Simulation of Hydrodynamic Semiconductor Device Models Using a Modified Scharfetter-Gummel Strategy

Anand L. Pardhanani and Graham F. Carey
Simulation of Hydrodynamic Semiconductor Device Models Using A Modified Scharfetter-Gummel Strategy

Anand L. Pardhanani & Graham F. Carey

Abstract

An efficient numerical solution scheme based on a new generalized finite difference discretization and iterative strategies is developed for submicron semiconductor devices. As a representative model we consider a non-parabolic hydrodynamic system. The discretization is formulated in a mapped reference domain, and incorporates a transformed Scharfetter-Gummel treatment for the current density and energy flux. This permits the use of graded, nonuniform curvilinear grids in the physical domain of interest, which has advantages when gridding irregular domain shapes or solution profiles. The solution of the discrete system is carried out in a fully-coupled, implicit form, and non-symmetric gradient-type iterative strategies are investigated. Numerical results demonstrating the performance and reliability of the scheme are presented for 1D and 2D test problems.

1 Introduction

A key requirement in the analysis and design of submicron semiconductor devices in two or three dimensions is the need to model drift-dominated transport effects reliably, accurately and efficiently. The traditional drift-diffusion equations have been very effective for larger devices, but more advanced carrier transport models such as augmented drift-diffusion, energy balance and variants of the hydrodynamic model \[2, 8, 18\] have been recently developed for submicron devices. These models are more complex, and both gridding and solution convergence are more sensitive problems. Efficient numerical simulation of these models is quite challenging, since the presence of steep gradients requires graded meshes, and the strong nonlinear coupling frequently leads to numerical oscillations, inaccuracy, instability and poor computational efficiency. These difficulties become more pronounced as device sizes shrink into the deep submicron regime.

Our primary objective in this work is to investigate new numerical methodology based on iterative and multigrid strategies on general structured grids in two-dimensions. This implies the numerical methods must be capable of handling nonuniform and curvilinear meshes, which impose the least restrictions on structured mesh design. In the present work we introduce a mapping strategy and generalize the Scharfetter-Gummel approach within this framework \[13\]. The ideas may also be directly extended to 3D and are generally applicable to a broad spectrum of advanced transport models. For clarity of exposition we focus on a representative non-parabolic variant of the hydrodynamic model which has been shown to predict velocity and energy effects more accurately in submicron devices \[3, 20\]. Our emphasis in the present work is on developing accurate, efficient and robust methods, and demonstrating their potential using relatively simple MOSFET-type model problems.
This enables us to explore the numerical formulations and solution behavior more readily, and to examine several key issues including gridding, scaling of the variables, discretization, and solution strategies.

We consider finite difference schemes for structured (but strongly nonuniform) grids, and use a mapping strategy to facilitate discretization. Fully-coupled, time-marching solution schemes are used in conjunction with iterative strategies to enhance stability and efficiency. Our use of the mapping approach for discretization is motivated by the need for curvilinear grids to accommodate a wider range of domain shapes and solution profiles in the finite difference formulation. The map introduces a mathematical transformation to a reference domain, which permits the use of more general grids in the original domain. We develop a modified Scharfetter-Gummel approach that uses the mapping strategy for discretizing the current density and energy flux in the non-parabolic hydrodynamic model. We remark that this approach can also be readily applied to the drift-diffusion analog of the hydrodynamic model (with constant energy) which we, in fact, routinely do to compute good starting iterates for the hydrodynamic model. The resulting sparse discrete systems are solved using the conjugate gradient squared (CGS) algorithm.

The rest of this paper is organized as follows: Section 2 introduces the hydrodynamic model, Section 3 outlines our numerical formulation and discretization method, Section 4 discusses the time-marching and iterative solution approach, Section 5 presents numerical results, and Section 6 offers some concluding remarks.

2 Hydrodynamic Model

In this work we use the hydrodynamic models developed at the Microelectronics Research Center at the University of Texas [1, 3]. Our emphasis is on the non-parabolic model, which is considered more physically appropriate since it is based on a more realistic energy band structure. For single-carrier devices it is mathematically described by the following set of coupled equations

\[
\nabla^2 \phi = \frac{q}{\epsilon} (n - N_D)
\]

\[
\frac{\partial n}{\partial t} + \nabla \cdot (nv) = 0
\]

\[
\frac{\partial (A(w)nv)}{\partial t} + 2 \frac{2}{3m^*} \nabla (B(w)nw) - \frac{q}{m^*} n \nabla \phi = -\frac{nv}{\tau_p}
\]

\[
\frac{\partial (nw)}{\partial t} + \nabla \cdot (\Omega nwv + Q) - qnv \cdot \nabla \phi = -\frac{n(w - w_0)}{\tau_w}
\]

where \( \phi \) is the electrostatic potential, \( n \) is the electron density, \( v \) is the drift velocity, and \( w \) is the total energy. The momentum and energy relaxation times, \( \tau_p \) and \( \tau_w \), are empirical functions, which are chosen from the work of Bordelon et al. [3] as

\[
\tau_p = \frac{0.007}{w} \times 10^{-12} \text{ s}
\]

\[
\tau_w = 0.46 \times 10^{-12} \text{ s}
\]
The system is closed by assuming a Fourier type constitutive relation for the heat flux, $Q$, of the form

$$Q = -\frac{2}{3} \left( \frac{\gamma R}{K_B} \right) n \nabla w$$  \hspace{1cm} (3)

The other quantities in equations (1) - (3) are defined as follows: $N_D(x,y)$ is given doping density, $A(w) = 1 + 2\Omega \alpha_w^m$, $B(w) = (1 + \alpha_w^m)/(1 + 2\alpha_w^m)$, $q = \text{electron charge} = 1.602 \times 10^{-19}$ C, $m^* = \text{effective mass} = 2.367 \times 10^{-31}$ kg, $\epsilon = \text{permittivity of silicon} = 11.9 \epsilon_0$, $\epsilon_0 = 8.854 \times 10^{-12}$ C$^2/(\text{joule m})$, $K_B = \text{Boltzmann constant} = 1.381 \times 10^{-23}$ joule/kelvin, $w_0 = \frac{3}{2} K_B T_L$ joule, $T_L = \text{lattice temperature in kelvin}$, and $\alpha = 0.5 eV^{-1}$, $\Omega = 1.3$, $\gamma = 4.2 \times 10^{-26}$ (watt m$^2$)/kelvin, $R = 0$ to 0.5 are empirical constants.

Note that in (1) it is also assumed that the drift kinetic energy is negligible, and the system reduces to the corresponding parabolic hydrodynamic model if we set $\alpha = 0$ and $\Omega = 3/2$.

3 \ Numerical Formulation

The first step prior to discretization is to scale the hydrodynamic system. We use the notation $t_s, x_s, \phi_s, n_s, v_s, w_s$ and $Q_s$ to denote normalizing constants for time, length, electrostatic potential, electron density, drift velocity, energy and heat flux respectively. The equations can then be written in the following non-dimensional form

$$\lambda^2 \nabla^2 \phi = (n - N_D)$$

$$\frac{\partial n}{\partial t} + C_1 \nabla \cdot (n v) = 0$$

$$\frac{\partial (A(w) n v)}{\partial t} + M_1 \frac{2}{3} \nabla (B(w) n w) - M_2 n \nabla \phi = -\frac{n v}{\tau_p}$$

$$\frac{\partial (n w)}{\partial t} + E_1 \nabla \cdot (\Omega n n w) + E_2 \nabla \cdot Q - E_3 n v \cdot \nabla \phi = -\frac{n (w - w_0)}{\tau_w}$$

$$Q = H_1 n \nabla w$$

where

$$\lambda^2 = \frac{c \phi_s}{q n_s x_s^2}, \quad C_1 = E_1 = \frac{t_s v_s}{x_s}, \quad M_1 = \frac{t_s w_s}{m^* v_s x_s}, \quad M_2 = \frac{q t_s \phi_s}{m^* v_s x_s},$$

$$E_2 = \frac{t_s Q_s}{n_s w_s x_s}, \quad E_3 = \frac{q t_s v_s \phi_s}{w_s x_s}, \quad H_1 = -\frac{2}{3} \left( \frac{\gamma R}{K_B} \right) \left( \frac{n_s w_s}{Q_s x_s} \right)$$

Note that the functions $A(w)$ and $B(w)$ are already non-dimensional by the way they are defined.

Although the scaling constants can be chosen in a variety of ways, our experience indicates that their choice has an impact on the convergence behavior and efficiency of the numerical algorithm [14]. Selberherr [16] and Markowich [9] have discussed some standard methods for choosing these constants, which we combine with numerical tests to determine their values in the present work. Table 1 summarizes typical values for the scaling constants. The main difference from the values given in [9] and [16] is in the choice of $\phi_s$. 

3
Table 1: Method of choosing scaling constants for the hydrodynamic system.

<table>
<thead>
<tr>
<th>Scaling constant</th>
<th>Selection method</th>
<th>Sample value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_s$</td>
<td>maximum domain width</td>
<td>1 micron</td>
</tr>
<tr>
<td>$n_s$</td>
<td>maximum doping density</td>
<td>$1.0 \times 10^{24} \text{ m}^{-3}$</td>
</tr>
<tr>
<td>$\phi_s$</td>
<td>maximum applied voltage</td>
<td>5 volts</td>
</tr>
<tr>
<td>$w_s$</td>
<td>$q \phi_s$</td>
<td>$8 \times 10^{-19}$ joule</td>
</tr>
<tr>
<td>$v_s$</td>
<td>$(w_s/m^*)^{1/2}$</td>
<td>$1.8 \times 10^6$ m/s</td>
</tr>
<tr>
<td>$t_s$</td>
<td>$x_s/v_s$</td>
<td>$5.5 \times 10^{-13}$ s</td>
</tr>
<tr>
<td>$Q_s$</td>
<td>$(n_s w_s x_s)/t_s$</td>
<td>$1.45 \times 10^{12}$ watt/m$^3$</td>
</tr>
</tbody>
</table>

and all the quantities that depend upon it. We found that choosing the maximum applied voltage instead of $K_BT_e/q$ gave better convergence behavior in our test cases. The general guideline we followed for scaling was to ensure that each scaled variable was bounded above by unity. However, we believe the best choice for the scaling parameters remains an open question that merits further research. The scaling issue is also closely connected to gridding questions, as the numerical impact of poor scaling appears similar to that of inadequate grid density or insufficient smoothness.

After scaling, a mathematical transformation of variables is introduced from the physical $(x, y)$ domain to a reference $(\xi, \eta)$ domain [4, 14]. This strategy yields a transformed hydrodynamic system in the mapped reference domain on which a rectangular grid can be defined. This transformation relates a curvilinear, nonuniform grid in the physical domain to a corresponding orthogonal grid in the reference domain. The transformed PDE's are then discretized in the reference domain using symmetric differencing in conjunction with a modified Scharfetter-Gummel type of scheme for the hydrodynamic model [12]. Similar schemes for finite volume discretization of the the current continuity equation in the standard hydrodynamic model have been proposed by Fukuma and Uebbing [5], Rudan and Odeh [15] and others. We have generalized this approach to allow its use in conjunction with 2D finite difference mapping strategies. The mapping approach allows greater flexibility in grid design when the domain shape is irregular or the solution contours are curvilinear. Although the mathematical formulation assumes the existence of a map, in practice it only appears in the form of transformation metric-coefficients that modify the original PDEs. These are discretized, along with the original PDEs, using finite difference approximations with respect to a uniform, rectangular reference grid. Thus the map is never explicitly constructed.

The Scharfetter-Gummel formulation is derived from the steady-state momentum conservation equations, which in the $(\xi, \eta)$ domain may be written as

$$ J = \left[ -\frac{2}{3} M_1 \tau_p \frac{\partial (B_n w)}{\partial \xi} + M_2 \tau_p n \frac{\partial \phi}{\partial \xi} \right] \nabla \xi |_{k_1} $$
where $J = n \nu = \text{current density (scaled)}$, $k_1$ and $k_2$ are unit vectors in the $\xi$- and $\eta$-direction respectively, and the other variables are defined previously. We make the usual assumptions about local behavior of the variables over a grid block: constant $J$ and $B(w)$, and linear $\phi$ and $w$. We also use the empirical relationship $\tau_p \propto \frac{1}{w}$, as described in Section 2. Equation (5) is projected along the respective $\xi$- and $\eta$-directions, and each component is integrated locally in the corresponding coordinate direction. The integration is performed along the sides of the local grid block using the nodal values of $n$, $w$ and $\phi$ for boundary conditions. This yields the following expressions for the $\xi$- and $\eta$-components of $J$

$$\frac{J_1}{|\nabla \xi|} = -\frac{2}{3} M_1 C_\tau B(w_{i+1}^{av}) \left[ \frac{n_{i+1}}{w_{i+1}} \beta(-\chi) - \frac{n_i}{w_i} \beta(\chi) \right] \frac{(w_{i+1} - w_i)}{\ln(w_{i+1}/w_i) \Delta \xi}$$

$$\frac{J_2}{|\nabla \eta|} = -\frac{2}{3} M_1 C_\tau B(w_{j+1}^{av}) \left[ \frac{n_{j+1}}{w_{j+1}} \beta(-\gamma) - \frac{n_j}{w_j} \beta(\gamma) \right] \frac{(w_{j+1} - w_j)}{\ln(w_{j+1}/w_j) \Delta \eta}$$

where

$$\chi = \left[ 2 - \frac{3}{2} \frac{M_2}{M_1 B(w_{i+1}^{av})} \right] \ln \left( \frac{w_{i+1}}{w_i} \right)$$

$$\gamma = \left[ 2 - \frac{3}{2} \frac{M_2}{M_1 B(w_{j+1}^{av})} \right] \ln \left( \frac{w_{j+1}}{w_j} \right)$$

and $\beta(x) = x/(e^x - 1)$ denotes the Bernoulli function. The other quantities introduced in (6) are

$$C_\tau = \text{scaled coefficient of } \tau_p = \frac{0.007q}{w_{st}} 10^{-12},$$

$$\Delta \xi = (\xi_{i+1} - \xi_i), \quad \Delta \eta = (\eta_{j+1} - \eta_j),$$

$$w_{i+1}^{av} = \text{average energy along } \Delta \xi = \frac{1}{2}(w_{i+1} + w_i),$$

$$w_{j+1}^{av} = \text{average energy along } \Delta \eta = \frac{1}{2}(w_{j+1} + w_j)$$

In practice it is also necessary to consider the limiting case where $w$ is locally constant ($w_i = w_{i+1}$ or $w_j = w_{j+1}$) and the expressions in (6) break down. For this case it can be shown that equations (6) reduce to

$$\frac{J_1}{|\nabla \xi|} = -\frac{2}{3} M_1 C_\tau B(w_i) \left[ \frac{n_{i+1}}{w_{i+1}} \beta(-\chi_i) - \frac{n_i}{w_i} \beta(\chi_i) \right]$$

$$\frac{J_2}{|\nabla \eta|} = -\frac{2}{3} M_1 C_\tau B(w_j) \left[ \frac{n_{j+1}}{w_{j+1}} \beta(-\gamma_i) - \frac{n_j}{w_j} \beta(\gamma_i) \right]$$

with

$$\chi_i = -\frac{3}{2} \frac{M_2}{M_1 B(w_i)} \left( \frac{\phi_{i+1} - \phi_i}{w_i} \right)$$

$$\gamma_i = -\frac{3}{2} \frac{M_2}{M_1 B(w_j)} \left( \frac{\phi_{j+1} - \phi_j}{w_j} \right)$$
A Scharfetter-Gummel type approach may also be applied for discretizing the transformed energy equation, for instance, by generalizing to the transformed case the schemes outlined by Meinerzhagen and Engl [10] and Tang [17]. In this case the energy flux is integrated analytically over local grid blocks, and the resulting expression is used in discretizing the energy equation. From equation (4), the energy flux in scaled form is

\[ S = E_1 \Omega J w + E_2 Q \]  

which can be expanded and written in the \((\xi, \eta)\) domain as

\[ S = [E_1 \Omega J_1 w + |\nabla \xi| E_2 H_1 n \frac{\partial w}{\partial \xi}] k_1 + [E_1 \Omega J_2 w + |\nabla \eta| E_2 H_1 n \frac{\partial w}{\partial \eta}] k_2 \]  

As before, we assume locally constant \(J\). In addition, \(S\) is assumed locally constant, and \(n\) is assumed to have an exponential behavior over a local grid block. Each component of equation (9) is then integrated analytically in the corresponding coordinate direction along the local grid block, and the nodal values of \(w\) are used for boundary conditions. This yields the following expressions for the \(\xi\) and \(\eta\) components of the energy flux on the sides of the grid block

\[ S_1 = D_\xi [w_{i+1} - w_i] \]
\[ S_2 = D_\eta [w_{j+1} - w_j] \]  

where

\[ D_\xi = \frac{|\nabla \xi| E_2 H_1}{\Delta \xi} \ln \left( \frac{n_{i+1}}{n_i} \right) \frac{n_{i+1}n_i}{n_{i+1} - n_i} \]
\[ D_\eta = \frac{|\nabla \eta| E_2 H_1}{\Delta \eta} \ln \left( \frac{n_{j+1}}{n_j} \right) \frac{n_{j+1}n_j}{n_{j+1} - n_j} \]

and

\[ x_s = \frac{E_1 \Omega J_1}{D_\xi}, \quad y_s = \frac{E_1 \Omega J_2}{D_\eta} \]  

Note that \(J_1\) and \(J_2\) in the above expressions correspond to those given in equations (6) - (7). In the degenerate case corresponding to the identity map, the metric coefficients are constant and we recover the standard Scharfetter-Gummel approximations.

Discretization of the current density and energy flux using equations (6) - (7) and (10) yields piecewise constant values for these quantities along the sides of each mesh element in the reference domain. These values may then be used in a standard central differencing scheme to discretize the \(\nabla \cdot J\) and \(\nabla \cdot S\) terms in the carrier continuity and energy equations. Central differencing is also used for discretizing the electrostatic potential equation. After completing the spatial discretization, we obtain a large, nonlinearly coupled system of ordinary differential equations (ODE's) in time.

We remark that this Scharfetter-Gummel treatment can also be extended to certain other more complex forms of the momentum and energy relaxation times. In particular, if the relaxation times can be written as some power of the energy, integration factors can be found to derive expressions analogous to those in (6).
4 Solution Approach

The ODE system that arises from semi-discretizing the hydrodynamic PDE system is strongly coupled and nonlinear, and we integrate it to steady-state in fully-coupled form using implicit methods. This makes it necessary to solve large systems of algebraic equations at each integration step, which can be computationally very expensive. The need for good preconditioning and iterative methods is crucial here, since they can significantly enhance the efficiency of algebraic system solution.

We consider the backward Euler method and a second-order semi-implicit Runge-Kutta method for performing the numerical integration. For convenience, we write the ODE system for \( U = [\phi, n, nu, nw]^T \) as

\[
\frac{dU}{dt} = F(U)
\]

Then, a backward Euler scheme applied to (12) yields

\[
[I - \Delta t A_n] (U_{n+1} - U_n) = \Delta t F(U_n)
\]

(13)

where \( I \) is the identity matrix, \( A_n = \frac{\partial F}{\partial U_n} \) is the Jacobian matrix, and the subscripts denote time steps. A two-stage semi-implicit Runge-Kutta method applied to (12) can be written in the form (see, for example, Lapidus and Seinfeld [7])

\[
U_{n+1} = U_n + \Delta t (\alpha_1 k_1 + \alpha_2 k_2)
\]

(14)

where

\[
\begin{align*}
  k_1 &= [F(U_n) + \Delta t a_1 A(U_n) k_1] \\
  k_2 &= [F(U_n + \Delta t b_1 k_1) + \Delta t a_2 A(U_n + \Delta t c_1 k_1) k_2]
\end{align*}
\]

(15)

Here \( \alpha_1, \alpha_2, a_1, a_2, b_1 \) and \( c_1 \) are constants that depend upon the specific integration scheme under consideration. For example, the second-order Rosenbrock version of the semi-implicit scheme used in the present work corresponds to the following choice of constants

\[
\begin{align*}
  a_1 &= a_2 = 1 - \frac{\sqrt{2}}{2}, \\
  b_1 &= \frac{\sqrt{2} - 1}{2}, \\
  c_1 &= 0, \quad \alpha_1 = 0, \quad \alpha_2 = 1
\end{align*}
\]

(16)

Note that both integration schemes require solution of linear algebraic systems during each step which, in the case of the hydrodynamic operator, are sparse, banded and nonsymmetric. The choice of iterative method depends upon the properties of the Jacobian matrix of the hydrodynamic system. Our experience indicates that the Jacobian matrix is very sensitive to the choice of grid and scaling. We have observed that minor perturbations in the grid can affect the success or failure of a given iterative scheme. In the present work the Jacobian systems are nonsymmetric, so we use the Conjugate Gradient Squared (CGS) and Bi-Conjugate Gradient (BCG) methods to solve the algebraic systems [6, 19]. The solution algorithm is set up to maximize efficiency by first trying iterative methods for solving each
algebraic system. Convergence is monitored continuously, and if it becomes unsatisfactory, the algorithm automatically switches to a band solver. Numerical tests indicate that with proper choice of grid most of the algebraic systems can be solved with the iterative methods (see Section 5 for more details).

5 Results and Discussion

Numerical studies using our approach have been carried out on a variety of graded meshes in one- and two-dimensions. The results presented in this section are computed using the non-parabolic form of the hydrodynamic model. The systems are integrated to steady state, and their corresponding solution profiles displayed. In all cases the Scharfetter-Gummel treatment is used for the current density and the energy flux. The drain voltage is applied directly, without using an incremental continuation strategy. However, our numerical studies indicate that the logarithmic expressions in the Scharfetter-Gummel form require a good starting iterate to ensure that their arguments remain positive. We provide this by first computing the solution to a constant-energy form of the hydrodynamic model, which is analogous to a drift-diffusion model. This approach is quite efficient, as the constant-energy solution typically requires less than 10% of the total computational time. When plotting $I - V$ characteristics it is used only for computing the starting iterate for the first bias point, after which the subsequent points needed for the plot constitute a natural bias continuation strategy. Our experience indicates that the constant-energy solution is a more efficient and robust starting iterate than bias continuation if the simulation is to be performed for a single point alone.

A series of 1D tests have been performed using $n^+ - n - n^+$ diode structures with channel-lengths ranging from 0.6µm down to 0.08µm, and applied drain bias up to 5 volts. Figure 1 shows simulated results for the 0.08µm structure at 3 volts bias [12]. In these calculations we use the value $R = 0.05$ in equation (3). For this case the doping density was $5 \times 10^{20}$ per cm$^3$ in the $n^+$ regions ($0 \leq x \leq 0.025$ and $0.125 \leq x \leq 0.15$), and $2 \times 10^{15}$ per cm$^3$ in the $n$ region ($0.035 \leq x \leq 0.115$), with a linear transition at the junctions. These computations were performed using a band solver, which is preferable in 1D since the matrices have small, non-sparse band-widths. In all our 1D test cases the simulation is efficient and robust, typically requiring less than one CPU minute on a Sun Sparc 1.

Two-dimensional studies have been carried out using MOSFET-type structures with gate-lengths from 0.5 - 0.6µm. In the first test case we use a simple doping profile with rectangular source and drain regions as shown in Figure 2. The grid used in this problem consists of $61 \times 41$ nodes, graded strongly toward the junction regions in the device. The non-dimensional time step (with reference to the scaling in Table 1) in this simulation was 0.01 units, and integration was carried out to steady-state, which was assumed to be reached when the absolute maximum change in the computed variables between consecutive time steps had dropped below $10^{-6}$ (non-dimensional) units. The total time required to reach steady-state depends on the initial conditions. The initial constant-energy solution is typically integrated through about 0.5 units of time, while the steady-state solution for the full hydrodynamic model generally requires 2 to 3 units of time per bias point. We use $T_L = 300K$ and $R = 0.05$ in equation (3) for these calculations. The numerical algorithm is implemented to optimize efficiency by using iterative methods for solving the linear systems.
whenever possible, and automatically reverting to the band solver if divergence is detected. For the 2D test problem, the ratio of CPU time for band solution to iterative solution is about 150 to 1. In practice we found that over 90% of the linear systems were successfully solved using the iterative method. However, we must emphasize that this is very much dependent on the grid – algebraic systems arising out of smoother grids are much more amenable to iterative solution. In fact, grid density and smoothness have a critical impact on the stability and efficiency of the numerical algorithms. Our experience underscores the importance of maintaining smooth grid transitions from low- to high-density regions over the domain.

Figure 3 shows steady-state surface plots for the carrier concentration, velocities and energy with 3 volts applied at the drain. The solution detail can be seen more clearly in a cross-sectional view of these results near the MOSFET surface, which is shown in Figure 4 for a section parallel to the oxide interface and about 40Å below it. The actual simulation was carried out over a range of drain voltages, and a plot of current versus voltage at the drain is shown in Figure 5. The calculations were performed on a Cray Y-MP, and the CPU time for computing a typical bias point on the $I_D - V_D$ curve is about 2 minutes.

Another MOSFET structure we consider has the doping profile and geometry shown in Figure 6. Here the peak donor doping in the $n$–type source and drain regions is $10^{18}$ cm$^{-3}$, and the acceptor doping in the $p$ region is $10^{16}$ cm$^{-3}$, with abrupt transition (over one mesh cell) along the junctions. The gate-length is 0.5µm. We use a numerical simulation strategy similar to the one outlined previously, using a 49 x 25 grid. Figures 7 show steady-state surface plots for electrostatic potential, carrier concentration and velocities for the constant-energy model, with applied bias $V_{DS} = 4$ volts and $V_{GS} = 1$ volt. We observe an overshoot in the x-velocity profile near the drain junction, which persists even as the solution converges to steady-state. We also observed that as the drain voltage is increased past about 3 volts, we start seeing a deterioration in computational efficiency, as the conditioning of the Jacobian...
Figure 2: Doping profile for 2D MOSFET test case (61 × 41 grid).

Figure 3: (a)-(e) Surface plots of steady-state results for 2D MOSFET test with $V_D = 3$ volts.
(b) X-velocity (cm/sec)

(c) Y-velocity (cm/sec)

Figure 3: (continued)
Figure 3: (continued)

(d) Energy (eV)

(e) Electrostatic potential (volts)
Figure 4: Projection of 2D results of Figure 3 on 1D cross-section 40Å below interface.
systems gets worse. Since the CGS iterative solver is less effective for such algebraic systems, the algorithm resorts to the band solver more frequently, which is considerably less efficient than the iterative solver.

Finally, we also performed some test runs using the parabolic form of our hydrodynamic model \((\alpha = 0 \text{ and } \Omega = 3/2)\). In general, for the same set of input conditions we saw no significant difference in convergence rate or performance between the two models.

6 Conclusion

We have developed an efficient finite-difference approach for simulating submicron semiconductor devices using a representative non-parabolic hydrodynamic model to illustrate the approach. The discretization scheme is formulated in a general framework, which permits the use of curvilinear, graded, nonuniform finite difference grids. It extends the Scharfetter-Gummel treatment for the current density and energy flux terms in the hydrodynamic model to the mapped system. It is important to note that the mapping ideas apply quite generally to other transport models and can be directly extended to 3D. We are currently working with Bell Laboratories and Stanford University on implementing some of these ideas in the PROPHET simulation platform.

Our numerical results demonstrate the efficiency and robustness of the present scheme for both 1D and 2D device structures. However, we have also shown that grid and scaling issues warrant attention, as they can have a significant impact on efficiency.

The primary challenge in grid design for submicron devices is the need to provide adequate resolution of very high gradients, while maintaining sufficient smoothness of the transition between regions of disparate grid densities. Although grid smoothness is desirable in most numerical applications, the semiconductor problem appears to be particularly
Figure 6: Doping profile and geometry for 2nd MOSFET test structure with $n$-type source and drain, $p$-type substrate.

sensitive. Adaptive grid techniques that can address these requirements are needed to minimize the frequent reliance on trial-and-error procedures. Adaptive schemes can be based on either mesh refinement or redistribution. We are currently investigating adaptive redistribution and optimization methods to handle this problem [11]. We are also developing multilevel preconditioning and continuation strategies to further improve the performance of iterative schemes. These investigations will be carried out using PROPHET, and will also expand PROPHET’s device simulation capabilities.

Acknowledgements

This research is supported by DARPA contract # DABT 63-96-C-0067.

References


Figure 7: (a)-(d) Surface plots of steady-state results for 2nd MOSFET test with $V_D = 4$ volts and $V_G = 1$ volt.
Figure 7: (continued)


