

Generalized Formula for the Electric Tunnel Effect between Similar Electrodes Separated by a Thin Insulating Film

Cite as: Journal of Applied Physics **34**, 1793 (1963); <https://doi.org/10.1063/1.1702682>
Submitted: 03 January 1963 . Published Online: 09 June 2004

John G. Simmons



View Online



Export Citation

ARTICLES YOU MAY BE INTERESTED IN

Tunneling Conductance of Asymmetrical Barriers

Journal of Applied Physics **41**, 1915 (1970); <https://doi.org/10.1063/1.1659141>

Generalized Thermal J-V Characteristic for the Electric Tunnel Effect

Journal of Applied Physics **35**, 2655 (1964); <https://doi.org/10.1063/1.1713820>

Comment: "Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film" [J. Appl. Phys. 34, 1793 (1963)]

Journal of Applied Physics **123**, 136101 (2018); <https://doi.org/10.1063/1.5019788>





Lock-in Amplifiers

Zurich Instruments

Watch the Video

value of p we obtain for the threshold currents for case (a)

$$i \geq eBw(0.064)^2(8mkT/h^2)^3,$$

and for case (b)

$$i \geq 2e(BD_p)^{1/2}(0.064)^{1/2}(8mkT/h^2)^{3/4}.$$

Taking at room temperature $D_p \sim 4 \text{ cm}^2/\text{sec}$ and w as $3 \mu^7$ we may estimate the threshold currents at room temperature.

$$\text{Case (a):} \quad i \geq 2.5 \times 10^4 \text{ A/cm}^2,$$

$$\text{Case (b):} \quad i \geq 4.0 \times 10^4 \text{ A/cm}^2.$$

If we assume D_p constant with respect to temperature, both formulas yield the same temperature dependence for the current, namely

$$i \sim T^{3/4}.$$

⁷ The value of 3μ for w is somewhat arbitrary; we have picked w as the distance in which a typical zinc-diffused junction would exhibit a variation in doping by a factor of 3. See F. A. Cunnell and G. H. Gooch, S. E. R. L. Tech. J. 10, No. 2, p. 83 (January 1960).

At liquid nitrogen temperatures the thresholds would be reduced by a factor of 8 below the values at room temperature into the range of 3–5 thousand A/cm². Such thresholds have been observed by most of the workers in this field.⁸

Note added in proof. A referee of this paper suggested that a strict application of the condition of Bernard and Durauffourg may lead to a less arbitrary assumption than $f_p = 0.1$. For band to band transitions, the condition of Bernard and Durauffourg yields $f_p = 0.15$, $f_n = 0.85$, and the calculated threshold currents are increased about 60% while the temperature dependence remains the same. For transitions involving an impurity level 0.04 eV from the band edge, the threshold current below room temperature varies faster than $T^{1/4}$, and the calculated threshold currents are decreased.

⁸ The threshold current at room temperature has been observed to be ten times the threshold current at liquid nitrogen temperatures. See G. Burns and M. I. Nathan, IBM J. Res. Develop. 7, 72 (January 1963).

Generalized Formula for the Electric Tunnel Effect between Similar Electrodes Separated by a Thin Insulating Film

JOHN G. SIMMONS

Burroughs Corporation, Burroughs Laboratories, Paoli, Pennsylvania

(Received 3 January 1963)

A formula is derived for the electric tunnel effect through a potential barrier of arbitrary shape existing in a thin insulating film. The formula is applied to a rectangular barrier with and without image forces. In the image force problem, the true image potential is considered and compared to the approximate parabolic solution derived by Holm and Kirschstein. The anomalies associated with Holm's expression for the intermediate voltage characteristic are resolved. The effect of the dielectric constant of the insulating film is discussed in detail, and it is shown that this constant affects the temperature dependence of the J - V characteristic of a tunnel junction.

INTRODUCTION

IF two electrodes are separated by a thin insulating film, and the film is sufficiently thin, current can flow between the two electrodes by means of the tunnel effect.¹ Sommerfeld and Bethe² were the first to make a theoretical study of this phenomena for very low voltages and for high voltages; later, Holm³ extended the theory to include intermediate voltages. The two studies were thus concerned with different voltage ranges, and the pertinent theory was derived separately and independently, using the WBK approximation as the starting point in each case.

Sommerfeld and Bethe first derived equations for the

current density transmitted by a rectangular barrier. Inclusion of the image potential in the theory resulted in equations that could be solved only numerically. To obtain an analytic solution, Sommerfeld and Bethe approximated the barrier by a symmetric parabola. Later, Holm and Kirschstein,⁴ using the same method, improved upon the results of Sommerfeld and Bethe by using a symmetric parabola that was a closer fit to the potential barrier. The range and applicability of this type of approximation are limited. Holm³ simplified the image potential problem for the intermediate voltage range by correcting the calculations based upon a rectangular barrier, using the results obtained by Holm and Kirschstein⁴ for the low-voltage case. The validity of this procedure is shown to be questionable.

¹ J. C. Fisher and I. Giaever, J. Appl. Phys. 32, 172 (1961).

² A. Sommerfeld and H. Bethe, *Handbuch der Physik von Geiger und Scheel* (Julius Springer-Verlag, Berlin, 1933), Vol. 24/2, p. 450.

³ R. Holm, J. Appl. Phys. 22, 569 (1951).

⁴ R. Holm and B. Kirschstein, Z. Tech. Physik 16, 488 (1935).

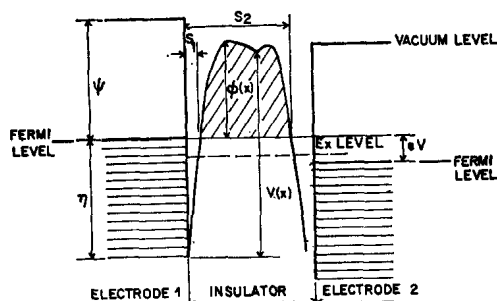


FIG. 1. General barrier in insulating film between two metal electrodes.

The purpose of this paper is to derive a single theory for the current flow through a generalized barrier. The theory is applied to the problem of the rectangular barrier, as studied by Sommerfeld and Bethe,² and by Holm.³ The generalized theory permits derivation of more accurate expressions for the high and low voltages, as well as resolution of the anomalies associated with the formulae derived by Holm for intermediate voltages.

Finally, a method is described for the application of the theory to a practical barrier—that is, to a rectangular barrier with the image potential included. The hyperbolic form of the image potential is used in the generalized formula, thus eliminating the need to resort to a parabolic approximation. The result is a more accurate theoretical current-voltage relationship for a tunnel junction.

NOTATION

- m = mass of electron,
- e = charge of electron,
- h = Planck's constant,
- s = thickness of insulating film,
- s_1, s_2 = limits of barrier at Fermi level,
- $\Delta s = s_2 - s_1$,
- J = tunnel current density,
- V = voltage across film,
- V_i = image potential,
- η = Fermi level,
- $f(E)$ = Fermi-Dirac function,
- ψ = work function of metal electrode,
- ϕ_0 = height of rectangular barrier,
- $\bar{\phi}$ = mean barrier height,
- ϵ = permittivity of insulating film,
- K = dielectric constant, and
- σ = tunnel resistivity ($\Omega\text{-cm}^2$).

THE TUNNEL EQUATION

When two metallic electrodes are separated by an insulating film, the equilibrium conditions require that the top of the energy gap of the insulator be positioned above the Fermi level of the electrodes. Thus, the action of the insulating film is to introduce a potential barrier between the electrodes which impedes the flow of electrons between the electrodes.

The electronic current can flow through the insulating region between the two electrodes if: (a) The electrons in the electrodes have enough thermal energy to surmount the potential barrier and flow in the conduction band. (b) The barrier is thin enough to permit its penetration by the electric tunnel effect.

Sommerfeld and Bethe, and Holm conducted analyses of these conditions for low temperatures so that thermal current could be neglected, thus restricting the electron transport between electrodes to the tunnel effect; a similar procedure is followed in this paper.

The probability $D(E_x)$ that an electron can penetrate a potential barrier of height $V(x)$ —the barrier is assumed to be in the x direction, as shown in Fig. 1—is given by the well-known WKB approximation⁵:

$$D(E_x) = \exp \left\{ -\frac{4\pi}{h} \int_{s_1}^{s_2} [2m(V(x) - E_x)]^{1/2} dx \right\}, \quad (1)$$

where $E_x = mv_x^2/2$, and is the energy component of the incident electron in the x direction. The number N_1 of electrons tunneling through the barrier from electrode 1 to electrode 2 is given by

$$N_1 = \int_0^{v_m} v_x n(v_x) D(E_x) dv_x = \frac{1}{m} \int_0^{E_m} n(v_x) D(E_x) dE_x, \quad (2)$$

where E_m is the maximum energy of the electrons in the electrode, and $n(v_x)dv_x$ is the number of electrons per unit volume with velocity between v_x and $v_x + dv_x$. For an isotropic velocity distribution, which is assumed to exist here inside the electrodes, the number of electrons per unit volume with velocity between the usual infinitesimal limits is given by

$$n(v)dv_x dv_y dv_z = (2m^4/h^3) f(E) dv_x dv_y dv_z, \quad (3)$$

where $f(E)$ is the Fermi-Dirac distribution function. Consequently, from Eq. (3),

$$\begin{aligned} n(v_x) &= \frac{2m^4}{h^3} \int \int_{-\infty}^{\infty} f(E) dv_y dv_z \\ &= \frac{4\pi m^3}{h^3} \int_0^{\infty} f(E) dE_r. \end{aligned} \quad (4)$$

In Eq. (4), the integrand is expressed in polar coordinates; that is,

$$\begin{aligned} v_r^2 &= v_y^2 + v_z^2, \\ E_r &= mv_r^2/2. \end{aligned}$$

Substituting Eq. (4) in Eq. (2) yields

$$N_1 = \frac{4\pi m^2}{h^3} \int_0^{E_m} D(E_x) dE_x \int_0^{\infty} f(E) dE_r. \quad (5)$$

⁵ D. Bohm, *Quantum Theory* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1951), p. 275.

The number N_2 of electrons tunneling from electrode 2 to electrode 1 is determined in a similar manner. The tunnel probability $D(E_x)$ is the same in either direction, and if electrode 2 is at a positive potential V , with respect to electrode 1, the Fermi-Dirac function is written as $f(E+eV)$; therefore,

$$N_2 = \frac{4\pi m^2}{h^3} \int_0^{E_m} D(E_x) dE_x \int_0^\infty f(E+eV) dE_r. \quad (6)$$

The net flow of electrons $N (= N_1 - N_2)$ through the barrier is

$$N = \int_0^{E_m} D(E_x) dE_x \times \left\{ \frac{4\pi m^2}{h^3} \int_0^\infty [f(E) - f(E+eV)] dE_r \right\}. \quad (7)$$

Writing

$$\zeta_1 = \frac{4\pi m^2 e}{h^3} \int_0^\infty f(E) dE_r,$$

and

$$\zeta_2 = \frac{4\pi m^2 e}{h^3} \int_0^\infty f(E+eV) dE_r,$$

and $\zeta = \zeta_1 - \zeta_2$, Eq. (7) becomes

$$J = \int_0^{E_m} D(E_x) \zeta dE_x. \quad (8)$$

Current-Voltage Relationship for a Generalized Barrier

Writing⁶ $V(x) = \eta + \varphi(x)$, with reference to Fig. 1, Eq. (1) becomes

$$D(E_x) = \exp \left[-\frac{4\pi}{h} (2m)^{\frac{1}{2}} \int_{s_1}^{s_2} (\eta + \varphi(x) - E_x)^{\frac{1}{2}} dx \right]. \quad (9)$$

Integrating Eq. (9), using Eq. (A5) from the Appendix, yields

$$D(E_x) \simeq \exp[-A(\eta + \bar{\varphi} - E_x)^{\frac{1}{2}}], \quad (10)$$

where $\bar{\varphi}$, the mean barrier height above Fermi level of the negatively biased electrode, is

$$\bar{\varphi} = \frac{1}{\Delta s} \int_{s_1}^{s_2} \varphi(x) dx,$$

and

$$A = (4\pi\beta\Delta s/h)(2m)^{\frac{1}{2}},$$

where β is defined in the Appendix. At 0°K, ζ_1 and ζ_2 are given by

$$\zeta_1 = (4\pi me/h^3)(\eta - E_x)$$

⁶ By the substitution $V(x) = \eta + \varphi(x)$, we have inherently assumed that the width Δs of the barrier in the range $E_x > V(x) > \eta$ is constant and equal to the barrier width at the Fermi level. This assumption is justified for practical barriers, because Δs varies slowly below the Fermi level [Fig. 4(a)], and the integral has effective values only when $E_x \simeq \eta$.

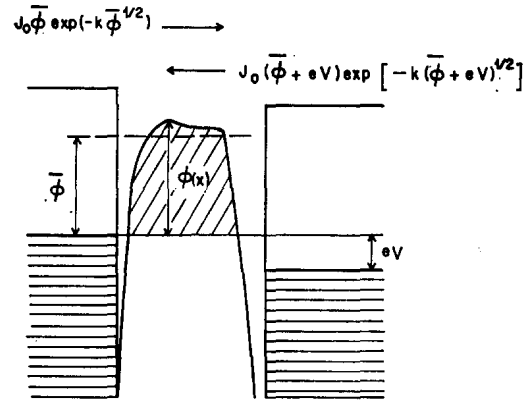


FIG. 2. Pictorial illustration of Eq. (20), showing current flow between the electrodes.

and

$$\zeta_2 = (4\pi me/h^3)(\eta - E_x - eV).$$

Hence,

$$\zeta = \begin{cases} (4\pi me/h^3)(eV) & 0 \leq E_x < \eta - eV \\ (4\pi me/h^3)(\eta - E_x) & \eta - eV < E_x < \eta \\ 0 & E_x > \eta \end{cases}. \quad (11)$$

Substituting Eqs. (10) and (11) in Eq. (8) gives

$$J = \frac{4\pi me}{h^3} \left\{ eV \int_0^{\eta - eV} \exp[-A(\eta + \bar{\varphi} - E_x)^{\frac{1}{2}}] dE_x + \int_{\eta - eV}^{\eta} (\eta - E_x) \exp[-A(\eta + \bar{\varphi} - E_x)^{\frac{1}{2}}] dE_x \right\}. \quad (12)$$

To facilitate integration, Eq. (12) is written in the form

$$J = \frac{4\pi me}{h^3} \left\{ eV \int_0^{\eta - eV} \exp[-A(\eta + \bar{\varphi} - E_x)^{\frac{1}{2}}] dE_x - \bar{\varphi} \int_{\eta - eV}^{\eta} \exp[-A(\eta + \bar{\varphi} - E_x)^{\frac{1}{2}}] dE_x + \int_{\eta - eV}^{\eta} (\eta + \bar{\varphi} - E_x) \times \exp[-A(\eta + \bar{\varphi} - E_x)^{\frac{1}{2}}] dE_x \right\}. \quad (13)$$

The first of the integrals in Eq. (13) yields

$$(8\pi mV/h^3)(e/A)^2 \{ [A(\bar{\varphi} + eV)^{\frac{1}{2}} + 1] \exp[-A(\bar{\varphi} + eV)^{\frac{1}{2}}] - [A(\bar{\varphi} + \eta)^{\frac{1}{2}} + 1] \exp[-A(\bar{\varphi} + \eta)^{\frac{1}{2}}] \}. \quad (14)$$

The second term in the braces is negligible compared to the first term and, usually, $A(\bar{\varphi} + eV)^{\frac{1}{2}} \gg 1$; thus Eq. (14) reduces to

$$(8\pi me^2/h^3 A) V(\bar{\varphi} + eV)^{\frac{1}{2}} \exp[-A(\bar{\varphi} + eV)^{\frac{1}{2}}]. \quad (15)$$

The second integral in Eq. (13) is of the same form as the first. Taking advantage of the approximations

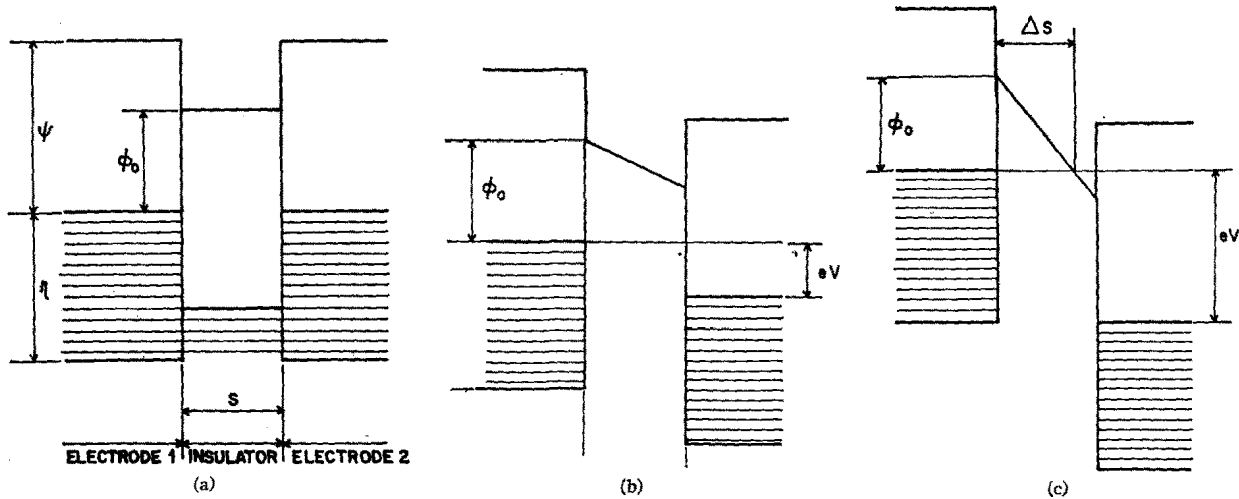


FIG. 3. Rectangular potential barrier in insulating film between metal electrodes for: (a) $V=0$; (b) $V < \phi_0/e$; (c) $V > \phi_0/e$.

that lead to Eq. (15), the second term integrates to

$$-(8\pi me/h^3 A^2) \bar{\phi} \{ [A \bar{\phi}^{\frac{1}{2}} + 1] \exp(-A \bar{\phi}^{\frac{1}{2}}) - [A(\bar{\phi} + eV)^{\frac{1}{2}} + 1] \exp[-A(\bar{\phi} + eV)^{\frac{1}{2}}] \}. \quad (16)$$

The third integral of Eq. (13) has the form

$$\int z^3 e^{-Az} dz = -e^{-Az} \left(\frac{z^3}{A} + \frac{3z^2}{A^2} + \frac{6z}{A^3} + \frac{6}{A^4} \right), \quad (17)$$

where

$$z^2 = \eta + \bar{\phi} - E_x.$$

The third and fourth terms in parentheses in Eq. (17) are negligible by comparison with the first two; therefore, the third integral in Eq. (13) integrates to

$$(8\pi me/h^3 A) \{ \bar{\phi}^{\frac{3}{2}} \exp(-A \bar{\phi}^{\frac{1}{2}}) - (\bar{\phi} + eV)^{\frac{3}{2}} \exp[-A(\bar{\phi} + eV)^{\frac{1}{2}}] \} + (8\pi me/h^3 A)(3/A) \{ \bar{\phi} \exp(-A \bar{\phi}^{\frac{1}{2}}) - (\bar{\phi} + eV) \exp[-A(\bar{\phi} + eV)^{\frac{1}{2}}] \}. \quad (18)$$

Summing Eqs. (15), (16), and (18) yields

$$J = (e/2\pi h)(\beta \Delta s)^{-2} \{ \bar{\phi} \exp(-A \bar{\phi}^{\frac{1}{2}}) - (\bar{\phi} + eV) \exp[-A(\bar{\phi} + eV)^{\frac{1}{2}}] \}. \quad (19)$$

Equation (19) can be expressed in the following form:

$$J = J_0 \{ \bar{\phi} \exp(-A \bar{\phi}^{\frac{1}{2}}) - (\bar{\phi} + eV) \exp[-A(\bar{\phi} + eV)^{\frac{1}{2}}] \}, \quad (20)$$

where

$$J_0 = e/2\pi h(\beta \Delta s)^2.$$

Equation (20) has the advantage that it can be applied to any shape of potential barrier providing the mean barrier height is known, or, alternatively, if the current-voltage characteristic of a tunnel junction is known, the mean barrier height can be determined.

Equation (20) can be interpreted as a current density $J_0 \bar{\phi} \exp(-A \bar{\phi}^{\frac{1}{2}})$ flowing from electrode 1 to electrode 2 and a current density $J_0 (\bar{\phi} + eV) \exp[-A(\bar{\phi} + eV)^{\frac{1}{2}}]$

flowing from electrode 2 to electrode 1, resulting in a net current density J , given by Eq. (20). (See Fig. 2.) When V is zero, a state of dynamic equilibrium can be considered to exist—that is, a current density of magnitude $J_0 \bar{\phi} \exp(-A \bar{\phi}^{\frac{1}{2}})$ flowing in either direction.

Low-Voltage Range

Although Eq. (20) can be used for very low voltages, a more convenient form can be deduced for this range. From Eq. (20),

$$J = J_0 \{ \bar{\phi} \exp(-A \bar{\phi}^{\frac{1}{2}}) - (\bar{\phi} + eV) \exp[-A(\bar{\phi} + eV)^{\frac{1}{2}}] \}. \quad (21)$$

It is observed that, since $eV \approx 0$, β [as defined in Eq. (A6)] takes the value unity. Since $\bar{\phi} \gg eV$, Eq. (21) can be written

$$J = J_0 [\bar{\phi} - (\bar{\phi} + eV) \exp(-AeV/2\bar{\phi}^{\frac{1}{2}})] \times \exp(-A \bar{\phi}^{\frac{1}{2}}). \quad (22)$$

Expanding $\exp(-AeV/2\bar{\phi}^{\frac{1}{2}})$, and neglecting terms containing V^2 and higher orders, Eq. (22) becomes

$$J = J_0 [\bar{\phi} - (\bar{\phi} + eV)(1 - AeV/2\bar{\phi}^{\frac{1}{2}})] \exp(-A \bar{\phi}^{\frac{1}{2}}) = J_0 eV [A \bar{\phi}^{\frac{1}{2}}/2 - 1] \exp(-A \bar{\phi}^{\frac{1}{2}}). \quad (23)$$

Since $A \bar{\phi}^{\frac{1}{2}}/2 \gg 1$, Eq. (23) reduces to

$$J = J_L \bar{\phi}^{\frac{1}{2}} V \exp(-A \bar{\phi}^{\frac{1}{2}}), \quad (24)$$

where

$$J_L = [(2m)^{\frac{1}{2}}/\Delta s](e/h)^2.$$

Since eV is very small, $\bar{\phi}$ is considered to be the zero-voltage mean barrier height. Thus, in this case, Eq. (24) expresses J as a linear function of V ; that is, the junction is Ohmic for very low voltages.

APPLICATION OF THE TUNNEL EQUATIONS

Consider a rectangular potential barrier [Fig. 3(a)]. This was the type of barrier studied by Sommerfeld and

Bethe, and by Holm. Sommerfeld and Bethe considered the low-voltage and high-voltage cases, and Holm the intermediate case. Equations for each are derived for these cases, using Eqs. (20) and (24).

Low-Voltage Range: $V \simeq 0$

From Fig. 3(a),

$$\Delta s = s,$$

and

$$\bar{\varphi} = \varphi_0.$$

Substituting these values in Eq. (24) gives

$$J = [3(2m\varphi)^{1/2}/2s](e/h)^2 V \times \exp[-(4\pi s/h)(2m\varphi)^{1/2}]. \quad (25)$$

This result is in agreement with the Sommerfeld-Bethe result for low voltages.

Intermediate-Voltage Range: $V < \varphi_0/e$

From Fig. 3(b),

$$\Delta s = s,$$

and

$$\bar{\varphi} = (\varphi - eV/2).$$

Substituting these values in Eq. (20),

$$\begin{aligned} J &= J_0 \left\{ (\varphi - eV/2) \exp[-A(\varphi - eV/2)^{1/2}] \right. \\ &\quad \left. - (\varphi + eV/2) \exp[-A(\varphi + eV/2)^{1/2}] \right\} \\ &= \frac{e}{2\pi h(\beta s)^2} \left\{ \left(\varphi - \frac{eV}{2} \right) \exp \left[-\frac{4\pi\beta s}{h} (2m)^{1/2} \left(\varphi - \frac{eV}{2} \right)^{1/2} \right] \right. \\ &\quad \left. - \left(\varphi + \frac{eV}{2} \right) \exp \left[-\frac{4\pi\beta s}{h} (2m)^{1/2} \left(\varphi + \frac{eV}{2} \right)^{1/2} \right] \right\}. \quad (26) \end{aligned}$$

It is now necessary to discuss the error associated with Eq. (26). It can be shown, using Eq. (A6), that β is given by

$$\beta = 1 - (eV)^2/96(\varphi_0 + \eta - E_x - eV/2)^2.$$

If this value of β is substituted in Eq. (26), the maximum error in the exponents in using the approximate integral Eq. (A5) is approximately 1% and occurs when $V = \varphi_0/e$ and $E_x = \eta$. For values of $eV < \varphi_0/e$, the error reduces rapidly. If β is chosen to be unity, the error in the value of the exponents is approximately 6% at $V = \varphi_0/e$. However, since the error reduces rapidly for values of $V < \varphi_0/e$, the error is only 1% at $V = 0.75\varphi_0/e$, and β can, therefore, be chosen to be unity to a reasonable approximation.

With $\beta = 1$, Eq. (26) becomes^{6a}

^{6a} Note added in proof. Simmons has shown [see J. G. Simmons, J. Appl. Phys. 34, 238 (1963)] that, at relatively low voltages, Eq. (27) reduces to

$$J = J_L(V + \beta V^2),$$

where

$$\beta = [(Ae)^2/96\varphi] - [Ae^2/32\varphi^3],$$

which is in good quantitative agreement with the experimental

$$J = \left(\frac{e}{2\pi h s^2} \right) \left\{ \left(\varphi_0 - \frac{eV}{2} \right) \exp \left[-\frac{4\pi s}{h} (2m)^{1/2} \left(\varphi_0 - \frac{eV}{2} \right)^{1/2} \right] \right. \\ \left. - \left(\varphi_0 + \frac{eV}{2} \right) \exp \left[-\frac{4\pi s}{h} (2m)^{1/2} \left(\varphi_0 + \frac{eV}{2} \right)^{1/2} \right] \right\}. \quad (27)$$

Equation (27) differs from Holm's result [Eq. (16) of reference 3]. In addition to the terms of Eq. (27), Holm includes an additional term,

$$\frac{(2m)^{1/2}}{h^2} eV \left(\varphi_0 + \frac{eV}{2} \right)^{1/2} \exp \left[-\frac{4\pi s}{h} (2m)^{1/2} \left(\varphi_0 + \frac{eV}{2} \right)^{1/2} \right] \quad (28)$$

which is in error. Holm recognizes that there are inconsistencies in his result (Sec. VIII of reference 3), for two reasons: (1) As $V \rightarrow 0$, his equation does not reduce to the Sommerfeld-Bethe relationship; that is, the equation does not predict the low-voltage Ohmic characteristic [Eqs. (24) and (25)]. (2) According to his equation, the resistance of the junction initially *increases* with increasing voltage.

Holm suggests that these anomalies are due to the approximate nature of $D(E_x)$. This is not the case, however, for the anomalies are removed when the extraneous term, Eq. (28), above, in his equation [Eq. (16) of reference 3] is neglected. (See Fig. 6.)

High-Voltage Range: $V > \varphi_0/e$

Figure 3(c) illustrates the energy diagram for this case; from this figure,

$$\Delta s = s\varphi_0/eV,$$

and

$$\bar{\varphi} = \varphi_0/2.$$

Substituting these values in Eq. (20) yields

$$\begin{aligned} J &= \left(\frac{2e^3(F/\beta)^2}{8\pi h \varphi_0} \right) \left\{ \exp \left[-\frac{4\pi\beta}{eF} m^{1/2} \varphi_0^{1/2} \right] - \left(1 + \frac{2eV}{\varphi_0} \right) \right. \\ &\quad \left. \times \exp \left[-\frac{4\pi\beta}{eF} m^{1/2} \varphi_0^{1/2} \left(1 + \frac{2eV}{\varphi_0} \right)^{1/2} \right] \right\}, \quad (29) \end{aligned}$$

where $F = V/s$ = the field strength in the insulator.

It is now necessary to determine the correction factor β appearing in Eq. (29). From Eq. (A6),

$$\begin{aligned} \beta &= 1 - \left[\frac{(eV/s)^2}{8\Delta s} \right] \int_0^{\Delta s = s\varphi_0/eV} (\Delta s/2 - x)^2 dx / (\varphi_0/2)^2 \\ &= 1 - 1/24 = 23/24. \end{aligned}$$

Therefore, for this case, β is independent of V . Sub-

results of Knauss and Breslow [see H. P. Knauss and R. A. Breslow, Proc. IRE 50, 1843 (1962)].

stituting this value in Eq. (29) gives

$$J = \frac{2.2e^3 F^2}{8\pi h \varphi_0} \left\{ \exp \left[-\frac{8\pi}{2.96heF} (2m)^{\frac{1}{2}} \varphi_0^{\frac{3}{2}} \right] - \left(1 + \frac{2eV}{\varphi_0} \right) \right. \\ \left. \times \exp \left[-\frac{8\pi}{2.96heF} (2m)^{\frac{1}{2}} \varphi_0^{\frac{3}{2}} \left(1 + \frac{2eV}{\varphi_0} \right)^{\frac{1}{2}} \right] \right\}. \quad (30)$$

For very high voltages (that is, where $V > (\varphi + \eta)/e$), the Fermi level of electrode 2 lies below the bottom of the conduction band of electrode 1. Under these conditions, electrons cannot tunnel from electrode 2 to electrode 1, since there are no empty levels available to them. The situation is reversed, however, for electrons tunneling from electrode 1 to electrode 2, since all of the available levels in electrode 2 are empty. The situation is analogous to that of field emission from a metal electrode, and, for this condition (that is, where $V > (\varphi + \eta)/e$), the second term in Eq. (30) is negligible; thus,

$$J = \frac{2.2e^3 F^2}{8\pi h \varphi_0} \exp \left[-\frac{8\pi}{2.96heF} (2m)^{\frac{1}{2}} \varphi_0^{\frac{3}{2}} \right].$$

This equation is similar to the Sommerfeld-Bethe² relationship for high voltages, except for the multiplicative factor 2.2; there is also a slight difference in the numerator, where 3 is replaced by 2.96. These differences arise because of the variation of Δs below the Fermi level.⁶ Because of the dominant influence of the exponential term, and since J , in Eq. (30), is a rapidly varying function of s , φ , and V , this difference is considered to be insignificant.

THE IMAGE FORCE

The effect of the image force is to reduce the area of the potential barrier by rounding off the corners and reducing the thickness of the barrier (Fig. 4) and, hence, increasing the flow of current between the electrodes. The image potential is a hyperbolic function which, when substituted in Eq. (9), results in an elliptic integral which can be solved only numerically. Sommerfeld and Bethe, and Holm solved the problem analytically by approximating the barrier by a symmetric parabola. This type of approximation is good only for the low-voltage range and high barriers, and is restricted in range of validity. This can be seen by reference to Fig. 5(a) (which illustrates the energy diagram with the electrodes at the same potential—that is, the Fermi levels coincide). When a voltage V is applied to the electrodes, the parabola is moved vertically down the energy diagram (that is, in the direction of negative energy) by an amount $eV/2$. The best correlation between the symmetric parabola and the true image force occurs at very low voltages, and is, at best, only fair, even where the parabola parameters are optimized to particular values of s and φ_0 . The correla-

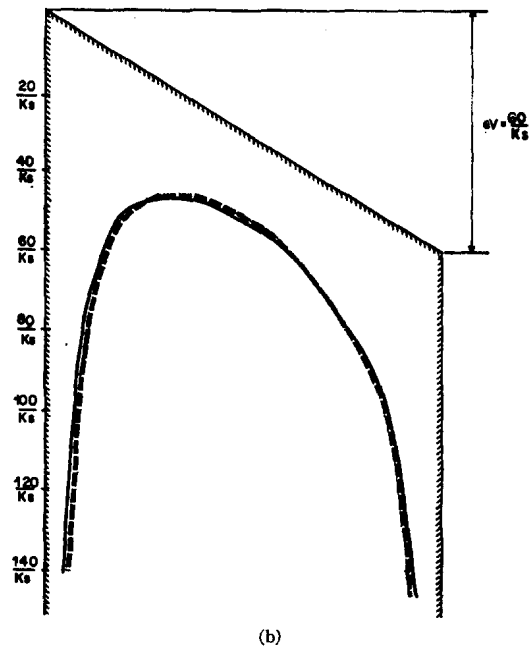
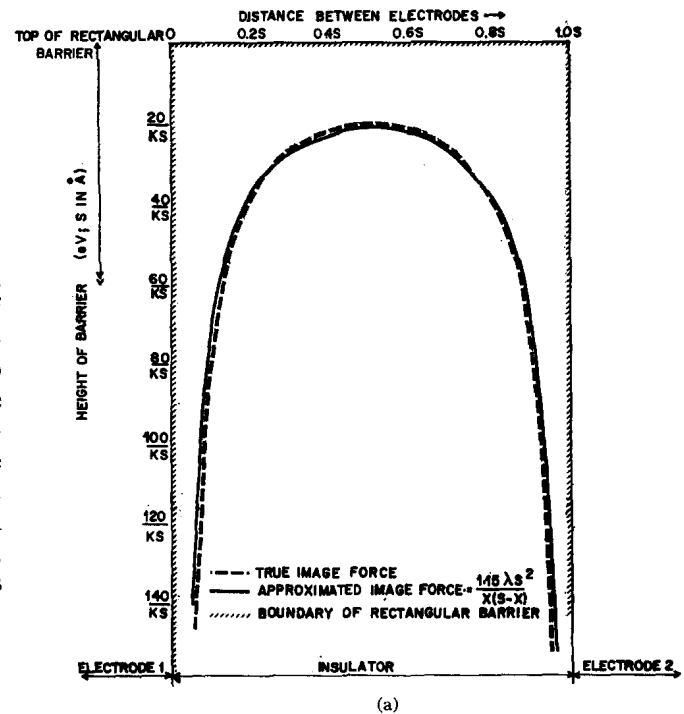


FIG. 4. Normalized energy diagram of a rectangular barrier with image forces included. Diagram compares the actual image potential with the approximate image potential for two cases: (a) zero voltage across film; (b) voltage of magnitude $60/eKs$ across film.

tion deteriorates for barriers having s and φ_0 different from those for which the parabola constants are optimized. For high voltages, the fit is very poor. [See Fig. 5(b).]

The true image force problem can be solved using Eqs. (20) and (24); however, the resulting expression, which is an infinite series, is awkward to handle. To

facilitate computation, the true expression can be approximated accurately by a simple hyperbolic function [Fig. 4(a)] which can also be readily solved by Eqs. (20) and (24). In contrast to the symmetric parabolic approximation, a very close fit is obtained to the actual barrier for all s , φ_0 , and V [See Fig. 4(b).]

The Image Potential

The image potential is readily determined using image force methods,⁷ and is given by

$$V_i = \left(-\frac{e^2}{4\pi\epsilon} \right) \left[\frac{1}{2x} + \sum_{n=1}^{\infty} \left\{ \frac{ns}{[(ns)^2 - x^2]} - \frac{1}{ns} \right\} \right], \quad (31)$$

where x is the distance of the electron from electrode 1. When $x = s/2$,

$$V_i = -\frac{e^2}{2\pi\epsilon s} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} = -\frac{e^2}{2\pi\epsilon s} \ln 2. \quad (32)$$

Equation (31) as it exists is extremely awkward to handle; a good approximation—see Fig. 4(a)— is given by

$$V_i = -1.15\lambda s^2/x(s-x), \quad (33)$$

where

$$\lambda = e^2 \ln 2 / 8\pi\epsilon s. \quad (34)$$

Transmission Through a Barrier with Image Force Included

When the image potential expressed by Eq. (33) is taken to account, $\varphi(x)$ is written—see Fig. 4(b)—as

$$\varphi(x) = \varphi_0 - eVx/s - 1.15\lambda s^2/x(s-x).$$

For this case,

$$\bar{\varphi} = \frac{1}{\Delta s} \int_{s_1}^{s_2} \left\{ \varphi_0 - \frac{eVx}{s} - \frac{1.15\lambda s^2}{x(s-x)} \right\} dx. \quad (35)$$

The limits s_1 and s_2 are given by the real roots of the cubic equation

$$\varphi_0 - eVx/s - 1.15\lambda s^2/x(s-x) = 0. \quad (36)$$

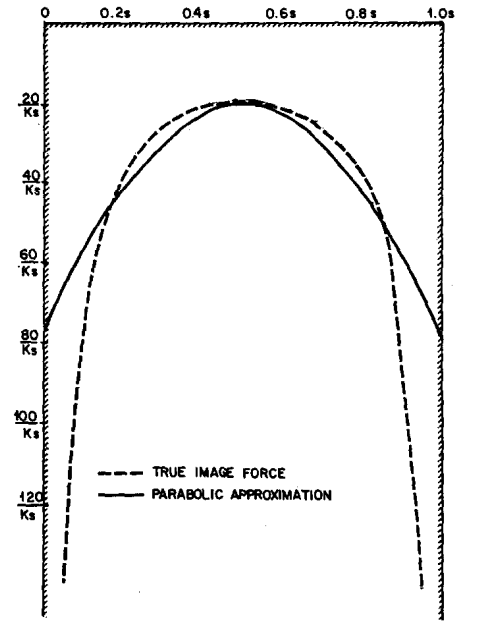
However, to facilitate an analytic solution of Eq. (35), the roots are written to a good approximation as

$$\left. \begin{aligned} s_1 &= 1.2\lambda s / \varphi_0 \\ s_2 &= s \left[1 - 9.2\lambda / (3\varphi_0 + 4\lambda - 2eV) \right] + s_1 \end{aligned} \right\} eV < \varphi_0, \quad (37)$$

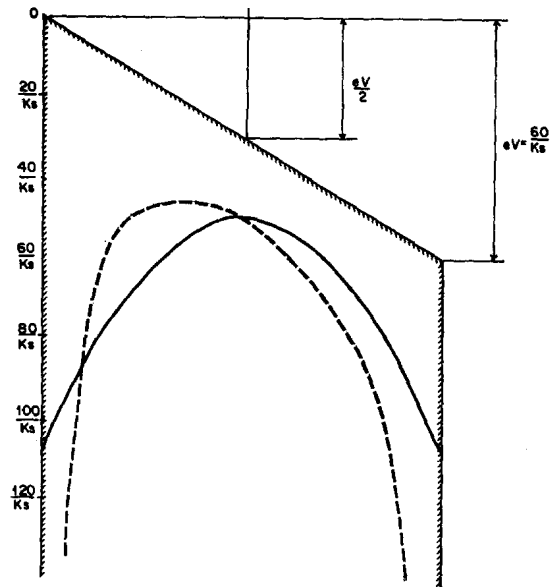
and

$$\left. \begin{aligned} s_1 &= 1.2\lambda s / \varphi_0 \\ s_2 &= (\varphi_0 - 5.6\lambda) / (s/eV) \end{aligned} \right\} eV > \varphi_0. \quad (38)$$

⁷ W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1950), Chap. IV.



(a)



(b)

FIG. 5. Normalized energy diagram comparing a parabolic approximation with the true barrier for: (a) zero voltage bias; (b) voltage bias of $V = 60/eKs$.

Integrating Eq. (35) yields:

$$\bar{\varphi} = \varphi_0 - (eV/2s)(s_1 + s_2) - [1.15\lambda s / (s_2 - s_1)] \times \ln[s_2(s - s_1) / s_1(s - s_2)] = \varphi_T. \quad (39)$$

Intermediate Voltages

For intermediate voltages (defined here as $0 < V < \varphi_0/e$), the tunnel current density is obtained by

substituting Eq. (39) into Eq. (20), giving

$$J = J_0 \{ \varphi_I \exp(-A \varphi_I^{\frac{1}{2}}) - (\varphi_I + eV) \exp[-A(\varphi_I + eV)^{\frac{1}{2}}] \}, \quad (40)$$

where s_1 and s_2 are given by Eq. (37).

It is necessary to comment here on the value of β appearing in A . It can be shown, using Eq. (A6), that $\beta > 0.96$ for all values of V . Thus, it is assumed that β takes the value unity for all V , to a good approximation. In this case, A is given by

$$A = (4\pi\Delta s/h)(2m)^{\frac{1}{2}}.$$

High Voltages

The high-voltage range is defined as $V > \varphi_0/e$. The current density equation is identical to Eq. (40), but, in this case, s_1 and s_2 are given by Eq. (38). For very high voltages—that is, where $V > (\varphi_0 + \eta)/e$ —the second term is negligible compared to the first.

Low Voltages

This range is defined as $V \simeq 0$; thus, Eq. (39) becomes

$$\bar{\varphi} = \varphi_0 - [1.15\lambda s/(s_2 - s_1)] \ln \times [s_2(s - s_1)/s_1(s - s_2)] = \varphi_L, \quad (41)$$

where s_1 and s_2 are given by Eq. (37), with V set equal to zero. Alternatively, because, in this case, the barrier

is symmetrical, s_2 can be written simply as

$$s_2 = s - s_1 = s - 1.2\lambda s/\varphi,$$

and

$$s_1 = 1.2\lambda s/\varphi.$$

Substituting Eq. (41) in Eq. (24) gives

$$J = J_L \varphi_L^{\frac{1}{2}} \exp(-A \varphi_L^{\frac{1}{2}}). \quad (42)$$

Comparison between Eqs. (42) and (40) shows that it is not possible to deduce information relevant to the higher voltage ranges from the low-voltage case, as Holm⁸ has suggested.

EQUATIONS EXPRESSED IN PRACTICAL UNITS

For convenience of numerical calculations, J is expressed in A/cm², φ_0 in V, and s , s_1 , and s_2 in Å units.

Generalized Barrier

(i) All V , Eq. (20):

$$J = [6.2 \times 10^{10}/(\beta\Delta s)^2] \{ \bar{\varphi} \exp(-1.025\beta\Delta s \bar{\varphi}^{\frac{1}{2}}) - (\bar{\varphi} + V) \exp[-1.025\beta\Delta s(\bar{\varphi} + V)^{\frac{1}{2}}] \}. \quad (43)$$

(ii) $V \simeq 0$, Eq. (24):

$$J = 3.16 \times 10^{10} \bar{\varphi}^{\frac{1}{2}} (V/\Delta s) \exp(-1.025\Delta s \bar{\varphi}^{\frac{1}{2}}). \quad (44)$$

Rectangular Barrier

(i) $V \simeq 0$, Eq. (25):

$$J = 3.16 \times 10^{10} \bar{\varphi}_0^{\frac{1}{2}} (V/s) \exp(-1.025s \bar{\varphi}_0^{\frac{1}{2}}). \quad (45)$$

(ii) $0 < V < \varphi_0$; Eq. (27):

$$J = (6.2 \times 10^{10}/s^2) \{ (\varphi_0 - V/2) \exp[-1.025s(\varphi_0 - V/2)^{\frac{1}{2}}] - (\varphi_0 + V/2) \exp[-1.025s(\varphi_0 + V/2)^{\frac{1}{2}}] \}. \quad (46)$$

(iii) $V > \varphi_0$; Eq. (30):

$$J = 3.38 \times 10^{10} (F^2/\varphi_0) \left\{ \exp(-0.689\varphi_0^{\frac{1}{2}}/F) - \left(1 + \frac{2V}{\varphi_0} \right) \exp \left[-0.689 \frac{\varphi_0^{\frac{1}{2}}}{F} \left(1 + \frac{2V}{\varphi_0} \right)^{\frac{1}{2}} \right] \right\}. \quad (47)$$

Rectangular Barrier with Image Forces Included

(i) $V \simeq 0$; Eq. (42):

$$J = (3.16 \times 10^{10}/\Delta s) \varphi_L^{\frac{1}{2}} V \exp(-1.025\Delta s \varphi_L^{\frac{1}{2}}), \quad (48)$$

where

$$\varphi_L = \varphi_0 - [5.75/K(s_2 - s_1)] \ln[s_2(s - s_1)/s_1(s - s_2)], \quad (49)$$

and

$$s_1 = 6/K\varphi_0,$$

$$s_2 = s - (6/K\varphi_0).$$

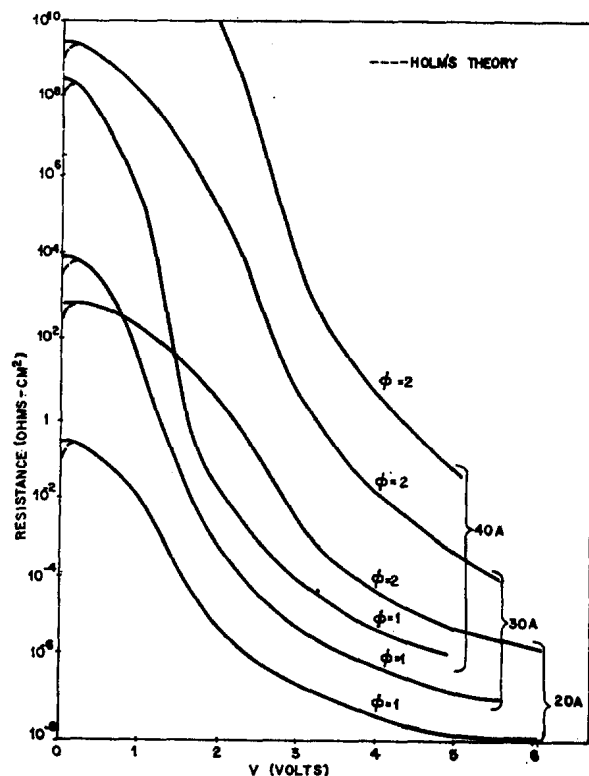


FIG. 6. Theoretical σ - V characteristic of a tunnel junction having a rectangular barrier; diagram also shows the anomalous behavior of Holm's equation.

⁸ R. Holm, *Electric Contacts Handbook* (Springer-Verlag, Berlin, 1958), 3rd ed., p. 433.

(ii) All V ; Eq. (40):

$$J = (6.2 \times 10^{10} / \Delta s^2) \{ \varphi_I \exp(-1.025 \Delta s \varphi_I^{\frac{1}{2}}) - (\varphi_I + V) \exp[-1.025 \Delta s (\varphi_I + V)^{\frac{1}{2}}] \}, \quad (50)$$

where

$$\varphi_I = \varphi_0 - (V/2s)(s_1 + s_2) - [5.75/K(s_2 - s_1)] \ln[s_2(s - s_1)/s_1(s - s_2)],$$

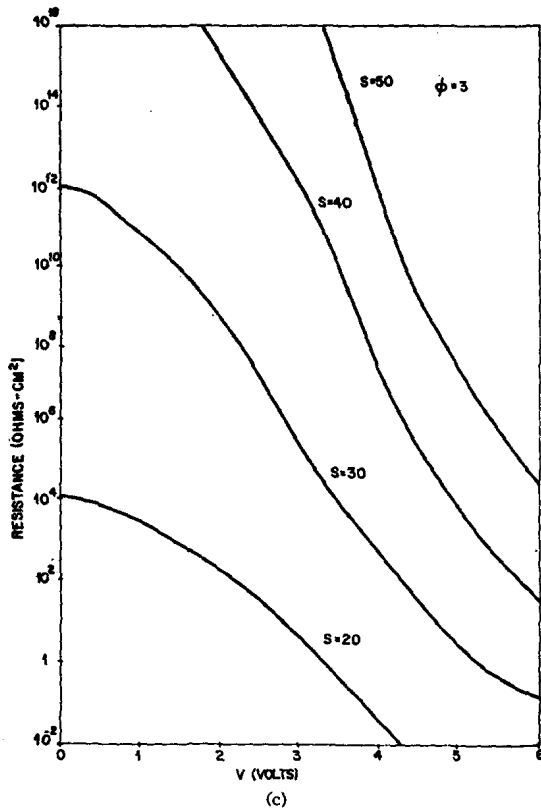
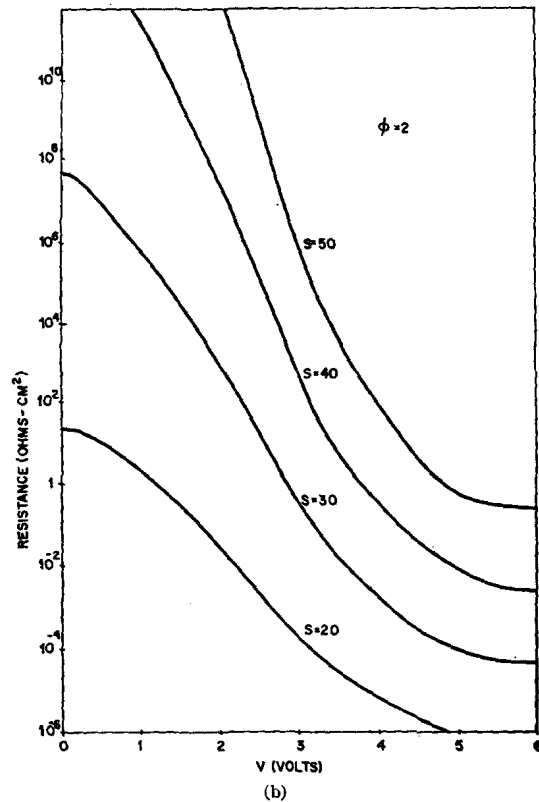
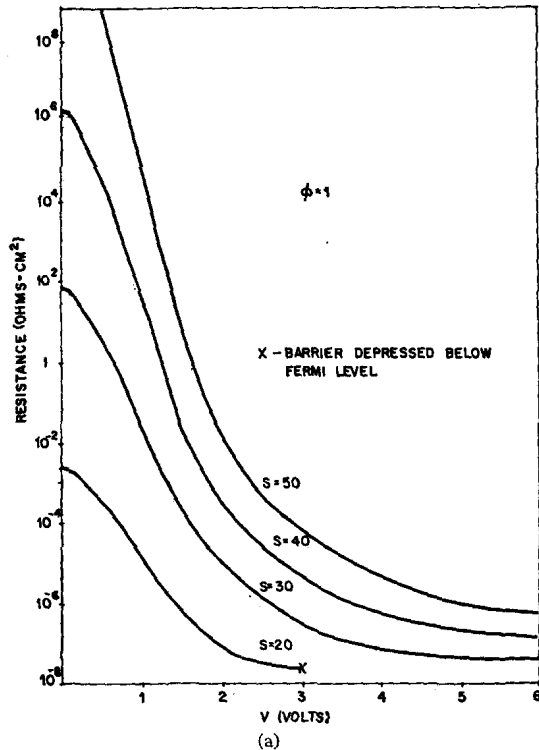


FIG. 7. Theoretical σ - V characteristic of a tunnel junction having a practical barrier for s with the values 20, 30, 40, and 50 Å and for: (a) $\varphi = 1$ eV; (b) $\varphi = 2$ eV; (c) $\varphi = 3$ eV.

and

$$\left. \begin{aligned} s_1 &= 6/K\varphi_0 \\ s_2 &= s[1 - 46/(3\varphi_0 Ks + 20 - 2VKs)] + 6/K\varphi_0 \end{aligned} \right\} V < \varphi_0,$$

$$\left. \begin{aligned} s_1 &= 6/K\varphi_0 \\ s_2 &= (\varphi_0 Ks - 28)/KV \end{aligned} \right\} V > \varphi_0.$$

NUMERICAL EVALUATIONS

Tunnel Resistivity

Tunnel resistivity $\sigma (= V/J)$, as a function of voltage, is illustrated in Fig. 6 and 7, using Eqs. (45) through (50). The curves shown in Fig. 6 are for a rectangular barrier without image forces. Figure 7 (wherein a dielectric constant of 6 has been assumed) depicts curves for a practical barrier—that is, a rectangular barrier with image forces included. It can be observed that, for a given s , φ , and V , the tunnel resistivity is lower for the practical barrier than for the ideal barrier,

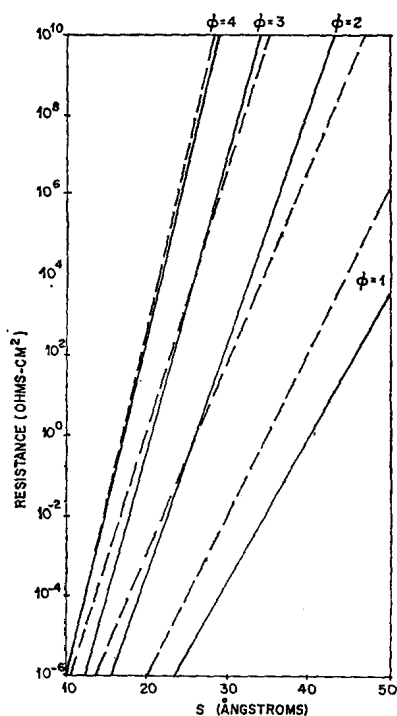


FIG. 8. Low-voltage σ - V characteristic for practical barrier; dotted lines indicate results of Holm and Kirschstein using the parabolic approximation. For these curves, $K=1$.

as anticipated, although the general shape of the curves is preserved.

For small values of V , all of the curves of σ versus V approach a horizontal asymptote; that is, the junctions exhibit an Ohmic characteristic which Holm's equations failed to predict. (See Sec. VIII, reference 3.) This Ohmic characteristic is well documented in the literature.^{1,9,10} Figure 8 illustrates the low-voltage (Ohmic) resistance as a function of thickness and barrier height, using Eq. (42). The broken lines indicate the results of Holm and Kirschstein⁴ for the low-voltage resistance in which the image force has been approximated by a symmetric parabola. The correlation between the two sets of curves is poor for small ϕ but good for high ϕ . This is because the image force has a greater influence on small barriers, and, hence, in this range, the approximate nature of the image force solution of Holm and Kirschstein⁴ becomes apparent.

If the voltage applied to the junction is great enough, the barrier is depressed below the Fermi level of the negatively biased electrode [Fig. 7(a)], and the current flows unimpeded in the conduction band of the insulator. The voltage at which this occurs is greater the larger the values of s , ϕ_0 , and K . (See Fig. 4.)

Effect of Dielectric Constant

The tunnel characteristics are dependent upon the dielectric constant of the insulating film; the smaller

⁹ R. Holm and W. Meissner, Z. Physik 74, 715 (1932); 86, 787 (1933).

¹⁰ J. G. Simmons, G. J. Unterkofer, and W. W. Allen, Appl. Phys. Letters 2, 78 (1963).

the value of K , the lower is the tunnel resistivity. Figure 9 illustrates the profound effect of the dielectric constant upon the low-voltage tunnel resistance. Since the dielectric constant of most materials is a function of temperature,¹¹ it follows that the tunnel characteristics are an intrinsic function of the thermal properties of the insulator, as well as of the electrodes. This fact appears to have been neglected in the literature.

SUMMARY

A generalized expression has been derived for the electric tunnel effect through an arbitrary barrier in a thin insulating film. The formula is applied to a rectangular barrier and the resulting expression compared with existing theories. The anomalies associated with Holm's expression for the intermediate voltage range are resolved and shown to arise from an extraneous term.

The formula is readily applied to the true image force problem for all voltage ranges. The resulting expressions are compared with the low-voltage characteristic derived by Holm and Kirschstein, who used the symmetric parabola approximation. It is shown that the approximation is good for high barriers, but poor for low barriers. Holm's suggestion for correcting the intermediate-voltage J - V expression for a rectangular barrier to include the image forces is shown to be questionable.

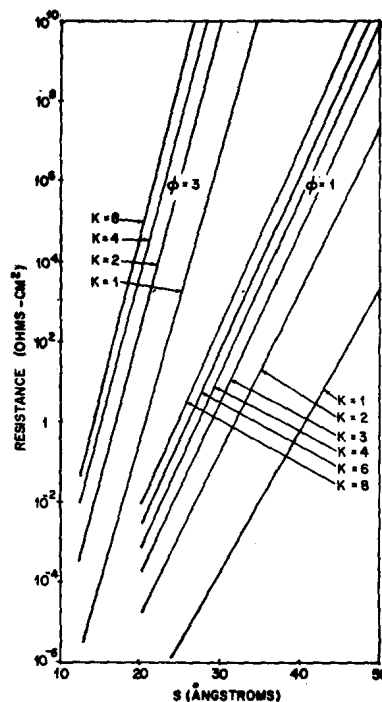


FIG. 9. Theoretical low-voltage σ - V curves showing the effect of the dielectric constant.

¹¹ A. R. von Hippel, *Dielectric Materials and Applications* (Tech. Press, Cambridge, Massachusetts, and John Wiley & Sons, Inc., New York, 1961), Pt. V.

The effect of the dielectric constant of the insulating film is discussed in detail, and is shown to have a profound effect on the J - V characteristic. The results suggest that if the dielectric constant is temperature-dependent, the J - V characteristic is also temperature-dependent.

ACKNOWLEDGMENTS

The author wishes to acknowledge G. J. Unterkofler for stimulating discussion and R. Proctor for programming the tunnel equations for the Burroughs 220 computer.

APPENDIX

To integrate an arbitrary function $f^{\frac{1}{2}}(x)$ —that is,

$$\int_{s_1}^{s_2} f^{\frac{1}{2}}(x) dx, \quad (\text{A1})$$

a function $\bar{f}$ is defined as

$$\bar{f} = 1/\Delta s \int f(x) dx; \quad (\text{A2})$$

that is, $\bar{f}$ is the mean value of $f(x)$. Equation (A1) is then written as

$$\int_{s_1}^{s_2} f^{\frac{1}{2}}(x) dx = \bar{f}^{\frac{1}{2}} \int_{s_1}^{s_2} \left\{ 1 + \frac{[f(x) - \bar{f}]}{\bar{f}} \right\}^{\frac{1}{2}} dx. \quad (\text{A3})$$

Expanding Eq. (A3) and neglecting terms in $[(f(x) - \bar{f})/\bar{f}]^3$ and higher powers,

$$\begin{aligned} \int_{s_1}^{s_2} f^{\frac{1}{2}}(x) dx &= \bar{f}^{\frac{1}{2}} \int_{s_1}^{s_2} \left\{ 1 + \frac{[f(x) - \bar{f}]}{2\bar{f}} - \frac{[f(x) - \bar{f}]^2}{8\bar{f}^2} \right\} dx. \quad (\text{A4}) \end{aligned}$$

The integral of the second term in the brackets is zero [as defined by Eq. (A2)]; thus, Eq. (A4) integrates to

$$\begin{aligned} \int_{s_1}^{s_2} f^{\frac{1}{2}}(x) dx &\simeq \bar{f}^{\frac{1}{2}} \Delta s \left\{ 1 - \frac{1}{8\bar{f}^2 \Delta s} \int_{s_1}^{s_2} [f(x) - \bar{f}]^2 dx \right\} \\ &= \beta \bar{f} \Delta s, \quad (\text{A5}) \end{aligned}$$

where

$$\Delta s = s_2 - s_1,$$

and

$$\beta = \text{correction factor} = 1 - \frac{1}{8\bar{f}^2 \Delta s} \int_{s_1}^{s_2} [f(x) - \bar{f}]^2 dx. \quad (\text{A6})$$

Usually,

$$1 \gg \frac{1}{8\bar{f}^2 \Delta s} \int_{s_1}^{s_2} [f(x) - \bar{f}]^2 dx;$$

that is $\beta \simeq 1$. Thus,

$$\int_{s_1}^{s_2} f^{\frac{1}{2}}(x) dx = \bar{f}^{\frac{1}{2}} \Delta s, \quad (\text{A7})$$

to a good approximation.