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

C. A. Hoffman, J. R. Meyer, and F. J. Bartoli
Code 5613, Naval Research Laboratory, Washington, D.C. 20375

A. Di Venere, X. J. Yi, C. L. Hou, H. C. Wang, J. B. Ketterson, and G. K. Wong
Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208
(Received 2 June 1994)

The suppression of intrinsic carrier generation does not in itself constitute proof that a semimetal-to-semiconductor transition has occurred. However, we present a realistic statistical analysis that demonstrates that the Bi film data do in fact imply the opening of a ≈ 56-meV energy gap in the thinnest sample (200 Å) studied. The standard vanishing-wave-function model accounts for the experimental observations, whereas the alternative vanishing-gradient model yields results that are inherently inconsistent with the data.

In his Comment[1] on our paper,[2] Chu correctly points out that the experimental finding of suppressed intrinsic carrier generation does not in itself constitute proof that we have observed a semimetal-to-semiconductor (SMSC) transition in a series of Bi thin films with varying thicknesses (d). Since the increase of the majority hole concentration with decreasing d would have necessarily reduced the density of minority electrons at a given temperature, our claim must ultimately rest on whether the *degree* of suppression is too large to be explained by the shift of the Fermi level with doping alone. We, therefore, present here a comparison of the data with a simple but realistic quantitative modeling of the free-carrier statistical properties. The results more rigorously justify that a SMSC transition has in fact occurred.

Our analysis is somewhat more general than Eqs. (1) and (2) of Ref. 1, in that we do not rely on the low-temperature limit and also include the effects of nonparabolicity. For a thin film in which the energy levels are quantized along the $\hat{z}$ axis, the electron and hole concentrations are given by

$$n=\sum_{j=1}^{\infty}\frac{2\gamma_n}{(2\pi)^2 d}\int_0^{\infty}d^2k f_0(E_{jn}(k)) \tag{1}$$

and

$$p=\sum_{j=1}^{\infty}\frac{2}{(2\pi)^2 d}\int_0^{\infty}d^2k f_0(E_{jp}(k))\ , \tag{2}$$

where the summation j is over quantized subbands, $\gamma_n=3$ is the number of degenerate electron valleys at the L point,

$$f_0(E_{ji})=\frac{1}{\exp[(E_{ji}-E_{Fi})/k_BT]+1}\ , \tag{3}$$

is the Fermi distribution function, and E_{Fi} is the electron or hole Fermi energy. Accounting for the nonparabolicity of the dispersion relations,[3] the energy of level j is given by

$$E_{ji}(k)=\Delta_{ji}+\frac{(E_{gdi}^2+2E_{gdi}\hbar^2k^2/m_{1i})^{1/2}-E_{gdi}}{2}\ , \tag{4}$$

where E_{gdi} is the direct energy gap for carrier i (i.e., the L-point gap for electrons[4] and the T-point gap for holes[5]), m_{1i} is the electron or hole effective mass in the plane,

$$\Delta_{ji}=\frac{\hbar^2\pi^2(j-\delta_G)^2}{2m_{3i}d^2} \tag{5}$$

is the quantization energy from the particle-in-a-box model, and m_{3i} is the effective mass along the growth (trigonal) axis. The quantity δ_G is zero if one imposes the usual condition that the wave functions should vanish at the boundaries of the quantum well, while it is unity for the alternative model in which the gradients of the wave functions vanish at the boundaries.[1] Thus quantum confinement does not induce an energy gap in the latter case. However, Ref. 1 notes that if p-type doping shifts the Fermi energy beyond the conduction-band minimum, one can nonetheless obtain an "effective energy gap."

If we define E_g to be the indirect energy gap between the conduction-band minimum at the L point and the valence-band maximum at the T point, the electron and hole Fermi energies satisfy the relation: $E_{Fp}=-E_{Fn}-E_g$, where

$$E_g=E_{g0}+\Delta_{1n}+\Delta_{1p} \tag{6}$$

and E_{g0} is the indirect energy gap in the absence of quantum confinement, which we take to be independent of temperature for lack of experimental information. Note that $E_g=E_{g0}$ in the vanishing-gradient model. Finally, we invoke the charge neutrality condition

$$p-n=N_A-N_D\approx p_s/d\ , \tag{7}$$

where N_D and N_A are the donor and acceptor concentrations, and the observed sheet density of $p_s\approx 2.75\times 10^{12}$ cm^{-2} is presumably associated with the ionization of p-type surface or interface states[6] (there were no exposed Bi surfaces, since the samples were capped with CdTe). The net doping levels for each sample were determined from the low-temperature carrier concentrations.

Despite some simplification of the anisotropic Bi band

structure, Eqs. (1)–(7) provide a realistic basis for modeling the main features of the experimental data discussed in Ref. 2. We will treat the thickest sample as a reference, since the quantum confinement effects under investigation are negligible at $d=5000$ Å (for that film, a three-dimensional (3D) density of states was employed rather than the 2D form specified above). In order to assure a good fit to the experimental $n(T)$ and $p(T)$,[7] E_{g0} and the 3D density-of-states mass product $(m_{1n}^{2/3}m_{3n}^{1/3}m_{1p}^{2/3}m_{3p}^{1/3})$ were allowed to vary slightly from the values usually quoted for bulk Bi.[8] By using $E_{g0}\approx 28$ meV (rather than 38 meV) and increasing all masses by a factor of 1.7, we calculate hole and electron densities that accurately reproduce the data as represented by the open and closed boxed in Figs. 1 and 2, respectively. The same E_{g0} was then used in modeling the data for the thinner samples.

We next fix the 2D density-of-states mass product $(m_{1n}m_{1p})$ by requiring n (300 K) in the 400-Å film to reproduce its experimental value. This is another region for which quantum confinement is relatively unimportant, since $k_BT>\Delta_{1n},\Delta_{1p}$. One final parameter is then used to account for the magnitude of the confinement shift of E_g. The growth-direction masses appearing in Eq. (5), m_{3n} and m_{3p}, have been varied by the same constant factor[9] until the calculated n (200 K) in the thinnest sample agrees with the experimental value. For the vanishing-wave-function model, this fitting procedure yields the solid curves in Figs. 1 and 2. Note that the calculation reliably reproduces all qualitative trends of the data, and yields relatively good quantitative agreement. For $d=200$ Å, the confinement shift derived from the fit corresponds to the presence of a positive energy gap of 56 meV. This value is in fair agreement with the result $E_g\approx 40$ meV obtained from the highly simplified analysis (based on the "law of mass action") discussed in Ref. 2, and clearly indicates that a SMSC transition has occurred. On the other hand, any attempt to fit the data with parameters giving $E_g(200\text{ Å})\leq 0$ leads to large discrepancies.

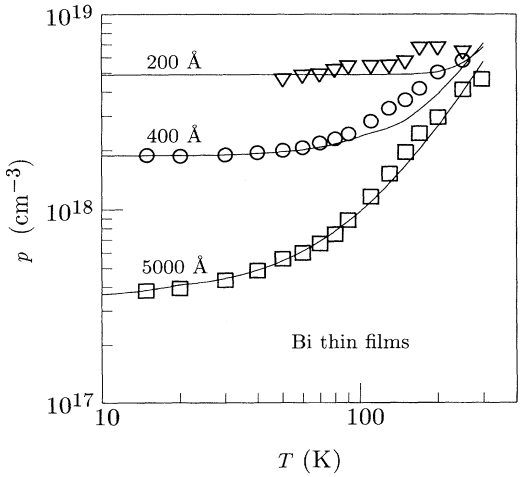

FIG. 1. Experimental (Ref. 2) (points) hole densities vs temperature compared to calculations based on the vanishing-wave-function model (curves), for Bi thin films with three different thicknesses.

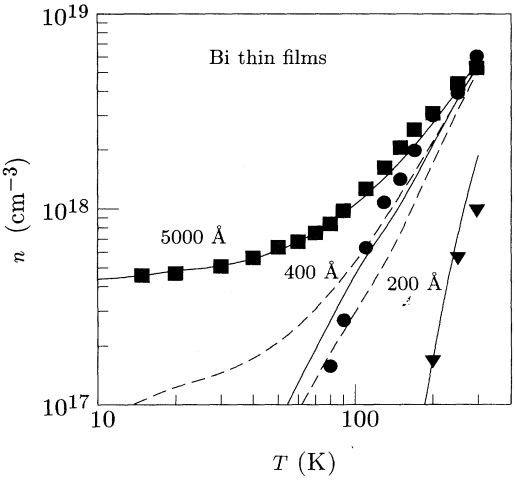

FIG. 2. Experimental (Ref. 2) (points) and calculated electron densities vs temperature for the same three films. The solid curves were obtained using the vanishing-wave-function model (with an energy gap of 56 meV in the 200-Å sample), while the dashed curves are results from the vanishing gradient model for the same two thicknesses.

Having demonstrated that the vanishing-wave-function model can account for the observed dependences of the carrier concentrations on temperature, we now consider the implications of the vanishing-gradient model. Again requiring that n(300 K) reproduce its experimental value in the 400-Å film, we obtain the $n(T)$ dependences shown as the dashed curves in Fig. 2. While we have taken m_{3n} and m_{3p} from the fit discussed above (which are comparable to bulk-Bi literature values) for definiteness, there are no values of those parameters that remove the large discrepancies with the data. No gap is induced by quantum confinement when the vanishing-gradient boundary condition is employed, and the "effective gap" due to doping is not nearly large enough to significantly suppress the minority electron concentration at temperatures above ≈ 100 K. The valence-band density of states is great enough that even for the net acceptor concentration of 4.9×10^{18} cm^{-3} in the 200-Å sample, the Fermi energy lies within 15 meV of the bottom of the conduction band. Since most of the electrons occupy the $j=1$ subband when $d=200$ Å, increasing the subband splitting by further decreasing m_{3n} and m_{3p} does not improve the poor fit.[10] Thus there is no way to reconcile the minority electron concentrations calculated within the vanishing-gradient model with the thin-film experimental data.

We conclude on the basis of a realistic statistical analysis that whereas the vanishing-gradient model gives results that are inherently inconsistent with the data from Ref. 2, the vanishing-wave-function model quantitatively accounts for the experimental $n(T)$ and $p(T)$ vs d. In or-

der to obtain agreement, the parameters must be chosen such that an energy gap of 56 ± 10 meV is present in the thinnest film studied (200 Å).[11] The analysis thus clearly confirms our contention in Ref. 2 that a confinement-induced SMSC transition has been observed. Reference 1 notes that a more sensitive study of quantum confinement effects could be carried out on samples with lower net doping levels. In fact, we plan to investigate compensated Bi films grown onto CdTe buffer layers that are intentionally n-doped near the top surface.

[1]H. T. Chu, preceding Comment, Phys. Rev. B **51**, 5532 (1995).
[2]C. A. Hoffman, J. R. Meyer, F. J. Bartoli, A. DiVenere, X. J. Yi, C. L. Hou, H. C. Wang, J. B. Ketterson, and G. K. Wong, Phys. Rev. B **48**, 11 431 (1993).
[3]B. Lax and J. G. Mavroides, *Advances in Solid State Physics* (Academic, New York, 1960).
[4]M. P. Vecchi and M. S. Dresselhaus, Phys. Rev. B **9**, 3257 (1974).
[5]J. P. Omaggio, J. R. Meyer, C. A. Hoffman, A. DiVenere, X. J. Yi, C. L. Hou, H. C. Wang, J. B. Ketterson, G. K. Wong, and J. P. Heremans, Phys. Rev. B **48**, 11 439 (1993).
[6]Reference 1 points out that in Ref. 2, we mistakenly stated that the Bi thin films discussed in Yu. F. Komnik and V. V. Andrievskii, Fiz. Nizk. Temp. **1**, 104 (1975) [Sov. J. Low Temp. Phys. **1**, 51 (1975)] were n type whereas ours are p type. In fact, the observation of positiive Hall coefficients at small d and low T (Fig. 4) implies that acceptorlike surface or interface defects also dominated in those films.
[7]We noted in Ref. 2 that the intrinsic carrier concentration determined for this sample was slightly larger than the bulk value, a trend also observed in previous electrical studies of films with intermediate thickness.
[8]J.-P. Issi, Aust. J. Phys. **32**, 585 (1979).
[9]Due to complexities in the anisotropic dispersion relations for Bi, one cannot define a single m_{3n} which is strictly appropriate for both confinement and density of states. However, in all cases the effective masses obtained from fitting our simplified model to the thin-film data agree to within a factor of 2 with values reported for bulk Bi (and there is often significant variation in the bulk values obtained by different methods).
[10]Actually, were it not for the increase of $N_A - N_D$ at small d, the electron densities in the 200-Å film would *exceed* those in the 400-Å film. In the vanishing-gradient model, decreasing d leads to a larger density of states at the band edge without an accompanying penalty in confinement energy.
[11]Experimental uncertainties and the use of a simplified anisotropic band structure in the data analysis may contribute an additional ±10 meV.