JPL Publication 11-3

Extension of the ADC Charge-Collection Model to Include Multiple Junctions

Larry D. Edmonds

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

April 2011

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

©2011 California Institute of Technology. Government sponsorship acknowledged.

TABLE OF CONTENTS

Abstract iii

I. Introduction 1

II. Review of The Drift-Diffusion Equations 3

III. The High-Injection Limit 7

IV. Regional Partitioning: Definitions and Equations 11

V. Relating Terminal Currents to Potential Boundary Values Under Ideal Boundary Conditions 17

VI. Several Inequalities 21

VII. Some Topological Properties 25

VIII. Necessary and Sufficient Conditions for the Existence of An HRR 29

IX. The Two-Terminal Problem 30

X. The Three-Terminal Problem 32

XI. Minimum Voltage Needed to Maintain a Reverse-Biasing Condition 37

XII. A Numerical Example Compared to TCAD 41

XIII. Conclusions 51

References 52

Appendices

A Invertibility of the Coefficient Matrix $C_{i,j}$ 53

B Inequalities Involving the Coefficients $C_{i,j}$ 55

C Inequalities Involving the Coefficients $\Phi_{i,j}$ 58

D Properties of Ohm's Law 62

ABSTRACT

The ADC model is a charge-collection model derived for simple p-n junction silicon diodes having a single reverse-biased p-n junction at one end and an ideal substrate contact at the other end. A steady-state ionization source liberates charge carriers in the device and the model estimates the terminal current produced by this carrier liberation. A recent paper confirmed that model predictions of collected charge are also correct for the transient problem in which carrier liberation is impulsive instead of steady-state. The present paper extends the model to include multiple junctions, and the goal is to estimate how collected charge is shared by the different junctions. Much of the theory is derived for an arbitrary number of junctions but a complete solution is given only for devices containing two junctions in a horizontal arrangement and exposed to a vertical line-source (or "track") of carrier generation. For each of three possibilities, the model identifies the conditions needed to produce that possibility. One possibility is that charge is shared by both junctions, and the total collected charge from the two junctions is less than the total amount of liberated charge. A second possibility is that collected charge is shared by both junctions, and the total collected charge from the two junctions is equal to the total amount of liberated charge. The third possibility is that all liberated charge is collected by one junction, and no charge is collected by the other. An example n^+-p device exhibited all three possibilities by varying the location of the track. The same example but with doping types interchanged to produce a p^+-n device exhibited only the first possibility regardless of track location. All examples show excellent quantitative agreement with TCAD simulations.

***Key words*:** ADC model, ambipolar diffusion, ambipolar diffusion with a cutoff, charge collection, drift-diffusion.

I. INTRODUCTION

A charge-collection model, called "ambipolar diffusion with a cutoff" (ADC) was reviewed in [1] (a more rigorous derivation is in an appendix in [2]) for a reverse-biased p-n junction silicon diode exposed to an ionization source that liberates carriers (electron-hole pairs) in the device. An important application of a charge-collection model is in the investigation of single-event effects (SEE) in which the ionization source is a single particle, such as a galactic cosmic ray heavy ion. The ADC model has some known limitations for this application. The first limitation is in the category of physical approximations. Carrier transport is described by the drift-diffusion equations with constant mobilities in the quasi-neutral region, carrier recombination in the device interior is neglected, and ideal boundary conditions are assumed. In contrast, real devices are more complex, e.g., mobilities depend on a variety of physical parameters. A second limitation is in the category of intended applications. The model is intended for high-injection conditions, i.e., the model is intended to become accurate when the ionization source is sufficiently intense. While SEE is a high-injection phenomenon, carrier liberation is still finite. The model is a mathematical limit that is an inexact approximation for the finite case. The last limitation also belongs to the category of intended applications. The model was derived from a steady-state analysis in which carriers are liberated at a quasi-constant rate (e.g., by photon irradiation). In contrast, SEE is a highly transient problem. In spite of these limitations, model predictions for example transient problems, in which the ionization source represents an ion track and the quantity of interest is total collected charge (terminal current integrated over all times), agreed well with predictions made by TCAD simulations [3]. However, the original model was derived for a simple silicon diode having only one charge-collecting junction. Motivated by the success of this model for a single-junction device, this paper presents a multi-junction version of the ADC model. A silicon device now contains a collection of reverse-biased p-n junctions and the goal is to determine how collected charge is shared between different junctions. To keep the model simple, the same limitations (physical approximations, high-injection, and steady-state) still apply. The contention here is that there is little hope of understanding charge-collection in a more complex problem if we have not yet understood charge-collection for this simpler problem, so an understanding of the simpler problem is a prerequisite to understanding a more difficult problem.

A silicon device having three (for example) junctions is illustrated in Fig. 1, which is explained as follows. All p-n junctions are formed with a common substrate. If the substrate is p-type, all junctions are n^+-p. If the substrate is n-type, all junctions are p^+-n. Each junction produces a depletion region (DR) and each DR has a boundary (DRB) in the substrate. The three DRBs are denoted S_1, S_2, and S_3 in the figure. The electrical contact to the substrate is treated as an ideal ohmic contact and is denoted S_0 in the figure. The quasi-neutral region (QNR) is that portion of the substrate that is outside the DRs. Hence, boundaries that enclose the QNR include S_0, S_1, S_2, and S_3. The remainder of the QNR boundary is regarded as reflective (the gradients of the potential and carrier density have zero normal components). Voltages applied to the terminals produce a reverse-biasing condition across each DR. A steady-state ionization source having an arbitrary spatial distribution (not shown in the figure) liberates carriers in the substrate, and some

of the liberated minority carriers move to and through a junction to produce a steady-state terminal current at that junction. The goal is to estimate the terminal current for each junction.

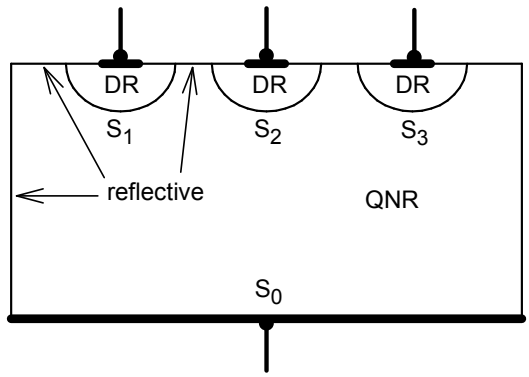

Fig. 1. Illustration of a device with three junctions. All three junctions are either n^+-p (for a p-type substrate) or p^+-n (for an n-type substrate). Each junction produces a DR, and the DR boundaries (DRBs) are denoted S_1, S_2, and S_3. The electrical contact to the substrate (a.k.a., electrode) is denoted S_0. The QNR is the portion of the substrate outside the DRs. Voltages applied to the terminals reverse bias each DR. An ionization source (not shown) produces an arbitrary spatial distribution of liberated carriers in the substrate and the goal is to calculate the terminal currents.

II. REVIEW OF THE DRIFT-DIFFUSION EQUATIONS

The well-known drift-diffusion equations can be found in any textbook on semiconductors. Under steady-state conditions with negligible recombination in the device interior, the equations reduce to

$$\frac{1}{q}\vec{J}_e = D_e \vec{\nabla} n - \mu_e n \vec{\nabla}\varphi \tag{1a}$$

$$\frac{1}{q}\vec{J}_h = -D_h \vec{\nabla} p - \mu_h p \vec{\nabla}\varphi \tag{1b}$$

$$\frac{1}{q}\vec{\nabla}\circ\vec{J}_e = -g \tag{2a}$$

$$-\frac{1}{q}\vec{\nabla}\circ\vec{J}_h = -g \tag{2b}$$

$$\varepsilon\nabla^2\varphi = q\left[(n-n_0)-(p-p_0)\right] \tag{3}$$

where $\boldsymbol{J}_e$ and $\boldsymbol{J}_h$ are the electron and hole current densities (functions of spatial coordinates), q is the elementary charge, D_e and D_h are the electron and hole diffusion coefficients (approximated as constants in this analysis), μ_e and μ_h are the electron and hole mobilities (also approximated as constants in this analysis), φ is the electrostatic potential (a function of spatial coordinates), n and p are the electron and hole densities (functions of spatial coordinates), n_0 and p_0 are the equilibrium electron and hole densities (constants for the uniformly-doped case considered), g is the carrier-generation rate density (a function of spatial coordinates), and ε is the permittivity constant of the medium. The same generation rate is used for both carriers because we consider the case in which electrons and holes are generated in pairs.

Junctions will be described by boundary conditions at the DRBs, so analysis of the drift-diffusion equations is needed only in the QNR. A region is quasi-neutral when the solution to the above equations can be approximated by the solution to the equations obtained from the limiting case as ε approaches zero. In this limit, (3) is replaced with

$$p = P + p_0\,, \qquad n = P + n_0\,, \tag{4}$$

where P is the excess carrier density that is the same for electrons and holes. We also use the Einstein relations $D_e = V_T\mu_e$ and $D_h = V_T\mu_h$, where V_T is the thermal voltage (sometimes written as kT/q and is about 0.026 V at room temperature). This allows us to write (1) as

$$\frac{1}{qD_e}\vec{J}_e = \vec{\nabla}P - \frac{P+n_0}{V_T}\vec{\nabla}\varphi \tag{5a}$$

$$-\frac{1}{qD_h}\vec{J}_h = \vec{\nabla}P + \frac{P+p_0}{V_T}\vec{\nabla}\varphi \tag{5b}$$

Note that (5) was obtained by replacing (3) with (4), but (4) cannot be used to solve for the electric field (the negative gradient of φ). However, (2) together with (5) is a

complete (when given boundary conditions) system of equations that will solve for both P and φ. It is important to note that (4) is a replacement for (3), not an equation that supplements (3). The system of equations consisting of (1) through (4) is over-determined and inconsistent, so (3) is discarded when (4) is used. However, Poisson's equation (3) would have a role in an iteration method if an improved estimate of the charge density is desired. Starting with an initial estimate of zero charge density, i.e., (4), we use this with (1) and (2) to obtain (2) and (5), then solve (2) and (5) for φ, and then substitute this into the left side of (3) and interpret the calculated value of the left side as an improved estimate of the charge density. This is analogous to the analysis of an ohmic medium with a specified inhomogeneous conductivity. For the ohmic problem, we would use Ohm's law, not Poisson's equation, to solve for the electric field. Having done that, we can then use Poisson's equation to calculate the charge density if desired. Similarly, for the semiconductor problem considered here, we use (2) and (5), not Poisson's equation, to solve for the electric field. Having done that, we can then use Poisson's equation to calculate the charge density if desired.

For an n-type material, we can neglect p_0 and set n_0 equal to the doping density. For a p-type material, we can neglect n_0 and set p_0 equal to the doping density. We can shorten the notation by including only one equilibrium density, and still represent both doping types with the same set of equations, by letting D_m denote the diffusion coefficient for minority carriers, D_M is the diffusion coefficient for majority carriers, and N denotes the doping density. We also define

$$\vec{J}_m \equiv \begin{cases} \vec{J}_e & \text{for p-type} \\ -\vec{J}_h & \text{for n-type} \end{cases}, \qquad \vec{J}_M \equiv \begin{cases} \vec{J}_h & \text{for p-type} \\ -\vec{J}_e & \text{for n-type} \end{cases} \tag{6a}$$

$$U \equiv \begin{cases} \varphi & \text{for p-type} \\ -\varphi & \text{for n-type} \end{cases}. \tag{6b}$$

Note that the subscript m to the $\boldsymbol{J}$ denotes minority-carrier current, while the subscript M denotes majority-carrier current. The sign convention was selected so that both doping types will be described by the same equations. With this sign convention, the direction of $\boldsymbol{J}_m$ is opposite to the direction of minority-carrier motion, while $\boldsymbol{J}_M$ has the same direction as the majority-carrier motion. Substituting (6) into (2) and (5) gives

$$\frac{\vec{J}_m}{q\,D_m} = \vec{\nabla}P - \frac{P}{V_T}\vec{\nabla}U, \qquad -\frac{\vec{J}_M}{q\,D_M} = \vec{\nabla}P + \frac{P+N}{V_T}\vec{\nabla}U \tag{7}$$

$$\vec{\nabla}\circ\vec{J}_m = -q\,g, \quad \vec{\nabla}\circ\vec{J}_M = q\,g\,. \tag{8}$$

A combination of the equations in (7) that is particularly useful is obtained by adding the two equations to get

$$\frac{\vec{J}_m}{2q\,D_m} - \frac{\vec{J}_M}{2q\,D_M} = \vec{\nabla}P + \frac{N}{2V_T}\vec{\nabla}U \tag{9}$$

and combining this with (8) gives

$$D^* \nabla^2 \left[P + \frac{N}{2V_T} U \right] = -g \tag{10}$$

where D^* is defined by

$$D^* \equiv \frac{2D_e D_h}{D_e + D_h} = \frac{2D_m D_M}{D_m + D_M} . \tag{11}$$

The square bracket in (10) can be expressed in terms of several functions that will each be regarded as known quantities in this analysis. The number of DRBs will be denoted K, and the DRBs will be labeled as the surfaces $S_1, \ldots, S_K$. The electrode surface will be denoted S_0. Corresponding to the ith DRB is the function denoted $\Omega^{(i)}$ defined by

$$\nabla^2 \Omega^{(i)} = 0 \quad \text{in QNR interior} \ (i = 1,\ldots,K) \tag{12a}$$

$$\Omega^{(i)} = 1 \quad \text{on } S_i \ \text{ and } \ \Omega^{(i)} = 0 \quad \text{on } S_j \text{ if } j \neq i \ (i = 1,\ldots,K : j = 0,\ldots,K) \tag{12b}$$

with reflective boundary conditions (the normal component of the gradient is zero) tacitly assumed on the reflective boundaries. The last function that is treated as a known is what the carrier density would be if carrier transport were a pure diffusion process. This function is denoted P^* and is defined by

$$D^* \nabla^2 P^* = -g \quad \text{in QNR interior} \tag{13a}$$

$$P^* = 0 \quad \text{on } S_j \quad (j = 0,\ldots,K) \tag{13b}$$

with reflective boundary conditions tacitly assumed on the reflective boundaries. We will call D^* defined by (11) the "ambipolar diffusion coefficient," and we will call (13a) the "ambipolar diffusion equation." In order to express the square bracket in (10) in terms of the functions just defined, it is necessary to introduce notation for the boundary values of P and U. The excess carrier density P is zero on the electrode surface S_0. This surface will also be selected as the reference for the potential, i.e., the potential is the potential relative to this surface, so we also have $U = 0$ on S_0. Notations used for the other boundary values are given by

$$P = 0 \ \text{ on } S_0, \quad P = P_i \ \text{ on } \ S_i \ \text{ for } i = 1,\ldots N \tag{14a}$$

$$U = 0 \ \text{ on } S_0, \quad U = U_i \ \text{ on } \ S_i \ \text{ for } i = 1,\ldots N \tag{14b}$$

with reflective boundary conditions for P and U tacitly assumed on the reflective boundaries. The solution to (10) for the square bracket that satisfies the same boundary conditions as the square bracket is given by

$$\left[P + \frac{N}{2V_T} U \right] = P^* + \sum_{i=1}^{K} \left[P_i + \frac{N}{2V_T} U_i \right] \Omega_i . \tag{15}$$

Substituting (15) into (9) gives

$$\frac{\vec{J}_m}{2q\,D_m} - \frac{\vec{J}_M}{2q\,D_M} = \vec{\nabla} P^* + \sum_{i=1}^{K} \left[P_i + \frac{N}{2V_T} U_i \right] \vec{\nabla} \Omega_i \ . \tag{16}$$

III. THE HIGH-INJECTION LIMIT

The analysis given here is intended to produce an approximation that becomes accurate under high-injection conditions, i.e., when carrier liberation is sufficiently intense. The DRs are assumed to be reverse-biased so high injection is produced by a large generation rate. Therefore, a high-injection condition can be mathematically represented as a large-g condition. The derivation given here is theoretical because it uses simplifying physical approximations (drift-diffusion equations with constant mobilities) and applies to a hypothetical device structure that is defined by these assumed governing equations. One distinction between this hypothetical structure and a real device structure is that, for the hypothetical case, it is possible to take the mathematical limit of an infinite carrier generation rate. In contrast, a real device would be destroyed before this limit is reached. However, it is assumed that the real device structure with a large but finite generation rate can be approximated by the hypothetical case with the same finite generation, so an approximation for this hypothetical case is also an approximation for the real case. The mathematical limit is an approximation for the hypothetical case having a sufficiently large but finite generation. Therefore, the limiting case is relevant to this investigation in spite of the fact that a real device structure could not survive this limit.

The task now is to give the "large-g limit" a precise definition. This is done by first selecting a non-negative (but otherwise arbitrary) reference function $g_{ref}(\boldsymbol{x})$. A large g is obtained by multiplying this reference function by a constant and positive scale factor γ that becomes large. In other words, if we define a generation rate with an adjustable scale factor by

$$g(\vec{x};\gamma) \equiv \gamma\, g_{ref}(\vec{x}) \tag{17}$$

then the large-g limit is the limit as $\gamma \to \infty$.

Because g depends on γ, the carrier density, electric field, and the currents satisfying the equations in the previous section will also depend on γ. We will be taking the limit as $\gamma \to \infty$, so we want the notation to show this dependence. Using more descriptive notation in (7) and (8) gives

$$\frac{\vec{J}_m(\vec{x};\gamma)}{q\,D_m} = \vec{\nabla}P(\vec{x};\gamma) - \frac{P(\vec{x};\gamma)}{V_T}\,\vec{\nabla}U(\vec{x};\gamma) \tag{18a}$$

$$-\frac{\vec{J}_M(\vec{x};\gamma)}{q\,D_M} = \vec{\nabla}P(\vec{x};\gamma) + \frac{P(\vec{x};\gamma)+N}{V_T}\,\vec{\nabla}U(\vec{x};\gamma) \tag{18b}$$

$$\vec{\nabla}\circ\vec{J}_m(\vec{x};\gamma) = -q\gamma\, g_{ref}(\vec{x}), \quad \vec{\nabla}\circ\vec{J}_M(\vec{x};\gamma) = q\gamma\, g_{ref}(\vec{x})\,. \tag{19}$$

There is no need to change notation in (12), because $\Omega^{(i)}$ does not depend on γ, but using more descriptive notation with (13) gives

$$D^*\,\nabla^2 P^*(\vec{x};\gamma) = -\gamma\, g_{ref}(\vec{x}) \quad \text{in QNR interior} \tag{20a}$$

$$P^*(\vec{x};\gamma) = 0 \quad \text{on } S_j \quad (j = 0,...,K)\,. \tag{20b}$$

The boundary values of the carrier density at the DRBs will be discussed at a later time, but are allowed to depend on γ. The potential at each DRB relative to the potential at the electrode will also depend on γ, so (14) is written in more descriptive notation as

$$P(\vec{x};\gamma) = 0 \text{ on } S_0, \quad P(\vec{x};\gamma) = P_i(\gamma) \text{ on } S_i \text{ for } i = 1,...N \tag{21a}$$

$$U(\vec{x};\gamma) = 0 \text{ on } S_0, \quad U(\vec{x};\gamma) = U_i(\gamma) \text{ on } S_i \text{ for } i = 1,...N\,. \tag{21b}$$

Finally, (15) and (16) are written in more descriptive notation as

$$\left[P(\vec{x};\gamma) + \frac{N}{2V_T}U(\vec{x};\gamma)\right] = P^*(\vec{x};\gamma) + \sum_{i=1}^{K}\left[P_i(\gamma) + \frac{N}{2V_T}U_i(\gamma)\right]\Omega_i(\vec{x})\,. \tag{22}$$

$$\frac{\vec{J}_m(\vec{x};\gamma)}{2q\,D_m} - \frac{\vec{J}_M(\vec{x};\gamma)}{2q\,D_M} = \vec{\nabla}P^*(\vec{x};\gamma) + \sum_{i=1}^{K}\left[P_i(\gamma) + \frac{N}{2V_T}U_i(\gamma)\right]\vec{\nabla}\Omega_i(\vec{x})\,. \tag{23}$$

Note that (19) implies that the currents become singular as $\gamma \to \infty$. This fact together with (18) implies that at any point $\boldsymbol{x}$ in the QNR, at least one of the two quantities, P or $\vec{\nabla}U$, becomes infinite as $\gamma \to \infty$. Similarly, P^* becomes singular. It is also possible that one or more boundary values of P or U might become infinite as $\gamma \to \infty$. However, (19) implies that these singularities are first order in γ, meaning that they can be removed by a normalization that divides by γ. To obtain quantities that remain finite in this limit, we define normalized quantities, indicated by script notation, by

$$\vec{\mathcal{J}}_m(\vec{x};\gamma) \equiv \frac{1}{\gamma}\vec{J}_m(\vec{x};\gamma), \quad \vec{\mathcal{J}}_M(\vec{x};\gamma) \equiv \frac{1}{\gamma}\vec{J}_M(\vec{x};\gamma) \tag{24a}$$

$$\mathcal{P}(\vec{x};\gamma) \equiv \frac{1}{\gamma}P(\vec{x};\gamma), \quad \mathcal{U}(\vec{x};\gamma) \equiv \frac{1}{\gamma}U(\vec{x};\gamma), \quad \mathcal{P}^*(\vec{x}) \equiv \frac{1}{\gamma}P^*(\vec{x};\gamma)\,. \tag{24b}$$

$$\mathcal{P}_i(\gamma) \equiv \frac{1}{\gamma}P_i(\gamma), \quad \mathcal{U}_i(\gamma) \equiv \frac{1}{\gamma}U_i(\gamma)\,. \tag{24c}$$

The normalized quantities have finite limits so we can define

$$\mathcal{P}(\vec{x};\infty) \equiv \lim_{\gamma\to\infty}\mathcal{P}(\vec{x};\gamma) = \lim_{\gamma\to\infty}\frac{1}{\gamma}P(\vec{x};\gamma) \tag{25}$$

with analogous definitions for the other normalized quantities. Also, because the normalized quantities remain finite, a second division by γ produces quantities that go to zero in the limit, i.e.,

$$\lim_{\gamma\to\infty}\frac{1}{\gamma}\vec{\jmath}_m(\vec{x};\gamma)=0,\quad \lim_{\gamma\to\infty}\frac{1}{\gamma}\vec{\jmath}_M(\vec{x};\gamma)=0 \tag{26a}$$

$$\lim_{\gamma\to\infty}\frac{1}{\gamma}\mathcal{P}(\vec{x};\gamma)=0,\quad \lim_{\gamma\to\infty}\frac{1}{\gamma}\mathcal{U}(\vec{x};\gamma)=0\,. \tag{26b}$$

Using (24) to express (18) through (23) in terms of normalized quantities gives

$$\frac{\vec{\jmath}_m(\vec{x};\gamma)}{\gamma\, q\, D_m}=\frac{1}{\gamma}\vec{\nabla}\mathcal{P}(\vec{x};\gamma)-\frac{\mathcal{P}(\vec{x};\gamma)}{V_T}\vec{\nabla}\mathcal{U}(\vec{x};\gamma) \tag{27a}$$

$$-\frac{\vec{\jmath}_M(\vec{x};\gamma)}{\gamma\, q\, D_M}=\frac{1}{\gamma}\vec{\nabla}\mathcal{P}(\vec{x};\gamma)+\frac{\mathcal{P}(\vec{x};\gamma)+\dfrac{N}{\gamma}}{V_T}\vec{\nabla}\mathcal{U}(\vec{x};\gamma) \tag{27b}$$

$$\vec{\nabla}\circ\vec{\jmath}_m(\vec{x};\gamma)=-q\, g_{ref}(\vec{x}),\quad \vec{\nabla}\circ\vec{\jmath}_M(\vec{x};\gamma)=q\, g_{ref}(\vec{x})\,. \tag{28}$$

$$D^*\nabla^2\mathcal{P}^*(\vec{x})=-g_{ref}(\vec{x})\quad \text{in QNR interior} \tag{29a}$$

$$\mathcal{P}^*(\vec{x})=0\quad \text{on } S_j\quad (j=0,\ldots,K)\,. \tag{29b}$$

$$\mathcal{P}(\vec{x};\gamma)=0\ \text{ on } S_0,\quad \mathcal{P}(\vec{x};\gamma)=\mathcal{P}_i(\gamma)\ \text{ on }\ S_i\ \text{ for } i=1,\ldots N \tag{30a}$$

$$\mathcal{U}(\vec{x};\gamma)=0\ \text{ on } S_0,\quad \mathcal{U}(\vec{x};\gamma)=\mathcal{U}_i(\gamma)\ \text{ on }\ S_i\ \text{ for } i=1,\ldots N\,. \tag{30b}$$

$$\left[\mathcal{P}(\vec{x};\gamma)+\frac{N}{2V_T}\mathcal{U}(\vec{x};\gamma)\right]=\mathcal{P}^*(\vec{x})+\sum_{i=1}^{K}\left[\mathcal{P}_i(\gamma)+\frac{N}{2V_T}\mathcal{U}_i(\gamma)\right]\Omega_i(\vec{x})\,. \tag{31}$$

$$\frac{\vec{\jmath}_m(\vec{x};\gamma)}{2q\, D_m}-\frac{\vec{\jmath}_M(\vec{x};\gamma)}{2q\, D_M}=\vec{\nabla}\mathcal{P}^*(\vec{x})+\sum_{i=1}^{K}\left[\mathcal{P}_i(\gamma)+\frac{N}{2V_T}\mathcal{U}_i(\gamma)\right]\vec{\nabla}\Omega_i(\vec{x})\,. \tag{32}$$

Taking the limit as $\gamma\to\infty$ of either equation in (27) while using (26) gives

$$\mathcal{P}(\vec{x})\vec{\nabla}\mathcal{U}(\vec{x})=0 \tag{33}$$

where the notation was shortened by writing $\mathcal{P}(\mathbf{x})$ in place of $\mathcal{P}(\mathbf{x};\infty)$ and writing $\mathcal{U}(\mathbf{x})$ in place of $\mathcal{U}(\mathbf{x};\infty)$. This notation denotes normalized quantities in the high-injection limit. Using similar notation (e.g., writing $\mathcal{P}_i$ in place of $\mathcal{P}_i(\infty)$) when taking the limits of (28) through (32) gives

$$\vec{\nabla}\circ\vec{\jmath}_m(\vec{x})=-q\, g_{ref}(\vec{x}),\quad \vec{\nabla}\circ\vec{\jmath}_M(\vec{x})=q\, g_{ref}(\vec{x}) \tag{34}$$

$$D^* \nabla^2 \mathcal{P}^*(\vec{x}) = -g_{ref}(\vec{x}) \quad \text{in QNR interior} \tag{35a}$$

$$\mathcal{P}^*(\vec{x}) = 0 \quad \text{on } S_j \quad (j = 0,\ldots,K) \tag{35b}$$

$$\mathcal{P}(\vec{x}) = 0 \text{ on } S_0, \quad \mathcal{P}(\vec{x}) = \mathcal{P}_i \text{ on } S_i \text{ for } i = 1,\ldots N \tag{36a}$$

$$\mathcal{U}(\vec{x}) = 0 \text{ on } S_0, \quad \mathcal{U}(\vec{x}) = \mathcal{U}_i \text{ on } S_i \text{ for } i = 1,\ldots N \tag{36b}$$

$$\left[\mathcal{P}(\vec{x}) + \frac{N}{2V_T}\mathcal{U}(\vec{x})\right] = \mathcal{P}^*(\vec{x}) + \psi(\vec{x}) \tag{37}$$

$$\frac{\vec{\jmath}_m(\vec{x})}{2q\,D_m} - \frac{\vec{\jmath}_M(\vec{x})}{2q\,D_M} = \vec{\nabla}\mathcal{P}^*(\vec{x}) + \vec{\nabla}\psi(\vec{x}) \tag{38}$$

where we further shortened the notation by defining

$$\psi(\vec{x}) \equiv \sum_{i=1}^{K}\left[\mathcal{P}_i + \frac{N}{2V_T}\mathcal{U}_i\right]\Omega_i(\vec{x})\,. \tag{39}$$

Note that ψ is the solution to Laplace's equation having a boundary value equal to the square bracket in (39) on S_i (for i = 1, …, K), and having a boundary value of zero on the electrode surface S_0.

IV. REGIONAL PARTITIONING: DEFINITIONS AND EQUATIONS

Note that (33) can be used to define a partitioning of the QNR. At any given point $\boldsymbol{x}$ in the QNR, either the normalized carrier density is zero or the normalized electric field is zero in the high-injection limit. The ambipolar region (AR) (the motivation for this name will be clear later) is defined to be the set of points in the QNR interior for which the normalized carrier density is positive in the high-injection limit. It will be argued later that the AR is never an empty set, i.e., there will always be some points in the QNR such that $P > 0$. The high-resistance region (HRR) (the motivation for this name will be clear later), when it exists, is defined to be the set of points in the QNR interior for which the normalized carrier density is zero in the high-injection limit. If there are no such points in the QNR interior, then the HRR is an empty set, i.e., there is no HRR. It will be seen later that there may or may not be an HRR, depending on the example.

Let us now consider the AR in more detail. Recall that either the normalized carrier density is zero or the normalized electric field is zero in the high-injection limit, but the normalized carrier density is not zero in the AR, by definition of the AR. Hence, the normalized electric field is zero in the AR in the high-injection limit, so the AR is characterized by

$$\mathcal{P}(\vec{x}) > 0 \qquad \text{(in AR)} \tag{40}$$

$$\vec{\nabla}\mathcal{U}(\vec{x}) = 0 \qquad \text{(in AR)}\ . \tag{41}$$

Note that (41) suggests that $\mathcal{U}$ is constant in the AR. However, the AR is not necessarily (or at least has not yet been shown to be) a connected region. It is possible that the AR might consists of several disjoint sub-regions, with $\mathcal{U}$ equal to a constant in each sub-region, but a different constant in different sub-regions. It is possible to define sub-regions (that may or may not be disjoint, and may or may not be distinct) that make up the AR, but it is necessary to first discuss carrier-density boundary conditions assumed at the DRBs. This analysis does not solve carrier transport equations within the DR, so boundary conditions at the DRB must be given in order to have a complete set of equations. A boundary condition assumed here is motivated by computer simulations of charge-collection in p-n junction silicon devices. These simulations show that, while the excess carrier density is much smaller at the DRB than at other locations in the QNR interior, the excess density at the DRB can still be much greater than the doping density when carrier generation is sufficiently intense, and the excess density on at least one of the DRBs increases with an increasing carrier-generation rate. We interpret this to mean that the normalized carrier density is greater than zero on at least one of the DRBs. It is small compared to the normalized carrier density at some interior points in the QNR, and the boundary value will be approximated as being zero in some selected equations that will be given later in Section V, but it is not exactly zero. Any DRB at which the normalized carrier density is greater than zero in the high-injection limit (there will be at least one such DRB) is surrounded by some region in which $\mathcal{P} > 0$. Hence, there will always be an AR in contact with at least one DRB (i.e., the AR is not an empty set). However, it is not clear that the un-normalized carrier density must increase without

bound with increasing γ on all DRBs (e.g., perhaps some DRBs are in contact with the HRR), i.e., $\mathcal{P}$ might be exactly zero on some DRBs. At any DRB (if there is one) such that $\mathcal{P}$ is exactly zero, there are two possibilities. One possibility is that $\mathcal{P}$ increases as the observation point moves from the DRB into the QNR interior, in which case the DRB is in contact with the AR. The other possibility is that $\mathcal{P}$ is zero throughout some region surrounding the DRB, in which case the DRB is not in contact with the AR (it is in contact with the HRR).

For each $i = 1, \ldots, K$, a subset of the AR is denoted AR_i and is defined to be the set of all points in the AR that can be connected to S_i by a path that lies entirely within the AR. If S_i is a DRB that is not in contact with the AR, then AR_i is an empty set. This implies that every pair of points within a nonempty AR_i can be connected to each other by a path that lies within AR_i, i.e., AR_i is a connected set. If two such sets, call them AR_i and AR_j , have any points in common, then they are the same set. In this case, we will say that "S_i is connected to S_j." In other words, if AR_i and AR_j have any points in common, then there is a connected portion of the AR that covers both S_i and S_j.

The fact that a nonempty AR_i is a connected set together with (41) implies that $\mathcal{U}$ is a constant in AR_i. The boundary of this set includes S_i, where $\mathcal{U} = \mathcal{U}_i$, so

$$\mathcal{U}(\vec{x}) = \mathcal{U}_i \qquad (in \quad AR_i)\,. \tag{42a}$$

Substituting this result into (37) gives

$$\mathcal{P}(\vec{x}) = \mathcal{P}^*(\vec{x}) + \psi(\vec{x}) - \frac{N}{2V_T}\mathcal{U}_i \qquad (in \quad AR_i)\,. \tag{42b}$$

The AR is characterized by (40) through (42), with (42) describing sub-regions within the AR. Note that (41) not only implies that the normalized potential is constant in each connected region within the AR, it also states that the normalized electric field is zero in the AR, suggesting that the un-normalized electric field is finite. To be technically correct, the mere fact that the limit of the un-normalized field divided by γ (i.e., the normalized field) is zero does not imply that the un-normalized field must have a finite limit, but it does have a finite limit at each interior point in the AR, as can be seen by deriving an equation for the un-normalized field. For this purpose, we write (27) as

$$\frac{\vec{\mathcal{J}}_m(\vec{x};\gamma)}{q\,D_m} = \vec{\nabla}\mathcal{P}(\vec{x};\gamma) - \gamma\frac{\mathcal{P}(\vec{x};\gamma)}{V_T}\vec{\nabla}\mathcal{U}(\vec{x};\gamma) \tag{43a}$$

$$-\frac{\vec{\mathcal{J}}_M(\vec{x};\gamma)}{q\,D_M} = \vec{\nabla}\mathcal{P}(\vec{x};\gamma) + \gamma\frac{\mathcal{P}(\vec{x};\gamma) + \dfrac{N}{\gamma}}{V_T}\vec{\nabla}\mathcal{U}(\vec{x};\gamma)\,. \tag{43b}$$

To obtain quantities that remain finite when taking the $\gamma \to \infty$ limit in the AR, where $\mathcal{P} >$ 0, it is necessary to pair the γ on the far right of each equation in (43) with the

normalized potential, because pairing γ with $\mathcal{P}$ produces a singular quantity. This converts the normalized potential into the un-normalized potential, i.e., the equations are written as

$$\frac{\vec{\jmath}_m(\vec{x};\gamma)}{q\,D_m} = \vec{\nabla}\mathcal{P}(\vec{x};\gamma) - \frac{\mathcal{P}(\vec{x};\gamma)}{V_T}\,\vec{\nabla}U(\vec{x};\gamma) \tag{44a}$$

$$-\frac{\vec{\jmath}_M(\vec{x};\gamma)}{q\,D_M} = \vec{\nabla}\mathcal{P}(\vec{x};\gamma) + \frac{\mathcal{P}(\vec{x};\gamma) + \dfrac{N}{\gamma}}{V_T}\,\vec{\nabla}U(\vec{x};\gamma)\,. \tag{44b}$$

When taking the $\gamma \to \infty$ limit, the un-normalized potential $U(\boldsymbol{x};\gamma)$ will sometimes become singular in the AR because the reference potential was taken to be at S_0 and the un-normalized potential drop across a region outside the AR can become singular (as seen later). However, this singularity is contained in an additive constant from the point of view of the AR because (44) implies that the gradient of the un-normalized potential has a finite limit in the AR. One way to interpret the limit of $\vec{\nabla}U(\vec{x};\gamma)$ is to interpret it as the limit of the gradient, which is defined in the AR. Another way to interpret the limit of $\vec{\nabla}U(\vec{x};\gamma)$ is to interpret it as the gradient of a limit, but with the reference potential changed as needed so that $U(\boldsymbol{x};\gamma)$ has a finite limit in the AR. Using either interpretation, taking the limit of (44) gives

$$\frac{\vec{\jmath}_m(\vec{x})}{q\,D_m} = \vec{\nabla}\mathcal{P}(\vec{x}) - \frac{\mathcal{P}(\vec{x})}{V_T}\,\vec{\nabla}U(\vec{x}) \qquad \text{(in AR)} \tag{45a}$$

$$-\frac{\vec{\jmath}_M(\vec{x})}{q\,D_M} = \vec{\nabla}\mathcal{P}(\vec{x}) + \frac{\mathcal{P}(\vec{x})}{V_T}\,\vec{\nabla}U(\vec{x}) \qquad \text{(in AR)} \tag{45b}$$

$$\frac{\vec{\jmath}_m(\vec{x})}{2q\,D_m} - \frac{\vec{\jmath}_M(\vec{x})}{2q\,D_M} = \vec{\nabla}\mathcal{P}(\vec{x}) \qquad \text{(in AR)} \tag{45c}$$

and substituting (42) into these results gives

$$\frac{\vec{\jmath}_m(\vec{x})}{q\,D_m} = \vec{\nabla}\left[\mathcal{P}^*(\vec{x}) + \psi(\vec{x})\right] - \frac{\mathcal{P}^*(\vec{x}) + \psi(\vec{x}) - \dfrac{N}{2V_T}\mathcal{V}_i}{V_T}\,\vec{\nabla}U(\vec{x}) \qquad (in \quad AR_i) \tag{46a}$$

$$-\frac{\vec{\jmath}_M(\vec{x})}{q\,D_M} = \vec{\nabla}\left[\mathcal{P}^*(\vec{x}) + \psi(\vec{x})\right] + \frac{\mathcal{P}^*(\vec{x}) + \psi(\vec{x}) - \dfrac{N}{2V_T}\mathcal{V}_i}{V_T}\,\vec{\nabla}U(\vec{x}) \qquad (in \quad AR_i)\,. \tag{46b}$$

Combining either equation in (46) with the corresponding equation in (28), while using (35) and the fact that ψ satisfies Laplace's equation, produces the same equation governing $U(\boldsymbol{x})$, which is

$$\vec{\nabla}\circ\left\{\left[\mathcal{P}^*(\vec{x}) + \psi(\vec{x}) - \frac{N}{2V_T}\mathcal{V}_i\right]\frac{\vec{\nabla}U(\vec{x})}{V_T}\right\} = \left[\frac{1}{2D_m} - \frac{1}{2D_M}\right]g_{ref}(\vec{x}) \qquad (in \quad AR_i)\,. \tag{47}$$

In summary, the AR is characterized by (40), (41), (42), (46), and (47). The name "ambipolar region" was chosen because (42), together with the fact that ψ satisfies Laplace's equation, implies that the normalized carrier density satisfies the same ambipolar diffusion equation that is satisfied by $\mathcal{P}^*$ and given by (35a).

We now consider the HRR in more detail. Recall that the HRR (when it exists) is defined to be the set of points in the QNR interior for which the normalized carrier density is zero in the high-injection limit. Depending on the specific example, there might not be any points in the QNR interior satisfying this condition, in which case there is no HRR and the entire QNR is the AR. That the points be interior points in the QNR is an essential part of the definition of the HRR, because there would otherwise always be an HRR containing (at least) the substrate electrode boundary. Combining $\mathcal{P} = 0$ with (37), we conclude that the HRR (when it exists) is characterized by

$$\mathcal{P}(\vec{x}) = 0 \qquad \text{(in HRR)} \tag{48}$$

$$\frac{N}{2V_T}\mathcal{U}(\vec{x}) = \mathcal{P}^*(\vec{x}) + \psi(\vec{x}) \qquad \text{(in HRR)}\,. \tag{49}$$

Note that defining the HRR to consist of interior points when it exists insures that the HRR has a nonzero width when it exists. This implies that (48) and (49) apply to an interval (as opposed to just a boundary), which implies that it is valid to take the gradients of these equations within the HRR when the HRR exists.

The fact that the normalized carrier density is zero in the HRR suggests that the un-normalized carrier density is finite. That this quantity is finite is seen by returning to (43). To obtain quantities that remain finite when taking the $\gamma \to \infty$ limit in the HRR, where $\vec{\nabla}\mathcal{U}(\vec{x}) \neq 0$ as seen by (49), it is necessary to pair the γ on the far right of each equation in (43) with the normalized carrier density, because pairing γ with the $\vec{\nabla}\mathcal{U}(\vec{x})$ produces a singular quantity. This converts the normalized carrier density into the un-normalized carrier density. Taking the $\gamma \to \infty$ limit gives

$$\frac{\vec{\jmath}_m(\vec{x})}{q\,D_m} = \vec{\nabla}\mathcal{P}(\vec{x}) - \frac{P(\vec{x})}{V_T}\,\vec{\nabla}\mathcal{U}(\vec{x})$$

$$-\frac{\vec{\jmath}_M(\vec{x})}{q\,D_M} = \vec{\nabla}\mathcal{P}(\vec{x}) + \frac{P(\vec{x}) + N}{V_T}\,\vec{\nabla}\mathcal{U}(\vec{x})\,.$$

As previously stated, it is valid to take the gradient of (48) and (49) within the HRR when the HRR exists, so we conclude that the gradient of $\mathcal{P}$ is zero in the HRR. Using this fact with the above equations, and also substituting (49) into the above equations, gives

$$\frac{\vec{\jmath}_m(\vec{x})}{2q\,D_m} = -\frac{P(\vec{x})}{N}\,\vec{\nabla}\left[\mathcal{P}^*(\vec{x}) + \psi(\vec{x})\right] \qquad \text{(in HRR)} \tag{50a}$$

$$-\frac{\vec{\jmath}_M(\vec{x})}{2q\,D_M} = \frac{P(\vec{x})+N}{N}\,\vec{\nabla}\left[\mathcal{P}^*(\vec{x})+\psi(\vec{x})\right] \qquad \text{(in HRR)}\,. \tag{50b}$$

Combining either equation in (50) with the corresponding equation in (28), while using (35) and the fact that ψ satisfies Laplace's equation, produces the same equation governing $P(\boldsymbol{x})$, which is

$$\vec{\nabla}\circ\left\{\left[P(\vec{x})+\frac{D_M}{D_m+D_M}N\right]\vec{\nabla}\left[\mathcal{P}^*(\vec{x})+\psi(\vec{x})\right]\right\}=0 \qquad \text{(in HRR)}\,. \tag{51}$$

In summary, the HRR is characterized by (48) through (51). The name "high-resistance region" was chosen because the carrier density remains finite (hence the electrical conductivity remains finite) in the HRR, compared to the infinite (in the $\gamma\to\infty$ limit) un-normalized carrier density (hence an infinite conductivity) in the AR. All of the electrical resistance associated with the QNR is in the HRR in the $\gamma\to\infty$ limit. This is consistent with another property, which is that the normalized electric field is nonzero (the un-normalized electric field is infinite in the $\gamma\to\infty$ limit) in the HRR, compared to a zero normalized field (a finite un-normalized field) in the AR.

If there is an HRR, then part of the boundary of the HRR will be the demarcation between the HRR and AR_i. This boundary will be denoted ARB_i and its location is determined by setting $\mathcal{P}(\mathbf{x})$ equal to zero in (42) to get

$$\mathcal{P}^*(\vec{x})+\psi(\vec{x})=\frac{N}{2V_T}\mathcal{U}_i \qquad (on \quad ARB_i)\,. \tag{52}$$

If there is no HRR, we will define ARB_i to be the electrode boundary S_0. With this definition, (52) will also apply when there is no HRR because the normalized electric field will be zero throughout the QNR, implying that $\mathcal{U}_i = 0$. We also have $\psi = 0$ and $\mathcal{P}^* = 0$ on S_0, so (52) correctly gives S_0 as the boundary.

The un-normalized electric field is finite at any point in the AR interior, but this does mean that this field is bounded in the AR. It is not bounded, which can be seen by selecting some point, call it $\boldsymbol{x}_0$, on the ARB. If there is no HRR, the ARB is S_0. If there is an HRR, the ARB is the demarcation between the AR and HRR. For either case, the normalized carrier density is zero at $\boldsymbol{x}_0$. Now select a point $\boldsymbol{x}$ in the AR interior so that (45) applies and implies

$$\frac{\vec{\jmath}_m(\vec{x})}{2q\,D_m}+\frac{\vec{\jmath}_M(\vec{x})}{2q\,D_M}=-\frac{\mathcal{P}(\vec{x})}{V_T}\,\vec{\nabla}U(\vec{x}) \qquad \text{(in AR)}$$

Now take the limit as the point $\boldsymbol{x}$ approaches the point $\boldsymbol{x}_0$ along a path in the AR interior. Excluding special cases, the limit of the left side will not be zero, but the $\mathcal{P}(\boldsymbol{x})$ on the right does approach zero, implying that the gradient of U increases without bound, i.e., the un-normalized electric field is not bounded in the AR. In contrast, the relevant electric field

in the HRR is the normalized electric field, which is not only finite at each point in the HRR interior, but also bounded in the HRR as implied by (49).

V. RELATING TERMINAL CURRENTS TO POTENTIAL BOUNDARY VALUES UNDER IDEAL BOUNDARY CONDITIONS

The goal is to derive equations expressing terminal currents in terms of other quantities, where the normalized terminal currents are defined by

$$I_{m,i} \equiv -\int_{Si} \vec{\mathcal{J}}_m \circ d\vec{S}, \quad I_{M,i} \equiv -\int_{Si} \vec{\mathcal{J}}_M \circ d\vec{S}, \quad I_{T,i} \equiv I_{m,i} + I_{M,i} \tag{53}$$

which applies to any of the surfaces S_0, S_1, …, S_K. The normal unit vector in the surface integrals is directed outward from the QNR because this is the customary convention when using Green's theorem. The signs in front of the integrals were selected to produce positive currents at each reverse-biased DR. For example, minority carriers will be flowing towards the DRB from the QNR interior, which is the direction of the outer-normal unit vector in the surface integral, but the direction of $\mathcal{J}_m$ is opposite to the direction of minority-carrier flow, so it is opposite to the direction of the normal unit vector; hence, a negative sign in front of the integral produces a positive current.

If the drift-diffusion equations had been completely solved, there would be enough information in that solution to calculate the minority and majority currents (and total currents) at each boundary when the carrier density and potential boundary values are given at all boundaries. Alternatively, if we are given the carrier density at each boundary and majority current (instead of potential) at each boundary, a complete solution would be able to calculate the total current and potential at each boundary from this given information. We have not yet obtained a complete solution. In terms of the original (un-normalized) quantities, the particular combination of carrier density and potential appearing on the left side of (22) has been solved (when all quantities on the right are regarded as known), but we have not yet solved for each quantity individually because we have not yet solved all of the differential equations that have been listed. Because the solution is not yet complete, we do not expect to be able to calculate (using only the analysis given so far) the total currents and potential boundary values, even after numerical values have been assigned to the majority currents and carrier density boundary values. Additional analysis in later sections is needed for that. However, the analysis given so far is at least enough to determine constraints between total currents and potential boundary values after numerical values have been assigned to the majority currents and carrier density boundary values. These constraints are equations relating the total currents to the potential boundary values, and some useful information can be extracted from these equations in spite of the fact that these constraints, by themselves, are not enough to uniquely solve for the total currents and potential boundary values individually.

A complete solution to the drift-diffusion equations would uniquely solve for all quantities after values have been assigned to each majority current and carrier density boundary value, so specifying any additional boundary conditions would be an over-specification of boundary conditions, but we are at liberty to assign a value to each majority current and carrier density boundary value. We will do this and then derive a partial solution, which is a set of equations relating the total currents to the potential

boundary values. The boundary values used as inputs are simplified approximations that are intended to have adequate accuracy when each DR is one-sided (n^+-p when the substrate is p-type, or p^+-n when the substrate is n-type) and is reverse biased. Such a junction blocks the majority current, so we use $I_{M,i} = 0$ when $i = 1, \ldots, K$. An approximation for the carrier density boundary values is motivated by computer simulations which show that, while the excess density at the DRB can be much greater than the doping density when carrier generation is sufficiently intense, the excess carrier density is still much smaller at the DRB (when the DR is reverse biased) than at the location in the QNR at which the density is maximum. This is interpreted to mean that $\mathcal{P}_i$ can be neglected compared to the other term in the square bracket on the right side of (39). In summary, the assumed boundary conditions for the DRBs, which will be called "ideal boundary conditions" (IBC) are

$$I_{M,i} = 0 \quad \text{and} \quad \mathcal{P}_i = 0 \quad \text{for each} \quad i = 1,\ldots,K \qquad \text{(IBC)} . \tag{54}$$

It should be noted that (54) is only a partial listing of the conditions that define IBC. A complete listing is given in Section VI. It should also be noted that ideal boundary conditions are not approximations that become exact in the high-injection limit. They are in the category of physical approximations, the same category as approximating mobilities as constants in the QNR. They describe an ideal reverse-biased DRB that is defined by these boundary conditions. The ideal DRB is not an exact representation of a real DRB even in the high-injection limit, but it is taken for granted that the ideal DRB resembles a real DRB close enough so that some useful physical insight can be obtained from an analysis of the ideal case.

IBC will be assumed throughout the remainder of this paper, so from this point on (39) reduces to

$$\psi(\vec{x}) = \frac{N}{2V_T} \sum_{i=1}^{K} \mathcal{U}_i \, \Omega_i(\vec{x}) . \tag{55}$$

Because (38) applies throughout the entire QNR, it can be used to relate surface integrals on any of the surfaces $S_0, S_1, \ldots, S_K$. Taking surface integrals while using (53) and (55) gives

$$\frac{I_{m,i}}{2q\,D_m} - \frac{I_{M,i}}{2q\,D_M} = -\int_{Si} \vec{\nabla}\mathcal{P}^* \circ d\vec{S} - \frac{N}{2V_T} \sum_{j=1}^{K} \mathcal{U}_j \int_{Si} \vec{\nabla}\Omega_j \circ d\vec{S} \quad for \quad i = 0,\ldots,K . \tag{56}$$

If we now restrict S_i to be any of the DRBs, i.e., $i = 1, \ldots, K$, we can use $I_{M,i} = 0$, which implies that $I_{T,i} = I_{m,i}$, to write (56) as

$$I_{T,i} = I_i^* - \sum_{j=1}^{K} \mathcal{U}_j \, C_{i,j} \quad for \quad i = 1,\ldots,K \tag{57}$$

where we shortened the notation by defining

$$I_i^* \equiv -2q\,D_m \int_{Si} \vec{\nabla}\mathcal{P}^* \circ d\vec{S} \quad for \quad i = 0,\ldots,K \tag{58}$$

$$C_{i,j} \equiv q\,D_m \frac{N}{V_T} \int_{Si} \vec{\nabla}\Omega^{(j)} \circ d\vec{S} \quad for \quad i = 1,\ldots,K; \quad j = 1,\ldots,K\ . \tag{59}$$

Regarding I_i^* defined by (58) as a known quantity describing the spatial distribution of the normalized carrier generation rate density, and $C_{i,j}$ defined by (59) as a known set of constants describing the QNR geometry, (57) becomes a set of constraints relating the normalized terminal currents to the normalized potential boundary values on the DRBs, which is the set of equations that was promised earlier in this section.

An alternate expression can be obtained for I_i^* defined by (58) by first using (12b), then (29b), then Green's theorem, and then (12a) and (29a) to get

$$-\int_{Si} \vec{\nabla}\mathcal{P}^* \circ d\vec{S} = -\oint \Omega^{(i)}\vec{\nabla}\mathcal{P}^* \circ d\vec{S} = \oint \mathcal{P}^*\vec{\nabla}\Omega^{(i)} \circ d\vec{S} - \oint \Omega^{(i)}\vec{\nabla}\mathcal{P}^* \circ d\vec{S} =$$

$$\int_{QNR} \mathcal{P}^*\nabla^2\Omega^{(i)}\, d^3x - \int_{QNR} \Omega^{(i)}\nabla^2\,\mathcal{P}^*\, d^3x = \frac{1}{D^*}\int_{QNR} \Omega^{(i)}\, g_{ref}\, d^3x$$

Combining this result with (58) while using (11) gives

$$I_i^* = q\left(1 + \frac{D_m}{D_M}\right)\int_{QNR} \Omega^{(i)}(\vec{x})\, g_{ref}(\vec{x})\, d^3x \quad for \quad i = 0,\ldots,K\ . \tag{60}$$

Except for material constants that determine units of measure, the $C_{i,j}$ defined by (59) is the same as the coefficient of induction defined in electrostatic treatments of capacitors consisting of K+1 conductors [4]. One conductor (S_0 in our notation) defines the reference potential and has a charge that balances the sum of the charges on the remaining K conductors consisting of S_1, …, S_K. It is intentional that the i and j subscripts in $C_{i,j}$ range from 1 to K (not zero to K) so that $C_{i,j}$ will be an invertible matrix (invertibility is proven in Appendix A).

The elements of the inverse of $C_{i,j}$ are denoted here as $\Phi_{i,j}$. These elements are the same (except for material constants that determine units of measure) as the coefficients of potential defined in electrostatic treatments of capacitors consisting of multiple conductors [4]. Important properties of $C_{i,j}$ and $\Phi_{i,j}$ are that they are symmetric. This is established in (A3) in Appendix A for $C_{i,j}$. Starting with the fact that the inverse of a transpose is the transpose of the inverse, it is easy to show that the inverse of a symmetric matrix is symmetric, implying that $\Phi_{i,j}$ is also symmetric. The conclusions are

$$C_{i,j} = C_{j,i} \quad for \quad i = 1,\ldots,K \quad and \quad j = 1,\ldots K\ . \tag{61a}$$

$$\Phi_{i,j} = \Phi_{j,i} \quad for \quad i = 1,\ldots,K \quad and \quad j = 1,\ldots K\ . \tag{61b}$$

The statement that these matrices are inverses can be written as

$$\sum_{k=1}^{K} \Phi_{i,k} C_{k,j} = \delta_{i,j} \equiv \begin{cases} 1 & if \quad i = j \\ 0 & if \quad i \neq j \end{cases} \quad for \quad i = 1,...,K \quad and \quad j = 1,...,K \ . \tag{62}$$

This can be used to invert (57) so that it solves for the potential boundary values in terms of the terminal currents instead of vice-versa. Changing the i subscript in (57) to k, multiplying the resulting equation by $\Phi_{i,k}$, summing in k, and then using (62) gives

$$\mho_i = \sum_{k=1}^{K} \left(I_k * - I_{T,k} \right) \Phi_{i,k} \quad for \quad i = 1,...,K \ . \tag{63}$$

VI. SEVERAL INEQUALITIES

Several inequalities derived here will be used in later sections to derive necessary and sufficient conditions for the existence of an HRR, and to derive some topological properties of the AR and HRR. The focus here is on the inequalities themselves, with applications postponed to later sections. However, some preliminary discussion is needed first before the derivation of the inequalities can proceed because (54) is a partial list but not a complete list of all of the properties that define IBC. It is necessary to list one more property, so we start the discussion by revisiting the ideal reverse-biased DR. Section IV mentioned the possibility that one or more (but not all) DRBs might be in contact with the HRR, at least as predicted by ideal boundary conditions. There are two possible ways in which a DRB can contact an HRR as predicted by ideal boundary conditions. The first possibility occurs when the condition $\mathcal{P}_i = 0$ is an inexact approximation while the exact condition is $\mathcal{P}_i > 0$, but $\mathcal{P}$ decreases as the observation point moves away from the DRB. The DRB is in contact with the AR, but the AR width is very narrow. In the IBC limit we would predict a zero AR width for such an example, but this is an approximation because the actual AR width is small but not zero. The exact description of a small but nonzero AR width allows us to substitute (45c) into the right side of

$$\frac{I_{m,i}}{2q\,D_m} - \frac{I_{M,i}}{2q\,D_M} = -\int_{Si} \left[\frac{\vec{J}_m}{2q\,D_m} - \frac{\vec{J}_M}{2q\,D_M} \right] \circ d\vec{S}$$

to get

$$\frac{I_{m,i}}{2q\,D_m} - \frac{I_{M,i}}{2q\,D_M} = -\int_{Si} \vec{\nabla}\mathcal{P} \circ d\vec{S} \,. \tag{64}$$

If we now take the IBC limit by letting $\mathcal{P}_i \to 0$, while noting that $\mathcal{P} \geq 0$ throughout the QNR, we conclude the direction of increasing $\mathcal{P}$ is opposite to the outer normal vector in the surface integral, which gives

$$\frac{I_{m,i}}{2q\,D_m} - \frac{I_{M,i}}{2q\,D_M} \geq 0$$

The second possibility is more subtle and occurs if the $\gamma \to \infty$ limit leads to the exact limiting condition $\mathcal{P}_i = 0$, and with $\mathcal{P}$ not increasing as the observation point moves away from the DRB. The final property used to define IBC states that this case is equivalent to a limit. The limit is obtained by starting with a DRB that is surrounded by an AR layer separating the DRB from the HRR, and taking the limit as the thickness of the AR layer shrinks to zero. In other words, the above inequality still applies. It also applies to any DRB in contact with the AR, so the last property used to define IBC states that the above inequality applies to every DRB. Combining this with the properties already listed in (54) gives the complete list of properties that define IBC, which are

$$I_{M,i} = 0 \quad \text{and} \quad I_{m,i} = I_{T,i} \geq 0 \quad \text{and} \quad \mathcal{P}_i = 0 \quad \text{for each} \quad i = 1,\ldots,K \qquad \text{(IBC)} \,. \tag{65}$$

One inequality is part of the definition of IBC and is $I_{m,i} = I_{T,i} \geq 0$. To obtain another inequality, go back to (64) and use $I_{M,i} = 0$, implying that $I_{m,i} = I_{T,i}$, to write (64) as

$$I_{T,i} = -2q\, D_m \int_{Si} \vec{\nabla}\mathcal{P} \circ d\vec{S} \ . \tag{66}$$

Now combine (45c) with (34) to get

$$D^* \nabla^2 \mathcal{P}(\vec{x}) = -g_{ref}(\vec{x}) \quad \text{in AR} \ . \tag{67a}$$

Under IBC we have

$$\mathcal{P}(\vec{x}) = 0 \quad \text{on } S_j \quad (j = 0,...,K) \ . \tag{67b}$$

If there is no HRR then the ARB is S_0. If there is an HRR, then the ARB includes points in the QNR interior, with these points being the union of ARB_i for each i such that AR_i is not empty. For all cases we have

$$\mathcal{P}(\vec{x}) = 0 \quad \text{on ARB} \ . \tag{67c}$$

Now compare (67) to (35). If there is no HRR, so that the AR is all of the QNR, we conclude that $\mathcal{P}$ and $\mathcal{P}^*$ satisfy identical boundary value problems, so

$$\mathcal{P}(\vec{x}) = \mathcal{P}^*(\vec{x}) \quad \text{throughout QNR if there is no HRR} \ . \tag{68}$$

Now suppose there is an HRR. First consider a DRB, call it S_i, that is in contact with the AR. This DRB is surrounded by a connected portion of the AR denoted AR_i. From (67b) and (67c) we conclude that all portions of the boundary of AR_i that are not reflective boundaries are sink boundaries. Part of the sink boundary is ARB_i. Note that $\mathcal{P}^*$ satisfies the same field equation as $\mathcal{P}$ in AR_i. Also, $\mathcal{P}^*$ satisfies the same boundary conditions as $\mathcal{P}$ on all portions of the boundary of AR_i consisting of DRBs (if any portion of the boundary consists of DRBs) or the electrode (if the boundary includes the electrode) or reflective boundaries (if the boundary includes reflective boundaries). However, if ARB_i is in the QNR interior, then this is a sink boundary for $\mathcal{P}$ but not for $\mathcal{P}^*$. The presents of this sink boundary can only reduce the surface integral on the right side of (66) compared to what the integral would be without the sink boundary, i.e.,

$$-2q\, D_m \int_{Si} \vec{\nabla}\mathcal{P} \circ d\vec{S} \le -2q\, D_m \int_{Si} \vec{\nabla}\mathcal{P}^* \circ d\vec{S}$$

and using (66) and (58) gives

$$I_{T,i} \le I_i^* \ .$$

This result was derived for a DRB that is in contact with the AR when the case considered is one in which there is an HRR. Interpreting a DRB in contact with the HRR as the limiting case of a DRB in contact with the AR, in the limit as the width of AR_i goes to zero, we not only obtain the above inequality, we obtain the stronger statement that $I_{T,i} = 0$, because the sink boundary is adjacent to the DRB. Therefore the above inequality applies to any DRB when there is an HRR. If there is no HRR, we conclude from (68)

that we not only have the above inequality (hence the above inequality applies to all DRBs with or without an HRR), we have the stronger statement that $I_{T,i} = I_i^*$. The conclusions are

$$0 \le I_{T,i} \le I_i^* \quad \text{for each} \quad i = 1,\ldots,K \tag{69}$$

$$\text{If there is no HRR then } I_{T,i} = I_i^* \quad \text{for each} \quad i = 1,\ldots,K \tag{70}$$

$$\text{If } S_i \text{ contacts the HRR for some } i = 1,\ldots,K, \text{ then } I_{T,i} = 0\,. \tag{71}$$

The next inequality involves the normalized potential boundary values. It is shown in Appendix C that the coefficients $\Phi_{i,j}$ are nonnegative (and the diagonal elements are positive). Using this fact together with (69), we conclude from (63) that

$$\mathcal{U}_i \ge 0 \quad \textit{for} \quad i = 1,\ldots,K\,. \tag{72}$$

The last topic in this section discusses the minority-carrier current at the electrode S_0. If all of S_0 is in contact with the HRR, we conclude from (50a) that the minority-carrier current through S_0 is zero because $P = 0$ on S_0. However, if some or all of S_0 is in contact with the AR, an analysis of the minority-carrier current using either (45a) or (46a) must deal with the fact that the electric field is unbounded, as pointed out at the end of Section IV. This inconvenience can be avoided by going back to the finite γ to evaluate the minority-carrier current at the electrode and then take the limit of the result as $\gamma \to \infty$. For a finite γ we can write (43a) as

$$\frac{\vec{\mathcal{J}}_m(\vec{x};\gamma)}{q\,D_m} = \vec{\nabla}\mathcal{P}(\vec{x};\gamma) - \frac{P(\vec{x};\gamma)}{V_T}\,\vec{\nabla}\mathcal{U}(\vec{x};\gamma)\,.$$

The ideal electrode is a perfect sink not only for the normalized carrier density $\mathcal{P}$, but also for the un-normalized carrier density P. Therefore the current at the electrode is given by

$$I_{m,0}(\gamma) \equiv -\int_{S0} \vec{\mathcal{J}}_m(x;\gamma) \circ d\vec{S} = -qD_m \int_{S0} \vec{\nabla}\mathcal{P}(\vec{x};\gamma) \circ d\vec{S}\,. \tag{73}$$

If the electrode is in contact with the HRR in the $\gamma \to \infty$ limit, the gradient on the right will be zero. This produces a zero current in the $\gamma \to \infty$ limit which is consistent with (50a). If the electrode is in contact with the AR in the $\gamma \to \infty$ limit, the result (73) is consistent with (45a) when taking the limit of (45a) as the observation point $\boldsymbol{x}$ moves to the electrode. In this limit, the electric field increases without bound as $\boldsymbol{x}$ moves to the electrode, but the normalized carrier density goes to zero faster than the electric field increases because the electrode is a sink for the un-normalized carrier density. (This property is not shared by an ARB separating the AR from the HRR because this ARB is a sink for the normalized carrier density but not for the un-normalized carrier density. Hence, the product of a normalized carrier density that goes to zero multiplied by an un-normalized electric field that increases without bound is more difficult to evaluate on any ARB that is not the electrode). Note that $\mathcal{P}$ is zero on the electrode and nonnegative in the

QNR interior, so the direction of increasing $\mathcal{P}$ is opposite to the direction of the outer normal vector in the surface integral in (73). This gives

$$I_{m,0}(\gamma) \geq 0$$

for all γ, and taking the limit as $\gamma \to \infty$ gives

$$I_{m,0} \geq 0 \,. \tag{74}$$

Another result implied by (74) is obtained by using the divergence theorem together with (53) and (34) to get the second equality in

$$I_{m,0} + \sum_{i=1}^{K} I_{T,i} = I_{m,0} + \sum_{i=1}^{K} I_{m,i} = q\int_{QNR} g_{ref}(\vec{x})\, d^3x \tag{75}$$

and combining this with (74) gives

$$\sum_{i=1}^{K} I_{T,i} \leq q\int_{QNR} g_{ref}(\vec{x})\, d^3x \,. \tag{76}$$

To shorten the notation, define

$$\mathcal{R} \equiv \int_{QNR} g_{ref}(\vec{x})\, d^3x \tag{77}$$

so that $\mathcal{R}$ is the normalized total rate of carrier generation in the QNR, and $q\mathcal{R}$ is the normalized charge generation rate in the QNR. This allows us to write (75) and (76) as

$$I_{m,0} + \sum_{i=1}^{K} I_{T,i} = q\mathcal{R} \tag{78a}$$

$$\sum_{i=1}^{K} I_{T,i} \leq q\mathcal{R} \,. \tag{78b}$$

VII. SOME TOPOLOGICAL PROPERTIES

This section begins with a review of some of the discussion that was given in Section IV, which is repeated here for easy reference. For each $i = 1, \ldots, K$, a subset of the AR is denoted AR_i and is defined to be the set of all points in the AR that can be connected to S_i by a path that lies entirely within the AR. If S_i is a DRB that is in contact with the HRR instead of the AR, then AR_i is an empty set. This definition implies that every pair of points within a nonempty AR_i can be connected to each other by a path that lies within AR_i, i.e., AR_i is a connected set, and this in turn implies that $\mathcal{V}(\boldsymbol{x})$ is a constant in AR_i. If two such sets, call them AR_i and AR_j , have any points in common, then they are the same set. In this case, we will say that "S_i is connected to S_j." An equivalent definition of the statement that a DRB S_i is connected to another DRB S_j is that there is a path within the AR interior that connects some point on S_i to some point on S_j. Because $\mathcal{V}(\boldsymbol{x})$ is constant on this path, and $\mathcal{V}(\boldsymbol{x}) = \mathcal{V}_i$ on S_i while $\mathcal{V}(\boldsymbol{x}) = \mathcal{V}_j$ on S_j, we conclude that if a DRB S_i is connected to another DRB S_j, then $\mathcal{V}_i = \mathcal{V}_j$. Furthermore, if a DRB S_i is connected to another DRB S_j, then $AR_i = AR_j$.

The remainder of this section derives new theorems regarding the topology of the AR and HRR. To simplify the analyses, the remainder of this paper considers "irreducible device geometries." These are geometries that are not "reducible geometries," where a reducible geometry is one in which S_0 is composed of multiple sections formed in such a way so that the DRBs can be divided into groups with one group completely surrounded by one section of S_0 and another group completely surrounded by another section. An example of a reducible geometry, which is excluded from consideration, is shown in Fig. 2. An example of an irreducible geometry, which is analyzed here, was already shown in Fig. 1. Note that there is no loss of generality by excluding reducible geometries because they decouple into separate and independent irreducible geometries. For example, the configuration consisting of S_3 and S_0 in Fig. 2 is separate and independent of the configuration consisting of S_1, S_2, and S_0, and each configuration can be treated using an analysis of irreducible geometries. An irreducible geometry has the property that there is some coupling between every pair of DRBs in the sense that all off-diagonal elements of $C_{i,j}$ are negative (not zero) and all of the elements $\Phi_{i,j}$ are positive (no elements are zero). These strict inequalities simplify the analysis and this is the motivation for confining our attention to irreducible geometries. The strict inequalities allow us to replace (B5b) in Appendix B and (C9b) in Appendix C with the stronger conditions

$$C_{i,j} < 0 \quad for \quad i = 1,\ldots,K \quad and \quad j = 1,\ldots,K \quad with \quad i \neq j \tag{79a}$$

$$0 < \Phi_{i,j} \leq \Phi_{j,j} \quad for \quad i = 1,\ldots,K \quad and \quad j = 1,\ldots,K\,. \tag{79b}$$

Implications from the condition that two DRBs are connected were already discussed, but now we derive implications from the condition that a DRB is connected to the electrode. We will say that a DRB is connected to the electrode if there is a path lying entirely within the AR that connects some point on the DRB to some point on the electrode. A topology theorem that will be proven below states that if there is any DRB

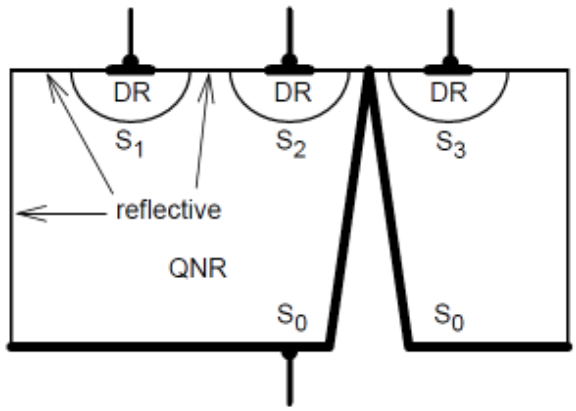

Fig. 2. Illustration of a device having a reducible geometry. This is excluded from consideration in the theory derived here. However, this problem decouples into two independent problems that can be treated by the theory derived here.

that is connected to the electrode, then there is no HRR anywhere. This implies that $\mathcal{U}_i = 0$ for each $i = 1, \ldots, K$, and that $I_{T,i} = I_i^*$ for each $i = 1, \ldots, K$. However, the proof will actually establish these conclusions in a different order. Starting with the given information that some DRB is connected to the electrode, we will first show that $I_{T,i} = I_i^*$ for each $i = 1, \ldots, K$, then show that $\mathcal{U}_i = 0$ for each $i = 1, \ldots, K$, and finally show that there is no HRR.

To prove the above conclusions, we start with the given information that there is some DRB, call it S_j for some $j = 1, \ldots, K$, with the property that there is some path in the AR that connects some point on S_j to some point on S_0. This path is a connected set of points so (41) implies that $\mathcal{U}_j$ is constant along this path. Note that $\mathcal{U}(\boldsymbol{x})$ equals zero on S_0 and it equals $\mathcal{U}_j$ on S_j, so the first conclusion is that $\mathcal{U}_j = 0$, and that $\mathcal{U}(\boldsymbol{x}) = 0$ along the path. Using $\mathcal{U}_j = 0$ with (63) gives

$$0 = \sum_{i=1}^{K} \left(I_i^* - I_{T,i}\right)\Phi_{j,i} .$$

From (79b) and (69), we see that the right side of the above equation is a sum of nonnegative terms, so each term in the sum must be zero. But the $\Phi_{i,j}$ coefficients are all positive, so each parenthesis on the right must be zero, i.e.,

$$I_{T,i} = I_i^* \quad \text{for each} \quad i = 1, \ldots, K .$$

Substituting this result back into (63) gives

$$\mathcal{U}_i = 0 \quad \text{for each} \quad i = 1, \ldots, K$$

and (55) gives $\psi(\boldsymbol{x}) = 0$ throughout the QNR. Therefore, (42b) reduces to

$$\mathcal{P}(\vec{x}) = \mathcal{P}^*(\vec{x}) \qquad \text{in} \quad AR_j$$

We next show that there is no HRR by contradiction. Assume that there is an HRR. Then there will also be a boundary, denoted ARB_j, that is in the QNR interior and separates AR_j from the HRR, and satisfies (52). Using $\psi(\boldsymbol{x}) = 0$ and $\mathcal{V}_j = 0$, (52) reduces to

$$\mathcal{P}^*(\vec{x}) = 0 \qquad \text{on} \quad ARB_i\,.$$

However, there is no boundary in the QNR interior satisfying this condition, because $\mathcal{P}^*(\boldsymbol{x}) > 0$ in the QNR interior. Therefore there is no HRR.

The conclusions from the above paragraph are:

If any DRB is connected to the electrode, then
(a) There is no HRR (hence, all DRBs are connected to the electrode). (80)
(b) $\mathcal{V}_i = 0$ and $I_{T,i} = I_i^*$ for each $i = 1,\ldots,K$.

An important implication of (80) is that there are only two possibilities. Either there is no HRR or the electrode is entirely covered by the HRR. We cannot have a situation in which part of the electrode contacts the HRR while another part does not, because the uncovered part establishes a connection with a DRB, which implies that there is no HRR. However, this conclusion was derived for an irreducible geometry in the $\gamma \to \infty$ limit. This limit may or may not be a good approximation for a large but finite γ because, depending on the geometry, there might be a problem with competing limits. For example, consider Fig. 2 and suppose that the point where S_0 contacts the upper plane is displaced slightly downward to produce a small gap between S_0 and the upper plane. The geometry now becomes irreducible. However, if the gap is small enough, we can expect almost no communication between the two sides of the device. In particular, if carrier liberation is confined to the left side, we might expect an HRR to cover the left side of S_0 without covering the right side, contradicting the claim that S_0 is completely covered if it is covered at all. This is a problem of competing limits. If the gap shrinks to zero first and then $\gamma \to \infty$ (a reducible geometry) we obtain a different result then we obtain if $\gamma \to \infty$ first (an irreducible geometry) and then the gap shrinks to zero. Stated another way, the smaller the gap is, the larger γ must be in order to be well approximated by the $\gamma \to \infty$ limit. From a practical point of view, it is recommended that the small-gap case, i.e., an irreducible geometry that is "almost reducible," be approximated as a reducible geometry. This will result in the $\gamma \to \infty$ limit being a better approximation of the finite γ case.

Note that there will always be an AR covering at least one DRB. Also, it was concluded in this section that if any DRB is connected to the electrode then there is no HRR. Furthermore, either there is no HRR or the electrode is entirely covered by the HRR. These conclusions limit the possible topologies of the AR and HRR. For illustration, consider a device containing two DRBs. There are five possible topologies for the AR and HRR, which are shown in Fig. 3. For the first possibility (Fig. 3a), there is no HRR, so the AR_1, the AR_2, the AR, and the QNR are identical sets. For the second

possibility (Fig. 3b), there is an HRR that covers S_2 as well as S_0, so AR_1 is all of the AR while AR_2 is an empty set. The third possibility (Fig. 3c) is the same as the second but with S_1 and S_2 interchanged. For the fourth possibility (Fig. 3d) there is an HRR but S_1 and S_2 are connected so the AR_1, the AR_2, and the AR are identical sets. For the fifth and last possibility (Fig. 3e), there is an HRR, and the AR_1 is disconnected from the AR_2, with neither set being empty.

The conclusion that an HRR covers the entire electrode when an HRR exists has an important implication regarding the minority-carrier current at the electrode. If there is an HRR, then it covers the entire electrode, so (50a) applies everywhere on the electrode. Because $P = 0$ on the electrode, there is no minority-carrier current through any part of the electrode. The conclusion is:

$$\text{If there is an HRR, then } I_{m,0} = 0. \tag{81a}$$

Using (78a), this can also be written as

$$\text{If there is an HRR, then } \sum_{i=1}^{K} I_{T,i} = q\mathcal{R}\,. \tag{81b}$$

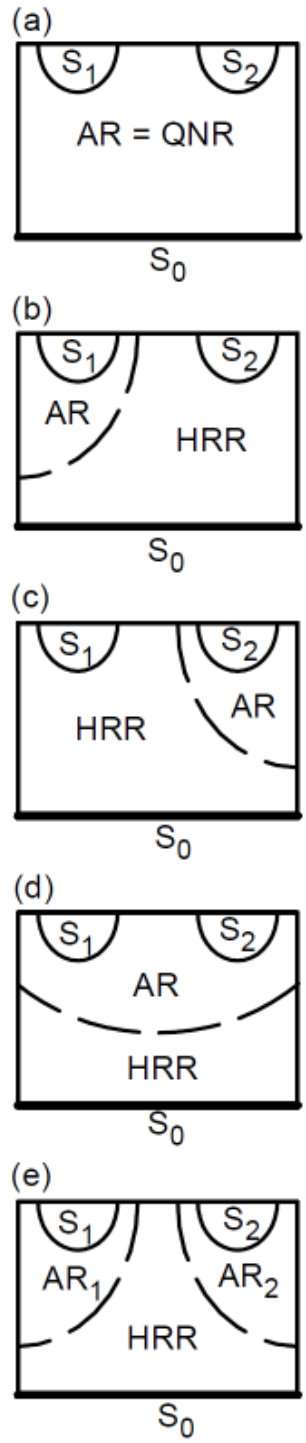

Fig. 3. The five possible topologies for an irreducible geometry with two junctions. See text for explanation.

VIII. NECESSARY AND SUFFICIENT CONDITIONS FOR THE EXISTENCE OF AN HRR

A necessary condition for the existence of an HRR can be derived by assuming that there is an HRR. From (81b) we conclude that

$$\sum_{i=1}^{K} I_{T,i} = q\mathcal{R}$$

while (69) gives

$$\sum_{i=1}^{K} I_{T,i} \leq \sum_{i=1}^{K} I_i^* .$$

Combining these results, we conclude that if there is an HRR then

$$\sum_{i=1}^{K} I_i^* \geq q\mathcal{R} .$$

Inverting this implication gives

$$\text{If } \sum_{i=1}^{K} I_i^* < q\mathcal{R} \text{ then there is no HRR.} \tag{82a}$$

Now assume that there is no HRR. Using (70) gives

$$\sum_{i=1}^{K} I_{T,i} = \sum_{i=1}^{K} I_i^* .$$

Combining this with (78b), we conclude that if there is no HRR then

$$\sum_{i=1}^{K} I_i^* \leq q\mathcal{R} .$$

Inverting this implication gives

$$\text{If } \sum_{i=1}^{K} I_i^* > q\mathcal{R} \text{ then there is an HRR.} \tag{82b}$$

IX. THE TWO-TERMINAL PROBLEM

The two-terminal problem is a K =1 problem with the DRB being one terminal and the electrode the other. With only one DRB, we will omit subscripts and write I_T and I^* instead of $I_{T,1}$ and I_1^*. Regarding I^* and the volume integral of $g_{ref}(\boldsymbol{x})$ as known quantities (note that I^* can be solved from (60)), enough information has been listed to solve for I_T. There are two cases to consider. The first is defined by

$$I^* > q\mathcal{R} \text{ (defines Case 1)} \tag{83a}$$

which implies an HRR via (82b), so (81b) gives

$$I_T = q\mathcal{R} \text{ (for Case 1)} . \tag{83b}$$

The second case is defined by

$$I^* < q\mathcal{R} \text{ (defines Case 2)} \tag{84a}$$

which implies no HRR via (82a), so (70) gives

$$I_T = I^* \text{ (for Case 2)} . \tag{84b}$$

In view of the discussion in the second paragraph in Section V, it might be surprising that the current can be solved for the two-terminal problem. In terms of the original (un-normalized) quantities, the only equation that has been solved is the equation corresponding to one combination of the two currents, with that combination being the left side of (23), and the solution solves for only one combination of P and U, with that combination being the left side of (22). The two quantities P and U were not individually solved because in order to do so it would be necessary to solve the equation corresponding to another, independent, combination of the currents, and that was not done (equations were listed but not solved). The complete solution to the boundary-value problem has not yet been obtained. When taking the limit as $\gamma \to \infty$ while recognizing a regional partitioning, the fact that the solution is not yet complete is reflected by the fact that the un-normalized electric field in the AR, and the un-normalized carrier density in the HRR, have not yet been solved (equations were listed for these quantities but the equations were not yet solved). As previously stated, it is necessary to solve the equation corresponding to some other combination of the currents, a combination that is independent of the combination on the left side of (23), to have a complete solution. Until that is done, there is one equation missing. This leads to the question of how it was possible to completely solve the two-terminal problem when there is an equation missing. The answer under Case-1 conditions is that we actually did supply an additional equation. The equation associated with the minority-carrier current, which is obtained by combining (18a) with the left equation in (19), was not completely solved, but we did extract enough information from these equations to conclude that $I_{m,0} = 0$. When this supplementary equation is included, we have a complete set of equations that were able to completely solve the two-terminal problem under Case-1 conditions. Unfortunately,

this approach does not generalize to multiple DRBs if the goal is to solve for each individual current $I_{T,i}$ (but the sum of these currents can be solved using the two-terminal analysis). For K DRBs, we would need K supplementary equations, and we only have one. The problem of multiple DRBs is going to require further analysis.

However, under Case-2 conditions, the approach used for the two-terminal problem does generalize to multiple DRBs. For the two-terminal problem we have (84b) under Case-2 conditions, and for multiple DRBs we have (70). The reason that it was possible to completely solve the problem under Case-2 conditions when there is an equation missing is that, while the equation associated with the minority-carrier current was not completely solved, we did extract enough information from it to conclude that $I_{m,0} \geq 0$. Case 2 is unique in that this small amount of supplementary information, when combined with the equations that were solved, is enough to conclude that $\mathcal{U}_i = 0$ and $I_{T,i} = I^*$ for each $i = 1, \ldots, K$.

X. THE THREE-TERMINAL PROBLEM

The three-terminal problem is a K =2 problem, i.e., there are two DRBs and an electrode. For this case, (57) can be written as

$$I_1^* = I_{T,1} + \mathcal{V}_1 C_{1,1} + \mathcal{V}_2 C_{1,2} \tag{85a}$$

$$I_2^* = I_{T,2} + \mathcal{V}_1 C_{2,1} + \mathcal{V}_2 C_{2,2} \,. \tag{85b}$$

A particular combination of these equations will be useful later in the analysis. This is obtained by multiplying the first equation by $C_{2,1} + C_{2,2}$, multiplying the second equation by $C_{1,1} + C_{1,2}$, and subtracting. The result is

$$\lfloor C_{1,1} + C_{1,2} \rfloor I_2^* - \lfloor C_{2,1} + C_{2,2} \rfloor I_1^* = \lfloor C_{1,1} + C_{1,2} \rfloor I_{T,2} - \lfloor C_{2,1} + C_{2,2} \rfloor I_{T,1} +$$

$$\lfloor C_{1,1} + C_{1,2} \rfloor \lfloor \mathcal{V}_1 C_{2,1} + \mathcal{V}_2 C_{2,2} \rfloor - \lfloor C_{2,1} + C_{2,2} \rfloor \lfloor \mathcal{V}_1 C_{1,1} + \mathcal{V}_2 C_{1,2} \rfloor .$$

Combining terms containing $\mathcal{V}_1$ and $\mathcal{V}_2$ gives

$$\lfloor C_{1,1} + C_{1,2} \rfloor I_2^* - \lfloor C_{2,1} + C_{2,2} \rfloor I_1^* = \lfloor C_{1,1} + C_{1,2} \rfloor I_{T,2} - \lfloor C_{2,1} + C_{2,2} \rfloor I_{T,1} +$$

$$\lfloor C_{1,1} C_{2,2} - C_{2,1} C_{1,2} \rfloor (\mathcal{V}_2 - \mathcal{V}_1) . \tag{86}$$

It was already shown, and stated in (82a), that there is no HRR if $I_1^* + I_2^* < q\mathcal{R}$. Furthermore, if this condition is satisfied, we conclude from (70) that $I_{T,1} = I_1^*$ and $I_{T,2} = I_2^*$. A more difficult case occurs when there is an HRR. This is more difficult because there are different possible topologies (S_1 might be connected to S_2, or it might not be) and the currents are calculated differently for different topologies. Instead of deducing the topology from a given set of conditions, we will work in the other direction. We will assume a topology and then determine the conditions that are implied by this assumption. If these conditions are not satisfied in a given example, the assumed topology is not the topology that is present in that example. Several assumed topologies are discussed separately below. An HRR is present in all cases below, so for all of these cases (81b) gives

$$I_{T,1} + I_{T,2} = q\mathcal{R} \qquad \text{(if HRR)} \,. \tag{87}$$

A. Case 1A: S_2 Contacts the HRR

The first case considered, illustrated in Fig. 3b, is one in which there is an HRR and S_2 contacts the HRR while S_1 contacts the AR. Inverting (82a), the presence of an HRR implies that $I_1^* + I_2^* \geq q\mathcal{R}$. The conditions that define Case 1A are

$$I_1^* + I_2^* \geq q\mathcal{R} \quad \text{and} \quad S_2 \text{ contacts the HRR} \quad \text{(defines Case 1A)} \,. \tag{88}$$

Note that (71) and (87) together imply

$$I_{T,2} = 0 \quad \text{and} \quad I_{T,1} = q\mathcal{R} \quad \text{(for Case 1A)} \tag{89}$$

so (86) becomes

$$\lfloor C_{2,1} + C_{2,2} \rfloor I_1 * - \lfloor C_{1,1} + C_{1,2} \rfloor I_2 * = \lfloor C_{2,1} + C_{2,2} \rfloor q\mathcal{R} + \lfloor C_{1,1}C_{2,2} - C_{2,1}C_{1,2} \rfloor (\mathcal{V}_1 - \mathcal{V}_2) \quad \text{(for Case 1A)} . \tag{90}$$

Recall from the discussion at the end of Section IX that the analysis in the main text is incomplete for the multi-junction problem, and another piece of information must be supplied in order to have a complete set of equations. The missing piece of information is extracted from (51), which is discussed in detail in Appendix D. Fortunately, it is not necessary to solve (51) for the un-normalized carrier density in the HRR. A simple conclusion from (51) provides enough information to complete the analysis for Case 1A. This conclusion, derived in Appendix D, is that floating terminals are at intermediate potentials. Using (72) and the fact that S_2 is a floating terminal (i.e., $I_{T,2} = 0$) we conclude that $0 \le \mathcal{V}_2 \le \mathcal{V}_1$ so

$$\mathcal{V}_1 - \mathcal{V}_2 \ge 0 \quad \text{(for Case 1A)} . \tag{91}$$

Note that (B5c) in Appendix B gives $C_{1,1} \ge -C_{1,2}$ and $C_{2,2} \ge -C_{2,1}$ with at least one of these inequalities being a strict inequality. Each side of each inequality is nonnegative so the inequalities can be multiplied, and using the fact that at one inequality is a strict inequality gives

$$C_{1,1}C_{2,2} - C_{2,1}C_{1,2} > 0 . \tag{92}$$

Combining (91) and (92) with (90) gives

$$\lfloor C_{2,1} + C_{2,2} \rfloor I_1 * - \lfloor C_{1,1} + C_{1,2} \rfloor I_2 * - \lfloor C_{2,1} + C_{2,2} \rfloor q\mathcal{R} \ge 0 .$$

The conclusions for Case 1A are

$$\begin{aligned} &\text{Under Case 1A conditions we have} \\ &(a) \quad \lfloor C_{2,2} + C_{2,1} \rfloor I_1 * \ge \lfloor C_{1,1} + C_{1,2} \rfloor I_2 * + \lfloor C_{2,2} + C_{2,1} \rfloor q\mathcal{R} \\ &(b) \quad I_{T,1} = q\mathcal{R} \quad \text{and} \quad I_{T,2} = 0. \end{aligned} \tag{93}$$

B. Case 1B: S_1 Contacts the HRR

The second case considered, illustrated in Fig. 3c, is one in which there is an HRR and S_1 contacts the HRR while S_2 contacts the AR. This is the same as Case 1A but with S_1 and S_2 interchanged. The conditions that define Case 1B are

$$I_1{}^* + I_2{}^* \geq q\mathcal{R} \quad \text{and} \quad S_1 \text{ contacts the HRR} \quad (\text{defines Case 1B}) . \tag{94}$$

The conclusion is the same as (93) but with S_1 and S_2 interchanged, i.e.,

$$\begin{aligned} &\text{Under Case 1B conditions we have} \\ &(a) \quad [C_{1,1} + C_{1,2}]I_2{}^* \geq [C_{2,2} + C_{2,1}]I_1{}^* + [C_{1,1} + C_{1,2}]q\mathcal{R} \\ &(b) \quad I_{T,2} = q\mathcal{R} \quad \text{and} \quad I_{T,1} = 0. \end{aligned} \tag{95}$$

C. Case 2: S_1 and S_2 are Connected

The next case considered, illustrated in Fig. 3d, is one in which there is an HRR and the AR is a connected set so that $AR_1 = AR_2$. For this case we have $\mathcal{U}_1 = \mathcal{U}_2$, so Case 2 can be defined by

$$I_1{}^* + I_2{}^* \geq q\mathcal{R} \quad \text{and} \quad \mathcal{U}_1 = \mathcal{U}_2 \quad (\text{defines Case 2}) . \tag{96}$$

For this case, (86) reduces to

$$[C_{1,1} + C_{1,2}]I_2{}^* - [C_{2,2} + C_{2,1}]I_1{}^* = [C_{1,1} + C_{1,2}]I_{T,2} - [C_{2,2} + C_{2,1}]I_{T,1} \quad (\text{for Case 2}) . \tag{97}$$

Note that $I_{T,1}$ and $I_{T,2}$ will each be strictly greater than zero for Case 2, and combining this fact with (87) gives

$$0 < I_{T,1} < q\mathcal{R} \quad \text{and} \quad 0 < I_{T,2} < q\mathcal{R} \quad (\text{for Case 2}) . \tag{98}$$

Note that (B5c) states that each square bracket on the right side of (97) is nonnegative, and at least one of them is positive. Combining this fact with $I_{T,1} > 0$ and $I_{T,2} < q\mathcal{R}$ gives the two inequalities

$$[C_{1,1} + C_{1,2}]I_{T,2} \leq [C_{1,1} + C_{1,2}]q\mathcal{R} \quad \text{and} \quad -[C_{2,1} + C_{2,2}]I_{T,1} \leq 0$$

with at least one of these inequalities being a strict inequality so adding inequalities gives the strict inequality

$$[C_{1,1} + C_{1,2}]I_{T,2} - [C_{2,1} + C_{2,2}]I_{T,1} < [C_{1,1} + C_{1,2}]q\mathcal{R} .$$

Similar steps using $I_{T,2} > 0$ and $I_{T,1} < q\mathcal{R}$ give a second inequality, and combining that result with the above inequality gives

$$-[C_{2,1} + C_{2,2}]q\mathcal{R} < [C_{1,1} + C_{1,2}]I_{T,2} - [C_{2,1} + C_{2,2}]I_{T,1} < [C_{1,1} + C_{1,2}]q\mathcal{R} .$$

Using this with (97) gives

$$-\lfloor C_{2,1}+C_{2,2}\rfloor q\mathcal{R} < \lfloor C_{1,1}+C_{1,2}\rfloor I_2{}^* - \lfloor C_{2,1}+C_{2,2}\rfloor I_1{}^* < \lfloor C_{1,1}+C_{1,2}\rfloor q\mathcal{R} \quad \text{(for Case 2)}\,. \tag{99}$$

To solve for the currents, note that (87) and (97) are two simultaneous equations containing $I_{T,1}$ and $I_{T,2}$. Solving these equations produces the bottom two entries in the conclusion below that summarizes Case 2 results.

Under Case 2 conditions we have

$$\begin{aligned}
&(a)\quad \lfloor C_{1,1}+C_{1,2}\rfloor I_2{}^* < \lfloor C_{2,2}+C_{2,1}\rfloor I_1{}^* + \lfloor C_{1,1}+C_{1,2}\rfloor q\mathcal{R}\\
&(b)\quad \lfloor C_{2,2}+C_{2,1}\rfloor I_1{}^* < \lfloor C_{1,1}+C_{1,2}\rfloor I_2{}^* + \lfloor C_{2,2}+C_{2,1}\rfloor q\mathcal{R}\\
&(c)\quad I_{T,1} = \frac{\lfloor C_{2,2}+C_{2,1}\rfloor I_1{}^* + \lfloor C_{1,1}+C_{1,2}\rfloor (q\mathcal{R} - I_2{}^*)}{\lfloor C_{1,1}+C_{1,2}\rfloor + \lfloor C_{2,2}+C_{2,1}\rfloor}\\
&(d)\quad I_{T,2} = \frac{\lfloor C_{1,1}+C_{1,2}\rfloor I_2{}^* + \lfloor C_{2,2}+C_{2,1}\rfloor (q\mathcal{R} - I_1{}^*)}{\lfloor C_{1,1}+C_{1,2}\rfloor + \lfloor C_{2,2}+C_{2,1}\rfloor}.
\end{aligned} \tag{100}$$

The denominator in items (c) and (d) in (100) are positive according to (B5c). Note that item (a) insures that the right side of item (c) is positive and that the right side of item (d) is less than $q\mathcal{R}$, while item (b) insures that the right side of item (d) is positive and that the right side of item (c) is less than $q\mathcal{R}$. Also, by combining $I_1{}^* + I_2{}^* \geq q\mathcal{R}$ with item (c), we verify consistency with $I_{T,1} \leq I_1{}^*$. Similar steps used with item (d) verify that $I_{T,2} \leq I_2{}^*$. Note that the conditions on $I_1{}^*$ and $I_2{}^*$ for the three cases (item (a) in (93) for Case 1A, item (a) in (95) for Case 1B, and items (a) and (b) together in (100) for Case 2) are mutually exclusive and all inclusive.

D. The Unfinished Case

The only possible case that still remains when there is an HRR is the case, illustrated in Fig. 3e, in which S_1 and S_2 both contact the AR, but they contact nonempty and disconnected portions of the AR. We will call this the "disconnected case" but the title of this subsection calls it the "unfinished case" to emphasize the fact that conclusions are not given here for this case. Additional work is needed to derive conclusions for this case, and perhaps this will be done in the future. However, we can include a discussion of the conditions needed for this case to be encountered so that the relevancy (or lack of relevancy) of this case can be recognized.

The case considered is one in which AR_1 and AR_2 are both nonempty and are distinct. Note that the close boundary of AR_1 can be divided into three sections. One section is a portion of the reflective QNR boundary adjacent to S_1. Another section is S_1, where $\mathcal{P}^* + \psi = N\mathcal{U}_1/2V_T$ (under IBC). The remaining section is ARB_1, where, again, $\mathcal{P}^* + \psi = N\mathcal{U}_1/2V_T$. Therefore, on all non-reflective portions of the boundary of AR_1 we have $\mathcal{P}^* + \psi = N\mathcal{U}_1/2V_T$, but in the interior of ARB_1 we have $\mathcal{P}^* + \psi > N\mathcal{U}_1/2V_T$. This implies that $\mathcal{P}^* + \psi$ has a relative maximum in the interior of AR_1. Similarly, $\mathcal{P}^* + \psi$ has a relative maximum in the interior of AR_2. Note that ψ satisfies Laplace's equation, so $\mathcal{P}^* + \psi$ satisfies the same ambipolar diffusion equation (35a) that is satisfied by $\mathcal{P}^*$, and this is satisfied throughout the QNR. Therefore, in order for AR_1 and AR_2 to be nonempty and

distinct, it is necessary (but not sufficient) for a solution to the ambipolar diffusion equation, with the driving term g_{ref}, to have relative maximums at distinct locations within the QNR. This places restrictions on what g_{ref} can be. For example, suppose carrier liberation is produced by a point source. The function $\mathcal{P}^* + \psi$ has a relative maximum only at the location of the source because in any region excluding this point the function $\mathcal{P}^* + \psi$ satisfies Laplace's equation which has no relative maximums. If AR_1 and AR_2 are both nonempty then they each contain a relative maximum, but a relative maximum occurs at only one point for this example, so AR_1 and AR_2 have a point in common, implying that they are connected, which implies that $AR_1 = AR_2$. Therefore, a point source cannot produce the disconnected case. Similarly, if the DRs are in a horizontal arrangement and carrier generation is confined to a vertical line segment, the disconnected case will not occur. However, if the DRs are in a horizontal arrangement and carrier generation is confined to a horizontal line segment, the disconnected case cannot be excluded based on these arguments because a horizontal arrangement of boundary values for $\mathcal{P}^* + \psi$ can result in $\mathcal{P}^* + \psi$ having relative maximums at distinct locations within the QNR. The arguments given here are inconclusive for a horizontal line source of carrier generation, and additional work is needed to reach conclusions for this kind of generation source. Fortunately, the analysis in this section is sufficiently complete to treat the specific examples discussed in Section XII.

XI. MINIMUM VOLTAGE NEEDED TO MAINTAIN A REVERSE-BIASING CONDITION

Previous sections considered the QNR alone. In an actual device, there are two sides to a p-n junction, and the power supply connections used to bias the junction are on the side opposite of the QNR. It was not necessary to analyze the interior of a DR because IBC, intended to represent boundary conditions at the DRB of a reverse-biased DR, supplied boundary conditions for the QNR so that a complete set of equations were obtained when investigating only the QNR interior. However, IBC implies voltages across the QNR when carriers are generated. Therefore, in order to maintain IBC, the power supply voltages connected to the device must be at least large enough to supply these QNR voltages. It is not enough for the power supply to have the correct polarity. To maintain a reverse-biasing condition, the biasing from the power supply must have a large enough magnitude to produce the QNR voltage drops plus a reverse-biasing voltage across the DR. Therefore, the power supply biasing voltage must exceed the voltage drop that occurs across the QNR. Stated another way, the QNR voltage drop implied by IBC is the minimum power supply voltage that can maintain a reverse-biasing condition. This minimum voltage is easily calculated for any example for which the currents have already been solved by using (63).

For a specific example, consider the two-terminal device. This is a K =1 device so there is only one $C_{i,j}$ element, which is $C_{1,1}$, and the inverse $\Phi_{i,j}$ has one element given by $\Phi_{1,1} = 1/C_{1,1}$. Omitting subscripts because there is only one DR, (63) reduces to

$$\mathcal{U} = \left(\mathcal{I}^* - \mathcal{I}_T\right)/C \qquad (K = 1)\,. \tag{101}$$

Recall from Section IX that if $\mathcal{I}^* < q\mathcal{R}$, then there is no HRR and $\mathcal{I}_T = \mathcal{I}^*$, implying that $\mathcal{U}$ = 0. However, $\mathcal{U}$ is the normalized potential in the limit as $\gamma \to \infty$, so a value of zero does not imply that the un-normalized potential with a finite γ is exactly zero. It does imply that the un-normalized potential is much smaller than it would be with the same finite γ but with a different spatial distribution of carrier generation as needed to produce an HRR, so let us consider that case. This occurs when $\mathcal{I}^* > q\mathcal{R}$, which gives $\mathcal{I}_T = q\mathcal{R}$, so (101) becomes

$$\mathcal{U} = \left(\mathcal{I}^* - q\mathcal{R}\right)/C \qquad (K = 1, \quad \mathcal{I}^* > q\mathcal{R})\,. \tag{102}$$

Again, $\mathcal{U}$ is the normalized potential in the limit as $\gamma \to \infty$, so the significance of (102) regarding the un-normalized potential produced by a finite generation rate requires some explanation. The definition of a limit implies that the approximation $\mathcal{U}(\gamma) \approx \mathcal{U}(\infty)$ can be made as accurate as desired by letting γ be sufficiently large. If accuracy is measured in terms of a relative or fractional error (a.k.a., percent error) instead of absolute error, the fact that $\mathcal{U}(\infty) \neq 0$ for the example considered (which was not true for the previous example in which there was no HRR) implies that the approximation $\gamma\mathcal{U}(\gamma) \approx \gamma\mathcal{U}(\infty)$ is accurate when γ is sufficiently large. Replacing the left side with U(γ) and using (102) to substitute for $\mathcal{U}(\infty)$ on the right side, the approximation is written as

$$U(\gamma) \approx \left(\gamma I^* - q\gamma\mathcal{R}\right)/C\ . \tag{103}$$

If we define $I^*(\gamma)$ and $R(\gamma)$ by

$$I^*(\gamma) \equiv \gamma I^* = q\left(1+\frac{D_m}{D_M}\right)\int_{QNR} \Omega(\vec{x})\,\gamma g_{ref}(\vec{x})\,d^3x = q\left(1+\frac{D_m}{D_M}\right)\int_{QNR} \Omega(\vec{x})\,g(\vec{x};\gamma)\,d^3x$$

$$R(\gamma) \equiv \gamma\mathcal{R} = \int_{QNR} \gamma g_{ref}(\vec{x})\,d^3x = \int_{QNR} g(\vec{x};\gamma)\,d^3x$$

then (103) becomes

$$U(\gamma) \approx \left(I^*(\gamma) - qR(\gamma)\right)/C\ .$$

The parameter γ has no further use here if we denote the actual (un-normalized) finite generation rate density as $g(\boldsymbol{x})$ instead of $g(\boldsymbol{x};\ \gamma)$, so the above equations become

$$U \approx \left(I^* - qR\right)/C \qquad (K=1,\quad I^* > qR) \tag{104}$$

where

$$I^* \equiv q\left(1+\frac{D_m}{D_M}\right)\int_{QNR} \Omega(\vec{x})\,g(\vec{x})\,d^3x \tag{105a}$$

$$R \equiv \int_{QNR} g(\vec{x})\,d^3x\ . \tag{105b}$$

The end result is equivalent to a simple substitution. If we replace the normalized generation rate density $g_{ref}(\boldsymbol{x})$ in both (60) and (77) with the actual (un-normalized) finite generation rate density $g(\boldsymbol{x})$, then $\mathcal{U}$ in (102) is replaced by an approximation for the actual (un-normalized) potential U.

To make the example more specific so that a numerical estimate can be obtained, suppose the diode is a cylinder (not necessarily circular), with the DRB at one end, the electrode at the other end, and the cylinder wall is reflective. Regardless of the spatial distribution of carrier generation, the device is one-dimensional from the point of view of Ω defined by (12), and the solution for Ω is $\Omega(x) = 1 - x/L$, where L is the distance between the DRB and electrode, and x is the distance between the DRB and the point of evaluation of Ω. The parameter C calculated from (59) for this Ω is given by

$$C \equiv C_{1,1} \equiv q\,D_m\,\frac{N}{V_T}\int_{DRB} \vec{\nabla}\Omega^{(1)} \circ d\vec{S} = -q\,D_m\,\frac{N}{V_T}\int_{x=0} \frac{d\Omega}{d\,x}\,dS =$$

$$-q\,D_m\,\frac{N}{V_T}\,A\left[\frac{d\Omega}{d\,x}\right]_{x=0} = q\,D_m\,\frac{N}{V_T}\,\frac{A}{L}$$

where A is the cross sectional area of the cylinder. To make the example still more specific, suppose the carrier generation rate is uniform within the cylinder. For a uniform g together with $\Omega(x) = 1 - x/L$, the parameters I^* and R calculated from (105) become

$$I^* = \frac{1}{2} q\, g \left(1 + \frac{D_m}{D_M}\right) A\, L, \qquad R = g\, A\, L\,.$$

A uniform g makes the example simple, but the example is interesting only if this produces an HRR, i.e., if $I^* > qR$. Comparing the above expressions, we conclude that a uniform g will produce an HRR if $D_m/D_M > 1$. This condition is satisfied in a p-type substrate (the minority carriers are electrons which have a larger diffusion coefficient than holes) which is the case considered, so (104) applies. Substituting the above expressions for C, I^*, and R into (104) gives

$$U \approx \frac{1}{2} V_T\, g \frac{L^2}{N}\left(\frac{1}{D_M} - \frac{1}{D_m}\right) = \frac{1}{2} g \frac{L^2}{N}\left(\frac{1}{\mu_h} - \frac{1}{\mu_e}\right).$$

An equivalent equation that expresses U in terms of qR/A instead of g for this example is

$$U \approx \frac{1}{2}\left(\frac{qR}{A}\right)\left(\frac{L}{qN}\right)\left(\frac{1}{\mu_h} - \frac{1}{\mu_e}\right).$$

To obtain a numerical estimate that can be compared to a TCAD simulation result that was previously reported, let the substrate depth L be 4μm and let the substrate doping density N be $8\times10^{14}/\text{cm}^3$. The mobilities for this doping density in silicon, used by the TCAD software that a prediction will be compared to, are $\mu_e = 1310\text{cm}^2/\text{V-s}$ and $\mu_h = 495\text{cm}^2/\text{V-s}$. We also select the carrier generation rate density to satisfy $qR/A = 1000\text{A/cm}^2$. This number was selected because it was concluded in [1] that this will produce the high-injection level condition that is needed for the above approximation for U to be accurate. Using these numbers in the above equation gives $U \approx 1.96$V. This is the minimum voltage that the power supply must have, according to the model, in order to force a reverse-biasing condition across the DR when the device is subject to this carrier generation rate. A 1V (for example) power supply voltage is not enough and will result in the DR becoming forward biased. This conclusion was confirmed by a TCAD simulation result reported in [1]. In fact, it was reported in [1] that the voltage across the QNR, at a power supply voltage of 1V, was 1.62V. This is roughly (but not exactly) the same as the 1.96V that the model predicts would occur if the power supply voltage was increased by the amount needed to produce a reverse-biasing condition. A QNR voltage of 1.62V in the simulated device when the power supply provides only 1V implies that the applied voltage across the DR is –0.62V (stated another way, the total voltage across the DR is the equilibrium voltage, a.k.a., built-in voltage, minus 0.62V, which agrees with TCAD results for the total voltage reported in [1]), implying that the DR was forward biased at 0.62V. However, in spite of this forward bias condition, terminal currents calculated by the model for this two-terminal example agreed with TCAD results reported in [1], even though the derivation of the model assumes a reverse-biasing condition. An explanation

as to why the model should give a correct prediction for such an example is outside the scope of the theoretical analysis given in this report, but an empirical observation from [1] is that a reverse-biasing condition is not always essential for the model to give a correct prediction of terminal current.

XII. A NUMERICAL EXAMPLE COMPARED TO TCAD

This section considers a specific example of the three-terminal (K =2) problem. Model predictions for this problem will be compared to predictions made by TCAD simulations performed by the ATLAS code, which is a device simulator developed by Silvaco [5]. The code solves the drift-diffusion equations governing carrier transport in a semiconductor device. Physical models include bandgap narrowing, Shockley-Read-Hall (SRH) recombination, Auger recombination, and the mobilities depend on doping density and on the electric field. The code accepts a user-specified low-concentration SRH lifetime (which was arbitrarily selected to be 1μs in the examples below) and automatically modifies this to account for doping density. Our version of the code is two-dimensional, and the example silicon device is two-dimensional in rectangular coordinates denoted x and y. An equivalent three-dimensional problem extends the device, as well as the spatial distribution of carrier generation, uniformly in the z direction. Therefore, if carrier generation is confined to a line section in the two-dimensional problem, this will be seen as a section of a plane in the equivalent three-dimensional problem. It is not possible, using our version of the code, to confine carrier generation to a line section in three-dimensions (i.e., a steady-state version of an ion track, which is relevant to studies of SEE) when using rectangular coordinates. However, numerical estimates of terminal currents are not the final products that are regarded as important here. What is considered important here is whether model predictions do or do not agree with TCAD predictions. A two-dimensional problem should be adequate for this comparison.

A. Device Description

TCAD inputs describe the device differently than model inputs, so both descriptions are given here. We begin with the description used for TCAD inputs. The two-dimensional n^+-p silicon device is shown in Fig. 4. The vertical walls at $x = 0$ and $x =$ 24μm are reflective boundaries needed to produce finite dimensions (a net width of 24μm) that can be represented by a finite number of grid points. The depth of the device from top to bottom (not including the aluminum contacts) was arbitrarily taken to be 4μm. Each n^+ region has a Gaussian doping profile with a peak concentration (at the top of the device) of $10^{20}/cm^3$, and the metallurgical (MJ) that separates the n-region from the p-region is at a depth of 0.1μm. The MJ on the left side extends horizontally from $x =$ 4.4μm to $x = 7.6$μm (a uniform in x doping between $x = 4.5$μm to $x = 7.5$μm but a Gaussian roll-off extends the width another 0.1μm on each side). The MJ on the right side extends horizontally from $x = 16.4$μm to $x = 19.6$μm. The doping of the p-region was uniform with a concentration of $10^{15}/cm^3$. Excluding locations near an MJ, the horizontal spacing between grid lines was about 0.25μm, and the vertical spacing was about 0.125μm. The grid was finer near either MJ, with a minimum spacing of 0.02μm. A 3V reverse-biasing voltage was applied to each upper contact. Carrier generation is along a vertical line as seen in the two-dimensional plot (which is a section of a plane in the equivalent three-dimensional geometry), and is vertically uniform. The horizontal location of this line, or "track," is a variable denoted X. Details regarding carrier generation are given later in Section XII-C. Another example considered, the p^+-n device,

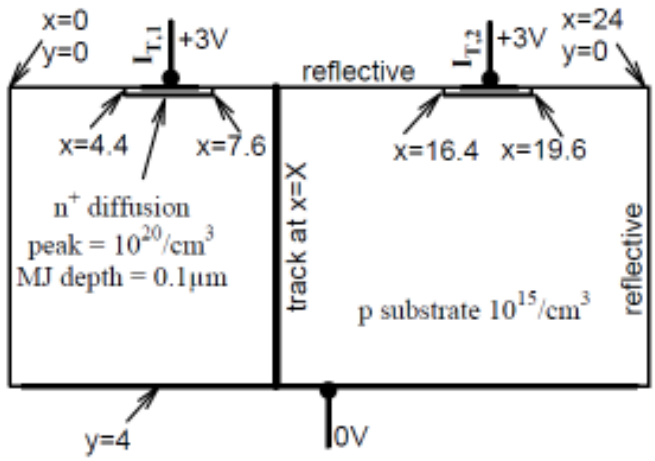

Fig. 4. The simulated two-dimensional n^+-p device. The p^+-n device is the same except that doping types are interchanged and the polarity of the applied voltage is reversed. Carrier generation is on a vertical line, or "track," and is uniform in the vertical (*y*) coordinate. The horizontal (*x*) location of the track is denoted *X* and is a variable. Dimensions denoted *x* or *y* are in μm.

is the same as Fig. 4 except that n-type and p-type are interchanged, and the polarity of the biasing voltage is reversed.

We now describe the device in the context of model inputs. This is a specification of the QNR geometry, which requires a specification of the DRB location. The lateral dimension of each DR produced by the doping profile in Fig. 4 will be about 4μm, but the DR thickness in the vertical direction requires some consideration. In the absence of carrier generation, a 3V biasing voltage with the doping densities shown in Fig. 4 would produce a DR thickness of about 2.25μm. However, intense carrier liberation results in the DR being sufficiently flooded with carriers so that the DR thickness is reduced even if the voltage across the DR is held fixed. The fact that some of the applied voltage appears across the QNR, which reduces the voltage across the DR, further reduces the DR thickness. The end result is that the DR thickness will be considerably less than the nominal value of 2.25μm. Furthermore, the actual DR thickness will depend on the location of carrier generation relative to the DR, and the thickness will not be uniform over the lateral dimensions of the DR. To keep the model simple, the DR thickness used in the model will be uniform over the DR lateral dimensions, and the same DR thickness will be used regardless of the location of carrier generation. The task now is to select a uniform thickness having the property that the calculated currents have adequate accuracy when the same thickness is used in all examples. The model does not calculate this thickness, so the selection was made by comparing model predictions to TCAD simulation results. Two trial values were considered for the DR thickness. One was 1μm and the other was 2μm. For each trial value, model predictions for the terminal currents were compared to TCAD simulation predictions of the terminal currents, and the agreement was found to be better for the first trial value (comparisons using the first trial value are discussed in detail later). Therefore, the DR thickness that will be used by the model is 1μm. The QNR geometry used for model predictions is shown in Fig. 5.

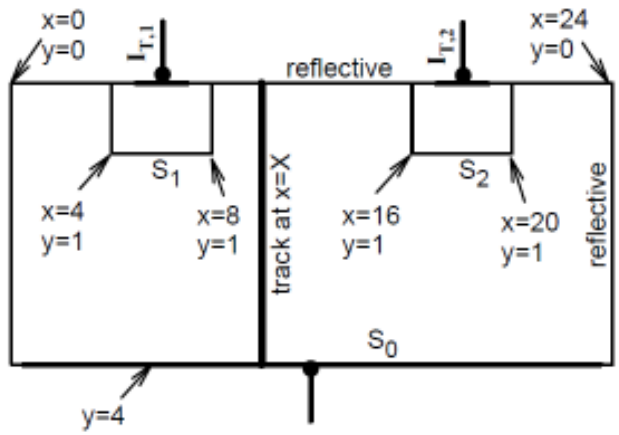

Fig. 5. The model two-dimensional device. S_1 and S_2 are DRBs while S_0 is the substrate contact. A simplified geometry was selected to represent the DRBs and is the same for both the n^+-p device and the p^+-n device. Dimensions denoted x or y are in μm.

B. Model Predictions

The device in Fig. 5 has geometric symmetry. The spatial distribution of carrier liberation need not conform to this symmetry, but symmetry in the device construction implies that

$$\lfloor C_{2,2} + C_{2,1} \rfloor = \lfloor C_{1,1} + C_{1,2} \rfloor > 0 \quad \text{(symmetry)}\,. \tag{106}$$

Recall from Section X-D that the "disconnected case" will not be encountered when carrier liberation in this example device is confined to a vertical line. Therefore, Cases 1A, 1B, and 2 (see Section X) are the only possible cases when there is an HRR. Using this fact in a proof by contradiction, we can invert some of the implications in (93), (95), and (100) to obtain

$$\begin{array}{l} \text{Given that there is an HRR, i.e., given that } I_1{}^* + I_2{}^* > q\mathcal{R}\text{, then} \\ \text{(a) } I_1{}^* \geq I_2{}^* + q\mathcal{R} \quad \text{implies Case 1A} \\ \text{(b) } I_2{}^* \geq I_1{}^* + q\mathcal{R} \quad \text{implies Case 1B} \\ \text{(c) } I_1{}^* < I_2{}^* + q\mathcal{R} \quad \text{together with } I_2{}^* < I_1{}^* + q\mathcal{R} \quad \text{imply Case 2} \end{array} . \tag{107}$$

where (106) was used to simplify expressions appearing in (93), (95), and (100). Combining (107) with other implications in (93), (95), and (100), and again using (106) to simplify expressions, gives

$$\begin{array}{l} \text{If } I_1{}^* > I_2{}^* + q\mathcal{R}\text{, then} \\ \text{(a) Case 1A applies} \\ \text{(b) } I_{T,1} = q\mathcal{R} \quad \text{and} \quad I_{T,2} = 0 \end{array} \tag{108a}$$

$$\begin{aligned}&\text{If } I_2^* > I_1^* + q\mathcal{R}, \text{ then}\\&\text{(a) Case 1B applies}\\&\text{(b) } I_{T,2} = q\mathcal{R} \quad \text{and} \quad I_{T,1} = 0\end{aligned} \tag{108b}$$

$$\begin{aligned}&\text{If } I_1^* + I_2^* > q\mathcal{R}, \quad \text{and} \quad I_2^* - q\mathcal{R} < I_1^* < I_2^* + q\mathcal{R} \text{ then}\\&\text{(a) Case 2 applies} \qquad .\\&\text{(b) } I_{T,1} = \frac{1}{2}\left[I_1^* - I_2^* + q\mathcal{R}\right] \quad \text{and} \quad I_{T,2} = \frac{1}{2}\left[I_2^* - I_1^* + q\mathcal{R}\right]\end{aligned} \tag{108c}$$

To complete the list, we include a reminder from (82a) and (70), which is

$$\begin{aligned}&\text{If } I_1^* + I_2^* < q\mathcal{R}, \text{ then}\\&\text{(a) There is no HRR} \qquad .\\&\text{(b) } I_{T,1} = I_1^* \quad \text{and} \quad I_{T,2} = I_2^*\end{aligned} \tag{108d}$$

In order to use (108) to calculate the currents, it is necessary to calculate I_1* and I_2*. Using the coordinates shown in Fig. 5, and with carrier generation being confined to a vertical line at $x = X$ and having a uniform intensity in the vertical direction, (60) reduces to

$$I_i^*(X) = q\, g_{ref}\left(1 + \frac{D_m}{D_M}\right)\int_{Ytop}^{4\mu m} \Omega^{(i)}(X, y)\, dy \quad for \quad i = 1,2$$

where Y_{top} is the y coordinate of the top of the substrate at $x = X$. If carrier generation is under a DRB, i.e., if either 4μm < X < 8μm or 16μm < X < 20μm, then Y_{top} = 1μm. Otherwise, Y_{top} = 0. Recall that $\Omega^{(1)}(x,y)$ satisfies Laplace's equation in the QNR, has a zero boundary value on S_0 and on S_2 (Fig. 5) and has a unit boundary value on S_1. Hence, $\Omega^{(1)}(x,y)$ is defined only in the QNR. Similarly for $\Omega^{(2)}(x,y)$, but with S_1 and S_2 interchanged. However, the above equation for I_i*(X) includes the contribution to collected charge only from carriers liberated in the QNR. It does not include carriers liberated in the DR. We can include carriers liberated in the DR, and also make I_i*(X) a continuous function of X, from the following considerations. First consider carrier liberation under S_1, i.e., 4μm < X < 8μm. Carriers liberated in the left DR will contribute to $I_{T,1}$ and can be included by extending the domain of $\Omega^{(1)}(x,y)$ to include the left DR by defining it to be 1 in the left DR. Now consider carrier liberation under S_2, i.e., 16μm < X < 20μm. Carriers liberated in the right DR will not contribute to $I_{T,1}$ so we can use Y_{top} = 1μm for this case. This is equivalent to extending the domain of $\Omega^{(1)}(x,y)$ to include the right DR by defining it to be 0 in the right DR. The result is

$$I_1{}^*(X) = \begin{cases} q\,g_{ref}\left(1+\dfrac{D_m}{D_M}\right)\left(1\mu m + \displaystyle\int_{1\mu m}^{4\mu m} \Omega^{(1)}(X,y)\,dy\right) & if \quad 4\mu m < X < 8\mu m \\ q\,g_{ref}\left(1+\dfrac{D_m}{D_M}\right)\displaystyle\int_{1\mu m}^{4\mu m} \Omega^{(1)}(X,y)\,dy & if \quad 16\mu m < X < 20\mu m \\ q\,g_{ref}\left(1+\dfrac{D_m}{D_M}\right)\displaystyle\int_{0}^{4\mu m} \Omega^{(1)}(X,y)\,dy & otherwise \end{cases} .$$

Note that $I_1{}^*(X)$ as given by this equation is continuous in X at $X = 4$, 8, 16, and 20μm. To shorten the notation, note that the total carrier generation rate, that includes carriers liberated in a DR, is given by

$$\mathcal{R} = g_{ref} \times 4\mu m \,. \tag{109}$$

We also define

$$W^{(1)}(X) \equiv \begin{cases} 1\mu m + \displaystyle\int_{1\mu m}^{4\mu m} \Omega^{(1)}(X,y)\,dy & if \quad 4\mu m < X < 8\mu m \\ \displaystyle\int_{1\mu m}^{4\mu m} \Omega^{(1)}(X,y)\,dy & if \quad 16\mu m < X < 20\mu m \\ \displaystyle\int_{0}^{4\mu m} \Omega^{(1)}(X,y)\,dy & otherwise \end{cases} . \tag{110a}$$

A similarly defined $W^{(2)}(X)$ can be calculated from symmetry by

$$W^{(2)}(X) = W^{(1)}(24\mu m - X)\,. \tag{110b}$$

Using this notation, the above equation for $I_1{}^*(X)$ can be written as

$$I_1{}^*(X) = q\mathcal{R}\left(1+\frac{D_m}{D_M}\right)\frac{W^{(1)}(X)}{4\mu m}\,. \tag{111a}$$

Similarly, for $I_2{}^*(X)$ we have

$$I_2{}^*(X) = q\mathcal{R}\left(1+\frac{D_m}{D_M}\right)\frac{W^{(2)}(X)}{4\mu m}\,. \tag{111b}$$

In order to calculate $W^{(1)}(X)$ and $W^{(2)}(X)$ given by (110), it is necessary to solve Laplace's equation subject to the boundary conditions previously stated for $\Omega^{(1)}(x,y)$. Approximate solutions to Laplace's equation can be found in the literature, especially for two-dimensional problems, but the method of solution is not the focus here because any available method that successfully produces the solution for $\Omega^{(1)}(x,y)$ can be used. Here we will use a numerical method. The same TCAD software previously discussed can be used to solve for $\Omega^{(1)}(x,y)$ by entering the geometry shown in Fig. 5. The source of carrier generation is removed and physical models affecting carrier mobilities are de-activated so

that the mobilities are spatially uniform in the code calculations. Also, the substrate is uniformly doped and (in this calculation) there are no p-n junctions. Instead, S_0, S_1, and S_2 are treated as ideal ohmic contacts. This will produce a uniform carrier density, and the potential will satisfy Laplace's equation. The boundary conditions are imposed by assigning a common terminal voltage to S_0 and S_2, and assigning a terminal voltage that is 1V larger to S_1. Finally, a suitably selected constant added to all three terminal voltages will compensate for the contact potentials between electrodes and semiconductor, with the result being a unit potential on S_1 and a zero potential on S_0 and S_2. The potential produced this way is $\Omega^{(1)}(x,y)$. The software has a provision for integrating the potential along a user-specified line, which is a convenient way to evaluate the integrals on the right side of (110a). Using this method to evaluate $W^{(1)}(X)$ and $W^{(2)}(X)$ given by (110) produces the values shown in Table 1.

We now estimate the terminal currents for the n^+-p device shown in Fig. 5. This is done in three steps. The first step calculates $I_1^*(X)$ and $I_2^*(X)$ from (111). The substrate is p-type, so the minority carriers are electrons. This gives $D_m/D_M = D_e/D_h$. Using the Einstein relation, this ratio is the electron mobility divided by the hole mobility. An accurate estimate of this ratio is needed only in the AR, which is the weak-field region, so low-field mobilities are used. We will use the same low-field mobilities that are used by the TCAD software, and these are given by $\mu_e = 1300\text{cm}^2/\text{V-s}$ and $\mu_h = 480\text{cm}^2/\text{V-s}$. This gives $D_m/D_M = 2.7$, so (111) becomes

$$\frac{I_1^*(X)}{q\mathcal{R}} = 0.925 \times \frac{W^{(1)}(X)}{1\mu m}, \qquad \frac{I_2^*(X)}{q\mathcal{R}} = 0.925 \times \frac{W^{(2)}(X)}{1\mu m} \qquad \text{for } n^+\text{-p}. \tag{112a}$$

Using Table 1 for $W^{(1)}(X)$ and $W^{(2)}(X)$, values calculated from (112) for $I_1^*(X)/q\mathcal{R}$ and $I_2^*(X)/q\mathcal{R}$ are shown in the second and third columns of Table 2. The second step uses these column entries with the tests in (108) to determine which case applies. The identified case is shown in the Comments column in Table 2. The last step uses the

Table 1. $W^{(1)}(X)$ and $W^{(2)}(X)$ Calculated from (110)

X (μm)	$W^{(1)}(X)$ (μm)	$W^{(2)}(X)$ (μm)		X (μm)	$W^{(1)}(X)$ (μm)	$W^{(2)}(X)$ (μm)
2	1.251	0.000		12.5	0.371	0.572
3	1.659	0.000		13	0.294	0.704
3.99	2.285	0.000		13.5	0.229	0.863
6	2.443	0.000		14	0.174	1.054
8.01	2.263	0.002		14.5	0.125	1.283
9	1.558	0.083		15	0.083	1.558
9.5	1.283	0.125		15.99	0.002	2.263
10	1.054	0.174		18	0.000	2.443
10.5	0.863	0.229		20.01	0.000	2.285
11	0.704	0.294		21	0.000	1.659
11.5	0.572	0.371		22	0.000	1.251
12	0.463	0.463				

Table 2. Model-predicted Currents for the N^+-P Example

X (μm)	I_1^*/qR	I_2^*/qR	$I_{T,1}/qR$	$I_{T,2}/qR$	Comments
2	1.157	0.000	1	0	Case 1A
3	1.535	0.000	1	0	Case 1A
4	2.114	0.000	1	0	Case 1A
6	2.260	0.000	1	0	Case 1A
8	2.093	0.002	1	0	Case 1A
9	1.441	0.077	1	0	Case 1A
9.5	1.187	0.116	1	0	Case 1A
10	0.975	0.161	0.907	0.093	Case 2
10.5	0.798	0.212	0.793	0.207	Case 2
11	0.651	0.272	0.651	0.272	No HRR
11.5	0.529	0.343	0.529	0.343	No HRR
12	0.428	0.428	0.428	0.428	No HRR
12.5	0.343	0.529	0.343	0.529	No HRR
13	0.272	0.651	0.272	0.651	No HRR
13.5	0.212	0.798	0.207	0.793	Case 2
14	0.161	0.975	0.093	0.907	Case 2
14.5	0.116	1.187	0	1	Case 1B
15	0.077	1.441	0	1	Case 1B
16	0.002	2.093	0	1	Case 1B
18	0.000	2.260	0	1	Case 1B
20	0.000	2.114	0	1	Case 1B
21	0.000	1.535	0	1	Case 1B
22	0.000	1.157	0	1	Case 1B

identified case to determine which calculations listed in (108) are to be used to obtain the terminal currents, and then performs the calculations. The results are listed in the fourth and fifth columns in Table 2. Similar steps are used for the p^+-n version of the Fig. 5 device, except that we flip the ratio D_m/D_M to obtain $D_m/D_M = 1/2.7$ so that (112a) is replaced with

$$\frac{I_1^*(X)}{q\mathcal{R}} = 0.343 \times \frac{W^{(1)}(X)}{1\mu m}, \qquad \frac{I_2^*(X)}{q\mathcal{R}} = 0.343 \times \frac{W^{(2)}(X)}{1\mu m} \qquad \text{for } p^+\text{-n}. \tag{112b}$$

Results for the p^+-n version are shown in Table 3. Note from Table 3 that an HRR is not created in the p^+-n version of this example regardless of the location of carrier generation. In contrast, Table 2 shows that the n^+-p version has an HRR for any carrier-generation location that is not very close to being midway between the DRs.

Table 3. Model-Predicted Currents for the P^+-N Example

X (μm)	I_1^*/qR	I_2^*/qR	$I_{T,1}/qR$	$I_{T,2}/qR$	Comments
2	0.429	0.000	0.429	0.000	No HRR
3	0.569	0.000	0.569	0.000	No HRR
4	0.784	0.000	0.784	0.000	No HRR
6	0.838	0.000	0.838	0.000	No HRR
8	0.776	0.001	0.776	0.001	No HRR
9	0.534	0.028	0.534	0.028	No HRR
9.5	0.440	0.043	0.440	0.043	No HRR
10	0.362	0.060	0.362	0.060	No HRR
10.5	0.296	0.079	0.296	0.079	No HRR
11	0.241	0.101	0.241	0.101	No HRR
11.5	0.196	0.127	0.196	0.127	No HRR
12	0.159	0.159	0.159	0.159	No HRR
12.5	0.127	0.196	0.127	0.196	No HRR
13	0.101	0.241	0.101	0.241	No HRR
13.5	0.079	0.296	0.079	0.296	No HRR
14	0.060	0.362	0.060	0.362	No HRR
14.5	0.043	0.440	0.043	0.440	No HRR
15	0.028	0.534	0.028	0.534	No HRR
16	0.001	0.776	0.001	0.776	No HRR
18	0.000	0.838	0.000	0.838	No HRR
20	0.000	0.784	0.000	0.784	No HRR
21	0.000	0.569	0.000	0.569	No HRR
22	0.000	0.429	0.000	0.429	No HRR

C. Simulation Results and Comparisons

While the model is a mathematical limit (a $\gamma \to \infty$ limit), which requires that certain quantities be normalized in order to remain finite, simulations treat the case of a finite generation rate. The relevant quantities for a simulated device are the actual (un-normalized) terminal currents as calculated by the simulation, and the actual (un-normalized) carrier generation rate R given by (105b). All simulated ionization sources were steady-state and selected to mimic the ideal case of a uniform line (a line as seen in the two-dimensional plot) source, but were actually given a Gaussian horizontal profile with a characteristic width of 0.5μm so that adequate resolution can be obtained without the need for excessively fine grid-line spacing. However, a finite grid-line spacing still produces some numerical error. During a simulation run, the code reports the generation rate as seen by the code with the finite grid spacing. To compensate for numerical error, the value assigned to qR is the generation rate reported by the code. Inputs given to the simulation were selected to make the code-reported value of the generation rate equal to 10^{-4}A/μm. The per-micron units appear because in the equivalent three-dimensional problem the current would be reported as amps per micron in the z dimension (terminal currents calculated by the code also have the units of A/μm). This generation rate was

selected because the model is a high-injection model and the intention is to produce high-injection conditions in the simulated device. It was pointed out in [1] that high-injection conditions are produced when the generation rate divided by the area of the charge-collecting junction is at least 1000A/cm^2. For the equivalent three-dimensional problem that extends 1μm in the z dimension, the generation rate becomes 10^{-4}A. With two DRBs that are each 4μm wide in the x dimension, the charge-collecting area in the equivalent three-dimensional problem is 8μm^2. The generation rate divided by the area slightly exceeds 1000A/cm^2, so we expect these simulation inputs to produce high-injection conditions. Plots of carrier density produced by simulation runs (not shown here) confirm this expectation, showing that the carrier density at locations of maximum density is between one and two orders of magnitude greater than the doping density.

Simulation predictions of terminal current divided by qR are obtained by dividing terminal current calculated by the code by 10^{-4}A/μm. The results are shown in Fig. 6 for the n$^+$-p device and in Fig. 7 for the p$^+$-n device. These plots also include model predictions so that they can be compared to the simulation predictions. The model predictions were obtained by using arguments similar to those given in the discussion surrounding (104) to conclude that the ratio $I_{T,i}/q\mathcal{R}$ of normalized quantities for the limiting case approximates the ratio $I_{T,i}/qR$ of actual (un-normalized) quantities for the finite case. Therefore, the model predictions in Figs. 6 and 7 are identical to the numerical entries in the fourth and fifth columns in Tables 2 and 3.

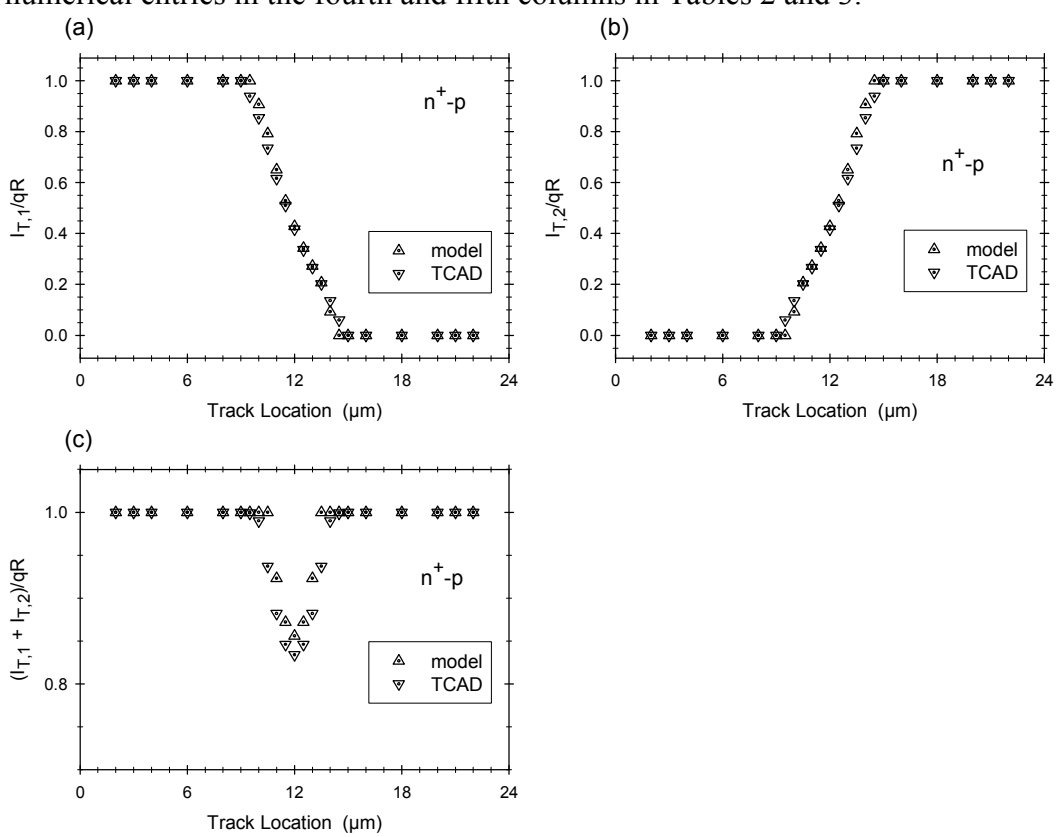

Fig. 6. Simulation predictions of terminal currents for the Fig. 4 device are compared to model predictions for the n$^+$-p version of the Fig. 5 device. Individual currents are compared in (a) and (b). The current sum in (c) is the best indication of whether an HRR is present.

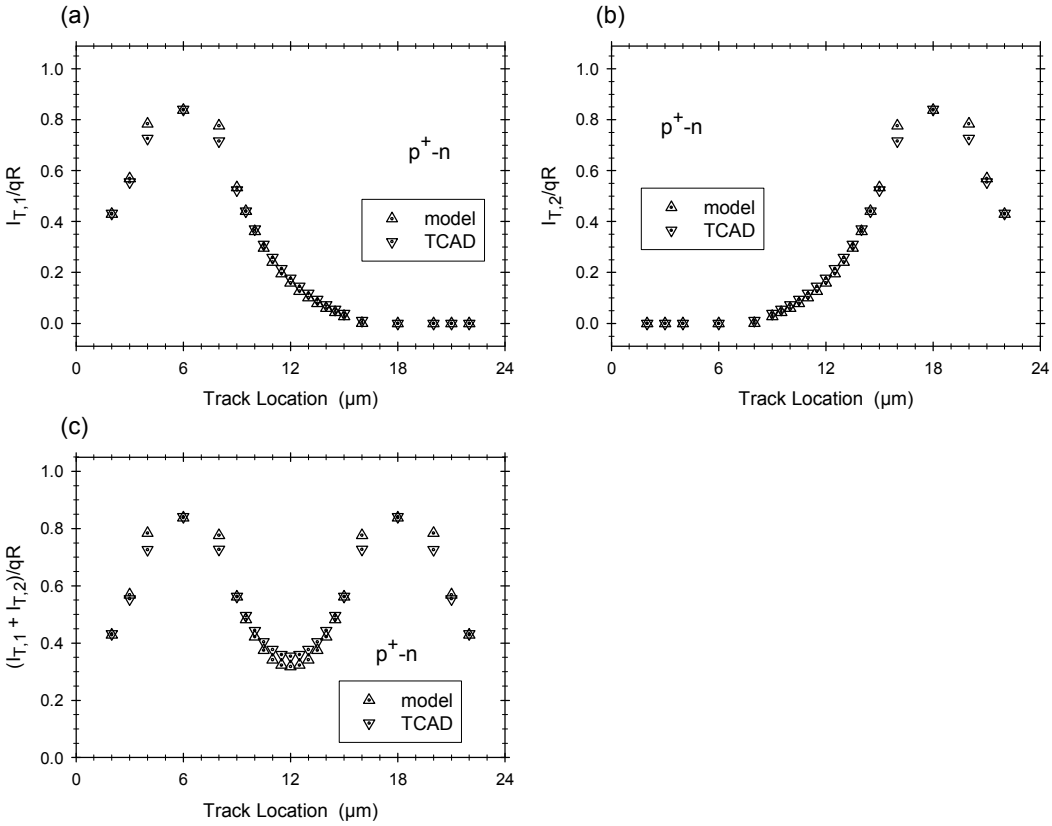

Fig. 7. Like Fig. 6 but for the p^+-n device.

Looking at results for the n^+-p device in Fig. 6, we see excellent agreement between model predictions and simulation results. The two terminal currents are shown individually in Figs. 6a and 6b. However, the sum of the two terminal currents in Fig. 6c is the best indication of whether an HRR is present. These plots combined show distinct charge-collection regimes, but the same conclusions can be seen with greater resolution by looking at the "Comments" column in Table 2. If the track is sufficiently close to one of the junctions (specifically, if $X \leq 9.5\mu m$ which produces Case1A, or $X \geq 14.5\mu m$ which produces Case1B), all of the liberated charge is collected by the closest junction. If the track is moderately close to one of the junctions (specifically, if either $9.5\mu m < X < 11\mu m$ or $13\mu m < X < 14.5\mu m$, which produces Case 2), liberated charge is shared between the two junctions and the two junctions together collect all of the liberated charge. However, if the track is sufficiently far from both junctions (specifically, if $11\mu m \leq X \leq 13\mu m$, which produces no HRR), liberated charge is shared between the two junctions and the two junctions together collect less than all of the liberated charge. In contrast, results for the p^+-n device in Fig. 7 do not show distinct charge-collection regimes. Liberated charge is shared between the two junctions for any track location between the junctions, and the two junctions together collect less than all of the liberated charge for any track location. This is consistent with the "Comments" column in Table 3, showing that an HRR is not present for any track location.

XIII. CONCLUSIONS

What is probably the most notable conclusion from this analysis is the theoretical prediction of distinct charge-collection regimes. This was demonstrated in an example representing a reverse-biased n^+-p silicon device with two junctions arranged horizontally and with carrier liberation along a vertical line (or "track") having a horizontal location that is different in different examples. Depending on where the ionizing track is relative to each of two junctions, either of three possibilities can occur in the example device according to the model. One possibility is that charge is shared by both junctions, and the total collected charge from the two junctions is less than the total amount of liberated charge. This occurs when the track is far enough from both junctions so that there is no strong-field blocking region (HRR) created anywhere in the substrate. The quantitative version of the last statement is that this occurs when $I_1^* + I_2^* < qR$. A second possibility is that collected charge is shared by both junctions, and the total collected charge from the two junctions is equal to the total amount of liberated charge. This occurs when the track is close enough to one of the junctions (but not too close to either junction) so that there is a strong-field blocking region (HRR) in the substrate and surrounding both junctions. The quantitative version of "the track is close enough to one of the junctions" is $I_1^* + I_2^* > qR$, while the quantitative version of "but not to close to either junction" is $I_1^* < I_2^* + qR$ and $I_2^* < I_1^* + qR$. The third possibility is that all liberated charge is collected by one junction, and no charge is collected by the other. This occurs when the track is close enough to one junction so that there is a strong-field blocking region (HRR) in the substrate surrounding just that one junction, with the other junction inside the HRR. The quantitative version of "the track is close enough to one junction" is either $I_1^* > I_2^* + qR$ or $I_2^* > I_1^* + qR$. The model correctly (according to TCAD simulations) predicted which track locations will produce each situation in the example n^+-p device. In contrast, the same example but with n-type and p-type interchanged did not display distinct charge-collection regimes. An HRR was not created by any track location. Collected charge is shared by both junctions, and the total collected charge from the two junctions is less than the total amount of liberated charge for this example. Quantitative agreement between model predictions and TCAD simulations was excellent for all of these examples.

REFERENCES

[1] L. D. Edmonds, "A theoretical analysis of steady-state charge collection in simple diodes under high-injection conditions," *IEEE Trans. Nucl. Sci.*, vol. 57, no. 2, pp. 818-830, April 2010.

[2] L. Edmonds, *A Theoretical Analysis of Steady-State Charge Collection in Simple Diodes under High-Injection Conditions*, Jet Propulsion Laboratory, Pasadena, CA, 2009, JPL Pub. 09-21.

[3] L. D. Edmonds, "A proposed transient version of the ADC charge-collection model tested against TCAD," *IEEE Trans. Nucl. Sci.*, vol. 58, no. 1, pp. 296-304, Feb. 2011.

[4] S. Ramo, J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*, John Wiley and Sons, New York, 1965, p. 315.

[5] http://www.silvaco.com/products/device_simulation/atlas.html

Appendix A: Invertibility of the Coefficient Matrix $C_{i,j}$

To show that $C_{i,j}$ is an invertible matrix, select two arbitrary sets of numbers, with one set denoted $X_1, \ldots, X_K$, and the other set denoted $X_1', \ldots, X_K'$, and then define another pair of number sets by

$$Y_i \equiv \sum_{j=1}^{K} C_{i,j}\ X_j, \qquad Y_i' \equiv \sum_{j=1}^{K} C_{i,j}\ X_j'. \tag{A1}$$

Invertibility of $C_{i,j}$ will be established if it can be shown that the condition that $Y_i' = Y_i$ for each $i = 1, \ldots, K$ implies the condition that $X_i' = X_i$ for each $i = 1, \ldots, K$. To prove this implication, first combine equations in (A1) to get

$$\sum_{i=1}^{K} (Y_i'-Y_i)(X_i'-X_i) \equiv \sum_{i=1}^{K} \sum_{j=1}^{K} (X_i'-X_i)(X_j'-X_j) C_{i,j}. \tag{A2}$$

The next step starts with (59), then (12b), then the divergence theorem, and then (12a) to get

$$\begin{aligned} C_{i,j} \equiv q\, D_m \frac{N}{V_T} \int_{Si} \vec{\nabla}\Omega^{(j)} \circ d\vec{S} &= q\, D_m \frac{N}{V_T} \oint \Omega^{(i)} \vec{\nabla}\Omega^{(j)} \circ d\vec{S} = \\ q\, D_m \frac{N}{V_T} \int_{QNR} &\left[\vec{\nabla}\Omega^{(i)} \circ \vec{\nabla}\Omega^{(j)} + \Omega^{(i)} \nabla^2 \Omega^{(j)}\right] d^3x = \\ q\, D_m \frac{N}{V_T} \int_{QNR} &\vec{\nabla}\Omega^{(i)} \circ \vec{\nabla}\Omega^{(j)} d^3x \quad for \quad i = 1,\ldots,K; \quad j = 1,\ldots,K. \end{aligned} \tag{A3}$$

Substituting (A3) into (A2) gives

$$\begin{aligned} &\sum_{i=1}^{K} (Y_i'-Y_i)(X_i'-X_i) = \\ &\qquad q\, D_m \frac{N}{V_T} \int_{QNR} \vec{\nabla}\left[\sum_{i=1}^{K} (X_i'-X_i)\Omega^{(i)}\right] \circ \vec{\nabla}\left[\sum_{j=1}^{K} (X_j'-X_j)\Omega^{(j)}\right] d^3x. \end{aligned} \tag{A4}$$

Note that (A4) applies to any set of numbers related by (A1) when $C_{i,j}$ is given by (59). As previously stated, invertibility of $C_{i,j}$ will be established if it can be shown that the condition that $Y_i' = Y_i$ for each $i = 1, \ldots, K$ implies the condition that $X_i' = X_i$ for each $i = 1, \ldots, K$, so we assume the former condition and the goal is to prove the latter condition. The assumed condition implies that the left side of (A4) is zero, which implies that the right side is zero. But the integrand on the right is the dot product of a gradient with itself, which is not negative anywhere, so the integrand must be zero everywhere, i.e.,

$$\vec{\nabla}\left[\sum_{i=1}^{K}(X_i'-X_i)\Omega^{(i)}\right]=0$$

which implies

$$\sum_{i=1}^{K}(X_i'-X_i)\Omega^{(i)}(\vec{x})=\text{constant} \quad \text{in} \quad \vec{x}\,.$$

The constant in the above equation is determined by noting that each function $\Omega^{(1)}(\boldsymbol{x})$, …, $\Omega^{(K)}(\boldsymbol{x})$ is zero on S_0, so evaluating the equation on this surface to determine the constant gives

$$\sum_{i=1}^{K}(X_i'-X_i)\Omega^{(i)}(\vec{x})=0\,.$$

Finally, we evaluate this equation on each surface S_1, …, S_K while using (12b) to conclude that $X_i' = X_i$ for each $i = 1, \ldots, K$, which completes the proof that $C_{i,j}$ is invertible.

Appendix B: Inequalities Involving the Coefficients $C_{i,j}$

We first derive inequalities for the $C_{i,j}$ defined by (59). These coefficients were already shown to be symmetric in (61). We start with a property of $\Omega^{(i)}(\boldsymbol{x})$ defined by (12). It is well known that functions satisfying Laplace's equation in a given region take on their minimum and maximum values on the boundary of that region. Using this fact together with (12) we conclude that $0 \le \Omega^{(i)}(\boldsymbol{x}) \le 1$ at any point $\boldsymbol{x}$ that is either in the QNR interior or on the QNR boundary. Therefore, if we first select the observation point $\boldsymbol{x}$ to be on S_i, where $\Omega^{(i)}(\boldsymbol{x}) = 1$, and then move the observation point into the QNR interior, $\Omega^{(i)}(\boldsymbol{x})$ cannot increase. It either remains constant or it decreases. Therefore, $\nabla\Omega^{(i)}(\boldsymbol{x})$ is either zero or is directed outward from the QNR when evaluated at S_i, implying that

$$\int_{Si} \vec{\nabla}\Omega^{(i)} \circ d\vec{S} \ge 0 \quad for \quad i = 1,\ldots,K\,. \tag{B1}$$

Similarly, if we consider a surface S_j with $j \neq i$, so that $\Omega^{(i)}(\boldsymbol{x})$ is zero on this surface, it increases (or remains constant) as the observation point is moved into the QNR interior, and we conclude that

$$\int_{Sj} \vec{\nabla}\Omega^{(i)} \circ d\vec{S} \le 0 \quad for \quad i = 1,\ldots,K \quad and \quad j = 1,\ldots,K \quad with \quad i \neq j\,. \tag{B2a}$$

Now consider the sum function

$$\sum_{j=1}^{K} \Omega^{(j)}(\vec{x})\,.$$

This function is the solution to Laplace's equation that is zero on the electrode S_0, and 1 on each DRB $S_1, \ldots, S_K$. Therefore, if the observation point starts on any DRB, this function decreases (or remains constant) as the observation point is moved into the QNR interior, and we conclude that

$$\int_{Si} \vec{\nabla}\left[\sum_{j=1}^{K} \Omega^{(j)}\right] \circ d\vec{S} \ge 0 \quad for \quad i = 1,\ldots,K$$

or

$$\sum_{j=1}^{K} \int_{Si} \vec{\nabla}\Omega^{(j)} \circ d\vec{S} \ge 0 \quad for \quad i = 1,\ldots,K\,. \tag{B2b}$$

The above derivation of (B1) had the advantage that the same derivation could be used to obtain (B2a) and (B2b), but an alternate derivation of (B1) will produce a stronger statement that the strict inequality applies. This derivation starts with the equality between the second term and the far right term in (A3) and evaluates the result at $i = j$ to get

$$\int_{Si} \vec{\nabla}\Omega^{(i)} \circ d\vec{S} = \int_{QNR} \vec{\nabla}\Omega^{(i)} \circ \vec{\nabla}\Omega^{(i)} d^3x > 0 \quad for \quad i = 1,\ldots,K \tag{B2c}$$

with the right side easily seen to be strictly greater than zero. Similar steps will derive a stronger conclusion (a strict inequality) then can be obtained by summing (B2b) in *i*. The boundary-value property of the sum function implies that

$$\sum_{i=1}^{K} \int_{Si} \vec{\nabla}\left[\sum_{j=1}^{K} \Omega^{(j)}\right] \circ d\vec{S} = \sum_{i=0}^{K} \int_{Si} \left[\sum_{j=1}^{K} \Omega^{(j)}\right] \vec{\nabla}\left[\sum_{j=1}^{K} \Omega^{(j)}\right] \circ d\vec{S}$$

and applying the divergence theorem to the right side while using the fact that the sum function satisfies Laplace's equation gives

$$\sum_{i=1}^{K} \int_{Si} \vec{\nabla}\left[\sum_{j=1}^{K} \Omega^{(j)}\right] \circ d\vec{S} = \int_{QNR} \vec{\nabla}\left[\sum_{j=1}^{K} \Omega^{(j)}\right] \circ \vec{\nabla}\left[\sum_{j=1}^{K} \Omega^{(j)}\right] d^3x \,.$$

The right side is seen to be strictly positive so

$$\sum_{i=1}^{K} \sum_{j=1}^{K} \int_{Si} \vec{\nabla}\Omega^{(j)} \circ d\vec{S} > 0 \,. \tag{B2d}$$

Therefore, while (B2b) states that each *j* sum is nonnegative, (B2d) implies that at least one of these sums is positive.

Another inequality is obtained from (A4) by setting the primed quantities equal to zero so that the equation reduces to

$$\sum_{i=1}^{K} Y_i\, X_i = q\, D_m \frac{N}{V_T} \int_{QNR} \vec{\nabla}\left[\sum_{i=1}^{K} X_i\, \Omega^{(i)}\right] \circ \vec{\nabla}\left[\sum_{i=1}^{K} X_i\, \Omega^{(i)}\right] d^3x \,. \tag{B3}$$

This applies to any set of numbers $X_1, \ldots, X_K$ and $Y_1, \ldots, Y_K$ that are related by the first equation in (A1). Note that if at least one of the numbers $X_1, \ldots, X_K$ differs from zero, each square bracket on the right will be a non-constant function having a gradient that differs from zero somewhere within the QNR, so the dot product of the gradient with itself, which is nonnegative everywhere, will be greater than zero somewhere. Therefore, the integral will be positive. With the understanding that the *Y*'s are related to the *X*'s by the first equation in (A1), the conclusion is

$$\sum_{i=1}^{K} Y_i\, X_i > 0 \quad \text{if at least one of the numbers } X_1,\ldots,X_K \text{ differs from zero} \,. \tag{B4}$$

Using (59) to express (B2) in terms of $C_{i,j}$, and using (A1) to express (B4) in terms of $C_{i,j}$, and repeating (61a) so that all properties are included in a single list, a summary of the properties of $C_{i,j}$ is given by

$$C_{i,j} = C_{j,i} \quad for \quad i = 1,...,K \quad and \quad j = 1,...K \tag{B5a}$$

$$C_{i,j} \le 0 \quad for \quad i = 1,...,K \quad and \quad j = 1,...,K \quad with \quad i \ne j \tag{B5b}$$

$$\sum_{j=1}^{K} C_{i,j} \ge 0 \quad for\ every\ i = 1,...,K,\ and \quad \sum_{j=1}^{K} C_{i,j} > 0 \quad for\ at\ least\ one\ i = 1,...,K\ . \tag{B5c}$$

$$C_{i,i} > 0 \quad for \quad i = 1,...,K \tag{B5d}$$

$$\sum_{i=1}^{K} \sum_{j=1}^{K} C_{i,j}\, X_i\, X_j > 0 \quad \text{if at least one of the numbers } X_1,...,X_K \text{ differs from zero .} \tag{B5e}$$

Note that the simple fact that $C_{i,j}$ is symmetric implies a reciprocity theorem. For any set of primed and unprimed quantities related by (A1) we have

$$\sum_{i=1}^{K} Y_i'\, X_i = \sum_{i=1}^{K} \left[\sum_{j=1}^{K} C_{i,j}\, X_j' \right] X_i = \sum_{j=1}^{K} \left[\sum_{i=1}^{K} C_{i,j}\, X_i \right] X_j' = \sum_{j=1}^{K} \left[\sum_{i=1}^{K} C_{j,i}\, X_i \right] X_j' = \sum_{j=1}^{K} Y_j\, X_j'$$

and changing dummy symbols on the right gives

$$\sum_{i=1}^{K} Y_i'\, X_i = \sum_{i=1}^{K} Y_i\, X_i'\ . \tag{B5f}$$

Note that (B5d) states that the diagonal elements of $C_{i,j}$ are positive, while (B5b) states that the off-diagonal elements are negative (or zero), but (B5c) implies that each diagonal element is at least as large as the sum of absolute values of the off-diagonal elements that are in the same row (or column because of symmetry) as the selected diagonal element.

Appendix C: Inequalities Involving the Coefficients $\Phi_{i,j}$

Inequalities for the coefficients $\Phi_{i,j}$ can be derived from properties of the $C_{i,j}$ coefficients listed in (B5). Recall that these are inverse matrices (invertibility of $C_{i,j}$ was already established in Appendix A), so if any set of numbers $X_1, \ldots, X_K$ and $Y_1, \ldots, Y_K$ are related by

$$Y_i = \sum_{j=1}^{K} C_{i,j} \, X_j \tag{C1}$$

then these numbers are also related by

$$X_i = \sum_{j=1}^{K} \Phi_{i,j} \, Y_j \, . \tag{C2}$$

Two conclusions are fairly obvious. The first is derived from the fact that the inverse of a transpose matrix is the transpose of the inverse, combined with the fact that $C_{i,j}$ is symmetric, to conclude that $\Phi_{i,j}$ is symmetric, as already pointed out in Section V. The second obvious conclusion is obtained by using (C1) to substitute for the sum in j in (B5e), then use (C2) to substitute for the remaining X_i in (B5e), and finally use the fact that at least one of the numbers $X_1, \ldots, X_K$ differs from zero if and only if at least one of the numbers $Y_1, \ldots, Y_K$ differs from zero (because $C_{i,j}$ is invertible) to conclude that

$$\sum_{i=1}^{K} \sum_{j=1}^{K} \Phi_{i,j} \, Y_i \, Y_j > 0$$

for any set of numbers $Y_1, \ldots, Y_K$ that are not all zero. A special choice for these numbers shows that the diagonal elements are positive, i.e.,

$$\Phi_{i,i} > 0 \quad for \quad i = 1, \ldots, K \, .$$

The remaining properties of $\Phi_{i,j}$ that are listed below are more difficult to derive. The analysis will be easier to follow if we use the same terminology that is used in a familiar physical problem so that visualization becomes easier. For this purpose, we will temporarily forget about the problem of charge collection in a semiconductor and think in terms of a simpler electrostatics problem that encounters the same equations as those discussed here. In the electrostatics problem, the surfaces $S_1, \ldots, S_K$ are conductors at potentials $X_1, \ldots, X_K$, while $Y_1, \ldots, Y_K$ are the charges induced on the conductors. The surface S_0 is a grounded conductor that defines the reference potential and has a charge that balances the sum of the charges on the remaining conductors. The remainder of this discussion will use the terminology of this electrostatics problem. Properties of $\Phi_{i,j}$ will be derived in several steps, with the conclusion derived in one step being used by the next step to derive another conclusion.

The first step shows that if at least one conductor has a positive potential, then the conductor at the highest potential has a positive charge. To show this, we consider the case where the maximum of the set { X_1, …, X_K} is positive, and let X_{max} denote this maximum value. It is possible for more than one conductors to be at this maximum potential, so we define the index set IND, a subset of { 1, …, K}, by the property that $i \in$ IND if and only if $X_i = X_{max}$. Note that if $i \notin$ IND then $X_i < X_{max}$. Summing (C1) in i over the index set, and breaking the j sum up into two parts gives

$$\sum_{i \in IND} Y_i = X_{\max} \sum_{i \in IND} \sum_{j \in IND} C_{i,j} + \sum_{i \in IND} \sum_{j \notin IND} C_{i,j} \, X_j \, . \tag{C3}$$

The goal is to show that the left side is positive. For this purpose, we first derive an inequality for the first term on the right. This is obtained by changing notation in (B5e) to get

$$\sum_{i=1}^{K} \sum_{j=1}^{K} C_{i,j} \, Z_i \, Z_j > 0$$

where the strict inequality applies if the Z's are not all zero. In particular, if we let $Z_i = 1$ when $i \in$ IND and $Z_i = 0$ when $i \notin$ IND we obtain

$$\sum_{i \in IND} \sum_{j \in IND} C_{i,j} > 0 \, .$$

Combining this with $X_{max} > 0$ gives

$$X_{\max} \sum_{i \in IND} \sum_{j \in IND} C_{i,j} > 0 \, . \tag{C4}$$

An inequality for the second term on the right side of (C3) is obtained from the fact that each $C_{i,j}$ in that term is an off-diagonal element, so

$$C_{i,j} \le 0 \quad when \quad i \in IND \quad and \quad j \notin IND \, . \tag{C5}$$

Combining this with $X_i < X_{max}$ when $i \notin$ IND gives

$$C_{i,j} \; X_j \ge C_{i,j} \; X_{\max} \quad when \quad i \in IND \quad and \quad j \notin IND \, . \tag{C6}$$

At this point it is necessary to consider two possibilities. The first would occur (for example) if S_0 were composed of multiple sections formed in such a way so the each conductor S_1, …, S_K is completely surrounded by a different section of S_0, so there is no interaction between the conductors S_1, …, S_K. More specifically, the first possibility is the case in which the equality in (C5) applies to every $i \in$ IND and every $j \notin$ IND. For this case, the right side of (C3) reduces to the first term, which is positive according to (C4), so the left side of (C3) is positive for this case. The only other possibility is that there is at

least one i, j pair, with $i \in$ IND and $j \notin$ IND, for which the strict inequality in (C5) applies. This implies the strict inequality in (C6) for at least one pair of indices, producing a strict inequality between the sums given by

$$\sum_{i \in IND} \sum_{j \notin IND} C_{i,j} \; X_j > X_{\max} \sum_{i \in IND} \sum_{j \notin IND} C_{i,j}$$

and substituting this into (C3) gives

$$\sum_{i \in IND} Y_i > X_{\max} \sum_{i \in IND} \sum_{j=1}^{K} C_{i,j} \, .$$

The right side is positive or zero, because of (B5c) together with $X_{max} > 0$, so for this second possibility we also reach the conclusion that the left side of (C3) is positive. The final conclusion is

$$\left.\begin{array}{l} \text{If the maximum of the set } \{X_1,...,X_K\} \text{ is positive,} \\ \text{and if IND is the index set that identifies those} \\ \text{conductors that are at the maximum potential, then} \\ \sum_{i \in IND} Y_i > 0. \end{array}\right\} \qquad \text{(C7a)}$$

A similar derivation will show that

$$\left.\begin{array}{l} \text{If the minimum of the set } \{X_1,...,X_K\} \text{ is negative,} \\ \text{and if IND is the index set that identifies those} \\ \text{conductors that are at the minimum potential, then} \\ \sum_{i \in IND} Y_i < 0. \end{array}\right\} \qquad \text{(C7b)}$$

The next step uses (C7) to conclude that uncharged conductors are at intermediate potentials. To be more specific, select one of the conductors and call it S_m for some m =1, …, K. Place a positive charge Y_m on S_m. All remaining conductors are uncharged, i.e., Y_i = 0 for each i = 1,…, K when $i \neq m$. The potential of S_m is X_m, and this potential is seen to be positive via (C2) combined with $Y_m > 0$ and $\Phi_{m,m} > 0$. The conclusion to be proven is that each uncharged conductor is at a potential that is somewhere between 0 and the potential of the charged conductor. In other words, the conclusion to be proven is that $0 \leq X_i \leq X_m$ for each i = 1,…, K. The proof for each inequality is by contradiction. We first prove that $X_i \leq X_m$ for each i = 1,…, K by assuming that there is at least one conductor satisfying $X_m < X_i$ and then look for a contradiction. This assumption implies that the conductor having the maximum potential is one of the uncharged conductors, or several of the uncharged conductors if they are at the same potential. Let IND be the index set that identifies all conductors at the maximum potential. The assumption implies that these are uncharged conductors, implying that

$$\sum_{i \in IND} Y_i = 0 .$$

The promised contradiction is obtained from (C7a). The contradiction implies that $X_i \leq X_m$ for each $i = 1,\ldots, K$. Similar steps that use (C7b) to obtain a contradiction conclude that $0 \leq X_i$ for each $i = 1,\ldots, K$. The conclusion is

$$\left.\begin{array}{l} \text{If } Y_m > 0 \text{ for some } m = 1,\ldots,K, \text{ and if} \\ Y_i = 0 \text{ for each } i = 1,\ldots,K \text{ when } i \neq m, \\ \text{then } 0 \leq X_i \leq X_m \text{ for each } i = 1,\ldots,K. \end{array}\right\} \tag{C8}$$

The last step uses (C8) to derive inequalities for $\Phi_{i,j}$. This is done by letting $Y_m = 1$ and letting $Y_i = 0$ for each $i = 1,\ldots, K$ when $i \neq m$. Substituting this into (C2) gives $X_i = \Phi_{i,m}$ for this choice of Y's, and combining this with (C8) gives $0 \leq \Phi_{i,m} \leq \Phi_{m,m}$.

In summary, properties of $\Phi_{i,j}$ are

$$\Phi_{i,j} = \Phi_{j,i} \quad \textit{for} \quad i = 1,\ldots,K \quad \textit{and} \quad j = 1,\ldots K \tag{C9a}$$

$$0 \leq \Phi_{i,j} \leq \Phi_{j,j} \quad \textit{for} \quad i = 1,\ldots,K \quad \textit{and} \quad j = 1,\ldots,K \tag{C9b}$$

$$\Phi_{i,i} > 0 \quad \textit{for} \quad i = 1,\ldots,K \tag{C9c}$$

$$\sum_{i=1}^{K} \sum_{j=1}^{K} \Phi_{i,j} Y_i Y_j > 0 \quad \text{if at least one of the numbers } Y_1,\ldots,Y_K \text{ differs from zero .} \tag{C9d}$$

Appendix D: Properties of Ohm's Law

We consider the case in which there is an HRR. If we define the total current density $\mathcal{J}_T$ to be $\mathcal{J}_m + \mathcal{J}_M$, then (50) gives

$$\vec{\mathcal{J}}_T(\vec{x}) = -2q\frac{D_m + D_M}{N}\left[P(\vec{x}) + \frac{D_M}{D_m + D_M}N\right]\vec{\nabla}\left[\mathcal{P}^*(\vec{x}) + \psi(\vec{x})\right] \qquad \text{(in HRR)}. \tag{D1}$$

The fact that the total current has a zero divergence was already stated in (51). To shorten the notation, define

$$\sigma(\vec{x}) \equiv q\frac{D_m + D_M}{V_T}\left[P(\vec{x}) + \frac{D_M}{D_m + D_M}N\right] \qquad \text{(in HRR)} \tag{D2a}$$

$$\mu(\vec{x}) \equiv \frac{2V_T}{N}\left[\mathcal{P}^*(\vec{x}) + \psi(\vec{x})\right] \qquad \text{(in QNR)} \tag{D2b}$$

so (D1) and (51) become

$$\vec{\mathcal{J}}_T(\vec{x}) = -\sigma(\vec{x})\vec{\nabla}\mu(\vec{x}) \qquad \text{(in HRR)} \tag{D3a}$$

$$\vec{\nabla}\circ\left[\sigma(\vec{x})\vec{\nabla}\mu(\vec{x})\right] = 0 \qquad \text{(in HRR)}. \tag{D3b}$$

This is Ohm's Law with a conductivity σ and a potential μ. Within the HRR, the potential μ is the same as the normalized potential $\mathcal{U}$ given by (49). Different symbols are used because (D2b) defines μ throughout the entire QNR (note that $\mathcal{P}^*$ and ψ are defined throughout the entire QNR), including the AR where (49) does not apply and $\mathcal{U}$ is not equal to μ. Note that σ is bounded above zero because $P \geq 0$. In the main text, $\mathcal{P}^*$ was regarded as known, and ψ was regarded as known via (55), so the purpose of (51) was to solve for P. This is equivalent to solving (D3b) for σ when μ is regarded as the known via (D2b). However, the mere existence of a σ that is bounded above zero and satisfies (D3b) implies some properties of μ. Stated another way, the HRR is some subset of that portion of the QNR in which μ has certain properties. This places a restriction on how large the HRR can be. One property, which is a well-known property of potentials satisfying Ohm's Law, is that μ has no relative maximums or minimums in the HRR interior. The defining equation (D2b) for μ together with (29a) and the fact that ψ satisfies Laplace's equation can be used to show that μ has no relative minimums anywhere in the QNR, but it can have a relative maximum in the QNR. Any point in the QNR at which μ has a relative maximum, if there is such a point, is outside the HRR.

Other properties of μ in the HRR interior can be obtained in analogy with the analysis already given in the first three appendices. To establish this analogy, first consider the closed surface that is the boundary of the HRR. Part of this surface is S_0, where the

boundary value of μ is zero. Another part is a portion of the reflective QNR boundary, which would be more appropriately be called "insulated" when discussing Ohm's Law, where μ satisfies reflective boundary conditions. The remainder of the boundary is a set of surfaces that have the role of contacts or terminals and each of these surfaces can be placed in one of two groups. The first group consists of those DRBs that contact the HRR. For any $i = 1, \ldots, K$ such that S_i contacts the HRR, S_i is one of the terminals in the first group. On this surface, the value of μ given by (D2b), with ψ satisfying (55) and $\mathcal{P}^*$ satisfying (35b), is $\mathcal{V}_i$. The second group of terminals consists of portions of the ARB. For any $i = 1, \ldots, K$ such that S_i contacts the AR, ARB_i is one of the terminals in the second group. On this surface, where (52) applies, the value of μ given by (D2b) is $\mathcal{V}_i$. Note that the number of distinct terminals in the second group might be less than the number of DRBs that contact the AR because if S_i is connected to S_j then $ARB_i = ARB_j$ and $\mathcal{V}_j = \mathcal{V}_i$. Let M ($\leq K$) denote the number of distinct terminals in the two groups combined, and let the terminals be denoted $\tilde{S}_1, \ldots, \tilde{S}_M$. Let $\tilde{\mathcal{V}}_i$ denote the value of μ on $\tilde{S}_i$. The set of numbers $\tilde{\mathcal{V}}_1, \ldots, \tilde{\mathcal{V}}_M$ is the same as the set of numbers $\mathcal{V}_1, \ldots, \mathcal{V}_K$, but they can differ in the way the subscripts were assigned because M can be less than K (in which case the latter set of numbers has repeated elements). The terminal currents are defined by

$$\tilde{I}_{T,i} \equiv -\int_{\tilde{S}_i} \vec{\mathcal{J}}_T \circ d\vec{S} \tag{D4}$$

with the unit normal vector in the surface integral directed outward from the HRR. Corresponding to the conductivity σ is a set of functions $\mu^{(1)}, \ldots, \mu^{(M)}$, defined by the boundary-value problems

$$\vec{\nabla} \circ \left[\sigma(\vec{x})\vec{\nabla}\mu^{(i)}(\vec{x})\right] = 0 \qquad \text{in HRR interior } (i = 1,\ldots,M) \tag{D5a}$$

$$\mu^{(i)} = 0 \quad \text{on } S_0 \ (i = 1,\ldots,M) \tag{D5b}$$

$$\mu^{(i)} = 1 \quad \text{on } \tilde{S}_i \ \text{ and } \ \mu^{(i)} = 0 \quad \text{on } \tilde{S}_j \text{ if } j \neq i \ (i = 1,\ldots,M : j = 1,\ldots,M) \tag{D5c}$$

with reflective boundary conditions tacitly assumed on the insulated boundaries. Comparing (D5) to (D3b) and to the boundary conditions previously stated for μ, we conclude that

$$\mu(\vec{x}) = \sum_{i=1}^{M} \tilde{\mathcal{V}}_i \, \mu^{(i)}(\vec{x}) \quad (\text{in HRR}) . \tag{D6}$$

Combining (D3a) with (D4) and (D6) gives

$$\tilde{I}_{T,i} = \sum_{j=1}^{M} \tilde{\mathcal{V}}_j \, \tilde{C}_{i,j} \quad \textit{for} \quad i = 1,\ldots,M \tag{D7}$$

where

$$C^{\sim}{}_{i,j} \equiv \int_{S^{\sim}{}_i} \sigma \vec{\nabla}\mu^{(j)} \circ d\vec{S} \quad for \quad i = 1,...,M; \quad j = 1,...,M \ . \tag{D8}$$

Conclusions previously derived for $C_{i,j}$ can also be applied to $C^{\sim}{}_{i,j}$ after showing that the latter satisfies the inequalities that were used to derive those conclusions. For example, (B5d) in Appendix B was derived from (B2c). The same result can be derived for $C^{\sim}{}_{i,i}$ by using (D8) with (D5) and the divergence theorem to get

$$C^{\sim}{}_{i,i} \equiv \int_{S^{\sim}{}_i} \sigma \vec{\nabla}\mu^{(i)} \circ d\vec{S} = \oint \mu^{(i)} \sigma \vec{\nabla}\mu^{(i)} \circ d\vec{S} = \int_{HRR} \sigma \vec{\nabla}\mu^{(i)} \circ \vec{\nabla}\mu^{(i)} \, d^3x \ .$$

Using the fact that σ is bounded above zero, we conclude from the above that $C^{\sim}{}_{i,i} > 0$. In fact, every statement listed in (B5) remains true when $C_{i,j}$ is replaced by $C^{\sim}{}_{i,j}$ and K is replaced by M. Therefore

$$C^{\sim}{}_{i,j} = C^{\sim}{}_{j,i} \quad for \quad i = 1,...,M \quad and \quad j = 1,...M \tag{D9a}$$

$$C^{\sim}{}_{i,j} \le 0 \quad for \quad i = 1,...,M \quad and \quad j = 1,...,M \quad with \quad i \ne j \tag{D9b}$$

$$\sum_{j=1}^{K} C^{\sim}{}_{i,j} \ge 0 \quad for\ every\ i = 1,...,M,\ and \quad \sum_{j=1}^{K} C^{\sim}{}_{i,j} > 0 \quad for\ at\ least\ one\ i = 1,...,M \ . \tag{D9c}$$

$$C^{\sim}{}_{i,i} > 0 \quad for \quad i = 1,...,M \tag{D9d}$$

$$\sum_{i=1}^{K} \sum_{j=1}^{K} C^{\sim}{}_{i,j} X_i X_j > 0 \quad \text{if at least one of the numbers } X_1,...,X_K \text{ differs from zero} \ . \tag{D9e}$$

The positive definite property (D9e) implies that $C^{\sim}{}_{i,j}$ has an inverse. This inverse will be denoted $\Phi^{\sim}{}_{i,j}$. This inverse allows us to invert (D7) to get

$$\mathcal{U}^{\sim}{}_i = \sum_{j=1}^{M} I^{\sim}{}_{T,j} \, \Phi^{\sim}{}_{i,j} \quad for \quad i = 1,...,M \ . \tag{D10}$$

Note that all properties listed in Appendix C for $\Phi_{i,j}$ were derived from the properties of $C_{i,j}$ listed in (B5). Because $C^{\sim}{}_{i,j}$ has the same properties listed in (B5) for $C_{i,j}$, we conclude that $\Phi^{\sim}{}_{i,j}$ has the same properties listed in Appendix C for $\Phi_{i,j}$. In particular,

$$\Phi^{\sim}{}_{i,j} = \Phi^{\sim}{}_{j,i} \quad for \quad i = 1,...,M \quad and \quad j = 1,...,M \tag{D11a}$$

$$0 \le \Phi^{\sim}{}_{i,j} \le \Phi^{\sim}{}_{j,j} \quad for \quad i = 1,...,M \quad and \quad j = 1,...,M \tag{D11b}$$

$$\Phi^{\sim}{}_{i,i} > 0 \quad for \quad i = 1,...,M \tag{D11c}$$

$$\sum_{i=1}^{M} \sum_{j=1}^{M} \Phi^{\sim}{}_{i,j}\, Y_i\, Y_j > 0 \quad \text{if at least one of the numbers } Y_1,\ldots,Y_K \text{ differs from zero .} \qquad \text{(D11d)}$$

Also, an inequality stated in Appendix C in the context of the electrostatics problem was the statement that uncharged conductors are at intermediate potentials. The same mathematical inequality applies to the Ohm's Law problem, but a more appropriate physical statement of this inequality is that floating terminals are at intermediate potentials.