On the phase diagram of an excitonic insulator in a strong magnetic field

A. A. Abrikosov

L. D. Landau Institute of Theoretical Physics, USSR Academy of Sciences

(Submitted March 18, 1973)
Zh. Eksp. Teor. Fiz. **65** 1508–1517 (October 1973)

It is shown that two phases should exist when an excitonic state is formed in a semimetal located in a strong magnetic field: (a) a "reduced-metal" phase in which, as a result of pairing, some of the carriers are removed from the free-carrier system at $T=0$, and (b) an "excitonic-insulator" phase in which there are no free carriers at $T=0$. These phases differ not only at $T=0$, but at finite temperatures as well, because the electron density in the reduced metal has a long-period superstructure along the magnetic field, the period becoming infinite upon transition to the excitonic-insulator phase. The qualitative shapes of the curves $T_c(E_g)$, on which $\Delta\to 0$, and $T_{c1}(E_g)$, which separates the two excitonic phases, are determined.

1. THE SPATIAL VARIATION OF THE ELECTRON DENSITY

In an earlier paper [1] we investigated the metal-excitonic insulator phase transition of a bismuth-type semimetal located in a strong magnetic field directed along the principal axis, and qualitatively considered the differences that arise when the magnetic field is oriented along other directions. Although a change in the direction of the magnetic field appreciably affects the quantitative characteristics of the new phase, the qualitative characteristics are more or less the same in all cases, and they can be followed on the example of the case when $\mathbf{H}\parallel C_3$. In [1] we found the specific heat, the conductivity, and the high-frequency permittivity of the new phase. These characteristics enable us to distinguish it sufficiently clearly from both the purely dielectric and the purely metallic phases.

An important characteristic feature accompanying the appearance of the excitonic phase is the change in symmetry, to wit, the appearance of a new spatial periodicity in the electron-density distribution and the electron potential.

Let us first of all find the change in the electron density as compared to the density in the metallic phase in the case considered in [1], i.e., when $\Delta \ll \epsilon_{Fe}, \epsilon_{Fh}$ (ϵ_{Fe} and ϵ_{Fh} are the electron and hole Fermi energies). The electron density is of the form

$$\rho(\mathbf{r})=\langle\Psi^{+}(\mathbf{r})\Psi(\mathbf{r})\rangle, \tag{1}$$

where (see [1])

$$\begin{aligned}\Psi(\mathbf{r})=\sum_k\{&[\overset{+}{\rho}_{3p_0+k}e^{-i(3p_0+k)z}+\overset{+}{\rho}_{-3p_0-k}e^{i(3p_0+k)z}]e^{i\mathbf{K}_0\mathbf{r}}\\&+[\psi_{p_0+k}e^{i(p_0+k)z}+\psi_{-p_0-k}e^{-i(p_0+k)z}]e^{i\mathbf{K}_1\mathbf{r}}+[\varphi_{p_0+k}e^{i(p_0+k)z}+\varphi_{-p_0-k}e^{-i(p_0+k)z}]e^{i\mathbf{K}_2\mathbf{r}}\\&+[\eta_{p_0+k}e^{i(p_0+k)z}+\eta_{-p_c-k}e^{-i(p_0+k)z}]e^{i\mathbf{K}_3\mathbf{r}}\}.\end{aligned} \tag{2}$$

Here ψ, φ, and η are the annihilation operators for the various electron clusters with centers at the points L on the twofold axes (the vectors $\mathbf{K}_1$, $\mathbf{K}_2$, and $\mathbf{K}_3$ in k space) and ρ^{+} is the creation operator for the holes (center at the point T whose vector is $\mathbf{K}_0$). It is assumed that only the vicinity of the Fermi limit (p_0, $-p_0$ for the electrons; $3p_0$, $-3p_0$ for the holes) will be important to us. Substituting (2) into (1), and taking only the "anomalous" terms giving the change in the period into account, we obtain

$$\begin{aligned}\delta\rho(\mathbf{r})=\sum_k\{&\langle\rho_{3p_0+k}\psi_{-p_0-k}\rangle e^{i(\mathbf{K}_1-\mathbf{K}_0)\mathbf{r}+2ip_0z}+\langle\rho_{-3p_0-k}\psi_{p_0+k}\rangle e^{i(\mathbf{K}_1-\mathbf{K}_0)\mathbf{r}-2ip_0z}\\&+\langle\overset{+}{\psi}_{p_0+k}\overset{+}{\rho}_{-3p_0-k}\rangle e^{-i(\mathbf{K}_1-\mathbf{K}_0)\mathbf{r}+2ip_0z}+\langle\overset{+}{\psi}_{-p_0-k}\overset{+}{\rho}_{3p_0+k}\rangle e^{-i(\mathbf{K}_1-\mathbf{K}_0)\mathbf{r}-2ip_0z}+\dots\end{aligned}$$

$$\begin{aligned}&+\langle\overset{+}{\psi}_{p_0+k}\varphi_{p_0+k}\rangle e^{i(\mathbf{K}_2-\mathbf{K}_1)\mathbf{r}}+\langle\overset{+}{\psi}_{-p_0-k}\varphi_{-p_0-k}\rangle e^{i(\mathbf{K}_2-\mathbf{K}_1)\mathbf{r}}\\&+\langle\overset{+}{\varphi}_{p_0+k}\psi_{p_0+k}\rangle e^{i(\mathbf{K}_1-\mathbf{K}_2)\mathbf{r}}+\langle\overset{+}{\varphi}_{-p_c-k}\psi_{-p_0-k}\rangle e^{i(\mathbf{K}_1-\mathbf{K}_2)\mathbf{r}}+\dots\}.\end{aligned} \tag{3}$$

We have not written out the corresponding terms for the other electron groups.

The averages occurring here are some definite Fourier components of the complete Green function $-i\langle T(\Psi(\mathbf{r},t)\Psi^{+}(\mathbf{r}',t'))\rangle$ for the electrons in the crystal. Let us use the time-reversal symmetry (see [2](:

$$G(\mathbf{r},t;\mathbf{r}',t')=G(\mathbf{r}',-t';\mathbf{r},-t)$$

or

$$G(\omega,\mathbf{r},\mathbf{r}')=G(\omega,\mathbf{r}',\mathbf{r}) \tag{4}$$

for the complete G-function. Hence we obtain

$$\langle\varphi_{+}{}^{+}\psi_{+}\rangle=\langle\varphi_{-}{}^{+}\psi_{-}\rangle,\quad\langle\psi_{+}\rho_{-}\rangle=\langle\rho_{+}{}^{+}\psi_{-}{}^{+}\rangle,$$

where the subscript + corresponds to the neighborhoods of p_0 and $3p_0$, and the − corresponds to those of $-p_0$ and $-3p_0$. Similar relations are valid for the other electron clusters. It follows, in particular, from this that Δ_1 and Δ_1' (see (19) in [1]) are connected by the relation: $\Delta_1=\Delta_1'^{*}$. We shall henceforth denote Δ_1 by Δ and consider it to be real.

Expressing our averages in terms of the functions G, F, and D (see [1]), and using the formulas (35) in [1], we obtain

$$\begin{aligned}\delta\rho=4i\frac{eH}{2\pi c}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\int_{-\infty}^{\infty}\frac{dk}{2\pi}\frac{\Delta}{(\omega+\xi_h)(\omega-\xi_e)-3\Delta^2}\\\times\cos 2p_0z[\cos(\mathbf{K}_1-\mathbf{K}_0)\mathbf{r}+\cos(\mathbf{K}_2-\mathbf{K}_0)\mathbf{r}+\cos(\mathbf{K}_3-\mathbf{K}_0)\mathbf{r}]\\-4i\frac{eH}{2\pi c}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\int_{-\infty}^{\infty}\frac{dk}{2\pi}\frac{\Delta^2}{(\omega+\xi_h)(\omega-\xi_e)-3\Delta^2}\frac{1}{\omega-\xi_e}\\\times[\cos(\mathbf{K}_1-\mathbf{K}_0)\mathbf{r}+\cos(\mathbf{K}_2-\mathbf{K}_0)\mathbf{r}+\cos(\mathbf{K}_3-\mathbf{K}_0)\mathbf{r}],\end{aligned}$$

where $\xi_e=v_ek$, $\xi_h=v_hk$, $v_e=p_0/m_e$, and $v_h=3p_0/m_h$; the poles are bypassed according to the following rule: $\omega\to\omega+i\delta\,\mathrm{sign}\,\omega$, $\delta>0$. The integration in the second term makes the term vanish, so that only the contribution from the first term remains. Recalling the definition of Δ in [1] (formula (19)), we obtain

$$\delta\rho=\frac{2eH}{\pi c}\frac{\Delta}{|\lambda_1|}\cos 2p_0z[\cos(\mathbf{K}_1-\mathbf{K}_0)\mathbf{r}+\cos(\mathbf{K}_2-\mathbf{K}_0)\mathbf{r}+\cos(\mathbf{K}_3-\mathbf{K}_0)\mathbf{r}]. \tag{5}$$

If we substitute into this $|\lambda_1|\sim e^2/\kappa$ (κ is the permittivity), and consider a sufficiently strong magnetic field, then $p_0\sim\sqrt{eH/c}$ and the vibration amplitude $\delta\rho$ turns out to be of the order of $(eH/c)^{3/2}\gamma e^{-\gamma}$, where $\gamma\sim\pi v\kappa/e^2$. In not too weak fields $(eH/c)^{3/2}$ determines the number of carriers (in this case $\gamma \gg 1$), so that $\gamma e^{-\gamma}$ is the relative amplitude of the variation of the electron density with respect to that of the carrier density.

The symmetry of $\delta\rho$ corresponds to the rhombohedral lattice with a doubled period and a superstructure in the form of a wave of long period π/p_0 superimposed on it along the z axis. In the typical alloy BiSb, in which this transition is observed [3], the period along the z axis should correspond to hundreds of interatomic distances.

2. THE DISAPPEARANCE OF THE SUPERSTRUCTURE (THE VICINITY OF T_c)

If we consider the transition to the excitonic phase from the insulator side, i.e., when $E_g > 0$ (E_g is the gap between the bands), then the question will be the pairing up of the individual carriers, and no Fermi degeneracy will be involved. Consequently, only the doubling of the rhombohedral period—without any superstructure along the z axis—should be observed in the new phase. Therefore, this superstructure should appear somewhere, i.e., another phase transition should exist. Since $p_0 \to 0$ as the carrier density decreases, it is natural to suppose that the period of the superstructure becomes infinite at the transition point.

It is difficult to obtain the $T_{c1}(E_g)$ curve of this transition in the general case, for when the condition $T, \Delta \ll \epsilon_{Fe}, \epsilon_{Fh}$ is not fulfilled there is no small parameter in the theory, and the effects connected with the interaction of the excitons become important (see [4,5]). This makes the problem practically insoluble.

Nevertheless, we shall use the self-consistent field method employed in [1]. This is justified on the following grounds. First of all, as we shall see below, if $m_e \ll m_h$, then it turns out that $T_{c1} \sim \epsilon_{Fh} \ll \epsilon_{Fe}$ at $\Delta = 0$, i.e., at the point of intersection of the curves $T_{c1}(E_g)$ and $T_c(E_g)$. This means that the electron gas is degenerate, and a small parameter remains in this case. This is not true for the beginning of the $T_{c1}(E_g)$ curve, since it turns out here that $\Delta \gg |E_g|$. In this case we shall employ a model approach that will help us understand the qualitative characteristics of the transition and make estimates. For quantitative computations this model approach will, of course, not do.

The terms with Δ_1 in the Hamiltonian (18) of [1] play the role of a potential that acts on the electrons and holes. To make the points at the electron and hole Fermi levels equivalent, this potential should contain Fourier components proportional to, say, $\exp[i(\mathbf{K}_1 - \mathbf{K}_0)\cdot\mathbf{r} + 2ip_0z)]$. Such an introduction of the self-consistent potential is possible in the neighborhood of the point of intersection of the curves $T_{c1}(E_g)$ and $T_c(E_g)$. We begin our investigation with the study of this point. For this purpose we assume that Δ_1 is proportional not to e^{2ip_0z}, but to e^{iKz}, where K is an unknown quantity that can be determined from the conditions of equilibrium (or, what is the same, from max $\Delta(T)$ or max T_c). The point is that the equation for Δ can be linearized in the vicinity of $\Delta \to 0$ and the individual harmonics do not interfere.

According to [1], the equation for the critical temperature is ($\lambda = \lambda_1$)

$$1 = \frac{\lambda}{2}\int_{-\infty}^{\infty}\frac{dp}{2\pi}\,\frac{\operatorname{th}[\xi_e(p)/2T] + \operatorname{th}[\xi_h(p+K)/2T]}{\xi_e(p) + \xi_h(p+K)}$$

where

$$\xi_e(p) = \frac{p^2}{2m_e} + E_g - \mu = \frac{p^2}{2m_e} - (|E_g| - |\mu|),$$

$$\xi_h(p) = \frac{p^2}{2m_h} + \mu = \frac{p^2}{2m_h} - |\mu|;$$

it is assumed that $E_g < 0$, $\mu < 0$, and $m_e \ll m_h$. Going over to integration over $p > 0$, we have

$$1 = \frac{\lambda}{2}\int_0^{\infty}\frac{dp}{2\pi}\left[\frac{\operatorname{th}(\xi_e/2T) + \operatorname{th}(\xi_{h1}/2T)}{\xi_e + \xi_{h1}} + \frac{\operatorname{th}(\xi_e/2T) + \operatorname{th}(\xi_{h2}/2T)}{\xi_e + \xi_{h2}}\right], \qquad (6)$$

where $\xi_{h1} = \xi_h(p+K)$ and $\xi_{h2} = \xi_h(p-K)$.

The chemical potential μ is determined from the condition of equality of the numbers of electrons and holes. We assume that $|\mu| \sim T \ll |E_g|$. Then the electrons will be degenerate, and we obtain

$$3\frac{eH}{2\pi c}\cdot 2\sqrt{2m_e|E_g|} = \frac{eH}{2\pi c}\int_{-\infty}^{\infty}\frac{dp/2\pi}{\exp[(p^2/2m_h - |\mu|)/T] + 1}.$$

Setting

$$\mu\Big/\left(\frac{m_e}{m_h}|E_g|\right) = \eta, \qquad T\Big/\left(\frac{m_e}{m_h}|E_g|\right) = \tau,$$

and introducing a new integration variable $x = p/\sqrt{2m_hT}$, we have

$$\frac{3}{\sqrt{\tau}} = \int_0^{\infty}\frac{dx}{e^{x^2 - \eta/\tau} + 1}. \qquad (7)$$

If $\eta \gg \tau$, the holes would be degenerate, and we should be back at the previous situation. We shall see later that in fact $\eta \sim 0.1\tau$. This enables us to expand the integral in the formula (7) in powers of η/τ. We then obtain

$$3/\sqrt{\tau} = 0.537 + 0.338\eta/\tau + 0.053(\eta/\tau)^2 + \ldots \qquad (7')$$

If the interaction is a δ-function interaction, then the upper limit of the integration in (6) is equal to infinity. It is easy to see that the integrand decreases when $p > \sqrt{2m_e|\mu|}$, and the integral converges. In order to avoid too complicated computations, we consider the interaction to be smeared out and introduce accordingly the upper integration limit according to the condition $|\xi_e| < L$, where $L < |E_g|$. This will allow us to replace ξ_e by $v_e(p - p_{0e})$ in the integral and set $p = p_{0e}$ in ξ_{h1} and ξ_{h2}. Let us add and subtract the same integral, but with $\xi_{h1} = \xi_{h2} = 0$. In the difference we can set the limit equal to infinity. In consequence, we obtain

$$\frac{\pi v_e}{\lambda} = \int_0^L \frac{\operatorname{th}(\xi/2T)}{\xi}\,d\xi + \frac{1}{2}\xi_{h1}\int_0^{\infty}\frac{\xi_{h1}\operatorname{th}(\xi/2T) - \xi\operatorname{th}(\xi_{h1}/2T)}{\xi(\xi^2 - \xi_{h1}^2)}\,d\xi$$
$$+ \frac{1}{2}\xi_{h2}\int_0^{\infty}\frac{\xi_{h2}\operatorname{th}(\xi/2T) - \xi\operatorname{th}(\xi_{h2}/2T)}{\xi(\xi^2 - \xi_{h2}^2)}\,d\xi.$$

For the integrals with ξ_{h1} and ξ_{h2} we can find expressions in the form of series:

$$\frac{\xi_{h1}}{2}\int_0^{\infty}\frac{\xi_{h1}\operatorname{th}(\xi/2T) - \xi\operatorname{th}(\xi_{h1}/2T)}{\xi(\xi^2 - \xi_{h1}^2)}\,d\xi = -\xi^{2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)}\,\frac{1}{\pi^2T^2(2n+1)^2 + \xi_{h1}^2}$$

We shall subsequently require quite large values of the ratio $(\pi T)^2/(\xi_h)^2 = (\pi\tau)^2/(\eta - 1)^2$. In view of this, we set ξ_{h1} (or ξ_{h2}) equal to zero in all the terms of the sum over n except the term with $n = 0$. We then obtain

$$\frac{\pi v_e}{\lambda} = \ln\frac{2\gamma L}{\pi T} - \frac{\xi_{h1}^2}{\xi_{h1}^2 + (\pi T)^2} - \frac{\xi_{h2}^2}{\xi_{h2}^2 + (\pi T)^2} - [{}^7/_8\zeta(3) - 1](\xi_{h1}^2 + \xi_{h2}^2)/(\pi T)^2. \qquad (8)$$

The coefficient in the last term is equal to 0.051, and since the significant values are $\pi T \gg \xi_h$, this term can be discarded.

If $K = 0$, then $\xi_{h1} = \xi_{h2} = \xi_h = (p_{0e}^2/2m_h) - |\mu|$. Using the previously introduced notation, we obtain

$$\tau \exp\{2(\eta-1)^2/[(\eta-1)^2+\pi^2\tau^2]\}=R\equiv\frac{4\gamma L m_h}{\pi p_{0e}^2}e^{-\pi v_e/\lambda}. \tag{9}$$

In fact, the upper limit is $L \sim p_{0e}^2/m_e$. Since $\tau\sim\eta\sim 1$ in the region of interest to us, we have $R\sim 1$. In other words, $(m_h/m_e)e^{-\pi v_e/\lambda}\sim 1$, or taking account of the fact that $\lambda\sim e^2/\kappa$ and $v_e\sim\sqrt{|E_g|/m_e}$, we have

$$E_g=E_{g_3}\sim-\frac{m_e e^4}{\varkappa^2}\ln^2\frac{m_h}{m_e}. \tag{10}$$

The entire expounded theory is valid in the region where $\pi v_e/\lambda \gg 1$, i.e., according to (10), $\ln(m_h/m_e)\gg 1$. Formulas (9) and (7) determine the $T_c(E_g)$ curve in the region $K=0$. Let us now consider the region of small but finite K. Expanding (8) in powers of K up to terms of order K^2, we obtain the correction to the right-hand side:

$$4\frac{K^2}{p_{0e}^2}(\pi\tau)^2\frac{-(\pi\tau)^2(3-\eta)+(\eta-1)^2(5+\eta)}{[(\eta-1)^2+(\pi\tau)^2]^3};$$

if the coefficient of K^2 is negative, then T_c decreases as K^2 increases, and, consequently, $\max T_c$ occurs at $K=0$. If, however, this coefficient is positive, then T_c increases with K. Thus, in the region with $K=0$ the quantity $(\pi\tau)^2$ is greater than the critical value, which is equal to

$$(\pi\tau_0)^2=\frac{(\eta-1)^2(5+\eta)}{3-\eta}. \tag{11}$$

Of course, it is assumed in this case that $\eta<3$ in the region of interest to us. In the opposite case the coefficient of K^2 is always positive.

Let us make the assumption that $(\pi\tau_0)^2\gg 1$. Then it follows from (11) that $\eta\approx 3$, i.e., $3-\eta\approx 32/(\pi\tau_0)^2$. It follows from (7′) that $\tau\approx 31$ for $\eta\ll\tau$. Thus, $3-\eta\approx 0.003$, and the correction to (7′) from the term with η/τ is 6%. Consequently, the assumption is justified. But it follows from Eq. (9) that $e^{-\pi v_e/\lambda}\sim 10 m_e/m_h$, and since this quantity should be small, the formulas obtained are strictly applicable to the quite rare case $m_e\ll 0.1 m_h$. But our aim is not so much to obtain rigorous formulas as to qualitatively analyze the phase diagram.

Equations (7), (9), and (14) determine the point of intersection of the $T_c(E_g)$ curve, on which $\Delta=0$, with the $T_{c1}(E_g)$ curve, on which $K\to 0$. For small $|E_g|$ (and high T_c), the $T_c(E_g)$ curve corresponds to $K=0$. For high $|E_g|$, we should have $K\neq 0$, which corresponds to a positive coefficient for the term with K^2. To find the $T_c(E_g)$ curve and the equilibrium value of K we must find the term with K^4. Setting in (8) $\pi T\gg\xi_h$, we obtain the value: $-2(\pi\tau_0)^{-2}(K/p_{0e})^4$. The coefficient of K^4 is negative. Determining K^2 from the $\max T_c$ condition, we find (with allowance for the condition $(\pi\tau_0)\gg 1$):

$$\frac{K^2}{p_0^2}=\frac{32\pi^2(\tau_0^2-\tau^2)}{(\pi\tau_0)^4}, \tag{12}$$

and the correction itself to the right-hand side of Eq. (8) is equal to

$$\frac{2\cdot(32)^2\pi^4(\tau_0^2-\tau^2)^2}{(\pi\tau_0)^{10}}.$$

Thus, as we move along the $T_c(E_g)$ curve from the point of intersection of the curves T_c and T_{c1}, the quantity K increases in proportion to $[(T_0-T)/T_0]^{1/2}$. Although the growth of K leads to a correction in the formula for the critical temperature, this correction is of the order of $(T_0-T)^2$ and is small compared to the change in T_c given by Eq. (9). Thus, Eq. (9) determines the $T_c(E_g)$ curve not only for $K=0$, but also in the region $K\neq 0$ where $T_0-T\ll T_0$.

3. THE $T_{c1}(E_g)$ CURVE AS $T\to 0$

It is must more difficult to find the $T_{c1}(E_g)$ curve in the vicinity of $T=0$. First of all, the commonly employed self-consistent potential approximation is applicable only for sufficiently small Δ, when the logarithmic situation is maintained. As we shall see below, this does not occur. Consequently, there is not a small parameter in the theory, and we should allow for the exciton interaction exactly (see [4,5]), which entails great mathematical difficulties.

Secondly, since we cannot expand in powers of Δ (even in the self-consistent field approximation), the theory is nonlinear and we cannot, as before, assume that $\Delta(z)\propto e^{iKz}$. In fact, Δ is some complex periodic function. We could, in principle, attempt to find it in the self-consistent field approximation, but the result can all the same serve only for estimates. In view of this, we consider a simpler (though unrealistic) model, in which the interaction term in the Hamiltonian is

$$H_{int}=-\Delta\sum_p\rho_{-p-K}(\psi_p+\varphi_p+\eta_p)-\Delta\sum_p(\psi_p^++\varphi_p^++\eta_p^+)\rho^+_{-p-K}, \tag{13}$$

where

$$\Delta=\lambda\sum_p\langle\rho_{-p-K}\psi_p\rangle.$$

This model does not correspond to the introduction of a true periodic potential, for in the latter case there would also be terms of the type

$$-\Delta\sum_p\rho_{-p+K}(\psi_p+\varphi_p+\eta_p),$$

which would lead to great complications. Since, however, the results obtained from this model are quite plausible, we can hope that it correctly describes the phase diagram at least qualitatively.

In this model the Green functions have the same form as in [1] (formulas (22)), where $\xi_e=(p^2/2m_e)-(|E_g|-|\mu|)$, $\xi_h=[(p+K)^2/2m_h]-|\mu|$. The energy spectrum consists of three branches:

$$\omega_1=\xi_e,\quad \omega_{2,3}={}^1/_2(\xi_e-\xi_h)\pm[{}^1/_4(\xi_e+\xi_h)^2+3\Delta^2]^{1/2}. \tag{14}$$

The equation for the determination of Δ is of the form

$$1=\frac{\lambda}{2}\int_{-\infty}^{\infty}\frac{dp}{2\pi}\frac{\operatorname{th}(\omega_2/2T)-\operatorname{th}(\omega_3/2T)}{\omega_2-\omega_3}. \tag{15}$$

We assume, as before, that $m_e\ll m_h$, and be interested only in small K. In this case the ξ_e and ξ_h curves are disposed as shown in Fig. 1. It is not difficult to see that in this case $\omega_2>0$ for all momenta, while ω_3 can change sign. In view of this we obtain from (15) at $T=0$

$$1=\lambda\int_{\omega_3<0}\frac{dp}{2\pi}\frac{1}{\omega_2-\omega_3}. \tag{16}$$

According to (14), the limits of the integration region ($\omega_3=0$) are determined by the equation

$$-\xi_e\xi_h=3\Delta^2. \tag{17}$$

Let us first suppose that the needed values of Δ are of the order of p_{0e}^2/m_h. Then we find, correct to terms of order $\Delta\sqrt{m_e m_h}/p_{0e}^2\sim\sqrt{m_e/m_h}$, that $\omega_3>0$ in the entire region where $\xi_e>0$ and $\xi_h<0$, i.e., for $p_{0e}<p<(p_{0h}-K)$ and $-(p_{0h}+K)<p<-p_{0e}$. It follows with the same accuracy from the electroneutrality condition that $p_{0h}=3p_{0e}$. Let us introduce for simplification a lim-

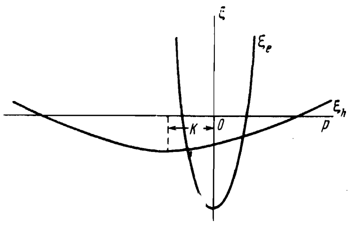

FIG. 1

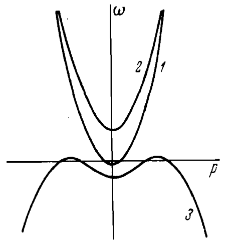

FIG. 2

itation on the integration domain over p in (16): $|\xi_e| < L$ (this does not change the qualitative result). We obtain from (16)

$$1=\frac{\lambda}{2\pi v_e}\int\limits_{-L}^{0} d\xi\{[(\xi+\xi_{h1})^2+12\Delta^2]^{-1/2}+[(\xi+\xi_{h2})^2+12\Delta^2]^{-1/2}\},$$

where ξ_{h1} and ξ_{h2} have the same meaning as in (6). Integrating and expanding in powers of K, we obtain

$$\frac{\pi v_e}{\lambda}=\ln\frac{Lm_e}{2p_{0e}^2}+\ln\frac{\sqrt{1+\zeta^2}-1}{\zeta^2}+\frac{1}{8}\frac{K^2}{p_{0e}^2}\left[(1+\zeta^2)^{-1/2}-\frac{1}{4}(1+\zeta^2)^{-3/2}\right],$$

where

$$\zeta=\frac{\sqrt{3}}{4}\Delta\Big/\frac{p_{0e}^2}{2m_h}=\frac{\sqrt{3}}{4}\Delta\Big/\frac{m_e}{m_h}|E_g|.$$

It follows from this that max Δ is necessarily attained at $K\neq 0$, and, consequently, the assumption that $K\to 0$ in the region of $|E_g|$ where $\Delta\sim p_{0e}^2/m_h$ is incorrect. We can establish in similar fashion, although in a more laborious manner, that $K\neq 0$ also in the region where $\Delta\sim p_{0e}^2/(m_e m_h)^{1/2}$.

Thus, we are obliged to assume that the equality $K=0$ is attained at large Δ of the order of $m_e e^4/\varkappa^2$. Let us now look at the question from a somewhat different standpoint. Let us consider Eq. (17) for $K=0$. Its solution is

$$p^2={}^1/_2(p_{0e}^2+p_{0h}^2)\pm[{}^1/_4(p_{0h}^2-p_{0e}^2)^2-12m_e m_h\Delta^2]^{1/2}.$$

Consequently, the region where $\omega_3>0$ exists until the expression under the radical sign can be positive, i.e., until

$$|m_h|\mu|-m_e|E_g||>2\sqrt{3m_e m_h}\Delta. \tag{18}$$

From the expression for the Green functions, (35) in [1], we can obtain the electroneutrality condition, which determines the chemical potential μ. It has the following form:

$$4p_{0e}=\int dp\,\theta(\omega_3) \tag{19}$$

(θ is a step function). The physical meaning of this condition consists in the following. It follows from [1] that the electron spectrum contains three branches having the schematic shapes shown in Fig. 2. The residue of the Green function corresponding to the branch 1 is equal to 2/3, while for the branches 2 and 3, the residue is equal to 1/3. Therefore, under the condition of particle conservation, the contribution from the branch 1 should double in comparison with the contribution from the branch 3. It follows from (19) that p_{0e} vanishes, i.e., $|\mu|\to|E_g|$, upon the disappearance of the region where $\omega_3>0$. It can be seen from Eq. (18) that in that case

$$|\mu|=|E_g|=2\sqrt{3m_e/m_h}\Delta\ll\Delta. \tag{20}$$

It is natural to suppose that the point $K=0$ corresponds to the tangency (with respect to ω) of the branches 1 and 3 in Fig. 2. Let us consider the vicinity of the point of contact from the standpoint that there is no intersection. In Eq. (16), the integration is over all momenta and, according to (20), we can assume $E_g=0$ in it. In consequence, we have

$$1=\frac{\lambda}{2\pi}\int\limits_0^\infty\left\{\left[\left(\frac{p^2}{2m_e}+\frac{(p+K)^2}{2m_h}\right)^2+12\Delta^2\right]^{-1/2}+\left[\left(\frac{p^2}{2m_e}+\frac{(p-K)^2}{2m_h}\right)^2+12\Delta^2\right]^{-1/2}\right\}dp$$

$$\approx\frac{\lambda}{\pi}\left[\int\limits_0^\infty\frac{dp}{[(p^2/2m_e)^2+12\Delta^2]^{1/2}}-\frac{K^2}{4m_e m_h}\int\limits_0^\infty\frac{p^2dp}{[(p^2/2m_e)^2+12\Delta^2]^{3/2}}\right]. \tag{21}$$

It follows from this that $\Delta\sim m_e\lambda^2$ and that max Δ is attained at $K=0$.

The region around the point of contact is somewhat more complicated when the tangency is considered from the standpoint that it is an intersection of the branches 1 and 3. We shall not present the detailed calculations. It turns out, as a result of the calculations, that the coefficient in the term with K^2 in the equation similar to (21) is positive, i.e., max Δ corresponds to $K\neq 0$. Thus, the assumption is confirmed.

Of course, it must once more be emphasized that the computation performed is not rigorous, because of the use of the nonphysical Hamiltonian and the use of the self-consistent field method. But the result obtained is physically the most probable, for otherwise we should have two phase transitions occurring at $T=0$, namely $K\to 0$ and the metal-insulator transition. Since, as already indicated above, the appearance of density oscillations along the z axis, i.e., of $K\neq 0$, itself entails a Fermi degeneracy, which, at $T=0$, will obtain as long as there is a band intersection, it is natural to suppose that the intersection and the superstructure disappear at the same point (at the same value of E_g).

Thus, it follows from the foregoing that the $T_{c1}(E_g)$ curve intersects the abscissa axis at the point where $\Delta\sim m_e\lambda^2\sim m_e e^4/\varkappa^2$ and

$$E_g=E_{g2}\sim-\sqrt{\frac{m_e}{m_h}}\Delta\sim-\sqrt{\frac{m_e}{m_h}\frac{m_e e^4}{\varkappa^2}}. \tag{22}$$

It should be noted that it makes sense to speak of a metal-insulator transition only for $T=0$, for at $T\neq 0$ there is always a finite number of carriers, and these phases are indistinguishable. The question under consideration here, however, is the disappearance of the superstructure along the z axis (the period becoming infinite), i.e., the problem of the crystal-symmetry change, which can be accurately determined at any temperature, and to which the entire $T_{c1}(E_g)$ curve corresponds.

The natural character of the results gives grounds for supposing that Eq. (16) can also be used to make qualitative predictions for low but finite temperatures. Let us therefore attempt to determine the shape of the curve $T_{c1}(E_g)$ near $T=0$. Expanding ω_3 about the maximum, we obtain

$$\omega_3\approx\omega_{3max}-a(p-p_0)^2/m_h,$$

where $a\sim 1$. By replacing $\tanh(\omega_3/2T)$ by −1, we incur an error of the order of $2n_F(|\omega_3|)$. In the region of interest to us, $\omega_{3max}\approx 0$. Thus, the temperature correc-

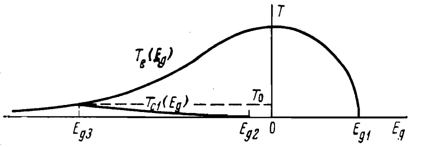

FIG. 3

tion will be proportional to $\sqrt{Tm_h}$. Consequently, we obtain

$$T_{c1}(E_g) \sim (m_e/m_h)\,(\Delta-\Delta_0)^2/\Delta_0.$$

If we allow for the fact that E_g is added to $p^2/2m_e$ in the integral (21), then expansion in powers of $|E_g| - |E_{g2}|$ yields $\Delta-\Delta_0 \propto |E_g| - |E_{g2}|$. Consequently, we have

$$T_{c1}(E_g) \sim \frac{(|E_g|-|E_{g2}|)^2}{m_h e^4/\varkappa^2} \tag{23}$$

(since $\Delta_0 \sim m_e e^4/\kappa^2$). It follows from this formula that $T_c \sim |E_g|\, m_e/m_h$ when $|E_g| \sim m_e e^4/\kappa_2 \gg |E_{g2}|$. This result can be matched to within logarithmic factors with the region where the $T_{c1}(E_g)$ curve intersects the $T_c(E_g)$ curve.

4. THE PHASE DIAGRAM

According to [4], the behavior of the $T_c(E_g)$ curve at $E_g > 0$ is essentially determined by the exciton interaction. If there is repulsion, then the curve $T_c(E_g)$ starts from $E_g = E_{g1} = \epsilon_0 \ln^2(\Omega/\epsilon_0)$ (here $\Omega = eH/m_e c$ and $\epsilon_0 = m_e e^4/2\kappa^2$), the binding energy of the isolated exciton; on the whole the phase diagram looks like the diagram in Fig. 3. If, however, the exciton interaction is an attractive one, then a liquid appears at the edge of the excitonic phase formation on the $E_g > 0$ side, and the phase diagram in this region has a more complicated shape.

Our analysis will also do qualitatively for substances with a different energy spectrum, and for other directions of the field. It follows from this analysis that with the exception of the very unlikely case when there are one group of electrons and one group of holes with the same density, there should necessarily exist in all cases two phase transitions in the band-overlap region ($E_g < 0$). The new phase, in which only a number of the free carriers are paired off, can be called a "reduced metal." The point of transition of this phase to the true excitonic insulator phase at $T = 0$ (i.e., E_{g2}) can be determined from the vanishing of the conductivity. As for the entire $T_{c1}(E_g)$ curve, it can be determined only from the disappearance of the long period along the z axis. The corresponding electron-density oscillations cannot exceed in amplitude the total carrier density, and therefore the relative intensity of the secondary x-ray peaks will be very low ($<10^{-6}$). The infrared radiation in the requisite wave band is strongly absorbed, and will therefore also not do.

Another possible method for detecting the superstructure is the high-frequency sound diffraction method. But this question requires further analysis.

In conclusion, we note that the problem considered by us here is close to the chromium phase diagram problem studied by Rice in [6] with the aid of a model band structure with isotropic electron and hole spectra. Rice obtains the dependence of the Néel point on the carrier-concentration disparity. As in our problem, two ordered phases are possible, in one of which a superstructure exists. Between these phases is an equilibrium curve corresponding to a phase transition with a change in symmetry. The transition can be of second, as well as of first, order. We may say apropos of this that according to our assumption, the T_{c1} curve corresponds to a second-order transition. Actually, the possibility of a first-order transition is not excluded. In that case the curve found by us would correspond to the stability limit of the reduced-metal phase, and the true transition curve would lie to the left of ours. The fact, however, that our curve ends at the metal-insulator transition point (at $T = 0$) points contrary to such a possibility.

The author is grateful to Yu. V. Kopaev, who drew his attention to Rice's paper and pointed out the possibility of a first-order transition

---

[1] A. A. Abrikosov, J. Low Temp. Phys. **10**, 3 (1973).
[2] A. A. Abrikosov, Zh. Eksp. Teor. Fiz. **43**, 1083 (1962) [Sov. Phys.-JETP **16**, 765 (1963)].
[3] H. B. Brandt and S. M. Chudinov, J. Low Temp. Phys. **8**, 339 (1972).
[4] S. A. Brazovskiĭ, Zh. Eksp. Teor. Fiz. **64**, 710 (1973) [Sov. Phys.-JETP **37**, 000 (1973)].
[5] L. V. Keldysh and A. N. Kozlov, Zh. Eksp. Teor. Fiz. **54**, 978 (1968) [Sov. Phys.-JETP **27**, 521 (1968)].
[6] T. M. Rice, Phys. Rev. **B2**, 3619 (1970).

Translated by A. K. Agyei
156