Many-body effects in n-type Si inversion layers. I. Effects in the lowest subband

B. Vinter
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 19 November 1975)

Some many-body effects on the electrons in the n inversion layer on the Si (100) surface of the metal-oxide-semiconductor structure have been calculated. Screening was treated in the Lundqvist-Overhauser approximation. In this paper we report calculations on the exchange and correlation energies and effective mass of the electrons in the lowest subband both in the limit of a two-dimensional interacting electron gas and for a finite thickness of the layer. Corrections due to finite oxide thickness and dispersion in the insulator have been investigated and found to have a very small influence on the effective mass. Finally, contributions from the electron-phonon interaction are estimated by deformation-potential theory and found to be negligible.

I. INTRODUCTION

Near the semiconductor surface in a metal-insulator-semiconductor structure the bands are bent. The amount of bending can be controlled by the gate voltage, and on p-type substrate an inversion layer of electrons can be formed. The potential well which confines the electrons in the direction perpendicular to the interface is so narrow that quantum effects occur. The properties of the inversion layer have been studied extensively both theoretically and experimentally.[1]

In this work we are concerned with the effects of electron-electron interactions in the inversion layer on a Si (100) surface. The metal-insulator-semiconductor (MIS) system is particularly interesting for such a study since the density of electrons can be varied continuously and in a controlled manner by simply changing the gate voltage. This paper treats many-body effects in the lowest subband both in the limit of an infinitely thin layer, in which case the system is a two-dimensional electron gas, and when the finite thickness of the layer is taken into account. Our results give the correlation energy and effective mass of an electron in the lowest subband. Apart from the direct electron-electron interaction, contributions to those quantities from the finite thickness and the dispersion of the insulator and the electron-phonon interaction are investigated but found to be quite small. In a later article we shall study effects which involve the higher subbands.

The basis of the calculations is the self-consistent results in the Hartree approximation obtained by Stern and Howard[2] and Stern.[3] They showed that within the effective-mass approximation the six valleys of Si in the bulk give rise to two ladders of subbands. The two valleys, which have their longitudinal mass perpendicular to the interface, lead to a set of subbands in which the wave functions are given by

$$\psi_{n,\vec{k}}(\vec{r}) = \zeta_n(z)\exp(i\vec{k}\cdot\vec{R}), \tag{1}$$

where $\vec{k}$ and $\vec{R}$ are two-dimensional vectors parallel to the interface and z is the coordinate perpendicular to the interface. n labels the subbands. The one-electron energy corresponding to the wave function (1) is

$$\epsilon_n(\vec{k}) = E_n + k^2/2m, \tag{2}$$

where m is the transverse mass of Si. E_n and $\zeta_n(z)$ are determined by a self-consistent solution of a one-dimensional Schrödinger equation and Poisson's equation.

The other four valleys give rise to another ladder of subbands with a structure similar to (1) and (2) but with different energies and a heavier mass for motion parallel to the surface. The lowest subband of that ladder has higher energy than the lowest subband of the doubly degenerate ladder. Throughout this work we shall work at absolute temperature $T=0$ and assume that the number of electrons in the inversion layer is so small that only the lowest subband is occupied in the ground state.

II. TWO-DIMENSIONAL LIMIT

As the wave function (1) shows, the electrons move freely parallel to the interface. The thickness of the inversion layer when only the lowest subband is occupied is of the order of 30 Å. For a first approximation we study the limit of zero thickness, i.e., $\zeta_0(z) \cong \delta(z)$. In that limit we have a completely two-dimensional electron gas and the methods which have been developed for a three-dimensional gas can be directly applied with proper changes. In particular we introduce the one-particle Green's function $G(\vec{k}, E)$,[4] where $\vec{k}$ is a two-dimensional wave vector parallel to the interface.

The one-particle Green's function satisfies the Dyson equation

$$G(\vec{k},E)=G_0(\vec{k},E)+G_0(\vec{k},E)\,M(\vec{k},E)\,G(\vec{k},E), \quad (3)$$

where G_0 is the Green's function for the noninteracting electron gas

$$G_0(\vec{k},E)=\frac{\theta(|\vec{k}|-k_F)}{E-\xi(\vec{k})+i\delta}+\frac{\theta(k_F-|\vec{k}|)}{E-\xi(\vec{k})-i\delta}, \quad (4)$$

with $\xi(\vec{k})=(k^2-k_F^2)/2m$, where m is the bulk effective mass for motion parallel to the surface, and $\delta=0+$.

In the random-phase approximation (RPA) the self-energy $M(\vec{k},E)$ is given by

$$M(\vec{k},E)=i\int\frac{d^2q\,d\omega}{(2\pi)^3}\,\frac{e^2}{2\epsilon_\infty q}\,\frac{\epsilon_\infty}{\epsilon(q,\omega)}\,G_0(\vec{k}-\vec{q},E-\omega), \quad (5)$$

where $e^2/2\epsilon_\infty q$ is the bare Coulomb interaction and ϵ_∞ is the effective permittivity of the surrounding medium. Its value and dependence on q are discussed later in this paper. The permittivity function $\epsilon(q,\omega)$ given by

$$\frac{\epsilon(q,\omega)}{\epsilon_\infty}-1=-\frac{e^2}{2\epsilon_\infty q}\,\tilde{\chi}(q,\omega)$$
$$=-\frac{e^2}{2\epsilon_\infty q}\,2n_v i\int\frac{d^2k\,dE}{(2\pi)^3}\,G_0(\vec{k}+\vec{q},E+\omega)\times G_0(\vec{k},E) \quad (6)$$

was evaluated by Stern.[5] n_v is the number of degenerate valleys in the bulk to which the lowest subband corresponds. For the (100) surface of Si, $n_v=2$.

The quasiparticle energy $E(k)$ can then be found as the pole of the Green's function $G(k,E)$:

$$E+\mu=\xi(k)+M(k,E), \quad (7)$$

where E is measured relative to the chemical potential μ, which is determined by setting $E=0$ and $k=k_F$ in the above equation.[6]

One further simplification is introduced. Instead of the RPA dielectric function [Eq. (6)] we use the plasmon-pole approximation suggested by Lundqvist[7] and Overhauser.[8] To this end we separate the screened interaction into an unscreened term which gives rise to the exchange part of the self-energy and a term which involves the coupling to density fluctuations:

$$\frac{e^2}{2\epsilon_\infty q}\,\frac{\epsilon_\infty}{\epsilon(q,\omega)}=\frac{e^2}{2\epsilon_\infty q}+\frac{e^2}{2\epsilon_\infty q}\left(\frac{\epsilon_\infty}{\epsilon(q,\omega)}-1\right). \quad (8)$$

The latter term is replaced by coupling to one effective plasmon

$$\mathrm{Im}\left(\frac{\epsilon_\infty}{\epsilon(q,\omega)}-1\right)=\alpha\,\delta(\omega-\omega_q), \quad (9)$$

where the energy of the plasmon ω_q and the coupling α are determined by the requirement that the f sum rule[5]

$$\int_0^\infty\omega\,\mathrm{Im}\left(\frac{\epsilon_\infty}{\epsilon(q,\omega)}-1\right)d\omega=-\tfrac{1}{2}\pi\,\omega_P^2(q), \quad (10)$$

and the zero-frequency Kramers-Kronig relation

$$\int_0^\infty\frac{1}{\omega}\,\mathrm{Im}\left(\frac{\epsilon_\infty}{\epsilon(q,\omega)}-1\right)d\omega=\frac{\pi}{2}\left(\frac{\epsilon_\infty}{\epsilon(q,0)}-1\right) \quad (11)$$

be fulfilled, where $\omega_P^2(q)=\frac{1}{2}qNe^2/\epsilon_\infty m$. We then have

$$\alpha=-\tfrac{1}{2}\pi\,\omega_P^2/\omega_q,$$
$$\omega_q^2=-\frac{\omega_P^2}{\epsilon_\infty/\epsilon(q,0)-1}, \quad (12)$$

$$\frac{\epsilon_\infty}{\epsilon(q,\omega)}-1=\frac{\omega_P^2}{2\omega_q}\,\frac{2\omega_q}{\omega^2-\omega_q^2+i\delta}. \quad (13)$$

For $\epsilon(q,0)$ we use the RPA result of Stern[5]

$$\epsilon(q,0)/\epsilon_\infty=1+s(q)/q$$
$$s(q)=(2n_v e^2m/4\pi\epsilon_\infty)\{1-\theta(q-2k_F)\times[1-(2k_F/q)^2]^{1/2}\}. \quad (14)$$

With these approximations we can integrate over ω and obtain the self-energy

$$M(k,E)=M_x(k)+M_c(k,E), \quad (15)$$

$$M_x(k)=\int\frac{d^2q}{(2\pi)^2}\,\frac{e^2}{2\epsilon_\infty q}\,\theta(k_F-|\vec{k}-\vec{q}|), \quad (16)$$

$$M_c(k,E)=\int\frac{d^2q}{(2\pi)^2}\,\frac{e^2}{2\epsilon_\infty q}\,\frac{\omega_P^2}{2\omega_q}\times\left(\frac{\theta(k_F-|\vec{k}-\vec{q}|)}{E-\xi(\vec{k}-\vec{q})+\omega_q}+\frac{\theta(|\vec{k}-\vec{q}|-k_F)}{E-\xi(\vec{k}-\vec{q})-\omega_q}\right). \quad (17)$$

Since the integrands are functions of the magnitudes of $\vec{q}$ and of $\vec{k}-\vec{q}$, it is natural to introduce $q=|\vec{q}|$ and $u=(\vec{k}-\vec{q})^2$ as integration variables:

$$M_x(k)=-\int_0^\infty dq\,\frac{e^2}{(2\pi)^2\epsilon_\infty}\int_{(k-q)^2}^{(k+q)^2}\frac{du\,\theta(k_F^2-u)}{\{-[u-(k-q)^2][u-(k+q)^2]\}^{1/2}}, \quad (18)$$

$$M_c(k,E)=-\int_0^\infty dq\,\frac{2e^2m}{(2\pi)^2\epsilon_\infty}\,\frac{\omega_P^2}{\omega_q}\int_{(k-q)^2}^{(k+q)^2}\frac{du}{\{-[u-(k-q)^2][u-(k+q)^2]\}^{1/2}}\left(\frac{\theta(k_F^2-u)}{u+p_+}+\frac{\theta(u-k_F^2)}{u+p_-}\right), \quad (19)$$

where

$$p_\pm = -2m(E \pm \omega_q) - k_F^2 .$$

The integrations over u can be performed analytically, and we have

for $k < k_F$:

$$M_x(k) = -\int_0^{k_F-k} \frac{e^2}{4\pi\epsilon_\infty}\, dq - \int_{k_F-k}^{k_F+k} \frac{e^2}{(2\pi)^2\epsilon_\infty} \left[\frac{\pi}{2} - \arcsin\left(\frac{k^2+q^2-k_F^2}{2kq}\right)\right] dq , \tag{20}$$

$$\begin{aligned} M_c(k,E) = &-\int_0^{k_F-k} dq \frac{2me^2}{(2\pi)^2\epsilon_\infty} \frac{\omega_P^2}{\omega_q} H((k+q)^2, k, q, p_+) \\ &- \int_{k_F-k}^{k_F+k} dq \frac{2me^2}{(2\pi)^2\epsilon_\infty} \frac{\omega_P^2}{\omega_q} [H(k_F^2, k, q, p_+) + H((k+q)^2, k, q, p_-) - H(k_F^2, k, q, p_-)] \\ &- \int_{k_F+k}^{\infty} dq \frac{2me^2}{(2\pi)^2\epsilon_\infty} \frac{\omega_P^2}{\omega_q} H((k+q)^2, k, q, p_-) . \end{aligned} \tag{21}$$

For $k > k_F$,

$$M_x(k) = -\int_{k-k_F}^{k+k_F} \frac{e^2}{(2\pi)^2\epsilon_\infty} \left[\frac{\pi}{2} - \arcsin\left(\frac{k^2+q^2-k_F^2}{2kq}\right)\right] dq , \tag{22}$$

$$\begin{aligned} M_c(k,E) = &-\int_0^{k-k_F} dq \frac{2me^2}{(2\pi)^2\epsilon_\infty} \frac{\omega_P^2}{\omega_q} H((k+q)^2, k, q, p_-) \\ &- \int_{k-k_F}^{k+k_F} dq \frac{2me^2}{(2\pi)^2\epsilon_\infty} \frac{\omega_P^2}{\omega_q} [H(k_F^2, k, q, p_+) + H((k+q)^2, k, q, p_-) - H(k_F^2, k, q, p_-)] \\ &- \int_{k+k_F}^{\infty} dq \frac{2me^2}{(2\pi)^2\epsilon_\infty} \frac{\omega_P^2}{\omega_q} H((k+q)^2, k, q, p_-) . \end{aligned} \tag{23}$$

The function H is given by, for $p < -(k+q)^2$,

$$H(x,k,q,p) = \frac{-1}{\{[p+(k-q)^2][p+(k+q)^2]\}^{1/2}} \left[\arcsin\left(\frac{[x-(k^2+q^2)](p+k^2+q^2)+4k^2q^2}{2kq(p+x)}\right) + \frac{\pi}{2}\right];$$

for $-(k+q)^2 < p < -(k-q)^2$,

$$\begin{aligned} H(x,k,q,p) = &\frac{-1}{\{-[p+(k-q)^2][p+(k+q)^2]\}^{1/2}} \\ &\times \ln \left| \frac{\{[p+(k-q)^2][p+(k+q)^2][x-(k-q)^2][x-(k+q)^2]\}^{1/2} + [x-(k^2+q^2)](p+k^2+q^2)+4k^2q^2}{2kq(p+x)} \right| ; \end{aligned} \tag{24}$$

for $-(k-q)^2 < p$,

$$H(x,k,q,p) = \frac{1}{\{[p+(k-q)^2][p+(k+q)^2]\}^{1/2}} \left[\arcsin\left(\frac{[x-(k^2+q^2)](p+k^2+q^2)+4k^2q^2}{2kq(p+x)}\right) + \frac{\pi}{2}\right].$$

By differentiation of the Dyson equation (7) we find the effective mass $m^* = k(dE/dk)^{-1}$:

$$\frac{m^*}{m} = \left(1 - \frac{\partial M}{\partial E}\right) \Big/ \left(1 + \frac{m}{k}\frac{\partial M}{\partial k}\right), \tag{25}$$

where the derivatives are to be evaluated at $E = 0$, $k = k_F$. The derivatives were also calculated as one-dimensional integrals. The rather complicated integrands will not be shown.

III. RESULTS IN THE TWO-DIMENSIONAL LIMIT

All calculations of this work have been done with the parameters corresponding to the SiO_2 – Si(100) interface. A list of these parameters is given in Table I. From a simple electrostatic consideration it follows that, in the two-dimensional limit where the electrons are right at the interface, the permittivity of the surrounding medium is the average of the permittivities of SiO_2 and Si.

In Fig. 1 we show the Fermi energy, exchange energy, correlation energy, and self-energy as a function of inversion-layer concentration. The exchange energy was calculated earlier by Chaplik[9] and Stern[10] and increases as the square root of concentration. The correlation energy is also negative, so the magnitude of the energy is larger than the exchange energy alone. This is in direct contrast to what would be obtained in a static screening approximation and demonstrates that the dynamical features of the screening are important.

For momenta different from the Fermi momentum the correlation energy changes in such a way that the variation with k of the exchange energy is almost cancelled. The resulting self-energy found by solving the Dyson equation varies very little with k, as shown in Fig. 2 for an inversion-layer density of 10^{12} cm^{-2}. Thus the quasiparticle energy band is very nearly parabolic. The mean mass $\bar{m}$ defined by $k_F^2/2\bar{m} = E(k_F) - E(0)$ is changed somewhat from the noninteracting value. These results are similar to what is seen in three dimensions both in the plasmon-pole approximation[7,8] and in the RPA.[6]

Figure 3 shows the mean mass as a function of density and the effective mass calculated from Eq. (25). For densities below 5×10^{13} cm^{-2} the electron-electron interaction enhances the effective mass, whereas it is lower than the bulk value for higher densities. For $n_v = 1$ and very high densities our results coincide with those of Janak[11] who used the static approximation and those of Chaplik.[9] It should be noted, though, that in the real system higher subbands will be populated at densities above ~5×10^{12} cm^{-2} so the model is not applicable there.

TABLE I. Parameters used in the calculations.

Valley degeneracy	n_v	2
Mass parallel to surface	m	$0.1905m_e$
Permittivity of Si	ϵ_s	$11.7\epsilon_0$
Permittivity of SiO_2	ϵ_i	$3.9\epsilon_0$
Average permittivity	ϵ_∞	$7.8\epsilon_0$

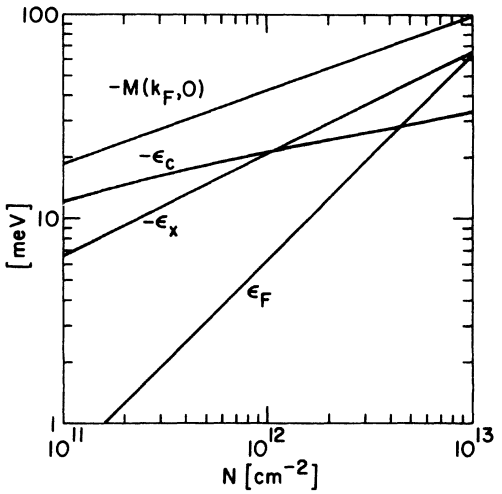

FIG. 1. Fermi energy ϵ_F, exchange energy ϵ_x, correlation energy ϵ_c, and self-energy M evaluated at the Fermi level as functions of inversion-layer density in the two-dimensional limit.

IV. THEORY FOR FINITE LAYER THICKNESS

The finite thickness of the inversion layer can be taken into account rather straightforwardly. The envelope wave function for the noninteracting electrons is

$$\psi(\vec{r}) = \zeta_0(z)\exp(i\vec{k}\cdot\vec{R}). \tag{26}$$

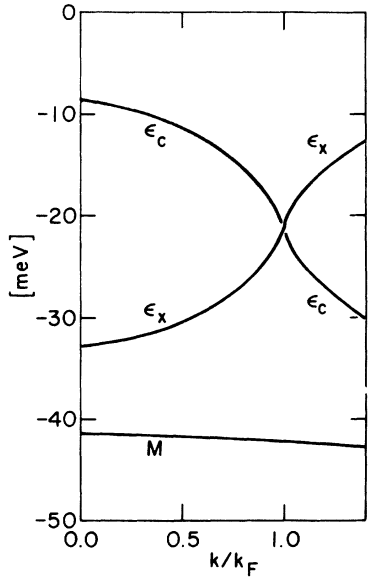

FIG. 2. Exchange energy ϵ_x, correlation energy ϵ_c, and self-energy M as functions of wave vector in the two-dimensional limit. $N_{inv} = 10^{12}$ cm^{-2}.

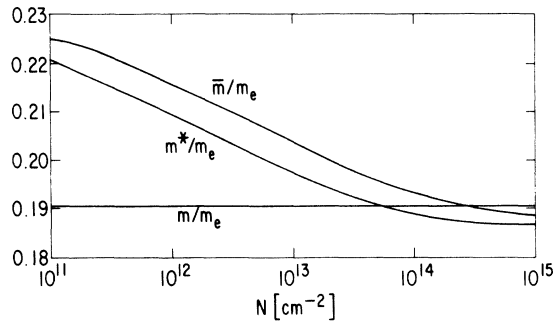

FIG. 3. Mean mass $\bar{m}[k_F^2/2\bar{m}=E(k_F)-E(0)]$, and effective mass m^* as functions of inversion-layer density in the two-dimensional limit. m is the bulk mass for motion parallel to the interface, and m_e is the free-electron mass.

The potential $\phi(\vec{R}, z; z_0)$ created by an electron at $(0, 0, z_0)$ is found by solving Poisson's equation

$$\nabla^2\phi = \frac{e}{\epsilon_s}\delta(\vec{R})\delta(z - z_0) \quad \text{for } z > 0,$$
$$\nabla^2\phi = 0 \quad \text{for } z < 0, \tag{27}$$

with the boundary conditions

$$\phi(\vec{R}, 0+; z_0) = \phi(\vec{R}, 0-; z_0),$$
$$\epsilon_s \left.\frac{\partial\phi}{\partial z}\right|_{z=0+} = \epsilon_i \left.\frac{\partial\phi}{\partial z}\right|_{z=0-},$$
$$\phi \to 0 \quad \text{for } |z| \to \infty,$$

where ϵ_s and ϵ_i are the permittivities of the semiconductor and insulator, respectively. Upon Fourier analysis in the directions parallel to the interface, we obtain the solution for $z > 0$:

$$\phi(\vec{R}, z; z_0) = \sum_{\vec{q}} \phi_{\vec{q}}(z; z_0) e^{i\vec{q}\cdot\vec{R}}, \tag{28}$$

$$\phi_{\vec{q}}(z; z_0) = \frac{-e}{2\epsilon_s q}\left(e^{-q|z-z_0|} + \frac{\epsilon_s - \epsilon_i}{\epsilon_s + \epsilon_i} e^{-q(z+z_0)}\right).$$

The bare Coulomb interaction between two electrons in the lowest subband having momentum difference $\vec{q}$ is then

$$v(\vec{q}) = \frac{e^2}{2\epsilon_s q} f(q),$$

$$f(q) = \int_0^\infty \int_0^\infty dz\, dz' \, |\zeta_0(z)|^2 \, |\zeta_0(z')|^2 \times \left(e^{-q|z-z'|} + \frac{\epsilon_s - \epsilon_i}{\epsilon_s + \epsilon_i} e^{-q(z+z')}\right). \tag{29}$$

This reduced interaction, which has been derived and used by several authors,[12-14] can be viewed as the Coulomb interaction between two electrons confined to a plane embedded in a medium with a q-dependent effective permittivity

$$\epsilon_\infty(q) = \epsilon_s / f(q). \tag{30}$$

The RPA screening and the Lundqvist-Overhauser approximation then follow from the expressions in the two-dimensional limit (Sec. II) by using this permittivity instead of $\epsilon_\infty = (\epsilon_s + \epsilon_i)/2$. A more rigorous justification for this will follow from the more general theory to be introduced in part II of this work. See also Siggia and Kwok.[15]

V. RESULTS

The simplest wave functions to use in (29) is the variational wave function applied by Fang and Howard[16]

$$\zeta(z) = (b^3/2)^{1/2} e^{-bz/2}. \tag{31}$$

b is determined by minimizing the total energy.[2] For this wave function $f(q)$ is given by

$$f(q) = \frac{8 + 9x + 3x^2}{8(1+x)^3} + \frac{\epsilon_s - \epsilon_i}{\epsilon_s + \epsilon_i} \frac{1}{(1+x)^6}, \tag{32}$$

where $x = q/b$.

In Fig. 4 are shown the exchange, correlation, and self-energies as functions of inversion-layer concentration. Dashed curves show the results when the variational wave function is used. The values $z_{av} = 3/b$ are those found[3] for a bulk doping of $N = 7\times10^{14}$ cm^{-3}. The full lines show results when we use the Hartree wave functions calculated numerically by Stern.[3,17] In neither case is the

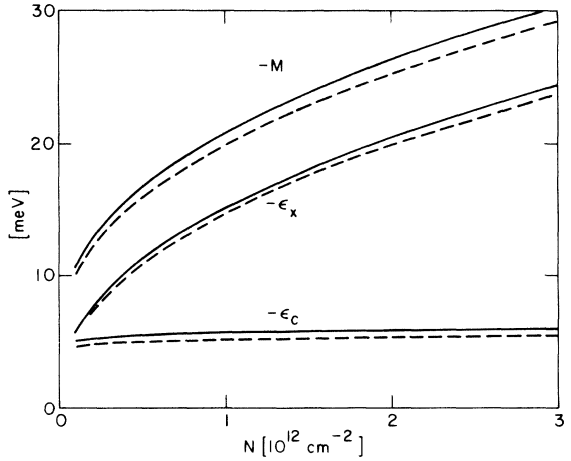

FIG. 4. Exchange energy ϵ_x, correlation energy ϵ_c, and self-energy M at the Fermi level as functions of inversion-layer density. Full curves: Results obtained with self-consistent Hartree wave functions. Dashed curves: Results obtained with the variational wave function [Eq. (31)].

interaction of an electron with its own image in the insulator taken into account in calculating the wave function. However, if we use Hartree wave functions determined from a Schrödinger equation which includes the image potential, the self-energy lies between the two and a little closer to the results for the variational wave function. Thus the exact shape of the wave function does not affect the results much, but comparison with Fig. 1 shows that the finite thickness of the layer strongly modifies the strictly two-dimensional results; the self-energy is roughly halved.

This is also the case for the effective mass. Results are shown in the discussion and comparison with experimental results, Sec. IX.

VI. FINITE INSULATOR THICKNESS

The possible effects of a small oxide thickness d can be treated in a very similar way. If the metal is considered infinitely conducting and we restrict ourselves to the two-dimensional case, Poisson's equation can be solved[18,19] and the unscreened Coulomb interaction between two electrons on the interface is $v(q)=e^2/2q\epsilon(q)$, where the effective permittivity $\epsilon(q)$ is

$$\epsilon(q)=\tfrac{1}{2}[\epsilon_s+\epsilon_i\coth(qd)]\,. \tag{33}$$

The main difference from the simplest case with $d=\infty$ is that the plasmon energy is proportional to q instead of $q^{1/2}$ for small q.

In Table II we show typical results on the self-energy of such a calculation. It is seen that the oxide must be extremely thin to show any perceptible effect. Since the finite thickness of the inversion layer will make the changes even smaller, they can be disregarded in normal experimental devices. The change in the effective mass is even smaller: At $d=100$ Å the mass is further enhanced by 5%, 1.5%, and 0.3% for the densities 10^{11}, 10^{12}, and 10^{13} cm^{-2}, respectively.

TABLE II. Effect of finite insulator thickness (d) on exchange (ϵ_x), correlation (ϵ_c), and self (M) energies as functions of inversion-layer density N calculated in the two-dimensional limit.

N (cm^{-2})	d (Å)	ϵ_x (meV)	ϵ_c (meV)	M (meV)
10^{11}	∞	− 6.6	−12.1	−18.7
	100	− 4.2	−11.5	−15.7
	50	− 3.1	−10.2	−13.3
10^{12}	∞	−20.8	−21.3	−42.1
	400	−20.1	−21.3	−41.4
	200	−19.3	−21.3	−40.6
	100	−17.9	−21.1	−39.0
	50	−15.5	−20.6	−36.1
10^{13}	∞	−65.9	−33.6	−99.5
	100	−62.8	−33.5	−96.3
	50	−59.9	−33.4	−93.3

VII. EFFECT OF OXIDE DISPERSION

It is known that amorphous SiO_2 shows dispersion and the lowest absorption band lies at a frequency ω_j of 440 cm^{-1} corresponding to an optical phonon energy of 55 MeV.[20] In the simple oscillator model the permittivity of SiO_2 is given by

$$\epsilon_i(\omega)=\epsilon_{i,\infty}+\frac{\epsilon_0 S}{1-(\omega/\omega_j)^2-i\gamma\omega/\omega_j}\,, \tag{34}$$

where S is the oscillator strength and γ the damping constant. In (34) we have only considered the lowest band; all higher bands are absorbed in the "high-frequency" dielectric function $\epsilon_{i,\infty}$, i.e., $\epsilon_{i,\infty}\equiv\epsilon_{i,\mathrm{st}}-\epsilon_0 S$ where $\epsilon_{i,\mathrm{st}}$ is the static permittivity. Experimentally, $S=0.84$ determined from the infrared absorption measurement by Miler.[20] The static dielectric constant is about 3.8.

Since the permittivity of the insulator enters in the bare Coulomb interaction because of the image potential, the dispersion can modify the previous results.

If we apply the rules [Eqs. (9–11)] for constructing the Lundqvist-Overhauser approximation to the permittivity $\epsilon_i(\omega)$, we obtain

$$\epsilon_i(\omega)=\epsilon_{i,\infty}-\epsilon_0 S\omega_j^2/(\omega^2-\omega_j^2+i\delta)\,, \tag{35}$$

i.e., the oscillator model without damping. In the two-dimensional limit the total permittivity function for interaction between electrons in the inversion layer is then

$$\epsilon(q,\omega)=\epsilon_\infty-\frac{\epsilon_0 S\,\omega_j^2/2}{\omega^2-\omega_j^2+i\delta}-\frac{\omega_p^2}{\omega^2+\omega_p^2-\omega_q^2+i\delta}\,, \tag{36}$$

where $\epsilon_\infty=\frac{1}{2}(\epsilon_{i,\infty}+\epsilon_s)$ and ω_p and ω_q were defined in Sec. II. For the inverse permittivity function the two dispersion relations mix and we obtain after elementary manipulations

$$\frac{\epsilon_\infty}{\epsilon(q,\omega)}-1=\frac{2\alpha_+\omega_+}{\omega^2-\omega_+^2+i\delta}+\frac{2\alpha_-\omega_-}{\omega^2-\omega_-^2+i\delta}\,, \tag{37}$$

where

$$\omega_\pm^2=\frac{\omega_{q0}^2+\omega_q^2}{2}\pm[\tfrac{1}{4}(\omega_{q0}^2-\omega_q^2)^2+\omega_s^2\omega_p^2]^{1/2}\,, \tag{38}$$

$$\alpha_\pm=\pm\frac{(\omega_\pm^2-\omega_q^2+\omega_p^2)(\omega_\pm^2-\omega_{q0}^2+\omega_s^2)}{2\omega_\pm(\omega_+^2-\omega_-^2)}\,, \tag{39}$$

$\omega_s^2=\frac{1}{2}(\epsilon_0/\epsilon_\infty)S\omega_j^2$, and $\omega_{q0}^2=\omega_j^2+\omega_s^2$. When Eq. (37) is inserted in the integrals for the self-energy, we see that the calculations are only slightly more complicated: to obtain the correlation energy we have to perform two integrals over q,

TABLE III. Effect of oxide dispersion on exchange (ϵ_x), correlation (ϵ_c), and self (M) energies in the two-dimensional limit ($z_{av}=0$) and when the finite layer thickness is taken into account through the variational wave function [Eq. (31)]. The values of $z_{av}=3/b$ correspond to a bulk doping of 7×10^{14} cm^{-3}. S is the oscillator strength of the lowest absorption band of SiO_2.

N (cm^{-2})	z_{av} (Å)	S	ϵ_x (meV)	ϵ_c (meV)	M (meV)
10^{11}	0	0	− 6.6	−12.1	−18.7
		0.84	− 7.0	−16.3	−10.1
	45	0	− 5.6	− 4.6	−10.1
		0.84	− 5.7	− 5.0	−10.7
10^{12}	0	0	−20.8	−21.3	−42.1
		0.84	−22.0	−25.4	−47.4
	30	0	−14.7	− 5.2	−19.9
		0.84	−15.3	− 5.5	−20.8

one for each branch. The same is true of the calculation of the effective mass.

It requires a few manipulations to convince oneself that these methods are applicable when we take the finite thickness of the inversion layers into account. Instead of the first two terms of Eq. (36) we have to introduce an effective-permittivity function of the surrounding medium defined by

$$\epsilon_m(q,\omega)=\epsilon_\infty(q)\left(1-\frac{\omega_1^2}{\omega^2-\omega_0^2}\right), \tag{40}$$

with

$$\epsilon_\infty(q)=\epsilon_s/f_\infty(q),$$

$$\omega_0^2=\omega_j^2\frac{\epsilon_s+\epsilon_{i,\mathrm{st}}}{\epsilon_s+\epsilon_{i,\infty}}\frac{f_0(q)}{f_\infty(q)}, \tag{41}$$

$$\omega_1^2=\omega_j^2\frac{\epsilon_s+\epsilon_{i,\mathrm{st}}}{\epsilon_s+\epsilon_{i\infty}}[1-f_0(q)/f_\infty(q)],$$

where $\epsilon_{i,\mathrm{st}}=\epsilon_{i\infty}+\epsilon_0 S$ is the static permittivity of the insulator, and f_0 and f_∞ are the functions in Eq. (29) calculated with the static and high-frequency value of ϵ_i, respectively.

In Table III we show typical results both in the two-dimensional limit and for finite layer thickness. The change is quite large in the former case for low densities but as one would expect the effect is much smaller in the latter since the image contribution to the bare Coulomb interaction decreases rapidly inside the semiconductor.

The effect of dispersion in the oxide on the effective mass is marginal. In the two-dimensional approximation the change is less than 0.6% for $10^{11}\leq N\leq 10^{13}$ cm^{-2}.

VIII. INFLUENCE OF ELECTRON-PHONON INTERACTION

In this section we describe a simplified theory for the contribution of the electron-phonon interaction to the self-energy. If we disregard the effect of the SiO_2-Si interface on the phonons, the deformation potential interaction is given by

$$H_{ep}=\int d^3\gamma\left(\Xi_d\nabla\cdot\vec{u}(\vec{r})+\Xi_u\frac{\partial u_z}{\partial z}\right)\psi^\dagger(\vec{r})\psi(\vec{r}), \tag{42}$$

where

$$\psi(\vec{r})=\frac{1}{\sqrt{A}}\sum_{\vec{k}}\zeta(z)e^{i\vec{k}\cdot\vec{R}}c_{\vec{k}}, \tag{43}$$

A is the area of the interface, $c_{\vec{k}}$ is the destruction operator for an electron with momentum $\vec{k}$, and in this section we use capital letters to denote the components of vectors parallel to the interface. For the displacement $\vec{u}(\vec{r})$ we have ($\alpha=x,y,z$)

$$u_\alpha(\vec{r})=\frac{1}{\sqrt{V}}\sum_{\vec{q},j}e^{i\vec{q}\cdot\vec{r}}u_{\vec{q}j}e_\alpha(\vec{q}j), \tag{44}$$

$$u_{\vec{q}j}=(2\rho\omega_{\vec{q}j})^{-1/2}(b_{\vec{q}j}+b^\dagger_{-\vec{q}j}), \tag{45}$$

where V is the volume of the sample, e_α denotes the direction of polarization, ρ is the mass density of the sample, $\omega_{\vec{q}j}$ is the phonon energy, $b_{\vec{q}j}$ and $b^\dagger_{\vec{q}j}$ are phonon destruction and creation operators, respectively, and j is the branch index.

Insertion of Eqs. (43)–(45) in (42) yields

$$H_{ep}=\sum_{\alpha,K,q,j}M_\alpha(\vec{q}j)[c^\dagger_{\vec{K}+\vec{Q}}c_{\vec{K}}(b_{\vec{q}j}+b^\dagger_{-\vec{q}j})]. \tag{46}$$

With $D_{xx}=D_{yy}=\Xi_d$ and $D_{zz}=\Xi_d+\Xi_u$ we have

$$M_\alpha(\vec{q}j)=\frac{iD_{\alpha\alpha}}{(2\rho V)^{1/2}}\frac{q_\alpha e_\alpha(\vec{q}j)}{\omega_{\vec{q}j}^{1/2}}\int dz\,|\zeta(z)|^2e^{iq_zz}. \tag{47}$$

If the variational wave function [Eq. (31)] is used, the last integral is $b^3/(b-iq_z)^3$.

With a phonon propagator $(\omega^2-\omega^2_{\vec{q}j})^{-1}$ we then obtain for the bare phonon contribution to the self-energy

$$M_{\mathrm{ph}}^{(\mathrm{bare})}(\vec{K},E)=\sum_j\int\frac{d^3q}{(2\pi)^3}\frac{|\gamma|^2}{2\omega_{\vec{q}j}}\left(\frac{\theta(|\vec{K}-\vec{Q}|-k_F)}{E-\xi(\vec{K}-\vec{Q})-\omega_{\vec{q}j}}+\frac{\theta(k_F-|\vec{K}-\vec{Q}|)}{E-\xi(\vec{K}-\vec{Q})+\omega_{\vec{q}j}}\right), \tag{48}$$

where

$$|\gamma(\vec{q},j)|^2=\left|\sum_\alpha\frac{i}{(2\rho)^{1/2}}D_{\alpha\alpha}q_\alpha e_\alpha(\vec{q}j)\frac{b^3}{(b-iq_z)^3}\right|^2. \tag{49}$$

To perform the integration over q_z we note that except for the factor $b^6/(b^2+q_z^2)^3$ the integrand does not vary extremely rapidly with q_z. We therefore crudely approximate it with that factor multiplied by the integrand evaluated at $q_z=0$. Then only longitudinal phonons contribute and we get

$$M_{\rm ph}^{(\rm bare)}(\vec{\rm K}, E) \simeq \int \frac{d^2Q}{(2\pi)^2}\frac{|g(Q)|^2}{2\omega_Q} \times\left(\frac{\theta(|\vec{\rm K}-\vec{\rm Q}|-k_F)}{E-\xi(\vec{\rm K}-\vec{\rm Q})-\omega_Q}+\frac{\theta(k_F-|\vec{\rm K}-\vec{\rm Q}|)}{E-\xi(\vec{\rm K}-\vec{\rm Q})+\omega_Q}\right), \tag{50}$$

with

$$|g(Q)|^2=\frac{3}{16}\frac{b}{\rho}\Xi_d^2Q^2, \tag{51}$$

and $\omega_Q=sQ$, where s is some average sound velocity for longitudinal phonons. The self-energy is now seen to represent the interaction between the two-dimensional electron gas and two-dimensional phonons with a propagator $(\omega^2-\omega_Q^2)^{-1}$ and an effective coupling $|g(Q)|^2$. Thus we can proceed in two dimensions, the third dimension being present only through b in the coupling (51).

The electron gas screens the electron-phonon interaction. We note that as long as we do not go beyond the RPA, we can simply add the bare phonon contribution to the bare Coulomb interaction and obtain the total screened Coulomb and phonon interaction in the RPA:

$$U(Q,\omega)=\frac{e^2/2\epsilon_\infty Q+|g(Q)|^2/(\omega^2-\omega_Q^2)}{1-[e^2/2\epsilon_\infty Q+|g(Q)|^2/(\omega^2-\omega_Q^2)]\tilde{\chi}(Q,\omega)}, \tag{52}$$

where $\tilde{\chi}$ was defined in Eq. (6). A few manipulations lead to

$$U(Q,\omega)=\frac{e^2}{2\epsilon_\infty Q}\Big/\left(1-\frac{2\epsilon_\infty Q}{e^2}g^2\frac{1}{\omega^2-\omega_Q^2+(2\epsilon_\infty Q/e^2)g^2}-\frac{e^2}{2\epsilon_\infty Q}\tilde{\chi}(Q,\omega)\right), \tag{53}$$

and if the Lundqvist-Overhauser approximation is applied for $\tilde{\chi}$ we see that we have the same mixed-mode problem as in the case of dispersion in the insulator, Sec. VII. We only have to substitute $|g(Q)|^2\times 2\epsilon_\infty Q/e^2$ for ω_s^2 and ω_Q^2 for ω_{q0}^2 in Eqs. (37–39).

As in Sec. IV we take the thickness of the inversion layer into account by using a q-dependent permittivity for the medium:

$$\epsilon_\infty(q)=\epsilon_s/f(q),$$

where $f(q)$ is given by Eq. (32).

TABLE IV. Effect of the electron-phonon interaction on exchange (ϵ_x), correlation (ϵ_c), and self (M) energies. $\Delta m^*/m$ shows the contribution to the mass enhancement from the electron-phonon interaction. Ξ is the deformation potential.

N (cm^{-2})	z_{av} (Å)	Ξ (eV)	ϵ_x (meV)	ϵ_c (meV)	M (meV)	$\left.\frac{\Delta m^*}{m}\right\|_{\rm ph}$
10^{11}	45	0	− 5.5	−4.6	−10.1	
		6	− 5.5	−4.9	−10.4	1.4×10^{-4}
5×10^{11}	35	0	−10.9	−5.0	−15.9	
		6	−10.9	−5.7	−16.6	1.9×10^{-4}
10^{12}	30	0	−14.7	−5.2	−19.9	
		6	−14.7	−6.4	−21.1	2.5×10^{-4}
2×10^{12}	25	0	−19.9	−5.3	−25.2	
		6	−19.9	−7.5	−27.4	3.4×10^{-4}
4×10^{12}	20	0	−27.0	−5.5	−32.5	
		6	−27.0	−9.2	−36.2	5.2×10^{-4}

In Table IV we list representative results. For the velocity of sound $s=8780$ m/sec was used, and for the mass density of Si $\rho=2330$ kg/m^3. It is seen that the electron-phonon interaction contributes little to the self-energy and almost nothing to the effective mass.

IX. DISCUSSION

Experiments which measure the self-energy invariably involve excitation from the lowest subband to a higher subband, and to make a comparison we have to know the self-energy of electrons in the excited subband. This calculation is the subject of a subsequent paper. Here we want to compare our results on the effective mass with experiments.

The cyclotron mass of electrons in the n inversion layer has been measured by several groups.[21-23] There seems to be agreement that for high frequencies and for concentrations $N_{\rm inv}>10^{12}$ cm^{-2} the cyclotron mass does not depend on $N_{\rm inv}$. This agrees with Kohn's theorem[24] that in a homogeneous system the cyclotron mass is not affected by the electron-electron interaction. There is not agreement on the actual value of the cyclotron mass, however, although all measurements have given results a little above the expected value $0.19\,m_e$. Kennedy *et al.*[23,25] have observed a frequency-dependent cyclotron mass with a variation from $0.193m_e$ at $\omega_c=52$ cm^{-1} to $0.207m_e$ at $\omega_c=25$ cm^{1}. At lower frequencies they have found a cyclotron mass which varies with $N_{\rm inv}$ in much the same way as measured by Smith and Stiles[26] from Shubnikov–de Haas oscillations.

They attributed this frequency dependence to the electron-electron interaction gradually made observable when the electron scattering gets relatively stronger so that the system can no longer be considered homogeneous. On the other hand the Munich group has seen no dependence of the cyclotron mass on $\omega\tau$,[27] where τ is the scattering time.

Magnetoconductivity oscillations in inversion layers were first observed by Fowler *et al.*[28] By measuring the temperature dependence of the amplitude of the oscillations and fitting to the corresponding theoretical expression they determined an effective mass of $0.2m_e$. Smith and Stiles[26] used the same method and found that the effective mass varies with density, as shown by their experimental points in Fig. 5. Most recently, in an effort to study the variation of the effective mass as a function of substrate bias, Lakhani *et al.*[29] have performed the same experiment. For zero substrate bias, they saw practically no enhancement of the mass with decreasing inversion-layer concentration. However, when a sufficiently high substrate bias was applied they found results a little above the values of Smith and Stiles. This was expected because substrate bias makes the inversion layer thinner, which increases the effective Coulomb interaction.

Also shown in Fig. 5 are the calculated values of the mass. The full curve shows our results in the two-dimensional limit. The curve is seen to follow the dependence on density measured by Smith and Stiles[26] very well. However, as the dashed curve shows, the agreement with the Smith and Stiles's measurement is only qualitative when the finite thickness of the inversion layer is taken into account. The doping of the samples used by Smith and Stiles was not recorded; the values for the parameter b of the variational wave function [Eq. (31)] used in the calculation correspond[3] to a doping of 10^{16} cm^{-3} ($b = 0.098$ Å^{-1} at $N_{\text{inv}} = 10^{11}$ cm^{-2}, and $b = 0.158$ Å^{-1} at $N_{\text{inv}} = 4\times10^{12}$ cm^{-2}). The values of b are not very sensitive to the doping.

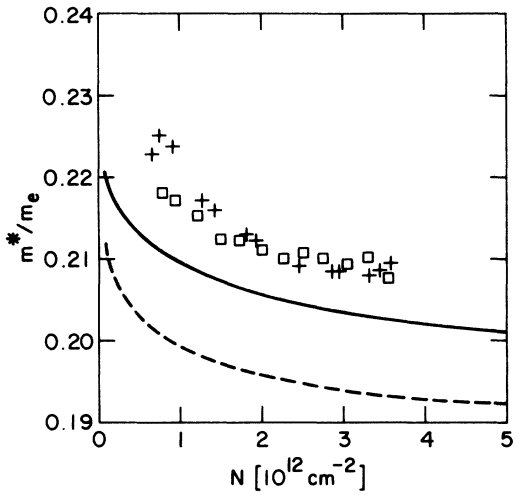

FIG. 5. Effective mass as a function of inversion-layer density. The full curve shows results in the two-dimensional limit. The dashed curve shows results for finite inversion-layer thickness. The variational wave function was used [Eq. (31)]. Experimentally determined values from Shubnikov–de Haas oscillations (Ref. 26): □ for the magnetic induction $B = 1.56T$, + for $B = 2.59T$.

By now a great many calculations of the effective mass have been done.[9,11,14,30-33] A rather complete collection of the results of different authors can be found in the paper by Uemura.[34] We wish to point out that the different approximations which employ dynamical screening give quite similar results. The fact that our results lie considerably lower than those of Ting, Lee, and Quinn[14,31] and those of Ohkawa[32] is not due to our use of the plasmon-pole approximation but to the fact that those authors use the bare particle energy in the Dyson equation (7) instead of the quasiparticle energy.[35]

Naturally it is hoped that sufficiently accurate measurements should make it possible to distinguish between the different approximations. However, as the experimental situation is not clear yet and the results are not too accurate, it is premature to draw conclusions about the accuracy of the different methods for calculating the many-body effects. Instead it is worth stressing the assumptions which have been made to obtain the theoretical results.

The surface has been considered perfectly plane and the barrier for penetration into the insulator was taken to be infinite. Furthermore, all effects of band structure have been treated in the effective-mass approximation. These three assumptions were discussed by Stern.[3] Recently the size of the mass enhancement due to nonparabolicity has been calculated by Falicov[36] to be only about 1% for $N_{\text{inv}} = 4\times10^{12}$ cm^{-2} increasing linearly with N_{inv}.

For the exchange and correlation we have used the Lundqvist-Overhauser approximation.[7,8] In three dimensions it is known that the difference between results of this approximation and the random-phase approximation is small.[37] As mentioned earlier this is also the case in the two-dimensional limit.

The critique which can be raised against the random-phase approximation at such rather low densities in three dimensions[38] applies to two dimensions as well. We do not know how crude this approximation is although arguments have

been given that it should work better in two than in three dimensions for comparable densities.[39] Results in the Hubbard approximation[31,32] do not differ much from the results on the effective mass in the random-phase approximation, and the contribution to the polarization from interband excitations is also very small.[32]

In this work we have studied a number of other interactions which contribute but they have all been shown to have a negligible effect on the mass. For the electron-phonon interaction this is in contrast with what has been calculated for metals in three dimensions,[40] for which the mass enhancement is even larger than the enhancement due to the electron-electron interaction.

ACKNOWLEDGMENTS

The author is indebted to L. J. Sham and F. Stern for many valuable discussions.

[1]Review articles have been published by G. Dorda, *Festkörperprobleme/Advances in Solid State Physics XIII*, (Pergamon, Oxford, 1973), p. 215; F. Stern, Crit. Rev. Solid State Sci. 4, 499 (1974); G. Landwehr, *Festkörperprobleme/Advances in Solid State Physics XV*, (Pergamon, Oxford, 1975), p. 49; J. F. Koch, *ibid.*, p. 79.

[2]F. Stern and W. E. Howard, Phys. Rev. 163, 816 (1967).

[3]F. Stern, Phys. Rev. B 5, 4891 (1972).

[4]A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs, N.J., 1963).

[5]F. Stern, Phys. Rev. Lett. 18, 546 (1967).

[6]L. Hedin, Phys. Rev. 139, A796 (1965).

[7]B. I. Lundqvist, Phys. Kondens. Mater. 6, 193 and 206 (1967).

[8]A. W. Overhauser, Phys. Rev. B 3, 1888 (1970).

[9]A. V. Chaplik, Zh. Eksp. Teor. Fiz. 60, 1845 (1971) [Sov. Phys.-JETP 33, 947 (1971)].

[10]F. Stern, Phys. Rev. Lett. 30, 278 (1973).

[11]J. F. Janak, Phys. Rev. 178, 1416 (1969).

[12]T. Ando and Y. Uemura, J. Phys. Soc. Jpn. 37, 1044 (1974).

[13]F. Stern, Jpn. J. Appl. Phys. Suppl. 2, 323 (1974).

[14]T. K. Lee, C. S. Ting, and J. J. Quinn, Solid State Commun. 16, 1309 (1975).

[15]E. D. Siggia and P. C. Kwok, Phys. Rev. B 2, 1024 (1970).

[16]F. F. Fang and W. E. Howard, Phys. Rev. Lett. 16, 797 (1966).

[17]F. Stern (private communication).

[18]A. V. Chaplik, Zh. Eksp. Teor. Fiz. 62, 746 (1972) [Sov. Phys.-JETP 35, 395 (1972)].

[19]M. Nakayama, J. Phys. Soc. Jpn. 36, 393 (1974).

[20]M. Miler, Czech. J. Phys. B 18, 354 (1968).

[21]S. J. Allen, Jr., D. C. Tsui, and J. V. Dalton, Phys. Rev. Lett. 32, 107 (1974).

[22]G. Abstreiter, P. Kneschaurek, J. P. Kotthaus, and J. F. Koch, Phys. Rev. Lett. 32, 104 (1974); J. P. Kotthaus, G. Abstreiter, J. F. Koch, and R. Ranvaud, *ibid.* 34, 151 (1975); H. Küblbeck and J. P. Kotthaus, *ibid.* 35, 1019 (1975).

[23]T. A. Kennedy, R. J. Wagner, B. D. McCombe, and D. C. Tsui (unpublished).

[24]W. Kohn, Phys. Rev. 123, 1242 (1961).

[25]T. A. Kennedy, R. J. Wagner, B. D. McCombe, and D. C. Tsui, Phys. Rev. Lett. 35, 1031 (1975); and Surf. Sci. (to be published).

[26]J. L. Smith and P. J. Stiles, Phys. Rev. Lett. 29, 102 (1972).

[27]J. F. Koch, Surf. Sci. (to be published).

[28]A. B. Fowler, F. F. Fang, W. E. Howard and P. J. Stiles, Phys. Rev. Lett. 16, 901 (1966).

[29]A. A. Lakhani, T. K. Lee, and J. J. Quinn, Surf. Sci. (to be published).

[30]K. Suzuki and Y. Kawamoto, J. Phys. Soc. Jpn. 35, 1456 (1973).

[31]C. S. Ting, T. K. Lee, and J. J. Quinn, Phys. Rev. Lett. 34, 870 (1975).

[32]F. J. Ohkawa, Surf. Sci. (to be published).

[33]T. Ando, Surf. Sci. (to be published).

[34]Y. Uemura, Surf. Sci. (to be published).

[35]B. Vinter, Phys. Rev. Lett. 35, 1044 (1975).

[36]L. M. Falicov (unpublished).

[37]B. I. Lundqvist, Phys. Kondens. Mater. 7, 117 (1968).

[38]K. S. Singwi, M. P. Tosi, R. H. Land, and A. Sjölander, Phys. Rev. 176, 589 (1968).

[39]R. K. P. Zia, J. Phys. C 6, 3121 (1973).

[40]L. Hedin and S. Lundqvist, Solid State Phys. 23, 108 (1969).