Theory and Modeling of Field Electron Emission
from Low-Dimensional Electron Systems

by

Alex Andrew Patterson

B.S., Electrical Engineering
University of Pittsburgh, 2011

M.S., Electrical Engineering
Massachusetts Institute of Technology, 2013

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2018

© Massachusetts Institute of Technology 2018. All rights reserved.

Author .................................................................
Department of Electrical Engineering and Computer Science
January 31, 2018

Certified by ............................................................
Akintunde I. Akinwande
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .............................................................
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Theses
Theory and Modeling of Field Electron Emission from Low-Dimensional Electron Systems
by
Alex Andrew Patterson

Submitted to the Department of Electrical Engineering and Computer Science on January 31, 2018, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering

Abstract

While experimentalists have succeeded in fabricating nanoscale field electron emitters in a variety of geometries and materials for use as electron sources in vacuum nanoelectronic devices, theory and modeling of field electron emission have not kept pace. Treatments of field emission which address individual deviations of real emitter properties from conventional Fowler-Nordheim (FN) theory, such as emission from semiconductors, highly-curved surfaces, or low-dimensional systems, have been developed, but none have sought to treat these properties coherently within a single framework. As a result, the work in this thesis develops a multidimensional, semiclassical framework for field emission, from which models for field emitters of any dimensionality, geometry, and material can be derived.

The effects of quantum confinement and emitter tip geometry on the properties of emission were investigated by utilizing the framework to derive models for: i) a highly-curved, nanoscale, metal emitter tip; ii) a bulk silicon emitter with a surface quantum well formed due to electric field penetration and a mechanism that limits the maximum conduction band emitted current density (ECD) to the bulk flux density supply; and iii) a cylindrical silicon nanowire emitter. Results from a highly-curved, nanoscale, metal emitter tip reveal that despite significant electron supply reductions as a result of quantum confinement, the emitted current density (ECD) increases as the emitter radius decreases due to the effects of electric field enhancement. Additionally, emitters with radii smaller than 5 nm exhibit a narrow total energy distribution and highly non-linear FN plots. Consistent with experimental observations, the saturation of the conduction band ECD in silicon emitters leads to the appearance of three distinct regions in FN plots, which signify conduction-band-dominated, valence-band-dominated, and transitional regimes of emission. Confinement of electrons to a nanowire emitter geometry further reduces the electron supply available for emission and, consequently, the conduction band saturation ECD.

Overall, findings show that the dimensionality, geometry, and material of field emitters all play a critical role in field emission processes at the nanoscale. Accordingly, the semiclassical framework for field emission is intended to form a solid foundation upon which more complete models of emission can be developed.

Thesis Supervisor: Akintunde I. Akinwande
Title: Professor of Electrical Engineering and Computer Science
Acknowledgments

There are many people who have supported me over the course of my time at MIT and I would like to thank them here. For those of you I may not have mentioned, thanks go out to you as well.

Foremost, I would like to thank my research advisor and mentor, Professor Tayo Akinwande, for his guidance, encouragement, and constant support. I am especially grateful for the independence he granted me during my time at MIT, particularly in choosing the direction of my research and by allowing me to leave Cambridge to undertake an internship in Germany near the end of my doctoral studies. He has instilled in me the importance of crafting a compelling narrative from my work, the resolve to leave no stone unturned, and the discipline to always “do my due diligence”. Under his supervision, I have grown immensely as a researcher and I look forward to his continued mentorship in my career.

I would also like to thank my other committee members, Professor Terry Orlando and Professor Dirk Englund, who have provided me with useful feedback during the course of my thesis work. I have known Professor Orlando as long as I have been at MIT; he has guided me as my housemaster, academic counselor, and supervising professor during my teaching assistantship and has supported me in both my research and teaching aspirations. Professor Englund has always advised me to tie the findings of my work to experimental results, from which I have developed a valuable understanding of the “other side” of field emission research.

From enlightening evenings at the library to the annual IVNC trips, the members of the Akinwande group have been great people to spend my time with daily. In particular, I would like to thank Melissa, Steve, Arash, Neptune, Andy, and Nedo. Thank you as well to others in the MTL community, particularly Dennis and Sabino, for the moments of comic relief and hilarious cookie social discussions.

I would also like to thank my friends in Cambridge, and elsewhere, for all of
the amazing memories over the years, especially Parker, Joel, Cait, Guha, Joey, Michael, Mike, Dave, Stephen, all the young lads, assorted degases, and the three-time-champion GAME softball team.

Finally, I would like to thank my parents, Pat and Steve, my brother Ryan, cousin Dan, cousin Nickie, Aunt Pam, Uncle John, Pete, Ksenia, Sonny, and the rest of my family for their continual support and for always reminding me of where to find home. May we all gather around the shagbark for years to come.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374.
## Contents

1 Introduction 35

2 Background 39

2.1 Current State of Field Emission Theory . . . . . . . . . . . . . . . . . . 39

2.2 Fowler-Nordheim Theory . . . . . . . . . . . . . . . . . . . . . . . . . 40

2.2.1 Elementary Fowler-Nordheim Equation . . . . . . . . . . . . . . 41

2.2.2 Fowler-Nordheim Plots . . . . . . . . . . . . . . . . . . . . . . . . . 44

2.2.3 Standard Fowler-Nordheim Equation: Murphy-Good Theory . . . 45

2.2.4 Fowler-Nordheim-Type Equations . . . . . . . . . . . . . . . . . . 47

2.3 Emission from Semiconductors . . . . . . . . . . . . . . . . . . . . . . . 48

2.3.1 Emission from the Conduction Band: $E_F > E_C(z_s)$ . . . . . . . 51

2.3.2 Emission from the Conduction Band: $E_F < E_C(z_s)$ . . . . . . . 53

2.3.3 Emission from the Valence Band: $E_F > E_V(z_s)$ . . . . . . . . . 53

2.3.4 Emission from Surface States . . . . . . . . . . . . . . . . . . . . . 55

2.3.5 Emission from Accumulation Layer Bound States . . . . . . . . . 57

2.4 Emission from Highly-Curved Emitter Tips . . . . . . . . . . . . . . . . 58

2.4.1 Field Enhancement Factors . . . . . . . . . . . . . . . . . . . . . . 59

2.4.2 Generalized Fowler-Nordheim Equation for Non-Planar Emitters 60

2.4.3 Semiclassical, Multidimensional Tunneling . . . . . . . . . . . . . 62

2.5 Emission from Low-Dimensional Emitters . . . . . . . . . . . . . . . . 69
2.6 Chapter Summary .................................................. 73

3 Semiclassical Framework for Field Electron Emission 75

3.1 Introduction ...................................................... 75

3.2 Semiclassical Solution to the Multidimensional Schrödinger Equation 77

3.2.1 Zeroth Order Wave Function: Classical Hamilton-Jacobi Equation ........................................ 80

3.2.2 First Order Wave Function: Continuity Equation ............ 81

3.3 Emitted Current Density Equation ................................. 83

3.3.1 Incident Current Density ................................. 85

3.3.2 Electron Transmission Probability ......................... 86

3.3.3 Bohr-Sommerfeld Quantization ............................. 91

3.3.4 Complete Expression for Emitted Current Density ......... 93

3.4 Model Assumptions and their Consequences ...................... 93

3.4.1 Quantum Potential ................................. 94

3.4.2 Local Density of States ............................. 96

3.4.3 Semiclassical, Multidimensional Tunneling Model ........ 98

3.5 Chapter Summary .................................................. 100

4 Field Electron Emission from a Low-Dimensional, Highly-Curved Metal Tip 101

4.1 Introduction ...................................................... 101

4.2 Definition of Paraboloidal Emitter System ........................ 102

4.2.1 Three-Dimensional Parabolic Coordinates ................. 102

4.2.2 Physical Definition of the Emitter Cathode .................. 104

4.2.3 Emitter System Electrostatics ............................ 106

4.3 Derivation of Emitted Current Density Equation ............... 110

4.3.1 Semiclassical Wave Functions in Parabolic Coordinates .... 111
6 Field Electron Emission from Cylindrical Silicon Nanowires

6.1 Introduction ........................................... 201

6.2 Cylindrical Silicon Nanowire Emitter Model ........................... 202
   6.2.1 Motivation for Cylindrical Silicon Nanowire Emitter Model . . 202
   6.2.2 Electron System of the Cylindrical Silicon Nanowire Emitter . 204
   6.2.3 Energetic Regimes of Emission from the Cylindrical Silicon
          Nanowire ........................................ 206
   6.2.4 Emitted Current Density Equations .......................... 208

6.3 Results ................................................................ 213
   6.3.1 Emitted Current Density vs. Applied Field ..................... 214
   6.3.2 Total Energy Distribution ................................. 218
   6.3.3 Fowler-Nordheim Plots .................................... 223

6.4 Comparison to Experimental Results .................................. 229
   6.4.1 I-V and FN Plot Data: n-type Single-gated Silicon FEAs . . 230
   6.4.2 FN Plot Data with Multiple Emission Regimes: n-type and
          p-type Single-gated Silicon FEAs .......................... 234

6.5 Limitations of the Model ......................................... 239
   6.5.1 One-dimensional Simulation of Silicon Electrostatics .......... 240
   6.5.2 Potential and Normal Bound State Energies Independent of
          Lateral Quantum Confinement ................................ 241
   6.5.3 Band Structure Independent of Emitter Dimensions .......... 242

6.6 Chapter Summary ............................................. 243
List of Figures

1-1 TEM image of a silicon field electron emitter with a radius of curvature of less than 1 nm [1]. Modern field emitter tip geometries are characterized by a gradual taper which terminates in a sharp, but rounded apex. ................................................................. 36

2-1 The major regimes of field emission theory can be classified via the properties of the emitter: i) electron system dimensionality, ii) emitter surface planarity, and iii) emitter material. The dashed lines connecting the emitter properties (triangles) to the regimes of field emission theory (ovals) represent the assumptions upon which the regime is based: Fowler-Nordheim theory describes emission from a bulk, planar, metal, emitter; sharp emitter theory is based upon a bulk, non-planar, metal emitter; semiconductor emission theory is based upon a bulk, planar, semiconductor emitter; and low-dimensional emission theory is based upon a non-bulk, planar, metal emitter. More general treatments of field emission, such as is the goal of this work, seek to unify multiple regimes of emission theory. .............................. 40
2-2 Inside the metal, a constant flux density of electrons $N$ impinges upon the emitter surface per unit area, per unit energy. Of the incident electronic flux density, only a small fraction $D$ manages to tunnel through the potential barrier $V(z)$ into vacuum and contribute to the emitted current density $J$ (straight arrow), whereas the rest is reflected back into the metal (curved arrows). 

2-3 Normal energy diagram for a bulk, planar, metal emitter with Fermi energy $E_F$, work function $\phi$, and an exact triangular potential barrier, defined by $V(z)$. $H$ is the zero-field barrier height experienced by an electron with normal energy $W$ and in adhering to customary field emission notation, the direction of the electric field $F$ is defined oppositely to that of the usual convention.

2-4 Normal energy diagram for a bulk, planar, metal emitter with the Schottky-Nordheim barrier (solid line), defined by $V(z)$, and exact triangular barrier (dashed line).

2-5 Constant energy surfaces of [001]-oriented silicon, projected onto the $k_z$ plane, where $k_i$ denotes momentum in the $i$ dimension. Assuming motion in the $z$ direction, conduction valleys lying along the $k_x$ and $k_y$ axes are elliptical with two transverse effective masses $m_t$ and $m_t$ and normal effective mass $m_{t'}$, while conduction valleys lying along the $k_z$ axis are circular, with transverse effective mass $m_t$ and normal effective mass $m_{t'}$. 

42

44

47

49
2-6 Energy band diagram for an n-type emitter that is degenerate in the CB at the surface, where $E_0$ is the vacuum level, $E_C$ denotes the bottom of the conduction band, $E_g$ is the magnitude of the forbidden energy gap, $E_F$ is the Fermi level, $E_V$ is the top of the valence band, and $z_s$ is the location of the emitter surface. Like in FN theory, most of the ECD originates from states close in energy to $E_F$. 52

2-7 Energy band diagram for an n-type emitter, which is non-degenerate at the surface. Most of the ECD originates from states at the bottom of the conduction band. 54

2-8 Energy band diagram for a semiconductor emitter, with a band of surface states located in the forbidden energy gap. The surface state band has minimum energy $E_s^0$ and a bandwidth of $\Gamma_s$. While $\theta^b_s$ denotes the difference between the conduction band edge at the surface and the bulk Fermi level, $\theta^a_s$ denotes the difference between the conduction band edge at the surface and surface state Fermi level. 55

2-9 Metal-oxide-semiconductor (MOS) structure, where the applied field has caused a quantum well (in the accumulation layer) and bound states to form near the Si-SiO$_2$ interface. 57

2-10 A non-planar protrusion between two parallel contacts. The macroscopic field $F_M$ far away from the emitter tip is equivalent to the constant electric field between the contacts of a parallel plate capacitor, while the surface field magnitude at the apex of the protrusion $F$ is enhanced due to the protrusion’s curved geometry. The ratio between $F$ and $F_M$ is the field enhancement factor at the apex $\gamma_a$. 59
2-11 The numerically-calculated tunneling barrier potential energy $U(z)$ for an ellipsoidal surface, as a function of the emitter radius of curvature. As can be seen, the potential barrier shape depends on the emitter’s radius of curvature $R$ and differs appreciably from the Schottky-Nordheim barrier (FN) [2].

2-12 The radius of curvature of the emitter $R$ is defined by the radius of the osculating circle, which touches the emitter surface at the apex.

2-13 Highly-curved emitter tip and anode, along with equipotential lines and the direction of the electric field $F$ at selected points in the vacuum. Due to the variation in the direction of the field $F$ as a function of position outside the emitter, electrons are acted upon by forces in multiple dimensions and may travel along multidimensional paths. A few example paths for electrons emitted from different locations on the emitter surface are shown and labeled as 1, 2, and 3.

2-14 Relative orientation of the wave fronts (surfaces of constant $W_{0R}$ and $W_{0I}$), along with the trajectories corresponding to the real $\hat{e}_R$ (R paths) and imaginary $\hat{e}_I$ (I paths) components of the electron’s wave vector $k_R(r) = 0$, only I paths exist, which are always normal to surfaces of constant $W_{0I}$ and, equivalently, the turning surfaces.
2-15 An electron tunneling from a classically-allowed Region I at point $\mathbf{r}_1$, through a classically-forbidden Region II, into a classically-allowed Region III at point $\mathbf{r}_2$ has infinitely many possible tunneling paths available, a few of which are shown. Under the semiclassical approximation, only the most probable path $s$ (dashed line) is considered, which coincides with the classical path of the particle, as determined by the classical equations of motion. $\hat{\mathbf{e}}$ is the unit tangent vector at a point along the tunneling path.

2-16 Examples of field emitter geometries of various dimensionalities with planar emitter surfaces: (a) bulk emitter (FN model), (b) normally unconfined nanowall emitter, (c) normally unconfined rectangular nanowire emitter, (d) normally confined nanowall emitter, and (e) normally confined rectangular nanowire emitter. The emission direction is along the $z$ axis.

2-17 The emitted current density as a function of emitter dimensions, normalized to the elementary Fowler-Nordheim equation, for (a) normally unconfined and (b) normally confined field emitters of various dimensionalities and geometries. [4].

2-18 The normalized emitted current density for a cylindrical nanowire emitter, with the field enhancement factor of a hemispherical cap, as a function of emitter diameter $d$. At a critical radius, the reduction in the electron supply due to quantum confinement overcomes the gains in transmission probability due to electric field enhancement and the ECD decreases as a function of decreasing emitter diameter [4].

3-1 Depiction of a tube of neighboring electron paths and its change in cross-sectional area $\sigma(s)$ as a function of position along the paths $s$. 
3-2 Schematic of a grounded, non-planar emitter tip (cathode) and anode, held at a potential $V = V_0$. At each location on the emitter surface $s$, the internal current density impinging on the emitter surface $\eta_Q$ has a probability of being transmitted through the potential barrier $D_Q$, resulting in a local emitted current density $J_Q$. Electrons tunnel through the classically-forbidden region between the emitter surface $s$ (solid curve) and the turning surface $s_t$ (dashed curve), along a path $s$.  

3-3 An electron, normally incident on the turning surface between the classically-allowed region and forbidden region, tunneling through the forbidden Region II ($E < V(r)$) follows a path $s$, which is governed by the classical equations of motion. The tunneling probability is determined by finding the ratio of the probability flux density exiting the forbidden region $J_{II}(r_2)$ to the probability flux density incident on the forbidden region $J_I(r_1)$. 

3-4 While the infinite square well potential (a) is characterized by two hard boundary conditions, the harmonic oscillator potential (b) is characterized by two soft boundary conditions. 

3-5 Numerically-calculated enhancement factor curves $R(\epsilon) = j(\epsilon)/j_0(\epsilon)$ for the (110) and (100) planes of tungsten, where $j(\epsilon)$ is the total energy distribution with the surface density of states included and $j_0(\epsilon)$ is the total energy distribution of the free electron model, without the surface density of states [5].

4-1 The surfaces of constant coordinate for $\eta = 2$ (blue, downward-opening paraboloid) and $\xi = 0.5$ (red, upward-opening paraboloid), along with the unit vectors $\hat{\eta}$ and $\hat{\xi}$ at selected points on their respective surfaces of constant coordinate.
4-2 Diagram of the paraboloidal emitter cathode defined between the surfaces of constant coordinate $\eta = \eta_1$ and $\xi = \xi_1$, with radius of curvature at the apex $R = \eta_1$, radius at the origin $a = \eta_1$, and height $L = (\eta_1 + \xi_1)/2$. The apex of the emitter is located at the point $s_a = (\eta_1, 0, \phi)$ and the emission direction $\hat{\eta}$ varies as a function of position on the emitter surface. 

4-3 Diagram of paraboloidal cathode-anode geometry. Like the emitter cathode, the anode is a downward-opening paraboloid defined by the surface of constant coordinate $\eta = \eta_2$, where $\eta_2 > \eta_1$. A voltage $V = V_0$ is applied to the anode, whereas the emitter cathode is kept grounded. The spacing between the anode and cathode apices is $d = (\eta_2 - \eta_1)/2$.

4-4 Tunneling barrier generated between a paraboloidal cathode with radius of curvature $R = \eta_1 = 1$ nm, work function $\phi_0 = 4.5$ eV, and Fermi energy $E_F = 5$ eV, and anode of radius $\eta_2 = 50$ nm, for applied voltage $V_0 = 10$ V, showing (a) the contours of the potential in the $x - z$ plane and (b) the profile of the tunneling barrier at the apex of the cathode, along the $z$ axis.

4-5 The magnitude of the electric field near the emitter surface for an emitter of radius $R = \eta_1 = 1$ nm, with work function $\phi_0 = 4.5$ eV and Fermi energy $E_F = 5$ eV, and an anode of radius $\eta_2 = 50$ nm, for applied voltage $V_0 = 10$ V. The plot in (a) shows the contours of the magnitude of the electric field in the $x - z$ plane, from which the electric field enhancement at the emitter tip apex is evident, and (b) shows the magnitude of the electric field as a function of position on the emitter surface, where $\xi = 0$ corresponds to the emitter apex and $\xi = 20$ nm is located closer to the emitter base. The apex field factor for this particular cathode-anode system is $\beta_a = 0.511 \text{ nm}^{-1}$. 

19
4-6 The total energy $E_Q$ vs. $q$ dispersion relation for the subbands \{$p, n$\} in a paraboloidal emitter of radius $\eta_1 = 1.5$ nm. Subbands with $n = 1$ are displayed as solid curves, while subbands with $n = 2$ are displayed as dashed curves.

4-7 The kinetic energy $T_{\eta,Q}$ associated with the $\eta$, $\xi$, and $\phi$ coordinates as a function of position on the surface $\xi$ of an emitter with radius $\eta_1 = 1.5$ nm, compared with the total energy $E_Q$. Plot (a) shows the kinetic energies for a state with $E_Q \approx E_F$, $p = 0$, and $n = 1$, while (b) shows the kinetic energies for a state with $E_Q \approx E_F$, $p = 1$, and $n = 1$.

4-8 Plots of the total state energy $E_Q$, kinetic energy associated with $\eta$ ($T_{\eta,Q}$), kinetic energy associated with $\xi$ ($T_{\xi,Q}$), and kinetic energy associated with $\phi$ ($T_{\phi,p}$) for subband \{$p = 0, n = 1$\}, as a function of $q$. The kinetic energy is plotted at four points: (a) at the apex, where $\xi = 0$ and $0 < q_{c,Q} = q_{max,Q} = 32$ nm$^{-1}$; (b) at $\xi = 0.375 \eta_1$, where $0 < q_{c,Q} < q_{max,Q}$; (c) at $\xi = \eta_1$ (or $z = 0$), where $q_{c,Q} = 0$; and (d) at $\xi = 3 \eta_1$, where $q_{c,Q} < 0$. The maximum value of $q$ allowed within a particular subband is $q_{max,Q}$.

4-9 Percentage of the total emitted current density contributed by (a) $p$ subbands and (b) $n$ subbands, as a function of emitter radius $\eta_1$. $n_{max}$ denotes the largest $n$ quantum number of subbands present in the emitter.

4-10 The (a) incident flux density $\eta_{\eta,Q}(\eta_1, \xi)$ and (b) transmission probability through the barrier as a function of position on the surface $\xi$ and quantum number $p$, for an emitter of radius $\eta_1 = 1.5$ nm, applied voltage $V_0 = 22.5$ V, and $E_Q = E_F$.

4-11 The (a) energy available for tunneling at the surface for the lowest 5 $p$ subbands as a function of $\xi$ and (b) tunneling paths for $p = 0$ and $p = 4$, for electrons with $E_Q = E_F$ in an emitter of radius $\eta_1 = 1.5$ nm.
4-12 The incident flux density at the apex \( \eta_{\theta,Q}(\eta_1) \) for \( p = 0 \) as a function of total energy \( E_Q \) and subband \( n \), for (a) an emitter of radius \( \eta_1 = 2.5 \) nm and (b) an emitter of radius \( \eta_1 = 5 \) nm.

4-13 (a) The effective emitting area \( A_{eff} \), the area of the emitter surface from which at least 99% of the emitted current originates; and (b) the emitting area efficiency, the percentage of the total emitter surface area \( A_{tot} \) represented by \( A_{eff} \), as a function of emitter radius \( \eta_1 \) and applied voltage \( V_0 \).

4-14 The (a) normalized emitted current density and (b) tunneling barrier width for electrons in subband \( p = 0 \) with \( E_Q = E_F \) as a function of \( \xi \), for a fixed applied voltage \( V_0 = 22.5 \) V and varying emitter radius \( \eta_1 \).

4-15 The (a) normalized emitted current density and (b) tunneling barrier width for electrons in subband \( p = 0 \) with \( E_Q = E_F \) as a function of \( \xi \), for various applied voltages \( V_0 \) and fixed emitter radius \( \eta_1 = 1.5 \) nm.

4-16 Emitted current density vs. emitter radius \( \eta_1 \) for the case of (a) fixed transmission probability at the apex, achieved by varying the applied voltage \( 15V < V_0 < 53V \), and (b) fixed applied voltage \( V_0 = 22.5 \) V. The ECDs from the paraboloidal emitters are plotted as red points, while the predictions of the elementary FN equation with the \( F_a \) of the paraboloidal emitters are plotted as black, dashed lines.

4-17 Plots of (a) the total electron density averaged across the emitter surface and (b) normal electron velocity at the surface averaged over all subbands, across the emitter surface, as a function of emitter radius \( \eta_1 \).

4-18 Plots of the (a) transmission probability at the apex, averaged over all \( p = 0 \) subbands, and (b) tunneling barrier width for an electron in subband \( p = 0 \) and with \( E_Q = E_F \), as a function of emitter radius \( \eta_1 \) for a fixed applied voltage \( V_0 = 22.5 \) V.
4-19 The normalized total energy distribution of emitted electrons with $T = 300$ K for the cases in which (a) the applied voltage is fixed at $V_0 = 22.5$ V and the emitter radius $\eta_1$ is varied, and (b) the emitter radius is fixed at $\eta_1 = 1.5$ nm and the applied voltage $V_0$ is varied. The corresponding elementary Fowler-Nordheim total energy distributions for equivalent $F_a$ are plotted as dashed lines in the same color as the paraboloidal emitter results, which are plotted as solid lines.

4-20 Plots of the tunneling barrier shape at the apex for (a) $V_0 = 22.5$ V and various emitter radii $\eta_1$ and (b) $\eta_1 = 1.5$ nm and various applied voltages $V_0$, as a function of distance along the $z$ axis. The dashed lines originating from the apex depict the exact triangular barrier defined by the electric field magnitude at the corresponding paraboloidal emitter apex $F_a$, while the dashed black line marks the Fermi energy $E_F$.

4-21 Fowler-Nordheim plots for emitters of varying radii and an anode radius of $\eta_2 = 50$ nm, expressed in terms of (a) the emitted current density $J$ vs. the apex electric field $F_a$ and (b) the emitted current $I$ vs. the applied voltage $V_0$.

4-22 Plots of the tunneling probability (a) as a function of the reciprocal of the magnitude of the electric field at the apex $1/F_a$ and (b) as a function of the reciprocal of the applied voltage $1/V_0$. The black dashed line is the tunneling probability for the exact triangular barrier of the elementary Fowler-Nordheim equation.
4-23 The tunneling barrier potential at the apex of a paraboloidal emitter of radius $\eta_1 = 1$ nm and an anode of radius $\eta_2 = 50$ nm for various applied voltages $V_0$, as a function of distance along the $z$ axis. The dashed lines correspond to the exact triangular barrier defined by the electric field magnitude at the emitter apex $F_a$ and the solid, gray line denotes the tunneling energy of the electron. 147

4-24 (a) Diagram of a single-gated field emitter array of molybdenum field emitter tips of radius 5 nm or 10 nm, and height 1.2 $\mu$m, with a gate aperture diameter of 2.3 $\mu$m. (b) I-V characteristics for three fabricated molybdenum field emitter arrays, the paraboloidal emitter model (circles), and the generalized nanoscopic Fowler-Nordheim model (dashed lines). While the device labeled “Single-gate FEA 1” (green circles) is composed of emitters with apex radii of curvature 5 nm, the device labeled “Single-gate FEA 2” (unfilled, blue squares) is composed of emitters with apex radii of curvature 10 nm. Model results for comparison with “Single-gate FEA 1” are plotted in black, while model results for comparison with “Single-gate FEA 2” are plotted in gray. 148

5-1 SEM images of (a) an ungated silicon field emitter array and (b) a single-gated silicon field emitter array [6, 7]. 154
5-2 (a) The band structure of silicon consists of a single conduction band $E_c$ and three valence bands: heavy hole $E_{v, hh}$, light hole $E_{v, lh}$, and split-off $E_{v, so}$, separated by an energy gap $E_g = 1.12$ eV, in which no states occur inherently. The electron affinity of silicon is $\chi_e = 4.04$ eV. (b) The conduction band of silicon has six conduction valleys, represented here by their ellipsoidal constant energy surfaces. Owing to the definition of the emission direction as along the $z$ axis of (001) silicon, conduction valleys which lie along the $k_z$ axis are defined as “on-axis”, while conduction valleys which lie along the $k_x$ and $k_y$ axes are termed “off-axis”.

5-3 Due to the change in the electron mass from $m_e$ to $m_0$ upon leaving the semiconductor, electrons in the conduction band of silicon acquire additional normal energy in vacuum $W'$ as compared to the normal energy inside the emitter $W$. For electrons in the valence band, the additional normal energy acquired from the change in electron mass shifts the normal energy in vacuum $W'$ downward and away from the top of the tunneling barrier.

5-4 Energy band diagram showing the relative lowering of the barrier height for electrons in the conduction band with normal energy $W$ and electrons in the valence bands with normal energy $\overline{W}$, due to a field-penetration-induced downward bending of the bands $\Delta E_c$. The barrier height for conduction band electrons $H_{CB}$ and for valence band electrons $H_{VB}$ is lowered by $\Delta E_c$, relative to the case in which field penetration does not occur. The effect on the normal energy in vacuum due to the change in the electron mass upon leaving the semiconductor is omitted here for clarity.
5-5 Energy band diagram for an electron in the valence band, with normal energy $W$, located just below the valence band maximum in the bulk. In order to be emitted, the electron must tunnel through the portion of the forbidden energy gap labeled as the internal tunneling barrier, with probability $D_{int}$, and through the usual surface barrier in the vacuum labeled here as the external tunneling barrier, with probability $D_{ext}$.  

5-6 Energy band diagram of the conduction band edge $E_c(z)$ and the discrete normal energy states that form near the surface $W_n$ as a result of electric field penetration. For the purposes of calculating the bound state energies $W_n$, the potential below the conduction band edge is taken to be infinite, while above the conduction band edge it is zero.  

5-7 Energy band diagram depicting the different energetic regimes of field emission from silicon in terms of the normal energy $W$ of electrons. For normal energies $W \geq E_{cb}$, the ECD originates from extended states in the conduction band; for $E_{cb} > W > E_{cs}$, the ECD originates from bound states in the conduction band with energies $W_n$; and for $E_{vb} > W$, emission originates from the valence band.  

5-8 Flow chart detailing the algorithm that calculates the emitted current density from conduction band states. The ECD is calculated for each total energy $E$ between an $E_{min}$ and $E_{max}$ with a spacing of $dE$. At each energy $E$, the ECD from bound states $J_n$ is calculated first, from lowest $W_n$ to highest, followed by the ECD from extended states $J_{ce}$. If any component of the ECD exceeds the remaining flux density supply from the extended states $\Phi_r$, the ECD of that component is limited to $\Phi_r$ and the ECD from all higher energy states is zero.
5-9 The simulation of field emission from bulk silicon was consisted of two steps. In the first step, *nextnano3* was used to find a self-consistent solution to Poisson’s equation and the Schrödinger equation across a range of surface electric fields, providing the electronic properties of the emitter. In the second step, the data provided by nextnano was used to calculate the ECD and TED contributions from each of the energetic emission regimes in the model.

5-10 Plots of the (a) conduction band and three valence band edges and (b) lowest 10 bound state energies $W_n$ in each type of conduction valley of the conduction band for n-type silicon with $N_D = 10^{15}$ cm$^{-3}$. Plots (c) and (d) display the same results for p-type silicon with $N_A = 10^{15}$ cm$^{-3}$. The energy reference level is the Fermi level.

5-11 Plots of the lowest bound state energy at the surface $W_1$, as measured from the conduction band edge in the bulk, vs. the applied electric field $F$, for both the on-axis (solid lines) and off-axis (dashed lines) conduction valleys of the conduction band in (a) n-type silicon and (b) p-type silicon.

5-12 The conduction band edge energy as a function of distance from the emitter surface $z$, as referenced to its value in the bulk for (a) n-type emitters and (b) p-type emitters of varying dopant density.

5-13 The total emitted current density $J$ from n-type (circles) and p-type (triangles) silicon emitters of varying dopant density, as a function of applied field on the vacuum side of the silicon surface $F$. 

26
5-14 The conduction band extended state (blue circles), conduction band bound state (blue triangles), and valence band extended state (red circles) emitted current densities as a function of applied field \( F \), for (a) an n-type emitter of dopant density \( N_D = 10^{15} \text{ cm}^{-3} \) and (b) a p-type emitter of dopant density \( N_A = 10^{15} \text{ cm}^{-3} \).

5-15 The normalized total energy distribution of emitted electrons from (a) an n-type silicon emitter of dopant density \( N_D = 10^{15} \text{ cm}^{-3} \) and (c) a p-type silicon emitter of dopant density \( N_A = 10^{15} \text{ cm}^{-3} \) for three different applied fields \( F \). Plot (b) shows a close-up of the conduction band TED of the n-type emitter, while plot (d) shows a close-up of the valence band TED of the p-type emitter. In plots (a) and (b), TED values of zero were not plotted to improve the clarity of the plots.

5-16 (a) Plot of the heavy hole valence band maximum energy (solid lines) and the potential barrier in vacuum (dashed lines) for various applied fields \( F \) as a function of \( z \), where \( z = 0 \) denotes the location of the surface. (b) For the same energies, the normalized transmission probability through the downward-bent valence band and external potential barrier for various applied fields \( F \), as shown in (a).

5-17 Fowler-Nordheim plots for n-type and p-type emitters of varying dopant density. Although the FN plot corresponding to an n-type emitter with \( N_D = 10^{15} \text{ cm}^{-3} \) is correctly located on the \( x \)-axis, all other FN plots were shifted positively along the \( x \)-axis for easier comparison across different dopant densities, as they would otherwise be located within the same range of \( x \) values.
5-18 Fowler-Nordheim plots for (a) an n-type silicon emitter of dopant density \( N_D = 10^{15} \text{ cm}^{-3} \) and (b) a p-type silicon emitter of dopant density \( N_A = 10^{15} \text{ cm}^{-3} \). The slopes of each of the non-zero linear regions were extracted via linear regression.

5-19 Energy band diagram of an n-type semiconductor emitter with an acceptor band of surface states (hatched area) at the emitter surface and a sample total energy distribution for emitted electrons. In the absence of an applied electric field, the upward bending of the bands is more dramatic. The peaks in the TED labeled as \( j_c, j_s, \) and \( j_v \) denote the contributions from the conduction band, surface state band, and valence band [8].

5-20 Qualitative diagrams comparing the (a) electron density at the surface and (b) emitted current density from the conduction band, as a function applied field, for the present model of field emission in this chapter (dashed lines) and a hypothetical model which incorporates the solution to the current continuity equation inside the emitter (solid lines).

6-1 (a) Infinite circular well of radius \( \rho_1 \), with constant potential \( V = 0 \) inside and (b) cylindrical nanowire emitter of radius \( \rho_1 \), with constant potential \( V = V_1 \) inside, and an electric field of magnitude \( F \) applied to its planar surface along the \( z \) direction.

6-2 Energy band diagram detailing the different energetic regimes of emission from the cylindrical silicon nanowire emitter. The regime to which an electron belongs is determined by the value of its normal energy \( W \) (solid lines) relative to the band edge energy to which the electron belongs, far from the emitter surface. For reference, the total electron energies \( E \) and ground state transverse confinement energies \( E_{0,1}^t \) are also shown (dashed lines).
6-3 The total emitted current density $J$ from (a) n-type cylindrical silicon nanowires of dopant density $N_D = 10^{15}$ cm$^{-3}$ and (b) p-type cylindrical silicon nanowires of dopant density $N_A = 10^{15}$ cm$^{-3}$, as a function of applied field (in vacuum) $F$ and for different nanowire radii $\rho_1$. The total ECD from a bulk emitter of the same dopant density is also shown (dashed line).

6-4 Total conduction band current as a function of applied field $F$ and emitter radius $\rho_1$ for (a) n-type cylindrical silicon nanowire emitters of dopant density $N_D = 10^{15}$ cm$^{-3}$ and (b) p-type silicon nanowire emitters of dopant density $N_A = 10^{15}$ cm$^{-3}$. The total conduction band ECD of the corresponding bulk silicon emitter is plotted as a dashed line.

6-5 As a function of emitter radius, plots of the (a) overall minimum state energy $E_{0,1}^t$ in each of the conduction valley type of the conduction band and in each of the valence bands and (b) the five lowest energies of the infinite circular well, labeled by subband $\{p, r\}$, where the electron mass was taken to be that of a free electron $m_0$.

6-6 The normalized total energy distribution of emitted electrons for various applied field strengths (in vacuum) $F$ from (a) n-type and (b) p-type cylindrical silicon nanowire emitters of radius $\rho_1 = 1$ nm; (c) n-type and (d) p-type cylindrical silicon nanowire emitters of radius $\rho_1 = 2.5$ nm; and (e) n-type and (f) p-type cylindrical silicon nanowire emitters of radius $\rho_1 = 5$ nm. All n-type emitters are of dopant density $N_D = 10^{15}$ cm$^{-3}$, while all p-type emitters are of dopant density $N_A = 10^{15}$ cm$^{-3}$. 
6-7 Plots of the total energy distribution of electrons emitted from the valence band of an n-type cylindrical silicon nanowire emitter of radius \( \rho_1 = 1 \text{ nm} \), for an applied field \( F = 2.3 \text{ V/nm} \). In plot (a), the TED is decomposed into the contributions from each of the valence bands of silicon and in (b) the TED of the heavy hole valence band is decomposed into the contributions from each of the subbands \{p, r\}, with the total TED of the heavy hole band shown by the dashed line.

6-8 Fowler-Nordheim plots for (a) n-type and (b) p-type cylindrical silicon nanowire emitters of dopant density \( N = 10^{15} \text{ cm}^{-3} \) for various emitter radii \( 1 \text{ nm} \leq \rho_1 \leq 10 \text{ nm} \). The FN plots for the corresponding bulk emitters are shown as dashed lines.

6-9 The extracted barrier heights for n-type (blue) and p-type (red) cylindrical silicon nanowire emitters of dopant density \( N = 10^{15} \text{ cm}^{-3} \) from the upper linear region \( H_U \) (circles) and lower linear region \( H_L \) (triangles) of FN plots, as a function of emitter radius.

6-10 Plot of the 20 lowest discrete total energies near the emitter surface in an n-type cylindrical nanowire emitter of dopant density \( N_D = 10^{15} \text{ cm}^{-3} \), for an applied field of \( F = 2.3 \text{ V/nm} \), as a function of emitter radius \( \rho_1 \).

6-11 The average electric field in the silicon \( F_{S,\text{avg}} \) as a function of the applied electric field at the emitter surface in vacuum \( F \), over the course of a single mean free path length of \( l = 70 \text{ nm} \) from the emitter surface, for n-type and p-type silicon of dopant density \( N = 10^{15} \text{ cm}^{-3} \). The dashed line denotes the minimum average electric field in the silicon needed for impact ionization \( F_{\text{min}} \). 
6-12 (a) Diagram of a typical single-gated silicon field emitter structure, with gate aperture radius $R_a$ and applied voltage on the gate electrode $V$ [1]; and (b) diagram of the adaptation of the cylindrical silicon nanowire model for comparison to experimental setups similar to (a). For electron supply purposes, the silicon nanowire is taken to have the same radius $\rho_1$ as the reported radius of curvature at the apex of the actual emitter, and the electric field at the surface is taken to be that of the electric field at the apex of a paraboloidal emitter of radius $\eta_1 = \rho_1$, inside a paraboloidal anode of radius $\eta_2 = R_a$. . . . . . . . . . . . . . . 230

6-13 (a) The experimentally-measured I-V data per tip for n-type FEAs of dopant density $N_D = 2 \times 10^{15}$ cm$^{-3}$ and average tip apex radius of curvature $\rho_1 \approx 5$ nm, for various array sizes. (b) Fowler-Nordheim plots of the data from (a). The predictions of the adapted cylindrical silicon nanowire emitter are shown as red circles, while Stratton’s bulk silicon model emitter predictions are indicated by a dashed, blue line [9].231

6-14 Fowler-Nordheim plots for (a) the prediction of the adapted silicon nanowire model and (b) experimentally-measured I-V data for a 60x60 n-type silicon FEA of dopant density $N_D = 2 \times 10^{14}$ cm$^{-3}$ and estimated average emitter tip radius of curvature of $\rho_1 = 9.84$ nm; and Fowler-Nordheim plots for (c) the prediction of the adapted silicon nanowire model and (d) experimentally-measured I-V data for a 20x20 p-type silicon FEA of dopant density $N_A = 10^{15}$ cm$^{-3}$ and estimated average emitter tip radius of curvature of $\rho_1 = 5$ nm [10]. The extracted FN slope plots $b_{FN}$ are shown in the plot insets and in Table 6.1. . . . . . . . 235
List of Tables

4.1 The extracted work function $\phi_0$ from the $J$ vs. $F_a$ FN plots in Figure 4-21a and the extracted apex field factor $\beta_a$ (with $\phi_0 = 4.5$ eV) from the $I$ vs. $V_0$ FN plots in Figure 4-21b, for each emitter radius. The actual value of the work function used in the calculations was $\phi_0 = 4.5$ eV and the actual field factors for the emitter tips are listed in the fourth row of the table. 144

5.1 The standard values for the effective masses of electrons $m_e$ in the “on-axis” and “off-axis” conduction bands and the effective masses of holes in the three valence bands of silicon $m_v$, in terms of the free electron mass $m_0$. Due to the orientation of the constant energy surfaces of the off-axis conduction valleys, the transverse effective mass of electrons $m_t^{off}$ is the geometrical mean of the normal and transverse effective masses of the on-axis valleys $m_t^{off} = (m_t^{on} m_t^{m})^{0.5}$ [11]. 158

6.1 The extracted FN plot slopes from the valence-band-dominated and conduction-band-dominated emission regimes and the ratio $r$ of the valence band $b_{FN}$ to the conduction band $b_{FN}$, for the adapted silicon nanowire emitter model and the experimentally-measured data of Ding [10]. 236
Chapter 1

Introduction

The physical process of electron emission is an integral aspect of many modern technologies. Injection of electrons into a material from a contact, electronic transport in tunnel field-effect transistors (TFETs), and field electron emission are all technological problems that rely on an electron supply to and transmission through or over a potential barrier. While the dimensions of today’s electronic devices reside in the realm of the nanoscale and may be small enough to manifest the effects of quantum confinement, the majority of the theory in the above areas of research rely on models that assume materials with bulk-like electronic characteristics. A large body of literature investigating the effects of quantum confinement of carriers in TFET devices due to transistor geometry and the confining potential along the channel has recently emerged [12–30], but only a handful of groups have studied quantum confinement effects in field electron emitters [4,31–42]. Instead, significant effort has been devoted to solving electrostatics and tunneling problems that aim to extend existing bulk field emission models to treat emission from highly-curved geometries [2,3,43–58], in order to resolve discrepancies between planar models and experimental results. Although the above studies investigated the effects of quantum confinement and the electrostatics relevant to nanoscale emitters, none have sought to develop a cohesive treatment
that models realistic emitter geometries which incorporate both.

Field electron emitters have primarily been employed as electron sources in vacuum nanoelectronic (VNE) devices, usually in the form of a field emitter array (FEA): a large number of closely-spaced, individual emitters. VNE devices are nanoscale, electronic devices which operate under vacuum conditions, whose current technological applications include terahertz radiation sources [59–62], x-ray radiation sources [63–74], vacuum channel nanotransistors [75–79], and spacecraft propulsion [80]. Since each of these applications requires high current outputs from the FEAs under low applied voltages [81], it is a conventional practice to reduce the dimensions of emitter tips to take advantage of electric field enhancement near highly-curved surfaces, which achieves the same local tip electric fields, but at lower applied voltages. Advances in device fabrication technology over the last few decades have enabled much progress in the scaling down of emitter tip dimensions, with modern field emitters typically having a radius of curvature in the 1-10 nm range, as illustrated by Figure 1-1. Nanoscale field emission devices have also been demonstrated in a variety of materials and geometries, most notably diamond films [82], carbon nanotubes [83], organic and inorganic semiconductor structures [84,85], vertically-aligned
graphene [86], and silicon [1, 7, 9, 87, 88].

Despite this technological progress, theory and modeling of cold field emission have not kept pace. Although specific to describing emission from a bulk, planar, metal emitter, Fowler-Nordheim (FN) theory [89–91] is still routinely used to analyze experimental data from nanoscale field emitters, even though: i) Cutler first reported decades ago that FN theory fails for nanoscopic emitter tip geometries from an electrostatics point of view [47], ii) Forbes found, via his test of “field emission orthodoxy”, that a significant number of emission parameters extracted from FN plots in the literature are spurious [92], and iii) the author has reported that emitters with dimensions below approximately 5 nm may suffer from the effects of quantum confinement, which reduce the emitted current density (ECD) and alter the slopes of FN plots [4]. While the continued use of FN theory is partially a result of its simplicity and accessibility, it is also due to the lack of a better alternative: a coherent physical treatment of cold field emission from realistic emitter tips. Consequently, a unifying and flexible framework for field emission from low-dimensional, non-planar emitter tips is greatly needed.

The primary goal of this work is to establish a semiclassical framework that treats cold field emission from structures of arbitrary dimensionality, geometry, and material. Unlike FN theory, the semiclassical framework incorporates quantum confinement effects and adopts a multidimensional approach to electrons tunneling through the potential barrier. Functionally, the framework provides the tools needed for deriving semi-analytical equations that describe the total energy distribution of emitted electrons and the total ECD from the emitter, through which the impact of emitter size, geometry, and material on field emission processes can be analyzed. Furthermore, since the leading-order semiclassical approximation to the Schrödinger equation is equivalent to the classical Hamilton-Jacobi equation, the semiclassical dynamics of electrons inside and outside the emitter are readily interpreted in terms of classical
mechanics. Additionally, it is hoped that the semiclassical framework will provide insights into current problems in the fields of low-dimensional contacts and TFETs.

This thesis is organized into seven chapters. Chapter 2 discusses the current state of cold field emission theory and reviews significant contributions to the field. In Chapter 3, the semiclassical framework for field emission is developed upon a multidimensional, semiclassical approximation to the wave function that is akin to the one-dimensional Jeffreys-Wentzel-Kramers-Brillouin (JWKB) approximation [93–96]. The semiclassical framework is then applied to a low-dimensional, highly-curved metal emitter tip in Chapter 4, in order to investigate the properties and physics underlying emission from realistic, nanoscale emitter tip geometries. In Chapter 5, a model for emission from bulk, planar silicon emitters is developed, including the effects of quantum confinement of conduction band electrons near the surface and tunneling through the downward-bent forbidden region (band gap) for electrons with energies near the top of the valence band, both of which occur due to field penetration, and an algorithm that limits the maximum conduction band emitted current density to the electron flux density able to be supplied by bulk states. Chapter 6 investigates the relationship between the lateral dimensions (orthogonal to the emitter’s axis of rotational symmetry) of cylindrical silicon nanowire emitters and their emission properties by extending the model of field emission from bulk silicon emitters to include the effects of lateral quantum confinement. Finally, a summary of the thesis work and suggestions for future work are given in Chapter 7.
Chapter 2

Background

2.1 Current State of Field Emission Theory

From a theory and modeling perspective, field electron emitters may be classified according to three primary properties: i) the dimensionality of the emitter’s electron system, ii) the planarity of the emitter surface, and iii) the material of the emitter. Figure 2-1 diagrams these emitter properties and their relationships to the different regimes of field emission theory that have been developed up to the present time. In the diagram, the emitter properties are represented by triangles and the field emission regimes by ovals, while the dashed lines connecting them indicate the emitter properties that define each regime. First chronologically and in the center of the diagram is Fowler-Nordheim (FN) theory [89], which was originally introduced in 1928 and is based upon the assumptions of a bulk, planar, metal emitter. Next came a model for field emission from bulk, planar, semiconductor emitters, developed by Stratton through the early 1960s [97–99], which incorporated emission from the conduction band and valence band, as well as corrections to the emitted current density due to the electron effective mass. As experimentalists succeeded in fabricating ever-smaller field emitter tips through the early 1990s, it became clear that FN theory fails for
The major regimes of field emission theory can be classified via the properties of the emitter: i) electron system dimensionality, ii) emitter surface planarity, and iii) emitter material. The dashed lines connecting the emitter properties (triangles) to the regimes of field emission theory (ovals) represent the assumptions upon which the regime is based: Fowler-Nordheim theory describes emission from a bulk, planar, metal, emitter; sharp emitter theory is based upon a bulk, non-planar, metal emitter; semiconductor emission theory is based upon a bulk, planar, semiconductor emitter; and low-dimensional emission theory is based upon a non-bulk, planar, metal emitter. More general treatments of field emission, such as is the goal of this work, seek to unify multiple regimes of emission theory.

highly-curved emitters [47, 49] and a body of work aiming to extend FN theory to bulk, non-planar, metal emitters emerged [2, 45, 51, 53]. Most recently, studies have focused on modeling the effects of quantum confinement in low-dimensional emitters, mostly in the context of planar, metal emitters [4, 38, 40, 41], but also for the non-planar case [42].

2.2 Fowler-Nordheim Theory

The Fowler-Nordheim (FN) equation serves as the foundation of field emission theory [89]. First introduced in 1928, FN theory utilized the new tools of quantum mechanics to develop a quantum-tunneling-based model that mathematically described the long-
observed phenomenon of field electron emission. Although the details of modern field emission models may have changed, the FN model captures the fundamental physical processes inherent to field emission, namely the transport of electronic flux to the surface and the probability of electrons being transmitted through the potential barrier at the surface, into vacuum. As a result, it has remained an important, relevant, and instructional model for the field emission community over the past 90 years.

2.2.1 Elementary Fowler-Nordheim Equation

The Fowler-Nordheim equation predicts the field emitted current density (ECD) originating from the planar surface of a metal emitter, whose electronic energy spectrum is assumed to be continuous. Fowler and Nordheim’s model consisted of a Sommerfeld-type metal at thermodynamic temperature $T = 0$ K, with an electric field applied normal to a planar surface of the metal [100]. A Sommerfeld-type metal is comprised of an ideal gas of free electrons of mass $m_0$ (the mass of an electron in free space), which obey Fermi-Dirac statistics. At $T = 0$ K, all energy states up to and including the Fermi energy $E_F$ are occupied, while all higher energy states are unoccupied. The momenta of electrons is assumed to be separable into components that are: i) perpendicular to the emitter surface $p_n$, known as the normal momentum, which corresponds to a normal kinetic energy $W$; and ii) parallel to the emitter surface $p_t$, known as the transverse momentum, which corresponds to a kinetic energy $E_t$. With these definitions, the total electron energy $E$ is expressed as

$$E = \frac{p_n^2}{2m_0} + \frac{p_t^2}{2m_0} = W + E_t. \quad (2.1)$$

The emitter surface is assumed to be perfectly planar, smooth, free of defects and to be characterized by a local, uniform work function $\phi$, the minimum energy needed
Figure 2-2: Inside the metal, a constant flux density of electrons \( N \) impinges upon the emitter surface per unit area, per unit energy. Of the incident electronic flux density, only a small fraction \( D \) manages to tunnel through the potential barrier \( V(z) \) into vacuum and contribute to the emitted current density \( J \) (straight arrow), whereas the rest is reflected back into the metal (curved arrows).

to remove an electron from the metal to a point in vacuum just outside the surface of the metal. The applied electric field at the surface \( F \) does not penetrate into the metal and creates an exact triangular potential barrier in the vacuum at the metal surface.

The current density emitted from a state with normal energy \( W \) is given by the product of the supply function \( N(W) \), the flux density of electrons per unit area, per unit energy impinging upon the metal side of the emitter surface; the transmission function \( D(F,W) \), the probability of electrons tunneling through the surface barrier; and the electronic charge \( e \), as illustrated in Figure 2-2. The total emitted current density \( J(F) \) is calculated by integrating over all normal energies \( W \)

\[
J(F) = e \int N(W)D(F,W)\,dW. \tag{2.2}
\]

The supply function is equal to the product of the normal electron group velocity, density of states with respect to normal momentum, density of states with respect to transverse momentum, and the Fermi-Dirac distribution, integrated over all trans-
verse momenta, which gives

\[ N(W) = \frac{4\pi m_0 k_B T}{h^3} \ln \left[ 1 + \exp \left( -\frac{W - E_F}{k_B T} \right) \right], \]  

(2.3)

where \( k_B \) is Boltzmann’s constant and \( h \) is Planck’s constant. The transmission probability of electrons through the triangular potential barrier is calculated using a zeroth order Jeffreys-Wentzel-Kramers-Brillouin (JWKB) approximation [93–96], giving:

\[ D(F,W) \approx \exp \left[ -g_0 \int_0^{z_1} \sqrt{V(z) - W} \, dz \right] \]  

(2.4)

where \( z \) denotes the emission direction (perpendicular to the emitter surface), \( g_0 = 4\pi\sqrt{2m_0}/h \), \( z = z_1 \) and \( z = z_2 \) are the classical turning points of the electron in the classically-forbidden region, and \( V(z) \) is the potential energy outside the emitter, also known as the tunneling barrier. In the case of the exact triangular barrier, \( V(z) = H_0 - eFz \), where \( H_0 = \phi + E_F \) is the barrier height as measured from the bottom of the conduction band, and

\[ D(F,W) = \exp \left[ -\frac{B}{F} H(W)^{3/2} \right], \]  

(2.5)

where \( H(W) = \phi + E_F - W \) is the zero-field barrier height and \( B = 8\pi\sqrt{2m_0}/3eh \) is known as the second Fowler-Nordheim constant. Figure 2-3 is a normal energy diagram for the case of the exact triangular barrier, which depicts the relationship between the normal energies inside the emitter, the Fermi energy, the work function, and the zero-field barrier height. Since most of the emitted electrons originate from states with normal energies close to the Fermi energy, it is a good approximation to expand Equation 2.5 about \( W = E_F \), giving

\[ D(F,W) = \exp \left[ -\frac{B}{F} \phi^{3/2} + c(W - E_F) \right], \]  

(2.6)
Figure 2-3: Normal energy diagram for a bulk, planar, metal emitter with Fermi energy $E_F$, work function $\phi$, and an exact triangular potential barrier, defined by $V(z)$. $H$ is the zero-field barrier height experienced by an electron with normal energy $W$ and in adhering to customary field emission notation, the direction of the electric field $F$ is defined oppositely to that of the usual convention.

where $c = 3B\sqrt{\phi}/2F$. Inserting Equation 2.3 and Equation 2.6 into the expression for the ECD in Equation 2.2, extending the lower limit of integration to $W = -\infty$, and performing the integral over all normal energies $W$ yields the elementary Fowler-Nordheim equation:

$$J_{FN}(F) = A\phi^{-1}F^2 \exp \left[ -\frac{B}{F}\phi^{3/2} \right], \quad (2.7)$$

where $A = e^3/8\pi h$ is the first Fowler-Nordheim constant.

### 2.2.2 Fowler-Nordheim Plots

The Fowler-Nordheim equation established a clear and simple relationship between the emitted current density $J(F)$ and the electric field at the emitter surface $F$, that characterizes field emission processes. More specifically, the FN equation predicts that if field emission data (in I-V form rather than in J-F form) is plotted on a set of axes defined by $\ln[I/V^2]$ for the y-axis and $1/V$ for the x-axis, it should form a straight line. This type of plot is termed a Fowler-Nordheim plot. Assuming that $I$ is linearly proportional to $J$ through the emission area $\Sigma_0$ and $V$ is linearly proportional to $F$ via a constant $\beta$, the equation for the line formed by the data on an FN plot,
given in terms of FN equation parameters is

\[
y = -\frac{B}{\beta V} \phi^{3/2} + \ln \left[ \frac{\Sigma_0 A \phi}{\beta} \right].
\] (2.8)

Depending on which of the above physical quantities is presumed to be known, one can extract either the field factor \( \beta \) or the local work function \( \phi \) (or more generally, the reference barrier height \( H_R \)) from the slope of a linear fit of the data. Less commonly, the emission area \( \Sigma_0 \) is extracted from the regression line offset.

### 2.2.3 Standard Fowler-Nordheim Equation: Murphy-Good Theory

Although instructive, the elementary FN model’s exact triangular barrier shape is physically unrealistic. As a result, the standard form of the Fowler-Nordheim equation that finds wide use today does not assume an exact triangular barrier at the emitter surface, but instead uses the Schottky-Nordheim barrier [90, 101], which is a nearly-triangular barrier whose top is rounded due to the inclusion of an image potential term. The Schottky-Nordheim potential energy, which is depicted in the normal energy diagram in Figure 2-4, takes the form

\[
V(z) = H_0 - eFz - \frac{e^2}{16\pi\epsilon_0 z},
\] (2.9)

where \( \epsilon_0 \) is the permittivity of free space. Inserting the barrier form of Equation 2.9 into Equation 2.4 and evaluating the integral, gives a new form for the transmission function which was first introduced by Murphy and Good [91]:

\[
D(F,W) = \exp \left[ -\frac{B}{F} \nu[y]H(W)^{3/2} \right],
\] (2.10)
where

\[ y = \frac{\sqrt{e^3 F}}{\phi + E_F - W}, \]

\[ \nu[y] = \frac{1}{\sqrt{2}} \left[ 1 + \sqrt{1 - y^2} \right]^{1/2} \left[ E(k^2) - \frac{y^2 K(k^2)}{1 + \sqrt{1 - y^2}} \right], \]

\[ k^2 = \frac{2 \sqrt{1 - y^2}}{1 + \sqrt{1 - y^2}}, \]

(2.11a, 2.11b, 2.11c)

\( K(k^2) \) is the complete elliptic integral of the first kind, \( E(k^2) \) is the complete elliptic integral of the second kind, and \( \nu[y] \) is known as the barrier shape correction factor. Performing the usual expansion of the exponent of the transmission function about \( W = E_F \) and integrating over \( W \) to calculate the ECD results in the standard Fowler-Nordheim equation:

\[ J_{MG}^{FN}(F) = A (\phi^{-1} t^{-2}[y_{EF}] F^2 \exp \left[ -\frac{B}{F} \nu[y_{EF}] \phi^{3/2} \right], \]

(2.12)

where

\[ t[y] = \nu[y] - \frac{2}{3} y \frac{d\nu[y]}{dy} \]

(2.13)

is the second barrier shape correction factor and \( y_{EF} \) indicates that \( y \) should be evaluated at \( W = E_F \). While numerical methods are required to evaluate \( \nu[y] \) and \( t[y] \) exactly, Forbes has derived analytical approximations for them [102]:

\[ \nu[y] \approx 1 - y^2 + \frac{1}{3} y^2 \ln[y], \]

(2.14a)

\[ t[y] \approx 1 + \frac{1}{9} y^2 - \frac{1}{9} y^2 \ln[y]. \]

(2.14b)

As can be seen in Figure 2-4, the introduction of the image charge term into the barrier potential decreases the barrier height for a given normal energy \( W \) compared to the exact triangular barrier. As a result, \( \nu[y] \) for \( y > 0 \) is less than unity, and the ECD for a given \( F \) is greater than that predicted by the elementary FN equation.
2.2.4 Fowler-Nordheim-Type Equations

Beyond the standard Fowler-Nordheim equation exists a class of *Fowler-Nordheim-Type equations*, which generalize FN theory to account for non-ideal effects. The goal of these FN-type equations is to improve the accuracy of emission models as compared with the elementary FN equation, since it has been consistently shown in experimental studies that the elementary FN equation significantly underestimates the measured ECD. Each non-ideality appears in the Fowler-Nordheim-type equation as an appended correction factor, giving the *physically-complete Fowler-Nordheim equation* [103]:

\[
J_{FN}^{PC}(F) = \lambda_T \lambda_B \lambda_P \lambda_T \lambda_T A \phi^{-1} F^2 P \exp \left[ -\frac{B}{F} \mu \phi^{3/2} \right] \tag{2.15}
\]

where the correction factors and their purposes are:

- \(\lambda_T\) → Temperature correction factor: accounts for the case in which the ECD is calculated for \(T \neq 0\) K [91]
- \(\lambda_B\) → Band structure correction factor: incorporates any of a wide variety of non-ideal effects due to the detailed band structure of the emitter, such as non-spherical constant energy surfaces [98, 104–106]
• $P$ and $\lambda_P \rightarrow$ Tunneling prefactor and tunneling prefactor correction factor: $P$ depends on the “ideality” (and shape) of the potential barrier, whereas $\lambda_P$ is a correction factor that arises due to the inclusion of $P(W)$, after integrating over normal energies in the ECD calculation \[107–109\].

• $\mu$ and $\lambda_\tau \rightarrow$ Barrier form and decay rate correction factors: $\mu$ accounts for deviations in the tunneling barrier shape from the exact triangular barrier and $\lambda_\tau$ appears due to the inclusion of $\mu$, after integrating over normal energies in the ECD calculation \[2,90,91\]. Since effects related to the tunneling barrier shape alone are located in the exponential function, all of the correction terms aside from the barrier shape correction factor appear as part of the equation’s pre-exponential term. Under the assumptions made in deriving the elementary Fowler-Nordheim equation, all of the correction factors are replaced by unity. For the case of the standard Fowler-Nordheim equation, $\mu \rightarrow \nu[y]$ and $\lambda_\tau \rightarrow t^{-2}[y]$, whereas the other correction factors are replaced by unity. A more detailed overview of the correction factors and their impacts on emission properties can be found in the work by Forbes \[103\] and in §2.2 of the author’s master’s degree thesis \[41\].

2.3 Emission from Semiconductors

With regard to the electronic properties that are relevant to field electron emission, metals and semiconductors differ in a few significant ways. Although in metals, current conduction is attributed to electrons moving in a single, partially-filled band, semiconductors may conduct electron current via multiple, mostly-empty conduction band valleys and hole current via multiple, mostly-occupied (by electrons) valence band valleys. Between these conduction and valence bands exists a forbidden energy gap, where no electronic states occur intrinsically. Additionally, each of the valleys
Figure 2-5: Constant energy surfaces of [001]-oriented silicon, projected onto the $k_z$ plane, where $k_i$ denotes momentum in the $i$ dimension. Assuming motion in the $z$ direction, conduction valleys lying along the $k_x$ and $k_y$ axes are elliptical with two transverse effective masses $m_t$ and $m_l$ and normal effective mass $m_l$, while conduction valleys lying along the $k_z$ axis are circular, with transverse effective mass $m_t$ and normal effective mass $m_l$.

may be characterized by different effective masses, which depend on the shape and relative orientation of the band structure to the direction of the electron’s momentum, as shown in Figure 2-5 for silicon. Furthermore, although the Fermi energy in a metal lies within the conduction band, in a semiconductor, the Fermi level may be located in the conduction band, in the valence band, or in the forbidden energy gap. Finally, unlike in the case of a perfectly-conducting metal, the finite permittivity $\epsilon_r$ of semiconductors allows the applied electric field to penetrate into the semiconductor, causing the conduction and valence bands to bend near the surface, and requiring the modification of the potential outside the emitter such that \[ V(z) = H_0 - eFz - \left(\frac{\epsilon_r - 1}{\epsilon_r + 1}\right) \frac{e^2}{16\pi\epsilon_0 z}. \] (2.16)

As a result, models for field emission from semiconductors are comparatively complex.

Most analytical treatments of field emission from semiconductors incorporate: i)
emission from each of the conduction band (CB) valleys; ii) emission from each of
the valence bands (VB); iii) band-dependent electron masses via the effective mass
approximation; and iv) alterations in transmission function parameters due to band
bending near the surface [97–99, 110–113]. More specifically, these models usually
provide equations for the emitted current density for the cases in which $E_F$ lies above
the CB edge or in the forbidden energy gap. Occasionally, models also treat emission
from a band of surface states (SS) in the forbidden energy gap [8, 114]. The details
of such analytical models for emission from the CB, VB, and SS are summarized in
the sections below. Since the details of field penetration and band bending near the
semiconductor surface are numerical in nature [115–118] and their effects on emission
can be incorporated into analytical models without explicit calculation, they are not
often included as a part of studies on field emission from semiconductors, although
analytical approximations do exist [115, 116, 118, 119]. For a comprehensive and de-
tailed overview of band bending calculations (and field emission from semiconductors
in general), see the textbook by Modinos [8].

The form of the equations for the ECD from the conduction band and valence band
depend on the location of the Fermi level (assumed to be constant throughout the
semiconductor) with respect to the conduction and valence band edges at the emitter
surface $z_s$. For field emission purposes, if $E_F$ is located above (below) the CB (VB)
edge at the surface, the semiconductor is considered to have a degenerate CB (VB),
and below (above) the CB (VB) edge, the semiconductor is considered to have a non-
degenerate CB (VB). For emission from the CB, the derivations of the ECD equations
in each of these cases are parallel to each other, except for the energy about which the
normal energy $W$ is expanded in the approximate transmission function. Emission
from the valence band is treated here only for the non-degenerate case, due to the
fact that in field emission processes, the CB and VB bend downwards to an extent
such that $E_F$ almost always lies above the VB edge at the surface. Traditionally,
energies in the conduction band are referenced to the conduction band edge at the emitter surface, while energies in the valence band are referenced to the valence band edge at the emitter surface, but are measured as positive, downward from the valence band maximum. With regard to surface states, although multiple bands likely exist, only one is assumed to exist above the VB maximum (in the forbidden energy gap) and to contribute significantly to the ECD.

2.3.1 Emission from the Conduction Band: \( E_F > E_C(z_s) \)

If a semiconductor is highly doped n-type or if band bending causes the Fermi level at the surface to be located above the CB edge at the surface \( E_C(z_s) \), as shown in Figure 2-6, the semiconductor is treated as metallic. For this case, the transmission function in Equation 2.5 is expanded around \( W = E_F \), giving the emitted current density [8]:

\[
J_C(F) \approx e^{\frac{4\pi m_0 \exp \left[-b_1\right]}{h^3}} \frac{1}{c_1^2} \exp \left[-b_1 \theta_s^b \right] \times \left\{ 1 - \exp \left[ -c_1 \theta_s^b \right] + \frac{\exp \left[ -c_1 \theta_s^b \right]}{1 - r_c} - \frac{\exp \left[ -r_c c_1 \theta_s^b \right]}{1 - r_c} \right\} \text{ for } \theta_s^b > 0, r_c \neq 1
\]

(2.17a)

\[
J_C(F) \approx e^{\frac{4\pi m_0 \exp \left[-b_1\right]}{h^3}} \frac{1}{c_1^2} \exp \left[-b_1 \theta_s^b \right] \times \left\{ 1 - \exp \left[ -c_1 \theta_s^b \right] - c_1 \theta_s^b \exp \left[ -c_1 \theta_s^b \right] \right\} \text{ for } \theta_s^b > 0, r_c = 1
\]

(2.17b)

where

\[
\theta_s^b = E_F - E_C(z_s), \quad (2.18a)
\]

\[
r_c = m_c / m_0, \quad (2.18b)
\]

\[
b_1 = \frac{B}{F} \nu[y_0^b] (\chi_e - \theta_s^b)^{3/2}, \quad (2.18c)
\]
Figure 2-6: Energy band diagram for an n-type emitter that is degenerate in the CB at the surface, where $E_0$ is the vacuum level, $E_C$ denotes the bottom of the conduction band, $E_g$ is the magnitude of the forbidden energy gap, $E_F$ is the Fermi level, $E_V$ is the top of the valence band, and $z_s$ is the location of the emitter surface. Like in FN theory, most of the ECD originates from states close in energy to $E_F$.

\[ c_1 = \frac{3B}{2F} t[y_{\theta_s}^b](\chi_e - \theta_s^b)^{3/2}, \quad (2.18d) \]

\[ y_{\theta_s}^b = \frac{\sqrt{\epsilon_r}F}{\chi_e - \theta_s^b} \left( \frac{\epsilon_r - 1}{\epsilon_r + 1} \right)^{1/2} \quad (2.18e) \]

$m_c$ is the effective mass of electrons in the conduction valley (assuming spherical constant energy surfaces), $\chi_e$ is the electron affinity of the semiconductor, $F$ is the magnitude of the electric field on the vacuum side of the interface, $\nu[y]$ and $t[y]$ are given by Equation 2.11b and Equation 2.13 from Murphy-Good theory, and $\epsilon_r$ is the relative permittivity of the semiconductor. It should be noted that in the limit of $c_1 \theta_s^b \gg 1$, which corresponds to $E_F \gg E_C(z_s)$ and/or $\chi_e \gg \theta_s^b$, both of the above expressions for the ECD simplify to the standard Fowler-Nordheim equation, with a work function equal to $\phi = \chi_e - \theta_s^b$. 

52
2.3.2 Emission from the Conduction Band: $E_F < E_C(z_s)$

When a semiconductor emitter is non-degenerate in the CB at the surface, either due to doping or applied electric field strength, the Fermi level lies below the conduction band edge at the surface, as shown in Figure 2-7. As a result, the conduction band at the surface is comparatively sparsely occupied and the occupation probability of states decreases rapidly with increasing energy. In this case, due to the exponential decay of the density of states in the conduction band, the transmission function in Equation 2.5 is expanded about the energy in the CB nearest to $E_F$: $W = E_C(z_s) = 0$. This results in the following ECD equation for emission [8]:

$$J_C(F) \approx e \frac{4\pi m_0}{\hbar^3} (k_BT)^2 \exp \left[ \frac{\theta_s}{k_B T} \right] \exp [-b_0] \text{ for } \theta_s < 0,$$

where

$$b_0 = \frac{B}{F} \nu[y_0] \chi_e^{3/2},$$

$$y_0 = \frac{\sqrt{e^3 F}}{\chi_e} \left( \frac{\epsilon_r - 1}{\epsilon_r + 1} \right)^{1/2}.$$

Unlike the ECD equations for emission from a degenerate CB, Equation 2.19 is temperature dependent and predicts that emission from the CB ceases when $T \to 0$ K, due to electrons lacking the thermal energy required for promotion from the VB to the CB.

2.3.3 Emission from the Valence Band: $E_F > E_V(z_s)$

In the case of emission from a non-degenerate valence band, the Fermi level $E_F$ lies above the VB edge at the emitter surface $E_V(z_s)$, which corresponds to both Figure 2-6 and Figure 2-7. Unlike the case of emission from a non-degenerate CB, the normal energy about which the transmission function is expanded is not determined by the density of states (approximately unity throughout the VB), but by the states that
Figure 2-7: Energy band diagram for an n-type emitter, which is non-degenerate at the surface. Most of the ECD originates from states at the bottom of the conduction band.

have the highest probability of tunneling through the barrier, which are located at the top of the valence band. Accordingly, the transmission function is expanded about the valence band maximum at the surface \( W = E_V(z_s) \), giving the ECD equation [8]

\[
J_V(F) = e \frac{4\pi m_0}{h^3} \exp \left\{ -\frac{b_{v0}}{c_{v0}^2} \right\} \left\{ \frac{1}{1 + r_v} - \frac{(c_{v0}k_B T)^2}{1 + \frac{E_g + \theta_s^b}{k_B T}} \right\} \text{ for } \theta_s^b > -E_g,
\]

(2.21)

where

\[
r_v = \frac{m_v}{m_0},
\]

(2.22a)

\[
b_{v0} = \frac{B}{F} \nu[y_{v0}](\chi_e + E_g)^{3/2},
\]

(2.22b)

\[
v_{v0} = \frac{3B}{2F} t[y_{v0}](\chi_e + E_g)^{1/2},
\]

(2.22c)

\[
y_{v0} = \frac{\sqrt{e^3 F}}{\chi_e + E_g} \left( \frac{\epsilon_r - 1}{\epsilon_r + 1} \right)^{1/2},
\]

(2.22d)

\( m_v \) is the effective mass of that particular valence band (assuming spherical constant energy surfaces), \( E_g \) is the magnitude of the forbidden energy gap, and it has been assumed that \( c_{v0}k_B T \ll 1 \). To a good approximation, emission from the valence band
under these conditions is independent of temperature, as Fermi-Dirac statistics are nearly temperature independent for $E \ll E_F$ and $T \approx 300$ K.

### 2.3.4 Emission from Surface States

Surface states arise in a semiconductor due to the termination of the infinite, periodic crystal by a surface and are treated as a separate band of states that are localized at the semiconductor surface. Although multiple surface state bands likely exist in the semiconductor, only a band that exists above the VB edge at the surface, as shown in Figure 2-8, contributes significantly to the ECD. In this model for field emission from surface states, a single band is assumed to exist in the forbidden energy gap, with energy

$$E_s = E_s^0 + \frac{\hbar^2 k_{||}^2}{2m_s}, \tag{2.23}$$

where $E_s^0$ is the minimum energy of the surface band, $k_{||}$ are a continuum of wave vectors parallel to the surface, and $m_s$ is the effective mass of electrons in the surface
band (assuming circular constant energy surfaces). The density of states in this band is two-dimensional and given by

\[ N_s = \frac{m_s}{\pi \hbar^2} \text{ for } E_s^0 < E_s < E_s^0 + \Gamma_s \]  

(2.24)

where \( \Gamma_s \) denotes the width of the surface state band in energy. Figure 2-8 illustrates the location of the surface state band in the context of a field emission energy band diagram. Modinos gives the total energy distribution of electrons emitted from the surface state band to be [8]

\[ j_s(E_s, \theta_s) = \frac{h}{2\pi m_0} N_s Q^2 \exp\left[-b(E_s - r_s(E_s - E_s^0))\right], \]  

(2.25)

where \( r_s = m_s/m_0 \) and \( Q \) is an effective wave vector that is treated as an adjustable parameter. Integrating the total energy distribution over all energies in the surface state band yields the ECD from the surface states:

\[ J_s(\theta_s^*, F) = e \int_{E_s^0}^{E_s^0 + \Gamma_s} j_s(E_s, \theta_s) dE_s \]  

(2.26)

where \( \theta_s^* = \theta_b(F_s) - \Delta \theta_s^* \), where \( \Delta \theta_s^* \) denotes the difference in the surface state occupation from that of the bulk state occupation. Although it is assumed that in steady state that the electrons emitted from the conduction or valence bands are immediately replaced by electrons from the crystal interior, this is not necessarily the case for electrons emitted from the surface state band. As a result, the electrons of the bulk bands and the surface state bands are not in thermal equilibrium with each other; however, the surface state electrons can be in “pseudoequilibrium” among themselves, giving rise to a description of their occupation using \( \theta_s^* \) instead of \( \theta_b^\circ \). For the case in which \( J_s \) is small, it is assumed that the surface states are replenished immediately and are in thermal equilibrium with the bulk band electrons, making
Figure 2-9: Metal-oxide-semiconductor (MOS) structure, where the applied field has caused a quantum well (in the accumulation layer) and bound states to form near the Si-SiO₂ interface.

$\theta^s_s \approx \theta^b_s$. In general though, $\theta^s_s$ and $\theta^b_s$ can be different even in the zero-current approximation and must be calculated self-consistently in order to obtain an exact solution.

### 2.3.5 Emission from Accumulation Layer Bound States

When an electric field is applied to the surface of an n-type semiconductor to induce field emission, it penetrates into the semiconductor and bends the bands downwards near the surface, forming an accumulation layer. Classically, this accumulation layer is characterized by a region of very high electron density, as compared with the bulk semiconductor far away from the surface. Quantum mechanically, the penetrating electric field forms a quantum well, bounded on one side by the tunneling barrier at the surface and on the other side by the CB edge (equivalently, the forbidden energy gap), which causes a two-dimensional electron gas to form near the surface, as shown in Figure 2-9. Although not a case of field electron emission, Rana has treated a strikingly similar problem: the tunneling of electrons from bound states in the accumulation layer of a MOSFET through a thin oxide (trapezoidal barrier), into a metallic gate electrode [120].
The total emitted current density from the accumulation layer is composed of contributions from the bound states $J_{\text{bnd}}$, which are the states confined to the accumulation layer well near the surface, and the extended states $J_{\text{ext}}$, which are states that extend throughout the bulk crystal. While emission from the extended states in the conduction band is identical to that of emission from bulk silicon for $E_F < E_C(z_s)$ in Equation 2.19, the current density emitted from the bound states is given by

$$J_{\text{bnd},n}(F) = e \left( \frac{4\pi m_z k_B T}{\hbar^2} \right) \sum_n \frac{1}{\tau_n(E)} \delta_n \ln \left[ 1 + \exp \left( \frac{E_F - E_n}{k_B T} \right) \right]$$  \hspace{1cm} (2.27)

where $n$ indexes the electronic subbands, $\delta_n$ is the Kronecker delta function, $m_z$ is the effective mass of the electron in the tunneling dimension, $E_n$ is the discrete subband energy, and $\tau_n$ is the lifetime of the state in the accumulation layer well, given by

$$\frac{1}{\tau_n(E)} = \frac{D(F,E)}{\int_0^{z_n} \sqrt{2m_z/|E_n - E_C(F,z)|} \, dz},$$  \hspace{1cm} (2.28)

where $D(F,E)$ is the probability of an electron tunneling through the potential barrier, $z_n$ is the location in the semiconductor where $E_n = E_C(F, z_n)$ (classical turning point) for subband $n$, and $E_C(F, z)$ is the conduction band edge. Summing the current density from all subbands $n$ results in the total bound state current density $J_{\text{bnd}}$.

### 2.4 Emission from Highly-Curved Emitter Tips

It has been reported in several studies that the Fowler-Nordheim equation fails to accurately describe field emission from highly-curved emitter tips [47, 49, 92]. From an electrostatics point of view, this is a direct result of the non-planarity of the emitter surface, which gives rise to an enhanced electric field at the emitter apex and a non-triangular barrier potential. Due to electric field enhancement, in which the magnitude of the local electric field at curved surfaces is increased relative to
Figure 2-10: A non-planar protrusion between two parallel contacts. The macroscopic field $F_M$ far away from the emitter tip is equivalent to the constant electric field between the contacts of a parallel plate capacitor, while the surface field magnitude at the apex of the protrusion $F$ is enhanced due to the protrusion’s curved geometry. The ratio between $F$ and $F_M$ is the field enhancement factor at the apex $\gamma_a$.

The field far from the surface, the field which appears in the FN equation may be much larger than in the assumed planar emitter case and may lead to higher emitted currents than predicted by FN theory. As a consequence of the non-triangular barrier, the ECD equation’s dependence on the surface field is no longer proportional to $F^2 \exp[-1/F]$, which causes FN plots become non-linear and to be a less reliable tool for parameter extraction from experimental field emission data [92]. In order to remedy these shortcomings of FN theory, two approaches have been adopted: i) the incorporation of a field enhancement factor; and ii) the extension of FN theory to non-planar emitters. Additionally, there also exist treatments of field emission from highly-curved tips that are external to the FN framework, as they are based on a multidimensional, semiclassical tunneling model instead of the one-dimensional tunneling model inherent to FN theory [44, 57, 121–123].

### 2.4.1 Field Enhancement Factors

The magnitude of the electric field that appears in Fowler-Nordheim theory is that of the local field at a planar emitter surface. However, due to the geometry of the
emitter tip, the surface field magnitude is generally different from that of the field far away from the emitter surface, which is sometimes referred to as the macroscopic field $F_M$ [52], as shown in Figure 2-10. The field at the surface of the emitter and the macroscopic field can be related by introducing a field enhancement factor $\gamma$, such that $F = \gamma F_M$. If the assumption is made that the overwhelming majority of the emitted current originates from the emitter apex, the surface field in the Fowler-Nordheim theory is replaced by

$$F = \gamma_a F_M,$$

(2.29)

where $\gamma_a$ is the field enhancement factor at the emitter apex, which depends on the specific geometry of the emitter. For example, the field enhancement factor at the apex of a hemisphere on a grounded plane, as shown in Figure 2-10, is $\gamma_a = 3$. For an overview of typical models for field emitter geometries and their associated field enhancement factors, see the work by Forbes [52]. For the case in which one is working with voltages instead of electric fields, as in experimental studies of field emission, an analogous parameter called the field factor $\beta$, which converts the applied voltage into an effective surface field, is defined such that $F = \beta_a V$.

### 2.4.2 Generalized Fowler-Nordheim Equation for Non-Planar Emitters

While the field enhancement factor approach to treating non-planar emitters has been widely employed [52] and may help close the gap between the current densities predicted by standard FN theory and the much higher, measured experimental ECDs, it does not solve the problem of non-linear FN plots. Due to the field enhancement factor (or field factor) actually depending on the applied electric field, the slope of FN plots is not constant and FN theory still does not apply in these cases [47, 49]. The failure of field enhancement factors to describe emission from highly-curved emitters
Figure 2-11: The numerically-calculated tunneling barrier potential energy $U(z)$ for an ellipsoidal surface, as a function of the emitter radius of curvature. As can be seen, the potential barrier shape depends on the emitter’s radius of curvature $R$ and differs appreciably from the Schottky-Nordheim barrier (FN) [2].

is a manifestation of the larger problem: the tunneling barrier shape near the surface of highly-curved field emitters is non-triangular and its exact shape depends on both the applied voltage and the tip’s radius of curvature [2,47,49], as depicted in Figure 2-11.

Recently, Kyritsakis and Xanthakis have developed an extension of FN theory that applies to nanoscopic emitters by using a generalized potential for the tunneling barrier that adds an additional 2nd-order term to the Schottky-Norhdeim barrier, which results from the emitter curvature [2]. The generalized potential in the vacuum is given by

$$V(z) = H_0 - eFz - \frac{e^2}{16\pi\epsilon_0 z(1+z/2R)} + \frac{eF}{R}z^2,$$

where $R$ is the radius of curvature of the emitter at the apex, which is depicted in Figure 2-12. Using JWKB theory to evaluate the electron transmission probability through the barrier within the Murphy-Good framework and calculating the emitted current density yields the ECD equation [2]:

$$J(F) = A\phi^{-1}F^2 \left( t[y] + \frac{\phi}{eFR} \psi[y] \right)^{-2} \exp \left[ -\frac{B\phi^{3/2}}{F} \left( \nu[y] + \frac{\phi}{eFR} \omega[y] \right) \right]$$
where \( y = 2\sqrt{BF/\phi} \), and \( \nu[y] \) and \( t[y] \) are the same functions as in the Murphy-Good framework,

\[
\omega[y] \approx \frac{4}{5} - \frac{7y^2}{40} - \frac{y^2 \ln[y]}{100}, \quad (2.32a)
\]

\[
\psi[y] = \frac{5}{3} \omega[y] - \frac{2}{3} y \frac{d\omega}{dy} \approx \frac{4}{3} - \frac{y^2}{500} - \frac{y^2 \ln[y]}{15}, \quad (2.32b)
\]

and \( \omega[y] \) and \( \psi[y] \) were approximated via an expansion that is equivalent to that for \( \nu[y] \) and \( t[y] \) by Forbes [102]. Based on the linearity of FN plots and comparisons of extracted \( R \) values from experimental data with SEM images of emitter tips, Equation 2.31 is valid for emitters with radii of curvature greater than approximately 5 nm, for surface fields in the range \( 3 \text{ V/nm} < F < 10 \text{ V/nm} \), and for work functions in the range \( 3 \text{ eV} < \phi < 5 \text{ eV} \). Also of note, in the limit of \( \phi/eFR \to 0 \), which corresponds to the case of a very large radius of curvature, Equation 2.31 reduces to the standard FN equation.

### 2.4.3 Semiclassical, Multidimensional Tunneling

Other investigations into modeling emission from highly-curved emitters have taken another approach, abandoning the one-dimensional tunneling theory of the FN model altogether. Since FN theory is based upon a planar cathode, the shape of the tunneling barrier is uniform at all points across the emitter surface, the electric field always points along the tunneling direction, and as a result, electrons experience a constant force only along the emission direction. Under these circumstances, one-dimensional
Figure 2-13: Highly-curved emitter tip and anode, along with equipotential lines and the direction of the electric field $F$ at selected points in the vacuum. Due to the variation in the direction of the field $F$ as a function of position outside the emitter, electrons are acted upon by forces in multiple dimensions and may travel along multidimensional paths. A few example paths for electrons emitted from different locations on the emitter surface are shown and labeled as 1, 2, and 3.

tunneling theory suffices. For curved surfaces, however, the surfaces of constant potential in the vacuum away from the emitter surface are not parallel to the emitter surface, which leads to differing tunneling barrier shapes, tunneling barrier widths, and surface field strengths at different points on the emitter surface. Consequently, the electric field does not point normal to the emitter surface at all locations in the vacuum, as illustrated in Figure 2-13. The differing direction of the electric field exerts forces on the electrons that cause them to travel in multiple dimensions and leads to a coupling between different components of the electrons’ momentum. Thus, the tunneling problem cannot be separated into independent, one-dimensional problems in each coordinate and a multidimensional treatment of tunneling is necessary.

In the case of tunneling in one dimension, the point of entry to and point of exit from the classically-forbidden region and the tunneling path between them are unique. In the case of multidimensional tunneling, there are infinitely many possibilities for
pairs of entry and exit points, and between each pair of points exist infinitely many
tunneling paths, each of which correspond to a finite probability of an electron tun-
neling through the classically-forbidden region. A few such tunneling paths, between
a pair of entry and exit points, are depicted in Figure 2-15. As a result, the procedure
for an exact solution to the multidimensional tunneling problem demands that a pair
of points be chosen and that the tunneling probability of all paths between them be
summed to find the total tunneling probability. This is precisely what is prescribed
by Feynman’s path integral formulation of quantum mechanics [124]. Repeating this
procedure for each pair of entry and exit points yields the total tunneling probability
between all pairs of points in the system. Fortunately, the complexity of the calcula-
tion can be greatly reduced if only a single path (or a few paths) which contributes
most significantly to the tunneling probability is considered. This is the core idea
behind the semiclassical approximation to quantum tunneling, in which a single path
is assumed to be the lone contributor to the total tunneling probability. The single,
dominant path coincides with the classical path of the particle and can be calculated
via one of the formulations of classical mechanics, as reviewed in Appendix A. Stated
in other words, under the semiclassical approximation, the probabilistic nature of
particles in quantum mechanics is exchanged for the deterministic behavior of classi-
cal particles. A review of semiclassical tunneling theory can be found in the work by
Takatsuka [125].

Zeroth Order Semiclassical Wave Function

The set of approaches taken to treating multidimensional tunneling which are directly
relevant to field emission are semiclassical in nature and reduce to the JWKB approx-
imation when considering the 1D Schrödinger equation. The most complete semiclas-
sical solution to the multidimensional Schrödinger equation is given by Huang [3],
which follows a derivation procedure parallel to that first introduced by Kapur and
Figure 2-14: Relative orientation of the wave fronts (surfaces of constant $W_{0R}$ and $W_{0I}$), along with the trajectories corresponding to the real $\hat{\mathbf{e}}_R$ (R paths) and imaginary $\hat{\mathbf{e}}_I$ (I paths) components of the electron’s wave vector [3]. Under the assumption that $k_R(\mathbf{r}) = 0$, only I paths exist, which are always normal to surfaces of constant $W_{0I}$ and, equivalently, the turning surfaces.

Peierls [126]. Starting with the three-dimensional Schrödinger equation, the wave function is assumed to have a pure-exponential form and is expanded in a power series in $\hbar$, such that

$$\Psi(\mathbf{r}) = \exp \left[ \frac{i}{\hbar} W(\mathbf{r}) \right] \approx \exp \left[ \frac{i}{\hbar} W_0(\mathbf{r}) + W_1(\mathbf{r}) + ... \right], \quad (2.33)$$

which, when inserted into the Schrödinger equation and terms equal in $\hbar$ are collected, yields the lowest order semiclassical approximation to the wave function phase

$$[\nabla W_0(\mathbf{r})]^2 = 2m_0[E - V(\mathbf{r})], \quad (2.34)$$

where $E$ is the total energy of the state and $V(\mathbf{r})$ is the potential energy function of the system. In general, $W_0(\mathbf{r})$ may be complex, such that $W_0(\mathbf{r}) = W_{0R}(\mathbf{r}) + iW_{0I}(\mathbf{r})$. 
and using this substitution, Equation 2.34 can be separated into a pair of equations:

\[
[\nabla W_{0R}(r)]^2 - [\nabla W_{0I}(r)]^2 = 2m_0[E - V(r)], \quad (2.35a)
\]

\[
[\nabla W_{0R}(r)] \cdot [\nabla W_{0I}(r)] = 0. \quad (2.35b)
\]

While the zeroth order approximation to the wave function phase \( W_0 \) in a classically-allowed region (where \( E - V > 0 \)) must be purely real (requiring \( \nabla W_{0I} = 0 \)), in the classically-forbidden region through which tunneling occurs (where \( E - V < 0 \)), \( W_0 \) may be complex. As shown in Figure 2-14, in the classically-forbidden region, \( W_{0R}(r) \) forms equiphase surfaces and \( W_{0I}(r) \) forms equiamplitude surfaces, whose gradients give orthogonal components of the electron’s wave vector \( \mathbf{k} \):

\[
\nabla W_{0R}(r) = \hbar \mathbf{k}_R(r) \tag{2.36a}
\]

\[
\nabla W_{0I}(r) = \hbar \mathbf{k}_I(r) \tag{2.36b}
\]

where \( \mathbf{k}_R(r) \) is parallel to and \( \mathbf{k}_R(r) \) is perpendicular to the turning surfaces, which separate the classically-allowed region from the classically-forbidden region. The orientation of these components of \( \mathbf{k} \) lend the interpretation that \( \mathbf{k}_I \) corresponds to the component of the electron’s momentum that is oriented along the electron’s direction of travel just outside the emitter surface in the vacuum and that \( \mathbf{k}_R \) corresponds to the component of the electron’s momentum that is oriented transverse to the direction of travel just outside the emitter surface in the vacuum. In terms of the components of \( \mathbf{k} \), the zeroth order wave function phase in the forbidden region is

\[
k_R^2(r) - k_I^2(r) = k(r)^2 = \frac{2m_0}{\hbar^2}[E - V(r)], \quad (2.37)
\]
which can be calculated by a step-wise, Huygens-like procedure explained in the work by Huang et al. [121]. Equation 2.37 can be greatly simplified if it is assumed that \( k_R(r) = 0 \), which restricts the possible tunneling paths to only those which are perpendicular to the emitter surface. This simplified approach was taken by Das and Mahanty in order to derive the tunneling probability for electrons through a multidimensional potential barrier [44]. Under this set of assumptions, the zeroth order wave functions in the classically-allowed \( \Psi^0_a(r) \) and classically forbidden-region \( \Psi^0_f(r) \) are given by

\[
\Psi^0_a(r) = A \exp \left[ i \sqrt{\frac{2m_0}{\hbar^2}} \int_{s_1}^{s_2} \sqrt{E - V(r)} \, ds \right] \quad (2.38a)
\]

\[
\Psi^0_f(r) = A \exp \left[ - \sqrt{\frac{2m_0}{\hbar^2}} \int_{s_1}^{s_2} \sqrt{V(r) - E} \, ds \right] \quad (2.38b)
\]

where \( A \) is a normalization constant, \( ds \) is the infinitesimal line element along the electron’s path of travel \( s \). In the forbidden region, the electron’s path \( s \) is determined by solving the equations of motion with \( E \to -E \) and \( V(r) \to -V(r) \), or equivalently, by varying the time parameter \( t \) along the imaginary axis in the complex \( t \) plane, such that \( t \to i\tau \) [44,127].

**Tunneling Probability**

Figure 2-15 depicts a typical tunneling situation, in which Region I and Region III are classically-allowed regions and Region II is the classically-forbidden region through which tunneling may occur. The transmission probability through the classically-forbidden region results from comparing the probability flux density exiting the classically-forbidden region at \( r_2 \) to the probability flux density entering the forbidden region at \( r_1 \). Taking the ratio of the exiting flux to the incident flux, the tunneling
Figure 2-15: An electron tunneling from a classically-allowed Region I at point \( r_1 \), through a classically-forbidden Region II, into a classically-allowed Region III at point \( r_2 \) has infinitely many possible tunneling paths available, a few of which are shown. Under the semiclassical approximation, only the most probable path \( s \) (dashed line) is considered, which coincides with the classical path of the particle, as determined by the classical equations of motion. \( \mathbf{e} \) is the unit tangent vector at a point along the tunneling path.

The probability is [44]

\[
D(r_1, r_2, E) = \left[ \frac{|dV(r)/ds|_{s=s_2}}{|dV(r)/ds|_{s=s_1}} \right]^{1/2} \exp[-T_{12}],
\]

(2.39)

where

\[
T_{12} = \frac{2}{\hbar} \int_{s_1}^{s_2} \sqrt{2m_0 [V(r) - E]} ds,
\]

(2.40)

and \( V(r) \) is the tunneling barrier potential energy. The tunneling probability is then calculated by choosing one endpoint (either \( r_1 \) or \( r_2 \) in Figure 2-15), calculating the path \( s \) of the electron and the other endpoint via the classical equations of motion, and evaluating the integral in Equation 2.40 along \( s \). In addition to the work of Das and Mahanty, a few other studies of field emission from highly-curved tips have been based upon this semiclassical multidimensional model of tunneling, mostly in the context of microscopy [57, 122, 123]. A more detailed derivation of the tunneling probability via the model of Das and Mahanty is provided in §3.3.2.
2.5 Emission from Low-Dimensional Emitters

Modern field emitter tips have been fabricated with lateral dimensions (orthogonal to the emission direction at the emitter apex) and radii of curvature that reach well into the nanoscale regime. When the dimensions of an electron system are sufficiently reduced, the electronic spectrum separates from a quasicontinuum of states into discrete energy levels. However, the traditional models for field emission, namely the FN model for metals and Stratton’s model for semiconductors, are based upon a bulk electron system with a quasicontinuous energy spectrum. In order to investigate and understand emission from nanoscale emitters, it is necessary to employ models which also incorporate the effects of quantum confinement.

The first elementary models for field emission from low-dimensional structures were motivated by describing emission from the faces [38] and edges [40] of two-dimensional graphene sheets. These studies introduced the basic ideas behind incorporating discretized energy states into field emission models and provided tools, such as the emission reference level, to facilitate the derivation of emitted current density equations for non-bulk electron systems. Based on these earlier works, the author developed a framework for modeling field emission from low-dimensional emitters, which unified the elementary FN equation and two-dimensional ECD equations, and provided the theoretical machinery needed to derive ECD equations for emitters of other dimensionalities [4,41,42]. At the core of the elementary framework for cold field emission from low-dimensional electron systems is a discrete ECD equation, which calculates the total ECD by summing over the contributions from each electronic state $Q$:

$$J(F) = e^2 \sum_0 \sum_Q f_{FD}(E_Q, T) \int \int D_Q(F, s) \left| \mathbf{j}_Q(s) \cdot \hat{n} \right| d\Sigma,$$

(2.41)

where $E_Q$ is the total energy of state $Q$, $\Sigma_0$ is the surface area of the emitter, $f_{FD}(E_Q, T)$ is the Fermi-Dirac distribution function, $D_Q(F, s)$ is the transmission
probability for electrons, \( s \) is the set of points that comprise the emitter surface, \( j_Q(s) \cdot \hat{n} \) is the electronic flux density incident normal to the metal side of the tunneling barrier, and \( d\Sigma \) is the area element on the emitter surface. Both the probability flux density \( j_Q(r) \), given by the expression for the probability current

\[
j_Q(r) = \frac{\hbar}{2m_0} \left[ \psi_Q^*(r) \nabla \psi_Q(r) - \psi_Q(r) \nabla \psi_Q^*(r) \right], \tag{2.42}
\]

and the transmission probability for each state \( D_Q(F,s) \), which follows the same formulation as in Equation 2.4, are based on a one-dimensional JWKB approximation for the wave function (see Appendix A in [42]). If the electronic spectrum of states associated with a particular degree of freedom can be approximated by a quasicontinuum, the sum over the corresponding quantum numbers in Equation 2.41 is converted to an integral via an appropriate density of states in energy.

Via the elementary framework, ECD equations for emitters of various geometries and orientations were derived, as shown in Figure 2-16, [4, 41, 42]. In addition to the overall geometry and dimensionality of the emitter, the orientation of the emitter surface to the dimensions of quantum confinement of the emitter are specified and fall into two categories: emitters that are \textit{normally unconfined} are characterized by
Figure 2-17: The emitted current density as a function of emitter dimensions, normalized to the elementary Fowler-Nordheim equation, for (a) normally unconfined and (b) normally confined field emitters of various dimensionalities and geometries. [4].

A quasicontinuum of momenta along the emission direction (Figures 2-16a, 2-16b, and 2-16c), while emitters that are normally confined are characterized by discrete momentum states along the emission direction (Figures 2-16d and 2-16e). Whether or not an emitter is normally unconfined or normally confined dictates the approach for calculating the incident flux density $j_Q(r)$. Although the probability flux density for normally unconfined emitters is calculated from Equation 2.42, the wave functions along the emission dimension of a normally confined emitter formally contribute zero incident probability flux density at the surface, since the wave functions are taken to be zero at the emitter surface under the assumption of an infinite potential outside the emitter. As a solution, the semiclassical method of Gurney is employed [128], which models the electrons as classical particles rebounding between the walls of a quantum well, incident at a frequency $\nu$ given by

$$\nu_n(F) = \left( \int_0^{z_n} \sqrt{2m_z/E_n - E_c(F,z)} \, dz \right)^{-1}. \tag{2.43}$$

The product of this tunneling attempt frequency and the spatial density of states in dimensions transverse to the emitter surface gives the incident flux density for normally confined emitters.
Figure 2-18: The normalized emitted current density for a cylindrical nanowire emitter, with the field enhancement factor of a hemispherical cap, as a function of emitter diameter \( d \). At a critical radius, the reduction in the electron supply due to quantum confinement overcomes the gains in transmission probability due to electric field enhancement and the ECD decreases as a function of decreasing emitter diameter [4].

In addition to deriving ECD equations for emitters of specific geometries and dimensionalities, the effects of lateral and normal (along the emission direction) quantum confinement on the emitted current density were studied. Figure 2-17a demonstrates that for normally unconfined emitter geometries, the ECD gradually decreases to zero from a value close to that predicted by the elementary FN equation as the emitter dimensions shrink, with a significant ECD decrease for emitters with lateral dimensions below 5 nm. The decrease in the ECD is directly a result of the sparser density of states in the emitter, which is caused by the increasing energy spacing between discrete energy levels, as lateral emitter dimensions decrease. For the normally confined emitter case, Figure 2-17b shows that the ECD oscillates as a function of normal emitter dimensions, with the oscillations gradually increasing in magnitude as the emitter dimensions decrease. The oscillations arise from the highest-energy subband below \( E_F \) contributing the majority of the ECD, and as normal emitter dimensions vary, the location of its energy relative to \( E_F \) changing and tracing out the individual oscillations [4, 41]. Also investigated was the interaction between the
competing influences of a reduced electron supply due to lateral quantum confinement and the increased transmission probability due to electric field enhancement at the apex. Figure 2-18 shows a plot of the normalized ECD vs. emitter diameter for a normally unconfined cylindrical nanowire, with the field enhancement factor of a hemispherical cap. As depicted, there exists a critical radius which separates the field-enhancement-dominated regime, in which the ECD increases as a function of decreasing emitter diameter due an increased transmission probability due to electric field enhancement, from the quantum-confinement-dominated regime, in which the gains from electric field enhancement are overcome by the lack of electrons available for emission and the ECD decreases with decreasing emitter diameter.

2.6 Chapter Summary

The different formulations of field emission theory that have been developed up to this point can be classified by the primary properties of the emitter systems they describe: i) dimensionality of the electron system, ii) the planarity of the emitter surface, and iii) the emitter material. The foundational theory of field emission, developed by Fowler and Nordheim for bulk, planar, metal emitters in 1928, captures the most fundamental aspects of field emission processes and has served as the basis for field emission theory for the last 90 years. Over the years, FN theory has been extended to account for different tunneling barrier shapes, emitter material non-idealities, and highly curved emitter surfaces, while also underlying Stratton’s theory of emission from semiconductors, as well as recent studies on modeling emission from low-dimensional emitters. However, each of these regimes of field emission theory investigates only a single deviation of real emitter properties from the FN model at a time and there currently exists no unified model of field emission that describes emission from emitters of arbitrary dimensionality, emitter surface geometry, and
material. The development of such a model for field emission is detailed in Chapter 3.
Chapter 3

Semiclassical Framework for Field Electron Emission

3.1 Introduction

As was discussed in Chapter 2, modern field emitters have been demonstrated in a diversity of dimensionalities, geometries, and materials. While separate regimes of field emission theory have been developed to address deviations of modern field emitter properties from those of the Fowler-Nordheim (FN) model, each regime only considers the effect of a single deviation from the FN model at a time. For example, Stratton’s theory of emission from semiconductors still assumes a bulk electron system and a planar emitter surface; Das and Mahanty’s model of emission from a scanning tunneling microscope tip [44] and Kyritsakis’ model for emission from nanoscopic emitters [2] preserve the bulk, metal electron system of the FN model; and the author’s framework for cold field emission from low-dimensional emitters applies to metal emitters with planar emitter surfaces (or geometries suited to a one-dimensional treatment of electron tunneling through the barrier) [4,41,42]. If one wants to study how parameters such as emitter size, geometry, and material affect the properties of modern
emitters, which are routinely nanoscale, semiconductor structures with highly-curved tips, a more general theoretical formulation of field emission that can unify the aforementioned field emission regimes is needed.

This chapter details the development of a semiclassical framework for field electron emission, which serves as the foundation of this thesis. The framework is based upon a semiclassical approximation to the multidimensional Schrödinger equation originally introduced by Kapur and Peierls [126] and incorporates the effects of quantum confinement, as well as multidimensional tunneling from non-planar emitter surfaces. The primary purposes of the framework are to provide a flexible, theoretical scaffolding which can be used to derive emitted current density (ECD) equations (and consequently total energy distribution (TED) equations) and to analyze the properties of nanoscale emitters of various dimensionalities, geometries, and materials. Since the framework is built up from contributions from individual energy states, employs a multidimensional model for tunneling from non-planar emitter surfaces, and can be applied to both metals and semiconductor material systems, it also serves as a fundamental formulation of field emission which unifies the existing regimes of field emission theory. Additionally, due to the nature of the semiclassical approximation, the Schrödinger equation reduces to a pair of equations that are well-known in classical physics: the classical Hamilton-Jacobi equation and the continuity equation. As a result, the long-established formulations of classical mechanics and fluid dynamics are readily adapted to help form intuitive physical understandings of the properties of nanoscale emitters.

In the sections below, the mathematical details of the semiclassical approximation to the Schrödinger equation are laid out, which provides expressions for the zeroth and first order approximations to the wave function. The semiclassical framework for field electron emission is subsequently constructed by introducing a fundamental form of the emitted current density equation and deriving expressions for each of its
3.2 Semiclassical Solution to the Multidimensional Schrödinger Equation

Although the first semiclassical approximation to the one-dimensional Schrödinger equation was developed independently by Wentzel, Kramers, and Brillouin in 1926 (based on a more general mathematical formulation by Jeffreys in 1925) [93–96], it wasn’t until over a decade later that a comparable semiclassical approximation to the multidimensional Schrödinger equation was introduced by Kapur and Peierls [126]. A semiclassical treatment of field emission is particularly appealing for a few reasons. Primarily, a semiclassical solution to the multidimensional Schrödinger equation often permits analytical (in the worst case, semi-analytical) solutions for the wave functions, which are later used to derive the relevant physical quantities for field emission calculations. This often obviates large amounts of numerical calculations. Secondly, a semiclassical treatment is compatible with previous treatments of field emission, namely the FN model for metals and Stratton’s model for semiconductors. Both of these models are based upon free-electron-like models for the electron system and capture the most basic physics of field emission. Finally, to zeroth order in the semiclassical approximation, the Schrödinger equation is equivalent to the classical
Hamilton-Jacobi equation, which enables interpretation and analysis of the properties of the quantum system in the more intuitive context of classical mechanics.

In multiple dimensions, the Schrödinger equation is

\[-\frac{\hbar^2}{2m} \nabla^2 \Psi(r, t) + V(r, t)\Psi(r, t) = i\hbar \frac{\partial}{\partial t} \Psi(r, t), \quad (3.1)\]

where \( \hbar = h/2\pi \) is the reduced Planck constant, \( \Psi(r, t) \) is the wave function, \( V(r, t) \) is the potential energy function, and the gradient can be expressed in generalized, curvilinear coordinates as

\[\nabla f(r) = \sum_{i=1}^{N} \frac{1}{h_i(r)} \frac{\partial f(r)}{\partial q_i} \hat{q}_i, \quad (3.2)\]

where \( q_i \) are the \( N \) generalized coordinates, \( i = 1, \ldots, N \) indexes the generalized coordinates, \( h_i(r) \) is the coordinate scale factor associated with \( q_i \), and \( \hat{q}_i \) is the unit vector along coordinate \( q_i \).

Although there are various, equivalent procedures for implementing the semiclassical approximation to the Schrödinger equation, the derivation that is often used to arrive at de Broglie-Bohm theory provides a very clear connection between semiclassical quantum and classical mechanics [130]. If the wave function is expressed in a polar form, such that

\[\Psi(r, t) = R(r, t) \exp \left[ \frac{i}{\hbar} S(r, t) \right], \quad (3.3)\]

and is inserted into the Schrödinger equation, a non-linear form of the Schrödinger equation results. This equation can be split into two equations: one depending only on the real terms and one depending only on the imaginary terms. The equation resulting from equating the real terms is

\[\frac{\partial S}{\partial t} = - \left[ \frac{(\nabla S)^2}{2m} + V - \frac{\hbar^2}{2m} \nabla^2 R \right], \quad (3.4)\]
which is the *quantum Hamilton-Jacobi equation* (HJE), since it is equivalent to the classical Hamilton-Jacobi equation, except for the addition of the term

\[ Q(r, t) = -\frac{\hbar^2}{2m} \frac{\nabla^2 R(r, t)}{R(r, t)}, \tag{3.5} \]

which is known as the *quantum potential*. The quantum potential \( Q \) is often interpreted as an informational potential, which contains information about the whole experimental setup and guides the motion of the quantum particle through mediating an instantaneous, non-local force on it [131]. The equation that results from the imaginary terms is

\[ \frac{\partial R}{\partial t} = -\frac{1}{2m} \left[ R \nabla^2 S + (2 \nabla R) \cdot (\nabla S) \right], \tag{3.6} \]

which, when the wave function amplitude is expressed in terms of its density \( R(r, t) = \rho^{1/2}(r, t) \) and \( \mathbf{v}(r, t) = \nabla S(r, t)/m \) is identified as the velocity, is more readily recognizable as the *continuity equation*

\[ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \tag{3.7} \]

In order to effect the semiclassical approximation, Equations 3.4 and 3.6 are evaluated in the limit in which \( Q \) is negligible compared to the other terms. While the continuity equation remains unchanged, the quantum HJE reduces to

\[ \frac{(\nabla S)^2}{2m} + V = -\frac{\partial S}{\partial t} \tag{3.8} \]

where the quantum potential \( Q \) of Equation 3.4 has vanished, yielding the *classical Hamilton-Jacobi equation*. The zeroth order wave function results from solving the classical HJE for the classical action \( S \), which is the wave function phase. The
first order wave function is then determined by inserting the solution for $S$ into the continuity equation and solving for $R$, the wave function amplitude.

### 3.2.1 Zeroth Order Wave Function: Classical Hamilton-Jacobi Equation

Via classical mechanics, $\nabla S$ is identified as the momentum $\mathbf{p}(\mathbf{r}, t)$ and the left-hand side of Equation 3.8 is the Hamiltonian for the system $H(\mathbf{p}, t)$. If the Hamiltonian is time-independent, such that $H = E$, where $E$ is the total energy of the system, the Hamilton-Jacobi equation (HJE) can be written as

$$
(\nabla S)^2 = 2m \left[ E - V(\mathbf{r}) \right].
$$

The solution to the time-independent HJE is

$$
S(\mathbf{r}) = \int \sqrt{2m \left[ E - V(\mathbf{r}) \right]} ds
$$

where $ds$ is the line element along the path of the electron $s$. Under the semiclassical approximation, the path that the electron takes $s$ is governed by the classical equations of motion, either the Euler-Lagrange equations of Lagrangian mechanics in §A.2 or, equivalently, Hamilton’s equations of Hamiltonian mechanics in §A.3. Additionally, details for obtaining separable solutions to the HJE can be found in §A.5.2. The resulting wave function to zeroth order in the classically-allowed region $\Psi^0_a(\mathbf{r})$ and classically-forbidden region $\Psi^0_f(\mathbf{r})$ is

$$
\Psi^0_a(\mathbf{r}) = A \pm \exp \left[ \pm \frac{i}{\hbar} \int \sqrt{2m \left[ E - V(\mathbf{r}) \right]} ds \right] = A \pm \exp \left[ \pm \frac{i}{\hbar} \int \mathbf{p}(\mathbf{r}) ds \right].
$$

$$
\Psi^0_f(\mathbf{r}) = A \exp \left[ -\frac{1}{\hbar} \int \sqrt{2m \left[ V(\mathbf{r}) - E \right]} ds \right] = A \exp \left[ -\frac{1}{\hbar} \int \mathbf{p}(\mathbf{r}) ds \right].
$$

80
where $A_{\pm}$ and $A$ are normalization constants and $p(r)$ is the magnitude of the electron’s momentum.

### 3.2.2 First Order Wave Function: Continuity Equation

Since the classical action $S$ only appears in Equation 3.6 in the form of its gradient, the time-independent continuity equation can be rewritten in terms of the momentum as

$$\frac{1}{2} \frac{(\nabla \cdot p)}{p} + \frac{\nabla R}{R} = 0,$$

(3.12)

where $\nabla S(r) = p(r) = p(r)\hat{e}(r)$ is the momentum of the electron, $p(r)$ is the magnitude of the momentum of the electron, and $\hat{e}(r)$ is the unit tangent vector along the path of the electron $s$. Solving for $R$ gives the amplitude of the wave function

$$R(r) = \frac{1}{\sqrt{p(r)}} \exp \left[ -\frac{1}{2} \int (\nabla \cdot \hat{e}) \, ds \right],$$

(3.13)

where the integral is evaluated along the path of the electron $s$. The integral over the divergence of the unit tangent vector of the path of the electron can be expressed in a more intuitive way if one instead considers a tube of neighboring classical electron paths with cross sectional area $\sigma(s)$ at a point $s$ along the path, as shown in Figure 3-1 [3]. In the limit of very small $\sigma$,

$$(\nabla \cdot \hat{e}) \sigma(s) \, ds = (\nabla \cdot \hat{e}) \, dV$$

(3.14)

where $ds$ is an infinitesimal displacement along the path of the electron and $dV$ is an infinitesimal change in the volume of the tube. Applying Gauss’s theorem yields

$$(\nabla \cdot \hat{e}) \, dV = \sigma(s + ds) - \sigma(s) = \frac{\partial \sigma}{\partial s} \, ds$$

(3.15)
Figure 3-1: Depiction of a tube of neighboring electron paths and its change in cross-sectional area $\sigma(s)$ as a function of position along the paths $s$.

which is the change in the cross-sectional area of the tube over the course of moving along the path a distance $ds$. This leads to the relationship

$$ (\nabla \cdot \hat{e}) ds = \frac{(\partial \sigma / \partial s)}{\sigma(s)} ds $$  \hspace{1cm} (3.16)

which, when inserted into the integral of Equation 3.13, yields a new expression for the wave function amplitude

$$ R(r) = \frac{1}{\sqrt{p(r)\sigma(r)}}, $$ \hspace{1cm} (3.17)

and the resulting wave function to first order in the classically-allowed $\Psi^1_a(r)$ and classically-forbidden region $\Psi^1_f(r)$ is

$$ \Psi^1_a(r) = \frac{A_\pm}{\sqrt{p(r)\sigma(r)}} \exp \left[ \pm \frac{i}{\hbar} \int \sqrt{2m} [E - V(r)] ds \right] $$ \hspace{1cm} (3.18a)

$$ = \frac{A_\pm}{\sqrt{p(r)\sigma(r)}} \exp \left[ \pm \frac{i}{\hbar} \int p(r) ds \right], $$

$$ \Psi^1_f(r) = \frac{A}{\sqrt{p(r)\sigma(r)}} \exp \left[ -\frac{1}{\hbar} \int \sqrt{2m} [V(r) - E] ds \right] $$ \hspace{1cm} (3.18b)

$$ = \frac{A}{\sqrt{p(r)\sigma(r)}} \exp \left[ -\frac{1}{\hbar} \int p(r) ds \right]. $$
In addition to the familiar $p^{-1/2}$ pre-exponential dependence that is also present in the first order one-dimensional JWKB approximation’s wave function, is an inverse-square-root dependence on the cross-sectional area of the tube of classical paths. While the first order wave function experiences a singularity when $p = 0$, due to the electron encountering a turning surface $s_t$ (a surface comprised of turning points), a singularity also occurs when $\sigma = 0$, which corresponds a tube with cross-sectional area of zero. When the cross-sectional area of the tube vanishes, the considered electron paths converge to a single point, which is known as a caustic [3]. As a result, in addition to the turning surfaces, there exist an additional set of caustic surfaces where the wave function’s value approaches infinity and the semiclassical approximation to first order is not valid.

### 3.3 Emitted Current Density Equation

The emitted current density (ECD) equation forms the core of the semiclassical framework for field emission. It not only provides a method for calculating the total current emitted by a system and the total energy distribution (TED) of the emitted electrons, but also serves as a structure for investigating how the physical processes inherent to field emission determine or influence emission properties. As illustrated in Figure 3-2, the emitted current density consists of two major physical quantities: i) the incident current density $\eta_Q$, the current per unit area in the material, impinging upon the tunneling barrier, and ii) the transmission probability of an electron through the potential barrier at the emitter surface $D_Q$. In the case of a non-planar emitter surface, both the incident current density and transmission probability vary as a function of position on the emitter surface. As a result, it is necessary to make a distinction between the local emitted current density from state $Q$, $J_Q$, and the average emitted current density from state $Q$, $\bar{J}_Q$. Whereas the local ECD is the current density
Figure 3-2: Schematic of a grounded, non-planar emitter tip (cathode) and anode, held at a potential $V = V_0$. At each location on the emitter surface $s$, the internal current density impinging on the emitter surface $\eta_Q$ has a probability of being transmitted through the potential barrier $D_Q$, resulting in a local emitted current density $J_Q$. Electrons tunnel through the classically-forbidden region between the emitter surface $s$ (solid curve) and the turning surface $s_t$ (dashed curve), along a path $s$.

emitted at a single point on the emitter surface $s$ and varies as a function of position on the emitter surface, the average ECD is the total current emitted from the tip, divided by the emitter surface area. The local ECD contributed by state $Q$ is

$$J_Q(V_0, T, s) = \eta_Q(V_0, T, s)D_Q(V_0, s),$$

(3.19)

where $V_0$ is the potential difference between the cathode and anode and $T$ is the thermodynamic temperature of the electron system. The current contributed by state $Q$ is determined by integrating the local ECD across the entire emitter surface

$$I_Q(V_0, T) = \int \int J_Q(V_0, T, s) d\Sigma$$

(3.20)

where $d\Sigma$ is the area element of the emitter surface. The contribution to the average ECD from state $Q$ is given by

$$J_Q(V_0, T) = \frac{I_Q}{\Sigma_0},$$

(3.21)
where $\Sigma_0$ is the emitter surface area. The total emitted current $I$ and average ECD $J$ are calculated by summing over the contributions from all states, giving

$$I(V_0, T) = \sum_Q I_Q(V_0, T), \quad (3.22a)$$

$$J(V_0, T) = \sum_Q J_Q(V_0, T). \quad (3.22b)$$

Each of the physical quantities that must be determined to calculate the ECD from an emitter system, such as the incident current density $\eta_Q$, electron tunneling probability $D_Q$, and approximate bound state energies $E_Q$ are discussed and derived in the sections below.

### 3.3.1 Incident Current Density

The incident current density contributed by state $Q$ at a point $s$ on the surface, is given by

$$\eta_Q(V_0, T, s) = 2e f_{FD}(E_Q, T) [j_Q(V_0, s) \cdot \hat{n}(s)], \quad (3.23)$$

where $e$ is the elementary charge, the factor of 2 accounts for the two-fold spin degeneracy of each state $Q$, $E_Q$ is the total energy of electrons in state $Q$, $j_Q$ is the *incident probability flux density* at $s$ for electrons in state $Q$, $\hat{n}$ is the unit normal vector of the emitter surface, and $f_{FD}$ is the Fermi-Dirac distribution function [132,133],

$$f_{FD}(E_Q, T) = \left\{ 1 + \exp \left[ \frac{E_Q - E_F}{k_B T} \right] \right\}^{-1}, \quad (3.24)$$

where $E_F$ is the Fermi energy and $k_B$ is the Boltzmann constant. The incident probability flux density measures the rate of probability flow through a surface per unit time, per unit area and, in the case of complex wave functions, can be determined
directly from the wave function $\Psi_Q$ via the expression [134]

$$j_Q(s) = \frac{\hbar}{2m_i} \left[ \Psi_Q^*(r) \nabla \Psi_Q(r) - \Psi_Q(r) \nabla \Psi_Q^*(r) \right] \bigg|_{r=s}. \quad (3.25)$$

In some special cases in which the potential in the emitter varies rapidly compared with the wavelength of the wave function, it may be more appropriate to adopt the method of Gurney for calculating the incident probability flux density [128]. Gurney’s model assumes that an electron in a quantum well is constantly in motion, rebounding between the walls of the well and attempting to tunnel each time it strikes the tunneling barrier. The rate $\nu_Q$ at which an electron in state $Q$ strikes the tunneling barrier at a location $s$ on the emitter surface is determined via

$$\nu_Q(V_0, E_Q, s) = \frac{1}{\left[ \int_{s_1}^{s_2} \sqrt{2m/(E_Q - V(r))} \, ds \right]^{-1}}, \quad (3.26)$$

where $s$ is the path of the electron’s orbit as it rebounds between the walls of the quantum well, $s_1$ and $s_2$ are classical turning points of the electron in the well, and $V(r)$ is the potential inside the emitter. Multiplying $\nu_Q$ by the appropriate areal density of states in dimensions transverse to the emitter surface $\rho_t$ at $s$ provides an alternative equation for the incident probability flux density

$$j_Q(V_0, s) = \nu_Q(V_0, E_Q, s) \rho_t(E_Q, s). \quad (3.27)$$

### 3.3.2 Electron Transmission Probability

Figure 3-3 shows a diagram of an electron, normally incident on the turning surface between the classically-allowed region and forbidden region, tunneling along a path $s$ from classically-allowed Region I, through forbidden Region II, into classically-allowed Region III. Although an approximation, assuming that the electrons attempting to tunnel are normally incident on the tunneling barrier significantly simplifies the cal-
Figure 3-3: An electron, normally incident on the turning surface between the classically-allowed region and forbidden region, tunneling through the forbidden Region II \((E < V(r))\) follows a path \(s\), which is governed by the classical equations of motion. The tunneling probability is determined by finding the ratio of the probability flux density exiting the forbidden region \(J_{III}(r_2)\) to the probability flux density incident on the forbidden region \(J_1(r_1)\).

Calculation of the tunneling probability through the forbidden region. Under this constraint, the action \(S(r)\) is purely imaginary in the forbidden region, allowing for the tunneling paths to be determined independently of each other, via modified classical equations of motion. Starting at the point \(r_0\) in Region I and traveling to another point \(r_\alpha\) along the path in Region I gives the action

\[
S(r_\alpha) = S(r_0) + \int_{s_0}^{s_\alpha} \sqrt{2m(E - V(r))} ds, \tag{3.28}
\]

and corresponding wave function at \(r_\alpha\)

\[
\Psi(r_\alpha) = A \exp \left[ \frac{i}{\hbar} S(r_0) + \frac{i}{\hbar} \int_{s_0}^{s_\alpha} \sqrt{2m(E - V(r))} ds \right], \tag{3.29}
\]

where the \(S(r_0)\) is the initial value of \(S\) at \(r_0\), \(s_0\) and \(s_\alpha\) are the distance traveled along the path \(s\) at points \(r_0\) and \(r_\alpha\) respectively, and \(A\) is a normalization constant. If the electron continues along the path and is not reflected at \(r_1\), it travels into the forbidden region, Region II. Within the forbidden region, the electron may also
experience reflections at $r_2$ after which it will travel back along the path towards $r_1$, where it can, in turn, be reflected again. If, after $n$ reflections at $r_2$, the electron is transmitted into Region III to a point $r_\beta > r_2$, the action is given by

$$S(r_\beta) = S(r_0) + \int_{s_0}^{s_1} \sqrt{2m[E - V(r)]}ds + i(2n + 1) \int_{s_1}^{s_2} \sqrt{2m[V(r) - E]}ds$$

$$+ \int_{s_2}^{s_\beta} \sqrt{2m[E - V(r)]}ds,$$

(3.30)

where the first integral is the contribution to the action from traveling along the path in Region I from $r_0$ to $r_1$; the second integral is the contribution to the action from traveling along the path in Region II from $r_1$ and $r_2$, plus the round trip distance between $r_1$ and $r_2$ due to $n$ reflections at $r_2$; and the last integral is the contribution to the action for the distance traveled from $r_2$ to $r_\beta$ after being transmitted into Region III. The corresponding wave function at $r_\beta$ is then given by

$$\Psi(r_\beta) = \sum_{n=0} B_n \exp \left[ \frac{i}{\hbar} S(r_0) + \frac{i}{\hbar} \int_{s_0}^{s_1} \sqrt{2m[E - V(r)]}ds + \frac{1}{\hbar} \int_{s_1}^{s_2} \sqrt{2m[V(r) - E]}ds + \frac{i}{\hbar} \int_{s_2}^{s_\beta} \sqrt{2m[E - V(r)]}ds \right],$$

(3.31)

where $B_n$ are constants. Since the wave function in the forbidden region decays exponentially with total distance traveled by the electron, its magnitude is significantly reduced with each reflection at $r_2$. As a result, only the case in which there are no reflections at $r_2$ ($n = 0$) is assumed to contribute significantly to the tunneling current, giving the approximate wave function

$$\Psi(r_\beta) \approx B_0 \exp \left[ \frac{i}{\hbar} S(r_0) + \frac{i}{\hbar} \int_{s_0}^{s_1} \sqrt{2m[E - V(r)]}ds + \frac{1}{\hbar} \int_{s_1}^{s_2} \sqrt{2m[V(r) - E]}ds + \frac{i}{\hbar} \int_{s_2}^{s_\beta} \sqrt{2m[E - V(r)]}ds \right].$$

(3.32)

The calculation of the transmission probability reduces to finding the ratio of the current density leaving the forbidden region at the exit point $J_{III}(r_2)$ to the current
density incident upon the forbidden region at the entry point $J_I(r_1)$, as shown in Figure 3-3. The expressions for the flux density in Region I and Region III are given by inserting the wave functions from Equation 3.32 and Equation 3.29 into the probability flux density expression of Equation 3.25, yielding

$$J_I(r_1) = \sqrt{\frac{2}{m}} [E - V(r_1)] \hat{e}(r_1), \quad (3.33a)$$

$$J_{III}(r_2) = \sqrt{\frac{2}{m}} [E - V(r_2)] \exp[-T_{12}] \hat{e}(r_2), \quad (3.33b)$$

where $\hat{e}(r)$ is the unit vector tangent to the electron’s path $s$ at $r$,

$$T_{12} = \frac{2}{\hbar} \int_{s_1}^{s_2} \sqrt{2m [V(r) - E]} ds, \quad (3.34)$$

and $s_1$ and $s_2$ are the locations of the entry and exit points of the forbidden region along the electron’s path $s$. Unfortunately, simply dividing the exiting flux density by the incident flux density at the turning points results in an undefined value for the tunneling probability, due to the singularities at the turning points $r_1$ and $r_2$. However, this problem can be avoided if the potential is linearized along the path about some point $s = s_0$, such that

$$V(r) \approx V(s_0) + \frac{\partial V}{\partial s} \bigg|_{s=s_0} \Delta s + ... \quad (3.35)$$

where $\Delta s$ is a small displacement along the path. Inserting the linearized potential into the expressions for the probability flux densities and taking the ratio of the exiting to the incident flux density yields the result of Das and Mahanty for the tunneling probability [44]

$$D(r_1, r_2) = \frac{J_{III}(r_2)}{J_I(r_1)} \approx \left[ \frac{(dV(r)/ds)|_{s=s_2}}{(dV(r)/ds)|_{s=s_1}} \right]^{1/2} \exp \left[ -g_e \int_{s_1}^{s_2} \sqrt{V(r) - E} ds \right], \quad (3.36)$$

89
where \( g_e = 2\sqrt{2m_0/\hbar} \).

In order to calculate the tunneling probability, one must determine the path that the electron takes through the forbidden region. The definition of the action in Equation 3.9 can be redefined in terms of the velocity \( \frac{dr}{dt} \) as

\[
\left( \frac{dr}{dt} \right)^2 = \frac{2}{m} [E - V(r)],
\]

which is a form of the equation of motion that gives the electron’s trajectory in the classically-allowed region. On the other hand, in the forbidden region, no classical trajectory is possible since the electron’s momentum is imaginary, which corresponds to an evanescent wave rather than a traveling wave. It is, however, possible to define a formal trajectory for the electron in the forbidden region, which is a continuation of the trajectory from the classically-allowed region into the forbidden region, if the electron is orthogonally incident on the barrier. This is accomplished by varying the time parameter \( t \) along the complex axis, which results in a new time parameter \( \tau \) defined by \( t = i\tau \) and gives the equation

\[
\left( \frac{dr}{d\tau} \right)^2 = \frac{2}{m} [V(r) - E],
\]

which is equivalent to the real time \( t \) motion of an electron in a potential \(-V(r)\), with total energy \(-E\). Under this time transformation, Equation 3.38 gives a minimized action in the forbidden region and, as a result, the trajectory in the forbidden region can be determined from the classical equations of motion with \( V(r) \rightarrow -V(r) \) and \( E \rightarrow -E \) [44].

In summary, the process for calculating the tunneling probability is as follows:

1. Fix either an entry point to or exit point from the forbidden region \( r_1 \), which lies on one of the turning surfaces \( s \) (emitter surface) or \( s_t \) (turning surface in vacuum).
2. Calculate the trajectory of the electron via the equations of motion with 
\[ V(\mathbf{r}) \rightarrow -V(\mathbf{r}) \text{ and } E \rightarrow -E, \] 
from \( r_1 \) to the point at which the electron encounters a turning surface and leaves the forbidden region at \( r_2 \).

3. Evaluate the integral between \( r_1 \) and \( r_2 \) along the path \( s \) in Equation 3.36 and calculate \((dV/ds)\) at the turning points \( r_1 \) and \( r_2 \) to determine the tunneling probability.

### 3.3.3 Bohr-Sommerfeld Quantization

In addition to providing approximate solutions for the wave functions, the semi-classical approximation also provides estimates for the bound state energies via the *Bohr-Sommerfeld quantization condition* [130]. Within the classically-allowed region of the emitter, the overall wave function consists of a forward-traveling and backward-traveling wave

\[
\Psi_Q(\mathbf{r}) = A_+ R(\mathbf{r}) \exp \left[ \frac{i}{\hbar} S(\mathbf{r}) \right] + A_- R(\mathbf{r}) \exp \left[ -\frac{i}{\hbar} S(\mathbf{r}) \right],
\]

where \( A_+ \) and \( A_- \) are the normalization constants for the forward- and backward-traveling waves. Imposing the boundary condition that the wave function must vanish at the two endpoints \( r_1 \) and \( r_2 \) gives the equations

\[
\Psi(r_1) = 0 = A_+ + A_-,
\]

\[
\Psi(r_2) = 0 = A_+ \exp \left[ \frac{i}{\hbar} S(r_2) \right] + A_- \exp \left[ -\frac{i}{\hbar} S(r_2) \right],
\]

where it has been chosen, without any loss of generality, that \( S(r_1) = 0 \). Combining the above equations results in the expression

\[
\Psi(r_2) = 0 = \sin \left( \frac{S(r_2)}{\hbar} \right) = \sin \left[ \frac{1}{\hbar} \int_{r_1}^{r_2} p(E_Q) \cdot dq \right],
\]
Figure 3-4: While the infinite square well potential (a) is characterized by two hard boundary conditions, the harmonic oscillator potential (b) is characterized by two soft boundary conditions.

where \( p(E_Q) \) is the momentum vector of the electron as a function of its total energy \( E_Q \) and \( dq \) is the differential displacement vector. As a result, Equation 3.41 establishes that in order for the overall wave function to be a standing wave, the phase of the wave function must satisfy

\[
(n - n_0)\pi = \frac{1}{\hbar} \int_{r_1}^{r_2} p(E_Q) \cdot dq = \int_{s_1}^{s_2} \sqrt{\frac{2m}{\hbar^2}} \sqrt{[E_Q - V(r)]} ds, \tag{3.42}
\]

where \( n \) is a positive integer which corresponds to the number of half wavelengths of the wave function that lie between the endpoints and \( n_0 \) is a constant that is determined by the “hardness” of the boundary conditions, such that \[134\]

\[
n_0 = \begin{cases} 
0 & \text{for 2 hard boundaries} \\
\frac{1}{4} & \text{for 1 soft and 1 hard boundary} \\
\frac{1}{2} & \text{for 2 soft boundaries.} 
\end{cases} \tag{3.43}
\]

While a hard boundary condition dictates that the wave function must vanish at the turning points, a soft boundary condition requires only that the wave function there be finite, as shown in Figure 3-4.

This process for quantization is also known as the old quantum condition and was used extensively for calculating energy levels of quantum systems, such as the Hydrogen atom, during the early years of quantum mechanics [130]. Like other aspects
of quantum mechanics in the semiclassical limit, the Bohr-Sommerfeld quantization
condition is very closely related to and can be derived from analogous concepts in
classical mechanics. An alternative derivation of the Bohr-Sommerfeld quantization
condition, based on a classical system that is described in terms of action-angle co-
ordinates, can be found in §A.6.2.

3.3.4 Complete Expression for Emitted Current Density

The complete form of the emitted current density equation is expressed as

\[ J(V_0, T) = 2e \sum_{Q} \int \int \left[ j_Q(V_0, s) \cdot \hat{n}(s) \right] D_Q(V_0, s) d\Sigma. \]  (3.44)

It should be noted, that in contrast to Fowler-Nordheim (FN) theory, the emitted
current density is not considered to be a function of the local electric field at the
surface \( F \), but instead, a function of the applied potential difference \( V_0 \). Unlike the
nearly-triangular, one-dimensional tunneling barrier models of FN theory, the field at
the surface of a non-planar emitter tip varies as a function of position on the emitter
surface and is usually not a suitable proxy for the shape of the barrier potential away
from the surface. As a result, the shape of the tunneling barrier is entirely determined
by the geometries of the cathode and anode, as well as the applied voltage \( V_0 \).

3.4 Model Assumptions and their Consequences

In the course of deriving the framework, several approximations and assumptions
were made in the interest of reducing the complexity of the model and retaining its
semi-analytical nature. While these simplifications are not expected to affect the
major outcomes predicted by the model in most cases, the potential consequences of
their omission are discussed below.
3.4.1 Quantum Potential

Due to invoking the semiclassical approximation, the quantum potential $Q$ in the quantum Hamilton-Jacobi equation (Equation 3.4) is neglected when solving for the wave functions. Before considering the effects of omitting the quantum potential $Q$ though, the role of the quantum potential should be discussed. The quantum potential does not appear in classical mechanics and, consequently, its presence accounts for most of the differences between classical and quantum mechanics. Unlike in the Copenhagen Interpretation of quantum mechanics, in de Broglie-Bohm theory, the properties of the quantum system are determined causally, not by probability theory. In de Broglie-Bohm theory, a quantum particle has a continuous, causally-determined trajectory and is inseparable from its quantum wave field $\Psi$, which satisfies the Schrödinger equation and influences how the particle behaves [135]. As in classical mechanics, the momentum in de Broglie-Bohm theory is defined as

\[ \mathbf{p} = (\nabla S), \tag{3.45} \]

however, the force acting upon the particle (the time derivative of the momentum) in de Broglie-Bohm theory contains an additional term compared to the classical case, the quantum potential $Q$:

\[ \frac{d\mathbf{p}}{dt} = -\nabla (V + Q), \tag{3.46} \]

resulting in a force with both a classical and quantum component, each of which play a role in determining the particle’s trajectory. The form of the quantum potential of Equation 3.5 reveals that since it depends on how the wave function varies in space (via the gradient operator), it mediates an instantaneous, non-local force on the particle. In addition, its $1/R(r)$ dependence explains its ability to exert strong, long-range effects on the particle that do not drop off with distance, such as quantum entanglement. Functionally, the quantum potential contains active information.
about the entire experimental arrangement, which modifies the behavior of the particle instantaneously through Equation 3.46, requiring that a quantum system be treated as a connected whole. As an example, in the double-slit experiment, it is the quantum potential that is responsible for producing the observed interference pattern, as even though each electron passes through only one slit, its trajectory is affected by information from both slits [136].

In the case of a single-particle quantum system, omitting the quantum potential $Q$ when solving for the wave functions neglects the non-local information about the experimental arrangement or environment. As a result, any connections between the quantum particle and distant features of the system are lost and the system behaves as if it were composed of individual, non-interacting parts and classical mechanics is recovered. Whether or not this is a valid approximation for the given system under consideration can be checked by inserting the wave function amplitude $R$ into the quantum potential, which gives the condition for the first order wave function [135]

$$-\frac{\hbar^2}{2m} \left( \frac{1}{[E - V(r)]^{-1/4}} \right)^2 \nabla^2 \left[ E - V(r) \right]^{-1/4} \ll 1,$$

(3.47)

where the condition is satisfied when $V(r)$ does not change appreciably compared with $[E - V(r)]$ over a single wavelength. This is normally satisfied in large-scale situations in which the quantum number is large and the wavelength is short enough that $V(r)$ changes little within it. These are also the conditions under which the Bohr-Sommerfeld quantization energies are a good approximation to the exact solutions, with the ground state energies usually being much less accurate than the energies corresponding to larger quantum numbers $n$. For cases in which the semiclassical approximation has been deemed invalid for the system, one needs to calculate the wave functions and energies with the quantum potential included and determine if any new quantum effects are revealed as a result. In most cases, this procedure
requires a fully numerical solution.

### 3.4.2 Local Density of States

Due to the framework being based upon a free electron model, there is no explicit inclusion of the local density of states (DOS) of the emitter at the emitter surface. As a result, other than effective masses derived from the curvature of the constant energy surfaces, specific information about the band structure of the emitter material is not included in the emission model. The local density of states \( \rho(E, r) \) is derived from knowledge of the wave function \( \Psi_Q(s) \) of state \( Q \) at the location of emission \( s \) via [8]

\[
\rho(E, s) = 2 \sum_Q |\Psi_Q(s)|^2 \delta(E - E_Q) \tag{3.48}
\]

where \( \delta(E - E_Q) \) is the Dirac delta function and \( E_Q \) is the energy of state \( Q \). The local DOS contains two components

\[
\rho(E, s) = \rho_b(E, s) + \rho_s(E, s), \tag{3.49}
\]

where \( \rho_b(E, s) \) is the contribution to the DOS from bulk (extended) states and \( \rho_s(E, s) \) is the contribution from the surface states. The bulk states contribute a DOS equal to [8]

\[
\rho_b(E, s) = \frac{2V}{4\pi^3} \int_{SBZ} d^2k || \left[ \sum_\alpha \left( \frac{\partial k_{\alpha z}}{\partial E} \right) |\Psi_B(s)|^2 \right] - E_B = E, \tag{3.50}
\]

where \( V \) is the volume under consideration, \( SBZ \) denotes the surface Brillouin zone, \( k_{||} \) are the wave vectors corresponding to momentum parallel to the emitter surface, \( k_{\alpha z} \) is the wave vector in the direction of emission for state \( \alpha \), \( B = (E_B, k_{||}, \alpha) \) denotes the \( B \)th bulk state and the sum over \( \alpha \) accounts for the fact that there may be zero or
Figure 3-5: Numerically-calculated enhancement factor curves $R(\epsilon) = j(\epsilon)/j_0(\epsilon)$ for the (110) and (100) planes of tungsten, where $j(\epsilon)$ is the total energy distribution with the surface density of states included and $j_0(\epsilon)$ is the total energy distribution of the free electron model, without the surface density of states [5].

multiple states for a given $E$ and $k_{||}$. The contribution from the surface states is [8]

$$\rho_s(E, s) = \frac{A}{2\pi^2} \int \int_{SBZ} \sum_{\mu} \delta(E - E_{\mu}(k_{||})) |\Psi_S(s)|^2 d^2k_{||},$$

(3.51)

where $S = (k_{||}, \mu)$ are the quantum numbers characterizing the surface state, $\mu$ is a surface band index, and the energy of the $S$th state is given by $E_S = E_{\mu}(k_{||})$. In each of the above equations, the band structure of the material is included via the specific relationship between $E$ and $k$, along with the degeneracy of such states, and to a lesser extent, the form of the wave function $\Psi(s)$.

Due to the DOS being directly dependent upon the total energy $E_Q$ of each state, it is suspected that the most significant effect of omitting a specific, local DOS would be observed in the total energy distribution (TED) of emitted electrons, which is plotted as a function of $E$. This results in TEDs that are highly dependent upon the emitter material and its crystal orientation, as well as the structure of the emitter surface. As a result, the TEDs differ from the rather simple shape predicted by FN theory, as shown for the case of emission from the (110) and (100) surfaces of
Tungsten in Figure 3-5. Whereas the TED for the (110) surface is similar to that of the free electron model, the TED for the (100) surface shows significant deviations, namely substantial peaks and valleys that differ from the free electron model by as much as an order of magnitude. Due to the integral over all total energies when evaluating the magnitude of the emitted current density, the consequence of excluding the DOS specific to the particular emitter surface under investigation is likely to have a less pronounced impact on the total ECD, especially for cases in which the DOS is considered to be quasi-continuous. The plots in Figure 3-5 suggest that the total ECD from the (110) surface of tungsten would be larger than that predicted by the free electron model by a factor of approximately two, while the total ECD from the (100) surface would be smaller by a factor of approximately one order of magnitude. However, in cases where the energy spectrum contains discrete components (low-dimensional emitters), the wave functions and energy dispersion relations for the states at the surface could greatly differ from the free electron case and, as a result, play a more prominent role in determining the magnitude of the total ECD, as each of the comparatively few energy states may account for a more significant fraction of the total ECD. Although it is not theoretically difficult to incorporate the local density of states into the semiclassical framework for field emission, its inclusion would require a large amount of computational effort and likely only reveal small deviations in emission properties as compared with the predictions of the free-electron model. For more detail on the role that the local density of states plays in the physics of field emission, see the textbook by Modinos [8].

3.4.3 Semiclassical, Multidimensional Tunneling Model

In deriving the tunneling probability of electrons in Equation 3.36, it was assumed that all tunneling electrons are normally incident upon the emitter surface, or equivalently, tunneling electrons in the forbidden region have no momentum in dimensions.
transverse to the emitting surface. This assumption greatly simplifies the tunneling model in the following ways: i) a single set of wave fronts, the equiamplitude surfaces of the action, are able to completely describe the semiclassical wave function in the forbidden region, instead of the two sets which are generally required; ii) the tunneling paths are completely independent of each other, as a lack of tangential momentum leads to tunneling paths with zero divergence, obviating the need to use the “tube of tunneling paths” concept in tunneling calculations; and iii) the path of the electron in the forbidden region can be determined by using the classical equations of motion with an inverted potential $V \rightarrow -V$ and total energy $E \rightarrow -E$.

Aside from the simplifications to the tunneling probability calculation, the justification for limiting tunneling electrons to normal incidence is that for electrons with a fixed total energy $E$, the electrons with normal incidence experience the smallest exponential decay in the forbidden region and dominate the tunneling current [3]. Of course, the major drawback to ignoring the tunneling of electrons with non-zero transverse momentum is a reduced total tunneling probability from a particular point on the emitter surface, due to the omission of the contributions from those electrons. For example, the tunneling probability of electrons with fixed total energy $E = 5$ eV from a planar emitter surface defined by $E_F = 5$ eV and $\phi_0 = 4.5$ eV, for an applied electric field of $F = 2.5$ V/nm, decreases by approximately one order of magnitude over the course of increasing the transverse energy from $E_t = 0$ eV to $E_t = 1$ eV. If the applied electric field is increased to $F = 5$ V/nm, the tunneling probability over the same range of transverse energies decreases only by a factor of two. This decrease in the tunneling probability occurs due to the longer tunneling path lengths traveled by electrons with non-zero transverse momentum components. Consequently, for realistic, highly-curved emitter tips, the reduction in the tunneling probability with increasing transverse energy would be exacerbated, due to the tunneling barrier being comparatively wider for electrons which do not tunnel along a path that is normal
to the emitter surface at the apex, thus helping to reduce the error introduced by this approximation.

3.5 Chapter Summary

Based upon a semiclassical approximation to the multidimensional Schrödinger equation, the semiclassical framework for field emission has been developed. Casting the wave function in the Schrödinger equation in a polar form results in a set of two equations which determine the approximate wave function solution: i) the quantum Hamilton-Jacobi equation of de Broglie-Bohm theory, and ii) the continuity equation. The semiclassical approximation was imposed by assuming that the quantum potential $Q$ of the quantum HJE is sufficiently small compared to the classical potential and electron kinetic energies, after which the quantum HJE reduces to the classical HJE, the equation that determines the zeroth order approximation to the wave function. The form of the continuity equation, which determines the first order approximation to the wave function, is unaltered by this assumption. The semiclassical wave functions were then used to derive the physical quantities needed for field emission calculations, such as the current density incident upon the surface barrier, the transmission probability of electrons through the surface barrier, and the bound state energy levels. From these quantities, a quantum-state-based equation for the current density emitted from the emitter surface was presented. Finally, the consequences of the simplifying approximations made during the derivation of the model were discussed.
Chapter 4

Field Electron Emission from a Low-Dimensional, Highly-Curved Metal Tip

4.1 Introduction

As was shown in Figure 1-1, the geometry of a typical modern field emitter is characterized by a rotationally symmetric emitter body that gradually tapers to a rounded emitter tip with a radius of curvature on the order of a few nanometers. In terms of the physics of field emission, such a geometry differs from the Fowler-Nordheim (FN) model in two significant ways: i) the nanoscale dimensions of the field emitter at the tip necessitate the inclusion of quantum confinement effects, and ii) the highly-curved emitter surface requires a multidimensional treatment of electron tunneling. While it has been shown by the author that lateral (transverse to the emission direction at the apex) quantum confinement reduces the electron supply inside the emitter relative to the bulk case [4, 41, 42], and it is well known that electric field enhancement at the apex of sharp emitter tips leads to an enhanced tunneling probability as compared
with the planar emitter surface case [52], no formulation of field emission treats them coherently and on equal footing with each other, as manifestations of the specific geometry of the emitter. As a result, it is not known how the competing influences of lateral quantum confinement (QC) and electric field enhancement (EFE) at the emitter tip govern the performance of modern emitters. For that reason, it is the task of this chapter to construct a model for field emission from a realistic geometry, a nanoscale paraboloid, and investigate the influence of the emitter radius (of curvature) on various properties of emission.

### 4.2 Definition of Paraboloidal Emitter System

An elliptic paraboloidal emitter geometry is a natural choice for modeling field emission from modern emitter tips. First, due to it being rotationally symmetric and having tapered lateral dimensions that terminate in a rounded apex, a paraboloid closely resembles the shape of a real emitter. Furthermore, since both the degree of QC and the magnitude of the EFE depend on the emitter radius of curvature at the apex, the influences of both can be studied simultaneously, as a function of emitter dimensions. Finally, since the paraboloid is a surface of constant coordinate in three-dimensional parabolic coordinates, both the Schrödinger equation, which determines the wavefunctions and energies inside the emitter (where \( V(\mathbf{r}) = 0 \)), and Poisson’s equation, which determines the electrostatic potential outside the emitter, can be solved analytically via separation of variables. Accordingly, the emitter geometry and field emission model are defined in terms of parabolic coordinates.

#### 4.2.1 Three-Dimensional Parabolic Coordinates

Circular, three-dimensional, parabolic coordinates consist of a set of coordinates \( \mathbf{r} = (\eta, \xi, \phi) \) defined in terms of their relation to the Cartesian coordinates \( \mathbf{r} = (x, y, z) \)
Figure 4-1: The surfaces of constant coordinate for $\eta = 2$ (blue, downward-opening paraboloid) and $\xi = 0.5$ (red, upward-opening paraboloid), along with the unit vectors $\hat{\eta}$ and $\hat{\xi}$ at selected points on their respective surfaces of constant coordinate.

\[ x = \frac{\eta^{1/2} \xi^{1/2}}{2} \cos[\phi], \]
\[ y = \frac{\eta^{1/2} \xi^{1/2}}{2} \sin[\phi], \]
\[ z = \frac{1}{2} (\eta - \xi), \] \hspace{1cm} (4.1)

or their relationship to cylindrical coordinates $r = (\rho, \phi, z)$

\[ \rho = \frac{\eta^{1/2} \xi^{1/2}}{2}, \]
\[ \phi = \phi, \] \hspace{1cm} (4.2)
\[ z = \frac{1}{2} (\eta - \xi). \]

While the coordinate $\phi \in [0, 2\pi]$ is identical to the azimuthal angle of cylindrical coordinates, the coordinates $\eta \in [0, \infty)$ and $\xi \in [0, \infty)$ can be visualized via their surfaces of constant coordinate, which are downward-opening paraboloids of revolution for $\eta$ and upward-opening paraboloids of revolution for $\xi$. The graphical relationship
between the parabolic coordinates and Cartesian coordinates, as well as the surfaces
of constant coordinate for $\eta$ and $\xi$ are shown in Figure 4-1.

The scale factors for parabolic coordinates are given by

$$
h_\eta = \sqrt{\frac{\eta + \xi}{4\eta}}; \quad h_\xi = \sqrt{\frac{\eta + \xi}{4\xi}}; \quad h_\phi = \sqrt{\eta \xi},
$$

(4.3)

from which the expression for the gradient operator in parabolic coordinates is ob-
tained

$$
\nabla f(r) = \sum_{i=1}^{N} \frac{1}{h_i} \frac{\partial f}{\partial q_i} \hat{q}_i = \sqrt{\frac{4\eta}{\eta + \xi}} \frac{\partial f}{\partial \eta} \hat{\eta} + \sqrt{\frac{4\xi}{\eta + \xi}} \frac{\partial f}{\partial \xi} \hat{\xi} + \frac{1}{\sqrt{\eta \xi}} \frac{\partial f}{\partial \phi} \hat{\phi},
$$

(4.4)

where $\hat{\eta}, \hat{\xi},$ and $\hat{\phi}$ are unit vectors. It is also useful to have the expression for the
Laplace operator in terms of parabolic coordinates

$$
\nabla^2 f(r) = \frac{4}{\eta + \xi} \left[ \frac{\partial}{\partial \eta} \left( \frac{\eta}{\partial \eta} \frac{\partial f}{\partial \eta} \right) + \frac{\partial}{\partial \xi} \left( \frac{\xi}{\partial \xi} \frac{\partial f}{\partial \xi} \right) \right] + \frac{1}{\eta \xi} \frac{\partial^2 f}{\partial \phi^2},
$$

(4.5)

as it is necessary for solving both the Schrödinger equation and Poisson’s equation.

### 4.2.2 Physical Definition of the Emitter Cathode

The geometry of the emitter cathode itself is a solid, downward-opening, elliptical
paraboloid of revolution, whose emitting surface is defined by the surface of constant
coordinate $\eta = \eta_1$ and the set of coordinates $\mathbf{s} = (\eta_1, \xi, \phi)$, where $\xi \in [0, \xi_1]$ and
$\phi \in [0, 2\pi]$, as shown in Figure 4-2. On the surface of the emitter, variation of the
free coordinates $\xi$ and $\phi$ corresponds to traveling vertically along lines of longitude
and horizontally along lines of latitude, respectively. The radius of curvature $R$ at
the apex, located at $s_a = (\eta_1, 0, \phi)$, is given by

$$
R = \eta_1,
$$

(4.6)
Figure 4-2: Diagram of the paraboloidal emitter cathode defined between the surfaces of constant coordinate $\eta = \eta_1$ and $\xi = \xi_1$, with radius of curvature at the apex $R = \eta_1$, radius at the origin $a = \eta_1$, and height $L = (\eta_1 + \xi_1)/2$. The apex of the emitter is located at the point $s_a = (\eta_1, 0, \phi)$ and the emission direction $\hat{\eta}$ varies as a function of position on the emitter surface.

which is also the radius of the emitter $a$ as measured horizontally from the origin of the coordinate system, and the height of the cathode $L$, as measured along the $z$ axis from the apex to the base (the surface of constant coordinate $\xi = \xi_1$) is

$$L = \frac{1}{2} (\eta_1 + \xi_1). \quad (4.7)$$

Additionally, the emitter has a volume equal to

$$V = \frac{\pi}{4} \eta_1 \xi_1 (\eta_1 + \xi_1), \quad (4.8)$$

and a surface area equal to

$$\Sigma = \frac{2\pi}{3} \eta_1^2 \left[ \left( \frac{\xi_1}{\eta_1} + 1 \right)^{3/2} - 1 \right]. \quad (4.9)$$
The emission direction \( \hat{\eta}(s) \) is defined as the unit vector normal to the emitter surface and can be expressed in terms of the unit vectors of cylindrical coordinates by

\[
\hat{\eta}(\eta_1, \xi) = \frac{\sqrt{\xi/\eta_1}}{\sqrt{\xi/\eta_1} + 1} \hat{\rho} + \frac{1}{\sqrt{\xi/\eta_1} + 1} \hat{z},
\]

as both parabolic and cylindrical coordinates are axially symmetric about the \( z \) axis. The emission direction varies as a function of location on the surface and coincides with the \( z \) axis at the apex \( s_a \).

The electron system of the emitter cathode consists of an ideal Sommerfeld-type metal \([100]\), which is composed of an ideal gas of free electrons that obey Fermi-Dirac statistics and have mass \( m_0 \), the mass of an electron in free space. The thermodynamic temperature of the metal is taken to be \( T = 0 \) K, at which all electronic states with energies (referenced to the conduction band edge) up to and including the Fermi energy \( E_F \) are occupied, and all higher energy states are empty. The surface of the emitter is assumed to be perfectly smooth, free of defects, and to have a uniform local work function \( \phi_0 \). Additionally, it is assumed that electric fields do not penetrate into the metal and that the potential everywhere inside the emitter is \( V(r) = 0 \), in accordance with the properties of perfect conductors.

### 4.2.3 Emitter System Electrostatics

In the interest of obtaining a separable, analytical solution to Poisson’s equation for the potential between the anode and emitter cathode, the anode geometry was also chosen to be a downward-opening paraboloid. The anode geometry is defined by a surface of constant coordinate \( \eta = \eta_2 > \eta_1 \), where \( \xi \in [0, \xi_2] \) and \( \phi \in [0, 2\pi] \), as shown in Figure 4-3. The shortest distance between a point on the anode and a point on
Figure 4-3: Diagram of paraboloidal cathode-anode geometry. Like the emitter cathode, the anode is a downward-opening paraboloid defined by the surface of constant coordinate $\eta = \eta_2$, where $\eta_2 > \eta_1$. A voltage $V = V_0$ is applied to the anode, whereas the emitter cathode is kept grounded. The spacing between the anode and cathode apices is $d = (\eta_2 - \eta_1)/2$.

the cathode occurs between their apices, and is equal to

$$d = \frac{1}{2} (\eta_2 - \eta_1).$$

(4.11)

The potential difference between the anode and cathode, which lowers the surface barrier and allows electrons to tunnel into vacuum to generate field emission current, is achieved by applying a voltage $V = V_0$ to the anode, while keeping the cathode grounded.

The potential profile $\Phi(r)$ between the anode and cathode is determined by Poisson’s equation in parabolic coordinates

$$\nabla^2 \Phi(r) = \frac{4}{\eta + \xi} \left[ \frac{\partial}{\partial \eta} \left( \eta \frac{\partial \Phi}{\partial \eta} \right) + \frac{\partial}{\partial \xi} \left( \xi \frac{\partial \Phi}{\partial \xi} \right) \right] + \frac{1}{\eta \xi} \frac{\partial^2 \Phi}{\partial \phi^2} = 0,$$

(4.12)

along with the boundary conditions at the anode and cathode surfaces.
Figure 4-4: Tunneling barrier generated between a paraboloidal cathode with radius of curvature $R = \eta_1 = 1$ nm, work function $\phi_0 = 4.5$ eV, and Fermi energy $E_F = 5$ eV, and anode of radius $\eta_2 = 50$ nm, for applied voltage $V_0 = 10$ V, showing (a) the contours of the potential in the $x-z$ plane and (b) the profile of the tunneling barrier at the apex of the cathode, along the z axis.

$$\Phi(\eta_1, \xi, \phi) = 0,$$

$$\Phi(\eta_2, \xi, \phi) = V_0. \tag{4.13}$$

Since the boundary conditions depend only on surfaces of constant coordinate in $\eta$, a separable, product solution of functions that only depend on a single coordinate can be found, taking the form $\Phi(r) = H(\eta)\Xi(\xi)F(\phi)$. However, if it is assumed that the anode and cathode extend infinitely downward, Poisson’s equation can be simplified to a one-dimensional problem, since Equation 4.12 is homogeneous and the boundary conditions only depend upon $\eta$. As a result, $\Xi(\xi)$ and $F(\phi)$ are constants and Poisson’s equation in $\eta$ becomes

$$\frac{d}{d\eta} \left( \eta \frac{dH}{d\eta} \right) = 0. \tag{4.14}$$

Solving for $H(\eta)$ and applying the boundary conditions yields the solution for the potential $\Phi(\eta)$

$$\Phi(\eta) = \frac{V_0}{\ln[\eta_2/\eta_1]} \ln \left[ \frac{\eta}{\eta_1} \right]. \tag{4.15}$$
Figure 4-5: The magnitude of the electric field near the emitter surface for an emitter of radius $R = \eta_1 = 1$ nm, with work function $\phi_0 = 4.5$ eV and Fermi energy $E_F = 5$ eV, and an anode of radius $\eta_2 = 50$ nm, for applied voltage $V_0 = 10$ V. The plot in (a) shows the contours of the magnitude of the electric field in the $x - z$ plane, from which the electric field enhancement at the emitter tip apex is evident, and (b) shows the magnitude of the electric field as a function of position on the emitter surface, where $\xi = 0$ corresponds to the emitter apex and $\xi = 20$ nm is located closer to the emitter base. The apex field factor for this particular cathode-anode system is $\beta_a = 0.511$ nm$^{-1}$.

from which the expression for the tunneling barrier potential energy is obtained

$$V(\eta) = \phi_0 + E_F - \frac{eV_0}{\ln[\eta_2/\eta_1]} \ln \left[ \frac{\eta}{\eta_1} \right].$$  \hspace{1cm} (4.16)

The contours of $V(\eta)$ in the $x - y$ plane and its profile at the emitter apex are plotted in Figure 4-4. In contrast to the Fowler-Nordheim model, the tunneling barrier at the surface of the emitter is not triangular, but instead has a logarithmic dependence on $\eta$. Taking the negative gradient of the potential gives the electric field at the surface of the cathode

$$\mathbf{F} = \frac{2V_0}{\ln[\eta_2/\eta_1]} \frac{1}{\eta_1} \sqrt{\frac{\eta_1}{\eta_1 + \xi \hat{\eta}}},$$  \hspace{1cm} (4.17)

which always points normal to the contours of constant potential in Figure 4-4. Contours of the magnitude of the electric field near the surface of the emitter and the magnitude of the electric field as a function of position on the emitter surface $\xi$ are
plotted in Figure 4-5. The magnitude of the electric field at the emitter apex $F_a$, which is also the maximum field anywhere on the emitter surface, is given by

$$F_a = \frac{2V_0/\eta_1}{\ln(\eta_2/\eta_1)},$$

(4.18)

where the convention in field emission theory of replacing the electric field $E$ with its negative, such that $E \rightarrow -F$, has been retained. Finally, the apex field factor $\beta_a$, defined as the ratio of the apex electric field $F_a$ to the applied voltage $V_0$, is given by

$$\beta_a = \frac{2/\eta_1}{\ln(\eta_2/\eta_1)},$$

(4.19)

which, like $F_a$, is independent of the aspect ratio of the emitter $L/R$.

### 4.3 Derivation of Emitted Current Density Equation

As presented in Chapter 3, the emitted current density (ECD) equation is given by

$$J(V_0, T) = 2^e \sum_Q f_{FD}(E_Q, T) \int \int [j_Q(s) \cdot \hat{n}] D_Q(V_0, s)d\Sigma,$$

(4.20)

where $e$ is the elementary charge, $\Sigma$ is the area of the emitter surface, $f_{FD}(E_Q, T)$ is the Fermi-Dirac distribution function, $E_Q$ is the total energy of state $Q$, $T$ is the thermodynamic temperature of the electron gas, $j_Q(s)$ is the probability flux density for state $Q$ at a point $s$ on the emitter surface, $\hat{n}$ denotes the unit vector in the emission direction (normal to the emitter surface), $D_Q(V_0, s)$ is the probability of an electron tunneling through the potential barrier at point $s$ on the emitter surface, $V_0$ is the voltage applied to the anode, and $d\Sigma$ is the surface element on the emitter surface. In order to construct the ECD equation, it is first necessary to solve the
Schrödinger equation for the approximate, semiclassical wave functions, from which expressions for the bound state energies \(E_Q\), probability flux density \(j_Q\), and the transmission probability of electrons through the barrier \(D_Q\) can be derived. These expressions and the resulting ECD equation for the paraboloidal emitter are derived in the sections below.

### 4.3.1 Semiclassical Wave Functions in Parabolic Coordinates

The Schrödinger equation in parabolic coordinates takes the form

\[
-\frac{\hbar^2}{2m_0} \left\{ \frac{4}{\eta + \xi} \left[ \frac{\partial}{\partial \eta} \left( \eta \frac{\partial \Psi_Q}{\partial \eta} \right) + \frac{\partial}{\partial \xi} \left( \xi \frac{\partial \Psi_Q}{\partial \xi} \right) \right] + \frac{1}{\eta \xi} \frac{\partial^2 \Psi_Q}{\partial \phi^2} \right\} + V(r)\Psi_Q = E_Q\Psi_Q,
\]

where \(\hbar\) is the reduced Planck constant, \(m_0\) is the mass of an electron in free space, \(\Psi_Q\) is the wave function for state \(Q\), \(E_Q\) is the total energy of state \(Q\), and \(V(r)\) is the potential of the system. Applying the semiclassical approximation as detailed in §3.2, the wave function is cast in a polar form \(\Psi_Q = R(r) \exp \left[ i S(r) / \hbar \right]\), where the wave function density is defined as \(\rho(r) = R^2(r)\) and the momentum is defined as the gradient of the classical action \(p(r) = \nabla S(r)\). Inserting the polar form of the wave function into the Schrödinger equation yields a pair of equations that determine the approximate wave function to first order

\[
\frac{(\nabla S)^2}{2m_0} + V = E_Q,
\]

which is the classical Hamilton-Jacobi equation (HJE) and

\[
R \nabla^2 S + (2 \nabla R) \cdot (\nabla S) = 0,
\]

is a form of the continuity equation.
Wave Function Inside the Emitter

In parabolic coordinates, the classical Hamilton-Jacobi equation (HJE) inside the emitter takes the form

$$\frac{4\eta}{\eta + \xi} \left( \frac{\partial S}{\partial \eta} \right)^2 + \frac{4\xi}{\eta + \xi} \left( \frac{\partial S}{\partial \xi} \right)^2 + \frac{1}{\eta \xi} \left( \frac{\partial S}{\partial \phi} \right)^2 = 2m_0E_Q,$$  \hspace{1cm} (4.24)

as $V(r) = 0$ everywhere inside the emitter. In this case, the HJE admits an additive, separable solution of the form $S(r)/\hbar = H(\eta) + \Xi(\xi) + \Phi(\phi)$, which corresponds to a zeroth order wave function $\Psi_Q(r) = \Psi_\eta(\eta)\Psi_\xi(\xi)\Psi_\phi(\psi)$, that when inserted into Equation 4.24 yields

$$\frac{4\eta}{\eta + \xi} \left( \frac{\partial H}{\partial \eta} \right)^2 + \frac{4\xi}{\eta + \xi} \left( \frac{\partial \Xi}{\partial \xi} \right)^2 + \frac{1}{\eta \xi} \left( \frac{\partial \Phi}{\partial \phi} \right)^2 = k_Q^2,$$  \hspace{1cm} (4.25)

where $k_Q = \sqrt{2m_0E_Q}/\hbar$ is the magnitude of the wave vector of the electron. A solution for $\Phi(\phi)$ can be found by multiplying Equation 4.25 by the factor $\eta \xi$, giving

$$\left( \frac{\partial \Phi}{\partial \phi} \right)^2 = p^2 \rightarrow \Phi(\phi) = \pm|p|\phi \rightarrow \Psi(\phi) = \exp[\pm i|p|\phi]$$  \hspace{1cm} (4.26)

where $p$ is a constant determined by the cyclic boundary condition for the azimuthal wave function

$$\Psi_\phi(0) = \Psi_\phi(2\pi) \rightarrow 1 = \exp[\pm i|p|2\pi],$$  \hspace{1cm} (4.27)

where $p = 0, \pm 1, \pm 2, \ldots$ is an integer and the *azimuthal quantum number*, corresponding to quantized angular momentum $\pi_{\phi,p} = \hbar p$. Inserting the solution for $\Phi$ back into Equation 4.25 and rearranging gives the equation that defines $H(\eta)$ and $\Xi(\xi)$

$$\left\{ \eta \left( \frac{\partial H}{\partial \eta} \right)^2 - \frac{k_Q^2}{4}\eta + \frac{p^2}{4\eta} \right\} = -\left\{ \xi \left( \frac{\partial \Xi}{\partial \xi} \right)^2 - \frac{k_Q^2}{4}\xi + \frac{p^2}{4\xi} \right\} = -q$$  \hspace{1cm} (4.28)
where $q$ is a constant of motion of the system. The solutions for $H(\eta)$ and $\Xi(\xi)$ are given by

$$H(\eta) = \int \kappa_{\eta,q}(\eta) d\eta, \quad \text{(4.29a)}$$

$$\Xi(\xi) = \int \kappa_{\xi,Q}(\xi) d\xi, \quad \text{(4.29b)}$$

where

$$\kappa_{\eta,q}(\eta) = \sqrt{\frac{k^2 Q}{4} - \frac{q - \frac{p^2}{4\eta^2}}{\eta}} \quad \text{and} \quad \kappa_{\xi,Q}(\xi) = \sqrt{\frac{k^2 Q}{4} + \frac{q + \frac{p^2}{4\xi^2}}{\xi}}. \quad \text{(4.30)}$$

Up to a constant, the solution for $S(r)$ is then given by

$$S(r) = \int_{\eta_0}^{\eta} \pi_{\eta,q}(\eta') d\eta' + \int_{\xi_0}^{\xi} \pi_{\xi,Q}(\xi') d\xi' + \pi_{\phi_p,\phi}, \quad \text{(4.31)}$$

where $\pi_{q,Q} = h\kappa_{q,Q}$ are the momenta conjugate to the coordinates $q_i$ after the transformation that produces the HJE (see §A.5), and the zeroth order wave function inside the emitter is

$$\Psi_Q^0(r) = A_Q \exp \left[ i \int_{\eta_0}^{\eta} \kappa_{\eta,q}(\eta') d\eta' \right] \exp \left[ i \int_{\xi_0}^{\xi} \kappa_{\xi,Q}(\xi') d\xi' \right] \exp [i\phi], \quad \text{(4.32)}$$

where $A_Q$ is a normalization constant defined by the normalization condition

$$1 = \frac{\pi}{2} \int_0^{\eta_1} \int_0^{\xi_1} |\Psi_Q(r)|^2 (\eta + \xi)d\eta d\xi. \quad \text{(4.33)}$$

Instead of using the form of the continuity equation of Equation 4.23 to calculate the amplitude $R(r)$ of the first order wave function explicitly, the result of Equation 3.17
can be used directly, yielding

\[ R(r) = \frac{1}{\sqrt{p(r)\sigma(r)}}, \tag{4.34} \]

where \( p(r) = \sqrt{2m_0E_Q} \) is the magnitude of the electron’s momentum and \( \sigma(r) \) is the cross-sectional area of a tube defined by neighboring classical electron paths, as discussed in §3.2.2. The cross-sectional area as a function of position along the electron’s path of travel can be calculated via the divergence of the unit vector tangent to the electron’s path of travel. In the case of free space, for which \( V(r) = 0 \), no forces are exerted on the electrons and they travel in straight lines (in Cartesian coordinates). This leads to zero divergence in the unit tangent vectors along the electron paths and a tube of constant area, such that \( \sigma(r) = \sigma \). Since \( p \) is also constant in free space, \( R \) becomes a normalization constant and the first order wave function is identical to the zeroth order wave function in Equation 4.32.

It should be noted, that even though \( V(r) = 0 \) throughout the emitter and there are no forbidden regions in the usual sense (where \( E_Q < V \)), the wave function has evanescent components in some regions of the emitter. It must be remembered that in deriving the HJE, a new set of coordinates and momenta \((q, \pi)\) were adopted to describe the system (see §A.5), where

\[ q = (\eta, \xi, \phi) \quad \text{and} \quad \pi = \left( \frac{\partial S}{\partial \eta}, \frac{\partial S}{\partial \xi}, \frac{\partial S}{\partial \phi} \right) = (\pi_\eta, Q, \pi_\xi, Q, \pi_\phi, p), \tag{4.35} \]

and \( \pi \) are constants of the motion. Since the Hamiltonian is time independent, the total energy \( E_Q \) is also a constant of the motion. As a result, the HJE and its solutions require that energy be conserved along each path, and when the constants of motion \( E_Q, q, \) and \( p \) are chosen as independent (fixed) variables, it leads to some regions in the emitter where the new momenta \( \pi \) take on imaginary values, corresponding to those regions being classically forbidden for that range of coordinates. Consequently,
in these forbidden regions, which may occur along $\eta$ or $\xi$, the wave function corresponding to that coordinate is an evanescent wave with a decaying amplitude. This leads to the turning points $\eta_{m,t,Q}$ and $\xi_{m,t,Q}$ inside the emitter, which are determined by solving the equations $\pi_{\eta,Q}(\eta_{m,t,Q}) = 0$ and $\pi_{\xi,Q}(\xi_{m,t,Q}) = 0$. Due to the wave function in these forbidden regions being an evanescent wave, no incident flux density is contributed by that portion of the wave function. However, since the turning points originate near the origin and move outward in $\eta$ with increasing $q$ and $p$, current is still emitted until $\eta_{m,t,Q} \geq \eta_1$, in which case the entire emitter from $0 \leq \eta \leq \eta_1$ is classically forbidden and the incident flux density is zero. Additionally, the presence of turning points inside the emitter plays a significant role in estimating the bound state energies $E_Q$ of the emitter via the Bohr-Sommerfeld quantization condition.

### 4.3.2 Bohr-Sommerfeld Bound State Energies: $E_Q$

For the purpose of estimating the bound state energy levels $E_Q$ in the emitter, it is assumed that the potential outside the emitter surface is infinite, such that

$$V(r) = \begin{cases} 
0 & \text{for } \eta < \eta_1 \\
\infty & \text{for } \eta \geq \eta_1
\end{cases},$$

(4.36)

and that $\xi_1$, which defines the location of the base of the emitter, is large enough that the momentum conjugate to the $\xi$ coordinate is quasi-continuous. Due to the infinite potential outside the emitter, the boundary condition at the surface is a hard wall, which requires that the wave function vanish at the surface such that $\Psi(\eta_1, \xi, \phi) = 0$. The other boundary condition occurs at $\eta = \eta_{m,t}$ and is a soft-wall boundary condition, since it is only required that the wave function there be finite. According to \S3.3.3, the standing wave that forms between these two boundaries along the coordinate $\eta$
must satisfy the condition

$$\left( n - \frac{1}{4} \right) \pi = \int_{\eta_{m}^{n}}^{n} \sqrt{\frac{k_{Q}^{2}}{4} - \frac{q}{\eta} - \frac{p^{2}}{4\eta^{2}}} \, d\eta,$$

(4.37)

where \( n = 1, 2, 3... \) is a positive integer quantum number and

$$\eta_{n}^{m} = \frac{q k_{Q}}{2} \left( 1 + \sqrt{\frac{k_{Q}^{2} p^{2}}{4q^{2}} + 1} \right)$$

(4.38)

is the turning point for which \( \kappa_{\eta,Q}(\eta_{n}^{m}) = 0 \). While \( n \) and \( p \) are discrete quantities due to the imposed boundary conditions in \( \eta \) and \( \phi \), \( q \) and \( k_{Q} \) are quantities that can vary continuously, which leaves a single dimension of freedom in the momentum and leads to the formation of one-dimensional energy subbands in the emitter. Although the integral has a closed form, it is not possible to obtain an explicit expression for \( k_{Q} \) and a numerical method must be employed to find valid sets of the parameters \((k, n, q, p)\). Furthermore, with the addition of \( n \) as a quantum number, each state is uniquely described by the set of labels \( Q = (n, q, p) \).

### 4.3.3 Incident Probability Flux Density: \([j_{Q} \cdot \hat{n}]\)

The incident probability flux at the emitter surface is evaluated using the wave function in Equation 4.32 and the \( \hat{n} \) component of the expression for the probability flux density from Equation 3.25, yielding

$$j_{\eta}(\eta_{1}, \xi) = [j_{Q} \cdot \hat{n}] = 2 |\Psi_{Q}(\eta_{1}, \xi)|^2 \sqrt{\frac{\eta_{1}}{\eta_{1} + \xi}} \frac{\pi_{\eta,Q}(\eta_{1})}{m_{0}}.$$

(4.39)

The maximum incident flux density at the surface occurs at the apex \((\xi = 0)\) and decreases as a function of increasing distance \( \xi \) from the apex.
4.3.4 Electron Transmission Probability: $D_Q$

In order to calculate the probability of electrons tunneling through the surface barrier, it is necessary to determine the paths of the electrons through the forbidden region. These paths are governed by the classical equations of motion with an inverted potential $V \rightarrow -V$ and total energy $E \rightarrow -E$. Although it was advantageous to solve the Schrödinger equation in parabolic coordinates, the form of the tunneling barrier potential in parabolic coordinates prohibits the HJE from fully separating into equations that are only dependent upon $\eta$ or $\xi$. As a result, the phase of the wave function in the forbidden region, which is required for the calculation of the tunneling probability in Equation 3.36, must be evaluated numerically and the choice of coordinate system is solely based on convenience for numerical calculation. Due to the symmetry about the $z$ axis shared by both parabolic and cylindrical coordinates and the simple form of the Lagrangian function in terms of the cylindrical coordinates $\rho$ and $z$, cylindrical coordinates are used for determining the path of the electron in the calculations below. The Lagrangian function $\mathcal{L}$ for the emitter system in the vacuum, between the cathode and anode is given in cylindrical coordinates by

$$\mathcal{L} = \frac{1}{2} m_0 \left( \dot{\rho}^2 + \rho^2 \dot{\phi}^2 + \dot{z}^2 \right) - \left( \phi_0 + E_F - \frac{e V_0}{\ln[\eta_2/\eta_1]} \ln \left[ \frac{z + \sqrt{\rho^2 + z^2}}{\eta_1} \right] \right)$$

(4.40)

where (\dot{} ) indicates a time derivative and the relationship $\eta = z + \sqrt{\rho^2 + z^2}$ has been used. The equations of motion are generated by inserting the Lagrangian into the Euler-Lagrange equations (see §A.2)

$$\frac{\partial \mathcal{L}}{\partial q_i} - \frac{d}{dt} \left( \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) = 0,$$

(4.41)

where the $q_i$ represent the coordinates $\eta$, $\xi$, and $\phi$. Since $\phi$ does not appear in the Lagrangian, it is a cyclic coordinate and its momentum is a constant of the motion,
such that
\[
\frac{d}{dt} \left( \rho^2 m_0 \dot{\phi} \right) = 0 \rightarrow \dot{\phi} = \frac{hp}{m_0 \rho^2},
\] (4.42)
where \( L = hp \) is the angular momentum. Substituting the constant angular momentum back into the Lagrangian function reduces the kinematics problem to a two dimensional one in \( \rho \) and \( z \), giving the new Lagrangian function
\[
\mathcal{L} = \frac{1}{2} m_0 \left( \dot{\rho}^2 + \frac{h^2 \rho^2}{m_0^2 \rho^2} + \dot{z}^2 \right) - \left( \phi_0 + E_F - \frac{eV_0}{\ln \left[ \eta_2/\eta_1 \right]} \ln \left[ \frac{z + \sqrt{\rho^2 + z^2}}{\eta_1} \right] \right). \] (4.43)
The Euler-Lagrange equations give the equation of motion for \( \rho \)
\[
\ddot{\rho} = \frac{h^2 p^2}{\alpha m_0^2 \rho} \left\{ \left[ 1 + \frac{z}{\rho} \left( \frac{z}{\rho} + \sqrt{\frac{z}{\rho}} \right) \right]^{-1} - \frac{\alpha}{\rho^2} \right\}, \] (4.44)
where \( \alpha = \frac{h^2 p^2 \ln \left[ \eta_2/\eta_1 \right]}{(eV_0 m_0)} \) and for \( z \)
\[
\ddot{z} = \frac{h^2 p^2}{\alpha m_0^2 \rho} \left[ \sqrt{\left( \frac{z}{\rho} \right)^2 + 1} \right]^{-1}. \] (4.45)
Although the equations of motion are expressed in terms of time, solving for \( \rho(t) \) and \( z(t) \) is not particularly useful, since the tunneling probability calculation only requires the shape of the path and not details related to how quickly the electron moves along it. Instead, it is more useful to solve for \( \rho(z) \), the differential equation for which is given by dividing Equation 4.44 by Equation 4.45
\[
\frac{d^2 \rho}{dz^2} = \frac{(\rho/z)}{1 + \sqrt{(\rho/z)^2 + 1}} - \frac{\alpha}{\rho^2} \sqrt{(z/\rho)^2 + 1}, \] (4.46)
which is solved numerically for the path \( s = \rho(z) \), with the initial position being a point on the emitter surface \( r_1 = (\rho_1, z_1) \), given in terms of \( \eta_1 \) and \( \xi \) as

\[
\begin{align*}
\rho_1 &= \frac{1}{2} \eta_1^{1/2} \xi^{1/2}, \\
z_1 &= \frac{1}{2} (\eta_1 - \xi),
\end{align*} \tag{4.47}
\]

and the direction of the initial velocity being normal to the emitter surface, given by

\[
\left. \frac{d\rho}{dz} \right|_{z=z_1} = \sqrt{\frac{\xi}{\eta_1}}. \tag{4.48}
\]

The end point of the path \( r_2 = (\rho_2, z_2) \) is located on the turning surface \( s_t \) in the vacuum, defined as the set of points for which \( E_\mathbf{Q} - V(r) = 0 \). For a given total energy \( E_\mathbf{Q} \) and corresponding set of quantum numbers \( \mathbf{Q} \), \( r_2 \) is found by calculating the path \( s \) over a large domain and truncating it at the point where \( V(s) = E_\mathbf{Q} \). The expression for the transmission probability is

\[
D(r_1, r_2) = \left[ \left. \frac{(dV(r)/ds)|_{s=s_2}}{(dV(r)/ds)|_{s=s_1}} \right] \right]^{1/2} \exp \left[ -g_e \int_{z_1}^{z_2} \sqrt{V[\rho(z), z] - E_\mathbf{Q}} \, dz \right], \tag{4.49}
\]

where \( g_e = 2\sqrt{2m_0}/\hbar \),

\[
V[\rho(z), z] = \phi_0 + E_F - \frac{eV_0}{\ln(\eta_2/\eta_1)} \ln \left[ \frac{z + \sqrt{\rho(z)^2 + z^2}}{\eta_1} \right], \tag{4.50}
\]

and the derivatives of \( V(r) \) with respect to the path \( s \) at the endpoints are also calculated numerically.
4.3.5 Total Energy Distribution and Total Emitted Current Density

The emitted current density $J(V_0)$ can be expressed as the sum (or integral) over all total energies in the system of the total energy distribution $j(V_0, E_Q)$, which is the current density emitted by a state with total energy $E_Q$,

$$J(V_0) = \sum_Q j(V_0, E_Q). \quad (4.51)$$

For the paraboloidal emitter, the total energy distribution is

$$j(V_0, E_Q) = e^{4\pi\eta_1} m_0 \Sigma f_{FD}(E_Q, T) \pi_{\eta, Q}(\eta_1) \int |\Psi_Q(\eta_1, \xi)|^2 D_Q(V_0, \xi) d\xi, \quad (4.52)$$

where $Q = (p, n, q)$, the surface area of the emitter $\Sigma$ is given by Equation 4.9, the wave function at the surface $\Psi_Q(\eta_1, \xi)$ is defined by Equation 4.32, $f_{FD}$ is the Fermi-Dirac distribution function, $\pi_{\eta, Q} = \hbar \kappa_{\eta, Q}$ is the momentum conjugate to the $\eta$ coordinate given by Equation 4.30, and the transmission probability $D_Q$ is given by Equation 4.49.

4.4 Results and Analysis

In this section, various properties of the paraboloidal emitter are investigated, including:

1. Bohr-Sommerfeld energies

2. Electronic subband origin of the emitted current density

3. Spatial origin of the emitted current density: effective emitting area

4. Emitted current density as a function of the emitter radius
5. Total energy distribution of the emitted electrons

6. Fowler-Nordheim plots

Unless otherwise stated, all results correspond to an emitter of height $L = 100$ nm, Fermi energy $E_F = 5$ eV, work function $\phi_0 = 4.5$ eV, and an anode of radius $\eta_2 = 100$ nm. Calculations were performed numerically using the Python programming language [137], with additional help from the NumPy [138], SciPy [139], and Matplotlib [140] packages.

### 4.4.1 Bohr-Sommerfeld Energies

**Results: $\eta_1 = 1.5$ nm**

The energies $E_Q = \hbar^2 k_Q^2/2m_0$ of the states $Q = (n, q, p)$ in the emitter were calculated numerically, according to Equation 4.37. However, due to the dependence of the energies on the radius of the emitter $\eta_1$, it is generally necessary to calculate the entire spectrum of energy states for each emitter radius. Fortunately, in this case, the Bohr-Sommerfeld quantization integral is able to be scaled such that the entire energy spectrum is calculated only once, for the largest emitter radius under consideration, and the energy spectrum for all emitters with smaller radii can be obtained via scaling relations. Scaling the spatial variable $\eta$ such that it is normalized to the emitter radius $\eta_1$ and substituting it into the Bohr-Sommerfeld condition yields

$$
\left( n - \frac{1}{4} \right) \pi = \int_{x_t}^{1} \sqrt{\frac{\tilde{k}_Q^2}{4} - \frac{\tilde{q}}{x} - \frac{p^2}{4x^2}} \, dx,
$$

where the scaled variables are defined as

$$
x = \eta/\eta_1, \quad x_t = \eta_t^\eta/\eta_1, \quad \tilde{k}_Q = \eta_1 k, \quad \tilde{q} = \eta_1 q,
$$

121
Figure 4-6: The total energy $E_Q$ vs. $q$ dispersion relation for the subbands \{p, n\} in a paraboloidal emitter of radius $\eta_1 = 1.5$ nm. Subbands with $n = 1$ are displayed as solid curves, while subbands with $n = 2$ are displayed as dashed curves.

While $n$ and $p$ are unaffected by the scaling procedure. Since the emitter system is taken to be at $T = 0$ K, only states with energies up to and including the Fermi energy $E_F$ are occupied and need to be considered. As a result, the valid energy states in an emitter of radius $\eta_1$ are determined by retaining only the scaled wave numbers $\tilde{k}_Q$ which fall below the scaled Fermi wave number $\tilde{k}_F = \eta_1 k_F$ and reversing the scaling procedure to find $k_Q$ and $q$.

Figure 4-6 shows the dispersion relation between $E_Q$ and $q$ for all subbands \{p, n\} in an emitter of radius $\eta_1 = 1.5$ nm. The dispersion relation illustrates that the electron system of the paraboloid is one dimensional, with $q$ being the only continuous quantity upon which the energy is dependent. Additionally, it is evident that the energy gap between states with differing $p$ quantum numbers is much smaller than that for the gap between states with differing $n$ quantum numbers, as there are eight $p$ subbands with $n = 1$, but only three subbands with $n = 2$. As the emitter radius increases, the spacing in energy between consecutive subbands rapidly decreases and the energy spectrum eventually forms a quasicontinuum in $p$ and $n$, corresponding to bulk-like emitter properties. For example, an emitter with radius $\eta_1 = 10$ nm
contains 96 $p$ subbands and only 18 $n$ subbands, all of which fall in the energy range $0 \leq E_Q \leq E_F$.

**Analysis: Classical Description of Electron Kinematics in the Emitter**

Understanding how the quantum numbers $Q$ govern the kinematics of electrons inside the emitter is vital to interpreting and explaining the predictions of the paraboloidal emitter model. Since the potential everywhere in the emitter is zero, the total kinetic energy $T_Q$ of the electrons in the emitter is fixed. However, the kinetic energy $T_{q_i,Q}$ associated with each coordinate $q_i$ is not necessarily fixed and varies as a function of quantum number and position in the emitter. The components of the kinetic energy are given by

$$
T_{\eta,Q}(\eta, \xi) = \frac{\hbar^2}{2m_0} \frac{1}{\eta + \xi} \left[ k_Q^2 \eta - \frac{p^2}{\eta} \right] = \frac{p^2_{\eta,Q}}{2m_0}, \tag{4.55a}
$$

$$
T_{\xi,Q}(\eta, \xi) = \frac{\hbar^2}{2m_0} \frac{1}{\eta + \xi} \left[ k_Q^2 \xi + \frac{p^2}{\xi} \right] = \frac{p^2_{\xi,Q}}{2m_0}, \tag{4.55b}
$$

$$
T_{\phi,p}(\eta, \xi) = \frac{\hbar^2}{2m_0} \frac{p^2}{\eta \xi} = \frac{p^2_{\phi,p}}{2m_0}. \tag{4.55c}
$$

From the expression for $T_{\phi,p}$, it is clear that $p$ is the *angular momentum quantum number*, where $L = \hbar p$ is the angular momentum. At a fixed location $(\eta, \xi)$, $T_{\phi,p}$ increases with increasing $p$ and for fixed $p$, $T_{\phi,p}$ increases with decreasing $\eta$ or $\xi$. At the surface of the emitter $\eta = \eta_1$ and the angular momentum for a given state with $p \neq 0$ increases with decreasing distance from the apex ($\xi = 0$), eventually tending toward infinity at the apex itself, as shown in Figure 4-7b. As a result, for states with a fixed total energy $E_Q$ and $p \neq 0$, the apex is a forbidden region, as the kinetic energy in $\phi$ exceeds the total energy of the state at that location. States with $p = 0$ have no associated angular momentum or kinetic energy and consequently, the apex is not excluded from being classically accessible, as is evident in Figure 4-7a.

The kinetic energies $T_{\eta,Q}$ and $T_{\xi,Q}$ are comparatively more complex in their de-
Figure 4-7: The kinetic energy $T_{\eta, \xi}$ associated with the $\eta$, $\xi$, and $\phi$ coordinates as a function of position on the surface $\xi$ of an emitter with radius $\eta_1 = 1.5$ nm, compared with the total energy $E_Q$. Plot (a) shows the kinetic energies for a state with $E_Q \approx E_F$, $p = 0$, and $n = 1$, while (b) shows the kinetic energies for a state with $E_Q \approx E_F$, $p = 1$, and $n = 1$.

Dependencies on the spatial coordinates and quantum numbers, as they both depend on $\eta$ and $\xi$, as well as $p$, $n$, and $q$. As can be seen in Figure 4-7a and Figure 4-7b, for fixed $p$ and $q$, $T_{\eta, \xi}$ increases with decreasing distance from the apex, along the emitter surface, while $T_{\xi, \eta}$ simultaneously decreases. For the case of Figure 4-7b in which $p \neq 0$, both $T_{\eta, \xi}$ and $T_{\xi, \eta}$ drop rapidly to zero near the apex, as $T_{\phi, p}$ approaches infinity.

Before exploring the dependence of the kinetic energies on $n$ and $q$, however, it is useful to develop a classical understanding of their roles in electron kinematics. The quantum number $n$ is a positive integer that arises from the imposition of boundary conditions on the wave function at the internal turning point $\eta^m_\ell$ and emitter surface $\eta_1$ and corresponds to the number of nodes in the wave function along the $\eta$ coordinate, plus one. Via the Bohr-Sommerfeld quantization condition, $n$ determines (along with $p$ and $q$) the total energies $E_Q$ of the states in the emitter. Functionally, increasing $n$ (with $p$ and spatial coordinates fixed) leads to an increase in the kinetic energy.
Figure 4-8: Plots of the total state energy $E_Q$, kinetic energy associated with $\eta$ ($T_{\eta,Q}$), kinetic energy associated with $\xi$ ($T_{\xi,Q}$), and kinetic energy associated with $\phi$ ($T_{\phi,p}$) for subband $\{p = 0, n = 1\}$, as a function of $q$. The kinetic energy is plotted at four points: (a) at the apex, where $\xi = 0$ and $0 < q_{c,Q} = q_{\text{max},Q} = 32 \text{ nm}^{-1}$; (b) at $\xi = 0.375 \eta_1$, where $0 < q_{c,Q} < q_{\text{max},Q}$; (c) at $\xi = \eta_1$ (or $z = 0$), where $q_{c,Q} = 0$; and (d) at $\xi = 3 \eta_1$, where $q_{c,Q} < 0$. The maximum value of $q$ allowed within a particular subband is $q_{\text{max},Q}$.

associated with the $\eta - \xi$ plane $T_{\eta\xi,Q}$, given by

$$T_{\eta\xi,Q}(\eta,\xi) = \frac{\hbar^2}{2m_0} \left[ k_Q^2 - \frac{p^2}{\eta \xi} \right], \quad (4.56)$$

which, notably, has no explicit dependence on $q$ outside of its relation to $k_Q$ in the Bohr-Sommerfeld quantization condition. Since $k_Q$ rises monotonically as a function of $q$, $T_{\eta\xi,Q}$ also increases as a function of $q$.

In addition to determining the total energy $E_Q$ of the electron, the quantum number $q$ governs the distribution of kinetic energy in the $\eta - \xi$ plane between $T_{\eta,Q}$ and $T_{\xi,Q}$. By equating $T_{\eta,Q}$ and $T_{\xi,Q}$, the value of $q$ that marks the point at which the kinetic energies are equal is

$$q_{c,Q}(\eta,\xi) = \frac{1}{8}(\eta - \xi) \left( \frac{p^2}{\eta \xi} + k_Q^2 \right) = \frac{z}{4} \left( k_Q^2 + \frac{p^2}{\rho^2} \right), \quad (4.57)$$
where ρ and z are the coordinates of the cylindrical coordinate system. If an electron has a value of \( q < q_{c,Q}(\eta, \xi) \), its kinetic energy in \( \eta \) is greater than its kinetic energy in \( \xi \), while if \( q > q_{c,Q}(\eta, \xi) \), the electron has more kinetic energy in \( \xi \) than in \( \eta \). Figure 4-8 plots \( T_{\eta,Q} \), \( T_{\xi,Q} \), \( T_{\phi,p} \), and \( E_Q \) as a function of \( q \) at four different locations on the emitter surface. Of particular note are the values of \( q_{c,Q} \) at the apex (\( \xi = 0 \)), where \( q_{c,Q} = q_{max,Q} \), the maximum value of \( q \) allowed in that subband, and at \( z = 0 \) (\( \xi = \eta_1 \)), where \( q_{c,Q} = 0 \). Since all values of \( q \) lie equal to or below \( q_{max,Q} \) in a given subband, \( T_{\eta,Q} \) will dominate at the apex up until \( q \) assumes its maximum value, as shown in Figure 4-8a. At \( z = 0 \), \( T_{\eta,Q} \) and \( T_{\xi,Q} \) are equal only at \( q = 0 \), after which \( T_{\xi,Q} \) dominates, as shown in Figure 4-8c. In between the apex and \( z = 0 \), \( q_{c,Q} \) assumes a positive value less than \( q_{max,Q} \) and there exist ranges of \( q \) where \( T_{\eta,Q} \) or \( T_{\xi,Q} \) dominate, which are visible in the plot of Figure 4-8b. For \( \xi > \eta_1 \), \( q_{c,Q} < 0 \) indicates that \( T_{\xi,Q} \) dominates for all \( q \), as shown in Figure 4-8d. In summary, increasing \( q \) has the effect of increasing both \( T_{\eta,Q} \) and \( T_{\xi,Q} \), while also controlling the ratio between them.

Additionally, \( q \) is a constant of the motion of the system, which arises from a symmetry and corresponding conservation law within the system. Similar to the Laplace-Runge-Lenz vector of celestial mechanics which describes the shape and orientation of planetary orbits [141], the symmetry related to \( q \) does not arise from a cyclical coordinate in the Lagrangian of the system. As a result, the conserved quantity associated with \( q \) is a \textit{dynamic conserved quantity} and must be directly derived via the method of Poisson brackets, instead of it naturally arising from the Euler-Lagrange equations of the system.
4.4.2 Electronic Subband Origin of Emitted Current Density

Results: p and n Subbands

The total emitted current density is comprised of contributions from each electronic subband \(\{p, n\}\). In Figure 4-9a, the percentage of the total ECD contributed by subbands with different \(p\) quantum numbers as a function of emitter radius \(\eta_1\) is shown. Across all radii, \(p = 0\) subbands contribute nearly 100% of the total ECD, while subbands with \(p > 1\) contribute a comparatively insignificant amount. Figure 4-9b plots the percentage of the total ECD contributed by subbands with different \(n\) indices as a function of emitter radius \(\eta_1\). The quantum numbers in the plot are referenced to \(n_{max}\), the subband with the maximum \(n\) value and highest energy in the emitter. While the total ECD is dominated by the subbands with \(n = n_{max}\) for smaller emitters, the lower \(n\) subbands begin to contribute more equally to the total ECD as the emitter radius increases. For example, for an emitter of radius \(\eta_1 = 10\) nm, states in the \(n_{max}\) subband contribute only 50% of the total ECD, as compared
with 100% for an emitter of radius $\eta_1 = 1$ nm.

**Analysis**

The electronic origin of the emitted current can be understood by investigating the incident flux density at the surface $\eta_{n,\mathbf{Q}}(\eta_1, \xi)$ and the transmission probability as a function of position on the emitter surface $D_{\mathbf{Q}}(\eta_1, \xi, V_0)$. The incident flux density at the surface contributed by a subband $\mathbf{Q}$ is given by the expression

$$
\eta_{n,\mathbf{Q}}(\eta_1, \xi) = 2ef_{FD}(E_{\mathbf{Q}}, T) \frac{h}{m_0} \kappa_{n,\mathbf{Q}}(\eta_1) 
\times |A_{\mathbf{Q}}|^2 \exp \left[ -2 \int_0^{\eta_1} |\kappa_{n,\mathbf{Q}}(\eta)| \, d\eta \right] \exp \left[ -2 \int_0^{\xi} |\kappa_{\xi,\mathbf{Q}}(\xi')| \, d\xi' \right],
$$

(4.58)

which decays exponentially as a function of the accrued wave function phase along $\xi$ (integral of the magnitude of $\kappa_{\xi,\mathbf{Q}}$) in the forbidden region $\xi < \xi_t$, where $\xi_t$ is the turning point along $\xi$. Outside the forbidden region, no additional phase is accrued because the purely imaginary wave function phases cancel each other upon taking the magnitude of the wave function and the upper limit of the integral in $\xi'$ is replaced by $\xi_t$. Thus, as shown in Figure 4-10a, the incident flux density has a maximum at the apex ($\xi = 0$) and decays extremely rapidly with distance from the apex along the emitter surface, with the exception of states with $p = 0$, for which $\xi_t = 0$ and the incident flux density is a constant.

Figure 4-10b shows a plot of the transmission probability through the barrier as a function of $\xi$, for the energy state $E_{\mathbf{Q}} = E_F$ of an emitter of radius $\eta_1 = 1.5$ nm. Across all subbands $p$, the transmission probability increases exponentially with decreasing distance from the apex, along the emitter surface. While the transmission probability of electrons in subband $p = 0$ reaches a maximum at the apex, the transmission probability of electrons in subbands $p > 0$ suddenly decreases in the immediate vicinity of the apex, with larger $p$ indices corresponding to lower maxi-
Figure 4-10: The (a) incident flux density $\eta_0 \cdot Q(\eta_1, \xi)$ and (b) transmission probability through the barrier as a function of position on the surface $\xi$ and quantum number $p$, for an emitter of radius $\eta_1 = 1.5$ nm, applied voltage $V_0 = 22.5$ V, and $E_Q = E_F$.

Since the maximum incident flux density for each state occurs at the apex and the transmission probability for all subbands $p$ reaches a maximum near the apex, the overwhelming majority of the emitted current originates from the immediate vicinity of the apex. Additionally, due to the incident flux density decaying extremely rapidly as a function of distance from the apex along the emitter surface, only the incident flux density at the apex itself is relevant. As a result, electrons in subband $p = 0$ essentially contribute all of the ECD, as it is the only subband with a non-zero transmission probability at the apex.

The shape of the tunneling probability as a function of position on the surface $(\eta_1, \xi)$ for different $p$ subbands is a direct result of the influence of the angular momentum quantum number $p$ exerts on the energy available for tunneling and the tunneling paths taken through the forbidden region. The energy available for tunneling is defined as the kinetic energy in the $\eta - \xi$ plane, given by the difference between the total energy $E_Q$ and the angular kinetic energy in Equation 4.56 and is shown as a function of $\xi$ in Figure 4-11a. Due to angular momentum, the electrons experience a force in the $\eta - \xi$ plane, arising from an effective potential term of the form $\hbar^2 p^2/(2m_0 \eta \xi)$, which tends towards infinity at the apex. As a result, the closer
Figure 4-11: The (a) energy available for tunneling at the surface for the lowest 5 $p$ subbands as a function of $\xi$ and (b) tunneling paths for $p = 0$ and $p = 4$, for electrons with $E_Q = E_F$ in an emitter of radius $\eta_1 = 1.5$ nm.

to the apex the electron is located, the larger its angular momentum and smaller its remaining momentum in the $\eta - \xi$ plane for tunneling. When the angular kinetic energy equals or exceeds the state’s total energy, corresponding to the region on the emitter surface where $\xi < \hbar^2 p^2 / (2m_0 \eta_1 E_Q)$, the electron has no momentum available for tunneling and the tunneling probability is reduced to zero.

When solving the tunneling problem in the $\eta - \xi$ plane as was done in this case, the angular momentum becomes a component of the tunneling barrier potential through adding $V_{eff}$

$$V_{eff}(\eta, \xi) = \phi_0 + E_F - \frac{eV_0}{\ln[\eta/\eta_1]} \ln[\eta/\eta_1] + \frac{\hbar^2 p^2}{2m_0 \eta \xi}. \quad (4.59)$$

Because the tunneling potential approaches infinity as $\xi \to 0$, electrons in states with $p \neq 0$ cannot tunnel from the apex in a straight line along the $z$ axis, as illustrated by the tunneling paths in Figure 4-11b. Instead, these electrons are deflected off axis by the additional term in the tunneling barrier potential and must travel a longer distance in the forbidden region before escaping, significantly reducing the transmission probability of these electrons at the apex. For the case of $p = 0$, no $\xi$-dependent effective potential term is added to the tunneling barrier potential and the highest tunneling probability occurs at the apex.
The incident flux density at the apex $\eta_0 Q(\eta_1)$ for $p = 0$ as a function of total energy $E_Q$ and subband $n$, for (a) an emitter of radius $\eta_1 = 2.5$ nm and (b) an emitter of radius $\eta_1 = 5$ nm.

Having established that the majority of the ECD originates from the apex and from subbands with $p = 0$, it is possible to investigate the distribution of current among the $n$ subbands by plotting the incident flux density at the apex as a function of subband $n$, as shown in Figure 4-12. As can be seen in Figure 4-12a for an emitter of radius $\eta_1 = 2.5$ nm, $n = n_{\text{max}} = 4$ is the highest energy subband in the emitter and contributes the largest incident flux density. Although the $n < n_{\text{max}}$ subbands contribute comparable incident flux densities at their maxima, they do so at lower total energies than $n_{\text{max}}$. At lower total energies, there is a comparatively small amount of available tunneling momentum for electrons, which severely limits their probability of tunneling and, as a result, emission from these $n < n_{\text{max}}$ subbands is greatly suppressed. As shown in Figure 4-9b, for an emitter of radius $\eta_1 = 2.5$ nm, over 90% of the total ECD is emitted by electrons in subband $n = n_{\text{max}} = 4$. As the emitter radius increases to $\eta_1 = 5$ nm, additional subbands are admitted into the emitter’s electron system, as shown in Figure 4-12b. Like the subbands for the $\eta_1 = 2.5$ nm emitter case, their incident flux density maxima are of a similar magnitude, but are offset in total energy from each other. However, more subbands contained in the same range of energies below $E_F$ results in subbands that are more closely spaced in energy, and in the case of an emitter of radius $\eta_1 = 5$ nm, the
\( n = n_{max} - 1 = 8 \) and \( n = n_{max} = 9 \) subband maxima are both located very close to \( E_Q = E_F \). This results in more balanced contributions to the ECD from electrons in the \( n_{max} \) and \( n_{max} - 1 \) subbands, respectively accounting for 60% and 35% of the total ECD. As the emitter radius is further increased, the density of states with energies near \( E_Q = E_F \) continues to increase and the lower \( n \) subbands account for a growing fraction of the total ECD.

### 4.4.3 Spatial Origin of Emitted Current Density

**Results: Effective Emitting Area and Emitting Area Efficiency**

The spatial origin of the current was investigated through calculating the *effective emitting area* \( A_{eff} \), defined as the area on the emitter surface (centered at the apex) which accounts for at least 99% of the total emitted current. The effective emitting area was determined by calculating the total ECD from an emitter of radius \( \eta_1 \), under an applied voltage \( V_0 \), and finding the point \( \xi_c \) on the emitter surface \( 0 < \xi_c < \xi_1 \) within which 99% of the total current originated. The effective emitting area as a function of emitter radius and applied voltage is shown in Figure 4-13a. For fixed \( V_0 \), increasing the emitter radius increases the effective emitting area and for fixed \( \eta_1 \), increasing the applied voltage also leads to an increase in the effective emitting area.

Of the emitter radii and applied voltages considered, the minimum of \( A_{eff} = 2.27 \text{ nm}^2 \) occurred for \( \eta_1 = 1 \text{ nm} \) and \( V_0 = 14.5 \text{ V} \) and the maximum of \( A_{eff} = 162.25 \text{ nm}^2 \) occurred for \( \eta_1 = 10 \text{ nm} \) and \( V_0 = 37.5 \text{ V} \). The *emitting area efficiency*, calculated by dividing the effective emitting area by the total emitter surface area and converting it to a percentage, is shown in Figure 4-13b as a function of emitter radius and applied voltage. The minimum emitting area efficiency of the parameters considered was \( 3.81 \times 10^{-4} \% \) for \( \eta_1 = 1 \text{ nm} \) and \( V_0 = 15 \text{ V} \), whereas the maximum of \( 8.13 \times 10^{-3} \% \) occurred for \( \eta_1 = 10 \text{ nm} \) and \( V_0 = 37.5 \text{ V} \). While it is acknowledged that the emitting area efficiency can be made arbitrarily small by increasing the height of the emitter.
Figure 4-13: (a) The effective emitting area $A_{eff}$, the area of the emitter surface from which at least 99% of the emitted current originates; and (b) the emitting area efficiency, the percentage of the total emitter surface area $A_{tot}$ represented by $A_{eff}$, as a function of emitter radius $\eta_1$ and applied voltage $V_0$.

$L_i$ as the emitter height does not significantly influence the total ECD in this model, the presented values convey the order of magnitude of the fraction of the total emitter surface area that is involved in significant emission activities.

Analysis

The dependence of the extent of the effective emitting area on the emitter radius $\eta_1$ and applied voltage $V_0$ can be explained by investigating the normalized ECD as a function of position on the surface. Figure 4-14a shows the normalized total ECD as a function of $\xi$, for various emitter radii and a fixed applied voltage. As can be seen, the maximum of the ECD profile along the surface occurs at the apex ($\xi = 0$) and the ECD decreases rapidly as a function of distance from the apex. Decreasing the emitter radius leads to a more rapid decline in the ECD with increasing $\xi$, leading to smaller emitters exhibiting an ECD profile that is more strongly peaked around the apex, as compared with larger emitter radii. This more rapid decline in the normalized ECD as a function of $\xi$ for smaller emitters is a direct consequence of the transmission probability profile along the surface. Since the majority of the ECD originates from electrons in subband $p = 0$ with energies near $E_Q = E_F$ and
Figure 4-14: The (a) normalized emitted current density and (b) tunneling barrier width for electrons in subband $p = 0$ with $E_Q = E_F$ as a function of $\xi$, for a fixed applied voltage $V_0 = 22.5$ V and varying emitter radius $\eta_1$.

electrons with zero angular momentum essentially follow a straight-line path through the forbidden region, the length of the tunneling path $s$ for such states is a good indicator of the transmission probability. The tunneling path length is defined as the distance the electron travels between leaving the surface at location $s_1 = (\eta_1, \xi_1)$ and arriving at a location on the turning surface $s_t = (\eta_v^*, \xi_t)$, where

$$\eta_v^*(\eta_1, \eta_2, V_0) = \eta_1 \left( \frac{\eta_2}{\eta_1} \right)^{\frac{\phi_0}{eV_0}},$$  \hspace{1cm} (4.60)$$

and it has been assumed that $E_Q = E_F$. The lengths of the tunneling paths for an electron in the $p = 0$ subband with $E_Q = E_F$ for various emitter radii, are shown as a function of $\xi$ in Figure 4-14b. Although the tunneling path length increases as a function of $\xi$ across all emitter radii, the smaller emitters experience a more rapid increase in $s$ near the apex, as compared with emitters of larger radii. Geometrically, the shortest distance between any two points on a pair of paraboloids which share the same focus is between their apices. As a function of $\xi$, the distance between the two points on the paraboloids connected by a straight line that is oriented normal to the surface of the inner paraboloid (approximately the tunneling path for electrons in subband $p = 0$) increases at a rate proportional to the difference between the radii.
Figure 4-15: The (a) normalized emitted current density and (b) tunneling barrier width for electrons in subband \( p = 0 \) with \( E_Q = E_F \) as a function of \( \xi \), for various applied voltages \( V_0 \) and fixed emitter radius \( \eta_1 = 1.5 \) nm.

of curvature of the inner \( \eta_1 \) and outer \( \eta_1^v \) paraboloids. Since decreasing \( \eta_1 \) linearly results in a decrease in \( \eta_1^v \) that is less than linear, given by \( \eta_1^v \propto \eta_1^{1-\phi_0/eV_0} \), the difference between the two paraboloids’ radii of curvature increases with decreasing \( \eta_1 \), causing \( s \) to increase more rapidly and the transmission probability to decrease more rapidly as a function of \( \xi \). As a result, emitters with smaller radii of curvature \( \eta_1 \) have an ECD profile that is more highly peaked about the emitter apex and have a correspondingly smaller effective emitting area.

The dependence of the effective emitting area on the magnitude of the applied voltage for a fixed emitter radius can be analyzed in a similar fashion. The normalized ECD from an emitter of radius \( \eta_1 = 1.5 \) nm as a function of \( \xi \) and applied voltage \( V_0 \) is plotted in Figure 4-15a. The plot reveals that the normalized ECD is much more highly peaked about the apex for smaller applied voltages than it is for larger applied voltages. A plot of the tunneling path length \( s \) as a function of \( \xi \) for various applied voltages, as shown in Figure 4-15b, reveals that \( s \) varies more rapidly as a function of \( \xi \) for smaller values of \( V_0 \), causing a more rapid decrease in the normalized ECD. According to Equation 4.60, the radius of curvature \( \eta_1^v \) for the paraboloid defined by the turning surface decreases exponentially with linearly increasing \( V_0 \). As a result, for the smallest considered value of \( V_0 = 15 \) V, the discrepancy between
the radii of curvature $\eta_1$ and $\eta_t^v$ is at a maximum, $s$ increases more rapidly with $\xi$, and the normalized ECD decays more rapidly as a function of $\xi$. As $V_0$ is increased, $\eta_t^v$ decreases and the discrepancy between the radii of curvature $\eta_1$ and $\eta_t^v$ is less significant, leading to a less rapidly increasing tunneling path length as a function of $\xi$, a normalized ECD that decays less rapidly in $\xi$, and a larger effective emitting area.

4.4.4 Emitted Current Density vs. Emitter Radius

Results: Fixed Apex Transmission Probability and Fixed Applied Voltage Cases

The emitted current density from the paraboloidal emitter was investigated under two different sets of constraints as a function of emitter radius $\eta_1$: i) constant transmission probability for electrons emitted from the apex with $E_Q = E_F$; and ii) constant applied voltage $V_0$ and anode radius $\eta_2$. The constraint of a constant transmission probability at the apex manifests the effects of quantum confinement on the ECD by minimizing changes in the electrostatics as a function of emitter radius. As is discussed in §4.4.2 and §4.4.3, the overwhelming majority of the ECD originates from the high-energy $p = 0$ subbands and from the immediate vicinity of the emitter apex. As a result, electrons which are emitted from the apex with $E_Q \approx E_F$ dominate the physics of emission. If the transmission probability of these electrons is held constant across all emitter radii, the influence of the variations in the electrostatics on the ECD as a function of emitter radius can be mitigated. Figure 4-16a shows the total ECD as a function of emitter radius $\eta_1$, with the applied voltage varied to maintain a constant transmission probability at the apex, as discussed above. As a reference, the transmission probability is fixed at $D_0 = 4.5 \times 10^{-14}$, which is the value corresponding to a surface field of $F = 1 \text{ V/nm}$ in the FN model, using an exact triangular barrier. Calculations show that as the emitter radius decreases, the
Figure 4-16: Emitted current density vs. emitter radius $\eta_1$ for the case of (a) fixed transmission probability at the apex, achieved by varying the applied voltage $15V < V_0 < 53V$, and (b) fixed applied voltage $V_0 = 22.5$ V. The ECDs from the paraboloidal emitters are plotted as red points, while the predictions of the elementary FN equation with the $F_a$ of the paraboloidal emitters are plotted as black, dashed lines.

ECD is reduced, decreasing a total of two orders of magnitude from $\eta_1 = 10$ nm to $\eta_1 = 0.75$ nm. This decrease in the ECD occurs despite the increase in the magnitude of the apex field (see tunneling barrier shape in Figure 4-20a), which, instead, causes the elementary FN ECD to increase by approximately four orders of magnitude over the same range of radii. While the discrepancy between the elementary FN ECD and the paraboloidal ECD is less than one order of magnitude for $\eta_1 = 10$ nm, it is approximately six orders of magnitude for $\eta_1 = 0.75$ nm.

The second set of constraints sought to reveal the effects of substituting an emitter of a different radius into a device in which the applied voltage and anode radius are fixed. The plot of the ECD as a function of emitter radius, with a fixed applied voltage of $V_0 = 22.5$ V is shown in Figure 4-16b. Contrary to the fixed transmission probability case, the ECD under a constant applied voltage increases rapidly as a function of decreasing emitter radius, at a rate similar to that of the elementary FN equation (black, dashed line). Overall, the ECD from the paraboloid increased by approximately 15 orders of magnitude during the course of decreasing the emitter radius from $\eta_1 = 10$ nm to $\eta_1 = 0.75$ nm, but remained approximately two to four
Figure 4-17: Plots of (a) the total electron density averaged across the emitter surface and (b) normal electron velocity at the surface averaged over all subbands, across the emitter surface, as a function of emitter radius $\eta_1$.

orders of magnitude lower than the elementary FN model prediction.

Analysis

With the transmission probability held constant, the reduction in the ECD must result from a decrease in the incident flux density (product of electron density and normal velocity) at the surface, as the emitter radius decreases. Plots of the surface-averaged total electron density and average electron normal velocity at the emitter surface in Figure 4-17 reveal a reduction in both quantities as the emitter radius decreases. In the case of the electron density, reducing the emitter radius decreases the number of electronic states supported by the emitter ($E_Q < E_F$), which reduces the density of states as a function of energy, and consequently, spatially as well. The average normal velocity of electrons at the emitter surface decreases as a function of decreasing radius for a similar reason: upon reducing the density of states, the density of states with energies near $E_Q = E_F$ is significantly reduced, as the spacing between subbands increases as a function of increasing energy, and the average energy of electrons at the surface decreases. For large emitter radii, the elementary FN and paraboloidal emitter model ECDs converge towards the same value, as the paraboloid’s electron system approaches that of a bulk emitter and the reduced curvature of the emitter.
As Figure 4-18a illustrates, the increase in the ECD as a function of decreasing emitter radius for a fixed applied voltage is a result of the increase in the transmission probability at the apex as the emitter shrinks in size. The transmission probability at the apex is enhanced due to a thinning of the tunneling barrier, as shown in Figure 4-18b. Although the incident flux density at the surface is reduced, the gains in the transmission probability at the apex are much more significant and the ECD increases as emitter dimensions shrink. The elementary FN equation overpredicts the ECD from paraboloidal emitters across all radii due to its omission of the reduced electron supply from quantum confinement and the paraboloidal emitter’s wider tunneling barrier for equivalent values of $F_a$.

4.4.5 Total Energy Distribution

Results: Fixed Applied Voltage and Fixed Emitter Radius Cases

The total energy distribution (TED) of emitted electrons was investigated as a function of emitter radius $\eta_1$ under two sets of conditions: i) for a fixed applied voltage $V_0 = 22.5$ V.
Figure 4-19: The normalized total energy distribution of emitted electrons with $T = 300$ K for the cases in which (a) the applied voltage is fixed at $V_0 = 22.5$ V and the emitter radius $\eta_1$ is varied, and (b) the emitter radius is fixed at $\eta_1 = 1.5$ nm and the applied voltage $V_0$ is varied. The corresponding elementary Fowler-Nordheim total energy distributions for equivalent $F_a$ are plotted as dashed lines in the same color as the paraboloidal emitter results, which are plotted as solid lines.

$V_0 = 22.5$ eV and varied emitter radius $\eta_1$; and ii) for a fixed emitter radius $\eta_1 = 1.5$ nm and varied applied voltage $V_0$. Figure 4-19a plots the TED for the former case, in which $V_0$ is held fixed for different values of $\eta_1$. As the emitter radius decreases, the shape of the TED becomes narrower, leading to a more highly peaked TED with a maximum closer to $E_Q = E_F$ than for emitters of larger radii. Additionally, as the emitter radius increases, the elementary FN TED and the paraboloidal emitter ECD converge towards a common TED profile, which would correspond to a paraboloidal emitter with $\eta_1 \gg 10$ nm. For the case of a fixed emitter radius and different applied voltages in Figure 4-19b, the TED profile narrows significantly and its maximum shifts closer to $E_Q = E_F$ with decreasing applied voltage. Compared with the FN TED for an equivalent applied voltage, the TEDs for paraboloidal emitters of the considered radii have the same general shape, but are narrower and have maxima located closer to the Fermi energy.
Figure 4-20: Plots of the tunneling barrier shape at the apex for (a) $V_0 = 22.5$ V and various emitter radii $\eta_1$ and (b) $\eta_1 = 1.5$ nm and various applied voltages $V_0$, as a function of distance along the $z$ axis. The dashed lines originating from the apex depict the exact triangular barrier defined by the electric field magnitude at the corresponding paraboloidal emitter apex $F_a$, while the dashed black line marks the Fermi energy $E_F$.

**Analysis**

The shape of the total energy distributions is a direct consequence of the shape of the tunneling barrier. Figure 4-20a plots the tunneling barrier at the apex as a function of distance along the $z$ axis for a fixed applied voltage and various emitter radii $\eta_1$. For $\eta_1 = 1$ nm, the tunneling barrier shape is highly non-linear and convex. As a result, as the tunneling energy decreases, the width of the tunneling barrier increases rapidly, leading to a very sharp decrease in the transmission probability and a TED that is highly peaked about energies near $E_F$. Due to this convexity, the shape of the barrier prefers higher energies for emission and suppresses emission from lower energies. As the emitter radius increases, the barrier shape becomes more linear and the reduction in tunneling probability with decreasing energy becomes less dramatic, leading to a broader peak in the TED with a maximum shifted slightly lower in energy. Also a result of the convex shape of the tunneling barrier, the triangular potential barrier defined by the electric field at the paraboloid apex $F_a$ severely underestimates the barrier width for a given tunneling energy, as shown by
the dashed curves in Figure 4-20a. The lack of curvature in the exact triangular tunneling barrier potential also causes the elementary FN TED to be much broader than for paraboloidal emitters when plotted for the same $F_a$. As the barrier shape becomes more linear with increasing $\eta_1$ however, the discrepancy between $F_a$ for an FN emitter and the paraboloidal emitter begins to vanish and the two TEDs begin to converge to a common profile.

The same rationale underlies the shape of the TED profiles as a function of varying applied voltage with fixed emitter radius. Figure 4-20b illustrates that for smaller applied voltages, the potential barrier shape is highly non-linear and convex, leading to strong suppression of the tunneling probability for electrons with lower energies and a TED that is more highly peaked about $E_Q = E_F$. As the applied voltage is increased, the potential drop between the emitter surface and the anode increases, necessitating a larger slope (larger $F_a$) near the emitter surface, which results in a smaller fractional change in the tunneling barrier width as a function of energy. The corresponding TED is broader than for smaller applied voltages and its peak shifts towards lower energies, as tunneling from lower energy states is less strongly suppressed. Increasing the applied voltage does not rapidly decrease the discrepancy between the elementary FN TED and paraboloidal TED because variations in the applied voltage do not significantly affect the basic shape of the potential barrier. As a result, for nanoscale paraboloidal emitters, across all electric field strengths relevant to field emission, the exact triangular barrier corresponding to the $F_a$ of a paraboloidal emitter is a poor approximation to the actual tunneling barrier potential and elementary FN theory is clearly an inadequate substitute for a geometry-specific treatment of field emission.
Figure 4-21: Fowler-Nordheim plots for emitters of varying radii and an anode radius of \( \eta_2 = 50 \) nm, expressed in terms of (a) the emitted current density \( J \) vs. the apex electric field \( F_a \) and (b) the emitted current \( I \) vs. the applied voltage \( V_0 \).

4.4.6 Fowler-Nordheim Plots

Results: J vs. F and I vs. V Plots

Fowler-Nordheim plots for the emitted current density \( J \) versus apex electric field \( F_a \) and emitted current \( I \) versus applied voltage \( V_0 \) from paraboloidal emitters of various radii are shown in Figure 4-21a and Figure 4-21b. While the FN plots for emitters of radii larger than \( \eta_1 = 5 \) nm appear relatively linear, as predicted by elementary FN theory (dashed line in Figure 4-21a), the curves associated with smaller emitters are characterized by an increasing downward slope for lower apex electric fields (\( F \approx 2 \) V/nm), and for a wide range of applied voltages. Table 4.1 lists the local work function \( \phi_0 \) and apex field factor \( \beta_a \) extracted from the FN plots in Figure 4-21 along with the actual values used in the calculations. The work function and field factors were extracted from the slopes of linear regressions of the most linear portions of the curves in the \( J \) vs. \( F_a \) and \( I \) vs. \( V_0 \) FN plots, via

\[
\phi_0 = \left( -\frac{m}{B} \right)^{2/3} \quad \text{and} \quad \beta_a = -\frac{B}{m} \phi_0^{3/2},
\]  

(4.61)
Table 4.1: The extracted work function $\phi_0$ from the $J$ vs. $F_a$ FN plots in Figure 4-21a and the extracted apex field factor $\beta_a$ (with $\phi_0 = 4.5$ eV) from the $I$ vs. $V_0$ FN plots in Figure 4-21b, for each emitter radius. The actual value of the work function used in the calculations was $\phi_0 = 4.5$ eV and the actual field factors for the emitter tips are listed in the fourth row of the table.

<table>
<thead>
<tr>
<th>Radius [nm]</th>
<th>1.0</th>
<th>2.5</th>
<th>5.0</th>
<th>7.5</th>
<th>10.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extracted Work Function [eV]</td>
<td>12.34</td>
<td>6.82</td>
<td>5.75</td>
<td>5.56</td>
<td>5.24</td>
</tr>
<tr>
<td>Extracted Field Factor [nm$^{-1}$]</td>
<td>0.135</td>
<td>0.155</td>
<td>0.133</td>
<td>0.110</td>
<td>0.103</td>
</tr>
<tr>
<td>Actual Field Factor [nm$^{-1}$]</td>
<td>0.511</td>
<td>0.267</td>
<td>0.174</td>
<td>0.141</td>
<td>0.124</td>
</tr>
</tbody>
</table>

where $B \approx 6.83 \text{ Vnm}^{-1}\text{eV}^{-3/2}$ is the 2nd FN constant and $m$ is the extracted slope of the appropriate FN plot. The slopes extracted from the $J$ vs. $F_a$ FN plots of the paraboloidal emitters consistently overestimate the local work function at the emitter surface, as compared with work function of $\phi_0 = 4.5$ eV actually used in the model calculations. The error in the extracted work function is more significant for smaller emitters, yielding extracted work functions as large as $\phi_0 = 12.34$ eV for the emitter of radius $h_1 = 1$ nm. The margin of error between the extracted and actual work functions decreases with increasing emitter radius, but is still appreciable for the emitter of radius $h_1 = 10$ nm, which has an extracted work function of $\phi_0 = 5.24$ eV.

In the case of the field factors, the extracted FN plot slopes consistently underestimate the actual values, with the discrepancy again being larger for smaller emitters. The FN plot for an emitter of radius $h_1 = 1$ nm yields a field factor of $\beta_a = 0.135 \text{ nm}^{-1}$, whereas the actual value derived from the model is $\beta_a = 0.511 \text{ nm}^{-1}$. Again, increasing the emitter radius leads to a decrease in the discrepancy between the extracted and actual field factors, with the extracted field factor of the emitter of radius $h_1 = 10$ nm being $\beta_a = 0.103 \text{ nm}^{-1}$, as compared with the actual value of $\beta_a = 0.124 \text{ nm}^{-1}$.
Fowler-Nordheim plots are linear for the exact triangular barrier because the transmission probability varies with the electric field at the surface as $1/F$ (and with the applied voltage as $1/V_0$); the non-linearities in the FN plots for the paraboloidal emitters are a direct result the tunneling probability through the paraboloidal cathode-anode barrier potential not having this functional form. The $1/F$ form of the tunneling probability expression arises from the barrier width having the same dependencies, given by

$$s_{tri} = \frac{\phi_0 + EF - W}{eF} = \frac{d}{eV_0} \phi_0 + EF - W,$$

(4.62)

where $W$ is the kinetic energy in the tunneling direction and $d$ is the spacing between the planar anode and emitter cathode. Additionally, the exact triangular barrier has the advantage that the potential at every point between the cathode and anode can be expressed in terms of the magnitude of the electric field at the surface $F$. The paraboloidal cathode-anode system, however, has a potential barrier that is non-linear as a function of distance traveled from the surface, and as a consequence, the apex field $F_a$ is not a reliable proxy for the potential at points away from the emitter surface. The tunneling barrier width at the apex of a paraboloidal emitter, for an electron with tunneling energy $W$ has an inverse exponential dependence on both the apex field $F_a$ and applied voltage $V_0$, given by

$$s_{par} = \eta_1 \left[ \left( \frac{\eta_2}{\eta_1} \right)^{\frac{\phi_0 + EF - W}{eF} - \frac{W}{eV_0}} - 1 \right] = \eta_1 \left[ \left( \frac{\eta_2}{\eta_1} \right)^{\frac{\phi_0 + EF - W}{eV_0} - 1} \right].$$

(4.63)

The immediate consequence of the tunneling barrier width having these dependencies is illustrated by the plots in Figure 4-22a and Figure 4-22b of the log of the transmission probability for $p = 0$ electrons with $W = EF$ vs. $1/F_a$ and $1/V_0$ for paraboloidal emitters of various radii. Similar to FN plots, these plots would all form straight lines as a function of either $1/F_a$ (the dashed line in Figure 4-22a) and $1/V_0$ if they
Figure 4-22: Plots of the tunneling probability (a) as a function of the reciprocal of the magnitude of the electric field at the apex $1/F_a$ and (b) as a function of the reciprocal of the applied voltage $1/V_0$. The black dashed line is the tunneling probability for the exact triangular barrier of the elementary Fowler-Nordheim equation.

were constructed for the exact triangular potential barrier; instead, like the FN plots for the paraboloidal emitters in Figure 4-21, they curve rapidly downward, roughly due to the barrier width’s exponential $1/V_0$ or $1/F_a$ dependence. The significant non-linearities for smaller emitter radii are due to a higher sensitivity of the convex barrier shape to variations in the applied voltage, as shown in Figure 4-23, which causes $s_{par}$ to vary more rapidly with changes in $V_0$ (and $F_a$) for a given tunneling energy $W$.

Due to the tendency of these FN plots to curve downward, the extracted slopes are larger than predicted by FN theory and provide larger extracted work functions and smaller extracted field factors than the actual values. For larger emitters, the barrier potential more closely resembles a triangular barrier and the variation in the log of the tunneling probability as a function of $1/V_0$ and $1/F_a$ appears more linear. As a result, the extracted work functions and field factors gradually approach the values used in the calculations as a function of increasing emitter radius.
Figure 4-23: The tunneling barrier potential at the apex of a paraboloidal emitter of radius \( \eta_1 = 1 \) nm and an anode of radius \( \eta_2 = 50 \) nm for various applied voltages \( V_0 \), as a function of distance along the \( z \) axis. The dashed lines correspond to the exact triangular barrier defined by the electric field magnitude at the emitter apex \( F_a \) and the solid, gray line denotes the tunneling energy of the electron.

### 4.5 Comparison to Experimental Data: Molybdenum Field Emitter Arrays

#### Results

The nanoscale, metal paraboloidal emitter model was compared to experimental data from molybdenum field emitter arrays (FEAs), fabricated by Tsujino et al. [142] In the study, Tsujino et al. fabricated and characterized two 40 × 40 arrays of single-gated FEAs: one composed of emitter tips with an apex radius of curvature of approximately 5 nm and one composed of emitter tips with an apex radius of curvature of approximately 10 nm. A diagram of the emitter array, including the emitter tips and single gate is shown in Figure 4-24a. The paraboloidal emitter model was adapted to the experimental setup by calculating the total emitted current from 1600 individual emitters with the paraboloidal cathode-anode geometry, using the emitter properties reported by the work and no fitting factors. Each emitter was defined by the following attributes: apex radius of curvature \( \eta_1 = 5 \) nm or \( \eta_1 = 10 \) nm, height \( L = 1.2 \) \( \mu \)m,
Figure 4-24: (a) Diagram of a single-gated field emitter array of molybdenum field emitter tips of radius 5 nm or 10 nm, and height 1.2 µm, with a gate aperture diameter of 2.3 µm. (b) I-V characteristics for three fabricated molybdenum field emitter arrays, the paraboloidal emitter model (circles), and the generalized nanoscopic Fowler-Nordheim model (dashed lines). While the device labeled “Single-gate FEA 1” (green circles) is composed of emitters with apex radii of curvature 5 nm, the device labeled “Single-gate FEA 2” (unfilled, blue squares) is composed of emitters with apex radii of curvature 10 nm. Model results for comparison with “Single-gate FEA 1” are plotted in black, while model results for comparison with “Single-gate FEA 2” are plotted in gray.

The results are plotted in Figure 4-24b, along with the measured FEA data and the ECD predicted by the generalized nanoscopic Fowler-Nordheim model of Kyritsakis et al. in Equation 2.31 [2], with the same electronic properties and apex field as the paraboloidal emitter above. Since the generalized nanoscopic FN equation gives the emitted current density, an effective emitting area of $A_{eff} = 68.4$ nm$^2$ was assumed for the $\eta_1 = 5$ nm device and $A_{eff} = 162.3$ nm$^2$ was assumed for the $\eta_1 = 10$ nm device, which were taken from the $V_0 = 37.5$ V curve of Figure 4-13a.
For the $\eta_1 = 5$ nm device “Single-gate FEA-1” (green circles), the paraboloidal emitter model (black circles) closely matches the measured I-V characteristics until approximately $V_{ge} = 90$ V, after which the measured current’s rate of increase with $V_{ge}$ slows and the metal paraboloid current overtakes it. While the paraboloidal emitter model assumes that all emitters contribute to the emitted current equally and suffer from no detrimental effects due to high voltages or currents, emitter tips in real FEAs may experience damage, blunting, or burn-out, which causes them to emit less current or cease emission altogether, resulting in flatter I-V characteristics. The generalized nanoscopic FN model (black dashed line) overestimates the measured FEA ECDs by approximately 2 orders of magnitude across the whole range voltages swept for this device, likely due to the assumption of a bulk emitter electron supply or the high curvature of the emitter tip being outside the range of validity for the potential expansion performed in the derivation of the equation. Kyritsakis et al. have reported that the equation is only appropriate to use for emitters with radii of curvature at the apex greater than 5 nm.

The paraboloidal emitter model (gray circles) does not match the I-V characteristics for the $\eta_1 = 10$ nm device “Single-gate FEA-2” (unfilled, blue squares) as closely. While Single-gate FEA-2 turns on around $V_{ge} = 50$ V, the paraboloidal emitter model predicts that the same current will be reached approximately 25 V later, thus underestimating the measured emitted current by approximately 2 orders of magnitude across the range of swept voltages. Although the generalized nanoscopic FN model (gray dashed line) predicts device turn on roughly 10 V later than Single-gate FEA-2, its current rises with a larger slope than the measured characteristics, eventually overtaking the measured I-V characteristics at around $V_{ge} = 105$ V.
Analysis

The fact that the paraboloidal emitter model and generalized nanoscopic FN model overestimate the emitted current from the $\eta_1 = 5$ nm device, but underestimate (for most of the voltage sweep range) the emitted current from the $\eta_1 = 10$ nm device points to a discrepancy between the models for emitter electrostatics being employed and the actual emitter electrostatics. As the transmission probability roughly depends exponentially on the tunneling barrier width, the emitted current is extremely sensitive to the shape of the potential near the emitter surface and any differences between the assumed and actual potentials are amplified in the ECD results.

The paraboloidal cathode-anode potential model clearly overestimates the effect of barrier thinning due to changes in emitter geometry (electric field enhancement), as evidenced by the larger gap between the model’s predicted I-V curves for the 5 nm and 10 nm devices, as compared with the gap between the measured I-V curves for the 5 nm and 10 nm devices. For example, to emit a current of 10 nA from the $\eta_1 = 10$ nm device, the paraboloidal emitter model predicts that the gate-emitter voltage must be approximately $V_0 = 95$ V. Shrinking the emitter radius to 5 nm reduces the gate-emitter voltage required to emit the same current to approximately $V_0 = 55$ V: a change of $\Delta V = 40$ V. In contrast, the same applied voltage difference for the experimental I-V curves was approximately $\Delta V = 10$ V. As a result of the model’s good agreement with the measured I-V curves for the 5 nm device and underprediction of the current emitted from the 10 nm device, the transmission probability of electrons in the $\eta_1 = 10$ nm device is underestimated by the paraboloidal cathode-anode model and the decrease in transmission probability with increasing emitter radius is overestimated. Therefore, the correct model for emitter electrostatics likely consists of a barrier shape that is somewhere in between that of the convex barrier of the paraboloidal cathode-anode geometry and the FN model’s triangular barrier (see Figure 4-23); it should be more linear than that of the paraboloidal cathode-anode...
model so that the tunneling barrier width changes less dramatically with variations in the applied voltage, but not triangular as in the FN case, due to the high curvature of the emitter tip.

A likely more appropriate cathode-anode geometry candidate for modeling the electrostatics of this particular device design lies within the prolate spheroidal coordinate system, as the surfaces of constant coordinate in the emission direction range from a hyperboloid of revolution to a plane in Cartesian coordinates. In this coordinate system, one could obtain a one-dimensional, analytical expression for the potential between a curved tip and a planar anode, which could serve as a more geometrically accurate approximation to emitter electrostatics in this case.

4.6 Chapter Summary

In this chapter, a realistic field emitter geometry was modeled by a nanoscale, metal, paraboloidal cathode and its emission properties were investigated as a function of the emitter radius and applied voltage. As the anode was also chosen to be a paraboloid, the tunneling potential barrier created between the anode and the cathode decays logarithmically as a function of distance from the surface, as apposed to the linear nature of the exact triangular barrier, and consists of paraboloidal surfaces of constant potential. The emitted current density equation for the paraboloidal emitter was derived from the semiclassical framework for field emission, as detailed in Chapter 3, and was employed to investigate various emitter properties. Results revealed that the majority of the emitted current originates from an area in the immediate vicinity of the apex, with an area between 2.27 nm² and 162.25 nm² contributing 99% of the total emitted current for emitters of radii between 1 nm and 10 nm. Additionally, regardless of the emitter radius, the entirety of the current originated from subbands with zero angular momentum, as they have the highest probability of tunneling from the apex.
For emitters with a radius less than 5 nm, the majority of the ECD originated from a single, high-energy subband denoted by $n_{\text{max}}$. Although quantum confinement of the electronic wave functions in the emitter reduces the density of states and supply of electrons to the emitter surface, the paraboloidal emitter ECD increased as a function of shrinking emitter dimensions under a fixed applied voltage, due to significant gains in the tunneling probability from electric field enhancement at the emitter apex. Due to the convex profile of the potential barrier which strongly suppresses tunneling from lower energies, the shape of the total energy distribution for the paraboloidal cathode-anode system was much narrower and more highly peaked about energies near the Fermi energy $E_F$ than predicted by the elementary FN equation, when calculated for equivalent apex electric fields. The non-linear nature of FN plots for paraboloidal emitters of radii below approximately 5 nm is also a consequence of the non-triangular tunneling barrier. Due to the downward curvature of these FN plots, slopes extracted from J-F or I-V data caused significant overestimations of the local work function and underestimations of the field factor at the apex, as compared with the actual values. In summary, for highly-curved emitters with radii of curvature below approximately 5 nm, the elementary Fowler-Nordheim equation fails from an electron supply and electrostatics standpoint, and instead, a geometry-specific model for emission must be employed.
Chapter 5

Field Electron Emission from Bulk Silicon

5.1 Introduction

One of the primary applications of field emitter arrays (FEAs) is as high-density current sources for nanoelectronic devices. From the standpoint of providing the highest emitted current density (ECD), FEAs consisting of semiconductor field emitter tips are not ideal, due to their relatively low electron densities ($n \approx 10^{15} - 10^{18} \text{ cm}^{-3}$), as compared with metals ($n \approx 10^{23} \text{ cm}^{-3}$). Unfortunately, the potential usefulness of metal FEAs as current sources in nanoelectronic devices is limited by their lack of compatibility with complementary metal-oxide-semiconductor (CMOS) processing technologies, with which almost all modern integrated circuits are fabricated. Accordingly, in competing with other technologies for widespread technological relevance, CMOS compatibility would be a major advantage for devices based on FEAs, as it would allow for the electron sources to be integrated directly into the devices and would be able to leverage the various advantages of the CMOS processing framework. Since silicon FEAs share fabrication technologies with silicon CMOS processes, they
Figure 5-1: SEM images of (a) an ungated silicon field emitter array and (b) a single-gated silicon field emitter array [6, 7].

are a prime candidate for integrated nanoscale current sources. Examples of silicon FEAs are shown in Figure 5-1.

As mentioned in Chapter 2, the first theoretical treatment of field emission from semiconductors was developed by Stratton in the 1950s and 1960s [97–99]. Like Fowler-Nordheim theory for emission from metals, Stratton’s model captured the most fundamental aspects of emission from semiconductors, such as contributions to the emitted current density from the conduction band and valence band, the influence of the effective mass of electrons, and regimes of field emission at higher temperatures. A few years later, Christov sought to unify the field and thermionic emission equations for semiconductors and provided equations for a wider range of temperatures and applied fields [112]. Perhaps the most complete treatment of field emission from semiconductors was presented by Modinos [8], which is based upon a more detailed physical description of semiconductor surfaces and includes contributions to the ECD from surface states. While some of the above-mentioned models include various physical phenomena that are related to downward band banding due to electric field penetration into the semiconductor, none of them incorporate the formation of a 2D electron gas near the surface or the portion of the forbidden energy gap through which electrons with energies near the valence band edge must tunnel. Additionally, due
to the analytical nature of these models, the effects of a limited electron flux density supply to the emitter surface from the bulk are neglected.

The goal of this chapter is to construct a model of field emission from bulk silicon, which includes emission from: i) free-electron-like extended states in the conduction band; ii) bound states (2D electron gas) in the conduction band, near the surface; and iii) free-electron-like extended states in the valence band. With regard to emission from the conduction band, an algorithm is implemented for calculating the ECD which ensures that the ECD does not exceed the incident flux density supplied by the emitter bulk. For valence band emission, an additional tunneling probability is introduced to account for the probability of electrons with energies near the valence band edge in the bulk to tunnel through a portion of the forbidden energy gap before reaching the external tunneling barrier in vacuum, which is due to downward band bending near the surface. In order to investigate the effects of the new additions to the silicon field emission model on various properties of emission, the total emitted current density, total energy distribution of emitted electrons, and Fowler-Nordheim plots are analyzed as a function of the applied electric field and across a range of dopant densities in n-type and p-type silicon emitters.

5.2 Bulk Silicon Emitter Model

5.2.1 Additional Physical Phenomena in Silicon Field Emitters

In order to construct a model for field emission from silicon, there are additional physical phenomena inherent to semiconductors that need to be considered, which are not present in metals. These phenomena can be categorized by the specific semiconductor property from which they arise: i) effects resulting from the band structure of silicon and ii) effects resulting from the finite dielectric constant of silicon. The
band structure of silicon necessitates that the model account for the: i) field-emitted current contributed from its conduction band and each of its three valence bands; ii) current contributions from each of its six conduction valleys in the conduction band; and iii) the different effective masses of electrons in each of these conduction valleys. The finite dielectric constant of silicon allows the electric field applied to the emitter surface in the vacuum to penetrate some distance into its interior, causing the bands near the surface to bend downward. The downward band bending causes the formation of a two-dimensional electron gas in the conduction band near the surface, a lowering of the surface barrier height for a fixed reference energy, and the appearance of an additional, internal tunneling barrier (forbidden energy gap) for electrons attempting to tunnel from near the top of the valence band. The finite dielectric constant also requires that the image charge correction in the Schottky-Nordheim barrier be slightly modified. These physical phenomena and how they affect the model of field emission from silicon are discussed in more detail in the following sections.

**Band Structure of Silicon**

As shown in Figure 5-2a, the band structure of silicon consists of a conduction band and three valence bands, designated the heavy hole, light hole, and split-off bands. The electron affinity of silicon is \( \chi_e = 4.04 \text{ eV} \) and the conduction band and valence bands of silicon are separated by a forbidden energy gap of approximately \( E_g = 1.12 \text{ eV} \). Due to the symmetry of the silicon crystal structure, the conduction band contains six equivalent conduction valleys with ellipsoidal constant energy surfaces located at the \( X \) point [11], as shown in Figure 5-2b. As the emission direction is defined as along the positive \( z \) axis of (001) silicon, the two conduction valleys which lie along the \( k_z \) axis in momentum space are termed “on-axis” conduction valleys and contain electrons with normal (parallel to emission direction) effective mass \( m_{on} \) and transverse (orthogonal to emission direction) effective mass \( m_{tm} \). The four conduction
Figure 5-2: (a) The band structure of silicon consists of a single conduction band $E_c$ and three valence bands: heavy hole $E_{v,hh}$, light hole $E_{v,lh}$, and split-off $E_{v,so}$, separated by an energy gap $E_g = 1.12$ eV, in which no states occur inherently. The electron affinity of silicon is $\chi_e = 4.04$ eV. (b) The conduction band of silicon has six conduction valleys, represented here by their ellipsoidal constant energy surfaces. Owing to the definition of the emission direction as along the $z$ axis of (001) silicon, conduction valleys which lie along the $k_z$ axis are defined as “on-axis”, while conduction valleys which lie along the $k_x$ and $k_y$ axes are termed “off-axis”. 

valleys which lie along the $k_x$ or $k_y$ axes are termed “off-axis” conduction valleys and are characterized by electrons with normal effective mass $m_n^{off}$ and transverse effective mass $m_t^{off}$. The constant energy surfaces of the valence bands are spherical and thus, holes in each valence band are characterized by a single effective mass: $m_{hh}$ in the heavy hole band, $m_{lh}$ in the light hole band, and $m_{so}$ in the split-off band. Table 5.1 lists the standard effective masses for each of the conduction valleys and valence bands for bulk silicon, in terms of the free electron mass. In the case of emission from other surfaces of silicon, such as the (111) surface, the normal and transverse effective masses change due to the new constant energy surface shapes, which correspond to electron energies in the (eight) conduction band valleys; properties of electrons in the valence bands would be unaltered, due to their spherical constant energy surfaces.
Table 5.1: The standard values for the effective masses of electrons $m_e$ in the “on-axis” and “off-axis” conduction bands and the effective masses of holes in the three valence bands of silicon $m_v$, in terms of the free electron mass $m_0$. Due to the orientation of the constant energy surfaces of the off-axis conduction valleys, the transverse effective mass of electrons $m_{t\text{off}}$ is the geometrical mean of the normal and transverse effective masses of the on-axis valleys $m_{t\text{off}} = (m_{n\text{on}} m_{t\text{on}})^{0.5}$ [11].

## Contributions from Multiple Conduction Valleys and Valence Bands of Silicon

Due to the multiple conduction valleys in the conduction band and the multiple valence bands of silicon, the emitted current density from each of these must be incorporated into the model. The contributions from each of these emission sources simply sum, such that the total emitted current density is given as follows:

$$J = 2J_{c\text{on}}^c + 4J_{c\text{off}}^c + J_{v\text{hh}}^h + J_{v\text{lh}}^l + J_{v\text{so}}^s,$$

(5.1)

where $J_{c\text{on}}^c$ is the conduction band current emitted from each on-axis conduction valley, $J_{c\text{off}}^c$ is the conduction band current emitted from each off-axis conduction valley, and $J_{v\text{hh}}^h$, $J_{v\text{lh}}^l$, and $J_{v\text{so}}^s$ represent the current contributions from the heavy hole, light hole, and split-off valence bands. Although the contributions from each of the conduction valleys in the conduction band are treated identically in the model, they differ in the values of the normal and transverse effective masses of electrons, which have a significant impact on the energies of bound states near the surface and the amount of normal energy available for tunneling, gained upon leaving the semiconductor (both discussed below). In addition to the different effective masses of holes in each of the
valence bands, the valence band maxima are slightly offset in energy, with the split-off band lying lower in energy than the degenerate heavy and light hole bands. As a result, the energy of the highest-energy occupied state is lower for the split-off band than for the heavy and light hole valence bands, and consequently, the energy gap between the split-off band maximum and the conduction band minimum is larger.

Change of Electron Effective Mass Upon Leaving the Semiconductor

The effective mass is a convenient way to group the different forces that the crystal structure exerts on electrons and holes in a semiconductor into a single term, which describes electron and hole kinematics within the crystal. The specific values for effective masses of electrons and holes result from the degree of curvature of the energy dispersion relation near the bottom of the conduction band and top of the valence bands respectively. The notion of a particle having an effective mass, however, loses its meaning outside of the semiconductor crystal, as an electron in vacuum is free of the forces from the crystal structure and its kinematics can be described using the free electron mass $m_0$. As a result, the moment the electron leaves the crystal and enters vacuum, the mass of the electron used for all relevant calculations must change from the effective mass to the free electron mass. The consequences of this “change” in the electron’s mass must be accounted for in models of field emission from semiconductor surfaces.

Consider an electron in the conduction band which has a transverse effective mass inside the semiconductor $m_t$, that the emitting surface is perfectly planar, and that everywhere the electric field is normal to the emitting surface. Inside the semiconductor, the electron has a total energy equal to

$$E = \frac{p_t^2}{2m_t} + \frac{p_z^2}{2m_n} = E_t + W$$

(5.2)
where $p_t$ is the total transverse momentum of the electron, $p_z$ is the normal momentum of the electron, $m_n$ is the electron’s normal effective mass, $E_t$ is the energy corresponding to the total transverse momentum, and $W$ is the energy corresponding to the normal momentum. Due to the fact that the electric field exerts a force on the electron only in the $z$ dimension, the transverse momentum of the electron must be the same on either side of the interface that divides the semiconductor crystal from vacuum, such that $p'_t = p_t$, where ($'$) denotes a quantity measured in vacuum.

As energy is also conserved in this system, the relationship between the momentum inside the semiconductor and in vacuum can be expressed as

$$E' = \frac{p_t^2}{2m_0} + \frac{p_z^2}{2m_n} = \frac{p_t^2}{2m_t} + \frac{p_z^2}{2m_n} = E.$$  \hspace{1cm} (5.3)

From the requirement that total energy be conserved and that the transverse components of momentum be equal on either side of the vacuum-semiconductor interface, the following expression for the normal energy in vacuum $W'$ results:

$$W' = \left(1 - \frac{m_t}{m_0}\right) E_t + W$$  \hspace{1cm} (5.4)

where $E_t = p_t^2/(2m_t)$ and $W = p_z^2/(2m_n)$. Equation 5.4 states that the normal energy in vacuum, or tunneling energy, is altered by an amount equal to the difference between the transverse energy inside and outside the semiconductor. In the case of silicon, all of the standard effective masses are smaller than that of the free electron mass, which results in an increase in the normal energy upon leaving the semiconductor crystal, as shown in Figure 5-3. As a result, in the following calculations, all quantities inside the silicon emitter are defined in terms of the internal normal energy $W$ and effective masses $m_c$, while all tunneling calculations are performed instead with the normal energy in vacuum $W'(W, E_t)$ and free electron mass $m_0$. In the case of the valence band, in which all energies and momenta are measured downward from
Figure 5-3: Due to the change in the electron mass from $m_e$ to $m_0$ upon leaving the semiconductor, electrons in the conduction band of silicon acquire additional normal energy in vacuum $W'$ as compared to the normal energy inside the emitter $W$. For electrons in the valence band, the additional normal energy acquired from the change in electron mass shifts the normal energy in vacuum $\bar{W}'$ downward and away from the top of the tunneling barrier.

the valence band edge in the bulk and are denoted by a bar above the variable, the change in the normal energy upon leaving the semiconductor is also given by Equation 5.4, but results in a normal energy in vacuum $\bar{W}'$ that is further from the top of the tunneling barrier than for the internal normal energy.

Modification of Schottky-Nordheim Barrier

As discussed in §2.2.3, the Schottky-Nordheim barrier consists of an exact triangular barrier with an image charge correction and is the tunneling barrier shape that is employed in standard FN theory [90,101]. Due to the finite relative dielectric constant of silicon $\epsilon_r = 11.7$, the form of the image charge term in the Schottky-Nordheim barrier must be modified as in Equation 2.16, yielding the new form of the potential.
energy outside the emitter (referenced to the conduction band edge)

\[ V(z) = \chi_e - eFz - \left(\frac{\epsilon_r - 1}{\epsilon_r + 1}\right) \frac{e^2}{16\pi\epsilon_0 z}, \]

(5.5)

where \( \epsilon_0 \) is the permittivity of free space and \( z \) is the distance traveled along the emission direction [98, 112]. The effect of this inclusion is to slightly lower the maximum barrier height as compared with the metal emitter case.

**Relative Barrier Height Lowering**

Due to the finite dielectric constant of semiconductors, the electric field penetrates into the semiconductor and leads to a downward bending of the conduction and valence bands near the emitter surface. Along with the energy bands, the vacuum level \( E_0 \) also bends downward to keep \( \chi_e \) constant everywhere in the semiconductor and leads to a reduction in the tunneling barrier height for all normal energies referenced to an energy level that is independent of the magnitude of the band bending, such as a band edge energy in the bulk. This effect is illustrated in Figure 5-4, in which the height of the tunneling barrier is shown for electrons in the conduction band with normal energy \( W = E_c(z \to -\infty) \) and for electrons in the valence band with normal energy \( W = E_v(z \to -\infty) \). In the diagram, the field penetration causes both the conduction and valence bands to bend downward by an amount equal to \( \Delta E_c = E_c(z = 0) - E_c(z \to -\infty) < 0 \) at the surface, pulling down the top of the barrier by the same amount relative to the case in which no field penetration occurs. As a result, for a given normal energy \( W \) in the conduction band, the new barrier height at the semiconductor surface \( H_{CB} \) is given by

\[ H_{CB}(F, W) = \chi_e + \Delta E_c(F) - W. \]

(5.6)
Figure 5-4: Energy band diagram showing the relative lowering of the barrier height for electrons in the conduction band with normal energy $W$ and electrons in the valence bands with normal energy $\bar{W}$, due to a field-penetration-induced downward bending of the bands $\Delta E_c$. The barrier height for conduction band electrons $H_{CB}$ and for valence band electrons $H_{VB}$ is lowered by $\Delta E_c$, relative to the case in which field penetration does not occur. The effect on the normal energy in vacuum due to the change in the electron mass upon leaving the semiconductor is omitted here for clarity.

For an electron in one of the valence bands with normal energy $\bar{W}$, the barrier height at the surface $H_{VB}$ is given by

$$H_{VB}(F, \bar{W}) = \chi_e + E_g + \Delta E_c(F) + \bar{W}.$$  \hfill (5.7)

In effect, this relative lowering of the tunneling barrier in vacuum provides a shorter tunneling path and higher transmission probability for conduction and valence band electrons of a given normal energy, as compared to an equivalent metal emitter or field emission models which omit the effects of band bending entirely.

**Appearance of an Internal Tunneling Barrier in the Valence Bands**

Although the downward band bending caused by field penetration into the semiconductor lowers the height of the barrier in vacuum, it also introduces an internal
tunneling barrier for electrons in the valence bands with energies near the band maximum. When the bands bend downward near the surface, the energies of the available states in each band also decrease, leading to a decrease in the maximum energy of available valence band states as a function of decreasing distance from the surface. As a result, an electron with normal energy equal to the valence band edge energy in the bulk will at some point encounter a region where there are no available or allowed states at that energy and the electron must either acquire enough normal energy on the way to the surface in order to continue to occupy available states of ever-increasing energy or it must tunnel through the band gap to reach the surface, as shown in Figure 5-5. Electrons tunneling through the band gap from the valence band to another set of electronic states is an experimentally-observed phenomenon and is the basis of a class of devices called tunnel field-effect transistors (TFET) [143], which are actively under research at the present time.

For the purposes of modeling tunneling through the band gap, the valence band edge $E_v(z)$ is treated as a tunneling barrier inside the semiconductor, and the tun-
neling probability through the band gap is evaluated within the JWKB framework to be

$$D_{int}(\overline{W}, F) = \begin{cases} 
\exp\left[-g_e \int_{z_t}^{0} [\overline{W} - E_v(z)]^{1/2} \, dz\right] & \text{for } 0 > \overline{W} > |\Delta E_c(F)| \\
1, & \text{for } \overline{W} \geq |\Delta E_c(F)|
\end{cases} \quad (5.8)$$

where $g_e = 2(2m_e)^{1/2}/\hbar$, $z_t$ denotes the turning point in the semiconductor at which $\overline{W} = E_v(z_t)$, $z = 0$ denotes the location of the semiconductor surface, and $\overline{W}$ is measured downward from the valence band edge in the bulk. When $\overline{W}$ is greater than or equal to the magnitude of the band bending at the surface, such that $\overline{W} \geq |\Delta E_c(F)|$, the internal tunneling barrier vanishes and the internal tunneling probability is thereafter unity. Naturally, the incorporation of this internal tunneling barrier into models of field emission leads to a greatly reduced tunneling probability for electrons near the top of the valence bands.

**Formation of Bound States Near the Emitting Surface**

Also a result of the downward band bending near the surface due to field penetration is the formation of a two-dimensional electron gas near the emitting surface. Since the conduction band bending downwards corresponds to a downward shift of available energy states near the surface, new, lower-energy states become available to electrons. Unlike the states above the conduction band edge in the bulk, however, these energy states do not form a continuum of momenta along the emission direction. Due to the spatial extent of the electronic wave functions of these states being confined to the nanometers-wide region between the edge of the conduction band and the surface of the emitter, discrete momenta states form along $z$, corresponding to the discrete normal energies $W_n$, as shown in Figure 5-6. As the normal momenta $p_z$ are discretized, but the momentum components transverse the emitting surface $p_x$ and $p_y$ still form a continuum, the electron system near the surface is two-dimensional,
Figure 5-6: Energy band diagram of the conduction band edge $E_c(z)$ and the discrete normal energy states that form near the surface $W_n$ as a result of electric field penetration. For the purposes of calculating the bound state energies $W_n$, the potential below the conduction band edge is taken to be infinite, while above the conduction band edge it is zero.

giving the total energy $E_n$ of each state as

$$E_n(p_x,p_y) = W_n + \frac{p_x^2}{2m_t} + \frac{p_y^2}{2m_t}.$$  \hspace{1cm} (5.9)

The discrete normal energies of electrons confined near the surface $W_n$ are calculated by solving for the wave functions $\Psi_n(z)$ and corresponding energies of the Schrödinger equation, assuming that the electron is confined to a quantum well with a well bottom shape defined by the conduction band edge as a function of $z$, $E_c(z)$:

$$-\frac{\hbar^2}{2m_n} \frac{d^2}{dz^2} \Psi_n(z) + V(z) \Psi_n(z) = W_n \Psi_n(z),$$  \hspace{1cm} (5.10)

where

$$V(z) = \begin{cases} 
0 & \text{for } 0 > z > z_t(W) \\
\infty & \text{for } z \leq z_t(W) \text{ and } z \geq 0, 
\end{cases}$$  \hspace{1cm} (5.11)

where $z_t$ is the location at which $E_c(z_t) = W$. The potential profile $\Phi(z)$ (negative of the conduction band edge profile $E_c(z)$) from which the potential energy in the
Schrödinger equation is derived, however, is defined by the density of free charge $\rho(z)$ in the domain over which it is solved, via Poisson’s equation:

$$\frac{d^2}{dz^2} \Phi(z) = -\frac{\rho(z)}{\epsilon}, \quad (5.12)$$

where $\epsilon$ is the permittivity of the region in which a solution is sought. The free charge density $\rho(z)$ of a two-dimensional electron gas is dependent upon the wave functions of the electrons that comprise it, via [120]

$$\rho(z) = -e \frac{m_e k_B T}{\pi \hbar^2} \sum_n \ln \left[ 1 + \exp \left( \frac{E_F - E_n}{k_B T} \right) \right] |\Psi_n(z)|^2. \quad (5.13)$$

From these expressions, it is clear that Poisson’s equation and the Schrödinger equation are coupled in this region and, for best results, a self-consistent solution for the conduction band edge profile $E_c(z)$ and bound state energies $W_n$ should be found.

### 5.2.2 Energetic Regimes of Emission from Bulk Silicon

As shown in Figure 5-7, the model for field emission from silicon consists of three energetic regimes, classified by the normal energy $W$ carried by the electron: i) emission from conduction band extended states for which $W \geq E_{cb}$, ii) emission from conduction band bound states for which $E_{cb} > W > E_{cs}$, and iii) emission from valence band extended states for which $E_{vb} > W$, where the subscript $b$ indicates the band edge energy in the bulk at $z \to -\infty$ and the subscript $s$ denotes the band edge energy at the emitting surface $z = 0$. Electrons in extended states in the conduction band and valence bands have wave functions that extend throughout the crystal and behave like free electrons, but are characterized by different effective masses. Electrons in bound states in the conduction band are spatially confined to a region near the emitting surface and may only have normal energies below $E_{cb}$, the conduction band edge in
Figure 5-7: Energy band diagram depicting the different energetic regimes of field emission from silicon in terms of the normal energy $W$ of electrons. For normal energies $W \geq E_{cb}$, the ECD originates from extended states in the conduction band; for $E_{cb} > W > E_{cs}$, the ECD originates from bound states in the conduction band with energies $W_n$; and for $E_{vb} > W$, emission originates from the valence band.

The total emitted current density from the silicon emitter $J$ is given by

$$J(F) = J_{ce}(F) + J_{cb}(F) + J_{ve}(F), \quad (5.14)$$

where $J_{ce}$ is the ECD from conduction band extended states and consists of contributions from all six conduction valleys; $J_{cb}$ is the ECD from the conduction band bound states, also consisting of contributions from all six conduction valleys; and $J_{ve}$ is the ECD from valence band extended states and consists of contributions from the heavy hole, light hole, and split-off bands.

### 5.2.3 Emitted Current Density Equations

Emitted current density equations for each of the energetic regimes discussed in §5.2.2 are derived in the sections below. As the model for field emission from silicon is meant to be implemented numerically, closed form expressions for the ECD are not sought. While the expressions for the ECD from extended states in the conduction band and valence band are closely related to the work of Christov [112], the expression for the
ECD from bound states in the conduction band is rooted in work by Rana [120]. In all cases below, the form of the tunneling barrier is given by Equation 5.5. As an additional note, all energies are referenced to a level that is independent of the applied field, such as the Fermi level or one of the band edges in the emitter bulk.

**Conduction Band Extended States**

The expression for the total energy distribution (TED) of electrons emitted from conduction band extended states of a single conduction valley is given by [112]

\[
j_{ce}(F, E) = e\frac{4\pi m_t}{h^3} f_{FD}(E) \int_{E_{cb}}^{E} D[F, W'(E, W')] dW, \tag{5.15}
\]

where \( m_t \) is the transverse effective mass of electrons in that particular conduction valley, \( h \) is the Planck constant, \( f_{FD} \) is the Fermi-Dirac distribution function, \( E_{cb} \) denotes the energy of the conduction band edge in the bulk, \( D(F, W') \) is the transmission probability through the surface barrier, given by

\[
D(F, W'(E, W)) = \exp \left[ -g_e \int_0^{zt} [V(z) - |\Delta E_c| - W'(E, W')]^{1/2} d\zeta \right], \tag{5.16}
\]

where \( g_e = 2(2m_0)^{1/2}/\hbar \), \( |\Delta E_c| \) is the magnitude of the downward band bending due to field penetration, \( W' \) is the normal energy in vacuum, given by

\[
W'(E, W) = \left(1 - \frac{m_t}{m_0}\right) E + \frac{m_t}{m_0} W, \tag{5.17}
\]

and \( E \) is the total energy of the electron. The ECD is obtained by integrating the TED in Equation 5.15 over all relevant total energies, yielding

\[
J_{ce}(F) = \int_{E_{cb}}^{E_{max}} j_{ce}(F, E) dE, \tag{5.18}
\]

169
where $E_{\text{max}}$ is the maximum total energy for the calculation. The total ECD from extended states in the conduction band consists of contributions from each of the conduction valleys, such that

$$J_{ce}(F) = 2J_{ce}^{\text{on}}(F) + 4J_{ce}^{\text{off}}(F), \quad (5.19)$$

where the superscripts “on” and “off” denote contributions from the on-axis and off-axis conduction valleys.

**Conduction Band Bound States**

The total energy distribution of electrons from the bound states in the conduction band near the surface consists of a sum of the TEDs of each of the bound states, labeled by $n$

$$j_{cb}(F, E) = e\frac{4\pi m_t}{\hbar^2} f_{FD}(E, T) \sum_n \nu(W_n) D[F, W'_n(E, W_n)], \quad (5.20)$$

where $\nu(W_n)$ is the tunneling attempt frequency of electrons in state $n$, given by [120]

$$\nu(W_n) = \left\{ \left. \int_{z_{t,n}}^0 \left[ \frac{2m_n}{W_n - E_c(z)} \right]^{1/2} \, dz \right\}^{-1}, \quad (5.21)$$

where $z_{t,n} < 0$ is the turning point of the electron inside the semiconductor with normal energy $W_n$, $z = 0$ denotes the emitter surface, $m_n$ is the normal effective mass of the electron, $E_c(z)$ is the conduction band edge as a function of $z$, and $D[F, W'_n]$ and $W'_n(E, W_n)$ are given by Equation 5.16 and Equation 5.17 respectively. The total ECD from the bound states in a single conduction valley is given by

$$J_{cb}(F) = e\frac{4\pi m_t}{\hbar^2} \sum_n \int_{W_n}^{E_{\text{max}}} dE \{ f_{FD}(E, T) \nu(W_n) D[F, W'_n(E, W_n)] \}, \quad (5.22)$$
and the total ECD from all conduction valleys in the conduction band is given by an expression identical to Equation 5.19.

**Valence Band Extended States**

In order to simplify the calculation of the ECD from extended states in the valence band, it is customary to first make a few variable substitutions. For convenience, the effective masses of electrons in the valence bands $m_e^*$ are replaced by the hole effective masses $m_v$, such that $m_e^* = -m_v$, and all energies are measured as positive, downward from the valence band edge in the emitter bulk $E_{vb}$, such that $E = -E$ and $W = -W$. Using these substitutions, the total energy distribution of electrons emitted from a single valence band is given by

$$j_{ve}(F, E) = e \frac{4\pi m_v}{h^3} f_{FD}(-E) \int_0^E D[F, -W'(E, W)] dW,$$  

(5.23)

where the transmission probability $D[F, -W'] \approx D_{int}[F, W] D_{ext}[F, -W']$, where $D_{int}$ is given by Equation 5.8, $D_{ext}$ is given by

$$D_{ext}(F, -W'(E, W)) = \exp \left[ -g_e \int_0^z \left[ V(z) - |\Delta E_c| + W'(E, W) \right]^{1/2} dz \right],$$  

(5.24)

and $W'$ is

$$W' = \left( 1 - \frac{m_v}{m_0} \right) E + \frac{m_v}{m_0} W.$$  

(5.25)

The total emitted current density from a single valence band is given by integrating the TED up to a maximum energy $E_{max}$

$$J_{ve}(F) = \int_0^{E_{max}} j_{ve}(F, E) dE,$$  

(5.26)
and the total ECD from valence band extended states results from summing the contributions from each of the valence bands, yielding

\[ J_{\text{ve}}(F) = J_{\text{ve}}^{\text{hh}}(F) + J_{\text{ve}}^{\text{lh}}(F) + J_{\text{ve}}^{\text{so}}(F), \tag{5.27} \]

where the superscripts “hh”, “lh”, and “so” denote contributions from the heavy hole, light hole, and split-off bands respectively.

### 5.2.4 Finite Electronic Flux Supply from Conduction Band States

As is often done in field emission calculations for simplicity, the effects on the potential and the electron density near the surface, due to current flowing both through and out of the semiconductor are omitted. In the case of this model, including a current continuity equation to treat these effects would require a simultaneous solution of the Schrödinger equation, Poisson’s equation, and current continuity equations, which would significantly add to the complexity of simulations. Consequently, it is possible in such models, that the predicted emitted current density from the emitter can exceed the supply of electrons to the surface from the bulk. While this is not an issue for emission from the valence bands due to the extremely high electron densities, it can occur in the conduction band. In the interest of capturing the most basic effects of a limited conduction band electronic flux density supply to the emitter surface, the ECD from conduction band states in the model is calculated via an algorithm which limits the ECD in the case that it would exceed the remaining flux density supply.

A diagram of the algorithm used to calculate the conduction band ECD is shown in Figure 5-8. The algorithm starts with the calculation of the total incident electron
Figure 5-8: Flow chart detailing the algorithm that calculates the emitted current density from conduction band states. The ECD is calculated for each total energy $E$ between an $E_{min}$ and $E_{max}$ with a spacing of $dE$. At each energy $E$, the ECD from bound states $J_n$ is calculated first, from lowest $W_n$ to highest, followed by the ECD from extended states $J_{ce}$. If any component of the ECD exceeds the remaining flux density supply from the extended states $\Phi_r$, the ECD of that component is limited to $\Phi_r$ and the ECD from all higher energy states is zero.

Flux density from the conduction band extended states $\Phi_b$, which is equal to

$$\Phi_b(T) = \frac{4\pi m_t}{\hbar^3} \int_{E_{cb}}^{\infty} f_{FD}(E, T)(E - E_{cb})dE,$$  \hspace{1cm} (5.28)$$

and initializes the remaining flux density supply to $\Phi_r = \Phi_b$. The ECD is calculated as a function of total energy $E$, starting from the minimum-energy state $E_{min} = W_1$ (lowest-energy bound state) and proceeding incrementally upwards by energy steps of size $dE$ until reaching some maximum energy $E_{max}$, after which the ECD is integrated across all total energies to obtain the total ECD. At each energy $E$, the algorithm identifies all of the bound states with normal energies $W_n < E$ and one-by-one, in order of ascending normal energy, the contribution to the ECD from each bound state $J_n$ is determined. After calculating the ECD contributed by a bound state of total energy $E$, the algorithm checks whether the ECD from that state exceeds the remaining flux density supply from the extended states $\Phi_r$. If $J_n < \Phi_r$, the full
ECD is emitted and the new remaining flux density supply becomes $\Phi_r = \Phi_r - J_n$. If $J_n > \Phi_r$, the flux density supply from the extended states has been depleted, so $J_n = \Phi_r$, $\Phi_r = 0$, and all ECD contributions from higher energy bound and extended states is set to zero: no more electrons can be emitted from the conduction band. The procedure of calculating the ECD from only the bound states continues until $E = E_{cb}$, which signals the onset of emission from the extended states in the bulk. In this regime, the bound state ECDs are calculated as described above, then the extended state ECD $J_{ce}$ for that particular total energy $E$ is calculated afterwards. Identical to the case of bound states, $J_{ce}$ is checked against the remaining flux density supply $\Phi_r$ to determine if the prescribed ECD can actually be emitted: if $J_{ce} < \Phi_r$, the full ECD is emitted, $\Phi_r = \Phi_r - J_{ce}$, and the calculation of the ECD for the next energy commences; if $J_{ce} > \Phi_r$, the flux density supply has been depleted, $J_{ce} = \Phi_r$, $\Phi_r = 0$, and all ECD contributions from higher energy bound and extended states is set to zero.

In order to implement such an algorithm, priorities for the distribution of the flux density supply among the emitting states needed to be established. The decision to distribute the supplied flux density first to the bound states, from lowest normal energy to highest, arises from the tendency of bound state electrons to be emitted more rapidly than extended state electrons. This results from bound states near the surface being characterized by a higher flux density of electrons attempting to tunnel, which decreases with increasing bound state normal energy $W_n$, hence the order of supplied flux density distribution among the bound states. While extended states in the conduction band have a higher transmission probability through the barrier owing to their higher normal energy, their tunneling attempt rate is significantly lower than for bound states, and thus, electron emission and the need to replace emitted electrons happens on a much slower timescale.
5.2.5 Approximations and Assumptions

In constructing the model for field emission from bulk semiconductors, several assumptions and approximations were made. While these approximations are briefly listed below, their possible effects on the accuracy of the model’s predictions are discussed in detail in §5.4.

- **Omission of Surface States:** Emission from the band of states at the surface of the semiconductor, located in the forbidden energy gap, has been ignored in the model. Although included in some models, the physical properties on which these states depend are not known a priori and must be treated as adjustable parameters within the model. As the current iteration of this model aims to provide general results that apply to all bulk silicon emitters, the details of emission from surface states have been omitted.

- **Zero Emitted Current Approximation:** The zero emitted current approximation assumes that the current flowing due to field emission does not alter the electron and hole distributions in the silicon from their equilibrium values. As a result, the potential inside the silicon and bound state energies can be determined by solving the coupled Poisson’s and Schrödinger equations using an equilibrium distribution of holes and electrons inside the emitter. This approximation is valid at relatively low emitted currents, but its accuracy decreases as the emitted current increases.

- **Zero Internal Current Approximation:** The emitted current density is also calculated under the assumption that there is no resistive voltage drop inside the semiconductor due to current flow. Realistically, current flow through the emitter due to field emission leads to downward band bending and a non-constant Fermi level along the entire length of the emitter. Under most circumstances, the voltage drop across the emitter amounts to only a few volts and is negligible.
compared to the anode voltage, which is typically dozens to thousands of volts. In such cases, the voltage drop has very little effect on the magnitude of the surface electric field and leaves the magnitude of the ECD unchanged, but shifts the entire total energy distribution down in energy equal to the internal voltage drop, relative to the Fermi energy in the metal contact [8].

- Crystal Structure of Semiconductor Bulk-like at Surface: Due to surface reconstruction, the crystal structure of a semiconductor near the surface differs from that of the bulk. As a result, the density of states at the semiconductor surface from which emission occurs can differ greatly from the density of states in the bulk. Since the reconstructed semiconductor surface and surface density of states either require experimental investigation or additional computational models, they are omitted from this model. The bulk density of states and effective masses are used in emission calculations instead.

5.3 Results and Analysis

In this section, the emitted current density, total energy distribution of emitted electrons, and Fowler-Nordheim plots of both n-type and p-type silicon emitters of varying dopant densities are presented and analyzed. As shown in Figure 5-9, the field emission simulation comprised two steps. The first step consisted of solving Poisson’s equation and the Schrödinger equation within the semiconductor in order to obtain the band edge profiles and the bound state energies near the surface. This step was performed in the nanoelectronic device simulator nextnano3 [144] by specifying the dopant type and dopant density of the silicon emitter, finding a range of surface potentials that produced electric fields at the surface typical for field emission, and performing a “voltage sweep” through the range of surface potentials. For each surface potential value, the band edge profiles, bound state energies near the surface, and
Figure 5-9: The simulation of field emission from bulk silicon was consisted of two steps. In the first step, *nextnano3* was used to find a self-consistent solution to Poisson’s equation and the Schrödinger equation across a range of surface electric fields, providing the electronic properties of the emitter. In the second step, the data provided by nextnano was used to calculate the ECD and TED contributions from each of the energetic emission regimes in the model. Equivalent surface field magnitude in vacuum were stored. In the second step of the simulation, the data supplied by nextnano3 was used to calculate the emitted current density and total energy distribution of emitted electrons from each of the energetic emission regimes of the model. The field emission calculations themselves were performed in python using the NumPy and SciPy packages, while the plots were created with Matplotlib [137–140]. Both n-type and p-type silicon emitters with dopant densities ranging between $N = 10^{15} \text{ cm}^{-3}$ and $N = 10^{18} \text{ cm}^{-3}$ were investigated, across a range of applied fields (in vacuum) between $F = 0.5 \text{ V/nm}$ and $F = 10 \text{ V/nm}$. The electron affinity of silicon was taken to be $\chi_e = 4.04 \text{ eV}$, the magnitude of the forbidden energy gap is calculated by nextnano in each case, but is approximately $E_g \approx 1.12 \text{ eV}$, the effective masses of electrons and holes are given in Table 5.1, and all calculations were performed for $T = 300 \text{ K}$.
Figure 5-10: Plots of the (a) conduction band and three valence band edges and (b) lowest 10 bound state energies $W_n$ in each type of conduction valley of the conduction band for n-type silicon with $N_D = 10^{15}$ cm$^{-3}$. Plots (c) and (d) display the same results for p-type silicon with $N_A = 10^{15}$ cm$^{-3}$. The energy reference level is the Fermi level.

5.3.1 Band Edge Profiles and Bound State Energies

Results

Figure 5-10a shows a plot of the conduction band minimum and valence band maxima as a function of distance into the semiconductor $z$ for n-type silicon with $N_D = 10^{15}$ cm$^{-3}$ and a vacuum-side applied field of $F = 3$ V/nm. Compared with the values in the bulk, the bands are bent downward by approximately $\Delta E_c \approx -0.5$ eV at the silicon surface $z = 0$ and form an accumulation layer (range of widths across all dopant densities $\approx 1 - 3$ nm), with significant electric fields occurring for $z < 5$ nm, but gradually beginning to approach a constant value at $z \approx 10$ nm, after which the electric field inside the semiconductor becomes negligible, signaling the onset of the
quasi-neutral bulk region. Figure 5-10b plots the lowest 10 bound state energies in the accumulation layer of the on-axis and off-axis conduction valleys of the conduction band at the surface indexed by \( n \), which form due to the downward band bending \((F = 3 \text{ V/nm})\) confining electrons near the surface. The energy spacing between bound state energies in both types of conduction valleys is approximately \( \Delta E \approx 0.1 \text{ eV} \) for the lowest-energy states and decreases rapidly with increasing \( n \), eventually falling below \( \Delta E = 0.01 \text{ eV} \) for the highest energies, at which point they form a quasi-continuum. At every index \( n \), the off-axis bound state energies are observed to be larger than the on-axis bound states, due to the off-axis states having a smaller normal effective mass \( m_{\text{off}}^{n} \approx 0.19m_{0} \) than the on-axis states \( m_{\text{on}}^{n} \approx 0.92 \).

Figure 5-10c and Figure 5-10d show the band edge energies and bound state energies under the same electrostatic conditions as in Figures 5-10a-5-10b for a p-type silicon emitter of dopant density \( N_{A} = 10^{15} \text{ cm}^{-3} \). As can be seen, the extent of the band bending due to field penetration in Figure 5-10c is significantly larger than in the n-type case, with the bands bending \( |\Delta E_{c}| \approx 1.3 \text{ eV} \) over a distance of approximately 1000 nm. An inversion layer forms at the surface of the p-type silicon, whose location is indicated by the steep slope near \( z \approx 0 \) in the plot, followed by a depletion layer for \( z < 1000 \text{ nm} \), indicated by the region with a lower, but non-zero slope in the band edges. The quasi-neutral bulk region in the p-type bulk, where the electric field is approximately zero, is located at \( z > 1000 \text{ nm} \). The bound state energies in the inversion layer at the surface, as shown in Figure 5-10d, are almost identical in magnitude to the energies in the accumulation layer of the n-type silicon sample and show the same trends with increasing state index \( n \). This correlation is explained by the shape of the inversion layer at the surface of the p-type silicon being nearly identical to the shape of the accumulation layer at the surface of the n-type silicon, as shown in the inset of Figure 5-10c. Above the top of the inversion layer in the depletion region, the size of the quantum well to which electrons near
Figure 5-11: Plots of the lowest bound state energy at the surface $W_1$, as measured from the conduction band edge in the bulk, vs. the applied electric field $F$, for both the on-axis (solid lines) and off-axis (dashed lines) conduction valleys of the conduction band in (a) n-type silicon and (b) p-type silicon.

The surface are confined greatly increases ($L \gg 10 \text{ nm}$) and the bound state energies form a quasi-continuum.

In Figures 5-11a and 5-11b, the lowest bound state energy $W_1$, as measured from the conduction band edge energy in the bulk, is shown as a function of the applied field $F$ and impurity dopant density for n-type and p-type silicon. The energies are referenced to the conduction band edge in the bulk due to its location being independent of both the applied field and dopant density, which allows for a clearer investigation into the changes in the lowest bound state energy across all dopant densities. The Fermi level, which is the usual zero of energy for these calculations, varies with the dopant density and prevents such a comparison. As can be seen, the lowest bound state energy $W_1$ (and by extension, all higher-energy bound state energies) of the on-axis and off-axis conduction valleys decrease significantly as a function of increasing field in both n-type and p-type silicon emitters.

**Analysis**

The increase in the lowest bound state energy in n-type silicon and the decrease in the lowest bound state energy in p-type silicon as a function of increasing impurity dopant density results from the change in the shape of the conduction band edge
near the surface, which forms the quantum well. Referenced to the conduction band edge in the bulk, the change in dopant density effects a change in the depth and width of the accumulation layer in n-type silicon, as well as the inversion layer in p-type silicon, as shown in Figure 5-12. As the dopant density in n-type emitters increases, the electric field at the surface is able to be screened more effectively by the higher density of electrons available near the surface, resulting in a reduction in the depth and width of the accumulation layer, as shown in Figure 5-12a. The reduced depth and spatial extent of the quantum well formed by the conduction band edge leads to a comparatively higher energy for the ground state and all other energy states. Increasing the dopant density in p-type silicon has the opposite effect on the conduction band edge profile. As the dopant density in p-type emitters increases, a lower density of electrons is available near the surface to screen the electric field and the depth and extent of the inversion layer at the surface increases, as shown in Figure 5-12b. Due to the spatially wider and energetically deeper quantum well in the inversion layer, the electrons are confined to a lesser extent and the lowest ground state energy (as well as all other energies) decreases.
5.3.2 Emitted Current Density vs. Applied Field

Results

Figure 5-13 shows the total emitted current density $J$ as a function of the applied field in vacuum $F$ for various dopant densities of n-type and p-type silicon emitters. At low-to-intermediate field strengths of approximately $F < 5 \text{ V/nm}$, n-type emitters across all considered dopant densities emit approximately the same ECD before the slopes of the curves of lower dopant density emitters begin to flatten and emit relatively lower current. For p-type emitters at fields below $F \approx 1.25 \text{ V/nm}$, the ECD is also approximately the same across all dopant densities. Within the range of applied fields $1.25 \text{ V/nm} < F < 2 \text{ V/nm}$, each of the curves exhibits a flat region where the ECD increases only slightly despite increases in $F$, with the onset of this flat region occurring at lower fields for higher dopant densities. For fields above $F \approx 2 \text{ V/nm}$, the curves corresponding to the different dopant densities converge again to a single profile and the ECD rises at a rate faster than that for n-type emitters, especially for the range of fields $F > 5 \text{ V/nm}$. As an additional note, the typical range of...
Figure 5-14: The conduction band extended state (blue circles), conduction band bound state (blue triangles), and valence band extended state (red circles) emitted current densities as a function of applied field $F$, for (a) an n-type emitter of dopant density $N_D = 10^{15} \text{ cm}^{-3}$ and (b) a p-type emitter of dopant density $N_A = 10^{15} \text{ cm}^{-3}$.

applied fields $F$ within which field emission is typically experimentally investigated and observed is $1 \text{ V/nm} < F < 6 \text{ V/nm}$.

Analysis

The appearance of a region of lower slope at high fields for n-type emitters and a region of no slope at low fields for p-type emitters in the total ECD plots is a result of the saturation of the conduction band ECD due to a depletion of the electron flux density supplied to the surface from the conduction band extended states in the bulk. Figures 5-14a and 5-14b show that the conduction band bound state ECD saturates to a constant value and the extended state ECD vanishes completely when the flux density supply is depleted, as per the prioritization of the distribution of the electronic flux supply detailed in §5.2.4. In n-type emitters, the conduction band current density saturates at fields above $F = 5 \text{ V/nm}$, leading to the corresponding decrease in the slope of the total ECD $J$ vs. applied field $F$ plot in Figure 5-13. The total ECD from n-type emitters does not completely saturate when the conduction band ECD saturates, due to the comparable contribution from the valence band at high fields. The conduction band ECD in p-type emitters saturates at much lower applied fields,
within which it is still the dominant contributor to the total ECD, which results in the saturation of the total ECD and appears as the flat region in the total ECD plot between approximately 1.25 V/nm and 2 V/nm. Once the valence band contribution to the ECD reaches the level to which the conduction band ECD has saturated, the total ECD begins to rise again with increasing applied field.

5.3.3 Total Energy Distribution

Results

The total energy distribution of electrons from an n-type silicon emitter of dopant density $N_D = 10^{15} \text{ cm}^{-3}$ and p-type silicon emitter of dopant density $N_A = 10^{15} \text{ cm}^{-3}$ for three different applied fields $F$ is shown in Figure 5-15. While the TED of the n-type emitter in Figure 5-15a has a maximum in the conduction band for lower applied fields of approximately $F < 5 \text{ V/nm}$, the peak is located in the valence band for higher applied fields. Correspondingly, the relative magnitude of the valence band TED peaks as compared to the conduction band TED peaks grows as a function of increasing field. Figure 5-15b shows a close-up of the conduction band TED for the n-type emitter, which reveals a TED that initially increases with increasing energy, then turns downward to decrease with increasing energy. At approximately $E - E_F = 0.25 - 0.5 \text{ eV}$, the TED abruptly increases to its maximum value and decays with increasing energy thereafter. For the high applied field result, the TED increases slightly with increasing energy $E$, but drops to zero at approximately $E - E_F = -0.9 \text{ eV}$. Additionally, the energy at which each of the conduction band TEDs first assume non-zero values decreases as a function of increasing applied field, as does the location of the TED peaks in the valence band. For p-type emitters, the TED maxima occurs in the valence band across all applied fields considered, as shown in Figure 5-15c. While for the lowest applied field $F = 2.3 \text{ V/nm}$, there exists a visible TED peak in the conduction band at approximately $E - E_F = 0 \text{ eV}$, no TED components are
Figure 5-15: The normalized total energy distribution of emitted electrons from (a) an n-type silicon emitter of dopant density $N_D = 10^{15} \text{ cm}^{-3}$ and (c) a p-type silicon emitter of dopant density $N_A = 10^{15} \text{ cm}^{-3}$ for three different applied fields $F$. Plot (b) shows a close-up of the conduction band TED of the n-type emitter, while plot (d) shows a close-up of the valence band TED of the p-type emitter. In plots (a) and (b), TED values of zero were not plotted to improve the clarity of the plots.

visible in the conduction band at higher fields. In the valence band, the general shape of the TED peaks remains the same across different applied fields, but the location of the peak shifts to lower energies and the width of the peak increases with increasing applied field. Changing the dopant density of an emitter for a fixed applied field resulted in only minor variations in the TED. Although not shown here, for n-type emitters, increasing the dopant density led to relatively higher peaks in the conduction band and lower peaks in the valence band; for p-type emitters, increasing the dopant density had the opposite effect.
Figure 5-16: (a) Plot of the heavy hole valence band maximum energy (solid lines) and the potential barrier in vacuum (dashed lines) for various applied fields $F$ as a function of $z$, where $z = 0$ denotes the location of the surface. (b) For the same energies, the normalized transmission probability through the downward-bent valence band and external potential barrier for various applied fields $F$, as shown in (a).

Analysis

The shape of the TED and how it changes as a function of applied field $F$ results from the flux supply limitations in the conduction band and the transmission probability of electrons from the valence band. As a function of increasing applied field, the conduction band TED peaks decrease due to the saturation of the conduction band ECD, which is caused by the limited flux density that is able to be supplied to the surface from the bulk. Conversely, the electron supply in the valence band is so large that the valence band ECD does not saturate for fields relevant to field emission and the relative magnitude of the valence band TED peaks continues to increase. The shape of the valence band TED is a direct consequence of the shape of the normalized transmission probability of electrons from the valence band, as a function of normal energy. As shown in Figure 5-16a, electrons that are emitted from the valence band may have to tunnel through two tunneling barriers: i) the for-
bidden energy gap, which lies above the valence band maximum energy, and ii) the usual potential barrier in vacuum. Starting from the valence band edge, the electronic transmission probability through the internal tunneling barrier increases as a function of decreasing normal energy, as the thickness of the downward-bent forbidden energy gap decreases. When the normal energy reaches the same energy as the valence band edge at the emitter surface \( z = 0 \), the transmission probability through the internal tunneling barrier reaches unity. The transmission probability for electrons through the potential barrier in vacuum, however, decreases as a function of decreasing energy, due to the increasing barrier width. Since the total transmission probability is approximately the product of the monotonically-increasing transmission probability through the internal tunneling barrier and the monotonically-decreasing transmission probability through the potential barrier in vacuum, the total transmission probability has a peaked shape that decays on either side, as shown in Figure 5-16b. The location of the peak in energy shifts downward with increasing applied field as a result of the valence band edge energy at the surface decreasing due to an increase in the amount of band bending from field penetration. The width of the peak increases with increasing applied field due to the thinning of the potential barrier in vacuum, which results in an overall increase in tunneling probability across all energies. For very high applied fields of approximately \( F > 7 \text{ V/nm} \), it may occur that the peak of the transmission probability shifts upwards in energy, as the thinning of the potential barrier in vacuum with increasing \( F \) increases the transmission probability for energies above the valence band edge at the surface more rapidly than the widening and deepening of the internal tunneling barrier decreases it.
Figure 5-17: Fowler-Nordheim plots for n-type and p-type emitters of varying dopant density. Although the FN plot corresponding to an n-type emitter with $N_D = 10^{15}$ cm$^{-3}$ is correctly located on the x-axis, all other FN plots were shifted positively along the x-axis for easier comparison across different dopant densities, as they would otherwise be located within the same range of x values.

5.3.4 Fowler-Nordheim Plots

Results

The Fowler-Nordheim (FN) plots for n-type and p-type silicon emitters of various dopant densities are shown in Figure 5-17. As can be seen, the FN plots corresponding to n-type emitters are linear as anticipated by FN theory, except for a small region near the maxima of the curves, where the curves have an apparently lower slope. No significant differences between the FN plots from n-type emitters of differing dopant densities are readily identifiable. FN plots for p-type emitters display three distinct regions: i) a region of lower slope at lower applied fields $F$ (larger $1/F$ in the FN plots); ii) a region with a slope of approximately zero for intermediate fields; and iii) a region of higher slope at higher applied fields. The zero-slope region is located within approximately the same range of applied fields regardless of p-type dopant density, but occurs at lower emitted current densities for p-type emitters with higher dopant
Figure 5-18: Fowler-Nordheim plots for (a) an n-type silicon emitter of dopant density $N_D = 10^{15}$ cm$^{-3}$ and (b) a p-type silicon emitter of dopant density $N_A = 10^{15}$ cm$^{-3}$. The slopes of each of the non-zero linear regions were extracted via linear regression.

densities.

Individual Fowler-Nordheim plots for an n-type and p-type silicon emitter and the barrier heights corresponding to the slopes extracted from each of the linear regions are shown in Figure 5-18. Although hardly visible in Figure 5-17, a close-up of the FN plot for an n-type emitter, as shown in Figure 5-18a, reveals that at high fields $F > 5$ V/nm, there are two distinct linear regions separated by a region of lower slope, as in the FN plots for p-type emitters. Extracted barrier heights for the n-type silicon emitter of dopant density $N_D = 10^{15}$ cm$^{-3}$ were $H_L = 3.87$ eV for the lower linear region and $H_U = 4.39$ eV for the upper linear region. The field at the boundary of the lower linear region and the transitional region is approximately $F = 5$ V/nm, while at the boundary of the upper linear region and the transitional region, the field is approximately $F = 8.5$ V/nm. The FN plot for a p-type emitter of dopant density $N_A = 10^{15}$ cm$^{-3}$ is shown in Figure 5-18b. The extracted barrier height for the lower region is $H_L = 3.9$ eV and the extracted barrier height for the upper region is $H_U = 5.08$ eV. The field at the boundary between the lower linear region and the transitional region is approximately $F = 1.4$ V/nm and the field at the boundary between the upper linear region and the transitional region is approximately $F = 2.5$ V/nm. The procedure for extracting the barrier heights from
FN plots is briefly outlined in §2.2.2.

Fowler-Nordheim plots with two distinct linear regions separated by a transitional region, as described above, have been experimentally observed by Ding [10]. A comparison between these experimental results and the predictions of a model of field emission from cylindrical silicon nanowires is presented in §6.4.2.

Analysis

The two distinct linear regions in the Fowler-Nordheim plots arise due to different components of the ECD dominating the total ECD at different ranges of applied fields \( F \). The linear region of FN plots that coincides with lower applied fields corresponds to emission dominated by the contribution from the conduction band and is marked by a lower slope. The linear region at higher fields corresponds to emission dominated by the valence band and is characterized by a larger slope in FN plots. The intermediate region between the two linear regions, which shows no identifiable functional form in FN plots for n-type emitters, but is almost flat in FN plots of p-type emitters, corresponds to a transitional region between the conduction-band-dominated and the valence-band-dominated regimes of emission. The transitional region in p-type emitters is nearly flat because the saturation of the conduction band current occurs before the valence band current begins to contribute significantly, causing the total ECD to saturate. In n-type emitters, the valence band ECD already contributes a significant amount to the total ECD when the conduction band current density saturates, resulting in a transitional region in the FN plot with various finite slopes as the valence band ECD gradually begins to dominate emission. Further evidence supporting the identification of the upper and lower linear regions of FN plots with valence-band- and conduction-band-dominated emission regimes is also found in the extracted barrier heights of the upper and lower regions \( H_U \) and \( H_L \) in Figure 5-17, which indicate that electrons emitted at higher fields (upper region) experience a
larger barrier height than electrons emitted at lower fields (lower region).

Due to band bending, the extracted barrier heights of the upper and lower regions of both n-type and p-type emitters are lower than the expected barrier heights in the absence of band bending, which are $H_{CB} = \chi_e = 4.04 \text{ eV}$ and $H_{VB} = \chi_e + E_g = 5.16 \text{ eV}$ respectively. The extracted barrier heights from the lower region are slightly smaller than $\chi_e$ due to emission originating from bound states, which lie above the conduction band edge at the surface, the energy to which the electron affinity $\chi_e$ is referenced. For the n-type emitter in Figure 5-17a, the extracted upper region barrier height is significantly smaller than $\chi_e + E_g$, which is a result of the relative barrier lowering associated with the large amount of band bending ($\Delta E_c \approx 0.75 - 1 \text{ eV}$) that occurs at such high applied fields. Under these conditions, the barrier height seen by electrons with energies around the valence band maxima is lowered by $\Delta E_c$.

Although the extracted barrier heights from the lower regions of the n-type and p-type emitters of Figures 5-17a and 5-17b are nearly identical, the extracted barrier heights from their upper regions differ by 0.69 eV. This is due to the difference in the shape of the depletion and inversion regions in the p-type emitter and the shape of the accumulation layer in the n-type emitter. As seen in Figure 5-10, while the band bending in the n-type accumulation layer occurs very close to the surface (within $\approx 5 \text{ nm}$), the band bending in the p-type emitter occurs in two phases: i) the rapid change in potential in the inversion layer (within $\approx 5 \text{ nm}$) and the ii) gradual change in potential between the inversion layer and bulk (between approximately 5 nm and 1000 nm). Because all of the band bending in an n-type emitter occurs so close to the surface, the internal tunneling barrier for electrons with normal energies close to the valence band maxima is relatively narrow. Conversely, even though the magnitude of the band bending in the p-type emitter is larger than in the n-type emitter, it occurs more gradually, over a much longer distance and leads to an extremely wide internal tunneling barrier for energies within about one half of the magnitude of the band
bending below the valence band edges. Consequently, the total electron transmission probability is highly suppressed at these energies and the extracted barrier height is much larger for the p-type emitter than for the n-type emitter.

5.4 Limitations of the Model

In this section, the limitations of the model which result from the approximations and assumptions that were made during its derivation are addressed. Where relevant, the effects of reversing such assumptions or approximations on the predictions of the model are discussed. Further details on the limitations of the fundamental physical aspects of the silicon emitter model are discussed in the context of comparing the predictions of the cylindrical silicon nanowire emitter model to experimental data in §6.4.

5.4.1 Omission of Surface States

Surface states usually occur as a band of states in the forbidden energy gap, located at the semiconductor surface. Some models include emission from surface states, such as Modinos’ treatment of emission from semiconductor surfaces in §2.3.4 [8], but they depend upon adjustable parameters to account for the lack of a priori knowledge about the properties of the surface states and, consequently, add an additional element of uncertainty to the model.

Surface states may either be of the acceptor type, which lie above the Fermi level and are empty at $T = 0$ K, or of the donor type, which lie below the Fermi level and are occupied by an electron at $T = 0$ K. In the case that the band of surface states is of the acceptor type, as shown in Figure 5-19, the conduction and valence bands must bend upward near the surface in order to account for the charge introduced by the surface band, in the absence of an applied electric field. Applying an electric field
Figure 5-19: Energy band diagram of an n-type semiconductor emitter with an acceptor band of surface states (hatched area) at the emitter surface and a sample total energy distribution for emitted electrons. In the absence of an applied electric field, the upward bending of the bands is more dramatic. The peaks in the TED labeled as $j_c$, $j_s$, and $j_v$ denote the contributions from the conduction band, surface state band, and valence band [8].

to the surface of the semiconductor attracts electrons to the surface and populates the surface states, causing the bands to bend downward and eventually to form a quantum well and bound states near the surface. The initial presence of these surface states would increase the applied field necessary for the formation of bound states at the semiconductor surface, as compared with the current iteration of the model which omits them. As a result, any effects on the ECD due to the bound states in the conduction band exhausting the electronic flux density supply from the bulk would occur at higher applied fields. In the case that the surface band is of the donor type, the opposite would be true.

Due to their location in the forbidden energy gap where no bulk states occur, the contribution of the surface states to the total energy distribution of emitted electrons should be prominent, also shown in Figure 5-19. According to Modinos [8], the experimental TED from the (100) plane of an n-type germanium emitter is dominated by contributions from surface states, which are extremely sensitive to the cleanliness and orientation of the emitter surface. The TED peak corresponding to emission from surface states was observed to be approximately 2-3 times larger than that of the peak corresponding to emission from the valence band, while a peak corresponding to
conduction band contributions is not even measurable. As the applied field increases, the relative height of the surface state peak decreases and contribution of the surface band to the ECD becomes less significant. In the cases of emission from germanium discussed by Modinos, the total ECD is comprised nearly equally of contributions from the valence bands and surface band, even for n-type emitters. Although the relative contributions to the ECD from a surface band, the conduction band valleys, and the valence bands in a silicon emitter is unknown, it is clear that the inclusion of surface states into the model would introduce an important channel for emission that has been experimentally observed to be of physical importance to field emission from semiconductors. For more details on theoretical studies of emission from surface states, see Modinos [8].

5.4.2 Zero Emitted Current Approximation

The zero emitted current approximation neglects the effect of current flow through the emitter on the distribution of electrons and holes, leaving the electron and hole distributions unaltered from their forms in thermal equilibrium and the Fermi level constant throughout the emitter. For cases in which field penetration into the semiconductor is significant, however, the effect of current flow should not be ignored, as the number of electrons that are being extracted from the emitter surface via field emission alters the distribution of holes and electrons significantly compared to thermal equilibrium. The potential profile near the emitting surface is also altered from thermal equilibrium and, in turn, affects the bound state energies near the surface.

Calculations of the approximate ECD from a planar semiconductor surface that include the effect of current flow in the semiconductor have been performed by Baskin et al. [113]. Under the assumptions that there are no surface states and that the only significant contributor to the ECD is from states in the non-degenerate conduction band, their calculations show that the incorporation of current continuity equations
into the model predicts that the ECD from n-type silicon emitters \((N_D = 10^{12} - 10^{13} \text{ cm}^{-3})\) saturates at applied fields of approximately \(F = 2.5 \text{ V/nm}\), while the ECD from p-type germanium emitters \((N_A = 10^{15} \text{ cm}^{-3})\) saturates at applied fields of approximately \(F = 2.9 - 3.5 \text{ V/nm}\), depending on the mobility model used. They report that initially, as the applied field increases, the conduction band electron density at the surface gradually increases until some critical field (approximately \(F = 3.5 \text{ V/nm}\) for the p-type germanium emitters), after which the electron density at the surface decreases as a depletion layer forms, leading to the saturation of the current density. The formation of a depletion layer and the corresponding saturated emitted current density are a result of the finite mobility (and velocity) of electrons and holes within the semiconductor, which limits the magnitude of the flux density that can be supplied to the surface to replace the emitted electrons.

The effect of including continuity equations into the model for field emission from silicon is likely to predict a lower ECD at intermediate-strength applied fields and a smoother transition to the current density saturation regime, due to the earlier and more gradual onset of electronic flux supply limiting effects. Due to the lack of current continuity equations, the present model predicts that as the applied field increases, the bound state energies at the surface continue to fall lower in energy relative to the Fermi level at the surface (which is constant throughout the emitter), which increases the occupation probability of these states and the corresponding electron density. This trend of an increasing electron density at the surface continues until the flux density supplied by the bulk is depleted and the conduction band emitted current density saturates abruptly. Inclusion of a current continuity equation into the calculation would likely reproduce the trend of an increasing electron density with increasing surface field for relatively low fields, but eventually dictate a leveling off of the electron density as a function of increasing applied field, due to enough electrons being supplied from the bulk to meet the demand for the ECD, but not enough to
cause a surplus of electrons to accumulate near the surface, as illustrated in Figure 5-20a. The leveling off of the surface electron density with increasing field would lead to a lower ECD than is predicted by the present model; this marks the onset of electronic flux supply-related effects on the ECD. At higher fields, a depletion layer would form, corresponding to the situation in which all of the electrons able to be supplied to the surface from the bulk, limited by the saturation velocity of electrons, are almost immediately emitted. Thus, the electron density at the surface would be severely reduced and the ECD would be limited to the flux density of electrons that are able to be supplied from the bulk. The complete saturation of the conduction band ECD would likely occur at higher fields than the present model predicts, as the ECD at intermediate-strength fields would be lower (due to the constant electron density at the surface with increasing field) and would consume the supplied flux density less rapidly with increasing applied field than the present model. Since the saturation ECD is ultimately limited by the electronic flux density able to be supplied to the surface, the present model and the model with the current continuity equation would saturate at the same ECD, but the present model would saturate at a lower field than the model incorporating current flow, as illustrated in Figure 5-20b.

It should also be remarked that while the work of Baskin et al. lays the groundwork for studying the physics of current density saturation in semiconductor field emitters [113], its direct application to modern semiconductor field emitters may be limited, due to it being a one dimensional calculation. In their calculations, they estimate that the extent of the depletion region can be on the order of millimeters when current density saturation occurs. While this treatment of field emission may be accurate for emitter tips with large radii of curvature, it is not appropriate for emitters with radii of curvature on the order of several nanometers, in which the geometry of the emitter tip plays a significant role. For analogous calculations in modern, nanoscale semiconductor emitter tips, a multidimensional model is required.
5.4.3 Zero Internal Current Approximation

In the zero internal current approximation, the resistive voltage drop in the semiconductor due to current flow via field emission is ignored. Unless the voltage drop across the semiconductor is comparable to the voltage applied at the gate (or anode), this results in no significant modifications to the physics of field emission compared to the case in which it is ignored, aside from the downward shift in energy of the total energy distribution. In the case that the voltage drop across the emitter is non-negligible compared to the gate (or anode) voltage, it results in a decrease in the magnitude of the surface field, as the potential difference between the gate (or anode) and the emitter surface has decreased.

5.4.4 Crystal Structure of Semiconductor Bulk-like at Surface

Surface reconstruction, in which atoms at the surface of a semiconductor arrange themselves differently than in the bulk, often occurs within the two or three layers closest to the surface of a semiconductor and depends highly on the bonding structure.
of the atoms within the crystal. The crystal structure of a reconstructed surface is not unique and different reconstructions can occur, depending on the crystal face at the surface and the temperature. Surface reconstruction alters the properties of the surface states, may lead to the appearance of additional bands of surface states, and has a significant effect on the density of surface states [8]. As the local density of states at the surface of a semiconductor, equal to the sum of the density of states in the bulk and the density of surface states, is of prime importance when treating field emission, the physics of field emission is likely affected by the surface reconstruction. In particular, the crystal surface from which electrons are emitted becomes another important parameter that affects emission, as discussed in §3.4.2.

5.5 Chapter Summary

In this chapter, a model for field emission from bulk silicon emitters, including emission from conduction band extended states, conduction band bound states near the surface, and valence band extended states was developed. The finite dielectric constant of silicon allows for the electric field to penetrate some distance into the material, resulting in the bands bending downward near the surface. As a result, electrons with energies below the conduction band edge in the bulk are spatially confined near the surface and a two-dimensional electron gas forms. The potential and bound state energies near the surface of the emitter were calculated via a self-consistent solution to Poisson’s equation and the Schrödinger equation, which was accomplished by the simulation software nextnano3. The emitted current density and other field-emission-related quantities were calculated using the python programming language and incorporated a electronic-flux-supply-limited-ECD algorithm for conduction band emission, which was included to account for the finite electronic flux able to be supplied to the surface by the bulk for emission, as current continuity equations were omitted
in the simulation. Results show that for n-type emitters, emission is dominated by conduction band contributions to the ECD at low and intermediate-strength applied electric fields \((F < 5 \text{ V/nm})\) and by valence band contributions at high applied electric fields, as a result of the saturation of the conduction band ECD, which results from the limited flux density supply from the bulk. In p-type emitters, the conduction band ECD saturates at much lower applied fields \((F \approx 1.25 - 2 \text{ V/nm})\) due to the highly-limited electron density in the conduction band and experiences a flat region in the total ECD \(J\) vs. applied field \(F\) plot before the valence band ECD contributions dominate the total ECD at higher fields. The total energy distribution of emitted electrons from n-type emitters has a dominant peak in the conduction band at lower applied fields, which disappears at higher applied fields due to valence band emission dominating the ECD. For p-type emitters, the dominant peak is located in the valence band across all applied fields relevant to field emission. Fowler-Nordheim plots for silicon emitters exhibit two distinct linear regions: i) a low-field region which corresponds to a lower extracted tunneling barrier height and conduction-band-dominated emission; and ii) a high-field region which corresponds to a higher extracted tunneling barrier height and valence-band-dominated emission. The primary limitations of the model are its omission of emission from surface states and current continuity equations that are simultaneously solved alongside Poisson’s equation and the Schrödinger equation, which would likely predict an additional peak in the forbidden energy gap of the TED and a more gradual saturation of the conduction band ECD, which would reach its limiting value at higher applied fields.
Chapter 6

Field Electron Emission from Cylindrical Silicon Nanowires

6.1 Introduction

In the previous chapter, a model for field emission from bulk silicon, which includes the effects of confinement of electrons in the conduction band to a quantum well at the surface and the appearance of an internal tunneling barrier for electrons in the valence band (both due to downward band bending from electric field penetration) was developed. Although a model of field emission from bulk silicon captures the major physical phenomena inherent to emission from semiconductors, it omits the effects related to the low dimensionality of modern silicon emitters, which often have radii of curvature at the apex smaller than 10 nm. In order to realistically model field emission from a modern silicon emitter, one would need to self-consistently solve Poisson’s equation, the Schrödinger equation, and the current continuity equation within a curved silicon emitter tip with a radius of curvature on the order of a few nanometers. However, qualitative conclusions about the effects of lateral (transverse to the emission direction at the apex) quantum confinement on the performance of
silicon field emitters can be gained through the development of a much simpler model: the cylindrical silicon nanowire.

This chapter presents an extension of the model of field emission from bulk silicon to include the effects of a low-dimensional electron supply by limiting the dimensions of the silicon emitter to a radius of less than 10 nm. In order to ascertain the effects of lateral confinement on the properties of emission, the emitted current density as a function of the applied electric field, the total energy distribution of emitted electrons, and Fowler-Nordheim plots are investigated as a function of emitter radius. Additionally, the predictions of the cylindrical silicon nanowire model are compared to various sets of experimental I-V data and Fowler-Nordheim plots from silicon field emitter arrays.

6.2 Cylindrical Silicon Nanowire Emitter Model

6.2.1 Motivation for Cylindrical Silicon Nanowire Emitter Model

Ideally, simulations of a modern silicon field emitter tip would be based on self-consistent solutions to Poisson’s equation, the Schrödinger equation, and current continuity equation both inside the emitter and in vacuum. A continuous, three-dimensional solution to Poisson’s equation across the vacuum-silicon interface is critical due to it determining the surface potential barrier shape and height in multidimensional tunneling calculations, as well as the three-dimensional shape and extent of the quantum well that forms at the emitter tip due to electric field penetration. The shape of the quantum well at the emitter tip determines the bound state energies near the surface, as well as the tunneling attempt frequency of electrons in each of these states. The incorporation of a current continuity equation ensures that there is continuity between the current density able to be supplied to the emitter surface.
from the extended states and the current density that is emitted at the surface. As these three equations are coupled to each other through the potential and electron density, they must be solved simultaneously. Additionally, since the extent of the field penetration into the silicon is roughly on the order of the lateral dimensions of the tip, a three-dimensional simulation is required for an accurate model.

Unfortunately, the implementation of such a model within nextnano3 is hindered by a few key limitations of the simulation platform. Although nextnano3 is equipped with a Poisson-Schrödinger-Current solver, the incompatibility between the band structures of silicon and vacuum (or air) as defined in the simulation platform prevents the calculation of bound state energies in models that contain a silicon-vacuum interface. As a result, the construction of a simulation which calculates a continuous solution for the potential across the vacuum-silicon interface and the bound state energies in the quantum well near the surface is ruled out. Consequently, a model which isolates the electrostatics in vacuum from the electrostatics in the silicon is adopted, with the (analytical) solution to Poisson’s equation in the vacuum ideally being applied as a boundary condition at the curved emitter surface in nextnano3. A challenge also stands in the way of implementing this model, as it is not possible to impose Neumann boundary conditions (applied electric field) in multidimensional simulations or position-dependent Dirichlet boundary conditions (surface potential or local emitted current density) in nextnano3. Since electrostatics in vacuum are not handled within nextnano3 in this case, the tunneling calculation must also be handled externally and, therefore, the inclusion of a current continuity equation in the model is also not possible.

The limitations of nextnano3 in multidimensional simulations also restrict the range of possible emitter geometries. The finite element solvers of nextnano3 are based on a rectangular grid, which makes it challenging to create a smooth, rounded surface with a nanoscale radius of curvature at the apex. In order to create a smoother
surface, one must include more grid points in the domain, which rapidly increases the computational effort required in multidimensional simulations. Additionally, if Poisson’s equation were to be solved across the vacuum-silicon boundary at the emitter surface, the corners of the mesh elements that comprise the emitter surface would lead to artifacts in the calculated potential, which would significantly affect the tunneling barrier shape and electronic transmission probability. Additionally, the roughness of the surface of a cylinder on the rectangular grid introduces further artifacts into the self-consistent, multidimensional solution to Poisson’s equation and the Schrödinger equation inside the silicon, which would introduce errors into the potential shape and bound state energies near the emitter surface.

As a result of the above limitations, a model of field emission from a nanoscale silicon structure in nextnano3 that includes bound states near the surface and reduced lateral dimensions is restricted to consisting of: i) the isolation of emission electrostatics in vacuum from emission electrostatics within the silicon emitter; ii) the exclusion of self-consistent solutions to the current continuity equation; iii) a completely planar emitter surface that allows for one-dimensional tunneling calculations; and iv) a one-dimensional solution to the coupled Poisson-Schrödinger equations inside the emitter. An emitter geometry that satisfies the above requirements of having a planar emitter surface, an electron system that can (approximately) be modeled in a one-dimensional simulation space, and lateral dimensions small enough to exhibit the effects of quantum confinement is the cylindrical silicon nanowire emitter.

6.2.2 Electron System of the Cylindrical Silicon Nanowire Emitter

The cylindrical silicon nanowire emitter model is identical to that of the bulk silicon emitter model from §5.2, except for the quantization of the transverse momenta (and energies), due to the emitter geometry being defined by a nanoscale cylinder. As-
Figure 6-1: (a) Infinite circular well of radius $\rho_1$, with constant potential $V = 0$ inside and (b) cylindrical nanowire emitter of radius $\rho_1$, with constant potential $V = V_1$ inside, and an electric field of magnitude $F$ applied to its planar surface along the $z$ direction.

Assuming a constant potential inside the emitter and an infinite potential outside the emitter, the physics transverse to the emission direction can be modeled as an infinite circular well, as shown in Figure 6-1a.

The solution to the two-dimensional Schrödinger equation for the infinite circular potential well of radius $\rho_1$, with the potential defined by

\[
V(\rho) = \begin{cases} 
0 & \text{for } \rho < \rho_1 \\
\infty & \text{for } \rho \geq \rho_1,
\end{cases}
\]  

(6.1)

where $\rho$ is the radial coordinate of cylindrical coordinates, consists of the normalized wave functions

\[
\Psi_\rho(\rho) = A J_p(k_\rho \rho),
\]  

(6.2)

where $A$ is a normalization constant, $J_p(\rho)$ is the $p$th order Bessel function of the first kind, $k_\rho = \zeta_{p,r}/\rho_1$, and $\zeta_{p,r}$ is the $r$th zero of the $p$th order Bessel function of the first kind.
kind, and eigenenergies $E_{p,r}$

$$E_{p,r} = \frac{\hbar^2 k^2_{\rho}}{2m} = \frac{\hbar^2 \zeta^2_{p,r}}{2m \rho_1^2}$$  \hspace{1cm} (6.3)

where $p = 0, \pm 1, \pm 2, \ldots$ is the azimuthal quantum number, $r = 1, 2, 3, \ldots$ is the radial quantum number, $\hbar$ is the reduced Planck constant, and $m$ is the mass of the electron.

Due to the lateral confinement of electrons to an infinite circular well, the electron system far away from the emitter surface in the nanowire is one dimensional, having a single degree of freedom along the axis of emission $z$, as shown in Figure 6-1b. As the simulation is one-dimensional, it is assumed that the potential inside the emitter is able to be separated into components along the direction of emission and components transverse to the emission direction, such that the bound state energies near the surface become

$$E_{n,p,r} = W_n + E^t_{p,r}$$  \hspace{1cm} (6.4)

where $W_n$ are the bound state energies near the surface of the emitter from the model of emission from bulk silicon in Chapter 5 and are indexed by $n = 1, 2, 3, \ldots$ and $E^t_{p,r}$ are the infinite circular well energies of Equation 6.3. Due to the electrons near the surface being confined along the emission direction as well, they form a zero-dimensional electron system: a quantum dot.

### 6.2.3 Energetic Regimes of Emission from the Cylindrical Silicon Nanowire

As in the model for emission from bulk silicon, the energetic regimes of emission for the cylindrical silicon nanowire are defined in terms of their internal normal energies $W$, as shown in Figure 6-2. Although lateral quantum confinement of electrons results in all total energies having a discrete component, the different energetic regimes of emission are still designated by the internal normal energies of electrons relative to
Figure 6-2: Energy band diagram detailing the different energetic regimes of emission from the cylindrical silicon nanowire emitter. The regime to which an electron belongs is determined by the value of its normal energy $W$ (solid lines) relative to the band edge energy to which the electron belongs, far from the emitter surface. For reference, the total electron energies $E$ and ground state transverse confinement energies $E_{0,1}^l$ are also shown (dashed lines).

The band edge energies far from the surface and whether or not the internal normal energies are discrete or continuous. Thus, emission from states with normal energies in the conduction band with values less than that of the conduction band edge energy far from the emitter surface is referred to as emission from conduction band bound states, since the normal energies of electrons take on discrete values. The electron system of conduction band bound states is zero-dimensional. Emission from states with normal energies above the conduction band edge energy far from the emitter surface is termed emission from conduction band extended states, as the normal energies of electrons are continuous and the electronic wave functions extend throughout the crystal. Finally, all emission from the valence band is from valence band extended states, whose normal energies lie below the valence band edge energies far from the emitter surface. All extended states in the cylindrical silicon nanowire form a one-dimensional electron system, as they have a single degree of freedom in momentum space along the $k_z$ axis.
6.2.4 Emitted Current Density Equations

Emitted current density equations for each of the energetic regimes of emission mentioned in §6.2.3 are derived below. The expressions for the ECD equations in each of the regimes are similar to those from the bulk silicon model, except for modifications resulting from the discretization of the transverse momenta, which reduces the dimensionality of the electron supply to one-dimensional for extended states and zero-dimensional for bound states near the surface. Additionally, the tunneling barrier potential is identical to that of the bulk silicon emitter and is given by Equation 5.5. Again, all energies are referenced to a level that is independent of the magnitude of the band bending in the silicon, such as the Fermi level or one of the band edge energies far from the emitter surface.

Conduction Band Extended States

The total energy distribution for electrons emitted from conduction band extended states in the cylindrical silicon nanowire consists of a sum of the TEDs of individual subbands of a single conduction valley and is given as a function of total energy $E$ by

$$j_{ce}(F, E) = e \frac{2}{\pi \hbar \rho_1^2} \sum_{p,r} f_{FD}(E, T) D_{p,r}(F, W'(E, E_{p,r}^t)),$$  \hspace{1cm} (6.5)

where $\rho_1$ is the radius of the nanowire, $\hbar$ is the Planck constant, $f_{FD}(E, T)$ is the Fermi-Dirac distribution function, the total energy of the electron is given by $E = W + E_{p,r}^t$, the discrete transverse energies are given by $E_{p,r}^t = \hbar^2 \zeta_{p,r}^2 / (2 m_t \rho_1^2)$, $m_t$ is the transverse effective mass of electrons in that particular conduction valley, the transmission probability of electrons through the potential barrier at the surface is given by

$$D_{p,r}(F, W'(E, E_{p,r}^t)) = \exp \left[ -g_e \int_0^{z_t} \left[ V(z) - |\Delta E_c| - W'(E, E_{p,r}^t) \right]^{1/2} dz \right],$$  \hspace{1cm} (6.6)
where \( g_e = 2(2m_0)^{1/2}/\hbar \), \( z_t \) is the classical turning point in vacuum, \( |\Delta E_c| \) is the magnitude of the downward band bending at the surface, and the normal energy of electrons in vacuum is given by

\[
W'(E, E_{p,r}^t) = E - \frac{m_t}{m_0} E_{p,r}^t.
\]  

(6.7)

The emitted current density from a single conduction valley is given by integrating over all total energies

\[
J_{ce}(F) = e \frac{2}{\pi \hbar \rho_1^2} \sum_{p,r} \int_{E_{p,r}^t}^{E_{max}} f_{FD}(E, T) D_{p,r} (F, W'(E, E_{p,r}^t)) dE
\]

(6.8)

and the total ECD from the conduction band extended states results from summing over contributions from each of the conduction valleys, such that

\[
J_{ce}(F) = 2J_{ce}^{on}(F) + 4J_{ce}^{off}(F),
\]

(6.9)

where the superscripts “on” and “off” denote contributions from the on-axis and off-axis conduction valleys.

**Conduction Band Bound States**

Unlike the energy spectra for extended states, the bound states near the emitter surface have no degrees of freedom and the total energy distribution for such states is not defined as a continuous function of energy. An analogous expression is the ECD contributed by each state of a single conduction valley as a function of total energy \( E_{n,p,r} \),

\[
j_{eb}(F, E_{n,p,r}) = e \frac{2}{\pi \rho_1^2} f_{FD}(E_{n,p,r}, T) \nu_n(W_n) D_{n,p,r} (F, W'(E_{n,p,r}, E_{p,r}^t)),
\]

(6.10)
where $\nu(W_n)$ is the tunneling attempt frequency given by

$$
\nu(W_n) = \left\{ \int_{z_{t,n}}^{0} \left[ \frac{2m_n}{W_n - E_c(z)} \right]^{1/2} \, dz \right\}^{-1},
$$

(6.11)

where $z_{t,n}$ is the classical turning point in the semiconductor, $m_n$ is the normal effective mass of electrons in the conduction valley, $W_n'$ is the vacuum normal energy given by Equation 6.7 with $E = E_{n,p,r}$, and $D_{n,p,r}$ is the transmission probability through the tunneling barrier given by Equation 6.6 with $E = E_{n,p,r}$. The total ECD from a single conduction valley is given by summing over all relevant state indices, giving

$$
J_{cb}(F) = e \frac{2}{\pi \rho_1^2} \sum_{n,p,r} f_{FD}(E_{n,p,r}, T) \nu_n(W_n) D_{n,p,r} \left( F, W_n'(E_{n,p,r}, E_{p,r}^b) \right),
$$

(6.12)

and the total ECD from the bound states, including contributions from all conduction valleys is given by Equation 6.9.

**Valence Band Extended States**

As in the model for emission from the valence band of bulk silicon, a few variable substitutions are made for convenience, such as the effective mass of electrons in the valence band $m_e^*$ being replaced by the effective mass for holes such that $m_e^* = -m_v$, and all energies $E$ being measured as positive, downward from the valence band edge far from the emitter surface, such that $E = -E$, etc. In terms of the substituted variables, the total energy distribution for electrons emitted from a single valence band is given by

$$
\tilde{j}_{ve}(F, \overline{E}) = e \frac{2}{\pi \hbar \rho_1^2} \sum_{p,r} f_{FD}(-\overline{E}, T) D_{p,r} \left( F, -W'(\overline{E}, \overline{E}_{p,r}^b) \right),
$$

(6.13)

210
where the transmission probability $D_{p,r}$ is the product of the probability of tunneling through the downward-bent portion of the forbidden energy gap $D_{p,r}^{\text{int}}$, which is given by

$$D_{p,r}^{\text{int}}(F, W) = \begin{cases} \exp \left[ -g_e \int_{z_t}^0 \left[ W - E_v(z) \right]^{1/2} dz \right] & \text{for } 0 > W > |\Delta E_c(F)| \\ 1, & \text{for } W \geq |\Delta E_c(F)|, \end{cases}$$  

(6.14)

where $z_t$ is the classical turning point inside the emitter and $E_v(z)$ is the valence band edge, and the probability of tunneling through the nearly-triangular barrier at the surface in vacuum $D_{p,r}^{\text{ext}}$ is given by

$$D_{p,r}^{\text{ext}}(F, -W'(E, W)) = \exp \left[ -g_e \int_0^{z_t} \left[ V(z) - |\Delta E_c| + W'(E, W) \right]^{1/2} dz \right],$$  

(6.15)

and the normal energy in vacuum is given by

$$W'(E, E_{p,r}^t) = E - \frac{m_v}{m_0} E_{p,r}^t.$$  

(6.16)

The ECD from a single valence band is obtained by integrating over the total energies $\overline{E}$ of electrons in all subbands, giving

$$J_{ve}(F) = e \frac{2}{\pi \hbar \rho_1^2} \sum_{p,r} \int_{E_{p,r}^t}^{E_{max}} f_{FD}(-\overline{E}, T) D_{p,r} \left( F, -W'(E, E_{p,r}^t) \right) d\overline{E},$$  

(6.17)

and the total ECD from the valence band consists of contributions from each of the valence bands

$$J_{ve}(F) = J_{ve}^{\text{hh}}(F) + J_{ve}^{\text{lh}}(F) + J_{ve}^{\text{so}}(F),$$  

(6.18)

where the superscripts “hh”, “lh”, and “so” denote contributions from the heavy hole, light hole, and split-off bands respectively.
Approximations and Assumptions

A few approximations, in addition to those on which the model for emission from bulk silicon were built (see §5.2.5), were made in order to develop the model of field emission from a cylindrical silicon nanowire. Although the approximations are briefly discussed below, their impact on the predictions of the model are discussed in §6.5.

- **One-dimensional Simulation of Silicon Electrostatics:** It has been assumed in the model that the electrostatics of the cylindrical silicon nanowire only vary along the axis of emission $z$ and are constant in transverse dimensions. This assumption ignores any non-uniform effects on the potential inside the emitter that would result from the variation of the potential (or electric field) across the emitter surface due its edges or electron depletion effects due to surface states along the shank of the emitter. This assumption enables the treatment of the applied field as uniform across the emitter surface and the decoupling of the transverse electron momenta from the normal electron momenta throughout the emitter.

- **Potential and Normal Bound State Energies Independent of Lateral Quantum Confinement:** Although Poisson’s equation and the Schrödinger equation are solved self-consistently in the model of emission from a cylindrical silicon nanowire, they are solved assuming bulk silicon emitter properties and do not include the consequences of a reduced electron density in the emitter caused by lateral quantum confinement. Quantum confinement has the effect of reducing the density of states in energy, and therefore, also reduces the spatial density of states compared to the bulk case. Because the electron density also appears in Poisson’s equation, changes in the electron density due to quantum confinement in dimensions transverse to the emission direction should influence the shape of the potential along the emission direction, which then affects the bound state ener-
gies near the surface in the conduction band. This assumption is a direct result of the one-dimensional domain of the nextnano3 simulation, which calculates the electronic properties of the silicon emitter at each applied field.

- **Band Structure Independent of Emitter Dimensions**: It is assumed in the model of emission from a cylindrical nanowire that the dimensions of the emitter have no impact on the band structure of silicon; instead, the bulk silicon band structure is used. In reality, nanoscale silicon structures may not have the same band structure as their bulk counterparts and the band structure may vary with the dimensions of the structure. Band structure variations may include changes to the width of the forbidden energy gap or different band curvatures near the band extrema, from which the effective masses of carriers are derived. The band structure specific to an emitter of a particular radius could in principle be calculated via k.p theory. However, such a calculation would significantly increase the required computational effort, as it would necessitate at least one additional iterative simulation step in the nextnano3 calculation flow for each applied field.

### 6.3 Results

Results for emission from cylindrical silicon nanowires were generated using the same calculation flow as was used for emission from the bulk silicon emitters, as outlined in §5.3 and by Figure 5-9: i) nextnano3 was used to determine the potential, bound states, and electron effective masses in the silicon; and ii) the python programming language was used for all subsequent field emission calculations. As mentioned in the section above, the self-consistent solution to Poisson’s equation and the Schrödinger equation is found in a one-dimensional domain, along the emission axis of the emitter $z$, and consequently assumes bulk silicon characteristics. The introduction of lateral
quantum confinement effects is handled by the Python code: the form of the total energy of extended state electrons is taken to be \( E_{p,r}(W) = W + E_{t,p,r} \), where the only degree of freedom is in the internal normal energy \( W \), and the total energy of bound states near the surface is \( E_{n,p,r} = W_n + E_{t,p,r} \), where there are no degrees of freedom. The same algorithm as was used in the calculations for the bulk silicon emitter is also used for the cylindrical silicon nanowire calculations to enforce the limitations of the finite electron flux density supply from the conduction band extended states.

The following results include plots of the total emitted current density vs. applied field, the total energy distribution of emitted electrons, and Fowler-Nordheim plots, as a function of emitter radius, for both n-type and p-type silicon nanowire emitters of dopant density \( N = 10^{15} \text{ cm}^{-3} \), across a range of applied fields (in vacuum) between \( F = 0.5 \text{ V/nm} \) and \( F = 10 \text{ V/nm} \), at \( T = 300 \text{ K} \). The electron affinity of silicon is taken to be \( \chi_e = 4.04 \text{ eV} \), the forbidden energy gap is approximately \( E_g = 1.12 \text{ eV} \), and the standard bulk values of carrier effective masses are given in Table 5.1.

### 6.3.1 Emitted Current Density vs. Applied Field

**Results**

The total emitted current density from n-type and p-type cylindrical silicon nanowire emitters of dopant density \( N = 10^{15} \text{ cm}^{-3} \) as a function of applied field (in vacuum) \( F \), for nanowires with radii in the range \( 1 \text{ nm} \leq \rho_1 \leq 10 \text{ nm} \) is shown in Figure 6-3. Figure 6-3a plots the total ECD for n-type emitters, which shows that decreasing the emitter radius results in a noticeable decrease in the ECD for emitters with radii less than \( \rho_1 = 5 \text{ nm} \), relative to the ECD from a bulk emitter of the same dopant density. The total ECD for the emitter of radius \( \rho_1 = 1 \text{ nm} \) is most significantly reduced, especially for the range of applied fields \( 1 \text{ V/nm} < F < 6 \text{ V/nm} \). For emitters with radii of \( \rho_1 = 5 \text{ nm} \) and larger, the total ECD is approximately that of a bulk emitter. Figure 6-3b shows the total ECD from p-type emitters and reveals a similar trend to
Figure 6-3: The total emitted current density $J$ from (a) n-type cylindrical silicon nanowires of dopant density $N_D = 10^{15}$ cm$^{-3}$ and (b) p-type cylindrical silicon nanowires of dopant density $N_A = 10^{15}$ cm$^{-3}$, as a function of applied field (in vacuum) $F$ and for different nanowire radii $\rho_1$. The total ECD from a bulk emitter of the same dopant density is also shown (dashed line).

that seen in the n-type emitter data: while the total ECD from emitters with radii of $\rho_1 = 5$ nm and larger is constant as a function of emitter radius and nearly identical to the bulk ECD, emitters with smaller radii exhibit visibly lower ECDs, especially the emitter with $\rho_1 = 1$ nm. As an additional observation, the total ECD from the p-type emitter with $\rho_1 = 10$ nm is lower than for the bulk case in the flat region ($1 \text{ V/nm} < F < 2 \text{ V/nm}$) of the $J$ vs. $F$ plot, but tracks the bulk ECD profile across all other applied fields.

Analysis

The reduction in the total ECD exhibited by cylindrical silicon emitters with smaller radii occurs due to the total conduction band current density saturating at a lower value than for the nanowire emitters of larger radii, as shown in Figure 6-4. As a function of increasing emitter radius, the conduction band saturation ECD increases toward the bulk value, but still remains lower than that of the bulk for n-type and p-type nanowire emitters of radius $\rho_1 = 10$ nm. The conduction band ECD saturating at a lower, radius-dependent value results from the maximum electronic flux density supplied to the surface by the conduction band extended states being reduced with
Figure 6-4: Total conduction band current as a function of applied field $F$ and emitter radius $\rho_1$ for (a) n-type cylindrical silicon nanowire emitters of dopant density $N_D = 10^{15}$ cm$^{-3}$ and (b) p-type silicon nanowire emitters of dopant density $N_A = 10^{15}$ cm$^{-3}$. The total conduction band ECD of the corresponding bulk silicon emitter is plotted as a dashed line.

the increasing degree of quantum confinement. The electronic flux density supplied by the extended states in the conduction band is given by the expression

$$\Phi_b(T) = \frac{2}{\pi h \rho_1^2} \sum_{p,r} \int_{E_{p,r}^t}^{\infty} f_{FD}(E,T)\,dE,$$

where all energies are referenced to the conduction band edge, $E_{p,r}^t$ are the transverse quantum confinement energies given by the infinite circular well energies, and the sum is performed over all values of $p$ and $n$. The lower limit of integration in Equation 6.19 is $E = E_{p,r}^t$ due to the transverse energy of electrons in subband $\{p, r\}$ being fixed at a discrete value, which dictates the lowest total energy $E$ possible for those states.

The discretization of the transverse energy spectrum due to lateral quantum confinement has two significant consequences for the flux density supplied by the extended states in the conduction band: i) it increases the minimum state energy in each subband $\{p, r\}$; and ii) it reduces the density of states in energy. Figure 6-5a shows the overall minimum state energy for each of the conduction valleys of the
conduction band and each of the valence bands $E_{0,1}$, as a function of nanowire radius. For emitters with $\rho_1 < 5$ nm, the minimum state energies are clearly nonzero, with significant increases in the minimum state energy occurring for all emitters of radius $\rho_1 < 2$ nm. The most significant effect of this upward shift in the minimum state energy is the reduction in state occupation probability, which causes a reduction in the electron density within each subband as the emitter radius decreases. In addition to the reduction of the electron density within each subband is the reduction in the total electron density due to a reduced density of states in energy. Figure 6-5b shows the lowest five energies of the infinite circular well as a function of circular well radius, denoted by their subband indices $\{p, r\}$. As can be seen, the spacing between energy levels is non-negligible for emitters with $\rho_1 < 5$ nm and is significant for emitters with radii less than approximately $\rho_1 = 2.5$ nm. As a result, even though the minimum state energy in the well $E_{0,1}$ may not lie far above zero, there may be a significant energy gap between it and the minimum energy of the next-highest-energy subband. For small emitters with radii below 2.5 nm, this energy gap can be large enough to reduce the occupation probability of the higher-energy states and, consequently, their contributions to the extended state flux density supply to a level that is comparatively
insignificant. Thus, the combined effects of an increased minimum state energy in each subband and increased spacing between consecutive subbands are a lower electron density within each subband and fewer subbands contributing meaningfully to the extended state flux density supply. Ultimately, this leads to a significant decrease in the flux density supplied to the surface as a function of decreasing emitter radius. Conversely, as can be seen in Figure 6-5b, as the nanowire radius increases, the energy spacing between consecutive subbands decreases and a quasi-continuum of energies forms, similar to the bulk silicon case.

6.3.2 Total Energy Distribution

Results

The total energy distribution of emitted electrons for n-type and p-type emitters of dopant density $N = 10^{15}$ cm$^{-3}$, for various applied fields $F$, and emitter radii $\rho_1$ are shown in Figure 6-6. The TEDs for the n-type and p-type emitters of radius $\rho_1 = 1$ nm are shown in Figures 6-6a and 6-6b. Despite the different impurity types, the TEDs from the n-type and p-type emitters seem to be nearly identical across all applied fields considered, with the dominant peaks occurring in the valence band. Different from the TEDs of bulk silicon emitters is the appearance of subpeaks in the valence band TEDs, whose number and relative magnitude increase as a function of increasing applied field. These valence band TED subpeaks vanish for all emitters of larger emitter radii investigated. The TEDs of n-type and p-type cylindrical silicon nanowire emitters of radius $\rho_1 = 2.5$ nm in Figures 6-6c and 6-6d are markedly different from each other. While the n-type emitter shows a TED that is dominated by peaks in the conduction band across all applied fields, the p-type emitter exhibits a TED that is dominated by peaks in the valence band. For the n-type emitter, the relative heights of the valence band peaks increase as a function of increasing applied field; likewise, the relative heights of the conduction band peaks in the p-type emitter.
Figure 6-6: The normalized total energy distribution of emitted electrons for various applied field strengths (in vacuum) $F$ from (a) n-type and (b) p-type cylindrical silicon nanowire emitters of radius $\rho_1 = 1$ nm; (c) n-type and (d) p-type cylindrical silicon nanowire emitters of radius $\rho_1 = 2.5$ nm; and (e) n-type and (f) p-type cylindrical silicon nanowire emitters of radius $\rho_1 = 5$ nm. All n-type emitters are of dopant density $N_D = 10^{15}$ cm$^{-3}$, while all p-type emitters are of dopant density $N_A = 10^{15}$ cm$^{-3}$.
decrease as a function of increasing applied field. The TEDs for n-type and p-type emitters of radius $\rho_1 = 5$ nm are shown in Figures 6-6e and 6-6f, revealing the same qualitative trends with applied field as seen in the emitters of radius $\rho_1 = 2.5$ nm. Of special note is that the dominant peak for the p-type emitter is located in the conduction band for the lowest applied field considered. As a function of increasing emitter radius, the n-type emitter displays a widening TED in the conduction band, with a greater number of and higher peaks, along with mostly decreasing valence band peak heights. The same trends are observed in the p-type emitter TEDs as a function of increasing emitter radius, but are less pronounced than in n-type emitters.

**Analysis**

The decrease in the number and relative heights of conduction band peaks as a function of decreasing emitter radius is a direct result of the increase in minimum state energies and decrease in the supplied flux density to the surface from the conduction band extended states with decreasing emitter radius, as discussed in §6.3.1. Although the internal normal energies $W$ (and transmission probability) are not directly affected by lateral quantum confinement in this model, the increase in the minimum total energy of electrons in each subband due to the discretization of the transverse energies decreases the occupation probability of the states, as compared with the bulk silicon emitter. For example, while the total energy of states in the quantum well near the surface can range from $W_n \leq E < \infty$ in a bulk emitter, the range is limited to $W_n + E_{0,1}^t \leq E < \infty$ in a cylindrical nanowire emitter. The result is an increase in the minimum total energy at which emission from conduction band states may occur, which leads to a decrease in the maximum conduction band current density as a function of total energy $E$, as only states with lower occupation probabilities (higher total energies) are contributing to the ECD. As the emitter radius increases, $E_{0,1}^t$ and the spacing between consecutive transverse energies $E_{p,r}^t$ decrease rapidly.
towards zero, the minimum total energy at which emission from the conduction band can occur decreases towards the discrete normal energies near the surface $W_n$, and the maximum ECD as a function of total energy increases due to the higher occupation probabilities of states with lower $E$. As a result, the relative heights of peaks in the conduction band TED increase.

There are fewer conduction band TED peaks for emitters with smaller radii due to the large spacing between the minimum energies of states in consecutive subbands and the lower flux density supplied to the surface from the extended states. As was discussed in §6.3.1, the large energy spacing between consecutive subbands in smaller emitters leads to higher-energy subbands with relatively low occupation probabilities and negligible contributions to the total ECD. This causes only the lowest-energy subbands to contribute significantly to emission and only peaks corresponding to these subbands to appear in the conduction band TED. As the emitter radius increases, the energy spacings decrease and the electronic flux density supplied to the surface increases, leading to additional higher-energy subbands contributing significantly to emission and appearing in the conduction band TED.

The origin of the subpeaks in the valence band TED of emitters with radius $\rho_1 = 1$ nm was first investigated by decomposing the valence band TED into the contributions from each of the valence bands of silicon, where it can be seen in Figure 6-7a that the TED of each valence band also exhibits subpeaks. Decomposing the TED of the heavy hole valence band into the contributions from each of its constituent subbands in Figure 6-7b reveals that the subpeaks originate from the peaks of the contributions from the individual subbands $\{p, r\}$ of the valence band. The displacement of the subpeaks in energy is due to the spacing between consecutive, discrete transverse energies $E_{p,r}^t$ and the lower relative heights of the lower-energy subpeaks are a result of the change of the effective mass of the electron upon leaving the silicon. Since valence band normal energies $W$ are measured downward from the valence band
Figure 6-7: Plots of the total energy distribution of electrons emitted from the valence band of an n-type cylindrical silicon nanowire emitter of radius $\rho_1 = 1$ nm, for an applied field $F = 2.3$ V/nm. In plot (a), the TED is decomposed into the contributions from each of the valence bands of silicon and in (b) the TED of the heavy hole valence band is decomposed into the contributions from each of the subbands $\{p, r\}$, with the total TED of the heavy hole band shown by the dashed line.

edge far from the emitter surface, an increase in $W$ results in a larger external tunneling barrier height and width and a decreased transmission probability, as shown in Figure 5-3. When considering the effect of the change of the electron effective mass, the normal energy in vacuum is

$$ W' = \left(1 - \frac{m_e}{m_0}\right) E_{p,r}^{t} + W, $$

(6.20)

which indicates that the tunneling energy $W'$ is further increased by a fraction of the discrete, transverse energy $E_{p,r}^{t}$. This leads to a reduction in the transmission probability and relative TED peak height for subbands with larger transverse energies. The subpeaks are not visible in the TEDs of emitters with radii $\rho_1 \geq 2.5$ nm due to the rapid decrease in the spacing between discrete transverse energies, as shown in Figure 6-5b.
6.3.3 Fowler-Nordheim Plots

Results

Figure 6-8 displays the Fowler-Nordheim plots for n-type and p-type cylindrical silicon nanowire emitters of dopant density $N = 10^{15}$ cm$^{-3}$ of varying emitter radii $1 \text{ nm} \leq \rho_1 \leq 10 \text{ nm}$. The FN plots for the corresponding bulk emitters are shown as dashed lines.

Although the FN plot for the emitter of radius $\rho_1 = 2.5 \text{ nm}$ is shifted downward slightly along the $y$ axis as compared with the bulk emitter results, the FN plot for the emitter of radius $\rho_1 = 1 \text{ nm}$ is shifted significantly lower in $y$ and exhibits two distinct linear regions, with the lower region slope being smaller than that of the bulk case. A similar set of trends is seen in the FN plots for p-type emitters, as shown in Figure 6-8b, where emitters with radii of $\rho_1 = 5 \text{ nm}$ and larger have FN plots that are very similar to that of the bulk case and emitters with smaller radii are shifted downward along the $y$ axis. While the FN plot for the p-type emitter with $\rho_1 = 2.5 \text{ nm}$ has the same shape as in the bulk emitter case, the emitter of radius $\rho_1 = 1 \text{ nm}$ has a lower region slope that is smaller than that of the bulk case and a transitional
Figure 6-9: The extracted barrier heights for n-type (blue) and p-type (red) cylindrical silicon nanowire emitters of dopant density $N = 10^{15} \text{ cm}^{-3}$ from the upper linear region $H_U$ (circles) and lower linear region $H_L$ (triangles) of FN plots, as a function of emitter radius.

region slope that is clearly non-zero. As was also seen in the total ECD plots for p-type cylindrical silicon nanowires, the FN plots for emitters with large radii are slightly shifted downward in $y$ in the transitional region, as compared with the bulk emitter case.

The extracted barrier heights from the upper linear region $H_U$ and lower linear region $H_L$ of the FN plots are shown for both n-type and p-type emitters, as a function of emitter radius in Figure 6-9. As can be seen, while the upper region barrier heights for both n-type and p-type emitters increase as a function of decreasing emitter radius, the lower region barrier heights decrease as a function of decreasing emitter radius, with almost the same values for n-type and p-type emitters. For large emitter radii, the upper region barrier height for p-type emitters converges to approximately $H_U = 5.1 \text{ eV}$, while the n-type emitter upper region barrier height converges to approximately $H_U = 4.7 \text{ eV}$. The lower region barrier heights for both n-type and p-type emitters converge to approximately $H_L = 3.8 \text{ eV}$. Except for the upper region barrier height for n-type emitters, the barrier heights for emitters with large radii agree with the values extracted from FN plots for bulk emitters, as shown
in Figure 5-18.

**Analysis**

The extracted barrier height from the upper region \( H_U \) increases as a function of decreasing emitter radius due to the increase in the magnitude of and the spacing between the discrete transverse energies \( E^t_{p,r} \) of subbands in the valence band. As the magnitude of \( E^t_{p,r} \) increases, the vacuum normal energy \( W' \) of states in subband \( \{p,r\} \) with internal normal energy \( W \) shifts downward (see Equation 6.20), away from the valence band edge far from the emitter surface, increasing the tunneling barrier height. The difference between the valence band tunneling barrier heights of n-type and p-type emitters for larger emitter radii is due to the shape of the band bending near the surface due to electric field penetration. Also discussed in §5.3.4, while the band bending occurs rapidly and over the distance of \( \approx 1 - 3 \) nm in the accumulation layer of n-type emitters (see Figure 5-10a), it occurs in two phases for p-type emitters: i) a rapid potential drop within several nanometers of the emitter surface (also \( \approx 1 - 3 \) nm), in the inversion layer; and ii) a more gradual potential drop over a length of hundreds of nanometers in the depletion region (see Figure 5-10c). As the electrons near the valence band edge far from the surface in p-type emitters must tunnel through an internal tunneling barrier (downward-bent forbidden energy gap) that is hundreds of nanometers wide, the maximum transmission probability as a function of the vacuum normal energy \( W' \) is shifted away from the valence band edge, increasing the tunneling barrier height. For n-type emitters, the width of the internal tunneling barrier is only tens of nanometers wide and the maximum transmission probability lies much closer to the valence band edge far from the surface, resulting in a smaller tunneling barrier height. As the emitter radius decreases, however, the increasing discrete transverse energies shift the maximum transmission probability for the n-type emitter downward in energy far
Figure 6-10: Plot of the 20 lowest discrete total energies near the emitter surface in an n-type cylindrical nanowire emitter of dopant density $N_D = 10^{15}$ cm$^{-3}$, for an applied field of $F = 2.3$ V/nm, as a function of emitter radius $\rho_1$.

enough that the width of the internal tunneling barrier is the same for both n-type (accumulation layer) and p-type (inversion layer) emitters, leading to the same barrier height for both.

The decrease in the barrier height experienced by electrons emitted from the conduction band is due to the effect of the change of the effective mass of conduction band electrons upon leaving the silicon, as electrons gain additional normal energy in vacuum equal to a portion of the electron’s transverse energy inside the emitter, as diagrammed in Figure 5-3. Accordingly, as the emitter radius decreases, the discrete transverse energies $E_{p,r}^t$ and discrete total energies of the states $E_{n,p,r}$ increase, as illustrated by Figure 6-10. This causes electrons in bound states near the surface with internal normal energy $W_n$ to experience an increase in vacuum normal energy $W_n'$, according to

$$W_n' = \left(1 - \frac{m_t}{m_0}\right) (E_{n,p,r} - W_n) + W_n,$$  \hspace{1cm} (6.21)

where $m_t$ is the appropriate transverse electron effective mass. An increase in the total state energy energy $\Delta E_{n,p,r}$ for a fixed internal normal energy $W_n$, then corresponds to a decrease in the barrier height equal to $\Delta H_L = (1 - m_t/m_0) \Delta E_{n,p,r}$. The lower
region barrier heights of the n-type and p-type emitters being almost identical is due to the accumulation and inversion layers near the emitter surface having nearly the same shape, as seen in Figure 5-10.

Although the difference between the extracted barrier heights suggests evidence of conduction-band- and valence-band-dominated emission regimes, other physical phenomena taking place in the silicon when large electric fields are present should also be considered. The linear region that appears in the FN plots at higher fields could also be caused by carrier generation from impact ionization. If electrons acquire enough energy from the electric field inside the silicon \( (E \approx 1.5E_g) \) [145], they may be able to promote valence band electrons to the conduction band through collisions, which could then contribute to the ECD. Due to regular collisions with the lattice, electrons normally do not accrue this much energy, however, if the electric field is large enough, electrons may gain enough energy between collisions to precipitate carrier generation. The mean free path length of an electron in silicon can be estimated by the product of the electron’s saturation velocity \( v_{sat} \) and the collision time \( \tau_c \), yielding \( l = v_{sat}\tau_c \). In silicon, the saturation velocity of electrons is approximately \( v_{sat} = 10^7 \) cm/s, whereas the collision time is equal to \( \tau_c = \mu_n m_e^*/q \), where \( \mu_n \) is the mobility of electrons and \( m_e^* \) is the effective mass of electrons. In either n-type or p-type silicon with a dopant density of \( N = 10^{15} \) cm\(^{-3} \), the electron mobility at room temperature is \( \mu_n = 1345 \) cm\(^2\)V\(^{-1}\)s\(^{-1} \), which yields a collision time of \( \tau_c = 7 \times 10^{-13} \) s and a mean free path length of approximately \( l = 70 \) nm, where an effective mass of \( m_e^* = 0.9163m_0 \) has been used [11]. If it is assumed that the electron has 70 nm in which to gain energy from the electric field and that the amount of energy needed for impact ionization in silicon is \( 1.5E_g \approx 1.7 \) eV, the electric field must average at least \( F_{avg} \approx 0.0243 \) V/nm over the mean free path length of the electron. The average magnitude of the electric field in the silicon over a distance \( l \) from the emitter surface can be calculated via
The average electric field in the silicon $F_{S,\text{avg}}$ as a function of the applied electric field at the emitter surface in vacuum $F$, over the course of a single mean free path length of $l = 70$ nm from the emitter surface, for n-type and p-type silicon of dopant density $N = 10^{15}$ cm$^{-3}$. The dashed line denotes the minimum average electric field in the silicon needed for impact ionization $F_{\text{min}}$.

$$F_{S,\text{avg}}(F) = \frac{1}{l} \int_0^l F_S(F, z) \, dz,$$  \hspace{1cm} (6.22)

where $F_S$ is the magnitude of the electric field in the silicon and $F$ is the magnitude of the electric field at the emitter surface in vacuum. Figure 6-11 shows the average magnitude of the electric field in the silicon in the 70 nm closest to the emitter surface as a function of the surface electric field in the vacuum $F$, for both n-type and p-type silicon of dopant density $N = 10^{15}$ cm$^{-3}$, and reveals that the electric field in the silicon is unlikely to be large enough for impact ionization to occur over the mean free path length of an electron at this dopant density. Furthermore, since the accumulation and inversion layers at the surface, where the largest fields occur ($F_S > 0.05$ V/nm), have a maximum width of about 5 nm, the electrons would have relatively low energies until they are essentially at the emitter surface itself, leading to a very short distance over which electrons could collide with sufficient energy for impact ionization to occur. Because the electron mobility decreases as the dopant density increases, the collision time also decreases, leading to an even shorter mean free path length (or equivalently, a larger required average electric field) in which the electron must acquire the requisite...
amount of energy; this makes impact ionization even more unlikely in emitters with higher dopant densities. For lower dopant densities, however, impact ionization may be a possibility, especially for large surface fields.

6.4 Comparison to Experimental Results

The model of field emission from a cylindrical silicon nanowire was adapted to the experimental setups of single-gated silicon field emitter arrays (FEAs), whose tips have radii of curvature less than 10 nm, in order to compare with experimental I-V curves of the ECD per tip and the extracted slopes of Fowler-Nordheim plots. A diagram of a typical device structure for a single-gated silicon tip of an FEA is shown in Figure 6-12a. The devices consist of a silicon emitter tip that protrudes into a circular opening of radius $R_a$ that is in the same plane as the gate electrode, called the gate aperture. Normally, the gate aperture is positioned such that the silicon tip is located at the center of the circular aperture and is approximately coplanar with the gate electrode. The voltage $V$ that is used to extract electrons from the emitter is applied to the gate electrode and the extracted electrons are collected by an anode located far from the emitter tip. Figure 6-12b shows how the cylindrical silicon nanowire emitter model was adapted to the device structure in Figure 6-12a. While the electron supply and emitter properties are that of a cylindrical silicon nanowire of radius $\rho_1$, the surface field used in the ECD calculations is that of the electric field at the apex of a paraboloidal emitter of radius $\eta_1 = \rho_1$, inside a paraboloidal anode of radius $\eta_2 = R_a$, the radius of the gate aperture. Additionally, since the experimental data is given in terms of the emitted current, an effective emission area $A_{eff}$ was interpolated from high-field results in Figure 4-13a, corresponding to the emitter radius under consideration. No adjustable parameters were used in generating the data from the silicon nanowire model; only parameters reported in the experimental
Figure 6-12: (a) Diagram of a typical single-gated silicon field emitter structure, with gate aperture radius $R_a$ and applied voltage on the gate electrode $V$ [1]; and (b) diagram of the adaptation of the cylindrical silicon nanowire model for comparison to experimental setups similar to (a). For electron supply purposes, the silicon nanowire is taken to have the same radius $\rho_1$ as the reported radius of curvature at the apex of the actual emitter, and the electric field at the surface is taken to be that of the electric field at the apex of a paraboloidal emitter of radius $\eta_1 = \rho_1$, inside a paraboloidal anode of radius $\eta_2 = R_a$.

work or theoretical material values were employed. In the sections below, the adapted model is compared to experimental I-V data and FN plots from n-type FEAs with emitter tip radii of approximately $\rho_1 \approx 5$ nm [7] and to experimental FN plots from both n-type and p-type FEAs, which exhibit the two, distinct linear regions predicted by the models of emission from bulk silicon and cylindrical silicon nanowires [10].

6.4.1 I-V and FN Plot Data: n-type Single-gated Silicon FEAs

Results

Guerrera fabricated n-type silicon FEAs of various sizes (1x1, 10x10, 25x25, 32x32, and 50x50) with a reported dopant density of approximately $N_D = 2 \times 10^{15}$ cm$^{-3}$, tips with an average radius of curvature at the apex of approximately $\rho_1 = 5$ nm and
Figure 6-13: (a) The experimentally-measured I-V data per tip for n-type FEAs of dopant density $N_D = 2 \times 10^{15} \text{ cm}^{-3}$ and average tip apex radius of curvature $\rho_1 \approx 5 \text{ nm}$, for various array sizes. (b) Fowler-Nordheim plots of the data from (a). The predictions of the adapted cylindrical silicon nanowire emitter are shown as red circles, while Stratton’s bulk silicon model emitter predictions are indicated by a dashed, blue line [9].

height of 10 $\mu$m [9], and a gate aperture radius of $R_a = \eta_2 = 175 \text{ nm}$. Accordingly, the model data was generated based on a cylindrical silicon nanowire of radius $\rho_1 = 5 \text{ nm}$ and effective emitting area $A_{eff} = 68.4 \text{ nm}^2$, which corresponds to the 5 nm emitter radius result on the 37.5 V curve of Figure 4-13a, with the magnitude of the electric field being given by the apex electric field of a paraboloidal cathode of radius $\eta_1 = 5 \text{ nm}$ inside a paraboloidal anode of radius $\eta_2 = 175 \text{ nm}$. Guerrera’s experimental I-V curves of the measured current per emitter tip are shown for five different array sizes in Figure 6-13a, along with the predictions of the adapted cylindrical silicon nanowire emitter model and Stratton’s model for emission from bulk silicon, as detailed in §2.3. For lower applied voltages $V < 45 \text{ V}$, the prediction of the adapted silicon nanowire emitter model follows the experimental I-V curve quite closely, but with a slightly smaller turn on voltage and, consequently, a slightly larger emitted current. Over this range of applied voltages, Stratton’s model predicts a larger turn on voltage and lower emitted current than the experimental results and silicon nanowire emitter model. For applied voltages larger than $V = 45 \text{ V}$, the silicon nanowire emitter model diverges
from the experimental results, as the saturation of the conduction band current leads to an emitted current that rises less rapidly with increasing applied voltage. With no built-in conduction-band-ECD-limiting mechanism, Stratton’s bulk silicon emission model continues to rise rapidly with increasing applied voltage and tracks the I-V curve as closely as it did for lower applied voltages.

Figure 6-13b shows the Fowler-Nordheim plots for each of the I-V curves in Figure 6-13a, along with the predictions of the adapted silicon nanowire emitter model and Stratton’s bulk silicon emitter model. Although not easily visible in the I-V curves of Figure 6-13a, the FN plots of Figure 6-13b suggest that the real FEAs experience a limited current saturation at higher applied voltages, evidenced by the apparently higher FN plot slopes for lower applied voltages than for higher applied voltages. This region of lower slope could correspond to the transitional region seen in the FN plots of the bulk silicon and silicon nanowire emitter models, which typically has a lower slope than both the valence-band-dominated and conduction-band-dominated regimes. While this transitional region is present in the predictions of the adapted silicon nanowire emitter model, it occurs for lower applied voltages and eventually transitions into a valence-band-dominated regime with a higher slope. The real FEAs do not seem to reach a valence-band-dominated regime within the range of applied voltages tested, but it is possible that one would occur at higher applied voltages. The slope of the FN plot corresponding to Stratton’s model appears to be larger than that of the experimental data across all applied voltages.

Analysis

Three major aspects of the adapted cylindrical silicon nanowire emitter model likely lead to the discrepancies between its predictions and the experimentally-measured data: i) the lack of an appropriate model for emitter electrostatics; ii) the lack of a self-consistently-solved current continuity equation inside the silicon emitter; and
iii) the assumption of a constant nanowire radius, equal to the radius of curvature at the apex of the actual emitter tip. While the paraboloidal cathode-anode model for electrostatics was practical for a rough comparison to experimental data, more rigorous comparisons should use a model that is as close as possible to the actual device geometry shown in Figure 6-12a. This is particularly important for semiconductor field emitters, as both the conduction band and valence band emitted currents depend on the extent of field penetration into the emitter and the transmission probability through the tunneling barrier exerts an exponential influence on the emitted current. As a result, small inaccuracies in the electrostatics model can lead to large changes in emitter properties such as turn-on voltage and the slope of FN plots for I-V data. In this case, the paraboloidal cathode-anode model predicted a slightly smaller turn-on voltage and larger FN plot slope at lower applied voltages than seen in the experimental data.

Since the silicon nanowire emitter model predicts that the electronic flux density that can be supplied to the surface from the extended states has no effect on emission until the ECD exceeds it, plots of the total ECD vs. applied field show an abrupt saturation of the ECD from the conduction band. As discussed in §5.4.2, the inclusion of a current continuity equation would lead to an earlier, but much more gradual onset of the saturation of the conduction band ECD, causing: i) a less abrupt current saturation effect, occurring over a wider range of applied voltages; and ii) the emitter to reach the same saturation current at much larger applied voltages. Additionally, the silicon emitters that comprise the FEA are only nanoscale at the very tip; although the tips have an apex radius of curvature of 5 nm, the nanowires upon which the tips are located have a radius of approximately 100 nm. This leads the adapted cylindrical silicon nanowire emitter model, which assumes a constant emitter radius of 5 nm, to vastly underestimate the flux density able to be supplied by the extended states, due to lateral quantum confinement effects. In reality, the larger lateral dimensions
of the actual emitter tip away from the apex lead to a lesser degree of quantum
confinement and a larger density of states, as compared with the cylindrical silicon
nanowire model. As a result, a silicon emitter tip model with tapered dimensions
would be able to supply a larger flux density of electrons to the surface, causing the
model to predict saturation at a larger ECD and for a larger applied voltage.

6.4.2 FN Plot Data with Multiple Emission Regimes: n-type
and p-type Single-gated Silicon FEAs

Results

Ding reported Fowler-Nordheim plots of n-type and p-type single-gated silicon FEAs,
which were characterized by two distinct linear regions, separated by a region of low
slope in his doctoral thesis work, as shown in Figure 6-14 [10]. The adapted silicon
nanowire emitter model was compared to FN plots from two different types of FEAs
from Ding’s work: i) a 60x60 n-type silicon FEA of dopant density \( N_D = 2 \times 10^{14} \)
cm\(^{-3}\) and estimated average emitter tip radius of curvature \( \rho_1 = 9.84 \) nm; and ii) a
20x20 p-type silicon FEA of dopant density \( N_A = 10^{15} \) cm\(^{-3}\) and estimated average
emitter tip radius of curvature \( \rho_1 = 5 \) nm. The gate aperture radius in Ding’s FEAs
was reported to be approximately \( R_a = 0.5 \) \( \mu \)m. Figures 6-14a and 6-14b show
the FN plots for the adapted silicon nanowire emitter model and experimentally-
measured I-V data of the n-type silicon FEA mentioned above. Although qualitatively
similar, with distinct, linear high-applied-voltage and low-applied-voltage regions,
the transitional region of the model’s FN plot has a larger slope than that of the
measured data. Additionally, as listed in Table 6.1, the extracted plot slopes \( b_{FN} \) of
the model’s prediction are as much as 50% larger than those for the experimentally-
measured data. Figures 6-14c and 6-14d show the FN plots for the p-type silicon
FEA described above. Again, both the model and the experimental data show two
Figure 6-14: Fowler-Nordheim plots for (a) the prediction of the adapted silicon nanowire model and (b) experimentally-measured I-V data for a 60x60 n-type silicon FEA of dopant density $N_D = 2 \times 10^{14} \text{ cm}^{-3}$ and estimated average emitter tip radius of curvature of $\rho_1 = 9.84 \text{ nm}$; and Fowler-Nordheim plots for (c) the prediction of the adapted silicon nanowire model and (d) experimentally-measured I-V data for a 20x20 p-type silicon FEA of dopant density $N_A = 10^{15} \text{ cm}^{-3}$ and estimated average emitter tip radius of curvature of $\rho_1 = 5 \text{ nm}$ [10]. The extracted FN slope plots $b_{FN}$ are shown in the plot insets and in Table 6.1.

distinct linear regions in the FN plot separated by a transitional region, but this time the model’s FN plot displays a transitional region with a slope of nearly zero. Again, the slopes extracted from the model’s predictions are significantly larger than those extracted from the experimental data. Although the discrepancies between the extracted slopes may result from inaccuracies in electrostatics calculations used in the adapted silicon nanowire emitter model, the ratio of the slope from the valence-band-dominated current regime to the slope from the conduction-band-dominated current regime should be comparable to the “reference” value of that ratio: $r = 1.44$, 235
Table 6.1: The extracted FN plot slopes from the valence-band-dominated and conduction-band-dominated emission regimes and the ratio $r$ of the valence band $b_{FN}$ to the conduction band $b_{FN}$, for the adapted silicon nanowire emitter model and the experimentally-measured data of Ding [10].

which originates from the assumption that the barrier height in the conduction band is equal to the electron affinity of silicon $\chi_e = 4.04$ eV and the barrier height in the valence band is equal to the $\chi_e + E_g = 5.16$ eV. Table 6.1 shows that the ratios of the extracted slopes are similar to the reference value.

**Analysis**

The difference between the extracted slopes from the FN plots of the model and the experimental data can mostly be attributed to a discrepancy between the model electrostatics and the actual emitter tip electrostatics. Since emission models based on planar emitter tip geometries naturally give the emitted current density as a function of the applied field, an effective emitting area $A_{eff}$ and a field factor $\beta$ must be introduced to convert the current density to a current and the applied field to a voltage. As a result, comparing the slopes of FN plots from experimental I-V data to the predictions of a field emission model requires that the model define a field factor $\beta$ (through establishing a cathode-anode geometry), which appears in the extracted slope from the FN plots of I-V data, as

$$b_{FN} = \frac{0.95B}{\beta}H^{3/2},$$  
(6.23)
where $B$ is the second Fowler-Nordheim constant, $H$ is the relevant tunneling barrier height, and the factor of 0.95 accounts for the lowering of the barrier due to the inclusion of the image charge potential. Consequently, discrepancies in $\beta$ due to inaccurate modeling of emitter electrostatics lead to differences in extracted FN plot slopes. Additionally, since Ding reports the emitter radii via comparison to simulation and not by direct measurement via microscopy, it is possible that the actual emitter tip radius at the apex is different from the reported value and it is likely that there is appreciable variation in emitter tip radii across the FEA, which would further increase the discrepancy between the model and actual emitter electrostatics. In the cases above, the larger extracted slopes of the model FN plots indicate a smaller $\beta$ and that the model underestimates the electric field at the emitter surface for a given applied voltage $V$.

The discrepancy between the model’s slope ratio and the measured slope ratio can be explained in terms of emitter electrostatics and the effects of lateral quantum confinement. When calculating the ECD from the adapted cylindrical silicon nanowire emitter model, it is assumed that the field factor $\beta$ is not a function of the applied voltage. Therefore, the ratio of the extracted slopes should also be the ratio of the barrier heights raised to the $3/2$ power. Real emitters, however, may experience blunting from damage at the large applied voltages associated with valence-band-dominated emission, which would result in a smaller $\beta$ and a relatively larger value for the slope in the upper linear region. For the n-type FEA, this leads to an extracted slope ratio that is larger than that of the model’s prediction. This is not the only phenomenon to consider in the case of the p-type FEA though, whose emitter radius of $\rho_1 = 5$ nm suggests that quantum confinement effects may play a role in it having a larger slope ratio than that of the experimental data, since the discretization of the transverse energies affects the barrier height seen by tunneling electrons. As mentioned in §6.4.1, real silicon emitter tips have a tapered geometry with a nanoscale...
radius of curvature at the apex, while the rest of the tip away from the apex has larger lateral dimensions. As a result, the effects of lateral quantum confinement play a more significant role in the model than in the actual emitter tip. For example, stronger lateral confinement of electrons in the model leads to a higher ground state transverse energy and larger spacing between energy levels. The non-zero ground state energy in the emitter leads to a reduction of the tunneling barrier height for electrons in the conduction band and an increase in the tunneling barrier height for electrons in the valence bands, due to the “extra” normal energy gained by tunneling electrons from the change of the electron’s effective mass upon leaving the silicon. Thus, a simultaneous decrease in the conduction band barrier height and increase in the valence band barrier height due to an overestimation of the effects of lateral quantum confinement leads to a larger extracted slope ratio for the model than for the ratio of slopes extracted from the experimental data. Quantum confinement likely plays no significant role in the barrier heights for the n-type FEA, as emitters of radius \( \approx 10 \text{ nm} \) have been shown in previous sections to display bulk-like characteristics.

As an additional note, the deviation of the extracted slope ratios from the reference ratio is not surprising. Due to the majority of the emitted current at low applied voltages originating from bound states in the conduction band, which lie above the conduction band edge at the surface, the extracted conduction band barrier height is expected to be less than that of the electron affinity of silicon \( \chi_e = 4.04 \text{ eV} \). Because of relative barrier height lowering due to downward band bending, it is also possible for the extracted valence band barrier height to be less than that of the sum of the electron affinity and forbidden energy gap of silicon \( \chi_e + E_g = 5.16 \text{ eV} \), as electrons with energies near the valence band edge can tunnel through the downward-bent forbidden energy gap and be emitted. For n-type emitters, the relatively lower valence band barrier height leads to an extracted slope ratio that is lower than that of the reference ratio, while the larger valence band barrier height in p-type emitters, due
to the wide internal tunneling barrier formed by the depletion region and the normal energy “gained” from the change of the electron mass upon leaving the silicon, leads to an extracted slope ratio that is larger than the reference ratio.

The FN plots corresponding to the model predictions also exhibit the sudden onset of the transitional region, characterized in n-type emitters by a rapid change of slopes and in p-type emitters by a slope of approximately zero. This behavior is a result of the abrupt saturation of the conduction band emitted current once the electronic flux supplied by the extended states is depleted. In reality, a depletion layer does not suddenly appear, but gradually forms with increasing applied voltage as the electron density near the surface is diminished due to emission. If a current continuity equation were incorporated into the model, the emitted current from the conduction band would saturate more gradually, over a much wider range of applied voltages. This would result in a less abrupt transitional between the conduction-band-dominated and valence-band-dominated regimes and a transitional region defined by a lower, but non-zero slope in both n-type and p-type emitters, as seen in the experimental data.

6.5 Limitations of the Model

In this section, the limitations of the model of field emission from a cylindrical silicon nanowire, due to the assumptions that were made during its derivation, are discussed. Only the assumptions specific to the cylindrical silicon nanowire emitter model are addressed, as the consequences of the assumptions made in the bulk silicon emitter model, upon which the silicon nanowire emitter model is also based, are discussed in §5.4.
6.5.1 One-dimensional Simulation of Silicon Electrostatics

A major assumption made while deriving the model is that the electrostatics inside and outside the silicon nanowire can be treated with a one-dimensional simulation. In order for this to be a good approximation, the depth into the emitter over which the potential varies must be much smaller than the dimensions parallel to the emitter surface over which the potential is approximately constant. In the case of a cylindrical nanowire emitter with a radius of less than $\rho_1 = 10 \text{ nm}$, this is certainly not a good approximation, as, for example, the total depth of the combined inversion layer and depletion region in a p-type emitter of dopant density $N_A = 10^{15} \text{ cm}^{-3}$ is about $1 \mu\text{m}$.

Such an approximation mostly neglects the effects of the emitter tip geometry on the potential inside and outside the emitter. In the case of the cylindrical nanowire geometry, the magnitude of the electric field would likely be enhanced at the edges of the circular emitter surface and would decrease across the emitter surface as a function of distance from the edge. Aside from invalidating a one-dimensional calculation of the transmission probability, this non-uniformity in the electric field across the emitter surface also corresponds to a potential inside the emitter that varies as a function of position in planes parallel to the emitter surface. As a result, the shape of the quantum well near the surface to which electrons are confined would no longer have transverse dimensions with a radius that is constant as a function of depth into the emitter; instead, the transverse dimensions of the quantum well would vary as a function of distance from the emitter surface. Not only would this couple electron momenta across all three dimensions and rule out the possibility of a one-dimensional model for calculating field emission quantities from an electron supply perspective, but it would also alter the energies of states in the quantum well and their tunneling attempt frequency. Consequently, a new model for field emission from silicon that does not rely on the ability to separate the normal momentum from the transverse momentum would need to be developed, most likely using a multitude of classical electron...
trajectories in the quantum well at the surface to determine the flux density incident upon the silicon side of the emitter surface. Although it is difficult to ascertain exactly how these alterations to the model would affect its predictions concerning field emitter performance and properties without first performing the multidimensional simulations, it is clear that the specific geometry of the emitter tip significantly affects all physical aspects of the emission process, and a one-dimensional calculation of electrostatics is especially invalid in the case of highly-curved emitter tips.

6.5.2 Potential and Normal Bound State Energies Independent of Lateral Quantum Confinement

It was also assumed in the model, since Poisson’s equation and the Schrödinger equation were solved self-consistently for a bulk silicon emitter, that the reduction in the electron density in the nanowire due to lateral quantum confinement has no effect on the emitter electrostatics. Since the solution to Poisson’s equation depends directly on the electron density, the overall decrease in the electron density due to lateral quantum confinement would lead to alterations in the potential in the silicon and the bound state energies in the quantum well near the surface. In an n-type silicon nanowire emitter, a lower electron density would lead to fewer electrons available near the surface to screen the electric field, allowing the applied field to penetrate further into the emitter and to form a wider accumulation layer near the surface. A wider accumulation layer would lead to weaker spatial confinement of electrons and lower bound state energies near the surface. For p-type silicon nanowire emitters, a lower electron density would lead to a narrower inversion layer, but a wider depletion layer, due to fewer electrons being available to form an inversion layer at the surface and fewer holes being available to screen the electric field in the depletion region. The stronger spatial confinement experienced by electrons in the inversion layer would lead to an increase in the bound state energies near the surface. However, in both
cases, the solution to the Schrödinger equation would provide wave functions, from which new electron densities are calculated and inserted into Poisson’s equation as the first iteration of finding a self-consistent solution. Due to the self-consistent nature of the calculations, accurate predictions about the effect on the properties of emission of including this coupling in the model as compared to the model based on bulk emitter properties are difficult to make. The decrease in the internal normal energy of bound states at the surface due to a wider accumulation layer would lead to an increase in the average occupation probability of each state, but could also cause a decrease in both the tunneling attempt frequency (depending on the width and depth of the quantum well for each individual energy) and transmission probability, due to the increased barrier height seen by electrons in lower normal energy states. On the other hand, a narrower inversion layer in p-type emitters would have an opposite set of effects on the properties of emission, as compared with the wider accumulation layer of n-type emitters. The overall effect of these competing influences on the emitted current density, however, is difficult to ascertain.

6.5.3 Band Structure Independent of Emitter Dimensions

No considerations were made in the model concerning how the band structure of silicon changes as a function of emitter dimensions. The continuous bands of available energy states that comprise the band structure of a solid arise due to the quasi-infinite extent of the crystal structure. A reduction of the silicon crystal dimensions below 5 nm causes the crystal to exist of approximately 10 atoms in that dimension, leading to a significant decrease in the degree to which the orbitals of the atoms overlap and the decomposition of the quasi-continuum of energy states into a set of discrete energy states. Some of the possible effects are an increase in the width of the forbidden energy gap between the valence band maximum and conduction band minimum, since the (now discrete) ground state energy in each of the bands increases
as quantum confinement effects begin to set in, and an increased effective mass due to
the flattening of the band extrema that often accompanies the increase in the width
of the forbidden energy gap [146]. An increase in the width of the forbidden energy
gap would lead to lower carrier concentrations due to the increased energy difference
between the band edges far from the surface and the Fermi level and a decreased
transmission probability for electrons tunneling from the valence band, due to an
increased barrier height. As a result of larger transverse effective masses, the discrete
transverse energies and total energies of electrons with normal energy $W$ would shift
downward, resulting in an increased average state occupancy and, possibly, a slightly
larger electronic flux density supply from the extended states for a given emitter
radius and applied field.

6.6 Chapter Summary

In this chapter, a model for field emission from cylindrical silicon nanowires, which is
based on the same underlying physics as the model for field emission from bulk sili-
con in Chapter 5, was developed. Results showed that the effects of lateral quantum
confinement were significant for emitters with radii below 5 nm, which exhibited: i)
conduction band emitted current density saturation at lower applied fields; ii) lower
TED peak heights in the conduction band; iii) a shift of the transitional region in
Fowler-Nordheim plots to lower applied fields; iv) smaller extracted barrier heights
for conduction band electrons; and v) larger extracted barrier heights for valence
band electrons, when compared to the corresponding bulk emitter properties. While
conduction band ECD saturation and the transitional region of FN plots occur at
lower fields due to a lower electron density in the emitter as a consequence of a lower
density of states, the conduction band TED peak heights are reduced as a result of the
lower average occupation probability of states that contribute significantly to emis-
sion, which arises from an increase in the minimum total energy within each subband. The effect of the change of the electron effective mass upon leaving the silicon leads to the decreased (increased) extracted tunneling barrier heights in the conduction (valence) band, as the states dominating emission have higher total energies and transfer larger quantities of transverse energy to the vacuum normal energy. A comparison of the predictions of a modified version of the cylindrical silicon nanowire emitter model to experimentally-measured I-V data from n-type silicon field emitter arrays showed excellent agreement with the experimental results without the use of any adjustable parameters, until the onset of the model’s current saturation effects. Likely due to the exclusion of a current continuity equation in the model and the underestimation of the available electron flux density supply from the extended states due to a cylindrical nanowire geometry, the adapted silicon nanowire model underestimates the voltage at which conduction band current saturation occurs and overestimates the rate at which it occurs as a function of applied voltage. When compared to FN plots from experimentally-measured data for n-type and p-type emitters, the model correctly predicted qualitative features, such as the appearance of distinct linear regions at low and high applied voltages, separated by a transitional region, but overestimated the extracted slope values, likely due to inaccuracies in the model of electrostatics employed for the comparison. The ratio of extracted slopes, however, should be nearly independent of the electrostatics model used and both the experimentally-measured data and the predictions of the model provided ratios close to the standard theoretical value. Major limitations to the model that should be addressed in future work are the inclusion of a current continuity equation to be solved self-consistently with Poisson’s equation and the Schrödinger equation and the expansion of the simulation domain to multiple dimensions, as a one-dimensional model for electrostatics is invalid for the case in which significant variations in the potential of the emitter reach depths that are on the order of the emitter radius.
Chapter 7

Thesis Summary and Suggestions for Future Work

7.1 Thesis Summary

Modern field emitters are structures with nanoscale dimensions, a highly-curved emitter surface, and are often fabricated from semiconductor materials. While experimentalists have succeeded in fabricating large arrays of ever-smaller field emitter tips to serve as electron sources in nanoelectronic devices, the theory and modeling work required to understand the physics of field emission in these devices have not kept pace. Although separate treatments of field emission that address individual deviations from Fowler-Nordheim theory, such as semiconductor emitters, highly-curved emitter tips, and emitters with low-dimensional electron systems have been developed, no studies have sought to treat these emitter properties coherently within a single framework or on equal footing with each other. As a result, it was the goal of this thesis work to develop a framework for field emission, from which models for field emission from structures of any dimensionality, geometry, and material can be derived.
Chapter 3 detailed the development of the semiclassical framework for field electron emission and its constituent parts. The framework is based upon a multidimensional, semiclassical approximation which reduces the Schrödinger equation to a pair of familiar classical equations: the Hamilton-Jacobi equation of classical mechanics and the continuity equation of fluid dynamics. These two equations define the electronic wave functions within the emitter to first order, which are subsequently used to derive expressions for the physical quantities in the emitter, such as the bound state energies and incident flux density attempting to tunnel into vacuum. The calculation of the transmission probability of electrons through the potential barrier at the emitter surface is also semiclassical and multidimensional, and consists of a phase integral along the “classical” path of the electron through the classically-forbidden region. An alternate formulation of the incident flux density is also given for cases in which the electrons attempting to tunnel are strongly confined to a quantum well of arbitrary geometry in one dimension. The final result is a general emitted current density equation consisting of physical quantities that are calculated from the semiclassical wave functions in the emitter and a quantum-mechanics-based framework within which the physics of field emission can be interpreted in terms of the often more intuitive concepts of classical mechanics.

In Chapter 4, the semiclassical framework for field emission was applied to derive a model of field emission from a low-dimensional, highly-curved, metal emitter tip: a nanoscale paraboloid. The simulated device structure consisted of a nanoscale paraboloidal cathode inside a larger paraboloidal anode. Results showed that despite a reduced incident flux density as emitter dimensions decreased, the emitted current density increased as a function of decreasing emitter radius and fixed applied voltage, as a result of gains in the transmission probability due to barrier thinning at the emitter apex (electric field enhancement). Additionally, it was shown that the overwhelming majority of the current originates from an area around the emitter apex
on the order of approximately $1 - 200 \text{ nm}^2$ and from subbands with zero angular momentum. Due to the convex shape of the tunneling barrier potential near the surface, the total energy distributions of electrons emitted from paraboloidal emitters with radii below 5 nm were much narrower and more highly peaked about energies near the Fermi energy than predicted by Fowler-Nordheim theory. Furthermore, the non-triangular barrier shape resulted in significant non-linearities in Fowler-Nordheim plots for paraboloidal emitters with radii smaller than approximately 5 nm, leading to sizable errors in the extracted work function and field factors. From these results, it is clear that the geometry and dimensionality of nanoscale emitters have a measurable effect on both the electron supply in the emitter and the electrostatics outside the emitter, and that one-dimensional models of field emission based on a bulk, planar emitter are not adequate at this scale.

A model for field emission from bulk silicon emitters was developed in Chapter 5, which includes physical phenomena new to field emission models, such as the confinement of electrons to a quantum well at the surface and an additional potential barrier (forbidden energy gap) through which electrons near the valence band edge in the bulk must tunnel. Both of these phenomena arise as a result of the downward bending of the conduction and valence bands near the surface, due to field penetration into the silicon. As the model includes emission from electrons confined to a quantum well at the surface, the potential and bound state electron energies near the emitter surface were calculated via a self-consistent solution to the coupled Poisson-Schrödinger equations. Additionally, the emitted current density from the conduction band was calculated via an algorithm which ensured that the predicted emitted current density did not exceed the total electronic current density able to be supplied to the surface from extended states in the bulk. Plots of the total emitted current density vs. applied electric field and the corresponding Fowler-Nordheim plots showed that the saturation of the conduction band ECD due to depletion of the supplied conduction band
electronic flux density leads to two distinct regimes of emission: i) a low field regime within which the total ECD is dominated by contributions from the conduction band and ii) a high field regime within which the total ECD is dominated by contributions from the valence band. Although present in both n-type and p-type emitters, the valence-band-dominated regime is much more pronounced in p-type emitters due to their relatively limited conduction band electronic flux density supply. Furthermore, precedence exists for such Fowler-Nordheim plots, as qualitatively similar plots have been observed in experimental data from both n-type and p-type silicon field emitter arrays.

In order to investigate the effects of quantum confinement in dimensions transverse to the emission direction, as experienced by nanoscale emitter tips, a model for field emission from cylindrical silicon nanowires was developed in Chapter 6, grounded in the same physical principles as the model for emission from bulk silicon in Chapter 5. Results revealed that the most significant consequence of lateral quantum confinement was a reduction in the electronic flux density able to be supplied to the surface from conduction band extended states. Further limiting of the conduction band flux supply led to an overall decrease in the emitted current density for emitters with radii smaller than 5 nm and a shift of the transitional region of FN plots to lower applied fields due saturation of the conduction band ECD at lower applied fields. Less significant consequences included the appearance of subpeaks in the valence band total energy distribution and an increase (decrease) in the extracted barrier height corresponding to the valence-band-dominated (conduction-band-dominated) regime, due to the effects of the change in the electron’s effective mass upon leaving the semiconductor. The cylindrical silicon nanowire emitter model was also adapted for comparison to experimentally-measured I-V data and FN plots from n-type and p-type, single-gated silicon field emitter arrays. The comparison showed that although the model’s predictions found excellent agreement with the experimental data at
low to intermediate applied voltages, the model predicted a premature saturation of the conduction band ECD, which led to an underestimation of the total ECD at larger applied voltages. The ratio of the slope extracted from the valence-band-dominated region to the slope extracted from the conduction-band-dominated region of Fowler-Nordheim plots generated by the model were found to be comparable to both the experimentally-determined values and the standard theoretical values. It is suspected that the discrepancies between the silicon emitter model’s predictions and the experimental results follow from the three aspects of the model: i) the omission of a current continuity equation that is solved self-consistently with Poisson’s equation and the Schrödinger equation; ii) inaccurate modeling of emitter electrostatics; and iii) the underestimation of the lateral emitter dimensions away from the emitter apex.

7.2 Suggestions for Future Work

Each of the models that comprise this thesis contain approximations that were invoked either on the grounds of practicality or due to the limits of the simulation platforms employed in the calculations. Some of these assumptions significantly limit the physical accuracy of the field emission models and if more realistic models of emission from nanoscale field emitters are to be developed, the technical challenges that made these assumptions necessary in this work must be overcome. The ultimate effort, of which this work is only a part, is to develop a fully multidimensional, semiclassical model of field emission from nanoscale emitter tips, and the suggestions for future work listed below directly support this goal.

Although the metal paraboloidal emitter model considered a cathode with a realistic geometry, the anode geometry may be unrealistic. The anode geometry was chosen to be a paraboloid so that the potential between the anode and cathode had a closed form in parabolic coordinates, which facilitated the eventual numerical calculations
of electron tunneling probabilities. A probably more realistic cathode-anode geometry for nanoelectronic device simulations would be that of a highly-curved emitter tip and a planar anode. For metal emitters, this electrostatics problem has a closed form solution in prolate spheroidal coordinates, as the hyperboloidal emitter tip and planar anode are both surfaces of constant coordinate. Accordingly, a future project should investigate the properties of emission from a hyperboloidal cathode, with the extraction voltage applied by a planar anode.

Actual silicon field emitters taper to a rounded tip with a nanoscale radius of curvature at the apex and are likely not well-represented by silicon nanowires with planar emitter surfaces. Therefore, future models of emission from silicon tips should be based upon a rounded silicon emitter tip and the solution to Poisson’s equation in multiple dimensions should be found both inside and outside the emitter, being continuous across the silicon-vacuum boundary. In silicon emitters, the geometry of the emitter tip is especially important for electrostatics calculations inside the emitter due to the properties of the quantum well created by band bending near the surface determining the bound state energies in the conduction band and their associated incident current densities. Furthermore, such a model would be able to determine how electric field enhancement and the field factor at the apex of a silicon field emitter tip are influenced by electric field penetration into the silicon and whether these customary simplifications to nanoscale emitter electrostatics are good approximations at all. It would also be useful to develop a multidimensional model for the electrostatics of single-gated field emitters that is compatible with multidimensional field emission calculations, which would consist of a nanoscale emitter cathode located in the center of the circular aperture of a planar gate electrode. Such a model could assist experimentalists in designing better semiconductor field emitter arrays and in analyzing their performance.

Due to the assumption that the potential in the silicon near the emitter surface
only varies along the emission direction, the incident flux density impinging on the emitter surface was able to be calculated using a one-dimensional model, based on the action-angle coordinates of classical mechanics (see \S 3.3.1 and \S A.6). In the situation in which the electrostatics calculations are performed in multiple dimensions, such as in the case of a nanoscale silicon emitter tip, a multidimensional model for calculating the incident flux density of electrons inside the quantum well near the surface must also be utilized. A semiclassical approach to determining the incident current density of electrons at a particular point on the emitter surface would require the calculation and sum of the tunneling attempt frequencies of all classical orbits of electrons in the quantum well that end at that particular point on the emitter surface. As this would need to be repeated for all points on the emitter surface, such a calculation would be very computationally intensive. However, it is possible that good approximations which obviate the need for the full calculation can be found or that an alternative formulation that is more practical from a computational perspective can be developed.

Future models of emission from semiconductor emitters should also include emission from a band of surface states, whose energies are located in the forbidden energy gap. As it has been shown experimentally that the total energy distribution of electrons emitted from a semiconductor can be dominated by contributions from surface states at lower applied fields, it is possible that emission from surface states plays an important role in emission from semiconductor emitter tips.

Finally, and possibly most importantly, in order to model the effects of conduction band flux density supply depletion and conduction band emitted current density saturation more accurately, a current continuity equation must be incorporated into the model. A current continuity equation would establish the vital link between the current density emitted at the surface and the electronic flux density able to be supplied from the extended states far from the surface. Since the current continuity equation involves the electron density, it will also influence the potential calculated
via Poisson’s equation and the bound state energies calculated via the Schrödinger equation. As a result, these three equations must be solved self-consistently and would give a more accurate description of the potential and bound states near the emitter surface. Poisson-Schrödinger-Current solvers are available in existing device physics simulation platforms, but few grant the flexibility required to model both the traditional, drift- and diffusion-based current flow through a semiconductor device and the non-traditional, tunneling-based current flow due to field emission simultaneously. Consequently, an entirely new simulation platform based on self-consistent solutions to the coupled Poisson-Schrödinger-Current equations, tailored specifically for field emission calculations, should be developed in order to enable more realistic modeling of field emission from nanoscale silicon field emitters.
Bibliography


[28] José L Padilla, Cem Alper, Francisco Gamiz, and Adrian Mihai Ionescu. Switching behavior constraint in the heterogate electron–hole bilayer tunnel FET: The


Appendix A

Overview of Classical Mechanics

This appendix gives an overview of the major aspects of classical mechanics on which the semiclassical framework for field emission is built. Although the content in this appendix was sourced from the textbooks by Landau and Lifshitz [147] and Joos [148], it pertains to standard treatments of Lagrangian and Hamiltonian mechanics and can be found in most introductory textbooks on classical mechanics.

A.1 Hamilton’s Principle

Modern formulations of classical mechanics are based upon Hamilton’s principle, which states that the true evolution $\mathbf{q}(t)$ of a system is described by $N$ generalized coordinates $\mathbf{q}$ between two specified states $\mathbf{q}_1$ and $\mathbf{q}_2$ at two specified times $t_1$ and $t_2$ respectively, is a stationary point of the action functional

$$ S[\mathbf{q}] = \int_{t_1}^{t_2} L (\mathbf{q}(t), \dot{\mathbf{q}}(t), t) \, dt, \quad (A.1) $$

where $L (\mathbf{q}, \dot{\mathbf{q}}, t)$ is the Lagrangian function of the system and $\dot{\mathbf{q}}$ is the time derivative of the generalized coordinates $\mathbf{q}$. Alternatively expressed, Hamilton’s principle states
that the true evolution of a physical system is a solution of the functional equation

$$\frac{\delta S}{\delta q(t)} = 0.$$  \hspace{1cm} (A.2)

### A.2 Essentials of Lagrangian Mechanics

Lagrangian mechanics is a reformulation of classical mechanics developed by Joseph-Louis Lagrange in 1788. Although it introduces no new physics compared to Newtonian mechanics, Lagrangian mechanics facilitates solving classical mechanics problems by the use of generalized coordinates, which when conveniently chosen, allow for the exploitation of symmetries in the system or the geometry of the constraints. Unlike Newtonian mechanics, Lagrangian mechanics also reveals conserved quantities and their related symmetries directly. The non-relativistic Lagrangian function of the system is defined as the difference between the kinetic energy $T$ and the potential energy $V$ of the system:

$$L(q, \dot{q}, t) = T(\dot{q}) - V(q, t) = \frac{1}{2} \sum_{k=1}^{N} m_k \dot{q}_k^2 - V(q, t),$$  \hspace{1cm} (A.3)

where $k$ indexes the generalized coordinate $q_k$ and $m_k$ is a coefficient for the $k$th generalized coordinate. In rectangular coordinates, $m_k$ is simply the mass corresponding to each generalized coordinate.

Inserting the Lagrangian function of the system into Equation A.1, performing the variational calculation, and equating the result to zero as required by Hamilton’s principle in Equation A.2 yields the Euler-Lagrange equations for the variational problem:

$$\frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} = 0,$$  \hspace{1cm} (A.4)

which determine the true path of the system as a function of time $q(t)$. The Euler-
Lagrange equations for a system are a set of $N$, 2nd order ordinary differential equations, whose solutions give the position of the system in each coordinate $q_k$, as a function of time. These are usually referred to as the equations of motion of the system.

### A.3 Essentials of Hamiltonian Mechanics

Hamiltonian mechanics, first introduced by William Rowan Hamilton in 1833, is a reformulation of classical mechanics that uses a different mathematical formalism compared to both Newtonian and Lagrangian mechanics, which provides a deeper and more abstract understanding of the underlying theory. Outside of classical mechanics, Hamiltonian mechanics also finds wide utility as the basis of statistical mechanics and quantum mechanics. Hamilton originally started from the Lagrangian formulation of classical mechanics, and as such, the Hamiltonian function, the basis of Hamiltonian mechanics, can be derived directly from the Lagrangian function via the Legendre transform:

$$H(p, q, t) = \sum_k \dot{q}_k \frac{\partial L}{\partial \dot{q}_k} - L = \dot{q} \cdot p - L,$$

(A.5)

where $r = (p, q)$ are the set of canonical coordinates that describe the state of the system: $q$ are the generalized coordinates and $p$ are the generalized momenta. Inserting the expression for the Lagrangian function in terms of the Hamiltonian function into Equation A.1 and repeating the process of finding stationary solutions to the variational problem results in Hamilton’s equations:

$$\frac{dq}{dt} = + \frac{\partial H}{\partial q}$$

(A.6a)

$$\frac{dp}{dt} = - \frac{\partial H}{\partial q}$$

(A.6b)
which are a set of $2N$, 1st order ordinary differential equations, whose solutions give
the position and momentum of the system as a function of time and are equivalent
to the Euler-Lagrange equations.

### A.4 Canonical Transformations

For a variety of reasons, it is often desirable to change to a new set of coordinates
$(P, Q)$, as functions of the old set of coordinates $(p, q)$. After performing such a
coordinate transformation, the original Hamilton $H(p, q, t)$ also undergoes a trans-
formation to a new Hamiltonian $K(P, Q, t)$, in terms of the new coordinates. Trans-
formations that preserve the form of Hamilton’s equations are known as canonical
transformations. The requisite constraints that will produce such transformations
can be determined from inserting the Lagrangian as a function of the Hamiltonian
into Equation A.1 and performing the variational calculation:

$$\delta \int [p \cdot dq - H(p, q, t)] = \delta \int [P \cdot dQ - K(P, Q, t)] = 0. \quad (A.7)$$

The variation of the time-integral action functional produces the constraint that the
new Lagrangian and the the old Lagrangian must be equal to within the total time
derivative of an arbitrary function $G$. This is a result of the time integral over such
a term yielding only the function $G$ evaluated at the endpoints, which is a constant
term that has no impact on the variational problem. Canonical transformations are
those that satisfy the following condition:

$$p \cdot dq - H(p, q, t) = P \cdot dQ - K(P, Q, t) + \frac{dG}{dt} \quad (A.8)$$

where $G$ is the generating function of the transformation.

This is known as the generating function approach. There are four primary types
of generating functions, which are categorized by the combination of new and old coordinates on which they depend. The major results that follow in this appendix are based upon a generating function of the 2nd type, denoted by \( G_2 \), in which the generating function is a function of the new momenta \( P \) and the old coordinates \( q \):

\[
G(P, q, Q, t) = -Q \cdot P + G_2(P, q, t), \tag{A.9}
\]

where the \(-Q \cdot P\) term has been added to produce the proper Legendre transform on the right-hand side of Equation A.8. Substituting Equation A.9 into Equation A.8 and equating the pairs of terms proportional to \( \dot{q} \) and \( \dot{P} \) results in the equations that define the transformation:

\[
\begin{align*}
Q &= \frac{\partial G_2}{\partial P}, \tag{A.10a} \\
p &= \frac{\partial G_2}{\partial q}, \tag{A.10b} \\
K(P, Q, t) &= H(p, q, t) + \frac{\partial G_2}{\partial t}. \tag{A.10c}
\end{align*}
\]

Hamilton’s equations in terms of the new coordinates and Hamiltonian are:

\[
\begin{align*}
\dot{Q} &= \frac{\partial K}{\partial P}, \tag{A.11a} \\
\dot{P} &= -\frac{\partial K}{\partial Q}. \tag{A.11b}
\end{align*}
\]

### A.5 Classical Hamilton-Jacobi Theory

When the time derivatives of the new momenta \( P \) and new coordinates \( Q \) are zero, they are called constants of motion, as they do not change with time as the system evolves. Sometimes it is convenient to transform coordinates such that both \( P \) and \( Q \) are constants and, according to Hamilton’s equations, this is achieved by making
the transformed Hamiltonian $K$ zero. Selecting $G_2(q, t)$ such that it is equal to Hamilton’s principal function $S(q, t)$ plus a constant $A$ gives the expressions:

\[ P = \alpha, \quad (A.12a) \]

\[ Q = \beta, \quad (A.12b) \]

\[ H(p, q, t) + \frac{\partial S}{\partial t} = 0, \quad (A.12c) \]

where $\alpha$ and $\beta$ are the designations for the new, constant generalized momenta and coordinates. Using the transformation relation from Equation A.10b allows for the expression of $p$ in terms of partial derivatives of $S(q, t)$ with respect to $q$, giving the Hamilton-Jacobi equation (HJE):

\[ H \left( \frac{\partial S}{\partial q}, q, t \right) + \frac{\partial S}{\partial t} = 0. \quad (A.13) \]

**A.5.1 Maupertuis’ Principle**

If the Hamiltonian is time-independent such that $H(p, q) = E$, the HJE becomes

\[ E + \frac{\partial S}{\partial t} = 0, \quad (A.14) \]

allowing for the immediate solution of the time-dependent part of $S$, such that $S(q, t) = W(q) - Et$, where the function $W(q)$ is known as Hamilton’s characteristic function. Incidentally, Hamilton’s principal function $S(q, t)$ is the classical action of the system up to an arbitrary constant. Using this knowledge, it can be shown that:

\[ W(q) = S(q, t) + Et = \int [L(q, \dot{q}) + H(p, q)] dt = \int p \cdot dq. \quad (A.15) \]
where the last step has used the Legendre transform relation between the Lagrangian function and Hamiltonian function of the system and $p$ are the conjugate momenta of the generalized coordinates, defined as $p = \partial L / \partial \dot{q}$. The integral on the far right of Equation A.15 is a functional that takes the path followed by the physical system without regard for its parameterization by time ($dq$) as the input and is known as the *abbreviated action functional*. The abbreviated action functional is the central element of *Maupertuis’ Principle*, which states that the true path of a system described by $N$ generalized coordinates $q$ between two specified states $q_1$ and $q_2$, is an extremum of the abbreviated action functional $W(q)$. Thus, Maupertuis’ principle is a time-independent, special case of Hamilton’s principle in which the extremization of $W(q)$ gives the path of the system only, providing no information about its state at any time $t$. Unlike Hamilton’s principle, each of the paths are constrained to conserve total energy $E$, as required by the time-independent Hamiltonian.

### A.5.2 Separable Solutions to the Hamilton-Jacobi Equation

Expressing the Hamiltonian explicitly in terms of partial derivatives of $S(q, t)$ with respect to $q$,

$$
H \left( q, \frac{\partial S}{\partial q} \right) = \frac{1}{2m} \sum_k \frac{1}{h_k^2} \left( \frac{\partial S}{\partial q_k} \right)^2 + V(q) = \frac{(\nabla S)^2}{2m} + V(q), \quad (A.16)
$$

where $h_k$ are the scale factors for each of the coordinates $q_k$ and $V(q)$ is the potential. In the time-independent case, this gives the following form for the Hamilton-Jacobi equation:

$$
\frac{(\nabla S)^2}{2m} + V(q) = E. \quad (A.17)
$$

In some cases, the system is fully solvable via the method of “separation of variables”. If some coordinate $q_1$ and the derivative of $S$ with respect to $q_1$ appear together only
in some combination

\[ \phi \left( q_1, \frac{\partial S}{\partial q_1} \right), \quad (A.18) \]

which does not involve any of the other coordinates, the HJE can be written in the form

\[ \Phi \left\{ q_k, \frac{\partial S}{\partial q_k}, \phi \left( q_1, \frac{\partial S}{\partial q_1} \right), E \right\} = 0, \quad (A.19) \]

where \( q_k \) includes all coordinates except \( q_1 \). As the next step, a solution in the form of a sum is sought:

\[ S(q, t) = S'(q_k, t) + S_1(q_1), \quad (A.20) \]

such that when inserted into Equation A.19, the resulting equation is

\[ \Phi \left\{ q_k, \frac{\partial S'}{\partial q_k}, \phi \left( q_1, \frac{\partial S_1}{\partial q_1} \right), E \right\} = 0. \quad (A.21) \]

If the solution to Equation A.20 is indeed found and then inserted into Equation A.21, it is clear that if Equation A.21 holds for any value of \( \phi \), which is only a function of \( q_1 \) and \( \partial S_1/\partial q_1 \), then \( \phi \) must be equal to a constant:

\[ \phi \left( q_1, \frac{\partial S}{\partial q_1} \right) = a_1, \quad (A.22a) \]

\[ \Phi \left\{ q_k, \frac{\partial S'}{\partial q_k}, a_1, E \right\} = 0, \quad (A.22b) \]

where \( a_1 \) is an arbitrary constant. If this process is successfully repeated for each coordinate, a full solution can be found that is a sum of the solutions in each of the coordinates:

\[ S(q, t) = \sum_{k=1}^{N} S_k(q_k; a) - E(a)t, \quad (A.23) \]

where \( a \) is the collection of all arbitrary constants \( a_k \). In the context of classical mechanics, these \( a \) are constants of motion, which correspond to a symmetry of the
system under some continuous transformation and imply a conservation law.

A.6 Action-Angle Coordinates

Given a time-independent Hamiltonian, there are other canonical transformations whose time-independent canonical coordinates provide useful information about the evolution of a classical system, without having to solve the equations of motion explicitly. One such piece of information is the frequency at which the system completes an orbit, which is readily obtainable by transforming to action-angle coordinates. Additionally, action-angle coordinates are integral to the Bohr-Sommerfeld quantization condition, also known as the “old quantum condition”, which was used to calculate quantum bound state energies early in the development of quantum mechanics.

A.6.1 Classical Frequency of Oscillation

Action-angle coordinates result from a canonical transformation, using Hamilton’s characteristic function $W(q)$ as the type 2 generating function. Since the Hamiltonian has no explicit time dependence, the new Hamiltonian $K$ is simply the old Hamiltonian $H$, expressed in terms of the new canonical coordinates. These new coordinates are chosen to be the action-angles $w$ and the conjugate momenta $J$, also called the action variables.

Instead of defining the action-angles directly, it’s more convenient to first define the action variables,

$$J_k = \oint p_k dq_k,$$

where the integral is evaluated over the course of an entire period of oscillation of the system and resembles the abbreviated action in each original generalized coordinate $q_k$. Since the motion of the system is not involved in this integral, the action variables $J_k$ are constants of the motion, implying that the new Hamiltonian $K$ does not depend
on the action-angle coordinates $w$, via $dJ_k/dt = \partial K/\partial w_k = 0$, but only on $J_k$. The definition of the action-angles is given in terms of the type-2 generating function $W(q)$:

$$ w_k = \frac{\partial W}{\partial J_k}. \quad (A.25) $$

The dynamics of the action-angles, most notably the frequency of oscillation or rotation along the path for the original coordinates $q_k$, are governed by Hamilton’s equations, yielding:

$$ \nu_k(J_k) = \dot{w}_k = \frac{\partial K}{\partial J_k}. \quad (A.26) $$

It is not always possible to express the new Hamiltonian $K$ in terms of the action variables $J_k$ so that the differentiation in Equation A.26 is readily evaluated. However, since the new and old Hamiltonians are both time independent and equal to $E$, if the action variables $J_k$ can be expressed in terms of the total energy, $\nu_k$ can be determined by reciprocating the relation in Equation A.26:

$$ \nu_k(J_k) = \left[ \frac{\partial J_k(E)}{\partial E} \right]^{-1}. \quad (A.27) $$

### A.6.2 Relation to the Bohr-Sommerfeld Quantization Condition

The Bohr-Sommerfeld quantization condition was the main tool of old quantum theory and was primarily used to determine the energies of quantized states. It was originally known that the difference between discrete energy levels is equal to Planck’s constant $\hbar$ times the frequency of the emitted light $\nu$. In the classical limit $\nu$ is a function of energy $E$, and in order for the correspondence principle to hold, the frequency of the emitted light and the classical definition must agree in the limit of large quantum numbers, such that

$$ \Delta E = \hbar \nu(E). \quad (A.28) $$
A direct connection between classical mechanics and the old quantum energies can be made by defining the function [129]:

\[
J_n = \sum_{n'}^{n} \frac{\Delta E_n}{\nu(E_n)} + J_{n'},
\]

(A.29)

where \(n'\) is a number large enough for the calculation to be done in the classical regime. According to Equation A.28, \(J_n\) must change by \(h\) when \(n\) changes by unity, and in the classical limit, both \(J_n\) and \(\nu(E_n)\) undergo only very small changes with this variation in \(n\). Under these circumstances, the small changes can be expressed in the form of a differential and the sum can be converted to an integral:

\[
J(E) = \int_{E_0}^{E} \frac{dE}{\nu(E)} + J(E_0).
\]

(A.30)

Differentiating Equation A.30 with respect to energy gives

\[
\frac{dJ}{dE} = \frac{1}{\nu(E)},
\]

(A.31)

which is the same definition of the frequency as in Equation A.27. Thus, in quantum mechanics, \(J(E)\) is a function that can only take discrete values that differ by \(h\), while in classical mechanics, it can take on continuous values and is precisely the action variable \(J\). Expressing Equation A.30 in terms of \(p\) and \(q\) via the definition of the \(\nu\) in Equation A.26 and \(J_k\) in Equation A.24 gives

\[
J(E) = \int_{E_0}^{E} \frac{\partial J_k}{\partial E} dE = \oint p_k dq_k.
\]

(A.32)
Restricting Equation A.32 to the discrete quantum values determined by the small \( n \) limit of Equation A.29 yields the Bohr-Sommerfeld quantization condition:

\[
\oint p_k dq_k = n_k \hbar, \tag{A.33}
\]

which is expected to give a good estimate of the energy when \( \hbar (\partial \nu / \partial E) \ll 1 \).