MODELING AND ELECTROSTATIC FOCUSING FOR A FIELD EMISSION ELECTRON SOURCE *

Vadim Jabotinski1, John Pasour, Khanh T. Nguyen1, John Petillo2, Baruch Levush, David Abe
Vacuum Electronics Branch, U.S. Naval Research Laboratory, Washington, DC 20375, USA

Abstract

This paper presents theory and analysis of single-tip field emission and electron beam propagation in the electrostatic focusing fields. It is shown that two gate apertures with a focusing anode allow transport of narrow electron beams over long distances without need for a confining magnetic field. Physical mechanisms of the beam formation, transport, field emission energy distributions, the effects of the emission properties, and parametric studies are discussed, emission current formula is derived, and new concept and model of the bandgap-spread multilevel field emission is given.

I. INTRODUCTION

Because of their small emitting area and high emission current density, tip-based field emitters produce beams of low intrinsic emittance and high-brightness. Most prospective applications, such as high-resolution X-ray imaging, free electron lasers, and high-frequency TWT and terahertz periodic structures, require a focused electron beam to be transported over distances many orders of magnitude larger than the typical field emitting geometry transverse dimensions.

This paper presents theory, simulations, and analysis that describe focusing and transport of the electron beam from a field emission tip in the electrostatic field produced by two gate apertures and a focusing anode. The idea of two gates was considered first by W. B. Hermannsfeldt [1] and further investigated by others [2-3]. However, the two-gate concept does not allow sufficiently long focused electron beams. We introduced a focusing anode into the geometry and obtained optimized configurations that provide the desired long aspect ratio focusing and beam transport. For more adequate description of the field emission, we derive an emission model with higher order corrections to the Fowler-Nordheim theory and suggest a new concept of bandgap-spread multilevel field emission.

II. FIELD EMISSION

The presence of electric field causes electrons to escape from a surface of the emitting matter. Such field emission of electrons can occur into vacuum or a medium and the emitting matter can be electrical conductor, semiconductor, dielectric, or more complex substance in the solid or other state. In solid-state electrical conductors, e.g. metals, the valence electrons possess the conduction energy band and are described by Sommerfeld free electron gas model with Fermi-Dirac statistics, which defines the electrons energy distribution. For the emission from not electrical conductors the Sommerfeld theory of metals with Fermi-Dirac statistics can still be applied as an approximation and with suitable modifications.

At sufficiently high temperatures the energy of thermal motion of the electrons allows them to overcome the binding potential barrier, described by the work function, thus producing the thermionic emission, which is a special case of the field emission. The Maxwell-Boltzmann statistics at such high temperatures determines the electrons energy distribution, which is also a special case of the Fermi-Dirac distribution. The Richardson-Dushman equation with the Schottky reduction of the effective work function describes the thermionic emission current density.

The high electric field triggers electron emission independently from the temperature. Such cold emission occurs due to quantum tunneling of electrons through the surface potential barrier. The electric field effects on the shape of the potential barrier by making it of a finite width, in the length units, reducing to zero with the higher electric field. Fowler and Nordheim derived a theory of the field emission in 1928 [4,5]. Burgess, Kroemer, and Houston in 1952 [6] corrected a mathematical error in the Nordheim analysis of the image charge effects. Murphy and Good in 1956 generalized the Fowler-Nordheim theory to describe the cold, thermionic, and intermediate emission regimes [7]. They also suggested an improved form for the Nordheim elliptic function and introduced a correction to the Fowler-Nordheim formula by adding into the analysis a linear term from the Taylor series expansion for the momentum integral determined by the Schottky-Nordheim tunneling potential barrier with the respect of the image charges.

* Work supported by DARPA/MTO and by ONR
1 e-mail: vadim.jabotinski.ctr@nrl.navy.mil
1 Beam-Wave Research, Bethesda, MD 20814
2 Scientific Applications Int’l Corp, Billerica, MA 01821
This paper presents theory and analysis of single-tip field emission and electron beam propagation in the electrostatic focusing fields. It is shown that two gate apertures with a focusing anode allow transport of narrow electron beams over long distances without need for a confining magnetic field. Physical mechanisms of the beam formation, transport, field emission energy distributions, the effects of the emission properties, and parametric studies are discussed, emission current formula is derived, and new concept and model of the bandgap-spread multilevel field emission is given.
The presented theory considers field emission starting from the energy distribution of the emitted electrons in both normal and transverse directions to describe the field emission phenomena including the intrinsic emittance and thermal effects.

### A. Normal and Transverse Energy Distribution

According to Sommerfeld free electron gas model the number of electrons in unit volume in the momentum range $dp_x, dp_y, dp_z$ is determined by the Fermi-Dirac distribution function and the total number of the energy states in the corresponding unit volume of the momentum phase space

$$N(p_x, p_y, p_z) dp_x dp_y dp_z = \frac{2}{\pi^{3/2}} \frac{h^3}{E} dp_x dp_y dp_z , \quad (1)$$

Here $p_x, p_y, p_z$ are the electron momentum components in the Cartesian coordinates $(X, Y, Z)$ respectively, $N(p_x, p_y, p_z)$ is in the units of $\frac{1}{m^3 (kg m/s)^2}$, $E$ is the electron kinetic energy measured relative to the bottom of the conduction band, $\mu$ is the Fermi level, $k$ is Boltzmann’s constant, $h$ is Plank’s constant, and the factor of 2 arises from the Pauli exclusion principle.

For the electron velocity phase space $(v_x, v_y, v_z)$ (1) yields

$$N(v_x, v_y, v_z) dv_x dv_y dv_z = \frac{2(m/\hbar)^3}{\pi^{3/2}} dv_x dv_y dv_z , \quad (2)$$

where $m$ is the electron mass, $N(v_x, v_y, v_z)$ is in the units of $\frac{1}{m^3 (kg/s)^2}$, and $E = \frac{m}{2} (v_x^2 + v_y^2 + v_z^2)$.

In cylindrical coordinates $(r, \theta, z)$ selected so that $v_r = v_r \sin \alpha$ and $v_z = v_r \cos \alpha$, for which $dv_r dv_z = dv_r dv \theta dv z$, integration of (2) over $\alpha$ from 0 to $2\pi$ gives

$$N(v_r, v_\theta) dv_r dv_\theta = 4\pi \left( \frac{m^3}{h^3} \right) \frac{v_r^2 dv_r dv_\theta}{m(v_r^2 + v_\theta^2)^{3/2}} , \quad (3)$$

where $N(v_r, v_\theta)$ is in the units of $\frac{1}{m^3 (kg/s)^2}$.

Combining (3) with the continuity equation for the electron flux in the electron energy phase space $(E_x, E_r)$ find the number of electrons striking a unit area of the emitting surface, $x$ is the normal direction, per unit time

$$N(E_x, E_r) v_x dE_x dE_r = \frac{4em}{h^3} \frac{dE_r}{1 + e^{-E_r/E_x}} dE_r dE_r , \quad (4)$$

where $N(E_x, E_r)$ is in the units of $\frac{1}{m^3 (kg/s)^2}$, $E_x = m v_x^2 / 2$, and $E_r = m v_r^2 / 2$.

With the account of the transmission coefficient $\Delta$ arising from the probability of quantum tunneling of the electrons through the surface potential barrier obtain for the emission current density $j$

$$\frac{\partial^2 (j(E_x, E_r))}{\partial E_x \partial E_r} = \frac{4em}{h^3} \frac{\Delta(E_x)}{1 + e^{-E_r/E_x}} , \quad (5)$$

where $e$ is the elementary charge, and the designation

$$\frac{\partial j(E_x, E_r)}{\partial E_x \partial E_r} = eN(E_x, E_r) v_x .$$

The transmission coefficient $\Delta$ is determined by solution of the Schrödinger equation. The Kemble solution [8] for the parabolic and so the Schottky-Nordheim potential barrier is

$$\Delta(E_x) = \frac{\Delta_0}{1 + e^{\mu/E_x}} , \quad (6)$$

where $\Delta_0$ is the JWKB integral

$$\Delta_0(E_x) = e^{-G} , \quad (7)$$

with the exponent

$$G = \frac{B}{F} (\mu + \phi - E_x)^{3/2} v(y) , \quad (8)$$

$F$ is the applied electric field, $B$ is the universal Nordheim constant given by $B = \frac{8\pi \sqrt{2m}}{3\epsilon h}$, $\phi$ is the work function, $v(y)$ is the Nordheim function defined by the integral

$$v(y) = \frac{3}{4\sqrt{2}} \int_{1 - y^2}^{1 + y^2} \sqrt{1 - \rho - y^2} d\rho , \quad (9)$$

$$\rho = \frac{2eF}{\mu + \phi - E_x} x ,$$

and $y$ is the Nordheim parameter given by

$$y = \frac{e^2}{4\epsilon \epsilon_0} \sqrt{\frac{\phi}{\mu + \phi - E_x}} . \quad (10)$$

For the strong potential barriers, $\Delta_0(E_x) \ll 1$, i.e. $G$ is large enough, (6) approaches to a simpler form

$$\Delta \approx \Delta_0 . \quad (11)$$

Note that with such approximation one can let

$$\Delta = 1$$

for $E_x \geq \mu + \phi_0$, where $\phi_0$ can be looked as a work function reduced due to the Schottky effect and determined by $E_x$ from (10) at $y = 1$,

$$\phi_0 = \phi - \sqrt{e^2 F/4\epsilon \epsilon_0} .$$

From (5) obtain the current density distributions over electron kinetic energy in the normal and transverse directions, respectively

$$\frac{dj(E_x)}{dE_x} = \frac{4em}{h^3} \frac{\Delta(E_x)}{1 + e^{-E_x/E}} \int_0^\infty \frac{dE_r}{1 + e^{-E_r/E_x}} \quad (12)$$

or upon integration

$$\frac{dj(E_x)}{dE_x} = \frac{4em}{h^3} \int_0^\infty \frac{dE_r}{1 + e^{-E_r/E_x}} \quad (13)$$

$$\frac{dj(E_x)}{dE_x} = \frac{4em}{h^3} \left( \int_0^\infty \frac{dE_r}{1 + e^{-E_r/E_x}} + kT \ln \left( 1 + e^{-E_r/E_x} \right) \right) . \quad (15)$$

Here $\mu + \phi$ defines the integration limit, at which the transmission coefficient $\Delta \approx 1$. In case of (11), $\phi = \phi_0$.

The normal energy distribution in the form of (12, 14) is similar to that obtained by Fowler and Nordheim for the cold emission calculations. The transverse energy distribution (13, 15) presented first in our recent theory [9] is key to understanding the intrinsic emittance and thermal effects. Second term in (15) shows the thermionic emission contribution, which becomes insignificant for the cold emission regimes. The following discussion is focused on the cold emission current density and corrections to improve the accuracy for the barriers satisfying (11).

### B. Emission Current Density

Integrating (14) over $E_x$ (or (15) over $E_x$) and using (11) find the field emission current density

$$j = \frac{4em}{h^3} \int_0^\infty \frac{dE_r}{1 + e^{-E_r/E_x}} \Phi(\delta) kT \ln \left( 1 + e^{\delta/E_x} \right) d\delta . \quad (16)$$
where \( \Phi(\delta) = (1 + \delta)^{3/2}v(y) \),
\[
\delta = \frac{\mu - E_x}{\phi}, \quad y = \sqrt{\frac{e^2}{4\pi\epsilon_0}} \cdot \sqrt{\frac{\phi}{1 + \delta}}, \quad \text{and} \quad \epsilon_0 \text{ is electric constant.}
\]
Case of the lower temperature
\[
kT \ln \left( 1 + e^{\frac{\delta \varphi}{kT}} \right) = \begin{cases} 
\delta \varphi & \delta > 0, \quad e^{\frac{\delta \varphi}{kT}} > 1 \\
\delta \varphi & \delta < 0, \quad e^{\frac{\delta \varphi}{kT}} < 1 
\end{cases}
\]
and (16) is simplified to
\[
j = \frac{4\pi e}{h^3} \varphi^2 \int_0^{\mu/\varphi} e^{-\frac{8\varphi^3}{\pi}} \Phi(\delta) \delta \delta \ .
\]
Obtain Taylor series expansion for \( \Phi(\delta) \) defined by (17)
\[
\Phi(\delta) = \nu + \frac{1}{2} \tau \delta + \frac{1}{6} f \delta^2 + O(\delta^3)
\]
where \( \nu, \tau, \) and \( f \) are the correction functions designated for brevity as \( \nu \equiv \nu(y_0), \quad \tau \equiv \tau(y_0), \quad \) and \( f \equiv f(y_0) \), respectively, and \( y_0 \) is the argument given by (18) at \( \delta = 0 \),
\[
y_0 = \sqrt{\frac{e^2}{4\pi\epsilon_0}} \cdot \sqrt{F/\varphi},
\]
\[
\tau = v - \frac{2}{3} y_0 v' , \quad f = v + \frac{4}{3} \left( y_0^2 v'' - y_0 v' \right),
\]
\[
v' \equiv \frac{dv}{dy} |_{y=y_0} \quad \text{and} \quad v'' \equiv \frac{d^2v}{dy^2} |_{y=y_0} .
\]
Here the Nordheim function \( v \) is given by (9) and can be expressed through the elliptic integrals [7]
\[
v(y) = \sqrt{1 + y} \left( E \left( \frac{1-y}{1+y} \right) - yK \left( \frac{1-y}{1+y} \right) \right),
\]
where \( K \) and \( E \) are the complete elliptic integrals of the 1st and 2nd kind, respectively, commonly defined as follows
\[
K(t) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1-t \sin^2 \theta}}, \quad E(t) = \int_0^{\pi/2} \sqrt{1 - t \sin^2 \theta} \ d\theta,
\]
\[
[\arg(1-t)] < \pi .
\]
Taking into account the found derivatives
\[
\frac{dv}{dy} = \frac{3}{2} \sqrt{1-y} \cdot K \left( \frac{1-y}{1+y} \right), \quad \frac{d^2v}{dy^2} = \frac{3/4}{(1-y)^{3/2}} \left( E \left( \frac{1-y}{1+y} \right) - (2-y) K \left( \frac{1-y}{1+y} \right) \right)
\]
obtain from (23)
\[
f = \frac{v}{1-y_0} .
\]
Considerations similar in principle to (19-25) have been conducted by Murphy and Good [7]. However their analysis is limited by the 1st order correction function \( \tau \). One can see though from (17-20) that in case of the smaller work function \( \varphi \) (e.g. reduced by the Schottky effect or the band bending effects in semiconductors, or due to the emitter properties), the variable \( \delta \) increases, which degrades accuracy of the series representation (20). Figure 1 shows the exact values of \( \Phi(\delta) \) obtained from (17) and from the expansion (20) calculated for the linear, using \( \nu \) and \( \tau \), and quadratic, using \( \nu, \tau \) and \( f \), approximations. Figure 2 shows plots of the correction functions. Cutler and Good in the follow-up paper [10] have included the \( f \)-function into the field emission calculations, due to different motivations. However, this was done with additional approximations by taking the \( f \)-term out from the exponent and other expansion terms to simplify the emission current formula for further integration, which in turn reduces the accuracy and imposes constraints on the allowable \( f \)-values. We have found that no additional simplifications are necessary for the correct integration and included the higher-order correction function \( f \) into the field emission formula.

From (19, 20) obtain
\[
j = \frac{e^3}{8\pi \hbar^2 q^2} \int e^{-\frac{8\varphi^3}{\pi}} \varphi \cdot Z(g, C) \ ,
\]
where \( Z(g, C) \) is the correction factor given by the integral
\[
Z(g, C) = \int_0^{C} \frac{e^{-\frac{g^3 \varphi^2}{4\pi}} y^2 \ dy} {\varphi},
\]
\[
v = \frac{3B \varphi^{3/2}}{2} \cdot \tau \cdot \delta, \quad \text{and} \quad g \text{ and } C \text{ are the parameters defined as}
\]
\[
g = \frac{3B \varphi^{3/2} \tau^2}{2} \quad \text{and} \quad C = \frac{\mu}{\varphi} \cdot \tau \ .
\]
The left part of (26) without the \( Z \)-factor is the known Fowler-Nordheim formula with the \( \nu \) (Nordheim [5]) and \( \tau \)
III. MODELING OF FIELD EMISSION

Based on the physical mechanisms and the number of the emitting energy bands, in describing and modeling the cold field emission we distinguish three types of the emission. These are the single-band and the two-band field emission, and the suggested new type of bandgap-spread multilevel field emission.

A. Single-Band Field Emission

For electron emission from a single energy band, that may occur for example in electrical conductors or dielectrics, we have implemented into the MICHELLE [11] particle optics code the improved emission model with the correction to the Fowler-Nordheim theory. The correction takes into account the higher order series expansion terms for the integral determined by the Schottky-Nordheim tunneling potential barrier. As discussed in the section above the approximation error without this higher order correction becomes critical for the lower work functions relative to the energy spread of the emitted particles. The current density \( j \) emitted from each site of the emitting surface is described by (26) with the correction factor \( Z(g, C) \) from (29) or its simplified form \( Z(g) \) from (31).

One can evaluate the Nordheim function \( \nu \), and the correction functions \( \sigma \) and \( \Delta \) from (24), (22), and (25), respectively. Alternatively, to avoid evaluation of the elliptic integrals in (24) we can use the series expansion representations. Dean and Forbes derived series expansion for \( \nu(l) \), where \( l = y^2 \) [12]. Upon regrouping present their series expansion in the following ready to use form

\[
\nu(l) = 1 - \frac{3}{16} \sum_{i=0}^{\infty} a_i (k_i - \ln l)^{i+1},
\]

where \( a_0 = 1 \), \( k_0 = 1 + 6 \ln 2 \),
\[
a_{i+1} = \frac{\frac{i+2}{16} a_i}{i+3/4}, \text{ and } k_{i+1} = k_i - \frac{1}{i+1} - \frac{1}{i+3/4} + \frac{1}{i+1} + \frac{1}{i+2}.
\]

We show here the recursive rather than explicit relations for the \( a_{i+1} \) and \( k_{i+1} \) coefficients since this is preferential for implementation into the computer simulation process.

From (22) and (34) obtain series expansion for \( \tau \)

\[
\tau(l) = 1 + \frac{1}{4} \sum_{i=0}^{\infty} a_i \left( i + \frac{3}{16} \right) (k_i - \ln l) - 2i - 1)^{l+1}.
\]

The correction function \( \nu \) can be evaluated from (25) and (24) or through a series expansion. From (23) find that

\[
f = \nu + \frac{16}{3} \int_{l=0}^{\infty} \frac{d^n \nu}{dl^n} \left| l = y^2 \right| \],
\]

which in combination with (34) yields

\[
f(l) = \frac{1}{1 - \sum_{i=0}^{\infty} a_i \left( i \left( i + 1 \right) + \frac{3}{16} \right) (k_i - \ln l) - 2i - 1)^{l+1}.
\]

B. Two-Band Field Emission

The two-band field emission model suggests emission from both conduction and valence energy band that may occur for example in silicon or other semiconductor materials. The model applies the aforesaid single-band field emission model to each of the emitting energy bands. The two-band model takes into account the field penetration.
and band bending effect on the effective work function as well as the supply limited conditions when emission, e.g., from the conduction band, is saturated due to the limited number of electrons, which is controlled to a substantial degree by doping.

C. Bandgap-Spread Multilevel Field Emission

For describing the field emission from diamond we introduce a new concept of bandgap-spread multilevel field emission, which involves emission from the multiple energy states that are either discretely or continuously distributed within the bandgap. The bandgap spread can be caused by various kinds of imperfections, intentional or accidental, such as lattice defects, dislocations, phase and chemical impurities, doping, and surface treatment. This emission model can be applied to other materials as well that emit electrons from multiple energy states.

As an example, the suggested bandgap-spread multilevel field emission model explains the voltage-current characteristics observed from a single diamond field emission nanotip. We can see in Fig. 5 good agreement between the simulations using the bandgap-spread multilevel field emission model (3-emission levels in this case) and experiment. On the contrary, the single- or two-band models cannot explain these experimental results.

As an example, the suggested bandgap-spread multilevel field emission model explains the voltage-current characteristics observed from a single diamond field emission nanotip. We can see in Fig. 5 good agreement between the simulations using the bandgap-spread multilevel field emission model (3-emission levels in this case) and experiment. On the contrary, the single- or two-band models cannot explain these experimental results.

V. ELECTROSTATIC FOCUSING

The main requirements for the focusing geometry are that it should capture as many as possible of the emitted electrons and transport and focus them with minimal loss over the distance and at the diameter required for the particular application. Power losses of the entire structure should be minimized. In addition, the focusing approach should be expandable to larger field emitter arrays and yet remain feasible for fabrication, taking into account the micron scale of the geometry. Simulation results of field emission nanotip using the new bandgap-spread multilevel field emission model (3-emission levels) show good consistency with the experiment. Experimental data shown are courtesy of C. A. Brau of Vanderbilt University.

IV. EFFECTS OF THE FIELD EMISSION PROPERTIES

It is imperative to understand the effects of the emission properties on the electron beam focusing and attainable characteristics. With such understanding one can determine and optimize requirements on the field emitters. Furthermore, the precise field emission properties may be not known or not achievable, so one can describe the beam focusing and propagation in the channel as a function of emitter parameters. For example, our simulations show that the effects of dopant level on emission properties tend to saturate with larger dopant amounts, as seen in Fig. 6. Also, a larger work function in general degrades the beam quality, affecting the emittance and brightness. However, in most cases the focusing structure remains functional and requires only minor voltage adjustments to accommodate the variation in field emission properties.

Figure 5. Simulations of the electron gun with diamond nanotip using the new bandgap-spread multilevel field emission model (3-emission levels) show good consistency with the experiment. Experimental data shown are courtesy of C. A. Brau of Vanderbilt University.

Figure 6. Effect of doping on field emission current in the double-gate geometry with focusing anode shown in Fig. 7.

Figure 7. Simulation of field emission with electrostatic focusing. Colors represent the current in the beam annular layers. Top: Entire model with particle trajectories. The image is stretched by x100 in radial direction. Middle: Overall view of the focusing geometry, in scale. Bottom: electron emission and beginning of the beam formation. Inset shows, magnified, the emitting 10-nm radius tip including the small but still occurring side emission.
emission from a single silicon-tip with electrostatic focusing are shown in Fig. 7. We derived this example geometry for a 10-100 kV high resolution X-ray imaging source. The focusing optics has extraction and focusing gates positioned near the emitter tip and a focusing anode downstream on the beam axis. The extraction gate 1 controls the emitted current. The focusing gate 2 pushes the expanding particle orbits toward the beam axis. The focusing anode bends the expanding particle orbits, thus allowing high aspect ratio confinement and beam transport. The current density distribution in the anode plane is shown in Fig. 8.

We have discovered that the suggested double-gate with focusing anode geometry can form fundamentally better quality electron beam with laminar trajectories. Such new laminar beam formation regime requires the sharper tip radius, lower work function, or shorter tip to anode length, e.g. 1 eV, 2 nm, and 5 mm, respectively, for which the simulation results are displayed in Fig. 9 versus the 4.05eV, 10 nm, and 9.8 mm example shown in Fig. 7. The sharper tip and shorter length lead to higher electric field formed over the tip surface due to the anode, including focusing anode, voltage that together with the lower work function reduce demand for the electric field from the extraction gate to achieve the required emission current. As the necessary gate field, and so the transverse component of the beam extraction field, is minimized the optics produces laminar electron beam of the smaller emittance, reduced spot size with the sharp boundary, and vanishing halo.

Acknowledgements
We appreciate helpful discussions with Kevin Jensen of NRL and Luis Velasquez-Garcia and Frances Hill of MIT.

VI. REFERENCES