Some problems in the theory of semimetals

A. A. Abrikosov

L. D. Landau Theoretical Physics Institute of the USSR Academy of Sciences
(Submitted June 25, 1973)
Zh. Eksp. Teor. Fiz. **65**, 2063–2074 (November 1973)

Two problems in the theory of semimetals are considered: (1) the fulfillment of the Luttinger theorem, and (2) the origin of the rhombohedral deformation of the sublattices. It is shown that the main assumptions of the theory of Abrikosov and Fal'kovskiĭ[1] can be explained in a natural and consistent manner. In conclusion, some considerations are presented regarding the possibility of applying the ideas of semimetal theory to explain first-order metal-insulator phase transitions involving small distortion of the structure in transition-metal oxides.

1. INTRODUCTION

In the time that has passed since the publication of the theory of semimetals[1], calculations of different physical effects have been performed using the theory. The predictions of the theory have been subjected to comparison with experimental data. It has been possible to explain theoretically such qualitative effects as the high dielectric permittivity and the transparency in the infrared region[2], the possibility of formation of a gapless state in bismuth-antimony alloys and the elongation of the electron "ellipsoids"[3], the relation between the "spin" and orbital splittings of the levels in a magnetic field[4,5], etc. A comparison of the theory with the experimental characteristics of the energy spectrum of bismuth, obtained from cyclotron resonance, the de Haas–van Alphen effect and optical measurements, displays agreement to within 5%[6].

The theory of[1] permits the possibility in which the Fermi surface goes out a long way from the points T and L; in this case, the formulas found for the energy spectrum are inapplicable. This is the case for holes in As and Sb. The "large" Fermi surfaces arising in this case are found completely successfully by the pseudopotential method[7,8]. On the other hand, when applied to bismuth, the pseudopotential method gives results differing from the experimental values by a considerable factor[9]. Thus, our approach and the pseudopotential method complement each other when we are concerned with the calculation of the energy spectrum. But the theory of[1] gives, in addition, the possibility of displaying qualitative effects that can be overlooked in numerical calculations.

The theory of semimetals[1] was based on certain assumptions. Some of these, e.g., the importance of the points T and L in the Brillouin zone and the appearance of a spontaneous displacement of the sublattices, were proved. There are certain others, however, which continue to raise doubts, as can be seen, e.g., from the recent paper of Gordyunin and Gor'kov[10][1)].

The main objections of the paper[10] to our approach are the following:

1. Luttinger's theorem is not fulfilled.

2. The origin of the rhombohedral deformation remains unclear. This is an important point, since in the absence of such a deformation the substance will not be a semimetal, but an insulator.

The present article is devoted to an analysis of these questions.

2. LUTTINGER'S THEORY AND EXTRA LEVELS

According to Luttinger[12], the density of electrons is connected with the volume bounded by the Fermi surface (and by the Brillouin zone boundaries) by the same relation as for interacting electrons in a periodic field:

$$n=\frac{2}{(2\pi)^3}(pV_B+V_F), \tag{1}$$

where V_B is the volume of the Brillouin zone, p is an integer, and V_F is the volume inside the Fermi surface. As applied to the group-V semimetals, this statement has the following meaning. As is well known, their lattices are obtained from the simple cubic lattice by a small displacement of the two face-centered sublattices along a body diagonal and by approximately the same rhombohedral deformation of the sublattices.

Since the number of electrons per cell in the original cubic lattice was odd, the sum of the volumes inside all the Fermi surfaces should be equal to $V_F = (p_1 + \frac{1}{2})V_{B0}$, where p_1 is an integer. In the displacement of the sublattices, a new Brillouin zone is obtained, with volume $V_B = \frac{1}{2}V_{B0}$. Consequently, $V_F = (2p_1 + 1)V_B$, i.e., V_F is a multiple of V_B. This corresponds fully to the fact that there is an even number of electrons per cell in the new lattice. But we note that the proportionality coefficient between V_F and V_B is an odd number.

In our work it was assumed that the principal role is played by a doubly degenerate level at the special points on the C_3 axes of the original cubic lattice. If we take only this energy level into account and assume that the Fermi boundary is in its vicinity, then the scheme of the Fermi surface and its occupation near these points has the form depicted in Fig. 1. The two surfaces in the vicinity of the special points are exactly superposed on reflection in a plane perpendicular to the C_3 axis. In[1], the assumption was made that a "dielectric" phase exists in which, in the limit of zero deformation, the same superposition of the surfaces occurs not only near the special points (which, in the new lattice, become the points T and L), but also far away from them.

To make the situation clear, we have illustrated a two-dimensional model in Fig. 2. The large square corresponds to the original Brillouin zone, and the dashed square to the new zone. Figures 2a and 2b illustrate our assumption. For an infinitesimal deformation, the new zone arises and the Fermi surface is reduced to it. The empty spaces in Fig. 2a then compensate the occupied spaces in Fig. 2b and an insulator is obtained. It is not difficult to see, however, that the volume inside all the Fermi surfaces is in this case equal to $V_F = V_{B0}$

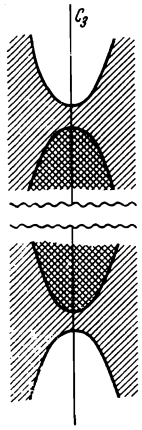

FIG. 1

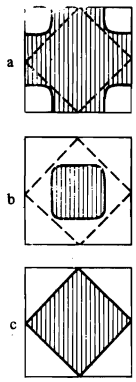

FIG. 2

$= 2V_B$, which does not correspond to Luttinger's theorem (the proportionality coefficient is even).

This circumstance was noted by Gordyunin and Gor'kov[10], and the following escape was proposed. According to Luttinger[12], V_F in formula (1) is the volume not only inside the Fermi surfaces, but in general inside all surfaces at which Re $G(\omega = 0, \mathbf{k})$ changes sign (G is the electron Green function). This can occur when Re $G(0, \mathbf{k})$ goes to infinity or to zero. The first case corresponds to the Fermi surface. However, in the opinion of the authors of[10], surfaces of the second type, which play no physical role but only rescue Luttinger's theorem, can also exist in a metal. It is difficult to agree with this opinion. In fact, Luttinger's theorem should be fulfilled not only for a normal metal, but also for a superconducting metal. In the latter case, if $\Delta \ll \epsilon_F$, the Green function near the Fermi surface will have the form

$$G=\frac{\omega+\xi(\mathbf{p})}{\omega^2-\xi^2(\mathbf{p})-\Delta^2(\mathbf{p})+i\delta}; \qquad (2)$$

it is clear that Re $G(0, \mathbf{p}) = 0$ at the Fermi surface. We cannot give a proof that the presence of surfaces at which Re $G(0, \mathbf{p}) = 0$ is necessarily connected with superconductivity, but this is highly probable.

We can, however, propose a more natural way out of this situation. We shall imagine that, in the original cubic lattice (in reality, we have in mind the "dielectric" phase in the limit of infinitesimal deformation) there is, in addition to the surfaces 2a and 2b, another surface, which coincides with the new Brillouin zone (Fig. 2c). In this case, when the new zone is formed, it will not give free carriers but, on the other hand, it will eliminate the difficulty with the Luttinger theorem.

At first sight, this assumption seems artificial. In fact, it corresponds completely to the philosophy of[1] and to calculations by the pseudo-potential method (cf., e.g.,[13]). As applied to semimetals, the corresponding analysis is carried through in[16], in which the origin of the spectrum of the semimetals is investigated. According to this work, in the construction of the Fermi surfaces in the free-electron approximation there arises on the C_3 axis a threefold degeneracy, which, being accidental, is lifted by the pseudopotential of the simple cubic lattice. Because of the usual weakness of the pseudopotential, the singlet level that is split off may be found at a comparatively small distance from the doubly degenerate level. The corresponding Fermi surfaces are at a small distance from the point at which the C_3 axis intersects the boundary of the new Brillouin zone arising when the sublattices are displaced (in the new lattice, these are the points T and L). With this construction, it is made clear that, of the five valence electrons, two correspond to a completely filled band (in a simple cube), and only three remain "active."

Arguing in the spirit of[1], we shall construct a new phase, which, in the limit of infinitesimal displacement of the sublattices, is an insulator with occupied bands corresponding to Figs. 2a, 2b and 2c. The free energy for such a phase is decreased on displacement of the sublattices, and has a minimum which may turn out to be lower than the energy of the "metallic" phase which concerned us above. However, as was noted in[1], the depth of the minimum is not great (it is proportional to the square of the displacement of the sublattices) and, therefore, the metallic and dielectric phases should not differ greatly from each other. In view of this, in calculating the spectrum in the new phase it is necessary, strictly speaking, to take into account not only the doublet level, but also the extra singlet level.

It is natural to suppose that the three active electrons correspond to the atomic p-electrons. The representation of the rotation group with $J = 1$ is split by a crystal field of symmetry C_{3v} into the symmetric one-dimensional representation A_1 (Fig. 2c corresponds to this) and the two-dimensional representation E (Figs. 2a and 2b).

Depending on the relative position of these levels and the Fermi boundary, there can be three cases: a) the case considered in[1], when the levels obtained from the singlet are positioned comparatively far from the Fermi level and therefore need not be considered; b) the case when, on the contrary, the doublet level is unimportant; in this case, it is not possible to obtain a spectrum of the bismuth type; c) all the levels are positioned close to the Fermi energy and should be taken into account. This case is more difficult to analyze and, therefore, we shall perform the corresponding analysis for the vicinity of the point T, which is simpler.

We shall reckon the energy from the level E. We denote the energy of the level A_1 (on the C_3 axis) by ν. Proceeding in complete analogy with[1], we take into account the spin-orbit coupling and construct a Hamiltonian matrix of the sixth rank. Using the functions $\psi_1 = \psi_+\chi_+$, $\psi_2 = \psi_-\chi_+$, $\psi_3 = \Psi\chi_+$, $\psi_4 = \psi_+\chi_-$, $\psi_5 = \psi_-\chi_-$, and $\psi_6 = \Psi\chi_-$ (where ψ_+, ψ_- is the basis of E, and Ψ is the basis of A_1), it has the form

$$D^{(6)}=\begin{pmatrix} D^{(3)}(\Delta) & \\ & D^{(3)}(-\Delta)\end{pmatrix}, \quad D_3(\Delta)=\begin{pmatrix} p+\Delta & q_+ & r_- \\ q_- & p-\Delta & r_+ \\ r_+ & r_- & p_1+\nu\end{pmatrix}, \qquad (3)$$

where $p = a\kappa_z$, $q_\pm = b\kappa_\pm$, $r_\pm = b_1\kappa_\pm$, and $p_1 = a_1\kappa_z$.

On displacement of the sublattices, the representations corresponding to $\pm\mathbf{k}_0$ merge and a 12th-rank matrix is obtained. The corresponding off-diagonal elements are $\gamma = cu_1$ for ψ_1, ψ_2, ψ_4 and ψ_5, and $\gamma_1 = c_1u_1$ for ψ_3 and ψ_6 (u_1 is the displacement of the sublattices). By interchanging the rows and columns, one can obtain the result that the full 12th-rank matrix takes the form

$$D^{(12)}=\begin{pmatrix} D_1^{(6)}(\Delta) & \\ & D_1^{(6)}(-\Delta)\end{pmatrix} \qquad (4)$$

where the order of the functions is as follows: $\psi_1, \psi_1', \psi_2, \psi_2', \psi_3, \psi_3', \psi_4, \psi_4', \psi_5, \psi_5', \psi_6, \psi_6'$, where ψ_i' is obtained from ψ_i by inversion.

The energy levels depend on Δ^2, so that it is sufficient to consider the matrix $D_1^{(6)}(\Delta)$:

$$D_1^{(6)}(\Delta)=\begin{pmatrix} p+\Delta & \gamma & q_+ & 0 & r_- & 0 \\ \gamma & -p+\Delta & 0 & -q_+ & 0 & -r_- \\ q_- & 0 & p-\Delta & \gamma & r_+ & 0 \\ 0 & -q_- & \gamma & -p-\Delta & 0 & -r_+ \\ r_+ & 0 & r_- & 0 & p_1+\nu & \gamma_1 \\ 0 & -r_+ & 0 & -r & \gamma_1 & -p_1+\nu \end{pmatrix} \quad (5)$$

(compared with[1], we have shifted the origin by f and have included the deformation term in ν).

The Fermi energy for the holes is bismuth is small compared with the characteristic parameters of the spectrum, and this makes it possible to expand the energy in the moments, i.e., to use the effective-mass approximation. For $\kappa = 0$, we have the following eigenvalues and eigenfunctions:

$$\begin{aligned} &\varepsilon_1(0)=-\Delta-\gamma, && \Psi_1=(\varphi_3-\varphi_4)/\sqrt{2}, \\ &\varepsilon_2(0)=\Delta-\gamma, && \Psi_2=(\varphi_1-\varphi_2)/\sqrt{2}, \\ &\varepsilon_3(0)=-\Delta+\gamma, && \Psi_3=(\varphi_3+\varphi_4)/\sqrt{2}, \\ &\varepsilon_4(0)=\Delta+\gamma, && \Psi_4=(\varphi_1+\varphi_2)/\sqrt{2}, \\ &\varepsilon_5(0)=\nu-\gamma_1, && \Psi_5=(\varphi_5-\varphi_6)/\sqrt{2}, \\ &\varepsilon_6(0)=\nu+\gamma_1, && \Psi_6=(\varphi_5+\varphi_6)/\sqrt{2}, \end{aligned} \quad (6)$$

where φ_i are functions arranged in the order corresponding to $D_1^{(6)}$, i.e., e.g.: $\varphi_1 = \psi_1$, $\varphi_2 = \psi_1'$, ... Differentiating $D_1^{(6)}$ with respect to κ, we obtain the velocity matrix

$$\hat{v}=\begin{pmatrix} an_z & 0 & bn_+ & 0 & b_1n_- & 0 \\ 0 & -an_z & 0 & -bn_+ & 0 & -b_1n_- \\ bn_- & 0 & an_z & 0 & a_1n_+ & 0 \\ 0 & -bn_- & 0 & -an_z & 0 & -b_1n_+ \\ b_1n_+ & 0 & b_1n_- & 0 & a_1n_z & 0 \\ 0 & -b_1n_+ & 0 & -b_1n_- & 0 & -a_1n_z \end{pmatrix}. \quad (7)$$

Using this, we can write the energy spectrum for small κ in the form

$$\begin{aligned} &\varepsilon_i(\varkappa)=\varepsilon_i(0)+{}^1/_2M_{i\alpha\beta}^{-1}\varkappa_\alpha\varkappa_\beta, \\ &M_{i\alpha\beta}^{-1}=\sum_{k\neq i}\frac{v_{ik}^\alpha v_{ki}^\beta+v_{ik}^\beta v_{ki}^\alpha}{\varepsilon_i-\varepsilon_k}. \end{aligned} \quad (8)$$

From (6) and (7), we obtain

$$\begin{aligned} \varepsilon_1&=-\Delta-\gamma-\frac{p^2}{2\gamma}-\frac{q^2}{2(\gamma+\Delta)}-\frac{r^2}{\gamma_1+\gamma+\nu+\Delta}, \\ \varepsilon_2&=\Delta-\gamma-\frac{p^2}{2\gamma}-\frac{q^2}{2(\gamma-\Delta)}-\frac{r^2}{\gamma_1+\gamma+\nu-\Delta}, \\ \varepsilon_3&=\gamma-\Delta+\frac{p^2}{2\gamma}+\frac{q^2}{2(\gamma-\Delta)}+\frac{r^2}{\gamma_1+\gamma-\nu-\Delta}, \\ \varepsilon_4&=\gamma+\Delta+\frac{p^2}{2\gamma}+\frac{q^2}{2(\gamma+\Delta)}+\frac{r^2}{\gamma_1+\gamma-\nu+\Delta}, \\ \varepsilon_5&=\nu-\gamma_1-\frac{p_1^2}{2\gamma_1}-\frac{r^2}{\gamma_1+\gamma-\nu+\Delta}-\frac{r^2}{\gamma_1+\gamma-\nu-\Delta} \\ \varepsilon_6&=\nu+\gamma_1+\frac{p_1^2}{2\gamma_1}+\frac{r^2}{\gamma_1+\gamma+\nu+\Delta}+\frac{r^2}{\gamma_1+\gamma+\nu-\Delta} \end{aligned} \quad (9)$$

In order not to obtain a large Fermi surface, it is necessary in all cases that all the levels have either a maximum or a minimum at p, q, r = 0 (by γ and Δ, we mean the absolute values, or, more simply, we choose γ and Δ to be positive). It follows from the expression for ϵ_6 that $\gamma_1 + \gamma + \nu - \Delta > 0$, and from the expression for ϵ_5 we find $\gamma_1 + \gamma - \nu - \Delta > 0$. Thus, although γ_1 and γ are proportional to the small displacement of the sublattices, Δ is the small spin-orbit splitting, and the quantity ν is not formally small, it is nevertheless necessary that the condition

$$|\nu|<\gamma_1+\gamma-\Delta. \quad (10)$$

be fulfilled. In reality, γ_1 and γ are the largest of the parameters. In experiment, $\gamma \sim 0.3$ eV and, as we shall see from the following, γ_1 should be still larger. On the other hand, as already noted above, the splitting ν is due to the pseudo-potential for the cubic lattice and cannot be large ($\sim$ 1 to 0.1 eV).

The inequality (10) places no restriction on the relative positions of $\epsilon_2(0)$ and $\epsilon_5(0)$. Of these levels, the higher one corresponds to the holes observable in bismuth. In order to choose between these two possibilities, we shall examine the question of the relation between the orbital and spin splittings of the levels in a magnetic field.

The orbital splitting is expressed by the formula (cf., e.g.,[13])

$$\Delta\varepsilon_{\text{orb}}=\frac{eH}{m^*c}=\frac{2\pi eH}{c\partial S/\partial\varepsilon},$$

where S is the area of the classical trajectory of the electron, ϵ = const, and p_z = const. If, in the quadratic approximation given by formula (8), the second term is written using the principal axes and the magnetic field is directed along the z-axis, then $m^* = \sqrt{M_xM_y}$. It follows from formula (9) that if the field is directed along the principal axis C_3, then for the bands ϵ_2 and ϵ_5 we have

$$\begin{aligned} \Delta\varepsilon_{2\,\text{orb}}&=\frac{eH}{c}\left(\frac{b^2}{\gamma-\Delta}+\frac{2b_1^2}{\gamma_1+\gamma+\nu-\Delta}\right), \\ \Delta\varepsilon_{5\,\text{orb}}&=\frac{eH}{c}\left(\frac{4b_1^2(\gamma_1+\gamma-\nu)}{(\gamma_1+\gamma-\nu)^2-\Delta^2}\right). \end{aligned} \quad (11)$$

If the field lies perpendicular to C_3, we obtain

$$\begin{aligned} \Delta\varepsilon_{2\,\text{orb}}&=\frac{eH}{c}\left[\frac{a^2}{\gamma}\left(\frac{b^2}{\gamma-\Delta}+\frac{2b_1^2}{\gamma_1+\gamma+\nu-\Delta}\right)\right]^{1/2}, \\ \Delta\varepsilon_{5\,\text{orb}}&=\frac{eH}{c}\left[\frac{a_1^2}{\gamma_1}\left(\frac{4b_1^2(\gamma_1+\gamma-\nu)}{(\gamma_1+\gamma-\nu)^2-\Delta^2}\right)\right]^{1/2}. \end{aligned} \quad (12)$$

According to the effective-mass approximation, the spin splitting is equal to

$$\Delta\varepsilon_{\text{spin}}=i\frac{e}{c}\mathbf{H}\sum_{k\neq i}\frac{[\mathbf{v}_{ik}\times\mathbf{v}_{ki}]}{\varepsilon_i-\varepsilon_k}. \quad (13)$$

Suppose the field lies along the principal axis C_3. We then obtain

$$\begin{aligned} \Delta\varepsilon_{2\,\text{spin}}&=\frac{eH}{c}\left(\frac{b^2}{\gamma-\Delta}-\frac{2b_1^2}{\gamma_1+\gamma+\nu-\Delta}\right), \\ \Delta\varepsilon_{5\,\text{spin}}&=-\frac{eH}{c}\left(\frac{4b_1^2\Delta}{(\gamma_1+\gamma-\nu)^2-\Delta^2}\right). \end{aligned} \quad (14)$$

For $\mathbf{H} \perp C_3$, we obtain

$$\Delta\varepsilon_{2\,\text{spin}}\approx 0, \quad \Delta\varepsilon_{5\,\text{spin}}\approx 0 \quad (15)$$

(in the present case, we are not considering small splittings of the order of $\mu_B H$, where μ_B is the Bohr magneton).

It follows from the experimental data of[14] that the spin splitting for the holes becomes small or a multiple of the orbital splitting for $\mathbf{H} \perp C_3$ and for $\mathbf{H} \parallel C_3$[2]. This agrees with formula (15) for both bands. However, it can be seen from formulas (12) and (14) that the band ϵ_5 does not have the necessary property, since for $\mathbf{H} \parallel C_3$ the spin and orbital splittings are completely different. On the other hand, if we assume that the holes originate from the band ϵ_2, and compare formulas (2) and (14), then the spin splitting will be equal to the orbital splitting if

$$\frac{b^2}{\gamma-\Delta}\gg\frac{2b_1^2}{\gamma_1+\gamma+\nu-\Delta}. \quad (16)$$

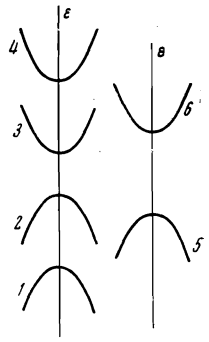

FIG. 3

Figure 3 schematically depicts the levels (9). The condition (10) means that the level ϵ_3 should be above ϵ_5, and the level ϵ_6 above ϵ_2. The additional requirement (16) means that the splitting between the levels ϵ_5 and ϵ_6, which is equal to $2\gamma_1$, should be sufficiently great, viz., $\gamma_1 \gg \gamma$ (it follows from the data of [14] that $\gamma_1 \sim 5\gamma$). But in this case the levels ϵ_5 and ϵ_6 lie far from the Fermi level and therefore have practically no substantial influence on the properties of bismuth (except on the optical properties at high frequencies). As regards the electron minimum at the point L, the position is evidently analogous. The point is that the coefficients of the spectrum in the regions of the points T and L are related to each other ($\gamma_L = -\gamma_T/3$, $\gamma_{1L} = -\gamma_{1T}/3$) and, therefore, a large splitting of the levels ϵ_5 and ϵ_6 at the point T means that there is also a large splitting at the point L. At the same time, the levels ϵ_2 and ϵ_3 are very close at the point L, and, most likely, are even "inverted" (cf. [3]).

Thus, the introduction of the extra level is justified from all points of view: a) its presence follows from pseudo-potential calculations, b) it leads to agreement with Luttinger's theorem, and c) it has no essential influence on the properties of bismuth and thus does not spoil the agreement of the theory with experiment.

3. THE RHOMBOHEDRAL DEFORMATION OF THE SUBLATTICES

First of all, we note that a certain inaccuracy was allowed to pass in [1]. As was correctly pointed out in [10], the self-consistent potential associated with the displacement u_1 of the sublattices cannot have "diagonal" matrix elements between the functions ψ_i, but only between the functions ψ_i and ψ_i'. This means that the coefficients f and β cannot depend on u_1. However, contrary to the statement in [10], in this case β does not vanish, since, as shown by Fal'kovskiĭ [15], the tensor u_{ik}, i.e., the rhombohedral deformation of the sublattices, makes a contribution to this coefficient. Thus, there is no need to invoke the superlattice in order to save this coefficient, as was proposed in the second variant of [10]. In experiment, the rhombohedral deformation of the sublattices in the lattices of semimetals is of the same order as the relative displacement of the sublattices. Moreover, since the coefficients f and β depend only on $w = (u_{xy} + u_{yz} + u_{zx})/3$, $w = 0$ would mean $f = \beta = 0$. But in this case, as remarked in [10] (cf. also [16]), the spectrum of an insulator, and not of a semimetal, arises.

We have shown in [1] how the energy minimum with respect to the displacement of the sublattices arises in the dielectric phase. The formula

$$\gamma_0 = Pe^{-(2\sigma+\chi)/2\chi}, \tag{17}$$

was obtained, where P is a large energy of the order of 10 eV, the coefficient σ is determined by the strong Coulomb interaction of the ions, and the coefficient χ by the much weaker electron-ion interaction. Even if $\chi \sim \frac{1}{3}\sigma$, then $\gamma_0 < 10^{-2}P$. The same result is obtained in [10] from a self-consistency condition. It should be remarked that, in obtaining the expression for the free energy, the principal logarithmic integral arises in the integration over momenta far from the points T and L, i.e., in the region where our formulas are inapplicable. The reasons why the result obtained is not too dependent on the model chosen were given in [1].

However, this same factor (the integration over the large momenta) does not give the possibility of a similarly definite answer to the question of the appearance of the rhombohedral deformation of the sublattices. The calculation of the free energy in [1] is rigorous under the assumption that the spectrum is expressed by the "hole" formula in the vicinity of both T and L, even when we are not very close to these points. In this case, a dependence on w does not arise in the free energy. An improvement of this calculation in the sense of taking into account the coefficients β and δ in the vicinity of the points L does not alter the position. Allowance for the third level (cf. the preceding section) also gives nothing[3)].

A dependence on w starts to appear only at sufficiently large values of w, when the valence band at the point T and the conduction band at the point L overlap strongly, so that the Fermi level corresponds to open isoenergetic surfaces in both bands (cf. [1,17]). This is effectively the critical value of w referred to in [10] (second variant). But this answer is not the solution of the problem, since in bismuth there is no strong intersection, whereas, at the same time, there is a rhombohedral deformation of the sublattices. We cannot yet give a unique explanation of this phenomenon. Different variants are possible.

a) It can be assumed that the fact that the electron-ion energy is independent of w is only a property of a particular model, since there are no reasons at all for our formulas for the spectrum to be valid far from the points T and L.

We might expect that, under very general assumptions about the spectrum, the electron-ion energy will have the form

$$-cu_1^2 \ln \frac{P/a}{u_1 f(w/u_1)},$$

where u_1 is the displacement of the sublattices and w is the rhombohedral deformation. Adding to this the elastic energy of the ions, we have

$$F = gu_1^2 + bw^2 - cu_1^2 \ln \frac{P/a}{u_1 f(w/u_1)}. \tag{18}$$

Minimizing the energy with respect to u_1 and w, we obtain the old result for u_1, and $w \propto u_1$.

As already stated, we have not succeeded in finding a model that leads to formula (18). It is not ruled out that it may be a completely general fact that the electron energy is independent of w for small w. The following reasoning argues for this. The logarithmic integral in (18) is obtained by integration over the momenta far from the points T and L. But in this case the treatment performed in [10], in which the authors assumed a spectrum of a general form and introduced a deformation potential, should apply. Even under such general as-

sumptions, the energy was found to be independent of w for small w.

b) Another explanation of the rhombohedral deformation of the sublattices seems to us more probable. If we regard the displacement u_1 of the sublattices as the main deformation, then, as a consequence of this, there will necessarily be a rhombohedral deformation $u_{\alpha\beta}$ of second order in u_1, since the cubic group permits an invariant of the type $u_{1\alpha}u_{1\beta}u_{\alpha\beta}$. Introducing $w = \frac{1}{3}(u_{xy} + u_{yz} + u_{zx})$ (in the cubic axes) and adding the elastic energy, we obtain terms in the free energy that depend on w:

$$F(w) = bw^2 + du_1^2 w.$$

By taking the minimum with respect to w, we have

$$w = -du_1^2/2b. \tag{19}$$

Of course, at first sight, this formula contradicts the experimental data, since w and the ratio of u_1 to the lattice constant are of the same order of magnitude experimentally and amount to 0.05—0.1. However, as we now show, there are good grounds for assuming that $b \ll d$ and, as a result, the second-order quantity simulates, as it were, a first-order quantity.

We shall find the coefficient b for a model of ions interacting in accordance with Coulomb's law. According to [18], for a simple cubic lattice with lattice constant a_0 we have

$$\lambda_{xyxy} = \frac{Z^2e^2}{\pi a_0^4} \lim_{Q\to\infty}\left[\frac{2\pi}{15}Q - \sum_{\substack{q=(n_1,n_2,l)\\ 0<|q|<Q}}\left(\frac{q_x^4}{q^6} - \frac{1}{6q^2}\right)\right] \approx -0.6\frac{Z^2e^2}{\pi a_0^4}.$$

The deformation energy is equal to

$$4\lambda_{xyxy}(u_{xy}^2 + u_{xz}^2 + u_{yz}^2) \approx -\frac{7}{\pi}\frac{Z^2e^2}{a_0^4}w^2. \tag{20}$$

The negative modulus is not especially surprising, since it is known that a cubic lattice with central pair interaction is unstable (cf. [19]). On the other hand, a calculation of the electron energy in the model of [1] with the coefficients β and δ taken into account, although it does not give an expression of the type (18), does give a positive term proportional to β^2 (i.e., to w^2). Hence it can be seen that the shear modulus b is determined both by the lattice and by the electrons, and it is best of all to make use of the experimental value [20].

The elastic constants of bismuth depend little on the temperature in the range from 4 to 300° K. The constants we need, in the principal axes ($z \parallel C_3$, $x \parallel C_2$), are equal to $c_{11} = 63.5$, $c_{33} = 38.1$, and $c_{13} = 24.5$ (in units of 10^{10} erg/cm^3). The expressions for $u_{\alpha\beta}$ in the principal axes in terms of the corresponding expressions $u^c_{\alpha\beta}$ in the cubic axes have the form $u_{\alpha\beta} = 0$ for $\alpha \neq \beta$,

$$u_{xx} = \tfrac{1}{2}(u_{yy}^c + u_{zz}^c - u_{yz}^c),$$
$$u_{yy} = \tfrac{1}{3}(2u_{xx}^c + \tfrac{1}{2}u_{yy}^c + \tfrac{1}{2}u_{zz}^c - u_{xy}^c + \tfrac{1}{2}u_{yz}^c - u_{zx}^c),$$
$$u_{zz} = \tfrac{1}{3}(u_{xx}^c + u_{yy}^c + u_{zz}^c + u_{xy}^c + u_{yz}^c + u_{zx}^c).$$

Putting $u^c_{xx} = u^c_{yy} = u^c_{zz} = v$, and $u^c_{xy} = u^c_{yz} = u^c_{zx} = w$, we obtain the elastic energy

$$F_{el} = (c_{11} + c_{33}/2 + 2c_{13})v^2 + (-c_{11} + c_{33} + c_{13})vw + (c_{11}/4 + c_{33}/2 - c_{13})w^2.$$

Substituting the values given above, we find

$$F_{el} = (131{,}5v^2 - 0{,}9vw + 10.4w^2)\cdot 10^{10} \text{ erg/cm}^3. \tag{21}$$

According to our philosophy, the term proportional to vw should be absent, and it is in fact practically equal to zero. A very interesting fact is the comparatively small value of the shear modulus (the coefficient of w^2) compared with the bulk modulus.

In order to determine the terms in F_{el} proportional to u_1^2 and $u_1^2 w$, it would be necessary to measure the frequency of the optical mode and its variation with the rhombohedral deformation. Such data are not known to us. But since the Coulomb energy of the ions in the present case gives a large positive constant multiplying u_1^2, it may be thought that the Coulomb contributions are the main part of the corresponding coefficients, or at least give the correct order of magnitude.

The new coefficients are expressed in terms of a formally divergent sum. In view of this, we introduce screening:

$$1/R \to f(R) = e^{-\varkappa R}/R.$$

In accordance with the properties of bismuth (cf. [2]), $\kappa^{-1} \sim 100\, a_0$ (a_0 is the edge of the original cube). The second- and third-order terms in the Coulomb energy, expanded in the small displacements of the ions, have the form

$$\frac{N}{V}\frac{Z^2e^2}{4}\sum_{R,\alpha,\beta}\left[\delta_{\alpha\beta}\frac{f'}{R} + \frac{R_\alpha R_\beta}{R^2}\left(f'' - \frac{f'}{R}\right)\right] \tag{22}$$
$$\times u_\alpha u_\beta + \frac{N}{V}\frac{Z^2e^2}{12}\sum_{R,\alpha,\beta,\gamma}\left[\frac{3\delta_{\alpha\beta}R_\gamma}{R^2}\left(f'' - \frac{f'}{R}\right) + \frac{R_\alpha R_\beta R_\gamma}{R^3}\left(f''' - \frac{3f''}{R} + \frac{3f'}{R^3}\right)\right].$$

The sum is taken over all the lattice sites. The origin is positioned at the site 0. The vector u is the difference of the displacements: $\mathbf{u} = \mathbf{u}_R - \mathbf{u}_0$.

Since we are interested in the terms in the elastic energy that contain u_1^2, the summation should run only over the atoms belonging to the sublattice that does not contain the site 0. The density of such atoms is $\frac{1}{2}N/V$.

In the first term, we take into account only the displacement u_1 of the sublattices. In the second term, we seek an energy proportional to $u_1^2 w$. For this we put $u_\alpha = u_{1\alpha} + (\partial u_\alpha/\partial x_\beta)R_\beta$ and find the corresponding terms. It is not difficult to see that if we had taken $f = 1/R$ we would have obtained formally logarithmically divergent expressions in both terms of (22). On the other hand, averaging over the directions of R makes each term vanish. This difficulty is circumvented by introducing screening. In the calculation, sums arise which converge for $R \sim 1/\kappa$. Since $\kappa^{-1} \sim 100a$, this makes it possible to replace the sums by integrals. The result does not depend on κ. With the assumption that $u_{1x} = u_{1y} = u_{1z} = u_1/\sqrt{3}$ and $u_{xy} = u_{yz} = u_{zx} = w$, we obtain

$$\frac{\pi}{6}\frac{Z^2e^2}{a_0^6}u_1^2 - \frac{14\pi}{15}\frac{Z^2e^2}{a_0^6}u_1^2 w. \tag{23}$$

First of all, we call attention to the fact that the coefficient $d < 0$, and consequently $w > 0$. This means that the rhombohedron is elongated, which is indeed the case in the semimetals. To find the coefficient d, we substitute $a_0 = 3.28$ Å, $u_1 = 4\sqrt{3} \times (0.25—0.237)a_0$ (cf. [16]). The value of Z is somewhat uncertain. Substitution of the coefficients d and b into formula (19) gives $w = 0.051Z^2$ and $\delta\theta = -4w/\sqrt{3} = -0.118Z^2 = -6°46'Z^2$. Even for $Z = 1$, this is greater than the true value for bismuth ($-2°54'$). Of course, the purely Coulomb value of the coefficient d cannot pretend to great accuracy. However, this estimate shows that the interpretation of the rhombohedral deformation as a secondary effect, quadratic with respect

to the displacement of the sublattices, does not contradict the experimental data.

4. CONCLUSION

Thus, the arguments and estimates given above make it possible to give an explanation of certain aspects of semimetal theory that are not completely obvious. As already remarked at the beginning of the article, the theory explains most of the experimental data well.

Certain questions arise only with regard to the behavior of the levels in a magnetic field. In our opinion, however, the existing experimental data admit various interpretations (see footnote[2)]) and should be refined. There are certain problems in connection with the closing of the gap at the point L by means of a magnetic field. A separate article, by S. D. Beneslavskiĭ and L. A. Fal'kovskiĭ, will be devoted to this.

It seems to us that, on the whole, there are no reasons to doubt the correctness of the proposed model of semimetals. At first sight, the assumption which lies at the basis of the model, viz., the complete "superposition" of two Fermi surfaces in the reduction to the new Brillouin zone, is somewhat exceptional. However, this is apparently not so[4)] and, in particular, it is not ruled out that metal-insulator phase transitions, together with the appearance of lattice distortion in transition-metal oxides (cf.[21]), can be explained by an analogous mechanism[5)].

In fact, we shall imagine a situation in which, in the metallic phase, there is a spectrum close to the "superposition" situation. We now construct "by hand" a new phase in which an exact superposition of the Fermi surfaces is obtained on infinitesimal displacement of the sublattices (on reduction of the spectrum to the new Brillouin zone). Of course, the energy of the new phase is higher than that of the old phase. However, in a finite displacement of the sublattices, there arises, as we have shown in[1], an energy correction of the type $-cu_1^2 \ln(P/au_1)$, this being in reality a consequence of the "superposition" and not specific to the actual situation in bismuth. The presence of this negative term leads to a minimum of the energy in the dielectric phase.

As a whole, the pattern of the thermodynamic potential may appear as in Fig. 3 of[1]. In this paper, we assumed that the minimum of Φ for the dielectric phase lies below the metallic minimum. However, in reality, the position of this minimum depends on the concrete conditions, e.g., on the pressure and temperature. Under certain conditions, it can be raised above the minimum for the metallic phase and a first-order transition then occurs.

One of the possible reasons for the raising of the minimum is the same mechanism as in superconductivity, viz., a redistribution of the electrons over the states. At a "critical temperature" of order γ, the minimum disappears completely, i.e., if there were no other phase, there would be a second-order phase transition here. The fact that it is first-order transitions that are observed in the vanadium oxides argues in favor of the existence of a metallic phase and justifies the philosophy of our approach[1].

In semimetals of the bismuth type, a temperature phase transition is not observed. This can be explained by the fact that the melting points of these semimetals (542° K for Bi and 903° K for Sb) lie below the corresponding transition points; it is, perhaps, a consequence of the instability of the simple cubic lattice.

In conclusion, the author takes the opportunity to express his gratitude to L. P. Gor'kov for numerous debates and to L. A. Fal'kovskiĭ for valuable discussions.

---

1) In this paper, an attempt was made to obtain the spectrum of bismuth by means of a microscopic approach. The authors take the same initial spectrum of the undeformed lattice as in [1], and, taking into account the interaction of the electrons with each other and with the phonons, treat the change of the spectrum in a manner analogous to that in the work of Keldysh and Kopaev [11], i.e., by applying the self-consistent field method. In fact, such an approach is completely equivalent to the work of [1], since the self-consistent field method was also applied in the latter. The only formal difference is that this field was determined from the minimum of the free energy in [1] and from the self-consistency condition in [10]. However, in the case of thermodynamic equilibrium, these procedures always give the same results.

2) It is stated in the paper [14] that for $\mathbf{H} \parallel C_3$ the spin splitting is twice the orbital splitting. In fact, the curves of the dependence of the levels of the angle (HC_3) can be drawn through the points given in [14] in such a way that, for $\mathbf{H} \parallel C_3$, the two splittings will be the same. The final solution of the problem can be achieved only with the aid of an experiment in a sufficiently strong field, which would make it possible to reach the zero level.

3) We shall not give the corresponding calculations here, as they are lengthy, but the energy does not depend on w in the result.

4) The idea that the "superposition" situation is not exceptional and can explain other structural transitions was put forward by L. P. Gor'kov (cf., partially, in [10]) and was formulated by him in the pairing language of the paper by Keldysh and Kopaev [11]. The model treated in [11] is a particular case of "superposition."

5) The limiting case of the "superposition" situation is the coincidence of the Fermi surface with the new Brillouin zone (cf. Fig. 2c). This is the analog of the Peierls instability for a one-dimensional odd metal. Such a case can evidently occur in the quasi-one-dimensional situation in which there are linear chains of closely-spaced metal atoms in the substance. In this case, we can expect the appearance of a Fermi surface with large planar parts. It is most likely that V_3Si is such an example (cf. [22]), and it is not ruled out that this is valid with regard to the vanadium oxides too. Naturally, all the general discussions on the "superposition" situation also apply to the given case.

[1] A. A. Abrikosov and L. A. Fal'kovskiĭ, Zh. Eksp. Teor. Fiz. **43**, 1089 (1962) [Sov. Phys.-JETP **16**, 769 (1963)].

[2] A. A. Abrikosov, Zh. Eksp. Teor. Fiz. **44**, 2041 (1963) [Sov. Phys.-JETP **17**, 1372 (1963)].

[3] A. A. Abrikosov, J. Low Temp. Phys. **8**, 315 (1972).

[4] L. A. Fal'kovskiĭ, Zh. Eksp. Teor. Fiz. **44**, 1935 (1963) [Sov. Phys.-JETP **17**, 1302 (1963)].

[5] L. A. Fal'kovskiĭ, Zh. Eksp. Teor. Fiz. **49**, 609 (1965) [Sov. Phys.-JETP **22**, 423 (1966)].

[6] L. A. Fal'kovskiĭ and G. S. Razina, Zh. Eksp. Teor. Fiz. **49**, 265 (1965) [Sov. Phys.-JETP **22**, 187 (1966)].

[7] L. M. Falicov and P. J. Lin, Phys. Rev. **141**, 562 (1966).

[8] P. J. Lin and L. M. Falicov, Phys. Rev. **142**, 441 (1966).

[9] S. Golin, Phys. Rev. **166**, 643 (1968).

[10] S. A. Gordyunin and L. P. Gor'kov, Zh. Eksp. Teor. Fiz. **63**, 1922 (1972) [Sov. Phys.-JETP **36**, 1017 (1973)]; J. Low Temp. Phys. **11**, 147 (1973).

[11] L. V. Keldysh and Yu. V. Kopaev, Fiz. Tverd. Tela **6**, 2791 (1964) [Sov. Phys.-Solid State **6**, 2219 (1965)].

[12] J. M. Luttinger, Phys. Rev. **119**, 1153 (1960).

[13] A. A. Abrikosov, Vvedenie v teoriyu normal'nykh metallov (Introduction to the Theory of Normal Metals),

"Nauka", M., 1972 (English Translation: Solid State Physics (Suppl. 12), Academic Press, N.Y., 1972).
[14] G. E. Smith, G. A. Baraff and J. M. Rowell, Phys. Rev. **135**, A1118 (1964).
[15] L. A. Fal'kovskiĭ, Zh. Eksp. Teor. Fiz. **53**, 2164 (1967) [Sov. Phys.-JETP **26**, 1222 (1968)].
[16] M. H. Cohen, L. M. Falicov and S. Golin, IBM J. Res. Develop. **8**, 215 (1964).
[17] A. A. Abrikosov, Zh. Eksp. Teor. Fiz. **44**, 1632 (1963) [Sov. Phys.-JETP **17**, 1099 (1963)].
[18] A. A. Abrikosov, Zh. Eksp. Teor. Fiz. **39**, 1797 (1960) [Sov. Phys.-JETP **12**, 1254 (1961)].
[19] R. D. Misra, Proc. Camb. Phil. Soc. **36**, 173 (1940); M. Born and M. Bradburn, Proc. Camb. Phil. Soc. **39**, 104 (1943); M. Bradburn, Proc. Camb. Phil. Soc. **39**, 113 (1943).
[20] Y. Eckstein, A. W. Lawson and D. H. Roneker, J. Appl. Phys. **31**, 1534 (1960).
[21] Proceedings of the International Conference on the Metal-Nonmetal Transition. Rev. Mod. Phys. **40**, 673–844 (1968).
[22] L. P. Gor'kov, ZhETF Pis. Red. **17**, 525 (1973) [JETP Lett. **17**, 379 (1973)].

Translated by P. J. Shepherd
214