Effective-mass approximation in the presence of an interface

L. J. Sham
Department of Physics, University of California San Diego, La Jolla, California 92093

M. Nakayama*
Department of Physics, Brown University, Providence, Rhode Island 02912
(Received 29 January 1979)

For an electron in an inhomogeneous system such as a junction with an interface, the validity of including a sharply varying interface potential in the effective-mass equation is questioned. A new scheme is presented where the electron wave function is expressed in terms of eigenfunctions of the junction system including the interface. An effective-mass equation with accompanying boundary conditions is then derived to account for a slowly varying external potential. The various results, depending on the different types of interface reflections, are categorized. The theory is applied here mainly to the space-charge layer in a MOSFET, in particular to the splitting of the valley degeneracy in the n inversion layer of Si(100)-SiO_2.

I. INTRODUCTION

In a junction of a semiconductor with another material, such as an oxide or another semiconductor, the electron motion in the presence of an external potential is usually treated in the effective-mass approximation (EMA), in which the effective potential is taken to include the jump at the interface. The inversion (or accumulation) layer in a metal-oxide–semiconductor junction furnishes such an example. An electric field normal to the semiconductor-oxide interface draws the electrons (or holes) from the bulk of the semiconductor to the interface where they are prevented from entering the oxide by a large work function. The electron energies and wave functions are usually calculated in the EMA with an effective potential consisting of the attractive potential on the side of the semiconductor and an abrupt potential barrier at the interface.[1,2] The usual derivation of the EMA[3] requires the effective potential to be slowly varying. It is, therefore, questionable whether the potential jump at the interface should be included in the effective-mass equation.

In this paper, we present a theory[4] which accounts for the interface barrier outside the EMA and which treats only the external potential within the assumptions of the usual EMA. The usual EMA is derived by expanding the electron wave function in terms of Bloch waves of the semiconductor crystal.[3] The coefficients of this expansion form an envelope wave function obeying a Schrödinger equation with an effective potential which includes any potential over and above the periodic potential of the lattice. As this EMA is usually applied to the interface problem, the interface potential has to be included in the effective potential. To circumvent this undesirable procedure, we construct first, instead of the Bloch waves of the bulk semiconductor, a basis set of energy eigenfunctions of the inhomogeneous system including the interface under no external potential (i.e., in the flat-band condition, a notion more carefully defined in Sec. II). This means that the interface boundary conditions are already satisfied by the basis functions. The construction of the basis set is discussed in Sec. II.

Then, in the presence of an external potential (or a self-consistent one), the electron wave function is expanded in terms of the new wave functions which already account for the interface. The usual condition of a slowly varying potential is applied only to the external potential. A set of effective-mass equations with concomitant boundary conditions is then derived in Sec. III. Special attention is given to the semiconductor-oxide interface. Different kinds of behavior in the reflection of the semiconductor Bloch waves by the interface lead to different effective-mass equations and boundary conditions.

A single band minimum (or maximum) can have two types of behavior: (A1) the nonresonant type which leads to the usual EMA with the boundary condition of vanishing envelope wave function at the interface, and (A2) the resonant type which leads to a zero-gradient boundary condition. This is not difficult to understand if one remembers that the wave function, which decays in the insulator region, is a product of the envelope function times the Bloch wave at the band edge. Either part alone can insure the decay beyond the interface, leading respectively to case (A1) and case (A2).

A multivalley band structure can also lead to two types of behavior: (B1) the nonresonant type in which the different valleys are weakly coupled and in

which the splitting of the valley degeneracy can be simply expressed, and (B2) the resonant type in which the different valleys are strongly coupled and in which the boundary condition for the envelope function can be either zero value or zero first derivative. It is shown that if the external potential is far away from the interface, our theory reduces to the EMA in the bulk.

In Sec. IV, the p-type silicon with an (001) surface bordering silicon dioxide is used as an example for the general theory. The S matrices for the surface reflection are evaluated for a model. The effects on the subband energy levels are calculated. A particularly important application is to the calculation of the splitting of valley degeneracy of the lowest subband. Comparison with experimental measurements is discussed.

A number of theories[5–7] besides our own have been advanced to account for the valley splitting. In Secs. III and IV, the relation of some of these works to ours is briefly discussed.

The most interesting result of our theory is the effect of the interface on the electron dynamics in the space-charge layer. We believe that our general theory forms the basis for a more realistic account in the future of the interface effect on the electrons. This raises the hope of obtaining some information on the interface by a study of the inversion and accumulation layers. Possible directions of future work along these lines are discussed in Sec. V.

II. ELECTRON STATES UNDER THE FLAT-BAND CONDITION

For our purpose, the semiconductor-insulator system is envisaged as a semi-infinite semiconductor and a semi-infinite insulator joined together by an interface region, about a couple of atomic layers thick.[8] The flat-band condition[9] is defined specifically as one in which the one-electron potential outside the interface region is either the periodic lattice potential of the semiconductor or the potential of the bulk insulator. In this section, the construction of energy eigenstates in the flat-band condition is considered.

In principle, the procedure is well known.[10,11] Outside the interface region, the electron wave function is a linear combination of the bulk eigenfunctions at the same energy. For a crystal, the combination includes the Bloch waves and evanescent waves at the same energy. The combinations in the two bulk regions must be joined smoothly through the interface region. To be specific, we confine our attention to the fine details in a narrow energy range (of the order of 0.01 eV) near the conduction- or valence-band edge of the semiconductor. The range of energy is assumed to lie well in the energy gap of the insulator such that the electron or hole work function is of the order 1 eV. Thus, a wave function of an electron or hole from the semiconductor region will decay rapidly in the insulator region.

It is assumed that the crystal translation symmetry parallel to the interface for the semiconductor still holds. The crystal momentum parallel to the interface is a good quantum number. Thus the wave function in the semiconductor region is a linear combination of Bloch waves and evanescent waves with the same parallel crystal momentum.

We shall adopt the coordinate system in which the positive z axis is perpendicular to the interface pointing towards the bulk semiconductor. Let $\psi_i^{(-)}$, $\psi_j^{(+)}$, χ_λ with sets of values for i,j,λ be the set of eigenfunctions in the semiconductor with the same energy E and parallel wave-vector $\vec{k}_{\parallel}$ in the x-y plane. $\psi_i^{(-)}$ denotes a Bloch wave with the velocity vector towards the interface. $\psi_j^{(+)}$ denotes a Bloch wave with the velocity vector away from the interface. χ_λ is an evanescent wave decaying exponentially in the positive z direction. In the semiconductor region, the energy eigenfunction is

$$\phi_i = \psi_i^{(-)} - \sum_j S_{ji}\psi_j^{(+)} - \sum_\lambda T_{\lambda i}\chi_\lambda \ , \tag{2.1}$$

where S_{ij} and $T_{\lambda i}$ are constant coefficients.

If $\bar{\chi}_\mu$ denotes the set of evanescent waves at energy E and wave vector $\vec{k}_{\parallel}$ in the insulator region decaying exponentially in the negative z direction, the wave function in the insulator region is

$$\phi_i = \sum_\mu R_{\mu i}\bar{\chi}_\mu \ . \tag{2.2}$$

These two expressions should be joined smoothly through the interface region. Since later we shall be mainly concerned with the asymptotic behavior of ϕ_i in the semiconductor region, we choose, for simplicity, to match the wave functions at a plane $z=0$ placed at where the crystal potential of the perfect semiconductor ends. At this plane the wave function (2.1) is joined smoothly to the function of the type of Eq. (2.2) extended to cover the interface region. In this manner a set of energy eigenfunctions ϕ_i which takes into account the influence of the interface is constructed, ready to be used in the next step of our theory.

The elements S_{ji} in Eq. (2.1) form the S matrix which plays an important role in the properties of the electron in the space-charge layer. We now examine a number of different types of behavior of the S matrix.

A. Band with a single minimum (or maximum)

Consider the simplest case where the band structure of the bulk semiconductor has a single minimum at $k_z = k_0$ for a given fixed value of $\vec{k}_{\parallel}$. For the energy E slightly above the minimum, there are two

Bloch waves, an incoming one (i.e., towards the interface), $\psi_{k_0-\kappa}$, with the z component of the wave vector at $k_0-\kappa$ and an outgoing wave, $\psi_{k_0+\kappa}$, with the z component of the wave vector at $k_0+\kappa$. The S matrix is just a number, of the form

$$S=e^{2i\delta} \ , \tag{2.3}$$

where δ is the phase shift, as a consequence of the conservation of flux.

S is determined from the boundary conditions at the interface for the wave function. The details are relegated to Appendix A. S is given by the ratio of two determinants

$$S=D(-\kappa)/D(\kappa) \ . \tag{2.4}$$

In the important case of $k_0=0$, $\vec{k}_{\parallel}=0$ from time-reversal symmetry

$$D(-\kappa)=D^*(\kappa) \ . \tag{2.5}$$

S indeed takes the form Eq. (2.3) with the phase shift given by

$$\tan\delta=-D_2(\kappa)/D_1(\kappa) \ , \tag{2.6}$$

where D_1 and D_2 are the real and imaginary part of D, respectively. Equation (2.5) dictates

$$D_1'(0)=0 \ \text{ and } \ D_2(0)=0 \ . \tag{2.7}$$

From the behavior of $D_1(0)$, two limiting cases for S can be distinguished

a. Case (A1) nonresonant scattering. $D_1(0)$ is finite. For small κ, the phase shift is

$$\delta=-\kappa D_2'(0)/D_1(0)\equiv -a_s\kappa \ , \tag{2.8}$$

defining a scattering length a_s. In the interior of the semiconductor where the evanescent waves have died down, the asymptotic form of the wave function (2.1) is

$$\phi_\kappa\sim[\sin\kappa(z-a_s)]\psi_{k_0} \ . \tag{2.9}$$

The amplitude modulation now vanishes at $z=a_s$ from the interface $z=0$. The approximation used,

$$\psi_{k_0+\kappa}=e^{i\kappa z}\psi_{k_0} \ ,$$

is found to be adequate for small κ.

b. Case (A2) resonance scattering. If $D_1(0)$ vanishes, the leading term in $D_1(\kappa)$ for small κ is $O(\kappa^2)$. Hence, $\delta\to\frac{1}{2}\pi$ and

$$S\simeq -1 \ . \tag{2.10}$$

The condition that $D(0)=0$ is the same for the existence of bound surface states extending from energies below the band minimum continuously to $k_z=k_0$. The asymptotic form of the wave function at E above the band edge is

$$\phi\sim(\cos\kappa z)\psi_{k_0} \ . \tag{2.11}$$

In contrast to case (A1), the envelope function does not vanish at the interface but its first derivative does.

The two cases enumerated above are of course only limiting cases. $D_1(0)$ may happen to be so small but nonzero that, for the range of values of κ which is of interest, $D_1(0)$ is comparable to $D_2'(0)\kappa$. This constitutes the intermediate case.

B. Band with two minima

Consider the example of silicon bounded by an (001) plane. For a given small $\vec{k}_{\parallel}$, the conduction band has two minima at $k_z=\pm k_0$. Label the valley around k_0 as 1 and the one around $-k_0$ as valley 2. At an energy E slightly above the band edge E_0, there are four Bloch waves, two incoming $\psi_1^{(-)}$, $\psi_2^{(-)}$ and two outgoing $\psi_1^{(+)}$, $\psi_2^{(+)}$ with the z components of wave vectors given by

$$k_1^{(\pm)}=k_0\pm\kappa \ , \tag{2.12}$$
$$k_2^{(\pm)}=-k_0\pm\kappa \ ,$$

where

$$\kappa=[2m_z^*(E-E_0)/\hbar^2]^{1/2} \ . \tag{2.13}$$

The S matrix is 2×2, with the form

$$S_{ij}(\kappa)=N_{ij}(\kappa)/D(\kappa) \ . \tag{2.14}$$

These expressions are derived in Appendix B by matching wave functions at the interface. Again, two limiting cases can be discerned.

a. Case (B1) nonresonant scattering. The denominator of the S matrix $D(0)$ is finite. The S matrix has the form, from Eq. (B2),

$$S=\begin{bmatrix}1-\beta\kappa & -\alpha^*\kappa\\ \alpha\kappa & 1+\beta^*\kappa\end{bmatrix} \ . \tag{2.15}$$

The Bloch wave propagating towards the interface is reflected back mostly as the Bloch wave in the same valley, but a small amount is reflected back as the Bloch wave in the other valley. This surface-induced intervalley scattering leads naturally to an intervalley coupling, the consequences of which will be considered later. In Appendix B, it is also shown that the S matrix is consistent with current conservation[12] and explicitly that the set of wave functions ϕ_i with various E satisfies the orthogonality relation.

If we separate β into its real and imaginary parts, $\beta=\beta_1+i\beta_2$, then it is shown in Appendix B that β_1 is related to the deviation of the energy dispersion from the quadratic form. β_1 can therefore be

neglected for small κ. β_2 is clearly related to the phase shift and the intravalley scattering length a_s is equal to $\frac{1}{2}\beta_2$. The asymptotic form of the energy eigenfunctions are, for small κ,

$$\phi_{1\kappa} \sim 2\sin\kappa(z-a_s)\psi_{k_0} - i\alpha\kappa e^{i\kappa(z-a_s)}\psi_{-k_0} , \quad (2.16a)$$

$$\phi_{2\kappa} \sim 2\sin\kappa(z-a_s)\psi_{-k_0} + i\alpha^*\kappa e^{i\kappa(z-a_s)}\psi_{k_0} . \quad (2.16b)$$

We have used the approximation $\psi_{k_0+\kappa} = e^{i\kappa z}\psi_{k_0}$. The correction to the first order in κ can be shown to play no role in subsequent calculations.

In order to treat bound states due to an external potential, it is more convenient to express the reflected wave in a different valley also as standing waves, by applying the transformation

$$\Psi_{1\kappa} = \phi_{1\kappa} - \tfrac{1}{2}\alpha\kappa\phi_{2\kappa} \sim 2\sin\kappa(z-a_s)\psi_{k_0} - i\alpha\kappa\cos\kappa(z-a_s)\psi_{-k_0} , \quad (2.17a)$$

$$\Psi_{2\kappa} = \phi_{2\kappa} + \tfrac{1}{2}\alpha^*\kappa\phi_{1\kappa} \sim 2\sin\kappa(z-a_s)\psi_{-k_0} + i\alpha^*\kappa\cos\kappa(z-a_s)\psi_{k_0} . \quad (2.17b)$$

These functions for positive values of κ form the basis set for later use.

b. Case (B2) resonant scattering. This occurs when $D(0)=0$. This is also the condition for the existence of a surface state at $E=E_0$. From the orthogonality or flux conservation requirement, one can see that the diagonal elements of S are zero and the off-diagonal elements can be made unity by an appropriate choice of the phase of the Bloch wave. Hence, the S matrix is of the form

$$S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \quad (2.18)$$

The wave functions in the asymptotic region are

$$\phi_{1\kappa} \sim \psi_{k_0-\kappa} - \psi_{-k_0+\kappa} , \quad \phi_{2\kappa} \sim \psi_{-k_0-\kappa} - \psi_{k_0+\kappa} . \quad (2.19)$$

A Bloch wave from one valley is reflected by the interface into the other valley. Intervalley coupling is very strong.

C. Infinite potential barrier at the interface

A numerical computation of the S matrix requires a knowledge of the band structures of the insulator and semiconductor as well as that of the interface region. In a metal-oxide–semiconductor junction, the oxide is amorphous and the band structure is not well known. We make use of the high work function (of the order of electron volts) and approximate the insulator and the interface region by an infinite potential wall placed at $z=0$. The matching problem is reduced to requiring the wave function (2.1) vanishes at $z=0$. The infinite potential barrier model not only simplifies the computation but retains the qualitative features of the interface scattering. Since the S matrix depends on the relative distance z_0 from the interface to the nearest crystal plane of the semiconductor, by varying z_0 we obtain a range of possible behavior for the interface reflection, including the limiting case of resonant scattering, as will be shown in Sec. IV.

III. EFFECTIVE-MASS EQUATION

Now, to the flat-band condition add a spatially slowly varying external potential $v(\vec{r})$. We confine our attention to the one-electron problem and simply regard $v(\vec{r})$ as the total self-consistent potential which includes the electron-interaction effects.[13] The slowly varying potential $v(\vec{r})$ mixes the eigenstates in the flat-band condition only from a small amount of phase space. In particular, only one band of Bloch waves will be included. The derivation of the effective-mass equation follows the standard procedure.[3] The only difference is that the electron wave function is expanded in terms of the basis set of eigenfunctions in the flat-band condition considered in Sec. II, which already satisfy the boundary conditions at the interface. This expansion is substituted in the Schrödinger equation and an equation for the envelope wave function is derived. We shall consider in this section a few specific cases.

For simplicity, the external potential is considered as a function of normal distance z from the interface only. The motion parallel to the interface is taken to be the usual effective-mass approximation. Possible deviation from this will be reserved for future study.

For the envelope wave function, we introduce the notion of an "effective-mass interface." This ideal plane $z=z_{\text{eff}}$ serves as a boundary for the external potential $v(z)$ and the envelope wave function which only exist in the region $z > z_{\text{eff}}$. Because the external potential is slowly varying, the electron wave function is dominated by the asymptotic region where the evanescent waves have become negligible. Then, the envelope function is an amplitude modulation of the Bloch wave. The "effective-mass interface" is chosen at a distance near the true interface such that the orthogonality of the asymptotic wave functions is not violated. For example, in case (A1) and case (B1), $z_{\text{eff}}=a_s$ and in case (A2) and case (B2), $z_{\text{eff}}=0$.

a. Case (A1) the single-valley nonresonant case.

The basis set is given by case (A1) in Sec. II. The electron wave function when the external potential is introduced can be expressed as

$$\psi(\vec{r}) = \int_0^\infty \frac{d\kappa}{2\pi} A(\kappa)\phi_\kappa \ , \tag{3.1}$$

The "effective-mass interface" is at $z = a_s$. For convenience, move the origin to the effective-mass interface, which is then given by $z = 0$. For ϕ_κ, the asymptotic approximation (2.9) is used for $z > 0$, and it is taken to be zero at $z < 0$.

To convert the Schrödinger equation in terms of this basis set, we require the matrix elements of the total Hamiltonian given by

$$H = H_0 + v \ , \tag{3.2}$$

where H_0 is the Hamiltonian for the system in the flat-band condition, with matrix elements given by

$$\langle\phi_\kappa|H_0|\phi_{\kappa'}\rangle = (E_0 + \hbar^2\kappa^2/2m_z^*)2\pi\delta(\kappa - \kappa') \ , \tag{3.3}$$

where κ and κ' take on only non-negative values. To evaluate the matrix element of the external potential, we make use of the slow spatial variation of v and take $|\psi_{k_0}|^2$ in the integrand to be averaged over a unit cell to the value of unity.[3] The integral is confined to the side $z \geq 0$. By introducing $v(-z) \equiv v(z)$, and the Fourier transform over all space,

$$v(k) = \int_{-\infty}^\infty dz\, e^{-ikz} v(z) \ , \tag{3.4}$$

we can write the potential matrix element as

$$\langle\phi_\kappa|v|\phi_{\kappa'}\rangle \simeq v(\kappa - \kappa') - v(\kappa + \kappa') \ . \tag{3.5}$$

Then, the Schrödinger equation is reduced approximately to

$$\left(\frac{\hbar^2\kappa^2}{2m_z^*} - \epsilon\right)A(\kappa) + \int_0^\infty \frac{d\kappa'}{2\pi}[v(\kappa - \kappa') - v(\kappa + \kappa')]A(\kappa') = 0 \ , \tag{3.6}$$

where $\epsilon = E - E_0$.

Now, we extend the values of κ to negative numbers as well and define

$$A(-\kappa) = -A(\kappa) \ . \tag{3.7}$$

Then Eq. (3.6) becomes

$$\left(\frac{\hbar^2\kappa^2}{2m_z^*} - \epsilon\right)A(\kappa) + \int_{-\infty}^\infty \frac{d\kappa'}{2\pi} v(\kappa - \kappa')A(\kappa') = 0 \ , \tag{3.8}$$

with κ in the range $(-\infty, \infty)$. If we define the envelope function as

$$A(z) = \int_{-\infty}^\infty \frac{d\kappa}{2\pi} A(\kappa) e^{ikz} \ , \tag{3.9}$$

then the actual wave function is given by

$$\psi(\vec{r}) \simeq A(z)\psi_{k_0}(\vec{r}) \ . \tag{3.10}$$

From Eq. (3.8), the envelope function satisfies the effective-mass equation

$$\left(-\frac{\hbar^2}{2m_z^*}\frac{\partial^2}{\partial z^2} + [v(z) - \epsilon]\right)A(z) = 0 \ . \tag{3.11}$$

From Eq. (3.7), at the effective-mass interface

$$A(z = 0) = 0 \ . \tag{3.12}$$

We note that this boundary condition is derived together with the effective-mass equation (3.11) from the integral equation (3.6) and that it is not imposed *ad hoc* or by appealing to an additional physical reasoning that the electron wave function must vanish at the interface.

The EMA derived here is identical in form to the one usually used in the inversion layer of a MOSFET (metal-oxide–semiconductor field-effect transistor).[2] There is a minor difference in energy of the amount $a_s\langle\partial v/\partial z\rangle$, because the effective-mass interface is taken at a distance a_s from the ideal interface.

b. Case (A2) the single-valley resonant case. The appropriate basis set is given by case (A2) in Sec. II. The procedure for the derivation of the effective-mass equation is the same as in case (A1). The cosine nature of the basis function (2.11), however, causes the matrix element of the potential to be

$$\langle\phi_\kappa|v|\phi_{\kappa'}\rangle \simeq v(\kappa - \kappa') + v(\kappa + \kappa') \ . \tag{3.13}$$

The difference from Eq. (3.5) that the two terms now add forces one to define for $\kappa > 0$,

$$A(-\kappa) = A(\kappa) \tag{3.14}$$

in contrast with Eq. (3.7), if one desires to arrive at the same form of the effective-mass equation (3.8) or (3.11). Hence, in the present case, with the envelope function defined by Eq. (3.9), $A(z)$ does not vanish at the effective-mass interface. Instead, from Eq. (3.14), the boundary condition is

$$\partial A/\partial z|_{z=0} = 0 \ . \tag{3.15}$$

If the external potential is extended to $z < 0$ with mirror symmetry at $z = 0$, then in case (A1) only odd parity states are allowed and in case (A2) only even parity states are allowed.

c. Case (B1) the two-valley nonresonant case. In terms of the basis set, Eq. (2.17), the electron wave

function is given by

$$\psi(\vec{r}) = \sum_{\sigma=1}^{2} \int_0^\infty \frac{d\kappa}{2\pi} A_\sigma(\kappa) \Psi_{\sigma\kappa} \ . \tag{3.16}$$

The matrix elements of the external potential are evaluated in the approximation of slowly varying potential and neglecting intervalley terms such as $\langle \psi_{k_0} | v | \psi_{-k_0} \rangle$. The diagonal element is

$$\begin{aligned} \langle \Psi_{\sigma\kappa} | v | \Psi_{\sigma\kappa'} \rangle &\simeq \langle 2\sin\kappa z | v | 2\sin\kappa' z \rangle \\ &= v(\kappa-\kappa') - v(\kappa+\kappa') \ , \end{aligned} \tag{3.17}$$

as in Eq. (3.5), with $v(-z) = v(z)$ also. The off-diagonal element,

$$\begin{aligned} \langle \Psi_{1\kappa} | v | \Psi_{2\kappa'} \rangle &\simeq \left\langle 2\sin\kappa z \left| \frac{1}{2}\alpha \left[\frac{1}{i}\frac{\partial}{\partial z}, v(z) \right] \right| 2\sin\kappa' z \right\rangle \\ &= w(\kappa-\kappa') - w(\kappa+\kappa') \ , \end{aligned} \tag{3.18}$$

where $w(\kappa)$ is the Fourier transform of $w(z)$, given by

$$w(z) = (\alpha^*/2i)(\partial v/\partial z), \quad z > 0 \tag{3.19}$$

and

$$w(-z) = w(z) \ . \tag{3.20}$$

Similarly

$$\langle \Psi_{2\kappa} | v | \Psi_{1\kappa'} \rangle = w^*(\kappa-\kappa') - w^*(\kappa+\kappa') \ . \tag{3.21}$$

Since we have included neither the intervalley contribution $\langle \psi_{k_0} | v | \psi_{-k_0} \rangle$, as was done in Ref. 5, nor the spin-orbit coupling which is the valley coupling mechanism of Kümmel,[6] the intervalley matrix element arises entirely out of S_{12}, the scattering from one valley to the other by the interface.

As in case (A1), the form of the potential terms (3.17) and (3.18) dictates the extension of the range of k to negative values by

$$A_\sigma(-\kappa) = -A_\sigma(\kappa) \ . \tag{3.22}$$

The Schrödinger equation is then reduced to a coupled set of effective-mass equations.

$$\left(-\frac{\hbar^2}{2m_z^*}\frac{\partial^2}{\partial z^2} - \epsilon \right) A_\sigma(z) + \sum_{\sigma=1}^{2} v_{\sigma\sigma'}(z) A_{\sigma'}(z) = 0 \ , \tag{3.23}$$

where

$$v_{\sigma\sigma'}(z) = \begin{pmatrix} v(z) & w(z) \\ w^*(z) & v(z) \end{pmatrix} \ . \tag{3.24}$$

The envelope wave function is given by

$$\begin{aligned} A_\sigma(z) &= \frac{1}{i}\int_{-\infty}^{\infty} \frac{d\kappa}{2\pi} e^{i\kappa z} A_\sigma(\kappa) \\ &= \int_0^\infty \frac{d\kappa}{\pi} \sin\kappa z \, A_\sigma(\kappa) \ , \end{aligned} \tag{3.25}$$

which yields the boundary conditions

$$A_\sigma(z=0) = 0 \ . \tag{3.26}$$

If the intervalley potential terms $w(z)$, $w^*(z)$ are neglected in the coupled set of effective-mass equations, the usual result of double valley degeneracy is recovered. If w and w^* are small, by the degenerate perturbation theory, the splitting of the valley degeneracy is given by

$$\Delta E = |\alpha \langle \partial v/\partial z \rangle| \ , \tag{3.27}$$

where

$$\left\langle \frac{\partial v}{\partial z} \right\rangle = \int_0^\infty dz \, |A(z)| \, \frac{\partial v}{\partial z} \ , \tag{3.28}$$

with $A(z)$ being the envelope function for each valley with the potential $v(z)$ uncoupled from the other valley, normalized in the range $z > 0$.

d. Case (B2) the two-valley resonant case. With the basis set of case (B2) in Sec. II, subject to the same approximation for the potential matrix element as case (B1), we have

$$\begin{aligned} \langle \phi_{\sigma\kappa} | v | \phi_{\sigma'\kappa'} \rangle &= v(\kappa-\kappa')\delta_{\sigma\sigma'} \\ &\quad - v(\kappa+\kappa')(1-\delta_{\sigma\sigma'}) \ . \end{aligned} \tag{3.29}$$

For this form of the potential terms, if we let

$$A(\kappa) = \begin{cases} A_1(\kappa), & \kappa > 0, \\ -A_2(-\kappa), & \kappa < 0 \ , \end{cases} \tag{3.30}$$

then the Schrödinger equation reduces to the same effective-mass equation as Eq. (3.8) or (3.11).

Now with the even-parity potential, $v(-z) = v(z)$, the envelope function $A(z)$, Fourier-transformed from $A(\kappa)$, can be either odd, i.e.,

$$A(z=0) = 0 \ , \tag{3.31}$$

or even, i.e.,

$$\partial A/\partial z|_{z=0} = 0 \ . \tag{3.32}$$

The even- and odd-parity wave functions are given by

$$\psi_\pm(\vec{r}) = \pm A_\pm(z)\psi_{k_0} - A_\pm(z)\psi_{-k_0} \ , \tag{3.33}$$

with the (+) sign for even and the (−) sign for odd parity. The coexistence of states of both parities is brought on by the resonance nature of the S matrix.

In the foregoing, we have demonstrated that, even with the usual assumptions made in the effective-mass approximation for the bulk semiconductor, there are additional effects due to the interface, provided that the interface potential is not included in the EMA. These interface effects should disappear if the external potential v is entirely confined in the

bulk of the semiconductor away from the interface. It can be shown explicitly that the effective-mass equations derived for all four cases above, indeed reduce to the bulk limit if v is moved into the interior. In Appendix C, this will be done for case (B1) and case (B2). The procedure for the single-valley cases is similar.

IV. APPLICATION TO INVERSION LAYER ON p-Si (001) SURFACE

Electron energies and wave functions in the inversion layer of the (001) surface of p-type Si have been calculated in the usual EMA.[2,14,15] The lowest subband comes from the two conduction-band valleys $(0, 0, \pm k_0)$ where $k_0 = 0.85 \times (2\pi/a)$, with lattice constant a, and has, therefore, a twofold valley degeneracy. In this section, the general theory described above is employed to calculate the energy splitting of this valley degeneracy. The computation consists of two parts: first the S matrix and then the solution of the coupled effective-mass equations. To avoid a large amount of numerical work, we compromise by employing (i) the $\vec{k}\cdot\vec{p}$ band-structure model of Cardona and Pollack,[16] (ii) a model for the interface of an infinite barrier in a plane at a distance z_0 from a (001) atomic plane, and (iii) the separation of motion parallel to the surface by the usual EMA.

The next subband,[14,15] E_0', is supposed to come from the conduction valleys $(\pm k_0, 0, 0)$ and $(0, \pm k_0, 0)$. This furnishes an example of the single-valley case and will be treated briefly at the end of this section.

A. Band-structure model for bulk silicon

Fifteen plane waves with reciprocal-lattice vectors $\langle 000\rangle$, $\langle 111\rangle$, $\langle 200\rangle$ are used to form symmetrized plane waves at Γ, consisting of $2\times\Gamma_1$, Γ_{15}, $2\times\Gamma_{25'}$, $2\times\Gamma_{2'}$, $\Gamma_{12'}$. These are diagonalized with pseudopotentials $V_3 = -0.105$ a.u., $V_8 = 0.02$ a.u., $V_{11} = 0.04$ a.u. to form the basis of 15 periodic wave functions at Γ. The energy and wave function throughout the Brillouin zone are given by the $\vec{k}\cdot\vec{p}$ method, with the momentum matrix elements given in Ref. 16.

We approximate the S matrix by the values at $\vec{k}_{\parallel} = 0$, because the Fermi wave vector for the inversion layer is small compared with the dimension of the Brillouin zone. Then, we need only consider the solutions in bulk Si with real and complex wave vectors $(0, 0, k)$ along the crystal axis, i.e., belonging to the Δ symmetry. The 15×15 $\vec{k}\cdot\vec{p}$ Hamiltonian matrix is reduced to a 3×3 matrix for doubly degenerate Δ_5 states, a 3×3 matrix for Δ_1 states, a 5×5 matrix for $\Delta_{2'}$ states and an isolated $\Delta_{1'}$ matrix (a scalar).

B. States for a given energy

Given the energy E, we need to find all the relevant Bloch waves and evanescent waves. The 15×15 $\vec{k}\cdot\vec{p}$ Hamiltonian yields 30 solutions of k for a given E. Now the number of linearly independent plane waves as a function of x and y from the set of 15 plane waves projected on the plane $z =$ constant is 9. In order to solve the boundary conditions $\phi_i = 0$ at the interface using Eq. (2.1), the number of outgoing Bloch waves plus evanescent waves decaying along $+z$ direction must be also 9. Including Bloch waves going in the opposite direction and evanescent waves decaying in the $-z$ direction, there should be 18 values for k. Thus, care must be exercised in selecting the 18 correct solutions out of the 30. In the general case of an infinite-sized Hamiltonian matrix, the excess solutions represent repetition of the same solution in different Brillouin zones. In the truncated matrix here, the superfluous solutions are likely to be inaccurate. Hence, solutions with Rek outside the first Brillouin zone are discarded.

For $E = E_0 + \hbar^2\kappa^2/2m_l$ just above the conduction-band minimum, the 3×3 matrix for Δ_1 yields four running waves with wave vectors $\pm k_0 \pm \kappa$. The 5×5 $\Delta_{2'}$ matrix yields four evanescent waves with wave vectors $\pm i\times 0.095$ a.u. and $\pm i\times 0.835$ a.u., neglecting terms of $O(\kappa^2)$. The 3×3 Δ_5 matrix yields four pairs of evanescent waves. The Δ_1' yields two evanescent waves.

For a fixed z, these solutions can be expanded in terms of the nine planes waves in two-dimensional x-y space, with wave vectors $\langle 00\rangle$, $\langle 11\rangle$, $\langle 20\rangle$. Since $\vec{k}_{\parallel} = 0$ has the $\bar{\Gamma}$ symmetry in two-dimensional space,[17] these 9 plane waves form into combinations of $4\times\bar{\Gamma}_1$, $\bar{\Gamma}_2$, $2\times\bar{\Gamma}_3$, $2\times\bar{\Gamma}_4$. Since only the three-dimensional Bloch waves with Δ_1 symmetry and evanexcent waves with Δ_2' symmetry are compatible with the $\bar{\Gamma}_1$ symmetry, the other evanescent waves can be ignored.

C. S matrix

Choose the origin of the Cartesian coordinates to be at Si atom. Let the idealized interface be at $z = z_0$. The four plane waves in x-y space with $\bar{\Gamma}_1$ symmetry at $z = z_0$ are

$$
\begin{aligned}
\xi_1 &= 1 \ , \\
\xi_2 &= 2\left(\cos\frac{2\pi}{a}x\cos\frac{2\pi}{a}y\cos\frac{2\pi}{a}z_0\right. \\
&\qquad \left. +\sin\frac{2\pi}{a}x\sin\frac{2\pi}{a}y\sin\frac{2\pi}{a}z_0\right) \ , \\
\xi_3 &= 2\left(\cos\frac{2\pi}{a}x\cos\frac{2\pi}{a}y\sin\frac{2\pi}{a}z_0\right. \\
&\qquad \left. -\sin\frac{2\pi}{a}x\sin\frac{2\pi}{a}y\cos\frac{2\pi}{a}z_0\right) \ , \\
\xi_4 &= \cos\frac{4\pi}{a}x + \cos\frac{4\pi}{a}y \ . \qquad (4.1)
\end{aligned}
$$

The Bloch waves ψ_k of Δ_1 symmetry are expanded in terms of them, at $z=z_0$,

$$\psi_k=\sum_{i=1}^{4} c_i(k)\xi_i \ . \tag{4.2}$$

The evanescent waves χ_λ of $\Delta_{2'}$ symmetry are given by, at $z=z_0$,

$$\chi_\lambda=\sum_{i=1}^{4} b_{i\lambda}\xi_i \ . \tag{4.3}$$

The phases of the bulk Si wave functions are chosen such that $\psi_k^* = \psi_{-k}$ and that the evanescent waves are real. Hence, the coefficients $c_1(k)$, $c_2(k)$, $b_{i\lambda}$ are real. $c_3(k)$ is purely imaginary. $c_4(k)$ vanishes because the lowest Δ_1 conduction band contains no $\langle 200\rangle$ plane waves.

For Eqs. (4.2) and (4.3), we select the incoming Bloch waves $\psi_{\pm k_0-\kappa}$ and the outgoing Bloch waves $\psi_{\pm k_0+\kappa}$ and two evanescent waves with wave vectors $i\times 0.095$ a.u. and $i\times 0.835$ a.u. which decay exponentially as $z\to+\infty$. Following the general procedure outlined in Appendices A and B, we substitute these expansions in Eq. (2.1). The vanishing of the wave function $\phi_i=0$ at $z=z_0$ leads to a set of linear equations, from which the S matrix is obtained in the form

$$S_{ij}(\kappa)=N_{ij}(\kappa)/D(\kappa) \ . \tag{4.4}$$

The denominator is

$$D(\kappa)=\begin{vmatrix} c_1(k_0+\kappa) & c_1(k_0-\kappa) & \tilde{b}_1 \\ c_2(k_0+\kappa) & c_2(k_0-\kappa) & \tilde{b}_2 \\ c_3(k_0+\kappa) & -c_3(k_0-\kappa) & \tilde{b}_3 \end{vmatrix} , \tag{4.5}$$

with

$$\tilde{b}_j=b_{j1}-b_{j2}b_{41}/b_{42} \ . \tag{4.6}$$

The 4×4 determinant for $D(\kappa)$ has been reduced to 3×3 using the fact that $c_4(k_0-\kappa)=0$. The numerators N_{11} and N_{21} are determinants with $c_1(k_0-\kappa)$, $c_2(k_0-\kappa)$, $c_3(k_0-\kappa)$ in place of column one and column two respectively of the determinant (4.5) for $D(\kappa)$.

The S matrix clearly varies with z_0, i.e., it is dependent on where the interface is placed related to an atomic plane. The denominator $D(0)$ vanishes at $z_0=0.047a$. This gives an example of the resonant case (B2). For values of z_0 where $D(0)$ is not small, expansion in κ yields the coefficient α for $S_{21}=\alpha\kappa$. In Fig. 1, $|\alpha|$ is plotted as a function of z_0, except for a small neighborhood near $z_0=0.047a$ where the resonance phenomenon occurs. The presence of resonance means that the surface states can be formed even for the infinite potential barrier model. In real silicon MOSFET devices, the density of charges trapped by the surface states is much smaller

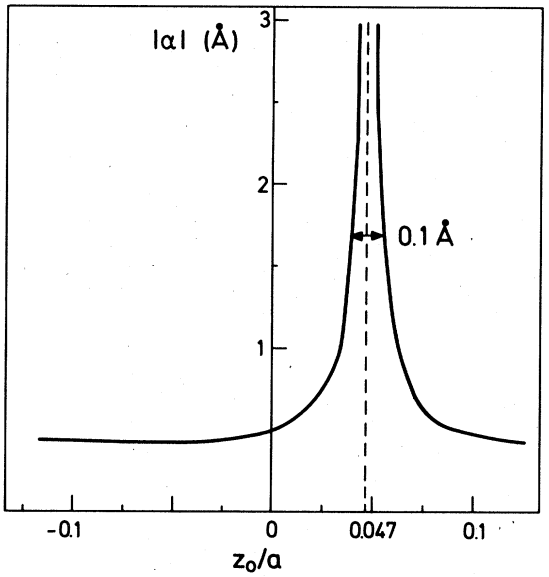

FIG. 1. Intervalley scattering matrix coefficient $|\alpha|$ for p-Si (001) surface as a function of position of the interface from a crystal plane at $z_0=0$ to midway between two crystal planes at $z_0=\pm\frac{1}{8}a$.

than that of the inversion layer. This means that no intrinsic surface states are formed at the interface, thus excluding the possibility of the resonance case. The value of $|\alpha|$, 0.43 Å, for the interface located midway between two planes of silicon atoms ($z_0=\frac{1}{8}a$) is typical of the nonresonant values. Until we have a better model for the interface region, this value of $|\alpha|$ will be taken as the estimate for the Si (001) surface.

D. Convergence problem

By adopting the Cardona-Pollack model[16] for the band structure, it is not feasible to extend the number of basis functions at Γ beyond fifteen. However, we can acquire some measure of convergence by investigating the results of using smaller basis sets.

Take the basis of nine lowest periodic functions at Γ consisting of $2\times\Gamma_1$, Γ_2', Γ_{15}, $\Gamma_{25'}$. Along the $[00k]$ axis, the Hamiltonian is reduced to a 3×3 matrix for Δ_1 states, a 2×2 matrix for $\Delta_{2'}$ states and two 2×2 matrices for the doubly degenerate Δ_5 states. There are three linearly independent functions of $\bar{\Gamma}_1$ symmetry in the $(x$-$y)$ space on a $z=$constant plane, given by the first three functions in Eq. (4.1). At a given energy E just above the conduction-band edge, there are four Bloch waves of symmetry Δ_1 and two evanescent waves of symmetry $\Delta_{2'}$ which are the only solutions compatible with $\bar{\Gamma}_1$. The one decaying wave has a wave vector $i\times 0.11$ a.u., to be compared with

the previous value of $i \times 0.095$ a.u. in the 15-plane-wave basis. The S matrix is given by Eqs. (4.4) and (4.5) with b_j in place of $\tilde{b}_j$. The value of $|\alpha|$ is changed from the 15 basis function value at $z_0 = \frac{1}{8}a$ by a negligible amount to two significant digits in units of Å.

A still simpler $\vec{k}\cdot\vec{p}$ approximation which reproduces the shape of the conduction band along Δ is that with two basis functions, one at Γ_{15} which is in the lowest conduction band and one at the upper Γ_1, just above the Γ_{15} state. The 2×2 $\vec{k}\cdot\vec{p}$ matrix for the Hamiltonian yields four Bloch waves at an energy just above the conduction-band edge. There is no evanescent wave in this model at all. Since only the two outgoing Bloch waves contribute to the right-hand side of Eq. (2.1), the basis set in the x-y plane now contains only two plane-wave combinations of $\bar{\Gamma}_1$ symmetry, i.e., ξ_2 and ξ_3 of Eq. (4.1) which originate from the $\langle 111\rangle$ plane waves. Then, the boundary condition is equivalent to setting the coefficients of the two basis functions Γ_1^{μ}, Γ_{15} to zero at $z = z_0$. We have

$$|\alpha| = [E(\Gamma_1^{\mu}) - E(\Gamma_{15})]/k_0 T^2 = 0.23 \text{ Å} \ , \qquad (4.7)$$

where T is the momentum matrix element between the Γ_1^{μ} state and the Γ_{15} state.[16] Note that the S-matrix element is independent of the position of the interface, contains no contribution from the evanescent waves and is smaller than the preceeding values by almost a factor of two. We conclude that the two-band model does not provide a satisfactory representation of the wave functions at the energy near the conduction-band edge. This two-band model is close to the Ohkawa-Uemura model[7] though not identical to it. The relation has been discussed in detail elsewhere.[18]

E. Valley splitting

With the S matrix in hand, we can calculate the energy splitting of the twofold valley degeneracy. In case (B2), the splitting is of the same order as the intersubband energy difference. However, we have already ruled out case (B2) as being unlikely to occur on Si (001) surface.

For case (B1) by Eq. (3.27) the valley splitting is proportional to α and also to $\langle \partial v/\partial z\rangle$. An estimate of the latter is made for a self-consistent potential

$$\left\langle \frac{\partial v}{\partial z} \right\rangle = \left(\frac{2\pi e^2}{\kappa_s} \right) (N_{\text{inv}} + 2N_{\text{depl}} - 2N_{\text{im}}) \ ,$$

$$\equiv \left(\frac{2\pi e^2}{\kappa_s} \right) N_{\text{eff}} \ , \qquad (4.8)$$

where κ_s is the Si dielectric constant, the terms proportional to the depletion layer density N_{depl}, to the inversion layer density N_{inv}, and to N_{im} are the contributions from the parts of the potential due to the depletion charges, the Hartree potential, and the image potential respectively. The exchange and correlation potential in the local density approximation[13,15] yields no contributions.

As is well known, the image potential, being inversely proportional[19] to z, has an unphysical divergence as $z \to 0$. This divergence becomes stronger still in $\partial v/\partial z$. It should be removed for two reasons. The microscopic reason is that, when a charge gets close to the interface region, the polarization distribution can no longer be neatly represented by a point image charge and thus dampens the $1/z$ dependence. The saturation should commence when z becomes a fraction of the lattice constant. The second and macroscopic reason is that the rapidly varying part of the $1/z$ potential for small z, which might still remain after the first defect is corrected, violates the assumption of a slowly varying effective potential and should be included in the interface potential barrier. The value of the potential in the effective-mass equation represents an average over at least one unit cell of the crystal. One plausible way to avoid the steep part of the image potential is to choose the "effective-mass interface," a notion introduced in Sec. III, a small distance into the region in which the charge resides. A rough way to account for both effects is to take the image potential to be

$$v_{\text{im}}(z) = 4\pi Q e^2/\kappa_s (z + \delta) \qquad (4.9)$$

with

$$Q = (\kappa_s - \kappa_0)/16\pi(\kappa_s + \kappa_0) \ , \qquad (4.10)$$

where κ_0 is the oxide dielectric constant and z is measured from the interface. δ is a positive distance of the order of the lattice constant. However, a definite value for δ deduced from first principles is at present lacking.

The image potential contribution to $\langle \partial v/\partial z\rangle$ is given by

$$N_{\text{im}} = \langle Q/(z+\delta)^2 \rangle \ . \qquad (4.11)$$

To estimate N_{im}, we take the accurate variational form of the wave function due to Takada and Uemura[20]

$$A(z) = (\tfrac{3}{2} b^3)^{1/2} z \exp[-\tfrac{1}{2}(bz)^{3/2}] \ . \qquad (4.12)$$

The variational parameter b is obtained by minimizing the ground-state energy (per particle) of the whole system in the Hartree approximation including the modified image potential

$$E = \tfrac{5}{16}\Gamma(\tfrac{2}{3}) \frac{\hbar^2 b^2}{m_z^*} + \frac{4\pi e^2}{\kappa_s b} \frac{35}{108} 2^{1/3}\Gamma(\tfrac{2}{3}) N$$

$$+ \frac{4\pi e^2 Q}{\kappa_s} b F_1(b\delta) \ , \qquad (4.13)$$

where

$$N = N_{\text{inv}} + \gamma N_{\text{depl}} \ , \tag{4.14}$$

$$\gamma = \tfrac{24}{7} x 2^{1/3} = 2.72 \ , \tag{4.15}$$

and

$$F_n(\beta) = \int_0^\infty dt\, t(t^{-2/3} + \beta)^{-n} e^{-t} \ . \tag{4.16}$$

Thus, b is a function of N_{eff} and so is N_{im}, given by

$$N_{\text{im}} = Qb^2 F_2(b\delta) \ , \tag{4.17}$$

with F_2 defined by Eq. (4.16).

Equation (4.8) can be rewritten

$$\langle \frac{\partial v}{\partial z} \rangle = (0.77 \text{ meV/Å})[N - 2N_{\text{im}} + (2-\gamma)N_{\text{depl}}] \ , \tag{4.18}$$

with all the densities N in units of 10^{12} cm^{-2}. The quantities b, N_{im}, and $N - 2N_{\text{im}}$ are evaluated for a range of values of δ as a function of N, defined in Eq. (4.14). In Fig. 2, $N - 2N_{\text{im}}$, which is a measure of $\langle \partial v/\partial z \rangle$ when N_{depl} is negligible, versus N is plotted. $\delta = 0$ corresponds to the usual usage of the image potential. $\delta = \infty$ corresponds to dropping the image potential term completely. The modification of the image potential by a finite δ of the order of a affects b and hence the unsplit energy levels rather weakly but changes $\langle \partial v/\partial z \rangle$ drastically. An example at $N = 2.4 \times 10^{12}$ cm^{-2} is shown in Table I. The valley splitting depends quite sensitively on the value of δ.

A recent measurement of the valley splitting of the lowest subband by Köhler, Roos, and Landwehr[21] yields $\Delta E = 0.75$ meV at $N_{\text{inv}} = 2.4 \times 10^{12}$ cm^{-2}. N_{depl} appears to be small. This splitting is deduced from the phase change of the Shubnikov–de Haas oscillations in a tilted magnetic field. Table I shows that while the full image potential gives too low a value for ΔE; shifting the potential through a distance of $\frac{1}{4}a$ or $\frac{1}{2}a$ yields a range of values in reasonable agreement with experiment.

An estimate of the valley splitting can also be obtained from the cusp structure in magnetoconductivity as a function of N_{inv} in a high magnetic field observed by Kawaji[22] and co-workers. At a field of 14 Tesla, the cusp occurs at the third Landau level (counting the lowest level as the zeroth) which gives $N_{\text{inv}} = 4.74 \times 10^{12}$ cm^{-2}. In a fixed field at low densities each Landau level consists of four lines with the spin splitting larger than the valley splitting. The valley splitting increases with the density until it equals the spin splitting. Then two states with opposite spins and different valley combinations coincide in energy and create a large peak bracketed by two smaller peaks on either side, instead of the four peak structure at low N_{inv}. From the spin splitting using a g factor of 2, we can get $\Delta E = 1.62$ meV. With a g factor of 2.5 at this density, according to Fang and Stiles,[23] $\Delta E \simeq 2.03$ meV. The value of 1.62 meV is consistent with that of Köhler *et al.* Our calculated values of 1.23–1.31 meV for $\delta = \frac{1}{4}a - \frac{1}{2}a$ (approximately double the values in Table I) are again in about the same sort of agreement as before with experiment. In the cusp structure, the density of states at the Fermi level is high and, therefore, the many-

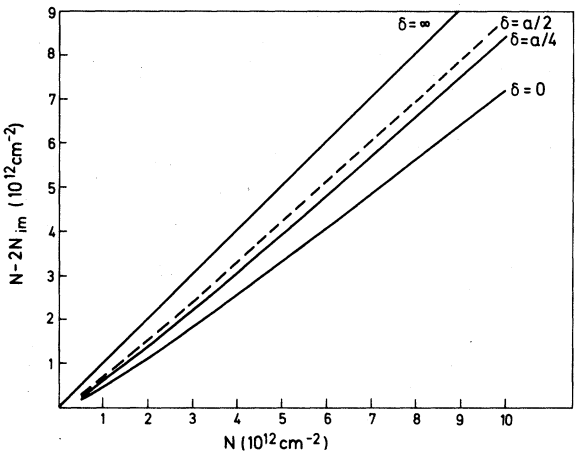

FIG. 2. Density dependence of the average force $\langle \partial v/\partial z \rangle$. The plot is actually $N - 2N_{\text{im}}$ vs N for various values of δ which modifies the image potential. The symbols are defined in Eqs. (4.18) and (4.14).

TABLE I. Effect of the image potential on valley splitting. $N_{\text{depl}} = 0$, $N_{\text{inv}} = 2.4 \times 10^{12}$ cm^{-2}, $|\alpha| = 0.43$ Å.

δ	b(10^{-2} Å)	N_{im}(10^{12} cm^{-2})	$0.77(N - 2N_{\text{im}})$ (meV/Å)	ΔE (meV)
0	5.99	0.51	1.08	0.47
$\frac{1}{4}a$	6.09	0.35	1.32	0.57
$\frac{1}{2}a$	6.16	0.28	1.43	0.62
∞	6.60	0	1.86	0.81

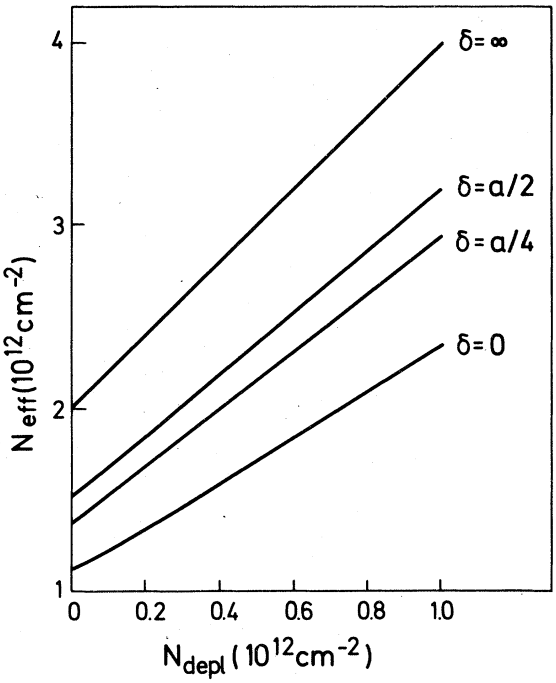

FIG. 3. Dependence of $\langle \partial v/\partial z \rangle$ on the depletion charge density at $N_{\text{inv}} = 2 \times 20^{12}$ cm^{-2}. N_{eff} is defined in Eq. (4.8). δ is defined in Eq. (4.9), modifying the image potential.

body enhancement is expected to be relatively weak.[7] The broadening of the Landau peaks by defects makes it possible for the cusp structure to occur even if the two central levels do not quite coincide. Thus, the estimated valley splitting can err on either side.

Ohkawa[24] has interpreted the cusp structure as due to the valley splitting equal to half the spin splitting and obtained for $g = 2$, $\Delta E = 0.81$ meV, in agreement with the smaller value of the Ohkawa-Uemura theory.[7] We believe that when the valley splitting is equal to half the spin splitting, one should expect a structure with four peaks and not a large central peak flanked by two smaller ones as was observed.[22]. That their theoretical value for valley splitting is almost half as large as ours is due to the use of the two-band model as shown in Sec. IV D.

To demonstrate the possible effects of substrate bias on the valley splitting, we plot in Fig. 3, for $N_{\text{inv}} = 2 \times 10^{12}$ cm^{-2}, the dependence of N_{eff}, which is proportional to ΔE as given by Eq. (4.8), on the depletion charge density for a range of values of δ.

F. E_0' state

As a purely academic exercise, we illustrate the application of our general theory to the case of a single band minimum, i.e., cases (A1) and (A2). Consider the eigenstate in the inversion layer composed of the Bloch states near the $(k_0, 0, 0)$ valley of Si with (001) surface bordering the oxide. For a given $\vec{k}_{\parallel}$ near this valley, the number of propagating states is 2, corresponding to case (A1) or (A2). For this valley, the mass in the normal direction to the interface is much lighter than that for the $(0, 0, k_0)$ valley. Hence, the lowest energy level is higher than that for the $(0, 0, k_0)$ valley and is termed the E_0' level.[2,14]

As a simple model of the Si bulk band struture near the $(k_0, 0, 0)$ valley, we use instead of the 15 basis functions at Γ, the four at $(k_0, 0, 0)$ closest to the band gap with wave functions approximately given by

$$\begin{aligned} \psi_{\Delta_1} &= \chi_1 e^{i(k_0 - G)x}, \quad G = 2\pi/a, \\ \psi_{\Delta_{2'}} &= \bar{\chi}_1 e^{i(k_0 - G)x}, \\ \psi_{\Delta_{5y}} &= \chi_4 e^{i(k_0 - G)x}, \\ \psi_{\Delta_{5z}} &= \bar{\chi}_4 e^{i(k_0 - G)x}, \end{aligned} \tag{4.19}$$

where χ_1, $\bar{\chi}_1$ are the Bloch functions of the lowest conduction band at $(2\pi/a, 0, 0)$ and χ_4, $\bar{\chi}_4$ are the ones of the highest doubly degenerate valence band at $(2\pi/a, 0, 0)$. All four functions at the X point are given in symmetrized combination of $\langle 011 \rangle$ plane waves. The Δ_1 and $\Delta_{2'}$ states are the lowest and next lowest conduction-band states. The Δ_{5y}, Δ_{5z} states are two degenerate states of the highest valence band. At an energy

$$E = E(\Delta_1) + \hbar^2 \kappa^2 / 2 m_t \tag{4.20}$$

slightly above $E(\Delta_1)$ and below $E(\Delta_{2'})$, the two Bloch wave functions are given by

$$\psi(k_0, 0, \pm\kappa) = \psi_{\Delta_1} \pm d_5 \kappa \psi_{\Delta_{5z}} . \tag{4.21}$$

The mixing with the top valence state Δ_{5z} alone is sufficient to account for the magnitude of the transverse mass m_t. The two evanescent waves are formed by the $\Delta_{2'}$ and Δ_{5y} states

$$\chi = B_{2'} \psi_{\Delta 2'} + B_5 \psi_{\Delta 5} . \tag{4.22}$$

At the interface plane $z = z_0$, all these functions are linear combinations of $\sin(2\pi y/a)$ and $\cos(2\pi y/a)$. By the procedure of Appendix A, we obtain the scattering length

$$a_s' = \frac{d_5 [B_5 \sin(4\pi/a) z_0 + i B_2' \cos(4\pi/a) z_0]}{B_5 \cos(4\pi/a) z_0 - i B_2' \sin(4\pi/a) z_0} . \tag{4.23}$$

The resonance occurs at $z_0 = z_r$ where

$$\tan \frac{4\pi z_r}{a} = \frac{B_5}{i B_{2'}} . \tag{4.24}$$

The coefficients d_5, B_5, $B_{2'}$ are evaluated by using the momentum matrix elements $\langle\Delta_1|p_z|\Delta_{5z}\rangle = 1.139$, $\langle\Delta_{2'}|p_s|\Delta_{5y}\rangle = 1.1332$ and the energy differences $E(\Delta_1) - E(\Delta_5) = 0.300$, $E(\Delta_1) - E(\Delta_{2'}) = -0.054$, all in atomic units calculated from the 15 basis $\vec{k}\cdot\vec{p}$ model. Then, the resonance occurs at $z_r/a = -0.028$.

The scattering length $a_{s'}$ for the $(k_0, 0, 0)$ valley is different from the intravalley scattering length a_s for the $(0, 0, \pm k_0)$ valleys. The difference contributes a term $(a_s')\langle\partial v/\partial z\rangle' - a_s\langle\partial v/\partial z\rangle$ to the energy difference $E_{0'} - E_0$ between the lowest states of the two types of valleys. Here v is the self-consistent potential in the presence of the electrons occupying the lowest subband of the $(0, 0, \pm k_0)$ valleys, and the averages are taken over the wave functions of the lowest states from the $(k_0, 0, 0)$ and $(0, 0, k_0)$ valleys respectively. While a_s is insensitive to the position of the interface except in the neighborhood where the resonance occurs, a_s' varies rather strongly with the position of the interface because of the large value of d_5 (2.0 Å). This simple calculation suggests that E_0' is more sensitive to the actual interface potential than E_0. A recent calculation by Stern[25] based on a graded surface EMA model shows a strong dependence of $E_0' - E_0$ on the width of the interface region Our result, arising out of the different nature of the Bloch waves involving in the surface reflection, has a different physical origin from Stern's.

V. DISCUSSION

We have given the general procedure of accounting for the interface effect outside the EMA but treating any spatially slowly varying external potential within EMA. Clearly such a procedure has many possible applications besides that of the inversion layer problem which we have treated. Even within the space-charge layer in MOSFET, we have yet to use this method to study other surfaces of Si and other semiconductors.

Our calculation with p-Si (001) surface shows that the interface has profound effects on the electron dynamics in the inversion layer. Because in our calculation, an idealized model of the interface is used, it does not do justice to the full range of possibilities of the general theory on how the nature of the interface affects the electron motion. By varying substrate bias or parallel component of the magnetic field, the average distance of the electron from the interface can be changed and therefore, there is the possibility of the variation of the interface effect on the electron. We should, therefore, study a more realistic interface potential. It would be very interesting if it should prove possible to invert the process, to use the electron properties to infer the nature of the interface.

As mentioned in Sec. III, we have neglected the dependence of the surface reflection on the transverse motion of the electron. A more careful inclusion of the transverse motion is desirable to study the possible deviation from the EMA.

Our use of the modified image potential in Sec. IV, though serves to demonstrate the necessity of rounding off the $1/z$ dependences, is entirely phenomenological. A microscopic theory for the image potential in the interface region and a more careful derivation of its expression in the effective-mass equation are clearly desirable.

ACKNOWLEDGMENT

L. J. Sham would like to thank Professor W. Hanke and Professor H. Bilz for the hospitality at Max-Planck-Institut für Festkörperforschung, Stuttgart and the Humboldt Foundation for an award. M. Nakayama would like to thank Professor J. J. Quinn for the hospitality at Brown University and the Ministry of Education of Japan for an Overseas Research Fellowship. This work was supported in part by the NSF Grant No. DMR 77-09595 and the Material Research Program at Brown University funded through the NSF.

APPENDIX A: S MATRIX NEAR A SINGLE BAND MINIMUM

This Appendix contains the details leading to the expression (2.4). Let $\{\xi_m\}$ be a complete set of orthornormal functions of coordinates x and y on the interface only. The boundary values at the interface ($z = 0$) of the Bloch waves and the evanescent waves in both media can be expanded in terms of this basis. Let $c_m(k_0 + \kappa)$ denote the coefficient of expansion for the Bloch wave $\psi_{k_0+\kappa}$, $b_{m\lambda}$ and $\bar{b}_{m\mu}$ the coefficients for the evanescent waves χ_λ and $\bar{\chi}_\mu$ respectively. The continuity of the wave function at the interface then yields the relation

$$c_m(k_0 - \kappa) = c_m(k_0 + \kappa)S + \sum_\lambda b_{m\lambda}T_\lambda + \sum_\mu \bar{b}_{m\mu}R_\mu \ . \tag{A1}$$

In the same way, the continuity of the z derivative of the wave function leads to the boundary condition

$$d_m(k_0 - \kappa) = d_m(k_0 + \kappa)S + \sum_\lambda e_{m\lambda}T_\lambda + \sum_\mu \bar{e}_{m\mu}R_\mu \ , \tag{A2}$$

where $d_m(\kappa)$, $e_{\lambda m}$, $\bar{e}_{\lambda m}$ are, respectively, the expansion coefficients in terms of the basis set $\{\xi_m\}$ of the derivatives of the Bloch wave $\partial\psi_{k_0+\kappa}/\partial z$ and the evanescent waves $\partial\chi_\lambda/\partial z$, $\partial\chi_\mu/\partial z$ evaluated at the interface $z = 0$.

Solution of the linear equations (A1) and (A2) yields the S-matrix expression given by Eq. (2.4), with

$$D(\kappa)=\det\begin{pmatrix} c(k_0+\kappa)\,b & \bar{b} \\ d(k_0+\kappa)\,e & \bar{e}\end{pmatrix} , \tag{A3}$$

where $c(k_0+\kappa)$ stands for a column vector with elements $c_m(\kappa)$ and b stands for a matrix with elements $b_{m\lambda}$.

For the special case of $k_0=0$ and $\vec{k}_{\parallel}=0$, the phase of the Bloch function can be chosen such that

$$\psi_{+k_0-\kappa}=\psi^*_{k_0+\kappa} \quad . \tag{A4}$$

The evanescent waves can be chosen to be real. Thus,

$$c_m(-\kappa)=c_m^*(\kappa) \ , \quad d_m(-\kappa)=d_m^*(\kappa) \tag{A5}$$

and $b_{m\lambda}$, $\bar{b}_{m\lambda}$, $e_{m\lambda}$, $\bar{e}_{m\lambda}$ are real and functions of κ^2. Hence,

$$D(-\kappa)=D^*(\kappa) \quad . \tag{A6}$$

APPENDIX B: S MATRIX FOR DOUBLE BAND MINIMA

The procedure parallels that in Appendix A. The only difference is that there are now two incoming Bloch waves at $k_0-\kappa$ and two outgoing ones at $k_0+\kappa$. The S matrix is of the form Eq. (2.14) with

$$D(\kappa)=\det\begin{pmatrix} c(k_0+\kappa) & c(-k_0+\kappa) & b & \bar{b} \\ d(k_0+\kappa) & d(-k_0+\kappa) & e & \bar{e}\end{pmatrix} . \tag{B1}$$

$N_{11}(\kappa)$ and $N_{21}(\kappa)$ are given by the determinants obtained by substituting $c(k_0-\kappa)$, $d(k_0-\kappa)$ as one column in the first and second column respectively of that for $D(\kappa)$. $N_{12}(\kappa)$ and $N_{22}(\kappa)$ are given by the determinants obtained by substituting $c(-k_0-\kappa)$, $d(-k_0-\kappa)$ as one column in the first and second column respectively of that for $D(\kappa)$.

From these forms for $N_{ij}(\kappa)$ and $D(\kappa)$, and time reversal symmetry it is straightforward to show that

$$S_{11}(\kappa)=S_{22}^*(-\kappa) \ , \quad S_{12}(\kappa)=S_{21}^*(\kappa) \quad . \tag{B2}$$

A small κ expansion then leads to the form (2.15).

The wave function ϕ_i represents a beam of particles impinging on the interface and two reflected beams. Conservation of particle flux[12] leads to the relations

$$\begin{aligned} &v(1-)+v(1+)|S_{11}|^2+v(2+)|S_{21}|^2=0 \ , \\ &v(2-)+v(1+)|S_{12}|^2+v(2+)|S_{22}|^2=0 \ , \end{aligned} \tag{B3}$$

where $v(1\pm)$ denote the z components of the group velocities of the Bloch waves at $k_1^{(\pm)}$. The properties (B2) are consistent with these conditions.

To show that the real part of β, β_1, from Eq. (2.15) is related to the deviation of the energy dispersion from the parabolic relation, let the energy around valley 1 be

$$E=E_0+(\hbar^2/2m_z^*)[(k_z-k_0)^2+2\gamma(k_z-k_0)^3] \quad . \tag{B4}$$

With the definition of κ given by Eq. (2.13), Eq. (2.12) has to be modified to

$$\begin{aligned} k_1^{(\pm)}&=k_0\pm\kappa-\gamma\kappa^2 \ , \\ k_2^{(\pm)}&=-k_0\pm\kappa-\gamma\kappa^2 \ . \end{aligned} \tag{B5}$$

The velocities are

$$\begin{aligned} v(1+)&=-v(2-)=(h/m_z^*)(\kappa+2\gamma\kappa^2) \ , \\ v(1-)&=-v(2+)=(h/m_z^*)(-\kappa+2\gamma\kappa^2) \ . \end{aligned} \tag{B6}$$

Substituting these expressions for the velocities and Eq. (2.15) for the S matrix into the flux conservation conditions (B3), we obtain

$$\beta_1=2\gamma \ . \tag{B7}$$

This relation is independent of the detail of the surface barrier.

The set of wave functions ϕ_i are solutions of the Schrödinger equation with the same boundary conditions and therefore are orthogonal to each other. We show here explicitly that for the asymptotic forms Eq. (2.16) with small κ approximation the orthogonality still holds to $O(\kappa)$. The contribution of the evanescent waves to the overlap integral is negligible because of their spatial confinement. The contribution from the overlap between intervalley Bloch waves are negligible. The overlap between intravalley Bloch waves is evaluated to the first order in κ. Then,

$$\begin{aligned} \langle\phi_{1\kappa}|\phi_{1\kappa'}\rangle&=2\pi(1-\beta_1\kappa)\delta(\kappa-\kappa') \\ &\quad+(\beta_1-2\gamma)4i\kappa\kappa'/(\kappa^2-\kappa'^2) \ , \\ \langle\phi_{2\kappa}|\phi_{2\kappa'}\rangle&=2\pi(1+\beta_1\kappa)\delta(\kappa-\kappa') \\ &\quad-(\beta_1-2\gamma)4i\kappa\kappa'/(\kappa^2-\kappa'^2) \ , \\ \langle\phi_{1\kappa}|\phi_{2\kappa'}\rangle&=0 \ . \end{aligned} \tag{B8}$$

The orthogonality follows from Eq. (B7), a consequence of flux conservation.

APPENDIX C: REDUCTION OF EMA TO BULK LIMIT

Consider the external potential $v(z-\lambda)$, centered at $z=\lambda$, vanishing outside the range $|z-\lambda|<l$. We wish to show here that, when $\lambda \gg l$, the effective-mass equations of Sec. III reduces to the bulk form.

Case (B1) is simple. It is instructive to see how the valley splitting disappears as we increase λ from zero to a larger value than l. When $\lambda >> l$, the unperturbed envelope function $A(z)$ for each valley without coupling will decay exponentially as $z \to 0$ from $z = \lambda - l$. By means of the effective-mass equation, the average field can be expressed as

$$\langle \frac{\partial v}{\partial z} \rangle = \frac{\hbar^2}{2m^*}[A'(0)]^2 \ , \qquad \text{(C1)}$$

where $A'(0)$ is the value of the z derivative of the envelope wave function at $z = 0$. Since $A'(0)$ diminishes exponentially as λ becomes much larger than l, the valley splitting vanishes in the bulk limit.

In case (B2) again the wave function $A(z)$ decays exponentially from $z = \lambda - l$ to $z = 0$ as λ becomes much larger than l. Then, the difference between making the envelope function vanishing at $z = 0$ or having zero slope there becomes unimportant. The even- and odd-parity states (under reflection in the plane $z = 0$) become degenerate. $A_{\pm}(z)$ become the same and Eq. (3.33) can be rewritten as two wave functions $A(z)\psi_{\pm k_0}$ which are the usual bulk expression.

---

*Permanent Address: College of Education, Kyusyu University, Ropponmatsu, Fukuoka 810, Japan.

[1]J. R. Schrieffer, in *Semiconductor Surface Physics,* edited by R. H. Kingston (University of Pennsylvania, Philadelphia, 1955), p. 55.
[2]F. Stern and W. E. Howard, Phys. Rev. 163, 816 (1967).
[3]W. Kohn, Solid State Phys. 5, 278 (1957).
[4]Brief preliminary account of our work was presented at the Second International Conference of the Electronic Properties of Two-Dimensional Systems, Surf. Sci. 73, 272 (1978).
[5]K. Narita and E. Yamada, *Proceedings of the Twelfth International Conference on Physics of Semiconductors, Stuttgart, 1974,* edited by M. H. Pilkohn (Teubner, Stuttgart, 1974), p. 719.
[6]R. Kümmel, Z. Phys. B 22, 223 (1975).
[7]F. J. Ohkawa and Y. Uemura, J. Phys. Soc, Jpn. 43, 907 (1977); 43, 917 (1977); 43, 925 (1977).
[8]J. A. Appelbaum and D. R. Hamann, Rev. Mod. Phys. 48, 479 (1976).
[9]A. Goetzberger and M. Schulz, in *Festkörperprobleme XIII,* edited by H. J. Queisser (Pergamon, Vieweg, 1973), p. 312.
[10]A. W. Maue, Z. Phys. 94, 717 (1935).
[11]J. B. Pendry, J. Phys. C 2, 2273 (1969).
[12]J. A. Appelbaum and E. I. Blount, Phys. Rev. B 8, 483 (1973), Appendix.
[13]W. Kohn and L. J. Sham, Phys. Rev. 140, 1133 (1965).
[14]F. Stern, Phys. Rev. B 5, 4891 (1972).
[15]T. Ando, Phys. Rev. B 13, 3468 (1976).
[16]M. Cardona and F. H. Pollak, Phys. Rev. 142, 530 (1966).
[17]R. O. Jones, Proc. Phys. Soc. London 89, 443 (1966).
[18]M. Nakayama and L. J. Sham, Solid State Comm. 28, 393 (1978).
[19]M. W. Cole, Phys. Rev. B 2, 4239 (1970).
[20]Y. Takada and Y. Uemura, J. Phys. Soc. Jpn. 43, 139 (1977).
[21]H. Kohler, M. Ross, and G. Landwehr, Solid State Comm. (to be published).
[22]S. Kawaji, Surf. Sci. 73, 46 (1978).
[23]F. F. Fang and P. J. Stiles, Phys. Rev. 174, 823 (1968).
[24]F. J. Ohkawa, Solid State Comm. 26, 69 (1978).
[25]F. Stern, Solid State Comm. 21, 163 (1977).