

JPL Publication 09-21



A Theoretical Analysis of Steady-State Charge Collection in Simple Diodes under High-Injection Conditions

Larry D. Edmonds

**National Aeronautics and
Space Administration**

**Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California**

September 10, 2009

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.

© 2009 California Institute of Technology. Government sponsorship acknowledged.

TABLE OF CONTENTS

ABSTRACT	1
I. INTRODUCTION.....	2
II. QUALITATIVE DISCUSSION OF THE AR AND HRR.....	5
III. HEURISTIC DERIVATION OF THE ADC MODEL IN 1D.....	8
A. Governing Equations	8
B. Derivation of the Charge-Collection Efficiency	9
C. Showing Consistency with the Symbolic Model.....	13
D. Derivation of the HRR Width.....	14
E. Voltage Needed to Reverse-Bias the DR.....	16
IV. HEURISTIC DERIVATION OF THE ADC MODEL IN 3D	18
A. Governing Equations	18
B. Derivation of the Charge-Collection Efficiency	20
C. An Example Geometry: The Isolated Disc	21
D. Voltage Needed to Reverse-Bias the DR	23
E. Terminology Clarification	26
V. COMPARISONS WITH COMPUTER SIMULATION RESULTS	27
A. First Example: A 1D n^+ -p Diode with Uniform g	27
B. Second Example: A 1D n^+ -p Diode with Localized Source near the QNR Center	29
C. Third Example: A 1D n^+ -p Diode with Localized Source near the Contact.....	29
D. Fourth Example: A 1D p^+ -n Diode with Uniform g	29
E. Fifth Example: A 3D n^+ -p Diode with a Custom g	30
VI. REFERENCES.....	32
APPENDIX A: A MORE RIGOROUS DERIVATION OF THE ADC MODEL	33
A1. The First Case.....	36
A2. The Second Case	39
A3. Ideal Boundary Conditions.....	40
APPENDIX B: ABBREVIATIONS, ACRONYMS, AND NOMENCLATURE.....	43

Figures

Figure 1. Diode with a localized carrier-generation site.	2
Figure 2. Charge-collection efficiency versus source location (a), and plots of carrier density for each of several source locations (b)–(d).	6
Figure 3. An example 3D geometry shows the notation. The surface S_1 is the DRB, and the surface S_2 is the electrode contact to the QNR. All other boundary surfaces surrounding the QNR are reflective. The spatial distribution of carrier generation within the QNR is arbitrary.	18
Figure 4. A more specific 3D geometry in which the DRB resembles an isolated disc.	22

A Theoretical Analysis of Steady-State Charge Collection in Simple Diodes under High-Injection Conditions

Larry D. Edmonds
Jet Propulsion Laboratory
California Institute of Technology
Mail Stop 303-220
4800 Oak Grove Drive
Pasadena, California 91109-8099
Phone: (818) 354-2778
FAX: (818) 393-4559
Email: larry.d.edmonds@jpl.nasa.gov

ABSTRACT

A previous rigorous mathematical analysis of drift-diffusion equations was used to investigate collected charge in a simple reverse-biased p-n junction diode exposed to an ionization source that liberates carriers (electron-hole pairs) in a quasi-neutral region within the diode. Each of two simple models was found to agree with the more rigorous analysis when carrier liberation is sufficiently intense. One is the sensitive volume (SV) model, and the other was called “ambipolar diffusion with a cutoff” (ADC). The earlier rigorous analysis was worked out in detail only for a localized source, i.e., a point source of carrier liberation, so it was able to validate the applicability of each simple model only for that case. The present paper treats an arbitrary spatial distribution of carrier generation and concludes that the ADC model remains valid for this more general case, but the SV model does not.

Key words: ADC model, ambipolar diffusion, ambipolar diffusion with a cutoff, charge collection, charge-collection efficiency, drift-diffusion, sensitive volume, SV model.

I. INTRODUCTION

A previous analysis [1] investigated collected charge in a simple reverse-biased p-n junction diode exposed to an ionization source that liberates electron-hole pairs in a quasi-neutral region (QNR) within the diode, i.e., carrier liberation was at a location outside the junction depletion region (DR). The analysis was worked out in detail only for a localized source as shown in Fig. 1. The region labeled “carrier source” in the figure is a thin layer in which electron-hole pairs are liberated. The confinement of carrier liberation to within a thin layer is a hypothetical case but it is an academically interesting case because an academically interesting question is how the collected charge varies as the source is moved to different locations within the QNR. The answer given in [1] to this question included plots of charge-collection efficiency (defined to be the charge-collection rate divided by the rate that carriers are generated by the source) versus the source location X_S measured from the DR boundary (DRB)¹. Different plots were given for different input parameters (e.g., carrier-generation rate, doping density, carrier mobility, etc.) and the combined set of plots was sufficiently complete to represent virtually any set of input parameters via interpolations. It was concluded in [1] that, when the source strength (i.e., carrier-generation rate) is sufficiently large, the collected charge is correctly predicted by each of two simple models. One is the sensitive volume (SV) model, which states that there is some portion of the QNR, the SV,² having the property that all charge liberated within this region is collected, while charge liberated outside is collected with an efficiency that is consistent with pure diffusion from the source to the SV boundary.

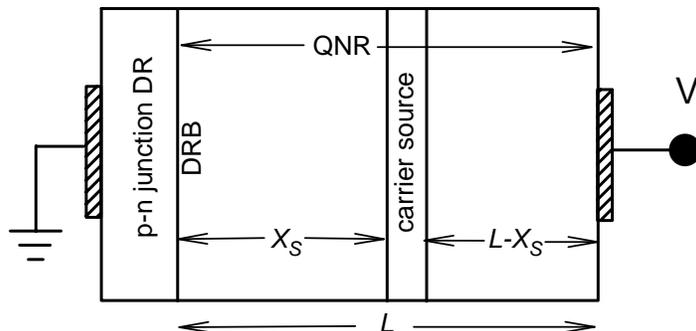


Figure 1. Diode with a localized carrier-generation site.

The other model was called “ambipolar diffusion with a cutoff” (designated ADC in this paper), which will be explained in the next paragraph. It was pointed out in [1] that the SV model is symbolic in the sense of producing correct predictions of collected charge (from a localized source) without explaining the relevant physics. Also, the SV model gives correct predictions only for a single localized source. If there are multiple sources at different locations, the presence

¹ The DRB can be recognized by plotting electron and hole densities, together on the same graph, against a spatial coordinate. This will show a reasonably well-defined boundary (the DRB) separating a space-charge region (the DR) from a quasi-neutral region (the QNR). Example plots with the DRBs indicated can be found in [2].

² The DR is also an SV, and the portion of the QNR that becomes an SV when the source strength is sufficiently large is adjacent to the DR, so we could think of the QNR contribution as an extension of the SV.

of one source influences the amount of charge collected from another,³ which is a nonlinear effect and is not consistent with the SV model. It was also pointed out in [1] that the ADC model gives exactly the same predictions for single localized sources as the SV model, but the ADC model has a closer connection with charge-collection physics. This closer connection with actual physics suggests that the ADC model might be more versatile in the sense of being applicable to an arbitrary spatial distribution of carrier generation. However, this assertion was not verified in [1] because a single localized source was the only spatial distribution analyzed in detail. The objective of the present paper is to show that the above assertion is correct.

A review of the ADC model rigorously derived in [1] is as follows. This model becomes exact⁴ in the limit of a large generation rate. It is only an approximation for a finite generation rate, but the approximation is accurate if the generation rate is sufficiently large. This model states that the collected charge can be calculated using three steps. The first step calculates what the collected charge would be if charge collection were purely by diffusion. The second step multiplies this estimate by a factor that is intended to include the contribution from drift. The diffusion coefficients used in these calculations motivated the name “ambipolar diffusion current” for the estimate calculated at the end of the second step. The factor used in the second step is a simple function of diffusion coefficients and is intended to convert a calculated number into the actual current, but this simple factor does not always accomplish its intended purpose, so a third step is needed to correct an error when there is an error. The third step applies a cutoff when needed. The charge-collection efficiency corresponding to the current calculated at the end of the second step can sometimes exceed 100 percent. If it does not exceed 100 percent, this calculated value is the final estimate. If it does exceed 100 percent, a cutoff changes the value to 100 percent, and this is the final estimate of the charge-collection efficiency. The ADC model can be used to derive the SV model when the source is localized. If the source is localized, the cutoff used in the ADC model applies when the source is sufficiently close to the DRB. In other words, the charge-collection efficiency is 100 percent when the source is within a certain distance from the DRB. The SV is defined to be the region contained within this distance from the DRB. The ADC model was derived in [1] for the special case of a localized source. As noted above, the objective of this paper is to establish the validity of the ADC model for an arbitrary spatial distribution of carrier generation by generalizing the derivation to include this more general case.

The analysis given here is theoretical because it uses a number of simplifying physical approximations (e.g., carrier transport is described by the drift-diffusion equations, electron and hole mobilities are constants and related to diffusion coefficients by the Einstein relation, carrier recombination in the device interior can be neglected, and ideal boundary conditions are assumed). The analysis applies to a hypothetical device that is defined by these assumed governing equations. Conclusions are correct for this hypothetical device if the mathematical analysis is valid, so this theoretical investigation is a mathematical analysis applied to assumed

³ An example starts with a localized source outside the SV so that the charge-collection efficiency is less than 100 percent. The addition of a second source of sufficient strength and sufficiently close to the DRB will make the charge-collection efficiency for both sources equal to 100 percent.

⁴ In this context, “exact” refers to the mathematical treatment of the drift-diffusion equations. These equations are themselves only approximations for real devices (e.g., electron and hole mobilities were regarded as constants in these equations), so accuracy claims refer to a hypothetical device that is defined by the assumed governing equations.

equations. Real devices are more complex (e.g., mobilities are functions of a variety of physical parameters), and there is a question as to whether this hypothetical device will be a good representation of a real device. Comparisons between model predictions and computer simulation results in Section V show that the model is accurate for the examples that were investigated, but no finite number of examples will be exhaustive, and there will never be guarantees that this hypothetical device will be a good representation of some selected real device. The contention here is that there is little hope of understanding charge collection in a real device if we have not yet understood charge collection in this simpler hypothetical device. Thus, an understanding of the simpler problem is a prerequisite to understanding charge collection in a real device.

Another limitation is that the analysis is steady-state, i.e., carrier generation is at a quasi-constant rate (e.g., by photon irradiation). An important application of a charge-collection model is in the investigation of single-event effects in which carriers are liberated by a single particle (e.g., a heavy ion) traveling through the device. This application makes the high-injection limit more than just a mathematical curiosity because the density of liberated carriers can be orders of magnitude larger than the doping density. However, this is a highly transient problem, and the relevancy of a steady-state analysis is questionable. The contention here is that the transient problem is no easier to understand than the simpler steady-state problem, and an understanding of the steady-state problem is a prerequisite to understanding the more difficult transient problem.

The analysis in [1] was specialized in the sense that carrier generation is confined to within a thin layer, but it was general in the sense that the source strength was arbitrary. However, the most interesting conclusions (e.g., that an SV is created) apply when the source has sufficient strength. Here we consider an arbitrary spatial distribution of carrier generation. The analysis is tractable only in the limiting case of a large carrier-generation rate,⁵ so conclusions are derived only for that case. However, this is also the most interesting case because nonlinear effects are fairly simple to quantify in this limit (nonlinear effects are represented by a simple cutoff in the limiting case). This simplicity leads to fairly simple approximations for cases in which the generation rate is finite but large enough for the approximations to apply.

Two versions of the analysis answer the same questions using different levels of mathematical rigor. Physical insight is most easily obtained from the less-rigorous version in the main text (a qualitative discussion in Section II followed by a quantitative discussion in Section III) because the physical concepts are discussed without being distracted by mathematical technicalities. However, three postulates (physically reasonable but without obvious guarantees) will be taken for granted in Section III, because this version of the analysis is not comprehensive enough to prove these postulates. That is the job of the more rigorous analysis in Appendix A. In fact, verification of these postulates is the only vital role of the more-rigorous analysis, so readers that are willing to believe the postulates in Section III without proof will not have to study this appendix.

⁵ The analysis is also tractable in the opposite limit of a small generation rate, but the limit of a large generation rate is more relevant to studies of single-event effects.

II. QUALITATIVE DISCUSSION OF THE AR AND HRR

A quantitative analysis will be easier to follow if some qualitative concepts are understood first. One of the examples in [1] is a good illustration. In this example, the location of the source in Fig. 1 was varied, and the charge-collection rate was calculated for each source location. This produced a plot of charge-collection efficiency (collection rate divided by creation rate) versus source location X_S shown in Fig. 2(a) (the other parts of Fig. 2 are discussed later). The example shown is characterized by the minority-carrier diffusion coefficient D_m being two times the majority-carrier diffusion coefficient D_M , and represents a p-type material in which the electron mobility is two (in this example) times the hole mobility. The labels distinguishing different curves in Fig. 2(a) refer to different source strengths. The source strength is denoted G and is the rate per area of electron-hole pair generation, so qG has the same dimensions as a current density and can be expressed in units of amperes per square centimeter (A/cm^2). The remaining symbols appearing in these labels consist of the doping density N and the length L of the QNR. The lowest curve in Fig. 2(a) agrees with textbook predictions for low-injection conditions, but our interest is in the upper-most curve. This curve is virtually identical to the limiting case of an infinite source strength, while the curve next to it is fairly well approximated by this limit. Our interest is in the high-injection limit, which we will call the “large-G limit” and which is well represented by the upper-most curve in Fig. 2(a). Note that this curve shows the presence of an SV. The SV width in this example (the width of the horizontal section of the curve) is $2L/3$. More generally, it was shown in [1] that the width is $D_m L / (D_m + D_M)$. Minority-carriers liberated within the SV are collected with a 100-percent efficiency in the large-G limit. The sloped portion of the upper curve in Fig. 2(a) implies that carriers liberated outside the SV are collected with an efficiency (in the large-G limit) that is consistent with pure diffusion from the source to the SV boundary. It will be seen in Section III that this is not a correct description of the actual charge-collection physics, but this model does correctly predict (in the large-G limit) the amount of collected charge from a localized source outside the SV; hence, the model is symbolic.

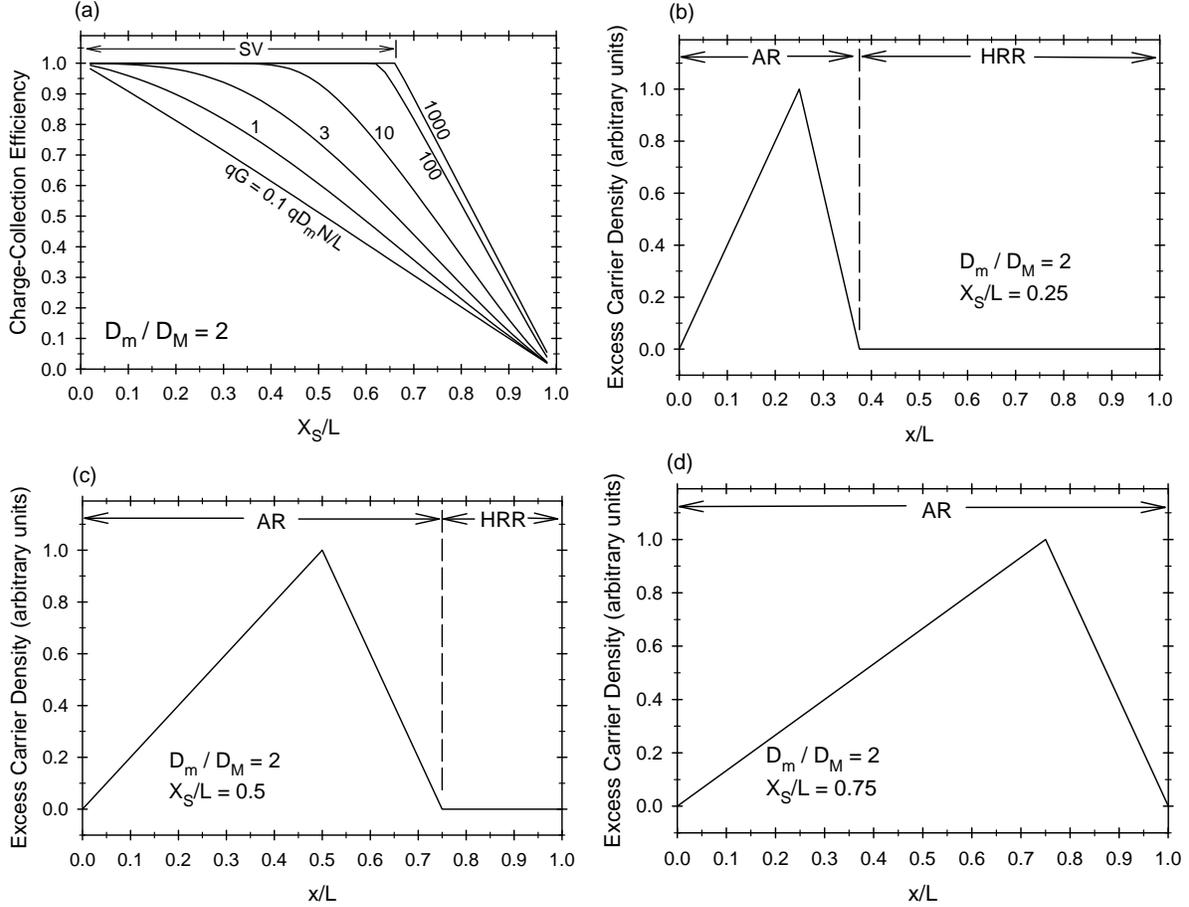


Figure 2. Charge-collection efficiency versus source location (a), and plots of carrier density for each of several source locations (b)–(d).

As already stated, the charge-collection model in the previous paragraph is symbolic in the sense of agreeing with the upper curve in Fig. 2(a) without explaining the physics. We now explain the physics. First consider a source location near the left end of the SV. This example source location is indicated by $X_S/L = 0.25$ in Fig. 2(b), which shows a qualitative plot of the excess carrier density versus the observation coordinate x within the QNR. Section III will explain why the carrier-density versus x curve is as shown in the figure. For this immediate discussion, the plot is regarded as a given. Note that there are two distinct regions in Fig. 2(b). The carrier density varies with x in the left region, but the excess carrier density is virtually zero in the right region. This partitioning of the QNR into distinct regions should not be confused with a partitioning into regions consisting of the SV and the region outside the SV. The latter partitioning is relevant to the location of the source, while the former partitioning is relevant to plots of carrier density versus x . If the source is within the SV, then a plot of carrier density versus x partitions into two regions as illustrated in Fig. 2(b). Also, partitioning the QNR into an SV and the region outside the SV is useful only for a localized source, while Section III will show that a partitioning into distinct regions illustrated for a localized source in Fig. 2(b) can be generalized to apply to an arbitrary spatial distribution of carrier generation. This partitioning was discovered in an earlier analysis [3], which assigned the name “ambipolar region” (AR) to the left region in Fig. 2(b), and the name “high-resistance region” (HRR) to the right region. The

first name was chosen because the carrier density satisfies the ambipolar diffusion equation in the left region. The second name was chosen because the electrical conductivity of the right region is comparable to the equilibrium conductivity, i.e., it is not significantly enhanced by excess carriers. The equilibrium QNR might not normally be thought of as “high resistance,” but it is high when compared to the AR. The large number of excess carriers in the AR gives this region a very large conductivity, i.e., a small resistance. Therefore, any voltage that is across the QNR (Section III derives an estimate for this voltage) is almost entirely across the HRR, with a much smaller voltage across the AR. A popular misconception is that there is a region (called a “funnel” in the literature) at the left side of the QNR that contains a strong electric field; a so-called “strong-field drift region.”⁶ In reality, the left side of the QNR is the AR, and it has a weak electric field. The strong field is in the HRR on the right.

As previously stated, there is a strong electric field at locations inside the HRR. With this fact taken as given for the present discussion, we can now see qualitatively the physical mechanisms that make the existence of an HRR possible. The strong electric field prevents minority carriers from entering the HRR from the left, and the ideal ohmic contact on the right side of the QNR (see Fig. 1) does not supply minority carriers; consequently, the HRR is nearly devoid of minority carriers. Majority carriers do enter the HRR from the left, but they move to the contact on the right without piling up inside the HRR because quasi-neutrality (together with a small number of minority carriers) keeps the majority-carrier density comparable to (or less than) the doping density within the HRR, which maintains the high resistance (“high” compared to the AR). The HRR is self sustaining; the low conductivity results in a strong electric field that maintains the low conductivity. This intuitive explanation makes the existence of an HRR credible, but it does not explain the conditions needed for an HRR to form and does not tell us how wide the HRR will be when it does form. Section III will answer these questions. A preview of an answer to come, for the special case of a localized source, is that an HRR forms (in the large-G limit) if the source is inside the SV.

Let us now return to Fig. 2(b). The AR on the left contains a weak electric field, as needed for the carrier density to be described by the ambipolar diffusion equation. The HRR on the right contains a strong electric field. This electric field prevents minority carriers from entering the HRR, so their only way out of the QNR is at the DRB.⁷ This explains why the charge-collection efficiency is 100 percent when the source is at the location indicated in Fig. 2(b). Similarly, if the source is moved farther from the DRB, but is still within the SV (as in Fig. 2(c)), we see from Fig. 2(c) that the HRR is narrower, but there is still an HRR, and the charge-collection efficiency is 100 percent. However, if the source is moved outside the SV, we see from Fig. 2(d) that there is no HRR. We can now expect the charge-collection efficiency to be less than 100 percent. The fact that the charge-collection efficiency follows the curve in Fig. 2(a) when the source is outside the SV will be explained in Section III.

⁶ This myth would not have started if investigators had paid more attention to the spacing between equipotential surfaces plotted by computer simulations, and to the fact that the electric field is weakest (not strongest) at locations where these plotted surfaces are farthest apart.

⁷ The AR has a weak electric field, but this does not mean that the electric field has no significance. The electric field is weak because the conductivity is correspondingly large (due to a large carrier density), so the electric field still influences the current. The influence is whatever is needed to make the minority carriers move to the DRB when they are blocked at the other end by the HRR.

III. HEURISTIC DERIVATION OF THE ADC MODEL IN 1D

Physical concepts are most easily understood from a one-dimensional (1D) version of the problem because simple illustrative plots make visualization easier. A comprehensive discussion of basic concepts is in this section, which considers the 1D problem. Once these concepts are understood, a 3D version discussed later in Section IV focuses on geometric issues because the physical concepts are the same as those discussed here.

A. Governing Equations

The quasi-neutral region is contained between two given points x_1 and x_2 on the x -axis, with the DRB at x_1 and an ideal ohmic contact at x_2 . Under steady-state conditions, and with recombination neglected in the QNR interior, the well-known drift-diffusion equations (which can be found in any textbook on semiconductors) reduce to

$$J_e(x) = q D_e \frac{dn(x)}{dx} - q \mu_e n(x) \frac{d\varphi(x)}{dx} \quad (1a)$$

$$J_h(x) = -q D_h \frac{dp(x)}{dx} - q \mu_h p(x) \frac{d\varphi(x)}{dx} \quad (1b)$$

$$\frac{dJ_e(x)}{dx} = -q g(x), \quad \frac{dJ_h(x)}{dx} = q g(x) \quad (2)$$

where $J_e(x)$ and $J_h(x)$ are the electron and hole current densities, 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), $\varphi(x)$ is the electrostatic potential, $n(x)$ and $p(x)$ are the electron and hole densities, and $g(x)$ is the carrier-generation rate density. The sign convention for the currents is such that J_h is positive when holes move in the direction of increasing x , and J_e is positive when electrons move in the direction of decreasing x . An additional equation is Poisson's equation relating the second derivative of φ to the carrier densities. A region is quasi-neutral when the solution to the complete set of equations can be approximated by the solution to the set of equations obtained by replacing Poisson's equation with

$$p = P + p_0, \quad n = P + n_0,$$

where P is the excess carrier density (taken to be the same for electrons and holes), p_0 is the equilibrium hole density, and n_0 is the equilibrium electron density. 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

$$\frac{J_e}{q D_e} = \frac{dP}{dx} - \frac{P + n_0}{V_T} \frac{d\varphi}{dx}, \quad -\frac{J_h}{q D_h} = \frac{dP}{dx} + \frac{P + p_0}{V_T} \frac{d\varphi}{dx} \quad (3)$$

which is a pair of simultaneous equations used to solve for both P and φ . 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

$$J_m \equiv \begin{cases} J_e & \text{for p-type} \\ -J_h & \text{for n-type} \end{cases}, \quad J_M \equiv \begin{cases} J_h & \text{for p-type} \\ -J_e & \text{for n-type} \end{cases} \quad (4a)$$

$$U \equiv \begin{cases} -\varphi & \text{for p-type} \\ \varphi & \text{for n-type} \end{cases}. \quad (4b)$$

Note that the subscript m to the 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, $J_m(x)$ is positive at any point x at which minority carriers are moving towards the DRB, and $J_M(x)$ is positive at any point x at which majority carriers are moving away from the DRB. Substituting (4) into (2) and (3) gives

$$\frac{J_m(x)}{qD_m} = \frac{dP(x)}{dx} + \frac{P(x)}{V_T} \frac{dU(x)}{dx}, \quad -\frac{J_M(x)}{qD_M} = \frac{dP(x)}{dx} - \frac{P(x) + N}{V_T} \frac{dU(x)}{dx} \quad (5)$$

$$\frac{dJ_m(x)}{dx} = -qg(x), \quad \frac{dJ_M(x)}{dx} = qg(x). \quad (6)$$

We will define G by

$$qG \equiv q \int_{x_1}^{x_2} g(\xi) d\xi \quad (7)$$

so qG is the total rate of charge liberation. It has the same dimensions as a current density, i.e., acceptable units are A/cm^2 . Our interest is in cases in which G is large, so we look for approximations that become exact in the large- G limit.

B. Derivation of the Charge-Collection Efficiency

Adding the two equations in (5) gives

$$\frac{J_m(x)}{2qD_m} - \frac{J_M(x)}{2qD_M} = \frac{dP(x)}{dx} - \frac{N}{2V_T} \frac{dU(x)}{dx}. \quad (8)$$

Unlike the two individual currents in (5), the particular combination given by (8) does not have P in the coefficient to dU/dx . In the large- G limit, P and its derivative become large, but the coefficient to dU/dx in (8) does not. This suggests that an approximation that becomes accurate in the large- G limit can be obtained by omitting the dU/dx term in (8). From the qualitative discussion in Section II, we can anticipate that omitting this term is correct when the point x is sufficiently close to the DRB because the point will be in the AR (a weak-field region). Omitting this term throughout the entire QNR may or may not give a correct result because there may or

may not be an HRR. The approach used here is to omit the term, find out what the implications are, and see where this takes us. With this term omitted, (8) reduces to

$$\frac{J_m(x)}{2qD_m} - \frac{J_M(x)}{2qD_M} \approx \frac{dP(x)}{dx} \quad (\text{if } E \text{ can be neglected}). \quad (9)$$

The electric field E , which is $-d\phi/dx$ in general, is dU/dx for the p-type material, or $-dU/dx$ for the n-type, so the qualifier “if E can be neglected” is interpreted as “if dU/dx can be neglected in (8).”

We now assume that the DR is reverse-biased. To produce a reverse-biasing condition, it is not enough that the power supply voltage have the correct polarity. The reason is that there can sometimes be a significant voltage across the QNR. The power supply must have enough voltage to supply the QNR voltage and still have enough left over to reverse-bias the DR. The required voltage will be estimated later. For now we take it as given that the DR is reverse-biased. We also specialize to the case in which all carrier liberation is in the QNR on the right side of the DR (see Fig. 1), i.e., carriers are not generated anywhere else within the device. With these assumptions, the DR blocks the majority-carrier current, i.e., $J_M(x_1) = 0$. The total current J_T (defined to be $J_m + J_M$) is constant in x and can be calculated from $J_T = J_m(x_1) + J_M(x_1) = J_m(x_1)$. Therefore, when (9) is evaluated at $x = x_1$ (where we expect (9) to be accurate in the large- G limit even though it has unknown accuracy when evaluated at points further in the QNR), we can write the result as

$$J_T = 2qD_m \left. \frac{dP(x)}{dx} \right]_{x=x_1}. \quad (10)$$

The minority-carrier diffusion current is defined to be $qD_m dP/dx$, so (10) states that the total current is twice the minority-carrier diffusion current evaluated at the DRB. This implies that the electric field becomes whatever is needed to make the minority-carrier drift current at the DRB equal to the minority-carrier diffusion current at the DRB. Another implication of (9) is obtained by differentiating while using (6) to get

$$\frac{d^2P(x)}{dx^2} \approx -\frac{1}{D^*} g(x) \quad (\text{if } E \text{ can be neglected}) \quad (11)$$

where the ambipolar diffusion coefficient D^* is defined by

$$\frac{1}{D^*} = \frac{1}{2D_m} + \frac{1}{2D_M}. \quad (12)$$

Note that the coefficient D^* that appears in (11), and is used to calculate the gradient of P , is different than the coefficient $2D_m$ that appears in (10) and relates the current to the gradient of P . Using terminology in [4], we would call the right side of (10), when P is calculated from (11), a “nonstandard diffusion current” because these coefficients are different. In contrast,

textbook treatments of low injection-levels would calculate the current by using the coefficient D_m (instead of $2D_m$ because drift is neglected) in (10), and would also use the same coefficient D_m (instead of D^* because the carrier density satisfies the minority-carrier diffusion equation instead of the ambipolar diffusion equation under low-density conditions) in (11). Using terminology in [4], these textbook equations describe a standard diffusion current. The traditional interpretation of a “minority-carrier diffusion current” refers to these textbook equations, so to emphasize that different coefficients are used here, we will use different terminology. The phrase “ambipolar diffusion current” will refer to the right side of (10) when P satisfies (11). Alternate terminology that will also be used is “a nonstandard diffusion current,” which emphasizes the fact that the coefficient D^* in (11) is different than the coefficient $2D_m$ in (10).

Computer simulations have shown that the carrier density at the DRB can be much larger than the doping density when G is sufficiently large, even when the DR is reverse-biased. However, the carrier density is still much smaller at the DRB (when the DR is reverse-biased) than it is elsewhere within the QNR. Therefore, for the purpose of solving (11) for P , we can use $P(x_1) = 0$. The excess density is also zero at the ideal ohmic contact, so $P(x_2) = 0$. Solving (11) subject to these boundary conditions and substituting the result into (10), and using (12) to write the coefficient $2D_m/D^*$ as $1 + D_m/D_M$, gives

$$J_T = \left(1 + \frac{D_m}{D_M}\right) q \int_{x_1}^{x_2} \frac{x_2 - \xi}{x_2 - x_1} g(\xi) d\xi \quad (\text{if } E \text{ can be neglected}). \quad (13)$$

The validity of (13) is conditional because it requires that the dU/dx term can be neglected in (8). A bound for the current that is unconditionally valid can be obtained by noting that the excess carrier density P is nonnegative, but it is zero at x_2 , which implies that dP/dx is less than or equal to zero when evaluated at $x = x_2$. Using this together with $P(x_2) = 0$ with the first equation in (5) gives $J_m(x_2) \leq 0$. Now use this result when integrating the first equation in (6) between x_1 and x_2 , and use $J_T = J_m(x_1)$, to get

$$J_T \leq q \int_{x_1}^{x_2} g(\xi) d\xi = qG. \quad (14)$$

In other words, the charge-collection rate cannot exceed the total carrier-generation rate. We can express (13) and (14) in terms of the charge-collection efficiency α defined by

$$\alpha \equiv \frac{J_T}{qG} = \frac{J_T}{q \int_{x_1}^{x_2} g(\xi) d\xi}. \quad (15)$$

We also define

$$\alpha^* \equiv \left(1 + \frac{D_m}{D_M}\right) \frac{\int_{x_1}^{x_2} \frac{x_2 - x}{x_2 - x_1} g(x) dx}{\int_{x_1}^{x_2} g(x) dx} \quad (16)$$

so that (13) and (14) can be written as

$$\alpha = \alpha^* \quad (\text{if } E \text{ can be neglected}) \quad (17)$$

$$\alpha \leq 1. \quad (18)$$

The integral in the numerator in (16) cannot exceed the integral in the denominator, but they can be nearly equal if the carrier generation is sufficiently concentrated near the DRB. For example, if $g(\xi)$ differs from zero only where $\xi \approx x_1$, the two integrals will be nearly equal. Because of the coefficient $1 + D_m/D_M$ in (16), α^* will be greater than 1 when the integrals are nearly equal. This example is an extreme case in which carrier generation is confined to within a narrow region adjacent to the DRB and was selected for illustration. More generally, α^* will be greater than 1 if the “center of the carrier generation” (defined the same way as a center of mass except that the generation rate density g replaces a mass density) is sufficiently close to the DRB. This condition results in a contradiction between the estimate in (17) and the upper bound in (18). This contradiction implies that it was not valid to omit the dU/dx term in (8). This suggests that there is an HRR, which in turn suggests that $\alpha = 1$. The heuristic analysis given here accepts these suggestions as a postulate, i.e., the postulate is:

Postulate 1: If $\alpha^ > 1$ then, in the large- G limit, there is an HRR, and $\alpha = 1$.*

Based on the analysis given here, the postulate is reasonable but not absolutely certain. The more rigorous analysis in Appendix A confirms that this postulate is correct. Now suppose $\alpha^* \leq 1$ so we have no obvious contradiction. The fact that we are not aware of a contradiction from (17) does not imply that the conclusion (17) is correct, but we still postulate that:

Postulate 2: If $\alpha^ \leq 1$ then, in the large- G limit, there is no HRR, and $\alpha = \alpha^*$.*

Again, based on the analysis given here, the postulate is reasonable but not absolutely certain. Again, the more rigorous analysis in Appendix A confirms that this postulate is correct. The final conclusion is that, in the large- G limit, the charge-collection efficiency α is given by

$$\alpha = \begin{cases} \alpha^* & \text{if } \alpha^* \leq 1 \\ 1 & \text{if } \alpha^* > 1 \end{cases} \quad (\text{large - } G \text{ limit}). \quad (19)$$

Note that α^* was calculated from the ambipolar diffusion current, which is the right side of (10) when P satisfies (11), but (19) includes a cutoff that sets α equal to 1 when α^* exceeds 1. This motivated the name “ambipolar diffusion with a cutoff” for the model in which α is calculated from (19).

C. Showing Consistency with the Symbolic Model

We will first show that the literal model (the ADC model) produces the upper curve in Fig. 2(a), and then explain why the symbolic model (the SV model) is also consistent with this curve in spite of its loose connection with actual physics.

A quantitative prediction of the upper curve in Fig. 2(a) is simple when using the literal model. For a localized source we have $g(\xi) = G\delta(\xi - X_S)$, where $\delta(\xi - X_S)$ is the Dirac delta function with source location X_S , which can be anywhere within the QNR in this analysis. Substituting this g into (16), and letting $x_1 = 0$ so that X_S is the distance from the DRB to the source, and x_2 is the QNR length L , we find that (19) becomes

$$\alpha = \begin{cases} \left(1 + \frac{D_m}{D_M}\right)\left(1 - \frac{X_S}{L}\right) & \text{if } X_S \geq \frac{D_m}{D_m + D_M}L \\ 1 & \text{if } X_S < \frac{D_m}{D_m + D_M}L \end{cases} \quad (\text{localized source}) \quad (20)$$

which agrees with the upper curve in Fig. 2(a).

If we were not already aware of the literal model, a casual inspection of the upper curve in Fig. 2(a) would strongly suggest the symbolic model. Recall that the symbolic model discussed in Section I applies when the source is localized in the sense of having a specific location X_S , which is the case represented in Fig. 2(a) and is the case considered here. Consider a source far from the DRB. In the symbolic model, this is interpreted to mean that the source is outside the SV illustrated in Fig. 2(a) (in the literal model, this is interpreted to mean that there is no HRR), so the source location is in the sloped portion of the curve. It was stated in Section I that the sloped portion of the upper curve in Fig. 2(a) implies that charge collection is consistent with pure diffusion from the source to the SV boundary. The meaning of “pure diffusion” in this statement is that the current is a standard diffusion current, and the meaning of “to the SV boundary” is that the SV boundary behaves as a sink for excess carriers. This is a symbolic model because the actual current is a nonstandard diffusion current⁸, and the actual sink is the DRB. The symbolic model will give a correct prediction of collected charge if the SV boundary location is selected so that a standard diffusion current to this artificial sink boundary is the same as the actual nonstandard diffusion current produced when the DRB is a sink. In other words, the SV boundary is an artificial boundary selected to make the symbolic model give correct predictions of collected charge. The symbolic model is not comprehensive enough to tell us where the SV boundary is, so a more literal model is needed to tell us where the SV boundary must be placed in order to make the symbolic model give correct predictions of collected charge. The collection efficiency for a standard current becomes 100 percent when the source location is moved to the charge-collecting sink, so the SV boundary location needed to make the symbolic model give correct predictions satisfies the condition that $\alpha = 1$ when the source is at this location, and α varies linearly with source location when the source is further from the DRB.

⁸ The right side of (10) when P satisfies (11) is not a standard current because the coefficient $2D_m$ in (10) is not the same as the coefficient D^* in (11).

The more literal model tells us, via (20), that the SV boundary location that satisfies this condition is at $x = D_m L / (D_m + D_M)$. In summary, the SV boundary is an artificial boundary selected to make the symbolic model give correct predictions of collected charge, but a more literal model is needed to predict where the SV boundary must be placed in order to do this.

D. Derivation of the HRR Width

The estimate of the amount of collected charge, derived in Section III.B, does not require that we know how to calculate the width of the HRR when there is an HRR. However, the HRR width is still interesting, at least from an academic point of view. When an HRR is created, there is an AR on the left and an HRR on the right (in Fig. 2(b) for example). The demarcation between these regions will be called the AR boundary (ARB). The goal is to solve for the location of this boundary, denoted x_{ARB} . There is an HRR if and only if the generation function g has the property that α^* defined by (16) is greater than 1, so we consider the case in which this condition is satisfied.

Anticipating that the AR on the left is a weak-field region suggests that (11) applies when the point of evaluation x is inside the AR. Also, with an HRR on the right, we expect a negligible excess carrier density in this region. The heuristic analysis given here accepts this as a postulate. That postulate is:

Postulate 3: If $\alpha^ > 1$, then, in the large- G limit, there exists an interval adjacent to the DRB such that (11) applies at each x in this interval, and $P(x)$ is negligibly small when x is outside this interval.*

The interval mentioned in this postulate is taken to be the definition of the AR, and “outside the interval” is taken to be the definition of the HRR. Based on the arguments given here, the postulate is reasonable but not absolutely certain. The more rigorous analysis in Appendix A confirms that this postulate is correct. The more rigorous analysis replaces the phrase “negligibly small” with a quantitative estimate of the carrier density in the HRR, but the implication of this estimate is that the excess carrier density is much smaller in the HRR than in the AR, even though the excess density in the HRR can be comparable to the doping density in the large- G limit.

The analysis that used (10), (11), and (12) to derive (13) is invalid when there is an HRR because (11) does not apply when x is outside the AR. However, Postulate 3 implies that the analysis becomes valid if the electrode coordinate x_2 is replaced by the ARB coordinate x_{ARB} . The result is (13) but with x_2 replaced by x_{ARB} . Also, the case considered is described by $\alpha^* > 1$ (so that there will be an HRR), which implies $\alpha = 1$, i.e., J_T is the total generation rate qG . Replacing J_T with qG on the left side of (13), and replacing x_2 with x_{ARB} on the right side of (13), the equation used to solve for x_{ARB} becomes

$$q \int_{x_1}^{x_2} g(\xi) d\xi = \left(1 + \frac{D_m}{D_M}\right) q \int_{x_1}^{x_{ARB}} \frac{x_{ARB} - \xi}{x_{ARB} - x_1} g(\xi) d\xi. \quad (21)$$

The point x_{ARB} has a physical interpretation. Recall that the electric field becomes whatever is needed to make the drift current become whatever is needed to make the total current equal to the nonstandard diffusion current, which is the right side of (10) when P satisfies (11). The diffusion current is (in turn) controlled by boundary conditions, not only at the DRB but also by the location of the other sink boundary (the boundary opposite the DRB), so the diffusion current depends on the location of the ARB. The left side of (21) is what the current is required to be, and it is the total generation rate. The right side is the nonstandard diffusion current when the other sink boundary is the ARB. In other words, the point x_{ARB} is the location at which the other sink boundary (the ARB) must be placed in order to make the nonstandard diffusion current equal to the total carrier-generation rate. The point x_{ARB} is on the left of the boundary x_2 because this reduces the diffusion current compared to what the current would be if $x_{ARB} = x_2$. This reduction is necessary because the latter current is too large (the charge-collection efficiency representing the latter current is α^* , which is greater than 1 and therefore is too large). The point x_{ARB} is on the left of x_2 so that the total current does not exceed the total generation rate.

Some of the above discussion of the ARB might be reminiscent of the discussion in Section III.C of the SV boundary. In both cases, the boundary location is what is needed to make some current satisfy some condition. An important distinction between the two cases is that the SV boundary is an artificial boundary because the location is selected to make the calculated current equal to the actual current when the calculated current is a fictitious standard diffusion current instead of the physically correct nonstandard diffusion current. In contrast, the ARB location makes the calculated current equal to the actual current when the calculated current is the physically correct nonstandard diffusion current. Hence, the ARB is a physical boundary instead of an artificial boundary. It separates physically distinguishable regions; the AR and HRR. Also, the SV boundary is a useful (albeit, artificial) concept for a single localized source, but not for an arbitrary spatial distribution of carrier generation. In contrast, the ARB is a physically definable boundary for the general case.

For the special case of a localized source, we can substitute $g(\xi) = G\delta(\xi - X_S)$ into (21) and rearrange some terms to get

$$\int_{x_1}^{x_{ARB}} \frac{x_{ARB} - \xi}{x_{ARB} - x_1} \delta(\xi - X_S) d\xi = \frac{D_M}{D_m + D_M}.$$

A necessary condition for this equation to be satisfied is that $x_{ARB} > X_S$, because otherwise the integral will be zero. Observing this condition when evaluating the integral, the resulting equation can be solved for x_{ARB} . Taking x_1 to be zero so that x_{ARB} and X_S each refer to a distance from the DRB, the result can be expressed as

$$x_{ARB} = \left(1 + \frac{D_M}{D_m}\right) X_S \quad (\text{localized source}). \quad (22)$$

Note that (22) is consistent with the requirement that $x_{ARB} > X_S$. However, another requirement is that the ARB be contained within the QNR, i.e., $x_{ARB} \leq L$. In order for the solution to (22) to

satisfy this condition, the source location X_S must satisfy $X_S \leq D_m L / (D_m + D_M)$. This is consistent with what we already know; that a localized source must be inside the SV in order for an HRR to form. For a more general spatial distribution of carrier generation, the concept of an SV no longer applies, and it is necessary to use the more general equation (21) to solve for x_{ARB} . There will be a solution consistent with $x_{ARB} < L$ (i.e., there will be an HRR) if g has the property that α^* defined by (16) is greater than 1.

E. Voltage Needed to Reverse-Bias the DR

The voltage across the QNR can be calculated by integrating (8) between x_1 and x_2 while using $P(x_1) \approx 0$ and $P(x_2) = 0$ to get

$$\Delta U \equiv U(x_2) - U(x_1) = \frac{V_T}{N} \int_{x_1}^{x_2} \left\{ \frac{J_M(\xi)}{q D_M} - \frac{J_m(\xi)}{q D_m} \right\} d\xi. \quad (23)$$

Using the endpoint values $J_m(x_1) = J_T = \alpha q G$ and $J_M(x_1) = 0$ when integrating (6) gives

$$\frac{J_M(x)}{q D_M} - \frac{J_m(x)}{q D_m} = \frac{2}{D^*} \int_{x_1}^x g(\xi) d\xi - \frac{\alpha G}{D_m} \quad (24)$$

and substituting this into (23) gives

$$\Delta U = \frac{L V_T}{N} \left\{ \frac{2}{D^*} \int_{x_1}^{x_2} \frac{x_2 - \xi}{x_2 - x_1} g(\xi) d\xi - \frac{\alpha G}{D_m} \right\}.$$

We can shorten the notation by using the definition (16) of α^* to write this as

$$\Delta U = (\alpha^* - \alpha) \frac{L V_T G}{N D_m}. \quad (25)$$

If we use (19) to evaluate α in (25) we will calculate a zero voltage (because $\alpha = \alpha^*$) when $\alpha^* \leq 1$ (i.e., when there is no HRR). Recall that a negligible electric field was assumed when deriving $\alpha = \alpha^*$, so it is not a surprise that we predict what we assumed when substituting $\alpha = \alpha^*$ back into (25). To look at this more carefully, note that (25) does not require that we take a large- G limit to become exact, while (19) does. As G is made larger, (19) becomes more accurate, so the term $\alpha^* - \alpha$ becomes smaller (when $\alpha^* \leq 1$), but this quantity that is becoming smaller is multiplying a G that is becoming larger on the right side of (25). Therefore, when $\alpha^* \leq 1$ (so there is no HRR), it is not obvious from (25) whether we can regard ΔU as large or small when compared to some absolute standard (e.g., one volt). However, it is clear that ΔU given by (25) is much smaller when $\alpha^* \leq 1$ (so that $\alpha \approx \alpha^*$) than it is when $\alpha^* > 1$ (so that $\alpha \approx 1$). In other words, even though it is not clear whether we can regard ΔU as large or small (when compared to some absolute standard) when there is no HRR, it is clearly much smaller when there is no HRR than when there is an HRR. When there is an HRR is also when (25)

becomes useful because we do not have to try to figure out how to multiply a quantity that becomes small by another quantity that becomes large. To calculate ΔU , we first use (16) to calculate α^* . If $\alpha^* > 1$ then there is an HRR, so we use $\alpha = 1$. We then calculate ΔU by substituting these numbers into (25). Note that in order for the DR to be reverse-biased, as assumed throughout the analysis, the power supply that biases the diode must have enough voltage to supply the QNR voltage ΔU , plus an additional amount of voltage that will appear as a bias voltage across the DR.

IV. HEURISTIC DERIVATION OF THE ADC MODEL IN 3D

With physical concepts already discussed in previous sections, a three-dimensional (3D) version of the analysis becomes fairly obvious from the point of view of physical concepts. Therefore, physical discussions in this section are brief. Most of the attention here is on the mathematical techniques suitable for a 3D geometry. An example 3D device geometry is illustrated in Fig. 3. The surface S_1 is the DRB, and the surface S_2 is the electrode contact. All other boundary surfaces are reflective. It will be tacitly assumed throughout this section that the medium identified as a QNR in Fig. 3 really is a QNR, i.e., there are no charge layers adjacent to reflective boundaries of the type seen under the insulated and biased gate of a metal oxide semiconductor field-effect transistor (MOSFET). Boundary conditions satisfied by P are $P(\vec{x}) = 0$ when the point \vec{x} is on the surface S_2 , $P(\vec{x}) \approx 0$ when the point \vec{x} is on the surface S_1 (given that the DR is reverse biased) and P satisfies reflective boundary conditions (the normal component of the gradient is zero) on the reflective boundaries.

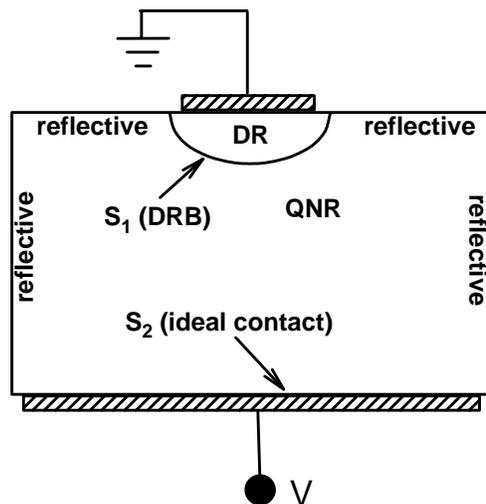


Figure 3. An example 3D geometry shows the notation. The surface S_1 is the DRB, and the surface S_2 is the electrode contact to the QNR. All other boundary surfaces surrounding the QNR are reflective. The spatial distribution of carrier generation within the QNR is arbitrary.

A. Governing Equations

The 3D version of (7) is

$$qG \equiv q \int_{QNR} g(\vec{x}) d^3x. \quad (26)$$

Note that qG has different dimensions here than it had in the 1D analysis. In the 1D analysis, qG had the dimensions of a current density, and acceptable units were amperes per square centimeter. Here, it has the dimensions of a current and acceptable units are amperes. This is

convenient because our interest now is in terminal currents instead of current densities. The terminal currents I_m and I_M (for minority carrier and majority carrier, respectively) are defined by

$$I_m \equiv - \int_{S_1} \vec{J}_m(\vec{x}) \circ d\vec{S}, \quad I_M \equiv - \int_{S_1} \vec{J}_M(\vec{x}) \circ d\vec{S}. \quad (27)$$

The normal unit vector in these 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 a positive current. Minority carriers will be flowing towards the DRB from the QNR interior, which is in the direction of the outer-normal unit vector in the surface integral, but the direction of \vec{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. Given a reverse-biasing condition we will have $I_M = 0$, so

$$I_T \equiv I_m + I_M = I_m. \quad (28)$$

The 3D version of (6) is

$$\vec{\nabla} \circ \vec{J}_m(\vec{x}) = -q g(\vec{x}), \quad \vec{\nabla} \circ \vec{J}_M(\vec{x}) = q g(\vec{x})$$

which gives

$$\vec{\nabla} \circ \left[\frac{\vec{J}_m(\vec{x})}{2q D_m} - \frac{\vec{J}_M(\vec{x})}{2q D_M} \right] = -\frac{1}{D^*} g(\vec{x}) \quad (29)$$

with D^* defined by (12). The 3D versions of (8) and (9) are

$$\frac{\vec{J}_m(\vec{x})}{2q D_m} - \frac{\vec{J}_M(\vec{x})}{2q D_M} = \vec{\nabla} P(\vec{x}) - \frac{N}{2V_T} \vec{\nabla} U(\vec{x}) \quad (30)$$

$$\frac{\vec{J}_m(\vec{x})}{2q D_m} - \frac{\vec{J}_M(\vec{x})}{2q D_M} \approx \vec{\nabla} P(\vec{x}) \quad (\text{if } E \text{ can be neglected}). \quad (31)$$

When E can be neglected, which is a good approximation at locations near the DRB, we can use (31) together with (27) and (28) to write I_T as

$$I_T = -2q D_m \int_{S_1} \vec{\nabla} P(\vec{x}) \circ d\vec{S}. \quad (32)$$

Also, when E can be neglected, we can combine (31) with (29) to get

$$\nabla^2 P(\vec{x}) \approx -\frac{1}{D^*} g(\vec{x}) \quad (\text{if } E \text{ can be neglected}). \quad (33)$$

B. Derivation of the Charge-Collection Efficiency

In order to continue with (32) and (33), it is helpful to define the function $\Omega(\vec{x})$ by the boundary-value problem

$$\Omega(\vec{x}) = 1 \text{ when } \vec{x} \text{ is on } S_1, \Omega(\vec{x}) = 0 \text{ when } \vec{x} \text{ is on } S_2 \quad (34a)$$

$$\nabla^2 \Omega(\vec{x}) = 0 \text{ in QNR interior} \quad (34b)$$

with reflective boundary conditions always assumed on the reflective boundaries. Using (34a) together with the boundary conditions for P , and using the fact that both functions satisfy reflective boundary conditions on the reflective boundaries, we have

$$\int_{S_1} \vec{\nabla} P \circ d\vec{S} = \oint \Omega \vec{\nabla} P \circ d\vec{S} - \oint P \vec{\nabla} \Omega \circ d\vec{S}$$

where the surface integrals on the right are over the closed surface containing the QNR. Green's theorem applied to the right side gives

$$\int_{S_1} \vec{\nabla} P \circ d\vec{S} = \int_{QNR} \Omega(\vec{x}) \nabla^2 P(\vec{x}) d^3x$$

where we used (34b) to eliminate an integral on the right. When E can be neglected, we can use (33) with the above equation to get

$$-D^* \int_{S_1} \vec{\nabla} P \circ d\vec{S} = \int_{QNR} \Omega(\vec{x}) g(\vec{x}) d^3x \quad (\text{if } E \text{ can be neglected}). \quad (35)$$

Combining (35) with (32) and using (12) gives

$$I_T = \left(1 + \frac{D_m}{D_M}\right) q \int_{QNR} \Omega(\vec{x}) g(\vec{x}) d^3x \quad (\text{if } E \text{ can be neglected}). \quad (36)$$

We now define α and α^* by

$$\alpha \equiv \frac{I_T}{qG} \quad (37)$$

$$\alpha^* \equiv \left(1 + \frac{D_m}{D_M}\right) \frac{q \int_{QNR} \Omega(\vec{x}) g(\vec{x}) d^3x}{qG} = \left(1 + \frac{D_m}{D_M}\right) \frac{\int_{QNR} \Omega(\vec{x}) g(\vec{x}) d^3x}{\int_{QNR} g(\vec{x}) d^3x} \quad (38)$$

so that (36) can be written as

$$\alpha = \alpha^* \quad (\text{if } E \text{ can be neglected}). \quad (39)$$

The two postulates in Section III.B give

$$\alpha = \begin{cases} \alpha^* & \text{if } \alpha^* \leq 1 \\ 1 & \text{if } \alpha^* > 1 \end{cases} \quad (\text{large - G limit}). \quad (40)$$

Note that the 3D results consisting of (38) and (40) are analogous to the 1D results consisting of (16) and (19). The $\Omega(\vec{x})$ that appears in (38) and is defined by (34) is the 3D version of the $(x_2-x)/(x_2-x_1)$ that appears in (16).

C. An Example Geometry: The Isolated Disc

An example 3D geometry considered here is good for illustration because it leads to a fairly simple analysis. It describes the case in which the DRB resembles a flat circular disc, and in which all other device structures are sufficiently far away from the disc so that the QNR below the DRB can be treated as having an effectively infinite extent. The geometry is shown in Fig. 4 (the “generation cylinder” in the figure is for a later discussion and is not relevant to this discussion). The DR is the shaded region, and the DRB (a.k.a., the surface S_1) is the lower boundary of this region. The DRB is a circular disc with radius r_D . The portion of the upper horizontal line that is above the DR represents a surface that includes an electrical contact to the DR, but the remainder of this horizontal line represents a reflective surface. The medium below this line, excluding the DR, is the QNR. The QNR extends to an effectively infinite distance in the lateral directions and in the downward direction, so the QNR contact (a.k.a., the surface S_2) is the infinite lower hemisphere. We will call this geometry the “isolated disc” geometry. Note that even with the infinite dimensions, there is still a finite spreading resistance between the DRB and the QNR contact, so these surfaces are not electrically isolated even though there is a large (effectively infinite) spatial separation. The spreading resistance for a sufficiently large spatial separation between S_1 and S_2 can be approximated by the spreading resistance for an infinite separation. The coordinate system is also shown in Fig. 4. The origin is at the disc center and the vertical coordinate z is positive at points within the QNR (below the disc). The radial coordinate r is $(x^2+y^2)^{1/2}$ and measures distance from the vertical center line.

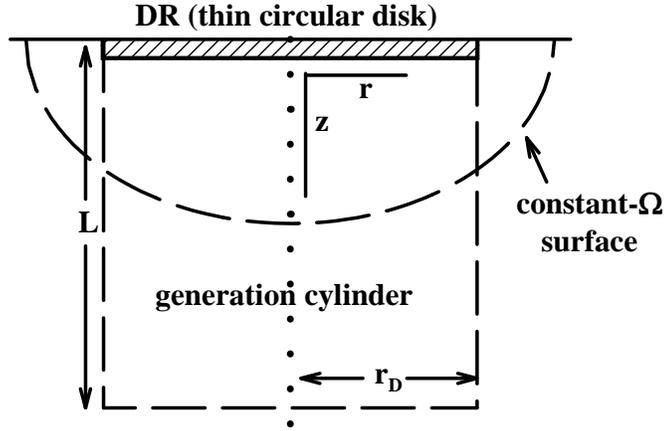


Figure 4. A more specific 3D geometry in which the DRB resembles an isolated disc.

The solution to (34) for this geometry is the electrostatic potential produced by a conducting disc at unit potential relative to the infinite distance at zero potential. The solution can be found in any textbook that treats electrostatic boundary-value problems, and one way to express $\Omega(x, y, z)$ as a function of the coordinates is with the equation

$$\left(x^2 + y^2\right) \sin^2\left(\frac{\pi}{2} \Omega(x, y, z)\right) + z^2 \tan^2\left(\frac{\pi}{2} \Omega(x, y, z)\right) = r_D^2 \quad (\text{isolated disc}). \quad (41)$$

This format for expressing the solution has an advantage over an alternate (and more cumbersome) format that expresses $\Omega(x, y, z)$ explicitly in terms of the coordinates. The advantage is that the geometry of constant- Ω surfaces can be seen by inspection. To be more explicit, select a positive v that is less than or equal to 1. We can recognize the set of points (x, y, z) satisfying $\Omega(x, y, z) = v$ because these are the points on the ellipsoid given by

$$\left(x^2 + y^2\right) \sin^2\left(\frac{\pi}{2} v\right) + z^2 \tan^2\left(\frac{\pi}{2} v\right) = r_D^2 \quad (\text{defines } \Omega = v \text{ surface}). \quad (42)$$

An example of a constant- Ω surface is shown in Fig. 4.

Fig. 2(a) plotted charge-collection efficiency versus source location for the 1D problem when the source is localized to a constant- x plane. This is the 1D version of a point source. In 3D, we can consider a true point source and identify those source locations that will produce a 100-percent collection efficiency in the large- G limit. In other words, we can determine the location of the SV boundary (but remember that the concept of an SV is useful only for point sources). We will do that now for the isolated disc geometry.

When carrier generation is produced by a point source with source coordinates (X_S, Y_S, Z_S) , (38) reduces to

$$\alpha^* = \left(1 + \frac{D_m}{D_M}\right) \Omega(X_S, Y_S, Z_S) \quad (\text{point source}). \quad (43)$$

The collection efficiency is 100 percent when α^* is greater than or equal to 1, and the SV boundary is the set of points (X_S, Y_S, Z_S) for which the right side of (43) is equal to 1. Therefore, SV boundary points are the points (x, y, z) satisfying

$$\Omega(x, y, z) = \frac{D_M}{D_m + D_M}.$$

In other words, the SV boundary is the constant- Ω surface obtained from (42) when $v = D_M/(D_m + D_M)$. The equation for this surface is

$$(x^2 + y^2) \sin^2 \left(\frac{\pi}{2} \frac{D_M}{D_m + D_M} \right) + z^2 \tan^2 \left(\frac{\pi}{2} \frac{D_M}{D_m + D_M} \right) = r_D^2 \quad (\text{SV boundary}). \quad (44)$$

For example, if $D_m = 2D_M$, then the SV boundary is the ellipsoid that intersects the axis of symmetry at a depth (measured from the DRB) of about $1.73 \times r_D$.

D. Voltage Needed to Reverse-Bias the DR

The QNR voltage ΔU , i.e., the voltage between S_1 and S_2 , can also be derived in 3D. This is the minimum voltage that a power supply must have in order to reverse-bias the DR. To determine the QNR voltage, given that the DR is reversed-biased so that the analysis given here is valid, we go back to (30). The 1D analysis prepared us for the fact that approximations that will be used will not be useful for estimating ΔU when there is no HRR. The reason is that the approximations were derived by neglecting ΔU in (30), so substituting these approximations back into (30) just brings us back to $\Delta U = 0$. However, a useful estimate can be obtained when there is an HRR, which is also the condition that produces the largest voltage. We start with an exact analysis and then use an approximation in the last step. For this purpose, we define P^* by the boundary-value problem

$$P^*(\vec{x}) = 0 \text{ when } \vec{x} \text{ is on } S_1, P^*(\vec{x}) = 0 \text{ when } \vec{x} \text{ is on } S_2 \quad (45a)$$

$$\nabla^2 P^*(\vec{x}) = -\frac{1}{D^*} g(\vec{x}) \text{ in QNR interior.} \quad (45b)$$

Note that approximations used for P when E can be neglected are the same equations that define P^* . The difference between P and P^* is that these equations are exact for P^* (by definition) regardless of whether there is an HRR. A result that will be needed later is obtained by using the same steps that produced (35) to get

$$-D^* \int_{S_1} \vec{\nabla} P^* \circ d\vec{S} = \int_{QNR} \Omega(\vec{x}) g(\vec{x}) d^3x \quad (46)$$

which is exact regardless of whether there is an HRR. Now use (45b) with (29) and (30) to get

$$\nabla^2 \left[P(\vec{x}) - P^*(\vec{x}) - \frac{N}{2V_T} U(\vec{x}) \right] = 0$$

i.e., the square bracket satisfies Laplace's equation. Setting the square bracket equal to the solution to Laplace's equation that has the same boundary values as the square bracket gives

$$P(\vec{x}) - P^*(\vec{x}) - \frac{N}{2V_T} U(\vec{x}) = -\frac{N}{2V_T} [(U_1 - U_2)\Omega(\vec{x}) + U_2]$$

where U_1 is the boundary value of U on S_1 , and U_2 is the boundary value of U on S_2 . Substituting this equation into (30) and using $\Delta U \equiv U_2 - U_1$ gives

$$\frac{\vec{J}_m(\vec{x})}{2qD_m} - \frac{\vec{J}_M(\vec{x})}{2qD_M} = \vec{\nabla} P^*(\vec{x}) + \frac{N\Delta U}{2V_T} \vec{\nabla} \Omega(\vec{x}).$$

Integrating this equation on the surface S_1 while using (27), (28), and (46), and solving for ΔU gives

$$\Delta U = \frac{V_T}{N} \frac{\frac{2}{D^*} \int_{QNR} \Omega(\vec{x}) g(\vec{x}) d^3x - \frac{I_T}{qD_m}}{\int_{S_1} \vec{\nabla} \Omega \circ d\vec{S}}.$$

Using (37) and (38), we can write this as

$$\Delta U = \frac{\alpha^* - \alpha}{\int_{S_1} \vec{\nabla} \Omega \circ d\vec{S}} \frac{V_T G}{N D_m}. \quad (47)$$

Note that the 3D result (47) is analogous to the 1D result (25). The reciprocal of the surface integral in (47), with $\Omega(\vec{x})$ defined by (34), is the 3D version of the L that appears in (25). This surface integral can be related to the electrical resistance through an ohmic medium having the same geometry as the QNR and with electrical contacts at S_1 and S_2 . If we take the conductivity to be the equilibrium (i.e., without excess carriers) conductivity of the QNR, the resistance will have a physical interpretation, at least in a hypothetical measurement. In this hypothetical measurement, we imagine that it is somehow possible to replace the DRB with an ideal ohmic contact at S_1 , so the electrical resistance of the QNR between S_1 and S_2 can be measured with the carrier-generation source turned off. When the source is turned off, a homogeneous QNR, surrounded by ideal contacts and reflective boundaries, will not contain

excess carriers, even with a voltage applied to the contacts so that the resistance can be measured. Therefore, the measured resistance will be the equilibrium resistance denoted R_{EQ} . To relate the surface integral in (47) to R_{EQ} , let v denote the voltage applied to S_1 (with S_2 grounded) during the resistance measurement. The electrostatic potential within the QNR during this measurement will satisfy Laplace's equation, and the solution to Laplace's equation that satisfies the same boundary conditions as this electrostatic potential is $v\Omega(\vec{x})$. Therefore, the electric field is $-v\vec{\nabla}\Omega(\vec{x})$. The equilibrium conductivity is $q\mu_M N$, which is the same as $qD_M N/V_T$, so the current density produced during this resistance measurement is given by $\vec{J}(\vec{x}) = -qD_M N v \vec{\nabla}\Omega(\vec{x})/V_T$. Therefore, the terminal current I produced during this resistance measurement is given by

$$I = -\int_{S_1} \vec{J} \circ d\vec{S} = v \frac{qD_M N}{V_T} \int_{S_1} \vec{\nabla}\Omega \circ d\vec{S}.$$

Combining this with $R_{EQ} = v/I$ gives

$$\int_{S_1} \vec{\nabla}\Omega \circ d\vec{S} = \frac{V_T}{qD_M N R_{EQ}}. \quad (48a)$$

Sometimes the capacitance between two geometric surfaces is easier to find in textbooks than the spreading resistance (e.g., the capacitance of an isolated conducting disc relative to a surface that is effectively infinitely far away can be found in any textbook on electromagnetic theory), in which case it may be more convenient to express the surface integral in terms of a capacitance. An alternative to (48a), which is equivalent to (48a), is

$$\int_{S_1} \vec{\nabla}\Omega \circ d\vec{S} = \frac{C}{\varepsilon} \quad (48b)$$

where C is the capacitance between S_1 and S_2 when the QNR is replaced by a dielectric with a permittivity constant ε . The value assigned to ε doesn't matter, because it divides out of the right side of (48b) as long as the ε in the denominator is the same ε used to calculate C , so we have the option of using the free-space permittivity constant for ε . For example, suppose the DRB resembles a flat circular disc of radius r , the electrode contact has an effectively infinite lateral extent, and the depth of the electrode relative to the DRB is much larger than r . The capacitance is the capacitance of an isolated flat conducting disc, except that we have to divide by 2 because we want the capacitance between the disc and the lower infinite hemisphere instead of the capacitance between the disc and the complete infinite sphere. The latter capacitance can be found in standard textbooks and is $8\varepsilon r$. Dividing by 2 gives $C/\varepsilon = 4r$ for this example.

Substituting (48) into (47) gives us two equivalent alternatives to (47), which are

$$\Delta U = (\alpha^* - \alpha) \frac{D_M}{D_m} qG R_{EQ} \quad (49a)$$

and

$$\Delta U = (\alpha^* - \alpha) \frac{\varepsilon}{C} \frac{G}{N D_m} V_T. \quad (49b)$$

In order to see the similarity between the 1D result (25) and the 3D result (49), we must remember that qG has the units of current density in the 1D problem, and it has the units of current in the 3D problem. The similarity becomes clear if we let A be the cross-sectional area of the 1D device and write (25) as

$$\Delta U = (\alpha^* - \alpha) \frac{L}{A} \frac{AG}{N D_m} V_T \quad (1D)$$

where AG in the 1D problem is the same as G in the 3D problem. With this interpretation of the symbolism, it is evident that the above equation is the 1D version of (49b).

As explained in the 1D analysis in Section III.E, the estimate (47), or either of the equivalent alternates in (49), is useful when there is an HRR (the condition that produces the largest voltage), in which case α^* calculated from (38) will be greater than 1, and we set α equal to 1.

E. Terminology Clarification

Before ending this section, some potential confusion regarding terminology should be clarified. The function $\Omega(\vec{x})$ was called a “charge-collection efficiency function” in some earlier papers, e.g., in [4] and [5]. This might be confusing because we are now calling α the charge-collection efficiency. To clarify this, we first consider the motive for the earlier terminology. The standard diffusion current corresponding to P^* defined by (45) is the left side of (46), because this contains the same diffusion coefficient that is in (45). We see from (46) that $\Omega(\vec{x})$ is a weight factor that reflects the effect that the location of ionization has on collected charge, when “collected charge” is governed by the standard diffusion current (a hypothetical case). It measures the relative importance of an increment of charge liberated at one location compared to the same amount of charge liberated at another location. To be more explicit, $\Omega(\vec{x})$ is the fraction of charge liberated at a point \vec{x} that is collected, i.e., it is the charge-collection efficiency (when the current is a standard diffusion current) for a point source located at \vec{x} . We now see two distinctions between α and $\Omega(\vec{x})$. The first distinction is that α refers to the actual current while $\Omega(\vec{x})$ refers to a standard diffusion current. The second distinction is that the charge-collection efficiency α is the fraction of the total liberated charge (integrated over the source distribution) that is collected. In contrast, the charge-collection efficiency *function* $\Omega(\vec{x})$ is associated with a point within the QNR. At a given point \vec{x} within the QNR, $\Omega(\vec{x})$ is the charge-collection efficiency (for the standard diffusion current) for a point source at that location.

V. COMPARISONS WITH COMPUTER SIMULATION RESULTS

Because the analytical model presented here was derived from assumed equations obtained from simplifying physical approximations, there is a question as to whether the model has any ability to represent a more realistic case. As an attempt to answer this question, we compare model predictions to the predictions obtained from computer simulations that give a more realistic representation of a device. The simulations were performed by the PISCES code [6], which solves the drift-diffusion equations subject to bandgap narrowing, Shockley-Read-Hall and Auger recombination, and with mobilities depending on the doping density and on the electric field. These comparisons will reveal both strengths and weaknesses of the analytical model.

In all examples listed here, the diode is either n^+p or p^+n , with the heavily doped region being a thin layer with a peak doping density of $10^{20}/\text{cm}^3$, and the lightly doped region (which contains the QNR of interest) having a uniform doping density of $8 \times 10^{14}/\text{cm}^3$. The Shockley-Read-Hall lifetime depends on the doping density and is shorter in more heavily doped regions. The code allows the user to input a low-concentration lifetime, and then the code modifies this lifetime to account for doping density. The input lifetime was arbitrarily chosen to be $1 \mu\text{s}$. When modified for the $8 \times 10^{14}/\text{cm}^3$ QNR doping, the lifetime used by the code became $0.984 \mu\text{s}$, which is very close to the input value. Mobilities are calculated by the code by starting with low-field mobilities, and then they are modified by the code to include the effect of the electric field. The simulation results discussed here were first presented in [3], which listed generic default values for the mobilities given by $\mu_e = 1000 \text{ cm}^2/\text{V}\cdot\text{sec}$ and $\mu_h = 500 \text{ cm}^2/\text{V}\cdot\text{sec}$. Using $D = V_T\mu$, the low-field diffusion coefficients become $D_e = 26 \text{ cm}^2/\text{sec}$ and $D_h = 13 \text{ cm}^2/\text{sec}$. These diffusion coefficients will be used as inputs when obtaining predictions from the analytical model that will be compared to simulation results.

A. First Example: A 1D n^+p Diode with Uniform g

The first example is a 1D diode with a spatially uniform carrier-generation rate function g . The demarcation between n-type and p-type will be called the metallurgical junction (MJ), and the DRB is slightly to the right (in the orientation shown in Fig. 1) of the MJ. The MJ is at a fixed location in the simulation (in contrast, the DRB location depends on bias voltage and carrier-generation strength). The device is described by Fig. 1 except that g is uniform between the MJ and the QNR contact, so carrier liberation is occurring not only within the QNR but also within the DR. The distance between the MJ and the QNR contact is $5 \mu\text{m}$, and $g = 1.25 \times 10^{25}/\text{cm}^3\cdot\text{sec}$ at all points between the MJ and the QNR contact. With G defined by (7), except that the integration limits are changed so the integration is from the MJ to the QNR contact, in order to include the DR, we find that this g produces a total charge liberation rate qG , which includes the contribution from the DR, of $1000 \text{ A}/\text{cm}^2$. The simulation found that this charge is collected with a 100-percent efficiency, i.e., the device current is $1000 \text{ A}/\text{cm}^2$, even when the biasing voltage at the device terminals is 0.4 V with a forward-biasing polarity. Any voltage greater than -0.4 V (the negative sign indicates a forward-bias polarity) produces a 100-percent charge-collection efficiency. The simulation also showed that when the power supply voltage is $+1 \text{ V}$ (a reverse-biasing polarity), the DR is still forward-biased. The QNR voltage is

1.62 V, while the power supply only provides 1 V, implying 0.62 V of forward-biasing across the DR⁹ (incidentally, the DRB is about 0.4 μm from the MJ in this example). In spite of this forward-biasing of the DR, the charge-collection efficiency is still 100 percent.

A weakness of the analytical model is that it reaches no conclusions when the DR is forward-biased (the derivation of this model was based on a reversed-biased DR), and we see from the simulation result that such cases can sometimes be encountered even when the power supply has a reverse-biasing polarity. Substituting numerical inputs from this example into (25), with α^* calculated from (16) and α calculated from (19), we estimate that the power supply voltage must be at least 1.95 V to reverse-bias the DR, and the analytical model will not reach any conclusions unless this biasing voltage is provided. However, the model does at least reach the correct conclusion under the reverse-biasing condition for which it was derived. When g is uniform, (16) gives

$$\alpha^* = \frac{1}{2} \left(1 + \frac{D_m}{D_M} \right) \quad (\text{uniform } g). \quad (50)$$

For a p-type QNR, we have $D_m/D_M = D_e/D_h$, which makes α^* greater than 1, which implies that $\alpha = 1$, so all charge liberated within the QNR is collected. All charge liberated within the DR is also collected, so the model correctly predicts a 100-percent collection efficiency under reverse-biasing conditions. Additional work is needed to generalize the analytical model so that it is able to reach conclusions under more general biasing conditions.

The derivation of the analytical model is limited to high-injection conditions (a.k.a., the large-G limit) and the generation rate used in the simulation was chosen to be large enough to produce such conditions. Another simulation considered the same case just described except that g was reduced by two orders of magnitude. Results from that simulation found that the device current is not adequately represented by approximations derived in the large-G limit. Instead, a better approximation is obtained from the opposite limit of low-injection conditions treated in elementary textbooks. The transition from low- to high-injection occurs when qG is varied between 10 and 1000 A/cm^2 , with the smaller value producing low-injection conditions and the larger value producing high-injection conditions. It is interesting that this assertion, which was originally discovered by performing simulations, can also be predicted from Fig. 2(a). This figure applies to a specific ratio D_m/D_M , but other figures in [1] indicate a universal (i.e., independent of D_m/D_M) trend; that the $qG = 100 qD_mN/L$ curve is well approximated by the large-G limit, while the $qG = 1 qD_mN/L$ curve is fairly well approximated by the small-G limit. Using numerical inputs from the example considered here, we find that the $qG = 10 \text{ A}/\text{cm}^2$ example is also a $qG \sim 1 qD_mN/L$ example, and the $qG = 1000 \text{ A}/\text{cm}^2$ example is also a $qG \sim 100 qD_mN/L$ example. We can therefore anticipate that the $qG = 10 \text{ A}/\text{cm}^2$ example can be

⁹ The DR has an equilibrium voltage (a.k.a., built-in voltage) of about 0.87 V for the doping levels in this example. About 0.62 V was removed from the DR and given to the QNR, where it was added to the 1 V from the power supply to produce 1.62 V across the QNR. The DR is now left with $0.87 - 0.62 = 0.25$ V. This voltage is 0.62 V less than the equilibrium voltage, so the DR is 0.62 V forward-biased. This example shows that the DR can be forward-biased even when the power supply has a reverse-biasing polarity, if the power supply voltage is not large enough to force a reverse-biasing condition.

approximated by the small-G limit, while the $qG = 1000 \text{ A/cm}^2$ example can be approximated by the large-G limit, which is consistent with simulation results.

B. Second Example: A 1D n^+ -p Diode with Localized Source near the QNR Center

The second example is just like the first except that carrier generation is changed from being spatially uniform to being a localized source midway between the MJ and the QNR contact. The total generation rate qG is still 1000 A/cm^2 , as in the first example. The simulation found the device current in this example to be virtually identical to the current in the first example for any biasing condition. The analytical model cannot predict this observation for arbitrary bias conditions, but it can predict this observation under the reverse-biasing condition for which it was derived. If the source was at the exact center of the QNR, α^* given by (16) would be the same for this source as it is for a uniform g , where we already concluded that $\alpha^* > 1$. The source is actually a little to the left of this midpoint (it is midway between the MJ and contact, with the MJ slightly to the left of the DRB), which increases α^* , so we still have $\alpha^* > 1$. The analytical model predicts a 100-percent collection efficiency for this example, the same as for the first example, so the analytical prediction is consistent with simulation results showing that the two cases are indistinguishable.

C. Third Example: A 1D n^+ -p Diode with Localized Source near the Contact

The third example is just like the first two except that carrier generation is now from a localized source that is $1 \mu\text{m}$ from the QNR contact. The total generation rate qG is still 1000 A/cm^2 , as in the first two examples. The simulation predicted the terminal current to be 740 A/cm^2 when the power supply voltage is 1 V with a reverse-biasing polarity. This is the first example in which the collection efficiency is less than 100 percent.

To obtain a prediction from the analytical model, we need to know the dimensions of the QNR and the location of the source relative to the DRB. The DRB location was not recorded when performing this simulation, but a rough estimate will do because the model prediction is not extremely sensitive to this estimate. A rough estimate of the DR width at a 1-V biasing voltage is $1 \mu\text{m}$. The length L of the QNR is then $4 \mu\text{m}$, and the distance X_S from the source to the DRB is $3 \mu\text{m}$. Substituting these numbers into (20) gives a model prediction of $\alpha = 0.75$. Note that this is less than 1, so we expect no HRR, and we therefore anticipate that the applied reverse-biasing voltage actually does reverse-bias the DR (the simulation confirms this). Therefore, the model prediction is expected to be applicable. The predicted current associated with this α is 750 A/cm^2 . This is very close to the simulation result of 740 A/cm^2 .

D. Fourth Example: A 1D p^+ -n Diode with Uniform g

This example is the same as the first example except that doping types are interchanged, so D_m/D_M is now $1/2$ instead of 2. The polarity of the power supply is reversed, so it still supplies 1 V with a reverse-biasing polarity. The simulation predicted the terminal current to be

791 A/cm². To obtain a prediction from the analytical model, we start with (50) to get $\alpha^* = 0.75$. Note that this is less than 1, so we expect no HRR, and we therefore anticipate that the applied reverse-biasing voltage actually does reverse-bias the DR (the simulation confirms this). Therefore, the model prediction is expected to be applicable. We therefore use $\alpha = \alpha^*$, so $\alpha = 0.75$. However, this applies only to the carriers liberated in the QNR. The carriers liberated in the DR are collected with a 100-percent efficiency. The simulation found the DR width to be 1.1 μm , producing a QNR length of 3.9 μm . Applying a 75-percent collection efficiency to the carriers liberated in the 3.9- μm region, and a 100-percent collection efficiency to the carriers liberated in the 1.1- μm region, produces a net collection of 80.5 percent of the total generation rate. The current is predicted to be 805 A/cm², which is very close to the simulation result of 791 A/cm².

E. Fifth Example: A 3D n⁺-p Diode with a Custom g

The 3D device in this example was constructed by starting with a cylinder with a 50- μm radius, with a 50- μm height and containing a p-type material. A thin layer of n⁺-type material (with an electrical contact) that extends to a 5- μm radius was added at the top end of the cylinder, and the remainder of the cylinder top is reflective. The bottom end of the cylinder is the electrical contact to the p-type material, and the vertical wall of the cylinder is reflective. The cylinder dimensions were chosen to be large so that the geometry will mimic the isolated disc shown in Fig. 4. The spatial distribution of carrier generation was an arbitrary choice, selected for no particular reason, and is confined to a smaller cylinder having the same radius as the n⁺-layer, i.e., the same radius as the MJ. This smaller cylinder where carriers are generated is the “generation cylinder” in Fig. 4. In this example, $L = r_D = 5 \mu\text{m}$. Carrier generation is spatially uniform within the cylinder and is given by $g = 6.25 \times 10^{24}/\text{cm}^3\text{-sec}$. This number was selected so that the surface-average current density through the MJ (terminal current divided by MJ area) will be 1000 A/cm² when the charge-collection efficiency is 100 percent. This is the same current density produced in the previous 1D examples (when the collection efficiency is 100 percent) for which it was found that the carrier generation was strong enough to produce high-injection conditions.

The power supply voltage was 1 V with a reverse-biasing polarity. As with the first example, the simulation found that the DR is actually forward-biased, but the charge-collection efficiency was 100 percent in spite of this DR bias. The charge-collection efficiency will also be 100 percent under the reverse-biasing condition for which the analytical model was derived, so to show agreement with the analytical model we merely have to show that the model predicts a 100-percent collection efficiency. This is done by showing that α^* defined by (38) is greater than or equal to 1.

It will not be necessary to accurately calculate the integrals in (38) because a simple inequality will be enough to reach the conclusion that $\alpha^* > 1$. An example of a constant- Ω contour is shown in Fig. 4. These contours are ellipsoidal surfaces, and the one that intersects the generation cylinder at its lower end (i.e., at the point $x=0, y=r_D$ and $z=L$) will completely enclose the cylinder. All constant- Ω contours contained within this contour have larger values of Ω , so if (x, y, z) is any point in the interior of the generation cylinder we will have $\Omega(x, y, z) >$

$\Omega(0, r_D, L)$. But $L = r_D$ in this example so $\Omega(x, y, z) > \Omega(0, r_D, r_D)$. Using this inequality together with $D_m = 2D_M$ in (38) gives $\alpha^* > 3\Omega(0, r_D, r_D)$. Using (41) we find that $\Omega(0, r_D, r_D)$ satisfies

$$\sin^2\left(\frac{\pi}{2}\Omega(0, r_D, r_D)\right) + \tan^2\left(\frac{\pi}{2}\Omega(0, r_D, r_D)\right) = 1$$

which can be solved for $\Omega(0, r_D, r_D)$ to give $\Omega(0, r_D, r_D) = 0.424$. Using this with $\alpha^* > 3\Omega(0, r_D, r_D)$ gives $\alpha^* > 1.27$, so the analytical model agrees with the simulation prediction that the collection efficiency is 100 percent.

VI. REFERENCES

- [1] L.D. Edmonds, *Analytical Solutions and Approximations for the Equation $ydy/dx = (ay+b)h(x)$ with Applications to Drift-Diffusion*, JPL Pub 09-13, Jet Propulsion Laboratory, Pasadena, CA, April 2009.
- [2] L.D. Edmonds, “Charge collection from ion tracks in simple EPI diodes,” *IEEE Trans. Nucl. Sci.*, vol. 44, no. 3, pp. 1448–1463, June 1997.
- [3] L. Edmonds, *A Theoretical Analysis of Steady-State Photocurrents in Simple Silicon Diodes*, JPL Pub 95-10, Jet Propulsion Laboratory, Pasadena, CA, March 1995.
- [4] L.D. Edmonds, “A time-dependent charge-collection efficiency for diffusion,” *IEEE Trans. Nucl. Sci.*, vol. 48, no. 5, pp. 1609–1622, Oct. 2001.
- [5] L.D. Edmonds, “SEU cross sections derived from a diffusion analysis,” *IEEE Trans. Nucl. Sci.*, vol. 43, no. 6, pp. 3207–3217, Dec. 1996.
- [6] M.R. Pinto, C.S. Rafferty, H.R. Yeager, and R.W. Dutton, “PISCES-IIB, Supplementary Rep.,” Stanford University, Stanford, CA, 1985.
(The code was revised to include cylindrical coordinates and photogeneration.)

APPENDIX A: A MORE RIGOROUS DERIVATION OF THE ADC MODEL

The pair of equations in (5) is a set of simultaneous equations used to solve for both P and U . We can eliminate U and obtain an equation containing P alone by multiplying the first equation in (5) by $P+N$, multiply the second equation by P , and add the resulting equations. This gives

$$\left[P(x) + \frac{N}{2} \right] \frac{dP(x)}{dx} = \left[\frac{J_m(x)}{2qD_m} - \frac{J_M(x)}{2qD_M} \right] \left[P(x) + \frac{N}{2} \right] + \left[\frac{J_m(x)}{2qD_m} + \frac{J_M(x)}{2qD_M} \right] \frac{N}{2}. \quad (\text{A1})$$

Boundary conditions that are included from the start were explained in the main text and are

$$P(x_2) = 0, \quad J_M(x_1) = 0. \quad (\text{A2})$$

Note that the second condition listed in (A2) makes $J_M(x)$ a known function of x by integrating (6). Similarly, if $J_m(x_1)$ is a known quantity, then $J_m(x)$ becomes a known function of x . If $J_m(x_1)$ is not a known quantity, then $J_m(x)$ becomes a known function of x plus an unknown constant. A complete solution to (A1) and (A2) can be obtained using either of two approaches. One approach regards the boundary value $P(x_1)$ as given and solves for $J_m(x_1)$. The other approach regards $J_m(x_1)$ as given (i.e., $J_m(x)$ is a known function of x) and solves for $P(x_1)$. If the given information is a statement about $P(x_1)$ instead of $J_m(x_1)$, we can still follow the second approach that solves for $P(x_1)$ in terms of $J_m(x_1)$, because this relation can then be inverted to solve for $J_m(x_1)$ in terms of the given $P(x_1)$. We follow that approach here, i.e., we proceed with the analysis as if $J_m(x_1)$ were known, even though the ultimate goal is to solve for $J_m(x_1)$. As in the main text, we use $J_m(x_1) = \alpha qG$, with G defined by (7), so specifying (or solving for) $J_m(x_1)$ is equivalent to specifying (or solving for) the charge-collection efficiency α .

The same steps that produced (24) in the main text give

$$\frac{J_m(x)}{qD_m} + \frac{J_M(x)}{qD_M} = \frac{D_m - D_M}{D_m D_M} \int_{x_1}^x g(\xi) d\xi + \frac{\alpha G}{D_m}. \quad (\text{A3})$$

It is convenient to define $H(x)$ by

$$H(x) \equiv \left(1 + \frac{D_m}{D_M} \right) \frac{\int_{x_1}^x \frac{x - \xi}{x - x_1} g(\xi) d\xi}{\int_{x_1}^{x_2} g(\xi) d\xi}. \quad (\text{A4})$$

We define $H(x_1)$ by taking the limit. Also, by comparing (A4) to (16) we find an expression for $H(x_2)$. The results are

$$H(x_1) = 0, \quad H(x_2) = \alpha^*. \quad (\text{A5})$$

Another property of H can be seen by using (A4), together with the fact that g is nonnegative, to show that the derivative of H is nonnegative, so H is increasing. These steps also show that $H(x)$ is strictly increasing (the derivative is positive) at any point x in the open interval (x_1, x_2) , having the property that there exists an x' between x_1 and x such that $g(x') > 0$. It is also easy to show from (A4) that H satisfies

$$\int_{x_1}^x g(\xi) d\xi = \frac{D_M G}{D_m + D_M} \frac{d}{dx} [(x - x_1) H(x)]. \quad (\text{A6})$$

Combining (A6) with (24), (A3), and (A1) gives

$$\begin{aligned} \left[P(x) + \frac{N}{2} \right] \frac{dP(x)}{dx} = \frac{G}{2D_m} \left\{ \alpha - \frac{d}{dx} [(x - x_1) H(x)] \right\} \left[P(x) + \frac{N}{2} \right] \\ + \frac{G}{2D_m} \left\{ \alpha + \frac{D_m - D_M}{D_m + D_M} \frac{d}{dx} [(x - x_1) H(x)] \right\} \frac{N}{2}. \end{aligned} \quad (\text{A7})$$

We now explain the meaning of the ‘‘large- G limit.’’ The shape of the function $g(x)$ is held fixed. The function is made large by multiplying it by a large scale factor. This is done by selecting some normalized function, denoted $g_{norm}(x)$, that has the desired shape. The normalization condition is arbitrary but is held fixed, i.e., the function $g_{norm}(x)$ is the same function throughout the analysis. We then select a positive γ and let $g(x) = g_{norm}(x)/\gamma$. The ‘‘large- G limit’’ of any quantity is obtained by taking the limit as γ approaches zero on the positive side. Note that the function H given by (A4) does not depend on γ because the scale factors divide out on the right side of (A4). Also, the independent parameter α will be held fixed when taking this limit, so it will not depend on γ . Similarly, the doping density and diffusion coefficients will be held fixed when taking this limit. The G that appears in (A7) is defined by (7) in the main text, and it does depend on γ . It is given by $G = G_{norm}/\gamma$, where G_{norm} is the integral of $g_{norm}(x)$. Also, the solution P to (A7) will depend on γ . The notation should show this dependence because we want to investigate the limiting behavior of the solution, so the solution will be denoted $P_\gamma(x)$. With these replacements, (A7) becomes

$$\begin{aligned} \left[P_\gamma(x) + \frac{N}{2} \right] \frac{dP_\gamma(x)}{dx} = \frac{G_{norm}}{2\gamma D_m} \left\{ \alpha - \frac{d}{dx} [(x - x_1) H(x)] \right\} \left[P_\gamma(x) + \frac{N}{2} \right] \\ + \frac{G_{norm}}{2\gamma D_m} \left\{ \alpha + \frac{D_m - D_M}{D_m + D_M} \frac{d}{dx} [(x - x_1) H(x)] \right\} \frac{N}{2}. \end{aligned} \quad (\text{A8})$$

Information about the limiting behavior of P_γ can be obtained by investigating the limiting behavior of ρ_γ defined by

$$\rho_\gamma(x) \equiv \gamma P_\gamma(x). \quad (\text{A9})$$

Substituting (A9) into (A8) gives

$$\left[\rho_\gamma(x) + \frac{\gamma N}{2} \right] \frac{d\rho_\gamma(x)}{dx} = \frac{G_{norm}}{2D_m} \left\{ \alpha - \frac{d}{dx} [(x-x_1)H(x)] \right\} \left[\rho_\gamma(x) + \frac{\gamma N}{2} \right] + \frac{G_{norm}}{2D_m} \left\{ \alpha + \frac{D_m - D_M}{D_m + D_M} \frac{d}{dx} [(x-x_1)H(x)] \right\} \frac{\gamma N}{2}. \quad (\text{A10})$$

Let $\rho(x)$ denote $\rho_\gamma(x)$ when γ equals zero. It is tempting to solve for $\rho(x)$ by solving the equation that (A10) becomes when γ equals zero, which is

$$\rho(x) \frac{d\rho(x)}{dx} = \frac{G_{norm}}{2D_m} \left\{ \alpha - \frac{d}{dx} [(x-x_1)H(x)] \right\} \rho(x). \quad (\text{A11a})$$

The boundary condition is

$$\rho(x_2) = 0. \quad (\text{A11b})$$

The problem with this approach is that solutions to (A11) are not unique, particularly if the derivative $d\rho(x)/dx$ is allowed to be discontinuous.¹⁰ Because solutions to (A11) are not unique, we cannot simply define $\rho(x)$ to be “the” solution to (A11). Instead, we have to define $\rho(x)$ as the limit of $\rho_\gamma(x)$. However, it is evident from (A11) that any point x in the open interval (x_1, x_2) satisfying $\rho(x) > 0$ is also a point where¹¹

$$\frac{d\rho(x)}{dx} = \frac{G_{norm}}{2D_m} \left\{ \alpha - \frac{d}{dx} [(x-x_1)H(x)] \right\} \quad (\text{at any } x \in (x_1, x_2) \text{ at which } \rho(x) > 0). \quad (\text{A12})$$

This analysis does not solve carrier transport equations on the left side of the DRB in Fig. 1, 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 which 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 at the DRB increases with an increasing carrier-generation rate. We interpret this to mean that $\rho(x_1)$ is not zero. It is small compared to $\rho(x)$ at some interior points, and $\rho(x_1)$ will be approximated as being zero in some selected equations that will follow, but it is not exactly zero. This means that there is some interval (x_1, x_c) for which (A12) applies. Integrating (A12) from x_1 to any point x contained in this interval gives

¹⁰ The limit of a sequence of functions, which is the interpretation of $\rho(x)$, often has a discontinuous derivative. Such a discontinuity is allowed by (A11) if it occurs at a point x at which $\rho(x) = 0$. Therefore, we should be prepared for the possibility that the derivative could be discontinuous.

¹¹ Mathematicians will complain that in order to establish (A12) we must first answer some questions, such as whether the derivative of the limit is equal to the limit of the derivative, and the question as to whether each of these quantities even exists. A more rigorous analysis was worked out by this author to address these issues but is not included here. The conclusion from that analysis is that (A12) is correct.

$$\rho(x) = \rho(x_1) + \frac{G_{norm}}{2D_m} \{\alpha - H(x)\}(x - x_1)$$

(at any $x \in [x_1, x_2]$ at which the right side is positive). (A13)

This applies at any point x such that the right side is positive.

Recall that we can regard α as given (but it was shown in the main text that α must be less than or equal to 1) with $\rho(x_1)$ the quantity to be solved, or vice versa. Some information regarding $\rho(x_1)$ is available (explained later) so the goal here is to use (A13), together with $\alpha \leq 1$, to solve for α in terms of $\rho(x_1)$. To solve for α it is necessary to also utilize the fact that $\rho(x_2) = 0$. Different steps are used to utilize this fact for different cases. There are two such cases and they are discussed separately below.

A1. The First Case

The first case considered is that in which $g(x)$ and $\rho(x_1)$ have the property that $H(x)$ defined by (A4) satisfies

$$H(x_2) > 1 + \frac{2D_m}{G_{norm}} \frac{\rho(x_1)}{x_2 - x_1} \quad (\text{defines Case 1}). \quad (\text{A14})$$

Using the fact that $\alpha \leq 1$, it is easy to show that (A14) implies that the right side of (A13) is negative when evaluated at $x = x_2$. The right side of (A13) is a continuous function of x , and it is positive when x is sufficiently close to x_1 , so there is some point in the open interval (x_1, x_2) , call it x_C , at which the right side of (A13) is zero at $x = x_C$ and negative at larger values of x . In other words, x_C satisfies

$$\left. \begin{array}{l} x_1 < x_C < x_2, \\ \rho(x_1) + \frac{G_{norm}}{2D_m} \{\alpha - H(x_C)\}(x_C - x_1) = 0, \\ \rho(x_1) + \frac{G_{norm}}{2D_m} \{\alpha - H(x)\}(x - x_1) < 0 \\ \text{when } x_C < x \leq x_2 \end{array} \right\} (\text{defines } x_C \text{ for Case 1}). \quad (\text{A15})$$

Note that $\rho(x) = 0$ at any point x to the right of x_C , which can be seen by contradiction. If $\rho(x) > 0$ at such an x , (A13) would apply. But the right side of (A13) is negative at this x , which contradicts the assumption $\rho(x) > 0$. Recall that $\rho(x)$ is the limit of $\gamma P_\gamma(x)$ as γ approaches zero, so the conclusion is that $\gamma P_\gamma(x)$ approaches zero as γ approaches zero if x is any point between x_C and x_2 . This information alone does not tell us whether $P_\gamma(x)$ is increasing or decreasing with γ . It just tells us that the rate of increase (if there is an increase at all) is slower than the rate at which γ approaches zero. To make a connection with the boundary condition $P_\gamma(x_2) = 0$, we must

investigate the limiting behavior of $P_\gamma(x)$ with greater resolution. For this purpose, we go back to (A8) and rearrange terms to write the equation as

$$\left\{ \alpha + \frac{D_m - D_M}{D_m + D_M} \frac{d}{dx} [(x - x_1)H(x)] \right\} \frac{N}{2P_\gamma(x) + N} = \frac{2D_m}{G_{norm}} \frac{d[\gamma P_\gamma(x)]}{dx} - \alpha + \frac{d}{dx} [(x - x_1)H(x)]. \quad (\text{A16})$$

We now take the limit as γ approaches zero. With the understanding that x is between x_C and x_2 , we have already established that $\gamma P_\gamma(x)$ approaches zero for any point x in this interval, so the derivative $d[\gamma P_\gamma(x)]/dx$ approaches zero. The limit of (A16) becomes

$$\left\{ \alpha + \frac{D_m - D_M}{D_m + D_M} \frac{d}{dx} [(x - x_1)H(x)] \right\} \lim_{\gamma \rightarrow 0} \left[\frac{N}{2P_\gamma(x) + N} \right] = \frac{d}{dx} [(x - x_1)H(x)] - \alpha \quad (\text{when } x_C < x < x_2). \quad (\text{A17})$$

An inequality will be needed to continue the analysis of (A17). This inequality is derived from the third condition in (A15), which implies

$$\{\alpha - H(x)\}(x - x_1) < 0 \quad (\text{when } x_C < x < x_2)$$

which can be written as

$$\int_{x_1}^x \frac{d}{d\xi} [(\xi - x_1)H(\xi)] d\xi > \alpha (x - x_1) \quad (\text{when } x_C < x < x_2). \quad (\text{A18})$$

It is clear from (A6) that the derivative inside the integral is increasing in ξ at any point ξ between x_1 and x_2 , so

$$\int_{x_1}^x \frac{d}{d\xi} [(\xi - x_1)H(\xi)] d\xi \leq \int_{x_1}^x \frac{d}{dx} [(x - x_1)H(x)] d\xi = (x - x_1) \frac{d}{dx} [(x - x_1)H(x)]$$

and combining this with (A18) gives

$$\frac{d}{dx} [(x - x_1)H(x)] - \alpha > 0 \quad (\text{when } x_C < x < x_2). \quad (\text{A19})$$

We can now return to (A17). We have just established that the right side of (A17) is positive, and this implies that the denominator inside the limit on the left side does not increase without bound

as γ approaches zero. The denominator also does not approach zero, because it contains the constant N , so the limit of the ratio in (A17) is the ratio of limits, i.e.,

$$\lim_{\gamma \rightarrow 0} \left[\frac{N}{2P_\gamma(x) + N} \right] = \frac{N}{\lim_{\gamma \rightarrow 0} [2P_\gamma(x) + N]} = \frac{N}{2P_0(x) + N} \quad (\text{when } x_C < x < x_2)$$

where $P_0(x)$ is the limiting value of $P_\gamma(x)$. Substituting this result into (A17) gives

$$P_0(x) = \frac{\alpha - \frac{D_M}{D_m + D_M} \frac{d}{dx} [(x - x_1)H(x)]}{\frac{d}{dx} [(x - x_1)H(x)] - \alpha} N \quad (\text{when } x_C < x < x_2).$$

The same result can be expressed in alternate notation by using (7) and (A6) to get

$$P_0(x) = \frac{\alpha \int_{x_1}^{x_2} g(\xi) d\xi - \int_{x_1}^x g(\xi) d\xi}{\left(1 + \frac{D_m}{D_M}\right) \int_{x_1}^x g(\xi) d\xi - \alpha \int_{x_1}^{x_2} g(\xi) d\xi} N \quad (\text{when } x_C < x < x_2).$$

It is evident from this equation that in order to satisfy the boundary condition $P_0(x_2) = 0$ we must have $\alpha = 1$, so the above equation becomes

$$P_0(x) = \frac{D_M \int_x^{x_2} g(\xi) d\xi}{D_m \int_{x_1}^x g(\xi) d\xi - D_M \int_x^{x_2} g(\xi) d\xi} N \quad (\text{when } x_C < x < x_2) \quad (\text{A20})$$

and the important conclusion is

$$\alpha = 1 \quad (\text{under Case 1 conditions}). \quad (\text{A21})$$

Note that α was held fixed when taking the limit as γ approaches zero, but the value assigned to α in (A21) satisfies boundary conditions for this limiting case. Therefore, even though the notation does not emphasize this fact, α is what the charge-collection efficiency becomes in the limit as γ approaches zero, i.e., in the large-G limit.

Before deriving (A21), we knew that there exists an x_C satisfying (A15) but we did not know how to calculate this x_C because the α appearing in (A15) was unknown. Now that α is known, the same condition that defines x_C can also be used to calculate x_C , and that condition is (A15) with $\alpha = 1$, i.e.,

$$\left. \begin{aligned}
& x_1 < x_C < x_2, \\
& \rho(x_1) + \frac{G_{norm}}{2D_m} \{1 - H(x_C)\} (x_C - x_1) = 0, \\
& \rho(x_1) + \frac{G_{norm}}{2D_m} \{1 - H(x)\} (x - x_1) < 0 \\
& \text{when } x_C < x \leq x_2
\end{aligned} \right\} \begin{array}{l} \text{(defines } x_C, \text{ which exists)} \\ \text{(under Case 1 conditions.)} \end{array} \quad (A22)$$

Similarly, (A13) now becomes

$$\rho(x) = \rho(x_1) + \frac{G_{norm}}{2D_m} \{1 - H(x)\} (x - x_1) \quad (\text{for Case 1 when } x_1 \leq x \leq x_C). \quad (A23)$$

A2. The Second Case

The second case considered is that in which $g(x)$ and $\rho(x_1)$ have the property that $H(x)$ defined by (A4) satisfies

$$H(x_2) \leq 1 + \frac{2D_m}{G_{norm}} \frac{\rho(x_1)}{x_2 - x_1} \quad (\text{defines Case 2}). \quad (A24)$$

We can find the α in (A13) that produces the boundary condition $\rho(x_2) = 0$ by guessing and then verifying that the guess does produce that boundary condition, and also satisfies $\alpha \leq 1$. The guess is

$$\alpha = H(x_2) - \frac{2D_m}{G_{norm}} \frac{\rho(x_1)}{x_2 - x_1} \quad (\text{requires verification}).$$

It is evident from (A24) that this tentative α satisfies $\alpha \leq 1$, so we now investigate the boundary condition. With this tentative α substituted into the right side of (A13), the right side becomes

$$\rho(x_1) + \frac{G_{norm}}{2D_m} \{\alpha - H(x)\} (x - x_1) = \frac{x_2 - x}{x_2 - x_1} \rho(x_1) + \frac{G_{norm}}{2D_m} \{H(x_2) - H(x)\} (x - x_1).$$

Note that H is increasing, so the curly bracket on the right is nonnegative for any x between x_1 and x_2 . Also, the first term on the right is positive for any x between x_1 and x_2 . Therefore, the left side, which is the right side of (A13), is positive for any x between x_1 and x_2 . Therefore, when evaluated at this α , (A13) applies to any x between x_1 and x_2 , and it becomes

$$\rho(x) = \frac{x_2 - x}{x_2 - x_1} \rho(x_1) + \frac{G_{norm}}{2D_m} \{H(x_2) - H(x)\} (x - x_1) \quad \text{for any } x \in (x_1, x_2). \quad (A25)$$

Boundary values are implied by (A25) by taking limits, producing $\rho(x_2) = 0$, which verifies that the tentative α satisfies the boundary conditions. The conclusion is

$$\alpha = H(x_2) - \frac{2D_m}{G_{norm}} \frac{\rho(x_1)}{x_2 - x_1} \quad (\text{under Case 2 conditions}). \quad (\text{A26})$$

As previously explained in the discussion of Case 1, the value given to α by (A26) is what the charge-collection efficiency becomes in the limit as γ approaches zero, even though the notation does not emphasize that this value is a limit.

A3. Ideal Boundary Conditions

As previously stated, this analysis does not solve carrier transport equations on the left side of the DRB in Fig. 1, so boundary conditions at the DRB must be given in order to have a complete set of equations. Assumptions used here regarding boundary conditions are 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 than at other locations in the QNR interior. One implication is that if we use either (A23) or (A25) to solve for $\rho(x)$, and if errors are considered acceptable when they are a small fraction of the maximum (in x) of $\rho(x)$, then we obtain acceptable accuracy by neglecting $\rho(x_1)$ in these equations. Similarly, the observation that $\rho(x_1)$ is small is interpreted to mean that x_C calculated from (A22), or α calculated from (A26), can be approximated by neglecting $\rho(x_1)$ in these equations. In other words, it is assumed that each of these quantities can be approximated by the quantities produced by ideal boundary conditions, where “the quantities produced by ideal boundary conditions” are defined by the equations obtained by omitting $\rho(x_1)$. To be more explicit, the quantities produced by ideal boundary conditions are α , $\rho(x)$, x_C (when it exists), and $P_0(x)$ given by

If $H(x_2) > 1$ then:

- (a) $\alpha = 1$.
- (b) There is an x_C in the open interval (x_1, x_2) satisfying $H(x_C) = 1$ and also satisfying $H(x) > 1$ for any x satisfying $x_C < x < x_2$.
- (c) For any x between x_1 and x_C we have $\rho(x) = \frac{G_{norm}}{2D_m} \{1 - H(x)\} (x - x_1)$.
- (d) For any x between x_C and x_2 we have $P_0(x) = \frac{D_M \int_x^{x_2} g(\xi) d\xi}{D_m \int_{x_1}^x g(\xi) d\xi - D_M \int_x^{x_2} g(\xi) d\xi} N$.

If $H(x_2) \leq 1$ then:

- (e) $\alpha = H(x_2)$.

(f) Throughout the entire QNR we have $\rho(x) = \frac{G_{norm}}{2D_m} \{H(x_2) - H(x)\}(x - x_1)$.

We will now change terminology and symbolism to conform to the main text. Instead of x_C we will now write x_{ARB} , which defines the ARB. The region to the left is called the AR, and the region to the right is called the HRR. Using (A5), we can write $H(x_2)$ as α^* . As previously stated, α given by either item (a) or item (b) is the limit as γ approaches zero, which we will call the large-G limit. Also, instead of writing $P_0(x)$, we will call this $P(x)$ in the large-G limit. Finally, consider Item (c). Recall that $\rho(x)$ is the limit of $\rho_\gamma(x)$, with $\rho_\gamma(x)$ defined by (A9), so Item (c) is equivalent to the statement that the approximation

$$\gamma P_\gamma(x) \approx \frac{G_{norm}}{2D_m} \{1 - H(x)\}(x - x_1)$$

becomes exact as γ approaches zero. If we divide both sides by γ , the resulting approximation becomes exact as γ approaches zero if error is defined as a fractional or relative error instead of an absolute error. This approximation can be written as

$$P_\gamma(x) \approx \frac{G}{2D_m} \{1 - H(x)\}(x - x_1).$$

and we will say that $P(x)$ becomes equal to the right side in the large-G limit, with the understanding that it is relative error rather than absolute error that goes to zero in the limit. Similar considerations apply to Item (f). With this change in notation and terminology, solutions produced by ideal boundary conditions are now described by

If $\alpha^* > 1$ then in the large-G limit we have:

(a) $\alpha = 1$.

(b) There is an ARB in the QNR interior that separates the AR on the left from the HRR on the right. The location of the ARB is x_{ARB} satisfying $H(x_{ARB}) = 1$ and also satisfying $H(x) > 1$ for any x satisfying $x_{ARB} < x < x_2$.

(c) For any x within the AR we have $P(x) = \frac{G}{2D_m} \{1 - H(x)\}(x - x_1)$.

(d) For any x within the HRR we have $P(x) = \frac{D_M \int_x^{x_2} g(\xi) d\xi}{D_m \int_{x_1}^x g(\xi) d\xi - D_M \int_x^{x_2} g(\xi) d\xi} N$.

If $\alpha^* \leq 1$ then in the large-G limit we have:

(e) $\alpha = \alpha^*$.

(f) There is no ARB.

(g) Throughout the entire QNR we have $P(x) = \frac{G}{2D_m} \{\alpha^* - H(x)\}(x - x_1)$.

APPENDIX B: ABBREVIATIONS, ACRONYMS, AND NOMENCLATURE

A	cross-sectional area of the 1D device
ADC	ambipolar diffusion with a cutoff
AG	the 1D quantity that is the same as G in the 3D problem
AR	ambipolar region
ARB	AR boundary
C	capacitance between S_1 and S_2 when the QNR is replaced by a dielectric with a permittivity constant ε
D_e	electron diffusion coefficient
D_h	hole diffusion coefficient
D_m	minority-carrier diffusion coefficient
D_M	majority-carrier diffusion coefficient
DR	depletion region
DRB	depletion region boundary
E	electric field
g	carrier-generation rate function
$g(x)$	carrier-generation rate density
G	source strength
HRR	high-resistance region
$H(x)$	(Defined by Eq. (A4))
I_m	minority-carrier terminal current
I_M	majority-carrier terminal current
I_T	total terminal current
J_m	minority-carrier current density (plus or minus depending on doping type)
J_M	majority-carrier current density (plus or minus depending on doping type)
$J_e(x)$	electron current density
$J_h(x)$	hole current density

J_T	total current density
KT/q	thermal voltage (written here as V_T)
large-G limit	high-injection limit
L	length of the QNR
MJ	metallurgical junction
MOSFET	metal oxide semiconductor field-effect transistor
n_0	equilibrium electron density
$n(x)$	electron density
N	doping density
p_0	equilibrium hole density
P	excess carrier density
P^*	(defined by Eq. (45))
$p(x)$	hole density
q	elementary charge
qG	total rate of charge liberation defined by Eq. (7) in 1D or by Eq. (26) in 3D
QNR	quasi-neutral region
r_D	radius of circular disc (in example geometry IV.C)
S_1	surface S_1 , also referred to as the depletion region boundary (DRB)
S_2	surface S_2 , also referred to as the quasi-neutral region (QNR) contact
SV	sensitive volume
U	plus or minus (depending on doping type) of the electrostatic potential
V_T	thermal voltage (sometimes written as KT/q)
x_1 thru x_n	points on the x axis
x_{ARB}	boundary location between an ambipolar region (AR) and a high-resistance region (HRR) (also called the AR boundary, ARB).

x_c	point at which the right side of (A13) is zero and negative at larger values of x
X_s	source location
1D	one dimensional
3D	three dimensional
α	charge-collection efficiency
α^*	defined by Eq. (16) in 1D or by Eq. (38) in 3D
ε	permittivity constant
μ_e	electron mobility
μ_h	hole mobility
$\varphi(x)$	electrostatic potential
$\Omega(\vec{x})$	weight factor that reflects the effect that the location of ionization has on collected charge from pure diffusion and is defined by Eq. (34)