need to cut down on computer time. For both of these reasons, we first did the calculations by approximating the $3s$ and $3p$ K^+ ion orbitals by single Slater functions with parameters chosen according to the usual prescription. This calculation gave not unreasonable results, but the approximation was too gross to be very satisfactory. Had we continued using it, the extent of the H^- polarization would most likely have been misleadingly small. The calculations were useful, though, in that they provided good starting values for the parameters β_a and β_b in the trial functions of the more exact calculations with HF ion orbitals. In connection with H^- polarization effects, we did a limited number of calculations to ascertain if it would be a good approximation to neglect the electronic structure on the four 1nn ions at right angles to the displacement of the H^- ion. We soon found out that these ions are quite important, especially to the polarization effect.

We hope to extend these calculations soon to the sodium halides and thus obtain further insight into the validity of some of the approximations.

ACKNOWLEDGMENT

We wish to thank R. Tseng for numerous, useful discussions.

PHYSICAL REVIEW VOLUME 162, NUMBER 3 15 OCTOBER 1967

Anisotropy Effects in the Excitonic Insulator*

JOHANNES ZITTARTZ†
University of California, San Diego, La Jolla, California

(Received 10 April 1967)

A new insulating phase in semiconductors and semimetals has been discussed recently in the literature. This paper extends the analysis of this excitonic phase to the more general case of an anisotropic band structure of the underlying two-band model. The thermodynamic properties of the excitonic phase are worked out. Whereas the anisotropy does not lead to qualitativly new features in the case of a positive band gap (semiconductor region), important changes occur in the case of a negative band gap (semimetallic region). The latter situation is discussed in detail. Finally, the phase diagram of the excitonic phase is calculated for a particular anisotropic band model.

I. INTRODUCTION

RECENTLY a new kind of insulating phase has been discussed in the literature. The idea is based on observations by Mott[1] and Knox[2] that in semimetals and semiconductors with small energy gaps the electrons and holes will under certain conditions form bound pairs of the exciton type. The properties of the new phase which has been called the "excitonic insulator"[3] have been investigated theoretically in several papers.[3-8] While most of these papers deal with the simple case of an isotropic band structure, Kopaev[7] has pointed out the importance of anisotropy in the description of the new phase.

In this article we investigate in detail the excitonic phase in the more realistic case of anisotropic energy bands. In Sec. II we derive the thermodynamic properties of the new phase quite generally. The influence of the anisotropy is discussed in detail in Secs. III and IV. For a positive band gap we find essentially the same results as in the isotropic case[3,5,6] apart from quantitative corrections. In the semimetallic region (band overlap), however, the description differs qualitatively from the corresponding isotropic case. The electron system condenses only partly in electron-hole pairs, whereas parts of the system remain in the normal state. It follows especially that the normal semimetallic state is stable against pair formation, if the anisotropy is large enough as has been shown first by Kopaev.[7]

We consider the familiar two-band picture of valence and conduction electrons and assume that Coulomb interactions of electrons within one band have been taken into account in the calculation of the energy bands. The dynamics of the system are then governed

---

* Supported in part by the Office of Naval Research and the National Science Foundation.

† On leave from Institute for Theoretical Physics, University of Cologne, Cologne, Germany until Oct. 1, 1967.

[1] N. F. Mott, Phil. Mag. 6, 287 (1961).

[2] R. S. Knox, in *Theory of Excitons, Solid-State Physics* (Academic Press Inc., New York, 1963), Suppl. 5, p. 100.

[3] D. Jerome, T. M. Rice, and W. Kohn, Phys. Rev. 158, 248 (1967).

[4] J. Des Cloizeaux, J. Phys. Chem. Solids 26, 259 (1965).

[5] L. V. Keldysh and Yu. V. Kopaev, Fiz. Tverd. Tela 6, 2791 (1964) [English transl.: Soviet State 6, 2219 (1965)].

[6] A. N. Kozlov and L. A. Maksimov, Zh. Eksperim. i Teor. Fiz. 48, 1184 (1965); 49, 1284 (1965) [English transls.: Soviet Phys.—JETP 21, 790 (1965); 22, 889 (1966)].

[7] Yu. V. Kopaev, Fiz. Tverd. Tela 8, 223 (1966) [English transl.: Soviet Phys.—Solid State 8, 175 (1966)]. In fact Kopaev shows from an analysis of the electron-hole vertex part that the instability of the normal state does not occur, if the anisotropy is large enough. However, his "gap equation" is in error as may be compared with our corresponding Eq. (32).

[8] E. V. Baklanov and A. V. Chaplik, Fiz. Tverd. Tela 7, 2768 (1965) [English transl.: Soviet Phys.—Solid State 7, 2240 (1966)].

by the many-particle Hamiltonian

$$H=\sum_{\mathbf{p}}\{\epsilon_a(\mathbf{p})a_{\mathbf{p}}^{\dagger}a_{\mathbf{p}}+\epsilon_b(\mathbf{p})b_{\mathbf{p}}^{\dagger}b_{\mathbf{p}}\} +\frac{1}{\Omega}\sum_{\mathbf{q}\mathbf{p}\mathbf{p}'} V(\mathbf{q})a_{\mathbf{p}+\mathbf{q}}^{\dagger}a_{\mathbf{p}}b_{\mathbf{p}'-\mathbf{q}}^{\dagger}b_{\mathbf{p}'}. \quad (1)$$

Here the a, $a^{\dagger}$ and b, $b^{\dagger}$ are annihilation and creation operators of valence and conduction electrons, respectively. The corresponding energies are denoted by ϵ_a and ϵ_b and are reckoned from the chemical potential. For simplicity we assume that the valence band has a single maximum at $\mathbf{p}=0$ and that the conduction band has a single minimum at $\mathbf{p}=\mathbf{w}$. [We can effectively eliminate the vector $\mathbf{w}$ from the Hamiltonian (1), if the operators $b_{\mathbf{p}}$ and $b_{\mathbf{p}}^{\dagger}$ and the energy $\epsilon_b(\mathbf{p})$ refer to conduction electrons with momentum $\mathbf{p}+\mathbf{w}$.] In the effective-mass approximation and taking into account anisotropy effects we describe the energies near $\mathbf{p}=0$ by

$$\begin{aligned}\epsilon_b(\mathbf{p})&=B+\frac{1}{2}\left(\frac{p_x^2}{m_{bx}}+\frac{p_y^2}{m_{by}}+\frac{p_z^2}{m_{bz}}\right),\\ \epsilon_a(\mathbf{p})&=-\left[A+\frac{1}{2}\left(\frac{p_x^2}{m_{ax}}+\frac{p_y^2}{m_{ay}}+\frac{p_z^2}{m_{az}}\right)\right].\end{aligned} \quad (2)$$

In the semiconductor case both A and B are positive and the gap separating valence and conduction band is

$$G=A+B. \quad (3)$$

If we deal with a semimetal both A and B are negative and G describes the overlap of the two bands. In this case the noninteracting ground state is partially filled with holes and electrons. For given A we may calculate B from the condition of electrical neutrality.

In the interaction part of the Hamiltonian (1) we retain only the direct Coulomb interaction of valence and conduction electrons with

$$V(\mathbf{q})=4\pi e^2/K(\mathbf{q})q^2, \quad (4)$$

where $K(\mathbf{q})$ is an effective dielectric constant.

II. GENERAL DESCRIPTION OF THE EXCITONIC PHASE

In this section we derive the basic thermodynamic properties of the new phase for the two-band model given by the Hamiltonian (1).

The approximation scheme we adopt for the present problem is the finite temperature version of the Hartree-Bogoliubov variational principle.

It has been realized that the normal ground state of (1) becomes unstable against the formation of excitons on a macroscopic scale. This leads, in obvious analogy to the theory of superconductivity, to the problem of recalculating the ground state and the energy spectrum. At finite temperatures excitons may be described by a nonvanishing thermal average

$$\langle b_{\mathbf{p}+\mathbf{Q}}^{\dagger}a_{\mathbf{p}}\rangle\neq 0, \quad (5)$$

where $\mathbf{Q}$ is the exciton momentum. From the theory of superconductivity it is well known that only "coherent pairing" (i.e., constant pair momentum $\mathbf{Q}$) leads to macroscopic effects. Therefore we have to choose a common momentum $\mathbf{Q}$ for all excitonic pairs, and as the $\mathbf{Q}=0$ excitons have lowest energy, we confine our attention to zero-momentum pairs.

The thermodynamic potential Φ can be written as a functional of the density matrix ρ[9]:

$$\Phi=\mathrm{Tr}[\rho H+(1/\beta)\rho\ln\rho], \quad (6)$$

with the property that $\Phi[\rho]$ attains its minimum for the correct density matrix. Thus

$$\partial\Phi/\partial\rho=0, \quad \mathrm{Tr}\rho=1 \quad (7)$$

leads to

$$\rho=e^{-\beta H}/\mathrm{Tr}e^{-\beta H}. \quad (8)$$

We obtain an upper bound to the true thermodynamic potential when ρ is restricted to a certain class of operators. Choosing

$$\bar{\rho}=\exp(-\beta\bar{H})/\mathrm{Tr}\exp(-\beta\bar{H}), \quad (9)$$

where $\bar{H}$ is an approximate Hamiltonian, we get the upper bound

$$\Phi=-(1/\beta)\ln\mathrm{Tr}[\exp(-\beta\bar{H})]+\langle H-\bar{H}\rangle, \quad (10)$$

with

$$\langle\mathcal{O}\rangle=\mathrm{Tr}\bar{\rho}\mathcal{O}. \quad (11)$$

Within the spirit of emphasizing bound pairs of valence and conduction electrons, we then consider the pair Hamiltonian

$$\bar{H}=\sum_{\mathbf{p}}\{\epsilon_a(\mathbf{p})a_{\mathbf{p}}^{\dagger}a_{\mathbf{p}}+\epsilon_b(\mathbf{p})b_{\mathbf{p}}^{\dagger}b_{\mathbf{p}}+[\Delta(\mathbf{p})b_{\mathbf{p}}^{\dagger}a_{\mathbf{p}}+\mathrm{c.c.}]\}, \quad (12)$$

which besides the noninteracting part of the true Hamiltonian (1) contains a term describing annihilation or creation of excitons. The parametric function $\Delta(\mathbf{p})$ can be chosen real and positive, as any phase factor $e^{i\varphi}$ can be removed by a corresponding phase transformation of the operators $a_{\mathbf{p}}(a_{\mathbf{p}}\to e^{-i\varphi}a_{\mathbf{p}})$. The function $\Delta(\mathbf{p})$ is fixed by imposing the stationary condition

$$\frac{\partial\Phi}{\partial\Delta(\mathbf{p})}=0; \quad (13)$$

thus the approximate thermodynamic potential Φ [Eq. (10)] achieves its lowest value.

As the Hamiltonian $\bar{H}$ can be diagonalized by a Bogoliubov transformation, the calculation of the thermodynamic potential (10) is straightforward. Introducing new Fermi operators

$$\begin{aligned}a_{\mathbf{p}}&=u_{\mathbf{p}}\alpha_{\mathbf{p}}+v_{\mathbf{p}}\beta_{\mathbf{p}},\\ b_{\mathbf{p}}&=u_{\mathbf{p}}\beta_{\mathbf{p}}-v_{\mathbf{p}}\alpha_{\mathbf{p}}, \quad u_{\mathbf{p}}^2+v_{\mathbf{p}}^2=1,\end{aligned} \quad (14)$$

---

[9] N. D. Mermin, Phys. Rev. 137, A1441 (1965).

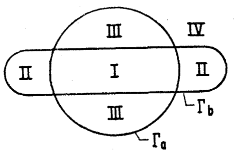

FIG. 1. Fermi surfaces and particle distribution of the anisotropic semimetal in the normal case. Γ_a is the Fermi surface of valence electrons; Γ_b is the Fermi surface of conduction electrons. I, II, III, IV: the different regions of distribution.

we get

$$u_{\mathbf{p}}^2=\frac{1}{2}\left(1+\frac{\xi}{E}\right),\quad v_{\mathbf{p}}^2=\frac{1}{2}\left(1-\frac{\xi}{E}\right),\quad u_{\mathbf{p}}v_{\mathbf{p}}=\frac{1}{2}\frac{\Delta}{E},\tag{15}$$

with the abbreviations

$$\xi_{\mathbf{p}}=\tfrac{1}{2}[\epsilon_b(\mathbf{p})-\epsilon_a(\mathbf{p})],\quad \eta_{\mathbf{p}}=\tfrac{1}{2}[\epsilon_b(\mathbf{p})+\epsilon_a(\mathbf{p})],\tag{16}$$
$$E_{\mathbf{p}}=(\xi_{\mathbf{p}}^2+\Delta_{\mathbf{p}}^2)^{1/2}.$$

In terms of the new operators (14) the pair Hamiltonian becomes

$$\bar{H}=\sum_{\mathbf{p}}\{(\eta_{\mathbf{p}}-E_{\mathbf{p}})\alpha_{\mathbf{p}}^\dagger\alpha_{\mathbf{p}}+(E_{\mathbf{p}}+\eta_{\mathbf{p}})\beta_{\mathbf{p}}^\dagger\beta_{\mathbf{p}}\},\tag{17}$$

and using this result we may calculate the thermodynamic potential

$$\begin{aligned}\Phi=\sum_{\mathbf{p}}\biggl[(\eta-E)&+\frac{1}{\beta}\ln\{(1-n_{E+\eta})(1-n_{E-\eta})\}\\&+\frac{\Delta^2}{E}(1-n_{E+\eta}-n_{E-\eta})\biggr]\\&-\frac{1}{\Omega}\sum_{\mathbf{pp}'}V(\mathbf{p}-\mathbf{p}')\frac{\Delta}{2E}(1-n_{E+\eta}-n_{E-\eta})\\&\qquad\times\frac{\Delta'}{2E'}(1-n_{E'+\eta'}-n_{E'-\eta'}),\end{aligned}\tag{18}$$

where primed and unprimed quantities refer to $\mathbf{p}'$ and $\mathbf{p}$, respectively, and where n_x is the Fermi distribution function. The stationary condition (13) leads to the gap equation

$$\Delta(\mathbf{p})=\frac{1}{\Omega}\sum_{\mathbf{q}}V(\mathbf{p}-\mathbf{q})\frac{\Delta(\mathbf{q})}{2E_{\mathbf{q}}}(1-n_{E+\eta}-n_{E-\eta})_{\mathbf{q}}.\tag{19}$$

Inserting this into the expression (18) we get a final result

$$\begin{aligned}\Phi=\sum_{\mathbf{p}}\biggl[(\eta-E)&+\frac{1}{\beta}\ln\{(1-n_{E+\eta})(1-n_{E-\eta})\}\\&+\frac{\Delta^2}{2E}(1-n_{E+\eta}-n_{E-\eta})\biggr].\end{aligned}\tag{20}$$

The last two equations in principle determine the thermodynamics of the system.

Another useful formula for the thermodynamic potential Φ is obtained from

$$\Phi(\lambda)-\Phi(0)=\int_0^\lambda\frac{d\lambda'}{\lambda'}\left(\frac{\partial\Phi}{\partial\lambda}\right)_{\lambda'},\tag{21}$$

where λ is the coupling constant and the integration goes over all intermediate values. From (18) and using (13) and (19) we get

$$\Phi(\lambda)-\Phi(0)=-\int_0^\lambda\frac{d\lambda'}{\lambda'}\left\{\sum_{\mathbf{p}}\frac{\Delta^2}{2E}(1-n_{E+\eta}-n_{E-\eta})\right\}_{\lambda'}.\tag{22}$$

As the temperature factor on the right-hand side is always positive, we conclude quite generally that Φ attains its minimum for the nontrivial solution of the gap equation, if it exists.

Finally we derive expressions for the average occupation of valence- and conduction-electron states in the new phase. Using the thermodynamic relations

$$\langle a_{\mathbf{p}}^\dagger a_{\mathbf{p}}\rangle=\partial\Phi/\partial\epsilon_a(\mathbf{p}),\quad\langle b_{\mathbf{p}}^\dagger b_{\mathbf{p}}\rangle=\partial\Phi/\partial\epsilon_b(\mathbf{p}),\tag{23}$$

we obtain from (18) using (15)

$$\begin{aligned}\langle a_{\mathbf{p}}^\dagger a_{\mathbf{p}}\rangle&=u_{\mathbf{p}}^2(1-n_{E-\eta})+v_{\mathbf{p}}^2n_{E+\eta},\\\langle b_{\mathbf{p}}^\dagger b_{\mathbf{p}}\rangle&=u_{\mathbf{p}}^2n_{E+\eta}+v_{\mathbf{p}}^2(1-n_{E-\eta}),\end{aligned}\tag{24}$$

whereas the excitonic pair is described by

$$\langle b_{\mathbf{p}}^\dagger a_{\mathbf{p}}\rangle=-(\Delta/2E)(1-n_{E+\eta}-n_{E-\eta})|_{\mathbf{p}}.\tag{25}$$

It should be remarked that we do not distinguish between singlet or triplet exciton states, as we have suppressed spin indices and neglected exchange interactions.[6]

III. INFLUENCE OF AN ANISOTROPIC BAND STRUCTURE

In the general derivation of the last section we did not have to consider the particular band structure. For a more detailed discussion we now have to distinguish between the two possible cases, namely, semiconductor or semimetallic case.

A. *Semiconductor.* In this case both conduction- and valence-electron energies ϵ_b and ϵ_a satisfy

$$\epsilon_b>0,\quad\epsilon_a<0\quad\text{for all }\mathbf{p},\tag{26}$$

which via (16) implies

$$E\geq\xi\geq|\eta|\quad\text{for all }\mathbf{p}.\tag{27}$$

This means that all arguments of the two Fermi functions in the formulas of the preceding section are positive, and the functions tend to zero at $T=0$. From (24) and (25) we especially obtain

$$\begin{aligned}\langle a_{\mathbf{p}}^\dagger a_{\mathbf{p}}\rangle_{T=0}&=u_{\mathbf{p}}^2,\\\langle b_{\mathbf{p}}^\dagger b_{\mathbf{p}}\rangle_{T=0}&=v_{\mathbf{p}}^2,\\\langle b_{\mathbf{p}}^\dagger a_{\mathbf{p}}\rangle_{T=0}&=-\Delta/2E,\end{aligned}\tag{28}$$

This means that the ground state which in the normal case is a completely filled valence band contains electron-hole pairs, if $\Delta\neq 0$. The positive kinetic energy we need to form this state is more than compensated by the binding energy we gain from the formation of pairs. It is to be remarked that a possible anisotropy of the band energies does not affect this picture qualitatively, but can only alter the quantitative calculations of thermodynamic quantities in the isotropic case considered in the literature.[6] Especially it follows that the gap equation (19) has a nontrivial solution as long as the band gap G satisfies

$$0\leq G\leq E_B\,, \tag{29}$$

where E_B is the binding energy of an exciton in the usual exciton theory, taking into account the anisotropic band structure.

B. *Semimetal.* In this case both constants A and B in the one-particle energies ϵ_b and ϵ_a are negative, and the normal ground state contains an equal number of holes and electrons. (If not stated otherwise, "holes" refer to the valence band and "electrons" to conduction electrons.) Due to the anisotropy in the energy dispersion the corresponding Fermi surfaces do not have the same shape. The situation is shown in Fig. 1. In this two-dimensional plot we denote the Fermi surfaces of holes by Γ_a and of electrons by Γ_b. The different regions I, II, III, and IV have the following meaning. Region I lies entirely within Γ_a and Γ_b and is therefore occupied by holes and electrons in the normal state. Region IV is outside of both Fermi surfaces and is therefore empty. Region II is occupied by electrons alone, but contains no holes; similarly region III is occupied by holes, but there are no electrons. Note that in the case of identical Fermi surfaces one only has to deal with the regions I and IV. This is the usual case in the theory of superconductivity and the situation so far considered for the excitonic insulator.[3,5,6]

In the excitonic phase the average occupation at $T=0$ can be read off from the formulas (24) and (25) as before. Figure 2 shows the modified situation. We now have to distinguish between the following regions in momentum space:

$$E\geq|\xi|\geq|\eta|\,,\quad n_{E-\eta}=n_{E+\eta}=0$$
$$\text{in I and IV}\quad(\text{I}:\xi<0,\quad\text{IV}:\xi>0)\,, \tag{30a}$$

which via (24) and (25) leads to the distribution

$$\langle a_{\mathbf{p}}a_{\mathbf{p}}{}^{\dagger}\rangle=\langle b_{\mathbf{p}}{}^{\dagger}b_{\mathbf{p}}\rangle=v_{\mathbf{p}}{}^{2}\,,$$
$$\langle b_{\mathbf{p}}{}^{\dagger}a_{\mathbf{p}}\rangle=-\Delta/2E\quad(\text{I and IV})\,. \tag{31a}$$

The former regions II and III each split into two subregions. First we have II_1 and III_1 with

$$\begin{aligned}-\eta\geq E\geq|\xi|\,,\quad n_{E-\eta}=0\,,\quad n_{E+\eta}=1&\\ \text{in }\text{II}_1(\eta<0)\,,&\\ \eta\geq E\geq|\xi|\,,\quad n_{E-\eta}=1\,,\quad n_{E+\eta}=0&\\ \text{in }\text{III}_1(\eta>0)\,,&\end{aligned} \tag{30b}$$

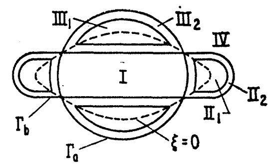

FIG. 2. Particle distribution in the excitonic phase. The broken line denotes the "common" surface $\xi=0$. I, IV, II_2, III_2: different regions of condensed particles. II_1, III_1: regions of normal distribution.

and the distribution

$$\begin{aligned}\langle a_{\mathbf{p}}a_{\mathbf{p}}{}^{\dagger}\rangle=0\,,\quad\langle b_{\mathbf{p}}{}^{\dagger}b_{\mathbf{p}}\rangle=1\,,\quad\langle b_{\mathbf{p}}{}^{\dagger}a_{\mathbf{p}}\rangle=0\quad\text{in }\text{II}_1\,,\\ \langle a_{\mathbf{p}}a_{\mathbf{p}}{}^{\dagger}\rangle=1\,,\quad\langle b_{\mathbf{p}}{}^{\dagger}b_{\mathbf{p}}\rangle=0\,,\quad\langle b_{\mathbf{p}}{}^{\dagger}a_{\mathbf{p}}\rangle=0\quad\text{in }\text{III}_1\,.\end{aligned} \tag{31b}$$

From this we see that the particles in the regions II_1 and III_1 do not participate in the condensation process, but remain "normal." This novel situation may be explained by the fact that in the normal state only one of the partners which form a pair is present. The amount of kinetic energy needed to create the missing partner would be greater than the gain in binding energy. Only in the small regions II_2 and III_2 which are near either of the two Fermi surfaces (Fig. 2) is the binding energy large enough to compensate the kinetic energy. The regions II_2 and III_2 are defined by

$$E\geq|\eta|\geq|\xi|\,,\quad n_{E+\eta}=n_{E-\eta}=0$$
$$\text{in }\text{II}_2(\eta<0)\text{ and }\text{III}_2(\eta>0)\,, \tag{30c}$$

and the distribution is the same as in I and IV [Eq. (31a)]. As can be seen from the inequality in (30c), the extent of these regions in momentum space is itself determined by Δ and vanishes with vanishing Δ. If on the other hand the Fermi surfaces Γ_a and Γ_b lie close together, the binding energy may be larger than the required kinetic energy in the whole regions II and III, and the whole system condenses in electron-hole pairs. This will be the case for very small Fermi surfaces when the negative band gap G is nearly zero, i.e., $G\lesssim 0$.

It should be emphasized that the pair function $\langle b_{\mathbf{p}}{}^{\dagger}a_{\mathbf{p}}\rangle$ vanishes in the regions II_1 and III_1 though the corresponding gap function $\Delta(\mathbf{p})$ is not zero. Actually the function $\Delta(\mathbf{p})$ no longer reflects a gap in the quasiparticle spectrum. The quasiparticle energies may be taken from the pair Hamiltonian (17) and are given by ω_α and ω_β, respectively:

$$\omega_\alpha=\eta-E\,,\quad\omega_\beta=\eta+E\,.$$

It is obvious from (30b) and (30c) that ω_α vanishes on the boundary between III_1 and III_2, whereas ω_β vanishes on the boundary between II_1 and II_2.

At finite temperatures this simple picture does not quite hold, as all states are thermally populated. This means especially that in II_1 and III_1 the pair function $\langle b_{\mathbf{p}}{}^{\dagger}a_{\mathbf{p}}\rangle$ starts out at $T=0$ with zero value, increases to a certain finite value, and decreases again until it vanishes at the critical temperature.

As we are dealing with a cooperative phenomenon, one finds that the system not only distinguishes between condensing and noncondensing particles, but also that the condensation process itself is greatly affected by the

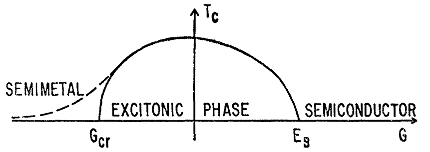

FIG. 3. Phase diagram of the anisotropic excitonic phase. The broken line corresponds to the isotropic case (Refs. 3 and 6).

presence of normal particles. This may be seen by looking for nontrivial solutions of the gap equation at $T=0$:

$$\Delta(\mathbf{p})=\frac{1}{\Omega}\sum_{\mathbf{q}}{}' V(\mathbf{p}-\mathbf{q})\frac{\Delta(\mathbf{q})}{2E_{\mathbf{q}}}. \tag{32}$$

There are two important differences from the familiar isotropic case. First, the temperature factor, $1-n_{E+\eta}-n_{E-\eta}$, in (19) vanishes in the regions II_1 and III_1 which restricts the integration in (32), as indicated by the prime. Roughly speaking this diminishes the gap function as compared to the isotropic case. Secondly we shall conclude that the condensation will not occur for a sufficiently weak interaction, as Eq. (32) will in that case only have the trivial solution; in other words the system is stable against formation of excitons, unless the attractive interaction between holes and electrons exceeds a critical value.

This is in obvious contrast to the situation in the isotropic case and to the situation in superconductivity. In the isotropic case the linearized gap equation

$$\Delta(\mathbf{p})=\frac{1}{\Omega}\sum_{\mathbf{q}}{}' V(\mathbf{p}-\mathbf{q})\frac{\Delta(\mathbf{q})}{2|\xi_{\mathbf{q}}|} \tag{33}$$

exhibits a logarithmic singularity which comes from an integration over the Fermi surface $\xi=0$, and which eventually is responsible for the existence of a nontrivial solution of the gap equation (32), no matter how weak the interaction is. In our anisotropic case the "common Fermi surface" $\xi=0$, entirely lies in the forbidden region as indicated in Fig. 2.[10] The integration in (33) extends over the regions I and IV which in three dimensions touch the surface $\xi=0$ in two lines only, namely, the intersection of the Fermi surfaces of the valence and conduction electrons, denoted by Γ_a and Γ_b. The corresponding line singularity in (33) is integrable in three dimensions. Therefore the right-hand side of Eq. (33) will be finite and, if the interaction is weak enough, will be smaller than the left-hand side. In this case we are left only with the trivial solution $\Delta(\mathbf{p})\equiv 0$. A quantitative estimate for nearly isotropic bands is given in the next section.

IV. PHASE DIAGRAM

In Fig. 3 we plot qualitatively the transition temperature T_c versus band gap G. This phase diagram was originally suggested by Kohn[11] and differs significantly from the isotropic case on the semimetallic side $(G<0)$.[3,6] In order to calculate the characteristics we confine ourselves to an isotropic valence band

$$\epsilon_a=-[A+p^2/2m_a], \tag{34}$$

whereas we assume an anisotropic conduction band

$$\epsilon_b=B+(1/2m_b)(p_x^2+p_y^2+\gamma p_z^2),\quad 0<\gamma<1 \tag{35}$$

with a larger effective mass m_b/γ in the z direction.

(a) The limit for the excitonic phase in the semiconductor region is determined by the linearized gap equation (19) at $T=0$. Introducing

$$\psi(\mathbf{q})=\frac{\Delta(\mathbf{q})}{\epsilon_b(\mathbf{q})-\epsilon_a(\mathbf{q})}, \tag{36}$$

and Fourier transforming we obtain the Schrödinger equation

$$\left[-\frac{1}{2\mu}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\delta\frac{\partial^2}{\partial z^2}\right)-\frac{e^2}{Kr}\right]\psi(\mathbf{r})=-G\psi(\mathbf{r}), \tag{37}$$

where we have explicitly introduced the interaction (4) with a static dielectric constant K, due to electrons in low-lying energy bands. μ is the reduced mass for the isotropic case:

$$\frac{1}{\mu}=\frac{1}{m_a}+\frac{1}{m_b}, \tag{38}$$

and

$$\delta=1-\alpha(1-\gamma),\quad \alpha=m_a/(m_a+m_b). \tag{39}$$

It is clear that the largest possible band gap G corresponds to the lowest eigenvalue of (37), i.e., the binding energy E_B. Kohn and Luttinger[12] have studied the solutions of Eq. (37) in connection with impurity states in Si. Writing

$$E_B(\delta)=f(\delta)E_B(1),\quad E_B(1)=\tfrac{1}{2}\mu(e^2/K)^2, \tag{40}$$

where $E_B(1)$ is the binding energy in the isotropic case, we conclude from their work that the function $f(\delta)$ increases monotonically between the limiting values

$$f(1)=1,\quad f(0)=4. \tag{41}$$

In our case, $\delta=0$ is not permissible, as in the extreme anisotropic limit $(\gamma=0)$, δ is still finite, $\delta=1-\alpha$; therefore the function $f(\delta)$ ranges between 1 and a value somewhat less than 4.

(b) *Touching bands.* Kozlov and Maksimov[6] have

---

[10] It should be remarked that portions of the surface $\xi=0$ actually lie within the "allowed" regions II_2 and III_2. But as the linear dimensions of these regions are themselves determined by Δ, as stated below Eq. (30c), their contributions to the right-hand side in (32) are of higher order in Δ and do not appear in the linearized equation.

[11] W. Kohn, in *Physics of Solids at High Pressures*, edited by C. T. Tomizuka and R. M. Emrick (Academic Press Inc., New York, 1965), p. 561.

[12] W. Kohn and J. M. Luttinger, Phys. Rev. 98, 915 (1955).

calculated the maximal energy gap which occurs for $G=0$. In terms of (40) the result for isotropic energies is

$$\Delta_0 = (8/\pi^2)E_B(1). \tag{42}$$

The corresponding transition temperature in the anisotropic case is determined by solving the linearized gap equation at the finite temperature T_c:

$$\Delta(\mathbf{p}) = \frac{1}{(2\pi)^3}\int d^3q \frac{4\pi e^2/K}{|\mathbf{p}-\mathbf{q}|^2} \frac{\Delta(\mathbf{q})}{\epsilon_b(\mathbf{q})-\epsilon_a(\mathbf{q})} \times \frac{1}{2}\left\{\tanh\frac{\beta\epsilon_b(\mathbf{q})}{2} - \tanh\frac{\beta\epsilon_a(\mathbf{q})}{2}\right\}, \tag{43}$$

where ϵ_a and ϵ_b are given by (34) and (35) with $A=B=0$. As the integrand drops off rather rapidly for large $|\mathbf{q}|$, we set $\Delta(\mathbf{q})\approx\Delta(0)$ on the right-hand side. Then setting $\mathbf{p}=0$, Δ drops out completely, and we are left with an integral relation for T_c. Introducing polar coordinates this can be written as

$$1 = \frac{\mu e^2/K}{2\pi^2}\int d\Omega \int_0^\infty \frac{dq}{q^2} \frac{1}{1-\alpha(1-\gamma)\cos^2\Theta} \times\left\{\tanh\frac{\beta q^2}{4m_a} + \tanh\left[\frac{\beta q^2}{4m_b}(1-(1-\gamma)\cos^2\Theta)\right]\right\}, \tag{44}$$

where we have used (38) and (39) and denote the inclination angle against the z axis by Θ. The integrations in (44) decouple when we transform the radial variable q to include the whole argument of the tanh function in both cases. Thus the remaining calculation is elementary and gives

$$KT_c = CF^2(\alpha,\gamma)E_B(1), \tag{45}$$

where the constant C is

$$C = 2\left[\frac{1}{\pi}\int_0^\infty \frac{dx}{x^2}\tanh x^2\right]^2 = \left(\frac{2}{\pi}\right)^3 [(1-2^{-3/2})\zeta(\tfrac{3}{2})]^2 \approx 0.72, \tag{46}$$

the integration being done by residue techniques. The function $F(\alpha,\gamma)$ is the result of the Θ integration and is given explicitly by

$$\begin{aligned} F(\alpha,\gamma) &= \int_0^{\pi/2} \frac{\sin\Theta d\Theta}{1-\alpha(1-\gamma)\cos^2\Theta} \times\{(1-\alpha)^{1/2}+\alpha^{1/2}(1-(1-\gamma)\cos^2\Theta)^{1/2}\} \\ &= \left[\frac{1-\alpha}{\alpha(1-\gamma)}\right]^{1/2}\left\{\tfrac{1}{2}\ln\frac{1+[\alpha(1-\gamma)]^{1/2}}{1-[\alpha(1-\gamma)]^{1/2}} + \frac{1}{(1-\alpha)^{1/2}}\arctan\left[\frac{1-\gamma}{\gamma}\right]^{1/2} - \arctan\left[(1-\alpha)^{1/2}\left(\frac{1-\gamma}{\gamma}\right)^{1/2}\right]\right\}. \end{aligned} \tag{47}$$

From (40) and (45) we may deduce a relationship between the binding energy $E_B(\delta)$ and the transition temperature for touching bands:

$$\frac{KT_c}{E_B} = C\frac{F^2(\alpha,\gamma)}{f(1-\alpha(1-\gamma))} = CR(\alpha,\gamma). \tag{48}$$

In the isotropic case ($\gamma=1$) the ratio $R(\alpha,1)$ is

$$\begin{aligned} R(\alpha,1) = [(1-\alpha)^{1/2}+\alpha^{1/2}]^2 &= 2, \quad \alpha=\tfrac{1}{2}(m_a=m_b) \\ &= 1, \quad \alpha=1(m_a \gg m_b). \end{aligned} \tag{49}$$

In the extreme anisotropic limit and with very large valence-band mass ($m_a \gg m_b$) we have

$$R(1,0) = \pi^2/16 \approx 0.62. \tag{50}$$

Quite generally we have concluded from the detailed form of (47) and the numerical calculation[12] of $f(\delta)$ that the ratio $R(\alpha,\gamma)$ has no large spread in the interval $0\leq\alpha$, $\gamma\leq 1$, and is numerically of the order 1.

(c) In the semimetallic case we consider the band energies

$$\begin{aligned} \epsilon_a &= (1/2m_a)(p_f^2-p^2), \\ \epsilon_b &= (1/2m_b)(p_x^2+p_y^2+\gamma p_z^2-\gamma^{1/3}p_f^2), \end{aligned} \tag{51}$$

the constants A and B chosen in such a way as to provide electrical neutrality. The number of holes or conduction electrons is

$$\rho = p_f^3/3\pi^2. \tag{52}$$

The critical limit G_{cr} of the excitonic phase once again is determined by the linearized gap equation (33) at $T=0$:

$$\Delta(\mathbf{p}) = \frac{1}{(2\pi)^3}\int d^3q\, V_{sc}(\mathbf{p}-\mathbf{q})\Delta(\mathbf{q}) \times\frac{\Theta(\epsilon_b)\Theta(-\epsilon_a)+\Theta(\epsilon_a)\Theta(-\epsilon_b)}{|\epsilon_b(\mathbf{q})-\epsilon_a(\mathbf{q})|}, \tag{53}$$

where the Θ functions [$\Theta(x)=1$ for $x>0$ and 0 for $x<0$] take care of the restricted integration, as discussed in Sec. III. As there are free carriers in the semimetallic region we use the screened Coulomb interaction

$$V_{sc}(\mathbf{q}) = \frac{4\pi e^2/K}{q^2+\kappa^2}, \quad \kappa^2 = \frac{e^2/K}{\pi}p_f(m_a+m_b), \tag{54}$$

due to both electrons and holes.

The solution of the integral equation (53) will in general have a strong angular dependence reflecting the anisotropy of the energies (51). For this reason we could not solve (53) for arbitrary γ between 0 and 1. However, a solution is possible in the nearly isotropic case, $1-\gamma \ll 1$.

We introduce polar coordinates in (53), and we integrate first the radial part. As this integral would diverge logarithmically at the Fermi surface in the isotropic case ($\gamma=1$), we expect a large contribution from the denominator in (53) when $1-\gamma\ll 1$. Doing the $|\mathbf{q}|$ integration by parts, we first integrate the denominator $|\epsilon_b-\epsilon_a|^{-1}$, thus separating a logarithmic term, $\ln(1-\gamma)^{-1}$. We may then pass to the limit $\gamma\to 1$ in the rest of the integral and neglect all contributions which are finite compared to $\ln(1-\gamma)^{-1}$ in this limit. This procedure leads to

$$\Delta(|\mathbf{p}|,\Omega_{\mathbf{p}})=\frac{1}{\pi}\frac{\mu}{M}\frac{\kappa^2}{p_f^2}\ln(1-\gamma)^{-1}$$
$$\times\int d\Omega_{\mathbf{q}}\,\Delta(p_f,\Omega_{\mathbf{q}})\left[\frac{\kappa^2}{p_f^2}+\left(\frac{\mathbf{p}}{p_f}-\frac{\mathbf{q}}{|\mathbf{q}|}\right)^2\right]^{-1}, \qquad (55)$$

where we have used (54) and set $M=m_a+m_b$; $\Omega_{\mathbf{p}}$ denotes the space angles. It is clear that we can solve (55) with a gap function independent of angles. Setting $|\mathbf{p}|$ equal to p_f, Δ drops out completely, and we have finally

$$\frac{1}{\ln(1-\gamma)^{-1}}=\frac{\mu}{M}\frac{\kappa^2}{p_f^2}\ln[1+4p_f^2/\kappa^2]. \qquad (56)$$

Using (40), (54), and $|G_{\mathrm{cr}}|=p_f^2/2\mu$, we can write Eq. (56) in terms of the critical overlap $|G_{\mathrm{cr}}|$ and the binding energy E_B:

$$\frac{\pi}{\ln(1-\gamma)^{-1}}=\left(\frac{E_B}{|G_{\mathrm{cr}}|}\right)^{1/2}\ln\left[1+\frac{4\pi\mu}{M}\left(\frac{|G_{\mathrm{cr}}|}{E_B}\right)^{1/2}\right]. \qquad (57)$$

Numerically, we have for $1-\gamma=\frac{1}{10}$, and equal band masses, $m_a=m_b$:

$$|G_{\mathrm{cr}}|\approx 1.14E_B. \qquad (58)$$

V. SUMMARY

In the preceding sections we have discussed the modifications in the description of the excitonic insulating phase in the more realistic case of anisotropic energy bands. We have seen that in the case of a positive band gap the results of earlier calculations[3,5,6] are essentially unaltered apart from some quantitative corrections. In the semimetallic region (band overlap), however, we have found a novel situation. For large enough anisotropy, the normal semimetallic state is stable against formation of excitonic pairs. Thus there exists a critical overlap beyond which the condensation will not occur. The possible existence of the excitonic phase is therefore limited to the finite region shown in the phase diagram (Fig. 3). The significant difference from the isotropic case[3,6] is the sharp cutoff on the semimetallic side.

In order to observe the phase experimentally one has therefore to consider materials which under normal conditions have a small band gap. Furthermore the exciton binding energy E_B which determines the transition temperature T_c should be large in order to have a significant value of T_c. Thus one has to look for materials with small dielectric constants and large effective masses. Jerome, Rice, and Kohn[3] have discussed the question of experimental realizability in detail. As possible candidates they suggest divalent metals and Group-V semimetals among which strontium and ytterbium seem to be the most promising though no clear experimental evidence for the occurrence of the phase has yet been reported.

ACKNOWLEDGMENT

The author is very grateful to Professor W. Kohn for many suggestions and helpful discussions, and especially for a critical reading of the manuscript.