Chapter 9
Renewing the Mainstream Theory of Field and Thermal Electron Emission

Richard G. Forbes

Abstract Mainstream field electron emission (FE) theory—the theory normally used by FE experimentalists—employs a Sommerfeld-type free-electron model to describe FE from a metal emitter with a smooth planar surface of very large extent. This chapter reviews the present state of mainstream FE theory, noting aspects of the history of FE and thermal electron emission theory. It sets out ways of improving the theory's presentation, with the ultimate aim of making it easier to reliably compare theory and experiment. This includes distinguishing between (a) emission theory and (b) device/system theory (which deals with field emitter behaviour in electrical circuits), and between ideal and non-ideal device behaviours. The main focus is the emission theory. Transmission regimes and emission current density regimes are discussed. With FE, a method of classifying different FE equations is outlined. With theories that assume tunnelling through a Schottky-Nordheim (SN) ("planar-image-rounded") barrier, a careful distinction is needed between the barrier form correction factor ν ("nu") and the special mathematical function v ("vee"). This function v is presented as dependent on the Gauss variable x. The pure mathematics of $\mathrm{v}(x)$ is summarised, and reasons are given for preferring the use of x over the older convention of using the Nordheim parameter y [$=+\sqrt{x}$]. It is shown how the mathematics of $\mathrm{v}(x)$ is applied to wave-mechanical transmission theory for basic Laurent-form barriers (which include the SN barrier). A brief overview of FE device/system theory defines and discusses different auxiliary parameters currently in use, outlines a preferred method for characterising ideal devices when using FN plots and notes difficulties in characterising non-ideal devices. The chapter concludes by listing some of the future tasks involved in upgrading FE science.

R. G. Forbes (✉)
Advanced Technology Institute & Department of Electrical and Electronic Engineering, University of Surrey, Guildford GU2 7XH, UK
e-mail: r.forbes@trinity.cantab.net

© Springer Nature Switzerland AG 2020

G. Gaertner et al. (eds.), *Modern Developments in Vacuum Electron Sources*, Topics in Applied Physics 135, https://doi.org/10.1007/978-3-030-47291-7_9

9.1 General Introduction

Electron emission from solid surfaces has many practical applications. Theoretical understanding of emission processes and of emitter characterisation helps electron source development. Understandings of how emitters can be fabricated and maintained in operation, and of how they fail, and of related theory, are also needed, but are not covered here.

In Chap. 8, Jensen has set out a detailed theory of field electron emission (FE) and of the transition between FE and thermal electron emission (TE). The present chapter's main aims are to discuss the wider background of this theory, including different conventions and terminology in use, and to discuss some issues relating to the theory of current-voltage characteristics. The chapter is set in the general context of FE textbooks/handbooks published in the last 30 years or so [1–14], and of recent reviews relating to FE theory (e.g. [15–21]).

The author now separates *emission theory* and *device/system theory*. The former provides a formula for the local emission current density (ECD), as a function of local work function, barrier field and temperature. Device/system theory converts this to a relation between measured current I_{m} and measured voltage V_{m}. Particularly in FE contexts, significant additional theory may be needed to accomplish this, and/or to validly interpret measured $I_{\mathrm{m}}(V_{\mathrm{m}})$ characteristics. Interpretation of FE $I_{\mathrm{m}}(V_{\mathrm{m}})$ data was recently discussed elsewhere [20, 21], and only a brief update is given here. Detailed interpretation theory for thermionic emitters is not covered.

In modern emission theory, it makes no sense to discuss field and thermal effects separately. This chapter presents an integrated approach, but at a more qualitative level than Chap. 8, and has the following overall structure. Section 9.2 deals with technical conventions; Sects. 9.3, 9.4, 9.5 and 9.6 deal with emission theory; Sect. 9.7 deals briefly with device and systems theory as these affect FE behaviour; and Sect. 9.8 with the interpretation of Fowler-Nordheim plots. Section 9.9 indicates some future needs. Appendix 9.1 lists universal constants used in field emission, and Appendix 9.2 lists data relating to the high-precision calculation of the FE special mathematical functions.

9.2 Technical Conventions

FE literature uses various technical conventions. This section describes the main alternatives, and aims to encourage a clearer formulation of FE theory. For reference, Table 9.1 presents a list of the various acronyms and abbreviations used in this chapter.

Table 9.1 Main acronyms and abbreviations used in Chap. 9

BT	Barrier top	L	Local
BTE	Barrier-top electron emission	LAFE	Large-area field electron emitter
C	Characteristic	LF	Low flyover
CFE	Cold field electron emission	m	Measured
CHTE	Classical high-temperature emission	M	Macroscopic
DT	Deep tunnelling	MG	Murphy-Good
ECD	Emission current density	PE	Potential energy
EPE	Electron potential energy	PPP	Parallel planar plate
ET	Exactly triangular	QMTE	Quantum-mechanical TE
FE	Field electron emission	SI	Système Internationale (of units)
FEF	Field enhancement factor	SN	Schottky-Nordheim
FN	Fowler-Nordheim	SPME	Smooth-planar-metal-like-emitter
GB	General barrier	ST	Shallow tunnelling
HF	High flyover	TE	Thermal electron emission
ISQ	International System of Quantities	VCF	Voltage conversion factor
JWKB	Jeffreys-Wentzel-Kramers-Brillouin	VCL	Voltage conversion length
k	Kernel	XC	Exchange-and-correlation

9.2.1 Equation Systems

In the 1970s, new International Standards defined a preferred system of quantities and equations for scientific communication (particularly between theoretical physicists and engineers) and a related unit system, the SI [22]. Since 2009 this system of quantities and equations, which was derived from the old metre-kilogram-second-ampere (mksa) system, and uses ε_0 in Coulomb's Law, has been called the *International System of Quantities* (ISQ) [22, 23]. Where clarity requires below, ISQ quantities are subscripted "ISQ".

The ISQ superseded earlier systems, particularly the Gaussian centimetre-gram-second (cgs) systems, and "atomic units" systems. The ISQ and SI are "master systems" that can define quantities used in earlier systems and the dimensionalities of units used. Thus, Gaussian elementary positive charge is denoted by e_s [$\equiv e_\mathrm{ISQ}/(4\pi\varepsilon_0)^{1/2}$] and would have the SI unit $\mathrm{J}^{1/2}\ \mathrm{m}^{1/2}$, and the magnitude of Gaussian electrostatic field is denoted by F_s [$\equiv (4\pi\varepsilon_0)^{1/2}F_\mathrm{ISQ}$] and would have the SI unit $\mathrm{J}^{1/2}\ \mathrm{m}^{-3/2}$.

In the 1960s, FE papers often employed a hybrid convention that used a Gaussian system for equations but an ISQ-like system to present experimental results (because laboratory equipment was calibrated in volts—a unit that does not exist in Gaussian systems). In experimental contexts, fields were often measured in "V/cm", which was a bastard unit created by taking one part (the "V") from the engineers' system, and other (the "cm") from the physicists' system.

The main FE equations that differ between the two systems are those describing (9.1) a Schottky-Nordheim (SN) tunnelling barrier of zero-field height H and (9.2) the related Nordheim parameter y

$$M^{\mathrm{SN}}(z) \;=\; H - e_{\mathrm{ISQ}}F_{\mathrm{ISQ}}z - e_{\mathrm{ISQ}}^2/16\pi\varepsilon_0 z \;=\; H - e_{\mathrm{s}}F_{\mathrm{s}}z - e_{\mathrm{s}}^2/4z, \tag{9.1}$$

$$y \;=\; (e_{\mathrm{ISQ}}^3 F_{\mathrm{ISQ}}/4\pi\varepsilon_0 H^2)^{1/2} \;=\; (e_{\mathrm{s}}^3 F_{\mathrm{s}}/H^2)^{1/2}. \tag{9.2}$$

Here, z is the distance measured from the emitter's electrical surface, and $M^{\mathrm{SN}}(z)$ is the electron motive energy for the SN barrier (see Sect. 9.3.3).

In papers, no subscripts were attached to symbols, and the same names and symbols were often used for both Gaussian and ISQ-like quantities. This led to two unhelpful 'features': context-dependent interpretations for the symbols 'e' and 'F', often in the same paper and sometimes in the same equation, and sometimes a need to convert a numerical value of F_{ISQ} to a numerical value of F_{s}, or vice versa, using a 'hidden conversion factor'. Values for such factors were rarely if ever written down in FE papers or FE textbooks, and were (and are) none too easy to find in general textbooks (though web search engines make conversions easier, nowadays). They were known to FE research workers of that era, and constituted 'secret knowledge' sometimes needed to interpret FE literature of that era in quantitative detail. One of many purposes of the 1970s changes was to abolish the need for 'features' like the above, by defining a single preferred system based firmly on quantity calculus/algebra.

Most new FE literature now uses the ISQ, but most older FE textbooks (including the reprinted seminal textbooks of Gomer [24] and Modinos [2]) and parts of some more recent ones [6, 9] are written using this 1960s-style hybrid system, and some FE material continues to be presented in this way. However, for most subjects in most universities in the world, the ISQ has been used almost exclusively for teaching for thirty years or more, and the knowledge needed to deal quantitatively with material written using the old hybrid system is no longer part of the background of many younger FE researchers. Thus, critical paper and textbook equations written using the old system are, in a certain sense, written in a theoretical language often not fully understood by many younger researchers.

The point is that it is now 40 years or more since the Governments of the world, via their national standards bodies, agreed to introduce (what is now called) the ISQ as a common language for the scientific communication of equation-based arguments. Editors of modern journals have no obligation to accept for publication FE papers that are written with equations in Gaussian or other obsolete equation systems, any more than they would have an obligation to publish papers partly written in Latin (the language of scientific communication of Newton's day). I accept the reality that obsolete conventions take a long time to fade away, but I strongly urge that all new FE literature is published with all equations written using the ISQ, especially when the publication is in English. Remaining discussion in this chapter uses the ISQ, except where otherwise indicated.

9.2.2 Other International Conventions About Equation Form

Two other parts of the international conventions about equations deserve mention. First, equations need to be dimensionally consistent, valid in all relevant units, and (except in the case of some formulae involving logarithms) to be 'stand alone' in the sense they can be exactly interpreted without reference to statements in the surrounding text. For example, the ISQ version of (9.2) is written correctly in any of the forms in (9.3), but forms (9.4a) and (9.4b) are considered formally incorrect.

$$y = (e^3F/4\pi\varepsilon_0H^2)^{1/2} = c_SF^{1/2}/H$$
$$\cong (1.199985\,\text{eV V}^{-1/2}\,\text{nm}^{1/2})\,F^{1/2}/H \cong (3.794686\times10^{-5}\,\text{eV V}^{-1/2}\,\text{m}^{1/2})\,F^{1/2}/H\,. \quad (9.3)$$

$$y \cong 1.199985\,F^{1/2}/H, \quad (9.4a)$$

$$y \cong 3.794686\times10^{-5}\,F^{1/2}/H. \quad (9.4b)$$

Here, c_S is the *Schottky constant*, as defined in Appendix 9.1 (also see Sect. 9.3.4). This international convention is not always observed in current FE literature, but is strongly recommended. Usually, the simplest procedure is to put into the equation a symbol denoting a constant, and give its value separately, using appropriate units.

A second convention [25] is that, with the symbol $\ln\{x\}$, and related logarithmic symbols, the curly brackets around x are allowed to mean 'the numerical value of x, when x is measured in specified units'. The specified units should be given in the text or (with diagrams) in the axis label or the figure caption, except when the discussion is valid in all relevant units. Alternatively, write $\ln\{x/(\text{'units of } x\text{'})\}$, but this approach is often cumbersome.

9.2.3 The Meaning of the Symbol 'e'

Present international conventions use the symbol 'e' (typeset italic) to denote the elementary (positive) charge and the symbol 'e' (typeset upright) to denote the base of natural logarithms. However, some publishers are not systematic about this in their typesetting and may use italic 'e' to denote the base of natural logarithms. This can cause confusion, particularly when authors do not define the meaning of their symbol 'e' and when the typesetting is changed (or not corrected) during publication.

In his Chap. 8, and elsewhere, Jensen has avoided this problem by using 'q' to denote the elementary charge, and other authors also use this convention. The following is an unambiguous convention more consistent with the International Standard. Use 'e' to denote the elementary charge. Use '$\exp(x)$' to denote exponents, rather than e^{x} or e^{x}, and use $\exp(1)$ rather than e or e, except when

typesetting incompatible with the international conventions cannot conceivably be misunderstood.

9.2.4 Conventions Concerning the Term 'Field'

In most cases, the 'field' discussed in ordinary FE contexts is an electrostatic field. It is better to describe these fields as 'electrostatic' rather than 'electric', because the latter term also includes the electric component of a travelling electromagnetic wave. Laser-induced effects are of increasing interest in FE contexts (e.g. [26, 27]), but there are unresolved fundamental issues (e.g. [28]) relating to whether the physical effects of an electrostatic field are the same as those of an electromagnetic E-field, and whether an electrostatic field is physically the low frequency limit of an electromagnetic E-field. It is safer to assume that they may not be.

For clarity in what follows, I use the term 'classical' to label quantities that are defined strictly in accordance with the conventions of classical electrostatics as used in standard textbooks (e.g. [29, 30]). In a one-dimensional (z) situation, these imply the following. If under the influence of a classical electrostatic field, a positive test particle (with a vanishingly small charge) moves in the positive z-direction, then the field is considered positive. In FE, electron emission is usually deemed to take place in the positive z-direction. Hence, the classical electrostatic field $\boldsymbol{E}^{\mathbf{ES}}$ applied to a field electron emitter is negative in value, and so, of course, are the classical emission current and current density. In FE literature, five different conventions are used to deal with 'fields'.

(1) In detailed discussions of electrostatics as it applies to field emitters (for example, the electrostatic depolarization effects that occur in emitter arrays), it usually clearest to stick firmly to classical electrostatic conventions, and (where appropriate) insert modulus signs around the symbols for field and for currents and related quantities.

However, it is more usual for FE papers (and other electron emission papers) to use the so-called *electron emission convention* in which fields and electron currents and current densities are treated as if they were positive. In FE, this convention about fields comes in three variants.

(2) Most modern experimental FE papers use the symbol E to denote the *absolute magnitude* of a classical electrostatic field that is actually negative. Since this E is usually simply called the 'field', this can be a confusing convention for non-experts. It can also be unhelpful in electrostatics contexts.

(3) An alternative uses the symbol F to denote the absolute magnitude of the classical electrostatic field. This convention goes back to at least the work of Dyke and colleagues (e.g. [31]), and was used in Gomer's influential textbook [24]. It was used widely in experimental FE literature in the period 1960–1990, and is still used in much theoretical FE literature, because theoreticians often find

it useful to have 'E' available to denote electron energy. Use of F to denote positive electrostatic fields is much older; in particular, J.J. Thomson used F in his paper [32] announcing the identification of the electron.

(4) A formal alternative—which, in practice, is usually equivalent to the convention (3)—is to define a new field-like vector quantity $\boldsymbol{F}$ by

$$\boldsymbol{F} = -\boldsymbol{E}^{\mathrm{ES}}, \tag{9.5}$$

and to specify that the symbol F represents the *signed magnitude* of $\boldsymbol{F}$. Obviously, in the context of electron emission theory, both $\boldsymbol{F}$ and F are positive quantities.

(5) A fifth convention is to define a quantity F^* by the equation

$$\boldsymbol{F}^* = -e\boldsymbol{E}^{\mathrm{ES}}. \tag{9.6}$$

This quantity F^* (usually F in the literature) is positive in the context of FE, and is best defined—essentially following Stern et al. [33])—as the negative of the gradient of the electrostatic component of the potential energy (PE) of a hypothetical classical point electron. F^* has the SI unit 'newton'.

The quantity F^* has no well recognised name in quantum mechanics. Jensen [14] calls F^* a 'field', but this name invites confusion with the ISQ quantities measured in V/m. The least worst alternative might be to call F^* the 'classical electrostatic force' on the electron, though in reality F^* is the force on a hypothetical classical point electron, not the force on a real distributed wave-like electron that is taking part in a tunnelling process. In Chap. 8, this quantity F^* is denoted by the symbol 'F' and is called 'force'.

Merits of using F^* are that it is a positive quantity, and the same in both the Gaussian system and the ISQ. Its use also has the typographical advantage of removing the symbol for an elementary positive charge from various formulae. However, this is also its main disadvantage. Any formula that in conventions (1)–(4) contains a 'field' measured in V/m (or dimensionally equivalent units), would need modifying for the convention (5). Also, new versions of several universal constants used in FE would need to be introduced: these new constants would have different dimensions and units, though numerical values in FE customary units (see below) would not change.

Classical electrostatic force was used to describe FE potential-energy barriers in the 1929 Stern et al. paper [33], and (before that) in the middle part of the 1928 Fowler and Nordheim paper [34] (the technical meaning of the symbol "F" changes twice in the FN paper). If this were still the prevailing convention in FE, then probably one would not seek to change it. However, the author's view is that its widespread readoption at this point in time would inevitably introduce unfamiliarity and further confusion into an already confused subject, and would add to the difficulties of explaining FE theory clearly. Also, there might be intellectual resistance from many

experimentalists to the need to learn two slightly different sets of theoretical language in order to read basic material relating to FE theory. I thus prefer conventions (3) and (4) above, and use convention (3) in this chapter.

9.2.5 Field Emission Customary Units

In many situations, SI units are inconveniently large for discussing atomic-level issues. Thus, field electron and ion emission traditionally use a set of atomic-level units based on the eV as the unit of energy, the V/nm as the unit of electrostatic field (though sometimes the V/μm is more convenient), the elementary (positive) charge e as the size of the unit of charge (which, for technical reasons, has to be written as 1 eV V^{-1}), and the second (s) as the unit of time. This system has been called *field emission customary units*, and is discussed in more detail in [35]. Its merit is that it allows certain frequently needed basic FE calculations to be carried out quickly and efficiently.

It needs stressing that this is a system of units only, not a distinct equation system, and that (like SI units) it uses the ISQ as the underlying system of quantities and equations. Further, since all units in the FE customary system are officially recognised [22] for continued use alongside SI units, the system is entirely compatible with the ISQ and SI.

It is usual to express certain universal constants, widely used in FE theory, in these FE customary units. A list is given in Appendix 9.1. In this chapter, all relevant universal constants are stated to a precision of seven significant figures. Values have—in a few places where it is necessary—been updated to be consistent with the May 2019 changes in the SI system (see Wikipedia articles accessible via web search).

9.3 Emission Theory—General Issues

9.3.1 Smooth-Surface Conceptual Models

Except for carbon-nanotube emission theory and some other notable exceptions, e.g. [18, 36–38], nearly all present FE theory is formulated using what are called here *smooth-surface conceptual models* or "*Sommerfeld conductor*" *models* (see below). In such models, the atomic structure is disregarded. The emitter is represented, in a Sommerfeld-like way, as a region of space where the electron potential energy (EPE) is constant at the 'conduction-band-edge value' E_c, with this region limited by a 'formal well boundary'. Outside the boundary, the EPE is modelled either as rising abruptly to the local vacuum level E_{vac}, or as rising more slowly to this level, as a result of exchange-and-correlation (XC) forces. These forces give rise to an XC contribution to the EPE, with the contribution usually modelled as a classical

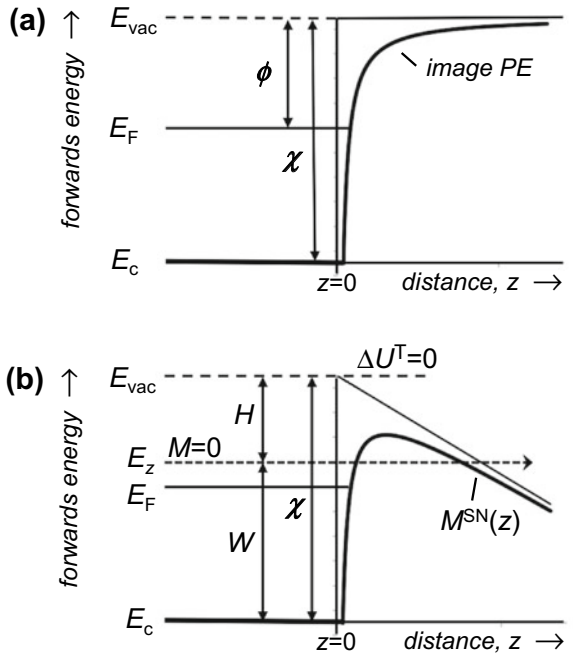

Fig. 9.1 Sommerfeld-type electron-energy diagrams, for an emitter with a smooth planar surface of large extent. **a** With image potential energy (PE) but in the absence of field. **b** The Schottky-Nordheim (SN) barrier (includes both image PE and field). Note that in both cases the vertical axis shows forwards energy, not total electron energy

image PE that cuts off at the point where it goes below E_c (see Fig. 9.1a, for the planar-surface case).

For simplicity, FE theory usually takes the local inner PE χ [$= E_{vac}–E_c$], and hence the local work function ϕ [$= (E_{vac}–E_F)$, where E_F is the emitter Fermi level], to be the same at all points on the well boundary. In fact, real emitters usually have many small surface areas (or "*patches*") with different local work functions, and (even in the absence of any applied voltage) are surrounded by a system of electrostatic fields called *patch fields*. With thermally induced emission, at very low applied voltages, these patch fields can cause significant local space-charge effects outside the emitter surface, well discussed in older review literature (e.g. [39]). In FE theory, the patch field adds to the voltage-induced component of surface field (and contributes to causing field inhomogeneities across the surface). However, for simplicity, patch fields are usually disregarded in mainstream FE theory.

Inside the well, electrons are treated, in the usual Sommerfeld way [40–42], as non-interacting electrons subject to the laws of quantum mechanics and statistical mechanics, with the electron distribution assumed to be in thermodynamic equilibrium and obedient to Fermi-Dirac statistics.

If the well boundary is assumed planar, and the well itself is assumed to be of infinite width (or of very large extent), then this is the usual well known 'Sommerfeld free-electron model'. However, advanced FE theory needs to consider equivalent models (without atomic structure) where the formal well boundary is curved and quantum confinement effects determine the electron wave-functions inside the well.

There seems to be no well recognised name for a model of this kind; I refer to it as a *Sommerfeld conductor*.

When an electrostatic field is created outside the emitter, by applying a voltage between it and a counter-electrode, the charge distribution induced at the emitter surface (by the voltage) is usually modelled (in a Sommerfeld conductor model) as confined to an infinitesimally thin surface layer. This surface layer defines the emitter's *electrical surface* within the framework of the model, and is taken as co-located with the formal well boundary. The resulting EPE variation is illustrated in Fig. 9.1b, for the Schottky-Nordheim barrier defined below. Note that the vertical axis represents the electron's *forwards energy*, i.e. the energy component associated with the direction normal to the emitter surface, which is called here the *forwards direction*. This name 'forwards energy' is an alternative name for what is more usually called the 'normal energy', and is used because it seems more instructive.

The electron wave-functions are assumed to take an appropriate non-atomic form, which will be a plane wave if the formal well boundary is planar or 'nearly planar'. These wave-functions are then used to evaluate values for the transmission probability D for an electron approaching the emitter surface from the inside, as a function of forwards energy and the *local barrier field* F_{L}. This field F_{L} is the local value of the parameter F, as taken in the emitter's electrical surface. For notational simplicity in what follows, we omit the subscript 'L' from local-field values.

9.3.2 The Problems of Smooth-Surface Models

Smooth-surface conceptual models disregard the existence of atoms, disregard the role of atom-level wave-functions in tunnelling theory and are not self-consistent with respect to the location of induced surface charge. Their use in FE theory can be described as *smooth-planar-metal-like-emitter (SPME) methodology.*

These physical assumptions are wildly unrealistic for real emitters, particularly for carbon nanotubes. Hence, the predictions of SPME methodology about current densities are not accurate. As discussed in Sect. 9.6.1, one can take this uncertainty into account by putting into core FE equations a general-purpose pre-exponential correction factor λ_{FE} of unknown functional dependence and unknown values. In the context of the Murphy-Good zero-temperature FE equation (see below), my present thinking [43] is that a smooth-surface model might do anything between under-predicting by a factor of around 14 and over-predicting by a factor of around 200. In other words, the relevant correction factor $\lambda_{\mathrm{FE}}^{\mathrm{SN}}$ has unknown functional dependence but probably lies in the range of $0.005 < \lambda_{\mathrm{FE}}^{\mathrm{SN}} < 14$. At present, we do not know how badly the models really behave, although one might speculate that over-prediction by a factor of 10 is more plausible than either extreme.

The problems of developing accurate FE theory are intensively difficult, and quite likely at the current boundaries of theoretical physics.[1] It seems likely that the *experimental* determination of λ_{FE} would be *relatively* easier and quicker. However, this is

[1] Prof. Marshall Stoneham (now deceased), a former President of the UK Institute of Physics, thought that some of the most difficult unsolved problems in theoretical physics were in field electron emission (private communication to the author, 2001).

relative. The necessary experiments seem likely to be difficult, time-consuming and expensive, and there seems no indication that funding bodies are currently inclined to fund basic physics work of this kind (or that experimentalists are prepared to carry it out).

Although sympathetic to the idea of advanced theoretical exploration, the author does not see how real progress will be made by adding purely theoretical investigation to purely theoretical investigation. Rather, I see the best route forwards as based on developing a more *scientific* approach to the smooth-surface conceptual models, that will enable them to be reliably tested against experiment, so that the actual size of discrepancies can be established. I see the immediate need as 'theoretical tidying up' of these models, starting with the planar-surface models. Later, one can investigate how to do reliable experiments, and seek support for this.

The main purpose of this chapter is to contribute to this "tidying up" of the planar-surface models. Beyond this, there are important issues relating to methods for more precise FE $I_{\mathrm{m}}(V_{\mathrm{m}})$ data interpretation for real emitters that have the shape of a pointed needle or a rounded post. This is an active topic of research (e.g., [14, 44, 45]), but details are beyond the scope of this chapter.

9.3.3 Barrier Form, Electron Motive Energy and Barrier Strength

The details of the barrier transmission theory depend on the barrier's mathematical form. This is best described via a quantity $M(z)$, called the electron motive energy, derived as follows: The Schrödinger equation is separated in Cartesian coordinates, and the component relating to motion normal to the emitter surface is written

$$\left[\hat{K}_z + \{U^{\mathrm{T}}(z) - E_z\}\right]\Psi_z = \left[\hat{K}_z + M(z)\right]\Psi_z = 0 \tag{9.7}$$

where z is the forwards direction, Ψ_z is the relevant wave-function component, $\hat{K}_z$ is the relevant kinetic-energy operator, $U^{\mathrm{T}}(z)$ is the total electron potential energy (EPE) and E_z is the forwards electron energy. $U^{\mathrm{T}}(z)$ and E_z (and also E_{c}, E_{vac} and E_{F}) can be taken as all measured relative to any well defined energy reference level (often the laboratory "Earth"). The symbol W is used to denote forwards energy measured relative to the PE well base (i.e. $W = E_z – E_{\mathrm{c}}$).

For an electron with forwards energy E_z, the *electron motive energy M(z)* is then defined by

$$M(z) = U^{\mathrm{T}}(z) - E_z = [U^{\mathrm{T}}(0) - E_z] + [U^{\mathrm{T}}(z) - U^{\mathrm{T}}(0)] \equiv H + \Delta U^{\mathrm{T}}(z), \tag{9.8}$$

where the *zero-field barrier height* $H \equiv \chi - W = U^{\mathrm{T}}(0) - E_z$ (see Fig. 9.1b). $\Delta U^{\mathrm{T}}(z)$ is defined by (9.8), and describes how the EPE varies outside the emitter surface, relative to the local vacuum level.

The term 'motive' was used by Newton [46], and also in early discussions of electron motion near conducting surfaces (e.g. [39, 47]), but the term 'motive energy' is used here in a slightly different and more specialised sense.

As is well known, different *barrier forms* are defined by selecting different mathematical forms for $\Delta U^{\mathrm{T}}(z)$ and hence $M(z)$. Those of most interest here are the planar forms called the *exactly triangular (ET)* barrier [defined by (9.9)] and the *Schottky-Nordheim (SN)* barrier [defined by (9.10)], but there is also current interest in forms that represent the EPE variation outside emitters with quasi-spherical apexes. For the ET and SN barriers, the motive-energy expressions are

$$M^{\mathrm{ET}}(z) = H - eFz \quad (z > 0) \tag{9.9}$$

$$M^{\mathrm{SN}}(z) = H - eFz - e^2/16\pi\varepsilon_0 z \quad (z > z_{\mathrm{c}}) \tag{9.10}$$

where the *cut-off distance* z_{c} is defined by the condition $M^{\mathrm{SN}}(z_{\mathrm{c}}) = H - \chi = -W$. To the left of these regions of validity we have $M(z) = H - \chi = -W$ (see Fig. 9.1).

9.3.4 Image Potential Energy and the Schottky Effect

This Section relates to the *image PE term* ($-e^2/16\pi\varepsilon_0 z$) in (9.10). As is well known, this describes the electrical PE of interaction between a classical point electron and a perfectly conducting infinite planar conductor, and is a model for quantum-mechanical exchange-and-correlation effects outside planar surfaces. The complicated historical origin of this term seems less well known, and is recorded here.

The idea of electrical images was introduced by W. Thomson (later Lord Kelvin) in 1847 [48], when discussing the reaction of an isolated conducting sphere to external charges. Later [49 but (better) see p. 67 in 50], he showed that, if the sphere were charged and both the sphere and an external point charge had the same sign, the force on the external charge would change from repulsive to attractive as the external charge is put closer to the sphere, i.e. (in modern terms) a PE barrier would exist.

In fact, the first person to use electrostatic potentials and electrostatic energies of mutual interaction, when discussing image effects relating to spherical conductors, was Maxwell [51]. These early treatments were algebraically complicated, and not fully correct in the 1st edition of [51]; a clearer discussion was given later by Landau et al. (see §3 in [30]).

Thomson [49 but (better) see p. 73 in 50] and Maxwell [51] also discussed image effects outside an (infinite) *planar* surface, but their thinking was too heavily linked to the spherical-conductor situation, and neither provided any formula useful to FE.

Thus, the Gaussian *planar* image-PE formula for an electron ($-e_s^2/4x$) is usually attributed to Schottky [52], although in fact, his 1914 discussion was in terms of the related positive *potential* due to the image charge. Schottky [52] also considered the case where the Gaussian classical electrostatic potential $\Delta\Phi_s$ (measured relative to the local vacuum *potential* level) increases as $\Delta\Phi_s = b_S z$. He found the Gaussian expression $(e_s b_S)^{1/2}$ for the height (above the vacuum potential level) of the minimum in the effective potential for an electron.

Obviously, the modern interpretation is that the image force causes a reduction, Δ_S, below the local vacuum *electron-energy* level, of the classical barrier for electron emission. This is the *Schottky effect* and we can write (using the ISQ)

$$\Delta_S = c_S F^{1/2} \equiv (e^3/4\pi\varepsilon_0)^{1/2} F^{1/2} \tag{9.11}$$

where c_S is the *Schottky constant* [$\cong$1.199985 eV $(\text{V/nm})^{-1/2}$, $\cong 3.794686\times10^{-5}$ eV $(\text{V/m})^{-1/2}$]. Schottky [52] gives, in his equation (6), the approximate numerical value of this constant, but for ISQ fields in V/cm, rather than V/m.

In fact, there had been an earlier suggestion that (what we now call) the Schottky effect could operate in electron emission. In his 1903 textbook (see p. 386 in [53]), J.J. Thomson suggested qualitatively that anomalies in spark physics, observed by Earhart [54] in 1901 for very small electrode gaps (less than 2 μm), might be explained if the combination of electrostatic and image forces made it energetically easier for electrons to leave the cathode. This proposal is clearly the Schottky effect (ten years before Schottky), but we would probably now attribute these anomalies to FE from nanoprotrusions, not observable in 1901, that had field enhancement factors of around 10–15.

Schottky [52] also noted the possible effect of this barrier reduction on electron emission, and then returned to the subject is his later (1923) paper [55], where he gives a formula[2] for (what I call) the *SN barrier reference field* F_R. This is the field that reduces to zero a SN barrier of zero-field height ϕ (i.e. the field that pulls the top of this barrier down to the Fermi level). This formula would now be written

$$F_R = c_S^{-2}\phi^2 = (4\pi\varepsilon_0/e^3)\phi^2 \cong (0.6944615\ \text{V/nm}) \cdot (\phi/\text{eV})^2. \tag{9.12}$$

For a material with $\phi = 4.50$ eV, $F_R \approx 14.1$ V/nm. Schottky [55] unsuccessfully proposed this kind of strong barrier lowering as an explanation of FE.

9.3.5 Scaled Field as a Modelling Parameter

In modern FE theory, this field F_R can be used to formally define a modelling parameter f that I now prefer to call the *scaled field* (for an SN barrier of zero-field height ϕ), by

[2]Reference [55] contains a typographical error, in that the numerical value in his equation (5) should be 6.97×10^6, as given correctly in Table 9.1 later in the paper. The modern value is slightly lower, as can be shown from (9.12).

$$f \equiv F/F_{\mathrm{R}} \equiv c_{\mathrm{S}}^2 \phi^{-2} F \cong (1.439965\ \mathrm{eV}^2\ \mathrm{V}^{-1}\ \mathrm{nm}) \cdot \phi^{-2} F. \tag{9.13}$$

For a barrier of zero-field height H, a parameter f_H is defined by a formula similar to (9.13), but with ϕ replaced by H. As seen later, f_H and f play important roles in modern FE theory. These parameters were previously called 'scaled barrier fields', but the simpler name 'scaled field' now seems better. (In the literature, f is also called 'dimensionless field'.)

Until recently, FE theory derivations used a modelling parameter y, called the *Nordheim parameter* and related to f_H by $y = +\sqrt{f_H}$. These modelling parameters enter FE theory in the following way. To evaluate the transmission probability for a SN barrier, using semi-classical methodology, the roots of the equation $M^{\mathrm{SN}}(z) = 0$ are needed, where $M^{\mathrm{SN}}(z)$ is given by (9.10). For a barrier of zero-field height H, these are

$$z = (H/2eF)\left[1 \pm \sqrt{1 - (e^3 F/4\pi\varepsilon_0 H^2)}\right]. \tag{9.14}$$

To simplify the mathematics, a modelling parameter is introduced to represent the combination $(e^3F/4\pi\varepsilon_0 H^2)$. In 1928, Nordheim [56] used a squared quantity, the modern y^2. (The original paper used the symbol x^2.) We now know that this combination is an expression for the scaled field f_H, and that—for reasons described near the end of Sect. 9.6.3—f_H would have been a better choice. In the author's view, there would be significant advantages in now making the difficult transition from the use of y to use of f_H, or (when $H = \phi$) to use of f—certainly when discussing current-voltage characteristics.

9.3.6 Barrier Strength and the Barrier Form Correction Factor

An important parameter derived from $M(z)$ is the so-called *barrier strength* (also called the 'Gamow factor' and the 'JWKB factor'). Jensen denotes this by 'θ', but the present author prefers 'G'. *For tunnelling*, G is defined by

$$G \equiv 2\kappa_{\mathrm{e}} \int M^{1/2}(z)\mathrm{d}z, \tag{9.15}$$

where κ_{e} is a universal constant (see Appendix 9.1), and the integral is taken 'across the barrier', i.e. over the range of z where $M(z) \geq 0$. The barrier strength G is a mathematically evaluated parameter, with a well defined numerical value, that relates to barrier transmission properties. Strong barriers (high-positive-G barriers) are difficult for an electron to tunnel through.

For an exactly triangular (ET) barrier of zero-field height H ($H \geq 0$), it is readily shown that

$$G^{\mathrm{ET}} = bH^{3/2}/F, \tag{9.16}$$

where b [$\equiv 4\kappa_e/3e$] is a universal constant often known as the *second Fowler-Nordheim constant* (see Appendix 9.1). For a general barrier of arbitrary (but well-behaved[3]) mathematical form (and with $H \geq 0$), the related barrier strength G is derived via (9.15), and a *barrier form correction factor* $\nu_H(H,F)$ ('nu$_H$') is then defined via

$$G(H,F) \equiv \nu_H(H,F) \cdot G^{\mathrm{ET}}(H,F), \tag{9.17}$$

where the functional dependences on H and F are here shown explicitly. There may also be other functional dependencies, for example on the emitter apex radius.

For the SN barrier, the barrier form correction factor $\nu_H^{\mathrm{SN}}(H,F)$ is given by an appropriate particular value (corresponding to barrier height H and local barrier field F) of a special mathematical function denoted here by $\mathrm{v}(x)$. This function is now known [57] to be a very special solution of the Gauss Hypergeometric Differential Equation; the *Gauss variable* x is the independent variable in this equation. It can be shown (see below) that the correction factor $\nu_H^{\mathrm{SN}}(H,F)$ is obtained by setting $x = f_H$, i.e. $\nu_H^{\mathrm{SN}}(H,F) = \mathrm{v}(x = f_H)$. Hence

$$G^{\mathrm{SN}}(H,F) \equiv \mathrm{v}(f_H) \cdot bH^{3/2}/F. \tag{9.18}$$

Note that the author now prefers to typeset 'v' upright, on the grounds that this symbol represents a special mathematical function (like 'sin' or 'Ai'). In Russian-language literature, the notation $\vartheta(y)$ is often used, rather than $\mathrm{v}(f_H)$, but the factor has the same numerical value. The mathematics of $\mathrm{v}(x)$ is discussed further in Sect. 9.6.3.

9.3.7 'Thermal Electron' Versus 'Thermionic'

When discussing the theory of high-temperature (low-field) effects, the author prefers the name 'thermal electron emission' rather than 'thermionic emission', for the following reason. Immediately before the well known work [34] of Fowler and Nordheim (FN) in 1928, it was thought by some scientists (in particular, Millikan [58]) that electron emission at high temperatures was the emission of a special type of electron called a 'thermion' that, inside the metal, was in a special high-energy state. This process was called 'thermionic emission', to distinguish it from the field-induced emission of conduction electrons at low temperatures, as occurred in what was then called 'autoelectronic emission'. This was a kind of 'two-band' theory of metal electron behaviour. Possibly the most important scientific outcome of the FN 1928 paper was the conclusion that 'thermions' in this sense did not exist, and that both thermally induced and field-induced emission phenomena could be understood as the emission

[3] A well-behaved barrier has no regions inside it where $M(z)$ is negative.

of 'one kind of electron' from a 'single band', under different conditions of field and temperature. (Hence, an integrated 'single-band' theory covering both effects could be created). Thus, the FN 1928 paper was one of the key papers that led to the development of the modern single-band theory of metals. Since FN argued successfully that thermions did not exist, it seems scientifically anomalous to continue to talk about 'thermionic emission' in FE theory papers. Whilst recognising that many still use this term, particularly in practical applications, I prefer the name 'thermal electron emission' (TE) in theoretical contexts.

Also, phenomena arising from the simultaneous application of field and temperature are more commonly called 'thermal-field' than 'thermionic-field'. This seems another good reason for preferring 'thermal' over 'thermionic'.

9.3.8 The Concept of 'Wave-Mechanical Flyover'

'Wave-mechanics' is the old name for 'quantum mechanics'. I prefer it when no quantisation or inelastic phenomena affect barrier transmission, because the name is a reminder of analogies between electron tunnelling and (what we now call) photon tunnelling. Both are consequences of the classical mathematical behaviour of second-order differential equations, where what is now called 'tunnelling' was once called 'evanescent-wave coupling'.

It is well known that the transmission of electrons over barriers is affected by wave-mechanical reflection effects, if the transmission is not too high above the barrier. Some years ago, Forbes and Deane [35] suggested that 'ballistic emission' was not a good name, and that this wave-mechanical transmission process should be called *flyover*. This term will be used in this chapter.

For reflection at a rectangular step of height-in-energy χ, the transmission probability $D(w)$ for an electron approaching with forwards energy $\chi + w$ ($w > 0$) is [35, 59]

$$D(w) = 1 - \left[\frac{(\chi + w)^{1/2} - w^{1/2}}{(\chi + w)^{1/2} + w^{1/2}}\right]^2. \tag{9.19}$$

If χ is taken as 10 eV, then $D(w) = 0.9$ when $w \approx 3.7$ eV, which corresponds to a temperature $T = w/k_B$ (where k_B is the Boltzmann constant) around 40000 K. This illustrates that all real technological thermal electron emitters are subject to wave-mechanical reflection effects, as assumed by Fowler long ago [60]. For this reason, the emission regime in which the devices discussed in the next section operate is called here *quantum-mechanical thermal electron emission* (QMTE).

9.3.9 The History of Thermal Electron Emission Theory

In this chapter, thermal electron emission (TE) is defined as the thermally induced emission of electrons in circumstances where the vast majority of electrons escape over the top of a barrier that may be lowered by the Schottky effect, and where very few escape by tunnelling. This definition is used by most thermionic-emission experimentalists. However, as discussed below, Murphy and Good [61] (and theoreticians who follow their approach) use a different definition of the term 'thermionic'.

In the QMTE regime, the local emission current density (ECD) J_{RS} is given by

$$J_{\mathrm{RS}} = \lambda_{\mathrm{TE}} A_{\mathrm{R0}} T^2 \exp[-\{\phi - c_{\mathrm{S}} F^{1/2}\}/k_{\mathrm{B}} T], \tag{9.20}$$

where T is the local thermodynamic temperature of the emitter surface, ϕ is local work function (NOT the Schottky-reduced barrier height), A_{R0} is a universal constant [60] called here the *universal theoretical Richardson constant* (see Appendix 9.1), and λ_{TE} is a general-purpose pre-exponential correction factor due mainly (but not exclusively) to wave-mechanical elastic reflection effects.

The author prefers to call this equation the *quantum-mechanical Richardson-Schottky (RS) equation*, on the grounds that, although others contributed, and although neither of these authors, in fact, devised the final forms of either the constant A_{R0} or the equation itself, these two played the biggest roles in generating the equation. The 'Richardson-Schottky' name was used by Murphy and Good [61]. One often sees a version of this equation in which the factor λ_{TE} has been set equal to unity and omitted. I call this reduced equation the *classical Richardson-Schottky equation.*

Modern TE discussions are sometimes superficial about the origins of (9.20). Thus, this chapter seems a good place to summarise its slightly complicated history. Development took place in stages. I deal first with the zero-field equation.

Investigation of what are now known to be TE effects goes back well into the eighteenth century, e.g. [62, 63] as noted in the books of Guthrie [64], Richardson [65] and Reimann [66], although Edison [67] usually gets the credit for the first really careful relevant experiments. However, little real progress could be made until after Thomson [32] had identified what (following Stoney [68]) we now call the 'electron', and until after vacuum technology had advanced sufficiently. Later experimental developments are described in many books and review articles. References [39, 47, 65, 66, 69–71] cover the period up to about 1950 in some detail, and [72, 73] are modern overviews.

Richardson always took the view that the electrons originated from inside the metal emitters and in 1901 published a formula for emission current density (here J) based on applying Maxwell-Boltzmann statistics to what he modelled as a classical electron gas inside the metal [74]. Nowadays we would write his 1901 formula as

$$J = A_1 T^{1/2} \exp[-\chi/k_{\mathrm{B}} T] \tag{9.21}$$

where A_1 is a universal constant (not equal to A_{R0}). Note this equation contains the local inner PE χ, not a modern work-function-like parameter.

However, this basic idea that the electrons came from inside the metal (rather than from gas molecules in the 'vacuum' space or adsorbed on the metal surface) encountered significant resistance. This inhibited acceptance of Richardson's ideas (and also those of Einstein about the origin of photoelectrons). It was not until 1913 that Richardson was able to prove, experimentally and convincingly, that his thermally emitted electrons must be coming from inside the metal [75], and that they must be supplied by electrical conduction inside the metal.

In the meantime, because of various inconsistencies that arose, in that era, in the physics of electron behaviour in metals (in particular, about the specific heat of electrons), Richardson had become unhappy with treating electrons inside the metal as a classical gas. Instead, he applied a thermodynamic argument [65, 76] to an electron atmosphere outside the emitter that was taken to be in equilibrium with the electrons inside, arguing that the external electron atmosphere better resembled a classical gas. This yielded [65]

$$J = A_2 T^2 \exp[-\phi_0/k_B T] \tag{9.22}$$

where A_2 is a universal constant very much less in value than A_{R0} (in fact less than $A_{R0}/20$), and ϕ_0 is, in effect, the 'typical activation energy needed for escape, at low temperatures' (my term).

In 1923, by a broadly similar thermodynamic argument, but based in detail on Nernst's heat theorem, Dushman [77] revised the pre-exponential in (9.22) to be equal to $A_{R0}/2$ (but also gave an alternative thermodynamic argument that made the pre-exponential about 10% less than this).

In reality, Richardson's work does not make the clear distinction that we can now make (post Sommerfeld) between inner potential energy χ and zero-temperature work function ϕ_{ZT}; nor does he make the clear modern distinction between gas kinetic theory and thermodynamics. With the benefit of hindsight, and guided by the clearly stated equations on p. 285 of [42], I have: (a) interpreted Richardson's parameter that he defines [74] as 'the work done by a corpuscle[4] in passing through the surface layer' as the inner PE; and (b) interpreted his parameter that he defines [65] as 'the change in energy of the system which accompanies the transference of each electron from the hot body to the surrounding enclosure' as a work-function-like parameter. However, even in the 1921 edition of his book, Richardson seems to regard the definitions as equivalent, because he uses the same exponent in both his '$T^{1/2}$' and 'T^2' equations. In contrast, Dushman [77] thinks of his parameter b_0 (which is effectively the same as Richardson's ϕ_0) as a 'thermodynamic heat of vaporisation', for the evaporation of a single electron. This is clearly a work-function-like parameter.

Next, in 1928, Fowler [60], and almost simultaneously Sommerfeld himself [78], made use of Sommerfeld's new electron theory of metals [40], based on Fermi-Dirac statistics, and of the idea of 'electron spin' (e.g. [79]), to make critical improvements.

[4] 'Corpuscle' was J.J. Thomson's original term for what we now call the electron.

Fowler suggested that (in the framework of a Richardson-Dushman-type thermodynamic argument, and in the ideal case of no wave-mechanical reflection effects) the pre-exponential should be $A_{R0.}$ He also clearly identifies ϕ_0 in (9.22) with the modern zero-temperature work function ϕ_{ZT}.

Fowler [60] also reinserted into the pre-exponential a factor (1–r), originally introduced empirically by Richardson [65], where r is a coefficient that allows for electron reflection at the surface. Fowler [60] interpreted this as a wave-mechanical effect. Then, in 1929 [80], he replaced (1–r) by the *mean transmission probability* $\bar{D}$. The best modern discussion of reflection effects is probably still the 1984 book by Modinos [2]. Later, it became clear that other effects, such as inelastic scattering and non-thermodynamic-equilibrium conditions, could in principle influence the pre-exponential; thus, I now prefer to use the general-purpose correction factor λ_{TE}.

However, note that some accounts of TE theory prefer to combine ($\lambda_{TE}A_{R0}$) into a single parameter A_{eff}, called the 'Richardson constant', which is not a universal constant. Notation and terminology in this area are somewhat variable, and potentially confusing to non-experts. Also, note that in experiments there are effects that can cause real or apparent temperature dependence in the exponent. These effects need to be taken into account when interpreting experimental data, particularly Schottky plots (e.g. see [39, 71]).

In summary, although the zero-field version of the classical RS-type equation is variously called the 'Richardson equation', the 'Richardson-Dushman equation', and the 'Richardson-Laue-Dushman' equation, the post-1920s form of it is really due to Sommerfeld and to Fowler.

As shown above, the Schottky effect reduces the zero-field barrier height by Δ_S. Hence, ϕ in the zero-field equation is replaced by (ϕ–Δ_S), and the Richardson-Schottky-type equations emerge. The earliest explicit formulation of (a Gaussian version of) (9.20) that I know of is in Compton and Langmuir's 1930 review ([47], see p. 149)—although, as already noted, J.J. Thomson seems to have been the first (in 1903) to suggest this as a physical electron emission effect.

A further logical step occurred when Fowler and Nordheim [34] in effect suggested that the best way of deriving TE theory would be to apply Sommerfeld-type metal electron theory to the electron population *inside* the emitter, and thereby create a unified theoretical approach that would describe both TE and FE. This approach was implemented by Sommerfeld and Bethe in 1933 [41] (but with incorrect tunnelling theory), and is discussed further below.

Later, Jensen and others showed that small temperature and field dependent changes in the inner PE χ mean that, in principle, emission equations should contain a temperature and field dependent 'effective work function' rather than the true zero-temperature, zero-field work function (see [81]), but this small effect is usually neglected.

A possible mild cause of confusion is that the term 'effective work function' is also used to describe the Schottky-effect-reduced barrier of height (above the Fermi level) (ϕ–Δ_S), particularly in the TE and photo-cathode communities. A possible solution

might involve the introduction of a new thermodynamic term, such as 'classical (or "canonical") work of emission', for this reduced barrier height.

9.4 Validity Regimes

9.4.1 Transmission Regimes and Emission Current Density Regimes

In a smooth-surface model, the FN approach is implemented as follows. Each travelling-electron wave-state in the well can be labelled with a set {Q} of quantum numbers. Each state that involves motion towards a chosen part of the emitter surface makes a contribution z_Q to the current density incident onto that part of the surface from the inside, with the size of the contribution depending (inter alia) on the state's occupation probability. The probability that an incident electron escapes is given by the transmission probability D_Q for the state. Hence, the state makes a contribution $z_Q D_Q$ to the emission current density (ECD) J, and the ECD is found by summing over all relevant states, thus

$$J = \sum\nolimits_{\text{relevant}\{Q\}} z_Q D_Q. \tag{9.23}$$

As discussed below, it is easiest (in a Sommerfeld free-electron model) to carry out this summation by a double integral over energy.

In practice, as shown below, the transmission probability is a function only of the local barrier field (here denoted by F) and of forwards energy, and can be written $D(F,W)$, where W (as above) denotes forwards energy *measured relative to the well base*. Hence, (9.23) can be converted to the single integration

$$J(F,T) = \int j_n(F,T,W)\,\mathrm{d}W = \int N(T,W)\cdot D(F,W)\,\mathrm{d}W, \tag{9.24}$$

where $j_n(T,W)$ is the *emitted normal energy distribution* (*e-NED*), and $N(T,W)$ is a recently introduced parameter called the *incident normal energy distribution (i-NED)*. If the electron distribution can be treated as if in local thermodynamic equilibrium at thermodynamic temperature T, which is usually done, then a mathematically explicit expression is easily obtained (below) for $N(T,W)$.

There are other ways in which the double integration could be carried out, but formula (9.24) is most convenient here. Note that the parameter N used here differs slightly from the parameter 'N' used in older work and denoted here by N_{old}, which was called the 'supply function'. The relationship is: $N = eN_{\text{old}}$. Thus, the new N, like j_n, has the units A m^{-2} eV^{-1}.

In the thermodynamic equilibrium situation, if valid values could be obtained for D across the whole range of W, then integral (9.24) could always be carried out

numerically. However, ideally, one wants analytical formulae. Unfortunately, exact analytical formulae can be obtained only under certain mathematical conditions, which may vary with the barrier form, and are certainly different for the ET and SN barriers. This gives rise to the idea of *validity regimes*. A *transmission regime* is a region of $\{F,W\}$ parameter space where a given formula for $D(F,W)$ is 'adequately valid', and an *emission current density (ECD) regime* (also called an *emission regime*) is a region of $\{F,T\}$ parameter space where (for given ϕ) a given formula for the local ECD J is 'adequately valid'. There are close links between transmission regimes and ECD regimes, but these are not straightforwardly one-to-one in all cases.

Because of the general difficulties with smooth-surface conceptual models, 'adequately valid' above really means 'adequately self-consistent' or 'is an adequate mathematical approximation'. However, a general weakness of electron emission theory is that even with these reduced meanings of 'adequately valid', the criteria for what one means by the term 'adequately', and the boundaries of the validity regimes are not yet as well investigated and defined as they probably ought to be.

9.4.2 Regimes for the Exactly Triangular Barrier

Currently, the best defined illustration of validity regimes involves the transmission regimes for the ET barrier. Essentially, this is because Forbes and Deane [35] managed to derive an exact general analytical formula for the transmission probability D^{ET} for the ET barrier, and to show that this had three very different forms of mathematical approximation in different parts of the *transmission regime diagram* shown in Fig. 9.2. The nature of the mathematics means that the most sensible choice is to plot w against $F^{2/3}$, where w is forwards energy measured relative to the top of the triangular barrier (which in this case coincides with the local vacuum level). In this special case, it is easy to draw the boundaries where the approximate solution departs from the exact analytical solution by 10%. For this diagram, the inner PE χ has been taken as 15 eV.

The exact general formula for D^{ET} is

$$D^{\mathrm{ET}} = 1/[\tfrac{1}{2} + \tfrac{1}{4}\pi\omega(A^2 + B^2) + \tfrac{1}{4}\pi\omega^{-1}(A'^2 + B'^2)] \tag{9.25a}$$

where A and B are the values of the Airy functions Ai and Bi, and A' and B' the values of their derivatives, at the PE step at $z = 0$ (see [32] for exact definitions), and ω is a dimensionless parameter given by

$$\omega \equiv c_\kappa F^{-1/3} W^{1/2} \tag{9.25b}$$

where c_κ is a universal constant (see Appendix 9.1), called here the *triangular barrier constant*. Earlier, Jensen ([7], and references therein) had reached an equivalent result.

Expression (9.25) has approximations that, for $w \ll 0$, $w \sim 0$, $w \gg 0$, are very different in their mathematical forms. This leads to three main transmission regimes,

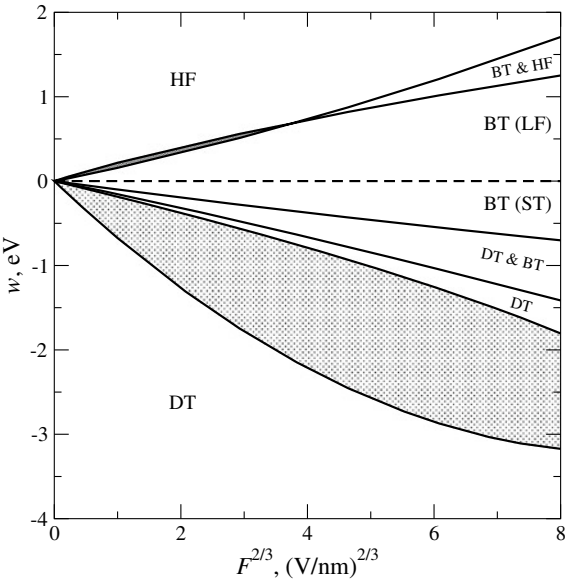

Fig. 9.2 Transmission regime diagram for the exactly triangular barrier ($\chi = 15$ eV, 10% error boundaries). Acronym meanings are: BT = barrier top; DT = deep tunnelling; HF = high flyover; LF = low flyover; ST = shallow tunnelling. The boundaries indicate a difference of 10% between the working formula for transmission probability D and the exact formula. In the two shaded regions, no simple working formula gets within 10% of the exact formula. Reprinted from [35, Fig. 5], with the permission of the Royal Society of London

each characterised by a useful working formula for D^{ET}. The main regimes are *deep tunnelling (DT)* ($w \ll 0$); *barrier-top emission (BT)* ($w \sim 0$); and *high flyover (HF)* ($w \gg 0$). For the $w \ll 0$ and $w \gg 0$ limits, the Airy functions and (9.25) have asymptotic expansions that each generate a sequence of approximate formulae of successively decreasing accuracy. The formulae used to define the transmission regimes are not the limiting formulae; rather they are the 'most useful working formulae'.

For the ET barrier, in the *deep tunnelling (DT) regime*, one can set the zero-field barrier height $H = -w$, and the working formula is the well known FN result

$$D_{\mathrm{DT}}^{\mathrm{ET}} \approx P^{\mathrm{FN}} \exp[-G^{\mathrm{ET}}] = \{4(HW)^{1/2}/\chi\} \exp[-bH^{3/2}/F], \tag{9.26}$$

where b is the second FN constant, as before, and the *FN tunnelling pre-factor* P^{FN} is given by $\{4(HW)^{1/2}/\chi\}$.

The *barrier-top (BT) regime* corresponds to $|w|$ small. This regime includes both shallow tunnelling (ST) and low flyover (LF), and a good working formula is [35]

$$D_{\mathrm{BT}}^{\mathrm{ET}} \approx 1/[\tfrac{1}{2} + c_0 F^{-1/3} W^{1/2} + c_\infty F^{+1/3} W^{-1/2} - c_1 F^{-1} W^{-1/2} w], \tag{9.27}$$

where c_0, c_∞ and c_1 are parameters, defined in [35], whose values are not of significance here. For *transmission at the barrier peak*, the result obtained by setting $w = 0$, $W = \chi$, in (9.27) is exact, and is

$$D_{\mathrm{BT}}^{\mathrm{ET}}(w=0) \;=\; 1/[\tfrac{1}{2} + c_0 F^{-1/3}\chi^{1/2} + c_\infty F^{+1/3}\chi^{-1/2}]. \tag{9.28}$$

Clearly, this barrier-peak transmission probability is a function of field, but goes to zero as the field goes to zero, as one would expect. It also goes to zero as $F \to \infty$, which was not anticipated.

In the *high flyover (HF) regime*, a suitable working formula is

$$D_{\mathrm{HF}}^{\mathrm{ET}} \;\approx\; \frac{4w^{1/2}W^{1/2}}{(W^{1/2}+w^{1/2})^2} \;=\; 1 - \left(\frac{W^{1/2}-w^{1/2}}{W^{1/2}+w^{1/2}}\right)^2. \tag{9.29}$$

This is also the standard result for transmission across a rectangular step of height $\chi = (W–w)$, as comparison with (9.19) shows. In the limit of *very* high flyover, where w and W become large, formula (9.29) tends to unity, as physically expected, and the regime limit is what one might reasonably call the *classical high-temperature emission (CHTE) sub-regime.*

Figure 9.2 shows that there is an area of the transmission regime diagram (enclosed within the DT regime) where none of the simple working formulae are adequately valid. Obviously, the details of this diagram depend somewhat on the value used for χ and on the numerical criterion used to define 'adequate validity' (10% error, here).

Older literature tends to discuss transmission in terms of two regimes: either tunnelling through the barrier or ballistic emission over it. What the above analysis shows is that, as far as the mathematics of the ET barrier is concerned, it is better to think in terms of *three* regimes rather than two (or possibly in terms of four rather than two regimes, if the CHTE sub-regime is treated as separate for pedagogical purposes). An important lesson is that, in the barrier-top regime, common theory applies both to (shallow) tunnelling and (low) flyover, but the sign of w is different in these two cases. In this BT regime, the value of the transmission probability increases smoothly as w changes from negative to positive, as one might expect physically. Another lesson from this diagram is that the BT regime is relatively narrow in its regime of validity, and that the width-in-energy of the regime diminishes with field.

Since (9.25) is the transmission-probability formula derived by FN, and then used by them to derive the original (zero-temperature) FE ECD equation, one can conclude that this ECD equation will not be adequately valid outside the boundaries of the DT regime. However, because the derivation of their ECD equation involves a Taylor expansion of $\ln\{\mathrm{D}_{\mathrm{DT}}^{\mathrm{ET}}\}$ about $H = \phi$, and then the neglect of terms higher than linear, there will be *additional* mathematical constraints on the adequacy of the resulting equation.

Because it is well known (or, at least, very firmly believed by nearly all FE theoreticians) that the ET barrier is not a good model for real surface barriers [82], finite-temperature versions of the original FN-type ECD equation have never been

formally developed, as far as the author is aware. (If they have, then knowledge has not survived.) However, it is known [83] that the temperature correction factor λ_T [$= p\pi/\sin(p\pi)$] developed below for the SN barrier, in fact, applies to all well-behaved barrier forms, and hence to the ET barrier (provided the parameter δ_F that appears in p is defined appropriately). Hence, if one were to attempt to construct an ECD regime diagram for a finite-temperature FN-type equation based on the ET barrier, then any mathematical constraints on the adequate validity of the expression $[(p\pi)/\sin(p\pi)]$ would also come into play when determining the boundaries of the relevant ECD regime, *in addition to* those indicated earlier (if the constraints are in fact different from those indicated earlier).

We can conclude that, in the framework of the 'SPME methodology' of using a planar smooth-surface conceptual model, the theoretical physical and mathematical questions that affect the 'adequate validity' of a particular ECD formula are as follows:

(1) Does a particular 'starting formula' for transmission probability D represent correct wave-mechanical transmission theory, as understood in a wider physics context? (This is not a significant problem for the ET barrier theory set out above, as we believe that the derivation of (9.25) represents correct wave-matching physics for a one-dimensional free-electron-metal model. However, for non-free-electron band-structures, and/or for other forms of the barrier, the situation is much more obscure.)
(2) Where relevant, has the starting formula for D been validly reduced to a simpler expression?
(3) If this 'simpler expression' needs to be further mathematically manipulated (e.g. by Taylor expansion), and some resulting terms are then neglected in order to provide a 'resulting expression', then is this mathematical approximation process adequately valid?
(4) If this 'resulting expression' then needs to be incorporated into an integral for which an analytical solution is known, then are the conditions for validity of the integration process adequately met? (In practice, questions (3) and (4) may relate to the same issue.)

A conclusion from this is that the 'failure analysis' of FE equations (and other electron emission equations) may be far from straightforward, and may deserve more attention than it has so far received.

Also, note that there are subtle differences in status between the Forbes-Deane transmission-regime diagram and the Murphy-Good ECD-regime diagram shown below as Fig. 9.3. The former shows regimes where the related formula is adequately valid as a description of the known correct physical analysis of the ET barrier model (to within 10%, at all points on each boundary). The latter simply shows regimes where certain mathematical approximations are considered adequately valid (with an error of between 15 and 40%, depending on the position in the boundary [61]).

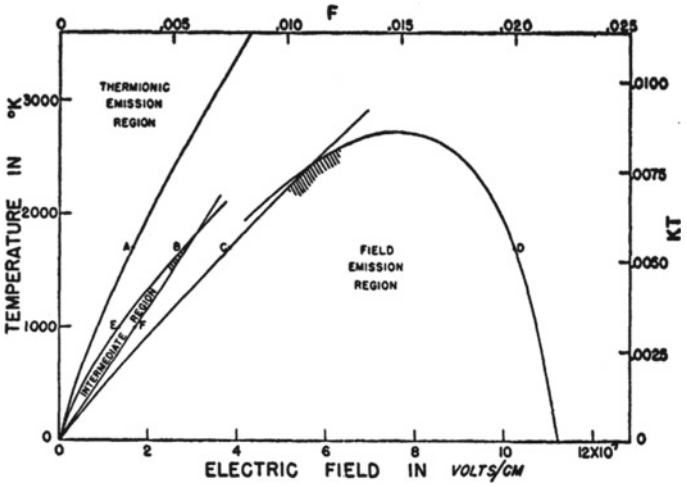

Fig. 9.3 Emission current density regime diagram for the Schottky-Nordheim barrier, as presented by Murphy and Good ([61Fig. 6]). See [61] for full discussion. Reprinted figure with permission from E.L. Murphy, R.H. Good, Jr, Phys. Rev. **102**, 1464 (1956). Copyright (1956) by the American Physical Society

9.4.3 Regimes for the Schottky-Nordheim Barrier—Qualitative Discussion

As just indicated, most FE theoreticians regard the Schottky-Nordheim barrier as the best simple model for the FE barrier at the surface of a metal emitter of moderate to large apex radius (say greater than 20 nm), for reasons recently set out in [82]. One can apply to the SN barrier the preceding Section's insights, but initially, this is done only at a qualitative level. Mathematical details form part of Sect. 9.5.

Confounding problems are the existence of two alternative top-level approaches and conflicting terminology. In this chapter I use 'thermal-field (TF)' as a general term to describe situations where both field and temperature play a role, but not as a term to describe any particular theoretical regime or sub-regime.

Although earlier discussions exist (e.g. [84]), the first 'reasonably modern' treatment of TF effects is the 1956 paper of Murphy and Good (MG) [61]. The MG treatment is the first substantial treatment after Burgess, Kroemer and Houston [85] had found, in 1953, a significant mathematical error in Nordheim's 1928 paper [56]. Unfortunately, the MG paper is heavily mathematical and written in a mixture of Gaussian equations and atomic units, and is now next to incomprehensible to many younger researchers. The Modinos textbook [2] has a simpler version of the MG treatment, though still with Gaussian equations. ISQ versions of MG's theory are given in [86] for the zero-temperature limit, in [87] for finite temperatures, and (slightly differently) in [10] and in [14].

Murphy and Good identified two main ECD regimes, which they labelled the 'field emission' and 'thermionic' regimes. Note that their definition of 'thermionic' does *not* correspond to the operating regime of the practical devices that electron source users call 'thermionic emitters', but also includes the devices now commercially called 'Schottky emitters' (my 'barrier-top emitters'). Jensen's chapter, and also the recent work of Kyritsakis and colleagues [88, 89], follow the MG two-regime approach and create bridging theory.

MG also identified what they called an 'intermediate' regime, but this is of very limited extent, and nowadays is not treated separately. MG's ECD regime diagram, showing all of their identified regimes, is reproduced as Fig. 9.3.

In the 1970s, Swanson and Bell [90] presented a more detailed classification of regimes/sub-regimes, and associated formulae for ECDs and energy distributions, although later [15, 91] they changed their nomenclature slightly.

The author's current classification of SN-barrier ECD regimes is based partly on the Swanson and Bell approach, partly on the findings of [35]. Nowadays, I identify three main ECD regimes. Using my currently preferred terminology, these are as follows:

(1) The *field electron emission regime* (also called the 'cold field electron emission (CFE)', the 'Fowler-Nordheim field electron emission (FNFE)' regime, or just the 'field emission regime'). In this regime, most electrons escape by deep tunnelling from electron states near the emitter Fermi level, and the local ECD is given by a FE equation. In MG's theory (which uses the simple-JWKB tunnelling formalism) the relevant FE equation contains a temperature correction factor λ_T given here by the linked equations

$$\lambda_T \equiv p\pi/\sin(p\pi), \tag{9.30a}$$

$$p \equiv k_B T/\delta_F^{SN}, \tag{9.30b}$$

$$1/\delta_F^{SN} \equiv [(\partial G^{SN}/\partial H)_F]|_{H=\phi} = t_F \cdot \tfrac{3}{2} b\phi^{1/2}/F, \tag{9.30c}$$

where p is the '*Swanson-Bell parameter p*' [90], δ_F^{SN} is a parameter I now call the *barrier-strength decay-width (or 'G-decay-width') for the SN barrier, taken at the Fermi level*, and t_F—here called the *G-decay-rate correction factor*—is the value, at $H = \phi$, of the correction factor (relative to the equivalent result for an ET barrier) in the first-order term of a Taylor expansion of $G^{SN}(H,F)$ about $H = \phi$ (with field F assumed constant.) The factor is given by $t_F = t_1(x = f)$, where $t_1(x)$ is a special mathematical function defined by (9.65b).

For numerics, it is convenient to write p in the form

$$p = C_p \cdot (t_F \phi^{1/2} T/F) \equiv (3bk_B/2) \cdot (t_F \phi^{1/2} T/F) \tag{9.30d}$$

$$C_p \cong 8.829607 \times 10^{-4}\ \mathrm{eV}^{-1/2}\ \mathrm{K}^{-1}\ \mathrm{V\ nm}^{-1}. \tag{9.30e}$$

The ECD regime boundaries are in principle influenced by (amongst other factors) the adequate validity of the derivation of λ_T. The Swanson and Bell view [90] was that use of the MG expression for λ_T is adequate in the range $p <$ ~0.7.

The FE regime has a zero-temperature theoretical limit at which $p = 0$, $\lambda_T = 1$. The original 1928 FE equation was, of course, derived as a formula applying in this limit. The high-T end of the FE regime has been called the 'T-F emission regime' (e.g. [90]), but this name has also been used in other contexts and is best avoided; the author's preferred name would be 'hot FE'.

(2) The *barrier-top electron emission (BTE) regime* (also called, in the case of the SN barrier, the 'extended Schottky regime' [15, 91], and before that the 'transition regime' [90]). In this regime, most electrons escape at a forwards energy level that is close to the top of the SN barrier. Significant numbers escape by tunnelling and significant numbers by wave-mechanical flyover. The local ECD for the BTE regime is usually given by an equation similar to the QM RS-type (9.20), but with the pre-factor λ_{TE} replaced by a pre-factor λ_{BTE} derived by using the Kemble tunnelling formalism (see Sect. 9.5.2). The pre-factor λ_{BTE} is given by the linked equations

$$\lambda_{\mathrm{BTE}} \approx \lambda_F \equiv q\pi/\sin(q\pi), \tag{9.31a}$$

$$q \equiv \delta_{\mathrm{b}}^{\mathrm{SN}}/k_{\mathrm{B}}T, \tag{9.31b}$$

$$1/\delta_{\mathrm{b}}^{\mathrm{SN}} \equiv [(\partial G^{\mathrm{SN}}/\partial H)_F]|_{H=H_{\mathrm{b}}} = \mathrm{t_b} \cdot \tfrac{3}{2} b H_{\mathrm{b}}^{1/2}/F = \mathrm{t_b} \cdot \tfrac{3}{2} c_{\mathrm{S}}^{1/2} F^{-3/4}, \tag{9.31c}$$

Here, q is the '*Swanson-Bell parameter q*' [90], H_{b} [$= \Delta_{\mathrm{S}} = c_{\mathrm{S}} F^{1/2}$] is the zero-field height of the transmission barrier experienced by an electron that crosses at the forwards energy level of the top of the barrier, and $\delta_{\mathrm{b}}^{\mathrm{SN}}$ is the G-decay-width for the SN barrier, taken for $H = H_{\mathrm{b}}$. The parameter $\mathrm{t_b}$ is the G-decay-rate correction factor for $H = H_{\mathrm{b}}$, and is given (from the theory of complete elliptic integrals, see below) as $\mathrm{t_b} = \mathrm{t_1}(x = 1) = (\pi\sqrt{2})/4 \cong 1.110721$.

For numerics, it is convenient to write q in the form

$$q = C_q F^{3/4}/T = [2/(3\mathrm{t_b} b c_{\mathrm{S}}^{1/2} k_{\mathrm{B}})]\, T^{-1} F^{3/4}, \tag{9.31d}$$

$$C_q \cong 930.8282\ \mathrm{K\ V}^{-3/4}\ \mathrm{nm}^{3/4}. \tag{9.31e}$$

This correction factor formula was first given by MG as part of their equation (9.33) (they use the symbol d rather than q), but the method of derivation indicated here is simpler, and open to generalisation.

Schottky electron sources, that operate in this BTE regime, are discussed in [11] and in Chap. 6, and in references therein.

On the high-q side, the original thinking of Swanson and Bell [90] was that this formula would be adequately valid for $q < \sim 0.7$. However, simulations based on the numerical solution of the Schrödinger equation suggest a lower limit of adequate validity, maybe around $q < \sim 0.4$ or less [15, 91]. Precise reasons for this lower limit of validity have not yet been established.

On the low-q side, it is clear that as $q \rightarrow 0$, then $\lambda_F \rightarrow 1$, and the ECD formula then goes over into the *classical* RS-type equation, which describes the classical high-temperature emission (CHTE) *sub-regime*, but not the main QMTE regime (except as an approximation). More generally, the precise limits of validity of formula (9.31a–9.31e) have not yet been clearly established.

(3) The third main regime is the *quantum-mechanical thermal electron emission* (QMTE) regime, already discussed. In this regime, almost all electrons escape by wave-mechanical flyover over the barrier, which may be lowered by the Schottky effect, and encounter wave-mechanical reflection effects in the whole surface + barrier potential structure. As already indicated, this has the CHTE sub-regime as the very-high-temperature limit. Both the main regime and the sub-regime in principle have 'pure thermal emission limits' as the field is reduced towards zero, but (as already indicated) related practical current densities may be limited by space-charge effects associated with patch fields.

Some general comments are now needed. Although the BTE and QMTE regimes have ECD equations of the same general mathematical form, the factors λ_{BTE} and λ_{TE} have different physical origins, and approximations for them have (or would have) different detailed mathematical forms. This arises because λ_{BTE} relates to wave-mechanical 'tunnelling and reflection' effects associated with the top part of the SN barrier, but λ_{TE} relates primarily to wave-mechanical reflection effects primarily associated with the sharp change in EPE gradient when the rising image PE (or, more generally, the surface EPE) takes over from the constant EPE inside the well. In reality, there may be interactions between these two effects, and these interactions might lead to a modern model for the anomalous Schottky effect [71, see p. 33]. (This might be seen as an improved version of Guth and Mullin's theory [84].) In the longer term, one might hope to have an integrated theory of λ_{TE} and λ_{BTE} that either separates into two or more regimes, or alternatively shows that the interactions between the two different sources of reflection effect must always be taken into account, certainly at low fields.

A different form of bridging theory is presented in Jensen's chapter, which is based on earlier work by himself and Cahay [92, 93], to which there have been some recent improvements [94]. Jensen has developed formulae that (in effect) have (9.30a–9.30e) and (9.31a–9.31e) as the low-temperature and high-temperature limits, and (for given ϕ-value) have adequate validity in the region of $\{F,T\}$ parameter space outside the regimes of adequate validity for the expressions for λ_T and λ_F. Such a formula is particularly useful in simulations when making an estimate of the total emission current from a device where the field varies strongly across the surface

and the temperature is high enough for the emission to be a mixture of field and thermally induced emission. As with formula (9.30a–9.30e), Jensen's formulae deal only approximately with the QMTE regime.

A slightly different approach to the development of bridging theory has been taken by Kyritsakis and colleagues, who have also extended the theory to cover charged, earthed spherical emitters—provided that the emitter apex radius is not so small that significant quantum confinement effects (e.g. [95, 96]) come in play. Readers are referred to the relevant Kyritsakis papers for details [88, 89].

A further point about regimes is that, at extremely high fields, where the top of the SN barrier is pulled down close to or below the emitter Fermi level, there are (from the point of view of emission theory) one or more additional ECD regimes. An expression $\frac{1}{2}z_S K_{F0}^2$, where z_S is the Sommerfeld electron supply constant (see below), and K_{F0} is the zero-temperature Fermi *energy*, exists for the limiting ECD in the hypothetical situation where all electrons approaching the surface escape. In reality, most emitters would heat up, melt and/or explode long before this limit is reached. These 'explosive emission' phenomena are covered briefly by Fursey [6] and in more detail by Mesyats [3], in their textbooks, and (for carbon) by Fursey in Chap. 11. They will not be discussed here. It further seems that sudden heating, even in the absence of applied voltages, can also cause explosive phenomena [97].

9.5 Emission Theory—Mathematical Background

9.5.1 Relevant Elements of Statistical Mechanics

Free-Electron Theory and the Sommerfeld Electron Supply Constant

To implement the summation in (9.23), or the integral in (9.24), basic understanding is needed of both free-electron statistical mechanics and wave-mechanical transmission theory. The next three sections cover this. Many treatments of free-electron theory still use the early 1930s starting point that the density of electron states in phase-space is constant, and proceed via a three-dimensional integration with respect to velocity, momentum or wave-number. In reality, there is an alternative starting point that is simpler, exactly equivalent physically, better for discussion of emission phenomena, and better for non-physicists. This is based directly on the use of energies, and is as follows.

Consider a mathematical plane in space inside a Sommerfeld well of large extent, and let the electron energies parallel and normal to the plane be E_p and E_n, with the total energy being given by $E_t = E_n + E_p$. In principle, E_n, E_t and the well base E_c can all be measured relative to *any* one energy reference level (e.g. the laboratory 'Ground'), but it is simplest to initially treat E_n and E_t as measured relative to the well base, and hence purely kinetic in nature. Any two of these three energies can be used to create a formal *energy-space* (call the axes '1' and '2'). The *supply density* $z(E_1,E_2)$ is defined as the current density crossing the mathematical plane in real

space, per unit area of energy-space, near (E_1,E_2). The *maximum supply density* (for the element) is achieved if all states in the relevant energy element are occupied with probability 1. There is a general statistical mechanical result that, for a free-electron model (of a metal), the maximum supply density is the same for all elements in all of the three alternative energy-spaces that can be created, and is equal to the *Sommerfeld electron supply constant* (or 'Sommerfeld supply density') z_S given by

$$z_S \equiv 4\pi e m_e/h_P^3 \cong 1.618311 \times 10^{14}\,\mathrm{A\,m^{-2}\,eV^{-2}} \tag{9.32a}$$

where h_P is Planck's constant and m_e is the electron mass in free space. The presence of this constant as part of the elementary RS-type equation [since $A_{R0} = z_S(k_B T)^2$] is particularly clear in [42]. However, some authors prefer to write it in the alternative form

$$z_S = e m_e/2\pi\hbar^3 \tag{9.32b}$$

Let $f_{FD}(E_t)$ be the Fermi-Dirac distribution function. It follows that, if states in the relevant energy elements are occupied in accordance with Fermi-Dirac statistics, then the contribution d^2Z to the total current density Z crossing the plane, made by states in the energy-space element concerned, can be written in any of the alternative forms

$$d^2Z = z_S f_{FD} dE_p dE_n = z_S f_{FD} dE_n dE_t = z_S f_{FD} dE_p dE_t. \tag{9.33}$$

This result is proved and discussed by the present author in [98], though I would not claim to have done anything other than provide an alternative proof and a more general statement of a result already known (e.g. §11.2 in [99]), and give a name to the universal constant involved. The point is that (for a free-electron Sommerfeld-type model) this is a proven statistical mechanical result that can be used as a starting point for further development. It is not necessary to repeatedly reprove some weaker specific version of it, starting from early 1930s formulations that many experimentalists may not fully understand—which is what tends to happen at present.

Derivation of the Incident Normal Energy Distribution

To apply (9.33) to an emission problem, one identifies the 'mathematical plane' with a planar or 'nearly planar' emitter surface, identifies d^2Z as an element of incident electron-current density approaching the surface from the inside, notes that all the states in the element will have the same transmission probability D (which will be a function only of barrier field and normal energy), and concludes that this element makes a contribution d^2J to the local *emission* current density J given by $d^2J = D(F,E_n)d^2Z$. The ECD J is then obtained by a double integral with respect to energy. In principle, there are six ways of carrying this out, and different methods are best for different purposes. For discussion of ECD regimes, the best method is to integrate first over E_p, then over E_n. This is sometimes called 'integrating via the NED', where NED stands for 'normal energy distribution'.

In the particular approach where E_n is measured *relative to the well base*, it is convenient to replace the symbol E_n with the symbol W, and to revert to using the term 'forwards energy'. (Jensen uses the alternative approach of replacing E_n by the symbol 'E'.) However, note that some older papers that use the symbol W for normal/forwards energy (in particular, MG) use the local vacuum level as their energy reference zero.

The double integral over E_p and W decomposes into the two single integrals

$$J(F,T) = \int_0^\infty j_n(F,T,W)\mathrm{d}W = \int_0^\infty N(T,W)\cdot D(F,W)\cdot \mathrm{d}W \tag{9.34}$$

$$N(T,W) = z_S \int_0^\infty f_{FD}(W,E_p)\,\mathrm{d}E_p \tag{9.35}$$

where N is the incident normal energy distribution, as before. Equation (9.34) here is (9.24) as stated earlier, and can be called the *forwards energy integral*. Substituting an explicit expression for the Fermi-Dirac function into (9.35) yields

$$\begin{aligned} N(W,T) &= z_S \int_0^\infty \left\{1+\exp[(W+E_p-W_F)/k_BT]\right\}^{-1} dE_p \\ &= -z_S k_B T\Big[\ln\{1+\exp[-(W+E_p-W_F)/k_BT]\Big]_0^\infty \\ &= z_S k_B T \ln\{1+\exp[-(W-W_F)/k_BT]\}, \end{aligned} \tag{9.36}$$

where k_B is the Boltzmann constant, and W_F denotes the Fermi level measured relative to the well base at total energy E_c, i.e. $W_F = E_F – E_c$ (see Fig. 9.1).

When $(W–W_F)$ is sufficiently greater than k_BT, then the final exponent in (9.36) becomes small, and the approximation $\ln(1+x) \approx x$ leads to the approximate expression for the i-NED

$$N(W,T) \approx N^{aprx}(W,T) = z_S k_B T \exp[-(W-W_F)/k_BT]. \tag{9.37}$$

9.5.2 *Elements of Barrier Transmission Theory*

The Status of Barrier Transmission Theory

Surprisingly, despite enormous efforts over many years, barrier transmission theory is still an underdeveloped branch of quantum mechanics, involving some very awkward physics and mathematics that is still far from being completely understood in full detail. The present situation (within the context of SPME methodology) is as follows:

For the ET barrier, an exact analytical solution (covering both tunnelling and flyover) is known for the free-electron *metal* case, as described above.

For the SN barrier, and for other field-influenced barriers relating to bulk emitters with *planar* surfaces, no exact analytical solutions are known in terms of the existing functions of mathematical physics, but approximate analytical solutions based on *semi-classical methods* (also called 'quasi-classical methods') have—until recently—been thought sufficient for nearly all technological purposes. These include 'JWKB-like' and 'phase-integral' methods, but only low-order forms of these methods are currently in use. Higher-order phase-integral formulas are given in [100]. For non-planar emitters, especially emitters of small apex radius, new apparent difficulties in developing semi-classical transmission theory have recently emerged [101], and are still subject to research.

Certainly for planar emitters, good numerical methods in principle exist for determining FE transmission probabilities, by direct numerical solution of the Schrödinger equation (e.g. [14, 102]) However, the most popular methodologies still rely on semi-classical approaches, as described next.

The Main Semi-Classical Formalisms

For the transmission probability D, the most general semi-classical formalism normally quoted in FE tunnelling contexts is the *Fröman and Fröman formalism* [103, 104]

$$D \approx D_{\mathrm{FF}} = \frac{P\exp[-G]}{1+P\exp[-G]} = \frac{1}{1+P^{-1}\exp[G]} \tag{9.38}$$

where P is a parameter, called the *transmission pre-factor*, that is thought to be of *order* unity in all normal circumstances. Though Mayer has made calculations that show this, and has provided a formula [102] that fits the result of numerical calculations, normal practice is still to use the *Kemble formalism* in which P is set equal to unity and hence D is given by

$$D \approx D_{\mathrm{K}} = \frac{\exp[-G]}{1+\exp[-G]} = \frac{1}{1+\exp[G]}. \tag{9.39}$$

This formula was first derived long ago by Kemble [105, 106], by a JWKB-type method. It was rederived by Miller and Good [107] in a slightly different way that some authors regard as technically not a JWKB-type method but an alternative semi-classical method (see [108]).

In the case of *deep tunnelling*, where the barrier strength G is sufficiently large ($G > 2.3$, for 10% error), (9.38) reduces to the so-called *Landau and Lifshitz formalism* [59, see (50.12)] :

$$D \approx D_{\mathrm{LL}} = P\exp[-G] \tag{9.40}$$

and (9.39) reduces to the *simple (or 'first-order') JWKB formalism*

$$D \approx D_{\mathrm{JWKB}} = \exp[-G] \tag{9.41}$$

This form can also be obtained by straightforward JWKB-type derivation methods, e.g. [59].

In all cases, for tunnelling through the SN barrier, the barrier strength G^{SN} is given by

$$G^{\mathrm{SN}} = \mathrm{v}(f_H) \cdot bH^{3/2}/F. \tag{9.42}$$

In the situation where the barrier has just vanished, $F = F_{\mathrm{R}}, f_H = 1$, $\mathrm{v}(f_H) = \mathrm{v}(1) = 0$, $G^{\mathrm{SN}} = 1$, and—in the Kemble formalism, from (9.39)—it is found that $D \approx 0.5$.

Barrier Strength for Transmission by Flyover above the Barrier

For the FE regime, simple (first-order) JWKB theory, as specified by (9.41), is sufficient. But when considering the other main regimes it is necessary to have formulae that give values of transmission probability D for wave-mechanical flyover over the top of the barrier.

In these circumstances, the motive energy is everywhere negative, but the equation $M^{\mathrm{SN}}(z) = 0$ does have mathematical solutions, although these are complex numbers, as illustrated in Fig. 9.4. There is a pole at the origin, and two complex roots of the form $(H/2eF)\{1 \pm i(f_H - 1)^{1/2}\}$, where $\mathrm{i} = +\sqrt{(-1)}$. Miller and Good [107], and Murphy and Good [61], assume that an appropriate procedure to evaluate G in these circumstances is to perform complex integration along the contour from the

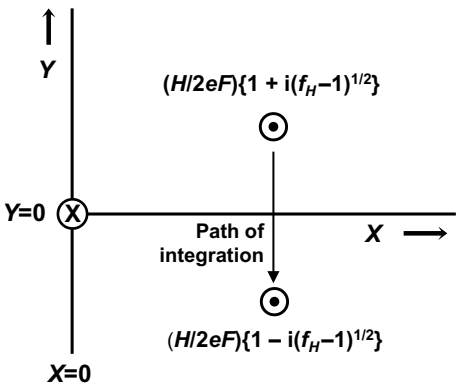

Fig. 9.4 To illustrate the procedure apparently used by Murphy and Good [61] to evaluate the barrier strength integral for a SN barrier in the case of flyover. In the complex-distance space associated with the z-direction, the expression $M^{\mathrm{SN}}(z) = 0$ has two complex zeros, and a pole at the origin. The path of integration in complex space is indicated

upper root to the lower root. The author has never seen a formal mathematical proof that this procedure is strictly valid, but it looks plausible enough (certainly as an approximation), at least for small values of $(f_H-1)^{1/2}$.

This integration again results in (9.42), but this now operates in the range $f_H > 1$. Since v(x) is a special solution of the Gauss Hypergeometric Differential Equation and is known to be well-behaved at $x = 1$, we can deduce that v(x) is well defined over the range $0 \leq x < \infty$; hence, *in principle,* it is straightforward *mathematically* to obtain values of v(x), and hence v(f_H), for the range $1 \leq f_H < \infty$. In this range the values of v(f_H) and G^{SN} are negative, and G^{SN} diminishes from 0 towards $-\infty$, causing D to increase from 0.5 towards 1.

In practice, due to the pole at the origin, this procedure for obtaining an expression for G^{SN} becomes invalid if the two complex roots are too far separated. MG take the validity condition to be $y \leq \sqrt{2}$, i.e. $f_H \leq 2$. This defines a forwards energy level $W^{\mathrm{lim}}(\phi,F)$ above which their theory for evaluating D becomes invalid, and they recommend the *physical* approximation $D = 1$ for higher forwards energies. As far as the author can establish, for the SN barrier, there are no valid, simple, semi-classical mathematical methods available for evaluating D-values for $W > W^{\mathrm{lim}}$.

In practice, MG then make mathematical approximations in order to create integrals that can be evaluated analytically, as discussed below (see [61] for details).

Evaluation of the Forwards Energy Integral

We now return to the evaluation of the forwards-energy integral (9.34). For each of the two components N and D in the integral, there is a 'full expression' and an 'approximated expression'. As shown in Table 9.2, in three of the resulting four cases an analytical solution exists. The combinations of taking one 'full' expression and one 'approximate expression' lead to expressions for λ_T and λ_F, respectively, as given by (9.30a–9.30e) and (9.31a–9.31e), both first derived (though not in these modern forms) by MG. Both these expressions have regimes of validity in $\{F,T\}$ parameter space, as partly shown in Fig. 9.3.

By contrast, the Jensen *general thermal-field (GTF) formalism*, described in detail in Chap. 8, is based on the integral that uses both 'full' expressions. The main aim is to derive a 'bridging formula' that will have adequate validity (for a given ϕ-value) in the region of $\{F,T\}$ parameter space outside the regimes of validity for the expressions for λ_T and λ_F.

Table 9.2 Results of evaluating the forwards-energy integral, (9.34), using different levels of approximation for the two terms in the integral

D-formula	N-formula	Analytical?	Result	Discussion
Approximate	Approximate	Yes	Not normally used	–
Approximate	Full	Yes	λ_T	Here (9.30a)
Full	Approximate	Yes	λ_F	Here (9.31a)
Full	Full	No	Jensen's GTF formula	See Chap. 8

9.6 Emission Theory

9.6.1 Field Electron Emission Equations

The next two Sections record further proposals for improving the presentation of electron emission theory. FE literature contains many different equations asserted to be '*the* Fowler-Nordheim equation'. In reality, this name applies to a very large family of related approximate equations, which I now prefer to call *field electron emission (FE) equations*. This Section first discusses how to classify them.

A *FE equation* (previously called a 'Fowler-Nordheim-type (FN-type) equation') is any FE-related equation with the mathematical form

$$Y = C_{YX} X^2 \exp[-B_X/X] \tag{9.43}$$

where X is any FE independent variable (usually a field, voltage or scaled field), and Y is any FE dependent variable (usually a current or current density). B_X and C_{YX} depend on the choices of X, Y and barrier form, and will often exhibit weak or moderate functional dependencies on the chosen variables and sometimes on others (e.g. apex radius).

The choices of X and Y determine the equation's *form*. *Core forms* give the local ECD J in terms of the local work function ϕ and local barrier field F and/or scaled field f. The simplest core form (which relates to the ET barrier) is the so-called *elementary FE equation*

$$J^{\text{el}} = a\phi^{-1}F^2 \exp[-b\phi^{3/2}/F], \tag{9.44}$$

where a and b are the FN constants (see Appendix 9.1).

For an arbitrary but well-behaved general barrier (GB), with barrier form correction factor ν^{GB}, (ν = 'nu'), one can define a *kernel current density* J_{k}^{GB} by

$$J_{\text{k}}^{\text{GB}} = a\phi^{-1}F^2 \exp[-\nu_{\text{F}}^{\text{GB}} b\phi^{3/2}/F] \tag{9.45}$$

where the subscript 'F' on $\nu_{\text{F}}^{\text{GB}}$ indicates that the correction factor applies to a barrier of zero-field height ϕ. A merit of defining this mathematical quantity J_{k}^{GB} separately is that, for given choices of barrier form and of ϕ and F, the value of J_{k}^{GB} can be calculated *exactly*; further, in the case of the SN barrier, reasonably accurate values of J_{k}^{SN} can be deduced from a FN plot if the emission situation is orthodox (see [20, 109]). The same is expected to be true for the newly introduced Murphy-Good plot [110].

The GB-model *local ECD* J^{GB} is then formally related to the corresponding kernel current density by

$$J^{\mathrm{GB}} = \lambda^{\mathrm{GB}} J_{\mathrm{k}}^{\mathrm{GB}}, \tag{9.46}$$

where the *pre-exponential correction factor* λ^{GB} *formally* takes into account all corrections other than those directly associated with the simple treatment of barrier form. These corrections include the use of atomic-level wave-functions and non-free-electron-like band-structure; better quantum mechanics of the transmission process; more accurate summation over states; temperature effects on the electron-energy distribution; field and temperature effects on the 'apparent zero-field work function'; and any factors currently unrecognised.

Equations (9.45) and (9.46) each contain a correction factor, and the *complexity level* of a FN-type-equation core form can be classified by specifying these two correction factors. For planar-surface emitter models, Table 9.3 shows the core forms (involving F) of most interest, historically and currently. These are classified, first, according to the assumed nature of the barrier, second (roughly) according to the sophistication of the assumptions made about the pre-exponential correction factor.

It needs emphasising that this is not a complete list, even for planar surface models. Further, in some ways, it is only an overview. For example, FE literature contains about 20 different mathematical approximations for the special mathematical function v, and there are various treatments of different individual components that go into the general-purpose pre-exponential correction factors.

Table 9.3 Names and complexity levels for field electron emission (FE) equations applying to emitter models that assume a smooth planar surface or (with "GB" models only) an atomically flat planar surface

Name of FE equation	Date	References	$\lambda_{\mathrm{C}}^{\mathrm{GB}}\rightarrow$	Barrier form	$\nu_{\mathrm{F}}^{\mathrm{GB}}\rightarrow$	Note
Elementary	1999	[111]	1	ET	1	a
Original	1928	[34]	P^{FN}	ET	1	
Fowler-1936	1936	[112]	4	ET	1	
Extended elementary	2015	[20]	$\lambda_{\mathrm{C}}^{\mathrm{ET}}$	ET	1	
Dyke-Dolan	1956	[113]	1	SN	$\mathrm{v_F}$	
Murphy-Good (zero-T)	1956	[61]	$\mathrm{t_F^{-2}}$	SN	$\mathrm{v_F}$	b
Murphy-Good (finite-T)	1956	[61]	$\lambda_T \cdot \mathrm{t_F^{-2}}$	SN	$\mathrm{v_F}$	
Orthodox	2013	[109]	$\lambda_{\mathrm{C}}^{\mathrm{SN0}}$	SN	$\mathrm{v_F}$	c
Extended Murphy-Good	2015	[20]	$\lambda_{\mathrm{C}}^{\mathrm{SN}}$	SN	$\mathrm{v_F}$	d
"Barrier-effects-only"	2013	[114]	$\lambda_{\mathrm{C}}^{\mathrm{GB0}}$	GB	$\nu_{\mathrm{F}}^{\mathrm{GB}}$	c
General	1999	[111]	$\lambda_{\mathrm{C}}^{\mathrm{GB}}$	GB	$\nu_{\mathrm{F}}^{\mathrm{GB}}$	

[a]Earlier imprecise versions exist, but the first reasonably clear statement seems to be in 1999
[b]Previously called the "Standard FN-type equation"
[c]The superscript "0" in column 4 indicates that the pre-factor is to be treated mathematically as constant
[d]Previously called the "New-standard FN-type equation"

9.6.2 Scaled Form for the SN-Barrier Kernel Current Density

For the SN barrier, the kernel current density is

$$J_{\mathrm{k}}^{\mathrm{SN}} = a\phi^{-1}F^2 \exp[-\mathrm{v}(f) \cdot b\phi^{3/2}/F], \tag{9.47}$$

where f is the scaled field for a SN barrier of zero-field height ϕ. On defining work-function-related parameters $\eta(\phi)$ and $\theta(\phi)$ by

$$\eta(\phi) \equiv b\phi^{3/2}/F_{\mathrm{R}} = bc_{\mathrm{S}}^2\phi^{-1/2} \cong 9.836238\ (\mathrm{eV}/\phi)^{1/2}, \tag{9.48}$$

$$\theta(\phi) \equiv a\phi^{-1}F_{\mathrm{R}}^2 = ac_{\mathrm{S}}^{-4}\phi^3 \cong (7.433979 \times 10^{11}\ \mathrm{A/m^2})(\phi/\mathrm{eV})^3\,, \tag{9.49}$$

the kernel current density $J_{\mathrm{k}}^{\mathrm{SN}}$ for the SN barrier can then be written exactly *in scaled form* as

$$J_{\mathrm{k}}^{\mathrm{SN}} = \theta f^2 \exp[-\mathrm{v}(f) \cdot \eta/f]. \tag{9.50}$$

Values of $\eta(\phi)$ and $\theta(\phi)$ are shown in Table 2 in [109] for a range of work-function values. For illustration, when $\phi = 4.50$ eV, then $\eta \approx 4.637$ and $\theta \approx 6.77 \times 10^{13}$ A/m^2. For algebraic simplicity in what follows, the explicit dependencies on ϕ are not shown.

Merits of the form (9.50) are that: (a) for a given ϕ-value (and hence given η-value), only a single independent variable (f) appears in the right-hand side; and (b) a good approximation for v(f) is known (see below). Hence, good approximate values for $J_{\mathrm{k}}^{\mathrm{SN}}$ are easy to obtain. Also, in some contexts, this scaled equation is easier to use than non-scaled equations, particularly since (in orthodox emission situations) f can also be treated as a scaled voltage.

By developing various different expansions for v(f), (9.50) can be written in various different *equation formats* [21].

9.6.3 The Principal Field Emission Special Mathematical Function v(x)

The Pure Mathematics of v(x)—Basic Formulae

As part of the tidy-up noted earlier, the author has advocated changes in the treatment of the function 'v'. One change is to make a clearer conceptual distinction between (a) the pure mathematical aspects of the special mathematical function v(x), and (b) its use in modelling barrier transmission. This subsection and the next four briefly summarise the pure mathematics of v(x); the following subsections consider its use in modelling. In line with this thinking, I now prefer to call v(x) the 'principal field emission special mathematical function'. The function v(f), where x has been

replaced by the modelling variable f ('the scaled field, for a barrier of height ϕ '), has previously been called the 'principal Schottky-Nordheim barrier function'. However, the function v(x) is useful in the mathematical analysis of various tunnelling barriers different from the SN barrier, in the contexts of both field electron and field ion emission.

The function v(x) has many alternative/equivalent definitions, but the most fundamental is as a very special solution of the Gauss Hypergeometric Differential Equation (HDE):

$$x(1-x)\mathrm{d}^2W/\mathrm{d}x^2 + \{\, c_\mathrm{G} - (a_\mathrm{G} + b_\mathrm{G} + 1)\mathrm{d}^2W/\mathrm{d}x - a_\mathrm{G}b_\mathrm{G}W \;=\; 0, \tag{9.51}$$

where W here is an abstract mathematical variable, and a_G, b_G and c_G are constants. Taking $a_\mathrm{G} = -3/4$, $b_\mathrm{G} = -1/4$, $c_\mathrm{G} = 0$, reduces (9.51) to the defining equation identified by Forbes and Deane [86], namely

$$x(1-x)\mathrm{d}^2W/\mathrm{d}x^2 \;=\; (3/16)\, W. \tag{9.52}$$

This is a special mathematical equation, like Airy's and Bessel's equations, but much more obscure.

Note that, if one recasts this equation to use the variable $y = +x^{1/2}$, the equation becomes

$$y(1-y^2)\frac{\mathrm{d}^2W}{\mathrm{d}y^2} - (1-y^2)\frac{\mathrm{d}W}{\mathrm{d}y} - \frac{3}{4}yW = 0. \tag{9.53}$$

This form is more complicated than (9.52), and less obviously a special case of the Gauss HDE.

The principal SN barrier function v(x) is a particular solution of (9.52) satisfying the unusual boundary conditions (derived in [57]):

$$\mathrm{v}(0) = 1; \quad \lim_{x\to 0}\{\mathrm{dv}/\mathrm{d}x - (3/16)\ln x\} \;=\; -(9/8)\ln 2. \tag{9.54}$$

An exact analytical solution to (9.52) is known [57], but is not given here because it involves obscure mathematical functions usually known only to mathematicians. This solution gives rise to the exact series expansion [57]

$$\mathrm{v}(x) = 1 - \left(\tfrac{9}{8}\ln 2 + \tfrac{3}{16}\right)x - \left(\tfrac{27}{256}\ln 2 - \tfrac{51}{1024}\right)x^2 - \left(\tfrac{315}{8192}\ln 2 - \tfrac{177}{8192}\right)x^3 + \mathrm{O}(x^4) + x\ln x\left[\tfrac{3}{16} + \tfrac{9}{512}x + \tfrac{105}{16384}x^2 + \mathrm{O}(x^3)\right]. \tag{9.55}$$

Recurrence formulae for determining higher-order terms are given in [57].

Note that this exact series expression contains no terms involving half-integral powers of x. This result, and the obvious superiority of (9.52) over the mathematically equivalent (9.53), shows that the Gauss variable x (rather than the Nordheim parameter y [$=+\sqrt{x}$]) is the natural mathematical variable to use. Also note that v(x) is mathematically unusual, in that its series definition needs TWO infinite power

series: this finding has been key to recent progress in FE theory, and will be key to further progress in FE data interpretation.

In the present chapter, the symbol x is used to denote the mathematical argument of v, whereas until recently I have used the symbol l'. The reason for the change is as follows. Until recently, the usual definition of v used complete elliptic integrals, as discussed below. In the theory of such integrals, one standard notation uses the symbols k, k', m and m'. The parameter l' was seen as part of an extended group of these parameters, related to m by $l' = [(1–m)/(1 + m)]^2$. With the discovery that, mathematically, a more fundamental definition of v is as a particular solution of the Gauss HDE, it makes better mathematical sense to use—as the argument of v—the independent variable in the Gauss HDE (here x), and call it the 'Gauss variable'. This is just a symbol change. All formulae previously expressed using l' are equally valid when expressed using x. The new notation is also less cumbersome.

The Pure Mathematics of v(x)—The 2006 Approximate Formula

Before the exact series expression was known, a good simple approximation had been found [115], namely

$$\mathrm{v}(x) \approx 1 - x + \frac{1}{6} x \ln x. \tag{9.56}$$

Over the range $0 \leq x \leq 1$, where v(x) varies from 1 to 0, this expression has absolute error everywhere less than 0.0024 and relative error everywhere less than 0.33% [86]. This performance is more than good enough for most technological purposes. Mathematically, v(x) is defined and well-behaved over the whole range $0 \leq x < \infty$, but the accuracy of formula (9.56) deteriorates rapidly above $x = 1$, and it should not be used above $x = 4$.

Formula (9.56) is not a truncated version of the exact series expansion, but was designed independently to have exactly correct behaviour at $x = 0$ and $x = 1$, namely $\mathrm{v}(0) = 1$ and $\mathrm{v}(1) = 0$. This could be achieved with any expression of the form $(1–x + qx\ln x)$. This form (involving an $x\ln x$ term) was partly suggested by numerical experiments using the computer algebra package MAPLE™, carried out with Dr. J.H.B. Deane. However, the choice $q = 1/6$ was made [115] using a spreadsheet [116], and was the 'best simple algebraic expression that gives good accuracy over the range $0 \leq x \leq 1$'. The author originally thought there might be some precise underlying mathematical reason for the appearance of the simple factor (1/6), but we now know there is not.

Exact formulae for v were first stated correctly by Burgess et al. [85] in 1953. Their formulae involve complete elliptic integrals, and can be evaluated by computer algebra to give very high-precision values. As shown in [86], slightly better fits to these 'exact' values can be obtained by choosing a numerical value for q by means of error minimization over the range $0 \leq x \leq 1$. However, since there are at least five different ways in which minimization can be carried out, all leading to slightly different values of q, it seemed best to opt for algebraic simplicity. If needed, a more accurate 'high precision' (numerical) formula for v(x) is given in Appendix 9.2.

As already noted, there are about 20 different approximate formulae for v in FE literature. When the whole range $0 \leq x \leq 1$ is considered, formula (9.56) is the 'best simple approximation' amongst those tested [117]. Presumably, this is largely because it is designed to be exactly correct at $x = 0$ and $x = 1$, and also mimics the form of the exact series expression, by including an $x\ln x$ term.

It is possible to get better accuracy over limited parts of the range $0 \leq x \leq 1$ with formulae of similar complexity [117], but such formulae often behave poorly or very poorly near $x = 1$. This makes them unsuitable for discussions of thermal-field phenomena, because transmission by flyover corresponds to the range $x > 1$. By contrast, formula (9.56) behaves quite well for forwards energy levels close above the top of the barrier. It is designed to give v(1) correctly, but also performs well in respect of the value of u(1), where $u(x) = -dv/dx$. From the mathematics of complete elliptic integrals (see below), the exact value $u^{exact}(1) = 3\pi/8\sqrt{2} \approx 0.83304$; from (9.56), $u^{aprx}(1) = (5/6) \approx 0.83333$; that is, agreement at $x = 1$ is within about 0.035%.

Another way of looking at this is as follows. By equating the formulae for $u^{exact}(1)$ and $u^{aprx}(1)$, we obtain a new numerical approximation for π, namely $\pi \approx (20\sqrt{2})/9 \approx 3.14270$. The exact value, to 6 significant figures, is $\pi \approx 3.14159$. The error, again, is about 0.035%.

The Pure Mathematics of v(x)—Definitions Using Complete Elliptic Integrals

The function $v(x)$ has multiple alternative (equivalent) definitions, obtained from each other by mathematical transformations. In particular, it has simple integral definitions, and also definitions in terms of complete elliptic integrals. The complete elliptic integral of the first kind (K) can itself be defined in alternative ways. In terms of the *elliptic modulus* k, K is given by

$$K = K[k] = \int_0^{\pi/2} (1 - k^2 \sin^2\theta)^{-1/2}\, d\theta \tag{9.57}$$

whereas in terms of the *elliptic parameter* m [$\equiv k^2$], K is given by the equivalent formula

$$K = K[m] = \int_0^{\pi/2} (1 - m \sin^2\theta)^{-1/2}\, d\theta. \tag{9.58}$$

A further complication has been that an older notation also used to exist, namely $K[k^2]$, which left it unclear to the non-expert what symbols such as '$K[0.4]$' really meant (it means put $k^2 = 0.4$, not put $k = 0.4$). It seems to have been a misinterpretation of this general kind that caused the error in the Nordheim (1928) paper [56] that persisted in FE theory until 1953.

The author's view has been (e.g. [116]) that use of the elliptic parameter m is less error-prone than the use of the elliptic modulus k, and consequently I use (9.56) for

K and the related definition for the complete elliptic integral of the second kind E:

$$\mathrm{E} = \mathrm{E}[m] = \int_0^{\pi/2} (1 - m \sin^2 \theta)^{+1/2} \, \mathrm{d}\theta \tag{9.59}$$

The next complication is that the part of FE literature that is mathematically correct provides three alternative (mathematically equivalent) choices for a formula for v(x) in terms of complete elliptic integrals. The most useful of these (in the updated notation used here) is considered to be the definition given by the linked equations

$$\mathrm{v}(x) \; = \; (1 + x^{1/2})^{1/2} \{\mathrm{E}[m_1(x)] - \mathrm{K}[m_1(x)]\}, \tag{9.60a}$$

$$m_1(x) = (1 - x^{1/2})/(1 + x^{1/2}). \tag{9.60b}$$

This should be the most useful formula for computer algebra, although specific packages may require E and K to be entered as functions of the elliptic modulus $k(x) = [m(x)]^{1/2}$. Equation (9.60a) here corresponds to (16) in [61], (26a) in [116], and (11) in [57]; m_1 here corresponds to m^* in [116] but m in [57].

As shown in [57], a formula for dv/dx can be derived from (9.60a), namely

$$\mathrm{dv/d}x \; = \; -(3/4)(1 + x^{1/2})^{-1/2} \mathrm{K}[m_1(x)] \tag{9.61}$$

For $x = 1$, $m_1(x = 1) = 0$. From standard tables (e.g. [118]), K($m = 0$) = $\pi/2$. Hence, u($x = 1$) = $3\pi/8\sqrt{2}$, where u(x) [=–dv/dx] is defined by (9.65a). This result is used earlier and below.

In the original derivations, (9.60) was derived, by means of a Gauss/Landen transformation, from a 'less convenient' standard result, namely (14) in [61] or (24b) in [116]. In present notation, this 'less convenient' result would be written

$$\mathrm{v}(x) \; = \; 2^{-1/2} \{1 + (1 - x)^{1/2}\}^{1/2} \{\mathrm{E}[m_2(x)] - \{1 - (1 - x)^{1/2}\} \mathrm{K}[m_2(x)] \, , \tag{9.62a}$$

$$m_2(x) \; = \; 2(1 - x)^{1/2} / \{1 + (1 - x)^{1/2}\}. \tag{9.62b}$$

(The older derivations use a or α to denote $(1–x)^{1/2}$ [=$(1–y^2)^{1/2}$]). MG considered that both (9.60) and (9.62) were inconvenient for discussion of flyover; thus, they provided a 'more convenient' third alternative formula, namely their (9.15). However, our modern experience is that the computer algebra package MAPLE™ has no problem in dealing with (9.60) when $x > 1$, and hence $m_1 < 0$, (and k_1 [=$m_1^{1/2}$] imaginary).

The Pure Mathematics of v(x)—Integral Definitions

The function v(x) also has integral definitions, related to each other by mathematical transformations, and related to definitions in terms of K and E by formulae in

mathematical handbooks. Possibly, the integral definition most useful for further mathematical manipulations is

$$\mathrm{v}(x) \;=\; (3\times 2^{-3/2})\int_{b'}^{a'} (a'^2-\omega'^2)^{1/2}(\omega'^2-b'^2)^{1/2}\,\mathrm{d}\omega', \tag{9.63a}$$

where ω', here, is a notional integration variable, and the limits of integration are defined by

$$a' \;=\; \{1+(1-x)^{1/2}\}^{1/2}; \quad b' \;=\; \{1-(1-x)^{1/2}\}^{1/2} \tag{9.63b}$$

This formula can be deduced using (19) in [116] and the definition of the second FN constant b in Appendix 9.1. The integral in (9.63) is a standard form that can be expressed in terms of complete elliptic integrals via a standard result in elliptic integral handbooks (see [116]).

An alternative integral definition of $\mathrm{v}(x)$, which is closer to the form derived from JWKB treatments of tunnelling, is obtained by writing $\omega' = \xi^{1/2}$, which yields

$$\mathrm{v}(x) = (3\times 2^{-5/2})\int_{q'}^{p'} (p'-\xi)^{1/2}(\xi-q')^{1/2}\xi^{-1/2}\,\mathrm{d}\xi \tag{9.64a}$$

$$p' \;=\; (a')^2 \;=\; 1+(1-x)^{1/2};\; q' \;=\; (b')^2 \;=\; 1-(1-x)^{1/2} \tag{9.64b}$$

In all cases, primes have been added to the symbols only to distinguish them from other uses of the symbols a, b, p, q and ω in FE theory.

Formulae for the Set of Special Mathematical Functions

In addition to $\mathrm{v}(x)$, several other special mathematical functions are used in connection with the SN barrier. These all have formal definitions in terms of $\mathrm{v}(x)$ and its derivative $\mathrm{dv}/\mathrm{d}x$, and approximate formulae for them can be found by applying these definitions to formula (9.56). The definitions and (where useful) approximate formulae are

$$\mathrm{u}(x) \;\equiv\; -\mathrm{dv}/\mathrm{d}x \;\approx\; \tfrac{5}{6} - \tfrac{1}{6}\ln x, \tag{9.65a}$$

$$\mathrm{t}_1(x) \;\equiv\; \mathrm{v}(x) - \tfrac{4}{3}x(\mathrm{dv}/\mathrm{d}x) \;\approx\; 1 + \tfrac{1}{9}x - \tfrac{1}{18}x\ln x, \tag{9.65b}$$

$$\mathrm{t}_2(x) \;\equiv\; \mathrm{v}(x)/(1-x), \tag{9.65c}$$

$$\mathrm{s}(x) \;\equiv\; \mathrm{v}(x) - x(\mathrm{dv}/\mathrm{d}x) \;\approx\; 1 - \tfrac{1}{6}x, \tag{9.65d}$$

$$\mathrm{r}(\eta, x) \equiv \exp[\eta \cdot \mathrm{u}(x)] \approx \exp\left[\eta\left\{\tfrac{5}{6} - \tfrac{1}{6}\ln x\right\}\right], \tag{9.65e}$$

$$\mathrm{w}(x) \equiv \mathrm{ds}/\mathrm{d}(x^{-1}) = x^3(\mathrm{d}^2\mathrm{v}/\mathrm{d}x^2) = \tfrac{3}{16}\{x^2/(1-x)\}\mathrm{v}(x) \tag{9.65f}$$

The function $\mathrm{r}(\eta,x)$ was first given in [114], and is denoted there by r_{2012}; it replaces a slightly different function denoted in [114] by r_{1999}. The modern expression for $\mathrm{t}_2(x)$ was derived in [119]. The remaining functions are derived from related functions given in [86]. (A typographical error in (5.1) in [86] has been corrected here.) It is suggested that these mathematical functions should simply be called by names of the type 'the field emission special mathematical function u(x)'.

At $x = 1$, the exact values of these functions can be obtained from the theory of complete elliptic integrals (see below). Relevant values are: $\mathrm{u}(1) = 3\pi/(8\sqrt{2}) \cong 0.8330406$; $\mathrm{t}_1(1) = \pi/(2\sqrt{2}) \cong 1.110721$.

If more accurate values of these functions are needed for other x-values, then formulae in Appendix 9.2 (derived from those in [86]), can be used. These have been derived as numerically optimised fits to MAPLE™ evaluations of formulae (9.60) and (9.61), and have a maximum absolute error of 8×10^{-10}. Sample MAPLE™ evaluations have been checked against independently programmed evaluations of integral definitions of v(x), with agreement to 12 decimal places.

Use of v(x) in Modelling Transmission Across the Laurent-form Barrier

The mathematical function v(x) is useful in modelling transmission across, not only the SN barrier, but also across barriers of similar mathematical form that appear in the theories of semiconductor FE, electrostatic field ionisation, field desorption and field evaporation. This section shows that this function is applicable to *any* barrier with motive energy of the form $M(z) = H - Cz - B/z$, where H, C and B are constants. The result can then be customised to the SN barrier.

A function of the form just stated seems not to have any well defined name, so I refer to it as the *basic Laurent-form barrier*. In practice, it is more convenient to write C in the form 'eF', where e (as usual) is the elementary positive charge and F is a parameter initially regarded as of general significance only. Thus, the barrier of interest has motive energy

$$M(z) = H - eFz - B/z \tag{9.66}$$

The zeros of $M(z) = 0$ are:

$$z = (H/2eF)\left\{1 \pm \sqrt{1 - 4eFB/H^2}\right\}. \tag{9.67}$$

By defining the *barrier parameter* μ and a mathematical parameter α' by

$$\mu = 4eFB/H^2, \quad \alpha' = +(1-\mu)^{1/2}, \tag{9.68}$$

(9.67) can be written in the simplified form

$$z = (H/2eF)\left(1 \pm \alpha'\right) \tag{9.69}$$

From a generalisation of (9.15), noting the analogy with the relation $2\kappa_e = 2(2m_e)^{1/2}/\hbar$, the barrier strength G for an entity of mass m (which is not necessarily an electron), for the basic Laurent-form barrier, can be written

$$G = (3e/2)\left[\tfrac{4}{3}(2m)^{1/2}/e\hbar\right] \times \int_{(H/2eF)(1-\alpha')}^{(H/2eF)(1+\alpha')} (H - eFz - B/z)^{1/2}\,\mathrm{d}z. \tag{9.70}$$

The transformation $z = (H/2eF)\xi$ yields

$$G = G_{0,m} \cdot (3 \times 2^{-5/2}) \int_{1-\alpha'}^{1+\alpha'} (-\xi^2 + 2\xi - \mu)^{1/2}\xi^{-1/2}\,\mathrm{d}\xi. \tag{9.71a}$$

where the symbol $G_{0,m}$ denotes the combination

$$G_{0,m} = [(4/3)(2m)^{1/2}/e\hbar](H^{3/2}/F). \tag{9.71b}$$

Since the zeros of the expression $(\xi^2 - 2\xi + \mu)$ are $\xi = 1 \pm (1-\mu)^{1/2} = 1 \pm \alpha'$, (9.71a) can be rewritten as

$$G = G_{0,m} \cdot (3 \times 2^{-5/2}) \times \int_{1-\alpha'}^{1+\alpha'} \{(1+\alpha') - \xi\}^{1/2}\{\xi - (1-\alpha')\}^{1/2}\xi^{-1/2}\,\mathrm{d}\xi. \tag{9.72}$$

Comparison with (9.64), taking $p' = (1-\alpha')$, $q' = (1+\alpha')$ shows that (9.72) involves one of the definitions of $\mathrm{v}(x)$, and hence can be written in the form

$$G = \mathrm{v}(\mu) \cdot G_{0,m}. \tag{9.73}$$

Hence, when μ is defined by (9.68), then $\mathrm{v}(x = \mu)$ is the barrier form correction factor for the basic Laurent-form barrier.

For the specific case of electron tunnelling through a Schottky-Nordheim barrier, F resumes its role as the local barrier field, $G_{0,m}$ becomes the quantity G^{ET} used earlier, and $B = e^2/16\pi\varepsilon_0$. Equation (9.68) then becomes a definition of the parameter f_H, or—in the case when $H = \phi$—the parameter f. Hence, the barrier form correction factor for the SN barrier becomes given by $\mathrm{v}(x = f_{\mathrm{H}})$ or (when $H = \phi$) $\mathrm{v}(x = f)$.

The Choice between Nordheim Parameter and Scaled Field

In the treatment above, the Nordheim parameter y does not appear. This is because I have chosen to define my barrier parameter (μ) by (9.68), rather than as a parameter

equal to $\sqrt{\mu}$. Either choice could be made to work mathematically, and in 1928 Nordheim took the opposite choice. However, there is no good mathematical reason to stick to his choice, particularly since this choice seems to have been part of the cause of the mathematical error made in his paper.

In the present treatment, it is probably better to leave μ as the barrier parameter for the basic Laurent-form barrier, and ask the alternative question: "When this theory is applied to the SN barrier, is it better to set $\mu = f_H$ (or f) or to set $\mu = y^2$?" Obviously, the latter is the more familiar convention, but the author's strong view is that use of f_H and f is in principle much the better convention. Arguments are as follows.

(1) For reasons noted earlier, the *natural* pure mathematical variable to use is the Gauss variable x: the modelling equivalents of x are μ, f_H and f, rather than y.
(2) The concept of 'scaled field' is probably easier to understand and use than the Nordheim parameter (which physically is 'scaled reduction in SN-barrier height').
(3) The parameters f_H and f are proportional to the barrier field F, whereas y is proportional to $\sqrt{F}$. This often makes f easier to use than y, particularly in the context of an orthodoxy test [109].
(4) The symbols f_H and f each have unique definitions, whereas, historically, depending on the paper and context in which it appears, the symbol y may mean $\sqrt{x}$, $\sqrt{f_H}$ or $\sqrt{f}$.
(5) The concept of 'scaled field' is, in principle, more general than the Nordheim parameter concept, and can easily be extended physically (when suitably modified in detail) to apply to real physical barriers. However, the parameter y represents the scaled barrier height ONLY for a SN barrier.

In the author's view, it would make for a simpler and clearer system if the use of y were gradually phased out, and relevant formulas and tables using y were replaced by formulas and tables using x or f.

In the present mixed system, a danger for non-experts is confusion between the meanings of 'v(f)' and 'v(y)', as this is not a change in symbol but a change in variable and (strictly) a change in mathematical function. To avoid confusion, when substituting (say) f for x in the function v(x), I often write v($x = f$). Also, again to avoid confusion, the argument-free symbols v_F and t_F are sometimes used as alternative notation in Murphy-Good FE equations, i.e. one writes: $v_F \equiv v(x = f)$, $t_F \equiv t_1(x = f)$.

9.7 Device and System Theory Issues

Many issues covered in Sects. 9.7 and 9.8 have recently been discussed in detail elsewhere [20, 21], and some are still topics of active research; hence, only an overview is given here. For details, readers are referred to the cited papers.

For clarity in Sect. 9.7, the local barrier field and local ECD are denoted by F_L and J_L. In practice, interest is usually in their values at some point 'C' regarded

as 'characteristic' of the emission. Values of barrier field and ECD at this point are denoted by F_C and J_C, respectively. In emission modelling (where it is usually assumed that local work function is constant across the surface), it is usual to take 'C' at an emitter apex, where (except in special circumstances) the field is locally highest.

9.7.1 Basic Auxiliary Parameters for Ideal FE Devices/Systems

To apply emission theory to FE devices and systems, *auxiliary parameters* are needed that relate F_C and J_C (or, more generally, F_L and J_L) to measured voltage (V_m) and measured current (I_m). Real FE systems involve electrical measurement and control circuits, of which the FE device forms part. Such systems may exhibit current-leakage and voltage-loss effects. The usual assumption is that current-leakage effects, once detected, can be eliminated by careful system design, but that voltage-loss effects are intrinsic to many forms of FE device, and cannot easily be eliminated.

An *ideal FE device/system* is one where it is adequate to assume that no such effects occur, and that no other distorting effects (such as, amongst others, field dependent changes in geometry, current dependence in field enhancement factors, temperature dependence in work function, or effects due to field emitted vacuum space-charge) occur. In an ideal system, the $I_m(V_m)$ relationship is determined completely by the following factors: the total system geometry (including unchanging emitter shape); unchanging emitter surface composition; and the emission process.

At this point in time (and probably for some years to come), in $I_m(V_m)$ data-interpretation theory it will be simplest to consider only the behaviour of *ideal* FE devices/systems. With real systems, it will be necessary to precede data analysis by an orthodoxy test [109] that uses a FN plot or MG plot to check whether the device/system is ideal.

For ideal FE devices/systems, a convenient theoretical approach uses two linked equations, the first an expression for measured current I_m in terms of characteristic field F_C (or corresponding scaled field f_C), and the second an expression for F_C (or f_C) in terms of measured voltage V_m. On omitting the superscript 'GB' used earlier, the equation for local ECD is written

$$J_L = \lambda_L J_{kL}. \tag{9.74}$$

A formal integration is done over the whole emitting area and the result is written

$$I_e = \int J_L \mathrm{d}A = A_n J_C = A_n \lambda_C J_{kC} = A_f J_{kC}. \tag{9.75}$$

Here A_n is the so-called *notional emission area*, and A_f [$\equiv \lambda_C A_n$] is the *formal emission area* (which will be substantially different in value from A_n if λ_C is substantially different from unity).

The quantity extracted from the corresponding 'ideal' FN or MG plot is a *formal* emission area A_f. Its type and value depend on what is assumed about the form of the tunnelling barrier. With SPME methodology, the best simple assumption is a SN barrier [82, 110]. Hence, for an ideal FE device/system, it is currently best to write

$$I_m = A_f^{SN} J_{kC}^{SN}, \tag{9.76}$$

where A_f^{SN} is the formal emission area for the SN barrier, and J_{kC}^{SN} is the characteristic kernel current density for the SN barrier, as given by (9.45), with $F = F_C$ and $f = f_C$, (or as given by an equivalent scaled equation).

Equation (9.76), in its various formats that incorporate the correction factor λ_C^{SN}, has been called [110] the *Extended Murphy-Good equation*. As indicated earlier, my current thinking [43] is that this correction factor, which has unknown functional dependencies and values, probably always lies within the range of $0.005 \leq \lambda_C^{SN} \leq 14$. But quite possibly the lower limit is too pessimistic.

The relation between field F_C and measured voltage V_m can be written in three alternative forms, all of which are in use. The formulae are

$$F_C = \beta_V V_m = V_m/\zeta_C = V_m/k_C r_a. \tag{9.77}$$

Here, β_V is the (characteristic) *voltage-to-local-field conversion factor (VCF)* (also denoted by 'β' in older literature) (unit: nm^{-1}, or equivalent); ζ_C is the (characteristic) *voltage conversion length (VCL)* (unit: nm, or equivalent); r_a is the emitter apex radius of curvature, and k_C (dimensionless) is called a *shape factor* or a *field factor*. The author now prefers the form involving the VCL ζ_C, although this is not the usual literature choice.

The VCL ζ_C is not a physical length (except in special cases). Rather, it is a system-geometry characterization parameter that assesses how easy it is to 'turn the emitter on'. For example, if we consider that an apex barrier field of 2 V/nm is needed to turn the emitter on, and $\zeta_C = 10$ nm, then turn-on occurs at $V_m = 20$ V.

For an *ideal* FE device/system, all these '*auxiliary parameters*' are constants (independent of field and voltage), and depend only on the (unchanging) total system geometry, including emitter shape. The literature contains many electrostatic analyses that lead to different formulae for the auxiliary parameters in (9.77), for different assumptions about system geometry and emitter shape. For single tip field emitters (STFEs), it is often thought that the 1956 *sphere-on-orthogonal-cone (SOC) model* of Dyke et al. [31] is the best simple choice for a needle-shaped STFE, and that the hemisphere-on-cylindrical-post (HCP) model (e.g. [120]) is the best simple choice for a post-shaped STFE.

9.7.2 Field Enhancement Factors

Additional auxiliary parameters, called field enhancement factors (FEFs) and denoted here by the symbol γ or γ_{M}, are used to describe the behaviour of single post-like emitters and of large-area field electron emitters (LAFEs). In the literature, FEFs are usually denoted by the symbol β. Due to the possibility (and sometimes reality) of confusion between this use and literature use of the symbol 'β' to denote a voltage conversion factor, I prefer to use the basic symbol γ.

A FEF can be defined at any emitter surface location, but interest is mainly in some characteristic value, γ_{MC}, usually thought of as the value at the apex of a single emitter or (with a LAFE) at the apex of a strongly emitting particular emitter, or at some other 'high point'.

Several different kinds of FEF are in use. These apply to slightly different geometrical situations, but in all cases, the characteristic FEF can be related to the characteristic VCL by an equation of the general form

$$\gamma_{\mathrm{MC}} = d_{\mathrm{M}}/\zeta_{\mathrm{C}} \tag{9.78}$$

where d_{M} is a *macroscopic distance* associated with the chosen mode of definition. The most common variants are the 'gap FEF' (replace 'M' by 'G'), where d_{G} is the length of the gap between characteristic location 'C' and the counter-electrode, and the 'plate FEF' (replace 'M' by 'P') where the emitter stands on one of a pair of well-separated parallel planar plates (PPP), separated by a distance d_{P}.

For an *ideal* PPP system, the voltage V_{P} between the plates is equal to the measured voltage V_{m}, so it follows that

$$F_{\mathrm{C}} = V_{\mathrm{m}}/\zeta_{\mathrm{C}} = V_{\mathrm{P}}/\zeta_{\mathrm{C}} = V_{\mathrm{P}}\gamma_{\mathrm{PC}}/d_{\mathrm{P}} = \gamma_{\mathrm{PC}}F_{\mathrm{P}}, \tag{9.79}$$

where the 'plate field' F_{P} is the mean field between the parallel plates. Thus, consistency is demonstrated between relation (9.79) and the more common definition of characteristic plate FEF γ_{PC}, namely $\gamma_{\mathrm{PC}} = F_{\mathrm{C}}/F_{\mathrm{P}}$.

For an *ideal* PPP device/system, the plate FEF is independent of the measured current and voltage. Provided that d_{P} is many times the emitter height h (preferably $d_{\mathrm{P}} > 10\,h$), this plate FEF characterises the 'sharpness' of the emitter alone. However, the value of a gap FEF may often depend on the whole system geometry.

Plate FEFs, gap FEFs and similar parameters are collectively called *macroscopic FEFs* here (previously, this name was applied only to plate FEFs).

LAFEs typically consist of large or very large numbers of individual emitters or emission sites, and are often modelled as arrays of upright posts, although this is not appropriate for all LAFE geometries. In LAFE array models, electrostatic interactions (often called 'shielding') occur between post-like emitters, and lead to significant electrostatic depolarisation effects if the emitters are sufficiently close together. There is now an extensive (and growing) literature on this topic, but details are outside the scope of this chapter. Gateways to it are provided by [121–125].

Note that, with resistive emitters, it is possible for current flow through the emitter to reduce the field enhancement factor, and hence make the voltage conversion length into a current-dependent parameter [126, 127]. The author is now of the view that this may be a common cause of non-linearity in experimental Fowler-Nordheim plots.

9.7.3 Macroscopic Current Density and Area Efficiency

The macroscopic or 'LAFE-average' emission current density J_M is defined by

$$J_M = I_m/A_M, \tag{9.80}$$

where A_M is the macroscopic area (or "footprint") of the device. From the general version of (9.76), we have

$$J_M = (A_f/A_M)J_{kC} \equiv \alpha_f J_{kC}. \tag{9.81}$$

where α_f [$\equiv A_f/A_M$] is the *formal area efficiency* of the LAFE. It is also possible, although less useful (due to the uncertainties in the value of λ_C), to define a *notional area efficiency* α_n by

$$\alpha_n \equiv A_n/A_M = (A_f/\lambda_C A_M) = \alpha_f/\lambda_C. \tag{9.82}$$

Values of formal area efficiency can be extracted from experimental data, but—as with formal emission area—the values extracted [using definition (9.81)] depend on the choice of barrier form. As of now, when most $I_m(V_m)$ data analysis is carried out in the context of SPME methodology, the best simple choice is the SN barrier, as already indicated.

The parameter α_f^{SN} is a measure of what fraction of the area of a LAFE is apparently emitting electrons. Relatively few values for α_f^{SN} have been reliably extracted, but it is thought that values for these dimensionless parameters may vary widely from material to material, and vary as between different LAFE processing regimes. Values are not well known, but are thought to perhaps typically lie in the range 10^{-7} to 10^{-4}. (See [128] for corresponding, but older, remarks about the possible value of α_n.)

Unfortunately, it has been relatively common practice in FE technological literature not to make a distinction between local current density J_L (or J_C) and macroscopic current density J_M, and to omit an area efficiency factor from published equations where it principle it ought to appear. This has given rise to situations where there are undiscussed large apparent discrepancies between theory and experiment in particular papers, and has apparently led some researchers to think that area information cannot be extracted from FN plots, even for ideal FE devices/systems. These issues have been considered elsewhere [20, 128, 129] but now require more careful discussion and further research, including reexamination of existing published data.

9.8 The Interpretation of Measured Current-Voltage Data

The commonest methodology of characterising FE materials and devices involves measurement of current-voltage $[I_{\rm m}(V_{\rm m})]$ characteristics, and then making some form of Fowler-Nordheim (FN) plot (see [20]), although there has been a recent suggestion that an alternative plot form—the so-called *Murphy-Good (MG) plot*—would provide greater precision, particularly in extracting formal emission area $A_{\rm f}^{\rm SN}$ [110], and/or that multi-parameter numerical fitting methods could be used. This section outlines the main points suggested in recent overviews [20, 21].

1. Both FN and MG plots should be made using raw measured $I_{\rm m}(V_{\rm m})$ data. It is advisable to discontinue the widespread community practice of pre-converting voltages to become (apparent) macroscopic fields, because the pre-conversion equation often used can be defective when applied to real systems [130], and this can lead to spurious results for characterization parameters [109].
2. Virtually all FE $I_{\rm m}(V_{\rm m})$ data analysis is currently carried out within the framework of SPME methodology. Exploration of how to carry out data analysis within a 'point-form emitter methodology' that treats the emitter as needle-shaped or post-shaped is long overdue and an active topic of research (e.g. [14, 44, 45]). However, the immediate practical need is to improve SPME methodology, in order to extract precisely defined parameters; the remainder of this section discusses this.
3. To analyse either a FN plot or a MG plot, within SPME methodology, it should be assumed that FE is adequately described by the Extended Murphy-Good FE equation. Formulae given here relate to this case.
4. An orthodoxy test [20, 109] should always be applied to the plot before detailed analysis; a form suitable for an MG plot is described in [131, 132].
5. For a FN plot that passes the orthodoxy test, the voltage conversion length (VCL) $\zeta_{\rm C}$ can be extracted by using the formula

$$\{\zeta_{\rm C}\}^{\rm extr} \; = \; -S_{\rm FN}^{\rm fit}/{\rm s_t} b\phi^{3/2}. \tag{9.83}$$

   where $S_{\rm SN}^{\rm fit}$ is the slope of a straight line fitted to the FN plot, and $\rm s_t$ is the fitting value [20] of the slope correction function $\rm s(\mathit{f})$. A related macroscopic FEF can then be derived by using formula (9.78).
6. For a FN plot that passes the orthodoxy test, the formal emission area $A_{\rm f}^{\rm SN}$ can be extracted by using the formula

$$\{A_{\rm f}^{\rm SN}\}^{\rm extr} \; = \; \Lambda_{\rm FN}^{\rm SN} R_{\rm FN}^{\rm fit} (S_{\rm FN}^{\rm fit})^2, \tag{9.84}$$

   where $\ln\{R_{\rm FN}^{\rm fit}\}$ is the intercept made by the fitted straight line with the $1/V_{\rm m} = 0$ axis, and $\Lambda_{\rm FN}^{\rm SN}$ is the *emission area extraction parameter for the SN barrier, using a FN plot,* given by

$$\Lambda_{\rm FN}^{\rm SN} = 1/[(ab^2\phi^2)({\rm r_t s_t^2})], \tag{9.85}$$

where r_t is the fitting value [21] of the intercept correction function $r(\eta,f)$. For $\phi = 4.50$ eV, A_{FN}^{SN} has the approximate value of 6 nm^2/A. A related formal area efficiency $\{\alpha_f^{SN}\}^{extr}$ (for the SN barrier) can be derived as $\{A_f^{SN}\}^{extr}/A_M$.

7. Equivalent formulae for the MG plot can be found in [110]. For ideal FE devices/systems, the extracted values of A_f^{SN} and and α_f^{SN} are significantly more precise when a MG plot is used.
8. Technological FE papers often use the elementary FE equation (see Table 9.3) to extract information from the FN plot slope. As compared with a MG-type equation, this omits the factor s_t in (9.83), and causes an error of about 5%. In practice, this equation is not used to extract emission area values, but if it were then an error of order 100 would arise, resulting from the omission of the factor $(r_t s_t^2)$ in (9.85). The existence of this potential error is a strong argument for using a Murphy-Good-type equation to interpret FE current-voltage data (whether using a FN plot or a MG plot).
9. As noted above, when writing an equation for macroscopic current density J_M, technological FE papers often omit the factor α_f. This can lead to very large apparent discrepancies between theory and experiment. A factor describing area efficiency should always be included in equations for macroscopic current density.
10. For experimental data that fail the orthodoxy test, it is possible to use the procedure of *phenomenological adjustment* [20], to extract empirical values of ζ_C and γ_{MC} that are more realistic than straightforward (but invalid) application of formula (9.83) would yield. The problem of extracting valid estimates of *area-like* characterization parameters for non-ideal devices/systems, when the cause of non-ideality is unknown, is so complicated and messy that it almost seems premature to attempt to do so.
11. There remains an urgent need for research into a comprehensive method of analysing data from non-ideal devices/systems, but only limited progress has been made. One route forwards is provided by the work of Bachmann et al. [133].

9.9 Future Needs

Field electron emission has many applications in technology, and good basic theory is needed in order to optimise both the interpretation of results and the prediction of technology behaviour, particularly in contexts such as electrical breakdown. More generally, FE has an enduring role within physics as one of the fundamental exemplars of quantum-mechanical tunnelling, and its theory deserves to be described clearly, correctly and in an integrated manner. As an exemplar of tunnelling, FE may have a role in new fields such as quantum biology.

This chapter has primarily been about the consolidation of FE theory and the interpretation of *ideal* FE measured current-voltage data, both of these within the framework of smooth-planar-metal-like-emitter methodology. This methodology is near-universally used by experimentalists as a first-approximation approach to interpreting FE current-voltage data, whatever the material used in experiments.

This chapter partially describes the first of several stages needed to put FE theory onto a more scientific basis by making it easier to link theory and experiment. There is already much relevant emission theory published, but future stages will need to develop a methodology for interpreting measured current-voltage data that takes into account: (a) the shape of emitters, as it affects the process of integration in (9.75); (b) the effect of emitter shape and surface curvature on local transmission probability; (c) the influence on transmission probability of including atomic structure, first for planar surfaces and then for curved surfaces; and (d), if it proves necessary, any weaknesses that may be confirmed/discovered in present formulations of basic quantum-mechanical transmission theory.

It would be helpful to extend/customise data-interpretation methodologies to deal with the specific problems of semiconductors (such as surface states, band-bending and field penetration), and to deal with any further specific problems related to carbon-based materials, in particular carbon nanotubes. However, it may be better to delay this until after the development of methodologies for interpreting data from point-form emitters, since new transport mechanisms may need to be considered, such as transport via surface states on the emitter shank. There is also the possibility that emitters that are not bulk metals may be both 'surface conductors' and 'surface-state emitters', as a result of the presence of active surface states, and thus 'quasi-metallic' in their behaviour.

It would be helpful to develop data-interpretation methodologies that apply to FE devices/systems that are non-ideal for various specific reasons, and to develop (if possible) some reliable method for identifying the precise reason for non-ideality. Finally, it would probably be helpful to develop variants of all the above that apply to thermal-field emission regimes. There would appear to be many years of work ahead.

Appendix 9.1. Fundamental and Universal Constants Used in Field Emission Theory

Table 9.4 presents values of the fundamental constants used in FE theory, both in SI units and in the 'field emission customary units' in which they are often given. Both sets of units are defined in the context of ISQ-based equations, and the normal rules of quantity algebra apply to their values. These customary units take the electronvolt, rather than the joule, as the unit of energy, and normally measure field in V/nm and charge in units equal to the elementary positive charge. Their use greatly simplifies basic calculations when energies are measured in eV and fields in V/nm, as is often

Table 9.4 The May 2019 values of the electronvolt (eV) and fundamental constants used in emission physics, given in SI units (either exactly or to 11 significant figures), and in field emission customary units (to 7 significant figures). Asterisks indicate constants where (since May 2019) the value in SI units is specified exactly

Name	Symbol	SI units		Atomic-level units based on eV	
		Numerical value	Units	Numerical value	Units
Electronvolt*	eV	$1.602\ 176\ 6340 \times 10^{-19}$	J	1	eV
Elementary (positive) charge*	e	$1.602\ 176\ 6340 \times 10^{-19}$	C	1	eV V^{-1}
Elementary constant of amount of substance[a]	n_1	$1.660\ 539\ 0672 \times 10^{-24}$	mol	1	Entity[a]
Unified atomic mass constant	m_u	$1.660\ 539\ 6660 \times 10^{-27}$	kg	$1.036\ 427 \times 10^{-26}$	eV nm^{-2} s^2
Electron mass in free space	m_e	$9.109\ 383\ 7015 \times 10^{-31}$	kg	$5.685\ 630 \times 10^{-30}$	eV nm^{-2} s^2
Vacuum electric permittivity	ε_0	$8.854\ 187\ 8128 \times 10^{-12}$	F m^{-1}	$5.526\ 349 \times 10^{-2}$	eV V^{-2} nm^{-1}
$\varepsilon_0 \times 4\pi$	$4\pi\varepsilon_0$	$1.112\ 650\ 0554 \times 10^{-10}$	F m^{-1}	0.694 4615	eV V^{-2} nm^{-1}
Planck's constant*	h_P	$6.626\ 070\ 15 \times 10^{-34}$	J s	$4.135\ 668 \times 10^{-15}$	eV s
Planck's constant $\div 2\pi$	$\hbar$	$1.054\ 571\ 8176 \times 10^{-34}$	J s	$6.582\ 120 \times 10^{-16}$	eV s
Boltzmann's constant*	k_B	$1.380\ 649 \times 10^{-23}$	J K^{-1}	$8.617\ 333 \times 10^{-5}$	eV K^{-1}

[a]Name used here, not 'official'. n_1 is equal to the reciprocal of the (exactly given) Avogadro constant

the case. The numerical values of the constants, measured in both sets of units, are based on the values of the fundamental constants given in October 2019 on the website of the US National Institute of Standards and Technology (NIST) (see http://physics.nist.gov/constants). These values incorporate the SI system changes made in May 2019.

Table 9.5 gives the values, in FE customary units, of many universal constants, and combinations thereof, that are used in field emission theory. Where necessary, the tabulated values take the May 2019 changes into account. The third column in Table 9.5 is interesting, because it shows that several of the universal constants used in field emission are related in a relatively simple way to the constants κ_e and z_S, which are very basic universal constants that appear centrally in the Schrödinger equation and in statistical mechanics, respectively.

Table 9.5 Some fundamental constants used in field emission and related theory. Values are given in field emission customary units

Name	Symbol	Derivation	Expression	Numerical value	Units
Coulomb law constant	–	–	$4\pi\varepsilon_0$	0.6944 615	eV V^{-2} nm^{-1}
Image potential energy constant	B	–	$e^2/16\pi\varepsilon_0$	0.3599 911	eV nm
Sommerfeld electron supply constant	z_S	–	$4\pi e m_e/h_P^3$	1.618 311×10^{14}	A m^{-2} eV^{-2}
Schrödinger equation constant for electron[a]	κ_e	–	$(2m_e)^{1/2}/\hbar$	5.123 168	$eV^{-1/2}$ nm^{-1}
JWKB constant for electron[b]	(g_e)	$2\kappa_e$	$2(2m_e)^{1/2}/\hbar$	10.24 624	$eV^{-1/2}$ nm^{-1}
Triangular barrier constant	c_κ	$(\kappa_e/e)^{1/3}$	$(2m_e)^{1/6}/(e\hbar)^{1/3}$	1.723 903	$eV^{-1/2}$ $(V/nm)^{-1/3}$
First Fowler-Nordheim constant	a	$z_S(e/2\kappa_e)^2$	$e^3/8\pi h_P$	1.541 434	μA eV V^{-2}
Second Fowler-Nordheim constant	b	$4\kappa_e/3e$	$(4/3)(2m_e)^{1/2}/e\hbar$	6.830 890	$eV^{-3/2}$ (V/nm)
–	ab^2	$4z_S/9$	–	7.192 493×10^{-5}	A nm^{-2} eV^{-2}
Universal theoretical Richardson constant	A_{R0}	$z_S k_B^2$	$4\pi e m_e k_B^2/h_P^3$	1.201 735×10^{6}	A m^{-2} K^{-2}
Schottky constant	c_S	–	$(e^3/4\pi\varepsilon_0)^{1/2}$	1.199 985	eV $(V/nm)^{-1/2}$
		c_S^2	$e^3/4\pi\varepsilon_0$	1.439 965	eV^2 $(V/nm)^{-1}$
		c_S^{-2}	$4\pi\varepsilon_0/e^3$	0.694 4615	eV^{-2} (V/nm)
		ac_S^{-4}	–	7.333 978×10^{11}	A m^{-2} eV^{-3}
		bc_S^2	–	9.836 239	$eV^{1/2}$
		$bc_S^{1/2}$	–	7.482 819	eV $(V/nm)^{3/4}$
–	C_p	$3bk_B/2$	$2(2m_e)^{1/2}k_B/e\hbar$	8.829 607×10^{-4}	$eV^{-1/2}$ K^{-1} (V/nm)
–	C_q	$(2/3t_b)\times(bc_S^{1/2}k_B)$	$(\hbar/\pi m_e^{1/2})\times(4\pi\varepsilon_0 e)^{1/4}$	930.8202	K $(V/nm)^{-3/4}$

[a] As introduced in the Fowler and Nordheim 1928 paper [34]
[b] Previously denoted by g_e, but notation $(2\kappa_e)$ is now preferred

Appendix 9.2. High-Precision Formulae for v(x) and u(x)

This Appendix provides formulae for estimating high-precision values of the FE special mathematical functions v(x) and u(x) [$\equiv -\mathrm{dv}/\mathrm{d}x$] (and hence of all the FE special mathematical functions). The parameter x is the *Gauss variable*. The two functions are estimated by the following series, derived from those given in [86] by replacing the symbol *l'* by the symbol x now preferred, and by slightly adjusting the form of the resulting series for v(x) (without changing its numerical predictions)

$$\mathrm{v}(x) \cong (1-x)\left(1+\sum_{i=1}^{4} p_i x^i\right) + x\ln x \sum_{i=1}^{4} q_i x^{i-1} \tag{9.86}$$

$$\mathrm{u}(x) \cong \mathrm{u}_1 - (1-x)\sum_{i=0}^{5} s_i x^i - \ln x \sum_{i=0}^{4} t_i x^i \tag{9.87}$$

Values of the constant coefficients p_i, q_i, s_i and t_i are shown in Table 9.6.

It can be seen that at the values $x = 0$, 1, formula (9.86) generates the correct values $\mathrm{v}(0) = 1$, $\mathrm{v}(1) = 0$, and that at $x = 1$, formula (9.87) generates the correct value $\mathrm{u}(1) = \mathrm{u}_1 = 3\pi/8\sqrt{2}$.

Equation (9.86) mimics the form of the lower-order terms in the (infinite) exact series expansion for v(x) [57], but the coefficients in Table 9.6 have been determined by numerical fitting to exact expressions for v(x) and u(x) (in term of complete elliptic integrals) evaluated by the computer algebra package MAPLE™. In the range $0 \leq x \leq 1$, v(x) takes values lying in the range $1 \geq \mathrm{v}(x) \geq 0$, and the maximum error associated with formulae (9.86) and (9.87) is less than 8×10^{-10} [57]. In Murphy-Good-type FE equations, these formulae are applied by setting $x = f_\mathrm{C}$.

Table 9.6 Numerical constants for use in connection with (9.86) and (9.87)

i	p_i	q_i	s_i	t_i
0	–	–	0.053 249 972 7	0.187 5
1	0.032 705 304 46	0.187 499 344 1	0.024 222 259 59	0.035 155 558 74
2	0.009 157 798 739	0.017 506 369 47	0.015 122 059 58	0.019 127 526 80
3	0.002 644 272 807	0.005 527 069 444	0.007 550 739 834	0.011 522 840 09
4	0.000 089 871 738 11	0.001 023 904 180	0.000 639 172 865 9	0.003 624 569 427
5	–	–	−0.000 048 819 745 89	–
$\mathrm{u}_1 \equiv \mathrm{u}(x=1) = 3\pi/8\sqrt{2} \cong 0.8330405509$				

References

1. M.I. Elinson (ed.), *Unheated Cathodes* (Soviet Radio, Moscow, 1974). (in Russian)
2. A. Modinos, *Field, Thermionic, and Secondary Electron Emission Spectroscopy* (Plenum, New York, 1984) (Reprinted by: Springer, online, 2013)
3. G.A. Mesyats, *Explosive Electron Emission* (URO-Press, Ekaterinburg, 1998)
4. V.M. Zhukov, *Surface Processes in Field Electron Emission* (Isdatelstvo, St Petersburg, 2007). (in Russian)
5. W. Zhu (ed.), *Vacuum Microelectronics* (Wiley, New York, 2001)
6. G.N. Fursey, *Field Emission in Vacuum Microelectronics* (Kluwer, New York, 2001)
7. K.L. Jensen, Electron emission physics. Adv. Imaging Electron. Phys. **149**, 1–338 (2007)
8. Y. Saito (ed.), *Carbon Nanotube and Related Field Emitters* (Wiley, Weinheim, 2010)
9. N.V. Egorov, E.P. Sheshin, *Field Emission: Principles and Practice* (Intellect, Russia, 2011). (in Russian) (Reference [13] is the English edition)
10. S.-D. Liang, *Quantum Tunnelling and Field Electron Emission Theories* (World Scientific, Singapore, 2014)
11. M. Bronsgeest, *Physics of Schottky Electron Sources* (Pan Stanford, Singapore, 2014)
12. A. Evtukh, H. Hartnagel, O. Yilmazoglu, H. Mimura, D. Paylidis, *Vacuum Nanoelectronic Devices* (Wiley, Chichester, 2015)
13. N.V. Egorov, E.P. Sheshin, *Field Emission Electronics* (Springer, 2017)
14. K.L. Jensen, *Introduction to the Physics of Electron Emission* (Wiley, Hoboken, NJ, 2018)
15. L.W. Swanson, G.A. Schwind, Review of ZrO/W Schottky cathode, in *Charged Particle Optics*, 2nd edn., ed. by J. Orloff (CRC Press, Baton Rouge, 2009)
16. L.W. Swanson, G.A. Schwind, A review of the cold-field electron cathode, in *Cold Field Emission and the Scanning Transmission Electron Microscope*, ed. by P. Hawkes. Adv. Imaging Electron Phys. **159**, 63–100 (2009)
17. R.G. Forbes, The theory of bright field electron and field ion emission sources, in *Nanfabrication Using Focused Ion and Electron Beams*, ed. by I. Utke, S. Moshkalev, P. Russell (Oxford University Press, Oxford, 2012)
18. Z.-B. Li, Density functional theory for field emission from carbon nanostructures. Ultramicroscopy **159**, 162–172 (2015)
19. V. Filip, L.D. Filip, H. Wong, Review on peculiar issues of field emission in vacuum nanoelectronic devices. Solid State Electron. **138**, 3–15 (2017)
20. R.G. Forbes, J.H.B. Deane, A. Fischer, M.S. Mousa, Fowler-Nordheim plot analysis: a progress report. Jordan J. Phys. **8**, 125–147 (2015). arXiv:1504.06134v7
21. R.G. Forbes, J.H.B. Deane, A.G. Kolosko, S.V. Fillipov, E.O. Popov, Reinvigorating our approach to field emission area extraction (because Murphy-Good plots are better than Fowler-Nordheim plots), in 32nd International Vacuum Nanoelectronics Conference & 12th International Vacuum Electron Sources Conference, Cincinnatti, July 2019. Technical Digest, p. 23. https://doi.org/10.13140/rg.2.2.32112.81927
22. Bureau International des Poid et Mesures (BIPM), *The International System of Units*, 9th ed. ("The SI Brochure") (BIPM, Sèvres, 2019). http://www.bipm.org/en/publications/si-brochure/
23. International Standards Organization (ISO), International Standard ISO 80000-1:2009. *Quantities and units. Part I: General* (ISO, Geneva, 2009, corrected 2011)
24. R. Gomer, *Field Emission and Field Ionization* (Harvard University Press, Cambridge, MA, 1961; Reprinted by: AIP, New York, 1993)
25. Adapted from an earlier version of, and consistent with, Reference [23], Section 3.20, Note 2
26. H. Yanagisawa, C. Hafner, P. Dona, M. Klöckner, D. Leuenberger, T. Greber, J. Osterwalder, M. Hengsberger, Laser-induced field emission from a tungsten tip: optical control of emission sites and the emission process. Phys. Rev. B **81**, 115429 (2010)
27. P. Hommelhoff, M.F. Kling, *Attosecond Nanophysics* (Wiley, Weinheim, 2015)
28. H.R. Reiss, Unsuitability of the Keldysh parameter for laser fields. Phys. Rev. A **82**, 023418 (2010)

29. J.D. Jackson, *Classical Electrodynamics*, 3rd edn. (Wiley, 1999)
30. L.D Landau, E.M. Lifshitz, L.P. Pitaevskii, *Electrodynamics of Continuous Media*, 2nd edn. (English translation published by: Butterworth-Heinemann, Oxford, reprinted 1995)
31. W.P. Dyke, J.K. Trolan, W.W. Dolan, G. Barnes, The field emitter: fabrication, electron microscopy, and electric field calculations. J. Appl. Phys. **24**, 570–576 (1953)
32. J.J. Thomson, Cathode rays. Phil. Mag. **44**, 293–316 (1897)
33. T.E. Stern, B.S. Gossling, R.H. Fowler, Further studies in the emission of electrons from cold metals. Proc. R. Soc. Lond. A **124**, 699–723 (1929)
34. R.H. Fowler, L. Nordheim, Electron emission in intense electric fields. Proc. R. Soc. Lond. A **119**, 173–181 (1928)
35. R.G. Forbes, J.H.B. Deane, Transmission coefficients for the exact triangular barrier: an exact general analytical theory that can replace Fowler & Nordheim's 1928 theory. Proc. R. Soc. Lond. A **467**, 2927–2947 (2011). See electronic supplementary material for details of universal constants used in field emission
36. A. Modinos, Theoretical analysis of field emission data. Solid State Electron. **45**, 809–816 (2001)
37. B. Lepetit, Electronic field emission models beyond the Fowler-Nordheim one. J. Appl. Phys. **122**, 215105 (2017)
38. B. Lepetit, A three-dimensional quantum-mechanical model of electronic field emission from metallic surfaces with nanoscale corrugation. J. Appl. Phys. **125**, 025107 (2019)
39. C. Herring, M.H. Nichols, Thermionic emission. Rev. Mod. Phys. **21**, 185–270 (1949)
40. A. Sommerfeld, Zur Elektronentheorie der Metalle. Naturwiss. **15**, 825–832 (1927)
41. A. Sommerfeld, H. Bethe, Electronentheorie der Metalle, in *Handbuch der Physik*, vol. 24, ed. by H. Geiger, H. Scheel (Berlin, Springer, 1933), pp. 333–675
42. A. Sommerfeld, *Thermodynamics and Statistical Mechanics* (Academic, London, 1964)
43. R.G. Forbes, Comparison of the Lepetit field emission current-density calculations with the Modinos-Forbes uncertainty limits, in 31st International Vacuum Nanoelectronics Conference, Kyoto, July 2018. Technical Digest, pp. 126–127. https://doi.org/10.13140/rg.2.2.17153.10080
44. D. Biswas, R. Kumar, Validation of current formula for a metallic nanotipped emitter. J. Vac. Sci. Technol. B **37**, 049603 (2019)
45. E.O. Popov, A.G. Kolosko, S.V. Fillipov, R.G. Forbes, Emission area extraction for needle-shaped and post-shaped emitters, in 32nd International Vacuum Nanoelectronics Conference & 12th International Vacuum Electron Sources Conference, Cincinnatti, July 2019. Technical Digest, p. 96. https://doi.org/10.13140/rg.2.2.35337.19041
46. I.S. Newton, *Philosophiæ Naturalis Principia Mathematica* (Royal Society, London, 1676) (in Latin) (English edition re-published by: Snowball Publishing, 2010). See Axioms, Law II (p. 19)
47. K.T. Compton, I. Langmuir, Electrical discharges in gases: part I. Survey of fundamental processes. Rev. Mod. Phys. **2**, 124–242 (1930)
48. W. Thomson, On electrical images, Report of the 17th Meeting of the British Association for the Advancement of Science, Oxford, June 1847, Transactions of the Sections pp. 6–7 (John Murray, London, 1848)
49. W. Thomson, On the mathematical theory of electricity in equilibrium: V—effects of electrical influence on internal spherical, and on plane conducting surfaces. Camb. Dublin Math. J. **4**, 276–284 (1849)
50. W. Thomson, *Reprint of Papers on Electricity and Magnetism* (Macmillan, London, 1872)
51. J.C. Maxwell, *A Treatise on Electricity and Magnetism*, 1st edn. (Clarendon, Oxford, 1873)
52. W. Schottky, Über den Einfluss von Strukturwirkungen, besonders der Thomsonschen Bildkft, auf die Electronenemission der Metalle. Phys. Z. **15**, 872–878 (1914). (in German)
53. J.J. Thomson, *Conduction of Electricity through Gases*, 1st edn. (Cambridge University Press, 1903)
54. R.F. Earhart, XI. The sparking distances between plates for small distances. Phil. Mag. **1**, 147–159 (1901)

55. W. Schottky, Über kalte und warme Electronenentladungen. Z. Phys. **14**, 63–106 (1923)
56. L.W. Nordheim, The effect of the image force on the emission and reflexion of electrons by metals. Proc. R. Soc. Lond. A **121**, 626–639 (1928)
57. J.H.B. Deane, R.G. Forbes, The formal derivation of an exact series expansion for the Principal Schottky-Nordheim Barrier Function v, using the Gauss hypergeometric differential equation. J. Phys. A Math. Theor. **41**, 395301 (2008)
58. R.A. Millikan, C.C. Lauritsen, Relations of field-currents to thermionic-currents. Proc. Nat. Acad. Sci. **14**, 45–49 (1928)
59. L.D. Landau, E.M. Lifschitz, *Quantum Mechanics* (Pergamon, Oxford, 1958)
60. R.H. Fowler, The restored theory of metals and thermionic formulae. Proc. R. Soc. Lond. A **117**, 549–552 (1928)
61. E.L. Murphy, R.H. Good, Thermionic emission, field emission and the transition region. Phys. Rev. **102**, 1464–1473 (1956)
62. E.H. Delaval, J. Canton, A letter from John Canton to Benjamin Franklin, containing some remarks on Mr Delaval's electrical experiments. Phil. Trans. **52**, 457–461 (1761–1762)
63. J. Priestly, *The History and Present State of Electricity, with Original Experiments* (1st edition 1775; 5th edition, corrected, published by: Johnson & Rivington, London, 1794) (5th edition reprinted by: ECCO Print Editions)
64. F. Guthrie, *Magnetism and Electricity* (Collins, London, 1876)
65. O.W. Richardson, *The Emission of Electricity from Hot Bodies* (Longmans Green, London, 1st edition 1916, 2nd edition 1921) (2nd edition republished as: *Thermionic Emission from Hot Bodies*, Watchmaker Publishing, online, 2003)
66. A.L. Reimann, *Thermionic Emission* (Chapman & Hall, London, 1934) (Reprinted by: Facsimile Publisher, Delhi, 2015)
67. T.A. Edison, Electrical indicator, US Patent Specification No. 307031 (Filed: 15 Nov. 1883; granted 21 Oct. 1884)
68. G.J. Stoney, Of the "electron", or atom of electricity. Phil. Mag. (Series 5) **38**, 418–420 (1894)
69. S. Dushman, Thermionic Emission. Rev. Mod. Phys. **2**, 381–476 (1930)
70. J.A. Becker, Thermionic electron emission and adsorption: part I. Thermionic emission. Rev. Mod. Phys. **7**, 95–128 (1935)
71. W.B. Nottingham, Thermionic emission, in: S. Flugge (ed.) *Encyclopedia of Physics, Vol. 21: Electron Emission Gas Discharges I* (Springer, Berlin, 1956); based on technical report of same title available on line via websearch
72. S. Yamamato, Fundamental physics of vacuum electron sources. Rept. Prog. Phys. **69**, 181–232 (2006)
73. J.A. Eichmeier, M. Thumm (eds.), *Vacuum Electronics* (Springer, Berlin, 2008)
74. O.W. Richardson, On the negative radiation from hot platinum. Proc. Camb. Phil. Soc. **11**, 286–295 (1900–1902)
75. O.W. Richardson, The emission of electrons from tungsten at high temperatures: an experimental proof that the electric current in metals is carried by electrons. Phil. Mag. **26**, 345–350 (1913)
76. O.W. Richardson, Some applications of the electron theory of matter. Phil. Mag. **23**, 594–627 (1912)
77. S. Dushman, Electron emission from metals as a function of temperature. Phys. Rev. **21**, 623–636 (1923)
78. A. Sommerfeld, Zur Elektronentheorie der Metalle auf Grund der Fermischen Statistik. Z. Phys. **47**, 1–32 (1928). (in German)
79. G.E. Uhlenbeck, S. Goudsmit, Ersetzung der Hypothese vom unmechanischen Zwang durch eine Forderung bezüglich des inneren Verhaltens jedes einzelnen Elektrons. Naturwiss. **13**, 953–954 (1925). (in German)
80. R.H. Fowler, The thermionic emission constant *A*. Proc. R. Soc. Lond. A **122**, 36–48 (1929)
81. K.L. Jensen, Exchange-correlation, dipole and image-charge potentials for electron sources: temperature and field variations of the barrier height. J. Appl. Phys. **85**, 2667–2680 (1999)

82. R.G. Forbes, Comments on the continuing widespread and unnecessary use of a defective emission equation in field emission related literature. J. Appl. Phys. **126**, 210191 (2019)
83. R.G. Forbes, Use of energy-space diagrams in free-electron models of field electron emission. Surf. Interface Anal. **36**, 395–401 (2004)
84. E. Guth, C.J. Mullin, Electron emission of metals in electric fields: III. The transition from thermionic to cold emission. Phys. Rev. **61**, 339–348 (1942)
85. R.E. Burgess, H. Kroemer, J.M. Houston, Corrected values of Fowler-Nordheim field emission functions $v(y)$ and $s(y)$. Phys. Rev. **90**, 515 (1953)
86. R.G. Forbes, J.H.B. Deane, Reformulation of the standard theory of Fowler-Nordheim tunnelling and cold field electron emission. Proc. R. Soc. Lond. A **463**, 2907–2927 (2007)
87. R.G. Forbes, Field electron emission theory (October 2016), v2. arXiv:1801.08251v2
88. A. Kyritsakis, J.P. Xanthakis, Extension of the general thermal field equation for nanosized emitters. J. Appl. Phys. **119**, 045303 (2016)
89. A. Kyritsakis, F. Djurabekova, A general computational method for electron emission and thermal effects in field emitting nanotips. Comput. Mater. Sci. **128**, 15–21 (2017)
90. L.W. Swanson, A.E. Bell, Recent advances in field electron microscopy of metals. Adv. Electr. Electron Phys. **32**, 193–309 (1973)
91. A.S. Bahm, G.A. Schwind, L.W. Swanson, Range of validity of field emission equations. J. Vac. Sci. Technol. B **26**, 2080–2084 (2008)
92. K.L. Jensen, M. Cahay, General thermal-field emission equation. Appl. Phys. Lett. **88**, 154105 (2006)
93. K.L. Jensen, General formulation of thermal, field, and photoinduced electron emission. J. Appl. Phys. **102**, 024911 (2007)
94. K.L. Jensen, A reformulated general thermal-field emission equation. J. Appl. Phys. **126**, 065302 (2019)
95. X.Z. Qin, W.L. Wang, N.S. Xu, Z.B. Li, R.G. Forbes, Analytical treatment of cold field electron emission from a nanowall emitter, including quantum confinement effects. Proc. R. Soc. Lond. A **467**, 1029–1051 (2011)
96. A.A. Patterson, A.I. Akinwande, Elementary framework for cold field emission from quantum-confined, non-planar emitters. J. Appl. Phys. **117**, 174311 (2015)
97. P.E. Mason, F. Uhlig, V. Vanek, T. Buttersack, S. Bauerecker, P. Jungwirth, Coulomb explosion during the early stages of the reaction of alkali metals with water. Nat. Chem. **7**, 250–254 (2015)
98. R.G. Forbes, Simple derivation of the formula for Sommerfeld supply density used in electron-emission physics and limitations on its use. J. Vac. Sci. Technol. B **28**, 1326–1329 (2010)
99. R.H. Fowler, *Statistical Mechanics*, 2nd edn. (Cambridge University Press, Cambridge, 1936; re-issued by Cambridge University Press, New York, 2011)
100. N. Fröman, P.O. Fröman, *Physical Problems Solved by the Phase-Integral Method* (Cambridge University Press, Cambridge, 2002)
101. R.G. Forbes, Derivation "in SI units" of the Landau and Lifschitz formula for field ionization rate-constant. arXiv:1412.1821
102. A. Mayer, Exact solutions for the field electron emission achieved from a flat metal surface using the standard Fowler-Nordheim equation with a correction factor that accounts for the electric field, the work function and the Fermi energy of the emitter. J. Vac. Sci. Technol. B **29**, 021803 (2011)
103. N. Fröman, P.O. Fröman, *JWKB Approximation: Contributions to the Theory* (North-Holland, Amsterdam, 1965)
104. R.G. Forbes, On the need for a tunneling pre-factor in Fowler-Nordheim tunneling theory. J. Appl. Phys. **103**, 114911 (2008)
105. E.C. Kemble, A contribution to the theory of the B. W. K. method. Phys. Rev. **48**, 549–561 (1935)
106. E.C. Kemble, *The Fundamental Principles of Quantum Mechanics with Elementary Applications* (Dover, New York, 1937)

107. S.C. Miller, R.H. Good, A WKB-type approximation to the Schrödinger equation. Phys. Rev. **91**, 174–179 (1953)
108. M. Razavy, *Quantum Theory of Tunnelling* (World Scientific, Hackensack, 2003)
109. R.G. Forbes, Development of a simple quantitative test for lack of field emission orthodoxy. Proc. R. Soc. Lond. A **469**, 20130271 (2013)
110. R.G. Forbes, The Murphy-Good plot: a better method of analysing field emission data. Roy. Soc. Open. Sci. **6**, 190212 (2019)
111. R.G. Forbes, Field emission: new theory for the derivation of emission area from a Fowler-Nordheim plot. J. Vac. Sci. Technol. B **17**, 525–533 (1999)
112. See p. 357 in Reference [99]
113. W.P. Dyke, W.W. Dolan, Field emission. Adv. Electr. Electron Phys. **8**, 89–185 (1956)
114. R.G. Forbes, A. Fischer, M.S. Mousa, Improved approach to Fowler-Nordheim plot analysis. J. Vac. Sci. Technol. B **31**, 02B103 (2013)
115. R.G. Forbes, Simple good approximations for the special elliptic functions in standard Fowler-Nordheim tunnelling theory for a Schottky-Nordheim barrier. Appl. Phys. Lett. **89**, 113122 (2006)
116. R.G. Forbes, Use of a spreadsheet for Fowler-Nordheim equation calculations. J. Vac. Sci. Technol. B **17**, 534–541 (1999)
117. R.G. Forbes, J.H.B. Deane, Comparison of approximations for the principal Schottky-Nordheim barrier function $v(f)$, and comments on Fowler-Nordheim plots. J. Vac. Sci. Technol. B **28**, C2A33–C2A42 (2010)
118. B.C. Carlson, Elliptic integrals, in *NIST Handbook of Mathematical Functions*, ed. by F.W.J. Olver et al. (NIST & Cambridge University Press, 2010), see formula 19.6.1
119. R.G. Forbes, ISQ derivation of the Murphy-Good second-order Taylor-expansion correction factor $t_2(f)$, in 32nd International Vacuum Nanoelectronics Conference & 12th International Vacuum Electron Sources Conference, Cincinnatti, July 2019. Technical Digest, p. 101
120. R.G. Forbes, C.J. Edgcombe, U. Valdré, Some comments on models for field enhancement. Ultramicroscopy **95**, 57–65 (2003)
121. A.I. Zhbanov, E.G. Pogolerov, Y.-C. Chang, Y.-G. Lee, Screened field enhancement factor for the floating sphere model of a carbon nanotube array. J. Appl. Phys. **110**, 114311 (2011)
122. J.R. Harris, K.L. Jensen, D.A. Shiffler, Dependence of optimal spacing on applied field in ungated field emitter arrays. AIP Adv. **5**, 087183 (2015)
123. R.G. Forbes, Physical electrostatics of small field emitter arrays/clusters. J. Appl. Phys. **120**, l054302 (2016)
124. T. de Assis, F.F. Dall'Agnol, Evidence of universal inverse-third power law for the shielding-induced fractional decreasee in apex field enhancement factor at large spacings: a response via accurate Laplace-type calculations. J. Phys.: Condens. Matter **30**, 195301 (2018)
125. D. Biswas, R. Rudra, Shielding effects in random large-area emitters, the field enhancement factor distribution, and current calculation. Physics Plasmas **26**, 083105 (2018)
126. E. Minoux, O. Groening, K.B.K. Teo et al., Achieving high-current carbon nanotube emitters. Nano Lett. **5**, 2135 (2005)
127. R.G. Forbes, Field emission: the theoretical link between voltage loss, reduction in field enhancement factor, and Fowler-Nordheim-plot saturation. Appl. Phys. Lett. **110**, 133109 (2017)
128. R.G. Forbes, Use of the concept "area efficiency of emission" in equations describing field emission from large area electron sources. J. Vac. Sci. Technol. **B27**, 1200–1203 (2009)
129. R.G. Forbes, Extraction of emission parameters for large-area field emitters, using a technically complete Fowler-Nordheim-type equation. Nanotechnology **23**, 095706 (2012)
130. R.G. Forbes, Why converting field emission voltages to macroscopic fields before making a Fowler-Nordheim plot has often led to spurious characterization results. J. Vac. Sci. Technol. B **37**, 051802 (2019)
131. M.M. Allaham, R.G. Forbes, M.S. Mousa, Applying the field emission orthodoxy test to Murphy-Good plots. Jordan J. Phys. 13, 101–110 (2020)

132. M.M. Allaham, R.G. Forbes, A. Knápek, M.S. Mousa, Implementation of the orthodoxy test as a validity check on experimental field emission data. J. Electr. Eng. Slovak 71, 37–42 (2020)
133. M. Bachmann, F. Dams, F. Düsberg, M. Hofmann, A. Pahlke, Extraction of the characteristics of current-limiting elements from field emission measurement data. J. Vac. Sci. Technol. B **35**, 02C103 (2017)