77°K and 300°K, within experimental error, in the range of photon energies 1.0–1.4 ev. The results at 300°K are in good agreement with those of Chynoweth and McKay.[23] Taking into account that $h\nu_0=1.093$ ev at 77°K, we conclude that there is no detectable contribution to the radiation at 77°K in these experiments from an exciton decay process. If the two processes outlined above are jointly responsible for the radiation, that from the intraband transitions must therefore extend to energies $\gtrsim 1.2$ ev, so that the transition from intraband radiation to interband radiation takes place at energies considerably in excess of the band gap. In this connection, from an analysis of the spectral distribution of avalanche breakdown radiation in germanium, Wolff[25] has concluded that the transition in germanium occurs at an energy of 1.3 ev, i.e., approximately twice the band gap energy.

Note added in proof. A further statement on the absence of exciton decay contributions to the avalanche breakdown radiation at 77°K has been privately communicated by A. G. Chynoweth. The critical field for dissociation of excitons is estimated to be of the order of the exciton binding energy (8×10^{-3} ev) divided by its effective radius (100A), i.e., 8 kv cm^{-1}. It is thus reasonable to expect the high field in the avalanche region to dissociate the exciton, if formed, so that no decay radiation is observed. The greatest field strength investigated here in p-π-n structures was 3.7 kv cm^{-1} (see Fig. 5), less than half the estimated critical field for exciton dissociation.

[25] P. A. Wolff, J. Phys. Chem. Solids (to be published).

ACKNOWLEDGMENTS

It is a pleasure to thank for their contributions to this work H. R. Moore, who designed much of the electronic equipment, and N. G. Cranna, who kindly supplied samples. The authors are grateful to many members of the Laboratories for discussions, particularly J. R. Haynes.

PHYSICAL REVIEW VOLUME 121, NUMBER 2 JANUARY 15, 1961

Energy Bands in the Bismuth Structure. I. A Nonellipsoidal Model for Electrons in Bi

MORREL H. COHEN
*Institute for the Study of Metals, University of Chicago, Chicago, Illinois**

and

Hughes Research Laboratories, Malibu, California

(Received September 6, 1960)

The band structure near a minimum at a point of no special symmetry is examined for energies small compared to all band gaps except that to the next lower band. Spin-orbit coupling is included. The theory is specialized to points having the three possible symmetries of electrons in Bi and further simplifications appropriate to Bi made. The resulting nonellipsoidal energy surfaces are studied in some detail. An experiment is suggested which is capable of distinguishing between the three possibilities. Fitting the model to existing information is not carried out in this paper.

I. INTRODUCTION

THE interpretation of the very considerable body of information bearing on the structure of the conduction band in bismuth has usually been on the parabolic ellipsoidal model of Shoenberg.[1,2] In this model, the electrons occupy six sets of ellipsoidal energy surfaces.[3] One set of ellipsoids is given by

$$E(\mathbf{p})=(\alpha_{xx}p_x^2+\alpha_{yy}p_y^2+\alpha_{zz}p_z^2+2\alpha_{yz}p_yp_z)/2m, \quad (1)$$

where x and z are chosen along a dyad axis and the triad axis, respectively, and the crystal momentum $\mathbf{p}$ is measured from the position of the nearest minimum in the conduction band. Two other sets are obtained by rotation of $\pm120°$ around the triad axis, the remaining three by inversion. Because one principal axis lies along a dyad axis, symmetry requires that the six energy minima lie either on the dyad axes or on the reflection planes normal to them. Typical values of the α_{ij} are those derived from a combination of de Haas-van Alphen[1,2,4] and cyclotron resonance data[5,6] by Aubrey and Chambers[6]:

$$\alpha_{xx}=202, \quad \alpha_{yy}=1.67, \quad \alpha_{zz}=83.3, \quad \alpha_{yz}=8.33, \quad (2)$$

* Permanent address.

[1] D. Shoenberg, Proc. Roy. Soc. (London) **A170**, 341 (1939).

[2] D. Shoenberg, *Progress in Low-Temperature Physics*, edited by C. J. Gorter (Interscience Publishers, Inc., New York, 1957), Vol. 2, Chap. 8.

[3] G. E. Smith, Phys. Rev. **115**, 1561 (1959).

[4] J. S. Dhillon and D. Shoenberg, Phil. Trans. Roy. Soc. (London) **A248**, 1 (1955).

[5] J. E. Aubrey and R. G. Chambers, J. Phys. Chem. Solids **3**, 128 (1957).

[6] J. E. Aubrey (private communication), and thesis, Cambridge University, 1959 (unpublished).

or in the principal axis system,

$$\alpha_1=202, \quad \alpha_2=0.83, \quad \alpha_3=84.2, \tag{3}$$

with the principal axis 3 tipped 5.8° from the triad axis. The Fermi energy derived from the de Haas-van Alphen effect on the basis of this model[4] is 0.0177 ev at 4°K.

For a parabolic relation between energy and momentum to hold at the Fermi energy, the latter must be small compared to the vertical energy gap to the valence band. One can estimate from the sum rule for the effective mass that a value of α of order 10^2 requires a band gap of order 0.1 ev. Since the Fermi energy is some 5 times smaller than this, the Shoenberg parabolic-ellipsoidal model (PE model) appears consistent. Lax and collaborators, however, have uncovered by various magneto-optic experiments[7–9] strong indications that the effective mass of electrons in bismuth is energy dependent in contradiction to the Shoenberg model. Lax[8,9] concludes, in effect, that the band gap must be smaller than 0.1 ev, and that the Fermi level must lie in the nonparabolic region of the conduction band. Evidence supporting this smaller value of the band gap derives from the optical absorption edge observed by Boyle and Rogers[10] at 0.06 ev, which Lax[9] interprets as indicating a gap E_g of 0.02 ev, and from the studies of Bi-Sb alloys of Jain,[11] who suggests a value of 0.007 ev for E_g.

The problem of deriving the deviation from parabolic behavior in the vicinity of a small band gap was solved by Kane for InSb.[12] Lax[9] has taken over Kane's results, modifying them slightly to apply to Bi, and proposes the ellipsoidal but nonparabolic model

$$\frac{1}{2m}(\mathbf{p}\cdot\boldsymbol{\alpha}\cdot\mathbf{p})=E(\mathbf{p})[1+E(\mathbf{p})/E_g], \tag{4}$$

for the electron energy surfaces in Bi, where the α_{ij} are similar in form to those of Eq. (1). Lax has shown that his model describes satisfactorily the observed energy dependence of the effective mass.[9]

The criterion for the validity of Eq. (4) is that all three principal components of $\boldsymbol{\alpha}$ be large because of a single small band gap. The values obtained by fitting the parabolic model to experiment, Eq. (3), do not satisfy this criterion, α_2 being of order unity. Therefore, the energy should be parabolic along axis 2, in contradiction to Eq. (4). Accordingly, in Sec. II of this paper we develop a treatment of the band structure near a minimum which does not require that all components of the effective-mass tensor be small. We use symmetry and numerical arguments to simplify the results in Sec. III and propose there a nonparabolic as well as nonellipsoidal model for the band structure of bismuth which should describe most of the Fermi surface to better than 5% accuracy. The details of the energy surfaces are worked out in Sec. IV for the three possible symmetries compatible with existing experimental information. Reinterpretation of the experiments in terms of the new model requires a lengthy point-by-point analysis which will be reserved for a subsequent paper, in which an attempt will be made to reconcile a number of apparently divergent results for both electrons and holes. However, in Sec. V a brief discussion of the implications for experiment is given, with emphasis on new experiments specifically designed to distinguish between the present proposals and previous models. Implications for Sb and As are also discussed.

II. BAND STRUCTURE NEAR A MINIMUM FOR SMALL BAND GAP

Suppose there to be a minimum in the conduction band (band index 1) of a crystal at a point $\mathbf{k}_0$ of no special symmetry within the Brillouin zone. To prepare for our study of Bi, we allow the spin-orbit coupling to be arbitrarily strong and assume the crystal to have inversion symmetry. In that case, the conduction band is doubly degenerate because of time-reversal symmetry. Suppose further that the gap E_g to the valence band (band index 0, also doubly degenerate) at the minimum is very small, whereas all other band gaps are large. While the method by which we analyze this situation is essentially that of Kane,[12] we shall follow more closely the notation and procedures of Cohen and Blount.[13]

The model Hamiltonian we use for the band structure is

$$\mathcal{H}=\frac{P^2}{2m}+V+\tfrac{1}{8}\lambda\!\!\!^{-2}\nabla^2V+\mathbf{P}\cdot\frac{\lambda\!\!\!^{-}}{2mc}\mathbf{s}\times\boldsymbol{\nabla}V, \tag{5}$$

where $\mathbf{P}$ is the actual momentum, $\mathbf{V}$ the crystal potential, $\mathbf{s}$ the electron spin, and $\lambda\!\!\!^{-}$ the Compton wavelength over 2π. The eigenfunctions of $\mathcal{H}$ are of the Bloch form

$$\mathcal{H}\psi_{\mathbf{k}n\rho}=E_n(k)\psi_{\mathbf{k}n\rho}, \tag{6}$$

with $\mathbf{k}$ the wave number, n the band index, and $\rho=1$ or 2 distinguishes the two independent eigenfunctions degenerate by time reversal. Let us measure crystal momentum $\mathbf{p}=\hbar\mathbf{k}$ relative to its value at the minimum $\hbar\mathbf{k}_0$ and ask for the explicit dependence of energy E_1 on crystal momentum $\mathbf{p}$ near the minimum.

To answer the question, expand the Bloch function at $\mathbf{k}$ in the set of Bloch functions at $\mathbf{k}_0$,

$$\psi_{\mathbf{k}1\rho}=\sum_{n\rho'}C_{n\rho'}(\mathbf{p})e^{i(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}}\psi_{k_0n\rho'}. \tag{7}$$

[7] R. J. Keyes *et al.*, Phys. Rev. **104**, 1805 (1956).
[8] B. Lax, Revs. Modern Phys. **30**, 122 (1958).
[9] B. Lax, Bull. Am. Phys. Soc. **5**, 167 (1960).
[10] W. S. Boyle and K. F. Rodgers, Phys. Rev. Letters **2**, 338 (1959).
[11] A. L. Jain, Phys. Rev. **114**, 1518 (1959).
[12] E. O. Kane, J. Phys. Chem. Solids **1**, 249 (1957).
[13] M. H. Cohen and E. I. Blount, Phil. Mag. **5**, 115 (1960).

Inserting (7) into (6) leads to

$$\left(\epsilon_n+\frac{p^2}{2m}\right)C_{n\rho}+\sum_{n'\rho'}\mathbf{p}\cdot\langle n\rho|\,\mathbf{v}\,|n'\rho'\rangle C_{n'\rho'}=E_1C_{n\rho}, \quad (8)$$

for the expansion coefficients, where ϵ_n is the energy of band n at $\mathbf{k}_0$ and $\mathbf{v}$ is the velocity operator

$$\mathbf{v}=\mathbf{P}/m+(\lambda/2mc)\mathbf{s}\times\nabla V. \quad (9)$$

Since all band gaps are large except that to band 0, only $C_{0\rho}$ and $C_{1\rho}$ are of order unity. All other coefficients are small and need be calculated only to lowest order in $\mathbf{p}$ with the result

$$C_{n\rho}=\sum_{\rho'}\left[\frac{\mathbf{p}\cdot\langle n\rho|\,\mathbf{v}\,|0\rho'\rangle C_{0\rho'}}{\epsilon_0-\epsilon_n}+\frac{\mathbf{p}\cdot\langle n\rho|\,\mathbf{v}\,|1\rho'\rangle C_{1\rho'}}{\epsilon_1-\epsilon_n}\right]. \quad (10)$$

Substitution of (10) into (8) enables us to express E_1 as the eigenvalue of an effective Hamiltonian $\mathcal{H}'$, a four-by-four matrix operating on C', a four-component wave function with components $C_{00}, C_{01}, C_{10}, C_{11}$ in that order. The form of $\mathcal{H}'$ can be simplified by use of time-reversal symmetry as in Eqs. (38) and (39) of reference 13, and its writing facilitated by the abbreviations

$$K_0=\mathbf{p}\cdot\langle 01|\,\mathbf{v}\,|01\rangle+\frac{p^2}{2m}+\mathbf{p}\cdot\sum_{n\rho}{}'\frac{\langle 01|\,\mathbf{v}\,|n\rho\rangle\langle n\rho|\,\mathbf{v}\,|01\rangle}{\epsilon_0-\epsilon_n}\cdot\mathbf{p}, \quad (11)$$

$$K_1=\frac{p^2}{2m}+\mathbf{p}\cdot\sum_{n\rho}{}'\frac{\langle 11|\,\mathbf{v}\,|n\rho\rangle\langle n\rho|\,\mathbf{v}\,|11\rangle}{\epsilon_1-\epsilon_n}\cdot\mathbf{p}, \quad (12)$$

$$t=\mathbf{p}\cdot\langle 01|\,\mathbf{v}\,|11\rangle, \quad (13)$$

$$u=\mathbf{p}\cdot\langle 01|\,\mathbf{v}\,|12\rangle, \quad (14)$$

$$E_g=\epsilon_1-\epsilon_0. \quad (15)$$

The first term in K_0 gives the slope of band 0 at $\mathbf{k}_0$, which need not vanish as it does in band 1; the remaining terms give the contribution of all bands other than 1 to the inverse effective mass at $\mathbf{k}_0$ for band 0, and similarly for K_1. In this way we arrive at

$$\mathcal{H}'=\begin{bmatrix}K_0-E_g & 0 & t & u\\ 0 & K_0-E_g & -u^* & t^*\\ t^* & -u & K_1 & 0\\ u^* & t & 0 & K_1\end{bmatrix}, \quad (16)$$

for $\mathcal{H}'$ after setting the zero of energy at the minimum in the conduction band.

It can readily be verified that $\mathcal{H}'$ satisfies

$$(\mathcal{H}')^2-(K_0+K_1-E_g)\mathcal{H}'+[K_1(K_0-E_g)-(|t|^2+|u|^2)]\mathbf{1}=0, \quad (17)$$

where $\mathbf{1}$ is the unit matrix. Equation (17) is immediately diagonalized; that is E_1 and E_0 satisfy the same equation:

$$E^2-(K_0+K_1-E_g)E+[K_1(K_0-E_g)-(|t|^2+|u|^2)]=0. \quad (18)$$

The solutions of (18) are

$$E_{1,0}=\tfrac{1}{2}(K_0+K_1-E_g)\pm\{[\tfrac{1}{2}(E_g+K_1-K_0)]^2+|t|^2+|u|^2\}^{\frac{1}{2}}, \quad (19)$$

where the positive sign goes with E_1. To identify the terms in (19), we go to the limit of small $\mathbf{p}$, i.e., K_1-K_0 small relative to E_g and $|t|^2+|u|^2$ small relative to $(E_g)^2$, and expand the root:

$$E_1=K_1+(|t|^2+|u|^2)/E_g. \quad (20)$$

Comparing (20) with

$$E_1=\frac{1}{2m}(\mathbf{p}\cdot\boldsymbol{\alpha}\cdot\mathbf{p}), \quad (1')$$

which also holds at the bottom of the band, we see that the second term in (20) gives the contribution of band 0 to the sum rule for the inverse effective mass at the bottom of band 1. Passing to larger $\mathbf{p}$ where $|t|^2+|u|^2$ becomes comparable to E_g, we can obtain the model of Lax only if we suppose that K_0 and K_0+K_1 remain smaller than E_g. Equation (4) then follows directly from Eq. (18). However, we shall show in the next section that for certain directions of $\mathbf{p}$, $-K_0$ and K_1 are comparable to E_F and hence possibly to E_g in Bi. Consequently Lax's model does not apply, and we must work directly with the more complete Eqs. (18) and (19).

III. A MODEL FOR BISMUTH

The experiments of Shoenberg[1,4] and of many others have demonstrated clearly that the Fermi surface of the electrons in Bi has either reflection symmetry in planes perpendicular to the binary axes and/or twofold rotation symmetry around the binary axes whatever model is used to interpret them. Such rotation or reflection symmetry of the Fermi surface can result if the minima of the conduction band lie at, on, or in any symmetry point, line, or plane in the Bi structure except the trigonal axis.[14] All the symmetry points of the Bi Brillouin zone are shown in Fig. 1 and listed in Table I together with their locations, symmetry elements, multiplicity of distinct minima which might be associated with them, and vanishing components of $\mathbf{A}$ [see Eq. (21)] required by symmetry. All those above the line have the twofold rotation and/or reflection symmetry required for the conduction band minima. The points Σ, K, Q, W, S, V, and Y have twofold rotation symmetry and a multiplicity of 6. The points σ, M, U, and N have reflection symmetry and a multiplicity

[14] An erroneous limitation of the possible locations of minima was made in reference 13.

TABLE I. Symmetry points in the brillouin zone of the bismuth structure.

Point	Location	Symmetry elements	Multiplicity	A
Σ	Binary axis (BA)	C_2	6	$A_2, A_3=0$
K	Centers of edges not parallel trigonal plane	C_2	6	$A_2, A_3=0$
Q	Projection of BA in trigonal face	C_2	6	$A_2, A_3=0$
W	Corners	C_2	6	$A_2, A_3=0$
σ	Reflection plane (RP)	IC_2	6	$A_1=0$
M	Intersection of RP and trigonal face	IC_2	6	$A_1=0$
U	RP, edge center	IC_2	6	$A_1=0$
N	Intersection of RP and hexagonal face	IC_2	6	$A_1=0$
L	Center, hexagonal face	C_2, I, IC_2	3	$\mathbf{A}=0$
S	Intersection of RP and rectangular face	C_2	6	$A_1=0$
X	Center, rectangular face	C_2, I, IC_2	3	$\mathbf{A}=0$
V	Projection of BA on rectangular face	C_2	6	$A_2, A_3=0$
Y	Projection of BA on hexagonal face	C_2	6	$A_2, A_3=0$
Γ	Center of zone	$2C_3, 3C_2, I, 2IC_3, 3IC_2$	1	$\mathbf{A}=0$
Λ	Trigonal axis	$2C_3, 3C_2$	2	$A_1, A_2=0$
T	Center, trigonal face	$2C_3, 3C_2, I, 2IC_3, 3IC_2$	1	$\mathbf{A}=0$

of 6. The points L and X have both rotation and reflection symmetry and a multiplicity of 3.

If the breakdown of the parabolic-ellipsoidal model does not invalidate Smith's conclusion that there are 6 minima,[3] we may be able to rule out the points L and X as possible locations. However, the differing behavior of

$$\mathbf{A}=\langle 01|\,\mathbf{v}\,|01\rangle, \tag{21}$$

for the three sets of points permits assigning the minima to one of the three sets by an independent experiment, as we shall show. In particular, A_1 vanishes for points having reflection symmetry, A_2 and A_3 vanish for rotation symmetry, and $\mathbf{A}$ vanishes for points having both symmetries. These results for $\mathbf{A}$ follow from the symmetry properties of the Bloch functions at such a point $\mathbf{k}_0$, as studied in reference 13. Additional consequences of the symmetry properties are that K_1, $[|t|^2+|u|^2]/E_g$, and the effective-mass terms in K_0 all have the form given in the right side of (1) with principal axis x or 1 along the binary axis.

Symmetry arguments like the above can carry us no further towards simplification and sharper specification of the model. We now proceed along a direction suggested by the actual numerical values obtained by fitting the parameters of the parabolic-ellipsoidal model to experiment. The small value of α_2 in Eq. (3) suggests that E varies very much more slowly with p in the direction of principal axis 2 than along 1 or 3. This implies that in the principal axis system of $|t|^2+|u|^2$ the coefficient of p_2^2 in $|t|^2+|u|^2$ either vanishes or is of order 1% of the coefficient of p_1^2 or p_3^2, which cannot be explained by a symmetry argument.[13] Thus, while the terms $|t|^2+|u|^2$ dominate the dependence of E on p_1^2 and p_3^2, the quadratic terms in K_0 and K_1 dominate the dependence of E_1 on p_2. For $|p_1|$ and $|p_3|$ small, therefore, $|t|^2+|u|^2$ may be neglected in (19) giving a parabolic dependence,

$$E_1=K_1, \quad |p_1| \text{ and } |p_3| \text{ small}, \tag{22}$$

of energy on crystal momentum quite distinct from that predicted by Eq. (4).

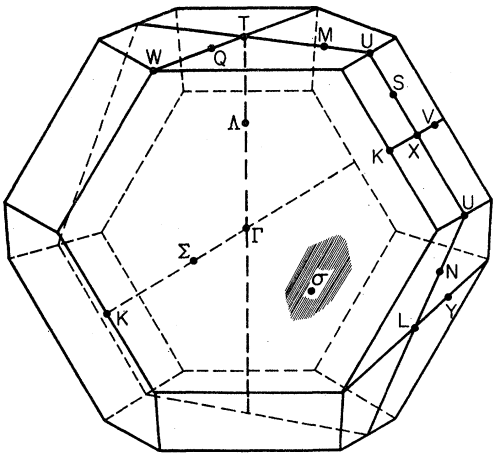

FIG. 1. Brillouin zone and symmetry points for the bismuth structure.

Let us indicate all velocity matrix elements entering our expressions which are of ordinary, i.e., atomic, magnitude schematically by v. The extreme values of $|vp_1|$ and $|vp_3|$ are of order E_F, so that the contributions of K_0 and K_1 are of order $E_F \times E_F/\Delta E$ where ΔE is a typical band gap and of order 1 ev. Because E_F is of order a few hundredths of an ev, the p_1 and p_3 dependence of K_1 and of the quadratic term in K_0 may be neglected. As a further simplification, we ignore the p_2 dependence of $|t|^2+|u|^2$; we estimate the error to be at most 10% in the energy and probably less.

Our conclusions are summarized by the following equations:

$$K_0=-(p_2^2/2m_2')+\mathbf{A}\cdot\mathbf{p}, \tag{23}$$

$$K_1=p_2^2/2m_2, \tag{24}$$

$$|t|^2+|u|^2=E_g\left(\frac{p_1^2}{2m_1}+\frac{p_3^2}{2m_3}\right) \tag{25}$$

which are estimated to reproduce the linear dimensions of the Fermi surface in **k** space to 5%, but may lead to a 10% error in the Fermi energy. Insertion of Eqs. (23)–(25) into (18) and (19) gives the dependence of E on **p** which makes up our model of the band structure of Bi. For all locations of minima listed in Table I and shown in Fig. 1, the cross sections of the constant energy surfaces perpendicular to axis 2 are ellipsoids with axes along directions 1 and 3, but with centers shifted away from the minima by an amount which may vary with p_2. The detailed properties of the energy surfaces are derived from the equations of the model in the next section separately for the three types of symmetry under consideration (rotation, reflection, and rotation plus reflection).

IV. CONSTANT ENERGY SURFACES

A. Rotation Plus Reflection Symmetry; A Vanishes

When **A** vanishes, as is the case for minima at L or X, Eq. (18) may be put into the simple form

$$\frac{p_1^2}{2m_1}+\frac{p_3^2}{2m_3}=\frac{1}{E_g}\left\{\left(E-\frac{p_2^2}{2m_2}\right)\left(E+E_g+\frac{p_2^2}{2m_2'}\right)\right\}. \quad (26)$$

The energy surfaces are ellipsoidal only for E significantly less than E_g. Otherwise the surface is such that its sections normal to axis 2 are ellipses with axes along directions 1 and 3. Its extrema along 2 are at $\pm(2m_2E)^{\frac{1}{2}}$ as for the ellipsoidal case. However, the areas of these normal sections vary differently with p_2 than in the ellipsoidal case. The area of the central section ($p_2=0$) may even be a local minimum, i.e., the areas of the normal sections increase with p_2 near $p_2=0$, provided that

$$y>1 \quad \text{and} \quad E>E_g/(y-1), \quad (27)$$

where

$$y=m_2/m_2'. \quad (27a)$$

Such a necking in of the Fermi surface in the vicinity of the central section would give rise to a large period in $1/H$ in the de Haas-van Alphen oscillations which would be quite easy to observe. Inasmuch as no doubling of the periods appears to have been observed in Bi for **H** near axis 2, we should require that

$$y<1 \quad \text{or} \quad E_F<E_g/(y-1), \quad (28)$$

if we are to apply this model to electrons in Bi.

B. Rotation Symmetry; A_2 and A_3 Vanish

When A_2 and A_3 vanish but not A_1, as is the case for rotation symmetry (points Σ, K, Q, W, S, V, and Y), Eq. (18) may be put into the form

$$\frac{(p_1')^2}{2m_1}+\frac{p_3^2}{2m_3}=\frac{1}{E_g}\left(E-\frac{p_2^2}{2m_2}\right)\times\left[\left(E+\frac{p_2^2}{2m_2'}+E_g\right)+\beta_1\left(E-\frac{p_2^2}{2m_2}\right)\right], \quad (29)$$

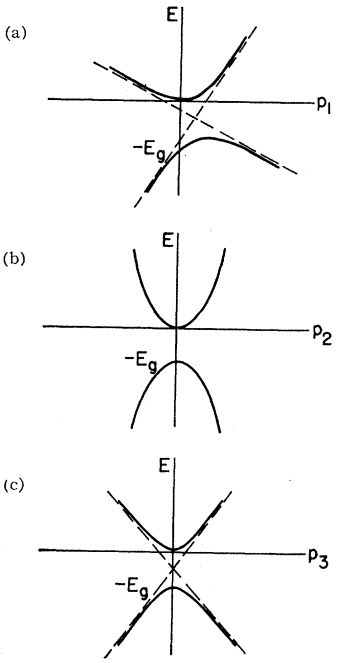

FIG. 2. Sketches of E vs p curves for the case of rotation symmetry (case B). (a) E vs p_1, (b) E vs p_2, (c) E vs p_3.

where

$$p_1'=p_1+\Delta p_1, \quad (30)$$

$$\Delta p_1=\frac{m_1A_1}{E_g}\left(E-\frac{p_2^2}{2m_2}\right), \quad (31)$$

$$\beta_1=\frac{\frac{1}{2}m_1A_1^2}{E_g}=\frac{\frac{1}{4}(\langle 01|v_1|01\rangle)^2}{|\langle 01|v_1|11\rangle|^2}. \quad (32)$$

The energy surfaces again have elliptical sections normal to direction 2 with axes along 1 and 3 and extreme values of p_2 of $\pm(2m_2E)^{\frac{1}{2}}$. The centers of the ellipses, however, are shifted along 1 by Δp_1, Eq. (31), the shift vanishing at the extrema of p_2 and increasing quadratically with decreasing $|p_2|$.

The area of these sections normal to 2 is simply

$$\mathfrak{A}=\frac{2\pi(m_1m_3)^{\frac{1}{2}}}{E_g}\left(E-\frac{p_2^2}{2m_2}\right)\times\left[\left(E+\frac{p_2^2}{2m_2'}+E_g\right)+\beta_1\left(E-\frac{p_2^2}{2m_2}\right)\right], \quad (33)$$

a formula which holds for the valence as well as the conduction band. The sections exist for those values of E and p_2 which give a positive $\mathfrak{A}$ when substituted into (33). If β_1 is less than y, then $\mathfrak{A}$ is positive for E positive (conduction band) and $p_2^2/2m_2$ less than E. For E negative (valence band), $\mathfrak{A}$ is positive if $p_2^2/2m_2'$ is such that

$$\left[E_g+\frac{p_2^2}{2m_2'}-\beta_1\frac{p_2^2}{2m_2}\right]\Big/(1+\beta_1)<|E|,$$

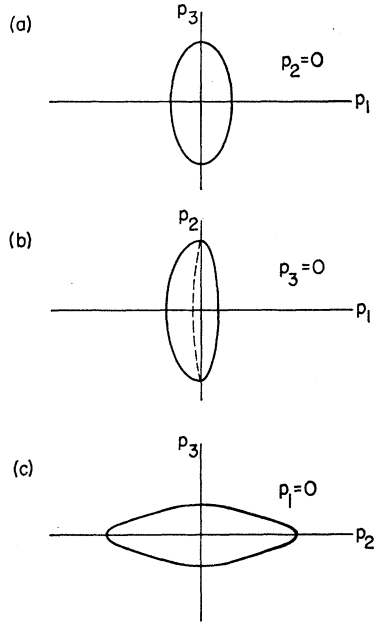

FIG. 3. Principal sections of the energy surfaces for the case of rotation symmetry (case B). (a) $p_2=0$, (b) $p_3=0$, (c) $p_1=0$.

Thus, the energy surfaces have maximum values of $|p_2|$, i.e., are closed, in both valence and conduction bands when y exceeds β_1. This remains true for the conduction band when $\beta_1>y$ but not for the valence band, where, for both positive and negative E, $\mathfrak{A}$ is positive only as long as

$$\frac{p_2^2}{2m_2'}>\frac{[E_g+(1+\beta_1)E]}{(\beta_1/y-1)}.$$

Thus when β_1 exceeds y the energy surfaces in the valence band are not closed, and there is a saddle point and not a maximum in the valence band, a situation unlike that in Bi.

From Eq. (33) for $\mathfrak{A}$ one can see that necking of the central section ($p_2=0$) will occur in the conduction band if

$$y>1+2\beta_1 \quad \text{and} \quad E>E_g/[y-(1+2\beta_1)].$$

Thus, to apply our model to Bi, we must require that

$$y>\beta_1, \tag{34}$$

and

$$y<1+2\beta_1 \quad \text{or} \quad E_F<E_g/[y-(1+2\beta_1)]. \tag{35}$$

One more restriction upon the possible values of the parameters derives from the requirement that the shift of the centers of the sections normal to axis 2 must not be so large that the section in the 1, 2 plane becomes sausage shaped, a geometric feature which would readily have been detected by the de Haas-van Alphen effect but was not. This restriction on y,

$$y+(2\beta_1 y)^{\frac{1}{2}}<1 \quad \text{or} \quad E_F<E_g/[y+(2\beta_1 y)^{\frac{1}{2}}-1], \tag{36}$$

is more severe than (35).

E vs p curves are sketched in Fig. 2. Note that the maximum in the valence band is displaced from the minimum in the conduction band by

$$p_0=m_1A_1/(1+\beta_1), \tag{37}$$

and that the gap in energy between the bottom of the conduction and top of the valence band is

$$E_t=E_g/(1+\beta_1). \tag{38}$$

Principal sections of the energy surfaces are sketched in Fig. 3. Note the possibility of a lens-shaped Fermi surface indicated by the section in the 1, 2 plane [Fig. 3(b)]. The de Haas-van Alphen effect, cyclotron resonance, and geometric resonance experiments already performed on Bi would not reveal such a distortion of the Fermi surface from an ellipsoid.

C. Reflection Symmetry; A_1 Vanishes

When A_1 vanishes but not A_2 and A_3, as is the case for reflection symmetry (points σ, M, U, and N), Eq. (17) may be put into the form

$$\frac{p_1^2}{2m_1}+\frac{p_3'^2}{2m_3}=\frac{1}{E_g}\left(E-\frac{p_2^2}{2m_2}\right)\times\left[\left(E+E_g+\frac{p_2^2}{2m_2'}-A_2p_2\right)+\beta_3\left(E-\frac{p_2^2}{2m_2}\right)\right], \tag{39}$$

where p_3', Δp_3, and β_3 are given by Eqs. (30)–(32) with the index 1 replaced by 3. The energy surfaces once again have elliptical sections normal to direction 2 with axes along 1 and 3; the centers of the ellipses are now shifted along 3 by Δp_3, see Eq. (31); but the extreme values of p_2 need not be $\pm(2m_2E)^{\frac{1}{2}}$ for the conduction band, as we shall show.

The area of the sections normal to 2 is

$$\mathfrak{A}=\frac{2\pi(m_1m_3)^{\frac{1}{2}}}{E_g}\left(E-\frac{p_2^2}{2m_2}\right)\times\left[\left(E+E_g+\frac{p_2^2}{2m_2'}-A_2p_2\right)+\beta_3\left(E-\frac{p_2^2}{2m_2}\right)\right], \tag{40}$$

for both valence and conduction bands. The condition following from (40) that there be a maximum and not a saddle point in the valence band is now

$$\beta_3<y. \tag{41}$$

A more perspicuous form for $\mathfrak{A}$ is

$$\mathfrak{A}=2\pi(1+\beta_3)(m_1m_3)^{\frac{1}{2}}E_g^{-1}[E-f(p_2)][E-g(p_2)], \tag{42}$$

where

$$f(p_2)=p_2^2/2m_2, \tag{43}$$

$$g(p_2)=E_t-\frac{y-\beta_3}{1+\beta_3}\frac{(p_2+\Delta p_2)^2}{2m_2}, \tag{44}$$

$$E_t=E_g\frac{\beta_2-(y-\beta_3)}{(1+\beta_3)(y-\beta_3)}, \tag{45}$$

$$\beta_2=\tfrac{1}{2}m_2A_2^2/E_g, \tag{46}$$

$$\Delta p_2=-m_2A_2/(y-\beta_3). \tag{47}$$

$\mathfrak{A}$ is positive in those ranges of p_2 for which the factors $(E-f)$ and $(E-g)$ are both positive or both negative, which are the ranges of p_2 spanning the constant-energy surfaces. Four cases may be distinguished on the basis of first the maximum value of g, E_t of Eq. (45), and second the associated value of Δp_2, Eq. (47), the position of the maximum. For energies both positive and greater than E_t, the energy surface consists of a single piece enclosing electrons with p_2 in the range $\pm(2m_2E)^{\frac{1}{2}}$ for all four cases. As this appears to occur in Bi, we must require that

$$E_t<0, \quad \text{i.e.,} \quad \beta_2<(y-\beta_3) \quad (\text{Case } a) \tag{48a}$$

or

$$E_t>0, \quad \text{i.e.,} \quad \beta_2>(y-\beta_3) \quad \text{and} \quad E_F>E_t, \tag{48b}$$

if we are to apply Eq. (39) to Bi. The differences between the four cases, though not relevant to Bi *per se*, may be important for Bi-Sb alloys and Sb. First, however, we discuss the properties of the energy surfaces common to all four cases when $E>0$, E_t.

We note first that the energy surfaces consist of a single piece but that the central section ($p_2=0$) is no longer an extreme section. For small y, β_2, and β_3, the displaced extreme section remains a maximum. As these parameters increase, the maximum can change over to a minimum. The exact relation among the parameters for the occurrence of this necking is complicated and unprofitable to write down here. Secondly, we note that a sausage-shaped Fermi surface[15] can occur for large values of y, β_2, and β_3 because of the shift of the centers of the ellipses by Δp_3 along axis 3. Once again the criterion for this is too complicated to write down. In Bi, presumably neither the neck nor the sausage shape occurs.

Now we return to the differences between the four cases:

(a) The condition

$$\beta_2<y-\beta_3 \tag{49}$$

holds, when the top of the valence band occurs at a finite energy $-E_t$ below the bottom of the conduction band as for the other two kinds of symmetry [Fig. 4(a)]. In this case all the energy surfaces in both conduction and valence band have the characteristics discussed in the previous paragraph.

(b) Condition (49) is violated but

$$\beta_2<1+y, \tag{50}$$

so that f and g do not intersect [Fig. 4(b)]. There is a simple overlap between the valence and conduction bands. For energies between 0 and E_t (>0), the energy surfaces consist of a piece containing electrons in the conduction band and a piece containing holes in the valence band. The shape of each piece is like that already discussed. For Bi-Sb alloys having a concentration greater than 5%, the electrical resistivity becomes temperature independent below 25°K.[11] Such behavior would occur for a band structure like that discussed here with $E_F<E_t$. This case may also be of interest for Sb, where the electrons and holes both appear to have reflection and/or twofold rotation symmetry.[16]

(c) If condition (50) is violated but

$$(\Delta p_2)^2/2m_2>E_t, \tag{51}$$

then the maximum of g falls outside of f, the situation depicted in Fig. 4(c). More explicitly, (51) is equivalent to

$$y<1+2\beta_3 \quad \text{and} \quad \beta_2<(y-\beta_3)^2/[y-(1+2\beta_3)]. \tag{51'}$$

In this case, for energies between 0 and E_t the energy surfaces once again consist of an electron piece and a hole piece. If E_a and E_b ($E_a<E_b$) are the common values of f and g at their intersection, then for $E<E_a$ and $E_b<E<E_t$ the energy surfaces are like those of case (b). The electron surface lies between $\pm(2m_2E)^{\frac{1}{2}}$ along axis 2, and the hole surface is similarly bounded by the two values of p_2 for which $g=E$. There is a point of contact between the valence and conduction band at the energy E_a and the point $p_2=-(2m_2E_a)^{\frac{1}{2}}\times\mathrm{sgn}(\Delta p_2)$, and a second point of contact at E_b and $p_2=-(2m_2E_b)^{\frac{1}{2}}\,\mathrm{sgn}(\Delta p_2)$. For $E_a<E<E_b$ the electron surface is bounded at one extreme by $f=E$ and at the other extreme by $g=E$, and similarly for the hole surface, changing the geometry of the energy surfaces somewhat from that of case (b).

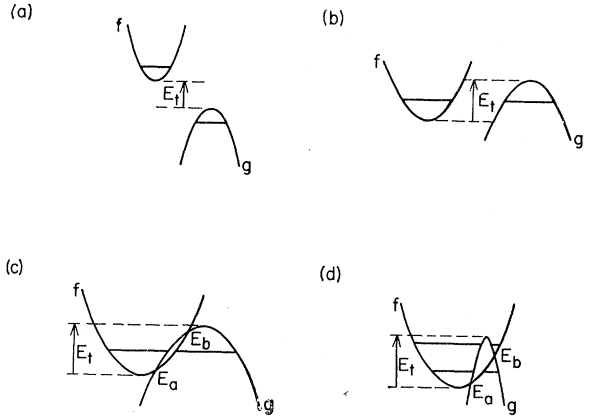

FIG. 4. p_2-dependence of f and g, the functions entering the area of sections normal to p_2. (a) Gap between valence and conduction bands. (b) Overlap. (c) Overlap with two points of contact between the electron and hole energy surfaces. (d) One point of contact occurs between the maximum in the valence band and a local minimum in the conduction band.

[15] We mean by a "sausage-shaped" Fermi surface one which is concave-convex in the p_2-p_3 plane because of a large Δp_3.

[16] H. Juretschke and S. Friedman (to be published).

(d) Because of the way we set up the model, the minimum of f cannot occur within the curve for g (see the Appendix), but, if condition (51) is violated, the maximum of g at $p_2=-\Delta p_2$ lies within f, Fig. 4(d). Up to energies somewhat greater than E_a, the energy surfaces are much like those of case (c). However, the electron surface does not contact the hole surface again at the energy E_b. Instead, the hole surface vanishes and a second electron surface appears at the energy E_b. Thus, although the conduction and valence bands come into contact at the energy E_b and $p_2=-(2m_2E_b)^{\frac{1}{2}}\times\mathrm{sgn}(\Delta p_2)$, the contact is between a maximum in the valence band and a minimum in the conduction band and not between two extended surfaces as at E_a. The two separate pieces of electron surface coalesce at $p_2=-\Delta p_2$ for energies equal to or greater than E_t, but a neck remains for a range of energies above E_t.

V. SUMMARY AND CONCLUSIONS

In Secs. II, III, and IV we have set up and discussed what we regard as the simplest model of the band structure of Bi capable of accurately portraying departures from the parabolic-ellipsoidal (PE) model. In doing so, we have used only the observed symmetry of the electron Fermi surface and the numerical values of the parameters of the PE model.

The extent of the departures from the PE model depends strongly on the ratio E_F/E_g. As discussed in the introduction, the existing information about the value of E_F/E_g is scarce at best. Since knowledge of E_F/E_g is crucial, direct measurement of the absorptivity of Bi in the infrared near where Boyle and Rodgers found the decrease in transmission at 0.060 ev would be of very great interest.

The existing experimental information does not distinguish between three possible kinds of symmetry for the electron energy surfaces, twofold rotation and/or reflection symmetry. The deviations from the PE model are quite different in the three different cases, and experiments set up to look for these deviations could thus further restrict the symmetry of the Fermi surface. One such experiment is the "tilt effect" in the ultrasonic attenuation,[17,18] which permits direct measurement of the Fermi velocity. The Fermi velocity so measured will show inversion symmetry in case A, rotation symmetry in case B, and reflection symmetry in case C, all readily distinguishable by studying the tilt effect as a function of orientation of magnetic field and direction of propagation. In contrast, it would be very difficult to infer these differences in symmetry from de Haas-van Alphen,[1,2,4] cyclotron-resonance,[5,6] geometric-resonance,[17] or anomalous-skin-effect[3] data. The anomalous-skin-effect data, when interpreted on the PE model, do strongly support 6 ellipsoids. Until we have explored the effect of departures from the PE model on the anomalous skin effect, however, it would be wise not to rule out the rotation and reflection symmetry of case A on this basis.

An overlap between the valence and conduction bands and even points of contact between them can occur if the conduction band minima have reflection symmetry. Although such features of the band structure could have been observed in the de Haas-van Alphen effect in Bi and were not, they may possibly provide the explanation for the low-temperature behavior of the resistivity of Bi-Sb alloys.[11] This in turn would suggest that the energy surfaces in Bi may have reflection symmetry, case C, with $E_F>E_t$. Case C may also apply to Sb, where the electrons and holes both appear to have one principal axis along the dyad axis.[16]

The models do have a large number of parameters: E_F, E_g, m_1, m_2, m_2', m_3, and the angle of tilt, i.e., 7, in case A, plus A_1 or 8 in case B, or plus A_2 and A_3 or 9 in case C. Nevertheless, the amount of experimental information is considerable. Reinterpretation of existing data together with the tilt effect and the infrared absorptivity should suffice to distinguish between the possible symmetries and to give numerical values to the parameters. The task of working out the theory of the various experiments for the new model will be tackled in a separate paper.

Finally, the model presents a single analytic expression for the constant-energy surfaces which permits one to illustrate in a clear and simple way a variety of general features of band structure, e.g., points of contact, to name but one.

ACKNOWLEDGMENTS

The author wishes to acknowledge that this work received considerable stimulation from the 1960 Buckley lecture given by Lax.[9] The work was supported in part by the National Science Foundation.

ADDED NOTE

After the above was written, the author received a preprint of a paper by Wolff[19] in which was developed independently the two-band model of Eq. (4) proposed also by Lax.[9] Wolff has made a detailed study of this model and infers from the de Haas-van Alphen, cyclotron resonance, and optical data that $E_F=0.022$ ev and $E_g=0.042$ ev in Bi. As emphasized in Sec. III of this paper, the two-band model of Eq. (4) cannot be applied directly to Bi because the effective mass is large along axis 2. Nevertheless, our own independent analysis of the experimental data had shown that the departures from (4) discussed in Sec. IV do not signficantly affect the magnetic energy levels for the specific orientation of magnetic field used in obtaining the data analyzed by Wolff, particularly for case A. Wolff's values for E_F and E_g should therefore remain unaffected by the refinements introduced in the present paper.

[17] D. Reneker, Phys. Rev. **115**, 303 (1959).
[18] H. Spector, Phys. Rev. **120**, 1261 (1960).
[19] P. A. Wolff (to be published).

Support for Wolff's values derives from new interband magnetoreflection experiments by Brown, Mavroides, Dresselhaus, and Lax,[20] who infer a value of 0.047 ± 0.003 ev for E_g, in excellent agreement with that of Wolff.

The new magnetoreflection experiments demonstrate that the departure from a parabolic E vs p relationship is significantly less than is implied by Eq. (4) with these values of E_F and E_g. Less departure from a parabolic E vs p relationship is implied by Eqs. (26), (29), or (39) than by Eq. (4) so that the experimental results are in accord with our present considerations.

Dresselhaus[21] has interpreted the transitions observed in the magnetoreflection experiments as indicating that case A holds, i.e., rotation plus reflection symmetry and only 3 ellipsoids. The tilt effect would provide a valuable check on this assignment.

Supposing that case A holds, the observation by Brown *et al.*[20] of equality of corresponding valence and conduction band cyclotron masses to within 10% for **H** along the binary and bisectrix axes would suggest that m_2' and m_2 do not differ by more than 20%.

Finally, Wolff's demonstration that models like these are adequate to explain the value of 100 reported for the dielectric constant by Boyle[19] lends general justification to such detailed studies as the present one.

The author is grateful to Dr. Wolff for a prepublication copy of his paper.

APPENDIX

The model as set up presupposes the existence of an ellipsoidal minimum in the conduction band at $\mathbf{k}_0$. This need not be the case; energy surfaces which are roughly ellipsoidal around $\mathbf{k}_0$ at energies near the Fermi energy are all that are required to fit the experimental data on Bi. Only a very minor modification of Eq. (38) suffices to introduce a point of contact with the valence band as the minimum of the conduction band, viz.,

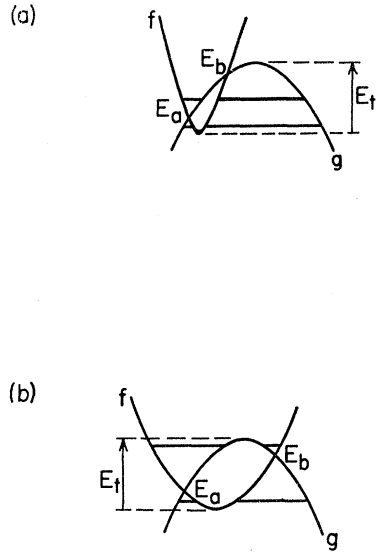

FIG. 5. p_2-dependence of f and g when no ellipsoidal minimum occurs in the conduction band (see Appendix). (a) One contact is between maximum and minimum. (b) Both contacts are between maximum and minimum.

$$\frac{p_1^2}{2m_1}+\frac{(p_3')^2}{2m_3}=\frac{1+\beta_3}{E_g}[E-f(p_2)][E-g(p_2)], \tag{52}$$

$$p_3'=p_3+\Delta p_3, \quad \Delta p_3=(m_3A_3/E_g)(E-f), \tag{53}$$

$$\beta_3=\tfrac{1}{2}m_3A_3^2/E_g =\tfrac{1}{4}(\langle 01|v_3|01\rangle)^2/|\langle 01|v_3|11\rangle|^2, \tag{54}$$

$$f(p_2)=p_2^2/2m_2, \tag{55}$$

$$g(p_2)=E_t-z(p_2+\Delta p_2)^2/2m_2, \quad z>0, \tag{56}$$

$$E_t=E_g\gamma/(1+\beta_3). \tag{57}$$

We now regard the parameter γ as disposable independently of Δp_2, which was not so in case $C(d)$ discussed previously. Thus, if

$$E_t>z(\Delta p_2)^2/2m_2, \tag{58}$$

holds, the minimum of f occurs within g as illustrated in Figs. 5(a) and (b). In either of these cases, the energy surfaces can have two hole pieces for $E<E_a$, one hole plus one electron piece for $E_a<E<E_b$, one hole plus one electron [Fig. 5(a)] or two electrons [Fig. 5(b)] for $E_b<E<E_t$, and one electron for $E>E_t$.

[20] R. N. Brown, J. G. Mavroides, M. S. Dresselhaus, and B. Lax, International Conference on the Fermi Surfaces of Metals, Cooperstown, 1960 (unpublished).

[21] G. F. Dresselhaus, International Conference on the Fermi Surfaces of Metals, Cooperstown, 1960 (unpublished).