Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 204 (2005) 131–156
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

A quantum corrected energy-transport model for nanoscale semiconductor devices

Ren-Chuen Chen, Jinn-Liang Liu *

Department of Applied Mathematics, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu 300, Taiwan

Received 29 June 2004; received in revised form 4 October 2004; accepted 4 October 2004

Abstract

An energy transport model coupled with the density gradient method as quantum mechanical corrections is proposed and numerically investigated. This new model is comprehensive in both physical and mathematical aspects. It is capable of describing hot electron transport as well as significant quantum mechanical effects for advanced devices with dimensions comparable to the de Broglie wave-length. The model is completely self-adjoint for all state variables and hence provides many appealing mathematical features such as global convergence, fast iterative solution, and highly parallelizable. Numerical simulations on diode and MOSFET with the gate length down to 34 nm using this model have been performed and compared with that using the classical transport model. It is shown that the I–V characteristics of this short-channel device is significantly corrected by the density-gradient equations with current drive reduced by up to 60% comparing with that of the classical model along. Moreover, a 2D quantum layer, which is only a fraction of the length scale of inversion layer, is also effectively captured by this new model with very fine mesh near the interface produced by an adaptive finite element method.

1. Introduction

Numerical simulation of charge transport in device structures is widely used for analysis of physical processes in the semiconductor devices and estimation of their electrical parameters. The major part of the activities in this field is based on drift-diffusion (DD) equations. However, there is a growing realization that technologist cannot ignore quantum effects much longer. The combination of thin gate oxides and heavy doping in the conventional MOSFETs, and the thin silicon body of the double-gate structures, will result in substantial quantum mechanical (QM) threshold voltage shift and transconductance degradation

* Corresponding author.
E-mail address: jinnliu@math.nctu.edu.tw (J.-L. Liu).

0021-9991/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2004.10.006

[21]. Computationally efficient methods to include QM effects are required for the purpose of practical Computer Aided Design of this generation of devices.

Some numerical methods employing full quantum models such as non-equilibrium Green's function [23] and Wigner's function [9] suffer from unsolved robustness problems and are still much too costly for device or circuit simulations. Another approach for including QM effects is to add quantum corrections to classical models [4,3,15–20,22,28,30,32,33]. In particular, the density gradient (DG) model developed by Ancona et al. is a more rigorous macroscopic transport model which avoids ad hoc assumption to the material parameters or imposing an artificial shape function [34]. It is demonstrated in [1,7,8,29] that this model is feasible and efficient to accurately and generally simulate multi-dimensional devices with gate lengths ranging from 30 nm down to 6 nm when combined with the DD model.

In this paper, we further extend the DG model to combine with the energy transport (ET) model proposed in our previous work [11] and show that this new combination (DGET) is capable of describing hot electron transport as well as significant QM effects for advanced devices. Our model is able to explain that electron temperature essentially differs from the lattice temperature. It is clear that this effect cannot be described by the DG model along. Quantum hydrodynamic (QHD) models give accurate simulation results, but the numerical methods to solve this system are too costly and time consuming to model real problems in semiconductor production mode where simulation results are needed in hours or minutes. The DGET model is of parabolic type so that its numerical solution needs less effort than QHD models which contain hyperbolic modes.

Moreover, our model is completely self-adjoint for all state variables and hence provides many appealing mathematical features such as global convergence, fast iterative solution, and highly parallelizable as demonstrated in our previous papers [11,12,24]. The global convergence is a consequence of monotone iterative methods used in solving the discrete systems of nonlinear algebraic equations resulting from adaptive finite element approximation of the model. It is shown here that the convergence analysis of these methods given in [11,12] can be straightforwardly carried over to the present model. Our numerical experiments on various device structures with high drain bias have shown that the monotone iteration do not suffer from the convergence difficulties as frequently encountered by the commonly used Newton's iteration since the Jacobian is either close to singular or poorly conditioned [29]. This is a fundamental issue constantly faced by the practitioner in device and circuit modeling. Numerical simulations on diode and MOSFET with the gate length down to 34 nm using the DGET model have been performed and compared with that using the ET model. It is shown that the I–V characteristics of this short-channel device is significantly corrected by the density-gradient equations with current drive reduced by up to 60% compared with that of the classical model along. Moreover, a 2D quantum layer, which is only a fraction of the length scale of inversion layer, is also effectively captured by this new model with very fine mesh near the interface produced by the adaptive finite element method.

The paper is divided into the following sections: Section 2 briefly recalls the ET model considered in [11] and the DG model. A full self-adjoint formulation of both models is then given in Section 3. For the sake of clearness, we also extend our previous adaptive finite-element algorithm [11] to the present model in Section 4. In Section 5, numerical results of simulation on various diodes to compare with the results in the literature and on MOSFET device structures to demonstrate the effectiveness of the proposed model. A short concluding remark is given in Section 6.

2. The energy transport and density gradient models

As in [11], we consider the following ET model

$$\Delta\phi = \frac{q}{\varepsilon_s}(n - p + N_A^- - N_D^+), \tag{1}$$

$$\frac{1}{q}\nabla\cdot\mathbf{J}_{\mathrm{n}} = R, \tag{2}$$

$$\frac{1}{q}\nabla\cdot\mathbf{J}_{\mathrm{p}} = -R, \tag{3}$$

$$\nabla\cdot\mathbf{S}_{\mathrm{n}} = \mathbf{J}_{\mathrm{n}}\cdot\mathbf{E} - n\left(\frac{\omega_{\mathrm{n}}-\omega_0}{\tau_{\mathrm{n}\omega}}\right), \tag{4}$$

$$\nabla\cdot\mathbf{S}_{\mathrm{p}} = \mathbf{J}_{\mathrm{p}}\cdot\mathbf{E} - p\left(\frac{\omega_{\mathrm{p}}-\omega_0}{\tau_{\mathrm{p}\omega}}\right), \tag{5}$$

where ϕ is the electrostatic potential, n and p are the electron and hole concentrations, q is the elementary charge, ε_{s} is the permittivity constant of semiconductor, N_{A}^{-} and N_{D}^{+} are the densities of ionized impurities, $\mathbf{J}_{\mathrm{n}}$ and $\mathbf{J}_{\mathrm{p}}$ are the current densities, R is the function describing the balance of generation and recombination of electrons and holes, $\mathbf{S}_{\mathrm{n}}$ and $\mathbf{S}_{\mathrm{p}}$ are the energy fluxes for carriers, $\mathbf{E}$ is the electric field, $\tau_{\mathrm{n}\omega}$ and $\tau_{\mathrm{p}\omega}$ are the carrier energy relaxation times, ω_0 is the thermal energy, and ω_{n} and ω_{p} are the carrier average energies. These physical variables are tightly coupled together with the following auxiliary relationships

$$\mathbf{E} = -\nabla\phi, \tag{6}$$

$$\mathbf{J}_{\mathrm{n}} = -q\mu_{\mathrm{n}}n\nabla\phi + qD_{\mathrm{n}}\nabla n = -qn\mathbf{v}_{\mathrm{n}}, \tag{7}$$

$$\mathbf{J}_{\mathrm{p}} = -q\mu_{\mathrm{p}}p\nabla\phi - qD_{\mathrm{p}}\nabla p = qp\mathbf{v}_{\mathrm{p}}, \tag{8}$$

$$\mathbf{S}_{\mathrm{n}} = \frac{\mathbf{J}_{\mathrm{n}}}{-q}\omega_{\mathrm{n}} + \frac{\mathbf{J}_{\mathrm{n}}}{-q}k_{\mathrm{B}}T_{\mathrm{n}} + \mathbf{Q}_{\mathrm{n}}, \tag{9}$$

$$\mathbf{S}_{\mathrm{p}} = \frac{\mathbf{J}_{\mathrm{p}}}{+q}\omega_{\mathrm{p}} + \frac{\mathbf{J}_{\mathrm{p}}}{+q}k_{\mathrm{B}}T_{\mathrm{p}} + \mathbf{Q}_{\mathrm{p}}, \tag{10}$$

$$\omega_0 = \frac{3}{2}k_{\mathrm{B}}T_{\mathrm{L}}, \tag{11}$$

$$\omega_{\mathrm{n}} = \frac{3}{2}k_{\mathrm{B}}T_{\mathrm{n}} + \frac{1}{2}m_{\mathrm{n}}^{*}|\mathbf{v}_{\mathrm{n}}|^2, \tag{12}$$

$$\omega_{\mathrm{p}} = \frac{3}{2}k_{\mathrm{B}}T_{\mathrm{p}} + \frac{1}{2}m_{\mathrm{p}}^{*}|\mathbf{v}_{\mathrm{p}}|^2, \tag{13}$$

$$\mathbf{Q}_{\mathrm{n}} = -\kappa_{\mathrm{n}}\nabla T_{\mathrm{n}}, \tag{14}$$

$$\mathbf{Q}_{\mathrm{p}} = -\kappa_{\mathrm{p}}\nabla T_{\mathrm{p}}, \tag{15}$$

$$\kappa_{\mathrm{n}} = 2\left(\frac{k_{\mathrm{B}}}{q}\right)^2 nq\mu_{\mathrm{n}}T_{\mathrm{L}}, \tag{16}$$

$$\kappa_{\mathrm{p}} = 2\left(\frac{k_{\mathrm{B}}}{q}\right)^2 pq\mu_{\mathrm{p}}T_{\mathrm{L}}, \tag{17}$$

$$R = \frac{np - n_i^2}{\tau_n^0(p + p_T) + \tau_p^0(n + n_T)}, \tag{18}$$

where $\mathbf{Q}_n$ and $\mathbf{Q}_p$ are the heat fluxes for carries, k_B is Boltzmann's constant, T_n, T_p, and T_L are the electron, hole and lattice temperatures, μ_n and μ_p are the field-dependent electron and hole mobilities, D_n and D_p are the electron and hole diffusion coefficients expressed by the Einstein relation with the mobilities, m_n^* and m_p^* are the electron and hole effective masses, $\mathbf{v}_n$ and $\mathbf{v}_p$ are the electron and hole velocities, κ_n and κ_p are the electron and hole heat conductivities, and (18) is the Shockley–Read–Hall (SHR) generation-recombination model with n_i being the intrinsic carrier concentration, τ_n^0 and τ_p^0 the electron and hole lifetimes, and p_T and n_T the electron and hole densities associated with energy levels of the traps. In the above equations, vectors are denoted by bold letters.

The DG theory was developed by observing that the electron gas is energetically sensitive not only to its density but also to the gradient of the density. It captures the nonlocality of quantum mechanics to the lowest-order of $\hbar^2$ where $\hbar$ is the reduced Planck constant and can be rigorously derived from quantum mechanics [4,3]. Specifically, a third order derivative term of quantum correction is added to the carrier current density as

$$\mathbf{J}_n = -q\mu_n n\nabla\phi + qD_n\nabla n - 2q\mu_n b_n n\nabla\left[\frac{\Delta\sqrt{n}}{\sqrt{n}}\right], \tag{19}$$

$$\mathbf{J}_p = -q\mu_p p\nabla\phi - qD_p\nabla p + 2q\mu_p b_p p\nabla\left[\frac{\Delta\sqrt{p}}{\sqrt{p}}\right], \tag{20}$$

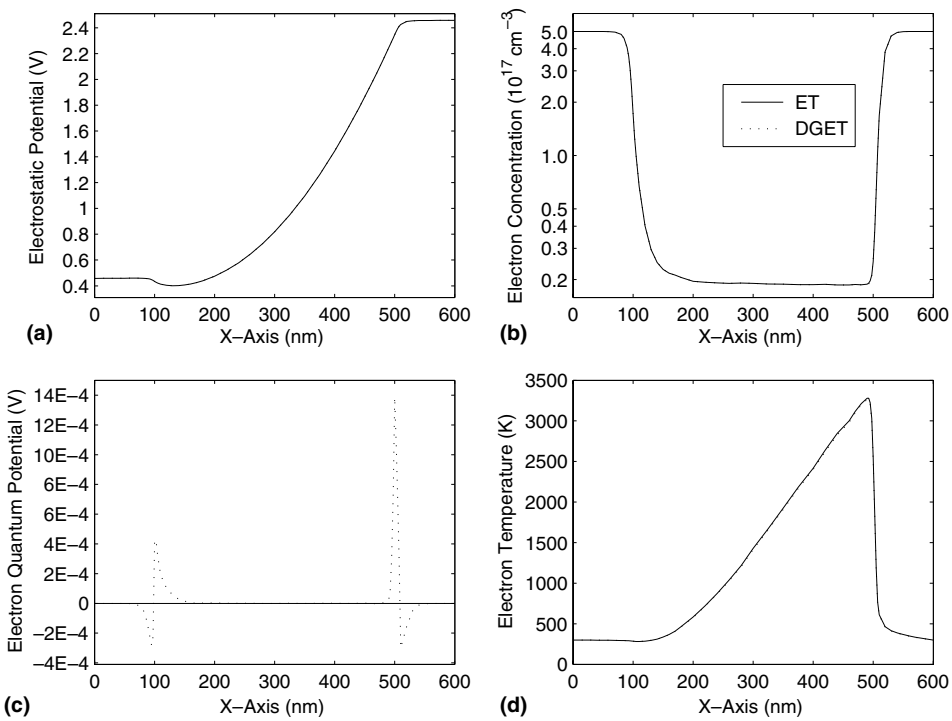

Fig. 1. The numerical results of the 600 nm silicon diode.

where the coefficients $b_n = \frac{\hbar^2}{12m_n^* q}$ and $b_p = \frac{\hbar^2}{12m_p^* q}$ are the material parameters measuring the strength of the gradient effects in the gas. To alleviate the difficulty in discretization caused by this higher order term, additional variables called the quantum potentials

$$\phi_{qn} \equiv 2b_n \left[\frac{\Delta\sqrt{n}}{\sqrt{n}}\right], \tag{21}$$

$$\phi_{qp} \equiv -2b_p \left[\frac{\Delta\sqrt{p}}{\sqrt{p}}\right] \tag{22}$$

have been introduced in [29] and thus can be lumped with the classical drift term to obtain

$$\mathbf{J}_n = -q\mu_n n\nabla(\phi + \phi_{qn}) + qD_n\nabla n, \tag{23}$$

$$\mathbf{J}_p = -q\mu_p p\nabla(\phi + \phi_{qp}) - qD_p\nabla p. \tag{24}$$

We thus have a complete set of seven PDEs (1)–(5) and (21) and (22) describing both ET and DG models with the seven state variables ϕ, n, p, ϕ_{qn}, ϕ_{qp}, T_n, and T_p.

Note that the coefficients in (21) and (22) result in a boundary layer near the silicon/silicon-oxide interface for short-channel devices. The layer is only a fraction of the length scale of the inversion layer, in which the electron density typically drops from its peak value of order 10^{18} at about 0.5–1.5 nm away from the interface to zero at the interface [1,7]. Numerical treatments for this boundary layer problem are evidently subtle and challenging. A more detailed description of our approach to this problem will be given in Section 5.

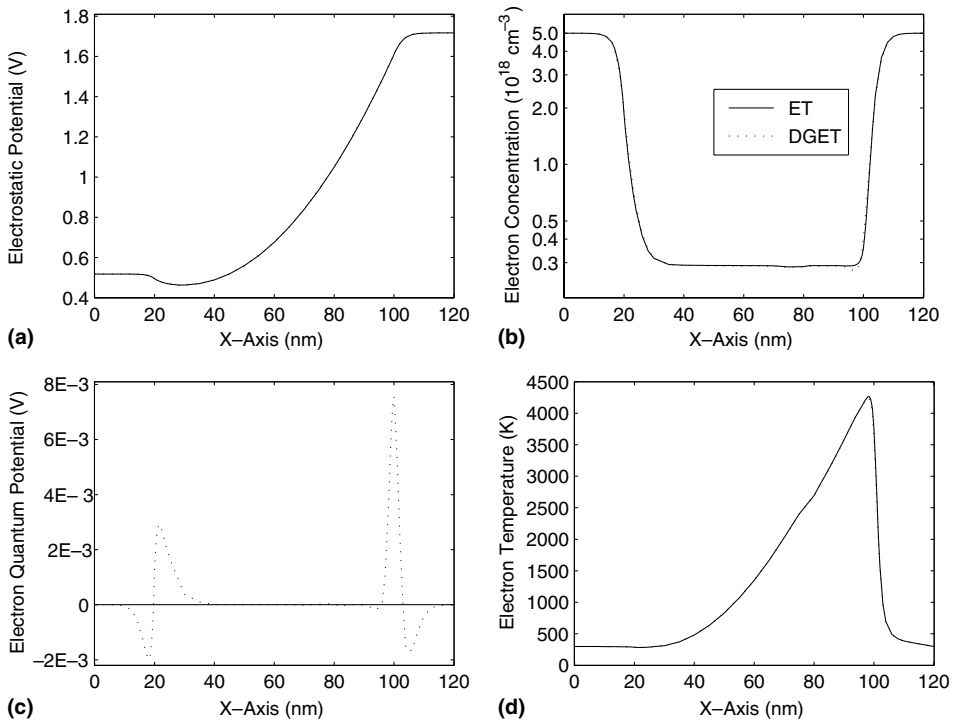

Fig. 2. The numerical results of the 120 nm silicon diode.

Remark 2.1. Taking first three moments of the Boltzmann transport equation (BTE) with conservation of particles, momentum, and energy, the classical hydrodynamic (CHD) model can be expressed as (for simplicity, we list the equations of electrons only) [6,17]:

$$\frac{\partial n}{\partial t} + \nabla \cdot (n\mathbf{v}_n) = \left(\frac{\partial n}{\partial t}\right)_c,$$

$$\frac{\partial \mathbf{p}_n}{\partial t} + \mathbf{v}_n \nabla \cdot \mathbf{p}_n + \mathbf{p}_n \cdot \nabla \mathbf{v}_n = -qn\mathbf{E} - \nabla(nk_B T) + \left(\frac{\partial \mathbf{p}_n}{\partial t}\right)_c,$$

$$\frac{\partial \omega_n}{\partial t} + \nabla \cdot (\mathbf{v}_n \omega_n) = -qn\mathbf{v}_n \cdot \mathbf{E} - \nabla \cdot (\mathbf{v}_n nk_B T) - \nabla \cdot \mathbf{Q}_n + \left(\frac{\partial \omega_n}{\partial t}\right)_c,$$

where $\mathbf{p}_n = m_n^* n\mathbf{v}_n$ is the momentum density. Considering the steady state and employing the collision terms

$$\left(\frac{\partial \mathbf{p}_n}{\partial t}\right)_c = -\frac{\mathbf{p}_n}{\tau_{pn}},$$

$$\left(\frac{\partial \omega_n}{\partial t}\right)_c = -\frac{\omega_n - \omega_0}{\tau_{n\omega}},$$

we have [11]

$$\mathbf{J}_n = q\mu_n \left[\frac{k_B T_n}{q}\nabla n + n\nabla\left(\frac{k_B T_n}{q} - \phi\right)\right]$$

and Eq. (4). Similarly the three conservation equations of the QHD model are [16,18]

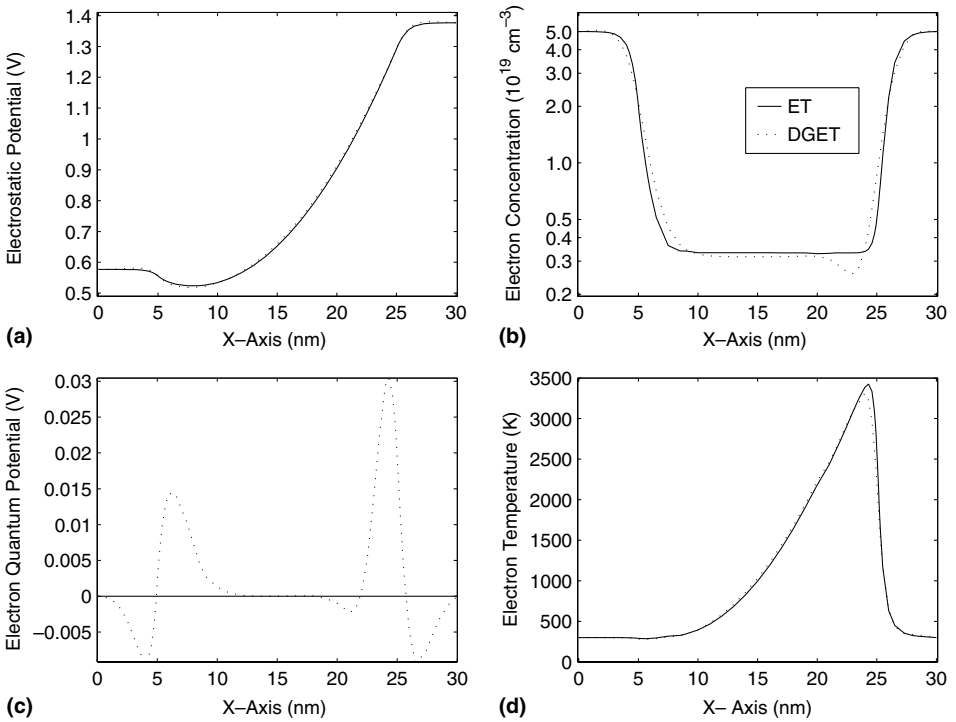

Fig. 3. The numerical results of the 30 nm silicon diode.

$$\frac{\partial n}{\partial t} + \nabla \cdot (n\mathbf{v}_n) = \left(\frac{\partial n}{\partial t}\right)_c,$$
$$\frac{\partial \mathbf{p}_n}{\partial t} + \mathbf{v}_n \nabla \cdot \mathbf{p}_n + \mathbf{p}_n \cdot \nabla \mathbf{v}_n + \frac{n}{3}\nabla Q = -qn\mathbf{E} - \nabla(nk_B T) + \left(\frac{\partial \mathbf{p}_n}{\partial t}\right)_c,$$
$$\frac{\partial \omega_n}{\partial t} + \nabla \cdot (\mathbf{v}_n \omega_n) = -qn\mathbf{v}_n \cdot \mathbf{E} - \nabla \cdot (\mathbf{v}_n nk_B T) - \nabla \cdot \mathbf{Q}_n + \left(\frac{\partial \omega_n}{\partial t}\right)_c.$$

The quantum correction to the momentum equation is related to the quantum potential of Bohm [27]

$$Q = -\frac{\hbar^2}{2m_n^*}\frac{\Delta\sqrt{n}}{\sqrt{n}},$$

and to the energy density given by

$$\omega_n = \frac{3}{2}k_B T_n + \frac{1}{2}m_n^*|\mathbf{v}_n|^2 - \frac{\hbar^2 n}{24m_n^*}\Delta\log(n).$$

Following the previous deductive procedure the quantum correction current density equation is

$$\begin{aligned}\mathbf{J}_n &= q\mu_n\left[\frac{k_B T_n}{q}\nabla n + n\nabla\left(\frac{k_B T_n}{q} - \phi\right) - \frac{\hbar^2}{6m_n^* q}\nabla\left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right)\right] \\ &= -q\mu_n n\nabla(\phi + \phi_{qn}) + qD_n\nabla n + \mu_n k_B n\nabla T_n.\end{aligned}$$

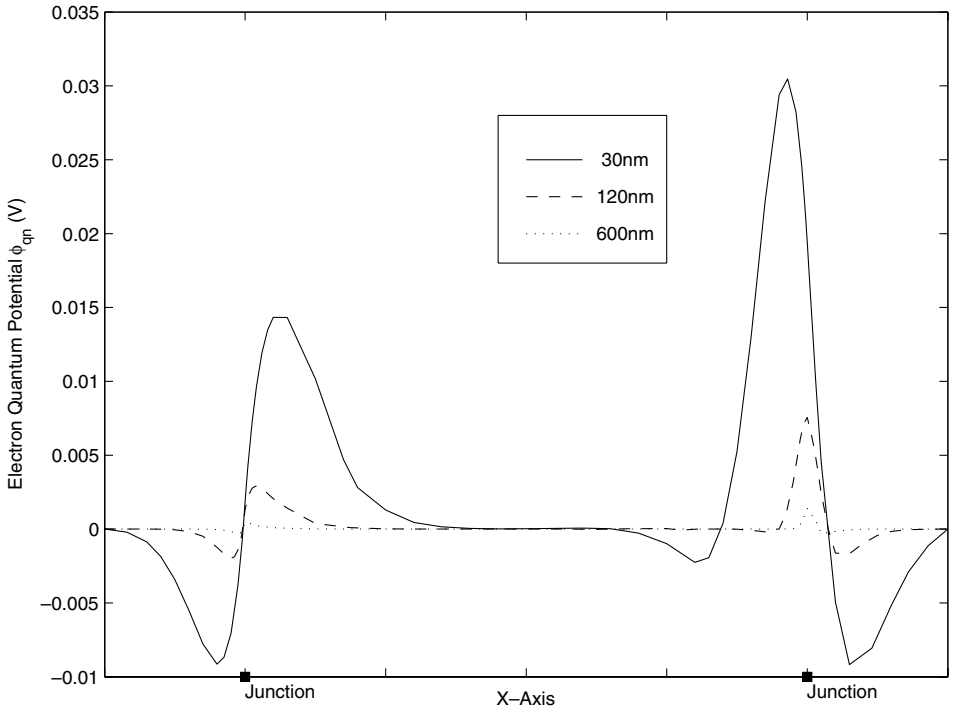

Fig. 4. The tendency of the quantum potential variation.

Compared to the DG model, there is a mechanism that can cause an increase in diffusion, i.e., particle diffusion is enhanced when T_n is significantly greater than T_L. Once we obtain the information of T_n from the DGET model, we will use this formulation to estimate the drain current and to sketch the I–V curves. On the other hand, since we do not add the quantum correction to the energy density, the difference of the temperature distribution between the ET and DGET models is not very significant.

Remark 2.2. A quantum energy balance model appears to be first proposed by Grubin and Kreskovsky in [18] for 1D mesoscopic structures. In their model, quantum correction terms are explicitly included in the carrier average energies (11) and (12). As a result, third order derivative terms of correction are associated not only with the carrier densities (see (19) and (20)) but also with the carrier energies. Putting these correction terms into our model, i.e. into (9) and (10), we will obtain a product of the correction terms in (9) and (10) which obviously makes computations more formidable for 2D simulation. Instead, the correction terms in our model are only explicitly added to the carrier density. The energy balance equations are implicitly and thus less corrected by the quantum effects via the carrier current densities.

3. A self-adjoint formulation of the DGET model

PDEs in self-adjoint form are analytically as well as numerically appealing. In [11,12], we give a rather thorough study of the self-adjoint DD and ET models in terms of mathematical analysis and numerical justification. We now consider the self-adjoint formulation of the above DGET model and, for this purpose, introduce the following new variables

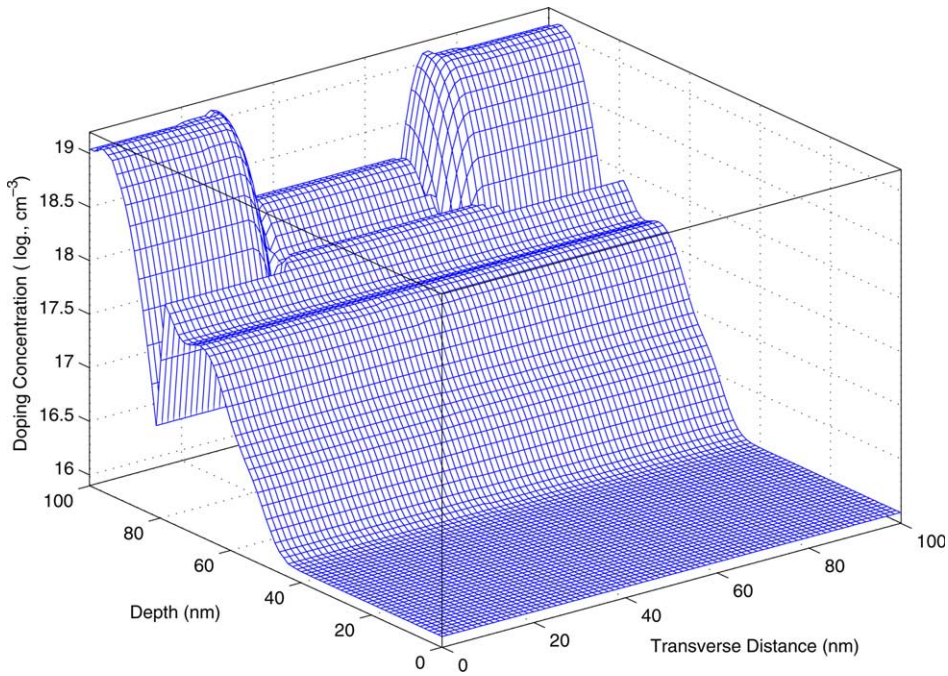

Fig. 5. Doping concentration.

$$u = \exp\left(\frac{-\varphi_n}{V_T}\right), \tag{25}$$

$$v = \exp\left(\frac{\varphi_p}{V_T}\right), \tag{26}$$

$$\zeta_n = \sqrt{n}, \tag{27}$$

$$\zeta_p = \sqrt{p}, \tag{28}$$

$$g_n = T_n / \exp\left(\frac{5\varphi_n}{4V_T}\right), \tag{29}$$

$$g_p = T_p / \exp\left(-\frac{5\varphi_p}{4V_T}\right), \tag{30}$$

where $V_T = (k_B T_L)/q$ is the thermal voltage and φ_n and φ_p are the generalized quasi-Fermi potentials that include the QM effects as shown below. Assuming a Maxwell–Boltzmann energy distribution of carriers, we have the quantum correction expressions of the carriers

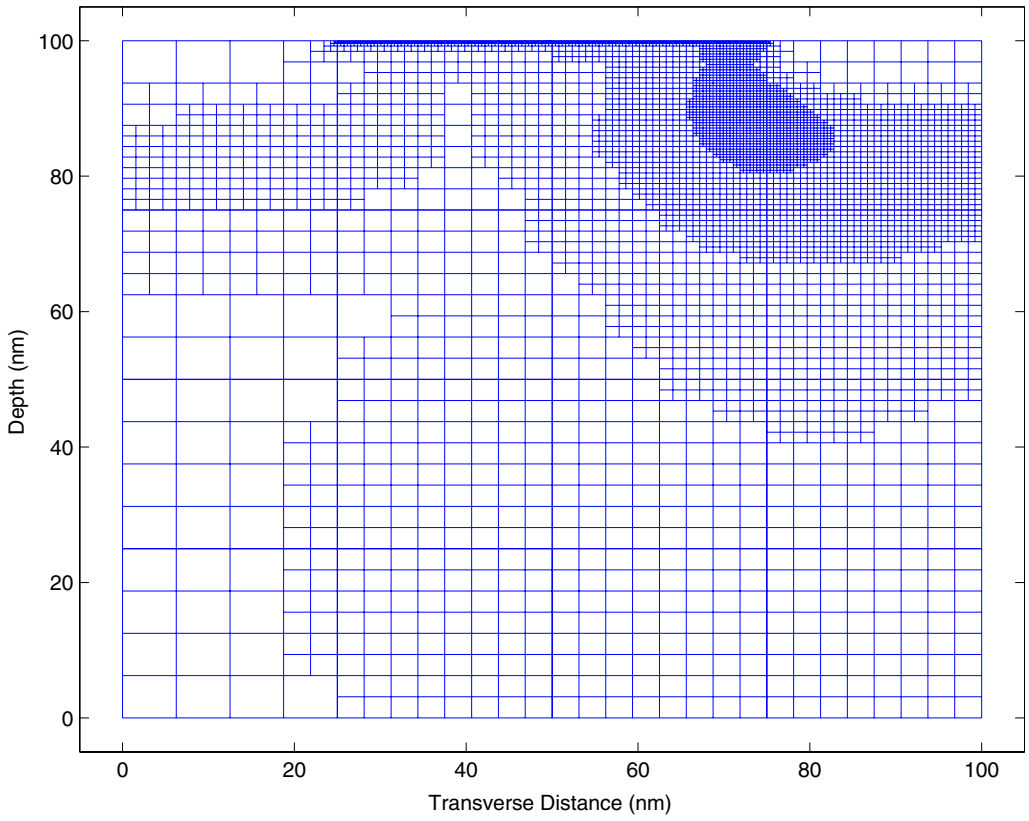

Fig. 6. The final adaptive mesh.

$$n = n_{\mathrm{i}} \exp\left(\frac{\phi - \varphi_{\mathrm{n}} + \phi_{\mathrm{qn}}}{V_{\mathrm{T}}}\right) = n_{\mathrm{i}} \exp\left(\frac{\phi + \phi_{\mathrm{qn}}}{V_{\mathrm{T}}}\right) u = \zeta_{\mathrm{n}}^2, \tag{31}$$

$$p = n_{\mathrm{i}} \exp\left(\frac{\varphi_{\mathrm{p}} - \phi - \phi_{\mathrm{qp}}}{V_{\mathrm{T}}}\right) = n_{\mathrm{i}} \exp\left(\frac{-\phi - \phi_{\mathrm{qp}}}{V_{\mathrm{T}}}\right) v = \zeta_{\mathrm{p}}^2, \tag{32}$$

and rewrite the quantum potentials as

$$\phi_{\mathrm{qn}} = V_{\mathrm{T}} \ln\left(\frac{\zeta_{\mathrm{n}}^2}{u n_{\mathrm{i}}}\right) - \phi, \tag{33}$$

$$\phi_{\mathrm{qp}} = -V_{\mathrm{T}} \ln\left(\frac{\zeta_{\mathrm{p}}^2}{v n_{\mathrm{i}}}\right) - \phi. \tag{34}$$

For Eq. (1) we have

$$\Delta\phi = F(\phi, u, v, \zeta_{\mathrm{n}}, \zeta_{\mathrm{p}}), \tag{35}$$

where

$$F(\phi, u, v, \zeta_{\mathrm{n}}, \zeta_{\mathrm{p}}) = \frac{q n_{\mathrm{i}}}{\varepsilon_{\mathrm{s}}}\left[u \exp\left(\frac{\phi + \phi_{\mathrm{qn}}}{V_{\mathrm{T}}}\right) - v \exp\left(\frac{-\phi - \phi_{\mathrm{qp}}}{V_{\mathrm{T}}}\right)\right] + \frac{q(N_{\mathrm{A}}^- - N_{\mathrm{D}}^+)}{\varepsilon_{\mathrm{s}}}. \tag{36}$$

Substituting (31) into the electron current equation (23), we obtain

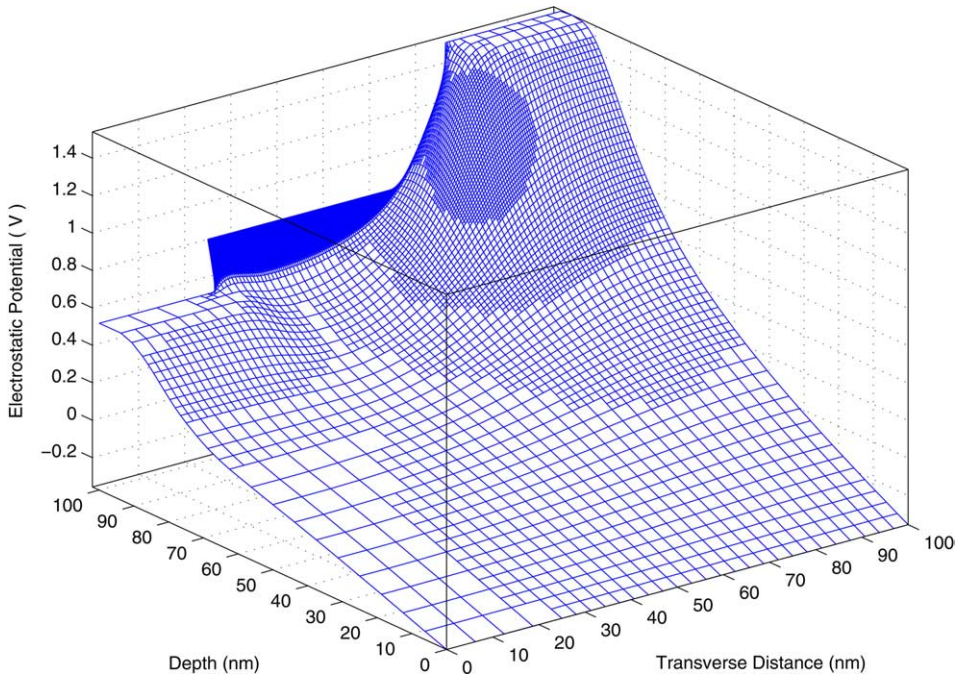

Fig. 7. Electrostatic potential.

$$\mathbf{J}_n = -q\mu_n n\nabla(\phi+\phi_{qn}) + qD_n\nabla\left[n_i\exp\left(\frac{\phi+\phi_{qn}}{V_T}\right)u\right] \tag{37}$$

$$= -q\mu_n n\nabla(\phi+\phi_{qn}) + q\frac{D_n}{V_T}\left[n_i\exp\left(\frac{\phi+\phi_{qn}}{V_T}\right)u\right]\nabla(\phi+\phi_{qn}) + qD_n\left[n_i\exp\left(\frac{\phi+\phi_{qn}}{V_T}\right)\right]\nabla u$$

$$= qD_n n_i\exp\left(\frac{\phi+\phi_{qn}}{V_T}\right)\nabla u, \tag{38}$$

which defines the generalized quasi-Fermi potential as in

$$\mathbf{J}_n = -q\mu_n n\nabla\varphi_n, \tag{39}$$

with the quantum correction in electron concentration. Boundary conditions for this potential can be easily specified. Similar expressions also exist for hole.

For the energy fluxes, we rewrite (9) more precisely as

$$\mathbf{S}_n = \frac{5\mathbf{J}_n}{-2q}k_B T_n - \kappa_n\nabla T_n + \frac{\mathbf{J}_n}{-q}\left(\frac{1}{2}m_n^*|\mathbf{v}_n|^2\right). \tag{40}$$

Substituting (16), (29), and (39) into this equation, we have

$$\mathbf{S}_n = \frac{5\mathbf{J}_n}{-2q}k_B g_n\exp\left(\frac{5\varphi_n}{4V_T}\right) - \kappa_n\left[\exp\left(\frac{5\varphi_n}{4V_T}\right)\nabla g_n + \frac{5}{4V_T}g_n\exp\left(\frac{5\varphi_n}{4V_T}\right)\nabla\varphi_n\right] + \frac{\mathbf{J}_n}{-q}\left(\frac{1}{2}m_n^*|\mathbf{v}_n|^2\right)$$

$$= -\kappa_n\exp\left(\frac{5\varphi_n}{4V_T}\right)\nabla g_n + \frac{\mathbf{J}_n}{-q}\left(\frac{1}{2}m_n^*|\mathbf{v}_n|^2\right). \tag{41}$$

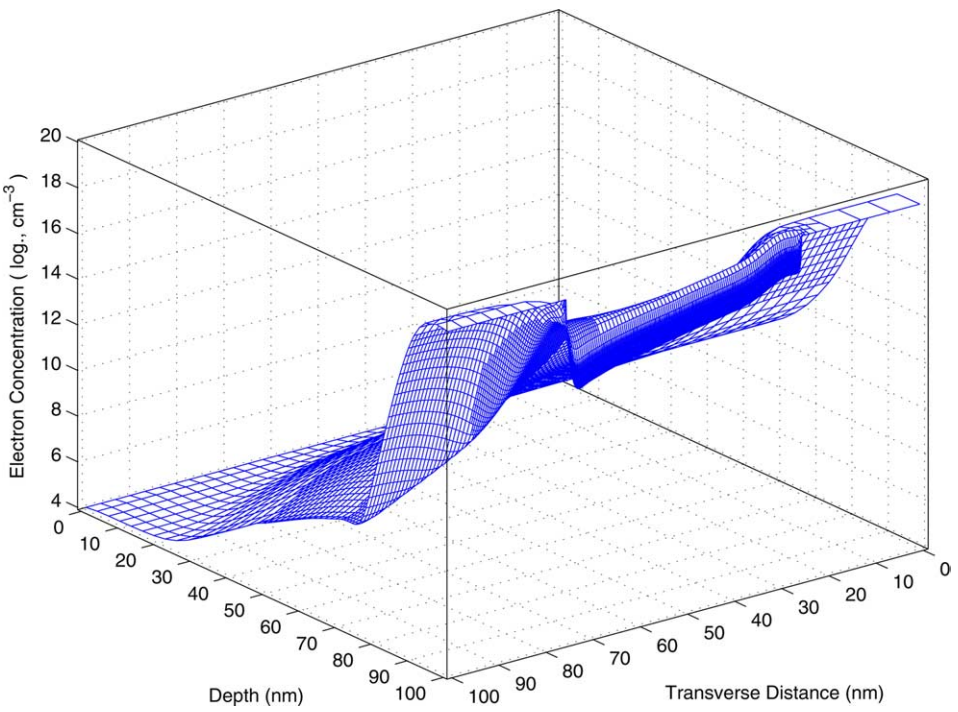

Fig. 8. Electron concentration.

Hence, we obtain the following self-adjoint form

$$\nabla \cdot \left(\kappa_{\mathrm{n}} \exp\left(\frac{5\varphi_{\mathrm{n}}}{4V_{\mathrm{T}}}\right) \nabla g_{\mathrm{n}} \right) = R_{\mathrm{n}}(g_{\mathrm{n}}), \tag{42}$$

where

$$R_{\mathrm{n}}(g_{\mathrm{n}}) = n\left(\frac{\omega_{\mathrm{n}} - \omega_0}{\tau_{\mathrm{n}\omega}}\right) - \mathbf{J}_{\mathrm{n}} \cdot \mathbf{E} - \frac{1}{q} \nabla \cdot \left(\frac{1}{2} m_{\mathrm{n}}^* \frac{|\mathbf{J}_{\mathrm{n}}|^2}{q^2 n^2} \mathbf{J}_{\mathrm{n}} \right). \tag{43}$$

We also have a similar equation for hole.

Our new model for both DG and ET equations with the seven state variables ϕ, u, v, ζ_{n}, ζ_{p}, g_{n}, and g_{p} and their associated boundary conditions (BCs) is re-organized as follows:

$$\Delta\phi = F(\phi, u, v, \zeta_{\mathrm{n}}, \zeta_{\mathrm{p}}), \tag{44}$$

$$\frac{1}{q} \nabla \cdot \mathbf{J}_{\mathrm{n}} = R(\phi, u, v, \zeta_{\mathrm{n}}, \zeta_{\mathrm{p}}), \tag{45}$$

$$\frac{1}{q} \nabla \cdot \mathbf{J}_{\mathrm{p}} = -R(\phi, u, v, \zeta_{\mathrm{n}}, \zeta_{\mathrm{p}}), \tag{46}$$

$$\Delta\zeta_{\mathrm{n}} = Z_{\mathrm{n}}(\phi, u, v, \zeta_{\mathrm{n}}, \zeta_{\mathrm{p}}), \tag{47}$$

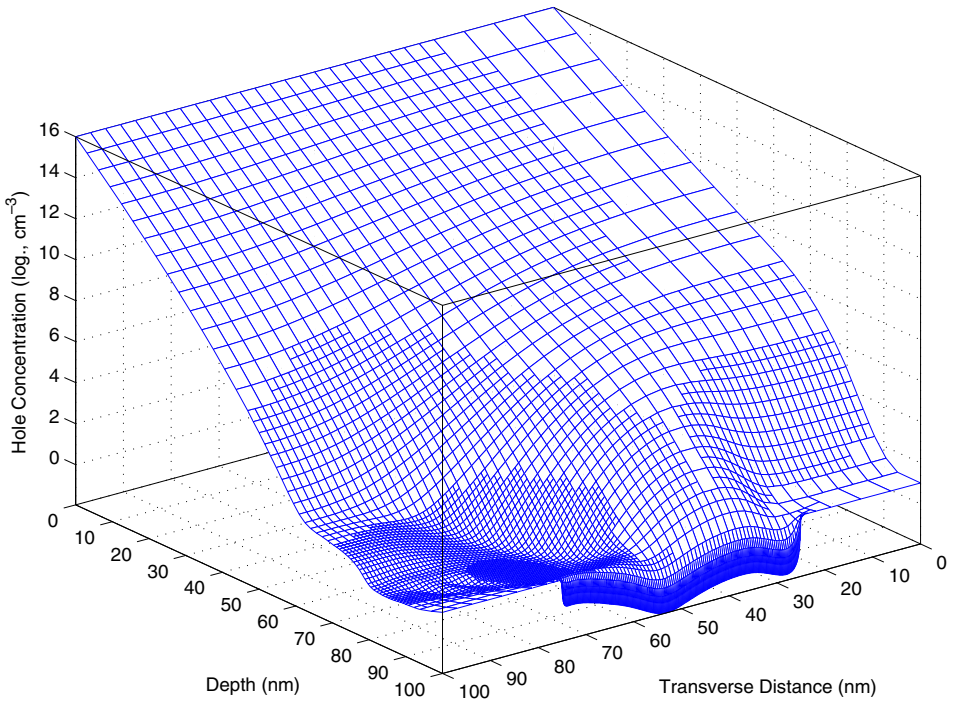

Fig. 9. Hole concentration.

$$\Delta\zeta_{\mathrm{p}} = Z_{\mathrm{p}}(\phi, u, v, \zeta_{\mathrm{n}}, \zeta_{\mathrm{p}}), \tag{48}$$

$$\nabla \cdot \mathbf{G}_{\mathrm{n}} = R_{\mathrm{n}}(g_{\mathrm{n}}), \tag{49}$$

$$\nabla \cdot \mathbf{G}_{\mathrm{p}} = R_{\mathrm{p}}(g_{\mathrm{p}}), \tag{50}$$

where

$$F(\phi, u, v, \zeta_{\mathrm{n}}, \zeta_{\mathrm{p}}) = \frac{qn_{\mathrm{i}}}{\varepsilon_{\mathrm{s}}}\left[u\exp\left(\frac{\phi+\phi_{\mathrm{qn}}}{V_{\mathrm{T}}}\right) - v\exp\left(\frac{-\phi-\phi_{\mathrm{qp}}}{V_{\mathrm{T}}}\right)\right] + \frac{q(N_{\mathrm{A}}^{-} - N_{\mathrm{D}}^{+})}{\varepsilon_{\mathrm{s}}}, \tag{51}$$

$$\mathbf{J}_{\mathrm{n}} = +qD_{\mathrm{n}}n_{\mathrm{i}}\exp\left(\frac{\phi+\phi_{\mathrm{qn}}}{V_{\mathrm{T}}}\right)\nabla u, \tag{52}$$

$$\mathbf{J}_{\mathrm{p}} = -qD_{\mathrm{p}}n_{\mathrm{i}}\exp\left(\frac{-\phi-\phi_{\mathrm{qp}}}{V_{\mathrm{T}}}\right)\nabla v, \tag{53}$$

$$R(\phi, u, v, \zeta_{\mathrm{n}}, \zeta_{\mathrm{p}}) = \frac{n_{\mathrm{i}}^2\left[uv\exp\left(\frac{\phi_{\mathrm{qn}}-\phi_{\mathrm{qp}}}{V_{\mathrm{T}}}\right) - 1\right]}{\tau_{\mathrm{n}}^0\left[n_{\mathrm{i}}v\exp\left(\frac{-\phi-\phi_{\mathrm{qp}}}{V_{\mathrm{T}}}\right) + p_{\mathrm{T}}\right] + \tau_{\mathrm{p}}^0\left[n_{\mathrm{i}}u\exp\left(\frac{\phi+\phi_{\mathrm{qn}}}{V_{\mathrm{T}}}\right) + n_{\mathrm{T}}\right]}, \tag{54}$$

$$Z_{\mathrm{n}}(\phi, u, v, \zeta_{\mathrm{n}}, \zeta_{\mathrm{p}}) = \frac{\zeta_{\mathrm{n}}}{2b_{\mathrm{n}}}\left[V_{\mathrm{T}}\ln(\zeta_{\mathrm{n}}^2) - V_{\mathrm{T}}\ln(un_{\mathrm{i}}) - \phi\right], \tag{55}$$

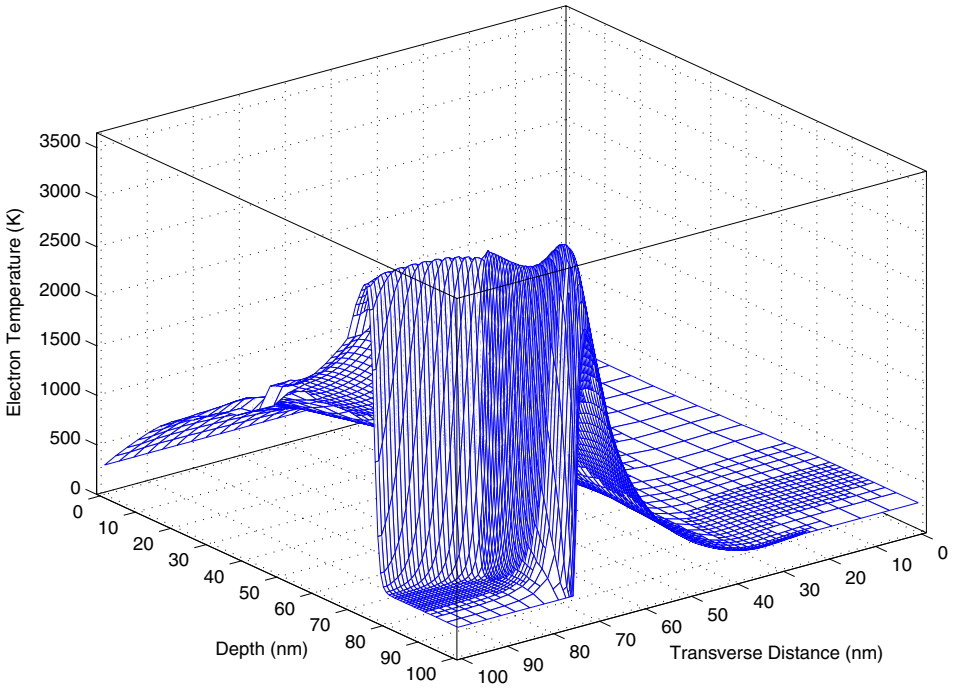

Fig. 10. Electron temperature.

$$Z_{\mathrm{p}}(\phi,u,v,\zeta_{\mathrm{n}},\zeta_{\mathrm{p}}) = -\frac{\zeta_{\mathrm{p}}}{2b_{\mathrm{p}}}\left[-V_{\mathrm{T}}\ln(\zeta_{\mathrm{p}}^2) + V_{\mathrm{T}}\ln(vn_{\mathrm{i}}) - \phi\right], \tag{56}$$

$$\phi_{\mathrm{qn}} = V_{\mathrm{T}}\ln(\zeta_{\mathrm{n}}^2) - V_{\mathrm{T}}\ln(un_{\mathrm{i}}) - \phi, \tag{57}$$

$$\phi_{\mathrm{qp}} = -V_{\mathrm{T}}\ln(\zeta_{\mathrm{p}}^2) + V_{\mathrm{T}}\ln(vn_{\mathrm{i}}) - \phi, \tag{58}$$

$$\mathbf{G}_{\mathrm{n}} = \kappa_{\mathrm{n}}\exp\left(\frac{5\varphi_{\mathrm{n}}}{4V_{\mathrm{T}}}\right)\nabla g_{\mathrm{n}}, \tag{59}$$

$$\mathbf{G}_{\mathrm{p}} = \kappa_{\mathrm{p}}\exp\left(-\frac{5\varphi_{\mathrm{p}}}{4V_{\mathrm{T}}}\right)\nabla g_{\mathrm{p}}, \tag{60}$$

$$R_{\mathrm{n}}(g_{\mathrm{n}}) = n\left(\frac{\omega_{\mathrm{n}}-\omega_0}{\tau_{\mathrm{n}\omega}}\right) - \mathbf{J}_{\mathrm{n}}\cdot\mathbf{E} - \frac{1}{q}\nabla\cdot\left(\frac{1}{2}m_{\mathrm{n}}^*\frac{|\mathbf{J}_{\mathrm{n}}|^2}{q^2n^2}\mathbf{J}_{\mathrm{n}}\right), \tag{61}$$

$$R_{\mathrm{p}}(g_{\mathrm{p}}) = p\left(\frac{\omega_{\mathrm{p}}-\omega_0}{\tau_{\mathrm{p}\omega}}\right) - \mathbf{J}_{\mathrm{p}}\cdot\mathbf{E} + \frac{1}{q}\nabla\cdot\left(\frac{1}{2}m_{\mathrm{p}}^*\frac{|\mathbf{J}_{\mathrm{p}}|^2}{q^2p^2}\mathbf{J}_{\mathrm{p}}\right). \tag{62}$$

The boundary conditions are changed accordingly to

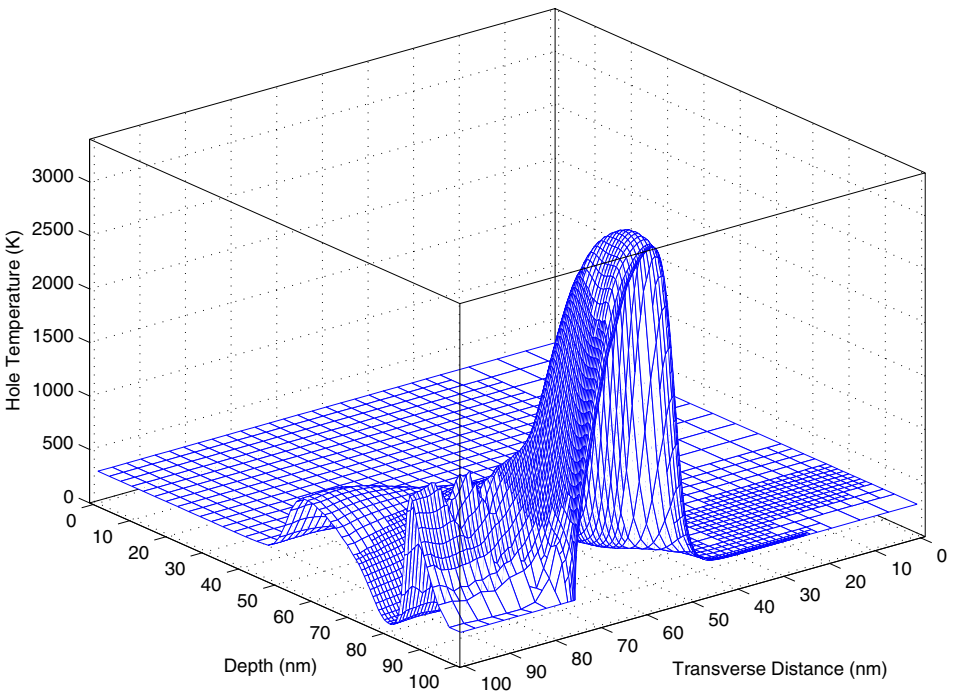

Fig. 11. Hole temperature.

$$\phi = V_O + V_b,$$
$$u = \exp\left(\frac{-V_O}{V_T}\right),$$
$$v = \exp\left(\frac{V_O}{V_T}\right),$$
$$\zeta_n^2 = \frac{1}{2}\left[(N_D^+ - N_A^-) + \sqrt{(N_D^+ - N_A^-)^2 + 4n_i^2}\right],$$
$$\zeta_p = n_i/\zeta_n,$$
$$g_n = \frac{300}{\exp\left(\frac{5V_O}{4V_T}\right)},$$
$$g_p = \frac{300}{\exp\left(-\frac{5V_O}{4V_T}\right)} \quad \text{on } \partial\Omega_D$$

and

$$\frac{\partial\phi}{\partial n} = \frac{\partial u}{\partial n} = \frac{\partial v}{\partial n} = \frac{\partial\zeta_n}{\partial n} = \frac{\partial\zeta_p}{\partial n} = \frac{\partial g_n}{\partial n} = \frac{\partial g_p}{\partial n} = 0 \quad \text{on } \partial\Omega_N,$$

where V_O denotes the applied voltage and V_b represents the built-in potential. Here, $\Omega \subset \Re^2$ denotes the bounded domain of the silicon. The boundary $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$ is piecewise smooth consisting of Dirichlet $\partial\Omega_D$ and Neumann $\partial\Omega_N$ parts. The Dirichlet part corresponds to the ohmic contacts on the device. Note

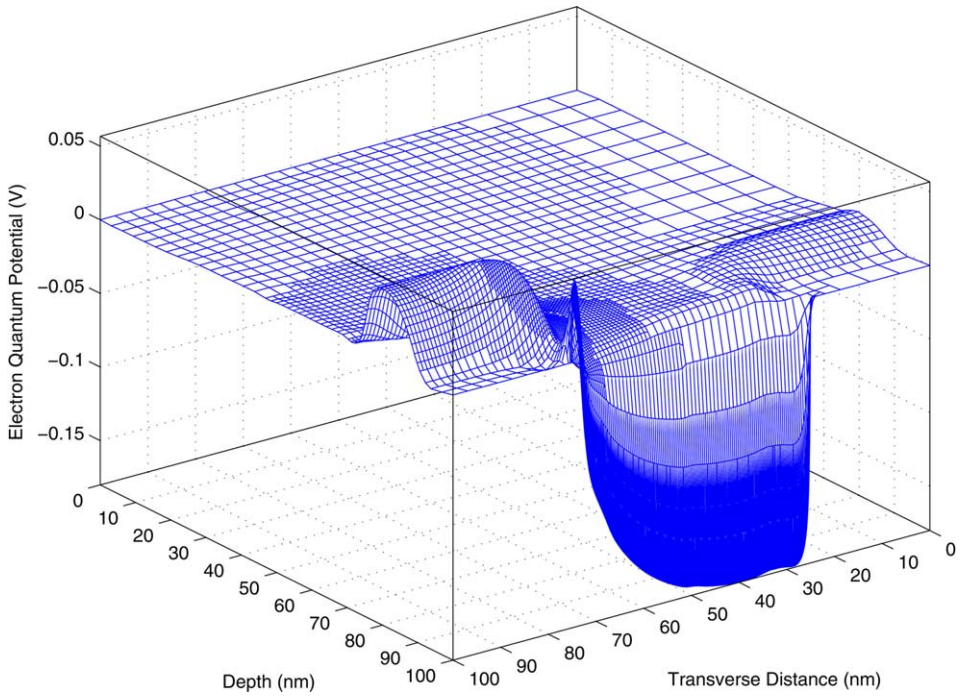

Fig. 12. Electron quantum potential.

that the above Neumann BCs for ζ_n and ζ_p do not hold on the entire $\partial\Omega_N$ excluding the oxide interface at which a zero Dirichlet BC is imposed. As mentioned in [8], the quantum potentials would have to be infinite at the interface to force the carrier densities to exactly zero there. Thus, a suitable constraint on the values of the quantum potentials at the interface is also not available. A small but non-zero value of the carrier densities is instead used in that paper. Our implementation of such non-exact zero Dirichlet BC at the interface will be specified in Section 5.

It should be noted that effective approximation of the gradient of current densities in formulas (61) and (62) is in general very difficult to acquire. Simplified models for these formulas based on physical consideration are possible. For example, by assuming that the drift energy is only a small part of the total kinetic energy [10], (61) and (62) can be reduced to

$$R_n(g_n) = n\left(\frac{\omega_n - \omega_0}{\tau_{n\omega}}\right) - \mathbf{J}_n \cdot \mathbf{E},$$
$$R_p(g_p) = p\left(\frac{\omega_p - \omega_0}{\tau_{p\omega}}\right) - \mathbf{J}_p \cdot \mathbf{E},$$

which will be used in our numerical simulations.

Remark 3.1. As observed in [5], the SRH generation-recombination model (18) should be modified for the DG model since this standard expression will produce spurious generation and recombination near the oxide barrier. We thus consider here the modified SRH (MSRH) proposed in [5] and extend it into the self-adjoint context as follows:

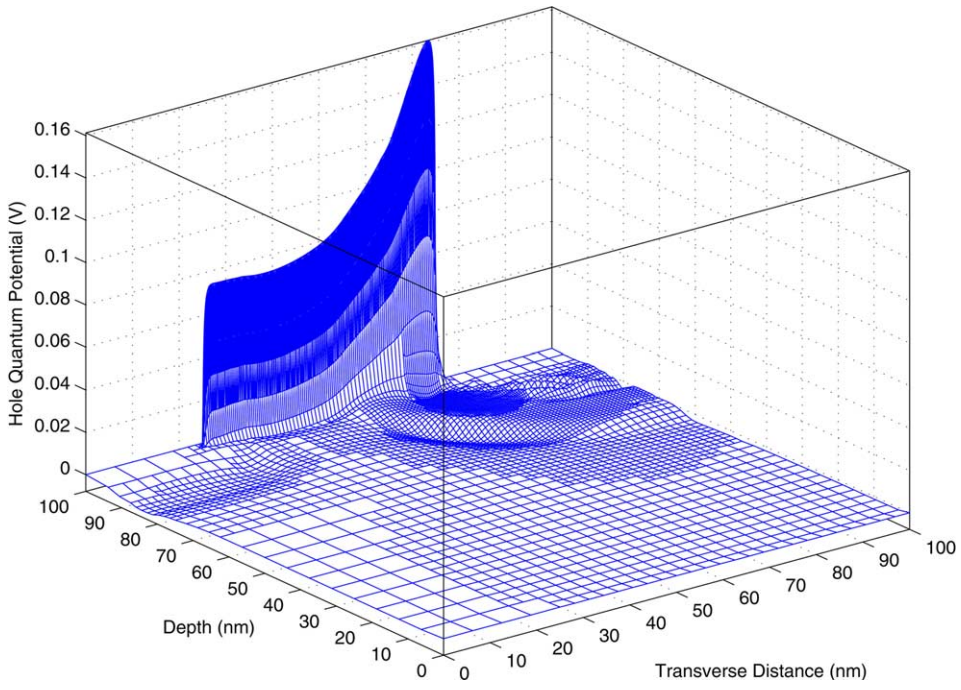

Fig. 13. Hole quantum potential.

$$\frac{np - n_{\mathrm{eq}}p_{\mathrm{eq}}}{\tau_{\mathrm{n}}^0\left(p + \sqrt{n_{\mathrm{eq}}p_{\mathrm{eq}}}\exp\left(\frac{\varepsilon_{\mathrm{t}}-\varepsilon_{\mathrm{i}}}{k_{\mathrm{B}}T}\right)\right) + \tau_{\mathrm{p}}^0\left(n + \sqrt{n_{\mathrm{eq}}p_{\mathrm{eq}}}\exp\left(\frac{\varepsilon_{\mathrm{t}}-\varepsilon_{\mathrm{i}}}{k_{\mathrm{B}}T}\right)\right)} = \frac{n_{\mathrm{i}}^2 uv\exp\left(\frac{\phi_{\mathrm{qn}}-\phi_{\mathrm{qp}}}{V_{\mathrm{T}}}\right) - n_{\mathrm{eq}}p_{\mathrm{eq}}}{\tau_{\mathrm{n}}^0\left[n_{\mathrm{i}}v\exp\left(\frac{-\phi-\phi_{\mathrm{qp}}}{V_{\mathrm{T}}}\right) + \sqrt{n_{\mathrm{eq}}p_{\mathrm{eq}}}\exp\left(\frac{\varepsilon_{\mathrm{t}}-\varepsilon_{\mathrm{i}}}{k_{\mathrm{B}}T_{\mathrm{L}}}\right)\right] + \tau_{\mathrm{p}}^0\left[n_{\mathrm{i}}u\exp\left(\frac{\phi+\phi_{\mathrm{qn}}}{V_{\mathrm{T}}}\right) + \sqrt{n_{\mathrm{eq}}p_{\mathrm{eq}}}\exp\left(\frac{\varepsilon_{\mathrm{t}}-\varepsilon_{\mathrm{i}}}{k_{\mathrm{B}}T_{\mathrm{L}}}\right)\right]}.$$

Note that the term n_{i}^2 in (18) is replaced by $n_{\mathrm{eq}}p_{\mathrm{eq}}$ in this MSRH model where ε_{t} and ε_{i} are the trapped and intrinsic energies. The quantities n_{eq} and p_{eq} are the spatially dependent equilibrium densities obtained from a separate numerical solution of the same DG problem, but with all voltages and R set to zero. Following that paper, we choose $\tau_{\mathrm{n}}^0 = \tau_{\mathrm{p}}^0 = 10^{-8}$ s with $\varepsilon_{\mathrm{t}} = \varepsilon_{\mathrm{i}}$ in our simulation. A comparison of numerical results based on both SRH and MSRH models will be given in Section 5.

Remark 3.2. For simplicity, we use fixed mobilities of $\mu_{\mathrm{n}} = 1500$ cm$_2$/V s and $\mu_{\mathrm{p}} = 500$ cm$_2$/V s which are roughly equal to the intrinsic values at room temperature for silicon as considered in [8]. In our numerical experiences in [11,12], the field-dependent mobility model of the Caughey–Thomas expression still can be used in the DGET simulation.

Remark 3.3. The above self-adjoint formulation is based on Maxwell–Boltzmann statistics. However, it is unclear to us whether the self-adjointness can also be derived for the case of Fermi–Dirac statistics which is more exact but more complicated to implement. Evidently, this issue deserves further investigation in the future.

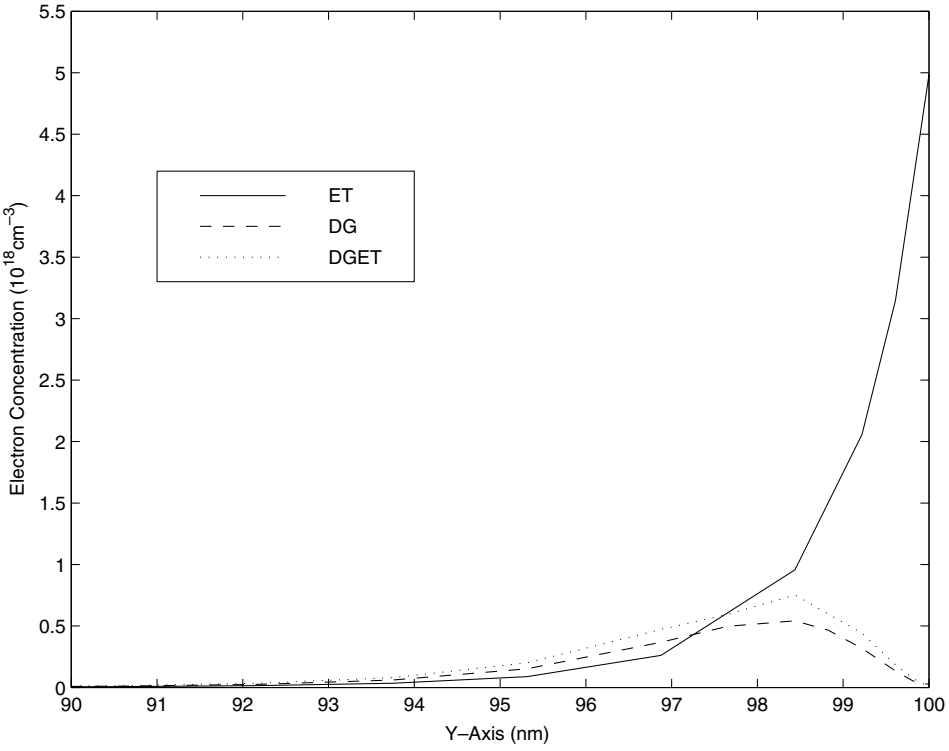

Fig. 14. Electron density profile perpendicular to the inverted channel.

4. An adaptive finite element algorithm for the DGET model

The main ingredients of the algorithm solving the DGET model are adaptive finite element approximation of the model, node-by-node and monotone iterative solution of the resulting nonlinear algebraic systems, and Gummel's iteration consecutively on the PDEs as described in [11] for the ET model. For the sake of clearness, we briefly illustrate the algorithm and refer to [11,12] for more details on the adaptive finite element formulation, monotone convergence analysis, and practical implementation issues.

Here, we use the notation l as Gummel's (outer) iteration index and m as the monotone (inner) iteration index.

Step 1. Initial mesh: create a coarse and structured mesh for which the number of nodes can be chosen as small as possible.
Step 2. Preprocessing: see [11].
Step 3. Gummel and Monotone iterations on (44)–(48).
Step 3.0. Set $l := 0$
Step 3.1. Solve the potential equation in (44).
Step 3.1.1. Set $m := 0$ and set the initial guess

$$\phi_j^{(m)} = \begin{cases} \widetilde{\phi_j} \text{ or } \widehat{\phi_j} & \text{if } l = 0, \\ \phi_j^{(l)} & \text{otherwise,} \end{cases} \quad \text{for all } (x_j, y_j) \in \overline{\Omega}^h,$$

where $\widetilde{\phi_j}$ and $\widehat{\phi_j}$ are constant values that can be easily verified to be an upper and lower solution of ϕ, respectively, and $\overline{\Omega}^h$ denotes the set of mesh points on the closure of the domain.

Step 3.1.2 If $l = 0$, set $u^{(l)}$ and $v^{(l)}$ by the charge neutrality condition.

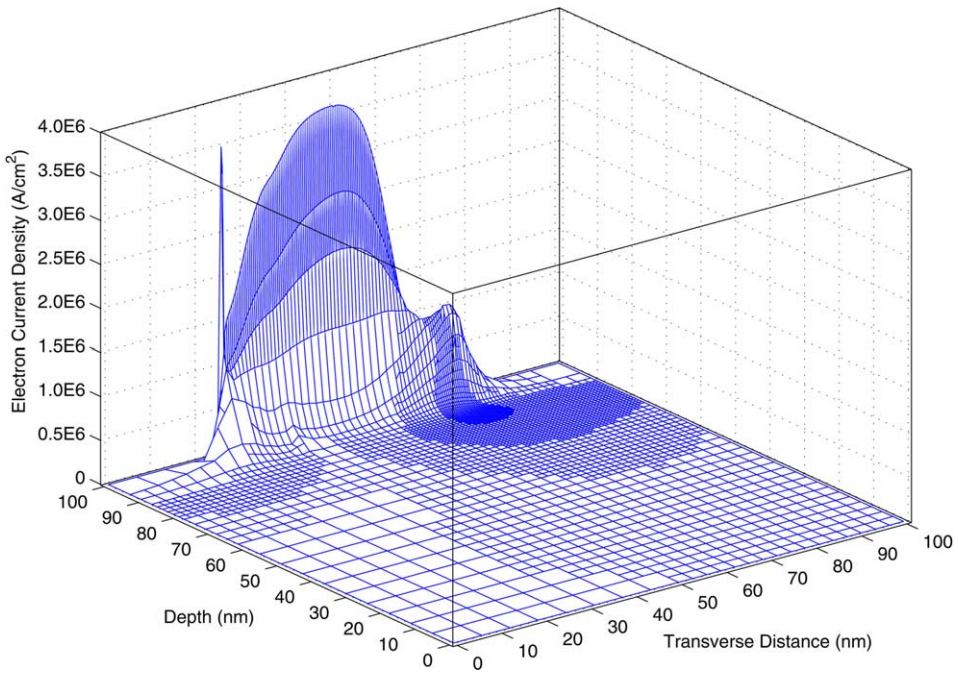

Fig. 15. Electron current density (ET).

Step 3.1.3. Compute $\phi_j^{(m+1)}$ by solving the discrete potential system of (44)

$$\begin{cases} \xi_j \phi_j^{(m+1)} + \gamma_j(\phi)\phi_j^{(m+1)} = \sum_{k\in V(j)} \xi_k \phi_k^{(m)} - F(\phi_j^{(m)}, u_j^{(l)}, v_j^{(l)}, \zeta_{\mathrm{n}}^{(l)}, \zeta_{\mathrm{p}}^{(l)}) + \gamma_j(\phi)\phi_j^{(m)}, & \forall (x_j, y_j) \in \Omega^h, \\ \phi_j^{(m+1)} = V_{\mathrm{O}} + V_{\mathrm{b}}, & \forall (x_j, y_j) \in \partial\Omega_{\mathrm{D}}^h, \\ \frac{\partial \phi_j^{(m+1)}}{\partial n} = 0, & \forall (x_j, y_j) \in \partial\Omega_{\mathrm{N}}^h, \end{cases} \tag{63}$$

where

$$\gamma_j(\phi) = \max\left\{ \frac{\partial F(\phi_j)}{\partial \phi}; \hat{\phi}_j \leqslant \phi_j \leqslant \tilde{\phi}_j \right\}, \tag{64}$$

ξ_k are the matrix elements of the discretization, and Ω^h, $\partial\Omega_{\mathrm{D}}^h$, and $\partial\Omega_{\mathrm{N}}^h$ represent the sets of mesh points in the interior, Dirichlet part, and Neumann part of the domain, respectively.

Step 3.1.4. Set $\phi_j^{(m)} := \phi_j^{(m+1)}$ $\forall j$ and $m := m + 1$. Go to Step 3.1.3 until the stopping criteria of the inner iteration are satisfied.

Step 3.1.5. Set $\phi_j^{(l+1)} := \phi_j^{(m+1)}$ $\forall j$.

Step 3.2. Solve the electron continuity equation (45).

Step 3.2.1. Set $m := 0$ and set the initial guess

$$u_j^{(m)} = \begin{cases} \widetilde{u}_j \text{ or } \widehat{u}_j & \text{if } l = 0, \\ u_j^{(l)} & \text{otherwise,} \end{cases} \quad \text{for all } (x_j, y_j) \in \overline{\Omega}^h,$$

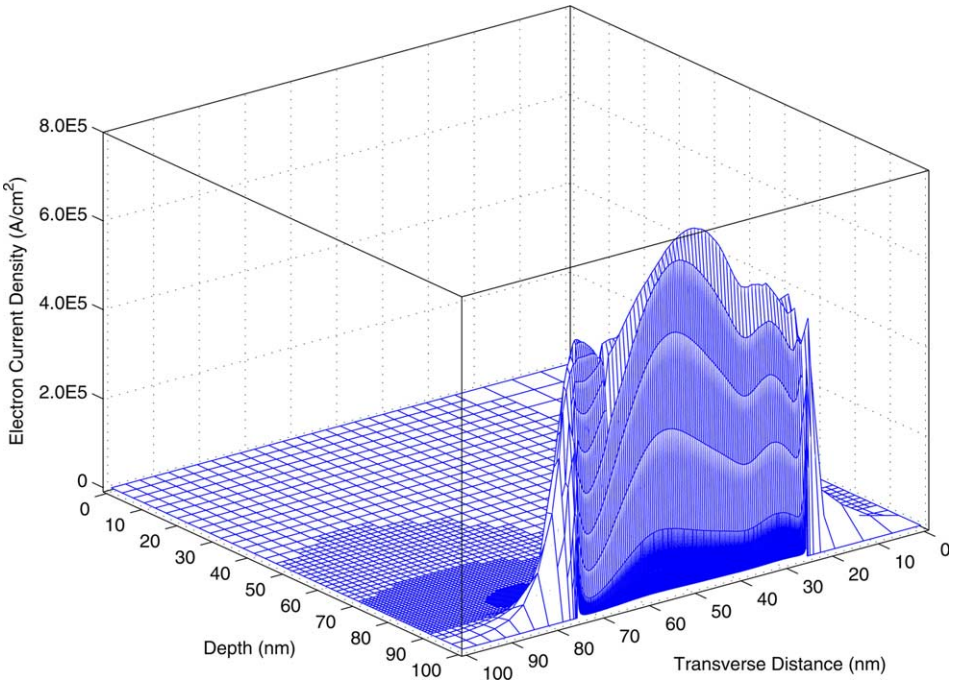

Fig. 16. Electron current density (DG).

where $\widetilde{u}_j$ and $\widehat{u}_j$ are constant values for all $(x_j, y_j) \in \overline{\Omega}^h$ that can be easily verified to be an upper and lower solution of u, respectively.

Step 3.2.2. Compute $u_j^{(m+1)}$ by solving the discrete electron system of (45).

Step 3.2.3. Set $u_j^{(m)} := u_j^{(m+1)}$ $\forall j$ and $m := m + 1$. Go to Step 3.2.2 until the stopping criteria of the inner iteration are satisfied.

Step 3.2.4. Set $u_j^{(l+1)} := u_j^{(m+1)}$ $\forall j$.

Step 3.3. Solve the hole continuity equation (46) similarly to that in Step 3.2.

Step 3.4. Solve the DG equation (47).

Step 3.4.1. Set $m := 0$ and set the initial guess

$$[\zeta_n]_j^{(m)} = \begin{cases} [\widetilde{\zeta_n}]_j \text{ or } [\widehat{\zeta_n}]_j & \text{if } l = 0, \\ [\zeta_n]_j^{(l)} & \text{otherwise,} \end{cases} \quad \text{for all } (x_j, y_j) \in \overline{\Omega}^h,$$

where $[\zeta_n]_j \approx \zeta_n(x_j, y_j)$ and $[\widetilde{\zeta_n}]_j$ and $[\widehat{\zeta_n}]_j$ are constant values for all $(x_j, y_j) \in \overline{\Omega}^h$ that can be easily verified to be an upper and lower solution of ζ_n, respectively.

Step 3.4.2. Compute $[\zeta_n]_j^{(m+1)}$ by solving the discrete system of (47).

Step 3.4.3. Set $[\zeta_n]_j^{(m)} := [\zeta_n]_j^{(m+1)}$ $\forall j$ and $m := m + 1$. Go to Step 3.4.2 until the stopping criteria of the inner iteration are satisfied.

Step 3.4.4. Set $[\zeta_n]_j^{(l+1)} := [\zeta_n]_j^{(m+1)}$ $\forall j$.

Step 3.5. Solve the DG (48) similarly to that in Step 3.4.

Step 3.6. Update $[\phi_{qn}]_j^{(l+1)}$ and $[\phi_{qp}]_j^{(l+1)}$ by the Eqs. (57) and (58).

Step 3.7. Set $l := l + 1$ and go to Step 3.1 until the stopping criteria of the outer iteration are satisfied.

Step 4. Monotone Iteration on (49) and (50).

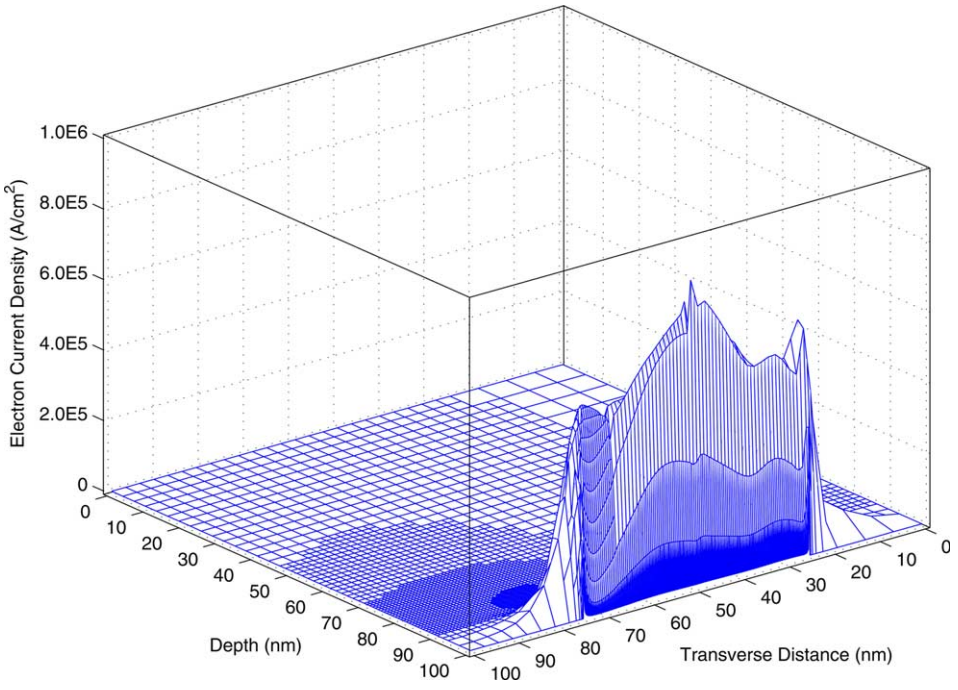

Fig. 17. Electron current density (DGET).

Step 4.1. Solve the energy equation (49) for g_n similarly to that in Step 3.2.
Step 4.2. Solve the energy equation (50) for g_p similarly to that in Step 3.2.
Step 5. Error estimation: See [11].
Step 6. Refinement: See [11]
Step 7. Postprocessing: All computed solutions are then postprocessed for further analysis of physical phenomena.

Note that, in each one of Steps 3.1–3.5, 4.1, and 4.2, a Jacobi (node-by-node) type of solution is performed for the corresponding discrete system (63), for example, in which the monotone parameters (64) can be easily obtained by means of lower and upper solutions. Two important factors that guarantee a global convergence with this kind of simple solutions as initial guesses are the diagonally dominant property of the matrices due to the self-adjoint formulation and the monotonicity of the parameters by the special nonlinearity of the formulation. The diagonally dominant property for (44)–(50) can be proved in exactly the same manner as that given in [11,12]. It can also be easily shown that each one of the nonlinear functionals in (44)–(50) is monotone in its respective state variable. It is thus a straightforward generalization from our previous theoretical analysis that all the nonlinear algebraic systems generated by this algorithm preserve these two factors. We thus summarize these important results in the following theorem.

Theorem. *For each one of the PDEs* (44)–(50) *with associated boundary conditions, the matrices resulting to the adaptive finite element approximation are diagonally dominant. Moreover, starting with suitable lower and upper solutions of the corresponding PDE, the Jacobi iteration in each of Steps* 3.1–3.5, 4.1, *and* 4.2 *generates a pair of lower and upper sequences which converge monotonically to the exact solution of the nonlinear algebraic system of equations of that PDE.*

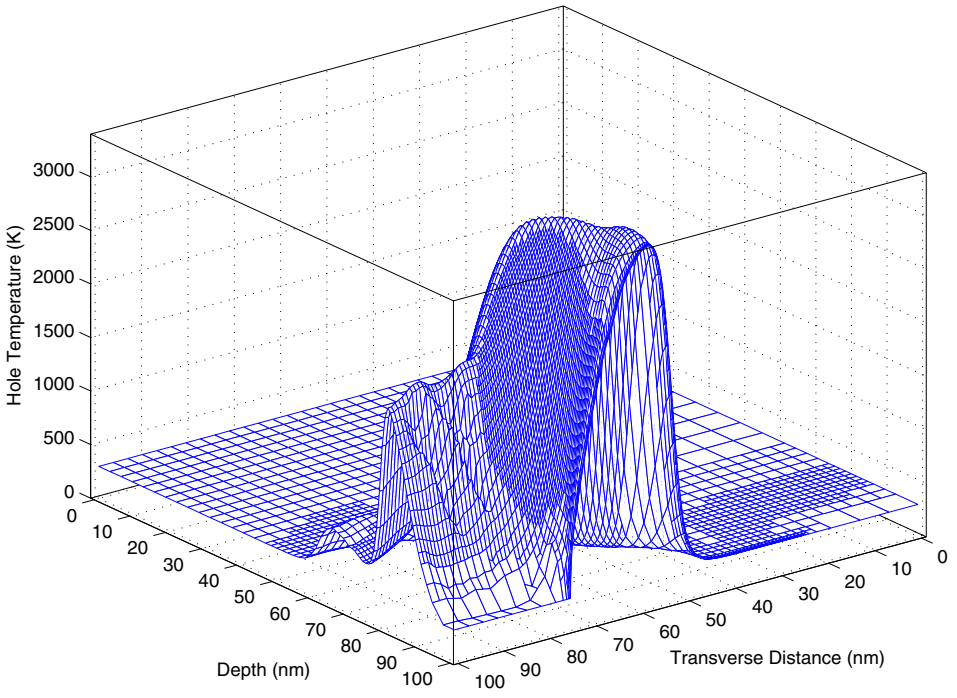

Fig. 18. Hole temperature (with the standard SRH model (18)).

5. Numerical examples

To demonstrate the effectiveness and accuracy of the DGET model, several numerical studies have been made for sample diode and MOSFET device structures. A benchmark device, namely, an abrupt n^+–n–n^+ silicon diode is first used to verify our methods and formulation with the results reported in literature. Numerical experiments are performed first on a 600 nm silicon diode at 300 K with $n^+ = 5.0 \times 10^{17}$ cm^{-3} and $n = 2.0 \times 10^{15}$ cm^{-3}. The length of the n-region is approximately 400 nm. The steady state results for this problem are illustrated by the dotted and solid curves with respective to the DGET and ET models in Figs. 1(a)–(d) where the applied voltage V_O is taken as 2.0 V. The dotted curve coincides with the solid curve. This represents that the new model can be applied to devices with larger size, i.e., where the QM effects are negligible. These results agree also very well with that previously reported in the literature [2,14,17,31].

To verify QM effects with our model, we then reduce the scale down to two cases. Case (1) is a 120 nm silicon diode with $n^+ = 5.0 \times 10^{18}$ cm^{-3} and $n = 2.0 \times 10^{15}$ cm^{-3}. The length of the n-region is approximately 80 nm. The applied voltage V_O is taken as 1.2 V. Case (2) is a 30 nm silicon diode with $n^+ = 5.0 \times 10^{19}$ cm^{-3} and $n = 2.0 \times 10^{15}$ cm^{-3}. The length of the n-region is approximately 20 nm. The applied voltage V_O is taken as 0.8 V. Figs. 2 and 3 show the significant change of the electron density predicted by the new model but for the electron temperature the change is not very significant. The maximal temperatures of ET and DGET models are $T = 3423$ K and $T = 3309$ K, respectively. The corresponding thermal energies are $E_{th} = \frac{3}{2} k_B T = 0.442$ eV and $E_{th} = 0.428$ eV. Therefore, the temperature reduced by the QM corrections of the DGET model is very similar to that by the nonparabolicity effects presented in [14]. Fig. 4 shows a visible tendency of the quantum potential ϕ_{qn} toward a large variation when the channel length is decreased. Here, we scale the figures into the same size for comparison.

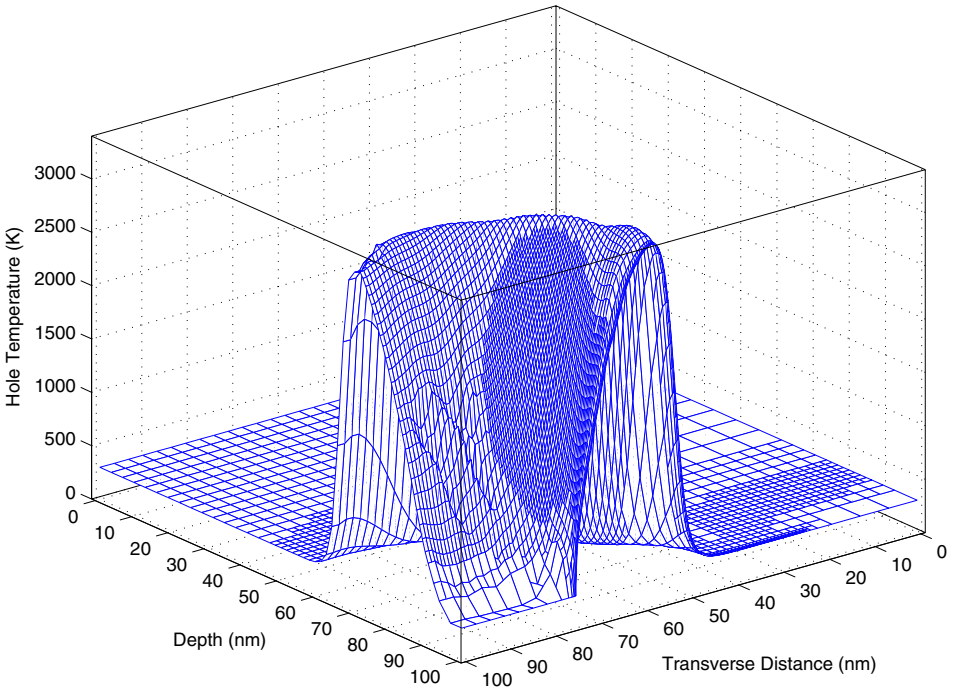

Fig. 19. Hole temperature ($R = 0$).

The second example of our simulation test on the model is a MOSFET device structure which 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 as shown in Fig. 5. 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. With $V_{BS} = 0$ V, $V_{DS} = 1.0$ V and $V_{GS} = 0.8$ V, Figs. 6–13 present the final adaptive mesh, electrostatic potential, electron concentration, hole concentration, electron temperature, hole temperature distribution, electron quantum potential, and hole quantum potential, respectively. Across the junction, Figs. 12 and 13 clearly show similar quantum potential profiles as that in Fig. 4 for the 1D diode device. Furthermore, in the direction perpendicular to the interface, a very thin boundary layer of about 6 nm appear in the inversion layer as shown in Figs. 12–17. The boundary layer as shown in Fig. 6 is effectively captured by the a posteriori error estimation with 1-irregular refinement strategy developed in [11,12,25,26].

As mentioned earlier, a suitable constraint on the values of the quantum potentials at the oxide interface is not available. One solution to this lack of quantum potential BCs is to solve the DGET model in the oxide as well as in the adjoining silicon and poly gate. This will allow us to simulate the tunneling effects across the oxide [13]. This issue is not addressed here and will be reported elsewhere in our future works. We do not impose zero Dirichlet BCs for the variables ζ_n and ζ_p exactly at the interface but instead at the grid points in silicon that are very close (about 0.13 nm) to the interface. In effect, these BCs are very similar to that in [8] where a small but non-zero value of the carrier densities is set at the interface. We found that if the BCs are prescribed exactly at the interface, the result of temperature will be very poor although the algorithm is still convergent.

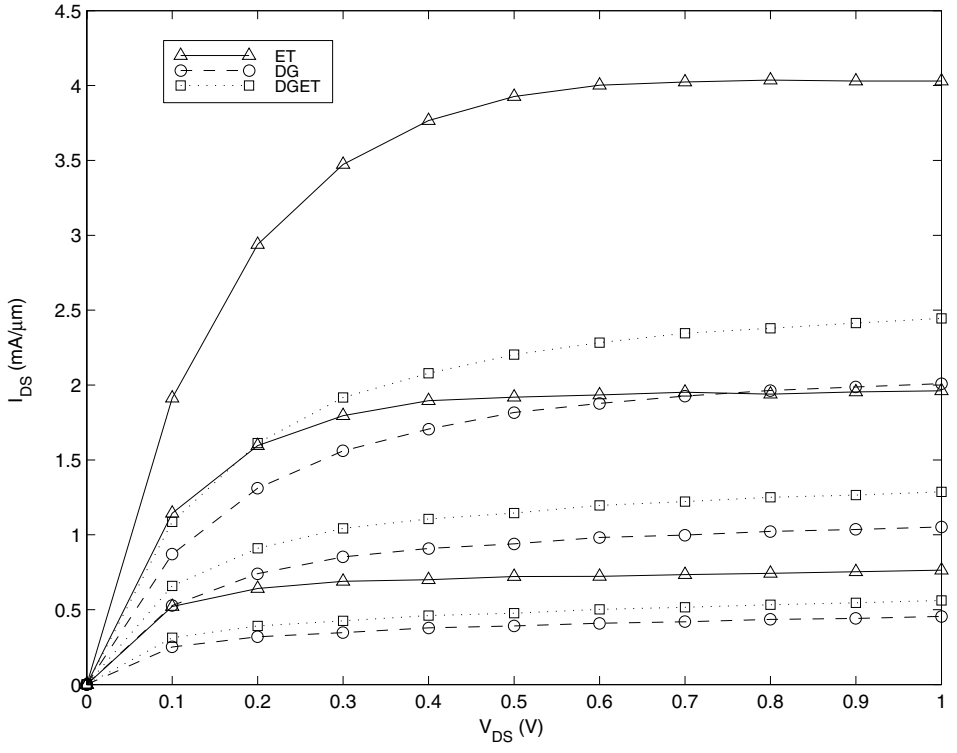

Fig. 20. Channel current for MOSFET, gate voltage $V_{GS} = 0.7$, 0.8, 0.9 V.

As noted in Remark 2.2, the quantum corrections are implicitly defined in the energy fluxes (9) and (10) via the carrier densities (19) and (20), the temperature distribution of carriers will not be very accurate as shown in Figs. 10 and 11. More specifically, the temperature peak for electron appears to be near the drain but the peak for hole is in the middle of the channel. We found that the generation-recombination model will influence the hole temperature distribution significantly. If the standard SHR model (54) is used instead of the MSHR model of Remark 3.1, the hole temperature is even much worse as shown in Figs. 18 and 19. To our knowledge, there are no numerical results of quantum corrected carrier temperatures available in the literature to be compared with our results. Evidently, efficient and effective numerical methods for handling energy fluxes with explicit quantum corrections are needed for future investigations.

The electron density profile shown in Fig. 14 is a cross section of the 2D profile at the middle point of the interface. The peak of the density is about 1.5 nm away from the interface, which agrees well with that in [7], see also [1]. Fig. 15 is the electron current density computed by the ET model, which clearly shows that the classical density is sharply peaked comparing with the smoothly peaked in Figs. 16 and 17 obtained by the DG and DGET models. The substantial QM effect of transconductance degradation is also evidently displayed in these figures. From these figures, we observe that the carrier temperature provides a mechanism to increase carrier diffusion as noted in Remark 2.1. This is the main justification to consider DGET instead of DG along.

Finally, Fig. 20 shows the simulated I–V curves in which the drain current obtained by the DGET model is about 20–60% below that predicted by the ET model for the gate biases of 0.7, 0.8, and 0.9 V. This result is also in good agreement with that of [8] where a MOSFET with 30 nm gate length and 2 nm gate oxide thickness is considered. Admittedly, this represents a serious decrease in the current drive capability of the device. Note also that the difference of the maximal temperatures between the ET ($T = 3677$ K) and DGET ($T = 3649$ K) models is not very significant. The corresponding thermal energies are 0.475 and 0.471 eV. The figure also shows that ET over estimates the current whereas DG under estimates.

6. Conclusion

A self-adjoint model combining both ET and DG models is proposed here for nanoscale semiconductor devices. This model is capable of describing hot carrier and quantum correction effects.

Moreover, due to the self-adjointness and monotonic nonlinearity, the present model enjoys many favorable mathematical properties such as global convergence with simple initial guesses, highly parallelizable, and fast iterative solution. Numerical convergence is a fundamental issue constantly faced by the practitioner in device and circuit simulation. This model and monotone iterative methods may offer an alternative in handling the convergence difficulties frequently associated with Newton's methods.

Our numerical simulations on diode and MOSFET with the gate length down to 34 nm using the DGET model have been performed and compared with that using the ET model. And the results are shown to be in good agreement with those reported in the literature. It is shown that the I–V characteristics of this short-channel device is significantly corrected by the density-gradient equations with current drive reduced by up to 60% comparing with that of the classical model along. Furthermore, a 2D quantum layer, which is only a fraction of the length scale of inversion layer, is also effectively captured by this model with very fine mesh near the interface generated by an adaptive finite element method.

Nevertheless, many improvements on our preliminary model can be further studied in future works. For example, the self-adjoint formulation of the present paper is based on Maxwell–Boltzmann statistics. It is however unclear to us whether the self-adjointness can be also derived for the case of Fermi–Dirac statistics which is more exact but more complicated to implement. Moreover, efficient and effective numerical methods for handling energy fluxes with explicit quantum corrections are also deserved for future investigations.

References

[1] H. Abebe, E. Cumberbatch, Quantum mechanical effects correction models for inversion charge and I–V characteristics of the MOSFET device, in: Proceedings of the 2003 Nanotechnology Conference, vol. 2, 2003, pp. 218–221.
[2] N.R. Aluru, A. Raefsky, P.M. Pinsky, K.H. Law, R.J.G. Goossens, R.W. Dutton, A finite element formulation for the hydrodynamic semiconductor device equations, Comput. Methods Appl. Mech. Engrg. 107 (1993) 269.
[3] M.G. Ancona, G.J. Iafrate, Quantum correction to the equation of state of an electron gas in a semiconductor, Phys. Rev. B 39 (1989) 9536.
[4] M.G. Ancona, H.F. Tiersten, Macroscopic physics of the silicon inversion layer, Phys. Rev. B 35 (1987) 7959.
[5] M.G. Ancona, Z. Yu, R.W. Dutton, P.J.V. Voorde, M. Cao, D. Vook, Density-gradient analysis of MOS tunneling, IEEE Trans. Electron. Dev. 47 (2000) 2310.
[6] Y. Apanovich, E. Lyumkis, B. Polsky, A. Shur, P. Blakey, Steady-state and transient analysis of submicron devices using energy balance and simplified hydrodynamic models, IEEE Trans. CAD 13 (1994) 702.
[7] A. Asenov, A.R. Brown, J.R. Watling, The use of quantum potentials for confinement in semiconductor devices, Model. Simul. Microstruct. (MSM'2002) (2002) 490.
[8] B.A. Biegel, M.G. Ancona, C.S. Rafferty, Z. Yu, Efficient multi-dimensional simulation of quantum confinement effects in advanced MOS devices, NAS Tech. Report NAS-04-008, 2004.
[9] B.A. Biegel, J.D. Plummer, Comparison of self-consistency iteration options for the Wigner function method of quantum device simulation, Phys. Rev. B 54 (1996) 8070.
[10] D. Chang, J.G. Fossum, Simplified energy-balance model for pragmatic multi-dimensional device simulation, Solid-State Electron. 41 (1997) 1795.
[11] 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.
[12] 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.
[13] C.-H. Choi, Z. Yu, R.W. Dutton, Modelling of MOS scaling with emphasis on gate tunnelling and source/drain resistance, Superlattices Microstruct. 27 (2000) 191.
[14] P. Degond, A. Jüngel, P. Pietra, Numerical discretization of energy-transport models for semiconductors with non-parabolic band structure, SIAM Sci. Comput. 22 (2000) 986.
[15] D.K. Ferry, J.-R. Zhou, Form of the quantum potential for use in hydrodynamic equations for semiconductor device modeling, Phys. Rev. B 48 (1993) 7944.
[16] C.L. Gardner, The quantum hydrodynamic model for semiconductor devices, SIAM J. Appl. Math. 54 (1994) 409.
[17] C.L. Gardner, J.W. Jerome, D.J. Rose, Numerical methods for the hydrodynamic device model: subsonic flow, IEEE Trans. CAD 8 (1989) 501.
[18] H.L. Grubin, J.P. Kreskovsky, Quantum moment balance equations and resonant tunnelling structures, Solid-State Electron. 32 (1989) 1701.
[19] W. Hänsch, T. Vogelsang, R. Kircher, M. Orlowski, Carrier transport near the Si/SiO_2 interface of a MOSFET, Solid-State Electron. 32 (1989) 839.
[20] T. Höhr, A. Schenk, A. Wettstein, W. Fichtner, On density-gradient modeling of tunneling through insulators, IEICE Trans. Electron. E86-C (2003) 379.
[21] S. Jallepalli, J. Bude, W.K. Shih, M.R. Pinto, C.M. Maziar, A.F. Tasch Jr., Electron and hole quantization and their impact on deep submicron p- and n-MOSFET characteristics, IEEE Trans. Electron Devices 44 (1997) 297.
[22] J.P. Kreskovsky, H.L. Grubin, VLSI Design 3 (2) (1995) 179.
[23] R. Lake, G. Klimeck, R.C. Bowen, D. Jovanovic, J. Appl. Phys. 81 (1997) 7845.
[24] Y. Li, J.-L. Liu, S.M. Sze, T.-S. Chao, A new parallel adaptive finite volume method for the numerical simulation of semiconductor devices, Comput. Phys. Commun. 142 (2001) 285.
[25] J.-L. Liu, On weak residual error estimation, SIAM J. Sci. Comput. 17 (1996) 1249.
[26] J.-L. Liu, I.-J. Lin, M.-Z. Shih, R.-C. Chen, M.-C. Hsieh, Object oriented programming of adaptive finite element and finite volume methods, Appl. Numer. Math. 21 (1996) 439.
[27] C. Philippidis, D. Bohm, R.D. Kaye, The Aharonov–Bohm effect and the quantum potential, Il Nuovo Cimento 71B (1982) 75.
[28] R. Pinnau, A. Unterreiter, The stationary current–voltage characteristics of the quantum drift-diffusion model, SIAM J. Numer. Anal. 37 (1999) 211.
[29] 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, Proc. SISPAD (1998) 137–140.
[30] R. Rios, N.D. Arora, C.-L. Huang, N. Khalil, J. Faricelli, L. Gruber, IEDM Technical Digest (1995) 937.
[31] M. Rudan, F. Odeh, J. White, Numerical solution of the hydrodynamic model for a one-dimensional device, COMPEL 6 (1987) 151.

[32] L. Shifren, R. Akis, D.K. Ferry, Correspondence between quantum and classical motion: comparing Bohmian mechanics with a smoothed effective potential approach, Phys. Lett. A 274 (2000) 75.
[33] M.J. van Dort, P.H. Woerlee, A.J. Walker, A.H. Juffermans, H. Lifka, A simple model for quantization effects in heavily-doped silicon MOSFET's at inversion conditions, IEEE Trans. Electron. Dev. 39 (1992) 932.
[34] Z. Yu, R.W. Dutton, R.A. Kiehl, Circuit/device modeling at the quantum level, IEEE Trans. Electron. Dev. 47 (2000) 1819.