Quantum Device-Simulation with the Density-Gradient Model on Unstructured Grids
Andreas Wettstein, Andreas Schenk, and Wolfgang Fichtner

Abstract—We describe an implementation of the density-gradient device equations which is simple and works in any dimension without imposing additional requirements on the mesh compared to classical simulations. It is therefore applicable to real world device simulation with complex geometries. We use our implementation to determine the quantum mechanical effects for a MOS-diode, a MOSFET and a double-gated SOI MOSFET. The results are compared to those obtained by a 1D-Schrödinger-Poisson solver. We also investigate a simplified variant of the density-gradient term and show that, while it can reproduce terminal characteristics, it does not give the correct density distribution inside the device.

Keywords—Box Method, Density Gradient Method, multidimensional device simulation, quantum corrections, semiconductor-device modeling

I. INTRODUCTION

The length scales in today’s semiconductor devices are so small that quantum effects become important. Apart from tunneling through the gate insulator, the most important effect is the modification of the gate capacity as a function of the gate voltage (CV-characteristics), where confinement leads to a threshold voltage shift and an apparent increase of the oxide thickness.

Various methods have been suggested to model these quantum effects. Among the approaches that are compatible with classical device simulators based on the drift-diffusion (or hydrodynamic) approach, the physically most accurate method is to include the Schrödinger equation into the self-consistent computation of the device characteristics [1]. As for current devices the quantization is only relevant for the direction perpendicular to the oxide-silicon interface, one restricts to a 1D treatment of the Schrödinger equation. 2D devices are treated by making a couple of 1D slices along the channel. Finite temperature is taken into account by computing multiple subbands and populating them according to their energy assuming Maxwell-Boltzmann or Fermi distribution functions. Solving the Schrödinger equation is not only time-consuming in itself, it also leads to direct coupling of all grid points within a slice. This destroys the sparsity structure in the Jacobian needed for the Newton-iteration used to solve the semiconductor equations. Therefore, these nonlocal couplings are typically neglected. Another serious drawback of this method is that it is one-dimensional in nature and thus realistic devices with nonplanar oxide-silicon interfaces can not be handled satisfactory. Using a 2D Schrödinger equation could help here, but at the price of much increased numerical cost to determine the eigensolutions and of yet higher nonlocality. An additional physical problem is the choice of appropriate boundary conditions.

Various simpler methods have been suggested. The probably most familiar is that by van Dort et al. [2], which expresses the quantum effect by an apparent band edge shift that is a simple function of the electric field. The model is based on the expression for the lowest eigenenergy of a particle in a triangular potential and reproduces the characteristics obtained with the Schrödinger equation quite well, and thus is widely used in practice. This model does not, however, give the correct charge distribution in the device. While in principle it can be used for 2- and 3-dimensional devices of nonrectangular geometry, it is unclear how well the model is still justified in such cases.

A third alternative are density gradient (or, generalized to the level of hydrodynamics, quantum hydrodynamic models) [3], [4], [5], [6], [7]. In these models, an additional term is introduced in the continuity equations. This term contains higher-order derivatives of the potential or of the charge density. It thereby introduces a limited degree of additional nonlocality into the equations which accounts for the effect of quantum mechanics. The model is simple and its ability to describe CV-characteristics [8] and even tunneling currents [9] is well established. To implement it, the Jacobian of the system has to be modified, but it remains sparse. Furthermore, the model is multidimensional by construction.

In this paper, we describe how the density gradient model can be implemented in a multidimensional device simulator using the box method on unstructured grids. While these grids are essential to keep the number of grid points in realistic devices small, the discretization of higher order derivatives is not straightforward. The strategy we have chosen is to introduce a new variable that encapsulates the higher order derivatives, thereby making two second order differential equations out of one fourth order equation. We will also motivate and investigate a simplified density-gradient method, which turns out to give correct terminal characteristics, but fails to predict the correct charge distribution in the interior of the device.

II. THE DRIFT-DIFFUSION EQUATIONS WITH DENSITY GRADIENT CORRECTION

In [10], it is shown that in lowest order Born approximation, the current equation reads

\[ e\mathbf{j} = -\mu k_B T \nabla n - \mu n \nabla \Phi - \mu \left( \nabla \frac{n^2 \mu}{4 mk_B T} \cdot \nabla \right) \nabla \Phi, \tag{1} \]
where we already have left out time derivatives and the convective term (terms quadratic in the particle current densities $\vec{j}$). $n$ is the particle density, $T$ the temperature (in contrast to [10], we do not consider second order momentum terms and assume $T$ to be constant), $k_0$ is the Boltzmann constant, $e$ the elementary charge, $m$ the effective mass and $\mu$ the mobility. $\Phi$ is the potential, including both the band edges $\Phi_{\text{band}}$ and electrostatic potential $\phi$, $\Phi = \Phi_{\text{band}} + q\phi$, with the carrier charge denoted by $q$. $\Phi_q$ is given by [10]

$$\Phi_q[\vec{R}] = \left(\frac{2m}{\pi \hbar^2 \beta}\right)^{d/2} \int d^dR' \Phi [\vec{R} + \vec{R}'] I_d \left[\frac{2mR'^2}{\hbar^2 \beta}\right]$$

(2)

where $d$ is the dimensionality of the mesh and $\beta = 1/k_0 T$. For a 1D problem, we find

$$I_1[\alpha] = \frac{\pi (1 + 2\alpha)}{4} \text{erfc} \sqrt{\alpha} - \frac{\sqrt{\pi} \alpha}{2} \exp[-\alpha].$$

For illustration, we consider a simple example. Assume that the system is infinite, one dimensional and in equilibrium ($\vec{j} = 0$), and the potential has the form

$$\Phi[z] = \begin{cases} -\Phi_0 & \text{for } z < 0, \\ \Phi_0 & \text{for } z > 0. \end{cases}$$

Under these assumptions, Eq. (1) simplifies to

$$0 = k_0 T \frac{\partial n}{\partial z} + n \frac{\partial}{\partial z} (\Phi + Q) + \frac{\partial n}{\partial z} Q,$$

(3)

where $Q$ evaluates to

$$Q = \frac{\Phi_0}{2} \left( z \sqrt{\pi} \text{erfc} \left[ \frac{z}{\sqrt{l_{qm}}} \right] - \frac{|z|}{z} \exp \left[ -\frac{z^2}{4l_{qm}^2} \right] \right),$$

(4)

and $l_{qm} = \sqrt{\hbar^2 / 2m}$. $\Phi + Q$ is a continuous function in $z$, that is, looking only at the second term in Eq. (3), we see that the quantum correction $\Phi_q$ smoothes out the jump in $\Phi$ on a length scale $l_{qm}$, see Fig. 1.

Eq. (3) is an ordinary first order differential equation which for $z \neq 0$ can be integrated to give

$$n[z] = \frac{C_+}{k_0 T + Q[z]},$$

where $C_+$ and $C_-$ are the integration constants for $z > 0$ and $z < 0$, respectively. If $\Phi_0 > 2k_0 T$ (which violates the assumptions used to derive Eq. (1), but is common in practice), the denominator of this equation becomes 0 for a $z > 0$ and the density in the barrier diverges. Therefore, for general structures, this model can not be used unaltered (This problem does not arise for the thin barriers in [10], [11], because $Q < -k_0 T$ throughout the barrier).

An obvious remedy would be to neglect the last term in Eq. (3). This leads in our example to $n = C_+ \exp(-\beta Q)$, which is well behaved. Using this form of $n$, the last term in Eq. (3) is by a factor $|\beta \Phi_0|$ smaller than the second term. As the derivation of Eq. (1) assumes that $|\beta \Phi_0|$ is small, neglecting the last term thus is also justified in the limit where Eq. (3) is strictly valid. Another approach to improve Eq. (3) has been described in Refs. [12], [13].

However, computing Eq. (2) is still numerically too expensive to be done in actual device simulation. Using a Fast Fourier Transform [11] reduces numerical complexity, but is impractical for general, unstructured grids. We therefore follow [10] and assume $\Phi$ to vary slowly on a length scale of $l_{qm}$, and simplify Eq. (2) to $\Phi_q = \Phi / 12m$. Furthermore, in equilibrium, $\log n = -\beta \Phi + O(m \hbar^2) + \text{const.}$ Hence, the last term in Eq. (1) becomes

$$\left(\nabla \cdot \frac{\hbar^2 n}{4m k_0 T} \nabla\right) \Phi_q = \frac{\hbar^2 n}{12m} \left(\nabla \cdot \left(\nabla^2 \Phi - \beta \left[\nabla \Phi\right] + \left(\nabla O \left[\hbar^2\right]\right) \cdot \nabla \left(\nabla \Phi\right)\right)\right).$$

(5)

The explicit $\nabla O \left[\hbar^2\right]$ on the right hand side gives only a $O(m \hbar^2)$ contribution to the continuity equation and can be omitted in lowest order quantum correction. Thus, Eq. (1) simplifies to

$$e\vec{j} = -\mu k_0 T \nabla n - \mu n \nabla (\Phi + \Lambda)$$

(6)

where $\Lambda$ is given by

$$\Lambda = \frac{\hbar^2 n}{12m} \left[\nabla^2 \Phi - \beta \left(\nabla \Phi\right)^2\right].$$

(7)

The quantum effect thus shows up as an additional force term derived from of a potential-like quantity $\Lambda$ that is comprised of higher order derivatives of the classical potential. As $\Phi$ might have abrupt quantity $\Lambda$ might be undefined in certain points. We therefore replace Eq. (7) by a slightly modified partial differential equation.

Fig. 1. Potential $\Phi$, correction $Q$ according to Eq. (4), their smooth sum, and $Q_s$ according to Eq. [10]
for \( \Lambda \),

\[
\Lambda = \gamma \hbar^2 \beta \left[ \nabla^2 (\Phi + \Lambda) - \frac{\beta}{2} (\nabla \Phi + \nabla \Lambda)^2 \right].
\]

As \( \Lambda \) is \( O(\hbar^2) \), this modification, like the approximation done on Eq. (3), does not contribute in lowest order quantum correction.

For a semiconductor with multiple conduction band minima and anisotropic effective mass it is not clear which value to take for \( m \) in Eq. (8). We take \( m \) as the density of states mass and handle the problem by introducing the fit factor \( \gamma \) above.

In order to take account of position dependent effective masses, we add a mass term contribution to the potential \( \Phi \) in the same manner it appears in the classical drift-diffusion equations,

\[
\Phi = \Phi_{\text{band}} + q\phi - \frac{3k_B T}{2} \log m.
\]

The gradient square term in Eq. (8) corresponds to the last term in Eq. (3) which we found to lead to problems under certain circumstances. It is therefore tempting to replace \( \Lambda \) by the simplified form

\[
\Lambda_s = \frac{\gamma \hbar^2 \beta}{12m} \nabla^2 (\Phi + \Lambda_s).
\]

For our 1D-potential step example discussed earlier (using \( \gamma_s = 1 \)) this leads to

\[
\Lambda_s = - \frac{\Phi_0}{2} \frac{2}{|z|} \exp \left[ - \frac{\sqrt{6}|z|}{l_{\text{qm}}} \right] \exp \left[ - \frac{\sqrt{6}|z|}{l_{\text{qm}}} \right]
\]

which for small \(|z|\) is close to what we have obtained before in Eq. (4), as can be seen from Fig. 1. As we will show later, while using \( \Lambda_s \) gives reasonable results for the CV-characteristics, it does not give the correct charge distributions in the channel.

In equilibrium, \( n \propto \exp[-\beta(\Phi + \Lambda)] \), therefore Eq. (8) can also be written as

\[
\Lambda = \frac{\gamma \hbar^2}{12m} \left[ \nabla^2 \log n + \frac{1}{2} (\nabla \log n)^2 \right] = \frac{\gamma \hbar^2}{6m} \nabla^2 \sqrt{n}
\]

which are the forms often used in the literature. We will continue to use the form (8), however. This is advantageous for cases where we are not interested in transport through barriers but only in properties like CV-characteristics. For these cases, (8) allows to treat the barrier as insulating and to exclude it from the solution of the continuity equation while still being able to correctly compute \( \Lambda \).

### III. Numerics

While implementations of the quantum hydrodynamic model using finite element methods have been reported [14], [15], the method most popular for implementation of the device equations are the Box Method and Finite Difference Methods (for an introduction, see [16]). In our case, the choice of the Box Method was predetermined by the target simulation environment [17].

The usual approach to discretize Eq. (6) is to substitute \( \Lambda \) according to Eq. (11) into this equation. The order of this differential equation is thereby increased by two. This means that to discretize it in a given grid point, not only nearest neighbor grid points need be examined, but also grid points further away. On grid data structures designed for use with the Box Method, the information needed to compute these derivatives might not be available at all. Therefore, this approach is only useful on simple grid structures like tensor grids.

Our approach is to introduce \( \Lambda \) as a new variable, that is, the number of unknowns of the nonlinear system is increased. On the other hand, as Eq. (8) is only of second order, it can conveniently be discretized combining function values only on nearest neighbor grid points.

While substituting \( \Lambda \) into Eq. (6) leads to additional off-diagonal elements in the Jacobian and changes the matrix structure, our approach adds an additional block to the Jacobian, but the sparsity structure of each block still reflects the nearest-neighbor relationship of the underlying mesh.

For Eq. (6), the Scharfetter-Gummel discretization [18] is used. Eq. (8) is discretized as

\[
\Omega_i \Lambda_i = \frac{\gamma \hbar^2}{6m_i} \sum_{j} \sigma_{ij} \left( 1 - \exp \left[ \frac{\Phi_i + \Lambda_i}{2k_B T_i} - \frac{\Phi_j + \Lambda_j}{2k_B T_j} \right] \right)
\]

The indices \( i \) and \( j \) are used to designate grid points. The \( j \)-sum runs over the nearest neighbors of grid point \( i \). \( \Omega_i \) is the volume of the box for grid point \( i \), \( l_{ij} \) is the distance of grid points \( i \) and \( j \), and \( \sigma_{ij} \) is the area of the box face between these two points.

As boundary conditions for Eq. (8), we assume that \( \nabla (\Phi + \Lambda) = 0 \) at the outer boundary of the device. This would corresponds to \( \nabla n = 0 \) if the density-based formula (11) for \( \Lambda \) was used. While this boundary condition is particularly easy to implement, it is probably not ideal at contacts where the electric field can be large, for example at metallic gates.

To solve the nonlinear system of equations, we perform a Newton iteration on the entire coupled system of equations (current equations Eq. (6), Eq. (8), and Poisson equation). For all examples discussed below, the linear systems that arise in each Newton step were solved by a direct method.

To our surprise, the convergence the Newton iterations of the quantum hydrodynamic models for the examples we consider in the following section was worse than for the Schrödinger equation, even though a full Jacobian was available. Also, due to the increased number of unknowns in the nonlinear system, the time needed to obtain the eigensolutions of the 1D-Schrödinger equation was counterbalanced.
IV. Examples

A. MOS diode

Our first example is a one-dimensional MOS diode. We compare 5 different approaches: 1D-Schrödinger-Poisson solver [19], van Dort model [2], the density gradient method based on Eq. (8), the simplified density gradient method using Eq. (9), and the classical model. In all the simulations, the oxide has been treated as a true insulator and no current flow has been allowed.

The upper graph in Fig. 2 shows the CV-characteristics for a MOS-diode with 4nm gate oxide thickness, a channel doping of $5 \times 10^{17}$ cm$^{-3}$ and a metal gate, computed at 300K. The fit parameters used were $\gamma = 3.6$ and $\gamma_s = 0.3$. The graph shows that the two density gradient models and the van Dort model describe the curve well, that is, the deviation from the Schrödinger result is negligible compared to the difference from the classical result.

For the same device, we have plotted the channel density profile at a gate-back contact voltage of 4V in the bottom of Fig. 2. While the CV-curves are quite accurate, the density profile for the van Dort model and the simplified Density Gradient model (9) are far off the Schrödinger result. In particular, the density for the latter decreases much too fast when approaching the oxide interface. On the other hand the density gradient model (8) reproduces the Schrödinger density well.

We varied the oxide thickness, the channel doping, and the temperature to check if the CV-curves can be reproduced with the same choice of $\gamma$ and $\gamma_s$. Fig. 3 shows the results. It can be seen, that all the models work well for different device parameters without need to readjust the fit constants.

B. MOSFET

In Fig. 4 we show a comparison of the five models for the current-voltage characteristics of a hypothetical MOSFET with 4nm gate oxide thickness, 300nm gate length, substrate doping $10^{17}$ cm$^{-3}$, an effective channel length of 200nm and a channel width of 1um. As channel mobility models make use of the field distribution [20] and would need model-specific recalibration of their fit parameters, we restricted to a nonrealistic, constant mobility to make interpretation of the results more straightforward. We used the parameters from section IV-A, $\gamma = 3.6$ and $\gamma_s = 0.3$. The graphs show all quantization models agree well for this example, too.

C. Double-gated SOI MOSFET

To investigate the limits of the models, we modeled a hypothetical double-gated SOI MOSFET with a very thin silicon layer of 5nm and a nearly undoped channel of 50nm length. While this is not a realistic device yet, it is useful to investigate how well the density gradient method and the van Dort model agree to the results obtained by the Schrödinger approach for a structure much different from an ordinary MOSFET.

In Fig. 5 we show the drain current as a function of gate voltage (both gates are ramped together) of the device in the sub-threshold regime. The van Dort model is found to agree with the classical curve here. This can be expected, because it was not intended to be applied in this case where potential is not triangular, but basically box shaped (note that at higher gate voltages the agreement is good again because the potential near the upper and lower interface becomes triangular). The density gradient and the simplified density gradient model give the same result and both underestimate the effect of quantization. We checked that the discrepancy remains if the channel length is increased to 500nm. Thus this is not a 2D effect. The behavior of
the drain current is reflected by the density profile given in Fig. 6, which shows a cut perpendicular to the interfaces in the middle of the channel at a gate voltage $V_{GS} = 1$V.

V. SUMMARY AND CONCLUSION

We have described a multidimensional implementation of the density-gradient method which does not depend on a tensor grid to work. The main aspect of this approach is to introduce the additional variable $\Lambda$ to the nonlinear system of unknowns. The equation for $\Lambda$ can easily be discretized even on unstructured grids using the Box Method.

We compared the results obtained to those of a full Schrödinger equation based quantization model, the van Dort model and a simplified density-gradient method. While the simplified model and the van Dort model are able to reproduce CV-characteristics well, only the full density gradient method also gives the correct charge distribution in the channel.

In our numerical examples we found no advantage of our implementation of the density gradient method over the 1D-Schrödinger equation approach in terms of stability and computation time for the examples where the latter is applicable. The higher flexibility regarding device geometry and mesh are therefore the primary advantages of the density gradient method over the Schrödinger method. The density-gradient method could be also advantageous for applications where the knowledge of the full Jacobian of the nonlinear system is mandatory, like, for example, small-signal analysis.

VI. ACKNOWLEDGMENTS

We thank Mario G. Ancona and Carl L. Gardner for useful discussion.

REFERENCES

Fig. 5. Drain current density vs. gate voltage for the SOI MOSFET described in the text. The drain bias was $V_D = 0.01V$, the temperature was 300K. The inset shows a zoom of the main graph.

[Graph of drain current density vs. gate voltage]

Schrödinger
Density Gradient
simplified Density Gradient
van Dort
classical

Fig. 6. Density profile in the middle of the channel of the SOI MOSFET at a gate voltage of 1V. The classical density and the density for the van Dort model agree.

[Graph of density profile]

Andreas Wettstein received his Diploma in physics from the Universität Karlsruhe in 1995. Since then he is with the Institut für Integrierte Systeme, ETH Zürich.

Andreas Schenk was born in Berlin, Germany, in 1957. He received the Diploma in physics and the Ph.D. in theoretical physics from the Humboldt University Berlin (HUB) in 1981 and 1987, respectively. In 1987 he became a research assistant at the Department of Semiconductor Theory of HUB, and 1988 he joined the R&D division of WF Berlin. From 1987 till 1991 he was working on various aspects of the physics and simulation of optoelectronic devices, esp. infrared detector arrays, and the development and implementation of physical models for modeling infrared sensors. He is now with ETH, Zürich, Switzerland, as a senior lecturer at the Integrated Systems Laboratory. His main activities are in the development of physics-based models for simulation of submicron silicon devices. He is a member of the German Physical Society (DPG).

Wolfgang Fichtner received the Dipl.ing. degree in physics and the Ph.D. degree in electrical engineering from the Technical University of Vienna, Austria, in 1974 and 1978, respectively. From 1975 to 1978, he was an Assistant Professor in the Department of Electrical Engineering, Technical University of Vienna. From 1979 through 1985, he worked at AT&T Bell Laboratories, Murray Hill, NJ. Since 1985 he
is Professor and Head of the Integrated Systems Laboratory at the Swiss Federal Institute of Technology (ETH). In 1993, he founded ISE Integrated Systems Engineering AG, a company in the field of technology CAD. Wolfgang Fichtner is a Fellow of the IEEE and a member of the Swiss National Academy of Engineering.