Properties of Semiconductor Surface Inversion Layers in the Electric Quantum Limit*

FRANK STERN AND W. E. HOWARD
IBM Watson Research Center, Yorktown Heights, New York
(Received 10 July 1967)

The strong surface electric field associated with a semiconductor inversion layer quantizes the motion normal to the surface. The bulk energy bands split into electric sub-bands near the surface, each of which is a two-dimensional continuum associated with one of the quantized levels. We treat the electric quantum limit, in which only the lowest electric sub-band is occupied. Within the effective-mass approximation, we have generalized the energy-level calculation to include arbitrary orientations of (1) the constant-energy ellipsoids in the bulk, (2) the surface or interface, and (3) an external magnetic field. The potential associated with a charged center located an arbitrary distance from the surface is calculated, taking into account screening by carriers in the inversion layer. The bound states in the inversion layer due to attractive Coulomb centers are calculated for a model potential which assumes the inversion layer to have zero thickness. The Born approximation is compared with a phase-shift calculation of the scattering cross section, and is found to be reasonably good for the range of carrier concentrations encountered in InAs surfaces. The low-temperature mobility associated with screened Coulomb scattering by known charges at the surface and in the semiconductor depletion layer is calculated for InAs and for Si (100) surfaces in the Born approximation, using a potential that takes the inversion-layer charge distribution into account. The InAs results are in good agreement with experiment. In Si, but not in InAs, freeze-out of carriers into inversion-layer bound states is expected at low temperatures and low inversion-layer charge densities, and the predicted behavior is in qualitative agreement with experiment. An Appendix gives the phase-shift method for two-dimensional scattering and the exact cross section for scattering by an unscreened Coulomb potential.

1. INTRODUCTION

An n-type inversion layer is produced at the surface of a p-type semiconductor when the energy bands near the surface are bent down enough that the bottom of the conduction band lies near or below the Fermi level. This band bending can be introduced by applying an electric field to the surface, in a configuration like that shown in Fig. 1, or by the presence of positive charges at or near the surface associated with impurity ions or other Coulomb centers.

The electric field associated with an inversion layer is strong enough to produce a potential well whose width in the z direction, the direction perpendicular to the surface, is small compared to the wavelengths of the carriers. Thus the energy levels of the electrons are grouped in what we call electric sub-bands, each of which corresponds to a quantized level for motion in the z direction, with a continuum for motion in the plane parallel to the surface.

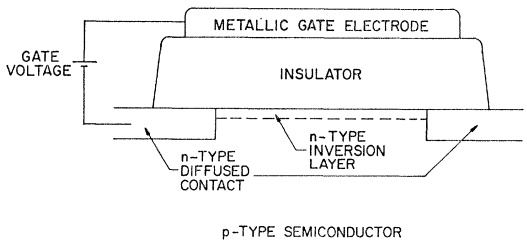

FIG. 1. Schematic metal–insulator–semiconductor structure used for inversion-layer experiments. The inversion-layer electron concentration is changed by changing the gate voltage. The inversion-layer conductance is measured by applying a small potential difference between the n-type contacts and measuring the resulting current.

The quantization of energy levels in inversion layers has been anticipated for many years,[1] but the two-dimensional nature of the electron gas when only one electric sub-band is occupied was only recently confirmed by experiments on n-type inversion layers on a (100) surface of silicon in the presence of a magnetic field perpendicular to the surface.[2]

The principal purpose of this paper is to study the effect of charged centers near the surface on the properties of electrons in the inversion layer. To do this we first find the average potential due to such a charge seen by the inversion layer electrons, including for the first time the effect of screening in this quasi-two-dimensional system. From this potential we find both the scattering cross section of the electrons and the inversion-layer bound states that result when the potential is attractive.

The calculated scattering rate is compared with experimental results on inversion-layer mobility at low temperatures in InAs[3] and in Si.[4] These experiments measure the conductance between the n-type contacts of Fig. 1, and determine the inversion-layer electron mobility by using the Hall effect[4] or magnetoresistance.[3] Thus an experimental curve of inversion-

* Some of the results of this work were presented at the Chicago meeting of the American Physical Society, March, 1967; Bull. Am. Phys. Soc. 12, 275 (1967).

[1] J. R. Schrieffer, in *Semiconductor Surface Physics*, edited by R. H. Kingston (University of Pennsylvania Press, Philadelphia, Pennsylvania, 1957), p. 55.

[2] A. B. Fowler, F. F. Fang, W. E. Howard, and P. J. Stiles, Phys. Rev. Letters 16, 901 (1966); in *Proceedings of the International Conference on the Physics of Semiconductors, Kyoto, 1966*, [J. Phys. Soc. Japan Suppl. 21, 331 (1966)].

[3] S. Kawaji and Y. Kawaguchi, in *Proceedings of the International Conference on the Physics of Semiconductors, Kyoto, 1966*, [J. Phys. Soc. Japan Suppl. 21, 336 (1966)].

[4] F. F. Fang and A. B. Fowler (to be published).

layer mobility versus inversion-layer free-carrier concentration can be constructed.

Because the inversion layer is thin—of the order of 100 Å or less—the mobility is sensitive to scattering associated with the surface, and thus provides a powerful tool for studying surface-scattering mechanisms.

If the Coulomb centers near the surface are positively charged, they attract electrons, and in general lead to bound states. The properties of these bound states depend strongly on the amount of screening by inversion-layer electrons, and are among the most interesting results of our work. In particular, we find that in Si inversion layers the bound states in the absence of screening are deep enough to trap the first electrons which enter the inversion layer at low temperatures. As the gate voltage is increased and more electrons are added, some will enter the electric sub-band and will contribute to screening, thus weakening the attractive potential. At a sufficiently high inversion-layer electron concentration, the screening will reach its full value, and the energy levels will be so shallow that the orbits of inversion-layer electrons bound to adjacent Coulomb centers overlap. Under these conditions the bound states merge with the bottom of the lowest electric sub-band and effectively disappear. In InAs, on the other hand, the bound states merge with the bottom of the lowest electric sub-band even in the absence of screening, and effects associated with bound states are not expected.

Except for the calculation of Landau levels in Appendix A, all of our results apply to zero magnetic field. In addition we restrict ourselves to low temperatures, for which only the lowest electric sub-band is occupied by electrons. We call this limiting case the electric quantum limit.

The organization of the remainder of the paper is summarized by the following Table of Contents:

2. ELECTRIC SUB-BANDS

In this section we consider the energy levels of inversion-layer electrons moving in a potential well which depends only on z, the distance from the surface. We find the effective masses for motion parallel to the surface and the corresponding one-dimensional Schrödinger equation for motion perpendicular to the surface for an arbitrary surface orientation and for arbitrary bulk ellipsoidal constant-energy surfaces.

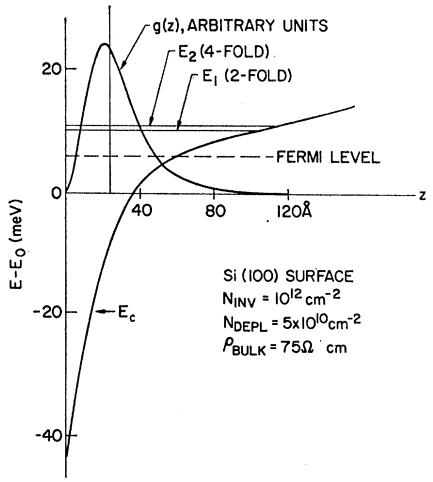

FIG. 2. Surface potential well and surface charge distribution for a representative Si surface. The curve labeled E_c gives the energy of the bottom of the conduction band near the semiconductor-insulator interface at $z=0$. Also shown are the position of the Fermi level for 10^{12} electrons/cm^2 in the inversion layer, and the positions of the first two excited states, as calculated by Howard (Ref. 10). The zero of energy is taken to be the bottom of the lowest electric sub-band. The upper curve is the charge distribution of carriers in the lowest electric sub-band, taken from Ref. 10. Also shown is a vertical line at the average inversion-layer thickness as computed from the approximate relations given in Eqs. (22) and (42).

The electron potential well near the semiconductor-insulator interface is shown for a typical case in Fig. 2. We wish to find the energy levels E and envelope functions ψ belonging to self-consistent solutions of the effective-mass equation

$$[T-e\phi(z)-E]\psi=0, \tag{1}$$

where T is the kinetic-energy operator and ϕ is the electrostatic potential, which in turn is the solution of Poisson's equation

$$\nabla^2\phi=-4\pi\rho/\kappa. \tag{2}$$

Here ρ is the charge density given by the fixed charge in the depletion layer plus the charge in the inversion-layer states, and κ is the dielectric constant. The boundary conditions on ϕ are

$$\phi\to 0 \quad \text{as} \quad z\to\infty, \tag{3a}$$

$$\kappa_{\text{ins}}(d\phi/dz)\big|_{z=0^-}=\kappa_{\text{sc}}(d\phi/dz)\big|_{z=0^+}, \tag{3b}$$

where κ_{sc} and κ_{ins} are the static dielectric constants of the semiconductor and the insulator, respectively.

The effective-mass equation (1) is perhaps a poorer approximation in the inversion-layer problem than in some other applications, since the width (in the z direction) of the inversion-layer envelope function ψ is only of the order of 20 Å in some cases, and therefore not much larger than atomic dimensions. The boundary

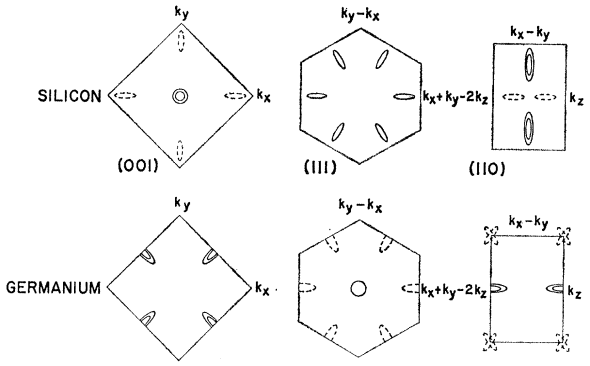

FIG. 3. Schematic representation of the constant-energy ellipses and Brillouin zones for the (001), (111), and (110) surfaces of Si and Ge in the effective-mass approximation. Where both solid and dashed curves appear, the solid curves correspond to the electric sub-band of lower energy. Concentric ellipses are shown to indicate doubly degenerate levels. Note that the dashed curves for the (110)-Si surface have been displaced by a reciprocal lattice vector to bring them within the first Brillouin zone. See also Table I. The horizontal and vertical axes have been marked to indicate their directions in reciprocal space. The radii of the circumscribed circles about the three Brillouin zones are $2\pi/a$, $(4\pi/3a)\sqrt{2}$, and $(\pi/a)\sqrt{3}$ for (001), (111), and (110), respectively, where a is the lattice constant in real space.

condition we use for ψ is

$$\psi(0)=0, \tag{4}$$

since the potential barrier at the semiconductor-insulator interface will lead to a small amplitude for the envelope function there.

The kinetic-energy operator in Eq. (1) can be written

$$T=\tfrac{1}{2}\sum_{i,j}w_{ij}p_ip_j, \tag{5}$$

where $p_j=-ih(\partial/\partial x_j)$, and w_{ij} is the reciprocal effective-mass tensor in a coordinate system in which the semiconductor-insulator interface is the plane $z=0$. In terms of the transformation matrix $\|a\|$ from the principal axes of a constant-energy ellipsoid of the semiconductor (indicated by primes) to the coordinate system we have chosen, we have

$$p_j=\sum_k a_{jk}p_k', \tag{6}$$

$$w_{ij}=\sum_k a_{ik}a_{jk}w_{kk}'=w_{ji}, \tag{7}$$

where $w_{kk}'=1/m_k'$, and the m_k' are the principal effective masses of the ellipsoid of constant energy in the semiconductor.

Since the potential energy $-e\phi$ in (1) is a function of z only, we take as a trial solution the product

$$\psi(x,y,z)=\xi(z)\exp(ik_1x+ik_2y). \tag{8}$$

If we substitute this in (1), using (5), we find that ξ must satisfy the equation

$$-\tfrac{1}{2}w_{33}\hbar^2(d^2\xi/dz^2)-\hbar^2(w_{13}k_1+w_{23}k_2)(d\xi/dz) -[e\phi(z)+E']\xi(z)=0, \tag{9}$$

where

$$E=E'+\tfrac{1}{2}\hbar^2(w_{11}k_1^2+2w_{12}k_1k_2+w_{22}k_2^2). \tag{10}$$

We now make the substitution

$$\xi(z)=\zeta(z)\exp[-iz(w_{13}k_1+w_{23}k_2)/w_{33}] \tag{11}$$

to eliminate the first derivative with respect to z, and find that $\zeta(z)$ must satisfy

$$d^2\zeta_i/dz^2+(2m_3/\hbar^2)[E_i''+e\phi(z)]\zeta_i(z)=0, \tag{12}$$

where $m_3=w_{33}^{-1}$, a subscript i has been introduced to label the solutions, and

$$E_i(k_1,k_2)=E_i''+\tfrac{1}{2}\hbar^2\Big[\Big(w_{11}-\frac{w_{13}^2}{w_{33}}\Big)k_1^2 +2\Big(w_{12}-\frac{w_{13}w_{23}}{w_{33}}\Big)k_1k_2+\Big(w_{22}-\frac{w_{23}^2}{w_{33}}\Big)k_2^2\Big]. \tag{13}$$

The Schrödinger equation thus reduces to an equation in z, in which the x and y motion enters in the energy only through (13), and through the boundary condition on ψ. Since we use the boundary condition (4), the eigenvalue E'' is independent of k_1 and k_2. If other boundary values at $z=0$ were used, both E_i'' and $\zeta_i(z)$ could depend on k_1 and k_2 unless $w_{13}=w_{23}=0$.

Because of the form of the kinetic-energy operator (5), our results do not apply to inversion-layer energy levels arising from a degenerate band like the valence band of Ge or Si or InAs. Furthermore, the strong electric field associated with the potential ϕ is expected to remove the degeneracy of the basis Bloch functions, as does a uniaxial strain,[5] thereby invalidating the simple effective-mass approach. We shall not deal with that case in this paper.

Three effective masses enter in the energy levels of electrons in the inversion layer. One is $m_3=w_{33}^{-1}$, the mass which determines the energy levels E_i'' for motion perpendicular to the surface via Eq. (12). The other two, m_1 and m_2, are the principal effective masses of the constant-energy ellipses associated with motion parallel to the surface, and can be deduced from Eq. (13). For a simple conduction band like that of InAs, all three of the effective masses are equal to each other and to the effective mass in the bulk, apart from the effect of nonparabolicity, which we have not considered.

It is easy to show from (13) that the product $m_1m_2m_3$ equals the product $m_1'm_2'm_3'$ of the principal effective masses in the bulk for an arbitrary ellipsoidal constant-energy surface and for arbitrary surface orientation. This assures that, as in the case of an isotropic mass,[1,6] the density of states when many electric sub-bands are occupied approaches the value one would find without considering surface quantization.

For conduction bands with multiple minima, like those of Ge and Si, the plane in **k** space containing the

[5] G. E. Pikus and G. L. Bir, Fiz. Tver. Tela **1**, 1642 (1959) [English transl.: Soviet Phys.—Solid State **1**, 1502 (1960)]. We thank J. J. Hall for discussions.

[6] R. F. Greene, Surface Sci. **2**, 101 (1964).

TABLE I. Effective masses for three surface orientations, for semiconductors having band structures like those of the conduction band of Si (six {100} ellipsoids of revolution) or of Ge (four {111} ellipsoids of revolution). The principal effective masses in the ellipsoids are m_t, m_t, and m_l. The derived values are m_3, the effective mass perpendicular to the surface, and m_1 and m_2, the principal masses of the constant-energy ellipse in the surface, defined by Eq. (13). The degeneracy of each set of ellipses is n_v.

Surface orientation	Si m_1	Si m_2	Si m_3	Si n_v	Ge m_1	Ge m_2	Ge m_3	Ge n_v
{100}	m_t	m_t	m_l	2	m_t	$(m_t+2m_l)/3$	$(3m_tm_l)/(m_t+2m_l)$	4
	m_t	m_l	m_t	4				
{110}	m_t	$(m_t+m_l)/2$	$(2m_tm_l)/(m_t+m_l)$	4	m_t	$(m_l+2m_t)/3$	$(3m_tm_l)/(m_l+2m_t)$	2
	m_t	m_l	m_t	2	m_t	m_l	m_t	2
{111}	m_t	$(m_t+2m_l)/3$	$(3m_tm_l)/(m_t+2m_l)$	6	m_t	m_t	m_l	1
					m_t	$(m_t+8m_l)/9$	$(9m_tm_l)/(m_t+8m_l)$	3

wave vectors for motion parallel to the surface will make equal angles with all the constant-energy surfaces of the bulk only in special cases, i.e., the {111} surfaces of Si and the {100} surfaces of Ge. On the other hand, for a (100) surface of Si, two of the six bulk constant-energy ellipsoids will give $m_3=m_l$, the longitudinal effective mass of the bulk, while the other four will give $m_3=m_t$, the transverse bulk effective mass. We list in Table I the values of the masses m_1, m_2, and m_3 for energy bands like those of Si and Ge for three high-symmetry surface orientations, and indicate in each case the degeneracy n_v, which gives the number of ellipsoids of the bulk that have equivalent sets of values. Where two different values of m_3 enter, the ith solution of (12) belonging to the larger mass will have a lower energy E_i'' than the ith solution belonging to the smaller mass. Note that the constant-energy ellipses for motion in the surface are not, in general, equivalent to slices through the constant-energy ellipsoids of the bulk.

The two-dimensional Brillouin zones and constant-energy ellipses for the cases of Table I are shown in Fig. 3, which indicates the kinds of degeneracy that can arise. Where a double degeneracy occurs at a point in the zone, as in the (100) surface of Si, we expect deviations from the effective-mass approximation to remove the degeneracy.[2] Since for low magnetic fields no splitting of the degeneracy is observed for Si,[2] we assume here that the two states can be described in the effective-mass approximation by the same envelope function, to which the boundary condition (4) applies. A more realistic treatment of the boundary conditions at the surface would give an estimate of the accuracy of this approximation, and would also give information about electron scattering associated with the surface.

The effective masses of the constant-energy ellipses given in Table I can be rather anisotropic, just as in the bulk. For most of the cases shown in Fig. 3, this anisotropy does not lead to an anisotropic conductivity because the ellipses are symmetrically placed. But for {110} surface orientations, which have lower symmetry, one should expect to see anisotropic conduction.[7]

[7] Dr. E. J. Walker (private communication) has informed us that measurements of the conductivity of (110) surface inversion layers on p-type Si have not so far shown any significant anisotropy.

Once we know the energy levels on the electric sub-bands of the surface inversion layer, we easily find the density of states for electrons, including a factor of 2 for spin, to be

$$\rho(E)=(\pi\hbar^2)^{-1}\textstyle\sum_i n_{vi}m_{di}H(E-E_i''), \qquad (14)$$

where the E_i'' are the various eigenvalues of (12), n_{vi} is the degeneracy of the electric sub-bands whose lowest energy is E_i'', m_{di} is the density-of-states mass per ellipse [in the notation of Table I, $m_{di}=(m_{1i}m_{2i})^{1/2}$], and $H(x)=1$ if $x>0$, $H(x)=0$ otherwise.

The number of carriers in the inversion layer at temperature T is

$$N_{\text{inv}}=\textstyle\sum_i N_i=(KT/\pi\hbar^2)\times\sum_i n_{vi}m_{di}F_0([E_F-E_i'']/KT), \qquad (15)$$

where K is Boltzmann's constant, E_F is the Fermi energy, and

$$F_0(x)=\ln[1+\exp(x)] \qquad (16)$$

is the Fermi-Dirac integral[8] with exponent zero. For large positive x, $F_0(x)\sim x$, and for large negative x, $F_0(x)\sim\exp(x)$. The charge density in the inversion layer can then be written

$$\rho_{\text{inv}}=-e\textstyle\sum_i N_i g_i(z), \qquad (17)$$

$$g_i(z)=|\zeta_i(z)|^2, \qquad (18)$$

where ζ_i is the normalized eigenfunction corresponding to the ith solution of (12), and N_i is given in (15). The function $g(z)$ is shown for a representative case in Fig. 2.

In the depletion layer, we suppose that all donors and acceptors are ionized, so that the charge density is equal to $e(N_D-N_A)$ in a layer whose thickness is approximately given by

$$d_{\text{depl}}=[\kappa|\phi(0)|/2\pi e|N_D-N_A|]^{1/2} \qquad (19)$$

in Gaussian units, i.e., with $\phi(0)$ given in terms of statvolts, where 1 statvolt $\simeq$300 V. If an n-type inversion layer forms, the conduction band edge at the surface is near the Fermi level, while the valence

[8] See, for example, A. H. Wilson, *The Theory of Metals* (Cambridge University Press, Cambridge, England, 1953), 2nd ed., p. 147.

band edge in the bulk is near the Fermi level, so that to a fair approximation we can replace $|\phi(0)|$ by E_g/e.[9] The total number of negative charges per unit area in the depletion layer is

$$N_{\text{depl}}=(N_A-N_D)d_{\text{depl}}. \tag{20}$$

If $N_A-N_D=10^{15}$ cm^{-3}, $\kappa=11.8$, and $e|\phi(0)|=E_g=1.15$ eV for Si, we find that $d_{\text{depl}}=1.2\ \mu$, and $N_{\text{depl}}=1.2\times10^{11}$ cm^{-2}.

The self-consistent equations (2), (12), and (17) for the inversion-layer energy levels and charge distributions have been solved for a range of temperatures and charge densities by Howard.[10] In this paper we use only the approximate expression[11]

$$g(z)=\tfrac{1}{2}b^3z^2\exp(-bz) \tag{21}$$

for the charge distribution when only the lowest inversion-layer energy level is occupied. A variational calculation[11] shows that for

$$b=([48\pi e^2m_3/\kappa_{sc}\hbar^2][N_{\text{depl}}+(11/32)N_{\text{inv}}])^{1/3}, \tag{22}$$

where m_3 has the same value as in (12), the energy of the lowest inversion-layer state lies within 6% of the energy found from a numerical self-consistent field calculation, for values of the ratio $N_{\text{inv}}/N_{\text{depl}}$ from 0 to at least 60.[10]

The average value of z, weighted by the charge distribution (21), is $3/b$, and we take this as a measure of the thickness of the inversion layer. It will be a maximum when $N_{\text{inv}}\to0$. In that case, if $m_3=m$ and $N_{\text{depl}}=1.2\times10^{11}$ cm^{-2}, we find that $3/b=45$ Å. Thus the inversion layer is considerably thinner than the depletion layer even in this case. As more charge is added by increasing the field at the semiconductor-insulator interface, almost all the excess goes into the inversion layer, while d_{depl} (and N_{depl}) remain essentially constant, provided that the surface states do not play a role.

Although most of the considerations of this paper are for zero magnetic field, we give in Appendix A some approximate results for the energy levels in the presence of a magnetic field making an arbitrary angle with the interface between the semiconductor and the insulator.

3. COULOMB POTENTIALS AND SCREENING

To calculate the effect of ionized impurities on the transport properties of carriers in the inversion layer, we must find the change in potential energy of the inversion-layer carriers because of the presence of an impurity. The carriers in the inversion layer will redistribute themselves in the presence of the impurity potential, increasing in number where their potential energy is lowered, and decreasing in number where their potential energy is raised, thus screening the charge of the impurity or other Coulomb center.

We shall calculate the effect of screening by analogy with the classical derivation of the screening length in a semiconductor in three dimensions.[12] We suppose that the impurity potential $\delta\phi$ depends only on the coordinates r and z, where $r=(x^2+y^2)^{1/2}$, and consider a region far enough away from the impurity that the potential $\delta\phi(r,z)$ is a slowly varying function of r. The potential then leads to an extra potential energy $-e\delta\phi(r,z)$ in (12), which to first order in $\delta\phi$ gives no change in the eigenfunction $\zeta_i(z)$, and gives a change

$$\delta E_i(r)=-e\bar{\phi}_i(r)=-e\int\delta\phi(r,z)g_i(z)dz \tag{23}$$

in the ith eigenvalue, where $g_i(z)$ is the normalized charge distribution (18). If we substitute the energy change associated with the impurity potential into (15), and keep only the terms linear in $\delta\phi$, we find that the induced charge density in the inversion layer is

$$\rho_{\text{ind}}(r,z)=-(\kappa/2\pi)\textstyle\sum_i s_i\bar{\phi}_i(r)g_i(z), \tag{24}$$

where

$$s_i=(2\pi e^2/\kappa)N_i/E_{d,i}, \tag{25a}$$

$$E_{d,i}/KT=F_0(x)/F_0'(x)$$
$$=[1+\exp(-x)]\ln[1+\exp(x)], \tag{25b}$$

$x=(E_F-E_i'')/KT$, and all other quantities have the same significance as in (15) and (17).

For the electric quantum limit at absolute zero, the screening constant (25) is independent of carrier concentration, and has the value

$$s=2n_vm_de^2/\hbar^2\kappa. \tag{26}$$

This agrees with the result obtained from the wave-vector-dependent longitudinal dielectric constant of a degenerate two-dimensional electron gas.[13]

Poisson's equation (2) for the potential change $\delta\phi$ thus becomes

$$\nabla^2\delta\phi(r,z)-2\textstyle\sum_i s_i\bar{\phi}_i(r)g_i(z)=-4\pi\rho_{\text{ext}}(r,z)/\kappa, \tag{27}$$

where ρ_{ext} is the external charge density, here associated with the impurity ions or other charged centers. Solutions of (27) for a number of cases of interest are given in Appendix B and are used in the following sections to calculate the energies of bound states in the inversion layer, and the differential cross section for screened Coulomb scattering.

[9] This approximation tends to overestimate d_{depl} in high-resistivity Si, for which the Fermi level in the bulk may be well inside the energy gap, and to underestimate d_{depl} in heavily doped InAs, for which the conduction band at the surface must lie well below the Fermi level before the inversion layer begins to be occupied.

[10] W. E. Howard (unpublished).

[11] F. F. Fang and W. E. Howard, Phys. Rev. Letters **16**, 797 (1966). We are indebted to J. F. Janak for pointing out a correction to the variational treatment indicated by Fang and Howard. The correction is used in finding Eq. (22) of the present paper.

[12] See, for example, F. Stern, Solid State Phys. **15**, 299 (1963), Sec. 48.

[13] F. Stern, Phys. Rev. Letters **18**, 546 (1967).

The simple screening constant given in (26) no longer applies if more than one electric sub-band is occupied, or at temperatures for which completely degenerate statistics cannot be assumed. In particular, the screening constant for higher sub-bands will be larger than for the lowest sub-band, an effect discussed for a three-dimensional case by Robinson and Rodriguez.[14] The increase in the screening constant for the higher sub-bands is balanced at least in part by the smaller values of $\bar{\phi}_i(r)$ and $g_i(z)$ associated with the greater spatial extent of the charge distribution in the higher sub-bands.

4. BOUND STATES

A. General Considerations

If the electrons in the inversion layer move in an average potential $\bar{\phi}(r)$, and if they have an isotropic[15] effective mass $m^*=m_1=m_2$, then their energy levels and wave functions will satisfy the two-dimensional effective-mass equation

$$[-(\hbar^2/2m^*)\nabla^2-e\bar{\phi}(r)-E]\psi(x,y)=0. \quad (28)$$

We assume that only the lowest electric sub-band is occupied, and identify the zero of energy in (28) with the bottom of this sub-band. The procedure for determining $\bar{\phi}(r)$ in (28) is discussed in Sec. 3 and in Appendix B. Thus we study here the effect of an additional potential, assumed to be attractive, whose z dependence has been treated by first-order perturbation theory and replaced by its average value, while its r dependence is to be treated using (28). The complete envelope function of bound states in this potential is given by the product of the envelope function found in (28) and the wave function $\xi(z)$ given in (11).

If we take advantage of the circular symmetry of $\bar{\phi}$, and make a transformation to eliminate the first derivative, then (28) becomes

$$P''+\{(\tfrac{1}{4}-m^2)r^{-2}+(2m^*/\hbar^2)[E+e\bar{\phi}(r)]\}P(r)=0, \quad (29)$$

where $P''=d^2P/dr^2$, and

$$\psi(x,y)=r^{-1/2}P(r)\exp(\pm im\theta). \quad (30)$$

The angular-momentum quantum number m here takes on the values 0, 1, 2, $\cdots$, and $P(r)$ must vanish at the origin and at infinity for a bound state.

If there were no screening by the inversion-layer electrons, if the inversion-layer charge distribution $g(z)$ were a delta function at the interface between the semiconductor and the insulator, and if the Coulomb center of charge e were also at the interface, then the effective potential energy in (28) would be[16]

$$-e\bar{\phi}(r)=-e^2/\bar{\kappa}r, \quad (31)$$

where

$$\bar{\kappa}=\tfrac{1}{2}(\kappa_{sc}+\kappa_{ins}) \quad (32)$$

is the average of the dielectric constants in the semiconductor and the insulator.

The solution of (29) proceeds by analogy with the three-dimensional case,[17,18] and the eigenvalues are the infinite set of $(2n-1)$-fold degenerate levels[19]

$$E_n=-(n-\tfrac{1}{2})^{-2}\mathrm{Ry}^*, \quad (33)$$

where the unit of energy is the effective rydberg

$$\mathrm{Ry}^*=m^*e^4/(2\bar{\kappa}^2\hbar^2), \quad (34)$$

and the quantum number n takes on the values $m+1$, $m+2$, $\cdots$. The ground-state envelope function is

$$\psi=(8/\pi)^{1/2}a^{*-1}\exp(-2r/a^*), \quad (35)$$

where

$$a^*=\bar{\kappa}\hbar^2/m^*e^2 \quad (36)$$

is the effective Bohr radius. Using parameters appropriate to Si-SiO_2 interfaces, as given in Table II, we find that $\mathrm{Ry}^*=42$ meV and $a^*=22$ Å.

For a Coulomb potential due to an ion at the surface of a semiconductor, with electrons free to move in the semiconductor, instead of being confined to an inversion-layer sheet as we assumed, the ground-state energy is $-\frac{1}{4}\mathrm{Ry}^*$.[20]

TABLE II. Values of parameters used in the numerical calculations for InAs and for the (100) surface of Si.

	InAs	Si-(100)
Electron effective mass parallel to surface (m)	0.025	0.19
Electron effective mass perpendicular to surface (m)	0.025	0.98
κ_{sc}, dielectric constant of semiconductor	14.3	11.8
κ_{ins}, dielectric constant of insulator	2.8[a]	3.8[b]
$\bar{\kappa}=(\kappa_{sc}+\kappa_{ins})/2$	8.55	7.8
n_v, sub-band degeneracy	1	2
n_1, factor for interband scattering[c]	1	1
a^*, effective Bohr radius[d] (Å)	181	22
Ry*, effective rydberg, Eq. (34) (meV)	4.6	42
δ, degenerate screening constant using $\bar{\kappa}$[e] (cm^{-1})	1.1×10^6	1.8×10^7

[a] Mylar
[b] SiO_2
[c] See Eq. (46).
[d] Equation (36).
[e] See Eq. (41).

[14] J. E. Robinson and S. Rodriguez, Phys. Rev. **135**, A779 (1964).

[15] When the effective mass is not isotropic, the methods described by W. Kohn [Solid State Phys. **5**, 257 (1957)] can be used.

[16] This found from the results of Appendix B or from the method of images [see, for example, J. H. Jeans, *The Mathematical Theory of Electricity and Magnetism* (Cambridge University Press, Cambridge, England, 1951), 5th ed., Sec. 22].

[17] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955), 2nd ed., Sec. 16.

[18] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1958), Sec. 36.

[19] S. Flügge and H. Marschall, *Rechenmethoden der Quantentheorie* (Springer-Verlag, Berlin, 1952), 2nd ed., Problem 24.

[20] J. D. Levine, Phys. Rev. **140**, A586 (1965), Eq. (10). See also E. P. Prokop'ev, Fiz. Tver. Tela **8**, 2770 (1966) [English transl.: Soviet Phys.—Solid State **8**, 2209 (1967)]. The dielectric constant κ should be replaced by $\bar{\kappa}=\frac{1}{2}(\kappa+1)$ in these papers if our Eq. (31) applies.

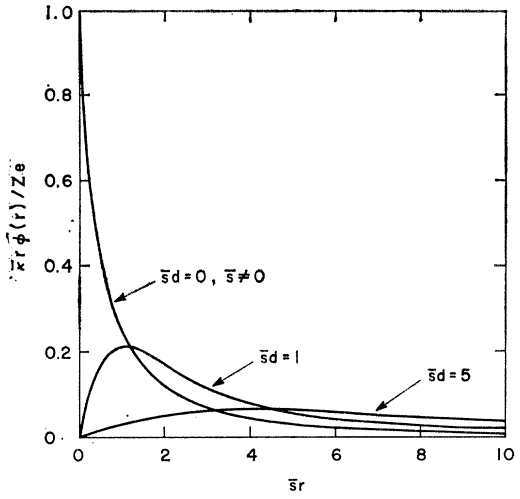

FIG. 4. Screened Coulomb potential in an inversion layer of zero thickness located at the semiconductor-insulator interface when a charge e is located a distance d away, as given by Eq. (B19). The screening constant is $\bar{s}=4/a^*$, where a^* is the effective Bohr radius of Eq. (36). The average dielectric constant of the two media is $\bar{\kappa}$, and r is the radial distance in the plane from the point closest to the charge.

For potentials differing from the Coulomb potential, we can infer something about the solutions from the strength integral

$$I=(2m^*e/\hbar^2)\int_0^\infty r|\phi(r)|\,dr, \tag{37}$$

which is finite when there is screening. To do this, we note that Eq. (29) for the function $P(r)$ is identical with the corresponding three-dimensional radial equation if we identify the two-dimensional orbital angular-momentum quantum number m with $l+\frac{1}{2}$, where l is the three-dimensional quantum number. We can then make use of a theorem proved by Bargmann,[21] which, restated for two dimensions, says that the number n_m of solutions of (29) for $m\neq 0$ obeys the inequality

$$2mn_m<I. \tag{38}$$

Bargmann's proof is not directly applicable when $m=0$, but it has been shown that for any attractive circularly symmetric potential there is always a two-dimensional bound state with orbital angular-momentum zero, no matter how weak the potential is,[22,23] which is not true in the three-dimensional case.[21] If we identify the range of the potential with twice the reciprocal of the screening constant $\bar{s}$, where $\bar{s}$ is obtained from (25) by replacing κ by the average dielectric constant $\bar{\kappa}$ of (32), we find that the estimated[22] energy of the lowest state is, for small I,

$$E\approx(\hbar^2\bar{s}^2/4m^*)\exp(-2/I). \tag{39}$$

The strength integral I can be calculated for the screened potential associated with charge Ze located at an arbitrary distance from the semiconductor-insulator interface, as calculated in Appendix B. We find, using (B17), that

$$I=2|Z|/a^*\bar{s}, \tag{40}$$

independent of the position of the ion and of the thickness of the inversion layer.

For the electric quantum limit at absolute zero, we find from (26) that

$$\bar{s}=2n_v/a^*, \tag{41}$$

so that $I=n_v^{-1}$ if $|Z|=1$. Thus the screened Coulomb potential is not strong enough under these circumstances to bind states with orbital angular-momentum quantum number greater than 0. However, I is not small enough to make (39) a reliable estimate for the binding energy when $m=0$. Thus a more detailed treatment is required to find the binding energy for the screened Coulomb potential.

B. Bound States for a Model Potential

To investigate the energies of bound states for a tractable potential $\bar{\phi}(r)$, we use the model that leads to Eq. (B19), with $Z=+1$ to make the potential attractive. The potential is calculated for an inversion layer of zero thickness located at the semiconductor-insulator interface, with a charge e located a distance d away, and is illustrated for a screening constant $\bar{s}=4/a^*$ in Fig. 4. This model does not accurately represent the more realistic potential given by (B9)–(B16) because it ignores the thickness of the inversion layer, but was used because of its considerably simpler form. The difference between our model potential and the more realistic potential is not likely to obscure qualitative features such as the dependence of the bound-state energies on the separation d and on the screening constant $\bar{s}$.

In applying our model potential, we identify the distance d between the inversion-layer sheet and the attractive Coulomb center with the average thickness,

$$z_{\rm av}=3/b, \tag{42}$$

of the inversion-layer distribution of Eqs. (21) and (22). This is intended to represent the average distance between an electron in the inversion layer and an attractive center at the semiconductor-insulator interface. For centers in the insulator, larger values of d should be used. Note that the ionized acceptors in the depletion layer of the p-type semiconductor are negatively charged, and thus constitute repulsive centers. There will be attractive centers in the depletion layer if it is compensated.

[21] V. Bargmann, Proc. Natl. Acad. Sci. (U. S.) 38, 961 (1952).
[22] See p. 156 of Ref. 18.
[23] This has been proved for an arbitrary attractive potential in two dimensions by D. Jepsen and T. D. Schultz (to be published) and by J. C. Slonczewski (unpublished), but disagrees with a remark in P. M. Morse and H. Feshbach [*Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), p. 1654].

We have obtained numerical solutions of the two-dimensional effective-mass equation (28) for the model potential (B19) with a number of values of d, and a number of values of the screening constant $\bar{s}$. In particular, when $\bar{s}=0$, the potential becomes the unscreened potential[16]

$$\bar{\phi}(r)=(e/\bar{\kappa})(r^2+d^2)^{-1/2}. \tag{43}$$

Results of the calculations for $\bar{s}=4/a^*$ and $\bar{s}=0$ are shown in Fig. 5. The integration procedure used is due to Cooley.[24]

The large ratio of binding energies for the unscreened and screened cases shown in Fig. 5 suggests the possibility of interesting freeze-out effects in inversion layers at low temperatures as the number of charges in the inversion layer is changed. We shall return to this point in Sec. 6.

Figure 6 shows how the binding energies vary with the screening constant for a number of values of d. From this figure we can estimate binding energies for InAs at absolute zero, for which $\bar{s}=2/a^*$, and can obtain approximations to deal with the screening at nonzero temperatures. The lower curve of Fig. 6, for $d=2a^*$, can be fairly well approximated by $E_b=0.54\times\exp[-2.8(a^*\bar{s})^{0.61}]\mathrm{Ry}^*$. The simple result (41) for $\bar{s}$ holds only at low temperatures. In general, as shown in (25), the screening constant depends both on temperature and on inversion-layer carrier concentration.

Thus far we have discussed only the lowest bound state associated with the lowest electric sub-band. For completeness, we must consider also three other kinds of states. The first of these are the higher-excited bound states associated with the lowest electric sub-band. We shall not deal with these further, since they are likely to be so close to the bottom of the sub-band that they do not have important experimental consequences.

Second, there are the bound states associated with higher electric sub-bands derived from the same bulk-energy surfaces that give the lowest sub-band. The higher sub-bands have more extended charge distributions, and therefore weaker potentials and smaller binding energies. In addition, the energies of the bound states will be raised by the amount of the splitting between sub-bands. Nevertheless they may lie lower in energy than the bottom of the lowest sub-band if the screening constant is small, and thus represent possible excited states of the attractive center. If these states lie in the first sub-band, they may affect the scattering of those carriers whose energy lies nearby.

The third type of state arises in a surface such as (100)-Si, in which there are higher electric sub-bands associated with bulk-energy ellipsoids not equivalent to those that give rise to the lowest sub-band. In this case, as in the previous one, the binding energy is reduced because of the greater value of z_{av} for the higher sub-band, and the state is raised by the sub-band separation. But now the effective rydberg is increased because of the larger effective masses associated with the second family of constant-energy ellipses for motion parallel to the surface. In general, these states could even lie lower than the states whose energies are shown in Figs. 5 and 6, and could be the primary bound states associated with the attractive center. However, estimates for the case of (100)-Si, based on calculations of the upper electric sub-bands,[10] indicate that these states are higher in energy than the ground state associated with the lowest electric sub-band, at least for conditions near the conductance threshold. From Fig. 6 it is clear that as the screening increases the last two types of bound states, being associated with larger

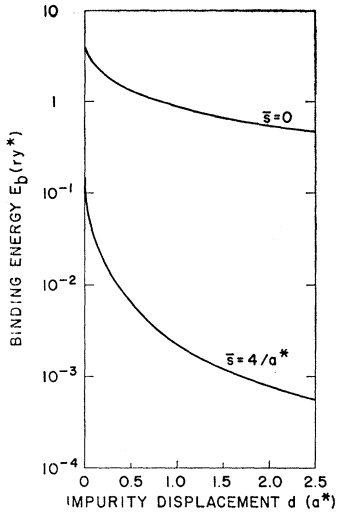

FIG. 5. Calculated binding energies for the model potential of Eq. (B19), with screening constants $\bar{s}=0$ and $4/a^*$. The unit of energy is the effective rydberg, given in Eq. (34).

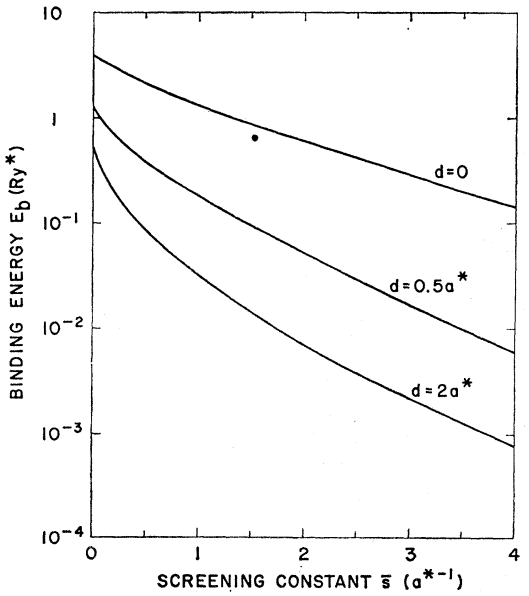

FIG. 6. Binding energy as a function of screening constant for three distances between the impurity and the inversion-layer plane, calculated from the model potential (B19).

[24] J. W. Cooley, Math. Comput. 15, 363 (1961). We are indebted to Dr. Cooley for providing the program for the integration.

z_{av}, will decrease in depth at a faster rate than the ground state associated with the lowest sub-band, and thus remain higher in energy if they are higher in the absence of screening.

Since all the kinds of states we have discussed arise from a single attractive center, they are statistically linked to each other. Only one state of the set may be occupied at a given time.

The considerations thus far, including the energies shown in Figs. 5 and 6, apply only to isolated Coulomb centers. For densities of practical interest, the overlapping of wave functions of the bound states will limit the range of binding energies for which these states can be represented as discrete levels separated from the bottom of the lowest electric sub-band. By analogy with the three-dimensional situation,[25] we estimate that overlap of the bound-state energies with the continuum occurs when

$$r_s/a^* \sim 2(E_b[\mathrm{Ry}^*])^{-1/2}, \tag{44a}$$

where $r_s=(\pi N_{\mathrm{int}})^{-1/2}$ is the radius of the Wigner-Seitz circle associated with a concentration N_{int} of attractive centers per unit area, and E_b is the magnitude of the binding energy of the bound state. We can rewrite (44a) in the form

$$N_{\mathrm{overlap}} \sim E_b[\mathrm{Ry}^*]/4\pi a^{*2}. \tag{44b}$$

For smaller concentrations of attractive centers the bound states will form a distinct impurity band, and for larger concentrations they will merge with the bottom of the continuum and form a tail in the density of states.

The density of states associated with random impurity centers is a complicated problem in three dimensions. The reader should recognize that Eq. (44) is a first attempt to estimate the overlap effects in the two-dimensional case, and that the specific values are subject to revision.

If we nevertheless use the estimate in (44), and take the binding energy from Fig. 5 for the *unscreened* potential, using $d=43$ Å for (100)-Si and $d=101$ Å for InAs [26] and taking a^* from Table II to obtain $E_b=0.5$ and 1.2 Ry*, respectively, in the two cases, we find that the concentrations for which overlap effects become important are 9×10^{11} cm^{-2} for (100)-Si and 3×10^{10} cm^{-2} for InAs. The values of N_{int} encountered in typical experiments[2,4,11] on (100)-Si surfaces are in the range 1 to 5×10^{11} cm^{-2}, while the value found in the InAs experiment[3] is 6×10^{11} cm^{-2}. We conclude that the bound states for the *unscreened* potential have fully merged with the continuum in InAs, and are more or less discrete in (100)-Si. As enough carriers are added to the inversion layer to introduce screening, the bound states in (100)-Si will also merge with the continuum.

5. IMPURITY SCATTERING

A. Born Approximation

Scattering of an electron in the inversion layer from an initial state ψ_i to a final state ψ_f by the potential of a charged scattering center is most easily treated in Born approximation. If the scattering potential is $-e\phi(r,z)$, then the differential cross section is found from time-dependent perturbation theory[27] to be

$$\sigma(\theta)=(\Omega^2 m^* e^2/2\pi\hbar^3 v_i)\,|\,(\psi_f|\phi|\psi_i)\,|^2, \tag{45}$$

where m^* is the effective mass (assumed circularly symmetric) for the final state, v_i is the carrier velocity for the initial state, and Ω is the area of normalization for the wave functions.

If we write

$$\psi_i=\Omega^{-1/2}h_i(z)\exp(i\mathbf{k}_i\cdot\mathbf{r}), \tag{46}$$

where h is normalized, and a similar expression for ψ_f, we find that the matrix element in (45) becomes

$$(2\pi/\Omega)\int_0^\infty J_0(Sr)rdr\int_0^\infty h_i(z)h_f^*(z)\phi(r,z)dz, \tag{47}$$

where $S=|\mathbf{k}_f-\mathbf{k}_i|$ is the wave-vector change during scattering.

When the lowest electric sub-band is degenerate, two kinds of complications may arise. The first, illustrated by the (111) surface of Si in Fig. 3, occurs if the n_v-fold degenerate sub-band has minima at n_v points equivalent by symmetry in the two-dimensional Brillouin zone associated with the surface periodicity. In that case, one would have to consider the anisotropy in the scattering associated with the anisotropic mass of each ellipse,[28] and the possibility of intervalley scattering by the Coulomb centers.[29] We shall not consider such cases further here.

A second complication is illustrated by the lowest sub-band for the (100) surface of silicon, which is at the center of the Brillouin zone and is doubly degenerate. The two states are associated with wave functions having different dependences on the coordinate perpendicular to the surface on an atomic scale, but the same over-all modulating function. Thus the sub-bands are degenerate in our effective-mass approximation, but the degeneracy could be removed in a more exact treatment. The experimental evidence[2] suggests that the splitting is small at low magnetic fields.

[25] See, for example, N. F. Mott and W. D. Twose, Advan. Phys. **10**, 107 (1961).

[26] These values of d are taken from (22) and (42), for 3×10^{11} inversion-layer electrons/cm^2, and for N_A equal to 1×10^{14} and 2.6×10^{16} cm^{-3} for Si and InAs, respectively.

[27] See, for example, Sec. 29 of Ref. 17.

[28] R. A. Laff and H. Y. Fan, Phys. Rev. **112**, 317 (1958); L. J. Neuringer and D. Long, *ibid.* **135**, A788 (1964); A. D. Boardman, *ibid.* **147**, 532 (1966); V. M. Bondar, O. G. Sarbei, and P. M. Tomchuk, Fiz. Tver. Tela **8**, 2511 (1966) [English transl.: Soviet Phys.—Solid State **8**, 2012 (1967)].

[29] G. Weinreich, T. M. Sanders, and H. G. White, Phys. Rev. **114**, 33 (1959).

We must make allowance for this double degeneracy in calculating the total scattering cross section, since carriers may scatter not only within the same sub-band, but also into the other sub-band of the degenerate pair. In the absence of a detailed theory, we allow for such scattering by replacing $h(z)$ in (46) by the envelope function $\zeta(z)$ of Eq. (12) for the lowest sub-band,[30] and then multiplying the cross section given by (45) by a factor n_s. If the z dependence of the perturbing potential is weak, then $n_s \approx 1$. When $\zeta(z)$ is used for $h(z)$, the integration over z in (47) gives $\bar{\phi}(r)$, the average impurity potential in the inversion layer, as discussed in Sec. 3 and in Appendix B.

The average potential $\bar{\phi}(r)$ in the inversion layer associated with an impurity located at $x=y=0$, $z=z_0$, is

$$\bar{\phi}(r)=\int_0^\infty \bar{A}(k,z_0)J_0(kr)kdk, \tag{48}$$

where the coefficients $\bar{A}(k,z_0)$ are given in (B9)–(B16). Thus the Born approximation differential cross section associated with an impurity at z_0 is

$$\sigma(\theta)=(2\pi m^*e^2n_s/\hbar^3v)\,|\bar{A}(S,z_0)|^2 \tag{49}$$

if we use (45)–(48), and make use of the relation (B7) to evaluate the integral over r in (47).

The rate of scattering through an interval $d\theta$ is taken to be the sum of the scattering rates of the individual scattering centers acting independently. It is given by $\tau^{-1}(\theta)d\theta$, where

$$\tau^{-1}(\theta)=(2\pi m^*e^2n_s/\hbar^3)\int_{z_1}^{z_2} N(z_0)\,|\bar{A}(S,z_0)|^2dz_0, \tag{50}$$

and $N(z_0)$ is the density-of-scattering centers per unit volume at a distance z_0 from the semiconductor-insulator interface.

To proceed further we must know $N(z_0)$, the distribution of charged scattering centers. For the purpose of this paper we assume that there is a uniform volume density N_{ins} of centers in the insulator, a density $N_{\text{sc}}=N_A+N_D$ in the semiconductor depletion layer, where N_A and N_D are the concentrations of ionized donors and acceptors, respectively, and a surface density N_{int} of charged scattering centers at the interface. The limits z_1 and z_2 correspond to the boundaries of the insulator and of the depletion layer, respectively, but can be taken to be infinite if, as in the cases we consider, $|Sz_1|\gg 1$ and $Sz_2\gg 1$ for most values of scattering angle.

With these approximations we find that the momentum-scattering rate due to charged scattering centers can be written

$$\tau_{\text{ion}}^{-1}=\int \tau^{-1}(\theta)(1-\cos\theta)d\theta$$
$$=(2\pi m^*e^2n_s/\hbar^3)\Sigma_jN_j\int_{-\pi}^{\pi} B_j(S)(1-\cos\theta)d\theta, \tag{51}$$

where θ enters through $S=2k\sin\frac{1}{2}\theta$. The index j refers to the interface, insulator, and semiconductor depletion-layer scattering centers, respectively, with surface or volume concentrations N_j. The coefficients B_j are given in (B21). At low temperature, the only case we evaluate in this paper, the mobility is given by $e\tau_{\text{ion}}/m^*$, where τ is evaluated for $k=k_F$, the wave vector at the Fermi surface.

Numerical examples for the mobility calculated from the foregoing expressions for (100)-Si inversion layers and for InAs inversion layers are given in the following section.

B. Validity of the Born Approximation

We now turn to the evaluation of the validity of the Born approximation for screened potentials. The example we analyze is that of an impurity of charge Ze, with $Z=\pm 1$, located a distance d from an inversion layer of zero thickness located at the interface between two dielectrics whose average dielectric constant is $\bar{\kappa}$. The screened potential for this case is the same one used in Sec. 4B, and is given by Eq. (B19) and illustrated for various values of d in Fig. 4.

The momentum scattering cross section, which includes the weighting factor $1-\cos\theta$, has been evaluated numerically for this potential using the phase-shift method and Eq. (C11). In addition, we have evaluated the momentum cross section using the Born approximation. From (C11), (49), and (B18) we find

$$\sigma_m=(2\pi m^*e^4/\bar{\kappa}^2\hbar^3v)$$
$$\times\int_{-\pi}^{\pi}(S+\bar{s})^{-2}\exp(-2Sd)[1-\cos\theta]d\theta. \tag{52}$$

The screening constant $\bar{s}$ was taken from (41) to equal either $4/a^*$ or $2/a^*$, corresponding to $n_v=2$ or $n_v=1$, respectively. The former case would apply to (100)-Si surfaces, and the latter to InAs surfaces. We take the scattering factor n_s of Eq. (49) to equal 1 here.

The momentum cross section, in units of a^*, is given in Fig. 7 as a function of Fermi energy $\frac{1}{2}m^*v^2$ for three values of $\bar{s}d$, with $\bar{s}=4/a^*$. The curves for the phase-shift calculation are for attractive ($Z=1$) and repulsive ($Z=-1$) potentials, while the Born-approximation curve applies for both signs of Z. The Born approximation gives quite good results over a wide energy range when $\bar{s}d\gtrsim 5$, but fails at small values of $\bar{s}d$ and at low energies.

[30] If the quantities w_{13} and w_{23} in Eq. (12) are nonzero, the expression for the cross section becomes somewhat more complicated. This complication does not arise for electrons in (100)-Si or in InAs inversion layers, and is ignored in the following discussion.

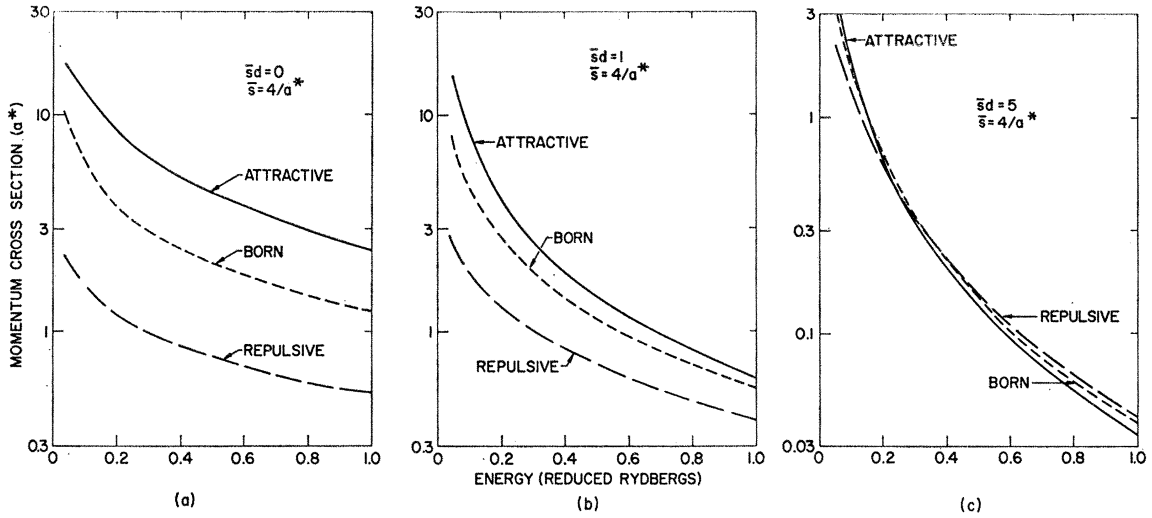

FIG. 7. Momentum cross section, calculated using Eq. (C11) and the model potential of Eq. (B19), for three values of the distance d between the external charge Ze and the inversion-layer plane, for $Z=1$ (attractive potential) and $Z=-1$ (repulsive potential). The Born-approximation cross section is independent of the sign of Z.

To see what values of the parameters correspond to typical experimental conditions, we identify the distance d between the scattering charge and the inversion layer with the average inversion-layer thickness given by Eq. (42). We consider a (100)-Si sample with 10^{12} inversion-layer carriers/cm^2, and with 10^{15} acceptor ions/cm^3 in the depletion layer. The momentum cross section, as given by the phase-shift method of Eq. (C11), and by the Born approximation, Eq. (52), is shown in Table III. We see that the phase-shift cross sections for attractive and repulsive potentials lie within 15% of each other, and that the Born-approximation cross section lies slightly higher.

Also shown in Table III is a similar comparison for InAs, using numbers representative of the experiment of Kawaji and Kawaguchi.[3] Here again we find the three cross sections to be rather close together.

TABLE III. Comparison of the momentum cross section as calculated by the Born approximation and by the phase-shift method of Eq. (C11) for attractive ($Z=1$) and repulsive ($Z=-1$) screened Coulomb scatterers of charge Ze. The potential used is that of Eq. (B19), in which the distance d between the scatterer and the inversion layer is taken to equal the quantity z_{av} of Eq. (42). The calculation is made for the indicated values of the inversion-layer electron concentration and bulk-acceptor concentration, using the parameters of Table II. Also given are the sums of the scattering phase shifts, taken over all values of the orbital angular-momentum quantum number m.

	Si (100)	InAs
N_{inv} (cm^{-2})	1.0×10^{12}	3.0×10^{11}
N_A (cm^{-3})	1.0×10^{15}	2.6×10^{16}
E_F (Ry*)	0.15	6.2
z_{av} (Å)	29	101
$\bar{s}d$	5.3	1.1
σ_m(Born) (a^*)	0.91	0.017
$\sigma_m(Z=1)$ (a^*)	0.89	0.015
$\sigma_m(Z=-1)$ (a^*)	0.78	0.019
$\sum\eta_m$(Friedel)	$\pi/4$	$\pi/2$
$\sum\eta_m(Z=1)$	0.80	1.57
$\sum\eta_m(Z=-1)$	0.75	1.57

Another test of the screened potential due to a charge Ze is the Friedel phase-shift sum rule, which we write in a slightly generalized form[31] for electrons in two dimensions:

$$2\pi^{-1}\sum_{m,i} n_{vi}\int_{-\infty}^{\infty}(d\eta_{m,i}/dE)f(E)dE=Z, \qquad (53)$$

where the sum goes over all orbital angular-momentum quantum numbers m from $-\infty$ to ∞, and over all electric sub-bands i, with orbital degeneracy n_{vi}. The integration is over all electron energies E, and $f(E)$ is the Fermi-Dirac occupation probability of a state of energy E.

Implicit in our use of a phase shift is the requirement that the constant-energy surfaces have circular symmetry, which is satisfied for InAs, and for the lowest sub-band in (100) surfaces. Then if only the lowest electric sub-band of electrons is occupied, (53) simplifies at absolute zero to

$$\sum_{-\infty}^{\infty}\eta_m(E_F)=\frac{\pi Z}{2n_v}. \qquad (54)$$

We have evaluated the phase-shift sum in (54) for the cases summarized in Table III, and give the results at the bottom of that Table. Phase shifts through $|m|=5$ were calculated for Si, and through $|m|=19$ for InAs. The contributions of higher terms were estimated by using the Born approximation (C16) for the phase shifts, the asymptotic expression $\bar{\phi}(r)\sim r^{-3}$

[31] F. Stern, Phys. Rev. 158, 697 (1967).

for $\bar{s}r \gg 1$ given in (B20a), and the relation[32]

$$\int_0^\infty J_\mu^2(kr) r^{-2} dr = (k/\pi)(\mu^2 - \tfrac{1}{4})^{-1}, \quad |\mu| > \tfrac{1}{2}, \quad (55)$$

which shows that the phase shifts fall off approximately as m^{-2} for large $|m|$. This behavior was exhibited by the calculated phase shifts.

The phase-shift sums for InAs agree to three figures with the requirement of the Friedel phase-shift sum rule, suggesting that we have chosen a self-consistent screened potential,[31] yet the Born-approximation momentum cross section differs from the phase-shift results by about 15%. One of us has shown elsewhere[31] that in three dimensions the Friedel sum rule is satisfied for the conventional screened Coulomb potential if the Born approximation is valid. A similar result applies in the two-dimensional case. The results of Table III show that the converse does not necessarily hold.

For Si (100) surfaces the phase-shift sums differ by less than 5% from the predicted value. They are smaller than those for InAs by about the factor of 2 which arises via (54) because $n_v = 1$ for InAs and $n_v = 2$ for Si (100) surfaces. We note that the Born-momentum cross section of Table III is quite close to the phase-shift value for an attractive potential. The dominant scatterers are the interface scatters, which for the experimental case considered[2,4] have an attractive potential. However, the agreement must be considered accidental, and the spread of about 15% between the attractive and repulsive values of σ_m must be taken as a better guide to the validity of the Born approximation.

On the basis of the results in Table III and in Fig. 7 we conclude that the Born approximation is satisfactory for the experimental range of N_{inv} in InAs, but becomes increasingly unreliable for determining the momentum cross section in (100)-Si surfaces from the screened Coulomb potential when the inversion-layer carrier density is much below 10^{12} cm^{-2}. On the other hand, for values near 10^{12} and above, which cover much of the range examined experimentally,[4] the Born approximation should give reasonable estimates for screened Coulomb scattering in the lowest electric sub-band.

In calculating the potential and the scattering cross section, we have here considered only screening by free electrons in the inversion layer. But there will also be screening by bound electrons, and modification of the screening by free electrons in the neighborhood of bound electrons. We have not treated these problems, although they are clearly important when the inversion layer is sparsely populated by electrons, some of which may be in bound states and some in the electric sub-band.

Although our calculations are restricted to the electric-quantum limit, in which only the lowest electric sub-band is occupied, the scattering rate may be affected by virtual transitions to higher-lying sub-bands as the carrier concentration increases and the Fermi level approaches the bottom of the first excited sub-band. Such processes have not been included in the cross sections calculated in this section.

[32] G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, Cambridge, England, 1962), 2nd ed., p. 403.

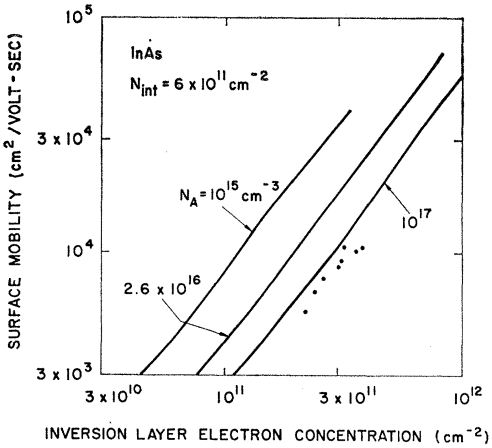

FIG. 8. Calculated low-temperature mobility for an n-type inversion layer on InAs with the indicated values of bulk-acceptor concentration N_A per cm^3, and with 6×10^{11} singly charged scatterers per cm^2 at the InAs-Mylar interface. The points were measured at 4.2°K by Kawaji and Kawaguchi (Ref. 3), on a sample with a net acceptor concentration of 2.6×10^{16} cm^{-3}. The calculated curves for $N_A = 10^{15}$ and 2.6×10^{16} cm^{-3} terminate where the Fermi level enters the next-higher electric sub-band (see Ref. 10) and additional scattering processes must be considered. For $N_A = 10^{17}$ cm^{-3}, the calculated crossover occurs at $N_{\text{inv}} = 1.3 \times 10^{12}$ cm^{-2}.

6. COMPARISON WITH EXPERIMENT

Since InAs and Si are the only materials for which low-temperature inversion-layer mobilities have been measured,[2-4] we have calculated surface mobilities for a number of cases for these materials, using the Born approximation as expressed in Eq. (51) and (B21), and material parameters given in Table II. Note that while the comparison of Born approximation and phase shift cross sections in Sec. 5 was made for a delta-function electron distribution in the z direction, the calculated mobilities given in this section have been obtained for electrons distributed according to the variational function given by (21) and (22). The results apply only to the low-temperature limit, and assume that screened Coulomb scattering within a single sub-band is the only important scattering mechanism.

A. InAs

We have calculated the electron mobility at absolute zero in inversion layers on uncompensated p-type InAs with 6×10^{11} interface Coulomb scatterers per cm^2, the value estimated by Kawaji and Kawaguchi[3] for their sample if all the centers are singly charged, and for several values of bulk-acceptor concentration N_A. The results are shown in Fig. 8 together with the

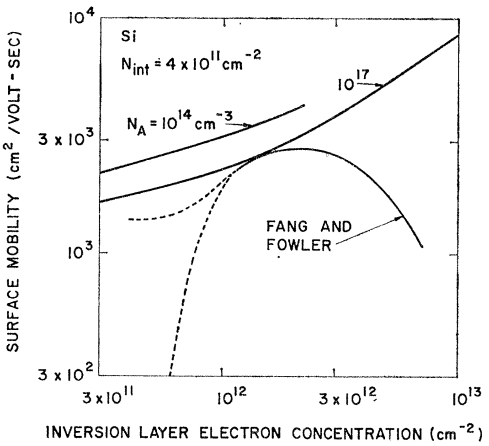

FIG. 9. Calculated and measured low-temperature mobility of electrons in inversion layers on (100) Si-SiO_2 interfaces. The two upper curves are calculated for 4×10^{11} interface scatterers per cm^2 and for the indicated bulk-acceptor concentrations, using Eq. (51). The curve for $N_A=10^{14}$ cm^{-2} is terminated where the Fermi level enters the next-higher electric sub-band. The experimental data are taken from the work of Fang and Fowler (Ref. 4), for a sample with $N_A=1\times10^{14}$ cm^{-3}. The lower dashed branch of the experimental curve is obtained from the surface conductance at 4.2°K, assuming that the carriers do not freeze-out. The upper dashed branch is obtained by extrapolating the exponential dependence of conductance on T^{-1} to zero reciprocal temperature. It represents the estimated mobility if the exponential dependence is ascribed to freeze-out of carriers into inversion-layer bound states, as discussed in the text.

mobilities measured at 4.2°K by Kawaji and Kawaguchi on a sample with 2.6×10^{16} net acceptors per cm^3.

The case of InAs is rather straightforward because, as noted in Eq. (44), we expect the bound states to merge with the lowest electric sub-band even in the absence of screening for $N_{int}\gtrsim3\times10^{10}$ cm^{-2}. Furthermore, the Born approximation is expected to give a reasonably good value for the scattering cross section for the range of values of N_{inv} covered by the measurements.

Over the experimental range, our calculated mobility has the same dependence on N_{inv} as does the measured mobility, and is larger by about a factor of 2. Since our approximations are not expected to have very high accuracy, and since there are no adjustable parameters in the calculation, the agreement is quite good. Compensation, either in the interface scattering charges or in the bulk doping, is not unlikely, and will tend to bring the calculated and observed mobilities into closer agreement.

As the family of curves in Fig. 8 suggests, the calculated mobility for a single sub-band is not very sensitive to the bulk doping; for $N_A=2.6\times10^{16}$ cm^{-3}, only about 10% of the calculated scattering is due to ions in the depletion layer, the remaining 90% being due to the interface charges.

The mobility calculated from the scattering cross section given by Kawaji and Kawaguchi is smaller than the observed mobility at 4.2°K if their value for the number of interface scatterers is used, and thus is hard to reconcile with the experimental results.[33] They used the classical Coulomb cross section (C29) for coplanar electrons and scatterers, and cut off the angular integration at a value corresponding to an impact parameter about half of the distance between scatterers. Their method gives a mobility $\mu\sim N_{inv}^{0.3}$ and our method gives $\mu\sim N_{inv}^{1.4}$ for the parameters corresponding to their experiments. Only the latter result is consistent with the observed variation, $\mu\sim N_{inv}^{1.2\pm0.2}$.

The agreement between calculated and observed mobilities for InAs must be viewed with caution because of the narrow range of values covered by the data. Other scattering mechanisms may well become important outside this range, and extension of the data to both higher and lower values of N_{inv} is highly desirable. But we may conclude that over the range of values covered by Kawaji and Kawaguchi,[3] screened Coulomb scattering by the known interface charge can account for the measured inversion-layer mobility of InAs.

B. Si

Silicon surfaces, particularly the (100) surface of Si, have been investigated extensively through transport measurements in metal–oxide–semiconductor structures. The annealed oxides used in such structures generally show no significant evidence of trapping and typically have fixed-positive surface-charge densities N_{ss} of a few times 10^{11} cm^{-2}. Although this number is an equivalent interface charge and may actually arise from a distribution of positive charge throughout the oxide, we assume that this charge is due to Coulomb centers at the interface and thus identify N_{int} in (51) with the experimental quantity N_{ss}.

Figure 9 shows the calculated mobility versus inversion charge in (100) surfaces for the case $N_{int}=4\times10^{11}/cm^2$ with two different substrate doping levels. The factor n_s in (51) has been taken as 1; that is, no intervalley scattering is assumed. This follows from the effective-mass approach.

The flatness of the curves at low inversion charge densities reflects the importance of screening in Si. On the other hand, our approach does overestimate screening at low charge densities.[34] As with InAs, we have terminated the calculated curve for $N_A=10^{14}$ cm^{-3} where the Fermi level enters the next higher

[33] Instead of using the average of the dielectric constants of the semiconductor and the insulator, as would be appropriate for a scatterer at the interface, the calculation in Ref. 3 used the dielectric constant of the semiconductor in calculating the cross section. In addition, Kawaji and Kawaguchi appear to have integrated the differential cross section only from 0 to π, instead of from $-\pi$ to π. If the indicated changes were made, their calculated mobility would be reduced, and would be in poorer agreement with experiment both in its magnitude and in its dependence on inversion-layer carrier concentration.

[34] To show that our expressions overestimate the screening at low charge densities, we note that at absolute zero the screening constant of Eq. (25) is independent of inversion-layer charge density. Thus a single mobile electron would fully screen an arbitrarily large number of Coulomb centers if this result were correct.

electric sub-band.[10] For $N_A=10^{17}$ cm^{-3}, the crossover occurs at $N_{inv}=1.2\times10^{13}$ cm^{-2}.

The measured mobility in Si at low temperatures is more difficult to describe than that in InAs, but we shall try to summarize the results of Fang and Fowler.[4] First, at high charge densities the mobility drops fairly rapidly, approximately as N_{inv}^{-1}, and in this region it is rather insensitive to temperature below $\sim$20°K. This region we show as a heavy solid line in Fig. 9. It is clear from the divergence between this line and our calculated curves that the first Born approximation for scattering by charged Coulomb centers within a single sub-band cannot explain the measured mobility at high inversion-layer carrier concentrations. The mobility turns down even before the inversion-layer carrier concentration is high enough that the Fermi level enters a higher electric sub-band, making possible real transitions between sub-bands and a reduction in mobility.[11] This discrepancy may be due to higher-order scattering involving virtual transitions to higher-lying electric sub-bands. We are at present attempting to evaluate the magnitude of these terms.

Another possible explanation of the decrease in mobility at high inversion-layer carrier concentrations may be an additional interface-scattering mechanism unrelated to the Coulomb centers.[35] Such interface scattering would become more pronounced at higher values of N_{inv}, as the inversion layer is drawn closer to the interface. It might also affect the boundary conditions for the wave functions at the interface. The apparent absence of interface scattering in the InAs data of Fig. 8 is not inconsistent with such a model, since—as shown in Table III—the inversion layer is considerably wider in InAs than in Si.

Below about 10^{12} electrons/cm^2 the surface conductance σ_s, and thus the surface mobility $\mu_{eff}\equiv\sigma_s/N_{inv}e$, is a strong function of temperature. If one simply takes the results at 4.2°K and assumes that all the inversion-layer electrons are mobile, then one obtains the lower dotted curve of Fig. 9, implying a strongly increasing mobility with increasing electron density.

On the other hand, we have pointed out in Sec. 4 that in Si surfaces the bound states associated with the interface Coulomb centers should be important. Thus one expects at low temperatures and low-electron densities some freeze-out or binding of electrons with its associated thermal-activation behavior. However, as shown in Fig. 5, the binding energy should depend strongly on the amount of screening present.

Fang and Fowler[4] have found that for constant surface charge in this low-temperature, low-density region, the surface conductance varies approximately as $\exp(-\Delta/KT)$ over fairly wide ranges of temperature, with the activation energy Δ depending upon total charge density and decreasing rapidly with increasing charge density.

An intriguing quantitative result of this analysis is that the binding energy obtained for the lowest charge levels measured (for which the measurements actually were extended above 200°K) is about 20 meV. As seen in Fig. 6, the binding energy for a Coulomb center when $d=2a^*$ and in the limit of no screening is 0.54 Ry* or about 23 meV. Although the evidence is by no means conclusive, the closeness of these numbers suggests that the bound states of Sec. 4 do play a role in the conductance of Si inversion layers. If we assume that the exponential factor in the conductance is associated with the number of mobile electrons and extrapolate the exponential regions to $T^{-1}=0$, then the intercept conductances or mobilities, as indicated by the upper dotted curve of Fig. 9, give a slowly varying function of charge density. The proximity of the calculated mobility curves to the upper dotted curve is consistent with our model, which assumes that Coulomb scattering within a single electric sub-band is the dominant scattering mechanism in (100)-Si surfaces except at the highest values of inversion-layer carrier concentration.

It must be conceded that our interpretation of the conductance curve is in apparent contradiction to the results of the Hall-effect measurements on these surfaces,[4] which indicate that the strong variation of conductance near the conductance threshold is dominated by a mobility change rather than a change in carrier density. However impurity-band conduction associated with the inversion-layer bound states is likely, especially after some carriers are induced and screening by these carriers increases the overlap of the bound-state wave functions. In such a case the interpretation of the Hall effect becomes very difficult, and further experimental and theoretical work is required to distinguish between carrier-concentration changes and mobility changes in the surface conductance near the conductance threshold.

Another manifestation of the thermally activated domain of conductance is the observed shift of conductance threshold to higher gate voltages with decreasing temperature. This is expected within the framework of our model, with each positive interface charge having a bound state. Indeed the total shift from room temperature to, say, liquid-helium temperature, when converted to the equivalent charge, should equal the density-of-interface charges. If our identification of N_{int} with N_{ss} is correct, then the threshold shift is a measure of N_{ss}, and the threshold at low temperatures is simply the voltage required to induce a charge density equal to the integrated charge per unit area in the depletion layer. For (100)-Si-SiO_2 interfaces, this appears to be the case.[4] Other surfaces show some extra positive charge which does not trap electrons at low temperatures.

[35] Indications of such scattering were also found by E. Arnold and G. Abowitz [Appl. Phys. Letters 9, 344 (1966)] in their room-temperature measurements.

The observed[4] temperature dependence of the threshold shift does not fit the simplest model of a single unscreened binding energy of 23 meV corresponding to ions right at the interface. One must accept, rather, a distribution of ions near the interface, yielding a distribution of bound states with a maximum binding energy of 23 meV. Such a situation would require that the calculated mobility curves of Fig. 9 be revised upward.

Freeze-out effects can also account for some observed[4] structure in the field-effect mobility $\mu_{FE} = d\sigma/dQ$, where σ is the surface conductance and Q is the induced charge. This can be seen from the following considerations: If at low to moderate temperatures one begins to add charge to a surface channel, initially there is no screening, binding energies are large, and most of the charge goes into bound states. As one continues to add charge, the fraction which goes into the band begins to screen the centers. The resulting decrease in binding energy of the localized states in turn leads to an increased fraction of electrons in the band. When enough carriers have been added to the inversion layer, the screening is sufficient to bring the binding energy down to such a small value that overlap effects lead to energy levels merged with the band, and bound states no longer exist.

We have attempted to model this situation numerically for the case $N_{int}=4\times10^{11}$ cm^{-2} by using the bound-state energy $E_b=0.54\exp[-2.8(a^*\bar{s})^{0.61}]$ Ry* found for $d=2a^*$ in Sec. 4, taking the screening constant $\bar{s}$ from (25) with κ replaced by $\bar{\kappa}$. Assuming the density of states in the lowest electric sub-band to be unperturbed, one can calculate the free-carrier concentration N in this sub-band as a function of the total (free plus bound) inversion-layer electron concentration N_{inv} and the absolute temperature T.

The calculated results generally yield a peak in the derivative of the inversion-layer free-carrier concentration with respect to total (free plus bound) inversion-layer electron concentration near the onset of inversion-layer conduction. This will also produce a peak in the field-effect mobility, and such peaks are generally observed at intermediate temperatures near the onset of conduction in Si-SiO_2 inversion layers.[4]

At some temperatures our calculations indicate bistability for N, i.e., two stable values of N for a single value of N_{inv}. Such bistability would, among other things, manifest itself in an abrupt increase in the surface conductance as the inversion-layer charge increased beyond a critical value. This behavior may be related to the sharp conductance jumps found in some early experiments.[36] We intend to pursue this question with an improved calculation of the bound-state energies with varying screening, using the more realistic potential given by Eqs. (B9)–(B16).

[36] W. E. Howard and F. F. Fang, Solid State Electron. 8, 82 (1965).

Our treatment of the inversion layer near the onset of conductance is inadequate for a number of reasons. First, we have used the simple screening relation (25), which is invalid at low inversion-layer carrier concentrations. Second, we have ignored the perturbation in the density of states of the electric sub-band which results from the presence of the attractive Coulomb centers which produce the bound states. Third, we have based our analytical modeling of freeze-out on a single curve of bound-state energy versus screening constant, ignoring the change in inversion-layer thickness with inversion-layer charge, and have used the curves of Fig. 6, which are based on a rather crude model potential. Finally, we have not treated the Coulomb scattering for the case in which some of the interface centers are ionized, while others have inversion-layer electrons bound to them, and are neutral. Specifically, when the screening parameter $\bar{s}$ is small, the bound electrons are tightly bound and provide a significant amount of screening. At higher inversion-layer carrier concentrations, electrons either are no longer bound [see Eq. (44)] or are so weakly bound that they provide a negligible amount of screening. Some of these approximations can be improved relatively easily, while others require major improvements in the theory. We are continuing the analysis of the freeze-out effects, and hope to present improved calculations later.

None of the approximations we have described is likely to invalidate the qualitative consequences of our freeze-out model, and we expect a peak in the field-effect mobility at intermediate temperatures to be present in any theory which includes screening effects on the inversion-layer bound states.

7. DISCUSSION AND CONCLUSIONS

In this final section we summarize some of the limitations of the approach we use to describe the inversion layer, indicate some directions in which the present calculation should be extended, and summarize the nature of the results already obtained.

The approximations used to derive the results of the present paper include the following:

(a) The effective-mass approximation, which may not be valid in the inversion layer because the potential there varies considerably in a distance of the order of 10 atomic spacings.

(b) The assumption that a charged center near the surface causes a weak perturbation of the average smooth surface potential. This assumption fails in the immediate neighborhood of the center, except when the center is in the insulator, and is also rather poor when the concentration of the centers and of the inversion-layer carriers is small.

(c) The linearized treatment of screening employed to construct the effective potential, which ignores the

oscillatory terms that would arise from a more exact treatment,[13] and which is likely to fail altogether at low inversion-layer carrier concentrations.

(d) The first Born approximation for treating the scattering, which—like our screening model—fails at low inversion-layer carrier concentrations. In addition, the higher-order scattering terms which we have ignored may become important when the Fermi level approaches the bottom of the first excited electric sub-band.

There are a number of directions in which our present results should be extended. First, the higher-order scattering terms mentioned in Sec. 6 and in (d) above should be evaluated. Second, a more realistic potential than the model potential of (B19) should be constructed to calculate the energies of bound states in the inversion layer for arbitrary screening constants. This might make possible a more quantitative treatment of the freeze-out effects discussed in Sec. 6. And third, the calculation should be extended to higher temperatures, both by the including lattice scattering[37,38] and by taking into account the population of more than one electric sub-band. When many electric sub-bands are occupied, it should be possible to relate the calculation to a conventional surface-scattering calculation.[6,39] We are at present working on the first two of these extensions.

In this paper we have studied some properties of thin semiconductor inversion layers under conditions such that they are effectively two dimensional. In particular, we have investigated the additional potential associated with charged centers near the inversion layer, including the contribution of screening to the potential. We have found some properties of bound states in the inversion layer associated with attractive charged centers, and have calculated the effect of charged scattering centers on the mobility of inversion-layer electrons at low temperature. The results appear to explain the low-temperature mobility in InAs inversion layers and some of the properties of (100)-Si inversion layers. Our approach has the advantage of simplicity, and can be applied to other inversion layers as experimental results become available.

Because of its location just under the surface of a semiconductor, the inversion layer provides a very useful tool for probing the surface. We believe that further study of semiconductor inversion layers will lead to new quantitative information about semiconductor surfaces and interfaces.

[37] V. Ya. Demikhovskii and B. A. Tavger, Fiz. Tver. Tela **6**, 960 (1964) [English transl.: Soviet Phys.—Solid State **6**, 743 (1964)].

[38] L. V. Iogansen, Zh. Eksperim. i Teor. Fiz. **50**, 709 (1966) [English transl.: Soviet Phys.—JETP **23**, 470 (1966)].

[39] R. F. Greene and R. W. O'Donnell, Phys. Rev. **147**, 599 (1966).

ACKNOWLEDGMENTS

We are indebted to R. W. Keyes, G. J. Lasher, P. J. Price, T. D. Schultz, and particularly A. B. Fowler and F. F. Fang, for many valuable comments and discussions. We are also indebted to T. N. Morgan for suggestions on the solution of Poisson's equation for the inversion layer, and to P. M. Marcus for providing the transformation from (B18) to (B19).

APPENDIX A: LANDAU LEVELS

We give here an approximate treatment of the Landau levels associated with a single-ellipsoidal energy surface in the bulk when both a strong electric field perpendicular to the surface and a magnetic induction of arbitrary orientation are present. Electron spin is not considered.

If the magnetic induction has components (B_x,B_y,B_z) in the coordinate system used in Sec. 2, in which the z direction is normal to the surface, then the Hamiltonian is

$$H=\tfrac{1}{2}\sum_{ij} w_{ij}(p_i+eA_i/c)(p_j+eA_j/c)+V(z), \quad \text{(A1)}$$

where $\mathbf{A}$ is the vector potential and $V(z)$ is the surface potential well. We use the gauge

$$\mathbf{A}=(B_y z-B_z y,\ -B_x z,\ 0). \quad \text{(A2)}$$

For simplicity we introduce $b_i=eB_i/c$.

The Hamiltonian (A1) with the gauge (A2) does not contain x, so that we may choose a wave function

$$\psi=\exp(ik_x x)f(y,z). \quad \text{(A3)}$$

The Hamiltonian acting on f then becomes

$$H=\tfrac{1}{2}w_{33}p_z^2+V(z)+w_{23}(p_y-b_x z)p_z +\tfrac{1}{2}w_{11}(\hbar k_x-b_z y+b_y z)^2+\tfrac{1}{2}w_{22}(p_y-b_x z)^2 +\tfrac{1}{2}iw_{23}b_x\hbar. \quad \text{(A4)}$$

We use coordinates such that $w_{12}=w_{13}=0$, which we can do without loss of generality provided the constant-energy ellipsoids of the semiconductor are ellipsoids of revolution.

We first find an approximate eigenfunction associated with the first three terms of (A4), which are the dominant terms, and then replace the remaining three terms by their expectation value. This treatment will only be valid if the Landau level spacing is small compared to the separation between the electric sub-bands, as is the case experimentally.[2,40]

The approximate eigenfunction associated with the first three terms of (A4) is

$$\exp[-iw_{23}(p_y z-\tfrac{1}{2}b_x z^2)/w_{33}\hbar]\zeta(z), \quad \text{(A5)}$$

[40] F. F. Fang and P. J. Stiles, Bull. Am. Phys. Soc. **12**, 275 (1967); also (to be published). We are indebted to Dr. Fang and Dr. Stiles for discussions of their work.

where the exponential factor is introduced to eliminate the term linear in p_z in (A4), and $\zeta(z)$ is the solution of (12) belonging to the lowest eigenvalue E''. The expectation value of (A4) for the wave function (A5) is

$$E''+\tfrac{1}{2}(w_{22}-w_{23}^2/w_{33})(p_y-b_xz)^2{}_{\rm av}$$
$$+\tfrac{1}{2}w_{11}[\hbar k_x-b_zy-(w_{23}/w_{33})b_zz+b_yz]^2{}_{\rm av}, \quad \text{(A6)}$$

where we used $\exp(iap_y/\hbar)f(y)=f(y+a)\exp(iap_y/\hbar)$.

Equation (A6) can be simplified if we introduce the new variable $y'=y+[(w_{23}/w_{33})-(b_y/b_z)]z_{\rm av}-\hbar k_x/b_z$, and if we introduce into the approximate eigenfunction a factor $\exp(ib_xz_{\rm av}y'/\hbar)$ to eliminate the term linear in p_y. The expectation value of (A4) then is

$$\tfrac{1}{2}m_1{}^{-1}b_z{}^2y'^2+\tfrac{1}{2}m_2{}^{-1}p_{y'}{}^2+\{\tfrac{1}{2}m_1{}^{-1}[b_y-(w_{23}/w_{33})b_z]^2$$
$$+\tfrac{1}{2}m_2{}^{-1}b_x{}^2\}(z^2{}_{\rm av}-z^2{}_{\rm av})+E'', \quad \text{(A7)}$$

where m_1 and m_2 are the principal effective masses of the constant-energy ellipse, as in Sec. 2.

The first two terms of (A7) give the harmonic oscillator Hamiltonian for the Landau levels, and the two terms in braces are small correction terms. The approximate eigenvalue of the ith Landau level relative to the energy of the bottom of the lowest electric sub-band in the absence of a magnetic field is

$$E=(n+\tfrac{1}{2})(e|B_z|\hbar/m_1{}^{1/2}m_2{}^{1/2}c)+\tfrac{1}{2}(e^2/c^2)(z^2{}_{\rm av}-z_{\rm av}{}^2)$$
$$\times\{[B_y-(w_{23}/w_{33})B_z]^2/m_1+B_x{}^2/m_2\}. \quad \text{(A8)}$$

The principal results embodied in (A8) are that the Landau-level spacing depends only on the component of the field perpendicular to the surface, and that the cyclotron mass which enters in the level spacing is $(m_1m_2)^{1/2}$, which is the density-of-states mass. These results are consistent with the interpretation of the oscillatory magnetoresistance experiments.[2,40]

Even when the magnetic induction parallel to the surface is 10^5 G, the correction terms in (A8) are only of the order of 0.1 meV if $m_1=m_2=0.2m$. These terms do not affect the level spacing, and they do not appear to be easily observable; however, for some surfaces with degenerate energy ellipses, a tilted magnetic field can lift the degeneracy. We do not expect our neglect of higher sub-bands to lead to significant errors until high enough magnetic quantum numbers n are reached to place the energy (A8) near the beginning of the second electric sub-band.

Not considered here is the effect of slight energy splitting between levels assumed to be degenerate in our work, like the doubly degenerate lowest electric sub-band of (100)-Si surfaces, as shown in Fig. 3. Such splittings cannot be treated within the framework used in this paper.

The Landau levels for the special case of a magnetic field normal to the surface and ellipsoid orientations such that the Hamiltonian (A1) contains only terms with $i=j$ have been given by Duke,[41] and are consistent with the results given here.

APPENDIX B: SCREENED COULOMB POTENTIAL

We want to find the additional potential produced by a charge in the oxide, in the semiconductor, or at the interface between them, for a structure like that of Fig. 1. For a charge Ze located at $r\equiv(x^2+y^2)^{1/2}=0$ and $z=z_0$, the additional Coulomb potential (in the quantum limit, when only the lowest electric sub-band is occupied) must satisfy

$$\nabla^2\phi(r,z)-2s\bar{\phi}(r)g(z)=-4\pi Ze\delta(x)\delta(y)\delta(z-z_0)/\kappa, \quad \text{(B1)}$$

as shown in Sec. 3. Here s is the screening constant given in (25);

$$g(z)=\tfrac{1}{2}b^3z^2\exp(-bz) \quad \text{(B2)}$$

gives the approximate distribution of inversion-layer charge, as in (18) and (21);

$$\bar{\phi}(r)=\int_0^\infty \phi(r,z)g(z)dz, \quad \text{(B3)}$$

as in (23); and κ is the dielectric constant, which equals $\kappa_{\rm sc}$ in the semiconductor ($z>0$) and equals $\kappa_{\rm ins}$ in the insulator.

To solve (B1) we take advantage of the cylindrical symmetry to write

$$\phi(r,z)=\int_0^\infty J_0(k'r)A_{k'}(z)k'dk'. \quad \text{(B4)}$$

The potential will satisfy (B1) if

$$d^2A_k/dz^2-k^2A_k-2s\bar{A}(k)g(z)=-2Ze\delta(z-z_0)/\kappa, \quad \text{(B5)}$$

where

$$\bar{A}(k)=\int_0^\infty A_k(z)g(z)dz. \quad \text{(B6)}$$

Equation (B5) is easily found from (B1) if we multiply the latter by $rJ_0(kr)$ and integrate over r, noting that[41]

$$\int_0^\infty J_\nu(kr)J_\nu(k'r)rdr=k^{-1}\delta(k-k'). \quad \text{(B7)}$$

Both A_k and $\kappa(dA_k/dz)$ must be continuous across the discontinuity in κ at $z=0$.

The impurity potential ϕ will go to zero at the metallic contact to the insulator, and in the semiconductor bulk, because of screening by free carriers. Such screening effects are absent in the insulator and in the semiconductor depletion layer. Since both the insulator thickness and the depletion-layer thickness are considerably greater than the surface screening

[41] C. B. Duke, Phys. Letters 24A, 461 (1967).

length s^{-1} and the carrier wavelength, which are the characteristic lengths of our problem, we can with negligible error use the approximate boundary conditions

$$A_k(z)\to 0 \quad \text{as} \quad |z|\to\infty . \tag{B8}$$

The solution for $A_k(z)$ under these conditions is straightforward, and will not be given here. Only $\bar{A}(k)$ enters in any of the calculations for which we need the potential, so we quote only the results for this average over the inversion-layer charge distribution. Charges located in the insulator or at the interface, or in the semiconductor, respectively, we find

$$\bar{A}(k,z_0)=(Ze/\bar{\kappa})P_0\exp(kz_0)/D, \quad z_0\leq 0, \tag{B9}$$

$$\bar{A}(k,z_0)=(Ze/\kappa_{sc})[P(z_0)+\delta P_0\exp(-kz_0)]/D, \quad z_0>0, \tag{B10}$$

where

$$D=k+sP_{av}+s\delta P_0^2, \tag{B11}$$

$$\delta=(\kappa_{sc}-\kappa_{ins})/(\kappa_{sc}+\kappa_{ins})=(\kappa_{sc}-\kappa_{ins})/2\bar{\kappa}, \tag{B12}$$

$$P_0=b^3(b+k)^{-3}, \tag{B13}$$

$$P_{av}=(8b^3+9b^2k+3bk^2)(b+k)^{-3}/8, \tag{B14}$$

$$P(z)=b^3(b-k)^{-3}[\exp(-kz)-\exp(-bz)(a_0+a_1z+a_2z^2)], \quad k\neq b,$$
$$a_0=2k(3b^2+k^2)(b+k)^{-3},$$
$$a_1=4bk(b-k)(b+k)^{-2},$$
$$a_2=k(b-k)^2(b+k)^{-1}, \tag{B15}$$

$$P(z)=(1/8)[1+2bz+2b^2z^2+(4/3)b^3z^3]\exp(-bz), \quad k=b. \tag{B16}$$

We find from these results, that

$$\int_0^\infty r\bar{\phi}(r)dr=Ze/\bar{\kappa}\bar{s}, \tag{B17}$$

where $\bar{s}=s\kappa_{sc}/\bar{\kappa}$. This integral enters in Eq. (37) in considering bound states. Note that (B17) is independent of the location of the Coulomb center and of the thickness of the inversion layer. This is a consequence of (B7), which shows that only the value of $\bar{A}(k,z_0)$ for $k=0$ enters, and of (B9)–(B16), which show that $\bar{A}(0,z_0)$ is independent of z_0 and of b.

The expressions (B9)–(B16) simplify considerably when $g(z)=\delta(z)$, i.e., when the inversion-layer thickness goes to zero. We reach this case in the limit $b\to\infty$, for which we find, for any impurity location z_0,

$$\bar{\phi}(r)=(Ze/\bar{\kappa})\int_0^\infty k(k+\bar{s})^{-1}J_0(kr)\exp(-kd)dk, \tag{B18}$$

where $d=|z_0|$ and $\bar{s}$ is the screening constant calculated from (25) using a dielectric constant equal to $\bar{\kappa}$. In the absence of screening this gives $\bar{\phi}(r)=Ze/\bar{\kappa}(r^2+d^2)^{1/2}$, as expected for that case from simpler considerations.[16]

For numerical purposes it is useful to transform (B18) to[42]

$$(\bar{\kappa}/Ze)\bar{\phi}(r)$$
$$=\left\{(d^2+r^2)^{-1/2}-\bar{s}\exp(\bar{s}d)\left[(\pi/2)\{\mathbf{H}_0(r\bar{s})-Y_0(r\bar{s})\}-\int_0^d(z^2+r^2)^{-1/2}\exp(-\bar{s}z)dz\right]\right\}, \tag{B19}$$
$$=\left\{(d^2+r^2)^{-1/2}-\bar{s}\exp(\bar{s}d)\times\int_d^\infty(z^2+r^2)^{-1/2}\exp(-\bar{s}z)dz\right\},$$

in which $\mathbf{H}(x)$ and $Y(x)$ are the Struve function and the Bessel function of the second kind, respectively. They have been tabulated by Watson,[43] and approximate numerical formulas for $\mathbf{H}_0-Y_0$ have been given by Luke,[44] but numerical values are easily obtained directly from (B19). Values of $\bar{\kappa}r\bar{\phi}(r)/Ze$ for various values of $\bar{s}d$ are shown in Fig. 4.

Asymptotic expressions for the potential (B18) can be given in a number of cases:

$$(\bar{\kappa}/Ze\bar{s})\bar{\phi}(r)$$
$$\sim R^{-3}\left[\zeta+\frac{\rho^2-2\zeta^2}{R^2}+\frac{\zeta(6\zeta^2-9\rho^2)}{R^4}+\cdots\right], \quad R\gg 1, \tag{B20a}$$

$$\sim(1+\zeta)[\rho^{-1}+\ln(C\rho)], \quad \zeta\ll\rho\ll 1, \tag{B20b}$$

$$\sim R^{-1}+(1+\zeta)\ln(2\zeta/C), \quad \rho\ll\zeta\ll 1, \tag{B20c}$$

where $\rho=\bar{s}r$, $\zeta=\bar{s}d$, $R^2=\rho^2+\zeta^2$, $C=\frac{1}{2}\exp(\gamma)=0.891$, and γ is Euler's constant. At large distances $\bar{\phi}$ falls off as r^{-3}, apart from oscillatory terms associated with a sharp Fermi surface,[13] while in three dimensions the corresponding screened Coulomb potential decays exponentially.

Finally, we give the coefficients which enter in (51) in the Born-approximation calculation of the scattering rate associated with N_{int} singly charged scatterers per cm^2 at the interface, and N_{ins} and N_{sc} singly charged scatterers per cm^3 in the insulator and the semiconductor, respectively. They are

$$B_{int}=|\bar{A}(S,0)|^2=(e/\bar{\kappa})^2P_0^2/D^2, \tag{B21a}$$

$$B_{ins}=\int_{-\infty}^0|\bar{A}(S,z_0)|^2dz_0=B_{int}/2S, \tag{B21b}$$

[42] See Chapter 14 of Ref. 32.
[43] See Table I of Ref. 32.
[44] Y. L. Luke, *Integrals of Bessel Functions* (McGraw-Hill Book Company, Inc., New York, 1962), pp. 80, 92; Y. L. Luke, J. Soc. Ind. Appl. Math. 3, 179 (1955).

$$B_{sc}=\int_0^\infty |\bar{A}(S,z_0)|^2 dz_0 = (e/\kappa_{sc})^2(C_0+C_1\delta P_0+\delta^2)/2SD^2, \quad \text{(B21c)}$$

$$C_0=(2b^6+24b^5S+48b^4S^2+43b^3S^3+18b^2S^4+3bS^5)/2(b+S)^6, \quad \text{(B21d)}$$

$$C_1=(2b^4+14b^3S)/(b+S)^4, \quad \text{(B21e)}$$

where $S=2k\sin\frac{1}{2}\theta$ is the momentum transfer in the scattering process, and the remaining symbols have the same meaning as in (B9)–(B16).

APPENDIX C: TWO-DIMENSIONAL SCATTERING

The conventional treatment of elastic scattering in three dimensions[45] must be revised for two-dimensional scattering. We give here a number of results for potentials with circular symmetry.

At large distances from a scattering center the wave function can be written as a plane wave plus an outgoing wave:

$$\psi(r,\theta)\sim\exp(ikx)+r^{-1/2}f(\theta)\exp(ikr), \quad \text{(C1)}$$

where $r=(x^2+y^2)^{1/2}$ is the radial coordinate in two dimensions and $x=r\cos\theta$. The scattering cross section is

$$\sigma(\theta)=|f(\theta)|^2. \quad \text{(C2)}$$

If the solution of Schrödinger's equation is expanded in a Fourier series

$$\psi(r,\theta)=\sum_{m=-\infty}^{\infty} R_m(r)\exp(im\theta) = \sum_{-\infty}^{\infty} r^{-1/2}P_m(r)\exp(im\theta), \quad \text{(C3)}$$

then $P_m(r)$ satisfies

$$P_m''+[(\tfrac{1}{4}-m^2)r^{-2}+k^2-U(r)]P_m(r)=0, \quad \text{(C4)}$$

where $k^2=2m^*E/\hbar^2$ and $U(r)=2m^*V(r)/\hbar^2$.

The expansion of a plane wave in two dimensions is

$$\exp(ikx)=\sum_{m=-\infty}^{\infty} i^m J_m(kr)\exp(im\theta), \quad \text{(C5)}$$

and for large kr

$$J_m(kr)\sim(2/\pi kr)^{1/2}\cos(kr-\tfrac{1}{2}m\pi-\tfrac{1}{4}\pi). \quad \text{(C6)}$$

At large distances from the scattering center the phase is shifted by η_m, and we can write

$$R_m(r)\sim A_m(2/\pi kr)^{1/2}\cos(kr-\tfrac{1}{2}m\pi-\tfrac{1}{4}\pi+\eta_m). \quad \text{(C7)}$$

By comparing (C3) and (C1), and using (C5)–(C7), we find

$$A_m=i^m\exp(i\eta_m)$$

and

$$f(\theta)=(2i/\pi k)^{1/2}\sum_{m=-\infty}^{\infty}\exp(i\eta_m)\sin\eta_m\exp(im\theta), \quad \text{(C8)}$$

from which the differential cross section is given by (C2). The total cross section is obtained by integrating over $\sigma(\theta)$ from $-\pi$ to π, giving

$$\sigma=(4/k)\sum_{m=-\infty}^{\infty}\sin^2\eta_m. \quad \text{(C9)}$$

Comparing (C8) and (C9), we find that the two-dimensional form of the optical theorem[46] is

$$\sigma=(8\pi/k)^{1/2}\,\mathrm{Im}[f(0)\exp(-\tfrac{1}{4}i\pi)]. \quad \text{(C10)}$$

The momentum cross section, needed for transport calculations, is

$$\sigma_m=\int_{-\pi}^{\pi}\sigma(\theta)[1-\cos\theta]d\theta = (4/k)\sum_{m=0}^{\infty}\sin^2(\eta_m-\eta_{m+1}). \quad \text{(C11)}$$

The phase shifts can be calculated in Born approximation. We put

$$R_m(r)=A_m'[J_m(kr)+\chi_m(r)]. \quad \text{(C12)}$$

If the potential is treated as a perturbation, we can write

$$\chi_m(r)=\int_0^\infty G_m(r,r')U(r')J_m(kr')r'dr', \quad \text{(C13)}$$

where the Green's function

$$G_m(r,r')=\tfrac{1}{2}\pi J_m(kr_<)Y_m(kr_>) \quad \text{(C14)}$$

is the solution of

$$d^2G_m/dr^2+r^{-1}dG_m/dr+(k^2-m^2r^{-2})G_m(r,r') = r'^{-1}\delta(r-r'), \quad \text{(C15)}$$

and $r_<$ and $r_>$ are the smaller and larger, respectively, of r and r'. We compare (C12) with (C7) and find that the phase shift is given in Born approximation by

$$\tan\eta_m=-\tfrac{1}{2}\pi\int_0^\infty J_m^2(kr)U(r)rdr. \quad \text{(C16)}$$

[45] See, for example, Chapters V and VII of Ref. 17, or Chapter XIV of Ref. 18.

[46] See, for example, A. Messiah, *Quantum Mechanics* (North-Holland Publishing Company, Amsterdam, 1962), Vol. II. Chap. XIX, Sec. 31; or Eq. (19.14) of Ref. 17; or Eq. (117.7) of Ref. 18. The extra phase factor in Eq. (B10) arises because of the asymptotic form of $J_0(x)$, Eq. (B6).

If we use the relation[47]

$$J_0(Sr)=\sum_{m=-\infty}^{\infty} J_m^2(kr)\exp(im\theta)\,, \qquad \text{(C17)}$$

where

$$S=|\mathbf{k}-\mathbf{k}'|=2k\sin\tfrac{1}{2}\theta \qquad \text{(C18)}$$

is the change in wave vector upon scattering, $|\mathbf{k}|=|\mathbf{k}'|=k$, and $\cos\theta=\mathbf{k}\cdot\mathbf{k}'/k^2$, and if we assume that all phase shifts are small, we can substitute in (C8) and (C2) to obtain

$$\sigma(\theta)=(\pi/2k)\left|\int_0^\infty U(r)J_0(Sr)rdr\right|^2. \qquad \text{(C19)}$$

This is the Born-approximation result for the scattering cross section, and can be obtained more directly without introducing the phase shifts.[48]

An exact solution of the scattering problem is possible for the Coulomb potential energy $V(r)=-Ze^2/\bar{\kappa}r$.[49] The Schrödinger equation then is

$$\nabla^2\psi+(k^2+2kGr^{-1})\psi=0\,, \qquad \text{(C20)}$$

where ∇^2 is the two-dimensional Laplacian, and

$$G=m^*Ze^2/k\bar{\kappa}\hbar^2=Z/a^*k\,, \qquad \text{(C21)}$$

where a^* is the effective Bohr radius, defined in (36). If we introduce $\xi=r-x$ as a coordinate, we find that a wave function of the form $h(\xi)\exp(ikx)$ satisfies (C20) if

$$h(\xi)=F(iG,\tfrac{1}{2},ik\xi)\,, \qquad \text{(C22)}$$

where $F(a,b,z)$ is the confluent hypergeometric function.[49] The asymptotic form of ψ for large ξ is

$$\psi(r,\theta)\sim\exp\{ikx-iG\ln[kr(1-\cos\theta)]\}+f(\theta)r^{-1/2}\exp[ikr+iG\ln(2kr)]\,, \qquad \text{(C23)}$$

where

$$f(\theta)=\frac{\Gamma(\frac{1}{2}-iG)\exp(2iG\ln\sin\frac{1}{2}\theta)}{\Gamma(iG)(2ik)^{1/2}\sin\frac{1}{2}\theta}\,. \qquad \text{(C24)}$$

[47] See Sec. 11.2 of Ref. 32.

[48] See, for example, Sec. 29 of Ref. 17.

[49] See, for example, Sec. 20 of Ref. 17, or Sec. 112 of Ref. 18 for the three-dimensional calculation, which we follow.

If we note that[50]

$$\Gamma(\tfrac{1}{2}+iG)\Gamma(\tfrac{1}{2}-iG)=\pi/\cosh(\pi G)\,, \qquad \text{(C25)}$$

$$\Gamma(iG)\Gamma(-iG)=\pi/G\sinh(\pi G)\,, \qquad \text{(C26)}$$

we find the differential scattering cross section for two-dimensional Coulomb scattering to be

$$\sigma(\theta)=G\tanh(\pi G)/2k\sin^2\tfrac{1}{2}\theta\,. \qquad \text{(C27)}$$

In the limit of large carrier velocity we can replace the hyperbolic tangent by its argument and obtain

$$\sigma(\theta)\sim(\pi/2k)(Ze^2/\bar{\kappa}\hbar v)^2/\sin^2\tfrac{1}{2}\theta,\quad \pi G\ll 1\,, \qquad \text{(C28)}$$

which is exactly the result obtained for this case from the Born approximation. This can be seen if we let the distance z_0 and the screening constant $\bar{s}$ go to zero and the inversion layer width constant b go to infinity in (B9), and substitute this result in (49), using $n_s=1$.

In the limit of small carrier velocity the argument of the hyperbolic tangent in (C27) goes to infinity, and the cross section becomes

$$\sigma(\theta)\sim(|Z|e^2/2\bar{\kappa}m^*v^2)/\sin^2\tfrac{1}{2}\theta,\quad \pi G\gg 1\,. \qquad \text{(C29)}$$

This is the classical differential cross section,[3] given by $|db/d\theta|$, where $b=(Ze^2/\bar{\kappa}m^*v^2)\cot\frac{1}{2}\theta$ is the classical impact parameter.

Two-dimensional Coulomb scattering thus provides a case in which the Born, exact, and classical cross sections are all different, while in three dimensions they happen to be the same. The conditions $G\ll 1$, or $G\gg 1$, are conditions for the validity of the Born and the classical approximations, respectively,[51] as is nicely confirmed by our example. It is interesting that the exact cross section for unscreened Coulomb scattering is independent of the sign of Z. This is not true for a general potential, as shown for example in Fig. 7.

To estimate the value of πG, we suppose that there are 3×10^{11} inversion-layer electrons per cm^2 in InAs and 10^{12} cm^{-2} electrons in the (100) surface of Si, and use the parameters of Table II. Then $\pi G=1.3$ for InAs, and $\pi G=8$ for Si, which would make the classical cross section slightly better in the former case, and essentially exact in the latter case if the unscreened Coulomb potential were valid.

[50] E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, Inc., New York, 1945), 4th ed., p. 11.

[51] See Eqs. (45.7) and (49.10) in Ref. 18.