Weierstraß-Institut
für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Generalized Scharfetter–Gummel schemes for electro-thermal transport in degenerate semiconductors using the Kelvin formula for the Seebeck coefficient

Markus Kantner

submitted: July 5, 2019

Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: markus.kantner@wias-berlin.de

No. 2605
Berlin 2019

2010 *Mathematics Subject Classification.* 35K57, 35Q79, 65N08, 80A20.

Key words and phrases. Finite volume Scharfetter–Gummel method, semiconductor device simulation, electro-thermal transport, non-isothermal drift-diffusion system, degenerate semiconductors, Fermi–Dirac statistics, Seebeck coefficient.

This work was funded by the German Research Foundation (DFG) under Germany's Excellence Strategy – EXC2046: MATH+ (project AA2-3). The author is grateful to Thomas Koprucki for carefully reading the manuscript and giving valuable comments.

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

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Generalized Scharfetter–Gummel schemes for electro-thermal transport in degenerate semiconductors using the Kelvin formula for the Seebeck coefficient

Markus Kantner

Abstract

Many challenges faced in today's semiconductor devices are related to self-heating phenomena. The optimization of device designs can be assisted by numerical simulations using the non-isothermal drift-diffusion system, where the magnitude of the thermoelectric cross effects is controlled by the Seebeck coefficient. We show that the model equations take a remarkably simple form when assuming the so-called Kelvin formula for the Seebeck coefficient. The corresponding heat generation rate involves exactly the three classically known self-heating effects, namely Joule, recombination and Thomson–Peltier heating, without any further (transient) contributions. Moreover, the thermal driving force in the electrical current density expressions can be entirely absorbed in the (nonlinear) diffusion coefficient via a generalized Einstein relation. The efficient numerical simulation relies on an accurate and robust discretization technique for the fluxes (finite volume Scharfetter–Gummel method), which allows to cope with the typically stiff solutions of the semiconductor device equations. We derive two non-isothermal generalizations of the Scharfetter–Gummel scheme for degenerate semiconductors (Fermi–Dirac statistics) obeying the Kelvin formula. The approaches differ in the treatment of degeneration effects: The first is based on an approximation of the discrete generalized Einstein relation implying a specifically modified thermal voltage, whereas the second scheme follows the conventionally used approach employing a modified electric field. We present a detailed analysis and comparison of both schemes, indicating a superior performance of the modified thermal voltage scheme.

1 Introduction

Thermal effects cause many challenges in a broad variety of semiconductor devices. Thermal instabilities limit the safe-operating area of high power devices and modules in electrical energy technology [1, 2], electrothermal feedback loops lead to catastrophic snapback phenomena in organic light-emitting diodes [3, 4] and self-heating effects decisively limit the achievable output power of semiconductor lasers [5–8]. The numerical simulation of semiconductor devices showing strong self-heating and thermoelectric effects requires a thermodynamically consistent modeling approach, that describes the coupled charge carrier and heat transport processes. In the context of semiconductor device simulation, the non-isothermal drift-diffusion system[9–14] has become the standard model for the self-consistent description of electro-thermal transport phenomena. This is a system of four partial differential equations, which couples the semiconductor device equations [15, 16] to a (lattice) heat flow equation for the temperature distribution in the device. On the step from the *isothermal* to the *non-isothermal* drift-diffusion system, additional thermoelectric transport coefficients must be included in the theory. The magnitude of the thermoelectric cross-effects is governed by the Seebeck coefficient (also *thermopower*), which quantifies the thermoelectric voltage induced by a temperature

DOI 10.20347/WIAS.PREPRINT.2605 Berlin 2019

gradient (*Seebeck effect*) [17, 18]. The reciprocal phenomenon to the Seebeck effect is the *Peltier effect*, which describes the current-induced heating or cooling at material junctions. As a consequence of Onsager's reciprocal relations, the Seebeck and Peltier coefficients are not independent such that only the Seebeck coefficient must be specified [19]. Over the decades, several definitions have been proposed for the Seebeck coefficient [20–23]; recent publications list at least five coexisting different (approximate) formulas [24, 25]. In the context of semiconductor device simulation, the Seebeck coefficients are typically derived from the Boltzmann transport equation in relaxation time approximation [26–28] or defined according to the adage of the Seebeck coefficient being the "(specific) entropy per carrier" [13, 14, 18, 29]. These approaches are often focused on non-degenerate semiconductors, where the carriers follow the classical Maxwell–Boltzmann statistics. This approximation breaks down in heavily doped semiconductors, where the electron-hole plasma becomes degenerate and Fermi–Dirac statistics must be considered to properly take into account the Pauli exclusion principle. Degeneration effects are important in many semiconductor devices such as semiconductor lasers, light emitting diodes or transistors. Moreover, heavily doped semiconductors are considered as "good" thermoelectric materials, i. e., materials with high thermoelectric figure of merit [17, 18], for thermoelectric generators, which can generate electricity from waste heat [30, 31].

In this paper, we will consider an alternative model for the Seebeck coefficient, which is the so-called *Kelvin formula for the thermopower* [32]. The Kelvin formula recently gained interest in theoretical condensed matter physics and has been shown to yield a good approximation of the Seebeck coefficient for many materials (including semiconductors, metals and high temperature superconductors) at reasonably high temperatures [32–42]. The Kelvin formula relates the Seebeck coefficient to the derivative of the entropy density with respect to the carrier density and therefore involves only equilibrium properties of the electron-hole plasma, where degeneration effects are easily included. To our knowledge, the Kelvin formula has not been considered in the context of semiconductor device simulation so far. In Sec. 2, we show that the Kelvin formula yields a remarkably simple form of the non-isothermal drift-diffusion system, which shows two exceptional features:

1 The heat generation rate involves exactly the three classically known self-heating effects (Joule, Thomson–Peltier and recombination heating) without any further (transient) contribution.
2 The thermal driving force in the current density expressions can be entirely absorbed in a (nonlinear) diffusion coefficient via a generalized Einstein relation, i. e., the ∇T term is eliminated in the drift-diffusion form.

The second part of this paper (Sec. 3) deals with the discretization of the electrical current density expressions, which are required in non-isothermal semiconductor device simulation tools. The robust and accurate discretization of the drift-diffusion fluxes in semiconductors with exponentially varying carrier densities is a non-trivial problem, that requires a special purpose discretization technique. The problem has been solved by Scharfetter and Gummel for the case of non-degenerate semiconductors under isothermal conditions [43]. Since then, several adaptations of the method have been developed to account for more general situations (non-isothermal conditions [44–51], degeneration effects [52–57]). The Kelvin formula for the Seebeck coefficients allows for a straightforward generalization of the Scharfetter–Gummel approach to the non-isothermal case. We take up two different approaches to incorporate degeneration effects into the non-isothermal Scharfetter–Gummel formula and give an extensive numerical and analytical comparison of both methods. This includes an investigation of limiting cases and structure preserving properties of the discrete formulas (Sec. 3.3), a comparison with the numerically exact solution of the underlying two-point boundary value problem (Sec. 3.4) and a comparison of analytical error bounds (Sec. 3.5). Finally, in Sec. 3.6, we present a numerical convergence analysis of both schemes based on numerical simulations of a one-dimensional p-n-diode.

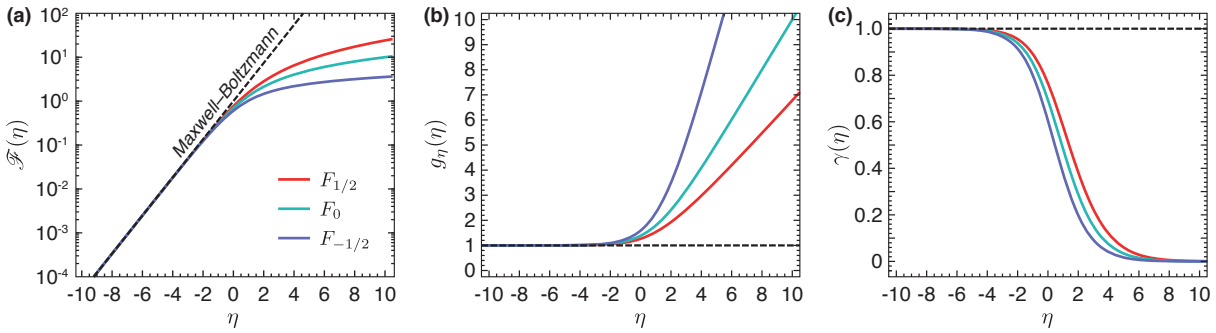

Fig. 1. (a) Fermi–Dirac integrals (7) of order $\nu = 1/2,\, 0$ and $-1/2$ as a function of the reduced Fermi energy η. For $\eta \ll -1$ the Fermi–Dirac integrals approach the Maxwell–Boltzmann distribution ($\mathscr{F}(\eta) = \exp(\eta)$, non-degenerate limit). (b) Plot of the degeneracy factor (16b) (or diffusion enhancement factor) for the Fermi–Dirac integrals in (a). For $\eta \ll -1$ the degeneracy factor approaches 1 (linear diffusion). (c) Correction factor (29) that quantifies the deviation of the Fermi–Dirac integrals from the exponential function. The non-degenerate limit corresponds to $\gamma(\eta) \equiv 1$.

2 The non-isothermal drift-diffusion system using the Kelvin formula for the Seebeck coefficient

In this section we briefly review the non-isothermal drift-diffusion system, which provides a self-consistent description of the coupled electro-thermal transport processes in semiconductor devices. The model has been extensively studied by several authors from the perspective of physical kinetics or phenomenological non-equilibrium thermodynamics [9–14]. The model equations read:

$$-\nabla \cdot \varepsilon \nabla \phi = q\,(C + p - n)\,, \tag{1}$$

$$\partial_t n - \frac{1}{q}\nabla \cdot \mathbf{j}_n = -R, \tag{2}$$

$$\partial_t p + \frac{1}{q}\nabla \cdot \mathbf{j}_p = -R, \tag{3}$$

$$c_V \partial_t T - \nabla \cdot \kappa \nabla T = H. \tag{4}$$

Poisson's equation (1) describes the electrostatic field generated by the electrical charge density $\rho = q\,(C + p - n)$. Here, ϕ is the electrostatic potential, n and p are the densities of electrons and holes, respectively, C is the built-in doping profile, q is the elementary charge and ε is the (absolute) dielectric constant of the material. The transport and recombination dynamics of the electron-hole plasma is modeled by the continuity equations (2)–(3), where $\mathbf{j}_{n/p}$ are the electrical current densities and R is the (net-)recombination rate. The latter includes several radiative and non-radiative recombination processes (Shockley–Read–Hall recombination, Auger recombination, spontaneous emission etc.) [15]. The carrier densities n, p are connected with the electrostatic potential ϕ via the state equations

$$n = N_c(T)\,\mathscr{F}\left(\frac{\mu_c + q\phi - E_c(T)}{k_B T}\right), \qquad p = N_v(T)\,\mathscr{F}\left(\frac{E_v(T) - q\phi - \mu_v}{k_B T}\right), \tag{5}$$

where k_B is Boltzmann's constant, $N_{c/v}$ is the effective density of states and $E_{c/v}$ is the reference energy level (typically the band edge energy) of the conduction band or valence band, respectively. The function $\mathscr{F}$ describes the occupation of the electronic states under *quasi-equilibrium* conditions, which is controlled by the quasi-Fermi energies $\mu_{c/v}$ of the respective bands. The quasi-Fermi energies are connected with the quasi-Fermi potentials $\varphi_{n/p}$ via

$$\mu_c = -q\varphi_n, \qquad \mu_v = -q\varphi_p. \tag{6}$$

In non-degenerate semiconductors (Maxwell–Boltzmann statistics), $\mathscr{F}$ is the exponential function $\mathscr{F}(\eta) = \exp(\eta)$. Taking the degeneration of the electron-hole plasma due to Pauli-blocking into account (Fermi–Dirac statistics), $\mathscr{F}$ is typically given by the Fermi–Dirac integral

$$\mathscr{F}(\eta) = F_\nu(\eta) = \frac{1}{\Gamma(\nu+1)} \int_0^\infty \mathrm{d}\xi \, \frac{\xi^\nu}{\exp(\xi-\eta)+1}, \tag{7}$$

where the index ν depends on the dimensionality of the structure. Isotropic, bulk materials with parabolic energy bands are described by $\nu = 1/2$; for two-dimensional materials (quantum wells) the index $\nu = 0$ applies. See Fig. 1 (a) for a plot of the Fermi–Dirac integrals for different ν as a function of the reduced Fermi energy. In the case of organic semiconductors, $\mathscr{F}$ is often taken as the Gauss–Fermi integral [58, 59] or a hypergeometric function [60, 61].

The heat transport equation (4) describes the spatio-temporal dynamics of the temperature distribution in the device. Here, T is the absolute temperature, c_V is the (volumetric) heat capacity, κ is the thermal conductivity and H is the heat generation rate. The non-isothermal drift-diffusion model assumes a local thermal equilibrium between the lattice and the carriers, i. e., $T = T_L = T_n = T_p$. The system (1)–(4) must be supplemented with boundary conditions for electrical contacts, semiconductor-insulator interfaces, heat sinks etc. We refer to Refs. [15, 62] for details.

The electrical current densities are driven by the gradients of the quasi-Fermi potentials and the temperature

$$\mathbf{j}_n = -\sigma_n(\nabla\varphi_n + P_n\nabla T), \qquad \mathbf{j}_p = -\sigma_p(\nabla\varphi_p + P_p\nabla T), \tag{8}$$

where $\sigma_n = qM_nn$ and $\sigma_p = qM_pp$ are the electrical conductivities (with carrier mobilities $M_{n/p}$) and $P_{n/p}$ denote the Seebeck coefficients.

2.1 Kelvin formula for the Seebeck coefficient

In this paper we will consider the so-called *Kelvin formula* for the Seebeck coefficient [32]

$$P_n = -\frac{1}{q}\frac{\partial s(n,p,T)}{\partial n}, \qquad P_p = +\frac{1}{q}\frac{\partial s(n,p,T)}{\partial p}, \tag{9}$$

which relates the thermoelectric powers to the derivatives of the entropy density $s = s(n,p,T)$ with respect to the carrier densities. The expression for the entropy density is easily derived from the free energy density $f(n,p,T)$ of the system, which is a proper thermodynamic potential if the set of unknowns is chosen as (n,p,T) ("natural variables") . The expressions for the quasi-Fermi energies and the entropy density then follow as

$$\frac{\partial f(n,p,T)}{\partial n} = +\mu_c(n,p,T), \quad \frac{\partial f(n,p,T)}{\partial p} = -\mu_v(n,p,T), \quad \frac{\partial f(n,p,T)}{\partial T} = -s(n,p,T). \tag{10}$$

Taking the second derivatives, this yields the Maxwell relations

$$\frac{\partial \mu_c(n,p,T)}{\partial T} = -\frac{\partial s(n,p,T)}{\partial n}, \qquad \frac{\partial \mu_v(n,p,T)}{\partial T} = +\frac{\partial s(n,p,T)}{\partial p}, \tag{11}$$

which allow for an alternative representation of Eq. (9). The free energy density includes contributions from the quasi-free electron-hole plasma (ideal Fermi gas), the lattice vibrations (ideal Bose gas) and the electrostatic (Coulomb) interaction energy. Throughout this paper, we assume the form [13]

$$f(n,p,T) = f_{\text{e–h}}(n,p,T) + f_L(T) + f_{\text{Coul}}(p-n). \tag{12}$$

The free energy density of the (non-interacting) electron-hole plasma reads [13, 63]

$$f_{\text{e–h}}(n,p,T) = k_BT\mathscr{F}^{-1}\left(\frac{n}{N_c(T)}\right)n - k_BTN_c(T)\,\mathscr{G}\left(\mathscr{F}^{-1}\left(\frac{n}{N_c(T)}\right)\right) + E_c(T)\,n \quad (13)$$
$$+ k_BT\mathscr{F}^{-1}\left(\frac{p}{N_v(T)}\right)p - k_BTN_v(T)\,\mathscr{G}\left(\mathscr{F}^{-1}\left(\frac{p}{N_v(T)}\right)\right) - E_v(T)\,p,$$

where $\mathscr{F}^{-1}$ is the inverse of the function $\mathscr{F}$ in the state equations (5) and $\mathscr{G}$ denotes its antiderivative: $\mathscr{G}'(\eta) = \mathscr{F}(\eta)$. Note that Eq. (13) implies

$$\frac{\partial f_{\text{e–h}}}{\partial n} = k_BT\mathscr{F}^{-1}\left(\frac{n}{N_c(T)}\right) + E_c(T), \quad \frac{\partial f_{\text{e–h}}}{\partial p} = k_BT\mathscr{F}^{-1}\left(\frac{p}{N_v(T)}\right) - E_v(T). \quad (14a)$$

The lattice contribution $f_L(T)$ yields the dominant contribution to the heat capacity c_V. It can be derived from, e. g., the Debye model for the free phonon gas [64]. The Coulomb interaction energy f_{Coul} must be modeled such that the state Eqs. (5) follow consistently from solving the defining relations for the quasi-Fermi energies (10) for the carrier densities. In order to supplement the missing electrostatic contributions in Eq. (14a), we specify the derivatives of f_{Coul} with respect to the carrier densities:

$$\frac{\partial f_{\text{Coul}}}{\partial n} = -q\phi, \qquad \frac{\partial f_{\text{Coul}}}{\partial p} = +q\phi. \quad (14b)$$

We refer to Albinus et al. [13] for a rigorous mathematical treatment of the Coulomb interaction energy.

The Seebeck coefficients (9) are evaluated using Eqs. (10)–(14). Since f_{Coul} is independent of the temperature and f_L does not depend on the carrier densities, the evaluation of Eq. (9) requires only the Maxwell relations (11) and the derivatives of Eqs. (14a) with respect to the temperature. One obtains

$$P_n(n,T) = -\frac{k_B}{q}\left(\frac{TN_c'(T)}{N_c(T)}g\left(\frac{n}{N_c(T)}\right) - \mathscr{F}^{-1}\left(\frac{n}{N_c(T)}\right) - \frac{1}{k_B}E_c'(T)\right), \quad (15a)$$

$$P_p(p,T) = +\frac{k_B}{q}\left(\frac{TN_v'(T)}{N_v(T)}g\left(\frac{p}{N_v(T)}\right) - \mathscr{F}^{-1}\left(\frac{p}{N_v(T)}\right) + \frac{1}{k_B}E_v'(T)\right), \quad (15b)$$

where the prime denotes the derivative, i. e., $N_{c/v}'(T) = \partial_T N_{c/v}(T)$ and $E_{c/v}'(T) = \partial_T E_{c/v}(T)$. For power law type temperature dependency $N_{c/v}(T) \propto T^\theta$ (with, e. g., $\theta = 3/2$), the factor in the first term reduces to a constant $TN_{c/v}'(T)/N_{c/v}(T) = \theta$. For temperature-dependent effective masses, the term is more complicated. The function

$$g(x) = x\,\frac{\mathrm{d}\mathscr{F}^{-1}(x)}{\mathrm{d}x} \quad (16a)$$

quantifies the degeneration of the Fermi gas. For non-degenerate carrier statistics (Maxwell–Boltzmann statistics), Eq. (16a) reduces to exactly $g \equiv 1$. For degenerate carrier statistics one obtains $g > 1$, which implies a nonlinear enhancement of the diffusion current (see Sec. 2.4 below). For later use, we also introduce the function

$$g_\eta(\eta) \equiv g(\mathscr{F}(\eta)) = \frac{\mathscr{F}(\eta)}{\mathscr{F}'(\eta)}, \quad (16b)$$

which is plotted in Fig. 1 (b). The last terms in Eq. (15) describe the contributions of the temperature dependency of the band edge energies to the Seebeck coefficients. The two terms are not independent as they are required to satisfy $E_g'(T) = E_c'(T) - E_v'(T)$, where $E_g(T)$ is the energy band gap. A plot of the Seebeck coefficients (15) as a function of the reduced Fermi energy η is shown in Fig. 2 (a) for $\mathscr{F}(\eta) = F_{1/2}(\eta)$ and $N_{c,v} \propto T^{3/2}$. The plot illustrates schematically the impact of the temperature derivatives of the band edge energies and the role of degeneration effects.

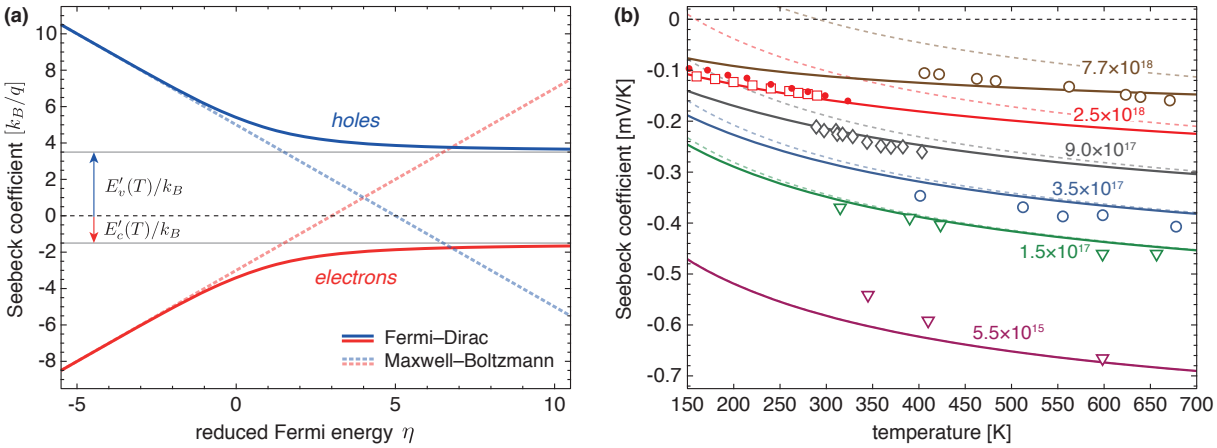

Fig. 2. (a) Seebeck coefficient according to the Kelvin formula (15) as a function of the reduced Fermi energy η for power law type effective density of states $N_{c/v} \propto T^{3/2}$ and $\mathscr{F}(\eta) = F_{1/2}(\eta)$ in units of k_B/q. The formula (15) takes degeneration effects (Fermi–Dirac statistics, solid lines) of the electron-hole plasma into account, which causes a deviation from the non-degenerate result (Maxwell–Boltzmann statistics, dashed lines) at $\eta \gtrsim -1$. The temperature dependency of the band gap energy yields an offset of $E'_c(T) = \left(\chi + \frac{1}{2}\right) E'_g(T)$ for electrons (red lines) and $E'_v = \left(\chi - \frac{1}{2}\right) E'_g(T)$ for holes (blue lines). The plot is for $\chi = -0.2$ and $k_B^{-1} E'_g(T) = -5$. (b) Seebeck coefficient for n-type GaAs. Solid lines are computed according to the Kelvin formula (15a) using Fermi–Dirac statistics, dashed lines indicate the corresponding non-degenerate limit. The respective ionized donor densities $C = N_D^+$ are given in the plot in units of cm^{-3}. The temperature-dependency of the band gap energy $E_g(T)$ is modeled by the Varshni model (17) with data from Ref. [62] and the effective mass is $m_c^*(T) = \left(0.067 - 1.2 \times 10^{-5}\,\mathrm{K}^{-1}\,T\right) m_0$ [62], where m_0 is the free electron mass. The fitting parameter is set to $\chi = -0.2$. Experimental data: ▽ Carlson et al. [65], ○ Amith et al. [66], ◇ Edmond et al. [67] (data from Ref. [68]), □ Homm et al. [69] and ● Emel'yanenko et al. [70] (data from Ref. [68]).

2.2 Comparison with experimental data

Several empirical models for the temperature dependency of the band gap energy have been proposed in the literature [71], including the commonly accepted Varshni model

$$E_g(T) = E_{g,0} - \frac{\alpha T^2}{\beta + T}, \tag{17}$$

where $E_{g,0}$, α and β are material specific constants [72]. In order to specify $E'_{c/v}(T)$ from Eq. (17), we introduce a parameter χ such that $E'_c(T) = \left(\chi + \frac{1}{2}\right) E'_g(T)$ and $E'_v(T) = \left(\chi - \frac{1}{2}\right) E'_g(T)$. In applications, χ can be used as a fitting parameter. It shall be noted that the terms involving $E'_{c/v}(T)$ in Eq. (15) are non-negligible and yield a significant contribution to the Seebeck coefficients at elevated temperatures. Indeed, some room temperature values of $k_B^{-1} E'_g(300\,\mathrm{K})$ for important semiconductors are -2.95 (Si), -4.47 (Ge), -5.32 (GaAs) [62], which are on the same order of magnitude as the first term $T N'_{c/v}(T) / N_{c/v}(T) \approx 1.5$ in Eq. (15).

In Fig. 2 (b), the Kelvin formula is plotted along with experimental data for n-GaAs. We observe a good quantitative agreement of the formula (15a) with the experimental data in both the weak and the heavy doping regime for temperatures above $150\,\mathrm{K}$. At high carrier densities ($N_D^+ \geq 9 \times 10^{17}\,\mathrm{cm}^{-3}$) the conduction band electrons become degenerate (see the deviation of the solid and dashed lines), where the experimental values nicely follow the degenerate formula (15a). See the caption for details. At low temperatures ($T < 150\,\mathrm{K}$, not shown), the Seebeck coefficient is increasingly dominated by the phonon drag effect [69], which is not considered in the present model.

2.3 Heat generation rate

A commonly accepted form of the heat generation rate H was derived by Wachutka [9]:

$$\begin{aligned} H = &\frac{1}{\sigma_n}\left\|\mathbf{j}_n\right\|^2 + \frac{1}{\sigma_p}\left\|\mathbf{j}_p\right\|^2 - T\,\mathbf{j}_n\cdot\nabla P_n - T\,\mathbf{j}_p\cdot\nabla P_p \\ &+ q\left(T\frac{\partial\varphi_n\left(n,p,T\right)}{\partial T} - \varphi_n - T\frac{\partial\varphi_p\left(n,p,T\right)}{\partial T} + \varphi_p\right)R \\ &- T\left(\frac{\partial\varphi_p\left(n,p,T\right)}{\partial T} + P_p\right)\nabla\cdot\mathbf{j}_p - T\left(\frac{\partial\varphi_n\left(n,p,T\right)}{\partial T} + P_n\right)\nabla\cdot\mathbf{j}_n. \end{aligned} \tag{18}$$

Here we omit the radiation power density contribution from the original work. The notation $\|\mathbf{x}\| = \left(\mathbf{x}\cdot\mathbf{x}\right)^{1/2}$ is the standard vector norm. It is important to note that the derivation of Eq. (18) does not involve any explicit assumptions on the Seebeck coefficient. Using the Maxwell relations (11) and the transport Eqs. (2)–(3), we rewrite Eq. (18) as

$$\begin{aligned} H = &\frac{1}{\sigma_n}\left\|\mathbf{j}_n\right\|^2 + \frac{1}{\sigma_p}\left\|\mathbf{j}_p\right\|^2 - T\,\mathbf{j}_n\cdot\nabla P_n - T\,\mathbf{j}_p\cdot\nabla P_p + q\left(\varphi_p - \varphi_n + \Pi_p - \Pi_n\right)R \\ &+ qT\left(P_p - \frac{1}{q}\frac{\partial s\left(n,p,T\right)}{\partial p}\right)\partial_t p - qT\left(P_n + \frac{1}{q}\frac{\partial s\left(n,p,T\right)}{\partial n}\right)\partial_t n, \end{aligned} \tag{19}$$

where we introduced the Peltier coefficients $\Pi_n = TP_n$ and $\Pi_p = TP_p$. Before we highlight the consequences of the Kelvin formula for the Seebeck coefficients on H, we give a brief interpretation of the individual terms in Eq. (19).

The first two terms $H_{J,\lambda} = \sigma_\lambda^{-1}\left\|\mathbf{j}_\lambda\right\|^2$ (for $\lambda \in \{n,p\}$) describe Joule heating, which is always non-negative and therefore never leads to cooling of the device. The terms $H_{\text{T–P},\lambda} = -T\,\mathbf{j}_\lambda\cdot\nabla P_\lambda$ (for $\lambda \in \{n,p\}$) describe the Thomson–Peltier effect, which can either heat or cool the device depending on the direction of the current flow. At constant temperature, this reduces to the Peltier effect $H_{\text{T–P},\lambda}|_{T=\text{const.}} = -\mathbf{j}_\lambda\cdot\nabla\Pi_\lambda$, which is important at heterointerfaces and p-n junctions. At constant carrier densities, one obtains the Thomson heat term $H_{\text{T}--\text{P},\lambda}|_{n,p=\text{const.}} = -\mathcal{K}_\lambda\,\mathbf{j}_\lambda\cdot\nabla T$ with the Thomson coefficient $\mathcal{K}_\lambda = T\frac{\partial P_\lambda}{\partial T} = \frac{\partial\Pi_\lambda}{\partial T} - P_\lambda$ (for $\lambda \in \{n,p\}$). The Thomson–Peltier effect combines both contributions. The recombination heat term $H_R = q(\varphi_p - \varphi_n + \Pi_p - \Pi_n)R$ models the self-heating of the device due to recombination of electron-hole pairs. The difference of the Peltier coefficients describes are the average excess energy of the carriers above the Fermi voltage. The last line in Eq. (19) is a purely transient contribution, that has been discussed by several authors [9–12, 25]. In simulation practice, this term is often neglected, since estimations show that it is negligible in comparison with the other self-heating sources, see Refs. [73, 74].

We observe that the transient term vanishes exactly if we choose the Kelvin formula (9) for the Seebeck coefficients. As a result, solely the classically known self-heating terms are contained in the model and all additional, transient heating mechanisms are excluded:

$$H = \frac{1}{\sigma_n}\left\|\mathbf{j}_n\right\|^2 + \frac{1}{\sigma_p}\left\|\mathbf{j}_p\right\|^2 - T\,\mathbf{j}_n\cdot\nabla P_n - T\,\mathbf{j}_p\cdot\nabla P_p + q\left(\varphi_p - \varphi_n + \Pi_p - \Pi_n\right)R. \tag{20}$$

Finally, we rewrite the recombination heating term using the Seebeck coefficients (15) and Eq. (5). One obtains

$$H_R = \left(E_g\left(T\right) - TE_g'\left(T\right) + \left[\frac{TN_v'\left(T\right)}{N_v\left(T\right)}g\left(\frac{p}{N_v\left(T\right)}\right) + \frac{TN_c'\left(T\right)}{N_c\left(T\right)}g\left(\frac{n}{N_c\left(T\right)}\right)\right]k_BT\right)R.$$

The term in the brackets describes the (differential) average thermal energy per recombining electron-hole pair. For an effective density of states function $N_{c/v} \propto T^{3/2}$ and non-degenerate carrier statistics, we recover the classical result $H_R \approx \left(E_g\left(T\right) - TE_g'\left(T\right) + 3k_BT\right) R$. This yields a clear interpretation of the degeneracy factor g (see Eq. (16)): It describes the increased average thermal energy of the Fermi gas due to Pauli blocking in comparison to the non-degenerate case at the same carrier density. We emphasize that the Kelvin formula immediately yields the correct average kinetic energy $2 \times \frac{3}{2} k_B T$ of the three-dimensional electron-hole plasma just from the temperature dependency of the effective density of states function $N_{c/v} \propto T^{3/2}$. This is in general not true for Seebeck coefficients derived from the Boltzmann transport equation in relaxation time approximation, where the average thermal energy of the electron-hole plasma in the recombination heat term depends on a scattering parameter, see e. g. Ref. [29].

2.4 Electrical current densities in drift-diffusion form

In this section we recast the electrical current density expressions from the thermodynamic form (8) to the drift-diffusion form. As we will see below, the Kelvin formula for the Seebeck coefficient allows to entirely absorb the thermally driven part of the electrical current density in the diffusion coefficient via a generalized Einstein relation. Thus, the ∇T term can be eliminated in the drift-diffusion form, which significantly simplifies the current density expression. Our derivation is based on rewriting the gradient of the quasi-Fermi potential using the free energy density (12) and further thermodynamic relations stated above. In the following, we sketch the essential steps for the electron current density, the corresponding expression for the holes follow analogously. We obtain

$$\begin{aligned} -q\nabla\varphi_n &\overset{\text{Eq. (10)}}{=} \nabla\frac{\partial f}{\partial n} \overset{\text{Eq. (12)}}{=} \nabla\left(\frac{\partial f_{\text{Coul}}}{\partial n} + \frac{\partial f_{\text{e–h}}}{\partial n}\right) \\ &\overset{\text{Eq. (14b)}}{=} -q\nabla\phi + \frac{\partial^2 f_{\text{e–h}}}{\partial n^2}\nabla n + \frac{\partial^2 f_{\text{e–h}}}{\partial n\,\partial p}\nabla p + \frac{\partial^2 f_{\text{e–h}}}{\partial n\,\partial T}\nabla T, \end{aligned}$$

where we have separated the contributions from the Coulomb interaction energy f_{Coul} (leading to drift in the electric field) and the quasi-free electron-hole plasma (yielding Hessian matrix elements of the ideal Fermi gas' free energy density $f_{\text{e–h}}$). The electrons and holes are decoupled in the non-interacting Fermi-gas such that

$$\frac{\partial^2 f_{\text{e–h}}}{\partial n\,\partial p} = 0.$$

Moreover, since (i) the Coulomb interaction energy is independent of the temperature and therefore does not contribute to the system's entropy and (ii) the lattice contribution f_L is independent of the carrier densities, it holds

$$\frac{\partial^2 f_{\text{e–h}}}{\partial n\,\partial T} = -\frac{\partial s}{\partial n},$$

where s is the entropy density of the full system (see the last formula in Eq. (10)). Thus, we arrive at

$$\nabla\varphi_n = \nabla\phi - \frac{1}{q}\frac{\partial^2 f_{\text{e–h}}}{\partial n^2}\nabla n + \frac{1}{q}\frac{\partial s}{\partial n}\nabla T,$$

which must be substituted in Eq. (8) to obtain

$$\mathbf{j}_n = -\sigma_n\left(\nabla\phi - \frac{1}{q}\frac{\partial^2 f_{\text{e–h}}}{\partial n^2}\nabla n + \left[\frac{1}{q}\frac{\partial s}{\partial n} + P_n\right]\nabla T\right) = -\sigma_n\nabla\phi + \sigma_n\frac{1}{q}\frac{\partial^2 f_{\text{e–h}}}{\partial n^2}\nabla n.$$

In the last step we have used the Kelvin formula (9) for the Seebeck coefficient. The temperature gradient term vanishes exactly, since reversing the order of the derivatives in the Hessian of the free energy density immediately yields the definition (9) and cancels with the Seebeck term in the thermodynamic form (8). The same result can be obtained by simply inverting the carrier density state equation (5) and using the explicit expression (15). With the electrical conductivities $\sigma_n = qM_n n$, $\sigma_p = qM_p p$ and

$$\frac{\partial^2 f_{\text{e–h}}}{\partial n^2} = \frac{k_B T}{n} g\left(\frac{n}{N_c(T)}\right), \qquad \frac{\partial^2 f_{\text{e–h}}}{\partial p^2} = \frac{k_B T}{p} g\left(\frac{p}{N_v(T)}\right),$$

(from Eq. (14a)), we finally arrive at the drift-diffusion form:

$$\mathbf{j}_n = -qM_n n\nabla\phi + qD_n(n,T)\nabla n, \qquad \mathbf{j}_p = -qM_p p\nabla\phi - qD_p(p,T)\nabla p. \tag{21}$$

The diffusion coefficients are given by the *generalized* Einstein relation [26, 75]

$$D_n(n,T) = \frac{k_B T M_n}{q} g\left(\frac{n}{N_c(T)}\right), \qquad D_p(p,T) = \frac{k_B T M_p}{q} g\left(\frac{p}{N_v(T)}\right). \tag{22}$$

Here the degeneracy factor g describes an effective enhancement of the diffusion current that depends nonlinearly on the carrier densities, which results from the increased average thermal energy of the carriers in the case of Fermi–Dirac statistics (see above). The diffusion enhancement due to carrier degeneracy has been found to be important in, e. g., semiconductor lasers [76] as well as organic field-effect transistors [77] and light emitting diodes [58]. We emphasize that the drift-diffusion form (21) of the current densities is fully equivalent to the thermodynamic form (8). Thus, even though the ∇T term is eliminated, the thermoelectric cross-coupling via the Seebeck effect is fully taken into account via the temperature dependency of the diffusion coefficient. For Seebeck coefficients that deviate from the Kelvin formula, additional thermodiffusion terms emerge.

3 Non-isothermal generalization of the Scharfetter–Gummel scheme for degenerate semiconductors

The typically exponentially varying carrier densities in semiconductor devices lead to numerical instabilities when using a standard finite difference discretization. In particular, the naive discretization approach results in spurious oscillations and may cause unphysical results such as negative carrier densities [78, 79]. A robust discretization scheme for the drift-diffusion current density was introduced by Scharfetter and Gummel [43], who explicitly solved the current density expressions as a separate differential equation along the edge between two adjacent nodes of the mesh. The resulting discretized current density expressions feature exponential terms that reflect the characteristics of the doping profile and allow for numerically stable calculations. Over the last decades, several generalizations of the Scharfetter–Gummel method have been proposed for either degenerate semiconductors [52–57] or non-isothermal carrier transport with included thermoelectric cross effects [44–51].

In this section, we derive two different generalizations of the Scharfetter–Gummel scheme for degenerate semiconductors obeying the Kelvin formula for the Seebeck coefficient. Both schemes differ in the treatment of degeneration effects and are obtained by extending the approaches previously developed in Refs. [54, 56] and [52]. First, we outline the finite volume method in Sec. 3.1 and then introduce the non-isothermal Scharfetter–Gummel schemes in Sec. 3.2. We study important limiting cases and structure preserving properties of the discretizations (Sec. 3.3), give a detailed comparison with the numerically exact solution of the underlying two-point boundary value problem (Sec. 3.4) and derive analytical error bounds (Sec. 3.5). Finally, we present a numerical convergence analysis by means of numerical simulations of a one-dimensional p-n-diode in Sec. 3.6.

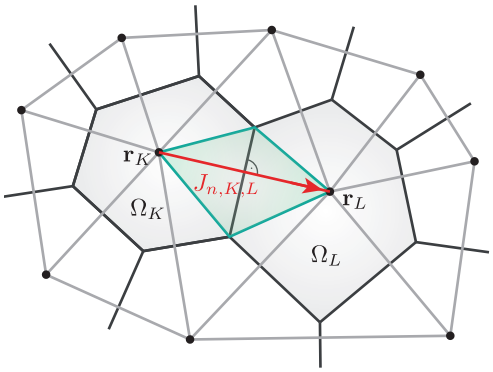

Fig. 3. Delaunay triangulation and construction of Voronoï cells. The red arrow indicates the discrete current $J_{n,K,L}$ between two neighboring control volumes Ω_K and Ω_L. The green area is the bi-hyperpyramid with height $\|\mathbf{r}_L - \mathbf{r}_K\|$ and internal face $|\partial\Omega_L \cap \partial\Omega_K|$.

3.1 Finite volume discretization

We assume a boundary conforming Delaunay triangulation [80] of the point set $\mathbf{R} = \{\mathbf{r}_K\}_{K=1\ldots N_{\text{nodes}}}$, $\mathbf{r}_K \in \Omega$, where $\Omega \subset \mathbb{R}^d$ is the computational domain with dimensionality d. The dual mesh is given by the Voronoï cells

$$\Omega_K = \{\mathbf{r} \in \Omega : \|\mathbf{r} - \mathbf{r}_K\| \leq \|\mathbf{r} - \mathbf{r}_L\| \text{ for all } \mathbf{r}_L \in \mathbf{R} \text{ with } \mathbf{r}_L \neq \mathbf{r}_K\},$$

which provide a non-overlapping tessellation $\Omega = \bigcup_K \Omega_K$ of the domain. This represents an admissible mesh in the sense of Ref. [81]. The finite volume discretization of the system (1)–(4) is obtained by integration over the cell Ω_K and usage of the divergence theorem [79, 81]. The discrete (stationary) non-isothermal drift-diffusion system reads

$$-\sum_{L\in\mathcal{N}(K)} s_{K,L}\varepsilon\left(\phi_L - \phi_K\right) = q|\Omega_K|\left(C_K + p_K - n_K\right) \tag{23a}$$

$$-\sum_{L\in\mathcal{N}(K)} s_{K,L} J_{n,K,L} = -q|\Omega_K|R_K, \tag{23b}$$

$$+\sum_{L\in\mathcal{N}(K)} s_{K,L} J_{p,K,L} = -q|\Omega_K|R_K, \tag{23c}$$

$$-\sum_{L\in\mathcal{N}(K)} s_{K,L}\kappa\left(T_L - T_K\right) = \frac{1}{2}\sum_{L\in\mathcal{N}(K)} s_{K,L}\left(H_{J,K,L} + H_{\mathrm{T--P},K,L}\right) + |\Omega_K|H_{R,K} \tag{23d}$$

with the flux projections on the edge $\overline{KL}$

$$J_{n,K,L} = (\mathbf{r}_L - \mathbf{r}_K)\cdot\mathbf{j}_n, \qquad J_{p,K,L} = (\mathbf{r}_L - \mathbf{r}_K)\cdot\mathbf{j}_p.$$

The geometric factors in Eq. (23) are the volume $|\Omega_K|$ of the K-th Voronoï cell and the edge factor

$$s_{K,L} = \frac{|\partial\Omega_K \cap \partial\Omega_L|}{\|\mathbf{r}_L - \mathbf{r}_K\|}.$$

The symbol $\mathcal{N}(K)$ denotes the set of nodes adjacent to K. For the sake of simplicity we restrict ourselves to the case of a homogeneous material. This limitation is not an important for the flux discretization, as the discrete fluxes appear only along possible heterointerfaces (edges of the primary simplex grid, see Fig. (3)) but never across. In the case of heterostructures, the currents along material

interfaces are weighted by the respective edge factors. Boundary terms on $\partial\Omega \cap \Omega_K \neq \emptyset$ are omitted in Eq. (23). The discrete electron density reads $n_K = N_c(T_K)\,\mathscr{F}(\eta_{n,K})$ with the reduced Fermi energy $\eta_{n,K} = -\left(E_c(T_K) - q\phi_K + q\varphi_{n,K}\right)/\left(k_B T_K\right)$ (holes analogously). The recombination rate is $R_K = R(\phi_K, \varphi_{n,K}, \varphi_{p,K}, T_K)$ and the discrete self-heating terms are

$$\begin{aligned} H_{J,K,L} &= -J_{n,K,L}\left(\varphi_{n,L} - \varphi_{n,K} + P_{n,K,L}\left(T_L - T_K\right)\right) \\ &\quad - J_{p,K,L}\left(\varphi_{p,L} - \varphi_{p,K} + P_{p,K,L}\left(T_L - T_K\right)\right) \end{aligned} \tag{24a}$$

$$H_{\mathrm{T}-\!-\mathrm{P},K,L} = -T_{K,L} J_{n,K,L}\left(P_{n,L} - P_{n,K}\right) - T_{K,L} J_{p,K,L}\left(P_{p,L} - P_{p,K}\right), \tag{24b}$$

$$H_{R,K} = q\left(\varphi_{p,K} - \varphi_{n,K} + T_K\left(P_{p,K} - P_{n,K}\right)\right) R_K. \tag{24c}$$

We remark that the discretization of the Joule heating and Thomson–Peltier terms in Eq. (20) is not straightforward, as this requires information about the discrete gradients on the control volumes and not just along the edges between the nodes [82–84]. The discrete heat equation (23d) together with Eqs. (24a)–(24b) was obtained in a careful way following Ref. [82]. The discretization of the edge current densities $J_{n/p,K,L}$, the edge-averaged temperature $T_{K,L}$ and the Seebeck coefficients $P_{n/p,K,L}$ along the edge $\overline{KL}$ are subject to the following sections.

3.2 Discretization of the current density expression

The discretization of $J_{n/p,K,L}$ is obtained by integrating the current density expressions (21) along the edge $\overline{KL}$ between two adjacent nodes of the mesh. Since the Kelvin formula implies a remarkably simple form of the electrical current densities in drift-diffusion form, where the thermal driving force is eliminated exactly (see Sec. 2.4), this allows for a straightforward adaptation of the Scharfetter–Gummel schemes developed for the isothermal case. We assume the electrostatic field $\mathbf{E} = -\nabla\phi$ and the temperature gradient ∇T to be constant along the edge $\overline{KL}$

$$\phi(x) = x\,\phi_L + (1-x)\,\phi_K, \qquad\qquad T(x) = x\,T_L + (1-x)\,T_K,$$

where $x \in [0,1]$ parametrizes the coordinate on the edge $\mathbf{r}(x) = x\,\mathbf{r}_L + (1-x)\,\mathbf{r}_K$. These assumptions have been already used above in Eqs. (23a) and (23d). Moreover, also the mobilities $M_{n/p}$ and the fluxes $J_{n/p,K,L}$ are assumed to be constant on the edge. For the electron current density, this yields the two-point boundary value problem (BVP)

$$\begin{aligned} k_B T(x)\, g\left(\frac{n(x)}{N_c(T(x))}\right)\frac{\mathrm{d}n}{\mathrm{d}x} &= q\left(\phi_L - \phi_K\right) n(x) + \frac{J_{n,K,L}}{M_n}, \\ n(0) = n_K, \qquad n(1) &= n_L, \end{aligned} \tag{25}$$

on $x = [0,1]$. The problem for the holes current density is analogous.

In the non-degenerate case (Maxwell–Boltzmann statistics) the degeneracy factor is exactly $g \equiv 1$, such that the problem can be solved exactly by separation of variables. One obtains

$$\int_{n_K}^{n_L} \frac{\mathrm{d}n}{\frac{J_{n,K,L}}{qM_n(\phi_L - \phi_K)} + n} = \frac{q\left(\phi_L - \phi_K\right)}{k_B}\int_0^1 \frac{\mathrm{d}x}{T(x)},$$

where the integral on the right hand side yields the (inverse) logarithmic mean temperature

$$\int_0^1 \frac{\mathrm{d}x}{T(x)} = \int_0^1 \frac{\mathrm{d}x}{x\,T_L + (1-x)\,T_K} = \frac{1}{\Lambda(T_L, T_K)} \equiv \frac{1}{T_{K,L}}, \quad \Lambda(x,y) = \frac{x-y}{\log(x/y)}, \tag{26}$$

with the logarithmic mean $\Lambda(x, y)$. Solving for the flux yields the non-isothermal Scharfetter–Gummel scheme

$$J^{\mathsf{ndeg}}_{n,K,L} = M_n k_B T_{K,L} \left(n_L B\left(X^{\mathsf{ndeg}}_{n,K,L}\right) - n_K B\left(-X^{\mathsf{ndeg}}_{n,K,L}\right)\right), \qquad X^{\mathsf{ndeg}}_{n,K,L} = \frac{q\left(\phi_L - \phi_K\right)}{k_B T_{K,L}}, \quad (27)$$

where $B(x) = x/(\exp(x) - 1)$ is the Bernoulli function. The Bernoulli function is closely related to the logarithmic mean: $B(x) = 1/\Lambda(x, 1)$. At isothermal conditions $T_K = T_L$, Eq. (27) reduces to the original Scharfetter–Gummel scheme [43].

In the case of Fermi–Dirac statistics $g \neq 1$, no closed form solution exists such that approximate solutions of the BVP (25) are required. As the degeneracy factor g depends on both the carrier density and temperature, the problem is not even separable

3.2.1 Modified thermal voltage scheme

Following Refs. [54, 56], we solve the BVP (25) by freezing the degeneracy factor $g\left(n/N_c(T)\right) \to g_{n,K,L}$ to a carefully chosen average. The resulting problem has the same structure as in the non-degenerate case (see above) with a modified thermal voltage $k_B T_{K,L}/q \to k_B T_{K,L} g_{n,K,L}/q$ along the edge, which takes the temperature variations and the degeneration of the electron gas into account. This yields the *modified thermal voltage scheme*

$$\begin{aligned} J^g_{n,K,L} &= M_n k_B T_{K,L} g_{n,K,L} \left(n_L B\left(X^g_{n,K,L}\right) - n_K B\left(-X^g_{n,K,L}\right)\right), \\ X^g_{n,K,L} &= \frac{q\left(\phi_L - \phi_K\right)}{k_B T_{K,L} g_{n,K,L}}, \end{aligned} \quad (28a)$$

where $T_{K,L}$ denotes the logarithmic mean temperature (26). In order to ensure the consistency with the thermodynamic equilibrium and boundedness $g_{n,K} \leq g_{n,K,L} \leq g_{n,L}$ (for $\eta_{n,K} \leq \eta_{n,L}$ or with $K \leftrightarrow L$ else), the edge-averaged degeneracy factor is taken as [54, 56]

$$g_{n,K,L} = \frac{\eta_{n,L} - \eta_{n,K}}{\log\left(\mathscr{F}(\eta_{n,L})/\mathscr{F}(\eta_{n,K})\right)} = \frac{\mathscr{F}^{-1}\left(n_L/N_c(T_L)\right) - \mathscr{F}^{-1}\left(n_K/N_c(T_K)\right)}{\log\left(n_L/N_c(T_L)\right) - \log\left(n_K/N_c(T_K)\right)}. \quad (28b)$$

In the limit of $\eta_{n,L} = \eta_{n,K}$, it approaches the common nodal value

$$\lim_{\eta_{n,L} \to \eta_{n,K} \equiv \bar{\eta}_n} g_{n,K,L} = \frac{\mathscr{F}(\bar{\eta}_n)}{\mathscr{F}'(\bar{\eta}_n)} = g\left(\mathscr{F}(\bar{\eta}_n)\right) \equiv g_\eta(\bar{\eta}_n).$$

For constant temperature $T_L = T_K = T$ the scheme reduces to the modified Scharfetter–Gummel scheme discussed in Refs. [54, 56]. It can thus be regarded as a non-isothermal generalization of this approach. In the non-degenerate limit $g = 1$, it reduces to the non-isothermal, non-degenerate Scharfetter–Gummel scheme (27).

3.2.2 Modified drift scheme

The traditional approach for the inclusion of degeneration effects in the Scharfetter–Gummel scheme, that is widely used in commercial software packages, is based on introducing the correction factors [52, 85, 86]

$$\gamma(\eta) = \frac{\mathscr{F}(\eta)}{\exp(\eta)}, \quad (29)$$

and rearranging the current density expression with nonlinear diffusion (21) (involving the generalized Einstein relation (22)) into a form with linear diffusion and a modified drift term:

$$\begin{aligned}\mathbf{j}_n &= \sigma_n \mathbf{E}_n + M_n k_B T \nabla n + \frac{k_B}{q}\sigma_n \rho_n\left(T,\eta_n\right)\nabla T,\\ \mathbf{E}_n &= -\nabla\left(\phi + \frac{k_B T}{q}\log\left(\gamma\left(\eta_n\right)\right)\right).\end{aligned} \tag{30}$$

Here, the degeneration of the electron gas yields a thermodiffusion term with the coefficient

$$\rho_n\left(T,\eta_n\right) = \log\left(\gamma(\eta_n)\right) - \frac{TN_c'(T)}{N_c(T)}\frac{\gamma'(\eta_n)/\gamma(\eta_n)}{1+\gamma'(\eta_n)/\gamma(\eta_n)} = \log\left(\gamma(\eta_n)\right) + \frac{TN_c'(T)}{N_c(T)}\left(g_\eta(\eta_n)-1\right), \tag{31}$$

that vanishes exactly in the non-degenerate limit $\gamma\left(\eta\right) \equiv 1 \equiv g_\eta\left(\eta\right)$. Hence, the function $\rho_n\left(T,\eta_n\right)$ quantifies the difference between the degenerate and the non-degenerate Seebeck-coefficient, see Fig. 2 (a). On the step from Eq. (21) to (30), we have used the relation $g_\eta\left(\eta\right) = \left(1+\gamma'\left(\eta\right)/\gamma\left(\eta\right)\right)^{-1}$. A plot of the correction factor (29) is given in Fig. 1 (c).

The current density expression (30) is discretized by projecting the current on the edge $\overline{KL}$, assuming the effective electric field $\mathbf{E}_n$ to be a constant along the edge, and freezing $\rho_n\left(T,\eta_n\right) \to \rho_{n,K,L}$ to a constant average value. Along the same lines as above, one arrives at the *modified drift scheme*

$$J^\gamma_{n,K,L} = M_n k_B T_{K,L}\left(n_L B\left(X^\gamma_{n,K,L}\right) - n_K B\left(-X^\gamma_{n,K,L}\right)\right), \tag{32a}$$

with

$$X^\gamma_{n,K,L} = \frac{q\left(\phi_L-\phi_K\right)}{k_B T_{K,L}} + \frac{T_L\log\left(\gamma\left(\eta_{n,L}\right)\right) - T_K\log\left(\gamma\left(\eta_{n,K}\right)\right)}{T_{K,L}} - \rho_{n,K,L}\log\left(\frac{T_L}{T_K}\right). \tag{32b}$$

Again, $T_{K,L}$ is the logarithmic mean temperature (26). The corresponding non-degenerate limit is easily obtained by $\gamma\left(\eta_{n,L/K}\right) \to 1$, $\rho_{n,K,L} \to 0$ and yields Eq. (27).

3.3 Limiting cases and structure preserving properties

In the following, we investigate some important limiting cases and structure preserving properties of the generalized Scharfetter–Gummel schemes (28) and (32). This includes an analysis of the consistency of the discrete expressions with fundamental thermodynamical principles (thermodynamic equilibrium, second law of thermodynamics). To this end, it is convenient to rewrite both expressions using the identity $B\left(-x\right) = \exp\left(x\right)B\left(x\right)$ and the logarithmic mean Λ (see Eq. (26)) as

$$\begin{aligned}J^g_{n,K,L} &= -\sigma^g_{n,K,L}\left(\varphi_{n,L} - \varphi_{n,K} + P^g_{n,K,L}\left(T_L - T_K\right)\right)\\ \text{with}\quad \sigma^g_{n,K,L} &= qM_n\frac{\Lambda\left(n_L\exp\left(-\frac{1}{2}X^g_{n,K,L}\right), n_K\exp\left(\frac{1}{2}X^g_{n,K,L}\right)\right)}{\operatorname{sinhc}\left(\frac{1}{2}X^g_{n,K,L}\right)},\end{aligned} \tag{33}$$

and

$$\begin{aligned}J^\gamma_{n,K,L} &= -\sigma^\gamma_{n,K,L}\left(\varphi_{n,L} - \varphi_{n,K} + P^\gamma_{n,K,L}\left(T_L - T_K\right)\right)\\ \text{with}\quad \sigma^\gamma_{n,K,L} &= qM_n\frac{\Lambda\left(n_L\exp\left(-\frac{1}{2}X^\gamma_{n,K,L}\right), n_K\exp\left(\frac{1}{2}X^\gamma_{n,K,L}\right)\right)}{\operatorname{sinhc}\left(\frac{1}{2}X^\gamma_{n,K,L}\right)},\end{aligned} \tag{34}$$

where $\operatorname{sinhc}(x) = \sinh(x)/x$. This representation directly corresponds to the continuous current density expression in the thermodynamic form (8), where the conductivity along the edge $\sigma^{g/\gamma}_{n,K,L}$ is determined by a "tilted" logarithmic average of the nodal carrier densities. Both expressions (33)–(34) have a common structure, but differ in the discrete conductivity $\sigma^{g}_{n,K,L} \neq \sigma^{\gamma}_{n,K,L}$ (due to $X^{g}_{n,K,L} \neq X^{\gamma}_{n,K,L}$) and the discrete Seebeck coefficients $P^{g}_{n,K,L} \neq P^{\gamma}_{n,K,L}$ along the edge, which are implicitly taken by the Scharfetter–Gummel discretization procedure. The latter are obtained as

$$P^{g}_{n,K,L} = -\frac{k_B}{q}\left[\log\left(\frac{N_c(T_L)}{N_c(T_K)}\right)\frac{g_{n,K,L}}{\log(T_L/T_K)} - \frac{(T_L - T_{K,L})\,\eta_{n,L} - (T_K - T_{K,L})\,\eta_{n,K}}{T_L - T_K} - \frac{1}{k_B}\frac{E_c(T_L) - E_c(T_K)}{T_L - T_K}\right] \tag{35a}$$

and

$$P^{\gamma}_{n,K,L} = -\frac{k_B}{q}\left[\log\left(\frac{N_c(T_L)}{N_c(T_K)}\right)\frac{1}{\log(T_L/T_K)} + \rho_{n,K,L} - \frac{T_L\log(\gamma(\eta_{n,L})) - T_K\log(\gamma(\eta_{n,K})) - T_{K,L}\log(\gamma(\eta_{n,L})/\gamma(\eta_{n,K}))}{T_L - T_K} - \frac{(T_L - T_{K,L})\,\eta_{n,L} - (T_K - T_{K,L})\,\eta_{n,K}}{T_L - T_K} - \frac{1}{k_B}\frac{E_c(T_L) - E_c(T_K)}{T_L - T_K}\right]. \tag{35b}$$

The discrete Seebeck coefficients (35) enter the discrete Joule heat term (24a) and thus the discrete entropy production rate (see Sec. 3.3.5 below). Outside of the thermodynamic equilibrium, the Seebeck coefficients (35) determine the point of compensating (discrete) chemical and thermal current flow $j_{n,K,L}|_{\varphi_{n,K}\neq\varphi_{n,L},T_K\neq T_L} = 0$. This compensation point is in general slightly different between both schemes (see inset of Fig. 4 (c)). In the limit of a small temperature gradient and a small difference in the reduced Fermi energy along the edge, both discrete Seebeck coefficients approach the continuous expression (15a)

$$P^{g/\gamma}_{n,K,L} = -\frac{k_B}{q}\left[\frac{\bar{T}N_c'(\bar{T})}{N_c(\bar{T})}g_\eta(\bar{\eta}_n) - \bar{\eta}_n - \frac{1}{k_B}E_c'(\bar{T})\right] + \mathcal{O}(\delta\eta_n^2) + \mathcal{O}(\delta\eta_n\,\delta\Theta) + \mathcal{O}(\delta\Theta^2),$$

where $\delta\Theta = (T_L - T_K)/\bar{T}$, $\bar{T} = \frac{1}{2}(T_L + T_K)$, $\delta\eta_n = \eta_{n,L} - \eta_{n,K}$ and $\bar{\eta}_n = \frac{1}{2}(\eta_{n,L} + \eta_{n,K})$.

3.3.1 Thermodynamic equilibrium

In the thermodynamic equilibrium (thermal $T_K = T_L$ and chemical equilibrium $\varphi_{n,K} = \varphi_{n,L}$), both the discrete current densities (28) and (32) are exactly zero. This is easily seen from Eqs. (33)–(34), where the discrete driving force $\left(\varphi_{n,L} - \varphi_{n,K} + P^{g/\gamma}_{n,K,L}(T_L - T_K)\right)$ vanishes under thermodynamic equilibrium conditions.

3.3.2 Strong electric field (drift-dominated limit)

Due to the asymptotics of the Bernoulli function $B(x \to \infty) = 0$ and $B(x \to -\infty) \sim -x$, the modified thermal voltage scheme (28) approaches the first-order upwind scheme

$$J^{g}_{n,K,L}(\delta\phi_{K,L} \to \pm\infty) \sim -qM_n\left(\frac{n_L + n_K}{2} + \frac{n_K - n_L}{2}\operatorname{sign}(\delta\phi_{K,L})\right)\delta\phi_{K,L} = J^{\text{upw}}_{n,K,L} \tag{36}$$

in the limit of a strong electrostatic potential gradient $\delta\phi_{K,L} = \phi_L - \phi_K \to \pm\infty$. The upwind scheme is a stable, first order accurate discretization for advection-dominated problems, where the coefficient is evaluated in the "donor cell" of the flow [87]. Hence, this asymptotic feature of the original Scharfetter–Gummel scheme, which is important for the robustness of the discretization as it avoid spurious oscillations, is preserved in the degenerate and non-isothermal case. The modified drift scheme (32) approaches the upwind scheme as well

$$J^{\gamma}_{n,K,L}\left(\delta\phi_{K,L}\to\pm\infty\right) \sim -qM_n\left(\frac{n_L+n_K}{2}+\frac{n_K-n_L}{2}\operatorname{sign}\left(\delta\phi_{K,L}\right)\right)\times \\ \times\left(\delta\phi_{K,L}+\frac{k_B}{q}\left(\log\left(\frac{\left[\gamma\left(\eta_{n,L}\right)\right]^{T_L}}{\left[\gamma\left(\eta_{n,K}\right)\right]^{T_K}}\right)-\left(T_L-T_K\right)\rho_{n,K,L}\right)\right),$$

however, in the case of strong degeneration the convergence is significantly slowed down if the nodal correction factors $\gamma\left(\eta_{n,K}\right) \neq \gamma\left(\eta_{n,K}\right)$ and temperatures $T_L \neq T_K$ are very different. This is illustrated in Fig. 4 (c), where the modified drift scheme shows a constant offset from the numerically exact solution of the BVP (25) for $\delta\phi_{K,L}\to-\infty$.

3.3.3 No electric field (diffusive limit)

In the case of a vanishing electrostatic potential gradient $\delta\phi_{K,L}=0$ the schemes take the form

$$\lim_{\delta\phi_{K,L}\to0} J^{g}_{n,K,L} = M_n k_B T_{K,L} g_{n,K,L}\left(n_L-n_K\right), \tag{37}$$

$$\lim_{\delta\phi_{K,L}\to0} J^{\gamma}_{n,K,L} = M_n k_B T_{K,L}\frac{1}{\Lambda\left(\frac{T_L^{\rho_{n,K,L}}}{\left[\gamma\left(\eta_{n,L}\right)\right]^{T_L/T_{K,L}}},\frac{T_K^{\rho_{n,K,L}}}{\left[\gamma\left(\eta_{n,K}\right)\right]^{T_K/T_{K,L}}}\right)}\times \\ \times\left(\frac{T_L^{\rho_{n,K,L}}}{\left[\gamma\left(\eta_{n,L}\right)\right]^{T_L/T_{K,L}}}n_L-\frac{T_K^{\rho_{n,K,L}}}{\left[\gamma\left(\eta_{n,K}\right)\right]^{T_K/T_{K,L}}}n_K\right).$$

The modified thermal voltage scheme (28) approaches the central finite difference discretization (37), which is a stable discretization for diffusion-dominated transport problems [87]. With the edge-averaged degeneracy factor $g_{n,K,L}$, Eq. (37) nicely reflects the structure of the diffusive part of the continuous current density expression (21) involving the generalized Einstein relation (22). For the modified drift scheme (32), the limiting expression is a weighted finite difference discretization. Due to the different treatment of the degeneracy of the electron gas via the correction factors (29), it does not yield a discrete analogue of the generalized Einstein relation.

3.3.4 Purely thermally driven currents

In the chemical equilibrium ($\varphi_{n,L}=\varphi_{n,K}$), the current is driven only by the temperature gradient. The corresponding expressions are easily obtained from Eqs. (33)–(34), which include the discrete Seebeck coefficients (35).

3.3.5 Non-negativity of the discrete dissipation rate

The continuous entropy production rate (dissipation rate) per volume (see Eq. (44))

$$\dot{s}_{\text{tot}} = \frac{1}{T}\left(\mu_c-\mu_v\right)R+\frac{\kappa}{T^2}\left\|\nabla T\right\|^2+\frac{1}{T}H_J,$$

has contributions from carrier recombination, heat flux and Joule heating $H_J = -\left(\nabla\varphi_n + P_n\nabla T\right)\cdot \mathbf{j}_n - \left(\nabla\varphi_p + P_p\nabla T\right)\cdot\mathbf{j}_p$. With the current density expressions (8) and a recombination rate of the form [63]

$$R = \left(1 - \exp\left(\frac{\mu_c - \mu_v}{k_B T}\right)\right)\sum_\alpha r_\alpha\left(\phi, \varphi_n, \varphi_p, T\right)$$

(where α labels the different recombination processes), all terms in $\dot{s}_{\text{tot}}$ (including H_J) are evidently non-negative (i. e., zero in the thermodynamic equilibrium and positive else) and therefore ensure the second law of thermodynamics. In order to rule out unphysical effects such as steady state dissipation [54, 88], it is highly desirable to preserve this important structural property of the continuous system in its discrete counterpart. Given the finite volume discretization described above, this is straightforwardly achieved for the contributions from the carrier recombination and the heat flux, however, it is less obvious for the Joule heating term. In fact, the non-negativity of the discrete Joule heating term is non-trivial and can be violated in general when using a naive discretization approach as in Ref. [89].

We show that the discrete dissipation rate is evidently non-negative for both generalized Scharfetter–Gummel schemes (28) and (32). This follows immediately from their consistency with the thermodynamic equilibrium (see Sec. 3.3.1) in conjunction with the discrete form (24a) of the heating term. Substituting Eq. (33) in (24a), one obtains

$$\begin{aligned} H^g_{J,n} &= -\left(\varphi_{n,L} - \varphi_{n,K} + P^g_{n,K,L}\left(T_L - T_K\right)\right) J^g_{n,K,L} \\ &= \sigma^g_{n,K,L}\left|\varphi_{n,L} - \varphi_{n,K} + P^g_{n,K,L}\left(T_L - T_K\right)\right|^2 \geq 0, \end{aligned}$$

which is zero only in the (discrete) thermodynamic equilibrium and positive else. The discrete conductivity $\sigma^g_{n,K,L}$ is positive by construction, see Eq. (33). Analogous expressions are obtained for the holes' current contribution and the modified drift scheme. In conclusion, the consistency of the discrete system (23) with the second law of thermodynamics relies on using the respective Seebeck coefficients $P_{n/p,K,L}$ implied by the current discretization (see Eq. (35)) consistently in the discretized Joule heating term (24a). Only then this structural property of the discrete system holds without any smallness assumption.

3.4 Comparison with numerically exact solution

We investigate the accuracy of the schemes (28) and (32) by comparing them with the numerically exact solution of the BVP (25). In the isothermal case, this has been carried out before in a similar way by Farrell et al. [90] (for different Scharfetter–Gummel schemes), which inspired the investigation of highly accurate Scharfetter–Gummel type discretizations based on the direct numerical integration of the arising integral equation using quadrature rules in Ref. [91]. In the present non-isothermal case, the problem is more complicated because of the spatially varying temperature distribution along the edge. It is convenient to recast the problem (25) into the form

$$\begin{aligned} \frac{\mathrm{d}y\left(x\right)}{\mathrm{d}x} &= \frac{\bar{T}}{T\left(x\right)}\left(\delta\Phi + \frac{N_c(\bar{T})}{N_c(T(x))}\frac{J}{\mathscr{F}\left(y\right)} - \delta\Theta\frac{T(x)N_c'(T(x))}{N_c(T(x))}\frac{\mathscr{F}\left(y\right)}{\mathscr{F}'\left(y\right)}\right), \\ y\left(0\right) &= \bar{\eta}_n - \frac{1}{2}\delta\eta_n, \qquad\qquad y\left(1\right) = \bar{\eta}_n + \frac{1}{2}\delta\eta_n, \end{aligned} \tag{38}$$

with the notations $T\left(x\right) \equiv \left(1 + \left[x - \frac{1}{2}\right]\delta\Theta\right)\bar{T}$, $\bar{T} = \frac{1}{2}\left(T_L + T_K\right)$, $\delta T = T_L - T_K$, $\bar{\eta}_n = \frac{1}{2}\left(\eta_{n,L} + \eta_{n,K}\right)$, $\delta\eta_n = \eta_{n,L} - \eta_{n,K}$ and the non-dimensionalized quantities

$$J = \frac{J_{n,K,L}}{M_n k_B \bar{T} N_c(\bar{T})}, \qquad \delta\Phi = \frac{q\left(\phi_L - \phi_K\right)}{k_B\bar{T}}, \qquad \delta\Theta = \frac{\delta T}{\bar{T}}.$$

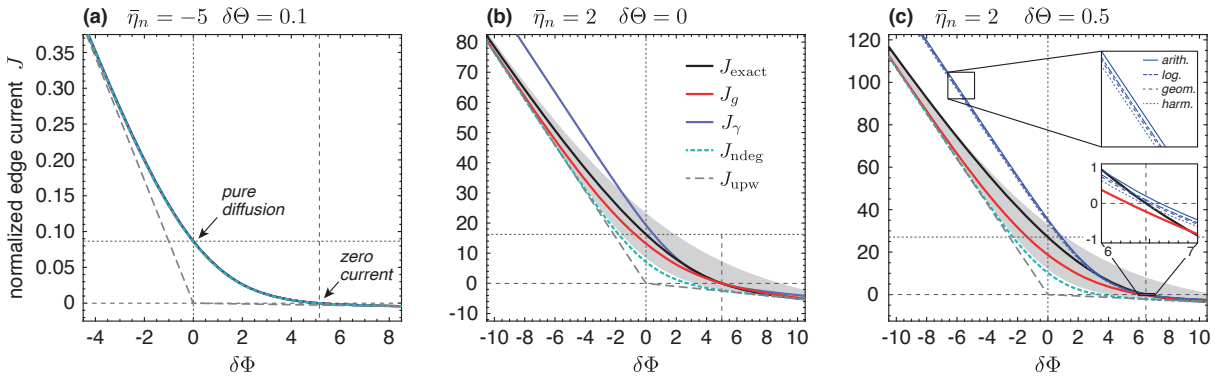

Fig. 4. Comparison of the non-isothermal Scharfetter–Gummel schemes for $\delta\eta_n = 5$. (a) In the non-degenerate regime ($\bar{\eta}_n = -5$) all schemes coincide, even in the presence of a temperature gradient. (b) At $\bar{\eta}_n = 2$ degeneration effects become significant. While J_g follows J_{exact} with an acceptable error over the whole range of $\delta\Phi$, the modified drift scheme J_γ has a significant offset at strong electric fields ($\delta\Phi \to -\infty$). The grey shaded area indicates the analytic bounds of the modified thermal voltage scheme determined using the nodal values $g_{n,K}$ and $g_{n,L}$ instead of $g_{n,K,L}$. (c) An additional temperature gradient increases the error in the degenerate regime. The insets show the effect of different averages $\rho_{n,K,L}$ of the nodal values $\rho_{n,K}$ and $\rho_{n,L}$ in the modified drift scheme and the different behavior of the schemes in the region of vanishing discrete currents.

The exact current $J_{\text{exact}} = J_{\text{exact}}(\delta\Phi, \delta\eta_n, \delta\Theta, \bar{\eta}_n, \bar{T})$ is a function of five parameters that satisfies the BVP (38). We solve Eq. (38) numerically using the shooting method [92], where we combine a 4th order Runge–Kutta method with Brent's root finding algorithm [93] . The problem is invariant under the simultaneous transformation

$$\delta\Phi \to -\delta\Phi, \qquad \delta\eta_n \to -\delta\eta_n, \qquad \delta\Theta \to -\delta\Theta, \qquad x \to 1 - x, \qquad J \to -J$$

(i. e., the sign of the current changes when changing the nodes $K \leftrightarrow L$), such that we can restrict our analysis to $\delta\Theta \geq 0$, when exploring the accuracy of the discrete current in the $(\delta\Phi, \delta\eta_n)$-plane. The comparison is carried out for $\mathscr{F}(\eta) = F_{1/2}(\eta)$ and $N_c = 2\left(m_c^* k_B T/(2\pi\hbar^2)\right)^{3/2}$. The Fermi–Dirac integrals $F_{1/2}(\eta)$ and $F_{-1/2}(\eta)$ are evaluated using MacLeod's algorithm [94].

Figure 4 shows the numerically exact current J_{exact} along with the approximations J_g (modified thermal voltage scheme (28)) and J_γ (modified drift scheme (32)) as a function of the normalized electric field $\delta\Phi$ along the edge. For weak degeneracy, both schemes agree with the numerically exact solution – even in the case of non-isothermal conditions. This is shown in Fig. 4 (a) for $\bar{\eta}_n = -5$ and $\delta\Theta = 0.1$. At strong electric fields $\delta\Phi \to \pm\infty$, all schemes approach the upwind scheme J_{upw} (grey dashed line, cf. Eq. (36)). The schemes (28) and (32) differ in the treatment of degeneration effects, which becomes apparent for increased $\bar{\eta}_n$. Figure 4 (b) shows the results for $\bar{\eta}_n = 2$ at isothermal conditions $\delta\Theta = 0$. The modified thermal voltage scheme J_g (red line) yields an acceptable deviation from the exact result J_{exact} (black line) over the whole range of $\delta\Phi$. The error vanishes at strong electric fields where both J_g and J_{exact} converge to the upwind scheme J_{upw}. The modified drift scheme (purple line), however, shows a significant error at large (negative) $\delta\Phi$, where it overestimates the current density significantly (about 33 % relative error at $\delta\Phi = -3$). This behavior results from the different treatment of the degeneration effects, that degrades the convergence of the modified drift scheme in the case of the strong degeneration (see Sec. 3.3.2). The plot highlights two other important exceptional points (pure diffusion and zero current), where both schemes show a similar accuracy. In the presence of an additional temperature gradient along the edge, see Fig. 4 (c), the approximation error of both schemes increases. The upper inset shows that the choice of the average of $\rho_{n,K,L}$ (see Eq. (31)) has only a minor impact on the modified drift scheme. The lower inset zooms on the region where the currents become zero. Here, all schemes provide a satisfying accuracy, but none of them is exact,

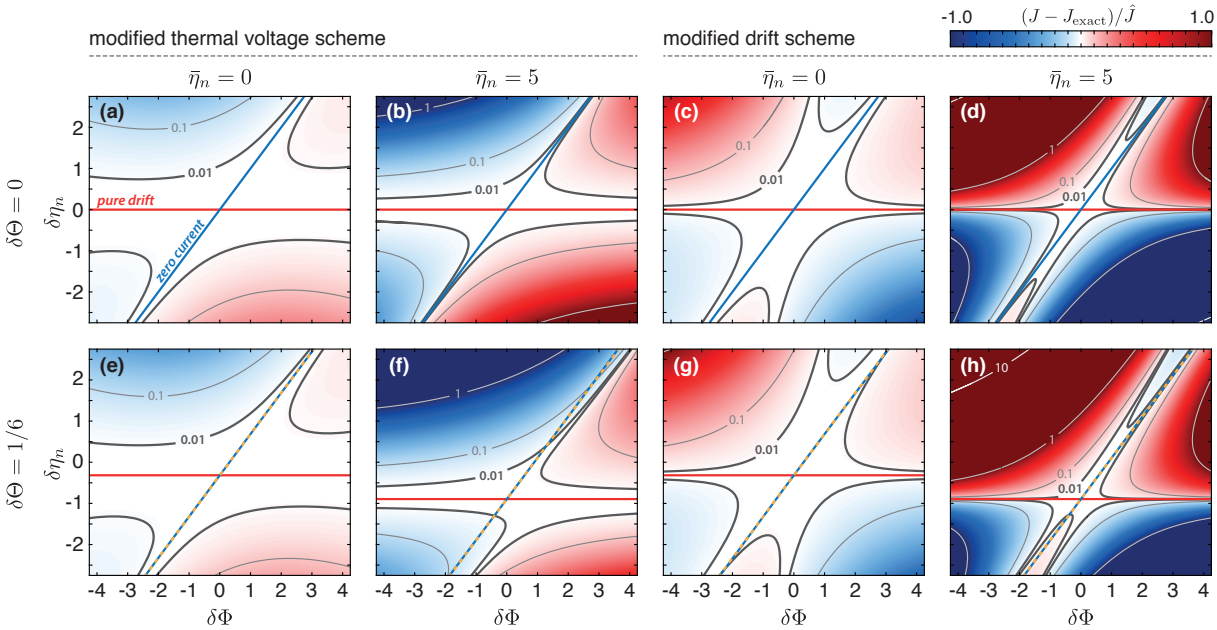

Fig. 5. Comparison of the two non-isothermal Scharfetter–Gummel schemes (28) and (32) (with arithmetic average $\rho_{n,K,L} = (\rho_{n,K} + \rho_{n,L})/2$) in the $(\delta\Phi, \delta\eta_n)$-plane at isothermal ($\delta\Theta = 0$, top row (a)–(d)) and non-isothermal ($\delta\Theta = 1/6$, bottom row (e)–(h)) conditions and different levels of degeneration (weak $\bar{\eta}_n = 0$ or strong $\bar{\eta}_n = 5$). The normalized absolute error $(J - J_{\text{exact}})/\hat{J}$ is color-coded for the range $[-1, 1]$ (dark colored regions correspond to larger errors, see the level lines). See the text for a discussion.

i. e., they yield a small spurious discrete current and intersect with the exact solution only in the vicinity of the exact zero current point. We observe that the modified drift schemes show a slightly better performance in this case, i. e., the Seebeck coefficient (35b) appears to be slightly better than (35a).

The normalized absolute errors $(J - J_{\text{exact}})/\hat{J}$ (with $\hat{J} = M_n k_B \bar{T} N_c(\bar{T})$) of the two schemes (28) and (32) are shown in the $(\delta\Phi, \delta\eta_n)$-plane in Fig. 5 under isothermal ($\delta\Theta = 0$, top row (a)–(d)) and non-isothermal ($\delta\Theta = 1/6$, bottom row (e)–(h)) conditions and for different levels of degeneration (weak $\bar{\eta}_n = 0$ and strong $\bar{\eta}_n = 5$). In the limit of very fine meshes $(\delta\Phi, \delta\eta_n, \delta\Theta) \to (0, 0, 0)$, both schemes coincide and the deviation from the exact current approaches zero. One observes that the size of the white regions with a normalized absolute error below 0.01 (in the following denoted as "low error domain") are generally larger for the modified thermal voltage scheme than for the modified drift scheme. Thus, the modified thermal voltage scheme is expected to yield a higher accuracy on sufficiently fine meshes. This will be evidenced by the numerical simulation of a p-n-diode in Sec. 3.6. The plots in Fig. 5 feature two additional lines, that refer to special limiting cases where both schemes yield a very high accuracy. The case of a pure drift current (i. e., no diffusion $n_L = n_K$) is indicated by a red line; the zero current line (blue) refers to the curve in the $(\delta\Phi, \delta\eta_n)$-plane where the exact current vanishes ($J_{\text{exact}} = 0$ in the BVP (38)). In the isothermal case ($\delta\Theta = 0$) the latter corresponds to the thermodynamic equilibrium, in the non-isothermal case $(\delta\Theta \neq 0)$ it refers to the situation of compensating chemical and thermal driving forces. Both schemes are exact in the case of a pure drift current, i. e., they asymptotically approach the upwind scheme (36), which is important for the robustness of the discretization in order to avoid spurious oscillations. The modified thermal voltage scheme shows a high accuracy also for slight deviations from the pure drift line, even in the case of strong degeneration, see Fig. 5 (b, f). In contrast, the modified drift scheme yields significant errors (much higher than 0.01) already for tiny deviations from the pure drift line in the strongly degenerate case, see Fig. 5 (d, h). This behavior has already been observed above in Fig. 4 (b, c) and was predicted analytically in Sec. 3.3.2. Note that in the non-isothermal case, the temperature gradient shifts the

pure drift line from $\delta\eta_n^{\text{pure drift}}|_{\delta\Theta=0} = 0$ (in Fig. 5 (a–d)) to $\delta\eta_n^{\text{pure drift}} \approx -\delta\Theta g_\eta(\bar{\eta}_n)\bar{T}N_c'(\bar{T})/N_c(\bar{T})$ (see Fig. 5 (e–h)). A prominent feature of the modified drift scheme is the additional intersection with the exact solution (see also Fig. 4 (c) at $\delta\Phi \approx 3.5$), which leads to additional "fingers" of the low error domain, see Fig. 5 (c, d, g, h), that are not associated with any special limiting case. The same feature has been observed for the so-called *inverse activity scheme* described in Ref. [90]. Finally, we study the consistency of the discretization schemes with the zero current line. In the isothermal case, both schemes are exact and therefore consistent with the thermodynamic equilibrium, see Fig. 5 (a–d). In the strongly degenerate case, however, the zero current line is only partially located within the low error domain, since the schemes intersect with the exact solution only in the vicinity of the zero current line and not exactly on it (see also the inset of Fig. 4 (c)). Nevertheless, the zero current lines of the discrete schemes, which are plotted as dashed orange lines in Fig. 5 (e–h), nicely overlap with the exact zero current line (blue). Thus, the spurious non-zero currents are very low and the little discrepancy in this limiting case is only of minor importance.

3.5 Analytical error estimate

We compare both schemes (28) and (32) by deriving an upper error bound. We follow the approach developed by Farrell et al. [90] and extend it to the non-isothermal case. Using the identities

$$B(x) - B(-x) = -x, \qquad B(x) + B(-x) = x\coth\left(\frac{x}{2}\right),$$

we obtain the series expansion of the discrete currents (28) and (32) at $(\delta\Phi, \delta\eta_n, \delta\Theta) = (0,0,0)$ up to second order as

$$J_g = -\mathscr{F}(\bar{\eta}_n)\,\delta\Phi + \mathscr{F}(\bar{\eta}_n)\left(\delta\eta + \frac{\bar{T}N_c'(\bar{T})}{N_c(\bar{T})}g_\eta(\bar{\eta}_n)\,\delta\Theta\right)\frac{\tilde{X}_g}{2}\coth\left(\frac{\tilde{X}_g}{2}\right) + \mathcal{O}(\delta^3),$$

$$J_\gamma = -\mathscr{F}(\bar{\eta}_n)\,\tilde{X}_\gamma + \mathscr{F}'(\bar{\eta}_n)\left(\delta\eta + \frac{\bar{T}N_c'(\bar{T})}{N_c(\bar{T})}g_\eta(\bar{\eta}_n)\,\delta\Theta\right)\frac{\tilde{X}_\gamma}{2}\coth\left(\frac{\tilde{X}_\gamma}{2}\right) + \mathcal{O}(\delta^3),$$

where

$$\tilde{X}_g = \frac{1}{g_\eta(\bar{\eta}_n)}\delta\Phi, \qquad \tilde{X}_\gamma = \delta\Phi - \frac{g_\eta(\bar{\eta}_n)-1}{g_\eta(\bar{\eta}_n)}\left(\delta\eta + \frac{\bar{T}N_c'(\bar{T})}{N_c(\bar{T})}g_\eta(\bar{\eta}_n)\,\delta\Theta\right),$$

and $\mathcal{O}(\delta^3) \equiv \mathcal{O}(\delta\eta_n^3) + \mathcal{O}(\delta\eta_n^2\,\delta\Theta) + \mathcal{O}(\delta\Theta^2\,\delta\eta_n) + \mathcal{O}(\delta\Theta^3) + \mathcal{O}(\delta\Phi\,\delta\eta_n^2) + \mathcal{O}(\delta\Phi\,\delta\eta_n\,\delta\Theta) + \mathcal{O}(\delta\Phi\,\delta\Theta^2)$ denotes the third-order corrections. The second-order expansion of the modified drift scheme is independent of the kind of average used for $\rho_{n,K,L} \approx \rho_n(\bar{\eta}_n, \bar{T}) + \mathcal{O}(\delta\eta_n^2) + \mathcal{O}(\delta\eta_n\,\delta\Theta) + \mathcal{O}(\delta\Theta^2)$, as only its zeroth-order contribution (where all means coincide) is relevant here. Using the inequality [90]

$$1 \le x\coth(x) \le 1 + |x|,$$

we arrive at the error estimates for the modified thermal voltage scheme (neglecting third-order terms)

$$|J_g - J_1| \le \frac{1}{2}\mathscr{F}'(\bar{\eta}_n)\left(|\delta\Phi\,\delta\eta_n| + \frac{\bar{T}N_c'(\bar{T})}{N_c(\bar{T})}g_\eta(\bar{\eta}_n)\,|\delta\Phi\,\delta\Theta|\right) \tag{39}$$

and the modified drift scheme

$$\begin{aligned}|J_\gamma - J_1| \le \frac{1}{2}\mathscr{F}'(\bar{\eta}_n)\Bigg(&|\delta\Phi\,\delta\eta_n| + \frac{\bar{T}N_c'(\bar{T})}{N_c(\bar{T})}g_\eta(\bar{\eta}_n)\,|\delta\Phi\,\delta\Theta| + \\ &+ \frac{g_\eta(\bar{\eta}_n)-1}{g_\eta(\bar{\eta}_n)}\left|\delta\eta_n + \frac{\bar{T}N_c'(\bar{T})}{N_c(\bar{T})}g_\eta(\bar{\eta}_n)\,\delta\Theta\right|^2\Bigg),\end{aligned} \tag{40}$$

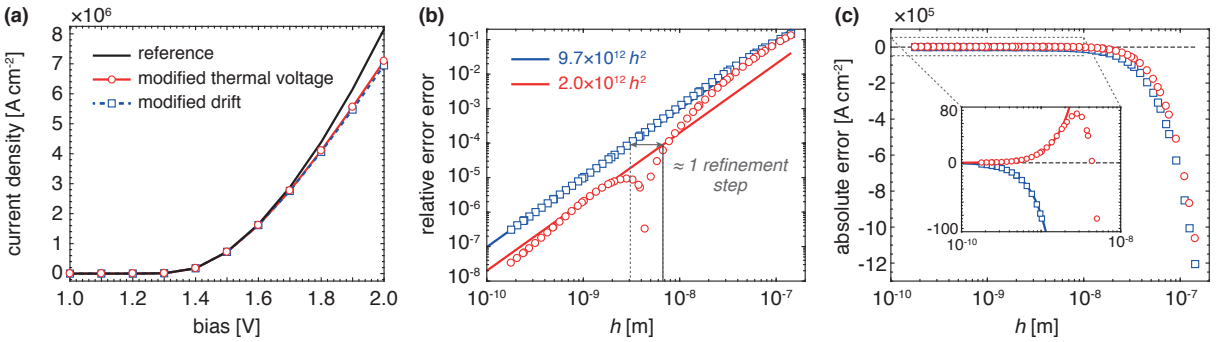

Fig. 6. Convergence of total current density in the p-n diode problem without self-heating effects ($H = 0$, isothermal case). (a) Current-voltage curves obtained by the two schemes (28) and (32) on an equidistant grid with 13 nodes. The reference solution (black line) was obtained on a fine grid (65535 nodes). (b) Convergence of the relative error with respect to the reference solution under mesh refinement at 2 V. The modified drift scheme (blue squares) shows a monotonous, quadratic convergence for decreasing h. The modified thermal voltage scheme (red circles) converges non-monotonously as it intersects with the reference solution at $h \approx 4.5$ nm. (c) Convergence of the absolute error of both schemes.

where $J_1 = \mathscr{F}(\bar{\eta}_n)\left(\delta\eta_n - \delta\Phi + \frac{\bar{T}N_c'(\bar{T})}{N_c(\bar{T})} g_\eta(\bar{\eta}_n)\,\delta\Theta\right)$ is the first-order exact solution of the BVP (38). Both schemes converge to the exact result as their first-order terms agree with J_1. The first two terms of Eq. (40) coincide with Eq. (39). The error bound for the modified drift scheme has an additional second-order contribution that becomes significant in the case of strong degeneration $\bar{\eta} \gg 1$ where $(g_\eta(\bar{\eta}_n) - 1)/g_\eta(\bar{\eta}_n) \to 1$. Therefore, the maximum error of the modified thermal voltage scheme is guaranteed to be smaller than that of the modified drift scheme in the case of degenerate carrier statistics. This analytical result is consistent with the numerical results shown in Figs. 4 and 5 and holds in both the isothermal and the non-isothermal case. For non-degenerate carrier statistics both error estimates (39) and (40) coincide, since both schemes reduce to the non-degenerate scheme (27).

3.6 Numerical simulation of a p-n-diode

We consider a one-dimensional GaAs-based p-n-diode and compare the convergence of the total current density ($\mathbf{j}_{\text{tot}} = \mathbf{j}_n + \mathbf{j}_p$) under mesh refinement using both discretization schemes. The device is composed of a $1\,\mu$m n-doped section with $C = N_D^+ = 2 \times 10^{18}\,\text{cm}^{-3}$ followed by a $1\,\mu$m long p-doped section with $C = -N_A^- = -2 \times 10^{18}\,\text{cm}^{-3}$. We use Fermi–Dirac statistics $\mathscr{F}(\eta) = F_{1/2}(\eta)$ and take Shockley–Read–Hall recombination, spontaneous emission and Auger recombination into account [15, 62]. The material parameters, mobility models (depending on temperature and doping density) and the temperature-dependent heat conductivity model are taken from Ref. [62]. The mobilities and thermal conductivity along the edges are taken as the harmonic average of the respective nodal values. We use Dirichlet boundary conditions on both ends of the diode, modeling ideal Ohmic contacts (charge neutrality at the boundary) and ideal heat sinks with $T_{\text{contact}} = 300\,\text{K}$. The simulations are carried out on equidistant grids with varying number of mesh points N_{nodes} and mesh size $h = 2\,\mu\text{m}/(N_{\text{nodes}} - 1)$. The nonlinear systems are solved using a full Newton iteration method [79].

Figure 6 (a) shows the current-voltage curves obtained by both discretization schemes on a coarse grid (13 nodes, $h \approx 1.7 \times 10^{-7}\,\text{m}$) under isothermal conditions, i. e., without self-heating and Seebeck effect. For the evaluation of the error, we use a reference solution that was computed on a fine grid with 65535 nodes ($h \approx 3.1 \times 10^{-11}\,\text{m}$), where the relative error between both schemes is about 9.6×10^{-9}. At 2 V the computed currents differ significantly from the reference result: The relative error is about $13\,\%$ for the modified thermal voltage scheme and $15\,\%$ for the modified drift scheme. The convergence of the computed current densities to the reference result under mesh refinement

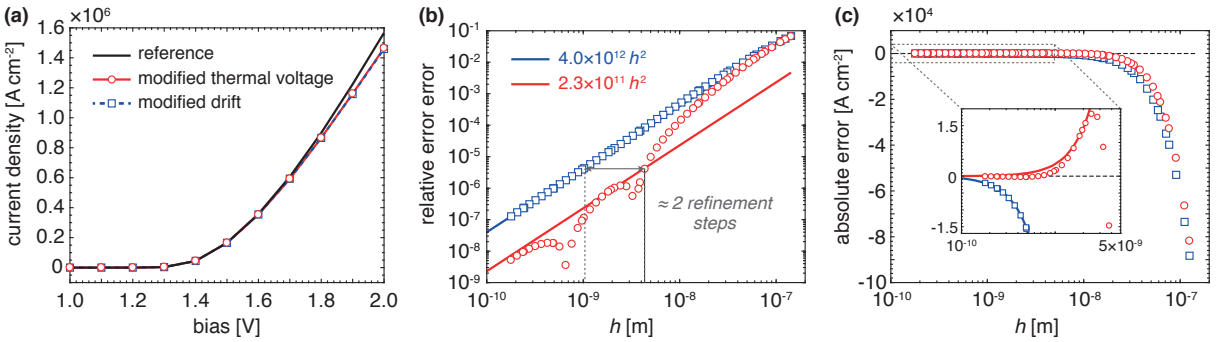

Fig. 7. Convergence of total current density in the p-n diode problem with self-heating effects. (a) Current-voltage curves obtained by the two schemes (28) and (32) on an equidistant grid with 13 nodes. As in Fig. 6, the reference solution (black line) was obtained on a grid with 65535 nodes. Self-heating lowers the mobilities such that the total current density is smaller than in the isothermal case, cf. Fig. 6 (a). (b) Convergence of the relative error with respect to the reference solution under mesh refinement at 2 V. As in the isothermal case, the modified thermal voltage scheme (red circles) shows a non-monotonous convergence behavior that is faster than the quadratic convergence of the modified drift scheme (blue squares). (c) Convergence of the absolute error of both schemes. The modified thermal voltage scheme intersects twice with the reference solution.

is illustrated in Fig. 6 (b). The modified drift scheme (Eq. (32), blue squares) shows a monotonous, quadratic convergence for decreasing h. The modified thermal voltage scheme (Eq. (28), red circles), however, shows a non-monotonous convergence behavior as it intersects with the reference solution at $h \approx 4.5 \times 10^{-9}$ m. On sufficiently fine meshes ($h < 10^{-8}$ m), the error of the modified thermal voltage scheme is almost one order of magnitude smaller than that of the modified drift scheme. Conversely, the modified thermal voltage scheme reaches the same accuracy as the modified drift scheme already on a coarse grid with less than half of the number of nodes. Thus, the modified thermal voltage scheme saves about one refinement step. The convergence of the absolute error is plotted in Fig. 6 (c), where the inset shows the origin of the non-monotonous convergence behavior of the modified thermal voltage scheme.

The numerical results for the non-isothermal case, where self-heating and the Seebeck effect are taken into account, are shown in Fig. 7. The results are qualitatively similar to the isothermal case shown in Fig. 6; quantitatively the advantage of the modified thermal voltage scheme over the modified drift scheme is even greater. On a coarse grid, the total current is underestimated by both schemes with a relative error of about $8\,\%$, see Fig. 7 (a). For sufficiently fine meshes ($h < 0.5 \times 10^{-8}$m), the error of the modified thermal voltage scheme is always more than one order of magnitude smaller than that of the modified drift scheme, see Fig. 7 (b). Hence, the modified thermal voltage scheme reaches the same accuracy already on an about four times coarser mesh (2 refinement steps), which is a substantial advantage for problems involving complex multi-dimensional geometries. Again, we observe a non-monotonous convergence behavior of the modified thermal voltage scheme, see Fig. 7 (b, c).

4 Summary and conclusion

We discussed the non-isothermal drift-diffusion system for the simulation of electro-thermal transport processes in semiconductor devices. It was shown, that the model equations take a remarkably simple form when assuming the Kelvin formula for the Seebeck coefficient. First, the heat generation rate involves exactly the three classically known self-heating effects (Joule heating, recombination heating, Thomson–Peltier effect) without any further transient contributions. Moreover, our modeling approach immediately yields the correct average kinetic energy of the carriers in the recombination heat term,

independently of any scattering parameter. Second, the Kelvin formula enables a simple representation of the electrical current densities in the drift-diffusion form, where the thermal driving force can be entirely absorbed in the (nonlinear) diffusion coefficient via the (generalized) Einstein relation. The Kelvin formula accounts for the degeneration of the electron-hole plasma (Fermi–Dirac statistics) and was shown to be in a good quantitative agreement with experimental data reported for n-GaAs.

We have derived two non-isothermal generalizations of the finite volume Scharfetter–Gummel scheme for the discretization of the current densities, which differ in their treatment of degeneration effects. The first approach is based on an approximation of the discrete generalized Einstein relation and implies a specific modification of the thermal voltage. The second scheme is based on including the degeneration effects into a modification of the electric field, which is similar to the conventional method that is widely used in commercial device simulation software packages [85, 86]. We presented a detailed analysis of both schemes by assessing their accuracy in comparison to the numerically exact solution of the underlying two-point boundary value problem. Moreover, we derived analytical error bounds and investigated important structure preserving properties of the discretizations, including the consistency with the thermodynamic equilibrium, the non-negativity of the discrete dissipation rate (second law of thermodynamics on the discrete level) and their asymptotic behavior in the drift- and diffusion-dominated limits. Finally, we performed a numerical convergence study for a simple example case (one-dimensional p-n-diode). Our results indicate a significantly higher accuracy and faster convergence behavior of the modified thermal voltage scheme in comparison to the modified drift scheme. This result holds under both isothermal and non-isothermal conditions. The higher accuracy — about one order of magnitude for sufficiently fine grids in the present case study — of the modified thermal voltage scheme makes it a favorable discretization method for problems exhibiting stiff solutions (internal layers at p-n junctions, boundary layers at semiconductor-metal interfaces) or devices with a complicated multi-dimensional geometry, where the number of nodes required to reach the asymptotic accuracy regime is extremely large and routinely exceeds the available computational power.

In more general situations, where the Seebeck coefficient deviates from the Kelvin formula (e. g., due to the phonon drag effect), we suggest to combine the two discretization techniques by decomposing the Seebeck coefficient into a Kelvin formula part and an excess contribution: $P_n = P_n^{\text{Kelvin}} + P_n^{\text{exc}}$. The first part can be absorbed in the generalized Einstein relation, which allows for the treatment described in Sec. 3.2.1 and inherits the improved accuracy of the modified thermal voltage scheme. The excess part P_n^{exc} must be averaged along the edge and plays a similar role as the $\rho_{n,K,L}$ term in Sec. 3.2.2 (leading to an additive correction in the argument of the Bernoulli function).

Appendix A Power balance

The balance equation for the total power is derived by integrating the heat transport Eq. (4) over the domain Ω:

$$\int_\Omega \mathrm{d}V\, c_V \partial_t T - \int_\Omega \mathrm{d}V\, \nabla \cdot \kappa \nabla T = \int_\Omega \mathrm{d}V\, H.$$

Using Eqs. (2)–(3) and partial integration, the heat generation rate (20) is rearranged as

$$H = -\nabla \cdot \left((\varphi_n + \Pi_n)\, \mathbf{j}_n + (\varphi_p + \Pi_p)\, \mathbf{j}_p \right) + q\, (\varphi_n + \Pi_n)\, \partial_t n - q\, (\varphi_p + \Pi_p)\, \partial_t p.$$

The internal energy density of the system reads $u\,(n,p,T) = f\,(n,p,T) + Ts\,(n,p,T)$, where the free energy density is given in Eq. (12). Using the defining relations for the quasi-Fermi potentials and the entropy density (10) as well as the Kelvin formula (9), the differential internal energy per carrier is

obtained as as

$$\frac{\partial u}{\partial n} = \frac{\partial f}{\partial n} + T\frac{\partial s}{\partial n} = -q\left(\varphi_n + \Pi_n\right), \qquad \frac{\partial u}{\partial p} = \frac{\partial f}{\partial p} + T\frac{\partial s}{\partial p} = +q\left(\varphi_n + \Pi_n\right),$$

where the Peltier coefficients $\Pi_{n/p}$ are defined via the Kelvin relation $\Pi_{n/p} = TP_{n/p}$. Using the definition of the heat capacity $c_V = \partial_T u$, the (integrated) heat transport equation is rearranged as a balance equation for the internal energy:

$$\begin{aligned}\int_\Omega \mathrm{d}V \left(\frac{\partial u}{\partial T}\frac{\partial T}{\partial t} + \frac{\partial u}{\partial n}\frac{\partial n}{\partial t} + \frac{\partial u}{\partial p}\frac{\partial p}{\partial t}\right) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \mathrm{d}V\, u \\ &= -\oint_{\partial\Omega} \mathrm{d}A\, \mathbf{n}\cdot\left(-\kappa\nabla T + \left(\varphi_n + \Pi_n\right)\mathbf{j}_n + \left(\varphi_p + \Pi_p\right)\mathbf{j}_p\right).\end{aligned} \tag{41}$$

In the following, we will recast the surface integral on the right hand side into an expression for the injected electrical power and the dissipated heat.

The local continuity equation for the internal energy (first law of thermodynamics, cf. Eq. (41)) reads

$$\partial_t u + \nabla\cdot\mathbf{j}_u = 0, \qquad \mathbf{j}_u = -\kappa\nabla T + \left(\varphi_n + \Pi_n\right)\mathbf{j}_n + \left(\varphi_p + \Pi_p\right)\mathbf{j}_p, \tag{42}$$

where $\mathbf{j}_u$ is the internal energy flux density. The corresponding continuity equation for the entropy density is obtained from Gibb's fundamental thermodynamic relation $\mathrm{d}u = T\mathrm{d}s + \mu_c\mathrm{d}n - \mu_v\mathrm{d}p$. By substituting Eqs. (2)–(3) and (42), one arrives at

$$\partial_t s = \frac{1}{T}\partial_t u - \frac{\mu_c}{T}\partial_t n + \frac{\mu_v}{T}\partial_t p = -\nabla\cdot\mathbf{j}_s + \dot{s}_{\text{tot}}, \qquad \mathbf{j}_s = \frac{1}{T}\left(\mathbf{j}_u - \varphi_n\mathbf{j}_n - \varphi_p\mathbf{j}_p\right), \tag{43}$$

where $\mathbf{j}_s$ is the the entropy flux density and the entropy production rate reads (cf. Sec. 3.3.5)

$$\dot{s}_{\text{tot}} = \frac{1}{T}\left(\mu_c - \mu_v\right)R + \frac{\kappa}{T^2}\left\|\nabla T\right\|^2 + \frac{1}{T}H_J. \tag{44}$$

Here, $H_J = \mathbf{j}_n\cdot\left(\nabla\varphi_n + P_n\nabla T\right) + \mathbf{j}_p\cdot\left(\nabla\varphi_p + P_p\nabla T\right)$ is the Joule heat term. The entropy flux density is closely connected with the heat flux density

$$\mathbf{j}_Q = T\mathbf{j}_s = \mathbf{j}_u - \varphi_n\mathbf{j}_n - \varphi_p\mathbf{j}_p = -\kappa\nabla T + \Pi_n\mathbf{j}_n + \Pi_p\mathbf{j}_p. \tag{45}$$

We return to the power balance Eq. (41) and rewrite the surface integral on the right hand side as

$$\oint_{\partial\Omega}\mathrm{d}A\,\mathbf{n}\cdot\left(-\kappa\nabla T + \left(\varphi_n + \Pi_n\right)\mathbf{j}_n + \left(\varphi_p + \Pi_p\right)\mathbf{j}_p\right) = \oint_{\partial\Omega}\mathrm{d}A\,\mathbf{n}\cdot\mathbf{j}_Q + \oint_{\partial\Omega}\mathrm{d}A\,\mathbf{n}\cdot\left(\varphi_n\mathbf{j}_n + \varphi_p\mathbf{j}_p\right).$$

The first term is the heat flux density leaving the device. The second term is evaluated for a device with mixed boundary conditions $\partial\Omega = \Gamma_D \cup \Gamma_N$, where the segments $\Gamma_D = \bigcup_i \Gamma_{D,i}$ are the electrical contacts with (ideal) Ohmic boundary conditions $\varphi_n = \varphi_p = \varphi_{D,i} = \mathrm{const.}$ on $\Gamma_{D,i}$ (Dirichlet conditions). The remaining facets Γ_N are semiconductor-insulator interfaces or artificial boundaries with "no-flux" boundary conditions $\mathbf{n}\cdot\mathbf{j}_{n/p} = 0$. One obtains

$$\oint_{\partial\Omega}\mathrm{d}A\,\mathbf{n}\cdot\left(\varphi_n\mathbf{j}_n + \varphi_p\mathbf{j}_p\right) = \sum_i \varphi_{D,i}\int_{\Gamma_{D,i}}\mathrm{d}A\,\mathbf{n}\cdot\left(\mathbf{j}_n + \mathbf{j}_p\right) = \sum_i \varphi_{D,i}I_i,$$

where $I_i = \int_{\Gamma_{D,i}}\mathrm{d}A\,\mathbf{n}\cdot\left(\mathbf{j}_n + \mathbf{j}_p\right)$ is the electrical current flux across the i-th electrical contact. For a two-terminal device with total current $I = I_2 = -I_1$ (Kirchhoff's current law) and applied voltage $U = \varphi_{D,1} - \varphi_{D,2}$, this is (cf. Ref. [12])

$$\oint_{\partial\Omega}\mathrm{d}A\,\mathbf{n}\cdot\left(\varphi_n\mathbf{j}_n + \varphi_p\mathbf{j}_p\right) = I_1\varphi_{D,1} + I_2\varphi_{D,2} = -UI.$$

At stationary conditions, we finally obtain the power balance equation as

$$UI = \oint_{\partial\Omega} \mathrm{d}A\, \mathbf{n} \cdot \mathbf{j}_Q. \tag{46a}$$

In conclusion, the injected electrical power UI is equal to the heat flux density leaving the device. With the divergence of the heat flux density $\nabla \cdot \mathbf{j}_Q = H + \nabla \cdot (\Pi_n \mathbf{j}_n + \Pi_p \mathbf{j}_p)$, this can also be written as

$$UI = \int_\Omega \mathrm{d}V H + \int_{\Gamma_D} \mathrm{d}A\, \mathbf{n} \cdot (\Pi_n \mathbf{j}_n + \Pi_p \mathbf{j}_p)\,, \tag{46b}$$

where the first term is the total heat generated on the full domain and the second term describes the “Peltier power”, that can be either positive or negative, depending on the direction of the current flow. In the case of optoelectronic devices (with open optical cavities), the power balance must be supplemented by additional terms describing the emitted radiation power [29]. The power balance equation (46) is consistent with the results previously reported in the literature, i.e., the result persists when assuming the Kelvin formula for the Seebeck coefficient.

References

[1] J. Lutz, H. Schlangenotto, U. Scheuermann, R. D. Donker, Semiconductor Power Devices – Physics, Characteristics, Reliability, Springer, Berlin, Heidelberg, 2011. doi:10.1007/978-3-642-11125-9.

[2] H.-J. Schulze, F.-J. Niedernostheide, F. Pfirsch, R. Baburske, Limiting factors of the safe operating area for power devices, IEEE Trans. Electron Devices 60 (2) (2012) 551–562. doi:10.1109/ted.2012.2225148.

[3] A. Fischer, T. Koprucki, K. Gärtner, M. L. Tietze, J. Brückner, B. Lüssem, K. Leo, A. Glitzky, R. Scholz, Feel the heat: Nonlinear electrothermal feedback in organic LEDs, Adv. Funct. Mater. 24 (22) (2014) 3367–3374. doi:10.1002/adfm.201303066.

[4] A. Fischer, M. Pfalz, K. Vandewal, S. Lenk, M. Liero, A. Glitzky, S. Reineke, Full electrothermal OLED model including nonlinear self-heating effects, Phys. Rev. Appl 10 (1) (2018) 014023. doi:10.1103/physrevapplied.10.014023.

[5] M. Osiński, W. Nakwaski, Thermal effects in vertical-cavity surface-emitting lasers, Int. J. High Speed Electron. Syst. 5 (04) (1994) 667–730. doi:10.1142/s0129156494000267.

[6] J. Piprek, J. K. White, A. J. SpringThorpe, What limits the maximum output power of long-wavelength AlGaInAs/InP laser diodes?, IEEE J. Quantum Elect. 38 (9) (2002) 1253–1259. doi:10.1109/jqe.2002.802441.

[7] M. Streiff, W. Fichtner, A. Witzig, Vertical-Cavity Surface-Emitting Lasers: Single-Mode Control and Self-Heating Effects, Springer, New York, 2005, Ch. 8, pp. 217–247. doi:10.1007/0-387-27256-9_8.

[8] H. Wenzel, P. Crump, A. Pietrzak, X. Wang, G. Erbert, G. Tränkle, Theoretical and experimental investigations of the limits to the maximum output power of laser diodes, New J. Phys. 12 (8) (2010) 085007. doi:10.1088/1367-2630/12/8/085007.

[9] G. K. Wachutka, Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling, IEEE Trans. Comput.-Aided Design Integr. Circuits Syst. 9 (11) (1990) 1141–1149. doi:10.1109/43.62751.

[10] U. Lindefelt, Heat generation in semiconductor devices, J. Appl. Phys. 75 (2) (1994) 942–957. doi:10.1063/1.356450.

[11] H. Brand, S. Selberherr, Two-dimensional simulation of thermal runaway in a nonplanar GTO-thyristor, IEEE Trans. Electron Devices 42 (12) (1995) 2137–2146. doi:10.1109/16.477772.

[12] J. E. Parrott, Thermodynamic theory of transport processes in semiconductors, IEEE Trans. Electron Devices 43 (5) (1996) 809–826. doi:10.1109/16.491259.

[13] G. Albinus, H. Gajewski, R. Hünlich, Thermodynamic design of energy models of semiconductor devices, Nonlinearity 15 (2) (2002) 367–383. doi:10.1088/0951-7715/15/2/307.

[14] U. Bandelow, H. Gajewski, R. Hünlich, Fabry–Perot lasers: Thermodynamics-based modeling, Springer, New York, 2005, Ch. 3, pp. 63–85. doi:10.1007/0-387-27256-9_3.

[15] S. Selberherr, Analysis and Simulation of Semiconductor Devices, Springer, Vienna, 1984. doi:10.1007/978-3-7091-8752-4.

[16] P. A. Markowich, The stationary Semiconductor device equations, Series in Computational Microelectronics, Springer, Vienna, 1986. doi:10.1007/978-3-7091-3678-2.

[17] H. J. Goldsmid, Introduction to Thermoelectricity, Vol. 121 of Springer Series in Materials Science, Springer, Berlin, Heidelberg, 2010. doi:10.1007/978-3-642-00716-3.

[18] C. Goupil, Ouerdane, K. Zabrocki, W. Seifert, N. Hinsche, E. Müller, Thermodynamics and Thermoelectricity, Wiley, Weinheim, 2011, Ch. 1, pp. 1–73. doi:10.1002/9783527338405.ch1.

[19] L. Onsager, Reciprocal relations in irreversible processes. I., Phys. Rev. 37 (4) (1931) 405. doi:10.1103/PhysRev.37.405.

[20] R. Kubo, M. Yokota, S. Nakajima, Statistical-mechanical theory of irreversible processes. II. Response to thermal disturbance, J. Phys. Soc. Jpn. 12 (11) (1957) 1203–1211. doi:10.1143/jpsj.12.1203.

[21] M. Cutler, N. F. Mott, Observation of Anderson localization in an electron gas, Phys. Rev. 181 (3) (1969) 1336–1340. doi:10.1103/physrev.181.1336.

[22] H. Fritzsche, A general expression for the thermoelectric power, Solid State Commun. 9 (21) (1971) 1813–1815. doi:10.1016/0038-1098(71)90096-2.

[23] P. M. Chaikin, G. Beni, Thermopower in the correlated hopping regime, Phys. Rev. B 13 (2) (1976) 647–651. doi:10.1103/physrevb.13.647.

[24] B. S. Shastry, Thermopower in correlated systems, in: V. Zlatc, A. Hewson (Eds.), New Materials for Thermoelectric Applications: Theory and Experiment, NATO Science for Peace and Security Series B: Physics and Biophysics, Springer, Dordrecht, 2013, pp. 25–29. doi:10.1007/978-94-007-4984-9_2.

[25] J. C. Freeman, Self-heating in semiconductors: A comparative study, Solid State Electron. 95 (2014) 8–14. doi:10.1016/j.sse.2014.02.005.

[26] K. M. Van Vliet, A. H. Marshak, Conduction current and generalized Einstein relations for degenerate semiconductors and metals, Phys. Status Solidi B 78 (2) (1976) 501–517. doi:10.1002/pssb.2220780209.

[27] A. H. Marshak, C. Van Vliet, Electrical current and carrier density in degenerate materials with nonuniform band structure, Proc. IEEE 72 (2) (1984) 148–164. doi:10.1109/proc.1984.12836.

[28] M. Lundstrom, Fundamentals of carrier transport, 2nd Edition, Cambridge University Press, Cambridge, 2000. doi:10.1017/CBO9780511618611.

[29] H. Wenzel, A. Zeghuzi, High-power lasers, in: J. Piprek (Ed.), Handbook of Optoelectronic Device Modeling and Simulation: Lasers, Modulators, Photodetectors, Solar Cells, and Numerical Methods, Vol. 2, CRC Press, Taylor & Francis Group, Boca Raton, 2017, Ch. 50, pp. 733–771. doi:10.4324/9781315152318-2.

[30] G. J. Snyder, E. S. Toberer, Complex thermoelectric materials, Nat. Mater. 7 (2008) 105–114. doi:10.1142/9789814287005_0006.

[31] N. S. Bennett, Thermoelectric performance in n-type bulk silicon: The influence of dopant concentration and dopant species, Phys. Status Solidi A 214 (7) (2017) 1700307. doi:10.1002/pssa.201700307.

[32] M. R. Peterson, B. S. Shastry, Kelvin formula for thermopower, Phys. Rev. B 82 (19) (2010) 195105. doi:10.1103/physrevb.82.195105.

[33] B. S. Shastry, Electrothermal transport coefficients at finite frequencies, Rep. Prog. Phys. 72 (1) (2008) 016501. doi:10.1088/0034-4885/72/1/016501.

[34] T. Silk, I. Terasaki, T. Fujii, A. Schofield, Out-of-plane thermopower of strongly correlated layered systems: An application to $Bi_2(Sr, La)_2CaCu_2O_{8+\delta}$, Phys. Rev. B 79 (13) (2009) 134527. doi:10.1103/PhysRevB.79.134527.

[35] A. Garg, B. S. Shastry, K. B. Dave, P. Phillips, Thermopower and quantum criticality in a strongly interacting system: Parallels with the cuprates, New J. Phys. 13 (8) (2011) 083032. doi:10.1088/1367-2630/13/8/083032.

[36] L.-F. Arsenault, B. S. Shastry, P. Sémon, A.-M. Tremblay, Entropy, frustration, and large thermopower of doped Mott insulators on the fcc lattice, Phys. Rev. B 87 (3) (2013) 035126. doi:10.1103/physrevb.87.035126.

[37] X. Deng, J. Mravlje, R. Žitko, M. Ferrero, G. Kotliar, A. Georges, How bad metals turn good: Spectroscopic signatures of resilient quasiparticles, Phys. Rev. Lett. 110 (8) (2013) 086401. doi:10.1103/physrevlett.110.086401.

[38] V. Zlatić, G. R. Boyd, J. K. Freericks, Universal thermopower of bad metals, Phys. Rev. B 89 (15) (2014) 155101. doi:10.1103/PhysRevB.89.155101.

[39] J. Kokalj, R. H. McKenzie, Enhancement of the thermoelectric power by electronic correlations in bad metals: A study of the Kelvin formula, Phys. Rev. B 91 (12) (2015) 125143. doi:10.1103/physrevb.91.125143.

[40] J. Hejtmánek, Z. Jirák, J. Šebek, Spin-entropy contribution to thermopower in the $[Ca_2CoO_{3-t}]_{0.62}(CoO_2)$ misfits, Phys. Rev. B 92 (12) (2015) 125106. doi:10.1103/PhysRevB.92.125106.

[41] I. Terasaki, Research update: Oxide thermoelectrics: Beyond the conventional design rules, APL Mater. 4 (10) (2016) 104501. doi:10.1063/1.4954227.

[42] J. Mravlje, A. Georges, Thermopower and entropy: Lessons from Sr_2RuO_4, Phys. Rev. Lett. 117 (3) (2016) 036401. doi:10.1103/physrevlett.117.036401.

[43] D. L. Scharfetter, H. K. Gummel, Large-signal analysis of a silicon Read diode oscillator, IEEE Trans. Electron Devices 16 (1) (1969) 64–77. doi:10.1109/t-ed.1969.16566.

[44] T.-W. Tang, Extension of the Scharfetter–Gummel algorithm to the energy balance equation, IEEE Trans. Electron Devices 31 (12) (1984) 1912–1914. doi:10.1109/t-ed.1984.21813.

[45] C. C. McAndrew, K. Singhal, E. L. Heasell, A consistent nonisothermal extension of the Scharfetter–Gummel stable difference approximation, IEEE Electr. Device L. 6 (9) (1985) 446–447. doi:10.1109/edl.1985.26187.

[46] M. Rudan, F. Odeh, Multi-dimensional discretization scheme for the hydrodynamic model of semiconductor devices, Compel. 5 (3) (1986) 149–183. doi:10.1108/eb010024.

[47] A. Forghieri, R. Guerrieri, P. Ciampolini, A. Gnudi, M. Rudan, G. Baccarani, A new discretization strategy of the semiconductor equations comprising momentum and energy balance, IEEE T. Comput. Aid. D. 7 (2) (1988) 231–242. doi:10.1109/43.3153.

[48] D. Chen, E. C. Kan, U. Ravaioli, Z. Yu, R. W. Dutton, A self-consistent discretization scheme for current and energy transport equations, in: IV Int. Conf. on Simulation of Semiconductor Devices and Processes (SISDEP), Zurich, 1991, pp. 235–240.

[49] K. Souissi, F. Odeh, A. Gnudi, A note on current discretization in the hydromodel, Compel. 10 (4) (1991) 475–485. doi:10.1108/eb051722.

[50] J. H. M. Ten Thije Boonkkamp, W. H. A. Schilders, An exponential fitting scheme for the electrothermal device equations specifically for the simulation of avalanche generation, Compel. 12 (2) (1993) 95–111. doi:10.1108/eb010116.

[51] A. W. Smith, A. Rohatgi, Non-isothermal extension of the Scharfetter–Gummel technique for hot carrier transport in heterostructure simulations, IEEE T. Comput. Aid. D. 12 (10) (1993) 1515–1523. doi:10.1109/43.256926.

[52] Z. Yu, R. W. Dutton, SEDAN-III – A generalized electronic material device analysis program, Tech. rep., Stanford Electronics Lab, Stanford Univ. (1985).

[53] H. Gajewski, Analysis und Numerik von Ladungstransport in Halbleitern, Mitt. Ges. Angew. Math. Mech. 16 (1) (1993) 35–57.

[54] M. Bessemoulin-Chatard, A finite volume scheme for convection-diffusion equations with nonlinear diffusion derived from the Scharfetter–Gummel scheme, Numer. Math. 121 (4) (2012) 637–670. doi:10.1007/s00211-012-0448-x.

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

[56] 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, Opt. Quantum. Electron. 47 (6) (2015) 1327–1332. doi:10.1007/s11082-014-0050-9.

[57] J. Fuhrmann, Comparison and numerical treatment of generalised Nernst–Planck models, Comput. Phys. Commun. 196 (2015) 166–178. doi:10.1016/j.cpc.2015.06.004.

[58] 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 (8) (2008) 085207. doi:10.1103/PhysRevB.78.085207.

[59] 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.

[60] M. C. J. M. Vissenberg, M. Matters, Theory of the field-effect mobility in amorphous organic transistors, Phys. Rev. B 57 (1998) 12964–12967. arXiv:cond-mat/9802133, doi:10.1103/PhysRevB.57.12964.

[61] K. Seki, K. Marumoto, M. Tachiya, Bulk recombination in organic bulk heterojunction solar cells under continuous and pulsed light irradiation, Appl. Phys. Express 6 (5) (2013) 051603. doi:10.7567/APEX.6.051603.

[62] V. Palankovski, R. Quay, Analysis and Simulation of Heterostructure Devices, Series in Computational Microelectronics, Springer, Vienna, 2004. doi:10.1007/978-3-7091-0560-3.

[63] M. Kantner, Modeling and simulation of electrically driven quantum dot based single-photon sources: From classical device physics to open quantum systems, PhD thesis, Technical University Berlin, Berlin (2018). doi:10.14279/depositonce-7516.

[64] G. Czycholl, Theoretical solid-state physics., Springer, Berlin, Heidelberg, 2008. doi:10.1007/978-3-540-74790-1.

[65] R. Carlson, S. J. Silverman, H. Ehrenreich, Nernst effect in n-type GaAs, J. Phys. Chem. Solids 23 (1962) 422–424. doi:10.1016/0022-3697(62)90112-9.

[66] A. Amith, I. Kudman, E. F. Steigmeier, Electron and phonon scattering in GaAs at high temperatures, Phys. Rev. 138 (1965) 1270–1276. doi:10.1103/PhysRev.138.A1270.

[67] J. T. Edmond, R. F. Broom, F. A. Cunnel, Report on meeting on semiconductors at Rugby, Physical Society, London, 1956, p. 109.

[68] S. K. Sutadhar, D. Chattopadhyay, Thermoelectric power of n-GaAs, J. Phys. C: Solid State Phys. 12 (9) (1979) 1693–1697. doi:10.1088/0022-3719/12/9/011.

[69] G. Homm, P. J. Klar, J. Teubert, W. Heimbrodt, Seebeck coefficients of n-type (Ga, In)(N, As),(B, Ga, In) As, and

GaAs, Appl. Phys. Lett. 93 (4) (2008) 042107. doi:10.1063/1.2959079.

[70] O. V. Emel'yanenko, V. A. Skripkin, U. A. Popova, Thermoelectric power and phonon drag in heavily doped III–V crystals, Sov. Phys.-Semicond. 7 (1973) 667–674.

[71] K. P. O'Donnell, X. Chen, Temperature dependence of semiconductor band gaps, Appl. Phys. Lett. 58 (1991) 2924–2926. doi:10.1063/1.104723.

[72] I. Vurgaftman, J. R. Meyer, L. R. Ram-Mohan, Band parameters for III–V compound semiconductors and their alloys, J. Appl. Phys. 89 (11) (2001) 5815–5875. doi:10.1063/1.1368156.

[73] K. Kells, S. Müller, G. Wachutka, W. Fichtner, Simulation of self-heating effects in a power pin diode, in: S. Selberherr, H. Stippel, E. Strasser (Eds.), Simulation of Semiconductor Devices and Processes, Vol. 5, Springer, Vienna, 1993, pp. 41–44. doi:10.1007/978-3-7091-6657-4_9.

[74] P. B. M. Wolbert, G. K. M. Wachutka, B. H. Krabbenborg, T. J. Mouthaan, Nonisothermal device simulation using the 2D numerical process/device simulator TRENDY and application to SOI-devices, IEEE T. Comput. Aid. D. 13 (3) (1994) 293–302. doi:10.1109/43.265671.

[75] P. T. Landsberg, On the diffusion theory of rectification, Proc. R. Soc. Lond. A 213 (1113) (1952) 226–237. doi:10.1098/rspa.1952.0122.

[76] K. Shore, M. Adams, The effects of carrier degeneracy on transport properties of the double heterostructure injection laser, Appl. Phys. 9 (2) (1976) 161–164. doi:10.1007/bf00903953.

[77] Y. Roichman, N. Tessler, Generalized Einstein relation for disordered semiconductors — Implications for device performance, Appl. Phys. Lett. 80 (11) (2002) 1948–1950. doi:10.1063/1.1461419.

[78] F. Brezzi, L. D. Marini, P. Pietra, Numerical simulation of semiconductor devices, Comput. Methods Appl. Mech. Eng. 75 (1-3) (1989) 493–514. doi:10.1016/0045-7825(89)90044-3.

[79] P. Farrell, N. Rotundo, D. H. Doan, M. Kantner, J. Fuhrmann, T. Koprucki, Drift-diffusion models, in: J. Piprek (Ed.), Handbook of Optoelectronic Device Modeling and Simulation: Lasers, Modulators, Photodetectors, Solar Cells, and Numerical Methods, Vol. 2, CRC Press, Taylor & Francis Group, Boca Raton, 2017, Ch. 27, pp. 15–58. doi:10.4324/9781315152318-25.

[80] H. Si, K. Gärtner, J. Fuhrmann, Boundary conforming Delaunay mesh generation, Comput. Math. Math. Phys. 50 (1) (2010) 38–53. doi:10.1134/S0965542510010069.

[81] R. Eymard, T. Gallouët, R. Herbin, Finite volume methods, in: P. Ciarlet, J. Lions (Eds.), Handbook of Numerical Analysis, Vol. 7 of Handbook of Numerical Analysis, Elsevier, 2000, pp. 713–1018. doi:10.1016/S1570-8659(00)07005-8.

[82] A. Bradji, R. Herbin, Discretization of coupled heat and electrical diffusion problems by finite-element and finite-volume methods, IMA J. Numer. Anal. 28 (3) (2008) 469–495. doi:10.1093/imanum/drm030.

[83] C. Chainais-Hillairet, Discrete duality finite volume schemes for two-dimensional drift-diffusion and energy-transport models, Int. J. Numer. Meth. Fl. 59 (2009) 239–257. doi:10.1002/fld.1393.

[84] J. Fuhrmann, A. Glitzky, M. Liero, Hybrid finite-volume/finite-element schemes for $p(x)$-Laplace thermistor models, in: C. Cancès, P. Omnes (Eds.), Finite Volumes for Complex Applications VIII – Hyperbolic, Elliptic and Parabolic Problems, Springer, Berlin, Heidelberg, 2017, pp. 397–405. doi:10.1007/978-3-319-57394-6_42.

[85] Synopsys, Inc., Sentaurus Device UserGuide, Mountain View, CA (Dec. 2010).

[86] Silvaco International, Atlas User's Manual, Santa Clara, CA (Aug. 2016).

[87] H. K. Versteeg, W. Malalasekera, An Introduction to Computational Fluid Dynamics: The Finite Volume Method, 2nd Edition, Prentice Hall, Harlow, 2007.

[88] M. Bessemoulin-Chatard, C. Chainais-Hillairet, Exponential decay of a finite volume scheme to the thermal equilibrium for drift–diffusion systems, J. Numer. Math. 25 (3) (2017) 147–168. doi:10.1515/jnma-2016-0007.

[89] A. Kato, M. Katada, T. Kamiya, T. Ito, T. Hattori, A rapid, stable decoupled algorithm for solving semiconductor hydrodynamic equations, IEEE T. Comput. Aid. D. 13 (11) (1994) 1425–1428. doi:10.1109/43.329272.

[90] P. Farrell, T. Koprucki, J. Fuhrmann, Computational and analytical comparison of flux discretizations for the semiconductor device equations beyond Boltzmann statistics, J. Comput. Phys. 346 (2017) 497–513. doi:10.1016/j.jcp.2017.06.023.

[91] M. Patriarca, P. Farrell, J. Fuhrmann, T. Koprucki, Highly accurate quadrature-based Scharfetter–Gummel schemes for charge transport in degenerate semiconductors, Comput. Phys. Commun. 235 (2019) 40–49. doi:10.1016/j.cpc.2018.10.004.

[92] H. B. Keller, Numerical Solution of Two Point Boundary Value Problems, SIAM, Philadelphia, Pennsylvania, 1976.

[93] R. P. Brent, An algorithm with guaranteed convergence for finding a zero of a function, Comput. J. 14 (4) (1971) 422–425. doi:10.1093/comjnl/14.4.422.

[94] A. J. MacLeod, Algorithm 779: Fermi–Dirac functions of order -1/2, 1/2, 3/2, 5/2, ACM T. Math. Software 24 (1) (1998) 1–12. doi:10.1145/285861.285862.