PHYSICAL REVIEW B VOLUME 15, NUMBER 8 15 APRIL 1977

Many-body effects in n-type Si inversion layers. II. Excitations to higher subbands

B. Vinter*
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 22 September 1976)

Effects of the electron-electron interaction on the excitation spectrum of the n-inversion layer on the Si (100) surface of the metal-oxide-semiconductor structure have been calculated. The correlation energy of an electron excited to the first subband of the doubly degenerate set or to the lowest subband of the fourfold degenerate set is of the order of -10 meV. The latter subband is found to be lower in energy for almost all inversion-layer concentrations and satisfactory agreement with magnetoconductivity measurements is obtained. A theory for infrared absorption which includes both resonant screening and exchange and correlation in the long-wavelength limit is developed and the results agree very well with experiments. The theory of a bound intersubband exciton is presented and binding energies of the order of 1 meV are calculated.

I. INTRODUCTION

In a previous article[1] (hereafter referred to as I) we investigated many-body effects on electrons in the lowest subband of an n-type inversion layer on p-type Si in a metal-insulator-semiconductor structure.

In this paper, we turn our attention to the effects of the electron-electron interaction on the excitation spectrum of the quasi-two-dimensional electron gas. The self-energy of electrons in the lowest subband was calculated in I, and it has been calculated by several other authors.[2-5] It is of the order of -20 meV. If there were no appreciable effect of the electron-electron interaction on the energy of an electron excited to the next higher subband, the excitation energy would be much larger than what is measured optically,[6] so it is important to consider many-body effects in the higher subbands.

The basis of the calculations is the self-consistent solution of the Schrödinger equation in the Hartree approximation.[7,8] From those solutions we derive the one-particle Green's function for the noninteracting electron gas to which we can apply the methods of many-body theory.[9] Here we have used the Lundqvist[10]-Overhauser[11] or plasmon-pole approximation, which gives results very close to those of the random-phase approximation.[12,13]

A similar method has been used by Ohkawa[2] both on n-type inversion and accumulation layers and on p-type layers, whereas Ando[3] applied a modified form of the local exchange-correlation potential method derived by Kohn and Sham.[14]

In Sec. II we derive the theory we have used and state the approximations. Sections III and IV show our results on the self-energy of electrons excited to higher-lying subbands. In Sec. V we compare the results with Shubnikov–de Haas measurements[15,16] of the inversion-layer density at which a higher-lying band begins to be filled. It is shown that our results give much better agreement with the experiments than the Hartree calculations.[3-8] In Sec. VI we present the theory we have used for infrared absorption. Our theory includes both the resonant screening,[17,18] and the exchange and correlation effects, and the comparison with experiments in Sec. VII shows very good agreement. Finally, we show in Sec. VIII that a two-particle bound-exciton-like state should be possible and estimate its binding energy. In the conclusion we point out some experimental facts that are still not understood.

II. THEORY

In the Hartree approximation the one-electron envelope wave functions have the form[7]

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

where z is the direction perpendicular to the interface and $\vec{k}$ and $\vec{R}$ are two-dimensional vectors parallel to the surface. $\zeta_i(z)$ is a solution of a one-dimensional Schrödinger equation and Poisson's equation. For the set of subbands originating from the Si valleys with high mass (m_3) perpendicular to the interface the corresponding energies are ($\hbar = 1$):

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

where m is the mass for motion parallel to the surface. We only consider a Si (100) surface, so m is the transverse Si mass.[7]

At absolute temperature $T = 0$ for sufficiently low densities only the lowest subband $i = 0$ is occupied up to the Fermi level determined by a Fermi wave-vector $k_F = (2\pi N/n_v)^{1/2}$, where N is the inversion layer density and n_v the valley degeneracy. We re-

strict our theory to this situation.

The Green's function[9] for the noninteracting electron gas is a sum of Green's functions in each subband. After Fourier analysis in the directions parallel to the surface we have

$$G(\vec{k}, E, z, z') = \sum_{i=0}^{\infty} G_i(\vec{k}, E, z, z'), \tag{3}$$

$$G_0(\vec{k}, E, z, z') = \zeta_0(z)\zeta_0^*(z') \times \left(\frac{\Theta(|\vec{k}| - k_F)}{E - \xi(k) + i\delta} + \frac{\Theta(k_F - |\vec{k}|)}{E - \xi(k) - i\delta} \right),$$

$$G_i(\vec{k}, E, z, z') = \zeta_i(z)\zeta_i^*(z')[E - \xi(k) - E_{i0} + i\delta]^{-1}, \quad i > 0, \tag{4}$$

where $\xi(k) = (k^2 - k_F^2)/2m$, $E_{i0} = E_i - E_0$, and $\delta = 0+$.

Since the wave functions $\zeta_i(z)$ form a complete set, we can use them to expand any function of z and z':

$$F(z, z') = \sum_{i,j} \zeta_i(z) F_{ij} \zeta_j^*(z'),$$
$$F_{ij} = \int_0^{\infty} \int_0^{\infty} dz\, dz'\, \zeta_i^*(z) F(z, z') \zeta_j(z'), \tag{5}$$

so that the noninteracting Green's function may be written

$$G(\vec{k}, E, z, z') = \sum_{i,j} \zeta_i(z) G_{ij}(\vec{k}, E) \zeta_j^*(z'), \tag{6}$$

$$G_{00}(\vec{k}, E) = \frac{\Theta(|\vec{k}| - k_F)}{E - \xi(k) + i\delta} + \frac{\Theta(k_F - |\vec{k}|)}{E - \xi(k) - i\delta},$$
$$G_{ij}(\vec{k}, E) = \frac{\delta_{ij}}{E - \xi(k) - E_{i0} + i\delta}, \quad i > 0. \tag{7}$$

In this basis the bare Coulomb interaction $v_{ijlm}(q)$ is given by

$$v_{ijlm}(q) = (e^2/2\epsilon_s q) f_{ijlm}(q),$$
$$f_{ijlm}(q) = \int_0^{\infty} \int_0^{\infty} dz\, dz'\, \zeta_i^*(z)\zeta_j(z)\zeta_l^*(z')\zeta_m(z') \times \left(e^{-q|z-z'|} + \frac{\epsilon_s - \epsilon_i}{\epsilon_s + \epsilon_i} e^{-q(z+z')} \right), \tag{8}$$

and is represented by the diagram (a) of Fig. 1. ϵ_s and ϵ_i are the permittivities of Si and SiO_2, respectively.

In the random-phase approximation an effective screened Coulomb interaction is introduced. It is determined by solving the set of equations shown diagrammatically in (c) of Fig. 1:

$$U_{ijlm}(q, \omega) = v_{ijlm}(q) + \sum_{n_1, n_2} v_{ijn_2n_1}(q) \chi_{n_2n_1}(q, \omega) U_{n_1n_2lm}(q, \omega), \tag{9}$$

$$\chi_{n_2n_1}(q, \omega) = -2in_v \int \frac{d^2k\, dE}{(2\pi)^3} G_{n_1n_1}(\vec{k} + \vec{q}, E + \omega) G_{n_2n_2}(\vec{k}, E). \tag{10}$$

The polarization χ is only different from zero when at least one index is zero.

Within the same approximation the dressed Green's function can be determined from the Dyson equation represented by diagram (d) in Fig. 1:

$$\mathcal{G}_{ij}(k, E) = G_{ii}(k, E)\delta_{ij} + G_{ii}(k, E) \sum_n M_{in}(k, E) \mathcal{G}_{nj}(k, E), \tag{11}$$

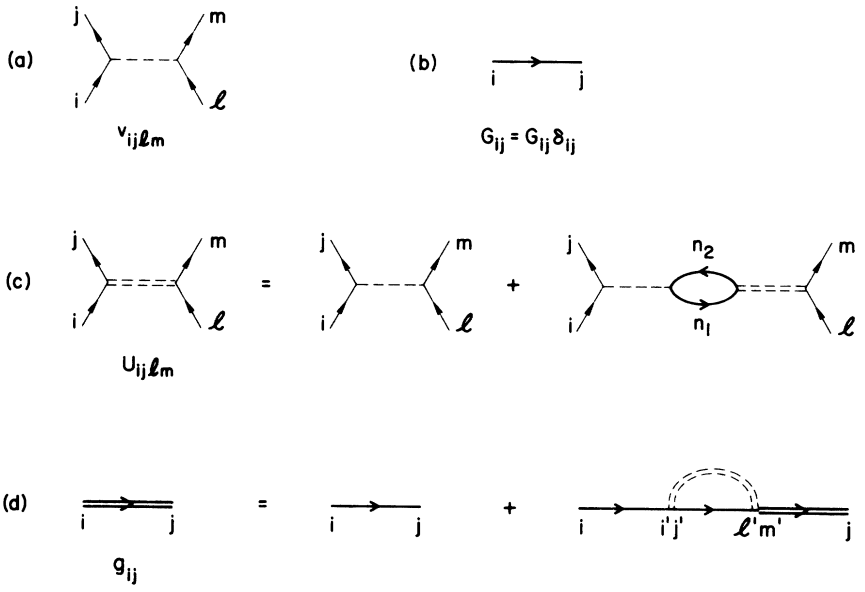

FIG. 1. Diagrammatic representation of the approximation used for calculating the self-energy. The indices denote subbands, $v(q)$ is the bare Coulomb interaction, $U(q, \omega)$ is the screened interaction, G is the Green's function for the noninteracting system, and $\mathcal{G}$ the dressed Green's function.

with a self-energy matrix

$$M_{in}(k,E)=\sum_l i\frac{d^2q\,d\omega}{(2\pi)^3}\times U_{iiln}(q,\omega)G_{ll}(\vec{k}-\vec{q},E-\omega). \quad (12)$$

Finally, if we restrict our attention to quasiparticle energies, we have to find the poles of $\mathcal{G}$, i.e., the solutions of the equation

$$\det[G_{ii}(k,E)\delta_{ij}+M_{ij}(k,E)]=0. \quad (13)$$

A considerable simplification of these equations can be obtained if we assume that the bare Coulomb interaction is negligible, when $i\neq j$ or $l\neq m$ in diagram (a) of Fig. 1. This is probably not a bad approximation: The orthogonality of the wave functions implies that $v_{ijlm}(q)$ in these cases is of the order of a constant for small q, while in the other cases $v_{ijlm}(q)$ goes like $1/q$. Furthermore, if screening is neglected, the self-energy matrix element M_{11} of an electron in the higher subband involves only the interaction v_{1001}, and that exchange contribution was calculated by Stern[19] to be less than 1 meV.

If we exclude contributions from such diagrams we find for the screened interaction:

$$\begin{aligned}U_{0000}&=v_{0000}/(1-v_{0000}\chi_{00}),\\ U_{ii00}&=U_{00ii}=v_{ii00}/(1-v_{0000}\chi_{00}),\\ U_{iiii}&=v_{iiii}+v_{ii00}\chi_{00}v_{00ii}/(1-v_{0000}\chi_{00}),\end{aligned} \quad (14)$$

while all other interactions are zero. In this approximation the self-energy and the dressed Green's function are diagonal, so only the lowest and one excited subband need be considered at one time.

For the screening we have used the Lundqvist[10]-Overhauser[11] approximation as described in I. Thus we separate U_{0000}:

$$U_{0000}=v_{0000}+v_{0000}[(1-v_{0000}\chi_{00})^{-1}-1], \quad (15)$$

and approximate the second term by an effective plasmon:

$$\mathrm{Im}[v_{0000}\chi_{00}/(1-v_{0000}\chi_{00})]=-\tfrac{1}{2}\pi(\omega_p^2/\omega_q)\delta(\omega-\omega_q), \quad (16)$$

where

$$\omega_p^2(q)=v_{0000}(q)Nq^2/m \quad (17)$$

and

$$\omega_q^2=-\omega_p^2\frac{1-v_{0000}(q)\chi_{00}(q,0)}{v_{0000}(q)\chi_{00}(q,0)} \quad (18)$$

are determined so that two sum rules are fulfilled. $\chi_{00}(q,0)$ was calculated by Stern[20]:

$$\chi_{00}(q,0)=-(2n_v m/2\pi)\times\{1-\Theta(q-2k_F)[1-(2k_F/q)^2]^{1/2}\}. \quad (19)$$

It is obvious that the same approximation can be used on the other matrix elements of the effective interaction.

For the self-energy we now obtain after integration over ω

$$M_{00}(k,E)=M_{00}^x(k)+M_{00}^c(k,E), \quad (20)$$

$$M_{00}^x(k)=\int\frac{d^2q}{(2\pi)^2}v_{0000}(q)\Theta(k_F-|\vec{k}-\vec{q}|), \quad (21)$$

$$M_{00}^c(k,E)=\int\frac{d^2q}{(2\pi)^2}v_{0000}(q)\frac{\omega_p^2}{2\omega_q}\left(\frac{\Theta(k_F-|\vec{k}-\vec{q}|)}{E-\xi(\vec{k}-\vec{q})+\omega_q}+\frac{\Theta(|\vec{k}-\vec{q}|-k_F)}{E-\xi(\vec{k}-\vec{q})-\omega_q}\right), \quad (22)$$

$$M_{11}(k,E)=\int\frac{d^2q}{(2\pi)^2}\frac{v_{1100}^2(q)}{v_{0000}(q)}\frac{\omega_p^2}{2\omega_q}\times[E-\xi(\vec{k}-\vec{q})-E_{10}-\omega_q]^{-1}. \quad (23)$$

Note that, since the excited band is unoccupied in the noninteracting case (G_{11} has no pole in the upper half of the E plane), there is no contribution from v_{1111}. For the same reason any static approximation to U_{1111} would give $M_{11}=0$, as long as we do not consider scattering processes which change the subband of one of the electrons.

The quasiparticle energies can finally be determined from the Dyson equations

$$E+\mu=\xi(k)+M_{00}(k,E), \quad (24)$$

$$E'+\mu=E_{10}+\xi(k)+M_{11}(k,E'), \quad (25)$$

where the chemical potential μ is found by inserting $k=k_F$ and $E=0$ in the first equation.[12]

III. RESULTS

The wave functions in the direction perpendicular to the interface were calculated numerically in the Hartree approximation by Stern.[8] A variational treatment of the Schrödinger and Poisson equations for the wave function in the lowest subband shows[21] that the image potential $(e^2/16\pi\epsilon_s z)(\epsilon_s-\epsilon_i)/(\epsilon_s+\epsilon_i)$, which represents the interaction of an electron with its image in the insulator, tends to increase the spatial extent of the wave function. On the other hand, the exchange interaction tends to compress the wave function. For low inversion-layer concentrations the exchange interaction is negligible. For higher concentrations the increase in the extent of the wave function due to the image potential is very nearly cancelled by the compression arising from the exchange interaction, so that the

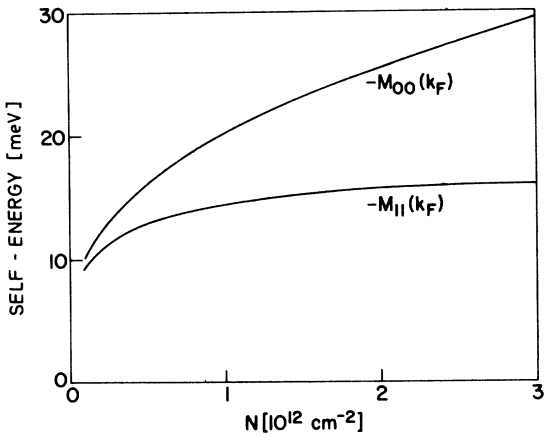

FIG. 2. Self-energy at the Fermi wave vector in the lowest (M_{00}) and the first excited (M_{11}) subbands for the parameters of Table I.

wave function when both interactions are used is close to the wave function calculated without image and without exchange.

In order to account for the influence of the exchange interaction in the numerical calculations, i.e., to improve the calculations in the direction of a Hartree-Fock approximation, the following correction to the Hartree energy separation between the lowest and first excited subband was made: The Schrödinger equation was solved (i) for a Hamiltonian which included the image potential and (ii) for a Hamiltonian without the image potential but with the image energy calculated as a perturbation. By comparison with the corresponding variational results the energy separation E_{10} was determined by linear interpolation between the results of (i) and (ii). The parameter for the interpolation was the average value of z, i.e.,

$$(E_{10}-E_{10}^{(i)})/(E_{10}^{(ii)}-E_{10}^{(i)})=(z_{av}-z_{av}^{(i)})/(z_{av}^{(ii)}-z_{av}^{(i)}), \tag{26}$$

where z_{av}, $z_{av}^{(i)}$, and $z_{av}^{(ii)}$ were taken from the variational results. Implicit in this correction is a disputable assumption that the many-body corrections affect the wave function and energy in the excited subband in the same way as in the lowest subband. Errors resulting from this method of estimating the one-electron energies are thought to be of the order of a meV or less. In Fig. 2, we show the self-energies at $k=k_F$ found by solving the Dyson equations (24) and (25). The parameters of the calculation are shown in Table I, and the wave functions used to obtain $f_{iiii}(q)$ [Eq. (8)] were those calculated with the image potential in the Hamiltonian. If wave functions calculated without the image potential are used, the self-energies change by less than 1 meV and the difference between the self-energies changes by at most 0.25 meV. Note that only the difference $E'-E_{10}$ enters in Eq. (23), so that the self-energy $M_{11}(k,E')$ is independent of E_{10}.

It is seen that the correlation energy in the excited subband is even larger than the correlation energy in the lowest subband which is about -5 meV (see Fig. 4 of I). Mathematically, this is due to the fact that the contribution to M_{00}^{c} from occupied states to some extent cancels the contribution from nonoccupied states, while in the excited subband no states are occupied in the noninteracting electron gas. The main contribution to the self-energy in the lowest subband comes from the exchange energy.

In Fig. 3, we show graphically the solution of the second Dyson equation (25) in the special case $N=10^{12}$ cm^{-2}. It is clear from the figure that the approximation to the Dyson equation

$$E'+\mu\cong\xi(k)+E_{10}+M_{11}(k,E_{10}+\xi(k)), \tag{27}$$

in which the noninteracting energy $E_{10}+\xi(k)$ instead of the quasiparticle energy E' is taken as the second argument of the self-energy would give quite different results.

It can be seen that the Dyson equation has two more solutions at higher energies. This kind of behavior has also been found in three dimensions in both the plasmon-pole and random-phase approximations by Lundqvist,[10,13] and by analogy with his results we expect those solutions to have little spectral weight.

If we turn to other k values, we find the self-energies shown in Fig. 4. It can be seen that they

TABLE I. Parameters used in the calculations.

Valley degeneracy	n_v	2
Mass perpendicular to surface	m_3	$0.916\ m_e$
Mass parallel to surface	m	$0.1905\ m_e$
Permittivity of Si	ϵ_s	$11.7\ \epsilon_0$
Permittivity of SiO_2	ϵ_i	$3.9\ \epsilon_0$
Doping concentration	N_A-N_D	7×10^{14} cm^{-3}
Depletion layer concentration	N_{depl}	1.01×10^{11} cm^{-2}

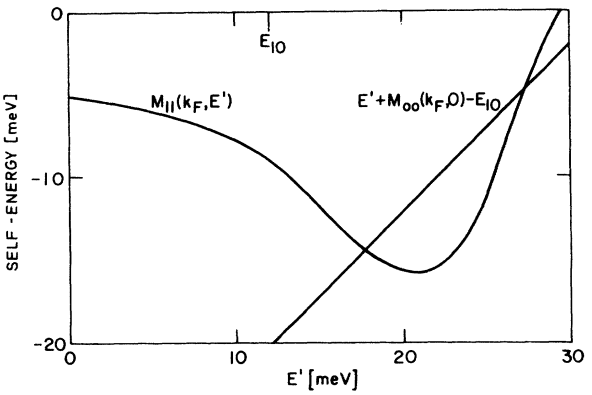

FIG. 3. Graphical solution of the Dyson equation for the first excited subband. E_{10} is the energy separation in the Hartree approximation. $N=10^{12}$ cm^{-2}. $N_A-N_D=7\times10^{14}$ cm^{-3}.

vary little with k in the excited band as found for the lowest subband in I, so that the quasiparticle bands are still nearly parallel. The mean mass in the excited band is a little larger than in the lowest band. For the case shown, the mean mass in the lowest subband is 8% higher than the bulk mass and it is 14% higher in the excited subband. The energy separation between the two subbands at the bottom of the bands is found to be only 0.3 meV larger than at $k=k_F$.

IV. OTHER SUBBANDS

From the general one-electron theory it is known[7] that the six equivalent valleys of Si give rise to two subband ladders. The lowest subband has its origin in the two valleys which have the longitudinal mass perpendicular to the surface. The lowest subband of the ladder coming from the four other valleys lies very close to the first excited subband in the self-consistent calculations.[8]

In the two-dimensional Brillouin zone, the four other valleys are situated far from the bottom of the doubly degenerate subbands. Since optical transitions are vertical because of momentum conservation it is not possible to excite an electron from the lowest subband to one of the fourfold degenerate subbands directly. Phonon-assisted indirect transitions are hardly possible because the relevant phonon energies are too large.[22] The only way to populate the fourfold degenerate subbands easily at low temperatures is by increasing the number of electrons until the Fermi energy lies above the bottom of those subbands.

In order to calculate the energy of the lowest subband of the fourfold degenerate set we proceed in the same way as for the excited subbands of the doubly degenerate set, with two modifications. First the envelope wave functions in the direction perpendicular to the interface in the fourfold and doubly degenerate sets are not orthogonal. However, we can still neglect diagrams which involve intervalley scattering, since the momentum transfer in such a process is of the order of k_0, the wave vector from the origin of the Brillouin zone to the bottom of one of the fourfold degenerate valleys. Thus the Coulomb interaction in such a process ($\sim 1/k_0$) is much smaller than in intravalley processes ($\sim 1/q$). Second, the subbands corresponding to the fourfold degenerate valleys are not isotropic.[7] We circumvent this problem by pretending they are isotropic, with a mass equal to the density of states mass (equal to $0.417m_e$). If we perform the calculations with the conductivity mass (equal to $0.315m_e$) instead, we find negligible changes in the energy of the bottom of the subband, which indicates that the isotropic approximation is not likely to introduce serious errors. The results of this calculation and a comparison with experiments are presented in Sec. V.

If we calculate the self-energy for an electron in the second excited subband of the doubly degenerate set, we get results which are a few meV smaller than for the first excited subband, when the density is below about 7×10^{11} cm^{-2}. This is because the interaction $v_{2200}(q)$ is smaller than $v_{1100}(q)$ since the envelope wave function $\zeta_2(z)$ is more spread out than the wave function $\zeta_1(z)$. Thus the energy difference between the first and second excited subbands is a little larger than what is found in the Hartree approximation. The calculated threshold for transition to the second excited subband varies from 16 meV at $N=10^{11}$ cm^{-2} to 22 meV at $N=5\times10^{11}$ cm^{-2} for the parameters shown earlier in Table I. For a higher doping $N_A-N_D=1.65\times10^{15}$ cm^{-3} the results are 20 meV at $N=10^{11}$ cm^{-2} and 26

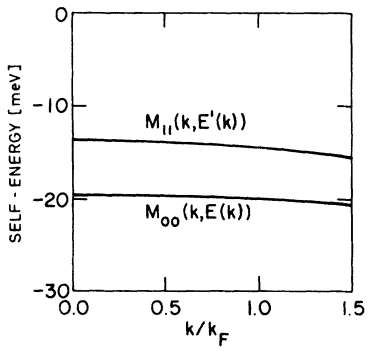

FIG. 4. Self-energies of quasiparticles in the lowest (M_{00}) and in the first excited (M_{11}) subbands as a function of the wave vector. $N=10^{12}$ cm^{-2}.

meV at $N = 5 \times 10^{11}$ cm^{-2}.

When we do the calculation for higher densities, the high doping case behaves normally and the energy separation E_{20} increases smoothly to 48 meV at $N = 3 \times 10^{12}$ cm^{-2}. For the lower doping concentration, however, a discontinuity occurs. What happens mathematically can be understood from the graphical solution of the Dyson equation shown in Fig. 3. At higher concentrations and lower doping the spatial extent of the wave function $\zeta_2(z)$ increases and the Coulomb interaction decreases. The self-energy as a function of E then decreases in magnitude. At some point the Dyson equation has no low-energy solution and we are left with the solution at high energy. When broadening is considered this would mean physically that the subband is gradually broadening with increasing density and disappearing in the background. It is difficult to decide whether the results bear any resemblance to reality in this case or whether we have pushed calculations farther than the approximations permit. However, it is clear that we would expect the second excited subband if observable to lie at least 10 meV above the first excited subband for $N \geq 10^{12}$ cm^{-2}.

V. COMPARISON WITH SHUBNIKOV-de HAAS MEASUREMENTS

Howard and Fang[15] and later, Tsui and Kaminsky,[16] have measured the density at which a higher subband starts to be populated. They measured the magnetoconductivity oscillations which occur when a magnetic field is applied perpendicular to the inversion layer. Each period corresponds to the filling of one Landau level in the lowest subband. At a certain inversion-layer density the period changes, which shows that not all added electrons go into the lowest subband. Some of them populate a higher subband.

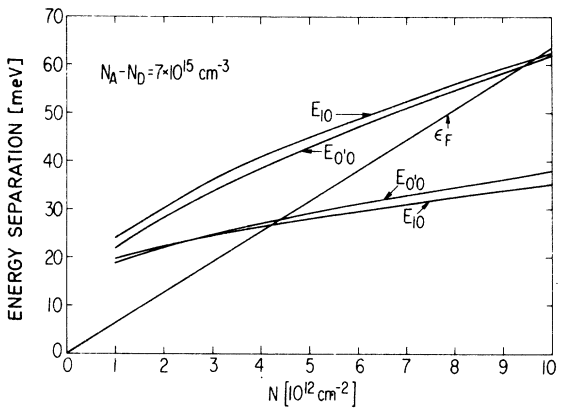

FIG. 5. Energy separation from the lowest subband to the first excited subband of the doubly degenerate set (E_{10}) and to the lowest subband of the fourfold degenerate set ($E_{0'0}$). The lower curves show the results in the Hartree approximation. In the upper curves exchange and correlation are included. ϵ_F is the Fermi energy. $N_A - N_D = 7 \times 10^{15}$ cm^{-3}.

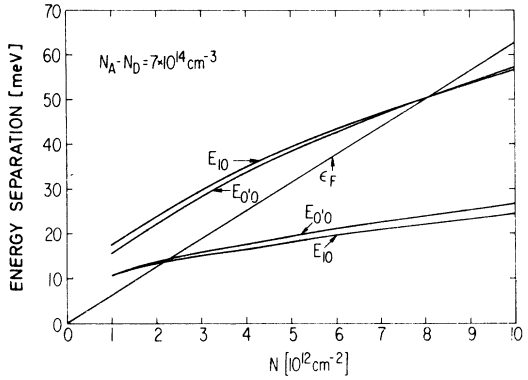

FIG. 6. Results as in Fig. 5 for $N_A - N_D = 7 \times 10^{14}$ cm^{-3}.

In Fig. 5 we show results of our calculation for a substrate doping of 7×10^{15} cm^{-3}. The lower curves show the energy separation between the lowest subband and the first excited subband (E_{10}) and the lowest subband of the fourfold degenerate set ($E_{0'0}$) in the Hartree approximation with the corrections for image potential and exchange mentioned in Sec. III. The higher-lying curves show the same separations when the many-body effects are added. It should be mentioned that all the calculations are made with the assumption that only the lowest subband is occupied, so the results are only valid up to the point when a higher subband is populated.

Two results are noteworthy. First, the density at which the Fermi level reaches the bottom of the first higher-lying subband is doubled when the many-body effects are taken into account. Second, in the Hartree approximation the first subband to be populated belongs to the doubly degenerate set whereas the many-body effects make $E_{0'0}$ smaller than E_{10} at all densities below the threshold density for population of higher subbands.

For the substrate doping of 7×10^{15} cm^{-3} the threshold density was found[15] to be $N = 7.6 \times 10^{12}$ cm^{-2} corresponding to an energy separation of 48 meV. For this density the calculation gives $E_{0'0} = 53$ meV which is significantly better than the Hartree result of $E_{10}^H = 32$ meV.

Figure 6 shows similar results for a lower doping, $N_A - N_D = 7 \times 10^{14}$ cm^{-3}. By accident the energies of subbands 1 and 0′ coincide at the concentration for which those levels start to be populated. Howard and Fang[15] did not obtain results for this doping but estimated by interpolation that the

threshold for population of higher-lying levels would be about $N = 4.7 \times 10^{12}$ cm^{-2}. The agreement with our results is a little worse than in the higher doping case. On the other hand Tsui and Kaminsky[16] measured the threshold to be at $N = 7.4 \times 10^{12}$ cm^{-2} in excellent agreement with our results. It is not possible to decide on the accuracy of our theory when two supposedly identical experiments give such different results. However, it is clear that inclusion of exchange and correlation considerably improved the agreement with the experiments compared to the Hartree results.

Tsui and Kaminsky[16] also found that the threshold is lowered when stress is applied, which indicates that $E_{0'0}$ is lower than E_{10} at the threshold density. This is at least consistent with our results and is qualitatively different from the Hartree results.

Other effects may be important for the separation energy $E_{0'0}$. Stern[21] has found that if penetration of the wave functions into the insulator is taken into account, the value of $E_{0'0}$ in the one-electron approximation may be reduced by several meV, whereas E_{10} does not change much. In the model of Kelly and Falicov[23] the interaction between electrons in the fourfold degenerate valleys could perceivably lift the degeneracy and change their position.

VI. THEORY OF INFRARED ABSORPTION

It was pointed out by Chen *et al.*[17] that in an infrared-absorption experiment one would not see maximum absorption at a radiation energy equal to the subband separation but at a somewhat higher energy. This is because the electrons in the inversion layer are polarized by the radiation field and thus create an additional field. A self-consistent solution in the one-electron approximation gives the result that at a radiation energy equal to the intersubband separation the electrons completely screen the external field so that the absorption vanishes. Chen *et al.* used a thin-slab model of the inversion layer and did not obtain quantitative results. A calculation which takes the inversion-layer wave functions into account was presented by Allen *et al.*[18] who found that the effect of resonant screening substantially changes the position of maximum absorption: The changes are of the same order of magnitude as the self-energy shifts calculated earlier. Results on the absorption spectrum furthermore showed that the position of the first maximum depends on the energy and broadening of higher-lying subbands. These results were all obtained in the long-wavelength limit, but in a real experiment retardation effects could be important in determining what modes actually propagate and absorb energy in the inversion layer.[17]

A complete theory of the absorption as a function of both wavelength and energy including exchange and correlation effects is desirable and important in order to make a good comparison to experiments. However, it would be a project beyond the scope of this paper. Here we have more modest goals and show two results. First, we want to demonstrate how the results of Chen *et al.*[17] and Allen *et al.*[18] come out of the Kubo formalism for the conductivity,[24] and second we show results of a calculation which to some extent takes into account exchange and correlation. In both cases we take the long wavelength limit, neglect subbands higher than the first excited subband, and neglect broadening of the two lowest subbands. Thus we only search for the lowest-energy pole of the conductivity for $q = 0$ in a two-subband model, or equivalently we search for the pole of the polarization operator $\Pi(0, \omega)$.

The simplest diagram for the polarization is shown in (a) of Fig. 7. It is straightforward to show that for $q = 0$ and $\omega \neq 0$, we have

$$\begin{aligned}\Pi^{(0)}(0,\omega) &= \chi(0,\omega) \\ &= \chi_{01}(0,\omega) + \chi_{10}(0,\omega) \\ &= 2NE_{10}/(\omega^2 - E_{10}^2),\end{aligned} \tag{28}$$

which has a pole at $\omega = E_{10}$, the subband separation for the noninteracting system χ_{ij} was defined in Eq. (10).

The self-consistent results of Allen *et al.*[18] are diagrammatically shown in (b) of Fig. 7. The summation is easily performed and gives

$$\Pi^{(1)}(0,\omega) = \chi(0,\omega)/[1 - v\chi(0,\omega)], \tag{29}$$

where

$$\begin{aligned}v &= v_{0101}(q=0) \\ &= -\frac{e^2}{2\epsilon_s}\int_0^\infty \int_0^\infty dz\, dz'\, \zeta_0(z)\zeta_1(z)\,|z - z'|\,\zeta_0(z')\zeta_1(z').\end{aligned} \tag{30}$$

A few manipulations show that $v = e^2 S_{11}/\epsilon_s$, where

$$S_{11} = \int_0^\infty \left(\int_0^z \zeta_0(z')\zeta_1(z')\,dz' \right)^2 dz. \tag{31}$$

The pole of the polarization in this approximation is now at a higher energy:

$$\begin{aligned}\omega_{10}^* &= (E_{10}^2 + \omega_{11}^2)^{1/2}, \\ \omega_{11}^2 &= 2e^2 N S_{11} E_{10}/\epsilon_s,\end{aligned} \tag{32}$$

the result given by Allen *et al.*[18]

Earlier works on many-body effects in this system implicitly assumed that the infrared absorption would show a maximum at a frequency corresponding to the intersubband energy separation.

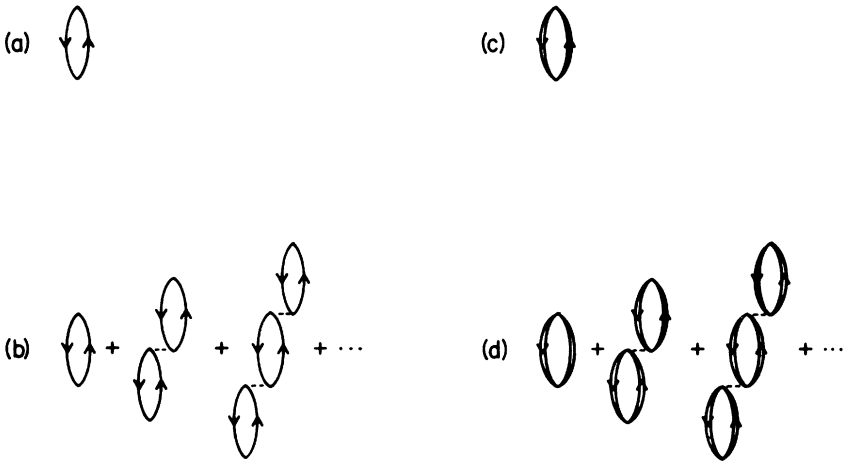

FIG. 7. Approximations used for determining the peak in conductivity: (a) for the noninteracting case; (b) including resonant screening; (c) including exchange and correlation; and (d) including both resonant screening and exchange and correlation.

This assumption is represented by the diagram (c) in Fig. 7. In that approximation the polarization is given by

$$\Pi^{(2)}(0,\omega)=\chi^{d}_{10}(0,\omega)+\chi^{d}_{01}(0,\omega), \tag{33}$$

where

$$\begin{aligned}\chi^{d}_{10}(0,\omega)&=-2n_v i\int\frac{d^2k\,dE}{(2\pi)^3}\mathcal{G}_{11}(k,E+\omega)\mathcal{G}_{00}(k,E)\\&=-2n_v i\int\frac{d^2k\,dE}{(2\pi)^3}\\&\quad\times[E+\omega-E_{10}-\xi(k)-M_{11}(k,E+\omega)]^{-1}\\&\quad\times[E-\xi(k)-M_{00}(k,E)]^{-1},\end{aligned} \tag{34}$$

and

$$\begin{aligned}\chi^{d}_{01}(0,\omega)&=-2n_v i\int\frac{d^2k\,dE}{(2\pi)^3}\mathcal{G}_{00}(k,E+\omega)\mathcal{G}_{11}(k,E)\\&=\chi^{d}_{10}(0,-\omega).\end{aligned} \tag{35}$$

Since only $\mathcal{G}_{00}(k,E)$ has a pole in the upper half plane and only for $k<k_F$ we obtain upon integration over E

$$\chi^{d}_{10}(0,\omega)=2n_v\int\frac{d^2k}{(2\pi)^2}\frac{\Theta(k_F-k)}{\omega-E_{10}-M_{11}(k,E_k+\omega)+M_{00}(k,E_k)}, \tag{36}$$

where E_k is the energy of a quasiparticle of momentum k in the lowest subband.

If we neglect broadening of the two subbands, it follows from Eq. (36) that $\Pi^{(2)}(0,\omega)$ has a finite imaginary part only if the frequency ω corresponds to an energy difference between a quasiparticle in the excited subband and a quasiparticle of the same momentum in the lowest subband below the Fermi level.

In order to calculate the combined effect of resonant screening and exchange and correlation, we perform the same summation as in diagram (b) but insert the dressed Green's functions and thus obtain the total diagram (d), for which we find the polarization

$$\Pi(0,\omega)=\Pi^{(2)}(0,\omega)/[1-v\Pi^{(2)}(0,\omega)]. \tag{37}$$

In the results to be presented in Sec. VII, we determine the frequency for maximum infrared absorption as the frequency for which the denominator of Π is zero. For $\Pi^{(2)}(0,\omega)$ we use the expressions (33)–(36) with the self-energies calculated as described in Secs. I–V. We thus neglect broadening of the subbands.

It is clear that in using the diagram (d) we do not overcount contributions from resonant screening and exchange-correlation effects. On the other hand, we are still neglecting some vertex corrections. In particular the interaction between the excited electron and the hole is left out. In Sec. VIII, we show that this interaction may even lead to a bound exciton, but we leave for future work an investigation of how the interaction affects the position of the maxima in infrared absorption.

VII. RESULTS AND COMPARISON WITH EXPERIMENTS

In Fig. 8 we show our results in the various approximations. The substrate doping is 7×10^{14} cm^{-3} corresponding to a depletion layer density of 1×10^{11} cm^{-2}. Curve (a) shows E_{10} in the Hartree approximation, and (b) shows the infrared absorption peak calculated from Eq. (32). In curve (c) is shown the energy separation between the lowest subbands at the Fermi wave vector when exchange and correlation are included. Finally, curve (d)

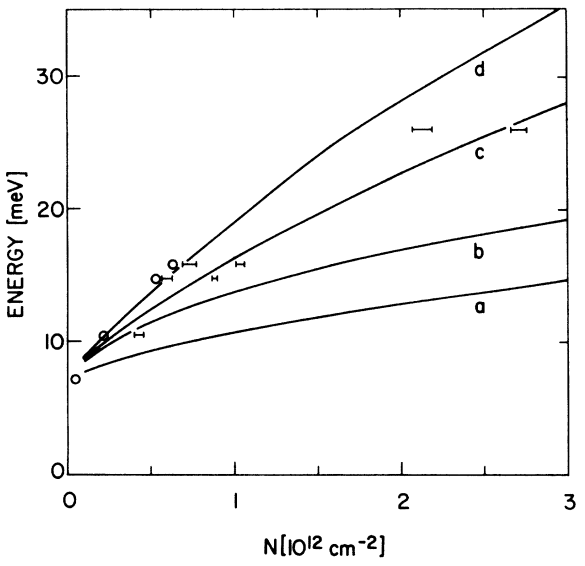

FIG. 8. Results on the peak absorption. The letters correspond to the approximations of Fig. 7. Experimental points from: Kneschaurek *et al.* (Ref. 6, O, infrared absorption) and Wheeler and Goldberg (Ref. 25, bars, photoresistivity). $N_A - N_D = 7\times10^{14}$ cm^{-3}.

shows results of the combined effect determined as the energy for which $\Pi(0,\omega)$ has a pole [Eq. (37)]. The experimental points are those of Kneschaurek *et al.*[6] whose sample had an effective depletion layer density of 10^{11} cm^{-2}. The agreement between experiment and curve (d) is seen to be excellent, the difference is less than 1 meV except for the very-low-density point. It is probably wise to be cautious about this good agreement. In particular a careful analysis of the experiment must be performed because there are many modes present in the waveguide and because the wave-vector dependence of the modes has not been included in our calculation. It is not necessarily certain that $q=0$ is a sufficient approximation for the mode which couples to the intersubband excitation in the inversion layer.

Another experiment which gives information on the subband structure of the inversion layer is the photoresistivity experiment of Wheeler and Goldberg.[25] They measured the change in the resistivity of the inversion layer when infrared radiation is beamed onto the layer through the silicon substrate. For each infrared frequency they saw two maxima in the resistivity as a function of inversion-layer density. In the standard interpretation of the results electrons are excited to a higher subband when the radiation energy is in resonance with the intersubband separation energy. In the higher subband the electrons presumably have a lower mobility and therefore the resistivity increases. If this interpretation is correct the photoresistivity peaks should be the same as the infrared absorption peaks.

Figure 8 also shows the photoresistivity results. If we compare with the infrared absorption results the most important difference is the fact that there are two maxima for each density. Furthermore it seems more reasonable to associate the maxima of higher energy with transitions from the lowest to the first excited subband. In that case we see that the theory shows good agreement except for the highest density. There are still minor quantitative differences between the two experimental measurements. It is not clear whether they are due to experimental uncertainties in determining the inversion-layer density for a given gate voltage or they are genuine differences because the experiments are different in principle and the interpretation of the photoresistivity results is too simplified. It should also be noted that the latter experiments the infrared radiation hits the inversion layer at an angle, whereas in the absorption experiment effort is made to propagate the radiation parallel to the layer.

The interpretation just presented leaves the resistivity maxima at low energy in Wheeler and Goldberg's results unexplained. We have earlier predicted[26] that a bound exciton might be formed in this system, and Sec. VIII describes the method which was used in that calculation. Still it would seem strange that such an exciton could be observed in the photoresistance experiment but not in the absorption experiment.

A different explanation is possible. Nakayama[27] has calculated the dispersion of a wave along the strip transmission line used in the infrared absorption measurement. He found a narrow band of frequencies bounded upwards by ω_{10}^*, the frequency for maximum conductivity at $q=0$, for which no wave can be transmitted. Since many modes are present in the waveguide the main contribution to the minimum in transmission could come from that forbidden band and from absorption by leaky surface modes.[17] In that case our comparison of the calculated ω_{10}^* with the experiment should be valid. In the photoresistivity measurement the incident radiation has a better defined wave vector parallel to the interface so the two peaks could correspond[17] to absorption by leaky surface modes and Berreman-type modes.[28] Quantitatively, however, this explanation does not agree with experiment, since the latter modes should lie above ω_{10}^* whereas both resistivity peaks lie below the infrared absorption maximum.

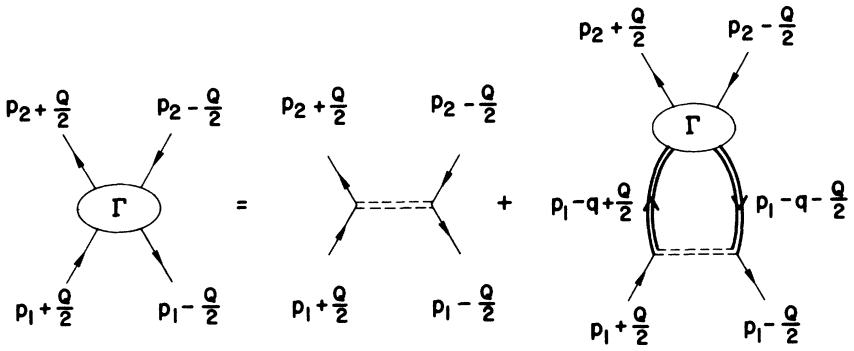

FIG. 9. Diagrammatic representation of the Bethe-Salpeter equation. Γ is the vertex; other symbols are the same as in Fig. 1.

VIII. THEORY OF INTERSUBBAND EXCITONS

Verdun and Drew[29] have found both experimental and theoretical evidence for an exciton between Landau levels in Bi. In this section, we apply the Bethe-Salpeter equation[30,31] and show that an analogous state is to be expected between the lowest and the excited subband.

We proceed along the lines described in Refs. 30 and 31, but retain the sign conventions of Ref. 9. In Fig. 9 is shown graphically the Bethe-Salpeter equation for electron-hole scattering. We use the notation $p=(\vec{p},E)$, $q=(\vec{q},\omega)$, and $Q=(\vec{Q},\Omega)$, where $\vec{Q}$ and Ω are the total momentum and energy, respectively, of the electron-hole pair. For the irreducible interaction we use the screened interaction in the state approximation $U_{ii00}(\vec{q},0)$ mainly to obtain manageable equations. The equation for the vertex is then

$$i\Gamma_{ii00}(p_1,p_2,Q)=iU_{ii00}(\vec{p}_2-\vec{p}_1,0)+\sum_j\int\frac{d^2q\,d\omega}{(2\pi)^3}iU_{jj00}(\vec{q},0)\mathcal{G}_{jj}(p_1-q+\tfrac{1}{2}Q)\mathcal{G}_{00}(p_1-q-\tfrac{1}{2}Q)i\Gamma_{jj00}(p_1-q,p_2,Q). \tag{38}$$

For a given $\vec{Q}$, bound states will reveal themselves as poles of the vertex.[30] Near a pole the first term and the terms from other subbands are negligible, so the energy Ω is determined as the eigenvalue of the homogeneous equation

$$\Gamma_{ii00}(p_1,p_2,Q)=i\int\frac{d^2q\,d\omega}{(2\pi)^3}U_{ii00}(\vec{q},0)\mathcal{G}_{ii}(p_1-q+\tfrac{1}{2}Q)\mathcal{G}_{00}(p_1-q-\tfrac{1}{2}Q)\Gamma_{ii00}(p_1-q,p_2,Q). \tag{39}$$

Since we are using the static screening approximation we can hardly expect very accurate results. Within a static approximation we can also neglect the E dependence of the self-energies for E near the quasiparticle energy, which means that we set quasiparticle renormalization factors equal to 1.

After integration over ω we obtain

$$\Gamma_{1100}(p_1,p_2,Q)=-\int\frac{d^2q}{(2\pi)^2}U_{1100}(q,0)\frac{\Theta(k_F-|\vec{p}_1-\vec{q}-\frac{1}{2}\vec{Q}|)\Gamma_{1100}(p_1-q,p_2,Q)}{\Omega-E_{10}-\xi(\vec{p}_1-\vec{q}+\frac{1}{2}\vec{Q})+\xi(\vec{p}_1-\vec{q}-\frac{1}{2}\vec{Q})-M_{11}+M_{00}}. \tag{40}$$

Since p_2 and E_1 are not affected by the integration we can write

$$\Gamma_{1100}(p_1,p_2,Q)=f(\vec{p}_1,Q)\,f_1(p_2,E_1)\times(\Omega-E_{10}-\xi_1+\xi_0-M_{11}+M_{00}), \tag{41}$$

where $\xi_1=\xi(\vec{p}_1+\frac{1}{2}\vec{Q})$, $\xi_0=\xi(\vec{p}_1-\frac{1}{2}\vec{Q})$ and

$$M_{11}=M_{11}[\vec{p}_1+\tfrac{1}{2}\vec{Q},E'(p_1+\tfrac{1}{2}\vec{Q})],$$
$$M_{00}=M_{00}[\vec{p}_1-\tfrac{1}{2}\vec{Q},E(\vec{p}_1-\tfrac{1}{2}\vec{Q})].$$

The equation for f is then

$$(\Omega-E_{10}-\xi_1+\xi_0-M_{11}+M_{00})f(\vec{p}_1,Q)=-\int\frac{d^2q}{(2\pi)^2}U_{1100}(\vec{q},0)\times\Theta(k_F-|\vec{p}_1-\vec{q}-\tfrac{1}{2}\vec{Q}|)f(\vec{p}_1-\vec{q},Q), \tag{42}$$

and if we assume that the quasiparticle bands are parallel, i.e., we neglect the small k dependence of $M_{11}-M_{00}$, we obtain for $\vec{Q}=0$

$$E_B f(\vec{p}_1) = \int \frac{d^2q}{(2\pi)^2} U_{1100}(\vec{p}_1 - \vec{q}, 0)\Theta(k_F - |\vec{q}|) f(\vec{q}), \tag{43}$$

where $f(\vec{p}_1) = f(\vec{p}_1; \vec{0}, \vec{\Omega})$, the binding energy $E_B = -(\Omega - E_{10}^c)$, and E_{10}^c is the energy separation between the quasiparticle subbands.

Equation (43) is a special form of the effective-mass equation. Since the bands have been assumed parallel, the reduced mass is infinite and we have no kinetic-energy term. This has the consequence that there is an infinite number of positive eigenvalues.[32]

Solutions of Eq. (43) can be separated:

$$f(\vec{p}_1) = u_m(p_1) e^{im\theta}, \tag{44}$$

where θ is a polar angle and m an integer. The effective potential

$$U^{(m)}(p,q) = \int_0^{2\pi} \frac{d\theta}{2\pi} e^{-im\theta} U_{1100}((p^2+q^2-2pq\cos\theta)^{1/2}) \tag{45}$$

for the wave functions with angular quantum number m,

$$E_B^{(m)} u_m(p) = \int_0^{k_F} U^{(m)}(p,q) u_m(q) \frac{dq}{2\pi}, \tag{46}$$

is much smaller for $m \neq 0$ than for $m = 0$ mainly because the screening wave vector s [see Eq. (14) in I] is of the same order of magnitude as the Fermi wave vector. The largest eigenvalue must, therefore, correspond to a solution with rotational symmetry.

We have solved Eq. (43) numerically and as was shown in Ref. 33 the binding energy varies from 0.9 to 1.8 meV for the inversion-layer density between 10^{12} and 3×10^{12} cm^{-2}.

The next eigenvalue of Eq. (43) for a rotationally symmetric eigenfunction was found to be about 20 times smaller than the largest eigenvalue, so the eigenvalue spectrum consists of one value of the order of 1 meV. All other eigenvalues are so small that they practically merge with the continuum and perhaps would not even lead to binding if the quasiparticle bands were not considered parallel.

It is obvious that we have made many simplifying assumptions. A more realistic calculation would have to take the imaginary parts of the self-energies into account, so that the lifetime of the exciton can be estimated. Furthermore the selection rules for exciting the exciton would have to be investigated. Nevertheless the estimated binding energy of the order of 1 meV is comparable to the separation of about 3 meV between the two resistivity peaks measured by Wheeler and Goldberg,[25] so we are led to speculate whether the lower-energy peaks correspond to excitation of the bound exciton.

IX. DISCUSSION AND CONCLUSION

We now compare our results to those of other calculations. Ohkawa[2] used a method very similar to ours. For the screening he used the Hubbard approximation. Furthermore, he included the contributions to screening from intersubband excitations but neglected the off-diagonal elements of the self-energy. If we compare his results for the self-energy at the Fermi wave vector in the lowest subband with ours, there is virtually no difference. This shows that the plasmon-pole approximation does not introduce errors compared with the Hubbard or with the random phase approximation. When we compare the self-energies at the Fermi wave vector in the excited subband we see a significant difference. This is not due to the different screening approximations but due to the fact that Ohkawa does not solve the Dyson equation (25). Instead he uses the value of the self-energy which one obtains by setting the quasiparticle energy equal to the energy of the noninteracting particle. As mentioned in Sec. III and shown in Fig. 3 this leads to different results. We still believe the method used here is better even though the self-energies are only approximate. Let us add that it would be difficult to extend the analysis for infrared absorption consistently to include resonant screening as in Sec. VI if we did not solve the Dyson equation.

The calculations by Ando[3] who applied a modified version of the local effective exchange correlation potential method[14] also in general give higher values for the subband separation than ours. The two calculations are widely different since Ando's method involves a perturbational calculation of the self-energy of a three-dimensional gas, but we believe that some of the difference can be ascribed to the fact that he used an exchange correlation potential which is independent of energy. Certainly in the case of the lowest subband of the fourfold degenerate set an energy dependent potential would change his results. Apart from the quantitative differences between his and our results there is also the qualitative difference that his calculations predict that with increasing inversion-layer density the first subband to be populated should belong to the doubly degenerate set. This is in contrast with our results and the experiments referred to in Sec. V.

Although there is generally good agreement between the experimental and theoretical results, some important facts are still not explained. It

was mentioned in Sec. V that the identical experiments[15,16] on the change in period of magnetoconductivity oscillations gave results that differ by more than experimental uncertainty. Most puzzling, however, is the fact that two maxima are observed in the photoresistivity experiment[25] but only one in the infrared absorption measurement.[6] Quantitatively, comparison of the two experiments seems to indicate that the higher-energy maxima are associated with the transition from the lowest to the first excited subband. An explanation would therefore involve (i) an excited state of lower energy which (ii) can be observed in photoresistivity but not in absorption. The possibility of an excitonic state as investigated in Sec. VIII fulfills the first criterion but probably not the second. The long response time measured in the photoresistivity (of the order 10^{-4}–10^{-3} sec) is not well understood. Döhler[22] has suggested that electrons are transferred from the first excited subband of the doubly degenerate set to the lowest subband of the fourfold degenerate set through impurity scattering. The probability for the electron to return to the ground state is very low if the latter subband is lower in energy than the former. Döhler's model depends crucially on E_0, being smaller than E_1 which is consistent with our results in Sec. IV. On the other hand his model does not directly explain the existence of two resistivity peaks and a full explanation of the photoresistivity results may have to await more experimental investigations.

In this work, we have calculated some effects of exchange and correlation on the excitation spectrum of the inversion layer. Although the model of screening that has been employed is rather simple, the results obtained show good agreement with several different experiments. We have stressed at several places, however, that the very good agreement should not be taken as proof that the theoretical description of the excitation spectrum in inversion layers is complete. There are still several experimental facts that are not explained, and on the theoretical side, other effects that those treated here could still turn out to be just as important.

ACKNOWLEDGMENTS

The author is indebted to F. Stern for supplying the Hartree wave functions and subband energies and for innumerous valuable discussions. He is also grateful to L. J. Sham and E. Burstein for discussions, and to the many authors who have sent copies of their work prior to publication.

*Present address: Institute of Physics, University of Aarhus, DK-8000 Aarhus C, Denmark.

[1]B. Vinter, Phys. Rev. B 13, 4447 (1976).
[2]F. J. Ohkawa, Surf. Sci. 58, 326 (1976); and thesis (Tokyo University, 1976) (unpublished).
[3]T. Ando, Phys. Rev. B 13, 3468 (1976).
[4]M. Jonson and G. Srinivasan (unpublished); and M. Jonson, J. Phys. C 9, 3055 (1976).
[5]A. K. Rajagopal and J. C. Kimball (unpublished).
[6]P. Kneschaurek, A. Kamgar, and J. F. Koch, Phys. Rev. B 14, 1610 (1976).
[7]F. Stern and W. E. Howard, Phys. Rev. 163, 816 (1967).
[8]F. Stern, Phys. Rev. B 5, 4891 (1972).
[9]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).
[10]B. I. Lundqvist, Phys. Kondens. Mater. 6, 193 and 206 (1967).
[11]A. W. Overhauser, Phys. Rev. B 3, 1888 (1970).
[12]L. Hedin, Phys. Rev. 139, A796 (1965).
[13]B. I. Lundqvist, Phys. Kondens. Mater. 7, 117 (1968).
[14]W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965); L. J. Sham and W. Kohn, *ibid.* 145, 561 (1966).
[15]W. E. Howard and F. F. Fang, Phys. Rev. B 13, 2519 (1976).
[16]D. C. Tsui and G. Kaminsky, Phys. Rev. Lett. 35, 1468 (1975).
[17]W. P. Chen, Y. J. Chen, and E. Burstein, Surf. Sci. 58, 263 (1976).
[18]S. J. Allen, Jr., D. C. Tsui, and B. Vinter, Solid State Commun. 20, 425 (1976).
[19]F. Stern, Jpn. J. Appl. Phys. Suppl. 2, 323 (1974).
[20]F. Stern, Phys. Rev. Lett. 18, 546 (1967).
[21]F. Stern (private communication).
[22]G. H. Döhler, Solid State Commun. 18, 633 (1976).
[23]M. J. Kelly and L. M. Falicov, Phys. Rev. Lett. 37, 1021 (1976).
[24]L. J. Sham (private communication).
[25]R. G. Wheeler and H. S. Goldberg, IEEE Trans. Electron Devices ED-22, 1001 (1975).
[26]B. Vinter, Phys. Rev. Lett. 35, 598 (1975).
[27]M. Nakayama (unpublished).
[28]D. W. Berreman, Phys. Rev. 130, 2193 (1963).
[29]H. R. Verdun and H. D. Drew, Phys. Rev. Lett. 33, 1608 (1974).
[30]P. Nozières, *Theory of Interacting Fermi Systems* (Benjamin, New York, 1964).
[31]L. J. Sham and T. M. Rice, Phys. Rev. 144, 708 (1965).
[32]W. Pogorzelski, *Integral Equations and Their Applications* (PWN-Polish Scientific, Warsaw, 1966), p. 132.
[33]B. Vinter and F. Stern, Surf. Sci. 58, 141 (1976).