ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Computational Physics 204 (2005) 533–561

JOURNAL OF COMPUTATIONAL PHYSICS

www.elsevier.com/locate/jcp

Quantum-corrected drift-diffusion models for transport in semiconductor devices

Carlo de Falco [a], Emilio Gatti [b], Andrea L. Lacaita [b], Riccardo Sacco [c,*]

[a] Dipartimento di Matematica "F.Enriques", Università degli Studi di Milano, via Saldini 50, 20133 Milano, Italy
[b] DEI – Dipartimento di Elettronica e Informazione, Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milano, Italy
[c] Dipartimento di Matematica "F.Brioschi", Politecnico di Milano, via Bonardi 9, 20133 Milano Italy

Received 12 May 2004; received in revised form 5 October 2004; accepted 18 October 2004
Available online 30 November 2004

Abstract

In this paper, we propose a unified framework for Quantum-corrected drift-diffusion (QCDD) models in nanoscale semiconductor device simulation. QCDD models are presented as a suitable generalization of the classical drift-diffusion (DD) system, each particular model being identified by the constitutive relation for the quantum-correction to the electric potential. We examine two special, and relevant, examples of QCDD models; the first one is the modified DD model named Schrödinger–Poisson–drift-diffusion, and the second one is the quantum-drift-diffusion (QDD) model. For the decoupled solution of the two models, we introduce a functional iteration technique that extends the classical Gummel algorithm widely used in the iterative solution of the DD system. We discuss the finite element discretization of the various differential subsystems, with special emphasis on their stability properties, and illustrate the performance of the proposed algorithms and models on the numerical simulation of nanoscale devices in two spatial dimensions.
© 2004 Elsevier Inc. All rights reserved.

Keywords: Quantum and drift-diffusion models; Density-gradient; Schrödinger–Poisson; Functional iterations; Finite element method; Nanoscale semiconductor devices

1. Introduction and motivations

The accurate prediction of the electrical behavior of up-to-date nanoscale semiconductor devices demands for the inclusion of quantum effects, such as, for example, increased equivalent oxide thickness due to strong electron confinement at silicon–silicon dioxide interface or direct tunneling through the

* Corresponding author. Tel.: +390223994540; fax: +390223994513.
E-mail address: riccardo.sacco@mate.polimi.it (R. Sacco).

0021-9991/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2004.10.029

channel potential barrier [1–3]. As these effects cannot be appropriately dealt with by the classical drift-diffusion (DD) model, alternative and more sophisticated mathematical models must be adopted. In this respect, two main approaches can be taken:

(i) employing a full quantum description of charge transport;
(ii) adding a suitable correction to the basic DD model to include quantum electrostatics.

The first approach typically relies on quantum models based on the non-equilibrium Green's function [4] or the Wigner function [5]. Another model is adopted in [6] to study the behaviour of devices in the quantum ballistic limit and requires strong assumptions on the wavefunctions in the metal leads. All of these models provide a very accurate and complete physical information on the quantum mechanical phenomena occurring in the device. However, a certain lack of numerical robustness and the intensive computational cost make these models still unsuitable for routine industrial semiconductor device simulation. These considerations strongly prompt towards investigating the approach (ii). This latter has the advantage to exploit all the benefits arising from the well established mathematical and numerical experience on the basic DD model, allowing at the same time to design a state-of-the-art simulation tool.

In this work, we focus on two relevant modifications of the basic DD system, namely, the quantum–drift-diffusion (QDD) model proposed in [7] and the Schrödinger–Poisson–drift-diffusion (SPDD) model proposed in [8].

The QDD model emanates from a self-consistent derivation of a generalized equation of state for the electron gas which includes a dependence on the gradient of electron density. This, in turn, allows to incorporate quantum phenomena description into the classical DD model by means of a quantum correction to the electric potential, the so called *Bohm potential* [9].

The SPDD model is based instead on a *fully two-dimensional* consistent solution of the Schrödinger–Poisson subsystem within a subregion of the semiconductor device, and incorporates quantum effects into the classical DD framework by means of a *modified density of states*. The SPDD model is more accurate than previous simulation schemes based on the coupling of the DD model with a one-dimensional solution of the Schrödinger equation [1,10–12, 43], requiring at the same time a reduced computational cost compared to the solution of a full quantum model.

In this paper, we show that the modification of the density of states in the SPDD model can indeed be regarded as a *suitable quantum correction of the electric potential* in the DD current relation. This allows to interpret the SPDD model as a variation of the QDD model, and, more generally, to view both the SPDD and the QDD models as special instances of a unified mathematical framework for nanoscale semiconductor device simulation, denoted henceforth as quantum-corrected drift-diffusion (QCDD) approach. This latter family of models thus represents a suitable generalization of the classical drift-diffusion (DD) system, each particular model being identified by the constitutive relation for the quantum-correction to the electric potential.

Following the derivation of a unified framework for QCDD transport models, we propose to apply a functional iteration technique, which is customarily and successfully used in standard DD-based semiconductor device simulation, to the decoupled numerical solution of the nonlinear boundary value problems deriving from both the SPDD and QDD models. The iteration procedure is a generalized Gummel algorithm [13–15] and, to our knowledge, represents the first fully decoupled functional iteration procedure applied to the QDD model (see also [16,17] for the use of other partially decoupled iterative maps). The advantage of adopting a Gummel-type iteration is twofold: on the one hand, it considerably saves computational effort and memory storage at each step, which is of relevant importance in multidimensional simulations; on the other hand, it leads to successively solving elliptic quasi-linear and linear boundary value problems for which efficient and stable discretization methods can be employed [18,19]. Moreover, the performance of the iterative procedure can be properly improved by resorting to suitable acceleration techniques, for example, Newton–Krylov subspace iterations [20,21].

The outline of the paper is as follows. In Section 2, we introduce the DD, SPDD and QDD transport models, while in Section 3, we present a generalized framework for QCDD models, that encloses as special cases both the SPDD and QDD models. Then, in Section 4, we introduce a functional iteration to construct a decoupled algorithm for the iterative solution of the SPDD and QDD systems, and in Section 5, we discuss the finite element discretization of the various differential subproblems obtained after decoupling. The detailed description of the numerical algorithms used in the computations is discussed in Section 6. Finally, in Section 7, we illustrate and compare the performance of the proposed algorithms and models on the numerical simulation of nanoscale devices in two spatial dimensions. Some concluding remaks and perspectives on future work are addressed in Section 8.

2. Physical models

In this section, we present the classical DD transport model and introduce the notation that will be used for the SPDD and QDD models. For ease of presentation, we develop in detail the derivation of the models only for electron carriers, as a similar treatment holds for hole carriers. Throughout the section, we shall assume that the device domain Ω is an open bounded set divided into two subregions, a semiconductor region, Ω_{Si}, and an oxide region, Ω_{OX}, such that $\overline{\Omega} = \overline{\Omega}_{\mathrm{Si}} \cup \overline{\Omega}_{\mathrm{OX}}$. In each of these two regions, we consider the material to be homogeneous and isotropic; this implies, in particular, that the electric permittivity ε is a scalar piecewise constant quantity over Ω. The above device structure is an appropriate model for a MOSFET (Metal–Oxide–Semiconductor Field-Effect Transistor), one of the most widely used component in nowadays semiconductor device technology (see Sections 2.2, 3.2 and 7 for a more detailed description).

2.1. The DD model

The classical DD model is the zeroth order moment in the expansion of the Boltzmann Transport Equation (see [5,15]) and, in the stationary case, it consists of the following system:

$$\begin{cases} \mathrm{div} J_n = qU & \text{in } \Omega_{\mathrm{Si}}, \\ \mathrm{div} J_p = -qU & \text{in } \Omega_{\mathrm{Si}}, \\ -\mathrm{div}(\varepsilon \nabla \varphi) = q(p - n + D) & \text{in } \Omega, \end{cases} \tag{1}$$

where $q > 0$ is the electron charge, J_n and J_p are the electron and hole current densities, D is the net doping profile in the device and U is the net recombination rate. In (1), the first two equations represent the continuity equations for electrons and holes n and p, while the third equation is the Poisson equation for the electrostatic potential φ. The issue of a suitable set of boundary conditions for (1) will be addressed in Section 3.2.

The following constitutive relation for the electron current density J_n in terms of n and of the electron quasi-Fermi potential φ_n can be derived through a linearization of the distribution function in phase space around its equilibrium value (see [13, Chapter 2])

$$J_n(x) = -q\mu_n n(x) \nabla \varphi_n(x), \tag{2}$$

where x denotes the spatial coordinate of a point in the device domain and μ_n is the electron mobility. In (2), φ_n describes the deviation of the distribution function from its equilibrium value and it reduces to the constant Fermi potential φ_{n_0} at equilibrium when no current flow is expected. The electron density n can be expressed as

$$n(x) = \int_{E_c}^{+\infty} g(E, E_c(x)) f(\varphi_n(x), E)\, \mathrm{d}E \quad \text{in } \Omega_{\mathrm{Si}}, \tag{3}$$

where E denotes energy, $g(E,E_c(x))$ is the density of states, $f(\varphi_n(x),E)$ is the occupation probability density and $E_c(x)$ is the energy at the bottom of the conduction band.

Let x be a fixed point in Ω_{Si}. Then, for a non-degenerate semiconductor and assuming a single parabolic energy band, we have

$$g(E, E_c(x)) = \frac{1}{2\pi^2}\left(\frac{2m_n^*}{\hbar^2}\right)^{3/2}(E - E_c(x))^{1/2}, \quad E \geqslant E_c(x) \tag{4}$$

and

$$f(\varphi_n(x), E) = \exp\left(-\frac{E + q\varphi_n(x)}{k_b T}\right),$$

where m_n^* is the (scalar) electron effective mass, $\hbar$ is the reduced Planck constant, k_b is the Boltzmann constant and T is the lattice temperature. The computation of the integral in (3) yields

$$n(x) = N_c \exp\left(\frac{-q\varphi_n(x) - E_c(x)}{k_b T}\right), \tag{5}$$

where N_c is the effective density of states in the conduction band. Let us introduce the intrinsic carrier concentration

$$n_{int} = \sqrt{N_c N_v}\exp\left(-\frac{E_{Gap}}{2k_b T}\right) \quad \text{in } \Omega_{Si},$$

where E_{Gap} is the (constant) energy band gap and N_v is the effective density of states for holes of scalar effective mass m_p^* in the valence band. Then, relation (5) can be written as

$$n(x) = n_{int}\exp\left(\frac{\varphi(x) - \varphi_n(x)}{V_{th}}\right) \quad \text{in } \Omega_{Si}, \tag{6}$$

where $V_{th} = k_b T/q$ is the thermal potential and

$$\varphi(x) = -\frac{E_c(x) + E_v(x)}{2q} + V_{th}\ln\left(\sqrt{N_c/N_v}\right), \tag{7}$$

$E_v(x) = E_c(x) - E_{Gap}$ being the energy at the top of the valence band. Replacing (6) into (2) yields the familiar DD expression of the electron current density

$$J_n(x) = -q\mu_n(n(x)\nabla\varphi(x) - V_{th}\nabla n(x)).$$

Repeating the above procedure for the hole density p, yields the following expression of the hole current density:

$$J_p(x) = -q\mu_p(p(x)\nabla\varphi(x) + V_{th}\nabla p(x)),$$

where μ_p is the hole mobility.

2.2. The SPDD model

The SPDD model proposed in [8] generalizes the DD system by replacing the explicit expression (4) of $g(E,E_c(x))$ in terms of E and $E_c(x)$ with a direct computation of the discrete spectrum of energy states for electron particles. The set of admissible energy states $\{E_i\}$, $i \in \mathbb{N}$, is computed through the solution of the stationary eigenvector/eigenvalue Schrödinger problem

$$-\text{div}\left(\frac{\hbar^2}{2m_n^*}\nabla\psi(x)\right) + E_c(x)\psi(x) = E\psi(x) \quad \text{in } \Omega_{Schr}, \tag{8}$$

where, by using (7), the relation between $E_c(x)$ and $\varphi(x)$ can be written as

$$E_c(x) = -q\varphi(x) + \left(\frac{E_{Gap}}{2} + k_b T \ln\sqrt{\frac{N_c}{N_v}}\right),$$

and $\Omega_{Schr} \subseteq \Omega_{Si}$ is a computational domain to be properly defined.

To simplify the presentation, we introduce the following assumptions:

(a) $\Omega_{Schr} \equiv \Omega_{Si}$;
(b) $\Omega_{Schr} \subset \mathbb{R}^3$.

We defer the discussion of these two assumptions and of their associated physical consequences to the end of this section, where we will also extend the presentation below to situations in which these assumptions do not hold.

To recover a density of states from the line spectrum obtained by solving (8), we introduce the following new definition for g

$$g(E,x) = 2\sum_i \delta(E - E_i)|\psi_i(x)|^2. \tag{9}$$

In (9), the summation is performed over all the possible energy states and $\delta(E - E_i)$ denotes the Dirac delta function centered at E_i. Notice that this *modified density of states* does not depend explicitly on $E_c(x)$, as in the case of the DD model. Rather, its spatial variation is given by the wavefunctions $\psi_i(x)$ which are *non-local* functions of $\varphi(x)$. Also notice that the factor 2 in (9) accounts for the spin degeneracy of the energy eigenstates. Inserting (9) into (3), yields the following modified expression of the free electron density:

$$n(x) = \int_{E_c}^{+\infty} g(E,x) f(\varphi_n(x), E)\, \mathrm{d}E = \sum_i |\psi_i(x)|^2 f(\varphi_n(x), E_i) \quad \text{in } \Omega_{Si}. \tag{10}$$

The choice of appropriate boundary conditions for (8) is a critical issue. A straightforward approach, which turns out to be computationally efficient, has been adopted in [8] and consists in setting

$$\psi(x)|_{\partial\Omega_{Si}} \equiv 0, \tag{11}$$

where $\partial\Omega_{Si}$ is the boundary of the semiconductor domain Ω_{Si}. Enforcing homogeneous Dirichlet boundary conditions for the wavefunction has the unphysical consequence that $n|_{\partial\Omega_{Si}} \equiv 0$. To solve this inconsistency, the classical expression (6) is forced to hold in regions of the device where quantum effects are assumed to be negligible. This, in turn, demands to define a quantum region $\Omega_Q \subset \Omega_{Si}$, where (10) holds, and a classical region $\Omega_{cl} \subset \Omega_{Si}$, where (6) holds instead, so that $\Omega_{Si} = \Omega_Q \cup \Omega_{cl}$. To properly account for the resulting *splitted* constitutive relation for n, a *quantum correction factor* is defined as follows. We introduce a *quantum electron density*

$$n_q(x) = 2\sum_i |\psi_i(x)|^2 f(\varphi_n(x), E_i) \quad \text{in } \Omega_Q \tag{12}$$

and a *classical electron density* $n_{cl}(x)$ given by relation (6). Then, we set

$$\gamma_n(x) = \begin{cases} 1 & \text{in } \Omega_{cl}, \\ \dfrac{n_q(x)}{n_{cl}(x)} & \text{in } \Omega_Q \end{cases} \tag{13}$$

and, accordingly, we define

$$n(x) = \gamma_n(x) n_{int} \exp\left(\frac{\varphi(x) - \varphi_n(x)}{V_{th}}\right) = \gamma_n(x) n_{cl}(x)(x) \quad \text{in } \Omega_{Si}. \tag{14}$$

Definition (13) introduces a jump discontinuity in the electron density n across the interface between Ω_Q and Ω_{cl}. A quantitative analysis of this issue is carried out in Section 3.4 where it is shown that a proper choice of the ratio between the sizes of Ω_Q and Ω_{Schr} leads to negligible effects of such discontinuity. By setting

$$G_n(x) = V_{\text{th}} \ln(\gamma_n(x)),$$

relation (14) can be rewritten as

$$n(x) = n_{\text{int}} \exp\left(\frac{\varphi(x) + G_n(x) - \varphi_n(x)}{V_{\text{th}}}\right) \quad \text{in } \Omega_{\text{Si}} \tag{15}$$

so that, repeating the derivation carried out in Section 2.1, with (14) in place of (6), yields the following *modified* constitutive relation for the electron current density

$$J_n(x) = -q\mu_n n(x)\nabla\varphi_n(x) = -q\mu_n(n(x)\nabla(\varphi(x) + G_n(x)) - V_{\text{th}}\nabla n(x)). \tag{16}$$

Following a completely similar procedure, the hole current density reads:

$$J_p(x) = -q\mu_p p(x)\nabla\varphi_p(x) = -q\mu_p(p(x)\nabla(\varphi(x) + G_p(x)) + V_{\text{th}}\nabla p(x)). \tag{17}$$

Let us now come back to assumptions (a) and (b) introduced at the beginning of the section. In most real-life applications both these conditions are too restrictive, so that it is convenient to extend the SPDD model to the case where neither of the two conditions (a) and (b) holds. As for assumption (a), we notice that the computation of the eigenvalues of the discretized Hamiltonian is a rather intensive numerical task. Therefore, it is convenient to solve (8) only in a smaller subdomain $\Omega_{\text{Schr}} \subset \Omega$ that needs to be properly defined in such a way that the closed boundary condition (11) does not relevantly affect the quantum charge $n_q(x)$ in Ω_Q, which is the device subregion under the interface between silicon and silicon dioxide where quantum effects are expected to be important. A representation of the device subdvision into the several physical subdomains described in this section is shown in Fig. 1.

As for assumption (b), it must be noted that if the Schrödinger equation is solved in $\Omega \subset \mathbb{R}^d$, $d = 1, 2$, then quantization effects are accounted for only in $\mathbb{R}^d$, so that a continuum approximation similar to (9) must be adopted for the energy levels associated with the remaining spatial directions. Therefore (9) and (10) only hold if $\Omega \subset \mathbb{R}^3$, while in the case $\Omega \subset \mathbb{R}^d,\ d = 1, 2$, we have

$$g(E, x) = \sum_i H(E - E_i) \mid \psi_i(x)|^2 g_{3-d}(E), \tag{18}$$

where $H(E - E_i)$ denotes the Heavyside step-function centered at E_i. Accordingly, (12) becomes

$$n_q(x) = \int_{E_c}^{+\infty} g(E) f(\varphi_n(x), E)\, \mathrm{d}E = \sum_i |\psi_i(x)|^2 \int_{E_i}^{\infty} g_{3-d}(E) f(\varphi_n(x), E)\, \mathrm{d}E, \tag{19}$$

where

$$\begin{cases} g_2 = \dfrac{m_n^*}{\pi\hbar^2} & \text{if } d = 1, \\ g_1 = \sqrt{\dfrac{m_n^*}{2\pi^2\hbar^2}}(E - E_i)^{-(1/2)} & \text{if } d = 2. \end{cases} \tag{20}$$

2.3. The QDD model

The QDD model was introduced in [22] and originally named Density Gradient model because it was obtained by allowing the electron gas equation of state to relate the quasi-Fermi potential φ_n not only to the electron density but also to its gradient. According to the single band version of the QDD model, the electron density can be expressed using the modified Maxwell–Boltzmann relation (15) and the quan-

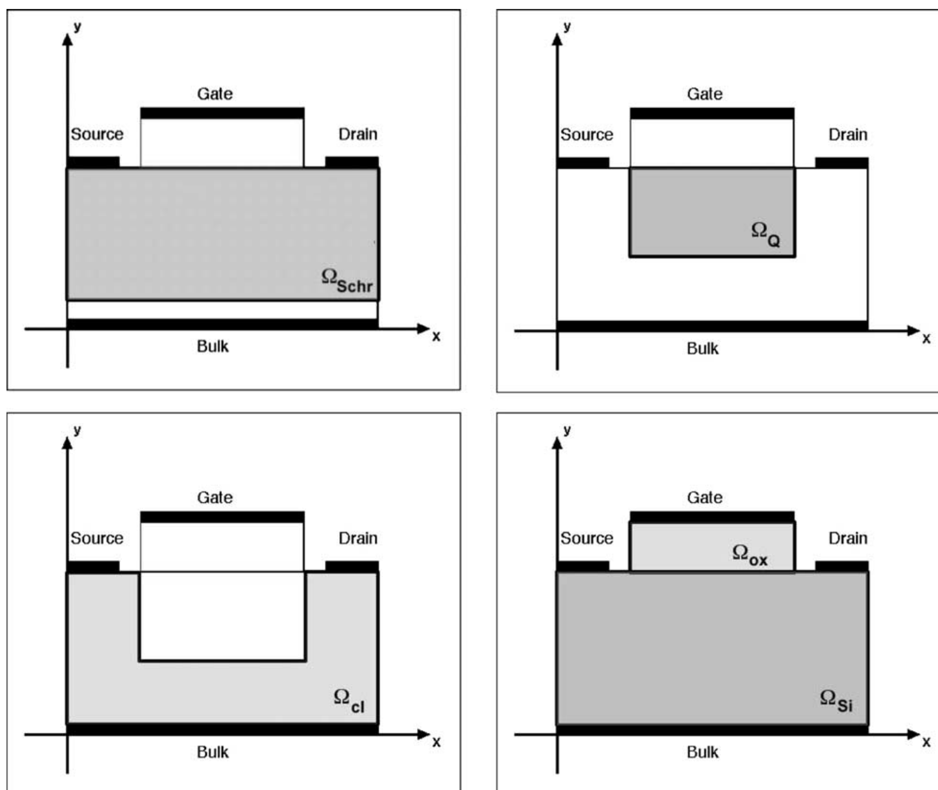

Fig. 1. Subdomain division of a two-dimensional cross-section of a MOSFET device.

tum correction G_n to the electric potential φ (usually referred to as the *Bohm potential*) is defined by the relation

$$G_n = \frac{1}{\sqrt{n}} \operatorname{div}\left(\frac{\hbar^2}{6qm_n^*} \nabla \sqrt{n}\right). \tag{21}$$

For a microscopic derivation of (21) from the Wigner equation, we refer to [9] and [7], while in [22] a macroscopic approach is adopted. Notice that the QDD expressions for the current densities are formally identical to (16) and (17). This formal analogy will be exploited in Section 6 to construct numerical algorithms for the solution of the SPDD and QDD transport models.

2.4. Scaling

Before proceeding, it is useful to rewrite the DD, QDD and SPDD systems in a scaled form. The scaling has the advantage to emphasize the *singularly perturbed* nature of the equations, which can be used as in [14] to perform an a priori qualitative analysis of their solutions.

Our choice for the scaling parameters is as follows [13,14]:

- the scaling factor for the carrier densities is chosen to be the maximum value of the doping throughout the device: $\overline{n} = \|D\|_{\infty,\Omega}$, where $\|\cdot\|_{\infty,\Omega}$ is the norm for the space $L^\infty(\Omega)$ of essentially bounded functions over Ω;

- voltages are scaled with respect to the thermal voltage, therefore $\overline{\varphi} = V_{\mathrm{th}}$;
- lengths are scaled with respect to the diameter of the device: $\overline{x} = \max_{x_1, x_2 \in \Omega} |x_1 - x_2|$;
- mobilities are scaled with respect to the zero field electron mobility: $\overline{\mu} = \mu_{n_0}$.

Accordingly, we define the following non-dimensional quantities:

$$\widehat{n} = \frac{n}{\overline{n}}, \quad \widehat{\varphi} = \frac{\varphi}{\overline{\varphi}}, \quad \widehat{\varphi_n} = \frac{\varphi_n + V_{\mathrm{th}} \ln(n_{\mathrm{int}}/\overline{n})}{\overline{\varphi}}, \quad \widehat{x} = \frac{x}{\overline{x}}, \quad \widehat{\mu_n} = \frac{\mu_n}{\overline{\mu}}, \quad \widehat{U} = U \frac{\overline{x}^2}{q \overline{\mu n}} \tag{22}$$

and the non-dimensional coefficients

$$\delta_n^2 = \frac{\hbar^2}{6 q m_n^* \overline{x}^2 \overline{\varphi}}, \quad \beta = \frac{V_{\mathrm{th}}}{\overline{\varphi}} = 1, \quad \lambda^2 = \frac{\overline{\varphi} \varepsilon}{q \overline{x}^2 \overline{n}}. \tag{23}$$

Using (22) and (23), Eq. (21) becomes

$$\widehat{G_n} = \frac{1}{\sqrt{\widehat{n}}} \widehat{\mathrm{div}} \left(\delta_n^2 \widehat{\nabla} \sqrt{\widehat{n}} \right).$$

Similarly, the Poisson equation $(1)_3$ reduces to

$$-\widehat{\mathrm{div}}(\lambda^2 \widehat{\nabla} \widehat{\varphi}) = \widehat{p} - \widehat{n} + \widehat{D},$$

while the continuity equation $(1)_1$ reads

$$-\widehat{\mathrm{div}}(\widehat{\mu_n}(\widehat{n} \widehat{\nabla}(\widehat{\varphi} + \widehat{G_n}) - \beta \widehat{\nabla} \widehat{n})) = \widehat{U}.$$

As for the Schrödinger equation (8), we set

$$\eta_n^2 = \frac{\hbar^2}{2 q m_n^* \overline{x}^2 \overline{\varphi}},$$

and we obtain

$$-\eta_n^2 \widehat{\mathrm{div}} \widehat{\nabla} \widehat{\psi} + \widehat{E_{\mathrm{c}}} \widehat{\psi} = \widehat{E} \widehat{\psi}.$$

Furthermore, we define the non-dimensional coefficient

$$\theta = \frac{n_{\mathrm{int}}}{\overline{n}}, \tag{24}$$

which will be useful in the definition of the boundary conditions.

In Table 1, we summarize the non-dimensional coefficients for electrons and holes appearing in the QDD and SPDD equations, providing their numerical values in the case of a nanoscale device.

Table 1
Non-dimensional coefficients in QDD and SPDD equations for a device with $\overline{n} = 10^{24}\ \mathrm{m}^{-3}$ and $\overline{x} = 10^{-7}$ m

Symbol	Meaning	Value
η_n^2	$\hbar^2/(2qm_n^* \overline{\varphi x}^2)$	1.46×10^{-5}
η_p^2	$\hbar^2/(2qm_p^* \overline{\varphi x}^2)$	4.25×10^{-5}
β	$V_{\mathrm{th}}/(\overline{\varphi})$	1
δ_n^2	$\hbar^2/(6qm_n^* \overline{x}^2 \overline{\varphi})$	1.88×10^{-4}
δ_p^2	$\hbar^2/(6qm_p^* \overline{x}^2 \overline{\varphi})$	5.48×10^{-4}
λ^2	$(\overline{\varphi}\varepsilon)/(q\overline{x}^2\overline{n})$	1.7×10^{-3}
θ	$n_{\mathrm{int}}/\overline{n}$	6.1401×10^{-9}

3. Unified framework for quantum-corrected DD models

We have shown in the preceeding sections that the QDD and SPDD models can be regarded as generalizations of the classical DD model, only differing by the choice of the constitutive relation for the quantum corrections G_n and G_p to the electric potential. In this section, we will present a unified framework for quantum-corrected DD models which includes both QDD and SPDD as special cases. These latter models can be derived by an appropriate definition of the constitutive relations for the quantum correction potentials G_n and G_p. For sake of completeness, we consider both electron and hole contributions to charge transport. For ease of notation, we will use henceforth for any scaled quantity the same symbol as in the unscaled case.

3.1. The unified form of quantum-corrected DD models

All of the three models discussed in Section 2 can be written in the following unified form:

$$\begin{cases} -\mathrm{div}(\lambda^2\nabla\varphi) + n - p - D = 0, \\ -\mathrm{div}(\mu_n(\nabla n - n\nabla(\varphi + G_n))) = -U, \\ -\mathrm{div}(\mu_p(\nabla p + p\nabla(\varphi + G_p))) = -U, \\ n = \exp((\varphi + G_n) - \varphi_n), \\ p = \exp(\varphi_p - (\varphi + G_p)), \\ G_n = G_n(\varphi, \varphi_n, n), \\ G_p = G_p(\varphi, \varphi_p, p). \end{cases} \tag{25}$$

In (25), the quantity U represents the scaled net recombination rate in the semiconductor and is a function of the carrier densities n, p and quasi-Fermi potentials φ_n and φ_p. A constitutive relation for this term in quantum-corrected models is not well established. In [23], the authors propose the following formal expression:

$$U = \frac{1}{a_0 + a_1 n + a_2 p}(\exp(\varphi_p - \varphi_n) - a_3), \tag{26}$$

where a_0, a_1, a_2 and a_3 are positive constants which should be chosen in such a way that U vanishes at equilibrium. An example of a recombination model which fits (26) and respects the latter condition is proposed in [24] as a QDD extension of the classical Shockley–Read–Hall theory [13]. As the focus of this paper is on comparing different choices for the quantum correction terms, in what follows we will neglect recombination phenomena and set $U = 0$.

Each particular model is characterized by the constitutive relations $(25)_6$ and $(25)_7$ for the correction terms G_n and G_p.

The DD model corresponds to setting

$$\begin{cases} G_n^{\mathrm{DD}}(\varphi, \varphi_n, n) = 0, \\ G_p^{\mathrm{DD}}(\varphi, \varphi_p, p) = 0, \end{cases} \tag{27}$$

while the QDD model corresponds to setting

$$\begin{cases} G_n^{\mathrm{QDD}}(\varphi, \varphi_n, n) = \delta_n^2 \dfrac{\mathrm{div}(\nabla\sqrt{n})}{\sqrt{n}}, \\ G_p^{\mathrm{QDD}}(\varphi, \varphi_p, p) = -\delta_p^2 \dfrac{\mathrm{div}(\nabla\sqrt{p})}{\sqrt{p}}. \end{cases} \tag{28}$$

Getting G_n from $(25)_4$ and G_p from $(25)_5$, and substituting them into (27) yields the following equivalent relations for computing the Bohm potentials, which are more amenable to numerical implementation

$$\begin{cases} -\mathrm{div}\left(\delta_n^2\nabla\sqrt{n}\right)+\sqrt{n}(\varphi_n-\varphi+\ln(n))=0,\\ -\mathrm{div}\left(\delta_p^2\nabla\sqrt{p}\right)+\sqrt{p}(-\varphi_p+\varphi+\ln(p))=0,\\ G_n^{\mathrm{QDD}}=(\varphi_n-\varphi+\ln(n)),\\ G_p^{\mathrm{QDD}}=(\varphi_p-\varphi-\ln(p)). \end{cases} \tag{29}$$

Finally, denoting by $\{E_{ni}\}$, $\{\psi_{ni}\}$, the eigenvalues and the corresponding eigenfunctions of the Hamiltonian $H_n=[-\eta_n^2\Delta+E_c]$ for electrons, and by $\{E_{pi}\}$, $\{\psi_{ni}\}$ the eigenvalues and eigenfunctions of $H_p=[-\eta_p^2\Delta-E_v]$ for holes, the SPDD model corresponds to setting

$$\begin{cases} G_n^{\mathrm{SPDD}}(\varphi,\varphi_n,n)=\begin{cases}\varphi_n-\varphi+\ln(\sum_i|\psi_{n_i}|^2 f(\varphi_n,E_{n_i})) & \text{in } \Omega_Q,\\ 0 & \text{in } \Omega_{\mathrm{cl}},\end{cases}\\ G_p^{\mathrm{SPDD}}(\varphi,\varphi_p,p)=\begin{cases}\varphi_p-\varphi-\ln(\sum_i|\psi_{p_i}|^2 f(\varphi_p,E_{p_i})) & \text{in } \Omega_Q\\ 0 & \text{in } \Omega_{\mathrm{cl}},\end{cases} \end{cases} \tag{30}$$

which can be cast in the following form more similar to (29):

$$\begin{cases} n=\begin{cases}2\sum_i|\psi_{n_i}|^2 f(\varphi_n,E_{n_i}) & \text{in } \Omega_Q,\\ \exp(\varphi-\varphi_n) & \text{in } \Omega_{\mathrm{cl}},\end{cases}\\ p=\begin{cases}2\sum_i|\psi_{p_i}|^2 f(\varphi_p,E_{p_i}) & \text{in } \Omega_Q,\\ \exp(\varphi_p-\varphi) & \text{in } \Omega_{\mathrm{cl}},\end{cases}\\ G_n^{\mathrm{SPDD}}=(\varphi_n-\varphi+\ln(n)),\\ G_p^{\mathrm{SPDD}}=(\varphi_p-\varphi-\ln(p)). \end{cases} \tag{31}$$

As already described in Sections 2.2 and 2.3, both QDD and SPDD systems modify the classical DD model by introducing a certain amount of non-locality in the relation that links the carrier densities to the electrical and quasi-Fermi potentials. More precisely, in the QDD model this non-locality effect is obtained through a dependance of φ_n on the gradient of the concentrations, while in the SPDD model it is obtained through a more detailed physical description of the density of states. A cross-validation and mutual comparison among the three models discussed in this section will be the object of thorough investigation in the numerical experiments shown in Section 7.

3.2. Boundary conditions

In this section, we define the proper boundary conditions for the QCDD models introduced in Section 3.1.

In the case of the QDD model, we let $\Omega\subseteq(0,1)^d$, $d=1, 2, 3$, be the scaled device computational domain, with $\Omega_{\mathrm{Si}}\subset\Omega$ and $\Omega_{\mathrm{OX}}\equiv\Omega\setminus\Omega_{\mathrm{Si}}$ (see Fig. 2). Moreover, let $\Gamma_{\mathrm{Si}}\equiv\partial\Omega_{\mathrm{Si}}$ be the boundary of Ω_{Si}, $\Gamma_{\mathrm{OX}}\equiv\partial\Omega_{\mathrm{OX}}$ the boundary of Ω_{OX} and $\Gamma_I\equiv\Gamma_{\mathrm{OX}}\cap\Gamma_{\mathrm{Si}}$ the interface between Ω_{Si} and Ω_{OX}, ν_{OX} and ν_{Si} being the outward unit normal vectors to Γ_{OX} and Γ_{Si}, respectively. These latter are divided into three disjoint subsets in such a way that $\Gamma_{\mathrm{OX}}\equiv\Gamma_{\mathrm{OX_D}}\cup\Gamma_{\mathrm{OX_N}}\cup\Gamma_{\mathrm{I}}$ and $\Gamma_{\mathrm{Si}}\equiv\Gamma_{\mathrm{Si_D}}\cup\Gamma_{\mathrm{Si_N}}\cup\Gamma_{\mathrm{I}}$. The subsets $\Gamma_{\mathrm{OX_D}}$ and $\Gamma_{\mathrm{Si_D}}$ model the ohmic contacts, where Dirichlet boundary conditions are given for carrier densities and potentials, while $\Gamma_{\mathrm{OX_N}}$ and

$\Gamma_{\mathrm{Si_N}}$ are the remaining part of the boundary where Neumann conditions are enforced. More precisely, the boundary conditions for the QDD model are as follows:

- at the Ohmic contacts a Dirichlet boundary condition is given for both the carrier concentrations and for the electric and quasi-Fermi potentials

$$\begin{cases} n|_{\Gamma_{\mathrm{Si_D}}} = n_{\mathrm{D}}, \quad p|_{\Gamma_{\mathrm{Si_D}}} = p_{\mathrm{D}}, \\ \varphi|_{\Gamma_{\mathrm{Si_D}}} = \varphi_{\mathrm{Si_D}}, \quad \varphi_n|_{\Gamma_{\mathrm{Si_D}}} = \varphi_{n_{\mathrm{D}}}, \quad \varphi_p|_{\Gamma_{\mathrm{Si_D}}} = \varphi_{p_{\mathrm{D}}}, \quad \varphi|_{\Gamma_{\mathrm{OX_D}}} = \varphi_{\mathrm{OX_D}}. \end{cases} \tag{32}$$

  Furthermore, we assume that thermal equilibrium conditions and charge neutrality hold at the contacts, and that quantum effects are negligible so that

$$\begin{cases} p_{\mathrm{D}} n_{\mathrm{D}} = \theta^2, \\ n_{\mathrm{D}} - p_{\mathrm{D}} - D|_{\Gamma_{\mathrm{Si_D}}} = 0, \\ \varphi_n|_{\Gamma_{\mathrm{Si_D}}} = \varphi_a - \ln(\theta), \\ \varphi_p|_{\Gamma_{\mathrm{Si_D}}} = \varphi_a + \ln(\theta), \\ \varphi|_{\Gamma_{\mathrm{Si_D}}} = \varphi_a - \ln(\theta) + \ln(n_{\mathrm{D}}), \end{cases} \tag{33}$$

  where φ_a is the scaled external applied potential at the ohmic contacts;
- at the artificial boundaries $\Gamma_{\mathrm{OX_N}}$ and $\Gamma_{\mathrm{Si_N}}$, the normal components of the current densities and of the electric potential and carrier density gradients are set equal to zero

$$\begin{cases} J_n \cdot \nu_{\mathrm{Si}}|_{\Gamma_{\mathrm{Si_N}}} = 0, \quad J_p \cdot \nu_{\mathrm{Si}}|_{\Gamma_{\mathrm{Si_N}}} = 0, \\ \nabla\varphi \cdot \nu_{\mathrm{Si}}|_{\Gamma_{\mathrm{Si_N}}} = 0, \quad \nabla\varphi \cdot \nu_{\mathrm{OX}}|_{\Gamma_{\mathrm{OX_N}}} = 0, \\ \nabla n \cdot \nu_{\mathrm{Si}}|_{\Gamma_{\mathrm{Si_N}}} = 0, \quad \nabla p \cdot \nu_{\mathrm{Si}}|_{\Gamma_{\mathrm{Si_N}}} = 0; \end{cases} \tag{34}$$

- at the boundary interface between oxide and semiconductor, imposing a proper set of boundary conditions is a more involved and delicate issue. On the one hand, in a quantum model, one would expect the carrier densities to become very small in the vicinity of the very high potential barrier given by the gate oxide, and in the limit of an infinite barrier one should predict that both carrier densities tend to zero. On the other hand, the condition of zero free charge carriers at the interface is incompatible with the modified Maxwell–Boltzmann statistics for electrons and holes, because it would require G_n and G_p to tend to

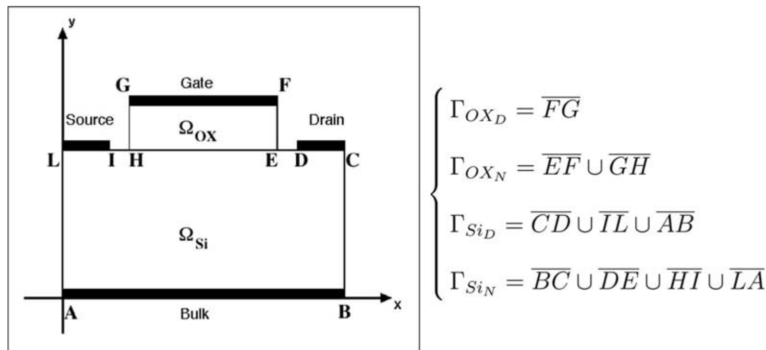

Fig. 2. Boundary subdivision for a two-dimensional cross-section of a MOSFET device.

$-\infty$ and $+\infty$, respectively. To circumvent this problem, one can set the interface densities equal to a small but non-zero value $c_{\rm I}$, which might be estimated by a priori 1D computations with a model including tunneling of free carriers through the oxide barrier [24]. In the numerical computations presented in Section 7, we have imposed a value of 10^{-2} m^{-3} for the charge densities at the interface, which corresponds to a non-dimensional concentration $c_{\rm I} = 10^{-27}$. Furthermore, at the interface we impose the normal component of the current densities to vanish and the normal component of the electric displacement vector to be continuous, i.e.,

$$\begin{cases} n|_{\Gamma_{\rm I}} = p|_{\Gamma_{\rm I}} = c_{\rm I}, \\ J_n \cdot \nu_{\rm Si}|_{\Gamma_{\rm I}} = 0, \quad J_p \cdot \nu_{\rm Si}|_{\Gamma_{\rm I}} = 0, \\ \left[\lambda^2 \nabla\varphi \cdot \nu_{\rm Si}\right]_{\Gamma_{\rm I}} = 0, \end{cases} \tag{35}$$

where $[f]_\gamma$ denotes the jump of a function f across the $d-1$-dimensional manifold γ.

In the case of the SPDD model, the mathematical formulation of the boundary-value problem requires a further subdivision of Ω as anticipated in Section 2.2. With this purpose, let $\Omega_Q \subset \Omega$ be the device subdomain in which we expect (by a priori physical considerations) the quantum effects to be relevant. Let $\Omega_{\rm cl} \equiv \Omega \backslash \Omega_Q$, and let $\Omega_{\rm Schr}$ be a suitably chosen subdomain such that $\Omega \supseteq \Omega_{\rm Schr} \;\; \Omega_Q$ (see Fig. 1). The choice of $\Omega_{\rm Schr}$ is the result of a careful trade-off. On the one hand, $\Omega_{\rm Schr}$ should be large enough to ensure that the closed boundary conditions (11) for Eq. (8) do not significantly affect the quantum charge distribution. On the other hand, the choice of a too wide $\Omega_{\rm Schr}$ can greatly increase the computational cost of the eigenvalue problem. Conditions (32)–(35) still hold for the SPDD model. In addition, homogeneous boundary conditions are enforced on the wavefunctions on $\partial\Omega_{\rm Schr}$ as in (11).

3.3. Recovering the classical limit

As we have presented both the QDD and SPDD models essentially as *perturbations* of the classical DD model, it is natural to briefly address the issue of how those two former models reduce to the latter in the classical limit.

In the case of the QDD model, we start noticing that, by formally setting $\delta_n^2 = \delta_p^2 = 0$, we immediately recover the DD model. The mathematical analysis of the limiting behaviour of the QDD model has been carried out in [23], where it was shown that the solution of the QDD model converges to the solution of the classical DD model as $\delta_n^2, \delta_p^2 \to 0$. From a quantitative point of view, letting

$$L_{\rm cl}^{\rm QDD} = \sqrt{\frac{\hbar^2}{6 m_n^* k_{\rm b} T}},$$

we note that $\delta_n^2 \ll 1$ when $\bar{x} \gg L_{\rm cl}^{\rm QDD}$, so that, at room temperature ($T = 300$ K), the classical limit is attained when the characteristic device length $\bar{x}$ is much larger than $L_{\rm cl}^{QDD} = 1.4$ nm.

For the SPDD case, it is well known (see for example [25]) that for a *very large* crystal and near equilibrium the summation in (9) reduces to the expression in (4) so that $\gamma_n = 1$. As an example, we consider a 1D case with $\varphi = 0$, for which the following exact expression of the energy eigenvalues is available

$$E_i = E_{\rm c} + \eta_n^2 \pi^2 i^2, \quad i = 1, 2, \ldots \tag{36}$$

Introducing the previous relation into (19) with $d = 1$ yields

$$g(E) = \langle g(E,x)\rangle = \sum_{i=1}^{\infty} H(E - E_i)\langle|\ \psi_i^2(x)\ |\rangle g_2(E) = \sum_{i<\frac{1}{\pi\eta_n}\sqrt{E-E_c}} H(E - E_i) g_2(E)$$
$$= \mathrm{int}\left(\frac{1}{\pi\eta_n}\sqrt{E - E_c}\right) g_2(E), \tag{37}$$

where $\langle f(x)\rangle$ is the integral mean value of a function f over the device domain and $\forall z \in \mathbb{R}$, $\mathrm{int}(z)$ stands for the greatest integer smaller than z. Thus, for

$$\frac{1}{\pi\eta_n}\sqrt{E - E_c} \gg 1, \tag{38}$$

one gets

$$g(E) \simeq \left(\frac{1}{\pi\eta_n}\sqrt{E - E_c}\right) g_2(E) = g_3(E),$$

which is the same density of states function used in (4). Letting

$$L_{\mathrm{cl}}^{\mathrm{SPDD}} = \sqrt{\frac{\pi^2\hbar^2}{2m_n^* k_b T}},$$

we note that condition (38) is attained for energies $E > E_c$ if $1 \ll \frac{1}{\pi\eta_n} = \bar{x}/L_{\mathrm{cl}}^{\mathrm{SPDD}}$, i.e., for $\bar{x} \gg L_{\mathrm{cl}}^{\mathrm{SPDD}} = 7.5$ nm, which is essentially the same condition obtained in the case of the QDD model.

Fig. 3 shows the quantum correction G_n in the channel of a 1D n^+–n–n^+ device for different values of the channel length. Concerning with the top figures, on the x-axis, we indicate the length of the device (in m), while on the y-axis we denote the scaled position along the device. The bottom figures display the quantity G_n at the middle of the channel as a function of the device length. Notice that, as expected, the quantum correction becomes negligible as $\bar{x}$ increases.

3.4. A proper choice for the quantum region

In the example discussed in Section 3.3, we have considered, for sake of simplicity, a spatial average of the *modified density of states* $g(E,x)$, which, for a uniformly doped material, leads to a constant quantum charge density n_q. Actually, even in the case $\varphi = 0$ and D constant, the quantum charge density is *not* constant in Ω_{Schr}, but, if the width $\bar{x}$ of Ω_{Schr} satisfies the condition $\bar{x} \gg L_{\mathrm{cl}}^{\mathrm{SPDD}}$, it is almost flat about the center of the well, where the value of n_q is close to n_{int}, and rapidly goes to zero at the boundaries. For this reason, as described in Section 3.2, one can obtain an almost constant charge density by forcing $\gamma_n = 1$, i.e., $n_q = n_{\mathrm{cl}} = n_{\mathrm{int}}$ in a subregion $\Omega_{\mathrm{Schr}} \backslash \Omega_Q$ of Ω_{Schr}. The choice of the subregion Ω_Q has to be made in such a way that the discontinuity suffered by the charge density n across its boundary $\partial\Omega_Q$ is small. In Fig. 4, the jump $|n_{\mathrm{int}} - n_q|/n_{\mathrm{int}} = |1 - \gamma_n|$ for the case $\varphi = 0$ is plotted as a function of the width $\bar{x}$ of the well. The white and red line show the width of the quantum region for which $|1 - \gamma_n| \leqslant 0.01$ and $|1 - \gamma_n| \leqslant 0.001$, respectively.

Notice that, as $\bar{x}$ increases, the distance from the boundary at which n_q approaches n_{int} decreases very rapidly. To understand this fact, we can use the exact expression of the wavefunctions

$$|\psi_i(x)|^2 = \frac{\sin^2(i\pi x)}{\int_0^1 \sin^2(i\pi x)\,\mathrm{d}x} = 2\sin^2\left(2\pi\frac{x}{\Lambda_i}\right), \quad i = 1, 2, \ldots,$$

where we have introduced the scaled wavelengths $\Lambda_i = 2/i$. It can be easily checked that, by inserting $(20)_2$ in (19) and applying the scaling procedure described in Section 2.4, one gets

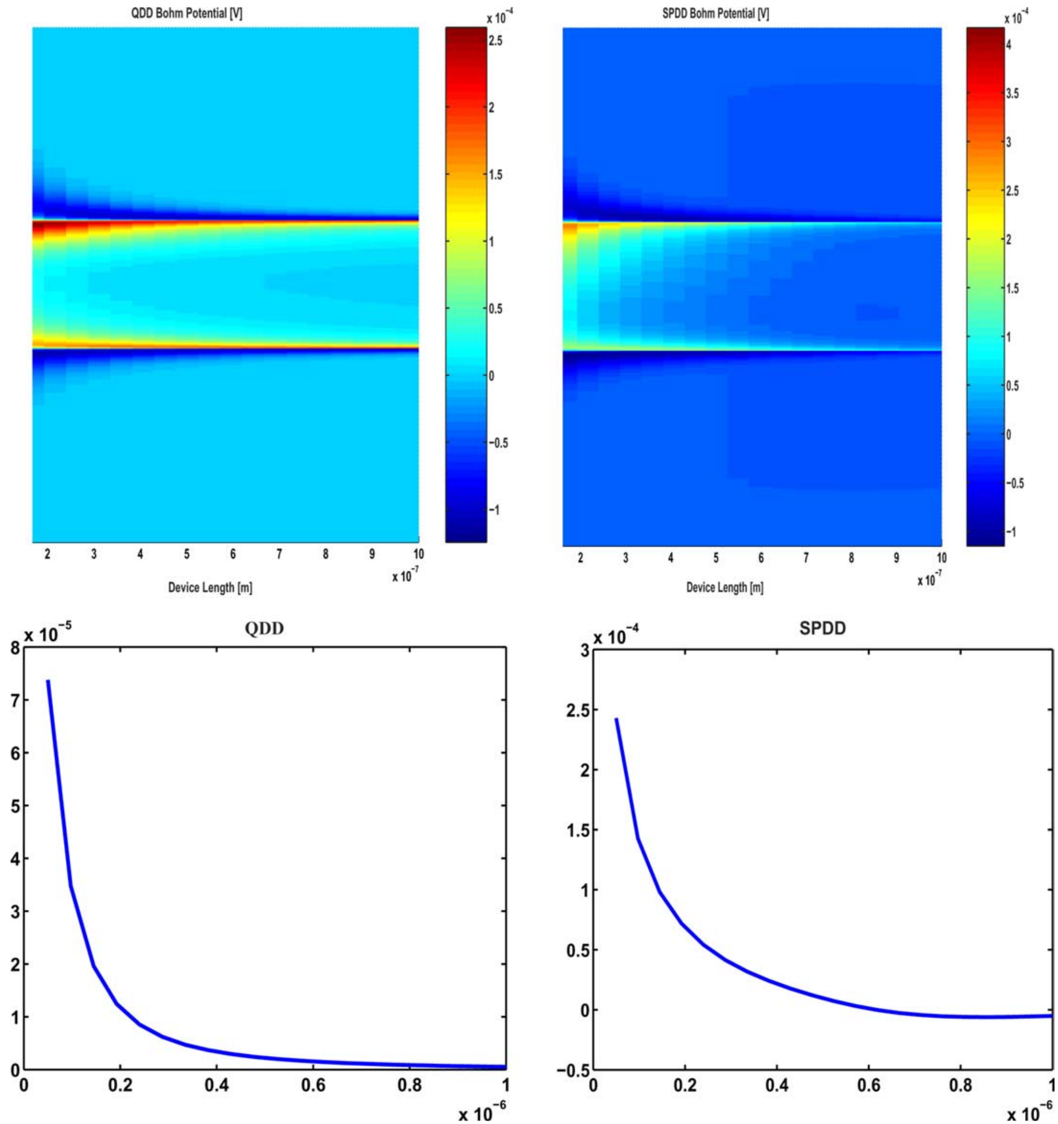

Fig. 3. The quantum correction potential G_n in the channel of a 1D n^+–n–n^+ device as a function of the channel length (figures on the left are obtained by QDD simulations and those on the right by SPDD simulations).

$$n_q = 2\sum_i \zeta_n \sin^2\left(2\pi\frac{x}{\Lambda_i}\right)\exp(-E_i - \varphi_n), \quad \zeta_n = \frac{m_n^* \pi k_b T}{\bar{x}\hbar^2}. \tag{39}$$

From (39) and (36) it becomes evident that, increasing the well width $\bar{x}$ the energy E_i associated with a given wavefunction ψ_i of wavelength Λ_i (with i fixed), decreases rapidly so that its contribution to the summation in (39) becomes more relevant. Therefore, while for small $\bar{x}$ low wavelength functions are strongly damped, as $\bar{x}$ becomes larger they become more important. This can be also observed in Fig. 5, where the product $f(E)\langle g(E)\rangle$ for DD and 1D SPPD models is compared for different values of $\bar{x}$.

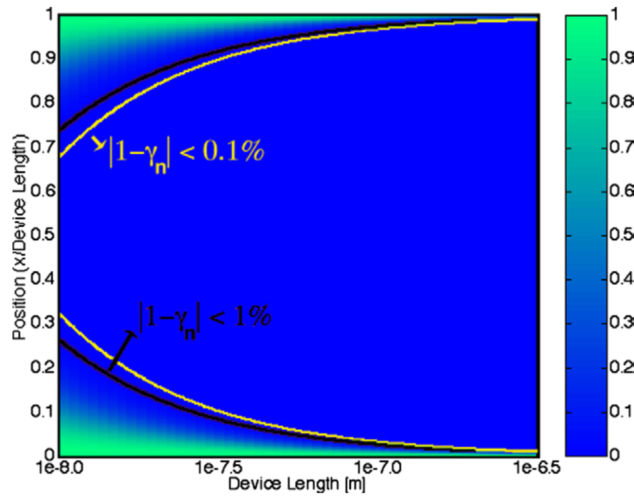

Fig. 4. Relative difference $|1-\gamma_n|$ between quantum and classic charge in an intrinsic Si sample with $\varphi = 0$, as a function of the width $\bar{x}$.

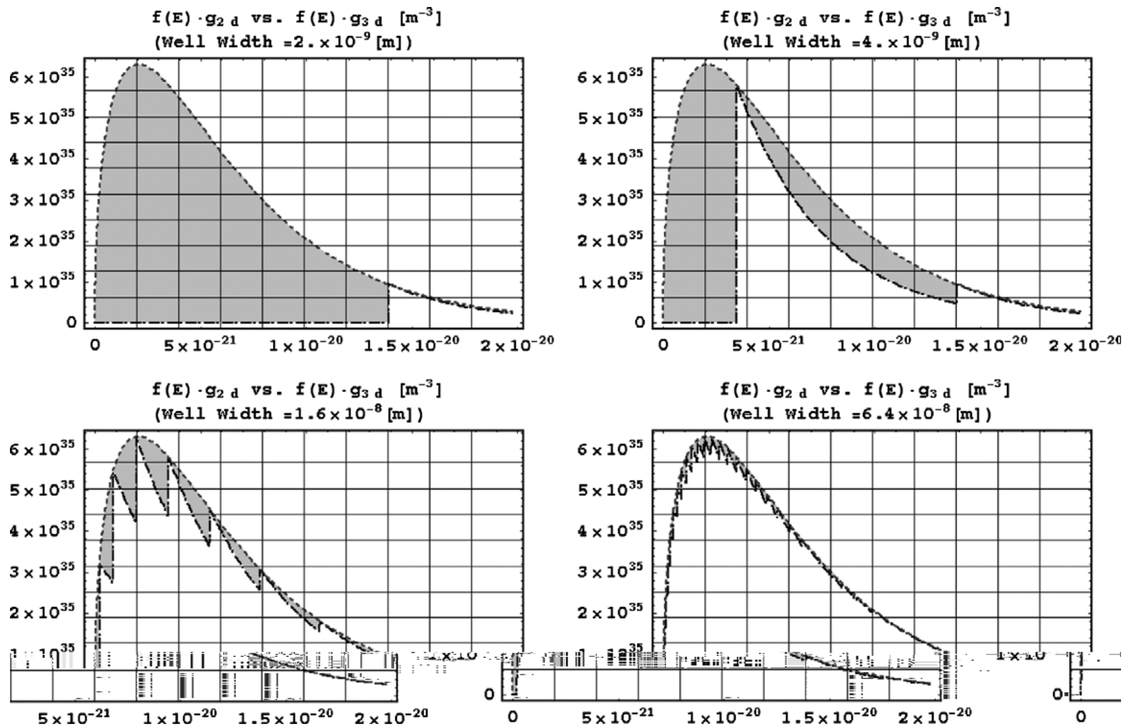

Fig. 5. Comparison of the product $f(E)\langle g(E)\rangle$ for DD and 1D SPPD models for different values of $\bar{x}$ (the gray area is the difference between $\langle n_q\rangle$ and n_{cl}).

4. Functional iterations

In this section, we discuss the decoupled functional iteration that will be used to solve the unified quantum-corrected model (25). This model constitutes a highly nonlinear system of boundary-value problems.

The choice of the Newton method for the linearization of (25), although attractive due to the quadratic convergence, is affected by several drawbacks. First, its high computational effort (in terms of memory storage, ill-conditioning of the Jacobian matrix and linear system solution at each step of the iterative procedure). Second, the need of availing (or constructing) a good initial guess to fully exploit second-order convergence. Third, the strong request on computing strictly positive carrier densities, which is a priori impossible to ensure when dealing with a fully coupled solution approach.

To overcome these limitations, a decoupled algorithm, well-known as *Gummel map* [26], is usually employed in the numerical approximation of the Drift-Diffusion model. The Gummel map can be interpreted as a nonlinear block Gauss–Seidel functional iteration, which consists in the successive solution of the nonlinear Poisson's equation $(25)_1$ coupled with $(25)_4$ and $(25)_5$, and of the two linearized continuity Eqs. $(25)_2$ and $(25)_3$. The method typically exhibits a rapid convergence and a good robustness with respect to the choice of the initial guess. Moreover, stable and monotone discretization schemes can be applied to the numerical solution of the linearized continuity equations, which allow to compute strictly positive carrier concentrations with appropriate current conservation properties (see [27–31]). A detailed study of the Gummel map can be found in the references [32,33,14,34,20,15].

In the following, we propose a generalization of the Gummel decoupled algorithm to the iterative solution of the quantum-corrected model (25).

With this aim, we construct a functional iteration

$$(\varphi_n, \varphi_p) \rightarrow T(\varphi_n, \varphi_p) \tag{40}$$

mapping a given couple (φ_n,φ_p) into $T(\varphi_n,\varphi_p)$, as follows:

(Step 1) compute

$$(\varphi, G_n, G_p) = \Phi(\varphi_n, \varphi_p) \tag{41}$$

by iteratively solving the subsystem $(25)_1$, $(25)_6$ and $(25)_7$, where the generalized Maxwell–Boltzmann statistics $(25)_4$–$(25)_5$ is used for n and p in $(25)_1$;

(Step 2) compute

$$\varphi_n = \Phi_n(\varphi, G_n) \tag{42}$$

by solving the linearized electron continuity equation $(25)_2$ with respect to the variable n and then setting

$$\varphi_n = (\varphi + G_n) - \ln(n);$$

(Step 3) compute

$$\varphi_p = \Phi_p(\varphi, G_p) \tag{43}$$

by solving the linearized hole continuity Eq. $(25)_3$ with respect to the variable p and then setting

$$\varphi_p = (\varphi + G_p) + \ln(p).$$

The boundary conditions for each differential subproblem involved in (40) are the same as described in Section 3.2. During the execution of Step 1, the functions φ_n and φ_p are given and kept fixed, while in Steps 2 and 3, the functions φ, G_n and G_p resulting from Step 1 are plugged into (42) and (43) and kept fixed, as in a standard Gauss–Seidel procedure (see [15] for a thorough discussion of this latter aspect). Special attention must be paid to the treatment of the net recombination rate in the solution of the linearized electron and hole continuity equations. As a matter of fact, the expressions (42) and (43) have rigorous validity only if, as is the case for this paper, the recombination term is neglected, otherwise one should also linearize U with respect to φ_n and φ_p. A possible strategy could be to resort to the lagging procedure described in [15],

although a more thorough investigation is needed about this issue (see [24] for some numerical results in the case of the QDD model). The above decoupled iteration consists of an *outer loop* (Steps 1, 2 and 3) and of an *inner loop* (Step 1). This latter loop can be regarded as a consistent generalization of the solution of the nonlinear Poisson equation $(25)_1$, where the algebraic Maxwell–Boltzmann relations $(25)_4$–$(25)_5$ for n and p are replaced by the solution of the differential problems (29) (in the QDD case) and (30) (in the SPDD case).

In the case of the DD model we have $G_n = G_p = 0$, and the algorithm (40) reduces to the classical Gummel map.

In the case of the QDD model, the inner loop (41) is defined by the following fixed point map:

(Step 1.1) Initialize G_n and G_p and compute

$$\varphi = \Phi_\varphi(\varphi_n, \varphi_p, G_n, G_p) \tag{44}$$

by solving with respect to φ the nonlinear Poisson equation $(25)_1$ coupled with $(25)_4$-$(25)_5$;

(Step 1.2) compute

$$G_n = G_n(\varphi, \varphi_n) \tag{45}$$

by solving the nonlinear equation $(29)_1$ with respect to the variable $\sqrt{n}$, and setting

$$G_n = \varphi_n - \varphi + \ln(n);$$

(Step 1.3) compute

$$G_p = G_p(\varphi, \varphi_p) \tag{46}$$

by solving the nonlinear equation $(29)_2$ with respect to the variable $\sqrt{p}$, and setting

$$G_p = \varphi_p - \varphi - \ln(p).$$

In the case of the SPDD model, Steps 1.2 and 1.3 are replaced by the computation of the eigenvalues and eigenvectors for the Hamiltonian operators H_n and H_p (see Section 3.1) and by the computation of G_n and G_p through relations (30). In the simulation of n-type devices, such as an n-channel MOS transistor, where current transport is mainly due to electrons, the solution of the SPDD equations for the hole carriers can be conveniently replaced by the solution of the corresponding DD equations. This allows to significantly reduce the high computational effort of the two-carrier SPDD model, maintaining at the same time a reasonable physical accuracy of the model. This simplified SPDD model is discussed in Sections 6.2 and 7. An analogous procedure can be obviously carried out when simulating p-type devices, by neglecting quantum transport effects due to electron contribution.

The convergence analysis of the generalized Gummel map (40) and of its corresponding finite element discretization (discussed in Section 5) will be object of a forthcoming paper. At the present stage, the main difficulty is to deal with Steps 1.2 and 1.3, that are the major (nonlinear) modification to the standard Gummel map used in the DD model. In particular, it is crucial to establish that a maximum principle is satisfied by the solutions $\sqrt{n}$ and $\sqrt{p}$ of $(29)_1$ and $(29)_2$, proving as a consequence that n and p are positive quantities through the functional iteration (40). The issue of positiveness of the carrier densities will be addressed in more detail in Section 5, where a (properly) damped Newton method will be adopted to construct iteratively a sequence of positive approximants to the solutions of the nonlinear equations $(29)_1$ and $(29)_2$.

Other functional block nonlinear iterations for the QDD model, related to (40), were proposed and analyzed in [16] and [17]. In the first reference, a generalized Gummel map is considered which consists of the successive solution of two *nonlinear* blocks, each one through a *fully coupled* approach. The first block amounts to carrying out Step 1 through the coupled solution of Eqs. $(25)_1$, $(29)_1$ and $(29)_2$ with respect to the variables φ, n and p, while the second block amounts to carrying out Steps 2 and 3 through the coupled

solution of Eqs. $(25)_2$ and $(25)_3$ with respect to the dependent variables. In the second reference, a unipolar 1D QDD model (for electrons) is dealt with. The structure of the proposed iterative mapping is similar to that in [16], the main difference being the use of the (generalized) Slotboom variable $u_n = n\mathrm{e}^{-(\varphi+G_n)}$, instead of the quasi-Fermi level φ_n, in the solution of the (linear) electron continuity equation.

5. Finite element discretization

In this section, we describe the finite element discretization of the differential subproblems involved in the generalized Gummel map introduced in Section 4.

5.1. Approximation of the QDD model

Step 1.1. amounts to solving the nonlinear Poisson equation

$$-\mathrm{div}(\lambda^2\nabla\varphi) + \exp((\varphi + G_n) - \varphi_n) - \exp(\varphi_p - (\varphi + G_p)) = 0 \quad \text{in } \Omega \tag{47}$$

with respect to the unknown φ, for given φ_n, φ_p, G_n, G_p, and subject to mixed Dirichlet–Neumann boundary conditions as discussed in Section 3.2. A damped Newton method is adopted for the linearization of (47), and the corresponding discretization is carried out using piecewise linear continuous finite elements as done in [35] in the case of the DD model. The resulting linear algebraic system to be solved at each step of the damped Newton iteration is characterized by having a symmetric positive definite and diagonally dominant coefficient matrix, provided that a suitable lumping procedure is employed to treat the zeroth order term in (47). The solution of the system is then efficiently performed by a preconditioned Conjugate Gradient method.

The iterative solution and the finite element discretization of Steps 1.2 and 1.3 is a critical issue, because of the need of maintaining positive solutions for the (square root) of the carrier densities. With this aim, we have modified the standard Newton procedure by introducing a relaxation parameter, to be chosen in such a way that the linearized boundary value problem (and the corresponding finite element approximation) enjoys a maximum principle. This is a sufficient condition to ensure positivity of the computed approximate carrier concentrations. We describe the novel methodology in the case of the Eq. $(29)_1$ for electrons, a completely similar treatment being used to solve the Eq. $(29)_2$ for holes.

Let $w = \sqrt{n}$; then, the standard Newton iteration applied to problem $(29)_1$ reads: given an initial guess $w^{(0)}$, at each step $k \geqslant 0$ solve for the unknown $w^{(k+1)}$ the *linear* elliptic problem

$$-\mathrm{div}(\delta_n^2\nabla w^{(k+1)}) + (\varphi_n - \varphi + 2\ln(w^{(k)}))w^{(k+1)} + 2w^{(k+1)} = 2w^{(k)} \quad \text{in } \Omega_{\mathrm{Si}} \tag{48}$$

subject to mixed Dirichlet–Neumann boundary conditions (for the variable $\sqrt{n}$ instead of n) as discussed in Section 3.2. If the following condition is satisfied

$$(\varphi_n - \varphi + 2\ln(w^{(k)})) + 2 \geqslant 0 \quad \text{in } \Omega_{\mathrm{Si}} \tag{49}$$

then (48) defines a linear continuous mapping from the function space $H^1_{\Gamma_{\mathrm{Si_D}}}(\Omega_{\mathrm{Si}})$ into itself, where $H^1_{\Gamma_{\mathrm{Si_D}}}(\Omega_{\mathrm{Si}})$ is the Sobolev space of order 1 accounting for non-homogeneous boundary conditions associated with $w^{(k+1)}$ (see [36,37] for a definition of this functional space and a discussion of its properties). Moreover, it can be checked that a maximum principle holds for the weak solution of (48), which implies that $w^{(k+1)} > 0$ in Ω_{Si} provided that $w^{(k)}$, $k \geqslant 0$, and the boundary data are positive.

In the case where condition (49) is not satisfied, we can still ensure positivity of $w^{(k+1)}$ by introducing a *damping parameter* t_k, $0 < t_k \leqslant 1$ in such a way that

$$t_k < \frac{2}{|\inf\limits_{\Omega_{\rm Si}}(\varphi_n - \varphi + 2\ln(w^{(k)}))|},$$

and modifying (48) as follows

$$-\mathrm{div}(\delta_n^2 \nabla w^{(k+1)}) + (\varphi_n - \varphi + 2\ln(w^{(k)}))w^{(k+1)} + \frac{2}{t_k} w^{(k+1)} = \frac{2}{t_k} w^{(k)} \quad \text{in } \Omega_{\rm Si}, \tag{50}$$

where the use of t_k on the right-hand side is required to maintain the consistency of the iterative procedure.

Problem (50) defines a modified fixed-point iteration with relaxation which, albeit not enjoying in general the quadratic convergence of Newton method, ensures that the square root of the electron carrier concentration remains positive at each step. The finite element discretization of problem (50) with piecewise linear continuous finite elements proceeds as in the case of the nonlinear Poisson equation (47). In particular, by lumping the mass matrix corresponding to the zeroth-order term in (50), the resulting linear algebraic system is characterized by having a symmetric positive definite and diagonally dominant coefficient matrix, with positive diagonal entries and non-positive off-diagonal entries, and by a positive right-hand side. These properties ensure that the coefficient matrix is an M-matrix and that the positivity property of w still holds on the discrete level.

Once Step 1 has been solved as described above, the numerical approximation of the remaining linear continuity equations (Steps 2 and 3) is carried out by a finite element scheme using piecewise linear continuous elements with exponential fitting as done in [31] in the case of the DD model. This approach is a consistent generalization of the classical Scharfetter-Gummel difference scheme to Delaunay-type triangular decompositions of the device domain. It has the advantage of automatically introducing an upwinding treatment of the carrier densities along triangle edges, which in turn ensures that the method satisfies a discrete maximum principle with positive nodal values of the carrier densities n and p (see [38,39] for a thorough discussion of this latter subject).

5.2. Approximation of the SPDD model

As it has been pointed out before, the algorithm used for solving the SPDD system only differs from that for the QDD system by the method used to compute the quantum correction potential. In particular, the solution of the nonlinear elliptic boundary value problem $(29)_1$ is substituted by the computation of the eigenvalues and eigenvectors of the Hamiltonian H_n followed by the summation $(31)_1$. This approach has two main effects which are relevant for the implementation of the numerical algorithm; on the one hand, $(31)_1$ guarantees that the quantum charge is always positive, on the other hand the heavy cost of the eigenvalue/eigenvector problem makes it very convenient to neglect the quantum correction term for one of the carriers, when it is not expected to have a high impact on the value of the currents. For example, in the device simulation presented in Section 7 we have treated the holes using the DD model. To discretize the operator H_n we have used piecewise linear finite elements, and we have solved the discrete eigenvalue/eigenvector problem using Arnoldi's method.

6. Numerical algorithms

In this section, we provide a detailed description of the algorithms emanating from the iterative procedure (40) for the approximate solution of the QDD and SPDD models.

6.1. Detailed description of the algorithm for the QDD model

- Input: $\{\varphi^{(0)}, n^{(0)}, \varphi_n^{(0)}, G_n^{(0)}, p^{(0)}, \varphi_p^{(0)}, G_p^{(0)}, \texttt{toll}, \texttt{kmax}, \texttt{jmax}\}$
- set $\gamma_n^{(0)} = \ln(G_n)^{(0)}, \gamma_p^{(0)} = \ln(G_p^{(0)})$
- `for` $k = 1,\ldots,$ `kmax` (k `is the outer iteration counter`)

(1) `for` $j = 1,\ldots,$ `jmax` (j `is the inner iteration counter`)
  (a) `Solve for` φ(`using a damped Newton method`):

$$-\mathrm{div}(\lambda^2 \nabla\varphi) + \gamma_n^{(k)}{}_{j+1} \exp(\varphi - \varphi_n^{(k)}) - \gamma_p^{(k)}{}_{j+1} \exp(-\varphi + \varphi_p^{(k)}) - D = 0$$

  (b) `set`:

$$\varphi_{j+1}^{(k)} = \varphi$$

  (c) `Solve for` w (`using the modified Newton method`):

$$-\mathrm{div}(\delta_n^2 \nabla w) + w(\varphi_n^{(k)} - \varphi_{j+1}^{(k)} + 2\ln(w)) = 0$$

  (d) `Solve for` v (`using the modified Newton method`):

$$-\mathrm{div}(\delta_p^2 \nabla v) + v(-\varphi_p^{(k)} + \varphi_{j+1}^{(k)} + 2\ln(v)) = 0$$

  (e) `set`:

$$\begin{aligned}
&n_q = w^2, && p_q = v^2 \\
&n_{\mathrm{cl}} = \exp(\varphi_{j+1}^{(k)} - \varphi_n^{(k)}), && p_{\mathrm{cl}} = \exp(\varphi_p^{(k)} - \varphi_{j+1}^{(k)}) \\
&G_n^{(k)}{}_{j+1} = \varphi_n^{(k)} - \varphi_{j+1}^{(k)} + 2\ln(w), && G_p^{(k)}{}_{j+1} = \varphi_p^{(k)} - \varphi_{j+1}^{(k)} - 2\ln(w) \\
&\gamma_n^{(k)}{}_{j+1} = \frac{n_q}{n_{\mathrm{cl}}} = \exp(G_n^{(k)}{}_{j+1}), && \gamma_p^{(k)}{}_{j+1} = \frac{p_q}{p_{\mathrm{cl}}} = \exp(G_p^{(k)})_{j+1} \\
&n_{j+1}^{(k)} = \gamma_n^{(k)}{}_{j+1} n_{\mathrm{cl}}, && p_{j+1}^{(k)} = \gamma_p^{(k)}{}_{j+1} p_{\mathrm{cl}}
\end{aligned}$$

  (f) `if` $\|\varphi_{j+1}^{(k)} - \varphi_j^{(k)}\|_{\infty,\Omega} \leqslant$ `toll set`:

$$\begin{aligned}
&\varphi^{(k+1)} = \varphi_{j+1}^{(k)} \\
&n^{\left(k+\frac{1}{2}\right)} = n_{j+1}^{(k)}, \quad p^{\left(k+\frac{1}{2}\right)} = p_{j+1}^{(k)} \\
&G_n^{(k+1)} = G_n^{(k)}{}_{j+1}, \quad G_p^{(k+1)} = G_p^{(k)}{}_{j+1} \\
&\gamma_n^{(k+1)} = \gamma_n^{(k)}{}_{j+1}, \quad \gamma_p^{(k+1)} = \gamma_p^{(k)}{}_{j+1}
\end{aligned}$$

  `and proceed to step 2`
  (g) `else : repeat steps a)` $\to$ `e)`

(2) `Solve for` n:

$$-\mathrm{div}(\mu_n \nabla n - \mu_n n \nabla(\varphi^{(k+1)} + G_n^{(k+1)})) = 0$$

(3) `set`:

$$n_{k+1} = n, \quad \varphi_n^{(k+1)} = \varphi^{(k+1)} - \ln\left(\frac{n^{(k+1)}}{\gamma_n^{(k+1)}}\right)$$

(4) `Solve for` p:

$$-\mathrm{div}\left(\mu_p \nabla p + \mu_p p \nabla\left(\varphi^{(k+1)} + G_p^{(k+1)}\right)\right) = 0$$

(5) `set`:

$$p^{(k+1)} = p, \quad \varphi_p^{(k+1)} = \varphi^{(k+1)} + \ln\left(\frac{p^{(k+1)}}{\gamma_p^{(k+1)}}\right)$$

(6) `if` $\|\varphi^{(k+1)} - \varphi^{(k)}\|_{\infty,\Omega} \leqslant$ `toll and` $\|\varphi_n^{(k+1)} - \varphi_n^{(k)}\|_{\infty,\Omega_{\mathrm{Si}}} \leqslant$ `toll and` $\|\varphi_p^{(k+1)} - \varphi_p^{(k)}\|_{\infty,\Omega_{\mathrm{Si}}} \leqslant$ `toll exit`
(7) `else : go back to step 1.`

6.2. Detailed description of the algorithm for the SPDD model

- Input: Input: $\{\varphi^{(0)}, n^{(0)}, \varphi_n^{(0)}, p^{(0)}, \varphi_p^{(0)}, \mathtt{toll}, \mathtt{kmax}, \mathtt{jmax}\}$
- `for` $k = 1,\ldots,kmax$ (k `is the outer iteration counter`)

(1) `for` $j = 1,\ldots,jmax$ (j `is the inner iteration counter`)

(a) `Solve for` φ(`using a damped Newton method`):

$$-\mathrm{div}(\lambda^2 \nabla\varphi) + \gamma_{n\ \ j+1}^{(k)} \exp(\varphi - \varphi_n^{(k)}) - \gamma_{p\ \ j+1}^{(k)} \exp(-\varphi + \varphi_p^{(k)}) - D = 0$$

(b) `set`:

$$\varphi_{j+1}^{(k)} = \varphi$$

(c) `Solve for the eigenvalues` $\{E_{ni}\}$ `and eigenvectors` $\{\psi_{ni}\}$:

$$-\mathrm{div}(\eta_n^2 \nabla\psi) + E_{\mathrm{c}}\psi = E\psi$$

(d) `set`:

$$n_q = \sum_i |\ \psi_{ni}|^2 \int_{E_{ni}}^{+\infty} f(E) g_{2D}(E) dE, \quad n_{\mathrm{cl}} = \exp(\varphi_{j+1}^{(k)} - \varphi_n^{(k)})$$

$$\gamma_{j+1}^{(k)} = \begin{cases} \dfrac{n_q}{n_{\mathrm{cl}}} & \text{in } \Omega_Q \\ 1 & \text{in } \Omega_{\mathrm{cl}} \end{cases}$$

$$G_{n\ \ j+1}^{(k)} = \ln(\gamma_{j+1}^{(k)}), \qquad n_{j+1}^{(k)} = \gamma_{j+1}^{(k)} n_{\mathrm{cl}}$$

(f) `if` $\|\varphi_{j+1}^{(k)} - \varphi_j^{(k)}\|_{\infty,\Omega} \leqslant$ `toll set:`

$$\varphi^{(k+1)} = \varphi_{j+1}^{(k)}, \quad n^{(k+\frac{1}{2})} = n_{j+1}^{(k)}$$

$$p^{(k+\frac{1}{2})} = p_{j+1}^{(k)}, \quad G_n^{(k+1)} = G_{n\ \ j+1}^{(k)}$$

$$\gamma^{(k+1)} = \gamma_{j+1}^{(k)}$$

`and proceed to step 2`

(g) `else : repeat steps a)` $\rightarrow$ `e)`

(2) `Solve for` n:

$$-\mathrm{div}(\mu_n \nabla n - \mu_n n \nabla(\varphi^{(k+1)} + G_n^{(k+1)})) = 0$$

(3) set:

$$n^{(k+1)} = n, \quad \varphi_n^{(k+1)} = \varphi^{(k+1)} - \ln\left(\frac{n^{(k+1)}}{\gamma^{(k+1)}}\right)$$

(4) `Solve for` p:

$$-\mathrm{div}(\mu_p \nabla p + \mu_p p \nabla \varphi^{(k+1)}) = 0$$

(5) `set`:

$$p^{(k+1)} = p, \quad \varphi_p^{(k+1)} = \varphi^{(k+1)} + \ln(p^{(k+1)})$$

(6) `if` $\|\varphi^{(k+1)} - \varphi^{(k)}\|_{\infty,\Omega} \leqslant$ `toll and` $\|\varphi_n^{(k+1)} - \varphi_n^{(k)}\|_{\infty,\Omega_{\mathrm{Si}}} \leqslant$ `toll and` $\|\varphi_p^{(k+1)} - \varphi_p^{(k)}\|_{\infty,\Omega_{\mathrm{Si}}} \leqslant$ `toll exit`
(7) `else : go back to step` 1.

7. Numerical results

In this section, we discuss the numerical results concerning the simulation of a nanoscale MOSFET device similar to that studied in [8]. In doing this, we will compare the computational performance and the physical accuracy of the DD, QDD and SPDD models introduced in Section 2. Devices similar to the one considered in this section are expected to be in mass production beyond 2010, as specified by the International Technology Roadmap for Semiconductors [40,41].

The device geometry, the finite element triangulation and the doping profile are shown in Fig. 6. In all the reported figures, the international system of units (SI) is used. The device has a 15 nm long channel with a uniform p-type doping of 2×10^{25} m^{-3}; the drain and source contact regions have a uniform n-type doping of 5×10^{25} m^{-3} and reach 20 nm down into the bulk. The gate oxide is 1.5 nm thick and is 25 nm long, so that it has an overlap of 5 nm with both the source and drain regions. The triangulation consists of 3284 elements and 1726 nodes.

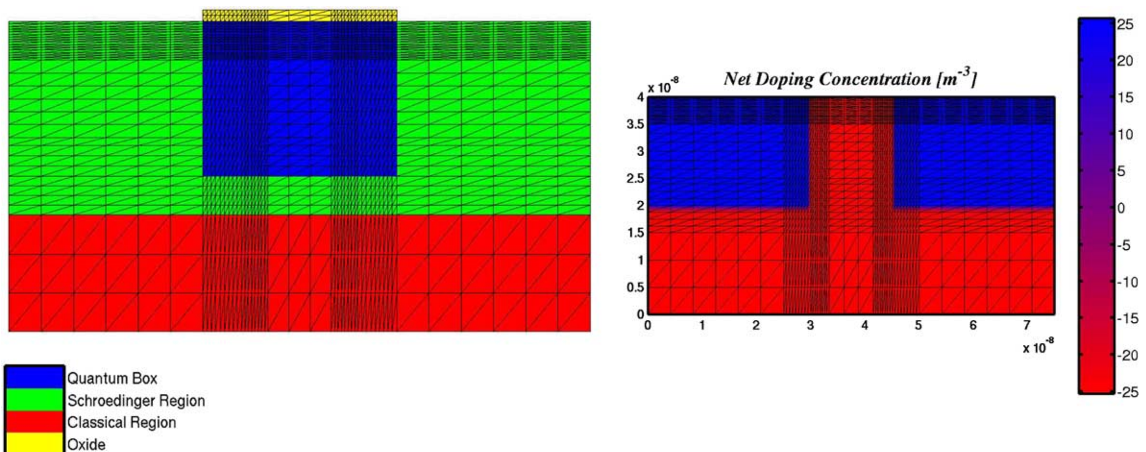

Fig. 6. Left: MOSFET Geometry and finite element triangulation, with highlighting of the different subdomains. Right: net doping profile (log-scale); positive doping is n-type, negative is p-type.

Fig. 7 shows the electric potential in the device with grounded bulk and source contacts and with 1.04 and 0.01 V voltages applied to the gate and drain contacts, respectively. Fig. 8 shows the electron density as computed by the QDD (left) and SPDD (right) models and Fig. 10 shows the same quantities along the following cross-sections:

- in the oxide-bulk direction at the middle of the channel (left);
- in the source-drain direction at a position corresponding to the electron density peak (right).

Both models exhibit a stiff boundary layer in the electron concentration at the Si/SiO_2 interface. No oscillations arise in the computed solution, due to the monotone exponentially fitted finite element method used to discretize the continuity equations. Notice that, although the potential distributions in the bulk-to-gate direction are very similar, the stiffness of the boundary layer computed by the QDD model is much stronger than for the SPDD model. This can be explained by looking at Fig. 9 that shows the computed Bohm potential distributions. Note that the Bohm potential computed by the SPDD model is much smoother near

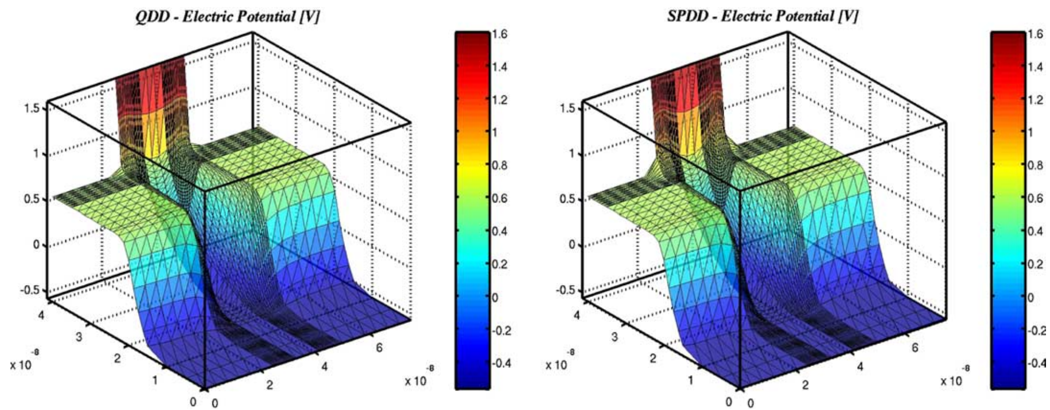

Fig. 7. Electric potential. The gate-to-source voltage is 1.04 V.

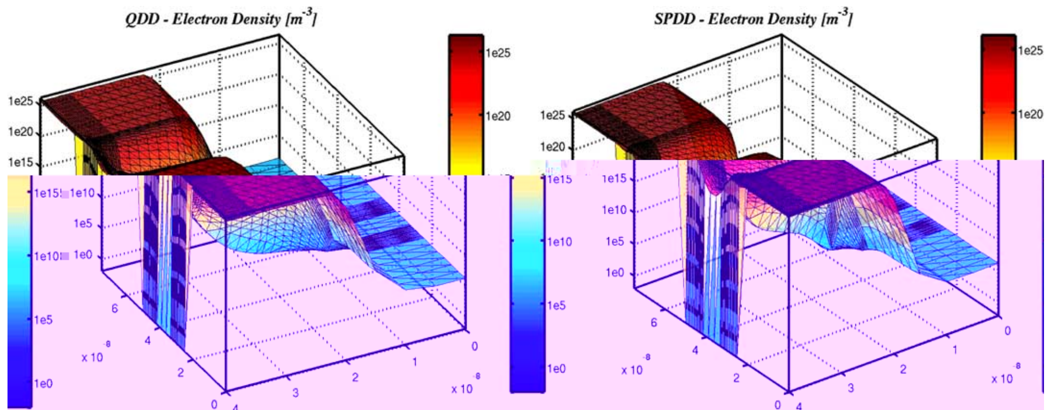

Fig. 8. Electron concentration. The gate-to-source voltage is 1.04 V.

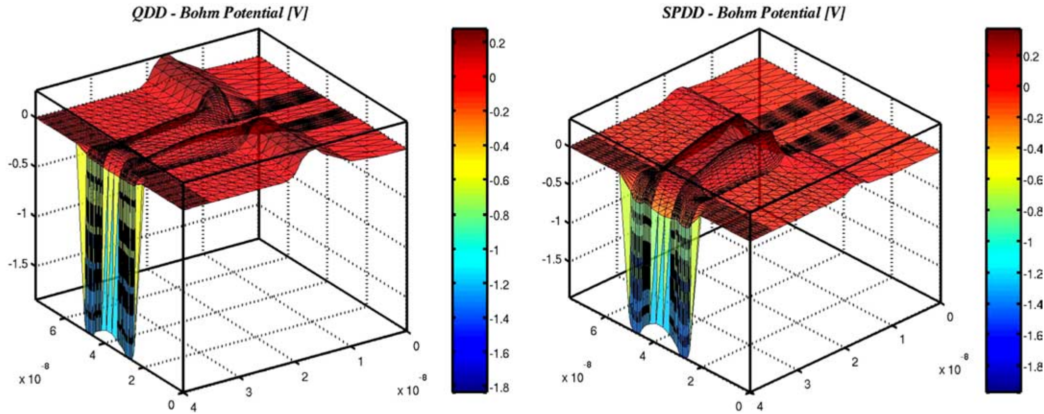

Fig. 9. Electron Bohm potential. The gate-to-source voltage is 1.04 V.

the Si/SiO_2 interface. For the QDD model, it is also to be noted that the strongly negative value of the correction factor ($G_n \simeq -1.5$ V $\Rightarrow$ $\widehat{G_n} \simeq -60$) forces the damping parameter t_k defined in Section 5 to become very small $\left(\frac{2}{t_k} > 60 \Rightarrow t_k < \frac{1}{30}\right)$. This has the drawback of slowing the convergence of the algorithm, but, at the same time, it ensures the strict positivity of the electron concentration which is a mandatory requirement in the physical problem.

A close-up on the channel of the transistor is given in Figs. 10 and 11. In particular, from Fig. 10 (left) it becomes evident that the inversion layer computed by the SPDD simulation is much wider than that of the classical simulation and the peak of the electron concentration is lower and it is attained a few *nm*'s away from the interface. The QDD solution exhibits the same qualitative features but the quantitative prediction of the charge peak shift is much smaller. The same phenomenon is observed by comparing the computed DD current field shown in Fig. 11 (left) with the QDD current field shown in Fig. 11 (middle) and the SPDD current field shown in Fig. 11 (right): in the former plot the current flow is concentrated at the interface while in the latter two it is moved down towards the bulk because very few carriers are available at the interface. By inspecting Fig. 10 (right) one may note that both quantum-corrected models predict a smaller

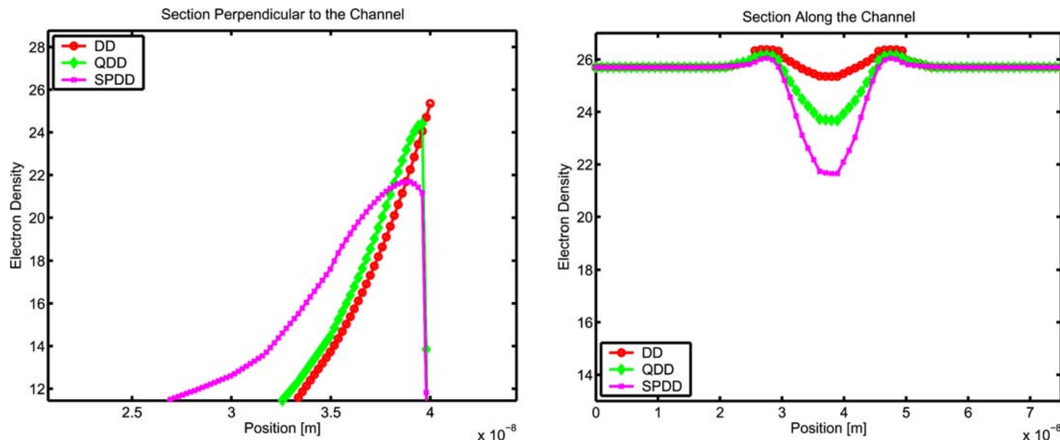

Fig. 10. Electron density in a section along the channel (right) and through the channel (left) of a MOSFET device close to the silicon/silicon-dioxide interface. The gate-to-source voltage is 1.04 V.

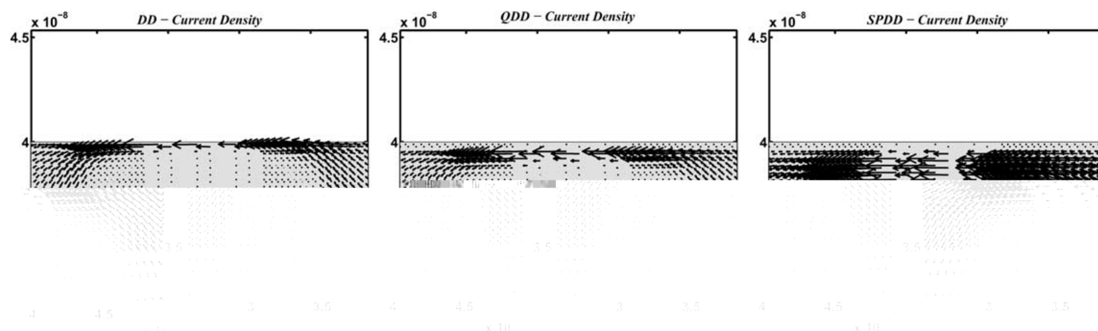

Fig. 11. Electron current density in the channel of the device (left DD model, center QDD model, right SPDD model). The gate-to-source voltage is 1.04 V.

charge density in the middle of the channel than classical simulations, but the QDD result appears to be smoother in this direction than the SPDD one.

Fig. 12 shows the *I–V* curve of the transistor. Notice that the QDD simulation results display both the relevant quantum-mechanical effects observed in [8] and in experimental measurements in [41], and consisting of a threshold voltage shift and a degraded subthreshold slope (see Fig. 14) in comparison with DD results. On the other hand, by comparing these results with the SPDD curve, we may note that the QDD model underestimates both effects. An explanation of these two latter phenomena can be attempted by comparing the log-scale curve in Fig. 12 (right) with Fig. 13 where we show on the left the total charge in the device ($q_{\mathrm{Si}} = \int_{\Omega_{\mathrm{Si}}} q(p - n + D)\,\mathrm{d}x$) against the gate voltage and on the right the capacitance ($C_{\mathrm{MOS}} = \frac{\mathrm{d}q_{\mathrm{Si}}}{\mathrm{d}V_g}$) against the gate voltage. As noted before, on the one hand, due to the strong confinement in the channel in the bulk-to-gate direction, the QDD and SPDD electron concentration in the channel are consistently lower than the DD value, on the other hand the QDD and SPDD inversion layers are wider and their width increases with the increase of gate voltage, because lowering the potential barrier between source and drain

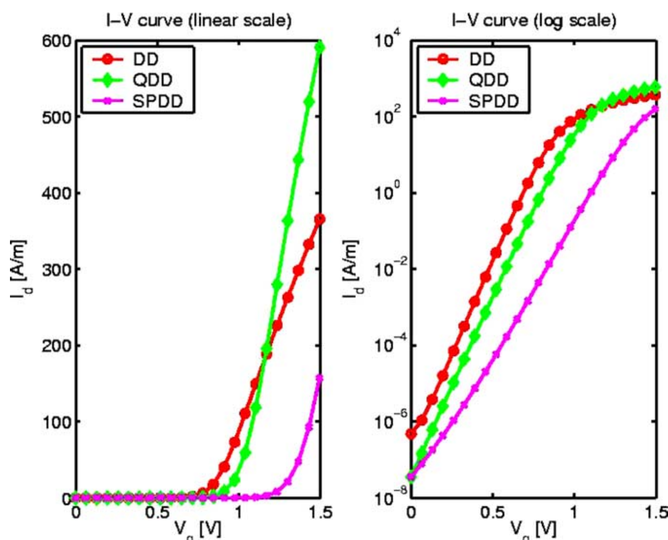

Fig. 12. *I–V* characteristics of a MOSFET device. Comparison between DD (dashed line) and bipolar QDD (solid line) models. The drain current is plotted versus the gate-to-source voltage at 0.01 V drain-to-source voltage. Left: linear scale; right: log-scale.

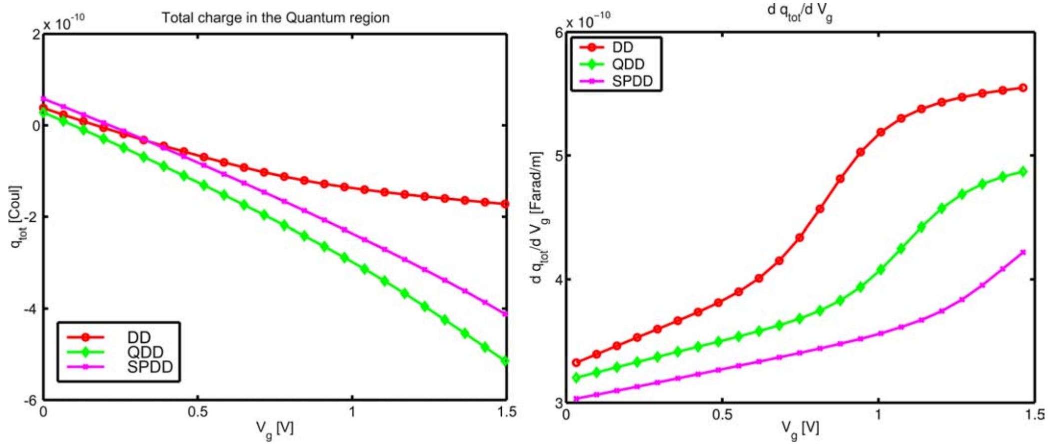

Fig. 13. Left: total charge in the quantum box. Right: capacitance.

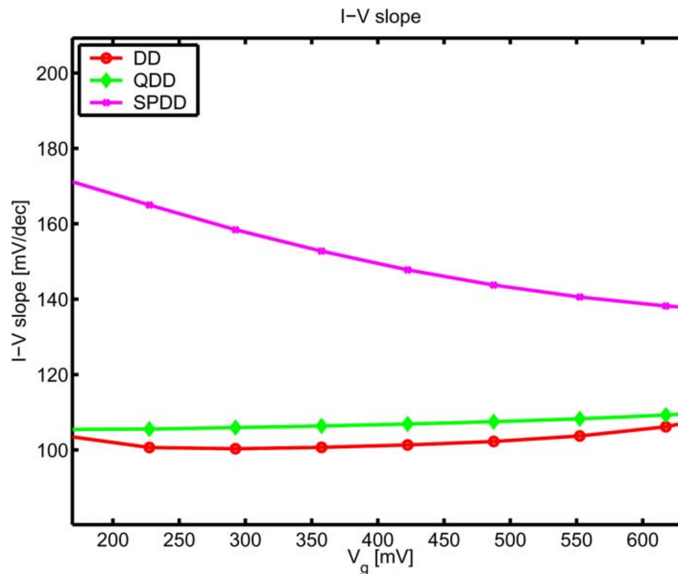

Fig. 14. Subthreshold slopes computed by DD and QCDD models.

regions allows carriers to directly tunnel through the channel, as is also observed in [3] for double-gate MOSFETs. Comparing QDD and SPDD predictions seems to indicate that the QDD model underestimates the peak shift effect while it overestimates the effect of charge penetration under the channel barrier. As a concluding modeling remark, we point out that all the results provided by the QDD simulations rely on a proper choice of the value of the quantum diffusion coefficient δ_n^2, which is typically used as a fitting parameter depending on the semiconductor material [42]. The weight of quantum-mechanical effects along different spatial directions suggests that to fit the QDD results with the SPDD predictions one might choose to take δ_n^2 as a rank-two tensor instead of a scalar quantity. This could have a strong beneficial impact in routine device simulations because, although the SPDD model is less expensive than more advanced quan-

tum models, the computational effort required for the simulation of one bias point with SPDD is about 10 times more than required by QDD (approximately 5 minutes running a `Matlab` computer code on a PC with a 700 MHz PPC-G3 processor).

8. Conclusions and future work

In this paper, we have proposed a unified framework for quantum-corrected drift-diffusion (QCDD) models in nanoscale semiconductor device simulation. QCDD models are presented as a suitable generalization of the classical drift-diffusion (DD) system, each particular model being identified by the constitutive relation for the quantum-correction to the electric potential. We have examined two special, and relevant, examples of QCDD models; the first one is the Schrödinger–Poisson–drift-diffusion model, and the second one is the quantum–drift-diffusion model. Both approaches provide a more accurate description of the density of states than the DD model, in order to represent the effect of two-dimensional quantum confinement on the charge densities in stationary regime. However, they neglect quantum effects in charge transport modeling because they use a classical DD expression for the current densities. For the numerical treatment of the two models, we have introduced a functional iteration technique that extends the classical Gummel decoupled algorithm widely used in the iterative solution of the DD system. This extension represents, to our knowledge, the first fully decoupled procedure applied to the QDD model that ensures strict positivity of the charge densities at each step of the iteration. We have discussed the finite element discretization of the various differential subsystems, with special emphasis on their stability properties, and we have successfully validated the performance of the proposed algorithms and models on the numerical simulation of nanoscale devices in two spatial dimensions.

Possible extensions in further investigation of QCDD models could be:

- the use of a Schrödinger equation with open boundary and direct quantum calculation of the current densities to simulate quantum transport effects in device regions where the ballistic limit holds [6];
- the use of Fermi–Dirac statistics to account for short-range scattering phenomena related to the Pauli exclusion principle;
- the inclusion of temperature effects connected with lattice heating (in the stationary case) and electron gas heating (in transient computations).

Acknowledgments

We gratefully acknowledge Dr. R. Gusmeroli, Dr. A. Pirovano and Prof. A. S. Spinelli, Dipartimento di Elettronica e Informazione, Politecnico di Milano, for many fruitful discussions on the research object of this paper.

References

[1] C. Moglestue, Self-consistent calculation of electron and hole inversion charges at silicon–silicon dioxide interfaces, J. Appl. Phys. 59 (1986) 3175–3183.
[2] W. Haensch, The Drift-Diffusion Equation and Its Applications in MOSFET Modeling, Springer, Wien, New York, 1991.
[3] J. Watling, A. Brown, A. Asenov, Can the density gradient approach describe the source-drain tunnelling in decanano double-gate MOSFETs?, J. Comput. Electron. (1) (2002) 289–293.
[4] S. Datta, Nanoscale device modeling: the Green's function method, Superlattices Microstruct. 28 (4) (2000) 253–278.
[5] P. Markowich, C. Ringhofer, C. Schmeiser, Semiconductor Equations, Springer, Wien, 1990.
[6] S. Laux, A. Kumar, M. Fischetti, Ballistic FET modeling using QDAME: Quantum device analysis by modal evaluation, IEEE Trans. Nanotechnol. 1 (4) (2002) 255–259.

[7] M. Ancona, G. Iafrate, Quantum correction to the equation of state of an electron gas in a semiconductor, Phys. Rev. B 39 (1989) 9536–9540.
[8] A. Pirovano, A. Lacaita, A. Spinelli, Two-dimensional quantum effects in nanoscale MOSFETs, IEEE Trans. Electron. Dev. 1 (47) (2002) 25–31.
[9] A. Jüngel, Quasi-hydrodynamic Semiconductor Equations Progress in Nonlinear DiffeBirkhauser, Birkhäuser, 2001.
[10] A. Spinelli, A. Benvenuti, A. Pacelli, Self-consistent 2D model for quantum effects in n-MOS transistors, IEEE Trans. Electron. Dev. 45 (1998) 1342–1349.
[11] N.B. Abdallah, F. Mehats, N. Vauchelet, Analysis of a Drift-Diffusion–Schrödinger–Poisson model, Comptes Rendus Math. 335 (12) (2002) 1007–1012.
[12] A. Wettstein, Quantum effects in MOS devices, Ph.D. thesis, ETH Zürich, 2000.
[13] S. Selberherr, Analysis and Simulation of Semiconductor Devices, Springer, Wien, 1984.
[14] P. Markowich, The Stationary Semiconductor Device Equations, Springer, Wien, 1986.
[15] J. Jerome, Analysis of Charge Transport, Springer, New York, 1996.
[16] R. Pinnau, A. Unterreiter, The stationary current–voltage characteristics of the quantum drift-diffusion model, SIAM J. Numer. Anal. 37 (1) (1999) 211–245.
[17] R. Pinnau, Uniform Convergence of an Exponentially Fitted Scheme for the Quantum Drift-Diffusion Model, to appear in SIAM J. Numer. Anal. (2004).
[18] A. Jüngel, R. Pinnau, A positivity-preserving numerical scheme for a nonlinear fourth order parabolic system, SIAM J. Numer. Anal. 39 (2) (2001) 385–406.
[19] S. Micheletti, R. Sacco, P. Simioni, Numerical Simulation of Resonant Tunnelling Diodes with a Quantum Drift-Diffusion Model, in: S.H.W.H.A. Schilders, E.J.W. ter Maten (Eds.), Scientific Computing in Electrical Engineering, Lecture Notes in Computer Science, Springer, 2004, pp. 313–321.
[20] T. Kerkhoven, Y. Saad, On acceleration methods for coupled nonlinear elliptic systems, Numer. Math. 60 (1992) 525–548.
[21] F. Bosisio, S. Micheletti, R. Sacco, A discretization scheme for an extended drift-diffusion model including trap-assisted phenomena, J. Comput. Phys. 159 (2000) 197–212.
[22] M. Ancona, H. Thiersten, Macroscopic physics of the silicon inversion layer, Phys. Rev. B 35 (1987) 580.
[23] N.B. Abdallah, A. Unterreiter, On the stationary quantum drift-diffusion model, Z. Angew. Math. Phys. 49 (1998) 251–275.
[24] M.G. Ancona, Z. Yu, R.W. Dutton, P.J.V. Vorde, M. Cao, D. Vook, Density-gradient analysis of MOS tunnelling, IEEE Trans. Electron. Dev. 47 (12) (2000) 2310–2319.
[25] C. Kittel, Introduction to Solid State Physics, third ed., Wiley & Sons, New York, 1967.
[26] H.K. Gummel, A self-consistent iterative scheme for one-dimensional steady-state transistor calculations, IEEE Trans. Electron. Dev. ED-11 (1964) 455–465.
[27] R. Bank, D. Rose, W. Fichtner, Numerical methods for semiconductor device simulation, IEEE Trans. Electron. Dev. ED-30 (1983) 1031–1041.
[28] F. Brezzi, L. Marini, P. Pietra, Two-dimensional exponential fitting and applications to semiconductor device equations, SIAM J. Numer. Anal. 26 (1989) 1342–1355.
[29] F. Brezzi, L. Marini, P. Pietra, Numerical simulation of semiconductor devices, Comp. Meths. Appl. Mech. Engrg. 75 (1989) 493–514.
[30] R. Sacco, F. Saleri, Mixed finite volume methods for semiconductor device simulation, Numer. Meth. Partial Differential Equations 13 (1997) 215–236.
[31] E. Gatti, S. Micheletti, R. Sacco, A new Galerkin framework for the drift-diffusion equation in semiconductors, East West J. Numer. Math. 6 (2) (1998) 101–135.
[32] T. Kerkhoven, A proof of convergence of Gummel's algorithm for realistic device geometries, SIAM J. Numer. Anal. 23 (1986) 1121–1137.
[33] T. Kerkhoven, A spectral analysis of the decoupling algorithm for semiconductor simulation, SIAM J. Numer. Anal. 25 (1988) 1299–1312.
[34] J.J. Jerome, T. Kerkhoven, A finite element approximation theory for the drift-diffusion semiconductor model, SIAM J. Numer. Anal. 28 (1991) 403–422.
[35] R. Sacco, A mixed problem for electrostatic potential in semiconductors, Numer. Meth. Partial Differential Equations 10 (1994) 715–738.
[36] R. Adams, Sobolev spaces, Academic Press, New York, 1975.
[37] J. Lions, E. Magenes, Problèmes aux limites non-homogènes et applications, Dunod, Paris, 1968.
[38] H.G. Roos, M. Stynes, L. Tobiska, Numerical methods for singularly perturbed differential equations, Springer, Berlin Heidelberg, 1996.
[39] R.E. Bank, W.M. Coughran, L.C. Cowsar, Analysis of the finite volume Scharfetter-Gummel method for steady convection diffusion equations, Comp. Vis. Sci. 1 (1998) 123–126.
[40] International technology road map for semiconductors, Tech. rep., ITRS, 2003.

[41] J.D. Lee, W.Y. Choi, B.Y. Choi, Y.J. Choi, D.-S. Woo, B.-G. Park, 30 nm MOSFET development based on processes for nanotechnology, in: Proccedings of the IEEE International Conference on Semiconductor Electronics, 2002, pp. 251–254.
[42] A. Schenk, Physical modeling of deep-submicron devices, in: Proceedings of ESSDERC'01, 2001, pp. 9–16.
[43] P. Degond, A.E. Ayyadi, A coupled Schrödinger drift-diffusion model for quantum semiconductor device simulations, J. Comput. Phys. 181 (1) (2002) 222–259.