Space-Charge Effects on Electron Tunneling

D. J. BENDANIEL AND C. B. DUKE
General Electric Research and Development Center, Schenectady, New York
(Received 2 June 1966)

The one-electron (Bethe-Sommerfeld) model of electron tunneling is formulated to describe tunneling when the curvature (electron mass) and centroid of the one-electron constant-energy surfaces vary across the junction. The conductance for an abrupt GaAs *p-n* tunnel diode is calculated and shown to exhibit minima near zero bias for highly asymmetrical doping ratios. The conductance of metal-oxide-semimetal (M-O-SM) tunnel junctions is evaluated both with and without the inclusion of space-charge effects and of surface states. All calculations are performed using solvable models for which the WKBJ approximation is not imposed. Neither the removal of the WKBJ approximation nor the space-charge effects give rise to maxima in the conductance of the M-O-SM junctions near a band edge.

I. INTRODUCTION

ALTHOUGH electron tunneling in metal-insulator-metal (M-I-M) junctions has been extensively studied,[1] junctions in which one or more of the components is a semimetal[2] or degenerate semiconductor have been systematically examined primarily within the context of *p-n* junctions.[3] In this paper we consider the modifications of the tunneling current in semimetal and *p-n* junctions caused by the dependence of the shape of the space-charge-induced barrier on the applied bias. We construct sufficiently simple models of the junctions to yield potential barriers for which the one-electron Schrödinger equation can be solved analytically. This procedure permits us to calculate the tunneling current without using the WKBJ approximation, and thereby achieve a more accurate description of the tunneling near band edges.[1,3] We also generalize the Bethe-Sommerfeld[4] model to incorporate the cases of differing masses and differing locations of the electron (hole) ellipsoids in the various component elements of the junction. The generalized model is employed to calculate the differential conductance due to tunneling current through the junctions.

In the case of *p-n* junctions, several authors[5] have used the WKBJ approximation to estimate corrections to the Zener-Keldysh-Kane uniform-field model.[3] They did not find the doping-sensitive structure in the differential conductance which we will demonstrate below, and Nathan concluded that the corrections generally reduced the agreement between the calculations and (room temperature) experimental data. Ivanchik[6] also has developed an effective-mass theory of tunneling in *p-n* junctions. However, he limited his attention to symmetric diodes and used the WKBJ approximation.

The theoretical discussion of metal-oxide-semimetal (M-O-SM) junctions given herein is, to the authors' knowledge, the first that has been presented. A treatment of metal-semiconductor contacts has been given by one of us elsewhere.[7] Although our models of the junctions are not as refined as those customarily used in metal-oxide-metal junctions,[1] they exhibit the advantage that the entire calculation of the tunneling probability can be performed in closed form. As the effects of the bias-induced changes in the space-charge region and the mass changes across the junction have not been examined previously, we feel that analytical simplicity is essential at this stage of the development of the theory in order that the consequence of these new features be separated from the consequences of approximations made on more elaborate models.

In the next section we introduce our generalized Bethe-Sommerfeld model for the calculation of the tunneling current. In Sec. III, we give the results for an abrupt-junction model of *p-n* junctions. Section IV contains a discussion of the M-O-SM junction which includes both the effects of space-charge in the semimetal and of surface states at the semimetal-oxide boundary.

II. THE EFFECTIVE-MASS MODEL

In order to apply the Bethe-Sommerfeld free-electron model to the calculation of tunneling probabilities in semimetal and semiconductor junctions, two aspects of the model must be generalized. First, we must formulate it in a fashion to account for the possibility of large changes in the effective mass of the tunneling electron as it passes from one component of the junction to another. Second, we must consider the possibility that the constant-energy ellipsoids are centered about different points in momentum space in the various components of the junction. The values of the effective masses and location of the constant-energy surfaces in

[1] See, e.g., W. A. Harrison, Phys. Rev. **123**, 85 (1961); J. G. Simmons, J. Appl. Phys. **34**, 1793 (1963); **34**, 2581 (1963). References to extensive early literature may be found in these papers.
[2] L. Esaki and P. J. Stiles, Phys. Rev. Letters **14**, 902 (1965); **16**, 574 (1966).
[3] References to an extensive literature may be found in R. T. Shuey, Phys. Rev. **137**, A1268 (1965).
[4] A. Sommerfeld and H. Bethe, *Handbüch der Physik* edited by S. Flügge (Springer-Verlag, Berlin, 1933), Vol. 24/2, p. 150.
[5] M. I. Nathan, J. Appl. Phys. **33**, 1460 (1962); P. N. Butcher, J. A. Hulbert, and F. K. Hulme, J. Phys. Chem. Solids **31**, 320 (1961).
[6] I. I. Ivanchik, Fiz. Tverd. Tela **3**, 103 (1961) [English transl.: Soviet Phys.—Solid State **3**, 75 (1961)]; See also P. André, in *Proceedings of the International Conference on the Physics of Semiconductors* (Dunod Cie., Paris, 1964), p. 593.
[7] J. Conley, C. B. Duke, G. D. Mahan, and J. J. Tiemann, Phys. Rev. **150**, 466 (1966).

momentum space are taken from models of the energy bands of the bulk materials.

It is well known[8] that if a slowly varying potential $V(x)$ is imposed in a solid, the resulting wave functions within a given band n are given by the (approximate) expressions:

$$\psi_E(\mathbf{x})=\sum_k f_{n,E}(\mathbf{k})\psi_{nk}(\mathbf{x}), \tag{2.1a}$$

$$f_n(\mathbf{k})=\frac{1}{\sqrt{\Omega}}\int e^{i\mathbf{k}\cdot\mathbf{x}}f_n(\mathbf{x})d^3x, \tag{2.1b}$$

$$[E_n(-i\nabla)+V(\mathbf{x})]f_{n,E}(\mathbf{x})=Ef_{n,E}(\mathbf{x}), \tag{2.1c}$$

where $E_n(\mathbf{k}-\mathbf{k}_0)$ specifies the bulk one-electron energies as a function of the distance in k space away from their minimum (or maximum) located at $\mathbf{k}=\mathbf{k}_0$ and Ω is the volume of the system. The effective-mass model consists of expressing the tunneling probability solely in terms of the envelope functions f rather than using the full wave functions ψ. It is evident that some information about the wave functions is lost in this approximation (although the approximation underlies the Bethe-Sommerfeld model). In order to examine the extent of this loss, we can compare the results of our effective-mass model with those of the full Bloch-wave calculation for a one-dimensional system comprised of an abrupt junction of two materials with dissimilar one-electron periodic potentials. One finds that current conservation is guaranteed in both cases by use of the boundary condition

$$\frac{1}{m_L^*}\left(\frac{\partial f}{\partial x}\right)_{x\to 0_L}=\frac{1}{m_R^*}\left(\frac{\partial f}{\partial x}\right)_{x\to 0_R} \tag{2.2}$$

in place of the usual continuity of the derivative of f at the boundary. In Eq. (2.2) the subscripts "L" and "R" denote the left- and right-hand sides of the boundary. We have considered a region of energies where both sides of the junction have allowed energy bands which are adequately approximated by

$$E_n(k)=E_0+\hbar^2k^2/2m_n^*. \tag{2.3}$$

An alternative statement of Eq. (2.2) is obtained by noting that, if we permit the change in energy-band structure with position to be simulated by a spatially varying effective mass $m^*(x)$, then the appropriate Hermitian one-electron Hamiltonian is

$$H=-\frac{\hbar^2}{2}\nabla\cdot\frac{1}{m(\mathbf{x})}\nabla+V(\mathbf{x}). \tag{2.4}$$

Equation (2.4) is the analog of Harrison's[1] Eq. (4) in which $\alpha=m^{-1}$ and $\beta=1$. Equation (2.2) represents the limiting case of (2.4) for an abrupt junction in one dimension.

Equations (2.2) or (2.4) guarantee that within the energy region for which Eq. (2.3) is an adequate approximation, we lose no information in the matching condition on the derivative of ψ by using only the envelope function f. However, we always lose the information contained in the Bloch functions when, in the effective-mass method, we apply the boundary condition of continuity of f rather than continuity of ψ. Thus the method is valid only when the Bloch functions do not change rapidly as a function of $\mathbf{k}$ over the energy range of interest. This condition is usually satisfied if Eq. (2.3) is an adequate approximation of the one-electron energies, (i.e., near a stationary point on the one-electron energy surface).

We use in this paper the abrupt-junction approximation for which Eq. (2.2) is relevant. In p-n junctions this boundary condition is most relevant for heterojunctions[9] although for purposes of illustration we apply it herein to a GaAs homojunction. In the uniform-field,[3] as well as junction-potential,[10] models of tunneling in homojunctions, one considers the tunneling to be induced by interband matrix elements of an appropriate slowly varying junction potential superimposed on a single underlying band structure. Like these models, for a uniform-field p-n junction the generalized Bethe-Sommerfeld model gives the Zener transition probability, $\exp(-CE_g^{3/2}/F)$, where E_g is the band gap, $F=eE$, E is the constant field strength, and C is a constant which depends on the point in the junction at which the change in mass is imposed according to Eq. (2.2). However, the generalized Bethe-Sommerfeld model is less relevant for p-n homojunctions than the junction-potential model. We propose it for abrupt heterojunctions and as a rough alternative to the junction-potential model of homojunctions on the basis of which one can obtain qualitatively correct results with much less labor.[3] The features of p-n tunneling discussed herein, i.e., the effects of the bias dependence of the space-change region and the behavior of the conductance near a band gap, have not previously been discussed using the junction-potential model.

Within the framework of a one-electron approximation, we expect our abrupt-junction boundary conditions to describe adequately tunneling in M-O-SM junctions for energies near the band edges in the semimetal.

We conclude this section by reiterating, for later reference, some well-known[1,3,10] formulas for the tunneling current density from the left to the right of a tunnel junction, with the boundary condition of con-

[8] See, e.g., J. C. Slater, Phys. Rev. **76**, 1592 (1949); H. M. James, *ibid.* **76**, 1611 (1949); W. Kohn, Solid State Phys. **5**, 238 (1957).

[9] D. M. Cook has recently reviewed the one-electron model of heterojunctions in General Electric Research Laboratory Report No. 63-RL-3516G, 1963 (unpublished).

[10] D. R. Fredkin and G. H. Wannier, Phys. Rev. **128**, 2054 (1962).

servation of the components $k_{\parallel}$ of $\mathbf{k}$ parallel to the face of the junction.

$$J_{LR}=\frac{4\pi e}{\hbar}\int_{-\infty}^{\infty}dE_R[f_0(E_R+eV)-f_0(E_R)]\times\int\frac{d^2k_{\parallel}}{(2\pi)^2}Z(k_{\parallel},E_R), \quad (2.5a)$$

$$f_0(E)=\{1+\exp[(E-\mu)/\kappa T]\}^{-1}. \quad (2.5b)$$

We use κ to denote Boltzmann's constant, μ the chemical potential, V the bias applied to the left-hand side of the junction, e the magnitude of the electron charge, and Z the barrier penetration probability calculated from the one-dimensional generalized Bethe-Sommerfeld model. The left-hand side of the junction is taken to have either electron (+) or hole (−) spherical surfaces of constant energy centered about $k=0$.

$$E_L=\pm\hbar^2k_L^2/2m_L. \quad (2.6a)$$

This side is the metal in the metal-semimetal junctions and the p-type side in the GaAs p-n junction. The right-hand side has the spherical energy surfaces

$$E_R=\Delta E\pm\frac{\hbar^2}{2m_R}[(k_{\parallel,R}-k_{\parallel,0})^2+(k_{\perp,R}-k_{\perp,0})^2] \quad (2.6b)$$

centered about $\mathbf{k}_0$, either electron(+)- or hole(−)-like, and with the bottom of the band shifted by ΔE relative to the bottom of the band on the left-hand side of the junction. We finally note that in accordance with Eq. (2.1c), only the shift of the energy spheres in the $k_{\parallel}$ component affects the kinematics of the tunneling in the M-O-SM junctions.

III. p-n TUNNEL JUNCTIONS

As stated in the Introduction, we reconsider the calculation of the conductance of tunnel diodes from the standpoint of the generalized Bethe-Sommerfeld[4] model rather than the uniform-field model[3,11] for two reasons. First, we examine the tunneling near a band edge without the use of either the WKBJ approximation or its equivalent.[3] Second, a previous discussion of space-charge effects in metal-semiconductor contacts[7] led to the conclusion that these effects caused a resistance maximum near zero bias. As similar conductance minima have been observed in p-n diodes,[12] we calculate the conductance in highly asymmetrical tunnel junctions for which, in the limit that one of the contacts is much more heavily doped than the other,

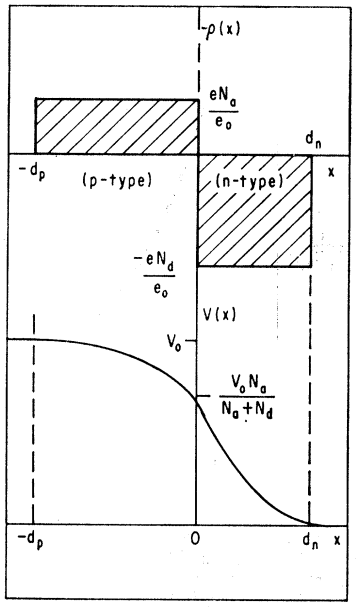

FIG. 1. A schematic illustration of the charge and potential distributions in an abrupt p-n junction. The symbols are defined following Eqs. (3.1) in the text.

we recover the metal-semiconductor Shottkey barrier case.

The abrupt-junction model of a p-n homojunction is illustrated in Fig. 1. The most important feature to be incorporated into any model of the junction is a description of the change in the junction-potential with applied bias. The abrupt-junction model reproduces this feature, and in addition leads to a potential for which the one-electron Schrödinger equation is exactly solvable in terms of Weber functions.[13] The junction between the p- and n-type material is taken at $x=0$. At this point we use the heterojunction joining condition, Eq. (2.2), and continuity of the envelope function f as the boundary conditions. In the abrupt step-junction model, the negative space-charge in the left-hand plane terminates at $x=-d_p$ and the positive space-charge in the right half plane at $x=d_n$. The negative of the space-charge and the corresponding potential energy of an electron is shown in Fig. 1. Continuity of both f and f' are required at $x=-d_p$ and $x=d_n$. When a bias V is applied to the p-type side of the junction, the potential energy is given by[14]

$$\begin{aligned}V(x)&=V_0-eV && x<-d_p\\&=V_0-eV-\left(\frac{2\pi N_ae^2}{\epsilon_0}\right)(x+d_p)^2 && -d_p<x<0\\&=\left(\frac{2\pi N_de^2}{\epsilon_0}\right)(x-d_n)^2 && 0<x\leq d_n\\&=0 && x>d_n\end{aligned} \quad (3.1)$$

[11] See also E. O. Kane, J. Phys. Chem. Solids, **12**, 181 (1959); P. N. Argyres, Phys. Rev. **126**, 1386 (1962).

[12] R. N. Hall, J. H. Racette, and H. Ehrenreich, Phys. Rev. Letters **4**, 456 (1960); R. N. Hall, in *Proceedings of the International Conference on Semiconductors, Prague, 1960* (Academic Press Inc., New York, 1961), p. 193; R. A. Logan and J. M. Rowell, Phys. Rev. Letters **13**, 404 (1965); R. M. Williams and J. Shewchun, *ibid.* **14**, 824 (1965); **15**, 160 (1965).

[13] J. C. P. Miller, in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (U. S. Department of Commerce, National Bureau of Standards, Washington, D. C., 1964), Appl. Math. Ser. 55, p. 685.

[14] See, e.g., L. P. Valdes, *The Physical Theory of Transistors* (McGraw-Hill Book Company, Inc., New York, 1961).

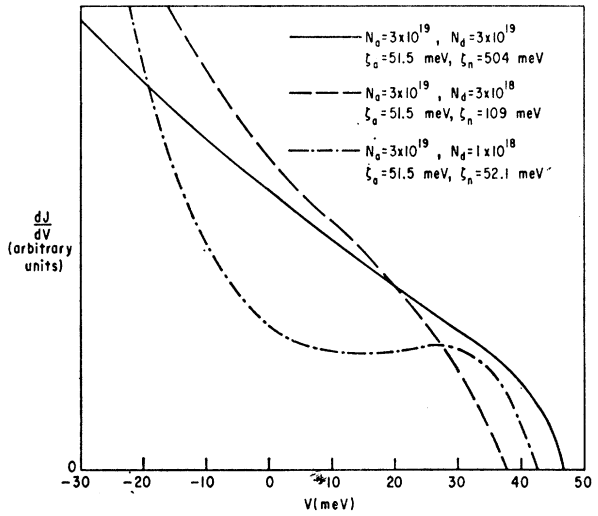

FIG. 2. Conductance as a function of applied bias calculated for a GaAs diode at 4.2°K. The doping ratios are indicated in the figure. The parameters $E_g=1.55$ eV, $(m_e/m)=0.07$, $(m_h/m)_l=0.17$, and $(m_h/m)_h=0.65$ were used in the calculation. (The quantity m denotes the free-electron mass.)

in which V_0 is the junction potential, ϵ_0 is the static dielectric constant of the junction material, N_a is the acceptor concentration, and N_d the donor concentration. We consider only degenerate semiconductors at low temperatures for which

$$V_0=E_g+\zeta_n+|\zeta_p|\,, \tag{3.2}$$

where E_g denotes the band gap, ζ_n the electron Fermi energy, and ζ_p the hole-Fermi energy. We do not incorporate small alterations in Eq. (3.2) associated with the details of the band theory of very degenerate materials[15] as they do not change our qualitative results. The continuity of $V(0)$ and $V'(0)$ imposes on d_n and d_p the conditions

$$N_a d_p^2+N_d d_n^2=\epsilon_0(V_0-eV)/2\pi e^2\,, \tag{3.3a}$$

$$N_a d_p=N_d d_n\,. \tag{3.3b}$$

From Eqs. (3.3) and (3.1) we compute the junction parameters for any given applied bias V.

Using the boundary conditions specified above, the transmission probability associated with a one-electron eigenstate of energy $E=E_{11}+E_\perp$ is given by

$$Z_{p-n}=\frac{16k_1k_4}{\pi^2}\frac{b_2^2b_3^2}{m_e m_h}\frac{1}{X^2+Y^2}\,, \tag{3.4a}$$

$$k_1=k_R=(2m_RE_{11}/\hbar^2)^{1/2}\,, \tag{3.4b}$$

$$k_4=k_L=(2m_LE_{11}/\hbar^2)^{1/2}\,, \tag{3.4c}$$

$$b_2=[16\pi m_RN_de^2/\hbar^2\epsilon_0]^{1/4}\,, \tag{3.4d}$$

$$b_3=[16\pi m_LN_ae^2/\hbar^2\epsilon_0]^{1/4}\,, \tag{3.4e}$$

in which m_R is the electron mass and m_L the (positive) hole mass. Only the contribution to the current due to the light holes has been calculated, although the Fermi energy is located by filling both the light and heavy hole bands. The quantities X and Y may be written in terms of the two linearly independent solutions $U(a,x)$ and $V(a,x)$ to Weber's equation.[13] The complete expressions for X and Y are given in Appendix A.

The conductance dJ/dV is calculated by using the expressions (3.4) in Eq. (2.5) and performing the two-dimensional integral numerically. The scale of $J(V)$ differs drastically from one doping level to the next. To get representative results on the same graph, we show in Fig. 2 the conductance at $T=4.2$°K associated with various doping ratios (N_a/N_d) plotted on an arbitrary scale. Comparison of these results with the experiments of Ref. 12 indicates that the observed conductance anomalies near zero bias are too narrow to be adequately described by our independent-particle model. The conductance minimum near zero bias for highly asymmetric junctions is a consequence of the fact that in this case, we begin to recover a metal-semiconductor contact[7] in which the dropoff in the density of states in the p-type material prevents a well-formed conductivity minimum at $eV=\zeta_n$.

Summarizing, we see from Fig. 2 that the bias-induced alteration in the transmission probability calculated from a single-particle model can introduce structure in the conductance both near zero bias and near the valley-current region. The Bethe-Sommerfeld model, as we have developed it, is thought to be applicable for the description of the prephonon current in direct-band-gap tunnel diodes. The experiments in

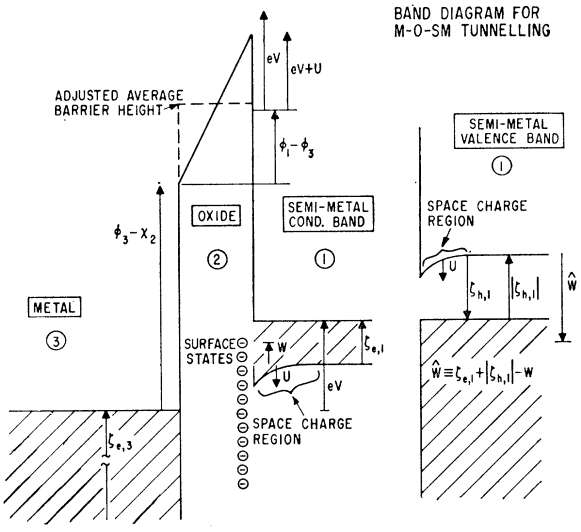

FIG. 3. The potential energy band diagram for M-OSM tunneling. The linear variation of the barrier was approximated by an average value, corrected for space-charge effects. The electron mass in the metal and in the barrier were taken as the free-electron mass m_0.

[15] W. Bernard, H. Roth, A. P. Schmid, and P. Zeldes, Phys. Rev. 131, 627 (1963).

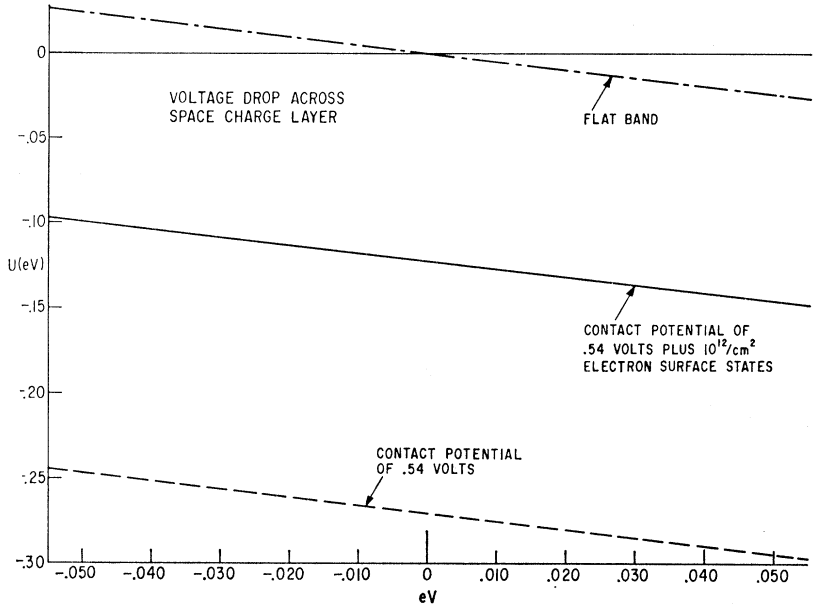

FIG. 4. The total voltage drop across the space-charge layer in the semimetal for: (a) flat band condition at zero bias; (b) contact potential of 0.54 V; (c) contact potential plus $10^{12}/cm^2$ electron surface states. We have used $\epsilon_{ox}=10$, $\epsilon_{sm}=100$, $x_b=10$ Å, $\zeta_e=17.8$ meV, and $\zeta_h=12$ meV. The work functions for bismuth and aluminum were taken as 4.28 and 3.74 eV, respectively, and $\zeta_{Al}=11.7$ eV.

Ge [5,16] junctions were not designed to study the details of the conductance as a function of bias and doping. They, at most, verified an approximate inverse-field dependence of the exponent in tunneling probability. We have pointed out that this dependence is a consequence of the Bethe-Sommerfeld model as well as the Keldysh-Kane model. The current-voltage curves in silicon diodes have been studied by Logan and Chynoweth[17] and interpreted in terms of Kane's more recent theory of tunneling in impure semiconductors.[18] Their results, like those of Hall *et al.*[12] without the zero-bias anomaly, are consistent with conductance-bias plots like the solid line in Fig. 2. Finally, we have made no corrections either for the Coulombic image forces in the space-charge region[19] or the Coulombic electron-hole final-state interaction.[20] As these corrections can be significant, they should be made before detailed comparison of the model with experiments are warranted.

IV. METAL-OXIDE-SEMIMETAL (M-O-SM) TUNNEL JUNCTIONS

Also within the framework of the generalized Bethe-Sommefeld model, we consider the description of tunneling current between metals and semimetals through a barrier of oxide. Of particular interest is:

(1) the existence of contributions to the tunneling current from both the conduction and valence bands in the semimetal.

(2) The division between the oxide and a space-charge layer in the semimetal of the total (applied plus contact potential) bias across the junction.

(3) The effect of differences in the metal and semimetal work functions (i.e., contact potential) and of surface states at the oxide-semimetal interface on the space-charge layer.

These features are illustrated in the band diagrams of Fig. 3. We adopt the convention of measuring the bias voltage in units of electron potential energy, and taking as our reference the bottom of the semimetal conduction band or the top of the valence band as appropriate.

The exact spatial behavior of the space-charge layer in the semimetal has been analyzed in Appendix B. We have demonstrated there that an adequate approximation to the exact solution for the potential in this region is an exponential decay into the interior whose characteristic length is given by the linearized Poisson equation. The total potential drop across the space-charge layer is determined by the interface boundary condition on the potential and the electric displacement vector. In the absence of a contact potential and of surface states, a flat band condition prevails at zero bias. At nonzero bias the total potential change between the metal and semimetal is shared between the oxide and the space-charge layers, the proportions being a sensitive function of the relative dielectric constants. Figure 4 shows the value of the potential drop across the space-charge layer as a function of applied bias for some typical experimental parameters. The role of a nonzero contact potential is identical with that of a

[16] D. Meyerhofer, G. A. Brown, and H. S. Sommers, Jr., Phys. Rev. **126**, 1329 (1962).
[17] R. A. Logan and A. G. Chynoweth, Phys. Rev. **131**, 89 (1963).
[18] E. O. Kane, Phys. Rev. **131**, 79 (1963).
[19] See, e.g., J. G. Simmons [J. Appl. Phys. **34**, 2581 (1963)] for an extensive study of these corrections in metal-insulator-metal junctions.
[20] W. Franz, Z. Naturforsch **14a** 415 (1959); C. B. Duke and G. D. Mahan, Bull. Am. Phys. Soc. **11**, 182 (1966).

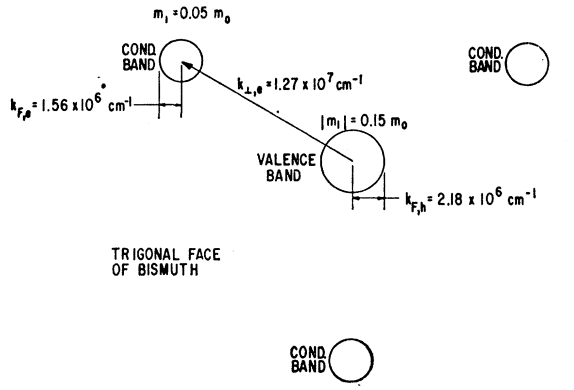

FIG. 5. The effective-mass electron and hole Fermi surfaces for the trigonal face of bismuth. The quantities $P_e = \hbar^2 k_{\perp,e}^2/2m_1 = 1.08$ eV and $P_h = 0$ used in Eq. (4.4) are obtained from this figure.

fixed reference bias voltage. Finally, surface states provide a discontinuity in the displacement vector for electron surface states; the potential barrier is enhanced for electrons tunneling from the semimetal and therefore, in that case, this will tend to diminish contact potential effects if $\varphi_{sm} - \varphi_m > 0$. We have assumed a density of surface states $\sim 10^{12}/cm^2$ as suggested by Bardeen.[21] Therefore, our model of the electrostatic potential-energy barrier in M-O-SM tunneling is, in the semimetal,

$$\varphi = Ue^{-\kappa_D x}, \quad 0_+ < x < +\infty, \tag{4.1a}$$

$$\kappa_D = \frac{6\pi n_0 e^2}{\epsilon_{sm}}\left[\frac{1}{\zeta_e} + \frac{1}{|\zeta_h|}\right], \tag{4.1b}$$

where $U = U(V)$ is discussed in Appendix B; n_0 is the electron (hole) density in semimetal; and ζ_e, ζ_h are the electron and hole Fermi energies, respectively. In the oxide we have the potential

$$\varphi = \tfrac{1}{2}(\varphi_m + \varphi_{sm}) - \chi_{ox} + (eV - U)[1 + (x/x_b)]; \quad -x_{b+} < x < 0_-, \tag{4.1c}$$

where φ_m, $\varphi_{sm} \equiv$ metal, semimetal work functions, V is the bias applied to the metal, χ_{ox} is the oxide electron affinity, and x_b is the barrier thickness. Finally, in the metal, we have the constant electrostatic potential

$$\varphi = -eV, \quad -\infty < x < x_{b-}. \tag{4.1d}$$

We have further approximated the barrier in the oxide by replacing the linear spatial variation by the value $\varphi(-x_{b+}) + (eV - U)/2$.

The exact solution of the free-electron wave equation in the semimetal is given by Bessel functions of imaginary order and real or imaginary arguments. The detailed discussion of these solutions is relegated to Appendix C. By invoking the boundary conditions discussed in Sec. II, one obtains an expression for the tunneling probability for use in Eq. (2.5a) as follows:

$$Z_{M\text{-}O\text{-}SM} = \frac{16 k_1 k_3 \gamma^2 e^{-2\gamma x_b} m_3/m_1}{(R^2 + T^2)(\gamma^2 m_3^2/m_2^2 + k_3^2)}, \tag{4.2}$$

where m_1, m_2, $m_3 \equiv$ effective masses in the semimetal, oxide, and metal, respectively; k_1, $k_3 \equiv$ normal component of the effective one-electron momentum in the semimetal and metal, respectively; $\gamma \equiv$ decay coefficient in the oxide;

$$R \equiv \gamma S_1 - (m_2/m_1) S_1', \quad T \equiv \gamma S_2 - (m_2/m_1) S_2', \tag{4.2a}$$

where we have defined for convenience

$$S_1 + iS_2 \equiv \psi(0), \quad S_1' + iS_2' \equiv \psi'(0). \tag{4.2b}$$

Appropriate series expansions for $\psi(0)$ and $\psi'(0)$ are given in Appendix C.

The expression for the tunneling probability given by Eq. (4.2) is exact except for neglect of terms in $e^{-\gamma x_b}$ in comparison with $e^{+\gamma x_b}$.

As an example of an M-O-SM junction, we have considered tunneling between the trigonal face of

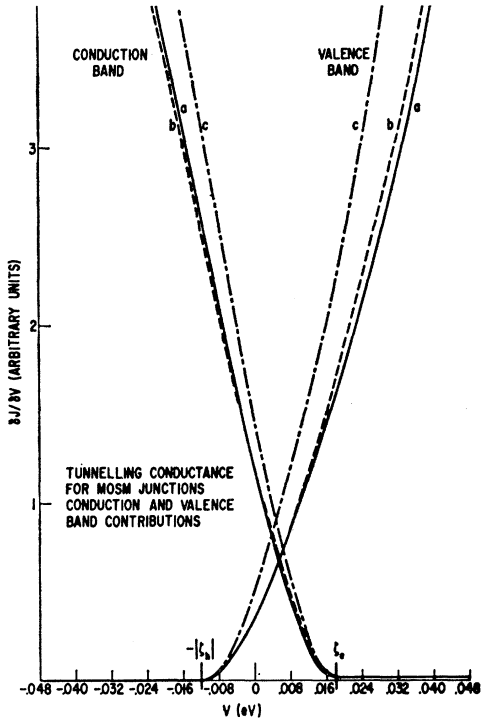

FIG. 6. This figure shows the separate conduction- and valence-band contributions to the tunneling conductance for the cases (a) square-barrier approximation neglecting space-charge effects; (b) space-charge effects; (c) space-charge effects including a contact potential of 0.54 V and $10^{12}/cm^2$ electron surface states. For positive bias the average barrier height is decreased whereas for negative bias it is increased. This effect is responsible for the larger space-charge-induced changes in the conductance at positive bias. The corrections to the effective-mass approximation for large values of $|V|$ flatten out these curves. The use of a multiple-band model, on the other hand, enhances their rise away from zero bias.

[21] J. Bardeen, Phys. Rev. 71, 717 (1947).

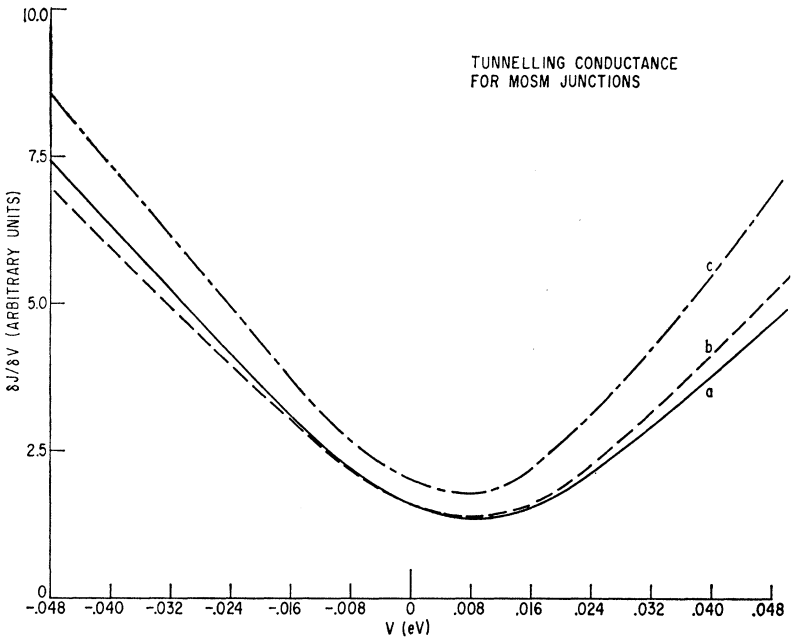

FIG. 7. Total tunneling conductance for: (a) square barrier approximation neglecting space-charge effects; (b) space-charge effects; (c) space-charge effects including a contact potential of 0.54 V and $10^{12}/\text{cm}^2$ electron surface states.

bismuth, through a barrier of oxide into aluminum. The projection of the effective-mass Fermi surfaces of bismuth as seen by the aluminum is shown on Fig. 5.[2,22] Following the discussion of Sec. II, the elongated and tilted Fermi-surface ellipsoids have been replaced by spheres. We recall that the shift in the normal component of $\mathbf{k}_0$ of the center of the Fermi spheres does not enter the calculation, in the effective-mass approximation.

In the kinematics, we have utilized conservation of energy and momentum in the direction parallel to the junction plane (specular reflection) and have replaced the small azimuthally dependent term in the calculation of $k_3(k_1)$ by its angular average, which is zero. The relations one obtains for k_1, k_3, and γ in terms of the total kinetic E and the component E_{11} parallel to the plane of the junction in the bismuth are

$$k_1=\left(\frac{2m_1}{\hbar^2}(E-E_{11})\right)^{1/2}, \tag{4.3}$$

$$k_3=\left(\frac{2m_3}{\hbar^2}\left\{\pm E+eV+\zeta_{e,3}-\frac{m_1}{m_3}\left[\frac{P_e}{P_h}+E_{11}\right]\right\}\right)^{1/2}$$

$$\approx\left(\frac{2m_3}{\hbar^2}\zeta_{e,3}\right)^{1/2}, \tag{4.4}$$

$$\gamma=\left(\frac{2m_2}{\hbar}\left\{\frac{\varphi_s+\varphi_{sm}}{2}-\chi_{ox}\pm\left[\frac{\zeta_{e,1}}{\zeta_{h,1}}-E+E_{11}\right]\right.\right.$$

$$\left.\left.-\tfrac{1}{2}(eV-U)\right\}\right)^{1/2}, \tag{4.5}$$

where the upper sign or quantity refers to the conduction band and, consistent with taking the top of the band as our reference in the case of the valence band, the lower sign or quantity refers to the holes. The quantity P is the displacement of the center of the Fermi spheres of the bismuth, as taken from Fig. 5.

Figures 6 and 7 show the behavior of the conductance $\partial J/\partial V$ as a function of V for the conduction and valence bands and their sum, respectively. On each figure, the calculation for: (a) the simple average-height square barrier; (b) the average-height adjusted square barrier plus the space-charge region of the semimetal utilizing a flat band condition at zero bias; and (c) the average-height adjusted square barrier plus the space-charge region of the semimetal including a 0.54 eV positive contact potential and $10^{12}/\text{cm}^2$ surface states distributed uniformly between ± 0.025 eV.

These and other results indicate that inclusion of space-charge, contact potential, and surface states does not significantly alter the fundamental character of the conductance curves. In particular, one effect is the slight lower of the adjusted average-height square barrier because that part of the voltage drop or of the contact potential always appears across the space-charge layer. Thus, in the case of positive V, the barrier does not rise quite as much with inclusion of the space-charge layer [cases (b) and (c)] as when this layer is neglected. This may be seen as the effect of U on γ in Eq. (4.5). Also the space-charge layer acts to modify the dynamics of the electrons in both bands in the bismuth (band bending). These effects are illustrated in Fig. 8. For any reasonable choice of parameters, the voltage drop across the space-charge layer is small in comparison with the barrier height provided by the

[22] Y. Kao, Phys. Rev. **129**, 1122 (1963).

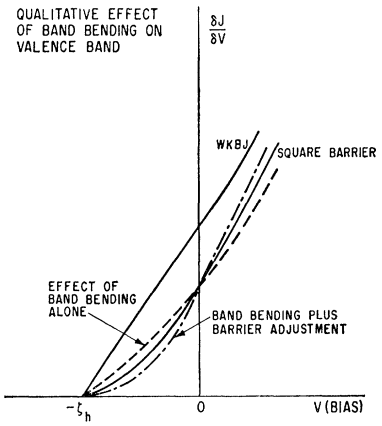

FIG. 8. The qualitative behavior is shown of the conductance curves for the valence-band component of M-O-SM tunneling. The WKBJ approximation rises $\sim(|\zeta_h|+V)$ from the band edge; the exact square-barrier model, which as an additional prefactor $\sim k_1$, rises as $(|\zeta_h|+V)^{3/2}$. The effect of band bending alone is to enhance this component of tunneling for negative bias and depress it for positive bias. The voltage drop across the space-charge region, however, is reflected as a modification of the barrier height, which overcomes the effect of band bending. For the continuum electrons in the conduction band, the average barrier effect tends to enhance the space-charge effect.

oxide. This fact provides the essential difference between the rather marked effects seen in *p-n* junctions and the relatively small effects in M-O-SM tunneling. The small effects of both removing the WKBJ approximation[1] and including the space-charge-induced "band-bending" demonstrate the inadequacy of the interpretation of their tunneling experiments on Al-oxide-Bi junctions suggested by Esaki and Stiles.[2] However, we have not included the contributions to the tunneling current due to electrons trapped near the surface by an attractive space-charge potential.[23]

The negative conductance exhibited by the valence band at negative biases $|eV|<|\zeta_h|$ is due to the lowering of the oxide barrier height with applied bias. Although the effect is small in our application, it is thought to dominate the conductivity at large bias values into degenerate semiconductors.[24]

Note added in proof. Since the submission of this manuscript we have performed an analysis of the contributions to the conductance of localized, discretely spaced two-dimensional energy bands for which the electron motion normal to the plane of the junction is described as a bound state in the accumulation region near the oxide—semimetal interface in Fig. 3.[25] When such a state occurs at zero energy, the conductance due to continuum electrons may be found from Eqs. (2.5) and (4.2) to be $G\alpha(\mathrm{eV}-\zeta_e)^{1/2}$ for an electron band. This result still does not justify the interpretation given in Ref. 2. However, for localized states with binding energy $E_B>0$ one obtains[25] an additional contribution to the conductance which is proportional to the two-dimensional density of states for motion parallel to the plane of the junction, and of magnitude roughly (E_B/ζ_e) relative to the continuum electron conductance at zero bias. Structure in the conductance near zero bias, similar to that reported in Ref. 2, can be attributed to such localized states associated with either a very narrow electron band or with several wider bands in a multiple band model.

[23] F. F. Fang and W. E. Howard, Phys. Rev. Letters **16**, 797 (1966); A. B. Fowler, F. F. Fang, W. E. Howard, and P. J. Stiles, *ibid.* **16**, 901 (1966).
[24] L. Esaki and P. J. Stiles, Phys. Rev. Letters **16**, 1108 (1966)
[25] D. J. Bendaniel and C. B. Duke, General Electric Research and Development Center Report 66-C-331, 1966 (unpublished).

ACKNOWLEDGMENTS

The authors are indebted to V. Scavullo for assistance in programming and to Dr. C. P. Bean, Dr. L. Esaki, and Dr. I. Giaever for discussions on the material presented in Sec. IV.

APPENDIX A: COEFFICIENTS FOR THE TRANSMISSION PROBABILITY IN AN ABRUPT *p-n* JUNCTION

The following expressions specify the quantities used in Eq. (2.4) in the text. All symbols are defined in the text.

$$X=-b_2b_3\{V'(a_2,0)V'(a_3,0)M_3+U'(a_3,0)U'(a_2,0)M_2-V'(a_2,0)U'(a_3,0)M_4-U'(a_2,0)V'(a_3,0)M_1\}+k_1k_4\{V(a_2,0)V(a_3,0)M_3+U(a_2,0)U(a_3,0)M_2-V'(a_2,0)U'(a_3,0)M_4-U'(a_2,0)V'(a_3,0)M_1\}, \quad \text{(A1)}$$

$$Y=-b_2k_4\{V'(a_2,0)V(a_3,0)M_3+U'(a_2,0)U(a_3,0)M_2-V'(a_2,0)U(a_3,0)M_4-U'(a_2,0)V(a_3,0)M_1\}, \quad \text{(A2)}$$

$$M_1=(b_3/m_L)V(a_2,b_2d_n)U'(a_3,b_3d_p)-(b_2/m_R)U(a_3,b_3d_p)V'(a_2,b_2d_n), \quad \text{(A3)}$$

$$M_2=(b_3/m_L)V(a_2,b_2d_n)V'(a_3,b_3d_p)-(b_2/m_R)V(a_3,b_3d_p)V'(a_2,b_2d_n), \quad \text{(A4)}$$

$$M_3=(b_3/m_L)U(a_2,b_2d_n)U'(a_3,b_3d_p)-(b_2/m_R)U(a_3,b_3d_p)U'(a_2,b_2d_n), \quad \text{(A5)}$$

$$M_4=(b_3/m_L)U(a_2,b_2d_n)V'(a_3,b_3d_p)-(b_2/m_R)V(a_3,b_3d_p)U'(a_2,b_2d_n). \quad \text{(A6)}$$

The forms of $U(a,x)$, $V(a,x)$, and their derivatives at $x=0$ are known (see Ref. 13). In numerical calculations their "thick-junction" asymptotic forms were used when $x=$"bd" in the above formulas. The parameters a_2 and a_3 are given by

$$a_2=-\frac{k_1^2}{2}\left[\frac{\hbar^2\epsilon_0}{4\pi m_R N_d e^2}\right]^{1/2}, \quad \text{(A7a)}$$

$$a_3=-\frac{k_4^2}{2}\left[\frac{\hbar^2\epsilon_0}{4\pi m_L N_a e^2}\right]^{1/2}. \quad \text{(A7b)}$$

APPENDIX B: SPACE-CHARGE LAYERS AT INSULATOR-SEMIMETAL CONTACTS

We adopt the model of a semimetal in which electrons and holes in parabolic bands are treated as a two-component electrically neutral plasma. At an oxide interface, space-charge regions occur just as in degenerate semiconductors[7] only in this case the space-charge is mobile rather than the charge on immobile ionized donors or acceptors. Denoting by a subscript e the electron parameters, by a subscript h the hole parameters, and by ζ the bulk Fermi energies of the semimetal at zero temperature, Poisson's equation for the one-electron potential energy u is given by

$$\frac{d^2u}{dx^2}=-\frac{4\pi e^2}{\epsilon_R}[n_e(x)-n_h(x)] \tag{B1}$$

in the absence of surface states. The quantities $n_e(x)$ and $n_h(x)$ are the particle densities of electrons and holes, respectively, and we have taken the semimetal to extend from $x=0$ to $x=\infty$. The boundary conditions on Eq. (B1) are

$$\lim_{x\to\infty} u(x)=0, \tag{B2a}$$

$$\lim_{x\to\infty} du/dx=0. \tag{B2b}$$

At zero temperature the particle densities are given by the Fermi-Thomas model to be

$$n_e(x)=n_0[1-u/\zeta_e]^{3/2}\theta(\zeta_e-u), \tag{B3a}$$

$$n_h(x)=n_0[1+u/|\zeta_h|]^{3/2}\theta(|\zeta_h|-u), \tag{B3b}$$

where n_0 is the bulk particle-density of electrons and holes and

$$\theta(x)=1;\quad x>0$$
$$=0;\quad x<0. \tag{B4}$$

In this case Eq. (B1) can be reduced to the quadrature

$$x=\left(\frac{5\epsilon_R}{16\pi n_0 e^2}\right)^{1/2}\int_{V(0)}^{u(x)} du/D(y), \tag{B5a}$$

$$D(y)=\left\{\theta\left(1+\frac{y}{|\zeta_h|}\right)|\zeta_h|\left[1+\frac{y}{|\zeta_h|}\right]^{5/2}-|\zeta_h|-\zeta_e+\theta\left(1-\frac{y}{\zeta_e}\right)\zeta_e\left[1-\frac{y}{\zeta_e}\right]^{5/2}\right\}^{1/2}. \tag{B5b}$$

The quantity $u(0)$ is determined by considering a particular model of the total metal-oxide-semimetal junction. In the absence of surface states the difference in the work function of the semimetal φ_{sm} and that of the metal φ_m is $\Delta\varphi=\varphi_{sm}-\varphi_m$. We take the model in which the voltage drop of $\Delta\varphi+eV$ is divided between an oxide layer of thickness x_b from $x=-x_b$ to $x=0$ and the semi-infinite semimetal. The boundary conditions at the oxide-semimetal interface are

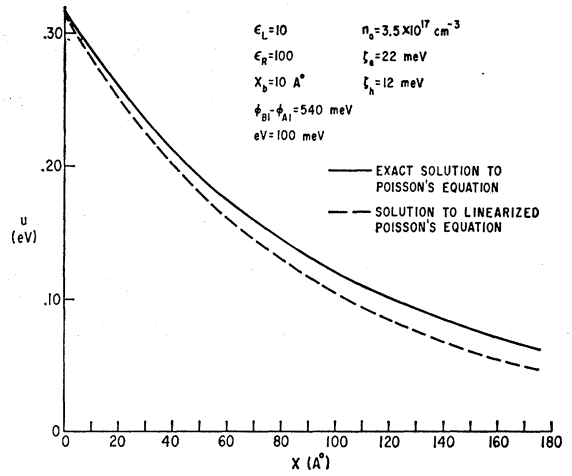

FIG. 9. Comparison of the exact space-charge-induced potential with the exponential approximation to it. The relevant parameters are indicated in the figure. The value of u at $x=0$ is determined from Eq. (B7). The exact potential is obtained from Eq. (B5) and the exponential approximation from Eq. (B8).

$$u(x\to 0_L)=u(x\to 0_R), \tag{B6a}$$

$$\epsilon_L u'(x\to 0_L)=\epsilon_R u'(x\to 0_R), \tag{B6b}$$

$$u(x)=F(x+x_b),\quad -x_b\leq x\leq 0 \tag{B6c}$$

$$\Delta\varphi+eV=-u(0)+Fx_b, \tag{B6d}$$

where F is the field due to the constant field in the oxide. $u(x)$ is just the electrostatic component of the potential and is augmented in the oxide by a square barrier of height $V_0=\frac{1}{2}(\varphi_m+\varphi_{sm})-\chi_{ox}$. By taking the derivative of Eqs. (B5a) and (B6c) we can eliminate F and obtain the set of coupled equations for $u(0)$ and $u'(0)$, i.e.,

$$U\equiv u(0)=-\Delta\varphi-eV+(\epsilon_R u'(0)x_b/\epsilon_L), \tag{B7a}$$

$$u'(0)=-\mathrm{sgn}(V)(16\pi n_0 e^2/5\epsilon_R)^{1/2}D[U]. \tag{B7b}$$

To obtain the space-charge potential in the semimetal for a given bias, Eqs. (B7) were solved iteratively on a GE235 computer and then the potential was calculated from Eq. (B5).

In calculating the tunneling current, we used the exponential barrier shape obtained by using the solution to the linearized form of Eq. (B1) for the particle densities, Eq. (B3). This solution is

$$u(x)=U\exp[-\kappa_D x], \tag{B8a}$$

$$\kappa_D^2=\frac{6\pi n_0 e^2}{\epsilon_{sR}}\left[\frac{1}{\zeta_e}+\frac{1}{|\zeta_h|}\right]. \tag{B8b}$$

Even for sufficiently large bias such that $u(0)\gg\zeta_e, \zeta_h$ the linear approximation represents the exact solution to within 10–20%. An example for the most extreme parameters used in the analysis is shown in Fig. 9 to

illustrate the accuracy of the linear approximation. The semimetal parameters are those appropriate for Bi at a Bi-Al_2O_2-Al junction.

Next, consider the modification of the above results imposed by a band of surface states which trap

$$n_s=\rho_s[(eV+E_B)\theta(eV+E_B)\theta(E_B-eV) +2E_B\theta(eV-E_B)] \quad \text{(B9a)}$$

states per unit area at the oxide-semimetal boundary, where ρ_s is the number of states per unit energy and area. The charge density associated with these states is

$$\sigma_s=(\mathrm{sgn}q)en_s, \quad \text{(B9b)}$$

where sgn$q$ denotes the sign of the trapped carriers. The effect of these states is to alter the condition of continuity of D in Eq. (B6b) to give

$$\epsilon_L u'(x\to 0_L)=\epsilon_R u'(x\to 0_R)-4\pi\sigma_s|e|. \quad \text{(B10)}$$

(Recall that u is the potential-energy seen by an electron.) Thus the net effect of surface states on our previous analysis is to replace Eq. (B7a) by

$$U\equiv u(0)=-\Delta\varphi-eV +(x_b/\epsilon_L)(\epsilon_R u'(0)-4\pi\sigma_s|e|) \quad \text{(B7c)}$$

which is solved iteratively with Eq. (B7b).

APPENDIX C: WAVE FUNCTIONS IN AN EXPONENTIAL POTENTIAL

The Schrödinger equation in the semi-metal is

$$\psi''=\frac{2|m_1|}{\hbar^2}(W\mp Ue^{-\kappa_D x})\psi, \quad \text{(C1)}$$

where the minus sign applies to the conduction band and the plus sign to the valence band where, in that case, the mass $|m_1|$ and the kinetic energy $|W|$ are used, with W measured from the top of the band.

The solutions of Eq. (C1) are linear combinations of the two Bessel functions[25] given by

$$\psi_\pm\sim J_{\mp 2ik_1/k_D}\left(-\frac{1}{\kappa_D}\left[\frac{2m_1Ue^{-\kappa_D x}}{\hbar^2}(\mp)\right]^{1/2}\right), \quad \text{(C2)}$$

where

$$k_1\equiv[(2m_1/\hbar^2)W]^{1/2}.$$

For the calculation of the barrier transmission it is sufficient to know the outgoing solution and it is convenient to normalize such that

$$\psi\to e^{ik_1x} \quad \text{as} \quad x\to+\infty. \quad \text{(C3)}$$

We obtain

$$\psi_+=[\tfrac{1}{2}G]^{-\nu}\Gamma(1+\nu)J_\nu(Ge^{-\kappa_D x}), \quad \text{(C3')}$$

where $\Gamma(z)$ is the complex gamma function,[26] and we adopt the notation

$$G^2\equiv\frac{8m_1(\mp)U}{\hbar^2(-\kappa_D)^2}, \quad \text{(C4a)}$$

$$\nu\equiv-2ik_1/\kappa_D. \quad \text{(C4b)}$$

Utilizing the general series expansion for the Bessel functions[25] the resultant expressions for the amplitude and slope of the wave function at $x=0$ are

$$\psi(0)=\sum_{l=0}^{\infty}\frac{(-\frac{1}{4}G^2)^l\Gamma(1+\nu)}{l!\Gamma(1+\nu+l)}, \quad \text{(C5)}$$

and

$$\psi'(0)=\sum_{l=0}^{\infty}\frac{(-\frac{1}{4}G^2)^{l-1}G^2\kappa_D\Gamma(1+\nu)}{4(l-1)!\Gamma(1+\nu+l)}-\frac{\kappa_D\nu\psi(0)}{2}. \quad \text{(C6)}$$

These summations are obtained by summing the series numerically using the standard complex two-term recursion formulas.

[25] F. W. J. Oliver, in *Handbook of Mathematical Functions* edited by M. Abramowitz and I. A. Stegun (U. S. Department of Commerce, National Bureau of Standards, Washington, D. C., 1964), Appl. Math. Ser. 55, p. 355.

[26] P. J. Davis, in *Handbook of Mathematical Functions* edited by M. Abramowitz and I. A. Stegun (U. S. Department of Commerce, National Bureau of Standards, Washington, D. C., 1964), Appl. Math. Ser. 55, p. 253.