Thermionic Emission, Field Emission, and the Transition Region

E. L. MURPHY* AND R. H. GOOD, JR.
Department of Physics, Pennsylvania State University, University Park, Pennsylvania

(Received January 16, 1956)

Although the theories of thermionic and field emission of electrons from metals are very well understood, the two types of emission have usually been studied separately by first specifying the range of temperature and field and then constructing the appropriate expression for the current. In this paper the emission is treated from a unified point of view in order to establish the ranges of temperature and field for the two types of emission and to investigate the current in the region intermediate between thermionic and field emission. A general expression for the emitted current as a function of field, temperature, and work function is set up in the form of a definite integral. Each type of emission is then associated with a technique for approximating the integral and with a characteristic dependence on the three parameters. An approximation for low fields and high temperatures leads to an extension of the Richardson-Schottky formula for thermionic emission. The values of temperature and field for which it applies are established by considering the validity of the approximation. An analogous treatment of the integral, for high fields and low temperatures, gives an extension of the Fowler-Nordheim formula for field emission, and establishes the region of temperature and field in which it applies. Also another approximate method for evaluating the integral is given which leads to a new type of dependence of the emitted current on temperature and field and which applies in a narrow region of temperature and field intermediate between the field and thermionic emission regions.

I. INTRODUCTION

THE current emitted from a metal increases with the temperature of the metal and the applied field strength. The thermionic and cold emission processes are very well understood on the basis of the Fermi-Dirac distribution for a free electron gas in the metal and the classical image force barrier at the surface. The temperature dependence arises in the distribution function and the field dependence in the shape of the surface barrier. For high temperature and low field strength, emission over the barrier predominates and the temperature dependence of the distribution function is mainly responsible for variations in the emitted current. This process is thermionic emission. For high field strength and low temperature, emission of electrons with energies below the Fermi level predominates; field dependence of the barrier shape is mainly responsible for variations in the emitted current and this process is called field emission.

The theoretical treatment of thermionic emission leads to the Richardson equation,[1] modified by the Schottky dependence[2] on the square root of the applied field. The theoretical treatment of field emission leads to the Fowler-Nordheim equation.[3,4] Contributions to the study of the emission in the transition region, based on this model, have been made by Sommerfeld and Bethe[5] and by Guth and Mullin[6] using series expansion methods and also have been made by Dolan and Dyke[7] and by Dyke, Barbour, Martin, and Trolan[8] using numerical methods.

In this paper the entire emission phenomenon is studied from a unified point of view. Extensions of the Richardson-Schottky and Fowler-Nordheim formulas are developed. The regions of temperature and field for which the extended formulas are valid are determined; these are referred to below as the thermionic and field emission regions. Also an expression for the current in a narrow intermediate region is developed. The calculations are based on a general expression for the emitted current as a function of temperature, field, and work function, in the form of a definite integral. The expression is found from the established model: the Fermi-Dirac distribution for the free electrons and the classical image force barrier at the surface. The form of the integrand suggests approximate evaluation techniques; these correspond to the various types of emission. The well-known formulas for thermionic and field emission currents come out as limiting cases. Roughly fields from 0 to 10^8 volts/cm and temperatures from 0 to 3000°K are considered, although not all of this range is at present experimentally accessible.

The basic equations, including the general expression for the current, are given in Sec. II. Thermionic and field emission are treated in Secs. III and IV and in Sec. V formulas for the intermediate region are derived. A general discussion of the results is given in Sec. VI.

II. BASIC EQUATIONS

The free-electron model gives the following for the number of electrons per second per unit area having

* National Science Foundation Predoctoral Fellow.
[1] O. W. Richardson, *The Emission of Electricity from Hot Bodies* (Longmans Green and Company, London, 1921).
[2] W. Schottky, Physik. Z. **15**, 872 (1914).
[3] L. W. Nordheim, Proc. Roy. Soc. (London) **A121**, 626 (1928).
[4] R. H. Fowler and L. W. Nordheim, Proc. Roy. Soc. (London) **A119**, 173 (1928).
[5] A. Sommerfeld and H. Bethe, *Handbuch der Physik*, edited by H. Geiger and K. Scheel (Verlag Julius Springer, Berlin, 1933), Vol. 24, No. 2, p. 442.
[6] E. Guth and C. J. Mullin, Phys. Rev. **61**, 339 (1942).
[7] W. W. Dolan and W. P. Dyke, Phys. Rev. **95**, 327 (1954).
[8] Dyke, Barbour, Martin, and Trolan, Phys. Rev. **99**, 1192 (1955).

energy within the range dW incident on the barrier[9]:

$$N(T,\zeta,W)dW=4\pi mkTh^{-3}\times\ln\{1+\exp[-(W-\zeta)/kT]\}dW, \quad (1)$$

where N is called the supply function, m is the electron mass, k is Boltzmann's constant, T is the absolute temperature, h is Planck's constant, and ζ is the Fermi energy. Energies are measured from zero for a free electron outside the metal, so that the work function ϕ is simply $-\zeta$. Here W is only the part of the energy for the motion normal to the surface:

$$W=[p^2(x)/2m]+V(x), \quad (2)$$

where x is the coordinate normal to the surface and out of the metal, $p(x)$ is the electron momentum normal to the surface, and $V(x)$ is the effective electron potential energy.

The assumed potential energy of the electrons is (see Fig. 1):

$$V(x)=-e^2(4x)^{-1}-eFx, \quad \text{when} \quad x>0, \quad (3)$$

$$=-W_a, \quad \text{when} \quad x<0, \quad (4)$$

where $-e$ is the charge on the electron, $-e^2(4x)^{-1}$ is the contribution from the image force, $-eFx$ is the contribution from the externally applied field F, and $-W_a$ is the effective constant potential energy inside the metal. In the region near $x=0$ it is assumed that $V(x)$ is regular and connects smoothly with the functions in Eqs. (3) and (4). The calculations below are independent of the details of the shape of the potential in that region. The maximum value of the potential energy $V_{\max}$ is $-(e^3F)^{\frac{1}{2}}$.

The following approximation for the probability $D(F,W)$ that an electron incident on the barrier emerges from the metal will be used:

$$D(F,W)=\left[1+\exp\left(-2i\hbar^{-1}\int_{x_1}^{x_2}p(\xi)d\xi\right)\right]^{-1}. \quad (5)$$

Here x_1 and x_2 are points where $p^2(x)$ becomes zero and $\hbar$ is $h/2\pi$. The branches of $p(x)$ to be used in the integrand are specified below. This formula was first proposed by Kemble[10] and also can be understood in terms of a parabolic WKB-type approximation.[11] It applies to the case of a simple potential barrier for which $p^2(x)$ has two zeros, possibly complex, and is not expected to be valid if $p(x)$ has any other zeros or singularities in their vicinity. When the energy W is below the peak of the barrier, the zero points are real and are to be chosen so that $x_1<x_2$; the argument of $p(x)$ is to be $\pi/2$.

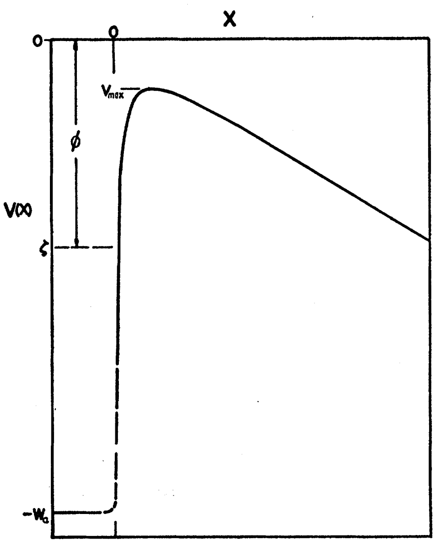

FIG. 1. Potential energy of an electron near the metal surface, Eqs. (3) and (4).

In consequence the entire exponent is positive. When the energy is above the peak of the barrier, the zero points are complex and are to be chosen so that the imaginary part of x_1 is positive and the imaginary part of x_2 is negative; the argument of $p(x)$ is to be in the neighborhood of zero. It develops, in consequence, that the entire exponent is negative. There is no difficulty in applying this transmission coefficient formula to the potential of Eqs. (3) and (4) as long as $W<V_{\max}=-(e^3F)^{\frac{1}{2}}$. The ξ integration is real and $p(x)$ is given by

$$p(x)=\{2m[W+e^2(4x)^{-1}+eFx]\}^{\frac{1}{2}}, \quad (6)$$

with argument $\pi/2$. When $W>-(e^3F)^{\frac{1}{2}}$, the ξ-integration is along a path in the complex plane and $V(x)$ must be defined for complex x. It is assumed that $V(x)$ is given by Eq. (3) for a range of W near $V_{\max}$. Then $p(x)$ is given again by Eq. (6) but with argument in the neighborhood of zero. In that case $p(x)$ branches at the points

$$x_1=(-W/2eF)[1+i(e^3FW^{-2}-1)^{\frac{1}{2}}], \quad (7)$$

$$x_2=(-W/2eF)[1-i(e^3FW^{-2}-1)^{\frac{1}{2}}], \quad (8)$$

and also at the origin. Since, as far as the actual potential $V(x)$ is concerned, the singularity at the origin does not apply, Eq. (6) should no longer be used when the origin is close to the other singularities relative to the distance between them. Therefore one expects that Eq. (6) may be no longer applicable for energies above the value given by

$$(e^3FW^{-2}-1)^{\frac{1}{2}}=1.$$

The limiting value W_l is

$$W_l=-\tfrac{1}{2}\sqrt{2}(e^3F)^{\frac{1}{2}}, \quad (9)$$

and the conclusion is that when W is below this value

[9] See, for example, R. Fowler and E. A. Guggenheim, *Statistical Thermodynamics* (Cambridge University Press, New York, 1952), p. 460.

[10] E. C. Kemble, *The Fundamental Principles of Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1937), first edition, Sec. 21j.

[11] S. C. Miller, Jr., and R. H. Good, Jr., Phys. Rev. **91**, 174 (1953), Sec. IV.

the transmission coefficient is given by Eqs. (5) and (6). When W is above this value the transmission coefficient may be taken to be one in calculating the current,

$$D=1 \quad \text{when} \quad W>W_l; \tag{10}$$

the justification is given following Eqs. (17) and (18).

The integral which arises when Eqs. (5) and (6) are combined can be evaluated in terms of complete elliptic functions. Only negative values of W need be considered. The integral is

$$-2i\hbar^{-1}\int_{x_1}^{x_2} p(\xi)d\xi=-2i\hbar^{-1}\int_{x_1}^{x_2} \times\{2m[W+e^2(4\xi)^{-1}+eF\xi]\}^{\frac{1}{2}}d\xi$$
$$=(4/3)\sqrt{2}(F\hbar^4/m^2e^5)^{-\frac{1}{4}}y^{-\frac{3}{2}}v(y), \tag{11}$$

where

$$y=(e^3F)^{\frac{1}{2}}/|W|, \tag{12}$$

$$v(y)=-\frac{3i}{4\sqrt{2}}\int_{1-(1-y^2)^{\frac{1}{2}}}^{1+(1-y^2)^{\frac{1}{2}}}[\rho-2+y^2\rho^{-1}]^{\frac{1}{2}}d\rho. \tag{13}$$

When $|W|<(e^3F)^{\frac{1}{2}}$ so that $y>1$, the argument of the integrand is in the neighborhood of zero and the argument of $(1-y^2)^{\frac{1}{2}}$ is $-\pi/2$. When $y<1$, the argument of the integrand is $\pi/2$ and the argument of $(1-y^2)^{\frac{1}{2}}$ is 0. The integral in Eq. (13) is well known.[12–14] The usual evaluation is

$$v(y)=2^{-\frac{1}{2}}(1+a)^{\frac{1}{2}}\{E[(2a)^{\frac{1}{2}}/(1+a)^{\frac{1}{2}}]-(1-a)K[(2a)^{\frac{1}{2}}/(1+a)^{\frac{1}{2}}]\}, \tag{14}$$

where

$$K[k]=\int_0^{\pi/2}(1-k^2\sin^2\theta)^{-\frac{1}{2}}d\theta,$$

$$E[k]=\int_0^{\pi/2}(1-k^2\sin^2\theta)^{\frac{1}{2}}d\theta,$$

and a is defined by

$$a=(1-y^2)^{\frac{1}{2}}.$$

[The argument of a is either 0 or $-\pi/2$ and the arguments of $(1+a)^{\frac{1}{2}}$ and $(1-k^2\sin^2\theta)^{\frac{1}{2}}$ are in the neighborhood of zero.] The following equivalent expressions, which considerably simplify discussions of this function, have been discovered:

$$v(y)=-(y/2)^{\frac{1}{2}}\{-2E[(y-1)^{\frac{1}{2}}/(2y)^{\frac{1}{2}}]+(y+1)K[(y-1)^{\frac{1}{2}}/(2y)^{\frac{1}{2}}]\} \tag{15}$$

when $\quad y>1,$

$$v(y)=(1+y)^{\frac{1}{2}}\{E[(1-y)^{\frac{1}{2}}/(1+y)^{\frac{1}{2}}]-yK[(1-y)^{\frac{1}{2}}/(1+y)^{\frac{1}{2}}]\} \tag{16}$$

when $\quad y<1,$

where the positive roots are to be used. Although Eq. (14) applies for all y, the modulus for the elliptic functions is complex when $y>1$ and this makes it awkward to evaluate $v(y)$ numerically in that region with this equation. In Eq. (15) the modulus for the elliptic functions is real and less than one; for this case the functions are well tabulated and numerical results are easily found. Equations (14) and (16) are equally convenient as far as the modulus range is concerned but Eq. (16) is usually preferable because of the simpler dependence on y. Burgess, Kroemer, and Houston have given a convenient table of $v(y)$ for $y\leq 1$.[13]

When Eqs. (5), (10), (11) are combined, the result for the transmission coefficient is

$$D(F,W)=\{1+\exp[(4/3)\sqrt{2}(F\hbar^4/m^2e^5)^{-\frac{1}{4}}y^{-\frac{3}{2}}v(y)]\}^{-1}$$

when $\quad W<W_l, \tag{17}$

$$D(F,W_l)=1 \quad \text{when} \quad W>W_l. \tag{18}$$

In order to justify setting $D=1$ in Eq. (18), one evaluates D by Eq. (17) at the limiting value of W. Evidently y is just $\sqrt{2}$ and Eq. (15) is to be used for $v(y)$. The result is

$$D[F,W_l]=\{1+\exp[-0.868(F\hbar^4/m^2e^5)^{-\frac{1}{4}}]\}^{-1}.$$

It develops that the energy range $W>W_l$ is significant only in the discussion of thermionic emission, Sec. III, and that there only fields smaller than 5×10^7 volts/cm are to be considered. At this extreme value $D[F,W_l]$ is already 0.94. Since also the actual transmission coefficient must approach one with increasing energy, one is justified in setting it equal to one over the whole range $W>W_l$. Any error in Eqs. (17) and (18) must be in the neighborhood of $W=W_l$, because below that value the approximation is dependable and above it the transmission coefficient is one. At the worst this will be only a minor part of the range of energies of the emitted electrons.

The total electric current per unit area $j(F,T,\zeta)$ is found by integrating, over all accessible energies, the product of the charge on an electron, the number per second per unit area incident on the barrier Eq. (1), and the penetration probability Eqs. (17) and (18):

$$j(F,T,\zeta)=e\int_{-W_a}^{\infty}D(F,W)N(T,\zeta,W)dW$$

$$=\frac{4\pi mkTe}{h^3}\int_{-W_a}^{W_l}\frac{\ln\{1+\exp[-(W-\zeta)/kT]\}dW}{1+\exp[(4/3)\sqrt{2}(F\hbar^4/m^2e^5)^{-\frac{1}{4}}y^{-\frac{3}{2}}v(y)]}$$
$$+\frac{4\pi mkTe}{h^3}\int_{W_l}^{\infty}\ln\{1+\exp[-(W-\zeta)/kT]\}dW. \tag{19}$$

Hartree units are used in the following sections. That is, j is redefined to mean the current per unit area divided by $m^3e^9\hbar^{-7}=2.37\times10^{14}$ amp/cm^2; F to mean the electric field strength divided by $m^2e^5\hbar^{-4}=5.15\times10^9$

[12] See reference 3, Sec. 5.

[13] Burgess, Kroemer, and Houston, Phys. Rev. **90**, 515 (1953).

[14] S. C. Miller, Jr., and R. H. Good, Jr., Phys. Rev. **92**, 1367 (1953).

volts/cm; and ζ, kT, W, W_a, W_l to mean the corresponding energies divided by $me^4\hbar^{-2}=27.2$ ev. In these terms,

$$j(F,T,\zeta)=\frac{kT}{2\pi^2}\int_{-W_a}^{W_l}\frac{\ln\{1+\exp[-(W-\zeta)/kT]\}dW}{1+\exp[(4/3)\sqrt{2}F^{-\frac{1}{4}}y^{-\frac{3}{2}}v(y)]}+\frac{kT}{2\pi^2}\int_{W_l}^{\infty}\ln\{1+\exp[-(W-\zeta)/kT]\}dW \tag{20}$$

is the complete expression for the current.

III. THERMIONIC EMISSION

One technique for evaluating the integrals in Eq. (20) applies when the conditions on the temperature and field given in Eqs. (34) and (35) below are satisfied. When these conditions hold, the emission will be called thermionic. The reason for this definition is that, as it develops, limiting values of the current in this region of temperature and field are given by the Richardson and Schottky formulas and the emission is strongly temperature-dependent. Within this thermionic emission region the integrals in Eq. (20) are similar to those evaluated by Miller and Good[15] and their procedure is also used here.

The basic idea of the approximation is to simplify the integrand in Eq. (20) by using the first term in an expansion of the logarithm above the Fermi energy and the first term in an expansion of the exponent in the denominator about the peak of the barrier. This leads to an integral which can be evaluated in terms of elementary functions. The expansions are

$$\ln\{1+\exp[-(W-\zeta)/kT]\}=\exp[-(W-\zeta)/kT]-\tfrac{1}{2}\exp[-2(W-\zeta)/kT]\cdots, \tag{21}$$

$$(4/3)\sqrt{2}F^{-\frac{1}{4}}y^{-\frac{3}{2}}v(y)=-\pi F^{-\frac{1}{4}}\epsilon+\tfrac{3}{16}\pi F^{-\frac{1}{4}}\epsilon^2\cdots, \tag{22}$$

where

$$\begin{aligned}\epsilon&=1+WF^{-\frac{1}{2}}\\&=(y-1)/y.\end{aligned} \tag{23}$$

The parameter ϵ is appropriate because it is zero at the peak of the barrier and linear in W. Equation (22) is easily found by expanding the elliptic functions in Eqs. (15) and (16) for small moduli. The small modulus expansions can be found, for example, in Byrd and Friedman's handbook,[16] Eqs. (900.00) and (900.07). When the first terms in Eqs. (21) and (22) are substituted into Eq. (20) the result is

$$j=\frac{kT}{2\pi^2}\int_{-W_a}^{W_l}\frac{\exp[-(W-\zeta)/kT]dW}{1+\exp[-\pi F^{-\frac{1}{4}}(1+WF^{-\frac{1}{2}})]}+\frac{kT}{2\pi^2}\int_{W_l}^{\infty}\exp[-(W-\zeta)/kT]dW. \tag{24}$$

[15] See reference 14, Sec. V.

[16] Paul F. Byrd and Morris D. Friedman, *Handbook of Elliptic Integrals for Engineers and Physicists* (Springer-Verlag, Berlin, 1954).

The conditions under which these substitutions are applicable yield the boundaries of the thermionic region. Before discussing these boundaries in detail, it is convenient to combine the two integrals by putting the extra factor

$$\{1+\exp[-\pi F^{-\frac{1}{4}}(1+WF^{-\frac{1}{2}})]\}^{-1}$$

into the second integrand. At the lower limit this has the value

$$\{1+\exp[-0.92F^{-\frac{1}{4}}]\}^{-1}$$

and it rapidly approaches unity with increasing W. Again anticipating that thermionic fields will be less than 5×10^7 volts/cm ($F<0.01$), one sees that even in the extreme case the value at the lower limit is 0.95 and so, as far as the entire integration is concerned, one is justified in inserting this factor. The expression for the current, Eq. (24), then becomes

$$j=\frac{kT}{2\pi^2}\int_{-W_a}^{\infty}\frac{\exp[-(W-\zeta)/kT]dW}{1+\exp[-\pi F^{-\frac{1}{4}}(1+WF^{-\frac{1}{2}})]}. \tag{25}$$

The boundaries of the thermionic region are established by the requirement that the approximations made to obtain the integrand in Eq. (25) should be valid in the range of W for which the integrand in Eq. (25) is appreciable. In terms of the abbreviation

$$d=F^{\frac{3}{4}}/\pi kT, \tag{26}$$

the energy η, defined to be the energy at which the integrand has its maximum value, is given by

$$\eta=-F^{\frac{1}{2}}-\pi^{-1}F^{\frac{3}{4}}\ln[d/(1-d)]. \tag{27}$$

For larger W the integrand behaves roughly like

$$\exp[-(W-\zeta)/kT]=\exp\{-\pi F^{-\frac{3}{4}}[dW-d\zeta]\}, \tag{28}$$

and for smaller W the integrand decreases certainly as fast as

$$\exp\{-\pi F^{-\frac{3}{4}}[(1-d)(-W)-d\zeta-F^{\frac{1}{2}}]\}. \tag{29}$$

The requirement is then that the approximations hold at least over the range

$$F^{\frac{3}{4}}/\pi d>W-\eta>-F^{\frac{3}{4}}/\pi(1-d). \tag{30}$$

The approximation of using the first term in Eq. (21) for the logarithm begins to apply when W becomes greater than $(\zeta+kT)$. Comparing with Eq. (30), one sees that a condition that Eq. (25) applies for the current is

$$\eta-[F^{\frac{3}{4}}/\pi(1-d)]>\zeta+kT. \tag{31}$$

One finds that the approximation of using the first term in Eq. (22) for the exponent in Eq. (20) applies when $\epsilon>-F^{1/8}$. In terms of W, this condition becomes $W>-F^{\frac{1}{2}}-F^{5/8}$, and comparing with Eq. (30) one obtains

$$\eta-[F^{\frac{3}{4}}/\pi(1-d)]>-F^{\frac{1}{2}}-F^{5/8} \tag{32}$$

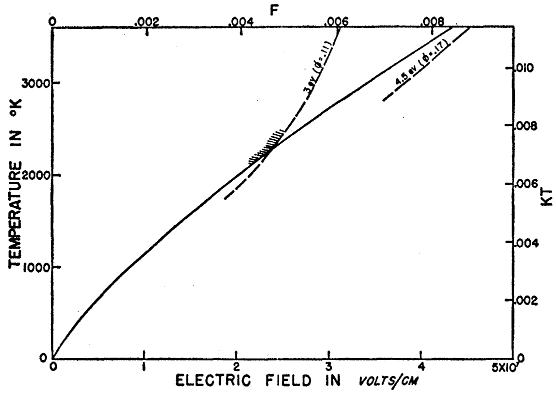

FIG. 2. Boundaries of the thermionic emission region as given by Eq. (34), the broken lines, and by Eq. (35), the solid line. For a work function of 3 ev, the region extends from the temperature axis out to the first line encountered as indicated by the shading. For work functions of 4.5 ev and higher, the corresponding broken line lies below the solid line at these temperatures and the region extends from the temperature axis out to the solid line.

for the second condition that Eq. (25) applies for the current.

Next the integral in Eq. (25) will be discussed. For metals the energy $-W_a$ is below the Fermi energy and so, according to Eqs. (30) and (31), is below the range where the integrand is appreciable; one can replace it by $-\infty$. By introducing the new integration variable

$$\mu=\exp[-\pi F^{-\frac{1}{4}}(1+WF^{-\frac{1}{2}})],$$

one brings the integral to a standard form[17]:

$$j=\frac{(kT)^2}{2\pi^2}\left(\exp\frac{F^{\frac{1}{2}}+\zeta}{kT}\right)d\int_0^\infty \frac{\mu^{d-1}d\mu}{1+\mu}$$
$$=\tfrac{1}{2}\pi^{-2}(kT)^2(\pi d/\sin\pi d)\exp[-(\phi-F^{\frac{1}{2}})/kT]. \quad (33)$$

Here ζ has been replaced by $-\phi$. From Eqs. (27), (31), and (32) one can write

$$\ln[(1-d)/d]-d^{-1}(1-d)^{-1}>-\pi F^{-\frac{3}{4}}(\phi-F^{\frac{1}{2}}), \quad (34)$$
$$\ln[(1-d)/d]-(1-d)^{-1}>-\pi F^{-1/8}, \quad (35)$$

as the conditions for the applicability of Eq. (33) for the current. Both these conditions also imply that d is bounded below one, the requirement for the convergence of the integral in Eq. (33). When d is so small that $\pi d/\sin\pi d$ can be replaced by one, Eq. (33) for the current becomes the Richardson-Schottky formula. The thermionic emission boundaries as given by Eqs. (34) and (35) are shown in Fig. 2 for representative values of the parameters. These boundaries are only to be taken as indications of where Eq. (33) for the current begins to apply. The thermionic region is always bounded by the $d=1$ line and is limited by the melting temperature of the metal. An upper limit on thermionic fields is provided by the conditions $d=1$ and $kT=0.011$ corresponding to the melting point of tungsten at 3600°K. This limiting field is roughly $F=0.01$ as was used in the earlier arguments.

IV. FIELD EMISSION

In parallel with the treatment of thermionic emission, one can base a discussion of field emission on a certain type of approximate evaluation of the current integral, Eq. (20). The approximation is to use the first term in an expansion of the denominator-factor below the peak of the potential barrier and the first two terms in an expansion of the denominator-exponent about the Fermi energy:

$$\{1+\exp[(4/3)\sqrt{2}F^{-\frac{1}{4}}y^{-\frac{3}{2}}v(y)]\}^{-1}$$
$$=\exp[-(4/3)\sqrt{2}F^{-\frac{1}{4}}y^{-\frac{3}{2}}v(y)]$$
$$\times\{1-\exp[-(4/3)\sqrt{2}F^{-\frac{1}{4}}y^{-\frac{3}{2}}v(y)]\cdots\}, \quad (36)$$

$$-(4/3)\sqrt{2}F^{-\frac{1}{4}}y^{-\frac{3}{2}}v(y)$$
$$=-b+c(W-\zeta)-f(W-\zeta)^2\cdots, \quad (37)$$

where

$$b=(4/3)\sqrt{2}F^{-1}\phi^{\frac{3}{2}}v(F^{\frac{1}{2}}/\phi), \quad (38)$$

$$c=2\sqrt{2}F^{-1}\phi^{\frac{1}{2}}t(F^{\frac{1}{2}}/\phi), \quad (39)$$

$$f=\tfrac{1}{2}\sqrt{2}F^{-1}\phi^{\frac{3}{2}}(\phi^2-F)^{-1}v(F^{\frac{1}{2}}/\phi), \quad (40)$$

and the function $t(y)$ is defined by

$$t(y)=v(y)-\tfrac{2}{3}y\,dv(y)/dy. \quad (41)$$

One can express $t(y)$ directly in terms of elliptic functions by using properties of the derivatives of the elliptic functions with respect to the modulus.[18] The results, corresponding to Eqs. (14) and (16), are as follows:

$$t(y)=2^{-\frac{1}{2}}(1+a)^{\frac{1}{2}}E[(2a)^{\frac{1}{2}}/(1+a)^{\frac{1}{2}}], \quad (42)$$

$$t(y)=(1+y)^{-\frac{1}{2}}\{(1+y)E[(1-y)^{\frac{1}{2}}/(1+y)^{\frac{1}{2}}]$$
$$-yK[(1-y)^{\frac{1}{2}}/(1+y)^{\frac{1}{2}}]\}. \quad (43)$$

Numerical values of $t(y)$ for $y\leq 1$ can be easily found from Burgess, Kroemer, and Houston's tables[13]; in terms of the functions $v(y)$ and $s(y)$ which they tabulate, $t(y)$ is given by

$$3t(y)=4s(y)-v(y).$$

One finds Eqs. (37)–(40) by making a straightforward Taylor's series expansion of the left hand side of Eq. (37) about $W=\zeta=-\phi$ and using the derivative-properties of the elliptic functions to simplify the third term. When the first terms in Eqs. (36) and (37) are substituted into Eq. (20), the result is

$$j=\frac{kT}{2\pi^2}\int_{-W_a}^{W_l} e^{-b+c(W-\zeta)}\ln(1+e^{-(W-\zeta)/kT})dW$$
$$+\frac{kT}{2\pi^2}\int_{W_l}^{\infty}\ln(1+e^{-(W-\zeta)/kT})dW. \quad (44)$$

[17] See, for example, Herbert Bristol Dwight, *Tables of Integrals* (The Macmillan Company, New York, 1947), revised edition, formula 856.2.

[18] See, for example, reference 16, Eqs. (710.00) and (710.02).

The region of temperature and field in which these approximations apply (discussed below) is here called the field emission region because in it the current is strongly field-dependent and the zero-temperature limit is the Fowler-Nordheim formula.

The second integral in Eq. (44) makes a negligible contribution to the current in the field emission region and so the conditions on the region are found from the first integrand. The energy η at which the integrand is a maximum is given by

$$ckT(1+e^{(\eta-\zeta)/kT})\ln(1+e^{-(\eta-\zeta)/kT})=1. \tag{45}$$

The dependence of η on ckT is illustrated by Fig. 3. By writing Eq. (45) in the limits of $(\eta-\zeta)/kT$ far below zero and far above zero, one finds that

$$\eta=\zeta-c^{-1} \quad \text{when} \quad ckT\ll 1, \tag{46}$$

$$\eta=\zeta+kT\ln[ckT/(2-2ckT)] \quad \text{when} \quad 1-ckT\ll 1. \tag{47}$$

Above the Fermi energy the first integrand in Eq. (44) has the energy dependence $e^{-W(1-ckT)/kT}$, and below the Fermi energy it has the energy dependence $(\zeta-W)e^{cW}$. In consequence of this and Eqs. (45) to (47) for the peak location, one writes

$$\zeta+kT(1-ckT)^{-1}>W>\zeta-2c^{-1}+Q \tag{48}$$

for an estimate of the energy range in which the contribution to the current is appreciable. Here Q is a positive quantity whose exact value does not influence the later calculations. The next question to be considered is the relation of the approximations made in obtaining the integrand to the energy range in Eq. (48). The approximation of disregarding the higher order terms in Eq. (36) applies roughly when

$$\exp[-(4/3)\sqrt{2}F^{-\frac{1}{4}}y^{-\frac{3}{2}}v(y)]<e^{-1}.$$

The extreme energy satisfying this condition can be adequately discussed in the small ϵ approximation of Eq. (22) which gives, for the extreme value,

$$\exp[\pi F^{-\frac{1}{4}}\epsilon]=e^{-1}, \tag{49}$$

or, equivalently,

$$W<-F^{\frac{1}{2}}-\pi^{-1}F^{\frac{3}{4}}. \tag{50}$$

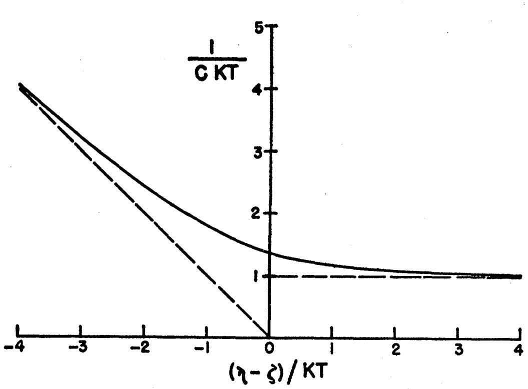

FIG. 3. Determination of peak energy η in the field emission region from the field, temperature, and work function, Eq. (45).

A condition on the field emission region is found by comparing this with Eq. (48):

$$\zeta+kT(1-ckT)^{-1}<-F^{\frac{1}{2}}-\pi^{-1}F^{\frac{3}{4}}. \tag{51}$$

The approximation of retaining only the linear terms in Eq. (37) must also be discussed. It will be assumed that the rest of the series is dominated by the quadratic term. Then the approximation applies when

$$f(W-\zeta)^2<\tfrac{1}{2},$$

since then the factor that it contributes to the integrand, $\exp[-f(W-\zeta)^2]$, is near unity. The requirement that this approximation be valid over the whole range given in Eq. (48) leads to two conditions:

$$\zeta-(2f)^{-\frac{1}{2}}<\zeta-2c^{-1}+Q, \tag{52}$$

$$\zeta+kT(1-ckT)^{-1}<\zeta+(2f)^{-\frac{1}{2}}. \tag{53}$$

One finds that, as long as Q is positive, Eq. (52) does not need to be considered because it is satisfied as long as Eq. (51) is satisfied.

Next the integrals in Eq. (44) for the current will be evaluated. For ordinary metals $-W_a$ is far below the lower boundary of Eq. (48); it will be assumed that it can be replaced by $-\infty$. From Eqs. (48) and (51) one sees that the important range of the first integral in Eq. (44) lies below the peak of the potential barrier at $-F^{\frac{1}{2}}$; the energy W_l lies above this value. Accordingly one assumes that only the first integral is significant in calculating the main contribution to the current and that further its upper limit may be extended to infinity:

$$j=\frac{kT}{2\pi^2}e^{-b}\int_{-\infty}^{\infty}e^{c(W-\zeta)}\ln(1+e^{-(W-\zeta)/kT})dW. \tag{54}$$

The expression for the current now becomes

$$j=\frac{kTe^{-b}}{2\pi^2 c}\int_0^{\infty}\frac{\nu^{ckT-1}}{1+\nu}d\nu =\frac{e^{-b}}{2\pi^2c^2}\frac{\pi ckT}{\sin(\pi ckT)}, \tag{55}$$

where the integration variable $\nu=e^{(W-\zeta)/kT}$ is introduced and an integration by parts is made in order to bring the integral to the same standard form[17] as arose in the thermionic emission discussion. The integral converges only if $ckT<1$; however this is guaranteed by Eq. (53). In summary, the current in the field emission region is given by the equation above and the region itself is defined by Eqs. (51) and (53). Using Eqs. (38) and (39) and simplifying, one can write the results as follows: The current is

$$j=\frac{F^2}{16\pi^2\phi t^2}\left(\frac{\pi ckT}{\sin\pi ckT}\right)\exp\left(-\frac{4\sqrt{2}\phi^{\frac{3}{2}}v}{3F}\right), \tag{56}$$

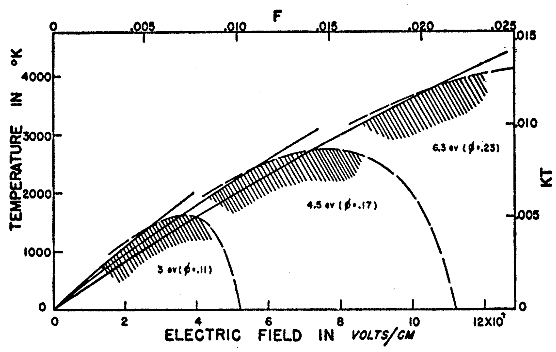

FIG. 4. Boundaries of the field emission region as given by Eq. (57), the broken lines, and by Eq. (58), the solid lines, for various values of the work function. The region lies under the two curves as indicated by the shading.

and the field emission region is defined by

$$\phi-F^{\frac{1}{2}}>\pi^{-1}F^{\frac{3}{4}}+kT(1-ckT)^{-1}, \tag{57}$$

$$1-ckT>(2f)^{\frac{1}{2}}kT, \tag{58}$$

where the arguments of v, t are $F^{\frac{1}{2}}/\phi$; c and f are defined by Eqs. (39) and (40). When ckT is so small that $\pi ckT/\sin(\pi ckT)$ can be replaced by one, Eq. (56) becomes the Fowler-Nordheim formula. The field emission boundaries as given by Eqs. (57) and (58) are shown in Fig. 4 for representative values of the parameters.

V. EMISSION IN THE INTERMEDIATE REGION

From Figs. 2 and 4 one can see that there is an appreciable region of temperature and field which is not covered by either of the above treatments. Another treatment of the integrals in Eq. (20) can be based on approximations suggested by Eq. (21) in Sec. III and Eq. (36) in Sec. IV. If only the first terms in these expansions are used, the expression for the current becomes

$$j=\frac{kT}{2\pi^2}\int_{-W_a}^{W_l}\exp\left[-\frac{W-\zeta}{kT}-\frac{4\sqrt{2}v(y)}{3F^{\frac{1}{4}}y^{\frac{3}{2}}}\right]dW$$
$$+\frac{kT}{2\pi^2}\int_{W_l}^{\infty}\ln\left(1+e^{-(W-\zeta)/kT}\right)dW. \tag{59}$$

The reason for studying these particular approximations is that the resulting expression for the current can be conveniently evaluated using the saddle-point method. The saddle point method is valid to a higher degree of accuracy than the approximations used to arrive at Eq. (59). Therefore, the conditions under which these approximations apply can be used to establish the boundaries of the temperature and field region for this type of emission. It turns out, as one would expect from the approximations used, that the range of temperature and field for this type of emission is intermediate between the thermionic and field emission regions. It develops that the second integral in Eq. (59) can be disregarded, and so the conditions on the region are found from the first integrand.

The condition on W so that only the first term in Eq. (36) can be used for the transmission coefficient is given by Eq. (50):

$$W<-F^{\frac{1}{2}}-\pi^{-1}F^{\frac{3}{4}}.$$

The ϵ approximation can be used to study the behavior of the first integrand in Eq. (59) for W in the neighborhood of $(-F^{\frac{1}{2}}-\pi^{-1}F^{\frac{3}{4}})$. Consequently the energy dependence of the integrand is represented by $\exp\{-W[(kT)^{-1}-\pi F^{-\frac{3}{4}}]\}$ for such W. So one condition on the temperature and field region for this type of emission is

$$\eta+[(kT)^{-1}-\pi F^{-\frac{3}{4}}]^{-1}<-F^{\frac{1}{2}}-\pi^{-1}F^{\frac{3}{4}}, \tag{60}$$

where η is the energy at the peak of the integrand and the term $[(kT)^{-1}-\pi F^{-\frac{3}{4}}]^{-1}$ takes account of the upper exponential tail. Differentiation of the exponent in Eq. (59) gives

$$\eta=-F^2/8(kT)^2t^2(F^{\frac{1}{2}}/-\eta). \tag{61}$$

Here t can be put equal to unity for a first approximation and for numerical calculations an iteration starting with $t=1$ will converge rapidly. It is convenient to rewrite Eqs. (60) and (61) in terms of $d=F^{\frac{3}{4}}/\pi kT$ as follows:

$$(F^{\frac{1}{2}}/-\eta)^{-1}>1+\pi^{-1}F^{\frac{1}{4}}d(d-1)^{-1}, \tag{62}$$

$$d=2\sqrt{2}\pi^{-1}(F^{\frac{1}{2}}/-\eta)^{-\frac{1}{2}}t(F^{\frac{1}{2}}/-\eta). \tag{63}$$

The first condition on T and F for this type of emission is found by eliminating $F^{\frac{1}{2}}/(-\eta)$ from these equations.

The approximation of using the first term in Eq. (21) for the logarithm is used in Eq. (59). This begins to apply when W becomes greater than $\zeta+kT$. First it will be shown that the condition imposed on the intermediate region by this approximation can be adequately discussed using only the linear terms in the expansion for $-(4/3)\sqrt{2}F^{-\frac{1}{4}}y^{-\frac{3}{2}}v(y)$ about $W=\zeta$, as given by Eq. (37). As argued following Eq. (51), this linear approximation applies when $W<\zeta+(2f)^{-\frac{1}{2}}$, and therefore holds in the region of $W=\zeta+kT$ as long as $(2f)^{-\frac{1}{2}}>kT$. In terms of F, kT, and ϕ, the condition

TABLE I. Values of the function $\Theta(y)$, as given by Eq. (76).

y	Θ	y	Θ
0	1.	0.55	1.6234
0.05	1.0104	0.6	1.7095
0.1	1.0362	0.65	1.7969
0.15	1.0739	0.7	1.8858
0.2	1.1215	0.75	1.9755
0.25	1.1772	0.8	2.0663
0.3	1.2398	0.85	2.1576
0.35	1.3084	0.9	2.2489
0.4	1.3816	0.95	2.3408
0.45	1.4590	1.	2.4318
0.5	1.5333		

$(2f)^{-\frac{1}{2}}>kT$ becomes

$$[\sqrt{2}F^{-1}\phi^{\frac{3}{2}}(\phi^2-F)^{-1}v(F^{\frac{1}{2}}/\phi)]^{-\frac{1}{2}}>kT, \qquad (64)$$

where Eq. (40) has been used for f. A useful numerical property of the functions v and t is that

$$(1-y^2)^{-1}v(y)t^2(y)\cong 1. \qquad (65)$$

The left-hand side is exactly unity when $y=0$; it departs from this value slightly as y increases, reaching a value of 1.03 when $y=1$. One can apply this in Eq. (64) and also put $t=1$ to obtain the simpler condition

$$\tfrac{1}{2}F^2\phi>(kT)^4. \qquad (66)$$

However Eqs. (62) and (63) imply that

$$d=F^{\frac{3}{4}}/\pi kT>1$$

for the region under study; also the condition $\frac{1}{2}F^2\phi>F^3/\pi^4$ is satisfied for all practical work functions and all attainable fields. The combination of these two gives Eq. (66) and this justifies the use of the linear approximation. Consequently the energy dependence of the integrand of Eq. (59) in the neighborhood of $W=\zeta+kT$ is $\exp\{W[c-(kT)^{-1}]\}$ and the corresponding condition on the intermediate region is

$$\eta-[c-(kT)^{-1}]^{-1}>\zeta+kT. \qquad (67)$$

Equation (67) can be rewritten as

$$\eta>\zeta+kT[1-(ckT)^{-1}]^{-1} \qquad (68)$$

and then, substituting from Eq. (61) for η, from Eq. (39) for c, and $-\phi$ for ζ, one finds

$$-\frac{F^2}{8(kT)^2t_\eta^2}>-\phi+kT\frac{1}{1-F(2\sqrt{2}\phi^{\frac{1}{2}}kTt_\phi)^{-1}}, \qquad (69)$$

where $t_\phi=t(F^{\frac{1}{2}}/\phi)$ and $t_\eta=t(F^{\frac{1}{2}}/-\eta)$. Equation (69) is the second condition for emission in the intermediate region.

The conditions on the intermediate region, as given in the preceding paragraphs, require that the energy range for which the integrand in Eq. (59) is appreciable lie between the Fermi energy and the peak of the potential barrier. This range lies between the limits of the first integral in Eq. (59) and so the second will be discarded. In addition it will be assumed that the saddle point method is adequate for evaluating the first integral. This requires an expansion of the second term in the exponent to three terms about the peak of the integrand η. The expansion is parallel to the one given in Eqs. (37) to (40); the result for the integrand is

$$\exp[-(W-\zeta)(kT)^{-1}-g+l(W-\eta)-n(W-\eta)^2],$$

where

$$g=(4/3)\sqrt{2}F^{-1}(-\eta)^{\frac{3}{2}}v(F^{\frac{1}{2}}/-\eta), \qquad (70)$$

$$l=2\sqrt{2}F^{-1}(-\eta)^{\frac{1}{2}}t(F^{\frac{1}{2}}/-\eta), \qquad (71)$$

and

$$n=\tfrac{1}{2}\sqrt{2}F^{-1}(-\eta)^{\frac{3}{2}}(\eta^2-F)^{-1}v(F^{\frac{1}{2}}/-\eta). \qquad (72)$$

When the current integration is performed, the result is

$$j=\tfrac{1}{2}(kT/\pi^2)(\pi/n)^{\frac{1}{2}}\exp[-g-(\eta-\zeta)(kT)^{-1}]. \qquad (73)$$

Equation (72) for n can be considerably simplified using Eq. (65), which applies very well in this region:

$$n\cong[\sqrt{2}F(-\eta)^{\frac{1}{2}}t^2]^{-1}. \qquad (74)$$

Equations (70) and (74) are to be substituted into Eq. (73) for the current. Then to show the primary dependence of the current on the temperature and the field one may substitute for η from Eq. (61) except where it appears in the argument of the relatively slowly varying functions v, t. The result is

$$j=\frac{F}{2\pi}\left(\frac{kTt}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{\phi}{kT}+\frac{F^2\Theta}{24(kT)^3}\right), \qquad (75)$$

where

$$\Theta=3t^{-2}-2vt^{-3}, \qquad (76)$$

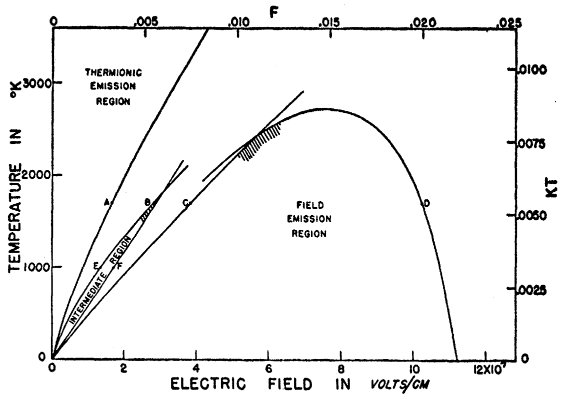

FIG. 6. The three emission regions for a 4.5 ev work function ($\phi=0.17$). The letters A to F indicate boundary points for which the approximate and exact energy distributions are given in Fig. 8. The points A to D are at 1700°K and E and F are at 1000°K. The values of the fields in volts/cm are as follows: A, 1.50; B, 2.50; C, 3.82; D, 10.26; E, 1.35; F, 1.74; all times 10^7.

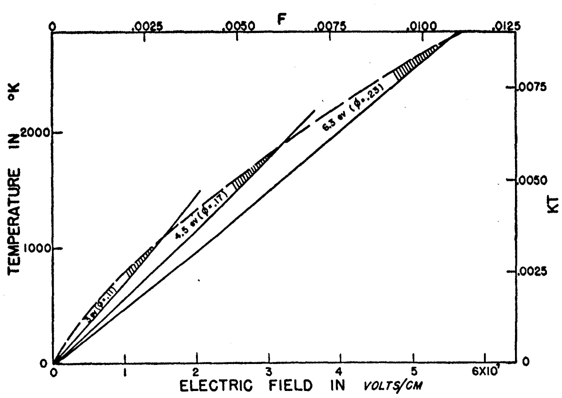

FIG. 5. Boundaries of the intermediate region as determined from Eqs. (62) and (63), the broken line, and from Eq. (69), the solid lines, for various values of the work function. The region lies between the two curves as indicated by the shading.

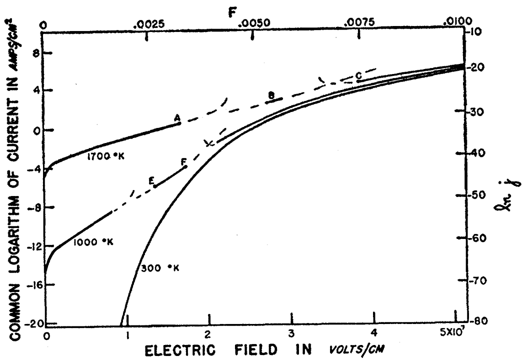

FIG. 7. Logarithm of the emitted current as a function of the applied field at various temperatures and for a work function of 4.5 ev ($\phi=0.17$). The curves are drawn according to Eqs. (33), (56), and (75). Solid lines are used for them in the regions where they apply (Fig. 6) and broken lines otherwise. The broken lines give a quantitative indication of the way the formulas depart from the correct values for the current outside the specified regions. The letters A to F indicate points for which the approximate and exact energy distributions are given in Fig. 8.

and the arguments of v, t are $F^{\frac{1}{2}}/(-\eta)$; η itself is to be found from Eq. (61). The function $\Theta(y)$ varies from 1 to 2.4 as y ranges from 0 to 1; some numerical values are given in Table I. This gives the final expression for the current in the intermediate region, where the boundaries of the region are given by Eqs. (62) and (63), and (69). Figure 5 shows the boundaries of the intermediate region for some representative values of the work function.

VI. DISCUSSION

The boundaries for the three types of emission are shown in Fig. 6 for a work function of 4.5 ev, a representative value for tungsten. One of the qualitative results of the above calculation is that there is an upper limit on the field for the applicability of the usual Fowler-Nordheim formula. These boundaries are only to be taken as an indication of the regions where the corresponding formulas for the current can be used. To illustrate their significance, the dependence of the logarithm of the emitted current on the applied field at several constant temperatures is given in Fig. 7. It is seen that for purposes of estimation the formulas can be applied somewhat outside the proper regions. One can see quantitatively how the approximate distributions of the current in energy correspond to the exact distributions from Fig. 8 in which the approximate and exact distributions are compared for several values of temperature and field on region boundaries. For points on the boundaries the error in the total current ranges between 15 and 40% and in the logarithm of the current between 0.1 and 1.0%.

The results obtained here have been compared with those of Dyke, Barbour, Martin, and Trolan[8] and of Dolan and Dyke,[7] found by numerical integration. Their results are mainly in the field emission region and extend occasionally into the intermediate and thermionic regions. In the field emission region a comparison is very easily made, for they give values of the current ratio $j(F,T,\phi)/j(F,0,\phi)$ and this, according to Eq. (56), is simply $\pi ckT/\sin(\pi ckT)$. The numerical agreement is satisfactory in the field and intermediate regions.

The criterion for the validity of the Richardson-Schottky emission formula within the thermionic emission region can easily be found by expanding the new factor in Eq. (33) for small πd as follows:

$$\begin{aligned}\pi d/\sin\pi d &= 1+\tfrac{1}{6}(\pi d)^2+\cdots \\ &= 1+F^{\frac{3}{2}}/6(kT)^2+\cdots.\end{aligned} \qquad (77)$$

Accordingly the fractional error involved in using the Richardson-Schottky formula is of the order of $F^{\frac{3}{2}}/6(kT)^2$. This is in complete agreement with Eq. (3) of Guth and Mullin[6] (to see the agreement one must expand for large values of their parameter $\mu=d^{-1}$ to order μ^{-2} and sum the resulting series). The factor $\pi d/\sin\pi d$ may also be expanded about $d=1$, giving in first approximation $(1-d)^{-1}$, so that Eq. (33) becomes

$$j=\tfrac{1}{2}\pi^{-2}(kT)^2(1-d)^{-1}\exp[-(\phi-F^{\frac{1}{2}})/kT], \quad (78)$$

in agreement with Sommerfeld and Bethe[5] and Eq. (4) of Guth and Mullin.[6] Since, as seen from Eqs. (34) and (35), the value $d=1$ lies completely outside the thermionic region, the expansion of Eq. (78) has a smaller domain of applicability than that of Eq. (77).

The criterion for the validity of the Fowler-Nordheim emission formula within the field emission region can easily be found by expanding the new factor in Eq.

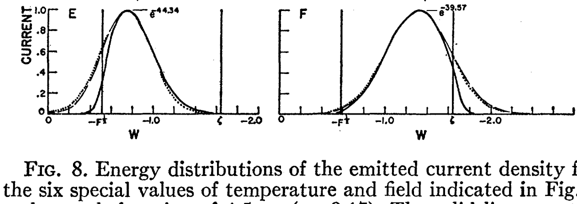

FIG. 8. Energy distributions of the emitted current density for the six special values of temperature and field indicated in Fig. 6 and a work function of 4.5 ev ($\phi=0.17$). The solid lines are the exact distributions according to Eq. (20). The broken lines are the approximate distributions according to Eqs. (25), (54), and (59). The dotted lines are the approximate distributions after furthermore applying the saddle-point method. The energy increases to the left. The energy at the peak of the potential $-F^{\frac{1}{2}}$ and the Fermi energy ζ are indicated by solid vertical lines. The curves are normalized to unit peak value and the actual peak values in Hartree units are given on the diagrams.

(56) for small πckT as follows:

$$\pi ckT/\sin(\pi ckT)=1+\tfrac{1}{6}(\pi ckT)^2+\cdots$$
$$=1+4\pi^2\phi(kT)^2t^2/3F^2+\cdots, \quad (79)$$

where here the argument of t is $F^{\frac{1}{2}}/\phi$. Accordingly the fractional error involved in using the Fowler-Nordheim formula is of the order of $4\pi^2\phi(kT)^2/3F^2$. This is in agreement with the results of Sommerfeld and Bethe[5] and Guth and Mullin, their Eq. (12)[6]; only the numerical factors differ slightly.