Electron structure of bismuth. Theory and experiment

A. E. Dorofeev and L. A. Fal'kovskiĭ

L. D. Landau Institute of Theoretical Physics, USSR Academy of Sciences
(Submitted 12 June 1984)
Zh. Eksp. Teor. Fiz. **87**, 2202–2213 (December 1984)

The electron structure of the bands made up in semimetals by atomic p states is determined for bismuth from experimental data on the carriers. A comparison is made with the results of optic measurements and pseudopotential calculations. Comparison of IV-VI semiconductors with bismuth confirms the notion that they have a common cubic parent phase.

1. INTRODUCTION

A general approach that relates the electron and phonon properties of IV-VI semiconductors, as well as of group V semimetals, with the features of their crystal structure was developed in Refs. 1 and 2. The theory is based on an idea advanced in Ref. 3 that the space lattice of a group-V semimetal can be obtained from a simple cubic one by a small shift of alternating layers of atoms located in planes perpendicular to the trigonal axis (doubling of the period), followed by rhombohedral deformation of the two resultant sublattices. The role of the small perturbation is played in IV-VI semiconductors by the ionicity difference between the atoms A and B.

In Ref. 3 were considered the vicinities of those Brillouin-zone points where band extrema can arise after the doubling of the period, and only a term that is doubly degenerate at these points was taken into account. Considerations connected with band filling, however, point out the need of allowing for one more closely located term. In addition, in a number of cases, for example in arsenic, the band extrema are displaced from symmetric points.

To cope with these difficulties, a tight-binding method was used in Ref. 2. This method makes it possible to take into account in the simplest manner the translational symmetry of the cyrstals. In contrast to pseudopotential calculations, that require a large volume of computer calculations and are not accurate enough (the effective masses calculated in Ref. 4 for bismuth differ from the measured ones by a factor of three), the results were obtained in analytic form. They are expressed in terms of several constants that have the meaning of overlap integrals.

In the present paper we obtain, by comparison with the experimental data, the parameters of the theory for bismuth. It is found that those parameters that bismuth has in common with IV-VI semiconductors coincide. This coincidence should be regarded as proof of the validity of the premise that they have a common cubic parent phase. The shape of the Fermi surface in the vicinity of the points T and L is obtained and the agreement between the calculated and observed extremal sections and cyclotron masses is shown. The parameters obtained are used to calculate the electron spectrum for the principal directions in the Brillouin zone and to interpret the optical-transition lines observed in the interval 0.7–4.5 eV. Since the comparison with the experimental data was accurate to within several percent, we needed, in contrast to Ref. 2, expressions for the observable quantities that were accurate within the framework of the approximation employed. It was also necessary to introduce some changes.

2. EFFECTIVE HAMILTONIAN. SPECTRUM AT SYMMETRIC POINTS

According to Ref. 2 the Hamiltonian of a group-V semimetal is of the form

$$H=\begin{pmatrix} \hat{h}_s+\hat{h}_a & i\hat{u} \\ -i\hat{u} & \hat{h}_s-\hat{h}_a \end{pmatrix}$$

$$\hat{h}_s=\hat{\eta}_0+\hat{\eta}_1+\hat{\eta}_2+\hat{\varepsilon}_0+\hat{\Delta},\quad \hat{h}_a=\hat{\xi}+\hat{\varepsilon}_1+\hat{\varepsilon}_2+\hat{\eta}_3, \tag{1}$$

where all the quantities are 6×6 matrices in the coordinate (x,y,z) and spin indices). The matrices $\hat{\xi}$, $\mathbf{x}\eta$, $\hat{\varepsilon}$, and $\hat{u}$ are diagonal in spin (we do not write out the corresponding unit matrix), and take in terms of the coordinate indices the form

$$\xi_{xx}=\xi_0 c_x+\xi_1(c_y+c_z),\quad (\hat{\eta}_1+\hat{\eta}_2)_{xx}=\eta_1 c_y c_z+\eta_2 c_x(c_y+c_z),$$

$$(\hat{\eta}_0+\hat{\eta}_3)_{xy}=\eta_0 s_x s_y+2\eta_3 s_x s_y c_z,$$

$$(\hat{\varepsilon}_0+\hat{\varepsilon}_1+\hat{\varepsilon}_2)_{xy}=\varepsilon_0+\varepsilon_1(c_x+c_y)+\varepsilon_2 c_z,$$

$$u_{xx}=u_1 s_x+u_2(s_y+s_z),\quad u_{xy}=u_3(s_x+s_y), \tag{2}$$

where $s_x=\sin k_x a$, $c_x=\cos k_x a$, etc.; a is the period of the cubic parent phase. In (2), ξ_0, and ξ_1 are the overlap integrals with the atoms in the first coordinate sphere in the cube, while η_0, η_1, and η_2 are the overlap integrals in the second sphere. For a reason discussed in detail below, account is taken also of an overlap integral η_3 in the third sphere. The quantities u and ε describe respectively the doubling of the period and the rhombohedral deformation.

The spin-orbit interaction $\hat{\Delta}$ has the matrix elements

$$\Delta_{xy}=-i\Delta\sigma_z/3, \tag{3}$$

where σ_z is a Pauli matrix. The remaining nonzero elements are obtained from (2) and (3) by cyclic permutation of the indices.

In addition to (1) we shall use also another form of the Hamiltonian:

$$\mathscr{P}^{-1}H\mathscr{P}=\begin{pmatrix} \hat{h}_s+\hat{u} & \hat{h}_a \\ \hat{h}_a & \hat{h}_s-\hat{u} \end{pmatrix}, \tag{4}$$

which is obtained by the unitary transformation

$$\mathscr{P}=\begin{pmatrix} \hat{c} & \hat{c} \\ \hat{c}^+ & -\hat{c}^+ \end{pmatrix},$$

where $\hat{c}$ is a 6×6 matrix with diagonal elements $c=(1+i)/$

 0038-5646/84/121273-07$04.00

2. The representation (4) is convenient for the analysis of the spectrum in the vicinity of points T and L (the Brillouin zone is shown in Fig. 1), since at these points themselves $\hat{\xi} = \hat{\varepsilon}_1 = \hat{\varepsilon}_2 = \hat{\eta}_2 = 0$, and in their vicinities the off-diagonal blocks of the matrix (4) can be regarded as a perturbation

$$V = \mathbf{k}\hat{\mathbf{p}}, \tag{5}$$

proportional to a small deviation of the quasimomentum $\mathbf{k}$.

Expressions (2) are written in the rectangular basis $\mathbf{a}_i^0$; k_x, k_y and k_z are orthogonal coordinates in it. The transition to the real deformed lattice with basis $\mathbf{a}_i'$ is effected with the aid of the strain tensor ε_{xy} (Ref. 5, p. 381 of Russ. original):

$$\mathbf{a}_i' = (\delta_{ij} + \varepsilon_{ij})\,\mathbf{a}_j^0, \tag{6}$$

where $\varepsilon_{xx} = 0$ and $\varepsilon_{xy} = 0.02$ for bismuth, and the corresponding $a = 3.289$ Å. These values were obtained from the known period $a_t = 4.746$ Å and the angle $\alpha = 57°19'$ between the elementary lattice vectors with the aid of the formulas

$$a_t = |\mathbf{a}_1' + \mathbf{a}_2'| = \sqrt{2}(1+\varepsilon_{xy})\,a,$$
$$\cos\alpha = (\mathbf{a}_1' + \mathbf{a}_2')\,(\mathbf{a}_2' + \mathbf{a}_3')/|\mathbf{a}_1' + \mathbf{a}_2'|^2 = 0.5 + 2\varepsilon_{xy}. \tag{7}$$

Conversion to the deformed basis means the substitution $k_x \to k_x + \varepsilon_{xy}(k_y + k_z)$ in Eqs. (2) with cyclic permutation of the indices for the two other projections.

We consider the most interesting points Γ, T, and L of the Brillouin zone.

Point Γ—center of zone. Here $\hat{u} = 0$ and the eigenvalues of the matrix in the upper left corner of (1) give the terms of the odd states

$$\Gamma_{45}^- = b_1 - b_2 + \Delta/3,$$
$$\Gamma_6^-(1,2) = b_1 + b_2/2 - \Delta/6 \pm [(b_2+\Delta)^2 + 8b_2^2]^{1/2}/2, \tag{8}$$

where $b_1 = \xi_0 + 2\xi_1 + \eta_1 + 2\eta_2$, $b = \varepsilon_0 + 2\varepsilon_1 + \varepsilon_2$. The even states are obtained from this by reversing the signs of ξ_0, ξ_1, ε_1, and ε_2.

Point T—the cubic coordinates of this points are $(1,1,1)\pi/2a$. The odd levels T_{45}^-, $T_6^-(1,2)$ are given by the same formulas (8), in which b_1 and b_2 are replaced respectively by

$$c_1 = u_1 + 2u_2, \quad c_2 = \eta_0 + \varepsilon_0 + 2u_3. \tag{9}$$

It is known that a hole extremum is located at the point T. The effective masses $m_\parallel$ and $m_\perp$ longitudinal and transverse relative to the trigonal axis, defined in the usual manner:

$$\omega = \omega(T) - k_\parallel^2/2m_\parallel - k_\perp^2/2m_\perp, \tag{10}$$

were measured for this extremum together with the spin splitting factor $\gamma = |\omega_s/\omega_c|$, which is the ratio of the spin and cyclotron splittings ω_s and ω_c in a magnetic field for the case when the field is parallel to the trigonal axis. A contribution to the effective masses is made by second order of perturbation theory in V [Eq. (5), and also by the quadratic terms of the expansion of the trigonometric functions in $\hat{u}$ and $\hat{\eta}$ [see (2) and (4)]. For the level T_{45} we obtain

$$-1/2m_\parallel = \mu^2(T_{45}^- - T_{45}^+)^{-1} + \alpha - \beta, \quad -1/2m_\perp = A + B + \alpha_1 - \beta,$$
$$\gamma = [(A - B + m_0^{-1})^2 + C^2]^{1/2}\,|A + B + \alpha_1 - \beta|^{-1}, \tag{11}$$

where

$$A = \nu_1^2\{\tilde{c}^2[T_{45}^- - T_6^+(1)]^{-1} + \tilde{s}^2[T_{45}^- - T_6^+(2)]^{-1}\},$$
$$B = \nu^2\{\tilde{s}^2[T_{45}^- - T_6^+(1)]^{-1} + \tilde{c}^2[T_{45}^- - T_6^+(2)]^{-1}\},$$
$$C = 2\nu\nu_1\tilde{s}\tilde{c}\{[T_{45}^- - T_6^+(1)]^{-1} - [T_{45}^- - T_6^+(2)]^{-1}\}, \tag{12}$$

and for the level $T_\varepsilon^{-1}(1)$:

$$\begin{aligned} -1/2m_\parallel &= (\mu s\tilde{s} + \mu_1 c\tilde{c})^2[T_6^-(1) - T_6^+(1)]^{-1} \\ &\quad + (\mu s\tilde{c} - \mu_1\tilde{s}c)^2[T_6^-(1) \\ &\quad - T_6^+(2)]^{-1} + \alpha + \beta(3c^2 - 1), \\ -1/2m_\perp &= A + B + \alpha_1 + \beta(3c^2 - 1), \\ \gamma &= |A - B - c^2/2m_0|\,|A + B \\ &\quad + \alpha_1 + \beta(3c^2 - 1)|^{-1}, \end{aligned} \tag{13}$$

where

$$A = [(sv)^2 + (cv_1)^2]\,[T_6^-(1) - T_{45}^+]^{-1},$$
$$B = \nu_1^2\{(c\tilde{s} - \tilde{c}s)^2[T_6^-(1) - T_6^+(1)]^{-1} + (c\tilde{c} + s\tilde{s})^2[T_6^-(1) - T_6^+(2)]^{-1}\} \tag{14}$$

In these expressions, the term with the electron mass m_0 takes into account the magnetic moment of the p state; the following notation was introduced:

$$\mu = 3^{-1/2}a(\xi_0 + 2\xi_1 - 2\varepsilon_1 - \varepsilon_2 - 2\eta_3),$$
$$\mu_1 = 3^{-1/2}a(\xi_0 + 2\xi_1 + 4\varepsilon_1 + 2\varepsilon_2 + 4\eta_3),$$
$$\nu = 6^{-1/2}a(\xi_0 - \xi_1 - 2\varepsilon_1 + 2\varepsilon_2 + 4\eta_3),$$
$$\nu_1 = 6^{-1/2}a(\xi_0 - \xi_1 + \varepsilon_1 - \varepsilon_2 - 2\eta_3),$$
$$\alpha = (-u_1/2 - u_2 + \eta_1 + 2\eta_2)/3, \quad \alpha_1 = (-u_1/2 - u_2 - \eta_1/2 - \eta_2)/3,$$
$$\beta = (-\eta_0 - u_3)/3, \quad r = 3[c_1 + 2c_2 - T_6^-(1)]/2^{1/2}\Delta,$$
$$s = r(r^2+1)^{-1/2}, \quad c = (r^2+1)^{-1/2}.$$

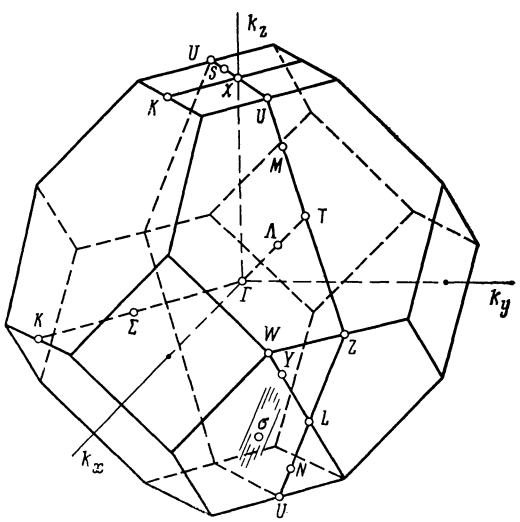

FIG. 1.

To obtain $\tilde{s}$ and $\tilde{c}$ it is necessary to reverse the signs of u_1, u_2, and u_3 in the expressions for s and c. The same substitution in (11) and (13) yields the masses and the γ factor for the even terms.

Point L—its cubic coordinates are $(-1,1,1)\pi/2a$. The even levels $L^+(1), L^+(2), L^+(3)$, which are doubly degenerate in spin, are defined by the cubic equation

$$\omega^3+a\omega^2+b\omega+c=0, \tag{15}$$

where

$$a=-\mathrm{Sp}\, w,$$
$$b=w_{22}w_{33}+w_{11}(w_{22}+w_{33})-w_{23}^2-\Delta^2/3,$$
$$c=-w_{11}(w_{22}w_{33}-w_{23}^2)+\Delta^2(-a+2\Delta/3)/9,$$
$$w_{11}=-\eta_0-\varepsilon_0+u_1-2u_3,\quad w_{22}=-\eta_0+(5\varepsilon_0-u_1+4u_2+2u_3)/3,$$
$$w_{23}=2^{1/2}(\varepsilon_0+u_1-u_2+u_3)/3,$$
$$w_{33}=2\eta_0+(-2\varepsilon_0+u_1+2u_2+4u_3)/3. \tag{16}$$

The column eigenvector corresponding to the level $L^+(n)$ and to the "up" spin contains three nonzero elements:

$$c_{1n}=[w_{11}-L^+(n)]\,[(w_{22}-w_{33})/2+iw_{23}]/N_n,$$
$$c_{2n}=[(w_{22}+w_{33})/2+\Delta/3-L^+(n)]\,[L^+(n)-w_{11}]/N_n,$$
$$c_{3n}=2^{1/2}\Delta\,[L^+(n)-(w_{22}+w_{33})/2-\Delta/3]/3N_n,$$

where N_n is determined by the normalization condition

$$\sum_i |c_{in}|^2=1.$$

Corresponding to the "down" spin is the complex-conjugate column c_{in}^*. The odd eigenfunction $\tilde{c}_{in}$ and the corresponding levels $L^-(n)$ are obtained by substituting $u\to -u$ in (16).

It is known that at the point L there are two close levels of opposite parity, $L^+(2)$, and $L^-(2)$ in our notation, and the distance between them

$$\varepsilon_g=L^+(2)-L^-(2) \tag{17}$$

is small both compared with the Fermi energy and with the remaining energy gaps. Therefore the **k·p** expansion in the vicinity of the point L should be carried out jointly for both bands.

The matrix elements that connect these two spin-degenerate bands can be written in the form of a 2×2 matrix:

$$V_2=\begin{pmatrix} t & v \\ -v^* & t^* \end{pmatrix},$$
$$t=v_{222}k_2+v_{322}k_3,\quad v=v_{122}k_1, \tag{18}$$

where the 1 axis is chosen along a twofold symmetry axis of the point L, the 3 axis coincides with the ΓL direction, i.e., with a threefold axis in the cubic lattice, and the 2 axis is perpendicular to 1 and 3. The matrix elements v_{jnm} (we shall need hereafter not only v_{j22}) are obtained in two steps. First, we expand the off-diagonal blocks of (4) in powers of k in the vicinity of the point L and find the explicit form of the perturbation (5). Second, we transform to the eigenfunctions c_{in} and $\tilde{c}_{im}$ and separate those matrix elements that connect the levels $L^+(n)$ and $L^-(m)$. We get

$$v_{1nm}=6^{-1/2}a\{(c_{1n}^*\tilde{c}_{3m}-c_{3n}\tilde{c}_{1m}^*)$$
$$\times[(2^{-1/2}+i)(\xi_0-\xi_1)+(2^{1/2}-i)(\varepsilon_1-\varepsilon_2+2\eta_3)]$$
$$+(c_{2n}\tilde{c}_{3m}-c_{3n}\tilde{c}_{2m})[(2^{-1/2}-i)(\xi_0-\xi_1)+(2^{1/2}+i)(\varepsilon_1-\varepsilon_2+2\eta_3)]\},$$
$$v_{2nm}=6^{-1/2}a\{(c_{1n}^*\tilde{c}_{1m}+c_{2n}\tilde{c}_{2m}-2c_{3n}\tilde{c}_{3m})$$
$$\times[(\xi_0-\xi_1)/2-\varepsilon_1+\varepsilon_2+2\eta_3]$$
$$+c_{1n}^*\tilde{c}_{2m}[(1/2-2^{1/2}i)(\xi_0-\xi_1)$$
$$+(5-2^{1/2}i)(\varepsilon_1-\varepsilon_2)/3+2(1+2^{1/2}i)\eta_3]$$
$$+c_{2n}\tilde{c}_{1m}[(1/2+2^{1/2}i)(\xi_0-\xi_1)$$
$$+(5+2^{1/2}i)(\varepsilon_1-\varepsilon_2)/3+2(1-2^{1/2}i)\eta_3]\},$$
$$v_{3nm}=3^{-1/2}a\{(c_{1n}^*\tilde{c}_{1m}+c_{2n}\tilde{c}_{2m})(-\xi_0-2\xi_1-\varepsilon_1-\varepsilon_2/2-\eta_3)$$
$$+c_{3n}\tilde{c}_{3m}(-\xi_0-2\xi_1+2\varepsilon_1+\varepsilon_2+2\eta_3)$$
$$-c_{1n}^*\tilde{c}_{2m}[(7+4\cdot 2^{1/2}i)(\varepsilon_1/3+\varepsilon_2/6)-3\eta_3]$$
$$-c_{2n}\tilde{c}_{1m}[(7-4\cdot 2^{1/2}i)(\varepsilon_1/3+\varepsilon_2/6)-3\eta_3]\}.$$

The effective Hamiltonian of the two-band approximation is

$$H=\begin{pmatrix} \varepsilon_g/2 & V_2 \\ V_2^+ & -\varepsilon_g/2 \end{pmatrix}, \tag{19}$$

and the energy origin is halfway between $L^+(2)$ and $L^-(2)$. For the spectrum we obtain the Kane equation

$$\omega^2-\varepsilon_g^2/4=|v|^2+|t|^2=Q_{11}k_1^2+Q_{22}k_2^2+2Q_{23}k_2k_3+Q_{33}k_3^2, \tag{20}$$

the application of which to electrons is bismuth is attributed to Lax. One of the principal values of the quadratic form $|t^2|$ that depends on k_2 and k_3 turns out to be small. Small ε_1 and ε_2 were therefore retained in v_{inm} alongside the large ξ_0 and ξ_1. This is why the term with η_3 was retained in the Hamiltonian (1). Added to the term $\varepsilon_g/2$ is the contribution of the second-order perturbation theory V (5):

$$H_c^{(1)}=\sum_{n=1,3}[|v_{22n}k_2+v_{32n}k_3|^2+(v_{12n}k_1)^2]\,[L^+(2)-L^-(n)]^{-1}, \tag{21}$$

and also the quadratic terms in the expansion of the trigonometric functions from $\hat{u}$ and $\hat{\eta}$ (2):

$$H_c^{(2)}=a^2[D(|c_{12}|^2+c_{22}^2)/4+Ec_{32}^2+Fc_{22}\,\mathrm{Re}\,c_{12}/6$$
$$+Gc_{22}\,\mathrm{Im}\,c_{12}/3\cdot 2^{1/2}], \tag{22}$$

where

$$D=-(\eta_0+u_3)(2l_1^2+k_1^2)+\eta_1(2l_1^2+2l_1l_2-k_1^2)$$
$$+\eta_2(2l_1^2+6l_1l_2-k_1^2)$$
$$-u_1(l_1^2-l_2^2+k_1^2/2)-u_2(3l_1^2-l_2^2+3k_1^2/2),$$
$$E=(\eta_0+u_3-u_1/2)(l_1^2+k_1^2/2)+\eta_1l_1l_2+\eta_2(l_1l_2+l_2^2-k_1^2/2)$$
$$-u_2(l_1^2-l_2^2+k_1^2/2)/2,$$
$$F=\eta_0(8l_2^2+10l_1^2+5k_1^2)-(\eta_2-\eta_1)(2l_1^2-2l_1l_2-k_1^2)$$
$$+u_3(8l_2^2-6l_1^2$$
$$-3k_1^2)+(u_1-u_2)(l_1^2-l_2^2+k_1^2),$$

$$G=\eta_0(2l_2^2-2l_1^2-k_1^2)+2(\eta_1-\eta_2)(2l_1l_2-2l_1^2+k_1^2)$$
$$-2u_3(3l_1^2-l_2^2+3k_1^2/2)$$
$$+2(u_2-u_1)(l_1^2+l_2^2+k_1^2/2),$$
$$l_1=3^{-1/2}k_3+6^{-1/2}k_2,\quad l_2=3^{-1/2}(k_3-2^{1/2}k_2).$$

We combine (21) and (22) to form

$$H_c=H_c^{(1)}+H_c^{(2)}=(\alpha_{11}{}^c k_1^2+\alpha_{22}{}^c k_2^2+2\alpha_{23}{}^c k_2k_3+\alpha_{33}{}^c k_3^2)/2m_0, \tag{23}$$

and denote the corresponding term added to $-\varepsilon_g/2$ in (19) by $-H_v$.

The left-hand side of (20) takes thus the form

$$(\varepsilon_g/2+H_c-\omega)(-\varepsilon_g/2-H_v-\omega). \tag{24}$$

If the right-hand side is left unchanged, we obtain the Cohen equation.[6] Since, however, (24) contains terms of fourth order in k, we must add terms of the same order also to the first part of (20). They appear above all as a contribution of third-order perturbation theory in V (5) to t and v (the general form of the third-order correction in the degenerate case is given in Ref. 5, p. 179), and the higher-order terms of the expansion of the trigonometric functions turn out here to be less significant. We denote this contribution by

$$P(k^4)=P_{1111}k_1^4+P_{2222}k_2^4+P_{3333}k_3^4+P_{1122}k_1^2k_2^2+P_{1133}k_1^2k_3^2$$
$$+P_{2233}k_2^2k_3^2+P_{1123}k_1^2k_2k_3+P_{2223}k_2^3k_3+P_{2333}k_2k_3^3, \tag{25}$$

and obtain ultimately the dispersion equation

$$(\varepsilon_g/2+H_c-\omega)(-\varepsilon_g/2-H_v-\omega)$$
$$=Q_{11}k_1^2+Q_{22}k_2^2+2Q_{23}k_2k_3+Q_{33}k_3^2+P(k^4). \tag{26}$$

Some coefficients of (26) were determined by McClure and Choi[7] by comparison with experimental data. In our case all the α, Q and P in (26) are not independent, and are uniquely expressed in terms of the parameters of the Hamiltonian (1). To find them we present expressions for the experimental quantities. The most extensive and accurate information is available at present on the cyclotron masses and areas of the extremal cross sections.[8]

To this end we transfer all the terms with to the right-hand side of (26). The coefficients of the quadratic terms turn out then to depend on the energy variable ω. For example

$$q_{33}=Q_{33}+[(\omega+\varepsilon_g/2)\alpha_{33}{}^c-(\omega-\varepsilon_g/2)\alpha_{33}{}^v]/2m_0. \tag{27}$$

The coefficients P acquire an increment connected with α:

$$p_{3333}=P_{3333}+\alpha_{33}{}^c\alpha_{33}{}^v/4m_0{}^2. \tag{28}$$

By rotating the coordinates in the 23 plane through the angle φ,

$$\operatorname{tg}2\varphi=q_{23}/(q_{33}-q_{22}), \tag{29}$$

we refer the form bilinear in k_2 and k_3 to the principal axes. We denote the principal axes $q_{\alpha\beta}$ by q_{y1}^2, q_{z1}^2, $q_{x1}^2\equiv q_{11}$, with

$$q_{y1}{}^2,\ q_{z1}{}^2={}^1/_2\{q_{22}+q_{33}\pm[(q_{22}-q_{33})^2+4q_{23}{}^2]^{1/2}\}, \tag{30}$$

where the z_1 axis is chosen in the elongation direction of the equal-energy surface (26). It is known from experiment that all the q_α^2 are positive and that $q_z/q_y\sim 0.1$; for brevity, hereafter $x_1\,y_1,z\to x,\,y,z$.

The terms with p are relatively small. All except p_{zzzz} (since the characteristic value $k_z\sim\omega/q_z$ is large) can be regarded as small when the sections and masses are calculated. The area of the central intersection by the plane $k_x=0$ will be designated as S_x, and the two other principal sections as S_y and S_z. It can be easily seen that the coefficients with odd numbers of z enter quadratically. Omitting them, we get

$$S_x=\frac{4(\omega^2-\varepsilon_g{}^2/4)}{q_yq_z}\int_0^{z_0}dzf(z)\left[1-\frac{1}{2}(\rho_{yyyy}f^2(z)+\rho_{yyzz}z^2)\right],$$
$$f(z)=(1-z^2-\rho_{zzzz}z^4)^{1/2}, \tag{31}$$

where p are taken in dimensionless form:

$$\rho_{\alpha\beta\gamma\delta}=p_{\alpha\beta\gamma\delta}(\omega^2-\varepsilon_g{}^2/4)/q_\alpha q_\beta q_\gamma q_\delta. \tag{32}$$

The upper limit z_0 is determined from the condition $f(z_0)=0$, and it is known from experiment that there exists one z_0. The integral (31) is expressed in terms of complete elliptic integrals of the first and second kind.

The cyclotron mass is determined from the formula

$$m_x=\frac{1}{2\pi}\oint dk_z|\partial\omega/\partial k_z|^{-1},$$

where the derivative must be calculated with the aid of (26), with allowance for the dependence of the small coefficient q_z on ω. We obtain

$$m_x=\frac{2\omega}{\pi q_yq_z}\int_0^{z_0}\frac{dz}{f(z)}(1-wz^2)$$
$$\times\left(1-\frac{1}{2}\rho_{yyzz}z^2-\frac{3}{2}\rho_{yyyy}f^2(z)\right),$$
$$w=\frac{1}{q_z\omega}\left(\omega^2-\frac{\varepsilon_g{}^2}{4}\right)\frac{\partial q_z}{\partial\omega}, \tag{33}$$

Actually, the corrections that distinguish Q from q (27) are small, and the angle of inclination of the "ellipsoid" (29) changes little with changing Fermi energy ω. Therefore w can be calculated with good accuracy by using expression (27) transformed to the rotated axes. It yields

$$\frac{\partial}{\partial\omega}q_z{}^2=\frac{1}{2m_0}(\alpha_{zz}{}^c-\alpha_{zz}{}^v). \tag{34}$$

The section S_y and the mass m_y are obtained from (31) and (33) by the obvious permutation of the indices, while the section S_z, the mass m_z, and the ellipsoid volume V_e are written out explicitly:

$$S_z=\frac{\pi(\omega^2-\varepsilon_g{}^2/4)}{q_xq_y}\left[1-\frac{1}{8}(3\rho_{xxxx}+\rho_{xxyy}+3\rho_{yyyy})\right], \tag{35}$$

$$m_z=\frac{\omega}{q_xq_y}\left[1-\frac{1}{4}(3\rho_{xxxx}+\rho_{xxyy}+3\rho_{yyyy})\right] \tag{36}$$

$$V^e=\frac{2\pi(\omega^2-\varepsilon_g^2/4)}{q_xq_yq_z}z_0\Big[\frac{4}{5}-\frac{2}{15}z_0^2$$
$$-(\rho_{xxzz}+\rho_{yyzz})z_0^2\Big(\frac{2}{3}-\frac{1}{5}z_0^2\Big)$$
$$-\frac{1}{45}(3\rho_{xxxx}+\rho_{xxyy}+3\rho_{yyyy})\Big(4-\frac{8}{7}z_0^2+\frac{1}{7}z_0^4\Big)\Big]. \quad (37)$$

If the dimensionless coefficient ρ_{zzzz} is also small compared with unity, Eqs. (31) and (33) become simpler in analogy with (35) and (36). For example,

$$m_x=\frac{\omega}{q_yq_z}\Big[1-\frac{1}{4}(3\rho_{yyyy}+\rho_{yyzz}+3\rho_{zzzz})$$
$$-\frac{1}{2}w\Big(1-\frac{15}{8}\rho_{zzzz}-\frac{3}{8}\rho_{yyzz}-\frac{3}{8}\rho_{yyyy}\Big)\Big]. \quad (38)$$

3. COMPARISON WITH EXPERIMENT

Our problem is to determine the parameters of the Hamiltonian (1) from experimentally known data on the Fermi surface in bismuth. The hole section is well described by the quadratic approximation (10) (see Édel'man's review[8]). The measured cyclotron masses, which are connected with the effective masses by the relations

$$m_3=m_\perp, \quad m_{1,2}=(m_\parallel m_\perp)^{1/2}$$

(the subscripts 1, 2, and 3 correspond to the direction of the magnetic field relative to the symmetry axes of the point T), and also the extremal sections and the γ factor, are given in the third column of Table I. Also given is the Fermi energy reckoned from the bottom of the zone and calculated from the measured m and S with the aid of the formula $\varepsilon_F^h=S/2\pi m$.

The experimental data on the electronic section are also gathered in the third column of Table I. The angle θ of inclination of the ellipsoid to the basal plane is given besides the cross sections and the masses for the three principal directions. The angle φ (29) is measured from the ΓL direction and is connected with θ by the relation

$$\varphi=\theta-90°+\arccos 1/3=-13.09°. \quad (39)$$

Table I gives also the values of the gap ε_g, obtained in Ref. 9 by extrapolating the data on the semiconducting Bi-Sb alloys with low antimony density. We agree to choose the sign of the gap in the following manner. It is known from experiment that the coefficients α_{zz}^c and α_{zz}^v of the quadratic forms H_c and H_v [see (23)] in terms of the principal axes (29) of the ellipsoid are of like sign and can be regarded as positive if the supscripts c and v in (26) are suitably chosen. The signs of ε_g is then also uniquely determined: at $\varepsilon_g>0$ the band higher in energy at the point L is the one ($L^+(2)$ in our notation) for which the quadratic form H_c has a positive coefficient α_{zz}^c.

We have thus at out disposal 12 independent experimental quantities: from among the data that determine the dimensions of the Fermi surface we can specify, besides the three electron sections, the Fermi energy ε_F^h of the holes or else the condition that the electron and hole densities are equal. The parameters of the Hamiltonian (1) were obtained by finding the minimum of the function

$$f=\Sigma(y_{\text{theor}}/y_{\text{exp}}-1)^2, \quad (40)$$

where y_{theor} and y_{exp} are the calculated and experimental values of any one of the foregoing 12 quantities (y was chosen to be $\tan 2\varphi$ (29) in place of θ). The first attempt to solve the problem has shown that, first, at the point Γ the valence band and the conduction band are not far from the Fermi level, and carriers can appear in the vicinity of this point at definite values of the parameters; second, the spin-orbit coupling Δ can vary in the interval 1.6 ± 0.3 eV without substantial deterioration of the agreement with experiment. We have therefore taken into consideration the results of optical measurements,[10,11] according to which the gap in Γ amounts to 0.71 eV, and in addition we put $\Delta=1.61$ eV, a value taken from calculations[12] for the free atom. It is known that the spin-orbit interaction is determined by small distances in the atom, and Δ should not change when a crystal is considered.

We present below the set of parameters (in eV) of the Hamiltonian (1), obtained as a result of the described optimization:

Δ	ξ_0	ξ_1	η_2	η_3	
1,610	3,389	−1,053	0,149	−0,092	(a)

u_1	u_2	u_3	ε_0
−0,493	0,220	0,257	−0,162

TABLE I. Carrier parameters in bismuth.

Carriers	Parameters	Experiment	Present work	[4]	[13]
Holes	m_3	0,0639±0,0003	0,0658	0,17	0,10
	$m_{1,2}$	0,212±0,0005	0,221	0,65	0,38
	S_3	6,76±0,01	6,45	18,4	7,66
	$S_{1,2}$	22,49±0,02	21,72	65,2	28,9
	γ	1,87 or 2,13	1,80	1,9	–
	$\varepsilon_F{}^h$	11,5	10,7	11,6	8
Electrons	m_x	0,119±0,0005	0,116	0,11	0,161
	m_y	0,088±0,002	0,102	0,091	0,148
	m_z	0,0082±0,00005	0,0085	0,018	0,015
	S_x	19,23±0,05	16,60	17,29	18,0
	S_y	14,48±0,04	14,80	14,6	14,4
	S_z	1,300±0,003	1,33	2,88	1,27
	θ	6,38	8,49	10	8,2
	ε_g	−11±2	−11,3	5,4	16

Note. The values of m are given in units of m_0, S are in units of 10^{-12} g²·cm²/s², ε in MeV, and θ in deg. The experimental values of ε_g were taken from Ref. 9, and all others from Ref. 8.

TABLE II. Parameters of two-band approximation in L.

Parameters	Present work	[7]	[9]	[13]
Q_x	0,905	0,995	0,986	0,496
Q_y	0,811	0,748	0,744	0,441
Q_z	0,0881	0,0704	0,0744	0,0391
α_{zz}^c	0,359	0,676	0,66	0,80
α_{zz}^v	0,901	0,735	1,15	2,78
α_{yz}^c	−1,194	–	–	–
α_{yz}^v	−2,002	−0,145	–	–
P_{xxzz}	1,076	−15,9	–	–
P_{yyzz}	−7,591	−17,0	–	–
P_{zzzz}	−0,106	–	–	–
ε_g	−11,3	−11,4	−9	+5,4
ε_F^e	35,19	35,1	34,7	18,7

Примечание. ε_F^e отсчитывается от середины запрещенной щели и измеряется, как и ε_g, в мэВ. Единицы измерения: Q — 10^8 см/с; P — m_0^{-1}.

The values of η_0, η_1, ε_1, and ε_2, which are not listed here have little effect on the known experimental data and are subject to large uncertainties. All that can be indicated are the approximations $\eta_0 \approx 0.01$ eV, $\varepsilon_2 \approx 0.1$ eV, and $|\varepsilon_1| \sim |\eta_1| \sim 0.05$ eV. If these parameters are set equal to zero, we obtain for the observable quantities the values listed in the fourth column of Table I. Nonzero η_0, η_1, ε_1, and ε_2 yield the somewhat better values $m_y = 0.098 m_0$ and $\theta = 7.63°$, which hardly change the remaining quantities.

We note that we have left out of (1), from the very beginning, one more overlap integral with the third coordination sphere, which makes a contribution of the form $\eta_4 c_x c_y c_z$ to the diagonal matrix element. It does not manifest itself at the points T and L, since it is of third order in k here, and at the point Γ it is a small increment compared with ξ.

It is of interest to compare the presented set of parameters of the Hamiltonian (1) with the corresponding values for IV-VI semiconductors. It was observed in Ref. 1 that ξ_0 does not depend on the chemical composition and varies with the lattice period as

$$\xi_0\,[\text{eV}] = 3{,}725 - 1{,}40\,(a - 3{,}00\ \text{Å}).$$

Substituting here the value $a = 3.289$ corresonding to bismuth, we obtain $\xi_0 = 3.32$ eV, which is close to that given above. With respect to the remaining parameters it is known that $\xi_1 \simeq -0.85$ eV independently of a and of the composition, and all the η are close to 0.1 in absolute value.

The last two columns of Table I contain the results of the pseudopotential calculations. Compared with Ref. 4, account was taken in Ref. 13 of pseudopotential screening. This yielded better values of the cyclotron masses, but the lower valence levels agree less with experiment.

The most important parameters of the two-band approximation (26) at the point L are listed in Table II. The values of Refs. 7 and 9 were selected as to describe best the experimental data. The influence of the term with P_{zzzz}, which compensates for the contribution of the product $\alpha_{zz}^c \alpha_{zz}^v$ [see (28)], is quite unexpected. For any of the variants considered by us, in which account was taken also of η_0, η_1, ε_1, ε_2, the dimensionless coefficient $|\rho_{zzzz}| < 0.03$ [see (32)]; it is equal to 0.1 at $P_{zzzz} = 0$ and at the same values of α_{zz}^c and α_{zz}^v. The two other coefficients, which characterize the deviation from the equal-energy surface from ellipsoidal, have values $\rho_{yyzz} \approx -0.07$, $\rho_{xxzz} \approx -0.01$. As a result, each term of fourth order in k alters the principal sections and the masses by not more than 2%. This explains in fact why the

TABLE III. Frequencies of optical transitions in the 0.7–4.5 eV range.

Experiment			Theory	
$T=10$ K [10]	$T=300$ K			
	[10]	[11]	transition	value
0,71	0,65	0,69	Near Γ	0,58
0,78	0,72	0,81	$\Gamma_{45}^+ \to \Gamma_6^-(2)$	0,76 (⊥)
0,88	0,80	–	$\Gamma_6^+(1) \to \Gamma_6^-(2)$	1,04
1,20	1,05	1,15	$T_{45}^- \to T_{45}^+$	1,13 (∥)
			$L^-(2) \to L^+(1)$	1,15
1,35	1,35	–	$L^+(2) \to L^-(1)$	1,33
			$L^+(3) \to L^-(2)$	1,47
1,55	1,52	1,68	$T_6^+(2) \to T_{45}^-$	1,75 (⊥)
			$T_6^-(2) \to T_6^+(1)$	1,76
2,10	2,10	1,92	$\Gamma_{45}^+ \to \Gamma_6^-(1)$	2,27 (⊥)
			$T_6^-(2) \to T_{45}^+$	2,51 (⊥)
			$\Gamma_{45}^+ \to \Gamma_{45}^-$	2,56 (∥)
			$\Gamma_6^+(1) \to \Gamma_6^-(1)$	2,56
2,45	2,45	2,49	$\Gamma_6^+(2) \to \Gamma_6^-(2)$	2,56
			$T_6^+(2) \to T_6^-(1)$	2,57
			$L^-(3) \to L^+(1)$	2,69
2,96	2,96	2,96	$L^+(3) \to L^-(1)$	2,79
			$\Gamma_6^+(1) \to \Gamma_{45}^-$	2,85 (⊥)
–	4,05	4,00	$\Gamma_6^+(2) \to \Gamma_6^-(1)$	4,08
4,70	4,70	4,38	$\Gamma_6^+(2) \to \Gamma_{45}^-$	4,37 (⊥)

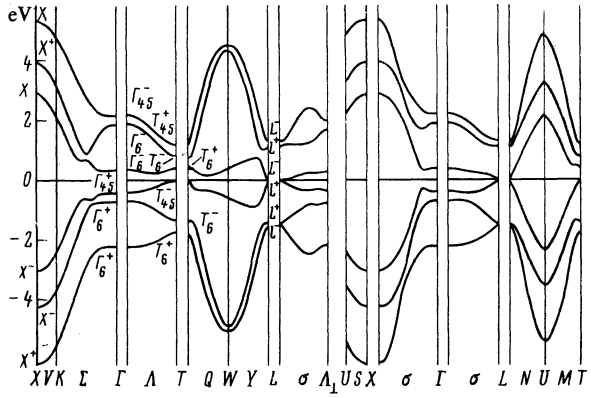

FIG. 2.

electron Fermi surface is so highly ellipsoidal, as noted in Édel'man's review.[8] More noticeable (about 1%) is the deviation of the spectrum from parabolicity, due to the influence of the coefficients α_{zz}^{c} and α_{zz}^{v} on q_z [see (27) and (34)] and are manifest in the cyclotron masses.

The parameters of the Hamiltonian (1) were used in its exact computer diagonalization for the principal directions in the Brillouin zone. The electron spectrum obtained in six bands that are doubly degenerate in spin are shown in Fig. 2. It can be seen that the electrons are indeed in L and the holes in T (the term T_{45}^{-}), and there are no other carriers.

We present the values (in eV) of the levels at the symmetric points of the Brillouin zone (reckoned from the Fermi level):

2,1593 Γ_{45}^{-}	1,1455 T_{45}^{+}	1,2911 $L^{-}(1)$	5,3063 $X^{-}(1)$
1,8699 $\Gamma_{6}^{-}(1)$	0,8310 $T_{6}^{-}(1)$	1,1198 $L^{+}(1)$	3,9226 $X^{+}(2)$
0,3512 $\Gamma_{6}^{-}(2)$	0,3906 $T_{6}^{+}(1)$	−0,0295 $L^{-}(2)$	2,8581 $X^{+}(1)$
−0,4050 Γ_{45}^{+}	0,0108 T_{45}^{-}	−0,0408 $L^{+}(2)$	−3,0370 $X^{-}(2)$
−0,6944 $\Gamma_{6}^{+}(1)$	−1,3668 $T_{6}^{-}(2)$	−1,4960 $L^{+}(3)$	−4,2247 $X^{-}(3)$
−2,2131 $\Gamma_{6}^{+}(2)$	−1,7371 $T_{6}^{+}(2)$	−1,5707 $L^{-}(3)$	−6,1470 $X^{+}(3)$

Compared with the pseudopotential calculations at T and L, the sequence of the two lower levels is reversed. In all other respect the qualitative picture at T is similar, and at L the level $L^{-}(2)$ is lower than $L^{+}(2)$, unlike in Refs. 4 and 13.

In Table III are compared the energies of the optical transitions in the interval 0.7–0.4 eV, observed and calculated by us (the required polarization of the electric field is indicated in the parentheses). Our interpretation differs from that given in Refs. 10 and 11 on the basis of the calcuation of Ref. 4. For example, to singularity at 0.88 (0.80) eV is attributed in Ref. 10 to the $T_{45}^{-}T_{6}^{-}$ transition (the calculated value[4] is 0.915 eV), even though this transition is parity-forbidden. The transition at 1.35 eV is also explained, but the cause of the singularity observed approximately at 3.30 eV is not clear. The situation might be clarified by analysis of the polarization dependence, but no such measurements have been made. A remark made in Ref. 11, that its authors were unable to observe transitions connected with the point T, is puzzling, inasmuch as this point has the same symmetry as the point Γ, whose contribution was observed. An important role might be played also by an elucidation of the temperature dependence, as well as by experiments on doped smaples, for in this case one can expect a substantial frequency shift of the transition connected with the Fermi level.

The authors thank S. V. Repin for help with the computer calculations and A. A. Abrikosov for a discussion of the work.

[1]B. A. Volkov and O. A. Pankratov, Zh. Eksp. Teor. Fiz. **75**, 1362 (1978) [Sov. Phys. JETP **48**, 687 (1978)].
[2]B. A. Volkov and L. A. Fal'kovskiĭ, *ibid.* **85**, 2135 (1983) [**56**, 1239 (1983).
[3]A. A. Abrikosov and L. A. Fal'kovskiĭ, *ibid.* **43**, 1089 (1962) [**16**, 769 (1963)].
[4]S. Golin, Phys. Rev. **166**, 643 (1968).
[5]G. L. Bir and G. E. Pikus, Symmetry and Strain-Induced Effects in Semiconductors, Wiley, 1975.
[6]M. H. Cohen, Phys. Rev. **121**, 387 (1961).
[7]J. W. McClure and K. H. Choi, Sol. St. Commun. **21**, 1015 (1977).
[8]V. S. Edel'man, Usp. Fiz. Nauk **123**, 257 (1977) [Sov. Phys. Usp. **20**, 819 (1977)].
[9]N. B.Brandt, R. German, G. I. Golyshev, L. I. Devyatkov, D. Kusnik, V. Kraak, and Ya. G. Ponomarev, Sov. Phys. JETP **83**, 2152 (1982) [Sov. Phys. JETP **56**, 1247 (1982)].
[10]O. Hunderi, J. Phys. **F5**, 2214 (1970).
[12]F. Herman and S. Skilman, At. St. Calc. New Jersey, 1963 [*sic!*].
[13]J. Rose and R. Schuchardt, Phys. Stat. Sol. (b) **117**, 213 (1983).

Translated by J. G. Adashko