RC 22092 (99048) June 12, 2001 **Physics**

IBM Research Report

Effective Electron Mobility in Si Inversion Layers in MOS Systems with a High-K Insulator: The Role of Remote Phonon Scattering

Massimo Fischetti, Deborah Neumayer, Eduard Cartier

IBM T. J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

LIMITED DISTRIBUTION NOTICE: This report has been submitted for publication outside of IBM and will probably be copyrighted if accepted for publication. It has been issued as a Research Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g. payment of royalties). Copies may be requested from IBM T. J. Watson Research Center , P. O. Box 218, Yorktown Heights, NY 10598 USA (email: reports@us.ibm.com). Some reports are available on the internet at http://domino.watson.ibm.com/library/CyberDig.nsf/home

Effective electron mobility in Si inversion layers in MOS systems with a high-κ insulator: The role of remote phonon scattering

Massimo. V. Fischetti, Deborah A. Neumayer, and Eduard A. Cartier
IBM Research Division, Thomas J. Watson Research Center
P. O. Box 218, Yorktown Heights, NY 10598, USA
(June 26, 2001)

The high dielectric constant of insulators currently investigated as alternatives to SiO_2 in metal-oxide-semiconductor structures is due to their large ionic polarizability. This is usually accompanied by the presence of soft optical phonons. We show that the long-range dipole field associated with the interface excitations resulting from these modes and from their coupling with surface plasmons, while small in the case of SiO_2, for most high-κ materials causes a reduction of the effective electron mobility in the inversion layer of the Si substrate. We study the dispersion of the interfacial coupled phonon-plasmon modes, their electron-scattering strength, and their effect on the electron mobility for Si-gate structures employing films of SiO_2, Al_2O_3, AlN, ZrO_2, HfO_2, and $ZrSiO_4$ for 'SiO_2-equivalent' thicknesses ranging from 5 nm to 0.5 nm.

I. INTRODUCTION

The study of insulators with a static dielectric constant higher than that of SiO_2 (so called 'high-κ' materials) is motivated by the necessity of increasing the gate capacitance of Si metal-oxide-semiconductor field-effect transistors (MOSFETs), while avoiding the problems which arise when the SiO_2 thickness is reduced below the 1.5-1.0 nm range, as demanded by device scaling[1,2]. While the focus of these efforts is, at least at present, mainly aimed at improving the chemical and physical properties of the insulating materials, in this paper we point out an intrinsic property of these materials, undesirable but hopefully not 'fatal', a property which would be present even if (or, more optimistically, when) 'ideal' high-κ films (*i.e.*, stoichiometric, thermally stable, free from electron and interface traps, leakage-free, reliable, reproducible, etc.) were (or will be) available: Namely, the reduction of the effective mobility of electrons in the inversion layer of Si MOSFETs fabricated using these dielectrics.

Unfortunately the origin of this undesirable property is intrinsically related to the high-κ itself. Before delving into technical details, it is worth outlining the basic physical elements which are the cause of the problem. The dielectric constant of a (non metallic) solid results from the contribution of the ionic and the electronic polarization. The latter scales roughly with the inverse of the (direct) band gap of the solid, averaged over the Brillouin Zone. Since an 'insulator' is, by definition, a large-gap material, there is little room left to increase the electronic polarization. Therefore, a higher dielectric constant can only originate from a larger ionic polarization. Indeed, in most of the high-κ materials being considered, the large (static) dielectric constant is due to highly polarizable – often metal-oxygen – bonds. It is the polarization of these 'soft' bonds which screens the external field, and results in the desired high κ. Associated with 'soft' bonds are low-energy lattice oscillations (phonons), 'optical' in nature because of the ionic character of the atomic bonds in most insulators. By contrast, the 'hard' Si-O bonds in SiO_2 yield a reduced ionic polarization. Associated with 'hard' bonds are 'hard' (that is, 'stiff') optical phonons.

As early as 1972, Wang and Mahan[3] showed that electrons in the inversion layer at the interface between a semiconductor of optical permittivity ϵ_s^∞ and a dielectric of static and optical permittivities ϵ_{ox}^0 and ϵ_{ox}^∞, respectively, can couple with the surface-optical (SO) modes (arising at the insulator/Si interface from the longitudinal-optical (LO) modes of the insulator) with a coupling strength proportional to

$$\hbar\omega_{SO}\left[\frac{1}{\epsilon_s^\infty+\epsilon_{ox}^\infty}-\frac{1}{\epsilon_s^\infty+\epsilon_{ox}^0}\right], \qquad (1)$$

where $\hbar$ is the reduced Planck constant and ω_{SO} is the frequency of the SO insulator-phonon, given by Eq. (40) below. Equation (1) is physically equivalent to the well-known Fröhlich electron/LO-phonon scattering strength, proportional to $\hbar\omega_{LO}[1/\epsilon^\infty - 1/\epsilon^0]$ in a material with static and optical permittivities ϵ^0 and ϵ^∞, respectively, and LO-phonon frequency ω_{LO}: The difference between the inverse of ϵ^0 and of ϵ^∞ is proportional to the squared-amplitude of the dipole field solely due to the oscillating ionic polarization of the material; that is, to the coupling between electrons and the bulk LO-phonons. Equation (1) has a similar nature, the coupling constant being modified by 'image-charge effects' at the insulator/semiconductor interface, which affect the decay of the dipole field of the insulator-phonons away from the bulk of the insulator into the semiconductor inversion layer. The effect of this scattering mechanism, called 'remote phonon scattering', on hot-electron transport in the Si/SiO_2 system was later studied by Hess and Vogl[4] and by Moore and Ferry[5], and its effect on the effective electron mobility by one of us (MVF)[6]. For the Si/SiO_2 system, and restricting our attention – now and throughout the rest of the paper – to the electron mobility, remote scattering does not play a major role. The reason for this is twofold. First, as we said above, the ionic polarizability of SiO_2 is not very large, because of the 'hard' nature of the Si-O bond. On the one hand, this results in an 'undesired' small static dielectric constant. On the other hand, this also results in a small difference between ϵ_{ox}^0 and ϵ_{ox}^∞, and so in a small electron/remote-phonon scattering strength. Second, the stiff Si-O bond also results

in a large LO (and SO) phonon energy ($\hbar\omega_{LO} \approx 0.15$ eV). Thermal electrons cannot emit excitations of such a large energy, and at room temperature there are too few thermally-excited phonons to be absorbed. (Another bulk SiO_2 phonon with $\hbar\omega_{LO} \approx 63$ meV could potentially have a larger effect, but its oscillator strength is too small: If it were not so, SiO_2 would have a significantly larger κ!).

Getting finally to the problem at hand, the high-frequency dielectric response of a high-κ insulator is mainly electronic – since heavier and 'slower' ions cannot respond fully at sufficiently large frequencies –and so it is not too different from that of SiO_2, provided the band gap is large enough. On the contrary, the large ionic response dominates at low frequency. This does indeed result in the desired large static dielectric constant, but also in both a large difference between ϵ^0_{ox} and ϵ^∞_{ox}, and so in a large scattering strength, Eq. (1), and in a low SO-phonon frequency, which triggers frequent emissions and absorption processes by thermal electrons. The net result is that the very same physical properties which are responsible for the higher κ of the insulator are also likely to yield (with some important exceptions we shall consider later) a degradation of the effective electron mobility in the inversion layers of MOS-systems using the high-κ insulator. This concept was very clear to Hess and Vogl[4], who concluded their article with the optimistic remark: *"In passing, we note that a reduction of the ionic polarizability of SiO_2, or better of the difference $\epsilon^0_{ox} - \epsilon^\infty_{ox}$, would reduce the electron-phonon coupling,* [...] *and correspondingly, enhance the field dependent electron mobility in MOS transistors"*. Unfortunately, as we move from SiO_2 to higher-κ materials with a higher difference $\epsilon^0_{ox} - \epsilon^\infty_{ox}$, we achieve the opposite effect of depressing the mobility.

To give a first idea of the size of the effect we are considering, we show in Fig. 1 the effective electron mobility in the inversion layer at the interface between Si and an infinitely-thick film of several insulators we have considered (SiO_2, HfO_2, ZrO_2, $ZrSiO_4$, AlN, and Al_2O_3). Full details will be given below. For now it suffices to say that the triangular well approximation has been used to treat the inversion layer, an anisotropic and non-parabolic band-structure model has been used to account for (anisotropic) scattering with acoustic phonons, as described in Ref. 7, surface roughness has been accounted for empirically using Matthiessen's rule, and scattering with remote SO modes had been treated using Fermi's Golden Rule with the Wang-Mahan matrix element proportional to the scattering strength given in Eq. (1). While this model is excessively oversimplified for the reasons stated below, it shows that effects as large as a factor of 2 or more can be expected.

The results shown in Fig. 1 are – in principle – no more than suggestive of what we should expect. In order to quantify more accurately the importance of remote phonon scattering in 'realistic' high-κ MOS systems, we must consider two additional complications related to the full dielectric response of the whole system. First, surface/interface optical modes can couple quite strongly – in the appropriate range of electron density, n_s, in the

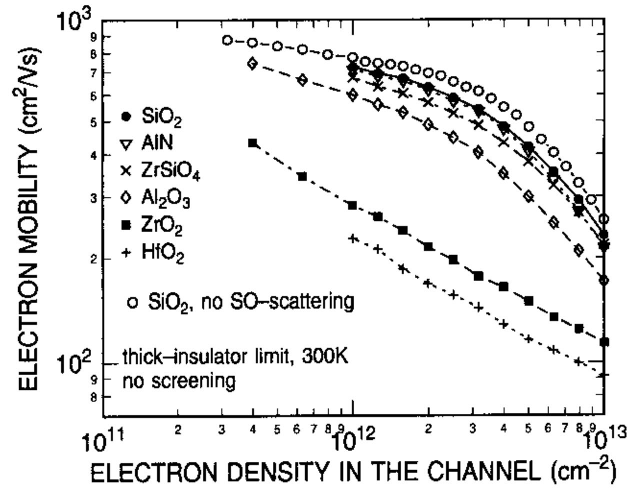

FIG. 1. Effective electron mobility in Si inversion layers of MOS systems with the insulators indicated. A triangular well approximation has been used to model the subband structure of the inversion layer. Anisotropic scattering with acoustic phonons and remote scattering with surface optical phonons has been accounted for (when indicated). Scattering with surface roughness has been added empirically using Matthiessen's rule and fitting the roughness-parameters to match the experimental 'universal' mobility for the SiO_2 system at an electron sheet density of 10^{13} cm^{-2}. The limit of infinite insulator thickness has been taken, and no additional dielectric effects (screening by substrate and gate electrons, plasmon-phonon coupling) have been considered here.

inversion layer – with the two-dimensional plasmons at the insulator/semiconductor interface. The second effect, most important in the thin-insulator MOS structures of present interest, is the coupling of insulator-optical-modes, interface substrate plasmons, and interface plasmons in the gate, typically consisting of heavily doped, n-type, poly-crystalline Si (poly-Si for short). We have previously investigated the coupling of substrate and gate interface plasmons, finding that a significant gate Coulomb-drag yields by itself a reduction of the electron mobility for SiO_2 layers thinner than about 2-1.5 nm.[7] Here we extend the treatment by including the coupling of surface plasmons to surface-optical modes, by accounting for electron scattering with the resulting phonon-like modes, and considering MOS systems with various thicknesses of different insulators selected among those of practical technological importance, thus covering a spectrum of medium-to-high dielectric constants. Anticipating our main result, the proximity of the heavily-doped gate has the beneficial effect of screening to a large extent the interaction between electrons and interface optical modes at the smallest insulator thickness considered for all but two (HfO_2 and ZrO_2) of the high-κ insulators we have considered, and at sufficiently large electron density in the depleted poly-Si gate.

In Sec. II we present our theoretical scheme. In Sec. III we present our results, discussing first, in Sec. III B, the non-trivial problem of selecting physical quantities (LO and/or transverse-optical, or TO, phonon energies, dielectric constants, and oscillator strengths) of the insulating films, comparing information available in the litera-

ture with data extracted from Fourier transform infrared spectroscopy. Finally, after a discussion in Sec. IV, in Appendix A we present some estimates about the role played by an interfacial SiO_2 layer. The role of remote phonon scattering on hot-electron (*i.e.* non Ohmic) transport will not be investigated here, although we should expect significant effects, along the line of our previous investigation of long-range Coulomb effects on the transconductance of Si n-MOSFETs (Ref. 8).

II. INTERFACE COUPLED PLASMON-PHONON MODES

Our goal is to calculate the dispersion of the coupled insulator-phonon/semiconductor-plasmon modes in an MOS system and the amplitude of the (scattering) field associated with each mode. The general framework has been already described in detail in Ref. 7, where we have discussed their dispersion in a general case, but have ignored the phonons in discussing the scattering field. Therefore, here we shall only summarize the results as far as the dispersion of the interface modes is concerned, but we shall discuss in detail the derivation of the plasmon- and phonon-content of each mode, as well as their scattering strength, because of the additional complication caused by the presence of the optical modes in the insulator.

A. Dispersion

Here, as in Ref. 7, we shall consider a structure consisting of degenerately-doped n-type Si (mimicking the poly-Si gate) in the half-space $z < 0$ (gate), an SiO_2 or high-κ insulating layer for $0 \leq z < t$, and a p-type Si substrate filling the half-space $z \geq t$. The latter is assumed to be inverted, and so it is treated as a two-dimensional electron gas (2DEG). The dielectric response of the gate, insulator, and substrate is described by the dielectric functions (in the long-wavelength limit discussed below) $\epsilon_g(\omega)$, $\epsilon_{ox}(\omega)$, and $\epsilon_s(Q,\omega)$, respectively. We denote by $\mathbf{Q}$ and $\mathbf{R}$ the two-dimensional wave vector and coordinate-vector in the (x,y)-plane of the interfaces, respectively.

We are interested in longitudinal electric eigenmodes of the system, since transverse modes (given by poles of the total dielectric response) correspond to a vanishing electric field, and so to a vanishing coupling with the charge carriers. These are transverse-magnetic solutions (TM- or p-waves) of Maxwell's equations. As described in Ref. 7, we can safely work in the non-retarded limit. Thus, the 'usual' boundary conditions require that the components of the electric field on the plane of the interfaces be continuous across the two interfaces at $z = 0$ and $z = t$, and similarly for the component of the displacement field normal to the plane of the interfaces. We can expand the electrostatic potential at frequency ω as:

$$\phi(\mathbf{R},z,t) = \sum_{\mathbf{Q}} \phi_{Q,\omega}(z)\, e^{i\mathbf{Q}\cdot\mathbf{R}}\, e^{i\omega t} , \quad (2)$$

where it must be understood that we shall take the real part of the complex exponentials here and in the following. Since in the following we assume an isotropic dielectric response everywhere, thanks to the cylindrical symmetry of the problem, $\phi_{Q,\omega}(z)$ depends only on the magnitude of the wave vector $\mathbf{Q}$. Employing the procedure and approximations described in Appendix B of Ref. 8, finding solutions of Mawell's equations reduces to finding the solution of the Laplace equation:

$$\frac{d^2\phi_{Q,\omega}(z)}{dz^2} - Q^2\phi_{Q,\omega}(z) = 0 . \quad (3)$$

The boundary conditions at the interfaces imply that a nontrivial physically acceptable (*i.e.*, finite as $z \to \pm\infty$) solution of Eq. (3) exists provided we satisfy the secular equation

$$e^{Qt}[\epsilon_g(\omega) + \epsilon_{ox}(\omega)][\epsilon_s(Q,\omega) + \epsilon_{ox}(\omega)] - e^{-Qt}[\epsilon_g(\omega) - \epsilon_{ox}(\omega)][\epsilon_s(Q,\omega) - \epsilon_{ox}(\omega)] = 0 , \quad (4)$$

which we can rewrite in the equivalent form:

$$\epsilon_{ox}(\omega)^2 + \epsilon_{ox}(\omega)[\epsilon_g(\omega) + \epsilon_s(Q,\omega)] \,\mathrm{cotanh}(Qt) + \epsilon_g(\omega)\epsilon_s(Q,\omega) = 0 . \quad (5)$$

By solving this equation we obtain the dispersion of the modes, $\omega(Q) = \omega_Q^{(i)}$, where the index i runs over the branches of the dispersion. In this case, the solution $\phi_{Q,\omega}(z)$ has the form:

$$\phi_{Q,\omega_Q^{(i)}}(z) = \begin{cases} a_{Q,\omega_Q^{(i)}}\, e^{Qz} & (z < 0) \\ b_{Q,\omega_Q^{(i)}}\, e^{-Qz} + c_{Q,\omega_Q^{(i)}}\, e^{Qz} & (0 \leq z < t) \\ d_{Q,\omega_Q^{(i)}}\, e^{-Qz} & (z \geq t) \end{cases} , \quad (6)$$

where:

$$b_{Q,\omega_Q^{(i)}} = \frac{\epsilon_{ox}(\omega_Q^{(i)}) - \epsilon_g(\omega_Q^{(i)})}{2\epsilon_{ox}(\omega_Q^{(i)})}\, a_{Q,\omega_Q^{(i)}} , \quad (7)$$

$$c_{Q,\omega_Q^{(i)}} = \frac{\epsilon_{ox}(\omega_Q^{(i)}) + \epsilon_g(\omega_Q^{(i)})}{2\epsilon_{ox}(\omega_Q^{(i)})}\, a_{Q,\omega_Q^{(i)}} , \quad (8)$$

$$d_{Q,\omega_Q^{(i)}} = \frac{\epsilon_{ox}(\omega_Q^{(i)}) - \epsilon_g(\omega_Q^{(i)})}{\epsilon_{ox}(\omega_Q^{(i)}) + \epsilon_s(Q,\omega_Q^{(i)})}\, a_{Q,\omega_Q^{(i)}} , \quad (9)$$

Thus, the field is determined up to an overall multiplicative constant $a_{Q,\omega_Q^{(i)}}$. We shall discuss below in Sec. II C how to determine this normalization constant.

As mentioned above, here we have taken the long-wavelength limits of the dielectric functions. In Ref. 7 we have shown that this is a satisfactory approximation in the range of Q of interest, of the order of the Fermi wave vector of the 2DEG. The lattice (ionic) response of the insulator in principle should include polarization effects due to a multitude of optical modes, functions not only of the chosen materials, but also of their chemical composition (stoichiometric or not, depending on deposition and annealing conditions), on their allotropic forms (amorphous or, if crystalline, on their crystallographic structure), etc. We shall discuss below in Sec. III B some

of these issues. In order to keep the treatment tractable, we consider only two bulk optical modes, obtained by averaging over direction (*e.g.*, over the A_{2u} and E_u modes for bc tetragonal ZrO_2 or $ZrSiO_4$), by considering only the two modes exhibiting the largest oscillator strength, by lumping 'bands' of modes into two groups, or by combining of all of these approximations.[9] Thus, we assume an ionic dielectric function of the form:

$$\epsilon_{ox}(\omega) = \epsilon_{ox}^{\infty} + (\epsilon_{ox}^{i} - \epsilon_{ox}^{\infty})\frac{\omega_{TO2}^2}{\omega_{TO2}^2 - \omega^2} + (\epsilon_{ox}^{0} - \epsilon_{ox}^{i})\frac{\omega_{TO1}^2}{\omega_{TO1}^2 - \omega^2}, \quad (10)$$

where ϵ_{ox}^0 and ϵ_{ox}^{∞} are the static and optical permittivity of the insulator, respectively, (so that $\kappa = \epsilon_{ox}^0/\epsilon_0$, where ϵ_0 is the permittivity of vacuum) and ω_{TO1} and ω_{TO2} are the angular frequencies of the only two TO-phonon modes we shall consider in the insulator. We assume $\omega_{TO1} \leq \omega_{TO2}$. Finally, ϵ_{ox}^i is an 'intermediate' insulator permittivity describing the dielectric response of the insulator at some intermediate frequency ω_{int} such that $\omega_{TO1} \leq \omega_{int} \leq \omega_{TO2}$. Physically, it is related to the oscillator strength of each mode. It must be determined from the energy splitting between longitudinal and transverse optical modes via the Lyddane-Sachs-Teller relation (or its trivial extension in the case of two optical modes), which allows us to rewrite Eq. (10) as:

$$\epsilon_{ox}(\omega) = \epsilon_{ox}^{\infty}\frac{(\omega_{LO2}^2 - \omega^2)(\omega_{LO1}^2 - \omega^2)}{(\omega_{TO2}^2 - \omega^2)(\omega_{TO1}^2 - \omega^2)}, \quad (11)$$

where the frequency of the two LO-modes is given by the generalized Lyddane-Sachs-Teller relation:

$$\omega_{LOi}^2 = \frac{1}{2\Delta}\left[b \pm (b^2 - 4\Delta c)^{1/2}\right] \qquad (1 = 1,2), \quad (12)$$

with:

$$\begin{aligned} \Delta &= \epsilon_{ox}^{\infty} \\ b &= \Delta(\omega_{TO1}^2 + \omega_{TO2}^2) + (\epsilon_{ox}^i - \epsilon_{ox}^{\infty})\,\omega_{TO2}^2 \\ &\quad + (\epsilon_{ox}^0 - \epsilon_{ox}^i)\,\omega_{TO1}^2 \\ c &= (\Delta + \epsilon_{ox}^0 - \epsilon_{ox}^{\infty})\,\omega_{TO1}^2\omega_{TO2}^2 . \end{aligned} \quad (13)$$

For the electronic response of the gate we take the usual long-wavelength expression:

$$\epsilon_g(\omega) = \epsilon_{Si}^{\infty}\left(1 - \frac{\omega_{p,g}^2}{\omega^2}\right), \quad (14)$$

where $\omega_{p,g}^2 = e^2N_g/(\epsilon_{Si}^{\infty}m_g)$ is the bulk plasma frequency of the polycrystalline-Si gate with an electron density N_g (obtained from some suitable average of the electron density over the depletion layer of the poly-Si gate, as discussed below in Sec. III A), with an effective mass m_g ($= 0.32\ m_0$, where m_0 is the electron mass), and optical permittivity ϵ_{Si}^{∞}. Finally, for the inverted substrate we assume:

$$\epsilon_s(Q,\omega) = \epsilon_{Si}^{\infty}\left[1 - \frac{\omega_{p,s}(Q)^2}{\omega^2}\right], \quad (15)$$

where $\omega_{p,s}(Q) = [\sum_\nu e^2n_\nu Q/(\epsilon_{Si}^{\infty}m_\nu)]^{1/2}$ is the plasma frequency of the 2DEG, n_ν and m_ν being the electron density and conductivity mass in each of the occupied subbands labeled by the index ν.

Equation (4) is an algebraic equation of sixth degree in ω^2, and so it admits 6 positive solutions, which we shall label as $\omega_Q^{(i)}$. Two of these solutions (which we shall arbitrarily label with the indices $i = 5,6$) are associated with a small scattering field and will be ignored: At small values of Qt, they have energies extremely close to the energies of the associated TO-modes and couple poorly with the electrons in the inversion layer. At large Q (or for very large insulator thickness), they are mainly localized at the 'far' gate/insulator interface – thus yielding a scattering strength depressed by a factor $\sim e^{-2Qt}$ – , with frequencies approaching the frequencies of the bare SO modes at that interface.[10] In the absence of electronic screening and assuming identical high-frequency electronic responses for the gate and the substrate (*i.e.*, setting $\epsilon_g(\omega) = \epsilon_s(Q,\omega) = \epsilon_{Si}^{\infty}$) these frequencies are given by Eq. (12) with Δ replaced by $\epsilon_{ox}^{\infty} + \epsilon_{Si}^{\infty}\coth(Qt)\{1 + [1 - \tanh^2(Qt)]^{1/2}\}$ in Eq. (13). In the same limit, $Qt \to \infty$, these frequencies coincide with those of the modes at the substrate/insulator interface whose frequencies, $\omega_Q^{(SOi)}$ (for $i = 1,2$), in turn are given by Eq. (12) with Δ now replaced by $\epsilon_{ox}^{\infty} + \epsilon_{Si}^{\infty}\coth(Qt)\{1 - [1 - \tanh^2(Qt)]^{1/2}\}$ in Eq. (13). The remaining 4 solutions (the only ones considered in Ref. 7 and which we shall label with the index i running from 1 through 4, ordered so that $\omega_Q^{(1)} \geq \omega_Q^{(2)} \geq \omega_Q^{(3)} \geq \omega_Q^{(4)}$) represent coupled interface plasmon-phonon modes. For the time being we shall assume that these solutions exist undamped at all wavelengths. We shall discuss below in Sec. II D how we handle them in the regime in which Landau damping makes it improper to talk about plasma oscillations.

B. Plasmon and phonon content

Each one of the four solutions $\omega_Q^{(i)}$ represents excitations involving both electronic – at both interfaces – and ionic– of both 'flavors' TO1 and TO2 – oscillations. Since we must deal separately with scattering with different components of the field (6), namely, the electronic component at the gate/insulator interface (the isurface plasmon, or SP, component) and the ionic, or SO, components[11], we must define first the gate- and substrate-plasmon content $\Pi^{(G)}(\omega_Q^{(i)})$ and $\Pi^{(S)}(\omega_Q^{(i)})$ respectively, of each solution. To do that, we first consider the three solutions $\omega_Q^{(-g,\alpha)}$) ($\alpha = 1,3$), obtained from the secular equation, Eq. (4), by ignoring the response of the gate-plasma (that is, by replacing $\epsilon_g(\omega)$ with ϵ_{Si}^{∞}). Then, extending the result of Kim and coworkers[12] to the case of interest here, the gate-plasmon content of mode i will be defined as:

$$\Pi^{(G)}(\omega_Q^{(i)}) \approx \left|\frac{(\omega_Q^{(i)2} - \omega_Q^{(-g,1)2})\,(\omega_Q^{(i)2} - \omega_Q^{(-g,2)2})\,(\omega_Q^{(i)2} - \omega_Q^{(-g,3)2})}{(\omega_Q^{(i)2} - \omega_Q^{(j)2})\,(\omega_Q^{(i)2} - \omega_Q^{(k)2})\,(\omega_Q^{(i)2} - \omega_Q^{(l)2})}\right| \quad (16)$$

where the indices (i,j,k,l) are cyclical. The 'approximate' sign above results from having neglected the two solutions mentioned above. However, at small Q these are perturbed by the absence/presence of the plasmas to such a small extent (and so are almost totally phonon-like) as to render Eq. (16) a very satisfactory approximation. Similarly, considering the three solutions $\omega_Q^{(-s,\alpha)})$ $(\alpha=1,3)$, obtained from the secular equation Eq. (4) by ignoring the plasma response of the 2DEG in the substrate (that is, by replacing $\epsilon_s(Q,\omega)$ with ϵ_{Si}^{∞}), we define the substrate-plasmon content of mode i as:

$$\Pi^{(S)}(\omega_Q^{(i)}) \approx \left| \frac{({\omega_Q^{(i)}}^2 - {\omega_Q^{(-s,1)}}^2)\,({\omega_Q^{(i)}}^2 - {\omega_Q^{(-s,2)}}^2)\,({\omega_Q^{(i)}}^2 - {\omega_Q^{(-s,3)}}^2)}{({\omega_Q^{(i)}}^2 - {\omega_Q^{(j)}}^2)\,({\omega_Q^{(i)}}^2 - {\omega_Q^{(k)}}^2)\,({\omega_Q^{(i)}}^2 - {\omega_Q^{(l)}}^2)} \right| \tag{17}$$

It can be verified that Eqns. (16) and (17) satisfy the normalization conditions[13]

$$\sum_{i=1}^{4} \Pi^{(G)}(\omega_Q^{(i)}) = 1 , \tag{18}$$

and

$$\sum_{i=1}^{4} \Pi^{(S)}(\omega_Q^{(i)}) = 1 . \tag{19}$$

Having ignored the two solutions $\omega_Q^{(5)}$ and $\omega_Q^{(6)}$ forces us to approximate the phonon content of each mode as follows. From Eqns. (16) and (17) it follows that the total phonon content of mode i will be:

$$\Phi(\omega_Q^{(i)}) = 1 - \Pi^{(G)}(\omega_Q^{(i)}) - \Pi^{(S)}(\omega_Q^{(i)}) . \tag{20}$$

In order to define separate phonon-1 and phonon-2 contents, we also consider the three solutions $\omega_Q^{(-TO1,\alpha)}$ $(\alpha=1,3)$, obtained from the secular equation, Eq. (4), but now ignoring the response of phonon-1 – that is, by replacing $\epsilon_{ox}(\omega)$ with $\epsilon_{ox}^{\infty}(\omega_{LO2}^2-\omega^2)/(\omega_{TO2}^2-\omega^2)$ – and the three solutions $\omega_Q^{(-TO2,\alpha)}$ $(\alpha=1,3)$ similarly obtained by ignoring the response of the TO-mode 2 by setting in Eq. (4) $\epsilon_{ox}(\omega)\rightarrow\epsilon_{ox}^{\infty}(\omega_{LO1}^2-\omega^2)/(\omega_{TO1}^2-\omega^2)$. Therefore, the relative phonon-1 content of mode i will be

$$R^{(TO1)}(\omega_Q^{(i)}) \approx \left| \frac{({\omega_Q^{(i)}}^2 - {\omega_Q^{(-TO1,1)}}^2)({\omega_Q^{(i)}}^2 - {\omega_Q^{(-TO1,2)}}^2)({\omega_Q^{(i)}}^2 - {\omega_Q^{(-TO1,3)}}^2)}{({\omega_Q^{(i)}}^2 - {\omega_Q^{(j)}}^2)({\omega_Q^{(i)}}^2 - {\omega_Q^{(k)}}^2)({\omega_Q^{(i)}}^2 - {\omega_Q^{(l)}}^2)} \right| \tag{21}$$

(where, as before, i,j,k,l are cyclical) so that, finally, the TO-phonon-1 content of mode i will be:

$$\Phi^{(TO1)}(\omega_Q^{(i)}) \approx \frac{R^{(TO1)}(\omega_Q^{(i)})}{R^{(TO1)}(\omega_Q^{(i)}) + R^{(TO2)}(\omega_Q^{(i)})} \times [1 - \Pi^{(G)}(\omega_Q^{(i)}) - \Pi^{(S)}(\omega_Q^{(i)})] , \tag{22}$$

and similarly for $\Phi^{(TO2)}(\omega_Q^{(i)})$. Once more, it has been verified numerically that these definitions satisfy the additional normalization conditions:

$$\sum_{i=1}^{4} \Phi^{(TO1)}(\omega_Q^{(i)}) = 1 , \tag{23}$$

$$\sum_{i=1}^{4} \Phi^{(TO2)}(\omega_Q^{(i)}) = 1 , \tag{24}$$

and, for each mode i:

$$\Pi^{(G)}(\omega_Q^{(i)}) + \Pi^{(S)}(\omega_Q^{(i)}) + \Phi^{(TO1)}(\omega_Q^{(i)}) + \Phi^{(TO2)}(\omega_Q^{(i)}) = 1 . \tag{25}$$

C. Scattering strength

The amplitude $a_{Q,\omega}$ of the field, Eq. (6), can be determined using the semiclassical approach originally proposed by Stern and Ferrel[14] which we also followed in Ref. 7 and described in a simple case in Appendix A of Ref. 8. We first consider the time-averaged total (electrostatic, including self-energy) energy, $< W_Q^{(i)} >$, associated with the field $\phi_Q^{(i)}(\mathbf{R},z,t)$ caused by the excitation of mode i oscillating at the frequency $\omega_Q^{(i)}$. (The bra-kets $< ... >$ denote time average). Let us write the electrostatic potential at a given wavelength as:

$$\phi_Q^{(i)}(\mathbf{R},z,t) = \phi_{Q,\omega_Q^{(i)}}^{(i)}(z) \cos(\mathbf{Q}\cdot\mathbf{R}-\omega_Q^{(i)}t) . \tag{26}$$

Since phonons and plasmons in the harmonic and linear-response approximations, respectively, are represented as harmonic oscillations, the time-averaged total energy associated with these excitations is simply twice the time-averaged potential energy, $< U_Q^{(i)} >$. This, in turn, is the electrostatic (self)energy of the interface polarization charge density $\rho_Q^{(i)}(\mathbf{R},z,t)$ in the presence of the potential $\phi_Q^{(i)}(\mathbf{R},z,t)$ caused by the interface charge itself. We may express this potential energy in two alternative equivalent ways: From the expression (6) for the potential, the density of the polarization charge associated with mode i is localized at the two interfaces and can be obtained from Poisson equation $\rho_Q^{(i)}(\mathbf{R},z,t) = -\nabla\cdot[\epsilon(\omega_Q^{(i)};z)\nabla\phi_Q^{(i)}(\mathbf{R},z,t)]$ (where the z-dependence in $\epsilon(\omega_Q^{(i)};z)$ reflects the fact the we must use the appropriate dielectric functions across the interfaces):

$$\rho_Q^{(i)}(\mathbf{R},z,t) = \{\delta(z)\,[\,\epsilon_{gate}(\omega_Q^{(i)})\,a_{Q,\omega_Q^{(i)}} + \epsilon_{insulator}(\omega_Q^{(i)})\,(b_{Q,\omega_Q^{(i)}} - c_{Q,\omega_Q^{(i)}})\,] + \delta(z-t)\,[\,\epsilon_{insulator}(\omega_Q^{(i)})(c_{Q,\omega_Q^{(i)}}e^{Qt} - b_{Q,\omega_Q^{(i)}}e^{-Qt}) + \epsilon_{substrate}(Q,\omega_Q^{(i)})\,d_{Q,\omega_Q^{(i)}}e^{-Qt}\,]\,\} \times Q\,\cos(\mathbf{Q}\cdot\mathbf{R}-\omega_Q^{(i)}t) , \tag{27}$$

having introduced the new functions $\epsilon_{gate}(\omega)$, $\epsilon_{insulator}(\omega)$, and $\epsilon_{substrate}(Q,\omega)$ which must be chosen in a way consistent with the component of the polarization charge $\rho_Q^{(i)}(\mathbf{R},z,t)$ we are considering, as discussed below. Therefore, for the energy $< W_Q^{(i)} >$ we can write:

$$< W_Q^{(i)} > = \; 2 \; < U_Q^{(i)} > =$$
$$\frac{2}{\Omega} \left\langle \int_\Omega d\mathbf{R} \int_{-\infty}^{\infty} dz \; \phi_Q^{(i)}(\mathbf{R},z,t) \; \rho_Q^{(i)}(\mathbf{R},z,t) \right\rangle , \quad (28)$$

where Ω is a normalization area. Alternatively, using Green's identity and accounting for the discontinuity of the electric and displacement fields across the interfaces, we can express $< W_Q^{(i)} >$ in terms of the electrostatic energy of the field $\mathbf{E}_Q^{(i)} = -\nabla\phi_Q^{(i)}$:

$$< W_Q^{(i)} > = \frac{2}{\Omega} \left\langle \int_\Omega d\mathbf{R} \int_{-\infty}^{\infty} dz \; \epsilon(\omega_Q^{(i)};z) \; |\mathbf{E}_Q^{(i)}(\mathbf{R},z,t)|^2 \right\rangle \quad (29)$$

From Eqns. (7-9) and either using Eq. (28) or performing the integrals in Eq. (29) using Eq. (6), we obtain:

$$< W_Q^{(i)} > = \; Q \; \epsilon_{TOT}(Q,\omega_Q^{(i)}) \left[\frac{\epsilon_{ox}(\omega_Q^{(i)}) - \epsilon_g(\omega_Q^{(i)})}{\epsilon_{ox}(\omega_Q^{(i)}) + \epsilon_s(Q,\omega_Q^{(i)})} \right]^2 \times a^2_{Q,\omega_Q^{(i)}} \; e^{-2Qt} \; . \quad (30)$$

Here the 'total' effective dielectric function of the substrate coupled to the gate and the insulating layer has been defined as:

$$\epsilon_{TOT}(Q,\omega) \; = \; \epsilon_{gate}(\omega) \left[\frac{\epsilon_{ox}(\omega) + \epsilon_s(Q,\omega)}{\epsilon_{ox}(\omega) - \epsilon_g(\omega)} \right]^2 e^{2Qt}$$
$$+ \; \epsilon_{insulator}(\omega) \left\{ \left[\frac{\epsilon_{ox}(\omega) + \epsilon_s(Q,\omega)}{2\epsilon_{ox}(\omega)} \right]^2 (e^{2Qt} - 1) + \left[\frac{\epsilon_{ox}(\omega) - \epsilon_s(Q,\omega)}{2\epsilon_{ox}(\omega)} \right]^2 (1 - e^{-2Qt}) \right\}$$
$$+ \; \epsilon_{substrate}(Q,\omega) \; , \quad (31)$$

having made repeated use of the relations (4) to reach one of the many possible equivalent algebraic forms. The semiclassical nature of the argument enters the final step of setting the quantity $< W_Q^{(i)} >$ equal to the zero-point energy, $\hbar\omega_Q^{(i)}/2$, of the quantized excitation. This finally determines the 'normalization constant', $a_{Q,\omega_Q^{(i)}}$, and thus the amplitude of the scattering field in the substrate ($z \geq t$):

$$\phi^{(i)}_{Q,\omega_Q^{(i)}} \; = \; \left[\frac{\hbar\omega_Q^{(i)}}{2 \; Q \; \epsilon_{TOT}(Q,\omega_Q^{(i)})} \right]^{1/2} \; e^{-Q(z-t)} \; , \quad (32)$$

up to the appropriate Bose factors of the excitations, $n_Q^{(i)\,1/2}$ and $(1 + n_Q^{(i)})^{1/2}$, which multiply the scattering potential for absorption and emission processes, respectively.

Care must be exercised in choosing the form of the dielectric functions $\epsilon_{gate}(\omega)$, $\epsilon_{insulator}(\omega)$, and $\epsilon_{substrate}(Q,\omega)$, which appear in the expression for $\epsilon_{TOT}(Q,\omega)$, since whenever we are interested in determining the potential energy due to a particular type of response of the system (ionic or electronic), we cannot include this response into these dielectric functions. For example, if we set $\epsilon_{gate}(\omega) = \epsilon_g(\omega)$, $\epsilon_{insulator}(\omega) = \epsilon_{ox}(\omega)$, and $\epsilon_{substrate}(Q,\omega) = \epsilon_s(Q,\omega)$, we effectively lump the whole dielectric response, ionic and electronic, into the dielectric functions, and we expect that the potential energy of 'whatever response is left' (none, in this case) in the field and charge, ϕ_Q and ρ_Q, should vanish. Indeed when so doing, the resulting $\epsilon_{TOT}(Q,\omega)$ vanishes for $\omega = \omega_Q^{(i)}$, the equation $\epsilon_{TOT}(Q,\omega) = 0$ being equivalent to the secular equation (4), as some tedious but trivial algebra reveals. Therefore, when taking $\epsilon_{gate}(\omega) = \epsilon_{Si}^{\infty}$, $\epsilon_{substrate}(Q,\omega) = \epsilon_{Si}^{\infty}$, and $\epsilon_{insulator}(\omega) = \epsilon_{ox}^0$ we consider only the plasmon contribution to the polarization charges. Indeed, the response of the insulator lattice is removed from the electrostatic field by being lumped into the insulator permittivity when setting $\epsilon_{insulator}(\omega) = \epsilon_{ox}^0$, while the electronic response is removed from the dielectric functions of the gate and substrate, and is included directly into the amplitude of the electrostatic field and polarization charge, ϕ_Q and ρ_Q. In this case Eq. (32) represents the amplitude of the field induced by plasma excitations. Therefore, defining as $\epsilon_{TOT}^{(PL)}(Q,\omega_Q^{(i)})$ the total plasma dielectric function so obtained (identical to what was obtained in Ref. 7 in which the insulator response was ignored[15]), scattering between electrons in the substrate and gate plasmons is described by the effective scattering field:

$$\phi^{(i,g,PL)}_{Q,\omega_Q^{(i)}}(z) \; =$$
$$\left[\frac{\hbar\omega_Q^{(i)}}{2 \; Q \; \epsilon_{TOT}^{(PL)}(Q,\omega_Q^{(i)})} \; \Pi^{(G)}(\omega_Q^{(i)}) \right]^{1/2} \; e^{-Q(z-t)} \; . \quad (33)$$

Scattering with the field induced by the polarization charges of the insulator lattice (*i.e.*, with the optical phonons in the insulator) can be evaluated in a way essentially identical to the approach followed by Kittel[16] to evaluate the Fröhlich coupling in bulk polar materials. The only difference between Kittel's and our approach consists in following Stern and Ferrel[14] in evaluating the ground-state energy semiclassically, rather than from second-order perturbation theory. In order to isolate the contribution of each phonon independently and consider only the lattice polarization, the squared, time-averaged amplitude of the scattering field is computed by lumping the electronic response into the dielectric functions of the gate and substrate, while letting one phonon (say, phonon 2 to fix the ideas) respond, but first by 'freezing' the other mode (TO1) and then by considering its full response. The difference between the two squared amplitudes so obtained constitutes the effect of the ionic polarization charge associated solely with the optical mode 1. To be explicit, in our case the amplitude of the field (32) when only phonon 2 responds is obtained by setting $\epsilon_{gate}(\omega) = \epsilon_g(\omega)$ (re-

sponse of the gate plasmons lumped into the gate dielectric function), $\epsilon_{substrate}(Q,\omega) = \epsilon_s(Q,\omega)$ (response of the substrate plasmons lumped into the dielectric function of the inversion layer), and setting $\epsilon_{insulator}(\omega) = \epsilon_{ox}^{\infty}(\omega_{LO2}^2 - \omega^2)/(\omega_{TO2}^2 - \omega^2)$ (phonon 2 responds at the frequency ω, while phonon 1 does not respond). Let $\epsilon_{TOT,high}^{(TO1)}$ be the resulting effective dielectric function. On the contrary, when phonon 1 is allowed to respond fully, we have $\epsilon_{insulator}(\omega) = \epsilon_{ox}^{\infty}[(\omega_{LO2}^2 - \omega^2)/(\omega_{TO2}^2 - \omega^2)](\omega_{LO1}/\omega_{TO1})^2$ (which reduces to ϵ_{ox}^0 in the simpler case of insulators exhibiting only one TO-mode), the full response of phonon 1 now being accounted for by the term $(\omega_{LO1}/\omega_{TO1})^2$. Let $\epsilon_{TOT,low}^{(TO1)}$ denote the resulting dielectric function. Thus, the interaction between electrons in the inversion layer and the TO1-phonon-content of the mode i will be described by the scattering field:

$$\phi_{Q,\omega_Q^{(i)}}^{(i,PH1)}(z) = \left\{ \frac{\hbar\omega_Q^{(i)}}{2Q} \left[\frac{1}{\epsilon_{TOT,high}^{(TO1)}(Q,\omega_Q^{(i)})} - \frac{1}{\epsilon_{TOT,low}^{(TO1)}(Q,\omega_Q^{(i)})} \right] \Phi^{(TO1)}(\omega_Q^{(i)}) \right\}^{1/2} e^{-Q(z-t)} \; . \quad (34)$$

A similar expression describing the scattering strength with phonon mode 2 can be trivially obtained by swapping the indices 1 and 2 in the discussion above.

It is interesting to consider a few simpler and/or limiting cases of Eq. (34). For a single LO-phonon of frequency ω_{LO} in a bulk polar crystal with static and optical permittivities ϵ^0 and ϵ^{∞}, respectively, in place of Eq. (34) our procedure yields for the phonon field the usual unscreened Fröhlich field of amplitude:

$$\left[\frac{\hbar\omega_{LO}}{2q^2} \left(\frac{1}{\epsilon^{\infty}} - \frac{1}{\epsilon^0} \right) \right]^{1/2} , \quad (35)$$

where q is the magnitude of the three-dimensional wave vector. When also accounting for the bulk plasma response at frequency ω_P, the total dielectric function takes the form $\epsilon_{TOT}(Q,\omega) = \epsilon^{\infty}[(\omega^2 - \omega_{LO}^2)/(\omega^2 - \omega_{TO}^2) - (\omega_P/\omega)^2]$ (where $\omega_{LO} = (\epsilon^0/\epsilon^{\infty})^{1/2}\omega_{TO}$), and the solutions of the equation $\epsilon_{TOT}(Q,\omega) = 0$ represent the two coupled TO-phonon/bulk-plasmons modes with frequencies

$$\omega_{\pm}^2 = \frac{1}{2}\{\omega_{LO}^2 + \omega_P^2 \pm [(\omega_{LO}^2 + \omega_P^2)^2 - 4\omega_{TO}^2\omega_P^2]^{1/2}\} \; . \quad (36)$$

The plasmon content of each mode, according to the reduction of Eq. (16) to the present case of only two modes, is

$$\Pi(\omega_{\pm}) = \left| \frac{\omega_{\pm}^2 - \omega_{LO}^2}{\omega_{+}^2 - \omega_{-}^2} \right| , \quad (37)$$

hence, from Eqns. (20) and (22), the phonon content will be:

$$\Phi(\omega_{\pm}) = 1 - \Pi(\omega_{\pm}) = \left| \frac{\omega_{\pm}^2 - \omega_P^2}{\omega_{+}^2 - \omega_{-}^2} \right| , \quad (38)$$

having used Eq. (36). Following the procedure we have employed to reach Eq. (34), we get for the scattering amplitude with the phonon content of each mode:

$$\left(\frac{\hbar\omega_{\pm}}{2q^2} \right)^{1/2} \left[\frac{1}{\epsilon^{\infty} - \epsilon^{\infty}\omega_P^2/\omega_{\pm}^2} - \frac{1}{\epsilon^0 - \epsilon^{\infty}\omega_P^2/\omega_{\pm}^2} \right]^{1/2} \times \left| \frac{\omega_{\pm}^2 - \omega_P^2}{\omega_{+}^2 - \omega_{-}^2} \right|^{1/2} , \quad (39)$$

which is identical to the expressions given by Kim and coworkers[12] and Ridley[17]. Note that it is not trivial to 'discover' this agreement, since Kim and coworkers do employ the same definition of phonon content, Eq. (38) (also obtained by Sanborn[18]) used here, but derive the scattering strength from the imaginary part of $1/\epsilon_{TOT}(Q,\omega_{\pm})$, while Ridley follows the equation-of-motion approach and employs a different definition for the phonon content and scattering strength of each mode, yet reaching the same identical result. This agreement has been verified numerically here, although presumably more efforts could produce a more elegant algebraic proof.

Closer to our present interest is the case of a single oxide/semiconductor interface (*i.e.*, the $t \to \infty$ limit) with a single insulator-TO-mode. In this case, ignoring the plasma response of the semiconductor, Eq. (4) yields a solution of frequency:

$$\omega_{SO} = \omega_{TO} \left[\frac{\epsilon_{ox}^0 + \epsilon_{Si}^{\infty}}{\epsilon_{ox}^{\infty} + \epsilon_{Si}^{\infty}} \right]^{1/2} . \quad (40)$$

Taking now the limit $Qt \to \infty$ and making use of the secular equation, Eq. (4), in order to retain only the leading terms in $\exp(-Qt)$, Eq. (34) gives:

$$\phi_{Q,\omega_{SO}}^{(PH)}(z) = \left\{ \frac{\hbar\omega_{SO}}{2Q} \left[\frac{1}{\epsilon_{Si}^{\infty} + \epsilon_{ox}^{\infty}} - \frac{1}{\epsilon_{Si}^{\infty} + \epsilon_{ox}^0} \right] \right\}^{1/2} \times e^{-Q(z-t)} \; . \quad (41)$$

which coincides with the unscreened strength for scattering with 'remote phonons' obtained by Wang and Mahan[3] and later employed for the Si/SiO$_2$ system by Hess and Vogl[4] and Moore and Ferry[5]. A simple screened expression follows immediately from our formulation, simply replacing ϵ_{Si}^{∞} with $\epsilon_s(Q,\omega_{SO})$ in Eq. (41) above. Clearly, our formulation would also yield the correct dispersion of the coupled phonon/interface-plasmon dispersion with the enhanced scattering amplitude with the high-frequency mode caused by a large amount of anti-screening. This has been discussed by Ridley[17], Varga[19], Das Sarma and coworkers[20] in the case of bulk systems, by Gersten[21] and Puri and Schaich[22] in the somewhat similar case of bulk-semiconductor/accumulated-semiconductor/vacuum systems.

D. Landau damping

At sufficiently short wavelengths plasmons cease to be proper excitations. In our context this may happen when

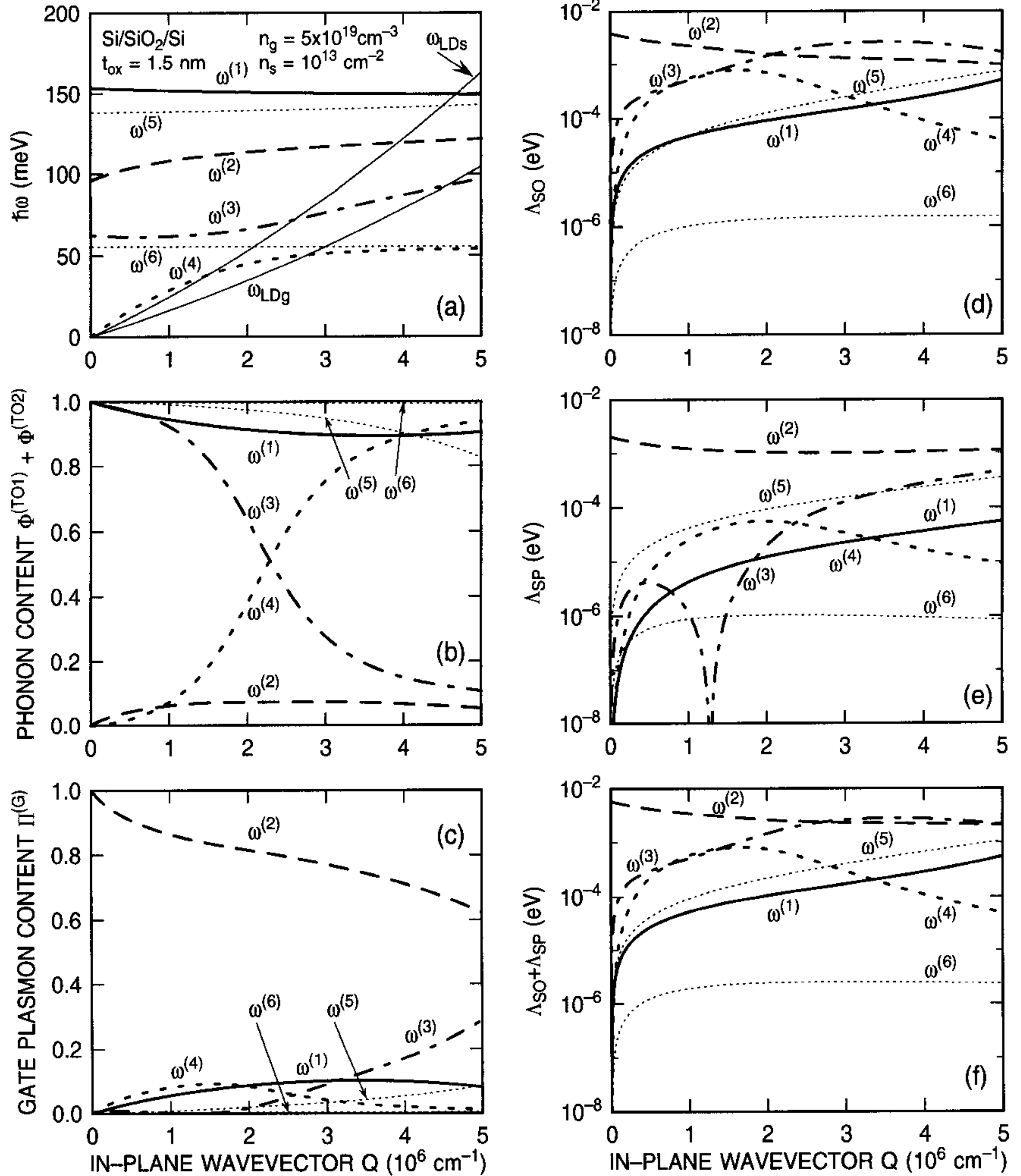

FIG. 2. Calculated dispersion (a), total phonon content (b), gate-plasmon content (c), scattering strength with the phonon-like content (d), gate-plasmon-like content (e), and total scattering strength (f) for the insulator-optical-phonon/substrate-and-gate-plasmons interface modes for the Si/SiO2/Si system. In (a), the curves labeled by ω_{LDs} and ω_{LDg} identify the boundary of the substrate and gate Landau-damping regions, respectively. Modes 5 and 6 (thin dotted lines) are related to bulk-insulator TO-modes (small Q) or to optical oscillations localized at the 'far' gate/insulator interface (large Q), and their effect on the electron mobility has been ignored.

the gate-plasma-like solution $\omega_Q^{(i_g)}$ (where usually $i_g = 1$ at large enough N_g) enters the single-particle continuum in the gate, the substrate-plasma-like solution enters the single-particle continuum in the substrate, or both. In order to account for Landau damping, albeit approximately, we proceed as follows. Whenever the substrate-

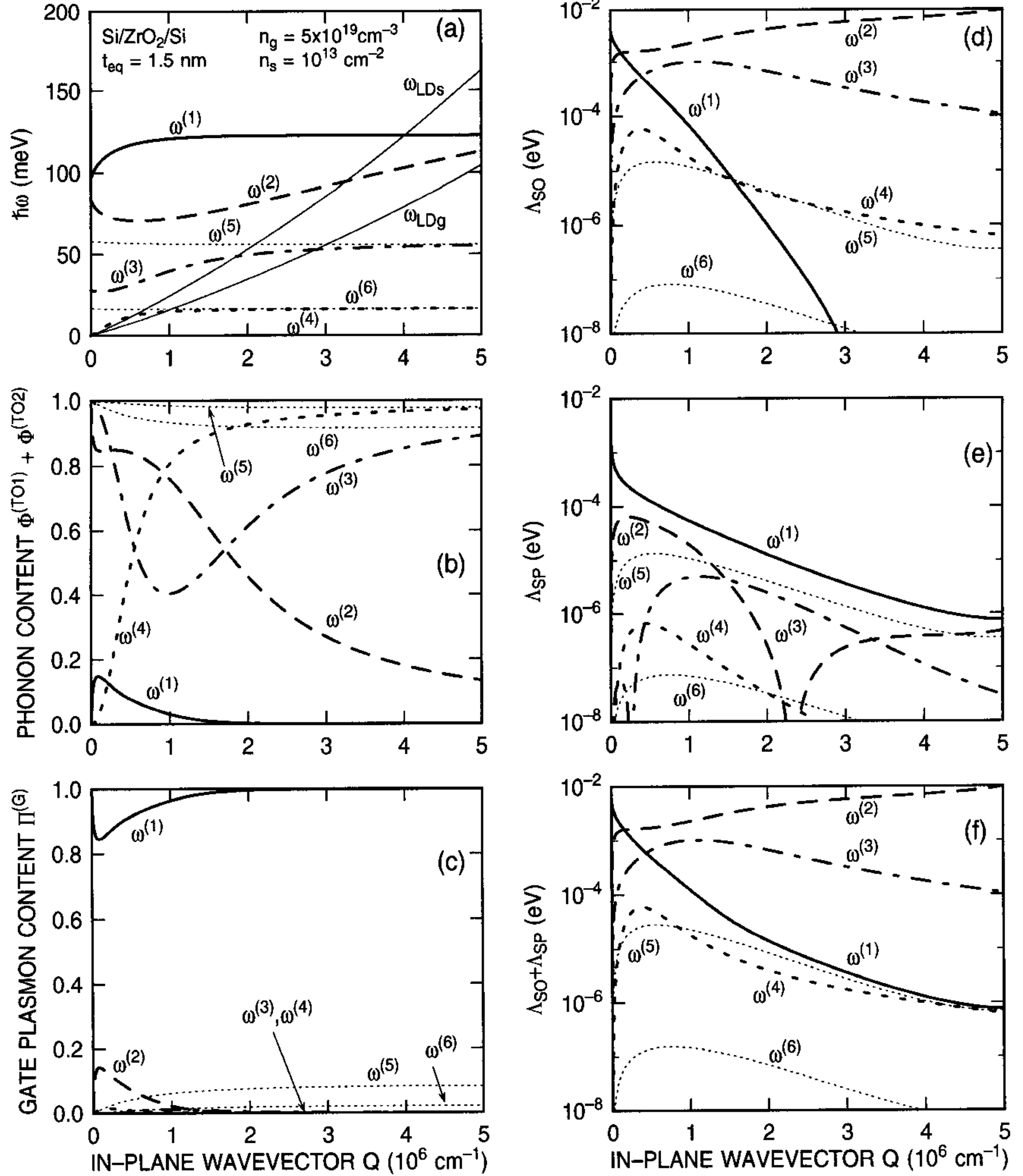

FIG. 3. As in Fig. 2, but for the Si/ZrO$_2$/Si system. Note that the two highest-energy modes are plasmon-like.

plasmon-like excitation $\omega_Q^{(i_s)}$ (where, usually, $i_s = 4$) enters the single-particle continuum of the 2DEG evaluated in the extreme quantum limit (*i.e.*, $\omega_Q^{(i_s)} \leq [(\hbar Q)/(2m_t)](Q + 2K_F)$, where $K_F = (\pi n_s)^{1/2}$, $m_t = 0.19 m_0$ being the transverse effective mass), we consider only the three solutions $\omega_Q^{(-s,i)}$ ($i = 1,3$) given above. These represent the three coupled gate-plasmon/insulator-TO-modes which exist when the substrate plasma is absent. The plasmon/phonon content and scattering strength for these modes are obtained in a way completely analogous to the one discussed so far. For example, in analogy with Eq. (16), the gate-plasmon content of each mode $\omega_Q^{(-s,i)}$ will be approximated by:

$$\Pi^{(G)}(\omega_Q^{(-s,i)}) \approx \left| \frac{(\omega_Q^{(-s,i)^2} - \omega_Q^{(SO1)^2})\,(\omega_Q^{(-s,i)^2} - \omega_Q^{(SO2)^2})}{(\omega_Q^{(-s,i)^2} - \omega_Q^{(-s,j)^2})\,(\omega_Q^{(-s,i)^2} - \omega_Q^{(-s,k)^2})} \right| , \quad (42)$$

where (i,j,k) are cyclical, and $\omega_Q^{(SOi)}$ – defined in the paragraph following Eq. (15) above – are the frequencies of the pure surface-optical modes in absence of any plasma response. The plasmon-like scattering field associated with these modes will be given by Eq. (33), while the phonon-like scattering field will be as in Eq. (34), but setting $\epsilon_{substrate}(Q,\omega) = \epsilon_{Si}^{\infty}$ into $\epsilon_{TOT,high}^{(TOi)}(\omega)$ and $\epsilon_{TOT,low}^{(TOi)}(\omega)$ to reflect the absence of the two-dimensional (2D) plasma.

Similarly, when the frequency of the gate-plasmon-like excitation enters the single-particle continuum of gate (that is, $\omega_Q^{(i_g)} \leq (\hbar Q)/(2m_d)(Q+2k_F)$, where k_F is the zero-temperature Fermi wave vector of the electron gas in the gate, $k_F = (\pi^2 N_g/2)^{1/3}$, and the index i_g takes a value of 1 or 2, depending on electron density in the gate, frequency of the high-energy TO-mode, ω_{TO2}, and dielectric constants of the material considered), we consider only the three solutions $\omega_Q^{(-g,i)}$ $(i = 1,3)$ representing the three coupled substrate-plasmon /insulator-TO-modes which exist when the gate plasma does not respond. In this case only phonon-like scattering with these modes is considered. Equation (34) describes the surface-phonon scattering field, setting $\epsilon_{gate}(\omega) = \epsilon_{Si}^{\infty}$ in $\epsilon_{TOT,high}^{(TOi)}(\omega)$ and $\epsilon_{TOT,low}^{(TOi)}(\omega)$ to reflect the absence of the gate plasma.

Finally, when both the gate- and the substrate-plasmons-like dispersions are within their respective Landau-damping regions, we consider only the two phonon-like modes of frequencies $\omega_Q^{(SOi)}$ $(i = 1,2)$, whose associated scattering field is given by Eq. (34) with $\epsilon_{gate}(\omega) = \epsilon_{substrate}(Q,\omega) = \epsilon_{Si}^{\infty}$ employed into $\epsilon_{TOT,high}^{(TOi)}(\omega)$ and $\epsilon_{TOT,low}^{(TOi)}(\omega)$. In this case we approach the high-Q unscreened limit in which Eq. (41) – suitably extended to the case of two insulator TO-modes – describes the scattering field of the two surviving surface phonon modes.

E. Discussion

We summarize graphically the results of this section by showing in Figs. 2 and 3 the significant properties of the interface modes for the $Si/SiO_2/Si$ and the $Si/ZrO_2/Si$ systems, respectively, as a function of the in-plane wave vector Q. These two MOS systems should be considered the two extreme cases of a low-κ (SiO_2) and a high-κ (ZrO_2) material, the $Si/SiO_2/Si$ system exhibiting some of the stiffest modes, the $Si/ZrO_2/Si$ some of the softest optical modes. The curves in the figures have been obtained using electron concentrations in the gate and in the Si substrate and an 'equivalent' insulator thickness, t_{eq}, (defined as $t\epsilon_{ox}^0/\epsilon_{SiO_2}^0$, *i.e.*, the thickness a layer of SiO_2 should have in order to yield the same low-frequency accumulation capacitance as for the layer of high-κ insulator of physical thickness t) which are representative of typical situations. The subband model employed has been described in Sec. II of Ref. 7. Basically, we have employed an anisotropic, non-parabolic band-structure, used a triangular-well approximation for the potential in the inversion layer, and embraced the long-wavelength approximation for the dielectric function discussed above, also ignoring intersubband-plasmons.

In the top-left frames (a) we show the dispersion of the modes, in (b) their phonon content, in (c) their gate-plasmon content, while in (d) and (e) we show the scattering strength for the phonon-like and gate-plasmon-like component of each excitation. Finally, frame (f) shows the total scattering strength for each mode. The scattering strength with the phonon-like component of each mode i, $\Lambda_{SO}^{(i)}(Q)$, has been defined, according to Eq. (34), by summing the scattering strength of both TO-modes α:

$$\Lambda_{SO}^{(i)}(Q) = \sum_{\alpha=1}^{2} \Lambda_{SO,\alpha}^{(i)}(Q) = \epsilon_0 \sum_{\alpha=1}^{2} \left| \frac{\hbar\omega_Q^{(i)}}{2} \times \left[\frac{1}{\epsilon_{TOT,high}^{(TO\alpha)}(Q,\omega_Q^{(i)})} - \frac{1}{\epsilon_{TOT,low}^{(TO\alpha)}(Q,\omega_Q^{(i)})} \right] \Phi^{(TO1)}(\omega_Q^{(i)}) \right| , \quad (43)$$

and similarly for the scattering strength for the gate-plasmon content of mode i, $\Lambda_{SP}^{(i)}(Q)$, defined, according to Eq. (33), as:

$$\Lambda_{SP}^{(i)}(Q) = \epsilon_0 \left| \frac{\hbar\omega_Q^{(i)}}{2\,\epsilon_{TOT}^{(PL)}(Q,\omega_Q^{(i)})} \Pi^{(G)}(\omega_Q^{(i)}) \right| . \quad (44)$$

(In both Eqns. (43) and (44) we have included a factor ϵ_0 to give the scattering strengths an 'intuitive' dimension of energy.) Note that, comparing Eq. (34) with Eq. (43) and Eq. (33) with Eq. (44), the (squared) scattering amplitude will be modulated by a factor Q^{-1} – which is compensated by the density-of-states factor Q while integrating over final scattering states. The scattering field will also exhibit the exponential decay $\exp[-Q(z-t)]$ into the substrate, away from the substrate/insulator interface.

Landau damping has been ignored in these figures. However, the two curves labeled ω_{LDg} and ω_{LDs} identify the region of strong damping of the gate and substrate plasma, respectively. As explained above in Sec. II D, Landau damping is approximately accounted for by ignoring the substrate plasma for values of Q such that $\omega^{(4)} \geq \omega_{LDs}$ (in both figures) and by ignoring the response of the gate electrons whenever $\omega^{(2)} \geq \omega_{LDg}$ (in Fig. 2) or $\omega^{(1)} \geq \omega_{LDg}$ (in Fig. 3).

In Fig. 2(a)-(c), the mode labeled $\omega^{(1)}$ appears clearly to be mostly a phonon-like mode, originating from the high-frequency SiO_2 TO-mode at about 0.135 eV. The mode $\omega^{(2)}$ is mainly a gate-plasmon mode, its coupling to the high-frequency phonon mode increasing at shorter wavelengths. The second, low-energy SiO_2 TO-mode at about 0.06 eV is strongly coupled to the substrate

plasmons. Indeed, the modes labeled $\omega^{(3)}$ and $\omega^{(4)}$ result from this strong coupling: The former is mostly phonon-like at small Q, but it quickly becomes mainly a substrate-plasma mode as Q grows, while the mode $\omega^{(4)}$ shows the opposite behavior. When examining the frames (d)-(f) we should keep in mind that the effect of each mode on the electron mobility depends not only on its scattering strength, but also on its energy: Modes of energy much larger than the thermal energy, $k_B T$, (where k_B is the Boltzmann constant and T the lattice temperature) and/or than the Fermi energy of the 2DEG in the Si substrate have a small effect, since, as already stated above, thermal electrons cannot emit excitations of such a large energy, while the thermal (Bose) population of these modes will be too small to induce any significant absorption at room temperature. For this reason the 'bulk' or 'gate/insulator-interface' modes, labeled $\omega^{(5)}$ and $\omega^{(6)}$ (for which we have computed phonon and gate-plasmon contents following a procedure similar to the one outlined above) are indeed expected to exhibit a negligible contribution: Mode $\omega^{(5)}$ has energy comparable to – actually, even slightly larger than – the phonon-like mode $\omega^{(2)}$, but it exhibits a much smaller scattering strength, especially at long and very short (not shown in the figure) wavelengths. Similarly, the scattering strength of mode $\omega^{(6)}$ should be compared with the scattering strengths of modes $\omega^{(3)}$ and $\omega^{(4)}$, which are also significantly larger.

Figure 3 conveys essentially the same information, but it shows how the stiffest mode, labeled $\omega^{(1)}$, is now mainly gate-plasmon-like, mode $\omega^{(2)}$ is mainly phonon-like at small Q, substrate-plasmon-like at shorter wavelength. The mode labeled $\omega^{(3)}$ is mainly phonon-like, but its phonon content switches from the low-energy (small Q) to the high-energy (large Q) insulator phonon mode, crossing the substrate-plasma mode at intermediate wavelength. The mode labeled $\omega^{(4)}$ starts as substrate-plasma-like at low Q, but it becomes almost completely phonon-like at larger Q. Finally, it can be seen that the modes $\omega^{(5)}$ and $\omega^{(6)}$ have a negligible scattering contribution, as it is the case for the SiO_2-system. A comparison between Fig. 2(e) and Fig. 3(e) shows weaker gate-plasmon scattering strength for the ZrO_2-system ($\omega^{(1)}$ in Fig. 3(e)) than for the SiO_2-system ($\omega^{(2)}$ in Fig. 2(e)). As explained before[7], this is simply due to the closer proximity of the gate in the SiO_2-system. Conversely, looking at the scattering strength of the $\omega^{(2)}$ mode, for example, at the largest values of Q in the undamped region, the ZrO_2-system exhibits a stronger scattering strength with phonon-like modes. Finally, in frames (d) of both figures, phonon-like scattering with modes $\omega^{(3)}$ (in Fig. 2(d)) and $\omega^{(2)}$ (in Fig. 3(d)) is significantly enhanced by the phonon-plasmon coupling. This effect – predicted also in bulk systems[12,17,19] – results from the anti-screening properties of the electron gas(es). Whenever the frequency of an excitation is larger than the frequency of the electron plasma, the coupling strength with the excitation is enhanced, while Landau damping gains strength. Indeed this effect is significant well within the region in which we must account for Landau damping. In our case the situation is noticeably complicated by the presence of two plasmas (the gate and the substrate) and by two optical phonons. As we shall see below discussing the effective electron mobility, each of the phonon-like excitations may be screened by one plasma and anti-screened by the other.

III. EFFECTIVE MOBILITY OF THE 2D ELECTRON GAS

A. The model

The calculation of the effective electron mobility in the inversion layer of the MOS systems considered here has been performed using the approximations and models described in detail in Sec. III of Ref. 7. We have employed Eq. (63) of that reference, using the total relaxation time computed by adding relaxation rates due to electron scattering with intravalley, intra- and inter-subband acoustic phonon using an anisotropic deformation potential (Eq. (86) of Ref. 7), scattering with intervalley phonons as described in Sec. III D of that reference, and considering scattering with the coupled plasmon/insulator-phonon interface modes employing Eq. (77) of Ref. 7, but substituting the 'effective' field amplitude $|\mathcal{A}_Q|^2$ with the scattering strengths $\Lambda^{(i)}_{SO,\alpha}(Q)/Q$ and $\Lambda^{(i)}_{SP}(Q)/Q$ given by Eqns. (43) and (44) above. The Si inversion layer has been modeled using the triangular well approximation. The number of subbands employed in the calculation of the mobility has been fixed by first considering only subbands containing a thermal (Fermi-Dirac) population of electrons larger than $10^{-4} n_s$. Having so determined the energy-bottom (*i.e.*, at $\mathbf{K} = 0$, where $\mathbf{K}$ is the 2D wave vector) of the 'topmost' subband required to account for the thermal electron occupation with a 10^{-4} accuracy, also those subbands whose energy-bottom lies at an energy $20 k_B T$ larger are included. This accounts correctly for absorption processes, since in the range of wavelengths and electron densities of interest the maximum energy which an electron can absorb from an interface excitation is smaller than $20 k_B T$ at room temperature,

The multiple integrals appearing in the Kubo-Greenwood-like expression for the mobility, Eq. (63) of Ref. 7, involving a two-dimensional integral over one energy and one angular variable, and in the expressions for the momentum relaxation times (Eq. (77) Ref. 7 for the SO-scattering processes, involving an integral over the angular variable, Eq. (86) for phonon-scattering) have been performed numerically using higher-order quadrature expressions. The complexity and computer requirements of the numerical procedure put a constraint on the maximum number of subbands (and so on the minimum electron density in the channel) which we could consider. Even so, some 'numerical noise' is evident in our results (up to 5% at the lowest density). Regarding the dispersion of the interface excitations employed, at the beginning of the calculation we have solved the secular equation (4) in the form (5) with a combination of methods: For a set of tabulated values of Q, a scan for sign-changes of the left-hand-side is done using a rather coarse ω-mesh. A bisection method is then used, followed

by Newton's method to converge quickly and accurately to each root. The roots of Eq. (5) so obtained (*i.e.*, the solutions $\omega_Q^{(i)}$) have been stored in look-up tables. In general there will be four of them in the undamped region of low Q, three or only two as Landau damping enters the picture at larger values of Q. We also tabulate, for each Q and branch i, the amplitude of the scattering fields and the group velocity of the modes, required to evaluate the Jacobian factor g (in the notation of Ref. 7). A linear interpolation of these functions of the magnitude of the momentum-transfer Q has been performed during the numerical integration of the SO-limited relaxation time.

The main significant difference with the procedure used before[7] is related to the different choice we have made for the electron concentration entering the evaluation of the gate plasma frequency. In Ref. 7 we had employed the depletion approximation for the polycrystalline gate and employed the electron density N_g obtained by averaging the electron concentration $N(z)$ ($z \leq 0$) over the width of the depletion layer. Here we have employed a slightly better approximation: Gauss' law requires that the surface field, F_s, used to determine the electron concentration, n_s, in the inversion layer of the Si substrate (see Eq. (41) of Ref. 7) be identical to the surface field at the gate/insulator interface ($z = 0^-$). We have integrated the Poisson equation into the gate up to a sufficiently large depth z_{max} using mixed boundary conditions for the electrostatic potential $V(z)$: Neumann at $z = 0^-$ (*i.e.*, $-dV(0^-)/dz = F_s$), and Dirichlet at $z = z_{max}$ (*i.e.*, $V(z = z_{max}) = -E_F^{(g)}$, where $E_F^{(g)}$ is the Fermi energy of the electron gas in Si at the density $n = N_{Dg}$, the donor concentration in the gate). Considering the problems arising when describing the plasma response of the inhomogeneous electron gas in the depleted gate region, we have performed the calculations making two different choices for the value of the electron concentration in the gate. The first choice consists in calculating the dispersion of the interface modes using $N_g = N(z = 0^-)$, the electron concentration at the gate/insulator interface obtained from the solution of Poisson equation. The second choice consists, instead, in employing for N_g the average quantity

$$< N_g(Q) > = \frac{\int_{z_{max}}^{0^-} N(z)\ e^{Qz}\ dz}{\int_{z_{max}}^{0^-} e^{Qz}\ dz} . \qquad (45)$$

This expression has a heuristic justification. The potential associated with the interface excitations has the form given by Eq. (6), exhibiting an exponential decay e^{Qz} for $z \leq 0$. Therefore, the 'effective' average electron density seen in the gate by an interface excitation will approach N_{Dg} at long wavelength and $N(z = 0^-)$ at short wavelength. Equation (45) empirically captures this behavior. Note that this Q-dependent average must be employed to compute the dispersion of the interface modes, since it appears via $\omega_{p,g}$, and so via $\epsilon_g(\omega)$, in the secular equation (4) (or (5)).

Finally, the nature of scattering with interface roughness causes significant problems. Even for the Si/SiO_2 system we have scant knowledge of the structural nature of the interface. Even worse, we must rely on a largely empirical model (Ando's, see Ref. 23) to handle electron scattering with interfacial roughness. Obviously, we know even less about the nature of the Si/insulator interface when considering an insulator other than silicon dioxide. Therefore, we have decided to follow a fully and shamelessly empirical approach: We have used Matthiessen's rule and added to the calculated mobility $\mu_{PH,SO,SP}$ – including scattering with Si phonons (PH), coupled surface-optical (SO), and interface plasmons (SP) modes - the contribution of a surface-limited mobility, μ_{SR} of the form $\mu_{SR} = \mu_0 (10^{13}/n_s)^2$, with n_s measured in cm^{-2}. The constant μ_0 has been determined by fitting the resulting 'total' mobility $\mu_{tot} = [1/\mu_{PH,SO,SP} + 1/\mu_{SR}]^{-1}$ calculated for thick (5nm) SiO_2-systems at $n_s = 10^{13}$ cm^{-2} to the experimental value in lightly-doped substrates of about 308 cm^2/Vs (see Ref. 24). The resulting value was $\mu_0 = 1.473 \times 10^3$ cm^2/Vs when using $N_g = N(z = 0^-)$, and $\mu_0 = 1.167 \times 10^3$ cm^2/Vs when using $N_g =< N_g(Q) >$. Converting the latter value to commonly used expressions, it implies a surface-roughness-limited mobility about a factor of 2 larger than what we obtained in the past[6] (using Monte Carlo simulations, and also used in Fig. 1, which assumed the values $\Lambda = 1.3$ nm and $\Delta = 0.48$ nm for the Ando's parameters), and within 10% of a typical empirical fit[25] to the effective electron mobility.

B. Insulator parameters

We have calculated the effective electron mobility in the inversion layer of MOS systems with the insulating layer consisting of 6 different materials: SiO_2, Al_2O_3, AlN, ZrO_2, HfO_2, and $ZrSiO_4$. These materials have been chosen as significant representatives of insulators being experimentally investigated at present, covering a wide range of static dielectric constant κ, and for which there is sufficient knowledge of the parameters required to evaluate the electron mobility.

In order to see which parameters are required to evaluate the electron/SO-modes scattering strength, it is convenient to rewrite Eq. (10) in the following more general form valid for insulators exhibiting N_{TO} TO-modes:

$$\epsilon_{ox}(\omega) = \epsilon_{ox}^{\infty} + \epsilon_0 \sum_{\alpha=1}^{N_{TO}} \frac{f_\alpha\ \omega_{TO,\alpha}^2}{\omega_{TO,\alpha}^2 - \omega^2} . \qquad (46)$$

In analogy with the notation used in Eq. (10), we can relate the oscillator strength f_α of the TO-mode α to 'intermediate' dielectric constants $\epsilon^{(\alpha)} \approx \epsilon_{ox}(\omega_\alpha)$, where $\omega_{\alpha-1} \leq \omega_{TO,\alpha} < \omega_\alpha$ for ω_α ordered so that $\omega_{\alpha-1} < \omega_\alpha$:

$$f_\alpha = \frac{\epsilon^{(\alpha-1)} - \epsilon^{(\alpha)}}{\epsilon_0} . \qquad (47)$$

It must be understood that $\epsilon^{(\alpha=0)} = \epsilon_{ox}^0$ and $\epsilon^{(\alpha=N_{TO})} = \epsilon_{ox}^{\infty}$. For a single TO-mode and, approximately, for TO-modes widely separated energetically, the Lyddane-Sachs-Teller (LST) relation provides the LO/TO splitting:

TABLE I. Parameters used to compute the electron-phonon coupling in polar insulators. The frequency of the optical phonons and/or the dielectric functions ϵ^0, ϵ^i, and ϵ^∞ are taken from the literature, when available, and the unavailable data are obtained using the Lyddane-Sachs-Teller relation. When more than two modes are present, only the two strongest modes (based on the magnitude of the LO-TO energy splittings) have been considered. When anisotropic quantities are given – such as in Refs. 34 and 47 for the energies of E_T and A_{2T} (bc tetragonal) or E_u and A_{2u} (hexagonal) modes propagating in directions perpendicular and parallel to the crystal c-axis, respectively, and also for the elements of the dielectric tensor – a simple average has been taken. The Fröhlich coupling constants for each mode are also indicated. Note the large coupling constants for the low-energy modes in ZrO_2 and HfO_2, which are indeed the materials yielding the lowest mobility.

Material Quantity (units)	$SiO_2^{(a)}$	$Al_2O_3^{(b)}$	$AlN^{(c)}$	$ZrO_2^{(d-f)}$	$HfO_2^{(f,g)}$	$ZrSiO_4^{(h)}$
ϵ_{ox}^0 (ϵ_0)	3.90	12.53	9.14	24.0$^{(i)}$	22.00	11.75
ϵ_{ox}^i (ϵ_0)	3.05	7.27	7.35	7.75	6.58	9.73
ϵ_{ox}^∞ (ϵ_0)	2.50	3.20	4.80	4.00	5.03	4.20
ω_{TO1} (meV)	55.60	48.18	81.40	16.67	12.40	38.62
ω_{TO2} (meV)	138.10	71.41	88.55	57.70	48.35	116.00
α_1	0.0248	0.0788	0.0248	0.2504	0.3102	0.0322
α_2	0.0113	0.0814	0.0423	0.0779	0.0362	0.2942

(a) Refs. 4 and 34 and references therein.
(b) Ref. 35.
(c) Ref. 27.
(d) Ref. 39.
(e) Ref. 40.
(f) Ref. 38.
(g) Ref. 44.
(h) Ref. 47 and references therein.
(i) Ref. 42.

$$\omega_{LO,\alpha}^2 \approx \frac{\epsilon^{(\alpha-1)}}{\epsilon^{(\alpha)}} \omega_{TO,\alpha}^2 . \quad (48)$$

For materials exhibiting two or more TO-modes at nearby frequencies, the energy of the LO-modes must be determined by the generalized LST-relation $\epsilon_{ox}(\omega) = 0$, such as Eq. (13) valid in the case of two TO-phonons. In order to determine completely the frequency dependence of $\epsilon_{ox}(\omega)$ in the model-form (46), for each bulk mode α we need knowledge of two of the quantities $\omega_{TO,\alpha}$, $\omega_{LO,\alpha}$, and $\epsilon^{(\alpha)}$ (or, equivalently, f_α). Ideally, experimental data can provide this information: Infrared absorption experiments can sense mainly $\omega_{TO,\alpha}$, Raman and electron tunneling-spectroscopy[26] can also sense $\omega_{LO,\alpha}$. The relative amplitude of each peak in these spectra can be correlated with the oscillator strength f_α. In addition, one could rely on theoretical calculations. Unfortunately, things are not so easy. We wish to calculate the electron mobility and, ultimately, compare our results with experimental data. But thin-insulator MOS structures are typically manufactured with processes which must be compatible with the current Si technology. The structure – and often even the composition – of the grown or deposited insulator can often be inferred only indirectly. The dielectric response of any given material can (and in general does) depend on its crystalline (or amorphous) structure: AlN, for instance, exhibits different properties in its wurtzite and zinc-blende structure[27]. It can also depend on its closeness to the ideal chemical composition: Incompletely oxidized Al_2O_3, for example, shows additional modes, possibly related to unoxidized Al ions[28]. Undesired, but so far unavoidable, 'native' interfacial layers (SiO_2 when dealing with oxides, Si_3N_4 when dealing with nitrides) can mask the response of the 'pure' dielectric under study, as discussed in Appendix A. Finally, the information available in the literature is incomplete, occasionally inconsistent. All we can do here is to try and make sense of the data and/or calculations reported elsewhere, and compare with Fourier transform infrared (FTIR) spectra we have obtained. Some of these have already been reported[29]. Additional spectra are shown in Fig. 4. Table I summarizes the values we have employed. In the table we also show the dimensionless coupling constant

$$\alpha_i = \frac{e^2}{4\pi\hbar} \left(\frac{m_t}{2\hbar\omega_{SOi}}\right)^{1/2} \left(\frac{1}{\epsilon_{Si}^\infty + \epsilon_{ox}^\infty} - \frac{1}{\epsilon_{Si}^\infty + \epsilon_{ox}^0}\right) , \quad (49)$$

for each of the two modes i, which corresponds to the dimensionless Fröhlich coupling constant usually defined in bulk materials[30]. These coupling constants have been obtained using for the energy of the SO-phonons the approximate expressions given by Eq. (40), with the appro-

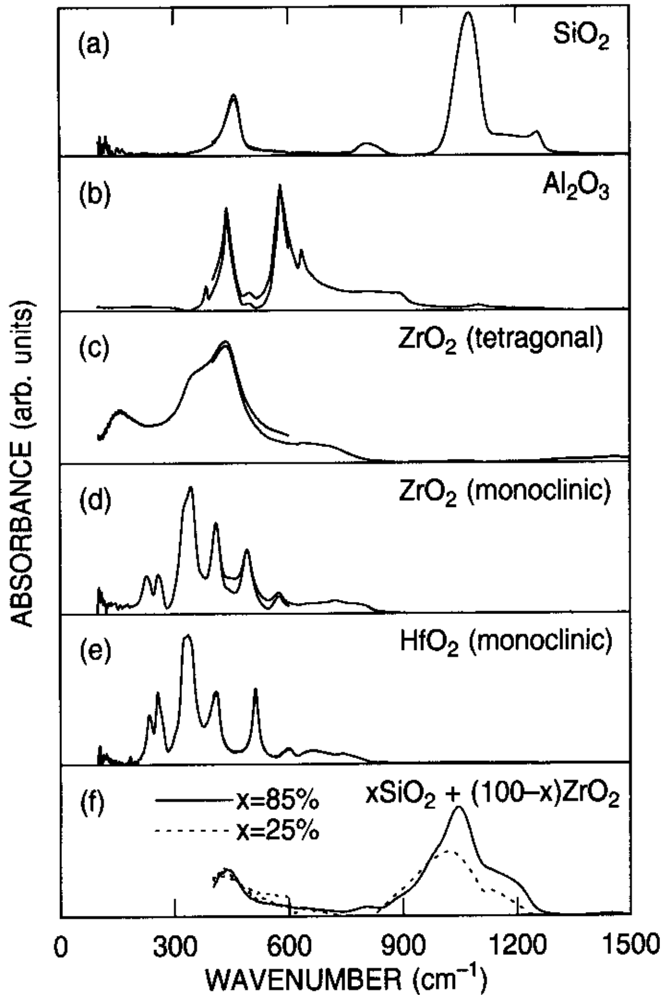

FIG. 4. FTIR spectra for some of the insulators considered. In all spectra (except for xSiO$_2$+(100-x)ZrO$_2$ with x=85% in the bottom panel) two curves are shown, one in the far-IR range of 50 to 600 cm^{-1} and one in the mid-IR range of 400 to 4000 cm^{-1}. In the bottom frame ('nominally' ZrSiO$_4$ obtained from chemical-solution deposition of xSiO$_2$+(100-x)ZrO$_2$) spectra for x=85% and x=25% are shown, to illustrate the decay of the SiO$_2$-mode at about 1080 cm^{-1} with decreasing x.

priate optical, static, and 'intermediate' dielectric constants. Note how well these values correlate with the mobility shown in Fig. 1. In particular, the high values of the coupling constants relative to the low-energy modes in HfO$_2$ and ZrO$_2$ hint very directly at the importance of remote scattering with SO-modes in MOS systems using these materials. Note also that the large magnitude of α_1 for these materials casts doubts on the accuracy of first-order perturbation theory (as implied by our use of the Fermi golden rule) to evaluate the momentum relaxation times relative to these processes.

The FTIR spectra were recorded with a Nicolet Nexus 670 spectrometer using a DTGS KBR detector and a KBR beam splitter for mid-IR (4000-400 cm^{-1}) and using a DTGS detector and a solid substrate beam splitter for far-IR (600-50 cm^{-1}) data collection. Spectra were collected from films deposited on a SiO$_x$N$_y$-coated Si wafer. Background spectra have been obtained using a portion of the SiO$_x$N$_y$-coated Si wafer on which the high-κ film had not been deposited. To minimize fringing effects and internal reflection, background and sample spectra were collected at $\sim$ 30^0 angle from the incident beam. The spectra were analyzed using OMNIC v. 5.2 (Nicolet) software. SiO$_2$ growth was observed from oxidation of the underlying Si wafer after annealing at temperatures $\geq$ 800 C, therefore a representative thermal SiO$_2$ film on Si was subtracted from the spectrum of the Al$_2$O$_3$ films annealed at 1200 C and from the monoclinic ZrO$_2$ films annealed at 900 C using the subtraction feature of the OMNIC v. 5.2 software. Peak position and area under the peak was determined using the peak-picking and peak-area functions of the software.

The FTIR spectra of SiO$_2$ were obtained from thermally grown SiO$_2$. The α-Al$_2$O$_3$ FTIR spectra were obtained from a chemical-solution-deposited film annealed at 1200 C for 60 minutes in oxygen. The FTIR spectra of tetragonal and monoclinic ZrO$_2$ were obtained from chemical-solution-deposited films annealed at 500 C and 900 C, respectively. The FTIR spectra of xSiO$_2$ +(100-x)ZrO$_2$ with x= 85% and 25% were obtained from a chemical-solution-deposited film annealed at 700 C. Additional details regarding deposition, phase formation, and FTIR analysis of the ZrO$_2$, HfO$_2$ and xSiO$_2$ +(100-x)ZrO$_2$ films can be found in the literature[29]. The monoclinic HfO$_2$ spectra were obtained from a film chemical-vapor-deposited at 700 C.

SiO$_2$. The SiO$_2$ FTIR spectrum shown in Fig. 4 exhibits two strong peaks at 1076 and 461 cm^{-1}, corresponding to TO-modes at 133.4 and 57.2 meV, associated with an asymmetric stretching of the SiO$_4$ unit and a bending of the Si-O-Si bond, respectively. A third mode at 806 cm^{-1} ($\approx$ 100 meV), due to a symmetric stretching mode of the Si-O-Si bond, is weak enough to be neglected. The shoulders at 1255 and 532 cm^{-1} are related to the corresponding LO modes. These values are in very good agreement with the experimental energies reported by Hess and Vogl[4] - from Refs. 31- 33. Also the LO/TO splittings are consistent with the oscillator strengths reported in the literature, but those derived from the areas under the FTIR peaks appear to give a stronger high-energy mode, the strength of the mode 1076 cm^{-1} being about 5 times stronger than the strength of the low-energy phonon. We have decided to comply with the most common values for the oscillator strengths reported in the literature. Indeed, recent calculations of the Raman-active intensities in α-quartz, based on a first-principle density functional approach, give a variety of modes.[34] Averaging the two strongest transverse modes over symmetry directions (the A_{2T} and E_T modes at the Γ point) gives two modes at about 1100 and 450-480 cm^{-1}. A similar average over the longitudinal modes (A_{2L} and E_L) provides the LO/TO energetic splitting and, via the LST relation Eq. (48) and Eq. (47), a ratio of 3:4 for the oscillator strengths of the modes.

The convincing agreement between our FTIR data and the published (experimental and theoretical) results for the energies and LO/TO splittings of the two dominant modes in silicon dioxide makes it for an easy choice of

parameters. The data listed in Table I are indeed in satisfactory agreement with all published results we are aware of.

Al_2O_3. The choice of dielectric parameters for Al_2O_3 is also a relatively easy task. Two peaks are clearly visible in the FTIR spectrum, at 579 and 437 cm^{-1} for a film deposited at 600 C and annealed in oxygen at 1200 C for 60 minutes to ensure complete oxidation. The areas under the peaks yield a ratio 56:44 for their respective oscillator strengths. High-resolution energy loss spectroscopy in thin Al_2O_3 films provides two sets of modes, in the plane of the film and off-plane[35]. The in-plane TO-modes (at 578 and 390 cm^{-1}) are in fair agreement with our FTIR results. Chen and co-workers[28] see a variety of modes as a function of annealing conditions of the thin films, typically grouped into three LO bands around 400-430, 600-655, and 850-895 cm^{-1}. The low-energy band is attributed to excess (unoxidized) Al. The remaining two bands are in satisfactory agreement with the TO energies and oscillator strengths which can be derived from the FTIR spectrum and Ref. 35. The selection reported in Table I can be viewed as a satisfactory compromise. We should add that we have used the bulk dielectric constant usually reported in the literature. Experiments on thin films can provide different results, Ref. 36 being just one example.

AlN. For AlN we rely exclusively on the results by Ruiz and co-workers[27] and by Gorczyca *et al.*[37]. The former is an *ab initio* Hartree-Fock study of the hexagonal (wurtzite) phase of AlN, the latter a muffin-tin theoretical analysis of both the wurtzite and zinc-blende phases. Additional theoretical and experimental results are reported in these references. In the hexagonal phase two almost degenerate TO modes at around 660 and 715 cm^{-1}, with oscillator strengths in a 59:41 ratio, originate from a single cubic TO mode at about 650 cm^{-1}, as a result of the doubling of the available optical modes moving from the cubic to the hexagonal structure. For the wurtzite phase, the modes reported in Table I have been obtained by averaging the modes over the various allowed symmetries. The particular structure selected (wurtzite or zinc-blende) is largely immaterial as far as electron scattering is concerned, since the total oscillator strength carried by the almost-degenerate modes in the hexagonal phase corresponds approximately to the oscillator strength of the the single mode in the cubic phase. A low-energy mode at 250 cm^{-1} seen in Raman spectra (reported in Refs. 27 and 37) is weak enough to be neglected.

ZrO_2. Desgreniers and Lagarec[38] have published Raman spectra for polycrystalline (cotunnite phase) HfO_2 and ZrO_2. For the latter insulator, they have observed two TO modes at wavenumbers of about 390 (an oscillation of the Zr-O bond) and 100 cm^{-1}, with the corresponding LO modes at about 430 and 170 cm^{-1}. Raman spectra by Morell and co-workers[39] for Y- and Ca-stabilized ZrO_2 give two LO modes at 620 and 160 cm^{-1}, the former mode possibly influenced by the dopants. Lattice-dynamics calculations for cubic and tetragonal lattices[40] give three transverse modes at wavenumbers 164 (E_u), 339, and 467 (A_{2u}) cm^{-1}, with corresponding longitudinal frequencies at 232, 354, and 650 cm^{-1}. These values, in rough agreement with the dispersions calculated by Mirgorodsky *et al.*[41] for cubic and tetragonal ZrO_2, show that the 339(TO)/354(LO)-cm^{-1} A_{2u}-mode is quite weak. In Fig. 4 we show FTIR spectra we have obtained for both tetragonal and monoclinic ZrO_2. The former exhibits a weak peak at 161 cm^{-1}, a weak shoulder at about 300 cm^{-1}, signature of the weak 339 cm^{-1} mode reported in Refs. 40 and 41, and a stronger structure around 439 cm^{-1}. These are in good agreement with both the experimental[38] and the theoretical[40,41] frequencies we have just discussed. The spectrum relative to tetragonal ZrO_2 is quite similar, while showing sharper peaks. Therefore, we have employed the LO-energies reported in Ref. 40, but have lowered the TO-frequency of the low-energy mode to account for the higher static dielectric constant ($\sim 24\ \epsilon_0$) observed in thin films[42]. Note that in FTIR spectra oscillations at wavenumbers below about 100 cm^{-1} remain elusive, as impurities (dopants) in the Si substrate render it opaque to the IR radiation.

HfO_2. The dominant high-frequency mode seen in the Raman spectra of Ref. 38 (TO at 395 cm^{-1} with corresponding LO at 450 cm^{-1}) – a vibration of the Hf-O bond – is also seen in the FTIR spectrum of this material in the monoclinic phase, as a double peak around 337-409 cm^{-1}. It corresponds to one of the modes of the monoclinic structure also reported in Ref. 43. The low-frequency mode seen in the spectrum of Ref. 38 (TO at 115 cm^{-1}, LO at 210 cm^{-1}), despite its strength is not easily visible in IR spectroscopy, as explained above. Thus, we have embraced essentially unaltered the results of Ref. 38, using the values for the refractive index from Refs. 45 and 46, and ignoring the weaker modes at 235 and 256 cm^{-1} seen in the spectrum of Fig. 4.

$ZrSiO_4$. FTIR spectra of chemical-solution deposited xSiO$_2$+(100-x)ZrO$_2$ (with x in percent) films, such as those shown in Fig. 4 or in Ref. 29, usually show two TO bands (a strong one at 430-460 cm^{-1}, a weaker one around 810-930 cm^{-1}). A strong signal around 1080 cm^{-1} can be attributed to residual 'unconverted' SiO_2, since, as shown by the dashed line in the bottom spectrum of Fig. 4, its intensity decreases with decreasing x. Averaging over the frequencies of the A_{2u} and E_u modes calculated for bc tetragonal $ZrSiO_4$ by Rignanese and co-workers[47], the two strongest modes appear to be at approximately 310 and 940 cm^{-1} (with corresponding LO wavenumbers at 410 and 1060 cm^{-1}). The relative oscillator strengths are approximately in the ratio 73:27. The low-energy mode can be assigned to an oscillation of the Zr-O bond, the high-energy mode to a vibration of the Zr-O-Si bond. Using the dielectric constants reported in Ref. 47 and the index of refraction from Ref. 48, we obtain the values shown in Table I.

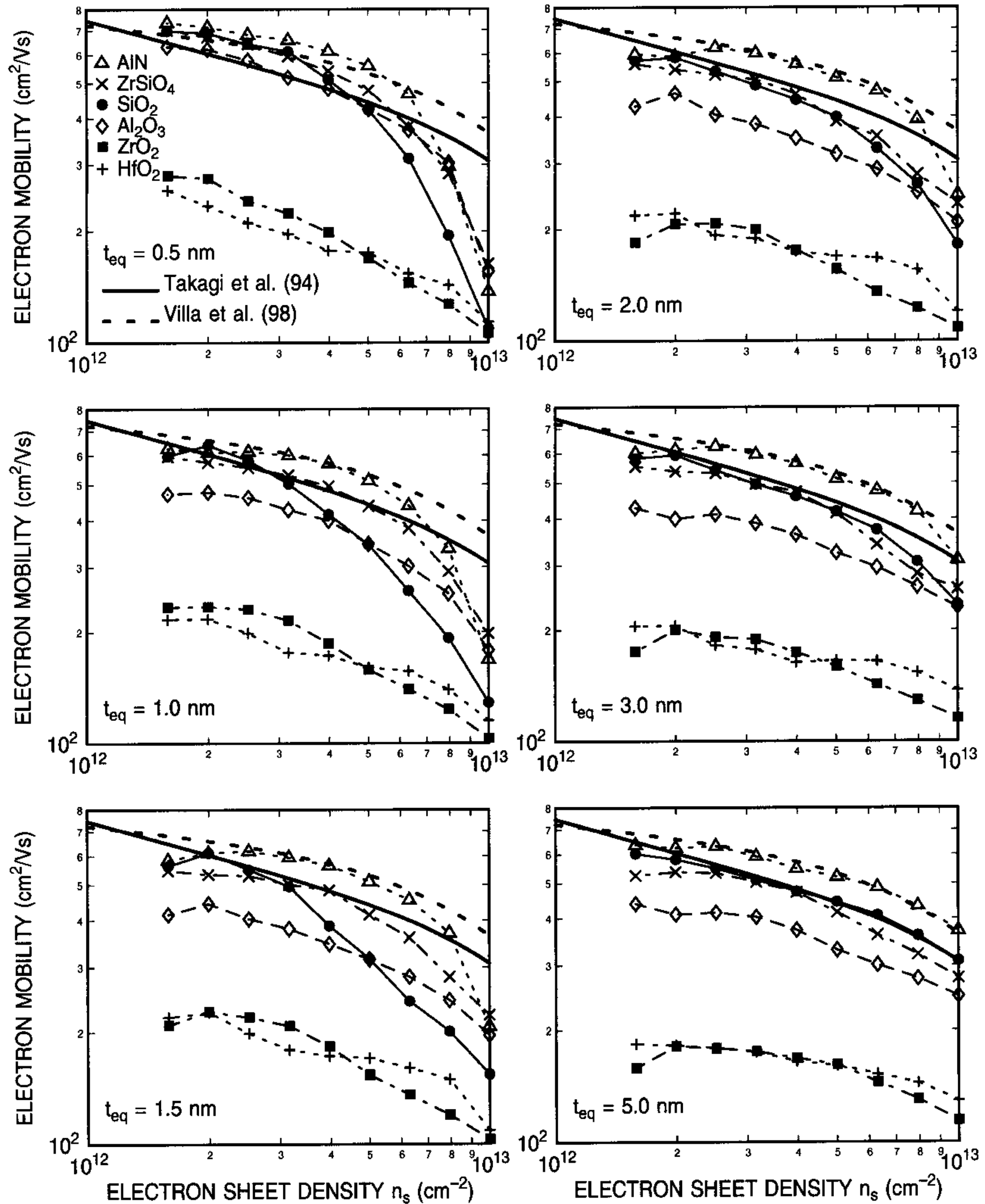

FIG. 5. Calculated effective electron mobility in the inversion layer of MOS systems with various insulators of the indicated 'SiO_2-equivalent' thicknesses. The Kubo-Greenwood expression for the mobility has been used by employing an anisotropic momentum relaxation time accounting for scattering with bulk Si phonons, gate/insulator interface plasmons (SP), and surface-optical (SO) insulator phonon. The effect of scattering with interface roughness at the substrate/insulator interface has been included by fitting the experimental mobility for the SiO_2-based system at an electron density of 10^{13} cm^{-2} and using Matthiessen's rule. The plasma response of the depleted poly-Si gate has been assumed to be given by the electron concentration at the gate/insulator interface determined by a numerical solution of the Poisson equation. In all frames the 'universal mobility curve' by Takagi *et al.*[24] (thick solid line) and an empirical fit from Ref. 25 (thick dashed line) show the range of experimental value. All data refer to systems with negligible scattering with dopants in the substrate or charges in the insulator.

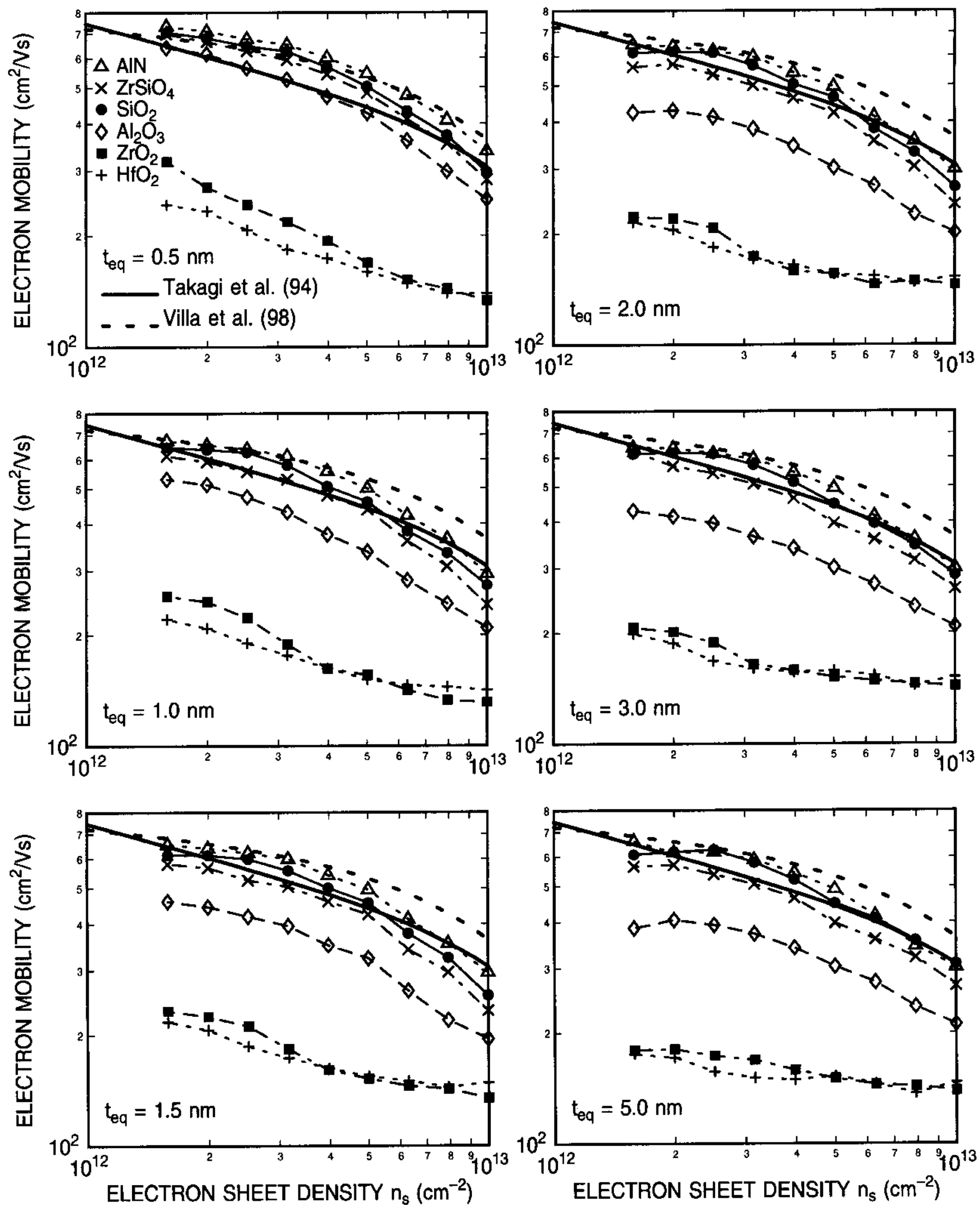

FIG. 6. As in Fig. 5, but with the plasma response of the depleted poly-Si gate now computed using a wave vector dependent average electron concentration in the depletion layer of the gate.

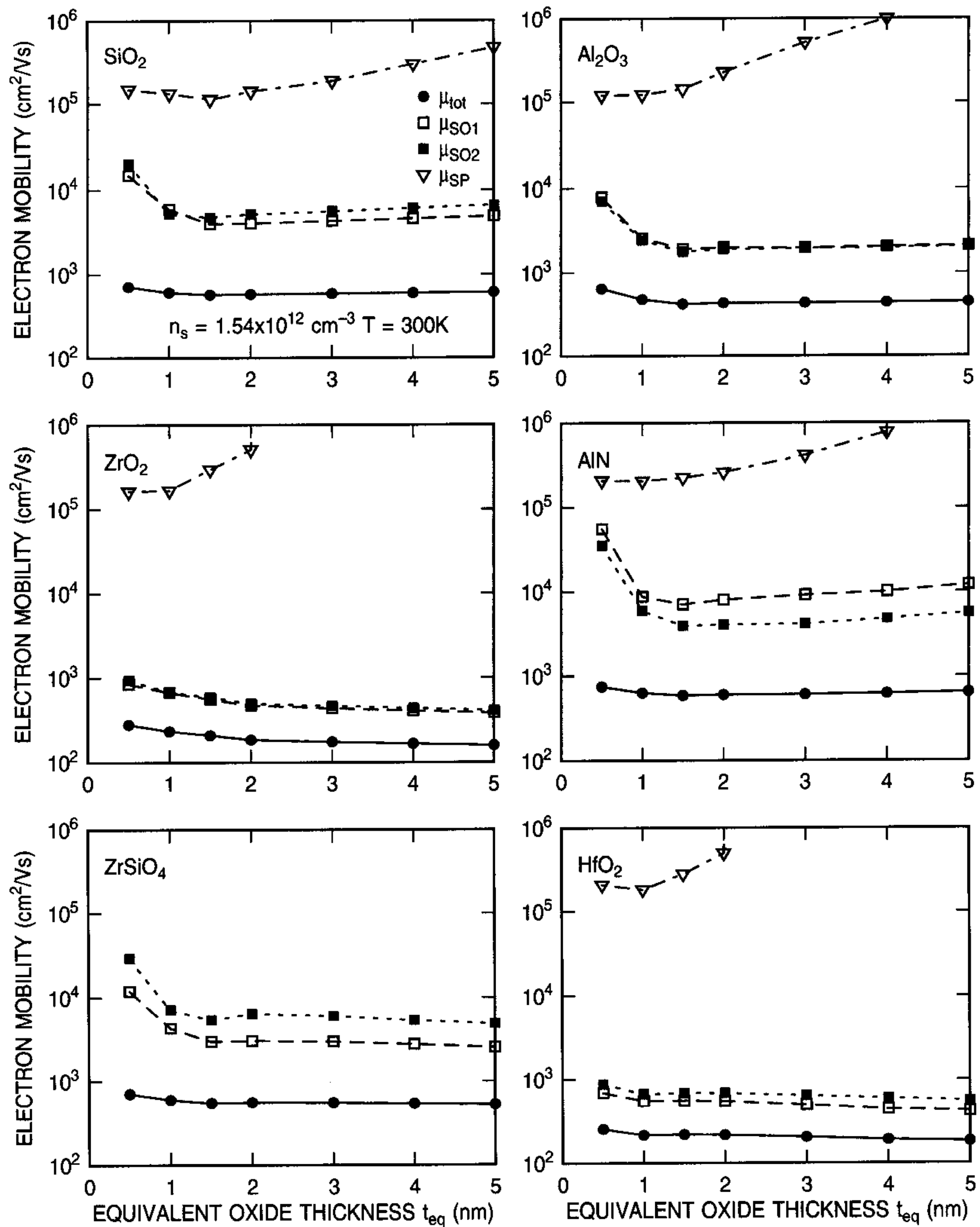

FIG. 7. Calculated components of the total mobility (μ_{tot}) at an electron sheet density in the substrate of 1.54×10^{12} cm^{-2} as function of the 'SiO_2-equivalent' thickness t_{eq} for MOS systems with various insulating materials. The curves labeled μ_{SO1} and μ_{SO2} refer to the components of the mobility limited by scattering with the TO1 and TO2 phonon-like components of the interface modes. The curve labeled μ_{SP} has been obtained from the total mobility and the SO-limited mobilities using Matthiessen's rule. The total mobility accounts for scattering with bulk Si phonons, SO- and SP-limited processes, but it does not account for scattering with interface roughness.

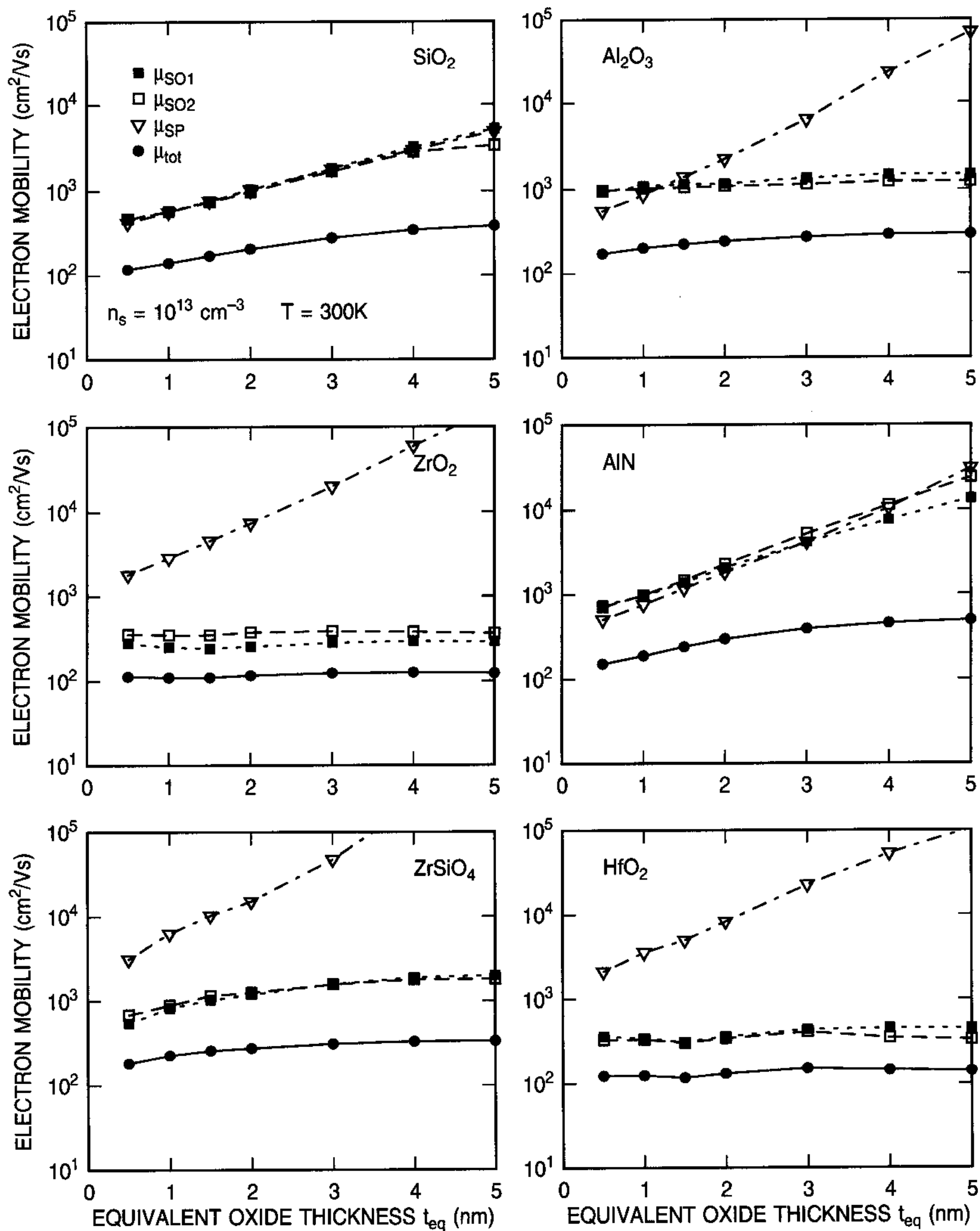

FIG. 8. As in Fig. 7, but at an electron density in the substrate of 10^{13} cm^{-2}.

C. Results

Before presenting the results of our calculations including the full dielectric response of the gate and substrate electron plasmas and their coupling with the optical modes of the insulating layer, it is interesting to revisit Fig. 1. The results shown in this figure illustrate the effect of the polar field of the optical modes of the insulators on the mobility of the electrons in the inversion layer, as determined by the parameters discussed in the previous section and by the simple Wang-Mahan scattering strength given by Eq. (1). The screening effect of the electrons in the inversion layer is ignored, as well as screening by the (infinitely far) gate.

SiO_2 (dots, solid line) is moderately affected by the presence of the SO modes, as a comparison with the curve calculated by neglecting them (open circles, dashed line) reveals. As stated in our introductory section, the stiffness of the Si-O bond results in a high-frequency LO-mode which couples poorly with thermal electrons, and a mode of lower frequency with a small coupling constant, Eq. (1). Thus, SO-modes have a very small effect, of about 5%, on the electron mobility. AlN has a somewhat larger dielectric constant (9.14) and in its wurtzite phase exhibits two almost degenerate modes at energies still larger than the thermal temperature. As for SiO_2, the electron mobility is only moderately affected by the presence of these phonons. But as soon as we consider materials with soft metal-oxygen bonds, the dielectric constant raises, and so does also the Wang-Mahan coupling. In addition, modes of lower energy – usually caused by oscillations of the oxygen ion in metal-O bonds – emerge and couple very effectively with thermal electrons. The insulators with the highest κ (ZrO_2 and HfO_2) can be seen in the figure to be negatively affected by the presence of low-energy modes and by the larger electron/SO-phonon coupling constant, showing the lowest mobility over the entire range of electron sheet densities. The dependence of the mobility on n_s is almost completely dominated by scattering with the SO-modes and by the electronic form-factor (overlap intergral) entering the expression for the relaxation rate associated with this process (see Eq. (79) of Ref. 7). Al_2O_3 and $ZrSiO_4$ are intermediate materials, both as far as their mobility as well as their dielectric constant are concerned.

In Figs. 5 and 6 we also show the effective mobility, but now accounting for the plasmon/TO-phonon coupling, and for the screening (or anti-screening) effects of the gate and substrate plasma. We show results relative to various SiO_2-equivalent insulator thicknesses. In Fig. 5 we have employed the gate/insulator surface electron concentration to determine the bulk plasma frequency of the gate, as determined by the solution of the Poisson equaton. In Fig. 6, instead, we have employed the Q-dependent average, $< N_g(Q) >$, given by Eq. (45) above. We see a few major differences between Fig. 1 on the one side, and Figs. 5 and 6 on the other: Screening of the electron-SO scattering by the substrate electron themselves results in a higher mobility for ZrO_2 and HfO_2 at large n_s: In these material the mobility is dominated by SO-scattering, the relevant SO-modes have relatively low energy, so their influence can be effectively screened by 2D-plasmons of higher frequency. At lower sheet densities, however, the frequency of the interface excitations becomes larger than the plasma frequency of the 2DEG in the substrate, and anti-screening takes effect, boosting the scattering rate and lowering the mobility with respect to the unscreend value shown in Fig. 1. At smaller t_{eq}, however, another interesting effect emerges: Screening by the electrons in the gate. For sufficiently small electron sheet densities in the substrate, $n_s \leq 5 \times 10^{12}$ cm^{-2}, SiO_2, AlN, Al_2O_3, and $ZrSiO_4$ exhibit mobilities approaching the value limited only by scattering with Si phonons. Even the mobilities of ZrO_2 and HfO_2 improve at these small densities, as a result of the competition between gate screening and substrate anti-screening.

The difference between the results shown in Fig. 5 and those shown in Fig. 6 is not qualitative, but only quantitative. Since the choice of a Q-dependent average $< N_g(Q) >$ results in a larger gate plasma frequency over most of the interesting range of values for Q, scattering with gate plasmons and the reduction of the SO-frequency is less pronounced in the dispersions used in Fig. 6. This results in large effective mobilities at large values of n_s.

In Figs. 7 and 8 we show the dependence on the thickness of the insulator of the various components of the mobility extracted from the data shown in Fig. 5. The mobilities limited by scattering with the SO1- and SO2-like components of the interface modes have been directly calculated from the knowledge of the (never vanishing) relaxation rate as a function of two-dimensional wave vector of the electrons in the inversion layer. On the contrary, the mobility limited by scattering with the gate-plasmon component of the interface excitations cannot be computed directly: In general, Landau damping will cause the relaxation rate to vanish for some particular value of the 2D electron wave vector. Therefore, this component (labeled SP below and in the figures) of the mobility has been extracted from the total mobility, from the SO1- and SO2-limited mobilities, and using Matthiessen's rule. In Fig. 7 the SP-component is very small even for SiO_2 and at the smallest t_{eq} investigated, because both of Landau damping and of the large gate-plasma frequency at the small electron density in the inversion layer (and so larger electron density in the gate) assumed in the figure ($n_s = 1.54 \times 10^{12}$ cm^{-2}).[7] Note also the screening effect of the gate plasma on the SO-limited mobility at small t_{eq}. On the contrary, at the larger electron density employed in in Fig. 8 ($n_s = 10^{13}$ cm^{-1}) we see all of the effects which our previous discussions, here and in Ref. 7, had anticipated: The SP-limited mobility is negligible in all materials, except obviously SiO_2 and also in the thinnest Al_2O_3 films. On the contrary, this 'advantage' of the high-κ materials is unfortunately more than compensated by a much stronger scattering with the TO-components of the interface excitation. The SO-limited mobility decreases at small t_{eq} – as also evident in Fig. 5 – because of the anti-screening effect of a strongly depleted gate: When using the gate/insulator interface electron concentration, the strongly depleted gate will exhibit a lower plasma frequency, and so it will be un-

able to screen the SO-component of the scattering field, actually anti-screening it.

Finally, in Fig. 9 we show the dependence of the SO-limited electron mobility on the static dielectric constant of the insulator. We have chosen a relatively small value for the electron concentration, in order to minimize the effect of scattering with surface roughness and with the gate-plasmon component of the interface excitations. Thus, the SO-limited component of the mobility is the major correction to the Si-phonon-limited component in the absense of Coulomb scattering with dopants and insulator charges. Note how the mobility decreases monotonically as κ increases, thanks to the softer oxygen bonds. AlN is indeed the single exception, thanks to the higher energy of the nitrogen-related optical phonons.

IV. DISCUSSION AND CONCLUSION

The mobilities shown in Figs. 5 and 6 show a clear trend, emphasized in Fig. 9. It appears that the price one must pay for a higher κ is a reduced electron mobility. Among the materials we have investigated, metal-oxides appear to be the worst, because of the soft modes caused by the oscillation of the oxygen ions, while AlN and, to some extent $ZrSiO_4$, show significant promise, albeit with the caveats we shall mention below.

Several sources of uncertainty affect the quantitative accuracy of our results: First and foremost is the choice of parameters for the insulators. This concerns both the overall quality of the parameters listed in Table I for 'ideal' materials, as well as their applicability to 'real' insulators, almost invariably of a 'non-ideal' composition and structure. This has been already discussed above and will be emphasized again in the following paragraph. Second, the difference between the results shown in Fig. (5) and those shown in Fig. (6) clearly points at the importance of knowing accurately the electron density in the depletion layer of the gate. Again, this source of uncertainty affects not only the 'ideal' calculations we have performed, but also the 'real world' complications we should expect: The poly-crystalline structure of the Si gate, for example, will undoubtely result in an electron concentration exibiting inhomogeneities not only in the z-direction, but also on the plane of the interface. Dopant segregation at grain boundaries is another possible cause of inhomogeneities. Nevertheless, our results stress at least qualititatively (but, hopefully, also quantitatively) the major role played by the gate both in screening the electron-SO interaction, as well as in triggering the gate-substrate Coulomb drag we had studied in Ref. 7. In particular, we should note that the use of metal gates should be beneficial in both cases, by inducing a more complete screening of the surface optical modes and by reducing the plasmon-mediated (long-range) component of the Coulomb drag. Finally, we have approximated the potential in the inversion layer as a triangular-well. This approximation is likely to be satisfactory at the large electron sheet densities of interest, but it will cause additional inaccuracies in the opposite limit of low n_s.

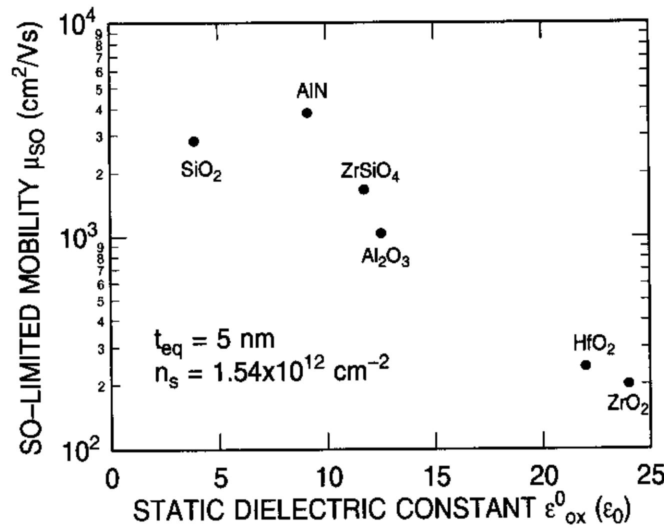

FIG. 9. Calculated SO-limited component of the electron mobility in the inversion layer of MOS systems with various insulators at an electron density of 1.54×10^{12} cm^{-2} as a function of the static dielectric constant of the insulator.

It would be interesting to support the results of our calculation with experimental evidence. Unfortunately, we have already alluded to the many experimental and processing complications which hamper a fair comparison. One could claim that, at least at present, the use of high-κ insulators has indeed resulted in disappointing performance in those few instances in which high-κ-based MOS transistors have been made in order to measure effective electron mobilities. ason *et al.*[49] have reported a peak mobility of 266 cm^2/Vs for MOSFET manufactured using aluminum gates and Al_2O_3 films of 'equivalent' thickness $t_{eq} \approx 2.9$ nm. Similarly, Qi and co-workers[50] have measured low mobilities when using 1.6 nm-thick $ZrSiO_4$ and 2.5 nm-thick ZrO_2 films, the former appearing to be about 40% better. Even lower values have been observed for other $ZrSiO_4$ films[42], and even for AlN films[51]. While some of these observations seem to agree quite nicely with our results (AlN being an exception we shall discuss shortly), it should be kept in mind that our calculations assume an ideal scenario: Perfectly stoichiometric films with no charges, electron traps (and the associated hysteresis) or interfacial layers. On the contrary, the structure or even the composition of the insulator itself is often unknown with the required accuracy. Charging effects, almost always seen, make it for a difficult, often impossible, accurate determination of the mobility (since an accurate determination of n_s becomes a hard task). Moreover, interfacial SiO_2 (or Si_3N_4) layers are almost always present. On the one hand this changes substantially the theoretical picture, with the additional complication arising from the coupling of the optical modes of two insulator and the presence of an additional interface (in Appendix A we estimate these effects in a simplified situation). On the other hand, the structural property of the interface, and not SO scattering, may dominate the experimental situation. This is probably the case for AlN-based MOSFETs, in which the Si_3N_4 interfacial layer, with the well-known associated electron

traps and instabilities, may completely mask the effects we are trying to observe. In conclusion, it is fair to say that, at least at present, we only have 'suggestive' and 'circumstantial' experimental evidence supporting our results. But no quantitative conclusions should be drawn.

ACKNOWLEDGMENTS

One of us (MVF) would like to acknowledge D. K. Ferry's suggestion to investigate in detail the role of coupled insulator-phonon/plasmon modes in MOS systems. Unpublished mobility data were kindly provided by Lars-Åke Ragnarsson and Sandro Callegari.

APPENDIX A: EFFECT OF A SILICON DIOXIDE INTERFACIAL LAYER

In this appendix we discuss the effect caused by the presence of a thin layer of SiO_2 between the Si substrate and the high-κ dielectric on the electron mobility in the Si inversion layer. In particular, we have in mind the *beneficial* effect of the interfacial layer in MOS systems based on materials (such as HfO_2 or ZrO_2) exhibiting a large ionic polarizability, so that removing the high-κ layer farther away from the Si substrate should reduce the strong interaction with the SO-modes of the high-κ dielectric. Therefore, the formation of a thin SiO_2 layer may be not only hard to avoid during the growth/deposition/annealing of the high-κ insulator, but also desirable. To formulate more precisely our expectations, note that the length-scale relevant for the calculation of the electron mobility is the Fermi wavelength of the 2DEG, $\lambda_F \sim K_F^{-1} \sim n_s^{-1/2}$. Since the scattering potential decays with increasing distance z from the high-κ insulator as $\exp(-Qz)$, the effect of an SiO_2 interfacial layer of thickness t_{ox} will be of 1. turning on the interaction with SiO_2 SO-modes, which we have seen is quite small, and 2. reducing the scattering strength of the high-κ SO-modes by a factor $\sim \exp(-2K_F t_{ox})$. At small electron sheet densities n_s, the small values of the Fermi wave vector means that unreasonably thick oxides are required in order to boost the electron mobility. Not so at large n_s (and so at large K_F), in which case even a 0.5-1.0 nm-thick SiO_2 layer can have a significant beneficial effect.

The analysis of the coupled plasmon-phonon modes of the full Si-gate/high-κ/SiO_2/Si system is quite cumbersome. Following a trivial generalization of the procedure described in Sec. II A, the secular equation – whose solutions yield the dispersion of the modes – takes the form:

$$\begin{aligned}&(\epsilon_g+\epsilon_\kappa)(\epsilon_s+\epsilon_{ox})(\epsilon_{ox}+\epsilon_\kappa)\\&+(\epsilon_g-\epsilon_\kappa)(\epsilon_s+\epsilon_{ox})(\epsilon_{ox}-\epsilon_\kappa)\,e^{-2Qt_\kappa}\\&-(\epsilon_g-\epsilon_\kappa)(\epsilon_s-\epsilon_{ox})(\epsilon_{ox}+\epsilon_\kappa)\,e^{-2Q(t_\kappa+t_{ox})}\\&-(\epsilon_g+\epsilon_\kappa)(\epsilon_s-\epsilon_{ox})(\epsilon_{ox}-\epsilon_\kappa)\,e^{-2Qt_{ox}}=0\,,\end{aligned}\quad\text{(A1)}$$

where ϵ_g, ϵ_κ, ϵ_{ox}, and ϵ_s (all functions of ω and, possibly, Q) are the dielectric functions of the Si gate, of the high-κ insulator of thickness t_κ, of the SiO_2 layer of thickness t_{ox}, and of the 2DEG in the Si substrate, respectively. Using the long-wavelength approximations for the dielectric functions, as in Sec. II A, and considering only two TO-modes in each of the two dielectric films, Eq. (A1) becomes an algebraic equation of 16-th degree in ω^2. The 16 positive solutions represent the dispersion of the 16 coupled modes resulting from 2 TO-like modes in each insulating film, and 12 surface modes (at large Q identifiable as groups of four modes mainly localized at each one of the three interfaces).

Our goal here is to investigate qualitatively the effects caused by the interfacial layer. Therefore, we abandon the idea of obtaining quantitatively correct information, and reduce the complexity of the problem by embracing the following set of approximations.

1. A comparison between Fig. 1 and Figs. 5 or 6 shows that the 'infinitely-thick insulator limit' captures the most important qualitative (and even quantitative) aspects of the problem. Thus, it seems appropriate to consider the simpler infinitely-thick-high-κ/SiO_2/Si system and ignore electronic screening effects. In this case, the dispersion of the modes is given by the solutions of the secular equation

$$\begin{aligned}\epsilon_{ox}(\omega)^2 \;+\; \epsilon_{ox}(\omega)[\epsilon_\kappa(\omega)+\epsilon_{Si}^\infty]\,\mathrm{cotanh}(Qt_{ox})\\+\;\epsilon_{ox}(\omega)\epsilon_{Si}^\infty \;=\; 0\,.\end{aligned}\quad\text{(A2)}$$

   This is exactly Eq. (5) with the role of the gate now played by the high-κ dielectric and with $\epsilon_s(Q,\omega)$ replaced by ϵ_{Si}^∞.

2. Low-energy SO modes are most important in determining the electron mobility. Thus, we approximate the response of the two insulators considering only their low-energy TO phonons and employ dielectric functions of the form:

$$\epsilon_{ox}(\omega) \;=\; \epsilon_{ox}^i\,\frac{\Omega_{LO}^2-\omega^2}{\Omega_{TO}^2-\omega^2}\quad\text{(A3)}$$

   and

$$\epsilon_\kappa(\omega) \;=\; \epsilon_\kappa^i\,\frac{\omega_{LO}^2-\omega^2}{\omega_{TO}^2-\omega^2}\quad\text{(A4)}$$

   where $\Omega_{LO} = (\epsilon_{ox}^0/\epsilon_{ox}^i)^{1/2}\Omega_{TO}$ and $\omega_{LO} = (\epsilon_\kappa^0/\epsilon_\kappa^i)^{1/2}\omega_{TO}$ are the longitudinal frequencies of the low-energy optical modes of the SiO_2 and high-κ layers, respectively.

3. Since we are most interested in understanding how 'low-mobility' materials behave in the presence of the interfacial oxide layer, and since these materials usually exhibit very soft optical modes, we typically have $\omega_{TO} << \Omega_{TO}$, as seen in Table I for HfO_2 and ZrO_2. Therefore, the SiO_2 and high-κ modes become largely decoupled. In the fully-decoupled limit, excellent approximations of the three solutions of Eq. (A2) are given by:

$$\omega_Q^{(\kappa)} \approx \left(\frac{\epsilon_\kappa^0 + \Delta_Q^{(\kappa)}}{\epsilon_\kappa^i + \Delta_Q^{(\kappa)}}\right)^{1/2} \omega_{TO} , \qquad \text{(A5)}$$

where $\Delta_Q^{(\kappa)} = \epsilon_{ox}^0[\epsilon_{ox}^0 + \epsilon_{Si}^\infty \coth(Qt_{ox})]/[\epsilon_{Si}^\infty + \epsilon_{ox}^0 \coth(Qt_{ox})]$, and

$$\omega_Q^{(\pm)} \approx \left(\frac{\epsilon_{ox}^0 + \Delta_Q^{(\pm)}}{\epsilon_{ox}^i + \Delta_Q^{(\pm)}}\right)^{1/2} \Omega_{TO} , \qquad \text{(A6)}$$

where $2\Delta_Q^{(\pm)} = (\epsilon_\kappa^i + \epsilon_{Si}^\infty)\coth(Qt_{ox}) \pm \{[(\epsilon_\kappa^i + \epsilon_{Si}^\infty)\coth(Qt_{ox})]^2 - 4\epsilon_{Si}^\infty \epsilon_\kappa^i\}^{1/2}$. The solution $\omega_Q^{(\kappa)}$ represents the SO-mode associated with the TO phonon of the high-κ film. For small Qt_{ox} the solution $\omega_Q^{(+)}$ approaches Ω_{TO}, thus being essentially a bulk SiO_2 TO-mode, while in the limit of large Qt_{ox} it approaches the frequency of the SiO_2 mode at the 'far' high-κ/SiO_2 interface. In either limit, this mode couples only weakly with the electrons in the inversion layer and – while retaining it – could safely be neglected, similarly to the modes labeled $\omega_Q^{(5)}$ and $\omega_Q^{(6)}$ in Sec. II A. Finally, the solution $\omega_Q^{(-)}$ is the SO-mode at the Si/SiO_2 interface associated with the SiO_2 TO-mode. All modes, $\omega_Q^{(\kappa)}$ and $\omega_Q^{(\pm)}$, exhibit a very weak dependence on Q. We shall ignore their dispersion and employ their short-wavelength limits, $\omega_Q^{(\kappa)} \approx \omega_{TO}[(\epsilon_\kappa^0 + \epsilon_{Si}^\infty)/(\epsilon_\kappa^i + \epsilon_{Si}^\infty)]^{1/2}$, $\omega_Q^{(+)} \approx \Omega_{TO}[(\epsilon_{ox}^0 + \epsilon_\kappa^\infty)/(\epsilon_{ox}^i + \epsilon_\kappa^\infty)]^{1/2}$, and $\omega_Q^{(-)} \approx \Omega_{TO}[(\epsilon_{ox}^0 + \epsilon_{Si}^\infty)/(\epsilon_{ox}^i + \epsilon_{Si}^\infty)]^{1/2}$.

The scattering strength associated with the high-κ mode can be obtained as in Sec. II C:

$$\Lambda^{(\kappa)}(Q) = \frac{\hbar\omega_Q^{(\kappa)}}{2} \left[\frac{1}{\epsilon_{TOT}^{\kappa,hi}(Q)} - \frac{1}{\epsilon_{TOT}^{\kappa,lo}(Q)}\right] , \qquad \text{(A7)}$$

where

$$\epsilon_{TOT}^{\kappa,hi/lo}(Q) = \epsilon_\kappa^{i/0} \left[\frac{\epsilon_{ox}(\omega) + \epsilon_{Si}^\infty}{\epsilon_{ox}(\omega) - \epsilon_\kappa(\omega)}\right]^2 e^{2Qt_{ox}} + \epsilon_{ox}(\omega) \left\{ \left[\frac{\epsilon_{ox}(\omega) + \epsilon_{Si}^\infty}{2\epsilon_{ox}(\omega)}\right]^2 (e^{2Qt_{ox}} - 1) + \left[\frac{\epsilon_{ox}(\omega) - \epsilon_{Si}^\infty}{2\epsilon_{ox}(\omega)}\right]^2 (1 - e^{-2Qt_{ox}}) \right\} + \epsilon_{Si}^\infty , \qquad \text{(A8)}$$

where the last expression is evaluated for $\omega = \omega_Q^{(\kappa)}$. Similarly, for the scattering strength of the modes $\omega_Q^{(\pm)}$ we get:

$$\Lambda^{(\pm)}(Q) = \frac{\hbar\omega_Q^{(\pm)}}{2} \left[\frac{1}{\epsilon_{TOT}^{ox,hi}(Q)} - \frac{1}{\epsilon_{TOT}^{ox,lo}(Q)}\right] , \qquad \text{(A9)}$$

where

$$\epsilon_{TOT}^{ox,hi/lo}(Q) = \epsilon_\kappa(\omega) \left[\frac{\epsilon_{ox}(\omega) + \epsilon_{Si}^\infty}{\epsilon_{ox}(\omega) - \epsilon_\kappa(\omega)}\right]^2 e^{2Qt_{ox}} + \epsilon_{ox}^{i/0} \left\{ \left[\frac{\epsilon_{ox}(\omega) + \epsilon_{Si}^\infty}{2\epsilon_{ox}(\omega)}\right]^2 (e^{2Qt_{ox}} - 1) + \left[\frac{\epsilon_{ox}(\omega) - \epsilon_{Si}^\infty}{2\epsilon_{ox}(\omega)}\right]^2 (1 - e^{-2Qt_{ox}}) \right\} + \epsilon_{Si}^\infty , \qquad \text{(A10)}$$

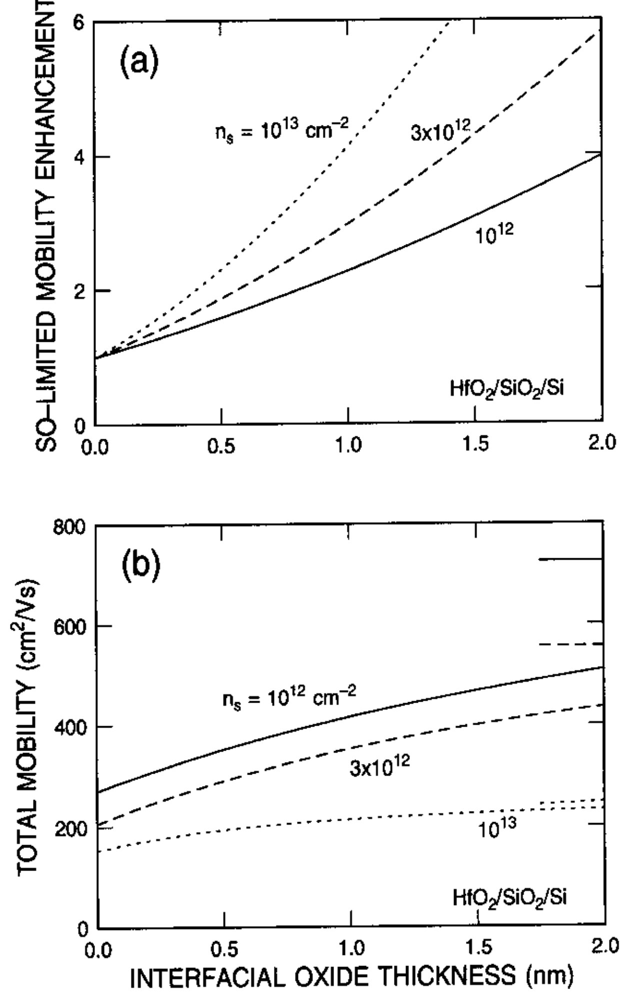

FIG. 10. (a) Calculated enhancement of the SO-limited mobility in the inversion layer of a HfO_2/SiO_2/Si system as a function of thickness of the interfacial SiO_2 layer for the three indicated values of the electron concentration in the inversion layer. An infinitely thick HfO_2 layer and only the low-energy TO-modes in the insulators have been considered. For a given thickness of the interfacial oxide, the enhancement is larger at large electron densities, since the length-scale is set by the Fermi wavelength of the two-dimensional electron gas. (b) Calculated total electron mobility (including scattering with Si phonons and interfacial roughness), as in (a). The horizontal lines at the far right are the asymptotic limits of infinite oxide thickness (*i.e.*, the mobility for the 'pure' Si/SiO_2 interface). As in (a), in order to approach the mobility at the Si/SiO_2 interface, a thicker SiO_2 interfacial layer is required at lower electron sheet density.

where $\epsilon_\kappa(\omega)$ and $\epsilon_{ox}(\omega)$ are evaluated at $\omega = \omega_Q^{(\pm)}$. Note that, as expected, $\Lambda^{(\kappa)}(Q) \sim e^{-2Qt_{ox}}$ while $\Lambda^{(\pm)}(Q) \to 0$ as $Qt_{ox} \to 0$.

4. Finally, we consider only one subband in the inversion layer. Therefore, the electron mobility can be obtained using Eqns. (68) and (85) of Ref. 7 (the latter equation unfortunately being mistyped in that reference: the factor $b^6 e^{-2Qt_{ox}}$ in the integrand should be replaced by unity), by substituting $|\mathcal{A}_Q|^2$ with $[\Lambda^{(\kappa)}(Q) + \Lambda^{(-)}(Q)]/Q$ in Eq. (85).

We show in Fig. 10 our results for the $HfO_2/SiO_2/Si$ system, similar results having been obtained also for the $ZrO_2/SiO_2/Si$ system. The top frame shows the enhancement of the SO-limited mobility resulting from the presence of the SiO_2 interfacial layer, *i.e.*, the ratio between the SO-limited mobility calculated for the $HfO_2/SiO_2/Si$ system with an SiO_2 layer of thickness t_{ox}, and the mobility calculated using the same approximations for the 'pure' infinitely- thick-HfO_2/Si system. As expected, a 1.0 nm-thick interfacial SiO_2 layer boosts the SO-limited mobility by a factor of more than 4 at the largest electron density considered (10^{13} cm^{-2}), but an SiO_2 layer as thick as 2.0 nm is required to obtain the same enhancement at the lowest density (10^{12} cm^{-2}). The bottom frame illustrates the dependence of the overall mobility (accounting also for scattering with Si phonons and surface roughness, included empirically using Matthiessen's rule). The horizontal lines at the far right indicated the asymptotic limit of an infinitely thick SiO_2 layer. Once more, at the largest density the mobility of the 'pure' infinitely-thick-SiO_2/Si system is recovered quickly even for thin interfacial layer. Not so at the lowest density: Even in the presence of a 1.0 nm-thick interfacial oxide layer, we remain almost a factor of 2 below the 'desired' SiO_2/Si limit.

---

[1] D. E. Kotecki, Integrated Ferroelectrics **16**, 1 (1996).

[2] T. Ohmi, K. Sekine, R. Kaihara, Y. Sato, Y. Shirai, and M. Hirayama, in *Material Research Society, Symposium Proceedings Volume 567*, H. R. Huff, C. A. Ritter, M. L. Green, G. Lukovsky, and T. Hattori editors (MRS, Warrendale, Pennsylvania, 1999), p.3.

[3] S. Q. Wang and G. D. Mahan, Phys. Rev. B **6**, 4517 (1972).

[4] K. Hess and P. Vogl, Solid State Comm. **30**, 807 (1979)

[5] B. T. Moore and D. K. Ferry, J. Appl. Phys. **51**, 2603 (1980).

[6] M. V. Fischetti and S. E. Laux, Phys. Rev. B **48**, 2244 (1993).

[7] M. V. Fischetti, J. Appl. Phys. **89**, 1232 (2000).

[8] M. V. Fischetti and S. E. Laux, J. Appl. Phys. **89**, 1205 (2000).

[9] We neglect here additional complications which may arise when dealing with layers of thickness comparable to the lattice constant or the size of the molecular bonds, such as localized/quantized phonon modes and electronic states.

[10] The physical reason why the 'far-interface' (*i.e.*, gate/insulator) SO-modes $\omega_Q^{(5)}$ and $\omega_Q^{(6)}$ do not contribute to the electron mobility in any appreciable way is twofold: At small values of Q the contribution to the momentum relaxation time is always negligible and, in addition, these modes are essentially bulk TO-modes with scattering strength vanishing in the limit $Qt \to 0$. At large Qt, instead, they are screened by the gate electrons at large electron densities in the gate, N_g, and so at small n_s. On the contrary, whenever the electron density in the (depleted) gate is lowered, the electron sheet density in the inversion layer, and so their Fermi wave vector K_F, is correspondingly raised. Thus, as discussed in Appendix A, the scattering strength of these modes is reduced by the factor $\sim e^{-2K_F t}$.

[11] As explained in Ref. 7, scattering with the substrate-plasmon component of the excitations is assumed to have no effect on the electron momentum relaxation rate, and so on the mobility, since it involves no direct loss of momentum by the 2DEG.

[12] M. E. Kim, A. Das, and S. D. Senturia, Phys. Rev. B **18**, 6890 (1978).

[13] This has been verified numerically to be true to a very high degree of accuracy (one part in 10^{15}). Presumably, an algebraic proof could also be found.

[14] E. A. Stern and R. A. Ferrel, Phys. Rev. **120**, 130 (1960).

[15] Using the present notation, in Ref. 7 the amplitude of the field caused by a coupled gate-substrate plasma oscillation of frequency ω_Q was given by:

$$\phi_Q(z) = \left[\frac{\epsilon_{ox}(\omega_Q) - \epsilon_g(\omega_Q)}{\epsilon_{ox}(\omega_Q) + \epsilon_s(Q,\omega_Q)}\right] \left(\frac{\hbar\omega_Q}{2Q\mathcal{D}_Q}\right)^{1/2} e^{-Qz}$$

where

$$\mathcal{D}_Q = \epsilon_g^\infty + \epsilon_i^\infty \left\{ \frac{[\epsilon_{ox}(\omega_Q) - \epsilon_g(\omega_Q)]^2}{4\epsilon_{ox}(\omega_Q)^2} (1 - e^{-2Qt}) + \frac{[\epsilon_{ox}(\omega_Q) + \epsilon_g(Q,\omega_Q)]^2}{4\epsilon_{ox}(\omega_Q)^2} (e^{2Qt} - 1) \right\} + \epsilon_{Si}^\infty \left[\frac{\epsilon_{ox}(\omega_Q) - \epsilon_g(\omega_Q)}{\epsilon_{ox}(\omega_Q) + \epsilon_s(Q,\omega_Q)}\right]^2 e^{-2Qt} .$$

Now dividing $\mathcal{D}_Q$ by e^{-2Qt} and by the square of the factor within square bracket in the expression for $\phi_Q(z)$ above, using the secular equation (4) to simplify the result, and finally noticing that in Ref. 7 we used $\epsilon_{ox}^0 = \epsilon_{ox}^\infty = 3.9\epsilon_0$, we recover the expression (33) of the text. The only remaining algebraic difference is due to the different definition for the gate-plasmon-content we have employed, since the definition used in Ref. 7 cannot be extended to the more complicated case of coupled phonon-plasmon excitations considered here. Notice, however, that despite the algebraic equivalence, physically the coupling of plasmons to phonons also alters the strength of the electron/gate-plasmon coupling via the number of dispersion branches present, via a shift of the plasma dispersion, and via a modification of the relative plasmon/phonon content of each branch.

[16] C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963).

[17] B. K. Ridley, *Quantum Processes in Semiconductors* (Clarendon Press, Oxford , 1988), pp. 332 ff.

[18] B. A. Sanborn, Phys. Rev. B **51**, 14256 (1995).

[19] B. B. Varga, Phys. Rev. **137**, A1896 (1965).

[20] S. Das Sarma, J. K. Jain, and R. Jalabert, Phys. Rev. B **37**, 6290 (1988).

[21] J. I. Gersten, Surf. Sci. **92**, 590 (1980); **97**, 206 (1980).
[22] A. Puri and W. L. Schaich, Phys. Rev. B **31**, 974 (1985).
[23] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
[24] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, IEEE Trans. Electron Device **41**, 2357 (1994)
[25] S. Villa, A. Lacaita, L. M. Perron, and R. Bez, IEEE Trans. Electron Device **45**, 110 (1998).
[26] See, for example, P. Balk, S. Ewert, S. Shmidtz, and A. Steffen, J. Appl. Phys. **69**, 6510 (1991).
[27] E. Ruiz, S. Alvarez, and P. Alemany, Phys. Rev. B **49**, 7115 (1994).
[28] P. J. Chen, M. L. Colaianni, and J. T. Yates, Phys. Rev. B **41**, 8025 (1990).
[29] D. A. Neumayer and E. Cartier, *Material characterization of ZrO_2-SiO_2 and HfO_2-SiO_2 Binary Oxides Deposited by Chemical Solution Deposition*, IBM Research Report RC 21932 01/15/01 (2001).
[30] Suitably modified (that is, removing ϵ_{Si}^{∞} and replacing m_t with the effective mass of the bulk polar material), this coupling constant α represents twice the number of LO-phonons constituting the 'polaron cloud' surrounding an electron in the bulk polar material. Equivalently, it expresses the ratio between the polaron and the LO-phonon energies. In our context we are more interested in the coupling between interface optical phonons and electrons in the 2DEG. Unfortunately, in this case the corresponding coupling 'constant' becomes a functional of the electron wave function in the inversion layer – via a 'form factor', or 'overlap factor'. Thus, we have decided to use the bulk expression for α even in our 2D-context, as it provides a qualitative measure of the strength of the interaction.
[31] D. A. Kleinman and W. G. Spitzer, Phys. Rev. **125**, 16 (1962).
[32] J. F. Scott and S. P. S. Porto, Phys. Rev. 161, 903 (1967).
[33] F. L. Galeener and G. Lucovsky, Phys. Rev. Lett. **37**, 1474 (976).
[34] P. Umari, A. Pasquarello, and A. Dal Corso, Phys. Rev. B **63**, 94305 (2001).
[35] B. G. Frederick, G. Apai, and T. N. Rhodin, Phys. Rev. B **44**, 1880 (1991).
[36] R. Ludeke, M. T. Cuberes, and E. Cartier, Appl. Phys. Lett. **76**, 2886 (2000).
[37] I. Gorczyca, N. E. Christensen, E. L. Peltzer y Blanca, and C. O. Rodriguez, Phys. Rev. B **51**, 11936 (1995).
[38] S. Desgreniers and K. Lagarec, Phys. Rev. B **59**, 8467 (1999).
[39] G. Morell, R. S. Katiyar, D. Torres, S. E. Paje, and J. Llopis, J. Appl. Phys. **81**, 2830 (1997).
[40] C. Pecharromán, M. Ocaña, and C. J. Serna, J. Appl. Phys. **80**, 3479 (1996).
[41] A. P. Mirgorodsky, M. B. Smirnov, P. E. Quintard, and T. Merle-Méjan, Phys. Rev. B **52**, 9111 (1995).
[42] A. C. Callegari, private communication.
[43] C. M. Philippi and K. S. Mazdiasni, J. Am. Ceram. Soc. **54**, 254 (1971).
[44] L. Kang, B. H. Lee, W.-J. Qi, Y. Jeon, R. Nieh, S. Gopalan, and J. C. Lee, IEEE Electron Dev. Lett. **21**, 181 (2000)
[45] M. Alvisi, S. Scaglione, S. Martelli, A. Rizzo, and L. Vasanelli, Thin Solid Films **354**, 19 (1999).
[46] T. Nishide, S. Honda, M. Matsuura, and M. Ide, Thin Solid Films **371**, 61 (2000).
[47] G.-M. Rignanese, X. Gonze, and A. Pasquarello, Phys. Rev. B **63**, 104305 (2001).
[48] J. D. Barrie and E. L. Fletcher, J. Vac. Sci. Technol. **A 14**, 795 (1996).
[49] L.-Å. Ragnarsson, S. Guha, N. A. Bojarczuk, E. Cartier, M. V. Fischetti, K. Rim, and J. Karasinski, *Electrical Characterization of Al_2O_3 n-channel Metal-Oxide-Semiconductor Field-Effect Transistors with Aluminum Gates*, unpublished (2001).
[50] W-J. Qi, R. Nieh, B. H. Lee, K. Onishi, L. Kang, Y. Jeon, J. C. Lee, V. Kaushik, B-Y. Prabhu, K. Eisenbeiser, and J. Finder, in 2000 Symp. on VLSI Technol. Dig., p. 40.
[51] L.-Å. Ragnarsson, private communication.