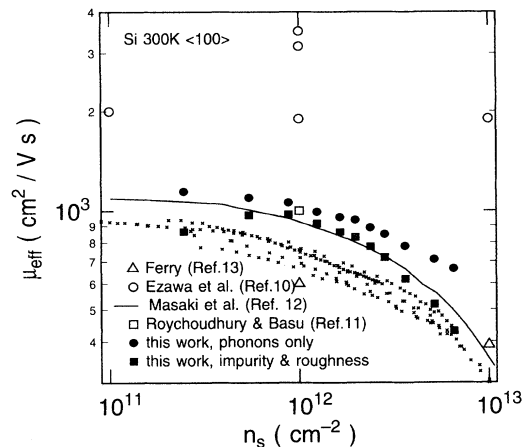


FIG. 19. Theoretical electron effective mobilities vs carrier density as obtained in this work and in previous studies, compared to the experimental data of Fig. 12 (small symbols).

by Roychoudhury and Basu,<sup>11</sup> who also accounted for the degeneracy of the 2DEG and obtained a result of  $\mu_{\text{eff}} \approx 1000 \text{ cm}^2/\text{Vs}$  at  $n_s = 10^{12} \text{ cm}^{-2}$ , which is quite consistent with our results. Ferry<sup>139</sup> introduced intervalley scattering and the interaction with first-order optical phonons to obtain good agreement with experiments. However, a very crudely approximated expression for scattering with interface charges and surface roughness had to be added, and the coupling parameters adjusted to experiments in order to achieve that agreement. A fair agreement was obtained recently by Masaki *et al.*,<sup>12</sup> particularly at high densities, when surface-roughness scattering (computed using  $\Lambda = 5 \text{ nm}$  and  $\Delta = 0.6 \text{ nm}$  in the curve we reproduce in Fig. 19) is important. They have assumed that Matthiessen's rule holds, have employed two spherical and parabolic subbands, while ignoring intersubband processes and—quite surprisingly—the strongest intervalley phonons. Impurity scattering and degeneracy effects were also ignored, but the problem was otherwise solved self-consistently. Despite these differences, their results are very close to ours, at least at densities where impurity scattering is not important: The absence in their calculations of nonparabolicity corrections appears to accidentally cancel the effect of the isotropic approximation employed for the deformation-potential interaction, both amounting to corrections of about 20% to the mobility, but acting in opposite directions. Similarly, the absence in the model of Ref. 12 of the strong *g*-type intervalley processes assisted by LO phonons appears to have an effect similar to the one caused by the lower values for the intravalley deformation potentials we have chosen. Thus, the agreement between the results of Ref. 12 and ours is purely accidental, the physical models differing substantially in the two approaches.

## V. DISCUSSION AND CONCLUSIONS

In conclusion, we have shown that the 300-K field-effect electron mobility in Si inversion layers can be explained with moderate success by accounting for all major known scattering processes (with acoustic, intravalley, and intervalley bulk phonons, with surface optical modes, with interface/oxide charges, with impurities and with interface roughness). Essential ingredients of the physical model we have employed are the band-nonparabolicity effects on the subband structure, leading to a 20% reduction of the mobility; the anisotropy of the intravalley deformation-potential interaction with acoustic phonons, leading to corrections of the same magnitude, but in the opposite direction; and a detailed analysis of screening effects. In particular, we have argued that while static, multisubband screening applies to the Coulomb interaction, a more complete dynamic picture must be employed to treat the deformation-potential interaction for a 2DEG in Si. Short of a rigorous formulation, ignoring screening altogether seems to be appropriate under the nondegenerate conditions in which phonon scattering dominates. On the other hand, our results on the hot-carrier-saturated velocity are less satisfactory. We have expressed our concern that numerical issues

(namely, the need to merge 2D and 3D transport models, as dictated by computational constraints) may actually play a non-negligible role in our computations which yield bulklike value for the saturated velocity. We have also presented a scheme to implement our physical model in general device simulations, accounting for the self-consistency between the Boltzmann, Poisson, and Schrödinger equations. Despite our efforts and partial success, fully satisfactory solutions to some problems of a general nature, such as the coupling of 2D and 3D transport models, are still lacking.

From the discussion of Sec. IV, we see that our results on the field-effect mobility are roughly consistent with the expectations expressed in Ref. 10 concerning a theory including intervalley scattering and the results of Ref. 11. The problem remains of ascertaining the importance of physical elements we have overlooked and of explaining the 20% overestimation of  $\mu_{\text{eff}}$  which affects our calculations. We have already remarked that surfon (as opposite to bulk acoustic modes) scattering may give perhaps a 5% reduction of the effective mobility. This may be compensated for by what we expect from corrections due to many-body effects: Nakamura and co-workers<sup>53,140</sup> have shown that accounting for the nonzero imaginary part of the electron self-energy due to electron-electron scattering may increase slightly the mobility. Overall, it appears that our theory is still subject to 5–10% corrections in either direction depending on the physical mechanism we add or neglect: We have seen already that nonparabolicity and anisotropy of the deformation-potential interaction have almost cancelling effects on the final result. Screening of TA phonons and antiscreening of LA phonons may also play competitive roles, enhancing and reducing the mobility by small amounts in opposite directions. Finally, we have already stressed the fact that the values of the deformation potentials themselves are somewhat uncertain, and different values may be appropriate, despite our efforts to pin down a particular set of values satisfying various experimental constraints. We conclude with the perhaps discouraging observation that we could not have expected anything better than a 20% agreement with the experimental data. The ability to improve the original calculations<sup>10</sup>—a factor of 4 too large—and obtain a modest 20% disagreement should leave us satisfied.

## ACKNOWLEDGMENTS

We have enjoyed many discussions with A. B. Fowler who has stimulated the work and originated many of the ideas used in this work. F. Stern has contributed actively by steering us throughout the relevant literature, and providing suggestions and help. One of us (M.V.F.) is particularly indebted to B. K. Ridley for conversations on the role of interface optical modes, to P. J. Price and F. Stern for conversations about screening effects on the deformation-potential interaction, to D. Ahn for his cooperation in developing the perturbative approach of Eqs. (11) and (12), and to F. Fang for a critical analysis of many experimental data. The manuscript has been critically read by P. J. Price, F. Stern, S. Tiwari, and T. N. Theis.

- <sup>1</sup>T. Ando, A. B. Fowler, and F. Stern, *Rev. Mod. Phys.* **54**, 437 (1982).
- <sup>2</sup>P. K. Basu, *Solid State Commun.* **27**, 657 (1978).
- <sup>3</sup>A. G. Sabnis and J. T. Clemens, *IEDM Technical Digest* (The Institute of Electrical and Electronics Engineers, Piscataway, NJ, 1979), p. 18.
- <sup>4</sup>S. C. Sun and J. D. Plummer, *IEEE Trans. Electron Devices* **ED-27**, 1497 (1980).
- <sup>5</sup>J. A. Cooper and D. F. Nelson, *J. Appl. Phys.* **54**, 1445 (1983).
- <sup>6</sup>S. Manzini, *J. Appl. Phys.* **57**, 411 (1985).
- <sup>7</sup>S. Takagi, M. Iwase, and A. Toriumi, *IEDM Technical Digest* (The Institute of Electrical and Electronics Engineers, Piscataway, NJ, 1988), p. 398.
- <sup>8</sup>K. Lee, J.-S. Choi, S.-P. Sim, and C.-K. Kim, *IEEE Trans. Electron Devices* **ED-38**, 1905 (1991).
- <sup>9</sup>H. Ezawa, *Ann. Phys.* **67**, 438 (1971).
- <sup>10</sup>H. Ezawa, S. Kawaji, and K. Nakamura, *Jpn. J. Appl. Phys.* **13**, 126 (1974); **14**, 921(E) (1975); H. Ezawa, *Surf. Sci.* **58**, 25 (1976).
- <sup>11</sup>D. Roychoudhury and P. K. Basu, *Phys. Rev. B* **22**, 6325 (1980).
- <sup>12</sup>K. Masaki, C. Hamaguchi, K. Taniguchi, and M. Iwase, *Jpn. J. Appl. Phys.* **28**, 1856 (1989); K. Masaki, K. Taniguchi, and C. Hamaguchi, *Semicond. Sci. Technol.* **7**, B573 (1992).
- <sup>13</sup>D. K. Ferry, *Phys. Rev. B* **14**, 5364 (1976).
- <sup>14</sup>D. K. Ferry, *Solid-State Electron.* **21**, 115 (1978).
- <sup>15</sup>C. Hamaguchi, *Physica* **134B**, 87 (1985).
- <sup>16</sup>M. Shihirata and C. Hamaguchi, *Jpn. J. Appl. Phys.* **25**, 1040 (1986).
- <sup>17</sup>W. Müller and I. Eisele, *Solid State Commun.* **34**, 447 (1980).
- <sup>18</sup>F. Fang and A. B. Fowler, *J. Appl. Phys.* **41**, 1825 (1970).
- <sup>19</sup>T. Sato, Y. Takeishi, H. Tango, H. Ohnuma, and Y. Okamoto, *J. Phys. Soc. Jpn.* **31**, 1846 (1971).
- <sup>20</sup>R. W. Cohen and R. S. Muller, *Solid-State Electron.* **23**, 35 (1980).
- <sup>21</sup>A. Modelli and S. Manzini, *Solid-State Electron.* **21**, 99 (1988).
- <sup>22</sup>M. V. Fischetti and S. E. Laux, *Phys. Rev. B* **38**, 9721 (1988).
- <sup>23</sup>S. E. Laux, M. V. Fischetti, and D. J. Frank, *IBM J. Res. Dev.* **34**, 466 (1990).
- <sup>24</sup>C. Jacoboni and L. Reggiani, *Rev. Mod. Phys.* **55**, 645 (1983).
- <sup>25</sup>P. J. Price, *Semicond. Semimet.* **14**, 249 (1979).
- <sup>26</sup>F. Stern, *Surf. Sci.* **73**, 197 (1978).
- <sup>27</sup>K. Yokoyama and K. Hess, *Phys. Rev. B* **33**, 5595 (1986).
- <sup>28</sup>M. Shihirata, K. Taniguchi, and C. Hamaguchi, *Jpn. J. Appl. Phys.* **26**, 1447 (1987).
- <sup>29</sup>K. Yokoyama and K. Hess, *J. Appl. Phys.* **59**, 3798 (1986).
- <sup>30</sup>It should be remarked explicitly that, since we solve the *one-dimensional* Schrödinger equation, we are tacitly embracing a "gradual-channel" approximation: We assume that the variation of the potential along the channel is small over an electron wavelength.
- <sup>31</sup>K. Nakamura, *Surf. Sci.* **58**, 48 (1976).
- <sup>32</sup>K. Hess and C. T. Sah, *J. Appl. Phys.* **45**, 1254 (1974); *Phys. Rev. B* **10**, 3375 (1974).
- <sup>33</sup>T. Hiroshima and R. Lang, *Appl. Phys. Lett.* **49**, 456 (1986).
- <sup>34</sup>M. Heiblum and M. V. Fischetti, in *Physics of Quantum Electron Devices*, edited by F. Capasso (Springer, Heidelberg, 1990), p. 275.
- <sup>35</sup>R. Sauer, T. D. Harris, and W. T. Tsang, *Phys. Rev. B* **34**, 9023 (1986).
- <sup>36</sup>K. Uomi, S. Sasaki, T. Tsuchiya, and M. Chinone, *J. Appl. Phys.* **67**, 90 (1990).
- <sup>37</sup>B. R. Nag, *Appl. Phys. Lett.* **59**, 1620 (1991).
- <sup>38</sup>L. M. Falicov, *Solid State Commun.* **18**, 669 (1976).
- <sup>39</sup>J. C. Slater, *Phys. Rev.* **76**, 1592 (1949).
- <sup>40</sup>J. M. Luttinger, *Phys. Rev.* **84**, 814 (1951).
- <sup>41</sup>J. M. Luttinger and W. Kohn, *Phys. Rev.* **97**, 869 (1955).
- <sup>42</sup>See also M. G. Burt, *Semicond. Sci. Technol.* **3**, 739 (1988); **3**, 1224 (1988) for a rigorous formulation of the envelope-function method.
- <sup>43</sup>M. V. Fischetti, S. E. Laux, and D. J. DiMaria, *Appl. Surf. Sci.* **39**, 578 (1989).
- <sup>44</sup>The nonparabolic perturbative approach has been developed by D. Ahn, and reported together with the pseudopotential approach in M. V. Fischetti and D. Ahn, IBM Research Report RC 14801 7/25/89 (unpublished).
- <sup>45</sup>J. R. Chelikowsky and M. L. Cohen, *Phys. Rev. B* **14**, 556 (1976).
- <sup>46</sup>S. E. Laux and M. V. Fischetti, *IEEE Electron Device Lett.* **EDL-9**, 467 (1988).
- <sup>47</sup>Density of states and electron-phonon-scattering rates employing a numerically tabulated band structure have been employed by us before in bulk semiconductors, as presented in Ref. 22. Density of states and scattering rates, in that case, could be computed once and for all and stored, since they do not depend on variables associated with the bias conditions and geometric configuration of the device under study. In the present case, unfortunately, eigenvalues and eigenfunctions of Eq. (13) depend on these variables. Therefore, while feasible in principle and easier than what was done in Ref. 22, an evaluation of density of states and scattering rates for a "numerical" subband structure is computationally prohibitive.
- <sup>48</sup>T. Ando, *Surf. Sci.* **58**, 128 (1976).
- <sup>49</sup>B. Vinter and T. Stern, *Surf. Sci.* **58**, 141 (1976).
- <sup>50</sup>T. Ando, *Phys. Rev. B* **13**, 3468 (1976).
- <sup>51</sup>B. Vinter, *Phys. Rev. B* **13**, 4447 (1976).
- <sup>52</sup>B. Vinter, *Solid State Commun.* **32**, 651 (1979).
- <sup>53</sup>K. Nakamura, H. Ezawa, and K. Watanabe, *Phys. Rev. B* **22**, 1892 (1980).
- <sup>54</sup>S. Das Sarma and B. Vinter, *Phys. Rev. B* **23**, 6832 (1981).
- <sup>55</sup>S. Das Sarma and B. Vinter, *Phys. Rev. B* **26**, 960 (1982).
- <sup>56</sup>F. Stern, unpublished results reproduced in Ref. 1.
- <sup>57</sup>P. J. Price, *Ann. Phys.* **133**, 217 (1981).
- <sup>58</sup>B. K. Ridley, *Rep. Prog. Phys.* **54**, 169 (1991).
- <sup>59</sup>C. Herring and E. Vogt, *Phys. Rev.* **101**, 944 (1956).
- <sup>60</sup>C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, and A. Alberigi-Quaranta, *Phys. Rev. B* **12**, 2265 (1975).
- <sup>61</sup>I. Baslev, *Phys. Rev.* **143**, 636 (1966).
- <sup>62</sup>K. Murase, K. Enjouji, and K. Otsuka, *J. Phys. Soc. Jpn.* **29**, 1248 (1970).
- <sup>63</sup>V. J. Tekippe, H. R. Chandrasekhar, P. Fischer, and A. K. Ramdas, *Phys. Rev. B* **6**, 2348 (1972).
- <sup>64</sup>J. C. Merle, M. Capizzi, P. Fiorini, and A. Frova, *Phys. Rev. B* **17**, 4821 (1978).
- <sup>65</sup>Note that Ref. 62 reports a value of  $-0.67$  for the ratio  $D_{\Sigma}$ .
- <sup>66</sup>L. Laude, F. H. Pollak, and M. Cardona, *Phys. Rev. B* **3**, 2623 (1971).
- <sup>67</sup>M. V. Fischetti and J. M. Higman, in *Monte Carlo Device Simulation: Full Band and Beyond*, edited by K. Hess (Kluwer Academic, Norwell, 1991), p. 123.
- <sup>68</sup>P. K. Basu, *J. Appl. Phys.* **48**, 350 (1977).
- <sup>69</sup>J. Zimmermann, R. Fauquembergue, M. Charef, and E. Constant, *Electron Lett.* **16**, 665 (1980).
- <sup>70</sup>Chu-Hao, J. Zimmermann, M. Charef, R. Fauquembergue, and E. Constant, *Solid-State Electron.* **28**, 733 (1985).
- <sup>71</sup>S. Imanaga and Y. Hayafuji, *J. Appl. Phys.* **70**, 1522 (1991).
- <sup>72</sup>S. Kawaji, *J. Phys. Soc. Jpn.* **27**, 906 (1969).
- <sup>73</sup>K. Hess and P. Vogl, *Solid State Commun.* **30**, 807 (1979).

- <sup>74</sup>B. T. Moore and D. K. Ferry, *J. Appl. Phys.* **51**, 2603 (1980).
- <sup>75</sup>B. T. Moore and D. K. Ferry, *J. Vac. Sci. Technol.* **17**, 1037 (1980).
- <sup>76</sup>S. Q. Wang and G. D. Mahan, *Phys. Rev. B* **6**, 4517 (1972).
- <sup>77</sup>I. Yokota, *J. Phys. Soc. Jpn.* **16**, 2075 (1961).
- <sup>78</sup>B. B. Varga, *Phys. Rev.* **137**, A1896 (1965).
- <sup>79</sup>H. R. Chandrasekhar and A. K. Ramdas, *Phys. Rev. B* **21**, 1511 (1980).
- <sup>80</sup>M. A. Hollis, S. C. Palmateer, L. F. Eastman, N. V. Dandekar, and P. M. Smith, *IEEE Electron Device Lett.* **EDL-4**, 440 (1983).
- <sup>81</sup>J. R. Hayes, A. F. Levi, and W. Weigmann, *Appl. Phys. Lett.* **48**, 1365 (1986).
- <sup>82</sup>M. V. Fischetti, *Phys. Rev. B* **44**, 5527 (1991).
- <sup>83</sup>R. Sirko and D. L. Mills, *Phys. Rev. B* **18**, 4373 (1978).
- <sup>84</sup>J. R. Lowney and H. S. Bennett, *J. Appl. Phys.* **69**, 7102 (1991).
- <sup>85</sup>N. S. Mansour, K. Diff, and K. F. Brennan, *J. Appl. Phys.* **70**, 6854 (1991).
- <sup>86</sup>R. H. Ritchie, *Phys. Rev.* **106**, 874 (1957).
- <sup>87</sup>R. A. Ferrell, *Phys. Rev.* **111**, 1214 (1958).
- <sup>88</sup>E. A. Stern and R. A. Ferrell, *Phys. Rev.* **120**, 130 (1960).
- <sup>89</sup>E. N. Economou, *Phys. Rev.* **182**, 539 (1969).
- <sup>90</sup>K. L. Ngai and E. N. Economou, *Phys. Rev. B* **4**, 2132 (1971).
- <sup>91</sup>R. Fuchs and K. L. Kliewer, *Phys. Rev.* **140**, A2076 (1965).
- <sup>92</sup>K. L. Kliewer and R. Fuchs, *Phys. Rev.* **144**, 495 (1966).
- <sup>93</sup>K. L. Kliewer and R. Fuchs, *Phys. Rev.* **150**, 573 (1966).
- <sup>94</sup>See the review by Ridley (Ref. 58) for recent work on interface and slab phonons with hydrodynamics and electromagnetic boundary conditions.
- <sup>95</sup>A. Caillé, M. Banville, and M. J. Zuckermann, *Solid State Commun.* **24**, 805 (1977).
- <sup>96</sup>J. Gersten, *Surf. Sci.* **92**, 579 (1980); **97**, 206 (1980).
- <sup>97</sup>A. Caillé and M. Banville, *Solid State Commun.* **19**, 951 (1976).
- <sup>98</sup>According to Ridley (Ref. 58), we are committing a sin here. Electromagnetic boundary conditions are known not to yield the right result, while hydrodynamic boundary conditions, as given by M. Babiker, *J. Phys. C* **19**, 683 (1986), approximate much better the results of microscopic models when there is a small mismatch of the optical-phonon frequency across the interface, as it is the case for the GaAs/Ge system [see H. Aker and T. Ando, *Phys. Rev. B* **40**, 2914 (1989)]. In our case, it is not clear whether even mechanical boundary conditions will work and we simply follow history, perhaps blindly and acritically, in using the Fuchs-Kliewer approach. Finally, we do not differentiate between the *interface phonons*, coupled to the electrons by a strong Fröhlich scalar coupling, and *interface polaritons*, coupled to the electrons rather weakly via a  $\mathbf{p} \cdot \mathbf{A}$  interaction, according to Ref. 58 and M. Babiker, *Semicond. Sci. Technol.* **7**, B52 (1992). This is still a controversial issue, debated by B. K. Ridley and M. Babiker, *Phys. Rev. B* **43**, 9096 (1991) on one side, and Gerald Weber, *Phys. Rev. B* **46**, 12 792 (1992) and Bang-fen Zhu, *Phys. Rev. B* **46**, 13 619 (1992) on the other.
- <sup>99</sup>D. A. Dahl and L. J. Sham, *Phys. Rev. B* **16**, 651 (1977).
- <sup>100</sup>In SiO<sub>2</sub>, the condition  $\omega_{\text{TO},2} \gg \omega_{\text{TO},1}$  is verified only approximately, as Table I shows. Therefore, the zeros of Eq. (32) are obtained by the expressions for  $\omega_{\text{LO},i}$  given in the text only within a few meV. The same considerations apply to the expressions given for the "flat" region of the dispersion of the SO modes.
- <sup>101</sup>See also Ref. 90, Eqs. (2)–(23).
- <sup>102</sup>M. E. Kim, A. Das, and S. D. Senturia, *Phys. Rev. B* **18**, 6890 (1978).
- <sup>103</sup>B. K. Ridley, *Superlatt. Microstruct.* **2**, 159 (1986).
- <sup>104</sup>C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963), p. 137ff.
- <sup>105</sup>S. E. Laux and M. V. Fischetti, in *Monte Carlo Device Simulation: Full Band and Beyond*, edited by K. Hess (Kluwer Academic, Norwell, 1991), p. 1.
- <sup>106</sup>F. Stern and W. E. Howard, *Phys. Rev.* **163**, 816 (1967).
- <sup>107</sup>See the brief review given in Ref. 1 and the relative references therein.
- <sup>108</sup>S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Liliental, D. Fathy, and O. L. Krivanek, *Phys. Rev. B* **32**, 8171 (1985).
- <sup>109</sup>S. M. Goodnick, R. G. Gann, J. R. Sites, D. K. Ferry, and C. W. Wilmsen, *J. Vac. Sci. Technol. B* **1**, 803 (1983).
- <sup>110</sup>R. E. Prange and T.-W. Nee, *Phys. Rev.* **168**, 779 (1968).
- <sup>111</sup>Y. Matsumoto and Y. Uemura, *Jpn. J. Appl. Phys. Suppl.* **2**, Pt. 2, 367 (1974) have discussed an analogous problem when using variational wave functions. See also P. J. Price and F. Stern, *Surf. Sci.* **132**, 577 (1983).
- <sup>112</sup>Jia Li and T.-P. Ma, *J. Appl. Phys.* **62**, 4212 (1987).
- <sup>113</sup>C. T. Sah, T. H. Ning, and L. L. Tschoop, *Surf. Sci.* **32**, 561 (1972).
- <sup>114</sup>T. H. Ning and C. T. Sah, *Phys. Rev. B* **6**, 4605 (1972).
- <sup>115</sup>F. Stern, *Phys. Rev. Lett.* **18**, 546 (1967).
- <sup>116</sup>E. D. Siggia and P. C. Kwok, *Phys. Rev. B* **2**, 1024 (1970).
- <sup>117</sup>A. L. Fetter, *Phys. Rev. B* **10**, 3739 (1974).
- <sup>118</sup>A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971), p. 305.
- <sup>119</sup>B. D. Fried and S. D. Conte, *The Plasma Dispersion Function* (Academic, New York, 1961).
- <sup>120</sup>A similar equation has been given by Stern and Howard (Ref. 106) and later used in the work by Yokoyama and Hess (Ref. 27), but with a slightly different approximation for the screening parameter  $q_s(q)$ , by ignoring its  $q$  dependence. Solutions of this equation assuming either zero thickness for the inversion charge, or a sinusoidal wave function, have been given by K. Hess, *Appl. Phys. Lett.* **35**, 485 (1979) and by J. Lee, N. H. Spector, and V. K. Arora, *J. Appl. Phys.* **54**, 6995 (1983), respectively.
- <sup>121</sup>M. J. Ziman, *Electrons and Phonons* (Clarendon, Oxford, 1960).
- <sup>122</sup>H. N. Spector, *Solid State Phys.* **19**, 291 (1966).
- <sup>123</sup>M. L. Cohen, in *Superconductivity*, edited by R. D. Parks (Dekker, New York, 1969).
- <sup>124</sup>P. J. Price, *Phys. Rev. B* **32**, 2643 (1985).
- <sup>125</sup>W. Walukiewicz, H. E. Ruda, J. Lagowski, and H. C. Gatos, *Phys. Rev. B* **32**, 2645 (1985).
- <sup>126</sup>P. F. Maldague, *Surf. Sci.* **73**, 296 (1978).
- <sup>127</sup>T. Kawamura and S. Das Sarma, *Phys. Rev. B* **45**, 3612 (1992).
- <sup>128</sup>P. Vogl, in *Physics of Nonlinear Transport in Semiconductors*, edited by D. K. Ferry, J. R. Barker, and C. Jacoboni, Vol. 52 of *NATO Advanced Study Institute, Series B: Physics* (Plenum, New York, 1980), p. 75.
- <sup>129</sup>P. Boguslawski and J. Mycielski, *J. Phys. C* **10**, 2413 (1977).
- <sup>130</sup>A. W. Overhauser, *Phys. Rev. B* **3**, 1888 (1970); B. I. Lundqvist, *Phys. Kondens. Mater.* **6**, 193 (1967); **6**, 209 (1967).
- <sup>131</sup>P. J. Price, *J. Vac. Sci. Technol.* **19**, 599 (1981).
- <sup>132</sup>Here we are assuming that  $K_{\text{sc}}^2 \equiv q^2 - \epsilon_{\text{sc}} \omega^2 / (\epsilon_0 c^2) \approx q^2$ , which is valid for  $q \gg 1.8 \times 10^3 \text{ cm}^{-1}$  when  $\omega$  is set equal to the LA-phonon frequency at  $\mathbf{Q} = (\mathbf{q}, \bar{q}_0)$ , with  $\bar{q}_0 \approx 1.7 \times 10^7 \text{ cm}^{-1}$ , as obtained from Eq. (72) of the text in our case.
- <sup>133</sup>The expression for the bulk dynamic polarizability given in

Ref. 118 is also valid in the case of a 2DEG (except for an overall factor  $2/\lambda_v$ ), as noted by Fetter in Ref. 117 for the static polarizability.

- <sup>134</sup>It is known that in bulk materials collisional broadening cannot be accounted for by simply replacing  $\omega$  in Eq. (73) with  $\omega + i/\tau$ , where  $\tau$  is the relaxation time: this would violate conservation of particle number. An extension to 2DEG's of the result given by N. D. Mermin, Phys. Rev. B **1**, 2362 (1970) is required. We are indebted to F. Stern for bringing this point to our attention.

- <sup>135</sup>A. M. Kriman and P. P. Ruden, Phys. Rev. B **32**, 8013 (1985).

- <sup>136</sup>M. V. Fischetti and S. E. Laux, IEEE Trans. Electron Devices **38**, 650 (1991).

- <sup>137</sup>D. M. Kim, F. Qian, Y. Tanng, and J. S. Blakemore, Appl. Phys. Lett. **56**, 2638 (1990).

- <sup>138</sup>G. Ghibaudo, Appl. Phys. Lett. **58**, 313 (1991).

- <sup>139</sup>D. K. Ferry, Surf. Sci. **57**, 218 (1976).

- <sup>140</sup>K. Nakamura, K. Watanabe, and H. Ezawa, Surf. Sci. **98**, 202 (1980).