A self-consistent solution of the Poisson, Schrödinger and Boltzmann equations by a full Newton-Raphson approach for nanoscale semiconductor devices

Dino Ruić* and Christoph Jungemann
Chair of Electromagnetic Theory
RWTH Aachen University
Kackertstraße 15-17, 52072 Aachen, Germany
*Email: dr@ithe.rwth-aachen.de

***Abstract*—We present a full Newton-Raphson approach for solving the Poisson, Schrödinger and Boltzmann equations in a deterministic framework with Fourier harmonics expansion for a 2D nanoscale device. The effects of the Schrödinger equation are included via first order perturbation theory and prove to have a significant impact. A comparison to the Gummel type iteration scheme yields superiority of the full Newton-Raphson method in convergence speed and solver time. The full Newton-Raphson method is also of particular relevance to small-signal analyses in this framework.**

I. Introduction

In recent years, deterministic solvers for the Boltzmann equation (BE) became increasingly popular as they do not suffer from the inherent deficiencies of Monte Carlo simulations with their inherently transient solutions and their stochastic nature. Solving a set of equations deterministically also removes the inability to simulate rare events or events that evolve on very different time scales [1]. In device simulations, the usual approach for deterministic solvers is to solve the BE together with the Poisson equation (PE) in a Gummel type iteration, similar to [2], until self-consistency is reached. In the case that electron confinement is considered, the Schrödinger equation (SE) is also solved within that loop until self-consistency of all three equations is achieved as e.g. in [3]. Including the SE leads to an indirect feedback of the PE to the BE as the potential only enters the SE. Thus, a full Newton-Raphson method (FNRM) to solve the PE and BE simultaneously is not straightforward to carry out. However, the lack of a FNRM also obstructs simulations of small-signal quantities [1]. In this work, we present a FNRM for the coupled system of PE, SE and BE for a two-dimensional electron gas, where the effects of the SE are included via time-independent first order perturbation theory.

II. Approach

In the following, we will consider a two-dimensional device. The BE will be solved in y-direction (transport direction), while the SE will be solved in x-direction (confinement direction). Throughout this work, we will assume that the device is homogeneous in z-direction.

The basic notion of the FNRM for the PE, SE and BE is to solve the PE and BE simultaneously while expressing the change in the subband energies and wave functions as first order perturbations of the potential. We will show how to set up each equation adequately in dependence of the other equations.

A. The Schrödinger Equation

The stationary Schrödinger equation (see e.g. [4]) will be solved in x-direction for all positions in y-direction. It reads

$$\left(-\frac{\hbar^2}{2m_x^v}\frac{\partial^2}{\partial x^2} - qV(x,y)\right)\Psi^{v,s}(x,y) = \varepsilon^{v,s}(y)\ \Psi^{v,s}(x,y), \tag{1}$$

where m_x^v is the valley dependent effective mass in confinement direction, $V(x,y)$ is the potential in the device, q is the electron charge, $\varepsilon^{v,s}(y)$ is the resulting energy eigenvalue and $\Psi^{v,s}(x,y)$ is the corresponding eigenfunction. The superscript s denotes the principal quantum number, or *subband*. The Dirichlet boundary conditions were chosen for the wave function to vanish at the borders of the semiconductor region.

The structure of the SE as an eigenvalue problem prevents us from including it directly into the FNRM. However, the change in subband energies and wave functions due to a small change in the potential can be expressed with time-independent perturbation theory (see e.g. [4]). Assuming the Hamiltonian of eq. (1) is perturbed by a small change in potential $V \rightarrow V + \delta V$, we can write down the first order correction to the subband energies and wave functions for the non-degenerate case:

$$\begin{aligned}\delta\varepsilon^{v,s}(y) =& -q\int dx\ |\Psi^{v,s}(x,y)|^2\,\delta V(x,y),\\ \delta\Psi^{v,s}(x,y) =&\\ -q\sum_{s'\neq s}&\frac{\int dx'\Psi^{v,s'}(x',y)\,\delta V(x,y)\,\Psi^{v,s}(x',y)}{\varepsilon^{v,s}(y)-\varepsilon^{v,s'}(y)}\ \Psi^{v,s'}(x,y).\end{aligned} \tag{2}$$

The correction for the wave function shows that all subband energies and all wave functions of the unperturbed Hamiltonian have to be known to achieve first order accuracy.

978-1-4673-5736-4/13/$31.00 ©2013 IEEE

B. The Boltzmann Equation

The stationary BE is greatly simplified by only considering one-dimensional transport. Moreover, k-space is reduced to two dimensions by solving the SE in confinement direction:

$$\begin{aligned} F^{\mathrm{BE}} := & \frac{1}{\hbar} F_y^{v,s}(y) \, \frac{\partial}{\partial k_y} f^{v,s}(y, k_y, k_z) \\ & + v_y(k_y, k_z) \, \frac{\partial}{\partial y} f^{v,s}(y, k_y, k_z) - S\{f\} \\ = & \, 0, \end{aligned}$$

where $F_y^{v,s}$ and v_y are the force and group velocity in transport direction, respectively, $f^{v,s}$ is the distribution function and $S\{f\}$ is the scattering term. We choose Neumann boundary conditions and include a simple generation/recombination rate proportional to a constant recombination velocity for each contact [5].

In this semi-classical approach, we treat the kinetic energy as a classical quantity, while the quantized momentum in confinement direction due to the SE generates a set of minimum energies. Thus, the total energy for a given subband is the sum of the respective quantized subband energy and the classical kinetic energy. We use a simple parabolic band structure according to the model of the Modena group [6].

We apply the Herring-Vogt transformation [7], the H-transformation [8] and project onto equi-energy surfaces. We also employ the Fourier harmonics expansion, which has proved rather useful for the k-space in semiconductors. As was shown in [1], after the box-integration in energy space and real space, we obtain equations with a structure as

$$F_\alpha^{\mathrm{BE}} = \mathcal{D}_\alpha(g, \varepsilon) - \mathcal{S}_\alpha(g, \varepsilon, \Psi),$$

where $\mathcal{D}$ is the free-streaming term, $\mathcal{S}$ is the scattering term and $g \equiv Zf$ is the 'modified distribution function' with the density of states Z. The index is a 5-tuple, $\alpha \in \{(v, s, y_i, H_j, m)\}$, running over all valleys v, subbands s, spatial grid points y_i, energy grid points H_j in the H-transformed energy space and Fourier harmonics m. The terms $\mathcal{D}_\alpha$ and S_α are indicated to depend on the subband energies, while the latter also depends on the wave functions due to the overlap integral of initial and final states. The crucial aspect is that they do not explicitly depend on the potential V. Furthermore, S_α depends non-linearly on g if the Pauli principle is included.

C. The Poisson Equation

The Poisson equation is solved in 2D, considering only the electrons. It is given by

$$F^{\mathrm{PE}} := \nabla \cdot [\kappa(x, y) \nabla V(x, y)] + q[n(x, y) - N_D] = 0,$$

where $n(x, y)$ is the electron density, N_D is the donator concentration and $\kappa(x, y)$ is the dielectric constant. We use Dirichlet boundary conditions for the gate contacts and Neumann boundary conditions everywhere else.

Discretization leads to a set of equations with a structure as

$$F_a^{\mathrm{PE}} = \mathcal{L}_a(V) - \mathcal{R}_a(g, \varepsilon, \Psi),$$

where $a \in \{(x_i, y_j)\}$ is a composite index of the spatial dimensions, $\mathcal{L}_a$ represents the discretization of the differential operators and $\mathcal{R}_a$ contains the charge density. Here, the electron density was expressed as

$$n(x, y) = 2\sqrt{2\pi} \sum_{v,s} \mu^v \int dH \; g_{m=0}^{v,s}(y, H) |\Psi^{v,s}(x, y)|^2,$$

where only the $m = 0$ harmonic of g contributes and μ^v is the valley multiplicity [1]. Thus, the term $\mathcal{R}_a$ is dependent on the subband energies and wave functions and it represents the direct feedback from the BE to the PE. Note, that the subband energies are the lower bounds of integration in the H-transformed energy space.

D. Setting up the full Newton-Raphson method

The FNRM yields a linear system of equations

$$\begin{aligned} -F_\alpha^{\mathrm{BE}} &= \sum_\beta \frac{\partial F_\alpha^{\mathrm{BE}}}{\partial g_\beta} \delta g_\beta + \sum_b \frac{\partial F_\alpha^{\mathrm{BE}}}{\partial V_b} \delta V_b, \\ -F_a^{\mathrm{PE}} &= \sum_\beta \frac{\partial F_a^{\mathrm{PE}}}{\partial g_\beta} \delta g_\beta + \sum_b \frac{\partial F_a^{\mathrm{PE}}}{\partial V_b} \delta V_b, \end{aligned}$$

with the unknown variables δg and δV. After solving these equations, g and V can be updated as

$$\begin{aligned} g_\alpha^{(n+1)} &= g_\alpha^{(n)} + \delta g_\alpha^{(n)}, \\ V_a^{(n+1)} &= V_a^{(n)} + \delta V_a^{(n)}, \end{aligned}$$

where the superscript (n) indicates the n-th iteration.

As noted earlier, the BE does not explicitly depend on the potential but rather on the subband energies and wave functions stemming from the SE. However, the FNRM requires us to express the derivative $\partial F_\alpha^{\mathrm{BE}} / \partial V_b$ in order to include the feedback from the PE to the BE. To this end, we can rewrite the derivative as

$$\frac{\partial F_\alpha^{\mathrm{BE}}}{\partial V_b} = \sum_u \frac{\partial F_\alpha^{\mathrm{BE}}}{\partial \varepsilon_u} \frac{\partial \varepsilon_u}{\partial V_b} + \sum_w \frac{\partial F_\alpha^{\mathrm{BE}}}{\partial \Psi_w} \frac{\partial \Psi_w}{\partial V_b}.$$

Again, the indices $u \in \{(v, s, y_i)\}$ and $w \in \{(v, s, x_i, y_j)\}$ are used as an abbreviation for their functions' domains of definition. The derivatives of the subband energies and wave functions can be evaluated using the first order perturbation theory of eq. (2):

$$\frac{\partial \varepsilon_u}{\partial V_b} = \frac{\partial(\delta \varepsilon_u)}{\partial(\delta V_b)}, \quad \frac{\partial \Psi_w}{\partial V_b} = \frac{\partial(\delta \Psi_w)}{\partial(\delta V_b)}.$$

Handling the PE analogously, we obtain a set of equations as

$$\begin{aligned} -F_\alpha^{\mathrm{BE}} = & \sum_\beta \left[\frac{\partial \mathcal{D}_\alpha}{\partial g_\beta} - \frac{\partial \mathcal{S}_\alpha}{\partial g_\beta} \right] \delta g_\beta - \sum_{b,w} \frac{\partial \mathcal{S}_\alpha}{\partial \Psi_w} \frac{\partial \Psi_w}{\partial V_b} \, \delta V_b \\ & + \sum_{b,u} \left[\frac{\partial \mathcal{D}_\alpha}{\partial \varepsilon_u} - \frac{\partial \mathcal{S}_\alpha}{\partial \varepsilon_u} \right] \frac{\partial \varepsilon_u}{\partial V_b} \, \delta V_b, \\ -F_a^{\mathrm{PE}} = & - \sum_\beta \frac{\partial \mathcal{R}_a}{\partial g_\beta} \, \delta g_\beta + \sum_b \frac{\partial \mathcal{L}_a}{\partial V_b} \, \delta V_b \\ & - \sum_{b,u} \frac{\partial \mathcal{R}_a}{\partial \varepsilon_u} \frac{\partial \varepsilon_u}{\partial V_b} \delta V_b - \sum_{b,w} \frac{\partial \mathcal{R}_a}{\partial \Psi_w} \frac{\partial \Psi_w}{\partial V_b} \delta V_b. \end{aligned} \tag{3}$$

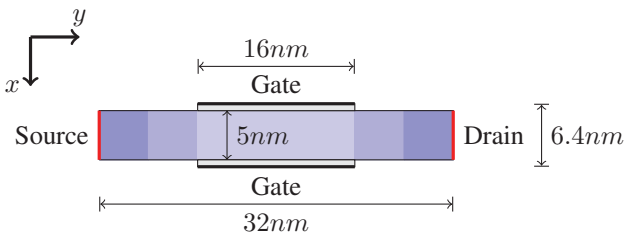

Fig. 1. Double gate nMOSFET with silicon channel. The shades in the channel indicate the doping density N_D.

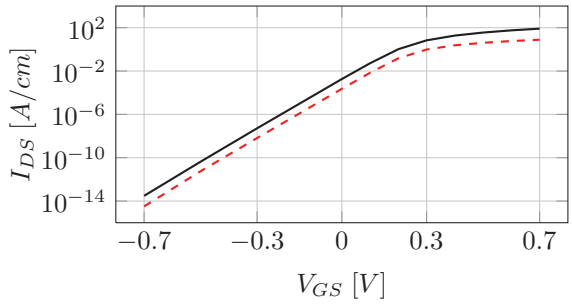

Fig. 2. Current I_{DS} vs. gate voltage V_{GS} with drain voltage $V_{\mathrm{DS}} = 0.01V$ (dashed line) and $V_{\mathrm{DS}} = 0.7V$ (solid line).

Note, that this means that the derivative w.r.t. to the potential of *every* term ocurring in the BE has to be taken, including all the boxes of integration over H-space and the scattering term with its overlap integral.

III. Results

We investigated a two-dimensional double gate nMOSFET with a silicon channel as shown in Fig. 1. For the following simulations, we used an equidistant two-dimensional rectangular grid with a gridpoint distance of $0.1nm$ and $0.5nm$ in x- and y-direction, respectively.

The SE was solved for all subbands to obtain the correct first order perturbation of the wave function as shown in eq. (2). However, the subbands included in the BE were limited to the 10 lowest ones. For the BE, we used a static H-grid with a grid spacing of $5.24meV$. The Fourier harmonics expansion was truncated after the 13th harmonic as no significant changes were observed beyond that order. The scattering term includes intra-valley elastic acoustic phonon scattering and inter-valley phonon scattering [1], while the Pauli principle was omitted.

To obtain feasible starting values for the FNRM we used a PE-SE loop until self-consistency was reached and thereafter a PE-SE-BE Gummel type iteration up to a predefined threshold, where the FNRM solver took over. The SE was solved before each FNRM iteration step to update the values of the subband energies and wave functions.

The resulting drain-source current I_{DS} vs. gate bias V_{GS} for a drain bias of $V_{\mathrm{DS}} = 0.01V$ and $V_{\mathrm{DS}} = 0.7V$ is shown in Fig. 2. The electron densities integrated in confinement direction for a few gate biases are depicted in Fig. 3.

The main aspect of this work is the comparison between the presented FNRM with the more conservative approach using a Gummel type iteration scheme to iterate over a PE-SE-BE-loop where each equation is solved separately. We compared

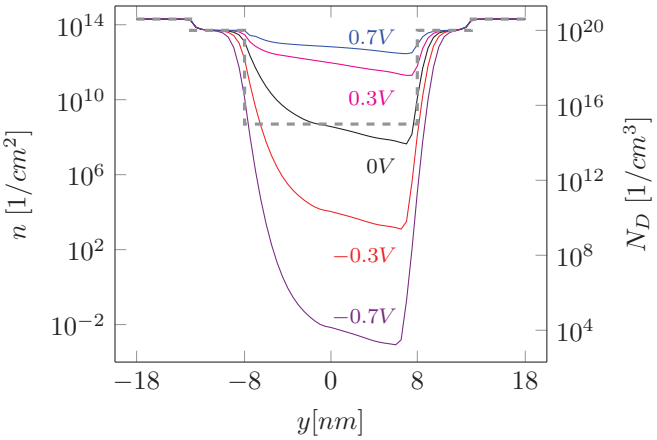

Fig. 3. Electron densities n integrated in direction perpendicular to transport direction for $V_{\mathrm{DS}} = 0.7V$ and $V_{\mathrm{GS}} = -0.7V,\ -0.3V,\ 0V,\ 0.3V,\ 0.7V$ (solid lines with adjacent values of the respective V_{GS}) together with the doping density N_D (dashed line).

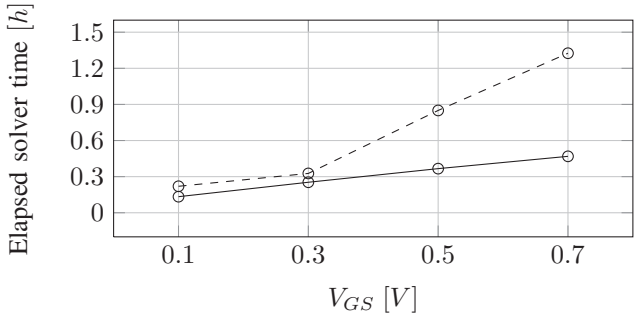

Fig. 4. Total elapsed solver time for Full Newton-Raphson approach (solid line) and Gummel type iteration (dashed line) for $V_{\mathrm{DS}} = 0.7V$.

the total elapsed solver time of both approaches for $V_{DS} = 0.7V$ and varying gate biases (see Fig. 4) and the number of iterations necessary to achieve an error of less than $10^{-11}V$ in the potential (see Fig. 5). An iteration step with the FNRM takes longer than an iteration step with the Gummel scheme. However, the quadratic convergence of the FNRM leads to a significant decrease in total solver time at high biases. At $V_{DS} = 0.7V$ and $V_{GS} = 0.7V$, the FNRM converges in only 17 steps while the Gummel scheme takes over 100 iterations. This is due to the strong coupling of the BE and PE and while in the Gummel Scheme the feedback has to propagate through incremental changes of the starting values of each equation, the FNRM includes the influence of the PE on the BE and vice versa directly. The downside of the FNRM is its instability, especially at high gate biases, and its tendency to diverge if the starting values are not in the vicinity of the solution.

During an inspection of the convergence behavior regarding the necessity of the derivatives w.r.t. to the subband energies and wave functions of eq. (3), we found that omitting the derivatives w.r.t. the wave functions in the BE has little effect far away from the solution, however, in the vicinity of the solution, the quadratic convergence is lost. In some cases we observed the solver to be more stable due to the weaker feedback. Omitting any of the other derivatives leads – in all cases we tested – to a diverging system of equations. Obviously, a combination of omittance of the derivatives w.r.t. the wave functions far away from the solution to gain stability

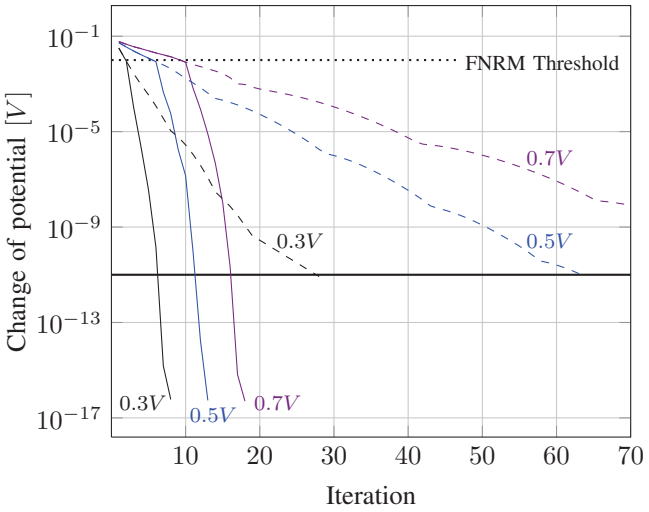

Fig. 5. Convergence given by the absolute change of the potential at each iteration step for $V_{\mathrm{GS}} = 0.3V,\ 0.5V,\ 0.7V$ (values indicated in figure) of full Newton-Raphson approach (solid lines) and Gummel type iteration (dashed lines) for $V_{\mathrm{DS}} = 0.7V$.

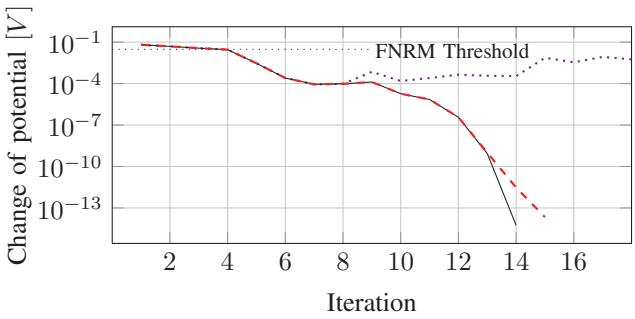

Fig. 6. Convergence behavior with higher FNRM threshold of FNRM with all derivatives (dotted line), without the derivatives of the wavefunction in the BE (dashed line) and a combination of both (solid line) for $V_{\mathrm{GS}} = 0.7V, V_{\mathrm{DS}} = 0.7V$.

and inclusion of the wave functions in the vicinity of the solution will yield the fastest convergence for those cases. In Fig. 6, the convergence behaviors are shown. Here, the FNRM threshold was chosen to be slightly higher to demonstrate the instability of the FNRM far away from the solution.

IV. Towards Small-Signal Analysis

The usefulness of the FNRM is limited in the context of sole DC calculations as the amount of programming prowess and time necessary to include all the derivatives correctly, vastly surpasses the gains in solver time. However, with the Gummel scheme, small-signal analysis is not manageable. But having implemented the FNRM, it is easily incorporated. Carrying out the linearization of the BE with a small AC perturbation yields – apart from the time-derivative – the very same equations as in eq. (3) with the l.h.s set to zero. Dealing with the time-derivative requires some carefulness as the H-grid depends on the subband energies which themselves depend on time. Thus, the H-grid itself is time-dependent and the time derivative in the BE takes the following form:

$$\frac{\partial g(y, H(t), t)}{\partial t} = \frac{\partial g(y, H, t)}{\partial H}\,\frac{dH(t)}{dt} + \frac{\partial g(y, H, t)}{\partial t},$$

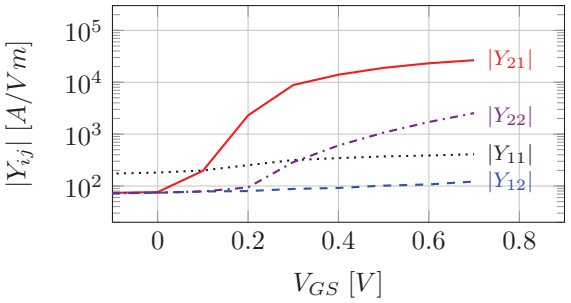

Fig. 7. Absolute values of Y-parameters at 100GHz in common source configuration. Top gate and bottom gate are short-circuited. Port 1 and 2 denote gates vs. source and drain vs. source, respectively.

where on the r.h.s., as an argument of g, H is considered an independent variable. Some preliminary results are shown in Fig. 7.

V. Conclusion

We have shown that by using time-independent first order perturbation theory a self-consistent quadratically converging full Newton-Raphson approach for the combined system of Poisson, Schrödinger and Boltzmann equations can be achieved. The full Newton-Raphson method is superior to the Gummel type iteration scheme in convergence behavior and total solver time. In addition we found that turning the derivatives w.r.t. the wave functions in the Boltzmann equation off has little effect far from the solution but – of course – the quadratic convergence is lost. However, in some cases the system becomes more stable. A working full Newton-Raphson approach marks an important milestone on the way towards small-signal analysis in this framework.

Acknowledgment

Financial support by the 'Deutsche Forschungsgemeinschaft' (DFG) is gratefully acknowledged.

References

[1] S.-M. Hong, A. T. Pham, and C. Jungemann, *Deterministic solvers for the Boltzmann transport equation*. Computational Microelectronics, Wien, New York: Springer, 2011.

[2] H. K. Gummel, "A self-consistent iterative scheme for one-dimensional steady state transistor calculations," *Electron Devices, IEEE Transactions on*, pp. 455–465, October 1964.

[3] A. Pham, C. Jungemann, and B. Meinerzhagen, "On the numerical aspects of deterministic multisubband device simulations for strained double gate PMOSFETs," *J. Computational Electronics*, vol. 8, no. 3, pp. 242–266, 2009.

[4] R. Shankar, *Principles of quantum mechanics*. New York: Springer, 1994.

[5] C. Jungemann, A.-T. Pham, B. Meinerzhagen, C. Ringhofer, and M. Bollhöfer, "Stable discretization of the Boltzmann equation based on spherical harmonics, box integration, and a maximum entropy dissipation principle," *J. Appl. Phys.*, vol. 100, pp. 024502–1–13, 2006.

[6] R. Brunetti, C. Jacoboni, F. Nava, L. Reggiani, G. Bosman, and R. J. J. Zijlstra, "Diffusion coefficient of electrons in silicon," *J. Appl. Phys.*, vol. 52, pp. 6713–6722, 1981.

[7] C. Herring and E. Vogt, "Transport and deformation-potential theory for many-valley semiconductors with anisotropic scattering," *Phys. Rev.*, vol. 101, no. 3, pp. 944–962, 1956.

[8] A. Gnudi, D. Ventura, G. Baccarani, and F. Odeh, "Two-dimensional MOSFET simulation by means of a multidimensional spherical harmonics expansion of the Boltzmann transport equation," *Solid–State Electron.*, vol. 36, no. 4, pp. 575 – 581, 1993.