A device model for intermediate band semiconductors

Eduard Christian Dumitrescu

Thesis submitted to

University of Ottawa

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in Physics

Ottawa-Carleton Institute of Physics

Department of Physics

University of Ottawa

Abstract

Semiconductors with an additional intermediate band (IB) have the potential to greatly improve solar cell efficiency. Their theoretical efficiency limit is over 50% higher than that of standard semiconductor solar cells at full concentration. In practice however, their efficiencies are low compared to this detailed balance limit. Part of the reason is that it has not been possible to optimize IB device geometry because no device model has existed that could capture all the effects present in IB materials (e.g., charge transport inside the IB and self-consistent optics). In this thesis I introduce my new device model for intermediate band semiconductors called Simudo. The software uses the finite element method to solve the coupled Poisson/drift-diffusion (PDD) system of equations that describe the carrier dynamics inside semiconductor (IB or not) devices, along with optical propagation. I benchmark its accuracy on standard semiconductor problems against Synopsys Sentaurus, and I find that not only does it give valid results but in fact converges to the solution with a smaller number of mesh points by having quartic rather than merely quadratic solution convergence with respect to the number of mesh points. I also demonstrate Simudo's immediate usefulness by answering the question of whether IB mobility can compensate for mismatched optical absorption processes in different regions of the device. The device model work is preceded by three introductory chapters bringing the reader up to speed on semiconductor device physics and providing them with a primer on the finite element method. The coupled PDD equations are numerically challenging to solve, and the road to development of Simudo tried a number of formulations of the problem that were not successful. In the final chapter I discuss some of these formulations and why they did not succeed.

Acknowledgments

Thank you: Jacob Krich, Karin Hinzer, Henry Schriemer, Matthew Wilkins, Kirill Stepanov, Michael Do, Sakif Khan, Emily Zhang, Mehan Haidari, Daisy Xia, Peter Rose, Luc Robichaud, Kyle Robertson, Meghan Beattie, Sanmeet Chahal. You've helped me get where I am, become who I am, and I am thankful for it.

Dedications

I dedicate this thesis, *and more*, to my lovely gf♡▷◇, who has supported me through the painful process of thesis.

Copyright

Statement of original contributions

I attest that I am the main author of this monograph, and that I am the sole author of all chapters except 5. The specifics are shown in Table 1.

I am the main developer and contributor to Simudo, and I wrote 100% of the core Poisson/drift-diffusion solver. I thank Jacob Krich and Matthew Wilkins for valuable feedback and discussions as I struggled to get the model working well.

Table 1: Breakdown of contributions and novelty by chapter of this monograph.

Chapter	Published elsewhere?	Novelty	My contribution
Preface	no	n/a (introductory chapter)	I am the sole author for everything, including all figures and models. I have received editing help and counsel from Jacob Krich.
1 - Introduction	no	n/a (introductory chapter)	
2 - Semiconductor physics	no	n/a (introductory chapter)	
3 - A more formal look at the law of the junction	no	Formal justification for the law of the junction, with explicit validity bounds. Independent derivation from that in [3].	
4 - Finite element method	[2] (preprint of [1])	n/a (introductory chapter)	
5 - The device model paper	[1]	World's first intermediate band drift-diffusion device model with self-consistent optics.	I am the main author of the paper, of the device model itself, and of every figure except for 5.4.
6 - Record of failures	no	Explains the challenges we faced in implementing the drift-diffusion equations.	I am the sole author for everything, including all figures and models. I have received editing help and counsel from Jacob Krich.
7 - Conclusion	no	n/a (conclusion chapter)	

Contents

Table 2: Common symbols used in this thesis, adapted from [1].

Symbol	Definition
u_k	Carrier density in semiconductor band k
$\mathcal{E}_k$	Band edge energy of band k; central energy of IB k
$\mathcal{E}_t$	Energy level of SRH trap
s_k	Sign of carriers in band k (+ for VB, − for CB)
w_k	Quasi-Fermi level ("imref") of carriers in band k
$\vec{j}_k$	Current density in band k
μ_k	Mobility of carriers in band k
$\mathcal{D}_k$	Diffusion coefficient of carriers in band k
S_k	Surface recombination velocity of carriers in band k
N_k	Effective density of states for nondegenerate band k Integrated density of states for intermediate band k
N_A	Acceptor doping concentration
N_D	Donor doping concentration
f_k	Filling fraction $f_k = u_k/N_k$ of IB k
f_A	Filling fraction of acceptor dopant
f_D	Filling fraction of donor dopant
$f_{k,\pm}$	$(1 - f_k)$ for positive sign and f_k for negative sign
$f_{k,0}$	Charge neutral filling fraction of IB k
g_k	Net generation in band k due to all generation and recombination processes
T	Temperature
β	Equal to $1/(k_B T)$, β is commonly known as the thermal energy scale
k_B	Boltzmann constant
q	Elementary charge
$\phi, \vec{E}, \rho, \varepsilon$	Electrostatic potential, electric field, charge density, and permittivity
$\alpha_{fi,\lambda}$	Optical absorption coefficient from band i to f at vacuum wavelength λ
σ_{fi}^{opt}	Optical cross section from band i to f
$\Phi_{\lambda,\hat{s}}$	Photon spectral flux density at wavelength lambda in direction $\hat{s}$
$\Phi_{[\lambda_1,\lambda_2],\hat{s}}$	Photon flux in direction $\hat{s}$ from λ_1 to λ_2, i.e., $\int_{\lambda_1}^{\lambda_2} \Phi_{\lambda,\hat{s}}\, \mathrm{d}\lambda$
$\hat{n}$	Surface normal vector
$\hat{s}$	Direction of light propagation

Chapter 1

Introduction

1.1 Context

Free energy means you can do anything. Cheaper renewable energy means you can still do a lot of things. One of them is addressing the modern climate crisis. For example, given sufficiently cheap and abundant energy production, CO_2 can be directly captured from the atmosphere. One strong contender for renewable energy generation is solar power via photovoltaic (PV) technologies. Although the PV market is booming now, standard cells have efficiencies limited by a fundamental thermodynamic limit called the Shockley-Queisser (SQ) or "detailed balance" limit [4]. The SQ limit is very general and only makes a few assumptions:

1. There is no nonradiative recombination;
2. All generated carriers either get collected at the contacts or recombine radiatively (emitting a photon with energy E_g);
3. An absorbed photon with energy $h\nu > E_g$ produces one carrier with energy E_g; and
4. Infinite carrier mobility (i.e., the quasi-Fermi levels are flat).

The SQ limit restricts standard solar cell efficiency to 33.7% for normal illumination [4], and 40.7% under full concentration[1] with a back reflector[2] [4, 5]. It would be better to generate more electricity from the same area on the ground. Several methods exist to break the SQ limit (by breaking one of the assumptions of the SQ limit). One exists as a real product (multijunction solar cells) and the rest are research activities[6]. In this thesis, we focus on an approach called intermediate band solar cells (IBSC).

[1] Full concentration means sunlight is concentrated onto the solar cell to the maximum extent allowable by optics and conservation of étendue.

[2] Mirror at the back of the device.

IBSCs have the potential to overcome the SQ limit by breaking the third assumption: the material can absorb two photons with energy less than E_g (which were previously not absorbed at all) and produce a single mobile carrier with energy E_g. Adding an intermediate band (IB) between the conduction and valence bands was first suggested in [7]. The idea languished in relative obscurity for about three decades until renewed interest in silicon sensitization to deeper infrared appeared [8]. This led [9] to recalculate the SQ limit for IB devices, calculating the fundamental Luque-Martí thermodynamic limit for IBSC as 42% under normal illumination [10] and 63.7% under full concentration with back reflector [9]. Physical realizations of IB materials include highly mismatched alloys [11, 12], quantum dots [13, 14], and bulk standard semiconductors with deep-level impurities [15, 16]. Although the IB effect (two photon non-resonant absorption giving photocurrent) has been observed, device efficiencies are unfortunately still low [17, 18] relative to the theoretical detailed balance efficiencies. What gives?

Contrary to the detailed balance assumptions, the reality of current IBSC devices is less than ideal. The non-ideal features can be roughly categorized as either *material* or *device design* issues. IB materials are not abundant: there are no known naturally occurring intermediate band materials, so they are all engineered materials. For an IB material to be appropriate for use in an IBSC, both sub-bandgap optical absorption processes must be strong, the carrier lifetimes must be sufficiently long, carrier mobility must be sufficiently large, and the semiconductor bandgap and the energy level of the IB must have values that are well matched to the sunlight spectrum [17]. In quantum dot systems, absorption can be increased by increasing dot density [19, 20, 21, 22]. To the extent that carrier lifetimes are determined by material defects, they can be improved. But if the IB is the fundamental cause of the short carrier lifetime [23], those IB materials may not be capable of making good IBSCs. The junctions between IB materials and standard semiconductors also need to be engineered to be high quality, allowing the free flow of electrons and holes but no conduction from the IB itself [11].

Most of the IBSC requirements of a material are encoded in the IB figure of merit

$$\nu = \frac{k_B T}{q} \min_i(\mu_i \tau_i) \min_i(\alpha_i)^2 \tag{1.1}$$

which can quickly predict whether an IB material is promising or not [23, 24, 25] given the material's conduction and valence band mobilities μ_C and μ_V, carrier lifetimes τ_C and τ_V, sub-bandgap absorptivities α_{IV} and α_{CI}, thermal energy $k_B T$, and elementary charge q.[3]

[3]There are actually two figures of merit, depending on whether transport inside the IB material is drift- or diffusion-dominated (which is a question about device design and can be achieved with appropriate doping profiles). The figure stated in Eq. 1.1 is the diffusive figure of merit. The drift figure of merit ν^{drift} is very similar, the only difference being that it replaces the thermal energy $k_B T$ in Eq. 1.1 with the bandgap E_g.

However, material parameters are only half the problem. The other half of the problem is device design. The detailed balance model cannot capture the effects of device design parameters (e.g., layer thicknesses), so a device model is required instead.

Device models are important for designing and optimizing all optoelectronic devices [26]. It would be prohibitively expensive to perform such design optimization by manufacturing candidate devices and analyzing them. Device models allow this semiconductor research to be performed *in silicio* (figuratively, rather than literally). Existing device models allow treatment of only 0-, 1-, and 2-band materials, but do not implement the PDE's required for 3-band IB materials. Moreover, unlike standard semiconductor photovoltaics, IB materials require self-consistent optics since the light propagation depends on the sub-bandgap optical absorption which depends on the IB filling fraction, which then depends on carrier densities and currents everywhere else inside the device [27]. For example, we present an effect in IB materials which requires self-consistent optics to be modelled: the sub-bandgap optical absorption processes depend on the IB's filling fraction (e.g., the IB→CB absorption process cannot happen if the IB is empty), so the sub-bandgap optical absorptions must be matched to each other or else the VB→IB→CB process will be limited by the weakest of its sub-processes. That is, unless the IB mobility is high enough to equalize the filling fraction throughout a wide IB region – we show an example of this phenomenon in Figures 5.8–5.10, and we invite the reader to see a more in-depth treatment in [28]. The absence of good device modeling (Table 5.1) has hindered the development of IBSCs, as it has been hard to simulate experiments accurately and impossible to systematically choose layer thicknesses, which is essential to making high-efficiency devices.

To address this glaring gap in IBSC device modeling, we developed our own device model named Simudo, a steady state Poisson/drift-diffusion device model supporting both standard and intermediate band semiconductors, as well as self-consistent optics and arbitrary user-defined inter-band processes (allowing the user to implement their own e.g., Auger recombination process). Simudo is written in Python and uses the FEniCS finite element platform [29, 30, 31], and has been released as free and open source software available for all to use at github.com/simudo/simudo.

Chapters 2–4 offer the background required for Chapter 5 which is a reproduction of the device model paper introducing Simudo for the first time [1]. In that chapter, we show that Simudo compares favourably and even outperforms commercial software Synopsys Sentaurus (Figure 5.4) on a reference problem by having quartic rather than quadratic convergence with respect to the number of mesh points. We also demonstrate in Section 5.4.3 Simudo's ability to deal with IB devices by using it to answer a legitimate research question: can IB mobility can compensate for mismatched optical absorption processes in different spatial regions of the device? (The answer is "yes, and here's some plots to tell you what mobility is necessary". Further work on this question using Simudo was performed in [28].) Finally, Chapter 6 shows some of the device model designs

that acted as precursors to Simudo, along with explanations as to why they failed.

1.2 Organization

The first three chapters are intended to be pedagogical, and hopefully useful to students new to photovoltaics. As such, we proudly display intermediate steps in derivations, and we prioritize qualitative behaviour and pedagogical value over quantitative accuracy.

Chapter 2 is a gentle introduction to semiconductor physics, with emphasis on carrier transport and solar cells. The target audience is the average physics BSc who didn't necessarily take solid state physics.

Chapter 3 examines a common assumption introduced in the previous chapter called the *law of the junction*, and derives rigorous bounds for its validity using minimal assumptions.

Chapter 4 is a gentle introduction to the finite element method, with emphasis on nonlinear equations. The target audience is the average physics BSc who has never touched FEM and who vaguely remembers Newton's method from second year calculus.

Chapter 5 is the device model paper introducing Simudo for the first time [1].

Chapter 6 describes some of the many things that we tried and failed before arriving at the working drift-diffusion formulation described in section 5.3.3.

Chapter 2

Semiconductor physics

This chapter is the solar cells physics primer that the author wishes they had read at the start of their Masters degree. Hopefully you will find it useful too.

We start by considering carriers in uniform semiconductors, with (uniform) doping and electric fields. We then move to a homojunction and introduce the depletion approximation. We derive the $J(V)$ (current as a function of applied potential) curve. Finally we compare the model against a full drift-diffusion model.

2.1 Homogeneous slab of semiconductor

Consider a homogeneous slab of standard semiconductor at thermal equilibrium, with a conduction band and a valence band. In the non-degenerate case where the Fermi level w is away from the band edges $\mathcal{E}_C, \mathcal{E}_V$ (also known as the Boltzmann approximation) the carrier concentration u_k in the bands can be written as

$$u_C = N_C e^{-\beta(\mathcal{E}_C - w)} \tag{2.1}$$

$$u_V = N_V e^{-\beta(w - \mathcal{E}_V)} \tag{2.2}$$

where w is the Fermi level, N_k is the effective density of states, and $\beta = 1/k_B T$ is the reciprocal of the thermal energy.

Relaxing the thermal equilibrium condition yields an extremely useful relationship: suppose that the carriers in each semiconductor band are at equilibrium with themselves (within the same band, but not necessarily across bands). This assumption is reasonable because the picosecond timescale for intra-band thermalization[1] is much shorter than the microsecond timescale for inter-

[1]"Thermalization" is the process that makes carriers' energy distribution approach the Fermi-Dirac or Boltzmann

band thermalization (which is usually referred to as "recombination"). In this case, each band k has its own pseudo-Fermi level w_k called a "quasi-Fermi level". We can then write

$$u_C = N_C e^{-\beta(\mathcal{E}_C - w_C)} \tag{2.3a}$$

$$u_V = N_V e^{-\beta(w_V - \mathcal{E}_V)} \tag{2.3b}$$

or more generally

$$u_k = N_k e^{-s_k \beta(w_k - \mathcal{E}_k)} \tag{2.4}$$

where $s_V = 1$ for holes and $s_C = -1$ for electrons.

In the following subsections will first consider an undoped ("intrinsic") semiconductor, and then a doped ("extrinsic") semiconductor. We will then introduce electric potential into Eq. 2.3.

2.1.1 Intrinsic semiconductor

Assume thermal equilibrium, so that $w_C = w_V$. In an undoped (intrinsic) semiconductor, every electron in the conduction band must have come from the valence band (and therefore left a hole behind), so $u_C = u_V$. We can solve for the intrinsic Fermi level $w_i = w_C = w_V$ and carrier concentration $u_i = u_C = u_V$.

$$1 = \frac{u_C}{u_V} = \frac{N_C e^{-\beta(\mathcal{E}_C - w_i)}}{N_V e^{-\beta(w_i - \mathcal{E}_V)}} \tag{2.5}$$

$$1 = \frac{N_C}{N_V} e^{-\beta(\mathcal{E}_C + \mathcal{E}_V - 2w_i)} \tag{2.6}$$

$$\frac{N_V}{N_C} e^{\beta(\mathcal{E}_C + \mathcal{E}_V)} = e^{2w_i\beta} \tag{2.7}$$

$$\beta w_i = \frac{1}{2} \ln \frac{N_V}{N_C} + \beta \frac{\mathcal{E}_C + \mathcal{E}_V}{2} \tag{2.8}$$

The intrinsic carrier concentration u_i is then simply

$$u_i = \sqrt{u_i^2} \tag{2.9}$$

$$= \sqrt{u_C u_V} \tag{2.10}$$

$$= \sqrt{N_C N_V} e^{-\beta(\mathcal{E}_C - \mathcal{E}_V)/2} \tag{2.11}$$

distributions. Typically this process is just electron-electron scattering.

where $\mathcal{E}_C - \mathcal{E}_V$ is the bandgap of the semiconductor.

2.1.2 Extrinsic semiconductor

2.1.2.1 Mass action law

We now relax the intrinsic condition (no longer require that $u_C = u_V$). This means that there could be more of one type of carrier, and for now we will defer to the next section the answer to "where did the excess carrier come from?".

Using Eq. 2.8 and 2.11, we obtain

$$\ln \frac{u_C}{u_i} = \ln u_C - \ln u_i \tag{2.12}$$

$$= \ln \left(N_C e^{-\beta(\mathcal{E}_C - w_C)} \right) - \ln \left(\sqrt{N_C N_V} e^{-\beta(\mathcal{E}_C - \mathcal{E}_V)/2} \right) \tag{2.13}$$

$$= \left[\ln N_C - \beta(\mathcal{E}_C - w_C) \right] - \left[\frac{1}{2} \ln N_C + \frac{1}{2} \ln N_V - \beta \frac{\mathcal{E}_C - \mathcal{E}_V}{2} \right] \tag{2.14}$$

$$= \underbrace{\frac{1}{2} \ln \frac{N_C}{N_V} - \beta \frac{\mathcal{E}_C + \mathcal{E}_V}{2}}_{-\beta w_i} + \beta w_C \tag{2.15}$$

$$= \beta(w_C - w_i) \tag{2.16}$$

and so

$$u_C = u_i e^{\beta(w_C - w_i)} \tag{2.17}$$

More generally,

$$u_k = u_i e^{s_k \beta(w_i - w_k)} \tag{2.18}$$

We can also write

$$u_i e^{s_k \beta(w_i - w_k)} = u_k = N_k e^{-s_k \beta(w_k - \mathcal{E}_k)} \tag{2.19}$$

$$u_i e^{s_k \beta w_i} = N_k e^{s_k \beta \mathcal{E}_k} \tag{2.20}$$

If the semiconductor is at thermal equilibrium, we have

$$u_C u_V = N_C e^{-\beta(\mathcal{E}_C - \cancel{w})} N_V e^{-\beta(\cancel{w} - \mathcal{E}_V)} \tag{2.21}$$

$$u_C u_V = N_C N_V e^{-\beta(\mathcal{E}_C - \mathcal{E}_V)} \tag{2.22}$$

and so the product $u_C u_V$ is independent of the Fermi level w. We note that at $w = w_i$ (intrinsic case) we have $u_{C,\text{intrinsic}} = u_{V,\text{intrinsic}} = u_i$, and so

$$u_C u_V = u_i^2 \tag{2.23}$$

which is called the *mass action law*.

2.1.2.2 Doping

Suppose that the semiconductor has been shallowly doped with dopant concentrations N_A and N_D (acceptors and donors respectively). We assume that donor atoms are charge neutral when filled (and positive when empty), and acceptor atoms are charge negative when filled (and neutral when empty). Donors enter with an extra electron at zero temperature, and acceptors enter with no extra electrons at zero temperature – the naming might seem counterintuitive, but "donors" and "acceptors" are named after their behaviour *as the temperature is increased from absolute zero* (at zero temperature, they haven't donated or accepted anything yet). Let f_A and f_D denote the electron filling fractions of the acceptor and donor levels (i.e. given a dopant atom, what's the probability that it has an electron). Then

$$u_V + N_D = u_C + f_A N_A + f_D N_D \tag{2.24}$$

$$u_V + (1 - f_D) N_D = u_C + f_A N_A \tag{2.25}$$

Proof. One way to find Eq. 2.24 is to note that (simply by electron conservation) we must have that

$$u_V - (u_C + f_A N_A + f_D N_D) = C \tag{2.26}$$

for some constant C. Changing the temperature should not change the number of electrons in the semiconductor, so C should not change either. In order to find the value of the constant C, we consider the carrier concentrations at $T = 0$. At zero temperature the conduction and valence band carrier concentrations go to zero, the acceptors are empty, and the donors are full. Therefore

$$\underbrace{u_V}_{0} - (\underbrace{u_C}_{0} + \underbrace{f_A N_A + f_D N_D}_{N_D}) = C \tag{2.27}$$

$$\implies \quad -N_D = C \tag{2.28}$$

$$\implies \quad u_V - (u_C + f_A N_A + f_D N_D) = -N_D \tag{2.29}$$

and so we recover Eq. 2.24. □

Rearranging Eq. 2.24 we obtain

$$u_C - u_V = (1 - f_D)N_D - f_A N_A \tag{2.30}$$

We can rewrite the left-hand side using Eq. 2.18 as

$$u_C - u_V = u_i e^{\beta(w - w_i)} - u_i e^{\beta(w_i - w)} \tag{2.31}$$

$$u_C - u_V = 2u_i \sinh \beta(w - w_i) \tag{2.32}$$

We now focus on the right-hand side of Eq. 2.30. In standard groups[2] IV, III/V, and II/VI semiconductors, the equilibrium f_A and f_D are Fermi factors

$$f_{D,\text{electrons}} = \frac{1}{1 + \frac{1}{2}e^{-\beta(w - \mathcal{E}_D)}} \tag{2.33}$$

$$f_{A,\text{holes}} = \frac{1}{1 + \frac{1}{4}e^{-\beta(\mathcal{E}_A - w)}} \tag{2.34}$$

where the $\frac{1}{4}$ and $\frac{1}{2}$ factors in f_A, f_D are due to degeneracy[3] and where $\mathcal{E}_A$ and $\mathcal{E}_D$ are the energy levels of the acceptor and donor dopants respectively. In this case, electrons are two-fold degenerate due to spin (two electrons with opposite spin can have the same energy level), and holes are four-fold degenerate (spin, and light/heavy hole) [32, Eq. 19.31-19.32]. Substituting, we obtain that

$$(1 - f_D)N_D - f_A N_A = (1 - \frac{1}{1 + \frac{1}{2}e^{\beta(\mathcal{E}_D - w)}})N_D - \frac{1}{1 + \frac{1}{4}e^{\beta(\mathcal{E}_A - w)}}N_A \tag{2.35}$$

$$(1 - f_D)N_D - f_A N_A = \frac{N_D}{1 + 2e^{-\beta(\mathcal{E}_D - w)}} - \frac{N_A}{1 + \frac{1}{4}e^{-\beta(w - \mathcal{E}_A)}} \tag{2.36}$$

We note that if w is far away from the dopant levels (a stronger version of the non-degenerate assumption), then the exponential factors in the denominators are negligible, and combining Eq. 2.32

[2]columns of the Mendeleev table

[3]The degeneracy factors may be different for other semiconductors. Different degeneracy factors end up being equivalent to just shifting $\mathcal{E}_A$ or $\mathcal{E}_D$ slightly.

and 2.36 we have

$$2u_i \sinh \beta(w - w_i) = \frac{N_D}{1 + 2\underbrace{e^{-\beta(\mathcal{E}_D - w)}}_{\ll 1}} - \frac{N_A}{1 + \frac{1}{4}\underbrace{e^{-\beta(w - \mathcal{E}_A)}}_{\ll 1}} \tag{2.37}$$

$$2u_i \sinh \beta(w - w_i) \approx N_D - N_A \tag{2.38}$$

$$w \approx w_i + k_B T \operatorname{arcsinh} \frac{N_D - N_A}{2u_i} \tag{2.39}$$

This isn't an unreasonable assumption. Dopants—intentional ones at least—tend to be shallow, i.e. $\mathcal{E}_D$ is close to $\mathcal{E}_C$ and $\mathcal{E}_A$ is close to $\mathcal{E}_V$. Intermediate band and unintentional/unwanted dopants may be deep instead (deeper into the bandgap).

As long as the doping doesn't exceed u_i by *too many* orders of magnitude[4], our revised non-degenerate assumption still holds.

We can solve for u_C and u_V in terms of N_D and N_A using the mass-action law $u_C u_V = u_i^2$ from Eq. 2.23.

$$u_C - u_V = N_D - N_A \tag{2.40}$$

$$u_C - \frac{u_i^2}{u_C} = N_D - N_A \tag{2.41}$$

$$u_C^2 - \big[N_D - N_A\big] u_C - \big[u_i^2\big] 1 = 0 \tag{2.42}$$

$$u_C = \frac{1}{2}\big[N_D - N_A\big] + \frac{1}{2}\big[(N_D - N_A)^2 + 4u_i^2\big]^{1/2} \tag{2.43}$$

$$\text{and} \quad u_V = \frac{1}{2}\big[N_A - N_D\big] + \frac{1}{2}\big[(N_D - N_A)^2 + 4u_i^2\big]^{1/2} \tag{2.44}$$

where we picked the positive roots.[5]

We note that if $N_D \gg u_i$ and $N_D \gg N_A$ (but not sufficiently to break the non-degeneracy condition), then $u_C \approx N_D$, and similarly for $N_A \gg u_i$ and $N_A \gg N_D$, we have $u_V \approx N_A$.

2.1.3 Potential energy

Suppose that now we have an additional potential $U(\vec{x})$, such that each conduction band carrier gains energy $U(\vec{x})$ in moving from $\vec{x}$ to ∞ (or some reference point), and each valence band carrier loses energy $U(\vec{x})$ in doing the same. This locally modifies the density of states at position $\vec{x}$ by shifting $D'(E) = D(E + U)$, which shifts $\mathcal{E}'_C = \mathcal{E}_C + U$ and $w' = w + U$.

If the entire semiconductor crystal is exposed to a spatially constant potential U, then we should have the same resulting carrier densities u_C and u_V (and the same u_i) as if the potential

[4]It can actually exceed by many orders of magnitude! Just not **too many** orders of magnitude.

[5]Negative roots correspond to negative carrier densities, which are unphysical.

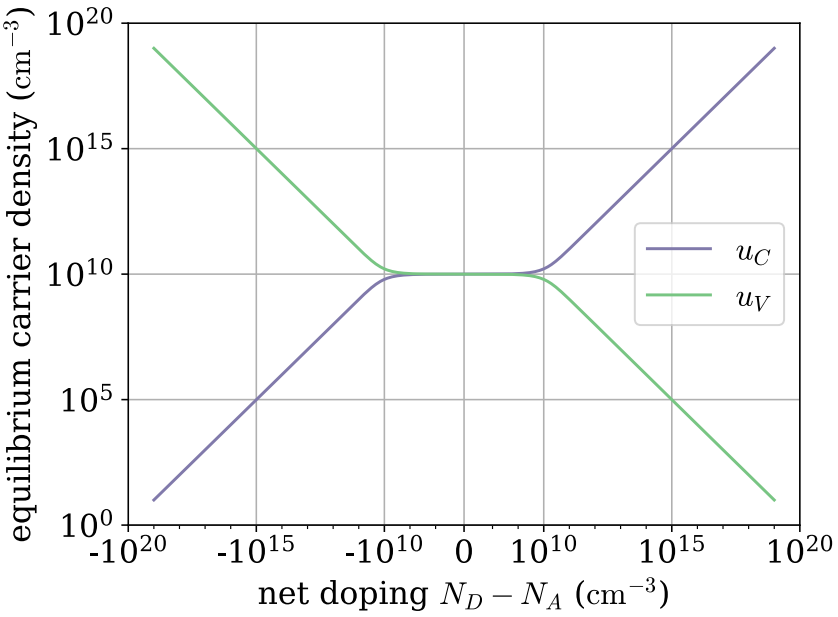

Figure 2.1: Silicon equilibrium carrier density in the conduction and valence bands as a function of the net doping $N_D - N_A$. Note that silicon has $u_i \approx 10^{10}\,\text{cm}^{-3}$ at $T = 300\,\text{K}$.

wasn't there. As a result, from Eq. 2.18 we must have that

$$\begin{aligned} u'_C &= u_C && (2.45)\\ &= u_i e^{\beta(w_C - w_i)} && (2.46)\\ &= u_i e^{\beta((w'_C - U) - w_i)} && (2.47) \end{aligned}$$

and

$$\begin{aligned} u'_V &= u_V && (2.48)\\ &= u_i e^{\beta(w_i - w_V)} && (2.49)\\ &= u_i e^{\beta(w_i - (w'_V - U))} && (2.50) \end{aligned}$$

We can expect the same equations to hold in a region where the potential is quasiconstant. Now let's take U to correspond to electric potential ϕ. Then $U = (-q)\phi$ (where q is the elementary charge) since U corresponds to the potential energy for a conduction band carrier (which has charge

$-q$), which results in

$$u_C = u_i e^{\beta(w_C + q\phi - w_i)} \tag{2.51a}$$

$$u_V = u_i e^{\beta(w_i - w_V - q\phi)} \tag{2.51b}$$

Using Eq. 2.20, this can also be written as

$$u_C = N_C e^{\beta(w_C + q\phi - \mathcal{E}_C)} \tag{2.52a}$$

$$u_V = N_V e^{\beta(\mathcal{E}_V - w_V - q\phi)} \tag{2.52b}$$

In general,

$$u_k = u_i e^{s_k \beta(w_i - w_k - q\phi)} \tag{2.53}$$

$$\text{or} \quad u_k = N_k e^{s_k \beta(\mathcal{E}_k - w_k - q\phi)} \tag{2.54}$$

Remark. It is useful to think about w_i (or $\mathcal{E}_k$) being modified by the addition of $-q\phi$. After all, w_i is defined as being the electrochemical potential w such that $u_C = u_V$, a definition which (under the replacement $w_i' = w_i - q\phi$) should remain unaffected by the addition of electric potential. This modification also preserves the requirement of w of being an energy quantity that is spatially constant at thermal equilibrium.

We note that the electrical potential ϕ itself is determined by u_C and u_V through Poisson's equation $\nabla \cdot (\varepsilon \nabla \phi) = -\rho$, where ε is the permittivity. Then

$$-\frac{\varepsilon}{q}\nabla^2 \phi = \frac{\rho}{q} \tag{2.55}$$

$$-\frac{\varepsilon}{q}\nabla^2 \phi = u_V - u_C + N_D - N_A \tag{2.56}$$

2.2 pn junction

An isolated homogeneous slab of semiconductor on its own isn't very useful (except maybe for producing materials science papers). Varying the material properties spatially (usually by selective doping) is what allows us to make useful devices like diodes and transistors and computer processors. Most of the interesting physics can be found in the simplest of these devices, the pn junction (also known as the simplest possible diode design).

We have p-doped semiconductor from $x_{pc} < 0$ to $x_J = 0$, and n-doped semiconductor from $x_J = 0$ to $x_{nc} > 0$. At x_{pc} and x_{nc} are located electrical contacts (whose exact behaviour we will hold off from describing for the moment). The dopant concentrations are N_A and N_D in each of

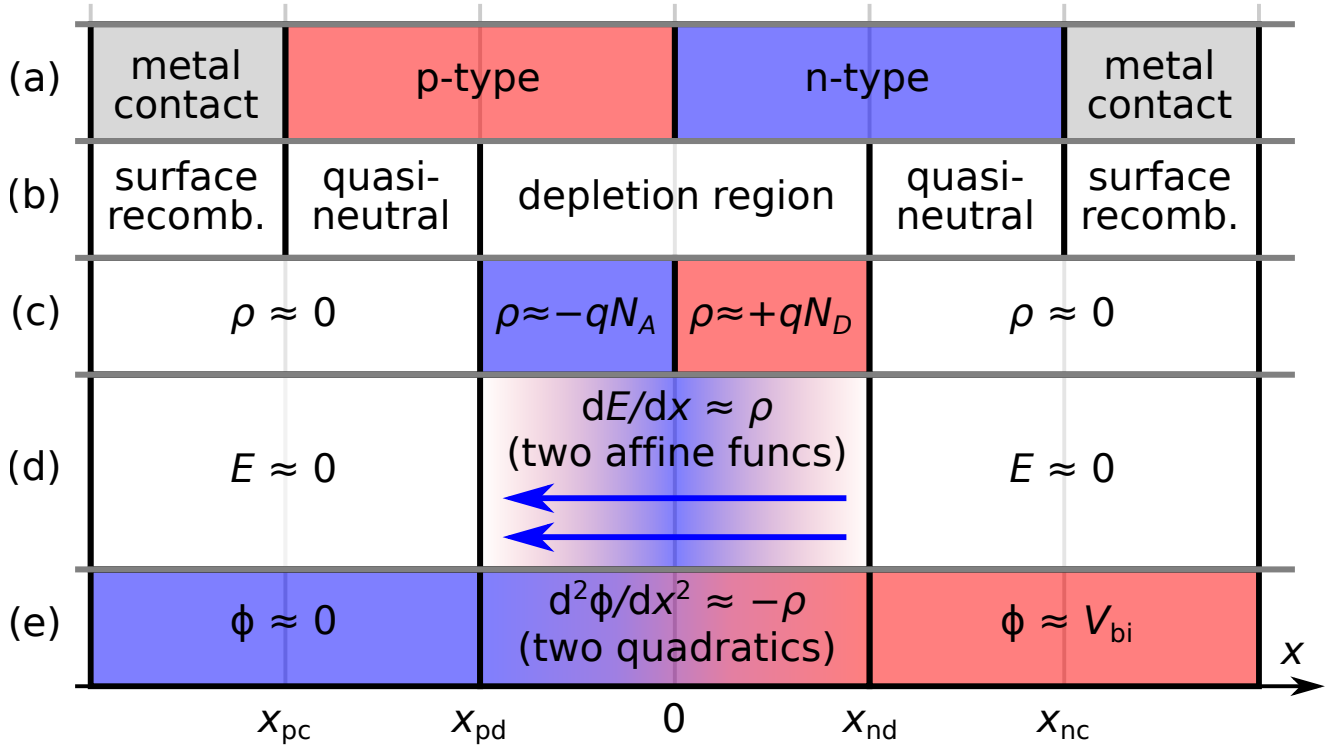

Figure 2.2: Diagram detailing the setup of this problem using the depletion approximation. We label (a) material type; (b) model used in each spatial region; (c) the electrostatic charge density, (d) electric field (given by Eq. 2.78), and (e) electrostatic potential (Eq. 2.90).

the two regions respectively. This geometry can be seen in Figure 2.2.

For now we are assuming equilibrium conditions (no external potential applied across contacts).

As we imagine gluing together the two halves of a diode, we see that there will be an electron flow from the n side to the p side since $w_{p\text{-side}} < w_{n\text{-side}}$ (the Fermi level for the p-side sticks close to the VB band edge, and the n-side Fermi level sticks close to the CB band edge). This electron flow will continue until the electric potential rises sufficiently on the n side to equalize the chemical potential.

We now derive the electric field and potential of a pn diode in the depletion approximation.

2.2.1 Depletion region

Consider $-x_{pc}$ and x_{nc} to be large and $N_A, N_D \gg u_i$. The chemical potential must be the same throughout the junction. The electrical potential must be a solution of

$$-\frac{\varepsilon}{q}\nabla^2\phi = u_V(x) - u_C(x) + \begin{cases} +N_D & x \geq x_J \\ -N_A & x \leq x_J \end{cases} \tag{2.57}$$

$$-\frac{\varepsilon}{q}\nabla^2\phi = u_i e^{\beta(w_i - w - q\phi)} - u_i e^{\beta(w + q\phi - w_i)} + \begin{cases} +N_D & x \geq x_J \\ -N_A & x \leq x_J \end{cases} \tag{2.58}$$

$$-\frac{\varepsilon}{q}\nabla^2\phi = 2u_i \sinh\beta\big(w_i - w - q\phi(x)\big) + \begin{cases} +N_D & x \geq x_J \\ -N_A & x \leq x_J \end{cases} \tag{2.59}$$

where we used Eq. 2.56 and 2.51a. At thermal equilibrium we can shift our definition of potential by a constant by letting $\phi' = \phi - \frac{1}{q}(w - w_i)$, and obtain

$$-\frac{\varepsilon}{q}\nabla^2\phi' = -2u_i \sinh\beta q\phi' + \begin{cases} +N_D & x \geq x_J \\ -N_A & x \leq x_J \end{cases} \tag{2.60}$$

Unfortunately, we are not aware of any analytic solution to this differential equation.

Assumption 1 (Depletion region approximation)**.** *On their own, the two semiconductor halves would have Fermi levels $w_V < w_C$. As a result, when the two halves of the junction are brought together, electrons flow from the n-type side to the p-type side. This results in a build-up of negative charge on the p-type side and positive charge on the n-type side.*

*The **depletion region approximation** states that in a region $x \in [x_{pd}, x_{nd}]$, there are almost no free carriers at all (and the dopant atoms are entirely ionized), and everywhere else ($x \in [x_{pc}, x_{pd}] \cup [x_{nd}, x_{nc}]$) we have charge neutrality and $\vec{E} \approx 0$. We logically define the depletion widths (always positive)*

$$W_p = x_J - x_{pd} = -x_{pd} \tag{2.61}$$

$$W_n = x_{nd} - x_J = x_{nd} \tag{2.62}$$

since $x_J = 0$.

We motivate the validity of the full ionization assumption underlying the depletion approximation by comparing the resulting solution against a real solution of a Eq. 2.60, as well as by noting that the carrier concentrations drop off exponentially inside the depletion region Figure 2.3.

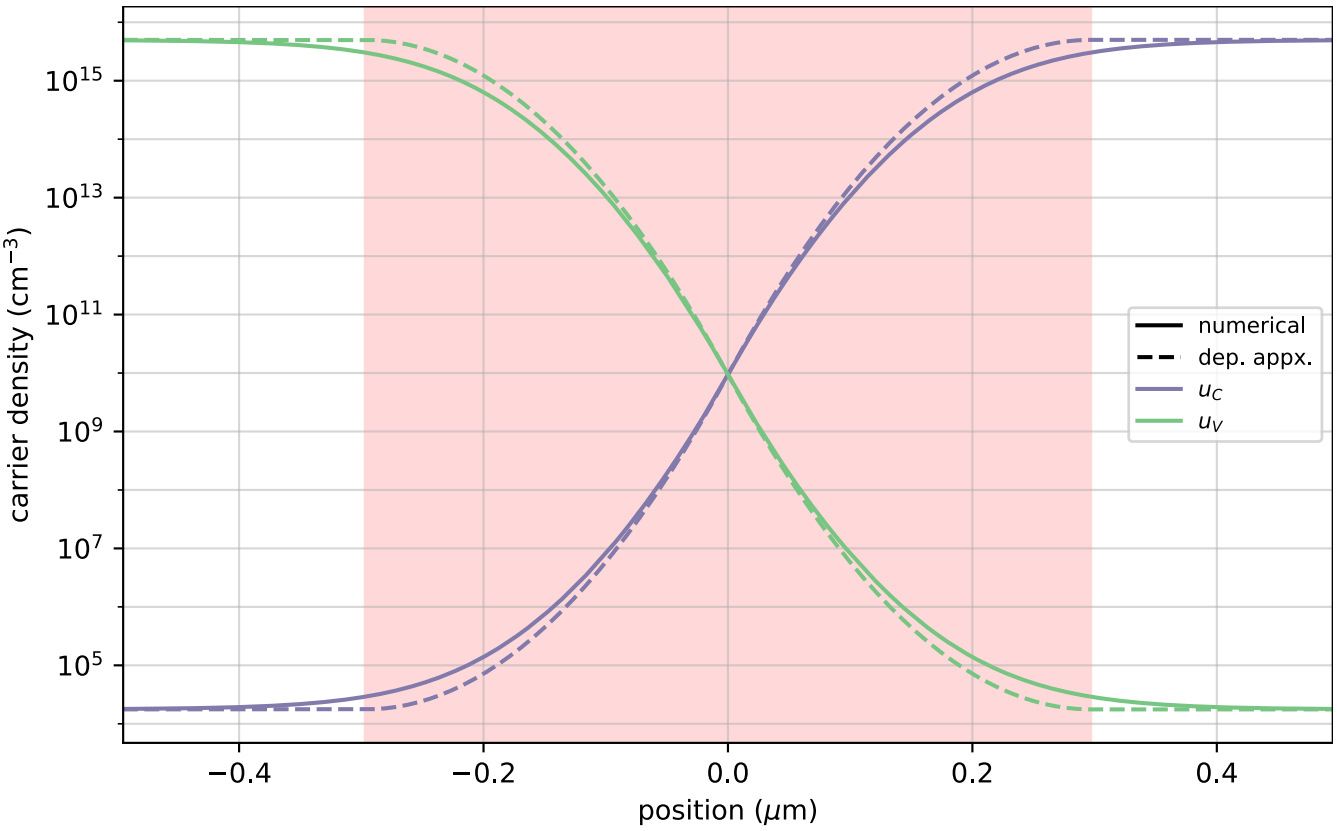

Figure 2.3: Carrier concentrations near the depletion region (shown in red) for a silicon junction with $N_D = N_A = 5 \times 10^{15}$ cm^{-3} at $T = 300$ K. The solid line is a true solution of Eq. 2.60 (evaluated numerically using Simudo) and the dashed line is the depletion approximation. Note that the true solution u_V falls off exponentially fast with x inside the depletion region, and falls to 10% of its p-type region equilibrium value $u_V^{(0)} \approx N_A$ within 0.10 μm of the depletion region edge x_{pd}, and falls to 1% of N_A within 0.17 μm of x_{nd}. For comparison, the width of the depletion region $W \approx 0.59$ μm.

At thermal equilibrium, w must be constant throughout the device. By invoking charge neutrality in the quasi-neutral regions, we have

$$N_A \approx u_V\Big|_{\text{p-type}} = u_i e^{\beta(w_i - w - q\phi_p)} \tag{2.63}$$

$$N_D \approx u_C\Big|_{\text{n-type}} = u_i e^{-\beta(w_i - w - q\phi_n)} \tag{2.64}$$

where ϕ_p and ϕ_n are electrostatic potentials in the p-type and n-type quasi-neutral regions respectively[6]. Then

$$N_A N_D = u_i e^{\beta(\cancel{w_i} - \cancel{w} - q\phi_p)} u_i e^{-\beta(\cancel{w_i} - \cancel{w} - q\phi_n)} \tag{2.65}$$

$$N_A N_D = u_i^2 e^{\beta q(\phi_n - \phi_p)} \tag{2.66}$$

$$\frac{1}{\beta q} \ln \frac{N_A N_D}{u_i^2} = \phi_n - \phi_p \equiv V_{\text{bi}}, \tag{2.67}$$

which is called the built-in potential.

Invoking the depletion region approximation, we then have a charge density

$$\rho(x) = q \begin{cases} 0 & x \in [x_{pc}, x_{pd}] \\ -N_A & x \in [x_{pd}, 0] \\ +N_D & x \in [0, x_{nd}] \\ 0 & x \in [x_{nd}, x_{nc}] \end{cases} \tag{2.68}$$

We set potential $\phi(x = -\infty) = 0$ (i.e. it is zero on the p-type side). These conditions can be visualized in Figure 2.2. To not have significant electric field outside the depletion region, it must be that the depletion region is overall neutral. Then

$$0 = \int \rho \, \mathrm{d}x \tag{2.69}$$

$$0 = (-q) N_A W_p + q N_D W_n \tag{2.70}$$

$$N_A W_p = N_D W_n \tag{2.71}$$

[6]Since the regions are quasi-neutral, the electrostatic potential must be (quasi-)constant inside each region, so ϕ_p and ϕ_n are (quasi-)well defined.

We apply Gauss' law[7] to get electric field within depletion region. Consider $x \in [x_{pd}, 0]$. Then

$$\int_{x_{pd}}^{x} \rho(x)\,\mathrm{d}x = Q = \int_{\partial} \varepsilon \vec{E} \cdot \mathrm{d}\vec{s} = \varepsilon(\vec{E}(x) - \overbrace{\vec{E}(x_{pd})}^{0}) \tag{2.72}$$

$$\implies \quad -qN_A(x - x_{pd}) = \varepsilon \vec{E}(x) \tag{2.73}$$

$$\implies \quad \vec{E}(x) = \frac{-qN_A}{\varepsilon}(x - x_{pd}) \tag{2.74}$$

Similarly, for $x \in [0, x_{nd}]$ we obtain

$$\vec{E}(x) = \frac{-qN_D}{\varepsilon}(x_{nd} - x) \tag{2.75}$$

In particular, we note that at $x = x_J = 0$ we have

$$\vec{E}(x_J) = \frac{-qN_A}{\varepsilon} W_p, \qquad \text{and} \tag{2.76}$$

$$\vec{E}(x_J) = \frac{-qN_D}{\varepsilon} W_n \tag{2.77}$$

which are equal to each other because of Eq. 2.71. This equality confirms that the electric field is continuous at the junction. To reiterate our results, we have used the depletion region approximation and obtained that

$$\vec{E}(x) = -\frac{q}{\varepsilon} \begin{cases} \approx 0 & x \in [x_{pc}, x_{pd}] \\ (x - x_{pd})N_A & x \in [x_{pd}, x_J] \\ (x_{nd} - x)N_D & x \in [x_J, x_{nd}] \\ \approx 0 & x \in [x_{nd}, x_{nc}] \end{cases} \tag{2.78}$$

[7] Gauss' law in 1D, also known as the fundamental theorem of calculus.

We now find the electric potential. We have $\phi(x \leq x_{pd}) = 0$. Now for $x \in [x_{pd}, x_J]$ we have

$$\phi(x) - \phi(x_{pd}) = -\int_{x_{pd}}^{x} \vec{E} \cdot \mathrm{d}\vec{x} \tag{2.79}$$

$$= -\int_{x_{pd}}^{x} -\frac{q}{\varepsilon}(\xi - x_{pd}) N_A \, \mathrm{d}\xi \tag{2.80}$$

$$= \frac{q}{\varepsilon} N_A \int_{x_{pd}}^{x} (\xi - x_{pd}) \, \mathrm{d}\xi \tag{2.81}$$

$$= \frac{q}{\varepsilon} N_A \left[\frac{(\xi - x_{pd})^2}{2} \right]_{\xi = x_{pd}}^{x} \tag{2.82}$$

$$= \frac{q}{\varepsilon} N_A \frac{1}{2} (x - x_{pd})^2 \tag{2.83}$$

For $x \in [x_J, x_{nd}]$, we have

$$\phi(x) - \phi(x_J) = -\int_{x_J}^{x} \vec{E} \cdot \mathrm{d}\vec{x} \tag{2.84}$$

$$= -\int_{x_J}^{x} -\frac{q}{\varepsilon}(x_{nd} - \xi) N_D \, \mathrm{d}\xi \tag{2.85}$$

$$= -\frac{q}{\varepsilon} N_D \int_{x_J}^{x} (\xi - x_{nd}) \, \mathrm{d}\xi \tag{2.86}$$

$$= -\frac{q}{\varepsilon} N_D \left[\frac{(\xi - x_{nd})^2}{2} \right]_{\xi = x_J}^{x} \tag{2.87}$$

$$= \frac{q}{\varepsilon} N_D \frac{1}{2} \Big((\underbrace{x_J - x_{nd}}_{W_n})^2 - (x - x_{nd})^2 \Big) \tag{2.88}$$

$$= \frac{q}{\varepsilon} N_D \frac{1}{2} \Big(W_n^2 - (x - x_{nd})^2 \Big) \tag{2.89}$$

Combining these expressions, we obtain

$$\phi(x) - \phi(x_{pd}) = \frac{q}{\varepsilon} \frac{1}{2} N_A (x - x_{pd})^2 \qquad x \in [x_{pd}, x_J] \tag{2.90a}$$

$$\phi(x) - \phi(x_J) = \frac{q}{\varepsilon} \frac{1}{2} N_D \Big[W_n^2 - (x - x_{nd})^2 \Big] \qquad x \in [x_J, x_{nd}] \tag{2.90b}$$

Then the built-in electrical potential V_{bi} across the depletion region is

$$V_{\text{bi}} = \phi(x_{nd}) - \phi(x_{pd}) \tag{2.91}$$

$$= \frac{q}{\varepsilon} N_D \frac{1}{2}\Big(W_n^2 - (x_{nd} - x_{nd})^2\Big) + \frac{q}{\varepsilon} N_A \frac{1}{2}(\underbrace{0 - x_{pd}}_{W_p})^2 \tag{2.92}$$

$$= \frac{q}{\varepsilon} N_D \frac{1}{2} W_n^2 + \frac{q}{\varepsilon} N_A \frac{1}{2} W_p^2 \tag{2.93}$$

$$= \frac{1}{2}\frac{q}{\varepsilon}(N_A W_p^2 + N_D W_n^2) \tag{2.94}$$

Invoking overall charge neutrality of the depletion region Eq. 2.71, $N_A W_p = N_D W_n$, we can solve for W_n and W_p in terms of V_{bi}. We have

$$W_p = \frac{N_D}{N_A} W_n \tag{2.95}$$

$$V_{\text{bi}} = \frac{1}{2}\frac{q}{\varepsilon}(N_A W_p^2 + N_D W_n^2) \tag{2.96}$$

$$V_{\text{bi}} = \frac{1}{2}\frac{q}{\varepsilon}\Big(N_A \frac{N_D^2}{N_A^2} W_n^2 + N_D W_n^2\Big) \tag{2.97}$$

$$V_{\text{bi}} = \frac{1}{2}\frac{q}{\varepsilon}\Big(\frac{N_D}{N_A} + 1\Big) N_D W_n^2 \tag{2.98}$$

$$W_n = \sqrt{\frac{2V_{\text{bi}}\varepsilon}{q\big(\frac{N_D}{N_A} + 1\big)N_D}} \tag{2.99}$$

$$\text{and} \quad W_p = \sqrt{\frac{2V_{\text{bi}}\varepsilon}{q\big(\frac{N_A}{N_D} + 1\big)N_A}} \tag{2.100}$$

We can solve for the width of the depletion region,

$$
\begin{aligned}
W_{\text{depletion}} &\equiv x_{nd} - x_{pd} \\
&= W_n + W_p && (2.101)\\
&= \sqrt{\frac{2V_{\text{bi}}\varepsilon N_A}{q\big(N_D + N_A\big)N_D}} + \sqrt{\frac{2V_{\text{bi}}\varepsilon N_D}{q\big(N_A + N_D\big)N_A}} && (2.102)\\
&= \sqrt{\frac{2V_{\text{bi}}\varepsilon}{q\big(N_D + N_A\big)}}\left(\sqrt{\frac{N_A}{N_D}} + \sqrt{\frac{N_D}{N_A}}\right) && (2.103)\\
&= \sqrt{\frac{2V_{\text{bi}}\varepsilon}{q\big(N_D + N_A\big)}}\left(\sqrt{\frac{N_A^2}{N_D N_A}} + \sqrt{\frac{N_D^2}{N_A N_D}}\right) && (2.104)\\
&= \sqrt{\frac{2V_{\text{bi}}\varepsilon}{q\big(N_D + N_A\big)}}\frac{N_A + N_D}{\sqrt{N_D N_A}} && (2.105)\\
&= \sqrt{\frac{2V_{\text{bi}}\varepsilon}{q}(N_A^{-1} + N_D^{-1})} && (2.106)
\end{aligned}
$$

Plugging in Eq. 2.67, we obtain

$$
\begin{aligned}
W_{\text{depletion}} &= \sqrt{\frac{2V_{\text{bi}}\varepsilon}{q}(N_A^{-1} + N_D^{-1})} && (2.107)\\
&= \sqrt{\frac{2\varepsilon}{\beta q^2}(N_A^{-1} + N_D^{-1})\ln\frac{N_A N_D}{u_i^2}} && (2.108)
\end{aligned}
$$

2.2.2 Depletion region under non-equilibrium conditions

Now consider that a bias is applied across the device. We continue to model this situation using the depletion approximation (the quasi-neutral regions remain quasi-neutral and the depletion region remains fully depleted), so the only thing that can change is the depletion width. Moreover, we continue to assert quasi-neutrality outside of the depletion region through the condition in Eq. 2.95 (for otherwise the quasi-neutral regions would no longer be so). Therefore 2.96 becomes

$$
\begin{aligned}
\Delta V &= \phi(W_n) - \phi(W_p) && (2.109)\\
&= \frac{1}{2}\frac{q}{\varepsilon}(N_A W_p^2 + N_D W_n^2) && (2.110)
\end{aligned}
$$

and so the depletion width is

$$W_{\text{depletion}} = W_n + W_p = \sqrt{\frac{2\Delta V \varepsilon}{q}(N_A^{-1} + N_D^{-1})} \tag{2.111}$$

We peek ahead to Eq. 2.146 to retrieve

$$\Delta V = V_{\text{bi}} - V_{\text{ext}} \tag{2.112}$$

which really just gives us the sign of V_{ext} that will make most sense in the future. Another way of justifying the minus sign in front of V_{ext} is that we expect to connect the positive lead of a multimeter or battery to the p-type side of the diode (instead of the rightmost contact which is on the n-type side), and that a positive V_{ext} should correspond to forward bias (which decreases the "diodeness" of the diode by lowering the potential barrier). Then

$$W_{\text{depletion}} = \sqrt{\frac{2(V_{\text{bi}} - V_{\text{ext}})\varepsilon}{q}(N_A^{-1} + N_D^{-1})} \tag{2.113}$$

Of course, this is invalid when $V_{\text{ext}} \geq V_{\text{bi}}$, i.e., when the external potential difference overcomes the built-in potential V_{bi}.

2.2.3 Drift-diffusion equations

Our ultimate goal is to use the depletion approximation to find the current flowing through the device as well as all other physical quantities (e.g., u_k, ϕ). We require a model to treat electric current.

The drift-diffusion model provides a set of equations that describe the movement of carriers within the semiconductor bands[32, Eq. 19.67-19.68]. This model is a very popular approximation based on the Boltzmann transport equation[33]. The Boltzmann transport equation is itself is an approximation to the Vlasov equation – there are *many* semiconductor transport models, as seen in Fig. 2.4. The model provides two equations for each band k,

$$\vec{j}_k = \overbrace{q\mu_k u_k \vec{E}}^{\text{drift}} - \overbrace{s_k q \mathcal{D}_k \vec{\nabla} u_k}^{\text{diffusion}} \qquad \text{drift-diffusion equation} \tag{2.114a}$$

$$\frac{\partial u_k}{\partial t} = -s_k \frac{1}{q} \vec{\nabla} \cdot \vec{j}_k + g_k \qquad \text{continuity equation} \tag{2.114b}$$

where $\vec{j}_k$ is the electric current flowing inside band k, μ_k is the carrier mobility, $\mathcal{D}_k$ is the carrier diffusivity, qs_k is the electric charge of a carrier in band k (so $s_C = -1$ and $s_V = +1$), and g_k is

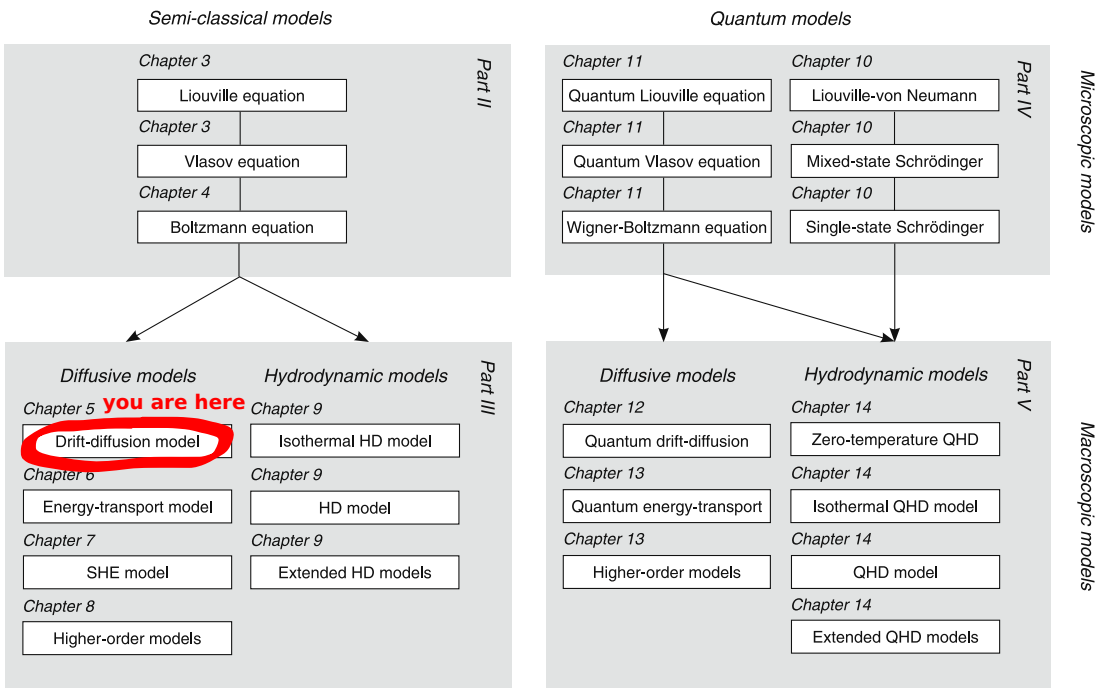

Figure 2.4: Diagram of the semiconductor transport models adapted from [33].

the net carrier generation (carriers received from other bands minus carriers lost to other bands – see Section 5.2.2).

For the conduction and valence bands in this section, we will use a simple relaxation model for minority carriers, in which

$$g_C = \begin{cases} -\frac{u_C - u_C^{(0)}}{\tau_C} & x \leq x_{pd} \quad \text{(minority carrier)} \\ 0 & \text{otherwise} \end{cases} \tag{2.115a}$$

$$g_V = \begin{cases} -\frac{u_V - u_V^{(0)}}{\tau_V} & x \geq x_{nd} \quad \text{(minority carrier)} \\ 0 & \text{otherwise} \end{cases} \tag{2.115b}$$

such that

$$\left.\frac{\partial u_C}{\partial t}\right|_{\text{minority}} = \frac{1}{q}\vec{\nabla}\cdot\vec{j}_C \underbrace{-\frac{u_C - u_C^{(0)}}{\tau_C}}_{g_C} \tag{2.116a}$$

$$\left.\frac{\partial u_V}{\partial t}\right|_{\text{minority}} = -\frac{1}{q}\vec{\nabla}\cdot\vec{j}_V \underbrace{-\frac{u_V - u_V^{(0)}}{\tau_V}}_{g_V} \tag{2.116b}$$

apply in the regions where the carrier is the minority carrier (e.g., p-type region for CB carriers). This simple relaxation model otherwise sets the generation to zero in the regions where the carrier is the majority carrier, as well as in the depletion region.

$$\vec{j}_C = q\mu_C u_C \vec{E} + q\mathcal{D}_C \vec{\nabla} u_C \tag{2.117a}$$

$$\vec{j}_V = q\mu_V u_V \vec{E} - q\mathcal{D}_V \vec{\nabla} u_V \tag{2.117b}$$

Remark 1. One would think that in Eq. 2.116 the minority $-\frac{u_C - u_C^{(0)}}{\tau_C}$ term in $\frac{\partial u_C}{\partial t}$ would also appear in the majority $\frac{\partial u_V}{\partial t}$ as $-\frac{u_C - u_C^{(0)}}{\tau_C}$ instead of being zero, since it takes two to recombine. In other words, these equations do not seem to preserve total charge. However, we will only use Eq. 2.116 to describe the flow of minority carriers (n carriers in p-type material in the stated example), where changes in the minority carrier are negligible compared to the majority carrier concentration. Accurate bimolecular recombination requires formulae that depend on both u_C and u_V. In some very important limits (e.g., looking at minority carrier under low injection), it reduces to a simple rate constant, as written above.

Lemma 1. *We have the following (Einstein) relations between mobility and diffusivity,*

$$\mu_C = \mathcal{D}_C \beta q \tag{2.118}$$

$$\mu_V = \mathcal{D}_V \beta q \tag{2.119}$$

for a non-degenerate semiconductor.

Proof. Suppose we have a uniformly doped semiconductor at equilibrium with an imposed electric field $\vec{E}$. Then $j_C = 0$, and so

$$\vec{0} = \vec{j}_C = q\mu_C u_C \vec{E} + q\mathcal{D}_C \vec{\nabla} u_C \tag{2.120}$$

$$\vec{0} = \cancel{q}\mu_C \cancel{u_i} e^{\beta(\cancel{\mu} + q\phi - \cancel{\psi_i})} \vec{E} + \cancel{q}\mathcal{D}_C \vec{\nabla} \cancel{u_i} e^{\beta(\cancel{\mu} + q\phi - \cancel{\psi_i})} \tag{2.121}$$

$$\vec{0} = \mu_C e^{\beta(q\phi)} \vec{E} + \mathcal{D}_C \vec{\nabla} e^{\beta(q\phi)} \tag{2.122}$$

$$\vec{0} = \mu_C \cancel{e^{\beta(q\phi)}} \vec{E} + \mathcal{D}_C \cancel{e^{\beta(q\phi)}} \beta q \underbrace{\vec{\nabla}\phi}_{-\vec{E}} \tag{2.123}$$

$$\vec{0} = \mu_C \vec{E} - \mathcal{D}_C \beta q \vec{E} \tag{2.124}$$

$$\mu_C = \mathcal{D}_C \beta q \tag{2.125}$$

Similarly, we obtain $\mu_V = \mathcal{D}_V \beta q$. □

The rest of this section will be spent solving equations Eq. 2.117 and 2.116 at steady state within the depletion approximation, with the additional assumptions that in the quasi-neutral regions,

the electric field is small so the drift currents $q\mu_k u_k \vec{E}$ of the minority carriers are negligible. Their transport is then purely diffusive, giving

$$\vec{j}_C\Big|_{\text{p-type}} = \underbrace{q\mu_C u_C \vec{E}}_{\approx 0} + q\mathcal{D}_C \frac{\partial u_C}{\partial x}, \qquad x < x_{pd} \tag{2.126}$$

$$\vec{j}_V\Big|_{\text{n-type}} = \underbrace{q\mu_V u_V \vec{E}}_{\approx 0} - q\mathcal{D}_V \frac{\partial u_V}{\partial x}, \qquad x > x_{nd} \tag{2.127}$$

There are called the minority carrier diffusion equations.

2.2.4 Law of the junction

The last missing piece of the puzzle is *some way* to relate the carrier concentrations on opposite sides of the depletion region. That relationship is the law of the junction, and in the following two sections we show two commonly used justifications for it. Chapter 3 shows a more formal derivation of the law of the junction, indicating clearly the requirements for its application.

2.2.4.1 With majority carrier quasi-Fermi level extension into depletion region

If the junction is not at equilibrium (e.g., there is an applied potential across it), we can still consider the quasi-neutral regions themselves to be (approximately) at thermal equilibrium far away from the junction, and therefore define quasi-Fermi levels for them, called w_C and w_V. We use the symbol w_k for the quasi-Fermi level rather than the traditional μ_k in order to avoid confusion with the mobility – using the same symbol for both would have made Eq. 5.5 very confusing.

Assumption 2. *We suppose that the quasi-Fermi levels for the majority carrier extend into the depletion region. As supporting evidence, see Figure 2.5. For a more formal treatment of this assumption, see [3]. To be more specific, we recall*

$$u_C = u_i e^{\beta(w_C(x) + q\phi(x) - w_i)} \tag{2.128}$$

$$u_V = u_i e^{\beta(w_i - w_V(x) - q\phi(x))} \tag{2.129}$$

and we assert that

$$w_C(x)\Big|_{x\in[x_{pd},x_{nc})} = w_C(x_{nc}) \tag{2.130}$$

$$w_V(x)\Big|_{x\in(x_{pc},x_{nd}]} = w_V(x_{pc}) \tag{2.131}$$

This can be visualized in table 2.1.

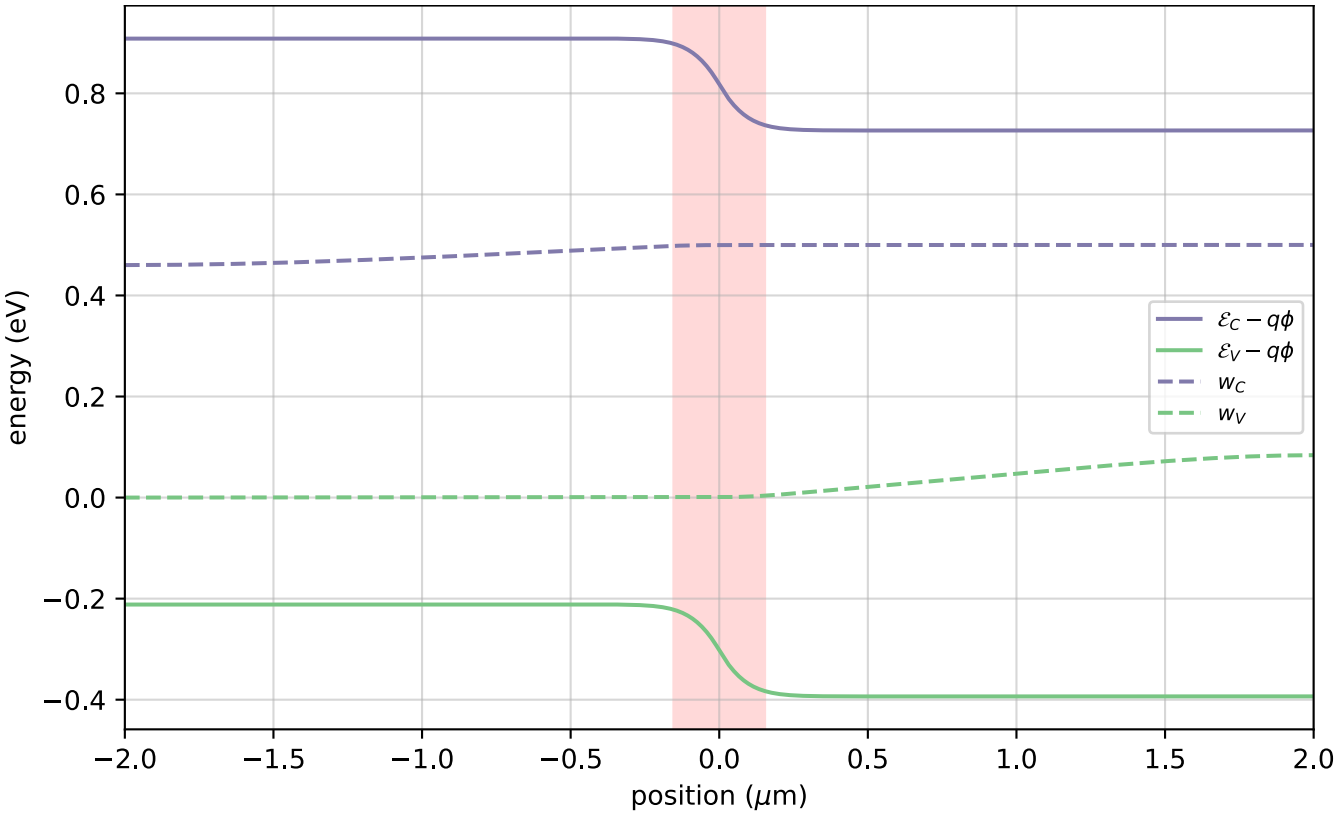

Figure 2.5: Quasi-Fermi levels w_k (shown in red) for a silicon junction with $N_D = N_A = 5 \times 10^{15}$ cm^{-3} with an applied bias of $V_{\text{ext}} = 0.5$ V, using a full drift-diffusion solver (Simudo). The parameters are fully described in the caption of Figure 2.8. The majority carrier quasi-Fermi levels extend throughout the depletion region and appear nearly flat.

Table 2.1: Summary of assumptions used in section 2.2.4.1.

$x \in$	$(x_{pc}, x_{pd}]$	$[x_{pd}, x_J]$	$[x_J, x_{nd}]$	$(x_{nd}, x_{nc}]$
Regime	quasi-neutral	depletion region		quasi-neutral
Doping	p-type N_A		n-type N_D	
w_C	TBD	constant by assu. 2		
w_V	constant by assu. 2			TBD
$\phi(x)$	constant	Eq. 2.90		constant
$\vec{E}(x)$	≈ 0	Eq. 2.78		≈ 0

Since the quasi-neutral regions are quasi-neutral, we also have

$$\phi(x_{nc}) = \phi(x)\Big|_{x\in[x_{nd},x_{nc})} \tag{2.132}$$

$$\phi(x_{pc}) = \phi(x)\Big|_{x\in(x_{pc},x_{pd}]} \tag{2.133}$$

Then

$$u_C(x_{pd}) = u_i e^{\beta(w_C(x_{pd})+q\phi(x_{pd})-w_i)} \tag{2.134}$$

$$= u_i e^{\beta(w_C(x_{nc})+q\phi(x_{pd})-w_i)} \tag{2.135}$$

$$= \underbrace{u_i e^{\beta(w_C(x_{nc})+q\phi(x_{nc})-w_i)}}_{=u_C(x_{nc})=u_C(x_{nd})\approx N_D} e^{\beta q(\phi(x_{pd})-\phi(x_{nc}))} \tag{2.136}$$

$$\approx N_D e^{\beta q(\phi(x_{pd})-\phi(x_{nc}))} \tag{2.137}$$

$$= N_D e^{-\beta q\Delta V} \tag{2.138}$$

and similarly

$$u_V(x_{nd}) = u_i e^{\beta(w_i-w_V(x_{nd})-q\phi(x_{nd}))} \tag{2.139}$$

$$= \underbrace{u_i e^{\beta(w_i-w_V(x_{pc})-q\phi(x_{pc}))}}_{=u_V(x_{pd})\approx N_A} e^{\beta q\overbrace{(\phi(x_{pc})-q\phi(x_{nd}))}^{-\Delta V}} \tag{2.140}$$

$$\approx N_A e^{-\beta q\Delta V} \tag{2.141}$$

Then at $x = x_{pd}$,

$$(N_D e^{-\beta q\Delta V})(N_A) = u_C(x)\overbrace{u_V(x)}^{\approx N_A} \tag{2.142}$$

$$= u_i e^{\beta(w_C(x)+q\phi(x)-w_i)} u_i e^{\beta(w_i-w_V(x)-q\phi(x))} \tag{2.143}$$

$$= u_i^2 e^{\beta(w_C(x)-w_V(x))} \tag{2.144}$$

so

$$\frac{N_A N_D}{u_i^2} e^{-\beta q\Delta V} = e^{\beta(\overbrace{w_C(x)-w_V(x)}^{\equiv qV_{\text{ext}}})} \tag{2.145}$$

$$\Delta V = \underbrace{\frac{1}{\beta q}\log\frac{N_A N_D}{u_i^2}}_{V_{\text{bi}}} - V_{\text{ext}} \tag{2.146}$$

From Eq. 2.146 we obtain

$$e^{-\beta q \Delta V} = \exp\left\{ -\beta q \left(\frac{1}{\beta q} \log \frac{N_A N_D}{u_i^2} - V_{\text{ext}} \right) \right\} \tag{2.147}$$

$$= \frac{u_i^2}{N_A N_D} e^{\beta q V_{\text{ext}}} \tag{2.148}$$

We combine this result with Eq. 2.138 to obtain

$$u_C(x_{pd}) \approx N_D e^{-\beta q \Delta V} \tag{2.149}$$

$$= N_D \frac{u_i^2}{N_A N_D} e^{\beta q V_{\text{ext}}} \tag{2.150}$$

$$= \frac{u_i^2}{N_A} e^{\beta q V_{\text{ext}}} \tag{2.151}$$

Similarly, from Eq. 2.141 we obtain

$$u_V(x_{nd}) \approx \frac{u_i^2}{N_D} e^{\beta q V_{\text{ext}}} \tag{2.152}$$

These two results constitute the law of the junction.

2.2.4.2 With informally justified drift-diffusion

We now rederive the law of the junction without the assumption that w_k is constant through the the depletion region. Instead, we assume that the current $\vec{j}_C$ is small in magnitude compared to the drift and diffusion terms taken separately (see Figure 2.6). Using the Einstein relation $\mu_C = \mathcal{D}_C \beta q$ (lemma 1), we write

$$0 \approx \vec{j}_C = q \underbrace{\mu_C}_{\mathcal{D}_C \beta q} u_C \vec{E} + q \mathcal{D}_C \vec{\nabla} u_C \tag{2.153}$$

$$0 = q \mathcal{D}_C \beta q u_C E + q \mathcal{D}_C \frac{\mathrm{d}u_C}{\mathrm{d}x} \tag{2.154}$$

$$-\beta q u_C E = \frac{\mathrm{d}u_C}{\mathrm{d}x} \tag{2.155}$$

$$-\beta q E \, \mathrm{d}x = \frac{\mathrm{d}u_C}{u_C} \tag{2.156}$$

$$\beta q (\phi - \phi_1) = \ln \frac{u_C}{u_{C,1}} \tag{2.157}$$

$$u_C = u_{C,1} e^{\beta q (\phi - \phi_1)} \tag{2.158}$$

where ϕ_1 and $u_{C,1}$ are the potential and CB carrier concentration at another arbitrary point x_1. In particular, the lack of net current between x and x_1 implies equal chemical potentials at the two points, since

$$u_C = u_{C,1} e^{\beta q(\phi - \phi_1)} \tag{2.159}$$

$$\cancel{u_i} e^{\beta(w_C(x) + q\cancel{\phi} - \cancel{w_i})} = \cancel{u_i} e^{\beta(w_C(x_1) + q\cancel{\phi_1} - \cancel{w_i})} e^{\beta q(\cancel{\phi} - \cancel{\phi_1})} \tag{2.160}$$

$$w_C(x) = w_C(x_1) \tag{2.161}$$

by recalling that (from Eq. 2.51a)

$$u_C = u_i e^{\beta(w_C(x) + q\phi(x) - w_i)} \tag{2.162}$$

$$u_V = u_i e^{\beta(w_i - w_V(x) - q\phi(x))} \tag{2.163}$$

We consider the points $x = x_{pd}$ and $x_1 = x_{nd}$, and so using Eq. 2.158 we obtain

$$u_C(x_{pd}) = u_C(x_{nd}) e^{\beta q \left[\phi(x_{pd}) - \phi(x_{nd})\right]} \tag{2.164}$$

$$= u_C(x_{nd}) e^{-\beta q \Delta V} \tag{2.165}$$

We now approximate $u_C(x_{nd}) \approx N_D$ (the majority carrier concentration is approximately the dopant concentration), and so

$$u_C(x_{pd}) = N_D e^{-\beta q \Delta V} \tag{2.166}$$

Then at the point $x = x_{pd}$ we have

$$(N_D e^{-\beta q \Delta V})(N_A) = u_C(x_{pd}) \overbrace{u_V(x_{pd})}^{\approx N_A} \tag{2.167}$$

$$= u_i e^{\beta(w_C(x_{pd}) + q\phi(x_{pd}) - w_i)} u_i e^{\beta(w_i - w_V(x_{pd}) - q\phi(x_{pd}))} \tag{2.168}$$

$$= u_i^2 e^{\beta(w_C(x_{pd}) - w_V(x_{pd}))} \tag{2.169}$$

We write $w_C^* \equiv w_C(x_{nd})$ and $w_V^* = w_V(x_{pd})$ for the majority carrier quasi-Fermi levels for brevity. We then have

$$e^{-\beta q \Delta V} = \frac{u_i^2}{N_A N_D} e^{\beta(w_C^* - w_V^*)} \tag{2.170}$$

$$\Delta V = \frac{1}{\beta q} \ln \frac{N_A N_D}{u_i^2} + \frac{1}{q}(w_V^* - w_C^*) \tag{2.171}$$

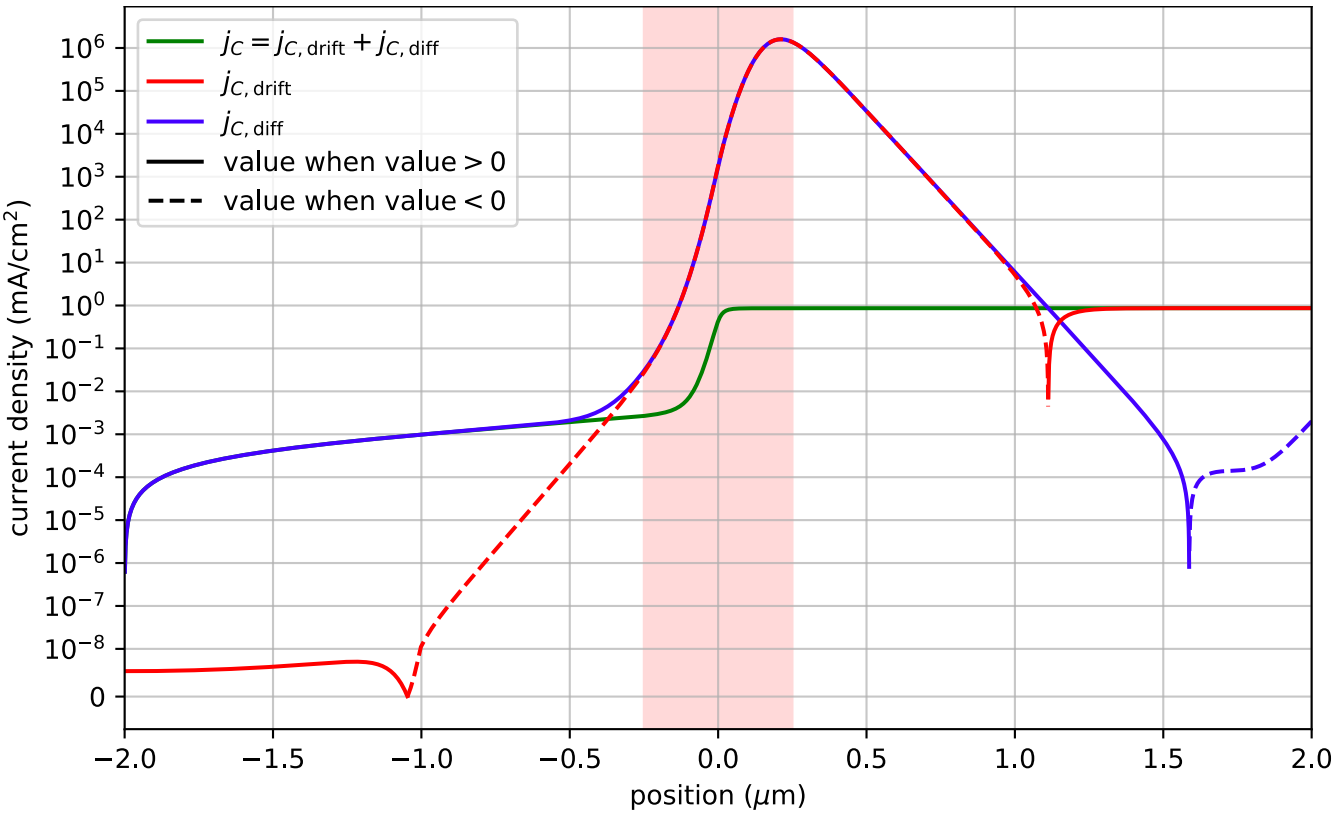

Figure 2.6: The drift and diffusion components of the current density $\vec{j}_C$ in a silicon pn junction at an applied bias of $V_{\text{ext}} = 0.2\,\text{V}$ (full parameters described in Figure 2.8). The two components oppose each other and very nearly cancel out in the depletion region and in part of the n-type quasi-neutral region nearby (where u_C is the majority carrier). In these regions, the total current $|j_C|$ is 1–6 orders of magnitude smaller than its individual components, justifying the starting assumption in Eq. 2.153. Diffusive transport dominates the p-type region (where u_C is the minority carrier).

Then $-qV_{\text{ext}} = w_V^* - w_C^*$. Note that V_{ext} is positive when p-type side has positive applied potential compared to n-type side, which matches our expected sign convention where V_{ext} is positive when a diode is in forward bias. Then

$$\Delta V = \frac{1}{\beta q} \ln \frac{N_A N_D}{u_i^2} + \frac{1}{q}(-q)V_{\text{ext}} \tag{2.172}$$

$$= \underbrace{\frac{1}{\beta q} \ln \frac{N_A N_D}{u_i^2}}_{V_{\text{bi}}} - V_{\text{ext}} \tag{2.173}$$

Then the (minority) CB carrier concentration at the p-type edge of the depletion region is

$$u_C(x_{pd}) = N_D e^{-\beta q \Delta V} \tag{2.174}$$

$$u_C(x_{pd}) = N_D \exp\left(- \beta q \left[\frac{1}{\beta q} \underbrace{\ln \frac{N_A N_D}{u_i^2}}_{V_{\text{bi}}} - V_{\text{ext}} \right] \right) \tag{2.175}$$

$$u_C(x_{pd}) = N_D \exp\left[- \ln \frac{N_A N_D}{u_i^2} + \beta q V_{\text{ext}} \right] \tag{2.176}$$

$$u_C(x_{pd}) = \frac{u_i^2}{N_A} e^{\beta q V_{\text{ext}}} \tag{2.177}$$

Similarly, we obtain

$$u_V(x_{nd}) = \frac{u_i^2}{N_D} e^{\beta q V_{\text{ext}}} \tag{2.178}$$

which is the law of the junction, again.

2.2.5 Total current in a device at steady state

The total current in a device at steady state must be divergence-free (i.e. $\vec{\nabla} \cdot \vec{j}_{\text{total}} = 0$). Recall the continuity equation Eq. 2.114b

$$\frac{\partial u_k}{\partial t} = -s_k \frac{1}{q} \vec{\nabla} \cdot \vec{j}_k + g_k \tag{2.179}$$

At steady state, $\frac{\partial u_k}{\partial t}$ is zero. By charge conservation, we must have that the net sum of generation processes must be zero, i.e.

$$\sum_k q s_k g_k = 0 \tag{2.180}$$

Then

$$0 = -s_k \frac{1}{q} \vec{\nabla} \cdot \vec{j}_k + g_k \tag{2.181}$$

$$\vec{\nabla} \cdot \vec{j}_k = q s_k g_k \tag{2.182}$$

$$\vec{\nabla} \cdot \vec{j}_{\text{total}} = \vec{\nabla} \cdot \sum_k \vec{j}_k = \sum_k \vec{\nabla} \cdot \vec{j}_k = \sum_k q s_k g_k = 0 \tag{2.183}$$

If the current in a device only flows along one axis given by unit vector $\hat{n}$ (which is true of a typical pn junction), then the total current is a constant since

$$\vec{\nabla} \cdot \vec{j}_{\text{total}} = 0 \tag{2.184}$$

$$\vec{\nabla} \cdot \hat{n} j_{\text{total}} = 0 \tag{2.185}$$

$$\hat{n} \cdot \vec{\nabla} j_{\text{total}} = 0 \tag{2.186}$$

$$\vec{\nabla} j_{\text{total}} = \vec{0} \tag{2.187}$$

2.2.6 Minority currents

We finally have all the parts we need. We now focus on finding the minority currents in the quasi-neutral regions by assuming diffusive transport. Once found, we can also find the total current without much work.

Note that we *cannot* re-use the j_C or j_V "found" using Eq. 2.117 in the depletion region, because we assumed them to be negligible in the depletion region in order to obtain the law of the junction boundary conditions.[32, §19.4.4]

Given the boundary conditions provided by the law of the junction (Eq. 2.151 and Eq. 2.152), we can solve for current. Under the assumption that drift is negligible in the quasi-neutral regions, we solve for current using Eq. 2.117 and Eq. 2.116. We shall perform the derivation for the minority carrier in the quasi-neutral n-type region $x > x_{nd}$.

Even after solving for the minority currents, we still have no good model for the majority currents. And we don't need one, because we can find the total current flowing through the device without ever calculating the majority currents. Since the pn junction is a 1d device and we are assuming steady state, the total current j must be spatially constant (Eq. 2.187). This is extremely useful: if we can find both the hole and electron currents at any point x, then we will have succeeded in finding total current everywhere throughout the device. We shall focus on finding both j_C and j_V at x_{nd}.

We substitute

$$0 = \frac{\partial u_V}{\partial t} = -\frac{1}{q}\vec{\nabla}\cdot\vec{j}_V - \frac{u_V - u_V^{(0)}}{\tau_V} \tag{2.188}$$

$$= -\frac{1}{q}\frac{\partial j_V}{\partial x} - \frac{u_V - u_V^{(0)}}{\tau_V} \tag{2.189}$$

$$= -\frac{1}{q}\frac{\partial}{\partial x}(\underbrace{q\mu_V u_V \vec{E}}_{\approx 0} - q\mathcal{D}_V\vec{\nabla}u_V) - \frac{u_V - u_V^{(0)}}{\tau_V} \tag{2.190}$$

$$= \mathcal{D}_V\frac{\partial^2 u_V}{\partial x^2} - \frac{u - u_V^{(0)}}{\tau_V} \tag{2.191}$$

Then

$$\underbrace{\mathcal{D}_V\tau_V}_{L_V^2}\frac{\partial^2 u_V}{\partial x^2} = u_V - u_V^{(0)} \tag{2.192}$$

This has solution $u_V = A\sinh\frac{x}{L_V} + B\cosh\frac{x}{L_V} + u_V^{(0)}$ where $L_V = \sqrt{\mathcal{D}_V\tau_V}$ is the diffusion length for the holes (minority carrier in the n-type region), since

$$\partial_x\cosh x = \partial_x\frac{e^x + e^{-x}}{2} = \frac{e^x - e^{-x}}{2} = \sinh x \tag{2.193}$$

and $\partial_x\sinh x = \cosh x$. We verify

$$\mathcal{D}_V\tau_V\frac{\partial^2 u_V}{\partial x^2} = u_V - u_V^{(0)} \tag{2.194}$$

$$\underbrace{\mathcal{D}_V\tau_V}_{L_V^2}\partial_x^2\left(A\sinh\frac{x}{L_V} + B\cosh\frac{x}{L_V} + u_V^{(0)}\right) = A\sinh\frac{x}{L_V} + B\cosh\frac{x}{L_V} + \cancel{u_V^{(0)}} - \cancel{u_V^{(0)}} \tag{2.195}$$

$$\cancel{L_V^2}\left(\frac{A}{\cancel{L_V^2}}\sinh\frac{x}{L_V} + \frac{B}{\cancel{L_V^2}}\cosh\frac{x}{L_V}\right) = A\sinh\frac{x}{L_V} + B\cosh\frac{x}{L_V} \tag{2.196}$$

which checks out.

We note that we could shift the solution ($x \mapsto x + x_0$) by any amount without changing the fact that it is a solution, since there is no explicit dependence on x in the differential equation Eq. 2.192. This is precisely what we will do for future convenience, so that we can apply the boundary conditions more directly; we shift this solution so that $\cosh x$ and $\sinh x$ receive a zero argument at the depletion region boundary x_{pd}; our solution then becomes

$$u_V = A_V\sinh\frac{x - x_{nd}}{L_V} + B_V\cosh\frac{x - x_{nd}}{L_V} + u_V^{(0)} \qquad x \geq x_{nd} \tag{2.197}$$

where we added subscripts on A and B so we know which minority carrier we're referring to.

We now use the boundary conditions given by the law of the junction (Eq. 2.151 and Eq. 2.152), i.e., the values of $u_C(x_{pd})$ and $u_V(x_{nd})$. Then

$$u_V(x_{nd}) = A_V \underbrace{\sinh \underbrace{\frac{x - x_{nd}}{L_V}}_{0}}_{0} + B_V \underbrace{\cosh \underbrace{\frac{x - x_{nd}}{L_V}}_{0}}_{1} + u_V^{(0)} \tag{2.198}$$

$$= B_V + u_V^{(0)} \tag{2.199}$$

which fixes $B_V = u_V(x_{nd}) - u_V^{(0)}$ while no condition is imposed on A_V (it will be fixed by the conditions at the contacts).

A similar derivation[8] follows for the n minority carrier in the quasineutral p region, resulting in

$$u_C = A_C \sinh \frac{-(x - x_{pd})}{L_C} + B_C \cosh \frac{-(x - x_{pd})}{L_C} + u_C^{(0)} \qquad x \leq x_{pd} \tag{2.200}$$

For convenience's sake, we have swapped the signs of the argument of $\sinh x$ and $\cosh x$ to make it positive; this is equivalent to multiplying A_C by -1 since $\sinh x$ is odd (the sign change does nothing to the $\cosh x$ since it's an even function). The main reason is that we expect the solution to be qualitatively symmetric on either side of the pn junction, so the A_C and A_V, and B_C and B_V coefficients would have the same sign (and in fact be equal to each other in the case perfectly symmetric case $L_C = L_V$ and $N_A = N_D$).

In a way similar to how we did it for the p minority carrier, we use the boundary condition $u_C(x_{pd})$ to obtain $B_C = u_C(x_{pd}) - u_C^{(0)}$.

Now that we have the minority carrier concentrations, to get the resulting current we simply plug Eq. 2.197 back into Eq. 2.117,

$$j_V\big|_{x \geq x_{nd}} = \underbrace{q\mu_V u_V \vec{E}}_{\approx 0} - q\mathcal{D}_V \vec{\nabla} u_V \tag{2.201}$$

$$= -q\mathcal{D}_V \frac{\partial u_V}{\partial x} \tag{2.202}$$

$$= -q\mathcal{D}_V \frac{\partial}{\partial x}\left[A_V \sinh \frac{x - x_{nd}}{L_V} + B_V \cosh \frac{x - x_{nd}}{L_V} + u_V^{(0)}\right] \tag{2.203}$$

$$= -q\frac{\mathcal{D}_V}{L_V}\left[A_V \cosh \frac{x - x_{nd}}{L_V} + B_V \sinh \frac{x - x_{nd}}{L_V}\right] \tag{2.204}$$

[8] Proof by analogy. Or more formally, *mutatis mutandis*.

and, by analogy,

$$j_C\big|_{x\le x_{pd}} = \underbrace{q\mu_C u_C \vec{E}}_{\approx 0} + q\mathcal{D}_C\vec{\nabla}u_C \tag{2.205}$$

$$= q\mathcal{D}_C\frac{\partial u_C}{\partial x} \tag{2.206}$$

$$= q\mathcal{D}_C\partial_x\left[A_C\sinh\frac{-(x-x_{pd})}{L_C} + B_C\cosh\frac{-(x-x_{pd})}{L_C} + u_C^{(0)}\right] \tag{2.207}$$

$$= -q\frac{\mathcal{D}_C}{L_C}\left[A_C\cosh\frac{-(x-x_{pd})}{L_C} + B_C\sinh\frac{-(x-x_{pd})}{L_C}\right] \tag{2.208}$$

All that's missing is A_V, and it is determined by the boundary condition at the contact between the n-type region and the (metallic) conductor. In general

$$qD_V\frac{\mathrm{d}u_V}{\mathrm{d}x} = -qS\Delta u_V \tag{2.209}$$

$$\frac{\mathrm{d}u_V}{\mathrm{d}x} = -\frac{S}{\mathcal{D}_V}\Delta u_V \tag{2.210}$$

at the contact, where S is the surface recombination velocity. For simplicity we (very optimistically) set the surface recombination velocity S to zero for the minority carrier, which then implies that $\frac{\mathrm{d}u_V}{\mathrm{d}x}\Big|_{x=x_{nc}} = 0$, i.e., $j_V(x_{nc}) = 0$. This assumption gives the boundary condition we've been yearning for. (For a much more general treatment where $S > 0$, see [34, section 4.4].) We then have

$$0 = j_V(x_{nc}) = -q\frac{\mathcal{D}_V}{L_V}\left[A_V\cosh\frac{x_{nc}-x_{nd}}{L_V} + B_V\sinh\frac{x_{nc}-x_{nd}}{L_V}\right] \tag{2.211}$$

$$0 = A_V\cosh\frac{x_{nc}-x_{nd}}{L_V} + B_V\sinh\frac{x_{nc}-x_{nd}}{L_V} \tag{2.212}$$

$$A_V = -B_V\tanh\frac{x_{nc}-x_{nd}}{L_V} \tag{2.213}$$

In a similar manner, we take

$$0 = j_C(x_{pc}) = -q\frac{\mathcal{D}_C}{L_C}\left[A_C\cosh\frac{-(x-x_{pd})}{L_C} + B_C\sinh\frac{-(x-x_{pd})}{L_C}\right] \tag{2.214}$$

and we obtain

$$A_C = -B_C\tanh\frac{-(x_{pc}-x_{pd})}{L_C} \tag{2.215}$$

for n carriers in the p-type quasineutral region.

2.2.7 Total current

We now know the minority currents at the edges of the depletion region $j_C(x_{pd})$ and $j_V(x_{nd})$. To relate the currents on either side of the depletion region, we require one extra assumption.

Assumption 3. *There is no recombination in the depletion region.*

This assumption immediately implies that

$$j_C(x_{pd}) = j_C(x_{nd}) \tag{2.216}$$

and so

$$\begin{aligned} j_V(x_{nd}) + j_C(x_{pd}) &= j_V(x_{nd}) + j_C(x_{nd}) && (2.217)\\ &= j(x_{nd}) && (2.218)\\ &= j(x) \quad \forall x && (2.219) \end{aligned}$$

where we used the constancy of total current at steady-state in a 1d device (Eq. 2.187).

We can now take Eq. 2.213 and use it to evaluate the current Eq. 2.204 at $x = x_{nd}$,

$$\begin{aligned} j_V(x_{nd}) &= -q\frac{\mathcal{D}_V}{L_V}\left[\underbrace{A_V}_{-B_V \tanh \frac{x_{nc}-x_{nd}}{L_V}} \underbrace{\cosh \frac{x_{nd}-x_{nd}}{L_V}}_{1} + B_V \underbrace{\sinh \frac{x_{nd}-x_{nd}}{L_V}}_{0}\right] && (2.220)\\ &= q\frac{\mathcal{D}_V}{L_V} \underbrace{B_p}_{u_V(x_{nd})-u_V^{(0)}} \tanh \frac{x_{nc}-x_{nd}}{L_V} && (2.221)\\ &= q\frac{\mathcal{D}_V}{L_V}\big(u_V(x_{nd}) - u_V^{(0)}\big) \tanh \frac{x_{nc}-x_{nd}}{L_V} && (2.222) \end{aligned}$$

Similarly we take Eq. 2.215 and Eq. 2.208,

$$\begin{aligned} j_C(x_{pd}) &= -q\frac{\mathcal{D}_C}{L_C}\left[\underbrace{A_C}_{B_C \tanh \frac{x_{pc}-x_{pd}}{L_C}} \underbrace{\cosh \frac{-(x_{pd}-x_{pd})}{L_C}}_{1} + B_C \underbrace{\sinh \frac{-(x_{pd}-x_{pd})}{L_C}}_{0}\right] && (2.223)\\ &= -q\frac{\mathcal{D}_C}{L_C} \underbrace{B_C}_{u_C(x_{pd})-u_C^{(0)}} \tanh \frac{x_{pc}-x_{pd}}{L_C} && (2.224)\\ &= -q\frac{\mathcal{D}_C}{L_C}\big(u_C(x_{pd}) - u_C^{(0)}\big) \tanh \frac{x_{pc}-x_{pd}}{L_C} && (2.225) \end{aligned}$$

Summing these up, we have

$$j = j_V(x_{nd}) + j_C(x_{pd}) \tag{2.226}$$

$$= q\left[\frac{\mathcal{D}_V}{L_V}\big(u_V(x_{nd}) - p_V^{(0)}\big)\tanh\frac{x_{nc}-x_{nd}}{L_V} + \frac{\mathcal{D}_C}{L_C}\big(u_C(x_{pd}) - u_C^{(0)}\big)\tanh\frac{x_{pc}-x_{pd}}{L_C}\right] \tag{2.227}$$

We now note that $u_C^{(0)} = \frac{u_i^2}{N_A}$ is the equilibrium minority carrier concentration. Using the law of the junction Eq. 2.151, we obtain

$$u_C(x_{pd}) = \frac{u_i^2}{N_A}e^{\beta q V_{\text{ext}}} = u_C^{(0)}e^{\beta q V_{\text{ext}}} \tag{2.228}$$

and so

$$u_C(x_{pd}) - u_C^{(0)} = u_C^{(0)}(e^{\beta q V_{\text{ext}}} - 1) \tag{2.229}$$

$$= \frac{u_i^2}{N_A}(e^{\beta q V_{\text{ext}}} - 1) \tag{2.230}$$

and similarly

$$u_V(x_{nd}) - u_V^{(0)} = \frac{u_i^2}{N_D}(e^{\beta q V_{\text{ext}}} - 1) \tag{2.231}$$

Therefore we can rewrite Eq. 2.227 as

$$j = q\left[\frac{\mathcal{D}_V}{L_V}\big(u_V(x_{nd}) - u_V^{(0)}\big)\tanh\frac{x_{nc}-x_{nd}}{L_V} + \frac{\mathcal{D}_C}{L_C}\big(u_C(x_{pd}) - u_C^{(0)}\big)\tanh\frac{x_{pc}-x_{pd}}{L_C}\right] \tag{2.232}$$

$$= qu_i^2\left[\frac{\mathcal{D}_V}{L_V N_D}\tanh\frac{x_{nc}-x_{nd}}{L_V} + \frac{\mathcal{D}_C}{L_C N_A}\tanh\frac{x_{pc}-x_{pd}}{L_C}\right](e^{\beta q V_{\text{ext}}} - 1) \tag{2.233}$$

2.2.8 Shockley equation

Let's assume that $x_{nc} - x_{nd} \gg L_V$ and $x_{pc} - x_{pd} \gg L_C$ (i.e., the contacts are far away from the depletion region compared to the diffusion length L_*).[9] Since $\tanh x \approx 1$ for $x \gg 1$, and Eq. 2.233 becomes

$$j = qu_i^2\left[\frac{\mathcal{D}_V}{L_V N_D}\tanh\frac{x_{nc}-x_{nd}}{L_V} + \frac{\mathcal{D}_C}{L_C N_A}\tanh\frac{x_{pc}-x_{pd}}{L_C}\right](e^{\beta q V_{\text{ext}}} - 1) \tag{2.234}$$

$$\approx qu_i^2\left[\frac{\mathcal{D}_V}{L_V N_D} + \frac{\mathcal{D}_C}{L_C N_A}\right](e^{\beta q V_{\text{ext}}} - 1) \tag{2.235}$$

[9] Though that would make for a very bad solar cell!

which is called the Shockley equation [32, Eq. 19.78].

2.2.9 Under illumination

We turn on the lights. We introduce the effects of illumination by considering the case where all generation is within the n-type region, and that it is spatially uniform. This modifies the minority carrier continuity equation within the n-type region Eq. 2.116,

$$0 = \frac{\partial u_V}{\partial t} = -\frac{1}{q}\vec{\nabla}\cdot\vec{j}_V - \frac{u_V - u_V^{(0)}}{\tau_V} + g \tag{2.236}$$

$$0 = -\frac{1}{q}\vec{\nabla}\cdot\vec{j}_V - \frac{u_V - (u_V^{(0)} + g\tau_V)}{\tau_V} \tag{2.237}$$

where g is the uniform generation in the n-type region (and has units of concentration per unit time).

By performing the $u_V^{(0)} \mapsto u_V^{(0)} + g\tau_V$ substitution, we can reuse almost all of our work from the previous (dark current) section. We therefore obtain

$$\begin{aligned}
j &= j_V(x_{nd}) + j_C(x_{pd}) \\
&= q\left[\frac{\mathcal{D}_V}{L_V}\big(u_V(x_{nd}) - u_V^{(0)} - g\tau_V\big)\tanh\frac{x_{nc}-x_{nd}}{L_V} + \frac{\mathcal{D}_C}{L_C}\big(u_C(x_{pd}) - u_C^{(0)}\big)\tanh\frac{x_{pc}-x_{pd}}{L_C}\right] && (2.238)\\
&= -\underbrace{(-q)u_i^2\left[\frac{\mathcal{D}_V}{L_V N_D}\tanh\frac{x_{nc}-x_{nd}}{L_V} + \frac{\mathcal{D}_C}{L_C N_A}\tanh\frac{x_{pc}-x_{pd}}{L_C}\right]}_{j_0}(e^{\beta q V_{\text{ext}}} - 1) + \underbrace{(-q)\frac{\mathcal{D}_V}{L_V}g\tau_V\tanh\frac{x_{nc}-x_{nd}}{L_V}}_{j_{\text{sc}}} && (2.239)\\
&= j_{\text{sc}} - \underbrace{j_0(e^{\beta q V_{\text{ext}}} - 1)}_{j_{\text{dark}}} && (2.240)
\end{aligned}$$

which is the current generated by a solar cell under illumination.

A little explanation for the names – we're at short circuit if and only if $V_{\text{ext}} = 0$, which implies

$$\begin{aligned}
j &= j_{\text{sc}} - \underbrace{j_0(e^{\beta q \overbrace{V_{\text{ext}}}^{0}} - \overbrace{1}^{1})}_{j_{\text{dark}}} && (2.241)\\
&= j_{\text{sc}} && (2.242)
\end{aligned}$$

If there is no illumination $j_{\text{sc}} = 0$ (since $g = 0$), so $j = j_{\text{sc}} - j_0(e^{\beta q V_{\text{ext}}} - 1) = j_0(e^{\beta q V_{\text{ext}}} - 1)$.

2.2.10 Depletion region recombination

In many real devices, a large fraction of the generation or recombination occurs inside the depletion region. We remove Assumption 3 (no recombination) and instead consider a nontrivial recombination model in the depletion region. We perform this last refinement to the depletion approximation model as a prelude to comparing it against a full drift-diffusion numerical model.

2.2.10.1 Shockley-Read-Hall recombination

A common type of nonradiative recombination is described by the Shockley-Read-Hall model. Using [35, Eq. 4.4] as the starting point, the steady state recombination rate from conduction to valence band through a trap[10] with energy level $\mathcal{E}_t$ is

$$U_{\text{SRH}} = \frac{K_C K_V (u_C u_V - \overbrace{u_{C,1} u_{V,1}}^{u_i^2})}{K_C (u_C + u_{C,1}) + K_V (u_V + u_{V,1})} \tag{2.243}$$

$$= \frac{u_C u_V - u_i^2}{\underbrace{K_V^{-1}}_{\tau_V}(u_C + u_{C,1}) + \underbrace{K_C^{-1}}_{\tau_C}(u_V + u_{V,1})} \tag{2.244}$$

$$= \frac{u_C u_V - u_i^2}{\tau_V (u_C + u_{C,1}) + \tau_C (u_V + u_{V,1})} \tag{2.245}$$

where K_k is the probability per unit time that a carrier from band k will be captured if the trap is available to it (i.e., the trap must be empty in order to accept an electron from the CB, and it must be full in order to accept a hole from the VB). The quantity $u_{k,1}$ represents the carrier density if the quasi-Fermi level of the band were equal to the trap energy level [36, Eq. 5.18]

$$u_{k,1} = u_i e^{s_k \beta (w_i - \mathcal{E}_t')} \tag{2.246}$$

where

$$\mathcal{E}_t' = \mathcal{E}_t + k_B T \ln g_t \cdot \begin{cases} +1 & \text{acceptor-like traps} \\ -1 & \text{donor-like traps} \end{cases} \tag{2.247}$$

and $\mathcal{E}_t$ is the actual energy level of the trap and g_t is its degeneracy.

[10] Usually a crystal defect or an undesirable impurity atom.

Remark 2. We note that in the limit $u_C \gg u_{C,1}$ and $\tau_V u_C \gg \tau_C(u_V + u_{V,1})$,

$$U_{\text{SRH}} = \frac{u_C u_V - u_i^2}{\tau_V(u_C + u_{C,1}) + \tau_C(u_V + u_{V,1})} \tag{2.248}$$

$$\approx \frac{u_V - u_i^2/u_C}{\tau_V} \tag{2.249}$$

In the high n-type doping limit we have $u_C \approx u_C^{(0)}$, and so we recover the simple relaxation model for minority carriers (Eq. 2.115)

$$U_{\text{SRH}} = \frac{u_V - u_i^2/u_C}{\tau_V} \tag{2.250}$$

$$\approx \frac{u_V - u_i^2/u_C^{(0)}}{\tau_V} \tag{2.251}$$

$$= \frac{u_V - u_V^{(0)}}{\tau_V} \tag{2.252}$$

A similar argument can be made for the p-type doping limit.

2.2.10.2 Current

Given a recombination model (e.g., Eq. 2.245) for the depletion region, we can use Eq. 2.114b to write

$$\underbrace{\frac{\partial u_C}{\partial t}}_{0} = -\underbrace{s_C}_{-1}\frac{1}{q}\vec{\nabla}\cdot\vec{j}_C + \underbrace{g_C}_{-U} \tag{2.253}$$

$$U = \frac{1}{q}\vec{\nabla}\cdot\vec{j}_C \tag{2.254}$$

Assuming current flows exclusively along the x-axis (without loss of generality), the fundamental theorem of calculus [37] gives us

$$U = \frac{1}{q}\vec{\nabla}\cdot\vec{j}_C \tag{2.255}$$

$$U = \frac{1}{q}\frac{\partial j_C}{\partial x} \tag{2.256}$$

$$\int_a^b U \,\mathrm{d}x = \frac{1}{q}\big[j_C(b) - j_C(a)\big] \tag{2.257}$$

In particular, we can now relate the current density on either side of the depletion region by

$$q\int_{x_{pd}}^{x_{nd}} U\,\mathrm{d}x = j_C(x_{nd}) - j_C(x_{pd}) \tag{2.258}$$

which replaces Eq. 2.216. We easily obtain the total current

$$j = j_V(x_{nd}) + j_C(x_{nd}) \tag{2.259}$$

$$= j_V(x_{nd}) + j_C(x_{pd}) + q\int_{x_{pd}}^{x_{nd}} U\,\mathrm{d}x \tag{2.260}$$

which allows us to amend Eq. 2.233

$$j = qu_i^2\left[\frac{\mathcal{D}_V}{L_V N_D}\tanh\frac{x_{nc}-x_{nd}}{L_V} + \frac{\mathcal{D}_C}{L_C N_A}\tanh\frac{x_{pc}-x_{pd}}{L_C}\right](e^{\beta q V_{\text{ext}}}-1) + q\int_{x_{pd}}^{x_{nd}} U\,\mathrm{d}x \tag{2.261}$$

The evaluation of the $\int U\,\mathrm{d}x$ term is often done numerically, since

$$\int U\,\mathrm{d}x = \int \frac{u_C u_V - u_i^2}{\tau_V(u_C+u_{C,1}) + \tau_C(u_V+u_{V,1})}\,\mathrm{d}x \tag{2.262}$$

$$= \int \frac{u_i e^{\beta(w_C^*+q\phi(x)-w_i)} u_i e^{-\beta(w_V^*+q\phi(x)-w_i)} - u_i^2}{\tau_V(u_i e^{\beta(w_C^*+q\phi(x)-w_i)}+u_{C,1}) + \tau_C(u_i e^{-\beta(w_V^*+q\phi(x)-w_i)}+u_{V,1})}\,\mathrm{d}x \tag{2.263}$$

$$= \int \frac{(e^{\beta(w_C^*-w_V^*)}-1)u_i^2}{\tau_V(u_i e^{\beta(w_C^*+q\phi(x)-w_i)}+u_{C,1}) + \tau_C(u_i e^{-\beta(w_V^*+q\phi(x)-w_i)}+u_{V,1})}\,\mathrm{d}x \tag{2.264}$$

probably doesn't have a clean analytic solution when $\phi(x)$ is a piecewise quadratic (Eq. 2.90).

2.2.10.3 Qualitative behaviour of $\int U\,\mathrm{d}x$ term

For the case where the recombination in Eq. 2.261 is dominated by SRH processes, we can get a qualitative sense of the $\int U\,\mathrm{d}x$ term. We can (over)estimate $U_{\text{SRH}}(x)$ it by assuming it to be equal to its value somewhere around the middle of the junction. We choose this to be the point x_m where $\phi = \frac{\Delta V}{2}$. Then

$$u_k(x_m) = u_i e^{s_k\beta[w_i - q\phi(x_m) - w_k(x_m)]} \tag{2.265}$$

$$= \underbrace{u_i e^{s_k\beta[w_i - q\phi(x_{(\text{maj},k)d}) - w_k(x_{(\text{maj},k)d})]}}_{u_k^{\text{majority}}} e^{s_k\beta[q\overbrace{\phi(x_{(\text{maj},k)d})}^{\Delta V/2} + w_k(x_{(\text{maj},k)d}) - q\overbrace{\phi(x_m)}^{\Delta V/2} - \overbrace{w_k(x_m)}^{=w_k(x_{(\text{maj},k)d})}]} \tag{2.266}$$

$$= u_k^{\text{majority}} e^{s_k\beta[q\phi(x_{(\text{maj},k)d}) - q\Delta V/2]} \tag{2.267}$$

where $x_{(\mathrm{maj},k)d}$ is the edge of the depletion region on the side where band k's carriers are the majority carrier (i.e., $x_{(\mathrm{maj},C)d} = x_{nd}$ and $x_{(\mathrm{maj},V)d} = x_{pd}$) and $u_k^{\mathrm{majority}} = u_k(x_{(\mathrm{maj},k)d})$ is the majority carrier concentration.

Since

$$\phi(x_{(\mathrm{maj},k)d}) = \begin{cases} 0 & k = V \\ \Delta V & k = C \end{cases} \tag{2.268}$$

we have

$$u_k(x_m) = u_k^{\mathrm{majority}} e^{s_k \beta [q\phi(x_{(\mathrm{maj},k)d}) - q\Delta V/2]} \tag{2.269}$$

$$= u_k^{\mathrm{majority}} e^{-\beta q \Delta V/2} \tag{2.270}$$

and so

$$U_{\mathrm{SRH}}(x_m) = \frac{u_C(x_m)u_V(x_m) - u_i^2}{\tau_V(u_C(x_m) + u_{C,1}) + \tau_C(u_V(x_m) + u_{V,1})} \tag{2.271}$$

$$= \frac{u_C^{\mathrm{majority}} u_V^{\mathrm{majority}} e^{-\beta q \Delta V} - u_i^2}{\tau_V(u_C^{\mathrm{majority}} e^{-\beta q \Delta V/2} + u_{C,1}) + \tau_C(u_V^{\mathrm{majority}} e^{-\beta q \Delta V/2} + u_{V,1})} \tag{2.272}$$

Assuming that $u_k^{\mathrm{majority}} e^{-q\Delta V/2} \gg u_{k,1}$ and using $u_C^{\mathrm{majority}} \approx N_D$ and $u_V^{\mathrm{majority}} \approx N_A$, we have

$$U_{\mathrm{SRH}}(x_m) \approx \frac{u_C^{\mathrm{majority}} u_V^{\mathrm{majority}} e^{-\beta q \Delta V} - u_i^2}{\tau_V(u_C^{\mathrm{majority}} e^{-\beta q \Delta V/2}) + \tau_C(u_V^{\mathrm{majority}} e^{-\beta q \Delta V/2})} \tag{2.273}$$

$$= \frac{u_C^{\mathrm{majority}} u_V^{\mathrm{majority}} e^{-\beta q \Delta V/2} - u_i^2 e^{\beta q \Delta V/2}}{\tau_V u_C^{\mathrm{majority}} + \tau_C u_V^{\mathrm{majority}}} \tag{2.274}$$

$$\approx \frac{N_A N_D e^{-\beta q \Delta V/2} - u_i^2 e^{\beta q \Delta V/2}}{\tau_V N_D + \tau_C N_A} \tag{2.275}$$

Since

$$\Delta V = \overbrace{\frac{1}{\beta q} \ln \frac{N_A N_D}{u_i^2}}^{V_{\mathrm{bi}}} - V_{\mathrm{ext}} \tag{2.276}$$

$$e^{\beta q \Delta V} = \frac{N_A N_D}{u_i^2} e^{-\beta q V_{\mathrm{ext}}} \tag{2.277}$$

$$e^{\beta q \Delta V/2} = \frac{\sqrt{N_A N_D}}{u_i} e^{-\beta q V_{\mathrm{ext}}/2} \tag{2.278}$$

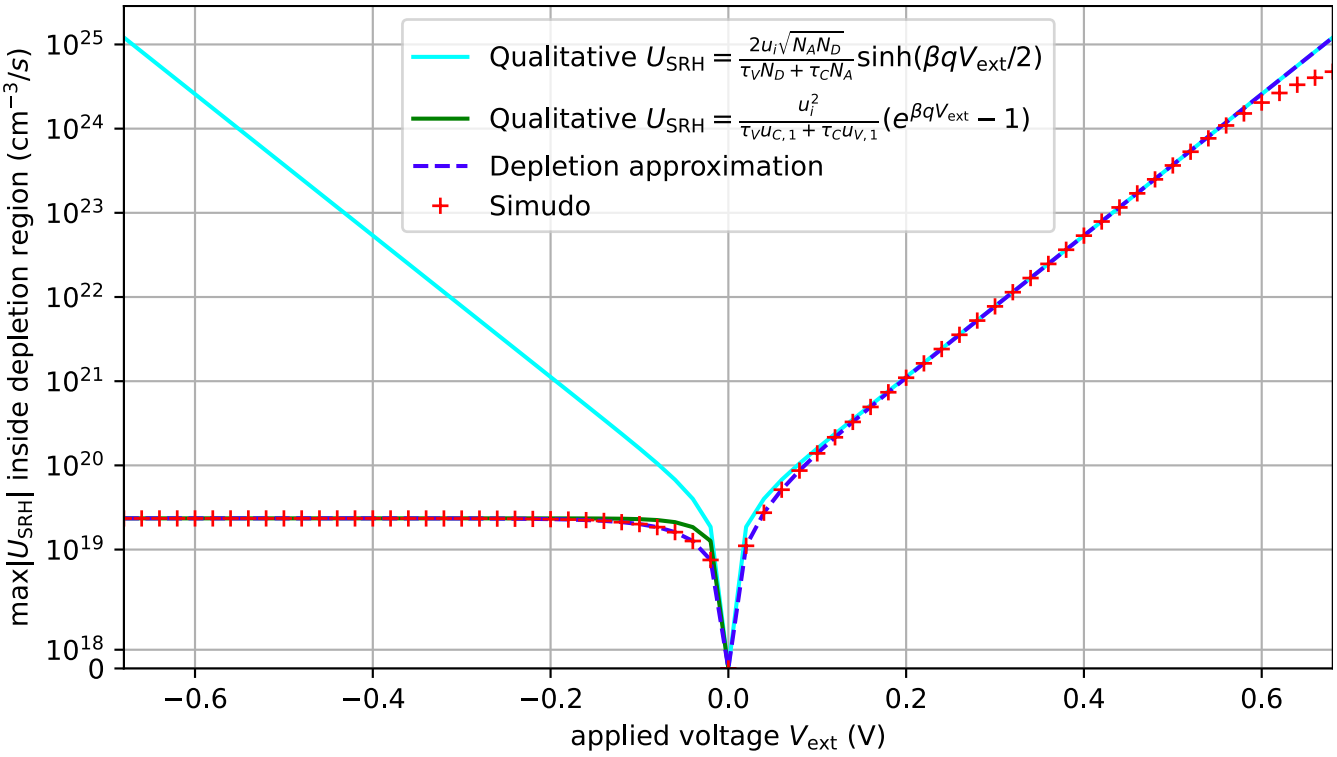

Figure 2.7: Comparison of the peak $|U_{\text{SRH}}|$ in the depletion region of a silicon pn junction between the depletion approximation, a full numerical drift-diffusion model (Simudo), and the qualitative formulas Eq. 2.282 and 2.287. The parameters are described in Figure 2.8. The qualitative result Eq. 2.282 is valid at forward bias but exponentially wrong at reverse bias, while the qualitative result Eq. 2.287 is valid only at reverse bias.

we get

$$U_{\text{SRH}}(x_m) \approx \frac{N_A N_D e^{-\beta q \Delta V/2} - u_i^2 e^{\beta q \Delta V/2}}{\tau_V N_D + \tau_C N_A} \tag{2.279}$$

$$= \frac{N_A N_D \frac{u_i}{\sqrt{N_A N_D}} e^{\beta q V_{\text{ext}}/2} - u_i^2 \frac{\sqrt{N_A N_D}}{u_i} e^{-\beta q V_{\text{ext}}/2}}{\tau_V N_D + \tau_C N_A} \tag{2.280}$$

$$= \frac{u_i \sqrt{N_A N_D}(e^{\beta q V_{\text{ext}}/2} - e^{-\beta q V_{\text{ext}}/2})}{\tau_V N_D + \tau_C N_A} \tag{2.281}$$

$$= \frac{2u_i \sqrt{N_A N_D}}{\tau_V N_D + \tau_C N_A} \sinh(\beta q V_{\text{ext}}/2) \tag{2.282}$$

As shown in Figure 2.7, this qualitative result is valid only at forward bias. This is because the assumption that $u_k^{\text{majority}} e^{-q\Delta V/2} \gg u_{k,1}$ is violated at reverse bias. We need to derive a different result for $V_{\text{ext}} < 0$.

We note that

$$u_C^{\text{majority}} u_V^{\text{majority}} e^{-\beta q \Delta V} \approx N_A N_D e^{-\beta q \Delta V} \tag{2.283}$$

$$= N_A N_D \frac{u_i^2}{N_A N_D} e^{\beta q V_{\text{ext}}} \tag{2.284}$$

$$= u_i^2 e^{\beta q V_{\text{ext}}} \tag{2.285}$$

Adding the assumption that $u_k^{\text{majority}} e^{-q\Delta V/2} \ll u_{k,1}$, we can write

$$U_{\text{SRH}}(x_m) = \frac{\overbrace{u_C^{\text{majority}} u_V^{\text{majority}} e^{-\beta q \Delta V}}^{u_i^2 e^{\beta q V_{\text{ext}}}} - u_i^2}{\tau_V(\underbrace{u_C^{\text{majority}} e^{-\beta q \Delta V/2}}_{\ll u_{C,1}} + u_{C,1}) + \tau_C(\underbrace{u_V^{\text{majority}} e^{-\beta q \Delta V/2}}_{\ll u_{V,1}} + u_{V,1})} \tag{2.286}$$

$$\approx \frac{u_i^2}{\tau_V u_{C,1} + \tau_C u_{V,1}}(e^{\beta q V_{\text{ext}}} - 1) \tag{2.287}$$

We can thus approximate Eq. 2.261 as

$$j = j_0(e^{\beta q V_{\text{ext}}} - 1) + qW \begin{cases} \frac{2u_i\sqrt{N_A N_D}}{\tau_V N_D + \tau_C N_A} \sinh(\beta q V_{\text{ext}}/2) & V_{\text{ext}} > 0 \\ \frac{u_i^2}{\tau_V u_{C,1} + \tau_C u_{V,1}}(e^{\beta q V_{\text{ext}}} - 1) & V_{\text{ext}} < 0 \end{cases} \tag{2.288}$$

where $j_0 = q u_i^2 \left[\frac{\mathcal{D}_V}{L_V N_D} \tanh \frac{x_{nc} - x_{nd}}{L_V} + \frac{\mathcal{D}_C}{L_C N_A} \tanh \frac{x_{pc} - x_{pd}}{L_C} \right]$. Note that Eq. 2.288 vastly overestimates the current produced in the depletion region by assuming that the generation/recombination is at its maximum value throughout the entire depletion region.

In the limit $\beta q V_{\text{ext}}/2 \gg 1$, we can further write

$$j = j_0(e^{\beta q V_{\text{ext}}} - 1) + qW \frac{2u_i\sqrt{N_A N_D}}{\tau_V N_D + \tau_C N_A}(e^{\beta q V_{\text{ext}}/2} - 1) \tag{2.289}$$

and note that this $j(V_{\text{ext}})$ curve corresponds to two diodes placed in parallel with "ideality factors" of 1 and 2 respectively. The ideality factor of a dark current term of the form $e^{\beta q V_{\text{ext}}/n}$ is n. Real devices often have non-integer ideality factors between 1 and 2 [38], but larger ideality factors are possible [39].

2.2.11 Solution quality

The depletion region approximation is great for pedagogical purposes. This model provides a precious shard of intuition on how semiconductor devices work qualitatively. However, it is very limited:

1. We assumed perfect Ohmic contacts ($S = 0$). They're usually not.
2. We assumed the optical generation to be spatially constant. It's usually not.
3. The relaxation time approximation may not be accurate in high injection, and may require Shockley-Read-Hall or Auger recombination.
4. We assumed the same semiconductor is on both sides of the junction, i.e., it is a homojunction. Some devices use heterojunctions.
5. We assumed the material properties to be uniform. Doping might not be perfectly sharp at the junction. Some materials are alloys, and their composition may vary spatially (intentionally or not).
6. We assumed there are only two bands (conduction and valence), which excludes intermediate band materials.

Some of these limitations can be overcome, to some extent, by refining the model. However, developing a more advanced model usually requires additional assumptions, and the resulting model (and its assumptions) often need to be validated against a higher source of truth, such as a numerical solver.

Most of these limitations cannot be overcome analytically. To predict electrical or other microscopic properties for a larger class of materials/devices than considered in this derivation, we *require* a numerical solver. One way of solving the drift-diffusion equations (Eq. 2.114) and Poisson's equation (Eq. 2.56) numerically is the finite element method, to which we give an introduction in Chapter 4. Finally, we introduce our device model called Simudo in Chapter 5.

To close this chapter, we present a direct comparison between the depletion approximation (including SRH recombination in the depletion region) and Simudo in Figures 2.8 and 2.9. Despite significant differences in u_k and $\vec{E}$ near the depletion region, the depletion approximation succeeds in finding total currents within 2–10% of Simudo as shown in Figure 2.10. Semianalytic models like the depletion approximation can be a powerful way to quickly estimate device properties (such as the $j_{\text{total}}(V_{\text{ext}})$ curve or solar cell efficiency). A semianalytic model for intermediate band solar cell efficiency was developed in [25] with the help of Simudo for validation.

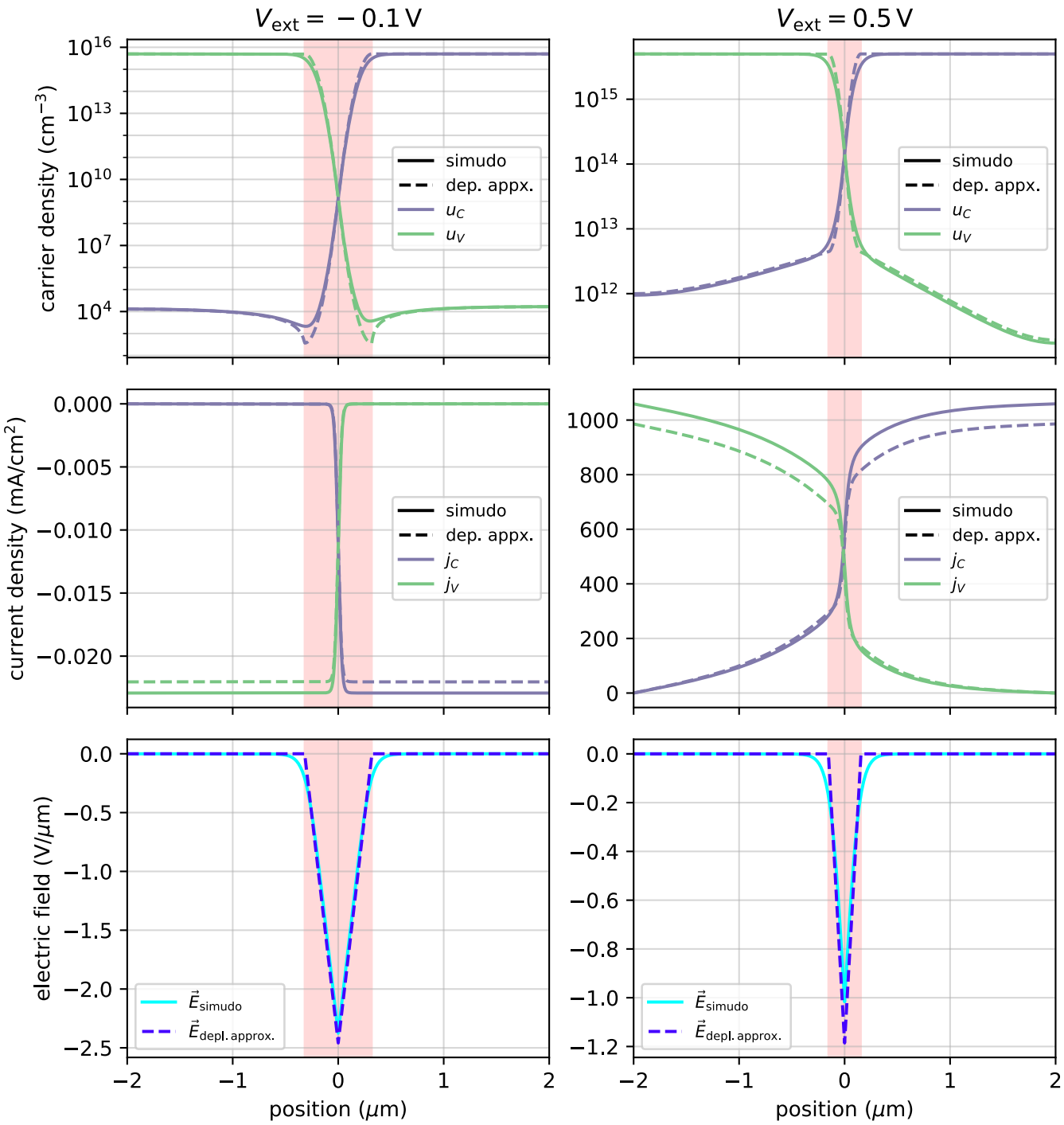

Figure 2.8: Comparison of $u_k, j_k, \vec{E}$ for the depletion approximation (including SRH recombination via Eqs. 2.245 and 2.257) versus a full numerical drift-diffusion model (Simudo). The depletion region is shown in red. The device is a silicon pn junction with $N_D = N_A = 5 \times 10^{15}$ cm^{-3} at $T = 300$ K with minority carrier lifetimes $\tau_C = \tau_V = 2 \times 10^{-10}$ s and SRH trap energy level $\mathcal{E}_t = \mathcal{E}_V + 0.5525$ eV. The depletion approximation solution underestimates the decline of the majority carrier into the depletion region and ends up significantly underestimating the magnitude of the generated current as a result.

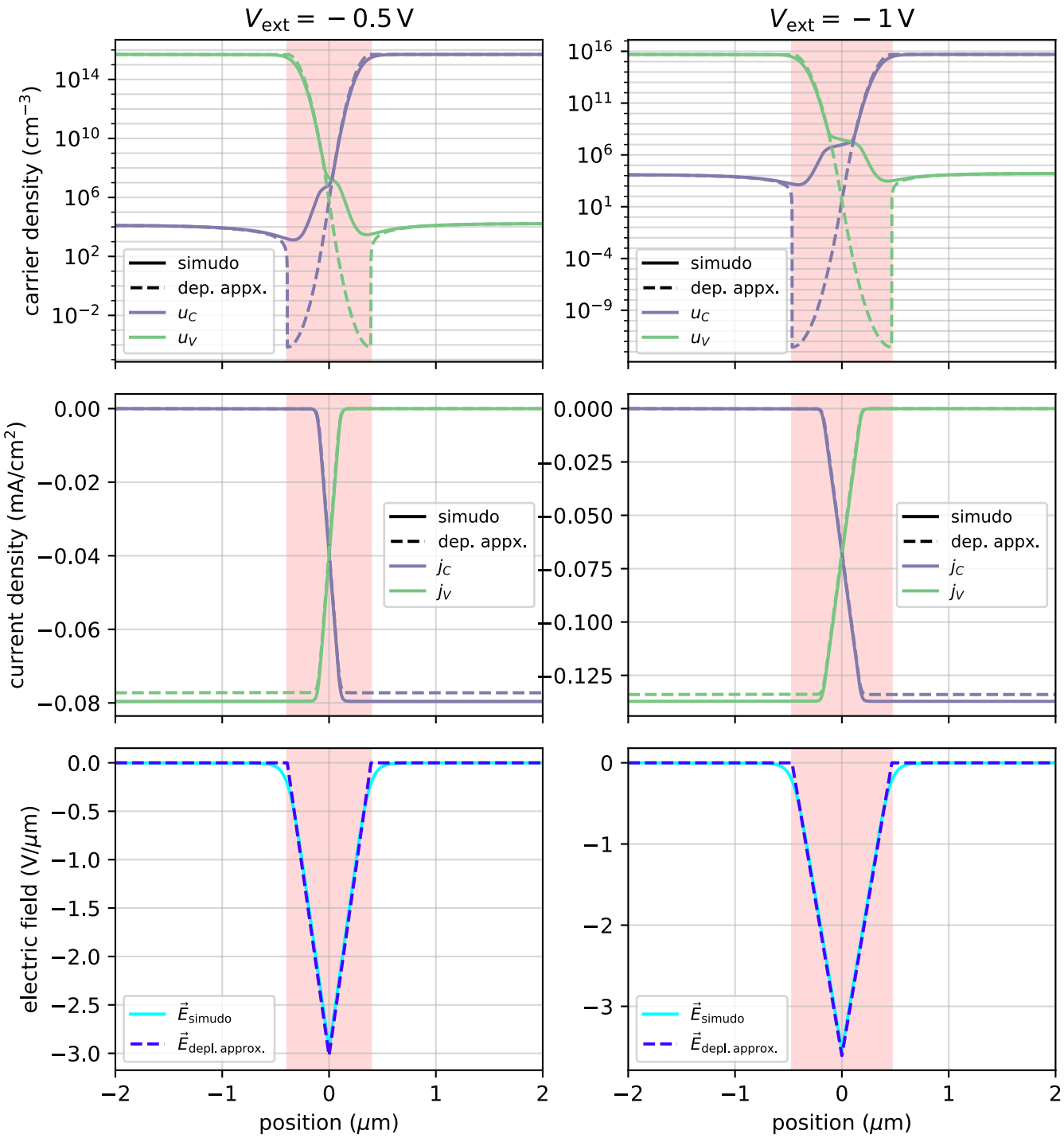

Figure 2.9: Comparison of $u_k, j_k, \vec{E}$ for the depletion approximation (including SRH recombination via Eqs. 2.245 and 2.257) versus a full numerical drift-diffusion model (Simudo) at larger-magnitude negative bias V_{ext}. The conditions and parameters are the same as in Figure 2.8. The law of the junction fails to determine the carrier densities inside and at the edges of the depletion region, so the minority carrier densities and currents in the quasi-neutral regions are incorrect. None of that ends up mattering because the vast bulk of the recombination happens in the middle of the depletion region, which is why the "depletion approximation" currents still agree with Simudo. Excluding the depletion region $\int U_{\text{SRH}}\,\mathrm{d}x$ term leads to vastly underestimating the total current at reverse bias as can be seen in Figure 2.10.

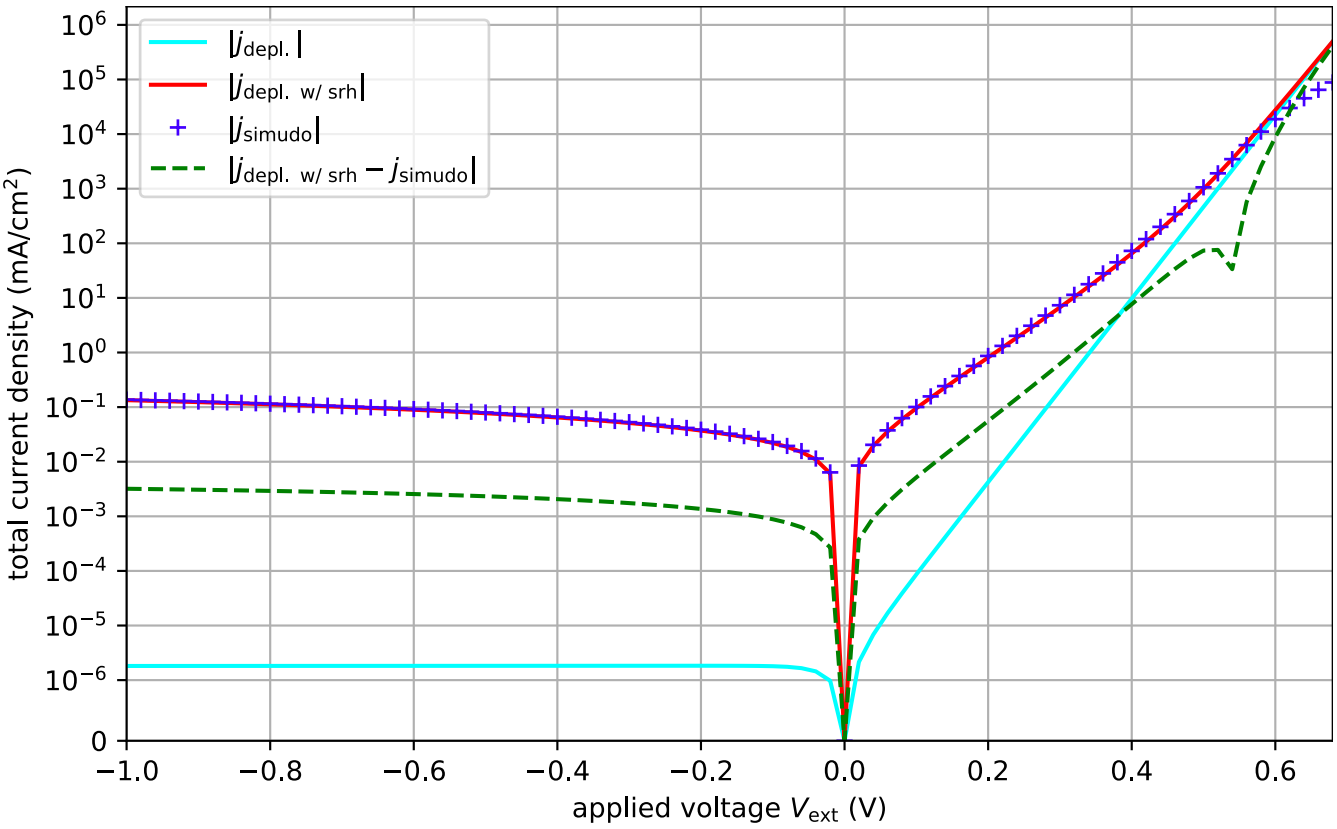

Figure 2.10: Comparison of j_{total} for the depletion approximation (with and without SRH recombination in the depletion region via Eq. 2.261) versus Simudo. The conditions and parameters are the same as in Figure 2.8. At reverse bias, most of the current is generated in the depletion region and so the no-SRH depletion approximation solution vastly underestimates the total current. As V_{ext} increases more of the current is generated in the quasi-neutral regions rather than in the (shrinking) depletion region, so the no-SRH depletion approximation solution becomes better. The depletion approximation solution including SRH yields total currents that are not just qualitatively correct but are only 2–10% away from the true (numerical drift-diffusion) solution.

Chapter 3

A more formal look at the law of the junction

Having powered through the depletion region approximation in sections 2.2.4.1 and 2.2.4.2 and convinced themselves of its validity, the mathematically-inclined reader may have grimaced at our rather informal justifications of the law of the junction. As a penance, we derive an exact solution to the drift-diffusion equations within the depletion region (under the simplifying assumptions of no recombination or generation within this region), and we find sufficient conditions for the validity of the law of the junction. A discussion of similar issues is in [3], but ours is a bit more general and hopefully more approachable to the curious but novice reader.

To simplify the notation, we shall focus on the valence band carrier and write $p(x) \equiv u_V(x)$ for the rest of this section. The differential equation we will solve is then

$$\vec{j}_V = q\mu_V p\vec{E} - q \underbrace{\mathcal{D}_p}_{\mu_V/\beta q} \vec{\nabla} p \tag{3.1}$$

$$\frac{j_V}{\mu_V} = qpE - \frac{1}{\beta}\frac{\mathrm{d}p}{\mathrm{d}x} \tag{3.2}$$

Since we are assuming the lack of generation or recombination in the depletion region, j_V must be constant (by the continuity equation).

3.1 Derivation of exact drift-diffusion in depletion region without recombination or generation

Theorem 1. *The solution $p(x)$ to the boundary value problem*

$$\frac{j_V}{\mu_V} = qp(x)E(x) - \frac{1}{\beta}\frac{\mathrm{d}p(x)}{\mathrm{d}x} \tag{3.3}$$

where $p(x_{nd})$ and $p(x_{pd})$ are fixed, $E(x)$ is given by Eq. 2.78, and j_V, μ_V, q, β are constants is given by

$$p(x_{nd}) + \zeta\xi_n \operatorname{erf}\frac{W_n}{\xi_n} = e^{-\beta q\Delta V}\Big(p(x_{pd}) - \zeta\xi_p \operatorname{erfi}\frac{W_p}{\xi_p}\Big) \tag{3.4}$$

where

$$\begin{aligned} \operatorname{erf}(x) &= \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,\mathrm{d}t, & \xi_n &= \sqrt{\frac{2\varepsilon}{\beta q^2 N_D}}, \\ \operatorname{erfi}(x) &= \frac{2}{\sqrt{\pi}}\int_0^x e^{t^2}\,\mathrm{d}t, & \xi_p &= \sqrt{\frac{2\varepsilon}{\beta q^2 N_A}}, \end{aligned} \tag{3.5}$$

$$\textit{and} \quad \zeta = \frac{\sqrt{\pi}}{2}\frac{\beta j_V}{\mu_V}. \tag{3.6}$$

Proof. We have a differential equation of the form

$$\frac{\mathrm{d}y}{\mathrm{d}x} + P(x)y = Q(x) \tag{3.7}$$

We need to use integrating factor, defined as $h(x) = e^{\int^x P(x)\,\mathrm{d}x}$ so that $\frac{\mathrm{d}}{\mathrm{d}x}h(x) = P(x)\,h(x)$, and so that

$$\frac{\mathrm{d}}{\mathrm{d}x}(hy) = (y\underbrace{\frac{\mathrm{d}}{\mathrm{d}x}h}_{Ph} + h\frac{\mathrm{d}}{\mathrm{d}x}y) \tag{3.8}$$

$$= (yPh + h\frac{\mathrm{d}y}{\mathrm{d}x}) \tag{3.9}$$

$$= (\underbrace{Py + \frac{\mathrm{d}y}{\mathrm{d}x}}_{Q(x)})h \tag{3.10}$$

We then obtain solution by integrating

$$\frac{\mathrm{d}}{\mathrm{d}x}(hy) = Qh \tag{3.11}$$

$$hy = \int^x Q(x)\, h(x)\, \mathrm{d}x \tag{3.12}$$

$$y = \frac{1}{h}\int^x Q(x)\, h(x)\, \mathrm{d}x \tag{3.13}$$

We identify $P(x) = -\beta q E(x)$ and $Q(x) = -\beta \frac{j_V}{\mu_V}$. Then $\int^x P(x)\,\mathrm{d}x = \int^x -\beta q E(x)\,\mathrm{d}x = \beta q \phi(x)$, and so

$$p(x) = -e^{-\beta q\phi(x)}\left(\int^x \beta \frac{j_V}{\mu_V} e^{\beta q\phi(x)}\,\mathrm{d}x + \text{const.}\right) \tag{3.14}$$

$$= e^{-\beta q\phi(x)}\left(c - \beta \frac{j_V}{\mu_V}\int^x e^{\beta q\phi(x)}\,\mathrm{d}x\right) \tag{3.15}$$

for some constant c. We fix lower endpoint of the integral,

$$p(x) = e^{-\beta q\phi(x)}\left(c - \beta \frac{j_V}{\mu_V}\int_{x_{pd}}^x e^{\beta q\phi(x)}\,\mathrm{d}x\right) \tag{3.16}$$

and plug in $x = x_{pd}$, resulting in

$$p(x_{pd}) = e^{-\beta q\phi(x_{pd})}\left(c - \beta \frac{j_V}{\mu_V}\int_{x_{pd}}^{x_{pd}} e^{\beta q\phi(x)}\,\mathrm{d}x\right) \tag{3.17}$$

$$p(x_{pd}) = c e^{-\beta q\phi(x_{pd})} \tag{3.18}$$

which yields $c = p(x_{pd}) e^{\beta q\phi(x_{pd})}$. Plugging in $x = x_{nd}$, we obtain

$$p(x_{nd}) = p(x_{pd}) e^{\beta q\left(\phi(x_{pd}) - \phi(x_{nd})\right)} - \beta \frac{j_V}{\mu_V} e^{-\beta q\phi(x_{nd})}\int_{x_{pd}}^{x_{nd}} e^{\beta q\phi(x)}\,\mathrm{d}x \tag{3.19}$$

Using $\phi(x_{pd}) = 0$ and $\phi(x_{nd}) = \Delta V$,

$$p(x_{nd}) = e^{-\beta q\Delta V}\left(p(x_{pd}) - \beta \frac{j_V}{\mu_V}\int_{x_{pd}}^{x_{nd}} e^{\beta q\phi(x)}\,\mathrm{d}x\right) \tag{3.20}$$

We now focus our efforts on evaluating the integral. We recall from Eq. 2.90,

$$\phi(x) - \phi(x_{pd}) = \frac{q}{\varepsilon}\frac{1}{2}N_A(x - x_{pd})^2 \qquad x \in [x_{pd}, x_J] \tag{3.21a}$$

$$\phi(x) - \phi(x_J) = \frac{q}{\varepsilon}\frac{1}{2}N_D\Big[W_n^2 - (x_{nd} - x)^2\Big] \qquad x \in [x_J, x_{nd}] \tag{3.21b}$$

where $W_n = x_{nd} - x_J$ and $W_p = x_J - x_{pd}$.

We rewrite the second line in terms of $\phi(x_{pd})$. For $x \in [x_{pd}, x_J]$,

$$\phi(x) - \phi(x_{pd}) = \big(\phi(x) - \phi(x_J)\big) + \big(\phi(x_J) - \phi(x_{pd})\big) \tag{3.22}$$

$$= \frac{q}{\varepsilon}\frac{1}{2}\left[N_D\Big[(x_{nd} - x_J)^2 - (x_{nd} - x)^2\Big] + N_A(x_J - x_{pd})^2\right] \tag{3.23}$$

$$= \frac{q}{\varepsilon}\frac{1}{2}\left[N_D\Big[W_n^2 - (x_{nd} - x)^2\Big] + N_A W_p^2\right] \tag{3.24}$$

We therefore have piecewise function

$$\phi(x) - \phi(x_{nd}) = \frac{q}{\varepsilon}\frac{1}{2}\begin{cases} N_A(x - x_{pd})^2 & x \in [x_{pd}, x_J] \\ N_D\Big[W_n^2 - (x_{nd} - x)^2\Big] + N_A W_p^2 & x \in [x_J, x_{nd}] \end{cases} \tag{3.25}$$

We must integrate over $x \in [x_{pd}, x_{nd}]$. We split up the integral into two parts, $x \in [x_{pd}, x_J]$ and $x \in [x_J, x_{nd}]$. We obtain

$$\int_{x_{pd}}^{x_J} e^{\beta q \phi(x)}\,\mathrm{d}x \tag{3.26}$$

$$= \int_{x_{pd}}^{x_J} \exp\left[\beta\frac{q^2}{\varepsilon}\frac{1}{2}\Big[N_A(x - x_{pd})^2\Big]\right]\mathrm{d}x \tag{3.27}$$

$$= \int_{x_{pd}}^{x_J} e^{\left((x - x_{pd})/\xi_p\right)^2}\,\mathrm{d}x \tag{3.28}$$

where $\xi_p = \sqrt{\frac{2\varepsilon}{\beta q^2 N_A}}$ and $\xi_n = \sqrt{\frac{2\varepsilon}{\beta q^2 N_D}}$. Using $\mathrm{erfi}(x) \equiv \frac{2}{\sqrt{\pi}}\int_0^x e^{x^2}\,\mathrm{d}x$ and the substitution $u =$

$(x - x_{pd})/\xi_p$,

$$= \int_{x_{pd}}^{x_J} e^{\left((x-x_{pd})/\xi_p\right)^2} \,\mathrm{d}x \tag{3.29}$$

$$= \left[\frac{\sqrt{\pi}}{2}\,\xi_p \,\operatorname{erfi}\frac{x - x_{pd}}{\xi_p}\right]_{x=x_{pd}}^{x_J} \tag{3.30}$$

$$= \frac{\sqrt{\pi}}{2}\,\xi_p \,\operatorname{erfi}\frac{x_J - x_{pd}}{\xi_p} \tag{3.31}$$

$$= \frac{\sqrt{\pi}}{2}\,\xi_p \,\operatorname{erfi}\frac{W_p}{\xi_p} \tag{3.32}$$

Now we do the second part over $x \in [x_J, x_{nd}]$,

$$\int_{x_J}^{x_{nd}} e^{\beta q\phi(x)} \,\mathrm{d}x \tag{3.33}$$

$$= \int_{x_J}^{x_{nd}} \exp\left[\beta\frac{q^2}{\varepsilon}\frac{1}{2}\left[N_D\left[W_n^2 - (x_{nd} - x)^2\right] + N_A W_p^2\right]\right] \mathrm{d}x \tag{3.34}$$

$$= e^{\beta\frac{q^2}{\varepsilon}\frac{1}{2}\left(N_D W_n^2 + N_A W_p^2\right)} \int_{x_J}^{x_{nd}} \exp\left[\beta\frac{q^2}{\varepsilon}\frac{1}{2}\left[- (x_{nd} - x)^2 N_D\right]\right] \mathrm{d}x \tag{3.35}$$

$$= e^{W_n^2/\xi_n^2 + W_p^2/\xi_p^2} \int_{x_J}^{x_{nd}} e^{-\left((x_{nd}-x)/\xi_n\right)^2} \,\mathrm{d}x \tag{3.36}$$

$$= e^{\overbrace{W_n^2/\xi_n^2 + W_p^2/\xi_p^2}^{\beta q\Delta V}} \left[\frac{\sqrt{\pi}}{2}\,(-\xi_n) \,\operatorname{erf}\frac{x_{nd} - x}{\xi_n}\right]_{x=x_J}^{x_{nd}} \tag{3.37}$$

$$= e^{\beta q\Delta V}\frac{\sqrt{\pi}}{2}\,\xi_n \,\operatorname{erf}\frac{W_n}{\xi_n} \tag{3.38}$$

where we used $\operatorname{erf}(x) \equiv \frac{2}{\sqrt{\pi}}\int_0^x e^{-x^2}\,\mathrm{d}x$, the substitution $u = (x_{nd} - x)/\xi_n$, and the equality

$$W_n^2/\xi_n^2 + W_p^2/\xi_p^2 = \frac{\beta q^2}{2\varepsilon}\Big(W_p^2 N_A + W_n^2 N_D\Big) \tag{3.39}$$

$$= \beta q\Delta V \tag{3.40}$$

from Eq. 2.110.

Therefore

$$\int_{x_{pd}}^{x_{nd}} e^{\beta q\phi(x)} \,\mathrm{d}x = \frac{\sqrt{\pi}}{2}\left(\xi_p \,\operatorname{erfi}\frac{W_p}{\xi_p} + e^{\beta q\Delta V}\,\xi_n \,\operatorname{erf}\frac{W_n}{\xi_n}\right) \tag{3.41}$$

$$p(x_{nd}) = e^{-\beta q \Delta V}\left(p(x_{pd}) - \beta\frac{j_V}{\mu_V}\int_{x_{pd}}^{x_{nd}} e^{\beta q\phi(x)}\,\mathrm{d}x\right) \tag{3.42}$$

$$= e^{-\beta q \Delta V}p(x_{pd}) - \beta\frac{j_V}{\mu_V}\frac{\sqrt{\pi}}{2}\left(\xi_p\, e^{-\beta q\Delta V}\ \mathrm{erfi}\,\frac{W_p}{\xi_p} + \xi_n\ \mathrm{erf}\,\frac{W_n}{\xi_n}\right) \tag{3.43}$$

or

$$p(x_{nd}) + \zeta\xi_n\,\mathrm{erf}\,\frac{W_n}{\xi_n} = e^{-\beta q\Delta V}\left(p(x_{pd}) - \zeta\xi_p\,\mathrm{erfi}\,\frac{W_p}{\xi_p}\right) \tag{3.44}$$

where $\zeta = \frac{\sqrt{\pi}}{2}\frac{\beta j_V}{\mu_V}$ has units of carrier concentration per length, i.e. cm^{-4}. □

3.2 Interpretation

Looking at the p side of the depletion region and translating the coordinate system such that $x_{pd} = 0$ (instead of $x_J = 0$ as usual), the drift current is then

$$j_{V,\mathrm{drift}} = q\mu_p \overbrace{(-ax)}^{E(x)} p(x) \tag{3.45}$$

We now look for the point where the drift current takes its maximum value

$$0 = \frac{\mathrm{d}j_{V,\mathrm{drift}}}{\mathrm{d}x} \tag{3.46}$$

$$0 = \frac{\mathrm{d}}{\mathrm{d}x}(px) \tag{3.47}$$

$$0 = p + x\,\frac{\mathrm{d}p}{\mathrm{d}x} \tag{3.48}$$

From the differential equation, we obtain

$$\frac{\mathrm{d}p}{\mathrm{d}x} = -\beta aqpx - \beta\frac{j_V}{\mu_V} \tag{3.49}$$

and so

$$0 = p + x\,\frac{\mathrm{d}p}{\mathrm{d}x} \tag{3.50}$$

$$0 = p + x\left(-\beta aqpx - \beta\frac{j_V}{\mu_V}\right) \tag{3.51}$$

$$0 = (1 - \beta aq\,x^2)p - \beta\,x\,\frac{j_V}{\mu_V} \tag{3.52}$$

At equilibrium $j_V = 0$, and

$$0 = (1 - \beta a q\, x^2) \underbrace{p}_{>0} \tag{3.53}$$

$$0 = (1 - \beta a q\, x^2) \tag{3.54}$$

$$x_K \equiv \sqrt{\frac{1}{\beta a q}} \tag{3.55}$$

This defines the length scale on which the $j_V \approx 0$ approximation may not be good since the electric field may not be strong enough compared to j_V. This length scale is precisely ξ_n, ξ_p, up to a factor of $\sqrt{\frac{1}{2}}$.

3.3 Revised law of the junction

We now process this result into an actual modified statement of the law of the junction. At equilibrium (dark), we most definitely have constant electrochemical potential throughout the device, and so

$$u_V^{(0)}(x_{pd}) = u_V^{(0)}(x_{nd}) e^{\beta q V_{\text{bi}}} \tag{3.56}$$

because

$$\frac{u_V^{(0)}(x_{pd})}{u_V^{(0)}(x_{nd})} = \frac{\cancel{\mu_i} e^{\beta(\cancel{\mu_i} - \cancel{\mu} - q\phi(x_{pd}))}}{\cancel{\mu_i} e^{\beta(\cancel{\mu_i} - \cancel{\mu} - q\phi(x_{nd}))}} \tag{3.57}$$

$$= e^{\beta q(\phi(x_{nd}) - q\phi(x_{pd}))} \tag{3.58}$$

$$= e^{\beta q V_{\text{bi}}}. \tag{3.59}$$

We further assume that $u_V(x_{pd}) \approx u_V^{(0)}(x_{pd})$ by quasi-neutrality. Then plugging Eq. 3.56 into Eq. 3.4 and using $\zeta = \beta \frac{j_V}{\mu_V} \frac{\sqrt{\pi}}{2}$ (Eq. 3.6), $\mu_V = \mathcal{D}_V \beta q$ (Eq. 1), and $\Delta V = V_{\text{bi}} - V_{\text{ext}}$ (Eq. 2.112),

$$u_V(x_{nd}) + \zeta\xi_n \operatorname{erf}\frac{W_n}{\xi_n} = e^{-\beta q\Delta V}\left(u_V(x_{pd}) - \zeta\xi_p \operatorname{erfi}\frac{W_p}{\xi_p}\right) \tag{3.60}$$

$$u_V(x_{nd}) + \zeta\xi_n \operatorname{erf}\frac{W_n}{\xi_n} = e^{-\beta q\Delta V}\left(u_V^{(0)}(x_{nd})e^{\beta q V_{\text{bi}}} - \zeta\xi_p \operatorname{erfi}\frac{W_p}{\xi_p}\right) \tag{3.61}$$

$$u_V(x_{nd}) = u_V^{(0)}(x_{nd})e^{\beta q V_{\text{bi}}}e^{-\beta q\Delta V} - \zeta\left[e^{-\beta q\Delta V}\xi_p \operatorname{erfi}\frac{W_p}{\xi_p} + \xi_n \operatorname{erf}\frac{W_n}{\xi_n}\right] \tag{3.62}$$

$$u_V(x_{nd}) = u_V^{(0)}(x_{nd})e^{\beta q V_{\text{ext}}} - \frac{j_V}{\mathcal{D}_V q}\Lambda \tag{3.63}$$

we obtain the modified statement of the law of the junction, where

$$\Lambda = \frac{\sqrt{\pi}}{2}\left[e^{-\beta q\Delta V}\xi_p \operatorname{erfi}\frac{W_p}{\xi_p} + \xi_n \operatorname{erf}\frac{W_n}{\xi_n}\right] \tag{3.64}$$

We may compare Eq. 3.63 against the usual statement of the law of the junction (Eq. 2.152),

$$u_V(x_{nd}) = u_V^{(0)}(x_{nd})e^{\beta q V_{\text{ext}}} \tag{3.65}$$

At equilibrium, Eq. 3.63 and Eq. 2.152 agree with each other, but they deviate as j_V becomes large. We could ask the question: how high can j_V become before the Λ correction term in $u_V(x_{nd})$ becomes too large? However, Eq. 3.63's $u_V(x_{nd})$ depends on j_V and j_V obviously depends on $u_V(x_{nd})$. This circular dependency makes it difficult to estimate the magnitude of the correction term. In the next section, we break this cycle and derive an error bound on j_V that does not depend on itself.

3.4 $j_V(V_{\text{ext}})$ curve and validity bound for law of the junction

To get an actual current (assuming uniform illumination and zero surface recombination velocity, both of which are optimistic for solar cell efficiency and pessimistic for the correctness of the law of the junction), we use Eq. 2.222

$$j_V(x_{nd}) = q\frac{\mathcal{D}_V}{L_p}\left(u_V(x_{nd}) - u_V^{(0)} - \tau g_0\right)\tanh\frac{x_{nc} - x_{nd}}{L_p} \tag{3.66}$$

$$= q\frac{\mathcal{D}_V}{L_p}\left(u_V^{(0)}(x_{nd})(e^{\beta q V_{\text{ext}}} - 1) - \tau g_0 - \frac{j_V}{\mathcal{D}_V q}\Lambda\right)\tanh\frac{x_{nc} - x_{nd}}{L_p} \tag{3.67}$$

We solve for j_V,

$$j_V\Big(1+\frac{\Lambda}{L_p}\tanh\frac{x_{nc}-x_{nd}}{L_p}\Big)=q\frac{\mathcal{D}_V}{L_p}\big(u_V^{(0)}(x_{nd})(e^{\beta qV_{\text{ext}}}-1)-\tau g_0\big)\tanh\frac{x_{nc}-x_{nd}}{L_p} \tag{3.68}$$

$$j_V=\frac{q\frac{\mathcal{D}_V}{L_p}\big(u_V^{(0)}(x_{nd})(e^{\beta qV_{\text{ext}}}-1)-\tau g_0\big)\tanh\frac{x_{\ell n}-x_{nd}}{L_p}}{1+\frac{\Lambda}{L_p}\tanh\frac{x_{nc}-x_{nd}}{L_p}} \tag{3.69}$$

where we note that the numerator is exactly j_V if using the law of the junction. Then

$$j_V^{\text{exact over depletion region}}=\frac{j_V^{\text{using law of the junction}}}{1+\frac{\Lambda}{L_p}\tanh\frac{x_{nc}-x_{nd}}{L_p}} \tag{3.70}$$

We note that $\tanh(x)\in[0,1]$ if $x\geq 0$. We recall Eq. 3.64 and we also note that $\operatorname{erfi}\frac{W_p}{\xi_p}\leq e^{(W_p/\xi_p)^2}<e^{\beta q\Delta V}$ and that $\operatorname{erf}(x)\in[0,1]$ if $x\geq 0$. Then

$$0\leq\Lambda<\sqrt{\pi}\frac{\xi_p+\xi_n}{2} \tag{3.71}$$

from which it follows that

$$j_V^{\text{exact over depletion region}}\leq j_V^{\text{using law of the junction}}. \tag{3.72}$$

The relative error between the currents yielded by the exact drift-diffusion solution and by the law of the junction is then bounded by

$$0\leq\frac{j_V^{\text{using law of the junction}}}{j_V^{\text{exact over depletion region}}}-1=\frac{\Lambda}{L_p}\tanh\frac{x_{nc}-x_{nd}}{L_p} \tag{3.73}$$

$$<\frac{\sqrt{\pi}}{2}\frac{\xi_p+\xi_n}{L_p} \tag{3.74}$$

Remark 3. The condition $\frac{\sqrt{\pi}}{2}\frac{\xi_p+\xi_n}{L_p}\ll 1$ given by 3.74 is a *sufficient* condition for the validity of the law of the junction, but not by any means a tight bound.

3.5 Estimating validity in typical devices

We can rewrite Eq. 3.74 by substituting in ξ_p,ξ_n.

$$\frac{\xi_p+\xi_n}{L_p}=\frac{\sqrt{\frac{2\varepsilon}{\beta q^2}}\Big(\frac{1}{\sqrt{N_A}}+\frac{1}{\sqrt{N_D}}\Big)}{L_p} \tag{3.75}$$

For extra generality, we now find a bound in terms of $W_{\text{depletion}}$ that does not depend directly on N_A or N_D. We recall from Eq. 2.106 that

$$W_{\text{depletion}} = \sqrt{\frac{2V_{\text{bi}}\varepsilon}{q}(N_A^{-1} + N_D^{-1})} \tag{3.76}$$

$$\begin{aligned}
\frac{\xi_p + \xi_n}{W_{\text{depletion}}} &= \frac{\sqrt{\frac{2\varepsilon}{\beta q^2}}\left(\frac{1}{\sqrt{N_A}} + \frac{1}{\sqrt{N_D}}\right)}{\sqrt{\frac{2V_{\text{bi}}\varepsilon}{q}(N_A^{-1} + N_D^{-1})}} && (3.77)\\
&= \sqrt{\frac{1}{\beta q V_{\text{bi}}}}\frac{\sqrt{N_A} + \sqrt{N_D}}{\sqrt{N_A + N_D}} && (3.78)\\
&= \sqrt{\frac{1}{\beta q V_{\text{bi}}}}\left(\frac{1}{\sqrt{1 + N_A/N_D}} + \frac{1}{\sqrt{1 + N_D/N_A}}\right) && (3.79)\\
&= \sqrt{\frac{1}{\beta q V_{\text{bi}}}}\Bigg(\underbrace{\frac{1}{\sqrt{1 + 1/r}} + \frac{1}{\sqrt{1 + r}}}_{f_1(r)}\Bigg) && (3.80)
\end{aligned}$$

where $r = N_A/N_D$.

The function $f_1(r)$ attains its maximum $\sqrt{2}$ at $r = 1$ over the interval $r \in (0, \infty)$ (see Lemma 2), so we can say

$$\begin{aligned}
\frac{\xi_p + \xi_n}{W_{\text{depletion}}} &= \sqrt{\frac{1}{\beta q V_{\text{bi}}}} f_1(r) && (3.81)\\
&\leq \sqrt{\frac{2}{\beta q V_{\text{bi}}}} && (3.82)
\end{aligned}$$

We rewrite the bound as

$$\begin{aligned}
\frac{j_V^{\text{using law of the junction}}}{j_V^{\text{exact over depletion region}}} - 1 &\leq \frac{\sqrt{\pi}}{2}\frac{\xi_p + \xi_n}{L_p} && (3.83)\\
&= \frac{\sqrt{\pi}}{2}\frac{W_{\text{depletion}}}{L_p}\frac{\xi_p + \xi_n}{W_{\text{depletion}}} && (3.84)\\
&\leq \frac{\sqrt{\pi}}{2}\frac{W_{\text{depletion}}}{L_p}\sqrt{\frac{2}{\beta q V_{\text{bi}}}} && (3.85)
\end{aligned}$$

Typically $W_{\text{depletion}} \ll L_p$, but we instead consider a pessimistic case in which $L_p = 10W_{\text{depletion}}$.

In typical devices $\beta q V_{\text{bi}} \gtrsim 30$, so

$$\frac{j_V^{\text{using law of the junction}}}{j_V^{\text{exact over depletion region}}} - 1 \leq \frac{\sqrt{\pi}}{2} \frac{W_{\text{depletion}}}{L_p} \sqrt{\frac{2}{\beta q V_{\text{bi}}}} \tag{3.86}$$

$$\leq \frac{\sqrt{\pi}}{2} (1/10) \sqrt{\frac{2}{30}} \tag{3.87}$$

$$\approx 0.023 \tag{3.88}$$

which means that using the law of the junction yields at most a 2.3% error in the obtained current.

In a more typical case where $W_{\text{depletion}} \ll L_p$, the error from using the law of the junction would be much less than 1%, showing why it is so widely used.

Chapter 4

Finite element method

The finite element method (FEM) is a computational method for solving systems of PDEs. We introduce the basics of the method here and refer the reader to many deeper introductions [40, 41]. Well-posed systems of PDEs have a domain Ω and boundary Γ. FEM divides Ω into cells, which are generally segments (1D), triangles (2D), or tetrahedra (3D). The solution to the PDEs is approximated by superpositions of local basis functions, which are each nonzero in only a small number of cells. The key difference between the FEM and finite difference methods is that FEM essentially averages over the cells to get the best approximation over a region rather than at individual points. We illustrate its key features by considering a solution of the Poisson equation.

4.1 Poisson equation

We wish to solve the equation

$$\nabla^2 u = f, \tag{4.1}$$

over a domain Ω with boundary Γ, where u is the function we want to find and f is a given function of space. Throughout Chapter 4, we will use u to denote a generic function that we wish to solve for (rather than the carrier concentration).

We consider standard Dirichlet and Neumann boundary conditions (BCs), which specify the value of u and its gradient $\vec{\nabla} u$ at the boundary, respectively. Suppose that the BCs we wish to apply are

$$u\big|_{\Gamma_D} = u_{\mathrm{BC}} \tag{4.2a}$$

$$\vec{\nabla} u \cdot \hat{n}\big|_{\Gamma_N} = g \tag{4.2b}$$

where $\hat{n}$ is the Γ_N surface normal. The Dirichlet condition applies on Γ_D, and the Neumann condition applies on Γ_N. Γ_D and Γ_N cannot overlap.

The core idea of FEM is to take a weighted average of Eq. 4.1 using a family of spatially highly localized weight functions $\{\varphi_i\}$ called *test functions*. This manipulation results in an equation called a *weak form*, imposing a weaker condition on the solution u than the *strong form* Eq. 4.1. To make a weak form, we take the spatial average of Eq. 4.1 with each test function φ_i to give the conditions

$$0 = \int (\nabla^2 u)\, \varphi_i \, \mathrm{d}v - \int f\, \varphi_i \, \mathrm{d}v \tag{4.3}$$

where the integrals are taken over the domain Ω.

For convenience, we introduce the notation $\int_\Omega \equiv \int \mathrm{d}v$, and write

$$0 = \int_\Omega \varphi_i \, (\nabla^2 u) - \int_\Omega \varphi_i \, f. \tag{4.4}$$

FEM seeks a solution that satisfies the weak form and can be written as $u = \sum_j a_j \psi_j$ for a set of basis functions $\{\psi_j\}$, such that $u = u_{\mathrm{BC}}$ on Γ_D. There are many choices of families of basis functions $\{\psi_j\}$, with various mathematical properties. An important property of basis functions is that they are highly localized, i.e., take nonzero values inside only one cell or a small number of adjacent cells. For this example, we will use continuous piecewise polynomials, named "Lagrange" or "Continuous Galerkin" (CG) basis functions, as shown in Fig. 4.1. In most cases (including this example), the family of test functions φ_i in Eq. 4.4 is precisely the family of basis functions ψ_j (without the ones that are nonzero on the Γ_D boundary).

We revisit the first term in Eq. 4.4. As $\vec{\nabla} u$ is discontinuous across cells, $\nabla^2 u$ is made meaningful by integrating by parts, using

$$\int_\Omega \vec{\sigma} \cdot \vec{\nabla} a + \int_\Omega a(\vec{\nabla} \cdot \vec{\sigma}) = \oint_\Gamma \vec{\sigma} a \cdot \hat{n}. \tag{4.5}$$

Then with $\vec{\sigma} = \vec{\nabla} u$ and $a = \varphi_i$ in the first term, the condition becomes

$$0 = \oint_\Gamma \varphi_i \, (\vec{\nabla} u \cdot \hat{n}) - \int_\Omega \vec{\nabla} \varphi_i \cdot \vec{\nabla} u - \int_\Omega \varphi_i \, f \tag{4.6}$$

Noting that the boundary is made up of Neumann and Dirichlet parts and recalling Eq. 4.2b,

the first term of Eq. 4.6 becomes

$$\oint_\Gamma \varphi_i \, (\vec{\nabla} u \cdot \hat{n}) = \int_{\Gamma_N} \varphi_i \, \underbrace{(\vec{\nabla} u \cdot \hat{n})}_{g} + \int_{\Gamma_D} \cancelto{0}{\varphi_i} (\vec{\nabla} u \cdot \hat{n}) \tag{4.7}$$

$$= \int_{\Gamma_N} \varphi_i \, g \, , \tag{4.8}$$

where we eliminated the Dirichlet boundary term by ensuring that the test functions φ_i are zero on the Γ_D boundary. This process yields our final weak form,

$$0 = \int_{\Gamma_N} \varphi_i \, g - \int_\Omega \vec{\nabla} \varphi_i \cdot \vec{\nabla} u - \int_\Omega \varphi_i \, f. \tag{4.9}$$

As the Neumann boundary condition g enters "naturally" through the weak form, it is called a *natural boundary condition*. On the other hand, the Dirichlet boundary condition must be imposed upon the solution space of u directly, and it is called an *essential boundary condition*.

Remark 4. Dirichlet/Neumann BCs aren't always matched up in this way with essential/natural BC application. For the mixed method in Section 5.3.2, the opposite occurs: the Dirichlet BC enters as a natural BC, and the Neumann BC enters as an essential BC.

Plugging $\varphi_i \mapsto \psi_i$ and $u \mapsto \sum_j a_j \psi_j$ into Eq. 4.9, we obtain

$$\begin{aligned} 0 &= \oint_{\Gamma_N} \psi_i \, g - \int_\Omega \vec{\nabla} \psi_i \cdot \vec{\nabla} \Big(\sum_j a_j \psi_j \Big) - \int_\Omega \psi_i \, f \\ 0 &= \sum_j a_j \underbrace{\left[\int_\Omega \vec{\nabla} \psi_i \cdot \vec{\nabla} \psi_j \right]}_{M_{ij}} - \underbrace{\left[\oint_{\Gamma_N} \psi_i \, g + \int_\Omega \psi_i \, f \right]}_{b_i} \end{aligned} \tag{4.10}$$

which forms a sparse[1] linear system for the a_j's, which could be written as $M\vec{a} = \vec{b}$. In the FEM literature, M is often called the "stiffness matrix", and $\vec{b}$ is called the "mass vector". Both M and $\vec{b}$ are evaluated numerically, and the resulting linear system is solved for $\vec{a}$. The essential (Dirichlet) boundary condition Eq. 4.2a is applied by manually forcing the appropriate a_j's to satisfy the boundary values u_{BC}, and removing the corresponding columns j from the matrix, as well as the corresponding rows from M, $\vec{a}$, and $\vec{b}$. Note that the rows removed from the $\vec{b}$ vector are precisely those corresponding to test functions that would have been nonzero on the Γ_D boundary, which ensures the condition used in Eq. 4.7.

Remark 5. Families of basis functions $\{\psi_j\}$ are typically "nodal". That is, each basis function ψ_j is associated to a point in space p_k called a node, such that $\psi_j(p_k) = \delta_{jk}$. This representation

[1] Since the basis functions $\{\psi_j\}$ and $\{\varphi_i\}$ and their derivatives are highly localized.

simplifies the application of essential boundary conditions – if the boundary condition dictates that $u\big|_{p_k} = c_k$, then the coefficient a_k is equal to c_k, and imposes no further restrictions on other coefficients a_j.

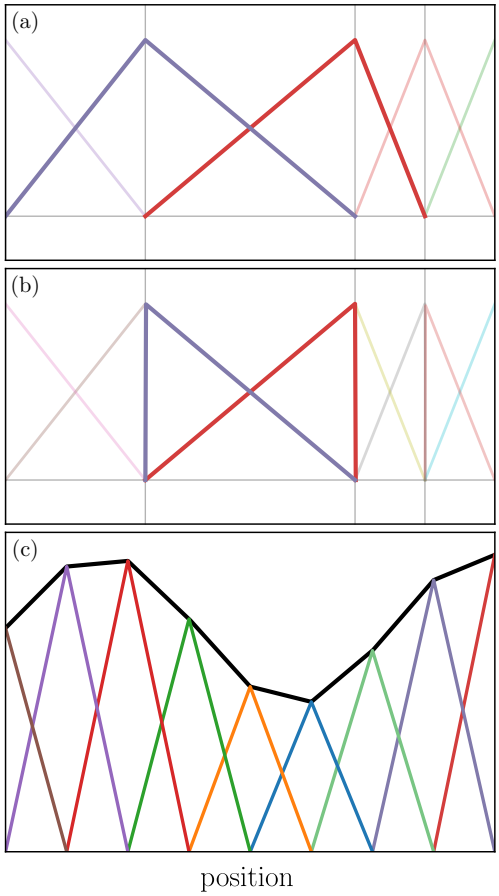

Figure 4.1: Examples of 1-dimensional basis functions, focusing on those that are nonzero in the central cell. Faint lines show basis functions in nearby cells. (a) Continuous Galerkin basis functions of order 1 (CG1), i.e., piecewise linear functions. Note that each basis function is nonzero in two cells. (b) Discontinuous Galerkin basis functions of order 1 (DG1), i.e., piecewise linear functions with no continuity requirement at cell boundaries. Note that each basis function is nonzero only in one cell. (c) Example of reproducing a function as a sum of CG1 basis functions.

4.2 Nonlinear Poisson equation

In the previous section we considered the linear problem $\nabla^2 u = f$, which reduced to a set of linear equations for the a_j. If the RHS were to depend nonlinearly on u, the resulting set of equations

would also be nonlinear. Consider the modified equation

$$\nabla^2 u = (1 + u^2) f, \tag{4.11}$$

and assume the same boundary conditions. Following the procedure as for the standard Poisson equation yields the weak form

$$0 = \int_{\Gamma_N} \varphi_i \, g - \int_\Omega \vec{\nabla}\varphi_i \cdot \vec{\nabla} u - \int_\Omega \varphi_i \, (1 + u^2) f. \tag{4.12}$$

Plugging $u = \sum_j a_j \psi_j$ into Eq. 4.12 yields $-\int_\Omega \varphi_i \Big(\sum_j a_j \psi_j \Big) \Big(\sum_j a_j \psi_j \Big) f$ for the last term, which is nonlinear as it contains a product of a_j's.

There are multiple methods to linearize Eq. 4.12. We sketch the Newton iterative method used in Simudo. The idea is to first replace u by $u_- + \delta u$ in the weak form Eq. 4.12, where u_- is the previous solution and δu is the new dynamical variable being solved for. Then, we pretend that δu is infinitesimal and drop all higher order terms such as $(\delta u)^2$, yielding a linear system in δu. Applying this procedure to the weak form above, we obtain

$$\begin{aligned}
0 &= \int_{\Gamma_N} \varphi_i \, g - \int_\Omega \vec{\nabla}\varphi_i \cdot \vec{\nabla}(u_- + \delta u) - \int_\Omega \varphi_i \left(1 + (u_- + \delta u)^2\right) f \\
&\approx \int_{\Gamma_N} \varphi_i \, g - \int_\Omega \vec{\nabla}\varphi_i \cdot \vec{\nabla} u_- - \int_\Omega \vec{\nabla}\varphi_i \cdot \vec{\nabla}(\delta u) \\
&\quad - \int_\Omega \varphi_i \, (1 + u_-^2) f - \int_\Omega \varphi_i \, 2(\delta u) u_- f
\end{aligned} \tag{4.13}$$

which we solve for δu the standard way. We then update $u_- \leftarrow u_- + \delta u$ and iterate to convergence. Note that inside Simudo, the linearized form Eq. 4.13 is derived automatically using FEniCS symbolic tools.

Chapter 5

Device model paper

5.1 Introduction

As noted in the Statement of original contributions, this chapter contains a reproduction of our published paper describing the device model[1], with minor adjustments.

Device models are essential components of the development of semiconductor devices, from transistors to solar cells to lasers. Standard semiconductor device models, such as Synopsys Sentaurus, treat materials with 0, 1, or 2 bands (i.e., dielectrics, metals, and semiconductors, respectively) along with an electrostatic potential. At a given location in a given material, each band has its own carrier concentration, with particle motion given by diffusion and electric-field-induced drift. Since the electric field itself depends on particle motion, the resulting Poisson/drift-diffusion (PDD) equations are nonlinear and require numerical solution in the general case [42, 43, 26, 44, 45].

A new class of materials, called intermediate band (IB) materials, has been developed over the last 20 years with the goal of improving solar cell efficiency and producing effective infrared photodetectors [9, 17, 46, 47, 48]. These IB materials are like semiconductors except they have an extra band of allowed electronic energy levels above the valence band (VB) and below the conduction band (CB), as shown in Fig. 5.1. Such a band structure permits optical absorption from VB to IB and from IB to CB, which is the key to the increased solar cell efficiency [9]. It is also possible to consider multiple IBs, though such materials have not yet been realized in practice [49].

Where IB devices have been made, they have not generally been highly efficient, which is believed to be largely due to fast nonradiative recombination processes [17, 50, 12, 11, 46, 47, 48, 24]. It has not been possible, however, to perform standard device modeling to optimize these devices, to determine the ideal layer thicknesses, doping levels, etc., since standard semiconductor device models do not allow the possibility of treating a third band. Therefore, we do not know what

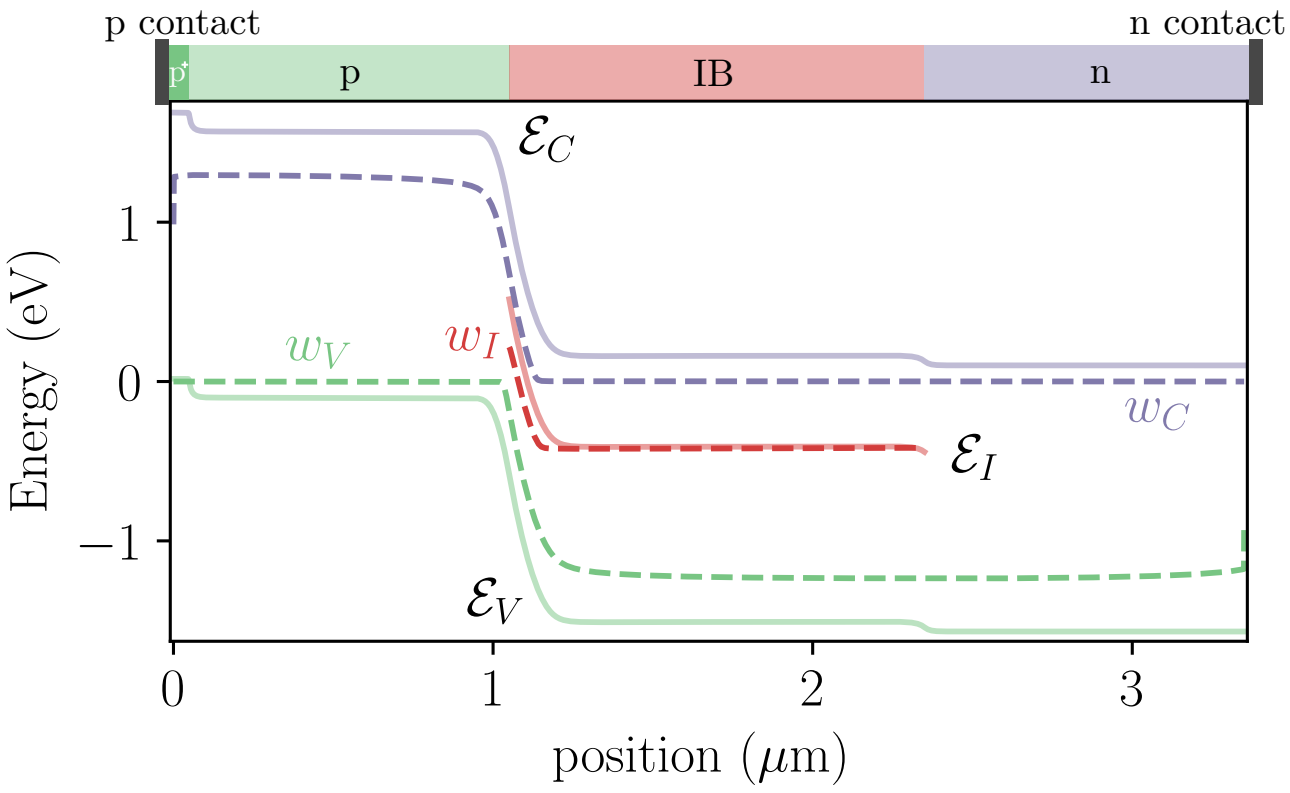

Figure 5.1: Band diagram of illuminated 1D intermediate band device at short circuit. Device structure of p-IB-n device shown at top, and the IB material has an extra band contained entirely inside the semiconductor band gap. Solid lines show the band edge energies $\mathcal{E}_C$, $\mathcal{E}_V$ and the IB energy $\mathcal{E}_I$. Dashed lines show the quasi-Fermi levels w_C, w_V, and w_I. Parameters as in Sec. 5.4.3.

efficiencies existing IB materials could permit, if they were optimized. Interpreting experiments on IB materials and designing the best devices require device modeling capabilities.

In order to describe the basic physics of IB devices, one must be able to describe

1. Optical processes between CB, VB, and IB, with rates dependent on IB filling fraction f,
2. Nonradiative processes between CB, VB, and IB, with rates dependent on f,
3. Carrier transport within the IB,
4. Junctions with standard semiconductors.

There is a large array of standard numerical semiconductor device models based on the coupled Poisson and carrier-continuity equations, including general purpose ones, such as Synopsys Sentaurus and Silvaco, as well as more specialized models such as Crosslight, which includes modelling of quantum well physics and a coupled treatment of carrier-density dependent optics for lasers, and TiberCAD [51]. Nextnano++ also includes features specific to quantum structures and can solve

Table 5.1: Comparison of selected device models

	IB optics with photofilling	IB nonradiative processes	IB transport	Junctions	2D
Sentaurus	N	Y	limited	Y	Y
PC1D [53]	N	Y	N	Y	N
SCAPS [56]	Y	Y	N	Y	N
TiberCAD [51]	N	Y	Y	Y	Y
Martí [58]	N	N	Y	N	N
Strandberg [27]	Y	N	Y	N	N
Tobias [59], Yoshida [60]	Y	N	Y	Y	N
Simudo	Y	Y	Y	Y	Y

an 8-band k.p model self-consistently with the Poisson and drift-diffusion calculations [52]. There are also more focused ones, such as PC1D [53, 54], AFORS-HET[55], SCAPS [56], and Solcore [57], which are 1D models focused on solar cells. Many of these models allow treatment of deep-lying states inside the semiconductor band gap, primarily as Shockley-Read-Hall (SRH) trapping and recombination centers [35]. Sentaurus and PC1D, for example, do not permit optical generation from the deep-lying states. SCAPS does permit both thermal and optical processes, but does not consider transport of carriers inside the defect band. We do not attempt a full characterization of all the available device models, but Table 5.1 shows which of these requirements are met by these device models.

There have been a number of device models developed specifically for IB materials, mostly for solar cells, all in steady state. These include traditional [9, 61, 62, 63]and Boltzmann-approximation [64, 65] detailed balance models, semianalytic models in the drift [66] and diffusion [58, 67, 68] limits, and PDD models [69, 27, 59, 60]. The semianalytic models are specific to either the drift or diffusive limits, while the PDD models allow treatment of IB regions that are neither fully depleted nor fully quasi-neutral. A comparison of the features of the PDD models is also included in Table 5.1. To our knowledge, none has been released as open-source software.

Here we introduce Simudo, a free and open source steady state PDD solver with self-consistent optics for arbitrary numbers of bands. Simudo uses the finite element method (FEM) to solve the coupled Poisson, drift-diffusion, and Beer-Lambert optical propagation equations self-consistently, when necessary including changing f according to local generation and recombination, with associated changes in the optical absorption coefficient. Simudo has built-in radiative recombination, Shockley-Read trapping, and SRH recombination models in the non-degenerate limit and is straightforward to extend to include other models of generation or recombination. All of the band parameters, from energies to mobilities to cross sections, can vary in space or as functions of other

parameters.

Simudo has a number of innovations in its formulation of the problem, described below, and allows high-accuracy simulation of benchmark semiconductor problems while working with 64-bit arithmetic, making it useful both for standard semiconductor simulations as well as for IB devices. It is written in the Python programming language, using the FEniCS platform to solve the FEM problem [29]. It exposes an easy-to-use API for defining problems and extracting results. It is designed for two-dimensional systems and is available for download at https://github.com/simudo/simudo.

Semiconductor device modeling is typically performed using the finite volume method (FVM), which ensures local charge conservation at each cell in the domain [70]. Finite element methods are less common in semiconductor device models, though there are a number of examples [71, 51, 72, 73]. FEM is widely used for related advection-diffusion problems in computational fluid dynamics (CFD) studies [74]. Commercial packages for CFD, as for semiconductor device modeling, are generally based on the FVM method. TiberCAD, a commercial device modeling package with many novel features, uses FEM with continuous basis functions[51]. Finite element methods including discontinuous local basis functions, called discontinuous Galerkin (DG) methods, also permit local charge conservation [74], and they have recently begun to be applied to semiconductor device problems [75, 76]. FEM methods simplify consideration of complicated simulation domains and in theory allow higher-order convergence of solutions, but performance of such methods can only be determined with testing. We use such a DG-FEM method here to produce a general purpose steady state PDD solver capable of treating IB systems, and we show that Simudo realizes the higher-order convergence with mesh size, converging much more rapidly than Synopsys Sentaurus as the mesh spacing is reduced. As shown in Sec. 5.3.5, for a reference pn-diode, Simudo demonstrates quartic self-convergence with mesh density while FVM-based Synopsys Sentaurus demonstrates only quadratic convergence. In the reference problem, Simudo achieves 5-6 digits of convergence with 193 mesh points while Sentaurus requires more than 3000. Simudo provides both a flexible framework for the study of IB devices and also a freely available example of a DG-FEM semiconductor device model.

In Section 5.2, we define the coupled partial differential equations (PDEs) Simudo solves. Section 5.3 describes the heart of Simudo, giving in detail the conversion of the equations of Section 5.2 to the weak forms solved using FEM. This section describes the choices for dynamical variables, the weak forms used for FEM, and how these choices enable Simudo to achieve accuracy despite the problems of finite precision arithmetic. This section concludes with a comparison to Synopsys Sentaurus on a benchmark pn-diode, showing the high quality of Simudo's results. Section 5.4.1 gives examples of setting up a simple problem using the API, including examples of its convenient topology definitions and Section 5.4.2 demonstrates the extensibility of Simudo to

include new physical processes (in this case, Auger recombination). Section 5.4.3 demonstrates the use of Simudo to analyze a system originally studied in [58], showing that its model works better than had been anticipated in the case with equal subgap optical absorption cross sections, but that unequal subgap absorption cross sections produce more complicated phenomena that require IB transport to describe properly.

5.2 Statement of problem

In this section, we describe the mathematical model of the steady state PDD and optical problems we use in Simudo. Carriers both drift in response to electric fields and diffuse. Carriers are generated optically and recombine using a variety of models. The local carrier concentration determines both the electric field and the optical absorption coefficients, so the transport, Poisson, and optical propagation equations are all coupled. Symbols used in this manuscript are summarized in Table 2.

5.2.1 Carrier transport and generation

We consider a CB, a VB, and some number of IBs under the assumption that the carrier population in each band is in local quasi-equilibrium with a temperature T and quasi-Fermi level w_k, where k can be one of $\{C, V, I\}$ for the CB, VB, and IB, respectively. In the case of multiple IBs, k can take values $I_1, I_2, \ldots$, indexing the various IBs, but we simplify the following discussion to consider the case of just one IB, indexed as I.

In the most common approximation of semiconductor device modeling, the carrier dynamics in each band can be described by the drift-diffusion equation and the continuity equation. Letting u_k represent the carrier concentration in band k, u_V and u_C are the hole and electron concentrations, respectively, which we use interchangeably with their standard symbols, p and n. We let $s_k = \pm 1$ give the charge of the carriers in band k, +1 for the VB and -1 for the CB. Then

$$\vec{j}_k = \overbrace{q\mu_k u_k \vec{E}}^{\text{drift}} - \overbrace{s_k q \mathcal{D}_k \vec{\nabla} u_k}^{\text{diffusion}} \tag{5.1a}$$

$$\frac{\partial u_k}{\partial t} = -s_k \frac{1}{q} \vec{\nabla} \cdot \vec{j}_k + g_k, \tag{5.1b}$$

where $\vec{j}_k$ is the current density of carriers in band k, μ_k is the carrier mobility, $\mathcal{D}_k$ is the carrier diffusion constant, $\vec{E}$ is the electric field, q is the elementary charge, and g_k contains all the generation, trapping, and recombination processes (see Section 5.2.2). As stated in Eq. 2.54, for

non-degenerate bands in which w_k is sufficiently far from the band edge $\mathcal{E}_k - q\phi$, we can write

$$u_k = N_k e^{-s_k(w_k+q\phi-\mathcal{E}_k)/k_BT}, \tag{5.2}$$

where N_k is the effective density of states of band k, ϕ is the electrostatic potential, and k_B is Boltzmann's constant.

Then, assuming $\mathcal{E}_k$ is spatially constant,

$$\vec{\nabla} u_k = -s_k N_k e^{-s_k(w_k+q\phi-\mathcal{E}_k)/k_BT} \frac{1}{k_BT} \vec{\nabla}(w_k + q\phi) \tag{5.3}$$

$$= -s_k \frac{u_k}{k_BT} \vec{\nabla}(w_k + q\phi), \tag{5.4}$$

For such nondegenerate bands, the Einstein relation gives $\mu_k = q\mathcal{D}_k/k_BT$, from which Eq. 5.1a gives [43]

$$\vec{j}_k = \mu_k u_k \vec{\nabla} w_k, \tag{5.5}$$

which we use instead of Eq. 5.1a. Equation 5.5 also applies to the case of degenerate bands, as shown in [77], even though the Einstein relation requires a modification. Moreover, Eq. 5.5 applies in the case of spatially-varying band structure (e.g., spatially-varying $N_c, \mathcal{E}_C$)[78], so it is considerably more general than this derivation.

Since an intermediate band is often partially filled, we cannot model it using the non-degenerate approximation of Eq. 5.2. We write $D_I(E)$ for the density of states of the IB, such that $N_I = \int \mathrm{d}E\ D_I(E)$ is the total density of IB states. If the IB has quasi-Fermi level w_I, the electron concentration is

$$u_I = \int \mathrm{d}E\ \frac{D_I(E)}{e^{(E-w_I-q\phi)/k_BT} + 1}. \tag{5.6}$$

If the bandwidth of the IB is narrow relative to k_BT, we can approximate the IB density of states as a Dirac delta $D_I(E) = N_I \delta(E - \mathcal{E}_I)$, and so

$$u_I = N_I \underbrace{\frac{1}{e^{(\mathcal{E}_I-w_I-q\phi)/k_BT} + 1}}_{f_I}, \tag{5.7}$$

where f_I is the filling fraction of the IB, and can be written as $f_I = f(\mathcal{E}_I - w_I - q\phi)$ where $f(E)$ is the Fermi function. We work in this limit for the remainder of this manuscript. Extending beyond this sharp-IB case is not difficult but requires more cumbersome notation.

5.2.2 Carrier generation and recombination

Each band's continuity equation (Eq. 5.1b) has a generation term g_k. This term is the sum of contributions from all generation and recombination processes to the band, which depend on which physical models are included in the simulations. We now specify the details of optical generation g_k^{opt} and a variety of recombination processes r_k, each of which enters either as a negative or positive contribution to g_k, as required for the process.

5.2.2.1 Optical carrier generation

Modeling optical carrier generation requires modeling the changing light intensity through the device. We use a simple Beer-Lambert model for optical propagation and absorption

$$\vec{\nabla}\Phi_{\lambda,\hat{s}} \cdot \hat{s} = -\alpha_\lambda \Phi_{\lambda,\hat{s}} \tag{5.8}$$

where $\Phi_{\lambda,\hat{s}}$ is the photon spectral flux at vacuum wavelength λ and direction of propagation $\hat{s}$ and α_λ is the total absorption coefficient, which can be written as

$$\alpha_\lambda = \sum_{i,f} \alpha_{fi,\lambda},$$

where $\alpha_{fi,\lambda}$ is the absorption coefficient for the optical process at wavelength λ that moves a carrier from band i to band f. In the usual semiconductor case, $\alpha_{VC,\lambda} = 0$ and $\alpha_{CV,\lambda}$ is finite for λ corresponding to energies larger than the band gap. Free-carrier absorption is included in $\alpha_{ii,\lambda}$. The carrier generation rate in band k due to optical processes is then

$$g_k^{\text{opt}} = -s_k \int \mathrm{d}\lambda \left(\sum_i \alpha_{ki,\lambda} - \sum_f \alpha_{fk,\lambda} \right) \Phi_{\lambda,\hat{s}}. \tag{5.9}$$

Further details of the optical propagation model are described in Section 5.2.3.

In nondegenerate bands, there are always enough carriers to excite in or out of a band. That is, the valence band always has electrons available, and the conduction band has empty states available to be filled, so the absorption coefficient $\alpha_{CV,\lambda}$ is insensitive to the free carrier density in the bands. In an IB, however, the VB→IB process requires empty states in the IB while the IB→CB process requires filled states in the IB. To capture this phenomenon, we write

$$\alpha_{CI,\lambda} = \sigma_{CI,\lambda}^{\text{opt}} u_I \tag{5.10}$$

$$\alpha_{IV,\lambda} = \sigma_{IV,\lambda}^{\text{opt}} (N_I - u_I) \tag{5.11}$$

where $\sigma_{fi,\lambda}^{\text{opt}}$ is the optical capture cross section from band i to f at wavelength λ of a single

intermediate state. We can combine these equations into a single expression,

$$\alpha_{kI,\lambda} = \sigma^{\text{opt}}_{kI,\lambda} u_{I,s_k}, \tag{5.12}$$

where $u_{I,-}$ is just u_I and $u_{I,+} = N_I - u_I$ is the number of holes in band I, and α_{kI} is understood to be α_{IV} when $k = V$. Since $\alpha_{fi,\lambda}$ depends on the carrier concentrations, and the carrier concentrations depend on $\alpha_{fi,\lambda}$ (through the generation rate g^{opt}_k), the transport and the optical models feed into each other, so they must be solved in a self-consistent manner.

5.2.2.2 Recombination and trapping

Simudo offers several built-in radiative and nonradiative recombination and trapping mechanisms using the non-degenerate limit for the CB and VB, each including an equivalent thermal generation. An example is the SRH recombination model with a single trap level at energy $\mathcal{E}_I$ [35], in which two trapping processes (of an electron and a hole) produce a recombination event, with recombination rate

$$r^{\text{SRH}} = \frac{pn - n_i^2}{(p + p_1)\tau_n + (n + n_1)\tau_p}, \tag{5.13}$$

where τ_p, τ_n are the carrier lifetimes and p_1, n_1 are the carrier concentrations of holes and electrons, respectively, if their quasi-Fermi levels were equal to $\mathcal{E}_I$. This r^{SRH} appears as a negative contribution to g_k for both CB and VB.

We can model traps as intermediate bands with w_I tracked explicitly, in which case we implement standard Shockley-Read trapping [35],

$$r^{\text{SR}}_{Ik} = \left[1 - e^{s_k(w_k - w_I)/k_B T}\right] f_{I,-s_k} u_k / \tau_k, \tag{5.14}$$

where $f_{I,-s_k}$ is the IB filling fraction of carriers with charge $-s_k$, and τ_k is the Shockley-Read lifetime for band k, as in Eq. 5.13 [35]. Note that r^{SR}_{IC} makes a negative contribution to g_C and a positive contribution to g_I, while r^{SR}_{IV} makes a negative contribution to both g_V and g_I.

Simudo also implements radiative trapping from band $k = C, V$ to I. When we use Boltzmann statistics rather than Bose statistics for the emitted photons, which is valid when $|w_k - w_I|$ remains at least a few $k_B T$ below $|\mathcal{E}_k - \mathcal{E}_I|$, as in Ref. [27], then the radiative trapping can be written

$$r^{\text{rad}}_{Ik} = \left[e^{-s_k(w_k - w_I)/k_B T} - 1\right] u_{I,s_k} \mathcal{I}_{Ik}, \tag{5.15}$$

where

$$\mathcal{I}_{Ik} = \frac{8\pi n_r^2}{h^3 c^2} \int_0^\infty \sigma_{Ik}^{\text{opt}}(E)\, E^2 e^{-E/k_B T}\, \mathrm{d}E, \tag{5.16}$$

where n_r is the index of refraction, and $u_{k,1}$ is either n_1 or p_1 for $k = C, V$, respectively. Note that Ref. [27] includes only the recombination term, and we add the corresponding thermal generation term, which is the -1 in Eq. 5.15. We can re-express Eq. 5.15 in a similar form to the nonradiative terms by using the relation

$$u_{I,s_k} = \frac{u_{I,-s_k} u_k}{u_{k,1}} e^{s_k(w_k - w_I)/k_B T},$$

which follows from Eqs. 5.2, 5.7. Then

$$r_{Ik}^{\text{rad}} = \left[1 - e^{s_k(w_k - w_I)/k_B T}\right] \frac{u_{I,-s_k} u_k}{u_{k,1}} \mathcal{I}_{Ik}. \tag{5.17}$$

As with Eq. 5.2, Eqs. 5.14-5.16 are valid in the non-degenerate limit where w_k does not approach $\mathcal{E}_k$ but full degenerate statistics are used for the IB. Extensions to the degenerate limit can be added, if desired. Simudo also treats standard radiative recombination between conduction and valence bands [79].

We also treat surface recombination at external surfaces Γ of the device, which imposes a boundary condition

$$\vec{j}_k \cdot \hat{n}|_\Gamma = S_k(u_k - u_{k0})|_\Gamma, \tag{5.18}$$

where S_k is the surface recombination velocity of carriers in band k at boundary Γ, $\hat{n}$ is the normal to Γ, and u_{k0} is the carrier concentration at equilibrium [80]. The current release of Simudo supports only $S_k = 0$ or ∞, which impose $\vec{j}_k \cdot \hat{n}|_\Gamma = 0$ or $(u_k - u_{k0})|_\Gamma = 0$, respectively.

5.2.3 Optical equations

For each wavelength, we need to solve the optical propagation according to Eq. 5.8. For stability of the numerical solution, it is convenient to use a second-order equation so that we can apply boundary conditions on both the inlet and outlet boundaries [81]. We take the derivative of Eq. 5.8 with respect to the direction of propagation,

$$\hat{s} \cdot \vec{\nabla}(\hat{s} \cdot \vec{\nabla}\Phi_{\lambda,\hat{s}}) + \hat{s} \cdot \vec{\nabla}(\alpha_\lambda \Phi_{\lambda,\hat{s}}) = 0, \tag{5.19}$$

With no reflection from the back, the boundary conditions are then

$$\Phi_{\lambda,\hat{s}} = \Phi^0_{\lambda,\hat{s}} \qquad \vec{x} \in \Gamma_i \quad \text{(inlet)} \tag{5.20a}$$

$$\hat{s} \cdot \vec{\nabla}\Phi_{\lambda,\hat{s}} + \alpha_\lambda \Phi_{\lambda,\hat{s}} = 0 \qquad \vec{x} \in \Gamma_o \quad \text{(outlet)} \tag{5.20b}$$

where $\Phi^0_{\lambda,\hat{s}}$ is a spectral photon flux at the inlet boundary.

In the case where α_λ is constant for λ in an interval $[\lambda_1, \lambda_2]$, the optical flux at all wavelengths in that range obeys Eq. 5.19 and can thus be treated together. We can write

$$\Phi_{[\lambda_1,\lambda_2],\hat{s}} = \int_{\lambda_1}^{\lambda_2} \Phi_{\lambda,\hat{s}} \mathrm{d}\lambda, \tag{5.21}$$

where $\Phi_{[\lambda_1,\lambda_2],\hat{s}}$ is a photon flux (where $\Phi_{\lambda,\hat{s}}$ is a spectral photon flux). In this case, we have

$$\hat{s} \cdot \vec{\nabla}(\hat{s} \cdot \vec{\nabla}\Phi_{[\lambda_1,\lambda_2],\hat{s}}) + \hat{s} \cdot \vec{\nabla}(\alpha_\lambda \Phi_{[\lambda_1,\lambda_2],\hat{s}}) = 0. \tag{5.22}$$

Simudo uses this form, which allows simple treatment of piecewise constant absorption coefficients with a small number of optical fields $\Phi_{[\lambda_1,\lambda_2],\hat{s}}$. When optical fields with only one propagation direction $\hat{s}$ are considered, we write the spectral flux density Φ_λ and the flux density $\Phi_{[\lambda_1,\lambda_2]}$.

5.2.4 Poisson's equation

In electrostatics, Poisson's equation relates ϕ, the charge density ρ, and the permittivity ε,

$$\vec{\nabla} \cdot (\varepsilon \vec{\nabla}\phi) = -\rho. \tag{5.23}$$

It can also be split into two equations

$$\vec{\nabla}\phi = -\vec{E}, \tag{5.24a}$$

$$\vec{\nabla} \cdot (\varepsilon \vec{E}) = \rho, \tag{5.24b}$$

where $\vec{E}$ is the electric field.

The charge density ρ is the sum of the static charge and the mobile charge in each band. In an IB material,

$$\rho = q[-n + p - N_I(f_I - f_{I,0}) + N_D - N_A], \tag{5.25}$$

where N_A, N_D are the shallow acceptor and donor doping concentrations, respectively, the mobile charge in the IB is $qN_I(f_I - f_{I,0})$, with $f_{I,0}$ the IB filling fraction of the bulk IB material at $T = 0$ K.

For a donor-type IB $f_{I,0} = 1$, and for an acceptor-type IB $f_{I,0} = 0$. Note that in writing the shallow dopant terms N_D and N_A, we are assuming complete ionization of these impurities.

5.3 Numerical method

Simudo uses the finite element method (FEM) to solve the coupled Poisson/drift-diffusion and optics problems, given by Eqs. 5.1b,5.5,5.19, and 5.24. The FEM method divides the simulation domain Ω into cells, which are generally triangles in 2D, and enforces a weak form of the desired PDE's with a set of test functions defined on those cells, with boundary conditions applied on the domain boundary Γ. The method is well-described in many reference texts [82, 41, 40]. In this section, we detail the weak forms used for these coupled equations and the solution method for the resulting nonlinear system. We benchmark Simudo against the industry standard Synopsys Sentaurus commercial simulator on a standard semiconductor problem to show the quality of our results.

5.3.1 Solution method

The PDD problem is a coupled nonlinear system of PDEs, which we solve iteratively using Newton's method as implemented in the FEniCS package. The solution procedure is outlined in Fig. 5.2. The goal is to find a solution $y = (\phi, \vec{E}, w_C, w_I, w_V, \vec{j}_C, \vec{j}_I, \vec{j}_V)$ that satisfies Eqs. 5.1b,5.5,5.24 and associated Φ_λ that obeys Eq. 5.19. The optical problem is solved alongside the PDD problem in a self-consistent manner. That is, the PDD subproblem produces the absorption coefficient $\alpha(\lambda)$ (which, for processes involving the IB, depends on the filling fraction). The optical subproblem is then solved using these absorption coefficients, yielding a new photon flux Φ_λ, which is fed back into the PDD where it enters in the optical carrier generation process, and the cycle iterates until a self-consistent solution is found.

The convergence of Newton's method depends on the quality of the initial guess. Steps 1 and 2 in Fig. 5.2 are the pre-solver, which is used once to make the initial guess for the main Newton solver, illustrated in step 3. The full procedure is:

1. For each point in space, calculate the equilibrium Fermi level E_F at that point assuming local charge neutrality and $\phi = 0$.

   Physically, this step finds the E_F in each location that makes it charge neutral, before any charge is allowed to flow.

2. Determine the built-in potential ϕ_{bi} of the equilibrium system, using $\phi_0 = E_F/q$ as the initial guess. That is, solve only Eq. 5.24 for ϕ while keeping all $w_k = 0$ and thus all $\vec{j}_k = 0$.

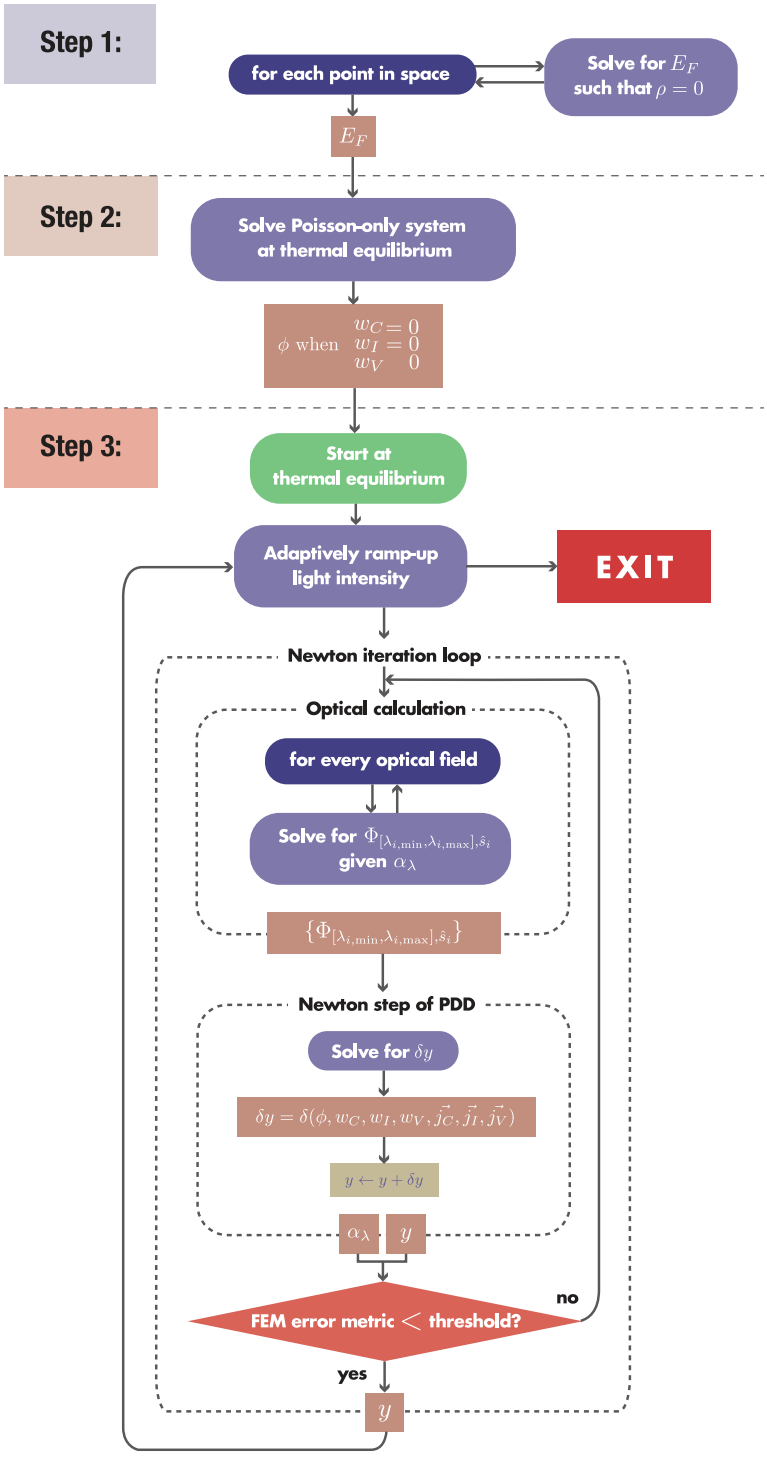

Figure 5.2: Workflow of the numerical method used in Simudo. Steps 1 and 2 are pre-solver steps, which construct an initial guess for the main Newton loop (step 3).

Physically, this step allows charge to move, forming depletion regions as the carriers move to achieve a zero-current configuration that satisfies Poisson's equation. The carrier density inside bulk-like regions of space changes little from the bulk equilibrium value in step 1, making ϕ_0 an excellent guess in large regions of space.

3. **Main solver loop** Adaptively ramp up light intensity and/or bias, starting with thermal equilibrium (dark, no bias). Each solution requires a loop of Newton iterations. Within each Newton iteration, do the following:

   (a) **Optical calculation** For each optical field λ, solve for the photon flux Φ_λ given the latest value of α_λ. Note that Eq. 5.19 is linear when α_λ is fixed.

   (b) **PDD Newton step** Perform one Newton step of the PDD problem.

      i. Solve for $\delta y = \delta(\phi, \vec{E}, w_C, w_I, w_V, \vec{j}_C, \vec{j}_I, \vec{j}_V)$. Use the value of Φ_λ (and thus optical carrier generation) computed in the previous step.

      ii. Update $y \leftarrow y + \delta y$.
      As an option, logarithmic damping can be applied to δy to prevent Newton's method from diverging, e.g., $y \leftarrow y + \text{LogDamping}(\delta y)$ where $\text{LogDamping}(z) = \text{sgn}(z)\log(1 + c|z|)/c$ for $c = 1.72$ or other user-defined value [83].

We now describe the weak forms that we use for each of Eqs. 5.1b, 5.5, 5.19, and 5.23. There is much flexibility in the choice of particular weak forms, all of which can be equivalent to the same strong form. In Section 5.3.3.2 we illustrate the use of a partitioned offset representation for w_k, which allows internal currents to be calculated accurately with double-precision arithmetic.

5.3.2 Poisson equation

Here we introduce the formulation we use to implement Eq. 5.24. We use a mixed method to solve for both ϕ and $\vec{E}$ explicitly [84, 85]. The potential ϕ is represented as a superposition of discontinuous Galerkin (DG) basis functions of order $d_{\text{poisson}} - 1$ (cell-wise discontinuous polynomials), and $\vec{E}$ is represented using Brezzi-Douglas-Marini basis functions of order d_{poisson} (cellwise discontinuous polynomials with continuous normal component on cell boundaries) [84]. We use the BDM space for all vector quantities, including $\vec{E}$ and $\vec{j}_k$. The BDM space is H(div) conforming, meaning the divergence is accurately calculated and fluxes between cells are preserved, which makes it a natural choice for conserved or almost conserved vector quantities.[1] While $\vec{j}_k$ is not a conserved quantity, due to generation and recombination that occur inside of cells, the BDM space ensures that $\vec{j}_k$ is

[1] In the BDM space, the normal fluxes are shared by adjacent elements. The flux exiting the perimeter of a collection of cells exactly equals the sum of fluxes out of each of the cells, with exact arithmetic.

accurately preserved when passing between cells. A method using CG or DG functions for $\vec{j}$ would be susceptible to numerical errors associated with non-conservation of currents between cells, and we show in Sec. 5.3.3.2 that Simudo conserves current well in a pn diode. In the results below, $d_{\text{poisson}} = 2$.

We multiply Eq. 5.24a by test function $\vec{\psi} \in \text{BDM}(d_{\text{poisson}})$ and Eq. 5.24b by test function $v \in \text{DG}(d_{\text{poisson}} - 1)$, then integrate each spatially, giving the weak forms

$$\underbrace{\oint_\Gamma \vec{\psi} \cdot \hat{n}\, \phi_{\text{BC}}}_{\text{natural BC}} - \int_\Omega (\nabla \cdot \vec{\psi})\, \phi + \int_\Omega \vec{\psi} \cdot \vec{E} = 0 \tag{5.26}$$

$$\int_\Omega v \left(\nabla \cdot (\varepsilon \vec{E})\right) - \int_\Omega v\, \rho = 0, \tag{5.27}$$

which must hold for every test function $\vec{\psi}$ and v, where Ω is the full domain and Γ is the boundary. Note that Eq. 5.26 includes an integration by parts. In this case, the electric field BC is an essential BC, imposed by reducing the set of test functions to those that satisfy the BC, while the potential BC is a natural BC.

5.3.3 Transport equations

The drift-diffusion equations are often numerically challenging to solve in semiconductors. In carrier density-based formulations, poor resolution of the gradients of u_k makes linear interpolation of current density unstable, which the Scharfetter-Gummel box method corrects for FVM methods [42]. Additionally, catastrophic cancellation can occur in Eq. 5.1a, e.g., for the majority carrier in a quasi-neutral region of a semiconductor, when the drift and diffusion contributions are nearly equal in magnitude. The current is given by the difference and can be hard to evaluate with finite precision arithmetic. We address these issues by using a quasi-Fermi-level-based representation for carrier density [86, 72]. Calculating $\vec{\nabla} w_k$ in finite precision for Eq. 5.5 can also be challenging when w_k is very flat, and in Sec. 5.3.3.2 we introduce a partitioned offset representation for w_k to allow accurate determination of $\vec{\nabla} w_k$ with essentially no extra computational cost. We use a mixed FEM method that solves explicitly for both w_k and the current density $\vec{j}_k$. As described in Sec. 5.3.2, the BDM space of basis functions enforces local current conservation in the solutions, which also enables local current densities to be well determined. Without the mixed method, local current conservation is enforced indirectly, and we were not able to obtain well-converged results for local currents.

5.3.3.1 Quasi-Fermi level formulation

The quasi-Fermi level w_k is represented as a superposition of DG basis functions of order $d_{\text{transport}} - 1$, and the current density $\vec{j}_k$ is represented using BDM basis functions of order $d_{\text{transport}}$. Section 5.3.2 contains a discussion of these functions' properties and of the mixed method. In the results below, $d_{\text{transport}} = 2$.

We derive weak forms of Eq. 5.1b and Eq. 5.5, multiplying Eq. 5.1b by test function $v \in \text{DG}(d_{\text{transport}} - 1)$, taking the dot product of Eq. 5.5 with the test function $\vec{\psi} \in \text{BDM}(d_{\text{transport}})$, and integrating each equation spatially, giving

$$0 = \int_\Omega v \, \vec{\nabla} \cdot \vec{j}_k - \int_\Omega s_k v \, q g_k \tag{5.28a}$$

$$0 = \int_\Omega \vec{\psi} \cdot \vec{j}_k / (\mu_k u_k) - \underbrace{\oint_\Gamma (\vec{\psi} \cdot \hat{n}) \, w_{k,\text{BC}}}_{\text{natural BC}} + \int_\Omega (\vec{\nabla} \cdot \vec{\psi}) \, w_k, \tag{5.28b}$$

where the second equation was obtained by a further integration by parts.

5.3.3.2 Quasi-Fermi level offset partitioning

As written, Eq. 5.28b still suffers from a form of catastrophic cancellation in its last term, which corresponds to the gradient term in Eq. 5.5. Since $\int_\Omega \vec{\nabla} \cdot \vec{\psi} = 0$ for $\vec{\psi} \in \text{BDM}(d_{\text{transport}})$, the last term is nonzero only if w_k varies within the domain where ψ is nonzero. w_k can be extremely flat, for example in quasineutral regions, which makes this integral hard to calculate with finite arithmetic precision. This difficulty is more apparent in Eq. 5.5: if $\vec{j}_k / \mu_k u_k = \vec{\nabla} w_k$ is small,[2] a representation of w_k that stores its value on the nodes of the mesh (as described in Remark 5)cannot resolve such small changes in w_k across space.

We circumvent this issue by using an offset representation for w_k. The idea is to give each cell in the domain its own (spatially constant) base quasi-Fermi level w_{k0} relative to which the new dynamical variable δw_k is expressed. That is, $w_k = w_{k0} + \delta w_k$ where δw_k is the quantity we actually solve for instead of w_k. Before every Newton iteration step, the w_{k0} of each cell is initialized to the cell average of w_k from the previous iteration. This representation allows small spatial changes of δw_k to be accurately represented, enabling accurate determination of the current.

The last remaining question is how to adjoin regions with different base w_{k0} values. We connect

[2] relative to $|w_k|$/(mesh size)

them by adding a surface integral jump term to Eq. 5.28b, resulting in

$$0 = \int_\Omega \vec{\psi} \cdot \vec{j}_k/(\mu_k u_k) - \oint_\Gamma (\vec{\psi} \cdot \hat{n})\, \delta w_{k,\mathrm{BC}} + \int_\Omega (\vec{\nabla} \cdot \vec{\psi})\, \delta w_k + \underbrace{\sum_{f \in \text{interior facets}} \int_f (\vec{\psi} \cdot \hat{n})\, [w_{k0}]}_{\text{region boundary term}}. \tag{5.29}$$

where $[w_{k0}]$ is the jump operator, which takes the difference between the values of a discontinuous expression on either side of a facet (see Appendix B). The rest of this section is dedicated to deriving that term and comparing the result to a formulation without the offset representation.

We substitute $w_k = w_{k0} + \delta w_k$ into Eq. 5.28b, and we obtain

$$0 = \int_\Omega \vec{\psi} \cdot \vec{j}_k/(\mu_k u_k) - \oint_\Gamma (\vec{\psi} \cdot \hat{n})\, \delta w_{k,\mathrm{BC}} + \int_\Omega (\vec{\nabla} \cdot \vec{\psi})\, \delta w_k \underbrace{- \oint_\Gamma (\vec{\psi} \cdot \hat{n})\, w_{k0,\mathrm{BC}} + \int_\Omega (\vec{\nabla} \cdot \vec{\psi})\, w_{k0}}_{(\star)}. \tag{5.30}$$

Our goal now is to rewrite the $(\star)$ term. Since w_{k0} is constant on each cell K, $\vec{\nabla} w_{k0} = \vec{0}$ within each cell. We integrate by parts using

$$\int_\Omega \vec{\sigma} \cdot \vec{\nabla} a + \int_\Omega a(\vec{\nabla} \cdot \vec{\sigma}) = \oint_\Gamma \vec{\sigma} a \cdot \hat{n}, \tag{5.31}$$

yielding

$$\int_K \vec{\psi} \cdot \underbrace{\vec{\nabla} w_{k0}}_{\vec{0}} + \int_K (\vec{\nabla} \cdot \vec{\psi})\, w_{k0} = \oint_{\partial K} (\vec{\psi} \cdot \hat{n})\, w_{k0} \tag{5.32}$$

Summing over all cells K,

$$\underbrace{\sum_K \int_K}_{\int_\Omega} (\vec{\nabla} \cdot \vec{\psi})\, w_{k0} = \sum_K \oint_{\partial K} (\vec{\psi} \cdot \hat{n})\, w_{k0} \tag{5.33}$$

$$\int_\Omega (\vec{\nabla} \cdot \vec{\psi})\, w_{k0} = \oint_\Gamma (\vec{\psi} \cdot \hat{n})\, w_{k0} + \sum_{f \in \text{interior facets}} \int_f (\vec{\psi} \cdot \hat{n})\, [w_{k0}] \tag{5.34}$$

$$\underbrace{- \oint_\Gamma (\vec{\psi} \cdot \hat{n})\, w_{k0} + \int_\Omega (\vec{\nabla} \cdot \vec{\psi})\, w_{k0}}_{(*)} = \sum_{f \in \text{interior facets}} \int_f (\vec{\psi} \cdot \hat{n})\, [w_{k0}] \tag{5.35}$$

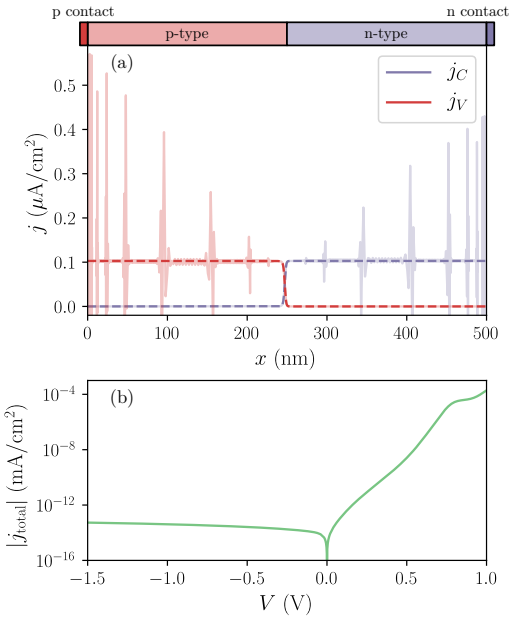

Figure 5.3: (a) Current densities j_C and j_V for a silicon pn-junction at a bias of 0.16 V. Simudo's partitioned offset representation (dashed, dark lines) allows robust determination of internal current densities, while calculations without the offset representation (solid, dull lines) cannot determine the majority current densities accurately. (b) Total charge current density at the contact. Results with and without the offset representation agree to 5 digits of precision.

which, plugged into Eq. 5.30, yields Eq. 5.29.

We perform a test of Simudo, which uses the partitioned offset representation, against the identical model without the offset representation. We consider a standard silicon pn-junction diode with symmetric doping of 10^{18} cm^{-3} and SRH lifetimes of 1 ns and 1 μs in the p- and n-type regions, respectively. Each region has a length of 250 nm, for total device length of 500 nm. Although the problem is one-dimensional, we consider a 2D region with a height of 1 μm. At each contact, the majority carrier has an infinite surface recombination velocity while the minority carrier has zero surface recombination. We use a mesh with 769 points in the x-direction, which is tightest near the contacts and junction and expands out geometrically toward the middle of the quasi-neutral regions. The mesh has 2 points in y-direction, and further details of the mesh are given in Sec. 5.3.5.

The offset representation allows internal current densities to be resolved accurately throughout the device. Figure 5.3(a) shows the electron and hole currents under 0.16 V bias. Without the offset representation, the majority currents are poorly resolved, due to the inability to resolve $\vec{\nabla} w_k$ with double-precision arithmetic. The majority currents in the no-offset model become worse as the mesh density increases (not shown), as expected for approximations of $\vec{\nabla} w_k$. Figure 5.3(b) shows the total charge current density at the contact, and the results with and without the offset

representation agree to 5 digits. We conclude that the offset representation allows robust extraction of internal current densities but does not seem to be important for the overall current density of the test device. The offset representation imposes essentially no extra computational cost on Simudo while enabling robust determination of internal current densities.

5.3.4 Optics

The optical problem is solved by self-consistently iterating through the optical flux variables $\Phi_{[\lambda_{i,\min},\lambda_{i,\max}],\hat{s}_i}$ and independently solving Eq. 5.19 for each one. For convenience, we write $\Phi = \Phi_{[\lambda_{i,\min},\lambda_{i,\max}],\hat{s}_i}$, $\hat{s} = \hat{s}_i$, and $\alpha = \alpha_{[\lambda_{i,\min},\lambda_{i,\max}]}$ for the remainder of this section. We represent Φ using CG basis functions of order $d_{\text{optical}} = 2$.

We now derive the weak form used in Simudo to solve each optical propagation problem. We follow closely the derivation in [81] of the modified second order radiative transfer equation (MSORTE) method, without the scattering matrix. Integrating Eq. 5.19 with a test function $v \in \text{CG}(d_{\text{optical}})$ gives

$$\int_\Omega \vec{v} \cdot \vec{\nabla} \zeta + \int_\Omega \vec{v} \cdot \vec{\nabla}(\alpha\, \Phi) = 0, \tag{5.36}$$

where $\vec{v} = v\, \hat{s}$ and $\zeta = \hat{s} \cdot \vec{\nabla}\Phi$. Using Eq. 5.31, we obtain

$$\oint_\Gamma (\vec{v} \cdot \hat{n})\, \zeta_{\text{BC}} - \int_\Omega (\vec{\nabla} \cdot \vec{v})\, \zeta + \int_\Omega \vec{v} \cdot \vec{\nabla}(\alpha\, \Phi) = 0. \tag{5.37}$$

Inserting the outlet boundary condition Eq. 5.20b into the first term, we obtain the final weak form

$$\oint_\Gamma (\vec{v} \cdot \hat{n})\, (-\alpha\, \Phi) - \int_\Omega (\vec{\nabla} \cdot \vec{v})\, \zeta + \int_\Omega \vec{v} \cdot \vec{\nabla}(\alpha\, \Phi) = 0 \tag{5.38}$$

The inlet boundary condition Eq. 5.20a is applied directly on Φ as an essential boundary condition.

5.3.5 Sentaurus benchmark comparison

To validate Simudo, we benchmark it against the industry standard Synopsys Sentaurus device simulator. Since Sentaurus does not support intermediate band materials, the benchmark is limited to standard semiconductors. Our test problem is the same silicon pn-junction as considered in Section 5.3.3.2, with the overall $J(V)$ as shown in Fig. 5.3(b). See below for a discussion of the differences in implementation of the Ohmic condition between Simudo and Sentaurus.

We study the convergence of the results as the mesh is refined, using a number of mesh points in the x-direction ranging from 45 to 12289, with the same meshes used for both Sentaurus and Simudo. The mesh spacing is nonuniform in the x-direction since the carrier and current densities vary most rapidly near the contacts and the junction. The meshes are generated by splitting the structure into two regions, one extending from the p-contact to the junction and one from the n-contact to the junction. In each region, a mesh spacing d_0 is applied to the cell adjacent to the contact and the cell adjacent to the junction. The mesh spacing increases geometrically toward the center of each region with a growth factor of 1.2. To generate the finest mesh sizes, these cells are further subdivided into 4, 16 or 64 equal parts. There are 2 points in the y-direction. The computational cost generally increases with the number of degrees of freedom rather than with the number of cells in the mesh. Simudo has more degrees of freedom associated with each cell than Sentaurus, due to its higher-order basis functions. For this mesh, Simudo has 36 degrees of freedom per triangle while Sentaurus has 9 per triangle, with 2 triangles per mesh point. In this mesh, each triangle has at least one edge on the boundary of the device, which increases the number of degrees of freedom per triangle compared to a mesh where most triangles share sides; this effect is similar for both methods.

Figure 5.4(a) shows the results of the study, where we plot the relative error in $J(V)$ for each simulation, with the reference current J_{ref} taken from Sentaturus with the densest mesh. With 193 mesh points, Simudo converges approximately as well as Sentaurus with 3073 mesh points. Above 769 mesh points, the Simudo results show no further improvement in error, indicating either that Simudo and Sentaurus converge to results that differ at the 10^{-5} level or that Sentaurus with 12289 mesh points is only converged to the 10^{-5} level, while Simudo may have converged more precisely. The simulations were performed on different machines, so we do not report timing data.

The figure makes clear that Simudo converges much more rapidly with mesh size than Sentaurus, which demonstrates the higher order convergence that FEM is supposed to provide over FVM. Figure 5.4(b) shows the scaling of the errors with the number of degrees of freedom in each simulation, at 0.4 V. The solid blue line shows that Sentaurus' self-convergence scales like N^{-2} with the number of degrees of freedom. The dashed green line shows Simudo's self-convergence, with J_{ref} taken from the Simudo simulation with 12289 mesh points. It shows that Simudo's self-convergence scales like N^{-4} with the number of degrees of freedom. For all but the smallest meshes, Simudo's convergence is superior to Sentaurus' at the same number of degrees of freedom. Taken together, these figures show that with 193 mesh points, Simudo's result is as good as Sentaurus' with 64 times as many mesh points, which is equivalent to 16 times as many degrees of freedom.

Note that the boundary conditions at the contacts are not precisely the same for the Sentaurus and Simudo simulations. Both are intended to simulate Ohmic contacts for the majority carrier and surface recombination velocities of 0 for the minority carrier. The Simudo simulations are

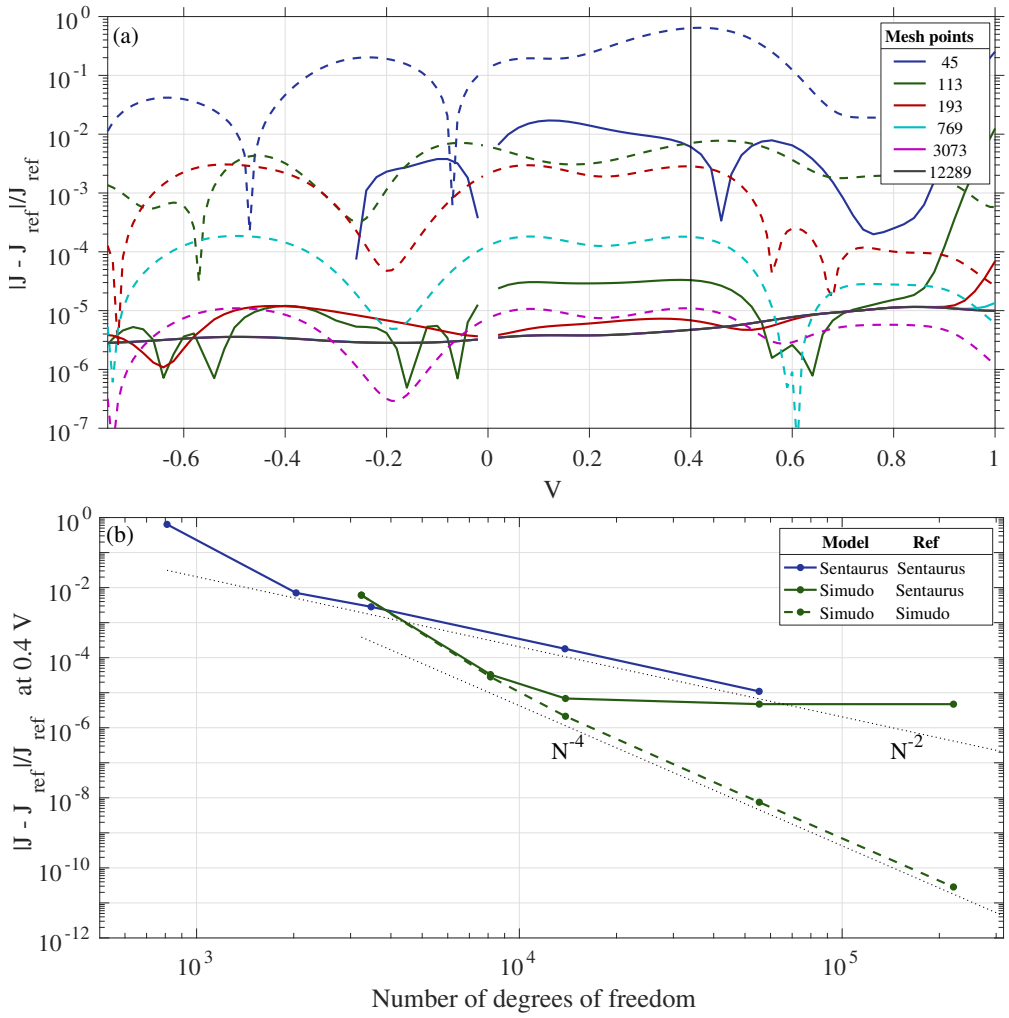

Figure 5.4: (a) Relative error in total current J for a test pn-junction in the dark, at several mesh densities. Dashed lines show results from Sentaurus, and solid lines show results from Simudo, with the same meshes. The Sentaurus result with the finest mesh is taken as J_{ref} for calculating relative error. Points at $V = 0$ are removed to avoid dividing by zero. (b) Relative errors at 0.4 V plotted against the number of degrees of freedom in each simulation. Solid lines use Sentaurus results for J_{ref}, and the dashed line shows Simudo self-convergence. Dotted lines show the scaling trends of the two methods.

performed with surface recombination velocity $S = \infty$ and 0 for the majority and minority carriers, respectively, imposing equilibrium carrier concentration at the boundary for majority carriers and setting $\vec{j}_k \cdot \hat{n}|_\Gamma = 0$ for minority carriers, as described in Sec. 5.2.2.2. The Sentaurus simulations are performed with $S = 0$ for the minority carriers, in agreement with Simudo, and the default "Ohmic contact" boundary condition for the majority carriers, which imposes charge neutrality and equilibrium carrier concentration at the contact. Under small and reverse bias, these two sets of boundary conditions should be equivalent, but under large forward bias, the default Sentaurus boundary condition is expected to give incorrect results due to its imposition of charge neutrality [87]. Sentaurus provides a "Modified Ohmic" boundary condition, which should be closer to the Simudo boundary condition, but we were unable to attain convergence using it. As a result, at larger biases the Simudo and Sentaurus results diverge from each other, and we do not include them in Fig. 5.4. For biases larger than 1 V, the diode is in high injection, and the Boltzmann approximation used in this calculation is not accurate, regardless.

5.4 Examples and results

In this section, we give examples of using Simudo. Section 5.4.1 shows how to set up a simple 1-dimensional pn-junction device and demonstrates the helpful tools that Simudo provides for defining regions and boundaries. Section 5.4.2 shows the extensibility of Simudo by illustrating the code required to add a new Auger recombination process. Section 5.4.3 illustrates the use of Simudo to study a system first considered in [58].

5.4.1 pn junction and topology definitions

We include in the supplementary material the code listing `equilibrium.py` describing a simple pn-junction device in Simudo. This example constructs the device and implements steps 1-4 of the pre-solver shown in Fig. 5.2. Here, we discuss some of the pieces of that code and illustrate the useful topology construction operations built in to Simudo.

In the 1-dimensional pn-junction example, the object `ls` contains information about the layers, including their sizes, positions, and mesh. The object `pdd` sets up the Poisson/drift-diffusion solver and has information about the bands in each material, including recombination processes and boundary conditions. In this example, there are only two bands (VB, CB); for a problem including an IB, `pdd` would have a third band, too.

Simudo is designed for 2-dimensional simulations, and it has sophisticated tools to define the arrangement of materials, dopings, contacts, meshing regions, or other user-defined spatial properties. In many FEM solvers, interfaces must be tracked manually, including their orientation,

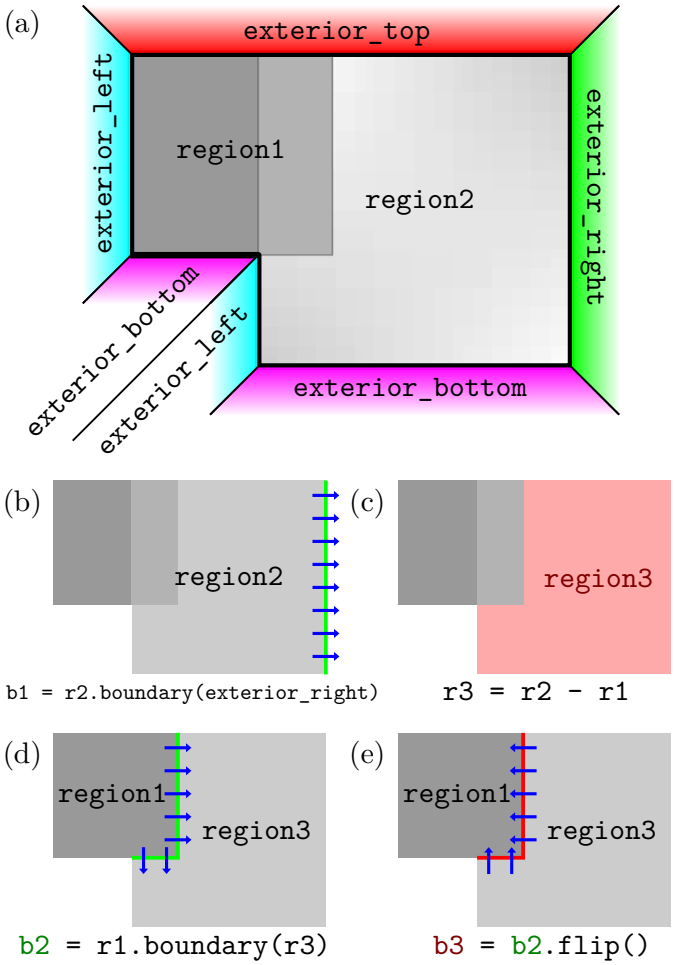

Figure 5.5: Simudo allows users to define complicated domains and regions within them, with useful tools to automatically keep track of the required cells and facets in each region. (a) A simulation domain divided into overlapping `region1` and `region2`. Simudo automatically defines the `exterior` outside the simulation domain, and provides the helper regions `exterior_left`, `exterior_right`, `exterior_top`, and `exterior_bottom`, which may not be helpful, depending on the shape of the domain. Even before the regions are given specific geometries in the simulation domain (e.g., those pictured in (a)), Simudo's topology tools allow construction of further regions and facets. (b-e) Derived regions and facets from the domain shown in (a), with the Simudo command to produce them shown underneath, where we shorten `regionX` to `rX`. Further example code is shown in Fig. 5.6. (b) The contact `FacetRegion` `b1` is the signed boundary from `region2` to `exterior_right`, with the sign indicated by the arrows. It can be used for determining the current flow out of the device. (c) `region3` is `region2` with `region1` removed. (d) The `FacetRegion` `b2` is the signed boundary from `region1` to `region3`. (e) The `flip` operation reverses the sign of the boundary.

```
from simudo.mesh import CellRegions, FacetRegions
R = CellRegions() ; F = FacetRegions()

# R.region1 is created automatically when referenced
print(R)
print(R.region1, type(R.region1))
print(R)
## CellRegions({})
## CellRegionByName('region1') <class 'simudo.mesh.topology.CellRegionByName'>
## CellRegions({'region1': CellRegionByName('region1')})

# Define new region as set difference of region2 and region1
R.region3 = R.region2 - R.region1
# The method (R.region1).boundary(R.region2) creates the signed boundary from
   region1 into region2, e.g., for calculating flux. R.region2 is created when
   referenced.
F.b1 = (R.region2).boundary(R.exterior_right)
F.b2 = (R.region1).boundary(R.region3)
F.b3 = (F.b2).flip()
```

Figure 5.6: Example of using `CellRegions` and `FacetRegions` to construct regions corresponding to those shown in Fig. 5.5. The `R` and `F` objects are containers for cell and facet regions respectively. They are both initialized empty, and regions in each one are created when referenced. The `boundary` method in `(R.region1).boundary(R.region3)` creates the signed boundary from `region1` to `region3`. The `flip()` method gives the boundary with the opposite sign. The actual mapping of these regions into the domain occurs in the mesh generation.

to ensure that integrals over those interfaces are added together properly. Simudo introduces a set of topology tools that instead allow users to define the regions and interfaces in which they are interested, and Simudo takes care of all the bookkeeping. The user defines regions as desired (e.g., emitter, base, defective-region), which can then be given properties, whether they be doping levels, recombination parameters, or other desired properties. These regions are initially defined abstractly, without having any coordinates in the device, using `CellRegions` and `FacetRegions`, and are later connected to geometry and materials by the mesh generator.

Full details are given in the documentation accompanying Simudo, but we give a further illustration of these methods in Figs. 5.5 and 5.6. That example illustrates the creation of arbitrary `CellRegion` objects, including unions and intersections, and edges that connect them. When `R` is a `CellRegions` container, accessing a nonexistent attribute (such as `R.domain`) causes its creation. The user can define new `CellRegion` objects by applying Boolean operations on previous ones and new `FacetRegion` objects by using the `boundary` method. For example, consider the region `R.region1`. Then
`R.region1.boundary(R.region2)` creates a signed boundary from `region1` to `region2`, as illustrated in Fig. 5.5. All of these custom regions are kept as symbolic expressions and evaluated by Simudo only when needed (e.g., when asked to apply a boundary condition or when asked to compute a volume or surface integral). This layer of abstraction allows the user not to worry about the details of mesh markers, entity indices, and facet orientations [88], and is described more fully in the documentation that accompanies Simudo.

The examples in Figs. 5.5-5.6 illustrate another useful concept. The mesh generation interprets the `external` region as being outside the simulation domain, allowing convenient definitions for boundary conditions and current flow. The `FacetRegions` are used in the pn-junction example shown in the supplementary material to define the boundary conditions, which – in step 2 – are conductive at the left and right contacts and nonconductive at the top and bottom surfaces. That example also shows how the mesh can be refined by adding extra mesh points near the contacts.

5.4.2 Extensibility: Adding Auger recombination

The initial release of Simudo contains radiative and Shockley-Read trapping and recombination processes in the non-degenerate limits for VB, CB. The user can easily add modified physics to their problems, which we demonstrate here with an example of adding an Auger recombination process to Simudo, with the form

$$U_A = C_n(n^2 p - n_0^2 p_0) + C_p(p^2 n - p_0^2 n_0), \tag{5.39}$$

where C_n, C_p are the Auger coefficients, and p_0 and n_0 are the hole and electron concentra-

```
from simudo.physics import (DarkEOPMixin, TwoBandEOPMixin,
   ElectroOpticalProcess)

class AugerRecombination(DarkEOPMixin, TwoBandEOPMixin, ElectroOpticalProcess):
  """Auger recombination"""
  name = 'Auger'

  def get_Auger_C(self, band):
    return self.pdd.spatial.get('/'.join((
           self.name, band.name, 'C')))

  def get_generation_user(self, band):
    # In this case, sign is +1 for CB, VB and None for all other bands
    sign = self.get_band_generation_sign(band)
    if sign is None:
      return self._zero_generation

    CB = self.dst_band ; C_n = self.get_Auger_C(CB)
    VB = self.src_band ; C_p = self.get_Auger_C(VB)

    n = CB.u ; n0 = CB.thermal_equilibrium_u
    p = VB.u ; p0 = VB.thermal_equilibrium_u

    r = (C_n*(n**2 * p - n0**2 * p0) +
         C_p*(p**2 * n - p0**2 * n0))

    return (-r) * sign  # Contribution to generation rate

...

spatial.add_rule('Auger/CB/C', R.domain,
     U('1.1e-30 cm^6/s'))
spatial.add_rule('Auger/VB/C', R.domain,
     U('0.3e-30 cm^6/s'))
```

Figure 5.7: Example Simudo code that implements Auger recombination (Eq. 5.39), including the `AugerRecombination` class. The last four lines show the Auger material parameters being set throughout the domain.

tions at thermal equilibrium, respectively [79]. The code is listed in Fig. 5.7. The function `get_generation_user(band)` adds a negative local generation rate in the CB and VB and returns 0 for all other bands. This recombination process moves particles between two bands, the `src_band` and the `dst_band`. In this case, where the electrons and holes have opposite charge, the Auger process destroys both particles simultaneously; if both carrier types involved in the process had the same charge (e.g., for a CB-to-IB trapping process), the process would represent a particle-number-conserving transfer (rather than a recombination) from the `src_band` to the `dst_band`, with the appropriate sign for the recombination process determined by the `get_band_generation_sign` method (inherited from `TwoBandEOPMixin`). This method's sign convention is that the `dst_band` always gains carriers through the generation process, while the `src_band` gains or loses as required by conservation of charge.

5.4.3 P[IB]N junction

In Ref. [58], the authors consider a quantum-dot-based IB solar cell with a p-n-IB-p-n structure. They present a drift-diffusion model for the IB region only, with the carrier density and current density boundary conditions obtained from a depletion approximation and law of the junction. This model assumes that transport is diffusion-dominated in the IB region, and drift can therefore be neglected. This early device model gave important insights into the behavior of IB devices.

Table 5.2: Parameters modeled on the device from [58], for Figs. 5.8–5.10.

Value	**Definition**
$\mathcal{E}_C = 1.67\,\text{eV}$	Conduction band edge energy
$\mathcal{E}_I = 1.10\,\text{eV}$	Intermediate band energy
$\mathcal{E}_V = 0\,\text{eV}$	Valence band edge energy
$N_C = N_V = 5 \times 10^{18}\,\text{cm}^{-3}$	CB and VB effective density of states
$N_I = 10^{17}\,\text{cm}^{-3}$	IB density of states
$\mu_C = \mu_V = 2000\,\text{cm}^2/\text{V/s}$	CB and VB mobility
$\mu_I = 0.001 - 300\,\text{cm}^2/\text{V/s}$	IB mobility
$\alpha_{CV} = 10^4\,\text{cm}^{-1}$	Absorption coefficient for CV process
$\sigma_{CI}^{\text{opt}} = (2-10) \times 10^{-13}\,\text{cm}^2$	Optical cross section for CI process
$\sigma_{IV}^{\text{opt}} = 2 \times 10^{-13}\,\text{cm}^2$	Optical cross section for IV process
$\varepsilon = 13\varepsilon_0$	Dielectric constant
$T_s = 6000\,\text{K}$	Sun temperature
$T_c = 300\,\text{K}$	Cell temperature
$X = 1000$	Solar concentration factor
$w_{\text{IB}} = 1.3\,\mu\text{m}$	IB region length
$f_{I,0} = 1/2$	Charge-neutral IB filling fraction

In testing the self-consistency of the model, the authors estimate the IB mobility required to

remain in the diffusion-dominated regime, finding that an IB mobility greater than $62\,\mathrm{cm}^2/\mathrm{V/s}$ is required to make their model consistent. This claim raises an immediate question: does something interesting happen when the IB mobility goes below that threshold? Since Simudo is a full drift-diffusion device model, we can directly answer that question.

We model a similar device with a simpler p-IB-n structure. This device has the same band and absorption parameters as the device described in [58], summarized in Table 5.2. The incident light is a blackbody spectrum at 6000 K with a solar concentration factor $X = 1000$. The device has equal subgap optical cross sections and nearly current matched (within 10%) incident photon fluxes for the two subgap transitions. In this case, the local IV and CI generation rates are nearly identical throughout the IB device. The code to set up this problem is included in the supplementary material, in `marti2002.py`.

We simulate this device with μ_I ranging from 0.001 to $300\,\mathrm{cm}^2/\mathrm{V/s}$, with resulting J-V curves shown as solid lines in Fig. 5.8a, which is tightly zoomed and still shows only minor effects of this over-10^5 change in μ_I. In fact, the IB and drifts currents contribute negligibly to the transport inside the IB region and the current remains diffusion-dominated throughout, as shown in Fig. 5.9a.

The device behavior is approximately independent of μ_I because the IV and CI generation rates are roughly equal at each point in the device, so IB transport is not required to enable sub-gap current matching, but this behavior is not generic for all IB devices. We illustrate this effect by increasing $\sigma_{CI}^{\mathrm{opt}}$ by a factor of five (while keeping $\sigma_{IV}^{\mathrm{opt}}$ unchanged). In this case, the device is still globally approximately current-matched, but the CI absorption process occurs preferentially at the top of the device while the IV generation occurs deeper in the IB region. The device thus relies on IB transport for the CI and IV generation rates to balance over the full device. These effects are shown in the dashed curves of Fig. 5.8, which show a stronger dependence on μ_I than in the matched case. When IB mobility is low, the excess CI generation in the front of the device instead causes local CI trapping, with equivalent local IV trapping toward the back of the device, reducing overall current. In the high-mobility case, the overall current is slightly larger with the mismatched absorptions, due to the increased optical depth. Figure 5.8b shows how J_{sc} varies with μ_I in both of these cases, where the greater dependence on μ_I in the mismatched case is apparent. Figure 5.9b shows that in the matched-σ^{opt} case, j_I is never particularly large, while it grows to be five times larger in the mismatched case, showing the role of IB currents in internally balancing the optical absorptions.

In the low-mobility limit, where j_I is always small, when local CI and IV current generations are imbalanced, the filling fraction f of the IB must shift to equalize generation and recombination at each point [89]. This effect is visible in Fig. 5.10, where at low mobility, the mismatched-σ^{opt} case has photodepletion at the front side and photofilling at the back side, consistent with excess CI generation at the front and excess IV generation at the back. Both the matched-σ^{opt} and the

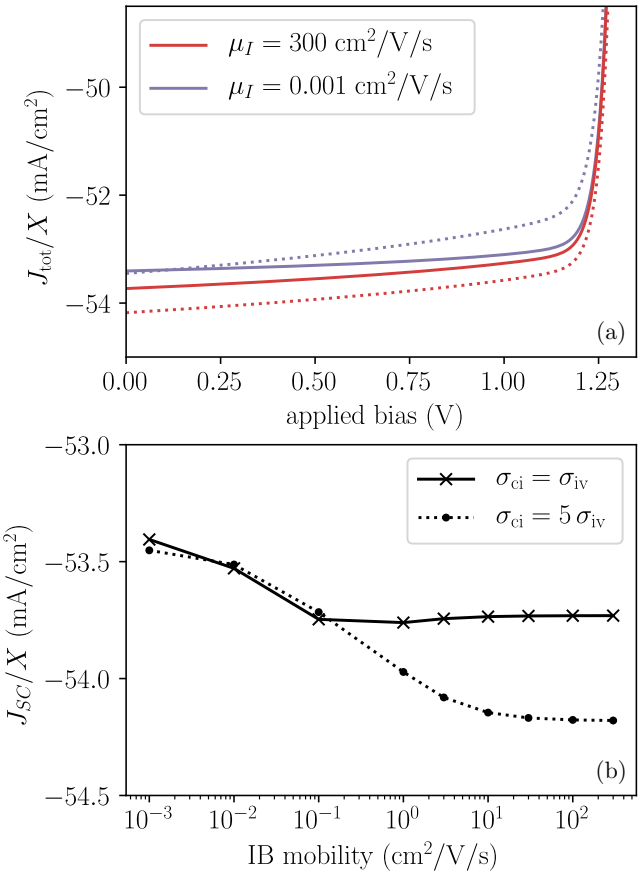

Figure 5.8: (a) $J(V)$ curves for devices with parameters of Table 5.2, under $X = 1000$ suns illumination, modeled on Ref. [58]. Solid lines show $\sigma_{CI}^{\rm opt} = \sigma_{IV}^{\rm opt}$, as in [58], while dotted lines show $\sigma_{CI}^{\rm opt} = 5\sigma_{IV}^{\rm opt}$, which causes the CI absorption to be preferentially at the top of the device. Note the small vertical scale. The mismatched-$\sigma^{\rm opt}$ case is more strongly influenced by the IB mobility μ_I, but the effects are relatively small throughout. (b) The short-circuit current J_{sc} for devices with varying μ_I shows that the matched-$\sigma^{\rm opt}$ case is independent of μ_I when μ_I is sufficiently large ($\gtrsim 0.1$ cm^2/Vs), while the mismatched-$\sigma^{\rm opt}$ case again shows a stronger μ_I-dependence, but note the small vertical scale.

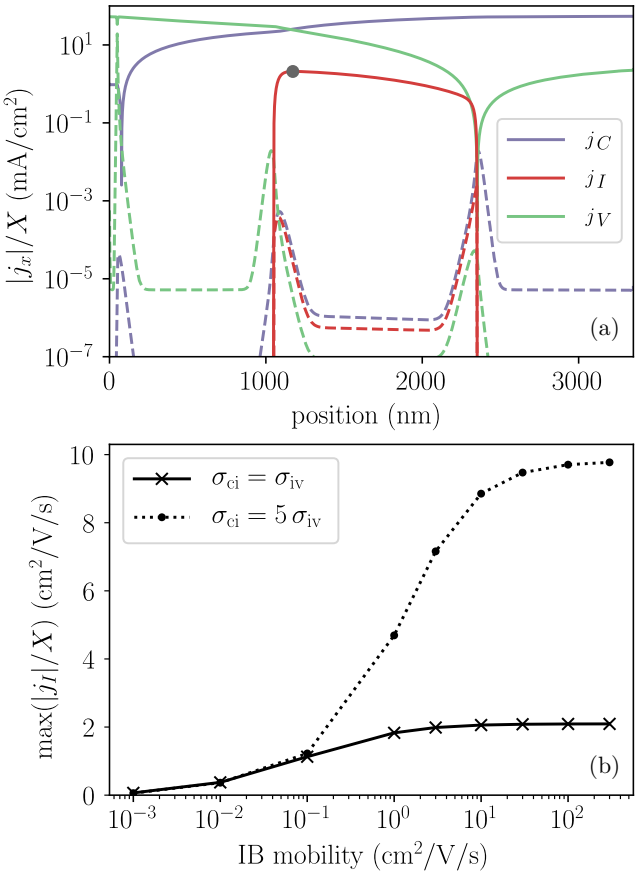

Figure 5.9: (a) Drift (dashed) and diffusion (solid) currents for each band, at $\mu_I = 100\,\mathrm{cm}^2/\mathrm{V/s}$ and $\sigma_{CI}^{\mathrm{opt}} = \sigma_{IV}^{\mathrm{opt}}$, with other parameters as in Table 5.2. The IB current density attains its maximum at the marked point. (b) Maximum IB current density as a function of μ_I for the matched and unmatched absorption cross sections, showing the increased importance of j_I and thus μ_I in the case where local absorptions are mismatched.

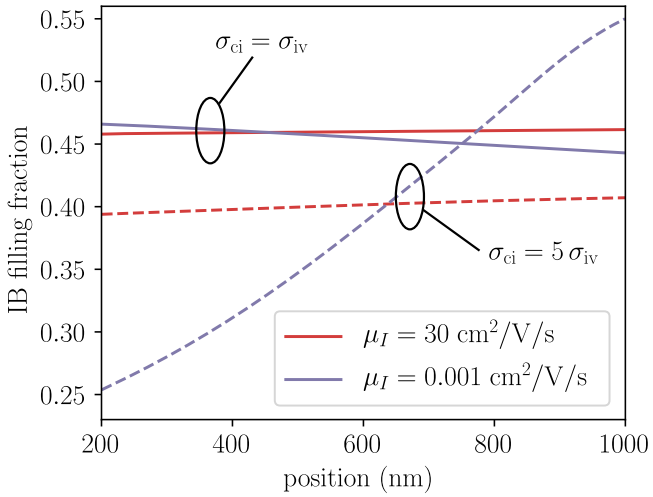

Figure 5.10: IB filling fraction $f(x)$ at the maximum power point with the parameters as in Table 5.2. With mismatched σ^{opt} and low μ_I, the IB filling fraction changes drastically through the depth of the device, as the local generation and recombination rates must come into balance; this balancing increases the local recombination and decreases the total current generation, as seen in Fig. 5.8b. With matched σ^{opt}, internal IB currents balance the absorptions, and $f(x)$ remains nearly constant.

high-mobility mismatched-σ^{opt} cases maintain an approximately uniform IB filling fraction.

These examples together show the utility of Simudo to explore the performance of IB devices and resolve an assertion made in earlier device models without the benefit of a coupled PDD/optics solver.

5.5 (Paper) Conclusion

The availability of a device model for intermediate band materials should enable both understanding of this new class of materials and optimization of IB devices. Simudo's use of the FEM and its methods for overcoming catastrophic cancellation may also prove useful in standard semiconductor device simulation. Simudo has been validated against Synopsys Sentaurus for standard semiconductor devices and shown to converge more rapidly with mesh size. This self-consistent solution of the Poisson/drift-diffusion and optical propagation equations provides a platform for studying a wide range of optoelectronic materials and devices, including solar cells and photodetectors, with tools to enable extensibility to arbitrary generation and recombination models, thermal effects, and more. The near-term roadmap for Simudo includes explicit heterojunction support and nonlocal tunneling, which will be available with future releases at github.com/simudo/simudo. We hope that the free and open source nature of this software will enable further development of IB materials and device simulation more broadly.

Acknowledgments

We acknowledge funding from US Army Research Laboratory (W911NF-16-2-0167), the Natural Sciences and Engineering Research Council of Canada TOP-SET training program, and computing resources from Compute Canada. We thank Emily Zinnia Zhang for alpha testing Simudo, contributing the first code implementing trapping processes, and valuable conversations.

Chapter 6

Record of failures

Note. This section is original content for the thesis.

The coupled nonlinear PDE's described in the previous chapters are frequently challenging to solve. Along the way to the solution presented in chapter 5, we tried a number of other (sometimes simpler) formulations of the problem. We record here those failed attempts and a brief discussion of why they failed. Hopefully our suffering can be educational (and if not, at the very least entertaining).

6.1 Standard density method failure – oscillations

The most direct way of applying the finite element method is to simply substitute Eq. 5.1a into Eq. 5.1b and discretize the resulting equation

$$\vec{j}_k = \overbrace{q\mu_k u_k \vec{E}}^{\text{drift}} - \overbrace{s_k q \mathcal{D}_k \vec{\nabla} u_k}^{\text{diffusion}} \tag{6.1}$$

$$0 = \frac{\partial u_k}{\partial t} = -s_k \frac{1}{q} \vec{\nabla} \cdot \underbrace{\vec{j}_k}_{} + g_k \tag{6.2}$$

$$0 = -s_k \frac{1}{q} \vec{\nabla} \cdot \overbrace{(q\mu_k u_k \vec{E} - s_k q \mathcal{D}_k \vec{\nabla} u_k)}^{} + g_k \tag{6.3}$$

Multiply by a test function v, integrate $\int_\Omega$, and voilà. Just solve for u_k and $\vec{E}$! It's simple, straightforward, and it doesn't work. The author spent more than a few weeks trying and failing miserably to get the numerical system to converge at all. What we call a drift-diffusion equation is more generally known as an advection-diffusion-reaction equation. There's a lot of literature [90, 42, 91] on why the standard CG1 elements numerical approach to these equations is bound to fail due to spurious oscillations in drift-dominated regions ("high Péclet number"). A highly simplified explanation is that FEM functions are usually linear or polynomial as a function of space, while

the advection-diffusion problems can be exponential, so a naive finite element approach would require a prohibitively fine mesh to tame the small-scale exponential nature of these problems. Many researchers have described methods to remedy the issue: artificial diffusion [71], upwinding [92], mixed elements [93], the (modified) finite volume method [90], and many others [94]. Note that these oscillations are *not* caused by numerical loss of precision; they would be present even if we used infinite-precision arithmetic. A better approach is required.

6.2 Density-mixed method failure – catastrophic cancellation

Our next attempt is a mixed method, with the mixed variables being the carrier density u_k and the current $\vec{j}_k$. On running this model, the resulting carrier density u_k looks like what we expect from the depletion approximation (and matches up with the benchmark result from Sentaurus reasonably well), but the current is extremely noisy in the regions where the carrier is the majority carrier (e.g., conduction band electrons in the n-type region). The noise is significant but presents itself as random oscillations *around* the correct value at forward bias (e.g., Figure 6.1). However, the noise completely drowns out the signal at reverse bias as shown in Figure 6.2. We in fact considered an entire family of solvers at this stage, including some desperate attempts involving streamline upwind Petrov–Galerkin (SUPG) [95], but the noise would not go away.

We believe this phenomenon is due to catastrophic cancellation in the drift-diffusion equation (Eq. 5.1a), where the magnitude of the total current is much smaller than the magnitudes of its drift and diffusion components (which have opposite sign and nearly cancel out inside the depletion region and part of the majority quasi-neutral region). Figure 2.6 showed a mild example of catastrophic cancellation where $|j_C|$ is "only" 6 orders of magnitude smaller than $|j_{C,\text{drift}}|$ and $|j_{C,\text{diffusion}}|$. With higher doping and (especially) small $|V_{\text{ext}}|$ it is easy for the drift and diffusion components of current to become over 16 orders of magnitude larger than the current itself. In such a case it is impossible to try to resolve j_C since 64-bit IEEE floats can only carry 16 digits of precision. A better approach is required.

6.3 qfl-mixed method failure – catastrophic cancellation redux

The next attempt is to rephrase the mixed method so that the carrier density is expressed as a quasi-Fermi level w_k instead of carrier concentration u_k. This replacement simplifies the drift-

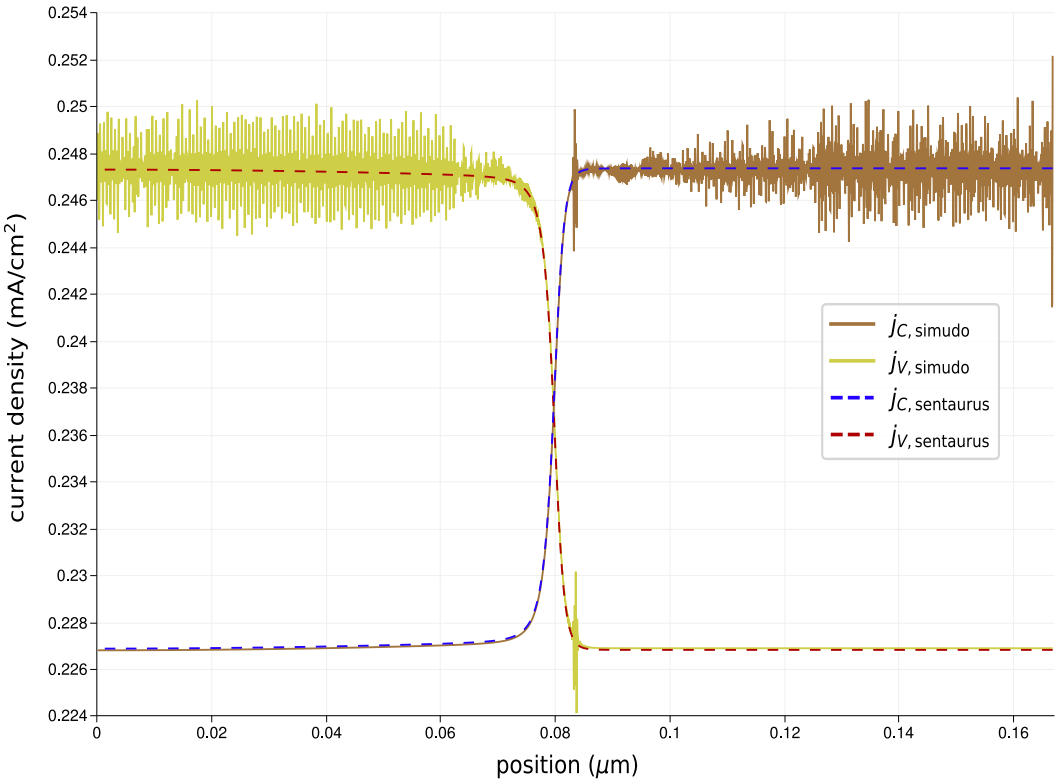

Figure 6.1: Comparison of currents in a silicon pn junction between the old Simudo density-mixed method and Synopsys Sentaurus. The diode has symmetric doping $N_A = N_D = 10^{18}\,\text{cm}^{-3}$ and has a forward bias $V_{\text{ext}} = 0.4\,\text{V}$ applied to it. Significant numerical noise can be seen throughout both of the majority regions. The two simulators agree well on the minority currents outside of the depletion region.

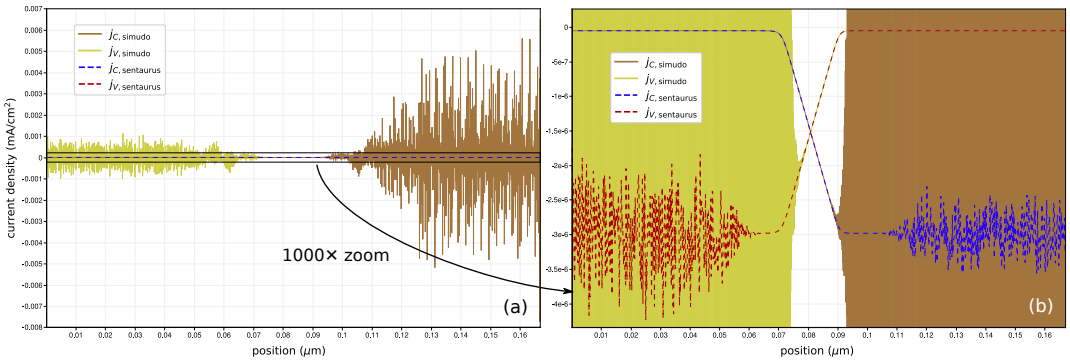

Figure 6.2: Reverse bias variant of Figure 6.1 at $V_{\text{ext}} = -0.6\,\text{V}$, showing the (a) magnitude of the noise and (b) a zoomed-in version showing the correct current densities (from Sentaurus). The numerical noise dominates the signal by a factor of 1000 in the majority regions, making it impossible to resolve the currents there. Sentaurus also has *some* numerical noise in the same regions.

diffusion equation (and resulting weak form) so that it only has one term

$$\vec{j}_k = \mu_k u_k \vec{\nabla} w_k, \tag{6.4}$$

as discussed in Section 5.2.1, and results in the weak forms shown in 5.3.3.1.

Unfortunately, noise is still present in the majority currents (as seen in Figure 5.3), albeit to a much lesser extent than in previous method. The reason is once again catastrophic cancellation, this time because w_k is *extremely* flat for the majority carrier. Computing the derivative ∇w_k in Eq. 5.5 numerically involves evaluating something like

$$\nabla w_k \approx \frac{w_k(x+h) - w_k(x)}{h} \tag{6.5}$$

The subtraction can suffer from catastrophic cancellation if $w_k(x+h)$ and $w_k(x)$ are too close together. Typically h is (approximately) the cell size, so we would expect to run into issues if $|w_k| \ggg h\vec{j}_k/(\mu_k u_k) = h\vec{\nabla} w_k$. Amusingly, the catastrophic cancellation happens because the mesh is *not coarse enough*, the exact opposite of Section 6.1. The minimum cell size h required to resolve w_k is shown in Figure 6.3, where it appears to be two orders of magnitude larger than the size of the entire device. The typical solution here would be to use higher precision floats, but we're already using 64-bit floats and going to 128-bit would incur a massive performance hit. And more importantly, the transition would require a massive development effort since 64-bit floats are hardcoded inside FEniCS. A better approach is required.

6.4 Delta-qfl-mixed method success

The story has a happy ending, as we found a workaround by representing w_k as a sum of two variables $w_{k0} + \delta w_k$ where w_{k0} is a cell-wise average (and is computed before each Newton iteration) and δw_k is an offset away from w_{k0} and is solved for as an FEM trial function. This method is fully described in section 5.3.3.2, and is the method currently used inside Simudo itself. This is the end of our journey, and the beginning of a new era in intermediate band device modeling.

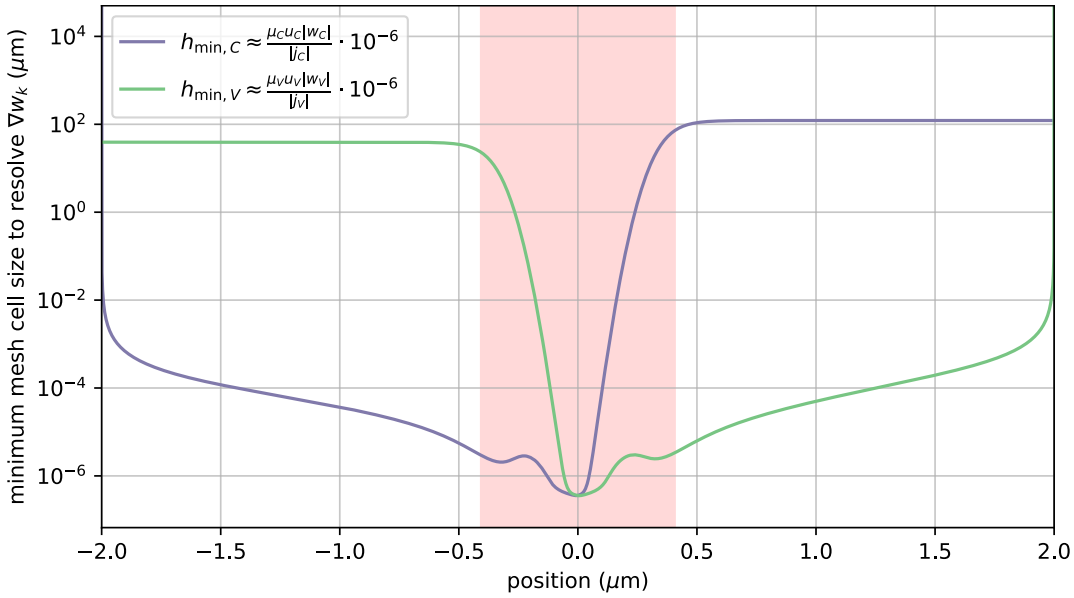

Figure 6.3: Minimum cell size required to accurately resolve ∇w_k throughout a pn junction, assuming that w_k is converged to 6 digits of precision. The device's full parameters are described in Figure 2.8. Since the device's overall width is $4\,\mu$m and a uniform Simudo mesh has on the order of 100 nodes, the cell size is $h \approx 0.04\,\mu$m. However, the plot shows that resolving the majority currents accurately using Eq. 5.5 would require a minimum cell size of $h_{\min} \gtrsim 30\,\mu$m, which is impossible.

Chapter 7

Conclusion

Simudo fills an important gap in semiconductor device simulation: it is extensible, supports both standard semiconductor and IB materials (with their unique effects), and it is available as free and open source software. Its immediate value is apparent, as it has been used to investigate the effect of IB mobility inside devices with mismatched sub-bandgap absorptions [28] and to validate and refine the IB material figure of merit [25].

Outside the IB device modeling niche, Simudo is competitive even in the more crowded space of standard semiconductor simulators. Simudo beat Sentaurus on a reasonable benchmark (section 5.3.5), and has better convergence with respect to the number of mesh points (figure 5.4).

An indispensable part of research is the realization that there are many more questions to be asked (and answers to be questioned). For instance, Simudo currently relies on the Boltzmann approximation for carrier densities. However, the approximation cannot be used for semiconductors with degenerate bands. Overcoming this limitation is a planned feature for the future, using efficient approximations to the Fermi-Dirac integral for parabolic bands [36, 96] to implement the relationship between u_k and w_k. Moreover, Simudo itself could use some more features such as tools for 2D/3D modelling and a graphical user interface. Other planned developments include writing a proper regression test suite (to quickly check that future code changes do not worsen Simudo's accuracy on existing problems), implementing reflecting and scattering boundary conditions for optics, and multiprocessing support via MPI.

Bibliography

[1] Eduard C. Dumitrescu, Matthew M. Wilkins, and Jacob J. Krich. "Simudo: a device model for intermediate band materials". In: *Journal of Computational Electronics* 19.1 (Mar. 2020), pp. 111–127. ISSN: 1572-8137. DOI: 10.1007/s10825-019-01414-3.

[2] Eduard C. Dumitrescu, Matthew M. Wilkins, and Jacob J. Krich. *Simudo: a device model for intermediate band materials*. Preprint, arXiv:1905.11303. 2019. arXiv: 1905.11303 [physics.app-ph]. URL: https://arxiv.org/abs/1905.11303v1.

[3] J.R. Hauser. "Boundary conditions at p-n junctions". In: *Solid-State Electronics* 14.2 (Feb. 1971), pp. 133–139. ISSN: 0038-1101. DOI: 10.1016/0038-1101(71)90088-8.

[4] William Shockley and Hans J. Queisser. "Detailed Balance Limit of Efficiency of p-n Junction Solar Cells". In: *Journal of Applied Physics* 32.3 (1961), pp. 510–519. DOI: 10.1063/1.1736034.

[5] Gerardo L. Araújo and Antonio Martí. "Absolute limiting efficiencies for photovoltaic energy conversion". In: *Solar Energy Materials and Solar Cells* 33.2 (1994), pp. 213–240. ISSN: 0927-0248. DOI: 10.1016/0927-0248(94)90209-7.

[6] Martin A. Green. *Third Generation Photovoltaics Advanced solar energy conversion*. Springer Series in Photonics. Springer, 2005. ISBN: 3-540-26562-7,978-3-540-26562-7.

[7] M. Wolf. "Limitations and Possibilities for Improvement of Photovoltaic Solar Energy Converters: Part I: Considerations for Earth's Surface Operation". In: *Proceedings of the IRE* 48.7 (1960), pp. 1246–1263. DOI: 10.1109/JRPROC.1960.287647.

[8] Jianming Li, Ming Chong, Jiancheng Zhu, Yuanjing Li, Jiadong Xu, Peida Wang, Zuoqi Shang, Zhankun Yang, Ronghua Zhu, and Xiolan Cao. "35% efficient nonconcentrating novel silicon solar cell". In: *Applied Physics Letters* 60.18 (1992), pp. 2240–2242. DOI: 10.1063/1.107042.

[9] Antonio Luque and Antonio Martí. "Increasing the Efficiency of Ideal Solar Cells by Photon Induced Transitions at Intermediate Levels". In: *Phys. Rev. Lett.* 78.26 (1997), pp. 5014–5017. DOI: 10.1103/PhysRevLett.78.5014.

[10] Yunlu Xu, Tao Gong, and Jeremy N. Munday. "The generalized Shockley-Queisser limit for nanostructured solar cells". In: *Scientific Reports* 5.1 (Sept. 2015), p. 13536. ISSN: 2045-2322. DOI: 10.1038/srep13536.

[11] N. López, L. A. Reichertz, K. M. Yu, K. Campman, and W. Walukiewicz. "Engineering the Electronic Band Structure for Multiband Solar Cells". In: *Phys. Rev. Lett.* 106.2 (2011), pp. 028701–. DOI: 10.1103/PhysRevLett.106.028701.

[12] Weiming Wang, Albert S. Lin, and Jamie D. Phillips. "Intermediate-band photovoltaic solar cell based on ZnTe:O". In: *Appl. Phys. Lett.* 95.1 (2009), p. 011103. DOI: 10.1063/1.3166863.

[13] A Martí, N López, E Antolin, E Cánovas, C Stanley, Cuadra Farmer, L Cuadra, and A Luque. "Novel semiconductor solar cell structures: The quantum dot intermediate band solar cell". In: *Thin solid films* 511 (2006), pp. 638–644.

[14] Ross Cheriton, Sharif M. Sadaf, Luc Robichaud, Jacob J. Krich, Zetian Mi, and Karin Hinzer. "Two-photon photocurrent in InGaN/GaN nanowire intermediate band solar cells". In: *Communications Materials* 1.1 (Sept. 15, 2021), p. 63. ISSN: 2662-4443. DOI: 10.1038/s43246-020-00054-6.

[15] Björn Marsen, Sascha Klemz, Thomas Unold, and Hans-Werner Schock. "Investigation of the sub-bandgap Photoresponse in CuGaS2: Fe for intermediate band solar cells". In: *Progress in Photovoltaics: Research and Applications* 20.6 (2012), pp. 625–629.

[16] Jinn-Kong Sheu, Feng-Wen Huang, Yu-Hsuan Liu, PC Chen, Yu-Hsiang Yeh, Ming-Lun Lee, and Wei-Chih Lai. "Photoresponses of manganese-doped gallium nitride grown by metalorganic vapor-phase epitaxy". In: *Applied Physics Letters* 102.7 (2013), p. 071107.

[17] Y. Okada, N. J. Ekins-Daukes, T. Kita, R. Tamaki, M. Yoshida, A. Pusch, O. Hess, C. C. Phillips, D. J. Farrell, K. Yoshida, N. Ahsan, Y. Shoji, T. Sogabe, and J.-F. Guillemoles. "Intermediate band solar cells: Recent progress and future directions". In: *Applied Physics Reviews* 2.2 (2015), p. 021302. DOI: 10.1063/1.4916561.

[18] Iñigo Ramiro and Antonio Martí. "Intermediate band solar cells: Present and future". In: *Progress in Photovoltaics: Research and Applications* 29.7 (Sept. 15, 2021), pp. 705–713. DOI: 10.1002/pip.3351.

[19] Antonio Luque, Antonio Martí, and Colin Stanley. "Understanding intermediate-band solar cells". In: *Nature Photonics* 6.3 (2012), p. 146. DOI: 10.1038/nphoton.2012.1.

[20] Koichi Yamaguchi and Toru Kanto. "Self-assembled InAs quantum dots on GaSb/GaAs (0 0 1) layers by molecular beam epitaxy". In: *Journal of Crystal Growth* 275.1-2 (2005), e2269–e2273. DOI: 10.1016/j.jcrysgro.2004.11.363.

[21] Naoki Kakuda, Takayuki Yoshida, and Koichi Yamaguchi. "Sb-mediated growth of high-density InAs quantum dots and GaAsSb embedding growth by MBE". In: *Applied surface science* 254.24 (2008), pp. 8050–8053. DOI: 10.1016/j.apsusc.2008.03.017.

[22] Edes Saputra, Jun Ohta, Naoki Kakuda, and Koichi Yamaguchi. "Self-formation of in-plane ultrahigh-density InAs quantum dots on GaAsSb/GaAs (001)". In: *Applied Physics Express* 5.12 (2012), p. 125502. DOI: 10.1143/APEX.5.125502.

[23] Jacob J. Krich, Bertrand I. Halperin, and Alàn Aspuru-Guzik. "Nonradiative lifetimes in intermediate band photovoltaics–Absence of lifetime recovery". In: *Journal of Applied Physics* 112.1, 013707 (2012), pp. -. DOI: 10.1063/1.4732085.

[24] J. T. Sullivan, C. B. Simmons, T. Buonassisi, and J. J. Krich. "Targeted Search for Effective Intermediate Band Solar Cell Materials". In: *IEEE Journal of Photovoltaics* 5.1 (2015), pp. 212–218. DOI: 10.1109/JPHOTOV.2014.2363560.

[25] Matthew M. Wilkins, Eduard C. Dumitrescu, and Jacob J. Krich. "Material quality requirements for intermediate band solar cells". In: *IEEE Journal of Photovoltaics* 10.2 (2020), pp. 467–474. DOI: 10.1109/JPHOTOV.2019.2959934.

[26] Peter A. Markowich. *The Stationary Semiconductor Device Equations*. Computational Microelectronics. Vienna: Springer Vienna, 1986. ISBN: 978-3-211-99937-0. DOI: 10.1007/978-3-7091-3678-2.

[27] Rune Strandberg and Turid Worren Reenaas. "Drift-diffusion model for intermediate band solar cells including photofilling effects". In: *Prog. Photovolt: Res. Appl.* 19.1 (2011), pp. 21–32. ISSN: 1099-159X. DOI: 10.1002/pip.983.

[28] Avya Stolow, David J. Gayowsky, Matthew M. Wilkins, and Jacob J. Krich. "Intermediate band mobility requirements in intermediate band solar cells". In: *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*. June 2020, pp. 1645–1650. DOI: 10.1109/PVSC45281.2020.9300909.

[29] Martin S. Alnæs, Jan Blechta, Johan Hake, August Johansson, Benjamin Kehlet, Anders Logg, Chris Richardson, Johannes Ring, Marie E. Rognes, and Garth N. Wells. "The FEniCS Project Version 1.5". In: *Archive of Numerical Software* 3.100 (2015). DOI: 10.11588/ans.2015.100.20553.

[30] Anders Logg, Kristian B. Ølgaard, Marie E. Rognes, and Garth N. Wells. "FFC: the FEniCS Form Compiler". In: *Automated Solution of Differential Equations by the Finite Element Method, Volume 84 of Lecture Notes in Computational Science and Engineering*. Ed. by Anders Logg, Kent-Andre Mardal, and Garth N. Wells. Springer, 2012. Chap. 11.

[31] Marie E. Rognes, David A. Ham, Colin J. Cotter, and Andrew T. T. McRae. "Automating the solution of PDEs on the sphere and other manifolds in FEniCS 1.2". In: *Geoscientific Model Development* 6 (2013), pp. 2099–2119. DOI: 10.5194/gmd-6-2099-2013.

[32] Michael P. Marder. *Condensed Matter Physics*. 2nd ed. Wiley, 2010.

[33] Ansgar Jüngel. *Transport equations for semiconductors*. 1st ed. Lecture Notes in Physics 773. Springer-Verlag Berlin Heidelberg, 2009. ISBN: 3540895256, 9783540895251.

[34] Stephen J. Fonash. *Solar Cell Device Physics*. Ed. by Stephen J. Fonash. Second Edition. Boston: Academic Press, 2010. ISBN: 978-0-12-374774-7. DOI: 10.1016/B978-0-12-374774-7.00001-7.

[35] W. Shockley and W. T. Read. "Statistics of the Recombinations of Holes and Electrons". In: *Phys. Rev.* 87.5 (1952), pp. 835–. DOI: 10.1103/PhysRev.87.835.

[36] R. F. Pierret. *Advanced semiconductor fundamentals*. 2nd ed. Modular Series on Solid State Devices vol. 6. Pearson, 2002. ISBN: 013061792X, 0-201-12295-2, 0-201-12296-0, 0-201-12297-9, 0-201-12298-7, 0-201-44494-1.

[37] Michael Spivak. *Calculus*. 2nd. Publish or Perish, 1980.

[38] Simon M Sze and Kwok K Ng. *Physics of semiconductor devices*. John Wiley & Sons, 2006.

[39] Aurelien David, Christophe A Hurni, Nathan G Young, and Michael D Craven. "Electrical properties of III-Nitride LEDs: Recombination-based injection model and theoretical limits to electrical efficiency and electroluminescent cooling". In: *Applied Physics Letters* 109.8 (2016), p. 083501. DOI: 10.1063/1.4961491.

[40] Antonio Huerta Jean Donea. *Finite element methods for flow problems*. Wiley, 2003. ISBN: 9780471496663,0-471-49666-9.

[41] Mark S. Gockenbach. *Understanding and implementing the finite element method*. SIAM, 2006. ISBN: 978-0-89871-614-6. URL: http://bookstore.siam.org/ot97/.

[42] Randolph E Bank, Donald J Rose, and Wolfgang Fichtner. "Numerical Methods for Semiconductor Device Simulation". In: *IEEE Transactions on Electron Devices* 30.9 (1983), pp. 1031–1041. DOI: 10.1109/T-ED.1983.21257.

[43] W. Fichtner, D.J. Rose, and R.E. Bank. "Semiconductor device simulation". In: *IEEE Transactions on Electron Devices* 30.9 (Sept. 1983), pp. 1018–1030. ISSN: 0018-9383. DOI: 10.1109/T-ED.1983.21256.

[44] J. Piprek, ed. *Handbook of Optoelectronic Device Modeling & Simulation*. Boca Raton, FL: CRC Press, 2018, p. 1720. ISBN: 9781498749466.

[45] Andreas Schenk. *Advanced Physical Models for Silicon Device Simulation*. Springer-Verlag Wien, 1998, p. 348. ISBN: 3211830529. DOI: 10.1007/978-3-7091-6494.

[46] Jonathan P. Mailoa, Austin J. Akey, Christie B. Simmons, David Hutchinson, Jay Mathews, Joseph T. Sullivan, Daniel Recht, Mark T. Winkler, James S. Williams, Jeffrey M. Warrender, Peter D. Persans, Michael J. Aziz, and Tonio Buonassisi. "Room-temperature sub-band gap optoelectronic response of hyperdoped silicon". In: *Nat Commun* 5 (2014), p. 3011. DOI: 10.1038/ncomms4011.

[47] Yonder Berencén, Slawomir Prucnal, Fang Liu, Ilona Skorupa, René Hübner, Lars Rebohle, Shengqiang Zhou, Harald Schneider, Manfred Helm, and Wolfgang Skorupa. "Room-temperature short-wavelength infrared Si photodetector". In: *Scientific Reports* 7 (2017), p. 43688. DOI: 10.1038/srep43688.

[48] Mao Wang, Y. Berencén, E. García-Hemme, S. Prucnal, R. Hübner, Ye Yuan, Chi Xu, L. Rebohle, R. Böttger, R. Heller, H. Schneider, W. Skorupa, M. Helm, and Shengqiang Zhou. "Extended Infrared Photoresponse in Te-Hyperdoped Si at Room Temperature". In: *Phys. Rev. Applied* 10.2 (2018), p. 024054. DOI: 10.1103/PhysRevApplied.10.024054.

[49] Andrew S. Brown and Martin A. Green. "Impurity photovoltaic effect: Fundamental energy conversion efficiency limits". In: *J. Appl. Phys.* 92.3 (2002), pp. 1329–1336. DOI: 10.1063/1.1492016.

[50] A. Martí, E. Antolín, C. R. Stanley, C. D. Farmer, N. López, P. Díaz, E. Cánovas, P. G. Linares, and A. Luque. "Production of Photocurrent due to Intermediate-to-Conduction-Band Transitions: A Demonstration of a Key Operating Principle of the Intermediate-Band Solar Cell". In: *Phys. Rev. Lett.* 97.24 (2006), pp. 247701–. DOI: 10.1103/PhysRevLett.97.247701.

[51] Matthias Auf der Maur. "A Multiscale Simulation Environment for Electronic and Optoelectronic Devices". PhD thesis. Universita' degli Studi di Roma Tor Vergata, 2008.

[52] Stefan Birner, Tobias Zibold, Till Andlauer, Tillmann Kubis, Matthias Sabathil, Alex Trellakis, and Peter Vogl. "nextnano: General Purpose 3-D Simulations". In: *IEEE Transactions on Electron Devices* 54.9 (Sept. 2007), pp. 2137–2142. ISSN: 0018-9383. DOI: 10.1109/TED.2007.902871.

[53] D. A. Clugston and P. A. Basore. "PC1D version 5: 32-bit solar cell modeling on personal computers". In: *Twenty Sixth IEEE Photovoltaic Specialists Conference*. 1997, pp. 207–210. DOI: 10.1109/PVSC.1997.654065.

[54] Halvard Haug and Johannes Greulich. "PC1Dmod 6.2 - Improved Simulation of c-Si Devices with Updates on Device Physics and User Interface". In: *Energy Procedia* 92.1876 (2016), pp. 60–68. ISSN: 18766102. DOI: 10.1016/j.egypro.2016.07.010.

[55] R. Varache, C. Leendertz, M.E. Gueunier-Farret, J. Haschke, D. Muñoz, and L. Korte. "Investigation of selective junctions using a newly developed tunnel current model for solar cell applications". In: *Solar Energy Materials and Solar Cells* 141 (2015), pp. 14–23. ISSN: 0927-0248. DOI: 10.1016/j.solmat.2015.05.014.

[56] M Burgelman, P Nollet, and S Degrave. "Modelling polycrystalline semiconductor solar cells". In: *Thin Solid Films* 361-362 (2000), pp. 527–532. ISSN: 0040-6090. DOI: 10.1016/S0040-6090(99)00825-1.

[57] D. Alonso-Álvarez, T. Wilson, P. Pearce, M. Führer, D. Farrell, and N. Ekins-Daukes. "Solcore: a multi-scale, Python-based library for modelling solar cells and semiconductor materials". In: *Journal of Computational Electronics* 17.3 (2018), pp. 1099–1123. ISSN: 15728137. DOI: 10.1007/s10825-018-1171-3. URL: 10.1007/s10825-018-1171-3.

[58] Antonio Martí, L. Cuadra, and Antonio Luque. "Quasi-drift diffusion model for the quantum dot intermediate band solar cell". In: *IEEE Transactions on Electron Devices* 49.9 (2002), pp. 1632–1639. ISSN: 0018-9383. DOI: 10.1109/TED.2002.802642.

[59] I. Tobías, A. Luque, and A. Martí. "Numerical modeling of intermediate band solar cells". In: *Semiconductor Science and Technology* 26.1 (2011), pp. 014031–. ISSN: 0268-1242. DOI: 10.1088/0268-1242/26/1/014031.

[60] Katsuhisa Yoshida, Yoshitaka Okada, and Nobuyuki Sano. "Device simulation of intermediate band solar cells: Effects of doping and concentration". In: *Journal of Applied Physics* 112.8 (2012), pp. -. DOI: 10.1063/1.4759134.

[61] L. Cuadra, A. Martí, and A. Luque. "Influence of the overlap between the absorption coefficients on the efficiency of the intermediate band solar cell". In: *IEEE Transactions on Electron Devices* 51.6 (2004), pp. 1002–1007. DOI: 10.1109/TED.2004.828161.

[62] Michael Y. Levy and Christiana Honsberg. "Intraband absorption in solar cells with an intermediate band". In: *Journal of Applied Physics* 104.11 (2008), p. 113103. DOI: 10.1063/1.3021449.

[63] W. G. Hu, T. Inoue, O. Kojima, and T. Kita. "Effects of absorption coefficients and intermediate-band filling in InAs/GaAs quantum dot solar cells". In: *Appl. Phys. Lett.* 97.19 (2010), p. 193106. ISSN: 0003-6951. DOI: 10.1063/1.3516468.

[64] R. Strandberg. "Analytic JV-Characteristics of Ideal Intermediate Band Solar Cells and Solar Cells With Up and Downconverters". In: *IEEE Transactions on Electron Devices* 64.5 (2017), pp. 2275–2282. ISSN: 0018-9383. DOI: 10.1109/TED.2017.2686359.

[65] R. Strandberg. "The JV-Characteristic of Intermediate Band Solar Cells With Overlapping Absorption Coefficients". In: *IEEE Transactions on Electron Devices* 64.12 (2017), pp. 5027–5033. ISSN: 0018-9383. DOI: 10.1109/TED.2017.2762761.

[66] Albert S. Lin, Weiming Wang, and Jamie D. Phillips. "Model for intermediate band solar cells incorporating carrier transport and recombination". In: *Journal of Applied Physics* 105.6, 064512 (2009), p. 064512. DOI: 10.1063/1.3093962.

[67] T.S. Navruz and M. Saritas. "Efficiency variation of the intermediate band solar cell due to the overlap between absorption coefficients". In: *Solar Energy Materials and Solar Cells* 92.3 (2008), pp. 273–282. ISSN: 0927-0248. DOI: 10.1016/j.solmat.2007.08.012.

[68] Jacob J. Krich, Anna H. Trojnar, Liang Feng, Karin Hinzer, and Alexandre W. Walker. "Modeling intermediate band solar cells: a roadmap to high efficiency". In: *Proc. SPIE 8981, Physics, Simulation, and Photonic Engineering of Photovoltaic Devices III*. 2014, 89810O. DOI: 10.1117/12.2041807.

[69] K. Yoshida, Yoshitaka Okada, and Nobuyuki Sano. "Self-consistent simulation of intermediate band solar cells: Effect of occupation rates on device characteristics". In: *Applied Physics Letters* 97.13 (2010). Ed. by Y. Okada and N. Sano, pp. 133503–. DOI: 10.1063/1.3488815.

[70] Robert Eymard, Thierry Gallouët, and Raphaèle Herbin. "Finite volume methods". In: *Handbook of Numerical Analysis*. Vol. 7. Elsevier, 2000, pp. 713–1018. DOI: 10.1016/S1570-8659(00)07005-8.

[71] Y. He and G. Cao. "A generalized Scharfetter-Gummel method to eliminate crosswind effects (semiconduction device modeling)". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 10.12 (1991), pp. 1579–1582. ISSN: 0278-0070. DOI: 10.1109/43.103508.

[72] Abdeljalil Nachaoui. "Iterative solution of the drift-diffusion equations". In: *Numerical Algorithms* 21 (1999), pp. 323–341. ISSN: 1017-1398. DOI: 10.1023/A:1019165614860.

[73] Pavel Bochev, Kara Peterson, and Mauro Perego. "A multiscale control volume finite element method for advection-diffusion equations". In: *Int. J. Numer. Meth. Fluids* 77.11 (2015), pp. 641–667. ISSN: 0271-2091. DOI: 10.1002/fld.3998.

[74] Bernardo Cockburn, George E. Karniadakis, and Chi-Wang Shu, eds. *Discontinuous Galerkin Methods: Theory, Computation, and Applications*. Vol. 11. Lecture Notes in Computational Science and Engineering. Berlin, Heidelberg: Springer, 2000. DOI: 10.1007/978-3-642-59721-3.

[75] Gaurav Kumar, Mandeep Singh, Anand Bulusu, and Gaurav Trivedi. "A Framework to Simulate Semiconductor Devices Using Parallel Computer Architecture". In: *Journal of Physics: Conference Series*. Vol. 759. IOP Publishing, 2016, pp. 012098–. DOI: 10.1088/1742-6596/759/1/012098.

[76] G. Kumar, M. Singh, A. Ray, and G. Trivedi. "An FEM based framework to simulate semiconductor devices using streamline upwind Petrov-Galerkin stabilization technique". In: *2017 27th International Conference Radioelektronika*. 2017, pp. 1–5. DOI: 10.1109/RADIOELEK.2017.7936644.

[77] Frédéric Poupaud and Christian Schmeiser. "Charge transport in semiconductors with degeneracy effects". In: *Mathematical methods in the applied sciences* 14.5 (1991), pp. 301–318.

[78] Alan H Marshak and CM Van Vliet. "Electrical current and carrier density in degenerate materials with nonuniform band structure". In: *Proceedings of the IEEE* 72.2 (1984), pp. 148–164.

[79] Jenny Nelson. *The Physics of Solar Cells*. Imperial College Press, 2003.

[80] Keith R. McIntosh and Lachlan E. Black. "On effective surface recombination parameters". In: *Journal of Applied Physics* 116.1 (2014). ISSN: 10897550. DOI: 10.1063/1.4886595.

[81] J.M. Zhao, J.Y. Tan, and L.H. Liu. "A second order radiative transfer equation and its solution by meshless method with application to strongly inhomogeneous media". In: *Journal of Computational Physics* 232.1 (2013), pp. 431–455. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2012.08.020.

[82] Claes Johnson. *Numerical solution of partial differential equations by the finite element method*. Cambridge University Press, 1987. ISBN: 9780521345149, 0521345146, 0521347580, 9780521347587.

[83] B. Gaury, Y. Sun, P. Bermel, and P. M. Haney. "Sesame: a 2-dimensional solar cell modeling tool". In: *Solar Energy Materials and Solar Cells* 198 (2019), pp. 53–62. DOI: 10.1016/j.solmat.2019.03.037.

[84] Franco Brezzi, Jim Douglas, and L. D. Marini. "Two families of mixed finite elements for second order elliptic problems". In: *Numerische Mathematik* 47.2 (1985), pp. 217–235. ISSN: 0945-3245. DOI: 10.1007/BF01389710.

[85] J.E. Roberts and J.-M. Thomas. "Mixed and hybrid methods". In: *Finite Element Methods (Part 1)*. Vol. 2. Handbook of Numerical Analysis Supplement C. Elsevier, 1991, pp. 523–639. DOI: 10.1016/S1570-8659(05)80041-9.

[86] Daniel J. Cummings, Mark E. Law, Steve Cea, and Tom Linton. "Comparison of discretization methods for device simulation". In: *International Conference on Simulation of Semiconductor Processes and Devices 2009* (2009), pp. 1–4. DOI: 10.1109/SISPAD.2009.5290236.

[87] Synopsys Inc. *Sentaurus Device User Guide, vK-2015*. Synopsys Inc., 2015, pp. 245–247.

[88] Anders Logg, Kent-Andre Mardal, Garth N. Wells, et al. *Automated Solution of Differential Equations by the Finite Element Method*. Springer, 2012. ISBN: 978-3-642-23098-1. DOI: 10.1007/978-3-642-23099-8.

[89] Rune Strandberg and Turid Worren Reenaas. "Photofilling of intermediate bands". In: *J. Appl. Phys.* 105.12 (2009), pp. 124512–8. DOI: 10.1063/1.3153141.

[90] D. L. Scharfetter and H. K. Gummel. "Large-signal analysis of a silicon Read diode oscillator". In: *IEEE Transactions on Electron Devices* 16.1 (Jan. 1969), pp. 64–77. ISSN: 0018-9383. DOI: 10.1109/T-ED.1969.16566.

[91] Franco Brezzi, Luisa Donatella Marini, and Paola Pietra. "Two-dimensional exponential fitting and applications to drift-diffusion models". In: *SIAM Journal on Numerical Analysis* 26.6 (1989), pp. 1342–1355.

[92] Alexander N. Brooks and Thomas J.R. Hughes. "Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations". In: *Computer Methods in Applied Mechanics and Engineering* 32.1 (1982), pp. 199–259. ISSN: 0045-7825. DOI: 10.1016/0045-7825(82)90071-8.

[93] Erik Burman and Cuiyu He. "Primal Dual Mixed Finite Element Methods for Indefinite Advection-Diffusion Equations". In: *SIAM Journal on Numerical Analysis* 57.6 (2019), pp. 2785–2811. DOI: 10.1137/18M1221473.

[94] Randolph E. Bank, Josef F. Bürgler, Wolfgang Fichtner, and R. Kent Smith. "Some upwinding techniques for finite element approximations of convection-diffusion equations". In: *Numerische Mathematik* 58.1 (Dec. 1990), pp. 185–202. ISSN: 0945-3245. DOI: 10.1007/BF01385618.

[95] G. Kumar, M. Singh, A. Ray, and G. Trivedi. "An FEM based framework to simulate semiconductor devices using streamline upwind Petrov-Galerkin stabilization technique". In: *2017 27th International Conference Radioelektronika (RADIOELEKTRONIKA)*. Apr. 2017, pp. 1–5. DOI: 10.1109/RADIOELEK.2017.7936644.

[96] J.S. Blakemore. "Approximations for Fermi-Dirac integrals, especially the function F12(η) used to describe electron density in a semiconductor". In: *Solid-State Electronics* 25.11 (1982), pp. 1067–1076. ISSN: 0038-1101. DOI: 10.1016/0038-1101(82)90143-5.

Appendix A

Miscellaneous proofs

Lemma 2. *The function* $f:\ (0,\infty) \to \mathbb{R}$ *defined by*

$$f(r) = \frac{1}{\sqrt{1+1/r}} + \frac{1}{\sqrt{1+r}} \tag{A.1}$$

attains its maximum at $r = 1$.

Proof. The function is smooth, so we look for extremal points by setting its derivative to zero.

$$0 = \partial_r f(r) \tag{A.2}$$

$$0 = \partial_r \left[\frac{1}{\sqrt{1+r^{-1}}} + \frac{1}{\sqrt{1+r}} \right] \tag{A.3}$$

$$0 = \frac{-1}{2}(1+r^{-1})^{-3/2} \cdot -r^{-2} + \frac{-1}{2}(1+r)^{-3/2} \tag{A.4}$$

$$0 = -(1+r)^{-3/2} r^{1/2} + (1+r)^{-3/2} \tag{A.5}$$

$$0 = -r^{1/2} + 1 \tag{A.6}$$

The function takes value $\sqrt{2}$ at $r = 1$, $\lim_{r\to\infty} f(r) = 1$, and $f(r) = f(r^{-1})$. Therefore $r = 1$ is the only extremum, and it is a maximum. □

Appendix B

Jump operator

Within the context of the finite element method, the jump operator is defined on cell boundaries, and it takes the difference in an expression between two cells. For example, consider a 1D context with two cells at $x \in [0, 1]$ and $x \in [1, 2]$, and consider a discontinuous finite element function defined on the two cells as

$$f(x) = \begin{cases} 20 & 0 \leq x \leq 1 \\ 50 & 1 \leq x \leq 2 \end{cases}$$

which could be a DG(0) function for example. The jump operator $[\cdot]$ is defined only at the interface of the two cells ($x = 1$). The jump operator applied on $f(x)$ at $x = 1$ is

$$\begin{aligned} [f]\Big|_{x=1} &= \lim_{h \to 1^+} f(x) - \lim_{h \to 1^-} f(x) \\ &= 50 - 20 \\ &= 30 \end{aligned}$$

The jump operator can be applied to any discontinuous expression, not only functions in the finite element space (such as DG or BDM). For example,

$$\begin{aligned} [e^{f(x)}]\Big|_{x=1} &= \lim_{h \to 1^+} e^{f(x)} - \lim_{h \to 1^-} e^{f(x)} \\ &= e^{50} - e^{20} \end{aligned}$$