Weierstraß-Institut für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint

ISSN 2198-5855

Computational and analytical comparison of flux discretizations for the semiconductor device equations beyond Boltzmann statistics

Patricio Farrell, Thomas Koprucki, Jürgen Fuhrmann

submitted: November 9, 2016

Weierstrass Institute Mohrenstr. 39 10117 Berlin Germany email: patricio.farrell@wias-berlin.de thomas.koprucki@wias-berlin.de jürgen.fuhrmann@wias-berlin.de

> No Berlin 2016



1

²⁰¹⁰ Mathematics Subject Classification. 35Q99, 82D37, 65M08, 65N08, 74S10.

Key words and phrases. finite volume method, flux discretization, Scharfetter–Gummel scheme, Fermi–Dirac statistics, degenerate semiconductors, van Roosbroeck system, semi-conductor device simulation, nonlinear diffusion, diffusion enhancement.

Edited by Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS) Leibniz-Institut im Forschungsverbund Berlin e. V. Mohrenstraße 39 10117 Berlin Germany

Fax:+493020372-303E-Mail:preprint@wias-berlin.deWorld Wide Web:http://www.wias-berlin.de/

Abstract

For a Voronoï finite volume discretization of the van Roosbroeck system with general charge carrier statistics we compare three thermodynamically consistent numerical fluxes known in the literature. We discuss an extension of the Scharfetter-Gummel scheme to non-Boltzmann (e.g. Fermi-Dirac) statistics. It is based on the analytical solution of a two-point boundary value problem obtained by projecting the continuous differential equation onto the interval between neighboring collocation points. Hence, it serves as a reference flux. The exact solution of the boundary value problem can be approximated by computationally cheaper fluxes which modify certain physical quantities. One alternative scheme averages the nonlinear diffusion (caused by the non-Boltzmann nature of the problem), another one modifies the effective density of states. To study the differences between these three schemes, we analyze the Taylor expansions, derive an error estimate, visualize the flux error and show how the schemes perform for a carefully designed p-i-n benchmark simulation. We present strong evidence that the flux discretization based on averaging the nonlinear diffusion has an edge over the scheme based on modifying the effective density of states.

1 Introduction

The van Roosbroeck system [1] (also known as the semiconductor device equations) has become a standard model to describe the flow of electrons and holes in semiconductor devices. Its numerical approximation is very well understood if Boltzmann statistics accurately describes the electron and hole densities. In this case, one refers to the semiconductor as *non-degenerate*. Scharfetter and Gummel [2] presented in the late 1960ies a flux discretization scheme which could deal with the numerical challenges such as stability and preservation of maximum principles posed by these equations, see for example [3, 4, 5]. The generalization to non-Boltzmann statistics (in *degenerate* semiconductor materials), however, presents very similar challenges which are not satisfactorily solved yet. Hence, the goal of this paper is to study the influence of three thermodynamically consistent flux approximations used in a Voronoï finite volume discretization of the van Roosbroeck system when assuming more general statistics functions. We will compare these schemes analytically and numerically to assess their quality for semiconductor device simulations.

A very general and computationally rather expensive finite volume flux approximation was studied in [6]. In order to determine the numerical flux between two control volumes this method needs to solve an integral equation. This integral equation is a reformulation of a nonlinear two-point boundary value problem which one obtains from projecting the continuity equation for the charge carriers onto the interval between two nodes belonging to neighboring cells. For so-called Blakemore statistics [7] the integral equation is known to simplify to a nonlinear algebraic equation which can be solved with a few Newton steps [8]. No corresponding simplification is known for more general statistics. For this reason several ideas were proposed. In [9], piecewise approximations (e.g. Padé interpolants) for the statistics function are discussed. However, this approach may still be rather costly. Hence, *modified* Scharfetter-Gummel schemes which only approximately solve the two-point boundary value problem may yield a good compromise between efficiency and accurateness by averaging certain quantities.

Several such schemes have been suggested to deal with more general statistics [10, 11, 12]. Unfortunately, these discretizations are not consistent with the thermodynamic equilibrium, i. e. they do not satisfy an analogous discrete version of the continuous property that the fluxes vanish if the quasi Fermi potentials are constant. Thermodynamic consistency is extremely important to avoid unphysical steady state dissipation. Furthermore, the consistent discretization of dissipative effects is crucial when coupling the semiconductor device equations to heat transport models.

Bessemoulin-Chatard suggested a scheme which averages the diffusion enhancement in such a way that the resulting flux approximation is thermodynamically consistent [13]. The diffusion enhancement can be interpreted as a measure for how far the system is from the Boltzmann regime. It leads to a nonlinear diffusion coefficient whose particular form is induced by the statistical distribution function for the charge carrier densities, e. g., the Fermi-Dirac integral of order one half. Using this flux, Bessemoulin-Chatard proved convergence of a semi-implicit finite volume scheme. This *diffusion enhanced* scheme was translated into the context of semiconductor device models in [14], making the dependency on the diffusion enhancement explicit.

It is also possible to derive another class of schemes by modifying the effective density of states. This so-called *inverse activity* scheme was introduced in [15] for the numerical solution of the generalized Nernst-Planck system which is similar to the van Roosbroeck system. A variant of this scheme for Fermi-Dirac statistics is described in [16, 17]. Here the adaption to general statistics is realized via averaging the inverse activity coefficient. Even though any such average will yield a thermodynamically consistent scheme, we will focus on two practical choices: an arithmetic and geometric mean of the inverse activity coefficients at neighboring nodes. However, some of our results apply to any average for the inverse activity coefficient that satisfies very mild additional assumptions.

We continue this paper by introducing the van Roosbroeck system in Section 2 and present its discretization as well as the different flux approximations in Section 3. In order to compare these schemes, we study their Taylor series expansions in Section 4, derive a general error estimate in Section 5, compare the flux error visually in Section 6 and finally analyze the influence of the different numerical fluxes to the coupled van Roosbroeck system by simulating a carefully chosen device setup consisting of a p-doped, intrinsic and n-doped region (p-i-n device) in Section 7.

This publication is supplemented with a Mathematica notebook, Matlab files and simulation data, which can be used to verify the presented results [18].

2 The van Roosbroeck system

The van Roosbroeck system describes the charge carrier flow and the electrostatic potential in a semiconductor device. It consists of three coupled nonlinear partial differential equations: one for the electrostatic potential ψ and two continuity equations, one for the electron and one for the hole density which we denote with n and p. We consider a homogeneous material and some domain $\Omega \subseteq \mathbb{R}^d$ for $d \in \{1, 2, 3\}$.

Then the stationary van Roosbroeck system is given by

$$-\nabla \cdot \left(\varepsilon_0 \varepsilon_r \nabla \psi\right) = q \left(p - n + C\right),\tag{1a}$$

$$\nabla \cdot \mathbf{j}_n = qR,\tag{1b}$$

$$\nabla \cdot \mathbf{j}_p = -qR. \tag{1c}$$

The constants q, ε_0 and ε_r denote the elementary charge, the vacuum dielectric permittivity and the relative permittivity of the semiconductor, respectively. The recombination rate R depends on the electron and hole densities and the doping profile C may vary spatially.

The electron density and the hole density are related to the electrostatic potential ψ as well as the quasi Fermi potentials of electrons and holes φ_n and φ_p via a statistical distribution function \mathcal{F} , namely by

$$n = N_c \mathcal{F}\left(\frac{q(\psi - \varphi_n) - E_c}{k_B T}\right) \quad \text{and} \quad p = N_v \mathcal{F}\left(\frac{q(\varphi_p - \psi) + E_v}{k_B T}\right).$$
(2)

The effective densities of states for electrons in the conduction band N_c and holes in the valence band N_v as well as the corresponding band-edge energies E_c and E_v are material parameters and assumed to be constant in this paper. However, in applications they can vary with the material (for example due to abrupt or graded heterojunctions). The temperature T is also assumed to be constant; in general it can be space or even time dependent. The Boltzmann constant is denoted with k_B .

Common examples for the statistical distribution function \mathcal{F} can be found in Figure 1. Our most important reference cases are the Boltzmann function $\mathcal{F}(\eta) = \exp(\eta)$, the Fermi-Dirac integral of order one half $\mathcal{F}(\eta) = F_{1/2}(\eta)$ and the Blakemore function $\mathcal{F}(\eta) = (\exp(-\eta) + \gamma)^{-1}$ with $\gamma \geq 0$. For large negative arguments all of these functions have a Boltzmann tail. Choosing a parameter of $\gamma = 0.27$, we can approximate the Fermi-Dirac integral with a Blakemore



Figure 1: Left: Standard statistical distribution functions for classical 3D semiconductors as well as organic semiconductors described by a Gaussian density of states with variance σ (the energetic disorder). Both can be approximated with the help of the Blakemore or the Boltzmann distribution function. Right column: visualization of different statistical distribution functions \mathcal{F} and corresponding diffusion enhancements g.

function fairly accurately up to $\eta \leq 1.5$, see Figure 1. For $\gamma = 0$, Blakemore reduces to Boltzmann. Thus, we can think of γ as a parameter controlling how close we are to the Boltzmann regime. Finally, we assume that \mathcal{F} is analytic with

$$0 < \mathcal{F}'(\eta) \le \mathcal{F}(\eta) \le \exp(\eta) \tag{3}$$

for all $\eta \in \mathbb{R}$. This assumption is reasonable for virtually all realistic device setups.

The current densities in (1b) and (1c) are given by

$$\mathbf{j}_n = -q\mu_n n \nabla \varphi_n, \qquad \mathbf{j}_p = -q\mu_p p \nabla \varphi_p. \tag{4}$$

These fluxes can also be written in drift-diffusion form

$$\mathbf{j}_n = -q\mu_n n\nabla\psi + qD_n\nabla n, \qquad \mathbf{j}_p = -q\mu_p p\nabla\psi - qD_p\nabla p.$$
(5)

The diffusion coefficients D_n and D_p are related to the carrier mobilities μ_n and μ_p by a generalized Einstein relation ($U_T = k_B T/q$ denotes the thermal voltage)

$$\frac{D_n}{\mu_n} = U_T g\left(\eta_n\right), \qquad \frac{D_p}{\mu_p} = U_T g\left(\eta_p\right) \tag{6}$$

where

$$\eta_n = \mathcal{F}^{-1}\left(\frac{n}{N_c}\right), \qquad \eta_p = \mathcal{F}^{-1}\left(\frac{p}{N_v}\right)$$
(7)

and g denotes a nonlinear function of the form

$$g(\xi) = \frac{\mathcal{F}(\xi)}{\mathcal{F}'(\xi)},$$

leading in general to a nonlinear diffusion coefficient. As proposed in [19], we call this function *diffusion enhancement*. If we want to stress the dependency on the charge carrier density, we can rewrite the diffusion enhancement equivalently to $g(n/Nc) = (n/N_c)(\mathcal{F}^{-1})'(n/N_c)$ in the case of electrons for example. For the Boltzmann approximation, we immediately see that $g(\xi) = 1$, which gives the classical Einstein relation $D_n = q\mu_n/(k_BT)$. For the Blakemore distribution function, we have $g(\xi) = 1 + \gamma \exp(\xi)$. In general, for non-exponential distribution functions we note by our assumptions (3) that $g(\xi) \geq 1$. The diffusion enhancement measures how well the semiconductor can be described by Boltzmann statistics. The larger the diffusion enhancement, the less accurate it is to assume Boltzmann statistics. For crystalline semiconductors assuming Boltzmann statistics becomes inaccurate for large densities (e.g. due to high doping or low temperatures). On the other hand, for organic semiconductors this effect becomes already prevalent for low or moderate densities [20, 21]. The diffusion enhancement functions for different distribution functions are shown in Figure 1.

The system (1) is supplied with mixed Dirichlet-Neumann boundary conditions. Since we focus on the flux approximation, we refer the reader for more details to [22, 23].

3 Discretization of the van Roosbroeck system

There are various approaches to discretize the van Roosbroeck system. Historically, the first numerical discretizations of the van Roosbroeck system were based on finite differences [2, 23]. Due to their inflexibility on arbitrary domains, another promising technique is based on finite elements [24, 25]. However, finite elements are known to run into stability issues when drift dominates diffusion [26, 27, 28] in which case one has to work with a different set of basic variables or stabilize this method. Unfortunately, there appears to be no stabilized finite element method which is provably guaranteed to satisfy the maximum principle or does not produce spurious oscillations [29]. In order to obtain stable discretizations of van Roosbroeck system even though the solution rapidly varies several order of magnitudes, finite volume schemes are well-established [23, 30]. Hence, we focus on a Voronoï finite volume technique here. More details concerning its use in the context of semiconductor device simulation can be found in [5, 16, 22, 23].

We start by partitioning the domain Ω into non-intersecting, convex polyhedral control volumes ω_K such that $\Omega = \bigcup_{K=1}^N \omega_K$. With each control volume we associate a node $\mathbf{x}_K \in \omega_K$. For every boundary intersecting control volume, we demand that this node lies on the boundary $\mathbf{x}_K \in \partial\Omega \cap \omega_K$. Assuming that the partition is *admissible* in the sense of [30], that is the *edge* $\overline{\mathbf{x}_K \mathbf{x}_L}$ of length h_{KL} is orthogonal to $\partial \omega_k \cap \partial \omega_L$, the normal vectors to $\partial \omega_K$ can be calculated by $\mathbf{n}_{KL} = (\mathbf{x}_L - \mathbf{x}_K)/||\mathbf{x}_L - \mathbf{x}_K||$. The notation is explained visually in Figure 2. In order to keep the following discretization of the van Roosbroeck system readable, we introduce two abbreviations used as arguments of the function \mathcal{F} ,

$$\eta_n\left(\psi,\varphi_n\right) = \frac{q\left(\psi-\varphi_n\right) - E_c}{k_B T} \quad \text{and} \quad \eta_p\left(\psi,\varphi_p\right) = \frac{E_v - q\left(\psi-\varphi_n\right)}{k_B T}.$$
 (8)

We note that these definitions are consistent with (7). With the help of both



Figure 2: Two adjacent control volumes ω_K and ω_L with corresponding notation.

of these expressions, we derive a finite volume based discretization of the van Roosbroeck system. For each control volume ω_K , we consider the three equations:

$$\sum_{\omega_{L}\in\mathcal{N}(\omega_{K})}\left|\partial\omega_{K}\cap\partial\omega_{L}\right|D_{K,L}=q\left|\omega_{K}\right|\left(C_{K}-N_{c}\mathcal{F}\left(\eta_{n}\left(\psi_{K},\varphi_{n;K}\right)\right)+N_{v}\mathcal{F}\left(\eta_{p}\left(\psi_{K},\varphi_{p;K}\right)\right)\right)$$

(9a)

$$-q|\omega_{K}|N_{c}\frac{d}{dt}\mathcal{F}\left(\eta_{n}\left(\psi_{K},\varphi_{n;K}\right)\right) + \sum_{\omega_{L}\in\mathcal{N}(\omega_{K})}|\partial\omega_{K}\cap\partial\omega_{L}|j_{n;K,L} = q|\omega_{K}|R_{K},$$
(9b)

$$q|\omega_{K}|N_{v}\frac{d}{dt}\mathcal{F}\left(\eta_{p}\left(\psi_{K},\varphi_{p;K}\right)\right) + \sum_{\omega_{L}\in\mathcal{N}(\omega_{K})}|\partial\omega_{K}\cap\partial\omega_{L}|j_{p;K,L} = -q|\omega_{K}|R_{K}.$$
(9c)

We denote with $\mathcal{N}(\omega_K)$ the set of all control volumes neigboring ω_K . In 2D, the measure $|\partial \omega_K \cap \partial \omega_L|$ corresponds to the length of the boundary line segment and in 3D to the area of the intersection of the boundary surfaces. Furthermore, in 2D the measure $|\omega_K|$ is given by the area and in 3D by the volume of the control volume ω_K . The unknowns ψ_K , $\varphi_{n;K}$ and $\varphi_{p;K}$ are function evaluations of the electrostatic potential as well as the quasi Fermi potentials for electrons and holes evaluated at node \mathbf{x}_K . Accordingly, R_K and C_K are defined as

$$C_K = C(\mathbf{x}_K)$$
 and $R_K = R(N_c \mathcal{F}(\eta_n(\psi_K, \varphi_{n;K})), N_v \mathcal{F}(\eta_p(\psi_K, \varphi_{p;k})))).$

Note that the doping profile C and the recombination rate R are known a priori. The numerical fluxes $D_{K,L}$, $j_{n;K,L}$ and $j_{p;K,L}$ approximate respectively $-\varepsilon_0\varepsilon_r\nabla\psi\cdot\mathbf{n}_{KL}$, $\mathbf{j}_n\cdot\mathbf{n}_{KL}$ and $\mathbf{j}_p\cdot\mathbf{n}_{KL}$ on the interfaces between two adjacent control volumes ω_K and ω_L . These fluxes can be expressed as functions depending nonlinearly on the values $\psi_K, \varphi_{n;K}, \varphi_{p;K}$ and $\psi_L, \varphi_{n;L}, \varphi_{p;L}$. The electrostatic displacement flux is approximated by

$$D_{K,L} = -\varepsilon_0 \varepsilon_r \frac{\psi_L - \psi_K}{\|\mathbf{x}_L - \mathbf{x}_K\|}.$$

3.1 Generalized Scharfetter-Gummel schemes

From now on, we restrict ourselves to the electron flux since analogous considerations hold true for the hole flux. By construction of our mesh, it suffices to study the one-dimensional flux j_n along the edge $\overline{\mathbf{x}_K \mathbf{x}_L}$. We make now the assumption that the flux is constant between both nodes and denote it again with j_n . This leads for general distribution functions \mathcal{F} to the ordinary differential equation

$$\frac{d}{dx}j_n = \frac{d}{dx} \left(-q\mu_n N_c \mathcal{F}\left(\eta_n\left(\psi,\varphi_n\right)\right) \frac{d}{dx}\varphi_n \right) = 0$$
(10)

along the edge $\overline{\mathbf{x}_K \mathbf{x}_L}$ with boundary conditions

$$\varphi_n(\mathbf{x}_K) = \varphi_K \text{ and } \varphi_n(\mathbf{x}_L) = \varphi_L$$
 (11)

where φ_K and φ_L are the values of the quasi Fermi potentials at the nodes \mathbf{x}_K and \mathbf{x}_L . Integrating twice leads to the integral equation [6, 8] for the constant and unknown current, namely

$$\int_{\eta_K}^{\eta_L} \left(\frac{j_n/j_0}{\mathcal{F}(\eta)} + \frac{\psi_L - \psi_K}{U_T} \right)^{-1} d\eta = 1,$$
(12)

where $j_0 = q\mu_n N_c \frac{U_T}{h_{KL}}$ and the integration limits are given by $\eta_K = \eta_n (\psi_K, \varphi_K)$ and $\eta_L = \eta_n (\psi_L, \varphi_L)$. For strictly monotonously increasing $\mathcal{F}(\eta)$ this equation has always a unique solution [9]. Since this scheme does not put any additional assumptions on \mathcal{F} other than (3), we will refer to it as *exact* or *generalized* Scharfetter-Gummel scheme.

This integral equation can be solved analytically for the Boltzmann approximation, $\mathcal{F}(\eta) = \exp(\eta)$, yielding the classical Scharfetter-Gummel scheme [2],

$$j_{\rm SG} = B\left(\frac{\psi_L - \psi_K}{U_T}\right)e^{\eta_L} - B\left(-\frac{\psi_L - \psi_K}{U_T}\right)e^{\eta_K},\tag{13}$$

for the non-dimensionalized edge current $j_{SG} = j_n/j_0$. The Bernoulli function is given by $B(x) := x/(e^x - 1)$.

In [8], it was shown that for the Blakemore approximation for the Fermi-Dirac integral of order one half, $\mathcal{F}(\eta) = \frac{1}{\exp(-\eta)+\gamma}$, the integral equation can be integrated, leading to a fixed point equation

$$j_{\text{GENSG}} = B\left(\gamma j_{\text{GENSG}} + \frac{\psi_L - \psi_K}{U_T}\right) e^{\eta_L} - B\left(-\left[\gamma j_{\text{GENSG}} + \frac{\psi_L - \psi_K}{U_T}\right]\right) e^{\eta_K}$$
(14)

for the non-dimensionalized edge current $j_{\text{GENSG}} = j_n/j_0$. The right-hand side is a Scharfetter-Gummel expression where the argument of the Bernoulli function is shifted by γj_{GENSG} . Hence, for $\gamma = 0$ the generalized flux j_{GENSG} reduces to the classical Scharfetter-Gummel scheme (13) since the Blakemore function reduces to the Boltzmann function.

Since the Bernoulli function is strictly decreasing, this fixed point equation possesses a unique solution j_{GENSG} . If we want to use the flux given by (14) in the discrete system (9), we need to solve for the flux j_{GENSG} twice (once for electrons and once for holes) on each discretization edge $\overline{\mathbf{x}_K \mathbf{x}_L}$. A few Newton steps are sufficient to solve this equation iteratively [8]. Even though the implicit equation (14) is restricted to the Blakemore approximation, it provides a useful scheme in the context of organic semiconductors. There the Blakemore function arises naturally as a model for materials with δ -shaped density of states [20, 31], describing a single transport level and approximating the Gauss-Fermi integral shown in Figure 1.

Unfortunately, for a general statistical distribution function (e.g. for Fermi-Dirac statistics), no corresponding equation has been derived so far. Therefore, in [9] it was proposed to use piecewise approximations of \mathcal{F} of Blakemore type or Padé functions in order to obtain piecewise integrable expression from the local boundary value problem. However, this is computationally not cheap. Hence, we introduce modified Scharfetter-Gummel schemes next.

3.2 Modified Scharfetter-Gummel schemes

Instead of trying to solve the integral equation (12) analytically for more general statistics such as Fermi-Dirac statistics, one can also average certain quantities in the flux expressions (4) and (5) in a clever way, which allows to keep the Scharfetter-Gummel structure. We are going to present two such *Boltzmannification* techniques, which can be physically interpreted as either modifying the thermal voltage U_T or the effective density of states N_c . For this reason we call the following schemes *modified* Scharfetter-Gummel schemes.

Bessemoulin-Chatard [13] derived a finite volume scheme for convectiondiffusion problems by averaging the nonlinear diffusion term appropriately. This idea was cast in a physical framework in [14], introducing a logarithmic average of the nonlinear diffusion enhancement

$$g_{KL} = \frac{\eta_L - \eta_K}{\log \mathcal{F}(\eta_L) - \log \mathcal{F}(\eta_K)}$$
(15)

along the discretization edge. Using the generalized Einstein relation (6), one immediately observes that the diffusion enhancement g can be seen as a modification factor for the thermal voltage U_T . Replacing U_T in the Scharfetter-Gummel expression (13) by $U_T^* = U_T g_{KL}$, we deduce the following (non-dimensionalized) modified Scharfetter-Gummel scheme

$$j_{\text{DESG}} = -g_{KL} \left(\mathcal{F}(\eta_K) B\left(-\frac{\psi_L - \psi_K}{U_T g_{KL}}\right) - \mathcal{F}(\eta_L) B\left(\frac{\psi_L - \psi_K}{U_T g_{KL}}\right) \right), \quad (16)$$

approximating the current along the edge. A problematic aspect of this scheme is that even though the diffusion enhancement factor g_{KL} is bounded from below by one, it is not straightforward to compute it on a computer when η_K approaches η_L . In this case one needs to use a suitable regularization strategy.

Instead of replacing the thermal voltage by a suitable average along the edge, it is also possible to approximate $\mathcal{F}(\eta)$ along the edge by an exponential (local Boltzmann approximation) and modify the effective density of states N_c accordingly, that is

$$N_c \mathcal{F}(\eta) \approx N_c^* \exp(\eta).$$

This choice makes it possible to keep the original Scharfetter-Gummel flux (13), only replacing N_c with N_c^* . One choice for the modified density of states is

$$N_c^*(\eta^*) = N_c \frac{\mathcal{F}(\eta^*)}{\exp(\eta^*)},$$

where $\eta^* \in [\eta_K, \eta_L]$, assuming $\eta_K \leq \eta_L$. In practice, we might consider taking the arithmetic or the geometric mean between $N_c(\eta_K)$ and $N_c(\eta_L)$, which leads to two more non-dimensionalized modified Scharfetter-Gummel schemes, namely

$$j_{\text{IACT}} = -\beta_{KL} \left(B \left(-\frac{\psi_L - \psi_K}{U_T} \right) e^{\eta_K} - B \left(\frac{\psi_L - \psi_K}{U_T} \right) e^{\eta_L} \right), \quad (17)$$

where the *inverse activity coefficient* is either given by

$$\beta_{KL} = \frac{1}{2} \left(\frac{\mathcal{F}(\eta_K)}{\exp(\eta_K)} + \frac{\mathcal{F}(\eta_L)}{\exp(\eta_L)} \right)$$
(18)

or

$$\beta_{KL} = \sqrt{\frac{\mathcal{F}(\eta_K) \mathcal{F}(\eta_L)}{\exp(\eta_K) \exp(\eta_L)}}.$$
(19)

Sometimes such a prefactor is also referred to as *degeneracy factor*. The idea behind this scheme was introduced in [15] for the numerical solution of the generalized Nernst-Planck system which is similar to the van Roosbroeck system (1). For the Boltzmann distribution function both inverse activity coefficients become one. For non-exponential distribution functions, on the other hand, they are less than one due to our assumptions (3).

3.3Thermodynamic consistency

All of the presented schemes are consistent with the thermodynamic equilibrium in the following sense: Setting the electron current in (5) to zero and using the generalized Einstein relationship, we deduce $\nabla \varphi_n = \mathbf{0}$ and thus for the electrostatic potential (scaled by U_T) that

$$\nabla \psi = \nabla \eta.$$

Thermodynamic consistency means that we require our numerical current approximation to satisfy an analogous relationship. Namely, setting any of the previously introduced fluxes to zero

$$j = j(\eta_L, \eta_K, \psi_L, \psi_K) = 0$$

shall imply

$$\delta\psi_{KL} = \delta\eta_{KL},\tag{20}$$

(00)

where

$$\delta\eta_{KL} := \eta_L - \eta_K \quad \text{and} \quad \delta\psi_{KL} := \frac{\psi_L - \psi_K}{U_T}.$$
 (21)

Indeed, setting the inverse activity current in (17) to zero and using the identity $B(-x) = e^x B(x)$ yields (20). This works for any average of the inverse activity coefficient β_{KL} . For the diffusion enhanced scheme (20) follows when setting the current in (16) to zero. Upon rearranging one obtains

$$\log\left(\frac{\mathcal{F}(\eta_L)}{\mathcal{F}(\eta_K)}\right) = \frac{\delta\psi_{KL}}{g_{KL}} = \frac{\log\left(\mathcal{F}(\eta_L)/\mathcal{F}(\eta_K)\right)}{\delta\eta_{KL}}\delta\psi_{KL}$$

which simplifies again to $\delta \psi_{KL} = \delta \eta_{KL}$. This scheme is only thermodynamically consistent when averaging the diffusion enhancement as done in (15). The consistency for the exact scheme can be directly inferred from (12) when setting the current j_n equal to zero. Even though the thermodynamic consistency is imperative to obtain accurate numerical simulation results, many authors introduced schemes which do no possess this property [10, 11, 12].

4 Flux expansions

In order to assess the quality of the different flux discretization schemes, we start by looking at their expansions. To this end, it will be useful to introduce the average

$$\bar{\eta}_{KL} = \frac{\eta_L + \eta_K}{2}.$$
(22)

The original values η_K and η_L relate to the new values $\bar{\eta}_{KL}$ and $\delta \eta_{KL}$ as follows

$$\eta_L = \bar{\eta}_{KL} - \frac{\delta \eta_{KL}}{2}$$
 and $\eta_K = \bar{\eta}_{KL} + \frac{\delta \eta_{KL}}{2}$.

Since only the electrostatic potential difference appears in the numerical fluxes, we can think of the non-dimensionalized flux depending on the following quantities

$$j = j(\bar{\eta}_{KL}, \delta\eta_{KL}, \delta\psi_{KL}) = j_n(\bar{\eta}_{KL}, \delta\eta_{KL}, \delta\psi_{KL})/j_0.$$

We expand the fluxes around $\bar{\eta}_{KL}$ in terms of the differences $\delta \eta_{KL}$ and $\delta \psi_{KL}$, which decrease on finer meshes. For a Mathematica notebook to verify the following series expansions, we refer to [18].

4.1 Generalized Scharfetter-Gummel scheme

We would like to derive expansions for the most general discretization scheme when the flux is given implicitly by the integral equation (12). Since in this case no additional assumptions on \mathcal{F} , apart from (3), are made, we can compare the exact flux to the modified ones to study their quality. We consider the generalized Scharfetter-Gummel as the reference flux. We transform integral equation (12) into an implicit equation for the non-dimensionalized flux $j = j_n/j_0$, namely

$$G(j,\delta\psi_{KL},\delta\eta_{KL}) := \frac{1}{2}\delta\eta_{KL} \int_{-1}^{1} \frac{\mathcal{F}\left(\bar{\eta}_{KL} + \frac{1}{2}\delta\eta_{KL}\xi\right)}{j + \delta\psi_{KL}\mathcal{F}\left(\bar{\eta}_{KL} + \frac{1}{2}\delta\eta_{KL}\xi\right)} d\xi - 1 = 0.$$
(23)

Expanding the flux around $\delta \psi_{KL} = 0$ and omitting the argument $\bar{\eta}_{KL}$, we find

$$j(\delta\psi_{KL},\delta\eta_{KL}) = j(0,\delta\eta_{KL}) + j'(0,\delta\eta_{KL})\delta\psi_{KL} + \frac{1}{2}j''(0,\delta\eta_{KL})\delta\psi_{KL}^{2} + \frac{1}{6}j'''(\xi,\delta\eta_{KL})\delta\psi_{KL}^{3}$$
(24)

for some $\xi \in [0, \delta \psi_{KL}]$. The derivative is taken with respect to $\delta \psi_{KL}$. The zeroth-order term on the right-hand side we can compute directly from (23) via

$$j(0,\delta\eta_{KL}) = \frac{1}{2}\delta\eta_{KL} \int_{-1}^{1} \mathcal{F}\left(\bar{\eta}_{KL} + \frac{1}{2}\delta\eta_{KL}\xi\right)d\xi$$

$$= \mathcal{F}(\bar{\eta}_{KL})\delta\eta_{KL} + \frac{1}{24}\mathcal{F}''(\bar{\eta}_{KL})\delta\eta_{KL}^{3} + \mathcal{O}(\delta\eta_{KL}^{5}).$$
 (25)

Since Equation (23) is only given in implicit form, we need to compute j' via implicit differentiation. We find

$$j'(0,\delta\eta_{KL}) = -\left(\frac{\partial G(j,\delta\eta_{KL},\delta\psi_{KL})}{\partial\delta\psi_{KL}} \middle/ \frac{\partial G(j,\delta\eta_{KL},\delta\psi_{KL})}{\partial j} \right) \bigg|_{\delta\psi_{KL}=0}$$

$$= -\left(\frac{\int_{-1}^{1} \frac{\mathcal{F}^{2}(\bar{\eta}_{KL}+1/2\delta\eta_{KL}\xi)}{(\bar{j}+\delta\psi_{KL}\mathcal{F}(\bar{\eta}_{KL}+1/2\delta\eta_{KL}\xi))^{2}}d\xi}{\int_{-1}^{1} \frac{\mathcal{F}(\bar{\eta}_{KL}+1/2\delta\eta_{KL}\xi)}{(\bar{j}+\delta\psi_{KL}\mathcal{F}(\bar{\eta}_{KL}+1/2\delta\eta_{KL}\xi))^{2}}d\xi}\right) \bigg|_{\delta\psi_{KL}=0} \qquad (26)$$

$$= -\frac{\int_{-1}^{1} \mathcal{F}^{2}(\bar{\eta}_{KL}+\frac{1}{2}\delta\eta_{KL}\xi)d\xi}{\int_{-1}^{1} \mathcal{F}(\bar{\eta}_{KL}+\frac{1}{2}\delta\eta_{KL}\xi)d\xi}$$

$$= -\mathcal{F}(\bar{\eta}_{KL}) - \left(\frac{\mathcal{F}'(\bar{\eta}_{KL})^{2}}{12\mathcal{F}(\bar{\eta}_{KL})} + \frac{\mathcal{F}''(\bar{\eta}_{KL})}{24}\right)\delta\eta_{KL}^{2} + \mathcal{O}(\delta\eta_{KL}^{4}),$$

where for the final line we have made the additional assumption that $\delta \eta_{KL}$ is small. Finally, by differentiating the implicit derivative (the term in parentheses in (26)) once more, we obtain

$$j''(0,\delta\eta_{KL}) = \frac{1}{6} \frac{\mathcal{F}'(\bar{\eta}_{KL})^2 \delta\eta_{KL}^2}{j(0)} + \mathcal{O}(\delta\eta_{KL}^4/j(0)) = \frac{1}{6} \frac{\mathcal{F}'(\bar{\eta}_{KL})^2 \delta\eta_{KL}}{\mathcal{F}(\bar{\eta}_{KL})} + \mathcal{O}(\delta\eta_{KL}^3).$$
(27)

Inserting (25), (26) and (27) into (24) leads to

$$j(\delta\psi_{KL},\delta\eta_{KL}) = -\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL} + \mathcal{F}(\bar{\eta}_{KL})\delta\eta_{KL} + \frac{1}{12}\frac{\mathcal{F}'(\bar{\eta}_{KL})^2}{\mathcal{F}(\bar{\eta}_{KL})}\delta\psi_{KL}^2\delta\eta_{KL} - \left(\frac{\mathcal{F}'(\bar{\eta}_{KL})^2}{12\mathcal{F}(\bar{\eta}_{KL})} + \frac{\mathcal{F}''(\bar{\eta}_{KL})}{24}\right)\delta\psi_{KL}\delta\eta_{KL}^2 + \frac{1}{24}\mathcal{F}''(\bar{\eta}_{KL})\delta\eta_{KL}^3 + \mathcal{O}(\delta\eta_{KL}^5) + \mathcal{O}(\delta\psi_{KL}\delta\eta_{KL}^4) + \mathcal{O}(\delta\psi_{KL}^2\delta\eta_{KL}^3) + \mathcal{O}(\delta\psi_{KL}^3).$$
(28)

Note we have expanded all third-order terms except the $\mathcal{O}(\delta \psi_{KL}^3)$ term as it does not appear in any of the following flux expansions. Furthermore, the first-order terms can be rewritten to $-\mathcal{F}(\bar{\eta}_{KL})\frac{\varphi_L-\varphi_K}{U_T}$ since by (8) and (21), we have

$$\frac{\varphi_L - \varphi_K}{U_T} = \frac{\psi_L - \psi_K}{U_T} - (\eta_L - \eta_K).$$

This constitutes a central difference approximation for the electron current (4):

$$j_n = j_0 j \approx -q\mu N_c \mathcal{F}(\bar{\eta}_{KL}) \frac{\varphi_L - \varphi_K}{h_{KL}}.$$
(29)

4.2 Inverse activity schemes

Before expanding the inverse activity fluxes, we rewrite the inverse activity scheme in terms of the new variables:

$$\begin{split} j_{\text{IACT}} &= \beta_{KL} e^{\bar{\eta}_{KL}} \left\{ e^{\frac{\delta \eta_{KL}}{2}} B\left(\delta \psi_{KL}\right) - e^{-\frac{\delta \eta_{KL}}{2}} B\left(-\delta \psi_{KL}\right) \right\} \\ &= \beta_{KL} e^{\bar{\eta}_{KL}} \left\{ \sum_{k=0}^{\infty} \frac{\left(\delta \eta_{KL}/2\right)^k}{k!} B\left(\delta \psi_{KL}\right) - \sum_{k=0}^{\infty} \frac{\left(-\delta \eta_{KL}/2\right)^k}{k!} B\left(-\delta \psi_{KL}\right) \right\} \\ &= \beta_{KL} e^{\bar{\eta}_{KL}} \left\{ \sum_{k=0}^{\infty} \frac{\left(\delta \eta_{KL}/2\right)^{2k}}{(2k)!} \left[B\left(\delta \psi_{KL}\right) - B\left(-\delta \psi_{KL}\right) \right] \right. \\ &\left. - \sum_{k=0}^{\infty} \frac{\left(-\delta \eta_{KL}/2\right)^{2k+1}}{(2k+1)!} \left[B\left(\delta \psi_{KL}\right) - B\left(-\delta \psi_{KL}\right) \right] \right\}. \end{split}$$

Using the properties of the Bernoulli function

$$B(x) - B(-x) = -x$$
 and $B(x) + B(-x) = x \coth(x/2)$,

we can simplify this expression to

$$j_{\text{IACT}} = \beta_{KL} e^{\bar{\eta}_{KL}} \delta \psi_{KL} \Big\{ -\cosh(\delta \eta_{KL}/2) + \sinh(\delta \eta_{KL}/2) \coth(\delta \psi_{KL}/2) \Big\}.$$

Assuming an arithmetic mean for the inverse activity coefficient β_{KL} leads to the expansion

$$\beta_{KL} = \frac{1}{2} \left(\frac{\mathcal{F}(\eta_K)}{e^{\eta_K}} + \frac{\mathcal{F}(\eta_L)}{e^{\eta_L}} \right)$$
$$= e^{-\bar{\eta}_{KL}} \left\{ \mathcal{F}(\bar{\eta}_{KL}) + \frac{1}{2} \left(\mathcal{F}''(\bar{\eta}_{KL}) - 2\mathcal{F}'(\bar{\eta}_{KL}) + \mathcal{F}(\bar{\eta}_{KL}) \right) \delta \eta_{KL}^2 + \mathcal{O}(\delta \eta_{KL}^4) \right\}.$$

On the other hand, a geometrically averaged inverse activity coefficient can be expanded to

$$\beta_{KL} = \sqrt{\frac{\mathcal{F}(\eta_K)\mathcal{F}(\eta_L)}{e^{\eta_K}e^{\eta_L}}}$$
$$= e^{-\bar{\eta}_{KL}} \left\{ \mathcal{F}(\bar{\eta}_{KL}) + \frac{1}{2} \left(\mathcal{F}''(\bar{\eta}_{KL}) - \frac{\mathcal{F}'(\bar{\eta}_{KL})^2}{\mathcal{F}(\bar{\eta}_{KL})} \right) \delta\eta_{KL}^2 + \mathcal{O}(\delta\eta_{KL}^4) \right\}.$$

Hence, using the following expansions

$$\cosh(x/2) = 1 + \mathcal{O}(x^2), \ \sinh(x/2) = \frac{x}{2} + \mathcal{O}(x^3), \ x \coth(x) = 1 + \frac{1}{3}x^2 + \mathcal{O}(x^4),$$
(30)

we derive for an arithmetically averaged activity coefficient

$$j_{\text{IACT}}(\delta\psi_{KL},\delta\eta_{KL}) = -\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL} + \mathcal{F}(\bar{\eta}_{KL})\delta\eta_{KL} + \frac{1}{12}\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL}^{2}\delta\eta_{KL} - \left(\frac{5}{8}\mathcal{F}(\bar{\eta}_{KL}) - \mathcal{F}'(\bar{\eta}_{KL}) + \frac{1}{2}\mathcal{F}''(\bar{\eta}_{KL})\right)\delta\psi_{KL}\delta\eta_{KL}^{2} + \left(\frac{1}{2}\mathcal{F}''(\bar{\eta}_{KL}) - \mathcal{F}'(\bar{\eta}_{KL}) + \frac{13}{24}\mathcal{F}(\bar{\eta}_{KL})\right)\delta\eta_{KL}^{3} + \mathcal{O}\left(\delta\psi_{KL}^{4}\right) + \mathcal{O}\left(\delta\psi_{KL}^{2}\delta\eta_{KL}^{3}\right) + \mathcal{O}\left(\delta\eta_{KL}^{4}\right)$$
(31)

and for a geometrically averaged activity coefficient

$$j_{\text{IACT}}(\delta\psi_{KL},\delta\eta_{KL}) = -\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL} + \mathcal{F}(\bar{\eta}_{KL})\delta\eta_{KL} + \frac{1}{12}\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL}^{2}\delta\eta_{KL} - \left(\frac{1}{2}\mathcal{F}''(\bar{\eta}_{KL}) - \frac{1}{2}\frac{\mathcal{F}'(\bar{\eta}_{KL})^{2}}{\mathcal{F}(\bar{\eta}_{KL})} + \frac{1}{8}\mathcal{F}(\bar{\eta}_{KL})\right)\delta\psi_{KL}\delta\eta_{KL}^{2} + \left(\frac{1}{2}\mathcal{F}''(\bar{\eta}_{KL}) - \frac{1}{2}\frac{\mathcal{F}'(\bar{\eta}_{KL})^{2}}{\mathcal{F}(\bar{\eta}_{KL})} + \frac{1}{24}\mathcal{F}(\bar{\eta}_{KL})\right)\delta\eta_{KL}^{3} + \mathcal{O}\left(\delta\psi_{KL}^{4}\right) + \mathcal{O}\left(\delta\psi_{KL}^{2}\delta\eta_{KL}^{3}\right) + \mathcal{O}\left(\delta\eta_{KL}^{4}\right).$$
(32)

We note that by our assumption (3) the fractions in this expansion are well-defined.

4.3 Diffusion enhanced scheme

Finally, we can apply similar ideas to the diffusion enhanced scheme. On expanding the distribution function \mathcal{F} , we obtain

$$j_{\text{DESG}} = g_{KL} \left\{ \mathcal{F}(\bar{\eta}_{KL} + \delta\eta_{KL}/2) B\left(\frac{\delta\psi_{KL}}{g_{KL}}\right) - \mathcal{F}(\bar{\eta}_{KL} - \delta\eta_{KL}/2) B\left(-\frac{\delta\psi_{KL}}{g_{KL}}\right) \right\}$$
$$= g_{KL} \left\{ -\frac{\delta\psi_{KL}}{g_{KL}} \sum_{k=0}^{\infty} \frac{\mathcal{F}^{(2k)}(\bar{\eta}_{KL})}{(2k)!} \left(\frac{\delta\eta_{KL}}{2}\right)^{2k} + \frac{\delta\psi_{KL}}{g_{KL}} \coth\left(\frac{\delta\psi_{KL}}{2g_{KL}}\right) \sum_{k=0}^{\infty} \frac{\mathcal{F}^{(2k+1)}(\bar{\eta}_{KL})}{(2k+1)!} \left(\frac{\delta\eta_{KL}}{2}\right)^{2k+1} \right\}.$$

The averaged diffusion enhancement coefficient g_{KL} can be expanded as follows

$$g_{KL} = \frac{\mathcal{F}(\bar{\eta}_{KL})}{\mathcal{F}'(\bar{\eta}_{KL})} - \left(\frac{1}{3}\frac{\mathcal{F}'(\bar{\eta}_{KL})}{\mathcal{F}(\bar{\eta}_{KL})} + \frac{1}{6}\frac{\mathcal{F}(\bar{\eta}_{KL})\mathcal{F}'''(\bar{\eta}_{KL})}{\mathcal{F}'(\bar{\eta}_{KL})^2} - \frac{1}{2}\frac{\mathcal{F}''(\bar{\eta}_{KL})}{\mathcal{F}'(\bar{\eta}_{KL})}\right)\delta\eta_{KL}^2 + \mathcal{O}(\delta\eta_{KL}^4).$$
(33)

Combining both equations, gives similarly as before

$$j_{\text{DESG}}(\delta\psi_{KL},\delta\eta_{KL}) = -\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL} + \mathcal{F}(\bar{\eta}_{KL})\delta\eta_{KL} + \frac{1}{12}\frac{\mathcal{F}'(\bar{\eta}_{KL})^2}{\mathcal{F}(\bar{\eta}_{KL})}\delta\psi_{KL}^2\delta\eta_{KL} - \frac{1}{8}\mathcal{F}''(\bar{\eta}_{KL})\delta\psi_{KL}\delta\eta_{KL}^2 + \left(\frac{1}{2}\mathcal{F}''(\bar{\eta}_{KL}) - \frac{1}{3}\frac{\mathcal{F}'(\bar{\eta}_{KL})^2}{\mathcal{F}(\bar{\eta}_{KL})} - \frac{1}{8}\frac{\mathcal{F}(\bar{\eta}_{KL})\mathcal{F}'''(\bar{\eta}_{KL})}{\mathcal{F}'(\bar{\eta}_{KL})}\right)\delta\eta_{KL}^3 + \mathcal{O}\left(\delta\psi_{KL}^4\right) + \mathcal{O}\left(\delta\psi_{KL}^2\delta\eta_{KL}^3\right) + \mathcal{O}\left(\delta\eta_{KL}^5\right).$$
(34)

4.4 Comparison

When comparing the expansions (31), (32) and (34) with the expansion of the exact scheme (28), we can make several observations. There are no zeroth and second-order terms. All first-order terms agree and constitute a central difference approximation of the flux in the sense of (29) thus showing that these discretizations are consistent. Almost all third-order terms are different with the interesting exception of the $\mathcal{O}(\delta \psi_{KL}^2 \delta \eta_{KL})$ term which appears in the diffusion enhanced as well as in the exact scheme. Introducing the diffusion enhancement factor g, the error between the prefactor in (28) and (34) as well as the corresponding term in (31) and (32) may be written as

$$\frac{1}{12}\left(1-\frac{1}{g(\bar{\eta}_{KL})^2}\right)\mathcal{F}(\bar{\eta}_{KL}).$$

We point out that the diffusion enhancement enters quadratically. For Boltzmann statistics the difference vanishes. However, since for physically relevant applications far from the Boltzmann regime the diffusion enhancement g is considerably larger than unity, we deduce that the error from this third-order term is larger for the inverse activity schemes. While this already indicates that the diffusion enhanced flux discretization might hold some advantage over the inverse activity flux discretizations, the argument is not fully conclusive since not all third-order terms cancel. For example, there are no $\mathcal{O}(\delta \psi_{KL}^3)$ terms in the expansions of the modified schemes. This follows from the final expansion in (30) and the structure of the fluxes (31), (32) and (34). Except for the linear term only even powers of $\delta \psi_{KL}$ enter these expansions.

The third-order prefactors are compared numerically in Figure 3 for Boltzmann, Blakemore, Fermi-Dirac statistics. For all third-order prefactors the deviation from the reference flux (28) is shown in the range of $-5 \leq \bar{\eta}_{KL} \leq 5$. The $\mathcal{O}(\delta \psi_{KL}^2 \delta \eta_{KL})$ prefactor is the same for both inverse activity schemes. In the plotted range the magnitude of all prefactors for the geometrically averaged inverse activity scheme is almost always larger than the prefactors for the arithmetically averaged inverse activity scheme. For Boltzmann statistics all schemes yield the same Scharfetter-Gummel flux (13). Hence, these errors (numerically) vanish. Similarly, since for large negative arguments all functions \mathcal{F} are approximated accurately by the Boltzmann approximation, all schemes agree in this regime and the error is small.



Figure 3: Errors of third-order prefactors between modified and exact schemes in the range of $-5 \leq \bar{\eta}_{KL} \leq 5$ for different choices of \mathcal{F} . For Matlab files to generate these figures see [18].

Since neither modified Scharfetter-Gummel scheme is third-order accurate, we derive a second-order accurate error estimate in the next section which helps us to better judge the quality of the different schemes.

5 Error estimate

Since from the previous section it is not really clear which modified Scharfetter-Gummel scheme is the best, we derive now an error estimate, bounding the error between the modified schemes and the exact one with a *second-order* $\mathcal{O}(\delta\psi_{KL}\delta\eta_{KL})$ term.

Theorem 1. Suppose the inverse activity coefficient satisfies

$$\beta_{KL} = e^{-\bar{\eta}_{KL}} \left\{ \mathcal{F}(\bar{\eta}_{KL}) + \mathcal{O}(\delta \eta_{KL}^2) \right\}.$$
(35)

Neglecting third-order terms, the error between the modified fluxes and the exact integral flux j as determined by (23) can be bounded by

$$|j_{IACT} - j| \le \frac{1}{2} \mathcal{F}(\bar{\eta}_{KL}) |\delta \psi_{KL} \delta \eta_{KL}|, \qquad (36)$$

$$|j_{DESG} - j| \le \frac{1}{2} \mathcal{F}'(\bar{\eta}_{KL}) |\delta\psi_{KL}\delta\eta_{KL}| = \frac{1}{2} \frac{\mathcal{F}(\bar{\eta}_{KL})}{g(\bar{\eta}_{KL})} |\delta\psi_{KL}\delta\eta_{KL}|.$$
(37)

Proof. First we note that using expansions (30) we can derive for the inverse activity flux as well as for diffusion enhanced flux (16) the expansions

$$j_{\text{IACT}} = -\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL} + \frac{\delta\psi_{KL}}{2}\coth\left(\frac{\delta\psi_{KL}}{2}\right)\mathcal{F}(\bar{\eta}_{KL})\delta\eta_{KL} + \mathcal{O}(\delta\psi_{KL}\delta\eta_{KL}^2) + \mathcal{O}(\delta\eta_{KL}^3)$$
(38)

as well as

$$j_{\text{DESG}} = -\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL} + \frac{\delta\psi_{KL}}{2}\coth\left(\frac{\delta\psi_{KL}}{2g_{KL}}\right)\mathcal{F}'(\bar{\eta}_{KL})\delta\eta_{KL} + \mathcal{O}(\delta\psi_{KL}\delta\eta_{KL}^2) + \mathcal{O}(\delta\eta_{KL}^3).$$
(39)

Next, it is useful to note that

$$1 \le \frac{x}{2} \coth(x/2) \le 1 + \frac{|x|}{2}$$
 (40)

for any $x \in \mathbb{R}$. Hence, we can bound the inverse activity flux discretization by

$$\begin{aligned} j_{\text{IACT}} &\leq -\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL} + \left(1 + \frac{1}{2}\left|\delta\psi_{KL}\right|\right)\mathcal{F}(\bar{\eta}_{KL})\delta\eta_{KL} + \mathcal{O}(\delta\psi_{KL}\delta\eta_{KL}^2) + \mathcal{O}(\delta\eta_{KL}^3) \\ j_{\text{IACT}} &\geq -\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL} + \mathcal{F}(\bar{\eta}_{KL})\delta\eta_{KL} + \mathcal{O}(\delta\psi_{KL}\delta\eta_{KL}^2) + \mathcal{O}(\delta\eta_{KL}^3). \end{aligned}$$

From this we deduce the following bounds for the error

$$\begin{aligned} j_{\text{IACT}} - j &\leq \frac{1}{2} \mathcal{F}(\bar{\eta}_{KL}) |\delta\psi_{KL} \delta\eta_{KL}| + \mathcal{O}(\delta\psi_{KL} \delta\eta_{KL}^2) + \mathcal{O}(\delta\eta_{KL}^3) + \mathcal{O}(\delta\psi_{KL}^3) + \mathcal{O}(\delta\psi_{KL}^2 \delta\eta_{KL}) \\ j - j_{\text{IACT}} &\leq \mathcal{O}(\delta\psi_{KL} \delta\eta_{KL}^2) + \mathcal{O}(\delta\eta_{KL}^3) + \mathcal{O}(\delta\psi_{KL}^3) + \mathcal{O}(\delta\psi_{KL}^2 \delta\eta_{KL}). \end{aligned}$$

That is, neglecting third-order terms and using (3), we have the error estimate

$$|j_{\text{IACT}} - j| \leq \frac{1}{2} \mathcal{F}(\bar{\eta}_{KL}) |\delta \psi_{KL} \delta \eta_{KL}|.$$

On the other hand, applying (40) to the diffusion enhanced flux (39), we derive (neglecting third-order terms straight away)

$$j_{\text{DESG}} = -\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL} + g_{KL}\left(\frac{\delta\psi_{KL}}{2g_{KL}}\right)\coth\left(\frac{\delta\psi_{KL}}{2g_{KL}}\right)\mathcal{F}'(\bar{\eta}_{KL})\delta\eta_{KL}$$
$$\leq -\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL} + \left(g_{KL} + \frac{1}{2}\left|\delta\psi_{KL}\right|\right)\mathcal{F}'(\bar{\eta}_{KL})\delta\eta_{KL}$$
$$= -\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL} + \mathcal{F}(\bar{\eta}_{KL})\delta\eta_{KL} + \frac{1}{2}\left|\delta\psi_{KL}\right|\delta\eta_{KL}\mathcal{F}'(\bar{\eta}_{KL}) + \mathcal{O}(\delta\eta_{KL}^3)$$

where we have used (33) in the last line. Similarly, we find

$$j_{\text{DESG}} \ge -\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL} + \mathcal{F}(\bar{\eta}_{KL})\delta\eta_{KL} + \mathcal{O}(\delta\eta_{KL}^3).$$

From both previous estimates we obtain, neglecting third-order terms,

$$|j_{\text{DESG}} - j| \le \frac{1}{2} \mathcal{F}'(\bar{\eta}_{KL}) |\delta \psi_{KL} \delta \eta_{KL}|.$$

Finally, we note that due to the definition of the diffusion enhancement

$$g(\eta_n) = \frac{\mathcal{F}(\eta_n)}{\mathcal{F}'(\eta_n)},$$

the claim follows.



Figure 4: Logarithmic absolute errors between the generalized Scharfetter-Gummel and the diffusion enhanced scheme (left), the arithmetically averaged inverse activity scheme (middle) and the geometrically averaged one (right) for $\delta\psi_{KL} = U_T$ (first row) and $\delta\psi_{KL} = 5U_T$ (second row). The dashed lines show where generalized and modified schemes agree exactly. The bold black lines highlight the same contour levels in all the plots. For Matlab files to generate these figures see [18].

The previous theorem shows that the main difference between both error estimates is that the diffusion enhancement appears favorably in the error bound of the diffusion enhanced scheme. So if we are far from the Boltzmann regime (where the diffusion enhancement is significantly larger than one), the second error bound is guaranteed to perform better. This is not just a theoretical result but has direct consequences for numerical computations as we will see in the next section with the help of an example. It is also important to note that in the proof the concrete form of the inverse activity coefficient did not matter. For any inverse activity coefficient which admits an expansion of the form (35) the result holds. Thus, this result applies to even more general averages for the inverse activity coefficient than just the arithmetic or the geometric mean.

6 Numerical flux comparison

In order to see the implications of Theorem 1, we plot the error between the modified and the generalized Scharfetter-Gummel flux approximation schemes in this section. We think of the generalized Scharfetter-Gummel as the reference solution since it solves the two-point boundary value problem exactly. So ideally, we would like to come as close as possible to the generalized Scharfetter-Gummel scheme using one of the modified ones. One of the simplest non-Boltzmann distribution function is the Blakemore approximation. As discussed before, it has the additional advantage that we do not need to solve the integral equation (12). It suffices to solve the implicit equation (14). For this reason, we focus on the Blakemore distribution function in this and the following section.

Figures 4 and 5 show the logarithmic absolute errors between the inverse



Figure 5: Logarithmic absolute errors between the generalized Scharfetter-Gummel and the diffusion enhanced scheme (left), the arithmetically averaged inverse activity scheme (middle) and the geometrically averaged one (right) for $\bar{\eta}_{KL} = 1.5$ (first row) and $\bar{\eta}_{KL} = 5$ (second row). The dashed lines show where generalized and modified schemes agree exactly. The bold black lines highlight the same contour levels in all the plots. For Matlab files to generate the figures see [18].

activity and the diffusion enhanced flux approximation scheme (for electrons) with respect to the generalized Scharfetter-Gummel scheme. To be more precise, for the diffusion enhanced and generalized Scharfetter-Gummel flux, for example, the error is defined by

$\operatorname{sign}(j_{\text{DESG}} - j_{\text{GENSG}}) \log(1 + |j_{\text{DESG}} - j_{\text{GENSG}}|).$

Figure 4 shows the error depending on η_K and η_L for a fixed value of the potential difference $\delta\psi_{KL}$. The figures illustrate the fact that for large negative η_K and η_L all schemes agree since they coincide with the classical Scharfetter-Gummel scheme in this regime. Additionally, they agree when there is no diffusion (pure drift current) in which case $\eta_K = \eta_L$ as well as when $\delta\psi_{KL} = \delta\eta_{KL}$, due to the consistency with the thermodynamic equilibrium which we discussed in Section 3.3. Both of these special cases are indicated by the dashed lines. The thick black lines highlight the same contour levels. The chosen value for $\delta\psi_{KL}$ has a moderate influence on the error, mostly by shifting the dashed line along which the schemes are thermodynamically consistent.

On the other hand, Figure 5 shows the errors in terms of $\delta\eta_{KL}$ and $\delta\psi_{KL}$ for a fixed value of $\bar{\eta}_{KL}$. Again, the dashed lines indicate where $\eta_K = \eta_L$ as well as $\delta\psi_{KL} = \delta\eta_{KL}$. This time, however, the value for $\bar{\eta}_{KL}$ drastically changes the behavior of the error. The diffusion enhancement for these values is given by g(1.5) = 2.21 and g(5) = 41.07, indicating that for these choices of $\bar{\eta}_{KL}$ the Boltzmann approximation is clearly no longer valid. For the first value the Blakemore distribution function is still a good approximation of the Fermi-Dirac distribution function [7], see Figure 1. The second choice illustrates the case of large diffusion enhancement as is typical for organic semiconductors with large disorder or semiconductors operating a cryogenic temperatures [32]. From Figure 5, it is clear that the diffusion enhancement is closest to the reference solution, the generalized Scharfetter-Gummel scheme. Both inverse activity schemes yield larger errors of approximately the same size which do not seem to depend on $\bar{\eta}_{KL}$ very much. This behavior is explained by our error estimates in Theorem 1 since the diffusion enhancement enters inversely in the bound for the diffusion enhanced flux whereas in the bound for the inverse activity flux it does not enter at all.

7 Benchmark simulation

Up to now we have only discussed local flux approximations but did not study the impact on the fully coupled van Roosbroeck system. In this section, we carefully design a relatively simple benchmark which shows that the local flux approximations can have a strong influence on the total current flowing through a device or even the electrostatic potential.

7.1 Designing the benchmark

Since all schemes are consistent with the thermodynamic equilibrium, we can only expect large differences between the schemes when considering a system which is far from equilibrium. In the case of electrons, this corresponds to maximizing the absolute value of the non-dimensionalized difference in quasi Fermi potentials $\delta \varphi_{KL} := (\varphi_L - \varphi_K)/U_T$, which for electrons is given by

$$\delta\varphi_{KL} = -\delta\eta_{KL} + \delta\psi_{KL}.\tag{41}$$

Assuming that the electrostatic potential is linear, we can relate the local potential difference $\delta \psi_{KL}$ between two neighboring cells in a one-dimensional device to the global bias $\delta \psi$, using the formula

$$\delta\psi_{KL} = \frac{h_{KL}}{h}\delta\psi.$$

As before h_{KL} denotes the local distance between node K and L. The total length of the device is given by h. From this we can deduce that we need a large global bias and a coarse mesh. On the other hand, for electrons the difference $\delta\eta_{KL}$ needs to have the opposite sign of the electrostatic potential difference $\delta\psi_{KL}$. This behavior can be clearly observed in Figure 5, where the signed logarithmic absolute errors between the different schemes are compared. Whenever $\delta\eta_{KL}$ and $\delta\psi_{KL}$ are large and have *opposite* signs, the flux error becomes large. A large difference in $\delta\eta_{KL}$ corresponds to a large density contrast.

We point out that the electrostatic potential difference in (41) is normalized with the thermal voltage $U_T = k_B T/q$. Hence, one can also increase the quasi Fermi potential difference by decreasing the temperature T.

7.2 A p-i-n device

To study the differences of the previous schemes, we simulate a simple diode, depicted in Figure 6, consisting of p-doped, intrinsic and n-doped regions. The example is chosen in such a way that we incorporate the observations from the previous subsection to guarantee that the device is sufficiently far from



Figure 6: The p-i-n device used for the numerical calculations. The p-contact is on the left and the n-contact is on the right.



Figure 7: The IV curves computed with the different schemes for different mesh refinements. The reference solution (black) was computed using the generalized Scharfetter-Gummel scheme on refinement level 12. For the data sets see [18].



Figure 8: Convergence studies for the L_{∞} errors of the total current (left) and the electrostatic potential for a fixed bias of 3V (right). The left graph shows the maximal error with respect to bias and the right one with respect to space. For the data sets see [18].

thermodynamic equilibrium (large global bias, coarse mesh and the fact that at least some local $\delta\eta_{KL}$ and $\delta\psi_{KL}$ have opposite signs). We study a gallium arsenide p-i-n diode, using the following parameters for the calculations: band gap $E_c - E_v = 1.424 \,\mathrm{eV}$, density of states $N_c = 4.35 \times 10^{17} \,\mathrm{cm}^{-3}$, $N_v = 9.14 \times 10^{18} \,\mathrm{cm}^{-3}$, mobilities $\mu_n = 8500 \,\mathrm{cm}^2/(\mathrm{Vs})$, $\mu_p = 400 \,\mathrm{cm}^2/(\mathrm{Vs})$, temperature $T = 300 \,\mathrm{K}$, relative permittivity $\varepsilon_r = 12.9$, n-doping $N_D = 4.35 \times 10^{17} \,\mathrm{cm}^{-3}$, p-doping $N_A = 4.20 \times 10^{18} \,\mathrm{cm}^{-3}$, intrinsic region $N_I = 0 \,\mathrm{cm}^{-3}$ and no recombination R = 0. Each subregion is $2 \,\mu m$ long. The calculations were carried out with the software package ddfermi [33]. All of the discussed schemes have been implemented in this package.

To be able to compare all three schemes, we choose the Blakemore distribution function since in this case the integral equation (12) simplifies to the implicit equation (14). In Figure 7, we show the IV curves (total current for a given bias) for the different schemes and mesh resolutions. The total current is a global and highly relevant quantity from the engineering point of view. For large biases, we can see that the schemes behave quite differently. The reference solution has been computed with the generalized Scharfetter-Gummel scheme using refinement level 12. The refinement level n corresponds to a uniform mesh width

$$h_n = \frac{2}{3 \cdot 2^n - 1} \mu \mathbf{m}.$$
 (42)

That is, for each new refinement level the uniform mesh size is halved.

For the coarsest refinement level the generalized Scharfetter-Gummel scheme already shows the classical p-i-n behavior, blocking the current up to a threshold value and then exhibiting a linear current growth. The diffusion enhanced and the inverse activity scheme (using the arithmetic mean for the inverse activity coefficient) do not capture the asymptotics accurately. The best seems to be the inverse activity scheme with geometrically averaged coefficient. However, this effect does not prevail on finer meshes as the next two refinement levels reveal. More clearly, this can be seen when studying the L_{∞} error for the total current taken with respect to the bias value, see the left graph in Figure 8. The first nine refinement levels are shown. Except for the first refinement level the diffusion enhanced scheme yields the smallest error among all modified schemes. However, we need to bear the following in mind: Whereas the exact scheme (which unsurprisingly gives the smallest error for any given mesh width) is already fairly accurate on the coarsest mesh, the modified schemes have still a relatively large error margin. Even though it is more costly to implement, this additional accuracy might be indispensable for 3D simulations where one may only work with coarse meshes. Eventually all schemes converge quadratically. However both inverse activity schemes converge more slowly than the diffusion enhanced and the exact scheme.

Figure 8 shows a convergence plot for the spatial L_{∞} error of the electrostatic potential. Here, the diffusion enhanced scheme seems to behave only significantly better on coarse meshes. This behavior is expected as all schemes should eventually converge to the same solution as the mesh width tends to zero. We point out that the Poisson equation is not directly influenced by the different current approximations. They only enter in the right-hand side via the charge carrier densities.

8 Conclusion and Outlook

In this paper, we have compared three charge carrier flux approximations for more general statistics than the Boltzmann approximation. The diffusion enhancement, a key indicator for how valid the Boltzmann approximation is, helped to judge the quality of the different flux approximations. We compared two types of modified schemes with the exact solution of the local two-point boundary value problem. It was shown that for the diffusion enhanced scheme even a third-order term agreed with the expansion of the exact flux. Furthermore, a second-order accurate error estimate revealed that the bound for the diffusion enhancement scheme depends more favorably on the diffusion enhancement than the inverse activity schemes, making a strong case for the former scheme. Finally, we analyzed the impact of the local flux approximations on the solution of the fully coupled van Roosbroeck system. For this we designed and studied a pi-n benchmark, which proved that under certain conditions the local fluxes have a direct impact on the solution and the overall current. The L_{∞} errors for the total current and the electrostatic potential showed that among the presented modified Scharfetter-Gummel schemes the diffusion enhanced one converged fastest. Hence, it appears that all in all the diffusion enhanced scheme bears significant advantages over the inverse activity scheme. These results are particularly important when the diffusion enhancement is large. This effect appears, for example, in organic semiconductors or when conventional semiconductors have to operate at cryogenic temperatures. Furthermore, if computational resources are limited (say for complex 3D geometries) the extra accuracy is crucial.

Several points can be the focus of future research: Firstly, the impact of the different flux approximations on multidimensional devices may be studied. Secondly, the stability of the linear systems which need to be solved during the Newton iterations can be analyzed for each flux approximation. And finally, the implementation aspect is worth looking into. This may not be trivial for the diffusion enhanced scheme as the computation of the diffusion enhancement factor (15) leads to numerical complications if the denominator becomes small.

References

- W. Van Roosbroeck, Theory of the flow of electrons and holes in germanium and other semiconductors, Bell System Technical Journal 29 (4) (1950) 560-607. doi:10.1002/j.1538-7305.1950.tb03653.x.
- [2] D. Scharfetter, H. Gummel, Large-signal analysis of a silicon read diode oscillator, IEEE Transactions on Electron Devices 16 (1) (1969) 64-77. doi:10.1109/T-ED.1969.16566.
- H. Gajewski, K. Gärtner, On the discretization of van roosbroeck's equations with magnetic field, ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik 76 (5) (1996) 247– 264. doi:10.1002/zamm.19960760502.
- [4] R. Kosik, P. Fleischmann, B. Haindl, P. Pietra, S. Selberherr, On the interplay between meshing and discretization in three-dimensional diffusion

simulation, Trans. Comp.-Aided Des. Integ. Cir. Sys. 19 (11) (2006) 1233–1240. doi:10.1109/43.892848.

- [5] K. Gärtner, Existence of bounded discrete steady-state solutions of the van roosbroeck system on boundary conforming delaunay grids, SIAM Journal on Scientific Computing 31 (2) (2009) 1347–1362. doi:10.1137/070710950.
- [6] R. Eymard, J. Fuhrmann, K. Gärtner, A finite volume scheme for nonlinear parabolic equations derived from one-dimensional local Dirichlet problems, Numerische Mathematik 102 (3) (2006) 463–495.
- [7] J. Blakemore, The Parameters of Partially Degenerate Semiconductors, Proc. Phys. Soc. London A 65 (1952) 460–461.
- [8] T. Koprucki, K. Gärtner, Discretization scheme for drift-diffusion equations with strong diffusion enhancement, Optical and Quantum Electronics 45 (7) (2013) 791–796. doi:10.1007/s11082-013-9673-5.
- [9] K. Gärtner, Existence of bounded discrete steady state solutions of the van Roosbroeck system with monotone Fermi–Dirac statistic functions, Journal of Computational Electronics 14 (3) (2015) 773–787. doi:10.1007/ s10825-015-0712-2.
- [10] O. W. Purbo, D. T. Cassidy, S. H. Chisholm, Numerical model for degenerate and heterostructure semiconductor devices, Journal of Applied Physics 66 (10) (1989) 5078–5082.
- [11] A. Jüngel, Numerical approximation of a drift-diffusion model for semiconductors with nonlinear diffusion, ZAMM 75 (10) (1995) 783–799.
- [12] S. Stodtmann, R. M. Lee, C. K. F. Weiler, A. Badinski, Numerical simulation of organic semiconductor devices with high carrier densities, Journal of Applied Physics 112 (11). doi:10.1063/1.4768710.
- [13] M. Bessemoulin-Chatard, A finite volume scheme for convection-diffusion equations with nonlinear diffusion derived from the Scharfetter-Gummel scheme, Numerische Mathematik 121 (4) (2012) 637–670. doi:10.1007/ s00211-012-0448-x.
- [14] T. Koprucki, N. Rotundo, P. Farrell, D. H. Doan, J. Fuhrmann, On thermodynamic consistency of a Scharfetter–Gummel scheme based on a modified thermal voltage for drift-diffusion equations with diffusion enhancement, Optical and Quantum Electronics 47 (6) (2015) 1327–1332. doi:10.1007/s11082-014-0050-9.
- [15] J. Fuhrmann, Comparison and numerical treatment of generalised Nernst-Planck models, Computer Physics Communications 196 (2015) 166 – 178. doi:10.1016/j.cpc.2015.06.004.
- [16] H. Gajewski, Analysis und Numerik von Ladungstransport in Halbleitern, WIAS Report No. 6, ISSN 0942-9077.
- [17] U. Bandelow, H. Gajewski, R. Hünlich, Fabry–Perot Lasers: Thermodynamics-Based Modeling, in: J. Piprek (Ed.), Optoelectronic Devices, Springer, 2005.

- [18] P. Farrell, T. Koprucki, J. Fuhrmann, Flux Discretizations beyond Boltzmann Statistics - Software and Data Sets (2016). doi:10.20347/WIAS. DATA.1.
- [19] S. L. M. van Mensfoort, R. Coehoorn, Effect of Gaussian disorder on the voltage dependence of the current density in sandwich-type devices based on organic semiconductors, Physical Review B 78 (8). doi:10.1103/physrevb. 78.085207.
- [20] G. Paasch, S. Scheinert, Charge carrier density of organics with Gaussian density of states: Analytical approximation for the Gauss-Fermi integral, J. Appl. Phys. 107 (10) (2010) 104501. doi:10.1063/1.3374475.
- [21] Y. Roichman, N. Tessler, Generalized einstein relation for disordered semiconductors—implications for device performance, Applied Physics Letters 80 (11) (2002) 1948–1950. doi:10.1063/1.1461419.
- [22] P. Farrell, N. Rotundo, D. H. Doan, M. Kantner, J. Fuhrmann, T. Koprucki, Electronics: Numerical methods for drift-diffusion models, in: Handbook of Optoelectronic Device Modeling and Simulation, Taylor & Francis, 2017, Ch. 51.
- [23] S. Selberherr, Analysis and simulation of semiconductor devices, Springer, Wien, New York, 1984.
- [24] M. Auf der Maur, M. Povolotskyi, F. Sacconi, A. Pecchia, G. Romano, G. Penazzi, A. Di Carlo, TiberCAD: towards multiscale simulation of optoelectronic devices, Optical and quantum electronics 40 (14-15) (2008) 1077–1083.
- [25] M. Harmon, I. M. Gamba, K. Ren, Numerical algorithms based on Galerkin methods for the modeling of reactive interfaces in photoelectrochemical (PEC) solar cells, Journal for Computational Physics (2016) accepted.
- [26] K. Morton, Numerical Solution Of Convection-Diffusion Problems, Applied Mathematics, Taylor & Francis, 1996.
- [27] H.-G. Roos, M. Stynes, L. Tobiska, Numerical methods for singularly perturbed differential equations : convection-diffusion and flow problems, Springer series in computational mathematics, Springer, Berlin, Heidelberg, New York, 1996.
- [28] M. Stynes, Steady-state convection-diffusion problems, Acta Numerica 14 (2005) 445–508. doi:10.1017/S0962492904000261.
- [29] V. John, P. Knobloch, On spurious oscillations at layers diminishing (SOLD) methods for convection-diffusion equations: Part I–A review, Computer Methods in Applied Mechanics and Engineering 196 (17) (2007) 2197–2215.
- [30] R. Eymard, T. Gallouët, R. Herbin, Finite volume methods, in: Solution of Equation in ℝⁿ (Part 3), Techniques of Scientific Computing (Part 3), Vol. 7 of Handbook of Numerical Analysis, Elsevier, 2000, pp. 713 – 1018.

- [31] M. Gruber, E. Zojer, F. Schürrer, K. Zojer, Impact of materials versus geometric parameters on the contact resistance in organic thin-film transistors, Advanced Functional Materials 23 (23) (2013) 2941–2952. doi:10.1002/adfm.201203250.
- [32] M. Kantner, T. Koprucki, Numerical simulation of carrier transport in semiconductor devices at cryogenic temperatures, WIAS Preprint No. 2296.
- [33] D. H. Doan, P. Farrell, J. Fuhrmann, M. Kantner, T. Koprucki, N. Rotundo, ddfermi – a drift-diffusion simulation tool, Version: 0.1.0, Weierstrass Institute (WIAS) (2016). doi:10.20347/WIAS.SOFTWARE.14.