Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 227 (2008) 6226–6240
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

# An accelerated monotone iterative method for the quantum-corrected energy transport model ☆

Ren-Chuen Chen [a], Jinn-Liang Liu [b,*]

[a] Department of Mathematics, National Kaohsiung Normal University, Kaohsiung 802, Taiwan
[b] Department of Applied Mathematics, National University of Kaohsiung, Kaohsiung 811, Taiwan

Received 20 September 2007; received in revised form 2 March 2008; accepted 3 March 2008
Available online 14 March 2008

**Abstract**

A non-stationary monotone iterative method is proposed and analyzed for the quantum-corrected energy transport model in nanoscale semiconductor device simulation. For the density-gradient equations, it is analytically and numerically shown that the convergence rate of the method is optimal in the sense of Gummel's decoupling iteration. This is a globally convergent method in the sense that the initial guess can be taken as a lower or an upper solution which is independent of applied voltages. The method integrates the monotone parameters, grid sizes, and Scharfetter–Gummel fitting in an adaptive and automatic way to treat the singularly perturbed nature of the model that incurs boundary, junction, and quantum potential layers in the device.


*Keywords:* Monotone iterative method; Quantum energy transport model; Semiconductor device; Convergence analysis

## 1. Introduction

In order to keep pace with the increasing speed of miniaturization of modern semiconductor technology, a great variety of device models that account for quantum effects, accuracy, robustness, and efficiency in real-life simulations have been intensively developed and tested in recent years, see e.g. [3–5,6,9–15,17,18,21,23,25,26,28] and references therein. Among them, a class of macroscopic quantum mechanical models based on the density-gradient (DG) theory of Ancona and Tiersten [1] have been shown to accurately simulate multi-dimensional MOSFET devices with gate lengths ranging from 50 nm down to 6 nm [3,5,9–13,18,21,23,25,28]. The DG theory is a continuum theory of quantum transport that exhibits the essential non-locality, confinement, and tunnelling of quantum effects by generalizing the equation of state for an ideal electron gas to include density-gradient dependences. In this paper, we consider in particular the quantum-corrected energy transport model (QCET) proposed in [9] which is one of the most advanced DG models in the sense that it

☆ This work was supported by NSC under Grants 96-2115-M-017-001 and 96-2628-M-390-006, Taiwan.
* Corresponding author.
*E-mail addresses:* rcchen@nknucc.nknu.edu.tw (R.-C. Chen), jinnliu@nuk.edu.tw (J.-L. Liu).

0021-9991/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2008.03.003

extends the commonly used quantum drift-diffusion (QDD) model [3,5,10,11,21,23,25,28] by including the energy balance equations in order to take the hot electron effects into account.

Since the power supply voltage is slowly scaled as compared with the gate length, the local field strength inside the device increases drastically and consequently incurs very large solution variations and sharp gradients. Moreover, the QCET model is a strongly coupled system of seven nonlinear partial differential equations (PDEs) in which the Poisson equation and the DG equations are singularly perturbed in terms of the Debye length [2] and the Planck constant [9,11,21], respectively. The singularities imply abrupt and thin boundary, interior, and quantum potential layers [9] in the device. All of these physical and mathematical properties associated with the model lead to various difficulties in solving numerically the device problem by appropriate discretizations such as the adaptive finite element method with a generalized Scharfetter–Gummel exponentially fitting developed in [7], which is also employed in this paper. One of the difficulties is to accurately and efficiently solve the nonlinear system of PDEs.

At functional level, there are two approaches to deal with this problem. One may solve the nonlinear system either by Newton's method that treats the coupled system as a whole or by Gummel's method that decouples the system into sub-systems. As noted in [2,11,24], Gummel's method is very robust, better-conditioned, memory efficient, and widely used in semiconductor simulations but only yields a linear convergence [24]. Since the system is even further enlarged by the DG equations, most multi-dimensional quantum-corrected models are treated by Gummel's method [9,11,21] as considered here.

For each Gummel's iteration, we have to solve seven decoupled nonlinear PDEs for the QCET model. At discrete level, each one of the seven decoupled nonlinear systems of algebraic equations can either be solved by using Newton's method or by the monotone iterative method proposed in [7,9]. As limited by the Gummel iteration, the optimal convergence of the nonlinear solver is linear by either method. For device simulation, the monotone iterative method is more robust than Newton's method since it is globally convergent, highly parallel, and very easy to implement, see [7,9] and references therein for more details. However, this method will be slowly convergent if the monotone parameters are not properly chosen. We propose here an accelerated monotone iterative method for the QCET model. It is optimal for the DG equations in the sense of Gummel's iteration. The main idea of the method is to set up a selection criterion of the monotone parameter for each PDE. The monotone parameter is chosen to be a diagonal entry of the Jacobian matrix for most of the solution domain and is relaxed in the layer regions so that the resulting matrix of the linear system is an $M$-matrix. This is a crucial condition to establish that the maximum principle is satisfied by the approximate carrier densities which are therefore guaranteed to be non-negative throughout the domain. The selection criterion is done automatically without any manual adjustment of relaxation parameters. In effect, the relaxation parameters (or equivalently the monotone parameters) are modified automatically with the adaptive grid sizes and the Scharfetter–Gummel fitting. In summary, the monotone parameters, adaptive grid sizes, Scharfetter–Gummel fitting, and two singular perturbation parameters are all tightly related to each other at the discrete level so that the approximate solutions can be analyzed by the singular perturbation theory as that of [16]. We do not provide any singular perturbation analysis in this paper. Nevertheless, our numerical results demonstrate that all approximate solutions of the decoupled system are uniformly convergent as proved in [16] for the classical DD model.

The plan of the paper is as follows. A short review of the DGET model is stated in Section 2. In Section 3, we present the accelerated monotone iterative method. Starting with an upper solution as the initial guess, it is also shown in this section that the method is globally convergent and optimal in the sense of Gummel's iteration for the DG equations. Section 4 presents our numerical experiments on a MOSFET device structure to demonstrate the accuracy and efficiency of the method.

## 2. The quantum-corrected energy transport model

Based on the density-gradient theory [1], we have proposed in [9] the following QCET model:

$$-\Delta\phi = F(\phi), \tag{2.1}$$

$$-\nabla\cdot\mathbf{J}_n = R(u,v), \tag{2.2}$$

$$-\nabla\cdot\mathbf{J}_{\mathrm{p}} = -R(u,v), \tag{2.3}$$
$$-\Delta\zeta_{\mathrm{n}} = Z_{\mathrm{n}}(\zeta_{\mathrm{n}}), \tag{2.4}$$
$$-\Delta\zeta_{\mathrm{p}} = Z_{\mathrm{p}}(\zeta_{\mathrm{p}}), \tag{2.5}$$
$$-\nabla\cdot\mathbf{G}_{\mathrm{n}} = R_{\mathrm{n}}(g_{\mathrm{n}}), \tag{2.6}$$
$$-\nabla\cdot\mathbf{G}_{\mathrm{p}} = R_{\mathrm{p}}(g_{\mathrm{p}}), \tag{2.7}$$

with the seven unknown functions $\phi$,

$$u = \exp\left(\frac{-\varphi_{\mathrm{n}}}{V_{\mathrm{T}}}\right), \tag{2.8}$$
$$v = \exp\left(\frac{\varphi_{\mathrm{p}}}{V_{\mathrm{T}}}\right), \tag{2.9}$$
$$\zeta_{\mathrm{n}} = \sqrt{n}, \tag{2.10}$$
$$\zeta_{\mathrm{p}} = \sqrt{p}, \tag{2.11}$$
$$g_{\mathrm{n}} = T_{\mathrm{n}}\exp\left(-\frac{5\varphi_{\mathrm{n}}}{4\,V_{\mathrm{T}}}\right), \tag{2.12}$$
$$g_{\mathrm{p}} = T_{\mathrm{p}}\exp\left(\frac{5\varphi_{\mathrm{p}}}{4V_{\mathrm{T}}}\right), \tag{2.13}$$

and the auxiliary relations

$$n = n_{\mathrm{I}}\exp\left(\frac{\phi+\phi_{\mathrm{qn}}}{V_{\mathrm{T}}}\right)u, \tag{2.14}$$
$$p = n_{\mathrm{I}}\exp\left(\frac{-\phi-\phi_{\mathrm{qp}}}{V_{\mathrm{T}}}\right)v, \tag{2.15}$$
$$\mathbf{E} = -\nabla\phi, \tag{2.16}$$
$$\phi_{\mathrm{qn}} = V_{\mathrm{T}}\ln(\zeta_{\mathrm{n}}^2) - V_{\mathrm{T}}\ln(un_{\mathrm{I}}) - \phi, \tag{2.17}$$
$$\phi_{\mathrm{qp}} = -V_{\mathrm{T}}\ln(\zeta_{\mathrm{p}}^2) + V_{\mathrm{T}}\ln(vn_{\mathrm{I}}) - \phi, \tag{2.18}$$
$$\mathbf{J}_{\mathrm{n}} = -q\mu_{\mathrm{n}}\nabla(\phi+\phi_{\mathrm{qn}}) + qD_{\mathrm{n}}\nabla n = qD_{\mathrm{n}}n_{\mathrm{I}}\exp\left(\frac{\phi+\phi_{\mathrm{qn}}}{V_{\mathrm{T}}}\right)\nabla u, \tag{2.19}$$
$$\mathbf{J}_{\mathrm{p}} = -q\mu_{\mathrm{p}}\nabla(\phi+\phi_{\mathrm{qn}}) - qD_{\mathrm{n}}\nabla n = -qD_{\mathrm{p}}n_{\mathrm{I}}\exp\left(\frac{-\phi-\phi_{\mathrm{qp}}}{V_{\mathrm{T}}}\right)\nabla v, \tag{2.20}$$
$$\mathbf{G}_{\mathrm{n}} = \kappa_{\mathrm{n}}\exp\left(\frac{5\varphi_{\mathrm{n}}}{4V_{\mathrm{T}}}\right)\nabla g_{\mathrm{n}}, \tag{2.21}$$
$$\mathbf{G}_{\mathrm{p}} = \kappa_{\mathrm{p}}\exp\left(-\frac{5\varphi_{\mathrm{p}}}{4V_{\mathrm{T}}}\right)\nabla g_{\mathrm{p}}, \tag{2.22}$$
$$F(\phi) = \frac{qn_{\mathrm{I}}}{\varepsilon_{\mathrm{s}}}\left(v\exp\left(\frac{-\phi-\phi_{\mathrm{qp}}}{V_{\mathrm{T}}}\right) - u\exp\left(\frac{\phi+\phi_{\mathrm{qn}}}{V_{\mathrm{T}}}\right)\right) - \frac{q(N_{\mathrm{D}}^{+}-N_{\mathrm{A}}^{-})}{\varepsilon_{\mathrm{s}}}, \tag{2.23}$$
$$R(u,v) = \frac{q(n_{\mathrm{eq}}p_{\mathrm{eq}} - np)}{\tau_{\mathrm{n}}^0\left(p+\sqrt{n_{\mathrm{eq}}p_{\mathrm{eq}}}\exp\left(\frac{\varepsilon_t-\varepsilon_i}{k_{\mathrm{B}}T}\right)\right) + \tau_{\mathrm{p}}^0\left(n+\sqrt{n_{\mathrm{eq}}p_{\mathrm{eq}}}\exp\left(\frac{\varepsilon_t-\varepsilon_i}{k_{\mathrm{B}}T}\right)\right)}, \tag{2.24}$$
$$Z_{\mathrm{n}}(\zeta_{\mathrm{n}}) = \frac{\zeta_{\mathrm{n}}}{2b_{\mathrm{n}}}\left[-V_{\mathrm{T}}\ln(\zeta_{\mathrm{n}}^2) + V_{\mathrm{T}}\ln(un_{\mathrm{I}}) + \phi\right], \tag{2.25}$$
$$Z_{\mathrm{p}}(\zeta_{\mathrm{p}}) = -\frac{\zeta_{\mathrm{p}}}{2b_{\mathrm{p}}}\left[V_{\mathrm{T}}\ln(\zeta_{\mathrm{p}}^2) - V_{\mathrm{T}}\ln(vn_{\mathrm{I}}) + \phi\right], \tag{2.26}$$

$$R_n(g_n) = \mathbf{J}_n \cdot \mathbf{E} - n\left(\frac{\omega_n - \omega_0}{\tau_{n\omega}}\right), \tag{2.27}$$

$$R_p(g_p) = \mathbf{J}_p \cdot \mathbf{E} - p\left(\frac{\omega_p - \omega_0}{\tau_{p\omega}}\right), \tag{2.28}$$

where $\phi$ is the electrostatic potential, n and p the electron and hole concentrations, $\varphi_n$ and $\varphi_p$ the generalized quasi-Fermi potentials, $V_T$ the thermal voltage, $T_n$ and $T_p$ the electron and hole temperatures, $n_I$ the intrinsic carrier concentration, $\phi_{qn}$ and $\phi_{qp}$ the quantum potentials, $\mathbf{E}$ the electric field, $\mathbf{J}_n$ and $\mathbf{J}_p$ the electron and hole current densities, $q$ the elementary charge, $\mu_n$ and $\mu_p$ the field-dependent electron and hole mobilities, $D_n$ and $D_p$ the electron and hole diffusion coefficients expressed by the Einstein relation with the mobilities, $\kappa_n$ and $\kappa_p$ the heat conductivities, $\varepsilon_s$ the permittivity constant of the semiconductor, $N_A^-$ and $N_D^+$ the densities of ionized impurities, $\tau_{n\omega}$ and $\tau_{p\omega}$ the carrier energy relaxation times, $\omega_0$ the thermal energy, $\omega_n$ and $\omega_p$ the carrier average energies, and $b_n = \frac{\hbar^2}{12m_n^* q}$ and $b_p = \frac{\hbar^2}{12m_p^* q}$ the material parameters measuring the strength of the gradient effects in the electron gas with $\hbar$ being the reduced Planck constant and $m_n^*$ and $m_p^*$ being the effective masses of electron and hole [1]. Note that each one of the PDEs (2.1)–(2.7)is semilinear and self-adjoint with respect to its own unknown function in the ordering of the variables $\phi$, $u$, $v$, $\zeta_n$, $\zeta_p$, $g_n$, and $g_p$. The model system is associated with a set of Dirichlet and Neumann boundary conditions (BCs) for various semiconductor device structures as described in [7,9] where a detailed description of physical parameters, doping profiles, generation–recombination models, and mobility models etc. can also be found. In particular, we consider here a MOSFET device as illustrated by Fig. 1 where the junction depth is 20 nm, the lateral diffusion under gate is 8 nm, the channel length is 34 nm, and the gate oxide thickness is 2 nm.

The system (2.1)–(2.7) models the stationary state of electron flow through the device by augmenting the macroscopic energy transport model (2.1)–(2.3), (2.6), (2.7) [7] with the density-gradient Eqs. (2.4) and (2.5) [1]. The square roots of carrier densities in (2.10) and (2.11) were introduced in [23] as extra unknown functions to define the quantum (Bohm) potentials (2.17) and (2.18) by means of the generalized quasi-Fermi potentials $\varphi_n$ and $\varphi_p$ [9,23]. These quantum potentials represent the first order quantum corrections of the drift-diffusion fluxes as shown in (2.19) and (2.20).

Note particularly that the right-hand side nonlinear functionals in (2.1)–(2.7), namely (2.23)–(2.28), are all expressed in terms of their respective unknown variables to illustrate that the functional derivative with respect

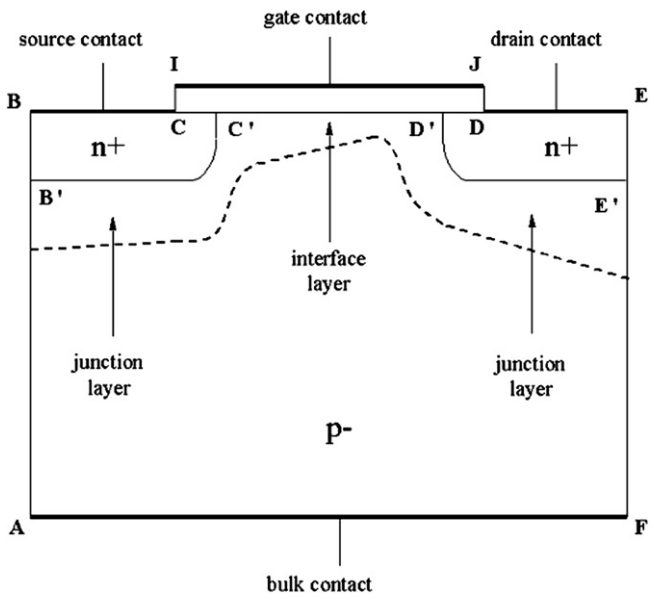


Fig. 1. Geometry of an *n*-MOSFET device.

to its variable for can be straighforwardly evaluated. All functionals are nonlinear due to the Slotboom-type transformations (2.8), (2.9), (2.12) and (2.13). Furthermore, these transformations also result in that all divergence operators in the left-hand sides of the system (2.1)–(2.7) are self-adjoint.

Due to the self-adjointness, each PDE with its prescribed BCs results in a similar nonlinear algebraic system from the adaptive finite element method. Moreover, it has been shown that the classical Gummel algorithm, i.e., a successively decoupling procedure for solving the DD model can be extended to the QCET model in [9] and the QCDD model in [11]. It is therefore sufficient to restrict our discussion below to only one of the decoupled PDEs. From our numerical experience on the QCET model and from the fact that the PDEs (2.4) and (2.5) are singularly perturbed with respect to the Planck constant, these two equations suffer from more ill-conditioning of the corresponding algebraic systems than that of other equations and consequently from slower convergence of the nonlinear iteration. We thus pay attention particularly to these two equations for our proposed monotone scheme. All other PDEs can be treated and analyzed in a similar way for which we refer to [9] for a detailed derivation of the corresponding Jacobian matrices.

Let $\Omega \subset \Re^2$ denote the bounded domain of the silicon in Fig. 1. The boundary $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$ is piecewise smooth consisting of Dirichlet $\partial\Omega_D$ and Neumann $\partial\Omega_N = \overline{AB} \cup \overline{EF}$ parts. The Dirichlet part corresponds to the Ohmic contacts $\partial\Omega_O = \overline{BC} \cup \overline{DE} \cup \overline{AF}$ and the silicon/oxide interface $\partial\Omega_I = \overline{CD}$, i.e., $\partial\Omega_D = \partial\Omega_O \cup \partial\Omega_I$. The BCs for the variables $\zeta_n$ and $\zeta_p$ are prescribed as

$$\zeta_n^2 = \frac{1}{2}\left[(N_D^+ - N_A^-) + \sqrt{(N_D^+ - N_A^-)^2 + 4n_I^2}\right] \quad \text{on } \partial\Omega_O, \tag{2.29}$$

$$\zeta_p = n_I/\zeta_n \quad \text{on } \partial\Omega_O, \tag{2.30}$$

$$\zeta_p = \zeta_n = 0 \quad \text{on } \partial\Omega_I, \tag{2.31}$$

and

$$\frac{\partial\zeta_n}{\partial\mathbf{n}} = \frac{\partial\zeta_p}{\partial\mathbf{n}} = 0 \quad \text{on } \partial\Omega_N, \tag{2.32}$$

where $\mathbf{n}$ is an outward normal unit vector to $\partial\Omega_N$. The BC (2.31) on the interface is so chosen that we do not consider tunneling effects [5] [9] across the interface. The monotone method presented below is not constrained by this simplification.

## 3. An accelerated monotone iterative method

The adaptive finite element method and the Scharfetter–Gummel exponential fitting scheme of [7] are used to approximate the boundary value problem (2.4), (2.29)–(2.32) and yield a system of nonlinear algebraic equations in matrix form as

$$\mathbf{A}U = F(U), \tag{3.1}$$

where $U \equiv (\zeta_1, \ldots, \zeta_N)^T$ is an unknown vector with $\zeta_i$ representing an approximation of $\zeta_n$ at the mesh point $(x_i, y_i) \in \overline{\Omega}$, i.e., $\zeta_i \approx \zeta_n(x_i, y_i)$, and $F(U) \equiv (F_1(\zeta_1), \ldots, F_N(\zeta_N))^T$ is a nonlinear vector in $U$ defined as

$$F_i(\zeta_i) = \begin{cases} Z_n(\zeta_i) & \text{if } (x_i, y_i) \in \Omega, \\ \frac{n_I}{\zeta_p(x_i, y_i)} & \text{if } (x_i, y_i) \in \partial\Omega_O, \\ 0 \quad \text{if } (x_i, y_i) \in \partial\Omega_N \cup \partial\Omega_I. \end{cases} \tag{3.2}$$

Note that the matrix $\mathbf{A} = [a_{ij}]_{N\times N}$ is diagonally dominant, i.e., $|a_{ii}| \geqslant \sum_{j\neq i}|a_{ij}|$ for all $i$ and the strict inequality holds for at least one $i$ as proved in [7] and hence it is an $M$-matrix (an inverse-positive matrix).

To develop monotone iterative schemes for (3.1), we need a pair of ordered lower and upper solutions $\underline{U} \leqslant \overline{U}$ such that

$$\mathbf{A}\underline{U} \leqslant F(\underline{U}) \quad \text{and} \quad \mathbf{A}\overline{U} \geqslant F(\overline{U}), \tag{3.3}$$

where the inequality between two vectors is always in the componentwise sense. Given any ordered pair, we define the sectors

$$\langle \underline{U}, \overline{U} \rangle \equiv \{W \in \Re^N; \underline{U} \leqslant W \leqslant \overline{U}\}, \tag{3.4}$$

$$\langle \underline{\zeta_i}, \overline{\zeta_i} \rangle \equiv \{w_i \in \Re; \underline{\zeta_i} \leqslant w_i \leqslant \overline{\zeta_i}\} \tag{3.5}$$

and stationary monotone parameters $\lambda_i$ such that

$$\lambda_i \equiv \max\left\{-\frac{\partial F_i(w_i)}{\partial \zeta_i}; w_i \in \langle \underline{\zeta_i}, \overline{\zeta_i} \rangle\right\}, \tag{3.6}$$

or in matrix form

$$\Lambda \equiv \mathrm{diag}(\lambda_i) \tag{3.7}$$

for $1 \leqslant i \leqslant N$. The stationary monotone iterative method for (3.1) generates a sequence of approximate solutions $\{U_S^{(m)}\}$ of an exact solution $U$ by iteratively solving the linearized system

$$(\mathbf{A} + \Lambda)U_S^{(m+1)} = F(U_S^{(m)}) + \Lambda U_S^{(m)} \tag{3.8}$$

for $m = 0, 1, 2, \ldots$. The initial iterate is chosen as either $U_S^{(0)} = \underline{U}$ or $U_S^{(0)} = \overline{U}$. The method is said to be *stationary* if all monotone parameters are kept fixed throughout the entire solution process and *non-stationary* otherwise.

The monotone iterative method has been widely used to show existence of solutions of some elliptic nonlinear PDEs. It may be used whenever a maximum principle is available. It depends on finding a super-solution (an upper solution for algebraic equations considered here), a sub-solution (a lower solution), and a convenient constant $\lambda$ (a constant matrix $\Lambda$). The monotone iterative method reduces to Newton's method if the monotone parameter is chosen to be the first derivative of the nonlinear functional of the PDE (the Jacobian matrix of algebraic system). Obviously, (3.8) is Newton's iteration if the matrix $\Lambda$ is the Jacobian matrix $\left[-\frac{\partial Z_n(\zeta_i^{(m)})}{\partial \zeta_j^{(m)}}\right]$. A drawback of Newton's method is its sensitivity to the initial guess. Another major difficulty for semiconductor device simulation using Newton's method is that the Jacobian matrix is highly unstable when the applied biases are large [2], i.e., far away from thermal equilibrium which is the case considered here. In contrast, the monotone iterative generates a monotonically convergent sequence from any one of a wide class of initial iterates. And the matrix $\Lambda$ can be simply a constant diagonal matrix.

With the nonlinear functional $Z_n(\zeta_n)$ in (2.25) and the matrix $\mathbf{A}$ being an $M$-matrix, it has been shown in [9] that the sequence $\{U_S^{(m)}\}$ converges monotonically to an exact solution. The convergence is also global in the sense that the lower or upper solution can be easily determined by means of the charge neutrality condition which is independent of applied voltages [9]. However, the convergence is rather slow as shown in the next section. We propose here an accelerated method by properly choosing the monotone parameters for each iteration. The method is non-stationary since the parameters are automatically updated in each iteration as follows.

We observe from (2.25) that

$$\frac{\partial Z_n}{\partial \zeta_n} = \frac{-1}{2b_n}\left[V_T \ln(\zeta_n^2) - V_T \ln(un_i) - \phi\right] - \frac{\zeta_n}{2b_n}\left[2V_T \cdot \frac{1}{\zeta_n}\right], \tag{3.9}$$

$$= \frac{-1}{2b_n}\left[V_T \ln(\zeta_n^2) - V_T \ln(un_i) - \phi + 2V_T\right], \tag{3.10}$$

$$= \frac{-1}{2b_n}\left[\phi_{qn} + 2V_T\right], \tag{3.11}$$

$$\frac{\partial^2 Z_n}{\partial \zeta_n^2} = \frac{-1}{2b_n}\left[2V_T \cdot \frac{1}{\zeta_n}\right] < 0, \tag{3.12}$$

where (3.11) follows from (2.17). Since the monotone parameters $\lambda_i$ in (3.6) are determined by the functional derivative (3.11), it is critical to study their numerics closely. We first note that the thermal voltage $V_T$ and the microscopic parameter $b_n$ are constant throughout the domain if the parabolic effective mass approximation is assumed. This microscopic constant is a singular perturbation parameter for the DG Eq. (2.4) with respect to the Planck constant $\hbar$ [11,21]. It is thus evident that the convergence behavior of the monotone iterative method is essentially harnessed by the quantum (Bohm) potential $\phi_{qn}$. As shown by Fig. 3 in the next section,

the quantum potential has a very thin boundary layer near the silicon/oxide interface. In fact, it is only a fraction of the length scale of the inversion layer. Since we use both the adaptive finite element method and the Scharfetter–Gummel fitting method to capture the layer as those analyzed in [16] for the Poisson Eq. (2.1) by means of the singular perturbation theory, it can hence be proved that these methods are uniformly convergent and layer jumps are automatically resolved for the DG Eqs. (2.4) and (2.5).

We are however concerned here with the convergence rate of the nonlinear solver. From Fig. 3, we observe that the electron quantum potential $\phi_{\rm qn}$ is negative in the layer with the average value about $-0.15$ V whereas $V_{\rm T} = 0.0259$ V and is non-negative in the rest part of the domain. This means that (3.11) will be negative for a very large portion of the domain except the thin layer region. It consequently motivates us to choose the monotone parameters as

$$\lambda_i^{(m)} = -\frac{\partial F_i(\zeta_i^{(m)})}{\partial \zeta_i^{(m)}} \quad \text{if} \quad -\frac{\partial F_i(\zeta_i^{(m)})}{\partial \zeta_i^{(m)}} > 0, \tag{3.13a}$$

$$\lambda_i^{(m)} = \max\left\{-\frac{\partial F_i(w_i)}{\partial \zeta_i^{(m)}}; w_i \in \langle \underline{\zeta}_i, \overline{\zeta}_i\rangle\right\} \quad \text{otherwise}, \tag{3.13b}$$

which are a hybrid form of the stationary case around the layer region and the non-stationary case elsewhere. The non-stationary parameters are the diagonal entries of the Jacobian matrix $\left[-\frac{\partial Z_{\rm n}(\zeta_i^{(m)})}{\partial \zeta_j^{(m)}}\right]$. In matrix form, we denote

$$\Lambda_{\rm N}^{(m)} = \mathrm{diag}(\lambda_i^{(m)}). \tag{3.14}$$

Our new monotone iterative method is the following iteration:

$$(\mathbf{A} + \Lambda_{\rm N}^{(m)})U_{\rm N}^{(m+1)} = F(U_{\rm N}^{(m)}) + \Lambda_{\rm N}^{(m)} U_{\rm N}^{(m)} \tag{3.15}$$

with the initial guess as above. Since the matrix $\mathbf{A}$ is an $M$-matrix and the matrix $\Lambda_{\rm N}^{(m)}$ is a non-negative diagonal matrix, the resulting approximation of $\zeta_{\rm n} = \sqrt{n}$ satisfies the maximum principle so that the computed quantities are non-negative and hence are physically feasible. Furthermore, if we assume that the Slotboom variable $u$ and the electrostatic potential $\phi$ are given exactly in (2.25), then (2.4) is an elliptic semilinear PDE with respect to the unknown function $\zeta_{\rm n}$.

The convergence analysis of the sequence $\{U_{\rm N}^{(m)}\}$ generated by (3.15) then follows the standard analysis for the monotone iterative method applied to the discrete semilinear PDEs as given in [7,8,19,20,27]. We now summarize all the above conditions that are related to the convergence results of the proposed method as follows:

(1) $\mathbf{A}$ is an $M$-matrix,
(2) an ordered pair of lower and upper solutions (3.4) exists,
(3) the matrices (3.14) are non-negative,
(4) the variables $u$ and $\phi$ are given exactly in (2.25),
(5) $\left(\frac{\partial Z_{\rm n}(\zeta_i^{(m)})}{\partial \zeta_i^{(m)}} < \lambda_0\right)$ where $\lambda_0$ is the smallest positive eigenvalue of $\mathbf{A}$,
(6) (3.12) is satisfied.

The following two theorems phrased in our model setting have been proven in [27] where the quadratic convergence analysis is extended from [20] to include a larger class of nonlinear PDEs. The error analysis is cast in the infinity norm which is also used for our numerical results in the next section.

**Theorem 3.1.** *If Conditions* (1)–(4) *hold, then the sequence* $\{\underline{U}_{\rm N}^{(m)}\}$ *with* $\underline{U}_{\rm N}^{(0)} = \underline{U}$ *converges monotonically from below to a minimal solution* $\underline{U}_{\rm N}^*$ *of* (3.1) *in* $\langle \underline{U}, \overline{U}\rangle$ *and the sequence* $\{\overline{U}_{\rm N}^{(m)}\}$ *with* $\overline{U}_{\rm N}^{(0)} = \overline{U}$ *converges monotonically from above to a maximal solution* $\overline{U}_{\rm N}^*$ *of* (3.1) *in* $\langle \underline{U}, \overline{U}\rangle$. *Moreover,*

$$\underline{U} \leqslant \underline{U}_{\rm N}^{(m)} \leqslant \underline{U}_{\rm N}^{(m+1)} \leqslant \underline{U}_{\rm N}^* \leqslant \overline{U}_{\rm N}^* \leqslant \overline{U}_{\rm N}^{(m+1)} \leqslant \overline{U}_{\rm N}^{(m)} \leqslant \overline{U} \tag{3.16}$$

*for all* $m = 0, 1, 2, 3, \ldots$ *If, in addition, Condition* (5) *holds for all* $i = 1, 2, \ldots, N$, *then* $\underline{U}_{\rm N}^* = \overline{U}_{\rm N}^* =: U^*$ *is the unique solution of* (3.1).

**Theorem 3.2.** *If Conditions* (1)–(6) *hold, then*

$$\|\overline{U}_{\mathrm{N}}^{(m+1)} - U^*\|_\infty \leqslant \sigma \|\overline{U}_{\mathrm{N}}^{(m)} - U^*\|_\infty^2 \quad \forall m = 0, 1, 2, 3, \ldots, \tag{3.17}$$

where

$$\sigma = c_1 \|(A + c_2 I)^{-1}\|_\infty, \tag{3.18}$$

$$c_1 = \max_i \max \left\{ \left| \frac{\partial^2 F_i(w_i)}{\partial \zeta_i^2} \right| : w_i \in \langle \underline{\zeta_i}, \overline{\zeta_i} \rangle \right\}, \tag{3.19}$$

$$c_2 = \min_i \min \left\{ -\frac{\partial F_i(w_i)}{\partial \zeta_i} : w_i \in \langle \underline{\zeta_i}, \overline{\zeta_i} \rangle \right\}. \tag{3.20}$$

Here the first maximum (or minimum) is obtained for the index $i$ ranging all over the gird points whereas the second maximum (or minimum) is evaluated by taking the variable $w_i$ within the sector $\langle \underline{\zeta_i}, \overline{\zeta_i} \rangle$.

Several remarks on the method and convergence results follow.

**Remark 3.1.** We first note that it is impossible to obtain the quadratic convergence if we take the lower solution $\underline{U}$ as the initial guess since the nonlinear functional $Z_{\mathrm{n}}(\zeta_{\mathrm{n}})$ is concave down as shown in (3.12). Moreover, in solving the whole system (2.1)–(2.7) by means of Gummel's iteration, the optimal convergence is linear [24] instead of quadratic since it is impossible to fulfill Condition (4). These indeed agree with our numerical experiments below.

**Remark 3.2.** It is well known that the continuous and the discrete QDD model may admit multiple solutions and that uniqueness can only be proved for the case near the thermal equilibrium state [21,22]. Although the mathematical analysis of the QDD model is in a rather mature state [21], it is mostly based on the small biasing conditions. We are interested in the global convergence of the QCET model in the sense of the upper or lower solution for more general biasing conditions. Following (3.11), Condition (5), and our numerical experiments, the above analysis in discrete setting may motivate a future analysis on the questions of existence, multiple solutions, and uniqueness of the model with practical biasing voltages in continuous setting.

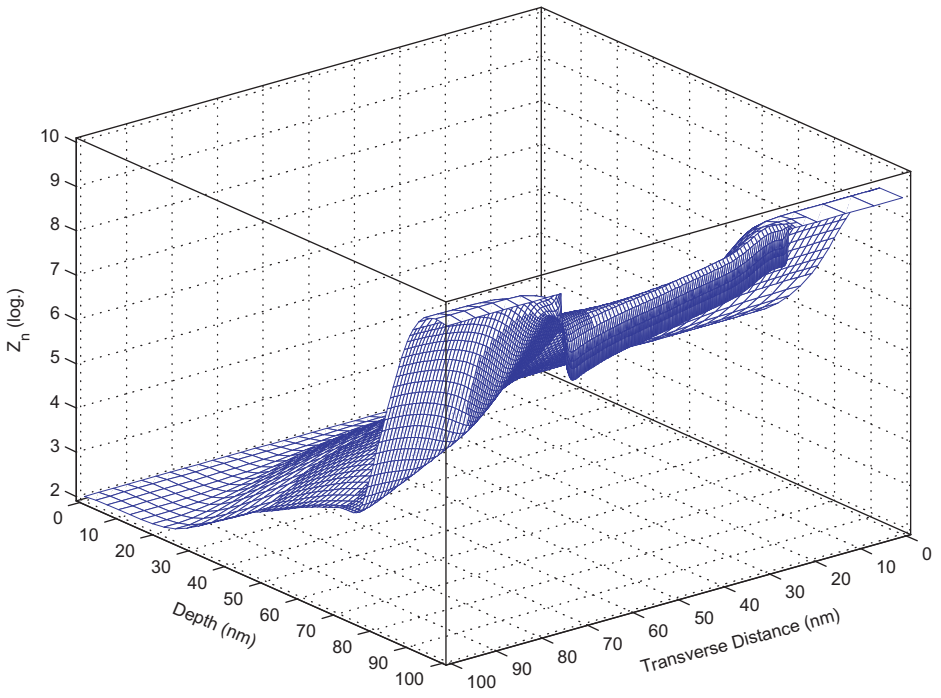


Fig. 2. The numerical solution $\overline{U}_{\mathrm{N}}$ for $\zeta_{\mathrm{n}}$.

**Remark 3.3.** Stability and conditioning are two of most important issues in developing numerical methods. In this paper, we use the standard linear finite elements with the Scharfetter–Gummel fitting scheme for all the unknown functions in (2.1)–(2.7). We refer to [7] for a very detailed account for the stability and conditioning of our methods when compared with other methods such as the mixed FEM for non-self-adjoint DD or ET models. We focus here the novel features on these issues specifically for the DG Eqs. (2.4) and (2.5) as

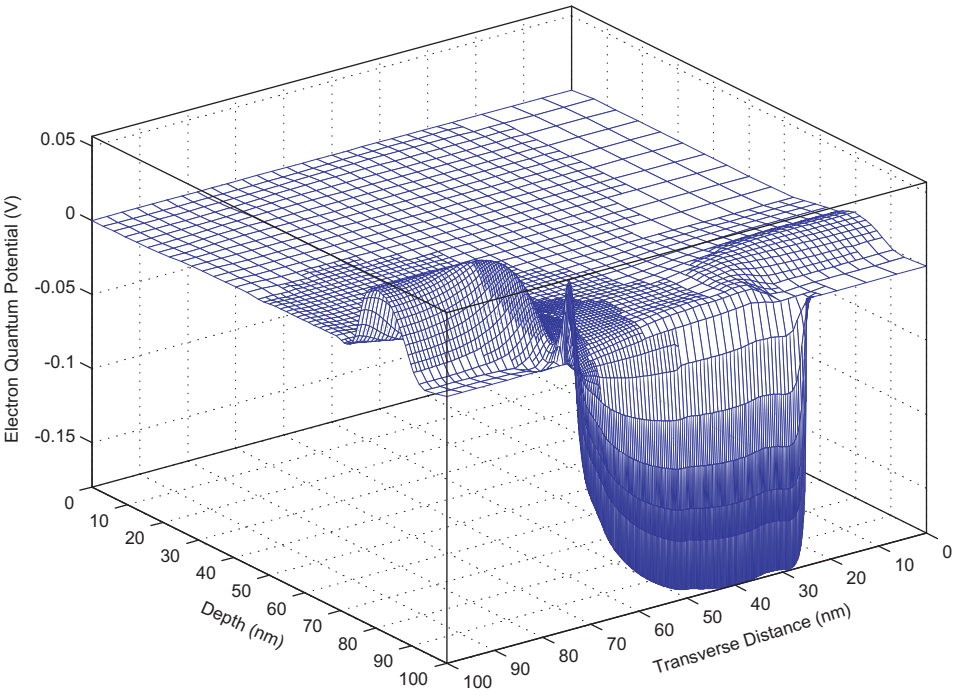


Fig. 3. Electron quantum potential $\phi_{qn}$.

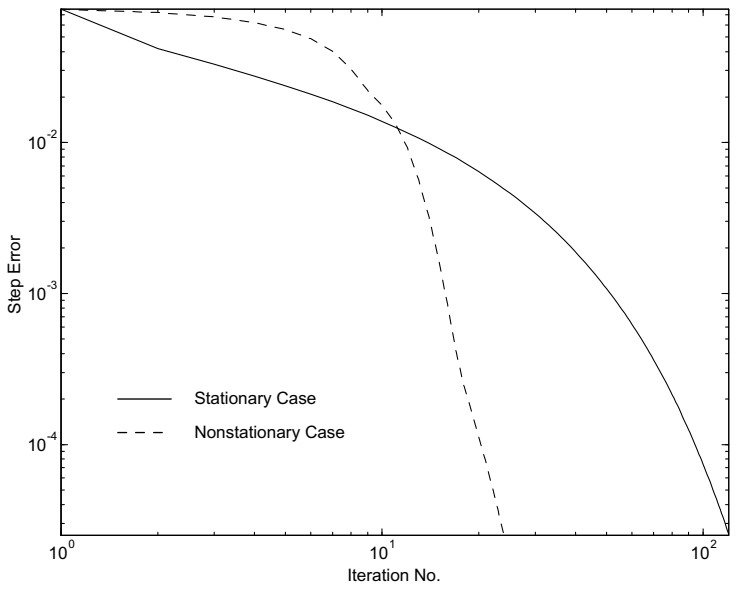


Fig. 4. The step error versus the number of iteration for $\zeta_n$.

Table 1
Number of monotone iterations converging to $\overline{U}_S^*$ and $\overline{U}_N^*$ for $\zeta_n$

| Gummel iteration | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\overline{U}_S^*$ | 120 | 39 | 31 | 23 | 18 | 15 | 13 | 10 | 8 | 6 |
| $\overline{U}_N^*$ | 24 | 12 | 12 | 11 | 10 | 8 | 6 | 5 | 5 | 4 |

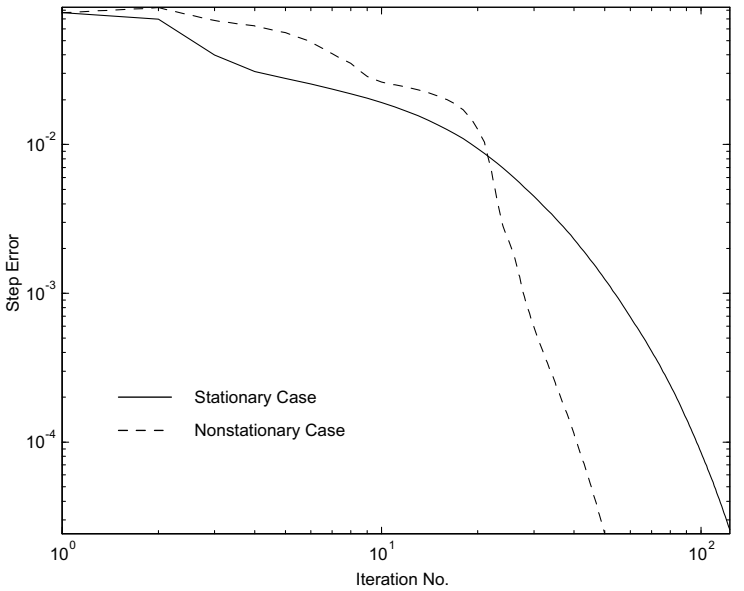


Fig. 5. The step error versus the number of iteration for $\zeta_p$.

Table 2
Number of monotone iterations converging to $\overline{U}_S^*$ and $\overline{U}_N^*$ for $\zeta_p$

| Gummel iteration | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\overline{U}_S^*$ | 123 | 52 | 37 | 29 | 23 | 22 | 22 | 24 | 22 | 22 |
| $\overline{U}_N^*$ | 50 | 22 | 18 | 18 | 16 | 13 | 7 | 3 | 3 | 3 |

Table 3
Rate of convergence of $\overline{U}_N$ for $\zeta_n$

| Monotone iteration ($m$) | $\|\overline{U}_N^{(m)} - U^*\|_\infty$ | $R$ |
|---|---|---|
| 1 | 9.0575e+009 | |
| 2 | 7.7524e+009 | |
| 3 | 4.0921e+009 | 4.10657 |
| 4 | 2.1913e+009 | 0.977493 |
| 5 | 1.1863e+009 | 0.982536 |
| 6 | 6.6188e+008 | 0.950875 |
| 7 | 3.8266e+008 | 0.939037 |
| 8 | 2.2492e+008 | 0.969823 |
| 9 | 1.3271e+008 | 0.992805 |
| 10 | 8.5969e+007 | 0.822967 |
| ... | ... | ... |
| 16 | 9.6363e+006 | 1.00468 |
| 17 | 6.6032e+006 | 1.00371 |
| 18 | 4.5237e+006 | 1.00064 |
| 19 | 3.0936e+006 | 1.00468 |
| 20 | 2.1163e+006 | 0.999135 |
| 21 | 1.4547e+006 | 0.987366 |

compared with the recent results in the literature. First of all, it has been observed in [28] that when solving the QDD model as a coupled system by Newton's method the convergence behavior is even worse than that for solving the Schrödinger–Poisson system in 1D case. In [11], a suitable lumping procedure and a damping parameter are devised for the damped Newton method for solving the QDD model by Gummel's iteration. For each Gummel iteration, a stability criterion for preserving the maximum principle is checked to determine the damping parameter which is a constant value throughout the silicon domain. The convergence rate is, of course, not quadratic and deteriorates as the Bohm potential gets larger [11]. This in principle verifies our

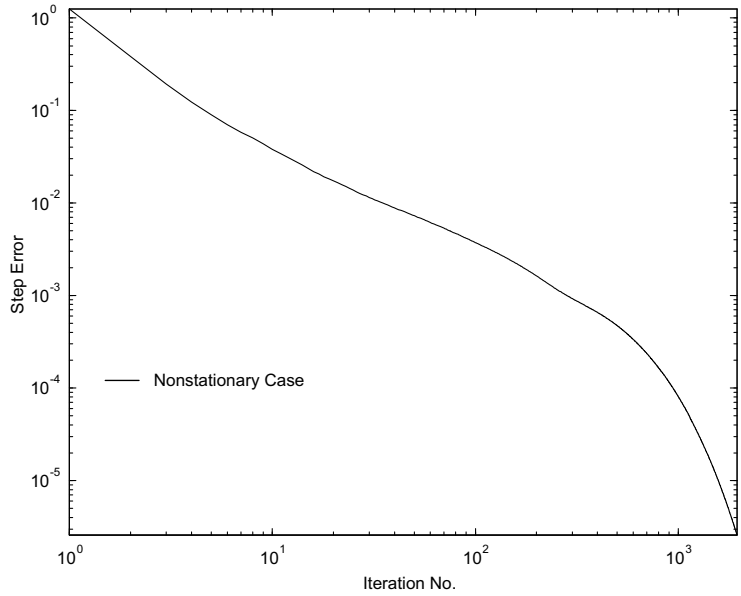


Fig. 6. The step error versus the number of iteration for $\phi$. Only the non-stationary method successfully converged to correct solutions.

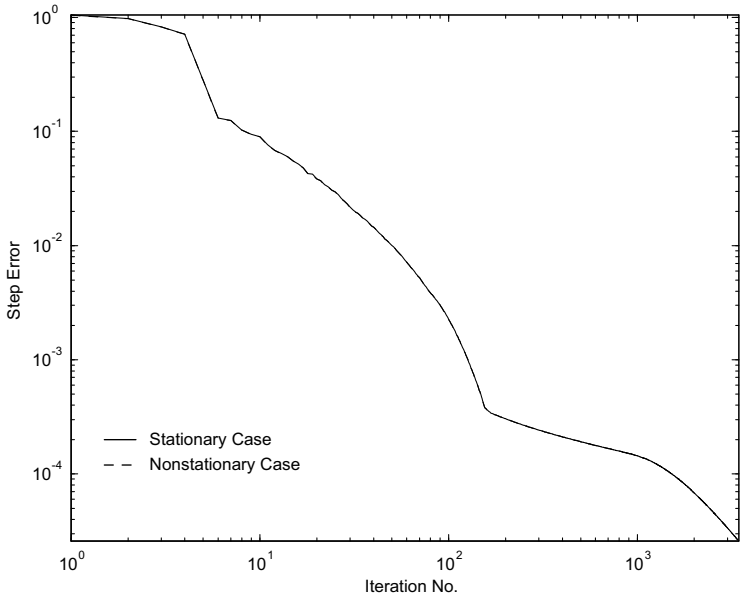


Fig. 7. The step error versus the number of iteration for $u$. Errors are indistinguishable between two cases.

formula (3.11) by which the monotone parameters (3.13) are instead determined not only iteration-by-iteration but also node-by-node (position dependence). The lumping procedure is not required for our method. In short, our method integrates the monotone parameters, grid sizes, and Scharfetter–Gummel fitting in an adaptive and automatic way to treat the singularly perturbed nature of the model.

## 4. Numerical results

In our numerical experiments with the proposed method, we consider the device of Fig. 1 with the applied voltages $V_S = 0$, $V_D = 1.0$ V, $V_B = 0$, and $V_G = 0.8$ V on the source, drain, bulk, and gate contacts, respectively. The device has an elliptical $10^{19}$ cm$^{-3}$ Gaussian doping profiles in the source and drain regions and $10^{16}$cm$^{-3}$ in the $p$-substrate region.

The whole system (2.1)–(2.7) has been solved by using the adaptive finite element and Gummel's iteration algorithms presented in [9] with both stationary and non-stationary monotone iterative methods. We first present the numerical results of the DG Eqs. (2.4) and (2.5) in more details.

The stopping criterion for the outer (Gummel) loop and the inner (monotone) loop is set to $\|e^{(k)}\|_\infty \leqslant 0.1V_T$ and to $\|e^{(m)}\|_\infty \leqslant 0.001V_T$, respectively, where $k = 0, 1, 2, \ldots$ is the Gummel iteration index and $\|e^{(m)}\|_\infty \equiv \|U^{(m)} - U^{(m-1)}\|_\infty$ is called the monotone step error [9]. The initial guesses are taken as either an upper solution $U^{(0)} = \overline{U} = 10^{10}$ cm$^{-6}$ for $\zeta_n$ and $U^{(0)} = \overline{U} = 10^{9}$ cm$^{-6}$ for $\zeta_p$ or a lower solution $U^{(0)} = \underline{U} = 10^{2}$ cm$^{-6}$ for $\zeta_n$ and $U^{(0)} = \underline{U} = 10$ cm$^{-6}$ for $\zeta_p$. Here the notation $\overline{U} = c$ means that all components in the vector $\overline{U}$ are equal to $c$. The stationary monotone parameters (3.6) can be easily computed by the formula (3.10) where the variable $\zeta_n$ is substituted by the upper solution and other variables are determined under the charge neutrality condition. We obtained the convergence sequences as predicted by Theorem 3.1 for both stationary and non-stationary cases. We also ran various cases with different biasing conditions and got similar results as the present one.

The approximate solution of $\zeta_n$ in logarithmic scale and the quantum potential $\phi_{qn}$ at the 10th Gummel iteration are given in Figs. 2 and 3. As shown by Fig. 4 (with Table 1 in numbers) and Fig. 5 (with Table 2), the convergence rate of the non-stationary method is much faster than that of the stationary method.

In order to study the convergence order, we define the rate of convergence as

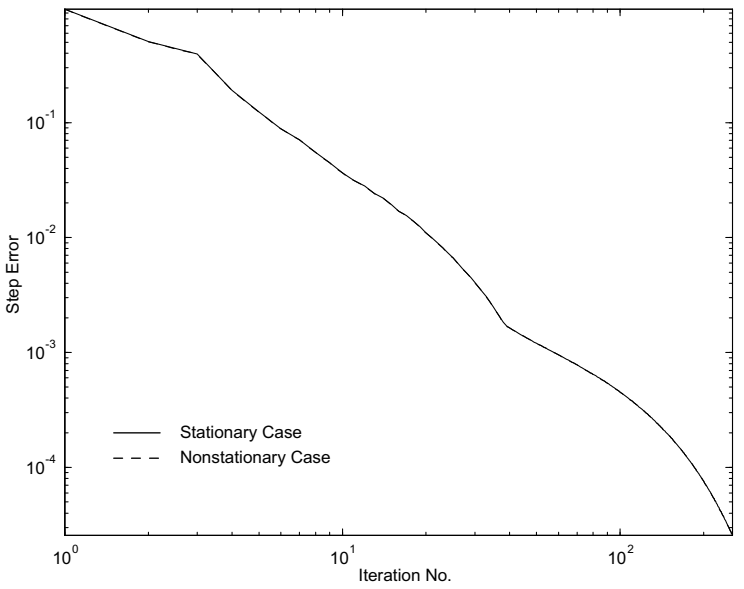


Fig. 8. The step error versus the number of iteration for $v$. Errors are indistinguishable between two cases.

$$R = \frac{\log(\|U_N^{(m-1)} - U^*\|_\infty) - \log(\|U_N^{(m)} - U^*\|_\infty)}{\log(\|U_N^{(m-2)} - U^*\|_\infty) - \log(\|U_N^{(m-1)} - U^*\|_\infty)},$$

where the exact solution is assumed as $U^* \equiv \overline{U}_N^{(24)}$. Table 3 shows that the convergence order is linear as analyzed above.

We next present the convergence results for the rest of the system, namely, Eqs. (2.1)–(2.3), (2.6), (2.7) with the variables $\phi$, $u$, $v$, $g_n$, and $g_p$ as shown in Figs. 6–10, respectively.

From these results, we note that the performance of the stationary and non-stationary methods is almost the same for $u$, $v$, $g_n$, and $g_p$ except for $\phi$ for which the stationary method fails to converge to correct approx-

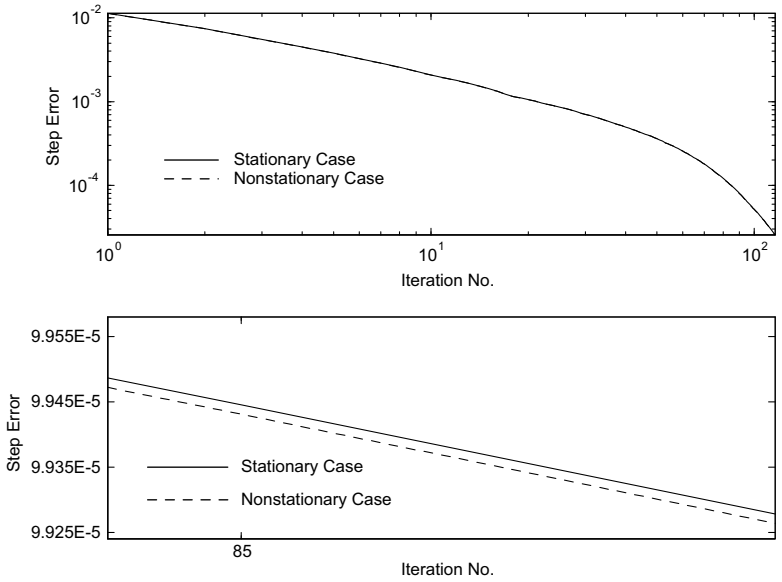


Fig. 9. The step error versus the number of iteration for $g_n$.

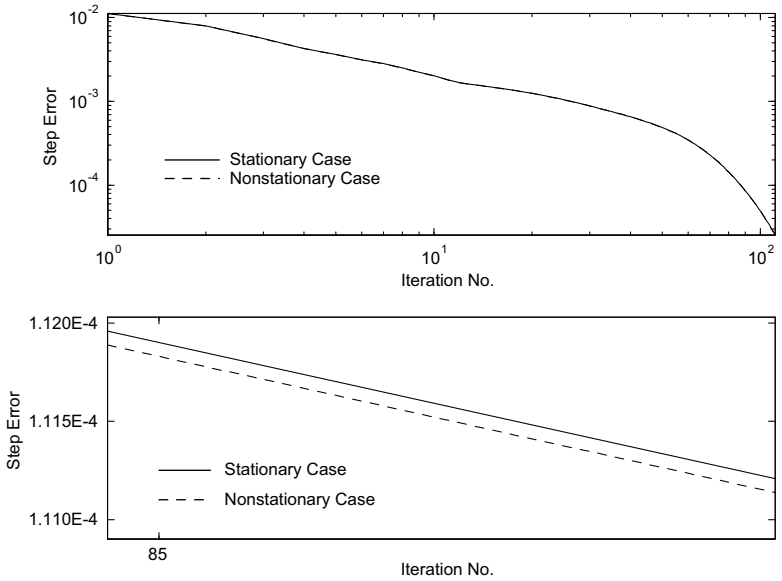


Fig. 10. The step error versus the number of iteration for $g_p$.

imate solutions in our simulations. In other words, we did not obtain physically meaningful solutions for the entire system (2.1)–(2.7) when the stationary method is used. Therefore, the non-stationary method not only leads to correct solutions but also accelerates the computation for the whole system when compared with the stationary method.

## 5. Conclusion

We have proposed and analyzed a non-stationary iterative monotone method for the quantum-corrected energy transport model [9]. It is a global method for the density-gradient equations in the sense that the initial guess can be taken as an upper or a lower solution which is independent of applied voltages. The key component of the method is to set up a selection criterion such that the monotone parameters are the diagonal entries of the Jacobian matrix in most part of the domain and the stationary values in the thin layer part. The method integrates the monotone parameters, grid sizes, and Scharfetter–Gummel fitting in an adaptive and automatic way to treat the singularly perturbed nature of the model. Its convergence rate is optimal in the sense of Gummel's iteration.

## References

[1] M.G. Ancona, H.F. Tiersten, Macroscopic physics of the silicon inversion layer, Phys. Rev. B 35 (1987) 7959.
[2] U. Ascher, P. Markowich, C. Schmeiser, H. Steinruck, R. Weiss, Conditioning of the steady state semiconductor device problem, SIAM J. Appl. Math. 49 (1989) 165–185.
[3] A. Asenov, A.R. Brown, J.H. Davies, S. Kaya, G. Slavcheva, Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs, IEEE Trans. Electron Dev. 50 (2003) 1837–1852.
[4] N. Ben Abdallah, M. Mouis, C. Negulescu, An accelerated algorithm for 2D simulations of the quantum ballistic transport in nanoscale MOSFETs, J. Comput. Phys. 225 (2007) 74–99.
[5] B.A. Biegel, M.G. Ancona, C.S. Rafferty, Z. Yu, Efficient multi-dimensional simulation of quantum confinement effects in advanced MOS devices, NAS Technical Report NAS-04-008, 2004.
[6] G.F. Carey, A.L. Pardhanani, S. Bova, Advanced numerical methods and software approaches for semiconductor device simulation, VLSI Des. 10 (2000) 391–414.
[7] R.-C. Chen, J.-L. Liu, An iterative method for adaptive finite element solutions of an energy transport model of semiconductor devices, J. Comput. Phys. 189 (2003) 579–606.
[8] R.-C. Chen, J.-L. Liu, Monotone iterative methods for the adaptive finite element solution of semiconductor equations, J. Comput. Appl. Math. 159 (2003) 341–364.
[9] R.-C. Chen, J.-L. Liu, A quantum corrected energy transport model for nanoscale semiconductor devices, J. Comput. Phys. 204 (2005) 131–156.
[10] D. Connelly, Z. Yu, D. Yergeau, Macroscopic simulation of quantum mechanical effects in 2-D MOS devices via the density gradient method, IEEE Trans. Electron Dev. 49 (2002) 619–626.
[11] C. de Falco, E. Gatti, A.L. Lacaita, R. Sacco, Quantum-corrected drift-diffusion models for transport in semiconductor devices, J. Comput. Phys. 204 (2005) 533–561.
[12] P. Degond, S. Gallego, F. Méhats, An entropic quantum drift-diffusion model for electron transport in resonant tunneling diodes, J. Comput. Phys. 221 (2007) 226–249.
[13] P. Degond, C. Ringhofer, Quantum moment hydrodynamics and the entropy principle, J. Stat. Phys. 112 (2003) 587–628.
[14] A. Jüngel, A. El Ayyadi, Semiconductor simulations using a coupled quantum drift-diffusion Schrödinger–Poisson model, SIAM J. Appl. Math. 66 (2005) 554–572.
[15] M. Karner, A. Gehring, S. Holzer, M. Pourfath, M. Wagner, W. Goes, M. Vasicek, O. Baumgartner, C. Kernstock, K. Schnass, G. Zeiler, T. Grasser, H. Kosina, S. Selberherr, A multi-purpose Schrödinger–Poisson solver for TCAD applications, J Comput. Electron 6 (2007) 179–182.
[16] P.A. Markowich, C.A. Ringhofer, S. Selberherr, M. Lentini, A singular perturbation approach for the analysis of the fundamental semiconductor equations, IEEE Trans. Electron Dev. ED-30 (1983) 1165–1180.
[17] A. Martinez, J.R. Barker, A. Asenov, A. Svizhenko, M. Bescond, A. Anantram, Development of a full 3D NEGF nano-CMOS simulator, IEEE SISPAD (2006) 353–356.
[18] S. Odanaka, Multidimensional discretization of the stationary quantum drift-diffusion model for ultrasmall MOSFET structures, IEEE Trans. Comput.-Aided Des. Integr. Circ. Syst. 23 (2004) 837–842.
[19] C.V. Pao, Accelerated monotone iterative methods for finite difference equations of reaction–diffusion, Numer. Math. 79 (1998) 261–281.
[20] C.V. Pao, Accelerated monotone iterations for numerical solutions of nonlinear elliptic boundary value problems, Comput. Math. Appl. 46 (2003) 1535–1544.
[21] R. Pinnau, A review on the quantum drift diffusion model, Transp. Theory Stat. Phys. 31 (2002) 367–395.

[22] R. Pinnau, A. Unterreiter, The stationary current–voltage characteristics of the quantum drift diffusion model, SIAM J. Numer. Anal. 37 (1999) 211–245.
[23] C.S. Rafferty, B. Biegel, Z. Yu, M.G. Ancona, J. Bude, R.W. Dutton, Multi-dimensional quantum effect simulation using a density-gradient model and script-level programming techniques, in: Proc. SISPAD, 1998, pp. 137–140.
[24] C. Ringhofer, C. Schmeiser, An approximate Newton method for the solution of the basic semiconductor device equations, SIAM J. Numer. Anal. 26 (1989) 507–516.
[25] S. Roy, A. Asenov, Where do the dopants go? Science 309 (2005) 388–390.
[26] D. Vasileska, S.S. Ahmed, Narrow-width SOI devices: the role of quantum–mechanical size quantization effect and unintentional doping on the device operation, IEEE Trans. Electron Dev. 52 (2005) 227–236.
[27] Y.-M. Wang, On accelerated monotone iterations for numerical solutions of semilinear elliptic boundary value problems, Appl. Math. Lett. 18 (2005) 749–755.
[28] A. Wettstein, A. Schenk, W. Fichtner, Quantum device-simulation with the density-gradient model on unstructured grids, IEEE Trans. Electron Dev. 48 (2001) 279–284.