Scaling properties of a non-Fowler–Nordheim tunnelling junction

A. Kyritsakis[1], J. P. Xanthakis[1] and D. Pescia[2]

[1]Department of Electrical and Computer Engineering, National Technical University of Athens, Zografou Campus, Athens 15700, Greece
[2]Laboratory for Solid State Physics, ETH Zurich, Zürich 8093, Switzerland

Research

CrossMark
click for updates

We have theoretically explained the experimentally observed scaling properties of the current–voltage (I–V) characteristics of a field-emission tunnelling diode with respect to the tip–anode distance d. All the I–V curves for different d-values collapse onto a single curve by a scaling transformation that keeps the electrical field at a given direction constant for different d. Our proof is applicable to more than just the obvious case where the electrostatic potential varies linearly with the distance from the cathode x (and the Fowler–Nordheim plot is also linear). It applies to any general nonlinear potential encountered in emitting tips of small radii of curvature R. Furthermore, we explain why the scaling property is excellent at $d \gg R$, but deteriorates when $d \sim R$. The scaling property is shown to derive from the simultaneous action of two factors: (i) in a Taylor expansion of the potential in the tunnelling region, the second-order coefficient is found to be proportional to the first-order coefficient and their ratio independent of d. (ii) The angular variation of the electric field is independent of d for $d \gg R$. Deviations from scaling at $d \sim R$ are attributed to both the dependence on d of the angular variation of the electrical field and/or the presence of a third-order term.

1. Introduction

Very recently, in Cabrera *et al.* [1], it has been shown that the experimental current–voltage (I–V) characteristics of a tunnelling diode are invariant with respect to the tip–anode distance d under a scaling transformation $R_{exp}(d)$ of the applied voltage, $V \rightarrow V R_{exp}(d)$. The invariance

Royal Society **Publishing**

2

with d was observed over a range of several orders of magnitude for d, from millimetres to several nanometres, but it began failing when d was only a few nanometres. At the outset of this work, it must be said that if the emitting surface is either planar or of large radius of curvature R, then the interpretation of the above observations is elementary. For emitting surfaces of large radii of curvature R the traditional Fowler–Nordheim (FN) equation is valid [2,3]. This equation contains only one physical variable, the electric field, and by scaling the voltage, one simply scales the electric field. Said alternatively, the electrostatic potential is adequately represented by the one-dimensional form $U(x) = W - Fx$, where W is the work function and F is the local electric field and hence scaling V, so that F remains the same is all that is needed.

However, the experiments in reference [1] were performed using tips of radii 4–6 nm. In such cases, the electrostatic potential is far from being linear with x [4,5], and, furthermore, the variation of the latter with the angle θ to the tip axis is quite significant. In short, the tunnelling potential U is three-dimensional in nature with the electric field different at different points so that—at first sight—there is no scalar quantity that can rescale the entire potential with respect to d and keep the current constant. Note that in tunnelling, the current density at any given point is a functional of the whole potential. In this paper, we show that if $d \gg R$, then there is, indeed, such a scaling parameter as experimentally observed. We present both numerical and algebraic evidence to this effect.

2. The potential: numerical calculation

The emitting tips used in reference [1] were macroscopically conical with a spherically rounded apex, see the inset of figure 1. For the purpose of numerical calculation, we have simulated this shape with a vertical stack of spheres of increasing diameter as shown again in figure 1. However, the algebraic–analytic arguments we will give do not need or necessitate this simulation which has been constructed only for the purpose of facilitating these calculations. We proceed with these first. As with our previous calculations [6,7], the general solution of the Laplace equation for this composite body will be the superposition of the solutions for each separate sphere (hence the choice for the construction simulating the tip). Then, we can write

$$\Phi(r, \theta) = \sum_{i=1}^{N} \sum_{n=0}^{\infty} C_{in}(r_i^{(-n-1)} P_n(\cos \theta_i)). \tag{2.1}$$

In expression (2.1), Φ is the electrostatic potential, (r_i, θ_i) are the spherical coordinates of each point at a coordinate system having its centre at the centre of the ith sphere and P_n are the Legendre polynomials of the first kind. Note that the centre of the first sphere is the centre of the coordinate system. The coefficients C_{in} are determined by the boundary conditions that $\Phi = 0$ at the cathode and $\Phi = V$ at the anode, where V is the applied voltage to the diode. Henceforth, the term potential will refer to the function Φ, whereas the term voltage to the scalar quantity is V.

Alternatively, one can introduce fictitious image spheres with respect to the anode plane and consider only the boundary condition $\Phi = 0$ at the cathode where now

$$\Phi(r, \theta) = V + \sum_{i=1}^{N} \sum_{n=0}^{\infty} A_{in}(r_i^{(-n-1)} P_n(\cos \theta_i) - r_i^{m(-n-1)} P_n(-\cos \theta_i^m)), \tag{2.2}$$

with (r_i^m, θ_i^m) the spherical coordinates with respect to the centres of the image spheres.

To evaluate the new coefficients A_{in}, we truncate the infinite index n to, say, M and consider $M \times N$ points on the cathode, where $\Phi = 0$. $M = 12$ was found to be more than enough.

To this electrostatic potential, we now add the image potential for a spherical surface and the work function discontinuity W, so that finally the tunnelling potential U is

$$U(r, \theta) = W - e\Phi(r, \theta) - \frac{e^2}{8\pi \varepsilon_0} \frac{R}{r^2 - R^2}. \tag{2.3}$$
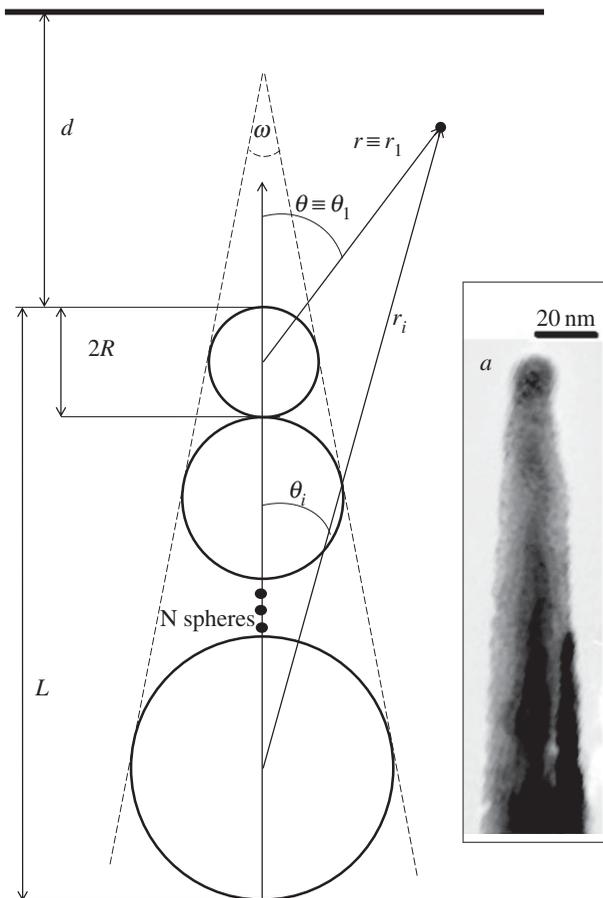
Figure 1. Geometrical model and notation. In the inset, a transmission electron microscopy image of one of the tips used in [1] is shown.

Figure 2a gives our calculated electrostatic potentials Φ as a function of $x = r - R$ for $\theta = 0$ at a constant applied voltage $V = 11.43$ V for various d-values. This calculation, such as all the calculations of this paper until §5, is performed with a conical cathode of radius $R = 4$ nm, full angle of planar section of cone $\omega = 7°$ and total length $L = 1.2456$ μm. The value of V has been chosen, so that the normal electric field at the apex of the $d_0 = 5$ nm curve is equal to $F(d_0) = 4$ V nm^{-1}. Now, we apply our scaling rule: we multiply V by a scaling factor $S(d)$, so that the electric field at the apex at each different d is equal to $F(d_0)$. The relation with the R_{exp} of [1] will become clear below. Our scaling rule is obviously

$$V'(d) = S(d)V = \frac{F(d_0)}{F(d)} V, \tag{2.4}$$

where $F(d)$ is the calculated electric field at the apex *before* scaling. We note that the idea of normalizing the potential distribution with respect to the apex field has been used previously [8]. The results of this scaling procedure are shown in figure 2b. We can see that all curves collapse onto the $d_0 = 5$ nm curve over a region of x that

1. is certainly wider than the region of linearity of Φ, and
2. is much wider than the tunnelling distance (usually 1–1.5 nm). Note that when the radius of curvature R is of this magnitude (less than 20 nm), the potential in the tunnelling region
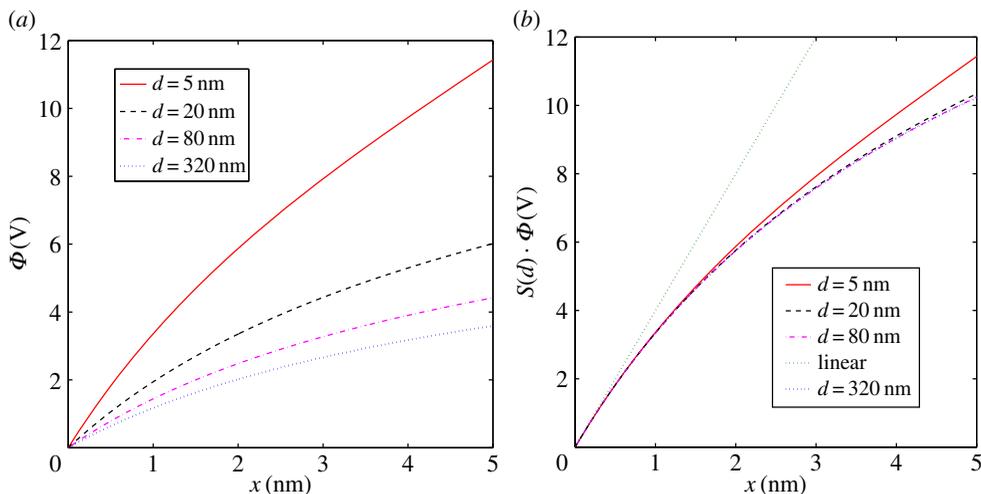
Figure 2. Calculated electrostatic potential along the vertical axis for various distances: (a) before scaling, i.e. with the same applied voltage $V = 11.43$ V and (b) after scaling, i.e. with applied voltage $V' = S(d)V$. A linear potential is also included only to show the curvature of the calculated potential. (Online version in colour.)
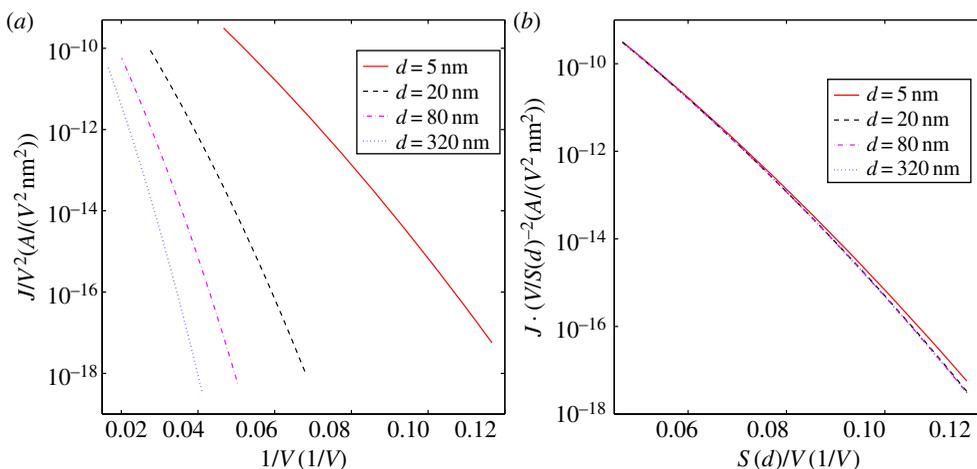
Figure 3. Current density versus voltage in the form of FN plots for various distances d: (a) before scaling and (b) after scaling. (Online version in colour.)

is no longer linear with x. Furthermore, the collapsing is very good when $d \gg R$ and conversely deteriorates as the tip approaches the anode.

It is important to stress that although the above scaling rule was applied for $V = 11.43$ V it holds true for any value of V, i.e. $S(d)$ does not depend on V.

Now if the potential becomes invariant under this scaling rule, then the current density should also become invariant under a similar scaling by virtue of the Wentzel–Krammers–Brillouin (WKB) approximation (and not by virtue of the FN theory). In figure 3a, we show the various $J_0(V, d)$ curves—in the form of FN plots—where J_0 is the current density along the $\theta = 0$ direction at voltage V at distance d of the tip from the anode. Again, we emphasize that these plots were obtained by the one-dimensional WKB approximation [2–4].

If we now plot in figure 3b these curves (again in FN form) in terms of $V/S(d)$, it can be seen that all the curves of figure 3a from $d = 5$ to $d = 320$ nm collapse again onto a single curve.

The reason that the voltage is divided and not multiplied by $S(d)$ this time is the following: in figure 3a, we get the same J_0 for a series of values of V, each one for each d. According to our previous arguments/results of figure 2, the voltages at which we obtain the same current at each d can be obtained by multiplying the voltage of d_0 by $S(d)$, i.e. $V(J_0, d) = S(d)V(J_0, d_0)$. Thus, if we divide every V value at each d-curve by $S(d)$, then we should obtain the same $d = d_0$ curve. In other words, all curves should collapse onto the $d = d_0$ curve when J_0 is plotted against the scaled variable $V/S(d)$ as indeed is shown in figure 3b. A very small deviation from collapsing appearing for small d and very small voltages is attributed to the fact that when the electric field becomes very small, the forbidden region extends to points where collapsing of the electrostatic potential is not perfect as shown in figure 2b.

Now the potential variation with r at other θ directions follows exactly the same trend as the $\Phi(r, \theta = 0)$ variation, only now to obtain the collapsing of the $\Phi(r, \theta)$ curves with d we need to scale them with the ratio

$$S_\theta(d) = \frac{F(\theta, d_0)}{F(\theta, d)}, \tag{2.5}$$

where $F(\theta, d)$ is the electric field on the emitter along the direction θ when the tip is at distance d from the anode. Likewise, the FN plots $J_\theta(V)$ collapse onto a single curve when they are scaled by the factor $1/S_\theta(d)$. This is shown in figure 3a,b. Given that the experimentally measured quantity is the current which is a surface integral of the current density, and given that $F(\theta, d)$ changes differently with θ at each d it is not evident how the experimentally observed scaling can appear. We tackle this problem immediately below, but before we do this, we want to use some general argument—as already stated—to show that this scaling behaviour of the potential depends on the properties of the nonlinear part of the potential Φ and is not particularly related to the conical shape of the tip we have used.

3. The potential: analytic approximation

At the beginning of this section, we assume that the shape of the tip in the region where the current is emitted is spherical. Later, we generalize to any smooth shape. Note that this assumption is distinct from the 'floating sphere' equipotential model of the cathode [9]. The shape of the emitter below the apex area may be, here, arbitrary. Now, we want a simpler and at the same time more general expression for the potential Φ than that of equation (2.3). Because the current is determined by the values of Φ only inside the tunnelling region which is small (1–1.5 nm), it is worthwhile to make a Taylor expansion for any θ_0 direction, including up to the second order to take account of the nonlinearities of Φ introduced by the Legendre polynomials or other functions. We write

$$\Phi(r, \theta_0) = \Phi(R, \theta_0) + \left[\frac{\partial \Phi}{\partial r}\right]_{(R, \theta_0)} \cdot (r - R) + \left[\frac{\partial^2 \Phi}{\partial r^2}\right]_{(R, \theta_0)} \cdot \frac{(r - R)^2}{2} + \cdots \tag{3.1}$$

The Laplace equation in spherical coordinates, for an object with rotational symmetry, takes the form

$$\frac{\partial^2 \Phi}{\partial r^2} + \frac{2}{r}\frac{\partial \Phi}{\partial r} + \frac{\cot(\theta)}{r^2}\frac{\partial \Phi}{\partial \theta} + \frac{1}{r^2}\frac{\partial^2 \Phi}{\partial \theta^2} = 0. \tag{3.2}$$

But, the emitting surface is an equipotential, so that $\partial \Phi/\partial \theta = \partial^2 \Phi/\partial \theta^2 = 0$ on the surface and hence equation (3.2) becomes

$$\left[\frac{\partial^2 \Phi}{\partial r^2}\right]_{(R, \theta_0)} = -\frac{2}{R}\left[\frac{\partial \Phi}{\partial r}\right]_{(R, \theta_0)}. \tag{3.3}$$

Equation (3.3) tells us that if we scale the electric field at the emitting surface F in any direction θ_0, then the second-order derivative in the same direction scales as well, or—in other words—their ratio is independent of the tip–anode distance d. However—as already noted—the above

relation which is in accordance with our previous numerical findings of figure 2 on the scaling of $\Phi(r, \theta = \theta_0)$ is not sufficient to guarantee scaling of the whole potential (and hence of the current), because the angular dependence of $F(\theta, d)$ is not independent of d.

The third (or higher)-order terms are, in general, dependent on d, but they are very small and except for extreme cases can be neglected. However, they produce small deviations from scaling observed both in figures 2 and 3. Small deviations from scaling of the potential when $x > 2$ nm (we remind the reader that, from §2, $x = r - R$) can be seen in figure 2b for small distances (comparable to R). In figure 3b, we observe the same deviations in current density again for small distances and for very small current in which cases the forbidden region extends up to 2–3 nm from the tip.

The above analysis can be extended to non-spherical surfaces. Actually, if we assume that an arbitrary system of curvilinear coordinates (u, v, φ) in which the emitting part of the tip is described by $u = u_0 = $ constant, then the above analysis of the Laplace equation can be very easily extended to this kind of surfaces. The Laplace equation in curvilinear coordinates with rotational symmetry can be expressed as

$$\frac{1}{h_u h_v h_\varphi} \frac{\partial}{\partial u} \left(\frac{h_u h_v h_\varphi}{h_u^2} \frac{\partial \Phi}{\partial u} \right) + \frac{1}{h_u h_v h_\varphi} \frac{\partial}{\partial v} \left(\frac{h_u h_v h_\varphi}{h_v^2} \frac{\partial \Phi}{\partial v} \right) = 0, \tag{3.4}$$

where h_i are the metric factors (they are functions of u, v, φ). The derivatives with respect to v are zero on the surface and hence the second term in equation (3.4) is zero. Thus, equation (3.4) becomes

$$\frac{1}{h_u h_v h_\varphi} \left[\frac{\partial \Phi}{\partial u} \frac{\partial}{\partial u} \left(\frac{h_u h_v h_\varphi}{h_u^2} \right) + \left(\frac{h_u h_v h_\varphi}{h_u^2} \right) \frac{\partial^2 \Phi}{\partial u^2} \right] = 0.$$

Reordering the above, we obtain

$$\left. \frac{\partial^2 \Phi}{\partial u^2} \right|_{(u_0, v)} = f(u_0, v) \left. \frac{\partial \Phi}{\partial u} \right|_{(u_0, v)}. \tag{3.5}$$

The above means that if we expand Φ near the surface point (u_0, v_0), along a line $v = v_0$ (equivalent to a radial line in the spherical case), the second-order coefficient of the Taylor expansion along this line will be proportional to the first-order coefficient and their ratio independent of d. Hence, if the potential is rescaled in order to keep the same first derivative at all distances d, the second order will also scale.

A note on the transmission coefficient T is worthy here. T within the WKB approximation is a line integral along the most probable path of the electron in the tunnelling region. In one dimension, this is necessarily a straight line segment, whereas in three dimensions, it is not always straight [10–12]. However, the curvilinear lines $v = $ constant are a better approximation to it compared with the straight lines segments [10].

4. The field θ-dependence

We have till now proved that the ratio of the first-order derivative (electric field) to the second-order derivative at the emitting surface is independent of d for all θ. But, both of them are individually dependent on θ. So, how is this fact consistent with scaling? Here, we can use the experimental observation along with our numerical findings of §3 that scaling with d works very well when $d \gg R$ and deteriorates as d diminishes. When $d \gg R$, the way the electric field changes with θ is unique and independent of d, i.e. $F(\theta, d)$ is no longer a function of d. This is shown in figure 4 where we plot $F(\theta)/F(0)$ as a function of θ for several d. It can be observed that as d increases the curves tend to collapse on each other. Hence, by virtue of equation (3.3), scaling only the applied voltage V scales automatically $F(0)$ and $F'(0)$ but also the electric field and its derivative at any other angle θ. Then, the whole potential to second order remains invariant.

Yet, the degree of collapsing of the curves in figure 4 is not complete except at very large d-values. In view of this, one would not expect such good scaling of the total current as obtained in reference [1] or as will be shown later in our results. The answer to this apparent contradiction
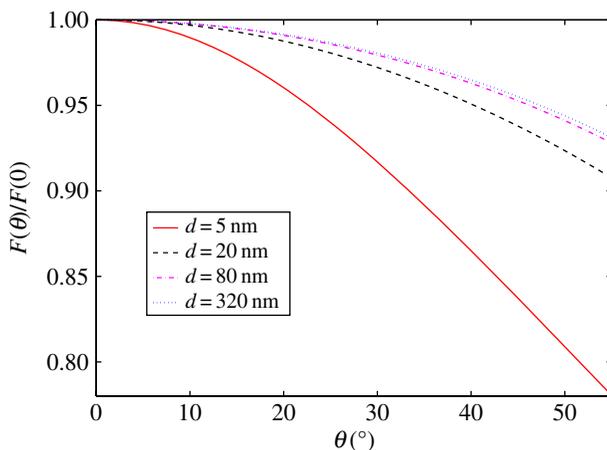
Figure 4. Ratio of electric field at angle θ to the vertical one ($\theta = 0$) as a function of the angle θ for various distances d. (Online version in colour.)
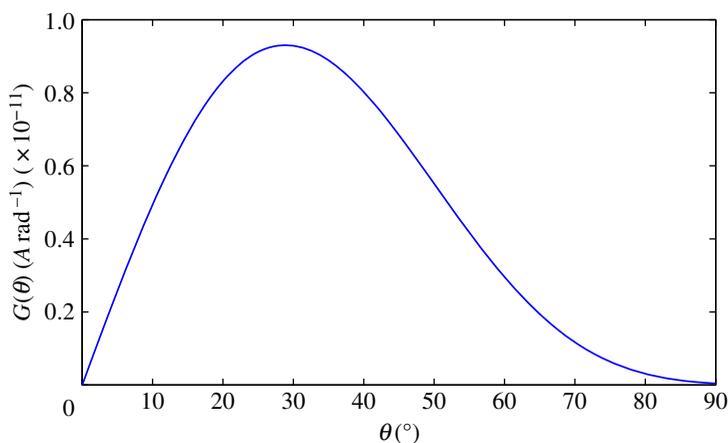

Figure 5. Current per unit angle emitted between θ and $\theta + d\theta$ as a function of θ for $d = 50$ nm and $V = 26$ V. The maximum is found at $30°$. (Online version in colour.)

lies in the fact that only a small range of values of θ contribute to the evaluation of the current I. This is shown immediately below.

Figure 5 gives the emitted current per unit angle between θ and $\theta + d\theta$, i.e. $G(\theta) = 2\pi R^2 J(\theta) \sin(\theta)$ (amps per rad). It is clear that the maximum of this quantity occurs not at $\theta = 0$ (where there is maximum F but zero area), but at $\theta = \theta_{max} = 30°$ (actually θ_{max} is d-dependent at small d). Note that $J(0) = 0.3$ pA nm^{-2}, whereas $J(\theta_{max} = 30°) = 0.19$ pA nm^{-2}. From figure 5, we conclude that the total current is determined by the θ values around θ_{max}. At these values, $F(\theta, d)/F(0, d)$ is almost independent of d.

5. The current-comparison with experiment

Finally, the total current as a function of the applied voltage V is shown in figure 6 as an FN plot for different d-values (a) before and (b) after the application of the scaling rule of equation (2.5) with θ chosen to be θ_{max}. It is clear that there is a very good degree of scaling analogous to that observed experimentally.
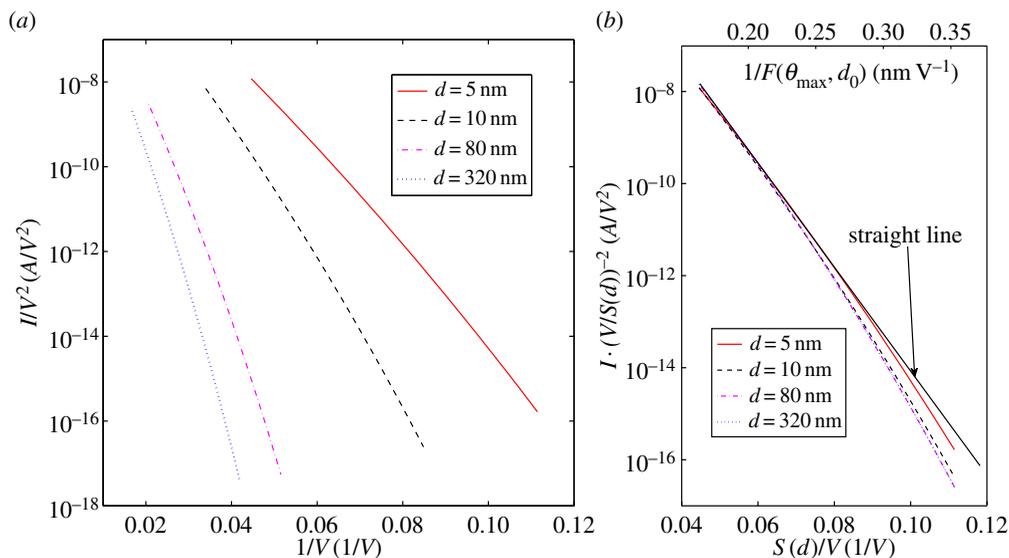
Figure 6. Total current as a function of the applied voltage for various distances d presented as an FN plot (a) before scaling and (b) after scaling. A straight line with the same slope as the initial slope of the diagrams is drawn in (b) to show the curving of the FN plots. At the top axis of the diagram (b), the common electric field at angle $\theta_{max} = 30°$ is shown. (Online version in colour.)

Furthermore, and most importantly, the FN plots collapse while being curved and not straight owing to the nonlinearity of the potential. If the FN plots were straight lines, then there would be nothing to explain as clearly stated in §1. To make this curvature clear, we have drawn the tangent at the highest point of the curves.

The deviations from scaling observed in figure 6 at small d are due to two factors. First, at small distances, the third-order coefficient of the Taylor expansion in equation (3.1) becomes significant and it contributes to the deviation of the electrostatic potential from the scaling behaviour as shown in figures 2 and 3. Note that the I–V characteristics of figure 6 reveal deviations from scaling that are bigger than those found in the J–V characteristics of figure 3. This is attributed to the loss of invariance of the $F(\theta, d)$ on d as shown in figure 4 when $d \sim R$.

Here, a comment is worthwhile regarding the scaling factor used here and that in the experiments in reference [1]. A careful reading of [1] reveals that the scaling factor used there $R(d)$ is associated with our $1/S_{\theta=0}(d)$. However, we have just shown that a better scaling factor (in the sense of producing more collapsing of the I–V curves for different d is the ratio of the electric fields at $\theta = \theta_{max}$). Therefore, we would assign

$$R_{exp}(d) = \frac{1}{S_{\theta_{max}}(d)} = \frac{F(\theta_{max}, d)}{F(\theta_{max}, d_0)}. \tag{5.1}$$

Obviously, the scaling variable $V/S_{\theta\,max}(d)$ is proportional to the electric field and that is why in figure 6b we have introduced $1/F(\theta_{max}, d_0)$ as a secondary x-axis at the top of the diagram.

6. The power law

In reference [1], apart from the collapsing of the curves under the scaling of the voltage, the following experimental observation has been made: the scaling factor $S(d)$ followed a power law of the form $S(d) = ad^\lambda$, where a is an arbitrary pre-factor, and λ is found experimentally to be positive, about 0.2. Our calculations confirm this observation for the intermediate scale of distances. This is shown in figure 7 where we plot in logarithmic scale the scaling function $S(d)$ as a function of d. The calculation is done for a conical-like stack of spheres with the radius of the
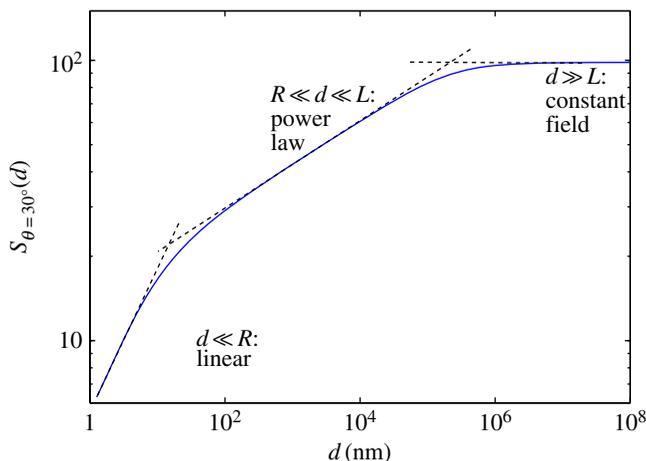
Figure 7. Voltage required to keep a constant field $F(\theta = 30°, d)$ as a function of d. A log–log scale is used. (Online version in colour.)

first sphere equal to $R = 5$ nm, full angle of planar section of cone $\omega = 10°$ and total length of the emitter $L = 326.4$ μm. We observe three regions:

(i) Very small distances $d \ll R$, where the field (and hence the scaling function) follows the parallel-plate capacitor law of $F = V/d$. In the figure, this corresponds to the straight line with slope equal to 1.
(ii) Intermediate distances $R \ll d \ll L$, where a power law is observed as a straight line with a slope of $\lambda = 0.153$.
(iii) Very large distances $d \gg L$, where the variation in d does not affect the electric field and the scaling function is a constant.

The exponent λ can be calculated for various configurations. Repeating the calculation of figure 7 for various conical geometries, we find that it depends only on the angle of aperture of the cone. Its numerical values are between 0.137 and 0.165 for apertures ω between 6° and 12°. The function $\lambda(\omega)$ seems to be increasing almost linearly. These values are in very good agreement with the theoretical analysis done in reference [1] (resulting values 0.14–0.17) on the exponent based on the solution of the Laplacian for a mathematically sharp conical emitter [13]. Our values of λ are also close to the experimental values which were found to be concentrated around 0.2 for apertures in the above range of values. The experimental value 0.2 is slightly higher than those given both by the theoretical analysis in reference [1] and our simulation. This small deviation could be attributed to the fact that the experimental tips have their angle increasing as one moves away from the apex, producing an effectively larger angle of aperture and hence a larger λ.

7. Conclusion

We have theoretically explained the scaling properties of the electrical characteristics of a field-emission tunnelling diode and its breakdown at small distances $d \sim R$. In doing so, we have discovered certain invariance properties of the tunnelling potential. These will be used in a forthcoming publication to derive a generalized FN equation.

References

1. Cabrera H *et al.* 2013 Scale invariance of a diode-like tunnel junction. *Phys. Rev. B* **87**, 115436. (doi:10.1103/PhysRevB.87.115436)

10

rspa.royalsocietypublishing.org *Proc. R. Soc. A* **470**: 20130795

2. Murphy EL, Good RH. 1956 Thermionic emission, field emission, and the transition region. *Phys. Rev.* **102**, 1464–1473. (doi:10.1103/PhysRev.102.1464)

3. Forbes RG. 2003 Use of energy-space diagrams in free-electron models of field electron emission. *Surf. Interface Anal.* **36**, 395–401. (doi:10.1002/sia.1900)

4. Cutler PH, He J, Miskovsky NM, Sullivan TE, Weiss B. 1993 Theory of electron emission in high fields from atomically sharp emitters: validity of the Fowler–Nordheim equation. *J. Vac. Sci. Technol. B* **11**, 387. (doi:10.1116/1.586866)

5. Fursey GN, Glazanov DV. 1998 Deviations from the Fowler–Nordheim theory and peculiarities of field electron emission from small-scale objects. *J. Vac. Sci. Technol. B* **16**, 910. (doi:10.1116/1.589929)

6. Xanthakis JP, Kokkorakis GC. 2007 Theoretical calculation of spatial variation of the transmission coefficient of closed carbon nanotubes. *Surf. Interface Anal.* **39**, 135–138. (doi:10.1002/sia.2476)

7. Kyritsakis A, Xanthakis JP. 2013 Beam spot diameter of the near-field scanning electron microscopy. *Ultramicroscopy* **125**, 24–28. (doi:10.1016/j.ultramic.2012.10.015)

8. Jensen KL. 2010 Space charge effects in field emission: three dimensional theory. *J. Appl. Phys.* **107**, 014905. (doi:10.1063/1.3272688)

9. Miller HC. 1967 Change in field intensification factor β of an electrode projection (whisker) at short gap lengths. *J. Appl. Phys.* **38**, 4501. (doi:10.1063/1.1709157)

10. Kapur PL, Peierls R. 1937 Penetration into potential barriers in several dimensions. *Proc. R. Soc. Lond. A* **163**, 606–610. (doi:10.1098/rspa.1937.0248)

11. Huang ZH, Feuchtwang TE, Cutler PH, Kazes E. 1990 Wentzel–Kramers–Brillouin method in multidimensional tunneling. *Phys. Rev. A* **41**, 32–41. (doi:10.1103/PhysRevA.41.32)

12. Das B, Mahanty J. 1987 Spatial distribution of tunnel current and application to scanning-tunneling microscopy: a semiclassical treatment. *Phys. Rev. B* **36**, 898–903. (doi:10.1103/PhysRevB.36.898)

13. Jackson JD. 1999 *Classical electrodynamics*, 3rd edn. New York, NY: Jon Wiley and Sons.