Stabilization of the Drift-Diffusion Model for Arbitrary Carrier Statistics

Max Renner
Chair of Electromagnetic Theory
RWTH Aachen University
Aachen, Germany
0000-0002-5412-4936

Tobias Linn
Chair of Electromagnetic Theory
RWTH Aachen University
Aachen, Germany
0000-0002-7681-7725

Christoph Jungemann
Chair of Electromagnetic Theory
RWTH Aachen University
Aachen, Germany
0000-0002-3423-4046

***Abstract*—Modern semiconductor simulation is still mostly based on the drift-diffusion model which is solved numerically using the finite volume method together with the Scharfetter-Gummel stabilization. Unfortunately, the Scharfetter-Gummel stabilization is only applicable if the charge carriers follow Maxwell-Boltzmann statistics. In this paper, we present a stabilization scheme generalizing the Scharfetter-Gummel scheme to arbitrary carrier statistics using no additional approximations. So far, no such scheme exists which is accurate and robust enough for a full-scale device simulation. In our method we solve an integral equation for the edge current using Tanh-Sinh-quadrature. This approach works for general carrier statistics, is easily implemented and can be effectively parallelized. Practical use of this method is demonstrated with 2D transient and small-signal simulation results for an exemplary MOSFET at 4K using Fermi-Dirac statistics. Parallelizing over 64 CPU cores, the edge current calculation takes up less than 6% of the total runtime.**

***Index Terms*—Drift-diffusion model, generalized Scharfetter-Gummel, Fermi-Dirac statistics, cryogenic electronics**

I. Introduction

The numerical robustness of the drift-diffusion (DD) model relies on the finite volume method in conjunction with the Scharfetter-Gummel stabilization scheme [1] which provides an analytical solution for the current density along an edge. Unfortunately, this expression is only valid for Maxwell-Boltzmann statistics. For degenerate semiconductors, where Fermi-Dirac statistics has to be used, no similar simple and reliable stabilization scheme exists. This is, for example, the case for cryogenic temperatures or high doping concentrations [2]. In order to obtain a stable simulation for these cases as well, a generalized Scharfetter-Gummel scheme is necessary.

In the derivation of the generalized scheme an implicit integral equation for the current density along an edge arises. In contrast to Maxwell-Boltzmann statistics, for Fermi-Dirac statistics it can not be solved analytically. Instead, the current has to be calculated by approximation or numerical means. The approximation-based schemes [2] usually preserve thermodynamic consistency (i.e. the edge current is exactly zero in thermodynamic equilibrium) but they suffer from reduced accuracy far from equilibrium and require very fine grids. A numerical scheme based on quadrature rules has been proposed in [2, 3]. While the approach is very promising, the specific implementation suggested delivers inaccurate results in a number of cases, in particular when using non-adaptive quadrature rules.

In this paper we build on the work in [2, 3] and improve the approach to be more numerically robust and to work for arbitrary (that is, continuous and strictly monotonically increasing) carrier statistics, using no further assumptions than the Scharfetter-Gummel scheme while preserving its excellent stability properties. Our approach is based on Taylor series expansions around critical points and Tanh-Sinh quadrature for the remaining cases.

II. Problem Description

In the DD model with arbitrary carrier statistics the electron density n is given by

$$n = N_\mathrm{C}\mathcal{F}\left(\frac{e\varphi - e\Phi_n - E_\mathrm{C}}{k_\mathrm{B}T}\right) =: N_\mathrm{C}\mathcal{F}\left(\eta_n\right) \tag{1}$$

where φ is the electrostatic potential, Φ_n is the electron quasi-Fermi-potential, η_n is the chemical potential, e is the elementary charge, k_B is the Boltzmann constant and T is the temperature. The effective density of states in the conduction band N_C and the conduction band edge E_C are material-dependent parameters. The function $\mathcal{F}$ depends on the band structure and carrier statistics: for parabolic bands and Fermi-Dirac statistics $\mathcal{F}(\eta) = \mathcal{F}_{1/2}(\eta)$, the Fermi-Dirac integral of order ½, which simplifies to $\mathcal{F}(\eta) = \exp(\eta)$ with the Maxwell-Boltzmann approximation.

The electron current density $\boldsymbol{J}_n$ is given as

$$\boldsymbol{J}_n = -en\mu_n\boldsymbol{\nabla}\Phi_n \tag{2}$$

with the electron mobility μ_n. This equation is usually discretized with the Scharfetter-Gummel stabilization. It provides an explicit expression for the edge current density under the assumption that the current density, electric field and mobility are constant along the edge and using Maxwell-Boltzmann statistics ($\mathcal{F}(\eta) = \exp(\eta)$). In contrast, for general $\mathcal{F}$ these assumptions yield an implicit integral equation [3]. For an edge between vertices L and R it reads

$$\int_{\eta_\mathrm{L}}^{\eta_\mathrm{R}} \frac{1}{\Delta\varphi - \frac{j}{\mathcal{F}(\eta)}}\mathrm{d}\eta = 1 \tag{3}$$

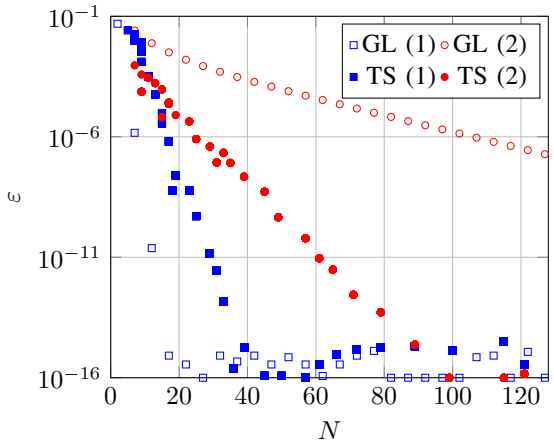

Fig. 1. Relative error in j for Gauss-Legendre (GL) and Tanh-Sinh (TS) quadrature depending on number of sample points N for two parameter sets.

where $\Delta\varphi = \frac{\varphi_\mathrm{R}-\varphi_\mathrm{L}}{V_\mathrm{T}}$ is the normalized potential drop, in which V_T is the thermal voltage, and $j = -\frac{J_n \Delta x}{e\mu_n N_\mathrm{C} V_\mathrm{T}}$ is the normalized edge current density. This equation relates the unknown edge current density J_n to the three input parameters $\varphi_\mathrm{R} - \varphi_\mathrm{L}$, η_L and η_R where the latter two can be calculated from the electron densities at vertices L and R.

Two factors make the reliable solution of this implicit equation difficult. First, we need to obtain a solution for every possible combination of the three input variables. In some scenarios, like cryogenic temperatures, these inputs may easily take extreme values of $\eta < -600$ where, using double precision, arithmetic underflow occurs for typical functions $\mathcal{F}$.

Second, if $\frac{j}{\Delta\varphi} > 0$, the integrand has a singularity at $\eta_0 = \mathcal{F}^{-1}(\frac{j}{\Delta\varphi})$. Even though in [4] it was shown that this pole can not be located inside the integration bounds, it can nevertheless move arbitrarily close. This prevents the straightforward use of Gaussian quadrature which becomes inaccurate for near-singular integrands. This is the case for the implementation from [2, 3] where non-adaptive quadrature rules were used without proper handling of the singularity. In consequence, the robustness of the scheme is deteriorated, making it infeasible for a full-scale device simulation.

III. Numerical Solution

First, we note that $j(\Delta\varphi, \eta_\mathrm{L}, \eta_\mathrm{R}) = -j(-\Delta\varphi, \eta_\mathrm{R}, \eta_\mathrm{L})$, thus we can limit the discussion to $\Delta\varphi > 0$.

A. Taylor series expansions

Three cases can be identified in which Eq. (3) has an analytical solution: $\Delta\varphi = 0$, $\Delta\eta = \eta_\mathrm{R} - \eta_\mathrm{L} = 0$ and $\Delta\varphi = \Delta\eta$. Around these as well as the combination with $\Delta\varphi = \Delta\eta = 0$ Taylor series expansions are used to calculate the edge current. This is necessary to ensure numerical stability (for $\Delta\eta \to 0$ the integral gets infinitesimally small) and thermodynamic consistency (for $\Delta\varphi = \Delta\eta$, the equilibrium case, j has to be exactly zero) [5]. Additionally, the runtime is shortened as the solution with a Taylor series is obtained non-iteratively.

B. Tanh-Sinh quadrature

For all remaining cases Tanh-Sinh (TS) quadrature [6, 7] is used, utilizing the transformation

$$\eta(u) = \frac{\eta_\mathrm{L} + \eta_\mathrm{R}}{2} + \frac{\Delta\eta}{2}\tanh(\sinh(u)) \tag{4}$$

which ensures that the integral boundaries are mapped onto $\pm\infty$ such that the quadrature is insensitive to the possible singularities near the endpoints. Afterwards, the trapezoidal rule with an adaptive number of sample points is optimal [7] to approximate the transformed integral.

Applying this technique to the integral in Eq. (3) results in

$$r(j) = \underbrace{\int_{\eta_\mathrm{L}}^{\eta_\mathrm{R}} \frac{\mathcal{F}(\eta)}{\mathcal{F}(\eta)\Delta\varphi - j}\,\mathrm{d}\eta}\, -1 \stackrel{!}{=} 0 \tag{5}$$

$$\approx \frac{h\Delta\eta}{2}\sum_{n=-\infty}^{\infty} \frac{\cosh(nh)}{\cosh^2(\sinh(nh))}\frac{\mathcal{F}(\eta(u=nh))}{\mathcal{F}(\eta(u=nh))\Delta\varphi - j}$$

where $\eta(u)$ is inserted from Eq. (4), and h is the stepsize for the trapezoidal rule, which is halved every refinement step, starting at $h = 1$. The sum is truncated when the terms become smaller than some suitable tolerance.

This equation is solved for j using Newton's method as in [3]. The Newton iteration is combined with the bisection method to ensure that it always converges to the correct solution. To that end, we keep track of an upper and lower bound between which the correct solution j must be located. If the Newton update would lead to a value outside that interval, we disregard it and use the interval midpoint instead. Due to the monotonicity of $\mathcal{F}(\eta)$ we always know that

$$\mathcal{F}(\eta^\mathrm{min})|\Delta\varphi - \Delta\eta| \le |j| \le \mathcal{F}(\eta^\mathrm{max})|\Delta\varphi - \Delta\eta| \tag{6}$$

with $\eta^\mathrm{min} = \min(\eta_\mathrm{L}, \eta_\mathrm{R})$ and $\eta^\mathrm{max} = \max(\eta_\mathrm{L}, \eta_\mathrm{R})$. Depending on the configuration of $\Delta\varphi$ and $\Delta\eta$ we can further reduce the range to speed up the iteration.

i) $\Delta\eta < 0$:

$$j > \Delta\varphi\mathcal{F}(\eta_\mathrm{L})\,. \tag{7}$$

ii) $0 < \Delta\eta < \Delta\varphi$:

$$j < \Delta\varphi\mathcal{F}(\eta_\mathrm{L})\,. \tag{8}$$

iii) $\Delta\eta > \Delta\varphi$:

$$j \ge \frac{\Delta\varphi(\Delta\varphi - \Delta\eta)}{\int_{\eta_\mathrm{R}-\Delta\varphi}^{\eta_\mathrm{R}} (\mathcal{F}(\eta))^{-1}\,\mathrm{d}\eta}\,. \tag{9}$$

The integral is once again calculated using Tanh-Sinh quadrature.

IV. Results

In Fig. 1 the error in j for $\mathcal{F}(\eta) = \exp(\eta)$ is compared between Gauss-Legendre (GL) [3] and TS quadrature for two parameter sets. The error is calculated in comparison with the analytical Scharfetter-Gummel solution. In the first case we choose $\Delta\varphi = 3$, $\eta_\mathrm{L} = -10$ and $\eta_\mathrm{R} = -5$ (no singularity) and in the second case $\Delta\varphi = -7$, $\eta_\mathrm{L} = -7$ and $\eta_\mathrm{R} = -5$ ($\eta_0 = -4.9992$). GL quadrature works well in the first case

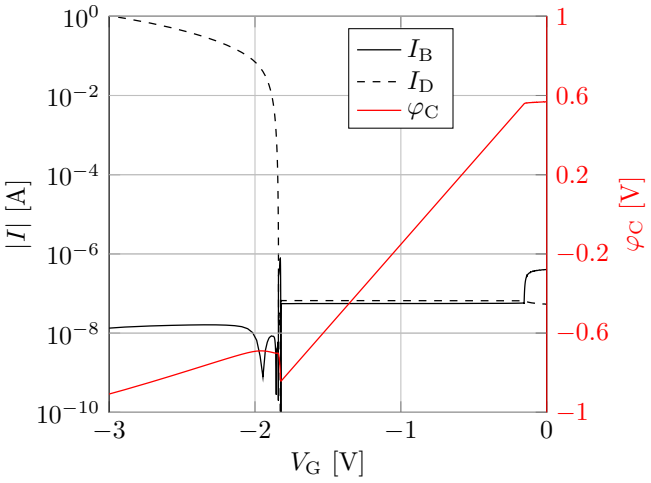

Fig. 2. Drain and bulk current and electrostatic potential at the middle of the channel surface of a pMOSFET for a transient gate voltage sweep from 0 V to −3 V taking 100 µs at 4 K and $V_{DS} = -1$ V.

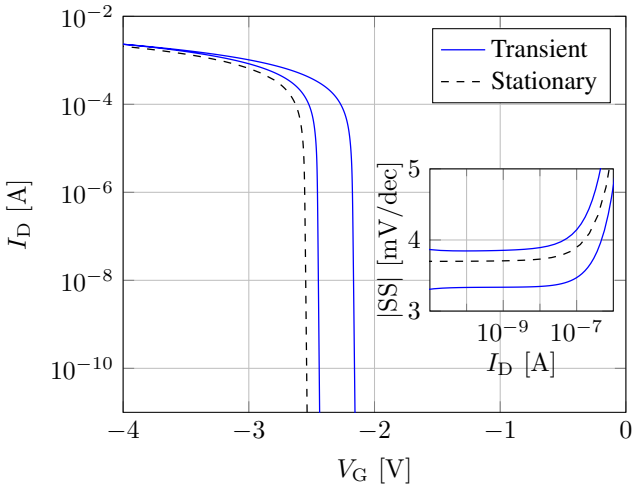

Fig. 3. IV-characteristic of a pMOSFET at 11.5 K for a dopant lifetime of 1 ps [9]. The right part of the curve belongs to the first part of the voltage ramp when the gate voltage is lowered from 0 V to −4 V, the left curve to the second part when the gate voltage is ramped back up to 0 V. The dashed line corresponds to the stationary state. The inset shows the subthreshold swing where the upper curve belongs to the first part, the lower curve to the second part and the dashed line to the stationary state.

but converges very slowly in the second case due to the singularity. If too few points are used no solution is found at all. In contrast, TS quadrature converges comparably in both cases. Note that the values chosen for the second case could easily occur in any actual device simulation, at 300 K they correspond to the densities $n_1 = 3 \times 10^{16}\,\mathrm{cm}^{-3}$ and $n_2 = 2 \times 10^{17}\,\mathrm{cm}^{-3}$ at a potential drop of $-7V_T$. Especially at low temperatures with smaller V_T the potential drop over an edge could easily be much larger making the convergence rate of GL quadrature even worse.

As an application, the stabilization is used in different 2D device simulations at cryogenic temperatures. Here, Fermi-Dirac statistics has to be used. Together with parabolic bands this yields $\mathcal{F}(\eta) = \mathcal{F}_{1/2}(\eta)$ which is calculated using the implementation from [8]. The considered device is an exemplary pMOSFET with 500 nm channel length and 10 nm oxide thickness. The donor concentration is $5 \times 10^{17}\,\mathrm{cm}^{-3}$ except close to the bulk contact where it is increased to $1 \times 10^{20}\,\mathrm{cm}^{-3}$. The acceptor concentration in the source and drain area is $1 \times 10^{20}\,\mathrm{cm}^{-3}$ which is high enough that these regions are fully ionized even at 4 K. The flatband voltage of the device is at about $V_{GS} = -0.18$ V and is approximately constant from 4 K to 77 K. For the terminal current calculation we use a device width of 1 cm.

To model the impurity freezeout occuring at cryogenic temperatures we use simple rate equations for the dopants similar to SRH recombination [9]. For electrons this reads

$$G_{DC} = \frac{1}{\tau_D} \exp\left(-\frac{e\varphi - e\Phi_n - E_D}{k_B T}\right) \frac{n}{N_C} \left(N_D - N_D^+\right) \tag{10}$$

$$R_{CD} = \frac{g_D}{\tau_D} \frac{n}{N_C} N_D^+ \tag{11}$$

$$\frac{\partial N_D^+}{\partial t} = G_{DC} - R_{CD} \tag{12}$$

where N_D and N_D^+ are the total and ionized donor concentration, respectively. G_{DC} and R_{CD} are the electron generation and recombination rates, respectively, between donor level and conduction band. The activation energy E_D is modeled according to Pearson and Bardeen [10, 11]. $g_D = 2$ is the degeneracy factor for the donor states. The rates for the acceptor states are analogous with $g_A = 4$. τ_D and τ_A are the average lifetimes of the ionized donator and acceptor states, respectively, which we assume to be 1 ps throughout this paper [9].

The first application is a time-dependent simulation of a gate voltage ramp at 4 K from 0 V to −3 V taking 100 µs. The transient simulation is performed using the TR-BDF2 method with adaptive time steps [12]. Stationary conditions are assumed at the beginning of the voltage ramp. In Fig. 2 the corresponding drain and bulk currents are shown for a fixed $V_{DS} = -1$ V. Different transient effects are visible in the resulting terminal currents. Around the flatband voltage the bulk current rises due to the accumulated electrons leaving the device through the bulk contact. While the hole channel is formed the potential overshoots causing sign changes in the bulk current. Meanwhile, the drain current rises with a very steep subthreshold slope. Running the simulation on a computer with 64 cores the calculation of the edge current densities can be easily parallelized and amounts to less than 6 % of the total runtime.

In a second simulation we qualitatively investigate a threshold voltage shift due to transient ionization. We consider a gate voltage ramp at 11.5 K from 0 V to −4 V back to 0 V, both parts taking 3 min, a typical duration for a bias sweep in experiments. The drain voltage is set to $V_{DS} = 1$ mV and stationary conditions are again assumed at the beginning. The

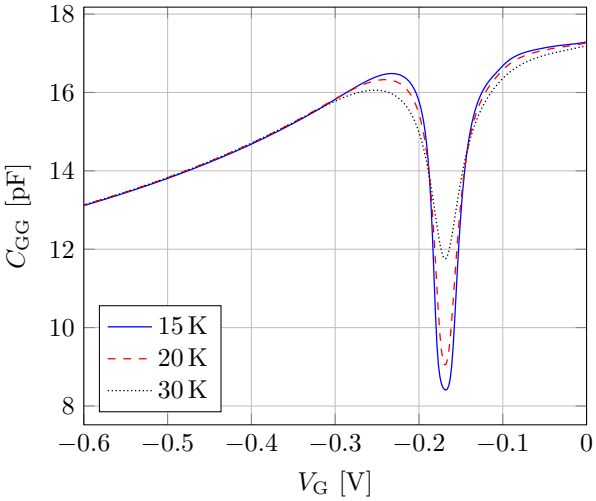

Fig. 4. Gate capacitance of the pMOSFET at $f = 1\,\mathrm{Hz}$ and $V_{DS} = 0\,\mathrm{V}$ for $T \in \{15\,\mathrm{K}, 20\,\mathrm{K}, 30\,\mathrm{K}\}$.

corresponding I-V characteristic shown in Fig. 3 includes the transient as well as the stationary solution. At the start of the simulation the channel dopants are frozen out completely. Due to the applied gate voltage they get ionized over time, leading to a transient shift of the threshold voltage. From the inset it can be seen that the subthreshold slope is influenced as well. This qualitative investigation shows that slow transient effects due to incomplete ionization can influence the device behavior at cryogenic temperatures.

To demonstrate small-signal results, we consider the gate capacitance of the exemplary pMOSFET which is extracted from its admittance parameters. The voltage dependence of the gate capacitance at low frequency ($f = 1\,\mathrm{Hz}$) is shown in Fig. 4 for three different temperatures. The sudden drop of the gate capacitance is an indicator of the impurity freezeout occuring at cryogenic temperatures and has been reported before to a lesser degree at 77 K [13]. The results indicate that a gate bias around the flatband voltage is optimal to investigate the impact of transient ionization as it results in the largest change of capacitance with temperature and frequency.

V. Conclusion

We presented a generalization of the Scharfetter-Gummel stabilization to arbitrary carrier statistics which works for coarse grids and without additional assumptions. Using this method we obtained a robust 2D solver for temperatures down to $4\,\mathrm{K}$ including Fermi-Dirac statistics which works for stationary, small-signal and time-dependent simulations. This demonstrates that the scheme is robust enough even for extreme situations. Even though we limited the discussion to parabolic band structure in this paper, the approach can easily be generalized to an arbitrary density of states.

References

[1] D. L. Scharfetter and H. K. Gummel, "Large-signal analysis of a silicon Read diode oscillator," *IEEE Trans. Electron Devices*, vol. ED-16, no. 1, pp. 64–77, 1969.

[2] P. Farrell, M. Patriarca, J. Fuhrmann, and T. Koprucki, "Comparison of thermodynamically consistent charge carrier flux discretizations for Fermi–Dirac and Gauss–Fermi statistics," *Optical and Quantum Electronics*, vol. 50, no. 2, p. 101, Feb 2018. [Online]. Available: https://doi.org/10.1007/s11082-018-1349-8

[3] M. Patriarca, P. Farrell, J. Fuhrmann, and T. Koprucki, "Highly accurate quadrature-based Scharfetter–Gummel schemes for charge transport in degenerate semiconductors," *Computer Physics Communications*, vol. 235, pp. 40–49, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010465518303527

[4] K. Gärtner, "Existence of bounded discrete steady state solutions of the van Roosbroeck system with monotone Fermi–Dirac statistic functions," *Journal of Computational Electronics*, vol. 14, pp. 773–787, 2015.

[5] P. Farrell, T. Koprucki, and J. Fuhrmann, "Computational and analytical comparison of flux discretizations for the semiconductor device equations beyond Boltzmann statistics," *Journal of Computational Physics*, vol. 346, pp. 497–513, 2017.

[6] G. Evans, R. Forbes, and J. Hyslop, "The tanh transformation for singular integrals," *International Journal of Computer Mathematics*, vol. 15, no. 1-4, pp. 339–358, 1984. [Online]. Available: https://doi.org/10.1080/00207168408803419

[7] H. Takahasi and M. Mori, "Double exponential formulas for numerical integration," *Publications of the Research Institute for Mathematical Sciences*, vol. 9, no. 3, pp. 721–741, 1974.

[8] T. Fukushima, "Precise and fast computation of Fermi–Dirac integral of integer and half integer order by piecewise minimax rational approximation," *Applied Mathematics and Computation*, vol. 259, pp. 708–729, 2015.

[9] C. Jungemann, B. Richstein, T. Linn, and J. Knoch, "Device modeling for admittance spectroscopy of PMOSFETs at cryogenic temperatures," *IEEE Transactions on Electron Devices*, vol. 71, no. 4, pp. 2322–2328, 2024.

[10] A. Akturk, J. Allnutt, Z. Dilli, N. Goldsman, and M. Peckerar, "Device modeling at cryogenic temperatures: Effects of incomplete ionization," *IEEE Transactions on Electron Devices*, vol. 54, no. 11, pp. 2984–2990, 2007.

[11] G. L. Pearson and J. Bardeen, "Electrical Properties of Pure Silicon Alloys Containing Boron and Phosphorus," *Phys. Rev.*, vol. 75, pp. 865–883, 1949.

[12] R. E. Bank, W. M. Coughran, JR., W. Fichtner, E. H. Grosse, D. J. Rose, and R. K. Smith, "Transient simulation of silicon devices and circuits," *IEEE Trans. Computer–Aided Des.*, vol. 4, pp. 436–451, 1985.

[13] A. Pirovano, A. Lacaita, A. Pacelli, and A. Benvenuti, "Novel low–temperature C–V technique for MOS doping profile determination near the Si/SiO_2 interface," *IEEE Trans. Electron Devices*, vol. 48, no. 4, pp. 750–757, 2001.