Effect of Electric and Magnetic Fields on the Self-Consistent Potential at the Surface of a Degenerate Semiconductor

G. A. Baraff and Joel A. Appelbaum
Bell Telephone Laboratories, Murray Hill, New Jersey 07974
(Received 14 July 1971)

We have performed a self-consistent calculation of the potential in the accumulation layer at the surface of a degenerate semiconductor and of the bound-state energies which that potential supports in the presence of a perpendicular electric field of appropriate sign and strength. The calculation is carried out in the Hartree approximation appropriate to the small r_s values found in moderately doped, narrow-band-gap materials, using a parametric scheme which enormously simplifies the calculation. This makes it feasible to study the effect of a quantizing magnetic field on the self-consistent potential and the bound-state energies. Bound states and mobile states are determined from the same Schrödinger equation. We derive and make extensive use of a one-dimensional Friedel sum rule, and show that there is a sudden rearrangement of mobile charge which compensates for what would otherwise be a discontinuous charge alteration when the number of bound states changes in response to changes in the external fields. Simple models are presented which reproduce the results of the self-consistent calculation and which are useful in interpreting the meaning of the detailed numerical results. Magneto-oscillations of the potential caused by a perpendicular magnetic field are studied in detail. The observable magneto-oscillations in the potential here arise from Landau-level quantization of the bound states only. We compare the results with earlier calculations in which the mobile states were treated using linear-response theory and we find that the role played by the Landau-level quantization of the mobile electron states is, in this nonlinear treatment, quite unimportant, supporting the conclusion which we drew from the linear surface-screening approximation of our earlier work.

I. INTRODUCTION

In this paper, as in the earlier ones in this series,[1,2] we study the self-consistent potential which exists near the surface of a degenerate semiconductor. Three broad questions present themselves. First, what is the shape of this potential in the absence of any external fields, and how does that shape depend on the doping of the semiconductor? Second, what is the effect of applying an electric field normal to the surface of the semiconductor? Third, how does a magnetic field applied normal to the surface affect the screening just described?

This program of investigation is closely related to the one which we carried out in Ref. 1(a), and indeed, the physical model is identical with the one we used there. In this model, the conduction-band electrons are the entities which rearrange themselves to provide the screening and they are treated within the simplest effective-mass formalism.[3] Recall, however, that in that earlier treatment we made the approximation of treating the mobile electrons (conduction electrons which are not bound in the self-consistent potential well near the surface) via linear-response theory. In retrospect, that earlier effort was productive in that it provided some insight into the role played by the magnetic field in altering the screening. The insight was valid primarily because in the cases we studied most of the screening was provided by bound electrons[4] (conduction electrons bound in the self-consistent potential well near the surface) and these bound electrons were treated exactly. Hence, we achieved a reasonable description of the potential well but we learned rather little about the true role the mobile electrons play in determining it. We undertake the present investigation with the objective of trying to understand the role of the mobile charge more fully.

In order to achieve this objective, it is necessary for us to treat the bound charge and the mobile charge on an exactly equivalent footing, which means a fully nonlinear treatment for both. This allows us to handle both strong and weak electric fields, i.e., to study in detail the departures from any of the linear screening theories. This we will do. One of the surprises we encounter here is that there is virtually no attractive electric field which gives rise to a weak potential. Even at zero electric field, it will turn out that at small values of r_s (the interelectronic spacing expressed in units of the effective Bohr radius) the self-consistent potential is of such depth and range as to support a weakly bound state. Hence, the zero-field wave functions are severely distorted from the sine waves on which linear screening theories are based. A second surprise is that in spite of the severe distortion of the mobile-state wave functions and the existence of a bound state which the

linearized theory ignores, i.e., in spite of its totally erroneous conception of the zeroth-order situation, the linear theory does provide a reasonably good accounting of the low electric field situation. Yet a third surprise is that it is possible to devise crude and easily manageable models which seem to provide insight into what is taking place. These models may ultimately prove to be the most useful feature of what we have done, in that the concepts they suggest may prove to be more generally applicable.

There is clearly much material to be covered here, and the paper, unfortunately, is a long one. Before going on to describe the contents of the remaining sections, it is perhaps in order to say a few words about how this calculation was performed. In brief, we are carrying out a Hartree self-consistent field calculation, but instead of striving for full self-consistency (which would be prohibitively costly for the number of cases we want to explore), we settle for a parametrized self-consistency. That is, instead of using the potential obtained from Poisson's equation for the input to Schrödinger's equation, we use a parametrized form which approximates the potential and solve for its eigenfunctions and eigenvalues. From these eigenfunctions a charge density is constructed and used in Poisson's equation to produce an iterated potential which is again approximated by the parametric form. The parameters of this form must be adjusted so that the new parameters and the old parameters agree. This condition expresses such self-consistency as can be achieved from a scheme in which the form of the potential is specified beforehand.[5]

This sort of procedure could be used with virtually any parametric form for the potential. The special form we choose has the advantage that its eigenfunctions are sums of exponentials. From this, it follows that all the spatial integrals which go into calculating the potential can be carried out in closed form, so that the only integrals we evaluate numerically are one-dimensional integrals over the continuum of mobile states. It is this feature which simplifies the problem enough to make it feasible to vary the electric field, add magnetic field, and study the concentration dependence.

Having described the program, we outline briefly the content of the following sections of this paper. In Sec. II, we present the model, describing why a Hartree calculation should be valid for the narrow-band-gap semiconductors to which this investigation is intended to apply. In Sec. III, we derive the Friedel-like[6] sum rule which plays a central role in our analysis. Section IV describes the running-wave (rather than the standing-wave) form for the mobile states. In Sec. V, we present the parametric method used for carrying out the calculation, exhibit the extent of self-consistency which this method achieves, and derive the working formulas which are needed. In Sec. IV, we show that the result of the Kohn-Majumdar[7] theorem, namely, the continuity of the properties of a fermion system in an external potential as the number of bound state changes, also applies to our self-consistent calculation. The importance of this theorem is that it probably provides part of the underlying reason for the success of the crude models mentioned above. Section VII contains the results and discussion of our numerical calculations. The crude models form a part of that discussion. We conclude by comparing the results obtained in Ref. 1(a) with those obtained here.

II. MODEL

We start by considering a uniformly doped, degenerate n-type semiconductor. The undoped crystal (ion cores plus filled valence bands) is to be regarded as a passive, neutral medium whose only electrical effect is to provide a background dielectric constant ϵ. We ignore true surface states here (states induced by the truncation of the crystal potential), or, alternatively, we lump any charge they may contain with that giving rise to the external field.

Suppose that there is a uniform electric field E outside the crystal and normal to its surface. It will be reduced in strength to E/ϵ inside the crystal. When we specify boundary conditions on the potential, we shall mean that it is the interior field which is fixed, and that the field outside the crystal is larger by a factor ϵ than that used in the boundary condition.

The dynamic system we are interested in, then, is the sea of degenerate conduction electrons which are free to move under the influence of various forces, and the fixed, uniform background of ionized donors, which are not. The forces acting on a conduction electron are basically of two types. The first is the electrostatic force arising from the charge density of other conduction electrons, the uniform background of ionized donors, and the externally imposed uniform fields. The second type is the exchange and correlation force. This latter may be regarded as arising from two sources. One is exchange and correlation with other conduction electrons. The strength of this force, relative to the electrostatic force, depends on r_s. Although r_s is generally large for a low density of electrons, the combination of small effective mass and large dielectric constant found in the narrow direct-gap semiconductors leads to small values of r_s at even very small conduction-electron densities. It is to these materials that we expect our work here to apply. Typical values of r_s might range from 0.1 to 0.5 (InAs with a conduction-electron Fermi en-

ergy of 0.4–0.02 eV). Under these conditions, exchange and correlation forces between conduction electrons themselves can be ignored relative to the Coulomb forces between them.

Exchange and correlation forces also arise between a conduction electron and the valence electrons. It is this which gives rise to the contact potential, and here, typically, we have potentials which change by 5 eV over distances of a few angstroms. The appropriate energy unit for the problem is the Fermi energy of the conduction electrons and the appropriate length unit is their inverse Fermi wave number. On these scales, the potential arising from exchange and correlation between a conduction electron and all the valence electrons can be well represented as an infinite potential step at $z=0$, the surface of the semiconductor. We are still left with the problem of determining the Coulomb potential self-consistently, but we are spared the nasty feature of calculating self-consistent exchange and correlation, a feature which enormously complicates similar investigations in metals.[8]

The system of equations which describe the model is the following: Schrödinger's equation for the states of the conduction electrons, a constitutive equation giving the charge density in terms of the occupied states and the uniform-background charge density, and Poisson's equation giving the potential in terms of the charge density.

We shall restrict our attention to the accumulation-layer situation, and in Fig. 1 we have sketched the potential energy an electron might experience in this case. It is convenient to let the potential be zero at large z, in the depths of the semiconductor, so that its value at the surface, at $z=0+$, is negative. The Schrödinger equation for this situation is separable. The energy associated with the electron's motion in the z direction may be negative, in which case we speak of it as a bound electron, or it may be positive, in which case we speak of it as a mobile electron.

It is convenient to introduce a magnetic field in the z direction at the outset and to use the appropriate states associated with it, since the transition to zero magnetic field is easy enough to accomplish. Accordingly, the equations of our model are those presented in Secs. IIA–IID.

A. Schrödinger's Equation for Bound States

$$\left(\frac{(\vec{p}+e\vec{A}/c)^2}{2m}+V(z)\right)\psi(\vec{r})=E\psi(\vec{r})\,, \tag{2.1a}$$

$$\psi(z=0)=0\,, \tag{2.1b}$$

$$\psi(z=\infty)=0\,, \tag{2.1c}$$

where $\vec{A}$ is the vector potential for a uniform magnetic field in the z direction, namely, $\vec{A}=(0, Hx, 0)$. The states ψ have the form

$$\psi_{n,m,k}(r)=(L_y)^{-1/2}\,e^{iky}\,\varphi_n(x-x_0)\,\chi_m(z)\,, \tag{2.2}$$

where L_y is a normalization length in the y direction, $\varphi_n(x)$ is the nth normalized harmonic-oscillator wave function

$$\left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}+\frac{m\omega_c^2}{2}\right)\varphi_n(x)=(n+\tfrac{1}{2})\hbar\,\omega_c\,\varphi_n(x)\,, \tag{2.3a}$$

$$\int dx\,\varphi_n(x)\,\varphi_m(x)=\delta_{nm}\,, \tag{2.3b}$$

and

$$\omega_c\equiv eH/mc\,, \tag{2.3c}$$

$$x_0\equiv-\hbar k/m\omega_c\,. \tag{2.3d}$$

The total energy of the state,

$$E_{n,m,k}=(n+\tfrac{1}{2})\hbar\omega_c+\epsilon_m\,, \tag{2.4}$$

and the bound-state wave function $\chi_m(z)$ are determined by the one-dimensional Schrödinger equation

$$\left(-\frac{\hbar^2}{2m}\frac{d^2}{dz^2}+V(z)\right)\chi_m(z)=\epsilon_m\chi_m(z)\,, \tag{2.5a}$$

$$\chi_m(0)=0\,, \tag{2.5b}$$

$$\chi_m(\infty)=0\,. \tag{2.5c}$$

B. Schrödinger's Equation for Mobile States

We take a quantization length L in the z direction, treating the states as though their eigenvalue spectrum is discrete, and normalize the states in this quantization volume. At the appropriate point in the calculation, we let L become infinite. This procedure might seem needlessly pedantic, but in the course of this work we encountered questions involving terms of order $1/L$ whose integrated ef-

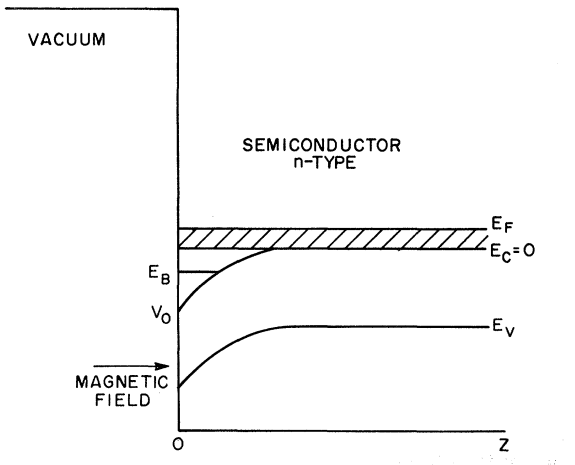

FIG. 1. Schematic representation of the potential in the neighborhood of an n-type semiconductor-vacuum interface in an electric field strong enough to support one bound state at energy $-E_B$. Dashed lines represent filled states of positive energy.

fect could not be ignored, and this admittedly pedestrian approach was the one which most satisfactorily answered these questions for us.

The x and y dependence of the mobile states is the same as for the bound states. The z-dependent part of the states will be labeled by a subscript j, and is determined by the one-dimensional Schrödinger equation

$$\left(-\frac{\hbar^2}{2m}\frac{d^2}{dz^2}+V(z)\right)\chi_j(z)=\epsilon_j\chi_j(z)\,, \tag{2.6a}$$

$$\chi_j(0)=0\,, \tag{2.6b}$$

$$\chi_j(L)=0\,, \tag{2.6c}$$

$$\epsilon_j\geq 0\,. \tag{2.6d}$$

C. Constitutive Equation for Charge Density

The charge density $\rho(x)$ which appears in Poisson's equation is composed of three parts:

$$\rho(x)=-2e\sum_k\sum_n\left(\sum_m f(\mu-E)\,\psi^*\psi+\sum_j f(\mu-E)\,\psi^*\psi\right)+eN_0\,. \tag{2.7}$$

The first term here represents the bound states, the second represents the mobile states, and the third represents the uniform background. The factor of 2 is the spin degeneracy and the function $f(\mu-E)$ is the probability that the state of energy E is occupied when the Fermi energy is μ. We consider only the zero-temperature limit and ignore lifetime broadening, so that $f=1$ if $\mu>E$, and $f=0$ otherwise.

Returning to (2.7), we replace the sum over k by an integral using periodic boundary conditions over the length L_y,

$$\sum_k\rightarrow\frac{L_y}{2\pi}\int dk\,.$$

Replacing the integration over k by an integration over x_0 which is carried out by using (2.3b), and using expression (2.4) for the energy, we obtain

$$\rho(z)=\frac{-e}{2\pi}\frac{2m}{\hbar^2}\hbar\omega_c\sum_n\left(\sum_m f[\mu-(n+\tfrac{1}{2})\hbar\omega_c-\epsilon_m]\chi_m^2(z)+\sum_j f[\mu-(n+\tfrac{1}{2})\hbar\omega_c-\epsilon_j]\chi_j^2(z)\right)+eN_0\,. \tag{2.8}$$

D. Poisson's Equation for Potential

We assume that the potential energy $V(z)$ appearing in Schrödinger's equation is just the energy of an electron in the electrostatic potential $U(z)$. The electrostatic potential satisfies Poisson's equation:

$$V(z)=-eU(z)\,, \tag{2.9a}$$

$$-\frac{d^2U}{dz^2}=\frac{4\pi\rho(z)}{\epsilon}\,. \tag{2.9b}$$

Boundary conditions on the potential are

$$U(z=\infty)=0\,, \tag{2.10a}$$

$$-(dU/dz)_{z=0}=E_0\,. \tag{2.10b}$$

Condition (2.10a) serves to determine the zero level of the potential, and Eq. (2.10b) corresponds to a given value of electric field at the surface of the semiconductor. Note that a condition at infinity, such as (2.10a), also implies that

$$(dU/dz)_{z=\infty}=0\,. \tag{2.10c}$$

Hence, if we integrate (2.9b) from $z=0$ to $z=\infty$, we obtain

$$\int_0^\infty\rho(z)\,dz=-\epsilon E_0/4\pi\,, \tag{2.11}$$

a condition on the charge density which must be *imposed* if boundary condition (2.10) is to be maintained.

Before proceeding further, it is convenient to introduce the dimensionless units which we shall employ for most of what follows. All energies will be expressed as multiples of E_F^0, the Fermi energy at zero magnetic field, and all lengths will be expressed as multiples of k_F^{-1}, the inverse of the Fermi wave number. That is, we have

$$E_F^0=\hbar^2k_F^2/2m\,, \tag{2.12a}$$

$$\epsilon_m=E_mE_F^0 \quad \text{(bound-state energy)}\,, \tag{2.12b}$$

$$\epsilon_j=E_jE_F^0 \quad \text{(mobile-state energy)}\,, \tag{2.12c}$$

$$\mu=E_FE_F^0 \quad \text{(Fermi energy)}, \tag{2.12d}$$

$$\hbar\omega_c=E_HE_F^0 \quad \text{(magnetic energy)}, \tag{2.12e}$$

$$V(z)=\phi(z)E_F^0 \quad \text{(potential energy)}\,, \tag{2.12f}$$

$$z=\hat{z}k_F^{-1} \quad \text{(unit of length)}\,. \tag{2.12g}$$

Then, we express the charge density in terms of the number density appropriate to this unit of length:

$$\rho(z)=-en(\hat{z})k_F^3\,. \tag{2.12h}$$

The wave functions are also renormalized to be compatible with the new unit of length:

$$\chi(z)=\hat{\chi}(\hat{z})k_F^{1/2}\,. \tag{2.12i}$$

(From this point on, we suppress the caret above $\hat{z}$ and $\hat{\chi}$.) Schrödinger's equations (2.5) and (2.6) now take the form

$$\left(-\frac{d^2}{dz^2}+\phi(z)\right)\chi_m(z)=E_m\chi_m(z)\,, \tag{2.13a}$$

$$\chi_m(0)=0\,, \tag{2.13b}$$

$$\chi_m(\infty)=0\,, \tag{2.13c}$$

$$\int_0^\infty \chi_m^2(z)\,dz = 1 \tag{2.13d}$$

and

$$\left(-\frac{d^2}{dz^2}+\phi(z)\right)\chi_j(z) = E_j\chi_j(z)\,, \tag{2.14a}$$

$$\chi_j(0) = 0\,, \tag{2.14b}$$

$$\chi_j(L) = 0\,, \tag{2.14c}$$

$$\int_0^L \chi_j^2(z)\,dz = 1\,. \tag{2.14d}$$

The constitutive equation (2.8) becomes

$$n(z) = \frac{E_H}{2\pi}\sum_n\Bigg(\sum_m f[E_F-(n+\tfrac{1}{2})E_H-E_m]\chi_m^2(z) + \sum_j f[E_F-(n+\tfrac{1}{2})E_H-E_j]\chi_j^2(z)\Bigg) - n_0\,. \tag{2.15}$$

Poisson's equation (2.9) becomes

$$-\frac{d^2\phi(z)}{dz^2} = Kn(z)\,, \tag{2.16}$$

where

$$K = 8\pi me^2/\epsilon\hbar^2k_F = 8\pi\alpha r_s\,, \tag{2.17a}$$

$$\alpha = (4/9\pi)^{1/3} = 0.52\,. \tag{2.17b}$$

The boundary conditions and constraint on the charge density become

$$\phi(z=\infty) = 0\,, \tag{2.18a}$$

$$\left(\frac{d\phi}{dz}\right)_{z=0} = KQ\,, \tag{2.18b}$$

$$\int_0^\infty n(z)\,dz = Q\,. \tag{2.18c}$$

The symbol "infinity" used in (2.18) means "as deep in the medium as we care to go, keeping in mind that the medium has a thickness L, established for quantization purposes, that is still longer." We shall always interpret an expression like (2.18c) to mean

$$Q = \lim_{\substack{R/L\to 0\\ R\to\infty}}\int_0^R n(z)\,dz\,. \tag{2.19}$$

III. ONE-DIMENSIONAL FRIEDEL SUM RULE

We now derive a Friedel-like[6] sum rule for a planar system in which bound states occur. The problems here are not associated with the presence of the bound states but rather with the one dimensionality of the potential, which introduces terms of order $1/L$ whose effect is negligible in the corresponding point-impurity situation. Parts of what follow are implicitly or explicitly in the standard literature on the Friedel sum rule. We have not, however, encountered a fully satisfactory derivation in the literature of the results we need.

At large values of z where $\phi(z)$ is negligible, Eq. (2.14) for the mobile charge states tells us that $\chi_j(z)$ has the form

$$\chi_j(z) = C(k_j)\sin[k_jz+\eta(k_j)] \tag{3.1a}$$

and that

$$E_j = k_j^2\,. \tag{3.1b}$$

In order to satisfy boundary condition (2.14c), we shall have

$$k_jL+\eta(k_j) = (M+j)\pi\,, \tag{3.2}$$

where M is some integer which we hold fixed.[9] The quantity $C(k_j)$ is a normalization constant, and is ultimately determined by (2.14d).

From (3.2) we see that successive values of k_j differ by order $1/L$, so that as L becomes large, k can be treated as a continuous variable and the wave functions can almost be treated as continuous functions of k. The "almost" proviso arises here because the wave functions can be treated as continuous only if the argument $k_jz+\eta(k_j)$ changes infinitesimally in going from one allowed value of k to the next. This restricts us to values of z such that $z/L\to 0$. However, this restriction allows all values of z which we will encounter in integrals such as (2.19).

With this proviso understood, we can make the standard replacement for going from a sum to an integral in the j summation in (2.15):

$$\sum f(k_j) = \sum f(k_j)\,\Delta j = \sum\frac{f(k_j)\,\Delta k}{\Delta k/\Delta j} \to \int\frac{f(k)\,dk}{\Delta k/\Delta j}\,, \tag{3.3}$$

where $\Delta k/\Delta j$ is the change in allowed values of k_j in going from one value of j to the next. Because η can be regarded as a continuous function of k, we can evaluate this quantity from (3.2), obtaining

$$\left(\frac{\Delta k}{\Delta j}\right)^{-1} = \frac{L+d\eta/dk}{\pi}\,. \tag{3.4}$$

Hence the mobile-state contribution to (2.15) is

$$n_{\rm mob}(z) = \frac{E_H}{2\pi^2}\sum_n\int_0^\infty f[E_F-(n+\tfrac{1}{2})E_H-k^2]\times\left(L+\frac{d\eta}{dk}\right)\chi_k^2(z)\,dk = \frac{E_H}{2\pi^2}\sum_n\int_0^{k_n}\left(L+\frac{d\eta}{dk}\right)\chi_k^2(z)\,dk\,, \tag{3.5}$$

where

$$k_n^2 = E_F-(n+\tfrac{1}{2})E_H \tag{3.6}$$

and the sum goes over all values for which $k_n^2 > 0$.

There is a useful relationship between the factor $L + d\eta/dk$ and the normalization of the states which markedly simplifies the evaluation of (3.5). Recall first that $\chi_k(z)$ is normalized according to (2.14d). Let us consider another function, $\xi_k(z)$, which is proportional to $\chi_k(z)$. That is, it satisfies the same equation (2.14a) and boundary conditions (2.14b) and (2.14c) as does $\chi_k(z)$ but its normalization is different, being set by the condition that at large z, where the potential $\phi(z)$ vanishes, $\xi_k(z)$ takes the asymptotic form

$$\xi_k(z) = 2^{1/2} \sin[kz + \eta(k)] \,. \tag{3.7}$$

Comparing the two asymptotic forms (3.7) and (3.1a) gives

$$\chi_k(z) = 2^{-1/2} C(k)\, \xi_k(z) \,, \tag{3.8}$$

which, inserted into the normalization condition (2.14d), yields

$$\tfrac{1}{2} C^2(k) \int_0^L \xi_k^2(z)\, dz = 1 \,. \tag{3.9}$$

The integral here can be evaluated using a minor variant of the mathematics which is usually employed in deriving the sum rule.[10] We write

$$\int_0^L \xi_k^2(z)\, dz = \int_0^R \xi_k^2(z)\, dz + \int_R^L \xi_k^2(z)\, dz \,, \tag{3.10}$$

where R is chosen large enough so that $\xi_k(z)$ has already taken on its asymptotic form (3.7) at $z = R$ but small enough so that $R/L \to 0$. This can always be done if L is taken large enough. The second integral on the right-hand side of (3.10) may be evaluated using the asymptotic form (3.7) and using boundary condition (2.14c). In the first integral, because $R/L \to 0$, we *are* permitted to treat $\xi_k(z)$ as a continuous function of k so that we may differentiate it with respect to k. Hence, we are allowed to write the Schrödinger equation for ξ_k,

$$\left(-\frac{d^2}{dz^2} + \phi(z) - k^2\right) \xi_k(z) = 0 \,,$$

to differentiate this equation with respect to k, to multiply from the left by $\xi_k(z)$, and to integrate from $0 < z < R$. We can also multiply this equation from the left by $d\xi_k(z)/dk$ and integrate from $0 < z < R$. Subtracting the two integrals from each other gives

$$2k \int_0^R \xi_k^2(z)\, dz = -\left(\xi \frac{d^2\xi}{dk\,dz} - \frac{d\xi}{dk}\frac{d\xi}{dz}\right)_0^R = 2k\left(R + \frac{d\eta}{dk}\right) - \sin 2(kR + \eta) \,. \tag{3.11}$$

The terms on the right-hand side have been evaluated using boundary condition (2.14b) and the asymptotic form (3.1). Combining (3.11) with the evaluation of the second integral on the right-hand side of (3.10) gives us

$$\int_0^L \xi_k^2(z)\, dz = L + \frac{d\eta}{dk}$$

or, using (3.9),

$$L + d\eta/dk = 2/C^2(k) \,. \tag{3.12}$$

Inserting this into (3.5) and using (3.8) gives

$$n_{\text{mob}}(z) = \frac{E_H}{2\pi^2} \sum_n \int_0^{k_n} \xi_k^2(z)\, dk \,, \tag{3.13}$$

a result which has the advantage that neither the factor $L + d\eta/dk$ nor the normalization of the states χ_k need be computed.

Now, consider the charge density (2.15) at values of z large enough that the potential is negligible and that the bound states have decayed away. We use the asymptotic form (3.7) and have

$$n(z) = \frac{E_H}{2\pi^2} \sum_n \int_0^{k_n} \{1 - \cos 2[kz + \eta(k)]\}\, dk - n_0 \,. \tag{3.14}$$

The charge density contains an osciliatory term and the sum of two constant (z-independent) terms. The constant term is forbidden, for a constant charge in a one-dimensional system of length L leads to a potential of order L^2. Hence, the two constant terms must cancel against each other to at least order L^{-2}. That is,

$$\frac{E_H}{2\pi^2} \sum_n k_n = n_0 \,. \tag{3.15}$$

The background charge density n_0 is fixed. Hence (3.15) and (3.6) together determine the Fermi energy in the presence of a magnetic field (see Appendix A). These are, to be sure, exactly the same as the equations usually written down for the bulk problem where the potential is zero everywhere. The point here is that even though there exists an oscillatory potential of order $1/L$, the Fermi energy is unshifted from its bulk value to an accuracy at least of order L^{-2}.

Making use of (3.15) and (3.13), we can now rewrite the charge density (2.15) as

$$n(z) = \frac{E_H}{2\pi} \sum_m N(E_F - E_m)\, \chi_m^2(z) + \frac{E_H}{2\pi^2} \sum_n \int_0^{k_n} dk\, [\xi_k^2(z) - 1] \,, \tag{3.16}$$

where

$$N(E_F - E_m) \equiv \sum_n f[E_F - (n + \tfrac{1}{2}) E_H - E_m] = \text{integer part of } \left(\frac{E_F - E_m}{E_H} + \frac{1}{2}\right) \,. \tag{3.17}$$

Note that $N(E_F - E_m)$ is the number of occupied Landau levels associated with the bound state whose energy is E_m.

Finally, we wish to evaluate the total charge Q, given by (2.19). Using the form (3.16) and the normalization of the bound states, we have

$$Q=\frac{E_H}{2\pi}\sum_m N(E_F-E_m)+\frac{E_H}{2\pi^2}\sum_n \lim_{\substack{R\to\infty\\ R/L\to 0}}\int_0^R dz \times\int_0^{k_n} dk\,[\xi_k^2(z)-1]\,. \quad (3.18)$$

We may interchange the order of integration here, and, using (3.11), the second term becomes

$$\frac{E_H}{2\pi^2}\sum_n \lim_{R\to\infty}\int_0^{k_n}\left(\frac{d\eta}{dk}-(2k)^{-1}\sin 2(kR+\eta)\right)dk\,. \quad (3.19)$$

At large R, the sine integrand oscillates rapidly and the only contribution is from the neighborhood of $k=0$. We can then expand $\eta(k)$ as

$$\eta(k)=\eta(0)+k\eta'\,, \quad (3.20a)$$

$$\eta'=(d\eta/dk)_{k=0}\,. \quad (3.20b)$$

At this point, we invoke Levinson's theorem,[11] namely, that if the phase shift $\eta(k)$ at $k=\infty$ is taken to be zero and if $\eta(k)$ is taken to be a continuous function of k, then

$$\eta(0)=\pi M_b\,, \quad (3.21)$$

where M_b is the number of bound states which the potential will support.

Using (3.21) and (3.20) in (3.19) allows us to perform the integral, take the limit, and finally, for the total charge Q, we arrive at

$$Q=\frac{E_H}{2\pi}\sum_{m=1}^{M_b} N(E_F-E_m)+\frac{E_H}{2\pi^2}\sum_n[\eta(k_n)-\eta(0)-\tfrac{1}{4}\pi]\,. \quad (3.22)$$

The sum over m is over the bound states, and the sum over n is over the occupied Landau levels of the mobile electrons. The number of terms in the n summation is clearly $N(E_F)$ [cf. Eq. (3.17)] and hence an alternative statement of (3.22) is

$$Q=\frac{E_H}{2\pi}\sum_{m=1}^{M_b}[N(E_F-E_m)-N(E_F)] +\frac{E_H}{2\pi^2}\sum_n[\eta(k_n)-\tfrac{1}{4}\pi]\,. \quad (3.23)$$

Both (3.22) and (3.23) are statements of the Friedel-like sum rule which will play an important role in our calculations.

IV. RUNNING-WAVE SOLUTIONS

It turns out to be more convenient to calculate the running-wave solutions for the mobile states than the standing-wave solutions. Accordingly, let us write $\psi_k^+(z)$ for the outgoing running wave; that is,

$$\left(-\frac{d^2}{dz^2}+\phi(z)-k^2\right)\psi_k^+(z)=0\,, \quad (4.1a)$$

$$\psi_k^+(z)=e^{i[kz+\eta(k)]}f_k(z)\,, \quad (4.1b)$$

$$f_k(z\to\infty)=1\,. \quad (4.1c)$$

Since the complex ψ^+ satisfies the real equation (4.1a), its real and imaginary parts must separately be solutions. Comparing the large-z behavior, we conclude that

$$\xi_k(z)=\sqrt{2}\,\mathrm{Im}\psi_k^+(z)\,. \quad (4.2)$$

The boundary condition $\xi_k(0)=0$ becomes

$$\mathrm{Im}e^{i\eta}f_k(0)=0$$

or

$$f_k(0)=R_k e^{-i\eta(k)}\,, \quad (4.3)$$

where R_k is a real number (of either sign). Equation (4.3) determines $\eta(k)$ up to an additive multiple of π. That multiple of π is fixed uniquely by Levinson's theorem (3.21) and continuity of the phase shifts.

V. PARAMETRIC APPROACH

Although the analysis to this point has been exact within the model, our procedure in carrying out the calculation and in analyzing the problem further will make use of a parametrized potential. We shall solve Schrödinger's equation, calculate charge densities, and use this charge density to compute a potential, making no approximations. The recomputed potential, when fitted by the parametric form, leads to new values of the parameters. Self-consistency then demands that the new parameters and the old be equal.

The form which we choose to represent the potential depends on three parameters:

$$V_p(z)=-\frac{V_0}{1-\beta}\,(e^{-\lambda z}-\beta e^{-2\lambda z})\,. \quad (5.1)$$

One of the parameters, β, can be expressed in terms of V_0 and λ by the requirement that the parametric form have the correct slope at the origin, i.e., that it satisfy (2.18b). We require

$$V_0\lambda(1-2\beta)/(1-\beta)=KQ\,. \quad (5.2)$$

In Fig. 2, we show how, with a fixed V_0 (a fixed well depth) and β determined by (5.2) (a fixed slope), varying λ changes the shape of the well from attractive to repulsive. The $\lambda=0.2$ curve is repulsive in that it moves charge away, inducing a deficit of charge. The $\lambda=0.8$ curve is the most attractive one shown, in that it moves the most charge inwards, inducing the largest surplus of charge. With V_0 fixed, we then sweep λ, running through a family of potentials such as that shown

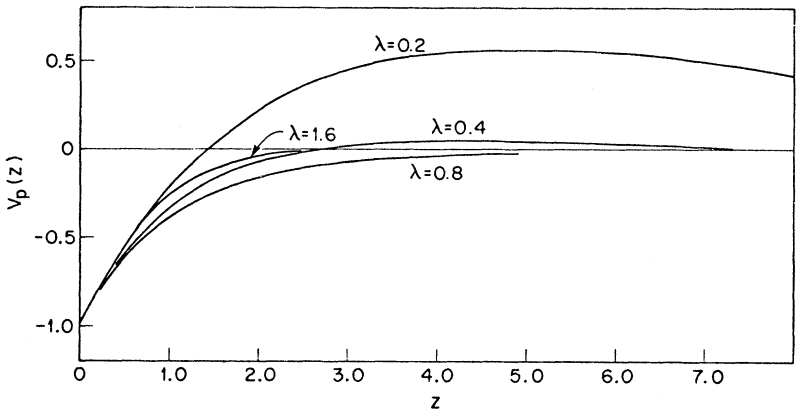

FIG. 2. Typical shapes of the parametric potential for fixed values of well depth and slope at the surface.

in Fig. 2, selecting that potential for which the total induced charge is indeed the value we have chosen in (5.2). This step, demanded by Eq. (2.18c), guarantees that the iterated potential will also have the same slope at the origin. To evaluate the induced charge, we need only calculate the bound-state energies E_m and phase shifts $\eta(k_n)$. We then use these in the sum rule (3.23). This leaves us with one parameter to adjust, namely, V_0, the depth of the parametrized potential at $z=0$. The fitting procedure we choose is to adjust V_0 so that the parametrized potential and the recomputed potential agree at $z=0$.

Before going into any detail about how this is carried out, it may be useful to demonstrate just what sort of self-consistency is available by the use of this rather simple scheme. In Fig. 3, we have plotted the parametrized potential and the resulting potential (with the parameters determined self-consistently) for two of the cases we have studied. One sees agreement which is good enough to ensure that the potential determined by our parametric scheme is an excellent approximation to the potential which would satisfy the fully self-consistent Hartree equations.

There are likely to be other three-parameter forms which do as well, or better, than this one in achieving self-consistency. This form was chosen solely because of the simplicity of its eigenfunctions. For its bound states, we obtain

$$\chi_m(z) = e^{-\alpha K_m} \sum_{j=0}^{\infty} a_j(K_m)\, e^{-j\lambda\alpha}\,, \tag{5.3a}$$

with

$$E_m = -K_m^2\,. \tag{5.3b}$$

The energy is determined by applying boundary condition (2.13b), which here becomes

$$\sum_{j=0}^{\infty} a_j(K_m) = 0\,. \tag{5.3c}$$

To calculate the coefficients a_j, we insert (5.3a) and (5.3b) into the Schrödinger equation (2.13a) and obtain a simple recursion relationship for the coefficients:

$$a_1 = -P a_0/(\lambda^2 + 2K_m\lambda)\,, \tag{5.4a}$$

$$a_j = -\frac{P(a_{j-1} - \beta a_{j-2})}{\lambda^2 j^2 + 2K_m j\lambda}\,, \quad j>1 \tag{5.4b}$$

$$P \equiv V_0/(1-\beta)\,. \tag{5.4c}$$

The coefficient $a_0(K_m)$ is fixed by the normalization requirement (2.13d):

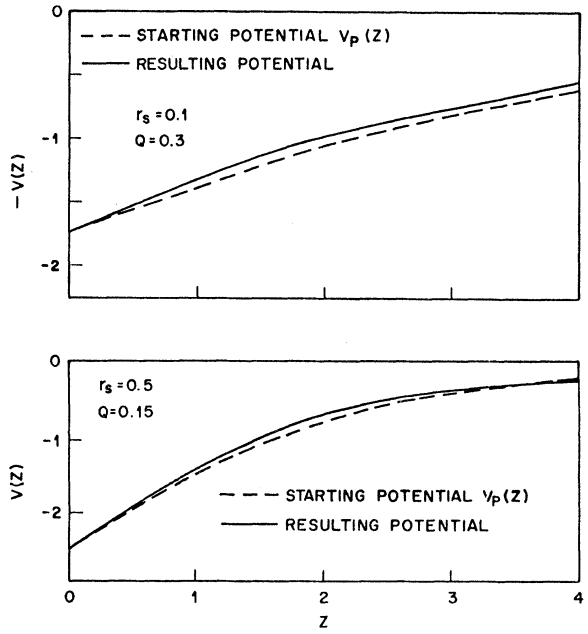

FIG. 3. Comparison between the self-consistently determined parametric starting potential (dashed line) and the iterated resulting potential (solid line) to which it gives rise, for two typical cases.

$$\sum_{r=0}\sum_{s=0}\frac{a_r a_s}{2K_m+(r+s)\lambda}=1\ . \tag{5.4d}$$

The solution for the running-wave states is equally simple. Here, we have (cf. Sec. IV)

$$\psi_k^+(z)=e^{i(kz+\eta)}\sum_{j=0}^{\infty}b_j(k)\,e^{-j\lambda z}\ , \tag{5.5a}$$

$$b_1=-Pb_0/(\lambda^2-2i\lambda k)\ , \tag{5.5b}$$

$$b_j=-\frac{P(b_{j-1}-\beta b_{j-2})}{\lambda^2 j^2-2ikj\lambda}\ , \tag{5.5c}$$

and the coefficient b_0 is also fixed by normalization, in this case by (4.1c):

$$b_0=1\ . \tag{5.5d}$$

The phase shift η is determined using (4.3):

$$\sum_{j=0}b_j(k)=R_k\,e^{-i\eta(k)}\ . \tag{5.6}$$

Let us turn now to the evaluation of the potential. First, of course, we need its value at the origin. Integrating Poisson's equation (2.16) twice gives us

$$-\phi(0)=K\int_0^\infty z'\,N(z')\,dz'\ . \tag{5.7}$$

It is now just a matter of using (3.16) for the density, using (4.2), (5.3), and (5.6) for the wave functions, and collecting together the terms which result. All the spatial integrals can be done in closed form, most of them trivially, one with a bit more work (see Appendix B), and the result is the expression

$$\begin{aligned}\frac{-2\pi\phi(0)}{KE_H}=&\sum_m^{M_b}N(E_F-E_m)\sum_{r=0}\sum_{s=0}\frac{a_r(K_m)\,a_s(K_m)}{[2K_m+(r+s)\lambda]^2}\\&+\frac{1}{\pi}\sum_n\left[\frac{\pi\eta'}{4}-\frac{\cos 2\eta(k_n)}{4k_n}\right.\\&\left.-\frac{1}{2}\int_0^{k_n}k^{-1}\frac{d\eta}{dk}\sin 2\eta(k)\,dk\right]\\&+\frac{1}{\pi}\sum_n\int_0^{k_n}h(k)\,dk\ ,\end{aligned} \tag{5.8}$$

where $\eta'=(d\eta/dk)_{k=0}$, and where $h(k)$ is an expression (again see Appendix B) involving the expansion coefficients $b_j(k)$ and phase shifts $\eta(k)$ of the mobile states.

The self-consistency condition of course is

$$-\phi(0)=V_0\ . \tag{5.9}$$

Assume that the parameters have been adjusted so that self-consistency has been achieved; i.e., we have satisfied (5.2), (3.23), and (5.9) in that order. It is now a simple matter to evaluate $\phi(z)$ for any value of z. Again, we integrate Poisson's equation twice, and arrange the result in the form

$$-\phi(z)=K\int_0^\infty z'\,N(z+z')\,dz'\ . \tag{5.10}$$

Note that the calculation of $\phi(z)$ is identical to the calculation of $\phi(0)$ except that $z+z'$ replaces z' as the argument of the density and, hence, of the wave functions. This substitution is equivalent to replacing

$$a_j(K_m)\ \text{ by }\ a_j(K_m)\,e^{-z(K_m+j\lambda)}\ , \tag{5.11a}$$

$$b_j(k)\ \text{ by }\ b_j(k)\,e^{-j\lambda z}\ , \tag{5.11b}$$

and

$$\eta(k)\ \text{ by }\ \eta(k)+kz\ . \tag{5.11c}$$

Hence, the same numerical program which evaluated $\phi(0)$ also serves to evaluate $\phi(z)$ if we first make the replacements (5.11). The potentials exhibited in Fig. 3 were computed in this way.

Finally, there is the matter of transition to the limit of zero magnetic field. It is easy to see that this limit is accomplished via the replacement of

$$E_H N(\epsilon)\ \text{ by }\ \epsilon\ , \tag{5.12a}$$

$$E_H\textstyle\sum_n f(k_n)\ \text{ by }\ 2\int_0^1 k\,dk\,f(k)\ , \tag{5.12b}$$

and

$$E_H\textstyle\sum_n\int_0^{k_n}h(k)\,dk\ \text{ by }\ \int_0^1(1-k^2)\,h(k)\,dk\ . \tag{5.12c}$$

VI. KOHN-MAJUMDAR THEOREM

Kohn and Majumdar[7] have proved that the density matrix for a system of noninteracting fermions is an analytic function of the strength of the external potential, even at those values of the strength where the number of bound states changes. In the model we are using here, the electrons interact via a Hartree self-consistent potential which, for the Schrödinger equation part of the problem, is the same as having noninteracting particles in an external potential. Thus, one might expect that the properties of the model system here would also be analytic, even as the number of bound states changes. In particular, this would mean that the expulsion of a bound state from the well produces no sudden change in the shielding properties of the semiconductor.

Without going into the question of analyticity, it is easy to investigate whether our model does in fact exhibit continuity when the number of bound states changes. Suppose that the external conditions, e.g., density of conduction electrons, electric field, and magnetic field, are such that there is a bound state whose binding energy is particularly small. We let the external conditions vary in the direction which further decreases the binding of this state. The internal parameters V_0, λ, and β determined by solving Eqs. (5.2), (3.23), and (5.9) will change, indicating a change in the shape

of the self-consistent well in response to the change in external conditions. Eventually the binding energy of the state goes to zero. Note that even under these conditions there is a finite amount of charge bound by this zero-energy state because the transverse states associated with it are still filled up to the Fermi level. With further change, the zero-energy bound state disappears, and with it goes a sizable amount of bound charge. Thus, potentially at least, there is the possibility of a sudden rearrangement of the system, a sudden change in the shape of the well and of the distribution of the mobile charge, to compensate for the loss of the bound-state charge. If this indeed happens, then there will be a sudden and discontinuous change in the internal parameters V_0, λ, and β, caused by the loss of the bound state. We want to know whether or not such a discontinuous change occurs.

Let us examine the self-consistent equations then. First consider (5.2). The external conditions here are embodied in K and Q and these change continuously. There is nothing in the structure of this equation to give rise to a discontinuity when we lose a bound state; so we pass on to (3.23). Here we have the external condition Q on the left-hand side changing continuously. On the right-hand side, we have a potentially discontinuous piece, namely, the sum over bound-state contributions. When the number of bound states is reduced by one, there is one less term in the sum. Note, however, that the term which is deleted, namely,

$$N(E_F - E_B) - N(E_F)$$

(where it is E_B, the energy of the Bth state, which goes to zero), is itself zero when the state is expelled. There is therefore nothing in (3.23) to cause a discontinuity, and we pass on to Eq. (5.8). Here, we find a bound-state contribution to $\phi(0)$ which *is* discontinuous and is even singular as the number of bound states decreases by one, namely,

$$N(E_F - E_B) \sum_{r=0} \sum_{s=0} \frac{a_r(K_B)\, a_s(K_B)}{[2K_B + (r+s)\lambda]^2} ,$$

where again it is $E_B = -K_B^2$, the energy of the Bth state, which is going to zero. However, a careful investigation of the mobile charge contribution to $\phi(0)$ (see Appendix C) reveals that it too is discontinuous, and in just such a way as to cancel the apparent discontinuity caused by the bound contribution above. None of the three equations determining the shape of the potential is discontinuous when a bound state is expelled from the well. It follows, therefore, that the shape of the self-consistent potential is also continuous even when a bound state is expelled from the well, even though the mobile charge density and the bound charge density separately change discontinuously. The analysis in Appendix C makes it evident that the continuity arises because there is a hole carved out of the mobile charge distribution, of just such shape and magnitude as to neutralize the charge contained in the bound state. (The hole arises because the bound states and mobile states are orthogonal.) When the bound state is expelled, the charge it contains is no longer counted, but the orthogonality hole then disappears from the mobile charge. This leads to a situation where, if we inadvertently overlook the existence of a weakly bound state *and also neglect to have the mobile states orthogonal to it*, very little error results.

VII. RESULTS AND DISCUSSION

A. Potential for No External Fields

To calculate the potential in the absence of external fields, we solve the self-consistency equations (5.2), (3.23), and (5.9) for $Q = 0$ and $E_H = 0$. In Fig. 4, we have plotted the parametrized potential and the iterated potential for the value $r_s = 0.325$ for the purpose of conveying some idea of how good the parametric self-consistency is when $Q = 0$. Although the self-consistency is not as impressive as it was for the cases illustrated in Fig. 3, it is still reasonably good. In Fig. 5, we have plotted V_0 and λ as functions of r_s. These two figures illustrate that even at zero external field, the requirements of self-consistency give rise to a potential well at the surface. We can account quantitatively for the existence of the well and for its r_s dependence by using a model based on the following simple considerations.

Note added in proof. After this paper was accepted for publication we were made aware of an earlier paper by O. V. Konstantinov and A. Ya. Shik, Zh. Eksperim i Teor. Fiz. 58, 1662 (1970) [Sov. Phys. JETP 31, 891 (1970)] which also treats the effect of the large work junction in causing the wave functions to vanish at the surface, establishing a deficit of shielding charge. There is great similarity between their reasoning and results and our reasoning and results for the field-free situation discussed here and in Ref. 2. We are sorry not to have been aware of their work.

Consider the electron density in the neighborhood of an infinite potential step, assuming that the electrons are in a uniform potential $V = 0$. The electron states are the products of sine waves in the z direction and the usual exponentials $e^{(i\vec{k}\cdot\vec{R})}$, where R is parallel to the surface. All states whose energy is less than the Fermi energy E_F will be occupied, and the electronic density will be the familiar[12]

$$\rho_{el} = 3N_0 e\left(\frac{\sin\theta}{\theta^3} - \frac{\cos\theta}{\theta^2} - \frac{1}{3}\right), \tag{7.1a}$$

where

$$\theta = 2k_F Z \tag{7.1b}$$

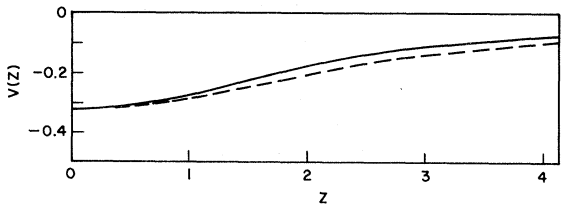

FIG. 4. Comparison between the self-consistently determined parametric starting potential (dashed line) and the iterated resulting potential (solid line) to which it gives rise for no external fields.

and

$$N_0 = k_F^3/3\pi^2 , \tag{7.1c}$$

$$k_F = (2M E_F^0/\hbar^2)^{1/2} . \tag{7.1d}$$

(In this section we shall use conventional units.) If we add to this the background charge density, the total density becomes

$$\rho(z) = 3N_0 e\left(\frac{\sin\theta}{\theta^3} - \frac{\cos\theta}{\theta^2}\right) . \tag{7.2}$$

This expression, plotted in Fig. 6, shows that there is a charge density at and near the surface of the semiconductor. This charge density arises because the uniform background is not neutralized by conduction electrons. (The conduction-electron wave functions vanish at $z = 0$.) This unneutralized charge creates an electric field which the mobile electron in the interior must screen. The spatial extent of the charge region is of order k_F^{-1}, the magnitude of the charge density is of order k_F^3, and so the magnitude of the charge to be shielded is of order k_F^2, which increases as r_s is made smaller. Hence, the well deepens as r_s is reduced. (This trend may not be obvious in the shape of Fig. 5 until one recalls that the unit of energy in Fig. 5 is E_F^0, which goes as r_s^{-2}.)

The essential features of this model are (a) that the electron density near the surface is assumed to be dominated by the requirement that the wave functions vanish at the surface and (b) that deeper in the interior the density is controlled by the necessity that the uncovered charge at the surface be shielded.

To make the argument quantitative, consider the semiconductor divided into two regions, $z < D$ and $z > D$. For $0 < z < D$, the conduction-electron density is less than that of the background, so that D marks the edge of the region where the unneutralized background is found. That is, $\rho(z = D) = 0$. The conduction-electron density vanishes at $z = 0$, and the charge density there is that of the unneutralized background; that is, $\rho(z = 0) = eN_0$. The simplest possible approximate description of the charge density in this region is a linear interpolation:

$$\rho(z) = eN_0\,(1 - z/D) , \quad z < D . \tag{7.3}$$

Beyond $z = D$ the electron density will exceed that of the background because of the presence of a potential energy to which the electrons respond. We assume that the response is local, and we use the linearized Thomas-Fermi description of it. In such case, the total charge density decays exponentially,

$$\rho(z) = -Ae^{-(z-D)/a} , \quad z > D \tag{7.4}$$

where A is a constant to be determined and a is the Thomas-Fermi shielding length.

The constant A is determined by the condition that the total induced charge be zero, corresponding to no external electric field:

$$Q = \int_0^\infty \rho(z)\,dz = 0 . \tag{7.5}$$

This gives

$$A = eN_0 D/2a .$$

We may insert the charge density, Eqs. (7.3) and (7.4), into Poisson's equation (2.9b) and, integrating, we find that the potential energy for an electron at $z = 0$ is

$$V(0) = -eU(z = 0) = -(4\pi e^2 N_0/\epsilon)\,(\tfrac{1}{3}D^2 + \tfrac{1}{2}aD) . \tag{7.6}$$

It is clear from Fig. 6 that when $V = 0$, $D \approx 2/k_F$. Now suppose that the Fermi momentum in the bulk, where $V = 0$, is k_F, but that there is a region near the surface where V has the value $V(0) = -V_0$. Near

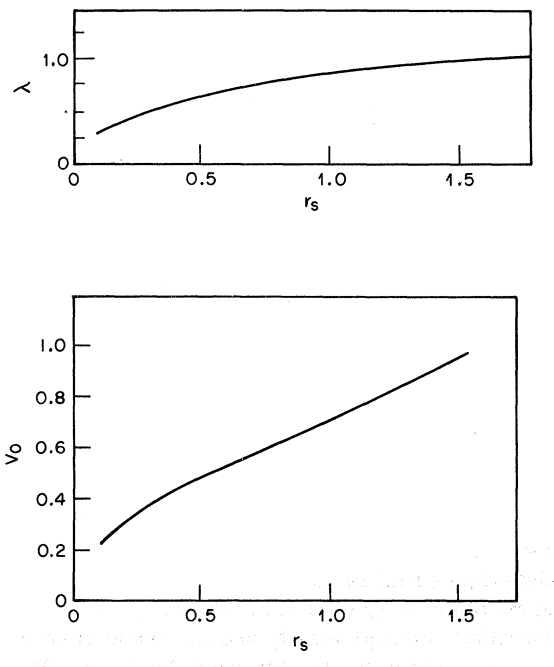

FIG. 5. Self-consistent potential parameters V_0 and λ plotted as a function of r_s for no external fields.

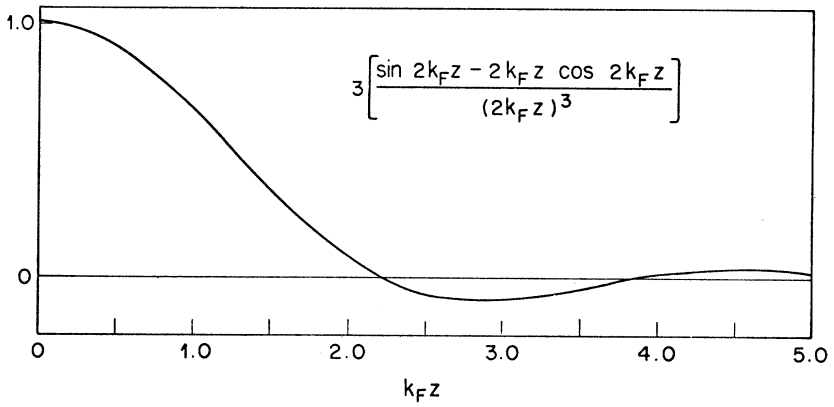

FIG. 6. Charge density in a region of zero potential in the neighborhood of an infinite potential step. Electronic states are filled to a Fermi energy determined by the condition that the uniform-background charge density is unity.

the surface, an electron at the Fermi energy will have a momentum k_F^* equal to

$$k_F^* = k_F(1 + V_0/E_F^0)^{1/2} . \tag{7.7a}$$

We assume that D scales like an inverse momentum; that is, we take

$$D = 2/k_F^* . \tag{7.7b}$$

Combining (7.6) and (7.7) gives

$$\frac{4\pi e^2 N_0}{\epsilon K_F k_F^2}\left(\frac{\frac{4}{3}}{1 + V_0/E_F^0} + \frac{ak_F}{(1 + V_0/E_F^0)^{1/2}}\right) - \frac{V_0}{E_F^0} = 0 . \tag{7.8a}$$

Note that

$$\frac{4\pi e^2 N_0}{\epsilon E_F^0 k_F^0} = \frac{8}{3\pi}\ \alpha r_s = 6\alpha^4 r_s \tag{7.8b}$$

[cf. (2.17)], while the standard expression for the Thomas-Fermi shielding length gives

$$ak_F = (9\alpha^4 r_s)^{-1/2} .$$

Therefore, (7.8) takes the form

$$\frac{8\alpha^4 r_s}{1 + V_0/E_F^0} + \left(\frac{4\alpha^4 r_s}{1 + V_0/E_F^0}\right)^{1/2} - \frac{V_0}{E_F^0} = 0 , \tag{7.9}$$

which is easily solved for $r_s^{1/2}$ vs V_0/E_F^0. This solution is plotted in Fig. 7. Also plotted in the same figure is the result of the parametric self-consistent calculation, a result which has already appeared as Fig. 5. The agreement between the two results suggests that the simple model is indeed a reasonable description of the situation. This provides a paradox, because the self-consistent potential we have calculated is of such range and depth that it supports a bound state, as we show in Fig. 8. Here, we have plotted the binding energy of the state vs r_s as obtained from the parametrized self-consistent solution. The paradox arises because the reasoning on which the simple model was based was appropriate to a situation with no bound states, and ignored both the possible existence of a bound state and the severe distortion of the mobile-state wave functions which will accompany its formation. It is likely that the success of the simple model rests on the cancellation of the bound charge by the orthogonality hole which we noted earlier in our discussion of the Kohn-Majumdar theorem.

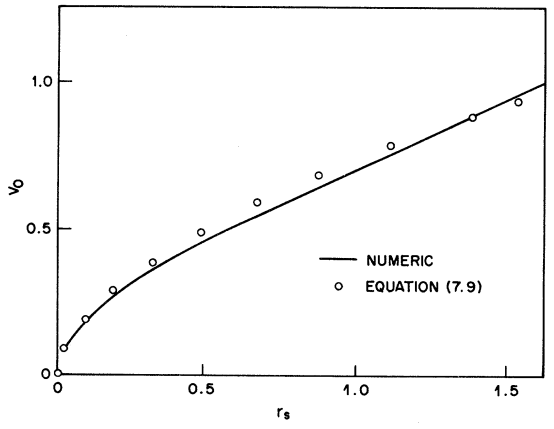

FIG. 7. Self-consistent potential parameter V_0 plotted as a function of r_s for no external field. The circles are the solutions to the model expressed by Eq. (7.9).

B. Shielding of External Electric Field

The simple model we have just described is

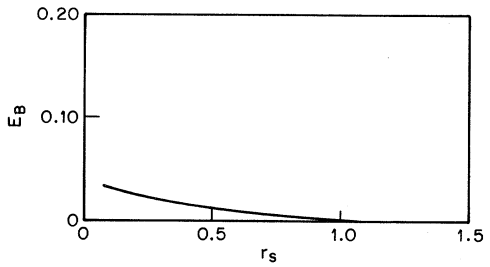

FIG. 8. Energy of the bound state supported by the self-consistent potential well in the absence of external fields.

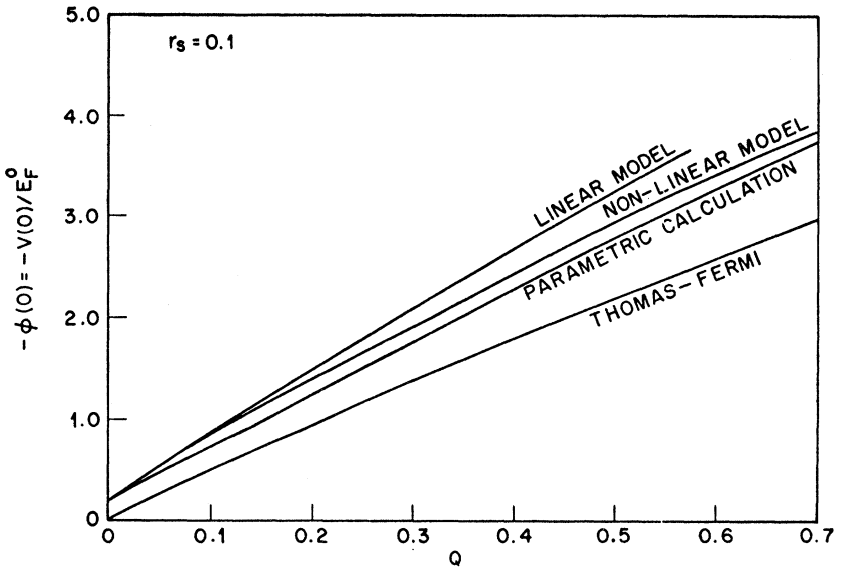

FIG. 9. Self-consistent well-depth parameter V_0 plotted as a function of the total screening charge Q (which is proportional to the external electric field) for $r_s = 0.1$.

easily extended to the case of finite electric fields merely by altering the $Q = 0$ condition of (7.5) and carrying through the derivation just as before. Results calculated in this way appear in Fig. 9 for $r_s = 0.1$, in Fig. 10 for $r_s = 0.5$, and in Fig. 11 for $r_s = 1.0$, and are labeled "Linear Model."

The well depths obtained are such that one wonders about the validity of linearized Thomas-Fermi

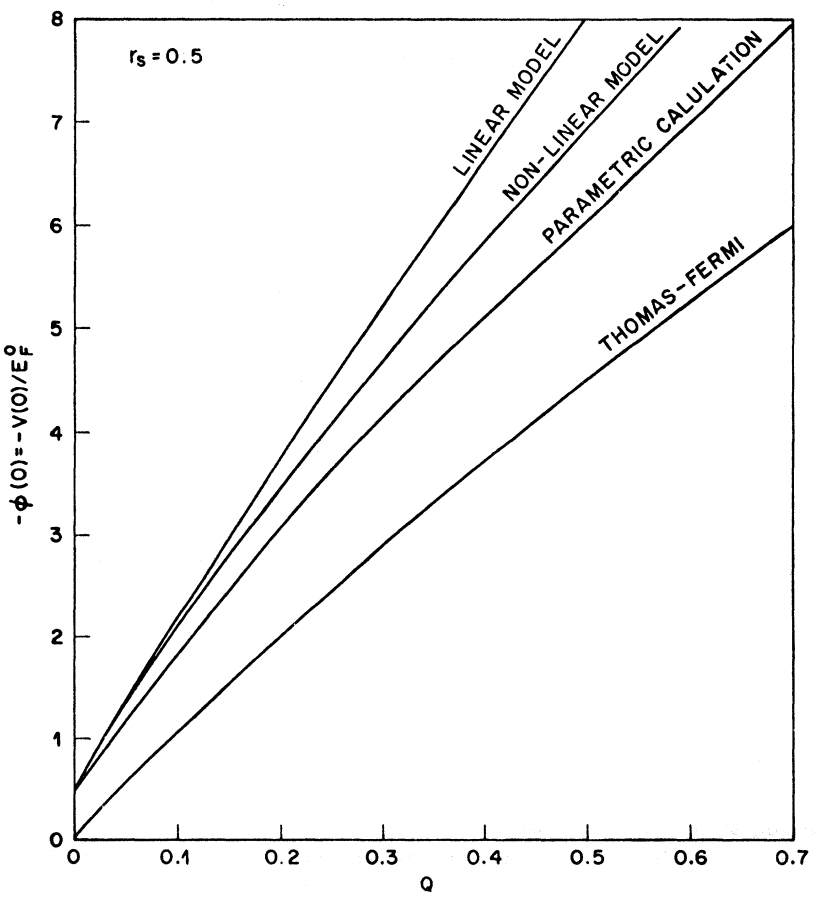

FIG. 10. Self-consistent well-depth parameter V_0 plotted as a function of the total screening charge Q (which is proportional to the external electric field) for $r_s = 0.5$.

theory in describing the response of electrons in the region $z > D$. Accordingly, let us digress to show how the model is altered by the use of nonlinear Thomas-Fermi theory in the region where $z > D$.

Again, we return to the dimensionless units of Eq. (2.12). The Thomas-Fermi density

$$n(z) = n_{\text{cond}}(z) - n_0 = (1/3\pi^2)\{[1-\phi(z)]^{3/2} - 1\} ,$$

inserted into Poisson's equation, Eqs. (2.16) and (2.17), gives the Thomas-Fermi equation for the potential,

$$-\frac{d^2\phi}{dz^2} = 6\alpha^4 r_s[(1-\phi)^{3/2} - 1] . \tag{7.10}$$

At large z, where $\phi(z) \to 0$, this equation becomes the linear Thomas-Fermi equation $\phi'' = 9\alpha^4 r_s \phi$, whose solution we write as

$$\phi(z) = ce^{-\lambda(z-d)} , \tag{7.11a}$$

where

$$\lambda = 3\alpha^2 r_s^{1/2} , \tag{7.11b}$$

$$d = Dk_F , \tag{7.11c}$$

and c is a constant in terms of which both $\phi(z=0)$ and Q will be expressed. By using $x \equiv z - d$ as an integration variable, and

$$\phi = ce^{-\lambda x}$$

as a condition at large x, we can integrate (7.10) numerically inward to $x = 0$, which gives us both $\phi(d)$ and $\phi'(d) \equiv (d\phi/dz)_d$ for any starting value of c.

In the region $0 < z < d$, we have the linear charge density (7.3), now expressed as

$$n(z) = -(1/3\pi^2)(1 - z/d) , \tag{7.12}$$

and, by integrating Poisson's equation, we have

$$\phi(0) = \phi(d) - d\phi'(d) + \alpha^4 r_s d^2 , \tag{7.13a}$$

$$\phi'(0) = \phi'(d) - 3\alpha^4 r_s d . \tag{7.13b}$$

The equation for d, (7.7), is

$$d = 2/[1-\phi(0)]^{1/2} ,$$

which, combined with (7.13a), yields d in terms of the known values of $\phi(d)$ and $\phi'(d)$. Using this value of d in (7.13) gives us

$$V_0 \equiv -\phi(z=0) , \tag{7.14a}$$

$$Q \equiv \int_0^\infty n(z)\,dz = \frac{\phi'(0)}{8\pi\alpha r_s} . \tag{7.14b}$$

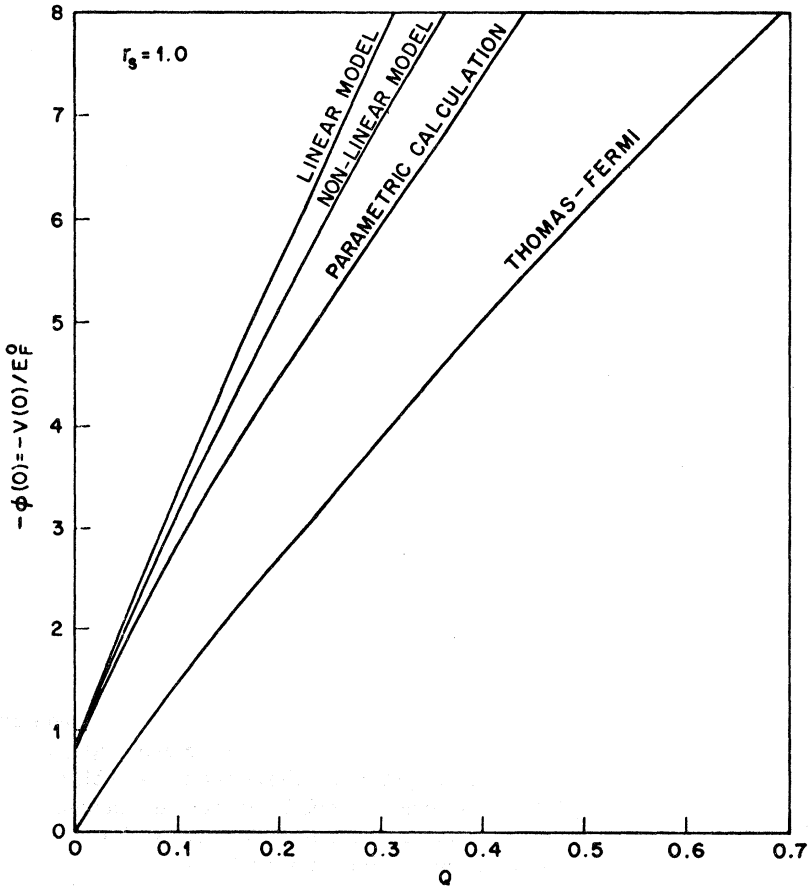

FIG. 11. Self-consistent well-depth parameter V_0 plotted as a function of total screening charge Q (which is proportional to the external electric field) for $r_s = 1.0$.

By running through various values of c, the V_0-vs-Q curves labeled "Nonlinear Model" in Figs. 9–11 were obtained.

The "Linear Model" calculation and "Nonlinear Model" calculations of course agree best with each other at small values of Q, where the potential throughout the well is small enough that linearization of the Thomas-Fermi equation is a mathematically valid procedure. The interesting question, though, is how well either of these simple models agrees with the results of the self-consistent calculation. The answer to this question also appears in Figs. 9–11, where the results of the self-consistent scheme we have been describing are labeled "Parametric Calculation." From these curves it appears that at small Q values agreement is excellent and that at large Q values the fractional error between the nonlinear model and the parametric calculation decreases. The difference between the two calculations is greatest at intermediate values of Q. Although it is not evident from the curves, the Q values where the fractional difference is greatest corresponds to well depths supporting between one (deeply bound) state and about three states. For well depths smaller than this, the single bound state is weakly bound or absent, and so the Kohn-Majumdar theorem is probably validating the model. For well depths larger than this, the number of bound states is large enough that the statistical reasoning on which the Thomas-Fermi theory rests does, in fact, apply. This is not to say that nonlinear Thomas-Fermi theory alone describes the shielding, because the region where $z < d$ plays an important role in determining the shape of the potential. To show that this is so, we have also plotted in Figs. 9–11 the results of a nonlinear Thomas-Fermi calculation in which the surface region is ignored, i.e., where the nonlinear Thomas-Fermi theory is assumed to hold all the way up to the surface. [This is accomplished by setting $d=0$ in Eq. (7.13).] The resulting curves, labeled "Thomas-Fermi" in Figs. 9–11, are evidently in poor agreement with either of the models or the parametric calculation.

C. Effect of External Magnetic Field

The magnetic field influences the potential through two separate mechanisms, both mechanisms arising from the quantization of the transverse motion into discrete Landau levels. The first mechanism is the quantization of the amount of charge associated with each bound state and the second is a change in the distribution of the mobile charge. Let us first consider what happens because of the bound states.

It appears from (3.22) and (3.17) that the amount of charge Q_m associated with the mth bound state must be an integer multiple of $E_H/2\pi$, the integer being equal to the number of occupied Landau levels. The amount of charge changes only when one of the Landau levels $E(m, n) = E_m + (n+\frac{1}{2})E_H$ sits just at the Fermi energy E_F. In what follows, we shall see several examples of how the potential well deforms in just such a way as to cause a bound-state Landau level to dwell at the Fermi energy. This will allow that particular level to fill or empty gradually in response to a continuous change in the external conditions, rather than fill or empty abruptly as it would do if it moved continuously through the Fermi energy.

The proper mathematical way of handling the case of a Landau level dwelling at the Fermi energy is to consider $f(\mu - E)$, the occupation factor which appears in (2.7), as the zero-temperature limit of the finite-temperature occupation factor. One arrives at the following prescription: Let $Q(V_0, \lambda)$ be the amount of charge induced in the system when the parameters of the potential are V_0 and λ, and where β is given by (5.2). [In this notation, Eq. (3.23), which expresses our effort to adjust the parameters to give the correct amount of induced charge, is $Q = Q(V_0, \lambda)$.] Suppose that as we sweep λ, the well deepens in a way that an extra Landau level fills at $\lambda = \lambda_0$. Then we shall have a discontinuity in $Q(V_0, \lambda)$ caused by the sudden filling of one Landau level:

$$Q(V_0, \lambda_0 - 0) - Q(V_0, \lambda_0 + 0) = E_H/2\pi \ .$$

If it turns out that the two values above bracket the value of charge we are trying to induce, i.e., if

$$Q(V_0, \lambda_0 - 0) < Q < Q(V_0, \lambda_0 - 0) \ ,$$

then we are to regard that topmost Landau level as being partially filled, the filling fraction x being given by solving

$$Q = xQ(V_0, \lambda_0 - 0) + (1 - x)\, Q(V_0, \lambda_0 + 0) \ . \qquad (7.15a)$$

In computing the potential at the origin, Eq. (5.8), we are then to treat the topmost Landau level as being fractionally filled by the same amount, i.e.,

$$\phi(0) = x\phi_0(V_0, \lambda_0 - 0) + (1 - x)\, \phi_0(V_0, \lambda_0 + 0) \ , \qquad (7.15b)$$

where $\phi_0(V_0, \lambda)$ is the potential at the origin, essentially the right-hand side of expression (5.8). This value of $\phi(0)$ is to be used in (5.9).

We can now describe how increasing the external electric field alters the well in the presence of a fixed magnetic field. Suppose that initially there is one bound state and that its Landau levels are either well above or well below the Fermi energy. As the electric field increases, the well deepens or broadens or both, so as to induce more charge. This change in potential alters the phase shifts $\eta(k_n)$ and the induced mobile charge [the second term in (3.22)] increases. The bound-state energy decreases (its binding energy $E_b = -E_m$ increases)

but the amount of charge in the bound state cannot change. As the bound-state energy decreases, the associated Landau levels also move downward in energy. Eventually one of the Landau levels reaches the Fermi energy. As that Landau level starts to fill, the charge it acquires tends to alter the potential in the direction of weakening the well, raising the bound-state energy and preventing the Landau level from filling further. This is the mechanism responsible for the tendency of the Landau level to dwell at the Fermi energy. Thus the downward motion of the bound state ceases during this dwell period, and the well deforms in such a way as to keep the bound-state energy fixed. In the example we shall exhibit, this will occur by having the range of the well become shorter even though the well becomes deeper as the field is increased. Eventually, the Landau level fills completely and the downward motion of the bound state resumes until the next Landau level reaches the Fermi energy and starts to dwell there. Finally, the well becomes sufficiently deep to support a second state. There are then two sets of Landau levels to try to follow; so the structure is more complicated, but the principle is the same. All of the features described above appear clearly in Fig. 12, where we have plotted V_0, λ, and E_B vs Q for $E_H = \frac{1}{4}E_F$ and $r_s = 0.5$. The range of Q extends from $Q = 0$ to just past $Q = 0.17$, where the onset of the second bound state occurs.

The tendency of a bound-state Landau level to dwell at the Fermi energy is also useful in understanding the effect of changing the magnetic field when the electric field is held fixed. As E_H increases, the amount of charge which resides in a single bound state increases, being given by [cf. Eq. (3.22)]

$$Q_m = N(E_F - E_m)E_H/2\pi \,. \tag{7.16}$$

When the field increases enough to force a Landau level through the Fermi energy, $N(E_F - E_m)$ drops by unity. This is shown in Fig. 13, and it occurs at magnetic field values such that

$$E_H = (E_F - E_m)/(N + \tfrac{1}{2}) \,. \tag{7.17}$$

Consider a magnetic field value such that there is no Landau level at the Fermi surface. An increase in E_H increases the amount of charge which resides in the bound state. This charge, being relatively close to the surface, alters the shape of the potential in the direction of making the well shallower. The binding energy of the state decreases and the energy of the Landau levels then rises with increase in magnetic field for two reasons; the energy difference between the Landau level and the bound state is increasing and, in addition, the energy of the bound state itself is also rising.

Eventually, a Landau level reaches the Fermi energy. As before, the abrupt emptying of a Landau level cannot occur. Instead, the Landau level must dwell at the Fermi energy, emptying gradually as the magnetic field increases. This emptying of the state means that charge relatively close to the surface is removed; hence the well tends to deepen, lowering the energy of the bound state. The rate at which the bound-state energy drops must exactly equal the rate at which the energy difference between it and the Landau level increases, in order that the dwell can occur. Hence, during the dwell period the binding energy depends on magnetic field as described by (7.17), namely,

$$E_b = -E_m = E_F\left(\frac{N+\frac{1}{2}}{E_F/E_H} - 1\right) \,. \tag{7.18}$$

Finally, the Landau level empties completely and there are no longer any occupied levels near the Fermi energy. The system then reverts to the

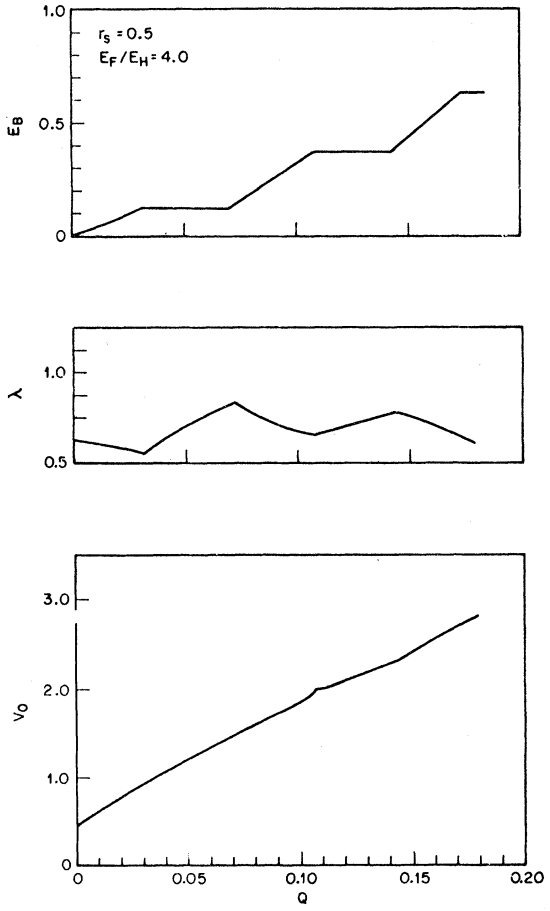

FIG. 12. Self-consistent well parameters V_0 and λ and the binding energy E_b of the state which that well supports plotted as a function of the total screening charge Q (which is proportional to the external electric field) in the presence of a strong magnetic field of strength $E_H = \frac{1}{4}E_F$.

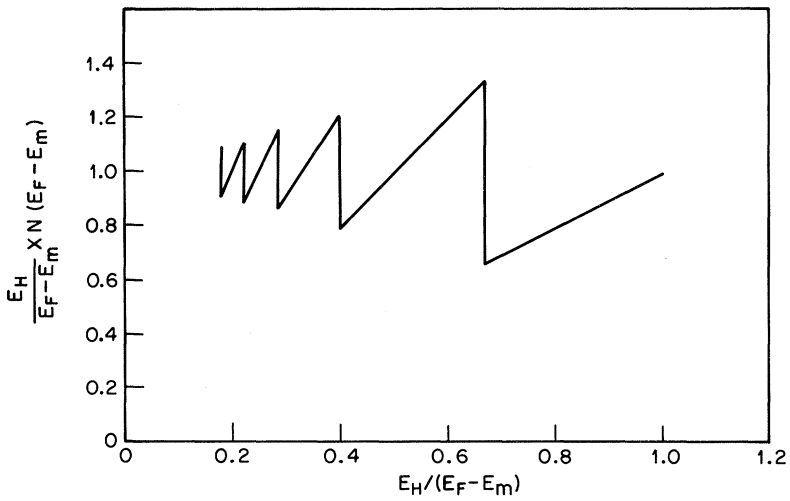

FIG. 13. Ratio of the charge, at magnetic field E_H, to the charge at zero magnetic field as a function of magnetic field for a single bound state of energy E_m (binding energy $E_b = -E_m$) and a Fermi energy of E_F.

previous behavior in which an increase in E_H decreases the binding energy.

The above behavior is apparent in Fig. 14, where we have plotted the well parameters V_0 and λ, and the binding energy $E_b = -E_m$, as a function of E_F/E_H for $r_s = 0.5$ and $Q = 0.15$, values which were selected so as to produce a single bound state. Also appearing in this graph is a horizontal bar which is blackened only over those ranges of E_F/E_H where a Landau level is exactly at the Fermi energy. We direct attention away from the values $E_F/E_H = \frac{3}{2}$, $\frac{5}{2}$, and $\frac{7}{2}$, where, seemingly there are discontinuities (to be discussed below), and we point out the oscillatory behavior of the binding energy E_b. Note that the binding energy increases with E_F/E_H (i.e., decreases as E_H increases), where there is no Landau level at the Fermi surface, and decreases with E_F/E_H (i.e., increases as E_H increases), where there is a Landau level at the Fermi surface. The effect of the magnetic field on the potential well is thus described in terms of the magneto-oscillations of the bound charge (which arises from counting considerations) as described graphically in Fig. 13 and the dwell feature (which arises from the self-consistency aspects of the potential) by which an emptying or filling of a Landau level alters the potential in such direction as to oppose that change.

The seeming discontinuities in Fig. 14 are artifacts of the method of plotting against E_F/E_H as an abscissa: They would not appear if the curves were replotted against the physical variable E_H. The cause of the seeming discontinuities is that at $E_F/E_H = \frac{1}{2}$, $\frac{3}{2}$, etc., E_F changes rapidly, so that E_H does not depend smoothly on E_F/E_H. Near these special values of E_F/E_H, there is a large range of E_H compressed into a tiny interval of E_F/E_H. The quantities exhibited really depend on E_H, and where there is a change in E_H, there is also a change in these quantities, but this change is compressed into a very small region of E_F/E_H. This is clearly shown by comparing Fig. 14 with Fig. 15, where we have replotted E_b, the binding energy exhibited in Fig. 14, vs E_H, instead of vs E_G/E_H. The near-vertical portions of Fig. 14 have been stretched into gentle slopes in Fig. 15 and there is no discontinuity at the special values of E_H, marked by arrows, where $E_F/E_H = \frac{3}{2}$, $\frac{5}{2}$, etc.

The special values of E_H marked by the arrows are the values for which there is a bulk Landau level at the Fermi energy—a value for which one of the k_n [Eq. (3.6)] is zero. At these field values, there are discontinuities in some of the bulk response properties of the mobile electron system which arise because the bulk density of states becomes infinite.[13] (These singularities are often suppressed from theoretical calculations by including damping or broadening effects of various sorts.) Hence, one might expect that in the present calculation (in which there are no damping or broadening effects) the effect of the magnetic field on the mobile electrons might have marked consequences at these special field values. A close inspection of Fig. 15 does, in fact, reveal a slight change of slope at these field values. This is not a major structural feature of the curve and, with any reasonable broadening, would probably pass unnoticed altogether.

This feature does arise from the effect of the magnetic field on the mobile charge. There are, in fact, two aspects to consider here. The first is operative in the region very close to the surface where the mobile electron density is decreasing rapidly. In this region, the system of mobile electrons behaves like a collection of one-dimensional fermion gases, each gas with its own Fermi momentum k_n. As the magnetic field varies the size and number of these k_n, the electron density in

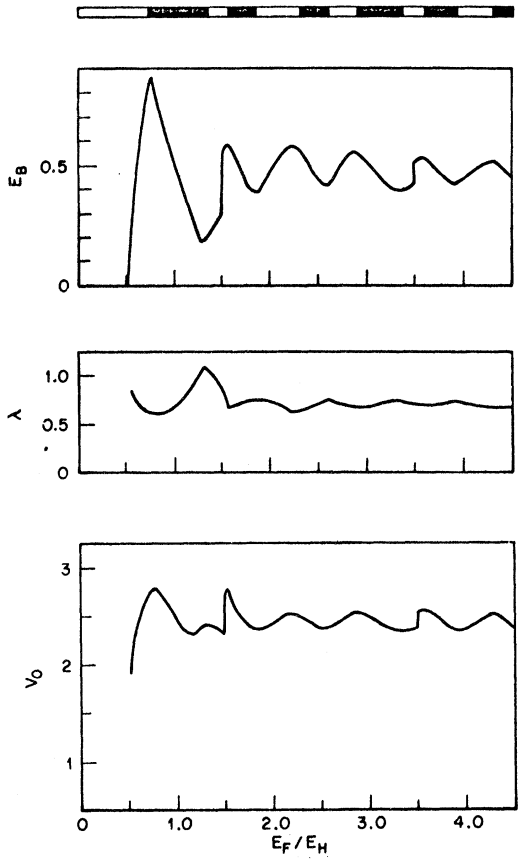

FIG. 14. Self-consistent well parameters V_0 and λ and binding energy of the state which the well supports plotted vs E_F/E_H for a fixed external electric field. In such a plot, oscillatory effects caused by bulk Landau-level effects will appear to be periodic with a period of one unit in E_F/E_H, with extremal points at $E_F/E_H = \frac{3}{2}, \frac{5}{2}, \frac{7}{2}$, etc. Oscillatory phenomena with a period other than unity and extrema at values other than these have to be attributed to some other phenomenon, in this case, to the Landau levels associated with the bound states. The horizontal bar at the top of the figure is blackened over these ranges of E_F/E_H for which a bound-state Landau level is dwelling at the Fermi surface. Note how the blacked portions correlate with the regions of negative slope in the binding-energy-vs-E_F/E_H curve.

this surface region undergoes magneto-oscillations. This is most easily seen by considering the zero-potential limit of (3.14), which we expand at small z assuming the phase shifts are zero. The result is

$$n(z) = (Cz^2 - 1)\, n_0 , \tag{7.19a}$$

where

$$C \equiv E_H \textstyle\sum_n k_n^3 . \tag{7.19b}$$

A plot of C vs E_H is given in Fig. 16 and it is clear that cusps appear at special values of E_H for which a bulk Landau level goes through the Fermi energy.

These magneto-oscillations in $n(z)$ will cause corresponding oscillations in the potential. A large density of charge near the surface tends to shield an attractive field and so the depth of the well is lessened at those values of E_H for which the cusps in C appear. Hence, we may expect this effect to contribute slope discontinuities in a plot of well depth or binding energy vs magnetic field, with the direction of the break always being toward lesser energy.

The effect we have been discussing is analogous to a magneto-oscillation of the driving charge in the region $z < d$ which we defined in the model describing the shape of the potential in the absence of any external field. That model also made use of the idea that screening occurs in the region $z > d$. Here, in the presence of a magnetic field, we can well imagine the screening length to be dependent on magnetic field and, if the bulk theories are any guide, the screening length should also show cusps as a function of E_H, with the shortest screening length at the cusp position. This effect should also contribute slope discontinuities in a plot of well depth vs magnetic field, with the direction of the break again always being towards lesser energy.

Again referring to Fig. 15, the slight breaks in slope are in the proper location and in the proper direction, but the effect is clearly far less important in determining the structure of the curve than are the magneto-oscillations which arise from the emptying and filling of a bound-state Landau level.

The final comment to be made about the magnetic field dependence of the potential concerns the extreme quantum region $\frac{1}{2} < E_F/E_H < \frac{3}{2}$, where there is but one occupied bulk Landau level. [This region covers fields greater than $E_H = (\frac{4}{9})^{1/3} = 0.763$.] At the onset of this region, there may be several filled bound-state Landau levels. As the magnetic field increases, these will empty in turn, giving rise to magneto-oscillations in the way we have described. Finally, at the highest fields, there must be only a single bound-state Landau level occupied. (Although the energy of that level increases indefinitely with field, it is always less than the Fermi energy, which, to provide a bulk density of carriers, must always be greater than the energy of the first bulk Landau level.) Hence, the amount of charge which the bound state then contains, $E_H/2\pi$, increases linearly with E_H. This causes the well to become shallower. Both the well depth V_0 and the binding energy E_B tend toward zero.

D. Comparison with Linear Screening Calculations

In Ref. 1(a), the effect of a magnetic field on the energy of a surface bound state was calculated in three ways, using three different approximations

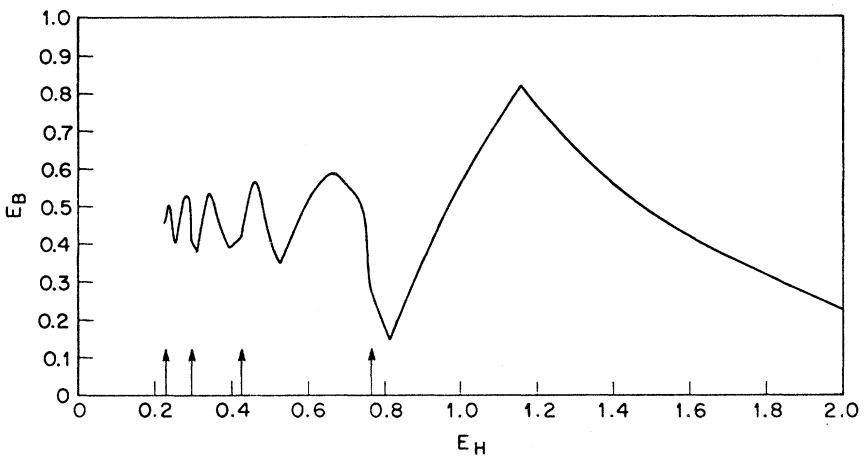

FIG. 15. Binding energy E_b from Fig. 14 plotted vs magnetic field strength E_H. The arrows are placed at values of E_H for which a bulk Landau level is at the Fermi energy, and the four shown correspond to the values $E_F/E_H = \frac{9}{2}$, $\frac{7}{2}$, $\frac{5}{2}$, and $\frac{3}{2}$. The well supports a second bound state (not shown) in the range $1.0 < E_H < 1.4$.

for the treatment of the mobile charge. All three approximations were made within a linear-response formalism, and were compared with each other. It is of some interest to compare the results of those calculations with what we have obtained here.

The first approximation was the bulk Thomas-Fermi approximation. In that case, we obtained a high density of states when $E_F/E_H = n + \frac{1}{2}$. This led to better shielding, hence a weakening of the binding energy at these values. However, the weakening of the binding energy was arbitrarily strong in the sense that the well would have disappeared altogether at these field values if we had not included lifetime broadening to limit the shielding. In the bulk Thomas-Fermi approximation, a high density of states occurred in the extreme quantum limit and hence, at very high magnetic field, screening by both surface states and mobile charge improved with magnetic field. Here, on the other hand, screening at high magnetic fields is provided mainly by the bound state. The well does not become as shallow nor does the binding energy drop so fast at high field as the bulk Thomas-Fermi approximation predicted.

The second approximation was the surface Thomas-Fermi approximation. It differed from the bulk Thomas-Fermi approximation in having zero mobile charge density at the surface. This gave rise to a response charge density which, very close to the surface, was independent of magnetic field and, further from the surface, exhibited the same sort of oscillations as did the bulk Thomas-Fermi approximation. In the extreme quantum limit, the mobile charge had a recovery distance which lengthened with field so that the mobile charge becomes less helpful in screening at high fields, leaving the screening to be done by the bound charge. The behavior of the mobile charge in the surface Thomas-Fermi approximation corresponded more closely to what we have described

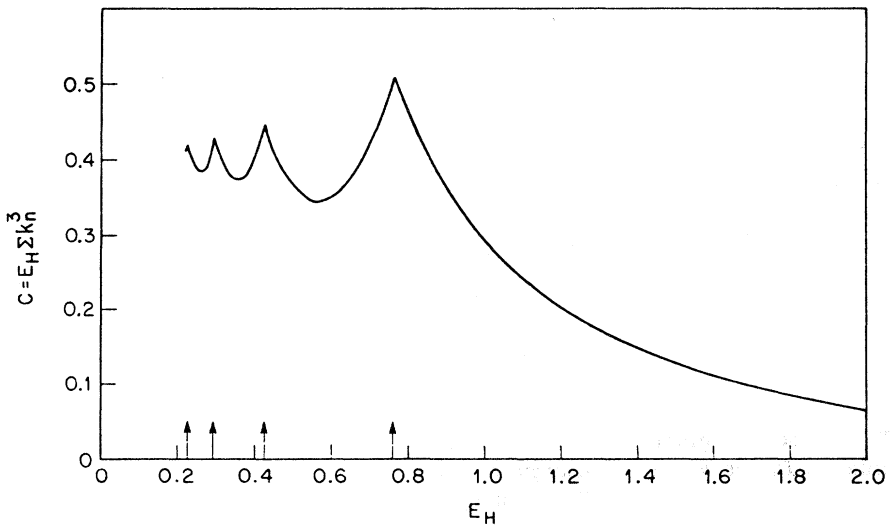

FIG. 16. Coefficient C of Eq. (7.18) plotted vs magnetic field E_H. In the limit of low fields, C goes to the value $\frac{2}{5}$.

in the present work. One therefore expects the surface Thomas-Fermi approximation to yield results very similar to what we find here. Comparing Fig. 15 of the present work with Fig. 6 of Ref. 1(a) (and dividing E_H there by $E_F^0 = 1.06$ to yield the units we use here), one does find close agreement (aside from the lack of low-field oscillations which were damped by the lifetime broadening introduced into the earlier calculation).

The third approximation of Ref. 1(a) was called the nonlocal screening approximation, and it apparently severely overestimates the effects of driving charge away from the surface in the extreme quantum limit. Its use, in the example chosen, leads to much poorer results than the surface Thomas-Fermi approximation.

ACKNOWLEDGMENTS

We should like to express our appreciation to V. Heine and to W. L. McMillan for useful discussions during the course of this work, to D. C. Tsui and J. M. Rowell for their interest and for a careful reading of this paper, and to Mrs. B. Chambers for assistance in programming.

APPENDIX A: CALCULATION OF FERMI ENERGY

From (3.15) and (3.6) we have

$$\frac{E_H}{2\pi^2}\sum_n [E_F - (n+\tfrac{1}{2})E_H]^{1/2} = n_0 \ . \tag{A1}$$

E_F here is a function of E_H. Let $E_H \to 0$. The sum becomes an integral which can be readily carried out, yielding

$$(1/3\pi^2)E_F(0)^{3/2} = n_0 \ , \tag{A2}$$

where $E_F(0)$ is the Fermi energy at zero magnetic field, i.e., 1 in the units we use here. We can combine (A1) and (A2) as

$$\sum_n \left(\frac{E_F}{E_H} - (n+\tfrac{1}{2})\right)^{1/2} = \frac{\frac{2}{3}}{(E_H)^{3/2}} \ . \tag{A3}$$

That is, given E_F/E_H, we evaluate the left-hand side of (A3), solve for E_H, and with that value of E_H and E_F/E_H we have E_F.

APPENDIX B: EVALUATION OF $\phi(0)$

We are concerned here with the evaluation of the integral on the right-hand side of Eq. (5.7). The bound-state contribution and that part of the mobile-state contribution which we have called $h(k)$ in Eq. (5.8) is trivially obtained. We merely state the result that

$$h(k) = \sum_{m=1}^{\infty} \frac{A_m(k)}{m^2\lambda^2} + \sum_{m=1}^{\infty} \frac{B_m(k)}{(m^2\lambda^2+4k^2)^2}\,[4mk\lambda\cos 2\eta(k) + (m^2\lambda^2 - 4k^2)\sin 2\eta(k)] + \sum_{m=1}^{\infty} \frac{C_m(k)}{(m^2\lambda^2+4k^2)^2} \times [(m^2\lambda^2 - 4k^2)\cos 2\eta(k) - 4mk\lambda \sin 2\eta(k)] , \tag{B1a}$$

where

$$A_m(k) \equiv \sum_{s=1}^{m-1} (Y_{m-s}\,Y_s + X_{m-s}\,X_s) + 2X_m \ , \tag{B1b}$$

$$B_m(k) = 2\sum_{s=1}^{m-1} Y_{m-s}X_s + 2Y_m \ , \tag{B1c}$$

$$C_m(k) = \sum_{s=1}^{m-1} (Y_{m-s}\,Y_s - X_{m-s}X_s) - 2X_m \ , \tag{B1d}$$

and the X_m and Y_m are the real and imaginary parts of the expansion coefficients appearing in (5.5). That is,

$$b_j(k) = X_j(k) + i\,Y_j(k) \ . \tag{B1e}$$

The terms included in (B1) all arise from short-range parts of the mobile charge density, that is, from parts which decay like $e^{-\lambda z}$, or faster. There is a long-range part which arises because

$$\xi_k^2(z) - 1 = -\cos 2(kz+\eta) + O(e^{-\lambda z}) \ ,$$

and its contribution to the right-hand side of (5.8) is

$$-\frac{1}{\pi}\sum_n \int_0^{\infty} z\,dz \int_0^{k_n} \cos 2[kz+\eta(k)]\,dk \equiv -\frac{1}{\pi}\sum_n \gamma_n(\infty) \ . \tag{B2}$$

The infinite upper limit here is to be interpreted in the sense of (2.19). Let us first replace the infinity by R, interchange orders of integration, and we obtain

$$\gamma_n(R) = \int_0^{k_n} \frac{dk}{4k^2}\,[\cos 2(kR+\eta) + 2kR\,\sin 2(kR+\eta) - \cos 2\eta] \ ,$$

which, upon integration by parts, becomes

$$\gamma_n(R) = -\left(\frac{\cos 2(kR+\eta) - \cos 2\eta}{4k}\right)_0^{k_n} - \frac{1}{2}\int_0^{k_n} \frac{dk}{k}\,\frac{d\eta}{dk}\,[\sin 2(kR+\eta) - \sin 2\eta] \ . \tag{B3}$$

In the integrated term we use Levinson's theorem again to write that, as $k \to 0$,

$$\eta(k) \to \pi M_b + k\eta' \ , \tag{B4a}$$

where

$$\eta' \equiv (d\eta/dk)_{k=0} \ . \tag{B4b}$$

Thus the integrated term is of order k as $k \to 0$ and there is no difficulty at the lower limit. At the upper limit, one of the integrated terms depends on R, and as R is taken to infinity, this term must be replaced by its average value, namely, zero.

Within the integral on the right-hand side of (B3),

the term which depends on R may be considered to be

$$k^{-1}\sin 2(kR+\eta)\ ,$$

which, at large R, oscillates so rapidly that it effectively limits the region of integration to the neighborhood of $k=0$ where we can make the replacement (B4a). That is, as $R\to\infty$, we have

$$\int_0^{k_n} dk\,\frac{d\eta}{dk}\,k^{-1}\sin 2(kR+\eta) \approx \int_0^{k_n} dk\,\eta'\,k^{-1}\sin 2k(R+\eta') = \tfrac{1}{2}\pi\eta'\ . \quad \text{(B5)}$$

Putting these results together, we are left with

$$\gamma_n(R\to\infty)=\frac{\cos 2\eta(k_n)}{4k_n}-\frac{\pi\eta'}{4}+\frac{1}{2}\int_0^{k_n}\frac{dk}{k}\,\frac{d\eta}{dk}\,\sin 2\eta(k) \quad \text{(B6)}$$

which completes the evaluation of (5. 8).

APPENDIX C: CONTINUITY OF $\phi(0)$

In order to continue the discussion of the Kohn-Majumdar theorem, we must develop a relation between the small-k behavior of the phase shifts $\eta(k)$ and the binding energy of the weakly bound state.

We have already made use of Levinson's theorem,

$$\eta(k=0)=\pi M_b\ , \quad \text{(C1)}$$

where M_b is the number of bound states. Now recall that the phase shifts themselves were calculated from (4. 3), which, using (5. 5), means that

$$R_k\,e^{-i\eta_l(k)}=\sum_{j=0}^{\infty} b_j(k)\equiv F(k)\ . \quad \text{(C2)}$$

There is a relation between the bound states and running-wave form of mobile states which arises from the fact that they both satisfy the same Schrödinger equation, in one case with $E=-K^2<0$, in the other case with $E=k^2>0$. If we replace k in (5. 5), the recursion relation for the running waves, by iK, we arrive at (5. 4), the recursion relation for the bound states. Thus, we have, to within a constant, that

$$F(iK)\equiv\sum_{j=0}^{\infty} b_j(iK)=\sum_{j=0}^{\infty} a_j(K)\ ,$$

so that the eigenvalue condition (5. 3c) which gives the bound-state energies can equally well be written

$$F(iK_m)=0\ . \quad \text{(C3)}$$

The form of (5. 3a) makes it apparent that K_m must be non-negative. It may happen that (C3) has a solution for $K<0$. Such a solution would result in a solution to Schrödinger's equation which grows exponentially with z, hence violates the condition $\psi(z\to\infty)\to 0$ and is therefore *not* an eigenstate. Nonetheless, the solutions to (C3), whether or not they correspond to an eigenstate of Schrödinger's equation, are of interest to us.

Suppose now that the potential is of such form that we have a weakly bound state. Then $F(iK_r)=0$ for a value of K_r which is small and positive. For k also small, we can expand $F(k)$ to first order, and, using (C2), we have

$$R_k\,e^{-i\eta(k)}=F(k)\approx F(iK_r)+(k-iK_r)F'(iK_r)+\cdots = (k-iK_r)F'(iK_r)\ .$$

Writing

$$k-iK_r=Be^{-i\xi}$$

and

$$F'(iK_r)=Ae^{-i\varphi}\ ,$$

where A, B, φ, and ξ are real, we have

$$\eta(k)=\varphi+\xi=\phi+\tan^{-1}(K_r/k)\ , \quad \text{(C4)}$$

and ϕ is independent of k. Let $k\to 0+$. Then, since $K_r>0$, $\tan^{-1}(K_r/k)\to\frac{1}{2}\pi$ and $\eta(0)=\varphi+\frac{1}{2}\pi$, which, using (C1), means that

$$\varphi=(M_b-\tfrac{1}{2})\,\pi\ .$$

Hence,

$$\eta(k)=(M_b-\tfrac{1}{2})\,\pi+\tan^{-1}(K_r/k)$$

or

$$\eta(k)=M_b\pi-\tan^{-1}(k/K_r)\ , \quad \text{(C5)}$$

$$\tan\eta(k)=-k/K_r\ . \quad \text{(C6)}$$

The form is to be used only for k and K_r both small. Now suppose that the potential changes in such a way that the binding energy of the weakly bound state is reduced. Then K_r approaches zero. Further change of the potential in the same direction takes K_r to zero, whereupon the bound-state energy becomes zero and the spatial extent of the state becomes infinite. Still further change of the potential in the same direction drives K_r still further in the same direction, i. e., onto negative values, but as we mentioned earlier, there is no eigenstate corresponding.

The number of bound states is one less than before, i. e., $M_b\to M_b-1$, but the same sort of analysis as before holds (even with $K_r<0$) and we are again led back to (C5) and (C6). Thus, whatever the sign of K_r, we have for small k and K_r

$$\tan\eta(k)=-k/K_r\ , \quad \text{(C7a)}$$

$$d\eta/dk=-K_r/(K_r^2+k^2)\ , \quad \text{(C7b)}$$

$$\eta'=-1/K_r\ . \quad \text{(C7c)}$$

We now turn to the study of Eq. (5.8) in the neighborhood of a region where the number of bound states changes. Suppose first that we have a weakly bound state and that we are changing the external conditions in such a way that the potential weakens, reducing the energy of this state still further. The normalization condition (5.4d) for this state (call it the Bth state) can be written

$$\frac{a_0^2(K_B)}{2K_B} + \sum_{r+s>0} \frac{a_r(K_B)\,a_s(K_B)}{2K_B+(r+s)\lambda} = 1 ,$$

which gives us

$$\lim a_0^2(K_B) = 2K_B \quad \text{as} \quad K_B \to 0 .$$

Accordingly, the contribution of this state to the right-hand side of (5.8) is, in the limit,

$$[-2\pi\phi(0)/KE_H]_B = N(E_F)/2K_B , \quad K_B > 0 . \qquad \text{(C8)}$$

That is, as the binding energy of the state drops and it becomes more and more spread out, the electrons on the average are further out and contribute more to the potential. In the limit that $K_B \to 0$, this piece becomes singular. All other parts of the bound-state contribution to (5.8) are continuous as $K_B \to 0$. However, there are mobile-state contributions to (5.8) which are also singular when $K_B \to 0$. The first of these, using (C7c), is

$$\frac{1}{\pi}\sum_n \frac{\pi}{4}\,\eta' = -\frac{1}{4K_B}\sum_n 1 = -\frac{N(E_F)}{4K_B} . \qquad \text{(C9)}$$

Unlike (C8), this piece is present for both signs of K_B. The other singular term in (5.8) is the integral involving $d\eta/dk$. The singularity in the integral occurs at its $k \to 0$ end, and hence we can use the small-k form, Eq. (C7), for the purposes of evaluating that singularity. Thus,

$$\frac{1}{\pi}\sum_n \frac{1}{2}\int_0^{k_n} k^{-1}\,\frac{d\eta}{dk}\,\sin 2\eta(k)\,dk$$

$$\approx \frac{1}{\pi}\sum_n \int_0^{k_n} \frac{dk}{K_B^2(1+k^2/K_B^2)^2}$$

$$= \frac{1}{|\pi K_B|}\sum_n \int_0^{|k_n/K_B|} \frac{du}{(1+u^2)^2} .$$

As $K_B \to 0$, the integral goes to $\frac{1}{4}\pi$, and this singular contribution to (5.8) is

$$-\frac{1}{|4K_B|}\sum_n 1 = -\frac{N(E_F)}{|4K_B|} . \qquad \text{(C10)}$$

In (C8)–(C10), we have isolated all the pieces of (5.8) which are either discontinuous or singular as $K_B \to 0$. The sum of these three pieces is

$$\frac{N(E_F)}{2K_B} - \frac{N(E_F)}{4K_B} - \frac{N(E_F)}{|4K_B|} = 0 \quad \text{if} \quad K_B > 0 , \qquad \text{(C11a)}$$

$$-\frac{N(E_F)}{4K_B} - \frac{N(E_F)}{|4K_B|} = 0 \quad \text{if} \quad K_B < 0 , \qquad \text{(C11b)}$$

which shows that as the parameters of the potential are changed infinitesimally in such a way as to expel a bound state from the well, the iterated potential (i.e., the first moment of the charge density) suffers no singularity or discontinuity.

Let us dwell for a moment on the implications of (C11). In (C11a), there is an near-infinite contribution associated with the bound state (the first term), a contribution which clearly arises because the electrons contained in the bound state are located further from the surface as the binding weakens. Yet this near infinity is canceled by a mobile charge contribution [the sum of the last two terms in (C11a)], which implies a corresponding deficit of mobile charge, i.e., a hole carved out of the mobile charge distribution. The hole arises, presumably, because bound and mobile states are orthogonal. In (C11b), the bound state has disappeared, and the two remaining terms, whose sum is the mobile charge contribution, also vanish.

---

[1] Joel A. Appelbaum and G. A. Baraff, (a) Phys. Rev. B 4, 1235 (1971) (this paper contains references to earlier work on this problem, which we omit reviewing here); (b) *ibid.* 4, 1246 (1971).

[2] Joel A. Appelbaum and G. A. Baraff, Phys. Rev. Letters 26, 1432 (1971).

[3] The validity of this approach just at the surface has been questioned; in particular, our choice of boundary conditions on the wave function at the surface is not the only one conceivable. It would take us too far afield to discuss this in detail here, but our preliminary investigation suggests that the procedure we used here is basically correct.

[4] J. R. Schrieffer, in *Semiconductor Surface Physics*, edited by R. A. Kingston (Pennsylvania U. P., Philadelphia, 1957), p. 55; F. Stern and W. E. Howard, Phys. Rev. 163, 816 (1967); C. B. Duke, *ibid.* 159, 632 (1967); D. C. Tsui, Phys. Rev. Letters 24, 303 (1970); Phys. Rev. B 4, 4438 (1971).

[5] Parametric self-consistency, in which the charge density is parametrized, has been used by Alan J. Bennett and C. B. Duke, Phys. Rev. 160, 541 (1967).

[6] J. Friedel, Phil. Mag. 43, 153 (1952); Nuovo Cimento Suppl. 7, 287 (1958).

[7] W. Kohn and C. Majumdar, Phys. Rev. 138, A1617 (1965).

[8] N. Lang and W. Kohn, Phys. Rev. B 1, 4555 (1970).

[9] Use of boundary condition (3.2) is, of course, merely a counting procedure. If one were to insist (unnecessarily) that self-consistency be maintained at $z = L$, then use of the asymptotic form (3.1) would require that an external charge distribution be present near $z = L$ in order that the total charge density (mobile plus background plus external) be identically zero there. Such an external charge will play no physical role; nor will it contribute to the sum rule we are deriving here.

[10] See, for example, C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963), p. 34.

[11]N. Levinson, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. 25, No. 9 (1949).
[12]W. J. Swiatacki, Proc. Phys. Soc. (London) A64, 226 (1951).
[13]A. H. Kahn and H. P. R. Freiderikse, in *Solid State Physics*, edited by F. Steitz and D. Turnbull (Academic, New York, 1959), Vol. 9, p. 257.

PHYSICAL REVIEW B VOLUME 5, NUMBER 2 15 JANUARY 1972

Interband Transitions and Exciton Effects in Semiconductors*

M. Welkowsky and R. Braunstein
Department of Physics, University of California, Los Angeles, California 90024
(Received 19 July 1971)

The band structures of Ge, Si, GaAs, GaP, GaSb, InAs, InP, InSb, and AlSb have been studied in reflectivity in the energy region 1.6–5.0 eV at temperatures ranging from 80 to 300 °K. Utilizing a double-beam, single-detector wavelength-modulation system, and ensuing Kramers-Krönig analyses, experimentally unambiguous line shapes have been obtained for the real and imaginary part of the dielectric constants, permitting the identification of the types of critical points involved in an optical transition, and the determination of the existence of hyperbolic exciton interactions. Such an interaction has been verified in all materials, except Si, as an M_1 critical point located at Λ in the Brillouin zone. The location and energy of the interband transition in these semiconductors correlates with existing band calculations. The interband transitions in Si are dominated by structure from a large region of the Brillouin zone. The high-energy E_2 transitions in all materials give evidence of a multiplicity of critical-point structure.

I. INTRODUCTION

The band structures of Ge, Si, GaAs, GaP, GaSb, InAs, InP, InSb, and AlSb have been studied in reflectivity in the energy region 1.6–5.0 eV at temperatures ranging from 80 to 300 °K. Utilizing a double-beam, single-detector wavelength-modulation system and ensuing Kramers-Krönig analyses, experimental line shapes have been obtained for the derivative of the real and imaginary parts of the dielectric constant. These analyses enable the identification of the types of critical points involved in the optical transition as well as the existence of hyperbolic exciton interactions.

Reflectivity derivative data using wavelength modulation[1] have been previously used to obtain interband transition energies of several of these semiconductors.[2–6] Shaklee *et al.*[2] observed the reflectance derivative spectra of InSb in the neighborhood of the E_1 and $E_1+\Delta_1$ structure; by studying the line shapes of the spectra, they obtained evidence for the contribution of exciton effects to the observed transitions. Evidence for the existence of hyperbolic excitons in GaAs by means of the polarization-dependent splitting of the E_1 and $E_1+\Delta_1$ structure was obtained by Rowe *et al.*[3] Wavelength-modulated reflectance measurements on the $InAs_{1-x}P_x$ alloys were studied by Thompson *et al.*[4] in the spectral region of the E_1 and $E_1+\Delta_1$ transitions. Zucca and Shen[5] observed the wavelength-modulation spectra of GaAs, GaSb, InAs, InSb, Ge, and Si and correlated the spectra with existing band calculations for these crystals. From the sharpening of the spectra with reduced temperature due to the reduction of lifetime broadening, it was indicated that hyperbolic excitions are associated with the E_1 peaks. Since Kramers-Krönig analyses were not performed on these reflectance data to obtain the derivative of ϵ_1 and ϵ_2, a direct comparison of line shapes with theory was not possible. Further evidence for excitonic effects on the E_1 and $E_1+\Delta_1$ transition in InAs was obtained by Antoci *et al.*[7] by studying the temperature dependence of the line shapes of the E_1 and $E_1+\Delta_1$ structure using thermoreflectance. The band structure features of Ge, Si, GaAs, GaSb, GaP, InAs, InSb, InP, and AlSb were studied by the present authors[6] in the spectral region 1.62–5.08 eV as a function of temperature using a double-beam single-detector wavelength-modulated reflectance system. Structures corresponding to various critical points were seen and related to existing band calculations. The above observations using wavelength-modulation techniques have all indicated that the reflectance spectrum in the E_1 region cannot be explained within the framework of the one-electron approximation with lifetime broadening corrections, and suggest that Coulomb interaction should be included to explain the observed structure. The prior work on these substances have involved either a number of substances in a restrictive spectral region or in most cases have reported only the derivative of the re-