HYDRODYNAMIC AND ENERGY-TRANSPORT MODELS FOR SEMICONDUCTOR DEVICE SIMULATION: REVIEW OF BASIC MODELS

© Grasser T., Kosina H. and Selberherr S., 2002

As the size of state-of-the-art devices is continually reduced, non-local behavior becomes a critical issue in the simulation of these structures. The well established drift-diffusion (DD) model [1] which is still predominantly used by engineers around the world cannot cover these effects as the electron gas is assumed to be in thermal equilibrium with the lattice temperature. In the DD approach the local energy can be estimated via the homogeneous energy flux equation (e.g., (53) with $\nabla \cdot (n\mathbf{S})$ and $\partial_t$ set to zero). However, for rapidly increasing electric fields the energy lags behind the electric field because it takes the carriers some time to pick up energy from the field. A consequence of the lag is that the maximum energy can be much smaller than predicted by the homogeneous energy flux 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. As the mobility $\mu$ has not yet been reduced by the increased energy but the electric field is already large, an overshoot in the velocity $\mathbf{v} = \mu \mathbf{E}$ is observed until the carrier energy comes into equilibrium with the electric field again. Thus, DD simulations predict the same velocity profile as for slowly varying fields which can dramatically underestimate the carrier velocities. Similar to the mobility many other physical processes are more accurately described by a local energy model rather than a local electric field model. Therefore, the assumption of a fixed energy-field relation can cause non physical results when used to predict, for example impact ionization. To overcome these limitations of the DD model, extensions have been proposed which basically add
an additional balance equation for the average carrier energy \([2, 3]\). Furthermore, an additional driving term is added to the current relation which is proportional to the gradient of the carrier temperature. Several different formulations have been proposed which vary considerably in complexity. Furthermore, these equations have been extended to handle non-homogeneous materials and non-parabolicity effects. In the following we review some of the basic assumptions underlying these models.

2 Boltzmann's Transport Equation

Transport equations used in semiconductor device simulation are normally derived from Boltzmann's transport equation (BTE) which reads [4]

\[
\partial_t f + \mathbf{u} \cdot \nabla_x f + \frac{\mathbf{F}}{\hbar} \cdot \nabla_k f = C[f]
\]

for a general inhomogeneous material with arbitrary band structure [5]. For inclusion of quantum effects equations based on the Wigner-Boltzmann equation have been considered [6]. The group velocity \(\mathbf{u}\) is

\[
\mathbf{u}(\mathbf{k}, r) = \frac{1}{\hbar} \nabla_k \mathcal{E}(\mathbf{k}, r)
\]

which defines the inverse effective mass tensor

\[
\mathbf{m}^{-1}(\mathbf{k}, r) = \frac{1}{\hbar^2} \nabla_k \otimes \mathbf{u}(\mathbf{k}, r) = \frac{1}{\hbar^2} \nabla_k \otimes \nabla_k \mathcal{E}(\mathbf{k}, r)
\]

where \(\otimes\) denotes the tensor product [5]. In the following we will only consider position-independent masses but permit energy-dependent masses. Generalizations to position-dependent band structures will be given in the appropriate context. The force \(\mathbf{F}\) exerted on the particles is generally given as

\[
\mathbf{F}(\mathbf{k}, r) = -\nabla_k \mathcal{E}(\mathbf{x}) x - q(\mathbf{x}) \mathcal{E}(\mathbf{x}) + \mathbf{u} \times \mathbf{B} - \nabla_k \mathcal{E}(\mathbf{k}, r)
\]

and depends both on \(\mathbf{k}\) and \(\mathbf{r}\). Omitting the influence of \(\mathbf{u} \times \mathbf{B}\) (see [7] for a treatment of this term) and assuming homogeneous materials, \(\mathbf{F}\) simplifies to

\[
\mathbf{F}(\mathbf{r}) = -q \mathcal{E}(\mathbf{r})
\]

The BTE is an equation in the seven-dimensional phase space which is prohibitive to solve for engineering applications. Monte-Carlo (MC) simulations have been proven to give accurate results but are restrictive time consuming. Furthermore, if the distribution of high-energetic carriers is relevant, or if the carrier concentration is very low in specific regions of the device, MC 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. We define the moments of the distribution function as

\[
\langle \Phi \rangle = \frac{1}{4\pi^2} \int \Phi f \ d^2k
\]
with a suitable weight function $\Phi = \Phi(k)$. In the following we will separate the group velocity $u$ into a random part $c$ and the mean value $v = (\mathbf{u})/(1)$ as $u = c + v$. We will write all moment equations introducing the following symbols [8]

$$n = \langle 1 \rangle$$

(7)

$$p = \frac{1}{n}(\hbar k)$$

(8)

$$v = \frac{1}{n}(u) = -\frac{J}{qn}$$

(9)

$$w = \frac{1}{n}(\mathbf{e})$$

(10)

$$S = \frac{1}{n}(u\mathbf{e})$$

(11)

$$Q = \frac{1}{2n}(m(k)c^2v)$$

(12)

$$T = \frac{1}{kbn}(m(k)c \otimes c)$$

(13)

$$\hat{U} = \frac{1}{n}(\hbar u \otimes k)$$

(14)

$$\hat{R} = \frac{1}{n}(\hbar u \otimes k\mathbf{e})$$

(15)

Furthermore, we will employ an isotropic effective mass approximation via the trace of the mass tensor as [9]

$$m^{-1} = \frac{1}{3} \text{tr}((\hbar k^{-1}))$$

(16)

### 3 Band Structure

The simplest approximation for the complex band structure is a parabolic relationship between the energy and the crystal momentum

$$\mathcal{E} = \frac{\hbar^2 k^2}{2m^*}$$

(17)

which is valid for energies close to the band minimum. A first-order non-parabolic relationship was given by Kane [10] as

$$\mathcal{E}(1 + \alpha \mathcal{E}) = \frac{\hbar^2 k^2}{2m^*}$$

(18)

with $\alpha$ being the non-parabolicity correction factor. This gives the following relationship between momentum and velocity [11]

$$\hbar k = m^*(1 + 2\alpha \mathcal{E})u$$

(19)

and between energy and velocity

$$\mathcal{E} = (1 - 2\alpha m^*u^2)^{-1/2} - 1$$

(20)
which reduce to their parabolic counterparts for $\alpha = 0$. Expansion of the square root in (20) yields terms in ascending powers of velocity which are not negligible when averaged. This is problematic because these quantities are additional unknowns representing higher-order moments of the velocity distribution. Although (18) is an improvement over (17) it is nevertheless a crude approximation for real band structures at higher energies.

4 Stratton's Approach

One of the first derivations of moment equations was performed by Stratton [2]. First, the distribution function is split into its even and odd parts as

$$f(k, r) = f_0(k, r) + f_1(k, r)$$

(21)

From $f_1(-k, r) = -f_1(k, r)$ it follows that $\langle f_1 \rangle = 0$. Assuming that the collision operator $C$ is linear and invoking a microscopic relaxation time approximation for the collision operator

$$C[f] = -\frac{f - f_0}{\tau(\mathcal{E}, r)}$$

(22)

the BTE can be split into two coupled equations. In particular, $f_1$ is related to $f_0$ via

$$f_1 = -\tau(\mathcal{E}, r)\left(\mathbf{u} \nabla \mathcal{E} f_0 - \frac{q}{h^2} \mathbf{E} \cdot \mathbf{k} f_0 \right)$$

(23)

The microscopic relaxation time is then expressed using a power law as

$$\tau(\mathcal{E}) = \tau_0 \left(\frac{\mathcal{E}}{k_B T_n}\right)^{-p}$$

(24)

which allows for an explicit integration over constant energy surfaces. When $f_0$ is assumed to be a heated Maxwellian distribution, the following equation system is obtained

$$\nabla \cdot J = q(\delta_n + R)$$

(25)

$$J = q\mu n \mathbf{E} + k_B \nabla (\nu_n T_n)$$

$$\nabla \cdot S = \frac{3}{2} k_B q(\delta_n T_n) + E \cdot J - \frac{3}{2} k_B q n \frac{T_a - T_b}{\tau_c}$$

(27)

$$S = -\left(\frac{5}{2} - p\right) \left(\mu n k_B T_n + \frac{k_B}{q} \nabla (\nu_n T_n)\right)$$

(28)

(26) is frequently written as

$$J = q\mu \left(\mathbf{n} \mathbf{E} + k_B T_n \nabla n + \frac{k_B}{q} n (1 + \nu_n) \nabla T_n\right)$$

(29)

with

$$\nu_n = \frac{T_a}{\mu} \frac{\partial \mu}{\partial T_n} = \frac{\partial \ln \mu}{\partial \ln T_n}$$

(30)

which is commonly used as a fit parameter with values $-0.5, ..., -1.0$. For $\nu_n = -1.0$, the thermal diffusion term disappears. Under certain assumptions [2, 12] $p = -\nu_n$. The problem with expression (24) for $\tau$ is that $p$ must be approximated by an average value to cover the relevant scattering processes. However, this average depends on the doping profile and the applied field and thus no unique value for $p$ can be given.
5 Blotekjaer's Approach

Blotekjær [3] derived conservation equations by taking the moments of the BTE using the weight functions 1, $hammer_{k}$, and $E$ without imposing any assumptions on the form of the distribution function. These weight functions $\Phi$ define the moments of zeroth, first, and second order. The resulting moment equations can be written as follows [8]

\[
\begin{align*}
\partial_t n + \nabla \cdot (nv) &= nC_n \\
\partial_t (nv) + \nabla \cdot (n\mathbf{v}) - nF &= nC_P \\
\partial_t (n\mathbf{v}) + \nabla \cdot (n\mathbf{S}) - n\mathbf{F} &= nC_E
\end{align*}
\]

(31)  \hspace{2cm} (32)  \hspace{2cm} (33)

Note that these expressions are valid for arbitrary band structures, provided that the carrier mass is position-independent. When $F$ is allowed to be position-dependent, additional force terms appear in (31)-(33) [13]. The collision terms are usually modeled with a macroscopic relaxation time approximation as

\[
C_n = -\frac{1}{n}(R - G) = -\frac{1}{n}U
\]

(34)

\[
C_p = -\frac{P}{\tau_p}
\]

(35)

\[
C_E = -\frac{w - w_0}{\tau_E}
\]

(36)

which introduces the relaxation times $\tau_p$ and $\tau_E$. A discussion on this approximation is given in [14]. This equation set is not closed as it contains more unknowns than equations. Closure relations have to be found to express the equations in terms of the unknowns $n$, $\mathbf{v}$, and $w$. Due to the strong scattering the temperature tensor is normally assumed to be isotropic and is approximated by a scalar $T_n$ as

\[
\mathbf{T} \approx T_n I = \frac{T_{xx} + T_{yy} + T_{zz}}{3}
\]

(37)

Traditionally, parabolic bands were assumed which gives the following closure relations for $p$, $\hat{u}$, and $w$

\[
p = m^* \mathbf{v}
\]

(38)

\[
\hat{u} = \frac{m^*}{n}(u \otimes u) = k_B T_n I + m^* \mathbf{v} \otimes \mathbf{v}
\]

(39)

\[
w = \frac{3}{2} k_B T_n + \frac{m^* v^2}{2}
\]

(40)

Note that the random component of the velocity has zero average $\langle \mathbf{c} \rangle = 0$. With (38) one obtains the following formulation for $C_p$

\[
C_p = -\frac{P}{\tau_p} = -\frac{m^* \mathbf{v}}{\tau_p} = -\frac{\mathbf{F}}{\mu}
\]

(41)

For modeling purposes it is advantageous to lump $m^*$ and $\tau_p$ into one new parameter, the mobility $\mu$. As signal frequencies are well below $1/(2 \pi \tau_p) \approx 10^{12}$ Hz the time derivative in (32) can safely be neglected.
Furthermore, a suitable approximation for the energy flux density $nS$ has to be found and different approaches have been published. Blotevogel used

$$nS = (w + k_BT_n)n\nu + nQ$$  \hspace{1cm} (42)$$

and approximated the heat flux $nQ$ by Fourier's law as

$$nQ = -\kappa(T_n)\nabla T_n$$  \hspace{1cm} (43)$$
in which the thermal conductivity is given by the Wiedemann-Franz law as

$$\kappa(T_n) = \left(\frac{5}{2} - p\right)\left(\frac{k_B^2}{q}\right)\rho \mu T_n$$  \hspace{1cm} (44)$$

where $p$ is a correction factor. As has been pointed out [8], this expression is problematic as (43) only approximates the diffusive component of $nQ$. For a uniform temperature $\nabla T_n = 0$ and thus $Q = 0$ which is not plausible. The convective component $Q_{conv}$ must be included to obtain physical results when the current flow is not negligible.

With these approximations (31)-(33) can be written in the usual variables as [15]

$$\nabla \cdot J = q(\delta n + U)$$  \hspace{1cm} (45)$$

$$\frac{\tau_p}{q} \nabla \cdot \left( J \otimes J \right) = \mu k_BT_n(nT_n) + qn\rho \nu$$  \hspace{1cm} (46)$$

$$\nabla \cdot (nS) = -\partial_t(n\nu) + \nu \cdot J - \frac{n\nu - w_b}{\tau_e}$$  \hspace{1cm} (47)$$

$$nS = \frac{1}{q}(w + k_BT_n)J - \kappa(T_n)\nabla T_n$$  \hspace{1cm} (48)$$
to give the full hydrodynamic model (FHD) for parabolic band structures. This equation system is similar to the Euler equations of gas dynamics with the addition of a heat conduction term and the collision terms. It describes the propagation of electrons in a semiconductor device as the flow of a compressible, charged fluid. This electron gas has a sound speed $v_s = \sqrt{T_n/m}$ and the electron flow may be either subsonic or supersonic. With $T_n = \xi T_L$ and $T_L = 300 \text{ K}$, $v_s = \sqrt{\xi 1.3 \cdot 10^5 \text{ cm/s}}$ while for $T_L = 77 \text{ K}$, $v_s = \sqrt{\xi 6.6 \cdot 10^6 \text{ cm/s}}$ [16].

In the case of supersonic flow, electron shock waves will in general develop inside the device [16]. These shock waves occur at either short length scales or at low temperatures. As the equation system is hyperbolic in the supersonic regions, special hyperbolic methods have to be used [16, 17, 18, 19]. Furthermore, the traditionally applied Scharfetter-Gummel [20] discretization scheme and its extension to the energy-balance and energy-transport models [21, 22, 23, 24] cannot be used for this type of equation. One approximation is to treat the convective term as a perturbation by freezing its dependence on the state variables at each linearization step and using the values from the last iteration [25]. However, this approach will degrade the convergence in cases where the variation in space or time is important [26]. Thus, to derive a spatial discretization, fluid dynamics methods known as upwinding are used [26]. Furthermore, the handling of the boundary conditions becomes more difficult [19, 27].

When the convective term

$$\frac{\tau_p}{q} \nabla \cdot \left( J \otimes J \right)$$  \hspace{1cm} (49)$$
is neglected, a parabolic equation system is obtained which only covers the subsonic flow regions. This is a very common approximation in today's device simulators. Furthermore, the contribution of the velocity to the carrier energy is frequently neglected

$$ u \approx \frac{3}{2} k_b T_n $$  \hspace{1cm} (50) $$

which then results in the following equation system

$$ \nabla \cdot \mathbf{J} = q (\partial_t n + U) $$  \hspace{1cm} (51) $$

$$ \mathbf{J} = \mu_0 k_b \nabla (nT_n) + qn \mu E $$  \hspace{1cm} (52) $$

$$ \nabla \cdot (n \mathbf{S}) = -\frac{3k_b}{2} \partial_t (nT_n) + E \cdot \mathbf{J} - n \frac{3k_b T_n - T_e}{2} $$  \hspace{1cm} (53) $$

$$ n \mathbf{S} = \frac{5k_b T_n}{2q} \mathbf{J} - \kappa(T_n) \nabla T_n $$  \hspace{1cm} (54) $$

(51)-(54) form a typical three moment energy-transport (ET) model which has been closed using Fourier's law.

To overcome the difficulties associated with the Fourier law closure (43), the fourth moment of the BTE has been taken into account [28] which gives

$$ \nabla \cdot (n \mathbf{F}) - n (u \mathbf{i} + \mathbf{U}) \cdot \mathbf{F} = n C_{pe} $$  \hspace{1cm} (55) $$

where the time derivative has been ignored using a similar argument to (32). The collision term in (55) can be modeled in analogy to (41) as

$$ C_{pe} = -\frac{qS}{\mu S} $$  \hspace{1cm} (56) $$

which gives

$$ \mathbf{S} = \frac{\mu S}{\mu_0} (u \mathbf{i} + \mathbf{U}) \cdot \mathbf{v} + \frac{\mu S}{\mu_0} \left( (u \mathbf{i} + \mathbf{U}) \cdot \nabla (n \mathbf{U}) - \nabla (n \mathbf{F}) \right) $$  \hspace{1cm} (57) $$

Now a closure relation for $\mathbf{F}$ has to be introduced, which can be, for example, obtained by assuming a heated Maxwellian distribution which gives

$$ \mathbf{F} = \frac{5}{2} k_b T_n \mathbf{i} $$  \hspace{1cm} (58) $$

Using closure (58) and the same approximations that led to the three moments ET model (51)-(54), a more accurate expression for $n \mathbf{S}$ is obtained from the fourth moment of the BTE

$$ n \mathbf{S} = -\frac{\mu S}{\mu} \frac{5k_b T_n}{2q} \mathbf{J} - \frac{\mu S}{\mu} \left( \frac{k_b}{q} \right)^2 qm T_n \nabla T_n $$  \hspace{1cm} (59) $$

which should be used to replace (54) to give a four moments ET model. Comparing (59) with (54) reveals that a consistent three moment ET model can be obtained with $\mu S/\mu = 1$ and $p = 0$. However, $\mu S/\mu$ strongly depends on the carrier temperature and shows a pronounced hysteresis as shown in Fig. 1 where the points B and D are from the rising and decreasing temperature regions, respectively. The energy relaxation time and the momentum relaxation time are shown in Fig. 2 and both are not single valued functions of the temperature.
6 Comparison

One of the extensively discussed differences between Blattekjaer's (A1) and Stratton's (A2) approach is that in A2 the mobility stands inside the gradient whereas in A1 it stands in front of the gradient in the current relation

\[
\text{Blattekjaer (A1): } \mu_1 \nabla (nT_n) \\
\text{Stratton (A2): } \nabla (n \mu_2 T_n)
\]

This issue was addressed by Stratton himself [29] and by Landsberg [30, 31]. It is important to note, that although this parameter is called mobility in both approaches, their definition differs significantly. Tang and Gan [32] compared both approaches and found that both formulations are justified, provided that the respective mobilities are modeled accordingly. For bulk simulations the mobilities are equal and can be properly modeled using conventional energy-dependent expressions [33, 34]. However, in inhomogeneous samples where the electric field varies rapidly, the mobilities are no longer single-valued functions of the average carrier energy. The advantage of the \( \mu_1 \) formulation lies in the fact, that for increasing values of the electric field, it can be roughly approximated by its bulk value whereas \( \mu_2 \) is always different. Thus \( \mu_1 \) can be expected to be more suitable because in most commercial simulators the mobility is modeled as a function of the carrier energy only. By expressing \( C_p \) empirically as

\[
C_p = C_{p}^* + \lambda_p \nabla \cdot \dot{U}
\]

(60)

where \( C_{p}^* \) is the homogeneous component and \( \lambda_p \) a dimensionless transport coefficient. Tang et al. [35] showed that A2 can be obtained from A1 with \( \lambda_p = -\nu \). Other comparisons of the two approaches can be found in [12, 32, 36, 37, 38].
7 Conclusions

Many different hydrodynamic and energy-transport models have been published so far. They rely on either Stratton’s or Blotekjær’s approach to find a suitable set of balance and flux equations. In Stratton’s approach there is no need to invoke Fourier’s law to close the equation system due to the relationship (23). Blotekjær used only three moments and closed the equation system by approximating the heat flux with Fourier’s law. This closure has frequently been replaced by equations obtained from the fourth moment of the BTE.

Uncertainties are introduced by the approximation of the collision terms which are modeled via relaxation times and by the derivation of closure relations. Expression for these are normally extracted from homogeneous MC simulations. As has been clearly shown, homogeneous MC 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. This used to be one of the advantages of the macroscopic transport models over the MC method because measured $\mu(E)$ characteristics could be directly incorporated into the simulation which is not possible for the microscopic approach taken in the MC method. Unfortunately, data for inhomogeneous situations are difficult to extract from measurements due to the complex interaction between the various parameters. Therefore, MC simulations of $n^+ - n - n^+$ test-structure were performed to extract the desired data.

Another problem is directly related to the MC simulations itself. As has been frequently reported, the results obtained by available MC codes differ significantly [39]. Especially impurity scattering is difficult to model [40] and any error in the mobility influences the simulated energy relaxation times were large differences were found in the published data.

References

15. K. Souissi, F. Odeh, H.H.K. Tang, and A. Gruzi, “Comparative Studies of