Comment on "Semimetal-to-semiconductor transition in bismuth thin films"

H. T. Chu
Department of Physics, The University of Akron, Akron, Ohio 44325
(Received 25 April 1994)

In a recent paper, C. A. Hoffman *et al.* reported field- and temperature-dependent magnetotransport measurements on bismuth thin films and concluded that the semimetal-semiconductor transition in ultrathin films of bismuth, caused by the quantum size effect, was confirmed based on the temperature-dependent behavior of the minority electrons. In this Comment, we discuss the fact that the existing experimental data are not sufficient to establish the picture that the energy overlap of the conduction and valence bands is replaced by an energy gap in sufficiently thin films. Two equally satisfactory pictures are provided for the interpretation of thermal excitations of the minority electrons. The deviation from the equality of electron and hole concentrations in bismuth thin films is also discussed.

In a recent paper,[1] Hoffman *et al.* reported a fine and interesting experimental work exploring the statistical properties of electrons and holes in thin films of pure bismuth. However, the conclusion that the semimetal-semiconductor transition, caused by the quantum size effect, has been confirmed remains, in our opinion, debatable at this point. We would like first to briefly summarize what appears to be the major conclusions drawn from this work.[1]

(1) The semimetal-semiconductor transition occurs in ultrathin bismuth films (thickness <280 Å), because minority electrons were not observed at temperatures <200 K and the electron density remained low compared with that in thicker films even at higher temperatures, e.g., 300 K. The temperature dependence of the carrier densities seems to suggest an energy gap in the 200-Å film [Eq. (2) in Ref. 1].

(2) Holes were found to be the majority carriers in thin films (of bismuth) and the density increased with decreasing film thickness [Eq. (1) in Ref. 1]. The majority carrier concentration would be insensitive to the introduction of an energy gap. The hole concentration remained undiminished even at $T\rightarrow 0$ and thus one should expect no abrupt changes in the low-temperature resistivity and/or Hall coefficient at the transition thickness.

(3) The numbers of the two charge carriers were found to be unequal and the density of the majority holes was generally larger than that of the minority electrons in thin films.

We would like to suggest a few points for discussion.

(1) To interpret the fact that minority electrons were not observed in ultrathin films at low temperatures and their thermal excitations at higher temperatures were somewhat proportional to $\exp(-E_g/kT)$, where E_g is a positive quantity, it is not necessary to introduce the transition from an energy overlap between the conduction and valence bands to an energy gap. We would like to start the discussion with checking over some theoretical background.

The commonly used boundary condition for thin films is the vanishing-wave-function condition. Since outside the boundary both the probability and the current are zero, the vanishing-wave-function boundary condition which ensures zero probability and current at the boundary seems certainly a good choice. However, if the gradient of the wave function vanishes at the boundary, the current is also zero at the boundary even though the wave function may not vanish. Zero current at the boundary is consistent with zero probability outside the boundary. Neither the vanishing-wave-function nor the vanishing-gradient boundary condition can provide continuities for both the wave function and its gradient at the boundary. Paskin and Singh[2] have concluded that "while the wave function vanishing at the boundary might be expected to be a good approximation for an infinite vacuum-metal potential barrier, the vanishing gradient might be expected to be a better approximation for a small barrier or at the rough surfaces that usually occur in thin films." Thus, it might be a good idea that both results derived from the different boundary conditions should be given a consideration when experimental results are available for comparison with theoretical calculations in thin films.

For holes in bismuth thin films normal to the trigonal axis, the energy spectrum may be given by

$$E=\frac{\hbar^2}{2M_1}(k_1^2+k_2^2)+\frac{\hbar^2}{2M_3}\left[\frac{\pi}{d}n\right]^2, \tag{1}$$

where M_1 and M_3 are effective masses,[3] d is the film thickness, and $n=1,2,\ldots$ or $0,1,2,\ldots$ depending on the application of the vanishing-wave-function or the vanishing gradient boundary conditions, respectively. At $T=0$, the hole concentration in the valence band can be readily evaluated and given by[4]

$$N_h=(\pi\hbar^2 d)^{-1}M_1\sum_n\left[E_F^h-\frac{\hbar^2}{2M_3}\left[\frac{\pi}{d}n\right]^2\right], \tag{2}$$

where the summation has a cap on n such that the quantity in the square brackets remains non-negative, and E_F^h is the hole Fermi energy which is the energy difference between the Fermi level and the valence-band top. Combining with Eq. (1) in Ref. 1,

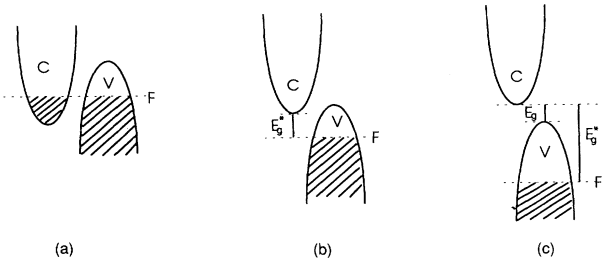

FIG. 1. Band edges and Fermi level in Bi thin films before and after an energy gap occurs (vanishing-wave-function-boundary condition). At $T=0$, only the shaded areas are occupied by electrons. C is the conduction band, V is the valence band, and F is the Fermi level.

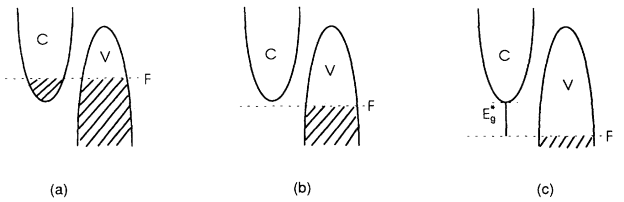

FIG. 2. Locations of the Fermi level with respect to stationary band edges (vanishing-gradient-boundary condition). The band overlap remains unchanged while the film thickness is reduced from (a) through (c). At $T=0$, only the shaded areas are occupied by electrons. C is the conduction band, V is the valence band, and F is the Fermi level.

$$N_h = N_i + N_s/d \ , \tag{3}$$

where N_i and N_S are positive constants, it can be seen that the hole Fermi energy E_F^h would increase with decreasing film thickness. As d becomes small, the summation in Eq. (2) is expected to include only the lowest integers of n and E_F^h would be roughly inversely proportional to d^2. This means the Fermi level would move continuously downward away from the valence-band top if the film thickness is continuously reduced.

Thus, assuming the vanishing-wave-function boundary condition, the relative positions of the conduction-band bottom, valence-band top, and the Fermi level before and after an energy gap occurs can be sketched as in Fig. 1. In thicker films, the energy overlap remains, this is sketched in Fig. 1(a). In thinner films, an energy gap E_g replaces the overlap, and there is no electron in the conduction band at $T=0$. At higher temperatures, thermal excitations of electrons from the valence band to the conduction band depend on the effective energy gap E_g^* instead of the gap E_g. This is sketched in Fig. 1(c). Figure 1(b) is an intermediate case where E_g^* exists even before E_g occurs.

On the other hand, applying the vanishing-gradient boundary condition, there would be a stationary lowest energy level in both the conduction and the valence band. Based on the fact that the majority hole density and thus the hole Fermi energy (the distance between the Fermi level and the valence band top) increase with decreasing film thickness, Figs. 2(a), 2(b), and 2(c), respectively, show the location of the Fermi level in a thicker film, an intermediate film, and a thinner film. While the energy overlap remains intact, there are no conduction electrons at $T=0$ in thinner films and the thermal excitation of an electron at higher temperatures would depend on E_g^*, the effective energy gap, as shown in Fig. 2(c).

It can be seen from the foregoing discussion that either Fig. 1 or Fig. 2 would be able to interpret the fact that the majority carrier (holes) concentration increases with decreasing film thickness and the minority electrons depend on thermal excitations in ultrafilm films. The thermally excited electron concentrations, however, should be proportional to $\exp[-E_g^*/kT]$ applying Fermi-Dirac statistics and assuming $E_g^* \gg kT$. The effective energy gap E_g^* is likely to increase with decreasing film thickness and starts to emerge as soon as the Fermi level moves down crossing the bottom edge of conduction band. However, the physical differences between the two pictures are also obvious. For instance, the effective energy gap E_g^*, which is the determining factor for thermal excitations of the minority electrons, can be quite different in value from the two pictures. From Fig. 1(c), $E_g^* = E_F^h + E_g$; and from Fig. 2(c), $E_g^* = E_F^h - E_0$, where E_0 is the energy overlap. Also, if the film is used as one of the electrodes in a tunneling junction, current can be quite different based on the different band pictures Fig. 1 or Fig. 2.

(2) The use is questionable of Eq. (2) in Ref. 1, the "law of mass action," where the Fermi energy or chemical potential once used in the evaluations of the electron and hole concentrations has been canceled in the product of these two based on that the Fermi-Dirac statistics can be approximated by the classical statistics. Assuming either the electrons or the holes are the majority carriers, the Fermi level at low temperatures would be either high in the conduction band or low in the valence band, and consequently either electrons or holes cannot be approximately described by the Boltzmann statistics. Since Fig. 3 in Ref. 1 was plotted using Eq. (2) (in Ref. 1), its physical significance may also be questioned. By the way, the slope of the 300-Å film in Fig. 3 (Ref. 1) looks more like a nonzero one.

(3) Assuming that the energy overlap is replaced by an energy gap in an ultrathin film of bismuth and that one of the charge carriers is the majority and remains undiminished as $T \to 0$, would semiconductor be the more appropriate name for the film? Since at $T \to 0$, there are free charge carriers in at least one band, of which the density is likely in the range[1,5] typically referred to semimetals and is nearly temperature insensitive (Fig. 2 in Ref. 1), it may be equally appropriate to retain the name semimetal.

(4) Both the authors of Ref. 1 and Komnik *et al.*[5] made direct measurements of the galvanomagnetic properties. The carrier concentrations and the mobilities, however, were calculated using theoretical equations that relate the resistivity, Hall coefficient, and magnetoresistance to the concentrations and mobilities. The carrier concentrations were not directly measured and the

theoretical equations are quite general and may not be quite suitable for the anisotropic band structure in bismuth. More direct and indirect experimental measurements plus more prudent analyses may still be needed before a confirmation of the semimetal-semiconductor transition can be made.

(5) The deviation from the equality of electrons and holes is expected in thin films due to the surface effect and crystal defects. But which carrier is the majority and what is the predominance of the majority over the minority? The authors of Ref. 1 made calculations based on their magnetotransport data and stated that holes were the majority carriers and Eq. (1) (Ref. 1) gave the quantitative variation of the holes with decreasing film thickness. The authors also stated that Komnik *et al.*[5] directly determined that the electrons were the majority carriers with very similar thickness dependence of the concentration. This was mistakenly misunderstood. Actually, Komnik *et al.*[5] made calculations of the carrier concentrations by assuming these two were equal (or assuming equal mobilities which gave rise to similar results). While being equal, both concentrations assumed a thickness dependence given by an equation identical to Eq. (1) in Ref. 1. They[5] in fact emphasized that their calculations showed no ground for assuming unequal concentrations. Only when they[5] tried to explain why no semimetal-semiconductor transition had been observed, they made the assumption that the concentration increasing with decreasing film thickness must be the result of surface effect and the increased concentration hindered the occurrence of the transition. In a previous work,[6] the Hall coefficient in thin films of bismuth was found generally negative at low temperatures. This means the electron contribution to the conductivity is predominant. Thus, it is our belief that the surface effect and/or crystal defects may vary in films and depend on many factors, like the fabrication process including the deposition method, choice of the substrate, etc. It is possible that in one case the electrons are the majority while in another case the holes are the majority. If the above thought is valid, it might be possible to fabricate a bismuth film with minimum charge neutrality violation. Should such films be materialized, the semimetal-semiconductor transition would then be equivalent to the observation of an abrupt change in the conductivity at low temperatures.

---

[1]C. A. Hoffman, *et al.*, Phys. Rev. B **48**, 11 431 (1993).

[2]A. Paskin and A. D. Singh, Phys. Rev. **140A**, 1965 (1965).

[3]G. E. Smith, G. A. Baraff, and J. M. Rowell, Phys. Rev. **135A**, 1118 (1964). In bulk bismuth, $M_1=0.64\times10^{-1}m_0$ and $M_3=0.69m_0$, m_0 being the free-electron mass.

[4]H. T. Chu and W. Zhang, J. Phys. Chem. Solids **53**, 1059 (1992).

[5]Yu. F. Komnik *et al.*, Zh. Eksp. Teor. Fiz. **60**, 669 (1971) [Sov. Phys. JETP **33**, 364 (1971)]; Fiz. Nizk. Temp. **1**, 104 (1975) [Sov. J. Low Temp. Phys. **1**, 51 (1975).

[6]H. T. Chu *et al.*, Phys. Rev. B **45**, 11 233 (1992).