100 years of the physics of diodes

Peng Zhang, Ágúst Valfells, L. K. Ang, J. W. Luginsland, and Y. Y. Lau

Citation: Applied Physics Reviews 4, 011304 (2017); doi: 10.1063/1.4978231
View online: http://dx.doi.org/10.1063/1.4978231
View Table of Contents: http://aip.scitation.org/toc/are/4/1
Published by the American Institute of Physics
100 years of the physics of diodes

Peng Zhang,¹,a) Ágúst Valfells,² L. K. Ang,³ J. W. Luginsland,⁴ and Y. Y. Lau¹

¹Department of Nuclear Engineering and Radiological Sciences, University of Michigan, Ann Arbor, Michigan 48109-2104, USA
²School of Science and Engineering, Reykjavík University, Menntavegi 1, 101 Reykjavík, Iceland
³Engineering Product Development, Singapore University of Technology and Design, Singapore, Singapore 487372
⁴Air Force Office of Scientific Research, Arlington, Virginia 22203, USA

(Received 29 September 2016; accepted 29 November 2016; published online 17 March 2017)

The Child–Langmuir Law (CL), discovered a century ago, gives the maximum current that can be transported across a planar diode in the steady state. As a quintessential example of the impact of space charge shielding near a charged surface, it is central to the studies of high current diodes, such as high power microwave sources, vacuum microelectronics, electron and ion sources, and high current drivers used in high energy density physics experiments. CL remains a touchstone of fundamental sheath physics, including contemporary studies of nanoscale quantum diodes and nano gap based plasmonic devices. Its solid state analog is the Mott–Gurney law, governing the maximum charge injection in solids, such as organic materials and other dielectrics, which is important to energy devices, such as solar cells and light emitting diodes. This paper reviews the important advances in the physics of diodes since the discovery of CL, including virtual cathode formation and extension of CL to multiple dimensions, to the quantum regime, and to ultrafast processes. We review the influence of magnetic fields, multiple species in bipolar flow, electromagnetic and time dependent effects in both short pulse and high frequency THz limits, and single electron regimes. Transitions from various emission mechanisms (thermionic-, field-, and photoemission) to the space charge limited state (CL) will be addressed, especially highlighting the important simulation and experimental developments in selected contemporary areas of study. We stress the fundamental physical links between the physics of beams to limiting currents in other areas, such as low temperature plasmas, laser plasmas, and space propulsion.

[http://dx.doi.org/10.1063/1.4978231]

TABLE OF CONTENTS

I. INTRODUCTION AND HISTORY ........................................ 2
   A. The Child–Langmuir law ........................................ 2
   B. Transit time view ............................................. 2
   C. Effects of nonzero initial velocity .......................... 3
   D. Induced current in a gap—The Ramo–Shockley theorem .... 4
   E. The Pierce diode ............................................. 4

II. THE ROLE OF SURFACE ELECTRIC FIELD ..................... 5
   A. Multidimensional CL law assuming uniform emission .......... 5
   B. Wing structures and protrusive surface ........................ 5
   C. General scaling with cathode vacuum electric field .......... 6
   D. Extension of multidimensional Mott–Gurney law ............. 8

III. TIME DEPENDENT PHYSICS ..................................... 9

IV. QUANTUM EXTENSION OF CL LAW ............................ 12
   A. 1D mean field model ..................................... 12
   B. Quantum SCL scaling ...................................... 12
   C. General scaling for quantum tunneling current ............. 14

V. TRANSITION FROM EMISSION TO CL CURRENT .............. 15
   A. Thermionic-, field-, photo-, and ferroelectric emission ...... 15
   B. Space charge limited emission below the Child–Langmuir limit. 16
   C. Emission and space charge in a microscale plasma gap .......... 18
   D. Computational issues ................................... 19

VI. DIODE PHYSICS IN REAL WORLD ............................ 19
   A. High power microwave (HPM) and x-ray sources .................. 19

a)Present Address: Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan 48824-1226, USA.
I. INTRODUCTION AND HISTORY

A. The Child–Langmuir law

The Child–Langmuir (CL) law\(^{1,2}\) is a statement on the maximum steady state current that can be transported in a gap, subject to the constraint imposed by the Poisson equation

\[ \nabla^2 \phi = \frac{\rho}{\varepsilon_0}, \]  

(1)

where \( \phi \) is the potential, \( \rho \) is the charge density, and \( \varepsilon_0 \) is the free space permittivity. For a single species, usually electrons, the current density is \( J = \rho v \), where \( v \) is the macroscopic velocity associated with the charge density \( \rho \). We may use the gap voltage, \( V_g \), and the gap spacing, \( D \), as the basic scales of the potential and distance, since these parameters are externally adjustable. From these two scales, a velocity scale \( v_l = \sqrt{eV_g/m} \) may be constructed, and the charge density scale \( \rho_s = \varepsilon_0 V_g/D^2 \) follows from the dimensions of the Poisson equation, (1). This leads to the current density scale \( J_s = \rho_s v_s = \varepsilon_0 \sqrt{e/m} V_g^{3/2}/D^2 \), expressed also in terms of \( V_g \) and \( D \). Multiplying this current density scale by a numerical constant of the order unity, we obtain the classical CL law for a one dimensional (1D) planar gap

\[ J_{CL} = 4\sqrt{\frac{2}{9}} \frac{\varepsilon_0}{\varepsilon_0} \frac{V_g^{3/2}}{m D^2}. \]  

(2)

Equation (2) gives the maximum steady state current density in a 1D, planar, vacuum gap of gap separation \( D \) and gap voltage \( V_g \), zero transverse magnetic field, assuming that the single-species charged particles are emitted with zero initial velocity, whose dynamics is governed by nonrelativistic, classical (non-quantum) mechanics under the electrostatic approximation. The italicized words represent idealization. Generalization on any one of the italicized terms has required substantial research effort. For example, changing the italicized word “planar” to “cylindrical” or “spherical” invalidates Eq. (2); the corresponding limiting current densities in these 1D cylindrical or spherical diodes are only given numerically, in tabulated form by Langmuir and Blodgett,\(^{3,4}\) until most recently.\(^{3} \)

We shall return to this problem later in our review. Here we note that the numerical coefficient in Eq. (2) is obtained by solving the 1D Poisson equation, \( d^2\phi/dx^2 = \rho/\varepsilon_0 \) \( = J/\varepsilon_0 v \) \( = J/(\varepsilon_0 \sqrt{e\phi/2m}) \), subject to the boundary conditions: \( \phi = 0 \) and \( d\phi/dx = 0 \) at \( x = 0 \) (cathode) and \( \phi = V_g \) at \( x = D \) (anode). The boundary condition of zero electric field on the cathode surface is known as the space charge limited (SCL) condition, as shown in Fig. 1. It determines the maximum injected current in steady state. Physically, any further increase of injection current leads to a negative potential near the cathode which would prevent further release of electrons (with assumed zero initial velocity) into the diode, under the italicized assumptions listed in this paragraph. Note that this space charge limited current density, Eq. (2), is independent of material properties. Regardless of the cathode material or cathode temperature, Eq. (2) gives the bound imposed by the Poisson equation.

The above dimensional argument may also be applied to charge injection into a solid, and the current density scale so constructed is known as the Mott–Gurney (MG) law.\(^6\) For a solid block of length \( D \) across which a voltage \( V_g \) is imposed, the charge density scale is still governed by the Poisson equation, (1), i.e., \( \rho_s = \varepsilon_0 V_g/D^2 \). However, because of their frequent collisions within the solid, the injected charges acquire a drift velocity when exposed to the electric field, \( V_g/D \). This gives the velocity scale of the injected charge, \( v_s = \mu \times (V_g/D) \), where \( \mu \) is the electron mobility. The current density scale, \( J_s = \rho_s v_s \), multiplied by a numerical coefficient also of order unity, is the MG law

\[ J_{MF} = \frac{9}{8} \frac{\varepsilon_0}{\varepsilon_0} \frac{V_g^2}{m D^4}. \]  

(3)

It gives the maximum current density that can be injected into a solid gap.

B. Transit time view

A useful way to view CL and MG is through the transit time model, also known as the capacitance model.\(^7–13\) While the transit time model has been used to interpret the CL and MG law over 60 years ago,\(^7,8\) it is only recently recognized that the transit time model may also be judiciously applied to a short pulse diode, to the quantum regime, and to non-planar gaps.

The idea follows. The maximum charge that a gap can hold is of order \( CV_g \), where \( C \) is the gap capacitance and \( V_g \) is the gap voltage. This is a very reasonable assumption, for example, the total amount of space charge associated with the CL solution, Eq. (2), may be shown to be \((4/3)CV_g\) for the planar gap. The limiting current is then approximately given by \( CV_g/T \), where \( T \) is the transit time of an electron to cross the gap subject to the vacuum electric field. Again, for the planar gap, the current density so obtained, upon multiplying by the numerical factor 8/9, becomes Eq. (2), the CL law.
This capacitance model proves to be very useful in predicting the maximum current density that can be injected in a very short pulse, so short that the bunch length (in space) is much shorter than the anode-cathode (AK) gap spacing, i.e., the pulse length (in time) is much less than the electron’s transit time across the gap.\(^{12}\) If the maximum charge within the gap is bounded by \(CV_g\), the maximum current density, \(J\), in a short pulse of length \(\tau \ll T\), is then bounded according to \(J \tau A < CV_g\), where \(A\) is the area for the planar gap. For a short pulse in a planar gap with \(\tau \ll T\), the maximum current density in this short bunch is in fact \(CV_g/(\tau A) = (8/9) \times (T/\tau) \times J_{CL}\), which can be much higher than \(J_{CL}\) for a short pulse length, \(\tau\). But the total charge in this bunch is still bounded by \(CV_g\). If there are multiple bunches within the gap, the maximum total charge is still on the order of \(CV_g\).

The transit time model has recently been applied to a cylindrical or spherical diode.\(^5\) This leads to useful scaling laws for these nonplanar diodes. In the transit time model, the limiting current density roughly describes the transport, over a transit time, of the charge on the cathode surface of charge density \(e_0E_c\), where \(E_c\) is the cathode surface electric field in a vacuum diode. The approximate limiting current density for both cylindrical and spherical diode reads

\[
J_{LB}(app) \approx \frac{4\sqrt{2}}{9} \frac{E_c^3/2}{m \sqrt{D}},
\]

where \(D\) is the anode-cathode separation, irrespective of whether the cathode is inside or outside the anode. The emphasis is then shifted from the gap voltage \((V_g)\) to the vacuum electric field on the cathode, \(E_c\), in the description of the limiting current density. This explicitly draws the link between the free charge in the gap and the resulting bound image charge needed to satisfy the boundary conditions in a capacitance model. Equation (4) is the lowest order approximation to the Langmuir–Blodgett (LB) law.\(^3,4\)

The transit time model has also been recently applied to a quantum diode. We shall comment on it in Section IV.

C. Effects of nonzero initial velocity

The transit time model naturally raises related questions of the limiting current in grounded structures. Of great interest is the limiting current of a drifting electron beam propagating inside some metallic structure. Let us first examine the Child–Langmuir 1D diode where the electrons have some initial velocity. Electrons emitted from a hot cathode possess some initial thermal velocity which leads to a potential minimum in front of the cathode. The space charge associated with this potential minimum has occupied a central role in microwave tube literature, as it is considered to play an important role in noise reduction in linear tubes.\(^14\) On the other hand, the space charge layer in front of the cathode is inevitable in many crossed field devices, and it is thought to be a cause for the excessive noise in crossed-field amplifiers and in magnetrons.\(^14,15\) This noise problem in crossed field devices remains poorly understood to this date. We may add that the electrons’ initial thermal velocity on a thermionic cathode places a fundamental limit on the quality of the electron beam, which is crucial to the generation of coherent high power millimeter and submillimeter waves.\(^16,17\)

When the injected velocity \((v_{th})\) is assumed constant, and orthogonal to the cathode surface, the limiting current density in a 1D planar gap is given by \(J_{CL}\), where \(J_{CL}\) is given by Eq. (2), and \(\Gamma\) is the modification factor\(^18\)

\[
\Gamma = \left[ \left( 1 + \frac{E_m}{V_g} \right)^{1/2} + \left( \frac{E_m}{V_g} \right)^{1/2} \right]^3,
\]

where \(E_m = mv_{th}^2/2e\) is the beam voltage associated with the initial velocity. It is clear that \(\Gamma = 1\) when \(E_m = 0\), as expected. Of interest is when \(V_g = 0\), i.e., the diode is short circuited (also known as drift space) and a 1D non-neutral electron beam with a nonzero initial velocity is injected into it. In this limit, the limiting current density, \(\Gamma J_{CL}\), is still given by Eq. (2), except the factor \(V_g^{3/2}\) in Eq. (2) is now replaced by the factor \((4E_m)^{3/2}\). Again, the (beam voltage)\(^{3/2}\) dependence in the limiting current density is observed. When the limiting current is reached, the beam is still propagating forward (no reflection), but at a reduced velocity from the initial injection velocity due to space charge on the beam. There is no solution to the Poisson equation in the steady state when the injection current is further increased.

If there is a background of stationary positive ions which provide complete charge neutralization of the electron beam in the short circuit diode, there is no constraint on the DC beam injection current from the Poisson equation, (1). That is, the trivial solution, \(\phi = 0\), and \(\rho = \rho\) is, \(\rho\) is, the only steady state solution to Eq. (1) when the beam is completely charge neutralized and the diode is short circuited. However, if one considers the small signal space charge waves on the neutralized electron flow in this short circuit diode, an instability arises (known as the Pierce instability\(^19\)) when the current density in this flow exceeds the limiting current density in the corresponding non-neutral electron flow that is discussed in the preceding paragraph. Subsequent work on the Pierce instability may be found in Refs. 14 and 20–27.

In addition to immobile neutralizing charge, applying the space charge limited condition to both the cathode (for electrons) and the anode (for ions) leads to a stable, steady-state solution called bipolar flow. This solution has similar scaling of \(V^{3/2}\) and \(D^{-2}\) as non-neutral CL, but prescribes a fixed relationship between the current carried by the electrons and the ions based on their respective mass, as well as a factor of 1.8 in the total transmitted current.\(^28\)

Returning now to a non-neutral electron beam drifting inside a grounded metallic structure, another important case to consider is that of a pencil beam, or an annular beam, guided by an infinite axial magnetic field, and propagating with a constant speed inside a hollow, circular metallic pipe, which is grounded.\(^22,29,33\) Under the assumption that the solution is \(z\)-independent and \(\theta\)-independent, we again have a 1D problem, in the radial direction for the potential function, \(\phi(r)\). For a non-neutral beam, \(\phi(r)\) is depressed, and the beam’s velocity is reduced, because of the beam’s space charge. Again, the Poisson equation does not admit a
solution in the steady state if the beam current exceeds a certain value. When this limiting current is reached, the beam still propagates in the forward direction, i.e., electrons are not reflected, similar to Jaffe's observation in his derivation of Eq. (5). The limiting current has a different form, depending on whether the beam is relativistic or not.\textsuperscript{28,34} From the experience gathered so far, we may see that an annular beam may provide a higher current than a pencil beam. First, an annular beam has a larger cross sectional area than a pencil beam, and therefore it can hold more current (for the same cathode current density). More importantly, the capacitance between the annular beam and the wall can be much larger than the capacitance between the pencil beam and the wall, by making the radius of the annular beam large, and close to the wall. Since the total charge that can be held is proportional to the capacitance (in the capacitance model), such an annular beam allows propagation of the highest current, as far as the constraint by the Poisson equation is concerned. Alternatively, a virtual cathode is likely to occur at locations where the beam is furthest away from the grounded drift tube, because at these locations, the beam’s space charge potential depression is the strongest.

D. Induced current in a gap—The Ramo–Shockley theorem

A very important concept on electron motion in a gap is conveyed by the Ramo–Shockley theorem (RS).\textsuperscript{35,36} As the space charge moves, the induced charge on the surrounding conductors also moves, causing a current to flow through these conductors. RS provided a very elegant formulation of this induced current, which is characterized mainly by the \textit{vacuum electrostatic} potential solution among the conductors. The induced current on electrodes then becomes a direct indication of charge motion within a detector, and RS provides the foundation of radiation detection,\textsuperscript{37} a subject of extreme importance for homeland security, national defense, and radiation protection, etc.

The induced current is also fundamental to microwave devices. The motion of a bunch of charge in the interaction space induces a current on the walls of the surrounding electromagnetic structure. It is this AC current induced on the walls of electromagnetic structures that drives the load (useful radiation) that is coupled out in high power microwave and millimeter wave sources. RS has been widely used in models and simulations of vacuum electronic devices.\textsuperscript{38}

Several interesting questions arise. Is RS relativistically correct? Is RS applicable to semiconductors? RS clearly refers to capacitive systems, what is the analogue of RS in inductive systems? Only the second question has recently been answered in the affirmative for semiconductor detectors.\textsuperscript{39}

E. The Pierce diode

The 1D CL law ignores the effects of beam expansion due to the beam’s space charge. The lack of proper control of beam expansion has significantly delayed the development of linear beam devices. A major step was taken, again, by Pierce.\textsuperscript{40} Instead of the 1D model, Pierce constructed a particular solution for a 2D cathode surface. The cathode consists of three segments, as shown in Fig. 2. The center segment, Segment 1, is a flat slab of a finite width from which electrons are allowed to emit uniformly and rectilinearly. This 1D electron flow from Segment 1 is governed by the familiar 1D CL solution. The potential along the two edges (top and bottom) of this electron flow is known, which serves as the boundary conditions for Segments 2 and 3. Both Segments 2 and 3 are non-emitting, and at an angle of 66.7° relative to the top and to the bottom edge of Segment 1, respectively. This angle is chosen so that the exact solution of the Laplace equation may be constructed so as to satisfy the boundary conditions. This laminar beam solution does not require an external magnetic field. A similar approach of shaping a 2D spherical cathode is proposed for a laminar, radially converging (pencil) beam without an external magnetic field in the cathode region. It was this advance in electron gun design that greatly sped up the beam optics development in linear beam devices. In fact, the Pierce solution was still used in recent sheet beam klystron experiments at the Naval Research Laboratory.\textsuperscript{41} Recent work on CL in other 2D diodes is highlighted in Section II.

This paper is organized as follows. Section II reviews the role of the surface electric field at cathodes in diodes including higher dimensional effects. The time dependent physics of CL law is addressed in Section III. The extension of diode physics to the quantum regime is summarized in Section IV. Section V reviews the transitions from various emission mechanisms (thermionic-, field-, and photoemission) to the space charge limited state (CL), highlighting important simulation and experimental developments in selected contemporary areas of study. Examples of the application of diode physics are shown in Section VI, where we stress the fundamental physical links between the physics of beams to limiting currents in other areas, such as low temperature plasmas, laser plasmas, and space propulsion. Concluding remarks and outlook are given in Section VII.
II. THE ROLE OF SURFACE ELECTRIC FIELD

In the analyses of Child and Langmuir, the electric field was assumed to be zero at or near the surface of cathode. In practical emitters, especially in the case of field emission, one must apply sufficiently large electric fields at the cathode surface in order to obtain electron emission. The role of surface electric field at the cathode is thus critical in diodes. Barbour et al.\textsuperscript{42} considered the effects of space charge in field emitters and compared their theory with experiments. Spindt et al.\textsuperscript{43} studied surface electric field and current emission in thin film field emission cathodes with molybdenum cones. Anderson\textsuperscript{44} compared the injected current density as a function of electric fields with and without space charge effects. Extensive investigations were performed to extend the fundamental understanding of current emission limits beyond one dimensional systems. Shiffler et al.\textsuperscript{45–48} performed a series of experiments with a variety of cathode materials and micro geometries and found that roughly the same total currents can be drawn from cathodes of the same bulk area. Haworth et al.\textsuperscript{49,50} showed that extremely small changes in local electric field near the cathode edge can drastically alter the current density being drawn from that edge, e.g., the “field shaper” cathode.

A. Multidimensional CL law assuming uniform emission

The Child–Langmuir (CL) law is based on a one dimensional (1D) model, which assumes that the emitting area of the cathode is much larger than the gap spacing. Although there are available analytic descriptions of two dimensional (2D) beam transport\textsuperscript{40} as are theoretical investigations into the effect of 2D geometrical cathode surface features,\textsuperscript{51} the seemingly simple problem of 2D planar SCL emission remained unsolved until much later. In 1996, the classical CL law was extended to two dimensions (2D) by Luginsland et al.,\textsuperscript{52} considering uniform emission of electrons over a finite strip of width $W$ in a planar diode of gap separation $D$. Using PIC simulations, it was found that the 2D CL law can be fitted by

$$J_{CL}(2D) = J_{CL}(1D) \left(1 + \frac{D}{W} \right),$$

for $D/W > 0.1$, where $J_{CL}(1D)$ is given by the 1D CL law, Eq. (2). It was also found that Eq. (6) is not sensitive to the applied magnetic field along the beam propagation. This is due to the fact that the virtual cathode always forms near the center of the emitting strip in a sheath very close to the cathode, where the velocity of the electrons is too small for the magnetic field to produce a large force on the electrons in the sheath. The empirical 2D CL law, Eq. (6), was later analytically derived by Lau\textsuperscript{53} to be

$$J_{CL}(2D) \approx J_{CL}(1D) \left(1 + \frac{D}{W} \right),$$

by considering the condition for the onset of virtual cathode formation at the center when a finite patch of cathode surface is allowed to emit, with a uniform emission current density across the patch. Using the same procedure, the 2D CL law for the case where electron emission is restricted to a circular patch of radius $R$ on the cathode is found to be\textsuperscript{53}

$$J_{CL}(2D) = J_{CL}(1D) \left(1 + \frac{D}{4R} \right),$$

which is confirmed by simulation\textsuperscript{54} to give a very good approximation for $R/D > 0.5$.

Adopting similar approaches in Ref. 53, the multidimensional CL law was later extended by Koh et al.\textsuperscript{55} in the classical, weakly relativistic and quantum regime for uniform electron emission in planar and cylindrical gap configurations with finite emission energy. It is proposed that the enhancement of 3D CL law (in terms of 1D CL law) in all the three regimes can be written in a general form of

$$J_{CL}(3D)/J_{CL}(1D) = 1 + F \times G,$$

where $F$ is a dynamical parameter to measure the mean position of the electrons (in different operating regimes) and $G$ is a geometrical correction factor depending on the shape and size of the emitting patches on the cathode. Analytical solutions were constructed for various emitting patches, for example, in classical regime, $F = 1/4$ and $G = \pi/4$ for the emitting long strip, $G = \pi/2$ for the circular patch, and $G = 4\sqrt{2}/\pi$ for a square of length $L$. Formulas of $G$ for other geometries can be found in Ref. 55. This general formulation was found to agree well with 3D PIC simulation.

Besides planar geometry, the space charge limited current from a finite length emitter in cylindrical diodes was also studied by both numerical simulations\textsuperscript{56} and analytical theory.\textsuperscript{57} Similar to the planar 2D CL law, it is found that the 2D limiting current in cylindrical diodes in the units of classical Langmuir–Blodgett law\textsuperscript{3} monotonically increases as the width of the emitter decreases.

B. Wing structures and protrusive surface

It is important to note that the models discussed in Sec. II A assumed that the emitting current density is uniform along the flat cathode surface and that the emission area cannot be too small compared with the gap spacing. Thus they may be considered as 2D uniform emission models. The uniform model may not be valid if there is a large supply of electrons in the emission patch to drive the electric field to zero everywhere on that patch. Intuitively, we may expect the emitting current density to be higher at the edge of the emission patch, since there is no space charge in the vacuum region right outside the edge. These kinds of wing like structures have been studied in various 2D non-uniform models,\textsuperscript{58–61} where the supply of the electrons along the emission surface is assumed to be unlimited and non-uniform electron emission is governed by the local fields. Figure 3 shows typical wing like structures of the 2D emission current from a flat emitting patch, where the current density emitted at the cathode surface is plotted as function of position along the cathode for a sampling of emission strip.
FIG. 3. Wing like structures in 2D SCL emission from flat patches. Simulated current density (normalized to the analytic 1D CL value) emitted at cathode versus position (normalized to gap distance) for a variety of emission strip widths (W) and gap distances (D) in cm; each trace is labeled with its corresponding (W, D) value. In the simulation, a 1 kV potential difference is applied across the AK gap with the cathode being held at zero. A confining magnetic field (0.5 T) is applied normal to the electrode surface to simplify electron motion and maintain a sharp transition between the beam and vacuum regions. Reproduced with permission from R. J. Umstandt and J. W. Luginsland. Phys. Rev. Lett. 87, 145002 (2001). Copyright 2001 American Physical Society.

widths (W) and gap distances (D), obtained in 2D PIC simulations. These 2D simulations also revealed the interesting result that 80% of 1D Child–Langmuir current may be emitted from a cathode surface if only 20% of its surface is actively emitting, in the form of isolated small patches. It is found that the high current density “wings” at the beam edge become less important as W/D increases. As the emission strip narrows (W/D < 1), the current density at the center of the beam begins to rise above the 1D CL value, which agrees with the simpler uniform emission 2D models described in Sec. II A. Hegeler et al. observed the wing like structure experimentally and introduced techniques to eliminate this edge effect/beam halo, including cathode recessing and introduction of a floating electric field shaper. The edge emission effects were also studied in an axisymmetric cylindrical system with axial emission from a circular cathode by PIC simulations.

The wing like structure in 2D SCL emission from flat patches was later studied using analytical approximations by Rokhlenko and Lebowitz. For a flat anode and a cathode field emitter with a periodic set of smoothly shaped parallel ridges, Rokhlenko and Lebowitz calculated the electric field and current density distribution semi-analytically. It was found that in spite of the non-uniformity of the current density in the flow direction, the total current is very close to that obtained with a strong magnetic field in the same geometry and applied voltage, which simplifies the computation and makes it more stable.

The non-uniform emission discussed so far is from a finite emission patch located on a flat cathode surface, where the wing structure is caused by a discontinuity between the vacuum electric field outside the emission edge and the electric field inside the diode where space charge dominates the electric field. Most times, however, electron emission is from cathodes with non-flat surfaces, either due to microscopic surface roughness of the cathode, or from intentionally placed protrusions, such as sharp emission tips, to provide strong local field enhancement to facilitate electron emission. The non-uniform SCL emission from protrusive surfaces has been extensively studied experimentally. Different models have been developed to study the electric field enhancement near surface protrusions, including conformal mapping, point charge model, perturbation theory, and numerical simulations.

Recently, Zhu and Ang developed a self-consistent model to study the SCL current emission from a hyperboloid tip in an otherwise flat AK gap, by solving the Poisson equation and electron trajectories (through equations of motion) self consistently. The prolate spheroidal coordinates (u, v, φ) are shown in Fig. 4(a), left, where the geometry is defined. At the apex of the tip, we have u = 0. The emission area S [Fig. 4(a), right] is characterized by a parameter u0, which denotes the outer boundary of the emission area. The size of S may be calculated from u0, D, and R, where D is the tip-anode spacing and R is the radius of curvature at the tip. Figures 4(b) and 4(c) show the dependence of the normalized SCL current density J [in terms of the 1D CL law] for different R and u0, at D = 1 µm and Vg = 1 kV. It is clear that larger emission area S (or higher values of u0) will provide more space charge effects and thus smaller limiting current can be emitted [see Fig. 4(b)]. Similar to the wing structures in flat cathodes of finite emission area, higher current density is found near the edge of the emission area. However, due to the significant field enhancement at the tip apex, the localized high value of SCL current density at the apex is much more significant than the edge effect.

These 2D models have extended fundamental understanding of electron emission limits beyond previously described one dimensional systems both in terms of prediction accuracy as well as basic understanding of cathode physics. While these efforts have been focused primarily on basic research and development, spin offs of this work have already shown themselves in confirmation of 2D effects from experimental cathode tests as well as from actual high power microwave device results.

C. General scaling with cathode vacuum electric field

As stated in Sec. 1B, the capacitive model, or the transit time model, can be used to approximate the SCL current by \( I = Q/T \), where \( Q = CVg \) is the total bound surface charge on the cathode, \( C \) is the diode capacitance, and \( T \) is the transit time of an electron to cross the gap subjected only to the vacuum field. Because of its simplicity and accuracy, this capacitive or transit time model is of great utility to study a short pulse diode and more recently the quantum regime.

In fact, though only with ad hoc justification, it has long been noticed by electron gun and (cathode ray tube) CRT designers that the SCL beam creation is ruled only by the vacuum electric field on the cathode (without beam) \( E_0 \) and is based on the very simple scaling law:

\[
J_{SCL} = \frac{4}{9} \left( \frac{2e}{m} \right) \left( \frac{E_0^{3/2}}{D} \right),
\]

where \( D \) is an “equivalent diode spacing” as an empirical parameter to fit the experimental results. Note that the
scaling in Eq. (10) is very different from the CL law,\textsuperscript{1,2} although it gives identical current density to the CL law for 1D planar diode with $E_c = V_g/D$.

Recently, by considering the surface electric field on the cathode of the vacuum diode,\textsuperscript{3,5,7} the above scaling laws have been demonstrated to give accurate approximations to the classical SCL current in all three geometries: CL law for planar diodes, and LB laws for cylindrical and spherical diodes (Eq. (4)). Consider a sheet of charge leaving the cathode of a vacuum diode, with the cathode (anode) located at $R_c$ ($R_a$) and a gap separation of $D = |R_a - R_c|$, in either planar, cylindrical, or spherical geometry. If all the bound charge on the cathode leaves then naturally the electric field directly in front of the cathode becomes zero, and the space charge limited condition applies. This charge sheet will then be accelerated by the vacuum field, $E_c$, that is set up by the anode voltage, at least initially. Applying the transit time argument to this charge sheet, as Valfells \textit{et al}. did on a charge sheet in a short pulse diode,\textsuperscript{12} it is straightforward to note that the major portion of the transit time is spent in the immediate neighborhood of the cathode surface, where the electrostatic potential may be approximated by

$$V(r) \approx \left| E_c (r - R_c) \right|.$$  

Using Eq. (11), the time of flight of an electron across the AK gap may then be given by\textsuperscript{7,13}

$$T = \int_{R_c}^{r_f} \frac{dr}{v(r)} = \sqrt{\frac{m}{2e}} \int_{R_c}^{r_f} \frac{dr}{\sqrt{V(r)}} \approx 2 \times \sqrt{\frac{m}{2e}} \times \sqrt{\frac{D}{E_c}}.$$  

In writing Eq. (12), we have used the energy conservation relation, $mv^2/2 = eV$. The transit time model yields

$$J_{\text{transit-time}} = \frac{\varepsilon_0 E_c}{T},$$  

because the surface charge density on the cathode of a vacuum diode is equal to $\varepsilon_0 E_c$ by Gauss’s law. Using the approximate transit time from Eq. (12) in Eq. (13), we obtain the “universal” scaling for SCL current density

$$J_{\text{SCL}} \approx \frac{8}{9} \times \frac{\varepsilon_0 E_c}{T} = \frac{4}{9} \varepsilon_0 \sqrt{\frac{2eE_c^3}{mD}},$$  

where the factor of 8/9 is inserted so that Eq. (14) becomes identical to the CL law for a planar diode, Eq. (2). Note that Eq. (14) is identical to Eq. (4). The magnitude of the vacuum electric field $E(r)$ and its potential function $V(r)$ are summarized as

$$E(r) = E_C, \quad V(r) = E_CR_r, \quad V_g = E_C D, \quad \text{planar}$$
$$E(r) = E_C R_c/r, \quad V(r) = E_C R_c/\ln (R_c/R_a), \quad V_g = E_C R_c/\ln (R_a/R_c), \quad \text{cylindrical}$$
$$E(r) = E_C (R_c/r)^2, \quad V(r) = E_C R_c (R_a/r - 1), \quad V_g = E_C R_a R_c/D, \quad \text{spherical}$$

whether $R_c > R_a$ or $R_c < R_a$, with gap separation $D = |R_a - R_c|$. It is shown that the general scaling, Eq. (14), which depends only on the vacuum electric field on the cathode surface $E_c$, obtained from Eq. (15) and the gap separation $D$, gives a very good approximation to the exact LB solutions for both cylindrical and spherical diodes (with error $\leq 30\%$, for $0.1 < R_c/R_a < 500$).\textsuperscript{5} In the cylindrical and spherical cases, Eq. (14) is further corrected by multiplying the RHS with $(1 + F)$, where the correction factor $F$ (different from that in Eq. (9)) is empirically fitted to the exact LB solutions\textsuperscript{5} and the resulting scaling laws are accurate to within $5\%$ for $10^{-5} < R_c/R_a < 500$. Very recently, Greenwood \textit{et al}.\textsuperscript{20} extended these scaling laws to the relativistic regime, with gap voltage up to 10 MV. They also provided highly
accurate numerical algorithms which can be used to benchmark 1D, 2D, and 3D emission algorithms.

This dominance of the vacuum field in the immediate vicinity of the cathode surface may have some implication in the contemporary development of electron gun codes, where modeling of electron emission in the first numerical grid proves most critical.\(^{85}\)

**D. Extension of multidimensional Mott–Gurney law**

As stated in Sec. IA, the equivalent of the CL law for SCL current flow in a trap free solid is known as the Mott–Gurney (MG) law, which is also a 1D uniform classical model. As pointed out toward the end of this subsection, SCL transport is important to light emitting diodes (LEDs), organic solar cells and organic semiconductors, etc., and the technique we developed for SCL transport in solid was able to explain some of the experimental data in these areas. In 2007, the 2D MG law including the effects of trapping was similarly derived in the form of \(1 + F \times G.\)\(^{100}\) Here, compared with \(F = 1/4\) (for CL law), we have \(F = 1/(c + 2)\) for the MG law, where \(c \geq 2\) is a parameter to account for the traps. At \(c = 1\) (trap free case), we have \(F = 1/3.\) The agreements of these scalings have been compared using a device simulator. The effects of Schottky contact between the cathode and solid were later included in a separate paper,\(^{101}\) which indicated that the same scaling remains valid at higher voltage to reach the SCL condition. Experimentally, the enhancement of the 1D MG law was observed in a nano rod (GaN) with an emitting area much smaller than the gap spacing that was verified using a separate model.\(^{102}\)

A model for non-uniform SCL current injection into a nano contact solid has been developed recently,\(^{103}\) similar to the non-uniform CL law.\(^{59}\) Figure 5 shows (a) the enhanced emission near to the edge and (b) overall enhancement for a finite emission width of \(W/D = 0.05\) to 2.5 for two different solids with \(e_r = 3\) and 9. This comparison indicates that the non-uniform edge emission that is observed in the 2D CL configuration\(^{58}\) is likely to be found in the SCL current transport in a dielectric solid and the enhancement factor will depend on the dielectric constant \(e_r\) of the material.

It is interesting to note that very similar non-uniform “wing like” current density distributions (Figs. 3 and 5(a)) are also observed at the contact interfaces between conductors with different cross sections,\(^{104–108}\) where the current non-uniformity is solely due to the current crowding effects near the constriction corners.

In a recent paper,\(^{109}\) an experiment was performed to understand the physical mechanism of the nitrogen incorporation in the high k dielectric hafnium oxide (HfO\(_2\))—5.1 nm, stacked with SiO\(_x\) —1.6 nm. Figure 6 shows the I–V characteristic measured by using the scanning tunneling microscopy (STM) for electron injection across the solid (5.1 + 1.6 = 6.7 nm) with a vacuum gap of 0.5 nm between the STM tip and the sample. In the high voltage regime, it is found that the transport is claimed to be space charge limited by fitting with a power law of \(I = V^n\) for \(n = 1.65.\) However, according to the MG law [Eq. (3)], the value should be \(n = 2\) for a trap free solid. This problem was solved by a recent paper on SCL transport in a gap combined of vacuum and solid.\(^{110}\) Consider a hybrid gap of \(7.2\) nm combined of solid (6.7 nm) and vacuum (0.5 nm), the model is able to reproduce \(n = 1.65.\) This experiment and model show the interesting phenomenon that SCL current conduction in a combination of free space [CL law] and solid [MG law] can be realized.

---

**FIG. 5.** (a) The non-uniform profile of the normalized SCL current density \(J\) for a trap free solid at \(W/D = 0.5\) and 1 (fixed \(D = 1\) \(\mu m\)) for PPV films (\(e_r = 3\)) and GaN (\(e_r = 9\)). (b) The geometrical enhancement of the 2D non-uniform MG law over the 1D MG law as a function of \(W/D = 0.05\) to 2.5 at \(D = 1\) \(\mu m\). The cases of no interface effect mean the change of the dielectric constant at vacuum-solid interface is not considered and only the dielectric constant of the material is used. Reproduced with permission from Y. B. Zhu and L. K. Ang, Sci. Rep. \(\textbf{5}, 9173\) (2015). Copyright 2015 Nature Publishing Group.

**FIG. 6.** The current as a function of bias voltage of a hybrid gap of a HfO\(_2\)/SiO\(_x\) stack (6.7 nm) with a vacuum gap of 0.5 nm from the STM tip, from three ramp-voltage stress (RVS) measurements. Reproduced with permission from Ong et al., J. Appl. Phys. \(\textbf{104}, 64119\) (2008). Copyright 2008 AIP Publishing LLC.
SCL current transport plays an important role in many novel devices involving either inorganic or organic materials, especially in solar cells and light-emitting diodes (LEDs). Joung et al.\textsuperscript{113} studied SCL conduction in graphene oxide sheets, with applications in organic solar cells\textsuperscript{112} and photovoltaic devices.\textsuperscript{113} Torricelli et al.\textsuperscript{114} developed an analytical model for the SCL current in single-carrier organic LEDs. The model was validated with experimental data collected from different materials in a wide range of operating conditions. Arkhipov et al.\textsuperscript{115} formulated a unified model of hopping carrier injection and SCL current in disordered organic materials. It was found that a metal contact with the injection barrier as high as 1 eV can be still an Ohmic contact at low temperatures, which provides conditions for the hopping SCL current. Fluctuations in the SCL current in organic semiconductors was investigated by Carbone et al.\textsuperscript{116} Mihailetchi et al.\textsuperscript{117} examined the SCL photocurrent in semiconductors, which is important for the design of new materials in organic photovoltaic devices. Leonard\textsuperscript{118} studied thermoelectric efficiency in the SCL transport regime in semiconductors and proposed that nanowires are promising candidates to realize SCL transport for thermoelectrics. Sha et al.\textsuperscript{119} proposed the breaking of the space charge limit in organic solar cells by a novel plasmonic-electrical concept. The understanding of SCL transport in solids provides new insights to the development of novel nanoscale devices involving new materials.

III. TIME DEPENDENT PHYSICS

A. Transient behavior in a diode

When the injected current into a diode exceeds the steady state limiting current, a virtual cathode is formed, from which some of the electrons are reflected, while others continue their forward motion. When this happens, the diode can never regain its steady state behavior. The reason follows. When the steady state injection current exceeds the limiting value for the gap, it means that the injected charge exceeds the amount that the diode can hold (which is of order $CV^2$), and the diode rejects the excess electrons. The rejection occurs at the location where the diode has roughly the highest potential depression locally. As the electrons are rejected, the potential depression will be relieved momentarily. After some time, a strong potential depression may appear again, as the injected current provides a steady supply of new electrons, forming a virtual cathode at a later time. This is the reason why the virtual cathode, once formed, always exhibits oscillatory behavior, known as virtual cathode oscillation. The oscillation frequency is therefore of the order of the electron transit time.

Virtual cathode was first discovered by Birdsall and Bridges in their pioneering work on particle simulation.\textsuperscript{14} Virtual cathode oscillations associated with a high current beam have been proposed to drive a high power microwave source, known as the vircator.\textsuperscript{120,121} While simple in concept and straightforward in implementation, vircators are characterized by poor efficiency and poor spectral characteristics.

If the injected current has a finite, but short rise time, before it reaches the steady state, a virtual cathode may also form even if the steady state value is lower than the corresponding limited current according to the time independent theory. The underlying reason is that there is an inductive voltage, $L\frac{di}{dt}$, associated with the rise time of the current where $L$ the inductance in the beam-diode assembly. Note that the inductive voltage always opposes the gap voltage or the equivalent beam voltage (Lenz law). Thus, the injected beam effectively experiences a lower voltage, possibly forming a virtual cathode before the steady state current is reached. This effect was discovered by Luginsland et al.,\textsuperscript{10,11} who also found that this transiently induced virtual cathode oscillation persists, long after the steady state current is reached, in line with the description given in the preceding paragraph. This virtual cathode is due to an electromagnetic effect, since the electrostatic approximation excludes the inductive voltage. This virtual cathode oscillation is not due to shock excitation, and is not caused by the self-magnetic field of the beam either.

A high voltage diode, which allows a higher limiting current, is more prone to the transiently induced virtual cathode oscillation, simply because $L\frac{di}{dt}$ is large, especially for a fast rise time. This transient effect on an ultrashort pulse diode remains to be analyzed.

B. Short pulse effects

The assumption made in derivation of the CL law that any variation in the electron stream has died out, is not necessarily valid in practical systems. For instance, in a photoinjector the beam pulse may be much shorter than the transit time through the diode. An experiment done at the University of Maryland on combined thermionic and photoemission illustrated some of the peculiarities of electron emission when the pulse length is comparable to the transit time, or even shorter. One important result is the distinction between a critical current density, at which a virtual cathode forms, and a maximum limiting current density which determines the highest average current density that can be transmitted through the diode. The maximum current density can be significantly higher than the critical current density. In the paper analyzing this experiment,\textsuperscript{12} a law for a short pulse equivalent of the CL law is derived. In the simplest case of a very short pulse, a capacitive model is used to derive the critical current density and it is found to be $J_{crit} = \frac{eV}{\tau_pD}$ where $V$ is the gap voltage, $D$ is the gap spacing, and $\tau_p$ is the pulse length (See Sec. II C for the capacitive model). Another model presented in the paper is based on an equivalent diode approximation where the electron beam is injected into the gap with a fixed current density $J_0$ over a short pulse length $\tau_p$, which is smaller than the gap transit time ($T_{CL}$) of the CL law. From this model, the normalized critical current density is given by the equation

$$J_{crit}/J_{CL} = 2\left[1 - \sqrt{1 - 3X_{CL}^2/4}\right]/X_{CL}^3$$

where $X_{CL} = \tau_p/T_{CL}$ is the normalized pulse length and $J_{CL}$ is the classic CL current density. This expression yields the same result as the capacitive model in the short pulse limit,
and gives the classical CL law, when \( X_{CL} = 1 \). Figure 7 shows a comparison of the critical current as predicted by the capacitive model, the equivalent diode model and from PIC simulations. Note that the time dependent PIC model gives slightly higher charge density than the equivalent diode model, and that the current density does not settle down to the value predicted by the CL law until the pulse length is a multiple of the gap transit time. Also, note that the effect of inductive voltage was not included in the model and simulation. 10

By combining the above mentioned effects of inductive voltage 10 and short pulse, 12 a detailed 2D electromagnetic model was developed recently, 122 which is able to cover a wide range of parameters. A numerical scaling was also given to calculate the ratio between the 1D and 2D models of the electromagnetic CL law, which provides a better understanding of the SCL current transport in the electromagnetic regime for a short pulse electron beam with finite rise time (but ignoring the self-magnetic field). It is important to note that SCL current in a relativistic diode including the consistent self-magnetic field remains unsolved. 123 The short pulse model has also been extended to study pulsed electron transport in a drift space. 124, 125

It is possible to examine the short pulse CL law at ultrafast time scale if we use ultrafast laser induced electron emission from metals. 126–132 Liu and Ang have considered the transition from ultrafast laser induced emission to space charge limited current 133 in a one dimensional system. They used a time dependent model for non-equilibrium multiphoton emission from a flat metallic surface with space charge effects included, and compared it with a multiphoton emission model without space charge effects, a time dependent PIC model with space charge effects, and the equivalent diode model described in the preceding paragraph. This work indicates that for high laser field strength the space charge effect cannot be ignored, particularly for low work function cathodes, and that the equivalent diode model gives a fair measure of the total charge that can be extracted in a single pulse. The reader should bear in mind that the one dimensional model used in this analysis is substantially different from the geometry of a practical emitter based on a sharp tip. Integration of the short pulse model and the work on space charge limited emission from a tip, as described in Section II B, is an open problem.

If the length of the short pulse is reduced into femtosecond time scale, we need to include quantum effects as the pulse width of such a short electron bunch may be comparable to the electron de Broglie wavelength (near the cathode). Ang and Zhang found that the classical short pulse CL current density is increased by a large factor due to electron tunneling through the space charge field. 134 The discussion of such quantum effects will be presented in Sec. IV B below. At high voltage, larger than the electron rest mass in a classical diode, relativistic effects will suppress the enhancement of short pulse CL law, which is confirmed by particle-in-cell simulation. 134 In general, the enhancement of the short pulse CL law over the long pulse CL law is proportional to the inverse power of the normalized pulse length in all regimes (classical, quantum and relativistic), when the pulse length is much smaller than the gap transit time. 134 Thus, the scaling of the short pulse SCL current density is \( J \sim V_g, D^{-1}, \) and \( \tau^{-1} \), as compared with the classical steady state CL law with a scaling of \( J \sim V_g^{3/2} \) and \( D^{-2} \).

### C. Oscillatory behavior at the small scale

The maximum number of electrons present within a diode gap, \( N_e \), can be easily estimated by \( CV_g/e \), where \( C \) is the gap capacitance and \( V_g \) is the gap voltage. Hence, \( N_e \approx 55E_0A \), where \( E_0 \) is the vacuum electric field in MV/m and \( A \) is the area of the emitting region in \( \mu m^2 \). From this it can be seen that in some microscopic systems of interest the number of electrons may be in the tens to thousands, at which point discrete electron effects are important.

Given such a microscopic system with a sufficient number of electrons available for emission at the cathode, it was found in a 2010 paper by Pedersen et al. 135 that space charge induced modulation of the current in the THz frequency range can occur. In this case the pulse length of the beamlets is smaller than the gap transit time. Furthermore it has been found that the frequency of the modulation is purely a function of emitter area and applied field. 136 Conceivably, the frequency of the modulated current can be roughly set by choosing a suitable emitter size, and tuned by varying the applied electric field. Figure 8 shows how the frequency varies as a function of the electric field for different emitter area radii.

Simulations indicate that an array of such finite emitters can synchronize via Coulomb interaction, resulting in higher power, although the synchronized system will oscillate at a lower frequency than an individual emitter. 137

![Figure 7](image-url)  
**FIG. 7.** Normalized short pulse CL law as a function of normalized pulse length \( X_{CL} \). The solid line shows the equivalent diode model, the dashed line is from the capacitive model, and the symbols are from the 1D electrostatic PIC simulations. Reproduced with permission from Valfells et al., Phys. Plasmas 9, 2377 (2002), Copyright 2002 AIP Publishing LLC.
dependent spread of emission energy leads to the broadening of electron bunches. On the other hand, the internal space charge force of a bunch will cause it to spread out as well. Both effects lead to a smearing out of the current modulation at the anode. As a result, the cathode temperature, the magnitude of the applied vacuum electric field, and the gap spacing all have an influence on the signal to noise ratio of the modulated current. In a recent paper by Ilkov et al.,\textsuperscript{138} it was found that the most favorable signal to noise ratio is obtained when there is only one bunch present in the diode gap at a time. For room temperature operation, this corresponds to a sub-micron diode gap and an applied field on the order of 10 MV/m.

It should be kept in mind that an experiment to establish this THz frequency modulation has not yet been carried out. However, the molecular dynamics code used for the simulations predicting this behavior has reproduced experimentally verified behavior, e.g., the short pulse SCL current dynamics observed at the University of Maryland.\textsuperscript{12}

**D. Time dependent space charge limited transport**

As can be seen from discussion of short pulse effects above, there are situations where the SCL current density is instantaneously much higher than predicted by the 1D CL law. Naturally, the question arises whether it is possible to construct a temporal profile for the current in a diode that it will, on average, exceed the classical CL current density yet not be accompanied by virtual cathode formation. Recent work\textsuperscript{139} provided an estimate for an upper limit to the time averaged SCL current density in a one dimensional diode, \( J_{\text{max}} \), such that \( J_{\text{max}} \leq 2.45 J_{\text{CL}} \), where \( J_{\text{CL}} \) denotes the classical CL current density. Furthermore, in this work the conjecture was made that the time averaged current density in a diode, with vanishing emission velocity from the cathode and fixed anode voltage, could not exceed \( J_{\text{CL}} \). This conjecture was later amended, to avoid effects of charge discreteness in systems with few electrons, to be solely applied to a continuous stream of current in the diode, assuming no field reversal at the cathode.\textsuperscript{140}

Caflisch and Rosin\textsuperscript{141} used a Lagrangian approach to reformulate the problem of one dimensional flow in a planar diode. Among the results presented in their work is the novel discovery that, when particles enter the diode with a finite velocity and periodic flux, the time averaged current density may exceed the CL limiting current density, Eq. (5) as derived by Jaffé,\textsuperscript{18} even by as much as 17%. Rokhleenko has also used a Lagrangian approach to examine the possibility of generating a time averaged current density beyond the CL limit, in this case by applying a periodically varying anode potential.\textsuperscript{142} In his analysis it was assumed that electrons coming from the cathode had no initial velocity and that the electric field vanishes at the cathode. This makes it possible to set up an inverse problem where the current density is varied periodically and the corresponding time varying anode potential is calculated. Hence it is possible to calculate the average SCL current for situations where the characteristic time scale for changes in anode voltage are comparable to the transit time (as opposed to the simpler approximation of adiabatic change in the anode potential where the time varying current density is calculated directly from the CL equation). The results indicate that under those conditions the average SCL current density can be considerably higher than the current density calculated from the classical CL law using the time averaged anode voltage as \( V_{\text{a}} \) in Eq. (2).

Griswold and Fisch have recently addressed the meaning of the CL limit with regards to a time varying potential across the diode.\textsuperscript{143} They point out that the CL limit is derived for a constant voltage, and argue that comparison of current in a diode with the CL current should be based on the maximum diode voltage over the duration of the current, rather than the time averaged voltage, leading to a higher value of the CL current density. In light of this definition their conjecture about the CL current providing an upper limit for the time averaged current has yet to be disproven.

Using ultrafast lasers to excite localized electron emission from metallic nano tips,\textsuperscript{126,128,144} it is possible to obtain a few emitted electrons per pulse. In 2011, Zhu and Ang extended the classical 1D CL law to the Coulomb blockade (single to few electrons) regime, including the effect of single electron charging.\textsuperscript{145} It was found that there is a threshold of voltage, \( V_{\text{th}} \), equal to one half of the single electron charging potential in order for electron injection into the gap to be possible (assuming zero initial velocity and no barrier at the interface). For voltages in the range of 1 to 2 times this threshold value, there is only 1 electron in the diode gap per transit, and the time average single electron injected current can be higher than the 1D CL law as shown in Fig. 9. This result was later confirmed by a PIC simulation which verified that it is due to the boundary electric field imposed by the Coulomb Blockade effect.\textsuperscript{140} The planar model has also been lately extended to a cylindrical model\textsuperscript{5} to confirm the
new scaling proposed for the LB law stated in Sections 1B and II C.

Recently, Liu et al. proposed a charge sheet model to study the space charge limited charge density of consecutive electron pulses injected into a diode with uniform temporal pulse separation.\(^{146}\) For \(N\) identical consecutive pulses and uniform injection interval \(\Delta t\), the SCL charge density per pulse is

\[
\sigma_N = \sigma_1 \frac{1}{N} \left( \frac{qE_0 \Delta t}{D} \frac{(N-1)}{2} \right),
\]

where the single pulse SCL charge density \(\sigma_1 = -\epsilon_0 \frac{V_g}{\sqrt{2}}\), the applied electric field \(E_0 = -\frac{V_g}{\sqrt{2}}\), and \(D\) and \(V_g\) are the gap separation and gap voltage, respectively. Equation (17) can be used to determine the upper limit of the charge density per pulse to avoid distortion in the time interval between pulses. The critical value turns out to be about 10% of a constant that can be calculated by using Eq. (17) for an applied voltage of up to a few MeV (including relativistic effects). The model was also extended to study multi pulse electron beam propagation in a drift space.\(^{147}\) The model may be useful in the design of Smith–Purcell radiation,\(^{148-151}\) multiple pulse electron beams for time resolved electron microscopy,\(^{152}\) free electron lasers,\(^{153}\) or other applications with charge pulse trains over a wide range of parameters.

IV. QUANTUM EXTENSION OF CL LAW

A. 1D mean field model

When the gap spacing is decreased significantly into nanometer scale, comparable to the electron de Broglie wavelength, quantum effects will become important. In this quantum regime, due to electron tunneling, it will require a higher potential barrier to block the injected current from the cathode, and thus it is expected that the SCL current in the quantum regime will be higher than the 1D classical CL law.

In 1991, Lau et al.\(^{154}\) constructed a simple 1D mean field model solving the coupled Poisson and Schrodinger equations, which indicated that, indeed, the classical CL law can be exceeded by a large factor due to electron tunneling. The current was expressed in terms of perveance (see, e.g., Refs. 15, 28, 31, and 38). In the dimensionless form, the perveance is \(\lambda = 2J/E^{3/2}\), denoted as \(\lambda_q\) for quantum model and \(\lambda_c\) for classical model, where \(J\) is the normalized current and \(E = E/eV_g\) is the normalized emission energy where \(V_g = h^2/2emD^2\) and \(h\) is the Planck’s constant. Figure 10 shows the enhancement of SCL current as a function of normalized emission energy \(E\) at different values of normalized gap voltage, \(\phi_g = eV_g/E\). Figure 10 clearly shows that when the gap spacing \(D\) becomes large, i.e., \(E\) becomes large, the classical CL law is recovered.

B. Quantum SCL scaling

In 2003, the 1D quantum model was extended by Ang et al.\(^{155}\) to include the effect of electron exchange correlation, of different geometries (cylindrical and spherical diodes) and of quasi 2D finite emitting areas. Here, the electron exchange correlation effect was found to be important, when the electron energy is comparable to the Hartree energy (27.2 eV). It is also explicitly pointed out that the new quantum scaling of the CL law is \(J_{QCL} \propto \sqrt{V_g}/D^3\). In Fig. 11, the quantum CL law indicates that the voltage scaling \((V_g^n)\) changes from \(n = 3/2\) (classical regime) to \(n = 1/2\) when the voltage is decreased for a planar gap with \(D = 1\) to 100 nm (top to bottom). Here \(\lambda_q = J_{QCL}/J_{CL}\) is the normalized quantum CL law in terms of the 1D classical CL law, and the energy of the source electrons injected into the gap is at Fermi energy level: \(\varepsilon = E - E_F = 0\). From the figure, we also see that the transition occurs at smaller voltage if the electron’s exchange correlation effect is not included (dashed lines). For example, at \(D = 10\) nm, the transition is at around

\[\varepsilon = \frac{E - E_F}{E}\]

\[\lambda_q = \lambda_c\]
10 V (solid line, red) as compared to about less than 0.01 V if the exchange correlation effect is not included.

The new quantum scaling $J_{QCL} \propto \sqrt{V/x^4}$ can be explained by a simple dimensional analysis.\textsuperscript{156} From the Poisson equation, the electrical potential can be dimensionally expressed as $V \propto en^2/\varepsilon_0$, where $n$ is the electron density. Substituting $V$ into the time independent Schrodinger equation, the electron density scale is given by $n = A/\sqrt{x^4}$, where $A = \hbar^2\varepsilon_0/(2m^2)$. Based on $J_{QCL} = enV \propto eA\sqrt{2e/m} \propto \sqrt{V/x^4}$, the quantum scaling is obtained dimensionally, which is $J_{QCL} \propto \sqrt{V/x^4}$.

Due to quantum effects, other properties of the SCL current are also different from their classical values. For example, the transit time of the SCL flow is found to be different,\textsuperscript{157} in contrast to the classical solution. The classical value of the SCL is greatly enhanced due to the tunneling through the space charge field created by both the electrons and ions, and it is no longer equal to the classical limit which is 1.86 times that of the classical CL law.\textsuperscript{158}

To realize the quantum SCL current in a nano gap, there were also studies to model the transition from field emission to SCL current in a nano gap at different conditions.\textsuperscript{159} By assuming the field emission to be governed by the Fowler–Nordheim (FN) law,\textsuperscript{160} it was found that the emission current can go beyond the classical CL law to reach the quantum CL law for a nanometer gap in the high current regime. As the FN law may not be valid for a nano gap, the model was later extended by using a modified Thomas–Fermi approximated (TFA) quantum model to calculate the transport of current in a nano gap, where the image charge potential, space–charge field and exchange–correlation effects of the emitted tunneling current are determined quantum mechanically.\textsuperscript{161} In this TFA model, it was also confirmed that the proposed quantum CL law remains valid for the high current regime.

It is clear that quantum effects for SCL current will be important when the characteristic length scale is comparable to the electron de Broglie wavelength. For steady state SCL electron beam transport in a nanometer gap, the length scale is the size of the gap in nanometer scale. The quantum effect can still exist for a large gap if the electron beam is emitted as a pulse such that the pulse length of the electron beam is comparable to the de Broglie wavelength. In this limit, the pulse duration is at the ultrafast time scale, which is much smaller than the classical transit time. Due to quantum effects at this ultrafast time scale, the enhancement can be a factor of 2 even for a large gap [cm] with a pulse length of 50 fs.\textsuperscript{134}

In addition to the theoretical modeling, there are two experimental papers, which have reported the observation of the new voltage scaling $\sqrt{V}$ of the quantum CL law in a nanometer gap.\textsuperscript{162,163} In the experiments, the electrons were injected into the gap due to field emission. For the low current regime, the measurements were first fitted by using the Fowler–Nordheim (FN) law to obtain the field enhancement factor at the emitting current level. By increasing the voltage, the emission reached the SCL regime and the total current collected at the anode was plotted as a function of the gap voltage. In the experiment, Aluminum based electrodes (work function is 4.08 eV) with a gap spacing of 30, 50, and 70 nm were fabricated.\textsuperscript{163} In Fig. 12, the SCL current $I$ [in pA] as a function of gap voltage $V$ [in volts] is plotted for both 50 and 70 nm gap. The results indicated that the transition from $V^{3/2}$ (classical regime) to $V^{1/2}$ is around 10 V for the 50 nm gap. It was also found that the transition is around 70 to 90 nm in another experiment.\textsuperscript{162}

Subsequent theoretical works include a nonlinear model of a quantum diode in a dense Fermi magnetoplasma by taking into account the quantum statistical pressure law for the electrons, and the quantum Bohm potential that causes the electron tunneling.\textsuperscript{164} The property of shot noise of SCL current was evaluated over a wide range of gap spacing, gap voltage, and pulse length.\textsuperscript{165} Effects of the boundary condition for quantum CL law have also been studied.\textsuperscript{166,167} Some important issues such as influence of image charge potential (near the emitting surface) due to different electron exchange correlation, the influence of the atomic surface on work function, and the scattering events on the surface were also discussed.\textsuperscript{168} The quantum CL model developed has been applied to describe electron tunneling induced charge transfer plasmons\textsuperscript{169} between nearly touching nano metal particles in the area of quantum plasmonics.
C. General scaling for quantum tunneling current

The quantum diode models discussed so far did not consider the current tunneling from the anode to the cathode. In fact, when the gap spacing between the two electrodes is reduced to the nanometer or even sub-nanometer scale, electron tunneling from the anode (electrode with higher bias) to the cathode (electrode with lower bias) may become significant, when the applied bias voltage is low. Tunneling effects from the anode in junctions separated by thin insulating films have been studied extensively by Simmons in the 1960s, but are usually ignored in quantum diode studies.

Recently, a general scaling law for quantum tunneling current in nano- and sub-nanoscale metal-insulator (including vacuum)-metal (MIM) junctions has been developed, by self-consistently solving the coupled Schrödinger and Poisson equations to include the effects of space charge and exchange correlation potential. The self-consistent model covers all the three regimes for the J-V characteristics of a MIM diode: direct tunneling, field emission, and space charge limited regime. As shown in Fig. 13, the general, the presence of space charge inside the insulator reduces current transfer across the junction, whereas the exchange correlation potential promotes current transfer. It is shown that these effects may modify the current density by orders of magnitude from the widely used Simmons’ formula, which is only accurate for a limited parameter space (insulator thickness > 1 nm and barrier height > 3 eV) in the direct tunneling regime. As shown in Fig. 13, the self-consistent model recovers various scaling laws in limiting cases: Simmons’s formula in the direct tunneling regime, Fowler–Nordheim law in the field emission regime, and Quantum Child–Langmuir law in the SCL regime. Note that in the SCL regime, the self-consistent current approaches the quantum version of CL law, which exceeds the classical Child–Langmuir law because of quantum tunneling. Smooth transition between various regimes has been demonstrated (Fig. 13).

The proposed model reveals the general scaling for quantum tunneling current in a nano- and sub-nanoscale diode and its dependence on the bias voltage, the dimension and material properties of the diode tunneling junction. It can be applied to broad areas involving tunneling junctions, for example, quantum plasmonics, transition voltage spectroscopy (TVS), molecular electronics, and resistive switching.

Recently, quantum plasmon resonances have received significant attention, because of potential applications in nanoscale optoelectronics, single molecule sensing, and nonlinear optics. Tunneling current is thought to support the charge transfer plasmon (CTP) mode when two plasmonic resonators are placed sufficiently closely together. The CTP resonance energy is very sensitive to the shape and conductivity of the junction. Savage et al. introduced a novel, simple, and fast approach. Their quantum corrected model (QCM) incorporated quantum effects in a classical electrodynamic framework. The QCM models the junction by its local dielectric response that includes electron tunneling and tunneling resistivity at the gap, which has been extensively used to model quantum plasmonic junctions. It is clear that the accuracy of the QCM approach is determined by the calculation of the tunneling current and tunneling resistivity of the plasmonic junction. Haus et al. derived a set of linear and nonlinear quantum conductivities when an ac field is applied across a nanoscale MIM plasmonic gap. Wu et al. considered Fowler–Nordheim (FN) tunneling that occurs at high electric fields for plasmon resonances, including the effects of the space charge and exchange correlation potential inside the gap. This FN tunneling approach suggests that with a sufficiently high intensity irradiation, the CTP can be excited via FN tunneling even for relatively large nanoscale gaps. Tan et al. have experimentally observed quantum plasmon resonances at length scales in the range 0.4 to 1.3 nm across molecular tunnel junctions made of two plasmonic resonators bridged by self-assembled monolayers (SAMs). The tunnel barrier width and height are controlled by the properties of the molecules, so is the resonant energy of the quantum plasmon mode. Recently, Wu et al. comprehensively modeled CTP resonances across silver-molecule-silver junctions to estimate the terahertz conductance of molecules at near infrared frequencies. Their analysis established a one to one relationship between the conductivity of the SAM and the resonant energy of the CTP modes, as shown in Fig. 14. This approach provides the guidance to use plasmonic oscillations for measuring the THz conductance of single molecules at near infrared frequencies.

The development of more accurate models on the quantum tunneling current and the extension of quantum CL law provides not only better understanding of the underlying physics but also insights in the development of novel quantum devices.

V. TRANSITION FROM EMISSION TO CL CURRENT

A. Thermionic-, field-, photo-, and ferroelectric emission

There are a number of mechanisms for emitting charged particles into a diode gap. In this section, a brief overview of the most commonly applied methods for electron emission will be given. Mechanisms such as secondary electron emission, beta emission, etc., will be omitted from the discussion. The emission physics are characterized by two main elements, the density and energy distribution of electrons in the cathode material, and the potential barrier which impedes the electrons entry into the gap. Electrons are emitted if they are energetic enough to overcome the potential barrier, or if they propagate through it via quantum tunneling.

Thermionic emission, which has a long history of application in vacuum tubes, is based on changing the energy distribution of the electrons in the cathode material by elevating the temperature, thus increasing the number of electrons that are energetic enough to overcome the surface barrier. Thermionic emission was first described by Richardson in 1901, but put on a more secure theoretical footing in the 1920s. The Richardson–Dushman equation, describing thermionic emission is given by:

$$J_{RD} = A_R T^2 \exp (-\phi / kT)$$

where the constant, $A_R = \frac{4\pi e^4}{h^2}$ is derived theoretically, but is in practice often multiplied by a correction factor depending on the cathode material. The work function, $\phi$, must be modified in the presence of a surface electric field to account for the Schottky effect which describes how the electric field at the cathode surface, $E_c$, decreases the effective work function by $\Delta \phi = \sqrt{e^2E_c / 4\pi \sigma_0}$. Photoemission is similar to thermoelectric emission in the sense that it relies upon electrons in the cathode material being energetic enough to overcome the surface barrier. In the most basic form of photoemission, a single photon imparts energy to an electron in the surface to such a degree that the electron’s kinetic energy can overcome the work function. Although the photoelectric effect was famously explained by Einstein in 1905, the theoretical framework for photoemission was constructed by Fowler and DuBridge in the 1920s and 1930s. With the advent of extremely short laser pulses it became possible to shine very intense laser light onto a surface without damaging it. Hence, it is possible to have multiple photons excite an electron so as to produce emission. Multiphoton photoemission was studied by Bechtel et al. and a generalized form of the Fowler–DuBridge photoemission current developed. The generalized Fowler–Dubridge equation for the density of the photocurrent, $J_{FD}$, is the sum of partial current densities for n-photon photoemission:

$$J_{FD} = \sum_{n=0}^{\infty} J_n$$

where the partial current for n-photon photoemission is given by:

$$J_n = a_n A_R (e/h\nu)^n (1 - R_c)^{1/2} F_b F_s^n$$

(note that $n = 0$ represents thermionic emission). Here, $a_n$ is a constant describing the probability of a n-photon process occurring; $R_c$ is the reflectivity coefficient for the photon frequency, $\nu$; $I$ is the laser intensity; $\phi$ is the effective work function; and $F$ represents the Fowler function $F(w) = \int_0^w \ln(1 + \exp(-w + s)) ds$. This generalized form of the Fowler–Dubridge equation has been verified experimentally.

Field emission differs from thermionic emission and photoemission in that it is based on free electrons tunneling through the surface barrier rather than energetic electrons overcoming it. By applying a strong electric field at the cathode the surface barrier is lowered and narrowed, whereby electron tunneling through it becomes more likely. Fowler and Nordheim developed the theory for field emission in a seminal paper in 1928 which built on experiments by Millikan and the Sommerfeld theory of free electrons in metals. The original work of Fowler and Nordheim ignored the effects of image charge on the shape and width of the surface barrier, thereby underestimating the current density to a great degree, but Nordheim later included the Schottky barrier into the field emission model. Forbes has investigated the properties of Fowler–Nordheim type equations, generalizing it for barrier shape and correcting for effects of temperature, electron band structure and occupation states. He gives the general structure of FN type equations for a one dimensional cathode in the absence of space charge as:

$$J_{FN} = \lambda a \phi^{-1} F^2 \exp (-\nu F b \phi^{3/2} / F)$$

where $\lambda$ is a correction factor ranging from 0.01 to 10, $a$ and $b$ are the FN coefficients, which depend on the cathode material, $\nu F$ is a correction term for the barrier shape due to image charge effects, $\phi$ is the local work function, and $F$ is the local field at the cathode surface. Note that the wide range in $\lambda$ is almost exclusively due to electronic band structure effects which strongly influence the supply function. Another strong influence on the solution is the image charge factor, $\nu F$, which may
increase the current density by a factor of 100 compared with the calculated current density when image charge is neglected.

The treatment of the three emission mechanisms described above is such that they are generally uncoupled. Even though the Richardson–Dushman equation takes into account lowering of the surface barrier due to an applied field, it does not include tunneling. Similarly, the generalized Fowler–DuBridge equation does not take tunneling into account, even though field and temperature effects are included to a degree. In reality, the effects of photoelectric excitation, electron temperature and enhanced tunneling due to the surface field can all be at play at once. In the past fifteen years, Jensen and his collaborators have done considerable work to generalize the emission model\textsuperscript{87,207–209} which has yielded an useful analytic model for generalized treatment of thermal, field, and photoemission\textsuperscript{210} which is valid over a wide range of parameters, thus bridging the transition regions between the familiar, Richardson–Dushman, Fowler–Nordheim, and Fowler–Dubridge approximations. An example of Jensen’s general thermal field formulation is shown in Fig. 15, in comparison with the Fowler–Nordheim (field) current density and the Richardson (thermal) current density.

Another emission mechanism, that is somewhat different in nature from thermal- field- or photoemission, is ferroelectric emission.\textsuperscript{211,212} Physically, the difference lies in the fact that, for ferroelectric emission (FEE) the polarization state of the cathode material is altered, rather than the energetics of the electron population or the shape of the surface barrier. A change in spontaneous polarization, of the cathode material, away from equilibrium leads to an internal electrostatic field which drives the current, which ultimately screens the spontaneous polarization. Hence, FEE is inherently pulsed in nature.

A distinction is made between “weak” FEE, which can deliver current densities in the range of $10^{-12}$ to $10^{-7}$ A/cm$^2$, and “strong” FEE which can deliver current densities of 100 A/cm$^2$. Weak FEE can only occur in the ferromagnetic phase and is due to tunneling of charges that move to the surface to screen the polarization charge. Weak FEE can be induced by pyroelectric or piezoelectric effects as well as polarization switching by an applied voltage. Strong FEE is a plasma assisted emission process, similar to explosive field emission (discussed in Sec. VI A) in the sense that an electron avalanche at the cathode surface creates a local plasma. This plasma has a negligible work function, and thus acts as a highly efficient source of electrons to be injected into the diode. The electron avalanche leading to plasma formation is in the form of a surface flashover, due to a voltage pulse applied to the ferroelectric ceramic, which can be initiated by FEE or field emission from a triple point on the cathode. The latter mechanism for flashover initiation is, of course not confined to ferroelectric materials, but is common to all dielectric material. Both mechanisms deliver similar current densities, but the delay time from voltage application to plasma formation is longer for avalanches caused by polarization switching rather than field emission from the triple point.\textsuperscript{211}

Although the theoretical groundwork for studies of electron emission was laid one hundred years ago, it is still an area of active research where technological advances, e.g., in nanostructured materials and ultrafast lasers have opened up new regimes to study, sometimes with intriguing results. For example, Liang and Ang have developed a formula for thermoelectric emission from graphene\textsuperscript{213,214} that explains experiments done by Zhu et al.\textsuperscript{215} The current scales with temperature as $T^3$ rather than as $T^2$ as would be predicted by the Richardson–Dushman equation.\textsuperscript{213,214} Another vibrant area of research, discussed in Sec. V B, has been the influence of space charge on the emission mechanisms.

B. Space charge limited emission below the Child–Langmuir limit

Derivation of the CL law is independent of the emission mechanism, the underlying assumption being that there is a sufficient supply of electrons available to form a virtual cathode. In this case the diode current is limited by the dynamics of charged particle propagation in the gap. In many cases the current is source limited rather than space charge limited, e.g., when the thermionic current from a cathode is lower than the CL current.\textsuperscript{216} From Sec. VA it is apparent that the electric field at the surface can affect the current source by changing the shape of the surface barrier, particularly in the case of field emission. Thus space charge can also influence current in the diode by directly affecting the source term.

Barbour et al.\textsuperscript{42} developed a theory based on finding the equilibrium electric field, $E_0$, at the cathode such that the space charge from the field emitted current in a one dimensional diode exactly shielded the applied vacuum field to such a degree that the surface field be equal to $E_0$. This theory involves several approximations, but was suitable to explain experimental results.\textsuperscript{42} Lau et al. further analyzed the transition from Fowler–Nordheim (FN) emission to CL current in a planar diode, developing a set of universal curves in normalized parameters, so that the flow regime can be immediately identified given the gap voltage, gap spacing and FN coefficients of the cathode material.\textsuperscript{217} Figure 16 shows these curves, from which the transition from FN to CL is apparent. The normalized parameters used in the figure

---

**FIG. 15.** Comparison of the general thermal field current density with the Fowler–Nordheim (field) current density and the Richardson (thermal) current density. The current scales with temperature as $T^3$ rather than as $T^2$ as would be predicted by the Richardson–Dushman equation. Another vibrant area of research, discussed in Sec. V B, has been the influence of space charge on the emission mechanisms.
are the gap voltage, $V$, current density, $J$, and gap spacing, $D$. Forbes used a slightly different, but equivalent approach with normalized parameters to examine space charge effects on field emission. 218 An interesting aspect of Forbes’ work is its applicability to ion emission, e.g., from liquid metal sources, and that it can accommodate any form of emission equation that is dependent on the surface electric field.

Rokhlenko and his collaborators conducted further investigations into space charge effects on FN emission in a planar diode, 219 using three different types of relations between the current density and equilibrium electric field at the cathode to find the behavior of the I–V characteristics of the diode in the small, and large current limits. They also showed that approximating the relationship between current density and electric field with a FN type equation using a tri-dimensional, planar theory to a field emitter array (which has a three dimensional structure). From this it is found that, in a field emitter array with an emitter spacing of 1 micrometer, that is subsequently reduced due to space charge effects, and that it can accommodate any form of emission equation that is dependent on the surface electric field.

Since field emission most commonly occurs at protrusions (either incidental or designed) because of the local electric field enhancement, it is naturally of interest to examine space charge effects on field emission for such three dimensional structures. Recent work by Zhu and Ang examined how space charge affects field emission from a single hyperbolic tip (Fig. 4). By iteratively solving the Poisson equation and particle trajectories for a given potential they designed a convergent algorithm to calculate the charge density of electrons emitted from the tip, both for FN emission and also for the extreme space charge limit where the field at the surface vanishes. The results showed that space charge reduction of the surface field was greatest at those locations on the tip where the Laplace field (or vacuum field) was strongest, resulting in more uniform, and lower current density from the tip than would be expected from considering only the vacuum field. An interesting feature of the model was the possibility of having a non-uniform work function along the tip (which could be a result of crystal orientation), in which case the point of highest current density is not necessarily at the tip apex. The temperature dependent material properties of the cathodes were also found to have profound effects on current emission. 220, 221

In an array of field emitters, it is apparent that one field emitter may influence the electric field at the surface of an adjacent emitter, either due to the surface charge of the first emitter or because of the space charge of current being emitted from it. This is commonly called screening or shielding of the field emitter, and has been studied to considerable degree over the past fifteen years. 222–225 These studies are very useful for design of emitter arrays to optimize emitter spacing for maximum current density (not including space charge effects) and avoiding edge effects. However, the analysis of shielding has not properly taken into account the effect of space charge from one emitter in an array on its neighbors, and vice versa. This calls for a combined model of the three dimensional environment in the immediate neighborhood of the emitting protrusion and a model of the neighborhood of the entire emitter array. As mentioned above, in a field emitter array with an emitter spacing of characteristic length, $H$, space charge effects on FN emission should be treated in a three dimensional manner at an elevation below $H$ above the emitter, but may be treated as a one dimensional problem above that elevation. 219

The analyses described heretofore are time-independent, that the Poisson equation and FN equation are simultaneously solved for a constant electric field at the cathode. The previous works also assume continuous, steady charge emission. To take into account temporal effects, the discrete nature of charge, and the statistical flavor of the tunneling process underlying field emission, simulations have been applied. Three types of simulations will be addressed here: Particle-in-cell (PIC); molecular dynamics (MD); and sheet models (SM). Feng and Verboncoeur investigated the transition from the FN regime to the CL regime in a one dimensional planar diode using PIC simulations. 227 Their results showed the temporal behavior of the current at (a) low values of applied field, where the FN current is much lower than the CL current; (b) intermediate applied field, where the FN and CL currents are comparable; and (c) high values of the applied field, where FN is much larger than CL. At higher values of the applied field the dynamics are characterized by a large initial burst of current from the cathode into the empty gap, that is subsequently reduced due to space charge effects. The current settles to an equilibrium value after a few damped oscillations. Another interesting result is that, in the transition region (intermediate field) the transmitted current is markedly lower than that predicted by the FN or CL law, as shown in Fig. 17, although it asymptotically converges to the CL scaling at very high values of the applied electric field. It is interesting to note that no virtual cathode was formed during the transient, and that the steady state
solution shown in Fig. 17 is consistent with Fig. 16 in the transition from FN to CL law as the voltage (or applied electric field) is raised.

Torrason et al. reported on simulations based on molecular dynamics for field emission in a planar diode. In these simulations every free electron in the gap is accounted for, and emission of a single electron is a statistical event determined by the Fowler–Nordheim equation and the local electric field at the cathode surface. The model used in this work included an emitter area of finite dimension on an infinite cathode, as opposed to the infinite emitter area of Feng and Verboncoeur, but shows very similar results for large emitter areas. The MD work also shows, that in the transition region the space charge is important in reducing the surface field, but has a much smaller effect on the transit time and beam spreading. Although the PIC and MD methods have considerable merit as simulation methods, they are still limited in applicability for overall design of electron guns based on field emission, due to the wide ranging length scales involved, and the number of particles in the system being a special concern for the MD approach. A recent paper addresses the problem of constructing a simple sheet model for simulation of space charge effects on field emission, with the aim of incorporating that model into a PIC based beam optics code. Four model types are studied: three one dimensional planar models and one three dimensional model where a ring of charge is emitted from a hemispheric protrusion. In the planar models two different approaches are taken to emission modelling. The first is based on sheets being injected at uniform time intervals, but varying charge density. The second, which is derived from the statistical nature of FN emission, is based on the interval between sheets being injected governed by Poisson distribution, which has a characteristic parameter, $\lambda$, dependent on the surface field. The hemispheric model is used to describe space charge effects of single electron emission from a protrusion. The authors show how the planar models reproduce the results from analytic studies, as well as those of the PIC and MD simulations described above. A description is also given of how to match the hemispheric model to the one dimensional model.

It should be noted that, even though most of the analytic work regarding space charge effects on field emission has been based on the assumption of equilibrium, Rokhlenko and Lebowitz have used analytic methods to study transient behavior. Their work shows the same overall behavior seen in the simulations that were discussed in the preceding paragraphs (i.e., settling to equilibrium after a few damped oscillations), and also addresses issues of stability of field emission.

The transition from thermionic emission to SCL emission or CL law (steady state) had been shown in the pioneering works by Langmuir and others. In 1993, a paper was also reported to show the transition of thermionic emission from metals to SCL condition. In the experiment, YAG pumped laser of about 90 fs and $4.9 \times 10^{11}$ W/cm$^2$ was used to induce thermionic electron emission from various metallic surfaces (Silver, Tungsten, and Aluminum). The measured electron yield is compared with a space charge limited extension of the Richardson–Dushman equation for short time scale as shown in Fig. 18 below.

C. Emission and space charge in a microscale plasma gap

Gas breakdown phenomena from nanoscale to microscale are of fundamental importance in device miniaturization. In the development of microelectromechanical systems effects of single electron emission from a protrusion. The authors show how the planar models reproduce the results from analytic studies, as well as those of the PIC and MD simulations described above. A description is also given of how to match the hemispheric model to the one dimensional model.

It should be noted that, even though most of the analytic work regarding space charge effects on field emission has been based on the assumption of equilibrium, Rokhlenko and Lebowitz have used analytic methods to study transient behavior. Their work shows the same overall behavior seen in the simulations that were discussed in the preceding paragraphs (i.e., settling to equilibrium after a few damped oscillations), and also addresses issues of stability of field emission.

The transition from thermionic emission to SCL emission or CL law (steady state) had been shown in the pioneering works by Langmuir and others. In 1993, a paper was also reported to show the transition of thermionic emission from metals to SCL condition. In the experiment, YAG pumped laser of about 90 fs and $4.9 \times 10^{11}$ W/cm$^2$ was used to induce thermionic electron emission from various metallic surfaces (Silver, Tungsten, and Aluminum). The measured electron yield is compared with a space charge limited extension of the Richardson–Dushman equation for short time scale as shown in Fig. 18 below.

C. Emission and space charge in a microscale plasma gap

Gas breakdown phenomena from nanoscale to microscale are of fundamental importance in device miniaturization. In the development of microelectromechanical systems effects of single electron emission from a protrusion. The authors show how the planar models reproduce the results from analytic studies, as well as those of the PIC and MD simulations described above. A description is also given of how to match the hemispheric model to the one dimensional model.

It should be noted that, even though most of the analytic work regarding space charge effects on field emission has been based on the assumption of equilibrium, Rokhlenko and Lebowitz have used analytic methods to study transient behavior. Their work shows the same overall behavior seen in the simulations that were discussed in the preceding paragraphs (i.e., settling to equilibrium after a few damped oscillations), and also addresses issues of stability of field emission.

The transition from thermionic emission to SCL emission or CL law (steady state) had been shown in the pioneering works by Langmuir and others. In 1993, a paper was also reported to show the transition of thermionic emission from metals to SCL condition. In the experiment, YAG pumped laser of about 90 fs and $4.9 \times 10^{11}$ W/cm$^2$ was used to induce thermionic electron emission from various metallic surfaces (Silver, Tungsten, and Aluminum). The measured electron yield is compared with a space charge limited extension of the Richardson–Dushman equation for short time scale as shown in Fig. 18 below.

C. Emission and space charge in a microscale plasma gap

Gas breakdown phenomena from nanoscale to microscale are of fundamental importance in device miniaturization. In the development of microelectromechanical systems
Field emission is found to play a profound role in gas breakdown in systems with AK gap separation at the microscale, leading to significant deviations from the traditional Paschen's law, especially in the low gas pressure regime. Because of very high electric fields at microscale dimensions, field emission provides an abundance of electrons from the cathode, which in turn create large ion concentrations through ionizing collisions. These positive ions generated in the gap increase the local electric field because of positive space charge, which further enhances the field emission process, leading to ion enhanced field emission. Thus, field emission is inherently coupled to the gas and discharge dynamics in microscale systems. This feedback mechanism ultimately leads to breakdown at applied voltages lower than those predicted by Paschen's law or vacuum breakdown theory.

Go and Pohlman provided a single, consistent formulation that combines Paschen’s curve and ion enhanced field emission to form the modified Paschen’s curve, which was also compared with microscale breakdown experiments conducted in atmospheric air. An analytical equation for the modified Paschen’s curve and microscale breakdown was later derived. A quantum model was also developed to study the ion enhanced field emission by determining the space charge of a gaseous ion on electron tunneling. A recent review on these studies can be found at Ref. 241. Most recently, Loveless and Garner derived simple analytic expressions to predict nanoscale to microscale breakdown voltage, by conducting a matched asymptotic analysis.

VI. DIODE PHYSICS IN REAL WORLD

A. High power microwave (HPM) and x-ray sources

High power diodes in the MV, multi KA range are used to generate high power microwaves (HPM). The performance of these HPM sources is dictated by the diode at least in three aspects: diode closure because of the plasma inside the diode, the rise and fall of the diode voltage pulse that leads to mode hopping in HPM, and the intense electron beam leads to unpredictable beam loading of the HPM structures.

The customary view of the electron beam production in pulsed power driven diodes is through an explosive emission process. The explosive field emission from isolated microscopic protrusions creates a local plasma that quickly spreads out over the entire surface. Plasma has virtually zero work function, therefore enabling the electrons to be emitted under the space charge limited condition everywhere on the cathode surface. It is this zero work function that generates current density at the cathode in the kA/cm² range, in contrast to thermionic cathodes yielding only 10 s of A/cm² at best. This cathode plasma expands and drifts toward the anode at a typical speed of 1–10 cm/μs, closing the gap, and during the process increasing the diode current significantly as the limiting current density increases rapidly with decreasing AK gap spacing. This “impedance collapse” is in sharp contrast to conventional tubes where the voltage and current waveform can be made very stable. As a consequence, the duration of the pulse length in a high power diode is limited.
The high current beam emerging from such a diode usually generates HPM over an even shorter duration. A major reason is that the intense beam may significantly load the resonant electromagnetic structures of the HPM device by itself, or by the plasma it generates out of wall materials. The detune and de-Q of the structure may lead to mode jumping at best, and to the extinction of the HPM pulse at worst. This premature termination of the HPM pulse when the diode is still operating has been called “pulse shortening” in the HPM literature.\textsuperscript{17,256–259} The combination of pulse length and peak power has put constraints on cathode and HPM source design. However, recent advances in cathodes\textsuperscript{45–48} and in PIC simulation models\textsuperscript{260,261} which can design cavities more resistant to mode jumping in the presence of plasma motions have, to a certain degree, alleviated some of the previous concerns in “pulse shortening.”\textsuperscript{7}

It is of interest to point out that the space charge in the intense beam has two advantageous effects in beam-gap interaction, as in a relativistic klystron.\textsuperscript{34,262} First the space charge provides a potential barrier across the gap. This potential barrier prevents stray electrons from crossing the gap, thereby reducing the tendency toward breakdown in the gap of a relativistic klystron. Second, this strong potential within the gap leads to a “gate effect” which bunches the electron beam very strongly.\textsuperscript{258,262} This is why a relativistic klystron usually requires only three cavities (including the input and output cavity) to fully bunch the beam.

Conventional tubes required clean cathode surfaces, good vacuum, and little or no plasma within the AK gap for good performance. Explosive emission in pulsed power diodes needs a plasma to form an emission surface with zero work function, and such a plasma is found to be easier to initiate with a not so clean cathode, e.g., one covered with layers of hydrocarbons. A high power diode is far less ready to emit after it undergoes rf discharge cleaning.\textsuperscript{263,264} It is significant to note that both the HPM and pulsed power communities are merging with the techniques used by the microwave tube community, such as surface and vacuum techniques, to alleviate the problem of diode closure.\textsuperscript{258,263}

A significant development in preventing gap closure, demonstrated in the relativistic magnetron, is the invention of the cesium-iodide coated carbon fiber cathode by Shiffler \textit{et al.}\textsuperscript{45,46,48} Presumably, these fibers emit electrons through field emission. Experiments show little evidence of plasma expansion during the diode operation, and there was no visible plasma light near the cathode. Despite the apparent lack of a plasma, the Shiffler cathodes can reliably reproduce kA current with current densities at many tens of A/cm\textsuperscript{2}. Cathode life times over 1 × 10\textsuperscript{8} shots have been achieved. The experimental performance of these cathodes warrants additional study of the role of low outgassing materials, field versus explosive emission physics, and the shielding effect\textsuperscript{13,77} of neighboring fibers.

Measurements\textsuperscript{265} from the cold cathode for a relativistic klystron amplifier showed that the I–V relationship deviates from the CL scaling (I ∝ V\textsuperscript{3/2}), especially during the initial period of voltage increase. This was explained by a voltage dependent perevance change of the cold cathode gun, as shown in Fig. 19. PIC simulations revealed that this change of perevance was due to the increase in the emission area of the cathode during the voltage increase.\textsuperscript{265}

Intense electron beam diodes are being extensively researched, primarily for the purpose of generating high brightness x-rays.\textsuperscript{266–273} In these diodes, because of high beam current (typically in hundreds of kA), the strong self-magnetic field pinches the electron beam to a few millimeters in diameter at the high atomic mass anode, where bremsstrahlung x-rays are generated.\textsuperscript{269,271} The heating of the anode by the intense electron beam enables space charge limited ion emission from the anode surface.\textsuperscript{271,273} The collective beam/plasma effects dominate the diode and beam characteristics, where the anode plasma expansion and increased ion space charge near the cathode surface are shown to cause rapid impedance loss of the diode and disruption of the radiation generation.\textsuperscript{273,274}

B. Heavy ion beams

Heavy ion beams are proposed as a driver for inertial confinement fusion (ICF). In comparison with ICF driven by lasers, a heavy ion beam driver has significantly higher efficiency.\textsuperscript{28} ICF driven by light ion beams would require mega amperes of current\textsuperscript{28} which makes beam focusing extremely difficult. On the other hand, ICF driven by heavy ion beams, known as heavy ion fusion (HIF), would require 10 kA beam current, but beam energy on the order of 1–10 GeV, and a pulse length of order 10 ns. In the direct drive HIF, tens of ion beams, each carrying a current of order kA, are illuminated symmetrically onto a spherical capsule.\textsuperscript{275} The ablation of the outer layers of the capsule provides the rocket effects which produce the implosion. As in laser fusion, or any ICF scheme, the Rayleigh–Taylor instability always presents a most serious threat, and in HIF, uniformity of illumination is a critical issue. Thus, studies of heavy ion beam transport, focusing, and emittance (a measure for the average
spread of particle coordinates in phase space, see, e.g., Refs. 31 and 32) growth became a central area of beam physics for the last 40 years.

Since a 1 GeV Pb beam has a velocity of 0.1c, Pb beams at kA level are necessarily space charge dominated. Prominent issues in space charge dominated beams are beam focusing and emittance growth. Thus both electrostatic focusing and magnetic focusing have been used in different stages of beam acceleration. Since an electron beam of 0.1c corresponds to 2.6 kV, a low energy electron beam may be used to simulate emittance growth in the space charge dominated regime. It might also be used to assess some potentially threatening instabilities that are anticipated during beam transport. Using a low energy electron beam to simulate space charge dominated ion beams and to study the resistive wall instability have become a major endeavor at the University of Maryland at College Park.32,275,276 Significant progress concerning halo formation, evolution of x-y phase space, emittance growth, and resistive instabilities, etc., has been made. These achievements have been summarized by Reiser277 and by Kishek et al.276 It is noteworthy that this university scale effort provided the much needed benchmark between experiments, theory, and simulation codes in space charge dominated beams, in addition to pushing into a space charge regime that is unprecedented in the HIF community.

While the design and control of low energy electron injection is a well-established practice for the University of Maryland experiments, development of ion sources and ion beam injection for HIF is a very different matter. The prevalent ion source used in HIF is the surface ionization source (which is a thermionic source), while the metal vapor vacuum arc source and plasma ion source have also been considered. Regardless of the source, the scaling of ion surface current density in terms of injector voltage and diode spacing is still governed by Eq. (2) in which $m$ is now the ion mass ($K^+$ for instance). Note that this Child–Langmuir law, due to space charge effects and obtained from a dimensional argument of the Poisson equation (1), imposes a constraint for HIF on the diode voltage (high), diode size (large), and achievable current (low). Additional considerations such as cost, size, achievable current, beam brightness, and electrical breakdown, etc., favor the creation of multiple ion beamlets which radially merge to form a single high current ion beam before this beam is transported into an electrostatic focusing channel for subsequent beam acceleration and longitudinal compression. Clearly, beam matching and emittance growth in the injector stage are serious concerns, and much studied.

Heavy ion beams, most notably carbon ions, have been deployed to treat deep-seated or radiation resistant tumors.278 Heavy ion beams have also been used to probe the electric potential, electron density, magnetic field component, and turbulent structures in a plasma. We mention in passing that the ion source is the central component that produces thrust for space propulsion.279

Not only the CL law, the Richardson–Dushman and Schottky corrections were also observed in ion sources. The thermionic emission of $K^+$ and $Cs^+$ ions from heated aluminosilicates was systematically studied by Kolling et al.280 The temperature dependence of the ion emission follows the classical Richardson–Dushman behavior. When the electric field was small (less than 1000 V/cm), the emitted ion density is space charge limited, as shown in Fig. 20. Ion emission transitioned to the Schottky regime (i.e., limited by the effective work function of the cathodes for ion emission) at higher electric fields (larger than 2000 V/cm).

C. Beam compression in THz sources and free electron lasers

Serious challenges arise when the HPM source is scaled to 100s of GHz to THz.16 If a pencil beam is to be used, as in conventional traveling wave tube (TWT) or klystron, the beam tunnel’s cross-sectional area is proportional to $\lambda^2$, where $\lambda$ is the free space wavelength. The beam radius, $a$, is usually chosen with $a/c = a/\lambda = 0.6$, i.e., $a = 0.1\lambda$, so that (a) the RF electric field is fairly uniform across the beam’s cross-section, and (b) the beam “fills up” a sizeable fraction of the beam tunnel without excessive beam interception with the wall. Thus, in the THz range, $a$ is exceedingly small. The electron beam current necessary for appreciable gain at moderate beam voltage (like 10–20 kV) at these high frequencies easily requires a current density of the order of 100s of A/cm², way beyond what is available from the most advanced thermionic cathodes. This implies aggressive radial convergence in the design of the Pierce diode. It also requires a very strong, and carefully designed focusing magnetic field for good beam transport within such a minuscule beam tunnel.

Because of the exceedingly high current density in THz sources, a less than perfectly designed diode leads to beam interception at the anode, along the drift tube, or in the collector region, all of which may lead to catastrophic failure of the tube, as shown in klystrino experiments.281 Even moderate scalloping of the beam can greatly diminish the power and efficiency, because it is equivalent to a large “beam emittance” which is detrimental to gain at high frequencies. This is compounded by the observed high cold tube loss rate.

![FIG. 20. Measured total ion current density as a function of the AK gap voltage $U_R^{3/2}$, indicating CL behavior for small voltage ($<1000$ V/cm). Open symbols indicate data points included in linear regression. Reproduced with permission from Kolling et al., J. Appl. Phys. 107, 014105 (2010). Copyright 2010 AIP Publishing LLC.](image-url)
at THz (which is yet to be understood), and by the proportionately larger effects of surface roughness which leads to a much higher surface resistance at high frequencies. The intrinsic high circuit losses in a THz tube make the control of beam emittance a critical issue.

The use of a sheet electron beam allows the beam tunnel’s cross-sectional area to scale as $\lambda$ (instead of $\lambda^2$ as in a pencil beam). Accordingly, much higher beam current may be transported in a larger structure, with less demanding focusing magnetic field, and a milder convergence ratio in the Pierce sheet beam diode for high frequencies. A novel design of the focusing magnetic field may eliminate the ditofocusing magnetic field, and a milder convergence ratio be transported in a larger structure, with less demanding nel’s cross-sectional area to scale as $k_{\text{DTEM}}$.284 The pulse length of interest in these applications (UED) and dynamic transmission electron microscopy (DTM).284 The pulse length of interest in these applications typically ranges from tens of femtoseconds to a picosecond, while the beam diameter is on the order of 10 $\mu$m. For the purposes of detection and resolution it is desirable to maximize the charge in the beam while minimizing its emittance.285,286 For example, in time resolved electron microscopy and in free electron lasers (FEL).

Ultrafast electron microscopy (UEM) is of growing importance to understand the kinetics and transient behavior of various processes in materials science, biology, and chemistry. UEM encompasses both ultrafast electron diffraction (UED) and dynamic transmission electron microscopy (DTEM).284 The pulse length of interest in these applications typically ranges from tens of femtoseconds to a picosecond, while the beam diameter is on the order of $10 \mu m$. For the purposes of detection and resolution it is desirable to maximize the charge in the beam while minimizing its emittance.285,286 For example, in time resolved electron diffraction, it is required to have short electron pulses with sufficient electron density for imaging. The propagation dynamics of femtosecond electron packets including space charge effects in the drift region of a photoelectron gun are important.285–288 It was found that space charge effects can broaden the electron pulse to many times its original length in both temporal and kinetic energy distributions.289,290

In the FEL,291 a relativistic electron bunch propagates through the periodic magnetic field of the undulator transferring the kinetic energy of the electron to an electromagnetic wave in a resonant condition. FELs can either operate as amplifiers or, more commonly, as oscillators. In most oscillator configurations the electromagnetic wave is confined within an optical cavity that encompasses the undulator, but in FELs designed for generation of hard X-rays the oscillator must be a single pass device due to the lack of appropriate reflectors for the X-rays. In both cases the electron beam must be bunched. In particular, the X-ray FEL calls for short pulse electron beams of very high brightness. Similarly, FEL amplifiers rely upon electron beams of short duration, high current, and low energy spread for good performance.

At this time RF photoinjectors are the state of the art sources for high brightness, short pulse electron beams. In these devices, the brightness is affected by transverse and longitudinal emittance at the point of emission, as well as emittance growth due to space charge forces and intra-beam scattering. The minimal emittance is dependent on the material properties of the cathode, emission process, and cathode temperature.292 For thermionic emission and delayed emision from semiconductor photocathodes, the temperature is of primary importance. For metal photocathodes and for prompt photoemission from semiconductor photocathodes it is the discrepancy between the photon energy and characteristic energy of extraction (work function, band gap energy, and electron affinity) that is most important in determining the emittance. However, space charge effects can rapidly lead to irreversible emittance growth, and must therefore be accounted for. A recent development for RF photoinjectors is the use of the so called “blowout” regime where the electron bunch is formed at the cathode by an ultrashort laser pulse, and is subsequently expanded by space charge forces into a uniformly filled elliptical shape. Musumeci et al.293 describe such an experiment where a 35 fs laser pulse is used to produce the beam in a photoinjector operated at a peak field of 80 MV/m. Among other things, they observe an asymmetry in the resulting pulse, at high laser intensity, due to the acceleration at the tail end of the pulse being diminished by the space charge field of the front end. Whether using a photoinjector in the “blowout” regime, multiphoton emission with short and intense pulses,294 or more conventional photoinjection with somewhat longer pulses,294 space charge forces will generally limit the brightness of the emitted bunches, drastically so if a virtual cathode is formed.

Since the relative magnitude of the space charge field compared with the applied accelerating field is a major determining factor on the upper limit of brightness, there is some promise that novel beam sources such as laser wakefield accelerators and electron beam wakefield accelerators with accelerating gradients that are orders of magnitude higher than those of RF accelerators could yield much brighter beams.295

D. Critical current in a crossed-field gap

The inclusion of a transverse magnetic field, $B$, that is orthogonal to the DC electric field greatly complicates the description of the physics of diodes, even in 1D. First, if $B$ is sufficiently large, electrons emitted from the cathode with zero initial velocity will not be able to reach the diode. For the 1D planar diode, this critical magnetic field, known as the Hull cutoff magnetic field, $B_{\text{Hull}}$, reads14,15,28,29

$$B_{\text{Hull}} = \frac{1}{D} \sqrt{\frac{2mV_i}{e}},$$

in the nonrelativistic limit. If $B > B_{\text{Hull}}$, the diode is magnetically insulated, and an electron emitted from the cathode is unable to reach the anode, (even if the diode contains an
arbitrary ion distribution that varies only in the direction of the external electric field). We shall separately consider the two cases $B < B_{HI}$ and $B > B_{HI}$.

1. $B < B_{HI}$ case

For $B < B_{HI}$, an electron emitted from the cathode will reach the anode. The maximum emitted current density, normalized to $J_{CL}$, is derived by Lau et al. and is shown in Fig. 21. This maximum current density is attained when the space charge limited condition is satisfied, i.e., when the cathode electric field vanishes. It is equal to $J_{CL}$ when $B = 0$, as expected for a nonmagnetized diode. It is equal to $0.718J_{CL}$ as $B \rightarrow B_{HI}$. The curve of $J$ vs. $B$ in Fig. 21 has an infinite slope as $B \rightarrow B_{HI}$.

We should stress that the results given in the preceding paragraph are valid only for a nonrelativistic diode. For a relativistic diode, the limiting current is expected to be high, in which case the self-magnetic field on the electron motion can no longer be ignored. In fact, this self-magnetic field is very important in self-pinched diodes. The self-pinched, intense E-beam diodes are the principal devices for X-ray radiography. The influence of the very strong self-magnetic field makes the orbital calculations analytically intractable. Thus, the self-consistent limiting current for a relativistic diode, for $B < B_{HI}$, has not been established. The solution to Ampere’s law, $\nabla \times \vec{B} = \mu_0 \vec{J}$, necessarily contains spatial non-uniformity of $B$ in a direction orthogonal to $\vec{J}$, making the problem three dimensional automatically. Jory and Trivelpiece’s relativistic correction to the 1D classical Child–Langmuir law ignores the self-magnetic field.

2. $B > B_{HI}$ case

For $B > B_{HI}$, an electron follows a cycloidal path that commences and finishes on the cathode surface. The maximum emission current, intuitively, is also attained by imposing the space charge limited condition of zero surface electric field on the cathode. This seemingly reasonable assumption turns out to be false, for either nonrelativistic or relativistic diodes. The maximum emission current density for the non-relativistic diode, for $B > B_{HI}$, is shown by the solid line in Fig. 21. At the maximum emission current density, the electric field on the cathode surface is non-zero and it accelerates the emitted electrons (with an assumed zero emission velocity).

The cycloidal electron flow, for $B > B_{HI}$, is very unstable, even when the emission current level is substantially below the solid curve, such as the point $C_1$ in Fig. 21. From numerical simulations and the accompanying analytic studies, it has been shown that this low current cycloidal flow is unstable in the presence of a small AC gap voltage, or of a small misalignment of the external magnetic field, or of some stray resistance across the AK gap. The final state is always the laminar Brillouin flow superimposed by a weak turbulent background. The underlying reason is that any of the above mentioned perturbations has a tendency to prevent an emitted electron from returning to the cathode. Thus, space charge in the gap keeps on accumulating until the Brillouin flow is established, for which the total space charge is of order $CV_2$ (again).

The cycloidal flow model is also known as the “multistream model.” The Brillouin flow, which is a laminar, shear flow in the direction of $E \times B$, is also known as the “single stream” model because the electron paths do not cross. Substantial controversy existed in the literature, as to whether the single stream model or the multistream model should be used to describe the basic state in magnetrons, and in high power ion diodes. In the study of magnetrons, the cycloidal orbits under space charge limited conditions are known as the Slater orbits, the $N$-th order Slater orbit refers to an emitted electron executing $N$ cycloidal hops before returning to the cathode, and the Brillouin flow corresponds to the Slater orbits of the order $N = \infty$. While the cycloidal orbits were named after J. C. Slater, Slater himself emphasized that the prevalent flow in a crossed field gap should be the Brillouin flow because it has the lowest energy state among all Slater orbits. Simulations by Palevsky and Bekefi confirmed this insight of Slater, which was subsequently further corroborated.

From the preceding discussions, it appears that there is no steady state, 1D equilibrium solution in a magnetically insulated crossed-field gap if the electrons’ emission velocity distribution is Maxwellian, regardless of the level of the emission current density. This conjecture of non-existence of equilibrium solution in a crossed-field gap, at any non-zero emission current level, remains to be proven. As indicated above, the prevalent state in a crossed-field gap is the Brillouin flow superimposed by a weak turbulent background.

While the planar Brillouin flow is fairly well understood, Brillouin flow in the cylindrical, inverted magnetron configuration (meaning the cathode is outside and the anode is inside) remains an open question for a long time, in view of the potential negative mass instability intrinsic to electron orbits in the inverted magnetron configuration. The stability of the Brillouin flow in a crossed-field diode was recently studied for planar, cylindrical (both conventional and inverted) magnetron with and without a slow wave structure on the anode.
Magnetic insulation is easy to defeat. The Hull cutoff condition is based on conservation of energy and conservation of canonical angular momentum in a strictly 1D analysis. Any non-uniformity in the $E \times B$ direction, due for example, to local ion production, or some electrostatic perturbations, etc., will violate conservation of canonical angular momentum and potentially leads to the leakage of electron current to the anode. This is why electron spokes are so readily formed in a magnetron making it highly efficient, with efficiency reaching 90%. This is also why gap closure persists in magnetically insulated transmission lines (MITL), why it is difficult to magnetically confine a plasma, and why eliminating the leakage currents in Hall thrusters is also difficult.

These narratives show the complexity of crossed field devices in general. Shortly before his death, the venerable John R. Pierce lamented, “Crossed field amplifiers are like the Lord. They are powerful, but difficult to understand.”

Besides the fascinating fundamental physics, the crossed-field gaps in practice exhibit electron and ion sputtering on the anode and cathode, heating and outgassing, arc and breakdown, etc. Reference may be made to excellent reviews by Cuneo, Latham, and Gilmour.

Using the quantum model of Sec. IV A, the SCL current density of a crossed field gap was studied for a nano gap. It was found that in the quantum regime, the crossed field gap is no longer magnetically insulated at magnetic field ($B$) larger than the Hull cutoff value ($B_{H}$) as predicted by the classical model. Due to electron tunneling, there is finite probability that the electrons will tunnel through the barrier to arrive at the anode even at $B > B_{H}$. Presence of ions also encourage diode closure in a crossed-field gap. Some unsolved problems and challenges in nanoscale and ultrafast diodes are recently addressed by Zhang and Lau.

E. Sheaths in thrusters and plasma processing

Because electrons in plasmas have higher mobility than the much heavier ions, when a quasi-neutral plasma is in contact with a solid wall, an electric space charge region—the plasma sheath—will build up near the surface. In the sheath region near the plasma boundaries, potential and density variations from the bulk plasma will be established, in order to balance the particle fluxes under the electrical conditions imposed at the solid walls. Understanding sheath phenomena is particularly important in applications such as ion and Hall Thrusters and plasma processing.

Typically, when electrons escape the plasma volume faster than the ions, more ions would be left behind in the plasma, resulting in a net positive charge potential. This positive potential will act as a retarding force for electrons moving towards the wall. As a result, the electrons would be slowed down and kept in the plasma. Near the plasma-wall boundary, potential gradients will be developed, forming a plasma sheath region, which confines the quasi-neutral plasma from the wall.

Under steady state conditions, the simplest, self-consistent model of a sheath in a plasma is to assume the potential drop across the sheath to be sufficiently large so that the electron density inside the sheath essentially goes to zero, which is expected to hold for very small electron temperature $T_{e}$ (as compared with the potential drop $eV$). By solving the coupled ion current continuity equation and Poisson’s equation, it is easy to show that the ion current density across the sheath of thickness $D$ with potential drop of $V$ is governed by

$$J_{i} = \frac{4\sqrt{2}}{9} e \frac{V^{3/2}}{M D^{2}},$$

which is the CL law, Eq. (2), with $M$ being the ion mass. This sheath is naturally called Child–Langmuir (CL) sheath, which is particularly important to ion thrusters. For a given plasma density, the CL sheath thickness varies as the potential to the 3/2 power. In ion thrusters, the accelerator structure can be designed, to first order, using the CL law with $D$ being the gap between the accelerator electrodes. An excellent discussion on the design of ion thruster accelerator grids using the CL scaling is given in Ref. 311. In plasma processing, CL scaling is substantially used to study sheath physics for both capacitive and inductive plasma discharges.

A comprehensive investigation by Hershkowitz showed that in low temperature collisionless and weakly collisional plasmas, the CL law provides a good fit to the sheath potential profile for both ion and electron sheaths when $eV \gg T_{e}$. Benilov provided a thorough discussion of the analytical models of space charge ion sheaths, including the CL law and model of collisionless ion sheaths, the Mott–Gurney (MG) law and model of collision dominated ion sheath, the Bohm model of a collisionless ion–electron sheath, the Su–Lam–Cohen model of a collision dominated ion–electron sheath, and ion sheaths with arbitrary collisionality. Collisional sheaths have been studied experimentally by Lisovskiy et al. They found that the collision-dominant CL law with constant ion mobility, rather than that with constant ion mean free path, was valid in a high electric field of glow discharge in hydrogen, as shown in Fig. 22. Note that the collision-dominant CL law with constant ion mobility is essentially the MG law, with the same scaling given in Eq. (3). A detailed review of sheaths in laboratory

![Image](https://via.placeholder.com/150)

**FIG. 22.** Experimental discharge current versus the $U^{m}/d^{n}$ ratio for the hydrogen pressure of 2 Torr, where $d$ is the cathode sheath thickness, and $U$ is the voltage drop across the sheath. Only the curve with the power scaling of $m = 2$ and $n = 3$ can be fitted with a straight line. Reproduced with permission from Lisovskiy et al., Phys. Scr. 91, 085601 (2016). Copyright 2016 The Royal Swedish Academy of Sciences.
and space plasmas was given by Robertson.\textsuperscript{320} Plasma sheaths with surface electron emissions have generated significant recent interest,\textsuperscript{321–324} for which the CL sheath scaling is subjected to further study.

In some plasma processing systems, it may not be possible to avoid the presence of dust particles. In a 2008 paper,\textsuperscript{325} the CL law is extended to include the effects of charged dust impurities. The paper provides analytical and numerical results of the 1D CL law in a planar diode with negatively charged particles. It was found that the scaling to the gap spacing remains as $D^{-2}$, but the scaling to the gap voltage deviates from the well-known $3/2$ power law at high voltages and high dust charge density. This model can be considered as the CL law for dusty plasmas. Further research on limiting currents in the presence of strongly coupled plasmas must be considered active and open questions.\textsuperscript{326}

\section{VII. CONCLUSIONS AND OUTLOOK}

This article reviews the advancement of diode physics since the pioneering works of Child and Langmuir. The classical CL law has been vastly generalized in various aspects, including higher dimensional effects, time dependent physics, quantum and relativistic extensions, and electron emission physics. Significant progress has been made in the understanding of the fundamental physics and in the development of simulation codes, for both scientific research and countless applications involving diodes. As systems become increasingly sophisticated, the basic scaling laws in diode physics still largely govern charge particle transport, because they represent the constraint imposed by the Poisson equation.

One important aspect about the CL law is that the physics is quasi-separable. That is, the revisions to the basic SCL scaling due to new physics can be made by introducing an additional term. For practical systems involving multiple physics, such as higher dimensions, pulsed operation, and quantum and/or relativistic physics, the “integrated” view of the SCL scaling may be approximately written in the form of

$$J = J_{\text{CL}}(1D) \times \left[\text{multi dimensions}\right] \times \left[\text{short pulse}\right] \times \left[\text{quantum}\right] \times \left[\text{relativistic}\right].$$

The above scaling may serve as a basic guide to the design of more complex devices or components involving diode physics.

While CL remains critical in high power diodes, sheath physics, and high current drivers used in high energy density physics experiments, it is intrinsic to any area involving space charge effects. For dense sources of charged particles from plasmas produced by intense lasers, much research is needed to understand the upper limit of beam brightness, which is perhaps one of the most fundamental questions in laser wakefield accelerators and electron beam wakefield accelerators. In the development of ultrafast electron microscopy (UEM) for the wide range of studies in the kinetics and various transient processes in materials science, biology, and chemistry, one of the major challenges is to maximize the charge per bunch while minimizing its emittance, i.e., to minimize the space charge effects.

Space charge is also found to be important in high order harmonic generation (HHG) from laser interactions with either atom clusters or solid density plasmas, which are considered to be promising ways of generating bright and ultrashort burst of X-rays. HHG in solids is also thought to bridge the gap between attosecond science and condensed matter physics. Compared with HHG from single atoms, space charge potential in solids is expected to play a much stronger role. When high intensity ultrafast laser pulse is used, electrons can be accelerated to relativistic velocities at ultrashort time scale, where relativistic and quantum effects have to be considered. The fundamental link between HHG and space charge deserves further investigation.

In nanoscale quantum diodes, quantum dots, and plasmonic devices and structures, such as charge transfer plasmon (CTP) tunneling junctions, photodiodes, and plasmonic electron emitters, CL becomes increasingly important when great efforts are being made to increase the current density. In particular, it is found that operating the CTP tunneling junctions in the SCL regime could significantly reduce its quantum tunneling damping parameter, which provides new insights in the design of quantum plasmonics.

The solid state analog of CL, the Mott–Gurney (MG) law, governing the maximum charge injection in solids, including organic materials and other dielectrics, is important to energy devices, such as solar cells and light emitting diodes. With the advancement of nanotechnology and ultrafast science, the extensions of the MG law to these new regimes are urgent.

While CL continues to influence new and emerging areas, there are various attempts to beat the CL limit. One such effort is to overcome CL in a time average sense. It is found that current in a traditional macroscopic diode could exceed the CL limit even by 50% if the voltage drop across the diode varied periodically in time in a carefully tuned process. Another method to overcome the CL limit is to modulate the current emission in time. Though these studies rest on a different definition of the CL limit, namely, a comparison between the time averaged diode current and the adiabatic average of the expression for the stationary CL limit, having a larger current can be potentially useful by itself in various systems.

Several open questions remain regarding the Ramo–Shockley theorem (RS).\textsuperscript{35,36} RS provided the basis for detecting charge motion within a detector by its induced current on electrodes, which is fundamental to radiation detection, microwave, and vacuum electronic devices. With the trend of miniaturization, and the push to higher power in these devices, it is natural to ask if RS is relativistically and/ or quantum mechanically correct.

At nano- or sub-nanometer scale, the boundaries among solid, plasma, and vacuum become blurred. Thus, the typical sharp boundary models adopted in CL studies require a new examination. In its essential form, anything where charge transport takes place under the influence of a voltage or electric field over some length may be considered as a diode, thereby encompassing atoms, laser plasmas, and even...
biological structures (e.g., lipids). In this sense, diode physics will continue to exert its impact for another hundred years.

**ACKNOWLEDGMENTS**

P. Zhang and Y. Y. Lau were supported by AFOSR Grant No. FA9550-14-1-0309. P. Zhang was also supported by AFOSR through a subcontract from the University of Michigan. L. K. Ang was supported by AFOSR-AOARD FA2386-14-1-0420, and J. W. Luginsland was supported by AFOSR.
