Weierstraß-Institut

für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Highly accurate quadrature-based Scharfetter–Gummel schemes for charge transport in degenerate semiconductors

Matteo Patriarca[1], Patricio Farrell[2], Jürgen Fuhrmann[3], Thomas Koprucki[3]

submitted: March 27, 2018

[1] University of Rome "Tor Vergata"
    Dept. Electronics Engineering
    Via del Politecnico 1
    00133 Rome, Italy
    E-Mail: matteo.patriarca@uniroma2.it

[2] TU Hamburg-Harburg
    Institut für Mathematik
    Am Schwarzenberg-Campus 3
    21073 Hamburg, Germany
    E-Mail: patricio.farrell@wias-berlin.de

[3] Weierstrass Institute
    Mohrenstr. 39
    10117 Berlin
    Germany
    E-Mail: juergen.fuhrmann@wias-berlin.de
           thomas.koprucki@wias-berlin.de

WI AS

Highly accurate quadrature-based Scharfetter–Gummel schemes for charge transport in degenerate semiconductors

Matteo Patriarca, Patricio Farrell, Jürgen Fuhrmann, Thomas Koprucki

Abstract

We introduce a family of two point flux expressions for charge carrier transport described by drift-diffusion problems in degenerate semiconductors with non-Boltzmann statistics which can be used in Voronoï finite volume discretizations. In the case of Boltzmann statistics, Scharfetter and Gummel derived such fluxes by solving a linear two point boundary value problem yielding a closed form expression for the flux. Instead, a generalization of this approach to the nonlinear case yields a flux value given implicitly as the solution of a nonlinear integral equation. We examine the solution of this integral equation numerically via quadrature rules to approximate the integral as well as Newton's method to solve the resulting approximate integral equation. This approach results into a family of quadrature-based Scharfetter-Gummel flux approximations. We focus on four quadrature rules and compare the resulting schemes with respect to execution time and accuracy. A convergence study reveals that the solution of the approximate integral equation converges exponentially in terms of the number of quadrature points. With very few integration nodes they are already more accurate than a state-of-the-art reference flux, especially in the challenging physical scenario of high nonlinear diffusion. Finally, we show that thermodynamic consistency is practically guaranteed.

1 Introduction

In 1950, van Roosbroeck formulated a nonlinear system of partial differential equations to model the flow of electrons and holes in semiconductor devices [1]. The numerical solution of these semiconductor device equations gained importance in the nineteen-sixties after Scharfetter and Gummel [2] proposed a 1D finite difference scheme to approximate the carrier flow in semiconductor devices assuming Boltzmann statistics, which results in a linear drift-diffusion equation for the carrier flux. The flux between two discretization nodes is obtained by solving a two point boundary value problem for the flux which is assumed to be constant between these nodes. In combination with a Voronoï finite volume method, this ansatz has been generalized to higher space dimensions [3]. The Scharfetter-Gummel flux is thermodynamically consistent, meaning that constant quasi Fermi potentials lead to vanishing currents. Unfortunately, this classical scheme approximates only charge-carrier fluxes in non-degenerate semiconductors well. When degeneracy effects are physically important, e.g. at cryogenic temperatures [4], for large doping concentrations or in organic semiconductor materials [5, 6], non-Boltzmann (i.e Fermi-Dirac or Gauss-Fermi) statistics are required. In this case, the diffusion enhancement – a function depending on the unknown carrier concentration – has to be multiplied with the diffusion coefficient, rendering the drift-diffusion problem nonlinear.

One promising technique to deal with this type of nonlinear diffusion is to introduce a logarithmic average of the diffusion enhancement function along a discretization edge [7, 8], leading to a *modified Scharfetter-Gummel scheme*. Just as the classical Scharfetter-Gummel flux it is thermodynamically consistent. The authors in [9, 10] give a comprehensive study of the finite volume approximation using

thermodynamically consistent numerical fluxes based on this and other ideas for modified Scharfetter-Gummel schemes.

In [11] the idea of Scharfetter and Gummel was generalized to a large class of scalar nonlinear drift-diffusion equations. In the context of degenerate semiconductors, this approach is referred to as *generalized Scharfetter-Gummel scheme*. The high accuracy of the resulting flux expression is especially important in higher space dimensions when one cannot use very many discretization nodes in each spatial direction. Moreover, it is useful to have such highly accurate schemes in oder to benchmark approximate flux schemes. This generalized Scharfetter-Gummel flux is given implicitly via a nonlinear integral equation. For the Blakemore approximation of the Fermi-Dirac statistic function, this integral equation conveniently reduces to an implicit scalar nonlinear equation [12]. Starting from this, Gärtner [13] proposed a piecewise approximation of general statistic functions by Blakemore type of functions. This approach preserves the thermodynamic properties of the original, continuous model. However, it can become cumbersome when many subdivisions are needed. Furthermore, robust subdivision strategies for different distribution functions or different potential value differences still have to be identified.

For this reason, we present in the following an algorithm to approximate the generalized Scharfetter-Gummel flux directly. It uses one of four quadrature rules to approximate the integral in the integral equation and Newton's method to solve the resulting approximate integral equation, yielding four different schemes. We will refer to them as *quadrature-based Scharfetter-Gummel schemes*. As we will show in this paper, for a moderate amount of quadrature points, they are practically exact.

This paper extends recent work [14] where the beneficial behavior of numerically approximating the integral equation via Gauss-Legendre quadrature was noted but not explained in detail. Now we not only study additionally Clenshaw-Curtis, Gauss-Lobatto and (adaptive) Gauss-Kronrod quadrature but also provide more comprehensive numerical convergence studies in terms of ℓ_2 and ℓ_∞ errors. Moreover, we compare how efficiently the different schemes perform for the physically relevant Blakemore, Fermi-Dirac and Gauss-Fermi statistics, using the diffusion enhanced scheme [7] as a reference. Finally, we see that our quadrature-based Scharfetter-Gummel schemes are thermodynamically consistent nearly up to machine precision.

The remainder of this paper is organized as follows: In Section 2, we introduce the van Roosbroeck system which we discretize via a finite volume method in Section 3. We put particular emphasis on the approximation of the the charge-carrier fluxes which we focus on in the following. In Section 4, we present the new quadrature-based Scharfetter-Gummel scheme and analyze them in Section 5 before we conclude in the final section.

2 Van Roosbroeck system

In statistical physics electron and hole densities, n and p, can be defined in terms of the electrostatic ψ and electrochemical potentials, φ_n and φ_p, via a general distribution function $\mathcal{F}$ by setting

$$n = N_c \mathcal{F}(\eta_n), \qquad p = N_v \mathcal{F}(\eta_p), \tag{1}$$

where N_c and N_v are the effective densities of state for electrons and holes, respectively. Assuming parabolic energy bands they are given by

$$N_c = \left(\frac{k_B T m_n^*}{2\pi\hbar^2} \right)^{\frac{3}{2}}, \qquad N_v = \left(\frac{k_B T m_p^*}{2\pi\hbar^2} \right)^{\frac{3}{2}}, \tag{2a}$$
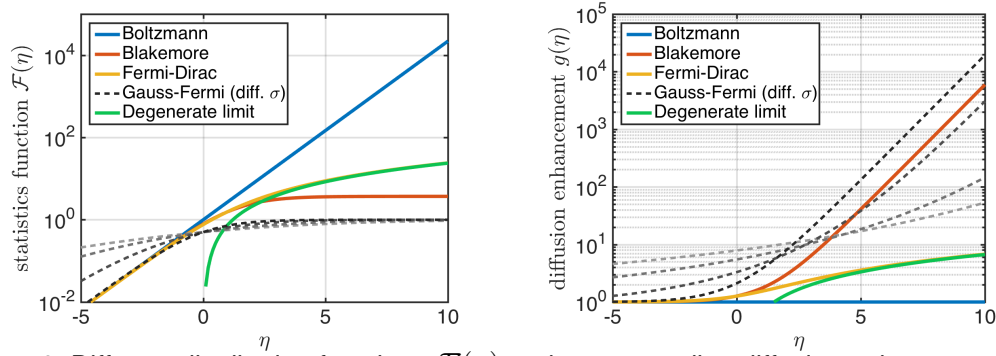
Figure 1: Different distribution functions $\mathcal{F}(\eta)$ and corresponding diffusion enhancements $g(\eta)$.

where $\hbar$ is the Dirac constant, m_n^* and m_p^*, are effective masses of holes and electrons dependent on the band structure of the semiconductor [15].

For both charge carrier densities and the electrostatic potential the stationary van Roosbroeck system is then given by

$$-\nabla \cdot (\varepsilon_0 \varepsilon_r \nabla \psi) = q\,(p - n + C)\,, \tag{3a}$$

$$\nabla \cdot \mathbf{j}_n = qR, \tag{3b}$$

$$\nabla \cdot \mathbf{j}_p = -qR. \tag{3c}$$

The constants q, ε_0 and ε_r describe the elementary charge, the vacuum dielectric permittivity and the relative permittivity of the material, respectively. Furthermore, C is the doping profile and $R = R(n, p)$ denotes the net recombination rate. The current densities can be expressed in terms of the constitutive equations

$$\mathbf{j}_n = -q\mu_n n \nabla \psi + qD_n \nabla n, \qquad \mathbf{j}_p = -q\mu_p p \nabla \psi - qD_p \nabla p, \tag{4}$$

where the electron and hole mobilities, μ_n and μ_p, and the diffusion coefficients, D_n and D_p, are linked via the generalized Einstein relations

$$\frac{D_n}{\mu_n} = \frac{k_B T}{q} g(\eta_n), \qquad \frac{D_p}{\mu_p} = \frac{k_B T}{q} g(\eta_p), \tag{5}$$

which depend on the diffusion enhancement function $g(\eta) = \mathcal{F}(\eta)/\mathcal{F}'(\eta)$. T is the absolute temperature and k_B the Boltzmann constant. For the Boltzmann approximation $\mathcal{F}(\eta) = \exp \eta$ the diffusion enhancement reduces to $g \equiv 1$. The variables, η_n and η_p, are given by

$$\eta_n = \frac{q(\psi - \varphi_n) - E_c}{k_B T}, \qquad \eta_p = \frac{q(\varphi_p - \psi) + E_v}{k_B T}. \tag{6}$$

For the sake of simplicity, in the following we restrict our considerations to electrons, partially omitting the index n.

2.1 Distribution functions

Distribution functions describe the relation between potentials and carriers densities. Considering the parabolic band approximation, for inorganic, 3D bulk semiconductors the link is established via the Fermi-Dirac integral of order 1/2,

$$\mathcal{F}(\eta) = F_{1/2}(\eta) := \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\xi^{1/2}}{\exp(\xi - \eta) + 1} d\xi, \tag{7}$$

which can be approximated in the low density limit by a Blakemore function ($\mathcal{F}(\eta) = (\exp(-\eta) + \gamma)^{-1}$ with $\gamma = 0.27$) that is reduced to the Boltzmann ($\mathcal{F}(\eta) = \exp(\eta)$) distribution for $\gamma = 0$. For large arguments, $F_{1/2}(\eta)$ can be approximated by the degenerate limit $\frac{4}{3\sqrt{\pi}}\eta^{3/2}$. For organic semiconductors the underlying statistical physics is described by the Gauss-Fermi distribution function [16]

$$\mathcal{F}(\eta) = \mathcal{G}(\eta;\sigma) := \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\frac{\exp(-\frac{\xi^2}{2\sigma^2})}{\exp(\xi-\eta)+1}d\xi, \tag{8}$$

The variance σ measures the disorder of the energy levels. The Gauss-Fermi integral reduces to a Blakemore distribution function (with $\gamma = 1$) for vanishing disorder σ, corresponding to a δ-shaped density of states, describing a single transport level. Figure 1 shows several distribution functions and their diffusion enhancement functions.

Next we present the implementational details of the half-order Fermi-Dirac and Gauss-Fermi distribution functions.

### 2.1.1 Numerical calculation of the Fermi-Dirac integral

For the Fermi-Dirac integral of order $1/2$ we use the implementation [17] by Natarajan and Mohanku-mar, who derived a quadrature scheme to estimate the Fermi-Dirac integral. A comprehensive description is presented in [18]. The algorithm evaluates the integral efficiently and reliably via a modified trapezoidal rule in which a correction for the poles of the integrand is applied to the trapezoidal integration sum after a change of variables in order to improve the convergence and accuracy.

2.1.2 Numerical calculation of the Gauss-Fermi integral

Even though some implementations of the Fermi-Dirac distribution $F_{1/2}$ are readily available, this does not hold true for the Gauss-Fermi distribution function. For this reason, we approximate the Gauss-Fermi integral (8) and replace the infinite integration domain with the finite domain $[-B, B]$, yielding the approximation $\widetilde{\mathcal{G}}(\eta;\sigma)$. A straightforward estimate gives

$$\begin{aligned}
\mathcal{G}(\eta;\sigma) - \widetilde{\mathcal{G}}(\eta;\sigma) &:= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\frac{\exp(-\frac{\xi^2}{2\sigma^2})}{\exp(\xi-\eta)+1}d\xi - \frac{1}{\sqrt{2\pi}\sigma}\int_{-B}^{B}\frac{\exp(-\frac{\xi^2}{2\sigma^2})}{\exp(\xi-\eta)+1}d\xi \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{B}^{\infty}\frac{\exp(-\frac{\xi^2}{2\sigma^2})}{\exp(\xi-\eta)+1}d\xi + \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{-B}\frac{\exp(-\frac{\xi^2}{2\sigma^2})}{\exp(\xi-\eta)+1}d\xi \\
&\leq \frac{1}{\sqrt{2\pi}\sigma}\int_{B}^{\infty}\exp\left(-\frac{\xi^2}{2\sigma^2}\right)d\xi + \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{-B}\exp\left(-\frac{\xi^2}{2\sigma^2}\right)d\xi \\
&= \frac{2}{\sqrt{\pi}}\int_{B/(\sqrt{2}\sigma)}^{\infty}\exp(-\xi^2)d\xi \\
&= 1 - \mathrm{erf}\left(\frac{B}{\sqrt{2}\sigma}\right)
\end{aligned}$$

Demanding that the final difference satisfies some tolerance (say ε_ξ), we would need to choose $B = \sqrt{2}\sigma\,\mathrm{erf}^{-1}(1-\varepsilon_\xi)$. Having replaced the infinite domain with a suitable finite domain, we would then proceed to approximate the integral $\widetilde{\mathcal{G}}(\eta;\sigma)$ via quadrature.
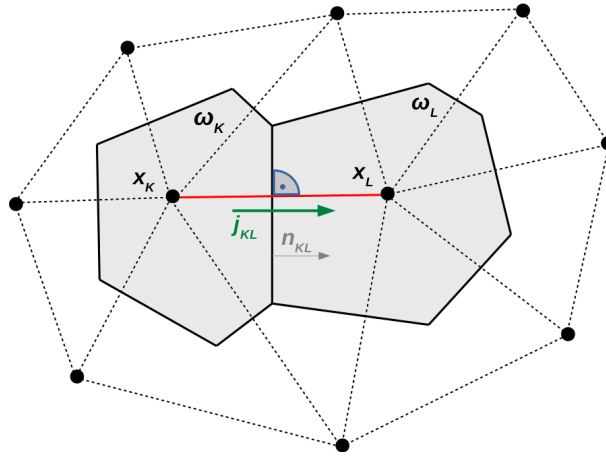
Figure 2: Definition of adjacent control volumes (Voronoï polygons) ω_K and ω_L with corresponding primary Delaunay grid.

3 Finite volume discretization

The finite volume method is a useful method for solving partial differential equations numerically as many physical properties which hold true at the continuous level can be incorporated in to it. We discretize the computational domain using Voronoï grids, for a detailed introduction of its use in the field of numerical device simulation we refer to [10].

We divide the domain Ω into control volumes ω_K and associate a node $\mathbf{x}_K \in \omega_K$ with each control volume such that $\Omega = \bigcup_{K=1}^{N} \omega_K$. Integrating over each control volume and applying Gauss's theorem to the integral of the flux divergence, we obtain the discretization of the continuity equation (3b), which describes how the carrier density changes within a control volume. Denoting with $\mathcal{N}(\omega_K)$ the set of neighboring control volumes to ω_K, we obtain

$$
\begin{aligned}
0 \quad &= \int_{\partial_K} \mathbf{j}_n \cdot \mathbf{n} ds + \int_K R \, d\mathbf{x} = \sum_{\omega_L \epsilon \mathcal{N}(\omega_K)} \int_{\partial_K \cap \partial_L} \mathbf{j}_n \cdot \mathbf{n} ds + \int_K R \, d\mathbf{x} \\
&= \sum_{\omega_L \epsilon \mathcal{N}(\omega_K)} \frac{|\partial_K \cap \partial_L|}{|\mathbf{x}_K - \mathbf{x}_L|} j_{n,KL} + |K| R(n_k, p_K)
\end{aligned}
\tag{9}
$$

where $\mathbf{n}$ is the outward normal at the interface of two neighboring cells and $j_{n,KL}$ denotes the approximated numerical flux between two control volumes. Similarly, the discretization for the hole continuity equation and for the Poisson equation can be deduced. Apart from choosing finer discretization meshes, accurate numerical fluxes significantly contribute to the accuracy and physical correctness of the overall discretization scheme. Therefore, the remainder of this paper focuses on these numerical fluxes.

3.1 Thermodynamic consistency

A scheme is referred to as thermodynamically consistent if

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

implies that $\delta\psi_{KL} = \delta\eta_{KL}$ where

$$
\frac{\psi_L - \psi_K}{U_T} =: \delta\psi_{KL}, \qquad \delta\eta_{KL} := \eta_L - \eta_K.
\tag{10}
$$

The constant $U_T = k_B T / q$ denotes the thermal voltage. Thermodynamic consistency is important when coupling the van Roosbroeck system to more complex models. In particular, it guarantees that for appropriate equilibrium boundary conditions, the solution of the discretized van Roosbroeck system is identical to the solution of the discretized nonlinear Poisson equation describing the thermodynamic equilibrium, see e.g. [10] for a more elaborate discussion.

3.2 Generalized Scharfetter-Gummel scheme

In this section, we discuss concrete choices for $j_{n,KL}$. If one assumes that the (unknown) flux j between two cells is constant and the electrostatic potential varies linearly, in the most general case it fulfills the integral equation [11, 12],

$$\int_{\eta_K}^{\eta_L} \left(\frac{j_n / j_0}{\mathcal{F}(\eta)} + \frac{\psi_L - \psi_K}{U_T} \right)^{-1} d\eta = 1, \quad j_0 = q \mu_n N_c \frac{U_T}{h_{KL}}, \tag{11}$$

where the integration limits are given by $\eta_K = \eta_n(\psi_K, \varphi_K)$ and $\eta_L = \eta_n(\psi_L, \varphi_L)$. In [13] it was shown that for strictly monotonously increasing $\mathcal{F}(\eta)$ this equation has always a unique solution. We will refer to it as the *generalized* Scharfetter-Gummel flux.

For the Boltzmann approximation we recover from (11) the classical Scharfetter-Gummel scheme [2],

$$j_{\text{sg}} = B\left(\delta \psi_{KL} \right) e^{\eta_L} - B\left(-\delta \psi_{KL} \right) e^{\eta_K}, \tag{12}$$

with the non-dimensionalized edge current $j_{\text{sg}} = j_n / j_0$ and the Bernoulli function $B(x) := x / (e^x - 1)$. In [12] it was shown that for the Blakemore approximation $\mathcal{F}(\eta) = \frac{1}{e^{-\eta} + \gamma}$ the integral equation (11) yields a fixed point equation

$$j_{\text{b}} = B\left(\gamma j_{\text{b}} + \delta \psi_{KL} \right) e^{\eta_L} - B\left(-[\gamma j_{\text{b}} + \delta \psi_{KL}] \right) e^{\eta_K}, \tag{13}$$

for the non-dimensionalized edge current $j_{\text{b}} = j_n / j_0$. The right-hand side is a Scharfetter-Gummel expression where the argument of the Bernoulli function is shifted by γj_{b}. Hence, for $\gamma = 0$ the generalized flux j_{b} reduces to the classical Scharfetter-Gummel scheme (12) since the Blakemore function reduces to the Boltzmann function.

3.3 Modified Scharfetter-Gummel schemes

For general distribution functions like (7) and (8), we cannot find simple fixed point equations for the unknown current similar to (11). For this reason one usually employs physically motivated approximate flux solutions. So called modified Scharfetter-Gummel schemes, namely the schemes based on averaging inverse activity coefficients [19] or the diffusion enhanced scheme [7], preserve the Scharfetter-Gummel structure and remain thermodynamically consistent. Especially the latter scheme we use as a reference to see how our new schemes perform. It is given by

$$j_d = -g_{KL} \left(\mathcal{F}(\eta_K) B\left(-\frac{\psi_L - \psi_K}{U_T g_{KL}} \right) - \mathcal{F}(\eta_L) B\left(\frac{\psi_L - \psi_K}{U_T g_{KL}} \right) \right) \tag{14}$$

where g_{KL} is a logarithmic average of the diffusion enhancement g as introduced in [8]:

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

Its usefulness was analyzed recently in [9] where it is shown that it results into a more accurate flux than the schemes based on averaging the inverse activity coefficients [19].

4 A family of new quadrature-based Scharfetter-Gummel schemes

In [14] it was shown that it is possible to numerically approximate the integral (11) via quadrature. This is particularly useful if highly accurate flux approximations are required. We introduce now a whole family of such approximation schemes by choosing different quadrature rules and explain in detail the algorithms used to obtain them. Solving (11) numerically yields effectively new modified schemes, which become practically exact schemes when the number of quadrature points is large enough. We refer to them as *quadrature-based Scharfetter-Gummel schemes*. The implementation is challenging due to two reasons: first one needs to approximate the integral accurately and then solve a nonlinear equation. We will discuss both steps separately.

4.1 Discretization of the integral equation

Denoting the integrand in (11) with $G(\eta; \delta\psi_{KL}, j_{gsg})$ for $j_{gsg} = j_n/j_0$, we can approximate (11) by

$$H(j_{gsg}) := \sum_{i=1}^{N} w_i G(\eta_i; \delta\psi_{KL}, j_{gsg}) - 1 = 0, \tag{16}$$

where w_i are some integration weights, η_i the quadrature nodes and N the number of quadrature nodes. We make explicit choices for the numerical integration in Section 4.2.

When implementing this method one needs to treat two limiting cases separately. For pure diffusive currents, i.e. $\delta\psi_{KL} = 0$ and arbitrary $\delta\eta_{KL}$, the integral equation (11) implies that one obtains the current simply by integrating $\mathcal{F}(\eta)$ from η_K to η_L. On the other hand, for small $\delta\eta_{KL}$ and arbitrary $\delta\psi_{KL}$, the mean value theorem for definite integrals yields

$$1 = \int_{\eta_K}^{\eta_L} \frac{\mathcal{F}(\eta)}{j_{gsg} + \delta\psi \mathcal{F}(\eta)} d\eta = \frac{\delta\eta_{KL}\mathcal{F}\left(\bar{\eta} + \frac{1}{2}\delta\eta_{KL}\xi\right)}{j_{gsg} + \delta\psi_{KL}\mathcal{F}\left(\bar{\eta} + \frac{1}{2}\delta\eta_{KL}\xi\right)}, \quad \bar{\eta}_{KL} := \frac{\eta_L + \eta_K}{2},$$

for some $\xi \in (-1, 1)$. Hence, in the pure drift limit when $\delta\eta_{KL} = 0$, we obtain $j_{gsg} = -\mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL}$ for any $\delta\psi_{KL} \in \mathbb{R}$. We point out that this case is incorporated in the low-order series expansion of (11) for small $\delta\eta_{KL}$ and $\delta\psi_{KL}$, derived in [10], namely

$$\begin{aligned} j_{gsg} = &- \mathcal{F}(\bar{\eta}_{KL})\delta\psi_{KL} + \mathcal{F}(\bar{\eta}_{KL})\delta\eta_{KL} + \frac{1}{12}\frac{\mathcal{F}'(\bar{\eta}_{KL})^2}{\mathcal{F}(\bar{\eta}_{KL})}\delta\psi_{KL}^2\delta\eta_{KL} \\ &- \left(\frac{\mathcal{F}'(\bar{\eta}_{KL})^2}{12\mathcal{F}(\bar{\eta}_{KL})} + \frac{\mathcal{F}''(\bar{\eta}_{KL})}{24}\right)\delta\psi_{KL}\delta\eta_{KL}^2 + \frac{1}{24}\mathcal{F}''(\bar{\eta}_{KL})\delta\eta_{KL}^3. \end{aligned} \tag{17}$$

Due to these considerations, we propose the following implementation for some suitable tolerances ε_η and ε_ψ:

Algorithm 1: Numerical solution of integral equation (11).

---
1 **if** $|\delta\eta_{KL}| < \varepsilon_\eta$ **then**
2   use series expansion (17);
3 **else if** $|\delta\eta_{KL}| \geq \varepsilon_\eta$ *and* $|\delta\psi_{KL}| < \varepsilon_\psi$ **then**
4 $j_{gsg} = \int_{\eta_K}^{\eta_L} \mathcal{F}(\eta)d\eta$;
5 **else if** $|\delta\eta_{KL}| \geq \varepsilon_\eta$ *and* $|\delta\psi_{KL}| \geq \varepsilon_\psi$ **then**
6   use Newton's method for (16) with the diffusion-enhanced scheme (14) as starting guess;
7 **end**

4.2 Numerical integration

Numerical integration, also known as quadrature, approximates definite integrals of a given function f. Without loss of generality, we confine the following considerations to the interval $[-1, 1]$. A general approximation by introducing a positive weight function is given by

$$\int_{-1}^{1} f(x)\omega(x)dx \approx \sum_{i=1}^{N} w_i f(x_i), \tag{18}$$

where x_i are the quadrature nodes, w_i the quadrature weights and N is the number of quadrature points.

A huge variety of quadrature rules are based on interpolation polynomials which are simple to integrate. To evaluate the integral (11) we use four different quadrature rules based on 'optimal' points: Clenshaw-Curtis, Gauss-Legendre, Gauss-Lobatto and adaptive Gauss-Kronrod quadrature [20, 21]. The nodes and the weights are chosen to minimize the approximation error. We use the code [22] for Gauss-Legendre quadrature and [23] for the other ones. There are several advantages to the quadrature rules studied here: The quadrature weights for these quadrature rules are positive, minimizing numerical (cancellation) errors. Moreover, these quadrature rules converge exponentially for sufficiently smooth integrands and there exist well-tested and well-documented implementations [24, 21].

4.2.1 Gauss-Legendre quadrature

Gauss-Legendre quadrature exactly integrates polynomials of degree $2N - 1$ by choosing suitable nodes x_i and weights w_i for $i = 1, \ldots, N$, where N is the number of quadrature points. The nodes x_i are precisely the ith roots of Legendre polynomials $P_N(x)$ on $[-1, 1]$ and the weights w_i are given by

$$w_i = \frac{2}{(1 - x_i^2)[P_N'(x_i)]^2}, \tag{19}$$

see [25] for details. The zeros of the N Legendre polynomial are computed by means of the Newton method. Gauss-Legendre quadrature is highly efficient for smooth integrands and also excludes the integration limits.

In [13] it is shown that no pole can appear within the integration limits of (11). However, poles can come close to the integration limits. If this happens, Gauss-Legendre quadrature might prove useful.

4.2.2 Clenshaw-Curtis quadrature

Clenshaw-Curtis quadrature is another numerical integration technique. It expands the integrand f by means of Chebyshev polynomials. Using the change of variable $x = \cos(t)$ and the discrete cosine transform (DCT) we can approximate our integrand with

$$f(t) = \frac{a_1}{2} + \frac{a_N}{2}T_N + \sum_{i=2}^{N-1} a_i T_i(t), \tag{20}$$

where a_i and $T_i(t)$ denote the Chebyshev coefficients and polynomials, respectively [26]. The jth node corresponds to a maximum or minimum of the Chebyshev polynomial $T_i(t)$ in the range $[-1, 1]$ given by

$$t_j = \cos\left(\frac{\pi j}{N}\right). \tag{21}$$

The Clenshaw-Curtis quadrature also evaluates the integrand at N points but integrates exactly only polynomials up to degree N rather than to degree $2N - 1$ like Gaussian quadrature. However, the Clenshaw-Curtis rule appears to be more than just half as effective as Gaussian quadrature. In [27] the authors show that the method is fast and as accurate as Gaussian quadrature.

4.2.3 Gauss-Lobatto quadrature

Gauss-Lobatto quadrature is another numerical integration technique which exactly integrates polynomials up to order $2N - 1$. In the early seventies, Golub [28] proposed a method to compute the zeros of orthogonal polynomials by using the eigenvalues of a tridiagonal Jacobi matrix. The idea is to generate the nodes and weights needed for the quadrature rule by computing eigenvalues and the first component of the respective eigenvectors of a modified tridiagonal matrix. The use of the Gauss-Lobatto rule is efficient for extrapolating the nodes but the calculation of the weights is quite complicated. A more simple and elegant approach is reported in [29], where the authors derive the nodes via a special Jacobi matrix based on the derivatives of Jacobi polynomial rather and also present an explicit formula to obtain the weights.

4.2.4 Adaptive Gauss-Kronrod quadrature

An adaptive quadrature rule is a technique where the integration interval is recursively split into two halves and the quadrature is applied to each subinterval if the integral is not evaluated with the desired accuracy. The Gauss-Kronrod quadrature, proposed by Alexander Kronrod in the 1960s, belongs to the set of adaptive quadrature rules and was formulated as an extension of the Gaussian quadrature formula. The integral approximation and resulting estimated error are calculated by adding $N + 1$ roots of Stieltjes polynomials to N nodes obtaining a formula of order $2N + 1$. These extra points allow to calculate a higher-order approximation of the integral and the difference between this value with the lower-order one is used as an estimate for the integration error to be compared with the required tolerance [30].

5 Numerical results

We analyze the accuracy and the computation time of different quadrature rules for the Blakemore, Fermi-Dirac and Gauss-Fermi distribution functions numerically by varying the number of quadrature points N. In the following discussion, we fix $\bar{\eta}_{KL}$ and impose $\delta\psi_{KL} \in [-6, 6]$ with step equal to 0.1, $\delta\eta_{KL} \in [-4, 4]$ with step 0.1 and the energetic disorder $\sigma = 5$ for the Gauss-Fermi distribution function, where $\bar{\eta}_{KL}$ denotes the arithmetic average

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

In particular, for the Blakemore distribution function we use the flux obtained from (13) as exact reference

$$err(\bar{\eta}_{KL}, \delta\eta_{KL}, \delta\psi_{KL}) := \left| j_{gsg}^N - j_b \right|, \tag{23}$$

whereas for the two other ones we estimate a highly accurate solution of the integral equation (11) via a large number of quadrature points ($N = 128$). We use this reference value to calculate the error resulting from solving the integral equation based on different numbers of quadrature points N for each

$\delta\psi_{KL}$, $\bar{\eta}_{KL}$ and $\delta\eta_{KL}$. Thus the error for the Fermi-Dirac and Gauss-Fermi distribution functions is given by

$$err(\bar{\eta}_{KL}, \delta\eta_{KL}, \delta\psi_{KL}) := \left| j_{gsg}^N - j_{gsg}^{128} \right|. \tag{24}$$

It is not possible to use only $N = 2$ quadrature points in the implementation of the Gauss-Kronrod quadrature [23] as the implementation requires at least two initial subintervals. In all figures the errors from the quadrature-based schemes are compared to the corresponding errors between the diffusion enhanced scheme (14). The error is again computed with respect to j_{gsg}^{128} for Fermi-Dirac and Gauss-Fermi distribution functions and j_b for the Blakemore distribution function.

5.1 Convergence studies

In the following, we consider for some error vector $\mathbf{e} = (e_1, e_2, \ldots, e_N)^T \in \mathbb{R}^N$ the discrete ℓ_∞ norm

$$\|\mathbf{e}\|_\infty = \max\{|e_k| \text{ for } 1 \leq k \leq N\}$$

as well as the ℓ_2 norm,

$$\|\mathbf{e}\|_2 = \sqrt{\frac{1}{N} \sum_{k=1}^N |e_k|^2}.$$

In Figure 3 and Figure 4 we study the convergence of the quadrature-based Scharfetter-Gummel fluxes via Algorithm 1 with respect to both of these error norms. Depending on the distribution function, the error is calculated either via (23) or (24). We notice that the convergence is exponential which does not come as surprise as the underlying quadrature rules also converge exponentially. However, this was not a priori clear as we solve an integral equation. Moreover, we see that for small values of the diffusion enhancement we need at least four quadrature points regardless of the quadrature rule to achieve the same accuracy as the diffusion enhanced scheme (14). The accuracy increases when $g(\eta)$ increases. We point out that the diffusion enhancement $g(\bar{\eta}_{KL})$ used in both figures is relatively low compared to physically challenging situations (cryogenic temperatures, high doping concentrations and organic semiconductors). However, we report here that we have tried even higher diffusion enhancement values and the positive trend (even faster convergence) which is already visible in both figures continues for larger $\bar{\eta}_{KL}$. In all figures the quadrature-based Scharfetter-Gummel schemes are compared with the diffusion enhanced flux (14).

We have used the following settings and parameters in Algorithm 1: As starting guess, we used the diffusion-enhanced flux (14). Moreover, the tolerances $\varepsilon_\eta = 10^{-12}$ and $\varepsilon_\psi = 10^{-3}$ as well as a tolerance of 10^{-12} for the absolute Newton update to stop the iterations were employed. On average approximately six Newton steps were needed.

5.2 Computation time

In Figure 5 we study the efficiency in terms of computation time needed to solve the integral equation (11) in order to obtain an approximated current j_{gsg}^N for the four different quadrature rules and different numbers of quadrature points N. Normally, Gauss-Kronrod quadrature takes more time than the other ones because it uses an adaptive distribution of nodes. The schemes based on Gauss-Legendre and Clenshaw-Curtis quadrature tend to be the most efficient.
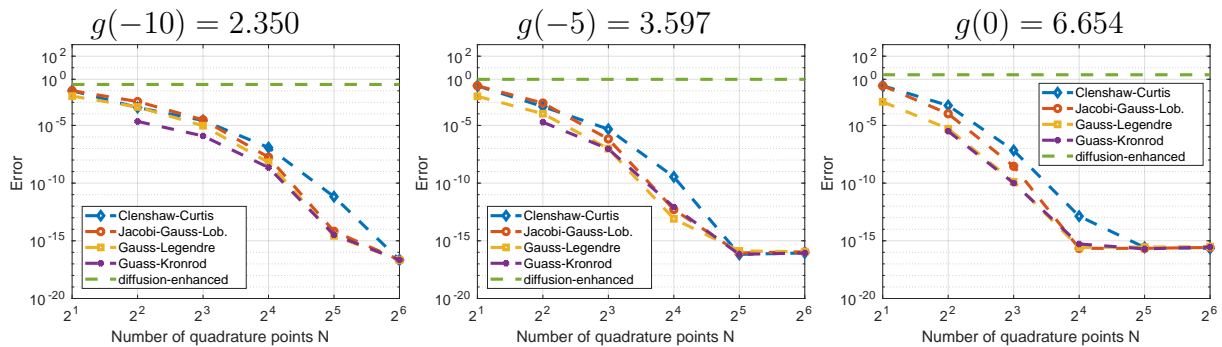
(a) Blakemore approximation for different diffusion enhancement $g(\bar{\eta}_{KL})$:
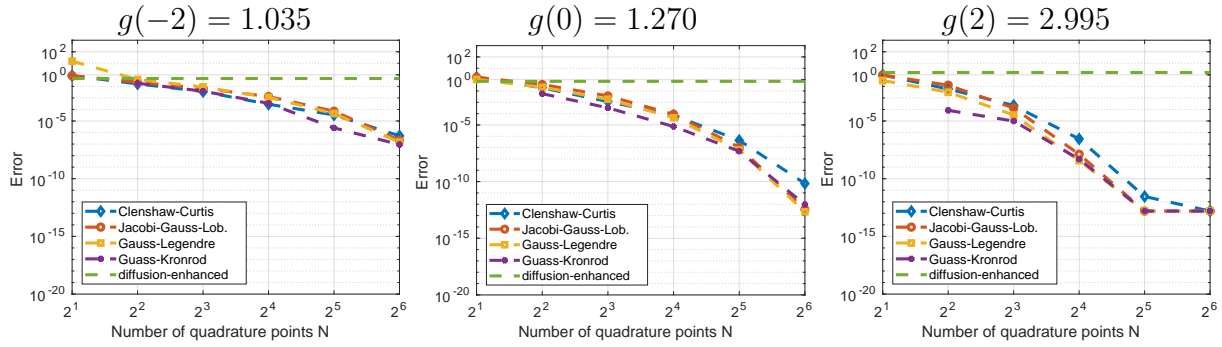
(b) Fermi-Dirac integral $F_{1/2}$ for different diffusion enhancement $g(\bar{\eta}_{KL})$:
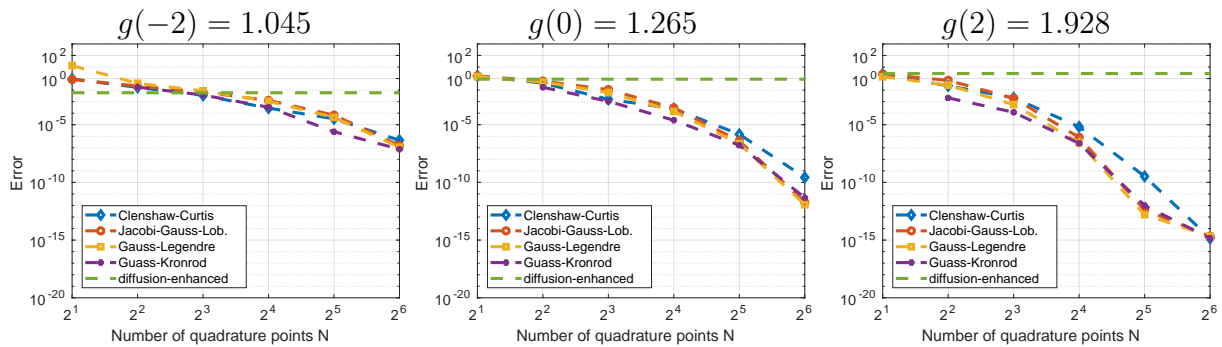
(c) Gauss-Fermi integral $G(\eta; \sigma = 5)$ for different diffusion enhancement $g(\bar{\eta}_{KL})$:
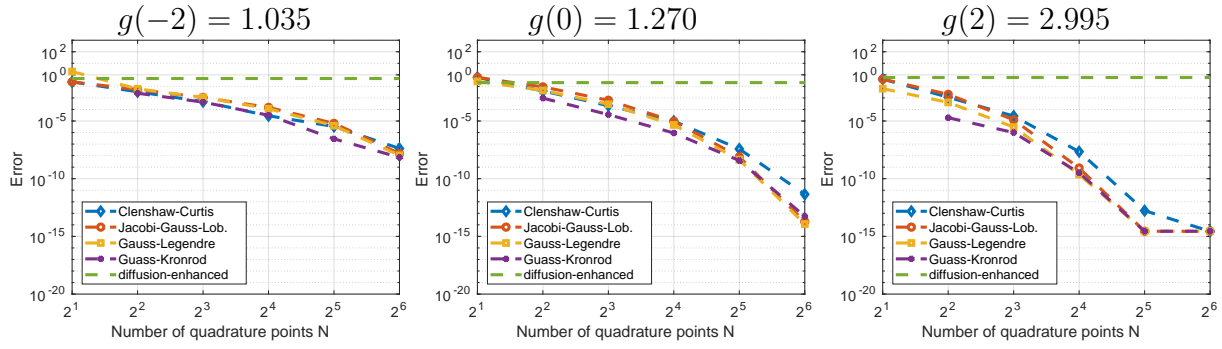

Figure 3: Logarithmic absolute ℓ_∞ errors depending on the number of quadrature points N for different fixed values of $\bar{\eta}_{KL}$ and for several quadrature rules. From left to right, errors for $\bar{\eta}_{KL} = -2, 0, 2$ are shown for Blakemore and Fermi-Dirac functions, rows (a) and (b), respectively. The row (c) shows the errors for the Gauss-Fermi function with $\bar{\eta}_{KL} = -10, -5, 0$. The ranges used to calculate the maximum error are: $\delta\psi \in [-6, 6]$ with step equal to 0.1 and $\delta\eta \in [-4, 4]$ with step 0.1. In all figures the errors from the quadrature-based schemes are compared to the corresponding error between diffusion enhanced scheme (14) and j_{gsg}^{128} (for Fermi-Dirac and Gauss-Fermi distribution functions) and j_b (for the Blakemore distribution function).

5.3   Effect of decreasing the range

In Figure 6 we show a detailed study of the error and the computation times in the reduced range $\delta\eta \in [-0.7, 0.7]$ with step equal to 0.1 and $\delta\psi \in [-1, 1]$ with step 0.1 and for $\bar{\eta}_{KL} = -2, 0, 2$ using the Fermi-Dirac distribution function. For the reduced range just two quadrature points are needed to
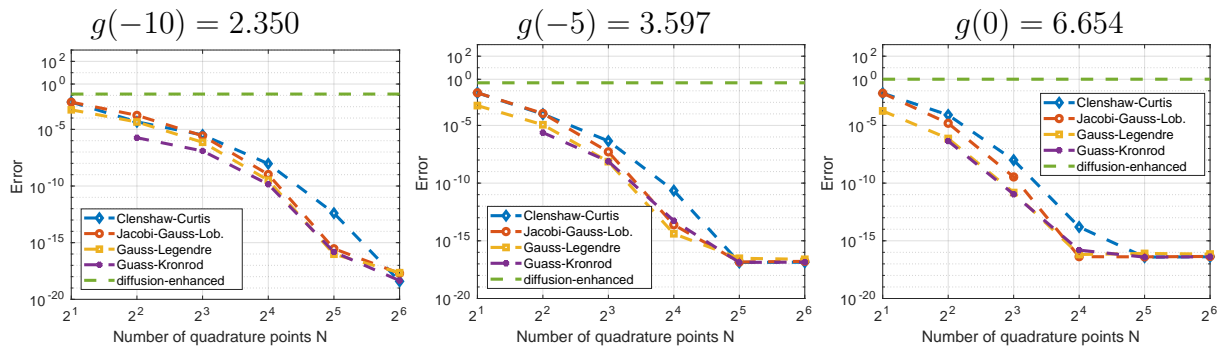
(a) Blakemore approximation for different diffusion enhancement $g(\bar{\eta}_{KL})$:

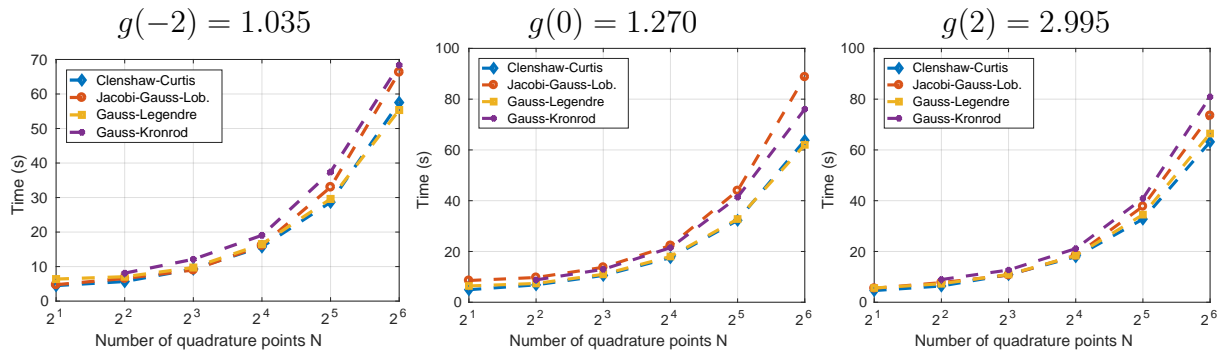(b) Fermi-Dirac integral $F_{1/2}$ for different diffusion enhancement $g(\bar{\eta}_{KL})$:
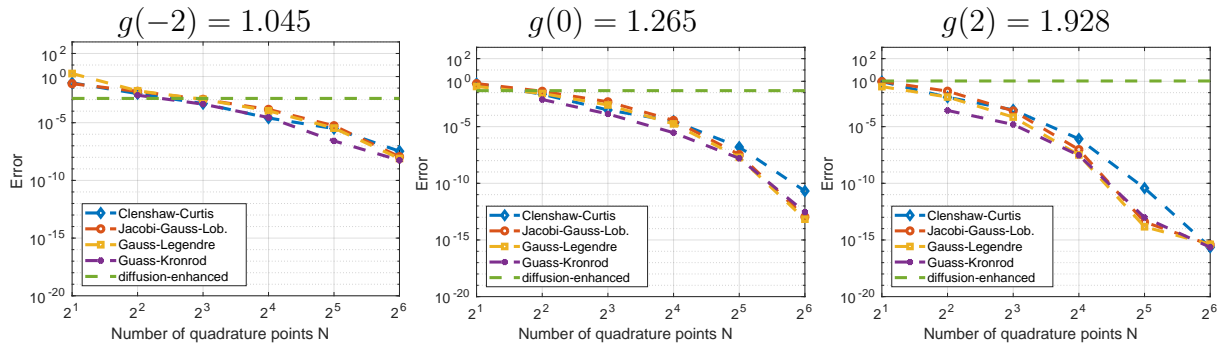
(c) Gauss-Fermi integral $G(\eta; \sigma = 5)$ for different diffusion enhancement $g(\bar{\eta}_{KL})$:
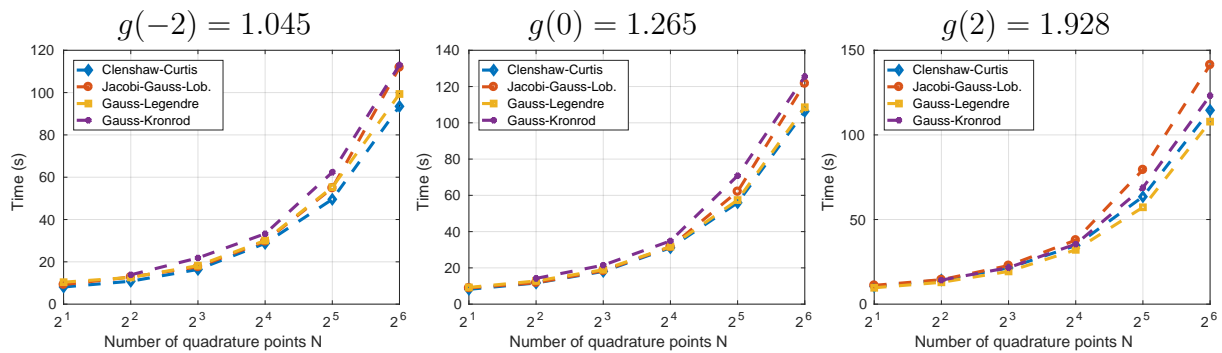

Figure 4: Logarithmic absolute ℓ_2 errors depending on the number of quadrature points N for different fixed values of $\bar{\eta}_{KL}$ and for several quadrature rules. From left to right, errors for $\bar{\eta}_{KL} = -2, 0, 2$ are shown for Blakemore and Fermi-Dirac functions, rows (a) and (b), respectively. The row (c) shows the errors for the Gauss-Fermi function with $\bar{\eta}_{KL} = -10, -5, 0$. The ranges used to calculate the maximum error are: $\delta\psi \in [-6, 6]$ with step equal to 0.1 and $\delta\eta \in [-4, 4]$ with step 0.1. In all figures the errors from the quadrature-based schemes are compared to the corresponding error between diffusion enhanced scheme (14) and j_{gsg}^{128} (for Fermi-Dirac and Gauss-Fermi distribution functions) and j_b (for the Blakemore distribution function).

obtain the same accuracy as the diffusion-enhanced scheme (14). For the larger parameter range, we needed at least $N = 4$ quadrature nodes. In addition, an error not too far from machine precision is obtained with only sixteen quadrature points.
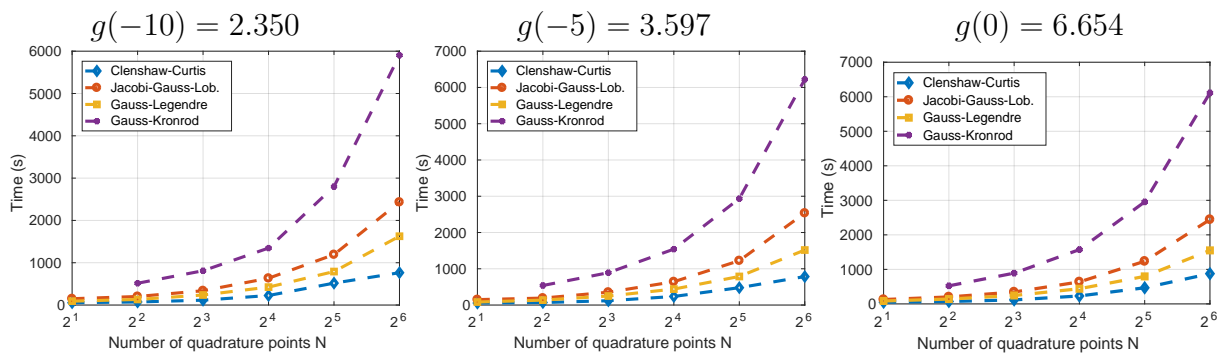
(a) Blakemore approximation for different diffusion enhancement $g(\bar{\eta}_{KL})$:

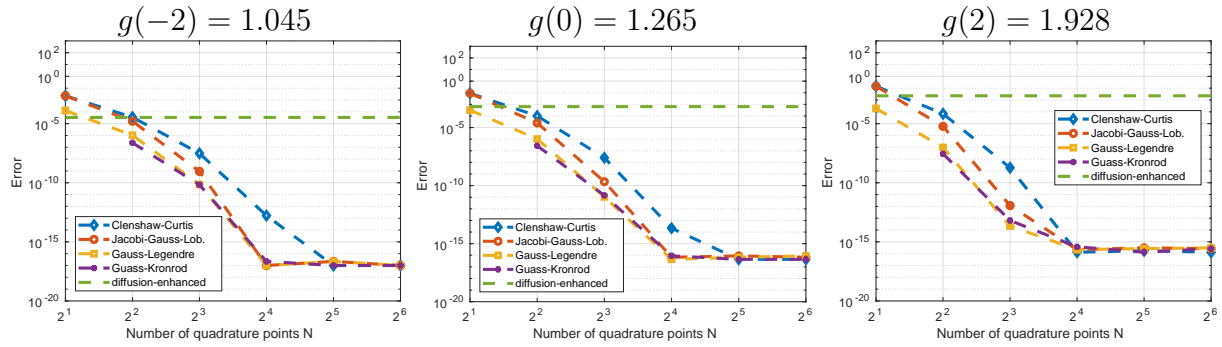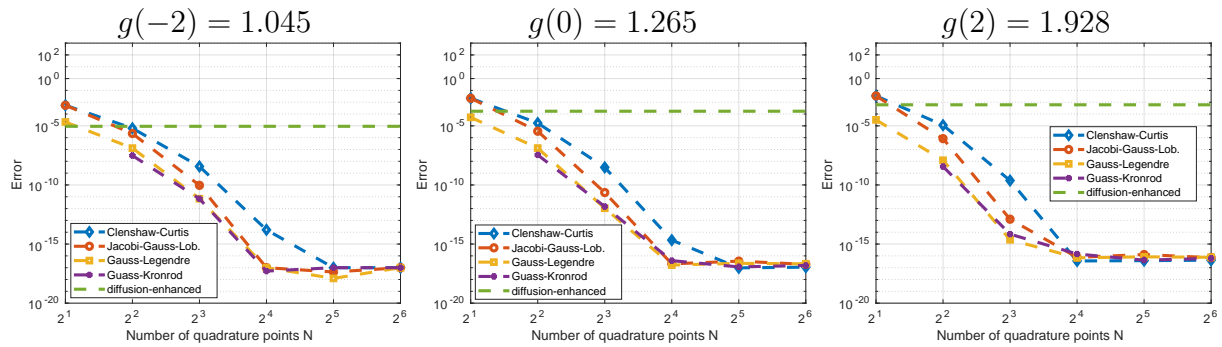(b) Fermi-Dirac integral $F_{1/2}$ for different diffusion enhancement $g(\bar{\eta}_{KL})$:

(c) Gauss-Fermi integral $G(\eta; \sigma = 5)$ for different diffusion enhancement $g(\bar{\eta}_{KL})$:


Figure 5: Computation times by varying the number of quadrature points N for different fixed values of $\bar{\eta}_{KL}$ and for several quadrature rules. From left to right are shown the running times for $\bar{\eta}_{KL} = -2, 0, 2$ for Blakemore and Fermi-Dirac, rows (a) and (b) respectively. The (c) row shows the computation times for Gauss-Fermi with $\bar{\eta}_{KL} = -10, -5, 0$. The ranges used are: $\delta\psi \in [-6, 6]$ with step grid equal to 0.1 and $\delta\eta \in [-4, 4]$ with step 0.1. That is a total of 9801 fluxes were computed.

5.4 Thermodynamic consistency

The generalized Scharfetter-Gummel scheme is consistent with the thermodynamic equilibrium. Considering the Einstein relation (5) and a current equal to zero, we obtain $\nabla\psi = \nabla\eta$ (from assuming constant quasi Fermi potentials), where the electrostatic potential is scaled by U_T. An analogous relation $\delta\eta_{KL} = \delta\psi_{KL}$ should lead to vanishing numerical currents, where $\delta\eta_{KL}$ and $\delta\psi_{KL}$ are defined in (10).

(a) accuracy for ℓ_∞

(b) accuracy for ℓ_2

Figure 6: Reduced ranges: Logarithmic ℓ_∞ (a) and ℓ_2 (b) errors, depending on the number of quadrature points N for different fixed values of $\bar{\eta}_{KL}$ and diffusion enhancement $g(\bar{\eta}_{KL})$. We compare several quadrature rules solving the integral equation (11) for the Fermi-Dirac integral $F_{1/2}$. From left to right are shown error data for $\bar{\eta}_{KL} = -2, 0, 2$. The ranges used to calculate the error and the computation times are: $\delta\psi \in [-1, 1]$ with step equal to 0.1 and $\delta\eta \in [-0.7, 0.7]$ with step 0.1. The error of the diffusion enhanced scheme (14) is shown for comparison.
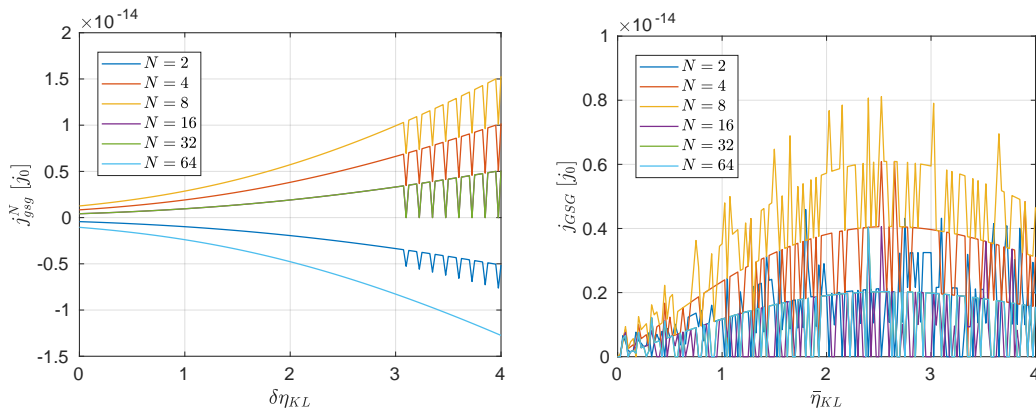
Figure 7: Numerical verification of validity of the thermodynamic consistency for $\eta \in [\eta_K, \eta_L]$ and the Fermi-Dirac distribution function via Gauss-Legendre quadrature, varying the number of quadrature points N. Left: $\delta\eta_{KL} = \psi_{KL} = 4$ and $\bar{\eta}_{KL} \in [0, 4]$ with step equal to 0.025. Right: $\bar{\eta}_{KL} = 2$ and $\delta\eta_{KL} \in [0, 4]$ with step 0.025

Figure 7 reports the degree to which the numerical flux $j_{gsg}^N(\delta\psi_{KL}, \bar{\eta}_{KL}, \delta\eta_{KL})$ violates the thermodynamic consistency when $\delta\eta_{KL} = \delta\psi_{KL}$. Unlike for the diffusion enhanced flux (14), the quadrature-based fluxes are not guaranteed to vanish. In the picture on the left, we fixed $\delta\eta_{KL} = \psi_{KL} = 4$ and

imposed $\bar{\eta}_{KL} \in [0, 4]$ with step equal to 0.025; whereas in the other one, we fixed $\bar{\eta}_{KL} = 2$ and varied $\delta\eta_{KL} \in [0, 4]$ with step 0.025. The integral equation (11) was computed via Gauss-Legendre quadrature for the Fermi-Dirac distribution function. We obtain numerical flux values comparable to machine precision in both pictures when the quasi-Fermi potentials are constant, thus for all practical purposes our schemes are likely to be thermodynamically consistent.

6 Conclusion

In this paper, we introduced a family of quadrature-based Scharfetter-Gummel schemes for non-Boltzmann statistics based on a numerical solution of the integral equation (11) via Newton's method and quadrature rules. In particular, we looked at four different quadrature rules for the evaluation of the integral in (11). This approach automatically lead to four different numerical schemes. Having noted the beneficial behavior for one of these schemes in [14], we investigated the convergence and efficiency more rigorously and compared these schemes to each other. In particular, we looked at ℓ_2 and ℓ_∞ errors with respect to the parameter space of possible arguments (and increasing numbers of quadrature nodes) for different diffusion enhancement values as well as different physically relevant distribution functions (Blakemore, Fermi-Dirac and Gauss-Fermi distribution functions). It turned out that the new quadrature-based Scharfetter-Gummel schemes for relatively few quadrature nodes are already more accurate than a state-of-the-art flux discretization. Furthermore, the convergence in terms of the number of quadrature nodes is exponential, especially in the physically more challenging case of high nonlinear diffusion. Even though the quadrature rules themselves converge exponentially, this was not a priori clear as we have to solve an integral equation. For the examples we studied here, nearly always the four quadrature-based Scharfetter-Gummel schemes have been able achieve a higher accuracy than the reference flux (14). For a smaller range of parameters $\delta\psi$ and $\delta\eta$, we obtained more accurate fluxes than the reference scheme (14) with just two quadrature nodes.

Furthermore, we compared the efficiency of the different quadrature-based Scharfetter-Gummel schemes. The quadrature-based Scharfetter-Gummel schemes based on Gauss-Legendre and Clenshaw-Curtis quadrature tend to be the most efficient. Finally, we verified that the quadrature-based Scharfetter-Gummel schemes (at least numerically) guarantee thermodynamic consistency. A test revealed fluxes close to machine precision when the quasi-Fermi potentials are constant.

In a more general context, our results hint at the possibility to obtain highly accurate numerical fluxes for a rather general class of nonlinear drift-diffusion problems as described in [11]. While the computational complexity in a practical context has yet to be tested, this approach is promising as a method to benchmark potentially computationally more efficient flux approximations like modified Scharfetter-Gummel schemes.

References

[1] W. van Roosbroeck, Theory of the flow of the electrons and holes in germanium and other semiconductor, Bell Syst. Tech. J 29 (4) (1950) 560–567. `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`.

[3] W. Fichtner, D. J. Rose, R. E. Bank, Semiconductor device simulation, SIAM Journal on Scientific and Statistical Computing 4 (3) (1983) 391–415.

[4] M. Kantner, T. Koprucki, Numerical simulation of carrier transport in semiconductor devices at cryogenic temperatures, Opt. Quant. Electron. 48 (12) (2016) 543. `doi:10.1007/s11082-016-0817-2`.
URL `http://dx.doi.org/10.1007/s11082-016-0817-2`

[5] 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, Phys. Rev. B 78 (2008) 085207. `doi:10.1103/PhysRevB.78.085207`.
URL `http://dx.doi.org/10.1103/PhysRevB.78.085207`

[6] E. Knapp, R. Häusermann, H. U. Schwarzenbach, B. Ruhstaller, Numerical simulation of charge transport in disordered organic semiconductor devices, J. Appl. Phys. 108 (5) (2010) 054504. `doi:10.1063/1.3475505`.
URL `http://dx.doi.org/10.1063/1.3475505`

[7] 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`.

[8] T. Koprucki, N. Rotundo, P. Farrell, D. 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`.
URL `http://dx.doi.org/10.1007/s11082-014-0050-9`

[9] P. Farrell, T. Koprucki, J. Fuhrmann, Computational and analytical comparison of flux discretizations for the semiconductor device equations beyond Boltzmann statistics, Journal of Computational Physics 346 (Supplement C) (2017) 497 – 513. `doi:https://doi.org/10.1016/j.jcp.2017.06.023`.
URL `http://www.sciencedirect.com/science/article/pii/S0021999\117304680`

[10] P. Farrell, N. Rotundo, D. Doan, M. Kantner, J. Fuhrmann, T. Koprucki, Mathematical methods: Drift-diffusion models, in: J. Piprek (Ed.), Handbook of Optoelectronic Device Modeling and Simulation, Taylor & Francis, 2017, Ch. 50, pp. 733–772.

[11] 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.

[12] T. Koprucki, K. Gärtner, Discretization scheme for drift-diffusion equations with strong diffusion enhancement, Opt. Quant. Electronics 45 (7) (2013) 791–796. `doi:10.1007/s11082-013-9673-5`.

[13] 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`.
URL `http://dx.doi.org/10.1007/s10825-015-0712-2`

[14] P. Farrell, M. Patriarca, J. Fuhrmann, T. Koprucki, Comparison of thermodynamically consistent charge carrier flux discretizations for Fermi-Dirac and Gauss-Fermi statistics, Optical and Quantum Electronics 50 (101), WIAS preprint 2424. `doi:10.1007/s11082-018-1349-8`.

[15] M. Green, Intrinsic concentration, effective densities of states, and effective mass in silicon, Journal of Applied Physics 67 (6) (1990) 2944–2954. `doi:10.1063/1.345414`.
URL `http://dx.doi.org/10.1063/1.345414`

[16] 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`.
URL `http://dx.doi.org/10.1063/1.3374475`

[17] Natarajan, Mohankumar, Fermi, MathWorks.
URL `https://mathworks.com/matlabcentral/fileexchange/13616-fermi`

[18] N. Mohankumar, A. Natarajan, The accurate numerical evaluation of half-order Fermi–Dirac Integrals, physica status solidi (b) 188 (2) (1995) 635–644. `doi:10.1002/pssb.2221880206`.
URL `http://dx.doi.org/10.1002/pssb.2221880206`

[19] 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`.

[20] L. M. Delves, J. L. Mohamed (Eds.), Computational Methods for Integral Equations, Cambridge University Press, New York, NY, USA, 1986.

[21] G. Dahlquist, Å. Björck, Numerical Methods in Scientific Computing, Volume I, Society for Industrial and Applied Mathematics, 2008. `arXiv:http://epubs.siam.org/doi/pdf/10.1137/1.9780898717785, doi:10.1137/1.9780898717785`.
URL `http://epubs.siam.org/doi/abs/10.1137/1.9780898717785`

[22] G. von Winckel, Legendre–Gauss Quadrature Weights and Nodes, MathWorks.
URL `https://mathworks.com/matlabcentral/fileexchange/4540-legendre-gauss-quadrature-weights-and-nodes`

[23] G. Papazafeiropoulos, Vectorized Numerical Integration Matlab, MathWorks.
URL `https://mathworks.com/matlabcentral/fileexchange/48931-vectorized-numerical-integration-matlab`

[24] P. K. Kythe, M. R. Schäferkotter, Handbook of computational methods for integration, CRC Press, 2004.

[25] M. Abramowitz, I. Stegun, Integration, in: A. M. Series (Ed.), Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, 55, Dover Publications, 1964, Ch. 25.4.

[26] C. Clenshaw, A. Curtis, A method for numerical integration on an automatic computer, Numerische Mathematik 2 (1960) 197–205.

[27] H. O'Hara, F. J. Smith, Error Estimation in the Clenshaw–Curtis Quadrature Formula, The Computer Journal 11 (2) (1968) 213–219. `doi:10.1093/comjnl/11.2.213`.

[28] G. Golub, Some Modified Matrix Eigenvalue Problems, SIAM Review 15 (2) (1973) 318–334.
URL `http://www.jstor.org/stable/2028604`

[29] Z. Zheng, G. Huang, Computation for Jacobi–Gauss–Lobatto quadrature based on derivative relation, Proceedings of Modeling in Geosciences.

[30] A. Kronrod, Nodes and Weights for Quadrature Formulae. Sixteen-place Tables, Nauka, Moscow, 1964 [English Translation: Consultants Bureau, New York, 1965].