Current Transport Models for Engineering Applications

Tibor Grasser and Siegfried Selberherr
Institute for Microelectronics, TU Wien, A-1040 Wien, Austria

1. Introduction

Numerical simulation of carrier transport in semiconductor devices dates back to the famous work of Scharfetter and Gummel. Since then the transport models have been continuously refined and extended to capture more accurately transport phenomena occurring in modern semiconductor devices. The need for refinement and extension is caused primarily by the ongoing feature size reduction in state-of-the-art technology. As the supply voltages cannot be scaled accordingly without jeopardizing the circuit performance, the electric fields inside the devices have increased. Large electric fields that rapidly change over small length scales give rise to non-local and hot-carrier effects that begin to dominate device performance. An accurate description of these phenomena is required and is becoming a primary concern for industrial applications.

Traditionally, the drift-diffusion model has been used to describe carrier transport in semiconductor devices. However, the drift-diffusion model assumes equilibrium between carrier energy and electric field, which is no longer valid in modern devices. Extended models have been proposed that consider the carrier energy an independent solution variable. These models are capable of describing non-local and hot-carrier effects to first order. Recent results, however, suggest that the average energy is in many cases not sufficient for accurate modeling. Both the transport models themselves and the models for the physical parameters are affected. In this article we review the most commonly used transport models and point out their most important limitations. In addition, an extended transport model based on six moments of the distribution function is presented, which seems to be a balanced trade-off between accuracy and complexity.

2. Boltzmann’s transport equation

Transport equations used in semiconductor device simulation normally are derived from Boltzmann’s transport equation, which provides a semiclassical description of carrier transport. For a general inhomogeneous material with arbitrary band structure it reads

\[ \frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla f + \frac{1}{h^2} \mathbf{F} \cdot \nabla_k f = C[f]. \]  \hspace{1cm} (1)
Here, \( u \) is the group velocity, \( F \) the force exerted on the particles, and \( C \) the collision operator. For inclusion of quantum effects, equations based on the Wigner-Boltzmann equation have been considered.\(^6\) Boltzmann’s equation needs to be solved in the seven-dimensional phase space, which is prohibitive for engineering applications. Monte Carlo simulations have been proven to give accurate results but are restrictively time consuming. Furthermore, if the distribution of high-energy carriers is relevant, or if the carrier concentration is very low in specific regions of the device, Monte Carlo simulations tend to produce high variance in the results. Therefore, a common simplification is to investigate only some moments of the distribution function, such as the carrier concentration and the carrier temperature. These moments of the distribution function are typically defined as

\[
\langle \phi \rangle = \frac{1}{(4\pi)^3} \int \phi \, d^3k ,
\]

with a suitable weight function \( \phi = \phi(k) \).

3. The drift-diffusion model

The drift-diffusion model is the simplest current transport model that can be derived from Boltzmann’s transport equation by the method of moments\(^2\) or from basic principles of irreversible thermodynamics.\(^7\) It has been the workhorse in industrial applications for over thirty years. Within the drift-diffusion model, the well known continuity and current equations have to be solved. In their static form, these equations read

\[
\nabla J = qR ,
\]

\[
J = q\mu n E + \mu k_B T_L \nabla n .
\]

Here \( \mu \) denotes the electron mobility, \( T_L \) the lattice temperature, \( E \) the electric field, \( J \) the current density, \( R \) the recombination rate, \( n \) the carrier density and \( k_B \) the Boltzmann constant. The average energy \( w \) can be estimated via the local energy balance equation. This method neglects the lag between the electric field and the average energy characterized by the energy relaxation time. One consequence of the lag is that the maximum energy can be much smaller than the one predicted by the local energy balance equation. Furthermore, this lag gives rise to an overshoot in the carrier velocity, because the mobility depends to first order on the energy and not on the electric field.

Therefore, modeling deep-submicron devices is becoming more and more problematic. Although successful reproduction of terminal characteristics of nanoscale MOS transistors has been reported,\(^8\) the values of the material parameters used significantly violate basic physical principles. In particular, the saturation velocity \( v_S \) has to be set to more than twice the value observed in bulk measurements.
4. The full hydrodynamic model

The full hydrodynamic model was first derived by Bløtekjær. In its original form the first three moments of Boltzmann's transport equation were considered. Closure was obtained by applying a heuristic model for the heat flux using Fourier's law. Furthermore, the band structure was assumed to be parabolic and the tensor quantities were approximated by scalars. The resulting equation set reads

\[ \nabla J = q R, \quad (5) \]
\[ J - (\tau_m/q) \nabla (J \otimes J/n) = q \eta n E + \mu k_B n \nabla (n T_n), \quad (6) \]
\[ \nabla S = \rho n J - n (w - w_0) \tau_e + G_E, \quad (7) \]
\[ S = -(w + k_B T_n) J / q - \kappa (T_n) \nabla T_n; \quad (8) \]

where \( \tau_m \) is the momentum relaxation time. The additional parameters are the energy relaxation time \( \tau_e \), the thermal conductivity \( \kappa \), the energy flux \( S \), the electron temperature \( T_n \), the average energy in equilibrium \( w_0 \), and the generation rate \( G_E \). For the thermal conductivity an empirical relation analogous to the Wiedemann-Franz law is used with a correction factor \( p \):

\[ \kappa(T_n) = (5/2 - p)(k_B/q)^2 q \mu n T_n. \quad (9) \]

This equation system is similar to the Euler equations of gas dynamics with the addition of a heat conduction term and the collision terms. Thus, the electron gas has a sound speed and the electron flow may be either subsonic or supersonic. In the case of supersonic flow, electron shock waves will in general develop inside the device. These shock waves occur at either short length scales or at low temperatures. Furthermore, the traditionally applied Scharfetter-Gummel discretization scheme and its extensions cannot be used for this type of equation, which makes handling the full hydrodynamic model quite difficult.

5. The energy transport model

As the closure of the full hydrodynamic model has been shown to be problematic, the fourth moment of Boltzmann's equation is added to give a more accurate description for the energy flux \( S \). For the closure of the equation system a heated Maxwellian distribution is generally assumed. Since the resulting equation system is difficult to handle, simplifications are generally considered. The four-moments energy-transport model is obtained by the simplification of the four-moments hydrodynamic model. The convective term,

\[ (\tau_m/q) \nabla (J \otimes J/n), \quad (10) \]
in the current relation is neglected, as are the corresponding convective term in the energy flux relation and the contribution of the kinetic energy to the total carrier energy:

\[ w = \frac{mv^2}{2} + 3k_B T_e/2 \approx 3k_B T_e/2 \ . \]  

(11)

This simplification gives the four-moments energy-transport model which reads as follows:

\[ \nabla \cdot J = qR , \]  

(12)

\[ J = q\mu n E + \mu k_B \nabla (nT_e) , \]  

(13)

\[ \nabla S = E \cdot J - 3nk_B(T_e - T_L)/2\tau_E + G_E , \]  

(14)

\[ S = -(5\mu_s/2\mu) [(k_B T_e/q)J + (k_B/q^2)\mu n T_e \nabla (nT_e)] . \]  

(15)

Here the energy flux mobility \( \mu_s \) appears instead of the thermal conductivity in Eq. \( (8) \). Considering the different definitions for the mobilities, the energy-transport model is equivalent to the energy-balance model proposed by Stratton.\(^3\) See for example Ref. 12.

A comparison of the energy flux equation of the hydrodynamic model and the four-moments energy transport model shows that the correction factor \( p \) in the thermal conductivity has to be set to zero to obtain a consistent equation set. Furthermore, the ratio of the mobilities \( \mu_s/\mu \) is assumed to be unity in the hydrodynamic model.

6. Problems of the hydrodynamic and energy-transport models

During the derivation of the models given above various approximations of different severity have been employed. The most important approximations will be summarized in the following.

• **Closure**

  The method of moments transforms Boltzmann’s equation into an equivalent infinite set of equations. One of the severest approximations is the truncation to a finite number of equations (normally three or four). The equation of highest order contains the moment of the next order, which has to be suitably approximated using available information, typically the lower order moments. Even though no form of the distribution function needs to be assumed in the derivation, an implicit coupling of the highest order moment and the lower order moments is enforced by this closure. One approach to derive a suitable closure relation is to assume a distribution function and calculate the fourth order moment, where a heated Maxwellian shape is almost exclusively used. Ramaswami and Tang\(^{13}\) gave a comparison of different closure relations available in the literature.
• **Tensor quantities**

An issue that has been only vaguely dealt with is the approximation of the tensors by scalar quantities, such as the trace of the tensors. For example, the carrier mass and the carrier temperature are approximations introduced that way. One-dimensional simulations show\(^{14}\) that the longitudinal temperature component is larger than the transverse temperature component. This result indicates that the distribution function is elongated along the field direction and thus that the normally assumed equipartition of the energy is invalid. A rigorous approach has been taken by Pejcinovic *et al.*\(^ {15}\) who model four components of the temperature tensor. They observed no significant difference between the scalar temperature and the trace of the temperature tensor for ballistic diodes and bipolar transistors but a 15% difference for aggressively scaled MOSFETs in the linear region of the transfer characteristics.

• **Drift energy versus thermal energy**

Another common approximation is that the contribution of the drift energy to the total carrier energy is neglected.\(^ {16}\) As has been pointed out by Baccarani and Wordeman,\(^ {17}\) the convective energy can reach values comparable to the thermal energy. The error introduced by this approximation can be significant in the beginning of the channel where the carrier temperature is still low and a velocity overshoot is observed. This effect has been studied in detail in Ref. 18.

• **Modeling of the physical parameters: mobility and impact ionization**

The relaxation times traditionally have been derived from homogeneous field measurements or Monte Carlo simulations. For homogeneous fields there is a unique relationship between the electric field and the carrier temperature via the local energy balance equation which can be used as a definition for \(\tau_E\). From Boltzmann’s equation it is clear, however, that the relaxation times depend on the distribution function through the collision operator. Since the distribution function is not uniquely described by the average energy, models based solely on the average energy are bound to fail.

Two models for the energy dependence of the mobility are frequently used, the model after Baccarani and Wordeman,\(^ {17}\)

\[
\frac{\mu(T_0)}{\mu_0} = \frac{T_L}{T_n}, \quad (16)
\]

and the model after Hänsch\(^ {19,20}\)

\[
\frac{\mu(T_n)}{\mu_0} = \left[1 - \left(3\mu_0/2\tau_E\nu_S^2\right)\left(k_B T_n/q + 2S/5J\right)\right]^{-1}. \quad (17)
\]

For homogeneous materials \((S/J) = 5k_B T_n/2q\), which can be used to simplify Eq. (17) to

\[
\frac{\mu(T_n)}{\mu_0} = \left[1 + \left(3\mu_0 k_B/2\nu_S^2 q\right)\left(T_n-T_L\right)\right]^{-1}. \quad (18)
\]

A comparison of these three expressions with Monte Carlo simulation results for an \(n^-n-n^+\) test structure with channel length \(L_C = 200\) nm is given in Fig. 1.
The analytical expressions were evaluated using the data from the Monte Carlo simulation. Note that the temperature dependence of the inverse mobility is frequently plotted because of the expected linear dependence of Eq. (16). The small hysteresis in the simplified Hänsch model and in the Baccarani model is due to the doping dependence of the zero-field mobility $\mu_0$.

The ratio of the mobilities $\mu_S/\mu$ as a function of the carrier temperature is shown in Fig. 2 for two $n^+\!-\!n\!-\!n^+$ test structures. To obtain comparable behavior the same doping profile has been used for both structures and the bias condition has been chosen to give a maximum electric field of 100 kV/cm. Note that in commercial device simulators the mobility ratio is normally assumed to be unity.

In Figs. 3 and 4 the error of the analytical models is shown for the two $n^+\!-\!n\!-\!n^+$ test structures. As has been pointed out in Ref. 14, Eq. (17) is the only expression that gives reasonable results in both increasing and decreasing field regions. However, at the beginning of the channel where the carrier temperature is still low, the mobility is considerably over- or underestimated. Furthermore, due to the quotient of the magnitudes of two vector quantities $S$ and $J$, Eq. (17) is rather difficult to handle in a multidimensional device simulator.

As for impact ionization, it is poorly described by models that use the local average energy as the only parameter. In general, ionization rates obtained by local-energy models start rising too early and fall off too sharply. Furthermore, local-energy models considerably overestimate the ionization rates if not calibrated for the investigated device. In particular, local-energy models cannot capture impact ionization caused by hot electrons in the drain because there the cold carriers dominate the average energy, which is close to the equilibrium value.

Several non-local models have been proposed which are, however, both difficult to implement in a conventional device simulator and difficult to justify on a theoretical basis, especially for multi-dimensional problems.
Current Transport Models for Engineering Applications

7. Possible solution: The six-moments model

One can include the kurtosis $\beta_n$ of the distribution function in addition to the temperature $T_n$ without making any assumption on the shape of the distribution function except that the diffusion approximation holds.\textsuperscript{22} With the new variables (the kurtosis $\beta_n$ and the kurtosis flux $K$ along with its generation rate $G_\beta$) the static flux and balance equations of the six-moments model for electrons read

$$\nabla \cdot J = qR ,$$  \hfill (19)

$$J = q\mu E + \mu k_B \nabla (nT_n) ,$$  \hfill (20)

$$\nabla \cdot S = E \cdot J - 3nk_B(T_n - T_L)/2\tau_E + G_E ,$$  \hfill (21)

$$S = -(5\mu k_B^2/2\mu q) \mu [qEnT_n/k_B] + \nabla (nT_n^2 \beta_n) ,$$  \hfill (22)

$$\nabla \cdot K = 2qE S - 15nk_B^2(\beta_n T_n^2 - T_L^2)/4\tau_E + G_\beta ,$$  \hfill (23)

$$K = -(35\mu k_B^3/4\mu q) \mu [qEn\beta_n T_n^2/k_B + \nabla (nT_n^3 \beta_n^3)] .$$  \hfill (24)

Only one additional second-order partial differential equation for $\beta_n$ is introduced.

The additional parameters are the kurtosis relaxation time $\tau_\beta$ and the kurtosis flux mobility $\mu_K$. The solution variables are defined as

$$T_n = 3\langle \varepsilon \rangle/2k_B$$ and $$\beta_n = 3\langle \varepsilon^2 \rangle/5\langle \varepsilon \rangle^2 .$$  \hfill (25)

For a heated Maxwell-Boltzmann distribution and parabolic bands $\beta_n = \beta_{MB} = 1$. When non-parabolicity is taken into account, the value of $\beta_{MB}$ depends on the energy but stays close to unity.
The main difference to the energy-transport equations is that the kurtosis $\beta_n$ appears in the equation for the energy flux. As a consequence, the energy flux equation cannot be written in the form frequently used for energy-transport models as proportional to the current density. This modification makes the coupled equation system difficult to solve and approximations have been used. Note that the six-moments model reduces to the standard energy-transport model when the equations for $K$ are dropped and a value of unity is assumed for $\beta_n$. The kurtosis assumes values in the range $0.75–2.5$ and gives the deviation from a heated-Maxwellian distribution. This deviation is vital and can be used to formulate more accurate models for the physical parameters.

8. Application of the six-moments model: Substrate currents

For reliability issues and for the calculation of substrate currents accurate modeling of impact ionization is required. As shown in Ref. 23, the kurtosis can be used to obtain an analytical expression for the distribution function that goes beyond the Maxwellian shape approximation. With this distribution function, microscopic scattering rates traditionally used in Monte Carlo simulations can be incorporated into macroscopic device simulators. For the purpose of demonstration we considered two MOSFETs with gate lengths $L_G = 1.0 \, \mu m$ and $L_G = 0.25 \, \mu m$. A comparison of the impact ionization rate predicted by a local-energy model and by a six-moments model is given in Fig. 5 for the short-channel device. Also shown is the metallurgical junction (fat line). For the local-energy model the

![Figure 5. Comparison of the impact ionization rate as predicted by a local-energy model (LE) and by a six-moments model (SM). Also shown is the metallurgical junction (fat line).](image)
maximum occurs at the junction where the average energy rapidly decreases, because the hot carriers from the channel meet the large pool of cold carriers in the drain. In the case of the six-moments model, the maximum is inside the drain region, which is in agreement with Monte Carlo simulations.

Simulated substrate currents are given in Figs. 6 and 7 for the long-channel and short-channel devices, respectively. Both characteristics were calculated using the same parameter values. Note that the local-energy model had to be calibrated for these devices. Although the local-energy model delivers reasonable substrate currents, the calculated ionization profiles inside the devices are at the wrong position and have a wrong shape that requires individual calibration.

9. Problems of higher-order moment equations

Even though the vast majority of routinely performed device simulations still employ the drift-diffusion model, hydrodynamic and energy-transport models have been investigated thoroughly during the last ten years. Accurate results have been obtained for a large variety of devices. However, the values of the physical parameters used vary considerably.

Uncertainties are introduced by the approximation of the collision terms that are modeled via relaxation times or mobilities and by the derivation of closure relations. Expressions for the parameters normally are calibrated with homogeneous Monte Carlo simulations. As has been clearly shown, homogeneous Monte Carlo simulation data are not sufficient for the simulation of state-of-the-art devices as neither the relaxation times nor the closure relations are single-valued functions of the average energy. Unfortunately, data for inhomogeneous situations
are difficult to extract from measurements due to the complex interaction between the various parameters. Therefore, Monte Carlo simulations of \( n^+ - n - n^+ \) test-structures were performed to extract the desired data. However, the results obtained by available Monte Carlo codes differ significantly.\(^{24}\) Impurity scattering is especially difficult to model\(^{25}\) and any error in the mobility influences the simulated energy relaxation times where large differences are found in the published data.

It is particularly important to note that all models should be able to reproduce correctly the homogeneous limit. Unfortunately, model parameters calibrated to particular devices frequently do not fulfill this basic requirement, indicating that some of the underlying assumptions need to be reconsidered.

From a practical point of view it has to be pointed out that convergence problems are still an issue and inhibit the use of higher-order moment equations in everyday engineering applications. Unfortunately, simulation codes based on these equations have never reached a robustness comparable to the drift-diffusion model. One reason may be that a consistent discretization of the current equation, the energy flux equation, and the heat source term in the energy balance equation is difficult.\(^{26}\) No generally accepted scheme like the Scharfetter-Gummel\(^1\) scheme for the drift-diffusion equations exists.

Furthermore it has been shown that hydrodynamic and energy-transport models capture velocity overshoot only to first order. In general, the velocity overshoot is overestimated. Moreover, a spurious peak in the velocity at the drain end is observed which has so far not been eliminated with a unique parameter set. Comparison with Monte Carlo simulations indicate that this is a principal problem caused by the truncation of the infinite moment series.\(^27\) In addition, inaccuracies in the physical parameters, such as the mobility, amplify the spurious peaks.\(^14\)

With the ongoing reduction of feature size, the influence of quantum-mechanical effects such as confinement in the channel and tunneling currents is increasing.\(^28\) Accurate modeling of these effects still requires large amounts of computation time. In particular, the influence of the surface still can not be modeled properly. For instance, the energy relaxation time is known to differ significantly from its bulk value.

10. Conclusions

Various transport models have been proposed so far. Apart from the drift-diffusion model, higher-order models based on either Stratton's or Bløtekjær's approach have been considered. Despite its well-known limitations, the drift-diffusion model is still predominantly used in engineering applications. The need for higher-order models is well understood and these models have delivered excellent results in carefully set-up simulations. However, handling of higher-order models still requires a lot of fine-tuning and a detailed understanding of the underlying physical phenomena.
References