Energy Distributions of Field Emitted Electrons

ROBERT STRATTON
Texas Instruments Incorporated, Dallas Texas
(Received 13 March 1964)

The total and normal energy distributions of electrons, emitted from a semiconductor conduction or valence band, by quantum mechanical tunneling through a surface potential barrier in vacuum, have been derived. General results, which apply for arbitrary emitter band structures, barrier shapes and distribution functions in momentum space, are given. Specific formulas are worked out for the case of spherical energy surfaces, an image force barrier and Fermi-Dirac statistics. The emitter band structure only affects the emitted electron energy distribution if the important energy surfaces in momentum space are very small. Thus, appreciable effects occur only for non- or semidegenerate semiconductors with low effective masses. The prevailing lack of agreement between theory and experiment and some difficulties in interpreting results of retarding potential measurements are discussed.

1. INTRODUCTION

THE application of a sufficiently intense electric field normal to the surface of a metal or semiconductor leads to the emission of electrons mainly by quantum mechanical tunneling through the surface barrier, rather than by thermionic emission over the top of the barrier. The fields required for measurable emission currents, at room temperature, are in the range 10^7 ro 10^8 V/cm or, equivalently, tunnel paths must be no more than several tens of angstroms long. Two ways of observing the phenomena experimentally have proved possible. In one, the emitter, in the form of a fine needle with a tip radius of about 10^{-4} cm, is placed inside an evacuated chamber, partly lined with a conducting film anode. In the other, a bias voltage is applied across a thin insulating film (10 to 100 Å) sandwiched between two electrodes. It is convenient to distinguish between the two arrangements by describing the phenomena as "field emission" in the former case and "thin film tunneling" in the latter case even though they are fundamentally the same.

The theory of field emission from metals was first derived by Fowler and Nordheim in 1928, who showed that, if the temperature is not too high, most of the emitted electrons originate from a small energy interval around the Fermi level of the metal. Henderson and his co-workers,[1] using a retarding potential analyzer, were the first to demonstrate that the electrons do originate from near the Fermi level but the half-widths of their measured distributions were too large. The latter was in part due to their use of a cylindrical arrangement (emitter in the form of a fine tungsten wire) which leads to field distortions and poor resolution.

In comparing their measured distributions with theory, Henderson and Dahlstrom[1] derived the "normal energy distribution" which can be generally defined as the number of electrons emitted whose "x-directed energy E_x" is in the interval E_x to E_x+dE_x. Here, the x direction is normal to the emitting surface and E_x is the difference between the total electron energy E and the energy $E_\perp$ associated with the tangential electron momentum $p_\perp$ in vacuum, just outside the metal surface, i.e., $E_\perp=p_\perp^2/2m$ where m is the free-electron mass. In the calculation by Henderson and Dahlstrom the free-electron model is assumed for the metal (with a free-electron mass) so that if $p_\perp$ is conserved, E_x is simply $(p_x^2/2m)$ which is their definition of E_x. Our more general definition applies for an arbitrary band structure of the emitter.

Müller[2] employed a spherical retarding potential arrangement and in later work[3] achieved good agreement between the emitted energy distribution from a tungsten tip and the calculated normal energy distribution. However, in 1959, Müller and Young[4] remeasured the energy distribution in an improved retarding potential tube and obtained distributions about one-third as wide as those predicted by the normal energy distribution theory. They pointed out that the previous agreement between theory and experiment, which arose because of the limited resolution of the older analyzer, was spurious since the experimental arrangement actually measures the "total energy distribution." This is defined as the number of electrons emitted whose total energy is in the range E to $E+dE$. Very good agreement was finally obtained between the new measurements and the calculated total energy distribution,[5] almost thirty years after the first experimental attempts were made.

Fischer[6] has derived total energy distributions for field emission from the conduction and valence bands of a semiconductor. His expression for valence-band emission is, however, incorrect since he based his calculations on a previous derivation by the present author[7] for the total valence-band emission current which needs to be modified. (cf. Sec. 4 and the Appendix where

[1] J. E. Henderson and R. E. Badgley, Phys. Rev. **38**, 590 (1931); J. E. Henderson and R. K. Dahlstrom, *ibid.* **45**, 764 (1934); **55**, 473 (1939).

[2] E. W. Müller, Z. Physik **102**, 734 (1937).

[3] E. W. Müller, Z. Physik **120**, 261 (1943); E. W. Müller and K. Bahadur, Phys. Rev. **102**, 624 (1956).

[4] R. D. Young and E. W. Müller, Phys. Rev. **113**, 115 (1959).

[5] R. D. Young, Phys. Rev. **113**, 110 (1959).

[6] R. Fischer, Phys. Stat. Sol. **2**, 1088 (1962); **2**, 1466 (1962).

[7] R. Stratton, Phys. Rev. **125**, 67 (1962).

correct expressions for the total emission current are given.)

In the present paper general expressions for the total energy distributions are derived for arbitrary band structures and internal distribution functions, in momentum space, for the electrons just outside the semiconductor surface. The expressions for conduction-band emission reduce to those given by Fischer as special cases (either nondegenerate or degenerate electron gas; constant effective mass different from the free electron mass).

As a further extension of the theory, the normal energy distribution for both conduction and valence band emission has also been derived. This is of interest in connection with thin film tunneling. So far, the only retarding potential experiments on thin film tunneling have involved emission from aluminum. Kanter and Feibelman[8] measured the normal energy distribution of electrons emitted from an Al-Al_2O_3-Au sandwich into vacuum. Here the electrons which tunnel through the oxide layer emerge into the conduction band of the oxide and then pass through the conduction band of the outer or gold metal electrode before being emitted or captured. Thus, although the tunneling calculation gives the initial energy distribution, the final energy distribution of the emitted electrons will generally be broader and displaced to lower energies because they suffer collisions, involving energy losses, in the oxide and base metal electrode. Collins and Davies[9] have made similar measurements on Al-Al_2O_3-Al sandwiches but involving a small cathode and spherical collector. They interpret their results on the total energy distribution entirely in terms of the collision phenomena.

In the next section, the principles involved in determining energy distributions by means of the retarding potential analyzer are discussed to bring out the difficulties in obtaining unambiguous results for semiconductors. The calculated results and some of the few experimental measurements for field emission from semiconductors that have appeared in the literature are discussed in the final section.

2. RETARDING POTENTIAL MEASUREMENTS

The total energy distribution of field emitted electrons can be experimentally determined in a retarding-potential analyzer.[4] Initially, a hypothetical one-dimensional plane structure will be assumed since this greatly simplifies the explanation of the principle involved in the experiment. Figure 1 shows the electron potential energy diagram for emission from a metal. A high field is applied to the surface of the cathode by means of an anode in the form of a grid. The field emitted electrons which pass through the grid are then collected by the collector provided the electron kinetic

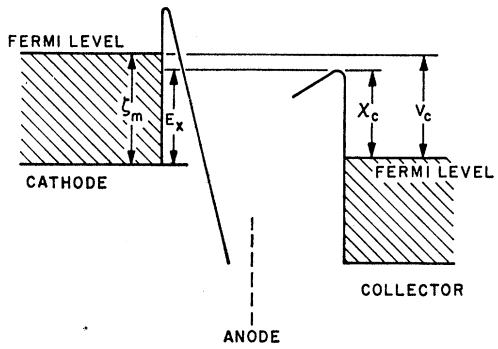

FIG. 1. Schematic electron potential energy diagram for the retarding potential analyzer with a metal emitter.

energy corresponding to motion in the x direction, normal to the plane of the collector, exceeds

$$E_x=\chi_c+\zeta_m-V_c \tag{1a}$$

or

$$E_x-\zeta_m=\chi_c-V_c, \tag{1b}$$

where ζ_m is the Fermi energy of the cathode, χ_c is the work function of the collector, V_c is bias (in energy units) between the cathode and the collector. If i_c is the current collected then

$$i_c=\int_{E_x}^{\infty} P_N(E_x')dE_x' \tag{2}$$

and

$$P_N(E_x)=-di_c/dE_x=di_c/dV_c \tag{3}$$

gives the normal energy distribution. Thus, $P_N(E_x)$ can be determined from the gradient of a plot of i_c versus V_c.

For a semiconductor with a clean surface, the electron potential energy diagram is given by Fig. 2 if the emission current is sufficiently weak to permit neglect of the internal potential drop arising from the bulk resistivity. An internal potential drop V_i has been assumed which depends on the applied field and the charge in surface states. Electrons emitted from the conduction band will be collected if their "x-directed" energy is in excess of

$$E_x=(\chi_c+\zeta_b+V_i)-V_c, \tag{4}$$

where ζ_b is the bulk Fermi energy of the semiconductor, measured with respect to the bottom of the conduction band. The current i_c collected will saturate to the value j_c of the total conduction band emission current when the collector bias is

$$V_{cn}=\chi_c+\zeta_b+V_i. \tag{5}$$

Thus, besides determining the normal energy distribution by means of Eq. (3), the retarding potential measurement also gives the value of any internal potential drop V_i.

If V_c is increased beyond V_{cn} by E_g, the forbidden

[8] H. Kanter and W. A. Feibelman, J. Appl. Phys. 33, 3580 (1963).

[9] R. E. Collins and L. W. Davies, Appl. Phys. Letters 2, 213 (1963).

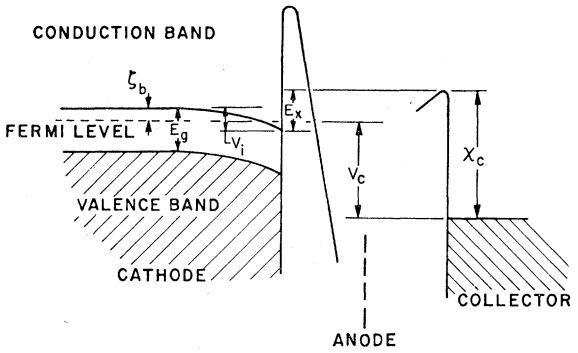

FIG. 2. Schematic electron potential energy diagram for the retarding potential analyzer with a semiconductor emitter having an internal potential drop V_i due to field penetration.

energy gap, electrons emitted from the top of the valence band can also be collected. Thus if V_c is greater than

$$V_{cp}=\chi_c+\zeta_b+V_i+E_g, \tag{6}$$

all the electrons emitted from the valence band whose "x-directed" energy, measured downwards from the top of the valence band, is less than

$$E_{vx}=V_c-(\chi_c+\zeta_b+V_i+E_g)=V_c-V_{cp} \tag{7}$$

will be collected. If i_v is the current collected, then

$$i_v=\int_0^{E_{vx}} P_N(E_{vx}')dE_{vx}' \tag{8}$$

and

$$P_N(E_{vx})=di_v/dE_{vx}=di_v/dV_c. \tag{9}$$

If the internal potential drop in the bulk of the semiconductor cannot be neglected the electron potential energy diagram will have the form shown in Fig. 3. Here it has been assumed that the semiconductor is in contact with a metal support and the collector bias is between the Fermi levels of the support and collector respectively. The metal-semiconductor work function is χ_{mn}. The internal potential drop now has contributions from three causes, field penetration (region 1), the bulk resistivity (region 2) and the metal-semiconductor barrier (region 3). In general, the boundaries of these regions are not well defined and the potential drop must be calculated for the semiconductor as a whole. However, in practice, estimates of the individual contributions may well be fairly good, particularly if one of them predominates.

The previous Eqs. (4) to (9) are also applicable to the situation depicted in Fig. 3 if ζ_b is replaced by $-\chi_{mn}$ which is the Fermi energy of the semiconductor in the vicinity of the metal support. Thus again the energy distribution and internal potential drop can be determined. This can be done even if the semiconductor surface is not clean. The presence of, say, an oxide layer will, of course, affect the tunnel probability and thereby the computed energy distribution but the experiment will still yield the internal potential drop and the actual energy distribution.

As pointed out in the Introduction, the use of a pointed field emission cathode in the retarding potential analyzer actually yields the total energy distribution. Hence, the "x-directed" energy E_x in the previous equations must be replaced by the total energy E and Figs. 1, 2, and 3 can be considered to give the electron potential variation along a line from the base of the field emission cathode to a point on its tip from which emission takes place. In general, the diagrams will be different for different locations of the point on this tip, i.e., both the internal potential V_i and the external field F will vary in a way determined by the cathode-anode geometry. This is one of the major difficulties in comparing measured field emission current-voltage characteristics with calculated current density-field characteristics. For metals ($V_i=0$) solutions of Poissons' equation for special geometries have been obtained which give the field distribution at the tip equipotential surface and good agreement between theory and experiment was found for the dependence of total field emitted current on the voltage applied between cathode and anode. The solutions can still be applied for semiconducting cathodes if the resistive potential drop *in the vicinity of the tip region* over which emission takes place is sufficiently small so that this portion of the cathode can be considered to have an equipotential surface.

In a well-constructed retarding-potential analyzer, only a small portion of the tip surface is used for collection. Thus, even if V_i and F vary along the surface of the tip, the experimental energy distribution should correspond to a definite value of F. However, particularly in the case of semiconductors, if electrons are simultaneously collected from various regions of the surface having different values of V_i and F, a series of overlapping energy distributions would lead to a very wide experimentally observed energy distribution. This might play a role in some of the experiments reported

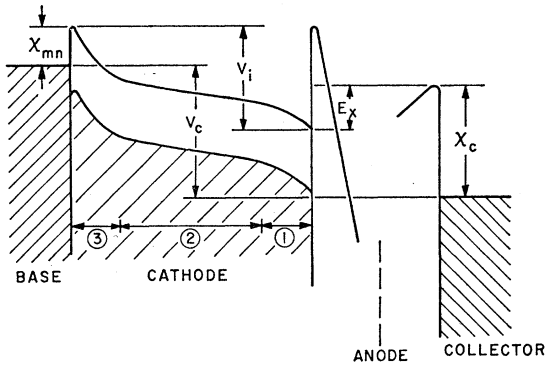

FIG. 3. Schematic electron potential energy diagram for the retarding potential analyzer with a semiconductor emitter having an internal potential drop V_i due to field penetration (region 1), the bulk resistivity (region 2) and the metal-semiconductor barrier (region 3).

where very wide distributions were ascribed to extremely large electron temperatures.

3. ELECTRON EMISSION FROM THE CONDUCTION BAND

3.1. Total Energy Distribution—Basic Equations

In this subsection general but rather formal equations will be derived which give the total energy distribution for arbitrary band structures and for the whole range of emission (field to thermionic) processes.

It will be assumed that for specular transmission through a barrier, depending only on the x coordinate, the transmission probability depends on the electron state only through the value of the "x-directed" energy *in the barrier region*. Thus, $D(E-E_\perp)$ is the probability that an electron of energy E incident on the surface barrier is transmitted, where

$$E_\perp = p_\perp^2/2m = (p_y^2+p_z^2)/2m \tag{10}$$

and p_y, p_z are the tangential components of the (conserved) quasielectron momentum components. This form of the transmission probability has been shown to apply for arbitrary band structures when the WKB approximation can be used to match the Bloch waves.[10]

The velocity of an electron, normal to the barrier, in the conduction band is

$$v_x = \partial E/\partial p_x. \tag{11}$$

Thus, if $f(E)$ is the electron distribution function in momentum space, the total energy distribution is given by

$$P_T(E)dE = \frac{2q}{h^3} f(E) \int^{E,E+dE} D(E-E_\perp) \frac{\partial E}{\partial p_x} \times dp_x dp_y dp_z, \tag{12}$$

where the integral is over the shell in momentum space lying between energies E and $E+dE$. (q=charge on the electron, h=Planck's constant.) This can be rewritten as

$$P_T(E) = (2q/h^3) f(E) \int D(E-E_\perp) dp_y dp_z, \tag{13}$$

where the integration in the p_y, p_z plane extends over all values corresponding to the energy surface E, i.e., p_y and p_z are inside the "shadow" of the energy surface E on a plane perpendicular to the x direction. Introducing polar coordinates $p_\perp$ and φ in the p_y, p_z plane and writing $E_m(E,\varphi)$ for the maximum value of $E_\perp = p_\perp^2/2m$ for a particular polar angle φ leads to

$$P_T(E) = K\frac{f(E)}{2\pi} \int_0^{2\pi} d\varphi \int_0^{E_m(E,\varphi)} D(E-E_\perp) dE_\perp \tag{14}$$

$$= Kf(E)\left[\int_0^E D(E_x)dE_x - \frac{1}{2\pi}\int_0^{2\pi} d\varphi \times \int_0^{E-E_m} D(E_x)dE_x\right], \tag{15}$$

where

$$E_x = E-E_\perp, \quad K = 4\pi mq/h^3. \tag{16}$$

In general, $D(E_x)$ will decrease very rapidly if E_x decreases. Thus, if $E_m(E,\varphi)$ is sufficiently large the second integral in Eq. (15) can be neglected and $P_T(E)$ is completely *independent of the conduction band structure*. The particular requirements for sufficiently large shadows will be discussed later; they must satisfy

$$D(E-E_m) \ll D(E). \tag{17}$$

Then also the upper limit $E_m(E,\varphi)$ in Eq. (14) and the lower limit 0 in the first integral in Eq. (15) will not affect the value of P_T. If this situation applies, Eq. (15) leads to

$$d[P_T(E)/f(E)]/dE = KD(E). \tag{18}$$

Thus, if $P_T(E)$ has been measured experimentally, Eq. (18) can be used to deduce the transmission probability if the distribution function is known. Alternatively, if the transmission probability is known, Eq. (15) can be used to deduce the distribution function $f(E)$ which could be of considerable interest for situations where $f(E)$ differs from the thermal equilibrium distribution due to high internal field effects.[11]

It must be emphasized that Eqs. (14) and (15) are general and apply to field, thermionic-field (T-F) or thermionic emission from a conductor of arbitrary band structure. In the Secs. 3.2 and 3.4 explicit results will be derived for field and T-F emission for a simple conduction-band structure.

3.2 Total Energy Distribution—WKB Approximation

Using the WKB approximation[10] gives

$$D(E_x) = \exp[-B(E_x)], \tag{19}$$

where

$$B(E_x) = 2[(2m)^{1/2}/\hbar]\int_{x_1}^{x_2} [\varphi(x)-E_x]^{1/2}dx > 1. \tag{20}$$

$\varphi(x)$ is the barrier potential energy, measured with respect to the conduction band edge and the limits x_1 and x_2 are the classical turning points.

For subsequent developments the Taylor expansion for $B(E_x)$ about an arbitrary energy $\mathcal{E}$ is required. This can be written as

$$B(E_x) = b(\mathcal{E}) - (E_x-\mathcal{E})c(\mathcal{E}) + (E_x-\mathcal{E})^2 a(\mathcal{E}) + \cdots, \tag{21}$$

where

$$b(\mathcal{E}) = B(\mathcal{E}), \quad c(\mathcal{E}) = -B'(\mathcal{E}), \quad a(\mathcal{E}) = B''(\mathcal{E})/2. \tag{22}$$

Murphy and Good[12] have derived the expressions for the coefficients appropriate to the image force barrier. General expressions, for potential barriers of arbitrary

[10] W. A. Harrison, Phys. Rev. **123**, 85 (1961).

[11] M. I. Elinson, Radiotekhn. i Elektron. 4, 140 (1959).

[12] E. L. Murphy and R. H. Good, Jr., Phys. Rev. **102**, 1464 (1956).

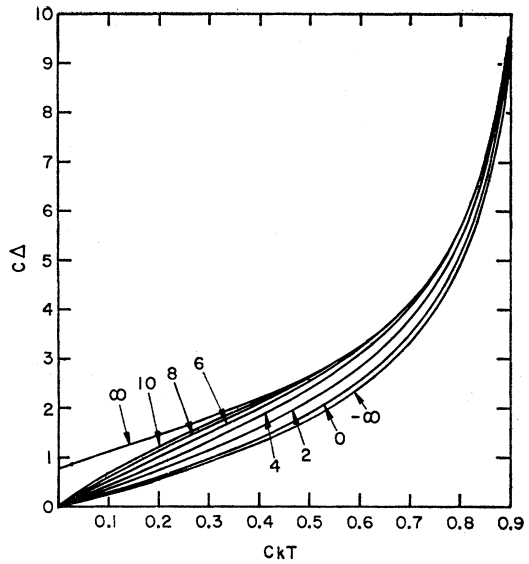

FIG. 4. Dependence of the reduced half-width $c\Delta$ of the total energy distribution on ckT for $m_c/m=1$. The numbers labeling each curve refer to the reduced Fermi energy ζ/kT.

shape, are given in Ref. 13. By inserting numerical values of the fundamental constants the expressions for the image force barrier can be written in the form

$$b=6.83\theta^{3/2}(10^7/F)v(\psi_i/\theta)\,,$$
$$c=10.25\theta^{1/2}(10^7/F)t(\psi_i/\theta)\ \text{eV}^{-1}, \tag{23}$$
$$a=2.56[\theta^{1/2}(F/10^7)\{1-(\psi_i/\theta)^2\}]^{-1}v(\psi_i/\theta)\ \text{eV}^{-2},$$
$$\theta=\psi-\mathcal{E}, \tag{24}$$

which gives some idea of their order of magnitude. Here, v and t are tabulated functions[14] whose magnitude is near one, ψ is the electron affinity of the cathode,

$$\psi_i=1.2(F/10^7)^{1/2}(\kappa-1)^{1/2}(\kappa+1)^{-1/2} \tag{25}$$

is the depression of the barrier height due to the image force, κ is the dielectric constant, energies are in eV and F is in V/cm.

Inserting the expression for $D(E_x)$ into Eq. (15) for $P_T(E)$ gives

$$P_T(E)=Kf(E)\frac{e^{-b}}{2\pi}\int_0^{2\pi}d\varphi\int_{E-E_m}^{E}e^{c(E_x-\mathcal{E})}dE_x, \tag{26}$$

provided that the quadratic and higher order terms on Eq. (21) can be neglected, say

$$a(E_x-\zeta)^2<\tfrac{1}{2}. \tag{27}$$

Since the major contribution to the integral over E_x comes from E_x values near E we choose $\mathcal{E}$ equal to E. Then the important range of integration is given by

$$0<E-E_x\lesssim 1/c,$$

so that the quadratic term in Eq. (21) is less than a/c^2 in this range. From Eq. (23)

$$a/c^2\approx 0.0244(F/10^7)/\theta^{3/2}, \tag{28}$$

which will always be considerably less than one. Thus, carrying out the integration,

$$P_T(E)=Kf(E)\frac{e^{-b(E)}}{c(E)}\left[1-\frac{1}{2\pi}\int_0^{2\pi}e^{-c(E)E_m}d\varphi\right], \tag{29}$$

if

$$a(E)/[c(E)]^2<\tfrac{1}{2}. \tag{30}$$

The specific band structure of the cathode affects only the integral which can be neglected when

$$c(E)E_m(E,\varphi)\gg 1\,, \tag{31}$$

i.e., a specific form of inequality (17). Since $c(E)$ has a magnitude of about 10 eV^{-1} only very "small" shadows will affect the energy distribution. This is unlikely to occur for metals with a large positive Fermi energy but band structure effects may be important for semiconductors and semimetals.

Fischer has derived $P_T(E)$ for the separate cases of positive and negative Fermi energies by expanding $b(E_x)$ about the Fermi level and conduction band edge respectively following Ref. 13. He uses a simplified version of Eq. (15), based on spherical energy surfaces with an effective mass $m_c=r_cm$ so that $E_m(E,\varphi)$ is then equal to r_cE. His results can be written in the more convenient forms

$$P_T(E)=(K/c_1)e^{-(b_1+c_1\zeta)}f(E)e^{c_1E}(1-e^{-r_cc_1E}), \tag{32}$$

if

$$a_1(E-\zeta)^2<\tfrac{1}{2},\quad \zeta>0 \tag{33}$$

$$P_T(E)=(K/c_0)e^{-b_0}f(E)e^{c_0E}(1-e^{-r_cc_0E}), \tag{34}$$

if

$$a_0E^2<\tfrac{1}{2},\quad \zeta<0, \tag{35}$$

which can also be derived from Eq. (29). Here $c_1\equiv c(\zeta)$ and $c_0=c(0)$, etc. The quantity $c(E)$, which is approximately equal to $c_1-2a_1(E-\zeta)$, has been replaced by c_1. The term $\exp[-r_cc(E)E]$ is only of importance if $r_cc_1E\lesssim 1$ when $2a_1|E-\zeta|r_cE<(2a_1)^{1/2}/c_1$ which is small [cf. inequality (28)]. Similarly, the factor $c(E)$ in the denominator can be replaced by c_1 and, for a negative Fermi energy, $c(E)$ can be replaced by c_0. Apart from the last factor in brackets Eq. (32) is the result first obtained by Young[5] for the case of metals.

By differentiating Eq. (32) it can be shown that the energy distribution peaks at an energy E_p which satisfies the equation

$$c_1(E_p-\zeta)=-c_1kT\ln[\{(1+\beta_1)c_1kT\}^{-1}-1],\ \zeta>0, \tag{36}$$

where

$$\beta_1=r_c(e^{r_cc_1E_p}-1)^{-1}, \tag{37}$$

and $f(E)$ is the Fermi-Dirac distribution function corresponding to a temperature T. [It should be

[13] R. Stratton, Phys. Chem. Solids **23**, 1177 (1962).
[14] R. E. Burgess, H. Kroemer, and J. M. Houston, Phys. Rev. **90**, 515 (1953).

observed that the conditions for which Eq. (32) applies will require $c_1kT<1$.[7,12] This actually is required to make $P_T(E)$ decrease for large E so that a peak E_p exists.]

From Eq. (37), β_1 decreases from a maximum value of $(c_1E_p)^{-1}$ as r_c increases from zero. Thus β_1 is very much less than one and can be neglected when

$$|c_1\zeta - c_1kT\ln\{(c_1kT)^{-1}-1\}|\gg 1. \tag{38}$$

Except for the limiting case where c_1kT tends to one this requires that $c_1\zeta$ be large compared to one. When β_1 can be neglected Eq. (36) is an explicit equation for E_p which shows that as c_1kT increases from zero, $c_1(E_p-\zeta)$ decreases from zero to a minimum value of $-(e+1)^{-1}$ at $c_1kT=(1+e)^{-1}$, then increases to zero when $c_1kT=\frac{1}{2}$ and then tends to infinity logarithmically as c_1kT approaches one.

Besides determining the location of the peak it is also of interest to calculate the half-width Δ of the distribution. Thus, if

$$P_T(E_{1,2})=\tfrac{1}{2}P_T(E_p), \tag{39}$$

then

$$\Delta=E_2-E_1. \tag{40}$$

An analytic expression for Δ can only be found in the limit of zero temperatures when

$$c_1\Delta \rightarrow \ln[2/(c_1+e^{-c_1\zeta})]. \tag{41}$$

For all other values of the parameters E_p and Δ must be determined numerically. The results of these computations will be discussed after the case of negative Fermi energies has been treated.

The conditions (20) and (33) for which Eq. (32) applies have been previously considered[7,12] in connection with the calculation of the total emission current for an image force barrier. Since the latter essentially involves an integration over the normal energy distribution the limits must be re-examined for the total energy distribution. For the case where the band structure term in Eq. (32) can be neglected it can be shown that approximately the previous limits on the field, derived in Ref. 13, are recovered. If the band structure term is important the inequalities (20) and (33) must be checked by inserting the numerical values for E_1 and E_2.

Since Eqs. (32) and (34) have identical analytical forms, Eq. (36) for E_p applies for the case of negative Fermi energies with c_1 replaced by c_0. However, it is more convenient to express the result in the form

$$E_p=\frac{1}{r_cc_0}\ln\left[1+\frac{r_cc_0kT}{1-c_0kT-f(E_p)}\right], \quad \zeta<0 \tag{42}$$

which gives an explicit relation for E_p in the classical limit when $f(E_p)\ll 1$. (The limiting relation has already been given by Fischer.[6]) If this applies, E_p increases monotonically from kT as c_0kT increases from zero.

There is no simple analytical expression for Δ which applies for any values of the parameters when the Fermi energy is negative.

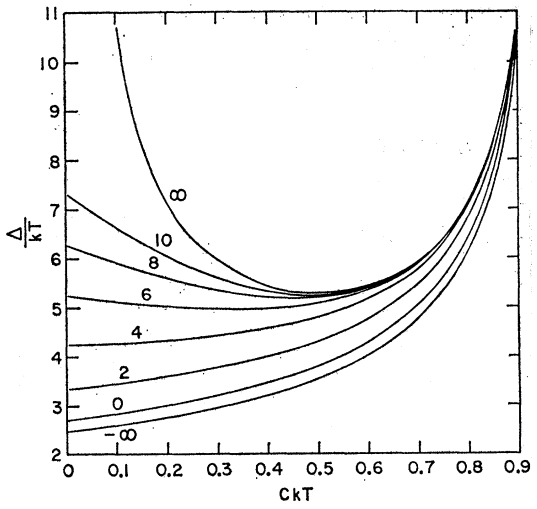

FIG. 5. Dependence of the reduced half-width Δ/kT of the total energy distribution on ckT for $m_e/m=1$. The numbers labeling each curve refer to the reduced Fermi energy ζ/kT.

It can again be shown that the conditions (20) and (35) for which Eq. (34) is valid lead to roughly the same restrictions on the value of F which were previously derived[13] for the normal energy distribution.

The peak values E_p and half-widths Δ have been derived for a variety of parameters by means of a numerical analysis of Eqs. (32) and (34). The results are given in the form of universal curves in Figs. 4 to 8. The numbers labeling each curve refer to the reduced Fermi energy (ζ/kT). For negative Fermi energies only the results for $(\zeta/kT)\rightarrow-\infty$ have been exhibited since the curves for $(\zeta/kT)=-2$ are already close to the limiting curves. Figure 4 gives the variation for Δ as a function of temperature T at a constant field (i.e., c is constant) for $r_c=1$. The half-width increases with T for all values of (ζ/kT). Actually, for a fixed ζ, (ζ/kT) decreases as T increases but Δ still increases. For a more complicated dependence of ζ on T additional numerical work would be required. Figure 5 (for $r_c=1$) gives the variation of Δ as a function of the applied field F at constant T since c is essentially inversely proportional to F. It is of interest that for fields exceeding the condition $ckT\approx\frac{1}{2}$ (i.e., the usual range in practice) Δ increases with F for large positive Fermi energies and decreases with F for small positive and negative Fermi energies. Figure 6 shows the effect of different effective masses on Δ. This is most noticeable for negative Fermi energies at large values of ckT.

The dependence of E_p/kT on ckT for $r_c=1$ is given in Fig. 7. Thus, for T constant, E_p decreases as F increases and for F constant, E_p increases as T increases. The variation of E_p with respect to r_c is similar to that of Δ, it is appreciable for negative Fermi energies and large values of ckT. This behavior is illustrated by the

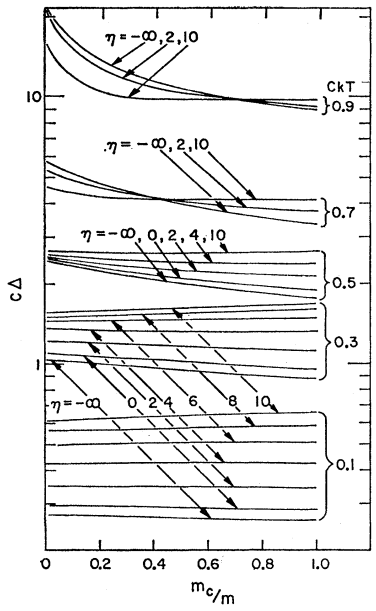

FIG. 6. Dependence of the reduced half-width $c\Delta$ of the total energy distribution on the reduced effective mass m_c/m for various values of the reduced Fermi energy $\eta=\zeta/kT$ and ckT.

dashed curves in Fig. 7 which correspond to the extreme value $r_c\to 0$ and $(\zeta/kT)=-\infty$, 4, and 10, respectively.

3.3 Normal Energy Distribution

Assuming the properties of the transmission probability outlined in Sec. 3.1 and using Eqs. (10) and (11) gives

$$P_N(E_x)dE_x=\frac{2q}{h^3}D(E_x)\int^{E_x,E_x+dE_x} f(E)\frac{\partial E}{\partial p_x}\times dp_x dp_y dp_z. \quad (43)$$

This is similar to Eq. (12) for $P_T(E)$ except for the region of integration which is now the shell in momentum space for which the "x-directed" energy is in the range E_x, E_x+dE_x. The equation can be rewritten as

$$P_N(E_x)=L\frac{D(E_x)}{2\pi}\int_0^{2\pi}d\varphi\int f(E_\perp+E_x)d(E_\perp/kT), \quad (44)$$

$$L=4\pi mqkT/h^3=KkT, \quad (45)$$

where the $E_\perp$ integration is over all values which satisfy

$$0\leqslant E_\perp\leqslant E_m(E,\varphi)\equiv E_m(E_\perp+E_x,\ \varphi) \quad (46a)$$

or

$$E_x\leqslant E_x+E_\perp\equiv E\leqslant E_m(E,\varphi)+E_x, \quad (46b)$$

for a given value of E_x. Now $E_m(E,\varphi)$ is non-negative. Thus for E_x greater than zero, there will be a range of integration for $E_\perp$ which extends from 0 to $E_{\perp u}$ where $E_{\perp u}$ either satisfies the relation

$$E_{\perp u}=E_m(E_{\perp u}+E_x,\ \varphi) \quad (47)$$

or

$$E_{\perp u}+E_x\equiv E_u=E_m(E_u,\varphi)+E_x, \quad (48)$$

or is infinite if no solution to Eq. (47) exists. The latter will arise if $E_m(E,\varphi)$ is always greater than E. For a complicated form of the function $E_m(E,\varphi)$ there could be additional ranges of integration corresponding to more than one solution of Eq. (47). These will generally lead to small contributions to the integral and can be neglected.

Carrying out the integration over $E_\perp$ and inserting the limits gives

$$P_N(E_x)=LD(E_x)\Big[\ln\{1+\exp(\langle\zeta-E_x\rangle/kT)\}-(1/2\pi)\int_0^{2\pi}\ln\{1+\exp(\langle\zeta-E_u\rangle/kT)\}d\varphi\Big],\quad E_x>0. \quad (49)$$

If there are ranges of E for which $E_m(E,\varphi)$ exceeds E, negative E_x values are possible. Then the lower limit on $E_\perp$ is given by the smallest positive solution of Eq. (47) say $E_{\perp u1}$ and the upper limit is given by the second smallest solution, say $E_{\perp u2}$ or infinity. Thus

$$P_N(E_x)=\frac{LD(E_x)}{2\pi}\times\int_0^{2\pi}\ln\left[\frac{1+\exp\{(\zeta-E_{u1})/kT\}}{1+\exp\{(\zeta-E_{u2})/kT\}}\right]d\varphi,\quad E_x<0 \quad (50)$$

where $E_{u1}=E_{\perp u1}+E_x$, etc.

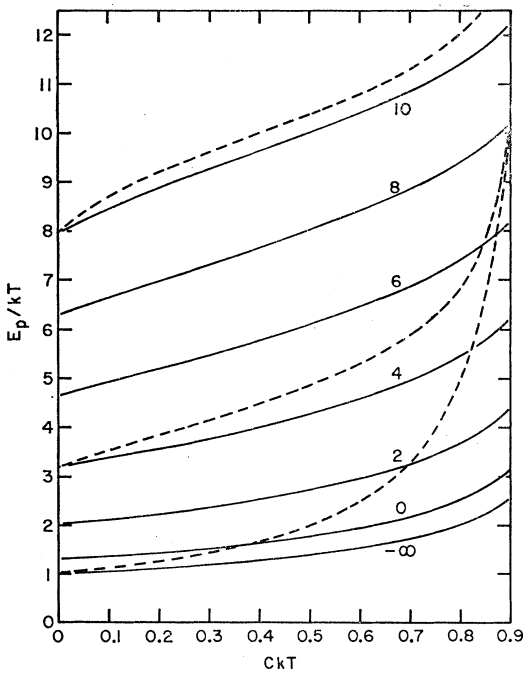

FIG. 7. Dependence of the reduced energy E_p/kT, at which the total energy distribution has a peak value, on ckT. Full line curves; $m_c/m=1$, dashed curves; $m_c/m\to 0$. The numbers labeling each curve are values of the reduced Fermi energy ζ/kT.

Considering again the example of spherical energy surfaces in the conduction band,

$$E_{\perp u}=E_x r_c(1-r_c)^{-1}, \quad E_u=E_x(1-r_c)^{-1}. \tag{51}$$

Suppose first that r_c is less than one. Then only positive E_x values are possible and the following special forms of Eq. (49) can be deduced.

If

$$1\ll(\zeta-E_u)/kT\ll(\zeta-E_x)/kT, \tag{52}$$

$$P_N=LD(E_x)(E_x/kT)r_c(1-r_c)^{-1}. \tag{53}$$

If

$$1\ll-(\zeta-E_u)/kT, \quad (\zeta-E_x)/kT, \tag{54}$$

$$P_N=LD(E_x)(\zeta-E_x)/kT. \tag{55}$$

If

$$1\ll-(\zeta-E_u)/kT\ll-(\zeta-E_x)/kT, \tag{56}$$

$$P_N=LD(E_x)\left[\exp\left(\frac{\zeta-E_x}{kT}\right)-\exp\left(\frac{\zeta-E_x(1-r_c)^{-1}}{kT}\right)\right]. \tag{57}$$

The intermediate range, Eq. (54), only exists if $(\zeta/kT)\gg r_c^{-1}$. For negative Fermi energies only the last range, Eq. (56), exists. If $r_c=1$, E_u is infinity and the band structure term in Eq. (49) vanishes. This result has been derived by Müller.[3] Only the ranges corresponding to Eqs. (54) and (56) then apply. If r_c is greater than one, E_u is also infinite for positive E_x values. However, negative E_x values are then also possible with

$$E_{u1}=-E_x(r_c-1)^{-1}, \quad E_{u2}\to\infty \tag{58}$$

leading to

$$P_N=LD(E_x)\ln\left[1+\exp\left(\frac{\zeta+E_x(r_c-1)^{-1}}{kT}\right)\right], \quad E_x<0. \tag{59}$$

Next the location of the peak E_{xp} and half-width of the normal energy distribution will be considered. The case $r_c\geq 1$ when $E_u=\infty$ will be dealt with first. Differentiating Eq. (49) gives

$$[c(E_{xp})kT]^{-1}=\ln[1+\exp\{(\zeta-E_{xp})/kT\}]/f(E_{xp}). \tag{60}$$

[This relation is given in Ref. 12 with $c(E_{px})$ replaced by c_1.] The condition that $E_{px}>0$ requires

$$(c_0kT)^{-1}\leqslant\ln[1+\exp(\zeta/kT)]/f(0), \tag{61}$$

which puts a lower limit on ζ, say $\zeta_0(c_0kT)$, for a given value of c_0kT. Thus, if $\zeta\geq\zeta_0$

$$E_{xp}=\zeta-c_1^{-1} \quad \text{if} \quad c_1kT\ll 1 \tag{62}$$

$$=\zeta-kT\ln\{2(1-c_1kT)\} \quad \text{if} \quad 1-c_1kT\ll 1, \tag{63}$$

replacing c_p by c_1 following Ref. 12. These equations correspond to the ranges specified by Eqs. (54) and (56), respectively.

A simple expression for the half-width can only be derived for the case $c_1kT\ll 1$ when, using Eq. (39), E_{x1} and E_{x2} are the solutions of

$$\exp\{c_1(\zeta-E_x)\}/\{c_1(\zeta-E_x)\}=2e, \tag{64}$$

so that

$$\Delta_x=E_{x2}-E_{x1}=2.45c_1^{-1}. \tag{65}$$

Comparing with Eq. (41) we see that (Δ_x/Δ) is about 3.5.

If $\zeta<\zeta_0(c_0kT)$, $P_N'(E_x)$ is negative for all positive E_x. But $P_N'(E_x)$ is positive for all negative E_x (if $r_c>1$) according to Eq. (59). Thus the normal energy distribution will have a cusp at $E_{px}=0$ with

$$P_N(0)=LD(0)[1+\exp(\zeta/kT)], \tag{66}$$

$$P_N'(0+)=\frac{LD(0)}{kT}[-f(0)+c_0kT\ln(1+e^{\zeta/kT})], \tag{67}$$

$$P_N'(0-)=\frac{LD(0)}{kT}\left[\frac{f(0)}{(r_c-1)}+c_0kT\ln(1+e^{\zeta/kT})\right]. \tag{68}$$

A simple result for the half-width exists only for the limit $-\zeta/kT\gg 1$ when

$$\Delta_x=kT\left[\frac{1}{1-c_0kT}+\frac{r_c-1}{r_cc_0kT+(1-c_0kT)}\right]\ln 2. \tag{69}$$

For the case $r_c<1$, E_u is finite and given by Eq. (51). The location of the peak is then given by

$$\frac{1}{c_pkT}=\frac{f(E_{xp})-f(E_{xp}[1-r_c]^{-1})(1-r_c)^{-1}}{\ln\left[\frac{1+\exp\{(\zeta-E_{xp})/kT\}}{1+\exp\{(\zeta-E_{xp}[1-r_c]^{-1})/kT\}}\right]}. \tag{70}$$

This reduces to Eq. (60) if the band structure terms are neglected which will be possible for sufficiently large positive Fermi energies. In the opposite extreme of classical degeneracy,

$$\frac{E_p}{kT}=\frac{1-r_c}{r_c}\ln\left[1+\frac{r_c(1-r_c)^{-1}}{1-c_0kT}\right], \quad -\frac{\zeta}{kT}\gg 1. \tag{71}$$

Thus, E_p decreases from $kT(1-c_0kT)^{-1}$ to zero as r_c increases from zero to one. (There are no simple results for Δ_x in this case.)

For the case $r_c<1$, $P_N(0)=0$ which should be contrasted with Eq. (66) for $r_c\geq 1$. This discontinuity seems unphysical. What actually happens is that as r_c tends to one from below, $P_N'(0)$ tends to infinity, the peak location tends to zero and the maximum value $P_N(E_{xp})$ tends towards $P_N(0)$ given by Eq. (66). Thus there is a continuous change of shape in the peak of the distribution from a smooth maximum ($r_c<1$) to a cusp ($r_c\geq 1$).

3.4. T-F Emission from the Conduction Band

The necessary criteria for field emission, $c_1kT<1$ if $\zeta>0$ and $c_0kT<1$ if $\zeta<0$, will not be obeyed if the applied field is too low or the temperature is too high. Equations (32) and (34) are then not applicable since $P_T(E)$ is still increasing at the greatest value of E for which they are valid. Another region of approximation then applies which requires that the quadratic term in the expansion of $b(E)$ be retained since the peak of the distribution is now well above either the Fermi level ($\zeta>0$) or the bottom of the conduction band. Thus, expanding about the peak of the distribution,

$$P_T=\frac{K}{c_p}\exp\left[\frac{\zeta}{kT}-b_p-c_pE_p-\frac{E}{kT}(1-c_pkT)-a_p(E-E_p)^2\right]\times[1-\exp\{-r_c[c_p-2a_p(E-E_p)]E\}], \quad (72)$$

if

$$\exp\{(E-\zeta)/kT\}\gg 1, \quad (73)$$

where

$$c_pkT=[1+\beta_p\{1-2(a_p/c_p)E_p\}]^{-1}\approx[1+\beta_p\{1-\tfrac{1}{2}E_p/(\psi-E_p)t_p^3\}]^{-1}. \quad (74)$$

and $c_p\equiv c(E_p)$ etc. [Eq. (84) has been deduced using Eq. (23) for $c(E)$ and $a(E)$ and the approximate relation, $v(y)t^2(y)\approx 1-y^2$, given by Murphy and Good.[12]] The term involving the factor β_p in Eq. (74) can actually always be neglected, so that

$$c_pkT=1, \quad (74a)$$

since $\beta_p\geq(c_pE_p)^{-1}$ [cf. Eq. (37)] so that $\beta_p\ll 1$ requires $(E_p/kT)\gg 1$ which is consistent with inequality (73). For an image force barrier, using Eq. (23) for $c(E_p)$, the "almost" explicit relation

$$E_p=\psi[1-(t_0/t_p)^2(c_0kT)^{-2}], \quad (75)$$

where only t_p involves E_p on the right hand side, can be derived.

Substituting into Eq. (72) for c_pkT from Eq. (74a), neglecting the band structure term, gives

$$P_T(E)=(K/c_p)e^{\zeta/kT}e^{-a_p(E-E_p)^2}; \quad (76)$$

a Gaussian distribution with a half-width

$$\Delta=2(\ln 2)^{1/2}a_p^{-1/2}. \quad (77)$$

This can also be written in the form

$$c_0\Delta=t_p^{1/2}t_0(\ln 2)^{1/2}(\psi/kT)^{1/2} \quad (77a)$$

for an image force barrier, which shows that Δ is essentially proportional to $F/T^{1/2}$.

The normal energy distribution [cf. Eq. (49)] for the T-F emission range is *identical* with the total energy distribution given by Eq. (76) if the band structure terms can be neglected.

The conditions on F and T for which the T-F region applies are given in Refs. 12 and 13. In addition, it can be shown that the cubic term in the expansion for $b(E)$, neglected in Eq. (72), is small. If the cubic term is written as $l(\mathcal{E})(E-\mathcal{E})^3$ then

$$l(\mathcal{E})=\frac{0.43[-3t+4v\{1-(\psi_i/\theta)^2\}^{-1}]}{\theta^{3/2}(F/10^7)\{1-(\psi_i/\theta)^2\}}\text{eV}^{-3} \quad (78)$$

$$\approx\frac{0.43}{(F/10^7)\theta^{3/2}}\left[\frac{4-3t^3}{vt^4}\right]\text{eV}^{-3}. \quad (78a)$$

Thus

$$l_p(\tfrac{1}{2}\Delta)^3\approx 0.06(F/10^7)^{1/2} \quad (79)$$

and is usually considerably less than one.

4. FIELD EMISSION FROM THE VALENCE BAND

4.1. Total Energy Distribution—Basic Equations

It will again be assumed that for specular transmission, through a one-dimensional barrier, the transmission probability is a function of only the "x-directed" energy in the *barrier region* and can be written as $D_v(E_v+E_{v\perp})\equiv D_v(E_{vx})$ where

$$E_{v\perp}=p_{v\perp}^2/2m=(p_{vy}^2+p_{vz}^2)/2m, \quad (80)$$

and E_v is *minus* the electron energy with respect to the top of the valence band as zero.[15] The sign convention is such that E_v and E_{vx} are positive. Further, as E_{vx} increases, the barrier height for tunneling increases, the argument of D_v increases and the magnitude of D_v decreases. This is opposite to the behavior of $D(E_x)$ which increases as E_x increases. In fact

$$D_v(E_{vx})\equiv D(-E_g-E_{vx}). \quad (81)$$

With this definition of D_v, the total energy distribution is given by

$$P_T(E_v)=\frac{f_v(E_v)}{2\pi}\int_0^{2\pi}d\varphi_v\int_0^{E_{vm}}D_v(E_v+E_{v\perp})dE_{v\perp} \quad (82)$$

$$=Kf_v(E_v)\left[\int_{E_v}^{EV}D_v(E_{vx})dE_{vx}-(1/2\pi)\int_0^{2\pi}d\varphi_v\int_{E_v+E_{vm}}^{EV}D(E_{vx})dE_{vx}\right], \quad (83)$$

where $E_{vm}(E_v,\varphi)$ is the maximum value of $E_{v\perp}$ and E_V is the width of the valence band.

From the form of Eq. (82) it is clear that the major

[15] In Ref. 7 the tunneling probability was erroneously written as $D_v(-E_v+E_{v\perp})=D_v(-E_{vx})$. Apart from the fact that we now *choose* to define D_v with a *positive argument*, the signs in front of E_v and $E_{v\perp}$ must be the same. As a result, some of the equations in that paper need modification and the final result for the emitted current density from a filled band must be multiplied by $(1-r_v)(1+r_v)^{-1}$. The necessary corrections and more general results for the emitted current density are given in the Appendix.

contribution to the total energy distribution again comes from electrons with small values of $E_{v\perp}$. If E_{vm} is sufficiently large, the second integral in Eq. (83) can be neglected and $P_{Tv}(E_v)$ is independent of the band structure.

4.2. Total Energy Distribution—WKB Approximation

Using the WKB approximation

$$D_v(E_{vx})=\exp[-B_v(E_{vx})], \tag{84}$$

where

$$B_v(E_{vx})=2\frac{(2m)^{1/2}}{\hbar}\int_{x_1}^{x_2}[\varphi(x)+E_g+E_{vx}]^{1/2}dx \tag{85}$$

$$=b_v(\mathcal{E}_v)+(E_{vx}-\mathcal{E}_v)c_v(\mathcal{E}_v)+(E_{vx}-\mathcal{E}_v)^2a_v(\mathcal{E}_v). \tag{86}$$

Here

$$b_v(\mathcal{E}_v)=B(\mathcal{E}_v),\quad c_v(\mathcal{E}_v)=B'(\mathcal{E}_v),\quad a_v(\mathcal{E}_v)=\tfrac{1}{2}B''(\mathcal{E}_v) \tag{87}$$

are given by Eq. (23), for an image force barrier, if θ is redefined as

$$\theta_v=\psi+E_g+\mathcal{E}_v. \tag{88}$$

Substituting into Eq. (82), choosing $\mathcal{E}_v$ equal to E_v, leads to

$$P_T(E_v)=Kf_v(E_v)\frac{e^{-b_v(E_v)}}{c_v(E_v)}\times\left[1-\frac{1}{2\pi}\int_0^{2\pi}d\varphi_v e^{-c_v(\mathcal{E}_v)E_{vm}}\right] \tag{89}$$

if

$$a_v(E_v)/[c_v(E_v)]^2<\tfrac{1}{2}.$$

The dominant factor here is $\exp[-b_v(E_v)]$, i.e., $D_v(E_v)$.

For the model of spherical energy surfaces with an effective mass $m_v=r_vm$, $E_{vm}(E_v,\varphi)$ reduces to r_vE. Following the treatment for conduction-band emission, the two cases of degenerate ($\zeta_v>0$) and nondegenerate ($\zeta_v<0$) *hole* distributions will be considered separately. Here ζ_v is *minus* the Fermi energy with respect to the top of the valence band as zero (i.e., $\zeta+\zeta_v=-E_g$). Expanding about the Fermi level and the top of the valence band respectively gives

$$P_T(E_v)=(K/c_{v1})f_v(E_v)e^{-(b_{v1}+c_{v1}\zeta_v)}\times e^{-c_{v1}E_v}[1-e^{-r_vc_{v1}E_v}],$$

if

$$a_{v1}(E_v-\zeta_v)^2<\tfrac{1}{2},\quad \zeta_v>0, \tag{90}$$

$$P_T(E_v)=(K/c_{v0})f_v(E_v)e^{-b_{v0}-c_vE_v}[1-e^{-r_vc_{v0}E_v}],$$

if

$$a_{v0}E_v^2<\tfrac{1}{2},\quad \zeta_v<0, \tag{91}$$

where $b_{v0}\equiv b_v(0)$ and $b_{v1}\equiv b_v(\zeta_v)$, etc.

Fischer has given a result which should be appropriate for the case $\zeta_v\to-\infty$ when $f_v(E_v)\to 1$. His result is, however, in error since he used the incorrect value of $D_v(E_v)$ in our earlier paper.[7] The correct expression for $P_T(E_v)$ has a peak given by

$$E_{vp}=(r_vc_{v0})^{-1}\ln(1+r_v) \tag{92}$$

and half-width which can be determined from our numerical calculations for conduction band emission. This follows since Eq. (91), with $f(E_v)=1$, has the same analytical form as Eq. (34), with $\eta\to-\infty$, if c_0kT is $\frac{1}{2}$ and r_c, c_0 are replaced by r_v, c_{v0}. Thus, from Fig. 6, $c_{v0}\Delta$ goes from 1.76 to 2.43 as r_v goes from 1 to 0.

The peak energy and half-width for arbitrary values of ζ_v could easily be derived as for the case of conduction band emission in Sec. 3. The results will not be presented here since the case $f_v(E_v)\approx 1$ is probably the most important in practice.

4.3. Normal Energy Distribution

By analogy with the result for conduction-band emission [cf. Eq. (44)]

$$P_v(E_{vx})=\frac{L}{2\pi}D_v(E_{vx})\int_0^{2\pi}d\varphi_v\int f_v(E_{vx}-E_{v\perp})\frac{dE_{v\perp}}{kT}, \tag{93}$$

where the $E_{v\perp}$ integration is over all values which satisfy

$$0\leqslant E_{v\perp}\leqslant E_{vm}(E_v,\varphi_v)\equiv E_{vm}(E_{vx}-E_{v\perp},\varphi_v). \tag{94}$$

Thus

$$P_N(E_{vx})=LD_v(E_{vx})\Big[\ln\{1+\exp(\langle E_{vx}-\zeta_v\rangle/kT)\}-(1/2\pi)\int_0^{2\pi}d\varphi_v\ln\{1+\exp(\langle E_{vl}-\zeta_v\rangle/kT)\}\Big], \tag{95}$$

where

$$E_{vm}(E_{vl},\varphi)=E_{vx}-E_{vl}. \tag{96}$$

In the limit of a completely filled band,

$$P_N(E_{vx})=KD_v(E_{vx})\left[E_{vx}-(2\pi)^{-1}\int_0^{2\pi}E_{vl}d\varphi_v\right]. \tag{95a}$$

In the spherical energy surfaces approximation

$$P_N(E_{vx})=KD_v(E_{vx})E_{vx}r_v(1+r_v)^{-1}, \tag{95b}$$

which has a peak at $E_{vxp}=c_{v0}^{-1}$ and a half-width equal to $2.45\ c_{v0}^{-1}$.

5. CONCLUSIONS

The numerical results presented in Sec. 3 indicate that the half-widths of the total energy distributions for electrons emitted from a conduction band are in the range of about 2 kT to 7 kT for ckT [cf. Eq. (22) and (23)] in the typical range of 0.2 to 0.5 and for all degeneracies. For valence-band emission the half-widths are in the range of about 3 kT to 12 kT for the same ckT

range. The experiments to be described have invariably led to distributions which are much wider than these predicted results if T is assumed to be the ambient temperature. This has led to a description of the results in terms of the hot electron process.[16] It should, however, be observed that limited resolution, bad alignment and emission from too large a tip area (cf. Sec. 2) will all lead to *spuriously wide distributions*; problems that plagued the early work on tungsten.

The assumption of specular as opposed to diffuse electron transmission (cf. Sec. 3) should apply reasonably well to clean semiconducting emitters where the tip surface consists of several flat regions corresponding to low index crystal planes which join smoothly. However, as pointed out by Harrison,[10] the converse might be the case for thin film tunneling since measurements of thin film sheet resistance indicate that electron reflection at the surface is predominantly diffuse[17] in many cases although predominantly specular reflection has also been observed.[18]

Russell[19,20] measured the total energy distribution field emitted from a Si tip in a spectrometer said to have a resolution[20] "lower than Young and Müller's[4] by a factor three" (i.e., from 0.06 to 0.09 eV). However, the width of the energy distribution from a tungsten tip was found to be 0.6 eV; about three times the correct value as measured by Young and Müller.[4] This would indicate that there are sources of experimental errors in Russel's equipment other than a relatively low resolution. [*Note added in proof.* Recent calculations (to be published) have shown that for nonspecular boundary conditions, the integrand in Eq. (16) for $P_T(E)$ contains an additional factor $U(E,\mathbf{p}_t)=\Sigma\rho(\mathbf{p}_t')\pi(\mathbf{p}_t',\mathbf{p}_t)/\rho(\mathbf{p}_t)$ where $\rho(\mathbf{p}_t)$=linear density of states for fixed tangential momentum $\mathbf{p}_t$, $\pi(\mathbf{p}_t',\mathbf{p}_t)$ is proportional to the probability that the diffuse boundaries change $\mathbf{p}_t'$ to $\mathbf{p}_t$ and the sum is over the "shadow" of the energy surface $E(\mathbf{p})$. Under reasonable assumptions for $\pi(\mathbf{p}_t',\mathbf{p}_t)$, no appreciable effect of the densities of state on U or on the field emission current should occur.]

Russell[19] finds that as the field on the tip is increased till emission occurs, total energy distributions are found which are similar to those from tungsten except that the critical collector bias is greater. In fact, up to a certain applied voltage, all the electrons emitted originate from the valence band with a constant internal potential drop V_i and the quantity $V_{cp}-\chi_c=(\zeta_b+E_g)+V_i\approx 0.4$ eV. Thus, the Fermi level at the surface is about 0.15 eV below the center of the gap, i.e., a p-type surface. Although the relevant parameters are not known, the observed width of the distribution is again far too wide. As the applied field is increased Russel finds that V_{cp} (or V_i) increases by over 3 V. As pointed out by Russell, the absence of any emission from the conduction band indicates that this increase in V_i cannot arise from pure field penetration as shown in Fig. 2. Russell therefore suggests an IR drop as indicated in Fig. 3. For a constant internal resistance this would imply that the logarithm of the increase in V_i is a linear function of the reciprocal applied voltage, assuming the usually observed empirical field emission relation. An attempt to roughly fit Russell's data in this way proves unsuccessful in that V_i increases much too fast as the applied voltage increases. Possibly part of the increase in V_i does correspond to field penetration. (In a private communication, however, Russell has informed the author that more complete measurements, involving also total emitted current, tend to verify his original suggestion that the increase in V_i can be accounted for by an IR drop.)

At still higher voltage, the Si tip fractures and exposes an irregular, but probably very clean, surface of the semiconductor. The observed energy distribution then consists of two separate contributions which Russell ascribes to emission from the conduction and valence bands respectively since their peaks are separated by an energy equal to the band gap of Si. Actually, it is the two energies corresponding to the onset of emission from the bottom of the conduction band and the top of the valence band which should be separated by the band gap whereas their measured separation is only about 0.5 eV. (As pointed out by Russell, part of this discrepancy may be accounted for by the finite time constant of the phase sensitive detector employed.) Further, both the distributions are again far too wide (about 0.4 and 0.9 eV, respectively) and it does not seem reasonable that this should result only from the relatively low resolution.

Zdhan[21] and his co-workers used a cathode consisting of a thin layer of SiO_2 on tungsten, activated with carbon. Although they claim a resolving power of 0.04 eV for their analyzer, their measured half-width for tungsten is 0.47 eV which again indicates additional sources of error. Under the conditions of their experiment (layer thickness of about a micron, high internal resistivity, low image force barrier height $\approx$0.8 eV) the electron gas in the bulk of the semiconductor is heated by the internal field ($1\sim 5.10^4$ V/cm) and a considerable portion of the emission is over the top of the barrier so that the process is more properly called hot electron thermionic emission. In trying to explain their observations of distribution widths of up to 4 eV (!) they predict very high electron temperatures which in energy units even *exceed the barrier height.* Such extreme conditions would require a complete reformulation of their analysis which, as usual, is based on a spherically symmetrical distribution function in mo-

[16] A. G. Zdhan and M. I. Elinson, Radiotekhn. i Elektron. **6**, 671 (1961).
[17] A. H. Wilson, *The Theory of Metals* (Cambridge University Press, New York, 1954), p. 248.
[18] M. S. P. Lucas, Appl. Phys. Letters **4**, 73 (1964).
[19] A. M. Russell, Phys. Rev. Letters **9**, 417 (1962).
[20] A. M. Russell and E. Litov, Appl. Phys. Letters **2**, 64 (1963).
[21] A. G. Zdhan, M. I. Elinson and V. B. Sandomirskiy, Radiotekhn. i Elektron. **7**, 630 (1962).

mentum space for the electrons just inside the surface of the emitter. Zdhan and his co-workers observe that the width of the energy distributions increases rapidly with rising field strength which is qualitatively consistent with a rising electron temperature.

Similar heating of the electron gas occurs in the experiments by Shcherbakov and Sokol'skaya[22,23] using a CdS tip. Their control experiment with a tungsten tip, while giving a distribution whose shape is still somewhat different from the correct one, does give half-width in substantial agreement with theory. They observed half-widths which increased as a function of the internal potential drop up to values of about 2.3 eV. An inexplicable aspect of the results is that even for a negligible internal potential drop, and presumably little electron heating, half-widths as large as 0.8 eV were observed. Also, further increases in the half-widths occurred before the internal fields were high enough to lead to changes in the conductivity which was measured simultaneously.

In summary, the experimental evidence on the high resistivity semiconductors SiO_2 and CdS strongly supports the concept of electron heating in the surface region due to very intense internal electric fields although detailed explanations of the experimental results are still outstanding. It must be pointed out that the hot electron distribution function in momentum space, just within the semiconductor surface, differs from that in the interior of the cathode if there are wide space charge regions.[24] The excessively wide distributions observed by Russel for Si cannot be ascribed to electron heating—at least not for the valence-band emission. Much more experimental work needs to be done on the high conductivity semiconductors to clarify the effect of electron heating.

The energy distributions for electrons emitted by thin film tunneling into vacuum have been measured by Kanter and Feibelman[8] (normal distribution, Al-Al_2O_3-Au structure) and Collins and Davies[9] (total distribution, Al-Al_2O_3-Al structure). In each case the emitted distributions had half-widths of about a volt. Collins and Davies gave an approximate analysis of their results which was based on electron scattering processes involving energy loss in the conduction bands of both the insulator and the counter electrode (Al). For the part of the calculation involving energy loss in the conduction band of the insulator they assume that all the electrons tunnel from the Fermi energy, i.e., the width of the normal energy distribution for tunneling is assumed small. For a more detailed analysis, the initial normal energy distribution of electrons that have tunneled into the conduction band would have to be considered since this distribution can be about 0.6 eV wide. (Although the calculations in Secs. 3 and 4 are based on a parabolic energy momentum relation, the results should still apply to tunneling through wide gap insulators if the barrier height is not too large and if the free electron mass in the barrier region is replaced by an effective mass, appropriate to the insulator.[13])

[22] I. L. Sokol'skaya and G. P. Shcherbakov, Fiz. Tverd. Tela **3**, 167 (1961) [English transl.: Soviet Phys.—Solid State **3**, 120 (1961)].

[23] G. P. Shcherbakov and I. L. Sokol'skaya, Fiz. Tverd. Tela **4**, 3526 (1962) [English transl.: Soviet Phys.—Solid State **4**, 2581 (1963)].

[24] R. Stratton, Phys. Rev. **126**, 2002 (1962).

APPENDIX: VALENCE BAND FIELD-EMISSION CURRENT

The drivation given in Ref. 7 needs to be modified by replacing $D_v(-\bar{E}+\bar{E}_\perp)$ by $D_v(E_v+E_{v\perp})$ wherever it occurs. Thus in Eq. (11) of Ref. 7, the sign in front of E_{vm} and E_{vm}' must be changed and a minus sign placed in front of the whole expression for the current density j_v, viz.,

$$j_v=L\Big[\int_0^{E_V} dE_{vx}\ln\{1+\exp[(E_{vx}-\zeta_v)/kT]\}\times\{D_v(E_{vx})-[1+E_{vm}'(E_{vx})]D_v[E_{vx}+E_{vm}(E_{vx})]\}+\ln\Big\{1+\exp\Big(\frac{E_V-\zeta_v}{kT}\Big)\times\int_{E_V}^{E_{mv}(E_V)+E_V} D_v(E_{vx})dE_{vx}\Big]. \quad \text{(A1)}$$

For the model of spherical energy surfaces we find, following Sec. 3 in Ref. 7

$$j_v=Ae^{-b_{v0}}\Big[\frac{r_v(1+r_v)^{-1}}{(c_{v0}kT)^2}+e^{\eta}\{H_0(-\alpha_{v0},\eta)-\gamma_v H_0(-\gamma_v\alpha_{v0},\eta)\}\Big], \quad \text{if} \quad \eta<0 \quad \text{(A2)}$$

$$j_v=A\Big[e^{-b_{v1}}\Big\{\frac{\pi/\alpha_{v1}}{\sin(\pi\alpha_{v1})}-\frac{H_0(\alpha_{v1},-\eta)}{e^{(1-\alpha_{v1})\eta}}\Big\}-\gamma_v e^{-\bar{b}_{v1}}\Big\{\frac{\pi/\bar{\alpha}_{v1}}{\sin(\pi\bar{\alpha}_{v1})}-\frac{H_0(\gamma_v\bar{\alpha}_{v1},-\eta)}{e^{(1-\gamma_v\bar{\alpha}_{v1})\eta}}\Big\}\Big], \quad \text{if} \quad \eta>0 \quad \text{(A3)}$$

where

$$\gamma_v=1+r_v, \quad \alpha_{v0,1}=c_{v0,1}kT, \quad \eta=\zeta_v/kT, \quad \text{(A4)}$$

H_0 is a slowly varying function defined in Ref. 7 and where the bars on b_{v1} and α_1 (or c_{v1}) indicate that the quantity $(\varphi+E_g+\zeta_v)=\theta_{v1}$, in their definitions [cf. Eq. (23)] must be replaced by $\theta_{v1}+r_v\zeta_v$, usually an unimportant distinction.