On the need for a tunneling pre-factor in Fowler–Nordheim tunneling theory

Richard G. Forbes^{a)}

Advanced Technology Institute (X1), Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom

(Received 18 December 2007; accepted 29 March 2008; published online 13 June 2008)

This paper argues that a tunneling prefactor should appear in expressions for the tunneling probability D relevant to cold field electron emission (CFE) and in Fowler–Nordheim (FN) type equations. Except in the case of the “ideally smooth” parabolic barrier, a prefactor is always present for barriers where D can be found by exact solution of the Schrödinger equation. A review of the Jeffreys–Wentzel–Kramers–Brillouin (JWKB) approach to solving the Schrödinger equation shows that tunneling barriers should be classified according to whether they are weak or strong and ideally smooth or not: there are four different JWKB-type formulas, depending on the nature of the barrier. CFE tunneling barriers are not ideally smooth but since the 1950s have usually been analyzed using the JWKB formulas for ideally smooth barriers. These analyses, and the standard Murphy–Good FN-type equation, seem mathematically and physically incomplete. The FN-type equations currently used to describe CFE should be revised to explicitly include a tunneling prefactor. Some implications are explored. © 2008 American Institute of Physics. [DOI: 10.1063/1.2937077]

I. INTRODUCTION

Fowler–Nordheim (FN) tunneling¹ is electric-field-induced tunneling through a roughly triangular barrier. When the barrier to emission from a material into vacuum is “strong,” in a sense defined below, there is a low-temperature electron emission regime (including room temperature) known as “cold field electron emission (CFE).” FN tunneling and CFE are processes of significant technological interest, especially in connection with the prevention of vacuum breakdown, the development of cold-cathode electron sources, and internal electron transfer processes in some types of electronic device. A more careful formulation of the theory would be helpful.

CFE is usually described by one or other of a large family of approximate equations, called Fowler–Nordheim-type (FN-type) equations. This paper is one of several by the present author^{2–5} that aim to progressively consolidate existing theoretical treatments of CFE into a more complete and coherent intellectual structure. It argues that a tunneling prefactor should appear both in the expression derived for the probability D of tunneling through any particular defined barrier, and in the FN-type equations derived by summing contributions to the emission current density (ECD) made by tunneling from all internal electronic states. Inclusion of the prefactor should eventually enable tunneling probabilities and current densities to be calculated slightly more consistently. However, more important, it helps make the theory physically more complete. The need to include this prefactor was briefly discussed elsewhere,⁵ as part of a wider discussion of correction factors. This paper presents detailed justification.

The following basic notation is used. Let e denote the elementary positive charge, m_e the electron mass in free

space, ϵ_0 the electric constant, h_p Planck’s constant (with $\hbar_p = h_p/2\pi$), and ϕ the local work function of the emitting surface. The field at the emitter surface, which determines the tunneling barrier, is denoted by F and called the “barrier field;” the ECD is denoted by J . In CFE theory, these positive quantities are the negative of the like-named quantities used in conventional electrostatics. The symbols $a[=e^3/8\pi\hbar_p]$ and $b[=(8\pi/3)(2m_e)^{1/2}/eh_p]$ denote the first and second FN constants.⁴

The structure of the paper is as follows. Section II provides theoretical background. Section III reviews barriers where the Schrödinger equation can be solved exactly to give an expression for D and shows that (except for the parabolic barrier) a tunneling prefactor is always present. Section IV re-examines Standard CFE theory in the light of this finding. Section V looks at the consequences of including this prefactor in the derivation of FN-type equations. Section VI notes other CFE contexts where a tunneling prefactor is needed. Section VII provides summary and conclusions.

II. THEORETICAL BACKGROUND

A. The Sommerfeld model

Arguments here are made in the context of emission from the flat planar surface of a bulk metal represented by a Sommerfeld-type free-electron model,^{6,7} as illustrated in Fig. 1. This simplifies much detail but does not affect the general validity of the conclusions. This model underlies many past discussions of CFE theory. The direction of the outward normal (from metal to vacuum) is termed the “forwards” direction.

The local depth of the potential-energy (PE) well is denoted by χ . For a given electron state A, the total electron kinetic energy is denoted by K , and its components parallel and normal to the emitter surface by K_p and K_n , respectively.

^{a)}Electronic mail: r.forbes@ieee.org.

FIG. 1. Schematic diagram illustrating the definitions of energies associated with the Sommerfeld model. The horizontal axis denotes distance measured from the emitter's electrical surface, and the vertical axis denotes "forwards" (or "normal") energy. See text for further explanation.

When $F=0$, an electron with a normal (forwards) kinetic energy K_n sees a barrier of "unreduced" (i.e., zero-field) height h ; hence, $K_n + h = \chi$.

The Fermi energy is denoted by K_F (hence $K_F + \phi = \chi$). The electron state with $K=K_F$, $K_p=0$, and $K_n=K_F$ is termed the "forwards state at the Fermi level" or "state F." An electron in state F sees a barrier of unreduced height ϕ . Parameters appropriate to state F, or the related barrier, are subscripted F .

The edge of the Sommerfeld well is aligned with the metal's electrical surface.^{8,9} Distance z normal to the emitting surface is measured outward from the electrical surface: this ensures that the electrostatic PE of the emitted electron has the usual limiting form $-eFz$. In simple models, the image plane coincides with the electrical surface:⁹ this allows the image PE to be written in the usual limiting form $-e^2/16\pi\epsilon_0 z$.

Let E denote the total electron energy and U the total electron PE, both measured relative to the same (arbitrary) reference level. The Schrödinger equation can be separated into components associated with motion parallel and normal to the emitter surface. That for motion normal to the surface can be written as

$$(\hbar^2/2m_e)\partial^2\psi/\partial z^2 + (E_n - U)\psi = 0, \quad (1)$$

where ψ is the electron wave function. The "normal energy" E_n (sometimes called the "total forwards energy") is the complete electron energy associated with motion normal to the emitter surface and is defined by $E_n = E - K_p$.

To simplify expressions below, we define

$$M(z) \equiv U - E_n \quad (2)$$

and call $M(z)$ the "motive energy" for the electron. This name derives from the term "motive" used by Herring and Nichols¹⁰ for a similar (but not exactly identical) concept. Inside the barrier, the motive energy is a positive quantity equal to the negative of the electron kinetic energy in the forwards direction.

The "shape" of a tunneling barrier is defined by $M(z)$. In CFE, the simplest shape models used are the exact triangular (ET) barrier $M^{\text{ET}}(z)$ and the Schottky–Nordheim barrier^{11,12} $M^{\text{SN}}(z)$ given by

$$M^{\text{ET}}(z) = h - eFz, \quad (3)$$

$$M^{\text{SN}}(z) = h - eFz - e^2/16\pi\epsilon_0 z. \quad (4)$$

B. FN tunneling from an individual state

For planar surfaces, the "core" forms of FN-type equation give J as a function of ϕ and F . (Nonplanar surfaces need at least one additional parameter to describe surface curvature.) Since neither F nor J can easily be measured directly, a complete theory also includes factors that relate J and F to the measured current I and voltage V , but we need not consider these factors here.

Derivation of the core form of FN-type equation has two main stages. First, the contribution made to J by an electron approaching the barrier in a particular electronic state is found. Then, these contributions are summed over all occupied electronic states.

To define individual states, we use the model of Ref. 3, Appendix. The emitter electrons are taken as free electrons confined/defined within a large rectangular box, where the front face of the box corresponds with the emitter surface. (To avoid problems in the statistics of counting states, a large box is needed.) Consider an electron in initial state i that has a component of current density in the forwards direction. Denote this current-density component by z_i when the state is fully occupied. When the state is partially occupied, with occupation probability f_i , the current-density contribution is $f_i z_i$. Denote the current density escaping from the state (by tunneling through the surface barrier) by $j_i(f_i)$. When state i is fully occupied ($f_i=1$), then a tunneling probability (or "escape probability") D_i is defined by

$$j_i(f_i=1) \equiv z_i D_i. \quad (5)$$

More generally, the contribution j_i made by the state to the ECD J is

$$j_i = f_i z_i D_i. \quad (6)$$

In most circumstances, it is adequate to take f_i as the Fermi–Dirac distribution function.

For some types of emitter, such as closed (capped) carbon nanotubes, the assumption that the box can be taken as large is obviously problematical. Such cases need a different approach, partly analogous to the treatment of field ionization,¹³ and are beyond the scope of this paper. (Although the conclusion is that a prefactor is present in such cases too).

In Eq. (6), the parameter j_i is easy to define because it represents current moving away from the emitter after escape. The parameter z_i needs a more careful examination. It is best interpreted as the current just inside the back face of the box. By implication, the model assumes that the only scattering/reflection process that affects the electron in state i is the partial reflection at the front face due to the tunneling

barrier. It is assumed that no inelastic scattering processes into or out of this state occur within the box and that there are no elastic reflection effects within the box as a result of potential-energy variations within it. Particularly for real emitters, both of these assumptions can be problematical. However, we need to make them at this stage of the discussion in order to focus on the physics of the barrier. These or similar assumptions have always been implicit in free-electron models of FN tunneling.

In this initial discussion, it is also assumed that the barrier is “well behaved” in the sense that tunneling probability decreases smoothly as h (the zero-field barrier height) increases. This will always be the case if there are no “resonance-type” effects associated with potential structure within the barrier and no back reflection into the barrier by structure beyond it. By making all these assumptions, we can focus on the physics of the tunneling barrier alone.

C. JWKB-type approximations

If expression (3) is inserted into Eq. (1), then the resulting ordinary differential equation (ODE) can be solved, as done by Fowler and Nordheim¹ in their original 1928 paper. However, if expression (4) is inserted into Eq. (1), then it is *impossible in principle* to solve the resulting ODE exactly in terms of the *ordinary* functions of mathematics. This conclusion seems to be a very firm result¹⁴ of continuing mathematical research that began nearly 200 years ago, and applies to most realistic barrier shapes.

In consequence, approximate solution methods have been developed. The best known are the class of approximations associated with names of Jeffreys,¹⁵ Kramer,¹⁶ Wentzel,¹⁷ and Brillouin¹⁸ (JWKB). These build on earlier mathematical results, including the work by Horn¹⁹ and by Fowler *et al.*²⁰ (see pp. 336–339), and since the 1920s have been developed further by very many other workers (for useful reviews, see Refs. 13, 21, and 22).

The name “JWKB approximation” is now ambiguous because it covers several different mathematical approximations and formulas that derive from a common historical origin. I prefer to write “JWKB-type” approximations and formulas. The oldest and most basic of these, formula (7) below, is often called the “JWKB approximation” (or the “WKB approximation,” although the work of Jeffreys¹⁵ is earlier); I now call this the “simple-JWKB formula.” With this formula, the tunneling probability D is given by

$$D \approx \exp[-G], \quad (7)$$

where the so-called JWKB exponent G is defined by

$$G \equiv g \int M^{1/2} dz. \quad (8)$$

The integral is taken over the range of z , where $M(z)$ is positive (i.e., over the barrier) and g is a universal constant (for a particle of given mass), given for an electron by

$$g = 2(2m_e)^{1/2}/\hbar_P \approx 10.24624 \text{ eV}^{-1/2} \text{ nm}^{-1}. \quad (9)$$

Since the work of Murphy and Good²³ in 1956, the simple-JWKB formula (7) has been very widely used in CFE theory.

WEAK ↓ STRONG G→large	Fröman and Fröman (FF) $D = Pe^{-G}/[1 + Pe^{-G}]$	Kemble $D \approx 1/[1 + e^G]$
	Landau and Lifschitz (LL) $D \approx Pe^{-G}$	Simple JWKB (s-JWKB) $D \approx e^{-G}$
NON-IDEAL $\Rightarrow$ IDEAL $P \rightarrow 1$		

FIG. 2. To show the relationships between the four formulas derived from JWKB-type approximations, for the calculation of electron escape probability D .

In the 1960s, Fröman and Fröman¹⁴ (FF) made a thorough mathematical exploration of JWKB-type approximations and developed a procedure for the exact evaluation of D . They concluded that descriptions of JWKB-type approximations in many textbooks of theoretical physics (including well-recognized ones) were incomplete and unsatisfactory, and that some were mathematically incorrect.

The FF procedure leads to the result [see their Eq. (9.10b)]

$$D = 1/[1 + |F_{22}(-\infty, +\infty)|^2 e^G], \quad (10)$$

where $F_{22}(-\infty, +\infty)$ is a parameter that, at least in principle, can be evaluated exactly by means of an algorithm involving repeated integrations of a matrix of coefficients along a path in complex space that stays well clear of the singularities at the barrier's classical turning points. They showed that their algorithm converges and that $|F_{22}(-\infty, +\infty) - 1|$ is bounded [see their Eq. (6.22b)].

Clearly, if a quantity p^{FF} is defined by

$$p^{\text{FF}} = |F_{22}(-\infty, +\infty)|^{-2}, \quad (11)$$

then Eq. (10) can be put in the form

$$D = Pe^{-G}/(1 + Pe^{-G}), \quad (12)$$

where P is given by p^{FF} . We call P the “tunneling prefactor,” and Eq. (12) the “Fröman and Fröman (FF) formula.”

As shown in Fig. 2, expression (12) is a unifying formula for three other JWKB-type formulas in the literature, namely, the simple-JWKB formula above, the Kemble formula,^{24,25} (which was rederived later, by a slightly different method, by Miller and Good²⁶), and a formula stated (without proof) by Landau and Lifschitz²⁷ (LL) [see their Eq. (50.12)]. It also allows the reflection coefficient R (for electrons reflected by the barrier) to have the easily derived form

$$R = 1 - D = 1/(1 + Pe^{-G}). \quad (13)$$

The three approximate formulas in Fig. 2 are obtained by making one or both of the approximations “ $e^{-G} \ll 1$ ” or “ $P \rightarrow 1$ ” in the FF formula. This suggests a two-way classification of barriers as (a) either “weak” ($e^{-G} \sim 1$) or “strong” ($e^{-G} \ll 1$) and (b) either ideally smooth (“ideal”) ($P=1$) or not ideally smooth (“nonideal”) ($P \neq 1$).

An earlier paper⁵ used the terminology “smooth” ($P=1$) or “sharp” ($P \neq 1$): the present terminology seems better, since some nonideal barriers are relatively smooth.

The question arises as to how to categorize the barrier models used in CFE. The CFE regime is defined as the re-

FIG. 3. To illustrate the definitions of parameters associated with tunneling through a rectangular barrier.

game where the barrier is strong, so the issue is whether CFE tunneling barriers are ideal or nonideal. The original FN treatment of an ET barrier generated a formula, Eq. (20) below, that shows that it is nonideal. However, for the last 50 years, most CFE theory has omitted the tunneling prefactor. The next section develops arguments to show that this is incorrect and that CFE tunneling barriers are nonideal. We do this by comparing the formulas shown in Fig. 1 with formulas derived in the literature for barriers for which the Schrödinger equation can be solved exactly.

III. EXACTLY SOLVABLE BARRIERS

The barrier shapes that yield exact solutions of the Schrödinger equation fall into two broad classes: those, such as the rectangular and triangular barriers, where $M(z)$ may have discontinuities and dM/dz and/or d^2M/dz^2 may have discontinuities or singularities, and those where there are no such discontinuities or singularities. The latter class has a case of special interest, the parabolic barrier, where $d^2M/dz^2 = \text{const.}$

A. Barriers involving discontinuities

Rectangular barrier. The tunneling probability for the symmetrical rectangular barrier shown in Fig. 3 is usually written (e.g., Ref. 27, p. 75) as

$$D^{\text{rb}} = 4k_1^2 k_2^2 [4k_1^2 k_2^2 + (k_1^2 + k_2^2)^2 \sinh^2(k_2 L)], \quad (14)$$

where L is the barrier length, $k_1 = [(2m_e K_n)^{1/2} / \hbar_p]$ is the electron wave number in regions 1 and 3, and $k_2 = [(2m_e h)^{1/2} / \hbar_p]$ is the decay constant in region 2. For this barrier, $G = 2k_2 L = gh^{1/2} L$, and we can rewrite Eq. (14) in terms of energies as

$$D^{\text{rb}} = 4K_n h [4K_n h + (K_n + h)^2 \sinh^2(G/2)]. \quad (15)$$

On defining $p^{\text{rb}} \equiv 16K_n h / (K_n + h)^2$, this can be rewritten as

$$D^{\text{rb}} = p^{\text{rb}} e^{-G} / [1 + (p^{\text{rb}} - 2)e^{-G} + e^{-2G}]. \quad (16)$$

Clearly, if G is sufficiently large (i.e., if the barrier is strong), then we may neglect the terms in e^{-G} and e^{-2G} in the denominator, yielding

$$D^{\text{rb}} \approx p^{\text{rb}} e^{-G}. \quad (17)$$

ET barrier. For the ET barrier, it can be shown that $G = bh^{3/2} / F = (2g/3)h^{3/2} / eF = (2/3)gh^{1/2} L$, where $L = [\hbar / eF]$ is the barrier length. In the present notation, the full FN result, as stated at the top of p. 178 in Ref. 1, gives the tunneling probability as

$$D^{\text{FN}} = 4e^{-G} h^{3/2} (eF)^{-1} \left\{ [K_n^{1/4} (h/eF)^{1/2} + h(eF)^{-1/2} K_n^{-1/4} e^{-G}]^2 + (2/g)^2 K_n^{-1/2} \left[\frac{1}{2} (eF/h)^{1/2} + h(eF)^{-1/2} (g/2)a \right]^2 \right\}, \quad (18)$$

where α is the quantity used by FN. Note that their $(C - W)$ is our h , their W is our K_n , their κ is our $g/2$, their β is our e^{-G} , and, at this point in their paper, their F is the energy gradient here denoted by eF . If we define

$$p^{\text{FN}} \equiv 4K_n^{1/2} h^{1/2} / (K_n + h), \quad (19)$$

and write $\chi = K_n + h$, then after some manipulation, including setting $\alpha \approx 1$ (as FN do), we obtain

$$D^{\text{FN}} \approx p^{\text{FN}} e^{-G} \left\{ 1 + (h/\chi) \left[\left(\frac{4}{3} \right) G^{-1} + \left(\frac{4}{9} \right) G^{-2} \right] + \left(\frac{1}{2} \right) p^{\text{FN}} e^{-G} + (h/\chi) e^{-2G} \right\}. \quad (20)$$

The factor (h/χ) is always less than 1, so we may neglect the terms in the denominator in e^{-G} and e^{-2G} when G is sufficiently large, obtaining

$$D^{\text{FN}} \approx p^{\text{FN}} e^{-G} \left\{ 1 + (h/\chi) \left[\left(\frac{4}{3} \right) G^{-1} + \left(\frac{4}{9} \right) G^{-2} \right] \right\}. \quad (21)$$

If G is large enough, then the terms G^{-1} and G^{-2} will be small in comparison with unity, and Eq. (21) yields

$$D^{\text{FN}} \approx p^{\text{FN}} e^{-G}. \quad (22)$$

Similar behavior is exhibited by Jensen's approximate result for an ET barrier [see Ref. 28, Eq. (16), and also Ref. 29], which can be rewritten as

$$D^J \approx p^{\text{FN}} e^{-G} \left[1 + \left(\frac{1}{2} \right) p^{\text{FN}} e^{-G} + \left(\frac{1}{4} \right) e^{-2G} \right]. \quad (23)$$

If G is sufficiently large, then the expression reduces to Eq. (22).

For this ET barrier, an alternative formula was derived by Frank and Young.³⁰ To represent the outgoing electron, they used a quasiclassical JWKB-type wave function (rather than FN's Hankel function), achieving the result [their Eq. (14)]

$$D^{\text{FY}} \approx 4K_n^{1/2} h^{1/2} e^{-G} / [(K_n^{1/2} + eF e^{-G} / 4g + h^{1/2} e^{-G/2})^2 + (h^{1/2} + eF / 2g + K_n^{1/2} e^{-G/2})^2]. \quad (24)$$

In practice, the terms involving F are always significantly smaller than $K_n^{1/2}$ and $h^{1/2}$ and may be disregarded in a first approximation. In the strong-barrier limit, the formula then reduces to the FN formula (22).

There is a pattern in these approximation sequences. The exact expression derived for D depends on the barrier shape and on the details of the mathematics used. However, applying the strong-barrier approximation (that G is "sufficiently large") then yields an expression of form $P e^{-G}$, where $P \neq 1$ and the approximated form for P depends only on barrier shape.

FIG. 4. The Eckart potential, as described by Eq. (26). The potential is illustrated for $A=-1$ and the values of B shown.

B. The parabolic barrier

The opposite extreme to these “sharp” barriers is the parabolic barrier $M^{\text{pb}}(x) = h - \frac{1}{2}\kappa x^2$, where x is measured from the barrier apex and κ is the constant curvature. For this barrier, $G = \pi gh / (2\kappa)^{1/2} = (\pi/4)gh^{1/2}L$, where L is the barrier length. The barrier can be described as ideally smooth because $d^2M/dx^2 = \text{const}$. It was analyzed by Kemble,^{24,25} but an elegant proof is given by LL (Ref. 27, p. 178), who, like FF, considered the wave-function behavior along a large semicircle in a complex space. The outcome is the formula

$$D^{\text{pb}} = 1/(1 + e^G) = e^{-G}/(1 + e^{-G}). \quad (25)$$

When G is large, this reduces to $D^{\text{pb}} \approx e^{-G}$. For this barrier, the prefactor in the FF formula (12) is, in some sense, *physically* equal to 1.

C. The Eckart barrier

A significant test case is the Eckart barrier,³¹ for which the motive energy is

$$M^{\text{EC}}(x) = Ae^{kx}/(1 + e^{kx}) + Be^{kx}/(1 + e^{kx})^2, \quad (26)$$

where k is a constant that can be written in the form $k = 2\pi/l$. The distance $2l$ is a measure of the length of the barrier, as shown in Fig. 4. For this barrier, there are no singularities or discontinuities in $M(x)$, but it is not as smooth as the parabolic barrier because d^2M/dx^2 varies with position. It is sufficient to consider the symmetrical barrier obtained by putting $A=0$. For this symmetrical Eckart barrier, it can be shown, from his Eq. (17b), that

$$D^{\text{EC}} \approx p^{\text{EC}} e^{-G}, \quad (27)$$

$$G = gl(\chi^{1/2} - K_n^{1/2}), \quad (28)$$

$$p^{\text{EC}} \approx 1 - \exp[-gIK_n^{1/2}], \quad (29)$$

where χ is the well depth (Eckart's V_m) and K_n is the forwards kinetic energy (Eckart's W). Obviously, for this bar-

rier, the prefactor is present in principle but would usually be very close to unity.

IV. RE-EXAMINATION OF STANDARD CFE THEORY

Thus, for all barriers tested here, except the parabolic barrier, the expression for D reduces to the Landau and Lifschitz¹ form $D \approx Pe^{-G}$ ($P \neq 1$), when the strong-barrier approximation is applied. In particular, this is the form of the original FN result. An important question is why the standard Murphy–Good (MG) treatment of CFE generates a formula with $P=1$. The immediate reason is that the MG treatment starts from the Kemble formula, which reduces to the simple-JWKB formula when G becomes large. MG used the Kemble formula because this is the stated outcome of the earlier analysis by Miller and Good.²⁶

Like the JWKB-type approximation, the Miller–Good approximation is a quasiclassical approximation. However, it takes a slightly different mathematical approach to the resolution of the difficulties associated with the classical turning points at ends of the barrier. (As a turning point is approached, the quasiclassical electron wavelength becomes infinite, and the conditions for the applicability of a normal JWKB approach break down.) The Miller–Good approach makes use of auxiliary functions, which must have the same general mathematical properties as the correct solutions but otherwise may be chosen arbitrarily. In practice, the auxiliary functions they chose were related to the solutions of the Schrödinger equation for a parabolic bonding well.

In the present notation, the MG stated result [their Eq. (37)] is $D \approx 1/(1 + e^G)$. That is, they confirm that D is approximately given by Kemble's²⁴ formula. Details of the final steps in this derivation are not given, and I am unable to establish precisely what mathematical approximations were made, but it is clear that this is not an exact result. This is confirmed by a test conducted by Miller and Good.²⁶ They applied their formula to a particular case of the Eckart potential, for which exact results can be derived. In circumstances where $D \approx 0.4$, the exact result for D is about 20% higher than their formula predicts.

In the strong-barrier limit, their formula becomes identical with the simple-JWKB formula $D \approx e^{-G}$. From the trend shown in their Fig. 3, it seems clear that their formula will underpredict the exact result. This means that the correct result should contain a correction factor that can be interpreted as a tunneling prefactor. However, the correction factor looks to be close to unity.

The following suspicion also exists but is difficult to confirm without extensive new analysis. Namely, their use of auxiliary functions that are the Schrödinger equation solutions for the (ideally smooth) parabolic barrier may have generated a result less general than originally believed.

Historically, what then seems to have happened is that, because the correction factor seemed close to unity, Murphy and Good adopted the Kemble/Miller Good formula as a starting point. MG do clearly describe it as an approximation [see MG's discussion relating to their Eq. (5)]. However,

users of their work have not sufficiently recognized that their results are not physically exact but are derived from an approximate starting formula.

V. THEORETICAL CONSEQUENCES FOR FN-TYPE EQUATIONS

A. Effect on form of equations

We next look at consequences of using the LL formula, rather than the simple-JWKB formula, for the integration over internal electronic states that leads to FN-type equations. The treatment here generalizes the Forbes³ 2004 approach and applies to a strong barrier of arbitrary but well-behaved shape.

Reference 3 shows that, for emission at zero temperature from a metal-like free-electron band, this integration yields a FN-type equation that can be written in the general form

$$J = Z_F D_F = z_S d_F^2 D_F, \quad (30)$$

where $z_S [=4\pi m_e/h_p^3 = 1.618\,311 \times 10^{14} \text{ A m}^{-2} \text{ eV}^{-2}]$ is a universal constant⁷ here called the Sommerfeld electron supply density, d_F and D_F are, respectively, the “decay width at the Fermi level” and the “escape probability at the Fermi level” (i.e., parameters for a barrier of unreduced height ϕ), and $Z_F [=z_S d_F^2]$ is a quantity I call the “effective electron supply.”

Reference 3 also shows that when the simple-JWKB formula is applied to the ET barrier, G_F , d_F , and Z_F are given (in this “elementary” theory) by the quantities G_F^{el} , d_F^{el} , and Z_F^{el} defined by

$$G_F^{el} \equiv G^{el}|_{h=\phi} = b\phi^{3/2}/F, \quad (31a)$$

$$1/d_F^{el} \equiv \partial G^{el}/\partial h|_{h=\phi} = (3/2)b\phi^{1/2}/F = (3/2)G_F^{el}/\phi, \quad (31b)$$

$$Z_F^{el} \equiv z_S (d_F^{el})^2 = a\phi^{-1}F^2. \quad (31c)$$

When the simple-JWKB formula is applied to the arbitrary well-behaved barrier, Ref. 3 shows that G_F , $1/d_F$, and Z_F are obtained by multiplying the three equations above by correction factors ν_F , τ_F , and τ_F^{-2} , respectively. The precise form of these correction factors depends on the barrier shape.

In deriving these formulas, the integration over occupied states requires that the escape probability $D(h)$ be expanded in the form

$$D(h) = D_F \exp(-\delta h/d_F), \quad (32)$$

where $\delta h = h - \phi$. When the LL formula (rather than the simple-JWKB formula) is used, the full formal definition of decay width d has to be used. As shown in the Appendix, when this full definition is applied to the LL formula for D , we obtain

$$1/d \equiv -\partial \ln\{D\}/\partial h = \partial G/\partial h - \partial \ln\{P\}/\partial h. \quad (33)$$

Obviously, the last term can be omitted when P is approximated as 1 (or as a constant), so the simplified version of Eq. (33) has been used in previous papers.

It is now shown how the present treatment (based on applying the LL formula to the arbitrary well-behaved barrier) relates to the elementary treatment (based on applying

the simple-JWKB formula to the ET barrier). It seems best to have τ retain its existing definition. By analogy with Ref. 3, if d_F can be written in the form

$$d_F = \lambda_D^{1/2} d_F^{el}, \quad (34)$$

where λ_D is a correction factor defined via Eq. (32), then Z_F can be written in the form

$$Z_F = \lambda_D a \phi^{-1} F^2. \quad (35)$$

Using definition (33) and noting that for an arbitrary well-behaved barrier $\partial G^{el}/\partial h|_{h=\phi} = \tau_F/d_F^{el}$, we obtain

$$1/d_F = \tau_F/d_F^{el} - \partial \ln\{P\}/\partial h|_{h=\phi}, \quad (36)$$

$$\lambda_D^{-1/2} = d_F^{el}/d_F = \tau_F[1 - \tau_F^{-1} d_F^{el} \partial \ln\{P\}/\partial h|_{h=\phi}] \equiv \tau_F \lambda_P^{-1/2}, \quad (37)$$

where λ_P is a correction factor defined by this equation.

In this new treatment, the escape probability D_F is given by

$$D_F = P_F \exp[-\nu_F b \phi^{3/2}/F]. \quad (38)$$

The FN-type equation that results when the LL formula is applied to a strong barrier of well-behaved but arbitrary shape thus takes the form

$$J = \lambda_D a \phi^{-1} F^2 P_F \exp[-\nu_F b \phi^{3/2}/F]. \quad (39)$$

Elsewhere,⁵ I have called this the “general (zero-K) free-electron FN-type equation.” (The designation “zero-K” indicates that the derivation is strictly valid only at a temperature of 0 K; in practice, no significant error arises from applying zero-K formulas to emission at room temperature.)

The analysis of FN plots is also affected. Some general results for FN-type equations were given some years ago² and may be applied here by identifying λ and μ in Ref. 2 with $(\lambda_P \tau_F^{-2} P_F)$ and ν_F here. In particular, the intercept correction factor (denoted here by ρ rather than r —see Ref. 4) is given by

$$\rho = \lambda_D P_F \exp[-b\phi^{3/2} d\nu_F/dF]. \quad (40)$$

B. Discussion

Numerical estimates are now needed. Due to mathematical difficulties noted earlier, accurate calculations of the parameters P_F and λ_D for realistic barrier models cannot easily be made. However, illustrative estimates can be found from the FN prefactor p^{FN} . Using expression (19) above to substitute for the parameter P in Eq. (37) and the definitions of G_F^{el} and d_F^{el} , we obtain

$$\lambda_P^{-1/2} \approx 1 - \tau_F^{-1} (1/3)(K_F - \phi)/(K_F G_F^{el}). \quad (41)$$

As illustrative values, we take $K_F = 10 \text{ eV}$, $\phi = 5 \text{ eV}$, $G_F^{el} = 5$, and $\tau_F \approx 1$; these yield $P_F \approx 1.9$ and $\lambda_P \approx 1.07$.

Simple smooth (but nonideal) barrier models might predict values of P_F and λ_P closer to unity than these values, but (as noted below) real emitters might have values further from unity. On the face of things, the value of P_F might sometimes

be of practical significance, but value of λ_P is unlikely to be. A factor very similar to λ_P was disregarded by FN in their original treatment.¹

Since the derivation of a FN-type equation involves the integration of D over occupied states, it seems best to use the combined factor λ_D (rather than the clumsy combination $\lambda_P \tau_F^{-2}$) in FN-type equations. In the author's view, it should be sufficient in most predictive calculations to put $\lambda_D = 1$.

The arguments here suggest that the standard (MG) FN-type equation should be replaced by Eq. (39). In practice, as argued elsewhere,⁵ one might want to generalize Eq. (39) further by multiplying by a factor λ_B representing band-structure effects on electron densities of states and (perhaps) by a factor λ_T representing temperature effects.

VI. OTHER USES OF THE PREFACTOR

Including the prefactor in evaluations of D , and in the resulting FN-type equations, also brings within the scope of a single form of equation various effects that until now have been ignored or treated separately. These effects arise when we note that, in reality, an electron moves in the whole potential structure relevant to its coherence length, and take into account potential-energy variations that are more realistic than those used in the usual simple models. It is beyond this paper's scope to discuss numerics, but relevant effects include the following.

- (1) If one takes the front surface of the "large box" used in Sec. II at the inner edge of the tunneling barrier, then the sharp changes in potential at the "emitter surface" (however these are represented) will give rise to reflection-type effects and change the value of P . It is well established that electron reflection from this "join" region can give rise to observable phenomena, for example, the reflection factor in the Richardson equation (e.g., Refs. 10 and 32), the periodic field-dependent deviations in Schottky emission (e.g., Refs. 10 and 33), and the Jason³⁴ effect in surface field ionization.
- (2) In more realistic models^{32,35} that include atomic structure, the need to match the wave function in the barrier to an atomic-type wave function, at the outer edge of the atomic bonding well, will influence the value of P . Modinos³⁶ discussed these effects.
- (3) If there is a thin conducting film at the emitter surface, of material different from that of the underlying substrate and of thickness less than the coherence length of electrons approaching it from the substrate, then scattering or resonance-type effects may affect the value of P . The existence of such effects is well established, both theoretically and experimentally (e.g., Ref. 37). However, they have weaker experimental consequences than one might expect, both on the total energy distribution and in current/voltage characteristics, because they tend to be "smoothed out" in the integration over electronic states.
- (4) With semiconductors (and similar materials), field penetration and band bending at the emitter surface may lead to wave-function modification that affects the value of P .
- (5) If there is a sharp structure in the barrier itself and/or back reflection from structure beyond the barrier, then the value of P will be affected (or it may be necessary to use a more complex expression for the tunneling probability D that does not reduce easily to the LL form). Jensen^{28,29,38} discussed effects of this type.

Values of P resulting from these effects can sometimes be substantially different from unity. So, it seems best to always use the LL formula as the physically relevant formula for a strong barrier, even in circumstances where simple barrier models would suggest that P is close to unity.

VII. SUMMARY AND CONCLUSIONS

This paper has argued for the (re)introduction of a tunneling prefactor into expressions for the escape probability for FN tunneling and into FN-type equations. An approach based on Fröman and Fröman's work suggests that barriers can be classified using two characteristics: whether they are weak or strong and whether they are ideal or nonideal. As summarized in Fig. 2, this suggests the existence of four related JWKB-type formulas, namely the FF formula and three approximations, applicable to the different cases.

For the past 50 years, CFE theory has been developed mainly by means of the simple-JWKB formula (7) that strictly applies only to the ideal parabolic barrier. This is mathematically incorrect, and the resulting formulas are conceptually incomplete. Examination of various cases where the Schrödinger equation can be solved exactly confirms that, in the "strong-barrier" (CFE) regime, approximate expressions for tunneling probability D and FN-type equations should contain a tunneling prefactor. For the simple Schottky–Nordheim barrier model, and for other relatively smooth simple barrier models, P may be relatively close to unity. However, there is no guarantee that this is true for CFE from real emitters, where electrons move in a much more complicated potential structure, as they approach the emitter surface. The discussion in Sec. VI notes physical circumstances where P can be substantially different from unity. So, the prefactor P should always be explicitly present in expressions for D , and the factor P_F always present in FN-type equations.

When the integration over internal states is carried out, this generates a further correction factor λ_D that takes the place of the factor τ_F^{-2} that appears in the standard FN-type equation. This factor λ_D should in principle be present in FN-type equations but can be approximated to unity when making predictive calculations. The product $\lambda_D P_F$ should be present in expressions for the intercept correction factor ρ .

At present, numerical values for P and P_F are difficult to predict accurately, in part because the motion of the electron in the whole of the relevant potential structure (not just the barrier) needs to be taken into account. However, the presence of the factor $\lambda_D P_F$ in FN-type-equation pre-exponentials (along with a factor λ_B caused by band-structure effects) might sometimes be relevant to the interpretation of apparently anomalous experimental results.

Broadly analogous considerations about prefactors apply to emission theory in the higher-temperature regime where

the strong-barrier approximation is not valid, and something more general than the Kemble^{24,25} formula is needed. These issues are beyond the scope of the present paper but have been discussed by Jensen.^{28,29,38}

APPENDIX: FORMULA FOR DECAY WIDTH

To sum the contributions to ECD from individual internal states, the escape probability D must be expanded as a function of unreduced height h . The expansion is about the value $h = \phi$, in terms of $\delta h = h - \phi$. Because D decreases approximately exponentially as h increases, one seeks an expansion in the form

$$D(h) = D(\phi)K(\delta h), \quad (\text{A1})$$

where $K(\delta h)$ is a factor that needs to be found.

Taking natural logarithms, we have

$$\ln[D(h)] = \ln[D(\phi)] + \ln[K(\delta h)]. \quad (\text{A2})$$

This is a Taylor-type expansion of $\ln[D(h)]$ with respect to h if we take

$$\ln[K(\delta h)] = (\delta \ln D / \delta h)|_{h=\phi} \delta h \equiv -\delta h/d_F, \quad (\text{A3})$$

where d_F is defined by this equation. The quantity $\ln[K(\delta h)]$ is clearly negative, so the minus sign is inserted into definition (A3) to make d_F positive. If we use the LL formula for escape probability, then d_F is the value, for a barrier of unreduced height ϕ , of the quantity d given by

$$1/d = -(\delta \ln D / \delta h) = \partial G / \partial h - \partial \ln\{P\} / \partial h. \quad (\text{A4})$$

d is termed the “decay width,” has the dimensions of energy, and in practice is measured in eV. When h increases by d , then $\ln D$ decreases by (approximately) 1, and D by (approximately) the factor e .

¹R. H. Fowler and L. W. Nordheim, *Proc. R. Soc. London, Ser. A* **119**, 173 (1928).

²R. G. Forbes, *J. Vac. Sci. Technol. B* **17**, 526 (1999).

³R. G. Forbes, *Surf. Interface Anal.* **36**, 395 (2004).

⁴R. G. Forbes and J. H. B. Deane, *Proc. R. Soc. London, Ser. A* **463**, 2907 (2007).

⁵R. G. Forbes, *J. Vac. Sci. Technol. B* **26**, 788 (2008).

⁶A. Sommerfeld, *Naturwiss.* **15**, 825 (1927).

⁷A. Sommerfeld and H. Bethe, in *Handbuch der Physik*, edited by H. Geiger and K. Scheel (Springer, Berlin, 1933), Vol. 24/2, p. 333.

⁸N. D. Lang and W. Kohn, *Phys. Rev. B* **7**, 3541 (1973).

⁹R. G. Forbes, *Ultramicroscopy* **79**, 25 (1999).

¹⁰C. Herring and M. H. Nichols, *Rev. Mod. Phys.* **21**, 185 (1949).

¹¹W. Schottky, *Phys. Z.* **15**, 872 (1914).

¹²L. W. Nordheim, *Proc. R. Soc. London, Ser. A* **121**, 626 (1928).

¹³S. C. Lam and R. J. Needs, *Appl. Surf. Sci.* **76–77**, 61 (1994), and references therein.

¹⁴H. Fröman and P. O. Fröman, *JWKB Approximation: Contributions to the Theory* (North-Holland, Amsterdam, 1965).

¹⁵H. Jeffreys, *Proc. London Math. Soc.* **s2–23**, 428 (1925).

¹⁶G. Wentzel, *Z. Phys.* **38**, 518 (1926).

¹⁷H. A. Kramers, *Z. Phys.* **33**, 828 (1926).

¹⁸L. Brillouin, *Compt. Rend.* **183**, 24 (1926).

¹⁹J. Horn, *Math. Ann.* **52**, 340 (1899).

²⁰R. H. Fowler, E. G. Gallop, C. N. H. Lock, and H. W. Richmond, *Philos. Trans. R. Soc. London, Ser. A* **221**, 295 (1921), and references therein.

²¹J. Heading, *An Introduction to Phase-Integral Methods* (Methuen, London, 1962).

²²M. V. Berry and K. E. Mount, *Rep. Prog. Phys.* **35**, 315 (1972).

²³E. L. Murphy and R. H. Good, *Phys. Rev.* **102**, 1464 (1956).

²⁴E. C. Kemble, *Phys. Rev.* **48**, 549 (1935).

²⁵E. C. Kemble, *The Fundamental Principles of Quantum Mechanics* (McGraw-Hill, New York, 1937).

²⁶S. C. Miller, Jr. and R. H. Good, *Phys. Rev.* **91**, 174 (1953).

²⁷L. D. Landau and E. M. Lifschitz, *Quantum Mechanics* (Pergamon, Oxford, 1958).

²⁸K. L. Jensen, in *Vacuum Microelectronics*, edited by W. Zhu (Wiley, New York, 2001), Chap. 3.

²⁹K. L. Jensen, *Adv. Imaging Electron Phys.* **149**, 1 (2007).

³⁰N. H. Frank and L. A. Young, *Phys. Rev.* **38**, 80 (1931).

³¹C. Eckart, *Phys. Rev.* **35**, 1303 (1930).

³²A. Modinos, *Field, Thermionic and Secondary Electron Emission Spectroscopy* (Plenum, New York, 1984).

³³E. Guth and C. J. Mullin, *Phys. Rev.* **59**, 575 (1941).

³⁴A. J. Jason, *Phys. Rev.* **156**, 266 (1967).

³⁵J. W. Gadzuk and E. W. Plummer, *Rev. Mod. Phys.* **45**, 487 (1973).

³⁶A. Modinos, *Solid-State Electron.* **45**, 809 (2001).

³⁷V. G. Litovchenko, A. A. Evtukh, Yu. M. Litvin, N. M. Goncharuk, and V. E. Chayka, *J. Vac. Sci. Technol. B* **17**, 655 (1999).

³⁸K. L. Jensen, *J. Vac. Sci. Technol. B* **21**, 1528 (2003).