Energy-Band Parameters and Relative Band-Edge Motions in the Bi–Sb Alloy System near the Semimetal–Semiconductor Transition*

LAWRENCE S. LERNER, KERMIT F. CUFF, LLOYD R. WILLIAMS

Lockheed Palo Alto Research Laboratory, Palo Alto, California

We report experimental information which elucidates important features of the $Bi_{1-x}Sb_x$ alloy system. By means of observations of the Shubnikov–de Haas effect on undoped and doped samples we have followed the transition from a band structure near the Fermi level which is essentially that of pure Bi to one which resembles that of pure Sb. Supplementary information has been obtained from transport measurements. In both Bi and Sb, the conduction band is at the L point in the Brillouin zone. A mirror-image band lies not far below. The upper L-point band is the conduction band throughout the alloy system. Our model of the transition involves two processes: (1) The descent of the T-point band (the valence band in Bi) relative to the L-point bands, a transition to six valence maxima at the nearby H points, and their subsequent rise past the Fermi level to become the valence bands in Sb; (2) the approach, crossing, and separation of the L-point bands. At 4.2°K, the descent of the T-point maximum below the L-point minimum and the L-point crossing both occur near $x=8\%$; consequently, there is a transition from semimetallic to direct-gap (L-point) semiconducting behavior. The most direct determination of the L-point gap is obtained from the nonparabolic behavior of the very small transverse mass components. Strong corroborative evidence is obtained from transport measurements in the temperature range 4°–30°K.

1. INTRODUCTION

It has been known for some time that the continuous system of alloys between the semimetals bismuth and antimony contains a semiconducting region at low temperatures.[1–3]

In bismuth, the conduction band is centered at the L point in the Brillouin zone, and the valence band is centered at the T point. The overlap is about 37 meV at 4.2°K.[4]

Upon addition of antimony, the T-point valence band edge descends relative to the L-point conduction band edge (the L_c band) and clears it at an antimony concentration C_{Sb} very close to 8 at.%.[1] This results in a transition from semimetallic to semiconducting behavior. A second transition, back to semimetallic behavior, occurs at higher Sb concentrations. It is apparently due to the rise of the H-point valence band edge relative to the L_c band edge; the H-point valence band edge is characteristic of pure Sb. In this paper, we concentrate our attention on the region near the first transition.

Lying about 15 meV below the L_c band edge is an approximately mirror-image valence band edge (the L_v band)[5]; they are separated by a band gap E_G. The L-point conduction band and its "mirror image" are characteristic also of the band structure near the Fermi surface in antimony, but E_G is much larger. The magnetoreflection data of Dresselhaus and Mavroides[6] indicate that the gap in Sb is about 100 meV. If the transition from Bi through the alloy system to Sb is a smooth one, either the L band edges must separate monotonically, or they must converge and subsequently separate, possibly crossing one another in the process. We set forth evidence that the latter process is the one which occurs, and that the closest approach (and probably crossing) takes place in the vicinity of $C_{Sb}=8\%$, very close to the first semimetal–semiconductor transition. In what follows, we present three arguments to support this point. The basic data are measurements made on a series of single-crystal samples with C_{Sb} in the region about 8%, whose Fermi levels have been adjusted by doping. In Sec. 2, the relative positions of the band edges and Fermi levels in these samples are obtained from Shubnikov–de Haas and galvanomagnetic measurements. In Sec. 3, supportive evidence for a small E_G is deduced from application of the nonparabolic band model to the effective masses derived from the temperature dependence of the Shubnikov–de Haas amplitudes. In Sec. 4, additional evidence for a small magnitude of E_G is derived from the temperature dependence of the resistivity and Hall coefficient in a semiconducting sample.

2. SHUBNIKOV–DE HAAS PERIODS AND HALL MEASUREMENTS

Figure 1 is a schematic representation of the relative motion of the band edges as a function of C_{Sb}. (The

* Work supported by the Lockheed Independent Research Fund.

[1] A. L. Jain, Phys. Rev. **114**, 1518 (1959).

[2] D. M. Brown and S. J. Silverman, Phys. Rev. **136**, A290 (1964).

[3] M. R. Ellett, R. B. Horst, L. R. Williams, and K. F. Cuff, Proc. Intern. Conf. Phys. Semiconductors, Kyoto, 1966; J. Phys. Soc. Japan Suppl. **21**, 666 (1966).

[4] L. S. Lerner, Phys. Rev. **127**, 1480 (1962).

[5] R. N. Brown, J. G. Mavroides, and B. Lax, Phys. Rev. **129**, 2055 (1963).

[6] M. S. Dresselhaus and J. G. Mavroides, Phys. Rev. Letters **14**, 259 (1966).

superposition of the T and L bands in Fig. 1 is meant to depict the relative energy levels only and not to imply denegeracy in **K** space.) The T–L_c band overlaps shown for pure Bi[4] and for the semimetallic 3.0% and 5.2% alloys[3] are taken from earlier work. It is evident that the T-point band edge descends in an approximately linear fashion with respect to the L_c band, at least in the semimetallic region, and that the semimetal–semiconductor transition should occur close to $C_{Sb}=8\%$. The rate of descent $\dot{T}$, in this linear approximation, is 4.5 meV/% Sb. We show in this section that the T-point band overlaps the L_c band at $C_{Sb}=8\%$, and lies below the L_v band at $C_{Sb}=9\%$. We conclude, as a consequence, that the value of E_G must be very small, since E_G is included in the small relative energy range traversed.

Table I summarizes the characteristics of four samples which are pertinent to the discussion below. The uncertainty in C_{Sb} is in each case $\pm 0.5\%$. Before discussing the results of measurements made on samples with C_{Sb} close to 8%, and consequently near the semimetal–semiconductor transition, we turn to a consideration of results from samples well into the semiconducting region, with C_{Sb} near 9%.

Sample 4, with 9.2% Sb, is moderately doped with Sn acceptors. Both L- and T-point holes are clearly evident and easily identifiable in the Shubnikov–de Haas oscillations; Fig. 2 is a plot of period vs magnetic field angle. For each individual piece of Fermi surface, the expression

$$E_F{}^0 = e\hbar/m_0 m_{ij}{}^* P_{ij} \qquad (1)$$

yields the Fermi energy, if the band is parabolic. Here P_{ij} is the period obtained with the magnetic field in the k direction, and $m_{ij}{}^*$ the associated cyclotron mass, obtained from the temperature dependence of the oscillation amplitude. For the T-point holes, Eq. (1) gives a T-point Fermi energy $E_T=10$ meV. The most

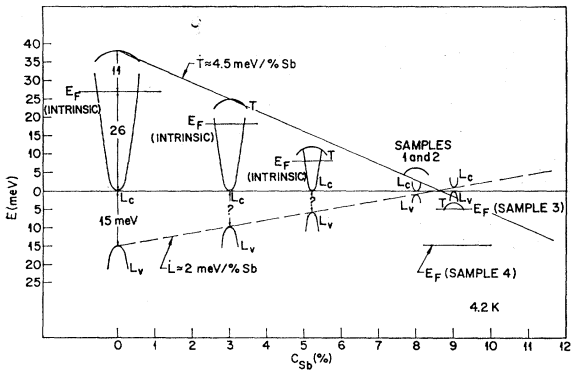

FIG. 1. Schematic representation of the relative energies of the L_c-, L_v-, and T-point band edges of Bi–Sb alloys as a function of Sb concentration C_{Sb}. The sloping lines give approximate values for $\dot{T}$, the rate of descent of the T point relative to the L_c point, and $\dot{L}$, the rate of ascent of the L_v point relative to the L_c point, in units of milli-electron volts per percent Sb.

TABLE I. Results of Shubnikov–de Haas and Hall measurements.

Sample	C_{Sb} (at.%)	Principal (bisectrix) L-point period (T^{-1})	L-point anisotropy $\kappa=m_2/(m_1m_3)^{1/2}$	Principal (trigonal) T-point period (T^{-1})	n_L (cm^{-3}) (type)	n_T (cm^{-3})	L-point $m_{13}{}^*$	T-point $m_{12}{}^*$	$E_F{}^0$ (L-point) (meV)	$E_F{}^0$ (T-point) (meV)
1	7.8	19.4	110	Not observed	2.0×10^{15} (n)	...	0.0017	...	3.6	...
2	8.0	27	140	See text	1.5×10^{15} (n)	$\approx10^{16}$	0.0016	...	2.7	≈1
3	9.0	30.5	70	Not observed	0.9×10^{15} (p)	$\approx10^{16}$	0.0014	...	2.6	≈1
4	9.2	2.65	170	0.12	4.8×10^{16} (p)	3.3×10^{17}	0.0051	0.098	8.5	10

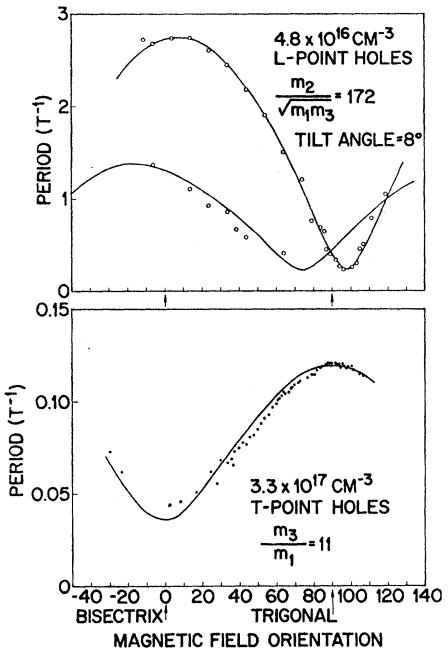

FIG. 2. Shubnikov–de Haas periods as a function of magnetic-field orientation for Sample 4, a 9.2% Sb alloy moderately doped with Sn acceptors. The upper curves are attributable to the L-point holes, and the lower curve to the T-point holes.

accurate results are derived from the principal (trigonal) period P_{12} and the associated value m_{12}* (see Table I). The latter is obtained from the temperature dependence of the oscillation amplitude. For the L-point carriers, the value of $E_L{}^0$ is best obtained from experimental measurements made with the magnetic field oriented along the bisectrix axis. To derive the Fermi energy for the L-point carriers, however, requires that we take the aparabolicity into consideration. The actual Fermi energy E_L depends upon $E_L{}^0$, as determined by means of Eq. (1), and upon the L-point gap energy E_G; to wit,

$$E_L = E_L{}^0 - \tfrac{1}{2}E_G + [(E_L{}^0)^2 + (\tfrac{1}{2}E_G)^2]^{1/2}. \qquad (2)$$

While we do not at this point know E_G, Eq. (2) indicates that the value of E_F is limited to the range $E_L{}^0 \leqq E_L \leqq 2E_L{}^0$. For Sample 4, we have 8.5 meV$\leqq E_L \leqq$17.0 meV. It follows that the T-point maximum lies between 1.5 meV above, and 7 meV below, the L_v maximum, depending upon the value of E_G. These limits are broader than necessary, and we narrow them later, when we discuss the results from Samples 3 and 4 together.

Sample 3 is substantially the same in antimony content as Sample 4, but it is less heavily p-type doped (see Table I). In Sample 3, only the L-point Shubnikov–de Haas oscillations are observed. Using Eq. (1) and Eq. (2), we obtain 2.6 meV$\leqq E_L \leqq$5.2 meV. The Fermi level cannot be above the T-point band edge, since the L_v carrier concentration, $n_L = 0.9\times10^{15}$ holes/cm^3, is so small that one can only account for the approximately 10^{16} holes/cm^3 added by doping if the Fermi level intersects the T-point band edge (see Table I). On the other hand, the absence of T-point oscillations indicates that the Fermi level cannot be more than about 1 meV below the T-point band edge. The T-point band edge consequently lies below the L_v band edge by approximately 1.6 meV$\leqq E_L - E_T \leqq$4.2 meV. These limits, while narrower than those obtained from Sample 4, can still be narrowed further, involving as they do the assumption of zero and infinite values of E_G. Any reasonable assumption as to the value of E_G restricts the position of the T-point band edge at $C_{Sb}=9\%$ to 3–4 meV below the L_v band edge.

We now show that at $C_{Sb}=8\%$ the T-point band edge overlaps the L_c minimum. Sample 2, with 8.0% Sb, is lightly Sn doped, but is n type at 4.2°K, as determined by the low-field Hall coefficient. (In this paper the term "Hall coefficient" refers always to the "large" Hall coefficient measured with current parallel to the binary axis and magnetic field parallel to the bisectrix axis.) The only carriers making a significant contribution to the low-field Hall coefficient are the L-point electrons, whose presence is also observed by means of the Shubnikov–de Haas oscillations. However, T-point holes must also be present. The sample contains $\approx10^{16}$ acceptors/cm^3; this would suffice to drain the L_c band of carriers in the absence of a T–L_c overlap. Since the density-of-states mass of the T-point band is $\approx0.2m_0$ at $C_{Sb}=8\%$, the Fermi level is about 1 meV below the band edge. The T-point holes, whose mobility is $\leqq10^{-2}$ that of the L-point electrons, make a negligible contribution to the low-field Hall coefficient. However, the Hall coefficient tends in a positive direction with increasing magnetic field and changes sign at about 1.7 T, the presence of T-point holes being manifested in the high-field Hall coefficient.

There is further evidence from Shubnikov–de Haas measurements that the T band overlaps the L_c band, i.e., that Sample 2 is semimetallic. The L-point oscillations are clearly observable through most of the binary plane. At magnetic field orientations more than about 15° from the trigonal axis, a second family of oscillations is observed. The angular dependence of the period and the magnitude of the cyclotron mass are strong evidence that the second family is attributable to T-point holes. Unfortunately, a quantitative analysis of the associated periods and effective masses is frustrated by three factors. First, the L-point electrons are in the extreme quantum limit at those orientations where the T-point oscillations are observed. Second, the putative T-point carriers are themselves observed near their quantum limit. Third, the oscillations appear near the upper limit of available magnetic field, 2.6 T. Nonetheless, the point which is of importance to our

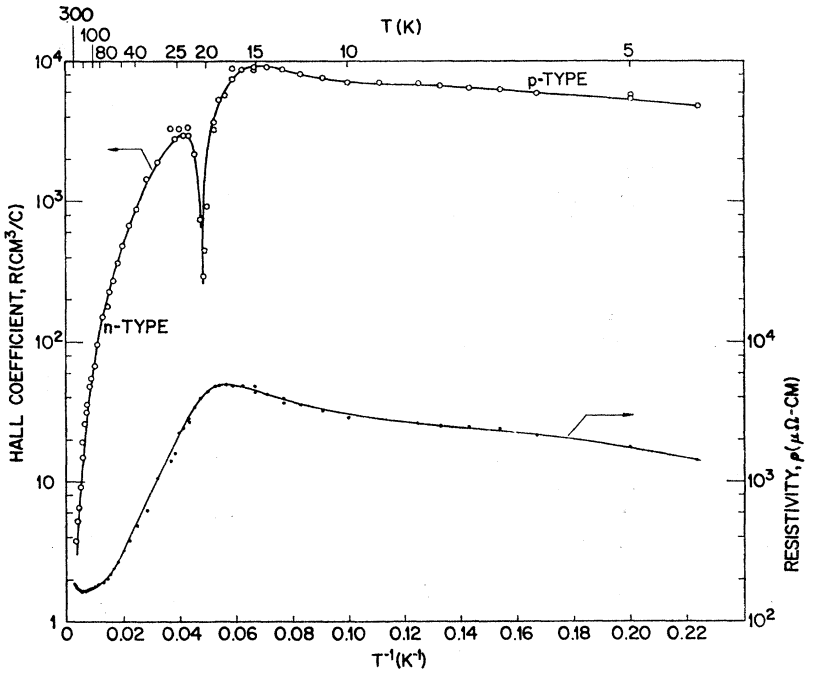

FIG. 3. Resistivity and weak-field Hall coefficient as a function of reciprocal temperature for Sample 3, a lightly acceptor-doped 9.0% Sb alloy. The Hall coefficient is measured with the current along the binary axis and the magnetic field along the bisectrix axis.

discussion, i.e., that there is a T–L_c overlap, appears to be well supported by the experimental evidence.

The appearance of the high-field oscillations under conditions of extreme quantum limit for the electrons is a novel one. It suggests that the depression of the Fermi level toward the bottom of the L_c band consequent on quantum-limit conditions has resulted in an increase in the hole Fermi energy, thus making possible the observation of the T-point oscillations.

Sample 1 has the same C_{Sb} as Sample 2 within experimental error; hence the T–L_c overlap should be about the same. The sample is not intentionally doped. The weak-field Hall coefficient is negative, and is 0.6 that of Sample 2. This, as well as the L-point oscillatory data, indicates that the Fermi level is somewhat higher than that of Sample 2. The Hall coefficient does not change sign up to a magnetic field intensity of 2.6 T, and no oscillations attributable to T-point holes are observed. The results for both Samples 1 and 2 indicate that a T–L_c overlap is present, but is not greater than 6 meV (see Table I).

We have now determined the approximate position of the T-point band edge relative to the L-point band edges in two sets of samples with known C_{Sb}. We can in principle employ our knowledge of $\dot{T}$ to determine the value of E_G since, as we have already noted, the gap E_G must be included in the excursion of the T-band edge. Unfortunately, the errors in C_{Sb} and the uncertainty involved in the extrapolation of $\dot{T}$ from the semimetallic region preclude a quantitative determination. We therefore limit ourselves to a discussion of reasonable bounds on the parameters of our model. Let us assume for the sake of argument that the extrapolation of $\dot{T}$ is justified. In that case, $E_G=0$ corresponds to an antimony concentration difference ΔC_{Sb} between the two pairs of samples of 2%; this is within the range of experimental error in the determination of C_{Sb}. On the other hand, to argue that E_G is unchanged from the 15-meV characteristic of pure Bi is to imply that $\Delta C_{Sb}\approx 5\%$, which is well outside the experimental limits. It is clear that, even allowing for errors in $\dot{T}$, E_G must be quite small, and in particular very much less than 15 meV.

3. THE L-POINT BAND GAP FROM EFFECTIVE-MASS CONSIDERATIONS

The L-point energy gap E_G can be determined from the nonparabolic behavior of the cyclotron mass component $m_{13}^*=(m_1^* m_3^*)^{1/2}$ of the L-point system.[3] For the dependence of the cyclotron mass m_c^* upon E_G and E_F, we have

$$m_c^*=(E_G+2E_F)/E_p. \tag{3}$$

For m_I^*, the inertial mass, we have

$$m_I^*=(E_G+E_F)/E_p. \tag{4}$$

In these expressions, E_p is the energy equivalent in electron volts of the $\mathbf{k}\cdot\mathbf{p}$ matrix element. From Eq. (1),

$$m_I^* E_F=e\hbar/m_0 P_{13}, \tag{5}$$

where P_{13} is the period associated with m_{13}^*. Eliminating

E_F from the above equations yields

$$E_G = E_p[m_c^{*2} - (4e\hbar/m_0 P_{13} E_p)]^{1/2}. \quad (6)$$

We have determined the value of E_p from the variation of m_c^{*2} with P_{13} in several sets of samples of nearly the same C_{Sb} doped to different Fermi levels. For both L_c and L_v bands, we obtain $E_p = 8.2$ eV quite accurately. For the 8% and 9% Sb compositions discussed here, the value of E_G determined in this manner is at most a few milli-electron volts. While this determination does not provide an accurate value of E_G, the results are consistent with those obtained in Sec. 2.

One additional feature of the nonparabolic behavior of the L-point bands is worth noting. The mass anisotropy $\kappa \equiv m_2^*/(m_1^* m_3^*)^{1/2}$ of both L_c and L_v bands at the low Fermi energies discussed here is relatively insensitive to variation in the Fermi energy (see Table I). Thus m_2, the large L-point ellipsoid mass, appears to vary with Fermi energy in a manner similar to that of the small-mass components m_1 and m_3. This indicates that at these small Fermi energies m_2 is determined largely by a relatively weak $\mathbf{k}\cdot\mathbf{p}$ interaction between L_c and L_v. Hence our results in this specific case tend to favor the Lax model[7] over the Cohen model.[8]

4. TEMPERATURE DEPENDENCE OF RESISTIVITY AND HALL COEFFICIENT

Figure 3 shows the temperature dependence of the weak-field "large" Hall coefficient R and of the binary resistivity for a sample with $C_{Sb} = 9.0\%$ and p-type doping very nearly equal to that of Sample 3. A detailed analysis of the transport coefficients in the range $4.2°K \leqq T \leqq 30°K$ shows that $E_G \leqq 3$ meV. We give here a simplified but adequate interpretation of the data which illustrates the smallness of E_G. The argument is based on the very low temperature at which the Hall coefficient changes sign to n type (21°K). We note first that at 4.2°K and below, the T-point band is approximately 4 meV below L_v. Even if there were no relative band motion as a function of temperature, the Fermi level would rise with increasing temperature; this effect accounts for a large part of the modest increase in weak-field Hall coefficient as the temperature increases from 4.2° to 15°K. In addition, the T-point maximum itself ascends slowly relative to the L_v maximum with increasing temperature, adding to the rise of the Fermi level. The two effects together, however, do not suffice to raise the Fermi level above the L_v band edge at $T = 21°K$. Now the ratio of the mobility of the L-point electrons to that of the L-point holes is approximately 2 (the T-point hole mobility is negligible). Hence the sign change in the low-field Hall coefficient at 21°K indicates that E_G cannot be significantly greater than kT; a reasonable upper limit is 3 meV, a value in close agreement with those obtained above by other means.

5. SUMMARY AND CONCLUSIONS

The semimetal–semiconductor transition in the alloy system Bi–Sb occurs when the T-point valence band descends below the L-point conduction band. The antimony concentration at which this takes place is close to 8 at.%. Occurring simultaneously with the steady descent of the T-point band edge is a shift in the relative positions of the L_c and L_v band edges. Such a shift is logically necessary if one assumes a smooth transition through the alloy system from a band structure like that of pure Bi (for which $E_G = 15$ meV) to a band structure like that of pure Sb (for which $E_G = 100$ meV). We have followed the L-point shift up to approximately 9% Sb, using the Shubnikov–de Haas effect, supplemented by the Hall effect, as a tool for the measurement of effective masses, carrier anisotropies, and Fermi energies. Our experimental results lead to three arguments which indicate that E_G decreases upon addition of Sb to Bi, and reaches a very small value in the vicinity of 8% Sb. The principal argument supporting this conclusion is that the T band overlaps the L_c band by about 6 meV for alloys with 8% Sb, but lies below the L_v band edge by 3–4 meV for 9% alloys. An upper limit of approximately 10 meV can be placed upon the energy range traversed by the T band; this range must include the gap E_G.

Supportive arguments are made on the basis of the effective masses of the L-point carriers in samples with C_{Sb} near 8% and with small Fermi energies, and on the basis of the galvanomagnetic coefficients as functions of temperature in a p-type doped 9% sample.

These results are consistent with the idea that the transition of E_G from pure Bi through the alloy system to pure Sb is accomplished by a monotonic relative shift of the L-point band edges. The small value of E_G near 8% and the monotonic shift together imply that the band edges cross and exchange roles as conduction and valence bands at a value of C_{Sb} in the vicinity of 8%.

An ancillary conclusion, based on the insensitivity of the anisotropy of the L-point oscillations to Fermi energy, is that the Lax model appears to be a more appropriate description of the band edges and their interactions than the Cohen model for the small Fermi energies encountered in this work.

ACKNOWLEDGMENTS

We thank M. K. Stafford, G. E. Huling, P. L. Jeffreys, and L. E. Burge for their aid in sample preparation and Dr. R. B. Horst for many valuable discussions.

[7] B. Lax and J. G. Mavroides, in *Advances in Solid State Physics*, F. Seitz and D. Turnbull, Eds. (Academic Press Inc., New York, 1960), Vol. 11.

[8] M. H. Cohen, Phys. Rev. **121**, 387 (1961).

Discussion of Lerner, Cuff, and Williams' Paper

R. S. ALLGAIER (Naval Ordnance Lab.): The crossing of the two bands at L is also seen in the corresponding alloys of cubically symmetric IV–VI compounds; for example, in alloys of PbTe and SnTe. From the cubic point of view, L and T are the same and wouldn't you therefore expect to see the same crossing of bands at T? If this isn't occurring, why not?

L. S. LERNER: The existing experimental data do not seem to support the hypothesis of T-point crossing. First, we have observed Shubnikov–de Haas electron oscillations in samples with Sb concentration up to about 30%, and the symmetry of the conduction band remains that of the L point throughout. Second, our observations on alloy composition up to 10% show a monotonic increase in the T-point transverse hole mass over that for pure Bi; this would suggest that the T-point band edges are separating rather than converging. Now we know that the lower T-point edge is descending as the L-point edges converge; a naive appeal to symmetry would suggest that the upper T-point edge is ascending, in agreement with the available experimental information. I think that, considering the very small energies involved, too many factors can influence the relative band motions to permit reliable extrapolation from cubic to rhombohedral (real) bismuth and Bi–Sb.

W. PAUL (Harvard University): Isn't this the case with a small L-point gap, the sort Halperin classified as one where you had bands of different symmetry but where there was a finite matrix element, and so this is a case where you certainly would not see any excitonic insulator transition?

L. S. LERNER: Yes. According to our understanding, a phase transition should not occur for the direct (L-point) gap. The possibility remains open for the indirect T–L gap in the alloy system near $C_{Sb}=8\%$.

W. KOHN (University of California, San Diego): Rice, Jérome, and I checked that system for a possible excitonic transition and it's quite hopeless because of the large dielectric constant, small masses, and, of course, since it is necessarily an alloy (unlike the case of grey tin that was discussed here), you have a rather short scattering time. There are many orders of magnitude from the possibility of realizing the excitonic transition.

L. S. LERNER: The scattering is not as large as one might expect. In fact, we see L-point mobility components in the alloys near $C_{Sb}=8\%$ which have been as high as 3×10^{7} cm^2/V·sec. This corresponds to $\tau\approx3\times10^{-11}$ sec. I might note in addition that samples have been prepared with carrier concentrations as low as 10^{14} cm^{-3}.

W. KOHN: As Dr. Sherrington said, the more serious effect of impurities is the lack of compensation rather than the scattering. He'll show you our calculations. We consider it not within the realm of possibility.